

Proceedings e report

87

Proceedings ECLAP 2012

Conference on Information Technologies
for Performing Arts, Media Access and
Entertainment

Florence, Italy
7 – 9 May 2012

edited by
PAOLO NESI
RAFFAELLA SANTUCCI

Supported by
ECLAP partners
European Commission

FIRENZE UNIVERSITY PRESS
2012

Proceedings ECLAP 2012 : Conference on Information Technologies for Performing Arts, Media Access and Entertainment : Florence 7-9 may 2012 / edited by Paolo Nesi, Raffaella Santucci. – Firenze : Firenze University Press, 2012.
(Proceedings e report ; 87)

<http://digital.casalini.it/9788866551287>

ISBN 978-88-6655-125-6 (print)
ISBN 978-88-6655-128-7 (online)

Peer Review Process

All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published on the website and in the online catalogue of the FUP (<http://www.fupress.com>).

Firenze University Press Editorial Board

G. Nigro (Coordinatore), M.T. Bartoli, M. Boddi, F. Cambi, R. Casalbuoni, C. Ciappei, R. Del Punta, A. Dolfi, V. Fargion, S. Ferrone, M. Garzaniti, P. Guarnieri, G. Mari, M. Marini, M. Verga, A. Zorzi.

© 2012 Firenze University Press
Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
<http://www.fupress.com/>
Printed in Italy



Supported by:

European Commission



ECLAP partners:

Department of Systems and Informatics, University of Florence, Italy, (coordinator)
Axmediatech Srl, Italy
Sound & Vision, Netherlands Institute for Sound and Vision, The Netherlands
Dario Fo & Franca Rame Archive, Italy
Escola Superior de Música e das Artes do Espectáculo do Porto, Portugal
Festival International de Films de Femmes de Créteil, France
Fondazione Rinascimento Digitale, Italy
The Institute of Polish Culture University of Warsaw, Poland
Museu de les Arts Escèniques Institut del Teatre de Barcelona, Spain
La Bellone, House of Performing Arts, Belgium
MUZEUM, Ljubljana, Slovenia
National Technical University, of Athens, Greece
Hungarian Theatre Institute, Hungary
Museum of Archaeology & Anthropology, University of Cambridge, UK
Universidad de Castilla La Mancha, Spain
History of Art Department at the University of Glasgow, UK
Centro Teatro Ateneo, Department of Performing Arts, University of Rome La Sapienza, Italy
Department of Theatre Studies, University of Amsterdam, The Netherlands
ARCHIBRAILLE - Associazione Culturale ArchiBraille, Italy
Fondazione Fabbrica Europa per le Arti Contemporanee, Italy
Grotowski Institute, Poland
The National Audiovisual Institute, Poland
DigiLab - Centre for Research in Digital Arts and Humanities, Italy
Irtem - Istituto di Ricerca per il Teatro Musicale, Italy
Accademia Nazionale di Santa Cecilia, Italy
Teatro Napoletano, Università degli Studi di Salerno, Italy

Table of Contents

Message from the Chairs	XI
Conference Organising Committee	XIV
International Program Committee	XV
<hr/>	
Keynote Speakers	
Cultural Heritage and Future Directions <i>Marco Marsella</i>	3
Making Siobhan Davies RePlay; the UK's first digital dance archive. The pleasure of preserving intangible cultural heritage <i>Sarah Whatley</i>	4
'You can't get there from here': Re-thinking the maps in a digital archive <i>Heather Nathans</i>	8
The search for sustainability in the cultural heritage domain - Building a knowledge-based infrastructure <i>Nikki Timmermans</i>	12
<hr/>	
General track on Performing Arts, Media Access and Entertainment	
A Model Driven Architecture approach for Open Archival Information Systems <i>Walter Allasia</i>	21
3D reconstructions and Cultural Heritage communication <i>Micaela Spigarolo, Maria Chiara Liguori, Antonella Guidazzoli</i>	29
Museum outside the Museum. Real environment Storytelling <i>Stefania Chipa, Gianluca Simonetta, Luca Toschi</i>	35
Curating Data, Disseminating Knowledge: Museums of the Digital Age <i>Sudeshna Guha</i>	41
MIMESIS: A Semantic-based Approach for Poetry in Music Digital Libraries <i>Ilaria Bartolini, Cecilia Luzzi, Angelo Pompilio, Marco Patella and Cesarino Ruini</i>	46
Indexing and Searching Cross Media Content in a Social Network <i>Pierfrancesco Bellini, Daniele Cenni, Paolo Nesi</i>	52
Interactive magazine about new technologies meeting culture worldwide <i>Antonella Fresa, Tania Masi, Claudio Prandoni, Valentina Bachi</i>	58
The Co-Creation of the City. re-programming cities using real-time user generated content <i>Salvatore Iaconesi, Oriana Persico</i>	62

Workshop: Educational Services for the Performing Arts

- Performance art studies at the dawn of the digital age
Dorota Sosnowska 71
- Open Networks for Culture and Education: ONCE
John Casey, Wolfgang Greller, Gareth Waller 75
- Knowledge Management in Performing Arts Education
Peter G.F. Eversmann 81

Workshop: New business models in the cultural heritage domain

- Model, format and services for the management of information on audiovisual rights
Laurent Boch, Annarita Di Carlo, Francesco Gallo 89
- Audiovisual Media Open Educational Resources. The case of SP-ARK: The Sally Potter Film Archive
Sarah Atkinson, Janko Čalić 95

Workshop: Performing Arts & ICT

- Languages for Augmented Choreography
Andrea Bene, Diego Bernini, Giorgio De Michelis and Francesco Tisato 101
- mConduct: Gesture Transmission and Reconstruction for Distributed Performance
Phoebe Bakanas, Joanne Armitage, Joel Balmer, Paul Halpin, Kyle Hudspeth, Kia Ng 107
- Mobile Augmented Reality for cultural dissemination
Raffaello Brondi, Marcello Carrozzino, Franco Tecchia, Massimo Bergamasco 113

Workshop: Authoritative Metadata Vs User-Generated Content

- A Database for Israeli Theater Archives In The University of Haifa Younas & Soraya Nazarian Library
Sharon Shapira 121
- Designing and implementing a user-generated content service for Europeana
Nicola Aloia, Cesare Concordia, Carlo Meghini 126
- The Open Data Semantics and the (re)use of open information in Cultural Heritage
Achille Felicetti, Andrea D'Andrea, Franco Niccolucci 131
- SKOSifying User Generated Content
Alessandra Donnini, Andrea Ciapetti, Maria De Vizia Guerriero, Matteo Lorenzini, Maria Emilia Masci, Davide Merlitti, Fabio Piro, Oreste Signore 137
- Making Europe's cultural treasures accessible: The Europeana Regia and Europeana Libraries projects
Gabriele Messmer 143

Workshop: Open Access 2002-2012

- Visions and Possibilities of Open Access Publishing in the Dissemination of the Hungarian Performing Arts Heritage
Attila Szabó, Piroska Ács 149
- The White Road to Open Access: A Proposal by the Open Access Publishing Workshop of the University of Florence
Beatrice Tottossy, Arianna Antonielli 155
- Two sides of the same coin: access to archival collections, a case study on HOPE
Gabriella Ivacs, Kathryn Mathe 162
- The Co-Laboratory: a tool for research and education at the University of Parma
Anna Maria Tammaro, Sara Valla, Elisabetta Longhi 166
- Opening up culture may not be sufficient: looking for a policy
Rosanna De Rosa 169
- Aggregation as a Model for a Sustainable Economy? The Case of Cultural Journals
Rosario Garra 175

Poster Session

- Nuova Consonanza Archive and short films of the Sixties and Seventies: an example of low cost preservation
Cecilia Montanaro and Laura Nicoletta Colabianchi 183
- The Documentation Centre and Museum of the Performing Arts
Mireia Roselló Gargallo, Roger Guasch Aramburu 187
- Rediscovery of the shape of artistic artefacts from moulds of Richard-Ginori collection: tools and methods for 3D archive implementation
Sergio di Tondo, Giulia Adembri, Monica Gherardelli 189

-
- Author Index**
- 191

Messages from the Chairs

Welcome to ECLAP 2012, the first international conference on Information Technologies for Performing Arts, Media Access and Entertainment.

Information Technologies have made possible many important changes in the field of cultural heritage and continue to provide dynamic and exciting media platforms through which new possibilities perpetually emerge. This wave of change has had particularly significant consequences in the field of the Performing Arts, where a vast array of possibilities for digital content fruition continues to reveal itself, constantly opening the doors to new and as-yet-unexplored synergies. Many technological developments concerning digital libraries, media entertainment and education are now fully developed and ready to be exported, applied, utilised and cultivated by the public.

ECLAP is a best practice network co-funded through the ICT Policy Support Programme of the European Commission. ECLAP's goal is to enable digital access to Performing-Arts resources, while providing guidelines and using metadata standards for searching and browsing. By creating a seamless and centralised online database, ECLAP is providing access to the Performing-Arts collections and archives of its project partners, amongst which are many of the leading institutions in the field. An ever-growing part of these resources is becoming accessible through a common, multilingual, easy-to-use ECLAP e-Library for the Performing Arts. The ECLAP metadata will become part of Europeana, the European multi-lingual online collection of millions of digitized items from museums, libraries, archives and multimedia collections.

The ECLAP 2012 conference aims to function as a forum in which progress-oriented individuals and institutions find a place to collaborate and present results. It also aims to provide an overview of the state of the art for Performing-Arts digital collections within the framework of the following best-practice themes: digital library tools, education and research facilities, IPR issues, cultural heritage and technologies.

The event consists of a set of workshops, sessions and panels that conform to our standard of excellence. We host a keynote-speaker lineup consisting of some of the most salient voices in the field: Marco Marsella, Deputy Head of Unit, eContent and Safer Internet, European Commission, Information Society and Media Directorate-General; Heather Nathans, Professor of Theatre Studies, School of Theatre, Dance and Performance Studies, University of Maryland, USA; Nikki Timmermans, Social Innovator, Kennisland, The Netherlands; Sarah Whatley, Director of the Centre for Dance Research (C-DaRE) at Coventry University, UK.

An international committee selected the best proposals. 36% of the submitted contributions have been accepted as full research papers. ECLAP 2012 features more than 40 presentations, coming from several countries: United States, Israel, United Kingdom, Netherlands, Italy, Spain, Germany, Hungary, etc.

The conference is open to researchers, professionals, industries, institutions, technicians, practitioners in the area of performing arts and information technologies, media-based entertainment, technology-enhanced learning, intelligent-media systems, acoustic systems, cultural heritage.

The ECLAP 2012 conference aims to become a place where institutions, industries, the European Commission and Europeana family projects in the areas of cultural heritage can find plenty of opportunities for networking, debating, sharing ideas as well as best practices.

As general chair, it is a pleasure to express my gratitude to the dedicated program co-chairs, workshop co-chairs, committee members and conference support staff who have contributed to making ECLAP a success. We hope that you will find the conference an exciting moment for exchanging research ideas and that you will also find the time to appreciate Florence, the wonderful location of the conference.

Paolo Nesi
ECLAP Chair

From our work on ECLAP as well as from examining the proposals that were sent in for the ECLAP Conference, it was apparent to us that an interdisciplinary approach to the application of IT to the Performing Arts is still both a need and a challenge. It is difficult, in fact, to meaningfully study all the ramifications of the sometimes troublesome marriage of IT with Performing Arts. On the one hand, the risk is that humanities scholars ignore, overlook or oversimplify technical issues; on the other hand, IT people are not necessarily aware of the problems and of the needs that are specific of the Performing Arts.

The ideal profile in this respect would be that of a person who combines a solid scientific background with considerable experience in the Performing Arts. However, regrettably, people with such an uncommon admixture of skills are still a rare find.

Our hope is that gatherings such as the ECLAP 2012 Conference will further promote dialogue between specialists in the different fields, as well as future fruitful collaborations. Other stated goals of the Conference are the promotion of networking and of knowledge transfer between the various EC projects belonging to the Europeana family.

I wish to thank those who enriched the ECLAP 2012 Conference through their contributions. Likewise, I wish to thank all those who made the conference possible through their dedication and work.

Raffaella Santucci

ECLAP Networking Coordinator

Performing arts are a fundamental facet of our shared European identity. This expression designates a plethora of human activities pertaining to our traditions, which share the common trait of having been usually transmitted orally, or by imitation, and recorded only scantily and/or occasionally, usually by theatre and music historians or anthropologists: examples include folk tales, traditional music, dance, popular festivities (religious or non religious), all forms of non-text-based performance.

The importance of safeguarding performing arts heritage is now acknowledged widely, and ratified by the latest workprogrammes of the European Commission.

The vast body of knowledge which characterises European performing arts is testament to the variety and specificity of European cultures. However, performing arts heritage is in danger of being lost, due to its ephemeral quality. Thanks to ECLAP the public can now reap the benefits that stem from being able to access to a vast unified repository.

We hope that ECLAP and ECLAP 2012 conference will work as the springboard for starting the digitization of other material in the field of performing arts, thus establishing a best practice example for the gradual future digital ingestion of their whole audiovisual heritage and advancing the Europeana's ongoing expansion by providing a critical mass of content pertaining to the area of performing arts.

Prof. Valentina Valentini
Centro Teatro Ateneo, Sapienza, University of Rome
ECLAP Content Coordinator Partner

Conference Organizing Committee

General chair

Paolo Nesi

DSI-DISIT, University of Florence, Italy

nesi@dsi.unifi.it, www.dsi.unifi.it/~nesi

Local Organization co-chairs

Paolo Nesi (ECLAP coordinator), University of Florence, IT

Ferruccio Marotti, University of Roma La Sapienza, IT

Pierfrancesco Bellini, University of Florence, IT

Alessandro Bernardi, University of Florence, IT

Beatrice Tottossy, University of Florence, IT

Fabrizio Argenti, University of Florence, IT

Workshops coordinator and networking

Raffaella Santucci, University of Roma La Sapienza, IT

Program co-chairs

Emanuele Bellini, Fondazione Rinascimento Digitale, IT

Peter Eversmann, University of Amsterdam, NL

Kia Ng, University of Leeds, UK

Johan Oomen, Institute for Sound and Vision, NL

Pierfrancesco Bellini, University of Florence, IT

International Programme Committee

Ad Aerts, NL
Fabrizio Argenti, University of Florence, IT
Jerome Barthelemy, IRCAM, FR
Emanuele Bellini, FRD, IT
Pierfrancesco Bellini, University of Florence, IT
Alessandro Bernardi, University of Florence, IT
Robin Boast, University of Cambridge, UK
Luciana Bordoni, ENEA, IT
Luigi Briguglio, Engineering Ingegneria, IT
Monica Carfagni, University of Florence, IT
Vittore Casarosa, CNR-ISTI, IT
Donatella Castelli, CNR-ISTI, IT
Tiziana Catarci, Sapienza, University of Rome, IT
ShiKuo Chang, University of Pittsburgh, USA
Martin Doerr, Center for Cultural Informatics FORTH, GR
Nasos Drosopoulos, NTUA, GR
Peter Eversmann, University of Amsterdam, NL
Antonella Fresa, PROMOTER, IT
Mariella Guercio, University of Urbino, IT
Susan Hazan, The Israel Museum, Jerusalem, IL
Geertje Jacobs, Rijksmuseum Amsterdam, NL
Anra Kennedy, Culture 24, UK
Yevgeniya Kovalchuk, University of Essex, UK
Nick Lambert, University of London, UK
Carlo Marinelli, IRTEM, IT
Ferruccio Marotti, Sapienza, University of Rome, IT
Michal Masa, CIANT, CZ
Carlo Meghini, ISTI-CNR, IT
Paolo Nesi, University of Florence, IT
Kia Ng, University of Leeds, UK
Johan Oomen, Sound&Vision, NL
Laurence Pearce, Xim Ltd., UK
Seamus Ross, University of Toronto, CA
Raffaella Santucci, Sapienza, University of Rome, IT
Josefien Schuurman, University of Amsterdam, NL
Jorge Solis, Karlstad University, SE
Beatrice Tottosy, University of Florence, IT
Yannis Tzitzikas, FORTH-ICS, GR
Seth van Hooland, Dublin Core, Université Libre de Bruxelles-ULB, BE
Kara Van Malsen, New York University, USA
Tillman Weyde, City University London, UK
Sarah Whatley, Coventry University, UK
Christine White, Nottingham Trent University, UK



Keynote speakers

Cultural Heritage and Future Directions

Marco Marsella

Deputy Head of Unit, eContent and Safer Internet, EUROPEAN COMMISSION, Information Society and Media,
Directorate-General (Cultural Heritage and Future Directions)

Marco Marsella is Deputy Head of the Unit eContent and Safer Internet of the European Commission Directorate-General for Information Society and Media. He coordinates Innovation (CIP Programme) and eContentplus Programme activities in the area of digital content. Prior to joining the eContent and Safer Internet Unit he has worked on EU research activities (RTD Framework Programme) on technology-enhanced learning where he has been responsible for coordinating research agendas, implementation of RTD Programmes and dissemination activities.

Making Siobhan Davies RePlay; the UK's first digital dance archive

The pleasure of preserving intangible cultural heritage

Sarah Whatley

Centre for Dance Research (C-DaRE), Coventry University, UK, s.whatley@coventry.ac.uk

Abstract— Siobhan Davies RePlay is the UK's first digital dance archive, and one of the first in the world to bring together the complete works of a single choreographer. Built over 30 months and launched in 2009, the archive is an ongoing collaboration between researchers at Coventry University and Siobhan Davies Dance in the UK. This paper describes some of the making of the archive and the various challenges and pleasures involved in working with audio-visual dance content. It will also discuss the various initiatives that developed in parallel with and subsequent to the project, to increase the impact of the archive and to enhance user-engagement. Siobhan Davies RePlay continues to grow and has informed other similar projects, playing an important role in promoting greater understanding of the digital environment as a site for knowledge organization and distribution.

Keywords— dance archives; metadata; virtual scrapbooks; online tools; archiving process (key words)

I. INTRODUCTION

Siobhan Davies is one of the UK's leading contemporary dance choreographers. She started her career in the late 1960s and has made more than 60 works for her own company, Siobhan Davies Dance, and with several of Britain's leading companies, including London Contemporary Dance Theatre, Rambert Dance Company and The Royal Ballet. RePlay was created to preserve and increase access to Davies' body of work, which otherwise would be vulnerable and remain inaccessible to the general public. What is important to emphasize is that RePlay is a born digital archive; there was no pre-existing hard copy archive, which was then digitized and organized online. This gave us a fair amount of freedom in terms of how to organize and structure the archive but also meant that there was a great deal of work involved in gathering and describing material, and agreeing permissions and licenses to ensure that the content could be freely distributed online.

II. BRINGING RESEARCH AND ARTS PRACTICE TOGETHER

The project, funded by the UK's Arts and Humanities Research Council (AHRC), was a collaboration from the beginning, between myself, as project leader and my research team at Coventry University, and Davies herself, working closely with her

company. This was hugely satisfying as a collaboration in terms of being able to source much of the content and for ensuring the accuracy of information about the objects. But partly because we were working with a living artist and not a historic collection there were interesting challenges in how to reconcile the need for a user-friendly, robust archive and Davies' desire for this to be a creative output; which represented her work as an artist in a fair and imaginative way. Building the archive was therefore a parallel process of ensuring the back-end architecture was fit for purpose and curating the content in such a way that the interface was elegant, inviting, simple to navigate and aligned well with Davies' choreographic 'signature'.

We launched the archive in 2009 at the end of the 30-month funding period. When we started the project in 2006 there were few models to consult so we had to make decisions about how to ensure we could complete the project in time. At that time, 'open source' options were limited so we decided to purchase a proprietary digital asset management (DAM) system. The advantage of this was that we had a very efficient system that provided us with the ready-made back-end architecture that we needed, allowing us to design an effective metadata schema (based on and developed from Dublin Core) and spend time designing the front-end. The disadvantage is that we remain 'locked-in' to a system that is now hard to change, although we are investigating the possibilities. We were also occupied for many months on securing permissions and licenses for the archive content. Whilst Davies' own collection of videos, scores and text-based documents were well-described, not all the metadata was recorded and the Company's view that they owned all the content that they held, turned out, through careful research, not to be the case.

III. WHO OWNS THE DANCE AND WHO IS THE ARCHIVE FOR?

As an art form, dance has little history of effective copyright, and has generated only partial and fragmented hard copy records. It is frequently co-created and collaborative in nature (typically bringing together dancers, musicians, composers, designers, film makers, photographers and so on), which means having to identify everyone who contributes and features in the

content before adding it to the archive. At the time there were no standards that we could reference or apply so we entered lengthy discussions with legal advisers, copyright experts and individual rights holders to agree terms. Various agreements were drawn up, for dancers, composers, photographers and music publishers. Through a process of careful negotiation we were able to reach agreements and arrange licenses to secure free access to the content. There are clear terms of use published on the archive and we have a ‘take-down’ policy in case we have inadvertently violated copyright.

Sustaining the archive, as with any digital resource, is always challenging. The difficulty is continuing to fund the maintenance of the archive, particularly as Davies continues to make work so new works are added approximately twice a year, requiring staff hours, design costs and time from our DAM providers to check and test the site. There are also ongoing costs for servers and licenses. At the point when the funding-period finished Coventry University agreed to continue to support the archive and the University Library now hosts, updates and maintains the archive. Discounting staff time provided by the University and Siobhan Davies Dance, which is donated in-kind to the project, there is an annual cost of maintaining the archive of approx £15k, which is not easy to find.

Since launching the archive our task has become in some ways even more challenging as Davies’ work has developed in new directions, no doubt partly influenced by the archive and seeing her work differently through its organization and distribution online. She now rarely makes work for theatres and has made a number of projects in collaboration with leading visual artists, film-makers, sculptors, poets, composers, ceramicists, writers and so on, for gallery spaces and other non-traditional theatre settings. Having established an archival structure for RePlay, which is fixed and built around a series of theatrical choreographic works as ‘series titles’, we have since had to consider how to organize content in a way that would reveal the multi-disciplinary, multi-level nature of her work, and work that may be less ‘choreographic’ in the traditional sense. This shift is characteristic of her career, which has developed through several phases. But this more radical change of direction in the last few years has meant exploring how we can retain access to her work through a predetermined archival structure, whilst keeping the web interface interesting and fresh. At present we regularly change the content on the landing page to feature different works and make subtle changes to the colour palette, to keep RePlay ‘alive’ whilst ensuring a measure of consistency.

IV. KNOWING OUR USERS

One of our first concerns when building the archive was to know more about our potential users. An initial

user survey told us that the archive would be popular with teachers and students, not least because Davies’ choreography has been on the list of ‘set works’ in the UK schools’ dance curriculum for many years. But we wanted to extend the audience for RePlay to prompt thought about how a digital archive could raise questions about the ontology of dance, which is so often regarded as an ephemeral, disappearing art form.

RePlay currently includes more than 5000 digital objects, ranging from video to photographic images, audio recordings, scholarly articles, marketing materials and associated text-based materials. The vast majority of the content is video, including many of the full productions in performance. Prioritizing these films was important to all of us, recognizing that accessing dance on film is generally very difficult and limited to short extracts on YouTube. Moreover, many dances of the past, if filmed at all, are lost or reside in collections that are very hard to access.

To maximize user uptake, the archive is publically available and free to access. One of Davies’ major motivations for the project was to increase knowledge about dance and offer ways for users to find out more about the choreographic process so the intelligence that is embodied in the dancer and in the dance making process could be accessed more easily. Ensuring open access was also important to generate audiences for dance who may not choose, or be able to see dance ‘live’ or were not regular dance viewers. However, users can register on RePlay and opt to receive updates and related information about Siobhan Davies. By registering, the user can also access additional content that was contributed on the basis that it would only be available to the ‘serious user’.

By testing the archive during its development and talking to users we realized that building an archive does not always mean that people will find it or do things with it, however simple it is to access. With that in mind we designed a number of simple tools to help users navigate through the archive, and build their own knowledge through collecting and reflecting on their searches.

V. TOOLS FOR USER ENGAGEMENT

RePlay includes a simple but effective virtual scrapbook, available for registered users. Users can collect their searches, name the collection and if they choose, share with other registered users. The scrapbook function has proved popular with students and teachers in particular; some telling us that they have used the tool as a template for students to create ‘visual essays’. Others have used the tool to construct scrapbooks of their own content using simple blogs such as WordPress or Tumblr, following the media-rich blog structure created by some of Davies’ dancers as part of the

Jerwood Bank project in 2007 [1]. Users can also explore two dance works, to see how they were made via the ‘kitchen’ micro sites, which are attached to two choreographies; *Bird Song* (2004) and *In Plain Clothes* (2006). Each ‘kitchen’ provides a visual map of the dance work, bringing together all the source materials that were drawn on to make the work from the composer, designer, dancers and so on (the ‘ingredients’) and to show how Davies constructed the dance (the ‘cooking’). Each offers a way of developing graphic ‘scores’ or representations of dance, which users find valuable as alternative methods to document the dance making process. As digital documents they are also the only tangible records of Davies’ making process. Davies is typical of many contemporary choreographers who have careers that began prior to the introduction of simple capture techniques in that there are few if any notated scores of her work and few other traces beyond her occasional personal notebooks and documents relating to the final performance event.

What does extend back over many years is the large collection of rehearsal tapes, which is a significant part of the content available for registered users. They not only provide a valuable insight to the choreographic process but also reveal how technology has developed over time and has entered the dance rehearsal, intervening in ways that support the individual artist and simultaneously capture and document a ‘private’ process. These tapes are now available for public viewing for the first time. In particular, we have included a large number of rehearsal ‘scratch tapes’, which were made over the last decade by the dancers themselves whilst in the rehearsal studio, mostly as rapid memory aids as they work through tasks and develop movement sequences. Filmed on simple hand-held cameras, the quality is variable but the short captures are replete with information for the user; about the dancer’s thinking, making and composing process – and about how a dance evolves over time. As a different kind of ‘memory object’ they bring back lost or discarded records and can be viewed alongside the videos of the final event, thereby enriching the user experience whilst simultaneously expanding the conventional notion of what we understand by ‘archive’.

Through building the archive we have generated a number of valuable learning resources for those who may be less familiar with how to navigate RePlay and which can be used to stimulate topic work and to support other areas of student learning. These resources are freely available in the micro site ‘The Learning Space’, principally to support teaching and learning. Whilst teachers and students value the archive for the volume of content and ease of access we were particularly keen to find out how the archive might also be reaching other users. A User Impact Analysis [2] gave us valuable information that helped us make improvements and set up related projects to increase

engagement. For example, at first we received little feedback from dance artists and we were interested to discover whether RePlay was influencing artists or having an impact on how artists were using archives more generally as research tools for their own practice. We knew from talking to artists that one of the most valuable aspects of RePlay is the rehearsal tape collection, which they found particularly appealing as valuable tools for supporting them in the studio, to unblock or to give them creative triggers for devising movement.

VI. EXPANDING THE COLLECTION AND NEW TOOLS FOR SEARCHING

Soon after the archive went ‘live’ we were able to secure further funding from the AHRC to do more with the archive to enhance sustainability. We partnered with the National Resource Centre for Dance (NRCD) and Surrey University in the UK to create a portal to a number of digital dance archives, including RePlay; Digital Dance Archives (DDA) [3]. NRCD held a number of hard copy archives, which had hitherto not been digitized; the portal provides users the chance to search across these different dance collections. One of the first challenges was aggregating the content and aligning the metadata schemas; our own was built for the digital domain whereas the NRCD archives were built on a different system, which didn’t easily translate to the online. Now resolved, users are able to make new connections between an expanded collection of dance objects that span more than a hundred years.

The DDA portal brings together visual content from each archive (video, still image, designs and drawings) and offers a scrapbook tool that is more sophisticated than that on RePlay, allowing users to move, annotate and write around content, as well as display and share their scrapbooks with groups of other users. This enhanced tool was built in response to user testing of the RePlay site. Users told us how much they valued the scrapbook function as a ‘thinking space’. The enhanced scrapbook tool on DDA takes this further by allowing the user to be more inventive and playful with the content. Significantly, DDA introduces a new visual search tool for searching across the collections by visual similarity. Users can search by similarity of colour (objects or backgrounds of similar colour) and pose (a similar outline of a body, or an object, which is fixed in a moment of time). Another search, by gesture (a movement over time) is in development. The search tool reveals unexpected links between disparate content and serendipitous outcomes, providing new information for researchers, archivists, learners and educators - and allows non-experts to take different journeys through the archives to discover new dance content.

To extend our interest in how artists engage with digital archives we then offered commissions to two

UK-based choreographers to spend some time with the archive, making a choreographic response to the content and producing an accompanying virtual scrapbook, which are now included on DDA (see <http://www.dance-archives.ac.uk/scrapbook/DDA/1> and <http://www.dance-archives.ac.uk/scrapbook/DDA/10>).

This project had several valuable outcomes. Firstly, the artists, Efrosini Protopapa and Oliver Scott, told us that they felt very inspired by their time immersed in the archives (Scott had said that previously he feared ‘getting lost in the archive’) and they were able to connect with their own histories through viewing and re-embodying the histories of other artists. Audiences for these new two projects talked about their excitement at experiencing the 3-D reconstruction of the archival collections and how archives, often thought as dry and inanimate could be made dynamic to reveal new readings and interpretations. The project also brought out the question of what it means to reuse archival content. Due to our terms of use we needed to be clear about what could and could not be done with the content, limiting the performances to an invited audience and within the frame of ‘research’.

VII. REPLAY INTO THE FUTURE: SUSTAINING, EXHANCING AND DOING MORE REPLAY

Siobhan Davies RePlay was, and continues to be, a very exciting collaboration between a University and a senior dance artist. By working together we have made a digital resource that has relevance to professional artists, the research community, teachers and learners, archivists, librarians and the general user. Significantly, as a UK-based artist, Davies has toured internationally infrequently but the archive means that she is now ‘exported’. We have visitors to the archive from more than 90 countries worldwide.

As we move forward we can reflect on what we have learnt. For Davies, she now secures permissions at the point of issuing contracts so traces of her process and documents of the finished work can be added to RePlay without further negotiations. Establishing efficient systems for updating are developed but rely on Coventry University’s ongoing commitment and technical infrastructure. Davies, inspired by her own archive and its capability of preserving the more intangible facets of her dance making process, now enthusiastically collects materials for adding to RePlay, which tests the capacity of the archival structure as well as those of us working to sustain the archive. So the project continues to be demanding and requires pragmatic decisions, driven by financial considerations, time limitations and what is possible within the constraints of IPR. The project is therefore ongoing and never ‘finished’; it continues as a living archive¹.

¹ A new project has just begun; the *Library of Processes*, which is another AHRC-funded project

Dance continues to be an art form that is hugely popular and because of its inherent collaborative and rhizomatic properties, provides access to other art form content. Through resources such as RePlay, users can more easily experience and enjoy dance but building digital archives is challenging and expensive. Once issues surrounding digitization, metadata and IPR are solved, the interface needs to be designed in a way that draws users in, providing easy-to-use tools that are robust, sustainable and are relatively simple to update to keep pace with technological advancements. Encouraging user interaction to support the reuse and remaking of content to support the democratization of cultural content is another important consideration. RePlay is pioneering as a digital dance archive but is also part of a broader development of digital dance object making. Together these projects, which are frequently built in partnership with dance artists and choreographers, offer new ways of engaging with dance and thinking about the rich contribution that dance makes to our cultural heritage.

ACKNOWLEDGMENT

I am grateful to my colleagues at Coventry University; Ross Varney and David Bennett, for their work on the project. RePlay is funded by the Arts and Humanities Research Council (www.ahrc.ac.uk).

REFERENCES

- [1] <http://jerwoodbank07.siobhandavies.com/>
- [2] <http://dancetraces.wordpress.com/>
- [3] <http://dancetraces.wordpress.com/>

focusing on a digital library, documenting the making processes of artists working in Siobhan Davies Studios and directly commissioned by Davies.

‘You can’t get there from here’: Re-thinking the maps in a digital archive

Heather S. Nathans⁽¹⁾

(1) Professor of Theatre, School of Theatre, Dance, and Performance Studies, University of Maryland College Park, MD, USA (hnathans@umd.edu)

Abstract—This paper explores the relation between archival structures and academic research. It questions what it means to be “lost” in an archive (whether accidentally or intentionally), and how scholars learn to develop archival maps to guide them through their research process. Yet successful research relies on an equal combination of skill and serendipity. To perform academic archival research, scholars need knowledge and skills are needed about archival structures as well as the ability to recognize the fortuitous discovery. This essay describes a creative research and performance workshop for first-year students at the University of Maryland entitled ‘The Avatar Project’, which offers an example of how these necessary skills and knowledge can be transferred to and internalized by students.

Keywords: *performing arts; higher education; creative learning; digital archives; archival search strategies*

Author’s note: What follows below is an excerpt from a larger work that will be presented at the 2012 ECLAP Conference in Florence, Italy.

I. GETTING LOST IN THE ARCHIVE

The phrase, “you can’t get there from here” is a colloquial American expression often used jokingly by local residents to explain complicated directions to visitors, and it reminds outsiders that, without specialized knowledge, navigating from Point A to Point B may be either hopelessly confusing or well-nigh impossible. Scholars who enter the hallowed halls of an archive on the hunt for a specific piece of information often face similar challenges. No clear path to that all-important document may exist in the library records. Similar problems abound in using the archives to map histories of live performance. Using an archived record may illuminate aspects of a particular theatrical event, yet archival finding aids for locating the work of specific performers, musicians, or designers may not be sufficiently developed to facilitate an easy search process. Even more frustratingly, while a recording may preserve a performance, it often ignores the audience -- still one of the most elusive targets in our research. We may hear the echoes of gasps or laughter or applause preserved in the archival footage of an event, but we have no way to *search* for those responses in any extant database. For example, can we imagine entering the keyword “laughter” for a record of physiological

responses to a certain play? Still, given enough time, researchers, like lost travelers, could locate the objects or emotional responses they seek in an archive in which those records were *physically* present. That eventual discovery might be a combination of prior knowledge and good luck, but eventually that scholar would produce a new trail of associations that could lead future seekers to that same material.

II. DIGITAL ANXIETIES

The theme of being “lost” in an archive has a different meaning for some scholars when discussing the development of the *digital* archive. For example, in a 2003 paper for the American Society for Theatre Research, Jane McGonigal (current Director of Game Research and Development at the Institute for the Future) noted that academics and librarians writing about the growing use of digital archiving often confessed to a tremendous fear of loss: “To archivists, digital technology presents itself simultaneously as a powerful tool for increased media accessibility and a dangerous threat to preservation efforts.” Yet she describes the sense of loss as more profound than simply the fear of losing the physical object – it extends to a larger anxiety about the danger of somehow losing a collective past. [1] As McGonigal suggests, that fear has not only permeated ongoing discussions about the development of digital archives, but has, in some cases, actually subverted or impeded the ways in which scholars learn to use these archives, and perhaps more importantly, to *teach* about them.

III. SIGNPOSTING POTENTIALITY

Ironically, most scholars who spend extended time in archives become extremely adept at intellectual map-making. We enter with our own list of “landmarks” and we try to reconcile those to the ones left by curators, archivists, and librarians of the past. We also, inevitably, benefit from a measure of happy accident in our searches – as Nancy Lusignan Schultz notes in her recent essay, “Serendipity in the Archive,” [2] or as Stephen Greenblatt suggests in his new work, *The Swerve, or How the World became Modern*. As

Greenblatt and Schultz point out, sometimes it is not the item we seek that proves vital to our argument, but rather the one nestling next to it on the same shelf. But our cognitive maps allow room for these digressions (or what scientists call “place learning” [3]), and most scholars doing archival research expect to take the scenic route towards their destination, rather than a linear path. Perhaps part of the challenge in making these maps occurs when the landscape becomes virtual rather than physical. As more and more of our research materials make the transition to the digital realm, how ready are we to re-think our approach to archival map-making? And how can we transmit these skills to students who arrive at college without either a ready-made map or their own map-making skills? Part of the challenge lies in the fact that most scholars know comparatively little about how an archive is actually created or how search algorithms are developed – they simply deal with extant models and adjust as needed. [4] Can we transform this reactive learning pattern into a more proactive process, or what researcher U. Neeser describes as, “plans for obtaining information from potential environments”? [5] For me, the key words in that phrase are “plans” and “potential environments.” After all, our ultimate goal as scholars and educators is not merely to chart an extant landscape, but to equip ourselves and our students to venture into *terra incognita*.

IV. COLLECTIVE CHARTS OF THE UNKNOWN

This paper explores how we might teach students to adjust to our new archival landscape. Many of our students already know how to use digital resources. Many professors know how to use physical archives. Can we “get there from here” and bridge the seeming gap between these two approaches to scholarship? What kinds of new training do we need to envision for ourselves *and* our students to reconcile these realms? How can we engage in a kind of communal map-making across both time and spatial regions? In addition to these strategies, how do we teach students to ask the kinds of questions that allow room for serendipitous discoveries in a digital archive? And how do we teach them to make the new intellectual-historiographical-archival maps that will blaze a trail for others to follow in the future? The remainder of this essay will focus on one specific pedagogical example that may offer a useful starting point.

V. AVATARS AS HISTORICAL MAPMAKERS

In 2011 I offered a seminar that combined archival research with a creative project as a way to introduce first-year students to in-depth archival research. Part of my goal in creating this class was to see how successfully I could combine archival research and various digital resources with a creative project that

entailed performance, and to see whether I could have students engage in a collaborative knowledge-building venture where they acquired not only new information, but new skill sets and new ways of approaching future research. The class was held in the Michelle Smith Performing Arts Library (housed in the Clarice Smith Performing Arts Center at the University of Maryland). I called it “The Avatar Project.” It centered on the year 1852 – a landmark year in American theatre as *Uncle Tom’s Cabin* exploded on the stage, focusing a spotlight on the antislavery movement and transforming the intellectual debates on the issue into passionate emotional appeals embodied by the characters of Uncle Tom, Eliza, Topsy, and Little Eva. At the beginning of the semester I offered students a choice from a list of fictional characters I created (Rachel Levy, Jewish immigrant, age 23; Sarah Butler, native of South Carolina, age 17, etc.). Using this very basic information, students had to develop some type of creative project, ranging from a diary to a piece of artwork to a performance piece that revealed information about their character. The key was that they had to base *every* choice they made on archival research, whether it was the language they used, the form of the project, or the topics they discussed. It would not be enough to know the general history of the period. They would need to be able to describe what their characters ate for breakfast or did on their days off, and to be able to document how they knew that information.

As a point of entry into this process, we launched the first section of the class in a very practical way by reading secondary sources about the history of the period. These readings were accompanied by a collection of primary sources including novels, newspapers, plays, pamphlet literature, diaries, letters, and speeches – all designed to immerse students in the language, issues, emotional and psychological world of the mid-nineteenth century. We also had special class sessions on music, dance, and design, which offered students the chance to learn about the soundscape and physical rhythms of the world they were imagining, as well as its visual appearance (these classes were *practical* classes – students took dance lessons, had to sing or analyze music, and got to work with a visual artist and costume designer).

The second section of the class involved immersion in primary sources. Students used online digital newspaper collections to document issues that would have been of interest to their particular characters. Perhaps the most important part of this exercise was not the material they viewed, but the fact that they had to log every search on our shared class website, documenting *how* they located the material so that others in the class would be able to find the same source. This process had two important goals: it demanded that students reflect on their own digital search processes and shape them into a narrative (they also noted terms, phrases, and combinations that

produced *negative* results); and it created a composite map for the class on how to locate material related to issues surrounding a specific time and location. Students were also required to cite fully and accurately any sources or websites they found particularly helpful, so throughout the semester we generated an ever-evolving list of digital archive resources for our website as well.

VI. STRATEGIES FOR SERENDIPITY

In addition to the solo quests that fed our communal archival map for the year 1852, I also designed class sessions to show them how a research scholar frames questions about where to look beyond the obvious keyword-centered places. By watching me conduct ‘live’ searches for their material, they were able to see the kinds of ‘informed guesses’ that can be made when searching a digital archive, and how that kind of serendipitous proximity so crucial in a physical archive can be replicated in a virtual one.

My goal was to allow students to become comfortable not only in searching digital archives, but in *failing* in their searches, re-mapping their route, and moving on. By documenting their process and witnessing their professor interpret why certain searches did or did not yield the desired results, they learned to become more efficient in their questions and ultimately in their own individual efforts at archival map-making.

VII. CREATING HISTORY

More importantly for the purposes of this particular class, mapping the digital archive became not an end in itself, or the source of raw data, but a springboard for interpretation and creativity. Because there was a *creative* project attached to the class, students had incentive to gather material that could feed their imaginations rather than fill their list of footnotes. Their results were astonishing. For example, one young woman created the character of Kathleen Mary O’Regan, an Irish Catholic immigrant, newly come to America and settled in a household with a Jewish servant. Because she had come over on a ship named *The Ivanhoe* (named for Sir Walter Scott’s famous novel that challenges issues of anti-Semitism), the character decided to go see the play, *Ivanhoe* (based on the novel). The ship, *The Ivanhoe*, and the play were real and the student documented both through her research. Her next step was to develop a creative project. Because *Ivanhoe* promotes an anti-discrimination message, the student used this to explore her character’s feelings about being an Irish Catholic living in a house with a Jewish servant. But she did it in an extraordinary way – through music. She wrote and recorded five original songs, based on digital archives of period music she had uncovered and based on our class with a music historian. These five songs reflected the

character’s journey from prejudice to acceptance, but not simply through the lyrics. Using the archival material she discovered, the student changed the musical *structure* of her songs, so that each one showed a subtle shift from purely Irish rhythms towards increasingly ‘American’ ones inflected by African American, Jewish American, and American folk music. The project was facilitated by the digital performance archives that allowed her to compare a wide range of materials and to use her computer to do analysis on the structure of each song. For example, the site below features a number of contemporary recordings of nineteenth century American and African American music re-created by the group, “Glimpses of Glory”: <http://utc.iath.virginia.edu/songs/glimpsehp.html>

Below is a brief sample of the lyrics Robinson wrote for the final class project. These are from a work entitled “Song of Leah” (Leah is the name of the Jewish servant in the household where her fictional character worked), as well as the student’s explanation of the musical influences that shaped the song.

Song of Leah by E.L. Robinson (excerpts)

Oh Father Son and Holy Ghost, am I led into sin?
For I have seen an angel, but she is not of my kin.
Michil and I were sitting in the rowdy Bow’ry pit
The curtain’s up and I nearly had me a gasping fit.
They say the Jews are rotten and I always thought it so,
How could this be, please answer me, for my soul longs to know.

For there she was so dazzling, most beautiful and brave
What can I do Dear Lord for my heart just shan’t behave
For here I stand so battered and so broken down you see
But I want to hold her close and have her never leave me be
On stage she was Rebecca, but now I know there’s more
In my heart I call her Leah and she sleeps behind that door...

...If she can be so lovely then how do I hate her people?
If loving hearts are full of sin, what’s in the church’s steeple?
Can people just be “different” without their being wrong?
For did not your Son himself once stand against the throng?
(Moment of silence)

(Spoken) Lord, what do I do? What do I say? I wish to tell her...

Sing me to sleep
I am yours to keep
For in your off-key
Lies hidden melody

Dance in my dreams
Wrong is right it seems

As Robinson noted in her description of her research process, the melody of the first several stanzas is influenced by Irish songs concerning the 1798 rebellion, especially “The Rising of the Moon,” which was written by John Keegan Casey (1846-70), the “Fenian Poet,” apparently as an homage to an older song, “The Wearing of the Green.” The sections following the spoken line are influenced by Jewish holiday songs from *Karp’s Hebrew Holiday and Folk Songs*, as well as prayer music from the Shabbat/Festival Siddur (the Jewish prayer book for the Sabbath and Holy Days). The song references anti-Semitic beliefs held by

Kathleen's priest, and it reflects her character's distress at the seeming disconnect between the teachings of her religious leaders and her own hard-won lessons in her new surroundings. The performance of *Ivanhoe* described in the song occurred at the Bowery Theatre February 8, 1852. The song plays on the aural difference between Irish and Jewish folksongs, hence Kathleen's describing Leah's singing 'off-key' despite its 'hidden melody' that she discovers as she attempts to emulate her singing style. [6] The choice to interpret the creative aspect of the project was a serendipitous discovery based on her exploration of the digital archives, as Robinson noted in her final essay,

"My project went through several critical changes during its development. Initially... I operated under the assumption that she would be a factory worker and that the final project would center around that experience. However, because I wanted to create an ostensible situation where she would rid herself of the anti-Semitism and anti-black attitudes commonly found in the Irish Catholic community and essentially learn tolerance, my project's focus shifted, as did Kathleen's profession, to a domestic setting (that of her caring for the Smith children as a domestic servant). This also allowed me to give Kathleen a more realistic creative outlet--that of singing while working with the Smith children, which meant that my research began to include a deeper look into what sounds, rhythms, patterns, melodies, would Kathleen have heard in her life that would have influenced her songs to sound the way that they did. This meant not only listening to a lot of the music that existed in America and Ireland during her lifetime, but also looking into how oral traditions work and how people adopt and adapt things they've heard into their own music often without fully realizing it." [7]



Caption: Some of the 58 workstations with remote audio listening capabilities and internet access in the Michelle Smith Performing Arts Library at the University of Maryland,
<http://www.lib.umd.edu/PAL/vt.html>

VIII. SIGNPOST FOR TOMORROW

As Patricia Ybarra has argued in *Theater Historiography: Critical Interventions*, writing good history takes time – not only the time spent in documenting an event and producing a publishable project, but the time spent in mapping one's route into the project as well. And yet it is this very luxury scholars often deny themselves. In a hurry to get there from here, we often feel pressured to focus on the specific goal of an individual project rather than the long-term one of assembling a collection of research maps and tools that will serve us, our colleagues, and our students in the years to come.

REFERENCES

- [1] Jane McGonigal, "A lost cause: performance and the Free Speech Movement Digital Archive, 2003 Annual meeting of the American Society for Theatre Research: Documentation & Technology/Technology & Performance," Accessed on February 20, 2012 at <http://www.avantgame.com/MCGONIGAL%20A%20Lost%20Cause%20ASTR%202003%20.pdf>.
- [2] Nancy Lusignan Schultz, "Serendipity in the archive," *Chronicle of Higher Education*, May 15, 2011.
- [3] Lucia F. Jacobs and Francoise Schenk, "Unpacking the cognitive map: The Parallel Map Theory.
- [4] My thanks to Aaron Tobiason, PhD candidate at the University of Maryland and web designer, for sharing his expertise in this field.
- [5] U. Neeser quoted in Lucia F. Jacobs and Francoise Schenk, "Unpacking the cognitive map: The Parallel Map Theory of Hippocampal Function," 286.
- [6] This material appears as part of a password-protected class website:
https://elms.umd.edu/webapps/portal/frameset.jsp?tab_id=2_1&url=%2fwebapps%2fblackboard%2fexecute%2flauncher%3ftype%3dCourse%26id%3d_626728_1%26url%3d.
- [7] This material appears as part of a password-protected class website:
https://elms.umd.edu/webapps/portal/frameset.jsp?tab_id=2_1&url=%2fwebapps%2fblackboard%2fexecute%2flauncher%3ftype%3dCourse%26id%3d_626728_1%26url%3d.

The search for sustainability in the cultural heritage domain

Building a knowledge-based infrastructure

Nikki Timmermans⁽¹⁾

(1) Stichting Nederland Kennisland, social innovator, KL, Amsterdam, The Netherlands, nt@kl.nl

Abstract— In order to be able to response to the new economic reality the cultural heritage sector is facing, it is argued in this paper that a more entrepreneurial mindset and new practices in business model innovation is needed. The first part of this paper focuses on the question of value creation. The ‘business model canvas’ is introduced as a tool to gain insight in the way social and economic value is created and can generate new ideas for innovation. A few new practices in business model innovation are explored and a few key principles underlying the design of these new business models are defined. The second part of this paper addresses what is needed for cultural heritage organisations to get innovation off the ground. The ‘step plan’ and the ‘fit and viability check’ are introduced as tools to help embed new concepts and the case of the Eemland Archive illustrates how this can work in practice. A few key principles are defined to help create a support infrastructure for cultural heritage institutions to spur innovation. Finally, the author gives two recommendations on how to build a knowledge-based infrastructure for the culture heritage sector.

Keywords: *business model innovation, heritage, culture, tools, emerging practices, knowledge infrastructure*

I. INTRODUCTION

Since the economic crisis set ground in Europe at the beginning of 2008, the subsidised cultural sector is facing turbulent times with severe governmental cutbacks and a radical change to cultural policy. In this changing landscape, one thing has become clear. The dominant business model for cultural heritage institutions, which has been ‘public funding’ since the end of World War II, is under ever increasing pressure. Even though cultural institutions have been working on gaining more own income – in the Netherlands cultural institutions have to earn at least 17,5 % of their own income by 2012, performing arts institutions 21,5 % – over the past years, the challenge to search for new ways to become (financially) sustainable is becoming very urgent.

In order to respond to this new reality, there is a need for the creation of experiment and new *practice*. And, most of all, a shift in *mindset* is needed. From a subsidy-mindset with a focus on the question how we can close the gap between income and cost towards a more entrepreneurial mindset focusing on the question: what is the essence of our story? What are the core qualities and assets? One of the core assets of cultural heritage

institutions is the stories behind their collection. For the past years cultural heritage institutions have been digitising this core of their reason of being: their collections. This offers new opportunities for strengthening the role of culture and heritage in our society. Through developing new digital services, we can build social and economic capital in new ways. In order to fully assess the possibilities for cultural heritage institutions, the search has to be as wide as possible.

The aim of this paper is to make a contribution to the changing discourse of the cultural sector by mapping out new and innovative practices in business model innovation and point out some principles on how this innovation can be spurred. In the first part of this paper (paragraph 2) a tool will be introduced to help cultural heritage institutions gain insight in the way they (can) create social and economic value and reflect on new practices in business model innovation. In the second part of this paper (paragraph 3) a step plan will be introduced as a tool to help cultural institutions embed new concepts in their current business model and reflect on how this can be done in practice. In paragraph 4 this paper is concluded by reflecting on what is needed for cultural heritage institutions and organisations to meet the challenge of finding (financial) sustainability.

II. ASSESSING VALUE CREATION

A. Business model canvas

A lot of tools have been created to help cultural heritage institutions assess their core qualities and gain insight in the way they create social and economic value. One of these tools is the ‘business model canvas’ developed by Osterwalder and Pigneur [23]. This canvas was also used in 2009 for a study of business model innovation in the heritage sector by Kennisland and the DEN Foundation [3]. The canvas provides a comprehensive framework for structuring ideas about business models and generating revenue. The framework consists of nine modules closely relating to each other, called the canvas. This canvas can be applied to both projects and entire organisations. We briefly discuss the nine building blocks, before identifying a few new practices in business model innovation in the cultural heritage sector [14]:

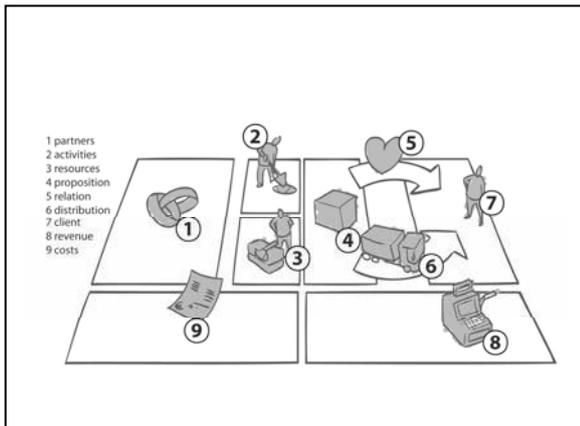


Figure 1. BIMCE business model canvas [14].

1) *Partners*: in an increasingly complex world, it appears to be increasingly attractive to engage in strategic partnerships to engage in activities that are not core competencies of the organisation.

2) *Activities*: in order to realise the proposition activities are needed.

3) *Resources*: the organisation may choose to carry out the activities themselves and deploy their own resources or outsource some of them to other partners.

4) *Proposition*: the distinctive character of an organisation. It solves a customer problem or fulfils a customer need.

5) *Customer Relationship*: customer type and the way the service is distributed, in large part determines how the organisation relates to its customers.

6) *Distribution*: the distribution channel that is used to provide the service determines the communication and customer experience.

7) *Customer*: the customer is always central to the business model. Without customers, there is no revenue. The organisation can focus on certain well-defined customer groups in order to make conscious choices about the deployment of people to reinforce the proposition.

8) *Revenue*: building blocks 1-4 together determine what the customer is willing to pay and therefore what the returns of the organisation are.

9) *Cost*: the combination of activities, the use of own resources and the cost of outsourcing activities to partners determine the cost structure of the organisation.

The 'business model canvas' can help to better understand the situation of your business model. You can structure the existing knowledge about the practices

in your institution but also discover blind spots. On the basis of a number of examples we researched in the cultural heritage sector [14], a quick overview of how cultural institutions have recently experimented with the implementation of new business - and business models will be given. Experiments which go beyond gifts, grants or funding, investments, sponsorships, traditional memberships or revenue generated from the rental of the real estate or café exploitation, which are well-known instruments in the cultural sector. The overview is not meant to be exhaustive, but is aimed at providing inspiration for innovating business models in the cultural heritage sector.

B. Practices in business model innovation

Firstly a few examples of the client side (building blocks 5-7) will be discussed, then a few examples of the organisational side (building blocks 1-3).

1) *Connecting new target-groups*: A relative simple way to innovate your business model is to rethink what you can offer to younger generations – people who will be your future (financial) supporters. What can you do to build a lasting relationship with them? Both the Van Gogh museum [30] and the Brooklyn Museum of Art [5] have managed to attract a new and young audience by opening up the museum at night and programme an informal event with a DJ, dance floor and special acts. Revenue is generated by selling tickets, drinks and in the case of the Brooklyn museum by selling exclusive memberships and offering VIP treatment, connecting a new generation to the museum and building a relationship with them. Design for Download [7] is a new way for design company Droog Design to cater for the new generation of design lovers. You can download blueprints of design products from a website and produce them at local manufacturers, or produce it yourself, instead of buying them in a (physical) shop. Droog Design offers a platform for young and upcoming designers and connects them to their brand. Revenue is generated by selling the blueprints and is shared by both the designer and the Droog Design label.

2) *Alternative pricing models*: Whereas in the cultural sector prices are most of the time determined by calculating the costs, commercial organisations determine price by what customers value it for. The most well known examples of this model are flight operators. The band Radiohead released their album 'In Rainbows' 'on the Internet' [24], giving people the option to choose if and how much they would like to pay for it and make a reservation for buying the physical CD-box. This new pricing strategy has led to substantial revenue for Radiohead. The city theatre of Amsterdam [27], de Stadsschouwburg, started an experiment to differentiate with ticket prices. The theatre determined the price of a ticket to the show Richard III on the basis of 'first come, pays the least'. The first one who buys a ticket will pay

less than people who buy a ticket at the last moment, thereby creating an incentive for people to be the first. Instead of the pricing strategy of the theatre, you can also base prices on things like popularity by good reviews or the programme slot or other relevant variables.

3) *Crowdfunding*: Crowdfunding is becoming a very popular way in the cultural sector to finance products, processes or projects by donations from people. The Louvre for example purchased the Three Graces by public donations [16]. The Sochi project [28], an initiative of photographer Rob Hornstra and writer and filmmaker Arnold van Bruggen, follows the development of the Sochi area in Russia, which is preparing itself for the Olympics in 2014. Funding for the project comes directly from people. The more you give, the more access you get to a protected part of a website and the more closely you will be informed and receive all publications and expenses arising from the project. In the Netherlands the initiative of a crowdfunding website Voordekunst [31] especially dedicated to support arts and culture projects was taken and acts now as a broker for individuals and arts organisations to get crowdfunding.

4) *Crowd financing*: Crowdfinancing is a way for private persons to invest structurally in culture and arts organisations and is something not so much experimented with yet. TwentseWelle [29] is an example of a museum where you can buy a 'museum share'. The museum attracts capital investment money and can add the interest to their capital to their assets. The Amsterdam based cultural hotspot Hannekes Boom [13] attracted private incubators. They could invest in the hotspot for small amount of money. The investors are given this amount of money plus interest as credit to spend at the hotspot for the forthcoming five years. As a private investor you are part of the cultural enterprise.

5) *Public-private partnerships*: Partnerships between the private sector and cultural organisations can create new opportunities to advance mutual interest, but best practices seem to be really rare for cultural institutions. In the Netherlands, the HEMA store entered partnerships with the Rijksmuseum [26], using artwork of the museum for everyday products like mugs, plates and vases and the Fotomuseum [11], making a selection of their collection available on canvas prints. A win-win situation was created offering a bigger public to the collections of the two museums and a more creative and meaningful product for the customers of the HEMA. Another example is UP – another initiative of design company Droog [8], introducing an investigative economic model that aims to increase the value of dead stock through re-design. As an alternative to recycling and disposal, UP treats leftover goods delivered by partner Van Gansewinkel, a supplier of raw materials and energy, for creative re-interpretation in order to

bring leftovers back into circulation. The concept invites partners from different sectors to join in.

6) *Physical capital based revenue*: Generating revenue with physical assets or the trading of physical goods is quite common in the cultural sector. Think about retail activities like gift shops, catering (restaurant, café) or the rental of real estate or venue. The Erfgoedlogies [10] foundation for example restored many historic buildings in the Netherlands, which are now used for overnight stays. Contemporary art museum Boijmans van Beuningen [21] was turned into a hotel for a period of time and people could spend the night in the museum. This area nevertheless could be explored more.

7) *Human capital based revenue*: Cultural institutions often build up a rare skill set, knowledge base or expertise, which could businesswise also be used in other contexts. For example by consultancy, education and training or trading services. The Dutch Museum Speelklok [22] entered a three-year partnership with the Chinese Palace Museum in Beijing to help restore the clocks in the Forbidden City. In return, the museum exhibits came to the Netherlands for an exclusive exhibition for which the number of visitors of the museum increased by 80%. The Concertgebouw entered a partnership with the ING bank [6], offering a management course for business professionals, given by the conductors of the orchestra.

8) *Cultural capital based revenue*: Many cultural institutions have valuable assets in house. It is known that 80% of the collection of a museum is not on show and stored away. Experiments are undertaken to get more out of this. The Museum Boerhaave recently started with an initiative to give objects away for adoption to support the museum [20]. Everybody could adopt an object for at least 1.000 Euro. The name of the beneficiary is put on a website and revealed in the museum.

C. Key principles in the design of new business models

The above mentioned upcoming alternative business models give an idea about what is possible in reshaping the traditional dominant business model of subsidies in the cultural sector. But, the list is not complete and it is hard to prove the success of the initiatives in terms of revenue. Besides a few thorough studies into practices in and the impact of new business models in the cultural heritage sector [17] [18] [19], more research, monitoring and most of all more practice in this area is needed. Based on the experiences in different projects on new business models in the cultural and heritage sector (Academy of Cultural Innovators, CATCH Plus, BMICE), four key principles in the design of new business models in the cultural heritage sector can be seen:

1) *Mix of financial resources*: In order to be more independent from governmental subsidies, a combination of different financial resources must be made. A combination of public funding, private capital and financial contributions by individuals can most likely guarantee (financial) sustainability.

2) *Value based and service oriented*: New cultural businesses are about creating the most social and economic value possible and creating high value services. The aim is to create value, not to maximize commercial profit or to spend public money. The success of new cultural businesses will be dependant on the question of whether there is a value added story to tell.

3) *Community driven*: Where traditional business models are mostly supply oriented ('we have a collection that is valuable'), the needs of the customer, citizen or end-user is becoming more central. Communities at the core of new business models, have an important voice and are given more (financial) control over cultural businesses.

4) *Crossing over*: Most highly valuable new business models can arise if you manage to make new connections or cross-overs to other sectors, organisations and (creative) industries. If you can reach out to others and create new ways of creating values together in an interdisciplinary setting of professionals.

These core principles in shaping new cultural businesses (value based and service oriented, community driven and interdisciplinary) take the shape of a business model that has been around since the late 1800s. It manifests itself mostly in times of crisis and is now getting more in vogue again [2] [25]. This business model is the cooperative model.

III. SPURRING BUSINESS MODEL INNOVATION

A. Step plan

There are many ways and forms to organise innovation and spur the implementation of new business models. What works? And why? The Dutch Innovators Network is an example that brings innovators together on a monthly basis to come up with innovative solutions. It has successfully influenced and spread innovative ideas and solutions for cultural institutions since its beginning [12]. Another example is the Academy of Cultural Innovators. This is a group of leading innovators in the cultural sector who developed and implemented new business ideas together [15].

In this paper two particular tools for organising innovation are now highlighted. Firstly the 'step-by-step plan' and secondly the 'fit and viability check' developed by the BMICE-consortium. Additional to the business model canvas, these tools were developed in

the past year to help cultural institutions embed new or existing digital service concepts in its business model [4]. The tools were especially designed for the cultural heritage sector. They are well documented and also transferable to other (cultural) organisations. The 'step-by-step plan' is comprised of seven steps. These steps can be repeated regularly on a long-term or occasional basis to help a heritage institution to permanently embed new or existing digital service concepts in its business model.

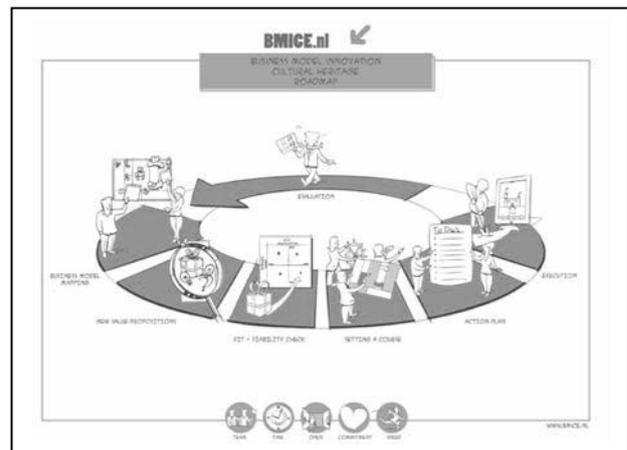


Figure 2. BMICE Step plan [4].

Step 1) Business model mapping: This step entails mapping out the existing business model and is absolutely essential in order to assess whether or not particular innovations are desirable.

Step 2) Development of product or service concept: The second step entails developing the product or service concept on which the innovation that the heritage institution wants to implement.

Step 3) Fit & viability check: Step 3 involves checking the fitness and viability of the product or service concept to determine whether it is appropriate to the organisation and viable in the market. The check indicates which parts of the existing business model need to be changed. This ensures that the new product or service concept can be successfully and sustainably executed. This involves filling out a checklist and indicating the extent to which particular criteria are applicable.

Step 4) Setting a course: The fourth step entails deciding whether or not to apply those changes. It needs to be decided whether to execute a particular product or service and determine what innovations are needed to make to the business model.

Step 5) Drawing up an action plan: The fifth step is to decide on the specific action needed to make the necessary changes to the business model.

Step 6) Execution: Step 6 focuses on actually making the changes and on the product or service concept. This is the phase in which the innovation actually occurs.

Step 7) Evaluation: Step 7 entails drawing up a diagram of the new business model and embedding the business model innovation process.

The ‘step-by-step plan’ and the ‘checklist’ can be downloaded from the website, www.bmice.nl. Although the plan has been especially designed for cultural heritage organisations in the Netherlands, it can also be applied to other creative industries around the world. The step-by-step plan was a joint approach with five key Dutch heritage institutions. The Amsterdam Museum, the Eemland Archive, Heritage Delft, the Glass Museum and the Regional Archive of Tilburg took up the challenge and brought in their cases. On the basis of the experiences of the professionals of these institutions, the case of the Eemland Archive and ‘Amersfoort on the Map’, will be highlighted to show how this ‘step-by-step plan’ works out in practice.

B. Practising new business models

In 2009, the Eemland archive had launched the website Amersfoort op de Kaart (Amersfoort on the Map) [1]. The website shows historical information (photos, movies, old maps) of the collection of the archive and of the Museum Flehite on a map and was received very well by the public.

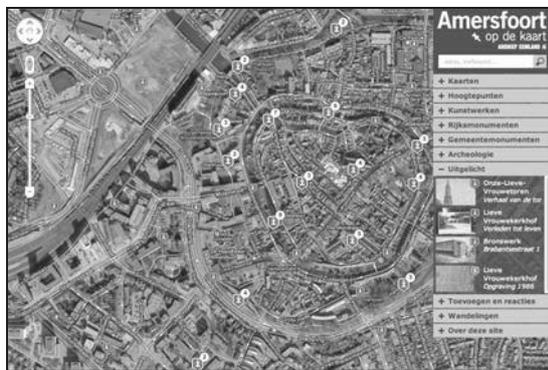


Figure 3. Amersfoort on the Map.

Immediately there were many people and organisations wanting something from the website. Professionals working in city marketing, tourism, archaeology, conservation, the creative sector in Amersfoort, other archives and educational partners. The Eemland Archive wanted to explore how it could further develop their service and joined the BMICE project. Using the business model canvas the archive defined four different value propositions to further explore, the website as:

- a platform for city marketing and tourism;
- as interactive gaming;

- as heritage portal for the city of Amersfoort and surrounding region;
- as a generic 'product'.

These different value propositions were tested in the ‘fit and viability test’, testing the fitness and viability of these propositions within the organisation. The proposition of the website as a platform for city marketing and tourism scored very highly on viability, but less on the fitness with the current organisation of the Eemland Archive. By contrast, the website as heritage portal for educational purposes scored very highly on both fit and viability. Based on these outcomes the Eemland Archive developed three projects around the website which were implemented:

- a mobile website and Layar application, adding information about monuments, archeology and art, videos and several web2.0 services for end users;
- a multitouch table application. The Eemland Archive had two multitouch tables which can be rented for institutions, such as libraries, nursing homes, schools, museums, etcetera;
- as a generic product for other archives, museums and heritage organisations to implement (www.erfgoedopdekaart.nl).

At this point, the project of the multi-touch table is developing itself as the most promising new business model of the archive. This has made a stronger proposition for the archive in the educational sector, delivering an interesting and valuable service for schools to learn more about the history of the Eemland region. Although this outcome might not be groundbreaking innovation, by going through the different steps, the archive could make considered choices about what to focus on and with whom to partner or not.

C. Key principles in organizing a support infrastructure

The ‘business model canvas’, the ‘step-by-step plan’ and the ‘fit and viability check’ can help cultural heritage organisations to structure existing knowledge. It offers a process for taking a structured approach to business model innovation. But, these tools will not instantly provide for a new revenue source or a new business model. Change nor innovation will happen by itself. It will not be there instantly by pushing a button or filling in an excel-sheet. Innovation needs to be organized and structurally embedded in the vision and policy of the organisation and the sector as a whole. Therefore a new infrastructure for the cultural heritage sector is needed that goes further than only supplying money and subsidies. The support infrastructure needs to be knowledge-based; giving access to new networks, knowledge and contacts and creating the conditions for

new practices to emerge. Based on the experiences in designing different innovation programs in the cultural heritage sector (Academy of Cultural Innovators, CATCH Plus, BMICE), eight key principles in the design of this infrastructure can be defined [15].

1) Work within a broad framework, in an open and interdisciplinary way. Invite outsiders to become involved, organise active access to other knowledge resources, expertise and new contacts from other sectors of society. This immediately extends the boundaries and scope.

2) Create a powerful vanguard of leaders in the sector. Who can become the drivers of innovation and change? Together they can form a front.

3) Work from content themes and questions from the sector itself. The sector itself has to take the lead in placing issues and questions on the agenda.

4) Offer a knowledge-sharing platform to innovators as ‘knowledge bearers’. Offer space, a (digital) platform, on which knowledge and experiences can be shared.

5) Organise ‘doing’ and ‘learning’ in close connection to each other. Link ‘Doing’ (working on projects) and ‘learning’ (reflecting on doing the projects) together. Many organisations have a tendency to get stuck in writing project plans, visions and policy. By just trying something you can test your theory and sharpen your proposition.

6) Highlight innovation and provide a stage for successful examples. Make sure that innovation is seen, identified and recognised in the sector so people can build from that and don’t have to reinvent the wheel.

7) Make connections that cut across different hierarchies and job structures. Let the knowledge extracted from practice, vision and policy circulate from staff, managers and decision-makers as effectively as possible so you have a broad-based support.

These key principles in the design of a new support infrastructure can take many different shapes and forms. These principles can be taken as a guideline for cultural heritage institutions to find a form to spur innovation in their own organization. They can also be used as guidelines for designing innovation programs for the cultural heritage sector as a whole.

IV. HOW TO BUILD A KNOWLEDGE-BASED INFRASTRUCTURE?

In order to be able to respond to the new (economic) reality the cultural heritage sector is facing, this paper has argued that we need a more

entrepreneurial *mindset* and new *practices* in business model innovation. The search for new ways to become (financially) sustainable starts with exploring the story you want to tell and the value you want to create (*what?*). In this paper the ‘business model canvas’ was introduced as a tool to gain insight in the way cultural heritage organisations create social and economic value and can generate new ideas for innovating business models. A few new practices in business model innovation were explored and a few key principles underlying the design of these new business models were defined. As we are in an explorative phase, sharing and exchanging stories of innovative business model practices, failures and successes could be of great value for cultural heritage institutions. Therefore the author suggests to start building an *international network of (best) practices*, stories to inspire and envision what is possible.

The second part of this paper addressed what is needed for cultural heritage organisations to get innovation off the ground (*how?*). The ‘step plan’ and the ‘fit and viability check’ were introduced as tools to help cultural institutions embed new concepts in its current business model. It was illustrated in this paper how this could work in practice. A few key principles were defined to help create a support infrastructure for cultural heritage institutions. Innovation (doing new things) will not happen by itself. Innovation goes hand in hand with learning, exploring and reflection. If you wish to learn, to develop a vision and to make certain choices as an institute or sector, it is important to systematically integrate reflection into the process of doing and organize the relevant knowledge. Therefore the author asserts that we should start building a *support infrastructure* to help organize knowledge, to share the stories and spur innovation in the cultural heritage sector across different departments, organisations, (creative) sectors and countries.

REFERENCES

- [1] Archief Eemland, Amersfoort on the Map: <http://www.amersfoortopdekaart.nl>.
- [2] Birchall, J. and L. Hammond Ketilson. “Sustainable enterprise programme. Resilience of the cooperative business model in times of crisis”. International Labour Organisation, 2009.
- [3] BMICE (Stichting Nederland Kennisland, Stichting DEN). Business model innovation cultural heritage, 2010. PDF available on <http://www.bmice.nl>.
- [4] BMICE (Stichting Nederland Kennisland, Stichting DEN, TNO, Stichting E30). “The Step-by-step plan”, 2011. PDF available on <http://www.bmice.nl>.
- [5] Brooklyn Museum of Art, First Fans: <http://www.brooklynmuseum.org/support/1stfans>.
- [6] Concertgebouw, ING partnership: <http://www.concertgebouworkest.nl/page.ocl?pageID=39>.
- [7] Droog Design, Design for Download: <http://www.droog.com/projects/events/design-for-download>.
- [8] Droog Design, UP recycling: <http://up.droog.com/>.
- [9] Elissen, G. “Archief Eemland en Businessmodel Innovatie”, 2011. Available at: <http://www.bmice.nl/?p=475>.

- [10] Erfgoedlogies Foundation, bed and breakfast: <http://www.erfgoedlogies.nl/>.
- [11] Fotomuseum, partnership with HEMA: http://www.nederlandsfotomuseum.nl/content/view/735/206/lan_g.nl/.
- [12] Grob, Bart and others. "Why reinvent the wheel over and over again? How an offline platform stimulates online innovation", 2011. Available at: http://www.museumsandtheweb.com/mw2011/papers/why_reinvent_the_wheel_over_and_over_again_how.
- [13] Hannekes Boom, private incubation: <http://www.hannekesboom.nl>.
- [14] Kennisland. "Businessmodel innovatie in de culturele sector", 2011. PDF available on <http://www.kennisland.nl>.
- [15] Kennisland. In connection. A conversation about art's connection with society, 2011. PDF available on <http://www.knowledgeland.org>.
- [16] Louvre, Les trois graces: <http://www.troisgraces.fr>.
- [17] MEAL. "Investing in culture and community. The social return on investing in work-based learning at the Museum of East Anglian Life". PDF available on <http://eastanglianlife.org.uk>.
- [18] MLNA. "Business models & financial Instruments for the museums, libraries and archives sector: review of the literature and survey results", 2011.
- [19] MMM. "Capital Matters. How to build financial resilience in the UK's arts and cultural sector", 2011. PDF available on <http://www.missionmodelsmoney.org.uk>.
- [20] Museum Boerhaave, adopt an object program: <http://www.museumboerhaave.nl/steun/adopteer-een-object/adopteer-een-object/>.
- [21] Museum Boijmans van Beuningen, sleeping in art: www.boijmans.nl/nl/10/press/pressitem/129.
- [22] Museum Speelklok, restauration partnership: http://www.museumspeelklok.nl/index.php?page=3_1_3.
- [23] Osterwalder, A., Pigneur, Y. Business model generation, 2009. www.businessmodelgeneration.com.
- [24] Radiohead, In rainbows: <http://www.inrainbows.com>.
- [25] Reeder, J. "Co-ops are Big: Charles Gould on the Int'l year of the co-op", 2012. Available at <http://www.shareable.net>.
- [26] Rijksmuseum, partnership with HEMA: <http://www.rijksmuseum.nl/pers/algemeen/hema?lang=nl>.
- [27] Stadsschouwburg Amsterdam, Dynamic pricing: <http://www.ssba.nl/page.ocl?pageid=315>.
- [28] The Sochi project, crowdfunding arts production project: <http://www.thesochiproject.org/home>,
- [29] Twentse Welle, museum share: <http://www.twentsewelle.nl/Subpagina.aspx?Pagina=1&Subpagina=74>.
- [30] Van Gogh Museum, Friday Night: <http://www.vangoghmuseum.nl/vrijdagavond>.
- [31] Voor de kunst, crowdfunding for arts and culture: www.voordekunst.nl.



**General track on Performing Arts, Media
Access and Entertainment**

Chair

Prof. Paolo Nesi, University of Florence (Italy)

A Model Driven Architecture approach for Open Archival Information Systems

Walter Allasia

Digital MediaTechnologies EURIXGroup 26 via Carcano 10153Torino, IT Email: allasia@eurixgroup.com

Abstract— This paper introduces a novel approach to the software design of an Open Archival Information System (OAIS): identify the needed abstract components and get the architecture as a result. Instead of starting writing the architecture first, the presented approach makes use of the *Object Oriented Design* and selects the suitable *Design Patterns* for defining the building blocks of the OAIS specifications. Models (the chosen *Design Patterns*) drive the architecture, providing a concrete software solution which is independent from programming languages and operating systems.

Index Terms - Digital Preservation, Open Archival Information Systems, Design Patterns

I. INTRODUCTION

The OAIS [1] specifications are currently the most widely adopted guidelines for setting up a preservation archive. Even though they provides a detailed functional analysis, the software implementation of the archive has been intentionally left to the user, in order to be able to address several different communities and contexts. This paper aims to provide a further level to the OAIS, enough abstract to be still able to address different topics, but at the same time enough detailed for the software implementation purposes: a candidate Platform Independent Model [2]. Making use of the Object Oriented Programming guidelines [3]–[5] and taking into account the best practices of commercial systems such as the Rosetta [6] (that is one of the current leader in the market place of digital preservation systems) an overall architecture is presented following the Model Driven Architecture (MDA [7]) approach. The main architecture follows the OAIS specification and the design involves its main six functional blocks mapped into software components: *ingest*, *access*, *administration*, *preservation planning*, *data management* and *archival storage*. In the following Sections, for each functional block a set of well-established design patterns has been chosen in order to solve the problems. A wrap up of the work done is presented in Section III whilst Section IV reports candidate future works that this paper enables as well as needed and missing parts that require further evaluation.

II. BACKGROUNDS

One way to avoid the act of design is to reuse existing designs [8]: this paper fully embraces the approach initially proposed by Beck and Johnson. According to the MDA[7], the overall analysis and design of an OAIS Preservation Platform maps the six OAIS main Functional Entities onto software components making use of available design patterns. *Generative patterns* [9], [10] are used for addressing the specific requirements of

each functional block. The design patterns solving the problems are identified and brought together in order to build up the software component diagram of the OAIS Functional Entities, eventually providing a candidate PIM. To the best of our knowledge very few works tried to follow the same approach. The most important among them is CASPAR [11]–[13], that produced a PIM with a straight forward approach, mapping the OAIS specifications directly onto software packages and classes but there are no explicit mentions to the adopted Design Patterns, which are indeed the starting point for the architecture design presented in this paper. The other international project working on this topic is PrestoPRIME [14] which is providing an Open Preservation Platform implementation based on the OAIS specs and partially funded the work presented in this paper (more details in Sections IV). Referring to the architecture overview of Figure 1 [15] each one of the six *OAIS Functional Entities*: **ingest**, **access**, **data management**, **preservation planning**, **archival storage**, **administration** are analyzed and described in the following Sections, pointing out the needed *Design Patterns* and identifying the Models.

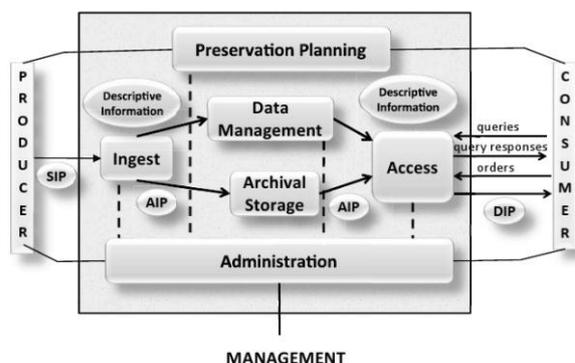


Fig. 1. OAIS Functional Entities [15]

A. Ingest

Identified Design Patterns: FrontController (Request Dispatcher, Chain of Responsibility, Filter)

The *Ingest* as well as the *Access* are the logical entities interacting mainly with the User. The *Ingest* is used by the *Producer* role, which submits the items to be preserved. A good candidate design pattern for dispatching the requests is the *FrontController* [16]: it gets the requests and dispatch to the associated component. In a more detailed view we can also split it into a Filter Cascade with a *Chain of Responsibility*. The

IngestFrontController is the first component contacted by the User that wants to send an Item to be preserved to the Archive. Actually the User sends a *Submission Information Package* (SIP). SIP is also the name of the “lollipop” symbol representing the exposed interface, on the top left of Figure 2. The Ingest component should be made up of 3 components elaborating the SIP: one is responsible for the Validation, making use of the Approver and the MetadataEnricher. As shown in Figure2 we have:

SIPValidator, which is responsible for the overall validation of the SIP submitted by the User contacting the SIP interface. This process requires the validation of the Formats (such as the check of wellformedness and validity), the Metadata, the Content (such as virus checks and checksums), and the Rights. Concerning the Metadata, a further inheritance is depicted, pointing out Metadata for Preservation, Rights, Content and Descriptive.

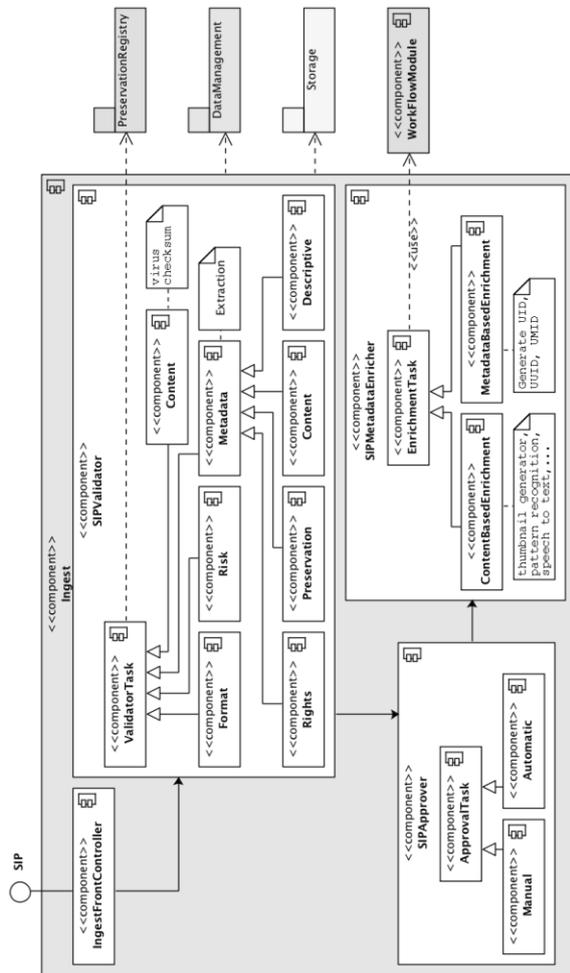


Fig. 2. **Ingest** Component Diagram

SIPApprover, which is responsible for approving the several Tasks performed in the Validation process, which can be Automatic (most of them) as well as Manual (requiring human intervention).

SIPMetadataEnricher, which is responsible for Enriching the Metadata describing the digital item submitted. These could be categorized into two main classes: Metadata extracted from the content, such as speech to text and visual descriptors and Metadata added to the item such as the UUID [17] or UMID [18] and format changes or aggregations.

Even though the metadata extraction can be performed during the creation of the SIP by the User (outside the scope of the OAIS), we can imagine that a simple metadata extractor is required in order to find out technical information about the digital content submitted. Once the SIP has been analyzed and validated, it can be stored into the Archive and the Preservation Process can take place. There are already several tools providing validation and metadata extraction features, such as JHOVE [19], DROID [20] and NZNL Metadata Extractor [21] and they will be taken into account also during the Preservation Platform implementation phase. On the right side of Figure 2 some dependencies are shown. They are modules that are not covered by the OAIS specification but should be taken into account in order to build a concrete software application. The Validation Task needs to contact a Preservation Registry component in order to get all the information related to a specific Format, whilst the Enrichment Task makes use of the Work Flow Module for processing the Content and Metadata Enrichment tasks (see Section IV for more details). The overall Ingest makes use of the OAIS DataManagement and Storage components that are described in the following Sections.

B. Access

Identified Design Patterns: Front Controller, Factory, Locator

As described in Section II-A, there are two main components exposing interfaces to the User: the Ingest and the Access. The latter is used by the *Consumer* role, which requests a *Dissemination Information Package* (DIP). As recognized in Section II-A, a *FrontController* is a good candidate for accepting the incoming requests and for dispatching them to the right software component in the Access package. In Figure 3, the *FrontController* is implemented by the *AccessFrontController* component which exposes a WS (WebServices) interface (the first “lollipop” at the top right).

In order to manage the creation of DIP packages, a *Factory* is needed, implemented by the *DIPFactory* component, which is linked to (and used by) the *AccessFrontController*. The Factory can also supply the functionality for searching the Information Packages in the system; indeed it can create *empty* package to be filled in with the retrieved information (searched by the

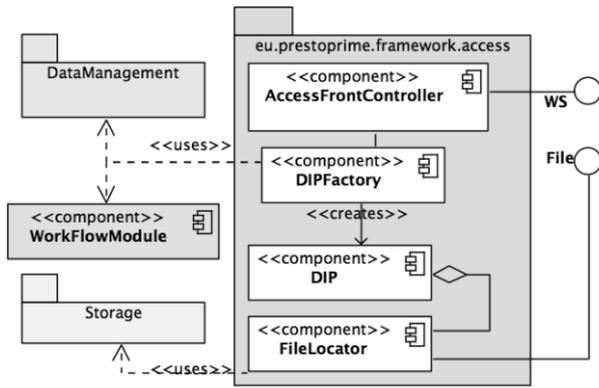


Fig. 3. Access Component Diagram

DataManagement component to which the Factory depends on). If some result is available, according to the Consumer request, a set of tasks is executed, submitted to a WorkFlowModule depicted at the left side of Figure 3 (see Section IV for further details). Examples of common tasks to be performed are format migration (resolution changes request) and metadata mapping.

Once the DIP is created by the factory, the consumer needs to access its contents. They are placed/handled in the Storage component (bottom left in Figure 3) and in order to locate them we need a decoupling service: a Locator [16] is introduced, implemented by the FileLocator component which locates the contents and resources associated to the AIPs. The FileLocator component exposes also a File interface (the second “lollipop” at the bottom right) to the user (Consumer) which will be used for accessing the bytes of the resources.

The Access component is also responsible for providing the user with functionalities for querying the Archive. If the user asks for a query, the AccessFrontController forwards the request to the DataManagement package where the Search component implements the different searching functionalities. There are no specific requirements requested by the OAIS. Even though a some protocols can be recommended such as the OAI-PMH [22], and the MPEG Query Format (MPQF) [23] and JPSearch [24], especially if the archive is going to manage multimedia contents because the latter two enable more complex queries. For example MPQF enables range and kNN queries and support the Content Base Information Retrieval (CBIR) approach, where the user can submit a digital content asking the system to search for similar ones.

C. DataManagement

Identified Design Patterns: Façade, Factory

(AbstractFactory), Entity Access Object, Composite, CollectingParameter

Within the DataManagement component, a Façade [4] design pattern is needed, in order to have a uniform and single interface for accessing AIPs that are managed internally. Usually a SIP (or a set of SIP) ingested is translated into an AIP, defining a specific Intellectual Entity. The AIPManagerFaçade is the Façade component, as displayed in Figure 4 which exposes an interface to the other OAIS software components. This is an internal interface, not published to the User. During its lifecycle, the AIPMangerFaçade contacts a Search component which is responsible for searching and aggregating AIPs. In order to get the AIPs, it makes use of a Factory for creating new AIPs as well as accessing preexisting ones, leading up to an AIPFactory implementation. The AIPFactory makes use of the Entity Access Object (EAO) [25], [26] design pattern for managing the access to the actual AIP, that adopts the AIP entities instead of simple Objects in the previous Data Access Object (DAO) [27] pattern.

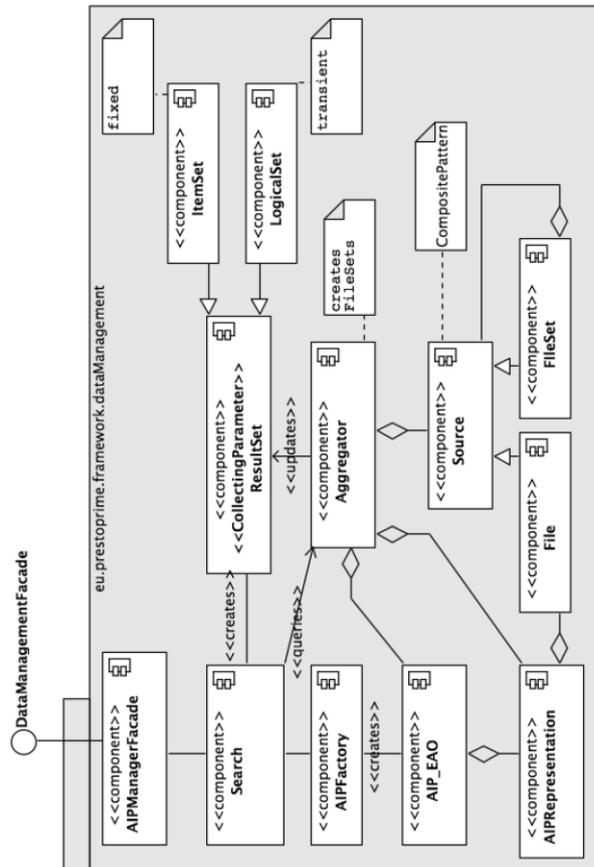


Fig. 4. Data Management Component Diagram

So far, this discussion has been focused on the creation and access of AIPs. The Data Management package is also responsible for supporting searches and aggregate results that are realized by a Search and an Aggregator components. The former is accessed by the

Facade and performs the queries on AIPs. The latter provides the aggregations of several items matching the search criteria. In order to have a common result in a single object, the *CollectingParameter* [4], [28] design pattern is introduced. In this pattern a *ResultSet* object is passed through the Search and Aggregator components, back to the Façade. It is a common techniques in order to collect results from different objects that execute methods in different times. The *ResultSet* is extended by *ItemSet*, which describes fixed and real items, and *LogicalSet*, which describes transient items such as aggregations of the same format type of items. The *Aggregator* component will be used in most of the internal processes executed for the preservation purposes such as evaluation of the risks associated to a specific format, or type, or the availability of a rendering application associated to a given media content. In order to deal with the aggregated sources as single items as well as item sets, a *Composite Pattern* [4] is adopted: the Aggregator aggregates Source which is inherited by File (the leaf) and FileSet (the aggregation of Source). The leaf File is associated and aggregated by the *AIPRepresentation* which is aggregated by the Aggregator.

D. PreservationPlanning

Identified Design Patterns: Façade, Composite, Collecting Parameter

The Preservation Planning package should expose a common interface to the other internal software components: a *Façade* pattern has been chosen, as already described in previous Sections. It is implemented by the *PreservationPlannerFaçade* component as shown in Figure 8. In the following Figure 5 it has been omitted for a better readability. The package is responsible for evaluating risks (*RiskAnalyzer* components), perform preservation simulations as well as the several preservation tasks required by the archive. The Risk Analyzer makes use of the DataManagement package, as shown by the exposed half circle attached to the “lollipop” of the DataManagement in Figure 5), where only the used components are displayed, the AIPSets and the Entity Objects. Particularly the AIPSets are the items that in Figure4 are more generically described as *ItemSet* requested to the DataManagement component, on which the preservation simulations are performed. Risk Analyzer aggregates the *PreservationRisk*, which asks for a preservation plan to the *PreservationPlanner*. The latter aggregates the *PreservationTest*. In order to have a common rule for managing single tests as well as aggregations of tests, a *Composite* pattern is introduced, where an abstract *PreservationTest* is inherited by the leaf *PreservationAlternative* and the aggregator *TestSet*.

The *PreservationTest* is aggregated by the *PreservationEvaluator*, which creates the Result objects making use of the *CollectingParameter* pattern,

implemented by the *PreservationPlanningReport*. It is

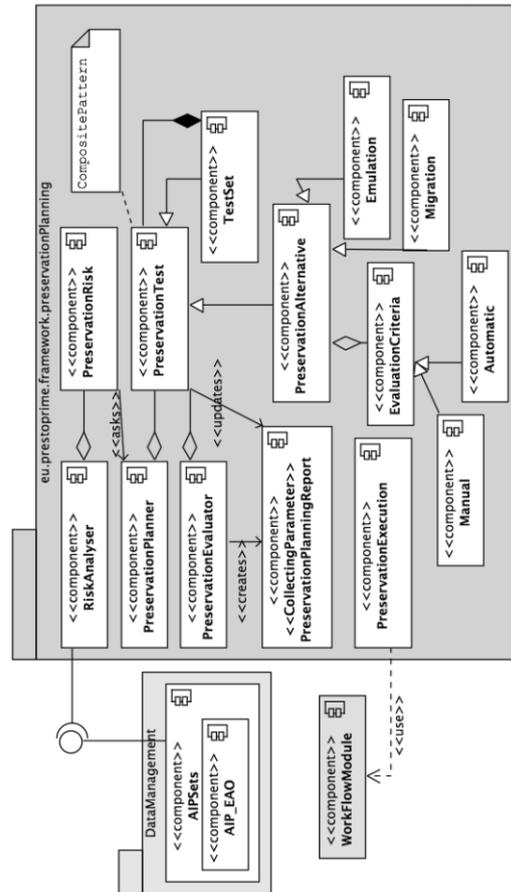


Fig. 5. Preservation Planning Component Diagram

passed to the *PreservationTest* abstraction, which updates it each time a leaf test is performed. For each Test there are alternatives to take into account. The *PreservationAlternative* is inherited by the *Emulation* and the *Migration* which are the two possible implementations and aggregates the *EvaluationCriteria*. The latter is inherited by the *Manual* and the *Automatic* criteria for the evaluation of the preservation alternatives. These couple of children Objects can be managed as a *State* design pattern, even if this is coding choice. Ideally the preservation platform should avoid the intervention of humans, which is the most expensive part of the archive management, and should make use of the *Automatic* component as much as possible. When the User has to be involved in the evaluation, the *Manual* component will interact with the appropriate UI (User Interface). As a result of the preservation planning activities, the *PreservationPlanningReport* is provided.

A further software component displayed in Figure 5 is the *PreservationExecution*, which makes use of the *Workflow* module; it is a complex component and has to expose hooks for custom implementations, enabling plug-in features of different executors.

components or systems can access the administration module by means of the *AdministrationFacade* which implements the **Façade** pattern (“lollipop” interface at the top of Figure 7). The Façade component aggregate¹ the others: the *SystemConfigurationManager*, the *UserManager*, the *PolicyManager*, the *Monitoring* and the *Audit* components.

The *SystemConfigurationManager* keeps the *DataManagement* updated whilst the *DataManagement* creates *Report* that by means of a “CollectingParameter” pattern are passed to the *Audit* which fills them with the extracted information. The *UserManager* is responsible for the management of the users accessing the Archive. They can be real as well as logical (for example other OAIS accessing the system). It makes use of the *UserPolicy* component which is figured out as a child of the *PolicyManager* and is aligned to the *Access* module. The latter is also the superclass of the *DigitalRightsPolicy* and the *PreservationPolicy* which are some (but not the complete list) of the available inheritances/specializations. If new needs arise, new children can be added to the *PolicyManger*, implementing their management. The *PolicyManager* makes use of the *Monitoring* component, which is also aggregated by the *AdministrationFacade*. It is worth to note that the *PolicyManager* child dealing with the preservation, the *PreservationPolicy*, is strongly connected to the *PreservationPlanning* package because it needs to get the policy information required by the *User*.

III. CONCLUSIONS

We have described an innovative approach for drawing the architecture of a software implementation of the OAIS specifications: identify the needed design patterns a building blocks of the overall model. The approach has been introduced by the OMG [32] as the Model Driven Architecture (MDA [7]) where the need of a specific behavior introduces models. In this paper models are made up of the identified design patterns suitable for solving the problem. Summing up the patterns and components realizing specific functionalities, the architecture comes consequently. Figure 8 shows the overall diagram representing the OAIS Functional Entities described in the previous Sections (II-A, II-B, II-C, II-D, II-E, II-F). It is a simplified overview of the software design (component diagram) and can be considered as a candidate Platform Independent Model [2] for an OAIS software implementation.

In order to achieve the highest flexibility and to

¹ Black diamonds mean “composition” that can be considered as a strong aggregation. In order to access the composed objects, every software component is required to pass through the aggregator object, the one with the black diamond attached.

enable an easy plug-in of new software components into the platform, an ESB [33] (Enterprise Service Bus) approach is recommended (and it is actually under development within the PrestoPRIME project). All the public interfaces should expose a SOAP [34] (Simple Access Object Protocol) listener each time a specific schema is required, leaving REST [35] (REpresentational StateTransfer) each time it is necessary to exchange simple documents without a too strict schema to be followed. These choice enables high flexibility to the overall system and an easy plug in of further components extending the current functionalities.

IV. FUTURE WORKS

As stated in the acknowledgment section this work has been partially undertaken within the PrestoPRIME project and an immediate future work is the actual implementation of the software components described in this paper, creating the Platform Specific Model (PSM [2]). Concerning the overall approach, an

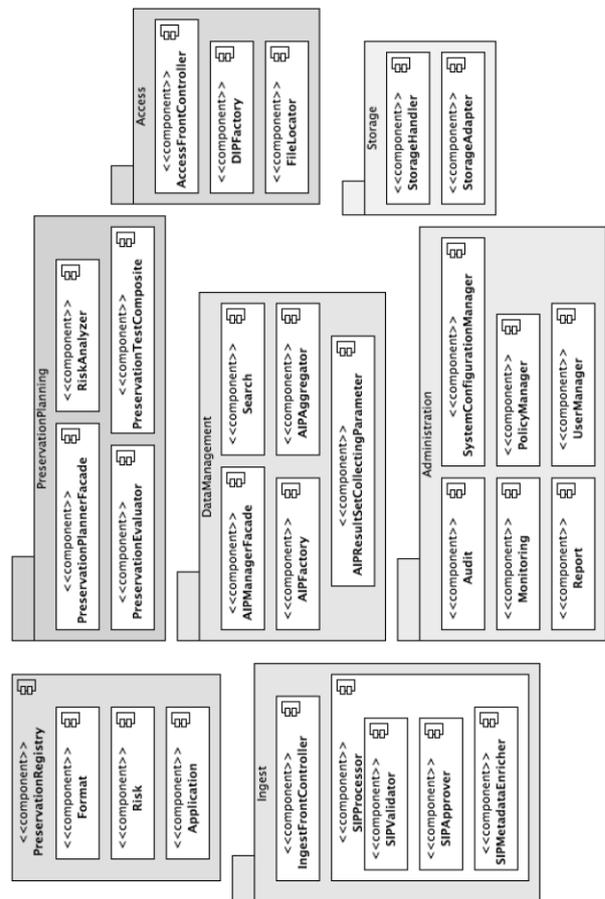


Fig. 8. The overall component diagram of the proposed OAIS Model Driven Architecture

interesting work under evaluation by the author is the description of the multimedia content preservation making use of Object Oriented DesignPatterns.

Following the approach described in this paper and the Object Oriented Design guidelines (especially the Liskov [36] principle for the respectful conversions [37] of the digital contents) it could be possible to provide a candidate design of OAIS systems tailored to multimedia contents specifically. This analysis is currently under deep evaluation.

Furthermore, some software components are not covered by the OAIS specifications, too general for this scope, but must be taken into account in a software implementation. In the paper have been identified at least the *Workflow Module* and the *Preservation Registry* software components.

The processing layer named *Workflow Module* has been introduced as a cross software component used by the others, responsible for managing the workflow that the several processes and tasks need during the runtime phases, such as the access, ingest and the preservation. It is pointed out in Figures 2,3,5.

The *Preservation Registry* is mandatory for getting the needed information about Formats and “Obsolescences”. The platform has to contact it during most of the processes undertaken, such as the ingestion (SIPProcessing) and the RiskAnalysis (Preservation Planning). As depicted in Figure 8 it can be conceptualized as made up of at least three main components:

the **Format** which is responsible for identifying the format and type of the digital content. Associated to the Format we can have the metadata extractors, that are available for a specific format and type (for example format=pdf and type=v.2).

the **Risk** which represents the risk associated to a specific content format and type.

the **Application** that can be split into Editing and Rendering Applications, for example respectively for editing/coding contents or render/play them.

It is mandatory to have and to manage these kind of information from the beginning, from the ingestion phase of digital content into the Archives. This topic is going to be addressed in the working group MPEG (AhG) on Multimedia Preservation [38], that is currently setting up the Multimedia Preservation Description Information (MPDI) as a “standard” that will make interoperable the ingestion/ submission and migration among different archives.

A. Design Evolution

Starting from the MDA approach the presented work has proposed an architecture addressing the PrestoPRIME’s needs, i.e. the preservation of audiovisual digital contents: making use of common design patterns, the functional blocks recommended in OAIS has been analyzed and a complete architecture has been generated accordingly (Figure 8). Next step of this work should analyze a more abstract level of the

architecture elements and should point out how to preserve themselves over the time, how to manage their evolution. The recognized patterns can be considered as parts of a complex information system that must be preserved as well, where each element can have dependences at different layers: the technical layer is just one aspect of the problem (mainly considered so far) and should be related to the application and business levels as well ([39]). A current work is the analysis of dependencies and criticality of used design patterns and the overall architecture, taking into account the work done within TOGAF [40] and Archimate [39] in order to define the semantic architectural blocks describing the specific domain of *Digital Preservation* and contextualize the *Information System*.

Moreover, cultural heritage is coming from the ability of archives to provide digital contents to the users in the long run: this implies that either the server side or the client side must be able to support the process. Every dependency represents a point of failure for the preservation point of view and a simple broken link should be considered as a preservation threat to be managed.

We can figure out the picture as made up of the Archival Information System *plus* client systems *plus* connected systems and dependencies are from and also to the OAIS. It is no more the only one leader in the architecture but it depends on the clients that can have different operating systems and software applications (browsers, plug-ins,..) as well as the other connected systems.

ACKNOWLEDGEMENTS

This work was partially supported within the *PrestoPRIME* project [14], an integrated project funded by the EC Seventh Framework Programme. *PrestoPRIME* involves the main broadcasters in Europe in order to set up an open digital preservation framework tailored to the multimedia content needs.

REFERENCES

- [1] CCSDS. Reference Model for an Open Archival Information System (OAIS), Blue Book. <http://public.ccsds.org/publications/archive/650x0b1.pdf>. Last visited: 2012 Jan 20.
- [2] Object Management Group. Platform Independent Model (PIM) and Platform Specific Model (PSM) for Super Distributed Objects (SDO) specification. http://www.omg.org/technology/documents/embed_intel_spec_catalog.htm#PIM_PSM_for_SDO, 2003. Last visited: 2012 Jan 20.
- [3] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. Design patterns: Abstraction and reuse of object-oriented design, 1993.
- [4] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. Design patterns: elements of reusable object-oriented software. Addison-Wesley Professional, 1995.
- [5] Erich Gamma. Design patterns -15 years later. In ECOOP, page 1, 2006.
- [6] ExLibris. Rosetta: Preserve your digital assets.

- <http://www.exlibrisgroup.com/files/Products/Preservation/Rosetta-A4LOWres.pdf>.
- [7] Object Management Group. Model Driven Architecture. <http://www.omg.org/mda/>. Last visited: 2012 Jan 20.
- [8] Kent Beck and Ralph E. Johnson. Patterns generate architectures. In ECOOP, pages 139–149, 1994.
- [9] Christopher Alexander, Sara Ishikawa, Murray Silverstein, Max Jacobson, Ingrid Fiksdahl-King, and Shlomo Angel. A Pattern Language: Towns, Buildings, Construction. Oxford University Press, New York, 1977.
- [10] Christopher Alexander. The Timeless Way of Building, volume 1 of Center for Environmental Structure Series. Oxford University Press, New York, NY, 1979.
- [11] CASPAR Project. <http://www.casparpreserves.eu/>. Last visited: 2012 Jan 20.
- [12] CASPAR Project. D2101: Prototype OAIS-Infrastructure. <http://www.casparpreserves.eu/publications/deliverables>. Last visited: 2012 Jan 20.
- [13] CASPAR Project. Overall Architecture, Components and Interfaces. <http://www.casparpreserves.eu/Members/cclrc/Deliverables/caspar-overall-architecture-components-and-interfaces/atdownload/file>. Last visited: 2012 Jan 20.
- [14] PrestoPRIME Project. <http://www.prestoprime.eu>. Last visited: 2012 Jan 20.
- [15] CCSDS. Reference Model for an Open Archival Information System (OAIS), Pink Book. <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf>. Last visited: 2012 Jan 20.
- [16] Deepak Alur, John Crupi, and Dan Malks. *Core J2EE Patterns: Best Practices and Design Strategies, 2nd Edition*. Prentice Hall, Sun Microsystems Press, 2003.
- [17] Sun. Universally Unique Identifier. <http://java.sun.com/j2se/1.5.0/docs/api/java/util/UUID.html>. Last visited: 2012 Jan 20.
- [18] Society of Motion Picture and Television Engineers. UMID Unique Material Identifier -SMPTE 330M-2000. <http://www.smpte-ra.org/umidreg/s330mex.html>. Last visited: 2012 Jan 20.
- [19] Harvard University. JHOVE [jstor/harvard object validation environment](http://hul.harvard.edu/jhove/). <http://hul.harvard.edu/jhove/>. Last visited: 2012 Jan 20.
- [20] National Archives. DROID Digital Record Object Identification. <http://droid.sourceforge.net/>. Last visited: 2012 Jan 20.
- [21] New Zealand National Library. Metadata extraction tool. <http://meta-extractor.sourceforge.net/>. Last visited: 2012 Jan 20.
- [22] Open Archives Initiative. Protocol for Metadata Harvesting. <http://www.openarchives.org/pmh/>. Last visited: 2012 Jan 20.
- [23] Moving Picture Experts Group. MPEG Query Format (ISO/IEC 15938-12). <http://www.fim.uni-passau.de/en/fim/faculty/chairs/chair-of-distributed-information-systems/research/mpeg-query-format-mpqf.html>. Last visited: 2012 Jan 20.
- [24] Joint Photographic Experts Group. Overview of JPSearch: a Standard for Image Search and Retrieval. http://infoscience.epfl.ch/getfile.py?docid=14159&name=2007cbmi_dufaux_et_al&format=pdf&version=1. Last visited: 2012 Jan 20.
- [25] DebuPanda, Reza Rahman, and Derek Lane. *Ejb3in Action*. Manning Publications Co., Greenwich, CT, USA, 2007.
- [26] Sun. The Java Persistence API. <http://java.sun.com/developer/technicalArticles/J2EE/jpa/>. Last visited: 2012 Jan 20.
- [27] Sun. Blue Prints: Data Access Object Design Pattern. <http://java.sun.com/blueprints/corej2eepatterns/Patterns/DataAccessObject.html>. Last visited: 2012 Jan 20.
- [28] Kent Beck, Erich Gamma, and David Saff. Junit documentation, cookstour. <http://junit.sourceforge.net/doc/cookstour/cookstour.htm>. Last visited: 2012 Jan 20.
- [29] IBM HP and Quantum. Linear Tape Open Technology. <http://www.lto-technology.com>. Last visited: 2012 Jan 20.
- [30] Sun. Java Database Connectivity. <http://java.sun.com/javase/technologies/database/>. Last visited: 2012 Jan 20.
- [31] Maydene Fisher, Jon Ellis, and Jonathan Bruce. *JDBC APITutorial and Reference, 3rd Edition*. Prentice Hall. Part of the Java Series series, 2003.
- [32] Object Management Group. Official website. <http://www.omg.org/>. Last visited: 2012 Jan 20.
- [33] David Chappell. Enterprise Service Bus. O'Reilly Media, Inc., 2004.
- [34] W3C. Simple Object Access Protocol. <http://www.w3.org/TR/soap/>. Last visited: 2012 Jan 20.
- [35] Roy Thomas Fielding. *Architectural Styles and the Design of Networkbased Software Architectures*. Dissertation for the doctor of philosophy in information and computer science, University of California, Irvine, 2000. <http://www.ics.uci.edu/~fielding/pubs/dissertation/dissertation.pdf> – `gepruft: 19.März 2004`.
- [36] Barbara H. Liskov and Jeanette M. Wing. A behavioral notion of subtyping. *ACMTransactions on Programming Languages and Systems*, 16:1811–1841, 1994.
- [37] Jeannette M. Wing and John Ockerbloom. Respectful type converters. *IEEETrans. Softw. Eng.*, 26:579–593, July 2000.
- [38] Moving Picture Experts Group. Ad Hoc Group on Multimedia Preservation -W. Chang and W. Allasia. <http://mpeg.chiariglione.org/meetings/geneval1/genevaahg.htm>. Last visited: 2012 Jan20.
- [39] Archimate specification -the open group, version2.0. <http://www3.opengroup.org/subjectareas/enterprise/archimate>, 2012. Last visited: 2012 Mar 28.
- [40] Togaf the open group architecture forum, version 9.1 <http://www.opengroup.org/togaf/>, 2011. Last visited: 2012 Mar 28.

3D reconstructions and Cultural Heritage communication

The modeling of the Sala Bologna as a case study for enabling emotional
philologically accurate experiences of a work of art

Micaela Spigarolo⁽¹⁾, Maria Chiara Liguori⁽²⁾, Antonella Guidazzoli⁽³⁾

(1) University of Bologna: dept. of Computer Science, Bologna, Italy, spigarol@cs.unibo.it

(2) Comitato di Bologna dell'Istituto per il Risorgimento, Bologna, Italy, mc.liguori@alice.it

(3) CINECA: Consorzio Interuniversitario, Bologna, Italy, a.guidazzoli@cineca.it

Abstract— This paper presents the experience of modeling a philological three-dimensional scenario (the Sala Bologna) as a set for a 3D stereo cartoon movie in the Cineca MDC ('Museo della Città', i.e. museum dedicated to the history of the city) cultural heritage project. In 2009 - 2011 Cineca was involved in the challenge of reconstructing three-dimensional historical scenarios to show Bologna in different ages as it probably was (the sets are philologically accurate). This movie will be part of the museum itinerary in 'Palazzo Pepoli' and displayed in the immersion room especially designed by Cineca. The aim is to take advantage of computer-based visualization methods to deliver information (culture) minimizing cognitive overload. The choice of Open Source software made the production pipeline a case-study highlighting interesting features such as model reusability. Cineca MDC Project is a case study for V-MusT.Net. The modeling of the Sala Bologna is proposed as a significant example of the issues dealt with in this new production pipeline which actually faces a twofold challenge: include philological constraints inside a traditional 3D movie pipeline production and test the multi-disciplinary ability of three-dimensional reconstructions to support both communication and research activities.

Keywords-component; virtual heritage; cloud computing; cultural communication

I. INTRODUCTION

Nowadays new technologies are carrying out a relevant role in Cultural Heritage (CH) to allow a work of art to correctly convey the message which it brings as a sign. A work of art is a semantic object (i.e. a sign [1]) originated with the specific aim of delivering a message.

To be correctly understood, such a message requires sender and receiver to share both a code (which is the way the message has been written), and a context (which relates to common information that can be assumed as generally known by people). However, the distance (both in time and in space) which today separates a sender (i.e. the artist who created the work of art) from receivers makes such conditions unfulfilled and people tend to consider cultural heritage as a matter of fact without being encouraged to disclose the embedded message.



Figure 1. The Etrurian Apa: from the 2D concepts to the 3D model.

3D models and computer-based visualization methods can help people to be more active in visiting a museum and to correctly interpret items on the basis of their original meaning, thus reaching a new audience. The goal is to involve people in emotional experiences which deliver the message concerning a work of art, while minimizing the cognitive effort, increasing motivation and enabling visual entities to be explained by visual tools ("visual is explained by visual", i.e. visual entities can be better explained by visual tools [1]).

For this reason 3D reconstruction can be employed not only to save works of art from the corruption of time but also to fill the gap between people and their ability to experience these works.

The Cineca MDC ("Museo della Città") Project is part of a wide Cultural Heritage project, Genus Bononiae (www.genusbononiae.it) 'Musei nella Città', which has renovated several buildings for public use in the historical centre of Bologna in order to define an integrated museum itinerary.

In this context, the MDC Project faced the challenge of producing a stereoscopic short animation movie, whose concept is "the big bang of the history of Bologna". It will be displayed in the immersion room



Figure 2. The Sala Bologna in the Vatican City, Rome.

(designed by Cineca) inside the museum (Palazzo Pepoli) dedicated to the history of Bologna. The aim is to intrigue the largest number of people (with different ages and background) in order to prompt them to discover items seen in the movie within both the museum itself, and the buildings in the historical centre of Bologna.

This aim raised a twofold challenge concerning both communicational and implementation issues. From a communicational point of view, the movie enables a philological approach within an emotional/narrative process. The implementation challenge was not only related to finding the most suitable software for a traditional 3D movie pipeline production, but also faced up to requirements which were specific for the project, such as integrating the traditional pipeline production with philological constraints.

Open Source provided the most suitable tools to manage such a complex project, first of all Blender. Furthermore, the Open Source model as a set of heuristics about how to encourage participation and innovation (O'Reilly [8]), enhanced a community actively engaged in the process (Cineca's partners for this work are Spark Digital - www.sparkde.com, which has already produced short movies with the software Blender, Lilliwood - www.lilliwood.eu, as the 3D stereography supervisor, and several experts such as Enrico Valenza - www.enricovalenza.com).

As a result, this production pipeline has been seen as a Blender Open Project, i.e. as a chance to study, in a real workflow production, requirements which were interesting for virtual heritage projects in general terms. For instance, a modeling approach which enhances reusability for 3D reconstructions. This made the MDC project a case-study for V-MusT.Net, a new European Network of Excellence dedicated to Virtual Museums (www.v-must.net).

The rest of this paper explains such aspects through an example: the modeling of a set for the MDC short movie, the Sala Bologna. In Section II we address communication issues introducing the relevance of the Sala Bologna for the history of the city. Section III summarize implementation issues, then delves into specific in Section III.A, III.B and III.C to outline some issues of general interest for virtual cultural projects,



Figure 3. Rendering of a shot of the scene 11: Apa come in front of the Mappa Vaticana.

which have been addressed as further work. Finally, we will outline some conclusions in Section IV.

II. THE SALA BOLOGNA

The room called Sala Bologna (depicted in Fig. 2) is in the Vatican City, Rome. Built in 1575, on the occasion of the Jubilee year (and consequently the enlargement of the third Loggia of the Northern wing at the Cortile di San Damaso - St Damaso court), it was explicitly required by Gregorio XIII, the Pope who reformed the calendar.

Ugo Boncompagni, who chose the name Gregorio when he became Pope, was from Bologna and the Sala Bologna takes its name after the many frescoes depicting Bologna in the room, one for each wall: a bird's-eye view of Bologna (also termed Mappa Vaticana) on the Southern wall, a plan on the Western wall and a view on the Northern wall. Other pictures may have decorated the Eastern wall too, but they have not been preserved due to deterioration caused by tapestries when this room was employed as a picture gallery. Further, the windows on this side were closed when the Sixtus V palace was built.

The Mappa Vaticana, which has been acquired in high resolution by the University of Bologna for Nu.M.E. ('Nuovo Museo Elettronico', i.e. New Virtual Museum [2]), is not only a work of art which could not be approached by people (it is a room next to the Pope's private apartment and near the Vatican Secretary of

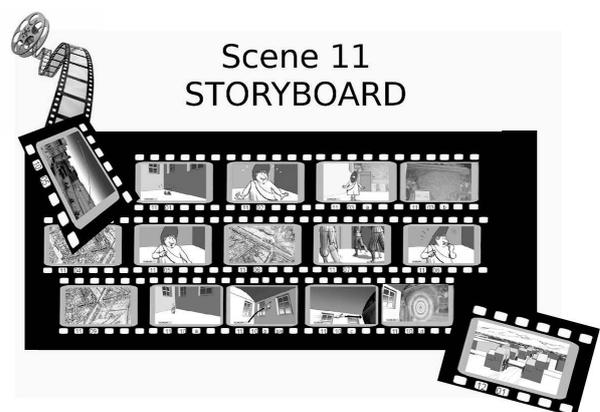


Figure 4. The entire storyboard of the scene 11.

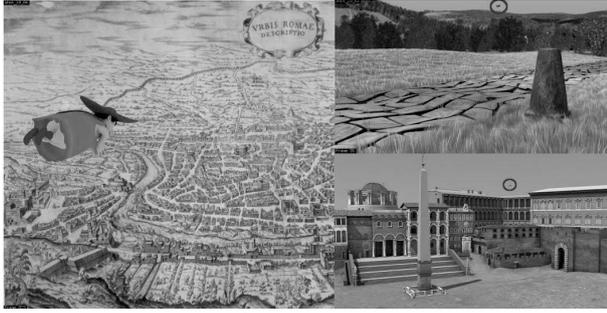


Figure 5. Apa coming to Rome: the ancient via Emilia highway and Peter's square before the dome.

State), but it is also a sort of Google Earth of the XVIth Century. The 'Mappa Vaticana' shows details of the buildings inside the 'circla' (i.e. the old city walls) in an iconographic way. This representation is precise, as verified by Dr. Ghizzoni in her PhD work [5]; in actual fact this work was able to verify how much weight the author's artistic choices had, and how much was determined by the Pope's will.

Cineca MDC favoured the experience of entering a room which is usually forbidden to people. Viewers travel with Apa, the main character of the movie (Fig. 1), from the top of the Garisenda tower (Torre Garisenda) in a Medieval Bologna, to Rome in 1578 and enter the Pope's apartment. This scene also gave us the opportunity to show a philological view of the ancient via Emilia highway, as well as St. Peter's square before the Bernini's dome was built (Fig. 5). Apa enters the Sala Bologna facing the 'Mappa Vaticana' which is suddenly shown in all its beautiful features (Fig. 3).

At the end of this scene (the full storyboard is shown in Fig. 4), Apa dives in the Mappa Vaticana and comes back to Bologna reappearing in one of the canals flanked by mills and part of the productive system of the Renaissance Bologna (Fig. 6).

III. IMPLEMENTATION APPROACH

Implementation issues faced the challenge of selecting the most suitable software tools for managing the traditional pipeline production for a short 3D stereoscopic movie, as well as for carrying out the aims which were specific to the project.

Open Source proved the most appropriate choice to manage all these aspects and the workflow in general. The traditional 3D movie pipeline has been almost completely realized with Open Source software, first of all Blender, a single product similar to other 3D tools such as 3D Studio Max or Maya, which allowed us to build the sets as well as characters, and to make the animations up to the final production of the frames on Cineca Render Farm.

This is a consolidated approach adopted by Cineca due to many benefits, such as:

- high availability and great compatibility;

- being always able to cope with problems, fix bugs, or adding new features;
- for students who attend training courses or a stage at Cineca, Open Source tools are immediately reusable;
- open format allows the reusability of the models even after many years.

This last property concerning file formats was a specially crucial issue for this project. In fact, a dialogue was required between these software which were often used in discordant contexts, such as reconstructing areas that are changing over the time, as well as the vegetation, to model cities, etc. Equally, the large amount of available documentary, iconographical and archaeological sources coming from projects of the past (i.e. made with the University of Bologna, with CNR ITABC, with the cultural department of Bologna and the Civic Museum) had to enter into the production chain.

The capability of exporting their output into a large range of different file format for Open Source software allowed us to find the right one to overcome the problem in any occasion it occurred. It is worth noting that such a capability enabled to enter a proprietary software (i.e. City Engine, for which there has not yet been implemented an open software alternative showing the same performance) in our Open pipeline.

Collaborative tools were also required to coordinate Cineca's partners, sometimes geographically distributed (i.e. Spark Digital is in Rome). In order to avoid to change internal IT design, we developed a mechanism implemented with the Cloud computing delivered by Google Docs. So the production pipeline was abstracted in terms of assets (i.e. any part of a 3D scene) life cycle. This enabled different teams to produce assets with the technologies most suitable for the existing internal IT. We delve into this subject in Section III.C.

Finally, it is worth noting the additional feature entered in the process by Open Source software: a rich developers community. This is not only a supports to problem solving, but also entails an Architecture of Participation that O'Reilly points out as a more representative property for the Open Source model today, than in fact the source availability [8]. So parallel with the production pipeline, the project developed a blog (<https://rvn05.plx.cineca.it:12001/php/MDC/portal/wordpress>) which collects internal communications as



Figure 6. Rendering of a shot of the scene 12, Apa come to the Zona delle Moline inside the Mappa Vaticana in the Sala Bologna, Vatican City, Rome.



Figure 7. A rendering of the Sala Bologna, Northern and Eastern walls, when the texturing was performing. Missed information is highlighted with a uniform colour.

well as information about the project, and the material produced above all the implementation phases.

A large section of the blog is dedicated to tutorials that represent a rich documentation of the contribution by every single participant involved in the project. These have also been employed as internal instruments for the transfer of knowledge, as well as material of interest to the Blender developers community.

As already mentioned in the Introduction, the Open-Source-like dynamic adopted also lead the project to be addressed as a Blender Open Project, i.e. the real production pipeline has been employed to highlight requirements which were interesting for virtual heritage projects in general terms, which made this project a case-study for the V-MusT.Net. The remainder of this Section delves into such issues which outline further work.

A. Integrating the traditional production pipeline with philological constraints

As a virtual heritage project, this movie was produced complying with the international guidelines and best-practices for cultural heritage 3D data (e.g. the London Charter [4]).

Sets in the movie are philologically accurate 3D reconstructions of monuments and locations representing the city of Bologna in different ages with the same relevance as the main character, the Etruscan Apa.

As a result, this work required a continuous check by historians and archaeologists to create a model that is as rigorous and fair as possible, on the other hand, even the virtual models can help to visualize and better understand their own work.

This leads to the problem of integrating the traditional pipeline production with philological constraints. Some aspects involve heuristics which is our expertise in preserving the most suitable trade-of between philological and movie requirements.

For instance, before starting modeling, all philological sources must be available. In fact, a traditional 3D movie pipeline production is composed of standard phases (see for example OpenPipeline project [7]) which requires an accurate

planning. So tasks, people involved in each task, and time needed to accomplish them are planned and scheduled once and for all. Any new philological information coming in the middle of the production

process risks compromising the delivery (which also means compromising the budget reserved to the project), so it should not be taken into account.

Equally, a movie cannot show uncertain information, which is a standard choice for insiders to highlight when the info is not known. The rendering of the Sala Bologna during the texturing shown in Fig. 7 can be taken into account as an example of models highlighting missed information with a uniform colour. One suitable choice would be to position the camera to look the parts of the model where valid information is available. As an example, the camera placement in the Sala Bologna photographed the walls which displayed most of the existing information. This limits the amount of information required to complete the set but not supported by sources.

So the employment of three dimensional historical scenarios as sets for a movie requires to making some choices which risk compromising the philological precision.

In order to preserve the philological features of the three-dimensional models (and according to the London Charter Principle 4), all these choices are recorded into documentation. This enables future improvements, too. We will deal with this subject in the next Section.

As a result, the philological approach inside a 3D movie pipeline mainly affects the architecture of the repository usually implemented to store assets. In fact, sources as well as internal documentation cause a sub tree which will be part of the repository.

This also is compliant with the Principle 3 in the London Charter [4] which states: “In order to ensure the intellectual integrity of computer-based visualization methods and outcomes, relevant research sources should be identified and evaluated in a structured and documented way”.

Data are basically ASCII or large bin files stored in a repository which is arranged as a sub-tree in the file system of the machine storing (and delivering) data (i.e. a server). So data can be accessed in a twofold way: with a browser to navigate them or with virtualization tools such as fusedav (0pointer.de/lennart/projects/fusedav) to enable the operating system's *breadcrumb* approach to file and folder navigation. This second approach also enables users to manage data, i.e. to performing operations such as creating, updating, deleting and so on.

B. Enhancing reusability for 3D reconstructions

Cultural experiences enabled by 3D technologies are meaning-driven, i.e. strongly connected with one specific meaning among all the ones concerning a work of art. This idea extends to a “one to n” relation the concept theorized by Manovich in [6] about new cultural forms and existing ones.

A database logic is a basic concept introduced by new media (with navigable spaces) which also entails new narrative forms as a method of accessing data among others. This changes the idea of work of art as unique, made within a particular medium when the tool was the medium and any level of an interface did not exist.

Today this is not true yet and a basic but crucial distinction must be made between art that uses digital technologies as a tool for the creation of traditional art objects and art that employs these technologies as its very own medium [9]. So the content of the work and the interface become separate. It is therefore possible to create different interfaces to with the same material.

In the same way, 3D technologies can be made interface for cultural information. A similar experience was pursued in the past by Cineca with the Certosa virtual museum [3]. Sources and documents, organised inside a multimedia database, get to the final user through the virtual interface in a continuous and bi-univocal interaction between virtual reconstruction and historical, artistic and archaeological documents. Equally, as observed in Section 1, the 3D reconstruction of the Sala Bologna employed in this movie delivered a specific meaning concerning the aim of relating the history of Bologna.

However, this room (and its frescoes) brings many other meanings that visual 3D technologies with a different communicational aim can deliver, employing the same model (or causing just some changes). Consider, for instance, motivations which drove into reproducing a landscape in a map (it was the time when the written works of Ptolomy came of age in Italy), or true author's artistic choices which affected the way Mappa Vaticana today presents itself.

As a result, a unique 3D model can be employed to achieve a wide range of different experiences, each of them directed to deliver a specific meaning among the ones concerning a work of art as a sign. So reusability is a property which should be carefully addressed. In order to address such capability, it is required to focus on the process in which models originated.

Generally speaking, careful planning before starting modeling is a good practice. It becomes of crucial importance for sets in a movie which attempts to be philologically accurate. Differently, there is the risk of employing a back tracking approach which entails starting anew, throwing the previous model away. It is a widespread opinion that it is a more suitable choice starting afresh than having to cause too many changes to an existing model.

This will lead to introducing a relevant delay in the delivery time, which is a crucial requirement for a 3D movie pipeline, as already mentioned in Section III.A.

This project has been the chance to study a way of producing models enhancing reusability.

The already mentioned method to document modeling and rendering choices in details used to

preserve the philological precision, proved suitable to address reusability, too. In fact, such a documentation can be employed among different virtual heritage projects, to introduce further improvements for the model when new information is available, or it is required (different contexts can also require a different degree of resolution of a 3D model). In fact, new information can be added to previous procedural annotations in order to be facilitated in modeling the new model.

For instance, when modeling the sala Bologna, no definite information was available for the cornice so it has been realized in probabilistically, as shown in Fig. 8, which is part of the documentation produced.

When (if) new information becomes available to refine the model, such a documentation will help to know how to add definition reducing the time required to plan alterations.

As a further work, we highlighted some course of action which should be suitable to automatically acquire and share information about modeling procedures followed to create a 3D model.

As an example, new Blender 2.5 allows developers to save in a Python script all the sequence of commands employed in modeling. However, a stable version of Blender 2.5 was delivered at the end of this project (which was developed with Blender 2.49). On the other hand, a sequence of procedural modeling can be shared as XML files (see for example [10]). It can be presumed that philological information could be preserved through XML procedural sequence of commands which can be implemented with the 3D modeling software preferred by the modeler.

C. Cloud Computing to support a collaborative work

As mentioned in the Introduction, Cineca involved partners for this work (i.e. Lilliwood and Spark DE). The participation of such teams, sometimes geographically distributed (i.e. Spark Digital is in Rome), required collaborative tools not only to share data but also to coordinate the pipeline process.

Although Blender was the common software we all employed, each team was dependent on an internal IT

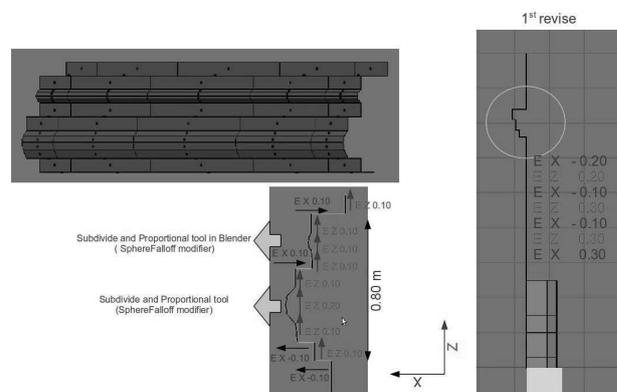


Figure 8. Part of the documentation concerning the reconstruction of the Sala Bologna: modeling choices for the cornice.

and got used to managing assets production in a consolidated way. For instance, Spark usually manages the production pipeline with a software implemented inside the company (SPAM), which had not been imported at Cineca because of the strength requirement for a network file system (NFS) architecture. This was not suitable for IT inside Cineca, which is basically composed of high performance super computers. Each team was required to produce assets with the technology most suitable for the existing internal IT. As an example, we implemented an architecture based on a centralized repository and a subversion system (SVN). So people at Cineca were allowed to checkout the repository in a local machine and then submit the new model to the repository on the server with a service which is able to automatically manage subversions of files and for each changes it maintains the history.

However, a framework was required to abstract the production pipeline in order to preserve the assets life cycle. In fact, an assets is involved in many production phases before being ultimated, i.e. wait, hold on, working, revision, approved, rejected, and closed.

A first prototype which was used in this project has been implemented with a paradigm of Cloud computing such as that delivered by Google Docs. However, this required a large amount of human intervention which could be avoided.

From now on, such a mechanism can be automatized with tools such as Google gadget API and iframe tag to embed them into a Web page.

IV. CONCLUSION AND FURTHER WORK

Relying on previous experiences on CH projects developed at Cineca, the MDC Project tries to test the ability of computer-based visualization methods to involve people in emotional historical experiences. The aim is to enable works of art to correctly deliver their messages as signs (in a semiotic sense), decreasing cognitive effort, while increasing a motivational drive.

The modeling of the Sala Bologna has been proposed as a significant example of the issues encountered in this new production pipeline which actually raised three challenges of general interest for virtual heritage project: (i) integrating the traditional pipeline production with philological constraints, (ii) studying a modeling pipeline which enhances reusability for 3D reconstructions, and (iii) studing an automate way to implement the Cloud computing mechanism designed for collaborative working.

Furthermore, a design for an interactive immersive space with Real time raytracing will be developed thanks to the GPU acceleration. A first prototype will be set up in Cineca virtual theatre using the new PLX-GPU supercomputer introduced in June 2011 by SCS S.r.l SuperComputing Solutions and CINECA .

ACKNOWLEDGMENT

We would like to acknowledge Dr. Ghizzoni, whose expertise and helpfulness have been crucial in reconstructing the Sala Bologna. Her Ph.D. Work [5] has been especially important because of the in-depth knowledge of the Sala Bologna it offered, the author having had the chance to see the subject of her PhD work in person. Equally, we would like to acknowledge Prof.ssa Francesca Bocchi. A few years she had already started looking ahead and seen the requirement for research to integrate Human and Computer Science disciplines, so she was the first in Bologna to support interdisciplinary Ph.D. projects which have been delivered as contribution to a larger Virtual Heritage project, Nu.M.E. [2].

REFERENCES

- [1] F. Antinucci, *Comunicare nel museo con DVD*. Roma, Italy: Perfororsi Laterza, Laterza, 2010.
- [2] [Bacd10] F. Bocchi and R. Smurra (a cura di): *La storia della città per il Museo Virtuale di Bologna*. Bologna, Italy: Bononia University Press, 2010.
- [3] C. Borgatti and M. Felicori and M. A. Mauri and L. Calori and A. Guidazzoli and S. Pescarin and T. Diamanri and M. C. Liguori and L. Valentini, "Databases and virtual environments: a good match for communicating complex cultural sites". In *ACM SIGGRAPH 2004 Educators program* (New York, NY, USA, 2004), *SIGGRAPH '04*, ACM, pp. 30–.
- [4] London Charter for the computer-based visualization of cultural heritage. URL: www.londoncharter.org. Last acces: 2011.
- [5] M. Gizzoni, "Analisi storica e informatica di fonti iconografiche: la pianta prospettica di Bologna in Vaticano (1575)". PhD work at the University of Bologna, History and Computer Science, 2001.
- [6] L. Manovich, *The Language of New Media*, Leonardo Books. MA, Cambridge, USA: Mit Press, 2001.
- [7] R O'neill and P. Mavroids and H. Hom, "Openpipeline: Teaching and implementing animation production pipelines in an academic setting". In *ACM SIGGRAPH 2007 educators program*.
- [8] T. O'reilly, "The Architecture of Participation" . URL: www.oreillynet.com/lpt/wlg/3017, Apr. 6, 2003.
- [9] C. Paul, *Digital art*. London, UK: Thames & Hudson Ltd., 2003.
- [10] C., Schinko and M. Strobl and T. Ullrich and D. W. Fellner, "Modeling procedural knowledge: a generative modeler for cultural heritage". In *Proc. of the Third international conference on Digital heritage* (Berlin, Heidelberg, 2010), *EuroMed'10*, Springer-Verlag, pp. 153–165.



The research leading to these results is partly funded under the European Community's Seventh Framework Programme [FP7 - 2007/2013] under the Grant Agreement 270404



Museum Outside the Museum. Real Environment Storytelling.

A case study: 150DIGIT – Italy of many schools

Stefania Chipa, Gianluca Simonetta, Luca Toschi

Communication Strategies Lab, Department of Science of Education and Cultural and Training Processes,
University of Florence, Florence, Italy
{stefania.chipa, gianluca.simonetta, luca.toschi}@unifi.it <http://www.csl.unifi.it>

Abstract—The Italian museums' main concerns are the number of visitors in decline and the exclusion of many interesting museums from the main stream of visitors. The user models museums usually propose appear to be weak as engines of social discourse. A new communication strategy (according to the “generative” communication model, visit www.comunicazionegenerativa.org) could engage visitors in a process of deconstruction/reconstruction of knowledge, creating dynamic and mobile communities. Inside and outside the space-time of the museums. Using knowledge in a very personal way we generate new knowledge. From this specific point of view we believe it is essential to rethink the environment of use: proposing new fruition models (to deconstruct and to reconstruct knowledge); redesigning the visit space and time in a way to extend the physical spaces of the museums to social areas in the territory. 150DIGIT experiments the possibilities offered by digital technology: from digital writing environments to Augmented Reality applications.

Keywords-component: museum, storytelling, digital writing, territorial knowledge, generative communication model, Augmented Reality, writing-in-action, reading-in-action, knowledge management system.

I. INTRODUCTION

Even the most recent statistics on the health of our museums present a scenario that is hardly comforting. The Italian museums have been confronted with many concerns for some time, of which the main ones can be summarized as: (a) the number of visitors is in steady and progressive decline and (b) many interesting museums are excluded from the great stream of visitors. A case in point is the archaeological area of Ravanusa, Sicily, which in 2009 did not sell a single ticket, and yet spent 340,000 euros in expenses for salaries and maintenance [1]).

It is true that in 2010 visitors of all state museums, monuments and archaeological sites in Italy increased by 15% compared to 2009 (37,336,961 against 32,380,144 the previous year), but the total was slightly higher than in 2006, when the visitors had stopped to 34,474,591. This is due to the fact that the number of non-paying visitors has significantly increased [2]. The comparison with the international scenario is revealing: in 2008 French museums have distributed more than

66.3% of tickets than the Italian museum (over 55 million tickets against 33,103,000 of the Italians).

If the receipt of the ticket offices in Italy manages to cover just 2% of budgets, in Europe this figure rises to 4.1%. To this we can add, for the purpose of assessing the sustainability of the museum systems, 8,7% coming from the so-called 'additional services' (bar, bookshop, car parks) and 8,8% from donations and sponsorships. In our country these values are notably lower: 6% additional services and 0.4% donations and sponsorships. The conclusion drawn is that Italian state museums weigh on taxpayers for about 89, 1% [1].

As for the second issue problem facing the Italian cultural heritage situation, namely the exclusion of many museums from the large influx of visitors, the measure of this difference between 'the usual suspects' (e.g. the Vatican Museums, the Colosseum, the Uffizi), and the rest of the museums, is revealed once more by the statistics. It is telling that 2% of Italian museums welcome 50% of visitors; the other 98% of museums reach the other half of museum attendance [3]. This disparity is of considerable importance in Italy, given the particular culture landscape of our country: Italy is in fact characterized by a constellation of museums, usually small or medium size, which are evidence of specific conditions and cultural traditions that have stratified in territories over centuries. Alessandra Mottola Delfino, current President of the Italian section of the International Council of Museums (ICOM), has captured this feature of the Italian situation with the expression "widespread museum" [4].

If, as the data indicates, this wealth (Italy has the greatest number of sites inscribed on the list of UNESCO World Heritage) is not being valued or, in the case of widespread museums, cannot be operated within regular channels of museum going, the result is fragility in our system, in terms of economic sustainability but also for reasons linked to our cultural identity broadly considered.

The concept of cultural sustainability must be analyzed in its dual nature. On the one hand, we must consider the museum's strategic direction so as to guide the mission and focus objectives with respect to

competitors. On the other hand there is the need to build dynamic and productive relationships that are a driving force for the economy of the area of direct affiliation. The fact that museums are rarely or very inadequately visited signifies that they are failing to exercise their primary role, as has been widely confirmed [5], [6], [7], [8], to act as an agency that intercepts knowledge, systematizing it on the basis of a precise global project and, finally, develops and disseminates it throughout the territory.

This perspective goes far beyond the simplistic view of communication as transmission, opening a new communication paradigm, the generative paradigm [9]. It becomes essential to rethink the strategic use and production of media, including the most recent developments in Augmented Reality. This article comes from this perspective.

This paper is structured as follows. Section II describes what, in our estimation, are the reasons why Italian museums are affected by the problematic situation outlined in this Introduction. Section III presents the *150DIGIT* overview, describing the project goals, the partners involved, and articulating a specific scenario wherein the solution has been applied. Section IV depicts the strategy we have followed. Section V describes the *150DIGIT* process, summarizing the tools and the achievements. Conclusions are drawn in Section VI.

II. DIRECTIONS FOR INTERVENTIONS

Italian museums could attract more visitors if they were to implement new ways of communication that are capable of the following: (a) engaging visitors in a process of deconstruction of knowledge offered in ways that take into account their needs and communication skills; (b) creating communities, fostering the processes of socialization of knowledge also outside of the classical space-time horizons of the tour, but extending outwards towards the territory and its social system.

We recognize that most visitors at Italian museums are strongly motivated; the appeal to those who are much less interested is small [10], [11]. Those individuals who possess a strong motivation are capable of moving beyond the difficulties that the visit of an Italian museum entails. The monitoring conducted on the behavior of many visitors [12] reveals how they felt at the end of the tour. They appear generally confused, tired, and unable to remember much about what they have just seen [3]. For many museums are perceived as harsh environments, in which knowledge is acquired through exhausting tours, the complexity of which is inadequate for the quantity and the quality of new information taken in. How can we alter this environment to encourage larger numbers of visitors? How can we enhance access for those who, from the outset, show little interest?

The second aspect concerns the ability of our museums to use their knowledge base to generate new communities, in accordance with data obtained from a careful monitoring of the visitors' experience [13]. Visitors to museums feel that the process of finding information (e.g. directions to the museum, buying a ticket, the opening hours) is not transparent or simply understood. This difficulty can be attributed to several causes: (a) the scarce visibility of many museums on the Internet. This involves, for example, the politics of positioning on the search engines that could be stronger or the semantic distance that sometimes exists between the name of the museum and its web address (the third largest Italian state museum by number of visitors, the Uffizi in Florence, until recently did not contain in its web domain the museum's name "uffizi"). (b) The indifferent attitude of many of our museums to activate communication relations with the local systems in place around them. We propose establishing closer connections, connections easily seen by visitors, among the system of public museums and the private museums, museums and education agencies, museums and municipalities, tourist promotion operators. A generative process of communication could be activated even within the commercial activities that they exert in the local areas [9]. One of the factors that significantly affect the visibility of museums is, in our opinion, represented by the connective capabilities that they exercise over their territorial system of reference that should not be confused with the territory in which the museum itself is physically located. If they are not well known, the reason is that communication relationships with potential users and the local institutions are negligible, faulty, or nonexistent. Except for the "usual known" museums that occupy a dominant position in attracting visitors (they have established themselves as a brand), the richness and variety of museums offered by the Italian territory is hardly known.

In other words, the problems faced by museums in defining their strategic identity can be found in the current difficulty in constructing narratives in terms of effective social representations and their economic impact. These narratives should be developed in production terms that are valid for the immediately surrounding territories and, moreover, in the prospect of a now unavoidable global economy.

The user models Italian museums usually propose appear to be weak as engines of social discourse. To overcome the current impasse we believe it is essential to develop a new model of conservation and enhancement that take into consideration the following factors: (a) to rethink the environment of use, enabling visitors to choose between the narrative paths that reflect different approaches according to cultural and economic profiles; (b) to propose fruition environments where it is possible to deconstruct and restructure knowledge, not only from a reading perspective but also

from the perspective of writing the narrative paths offered by museums (this is done according to an idea of communication that generates knowledge); (c) to radically reconsider the use of space and time by going beyond the classical limit of space-time visits and widen them to social territories, including the use of new technologies and Augmented Reality.

III. 150DIGIT OVERVIEW

These premises constitute the scenario in which our research lab, the Communication Strategies Lab of the University of Florence, imagines the project *150DIGIT – Italy of many schools*. At the same time these reflect the issues that we have addressed during the first steps of our experiment.

150DIGIT is a project that has been created to support *Esperienza Italia*, four important exhibitions organized to celebrate the 150 year anniversary of the unification of Italy: “*Fare gli Italiani*” (a journey through 150 years of national history), “*Stazione futuro*” (how will Italy be in ten years?), “*Il Futuro nelle mani*” (the excellence of craftsmanship), “*La bella Italia*” (150 years of masterpieces of art from all over the peninsula).

The *Comitato Italia 150* (www.italia150.it), organizer of the celebration, has been our first partner in this experience. Furthermore, we have involved other national institutions dealing with research at various levels: in regards to training issues, the National Institution for Documentation and Educational Research (INDIRE - www.indire.it); in regards to technological aspects, the Interdepartmental Center for the Research on Multimedia and Audiovisual (CIRMA - www.cirma.unito.it) and the Virtual Reality & Multimedia Park (VRMP - www.vrmp.it) in Turin.

150DIGIT has been recognized by the Ministry of Education, Universities and Research (MIUR - www.istruzione.it) as an action for innovation in training and education with the help of Communication Technology.

Starting from the generative communication model [9], we have used the exhibitions of the 150 year anniversary of the unification of Italy to create different narrative strategies: visitors can deconstruct the contents of the exhibitions to generate new stories. These narrations are then published in an online environment structured as a social network; this network has the ability of creating an engaging social environment wider than the physical space of the exhibitions.

We have focused our attention toward schools for their capacity to be a ‘mirror’ of our society [14], as they reflect in their organization the forms of communication that we use. Moreover, schools act as *agenzie di conoscenza* (‘knowledge agencies’): they play a vital communicative role by collecting, organizing and disseminating stories from local families and social communities.

IV. 150DIGIT STRATEGY

150DIGIT has set out the aim to redefine the relationship between the visit space and time, redesigning different time-space environments that are capable of extending the physical spaces of the museum to other social areas in the territory.

Tracing the origins of the first museums, we see how these objectives were also those who had guided their constitution. The stories of two of the symbols of our museum culture, the Louvre (still the most visited in the world) and the Natural History Museum of Florence (the oldest science museum in Europe), were built on a similar foundations. At the base of the first there was a declared need to build an integrated system with the territory and for this reason it was “placed in the national education system” [15]. What inspired the second connection was the need to establish a relationship with users based on engagement. In fact, the Museum of Natural History was created to allow students to manipulate the herbs and the natural objects, as if it were a real laboratory.

The concept of manipulation is central to the theories of knowledge. The possibility to work with objects, structuring and deconstructing them from our system of knowledge, promotes the acquisition of new knowledge and the more durable storage of such knowledge. The perceptual motor learning model [15] has indeed a strong experiential component: it allows you to test the effects of an action and to put content within the socio-cultural context most familiar to us, making it more comprehensible. It is no coincidence that this is the model that we use in the early years of life to know the world. For this reason, to promote more effective communication and a more rewarding, educational experience for even for those visitors less motivated to visit, with *150DIGIT* we tried to give life to an environment of use that would allow visitors to take action on the works on display, structuring and deconstructing the content. In parallel, since the works are included in a structured online environment like a community, each element has become a territorial knowledge that has helped to extend the exhibition to a wider social space. The exhibitions have become the engine of a new society that is not structured around a physical territory (space and time of visit), but around the knowledge generated from the enjoyment had by the exposure.

All this was accomplished by using the possibilities offered by digital technology. Manipulative and connective techniques have enabled us to work on creating a new environment of use without having to physically act on the physical environment of the exhibits. Together with project partners, we constructed an online environment (www.150digit.it) that expands the ability to communicate traditionally offered by an exhibition with new perspectives based on the possibility of use for:

- visitors to manipulate objects: deconstructing and reconstructing the 'digital transpositions' of the exhibits and put it in a new context within their system of knowledge;
- to create exhibitions with visitors and with educational institutions around the area, a community based on knowledge, capable of extending the narrative space of the museum beyond its physical location, yet within a an integrated system.

For this reason the core of *150DIGIT* consists of two major conceptual environments:

- the environment for the composition of original contributions. Within it, as we shall see further below, teachers and students were allowed access to the content of exhibitions and rewrite them according to their perspective of knowledge;
- the environment of use, implemented for viewing in 3D mode the works on display and for sharing and dissemination on the territory of new narratives made by schools and visitors.

V. 150DIGIT ENVIRONMENT

It was therefore implemented a communication environment in which the participants could find the means of expression (ICT and its related literacy) necessary to make their voices heard with original contributions.

A. Communication environment and workflow.

Fig. 1 shows *150DIGIT*'s workflow. Full environment is conceived as a crowdsourcing environment for national identity; it is composed of sub-environments below:

DigitViewer conceived as a consultation platform to play online works and installations from exhibitions of *Esperienza Italia*;

DataSet conceived as repository of digital resources focused on the themes of exhibitions;

EduLab (D-Training + D-Writing) conceived as training environment that guides users to become aware of process of creating a content via *compositing tools*;

DigitalPublishing conceived as social publishing and territorial dissemination of experiences, specifically oriented towards the experimentation of innovative solutions of *Augmented Reality*.

B. Focus on D-Writing: reading & writing in action.

D-Writing is the main *compositing tool* of *150DIGIT*'s environment and is focused on basic process of selection and combination of assets and ideas: starting contents comes from exhibitions of *Esperienza Italia* via reading action; ending contents proceeds from user's works via (re)writing action.

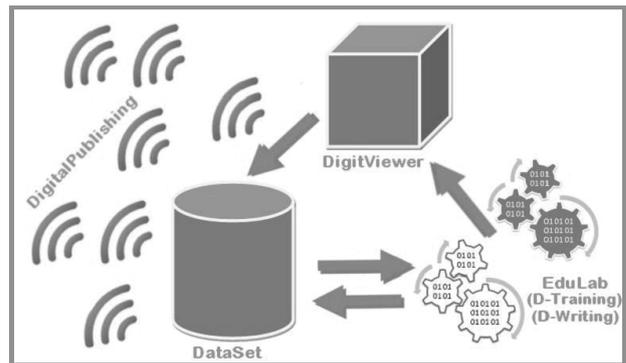


Figure 1. Sub-environments and workflow

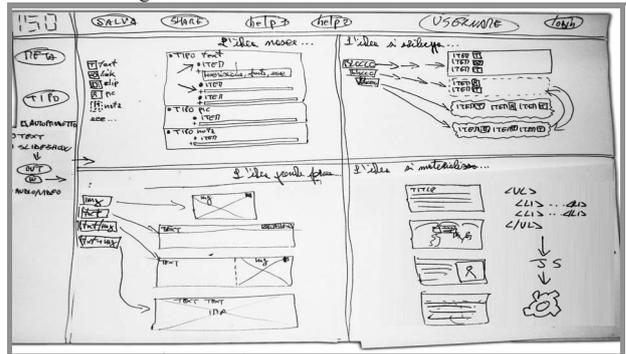


Figure 2. Mockup for "rich tech" approach

1) A framework for Digital Writing: rich tech and poor tech.

Digital Writing starts in reading, before manipulating assets, just ordering selection of contents, and proceeds in rewriting, using different combination of assets to create new contents that are ready for new reading in action.

D-Writing allows two directions of work, served by the same generative concept: a "rich tech" approach and a "poor tech" approach. Both follow selection and combination process in mutual alternation; second approach can integrate the first.

2) "Rich tech" approach: four digital setting.

"Rich tech" approach (Fig. 2) provides four digital setting. Writing in action swings through four (non strictly consequential) steps that attends users to accomplish goal in reading contents (for generating ideas), pull out assets (selecting ideas), match and arranges in re-writing new contents (to create a forceful combination) ready for publishing.

Setting one (Fig. 3) - The idea was born... Panel for collecting materials, documents, resources, ideas, personal notes, etc. It allows an inventory ordered at the moment of reading: because ideas come from research and documentation.

Setting two (Fig. 4) - The idea takes shape... Panel for the spatialization of structured collected materials. Because thoughts and intentions need to be developed and organized even through several attempts.

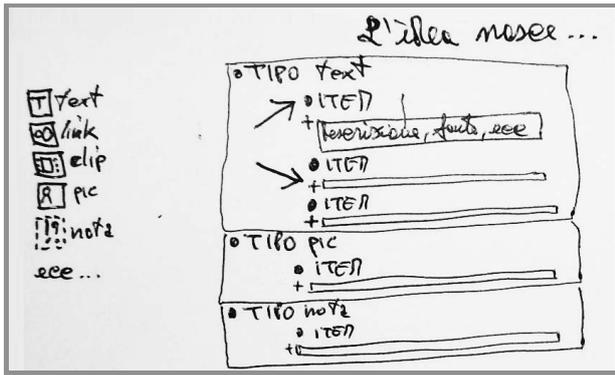


Figure 3. Mockup for setting one (*The idea was born*)

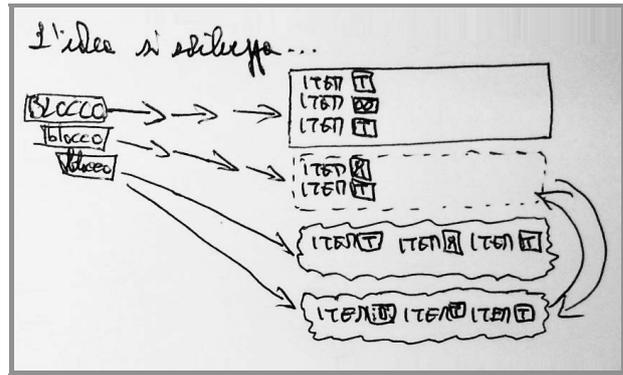


Figure 5. Mockup for setting three (*The idea takes shape*)

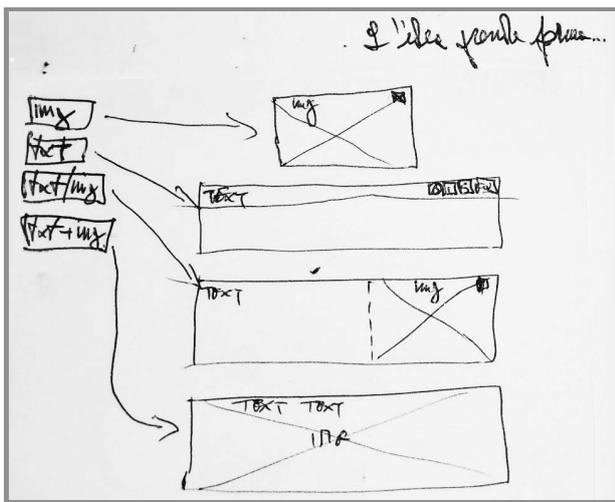


Figure 4. Mockup for setting two (*The idea develops*)

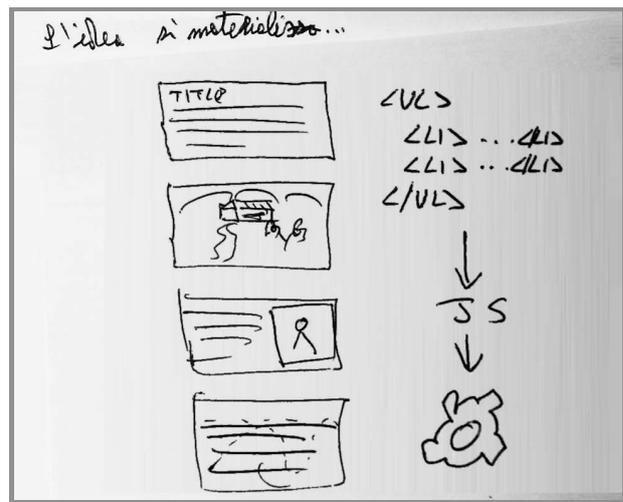


Figure 6. Mockup for setting four (*The idea is materialized*)

Settings Three (Fig. 5) - The idea develops... Editing panel. Because the eye wants its part and then at some point you start working on real editing.

Setting four (Fig. 6) - The idea is materialized... Instant preview panel. Only in this way the ‘visual’ layout of the idea is materialized. This visual effect may highlight some points of the process which need to be revised.

1) “Poor tech” approach: where good ideas come from

“Poor tech” approach attends users where good ideas come from. A writing environment can be immaterial. The first digital writing lesson is: *let's script ideas... ICT will follow!* The main point is to regulate the steps leading from idea to realization in any form or language.

Equipment for reading and writing in action consist in a *Kit for the Script*.

The kit contains a structured operational proposal in three simple steps that help you organize the work of composition. It is a lightweight, minimalist approach, designed as a rough indication, the user is encouraged to break rules and to customize the procedure: *your creativity is different from all other...*

2) *Inside the Kit*

The kit contains three .rft documents for the screenplay, to navigate and complete, on the computer or in a printed version to create the script for a content.

a) *Raccogli-le-tue-idee.rft - Collect your ideas...* A simple grid for collecting and sorting the ideas, insights, purposes and materials (Fig. 7).

b) *Andiamo-con-ordine.rft - First things first...* A smart scheduling table for the phases of work, step by step, day by day (planning what to do today and what tomorrow) and user by user (to distribute the tasks when the work is group) (Fig. 8).

c) *Ogni-cosa-al-suo-posto.rft - Everything at their place...* Just another table to keep ordered your resources when repository is the world wide web... and to find the right thing at the right moment (Fig. 9): “Do you still remember the URL of that video?”; “What is the image you wanted to add at that point of your text?”; “In which directory did you download that pdf?”

C. *From Word editing to Wor(l)d editing: contents for Augmented Reality.*

Reading and writing in action find a deep mean at the end of working process, when *DigitalPublishing* sub-

Curating Data, Disseminating Knowledge: Museums of the Digital Age

Sudeshna Guha

Research Assistant, Museum of Archaeology and Anthropology, University of Cambridge, U.K.
sg10012@cam.ac.uk

Abstract — This paper substantiates the premise that cultural heritage is a construct, and therefore engagement with theoretical issues are mandatory for understanding the ways in which heritage gets created, nurtured and preserved. It demonstrates that although digital media allow the nurture of repositories of culture heritage, they require curatorial directions for establishing notions, and illuminating the changing social lives of the phenomenon.

Keywords: *Curation, Heritage-making, Materiality, Context, Meaning.*

I. INTRODUCTION

Museums in the digital age are no longer conjured by the ‘public’ as dark and alienating places with mile-long cabinets of other people’s creations and possessions, bearing down with curatorial authority upon hapless visitors, thrusting pedagogy and histories, and displaying the power and force of institutional mediations. Most museums, be they public or privately owned, of local or national intent, in the western or ‘other’ worlds, now vie with their chosen peers for establishing dialogical relationships with ‘source communities’ and people, and making conscious endeavours in conveying their curatorial practices. The open ended, semantically orientated and user-driven characteristics of the digital media present infinite possibilities for negotiating relationships between democratic communications and museum practices. In turn, museum uses of digital technologies engender ideological shifts regarding perceptions of communication, and transformations of digital media into museum objects. For example, by enhancing the visual and sensory experiences of the exhibited objects digital technologies often mould into the forms of the exhibitions they showcase (Fig. 1). The growing realms of the digital age within museums therefore, implicate upon the epistemology of representation, and as most museum practitioners would concur, aid self-reflexive curation and curatorial practices.

Yet, in celebrating the value of digital technologies, we often tend to forget that the core aims that sustain and nurture their myriad creations are very similar to those that were foundational to the establishment of museums, and curatorial practices; namely, the access, preservation and increase of knowledge. By the late eighteenth century museums heralded the transformation of cabinets of curiosities into public objects, and intellectual moves towards more scientific collecting practices. Subsequent

curation of objects and collections demonstrated that the ordering and storage of information required complex decisions regarding attributions of authorship and custodianship, creations of classification schemes and taxonomies, and policies of preservation and access. An analytical view of the categorization of information, data and knowledge, is of significance to critically review the shifts in knowledge making processes over time, and the relational encounters between robust curatorial practices and digital media.

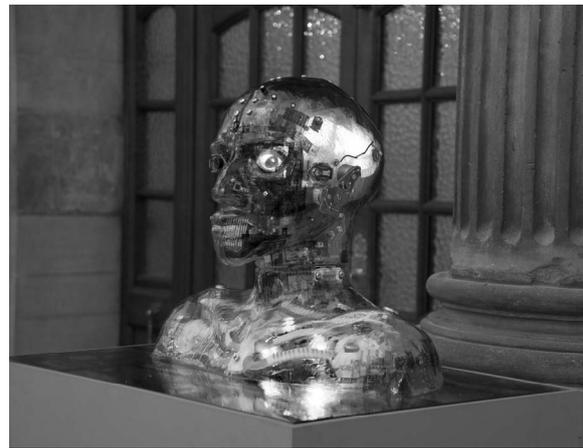


Figure 1. Head of the Blue Chip II. Dianne Harris, 2009. Mixed media, electronics. © MAA.

Within the last decade, the moves made by museums towards digitization have greatly impacted upon the sociality of museum worlds. They have brought to fore the different requirements of the collections of the humanities from those of natural sciences, and the need for devising methodologies that facilitate synergy between them. The increasing presence of digital technologies within the curatorial domain also highlights an intrinsic feature of all technologies, namely, that they only ‘work’ within philosophies. Through case studies from the Museum of Archaeology and Anthropology at the University of Cambridge (henceforth MAA), this paper focuses upon the challenges faced by curatorial practices in their encounters with the digital age. The challenges have referential values for projects such as ECLAP, which create and nurture large ‘public-centered’ digital repositories, often through the premise that those creations are potentially neutral acts. The examples document the reception of digital projects as sites of

engagement and experiences, and encourage reflection upon the promises of digital technologies beyond notions of use values and exploitation.

Case I: A Digital Research Environment

MAA was formally established in 1884 when ‘two Cambridge men’, Arthur Gordon (1829–1912, the first British governor of Fiji), and Alfred Maudslay (1850–1931, British colonial officer, archaeologist, collector, biographer) deposited their large collections of artifacts from the South Seas (see [3]). The histories of MAA’s collection management represent the ways in which objects, collecting practices and collections have fed into creations of anthropological and archaeological knowledge during the twentieth century. The exhibitions since the centenary, in 1984, have specifically focused on the

Figure 2. A page from MAA’s photo catalogue

practices of ‘new museology’ (for definition see [6]), which has entailed ‘opening’ collections for eliciting wider public participation within museum projects. Through digital catalogues, MAA has also created vibrant technological interfaces for structuring information regarding aspects of collections (Fig.2).

Yet, new knowledge, and new forms of sociality and relationships rarely emerge through exploitations of new technologies, however creative the powers of the latter may be. Data curation in the digital world creates an acute awareness of the constructed nature of culture heritage, and MAA’s examples inform us of some of the ways in which this is framed through practices and ideologies. They also show that source communities ‘do not wait passively ‘out there within the public body to be identified and included’, and illuminate on the ways in which communities are redefined and re-constructed through experiences, and engagements ([12]: 108; also [1] 2002). In ‘presenting’ source communities with artifacts of their cultural heritage curatorial practices show rather clearly that digital technologies often contribute to ‘facts’ about heritage, enable greater democratization in decisions of heritage-making, and greatly increase the scope of dissemination. Technologies allow in uniting material and collections,

including those that are conceptually dispersed, for fostering heritage, and MAA’s ‘Artefacts of Encounter’ (<http://maa.cam.ac.uk/aofe/about.html>), an ambitious project launched in 2010, provides significant insights.

The project attends to the ontology of ‘thinking through things’ (Henare, Holbraad and Wastell, 2007), and connects data in many different ways to construe the Pacific of the late eighteenth century as a cultural artefact of diverse encounters. It has demanded a creative digital research environment, designed as KIWA, that enables the sharing of data and research among a geographically dispersed project team (based in the UK, New Zealand, Australia and Brazil) and between dispersed institutions holding related collections. A foundational component is the building of a ‘digital repository of the Hauri Tonga (treasured ancestral artifacts) that are held in museums internationally to support the Maori iwi [communities]’ (website). In this, the project has a larger ambition than some other contemporary creations of uniting object collections and documentary archives to which MAA lends support, one being the ‘VKS archive’ at the Vanuatu Cultural Centre (on this [4]). However, in facilitating live-feed, and striving for democracy in forging relationships between communities and their ‘ancestral’ objects, the database Te Rauata (maori for ancestor) is similar to the VKS in eliciting description entries, permitting the modification of earlier catalogue entries, facilitating continuous restructuring of templates for different uses, and for inputting information in various forms and manner. With respect to the exhibition of information, the databases ‘hide’ and ‘show’ that which is considered ‘appropriate’, and is one among many others that are being increasingly created for museums with substantial anthropological collections. Such databases represent the participation of digital networks in cultural codes of place, secrecy, information and knowledge, and within this context the Te Rauata holds saliency as a cultural object.

‘Artefacts of Encounter’ has been conceptualized as ‘digital subjects, cultural objects,’ (conference of the same name was held in 2011) and on its completion in 2013, this may, indeed, be the project’s legacy. The point, however, is that the project relates to information technologies besides their parameters of utility. The digital research environment which it fosters, is rather similar to all other research environments in that it thrives upon the targeting and nurturing of potential interests. Designed by specialists of the Pacific, who are well aware of the implications of investing in, and receiving, custodial rights and authority, KIWA and Te Rauata, encourages us ask the question ‘who speaks for culture?’ when we celebrate the innovations of information technologies for serving the cultural heritage sector. For, they show that technologies function because of social interconnections. We are guided to look beyond notions of ‘exploitation’ for gauging the kinds of cultural sensitivity information technologies permit, and for seeking the ways in which they redefine the ‘public’, create a sense of sociality, draw people in towards

knowing, and encode different imaginaries within museum managements of culture.

Case 2: Questions of Translation: Protocols and Access

Alana Jelinek, artist in residence at MAA (2009–2014), shares the belief of many peers in reckoning that ‘since performance is about life experiences in a specific context, recording can be an abstracting and distilling act, which can potentially engender de-contextualization’ (personal communication, 23 February 2011). Jelinek is apprehensive (quite rightly) that meanings may be lost, and unintended meanings reified in acts of archiving—aspect of recording—and cites instances of the obliteration of audience responses as examples. Like most contemporary artists, she separates her practices carefully, and creates specific artforms for the web that can negotiate its 2-D, pixallated visual format, and the one singular characteristic of the internet which gets overlooked within celebrations of its reach, namely, the low sound quality (see <http://www.alanajelinek.com>). Although, Jelinek’s apprehension of ‘loss’ may sound unduly harsh to archivists, increasingly sensitive to unintended slippage of information, and to transformations of experiences as archival copies, it conveys the force of translation in the semiotic of digital technology. The digital is a continual process of translation by which data is converted into binary information, and reconverted into multiple representations. Technologies of replication, such as computers and web sites, are now intrinsically linked to ideas about increasing access to knowledge. Yet few studies engage with the ways in which transactions over knowledge occur between cultures, and across chronologies. In this respect curatorial practices with anthropological intent contribute to instances of alterity in reproductions of knowledge.

A recent article on the overlaps between Zuni and Euro-American approaches to reproduction of knowledge unfolds the promise of technologies in exploring the relationships between materiality and orality [8]. The MAA’s ‘World Oral Literature Project’, which has been developed since 2009, creates a more comprehensive opportunity for focusing on such relationships. The project documents and makes accessible ‘endangered oral literatures before they disappear without record’ and ‘collaborates with local communities to develop their own oral narratives’ (<http://www.oralliterature.org>). The collection of films and audio recordings, of interviews, dances and acts of story-telling, elucidates the social force of materiality in transmissions of oral literature which mutate—fuse, change and modify—into the ‘born digital’ collections that are taken back to the ‘ethnographic field’ for local custodianship (Fig. 3). The project was established following the success of ‘Digital Himalaya’, which involved the digitization of the early-twentieth-century photographs from the Himalayan states of Bhutan, Sikkim and Tibet, in MAA’s Williamson Collection. This start-up project (<http://www.digitalhimalaya.com>) continues to remain theoretically useful for probing into

relationships between digital replications and productions of cultural knowledge.



Figure 3. Shaman Dirga Bahadur Dumi reciting for the ethnographers (Baksila, Khotang District, Nepal. Photo: Marion Wettstein, 2006).

Translations are dependent upon standardized protocols that are derived through understandings of best practice. However, research on aspects of MAA’s catalogues has shown that the elicitation of criteria for best practice is not a simple matter of arriving at consensus with chosen professionals and select members of user communities. A project hosted by the A:shiwi A:wani Museum and Heritage Centre of Zuni (New Mexico, USA), which focused on collections of Zuni objects in MAA, has succinctly demonstrated the inadequacy of the MAA catalogues in offering sufficient descriptions of the objects [13]. The researchers report that ‘the entries from the MAA catalog for the objects we included in our study are typical of what is usually found in a museum’s catalog with “name”, “description”, “material”, and “notes” containing the most illustrative information about each object. ... However, a simple comparison between the descriptive information in the Cambridge catalog and the data collected during our interview shows that for the most part the Cambridge catalog entries are sparse, clinical, and use highly specialized language to describe everyday objects.’ (Ibid: 751). The creation and dissemination of standardized protocols is ‘ontologically [...] the heart of the story’ of the internet’, and of its ‘technical configuration and modes of ordering’ [9: 306]. Moves towards democratization by museums, which encourages the reasons for the scrutiny of MAA’s catalogue have meant redrawing protocols for greater openness of catalogued entries to modifiability. Tagging, objectwikis, folksonomies and other means of creating ‘networked objects’ are seen as the panacea. However, specific curatorial guidance is mandatory for projects that engage with people, and they often replicate structures of authority that tagging and other ‘utopic visions’ of openness try to dismantle (Geismar 2012). Besides, MAA’s Facebook-wall exemplifies rather well that tagging is often a form of self-representation, used mainly for self-promotion. Social tagging has yet to be an act of intervention within museum cataloguing and selection procedures, and museum practices show that the celebration of the democracy of digital technologies

needs qualifications, as democracy comes with responsibilities.

A recent history of access in India of two photograph collections in the UK, at SOAS (School of Oriental And African Studies) and MAA, also alerts us to the fact that open access and greater sourcing neither equate as immediate, nor are covalent, since user communities often show preferences for curatorial guidance in their navigation of institutional archives. Both institutions gave access to their vast photographic collections for an exhibition designed and curated by adivasi communities at Tejgadh (Gujarat, India). SOAS agreed to lend images from the Fürer-Haimendorf Collection, which has been fully digitized in recent years and is open to public access on the web. MAA sent copy prints of select photographs from the recently acquired Archer Collection, and was unable to emulate the kind of access that SOAS gave, as this Collection has yet to be digitized completely. The SOAS material was, however, not used at the Tejgadh exhibition, while images from the undigitized Collection at MAA (and Grassi Museum für Völkerkunde, Leipzig) were selected and displayed in January 2012. As Mark Elliott, one of the project coordinators, suggests this could be because the Fürer-Haimendorf Collection appeared to the adivasis as too vast an archive for comprehensive perusal, or perhaps because the MAA and Leipzig images appeared to them as mediated—through personal exchange between individuals. ‘Open access’ to images online, can therefore, appear alienating, and in this aspect they do not replace the intellectual force of curation.

Case 3: The Crafting of Heritage as Digital Objects

A digital project in which MAA participates, and which demonstrates ‘a fundamental shift in the way cultural heritage research is conducted’ is the Reciprocal Research Network (<http://www.moa.ubc.ca/RRN>). The RRN has been specifically designed to link catalogue information and collection databases of all partner institutions to communities via research hubs, and therefore utilizes the digital to create openness—it has recently launched a mobile version of its website, optimized for phones, tablets and older browsers—but is not openly public. The networking serves the cultural heritage of the First Nations and ‘creates collaborative relationships between communities, researchers, institutions where [their] culture heritage is held.’ It allows comments, discussions, and critiques of communities who are unable to access the collections in situ. GRASAC (Great Lakes Research Alliance for the Study of Aboriginal Arts and Cultures, https://grasac.org/gks/gks_home.php), is another digital networking platform which hosts a selection of objects of MAA from the Great Lakes, and shares aims that are similar to RRN. Both create a methodology of ‘open source’ that builds upon notions of cataloguing.

Digital objects allow different ways of approaching questions of ontology and epistemology. The understanding that the mutability of digital objects is not



Figure 4. Santhal Women showing the poses of a heron or vulture during the ‘Golware’ dance. Bihar, India, Archer Collection, photo: William Archer, ca. 1926–39. MAA P.111910.WARC.

‘just the mutability of its representation, but of its agency’ has been charted in depth during the ‘Zuni project’ where the use-value of digital objects were vested by the Zuni in teaching and learning [13: 754]. In debates regarding originals and copies, which theorize on cultural authorship and values, we find that digital materiality can accrue many different receptions from stake-holders. The portrait images of digital prints from the Archer Collection were received as ‘our ancestors’ by adivasis of Tejgadh, who show no qualms about the prints being a digital copy (Fig 4). The status of the ‘copy’ has been a site of much post-colonial angst in the Western world. However, the above receptions of digital objects lend a balm to intellectual expressions of this angst. The value of the copy, as Barbara Kirschenblatt-Gimblett [10] has succinctly shown through examples of the Museum of the History of the Polish Jews and the Jingu Shrine in Ise (Japan), points to a very different tradition of thinking about what constitutes or is defined as an ‘actual’ object (2009). Digital objects allow us to rethink the ‘conceit of the copy’ [5], which inundates post-colonial histories and cultural theories of heritage and its management (e.g. Ibid).

IN CONCLUSION

The MAA’s projects clearly demonstrate that information ‘out there’ does not necessarily translate as accessible knowledge. In this respect, the dictates of ‘information first analysis later,’ often expressed for embracing the causes of the Semantic web and its scheme of linked data, appear hollow. The projects also show that collections are inevitably enriched when their custodians devise shared thesauri and dialogic documentary catalogues, are accommodative of live feed, and create social networking outlets. However, foundational to their enrichment and preservation are responsible formulations of policies, which are derived from the understanding that a) culture and heritage are constructs, b) their notions, descriptions and constructs shift and change through time and over geographies, and that c) digital technologies perform in the manner in

which we make them perform. The infinite ways in which the digital media encourages the tagging of entities makes it equally mandatory for refining judgements regarding the quality of data. Quality implicates intellectual property and traditional cultural rights (see [2]) and the ‘progress’ of MAA’s projects illuminates that heritage is as much a local matter as a global one.

The proliferation of information that now attends the digital age makes us sensitive to distinctions between data and knowledge. In this respect, curatorial practices are informative guides of the ways in which we achieve, and think through notions of classifications and devise taxonomical schemes. Protocols, such as Web Description Resources, which sustain the technologies, are usually established and managed by select groups and institutions. Hence, significant changes in their characteristics are usually indices of replacements of hegemony. Although the promise of empowering those whose ancestors once made and used the objects guides all MAA’s projects now-a-days, the histories of the Museum’s collections, which are imbued within the politics of imperialism, paradoxically poses as constraints upon the meaning of open access. Curating data in the digital age therefore creates a critical outlook towards the manner in which information is established, nurtured, shared, presented and preserved, without transgressing the rights of those who make claims on them as cultural heritage, or, conversely, by acceding to all kinds of claims that are made in the name of culture and heritage.

- [1] Alleyene, B., 2002, ‘An Idea of Community and its Discontents: Towards a more Reflexive Sense of Belonging in Multicultural Britain’, *Ethnic and Racial Studies*, 25 (4), 607–27.
- [2] Boast, R., 2011 ‘Inappropriate Use: Intellectual Property Rights And Traditional Cultural Material’ (ECLAP Workshop, Brussels, June).
- [3] Ebin, V., and Swallow, D.A., 1984, *The Proper Study of Mankind: Great Anthropological Collections in Cambridge*, MAA.
- [4] Geismar, H., 2012, ‘Museum + digital = ?’ in D. Miller, and H.A. Horst (eds.) *Digital Anthropology*, Berg Publications, Oxford (forthcoming).
- [5] Guha-Thakurta, T., 2011, ‘Conceits of the Copy: Traveling Replicas in Colonial & Contemporary India’, lecture at Smith College, 18 April.
- [6] Hooper-Greenhill, E., 1992, *Museums and the Shaping of Knowledge*, Routledge, London, New York.
- [7] Isaac, G., 2008, ‘Technology Becomes the Object: The Use of Electronic Media at the National Museum of the American Indian’, *Journal of Material Culture*, Vol. 13, pp. 287–310.
- [8] Isaac, G., 2011, ‘Whose Idea Was This? Museums, Replicas and the Reproduction of Knowledge’, *Current Anthropology*, Vol. 52, no. 2, pp. 211–33.
- [9] Kelty, C.M., 2008, *Two Bits: The Cultural Significance of Free Software*, Duke University Press, Durham.
- [10] Kirschenblatt-Gimblett, B., 2009, ‘Materiality and Digitization in the Museum of the History of Polish Jews’ (http://blogs.nyu.edu/projects/materialworld/2009/07/materiality_and_digitization_i.html)
- [11] McLemee, S., 2008, ‘Its All Geek to Me,’ *Inside Higher Ed*, <http://www.insidehighered.com/views/mclemee/mclemee84> (accessed 2 March 2012, 1508).
- [12] Smith, L., and Waterton, E., 2009, *Heritage, Communities and Archaeology*, Duckworth, Wiltshire.
- [13] Srinivasan, R., Becvar, K., Boast, R., and Enote, J., 2010, ‘Diverse Knowledges and Contact Zones within the Digital Museum’, *Science, Technology and Human Values*, Vol. 35 (5), pp. 735–68.

MIMESIS: A Semantic-based Approach for a Digital Library of Poetry in Music

Application to the Repertorio della Poesia Italiana in Musica 1500-1700 (RePIM) Repository

Ilaria Bartolini⁽¹⁾, *Cecilia Luzzi*⁽²⁾, *Marco Patella*⁽¹⁾, *Angelo Pompilio*⁽³⁾ and *Cesarino Ruini*⁽⁴⁾

(1) DEIS, University of Bologna, Bologna, Italy, {i.bartolini, marco.patella}@unibo.it

(2) Liceo Musicale of Arezzo, Arezzo, Italy, cecluzzi@tin.it

(3) DISMEC, University of Bologna, Bologna, Italy, angelo.pompilio@unibo.it

(4) MUSPE, University of Bologna, Bologna, Italy, cesarino.ruini@unibo.it

Abstract— In this position paper we discuss matters relative to the development of a semantic-based model for an on-line digital library of poetry in music. In particular, we present a general approach called MIMESIS (Multimedia Interoperable Model of poEtry and muSic Sources). Using semantic technologies and methodological solutions, MIMESIS enables different types of users to creatively enjoy digitized cultural resources, not only in the specific case of poetry in music, but also in the general field of the performing arts. The MIMESIS approach is then applied to an existing repository, RePIM (*Repertorio della Poesia Italiana in Musica 1500-1700*), which represents an interesting test case for our study. Other state-of-the-art initiatives utilizing semantic Web technologies for digital collections exist, and some show some encouraging results. Yet although these projects represent an improvement over previous systems, they still fall short in terms of complexity and usability. MIMESIS is built on recent definitions of ontology for cultural heritage, such as FRBRoo (Functional Requirements for Bibliographic Records object-oriented), and explicitly addresses issues of interoperability with current European initiatives.

Keywords: *Digital libraries, Semantic web, Performing arts, Poetry and Music (1500-1700)*

I. INTRODUCTION

The increasing adoption of digital multimedia technologies is producing massive amounts of cultural content. Wide-ranging initiatives, such as *Europeana* [1] and the realization of a *European Digital Library* (EDL) [2], promise to deliver access to digitized objects, including books, photos, maps, audio, films, and archival records from Europe's libraries, archives, museums and audio-visual collections.

The realization of this new global network of resources has quickly become an inspiration and source for new learning scenarios and challenges. For example: how to handle large sets of raw and edited reference material, and local and remote multimedia content? How to

extract actionable meaning from structured and unstructured information and social interaction patterns? How to provide functionalities across different content types, improving access to information for humans and machines alike? How to enable access to, and engagement with, rich distributed information sources?

Despite covering huge amounts of digital content, these projects have not managed to achieve the anticipated benefits and exposure. Widespread access to these collections and an appropriate level of visibility have been compromised by confusing standards and changing technologies. Many Web sites have poorly developed search and display functions that are difficult to use and, in many cases, fail to appeal to users; and often, multilingual access is not well-developed.

The MIMESIS (Multimedia Interoperable Model of poEtry and muSic Sources) approach discussed in this position paper is aimed at overcoming these limitations. A semantic-based method for developing an on-line digital library model for poetry and music multimedia, MIMESIS is built on recent definitions of ontologies for cultural heritage such as FRBRoo. The MIMESIS approach is applied to an existing repository, *Repertorio della Poesia Italiana in Musica 1500-1700* (RePIM), a digital archive of poetry and music from 1500 to 1700, comprising analytical descriptions of the literary and musical sources of circa 43,000 texts set to music (for a total of about 66,000 musical settings), and digital items, such as images from historical source material and modern editions of the texts, and the music for samples of particular significance.

The MIMESIS approach allows for the development of methodological solutions that enable all users to explore digitized cultural resources, and that explicitly address issues of interoperability with other ongoing European initiatives.

With its strong interdisciplinary nature, MIMESIS bridges the boundaries between cultural heritage, poetry, music and the performing arts, new information technologies and digital libraries. MIMESIS must be considered a pilot project, aimed at achieving the following objectives:

- to ascertain FRBRoo model suitability through the case study under consideration;
- to develop access to digitized content;
- to enable the user to make full use of digitized cultural resources with powerful, streamlined, and creative search and display functions.

The rest of the paper is organized as follows: Section II offers an overview of related work; Section III describes the MIMESIS approach; Section IV concludes the paper.

II. THE STATE-OF-THE ART

A. Previous work

Current efforts in digitization have resulted in the creation of international, unified portals which provide access to individual libraries, archives, and museums.

- The *World Digital Library* (WDL) [3], a project fostered by the Library of Congress and sustained by UNESCO, provides Internet access, free of charge and in multilingual formats, to significant primary materials from countries and cultures around the world.
- The *Multilingual Inventory of Cultural Heritage in Europe* (MICHAEL) [4], a ground-breaking project funded by the European Commission, provides access to digital collections from European archives, museums, and libraries.
- *Europeana*, the ongoing European portal launched as a proof of concept in 2008, provides access to as many as two million books, photographs, maps, sound recordings, films, and archival records from libraries, archives, museums, and audio-visual collections, that have been digitized throughout Europe. Its aims are to provide access to all of Europe's digitized cultural heritage by 2025, to improve the quality of the metadata – e.g. detailed description of a digital object's contents – and to make content searches easier [1].

Other important digital collections exist within the Web sites of national cataloguing organizations such as Italy's *Internet Culturale* [5], or large national libraries, such as *Gallica* of the Bibliothèque Nationale de France [6], *Online Gallery* of the British Library [7] and *Münchener Digitalisierungs Zentrum* of the Bayerische Staatsbibliothek [8]. Further examples are *Digital.csic*, the digital library of the Spanish Consejo Superior de Investigaciones Científicas [9]; and *Early European*

Books [10] – a digitalization project of printed books in Europe from their earliest origins through the close of the seventeenth century. *Early European Books* complements *Early English Books Online* (EEBO), collected by Biblioteca Nazionale Centrale (Firenze), Det Kongelige Bibliotek (Copenhagen), Koninklijke Bibliotheek (Den Haag) and the Wellcome Library (London).

At present, the most urgent challenge facing the digital library field is the demand for powerful and efficient systems which allow for integrated access to vast amounts of digital content. The current available repositories have not fulfilled expectations from this point of view because of limits in resource description (e.g., *Europeana*, based on Dublin Core) and interoperability (proprietary format). In addition, these initiatives feature poorly developed search and display functions which are difficult to use and which, in many cases, fail to appeal to users. The massive amounts of cultural content in these repositories cannot be captured through the traditional systems, typically based on keywords, that are commonly used to search the Web. Instead, this material could be made more accessible to the user by means of semantic digital library technologies.

B. Semantic digital library technologies

III. The recent definition of ontologies – those concepts and relationships used to describe and represent an area of knowledge – for cultural heritage developed by the International Council for Museums-International Committee on Documentation (ICOM-CIDOC) and the International Federation of Library Associations and Institutions (IFLA) provide an important contribution to this undertaking.

The CIDOC Conceptual Reference Model (CRM) supplies definitions and a formal structure for describing concepts and relationships used in cultural heritage documentation. It promotes a shared understanding of cultural heritage data by providing a common and extensible semantic framework by which any cultural heritage information can be mapped. In this way, CIDOC CRM fills in the gaps that otherwise occur in large collections of cultural heritage data gathered from various sources, such as museums, libraries, and archives [11].

The FRBR model [12] was originally designed as an entity-relationship standard by a study group appointed by the IFLA [13] during the period 1991-1997, and published in 1998 [14]. FRBR specifies that intellectual or artistic products include four types of entities. In “Group One” entities, the FRBR model distinguishes between immaterial (“work” and “expression”) and material (“manifestation” and “item”) entities. Work and expression are abstract concepts; the manifestation and the items are concrete objects related to them. In particular, “work” is a distinct intellectual or artistic

creation; and “expression” is an intellectual or artistic realization of a work. A “manifestation” is the physical embodiment of an expression of a work; and finally, an “item” is a single exemplar of a manifestation. FRBR also specifies particular relationships between classes of “Group One” entities: a work is realized through one or more expressions, each of which is embodied in one or more manifestations and exemplified by one or more items.

Initial contacts, in the year 2000, between the two communities (CIDOC and IFLA) eventually led to the formation, in 2003, of the International Working Group on FRBR/CIDOC CRM harmonization. In May 2009, this Working Group published a final complete draft version of FRBRoo; that is, the object-oriented version of FRBR, harmonized with CIDOC CRM [15]. The common goals were to express the IFLA FRBR model with the concepts, ontological methodology, and notation conventions provided by the CIDOC CRM; and to merge the two object-oriented models thus obtained [16].

FRBRoo is a formal ontology intended to capture and represent the underlying semantics of bibliographic information as well as facilitate the integration, mediation, and interchange of bibliographic and museum information. Such a common view is necessary to provide interoperable information systems for those users interested in accessing common or related content.

FRBRoo is an exhaustive ontology that goes beyond the expressive limits of existing systems such as *Europeana*; though, quite obviously, at the expense of a greater complexity. Until now, this model has not been implemented in actual digital collections, such as libraries, museums, archives, research centers, music music stores, and record labels.

C. *Relevant projects*

Presented below are projects that apply semantic technology to the field of music; efforts which, in many cases, are still at an embryonic stage.

EASAIER (Enabling Access to Sound Archives through Integration, Enrichment and Retrieval) is a European-funded project aimed at meeting the challenges of the increasing amount of digitized audio and audio-visual material in archives across the UK and Europe.¹ Many digital sound archives still suffer from issues of accessibility. The EASAIER system has been designed with sound and broadcast archives, libraries, museums, and music schools in mind. It has developed access, retrieval, and interactive software – license free – in direct response to the needs of its proposed users. These users have been defined as anyone, amateur or

professional, interested in accessing archived data regardless of the type of archival material involved [17].

FictionFinder is an FRBR-based prototype that provides access to over 2.9 million bibliographic records for fiction books, eBooks, and audio materials described in WorldCat, a union catalog – or combined library catalog describing the collections of a huge number of libraries from all over the world which participate in the Online Computer Library Center (OCLC) global cooperative [18].

Music Ontology is a project carried out by Yves Raimond.² It provides main concepts and properties for describing music (i.e. artists, albums, tracks, but also performances, arrangements, etc.) in the semantic Web vocabulary, linking a wide range of music-related information in sharing systems like Napster, communities like MySpace, music services like Last.FM or music stores like iTunes. With a goal of helping to create a music-related web of data, Music Ontology represents an online community effort to express music-related information on the semantic Web [19].

Variations3 is a project of the Indiana University Digital Library Program. Developed a digital music library software system for online access to streaming audio and scanned score images, Variations3 features a flexible access control framework that ensures respect for intellectual property. It also includes analysis and annotation tools which are useful for music teaching, learning, and research. In addition, the Variations3 project is conducting research on a music metadata model and search system centered on the notion of the musical work, which, as compared to traditional catalog systems, improves the music search experience for users [20].

The Digital Library for Poetry in Music (1500-1700) project could benefit considerably from the use of FRBRoo. This object-oriented model could be used not only as a point of departure for the definition of specialized ontologies in the domains of poetry and music, but also for the exhaustive description of the intellectual content of poetry and music repertoires. The groundbreaking feature of this project will be the application of the theoretical model in order to assess the potential of FRBRoo for providing access to the network of digital projects related to poetry and music from 1500 to 1700.s

MIMESIS will seek to develop methods that put relatively abstract models of FRBRoo into practice. In addition, MIMESIS aims to demonstrate how these models may be connected to detailed domain ontologies

¹ This project was funded by the European Commission under project EU-FP6-IST-033902, coordinator Josh Reiss (Queen Mary, University of London).

² Raimond carried out the Music Ontology project as a Ph.D. student at the Centre for Digital Music, Queen Mary, University of London; and, from 2008, as Senior Technologist at BBC Research & Development.

(such as those for poetry set to music, and the performing arts), integrating various types of content under these standards within a system that enables advanced semantic Web reasoning methods.

III. MIMESIS APPROACH

In practice, the MIMESIS approach will be realized through the implementation of a framework consisting of software tools and the population of a domain knowledge base. It will comprise bibliographical data and digital items relative to Italian poetry by Petrarch and Petrarchist poets set to music in Europe between 1500 and 1700, making the RePIM database available online.

RePIM, a digital archive of poetry and music from 1500 to 1700, represents an ideal setting for the implementation of FRBRoo in a delimited field. As a case study, the use of RePIM restricts complexity while also providing information on the feasibility and scalability of the MIMESIS system.

In the late 1970s, the musicologist Lorenzo Bianconi conceived the idea of a checklist as a means of making an analytical description of Italian poetry set to music from 1500 to 1700. In collaboration with Antonio Vassalli and Angelo Pompilio, the checklist was drawn up on paper support in the 1980s. It was later developed into a database – the *Repertorio della Poesia Italiana in Musica 1500-1700* (RePIM) – edited by Pompilio.³ The project attempted to identify as many authors of poetic texts set to secular and spiritual vocal music from 1500 to 1700 as possible. And, given that music sources rarely reveal the author of poems set to music, this required a systematic examination of the printed poetic sources of the period. In the first decade of research, the project produced a bibliography of about 2,000 printed literary sources. This allowed the identification of the sources and authors of about 7,000 poems set to music, and prompted the creation of a first checklist on paper support [21].

The development of a database in the early 1990s allowed for better management of the collected information.⁴ This database comprises circa 43,000 texts (variant forms and the single parts of a text divided into more sections have been included in this estimate), for a total of about 66,000 musical settings. Of the 43,000

texts set to music, approximately 13,000 have been identified through citation in roughly 1,500 literary sources. The electronic collection consists of two different parts: the bibliographical data of the sources of the poems, which are described in detail; and the data relative to some 3,500 music sources.

For the poetry sources, the database provides a complete transcription of the title page, imprint, pagination, and shelfmark of the exemplars consulted; descriptions of introductory texts (dedications, prefaces, printer's notes); the list of poets; a brief summary of literary content; and analytical descriptions of the poetic texts (limited to those set to music). For each text an extended incipit comprising at least the first two verses (and, often, three or four verses; in a few cases, the complete text is reproduced) is provided. In addition, the author's instructions, poetic meter, and the transcription of any titles and inscriptions are recorded. The data referring to music, extracted from existing repertoires [22], revealed some gaps or incorrect transcriptions of the incipit. These therefore require analytical verification of content directly from the sources.⁵

For the musical editions, direct consultation of sources, in original or digital format from archives and libraries throughout Europe, is foreseen. This will ensure that an analytical examination and verification of content be made for each item. An analytical description comprising source content will be provided for each of the approximately 3,500 musical sources indexed. For the moment, the database contains only data relative to the description of the content: extended incipit comprising at least the first two verses, poetic form, musical form, composer, eventual inscription or title of the composition, voices and instrumentation.

The MIMESIS approach will permit researchers to trace the tradition of sixteenth to eighteenth century poetic texts set to music – which have never been studied by

³ The project was mainly implemented at the Dipartimento di Musica e Spettacolo of the University of Bologna: in different phases it has benefited from financial, logistic, and operative support provided by other organizations, such as the Fondo Nazionale Svizzero per la Ricerca Scientifica (FNS), the Italian Consiglio Nazionale delle Ricerche (CNR), Ministero Italiano dell'Università e della Ricerca (MURST, after MIUR), Istituto di Studi Rinascimentali di Ferrara, Dipartimento di Storie e Metodi per la Conservazione dei Beni Culturali of the University of Bologna (seat of Ravenna).

⁴ The database project was realized by Angelo Pompilio, with Thomas Walker and Silva de Marchi; and implemented by Livio Aragona, Cecilia Luzzi, Gianmario Merizzi, Roberta Ziosi and Concetta Assenza.

⁵ The examples of incomplete descriptions of the sources regard mainly the seventeenth century books for more than one voice transcribed by Emil Vogel in which the content was derived only from the table of contents of a single voice's partbook and therefore do not include some compositions. See, for example, the Nuovo Vogel index of *Madrigali concertati a due, tre, quattro e cinque voci*, Venezia, Vincenti, 1627, by Domenico Obizzi (NV 2046) where seven musical settings are lacking (*Poich'a tanti tormenti; Filli, se vuoi ch'io mora; Se sete Beatrice; Porta sul biondo crine; Come poss'io mirare; Mentre la bella mia nemica altera; Se de' miei giusti prieghi*). There are also incorrect or inaccurate transcriptions of the first lines, such as *Dolei leggiadri e belli* instead of *Occhi leggiadri e belli* (NV 2552) or *O speranza fallace a che turbar mia pace* instead of *O speranza fallace dunque sperando a che turbar mia pace*, NV 1884; *Sol per fuggir da voi mio sole amato* instead of *Sol per fuggir Amore da voi mio sole amato* (NV 2235). In other cases, Vogel completed the first line in the table of contents by referencing similar lines of other works: *O d'amor meraviglia, quel dolce canto* instead of *O d'amor opre rare col torme da me stesso* (NV 1534); *Se il dolce sguardo di costei m'ancide* instead of *Se il dolce sguardo del divin tuo volto* (NV 2775); *Cigni bianchi e canori del Mincio a cantar meco* instead of *Cigni bianchi e canori cantate i casti amori* (NV 3002).

philologists because they were contained in musical sources – by basing itself on the FRBRoo model requisites. For each “work” (abstract work), the model will allow for the identification of its morphological anatomy (horizontal scan: the articulation in parts, sections, stanzas, quatrains, tercets, etc.) and its tradition (chronological scan: the different versions and forms the work has assumed over time).

An ontology which satisfactorily accounts for the tradition of poetry set to music from 1500 to 1700 must address the variable nature of the sources which make identification of the poems difficult. The text set to music can vary from the archetype attested to by literary or musical sources. It may be broken up and recomposed; the order of the lines may be changed; the text may be excerpted arbitrarily from a longer poem. Composers, or even poets, may replace one or more words, introducing significant or minor variants. They may also substantially re-arrange a part of the poem. We can offer some instances which better illustrate this variational nature: the madrigal *Chi volesse saper che cosa è Amore* set by Cimello (NV 578) is a cento of the first lines of settings collected in *Il primo libro de madrigali de diversi autori* (RISM 1542¹⁷). The madrigal by Jachet Berchem *Signor, Lidia son io* (NV 328) omits the first part of the first line of Ariosto’s poem *E cominciò – Signor, Lidia son io* (Orlando furioso, XXXIV,11). The canzonetta *Mia Filli, io non so dire* set by Tommaso Pecci (NV 2164) is a reworking of the madrigal by Battista Guarini published in two different versions: *Mia diva, io non so dire* (in *Rime de’ diversi celebri poeti dell’età nostra*, Bergamo, Comin Ventura, 1587) and *Dov’hai tu nido, Amore* (in his *Rime*, Venezia, Ciotti, 1598). The madrigal *Sospir, dimmi, dimmi che fa quel core*, set to music by Pietro Paolo Torre (NV 2736), is a shortened version of Giovan Battista Marino’s madrigal *Sospir che dal bel petto*.

With this repertory, which contains so much descriptive data relative to this particular poetic and musical tradition, it will be possible to query numerous digital collections at the same time, and access the content on the Web. In addition to the names and professional details of the poet, composer, dedicatee, editor, etc., descriptions of each abstract work, understood as the intellectual or artistic creation free from material entities, and of all its particular manifestations (the various versions of the text) are available. Bibliographic descriptions of the poetic and musical collections to which the various manifestations of the work belong are included; as well as references to single examples of the various manifestations, and to the contents themselves (digital images of sources, transcriptions of the poetic texts, musical texts, audio files in MIDI and MP3 formats). The repertory also comprises a list and description of the various libraries in which single examples from the poetic and musical collections are

conserved; and, of course, relative bibliographic and discographic information.

As an example of the kinds of digital collections to which MIMESIS will provide access, it is sufficient to cite not only the impressive digital library projects mentioned earlier – *Europeana*, *World Digital Library*, *MICHAEL*, *Internet Culturale*, *Gallica*, *Online Gallery*, *Münchener Digitalisierungs Zentrum* and *Early European Books* – but also some projects from the area of lyric poetry studies. *Biblioteca Italiana* (BibIt) [23], *Liberliber* (Progetto Manuzio) [24] contains versions of texts representative of Italian literature from the Middle Ages to the twentieth century. *Antologie della Lirica Italiana* (ALI) [25] of the University of Pavia collects information about anthologies of Italian lyrical works from the sixteenth to the eighteenth century, both handwritten (MAMIR) and printed (RASTA) [26]; and is analogous to the University of Torino project, *Antologie di rime del Cinquecento* [27], [28]. With reference to music, the enormous quantity of digital musical content published online by European national libraries is complimented that of specialized libraries. The Museo internazionale e biblioteca della musica di Bologna [29], for example; and DIAMM (Digital Image Archive of Medieval Music) [30], a portal to information about, and images of, medieval polyphonic music manuscripts dating from circa 800 to 1550 gathered from collections throughout the world. MIMESIS will make it possible to integrate all of the content disseminated throughout the Internet (which may include versions of poetic texts, musical scores, and audio files in MIDI or MP3 format found in music sharing systems, Web communities, or stores) relative to a single theme.

To support interoperability with existing and emerging standards, the MIMESIS approach will develop methods for semantic mapping and tools for interconnections, taking into consideration international and national standards for metadata of cultural artifacts. It will deliver an authoritative and comprehensive semantic knowledge base in the domain of poetry set to music. At project completion, the Repository will contain a domain-specific knowledge base, together with the semantic representation of work, expression, and segmentation of all poems set to music from 1500 to 1700; and of an Italian secular vocal music repertory of these two centuries.

This represents a huge step forward in the state-of-the-art. The MIMESIS project will develop scalable semantic-based algorithms and techniques for exploiting massive distributed repositories and analyzing large volume data streams for actionable knowledge. In particular, it will support semiautomatic semantic annotation of heterogeneous media forms: video, audio, images, etc. The library will explicitly address interoperability with other European initiatives, and in particular with *Europeana*. All the content and the

metadata created in the project will be accessible as well through the *Europeana* portal.

IV. CONCLUSIONS

In this position paper we present MIMESIS, a semantic-based approach for the creation of a model for an on-line digital, multimedia library of poetry in music, built on recent definitions of ontologies for cultural heritage such as FRBRoo. We describe the application of the MIMESIS approach to an existing repository, *Repertorio della Poesia Italiana in Musica 1500-1700* (RePIM), a digital archive of poetry and music from 1500 to 1700, which comprises the analytical description of the literary and musical sources of circa 43,000 texts set to music (for a total of about 66,000 musical settings), and digital items such as images of the ancient sources and modern editions of the texts, as well as the music of particularly significant samples. The MIMESIS approach allows for the development of methodological solutions which enable all users to explore and creatively enjoy digitized cultural resources, explicitly addressing issues of interoperability with ongoing European initiatives.

REFERENCES

- [1] *Europeana*: <http://www.europeana.eu/>
- [2] *European Digital Library*: <http://www.edlproject.eu/>.
- [3] *World Digital Library*: <http://www.wdl.org/en/>.
- [4] *Multilingual Inventory of Cultural Heritage in Europe* (MICHAEL): <http://www.michael-culture.eu>.
- [5] *Internet Culturale* is the Web portal that provides access to the digitised cultural heritage preserved by the Italian Public Libraries: <http://www.internetculturale.it/opencms/opencms/it/>.
- [6] *Gallica* is the digital library of the Bibliothèque Nationale de France: <http://gallica.bnf.fr/?lang=EN>.
- [7] *Online Gallery* is the digital library of the British Library: <http://www.bl.uk/onlinegallery/ttp/tpbooks.html>.
- [8] *Münchener Digitalisierungs Zentrum* (MDZ) handles the digitization and online publication of the cultural heritage preserved by the Bayerische Staatsbibliothek and by other institutions: <http://www.digitale-sammlungen.de/>.
- [9] *Digital.csic* is the digital library of the Spanish Consejo Superior de Investigaciones Científicas: <http://digital.csic.es>.
- [10] *Early European Books* provides scholars with new ways of accessing printed sources from the fifteenth to the seventeenth centuries: <http://eeb.chadwyck.com/marketing/about.jsp>. *Early English Books Online* (EEBO): <http://eebo.chadwyck.com/home>.
- [11] N. Crofts, M. Doerr, T. Gill, S. Stead and M. Stiff, Definition of the CIDOC Conceptual Reference Model, ICOM/CIDOC CRM Special Interest Group, May 2011 (http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.3.pdf).
- [12] *Functional Requirements for Bibliographic Records*: <http://www.ifla.org/en/frbr-rg>.
- [13] *International Federation of Library Associations and Institutions*: <http://www.ifla.org>.
- [14] K. G. Saur, *Functional Requirements for Bibliographic Records: Final Report*, Munich: UBCIM Publications, 1998 (vol. 19).
- [15] International Working Group on FRBR/CIDOC CRM harmonization, object-oriented version of FRBR, harmonized with CIDOC CRM (May 2009): http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0_2009_june.pdf.
- [16] M. Doerr and P. LeBoeuf, "Modelling Intellectual Processes: The FRBR-CRM Harmonization", in *Digital Libraries: Research & Development*, First International DELOS Conference, C. Thanos, F. Borri and L. Candela Eds. Berlin-Heidelberg: Springer Verlag, 2007, pp. 114-123.
- [17] *Enabling Access to Sound Archives through Integration, Enrichment and Retrieval* (EASAIER): <http://www.elec.qmul.ac.uk/easaier/index.html>.
- [18] *Online Computer Library Center* (OCLC) <http://www.oclc.org/research/activities/fictionfinder/default.htm>.
- [19] *Music Ontology*: <http://musicontology.com> and Raimond's site, <http://moustaki.org/>.
- [20] *Variations3*: <http://www.dlib.indiana.edu/projects/variations3/index.html>.
- [21] A. Pompilio, "RePIM - Repertorio della Poesia Italiana in Musica, 1500-1700" in *Petrarca in musica*, Proceedings of the International Conference promoted by the National Committee on the Seventh Centenary of the Birth of Francesco Petrarca (2004), Arezzo, 18-20 March 2004, A. Chegai and C. Luzzi Eds. Lucca: LIM, 2005, pp. 391-396.
- [22] E. Vogel, *Bibliothek der gedruckten weltlichen Vocalmusik Italiens aus den Jahren 1500-1700*, reprint Hildesheim: G. Olms, 1962 and *Nuovo Vogel*, E. Vogel, A. Einstein, F. Lesure and C. Sartori, *Bibliografia della musica italiana vocale profana pubblicata dal 1500 al 1700*, Pomezia: Staderini-Minkoff, 1977.
- [23] *Biblioteca Italiana* (BibIt) is available at: http://www.internetculturale.it/opencms/opencms/it/collezioni/collezione_0061.html.
- [24] *Liberliber* (Progetto Manuzio): <http://www.liberliber.it/libri/>.
- [25] *Antologie della Lirica Italiana*: <http://ali.unipv.it/>.
- [26] See S. Albonico, "Rasta. Raccolte a stampa antologiche del petrarchismo e della poesia italiana dal Cinque al Settecento", in *Il Petrarchismo. Un modello di poesia per l'Europa*, vol. I, L. Chines Ed. Roma: Bulzoni, 2006, pp. 311-348.
- [27] *Antologie di rime del Cinquecento* is available at: <http://www.sursum.unito.it/archivi/>.
- [28] C. Perelli Cippo, "Antologie di rime del Cinquecento: un progetto di schedatura elettronica" in *Il Petrarchismo. Un modello di poesia per l'Europa*, F. Calitti and R. Gigliucci Eds. Roma: Bulzoni Editore, 2006, pp. 583-588.
- [29] Museo internazionale e biblioteca della musica di Bologna: www.museomusicabologna.it/
- [30] DIAMM: <http://www.diamm.ac.uk/index.html>.

Indexing and Searching Cross Media Content in a Social Network

Pierfrancesco Bellini, Daniele Cenni, Paolo Nesi

Department of Systems and Informatics, Faculty of Engineering, University of Florence, Florence, Italy
{pbellini, cenni, nesi}@dsi.unifi.it, <http://www.disit.dsi.unifi.it>

Abstract—Recent challenges in information retrieval are related to information in social networks and rich media content. In those cases, the content is associated with multilingual, user generated aspects and content, scalability, robustness and resilience to errors. Moreover, fast and efficient indexing and searching services are needed, in order to scale digital content distribution and video on demand, where huge amount of queries and content related tasks are performed by users. In this paper, an indexing and searching solution for cross media content is presented. It has been developed addressing several problems of the ECLAP social network, collecting and managing content and users in the domain of Performing Arts. The research is conducted in the scope of ECLAP (<http://www.eclap.eu>). The effectiveness of the proposed solutions has been also measured and presented.

Keywords: *content, distribution; cross media content; automated production; mpeg-21; mobile; iphone; indexing, searching, search engines, information retrieval; cultural heritage.*

I. INTRODUCTION

In recent years, the rapid growth of digital resources on the Web has opened new challenges in developing efficient and robust information retrieval solutions. A wide variety of contents, with different formats and metadata types, constitute a heterogeneous set of resources difficult to deal with. An important example is given by cross media resources, which often include a rich set of metadata, mixed media, addressing serious issues when building a digital content index. Typically, there is a need of tools for metadata extraction, schemas and metadata mapping rules and tools, multilingual metadata and content translation and certification. Information retrieval (IR) systems are required to give coherent answers with respect to typos or inflexions, and must be efficient enough while sorting huge result lists. Search refinement, sorting and/or faceting techniques are major topics, especially in large multimedia repositories. Document parsing algorithms have to be fast enough in order to index high volumes of rich text documents, and to support different types of content descriptors. Indexes and repositories have to be fully accessible, without significant downtime, in case of failures or major updates of the index structure, in production services (e.g., redefinition of index schema, index corruption etc.).

Multilingual documents require query or metadata translation for information retrieval. The first approach reduces the memory usage and each document is stored only once in the index [1], while the second produces larger indexes and a voids query translation issues.

Indeed the automatic query translation process could create word ambiguity, polysemy, inflection and homonymy issues [2], especially in the case of short queries [3]. Disambiguation techniques can be used, for example using co-occurrences of pair terms [4], or a general statistical approach. Query expansion [5], for example pseudo relevance feedback technique [6] [7], thesauri such as WordNet [8] or structured translation [9] can be applied to increase the retrieval efficiency. A retrieval system's effectiveness can be measured with basic estimation metrics such as precision, recall, mean average precision, R-precision, F-measure and normalized discounted cumulative gain (NDCG) [13]. Relevant test collections and evaluations series for a *ad hoc* IR system assessment include TREC [34], GOV2 [35], NTCIR [36] and CLEF [37].

Other possible approaches for dealing with multilingual documents refer to Self Organizing Maps (SOMs) [10] or make use of sentence clustering before the translation process [11]. An alternative query translation approach involves the use of parallel or comparable Corpora [12]. They consist in a collection of natural language texts, where each document is translated in various languages; aligned parallel corpora are annotated to match each sentence in the source document with their respective translations. Thus, documents are comparable when they use the same vocabulary and deal with the same topic [13]. Ranking algorithms consist in ordering the output results list from the most to the least likely item [14]. Generally, ranking is based on location and frequency; documents with higher term occurrences are ranked higher. A notable example is the PageRank algorithm [15], which determines a page's relevance with a link analysis. Relevance feedback algorithm is based on the concept that a new query follows a modified version of the old one, derived by increasing the weight of terms in relevant items, and decreasing the weight of terms in non-relevant items. In order to overcome the limitations of traditional keyword based search engines, fuzzy approaches are used; synonyms or typos are evaluated in terms of similarity with current indexed tokens, to provide more complete results. Relevant examples of fuzzy techniques application include semantic search [16], ontologies [17], cloud computing [18], image text analysis [21], query expansion [20], clustering [19], and popular search platforms [22]. Multidimensional dynamic taxonomies models (i.e., faceted search [24] [25]) are also very popular, especially in e-commerce sites, where the user needs a way to easily explore the contents, and each facet can be represented with a taxonomy [23]. Document type detection and parsing

algorithms for metadata extraction are a valuable key factor for integrating rich text resources (e.g., semi-structured or unstructured documents) in digital indexes, with the aim of Natural Language Processing (NLP) techniques; example approaches include machine learning methods [26], table metadata extraction (e.g., from PDFs [27]), context thesauri in conjunction with document analysis [29], DOM based content extraction [28]. Typically, extracted information from unstructured documents can be organized as entities (i.e., noun phrases) and relationships between them, adjectives, tables and lists [30].

In this article, an indexing and searching solution for cross media content is presented. It has been developed addressing several problems in the area of cross media content indexing for social networks. It has been developed for the ECLAP social service portal, in the area of Performing Arts. The solution is robust with respect to typos, runtime exceptions, index schema updates, different metadata sets and content types, that constitute the ECLAP information model. It enhances and facilitates the user experience with full text multilingual search, for a large range of heterogeneous type set of content, with advanced metadata and fuzzy search, faceted access to query, content browsing and sorting techniques. The defined indexing and searching solution for ECLAP portal enabled a set of features involving a large range of rich content as: MPEG-21, web pages, forums, comments, blog posts, images, rich text documents, doc, pdf, collections, play lists, etc. Most of the activities are performed in the system back office developed with AXCP (AXMEDIS Media Gris solution for semantic computing) tools. Thus, the indexing service can be implemented in a distributed parallel architecture for massive ingestion and indexing. The proposed solution includes monitoring and logging facilities, providing data for further investigations (e.g., IR system effectiveness and user behavior assessment).

This paper is structured as follows. Section II depicts the ECLAP overview, describing the projects goals, and giving the general scenario where the solution has been integrated. It summarizes the main ECLAP tools and services. Section III presents the ECLAP indexing core services with respect to the main identified requirements and discusses the indexing related issues. Section IV depicts the ECLAP searching solution and discusses the searching strategies to increase retrieval efficiency, in the context of Performing Arts. Conclusions are drawn in Section V.

II. ECLAP OVERVIEW

The goal of the ECLAP project is to create an online archive and portal in the field of the European Performing Arts, which will also become indexed and searchable through the Europeana portal in the so called EDM data model [33]. ECLAP main objectives are to: make accessible on Europeana a large amount of Performing Arts related material as cross media content (e.g., performances, lessons, master classes, teaching material in the form of videos, audio, documents, images

etc.); bring together Europe's relevant Performing Arts institutions, to provide their content on Europeana; create a stable best practice network of European Performing Arts institutions. ECLAP provides solutions and services for: Performing Arts institutions, to bridge the gap between them and Europeana, via guidelines and solutions; final users (teachers, students, actors, researchers, and Performing Arts lovers for edutainment, infotainment and entertainment). The ECLAP mission is to develop new technologies, to create a virtuous self-sustainable mechanism, to provide continuously access to the content, and to increase the number of online materials. ECLAP can be seen as a support and tool for: content aggregators (e.g., for content enrichment and aggregation, preparing content for Europeana, for content distribution); working groups on best practice reports and articles, about tools for education and training, intellectual property and business models, digital libraries and archiving. Many other thematic groups and distribution channels are also defined. ECLAP networking and social services facilities include: user group, discussion forums, mailing lists, connection with social networks, suggestions and recommendations to users, as intelligence tools (e.g., potential colleagues, using metrics based on static and dynamic user aspects, similar contents), etc. Content distribution is available toward several channels: PC/Mac, iPad and Mobiles. ECLAP includes a back office intelligence mechanisms for: automated ingestion and repurposing for metadata and content items; multilingual indexing and querying, content and metadata enrichment, IPR wizard for IPR modeling and assignment, content aggregation and annotations, e-learning support, production of suggestions. ECLAP content portal features a large set of item formats, accessible through a search service with faceting refinement and ordering. Monitoring services allow content providers to assess user data and user behavior analysis (e.g., download of contents, user satisfaction about search), reports about user preferences, with visual statistical analysis overviews on the administrative section of the portal; promotion on all indexing portals, to make more visible each partner's content. The ECLAP solution would result in cultural enrichment and promotion of European culture, in learning and research improvements. In ECLAP, users are able to deal with forums, groups, blogs, events, pages, archives, audios, blogs, braille music, collections, documents, epub, excel, flash, html, images, pdf, playlists, slides, smil, tools, videos, etc. Depending on credentials and a set of grants, each user can upload, create, improve and/or edit digital resources and their corresponding metadata. In this context, an indexing and searching service has been developed and integrated into the ECLAP Web portal.

III. INDEXING

The ECLAP content model has been designed to cope with several types of digital resources with different metadata, which require a suitable metadata mapping schema to be used for content indexing, thus enabling the whole set of content related metadata to be stored in

the same index instance. Each resource category has to map its metadata into the same set of fields, adding its specific ones into a separate set, for advanced search purposes; this results in a unified and flexible indexing schema describing the whole set of heterogeneous contents. The metadata schema is categorized in different sections (see Table I): Dublin Core (DC [32]), Technical (e.g., type of content, partner acronym providing the data, IPR model, duration, video quality, GPS position, sources, formats, etc.), Performing Arts specific metadata (such as roles of actors, relationships, information related to recording situation, etc.), ECLAP Distribution and thematic Groups, and assignment of Taxonomical terms to content. Moreover, for some of the content types, full text is accessible; comments, tags and votes may be added as user generated content.

Multilingual aspects can be at metadata level and at content body level. For example, a multilingual web page or multilingual taxonomy terms. DC and DCTerms may include a metadata language (*xml:lang* attribute that can be either mandatory, optional or not necessary). Elements that are mapped to a Performing Arts metadata are mapped to one of the generic DC/DCTerms metadata when providing metadata to Europeana. Cross media content can be MPEG-21, animations, intelligent content etc.; they share the same set of metadata, while contents that provide full text data for indexing such as blog posts, web pages, events, etc. typically do not have metadata, though they may have Groups, Taxonomical associations and comments. The ECLAP multilingual index includes the languages of the ECLAP partners (Catalan, Greek, English, Spanish, French, Hungarian, Italian, Dutch, Portuguese, Slovenian). Multilingual metadata are automatically translated from source language, and mapped into their respective language schema fields, in order to avoid query translations issues.

TABLE I. ECLAP INDEXING MODEL

MEDIA TYPES	DC+Multilingual	Technical	Performing Arts	Full Text	Tax Group + ML	Comments, TAGS+ML	Votes
Cross Media (html, MPEG21, animations, etc.)	<i>Y_n</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y_n</i>	<i>Y_m</i>	<i>Y_v</i>
Info text: blog, web pages, events, forum, comments	<i>T</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>Y_m</i>	<i>N</i>
Document: PDF, DOC, ePub, ...	<i>Y_n</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y_n</i>	<i>Y_m</i>	<i>Y_v</i>
Audio/video/image	<i>Y_n</i>	<i>Y</i>	<i>Y</i>	<i>N</i>	<i>Y_n</i>	<i>Y_m</i>	<i>Y_v</i>
Aggregations (play lists, collections, courses, etc.)	<i>Y_n</i>	<i>Y</i>	<i>Y</i>	<i>I</i>	<i>Y_n</i>	<i>Y_m</i>	<i>Y_v</i>

In Table I: *Y_n* means yes with *n* possible languages (i.e., *n* metadata sets); *Y_v* means yes with *v* votes; *Y* means only one set of those metadata; *I* means that the aggregation itself does not present any full text indexing, while the aggregation elements are indexed as full text

entities according to their type; *T* means only title of the metadata set, *Y_m* means that for that content type *m* different comments can be provided and each of them in different language. Comments may have comments as well, thus resulting in a discussion even out of a regular forum.

The above index model has been designed in order to meet the metadata requirements of the digital contents, while indexing follows the ECLAP metadata ingestion schema. 20 different partners are providing their archives by using their own custom metadata that only partially meet with standard DC. Thus, in ECLAP, digital contents are indexed in a single multilingual index, for fast access, easy management and optimization.

Catchall fields for main metadata are automatically populated at indexing time (see Table II), for full text general search (e.g., title, description, subject, content taxonomies, and Performing Arts classification), to allow multilingual metadata retrieving through a single index instance, and to build more compact Boolean queries. In this model, the catchall field includes full text of each content element. This means that in the case of Text-info media types, title and body are the only accessible fields.

TABLE II. INDEX CATCHALL FIELDS^A

Catchall field	Component fields	Boost Weight of component fields
text	pdf_*, doc_*, docx_*, ppt_*, pptx_*, htm_*, html_*, txt_*, rtf_*	1
title	title *	3.1
body	body *	0.5
description	description *	2.0
contributor	contributor *	0.8
subject	subject *	1.5
taxonomy	taxonomy *	0.8
Performing Arts	PerformingArtsMetadata.#	1

A. Index fields with a _* suffix indicate language field declinations (e.g. title_english etc.)

Performing Arts metadata list

Relevant metadata are only indexed and not stored in the index, in order to keep the index smaller, easier to manage and faster to be accessed; technical metadata are stored verbatim and not analyzed (e.g., video resolution, quality, format), while descriptive metadata are fully processed (i.e., tokenized, lowercased). Metadata and structured text content from attached resources (i.e., doc, docx, ppt, pptx, xls, xlsx, pdf, html, txt) are extracted after detecting magic bytes (i.e., a prefix that identifies the file format), file extension, content type and encoding, with the aim of an internal MIME database and parsing libraries provided by Apache Tika [31]. Multilingual taxonomies are hierarchically managed in a MySQL database; each resource can be linked to a set of taxonomies, which come as a part of the resource's metadata in the index (each taxonomy path is serialized into a string, before indexing in the document structure).

Each object metadata set can be enriched and edited on the portal, to arrive at the validated/certified version to be published on Europeana. Corrections can be applied to multiple objects at the same time. Technical Metadata are used to search and retrieve them in the back office.

Avoiding downtime

The index structure is rebuilt automatically from scratch in case of corruption or major schema updates. To avoid significant service downtime, the production service keeps running while a separate instance of the index is being built, and accessory calculations are performed (e.g., taxonomy/group extraction related to each digital resource). During re-indexing, the production indexing service keeps trace of the newly uploaded contents, to index them in the new generated index; every possible thrown exception is notified by mail to the administrator, with all the relevant information (i.e., type of error, stack trace, timestamp and resource involved); mail recipients are customizable from the administrative panel settings. At the end of the re-indexing process, the newly created index is transferred over the old one.

IV. SEARCHING

The goal of the searching service is to allow the users to easily locate and sort each type of content in the ECLAP portal, and to refine their queries for a more detailed result filtering, through a fast search interface, robust with respect to mistyping; high granularity of data has to be offered to the users (i.e. advanced metadata search), with a detailed interface.

Text-Info contents (such as web pages, forums, comments, groups, events, etc.) and media contents have to be searchable in the ECLAP portal; after a query heterogeneous results may be obtained. For example, a blog post, a group, an event, a comment, a tag, a PDF, etc. In order to reduce this kind of complexity and simplify the readability of the results, in terms of play of the obtained results, the identification of comments is manifested as the presentation of the original content element, at which the comment has been provided. This means that querying for a term contained in a page, blog, forum or cross media content, has to match the set of resources containing that search term, producing a list of formatted results. Querying by taxonomy related to the content has to provide a pertinent match too. Relevance scoring has to take into account different weights for each document's metadata field; a same term occurring in different document fields is expected to provide different scoring results (i.e., a higher field's weight means a higher relevance of that field).

In order to simplify the work of users, the search service is provided as a n easy to use full text, and as a n advanced search. The easy to use full text frontal search is in the top center of the portal; it is offered as a text box with a search button and a drop-down menu for resource type filtering (video, audio, images, initially alphabetically ordered, while now ordered presenting on

the first 5 those that are the most chosen). Queries are automatically tokenized and lowercased, before assembling the query string (i.e., a combination of weighted OR Boolean clauses, with escaping of special characters), and sent to the indexing service. Depending on the enabled languages in the ECLAP portal, each active language field is included in the query string for full text search. Advanced search is reachable from the top center portal menu, and provides language, partner and metadata filtering. The user is allowed to compose an arbitrary number of Boolean clauses in the advanced search page, thus allowing the building of a rich metadata query; for example by restricting the search to some metadata fields that only match any or all of them (OR/ALL).

In both cases, fuzzy logic facility is transparently applied to both simple and advanced search; even a query with typos can return coherent results. The query term is compared to similar terms in the index, for retrieving documents with a high degree of similarity (e.g., “*documant*” should match “*document*”), thus allowing an efficient search in case of mistyping. String metric used (Levenshtein [39] or edit distance) allows measuring the similarity between two search terms, by evaluating the minimum number of transformations needed to change one search term into another. This fuzzy similarity weight is customizable by the administrator in the portal (a weight < 1 means fuzzy logic, while a weight = 1 means Boolean logic).

In the frontal search service, a deep search checkbox is available. It allows the user to enable/disable such functionality, thus the query string is prefixed and suffixed with a special wildcard, in a transparent way to the user, to allow searching of substrings in the index (e.g., query “*test*” matches “*testing*”).

Boosting of terms is configurable on the portal. This allowed us to tune and stress the importance of certain metadata. On the basis of the performed experiments, the best appreciation has been obtained by giving more relevance to some fields with respect to others (i.e., title, subject, description, see Table II for boosting weights used in the ECLAP index). The administrator is able to change the boosting of the main search fields (i.e., title, body, description, subject, taxonomy); boost values can be extended to the whole set of metadata, though. Boosting and weighting of metadata are better tuned when the portal is more populated with significant contents. Each field of the ECLAP document structure is boosted with its predefined value at query time.

Faceted search is activated on the results of both simple frontal search and a dvanced search. Each faceted term is indexed un-tokenized in the ECLAP index, to accomplish a faceting count based on the whole facet. Facet parameters are appended to the query term; facet counts are evaluated from the output result by a Drupal service module, before rendering. The user can select or remove any facet in any order to refine the search. Adding or removing a facet results in adding or deleting

a search filter, and performs again the search query with or without it. Relevant facets include:

- DC: resource category, format, type, classification, creator, content language, etc.
- Technical: duration, video quality, device, publisher, source metadata language and upload time
- Group, taxonomy: genre, historical period, performing arts, coded subject

These facets can be subject to change. For instance, locations and dates, different for each historical period, can be added.

Search results are listed by relevance in descending order; this means that the first document is the most relevant with respect to the query. Results can be sorted by uploading or updating time too. The relevance is based on the occurrence of the query term in the indexed document fields: a higher number of term's occurrences (or similar terms) gives a higher score for the document. Each result item is presented with a thumbnail, relevant metadata (i.e., title and description), rating, relevance score and number of accesses; data is presented in the same language chosen by the user among the available portal localizations (or, if not available, in English or the source metadata language of the content). Results are paginated, typically 10 per page; this setting can be changed by the administrator in the settings panel. Suggestions can be enabled from the settings panel; while typing a query the system searches in the ECLAP index and may suggest similar terms to the typed one.

Search Facility Assessment

The analysis has been performed in the period from September 1st 2011 to November 30th 2011. Some of the data have been collected by using Google Analytics while others have been directly collected with internal logs. In that period, a total number of 11294 visits to the portal (of which 6032 unique visitors) have been registered. A total of 62768 views, and thus we had 5.56 pages for visit. These data were associated with 7.29 minutes of mean time of permanence on the portal. These values were substantially better than those that have been published for similar social network portals. In that period, a total of 30502 contents accesses have been registered (view, play and download, download is the 0.035%).

Moreover, Table III depicts some data about searching activities performed on ECLAP community (sorted by partnership), through queries and static menu lists available on the ECLAP portal. The numbers are referred to the same period. The first column reports the number of performed full text queries, obtaining a high rate of query per visit ratio of 37%. This means that the 37% of visitors have performed at least a query, and from these: 4051 full text queries were performed (94.56% of total), 192 faceted queries, and 41 advanced queries. Most of the queries have been performed by anonymous users. Registered users are those that are regularly registered on the portal as single users, and thus

do not belong to one of the 25 institutions that have signed an agreement with ECLAP as partners or affiliated partners.

In Table III, the data related to other search results is reported to put in evidence the usage of: faceted (used only in the 5% of cases); ready-made queries to propose last posted, featured and the most popular content. The last line of the table reports the number of clicks performed on those search and list results. Thus, clicks on Last Posted Contents and Featured Contents were performed through the ECLAP menu, at the top of the home page; Clicks on Featured Contents list were performed when that list was in the home page. Moreover, it is also evident that over the 4051 queries only in 1564 cases the user has proceeded to click on the provided lists of objects.

TABLE III. QUERIES / CONTENT LISTS^A

users	# Full Text Queries	# of Faceted Queries	# Last Posted Contents	# Featured Contents	# Popular Contents
simple registered	323	24	4	22	17
Registered as partners	1094	21	27	19	9
anonymous	2634	147	234	302	213
Total	4051	192	265	343	239
Clicks after query	1564	200	318	2799	231

A. Sample Period: September 1st 2011 – November 30th 2011

It can be useful to see where the users have clicked into the lists of query results, coming from the full text and advanced query. This distribution is reported in Figure 1 in which the first 14th positions are reported. 10 of them are in the first page and the others in the first part of the second page. From these data, it can be noted that after a query on the portal, the 92.65% search results clicks were performed in the first page (first ten results). 42.27% of clicks on search results have been performed to the first proposed result. The second has received only the 14% clicks.

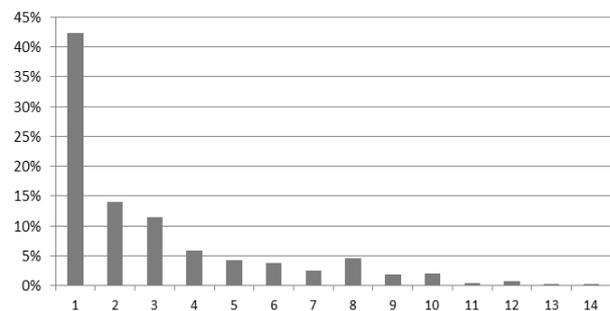


Figure 1. Clicks order distribution (first page results and a part for the second)

V. CONCLUSIONS AND FUTURE WORK

An integrated searching and indexing solution for the ECLAP Web portal has been developed. The ECLAP index has been designed to scale efficiently with thousands of contents/accesses; the proposed searching

facilities contribute to enhance the user experience, speed up and simplify the information retrieval process. A preliminary analysis of user sessions has been conducted, to put in evidence the user behavior on the portal. A deeper analysis is in progress to better understand the appreciation, the effective satisfaction, and user preferences.

VI. ACKNOWLEDGMENTS

The authors want to thank all the partners involved in ECLAP, and the European Commission for funding the project. ECLAP has been funded in the Theme CIP-ICT-PSP.2009.2.2, Grant Agreement No. 250481.

REFERENCES

- [1] J.S. McCarley, "Should we translate the documents or the queries in cross-language information retrieval?", in Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. 1999, Association for Computational Linguistics: College Park, Maryland.
- [2] Abusalah, M., J. Tait, M. Oakes, "Literature Review of Cross Language Information Retrieval", World Academy of Science, Engineering and Technology 4, 175-177, 2005.
- [3] Hull, David A., G. Grefenstette, "Querying across languages: A dictionary-based approach to multilingual information retrieval", Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR96), pages 49 – 57, Zürich, Switzerland, 1996.
- [4] Yuan, S., S. Yu, "A new method for cross-language information retrieval by summing weights of graphs", in Fourth International Conference on Fuzzy Systems and Knowledge Discovery, J. Lei, Editor. 2007, IEEE Computer Society. p. 326 - 330.
- [5] L. Ballesteros, B. Croft, "Phrasal translation and query expansion techniques for cross language information retrieval", in Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR97), pages 84 – 91, Philadelphia, PA, 1997.
- [6] Attar, R., A.S. Fraenkel, "Local feedback in full-text retrieval systems", Journal of the Association for Computing Machinery, 1977. 24: p. 397-417.
- [7] L. Ballesteros, W.B. Croft, "Resolving ambiguity for cross-language retrieval", Proceedings of the 21st International ACM SIGIR Conference on Research and Development in Information Retrieval. 1998. New York: ACM Press.
- [8] Fellbaum, C., "WordNet: an electronic lexical database", Cambridge, Mass: MIT Press, 1998.
- [9] Sperer, R., D.W. Oard, "Structured translation for cross-language information retrieval", in Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR00), Athens, Greece, 2000.
- [10] Lee, C.-H., H.-C. Yang, "Towards multilingual information discovery through a SOM based text mining approach", T. Ah-Hwee and P. Yu, editors, Proceedings of the International Workshop on Text and Web Mining (PRICAI 2000), pages 80–87, Melbourne, Australia, August 28 - September 1 2000. Deakin University, Australia.
- [11] Chen, H.-H., J.-J. Kuo, T.-C. Su., "Clustering and visualization in a multi-lingual multi-document summarization system", in Proceedings of the 25th European Conference on Information Retrieval Research (ECIR 2003), number 2633 in LNCS, pages 266–280, Pisa, Italy, April 14-16 2003. Springer.
- [12] Picchi, E. and C. Peters, "Cross-language information retrieval: a system for comparable corpus querying", Cross-Language Information Retrieval, G. Grefenstette, Editor. 2000, Kluwer Academic Publishing: Massachusetts. p. 81-90.
- [13] D. Manning, C., P. Raghavan, H. Schütze, "Introduction to Information Retrieval", 2009, Cambridge: Cambridge University Press.
- [14] Belkin, N. J., W. B. Croft, "Retrieval Techniques", Williams, M. (ed.), Annual Review of Information Science and Technology, Elsevier Science Publishers, New York, 1987, Pages 109–145.
- [15] Brin, S., Page, L., "The anatomy of a Large-Scale Hypertextual Web Search Engine", Computer Networks and ISDN Systems 30: 107–117, 1998.
- [16] L.-F. Lai, C.-C. Wu, P.-Y. Lin, L.-T. Huang, "Developing a fuzzy search engine based on fuzzy ontology and semantic search," Fuzzy Systems (FUZZ), 2011 IEEE International Conference on , vol., no., pp.2684-2689, 27-30 June 2011.
- [17] Akhlaghian, F., Arzaniyan, B., Moradi, P., "A Personalized Search Engine Using Ontology-Based Fuzzy Concept Networks," Data Storage and Data Engineering (DSDE), 2010 International Conference on , vol., no., pp.137-141, 9-10 Feb. 2010.
- [18] Jin Li; Qian Wang; Cong Wang; Ning Cao; Kui Ren; Wenjing Lou; , "Fuzzy Keyword Search over Encrypted Data in Cloud Computing," INFOCOM, 2010 Proceedings IEEE , vol., no., pp.1-5, 14-19 March 2010.
- [19] Matsumoto, T.; Hung, E. , "Fuzzy clustering and relevance ranking of web search results with differentiating cluster label generation", Fuzzy Systems (FUZZ), 2010 IEEE International Conference on , vol., no., pp.1-8, 18-23 July 2010.
- [20] Takagi, T., Tajima, M., "Query expansion using conceptual fuzzy sets for search engine", The 10th IEEE International Conference on Fuzzy Systems, vol.3, pp.1303-1308, 2001.
- [21] Berkovich, S., Inayatullah, M., "A fuzzy find matching tool for image text analysis", Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on , vol., pp. 101 - 105, 13-15 Oct. 2004.
- [22] <http://lucene.apache.org/solr/>
- [23] G. M. Sacco, "Research Results in Dynamic Taxonomy and Faceted Search Systems", 18th International Workshop on Database and Expert Systems Applications, 2007.
- [24] D. Tunkelang, "Faceted Search", (Synthesis Lectures on Information Concepts, Retrieval, and Services), Morgan and Claypool Publishers, 2009.
- [25] G. M. Sacco, Y. Tzitzikas, "Dynamic Taxonomies and Faceted Search", Springer, 2009.
- [26] Hui Han, C. Lee Giles, Eren Manavoglu, Hongyuan Zha, "Automatic Document Metadata Extraction using Support Vector Machines", Proceedings of the 2003 Joint Conference on Digital Libraries, 2003.
- [27] Y. Liu, P. Mitra, C. L. Giles, K. Bai, "Automatic Extraction of Table Metadata from Digital Documents", Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, 2006.
- [28] S. Gupta, G. Kaiser, D. Neistadt, P. Grimm, DOM-based Content Extraction of HTML Documents, Columbia University Computer Science Technical Reports, 2002.
- [29] Shepherd, M., Watters, C. Young, J., "Context thesaurus for the extraction of metadata from medical research papers", Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.
- [30] S. Sarawagi, "Information Extraction", Foundations and Trends in Databases, Vol. 1, No. 3, 261–377, 2007.
- [31] <http://tika.apache.org/>
- [32] <http://dublincore.org/>
- [33] <http://www.europeana.eu/portal/>
- [34] <http://trec.nist.gov/>
- [35] <http://www-nlpir.nist.gov/projects/terabyte/>
- [36] <http://research.nii.ac.jp/ntcir/data/data-en.html>
- [37] <http://www.clef-initiative.eu/>

DIGITALMEETSCULTURE.NET

Interactive magazine about new technologies meeting culture worldwide

Antonella Fresa, Tania Masi, Claudio Prandoni, Valentina Bachi

Promoter via Boccioni 2 - 56037 Peccioli (Italy) promoter@promoter.it

Abstract— In the context of global growing access to the Internet, and over-flooding of on-line information about any topic, it is sometimes hard for users to collect materials and news about their interests.

digitalmeetsculture.net is a project run by company Promoter, and it was imagined as an on-line magazine about the digital culture.

The portal was also built in order to create a useful tool for the dissemination of digital cultural heritage projects and initiatives: by giving wide visibility to the activities and achievements, by deepening their topics through interviews and related articles, and by offering a repository service for the projects' results.

I. INTRODUCTION

The concept of digital culture enlarged very much from its initial aspect of technical-scientific knowledge and values related to the information technology and its application. Nowadays, digital culture is referred to the application of digital technologies to any field of human expression. New ideas like Digital Art, Digital Heritage, Digital Native, Digital Age are currently accepted and used.

Internet is indeed “the” tool for communication at every level and with every purpose: it is used for disseminating project activities, for business promoting, and more in general for finding or sharing any information. Beside this, Internet is also a powerful resource for entertainment and amusement. In the context of global growing access to the Internet, and over-flooding of on-line information about any topic, it is sometimes hard for users to collect materials and news about their interests.

digitalmeetsculture.net is a project run by company Promoter, and it was imagined as an on-line magazine about the digital culture for collecting and sharing information and events, in a global dimension, taking into account the different approaches that scientific, humanistic and artistic culture have to the digital age.

The use of the network and of digital technology is evolving towards more and more interactive forms, with a clear trend towards the interchangeability of roles between the sender and the receiver of the information (Web 2.0). In addition, more and more artists of every discipline, from visual to performance arts, use computer technology as a tool and the web as a space for displaying and disseminating their works.

To this end, the portal digitalmeetsculture.net will discover, analyse, promote and disseminate the new achievements in the field of digital culture. It is already rising as a remarkable meeting point among different cultural fields, that takes into account the different approaches that scientific, humanistic and artistic cultures have to the digital age. The portal aims to act as a landmark and as a valuable mean of information and communication for different users in a global dimension.

Moreover, the speed and the amount of information offered by new technologies, allow, through specific projects which are promoted by academies, organizations and companies with the support of governments, to increase the accessibility to cultural heritage: works of art, texts and documents - but also video and audio repositories and archives - whose physical accessibility is limited by many factors, can be readily available to the global network, thus increasing the possibilities of enjoyment, and approaching researchers, experts, enthusiasts and curious ones to a field which initially was elitist.

Such projects require indeed great visibility in order to be shared with the community, and also support for disseminating their results: the digitalmeetsculture.net portal can be an useful, customizable and flexible tool to this purpose.

II. GENERAL OVERVIEW

A. The Sections

The portal is composed of two specific areas dedicated to the cultural heritage in general, and to the digital art.

The screenshot shows the website's layout. At the top, there's a navigation bar with links like 'Home', 'Our Mission', 'Events', 'Register / Edit Profile', 'Contacts', 'People', and 'Sitemap'. The main header features the 'DIGITAL MEETS CULTURE .net' logo. Below the header, there are sections for 'DIGITAL HERITAGE' and 'DIGITAL ART'. On the left, there is a 'Login Status' section for user 'adrian' and a 'Send your NEWS' button. The main content area includes a 'WELCOME TO digitalmeetsculture!' message, a 'Welcome to digitalmeetsculture!' section, and a 'Welcome to digitalmeetsculture!' section. On the right, there are sections for 'EDITORIALS' and 'INTERVIEWS'. At the bottom, there are sections for 'Upcoming Events' and 'EUROPHOTO'.

DIGITAL ART: The impact of digital technology transformed human expressions as painting, drawing and sculpture, but also music, poetry and the performances; first by implementing new tools which can be very helpful for the artists and which can enhance the creative process towards innovative and unexplored paths. In an expanded sense, "digital art" is a term applied to contemporary art that uses the digital technologies and the digital media, also to disseminate and share the artworks with the community. The section DIGITAL ART wishes to focus on the latest, innovative forms of the digital art in any expression, and to give space and visibility to upcoming events and discussions about this large, open and evolving topic. Conferences and seminars, interesting dates and case-studies will be collected; and nice showcases dedicated to associations, artists and companies, provided with many information and amazing multimedia contents in every format will offer a wide overview about what's going on.

DIGITAL HERITAGE: The digital technologies offer new modern tools for cultural heritage preservation; they also play a leading role about key issues as providing access, interaction and sharing knowledge.

Furthermore, our society is accumulating a large amount of born-digital heritage, especially documents, artworks, softwares, and the Web itself. Worldwide, preservation – both the preservation of tangible heritage through the digital technologies and the preservation of the digital cultural content – is felt as a matter of the highest importance, being so strictly connected to the true essence of the human culture.

A methodological and coordinated approach to digitization involves archives, libraries and museums and many other actors as academies, researcher centers, national and local institutions, private companies, to support properly this on-going process.

Last, but definitely not least, digital technologies represent the future of cultural heritage not only for

preservation intended as a Memory Institutions' concern, but also as a device the community will benefit from:

- by providing a growing and open access for investigating and/or general purposes to researchers, students and teachers, impaired people, and users in general;
- by enhancing the interaction with every user, thus enhancing a deeper intellectual enjoyment of cultural heritage;
- by developing new and challenging learning resources which will improve knowledge.

The section DIGITAL HERITAGE collects articles and information about projects and initiatives for the digitization and access to the digital cultural heritages all over the world. The key point at the base of this section is to give visibility to the institutions which work for preservation, to companies which develop tools and problem solving, to users who get benefit or gratification from digital cultural heritage.

B. Dedicated Services to EU Projects

digitalmeetsculture.net is a project run by Promoter, a relevant company based in Italy which is cleverly committed, with an open and coordinated approach, to several fields as technical development of ICT platforms and web-design, multimedia production and publishing. Promoter's general management is composed of skilled professional people who have been managing or contributing to EU projects since the early 1990s, in different sectors belonging to several programs as Esprit, ACTS, eContent, eTen, and RTD Framework Programmes FP5, FP6 and FP7. Promoter is currently partner of several projects in the digital cultural heritage sector, with the role of technical coordinator and an active attendance to dissemination. For example, recently, EuropeanaPhotography, a digitization project for over 500.000 artistic and ancient photos, has just started and DCH-RP, devoted to the development of a new research e-infrastructures for the digital cultural heritage, is going to have its kick-off very soon.

It is possible to comply with the dissemination packages of such projects thanks to a dedicated service in the portal: each project is provided with a customizable showcase for presentation, RSS, general information and contacts. The showcase is easily accessible thanks to an evident button with logo, which is present in the Home Page and in any other page. Clicking on the project's logo opens a dedicated page which contains a presentation of the project, beside related articles and interviews, and focused issues; the page contains also other useful information like contact details, auto-refreshing news (via RSS), multimedia galleries, and much more. Another service which is going to be fully developed in the next time is the Repository Area, where any kind of documents,

prototypes, websites which had been produced during the projects' life will be stored and preserved.

To sum up, therefore, the portal was also built in order to create a useful tool for the dissemination of digital cultural heritage projects: by giving wide visibility to the projects' activities and achievements, by deepening their topics through interviews and related articles, and by offering a repository service for the projects' results.

C. Target Users

digitalmeetsculture.net is particularly addressed to people belonging to the cultural heritage sector, who are interested in the digitisation process and technologies (e.g. museums and libraries, researchers, technicians, etc.), people interested in the creative uses of the digital art (e.g. artists, photographers, performers, art critics, etc.), people interested in the digital data from the technical and scientific point of view (e.g. professionals, researchers, information technology enthusiasts, etc.); people belonging to the educational sector (e.g. teachers, students, vocational trainers...) and also general users who want to be informed and up-to-date on these matters.

III. THE PORTAL'S STRUCTURE

The digitalmeetsculture.net portal is designed and actually working as an on-line magazine about the digital culture in all its meanings. The contents of the portal are organized in separate but correlated sections:

- The “**Editorials**” section contains articles of general interest and of theoretical – critical character pertaining to any disciplinary area.
- The “**Interviews**” section contains the report of interviews to key people in the field of digital culture.
- The “**Events**” section gives an overview of what it is happening all over the world and of the most important events related to the digital culture.
- The “**Showcases**” section is devoted to give

visibility to projects and organizations involved in the digitization process of culture, heritage and art.

- The **two specific areas** – digital heritage, digital art – host articles which have a specific focus on the two main topics of the portal; the writing style changes according to the specific area but it keeps the same high level of disclosure and scientific rigor.

Any article is published with links, attachments and every kind of multimedia content, and associated to one or more topics and keywords or free tags. At the end of each article there is the possibility for registered users to leave a comment, and a list of related articles is displayed. All the articles are at first highlighted in the Home Page, and later on they are visible in the main pages of each section. When items are no more immediately visible because they get old, they are still accessible through several search options: free text, by topic and by most popular keywords.

The editorial staff is assisted by Correspondents from all over the world, who cooperate to harvesting news and information, while a particular role is played by the Referents: key people who act as main contact inside relevant organizations or institutes, to keep the portal always up-to-date on the most prominent initiatives, projects, events and progresses.

Besides these fully working areas, there are 2 more that will be perfectly developed in the next time:

- The “**Repository**” section is dedicated to featured projects and will host:
 1. *a repository of **documents**, to archive any kind of documents;*
 2. *a repository of **prototypes**, to store all the software which had been produced during the project's life;*
 3. *a repository of **websites**, to collect and preserve all the project's websites.*



- The “**Search and Offer**” section will actually be a virtual pin-board to post announcements and ads, just like any modern magazine.

On a later stage, the portal will also host a virtual gallery of photography and art, where photographers and artists will display their works and give extra visibility to their personal websites, thus promoting their activity.

IV. TECHNICAL INFRASTRUCTURE

A. *digitalmeetsculture.net* portal base technology

The CMS which has been selected as base technology upon which to implement the *digitalmeetsculture.net* portal is Wordpress[1]. Wordpress is an open source blog tool and publishing platform licensed under the GNU General Public License (GPL), powered by PHP and MySQL and easily customizable into a Content Management System (CMS). Wordpress has been selected as the base technology for the implementation of the *digitalmeetsculture.net* portal because of its flexibility and of its easy and user friendly setup and usage, which nevertheless allows a high level of personalization, making it a very versatile CMS. Wordpress has a web template system which uses a template processor. It is therefore possible and easy to re-arrange widgets as well as to install and switch between themes. The PHP and HTML code in themes can also be edited for more advanced customizations. Wordpress features integrated link management; a search engine-friendly, clean permalink structure; the ability to assign nested, multiple categories to articles; support for tagging of posts and articles. Automatic filters are also included, providing standardized formatting and styling of text in articles. Wordpress supports the Trackback[2] and Pingback[3] standards for displaying links to other sites that have themselves linked to a post or article. Finally, Wordpress has a rich plugin architecture which allows users and developers to extend its functionality beyond the features that come as part of the base install. The features which have been implemented in *digitalmeetsculture.net* by making use of Wordpress core functionalities and plugins and by customising them according to the portal’s needs are: Management and Administration, Portal visibility, Design customization, Content creation, Archiving and Searching, Moderation.

B. *Users updating and interactivity*

The *digitalmeetsculture.net* portal includes a set of tools aimed at keeping users always involved and up-to-date, such as mailing lists, newsletters, accounts on the major social networks. The continuous renewal of the Home Page and of the main pages of the different specific areas with new articles, as well as the presence of an always updated Events section, motivates the users to visit the site several times.

Interactivity is a key-point of the portal: like any modern communication website, users can register to the portal, thus being allowed not only to simply comment the articles, but mostly to play an active part and send events and articles, until the point of joining the staff as correspondents or even referents if they wish.

The global dimension of the project is therefore sustained and implemented by enlarging the group of involved people who act as content providers, in order to open the magazine to any input.

V. FUTURE STEPS

Being the portal still young, as its official launch dates back to 2011, a wide number of dissemination activities are currently carried out to make the portal more and more renowned and visited. These include: presentations and demo of the portal to the events organized by the featured projects and on the occasion of conferences, exhibitions and international meetings; links and RSS feeds included in the featured projects’ sites and in other sites; distribution of brochures and other dissemination materials, etc.

Promoter is currently carrying on financial investments for empowering the visibility of the website with the aim of positioning the website on many search engines placed in the 5 continents.

The company will also make use of its long-time experience in European projects for networking the portal and achieve proper dissemination.

REFERENCES

- [1] ¹ <http://wordpress.org>
- [2] ¹ http://www.sixapart.com/pronet/docs/trackback_spec
- [3] ¹ <http://www.hixie.ch/specs/pingback/pingback>

WWW.DIGITALMEETSCULTURE.NET

WWW.PROMOTER.IT

© PROMOTER 2012

The Co-Creation of the City

re-programming cities using real-time user generated content

Salvatore Iaconesi⁽¹⁾⁽²⁾⁽³⁾, Oriana Persico⁽²⁾

(1) Faculty of Architecture, La Sapienza University of Rome, Rome, Italy, salvatore.iaconesi@artisopensource.net

(2) Fake Press Publishing, Rome, Italy, oriana.persico@gmail.com

(3) ISIA Design, Florence, Italy

Abstract—Is it possible to imagine novel forms of urban planning and of public policies regulating the ways in which people use city spaces by listening to citizens' expressions, emotions, desires and visions, as they ubiquitously emerge in real-time on social networks and on other sources of digital information?

This paper presents the theoretical and methodological approach, the investigation and research phases, the design and prototyping processes constituting the ConnectiCity initiative, a collaborative, multi-disciplinary series of projects in which artists, scientists, anthropologists, engineers, communicators, architects and institutions participated to the design of innovative ubiquitous and pervasive systems which were able to transform the ways in which the concepts of urban planning and city-wide decision-making are defined. Novel forms of urban life were imagined, in which cities became the time/space continuum for multiple, stratified layers of information expressing the ideas, goals, visions, emotions and forms of expression for multiple cultures and backgrounds, producing new opportunities for citizenship: more active, aware and engaged in the production of urban reality, and in the transformation of city spaces into possibilistic frameworks.

Urban planning, public space, ubiquitous technologies, participatory practices, urban ecosystems

I. INTRODUCTION

As we move through cities in our daily lives, we are in a constant state of transformation of the spaces around us.

Information, encounters, sounds, to-do lists, things which enter our field of vision recalling memories and thoughts, people to contact, images: all of these things are just a small part of the set of stimuli which push us to re-interpret urban spaces in novel ways. [1]

An unexpected phone call can transform a park bench into a temporary, ubiquitous office; an encounter can mutate a street corner into a theatrical performance; a doubt while choosing a shampoo among the hundreds of available ones in a supermarket aisle can enact a nomadic participative forum on ecology, health and sustainability; an intense traffic jam and the messages sent by distressed drivers to alert about delays and possible alternative routes can generate an emergent expert system on mobility and urban transportation; a bar conversation can create a useful source of information about urban security and on the visions of people about

the possible, desired transformations of their neighborhood, promoting wellness, culture, business and, more in general, opportunities for multiple types of citizens.

On top of that, the ways in which we constantly personalize the spaces which we traverse and in which we perform our daily routines is also able to communicate a wide array of truly significant information about our emotional states, our working methodologies, about our knowledge, skills, cultural backgrounds, desires, visions and perception of our cities, shops, parks, offices, houses, schools, as suggested in [2][3][4] among many others.

Furthermore, this process is far from being something which only effects single individuals, as it enacts transformations that are visible at the levels of households, communities, neighborhoods and entire cities.

“In the course of time every section and quarter of the city takes on something of the character and qualities of its inhabitants. Each separate part of the city is inevitably stained with the peculiar sentiments of its population.” [5]

On top of that, the form and essence of urban space directly affects people's behavior, describing in their perception what is possible or impossible, allowed or prohibited, suggested or advised against. [6]

This mutual feedback and transformative loop describes a second order cybernetic system whose observation can offer innovative perspectives for the processes which are concerned with the conception, design and implementation of products, services, practices or entire environments aiming at proposing visions of public space as possibilistic infrastructure dedicated to the development of human possibilities.

Ubiquitous technologies push all these tendencies and attitudes to the extreme.

We are now able to fill and stratify space/time with digital information layers, completely wrapping cities in a membrane of information and of opportunities for interaction and communication. [7][8][9][10][11][12]

We see a series of opportunities in all this, as we approach these issues from the points of view of art, design, architecture, anthropology, urban planning, communication science: the possibility to grasp and understand human beings' continuous state of re-

interpretation of the world to infer information, suggestions and visions about the ways in which people transform their reality, in both plan and action, to achieve a set of possible goals, such as:

- ^ gaining a better understanding of human presence in urban spaces;
- ^ producing tools to be able to listen to all the richness represented by the possibility to listen in real-time to the city, across cultures, languages, backgrounds, religions, nationalities and political creeds;
- ^ create methodologies, for all actors involved, according to which the possibility to listen and understand the expressions and emotions of people can become a valuable tool to create new, ethical, sustainable, participatory policies, plans, businesses, initiatives, processes;
- ^ enact initiatives in which citizens, organizations and institutions take part into a choral effort to actively accomplish a different form of urban space in which both citizens and administrators become active, aware and insightful agencies;

II. THE PERSONALIZATION OF SPACE/TIME

Portable technological devices have been known to enable radical practices for the possibility of human beings to personalize the spaces which they traverse.

For example, the Sony Walkman has represented a breakthrough innovation in this kind of scenario: “being in two places at once, or doing two different things at once: being in a typically crowded, noisy, urban space while also being tuned in, through your headphones, to the very different, imaginary space or soundscape in your head which develops in conjunction with the music you are listening to”. [13]

The possibility to move through urban spaces – with their cognitive, aesthetic and moral significance – using personal sound devices such as the Sony Walkman, allow us to benefit from the use of a critical tool in the management of our space and time, in the construction of boundaries around ourselves, and in the creation of sites of fantasy and memory. [14]

More recently, mobile devices, smartphones, wearables, digital tags, near field communication devices, location based services and mixed/augmented reality have gone much further in this direction, turning the world into an essentially read/write, ubiquitous publishing surface. [15]

The usage of mobile devices and ubiquitous technologies alters the understanding of place. [16]

As Morley describes: “The mobile phone is often understood (and promoted) as a device for connecting us to those who are far away, thus overcoming distance – and perhaps geography itself”. [17]

This description emphasizes only one of the ways in which the use of ubiquitous and mobile technologies alter our perception of time and space, by compressing

them into entities within reach and, thus, redefining our possibilities to interconnect and relate with objects, processes, places and other people. But there are more perspectives in which mobile devices transform our perception of physical reality, and Morley himself gives an account for that a few paragraphs after the previous remark, when he states how the mobile phone fills “the space of the public sphere with the chatter of the hearth, allowing us to take our homes with us, just as a tortoise stays in its shell wherever it travels”.

In this process, the definition of (urban) landscape powerfully shifts from a definition which is purely administrative (e.g.: the borders of the flower bed in the middle of a roundabout) to one that is multiplied according to all individuals which experiences that location; as a lossless sum of their perceptions; as a stratification of interpretations and activities which forms our cognition of space and time, just as suggested in the theories expressed in [18] and [19].

Furthermore, the possibilities for expansion do not end in this single direction, as the opportunity to access and experience the multiplicity of layers of information stratified by people and organizations onto the spaces we traverse constitute the chance to contaminate our perception of space/time according to entirely different sets of assumptions, originating, possibly, from cultures, backgrounds, beliefs and visions which are completely different from our own and which, at the same time, are intimately interconnectable to our perception, as they become easily accessible from devices we hold in our pockets, wear or, in easily imaginable future scenarios, directly see superimposed onto our field of vision.

When Derrick de Kerckhove in his [20] describes a scenario in which it is possible to expand the concepts of the world wide web to the architectures of our cities, he describes the possibility to expand our possibilities for awareness and consciousness, through the wide, ubiquitous availability of multiple sources of information which can be interrelated through hyperlinking and what today would be probably referred to as a personal curation tool, defining our point of view on the world through a collection of interconnected (real-time) flows of information. This approach is twofold, as the action of forming such a point of view on reality is an act of construction and of collection at the same time [21] and, thus, the possibility to form knowledge and awareness about places by performing these forms of cognitive mash-up and remix processes among our knowledge and awareness and the ones of other individuals and organizations (and of the place itself, if we bring into the scene the possibility for the territory to “speak for itself” by using sensors) represents a possibility to define a novel form of intelligence, coagulating and interconnecting in multiple ways the diffused intelligence expressed through ubiquitous digital devices and networks.

In [22] Green investigates these phenomena from the point of view of time: the emergence of new temporalities through the usage of mobile devices, at institutional, social and subjective levels, through the evaluation and interpretation of extensive ethnographical

data sets. What emerges is that mobile devices act as spatial/temporal mediators, exposing alternative perceptions and behaviors in human beings using them and, thus, proposing different usage grammars for spaces and timeframes.

The concept analyzed in [22] over the domain of time can, in fact, be extended to space, as Berry and Hamilton have done in [23], among many others. By analyzing mobile devices usage on trains, Berry and Hamilton highlight how “public places and spaces are being transformed into hybrid geographies through the introduction of new spatial infrastructure” and, thus, how the use of these devices are enabling people to systematically re-program their surroundings according to the opportunities offered by their possibility to access ubiquitous technologies and networks.

III. CONNECTICITY

In our research we investigated the possibilities to use the scenario which sees urban spaces progressively filling with multiple layers of real-time, ubiquitous, digital information to conceptualize, design and implement a series of usage scenarios which would pursue the following goals:

- △ to create a set of tools that would allow us to capture in real-time various forms of city-relevant user generated content from a variety of types of sources, including social networks, websites, mobile applications, to interrelate it to the territory [8][24] using Geo-referencing, Geo-Parsing and Geo-Coding techniques [25][26][27][28], and to analyze and classify it using Natural Language Processing and Named Entity Recognition techniques to identify users' emotional approaches, forms of expression, topics of interest, discussion graphs, networks of attention and of influence, trending issues, evaluations of satisfaction, well-being and happiness, and other forms of expression [29][30][31][32][33][34][35][36][37][38];
- △ to make this information available and accessible both centrally and peripherally, to enable the creation of novel forms of decision-making processes [39][40] as well as to experiment innovative models for participated, peer to peer, user-generated initiatives that would be able to reflect onto the life and expressions of the city and of its inhabitants, to identify new policies, new sustainable, ethical business models, new urban planning processes, new grass-roots initiatives, new operative models along the lines described by concepts such as living labs and other user-centric innovation processes [41][42][43][44];
- △ to reflect on the themes of cognitive accessibility for this kind of information, analyzing visual and multi-modal representation and interaction metaphors that would allow to maximize the effectiveness and ease of use and understanding of the complex information

scenarios produced by the possibility to capture and ubiquitously display large quantities of geo-located, realtime information layers coming from multiple sources [45][46][47][48];

- △ to confront with validation models that would allow us to assess the quality, relevancy and reliability of harvested data, affected by information noise, digital-divide related issues (e.g.: not all citizens use social networks or imagine that they can use them to express opinions about their city), interpretation errors;
- △ to imagine and implement an openly accessible service layer based on widely accepted open standards that would allow other agencies to use the technologies and processes to easily design and build their own systems and services.

Along the process we focused on the implementation of a multidisciplinary research agenda involving designers, artists, architects, engineers, anthropologists, public administrations, organizations, private organizations and active grassroots communities, each providing knowledge, skills and visions on multiple perspectives of the same issues.

IV. RESULTS

The research project started in 2008 and is still ongoing. Up until the current date, it produced a series of different results.

Presented at the 7th International Meeting for the Revitalization of Historical Centers, in 2008 in Mexico City, *architettura rel:attiva* presents the initial scenario for *ConnectiCity*, in which ubiquitous technologies and hybrid practices offer a different vision about what is/can become architecture, together with an idea on the new models for political representation and interaction for urban ecosystems[49].

The Atlas of Rome[50], a 35 meters long architectural surface created in 2010 in occasion of the *Festa dell'Architettura* of Rome, starts from this initial experience, and thanks to a progressively wider availability of ubiquitous technologies.

The interactive surface realized a promenade allowing citizens to have a real-time experience about the information generated by other citizens about the city.

Information was gathered and visualized around the visions on the city, creating an emblematic space in which to come into contact with the desire for evolution and mutation constantly expressed throughout our day-to-day activities. Ideas for the re-use of the spaces of the city; events and occasions for debate; expressions about the fundamental issues of the city; and description of the emotional states of citizens.

Information was represented through several infoaesthetic visualizations that were able to make it accessible and suggestive, so as to establish processes of cognitive facilitation and emotional engagement which

would activate citizens that, by using their mobile phones and some touch screens, could take part in the urban conversation.

A local, smaller version, ConnectiCity Neighborhood edition [51], implemented an awareness mechanism for the life in the single neighborhoods of the city. Presented in 2010, the urban screen prototype allowed reading in real-time the conversations that the inhabitants of the neighborhood were exchanging using social networks, highlighting the more influential themes and the main flows of interaction. The digital life of neighborhoods.

CoS, Consciousness of Streams [52], was presented at the 2010 edition of the Transmediale festival in Berlin, as an experiment in which the techniques for ubiquitous and participative publishing became tools for the observation of the emotional states expressed by urban space dwellers.

An infoaesthetic real-time representation showed the emotions expressed by people across different cities using smartphones and digital kiosks. An emergent global map of the spatial and temporal emotions of the planet emerged from the process.

The most recent addition to the ConnectiCity project is VersuS, started in occasion of the riots which took place on October 15th 2011 in Rome. It has been officially presented at the Regional Museum for Natural Sciences in Turin, during the Piemonte Share Festival [53].

VersuS is a real-time system which harvests the conversations that the citizens of a certain cities engage using social networks.

The conversations are analyzed using Natural Language Analysis techniques, to understand the themes and approaches at the level of emotion and interrelation. Through the system it is possible to select themes, emotions and entire discussions, using them to visualize geo-referenced three-dimensional surfaces whose elevation corresponds to the instantaneous intensity of the selected theme in different parts of the city.

Moving back and forward through time it is, thus, possible to observe the elevation of the surface over the city, showing the mutation of the intensity according to which the topic is being discussed in that location by social network users [54].

V. CONCLUSIONS

The possibility to listen to the ideas, visions, emotions and proposals which are expressed each day by citizens – either explicitly or implicitly by the ways in which they use their cities, workplaces, malls... – suggests the emergence of positive scenarios.

Harvesting systems allow us to continuously sense the public discussion and to correlate it to cities, transport systems, infrastructures, architectural spaces, neighborhoods.

“Sensibility Networks” can be established using natural language analysis processes allowing us to “read”

cities, for how it is “written” by people, traversing languages and cultures.

Sensor networks can be included in the scenario to record in real-time information about pollution, traffic and the other measurements which shape the ecological, social, administrative and political lives of our cities.

It is possible to create multiple layers of narratives which traverse the city and which allow us to read them in different ways, according to the different strategies and methodologies enabling us to highlight how cities (through their citizens or even on their own, expressing through sensors) express points of view on the environment, culture, economy, transports, energy and politics.

These methodologies for real-time observation of cities can be described as a form of “ubiquitous anthropology”, based on the idea that we can take part in a networked structure shaped as a diffused expert system, capturing disseminated intelligence to coagulate it into a framework for the real-time processing of urban information.

In this context infoaesthetic representations become enablers to enact radical strategies to maximize the accessibility and usability of this information.

Together with the idea of “user generated search engine” they contribute to give shape to a scenario in which the concepts of citizenship and political representation can be reinvented, tending towards a vision in which people can be more aware and benefit from added opportunities for action, participating to an environment designed for ubiquitous collaboration and knowledge which is multi-actor, *multi-stakeholder*, in real-time: the city.

A series of critical issues have also been identified throughout the experiments and their assessments, establishing the future paths of the research:

- ▲ the lack of extensive research in the field, leading to the necessity of performing massive tests and to engage multiple other researchers in validating the results which progressively are produced;
- ▲ the need to attribute major importance to the issues of digital divide in understanding the relevance of the produced results, now and in the near future;
- ▲ the need to assess the issues of digital inclusion in the research scenarios, addressing the cultural dimension of “inclusion” and the need to understand diversity in cities, the effects it produces and the opportunities which it opens up;
- ▲ the need to go beyond the idea of “buzz” in understanding of the ways in which it is possible to conceive inference methodologies which allow to identify and classify emotions, wishes, visions, desires and expectations expressed by people, in multiple languages and

contexts, accepting the complexity which is needed to approach the important issues of our cities;

- ▲ the need to seriously confront with the ethical issues raised by these kinds of processes, including the changing ideas of privacy, intellectual property, access and the mutating definitions of public and private spaces.

ACKNOWLEDGMENT

We would like to thank all the researchers, artists, professionals, public and private organizations which made this research process possible, with specific highlight to the FakePress Publishing Think Tank, Rome's City Administration, Franco Aviccolli and Massimo Canevacci Ribeiro.

REFERENCES

- [1] J. D. Porteous, "Home: the territorial core", *Geographical Review*, Vol. 66, N.4, pp. 383-390, Oct. 1976.
- [2] S. L. Scheiberg, "Emotions on display: The personal decoration of work space", *American Behavioral Scientist*, Vol 33(3), 330-338, Jan-Feb 1990.
- [3] E. Krupat, W. Guild, "Defining the City: The Use of Objective and Subjective Measures for Community Description", *Journal of Social Issues*, Issue 36 pp.9-28, 1980.
- [4] J. L. Nasar, "Perception, cognition and evaluation of urban places", in I. Altman and E. Zube (Eds.) *Human Behavior and Environment: Public places*, pp. 31-56, 1989. New York: Plenum.
- [5] M. Gottdiener, "The social production of urban space", pp. 28. Austin: University of Texas Press, 1994.
- [6] F. E. Horton, D. R. Reynolds, "Effects of Urban Spatial Structure on Individual Behavior", *Economic Geography, Perspectives on Urban Spatial Systems*, Vol. 47, No. 1, pp. 36-48, Jan. 1971.
- [7] M. Zook, M. Graham, "From Cyberspace to DigiPlace: Visibility in an Age of Information and Mobility", in: H. J. Miller, Ed., *Societies and Cities in the Age of Instant Access*. London: Springer, 2007.
- [8] M. Zook, M. Graham, "Mapping DigiPlace: Geocoded Internet Data and the Representation of Place", in *Environment and Planning B: Planning and Design*, 34, 2007.
- [9] J. Zeisel, "Inquiry by Design: Environment/Behavior/Neuroscience in Architecture, Interiors, Landscape, and Planning", 2006. W.W. new York: Norton & Company.
- [10] M. McCullough, "Digital ground: architecture, pervasive computing, and environmental knowing". Cambridge Mass.: MIT Press, 2004.
- [11] K. Fattahi, H. Kobayashi, "New era, new criteria for city imaging", *Theoretical and Empirical Researches in Urban Management*, Vol. 3, N. 12, Aug. 2009.
- [12] W. J. Mitchell, "City of Bits: Space, Place, and the Infobahn". Cambridge MA: MIT Press, 1995.
- [13] P. Du Gay, "Doing Cultural Studies: the story of the Sony Walkman", pp.17. London: Sage, 1997
- [14] M. Bull, "Sounding out the city: personal stereotypes and the management of everyday life". London: Berg, 2000.
- [15] S. Iaconesi, O. Persico, "RWR Read/Write Reality Vol. 1", Rome: FakePress Publishing, 2011.
- [16] R. Wilken, "From Stabilitas Loci to Mobilitas Loci: Networked Mobility and the Transformation of Place", *Fibreculture Journal*, Issue 6, Mobility, New Social Intensities and the Coordinates of Digital Networks, 2005.
- [17] D. Morley, "What's "Home" Got to Do with It?: Contradictory Dynamics in the Domestication of Technology and the Dislocation of Domesticity", *European Journal of Cultural Studies* 6.4 (2003), 435-458.
- [18] J. P. Eberhard, "Brain landscape: the coexistence of neuroscience and architecture", Oxford: Oxford University press, 2009
- [19] A. Farina, "Ecology, Cognition and Landscape", New York: Springer, 2010.
- [20] D. de Kerckhove, "The architecture of intelligence", Basel: Birkhäuser, 2001.
- [21] J. Piaget, "The psychology of Intelligence", London: Psychology Press, 2001.
- [22] N. Green, "On the Move: Technology, Mobility, and the Mediation of Social Time and Space", *The Information Society*, Vol. 18, N. 4, pp. 281-292, London: Routledge, 2002.
- [23]
- [24] M. Berry, M. Hamilton, "Changing Urban Spaces: Mobile Phones on Trains", *Mobilities*, Vol. 5, N.1, pp. 111-129, London: Taylor & Francis, 2010.
- [25] M. F. Goodchild, "Citizens as sensors: the world of volunteered geography," *GEOJOURNAL*, p. 211-221.
- [26] M. D. Lieberman, H. Samet, "Multifaceted toponym recognition for streaming news", in *SIGIR '11 Proceedings of the 34th international ACM SIGIR conference on Research and development in Information, Proceedings*, pp. 843-852. 2011. New York: ACM.
- [27] T. Quin, R. Xiao, L. Fang, X. Xie, L. Zhang, "An efficient location extraction algorithm by leveraging web contextual information", in *GIS '10 Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, Proceedings*, pp. 53-60. 2010. New York: ACM.
- [28] J. L. Leidner, M. D. Lieberman, "Detecting geographical references in the form of place names and associated spatial natural language", in *SIGSPATIAL Special, Newsletter, Special Issue, Volume 3, Issue 2*, pp. 5-11. July 2011. New York: ACM.
- [29] G. Shi, K. Barker, "Thematic data extraction from Web for GIS and applications", in *Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on, Proceedings*, pp. 273-278. 2011. Fuzhou: IEEE.
- [30] M. N. Kamel Boulos, A. P. Sanfilippo, C. D. Corley, S. Wheeler, "Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance", in *Computer Methods and Programs in Biomedicine*, Vol. 100, Issue 1, pp. 16-23. October 2010. London: Elsevier.
- [31] S. D. Kamvar, J. Harris, "We feel fine and searching the emotional web", in *WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining, Proceedings*, pp. 117-126. 2011. New York: ACM.
- [32] S. Abe, K. Inui, K. Hara, H. Morita, C. Sao, M. Eguchi, A. Sumita, K. Murakami, S. Matsuyoshi, "Mining personal experiences and opinions from Web documents", in *Web Intelligence and Agent Systems*, Vol. 9, Number 2/2011. 2011. Amsterdam: IOS Press.
- [33] L. Gentile, V. Lanfranchi, S. Mazumdar, F. Ciravegna, "Extracting Semantic User Networks from Informal Communication Exchanges, in The Semantic Web", *ISWC 2011, Lecture Notes in Computer Science*, 2011, Volume 7031/2011, pp. 209-224. 2011. New York: Springer Link.
- [34] D. Agrawal and C. C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms", In *Proceedings of the twentieth ACM PODS*, 2001. 2001. New York: ACM.
- [35] A. McCallum, "Information Extraction: Distilling Structured Data from Unstructured Text", in *Queue - Social Computing*, Vol. 3, Issue 9, pp. 48-57. November 2005. New York: ACM.
- [36] J. Lin, R. Snow, W. Morgan, "Smoothing techniques for adaptive online language models: topic tracking in tweet streams", in *KDD '11 Proceedings of the 17th ACM SIGKDD*

- international conference on Knowledge discovery and data mining, pp. 422-429. 2011. New York: ACM.
- [37] J. Yi, T. Nasukawa, R. Bunescu, W. Niblack, "Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques", in Data Mining, 2003. ICDM 2003. Third IEEE International Conference on , pp. 427-434. 2003. Melbourne, FL, USA: IEEE.
- [38] C. Strapparava, R. Mihalcea, "Learning to identify emotions in text", in Proceeding SAC '08 Proceedings of the 2008 ACM symposium on Applied computing, pp. 1556-1560. 2008. New York: ACM.
- [39] C. Strapparava, R. Mihalcea, "Annotating and Identifying emotions in text", in INTELLIGENT INFORMATION ACCESS, Studies in Computational Intelligence, 2010, Volume 301/2010, pp. 21-38. 2010. New York: Springer.
- [40] K. Snyder, "Tools for Community Design and Decision Making", Planning Support Systems in Practice, Springer. 2003
- [41] C. Alexander, S. Ishikawa, M. Silverstein, M. Jacobson, I. Fiksdahl-King, S. Angel, "A pattern language. Towns, buildings, construction", Computer and Information Science, Oxford: Oxford University Press. <http://www.amazon.fr/exec/obidos/ASIN/0195019199/citeulike04-21>.
- [42] N. A. Salingaros, "The structure of pattern languages", Architectural Research Quarterly, N. 4, pp. 149-162, Cambridge: Cambridge University Press, 2000.
- [43] H., Schaffers, A. Sallstrom, M. Pallot, J. M. Hernandez-Munoz, R. Santoro, B. Trousse, "Integrating Living Labs with Future Internet experimental platforms for co-creating services within Smart Cities", Concurrent Enterprising (ICE), 2011 17th International Conference on, pp. 1-11: Aachen: IEEE, 2011
- [44] M. Mulvenna, B. Bergvall-Kåreborn, J. Wallace, B. Galbraith, S. Martin, "Living labs as engagement models for innovation", eChallenges, 2010, pp. 1-11, Warsaw: IEEE, 2010.
- [45] E. Tufte, "Visual Explanations: Images and Quantities, Evidence and Narrative", Cheshire, CT: Graphics Press. 1997.
- [46] W. S. Cleveland, R. Mc Gill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods", Journal of the American Statistical Association, 79,387(1984), pages 533, 554.
- [47] C. Ware, "Information visualization: Perception for design", Morgan Kaufmann, second edition, 2004.
- [48] B. Wünsche, "A survey, classification and analysis of perceptual concepts and their application for the effective visualisation of complex information", APVis '04: Proceedings of the 2004 Australasian symposium on Information Visualisation, Darlinghurst, Australia, 2004, Australian Computer Society, Inc., pages 17-24
- [49] S. Iaconesi, O. Persico, "Architettura rel:attiva", Proceedings of The Seventh International Meeting on the Revitalization of the Historical Centers. 2008. Mexico City: INAH publishing.
- [50] Atlante di Roma, <http://www.artisopensource.net/2010/06/07/atlas-of-the-visions-festa-dellarchitettura-index-urbis-rome/>, <http://www.artisopensource.net/2010/06/13/atlante-dell-visioni-atlas-of-the-visions/>, <http://www.artisopensource.net/2010/07/02/a-video-atlante-di-roma-atlas-of-rome/>
- [51] ConnectiCity Neighborhood edition, <http://www.artisopensource.net/2010/09/20/cities-telling-stories-atlas-connectivity-reality-and-a-nice-event/>
- [52] CoS, Consciousness of Streams, <http://www.artisopensource.net/category/projects/consciousness-of-streams-projects/>
- [53] VersuS, the realtime life of cities, <http://www.artisopensource.net/2011/11/06/versus-the-realtime-lives-of-cities/>
- [54] VersuS, loveVSturin, a video, <http://vimeo.com/30976473>



Workshop on Educational Services for the Performing Arts

Chair

Peter Eversmann, University of Amsterdam (The Netherlands)

Workshop Committee:

Ad Aerts, independent IT and arts advisor, NL

Robin Boast, University of Cambridge, UK

Josefien Schuurman, University of Amsterdam, NL

Christine White, University of Derby, UK

Performance Art Studies At The Dawn Of The Digital Age

Dorota Sosnowska

Warsaw University, Institute of Polish Culture, IKP, Warsaw, Poland, d.e.sosnowska@gmail.com

Abstract—This paper addresses the issue of the material source in the performance art studies in the era of digital media. My goal is to define the challenges and questions, which emerge from usage of platforms such as ECLAP in the educational process. I use the context Polish theater studies to show what are possibilities and dangers in this subject. Firstly, I present the theoretical perspective on this subject, referring to Walter Benjamin's theory and new materialism. Then, I analyze the usage of new tools, or rather their problematic status, in the context of Polish academic education.

Keywords-source, digital era, education, ECLAP, Poland

I. INTRODUCTION

Walter Benjamin was the first to name the “mechanical reproduction era” as the one that radically changes the status of the piece of art. Photography doesn't have the same positing as painting or sculpture - something that exists in one copy, so there can be no doubt about its originality and thus, status. Benjamin called it *aura* - something close to a metaphysical feeling of the relation between the spectator and something that is original and unique. With the development of the advance mechanical reproduction techniques, the copy and the original are hard to distinguish. In the eyes of many, art is losing its' aura. For culture it means more complexity. Implications of this phenomenon are known in the history of art and visual theory, but my impression is, that it hasn't been recognized in the field of performance art studies yet. The first questions that arise, when we think about Benjamin's theory is, what is the actual source in the performing arts studies? What is the subject of our research? Is *aura* possible? Is there an original and therefore can copy exist at all? What would “reproduction” mean?

II. FROM THEATRE STUDIES TO NEW MATERIALISM

In Poland theatre studies were defined as a scientific discipline at the beginning of the post second world war period. One of many discussions among the researchers, taking place in 70's, was about the need for specific methods to study performances. The posed problem was: In order to start the analysis and interpretation of a spectacle one needs to reconstruct it somehow - describe it, transcribe it into a text form. This “description” that would qualify as a scientific method was something that

researchers needed to establish. The most obvious one was that the researcher should see the spectacle more than once, each time taking a different place in the theatre. Then, one should collect interviews with the public, choosing different age groups and social classes. One should also have three recordings of the spectacle at disposition: one made from the left corner, one from the right corner and the last made from the place in front of the scene. The copy of texts used by the director should be collected, together with set design plans, photos, different essays and reviews about the spectacle. Collecting this documentation would be the first step in the process of the *description* of the spectacle. Additionally, researchers have requested that theaters should collect all these materials and preserve them in their archives for the sake of historical studies. Apart from the need of being accurate and not to lose anything that could help to reconstruct the spectacle, this great effort was also a sign, that this new discipline - theatre studies - wanted to be treated as a fully scientific and serious discipline, one that is able to give some objective results of the inquiry. Researchers using methodologies like structuralism and semiotics believe that it is possible to establish such a *description*, to collect such amounts of documents that it would preserve the definitive and objective form of the spectacle. The heart of this proposed methodology was finding the material source – the documents that could always be a reference and could abolish the problematic non-existence of the spectacle. But what was the effect of this scrupulous collection? That it didn't really bring us any closer to the spectacle itself. It was a big closet of different documents presenting separate linear narrations.

When the cultural studies appeared in Poland, a huge shift in attitude towards the performance art studies could be observed. At the Institute of Polish Culture the Section for Theatre and Performance had been established. Its work concentrates on the cultural performances, particularly the theatre performances. Cultural spectacles and performances, defined as organized events, that proceed according to definite dramatic scenarios – events like: ceremonies, sports competitions, cult celebrations, carnival processions, and theatre and dance performances - where the performativeness of social agency finds its climax. Cultural performances materialize the myth of society's integrity and legitimize transformations of its structures; hence, they embody the social life in its most intensive

manifestations. In the essence they embody culture, perceived as a whole, showing its most dynamic and intensive powers, as a centre of festive exultation. In this context, theatre is embedded in social life as an important kind of meta-social commentary - a story that each community tells itself about itself. Thanks to this methodology, the spectacle is seen as the representation of social tensions. The researcher, as a member of the society, can base the analysis on the personal experience, using tools of cultural anthropology - theatre becomes field as it is understood in ethnography.

However, when it comes to the history of theatre, once again we face the problem of the source. It's not only the impossible description of the spectacle that is the problem; we need also to find accurate social context, which needs different kind of material source, than reviews and recordings. With anthropology of performance, the problem of the research source becomes even more important.

The cultural studies, that brought to the humanistic disciplines certain laxity of rules, today are trying to bring to life old-new methodologies like new empiricism, new historicism, new materialism. New empiricism, which derives from the works of Gilles Deleuze, brings the subject of research to the process of perceiving something. As Deleuze was writing in *Repetition and Difference*: "Empiricism is by no means a reaction against concepts, nor a simple appeal to lived experience. On the contrary, it undertakes the most insane creation of concepts ever seen or heard. Empiricism is mysticism and mathematicism of concepts, but precisely one which treats the concept as an object of an encounter, as a here-and-now, or rather as an Erehwon from which emerge inexhaustibly ever new, differently distributed *heres* and *nows*. Only an empiricist could say, concepts are indeed things, but things in their free and wild state, beyond *anthropological predicates*. I make, remake and unmake my concepts along a moving horizon, from an always decentered center, from an always displaced periphery which repeats and differentiates them..." [2] The philosopher is trying to understand how he sees, perceive his own perception, and further to conceptualize his process of thinking. What is empirically perceived, is the process of that perceiving, what is empirically approachable is the concept itself. The source, as a subject of perception, gains new importance and takes the role of the Derridian trace.

"The incipience of the New Historicism approach occurred as a response to several factors, among them the New Criticism's tendency to analyze literary works as if they had been produced in a historical vacuum, i.e., as if a poem or a novel would hold no relation to the historical context that gave birth to it. Furthermore, the 1960's recognized the need for political developments that materialized in the form of a desire, especially on the part of American literature professors, to comprehend the role of literature in the understanding of social and historical issues. Roughly at the same time, there was an influx of professors in American and British universities, mainly from two different backgrounds, whether they

were from continental Europe, carrying continental modes of thinking literary theories, or they were female, working class, Italian-American, African-American, Asian-American or Latin-American scholars, who were holding high teaching positions for the first time, with the fresh ideas and ideals. These professionals sought to understand their historical context in many ways, and literature arose as one of them, a particularly elegant solution and an incredibly intelligent source of information".[2] In new historicism then, sources are understood in a very broad way. Documents, images, texts but also discourses and stories - everything is used to build the context of the research. This great historical apparatus is used here to trace the changes and processes in culture. The source is important, not only as a medium that brings some light on the facts, but it's material status is also interesting - what it is and how it appears? Shape, font, handwriting character, colors, paint, type of used film etc. - everything is meaningful for those, who use new historicism as a methodology.

With the possibilities of digital era the source is no longer the same as it was. Still, of course, we have recordings, reviews, interviews, trailers, texts, set design projects and photos, but as we see on ECLAP, they don't function separately anymore. This is not a closet full of different things but a new kind of medium that gives us possibility to see those documents in synchronic manner. We can build a collection, we can use annotations to link chosen parts of one document to another. This gives us something that no longer resembles old kind of sources. Before we start writing our own thesis using ECLAP tools, we build something new - a secondary source - that has some features and characteristics on its own. The problem that we have to face is that it is not material anymore. We are dealing more with a medium than with the document - what is provided by digital platform works like intermediary between us and the content we are looking for. The old questions, inspired by Benjamin's theory are valid again - this time it's about the digital or digitized source. What is the status of it?

New materialism seems to bring especially interesting responses. With this perspective the difference between material and virtual is abolished and a medium ceases to be something that mediates between the real and us - the viewers - and becomes something of material nature by itself. "Here the world is not reduced to symbolic, signifying structures, or representations, but is seen by such writers as Friedrich Kittler (and more recent theorists such as Wolfgang Ernst in a bit differing tone of media archaeology) as a network of concrete, material, physical and physiological apparatuses and their interconnections, that in a Foucaultian manner govern whatever can be uttered and signified" [1]. In virtual environments the reflection about the difference between original and its' copy becomes out of scope. We deal with different material objects that are not "reproductions" - they have their own status and characteristics not additional to the "original". This means, that digital platforms, like ECLAP, offer new materiality that can't be read in an old manner and needs new integration in the learning and academic process.

Returning to the first question: how should we then understand the material source in the performing art studies? What is its status and how does it change with the emergence of new technologies? I think that we should be conscious of the fact, that the ideal description of the performance is not possible. Even if we would gather all the possible documents, the spectacle is something that is only temporarily available. It is finished before we start our analysis. There is no possibility to grasp it in any medium we have at our disposition. In my opinion the solution is to treat the sources we can get as separate emanations of one artistic idea. We should use tools and theories that are accurate for this specific kind of medium we are working with and we should try to discover that idea – not the perfect description of the performance as it happened. The Benjamin's dilemma is important to see that in the specific situation of performing arts studies there is no original available, although it's not the mechanical reproduction that makes it disappear. We could even say that performance is in fact the medium itself - it mediates between some possible reality and viewers. "Theater is a medium that, from Plato and Aristotle to the present, has been regarded with suspicion, fear, and contempt – but also with fascination and desire – by a tradition seeking at all costs to keep the ground from slipping out from under its feet. The twist and turns of this medium, in its theory as well as its practice, are perhaps even more acute today, when the notion of *media* has become more ubiquitous and more elusive than ever before. What we call *theatre* and, even more, *theatricality* provides an instructive arena for the examination of those *twins and turns*. *Multimedia* long before the word became a cliché, Western theater long occupied an uneasy position between *art* and *entertainment*, between discovery and manipulation, and this situation has not changed"[2]

In this context the digital era brings possibility to create even more adequate type of material source than it was ever possible. With Internet, new tools had been offered: platforms like ECLAP. Its' not-linear and not-textual character of narrations bring us even closer to the multi-media and interactive vibe of the spectacle.

But how the possibilities offered by digital platforms are in fact used in education? Can we observe any impact of this new reality on the academic practice?

III. E-LEARNING AND DIGITAL LIBRARIES IN POLISH CONTEXT

In Poland, websites like ECLAP and other digital libraries are commonly used by individuals (academic researchers and students). They are not incorporated however, in educational programs and classes.

E-Learning at Warsaw University - the biggest university in Poland - consists mostly of virtual classes that provides different materials in combination with some questions to answer, tests, short texts to be prepared by student or forum discussions. At the end of these courses, there is a face-to-face meeting with the teacher during which students pass their exam in written or spoken manner. These courses are independent from other classes. They use Internet as transparent (not

problematized during learning process) medium of communication and have the ambition of being analogous to the traditional classes. They are based on the Moodle platform. From what I know it is not a favorite type of classes for our students. The whole issue of virtual tools for education seems fairly neglected in the everyday life of the university.

The fact that both sides – students and academics – seem to ignore the new educational (virtual) reality results in a rather unscientific situation: students write papers about spectacles, apparently using some materials that registered the event, although there is no sign of the selected source in the text. The fact that it wasn't a personal experience is hidden. In the analysis, there is no sign of the public, the specific context or the uniqueness of the concrete performance. The spectacle becomes objectified like its registration would be its definitive and obvious form. Writing about performance without proper acknowledgement of the used sources results in a methodological disorder between old and new paradigms, that in combination with a fundamental change of nature of 'the material source' can create a crisis in performing arts academic learning.

Even more striking situation can be observed, when students perform their analysis based on vast amounts of reviews, in their papers providing references to concrete newspapers and magazines. It's obvious however, that they are using a popular digital platform: www.e-teatr.pl. This is the site of Theatre Institute in Warsaw. This public institution was created in 2003 with an aim to collect the main theatre archive for whole Poland. They obtained the materials and documents collected by other institutions that ceased to exist, as they were founded in the communist times. Those documents are held in magazines of the Institute, which is trying to digitalize them consecutively. Documents are scanned, converted to the text files and put on the website where, cut out from the original source, they are listed by the name of an actor, director or title of the spectacle. The academic system puts a sense of guilt on the students that they use sources stripped from original context. They try to hide the fact that they are using such digital platforms. Both sides (students and academics) not only neglect the importance of e-learning and don't take the responsibility for their quality, but also blur out the obvious difference between the scientific significance of traditional sources and its digital version. Although digitized materials have neither print nor characteristics, they can't be perceived in the old-fashioned context of the newspapers or magazines, where we are dealing with something that can be *described*, with something that has its own "materiality". That's why the change of perspective seems to be so important. We can't treat those objects like "worse" kind of source, we can't hide the fact that they are something else than newspaper article and simply proceed with our analysis like nothing has changed. Neglecting the fact that digital objects are something different, although still material, can result in falsification of the source and thus can endanger scientific discourse. I would like to emphasize that I don't want to say that the decontextualized sources on

platforms like e-teatr are scientifically less useful – we just need to look at these materials in a new way.

The other question concerns, of course, the scientific quality of metadata that are provided by any digital platform. The e-teatr, like most of the similar platforms, has low standards of metadata. This is the difference that distinguishes ECLAP. I think that with those new, high standards of scientific description, ECLAP will be treated as reliable source of knowledge, which will also help to integrate this platform with academic education and thus, conceptualize character of digitized sources in academic research.

On a digital platform we are provided different kinds of objects: photographs, reviews, posters, set design plans - all mediated by the platform, all “not-original” and all brought together to construct something new - one big “representation” of the spectacle. Because these materials are not put in an objective order (I can choose how I want to look at them), they are not a part of the linear narratives I suppose to look at these materials (I would like to emphasize the word “look”) not like at a book or collection of texts, but as at visual fields (as it is understood in cultural anthropology) As W.T.S. Mitchell states formulating the definition of the new approach to the visual culture: “I propose what I hope is a more nuanced and balanced approach located in the equivocation between the visual image as instrument and agency, the image as a tool for manipulation, on the one hand, and as an apparently autonomous source of its own purposes and meanings on the other. This approach would treat visual culture and visual images as go-betweens in social transactions, as a repertoire of screen images or templates that structure our encounters with other human beings”.[3].

What differs so called “visual fields” from the traditional approach to the image is, that accent is put on the relations between looking, being looked at, between the objects, the viewer and the one who makes you view. This is the field, where all those different angles and interest are crossed, so the analysis consists of finding and describing those relations. They are like a line that in a simple picture separates the object from the background. In visual studies this object is the culture itself. In the performance studies that would be the performance we are looking for. In my opinion the relations between digital objects cut from their original context and put together in the virtual space could be the key to the interpretation of the new kind of source that we can access with ECLAP.

With tools such as collections and annotations we can establish those relations literally. Thanks to that process we can see digitized objects contextualized again – among themselves. This is the kind of exercise that seems very productive within the field of education: to ask students to make their own “visual fields” with the tools (like collections, annotations, mystoryplayer) that ECLAP provides. This work could be discussed and problematized during the classes - then the e-learning tools would be integrated in the learning process that takes place at the university. I think that using ECLAP in this manner would help to overcome the potential learning crisis giving students and academics possibility to work with new kind of source without neglecting it's digital nature.

When this kind of “picture” of the spectacle is made we need to ask another question: what is its’ status? It’s not transparent, it’s not objective, it is not something we can “hide” in our inquiry making the impression, that we still write about the spectacle, we’ve just seen, or we have some irrefutable knowledge about. That’s why, I think that usage of this kind of educational tools, which is indispensable, results in the big change of the performance art studies itself. The written interpretation it’s not enough to show the complex nature of the subject we are dealing with. The visual need to enter the text itself - need to revile what kind of source we are working with, show pictures, maybe even the fragments of registrations, the relations between documents. After the work we are doing to construct our source, it becomes the part of interpretation and should be treated like the written text itself. Soon platforms like ECLAP won’t be only treated like libraries but they will be the tool to write the actual thesis and texts. I believe that without the possibility of multi-media narrations, we won’t be able to write about theatre anymore.

REFERENCES

- [1] Gilles Deleuze, “Difference and Repetition”, Clumbia University Press 1994.
- [2] Gustavo Vargas Cohen, “Stephen Greenblatt's New Historicism: Origins, Premises, Complaints, Critics, Texts and Trends”, <http://voices.yahoo.com/stephen-greenblatts-historicism-origins-premises-5880948.html>
- [3] Milla Tiainen and Jussi Parikka, “New Materialism and Digital Culture”<http://machinology.blogspot.com/2010/06/what-is-new-materialism-opening-words.html>
- [4] Samuel Weber, “Theatricality as Medium”, Fordham Univ Press 2005.
- [5] W.J.T. Mitchell, “Showing seeing: a critique of visual culture”, *Journal Of Visual Culture*, 2002, Vol 1(2), p. 170.

Open Networks for Culture and Education: ONCE

John Casey⁽¹⁾, Wolfgang Greller⁽²⁾, Gareth Waller⁽³⁾

(1) Centre for Learning & Teaching in Art & Design (CLTAD), University of the Arts London, UAL, London, United Kingdom, j.casey@arts.ac.uk

(2) Open University of the Netherlands, OUNL, Heerlen, Netherland, wolfgang.greller@ou.nl

(3) EDINA, University of Edinburgh, Edinburgh, United Kingdom, g.waller@ed.ac.uk

Abstract— Open cultural collections and open educational resources face many similar challenges and occupy the same problem space in terms of user needs and technical requirements. To meet these needs in a sustainable and scalable way a simple shared technical architecture is proposed that moves away from the narrow focus on technical concerns of current repository software platforms and their associated expert communities. Instead, the ONCE philosophy envisages that experts and specialist services will work in tandem with the open user-controlled systems described in this paper. The role of the digital commons and licensing is a key enabler for the technical architecture. The paper is based on the experience of an open education project at the University of the Arts London and draws on the experience of establishing the proposed architecture in a real institutional setting.

Keywords— open education; open educational resources; culture, repositories; metadata, creative commons licences; socio-technical; ALTO; University of the Arts London.

I. A DIGITAL COMMONS FOR EDUCATION AND CULTURE

The idea of ‘The Commons’ as a shared space where resources are held for the common good is an enduring one in human history. After agitation and social conflict in the 13th century in England, the monarchy was forced to cede control of large parts of land for common use by the population under the *Carta de Foresta* (Charter of the Forest). This is the lesser-known companion to the more famous Magna Carta that is widely regarded as a milestone in the evolution of the legal articulation and protection of human rights. More recently, during the current global economic crisis various protest movements such as ‘Occupy Wall Street’ have been articulating the need for common ownership of resources for the benefit of society as a whole, including the banking system.

The debate about common ownership and/or access to resources has also been revitalised in the digital realm by developments to promote open access to research and open educational resources (OERs) and open educational practices. Questions and issues raised include:

- What are the implications of the commons for students, teachers, institutions and society? If the commons are characterised by an economy of ‘plenty’ and open access, what is the future for

the current institutions of an educational and cultural economy that are based on scarcity and restricted access? How may they evolve to include the Commons?

- How can accreditation of learning be effectively endorsed in the Commons in a trustworthy way?
- How might the growth of the commons change the roles of; the creators of cultural artefacts, teachers, librarians, archivists, educationalists and technical staff etc.?
- What kind of knowledge and skills are needed to navigate the Commons, both for individuals and institutions?
- Engagement with the commons acts as a ‘lightening rod’ to raise debate about a broad range of issues such as; ownership, control, power and identity. How do we deal with this in our private and institutional roles?

A. Establishing a legal basis for The Commons

The ALTO (Arts Learning and Teaching Online) project, in common with many other open education projects around the world is making use of the Creative Commons licensing system. This simple, but imaginative, application of existing copyright law has far reaching effects including, as we shall explain, the shaping of the proposed technical architecture of ONCE.

In the UK the current institutional IPR (Intellectual Property Rights) policy picture is confused and represents pre-digital and traditional working methods when publishing technology was well beyond the means of many institutions. Given this cultural background any discussion of licensing of learning resources in a university, especially for open distribution, is likely to be a sensitive issue. However, we have found the Creative Commons Licence system very well suited to our purposes.

The background to the development of the Creative Commons licence system is well described on their website [1]. In essence the founders were dissatisfied with the effects of the application of traditional restrictive copyright law being applied in the digital realm of the World Wide Web. The underlying vision was the realisation that the traditional economy of ‘pay per copy or view’ of older media such as books and films

that had created the basis of current copyright law was not going to suit everyone who publishes their work on the web. While big business has sought to take advantage of the web as a new distribution medium it has also tended to try and enforce the old model of restricted access and lobbied for tougher legislation to protect their IPR. At the other end of the economic spectrum individuals, public bodies and small companies, especially those in the creative industries, cannot easily afford to hire lawyers to handcraft licenses to stipulate how their IPR may be used.

The Creative Commons licence system was developed to meet the needs of people who wanted to publish their content to the World Wide Web openly and freely and yet retain some legal control and protection for their work. Whereas, traditional copyright reserves all rights to the owner, the Creative Commons licence system uses existing law to adopt a ‘some rights reserved’ approach to support open publishing to the web. It is also important to understand that the Creative Commons licence system is actually based on existing copyright and contract law. There was, and is, a social and political agenda attached to this – the broad aim is to lower the legal barriers to sharing and reuse. The Creative Commons licensing system was envisaged as a means to provide a shared common space on the web where people could publish their works under simple, easily understood licence terms, in a way that helps them take advantage of the networking properties of web technologies. This vision is simply summed up by this statement from the organisation’s website:

“Our vision is nothing less than realising the full potential of the Internet — universal access to research and education, full participation in culture — to drive a new era of development, growth, and productivity.” [2]

The use of Creative Commons licences for content and metadata has the considerable advantage of greatly lowering the transactional costs involved in sharing, reuse and adaptation. As we shall see, this has a considerable potential to simplify technical systems designed to support open educational and cultural practices.

B. The Open Education Movement – a brief overview

The best known UK example of open education in the UK is the Open University (OU), set up in the 1960’s to make university education more accessible and open to students without the formal qualifications normally required by traditional universities or the ability to attend full time education. In addition, the OU aimed to deliver its programmes of study using distance learning techniques to offer study opportunities that were flexible in terms of the place and time of study, so that students could fit learning around their personal and working lives. The OU was (and is) an example of ‘official’ and certificated education delivered via innovative means to overcome the barriers of time, place, work and family commitments and the lack of formal qualifications.

More recently, interest has grown in taking this model a step further by providing open access to the course study materials of universities on the web so that ‘independent learners’ can use them. The acknowledged pioneer in this area has been MIT in the USA who, with the help of large charitable grants, has made all its undergraduate course materials openly and freely available under the terms of a Creative Commons licence. A useful overview about OERs has been produced by JISC [3], which provides this definition:

“The term Open Educational Resources (OER) was first introduced at a conference hosted by UNESCO in 2000 and was promoted in the context of providing free access to educational resources on a global scale. There is no authoritatively accredited definition for the term OER at present; the most often used definition of OER is, ‘digitised materials offered freely and openly for educators, students and self-learners to use and reuse for teaching, learning and research’.”

With regard to this working definition, it is important to note that “resources” are not limited to content but can include software, and legal licences etc.

II. THEORETICAL AND METHODOLOGICAL PERSPECTIVES ON USING ICT

In the UK, activity and discourse in the area of sharing and reusing learning resources has tended to have been dominated by narrow technological concerns with interoperability standards, learning objects, metadata and the creation of specialist repository software – sometimes becoming an end in itself rather than linked to real human needs [4]. As a result, things have not worked out as expected; Fini [5] describes it this way:

“This way of interpreting e-learning is running into a crisis: the promised economic effectiveness of content re-use is often hard to demonstrate or it is limited to specific contexts, while a general feeling of discontent is arising.” [5]

To understand this apparent impasse Friesen [6] and Friesen & Cressman [7] helpfully point out there is a set of important political and economic sub-texts connected to the proposed uses of technical standards and technologies in education that still need to be explored. Neglecting such ‘soft’ issues is a major cause of the problems cited above by Fini [5]. While Harvey [8] notes a prevailing belief in neo-liberal thinking that there can be a technological fix for any problem and that products and solutions are often developed for problems that do not yet exist. In education, one of the materializations of this tendency was the proposition that interoperability standards and techniques developed in the military and aviation sectors could be adopted in the mainstream public education system [6]. But, despite the large amounts of money spent by public bodies in this area there has been little progress; Friesen [9] predicted (correctly) the high chance of failure of these activities. In retrospect it is not surprising that technologies and approaches originating in the military and industrial

sectors have not taken root in mainstream public education systems. Here, teaching and learning is inevitably a far more messy, less controlled and contingent enterprise. Influenced by these experiences, the ALTO project has sought to move away from such technical determinism in its approach.

Pioneering work about introducing technology into workplaces by Mumford [10] and others at the Tavistock Institute has shown that successful innovation has to address the contextual and social aspects of using new technologies. This applies especially to higher education organisational and teaching cultures, which can be notoriously resistant to change. This approach, which became known as the Socio-Technical model, has been adapted successfully for the introduction and integration of information technologies into modern knowledge-based workplaces, notably by Sharples [11] as ‘Socio-Cognitive Engineering’ and Wenger [12] & [13] as ‘Communities of Practice’ and ‘Technology Stewards’. These approaches draw on traditional ethnographical approaches, where project fieldworkers interact with the groups under study to understand better how they work and live. This information is then used in the iterative construction of prototypes and systems that are tested with people to understand how both the tools and human systems may be improved. One way of describing this approach is that it is investigative and human-centred, as well as being contextually and culturally sensitive.

A. Putting the Social into Socio-Technical System Development

To begin with, and in accordance with the existing technological hegemony in education in the UK, the ALTO project initially started out by committing to acquire and install a repository software package. Repository software is optimized for storage and management and operates using a library paradigm, but is not good at presenting or publishing information. These limitations rapidly became apparent in the context of ALTO and the Art and Design academic community, who traditionally place a high importance on ‘look and feel’ i.e. affective and usability issues.

We realized that while a repository might be a solution for meeting institutional storage requirements, it alone would not be enough for open education development. We came to understand that ALTO needed to be more than just one software tool – it would need to be a system of connected and related tools. A repository gave us a place to safely and reliably store resources in the long-term for which there was already a strong institutional need. We came to see that the ALTO repository needed to fit into and be a part of a wider and dynamic ‘ecosystem’ for creating open online resources and supporting their associated communities of practice. Two things became clear. First, was that resources in the repository would need to be easily ‘surfaced’, in a variety of social media to aid dissemination and impact. Second, that the other components of such a UAL ecosystem would want to use the repository to deposit

some of their outputs now that the a long term storage service was possible.

Fortunately, a communal social media platform was already available through an existing UAL initiative called Process.Arts: “an open online resource showing day-to-day arts practice of staff and students at UAL” [14]. This originated as a small personal research project to explore how to meet the need for staff and students to show and discuss aspects of their practice as artists and designers by providing a collaborative space using the Drupal web content management system, which includes many common Web 2.0 features. The ALTO project decided to support this initiative and it has since been very successful in a short time, with users uploading images and videos and discussing each other’s work. User numbers and interactions are high and growing with considerable interest from abroad.

Through this experience, we came to understand that if the repository were to be the officially branded ‘library’ part of ALTO then Process.Arts would provide the ‘open studio and workshop’ where knowledge and resources are created and shared. As a result, the project board took a decision to redesign the initial architecture of ALTO to add a ‘social layer’ to the initial repository, which incorporated Process.Arts. As at 2012, the evolving institutional infrastructure can be viewed at this web site <http://alto.arts.ac.uk/>. The digital library component can be viewed at this web site <http://alto.arts.ac.uk/filestore/>, and the social layer which provides an open collaborative studio/workshop space can be viewed at this web site <http://process.arts.ac.uk/>.

III. ONCE: A DRAFT SOCIO-TECHNICAL ARCHITECTURE

Simply put, the proposal is to make available a set of simple free and open source tools that build on the experience of the ALTO project. These tools could support both engagement with open education in the arts and also meet existing institutional needs to manage collections of cultural artefacts, a combination that would greatly encourage adoption. At an institutional level this would include a simple storage system (that would be comprised of a ‘digital gallery’ and a filestore management tool) together with a social media platform. At a national or regional level there could be a shared hosted service option for those institutions and departments that lack the ability to install and support such tools locally but still need the ability to project their distinct identity online. In addition, there could be an aggregation service that combined feeds from institutionally based content together with content from the shared system. This aggregation service would allow users to browse and search through the service by topic, subject, institution, community, person, region etc.

Fig. 1 provides a simple illustrated conceptual model that describes the main components and data flows. The term ‘open data’ is used here as an umbrella term to include both content and metadata. The metadata generated will be of a simple and lightweight nature. The

emphasis on simplicity of design is based on the experience gained by the author's involvement in the national UK learning resource repository project Jorum <http://www.jorum.ac.uk/> and other repository projects at the UAL and the University of the Highlands and Islands in Scotland. An essential feature of this model is that the social space and simple storage area is completely under the control of the users. As opposed to many institutional research repository systems, there is no mediated deposit

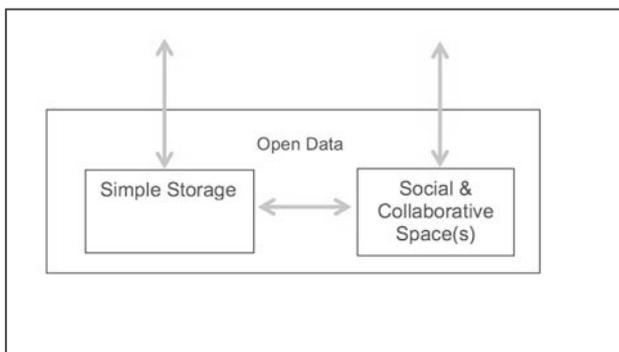


Figure 1. ONCE Conceptual Model with Data Flows

of content. Nor are there any specialist workflows or procedures or the use of complex metadata schema, or controlled vocabularies. The social space in the system makes use of a basic Dublin Core metadata [15] schema for each 'post' in the system, in addition each post is categorized according to a simple subject taxonomy (ceramics, textiles, drawing, etc.) plus there is the use of user-created tags. The simple storage area, is if anything even simpler and uses a simple Dublin Core schema and user created tags. This combination of simple metadata, categories and tags constitute what Lambe [16] describes as the *knowledge infrastructure tools* our system uses. This places our system towards the low cost, high serendipity, low precision end of Lambes's continuum of knowledge management tools.

As Lambe observes there is no single information management tool that satisfies effective knowledge management, although the tendency to assume that there is can be a common failing in 'traditional' academic repository based projects. This is where we think our emphasis on the use of Creative Commons licences and user-controlled spaces becomes very useful. Because the content and metadata of ONCE is licensed under Creative Commons, external specialists and experts are free to add any services or enhancements they choose. This allows the creators of cultural artefacts and open educational resources the freedom they need to manage their own content and interact with their communities without being encumbered by unworkable tools. Those who are interested in adding the specialist services of preservation, curation, metadata and interpretation to the content of ONCE can do so easily using their own separate systems. This opens the door to the use of linked data and semantic web technologies using the content of ONCE either as a 'target' or through harvesting the content for use elsewhere. This arrangement, in our opinion, involves a much needed

rebalancing of the power relations between such information specialists and system users.

Fig. 2 describes how a mix of institutionally hosted and shared systems might interact with a national aggregation service and, potentially, Europeana. We think the free and open source software Drupal could serve as the basis for both the social platform and, potentially, the national aggregation service.

A. Simple Storage - Gallery and Filestore Options

As already observed, specialist repository software in the UK has been the presumed solution to this need. This has resulted in a range of both commercial and free and open source repository software tools being made available, with varying degrees of success. The development of the free and open source repository tools (e.g. Eprints, Dspace and Fedora) has, to date, been dominated by a focus on the needs of the research and archive communities with a development paradigm that has concentrated on the information science aspects of resource management. As a result, the usability and widespread take up of these systems has been quite limited, especially when applied to the field of learning resources, while commercial solutions have made little progress due to their cost factor. The currently available free and open source repository tools, require a high degree of specialised technical expertise to deploy, maintain and customize, while suffering from poor usability. Another important problem with these tools is that the associated developer communities are very small, raising issues of sustainability, resilience and lock-in to specialist suppliers and developers. More fundamentally, these tools were not designed for learning resources which have a very different user community to those of research repositories. Combined, these factors make these systems unsuitable for large-scale adoption in mainstream public education.

There are other concerns about trying to use these kinds of repository platforms. It is important to remember that storage and retrieval (i.e. finding a resource) are not the same problem and using one solution for both results in a solution that may not be ideal. The modern internet places great emphasis on social networking, vibrant user interfaces, high performance, usability and growing support for usage via mobile devices. This raises the question - is the ideal solution to manage educational and cultural collections a software platform that was designed for a different purpose and at a different time, without features which modern users expect? A fresh and more modern approach is required, building on experiences learnt over time in the wider software industry to produce solutions which users are more likely to find acceptable:

- Simple metadata profiles that are kept to the bare minimum
- Tight integration with a social media space (the social layer)

- Solid software architectural design which separates out functionality clearly i.e. separate the storage layer from the user interface layer
- Rich user interface
- Low technical barrier of entry i.e. easy install, minimum effort required to maintain
- Support for modern browsing environments e.g.

simple and scalable manner by supplying components that meet different user needs, rather than by building monolithic systems dominated by specialist concerns of experts.

This paper has described how the perspective of ‘the commons’ has enabled the project to move away from narrow technical determinism to articulate more realistic solutions based on the real-life context of the UAL. The system described in this paper is currently being

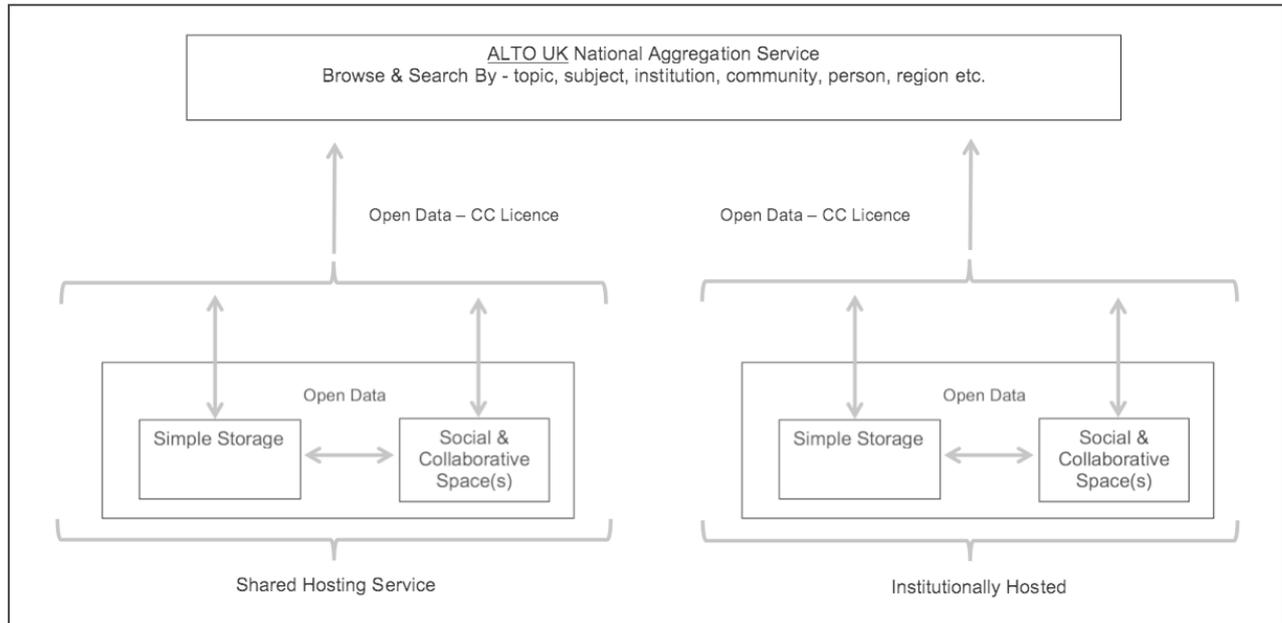


Figure 2. Showing how a mix of institutionally hosted and shared systems might interact with a national aggregation service

smart phones, tablets etc.

In addition, there already exists many well-used software components in the modern internet ecosystem, perhaps the solution is not a wholly new product but includes a fusion of these products e.g. YouTube, Drupal, FaceBook, Amazon Web Services & S3 etc.

IV. CONCLUSIONS

The ALTO project has found it's simply not enough to provide a repository mechanism of storage (important as that may be), this needs to be accompanied by a ‘social layer’ that enables the important human factors of communication, collaboration, and participation that are needed for sustainable resource creation and sharing within community networks. The technical solutions provided should help not hinder these activities; the guiding design principle for these socio-technical systems should be that of the concept of *conviviality* [17], [18] and *stewardship* [13]. This simply means that such systems are responsive to users real needs. Achieving this will involve changing the current power relations between the information specialists and software developers who currently control these systems and the users of such systems in academia. The proposed ONCE architecture aims to meet these requirements in a

implemented at the UAL. It is becoming clear that it also has the capability to host and support the significant cultural collections (together with their communities of practice) for which the university is responsible. Previously, such collections have had to be hosted and supported externally, resulting in lost opportunities, unsustainable costs and the danger of losing some of the intellectual capital of the university over time. The combination of the ability to manage and share collections of learning resources and cultural artefacts effectively and economically should be an attractive proposition for institutions in the developed and developing world.

Looking forwards, the project would like to continue this development work in collaboration with external partners. One of the aims of such work would be to create a ‘living laboratory’ [19] of users to support future joint research and development opportunities in the area. Areas for further investigation and development would include:

- Prototype of a simple storage solution including the gallery and filestore tools taking the existing ALTO Filestore system (<http://alto.arts.ac.uk/filestore/>), as a starting point

- Prototype of a social media platform that can be installed and customized at an institutional level
- Prototype of a shared service to provide a hosted simple storage and social media platform that can operate at regional and national levels
- Prototype of an aggregation service
- Prototype of a system to explore different approaches to the trustworthy accreditation of independent learning
- Development and release of open source software tools for identified communities of users
- Development of a knowledge organisation and management model along the lines advocated by Lambe [16] to guide and underpin system development as opposed to the current emphasis on information management.
- Explore and articulate the philosophical and economic aspects of open educational and cultural practice in relation to existing power structures in education and technology. Use the concept of the accreditation of independent learning using OERs and open cultural artefacts to drive this debate
- Continue to make use of the socio-technical school to guide system development, particularly the ‘socio-cognitive engineering model. In addition, use the insights of the industrial and product design disciplines as exemplified by the work of Norman [20] and Castiglioni [21].

REFERENCES

- [1] <http://creativecommons.org>
- [2] <http://creativecommons.org/about>
- [3] JISC (2008), Open Educational Resources – Opportunities and Challenges for Higher Education, JISC CETIS. Available at http://wiki.cetis.ac.uk/images/0/0b/OER_Briefing_Paper.pdf
- [4] P. Barker, An Open and Closed Case for Educational Resources, 2010 <http://blogs.cetis.ac.uk/philb/2010/09/10/open-closed-case/> accessed March 6 2011
- [5] A. Fini, “Focus on e-learning 2.0”, in Journal of e-Learning and Knowledge Society: The Italian e-Learning Association Journal, Vol. 3, n.2, 2007 http://je-lks.maieutiche.economia.unitn.it/index.php/Je-LKS_EN/issue/view/33 accessed August 16 2011
- [6] N. Friesen, N. (). The E-Learning Standardization Landscape. Canada: Cancore, http://cancore.athabasca.ca/docs/intro_e-learning_standardization.html, 2004, accessed August 16 2011
- [7] N. Friesen, & D. Cressman, The political economy of technical e-learning standards in Learning Objects: Theory, Praxis, Issues & Trends, A Koolhang, & K. Harman, Eds Warsaw: Informing Science Press. pp. 507-526 2007.
- [8] D. Harvey, A short history of neoliberalism Oxford, University Press, 2007.
- [9] N. Friesen, “Three objections to learning objects and e-learning standards” in Online Education Using Learning Objects, R. McGreal, Ed., London: Routledge. pp. 59-70, 2004.
- [10] E. Mumford, Effective Systems Design and Requirements Analysis: The ETHICS Approach. Basingstoke: Macmillan, 1995.
- [11] M. Sharples, N. Jeffery, J.B.H. du Boulay, J.B.H., D. Teather, B. Teather, and G.H. du Boulay, “Socio-cognitive engineering: a methodology for the design of human-centred technology”, European Journal of Operational Research, 136, 2, pp. 310-323 2002. Preprint available as 104Kb pdf file at http://www.eee.bham.ac.uk/sharplen/Papers/mediate_ejor.pdf
- [12] E. Wenger, Communities of Practice. Cambridge: Cambridge University Press, 1998.
- [13] E. Wenger, N. White, J.D. Smith, Digital Habitats: stewarding technology for communities Portland: CPsquare, 2009.
- [14] C. Follows, About Process.Arts <http://process.arts.ac.uk/content/about-processarts> accessed August 16 2011
- [15] <http://dublincore.org/about-us/>
- [16] P. Lambe, Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness. Oxford: Chandos, 2007.
- [17] I. Illich, Tools for Conviviality, London: Harper & Row, 1973.
- [18] M. Hardt, A. Negri., () Commonwealth, Cambridge, Massachusetts: Harvard University Press, 2009.
- [19] http://en.wikipedia.org/wiki/Living_lab
- [20] D. Norman, The Design of Everyday Things. Cambridge, Massachusetts: MIT Press, 1999.
- [21] P. Antonelli, Curator’s Essay. In Achille Castiglioni: Design! (A retrospective of the Italian architect and industrial designer), The Museum of Modern Art, New York. <http://www.moma.org/exhibitions/1997/castiglioni/index.html>, 1997

Knowledge Management in Performing Arts Education

Peter G.F. Eversmann

Department of Theatre Studies, University of Amsterdam, UvA, Amsterdam, The Netherlands, p.eversmann@uva.nl

Abstract—In this paper two models for knowledge creation and management are introduced: that of Nonaka and Takeuchi and that of Csikszentmihalyi. Application in the domain of performing arts education shows that they complement each other: the former focuses more on organizational learning while the latter is better able to describe the dynamical relationships between individual (student), field (teachers) and domain (performing arts studies). A third model by Koehler and Mishra addresses the specific problem of integrating information technology in education. Together the models provide a conceptual framework for assessing ECLAP's educational potential.

Keywords: *Education; Performing Arts; Knowledge Management; Systems Model of Creativity; TPACK*

I. INTRODUCTION

ECLAP positions itself as an e-library for the performing arts. As such it will be indispensable to researchers in the field and to tertiary education – whether this is geared towards a practical or a theoretical approach. This library function and its relation to research and education poses the question of knowledge management in the performing arts: what do we preserve, how is it made accessible and how is knowledge generated and improved?

II. NONAKA AND TAKEUCHI: MODEL OF KNOWLEDGE CREATION

When one talks about knowledge management the famous model of Nonaka and Takeuchi (1995) immediately springs to mind. The basis of this model is the dynamic transformation of tacit knowledge into more explicit forms and then back again – forming knowledge spirals for individual, group and organizational innovation and learning. The spiral loops continuously through a four step process. Knowledge formation starts with Socialization in which tacit knowledge is shared through social interactions. In this step one gathers knowledge through face-to-face communication and/or through sharing experiences directly with each other. The tacit knowledge that is referred to here is knowledge that lies in people's brains, is highly personal and is hard to formalize. As Nonaka, Toyama and Konno (2000:7) write: "Subjective insights, intuitions and hunches fall into this category of knowledge. Tacit knowledge is deeply rooted in action, procedures, routines, commitment, ideals, values and emotions." Typically the process of Socialization is characterized by the roles of tutor and apprentice where one learns through live communication (rather than through textbooks or manuals) and arrives at a mutual understanding through

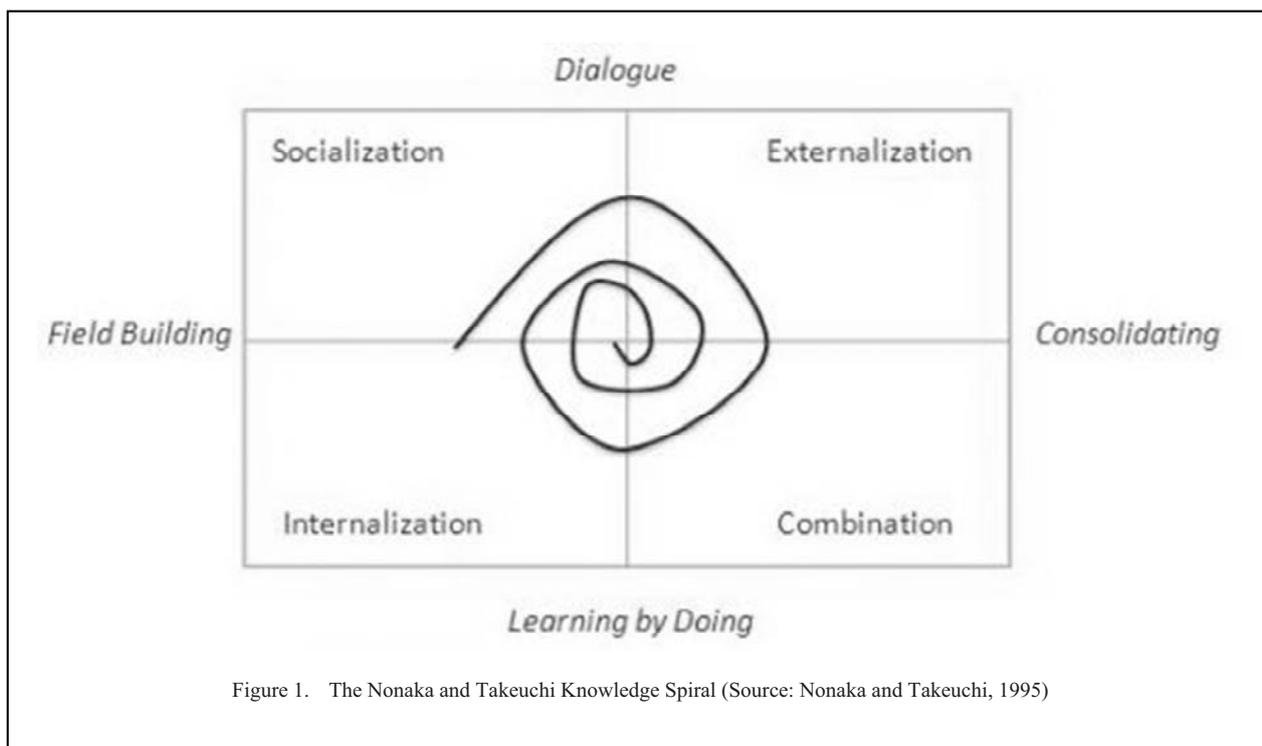


Figure 1. The Nonaka and Takeuchi Knowledge Spiral (Source: Nonaka and Takeuchi, 1995)

the sharing of mental models. In the second step - Externalization- the tacit knowledge is converted into explicit knowledge. Knowledge is articulated into various shapes, such as metaphors, analogies, concepts, hypotheses, or models. The next step is called Combination: discrete pieces of articulated knowledge are recombined and synthesized into a new form so that it can be disseminated among the members of an organization. The outcome of this process is systematized and packaged explicit knowledge such as documents, manuals, databases, patents and so on. In the fourth step -Internalization- the acquired and consolidated knowledge is embodied into tacit knowledge and shared throughout an organization. This process of internalization is strongly related to 'learning by doing': through action and practice one acquires the knowledge that is inherent to the field one is engaged in. Once internalized the new knowledge can then be used, broadened, extended and reframed within one's own tacit knowledge. Thus it can set off a new spiral of knowledge creation when it is shared with others through socialization.

Although Nonaka and Takeuchi developed this model within management studies for commercial organizations, corporations and companies and not in the context of education it is not too hard to apply the model to our field: tertiary education within the performing arts. Let me give a few examples of how the model might illuminate what goes on in this domain.

A teacher at a university department for performance studies can have a personal idea about a new form of theatre that is emerging and that connects various groups and artists. As yet he cannot precisely define the characteristics of this 'new' form of theatre, but together with a group of students and colleagues he views a number of such performances, analyzing them, comparing them with each other and trying to define commonalities (socialization). From this process slowly emerges a number of characteristics and concepts that bind the different performances together. He can now give a name to the 'newness' of these performances and define more precisely what constitutes the genre in question. In a course on the subject he presents his findings, discusses them with students, analyzes more examples of performances and refines his concepts. (Externalization) Our teacher then decides to formalize his knowledge and he publishes a textbook on the new theatre form. In order to do this he has to restructure his somewhat informal lectures and combine several of the analyses into a coherent, didactic text complete with footnotes, indexes, bibliography and so on. (Combination). Finally, through disseminating the textbook among students and colleagues, the new genre becomes 'internalized' within the field: one recognizes the form and can use the analytical concepts that go with it (Internalization). In a further process of socialization the new genre can then give rise to further enhancements, extensions and critical reflections.

A piano student arriving at the academy of music will be trained by a teacher to master her instrument. Apart

from studying by herself she has regular face-to-face sessions in which the teacher will give hands on instruction, she will also learn a lot more from him: how to deal with and analyze certain musical scores, how to listen to and talk about other musicians, how to prepare for concerts, etc. (Socialization). Although the phase of converting this tacit knowledge into explicit knowledge is maybe not so clearly defined, one could nevertheless say that the student will develop the ability to conduct herself as a professional musician. Her self-awareness and her confidence in expressing opinions related to music will grow, as will her capability to assess her own place in the musical landscape (Externalization). Also the process of consolidation is not very clear with this example, but one could say that the demands of the school (as far as they are explicitly stated in the educational goals) in conjunction with the results of various exams and finally the diploma she receives at the end of her studies formally confirm her progress as a mature musician who is sufficiently versed in all aspects of the profession (Combination). And as she then proceeds with her career the acquired skills and attitudes will become more internalized and will be shared with other musicians (Internalization).

III. APPLICATION PROBLEMS

As becomes clear from these examples the Nonaka and Takeuchi model is quite capable of describing processes of education and research in the performing arts. However, there are also some problems:

On a personal level application of the model seems not always successful. In the second example one experiences some difficulties with formulating the phases of externalization and consolidation. This is less problematic in the first example because there the focus lies more on the field of theatre studies and less on the individual teacher. One should realize that this comes about because basically the Nonaka and Takeuchi management model was designed for analyzing larger entities: institutions, companies, corporations or even fields. So, when one would look at how the school of music in the latter example generates knowledge as an organization it would be easier to apply the model and describe the development of educational knowledge in this institution in terms of a series of conversions between tacit and explicit knowledge.

A second problem lies in the ontological status of knowledge within an organization. The model doesn't address the question what actually happens with the generated knowledge other than presuming that it can become the starting point of a new cycle. But when one looks at the sort of organizations that the model was designed for and at the applications in which it is used one realizes that one is in the realm of larger commercial corporations and industries. The knowledge that is generated in these organizations is not enduring or everlasting but dynamic and changing all the time. Even when the knowledge is explicitly consolidated and solidified in documents, reports, specifications, manuals, databases, patents, licenses and so on, after some time it

will become obsolete and the company will discard it as no longer being a valuable asset for its daily functioning. At most this knowledge will be kept in an archive or library, but it will have progressively fewer consequences for the actual bearings of the people in the organization and after a while it will only be of historical interest, functioning as a status symbol for the age and respectability of the company. As such the model complies with how knowledge is treated in large sections of universities – especially scientific and medical knowledge. Also in these areas knowledge is discarded when it has become obsolete - for example through the formulation of a new paradigm – and becomes only interesting for those who study history of science or history of medicine. Today no student of physics in his right mind will study Newton's work in earnest – as something bearing upon the contemporary discourse in the field. Although his work is of course recognized as most important for the development of the discipline, the impact upon present day problems and theories is nil. Also in medicine young doctors and nurses will not learn from older books on for example anatomy, but will rather have new insights and visualizations in order to get an idea of how the body is structured and functions. But this is quite different in the disciplines belonging to the humanities: here one does not throw away older theories and knowledge. The criterion for 'truth' (or even for subjective consensus) is not very clear cut and there is no methodological procedure to absolutely value one theory over the other. Although one can discern 'paradigms' these are more like academic fashions than that the most recent one makes earlier ones superfluous. So, within theatre studies Aristotle's theories can be just as 'actual' as those of Schiller, Brecht or Brook.

Finally a third problem lies in the assessment of the value of knowledge within a corporation. As such the model doesn't address the question of 'better' or 'worse' knowledge. It assumes that in the recurring

process of conversion from tacit to explicit and solidified knowledge somehow the relevant and valuable knowledge for an organization will prevail. But how less relevant and less valuable knowledge is filtered out (or even: how faulty knowledge is corrected) and who is responsible for quality control with regard to the generated knowledge is not clear. One might assume that in the end knowledge that is leading to more efficient behavior of an organization (read: leads to more profit) will manifest itself and that management will have the capabilities of deciding what knowledge is likely to be the most valuable, but experience tells us that this is not always the case.

IV. CSIKSZENTMIHALYI: SYSTEMS MODEL OF CREATIVITY

In view of these problems I would like to introduce another model that might provide a valuable addition in describing what happens in academic fields such as the humanities. In the systems model of creativity developed by Mihaly Csikszentmihalyi it is proposed that creativity results from the dynamic interaction of three elements: "a culture that contains symbolic rules, a person who brings novelty into the domain, and a field of experts who recognize and validate the innovation" (Csikszentmihalyi 1997: 6). It is important to realize that each component in the system is integral to it with one being no more important or less necessary than the other. In short 'each of the three main systems – person, field and domain – affects the others and is affected by them in turn' (Csikszentmihalyi 1988: 329). Each component is necessary for creativity -or new knowledge- to occur but not sufficient, in and of itself, to produce novelty. A closer look at the components of the system and their interactions can reveal how it works in practice and how it might be applied to the domain of performing arts studies. Starting with an individual person we note how he/she is determined by environmental and biographical

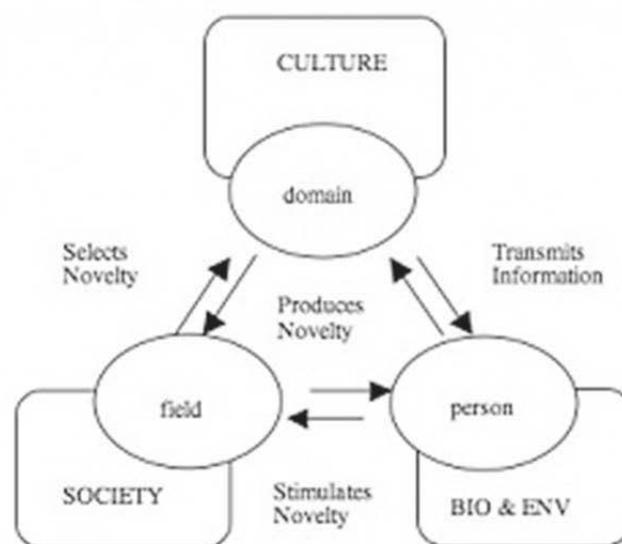


Figure 2. The systems model of creativity (source: Csikszentmihalyi, 1999)

circumstances. In order to function creatively within the domain of the performing arts it is necessary that one has access to this domain because one should become acquainted with the 'symbolic rules' that govern the domain. In other words: before one can play a musical score one has to master reading notes, before one can analyze a performance one has to become familiar with theatrical means, with theoretical notions, with a theatrical environment and with some of theatre's history, before one can dance one has to master certain steps and movement that form the 'language' of a choreography, and so on. Becoming acquainted with a domain and its rules and practices usually comes about by learning from teachers and/or by accessing stored information in writings, recordings, objects and so on. It is thereby important to realize that a domain can contain knowledge outside human agents but is at any time of its existence dependent upon and shaped by a conglomeration of people. This is what Csikszentmihalyi calls the field: the expert 'gatekeepers' who decide what elements belong to a certain domain and how important these elements are within that domain. So the field of performing arts may consist of artists, critics, teachers, theatre managers, sponsors, archivists, government officials and so on. Together they 'decide' what can be seen in the domain and what is important to keep for posterity. A person that has a new idea -say: a composer that has written a new musical score or a student in theatre studies that has 'discovered' a new genre- has to be accepted by the gatekeepers in order to become incorporated in the domain. The new musical score has to find musicians that want to play it, a concert hall that is willing to put it on the program, a publisher who is willing to print it and a librarian that puts it in a library.

Similarly the paper with the description of the new genre has to be accepted by the university professor, find a greater audience at a conference, a publisher that prints the manuscript and disseminates it further and university libraries that are buying it. To summarize in the words of Csikszentmihalyi: "For creativity to occur, a set of rules and practices must be transmitted from the domain to the individual. The individual must then produce a novel variation in the content of the domain. The variation then must be selected by the field for inclusion in the domain" (Csikszentmihalyi 1999: 315).

In a way then this model is giving insight in the tasks and responsibilities of educators within the field of the performing arts. On the one hand they introduce the students to the domain: outlining its rules and practices, pointing to important examples and, not to forget, stimulating the students to make their own innovative contributions. On the other hand they are important gatekeepers that are responsible -together with other experts- for shaping the domain and, in the case of libraries and archives, of deciding what events, objects and documents are worth storing in the memory banks of the domain and (maybe even more of a responsibility) what can be discarded.

V. KOEHLER AND MISHRA: TECHNOLOGICAL PEDAGOGICAL AND CONTENT KNOWLEDGE (TPACK)

The flexibility of Csikszentmihalyi's system model of creativity comes to the fore when we realize that also education in the performing arts can be seen as a domain with its concomitant field (teachers, heads of schools, administrators, etc.) and with persons that are creatively

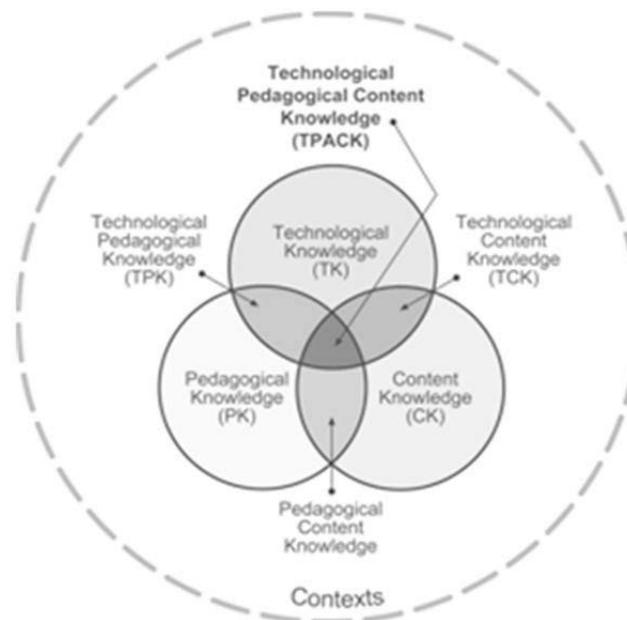


Figure 3. The TPACK framework and its knowledge components (source: Koehler & Mishra, 2009)

trying to bring about innovations. One of these innovations is, of course, the introduction of information technology and e-learning in the domain. At this point it is therefore advantageous to introduce yet a third model: TPACK.

TPACK stands for Technological Pedagogical and Content Knowledge and was introduced by Koehler and Mishra (2008) who describe the model as a framework for teacher knowledge for integrating technology in the classroom. The idea builds on Shulman's concept of PCK or pedagogic content knowledge (Shulman 1987) in which he distinguishes between knowledge about a subject ('content knowledge') and about teaching ('pedagogic knowledge'). A teacher may have both, but still can experience problems in teaching a subject effectively. The ideal is true integration of the two forms of knowledge or 'pedagogic content knowledge'; knowing how to teach in the specific context of a subject or discipline for a specific group of students. Koehler and Mishra extend this model with a third circle representing technological knowledge. The advantage of this addition is described by Martin Oliver as:

"(...) the proliferation of overlaps: it allows distinctions to be drawn between knowledge about technology, knowledge about technology used in teaching (technological pedagogical knowledge), knowledge about the technologies used within particular disciplines (technological content knowledge) and the central overlap, knowledge about how to teach a particular subject using appropriate technologies ... Analytically, this allows practical distinctions to be drawn between (say) a teacher's understanding of a concordance database ..., their understanding of how virtual learning environments work, and whether either of these would be useful in teaching their classes." (Oliver 2011, 91).

It is important to realize that this model was constructed in order to help integrating technology in classrooms and that it functions within a broader context of ideas on teaching and on the role of teachers. This perspective on teaching and learning is therefore particularly relevant for understanding the position of an e-library like ECLAP and its services in relation to performing arts education.

Koehler and Mishra argue that "teachers practice in a highly complex, dynamic environment that asks them to integrate knowledge of student thinking and learning, knowledge of the subject matter, and increasingly, knowledge of technology." (Koehler & Mishra 2008: 3-4)

The integration of the newer, digital technologies in the teaching process is complicated by the fact that these technologies are characterized by affordances and constraints that are protean, instable and opaque in nature. The protean aspect refers to the fact that computers are multi-tools; capable of manipulating visual, acoustic, textual and numerical sign systems. In other words: the digital revolution is that of a meta-medium resulting in the fact that computers find a

multitude of applications in almost every field of human endeavour. In the words of Koehler and Mishra: "...computers) can be a tool for communication (...), a tool for design and construction (through software for scientific modelling or software for designing websites, themselves very different activities), a tool for inquiry (such as through digital libraries and digital probes), and a tool for artistic expression through image, movie, and audio design software programs)." (Koehler & Mishra 2008: 7) And of course this certainly applies also to the multimedial world of performing arts education where examples of all of these uses -and others! - come readily to mind. The opacity of digital technologies (i.e.: the fact that the inner workings of computers and programmes are hidden) complicates things even further and makes learning a quite difficult process. Besides, most software tools are designed for business and work so that before they can be used in the classroom they often have to be adapted for educational purposes. A third characteristic of digital technologies is the fact that they are instable, expanding and changing constantly, so that users have to keep up continuously with ever new hardware and software (often not very compatible with former versions). A consequence of these rapid developments is also that the 'new' tools are often unreliable, not very well tested and full of bugs. As a result the "instability of digital technologies requires that "teachers become life-long learners who are willing to contend with ambiguity, frustration and change." (Koehler & Mishra 2008: 8) And all this in a context where (university-) teachers have increasingly busy time schedules and where a lot of them -still- grew up at a time when the digital revolution was in its early infancy. Finally, when enumerating the difficulties of integrating digital technologies in the learning process, one should take into account:

- that there are important social and psychological differences between the worlds of technology and pedagogy that may hamper fruitful cooperation between the two.
- that classroom contexts are varied and diverse and that hence " integration efforts should always be custom-designed for particular subject matter ideas in specific classroom contexts." (Koehler & Mishra 2008: 10)
- that there exist 'divides' between digital natives and digital immigrants as well as the digital divide between those who have access to the latest technology and those who do not.

The complexity of integrating technology in the educational process of teaching and learning, which in itself is variable and complex looking at the two former models, is then the main reason for Koehler and Mishra to develop the analytical model that enables them to confront that complexity. This TPACK model is, as we have seen, characterized by the three main knowledge components (content, pedagogy and technology), the interactions between them and the various contexts in which these interactions play themselves out. In the end it is argued that what is needed for good teaching is

acquiring technological pedagogical content knowledge – a true integration of all three knowledges and not just understanding these concepts individually.

In order to reach a satisfactory equilibrium among the three components one should realize that the key figure is the teacher. In the end he or she is the curriculum designer par excellence and “teachers constantly negotiate a balance between technology, pedagogy and content in ways that are appropriate to the specific parameters of an ever-changing educational context” (Koehler & Mishra 2008: 21) This perspective on teaching implies the following consequences:

- Merely having teaching skills (technological or otherwise) is not enough. One should also have an understanding of what to do with these skills and how they relate to the whole TPACK model
- There is a need for greater emphasis on the demands of subject matter. One should not apply technological tools uniformly but instead adapt to the various demands of the curricular subject matter and/or the pedagogical approach that is deemed suitable.
- Practice (in curriculum design and teaching) is an important route to learning. Simulation, case studies or problem based learning scenarios are important strategies to come to ‘acceptable’ solutions of technology integration before applying these solutions in real classrooms
- Context matters. How TPACK is implemented in specific educational situations cannot be generalized. Efficient implementation needs a thorough understanding of the context and how it relates to the different components of the model.

VI. CONCLUSION

All in all one should conclude that the three models we have explored together can provide a conceptual basis with which to identify and understand the

processes that are at work within performing arts education. As such they should provide the framework for assessing ECLAP’s educational potential.

BIBLIOGRAPHY

- [1] Nonaka, I., Takeuchi, H (1995), *The Knowledge Creating Company*, Oxford University Press, New York, NY.
- [2] Nonaka, I., Toyama, R. and Konno, N. (2000). ‘SECI, Ba, and leadership: a unified model of dynamic knowledge creation’. *Long Range Planning*, Volume 33, Number 1, 1 February 2000 , pp. 5-34.
- [3] Csikszentmihalyi, M. (1997), *Creativity: flow and the psychology of discovery and invention*. New York. Harper Collins.
- [4] Csikszentmihalyi, M. (1988) Society, culture, person: A systems view of creativity. In R.J. Sternberg (ed.) *The Nature of Creativity*. Cambridge: Cambridge University Press. pp. 325–339.
- [5] Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R.J. Sternberg (ed.) *Handbook of Creativity*. Cambridge. Cambridge University Press. pp. 313-338.
- [6] Koehler, M. J., & Mishra, P. (2008). Introducing TPACK. In American Association of Colleges for Teacher Education. Committee on Technology and Innovation. (ed.), *Handbook of technological pedagogical content knowledge (TPCK) for educators* New York: Routledge. pp. 3-30.
- [7] Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57 (1), pp. 1-22.
- [8] Oliver, M. (2011). Handbook of technological pedagogical content knowledge (TPCK) for educators. *Learning, Media and Technology*, 36(1), 91-93. Retrieved 21/12, 2011 from <http://www.tandfonline.com/doi/abs/10.1080/17439884.2011.549829#tabModule>
- [9] DePlatchett, N. (2008). Placing the magic in the classroom: TPACK in arts education. In American Association of Colleges for Teacher Education. Committee on Technology and Innovation. (ed.), *Handbook of technological pedagogical content knowledge (TPCK) for educators*. New York: Routledge. pp. 167-192.
- [10] Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9 (1), Retrieved 27/12, 2011, from <http://www.citejournal.org/vol9/iss1/general/article1.cfm>



Workshop on New business models in the cultural heritage domain

Chair

Johan Oomen, Netherlands Institute for Sound and Vision (The Netherlands)

Workshop Committee:

Geertje Jacobs, Rijksmuseum Amsterdam, NL

Anra Kennedy, Culture 24, UK

Kara Van Malsen, NYU. USA

Model, format and services for audiovisual rights management

Laurent Boch⁽¹⁾, Annarita Di Carlo⁽²⁾, Francesco Gallo⁽³⁾.

(1)RAI – Radiotelevisione Italiana, Centro Ricerche e Innovazione Tecnologica, Torino, Italy, l.boch@rai.it

(2)RAI – Radiotelevisione Italiana, Teche, Roma, Italy, a.di_carlo@rai.it

(3) EURIX, Digital Media Technologies, Torino, Italy, gallo@eurix.it

Abstract— A substantial bottleneck to the exploitation of AV content in the professional domain, in particular for AV archive organizations, is the ineffective management of rights information. In this paper we present the recent achievements of PrestoPRIME, a EU project focused on the preservation of digital AV content and metadata. We analysed a sample of narrative media contracts, which resulted in the definition of a glossary of rights terms and in the creation of an ontology model and a new MPEG standard format for representing contract instances. In order to support the whole life cycle of AV rights, we implemented a set of services for creating, modifying, storing, and accessing rights information associated with archival content.

Keywords: audiovisual rights, ontology model, rights management services, audiovisual content exploitation

I. INTRODUCTION

The ineffective management of rights information results in an hold-up for the exploitation of audiovisual (AV) content preserved by the archive organisations. Several initiatives have been carried out in the past, which resulted in the definition of rights expression languages (RELs), often as part of wider standards, for representing contracts and licenses. An exhaustive overview of projects and initiatives related to rights representation and management is beyond the scope of the present paper. We just mention here MPEG-21 REL and the Open Digital Rights Language (ODRL), the latter used as a basis for the Open Mobile Alliance (OMA) REL. Recently the AXMEDIS project produced relevant solutions including tools and documents related to the support and automation of digital content management along the whole value chain.

The European Project PrestoPRIME [1] included rights modelling as an essential task in its research programme. The main purpose of this task was to try to define an AV rights ontology which could be used on a European level. In order to overcome the limited expressivity of REL licenses with XML representation of contracts, an ontologies based solution was sought. An overview of projects, initiatives, and technologies related to rights, including available systems for digital rights management, was carried out in PrestoPRIME and can be found in part B of [3] and the references therein.

In this paper we present some of the project outcomes, including a glossary of rights terms [2], an audiovisual rights ontology and the implementation of various services for the management of rights. The

PrestoPRIME AV Rights Ontology (PPAVRO) can be used for developing a flexible rights model, consistent with the general standard model, and a set of services that can actually support rights management.

The PrestoPRIME solution is based on the MPEG-21 standard (ISO/IEC 21000), resulting in new parts recently submitted for standardisation. In fact, by adopting the MPEG-21 Media Value Chain Ontology as the basis of the PrestoPRIME rights model, standardisation activities in this domain was fostered, notably in MPEG for the MPEG-21 Contract Expression Language. The coverage of the media contracts scenario is of particular importance in our case.

The document is structured as follows: in Section II we provide an overview of the current situation concerning rights management in a typical archive or broadcast environment, based on RAI's experience; in Section III we discuss the creation of a glossary of rights based on natural language contracts; in Section IV we describe the PrestoPRIME rights ontology and its relationship to MPEG-21; finally in Sections V and VI we discuss how services for rights management have been built on top of the ontology model.

II. THE CURRENT ORGANISATIONAL APPROACH AND ITS LIMITATION: THE CASE OF RAI

The issue of rights management became important for RAI (the Italian public broadcasting company) in the early 90's. Before that time RAI was rarely limited in its use of AV products as there were very few exceptions to the case of complete ownership of all rights. This was due to the fact that historically broadcast television was the only transmission medium, with the public broadcaster as the main interlocutor for artists and authors in the country. Therefore most production activities were carried out in house. Typical rights issues were represented by cases where another production company was involved, such as for movies, filming of theatrical performances and international co-productions. The amount of work in the area of rights management, completely paper-based, was easily under control.

This scenario changed radically after a few years, for a combination of reasons. The number of competitors acting in the same territory increased considerably together with the development of new delivery technologies (satellite, cable, internet) and the rapid growth of the number of transmission channels. This, on

the one hand, fostered the use of archive material, as the cost of good quality production remained high despite technical progress in this field. On the other hand, it pushed the rights holders to restrict the exploitation rights that they granted more and more, in order to maximise their revenues. The text of contracts became more and more complex and included a change in terminology with expressions such as “free TV”, “pay TV”, “satellite rights”, “VOD rights”, “internet rights”, and so on. As a consequence, rights clearance, namely the verification of the possibility to exploit an audiovisual asset in a target context, became more and more difficult. The teams assigned to this task became larger and had to include several law experts.

In 1996, in order to address this issue, RAI decided to set up a database of rights, aimed at supporting the exploitation of the AV archive. In 1997, the archive had the benefit of a new “multimedia catalogue”, a system providing intranet access to browsing quality copies of AV material, with metadata, annotations and indexing. The text of contracts formed the basis of that work. Contracts were collected and analysed in order to extract the rights information to be saved in the database. A specific archive department in charge of rights clearance was established in the organisation. Nowadays, that team is made up of 15 people and supports over 50,000 requests per year (not taking into account the trivial ones).

The basic problem with this approach is that the model on which the database was built cannot provide the flexibility required by the evolution of the real rights contract domain and the European legal framework over time. Anytime a new rights context appears, it is possible that the current system is unable to provide an absolutely reliable clearance answer. The consequence is that the members of the team have to manually go back to the referenced contracts and make another round of legal analysis, without the ability of consolidating the results in the IT system, for repeated queries.

It is also difficult to overcome the obsolescence of the current software tools running on a database because the tools themselves implement an inflexible rights model in their structure. Therefore any modification to the model, and also to the user interfaces (concerning the way the information is presented and how the user has to interact with the system for executing tasks), requires a considerable but not definitive re-engineering effort.

The work described in this paper tries to address the issues described above through the definition of a rights model and services, which are general and flexible enough to reduce maintenance costs.

III. GLOSSARY OF RIGHTS TERMS USED IN CONTRACTS

A. *Contract terminology in the trade of rights*

Any agreement on AV rights is the result of a negotiation which is bounded, in fact, by the purpose of

the purchaser, by the rights actually owned by the seller, and by the legal framework. The terminology of the narrative contracts (i.e. written in natural language) is left to the freedom of the parties involved. It is possible to observe the presence of technology related terms, used to distinguish among permitted exploitation; however the same term may not have the same meaning for all operators and often the term’s meaning is different than that used in a pure technical environment.

The glossary of rights terms [2] is the result of the analysis of a set of narrative contracts [3], most of them signed by RAI and involving international operators. The goal was to make available a broad list of commonly used terms and their definitions as a source of shared terminology for the negotiation of rights. The criteria for inclusion is consistent with the current evolution of the rights’ market and the legal framework of the European Union on copyright. In order to allow a more effective use of the glossary in various European countries, the original version of the glossary (in English) was also translated to French, German, and Italian.

B. *Exploitation Rights*

Many relevant glossary terms refer to exploitation rights. By analyzing the relationships among terms, typical patterns were identified that define an exploitation right as a combination of an action related to the particular exploitation of the intellectual property and a set of conditions and constraints used to restrict the allowed exploitation context.

The exploitation rights of the intellectual property (IPR) are those that arise from the legal framework: fixation, transformation, duplication, distribution, public performance (to perform or show or play the work in public), and communication to the public (to make available the work to the public through a communication mean).

IV. PRESTOPRIME AUDIOVISUAL RIGHTS ONTOLOGY

While the glossary presented above might be considered a useful reference for establishing the agreed meaning of the terms to be used within a narrative contract, it does not represent a model of the AV rights domain. The aims of the PrestoPRIME Audio Visual rights ontology (PPAVRO) [6] are to provide such a model and to permit the unambiguous representation of the rights situation of any audiovisual editorial entity.

A. *Basis of the rights ontology*

PPAVRO is conceived as an extension of the MPEG-21 Media Value Chain Ontology (MVCO)[4].

In particular, the general mechanism shown in Figure 1. , is used, in which a licensee is granted by a licensor the right to carry out a given action on an Intellectual Property entity through the Permission entity, which can be restricted by a number of Facts required to be true (or false in case of negation).

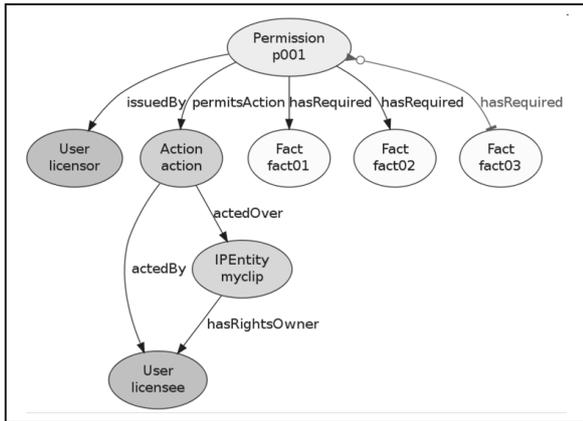


Figure 1. MVCO rights representation with permission

B. Main characteristics

With respect to MVCO, the major innovations of PPAVRO are:

- the definition of a hierarchy of Actions reflecting the intellectual property exploitation rights;
- the definition of a hierarchy of Facts reflecting the possible types of conditions and constraints which apply to a Permission;
- a semantic structure allowing the expression of complex logical constructs of Facts, in terms of unions and intersections.

The hierarchy of Facts is derived from the analysis of the glossary of rights, which in turn is based on the study of real narrative contracts. The resulting model provides the possibility to specify restrictions on the way content is delivered to the final user, including a modality of access, time of availability, devices, and others. The common set of exploitation constraints on territory, license periods, languages, number of runs, with the related and necessary data properties, are clearly supported, as well as the representation of exclusivity and sub-licensing.

C. Rights representation instances

The rights management environment that we address is related to the archive domain where the audiovisual material composing an intellectual property entity is submitted for preservation and future access and reuse. For each “archival information package”, according to terminology of the Open Archival Information System (OAIS) model [5], the rights situation of the related intellectual property entity must be considered. The instances of rights are written in the same language as the model, the web ontology language (OWL). As the PPAVRO is defined to “import” MVCO, so each instance document is defined to “import” PPAVRO, making the whole model logically available for processing the instances. The specific rights situations are clearly defined by means of individuals belonging to

the classes defined in the ontology, together with the related object properties (relations between individuals of two classes) and data properties [10].

D. Current status: becoming a part of a standard contract expression language

The PPAVRO model has been used as a reference for the contributions to MPEG (see e.g. [7]). The current draft of the MPEG-21 Contract Expression Language (CEL) actually includes a work derived by PPAVRO as an extension for media contracts. Reference [9] explains in greater detail how the contract expression language is going to provide an electronic counterpart to the traditional narrative contracts that supported the agreements on digital items in the MPEG-21 framework or material and services representable as such. It is based on an XML schema standard specification. Therefore, in the future, it will not be necessary to manually analyse contracts in order to extract a machine readable rights situation because the operative part of the new electronic contracts will be able to directly carry that information in the OWL/XML coding.

V. THE BASIC SERVICES FOR RIGHTS MANAGEMENT

The PrestoPRIME scenario relates to a digital AV preservation system, responsible for both the content and the metadata. The users of such a system can play the “producer” role when submitting information for ingest or update, or the “consumer” role, when posing queries, browsing content and information, and gaining access to materials. The same model applies to rights information. However, services for rights management have their peculiarities, specifically when creating and modifying the rights information, but also when making queries based on rights criteria. In addition, working on rights is a specialised task and the information has a degree of confidentiality, even within a single organisation. Rights editing services are therefore deployed separately from archive services, which must include services for rights indexing and retrieval and must support policies for access control. The proposed approach is to offer a remote service with a web user interface, consumed by using a browser, in order to avoid dependencies on the client operating system and without the need to install any application local to the client. Such a service can support multiple users, each one with its document storage area, with access granted through a basic authentication mechanism.

A. Basic functionalities

The remote service has been conceived as a rights editor allowing CRUD (create, read, updated, delete) operations of documents containing rights instances, represented by OWL files. The back end of such a rights editing service is implemented by a few very basic services, which can be invoked by a higher level application, for adding and removing individuals, object

properties, and data properties. The complete service interfaces are described in [8]. They also include interfaces for exporting the rights document locally or for ingesting it through the archive services.

B. Graphical representation

The ontology OWL document has the benefit of providing an unambiguous representation of the rights situation while enabling software processing of the information. The drawback of OWL is that, although the document itself is XML, either with OWL/XML [12] or Resource Description Framework RDF/XML [11] serialisations, it is very difficult for a human user to easily understand what the rights situation is.

Since the RDF represents a graph, it is possible to automatically derive a diagram which can graphically represent the rights situation. To achieve this goal we developed an XSL tool in order to transform the OWL document into a representation that can be consumed by graphic tools (e.g. Graphviz [13]). Figure 1. and Figure 2. were obtained through this process. The ellipses represent the individuals, the class name is also given, the arrows represent the object properties, and the data properties with their values are represented by grey boxes linked to the individual to which they are related.

The meaningfulness of such a representation was evaluated in the 2010 project test-beds, reported in [14]. It was found that the graphical notation is effective in representing the rights situation, it can be clearly understood by the users and requires only minimum training on the notation and OWL technologies. The diagrams are compact and expressive for simple rights situations, but they can become too detailed and difficult to understand in the case of very complex instances, for example when many, various media temporal fragments have their own specific restrictions.

C. Integration with archive services

In PrestoPRIME we implemented an open platform for digital preservation integrating several tools and services, [16]. The rights services described below have been integrated into the preservation platform and can be used to ingest, update or access archived content through the platform REST APIs.

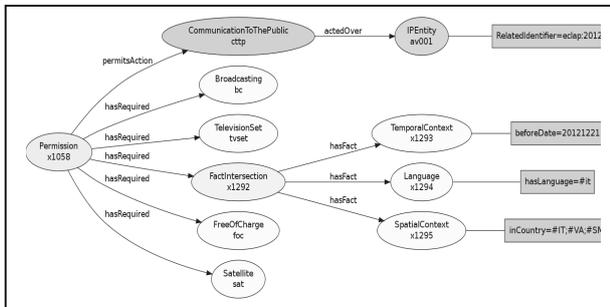


Figure 2. Graphical representation of an example of rights to satellite broadcast in a constrained context.

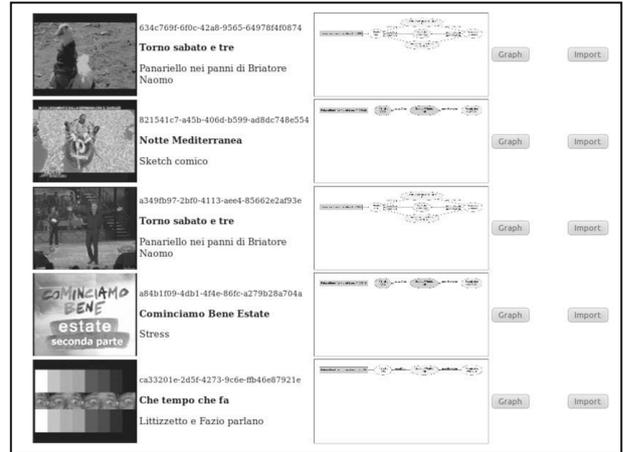


Figure 3. Presentation of results of a query on rights, with the link to the video preview, some descriptive metadata and the rights graph, and the buttons for retrieving the selected owl document.

The archive services interact with the rights manager for ingest and update, for answering the submitted queries and for providing access to the rights information, upon verification of access control properties. The OWL files are exchanged through the archive interfaces in the following way:

- during ingest, update, or access, where the rights information document is part of the “submission information package”. The PrestoPRIME project adopted the information package wrapper format defined in [15], which includes a specific rights metadata section. The OWL document can be inserted “as is” or included by reference. Another metadata section is used for setting the values of access control properties.
- during query submissions, where the query-by-sample paradigm is supported. The target rights situation is submitted in the form of an OWL document (not specifying any intellectual property entity), for comparison with the rights of the archival items. The answer is provided in XML format according to a schema defined in [8] and is presented to the user of the rights management application as in Figure 3.

VI. TOWARDS AN ADVANCED SET OF SERVICES FOR RIGHTS MANAGEMENT

The first set of basic services, including the integration with archive services, has been positively evaluated in 2011 project test beds, reported in [17]. Now the challenge clearly moves to designing a set of reliable and user effective services, beyond the prototype developed in the proof-of-concept context of an evaluation event. The following requirements must be fulfilled: (i) allowing quick migration from legacy systems when adopting the new services; (ii) enabling the users to operate in a simple, fast, and secure manner; (iii) defining a process for integrating new extensions,

without having to re-engineer the tools or analyse again the rights instances.

A. A set of key rights patterns from an organisational perspective

After having analysed a wide set of contracts, we could conclude that although the contract clause text has complex semantics, in which a minor detail can completely change the meaning of the agreed terms, in most cases only a limited number of rights patterns occur. Our current effort is on trying to produce precise figures on this, although our preliminary evaluation estimated that more than 90% of rights situations will fall under a pre-defined key pattern with a limited number of differences, i.e. without any or with very few additional clauses. The reason for that is simple: most contracts are based on templates, which are then modified according to the specific needs of the parties and the context. Within a single organisation or within a particular market of rights (such as those of movies, fictional TV, sports and theatrical events) contract templates are common practice. Such templates are subject to change over time, following the evolution of the legal framework, the technologies related to delivery and fruition, and market drivers.

The patterns which are most relevant to one organisation's business can be different from those of another. Each organisation may prefer to make reference to the various patterns with names or labels which are familiar to the users of their teams.

Working on a key rights-pattern definition by using the PPAVRO model is relatively easy and can be done by means of the basic rights editor described above, although it may require a deeper knowledge of the legal aspects of the contract clauses. The professional user can validate a pattern through the examination of the graphical notation because each single pattern is quite simple and thus the limitations identified above for the complex cases are not relevant here.

The approach is based on the conclusion that once patterns are well defined, it is easy and fast for the users to recognise them during the analysis of a contract. Then, the identification and extraction of the key information which defines the required exploitation constraints can be an easy task.

We defined a set of patterns to be included in our prototype with the following labels: all-rights, audiovisual-distribute, commercial-video, DVBH, FreeTV, IPTV, SVOD, satellite-free, satellite-pay, VOD-free, VOD-pay, video-rental, video-sell-through, WebTV.

B. An architecture for the advanced services

The use of the pre-defined key rights patterns during the creation or modification of OWL instances doesn't have any impact on the resulting documents, which are

serialised independently from the pattern definition. The names or labels associated with the patterns are only used for the user interface presentation but they are not recorded within the OWL document. This prevents any compatibility problem with patterns adopted in other organisations or future pattern modifications.

Here we provide an overview of the necessary adjustments to the rights services architecture, in order to support the advanced management features. The basic core functionalities are maintained, but deployed mostly in the back end. There is a package dedicated to key patterns with various components supporting the following functions: (i) creation and development of new patterns (ii) pattern validation, (iii) a repository of validated patterns, with version track support, (iv) deployment of the patterns for use in the appropriate contexts (v) access control. The package for rights instances creation and update is made up of a component for the creation from scratch, a component for the presentation of the rights information and one for supporting the updates.

The development of the creation from scratch service is the simplest because it relies on the patterns defined and available at the moment of work. Conversely, a few issues need to be addressed concerning the presentation of rights with respect to the defined key-patterns. Indeed, when a rights OWL instance is parsed, the fact that a target pattern matches or not has to be verified each time. This is possible with our rights compare tool, which is also used for answering the queries. Rights matching, as explained in part A of [8], is not dependant on the pattern definition and the set of patterns matching a rights situation may be wider than the set of patterns used for defining it.

C. Support to purchases and sales

A rights management system has to provide support to the trade of rights for purchases and sales. The former is straightforward and is defined as follows. The purchase agreement is expected to result in an OWL instance document including some permissions to exploit an IP entity; assuming that the buyer already has some rights on the same IP entity, and thus the system contains the related rights information within another OWL instance document, the operation to be performed is to simply merge the two documents and validate the result.

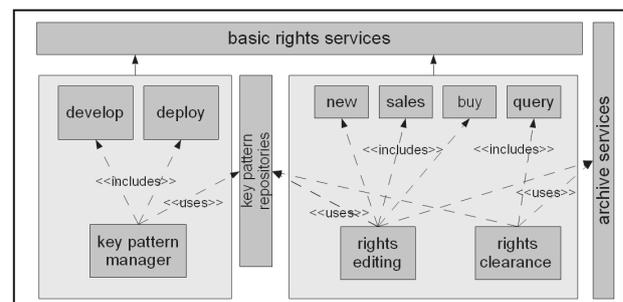


Figure 4. Architecture for advanced rights management services

The scenario regarding sales of rights is more complex, as it is required to check if the requested rights are owned by the seller and if sub-licensing is allowed. Assuming that the trade will result again in an OWL document, the problem is to update the rights situation regarding the seller. If exclusivity is not granted nothing has to be done, because the seller keeps the same rights as before the trade. Otherwise it is necessary to modify the relevant context of constraints according to the result of the deal. For example, an expiration date for a given right should be set to the date after which it becomes valid for the buyer. Figure 4. shows our architecture design for advanced services. The services supporting rights editing and rights clearance are conceived to optimise their specific task, with greater integration with the archive services, also for the examination of the AV materials. The basic services are required for the development of the key patterns and for handling the possible exceptions.

VII. CONCLUSIONS

The experience of the legacy approach to audiovisual rights management showed that effective and efficient services must rely on a well defined model supporting a clear format for information exchange. In the work carried out within PrestoPRIME, we produced an ontology model, based on an MPEG-21 specification. The goal is to become the operative part of the future MPEG-21 CEL standard. The rights instance documents use the same language used for the model and can benefit from its expressiveness and clarity, enabling the introduction of automated information processing. We were able to develop a set of services implementing the very basic functionalities and have proven that it is also possible to create a wider architecture for advanced and powerful services. The ability to provide the required flexibility to user applications and easily adapt them to changing contexts is the key factor for success.

Digital AV archive and rights management services described in this paper are under testing in RAI archive, where they will be integrated with the existing systems in order to optimise the mutual benefits. A correct and efficient management of rights will enable a wider exploitation of archive content, from which the costs of preservation are also financed. On the other hand, access to archival browsing copies is very useful to the work on rights clearance and rights editing. The appropriate definition of the service interfaces permits such integration, while at the same time, addressing the specific requirements of both areas.

ACKNOWLEDGMENTS

This work was partially supported by the PrestoPRIME project [1], funded by the European Commission under ICT FP7 (Seventh Framework Programme, Contract No. 231161).

The authors warmly thank Beth Delaney for her valuable suggestions and help with proofreading.

REFERENCES

- [1] PrestoPRIME, “Keeping audiovisual content alive”, <http://www.prestoprime.eu>
- [2] A. Di Carlo, M. Bartoleschi, G. Picciotti, “Glossary of rights”, <http://purl.org/NET/pprights/glossaryCS>, May 2011, revised February 2012.
- [3] J.H. Chenot, A. Di Carlo, “Content tracking system proposal and common rights ontology”, PrestoPRIME deliverable D4.0.5, [part B: Common Rights Ontology] <http://purl.org/NET/pp-ID405B-CommonRightsOntology>, 2010.
- [4] ISO/IEC/JTC1/SC29, ISO/IEC 21000-19, “Information Technology Multimedia Framework (MPEG-21) part 19: Media Value Chain Ontology”, 2010.
- [5] CCSDS 650.0-B-1, “Reference model for an open archival information system (OAIS)”, January 2002.
- [6] PrestoPRIME audiovisual rights ontology, <http://purl.org/NET/ppavro.owl>
- [7] J. Delgado, V. Rodriguez Doncel, E. Rodriguez, W. Allasia, F. Gallo, E. Todarello, L. Boch, A. Di Carlo, A. Messina, “A Proposal for an MVCO Extension for Audiovisual Contracts”, ISO/IEC JTC1/SC29/WG11MPEG2010/M17749. July 2010.
- [8] L. Boch, W. Bailer, M. Hildebrand, J.H. Chenot, “Metadata and rights APIs and service interfaces”, PrestoPRIME deliverable D4.0.2, 2011.
- [9] Boch, A. Di Carlo, J. Delgado, E. Rodriguez, V. Rodriguez, F. Gallo, “Towards a contract expression language for audiovisual content”, Virtual Goods+ODRL 2011 [9th international workshop for technical, economic and legal aspects of business models for virtual goods], Barcelona, September 2011.
- [10] OWL 2 Web Ontology Language Document overview, W3C Recommendation, <http://www.w3.org/TR/owl2-overview/>, 27 October 2009.
- [11] RDF/XML Syntax Specification (Revised), W3C Recommendation, <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>, 10 February 2004
- [12] OWL 2 Web Ontology Language XML Serialization, W3C Recommendation, <http://www.w3.org/TR/2009/REC-owl2-xml-serialization-20091027/>, 7 October 2009
- [13] Graphviz, “Graph visualization software”, <http://www.graphviz.org>
- [14] C. Bauer, “Report on evaluation phase 1”, PrestoPRIME deliverable D5.4.2, January 2011.
- [15] The Library of Congress, “METS: Metadata Encoding and Transmission Standard”, <http://www.loc.gov/standards/mets/>
- [16] F. Gallo, “First prototype of open PrestoPRIME reference implementation”, PrestoPRIME deliverable D5.2.2, Dec. 2011.
- [17] C. Bauer “Report on evaluation phase 2”, PrestoPRIME deliverable D5.4.3, December 2011.

Audiovisual Media Open Educational Resources

The case of SP-ARK: The Sally Potter Film Archive

Dr Sarah Atkinson

Faculty of Arts: University of Brighton:
Brighton, UK,
s.a.atkinson@brighton.ac.uk

Dr Janko Čalić

I-Lab, Centre for Vision, Speech and Signal
Processing, University of Surrey, Guildford, UK
j.calic@surrey.ac.uk

Abstract This paper provides insights into a research project in progress that investigates the collaborative potential and opportunities between open educational practices and online open audiovisual archives. As part of this research, an in-depth evaluative case study is being undertaken in conjunction with a number of University's into the pedagogic potential of the Sally Potter Archive, SP-ARK. This paper investigates the SP-ARK case study within the wider context of audiovisual OERs and audiovisual archives, and policy issues surrounding their access, use and licensing.

Keywords- SP-ARK, Sally Potter, Open Educational Resources, OER, Audiovisual Archives, Creative Commons, Participatory Learning, Social Learning.

I. OPEN EDUCATIONAL RESOURCES

‘Open Educational Resources (OER) are teaching and learning materials that are freely available online for everyone to use, whether you are an instructor, student or self-learner. Examples of OER include: full courses, course modules, syllabi, lectures, homework assignments, quizzes, lab and classroom activities, pedagogical materials, games, simulations, and many more resources contained in digital media collections from around the world’ (OER commons definition). There are a number of Audiovisual OERs housed within some of the general resource repositories in the UK, most notably The Joint Information System Committee’s Online Repository for [Learning and Teaching] materials (JORM), but these tend to be text or still-image only based, primary sources of audiovisual matter that go beyond tutor-led presentations and instructional screen-captures are notoriously difficult to identify, locate, access and repurpose. The reasons for this will be more fully explicated within this paper.

II. ON-LINE AUDIOVISUAL ARCHIVES

Traditionally, the primary sources generated by filmmaking are housed within specific physical locations that are not easily accessible and not always open to the public. In the UK, the national film repository is the British Film Institute (BFI), some of the materials are available online, but only those that are

not restricted by copyright issues. There are also a growing number of both subscription-based and open-access broadcast archives online, which include *Box Of Broadcasts (BOB)* a service delivered by the British Universities Film & Video Council (*BUFVC*) in the UK and *EU Screen and Europeana* in Europe. These house audiovisual assets such as newsreels, reports and documentary footage from television broadcasts. The equivalent archives within the fictional and dramatic realm of cinema are extremely limited, yet could be invaluable tools in the maintenance and preservation of cinematic and cultural heritage. As Gerhardt and Kaufman have noted ‘This disconnect – perhaps we call it an ‘A/V gap’ – is largely a function of attitudes and behaviours within teaching, production, and publishing. It is also an outcome of the paucity of quality audiovisual work now available for educators. As we note in our 2010 Film & Sound Think Tank film, *Knowledge Is...*, despite the leading investments of JISC and others worldwide, only 5% of our audiovisual history is digitized and available to educators and the public online’ (2011:3). There are a limited number of disparate and disassociated online repositories and databases, which house the culturally rich resources of filmmaking and cinema across the globe. In the UK, the only known openly accessible online version of such an archive is SP-ARK. Internationally, in Japan the entire archive of prolific Director Akira Kurosawa has been uploaded and is freely available to view, access and download online, but is only currently available in the Japanese language. These exemplars are invaluable and enriching resources within film and cinema education, since in their exposure of all the materials related to a film production; photographs, video assist footage, casting recordings, scripts, storyboards and developmental paperwork, they reveal the often hidden creative practices of fictional filmmaking. As Mayer contends in relation to SP-ARK; ‘It reflects the dailiness of labour involved in filmmaking as opposed to the heroic narrative portrayed in mainstream films’ (2008:201). Within educational contexts, the access to such primary resources facilitates the close textual analysis and in-depth examination of films, practices that could not be achieved without such access. Such resources also have the potential to provide unique and unprecedented sites for communication, collaboration

and the establishment of both online and physical networks. It is the key issue of licensing that appears to be foreclosing the development of open access to film and cinema resources.

In an attempt to address this issue, Creative Commons Licenses were established and were first accessible in 2002, (these tend to be the licenses that are incorporated into open educational practices). Lawrence Lessig, Creative Commons founder and Stanford Law Professor states that ‘its aim is to build a layer of reasonable copyright on top of the extremes that now reign. It does this by making it easy for people to build upon other people’s work, by making it simple for creators to express the freedom for others to take and build upon their work’ (2004:282). Although this continues to be a problem, fortunately it is not so for the case of SP-ARK, since the copyright to all of the materials belongs to *Adventure Pictures*, and they have chosen to allow access and use of the materials via a Creative Commons licensing model.

III. SP-ARK

The SP-ARK archive provides a unique example of the successful marriage between the principles of open educational resources and open archives. SP-ARK was first established in 2007, and is an interactive online project based on the multi-media archive of filmmaker Sally Potter. Potter is a world-renowned film director, known for her explorations into the potential of nascent technologies to enhance audience engagement and participation in her work. Her 2009 film *Rage* was the first feature film to be launched and distributed on mobile phones. Over the past five years, the archive has been developed to a Beta-testing level, and includes the intuitive visual navigation of one of Potter’s films, *Orlando* (1992), and all of the related assets. All of the resources have been digitised and meta-data has been added relating to the items description and association with other assets. In the latest iteration, which was officially released in February 2012, users are able to view clips from the film as well as a myriad of associated materials including the scripts, storyboards, still images, location and developmental paperwork, using the intuitive visual browsing interface (see Figure 1 and section V for further technical details). Resources are initially organised in a linear taxonomy which aligns with the sequence of the film production process; Development, Preproduction, Production, Postproduction, Finished Film and Distribution (see the indexical sidebar in Figures 1 and 2). The materials within these processes are then organised within further drop-down subsections. Once an asset is opened, users are then able to continue browsing the archive in a non-linear and exploratory fashion by linking to the asset’s ‘related items’ (for every item is linked to other associated items) for example, a page of script is linked to its corresponding clip, which could then be linked to a call sheet, a continuity report, production design images, location notes etc.

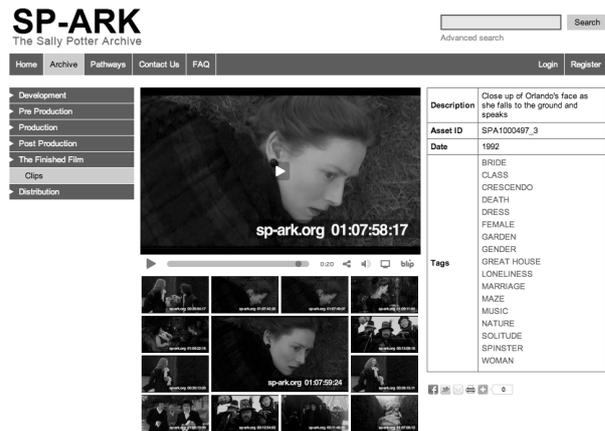


Figure 1. The SP-ARK Visual Browsing Interface

IV. PATHWAYS

This mode of browsing allows the user to build his or her own unique ‘pathway’ through the archive’s content as they explore a particular theme or process; they are able to save items that they have viewed. This type of archival browsing which is embedded into the infrastructure of the latest version of the site is not so easy to achieve through the boxed presentation of materials within a traditional physical archive. This intuitive browsing is extended and supported by the fact that users are able to annotate each individual item in their pathway, with their own comments, observations and streams of thought, as well as to describe and save the pathway itself. Other users are then able to access each other’s pathways (when they click on an item all associated pathways are displayed) and to link to them (see Figure 2), which offers a further level of user-led archival exploration. Users are also able to directly communicate with one another using the messaging tool. This type of interaction leads to the deeper engagement with the materials, encourages the sharing of ideas and practices, and fosters the creation of a user-community around the archive’s content. The SP-ARK resource exemplifies the successful combination of an archive and an educational resource within this feature, providing a unique model for social and participatory learning. The benefits that such a resource can bring to higher education academics and students are invaluable and as yet unprecedented. The pathways tool lends itself to the critical and analytical study of primary materials as intrinsic to both undergraduate and postgraduate study within numerous disciplines. The successful development and utilisation of such a resource has the potential to enhance and enrich teaching and learning practices within these disciplines, as well as to encourage other high-profile filmmakers and organisations to allow online access to their work in the future.

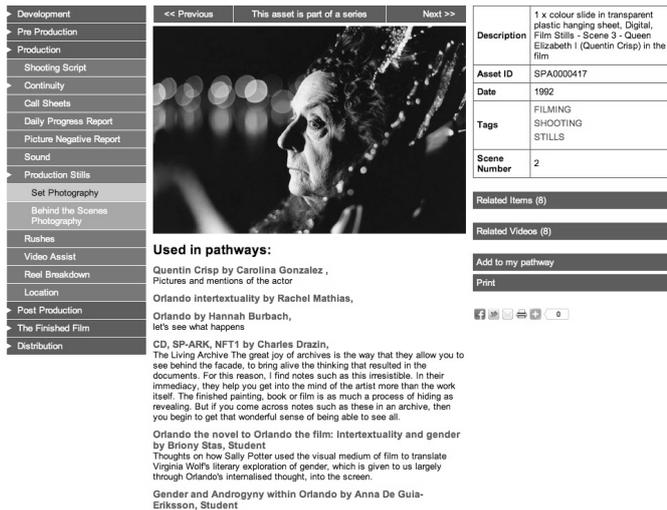


Figure 2: An example of the pathways tool in action

Dr Charles Drazin has already used the pathways feature as an assessment tool on the Film Studies Programme at Queen Mary University of London (QMUL). The students were set a discussion topic through which they had to construct their answer within a pathway rather than a traditional essay format. Drazin noted that ‘from a teacher’s perspective what was great about the site was to be able easily to visit students’ pathways and to see their thoughts take shape. It facilitated the provision of on-going feedback as students worked on their assignment in a way that is not feasible in traditional coursework’. In addition, students of the exercise also responded positively. The assets that the students identified, along with content of the pathway could then be used as OERs themselves; as envisaged by the OER impact study suggesting ‘validating the sharing of online resources discovered by students’ (JISC, 2011: 25).

V. VISUAL BROWSER

The latest version of SP-ARK incorporates a visual browser which was designed and integrated as part of a Knowledge Transfer Partnership between *Adventure Pictures* and the Essex University’s Department of Literature, Film and Theatre Studies and the University of Surrey’s Centre for Vision, Speech and Signal Processing. Full explications can be read in Ren, Sarvas and Čalić (2010).

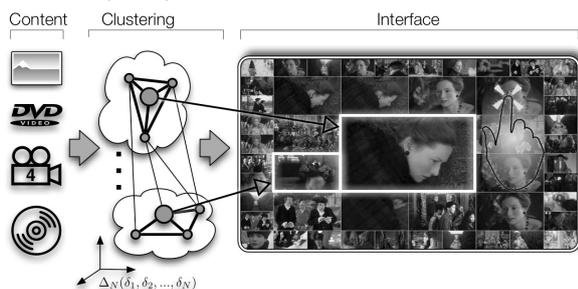


Figure 3. Block scheme of the interactive image browsing system

The visual browser comprises two modules: an image clustering engine, that derives the underlying structure of the database, and a hierarchical interactive interface depicted in the Figure 3. The size of every image in a generated interface layout is proportional to its similarity to the central image. The choice of the similarity metric is invariant to the type of clustering engine and/or the interface design, enabling generic application of this system. In case of SP-ARK visual browser, a chi-square distance between three-dimensional RGB colour histograms was utilized as the similarity measure. The shots were represented by a set of key-frames efficiently extracted using a method for video summarisation introduced by Čalić et.al (2007).

VI. USER TESTS & DATA

In order to evaluate the usability of the SP-ARK web interface a focus group session was held at the School of Languages, Linguistics and Film at Queen Mary, University of London. The user study comprised an interactive design session, user interviews and a questionnaire. The possibility to browse visual content in an intuitive way as well as the pathway functionality were the strongest aspects highlighted by this user group. These are some representative subjective feedback responses: i) “*Very easy to navigate and understand. Colour effectively used to highlight different aspects of the site,*” ii) “*It was very accessible and easy to use. It is also very bright and inviting,*” iii) “*The layout and design made the website/archive extremely simple to use.*”



Figure 4. Visual browser representing semantically coherent content from the SP-ARK archive

During the discussion at the SP-ARK project launch as a part of Sally Potter retrospective at the British Film Institute in London in December 2009, Sally Potter specifically responded to the instance of visual browser depicted in the Figure 4. She recalled a breakfast with the director of cinematography Aleksei Rodionov in a hotel in Moscow during the production preparations, when they discussed the visual links between mist, snow and pale yellow palette in various scenes to maintain

visual and narrative coherence. During the launch, the very scenes they were planning to link were represented in the single instance of SP-ARK visual browser. This symbolic, but at the same time very real, proof of deep semantic links represented in the visual browser acknowledged by the author herself demonstrated that the user-centric approach to interaction with large-scale visual repositories has got an immense potential to augment, if not replace, the semantic efforts in the visual content management research.

VII. CONCLUSIONS & NEXT STEPS

This case study will go on to test and report upon the educational potential of this open resource within the fields of media and film theory and practice. The case study will include the facilitation of focus groups with students and staff at various universities. The focus groups will be used to demonstrate, explore and evaluate the potential of the archive as a teaching, learning and assessment device; and to collaboratively generate, develop and share open educational resources around the content of the archive. Currently, in addition to QMUL, students from Bucknell University, Pennsylvania USA studying Film and Media Studies and students on the Historiography course at the New York University are actively engaging with the archive. Their assessments are related to creating critically informed thematic reflections using the Pathways tool.

The next step for the development of SP-ARK is already underway, and aims to extend the reach of the archive into vocational and professionally based education. Potter's latest film *BOMB* has just completed production, and during the process, key members of the production crew were furnished with portable flip cameras to record their daily activities on set. *The Anatomy of a Film Set: Exposing the people, the roles, the processes and the careers on set of Bomb* will create an interactive audiovisual ecology of the film production process, capturing all of the individual crewmembers contributions to the creation of the film through personal testimony. The project exposes *all* the roles on set from runners and caterers to camera, sound, costume, make-up, continuity, set builders, sparks, extras to heads of all the departments. This project will extend beyond the production phase into the postproduction, marketing and distribution of the film. This will be presented in visual form whereby the user can intuitively access the video diaries, photos and testimonies. This will extend SP-ARK as an invaluable and innovative resource for young people and students seeking careers in the film industry, and to educators and academics teaching film production and processes.

These activities seek to draw out the benefits and efficiencies of collaborative resource generation, exploring the challenges of sustainability and expansion of both the resources and the encompassing user-group community. The findings of the case study will not only inform the future direction of SP-ARK; an endeavor which has always placed the educational community at

the core of its development (initially at the Screen School at Goldsmith's College, see Mayer: 2008) but also has the potential to support and inform the approaches of emerging online film-based repositories as they grapple with the issues of openness, reuse and licensing. The project ultimately provides an innovative example of Higher Education Institution and archive collaboration in action, which could in turn provide a compelling model for the development of open academic practice.

ACKNOWLEDGMENT

The authors would like to thank Sally Potter, Christopher Sheppard and Clare Holden at *Adventure Pictures* for pioneering and sharing such an inspirational and groundbreaking resource, the University of Brighton, Essex University: Department of Literature, Film and Theatre Studies, University of Surrey Centre for Vision, Speech and Signal Processing, Charles Drazin and Lucy Bolton; Queen Mary University of London, Eric Faden; Bucknell University, Christine Gledhill; New York University, The Open University's Support Centre for Open Resources in Education (SCORE) and JISC for funding the SP-ARK case study.

REFERENCES

- [1] British Universities Film & Video Council [Online] <http://beta.bufvc.ac.uk/> Accessed 02/02/12
- [2] Čalić, J., Gibson, D., Campbell, N., Efficient Layout of Comic-Like Video Summaries. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(7). ISSN 1051-8215, pp. 931–936. 2007.
- [3] EU Screen: Europe's Television Heritage [online] <http://www.euscreen.eu/> Accessed 02/02/12
- [4] Europeana: Europe's Cultural Collections <http://www.europeana.eu/portal/> Accessed 02/02/12
- [5] Gerhardt, P., and P.B. Kaufman. *Film and Sound in Higher and Further Education: A Progress Report with Ten Strategic Recommendations*: HEFCE/JISC, 2011. Print.
- [6] JISC, Open Educational Resources: *The value of reuse in higher education* [online] http://www.jisc.ac.uk/whatwedo/programmes/elearning/oe2/oe2_impact.aspx Accessed 12/1/12
- [7] JORUM, Free Learning Resource for Teachers, [online] <http://www.jorum.ac.uk/> Accessed 15/12/11.
- [8] Kurosawa, A. Online Film Archive, [online] <http://www.afc.ryukoku.ac.jp/Komon/kurosawa/index.html>. Accessed 15/12/11.
- [9] Lessig, L. *Free Culture: The Nature and Future of Creativity*. New York: Penguin Books, 2004. Print.
- [10] Mayer, S. Expanding the frame: Sally Potter's digital histories and archival futures in *Screen* 49:2 Summer 2008. Print.
- [11] OER Commons, [online] <http://www.oercommons.org/>. Accessed 20/12/11.
- [12] Ren, K., Sarvas, R. and Čalić, J. Interactive search and browsing interface for large-scale visual repositories in *Multimedia Tools and Applications*, Volume 49, Number 3, September 2010, pp.513-528.



Workshop on Performing Arts & ICT

Chair:

Kia C. Ng, University of Leeds, UK

Workshop Committee:

Jerome Barthelemy, IRCAM, FR,
Tillman Weyde, City University London, UK
Laurence Pearce, Xim Ltd., UK
Jorge Solis, Karlstad University, SE
Michal Masa, CIANT, CZ

Languages for Augmented Choreography

Andrea Bene, Diego Bernini, Giorgio De Michelis and Francesco Tisato

DISCo, University of Milano-Bicocca,
viale Sarca 336, 20126 Milano, Italy
{bene, bernini, demichelis, tisato}@disco.unimib.it

Abstract— Choreographers may exploit digital technologies to create augmented choreographies where the behavior of dancers determines the behavior of projections, lighting and other scenographic effects. The paper recalls our ongoing activity concerning the development of an end-user software framework for augmented choreography and it summarizes the state-of-the-art of choreographer-oriented notation systems and tools. Finally the paper describes our strategy for building an end-user language for augmented choreography.

Keywords— augmented choreography; interactive performance; choreographic notations; choreographic tools

I. INTRODUCTION

Choreography may be defined as the art of designing sequences of movements performed by several dancers [1]. *Augmented Choreography* is a Choreography whose movements and effects are realized by digital technologies (Fig. 1). For example, one or more virtual dancers i.e., computer-projected body animations, can interact with the real dancers. Digital-based lighting and scenographic effects are other examples of interactive elements. Merce Cunningham Biped is a famous early example of augmented choreography [2].

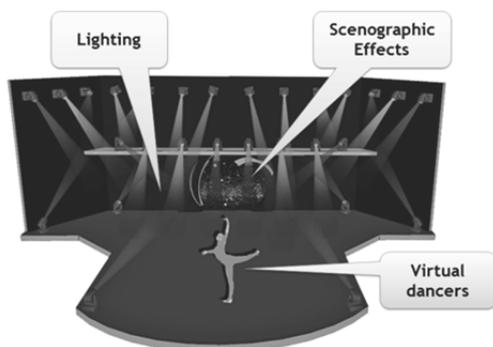


Figure 1. Digital-based choreographic effects.

Currently, Choreographers aiming to realize augmented choreographies are constrained by several factors, related to the current technology-driven approach, as listed below.

Technologist-dependency. Choreographers strongly need the assistance of one or more technologist. Software tools proposed for multimedia and creative purposes (e.g., data-flow visual programming environments like MAX/MSP and Pure Data) have simplified the development of such applications.

However, they are far to be directly accessible by non-technical people.

Technology-dependency. Too often augmented choreographies are driven by the available hardware/software technologies rather than by what choreographers want.

Ad-hoc technologies. The available solutions for this field are vertical ad-hoc solutions, developed for particular aims and/or for specific choreographies/choreographers and bound to specific technologies.

As a solution to these issues, we present an architecture proposal for an *end-user software framework* where choreographers can define augmented choreographies via user-friendly languages which hide technological aspects (Fig. 2). Furthermore we started an ethnographic study about choreographers that aims to highlight relevant features of dance languages.

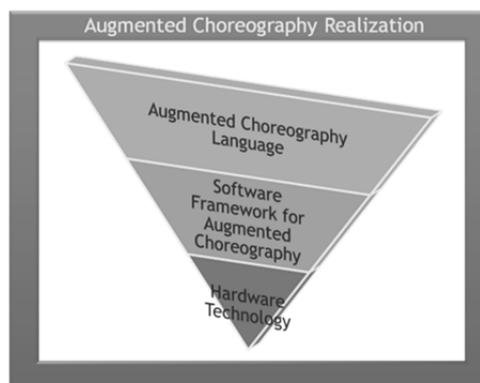


Figure 2. End-user augmented choreography framework.

In a previous study [3] we presented our vision and some first experimentations about the development of the framework. In this paper we present our refined perspective after studying the existing notation systems referring to choreographies and the available tools to support the end-user realization of digital choreographies. Moreover, starting from a critical analysis of the previous solutions in collaborations with several Italian choreographers, the paper proposes a different strategy for the end-user language which will be supported by the proposed framework.

The paper is organized as follows. Section 2 briefly summarizes our proposal for the augmented choreography framework, Section 3 presents the most

significant choreographic notations, while Section 4 presents the state-of-the-art IT choreographic tools. Section 5 describes our strategy for the end-user language. Finally, Section 6 outlines ongoing and future work.

II. AUGMENTED COREOGRAPHY FRAMEWORK

The framework we propose aims to simplify the development of augmented choreographies by providing an abstract design level hiding technological details, so that choreographers can focus on what they want, rather than on how to realize it.

A. Features

The framework we are developing must be capable of *a)* recognizing events and behaviors of the real performers *b)* executing actions/commands on *virtual performers* and *c)* let choreographers define the augmented choreography, i.e., the relations between events/behaviors and actions/commands.

With virtual performers we include virtual human-like projections, like virtual dancers but also projections of different forms or objects and different stage effects (lights, scenography, and so on).

Choreographers will define augmented choreographies as *mappings* between what happens on the stage (like the movements of the real dancers) and actions of virtual performers (Fig. 3).

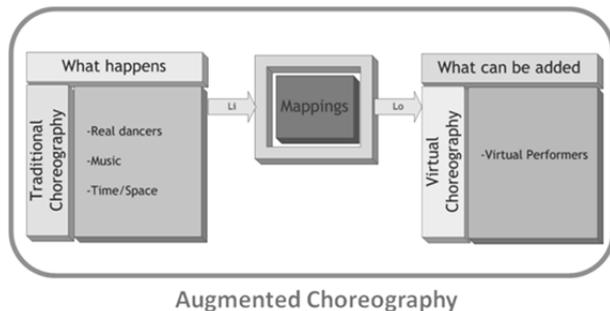


Figure 3. Augmented Choreography as mappings.

The framework must provide a suitable end-user language to express the mappings. Next sections of the paper discuss the topic of what kind of language could be exploited. The method we follow to build and define languages is based on a design approach for technological artifacts, based on a new concept of the artifact itself [4]. The importance of context is determinant to understand what the user want and which tools can be used. Fundamental are the studies about context by Dey [5] and Dourish [6]. To better define the most suitable design strategy, based on links between practices and meanings, we are studying languages and methods of choreography.

B. Architecture

The framework is organized accordingly to five abstraction levels (Fig. 4). Hardware technologies like sensors and output devices are managed by a proper

software level (the driver level). Over the driver level basic event detection and command execution are developed. Starting from sequence of events, significant behaviors may be recognized; analogously high-level actions may be defined in terms of one or more commands. Finally, the *choreography mapper level* allows to map recognized behaviors to actions according to the end-user language.

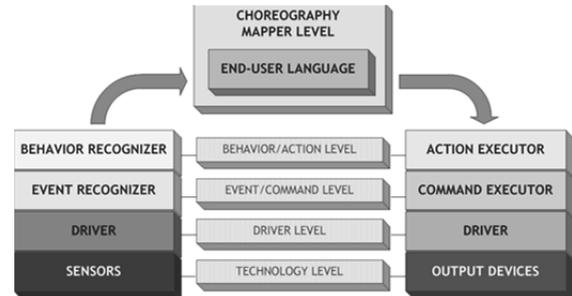


Figure 4. Abstraction levels for the framework.

From a software engineering point of view each level is reified by an extendible software component library [7], in order to manage several heterogeneous sensors, recognize manifold events & behaviors and execute manifold actions and commands.

C. Prototype & Experimentation

We have developed a first prototype of the framework which actually allows to manage mappings between events and commands. We are working towards supporting behaviors, actions and a suitable end-user language.

We experimented the framework with an Italian theatrical residency w.r.t. basic scenarios, as discussed in [3]. Next experimentations will regard the realization of specific choreographies in collaboration with Italian choreographers.

III. CHOREOGRAPHIC NOTATIONS

Notation systems represent the first techniques to hand down dances and ballet. From the fifteenth century to the present days we had a proliferation of notations. In the landscape of choreography of Western tradition they have traced the history of dance, being influenced by the most popular type of dance in each period.

Of course, different notation systems have different goals: some are particularly accurate in reporting steps and movements, others prefer the ease and speed of reading; though some were created as an aid to memorization of the steps, others were imagined to be used when creating new choreographies. In order to better clarify the different solutions Table 1 proposes a division into five main categories based on the technique used for transcription [8], presenting pros and cons for each category: words and word abbreviations; track drawing, stick figures (visual) systems, musical note systems, abstract symbols system

In the following we present the most used notation systems derived from the two categories of Stick Figures

and Abstract Symbols. The presented notation systems are important also because they have been applied for developing informatics tools to compose virtual choreographies.

TABLE I. DANCE NOTATION SYSTEMS

Categories	Features		
	Systems	pro	cons
Word and Words abbreviation	Cervera; Arbeau; Playford; Meunier System; Saunders System.	Intuitive; useful for memory; fast.	Few details; very expensive in terms of space; time-related; ambiguous terminology.
Track Drawing	Feuillet System; Lorin; Landrin;	Easy to read; clear description of movement in two dimensions.	Few details on the positions of the dancers.
Stick figure (visual) systems	Saint-Léon System; Zorn System; Benesh System; Sutton System;	Ready and willing; perfect description of positions; ideal for ballet; customization possibilities.	Movements unclear; difficulties with the third dimension; difficulty in describing the positions of modern dance.
Music note systems	Stepanov System; Conté System; Nikolais System;	Perfect synchronization of movements and music.	Risk of ambiguity; difficulty in determining duration of the movements.
Abstract symbol systems	Théleur System; Laban System; Morris System; Loring System; Eshkol-Wachmann System.	Can describe any movement; easy to read (except for Eshkol-Wachmann).	Wasteful in terms of space

A. Benesh Movement Notation

Benesh Movement Notation has been adopted by Royal Academy of Dance because of its rigour and flexibility. Born in 1956 just for dance uses, it is suitable to represent almost any sequence of movement allowing to describe the majority of the choreographic styles of recent years.

This notation system uses the music staff where each line corresponds to a section of the body (Fig. 5). The first one corresponds to the top of the head, the second to the shoulders, the third to the pelvis, the fourth to the knees and the fifth represents the ground. It uses three distinct signs to draw in three dimensions the location of anatomical parts. This simple system of symbols leads to a vocabulary of about 355 descriptive options.

Figure 5. A Benesh manuscript extracted by the *Voluntaries* by Glen Tadelly of 1975 [8]

This system is a great help for sketching and for those who have a choreography background is also relatively easy to learn. One of the main limitations of visual symbols in transcription is that, while there is a strictly defined ballet movement vocabulary, the same cannot be said for modern dance, where instead a wide freedom of customization is left to the choreographer. Over the years the *Benesh* has exceeded this limit because its flexibility allowed him to adapt to any new positions, even if its “ballet-oriented” nature still makes it stern to use in an augmented choreography.

B. Eshkol-Wachmann Movement Notation

A system that fits well for a computer use is the *Eshkol-Wachmann* (1958). This method is based on a spherical System of Reference to represent the general body pointed on the pelvis and a System of Reference for each joint. In this way, every movement can be detected by a set of coordinates in space and they can be summarized through simple mathematical expressions.

The final result of this representation is a matrix of numbers and symbols that express particular direction and expressiveness (Fig. 6). Each row corresponds to different parts of the body involved in the movement. The time is identified by separation lines between the columns of the matrix. Although it is extremely rigorous and hence computable, this solution is very expensive in terms of time, because the slightest movement requires a large number of annotations.

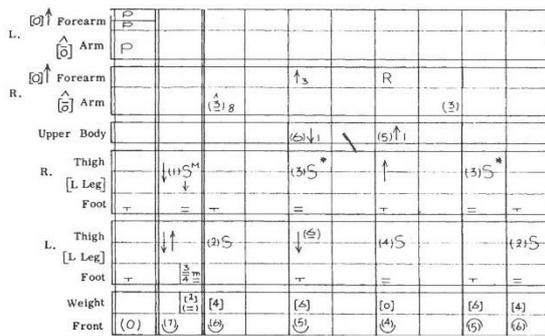


Figure 6. Manuscript extracted by the Diminishing Series suite by Noa Eshkol, as reported by [8]

C. Labanotation

All the notation systems arise from a general idea (as for musical notes or visual images) that is enriched and tailored to suit the characteristics of the movement in the dance. Conversely, the abstract symbols-based systems are born with the intention of classifying and describing the movements in general, then adapting to the particular case of choreography. This makes them very accurate and portable. We have chosen to treat *Labanotation* (1930) because it represents the right trade-off between similarity to choreographers language and strictness in terms of computation.

Labanotation is more intuitive [9] that it appears at first glance. The vertical positioning of the score for example, is already perceived as a representation of the human body in its symmetry (Fig. 7). This strategy also allows an easy reading by the performer, as the choreography is described like a stream without interruption.

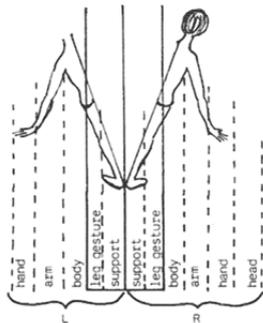


Figure 7. The Laban's staff structure

The variability of the block length provides an intuitive and flexible expedient to represent the time flow for different parts of the body allowing to overlay different durations referring to different limbs.

Movements made in one direction are represented in the same cluster, to facilitate further reading. The "clustered block" allows to immediately identify the key components of the movement, which is enhanced by the addition of dots, circles and other minor symbols (Fig. 8).

Another element that works in Labanotation's favor is its extreme "cheapness" of the transcription technique: each symbol on the score represents the part of the body to be moved, the direction and level, the exact time when the movement begins and the exact length. All these aspects seem to make *Labanotation* the ideal end-user language for Augmented Choreography.

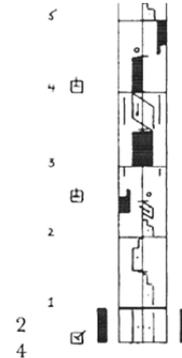


Figure 8. Examples of Laban staff [10]

D. Syllabus

Although it cannot be classified as a notation system, we have to mention here *Syllabus*, the basic vocabulary to define choreographies. *Syllabus* describes each dance pose with a French name, a graphic representation and a verbal description. The choreographer and the dancers must both know every term and its meaning.

IV. CHOREOGRAPHIC TOOLS

Most computer tools designed for the dance exploit the previous notation systems to digitize choreography. In particular, the most common solutions use *Labanotation*, *Benesh Movement Notation* and *Eshkol-Wachmann Notation*. Among the analyzed systems, we must distinguish those which act only as simple text-editors from those which make possible to reproduce the written choreography as digital animation.

It is important to highlight that none of these tools allows to create digital choreographies where some elements depend from real-time behaviors of the dancers on stage. They allow to interpret and visualize a choreography described by a notation system, thus creating virtual dancers. Nevertheless, the resulting animation cannot interact with the sensing data about the (real) dancers.

A. Eshkol-Wachmann Tools

As for the tools using *Eshkol-Wachmann* notation, which we know to be based on a geometrical approach, there is the EW Notator©, created in 2008 by Drewes Enner in collaboration with the Universitat Salzburg [11]. The EW Notator allows to translate the notations in a digital file exported in XML format. This software is complemented by the Movement-Oriented Animation Engine which consists of a graphic engine that should be able to render files created by EW Notator © [12], but the latest publications on the project date back to 2008.

B. Benesh Movement Notation Tools

As for the use of notation systems more closely related to the dance world we propose the Benesh Notation Editor, a program for Windows PCs that allows you to write using the Benesh method. The Benesh Editor has been particularly successful at the Royal Academy of Dance in London which, as stated above, still uses it [13]. Also there is the possibility of reproducing these files exported in PDF format, through the use of a second animation software specifically for dance, Dance Forms 1.0 [14].

C. Labanotation Tools

A first application was developed by Ohio State University who created a software package for Macintosh. This solution consists of a LabanWriter [15] that allows you to write and compile Labanotation movements, and a viewer, originally designed as a plug-in for another animation software (DanceForms 2.0), which then became a stand-alone application called LabanDancer. There are also numerous other experiments that allow the creation of digital Labanotation transcripts [16] [17] and generate lifelike animations of these [18][19][20].

D. DanceForms 2.0

Finally there is an animation software dedicated to dance, Dance Forms 2.0 ©, which uses a completely new system for transcribing the choreography, based on palettes of positions of dance [21]. It also allows you to render files obtained through motion capturing techniques formats BioVision, Acclaim, and HTR. It can be used as a visual idea generator to sketch out basic choreographic concepts, or to plan and record even the most intricate and detailed dance steps.

DanceForms' specialized interface makes it easy to create realistic dance poses and motion, with single or multiple figures. It presents an innovative solution to control a virtual dancer independently from all the notation systems. This graphical strategy is composed by preset dance positions extracted from dance's Syllabi and positioned in a timeline (Fig. 9). This method solves a pernicious problem of usability and for this reason it has had a strong influence on our studies on languages. Furthermore, the language used by this software has been approved by Dance Notation Bureau.

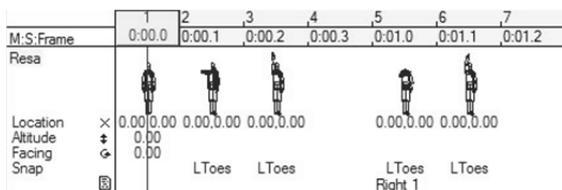


Figure 9. Timeline to compose a virtual choreography on DanceForms 2.0

V. OUR PROPOSAL

Dances in history have clearly influenced the notation systems. In more recent times the need to

legally protect the choreography is felt stronger, generating a further leap in description quality, generating a series of many variations and interpretations of existing systems. From the beginning of the twentieth century until the mid-70s, there was a proliferation of notation tools that used solutions more specific in relation to various dance styles. The explanation of the 70s collapse must be sought in the development of audio-visual technologies, that began to prevail in that period and in the years to come.

A. Usability Problems

Studying the literature it appeared plausible that end-users could directly use the Labanotation to control behaviors and movements of virtual dancers. Nevertheless, after having held meetings with several choreographers and dancers, we realized that the introduction of another language layer was needed. The question that all the choreographers pose today when they are confronted with a notation system, even if the most complete and efficient, is: why should I waste time learning to read a score of a choreography, when I can use the numerous video clips, already available? What could be more explicit and detailed of a video recording? The ineluctable simplicity of the answers to those questions made us realize that although we found in Labanotation the best system for digital use, it cannot be used as a final language for Augmented Choreography. This because the choreographers are no longer able to use the notation systems, and it will take too much work and time to create a complex choreography.

A. Syllabus and Movement Modulation Parameters

The idea we are considering is that of a structured language based on two dimensions named respectively "Syllabus and Movement Modulation Parameters". If the first represents a successful solution adopted by DanceForms, the second is totally new. The Syllabus consists of a set of steps and movements, described as a graphic primitives, which will be used on one hand by the system to "understand" what it is happening in reality, namely the steps of the dancers on stage, and on the other hand to allow the choreographer to associate the movements of the real performers to the behaviors of virtual actors (Fig. 10).



Figure 10. Example of four graphical primitives in a Syllabus

At this point, the choreographer will prepare a series of actions and behaviors of virtual performers to be associated with those of the real dancers using simple icons that represent the steps. Furthermore, if the choreographer will want to emphasize some particular

moments of the performance, he/she can associate, to one or more portions of choreography, particular values for the Movement Modulation Parameters of virtual performers (Fig. 11).

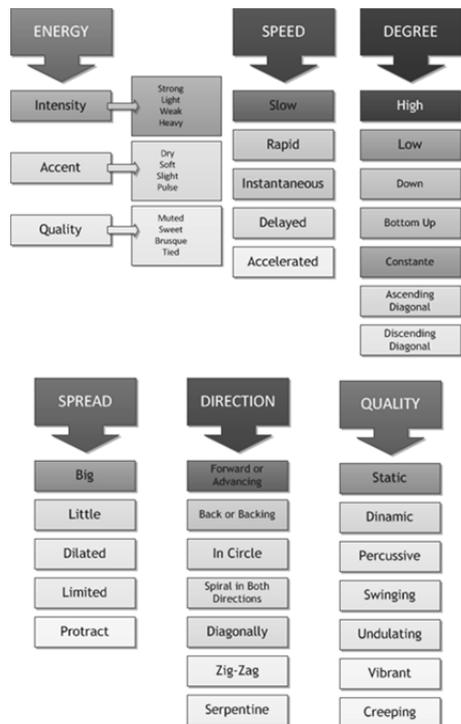


Figure 11. Movement Modulation Parameters

If the choreographer aims to express a nervous behavior, he will just need to change the Energy variable setting Intensity, Accent and Quality respectively on Strong, Dry and Brusque, and Speed on Accelerated. It could further simplify the selection of behaviors through preset options. These variables also allow the choreographer to manipulate in real time the augmented choreography, making it dynamic. In conclusion the Movement Modulation Parameters can be applied to every kind of virtual performer.

VI. ONGOING AND FUTURE WORK

To better proceed in the development of the project features, it is necessary to test them together with choreographers and dancers. Various proposals for collaboration with choreographers and technical experts of interactive installations are under consideration. The system will be developed in the world of dance and performing arts, surely, but it could be also used for pure artistic experimentation.

Looking at the project from this different point of view, many contexts of use can be exploited, namely, theaters, museums, exhibitions, galleries, and events that may become ideal places and opportunities to present an Augmented Choreography, employing disparate strategies of interaction. With experimentations, the concepts of Choreography and Dance will take on new and unexpected connotations, the art world should be immediately involved in testing,

in a free and open manner, so that every idea can be shared and exploited for the benefit of increasingly rich interactivity.

REFERENCES

- [1] L. A. Blom and L. T. Chaplin, *The Intimate Act Of Choreography*. University of Pittsburgh Press, 1982, p. 256.
- [2] "Biped," *Merce Cunningham Dance Company*, 1999. [Online]. Available: <http://www.merce.org/company/biped.php>.
- [3] D. Bernini, G. De Michelis, M. Plumari, F. Tisato, and M. Cremaschi, "Towards Augmented Choreography," in *Arts and Technology, Second International Conference, ArtsIT 2011 (In Press)*, 2011.
- [4] G. De Michelis, "The 'Swiss Pattada,'" *interactions*, vol. 10, no. 3, pp. 44-53, 2003.
- [5] A. K. Dey, "Understanding and Using Context," *Personal and Ubiquitous Computing*, vol. 5, no. 1, pp. 4-7, Feb. 2001.
- [6] P. Dourish, "What we talk about when we talk about context," *Personal and Ubiquitous Computing*, vol. 8, no. 1, pp. 19-30, Feb. 2004.
- [7] G. T. Heineman and W. T. Council, Eds., *Component-based software engineering: putting the pieces together*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001.
- [8] A. Hutchinson Guest, *Choreographics: A Comparison of Dance Notation Systems from the Fifteenth Century to the Present*. Routledge, 1989, p. 194.
- [9] A. Hutchinson Guest, *Labanotation: The System of Analyzing and Recording Movement*. Routledge, 2005, p. 504.
- [10] A. K. Brown, *Labanotation for Beginners*. Princeton Book Co Pub, 2008, p. 84.
- [11] "EW Notator 2.1," *Universitat Salzburg*, 2006. [Online]. Available: <http://www.movement-notation.de/>. [Accessed: 2012].
- [12] E. Drewes, "Project Structure," 2008. [Online]. Available: <http://www.movement-notation.org/VisualizingArchives/development/project-structure.html>. [Accessed: 2012].
- [13] "The Benesh Institute & Benesh Movement Notation," *Royal Academy of Dance*, 1997. [Online]. Available: <http://www.rad.org.uk/article.asp?id=114>. [Accessed: 2012].
- [14] "Blue Bird, Male Solo," *Dance Write*, 2011. [Online]. Available: <http://www.dancewrite.com/OpenBenesh/Classical.html>. [Accessed: 2012].
- [15] "Laban Writer," *Ohio State University*, 2006. [Online]. Available: http://dance.osu.edu/3_research_gallery/labam_writer.html. [Accessed: 2012].
- [16] M. D. Brown and S. W. Smoliar, "A graphics editor for labanotation," *ACM SIGGRAPH Computer Graphics*, vol. 10, no. 2, pp. 60-65, Aug. 1976.
- [17] I. Fox, "Documentation Technology for the 21st Century," in *World Dance 2000 Academic Conference*, 2000, pp. 137-142.
- [18] L. Wilke, T. Calvert, R. Ryman, and I. Fox, "From dance notation to human animation: The LabanDancer project: Motion Capture and Retrieval," *Computer Animation and Virtual Worlds*, vol. 16, no. 3-4, pp. 201-211, Jul. 2005.
- [19] C. Griesbeck, "Limelight," 1996. [Online]. Available: <http://user.uni-frankfurt.de/~griesbec/LIMEE.HTML>. [Accessed: 2012].
- [20] J. Hatol, "MovementXML: A Representation of Semantics of Human Movement Based on Labanotation," Simon Fraser University, 2006.
- [21] "Dance Forms 2.0," *Credo Interactive*, 2006. [Online]. Available: <http://charactermotion.com/products/danceforms/index.html>. [Accessed: 26-Jan-2012].

mConduct: Gesture Transmission and Reconstruction for Distributed Performance

Phoebe Bakanas, Joanne Armitage, Joel Balmer, Paul Halpin, Kyle Hudspeth, Kia Ng

ICSRiM – University of Leeds,
School of Electronic and Electrical Engineering, School of Music & School of Computing, Leeds LS2 9JT, UK
mconduct@icsrim.org.uk, www.icsrim.org.uk

Abstract— Conductors play an important role in an ensemble. They are essential to an ensemble functioning as a single coherent unit, being responsible for controlling tempo and dynamics, whilst also channeling the emotional intensity of a piece. Traditional conducting requires visual cues and is challenged by contemporary composers using distributed performance techniques, where the performers' view of the conductor may be obscured.

This paper proposes an interface to enhance coordination in distributed performance. The interface is a non-intrusive system for communicating conductor information via haptic feedback. Data is collected using sensors and mapped into haptic feedback and transmitted via a wireless connection. To verify its effectiveness an experiment is designed where participants attempt to discern the characteristics of beat and time signature from a vibrating motor. The paper concludes with the latest findings and discusses future directions and impacts for the system.

Keywords— vibrotactile; conductor; distributed performance; feedback; gesture; music.

I. INTRODUCTION

The art of conducting has a long and well-established history of using arm and hand gesture to convey musical intent. The visual communications required for conducting can be limiting in contemporary music due to the practice of distributed performance, a compositional technique that facilitates the spatialisation of sound within an ensemble. This paper proposes a tool to remove the visual constraint between the conductor and the performers and thereby allowing composers to write for a greater variety of potential ensemble distributions and give conductors greater control over the synchronicity of a distributed ensemble.

The interface captures and analyses the hand gestures of conducting to offer real-time interactive multi-modal feedback in a number of application contexts including vibrotactile feedback. This paper presents the design and development of the interface involving both multiple hardware sensors and software analysis modules, and discusses its use in vibrotactile feedback for distributed performance.

The potential benefits of such a system range from expanding artistic possibilities of distributed performance

to providing a means to understand the conductor's signals for performers with visual limitation/impairment.

There are four key stages in achieving this goal: (i) capturing conducting gesture data with sensors; (ii) analysing the sensor data captured to detect and extract features such as beat points; (iii) mapping the data and features detected onto tactile feedback control parameters and broadcast the output; (iv) one or more receiver(s) that generate tactile feedback based on the signals received to inform the musician of the conductor's gesture.

II. BACKGROUND

A. Conducting

Conducting is required to coordinate a group of players to perform a set piece with precise and delicate synchronisation and timing. The gesture communication is required to control a range of musical parameters; for example, indicating entries, setting tempo, synchronising the various instruments and parts, shaping the ensemble's sound and more. Conducting has not always been rhythmically focused. Cheironomy, the earliest form of conducting, can be traced back to early Egyptian performances [1]. Before melody was notated in written score, Cheironomy hand signs indicated the melodic shape. Cheironomy was widespread in the ancient world and it endured into medieval times, being used to direct singers of Gregorian chant [2]. However, during the Middle Ages, music became more complex. The melodic line expanded to have more pitches sung upon a single syllable as well as more pitches in the contour of the melody. The potential pitches within the octave also expanded allowing ensembles to have looser tonality. In addition to the development of polyphony, music also became more rhythmically complex. Cheironomy could not keep up and lost its effectiveness and the complexity of the polyphonic music compelled the development of staffed musical notation [2, 3].

During this transitional time, in the Christian church the person giving musical signals held a staff. As music became more complex, the staff was moved up and down to indicate the beat instead of pitch, thus acting as an early form of a baton keeping the ensemble together [4]. During the 17th century, rolled up sheets of paper or smaller sticks were used to signal beat and tempo. The first conductor to utilize a baton can be traced back to 1794,

when the Belgian composer Guillaume-Alexis Paris used a baton while conducting opera in Hamburg, Germany.

As orchestra size increased, conducting became a fixed role in the orchestra. By the early nineteenth century, it became common practice to have a dedicated conductor, an individual who did not play in the ensemble. Also in the 19th century Wagner developed a theory that shaped the role of conductors today [4]. He believed that conductors should not only keep time but also impose their own interpretation of the piece. Modern conducting is built upon the traditional techniques founded by Wagner and other early conductors. However, conducting is still an evolving art form. Not only does modern music necessitate the use of new conducting techniques but also technology is opening doors for new types of music and new interpretations of conducting.

B. Distributed Performance

The spatial dispersion of performers is apparent in any performance whether deliberate or not, due to the inherent positional relationship between both performers and the audience. The earliest recorded example of deliberate spatialisation in Western music dates back to the liturgical practice of *psalmody* in the 9th century [5]. Antiphony was created through separating the choir into two distinct sections [6]. During the 16th century this developed into poly-choric works that were composed for St Mark's Church, Venice to enhance both echo effects and antiphony [7].

The technique was employed by classical composers including Mozart, who composed two notable spatial works; *Serenade No. 6* and (K. 239) and *Serenade No. 8* (K. 286). *Serenade No. 6* was composed for two orchestras, one acting as a distinct solo string quartet. It has been suggested that the two groups would have been placed in opposite corners of a room to create a spatial distinction between them [8]. In *Serenade No. 8*, Mozart wrote for four orchestras each consisting of four strings and two horns. He does not specify the spatial distinction between the performers. However he creates an artificial space through dynamics and development of material. In a 1964 performance of the piece at Symphony Hall, Boston, one orchestra was positioned on the stage, two on the floor and the final orchestra under a canopy. An additional conductor was required to direct this orchestra [9].

During the Classical period and romantic periods, spatialisation was generally applied to a standardised symphony orchestra in the form of notated dynamic differences and echo effects through repetition. Through the late 19th century composers began to specify spatiality, an example of this is Mahler's, *Symphony No. 2* where brass and percussion instruments are offstage and *Symphony No. 8* where offstage brass appear at the end of each movement [6].

Stockhausen identifies location as a musical parameter that can affect the tonal characteristics of a sound alongside pitch, duration, loudness and timbre [10]. Many contemporary composers assign instrumentalists specific positioning in their composition performance notes.

Through specifying sound paths and sound source locations the composer has direct control of the piece's overall acoustic balance.

Spatialisation of sound can be applied in a number of ways in both acoustic and electronic compositions. With the development of electroacoustic music in the 20th century, composers were able to directly control sound and sound trajectories. Arrangements of loudspeakers can be used to form the spatial characteristic of a sound. An example of this is Barret's recent composition *Construction*; involving 22 musicians, live electronics and a 16-channel sound installation [11]. Positioning acoustic performers in a predefined, distributive arrangement eliminates the need for additional equipment such as loudspeakers and microphones. The separation between performers can obscure their view of the conductor leading to logistical difficulties in performance. The composer Henry Brant's observation on spatiality in music reflects this:

"Separated groups are difficult to coordinate - exact rhythmic simultaneities are almost impossible because of the distances between the musicians."

Brant, 1967 [12: 224]

With the proposed system this difficulty will be addressed. The three orchestras in Stockhausen's, *Gruppen fur Drei Orchestra*, could be controlled by one conductor [13]. Thus, the simultaneity of the separate orchestras could be enhanced without affecting the spatial distinction between them. In this application the system approaches the coordination difficulties in distributed performance by offering real-time vibrotactile feedback of a conductor's gestures.

C. Conductor Tracking Systems

A variety of conducting gesture analysis and performance systems have been developed. The most common objective for these systems is conducting electronic instruments or a virtual orchestra. The term 'virtual orchestra' was first introduced into the musical lexicon in the early 1990s by Bianchi and Smith [14]. Bianchi and Smith developed an interactive computer music system that was used in the Kentucky Opera's 1995 production of *Hansel and Gretel*. This marked one of the first uses of technology by a major performing arts organisation.

Another example of an early virtual orchestra is the system created in 1991 by Morita, Hashimoto and Ohteru [15]. Their electronic orchestra responded to a conductor's gestures whose movements were tracked through a Charge-Coupled Device (CCD) camera and a sensor glove. Morita *et al.* categorise the tracked conducting information into two main functions:

- i) Basic that includes notes, pitch, frequency, duration.
- ii) Musical performance expression (Mpx), such as ritardando, sostenuto, dolce.

The basic information is quantifiable and necessary when performing a piece. The Mpx information is subjective and creates the artistic essence of the performance [15]. Ascertaining beat points to indicate tempo is a minimum requirement for this system. Expanding upon this to measure gestural expression is fundamental in creating an authentic reconstruction.

In 1996, the ‘Digital Baton’ [16, 17] was designed as a multipurpose device to control electronic music through traditional conducting parameters such as tempo, dynamic and duration alongside individual notes and details of particular sounds. Gestures are tracked using accelerometers, infrared LED and piezo-resistive strips. A similar array of sensors will be used in this project. Due to advancements in technology, the dimensions and weight of the device will be significantly reduced allowing the conductor greater and more traditional movement.

While many conducting gesture analysis systems focus on controlling a virtual orchestra, in other systems conductor analysis is a research motivation. In 1998, Nakra improved upon the digital baton with the ‘Conductor’s Jacket’. The ‘Conductor’s Jacket’ is a physiological monitoring system built into the clothing of a conductor; it is designed to study conductors’ techniques in their working environments [18]. Whilst Nakra was interested in mapping the conductor’s expressive features to a musical score, the aim of analysis in this system will enhance the conductor’s control of a distributed ensemble.

Other systems aim at conducting pedagogy. Examples include Peng’s 2009 [19] and Bradshaw and Ng’s 2008 [20, 21] conducting analysis systems. These two projects used a Wii-based tracking system to capture the conducting gesture. Bradshaw and Ng explain how their conducting tracking system can be used in different pedagogical scenarios. Their system allows features of conducting movements to be clustered and compared. Additionally, a conductor can also compare recordings of his/her gestures to highlight differences in different contexts or over time. The system also allows conductors to view how consistently they conduct by comparing the tempo between their own beat points to a set tempo on a metronome.

Many other conducting tracking systems have been developed including: Lee *et al.* [22]; Borchers *et al.* [23]; Katayose and Okudaira [24]; Bruegge *et al.* [25]; Nakra *et al.* [26]; Baba *et al.* [27].

D. Reconstruction of Data

Data mapping is used widely in multimedia projects and involves translating data across different domains.

Eacott [28] sonifies tidal speed information through mapping it to produce live notation. Bradshaw and Ng [29] analyse the conducting gesture through sonification and visualisation. Robertson *et al.* [30] designed a beat tracking algorithm for real-time beat visualisation.

Raisamo *et al.* [32] utilised haptic feedback in their experiments to understand emotional communication using *cutaneous saltation*. Similarly, Lemmens *et al.* [33]

designed a vibrotactile jacket with 64 actuators to enhance the emotion and immersion of the cinematic experience. The test subjects wearing the vibrotactile jacket interpreted vibrations as perceived movement that correlated to emotive visual stimuli.

These research examples informed the design of this system. However their focus was to improve the experience for the audience, whereas the proposed system aims to convey gestural signals from the conductor to the performer.

III. DESIGN AND DEVELOPMENT

In this system, data is mapped to convert gestural conducting data into a different domain, specifically vibrotactile.

The design of the system can be broken down into four distinct sections: data capture, analysis, mapping, and reconstruction.

While designing the system, the weight and dimensions of both the baton and the receivers were considered to be important. The baton device has to be both small and non-intrusive in order for the gestures to be performed comfortably and naturally by the conductor. The receiving device that provides feedback to the performers has to be similarly non-intrusive so as to not distract the performers.

The device for gesture capture requires a number of sensors to effectively measure the motion of the conductor’s baton in multiple dimensions. The primary sensors used in the baton are an accelerometer that provides acceleration information in three dimensions, and a gyroscope that provides orientation information. Tests were run that demonstrate the accelerometer and gyroscope sensors accurately measure a beat. Other sensors, such as magnetometers, flex sensors, and distance sensors are currently being tested within specific use cases. The sensors in the baton are connected to an Arduino Uno [34] board that interprets the received information and sends it wirelessly to a computer for analysis.

Wireless transmission of data is essential for this system. In terms of system performance, using wired connections for both the baton and receiving units would be detrimental to the non-intrusivity of the system; having cables running from conductors and performers during a performance would be likely impact their movements. Wired connections also contain many more practical problems in terms of cost and aesthetics; long trailing cables are unsightly, and can become expensive, particularly if a number of receivers are used. For these reasons, several wireless protocols were considered. ZigBee was chosen due to its low power consumption and appropriate range for this application.

Once captured, gestural data is analysed to extract specific features particularly the beat points to identify tempo/rhythmical patterns and mapped onto vibrotactile controls, and broadcast wirelessly.

One or more receiver units receive live broadcast data and translate the data back to the physical world with actuators, thereby reconstructing the gesture data with vibration for the performer to interpret.

The system uses a coin type Eccentric Rotating Mass (ERM) vibration motor. This actuator is has a diameter of 10 mm and thickness of 3.4 mm. The strength of its vibration is controlled by a pulse width modulation (PWM) signal, the duty cycle ratio of which is determined by the data captured from the baton device.

The overall design of the system has adapted the ‘Music via Motion’ (MvM) [31] framework approach that facilitates the trans-domain mapping of data (such as movement) to another domain (e.g. sound and visual) with a modular architecture (see Figure 1).

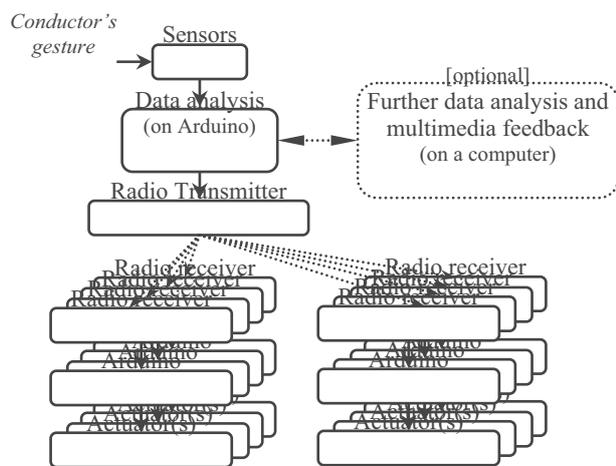


Figure 1. Overall system architecture design.

IV. EXPERIMENTATION AND VALIDATION

A. Experimentation

An experiment has been designed and carried out to assess whether the conductor's beat points and time signature could be understood through vibrotactile feedback. The prototype was used to convey this information to test subjects. The intensity of the vibration feedback was linearly relative to the amount of acceleration applied to the prototype baton. The vibrations were between 200-300 Hz as the skin is most sensitive to haptics within this range [35].

The experiment consisted of a ‘conductor’ and ‘performer’. The conductor beat 30 bars at around 75 beats per minute (bpm) for each test. The performer was blindfolded and the actuator was attached to their lower leg. Optimal positioning of the actuator was assessed through considering the sensitivity of the skin and intrusivity to the instrumentalist. It was also easy and non-invasive to place the actuator on the lower leg.

A total of 10 subjects were tested, all with an understanding of basic music theory and conducting. Each performer participated in two tests. In the first test the

conductor beat time with smooth motion. The performer was instructed to say, “B” whenever they felt a beat. The performer’s vocal reactions were recorded. The test determined how accurately the performer could recognize the beat from the vibrations.

During the second test the conductor beat 10 bars of 2/4, 3/4 and 4/4 time. Each bar was given an accented first beat. The performer had to identify the time signature. The time signature was randomly changed each trial so there was not a pre-disclosed pattern the performer could preempt. The data from this test determined whether participants could recognise a time signature through vibration.

In the first test analysis, the vocal recordings of performers’ reactions were compared with a plot of the acceleration values from the conductor’s beat. A click track was used to align the accelerometer data with the vocal waveforms. A sample of the accelerometer data was displayed in a line graph and scaled to the same range as a waveform of the recorded audio. The alignment of the signals determined the accuracy of the performer’s response to the conductor’s beats.

The second test analysis involved analysing the accuracy of the subjects’ time signature identification. This data was used to assess the accuracy of each trial and determine trends within the inaccurate answers.

B. Results

The experiment showed that beat points and time signatures can be successfully conveyed through vibrotactile feedback. The results from the first test found that subjects could identify the conducted beat. Figure 2 illustrates a sample result with the acceleration data and the audio recording of one of the subjects. While all subjects accurately identified all beat points, it has been observed that trained musicians identified the beat with greater consistency and accuracy. Subjects with less musical background identified beat points in a more sporadic manner. Discrepancies in reaction time can be accounted for by delays in the participants’ reflexes; some subjects anticipated the beat while others waited to feel it before responding. One subject noted that in some instances they could not feel the vibration motor enough to accurately identify a beat point.

Since this experiment, feedback has been improved through non-linear mapping techniques. Improvements in the data mapping create a greater gradient of vibration intensity, enhancing the subject’s ability to distinguish between strong and weak beats.

The second test found the system to be effective at haptically reconstructing time signature, as indicated by the distribution of the subject’s score. The median score for the test subjects was 80%, which means the majority of participants only answered one trial incorrectly. Another positive result is that subjects with musical training and experience playing in large ensembles identified the time signature most quickly. This indicates that trained

musicians, familiar with conducting technique, can understand the gesture through haptics.

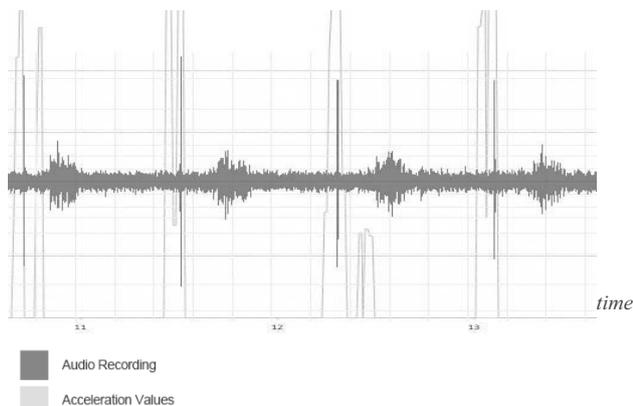


Figure 2. A comparison of the recorded voice audio and the recorded sensor data to measure their timing differences.

V. CONCLUSION AND FUTURE DIRECTION

The proposed system has a unique application context that focuses on the translation of gestural communication to bypass the visual constraint of conducting. It performs a trans-domain mapping between the baton's movement and the vibration intensity of the actuator(s) over a wireless system that allows sensor data to be received simultaneously by multiple devices to reconstruct the gesture signal through vibration.

With an initial evaluation, the system has been found to be effective to allow the conductor's communication with a distributed ensemble in real time through haptic feedback without visual contact.

Further processing and multimodal feedback including visualization is being added to the system. Additional features exploring different vibration patterns and movements with multiple actuators are being tested.

Further qualitative and quantitative evaluations have been planned to ascertain the systems performance. Quantitative validation will involve analysing measurable hardware data such as speed, accuracy and jitter. A final validation of the system will be implementing the full system in a distributed performance. In the performance the musicians will be distributed throughout the concert hall and will not be able to see the conductor. The conductor will use the prototype developed to transmit gesture through the system as vibrations to the performers. Rehearsals combining both the piece and system are underway in Leeds. Feedback from the performers is being used to further optimize the system particularly in relation to the calibration of the intensity and directionality of the vibration patterns.

While the main objective is to use the system prototype to enhance distributed performance, it can also be utilised in other applications. It is said that two thirds of

communication is through non-verbal channels [36]. Gestures provide information about a person's emotions and relationships with others. This research can help further the understanding of gesture communication. The data collected from the sensors can be analysed to explore the relationship between conductor and musicians and their dependence on each other. Then the knowledge about the communication between conductor and musicians can be applied to gesture communication at large. The application of this system will have an important role in human computer interface.

The system can also aid the pedagogy of conducting. Virtual orchestra training programs can use the empirical data on how performers synchronise with a conductor in their mappings of gestural interaction. Recorded conducting gesture can also be preserved for future generations to understand the different interpretation and expression of a specific time period or a conductor.

REFERENCES

- [1] R. Randhofer, "By the rivers of babylon: echoes of the babylonian past in the musical heritage of the iraqi jewish diaspora," *Ethnomusicology Forum*, vol. 13, pp. 21-45, 2004.
- [2] S. Haik-Vantoura. "Chironomy in the ancient world." Internet: <http://www.rakkav.com/biblemusic/pages/chironomy.htm>, Dec. 27 2011 [28 Jan. 2012].
- [3] "Medieval and Renaissance Music." Internet: <http://www.essentialhumanities.net/index.php>, 2011 [Jan. 20 2012].
- [4] "Conducting (music).", Internet: http://www.newworldencyclopedia.org/entry/Conducting_%28music%29, 2011 [Dec. 12 2011].
- [5] P. Kollias, "An Introduction to Henry Brant's Spatial Music through his Work Ice Field", MA Thesis, City University London, 2004.
- [6] J. W. Solomon, "Spatialization in Music: The Analysis and Interpretation of Spatial Gestures," PhD dissertation University of Georgia, 2007.
- [7] K. Stockhausen, "Music in Space," in *die Reihe*, Vol 5, pp67-82, 1959.
- [8] R. E. Rodder, "Serenade No. 6 in D major, K. 239, Serenata Notturna" Internet http://www.kennedy-center.org/calendar/?fuseaction=composition&composition_id=4358 [Feb. 19 2012].
- [9] "A Choice and an Echo," *Time Magazine*, 7 August 1964, available at: <http://www.time.com/time/magazine/article/0,9171,871353,00.html> [Feb 19, 2012].
- [10] M. A. Harley, "Space and spatialization in contemporary music: History and analysis ideas and implementations," unpublished.
- [11] R. Barrett, "Construction." Internet: <http://www.hcmf.co.uk/Richard-Barrett-a-monumental-CONSTRUCTION>, 2011 [Jan. 2012].
- [12] H. Brant, "Space as an essential aspect of musical composition." *Contemporary Composers on Contemporary Music*. E Schwarz, B. Childs, Eds. New York: Holt, Rinehart and Winston, 1967 pp. 221-242.
- [13] K. Michael, Stockhausen: A Biography, Transl. R. Toop, London and Boston: Faber and Faber, 1992, p 80.
- [14] F. Bianchi and D. B. Smith. "Virtual Orchestra." Internet: <http://www.virtualorchestra.com/index.html>, [Dec. 2011].
- [15] H. Morita, S. Hashimoto, and S. Ohteru, "A Computer Music System that Follows a Human Conductor," *Computer*, vol. 24, pp. 44-53, 07/1991 1991.

- [16] T. Marrin. "Possibilities for the Digital Baton as a General Purpose Gestural Interface." In CHI, 1997, pp. 311-312.
- [17] T. Marrin and J. Paradiso, "The Digital Baton: a Versatile Performance Instrument," presented at the International Computer Music Conference, Thessaloniki, Greece, 1997.
- [18] T. M. Nakra, "Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture," Doctor of Philosophy, Philosophy, Massachusetts Institute of Technology, Boston, MA, 2000.
- [19] L. Peng and D. Gerhard, "A wii-based gestural interface for computer-based conducting systems," in *New Interfaces For MusicalExpression*, 2009.
- [20] D. Bradshaw and K. Ng, Motion Capture, Analysis and Feedback to Support Learning Conducting, in *Proceedings of the International Computer Music Conference (ICMC)*, McGill University, Montreal, Quebec, Canada, 16-21 August 2009
- [21] D. Bradshaw and K. Ng, "Analyzing A Conductors Gestures With The Wiimote," in *EVA 2008*, London, UK, 2008.
- [22] E. Lee, T. M. Nakra, and J. Borchers, "You're the conductor: a realistic interactive conducting system for children," in *The 2004 Conference On New Interfaces For Musical Expression Hamamatsu, Japan, 2004*, p. 6.
- [23] J. Borchers, E. Lee, W. Samminger, and M. Mühlhäuser, "Personal orchestra: a real-time audio/video system for interactive conducting " *Multimedia Systems*, vol. 9, pp. 458-465, 2004.
- [24] H. Katayose and K. Okudaira, "iFP A Music Interface Using an Expressive Performance Template," *Entertainment Computing 2004, Lecture Notes of Computer Science*, vol. 3166/2004, pp. 529-540, 2004.
- [25] B. Bruegge, C. Teschner, P. Lachenmaier, E. Fenzl, D. Schmidt, and S. Bierbaum, "Pinocchio: conducting a virtual symphony orchestra," in *The International Conference On Advances In Computer Entertainment Technology*, 2007.
- [26] T. M. Nakra, Y. Ivanov, P. Smaragdis, and C. Ault, "The USB Virtual Maestro: an Interactive Conducting System," presented at the NIME, 2009.
- [27] T. Baba, M. Hashida, and H. Katayose, "'VirtualPhilharmony': A Conducting System with Heuristics of Conducting an Orchestra," in *NIME 2010*, 2010, pp. 263-270.
- [28] J. Eacott, "Flood tide see further: Sonification as a musical performance." in *ICMC*, 2011, pp 69-74.
- [29] D. Bradshaw, K. C. Ng, "Analyzing a conductors gestures with the Wiimote." in *EVA*, 2008, pp. 35-42.
- [30] A. Robertson, A. M. Stark , M. D. Plumbley, "Real-time visual beat tracking using a comb filter matrix." in *ICMC*, 2011, pp. 617-620.
- [31] K. Ng, "Music via motion: Transdomain mapping of motion and sound for interactive performances." in *IEEE*. 92(4), pp. 645-655, Mar. 2004.
- [32] J. Raisamo, R. Raisamo, and V. Surakka, "Evaluating the effect of temporal parameters for vibrotactile saltatory patterns," in *Proceedings of the 2009 ICMI*, 2009, pp. 319-326.
- [33] P. Lemmens, F. Cromptvoets, D. Brokken, et al. "A body-conforming tactile jacket to enrich movie viewing" *World Haptics Conference*, Salt Lake City, UT, USA, 2009
- [34] "Arduino." Internet: <http://www.arduino.cc/>, Jan 2012 [Jan. 2012].
- [35] B. J. P. Mortimer and R. W. C. Gary A. Zets, "Vibrotactile transduction and transducers," *Journal of the Acoustical Society of America*, vol. 121, 2007, p. 2971.
- [36] K. Hogan and R. Stubbs, *Can't get Through 8 Barriers to Communication*. Grenta, LA: Pelican Publishing Company, 2003.

Mobile Augmented Reality for cultural dissemination

Raffaello Brondi⁽¹⁾, Marcello Carrozzino⁽²⁾, Franco Tecchia⁽³⁾, Massimo Bergamasco⁽⁴⁾

(1) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, r.brondi@sssup.it

(2) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, m.carrozzino@sssup.it

(3) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, f.tecchia@sssup.it

(4) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, m.bergamasco@sssup.it

Abstract— This paper presents the development of a Mobile Augmented Reality (MAR) framework, named XMAR, and of a prototype application, built on top of XMAR, and oriented to the cultural heritage field. The MAR framework has been developed to address the needs of outdoor markerless applications to be used in the field of cultural heritage. Furthermore, the framework had to meet the requirements of providing real time registration using inertial sensors and GPS and of efficiently managing non trivial tridimensional environments, transforming a typical smartphone in an augmented reality device. An application called LiTe has been designed and developed on top of the XMAR framework to provide an augmented view of Piazza dei Miracoli, one of the most famous artistic sites in Italy.

Mobile Augmented Reality; Markerless; Cultural Heritage;

I. INTRODUCTION

A huge amount of information about many cultural heritage sites has been extracted over the years by theoretic studies, though all these information is practically invisible to visitors. Augmented Reality (AR) systems can provide a solution to this lack, providing a valuable instruments for cultural heritage applications. They offer for example the possibility to see virtual reconstructions of ancient ruins in the same environment where the original buildings were. In addition, visual information can be combined with texts and multimedia information, providing interactive instruments and changing the learning process. The use of Mobile Augmented Reality (MAR) applications adds the ability to move within the sites while looking at the environment around you as it was when it was built.

Augmented Reality is a branch of the bigger research field of the Mixed Reality, which also includes Virtual Reality (VR), distinguished by the blending of virtual and real objects with the aim of augmenting the human perception of the surrounding environment and developing new metaphors of interaction. The sector of AR, as like as the one of VR, is closely related to the technological and technical progress. The goal of the first AR systems was the augmentation of an environment located in proximity of the user position. The technological barriers resulting from the devices used to implement such a system was the cause of that limit. Recently the improvements of mobile platforms, and in particular of smartphones, defined the new breed

of the Mobile Augmented Reality (MAR) and opened a new thread in this research field. These devices include processors able to make relatively complex calculation and powerful graphic cards, inertial sensors and usually also GPS modules, the opportunity of fast connection via third-generation mobile telecommunications technologies and wireless connectivity in a unique small device. This set of characteristics offers a significant starting point on which it is possible to build MAR systems, which do not have mobility limits caused by the hardware used. Therefore, these systems can aim to provide the users with an augmentation of potentially unbounded environments, allowing him to move inside a large augmented environment.

The main motivation for the development of the presented framework has been the need for a tool for mobile devices capable of providing both a sensor based registration of the virtual environment, and able to handle and render non-trivial three-dimensional environments in real-time. The prototype application LiTe has been developed on top of this framework to verify the real capabilities of the framework. It is a MAR application for the "history browsing" of Piazza dei Miracoli in Pisa. The application resumes a previous research which has produced several multimedia information, included a tridimensional reconstruction of the artistic complex in several historical periods, from the 11st century up to now. The purpose of LiTe has been to provide the multimedia material produced by the previous research work through the use the Augmented Reality. The tridimensional models of the monuments can be seen overlaid on the video taken from the smartphone camera while the user is visiting the square. The user can then learn about the monuments around him through a new interaction paradigm characterized by a deep interactivity.

II. STATE OF THE ART

As part of the enhancement of cultural communication, there are many researches and applications aimed at exploiting, disseminating and making the artistic heritage easily accessible to non-specialist public. The use of new technologies is bringing about a substantial contribution to this field, and in particular the use of AR systems is increasingly proving to be valuable to this purpose.

Among the first researches in the field of Augmented Reality for Cultural Heritage it is noteworthy to mention ARCHEOGUIDE, a project aiming to provide a personalized electronic guide and tour assistant to cultural site visitors. It provides customized Augmented Reality tours and reconstructions of ruined cultural heritage sites in order to help visitors and scientists to better appreciate and enjoy the past glory of these sites. The system hardware is composed of a laptop and a head mounted display which communicate with a remote server containing the multimedia contents[6].

Another recent project with the aim of supplying a culture-oriented information system based on an AR system is iTACITUS. It combines itinerary planning, navigation and rich content on-site information, based upon a dispersed repository of historical and cultural resources. It enables a complete information system and media experience for historical interested travellers. It uses mobile computers to supply an augmented view of the cultural sites and spatial acoustic sounds to provide the user with an acoustic impression about how the place has been before[9]. The same team proposed an AR systems to be installed in museums to augment large wall-filling photographs of the real sites with interactive contextual annotations like 3D reconstructions, images and movies[8].

In the field of museums communication, it is historically important to mention "The Virtual Dig," made in 2003 by the Seattle Art Museum and the university of Washington Human Interface Technology laboratory. This interactive experience is based on an AR system for the presentation and exhibition of artifacts from Sichuan, and allows the visitors to perform some simple tasks, guided by a narrative, replicating the process of excavation and retrieval of archaeological finds. More recently, in 2008, the DNP Museum Lab has developed for the Louvre museum an AR guide [10], based on the Unifeye technology. The subsequent experience evaluation has shown a great public appreciation and the recognition of a substantially usefulness of the device. A subsequent evaluation experiment about the use of AR guides has been presented by CNAM [11].

The previous solutions use powerful and/or customized hardware to achieve their result, therefore they cannot be straight-forwardly implemented on devices with low capabilities such as smartphones. On the other hand, the system MARCH (Mobile Augmented Reality for Cultural Heritage) uses AR to assist visits to prehistoric caves using a smartphone [4]. The application is able to show a possible reconstruction of a painting on the cave walls but it relies on visual markers to perform the registration of the virtual painting with the real environment.

On the contrary, the system presented in this paper, does not make use of markers and can rely just on the hardware available on a commercial smartphone device to perform the registration of the virtual environment.

III. MOBILE AUGMENTED REALITY CHALLENGES

Any AR-based application presents a great deal of challenges that must be solved. There are two main common issues that must be addressed when developing an Augmented Reality system:

- real time interaction of the virtual environment with the user;
- registration of the virtual environment with the real world.

These issues are a direct consequence of the definition of AR as given by Azuma [1], and they have been studied ever since the advent of the first Augmented Reality systems. The impact of these problems on the final result is strictly related to the hardware configuration used and their incidence may vary for different hardware solutions.

Mobile devices, generally defined as handheld display, represents an excellent alternative to other kind of Augmented Reality displays, like wearable head-mounted displays, particularly because they are minimally intrusive, socially acceptable, readily available and highly mobile; moreover, they are usually cheaper and often present in everyday life. The typical use of such displays is to generate images at arm reach and video-through is the preferred paradigm. Integrated video cameras capture live video streams of the environment that are overlaid by graphical augmentations before being displayed. However, these devices, due to their intrinsic structural characteristics, suffer from a number disadvantages that have to be taken into account when developing an AR system:

- Power consumption: they have a limited autonomy.
- Computational power: they cannot run applications requiring heavy computations or being memory-intensive.
- Screen: their small screen results in a restricted FOV.
- Components quality: especially in consumer products, components, such as the camera or the sensors equipped on the mobile, might not suitable for AR.
- Hands-free: as opposed to head-mounted or projection based displays, handhelds force users to always have at least a busy hand.

IV. THE XMAR FRAMEWORK

The XMAR framework, so far available for the Android Operating System, has been designed and developed [16] to allow the easy and rapid development of Augmented Reality applications overlaying 3D co-located information to outdoor unprepared environments. This kind of scenario proposes several issues especially on mobile platforms.

Concerning the registration of virtual environments with the real world, there are many Computer Vision techniques which have shown excellent results in similar contexts, although they are computationally too expensive and to date not suitable for real time usage on mobiles. Thus, given the actual state of the art of smartphone technologies, XMAR bases the registration process on the combination of data given by the GPS and inertial sensors, providing a good compromise between precision of the registration and computational cost.

The real-time rendering of the augmented 3D information takes place through a library purposely developed on top of OpenGL ES 2.0, taking full advantage from the several structures supplied by this API.

A. Registration

The registration is the procedure through which an AR system brings the digital information to be correctly co-located with the view of the real environment. Without this, it is usually difficult to trick the human senses into believing that computer-generated virtual objects co-exist in the same physical space as the real world objects. To address this issue the registration process has to estimate both the position and the orientation of the device in order to place the virtual camera in the same reference system of the real environment.

To estimate the geographical position of the user, the framework uses the GPS sensors now almost ubiquitously available in the last generation of smartphones. The acquired geographic coordinates are defined in the WGS84 coordinate reference system, used as default reference system by almost all the GPS, which defines the user position using the angular values latitude and longitude. These values are converted to the UTM coordinate reference system, which provide distances in meters.

Each measurement contains a confidence value, indicating how accurate that particular measurement is. To reduce repositioning of the virtual camera due to inaccurate locations, an incoming value read from the GPS is used only:

- if it is significantly newer (more than a minute) than the current one, else
- if it is more accurate than the current one, else
- if it is not older neither less accurate than the current

otherwise it is discarded. Furthermore, the framework applies a low-pass filter to these values to reduce the noise of the measurements.

The orientation of the smartphone is obtained as a function of the readings of the magnetometer and accelerometer sensors applying the deterministic attitude estimation algorithm TRIAD [5]. The attitude matrix is then used to evaluate the view vector in the virtual

environment. Like the GPS, magnetometer and accelerometer are not usually sufficiently precise and the signal coming from them is heavily noisy, affecting the evaluation of the virtual camera placement. To reduce the noise effect all the values read are filtered through a low pass filter.

The filtering processes used to reduce the errors of both GPS and inertial sensors make the registration more stable but of course introduce a certain delay in the update of virtual objects. However, since during the use of the application the user is supposed to make relatively slow movements, therefore a small delay is tolerable and is supposed not to be strongly noticeable.

B. Rendering

Although recent smartphones have processors and graphic cards much more powerful of the first MAR systems, the amount of graphic information that need to be handled and their quality has greatly increased. While the first systems limited the augmentation to text labels or trivial geometries, today the expectation for these kind of applications is much greater than before, due both to the increased availability of computing power and to our addiction to a more and more realistic virtual reality. On a smartphone application the management and rendering of an heavy tridimensional scene can still be a serious bottleneck, especially if it has to react to the user stimuli and being updated in real time.

The XMAR framework uses a library, called MVRLib, to manage and render the virtual environments. The library, written in C++ and representing a partial porting of the visual features of the XVR technology [15], interacts with the graphic hardware through OpenGL ES 2.0 and offers a compact object oriented interface to handle the complexity of a 3D environment. To avoid data redundancy, MVRLib uses four different static manager classes which respectively handle shapes, materials, textures and shaders. The virtual environment is hold in a Scene Graph which allows to structure the scene in a bounding volume hierarchy. On top of this data structure the library efficiently implements collision detection [13] and view frustum culling [12] algorithms to respectively provide a method for the selection of 3D objects through ray casting technique, and to boost rendering performance avoiding unnecessary visibility test in the graphic pipeline.

V. LiTe

LiTe is a prototype application developed on top of the XMAR architecture. The application is based on a previous work carried out in June 2001 when, in concurrence with the reopening of the leaning Tower, a large-scale multi-media project was launched on the Cathedral Square of Pisa, universally known as Piazza dei Miracoli.



Figure 1. Screenshot of the web application available at <http://piazza.opapisa.it/3D/>

One of the resulting products of the project was the creation of a web site (available online at the address <http://piazza.opapisa.it/3D/>) which contains a desktop Virtual Reality application enabling the exploration in space and time of the famous historical site. Based on the various surveys performed over the years on the Cathedral Square, the final 3D model was built and georeferenced inside the urban system, depicting the whole site starting from the 11st century up to nowadays [3], identifying six different relevant historical periods virtually reconstructed.

LiTe, starting from this previous work, proposes a new form of interaction allowing to interactively access all the available multimedia material produced in the context of the described project. The main goals of the LiTe application are to:

- provide an augmented view of the site;
- allow the user to travel through different historical ages;
- allow the user to interactively request additional information about what he/she is currently seeing.

The application has been designed to be easily portable. It does not make any assumptions on the virtual environment it has to deal with. Once a georeferenced polygonal mesh has been supplied, it is able to provide an augmentation of the environment. For this reason it is possible to use the application in all those situations concerning outdoor monuments or archaeological sites. The artistic heritage is indeed full of sites, like Piazza dei Miracoli, whose communication can be greatly enhanced by the adoption of such instruments.

The application is structured on three layers. The deepest one displays the images captured by the device camera. The middle layer defines an OpenGL surface where the graphic library draws the virtual environment. The virtual objects are rendered overlaid on the video and are placed coherently with the real environment thanks to the registration process. The uppermost layer

contains a set of widgets and buttons forming the GUI allowing the user to control the application.

A. User interface

When LiTe starts, it requests to the user to enter an internet address from where to download the tridimensional model to display. The application assumes that the specified data defines a 3D environment containing a georeferenced object to be used as an augmented content for the images acquired by the camera, using the position of the virtual camera as described in IV.A.

After loading the virtual environment, the application turns on the smartphone camera and shows the user the video streaming overlaid with the registered virtual environment.

Users can use the provided GUI to switch between the available historical periods (and therefore change the displayed 3D environment), to enable or disable the GPS positioning (a feature that can be useful when the noise of the acquired data is too high, resulting in a disturbing swing of the virtual camera), and to switch between the two implemented augmented modes, which will be detailed in the following section. The application also allows to use the touch screen to obtain information about the displayed virtual object. When a touch is retrieved, the system generates two screen coordinates and queries the XMAR framework to verify if the click has occurred on a region of the screen occupied by a virtual object. The framework performs a pick correlation on that coordinate and, if the test succeeds, it returns a label describing the touched object and, if available, further information registered for that element. The retrieved content is then displayed on the screen.

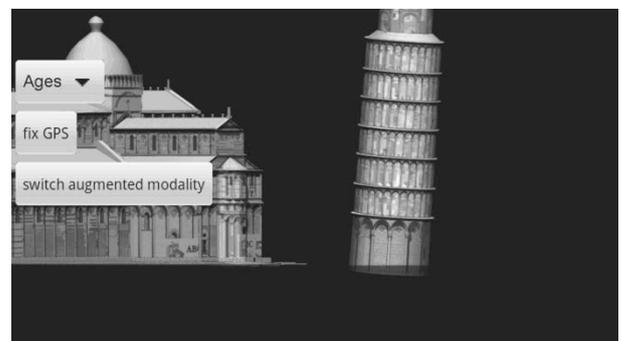


Figure 2. Screenshot of the LiTe application augmented content



Figure 3. Picture of the LiTe application running in Piazza dei Miracoli.

B. Full VS semi augmented view

The tests carried out in the real site show that the registration process still suffers from a lack of precision. In the worst cases, the virtual environment was too unstable due to the fluctuation of the acquired values, and the user immersion experience was damaged.

The precision of the registration process is strictly related to the hardware the smartphone is equipped with, and is also influenced by the surrounding environment. The GPS, for instance, may present precision issues due to the presence of buildings or other entities blocking radio signals from satellites.

To improve its usability, LiTe implements two different paradigms of navigation:

- fully augmented.
- semi augmented.

The first one represents the full augmentation of the square, where the tridimensional objects are shown overlaid on top of the video data coming from the camera. The latter, instead, does not render at all the virtual environment but is used to navigate through the real environment and query the application to obtain information about what is currently observed. The virtual environment indeed, even if not rendered, is still registered and when the user touches the screen to obtain information, the framework performs the pick correlation as previously described and displays the returned information as labels overlaid to the video stream. Thus, the semi augmented solution allows the user to receive only 2D augmented information about the surrounding environment; disabling the rendering of the 3D environment, however, although resulting in a somehow reduced experience, can be useful in case of a poor accuracy of the registration. The accuracy needed to obtain a good correlation between the touch zone and the virtual elements is in fact much less challenging, also thanks to the small dimension of the screen.

VI. RESULTS AND FUTURE WORKS

The application, tested on a Motorola Milestone A583 equipped with Android 2.1 Operating System, was generally very well accepted by test users, despite of noticeable precision issues. The GPS module equipped

on the smartphone presented a poor accuracy and is heavily influenced by the environment. Position values coming from the GPS and inertial sensors are filtered by the architecture to reduce the effect of errors, but the results are not always satisfactory. These limitations are mainly related to the currently available low-price components and should be likely addressed by technology advancements in the immediate future. Smartphones are in fact conceived to be a product for a wide market and have to satisfy many different demands. Price and form factor considerations often lead to use cheaper or smaller components instead of more advanced solutions. As far as the 3D real-time rendering is concerned, the XMAR architecture has proved to be able to manage complex virtual environments. The frame rate obtained while rendering the tridimensional model of Piazza dei Miracoli, composed by about 50K triangles held in a scene graph of 561 different objects, is about 17fps, when using textures, and about 20fps, without textures.

Future work will focus on the improvement of the registration process. The latest smartphones are equipped, in addition to GPS and accelerometers, also with gyroscopes. These sensors allow the creation of more sophisticated filters, such as the Kalman filter[14], which should help to achieve better results in terms of precision and stability. Given the steady increase in computing power in recent smartphones, we will also consider and evaluate computer vision algorithms (such as SIFT tracking, PTAM, etc.) for the registration of virtual environment to test their applicability in this context.

VII. ACKNOWLEDGMENTS

The 3D Website of Piazza dei Miracoli was funded by Opera della Primaziale Pisana and Fondazione Cassa di Risparmio di Pisa, and realized together with VRMedia S.r.l, Andrea Brogi, Scuola Normale Superiore, Liberologico, Soprintendenza ai Beni Culturali di Pisa, Delcaldo WebStudio.

The research on Mobile Augmented Reality was partially carried on in the framework of the St@rt project, funded by Regione Toscana.

REFERENCES

- [1] R.T. Azuma et al. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments*, 6(4):355–385, 1997.
- [2] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, 2001.
- [3] M. Carrozzino A. Brogi F. Tecchia and M. Bergamasco, 2005. The 3D interactive visit to Piazza dei Miracoli, Italy. In *ACE '05: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM Press, New York, NY, USA, 192–195.
- [4] O. Choudary, V. Charvillat, R. Grigoras, and P. Gurdjos. MARCH: mobile augmented reality for cultural heritage. In *Proceedings of the seventeen ACM international conference on Multimedia*, pages 1023–1024. ACM, 2009.
- [5] M.D. Shuster and SD Oh. Three-axis attitude determination from vector observations. *Journal of Guidance and Control*, 4(1):70–77, 1981.

- [6] V. Vlahakis, et al. Archeoguide: first results of an augmented reality, mobile computing system in cultural heritage sites. In Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage, November, pages 28–30, 2001.
- [7] F. Zhou, H.B.L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pages 193–202. IEEE Computer Society, 2008.
- [8] M. Zöllner, J. Keil, H. Wuest, and D. Pletinckx. An Augmented Reality Presentation System for Remote Cultural Heritage Sites. 2009.
- [9] M. Zoellner, D. Stricker, G. Bleser, and Y. Pastarmov. iTACITUS—Novel Interaction and Tracking Paradigms for Mobile AR. In VAST, pages 26–30, 2007.
- [10] T. Miyashita et al., “An Augmented Reality museum guide: Mixed and Augmented Reality,” In Proc. of ISMAR 2008, pp. 103-106
- [11] A. Damala, P. Cubaud, A. Bationo, P. Houlier and I. Marchal. Bridging the gap between the digital and the physical: design and evaluation of a mobile augmented reality guide for the museum visit. Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts, pages 120–127, ACM New York, 2008
- [12] G. Gribb and K. Hartmann. Fast extraction of viewing frustum planes from the world-viewprojection matrix. Online document, 2001.
- [13] S. Gottschalk, M. C. Lin, and D. Manocha. Obbtree: A hierarchical structure for rapid interference detection. In SIGGRAPH '96 Proc., 1996.
- [14] G. Welch and G. Bishop. An introduction to the Kalman filter. University of North Carolina at Chapel Hill, Chapel Hill, NC, 1995.
- [15] F. Tecchia et al., A Flexible Framework for Wide-Spectrum VR Development, in Presence-Teleoperators And Virtual Environments 2010 – 19, 302-312
- [16] R. Brondi, Design and realization of a module for the real-time visualization of augmented 3D environments on mobiles, Master Thesis, Università di Pisa, 2010.



Workshop on Authoritative Metadata Vs User-Generated Content

Chair

Emanuele Bellini, Fondazione Rinascimento Digitale (Italy)

Workshop Committee:

Luigi Briguglio, ASSETS-EU, IT

Martin Doer, FORTH, GR

Mariella Guercio, University of Urbino, IT

Seamus Ross, iSchool, University of Toronto, CA

Carlo Meghini, ISTI, IT

Seth van Hooland, Dublin Core MI and ULB, BE

Yannis Tzitzikas, FORTH, GR

A Database for Israeli Theater Archives (DITA) In The University of Haifa Younes & Soraya Nazarian Library

Sharon Shapira-Glaubach

Media Department, The Younes & Soraya Nazarian Library, University of Haifa, Haifa, Israel
shapira@univ.haifa.ac.il

Abstract— The Database for Israeli Theater Archives (DITA - <http://digitool.haifa.ac.il>) is a database initiated at the University of Haifa Younes & Soraya Nazarian Library in order to digitally preserve, document, and make accessible materials from Israeli theater physical archives. DITA is a "full text" database maintaining objects of various types and formats (text, graphic, sound and video) digitized and cataloged by the library. This paper introduces some of the concepts that guided the library in structuring the database, focusing on two main aspects – the library's "back-stage activity" regarding metadata issues, and its "on-stage activity" regarding the user-oriented-access-design to the digital archives and their contents.

Keywords- *Israeli theater archives; Digital theater archives; Documentation, cataloging and metadata; User-oriented access design, Academic libraries' digital projects; Academic libraries cultural heritage initiatives.*

I. INTRODUCTION

The Database for Israeli Theater Archives (i.e. DITA) is a digital database for Israeli theater archival materials, initiated and maintained at the Digital Media Center (DMC) of the University of Haifa Younes & Soraya Nazarian Library, as part of its Israeli Heritage Preservation activities. The database holds digitized items of various formats from several physical theater archives, as well as bibliographic metadata records for each object.

Being a product originating in the academic environment the database is designed to support the information needs of scholars and students, however, it is widely embraced by the professional theater community, the media, and the general public.

The databases' equally important three main components, which are generated by the library, are: Digital preservation of all materials preserved in the physical archives, documentations & metadata, and user-oriented access design. This paper will focus on the last two.

II. THE BEHIND-THE-SCENES ARENA: METADATA ISSUES

A. Cataloging concepts, resources & tools

One of the strengths of libraries is their professional and technical resources enabling them to organize information. The first part of this "information organizing" has to do with organizing the information "behind the scenes" within metadata records, according to accepted conventions and standards. The second part is organizing the "on stage" user-access to those records, again, according to accepted conventions and standards. One could simply argue that for libraries, organizing means conventions and standards.

Controlled vocabularies, structured syntax and the usage of authority files are the fundamentals to the established control libraries have over the information they handle. In order to benefit from this deep and refined treatment of information we decided to catalog the DITA records as we would catalog a monograph in the library's catalog, referring to every document as a unique unit.

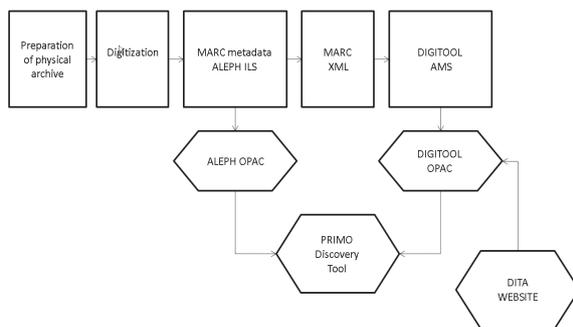
The DITA metadata records are supported by LC Name Authority files, LC Subject Headings Authority files [1], LC Uniform Titles, an in-house developed classification schema, and on the IHP (Index to Hebrew Periodicals) Names & Subject Headings Authority files. The cataloging is performed in the library's ALEPH system and the records are ingested along with the digital objects into the library's digital asset management system DIGITool.

The DITA records are cataloged according to AACR2 cataloging rules, using MARC cataloging standards [2]. MARC does not supply a perfect solution for the metadata elements needed to sufficiently describe digital archival materials, and it sets various limitations to their retrieval. The pros and cons for using MARC in order to describe digitized representations (of music, in this case), were thoroughly discussed by Harriette Hemmasi in her "Why not MARC?" (2002) paper [3]. I find her analysis and observations applicable to the DITA materials as well, especially the ones pointing to the lack of adequate structural and administrative metadata, the weak relationships between

fields describing separate works, and the dependency on format definitions.

However, the decision to use MARC despite its weaknesses for the cataloging of the DITA records was made subject to three considerations: We wanted to benefit from the authoritative references included in our authority files (MARC based), and we wanted the DITA records to be integrated in the library catalog (also MARC based). We try to compensate for potential searches and retrievals difficulties, by formulating contextualized descriptive title statements for each object, and by offering structured accesses within our digital asset management system interface [4], and in the DITA website [5]. The third significant consideration was that other existing standards (such as Dublin Core) had their own disadvantages and weaknesses, and the library had but limited resources which could be devoted to the development of a satisfactory comprehensive alternative.

DITA Creation Process



Another major issue that needed to be addressed was the cataloging and description language. Since the content of DITA is Israeli theater, it reflects a culture active, mostly, in the Hebrew language. Most of the textual documents preserved in the archives are in Hebrew, and therefore the chosen cataloging language is Hebrew. However, every record contains a number of cataloging fields in English and other languages in order to support the accuracy of documentation (the original title of a play, a playwrights' name et cetera) and the retrieval for non Hebrew speakers.

The most demanding component in the DITA project is the cataloging and metadata section. The content described in the metadata records of the DITA objects demands time-consuming original-cataloging, professional metadata specialists, knowledge in the general history of theater and in the history of Israeli theater. The necessity of these qualifications and the massive investment the library is putting in metadata creation, emerges mainly from the need to supply a unique bibliographic metadata record for each digitized item (in opposed to a single metadata record representing

a production as a collection of documents, as was the documenting practice in the physical archives).

B. Content & information resources for cataloging

Technically the cataloging work is done when both the physical and the digitized item are in front of the eyes of the cataloger, thus, relevant details can be drawn from both manifestations.

Content wise, the resource for the most comprehensive and authorized description of a certain production is the production program. The DITA cataloger rely, heavily, on the information supplied by the program (which is, usually, a document that was preserved in the physical archive) to document all the objects relating to a specific production. Complementary information is gathered by the cataloger from all the documents preserved, from research publications and history and theater books, from contacts with the theaters' staffs, and from personal knowledge.

C. Approaching textual documents & non-textual documents

In general, textual materials tend to supply sufficient cataloging details about themselves (the name of an author of a review, the title of an article, the plays' title et cetera). However, in the physical world there is no need to write on each play text placed, for example, in the Hamlet container: "this is the play text of Hamlet and it is placed in the Hamlet container". In the metadata record of the digital manifestation of this play text, this sort of information is necessary in order to contextualize it correctly. Normally, the information gathered from the textual document along with the complementary information and the physical placement of the document in a physical container, enables the cataloger to responsibly describe the item itself, as well as its context and relations to the production in a whole.

The main challenge we encounter is in cataloging the vast amount of non-textual and graphic materials preserved in the archives. Those materials are "wordless" in nature, and they are not always identified by the theaters' archive beyond what the eye can see (1 stage model; 3 posters; 2 audio cassettes). Sometimes all we have is a bunch of pictures in the container, a film role with no title, a sketch; yet, we need to supply them with information in the shape of words in order to document them and to retrieve them later on. But what are the right words? What information should we describe and document? Where can we find this information? Just to name a few obstacles, what should we do with materials such as:

- Unidentified photographs - who are in them? Who was the photographer? What is the scene seen in the photograph? Is this a rehearsal or a live performance?
- Costume designs - who is the designer? what character does this sketch represent? Who played this character?

- Sound recordings - what were the cues for the sound effects? What musical instruments were played? Who were the music performers?

How do we relate to those materials' content beyond stating that "this is a picture preserved in the Hamlet container"? And, the heaviest question of all, words in what language should we use? Does it really matter to the costume designer looking for an inspiration for the design of Juliet's dress, to the actor looking for visuals of Arlecchinos' gestures, or to the musician looking for sound effects used for Brecht's plays, that the play production was in Hebrew? Not necessarily. The audio and the visual information stored in DITA can be of interest to people with no knowledge of Hebrew what so ever – but how would they retrieve it if the metadata language is Hebrew and they don't speak/write Hebrew?

D. Dissecting the "production" to its single units

Most of the physical archives included in our project were organized according to productions – each production had a physical container that held all the materials together, and the container served as the context creator: If a document is placed in a certain container it means that it belongs to this specific production. In a physical archive, this is probably the most convenient way to organize the materials in order to browse through "everything" preserved: the users enter the archive and all the containers are in front of them; the differentiation between a brochure, a playbill, a newspaper clip, a play text, a picture, an audio role and a video cassette is apparent to the eye, and the users just have to browse through the container and pull out whatever they are looking for. This is, in no way, the reality of the digital archive where everything is stored files, and physical differences such as physical format, size, carrier or the tactility of the materials do not exist at all. Furthermore, a glimpse at the item and its content (which is the natural act we perform in order to select relevant items in the physical environment), is not available to us in the digital environment we are working in without conducting a search using a written language. In this case the structure and the quality of the metadata record are crucial since this obligatory textual search is performed not on the object itself (as is in looking at the physical object), but on the textual representation of the object (i.e. the metadata record). Therefore, from the users point of view, the retrieval (and therefore the existence) of the digital object is absolutely dependent on the metadata attached to it.

In order to supply relevant, accurate and targeted metadata for each item and to be able to contextualize it correctly, we needed to dissect the "production" to its single-item components. In practice that means referring to the characteristics of each item (of the original physical item as well as to its digital manifestation), and mostly, to its specific and unique content and context. The kind of information and data needed in order to catalog the archival materials "per-document", was not

available to the library from the documentation supplied by the physical archives, because their documentation related to the whole production, not to a single document within a production. Thus, the library needed to create original records (based on the catalogers' identification and research) by describing each digitized object in a unique bibliographic record of its own.

E. Extent of the cataloging & metadata creation work

Every "production" is cataloged to the full extent of the information we could establish about it: title (original and translated), playwright, director, all the actors, designers, technicians and other professionals involved in the production, dates, all the characters in the play, and many other details. A single metadata record in DITA serves as master-record for the "production" - subordinating all the other documents since it contains most of the names and functions that were involved in the production. All the subordinate documents relate (to some extent) to the information contained in this master-record and necessary additions and details can be easily added per document. Since an average scope of the archives we manage is 350 productions of about 150 documents each, and an average metadata record in DITA contains 50-100 cataloging fields, the cataloging & metadata creation work is quite extensive.

F. Questions of necessity

One could argue whether the massive investment the library puts in producing such deep, controlled and researched metadata is really necessary. I believe it is, for two main reasons:

The first, which was already mentioned above, is the dependency on the metadata records for retrieval purposes. The richer the metadata, the better chances the records will be retrieved when searched upon. This became a clear and practical conclusion once we studied the enormous gap between retrievals resulting from searches performed over rich metadata records vs. retrievals resulting from basic metadata records in the library's catalog.

The second reason is more profound, and it has to do with the slightly ambiguous concept of quality. When we started working on the DITA project, we needed to define what would be its desirable quality. Since a quality of a product is, basically, a standard of the product's characteristics defined (first) by its creator (and only later judged by the users), we had to decide what those characteristics would be and how to achieve them.

As mentioned in the introduction, DITA is a product of the academic world, and as such, we believed it should share its standards and speak its language. We believed that an academic library product should be based on researched, examined, confirmed and authoritative information because when users

(especially scholars and students) turn to academic information resources they, justifiably, expect to find researched, examined, confirmed and authoritative information. We wanted to supply that to the best of our abilities, and strict metadata treatment was our professional behind-the-scenes tool to fulfill this goal.

III. THE ON-STAGE ARENA: USER-ORIENTED ACCESS DESIGN

A. Concept

DITA is a digital archive comprised of several independent theater archives. It now holds four institutions' archives: the Haifa Theater, the Cameri Theater of Tel-Aviv, The Acco Festival of Israeli Alternative Theater and the Mario Kotliar Archive. We wanted to present a profile of each archive on its own as well as being a part of a master archive where all participating parties are represented.

Implemented in the metadata records, the information is now in the state of promising potential. It is not, yet, apparent to the users, for it is stored (along with the digital objects themselves) inside the asset management system. In order for it to be discovered, the users must perform a search. But how would they know what to expect to find? And how would they formulate the most accurate query that would result in the most accurate and comprehensive retrieval?

In order to expose the users to the content stored in DITA and to suggest an optimal access to it, we offer a user-oriented access design of browsing by various types of contents. Alongside the option of a free text search, the pre-defined access points allow the users to skip the forming of search queries and to follow a guided path.

B. Structuring the scope view & contents of a single archive

The user-oriented designed access option is structured to reveal each archive with its own productions and contents.

After the digital objects (item + metadata record) are ingested into the asset management system, we gather them in labeled collections reflecting their context. Every document is connected to and collected with similar contents and formats (such as "reviews", "photographs", "play texts", et cetera). These collections are created by "planting" a defined optimal search query (based on structured metadata fields) behind their collections' labels. The queries are activated every time a collection is clicked upon by the users. Every such collection is subordinated to its specific production (again, collected by pre-defined queries); every production is subordinated to its specific theater archive (again, pre-defined by structured queries). Thus, a structured, comprehensive digital archive is made available with its general contents logically structured,

exposed, and easily accessed. This structure imitates, in fact, the primary physical archive where users enter a room and all the productions' containers are in front of them, full of apparent characteristics. However, in the digital archive environment we needed to verbalize those characteristics in order to systematically label the collections. This is, in no way, a trivial task, for it involves identifying and defining the common characteristics of theater archival materials and to embed them, in advance, in the metadata records.

In addition to the **per-production** access, we also offer direct access to another five major domains:

- **Chronological** access points (all the productions opened in a specific year and in a decade)
- Name index of **playwrights** (access to all the productions based upon their plays);
- Name index of **directors** (access to all the productions directed by a specific director);
- Name index of **actors** (access to all the productions an actor participated in); and
- **Character-in-a-play** index (access to every character in a specific play and all the archival materials relating to it - photos, prop lists, et cetera).

Other features are represented according to the specific theater activities (such as main stage productions, small stage productions, original plays, outdoor activities, et cetera).

C. Structuring the Master-archive

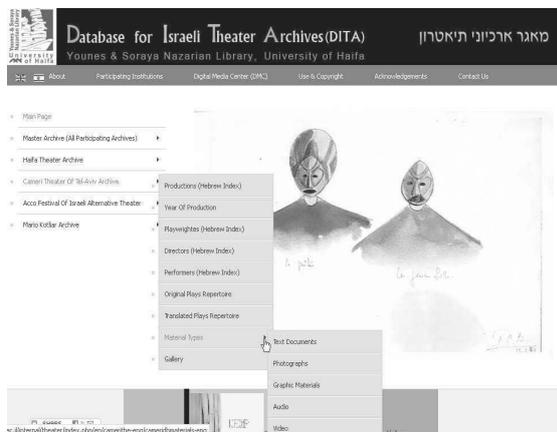
In addition to the per archive direct access we wanted to emphasize the connections between activities taking place in the individual institutions.

Our intention by structuring a master-archive was to create a dynamic map of the theatrical activity taking place in Israel's cultural arena as a whole. Therefore, all the productions master-records (discussed in the cataloging section of this paper) are connected to the master-archive. There, a direct access is offered to chronological and name indexes where one could find who directed or played in which theater and when, how many productions of a certain play were presented over the years, and so on. The master archive maps major features of theater research over a wide corpus of productions and institutions: it offers a chronological view over the events, it offers a view of the choices of repertoire, and it offers a view on the people involved and their professional activities.

IV. THE BEHIND-THE-SCENES AND ON-STAGE ARENAS COMBINED

At this point, it is almost trivial to state that the ability to create these customized archives and sophisticated collections in the user-oriented access design is dependent upon the controlled and structured

information that was embedded in the metadata records. Thus, the effort we put in the data-seeds we plant is maximized and bears potential fruits for the users.



The DITA web site

V. EPILOGUE

Now the process is complete, and both protagonists are well off - the digital items are treated accurately, responsibly and extensively, and the users are served with tailored access to a wide corpus of relevant materials.

The library, as always, is (just) the middleman.

ACKNOWLEDGMENTS

The Haifa Theater; The Cameri Theater of Tel-Aviv; The Acco Festival of Alternative Israeli Theater; The

Mario Kotliar Archive; The Buchsweiler family; Dr. Dorit Yerushalmi of the University of Haifa Theater Department; Prof. Yossi Ziegler, Academic head of the Younes & Soraya Nazarian Library; Mrs. Pnina Erez, director of the Younes & Soraya Nazarian Library; The Library's Media Department staff – especially Ms. Keren Barner, Ms. Tahel Gover and Mr. Amir Dekel; Mrs. Lia Koffler, former head of the Library's Media Department; The Library's Information Technology Unit – especially Mrs. Marina Kalach and Mr. Boaz Dotan.

I would like to thank Mrs. Ora Zehavi, head of the Library's Media Department, for her invaluable personal and professional support.

REFERENCES

- [1] Library of Congress. Authorities web site <http://authorities.loc.gov/>
- [2] Library of Congress. MARC standards web site <http://www.loc.gov/marc/>
- [3] H. Hemmasi, "Why Not MARC?", 2002, IRCAM – Centre Pompidou <http://variations2.indiana.edu/pdf/hemmasi-ismir2002.pdf>
- [4] Theater Archives Collections in the Digital Media Center (DMC), Younes & Soraya Nazarian Library, University of Haifa <http://digitool.haifa.ac.il>
- [5] Database for Israeli Theater Archivs (DITA), The Younes & Soraya Nazarian Library, University of Haifa web site <http://lib.haifa.ac.il/collections/dita>

Designing and implementing a user-generated content service for Europeana

Nicola Aloia, Cesare Concordia, Carlo Meghini.

Istituto di Scienza e Tecnologie dell'Informazione, ISTI-CNR, Pisa, Italy,
(nicola.aloia,cesare.concordia,carlo.meghini)@isti.cnr.it

Abstract— The paper presents an overview of the user generated content service that the ASSETS Best Practice Network has designed, implemented and evaluated with the user for Europeana, the European digital library. The service will allow Europeana users to contribute to the contents of the digital library in several different ways, such as uploading simple media objects along with their descriptions, annotating existing objects, or enriching existing descriptions. The user and the system requirements are outlined first, and used to derive the basic principles underlying the service. A conceptual model of the entities required for the realization of the service and a general sketch of the system architecture are also given, and used to illustrate the basic workflow of some important operations.

Keywords— component; user-generated content; digital library; Europeana; UGC; EDM; ESE.

I. INTRODUCTION

In the 2011-2015 Strategic Plan, Europeana [1] announces User Engagement to be one of the strategic tracks by which the organization will deliver value. By the term 'Engage' Europeana refers to cultivating new ways for end user to participate in their cultural heritage. The Europeana network comprises communities of archivists, curators and librarians who show a growing interest in exploring new methods of access and dialogue. Europeana intends to enhance the user experience and offer services that allow users to interact and participate.

User-generated-Content (UGC) is one aspect of this renewed way of participating. Information about cultural heritage exists outside the heritage institutions; artifacts, written sources and memories of individuals complement collections held in institutions. UGC services are designed to provide users with means to support and interpret content. They will be involved in storytelling, curating of virtual exhibitions, reviews and even the creation of new collections. Greater participation will increase users' interest and loyalty. Europeana is therefore devoting increasing resources to initiatives that bring out the value of the contribution those users can make.

In response to these needs, the ASSETS [2] Consortium has included the support of user-generated content amongst the services it is going to develop for Europeana. ASSETS is a two-year Best Practice Network co-funded by the CIP PSP Programme to improve the accessibility and usability of Europeana. The ASSETS Consortium comprises 24 partners,

including institutions from ten different European countries and Japan, which are active in the field of cultural heritage and digital libraries.

Rather than focusing on a specific set of UGC applications, ASSETS developed a general purpose, back end component that aims at supporting any UGC service Europeana will want to offer to its users. To this end, the ASSETS back end component implements an Application Programming Interface (API) for creating, storing and manipulating UGC Units of Work, and for submitting these Units of Work to Europeana, in the form of Europeana Submission Information Packages (SIPs). Final users will interact with their Units of Work through client interfaces, which will hide the unnecessary technical details and complexities of the back end to them, providing them with the level of representation that is most suitable for the specific UGC task at hand. Indeed, it is expected that every UGC task will be supported by a different final user interface. But this will have no impact on Europeana, since every different front end will talk to Europeana through the same API.

The API relieve future UGC applications from implementing any server side functionality; they will have to code only the client side, connecting the user world to the API. At the same time, the API move away from Europeana the technical interoperability problems that would arise upon integrating into its database the possibly different objects coming from future UGC applications. The service rely on the Europeana Data Model (being developed by the Europeana version 1.0 project [3]) in order to tackle the more serious semantic interoperability problems.

The definition of the conceptual model underlying the UGC API is the most difficult challenge that the ASSETS UGC team has faced. The proposed model attempts to provide the optimal balance between simplicity, so to be quickly learned and easily coded against by the future UGC service developers, and generality, so to satisfy the needs of any possible future UGC service. This conceptual model has been defined during the first year of the ASSETS project, based on an analysis of the different types of requirements that are in place. The model has been subsequently used to define the UGC API, which has been implemented and demonstrated at the first year project review, in June 2011. On top of this API, a few, initial UGC tasks will be implemented through specific front ends. These tasks will allow Europeana users to contribute to the contents

of the digital library in several different ways, such as uploading new objects along with simple descriptions, annotating existing objects, or enriching existing descriptions.

The paper presents an overview of the UGC service. The user and the system requirements are outlined first, and used to derive the basic principles underlying the service. The conceptual model required for the definition of the API and a general sketch of the system architecture are also given, and used to illustrate the basic workflow of some important operations.

II. REQUIREMENTS

A. User Requirements

User requirements can take different forms:

- First of all, the submission of objects to a repository. Through a user interface (typically a website), users upload digital objects of various formats (images, audio and video). These must be objects that they actually hold rights over (for instance, by creating them). A minimal set of metadata will have to be provided in order to support the interpretation, discovery and management of the object. The user requires to be free of choosing which metadata format to use, but the system must propose a default one.
- Secondly, metadata enrichment. Users contribute factual metadata to an object, such as location, date, names, or tags. The object can be created by the user, but also by another user; the object may also enrich existing content in Europeana.
- Thirdly, annotations: users are contributing their views, comments, opinions to an object
- And finally, there is contextualization: through storytelling or creating virtual exhibitions and galleries and possibly adding narratives to them, users are combining existing objects into a new context (without changing the objects and metadata itself).

Before publishing user generated content, moderation may be added as an intermediate step in the process. Authorized users review the UGC and decide to accept and publish it. In most cases, this includes a feedback loop to the user who originally contributed the data.

B. User Interfaces

Europeana audiences include academic researchers with a high level of language and computer skills but also people who are hardly familiar with foreign languages or using the internet. Typical target groups are secondary school students and their teachers as well as cultural explorers and travelers. While possessing intermediate to good knowledge of foreign languages and online search, these groups generally expect services to be easy and intuitive. At the same time, they want to understand what happens with their contribution and

who keeps control over their content. Secondary school students are generally known for focusing on images rather than digesting long texts.

User Interfaces should therefore preferably be simple, straightforward and visual. Additionally, clear information must be provided about rights regarding the content.

C. System Requirements

The back-end component has to comply with the Europeana architecture, which is based on an Open Source policy. Europeana has also defined a set of guidelines regarding the coding, the testing and the deployment of the components that make up its architecture. These guidelines are publicly accessible on the wiki of the EuropeanaLabs [4] and are been followed by the ASSETS UGC development team.

One important requirement concerns the management of multimedia content. As outlined in the previous section, Europeana expects users to contribute media files of various formats including text, images, videos and the like through its UGC service. However, Europeana was not initially meant to deal with (i.e., collect, store and make accessible) media files. With UGC, this has to change, of course. Moreover, the change has to happen gradually, by evolving the current architecture in a stepwise manner, so as not to compromise the operation of the existing components. In order to meet this requirement, ASSETS provides a simple media object repository, and deploy it on a separate server, the ASSETS server, to be later integrated into the Europeana architecture. More details on the media object repository are given in the architecture Section.

III. THE CONCEPTUAL MODEL OF THE UGC SERVICE

In order to meet the user and system requirements, the ASSETS team designed a UGC service based on the concepts outlined below and presented in Fig.1 as a UML class diagram. In the diagram, boxes represent classes and arcs represent associations. Every association is bi-directional, but only the name in one direction is given for readability purposes. The name given refers to the association going from the class closest to the name to the other class. The reader can derive the name of the association in the opposite direction by following any notational convention.

From the UGC server point of view, at any point in time there exists a set of users of the UGC service. Each user is in fact a role, identified by an id and a password, behind which a whole community may actually operate. Each user has its own Workspace (WS) on the UGC server. A WS is simply a container of the objects that the associated user needs to perform UGC tasks.

The creation of a single UGC object may take a long time and span several sessions of work. In between one of these sessions and the next, the partial results achieved so far have to be persisted, in order not to be lost and to be resumed at the beginning of the next session. The

concept of “partial” UGC is captured by the notion of Unit of Work (UoW). The UoWs of a user are maintained in the user’s WS.

A single UoW contains objects, identified by URIs, and their accompanying descriptions. The objects in a UoW can be of two kinds:

- Existing Europeana objects, that the user has included in the UoW in order to link them to new objects (see below) as values of some property, or in order to enrich them with new descriptions. Existing Europeana objects can be retrieved for inclusion in a UoW via a query issued to Europeana.
- Newly created objects, which are called UGC objects. These objects are original contributions to Europeana, and can be of three kinds:
 - i. digital objects having an associated media file with the content;
 - ii. digital objects for which no media file is available;
 - iii. non-digital objects.

consisting of a property and a value. Different attributions can have the same property with a different value. A value can be itself an object, or a literal or another resource, external to the digital library.

When a UoW is ready to be submitted to Europeana, the user can do so by using an operation that transforms the UoW into a well-formed Submission Information Package (SIP) and places a message signalling the existence of the SIP into the Outbox. Each user WS is endowed with an Outbox. Europeana retrieves messages from Outboxes in order to harvest the corresponding SIPs.

As already mentioned, users can issue queries to Europeana in order to retrieve objects. Each query returns a result, in the form of a message stored in a special area of the user WS called the Inbox. For each retrieved object, the result of a query contains a subset of metadata of the ESE schema of the object, a URI to the full set of metadata and a URI to the digital object on the content provider’s web site, if any.

Each user WS is endowed with an Inbox. Messages in the Inbox are of two kinds: query results and notifications that communicate the result of submissions. These latter notifications may be positive and report the successful ingestion of a SIP, or may be negative and report the reasons why a certain SIP could not be

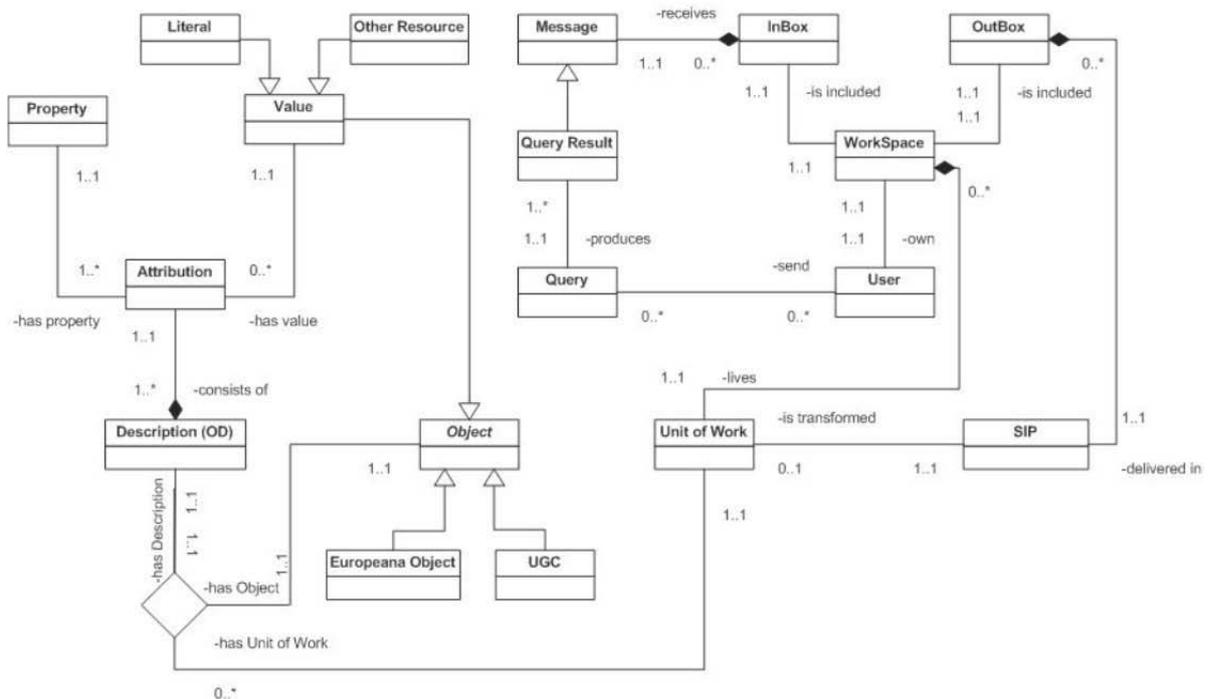


Figure 1 UML class diagram of the UGC service conceptual model

Every object in a UoW has an associated description. A description represents a metadata record of the object and is modeled as a set of attributions, each attribution

ingested. In the latter case, the rejected SIP can be retrieved and re-transformed into a UoW so to allow the user to perform the necessary repairing actions.

It is important to notice that these concepts define a general-purpose schema, whose machinery need not be used by every UGC application. For instance, a simple UGC task that takes place in a single session, such as an image upload, may be implemented by directly building the corresponding SIP, so by-passing the UoW stage. On the other hand, another UGC application may decide to publish a finished UoW to a community of users in order to perform a socially oriented form of mediation before submitting the UoW to Europeana. These decisions will be taken by the client side of the applications, relying on appropriate shortcuts offered by the UGC API.

IV. ARCHITECTURE

For the purposes of developing, testing and evaluating with users the UGC functionality, the UGC server will be deployed on the ASSETS Server. After successful evaluation, the Server will be moved into the Europeana production server and its functionality will be made available to the Europeana users. Fig. 2 shows the architecture of the UGC service. The three main components of this architecture are:

1. *Europeana Server.* This is the server implementing the Europeana search functionality. It provides an API to search for content in the digital library, using optional filters on metadata fields. Europeana query API implements the Open Search directives (<http://www.opensearch.org/Home>). The Europeana server also provides an application

2. *ASSETS UGC Server.* In the context of UGC, the ASSETS UGC server provides functionality to communicate with Europeana and with the UGC client. The communication with the Europeana server is by invoking the Query API module (namely OpenSearch API). The ASSETS UGC server manages the workspaces of the users, and provides an API to manipulate the UoWs in the workspace. The UGC server provides API as REST Web services and is independent from any specific UGC client. The User WS contains an Inbox, an Outbox and the set of Units of Work that the user is currently playing with. In addition, the UGC Server maintains the ASSETS Media Object Repository (AMOR), logically partitioned into two sub-repositories: the repository storing the SIPs not yet harvested by Europeana, and the repository storing the SIPs already harvested by Europeana but not yet ingested by it. On the former repository, AMOR implements OAI-PMH functionality to allow Europeana to harvest SIPs.

3. *UGC Client.* Is a browser-based GUI supporting the user in a specific content generation tasks. The UGC Client interacts with the ASSETS Server via REST web services provided by the UGC Server module

V. THE ASSETS UGC DEMONSTRATORS

To verify the functionality of the UGC services, two demonstrators have been implemented. The first one is a simple HTML page with examples of use of the API

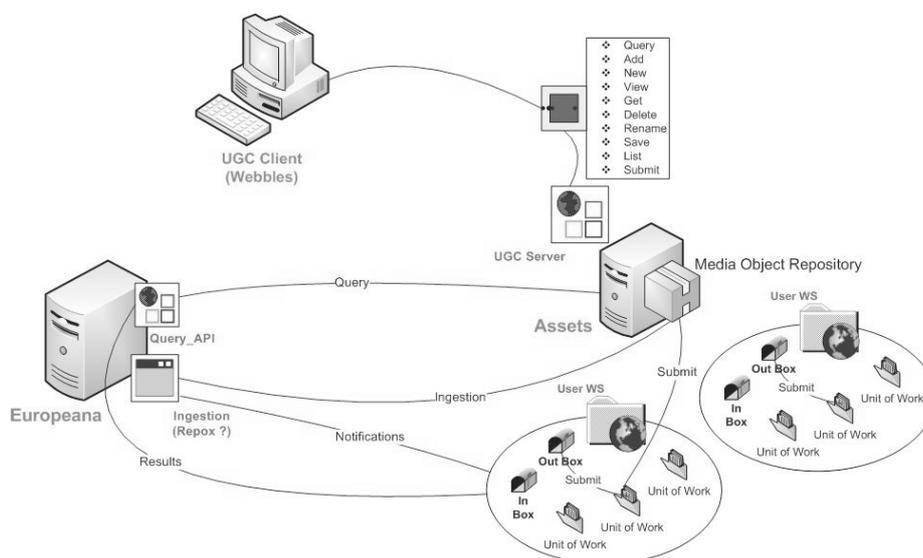


Figure 2 UGC service architecture

for harvesting of data (e.g. Repox, is a GUI based tool for managing these activities).

provided by the UGC service. This demo is useful for developers who want to implement GUI using HTML and is available at [5]. The second consists of a GUI that

uses the "webble" technology developed at the MEME Media Lab and is available at [6]. In Fig. 3 a screenshot of the last demonstrator is shown. On the left side of the GUI there is a tabbed pan having the following sections:

- *Uow* -- Shows the list of existing UOWs (possibly empty) on the Assets server. The user can create a new UOW by dragging objects from "Europeana search" and/or from "User Object" Tab.
- *Search Europeana* -- The user can search Europeana objects by querying in Google style (also an advanced search using Boolean operators among formatted fields condition is supported by the backend). The user will get a set of objects that can be dragged in the yellow area of the GUI, to create UOW.
- *User Object* -- The user can upload an object from his workstation, or can get an object from the web (by providing the object URI). The user can drag these objects in the yellow area of the GUI, to create a UOW.
- *Property* -- The user will see on the left side the list of the ESE properties that he can assign to the objects that are in the yellow area of the GUI. Complex UGC object can be created by using ESE properties like "Has part"/"Is partOf"/"Is VersionOf/etc.

VI. CONCLUSIONS AND OUTLOOKS

The main concepts and architectural features of the user-generated content service have been illustrated. The service has been implemented by the ASSETS Best Practice Network and will be evaluated within the lifetime of the project.

The UGC service developed by ASSETS is based on a general-purpose back end, which is meant to relieve Europeana from dealing with the specificities of the possibly very many UGC tasks that may be offered to users. At the same time, the back end relieves developers of UGC tasks from implementing the server side of their applications.

ACKNOWLEDGMENT

This work has been partially supported by the PSP Best Practice Network ASSETS (CIP-ICT PSP-2009-3, Grant Agreement n. 250527).

REFERENCES.

- [1] www.europeana.eu.
- [2] www.assets4europeana.eu
- [3] version1.europeana.eu/web/europeana-project/
- [4] europeanalabs.eu/wiki/.
- [5] <http://146.48.83.12:8983/assets/communityservices-ugc/>.
- [6] <http://www.meme.hokudai.ac.jp/WebbleWorld/WebbleWorldIndex.html>.

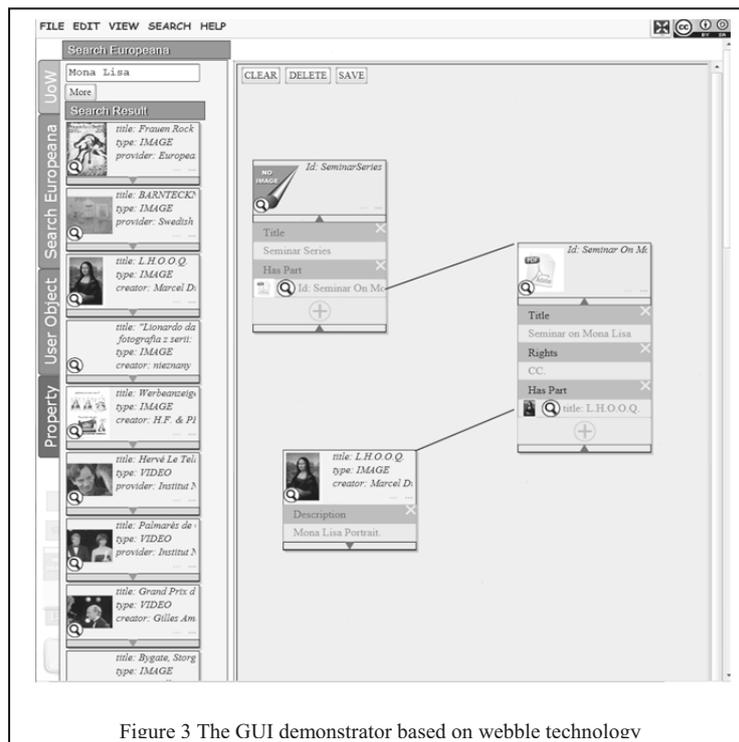


Figure 3 The GUI demonstrator based on webble technology

The Open Data Semantics and the (re)use of Open Information in Cultural Heritage

Achille Felicetti⁽¹⁾, Andrea D'Andrea⁽²⁾, Franco Niccolucci⁽³⁾

(1) PIN S.c.R.L., Università degli Studi di Firenze, Prato, Italy, achille.felicetti@pin.unifi.it,

(2) Università degli Studi di Napoli "L'Orientale", Italy, dandrea@unior.it

(3) PIN S.c.R.L., Università degli Studi di Firenze, Prato, Italy, niccolucci@unifi.it

Abstract — This paper describes the different issues and challenges IT and CH experts have to face when trying to put their information in Linked Data format on the web, issues that mainly concern the creation, use and reuse of such information. We first of all report a survey of the various attempt already done in this direction by different institutions by reporting examples of important CH datasets available. We afterward try to describe the various ways of publishing information in a semantic format carried out by institutions and project involved. Finally we will summarize the results of our activity in this field to illustrate the different techniques we used to create standard information accessible on the Web.

Keywords-component; Linked Open Data; Semantic Web; Mapping; Ontologies; Standard; RDF

I. INTRODUCTION

The creation and use of online CH digital information is becoming widespread in many areas and comprises nowadays a wide range of applications covering every aspect of the encoding work.

In the past years the systematic production and diffusion of CH digital content did not set in so quickly and was in most cases originated from specific institutions and areas whereas today the problem is inverted, raw data are making their way on the web through a number of repositories offering online access to the users.

The Web becomes more and more open and linking data is easier. In a very short time we have passed from scarcity of digital information to abundance and this has created new challenges in designing and implementing new ways to use and reuse valuable data and to find suitable ways to manage the large amounts of information available today.

The best-known example in the field of Cultural Heritage is Europeana, which currently stores metadata concerning more than 10 million digital objects. However, what such a wide-range collection gains in coverage, it may lose in focus, serving research communities less well than the public at large [9].

While we are witnessing the creation and sharing of cultural contents, in some cases accessible online without restrictions, a more widespread liberalization process seems to be obstructed by problems concerning copyright, as institutions and research centers are reluctant to release their data freely before it can be

published in paper format according to the traditional canons of research in the Humanities.

In this paper we will try to describe the different issues and challenges to face when dealing with the complexity of Linked Data, issues that concern mainly the creation and use of such information. Consequently we will report examples of important CH datasets available and describe the various ways of publishing information in a semantic format carried out by the various institutions and project involved. Finally we will summarize the results of our activity in this field to illustrate the different techniques we used to create standard information accessible on the Web.

II. ONLINE INFORMATION AND ITS REUSE

A. User Generated Content

A huge part of the online information nowadays is “user generated information”. The wide public is not only (and no longer) a passive spectator of what is going on online as it was at the beginning of the Internet era, but has become also a creator of information without being limited to being a consumer. The possibilities offered by modern tools and technologies are potentially infinite: from social networks to forums, from blogs to galleries and other online collaboration platforms, the only limit is the fantasy. Content is not only textual, but may include user generate images and videos. Additionally the new mobile phones, very sophisticated objects, are often equipped with GPS and similar tools able to geolocate the user everywhere, in every moment of his digital life and to add geographic information on the created content.

What is possible to do nowadays by gathering and putting together the huge amount of user generated content available online seems to have no limit: the Internet is full of aggregators able to gather information from social networks, blogs and other similar sites and to present it in a plethora of different forms by combining and reusing existing information according to the needs, either by presenting it on a map or on top of a framework of augmented reality.

Unfortunately user generated content is not Aladdin's Lamp, able to satisfy every desire. There are fields of application where this kind of content loses its validity and becomes totally useless. In professional contexts, e.g. in the Cultural Heritage field, opinions and reviews provided by amateurs are of scarce or no importance

compared with the required quality of data provided by experts. In fields such as these, the availability of scientific information becomes paramount. Even if, on the other hand, when views and opinions expressed in the evaluation of shows and exhibitions or other forms of cultural heritage communication, thoughts and convictions formulated by enthusiasts and/or amateurs are crucial when it comes to measuring the level of satisfaction of the initiatives, thus contributing to a process of analysis, verification and possible rethinking of the chosen forms of dissemination.

B. Legacy Information

Since it is still not possible to locate valuable content within the mess of the user generated content, the only place where we can (still) find scientific information for Cultural Heritage is legacy archives. Traditional data derived from museum databases, archaeological excavation recordings, restoration reports, lab analysis *et similia* remain the only trustworthy sources until the advent of more sophisticated tools such as, for instance, ad hoc social platforms where experts will be able to exchange their knowledge in a social-style environment.

But legacy archives also present a set of technical challenges to confront. The first concerns the different retrieval systems and metadata grammars used to describe and index data, sometimes in a such a specific way that it is totally impossible to make it work in a different environment. Not one of these metadata systems will scale up to analyse all the information on the Web.

The second relevant challenge to face is information itself. The huge variety of formats (very often proprietary formats) used to index data is a big obstacle to integration and seriously needs to be overcome. Proprietary data formats have always created barriers to the integration of digital information. The war of the "closed formats" is far from over and at present many closed formats are still considered *de facto* standards, although they frustrate any chance of efficiently sharing and exchanging data.

To create a uniform conceptual layer, information should be extracted from databases, HTML pages, descriptive texts and metadata tags and put in a standard format. These operations are made easy today by the constant activity of the W3 Consortium in defining new standards for web information encoding, such as RDF and SPARQL, by the use of ontologies and by the power of the available mapping, exporting and encoding tools, able to physically or logically extract and re-encode information from legacy archives in a semi-automatic way [13].

C. Newly Created Information

Information created afresh is easier to manage than one obtained as legacy, especially when the tools used for its creation are ready to encode it in a suitable format.

The treatment of new information is quite straightforward when it concerns new records created in

databases already mapped and/or converted. Since the mapping and conversion operations are driven by a schema mapping, each time a new record is created it can be automatically made available on the web by just updating/re-exporting the new database content according to that schema.

The 3D-COFORM project has devoted a lot of attention to the problem of the creation of new compliant metadata, focusing especially on the challenge of the production of valuable provenance metadata able to document each phase of acquisition and creation of digital (3D) objects [1]. The tools developed by the project are already able to produce standard encoded metadata suitable to be shared as Linked Data and to be reused and shared in different contexts.

III. 3. AVAILABLE CH REPOSITORIES

We present here, as an example, a survey of some of the most important national institutions who are moving towards an interoperable world by trying to put their datasets in a standard format. We have chosen examples from Italy, France and England.

It is not a coincidence that some of the most important national institutions are showing more attention to the world of Open Data, whereas in academic circles the researchers' interest seems to concentrate more on the forms of publication and circulation of the online digital archives, also in view of the new types of research from which the scientific community benefits. The central offices of the national Ministries of Culture are responsible for launching an open-data policy mainly for those databases which document the actual existence of cultural heritage (files, photos, drawings, etc).

A. The ICCD CulturalItalia Portal

CulturalItalia is a portal promoted and managed by the Ministry of Cultural Heritage and Activities (Italy). Its repository hosts the digital resources from several Museums, Public and Private Cultural Institutions and part of the data from the Istituto Centrale per il Catalogo e la Documentazione (ICCD) catalogue. It contains about 2.000.000 records.

The digital resources related to the national CH can be consulted through the "Metadata Index", an ordered metadata set with a tree structure classified according to a thesaurus (PICO 4.2) [2] [23].

The index metadata is structured according to the Dublin Core Culture profile, a qualified Dublin Core subset.

ICCD and the Central Office for Documentation of the Ministry of Culture Heritage has developed many tools and services to support all stages of production and use of information regarding CH assets. The latest generation of tools provides access to data through web interface and services in order to facilitate the cataloguing activities of the organizations involved in

CH Conservation and the exchange of standard information through the OAI-PMH protocol.

B. *The Getty Research Institute*

The Getty Provenance Index Databases provide instant access to data from primary sources relevant to the history of collecting [13]. The Getty repository also hosts a very important and rich thesaurus, standard encoded and accessible online in a Linked Data-like modality [3]. The index contains nearly one million records of western European works of art from the late 16th to the early 20th century, including:

- Archival Inventories: it contains 5,200 inventories and more than 260,000 individual records about artworks existing in a specific collection at the time the inventory was made.
- Sales Catalogues: it contains 8,400 catalogues and more than 740,000 individual records of works of art for sale and can be used to trace the sales and ownership history for a work of art.
- Dealer Stock Books: maintained by galleries and art dealers, the database contains 15 stock books and more than 43,700 individual records.
- Payments to Artists: it contains 1,000 recorded payments to artists made in Rome between 1576 and 1711.
- Public Collections: descriptions and provenances of paintings held by public institutions.

C. 3.3. *English Heritage*

PastScape is an online resource from English Heritage's National Monuments Record, which provides information about many of England's ancient and historical sites, buildings and monuments. The information within PastScape is taken directly from the NMR's national historic environment database that contains nearly 400,000 records on the archaeology and buildings of England and its territorial waters [19]. The records are referred to the Thesaurus of Monument Types [17].

Images of England is another English Heritage database containing a photographic record of England's 370,000 listed buildings. All images are stored in jpeg.

All the lists of heritage assets are being unified into the National List for English Heritage [10].

D. 3.4. *The European Heritage Network (HEREIN)*

The database on heritage policies in Europe (HEREIN) provides a multilingual overview of the heritage policies pursued in European countries. The new version of the system, HEREIN 3, focuses on an integrated approach between Archaeological Heritage and Landscape Heritage, which still in many countries falls under different policies.

IV. 4. LINKED OPEN DATA AND THE SEMANTIC WEB

A. *Meaning of Data and Semantics*

Exporting and encoding information is not enough. To reach a full interoperability there is another necessary step to follow to make the information truly reusable in different contexts: semantics. Semantics can be seen mainly as a way to establish a direct relationship between digital data and real objects in the real world. Only when every element is described with the words of the real world, integration will naturally follow, as we live in a relational reality, and the connection of distributed data across the web will be possible.

Some preliminary considerations are necessary when dealing with the use and reuse of online information: first of all we, should observe that usually there is more than one way to describe a given resource. This very important consideration remains true not only for digital information, but for every aspect of communication. In a digital world this essentially means that potentially there are an infinite number of ways (schemas) used worldwide by data creators to encode their information.

Another consideration concerns the meaning of information itself, since every data or metadata is strongly context dependent. This means that in some way the context must be preserved every time we reuse, transfer or exchange data, to avoid the loss of meaning of valuable information [15].

B. *The URIs mechanism*

Linked Data provides the necessary tools to overcome the issues mentioned above. The main and most important one is the URI mechanism, used to identify and interconnect things over the web to guarantee a reuse of information in unexpected ways. The "unexpected reuse" is the value added by the web and represents the most powerful feature provided by the LOD approach.

But the URI mechanism also carries some risks. In order to avoid different institutions generating competing URIs for the same object, each object (or set of objects) should have one preferred authority that assigns the URI for the object. The URI authority for the object must be known to all interested parties or be easy to discover.

Since we are dealing with legacy information, the most natural candidate for the URI authority for an object is the institution that curates the object and its metadata, which is the only institution that can absolutely determine that two different object URIs actually describe the same thing. In conclusion, the museum or institution holding the information is the one that must take control of how their physical holdings are to be identified on the Web.

V. PUBLISHING LINKED OPEN DATA

As we can see from the above list of institutions and archives, the information is manifold as manifold is the way it is stored in the various archives. To capture

information and put it in Linked Data format it is necessary to make distinctions according to the different ways and places the information is stored in and, based on that, to decide the techniques to apply for its standardization and publication on the Web [21] [22].

A. *Static RDF*

A lot of legacy information is still represented in formats such as CSV, Microsoft Excel or text documents. Another typical case regards information stored in relational databases with scarce or no access from the web, i.e. databases having no web interface to be used for the publication of the data.

The simplest way to publish Linked Data for this kind of information is to generate static RDF files by using mapping or annotations and upload it on a semantic server. This method is required also when the RDF files are created manually from scratch.

B. *Views on Existing Databases*

Another typical case is when data is stored in a relational database exposed on the internet. In this case, information can be left on it and, instead of creating new RDF to be placed somewhere, it is more favourable just to publish a Linked Data view on the existing database using a tool for serving Linked Data views on relational databases, for instance D2R Server [7].

The mechanism is very easy: everything is based on a declarative mapping between the schemata of the database and the target RDF terms. Based on this mapping, D2R Server creates the Linked Data view on a legacy database providing a SPARQL endpoint for it.

C. *Wrappers around Applications and APIs*

A certain type of information already exists on the web but in a different form than Linked Data. This information is usually accessible via web APIs providing various query and retrieval interfaces, but not using generic data browsers. Usually they return results in a number of different formats such as XML, OAI-PMH, JSON etc.

This kind of information can be accessed and converted in Linked Data using a set of wrappers able to interface the APIs and to rewrite the client's request into a request against the underlying API, using once gain a mapping framework. The results of the API request can be turned into RDF and sent back to the client.

VI. OUR EXPERIENCE

For more than a decade we have tried to face the various problems related to standardization and interoperability in the Cultural Heritage field. Under the framework of various projects (EPOCH [8], COINS [5], 3D-COFORM) we have designed and experimented many solutions to encode information and tested different kinds of approaches on different kind of archives.

From this experience we've learned that there are many solutions for publishing Linked Open Data on the web, all depending on the nature of the original information.

A. *Working With URIs*

One of the main issues we tried to solve is the definition of a mechanism for the automatic creation of valuable and unambiguous URIs to identify the various resources. In order to avoid different institutions generating competing URIs for the same object, each object (or set of objects) should have one preferred authority that assigns the URI for the object. The URI authority for the object must be known to all interested parties or be easy to discover.

Under the framework of the 3D-COFORM project a recommendation on Linked Open Data was created in association with the CIDOC-ICOM consortium, the recommendation which aims to provide a practical approach for museums to tackle this issue is motivated by the fact that museums should take control of how their physical holdings and identify them on the Web [4]

The created URIs should have a form that enables any museum to provide a Linked Open Data service that resolves to the associated description of that object.

The document describes in detail all the necessary operations each institution should perform to create valuable URIs.

B. *Managing Static RDF*

We have tested the possibility to create static RDF from various sources under the 3D-COFORM framework by developing a set of tools for annotations and mapping able to export and create RDF content from the various data sources described above. The AnnoMAD annotation tool, in particular, is very useful when dealing with textual or semi-structured information [11] [18]. A mapping and exporting tool has also been developed as part of the 3D-COFORM toolset to deal with information stored in non-exposed databases.

The tool gives the user the possibility to create matching schemas to describe the matching between legacy database structures and Linked Data semantics to assist the other tool, the Legacy Database Extractor, in the operations of semi-automatic encoding of the legacy data starting from a schema representing the DB structure.

As a case study we have used the archives of the Victoria & Albert museum from which we have extracted knowledge concerning various objects, put in RDF and stored in the 3D-COFORM semantic repository so it may be integrated and exchanged (see Figure 1) [26]. All the extracted knowledge has been transformed into Linked Open Data format by following the recommendations provided by the CIDOC-ICOM consortium on Linked Open Data (see Figure

2).

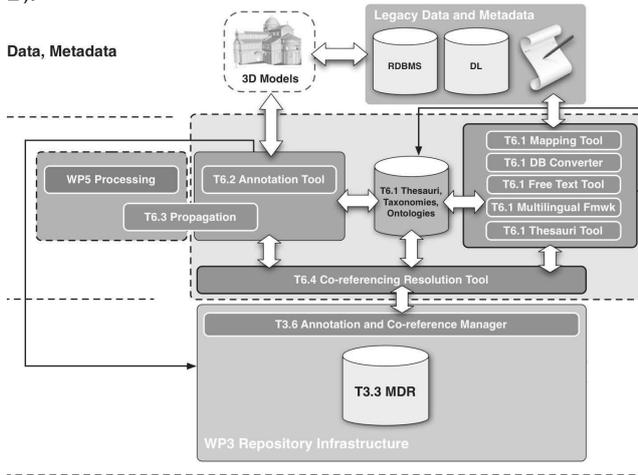


Figure 1. The 3D-COFORM Workflow and the Repository Infrastructure

```

<E21.Person
rdf:about="http://www.vam.ac.uk/res/A13665">
  <L4F.has_preferred_label>
    Smith, Albert
  </L4F.has_preferred_label>
  <L53F.is_uniquely_identified_by>
    A13665
  </L53F.is_uniquely_identified_by>
  <L62F.in_the_role_of>
    <E55.Type rdf:about="http://3dc.org/maker"/>
  </L62F.in_the_role_of>
</E21.Person>

```

Figure 2. Example of RDF fragment generated from the V&A Museum information.

The mapping tool developed for 3D-COFORM does not necessarily need to be used in combination with the exporting tool to generate static RDF, but it can also be employed to create mapping files to be used, for instance, by tools like D2R-Server for the on-the-fly view of relational databases (see below).

C. Creating Views on Existing Databases

We have already tested the D2R Server features to expose archaeological databases in a semantic format and noticed that there is no lack in performances and quality of the information provided. The mapping tool we have developed for 3D-COFORM could also be used to create the mapping for the D2R Server and other similar tools for the on-the-fly view of relational databases.

We are now going to test more complex tools able to perform a wide set of operations, like the OpenLink Virtuoso semantic database [20], specifically designed to publish relational database as Linked Data and Triplify [25], a small plug-in for Web applications able to capture the semantic structures encoded in relational databases by turning database content published as RDF into Linked Data [12].

D. Application APIs Wrappers

We are experiencing this method of serving data and the mapping and rewriting problems related to it while dealing with the information distributed by the Italian Ministry of Cultural Heritage in OAI-PMH format. The Culturaitalia portal has implemented a sophisticated mechanism to provide CH metadata towards Europeana, which in turn is in charge of aggregating and redistributing the collected information in a semantic format.

By interfacing the Culturaitalia engines and using ad hoc mapping scripts, is also possible to create a wrapper around the OAI-PMH API, to collect and reuse the PICO and DC information stored on the portal under a Linked Data format. This will give access to a large set of valuable information since the portal is collecting metadata provided by a large number of Italian institutions.

VII. CONCLUSIONS

In the course of time, the availability of sophisticated hardware and software tools has allowed a greater digitization of cultural resources. The development of recording systems with proprietary programs or languages has marked an important step in the acquisition of a new awareness in data management and flow. The subsequent evolution of Web 2.0 greatly improved accessibility to digital archives, although the technological approaches of the time did not go beyond simple passive consultation of the data. Often, without in-depth knowledge of the data encoding forms, query operations were hindered by objective difficulties in reading and comprehending the available information. To overcome this limitation, therefore facilitating technical and semantic interoperability, these past years have been characterized by the development of standard metadata systems to which *harvesting* protocols have been added. In parallel we've witnessed the development of a new system for encoding archives based on rendering explicit implicit knowledge. An uncontrolled development of ontologies, i.e. formalized and reusable knowledge based on entity, property and relationships, was followed by a more recent phase dedicated to the realignment, or mapping, of the different ontologies created in the meantime for CH. This "race" for integration was in some instances disastrous and above all not entirely measure up to the efforts. We are still far away from fully exploiting the resources available on the web. There is an awareness that the conceptual tools are now widely shared by the community of users, but a strategy capable of making a critical mass of the enormous potentialities of online data is missing. Can Linked Open Data represent a new model for integration and data interoperability, acting as a glue not only for the data, but also for metadata and dominion ontologies? Once integration is achieved, at what level will it be represented: at a high level of specialization, expressiveness and semantic richness, or only at the level of generic, summary and incomplete concepts? Without a doubt, we can already count on many benefits:

1. LOD exceed the narrow-minded logic of ontologies, viewed as the specialized formalization of knowledge, applicable exclusively to a circumscribed sector of knowledge. The integration (or aggregation) of different taxonomies and ontologies allows to place the relationships between foundational and dominion ontologies, characterized by a progressive level of semantic richness and therefore of coarseness, in an innovative way. Until now interoperability could only be obtained by exploiting the poorest level of relationships with a substantial lack of true integration.
2. Modeling knowledge through forms of LOD publication is not limited to the kind of research executable in a single archive. On the contrary, overcoming the barriers of a limited circle of competences benefits the growth of a truly integrated community in which the final exploitation of the data is not necessarily considered as a basic prerequisite of the system.
3. Lastly, LOD could represent an important goal within that process of open and transparent communication inaugurated with the introduction of Web 2.0 and in which the user plays a role which is no longer exclusively passive. This perhaps represents the most interesting aspect, beyond the technical issues pertaining to the implementation and publication of the data and of single archives in the form of RDF triplets. Thanks to LOD we may be able to unveil a new form of Humanities, more collaborative, open and truly “cumulative” without having to wait for the whims of those who would rather keep their data hidden away in basements, data often acquired through public funding.

VIII. REFERENCES

- [1] 3D-COFORM Project: <http://www.3d-coform.eu/>
- [2] AMATURO M., BERARDI E., BUONAZIA I., LATTANZI M., MASCI M., MERLITTI D., T. T. S.: Mapping dei metadati tra le schede ICCD e PICO Application Profile per il Portale della CulturalItalia. ICCD, 2007. 2.
- [3] CARLISLE P.: HEREIN Thesaurus Editing Software. Thesaurus Guidance Manual, March 29, 2011. 3
- [4] CIDOC-ICOM Recommendation on Linked Open Data for Museums: http://www.collectionslink.org.uk/attachments/article/1082/LoD_For_Museums%20v1.6.pdf
- [5] COINS Project: <http://www.coins-project.eu>.
- [6] CulturalItalia Portal: <http://www.culturaitalia.it>.
- [7] D2R Server. Publishing Relational Databases on the Semantic Web. <http://www4.wiwiw.fu-berlin.de/~bizer/d2r-server/>, July 2011.
- [8] EIDE O., FELICETTI A., ORE C., D'ANDREA A., HOLMEN J.: Encoding Cultural Heritage Information for the Semantic Web. Procedures for Data Integration through CIDOC- CRM Mapping. In Proceedings of the EPOCH Conference on Open Digital Cultural Heritage Systems '08 (2008), Arnold D., Niccolucci F., Pletinckx D., Gool L. V., (Eds.).
- [9] Europeana – European Digital Library, Museum, Archive: <http://www.europeana.eu>.
- [10] English Heritage. <http://www.imagesofengland.org.uk/>.
- [11] FELICETTI A.: MAD: Managing Archaeological Data. In The e-volution of Information and Communication Technology in Cultural Heritage (2006), Ioannides M., Arnold D., Niccolucci F., Mania K., (Eds.), Archaeolingua, Budapest, pp. 124–131.
- [12] FELICETTI, A., MARA, H.: Semantic maps and digital islands: semantic web technologies for the future of cultural heritage digital libraries. In proceedings of the Semantic Interoperability in the European Digital Library workshop (ESWC 2008), pages 51-62, June 2008.
- [13] Getty Provenance Index Databases. <http://piprod.getty.edu/starweb/pi/servlet.starweb?path=pi/pi.web>, July 2011.
- [14] ISAAC A., VAN DER MEIJ L., SCHLOBACH S., WANG S.: An Empirical Study of Instance Based Ontology Matching. In Proceedings of ISWC '07 (2007), pp. 253–266.
- [15] Linked Data - Connect Distributed Data across the Web: <http://linkeddata.org/>
- [16] MARTINO M. D., ALBERTONI R.: A multilingual/multicultural semantic-based approach to improve Data Sharing in an SDI for Nature Conservation. International Journal of Spatial Data Infrastructures Research 6 (2011), 206–226.
- [17] National Monuments Record. <http://www.english-heritage.org.uk/professional/archives-and-collections/nmr/>, July 2011.
- [18] NICCOLUCCI F., FELICETTI A., SAMAES M., HERMON S., NYS K.: Ontologies and Semantic Tools For The Management Of Full-Text Archaeological Documentation. Assessments From The Hala Sultan Tekke Case-Study. In Fusion of Cultures. Proceedings of CAA2010 (2010), Contreras F., Melero F. J., (Eds.), Archaeolingua, Budapest.
- [19] NMR Monument Type Thesaurus. http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=1, July 2011.
- [20] OpenLink Virtuoso, Universal Server Platform for Data Integration: <http://virtuoso.openlinksw.com/>
- [21] PLINI P., DI FRANCO S., DE SANTIS D., URICCHIO V. F., DE CARLO D., D'ARPA S., DE MARTINO M., ALBERTONI R.: A Joint Initiative to Support the Semantic Interoperability within the GIIDA Project. In Proceedings of Gi4DM '10 (2010).
- [22] SIGNORE, ORESTE: Representing Knowledge in the Semantic Web - Open Culture: accessing and sharing Knowledge - scholarly production and education in the digital age (<http://openculture.org/milan-2005>) - June 27–29, 2005 - Milano, Italy.
- [23] Thesaurus PICO 4.2, CulturalItalia Portal. http://www.culturaitalia.it/pico/thesaurus/4.2/thesaurus_4.2.0.skos.xml. July 2011
- [24] TORDAI A., OMELAYENKO B., SCHREIBER G.: Thesaurus and Metadata Alignment for a Semantic ECulture Application. In Proceedings of K-CAP '07 (2007), pp. 199–200.
- [25] Triplify: <http://triplify.org>
- [26] XUEMING PAN, PHILIPP BECKMANN, SVEN HAVEMANN, KATERINA TZOMPANAKI, MARTIN DOERR, DIETER W. FELLNER: A Distributed Object Repository for Cultural Heritage. In Proceedings of VAST'2010. pp.105-114.

SKOSifying User Generated Content

Alessandra Donnini⁽¹⁾, Andrea Ciapetti⁽¹⁾, Maria De Vizia Guerriero⁽²⁾, Matteo Lorenzini⁽¹⁾, Maria Emilia Masci⁽¹⁾, Davide Merlitti⁽¹⁾, Fabio Piro⁽¹⁾, Oreste Signore⁽²⁾

(1) EtcWare, Roma, Italy, mail@etcware.it

(2) CNR-ISTI, Pisa, Italy, oreste.signore@isti.cnr.it

Abstract— The increasing social approach to information sharing causes expectations of growth of the information assets and stimulates contributions to the creation of new information and knowledge. User Generated Content (UGC) is lacking a uniform and agreed structure and tagging system (metadata), which makes uneasy to share knowledge on the web. The challenge is to combine the advantages and flexibility of Web 2.0 and folksonomies with a rigorous approach inherent to Semantic Web and ontologies. The LABC (LABoratories for Culture) project is aimed to develop an innovative multimedia web 3.0 platform which enables *capturing, sharing, and enjoying data on tangible and intangible cultural and environmental heritage*. The platform is designed to collect resources from a broad community of users, and to reformulate such contents into scientifically structured information using Semantic Web tools and technologies. The platform allows for the creation of various specialized environments (*Laboratories*), where communities of different types of users, with appropriate rights, can contribute information. CIDOC CRM core ontology has been chosen as reference model, and SKOS standard has been adopted to codify, export and exchange the terminology. A specific component (SKOSware) manages vocabularies and thesauri encoded using SKOS and allows for consultation and handling of vocabularies and thesauri, as well as validation by domain experts.

Keywords - Web 2.0, Semantic Web, Ontologies, SKOS, Thesauri, CIDOC CRM

I. INTRODUCTION

The so called Web 2.0 and the consequent social approach to information sharing is producing large expectations of growth of the information assets and is stimulating contributions to the creation of new information and knowledge [5]. However we have to consider that in several communities the resulting information must be *reliable*, which implies a sort of control and/or validation, contrasting with the freedom characterizing social networks. This issue is even more relevant in contexts like Cultural Heritage domain, which shows, by its own nature, plenty of *interdisciplinary connections*. In fact the communities interested in this domain are composed by people with very different interests, competences and skills, while a

community is normally intended as a social group of organisms sharing an environment, normally with *shared* interests. A problem to be solved on the User Generated Content (UGC) is the lack of a uniform and agreed structure and tagging system (metadata), which makes uneasy to share knowledge on the web, while since many years several communities are converging towards the idea that unified metadata can be a key for combining the large amount of information and data stored on many different specialized sites.

In the present paper we will briefly discuss the characteristics and the differences among the Social Web, the Semantic Web and the traditional core metadata approach. Moreover we will point out that a main issue is to combine the advantages and flexibility of Web 2.0 and folksonomies with the rigorous approach inherent to Semantic Web and ontologies. Finally, we will explain how the LabC project designed and implemented an environment and a technical platform based on the synergy of these different approaches, using as test-bed data on tangible and intangible cultural and environmental heritage.

II. FOLKSONOMY, CORE METADATA AND ONTOLOGY

The main characteristic of Web 2.0 is its *social* aspect, as users provide content and add value. The collaborative tagging, also referred as *social indexing*, *social classification* or *folksonomy*, is one of the most appealing activities performed by users, essential to combine and search for information. Within folksonomies, users add metadata to contents, freely selecting the terms, eventually looking to the tags used by other users. Subsequently, contents with the same tag can be retrieved and linked. This is the power of a folksonomy: flexible, based on user preferred terms, the result and at same time the trigger to social participation. Adding a term to a folksonomy has an almost null cost, while the effect is immediately perceivable, in contrast with the difficulties and long term discussions in building or modifying authority files and controlled vocabularies. Folksonomies exhibit the interesting characteristic that they are built around users' preferred terms; therefore the vocabulary adopted by users emerges. However, we can't ignore an important drawback: they are essentially *terms oriented*. This is the origin of a certain level of *ambiguity*, as users can apply

the same tag to different concepts. At the opposite end of the spectrum, the lack of synonym control can lead to different tags being used for the same meaning. In addition, folksonomies seem designed to support tags made by a *single word*, ignoring spaces and phrases, and are *case insensitive*.

The Web must be seen as a Universal Information Space, navigable, with a mapping from URI (Uniform Resource Identifier) to resources. The Semantic Web is an extension of the current Web, where computers can access structured collections of information and set of inference rules that they can use for automated reasoning. The foundation of Semantic Web is the Resource Description Framework (RDF) based upon a model for representing named properties and property values. The basic RDF model consists of three object types: Resources, Properties and Statements. All things being described by RDF expressions are called *Resources*, and are always named by URIs plus optional anchor ids. A *Property* is a specific aspect, characteristic, attribute, or relation used to describe a resource. Each property is identified by a name, and takes some values. A specific resource together with a named property plus the value of that property for that resource is an RDF *Statement*. These three individual parts of a statement (sometimes addressed as s-p-o statement) are called, respectively, the *subject*, the *predicate*, and the *object*. The object of a statement (i.e., the property value) can be another resource or it can be a literal. A more complete description of Semantic Web and its technologies is out of scope of this paper, and we will not go into details. Details can be found in the vast literature about this topic [1]. A building block of Semantic Web is the *ontology*, intended as an *explicit, formal and shared* representation of the concepts and their relationships. An ontology contains the vocabulary and the rules to infer new knowledge.

From this definition it appears evident that without a human intervention **a folksonomy will never become ontology**, but will just remain a flat representation of the used vocabulary.

An important step towards the interoperability has been the definition of the Dublin Core Metadata Element Set (DCMES) [8]. As clearly explained in [2] Dublin Core is often presented as a modern form of *cataloguing card*, a *set* of elements (and now qualifiers) that describe resources in a complete *package*. Sometimes it is proposed as an *exchange format* for sharing records among multiple collections. A founding principle is that "every element is optional and repeatable". The fifteen elements of the Dublin Core Element Set are the defining feature of Dublin Core as a language, and correspond to fifteen broadly defined *properties* of resources that are generally useful for searching across repositories in multiple domains. Dublin Core is in effect a class of statements of the pattern "*Resource has property X*," where "resource" is the *implied* subject; followed by an implied verb ("has"); followed by one of fifteen properties from the Dublin Core element set; followed

by a property value, an appropriate literal such as a person's name, a date, some words, or a URI.

Among the advantages of Dublin Core we must note the wide acceptance of this standard, which, in turn, is reasonably simple and general. There is a plenty of projects adopting DCMES to achieve interoperability among different documents. However, there are some critical points to consider. First of all, the fifteen elements, sometimes, risk being too generic, and can't represent some typical and complex cases: hence the "qualified" DCMES was born. The second issue is that DC remains an *item-centric* approach and, even more important, the approach by itself is unsuitable to support *automated reasoning*. In fact, after specifying some DC metadata such as title, creator, date and subject, we intend to say that a particular artifact (a book, for example) has been made by a given author, dated at certain date, and has some subjects (e.g. "engineering" and "steam"). We can add controlled vocabularies to be sure that we specify normalized terms for "creator" or "subject", but only humans can search for contemporary artifacts, check the consistency between *dc:creator* and *dc:date*, look at the historical context, and so on.

Hierarchical classification systems and structured vocabularies do not lend themselves easily to rich inter-linking of conceptual "trees". A major step further in this direction is "CIDOC object-oriented Conceptual Reference Model" (CRM). This provides an ontology of 81 classes and 132 unique properties, which describes with a formal language concepts and relations relevant to the documentation of cultural heritage [4]. CIDOC CRM is a formal ontology for cultural heritage information specifically intended to cover contextual information. It can be used to perform reasoning (e.g. spatial, temporal). CIDOC CRM is an International Standard and is extensible. However, the full approach is at a very high abstraction level, and requires a careful analysis of the information to be managed.

Considering the agreement upon a common metadata set, *why should we consider an ontological approach?* First of all ([6], [7]), even if both a core ontology and core metadata, such as Dublin Core, are intended for information integration, they differ in the relative importance of human understandability. Metadata is in general thought for human processing, while a core ontology is a formal model for automated tools that integrate source data and perform a variety of functions. Vocabularies based on ontologies that organize the terms in form that has a clear and explicit semantics can be reasoned over, which is a fundamental process in enriching knowledge, inferring new information about resources. Secondly, there is a drawback in the implicit assumption made with the metadata approach. In short, it should become evident as adding metadata to the description of an artifact implicitly means that we assume a one-to-many (or possibly many-to-many) relationship between the object and the items identified by the metadata.

III. FROM TAGS TO ONTOLOGY

Web 2.0 is the most suitable and popular way to collect information from a broad range of users. In fact, many users would like to participate in contributing to the creation of new information and knowledge. The possibility of being free to add contents and tags would increase the participation level and extend the knowledge base. However, as it has been already pointed out, this freedom will not guarantee the quality of information and, if in case content could be considered correct, its structure and the completeness of data would be questionable.

On the other hand, a fully ontological approach could be a too rigid process, with an awesome number of forms and fields. It would be a highly formalized process (as the ontologies are) and, finally, it would be reserved to experts in the field, thus losing all the contributions that could origin from a broader community of users.

Therefore the challenge is to *combine the advantages of Web 2.0 and Semantic Web*, namely the freedom and participation energy of the first one together with the open, rigorous, interoperable and supporting knowledge creation approach of the last one.

Additionally, we have to consider the enormous heritage of legacy catalogues and cataloguing tools like thesauri, vocabularies and authority files. Making them available in a new (and *semantically interoperable*) environment can save decades of scholars' efforts and a lot of resources, in terms of money and knowledge.

To achieve this ambitious task we have to:

- map the legacy resources to ontologies;
- organize free tag (folksonomies) into ontologies;
- support users with ontologies, to speed up the use of appropriate tags.

The final result will be to make knowledge available worldwide, and **lift communities towards Semantic Web**.

Transforming legacy resources into ontologies is not a trivial task. Taking as an example cataloguing cards of cultural heritage, it is easily seen that the main issue is that traditional cataloguing cards are based on an item-centric approach, and the sequence of (repeating) fields and subfields implicitly models some kinds of relationships (e.g. *partOf*, *createdBy*, *ownedBy*, etc.) or a cascade of 1:N relationships. In addition, some related fields represent a (faceted) thesaurus. This kind of problems is quite general and requires an in depth analysis of the structures of various types of cataloguing cards; therefore it will not be discussed furthermore in this paper. We stress here that switching towards the Semantic Web approach doesn't mean that all the existing information must be transformed into a set of RDF triples (s-p-o statements). According to the Linked Open Data principles [3], it's enough that repositories show a SPARQL endpoint, which means that any application can interrogate the repository making

reference to some ontologies. There are many software packages that support mapping from internal data organization towards an ontology, so enabling the implementation of adequate SPARQL endpoints.

In general we will recall that even the transformation of available thesauri is neither trivial, nor automatic, as some relators (typically the NT relator) model sometimes a subclass (IS-A) hierarchy, while in other cases they just model an association between two terms, useful for a more effective search. For the last case, take the example of two terms like "*Renaissance*" and "*16th century*". It is evident that both are "time period" which have a start and an end date, with an overlap, which is much different from saying, e.g. that "*statue*" is a NT for "*sculpture*". Other relators, like RT, are even more generic, as they are lacking the semantics of the relationship, and this could bring to very strange consequences if the RT relator is taken as transitive. The main concern is that thesauri are essentially based on a terminological approach, while it would be much more important to enhance the conceptual aspect. In this direction moved the World Wide Web Consortium in defining a standard to represent dictionaries and thesauri, named SKOS - Simple Knowledge Organization System which provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary. As an application of the Resource Description Framework (RDF), SKOS allows concepts to be composed and published on the World Wide Web, linked with data on the Web and integrated into other concept schemes. In basic SKOS, conceptual resources (concepts) are identified with URIs, labeled with strings in one or more natural languages, documented with various types of note, semantically related to each other in informal hierarchies and association networks, and aggregated into concept schemes. In advanced SKOS, conceptual resources can be mapped across concept schemes and grouped into labeled or ordered collections. Relationships can be specified between concept labels. Finally, the SKOS vocabulary itself can be extended to suit the needs of particular communities of practice or combined with other modeling vocabularies. The SKOS data model provides a standard, low-cost migration path for porting existing knowledge organization systems to the Semantic Web. SKOS also provides a lightweight, intuitive language for developing and sharing new knowledge organization systems. It may be used on its own, or in combination with formal knowledge representation languages such as the Web Ontology Language (OWL).

In short terms, the basic idea is to collect information both from institutional, authoritative sources and from users, and organize the user preferred terms into an ontology, making use of the reference scheme which is the basis of SKOS. To achieve this goal we need a framework to manage SKOS files and an environment to support and manage users.

collected through this web portal will automatically be structured using Semantic Web tools and technologies. The platform allows for the creation of various "Laboratories", i.e. *specialized environments*, where a restricted community of users can collaborate by entering and/or enriching information, which will then be made available to the wider community of all of the portal users, and to the entire web community in general.

Several types of actors, with different functions and privileges, have been defined. Generally speaking, the actors can belong to two main classes, depending on if they are playing a technical/management role (i.e. the *Community Administrator* and the *Portal Administrator*) or can be seen mainly as users, at different levels (namely, the *Guest User*, the *Registered User*, and the *Expert*). The actors belonging to these two broad classes are ordered in a hierarchical IS-A relationship. Therefore the Portal Administrator can perform all the actions available to the Community Administrator, plus some additional functions specific to his/her role. The same happens for the different types of users.

The role played by the *Community Administrator* is crucial: (s)he performs the most relevant tasks contributing in the growth of the community and in the organization of contents, mainly through:

- managing dictionaries: creating, importing, exporting;
- managing users: changing their status and relative privileges;
- importing and transforming contents from external sources;
- defining conceptual organization of contents, to build up "*conceptual paths*".

All of the users can add User Generated Content (UGC) and terms to the vocabularies, but the type of UGC is different for the various types of users. The *Guest* can insert new contents or enrich the existing ones with comments and tags. However (s)he must comply with some limitations in the length of inserted texts. It is mandatory to insert some information, namely a *Title*, a short *Description*, the *Location* and *Time*. This allows the georeferencing and chronoreferencing of supplied information, and the implementation of a simple *What-Where-When* interaction paradigm. The input must be approved before being visible to the whole community. The user is free to use any tag, but can also browse existing dictionaries or thesauri, to select more precise terms. To avoid abuses, and keep contents to an adequate scientific level, the UGC and the new terms added to the vocabularies must be validated by an Expert, who is also in charge of structuring terms, thus ordering into ontologies terms and concepts taken from the vocabulary created in the folksonomies. As the vocabularies are SKOS compliant, and the contents are structured making reference to an ontology, these new contents are immediately visible and reusable by the whole Web community.

The Registered User and the Expert can supply more complex UGC. The *Registered User* can enter more structured info, filling in appropriate forms, with online help to select the classes to instantiate. This kind of content is fully compliant with the Semantic Web, hence can be immediately stored in the triple store.

The *Expert* can fill-in contents using even more complex forms, consistent with the national cataloguing standards and requiring appropriate skills. The Community Administrator can promote a registered user to the level of Expert according to the level and correctness of her/his contributions. This will enhance the sense of community and enlarge the base of expert users, on the basis of the effectiveness of their participation, implementing the principle of being authoritative when the rest of community is trusting in the contributed content.

The two important aspects of *sharing of information* and *language standardization* have been faced through the adoption of CIDOC CRM core ontology as reference model in order to effectively share information, and SKOS to codify, export and exchange the terminology used within the portal domain.

C. The case study

A case study has been conducted on data on tangible and intangible cultural and environmental, whose description standard (the BDI card) is defined by the Italian national standard body in [16]. This area has been selected as a test bed because it is plausible that a broad community of user could contribute contents. In the implementation of the prototype we tested all the steps described before. First of all, an in depth analysis of the structure and semantics of the BDI card lead to detect explicit and implicit relationships between descriptive fields, and find an appropriate mapping of the structure onto CIDOC CRM. Thesauri directly referenced by the descriptive fields, or implicitly modeled by related fields were SKOSified and a set of cards was imported into the triplestore, making use of an appropriate XSLT transformation. Subsequently, users were allowed to search and browse data following semantic associations defined in the ontology which was extending CIDOC CRM.

Several other thesauri have been loaded and managed by SKOSware, to fully test this component. Additional info about SKOSware can be found at [17].

V. CONCLUSIONS

The Web 2.0 scenario offers a great opportunity to share and enhance information, and is a trigger towards participation of users. However the social approach to build new knowledge involves the risk of creating babel of terms, which is often inappropriate. On the other hand, a fully ontological approach can be too rigid and formalized, and could discourage general users to contribute. The challenge is then to combine the benefits of the two approaches, after collecting information by a broad community of users, and reorganizing content in a more structured way. The use of Semantic Web

technologies allow to make this new knowledge available to the entire web community.

The process requires a deep examination of the existing patrimony of scholar experiences and the mapping towards an ontological view of data.

The LabC project implemented a platform which enables *capturing, sharing, and enjoying* data on tangible and intangible cultural and environmental heritage.

An important component of the implemented platform is a module enabling the management of thesauri and dictionaries using a graphic interface that can be adapted to various Content Management Systems.

ACKNOWLEDGMENTS

The LabC project has been financially supported by a SME RTD project POR FESR Lazio 2007/2013 – Asse I – Attività I.1.

This work is the result of a collective and cooperative effort by all the authors, but we can somehow assess the different contributions by everyone of us: A. Donnini was the manager of the project, O. Signore was most involved in design and analysis of basic issues and in the mapping of legacy resources into ontologies, A. Ciapetti and F. Piro mainly worked on the implementation issues, D. Merlitti implemented a prototype as proof of concept, M. Lorenzini was mainly involved in the transformation of existing thesauri in ontologies and their representation in SKOS, M.E. Masci worked essentially on the issues related to the mapping of existing cataloguing cards into CIDOC CRM compliant structure, M. De Vizia Guerriero cooperated in the representation of national cataloguing standards into ontologies.

REFERENCES

- [1] A. Grigoris, F. van Harmelen, A Semantic Web Primer, The MIT Press, Cambridge, Massachusetts, April 2004, ISBN 0-262-01210-3.
- [2] T. Baker, A Grammar of Dublin Core, D-Lib Magazine, October 2000 Volume 6 Number 10, available at: <http://www.dlib.org/dlib/october00/baker/10baker.html>
- [3] T. Berners-Lee, Linked Data, available at: <http://www.w3.org/DesignIssues/LinkedData.html>.
- [4] CIDOC-CRM, <http://cidoc.ics.forth.gr>
- [5] K. Coyle, “MANAGING TECHNOLOGY - The Library Catalog in a 2.0 World”, The Journal of Academic Librarianship, Volume 33, Number 2, pages 289–291 (March 2007), preprint at: http://www.kcoyle.net/jal_33_2.html
- [6] M. Doerr, J. Hunter, C. Lagoze, Towards a Core Ontology for Information Integration, Journal of Digital Information, Volume 4 Issue 1, Article No. 169, 2003-04-09, (April 2003) http://www.cs.cornell.edu/lagoze/papers/core_ontology.pdf
- [7] M. Doerr, The CIDOC CRM—An ontological approach to semantic interoperability of metadata. AI Magazine 24, 3, 75–92 (2003)
- [8] Dublin Core Home Page, URL: <http://dublincore.org/>
- [9] A. Isaac, E. Summers, SKOS Simple Knowledge Organization System Primer. W3C Working Group Note, World Wide Web Consortium . <http://www.w3.org/TR/skos-primer/>
- [10] O. Signore, Representing Knowledge in Semantic Cultural Web - EVA 2004 Jerusalem Conference on the Digitisation of Cultural Heritage - Jerusalem, 11-12 October 2004, <http://www.w3c.it/talks/eva2004Jerusalem/overview.htm>
- [11] O. Signore, The Semantic Web and Cultural Heritage: Ontologies and Technologies help in Accessing Museum Information, in Information Technology for the Virtual Museum (December 6-7, 2006 – Sønderborg, Denmark), p. 1.31 (published 2008), Accessible at: <http://www.weblab.isti.cnr.it/papers/public/itvm2006.pdf>
- [12] O. Signore, Ontology Driven Access to Museum Information - CIDOC 2005 Annual Conference of the International Committee for Documentation of the International Council of Museums ICOM-CIDOC - May 24 -27, 2005 Zagreb, Croatia - ISBN 953-6942-15-1 (<http://www.w3c.it/talks/2005/cidoc2005/overview.html>)
- [13] O. Signore, Representing knowledge in archaeology: from cataloguing cards to semantic web, Archeologia e calcolatori, Vol. 20 (2009), p. 111-128
- [14] A. Miles, S. Bechofer, SKOS Simple Knowledge Organization System Reference, W3C Recommendation 18 August 2009, available at: <http://www.w3.org/TR/skos-reference/>
- [15] M. van Assem, V. Malaisé, A. Miles, G. Schreiber, A Method to Convert Thesauri to SKOS, Y. Sure and J. Domingue (Eds.): ESWC 2006, LNCS 4011, pp. 95–109, 2006 (<http://www.springerlink.com/content/4811513015750q25/fulltext.pdf>)
- [16] ICCD, Beni EtnoAntropologici, cataloguing standard, <http://iccd.beniculturali.it/index.php?it/254/beni-etnoantropologici>
- [17] Etcware site: <http://www.etcware.it>

Making Europe's cultural treasures accessible

The Europeana Regia and Europeana Libraries projects

Gabriele Meßmer¹

(1) Munich Digitization Center, Bayerische Staatsbibliothek (BSB), Munich, Germany, messmer@bsb-muenchen.de

Abstract— The key to ensure that resources can be found in catalogs and portals are metadata. In the context of cultural heritage there are a lot of different metadata formats. Among them is the Europeana Data Model (EDM) used as rich format for describing digital objects in the EU-funded projects Europeana Regia and Europeana Libraries.

Keywords: Metadata; Munich Digitization Center; Europeana Regia; Europeana Libraries

I. INTRODUCTION

It is really trivial: We all use metadata when searching in a catalogue or a portal. The key to ensure that resources will survive and continue to be accessible in the future are metadata. They describe different resources and provide information about a certain item's content – in former times stored physically in a library, now accessible on the Web in electronic format. In addition to resource discovery, metadata can help to organize electronic resources, to facilitate interoperability and legacy resource integration, to provide digital identification and to support archiving and long-term preservation.

Metadata aim at the same functions in resource discovery as good cataloging does and those functions are defined in the “Functional requirements for bibliographic records” (FRBR): The FRBR is a conceptual model based on an entity-relationship model with a set of defined elements (i.e. attributes) that are required for a bibliographic record. The main purpose of the FRBR is to ensure that the users' needs are met, which are to find, to identify, to select and to obtain a specific resource.

In the context of electronic resources, there are three main types of metadata:

- Descriptive metadata – these describe a resource for purposes such as discovery and identification. They include elements such as title, abstract, author, and keywords.
- Structural metadata – these indicate how compound objects are put together, for example, how pages are ordered to form chapters.
- Administrative metadata – these provide information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. There are subsets of administrative data; two of them that are sometimes listed as separate

metadata types are: rights management metadata, which deal with intellectual property rights, and preservation metadata, which contain information needed to archive and preserve a resource.

Metadata can describe resources at any level of aggregation, for example they can describe a collection, a single resource, or a component part of a larger resource (i.e. a page in a book) and they can be embedded in a digital object itself or be stored separately.

II. METADATA FORMATS

Cultural heritage institutions use a lot of different metadata formats, among them are:

- MAB: *Maschinelles Austauschformat fuer Bibliotheken* is a specific metadata exchange format used only in German and Austrian libraries. It will be replaced by MARC 21 in the next few years.
- MARC 21 and MARCXML are standard formats used in library catalogs and for data exchange in libraries worldwide.
- MODS: the *Metadata Object Description Schema* is a derivative of MARC 21 and is intended to either carry selected data from existing MARC 21 records or to enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than the numeric tags used in MARC 21 records. In some cases, it regroups elements from the MARC 21 bibliographic format.
- METS: the *Metadata Encoding and Transmission Standard* is a container format that was developed to fill the need for a standard data structure for describing complex digital library objects. METS is an XML schema for creating XML document instances that express the structure of digital library objects, the associated descriptive and administrative metadata, and the names and locations of the files that comprise the digital object.
- DC: the original objective of the *Dublin Core* set was to define a set of elements that could be used by authors to describe their own web resources. Today DC has 15 elements: title,

creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, and rights.

- TEI: the *Text Encoding Initiative* is an international project developing guidelines for marking up electronic texts primarily to support research in the humanities. The bibliographic information is similar to that recorded in a library catalog and can be mapped to and from MARC. Furthermore there are elements defined to record details about how the text was transcribed and edited, how mark-up was performed, what revisions were made, and other non-bibliographic facts.

The Bavarian State Library actually uses TEI version 5 (TEI P5) as metadata format for its digitized objects.

- LIDO: the *Lightweight Information Describing Objects* schema is a reference model developed especially for the needs of museums.
- EAD: the *Encoded Archival Description* is mainly used for the description of archival materials.
- Cidoc CRM: the *Cidoc Conceptual Reference Model* will be used as format for the German Digital Library. Metadata from the different German cultural heritage institutions (libraries, archives, museums and offices for historical monuments) will be mapped to this format.

III. MUNICH DIGITIZATION CENTER OF BAYERISCHE STAATSBIBLIOTHEK

The Bayerische Staatsbibliothek (Bavarian State Library) is partner in two projects within the Europeana framework funded by the European Union: Europeana Regia and Europeana Libraries. The strategy of the Bavarian State Library is to support the digitizing of its rich cultural treasures in different ways, for example digitization on demand by request of a patron, digitization projects funded by different partners (e.g. the German Research Foundation) and the public-private partnership with Google. Since its beginnings in 1997 the Munich Digitization Center (MDZ) has gained a lot of experience in all sorts of digitization projects, as there are the digitization of old and precious manuscripts, of incunabula, the early printings of the 15th century, of more than 40.000 books published in the 16th century, as well as the digitization of more than 5,000 books from the 20th century including full-text generation. MDZ provides one of the largest and fastest growing digital collections in Germany, now (as of 1 April, 2012) comprising 770,312 titles available online and all freely accessible.

The digitization policy reflects the traditional special collection fields of the library: history, classical antiquity, Eastern Europe and musicology. It comprises manuscripts, early prints, modern books, maps and

photographic collections as well as journals and newspapers. MDZ cooperates with many regional, national, and international partners in a lot of different projects not only in digitization but also in storing and preservation. At present more than 358 terabyte are kept in the storage system of the technical partner for preservation activities, the Leibniz Supercomputing Center.

IV. EUROPEANA REGIA

Europeana Regia is a project involving five partners from four European countries, in which nearly 900 important and precious medieval manuscripts will be digitized and presented as a separate exhibition in Europeana. Partners are the Bibliothèque nationale de France (BnF), Bibliothèque royale de Belgique (KBR), Herzog-August Bibliothek Wolfenbuettel (HAB), the Bayerische Staatsbibliothek Muenchen (BSB), and the Biblioteca historica, Universitat de Valencia (BHUV). There are three sub-projects: Bibliotheca Carolina (425 manuscripts), The Library of King Charles V (167 mss.) and The Library of the Aragon Kings of Naples (282 mss.). The Bibliotheca Carolina involves masterworks from the main abbey and bishop schools of the Carolingian Empire (8th-9th centuries), including Reichenau, Saint-Denis, Corbie, Reims, Saint-Amand, Freising and Wissembourg. The manuscripts illustrate the intellectual and artistic activity of these centers of religious life, ecclesiastical and imperial power, and their numerous exchanges of texts and patterns. The library assembled by Charles V and lodged in the Falconry Tower at the Louvre was more than a fabulous collection of books (over 900 volumes in 1380) put at the disposal of the king and his counsellors. It was quite extraordinary for its time and so was the library assembled in the 15th century by the Aragonese kings of Naples is the third component of the *Europeana Regia* project.



Figure 1. Website of the Europeana Regia project

All manuscripts digitized in the project are freely accessible on the Web: in the digital collections of the Munich Digitization Center¹, in Europeana and via the project's website.

V. EUROPEANA LIBRARIES

The goals of the multi-national Europeana Libraries project are to find and to define an efficient model for the delivery of metadata and full-text to Europeana that can be used by further institutions and, moreover, to add more than 5 million records of objects like digitized books, manuscripts, maps, photos and other material to the Europeana portal. Among the partners of the project are 19 leading European research libraries, the Europeana Foundation, the Conference of European National Libraries (CENL), the Consortium of European Research Libraries (CERL) and the Ligue des Bibliothèques Européennes de Recherche (LIBER). The project that involves six work packages (WP) started in January 2011 and will come to an end in December 2012.

Each of the cultural heritage sectors supports its own metadata format. Therefore it is indispensable to agree upon a common format and to map the different - partly very complex - formats to one that can be used by as many providers as possible, by libraries as well as by other cultural heritage institutions. As Dublin Core is a well-established data format today it was chosen as the basis for the Europeana Semantic Element (ESE) set, which was additionally expanded by thirteen tags defined by Europeana. ESE defines a small set of information and so reduces the large variety of formats to a common denominator. ESE can thus be regarded as a first European standard for metadata exchange. The task of the data providers and aggregators is to map their specific format to the Europeana format in due consideration of the "Europeana semantic element specification" and the "Metadata mapping & normalisation guidelines for the Europeana semantic elements".

The work packages 3 to 5 of the Europeana Libraries project are essential for the delivery of data to Europeana. The objectives of work package 3 are as follows²:

- "To aggregate the digital content from all content partners and facilitate its ingest into Europeana.
- To ensure that all the aggregation-related work of the project is carried out cohesively and efficiently.
- To promote and build consensus on the adoption of standards related to the aggregation of digital content.

¹ The digital collections of the Bavarian State Library are accessible via the homepage of the Munich Digitization Center <http://www.digitale-sammlungen.de/index.html?c=startseite&l=en>

² Europeana Libraries, description of work, p. 20

- To represent the European library-domain within the Europeana network and beyond."

The objectives of work package 4 are³:

- "Extend The European Library's existing aggregation infrastructure to enable the aggregation of digital content from libraries in Europe for Europeana, including full-text content.
- Promote and build consensus on the adoption of standards related to the aggregation of digital content.
- Represent the European library-domain within the Europeana network and beyond."

Work package 5 defines⁴:

"Enhance the searchability of existing library-domain content in Europeana by defining transformations from ESE metadata to EDM and establishing best practice taking account of the different types of library contributing to Europeana."

The Bavarian State Library is involved in five of the six work-packages, among them the three work-packages listed below.

VI. METADATA INGEST

To achieve the main goal of work package 3, that is to aggregate the digital content from the library partners, the Bavarian State Library is setting up a new OAI-PMH repository from which Europeana can harvest metadata in defined time intervals. The repository will facilitate aggregating and providing not only digitized material from the Bavarian State Library, but also from other institutions that are partners of the Bavarian Library Network, and also born digital material. The records will be provided in the richest available metadata format - that is MARCXML - and will contain bibliographic and - if existing - structural data.

With work package 5, the libraries are preparing the metadata delivery of the future. Up to now, libraries deliver their records in the Europeana Semantic Element format, a quite simple and robust format. But library data are usually more complex: especially the German format includes advantageous information that is of great value for the linked data world.

Therefore in the next step a new Europeana Data Model (EDM) was created. EDM "adopts an open, cross-domain Semantic web-based framework that can accommodate the range and richness of particular community standards such as LIDO for museums, EAD for archives or METS for digital libraries."⁵ EDM is meant to be the core of the Europeana Libraries Project, as it structures all the data that are ingested, managed and published within the project.

³ Europeana Libraries, description of work, p. 23

⁴ Europeana Libraries, description of work, p. 26

⁵ Europeana data model primer, pp. 4-5

The challenge within work package 5 is to align the library metadata in their richness and complexity with EDM. A lot of different types of national, regional, university, research, municipal and public libraries, small and large ones, with different scopes and approaches have to be taken into account. That is why the WP5 working group is developing a library-domain specific application profile for EDM, addressing the different types of library resources such as monographs, multi-part works and serials, and distinguishing between born digital and digitized objects. Other resources such as still images, videos, manuscripts, incunabula, theses, maps and music scores have to be considered as well. Another challenge is to map the different formats used by the libraries - MARC21, UNIMARC, MODS, METS, TEI, DC - to this application profile and thus prove its validity with real library metadata.

The expected advantage of the new data model EDM is to enrich the records in Europeana and thus make them interoperable and ready for the linked open data world. EDM will use already established standards in the Semantic Web among them: *Dublin Core* to describe the core elements of a digitized object (i.e. title, creator, subject), *Simple Knowledge Organization System* (SKOS) to represent elements of thesauri or classification systems and *Friend of a Friend* (FOAF) to describe all kinds of persons (creators, distributors, publishers and so on). EDM will be based on the *Resource Description Framework* (RDF) with its three object types: resources, properties and statements.

Important findings of the Europeana Libraries project are the so-called deliverables such as “requirements infrastructure and harvester” that presents the wanted technical infrastructure and “report on the alignment of library metadata with the Europeana Data Model (EDM)” that describes best practices for the alignment of library records with EDM. The outcome of work package 4 is the “content ingestion plan” with the requirements for the aggregation process into Europeana.

In addition, a brand strategy will be developed that clarifies the branding of ingested data. It is estimated that the new European Library will be established as one of the main data aggregator for libraries.

Currently Europeana holds open metadata on 2.4 million objects. In February 2012 Europeana released a

pilot data set of linked open data under the open metadata licence Creative Commons Public Domain Dedication (CC0). These data are freely available and linked to external data sources such as GeoNames. They can be re-used for example by the providers for improved user services. To support the free re-use of metadata, Europeana launched a new Europeana data exchange agreement that was adopted by the Europeana Foundation in 2011.

Europeana Regia and Europeana Libraries are two projects that help libraries to present their usually hidden treasures in a new way and to make them visible and free accessible worldwide. Hence new users will be pointed to libraries as well as to all the other cultural institutions represented in Europeana.

REFERENCES

All links seen 1 April 2012

- [1] Definition of the Europeana Data Model elements, version 5.2.3 <http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22>
- [2] Europeana data model primer, ed. by Antoine Isaac, 26.10.2011 version http://pro.europeana.eu/documents/900548/974457/EDM_Primer_111026.pdf
- [3] Europeana Libraries, aggregating digital content from Europe's libraries, competitiveness and innovation framework programme, annex I, “Description of work” <http://www.europeana-libraries.eu/documents/868553/db047daf-25e9-4347-876e-ef11dbd82465>
- [4] Europeana semantic elements specification, version 3.4.1 <http://pro.europeana.eu/documents/900548/dc80802e-6efb-4127-a98e-c27c95396d57>
- [5] Functional requirements for bibliographic records, final report, IFLA Study Group on the Functional Requirements for Bibliographic Records. München: Saur, 1998, viii, 136 p. (UBCIM publications ; new series, vol. 19. Access to the current text, including amendments and corrections to date: http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf
- [6] Metadata mapping & normalisation guidelines for the Europeana semantic elements, version 2.1.1 <http://pro.europeana.eu/documents/900548/dc80802e-6efb-4127-a98e-c27c95396d57>
- [7] Understanding metadata. Bethesda, MD: NISO Press, 2004, 16 p. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>



Workshop: Open Access 2002-2012

Chair

Beatrice Tottosy, University of Florence, Italy

Visions and Possibilities of Open Access Publishing in the Dissemination of the Hungarian Performing Arts Heritage

Attila Szabó⁽¹⁾, dr. Piroska Ács⁽²⁾

(1) Hungarian Theatre Museum and Institute, OSZMI, Budapest, Hungary,
szabo.attila@szinhaziintezet.hu

(2) Hungarian Theatre Museum and Institute, OSZMI, Budapest, Hungary, piroska.acs@szinhaziintezet.hu

Abstract—Thinking over the special nature of open access publication in case of performing arts content the paper investigates the options and difficulties national collections have to face when attempting to disseminate digital copies of their performing arts content. The paper presents a case study of such an attempt, namely the endeavours of the Hungarian Theatre Museum and Institute, the largest theatre, dance and puppet theatre collection in Budapest, to disseminate their content through the ECLAP digital library. The paper examines the possibilities of how to face the biggest obstacles in this process, the difficult process of clearing of intellectual property rights. Several working solutions are presented which enable similar collections to publish a significant and relevant part of their materials.

Keywords- performing art representations; intellectual property; open access publishing; theatre history; ECLAP

Theatre records as shadows

The concept of open access publication of performing arts heritage appears as both indispensable for the given artistic field but also significantly different from the publication of other types of content. Stemming from the specific nature of performing arts, it is often argued that the real substance of a performing arts event cannot be reproduced by any technological means, and that the spacio-temporal bounds of a performance cannot be truly broken up by any device, no matter how fast the mechanisms of image and sound capture seem to be developing. Therefore any document of such an artistic event could be no more than a distant shadow of the original and irreproducible work of art. These considerations have very strong bearings on theatre theory and historiography, but they must also be considered as a very pragmatic point when it comes to judging the publication of performing arts records. The different stakeholders of the field (and to some extent also the authors of intellectual property rights legislation) seem to forget about the fact that any published copy of a performing arts event can only be an annotation of the given event, a reconstruction attempt of a bygone spatiality and temporality. Thus the case is very different from the online publication of, e.g. a scientific article or a reproduction of a painting, where a certain public access

is given to the original document in its totality, since the online version reproduces a different degree of the experience given by the observation of the original piece. Whereas a photograph of a theatre performance or even a video recording of a theatre play can be no more than some ruinous landmark in the reconstruction of a real-time experience which cannot ever be repeated.

Yet, on the other hand, it is obvious that the performing arts field has no other means than these feeble records to step outside the limits of its spacio-temporal limitations, and spread the news of its existence to a much larger number of people. Some of them might consequently choose to attend the show, if they are lucky enough to live in that very limited timeframe where the given event is performed. For the rest of us it's just history. Therefore it seems reasonable to say that performing arts records can by definition only be considered observations of the second order, that is, observations of other observations, meta-images which refer to other images which can only be captured fully by the memory of the human mind [1]. And of course, these records are excessively subjective, representing the viewpoint of the person (or people) making the recording, which leads to other difficult theoretical insights, testifying of a conspicuous distance between the event and the record. Consequently, records of performing arts events serve either the role of stimuli for historical research or tools for marketing, rather than being entertainment products on their own. In this sense, a video recording of a theatre performance on the internet would more likely increase the number of tickets sold for the particular show, acting more like music records stimulating concert attendance, as people would be interested in the experience of what has been left out from the video recording, of that live touch which cannot be captured. And also very different from the model of films or texts published online, which, to the general belief, tend to discourage purchase of cinema tickets or books. Against the official lobbies against (copyright infringing) peer-to-peer sharing of performing arts content an increasing number of studies give evidence of the 'sampling' use case of the downloaded materials, which precedes and even encourages purchase.[2]

Of course, if the content is available at all in a commercial form. The Hungarian practice shows that the performing arts institutions in our country are not yet fully aware of the benefits of offering records of their present and earlier activity to the public. In most cases video recordings are made but kept locked in drawers, never digitised, never published. And naturally it is also very rare for the theatres to build their own archives, catalogued and accessible, at least for research purposes. This is of no surprise, however, if one looks at the present practice of theatre institutions: usually they are mostly focused on their artistic activities, while the means and personnel are insufficient for a meticulous archive to be kept. Furthermore, they lack the marketing tools and the experience in this field.

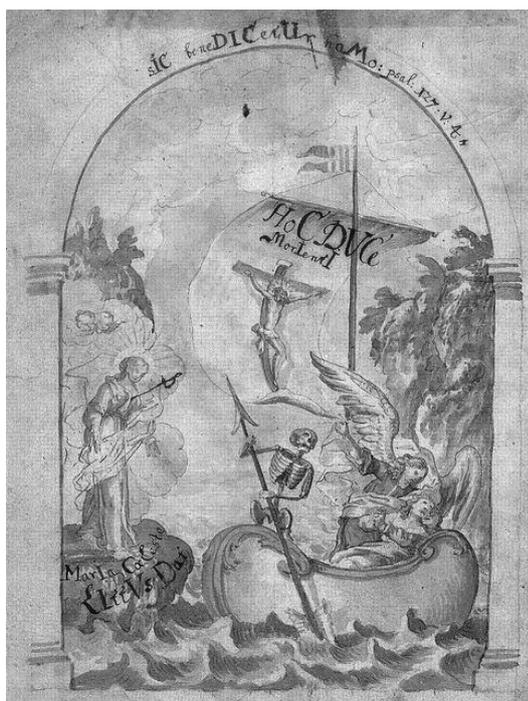


Figure 1. The Allegory of Salvation. (From the Sopron Collection of Jesuit Stage Designs, 1728, Hungarian Theatre Museum and Institute, Secnography Collection)

Open access publication attempts at the Theatre Institute

The Hungarian Theatre Museum and Institute is a government-funded archive which, among its other functions, has the duty to keep all records of the Hungarian theatre life (including puppet theatre and dance). The Institute is a place which could not only take over some of the burden of archiving from the performing arts institutions, but it is also a principal part of its mission statement to make this information accessible to a wide public of researchers, students and very importantly to theatre practitioners themselves. As a central spot dedicated to the Hungarian theatre memory, located in Budapest, our theatre archives not only support academic research on theatre history, but often

make sure that the theatrical tradition is reflected upon and has a certain continuity in the new theatrical output. Before putting up a new show many dramaturges, director's assistants, set and costume designers carry out extensive research in our press archives, library, video archive, scenography collection etc. [3]

Yet even the process of archiving has its substantial difficulties. Some of these stem from the transformation, i.e. expansion of the theatre field, the blurring of genres, the dissolution of the boundary between professional and amateur, establishment and independent forms. In the fifties when systematic collection started all theatres and media could be covered, but after the turn of the Millennium information boom and diversification of performing arts forms made it impossible to achieve a comprehensive and all-embracing collection. A theatre survey issued in 2005 shows that in Hungary the number of theatre venues has been doubled from 1989 to 2005 in Budapest, with an increase of 3500 seats in 32 venues, resulting mostly in an increase of the small playing sites. In the last years the boom in mostly smaller independent companies has been even more intensive and a shortage of space can be perceived today [4].

Conversely, funding for the Institute has been constantly decreased making it virtually impossible for us by today to commission or produce any recordings of our own. It can also be mostly written on the account of the lack of qualified personnel that the Theatre Institute has no (official) channels or power today to collect new materials from the theatres and companies, on a constant, systematic, regular and up-to-date basis. On the other hand, the producers and owners of the recordings and the theatres themselves seem to be holding on to their records – partly for financial reasons, e.g. photographers hoping for revenues for usage rights of their theatre photos. And they cannot be judged so bluntly given that the whole arts field is severely underfinanced, and the situation is only getting worse. And thirdly, there is at stake a certain symbolic power perceived in keeping some records restricted or non-accessible to the widest public, thus protecting the somewhat mystical communion of the theatre experience. When the opening performance of the new national theatre in Budapest was broadcast live in 2002 (*The Tragedy of Man* by Imre Madách) there was significant protest coming mostly from the theatre profession for destroying thus the very essence of a theatre performance.

The real problems, however, arise when we attempt to make this heritage publicly accessible. Looking at most parts of our collections we have to realize that almost nothing we store is owned by us, that is the IPR are not ours or not even cleared for the biggest part of our materials. Although we managed to purchase the rights of some valuable items (for instance many set and costume designs of our scenography collection), using some occasional extra application funding, most of the performance photos belong to individual artists, most of the videos have been recorded by the Hungarian Television or some private companies and we only store

a copy of them – sometimes even not officially. Several hundred audio tapes contain performance recordings from the pre-video times, produced by the Hungarian Radio, and we have thousands of high-value performance photos from the period of 1950-2000 copyrighted to the Magyar Távirati Iroda (Hungarian News Agency). Even if we are both national public archives, and have a hard copy of these performance photos in our collection, we are charged full price by the Agency whenever we wish to publish any of them, as MTI has its own very strict sales policy.

Even these short examples make it obvious that any attempt from the Theatre Institute's part to build and publish a digital library of performing arts, – which could offer a comprehensive overview of professional performing arts history in Hungary from the 19th century to our days – would be a Herculean task, to say the least. And the biggest challenge would not be the production of the digital copies but the disentanglement of the complicated and dispersed IPR status. And the Hungarian IPR law does not make things any easier. As in most European countries performing arts records activate possibly the most complicated IPR scenario, as the rights are shared between the producers (and actors) of the original show, the composers and performers of the music and the producers of the recording itself. Obviously clearing all these rights for thousands of video recordings is not a realistic scenario.

To publish or not to publish

When considering the spread of open access publication practice in the context of performing arts, it must first be noted that unfortunately this field is still very much dominated by the scientific field, while its spread towards the arts and humanities is a slow and gradual process. As a study from 2008 remarks:

Readers of journal articles, 18% of arts and humanities researchers (compared with 44% in the life sciences and 32% in the physical sciences) say they are familiar with methods for finding open access material. Just 6% say they use open access journals frequently (compared with the 35% of life science researchers who use BioMed Central journals alone). The proportion of arts and humanities researchers who say that they visit their own institution's repository frequently is in line with the average for all disciplines, however, at 7%. [5]

However, an increasing number of articles, also focusing the field of sciences, seem to provide statistical evidence that publications openly accessible are cited substantially more than those that are not. [6]

On the other hand, with the spread of the internet-based publishing one might ask if there is any need for a public collection to put serious efforts in archiving and publishing theatre-related sources. Most of the Hungarian specialist performing arts magazines have their own digital archives which also work on a delayed

open access basis (*Színház, Ellenfény, Kritikai Lapok*). Theatre articles in the dailies are available online even sooner, while there is an increasing number of exclusively online, open access cultural journals (e.g. www.szinhaz.hu). The large community content portals (like youtube) also store and broadcast an increasing number of data related to performing arts. Despite of all these, several factors still strongly justify our efforts in collecting the records. For instance we are building a database of theatre reviews which aggregates all the different content according to the same metadata. The database is easily searchable along several dozens of criteria, which serves research purposes more than ad-hoc Google search. Furthermore, as the newspaper publishers are not primarily interested in the specific field of performing arts some of the articles become inaccessible after several years, or there is no guarantee that their own archives will still be accessible after an eventual liquidation of a newspaper. Probing at random our press database we found that from 50 articles recorded in 2008, which were originally published online, as many as 24 already have dead links in our database (meaning that the articles had been either moved or deleted), which justifies the need to archive the content itself in pdf form.

On the other hand, public interest constantly urges the Institute to provide an increasing online digital content of the older documentation, which is still only stored in form of paper cuts in thousands of dusty folders. We joined ECLAP with the hope to be able to fulfil some of these needs. Digitisation has been one of our core activities for the past years, also a strong expectation from the supporting ministry, yet without making the virtual contents publicly accessible the whole process is prone to become an end in itself. And some of our hopes seem to have been met so far: we were offered the chance to publish our performing art content alongside with more than a dozen prestigious archives from Europe, also being forwarded to Europeana, which would give a great world wide access also to the most crucial phenomena of Hungarian theatre. This could stimulate further outputs like specialist research publications using our content or exhibitions in a specific field of interest, even using theatre content from all over Europe. Also the content organization, annotation and presentation tools which are being embedded in the ECLAP portal offer cutting edge options, which could not be implemented in our own systems in the near future. [7]



Figure 2. Poster of an operetta performed open-air at the Budapest Zoo, 1959, Hungarian Theatre Museum and Institute, Collection of Posters and Playbills

On the other hand, however, we are meeting strong challenges in the implementation of the ECLAP project. The biggest of these is the issue of IPR limitations, which made us have to fundamentally rethink the content we will be able to offer to the project. Our hopes to find a quick and cost effective way of clearing property rights for a large number of items for the purpose of publishing them on the ECLAP portal seemed to be too naive from our part. The workshops we attended so far on these topics confirmed that the situation in the other European countries is not easier at all. Until some common European agreement will decide to ease the access of public archives to offer some form of online access to the digital items of their collection each country will stay responsible for clearing the rights for their content uploaded to ECLAP for every country of the world. Sadly, the database itself cannot offer the level of protection for the digital content to avoid the necessity to clear rights for each item individually. It also turned out that limiting the access of some restricted content to educational and research users only, within the ECLAP system, could still not guarantee that only these institutes could access this content exclusively from their own premises or staff and that the content could not “escape” to the uncontrollable flow of the world wide web. Also with these limitations the individual clearing of IPR rights for each item cannot be avoided. Generally, programmers argue that content protection methods, no matter how sophisticated they may be, offer no ultimate safety to content once uploaded to the internet. For the

skilled and motivated there is always an easily accessible workaround available against all methods of protection. But first and foremost, especially in the context of open access possibilities, it has to be said that the internet is not the place where restrictions are desirable, stimulating or helpful in any way. As content providers we really do not wish to be akin to the lady from the Hungarian folk tale who was asked by the king to visit him on foot yet still on horseback, to bring some present and yet not to bring anything, to be dressed and yet to be naked.

The biggest challenge of ECLAP from our part seems to be the fact that the consortium has undertaken to provide, by the end of the project in June 2013 a very large amount of digital items. This number, however, has not been divided proportionally among the partners. Originally we wished to give a comprehensive selection of the Hungarian performing arts history combining several different types of content (with a natural emphasis on motion picture and sound) which could give a balanced representation of different eras of our theatre history. From 18th century set design, 19th century photos and playbills, 20th century videos, reviews, to the newest trends in director’s theatre, post-dramatic and site-specific theatre. We also wish to offer a balanced representation of the different performing arts subforms: drama, dance, puppet theatre. Yet within this time frame and the amount of financial resources available it is impossible to prepare and clear the rights for such a high number of items. This is especially the case with videos, for the reasons explained above. Unfortunately, this way quality seems to fall prey to quantity, a truly unfortunate thing in case of such an audacious project wishing to stimulate the finding of synaptic links, abundant cross references between the performing arts items from so many different countries and cultures of Europe. Especially if these links (annotations, references) should be added by a community of professional researchers, university teachers and theatre students, who will focus mainly on the quality, relevance and ease of access of the content. With a more modest (yet balanced) number of items more effort could have been invested in creating good quality English language translations to the uploaded content and metadata, which would have facilitated access of researchers and increase professional quality against other, uncontrolled social content sites widely used. Generally, we believe that a more meticulous and extended preparatory work on the content to be implemented (types, taxonomy, eras, events, points of focus) and the structure of the portal (obligatory and optional metadata fields) should have been carried out with the involvement of all partners before any uploading or division of item numbers was started.



Figure 3. László Gyenes as Lucifer in Imre Madách's *The Tragedy of Man*, 1883. Photo Collection, OSZMI

IPR issues and suggested remedies

However, in order to try to fulfil our obligations towards the consortium and still be as faithful as possible to our original open access publishing aspirations we managed to find several workarounds of which the two most important are: a) to focus on content which has already entered public domain, b) to try to invoke the option of free use by selecting excerpts from full representations of performing arts events. According to the 70-years limit specified by Hungarian Intellectual Property Act we can freely publish any item in our collection which was created by an author who died before 1942. Article 33 of same Act specifies which are the instances of free use, which “ shall not be subject to the payment of any consideration and to any authorization of the author. Only works disclosed to the public may be used freely pursuant to the provisions of this Act. [8]” The next article specifies that citations and borrowing could be justified as instances of free use, providing that the source is indicated and is not used for commercial purposes. However the Act does not specify any exact length of how long the citation can be, only mentioning that its length should be justified by the nature and purpose of the borrowing work. This way we decided to judge the length of the excerpt bearing in mind the full

extent of the given performance, which is often two-three hours long. Therefore a sample of ten minutes from different scenes could well fall into this category of usage. In the metadata section of the portal we will publish comprehensive data about the authors and recorders of the given performance. The same principles will be applied to the audio recordings.

Act No. LXXXVI of 1999 on copyright	
<i>Article 31</i>	The authors' rights shall enjoy protection during the lifetime of the author and for seventy years following his death. (2) The seventy years' term of protection shall be counted from the first day of the year following the death of the author and, in the case of joint authors, from the first day of the year following the death of the joint author dying last. (3) In case the person of the author is unknown, the term of protection shall be seventy years and shall be counted from the first day of the year following the first disclosure of the work. However, should the author become known during this period of time, the term of protection shall be counted as in Paragraph (2).
<i>Article 34</i>	(1) From a work any part may be cited by indication of the source and naming the author indicated therein. Such citation shall be true to the original and its scope shall be justified by the nature and purpose of the borrowing work. (2) Part of a literary or musical work or such a type of an entire work of a minor volume disclosed to the public may be borrowed for the purposes of illustration for teaching and scientific research, with the indication of the source and the author named therein, to the extent justified by the purpose on condition that the borrowing work is not used for commercial purposes.. Borrowing shall mean the use of a work in another work to an extent that goes beyond citation.

Another workaround we try to apply is to publish tri-dimensional content items present in our collections (puppets, costumes, models), which have been purchased by the Institute or given to us free of charge. This would be a very significant step forwards in our own digitisation process too, since these items are usually very difficult to present, as we have no puppet or costume displays and most of these items are stored in special warehouses without any public access. The only downside of this process is the high cost of creating our own digital representations of these objects, which requires an experienced photographer, also familiar with the theatre field. Funding for this work must come from other sources as ECLAP as an ICT-PSP project does not directly fund digitisation, only the dissemination of the already existing digital content.

Finally, we are striving to select the presented content in a way to focus, on the one hand, on the most valuable items of our collections, which are by themselves relevant in an European context, and underline the adherence of the Hungarian theatre life to the stream of

the European theatre history (like the Baroque Jesuit Stage designs from the Storno Collection in Sopron or the work of puppeteer Géza Blattner in Paris, or the first dance performances on Béla Bartók's music). On the other hand, we also try to single out the relevant and specifically Hungarian (or at least Central European) aspects of our theatre history (the operetta tradition, the strong influence of the Romantic dramaturgy on the stage, folk plays, different ways of avoiding censorship during the Communist Dictatorship, the fundamental renewal of the dramatic and theatrical language after 1989 etc. We also try to ingest some of our collections in full (e.g. the graphic posters from the 19th century to our days), thus offering the possibility of a comparative analysis with content from the same era from other partners in the ECLAP project, even using only non-verbal ways of approach (analysis of iconography, design, repertory etc). On the other hand, we try to offer, mostly to our Hungarian readers, full 'source packages' for the reconstruction of the most significant Hungarian performances from the past. Full sets of reviews, video or audio fragments, posters, set and costume designs were selected to complement each other and facilitate research and educational use. This is especially relevant for the great milestone performances of our theatre and dance history, which have had no extensive documentation available online before.

Because soon what is not out there will cease to exist; but turning it inside out: those things which do make it to the public forum will have a good chance to be born again.



Figure 4. Gizi Bajor and Oszkár Beregi as Juliet and Romeo, National Theatre, 1918, OSZMI, Photo Collection

REFERENCES

- [1] Luhmann, Niklas, *Art as a Social System*, (Die Kunst der Gesellschaft) 1995 transl. EvaM. Knodt. Stanford: Stanford University Press.
- [2] Andersen, Brigitte – Frenz, Marion, *The Impact of Music Downloads and P2P File-Sharing on the Purchase of Music: A Study for Industry Canada*, [http://www.ic.gc.ca/eic/site/ippd-dppi.nsf/vwapj/IndustryCanadaPaperMay4_2007_en.pdf/\\$FILE/IndustryCanadaPaperMay4_2007_en.pdf](http://www.ic.gc.ca/eic/site/ippd-dppi.nsf/vwapj/IndustryCanadaPaperMay4_2007_en.pdf/$FILE/IndustryCanadaPaperMay4_2007_en.pdf)
- [3] Our webpage <http://szinhaziintezet.hu>, also available in English, gives a detailed description of our collections.
- [4] Szabó, István, *A színházi struktúra helyzete, a jövőre vonatkozó elképzelések*, in *Színházi jelenlét, színházi jövőkép*, NKA kutatások, Budapest: OSZMI, 2005
- [5] Heath, Malcolm – Jubb, David – Robey, David, *E-Publication and Open Access in the Arts and Humanities in the UK*, *Ariadne*, 2008.01, <http://www.ariadne.ac.uk/issue54/heath-et-al/>
- [6] Hajjem et al, *Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact*, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2005, <http://arxiv.org/ftp/cs/papers/0606/0606079.pdf>
- [7] A partial open access output of our digital databases can be accessed at: www.szinhaziadattar.hu.
- [8] Republic of Hungary, Act No. LXXVI of 1999 on copyright, official English translation published at: http://portal.unesco.org/culture/en/files/39703/12519733393hu_copyright1999_Conso2007_en.pdf/hu_copyright1999_Conso2007_en.pdf

The White Road to Open Access: A Proposal by the Open Access Publishing Workshop of the University of Florence

Beatrice Tottossy⁽¹⁾, Arianna Antonielli⁽²⁾

(1) Dept. of Comparative Languages, Literatures and Cultures, University of Florence, Florence, Italy,
<tottossy@unifi.it> <bsfm@comparate.unifi.it>

(2) Dept. of Comparative Languages, Literatures and Cultures, University of Florence, Florence, Italy,
<arianna.antonielli@unifi.it>

Abstract—This paper aims to present the current development of the open access routes within an Italian academic context: the Open Access Publishing Workshop (hereafter OAPW) of the Department of Comparative Languages, Literatures and Cultures of the University of Florence. The OAPW aims to move along the gold, green and platinum roads, and to interrelate them into a third route, which we have called “the White Road” to Open Access. The latter attempts to promote, within a new possible context and social net given and made possible by Open Access Publishing Workshops (both academic and non), three ways of achieving open access: self-archiving (Green Road), author charging and users free (Gold Road) and completely free (Platinum Road). The OAPW, as a White Road, is currently working behind the curtains as a back-end service and facility to the frontline OA roads. It is currently engaged in three main fields: 1) **Research**: the OAPW is working on a) the identification of good practices in copyright management in Open Access volumes, journals, and websites, in the Humanities Area of Cultural Heritage, with a particular focus on the academic authors’ attitudes and viewpoints; b) online literary books (e-books) ‘enriched’ by reading and interpretation tools). 2) **Teaching**: the OAPW promotes: a) publishing courses and trainings to university students; b) an introduction to scientific and cultural activities related to the digital book (e-books). 3) **Publishing**: the OAPW aims to develop an integrated process of editing, producing, archiving, self-archiving, and promotion of new, fully or partly, open access products.

Keywords-component;

Keywords-component *Cultural Heritage in the Humanities; White Road to OA; OA Publishing, Research, Teaching on OA-books; Open Access Publishing Workshops*

I. AN INTRODUCTION TO THE OPEN ACCESS

“Promoting innovation through, inter alia, open access to scientific publications, a database for knowledge management and dissemination and promoting technology topics in educational programmes” [1].

A. Editorial Market Strategies (1960-2002)

Ensued by digital technologies, the digital revolution of the twentieth-century has improved literary and communicative systems and enabled the so-called digital preservation of cultural heritage, as well as its expan-

sion in terms of accessibility and availability. Preserving cultural heritage through oa publishing, collecting and archiving, as well as providing open access to data created by 3D laser scanning and digital modeling, is the new frontier to be achieved and improved. The notion of scholarly communication and publishing in the Humanities is deeply connected with the investigation on and preservation of cultural heritage in the past and present. It is currently facing a series of topics and problems sometimes very hard to deal with, mostly related to scholarly and commercial publishing business, to traditional and online publishing, as well as to issues of copyright and copyleft counter-movements, pre-print and post-print standard commercial publisher licenses, subject and/or institutional repositories, and strategies of self-archiving and self-publishing. State-of-the-art technologies and methods have opened up new ways for improving scholarly communication, concurrently, new market strategies have paved the way for its unequal accessibility and transmission. “Beginning in the late 1960s,” Jean-Claude Guédon argues, “commercial publishers began to acquire a variety of journals that belonged to the enchanted ISI set with a view to recouping [...] investments as quickly as possible through steep price rises”. [2] This process provoked the so-called “serial pricing crisis” or “journal crisis” (1990s) in the academy, whereby the rise in journal subscriptions fees applied by commercial publishers led to a decrease in the number of subscriptions that academic libraries were able to afford [3]. Thus, as reported by CRUI Guidelines (2009), “The current struggling conditions of libraries have worsened [...] as [...] subscriptions are interrupted for on-line journals. There is no guarantee for the regular access to issued under regular subscription, unless this is explicitly mentioned in the libraries’ contractual arrangements” [4].

Even though the rise in prices, especially for core journals, was still a favourable way for majors to pay off the investments demanded by the fusion process, this editorial strategy clearly had negative effects on scientific communities at large; to such an extent that, by now, the process accounts to huge sums of money to be paid for the regular circulation and distribution of

knowledge. In addition, this gap between the circulation and availability of information among the different institutions reduced their potential to develop and maintain research within themselves, since authors chose to publish their research in journals linked to leading publishing houses, in order to enjoy the privilege of top refereed publishers.

This vicious circle was also linked to another equally negative and destabilising market strategy, namely, the so-called “circle of gifts” [5], according to which scientific authors hand over their research for free to commercial publishers (losing almost all control on the distribution of their work), and get a wider circulation of their research in turn. Universities were thus forced to purchase the finished product to make it accessible to their own research entourage. These market strategies created a severe inequality between the actors in the publishing field. But they also promoted initiatives oriented towards a greater freedom in the circulation and fruition of knowledge.

B. Developments and Reactions: Library Consortia (1990s) and the Open Access Movement (2002)

The creation of library consortia was the first real effort to tackle the situation mentioned above [6]. In the 1990s, libraries increased their ability to afford scientific works and traditional and/or electronic journals, and to retain rates of access through cooperation agreements. As a matter of fact, library consortia succeeded in negotiating licence terms for journals subscription, on behalf of the institutions they represented. Library consortia also helped universities to contrast the big commercial publishers’ monopoly on the publication of their research, by supporting the birth of university presses and implementing free access to scientific journals and works *tout court*. University presses nowadays are the primary publishing location of many universities. Their premise is to rationalize expenses so as to guarantee a more influential scientific research as well as a safer working environment for researchers, mainly as regards intellectual property. Moreover, university presses aim to storage scientific data into National Libraries and digital archives and to disseminate them open access on the web. It is worth noticing that publishing policies of many university presses go hand in hand with the philosophical and political concept of the Open Access Movement. The Florence University Press represents an exemplary case of interaction with oa policies and models.

Firstly ratified by the Budapest Open Access Initiative (February 2002), the Open Access Movement was further developed by Bethesda (June 2003) and Berlin (October 2003) Declarations [7]. The Budapest Open Access Initiative reads [8]:

“1) Open access is intended as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community. In order

to realize the vision of a global and accessible representation of knowledge, in the future Web content and software tools will have to be openly accessible and compatible.

2) This idea is related to Public-funded scientific results, that authors publish for free; many initiatives for OA are promoted by public and private associations as well as Library networks, Academic Institutions and Research Centres, and by the Soros Foundation.

3) “open access” to this literature, means its free availability on the public internet in a wide sense. The only relevant constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.”

C. The Three Roads to Open Access

In order to make the academic scientific information openly accessible online and to thwart the politics of copyright, the Open Access Movement first developed, two different strategies to improve accessibility and availability to online scientific knowledge, as well as to change the transfer of the copyright policies: the so-called gold and green roads to open access. Esther Hoorn and Maurits van der Graaf explain that, “The green road refers to (subscription-model) journal publishers that allow some form of the article to be archived in institutional repositories and to be made accessible either directly after publication or with a waiting period of 6 to 12 months. The golden road refers to a change in academic journal publishing: the academic journal itself is an Open Access journal and the business model has to change from the subscription model to the ‘author pays’ model” [9].

Open access repositories and journals are now a reality and their copyright policies should be considered in the interests of maintaining research quality standards. It is worth mentioning, in this field, the working conference on the management of copyright issues for universities that took place in June 2001 in Zwolle (the Netherlands). An international delegation of participants agreed to collaborate on Copyright Management for Scholarship and to develop a set of principles aimed to improve access to scholarly communication and provide a guide to good practices on copyright policies in universities. A programme was subsequently developed by the Joint Information Systems Committee (JISC) from the United Kingdom and by the SURF Foundation from the Netherlands in order to follow the Zwolle Principles in assisting “stake-holders—including authors, publishers, librarians, universities and the public—to achieve maximum access to scholarship without compromising quality or academic freedom and without denying aspects of costs and rewards involved” [10].

Among disciplinary or subject repositories, it is worth recalling ArXiv, PubMed Central, Cogprints, CSeARCH, RePEc, E-LIS, and NCSTRL. As to institu-

tional repositories, we can mention OpenDOAR, the Directory of Open Access Repositories; the Registry of Open Access Repositories (ROAR); and Pleiadi (<<http://www.openarchives.it/pleiadi/>>). Open Source softwares, such as EPrints (Southampton, 2000), CDSware (CERN, 2002), DSpace (MIT-HP, 2003); FEDORA (Virginia and Cornell University, 2003) are currently used. They are based on the OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting, 2.0, <<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>>).

The number of peer-reviewed OA journals (Gold OA) listed in the Directory of Open Access Journals (DOAJ, <<http://www.doaj.org>>) as of November 2011 is 7,311; 1,728 the number of CC-licensed journals in the DOAJ; 6,502 the number of peer-reviewed OA journals listed on December 12, 2011 in Open J-Gate. Concerning OA repositories (Green OA), the number of OA, OAI-compliant repositories listed by ROAR (Registry of Open Access Repositories) on November 28, 2011 is 2,584; 3,946 the number of OA, OAI-compliant repositories listed by OpenArchive.edu; and the number of full-text items on deposit at E-LIS, the Open Archive for Library and Information Studies is 12,473 on November 28, 2011 [11].

These figures show how it is possible to assume that open access journals are in rapid growth. Yet, among these, the number of journals charging authors in order to recover publication expenses no longer covered by subscriptions is equally growing, and amounts to less than half the total [12]. According to Tom Wilson (Publisher/Editor in Chief of the electronic journal *Information Research*, <<http://informationr.net/ir/>>), such tendency might be prevented by the so-called Platinum Road. On April 19, 2007 Wilson posted an item to the BOAI Forum, on the thread: “Re: Independent open-access Canadian medical journal launches”. He defines the Platinum Route, in the academic context, as “the voluntary, collaborative, no-charge model that is usually overlooked in the debates on OA. Usually that debate concerns itself with the choice between author charging and open archiving - one restricts access to authors, the other is crucially dependent upon the acquiescence of the commercial publishers. The only true open access, which we can perhaps call TOA, is the Platinum Route [...]” (<<http://threader.ecs.soton.ac.uk/lists/boaforum/1078.html>>). He also provides links to other case studies of the Platinum Route by Bo-Christer Björk, David J. Solomon and John Willinsky & Ranjini Mendis [13].

On Nov 7, 2007, Wilson posts an additional comment in the blog following the spin-off of the e-journal *Information Research*: “[...] I distinguish between open access through author charging, which is what the Gold Route is usually promoted as being (and which all official bodies from the NIH to the UK research councils assume as 'open'), and the Platinum Route of open access publishing which is free, open access to the publi-

cations and no author charges. In other words the Platinum Route is open at both ends of the process: submission and access, where as the Gold Route is seen as open only at the access end” [14].

D. The Open Access in Italy. An Outline [15]

- 4-5 Nov. 2004: Conference on Open Access to scholarly literature held in Messina, promoted by the Library Committee of the Italian Council of Rectors (CRUI), in collaboration with the University of Messina. During the conference more than 30 Italian Universities signed the “Messina Declaration” in support of the Berlin Declaration on Open Access.
- 2004: creation of PLEIADI, the Italian service provider. Publishing of the Creative Commons Public Licenses in Italian.
- 2006: the Conference of Italian Universities Rectors (CRUI) decided to create a working group on OA as part of the CRUI Library Committee.
- October 2007: publication of the “Linee guida per il deposito delle tesi di dottorato negli archivi aperti” (Guidelines on Doctoral Dissertations in OA Repositories).
- April 9, 2008: the Italian Wiki on Open Access (<http://wiki.openarchives.it/index.php/Pagina_principale>) was created and is currently managed by a team of OA experts. It includes: OA definitions, history of the movement, the intellectual property rights, FAQ on OA and a rich bibliography on Open Access in the Italian language (<http://wiki.openarchives.it/index.php/Bibliografia_in_lingua_italiana>).
- April 2009: the CRUI OA Working Group published the following guidelines: 1) “Raccomandazioni su Open Access e la valutazione dei prodotti della ricerca scientifica” (<<http://www.crui.it/HomePage.aspx?ref=1782>>); 2) “Riviste ad accesso aperto: linee guida” (<<http://www.crui.it/HomePage.aspx?ref=1789>>); “Linee guida per gli archivi istituzionali” (<<http://www.crui.it/HomePage.aspx?ref=1781>>).

In terms of OA publishing, the Italian situation is quite old-fashioned. Among the few publishing houses which have embraced the OA Movement in Italy, it is worth mentioning Casalini Libri (<<http://www.casalini.it/>>) with its Digital Division branch and the Florence University Press (<<http://www.fupress.com>>). Both of them have taken part into OAPEN (<<http://www.oapen.org/>>), the European project for OA monographs. It is also worth mentioning the project “Libri gratuiti in Ateneo” of the Polimetrica Publishing House (Monza). A project which aims to improve Italian scholarly communication and to provide easy access for free to users (especially students) [16]. The number of institutional repositories in Italy is 72 (see OpenDOAR). They belong to universities, research institutions (CNR, INGV), and postgradu-

ate schools (SISSA, SSPAL). The most widespread software is EPrints, followed by DSpace, CDSWare (now CDSInvenio), and ETD-db.

II. THE OPEN ACCESS PUBLISHING WORKSHOP

A. Foreword

Within the aforementioned three-layered differentiated context of developing open access (gold, green and platinum roads) in Italy and trying to keep high quality scientific research freely accessible, a fourth route to OA has been implemented since 2005 by the Department of Comparative Languages, Literatures and Cultures of the University of Florence: the “White Road” to Open Access. The latter attempts to promote, develop and interrelate, within the context of the Department Open Access Publishing Workshop (OAPW), the aforementioned ways of achieving open access, i.e. self-archiving (Green Road), author charging and users free (Gold Road), and completely free (Platinum Road). The Florence OAPW, based on a new integrated system of research, teaching, and publishing, is also directed towards the creation of a social network of Open Access Publishing Workshops (both academic and non). The Florence OAPW, as a White Road to OA, is currently working as a back-end service and facility to the front-line OA roads.

B. History

In 2000-2005 the OAPW has represented and supported the creation of a digital scholarly community, of its very infrastructure and first projects. Among these projects, it is worth mentioning the creation of a first digital journal, co-edited with the academic-oriented Publishing House Carocci in Rome and the Florence University Press, “LEA – Letterature d’Europa e d’America”; and the organization of a first training and self-training in digital scholarly communication undertaken by all the Department scholars. In 2005, the OAPW was enriched with a research project titled “Projecting and managing online contents in the Humanities: e-publishing and related websites”. This project was mainly committed to a first computerisation of literary-linguistic studies and saw the entire Department of Comparative Languages, Literatures and Cultures (with its thirty linguistic areas) collaborating towards the creation of a digital scientific community. With the enforcement of the new law on Education, started in 2000, and with the expansion of the European Community, the Department endorsed a series of new bilateral and multilateral curricula connected with other European Universities (Bonn, Budapest, Paris and Saint Petersburg). For these new curricula, the OAPW proposed the management of a professional training devoted to BA, MA and PhD students in the Humanities, in order to promote professional roles within the cultural work (in particular, online philology; best practice networks; strategies of CCPL; and the post-Gutenberg book).

C. The White Road to OA. Current and Future Projects

On the wake of the aforementioned initial projects, the OAPW can be described as a research and development unit within the Department of Comparative Languages, Literatures and Cultures. The OAPW is devoted to the advancement and accessibility of digital scholarly communication within the academic community. It is currently engaged in establishing a fourth road, the one we have named the ‘White Road’ to open access, running in the backstage to consolidate the Gold, Green and Platinum Roads by creating a social network of OA Publishing Workshops whereby it will be possible to develop and follow an integrated editorial model which the Florence OAPW is already pursuing within its very infrastructure. This integrated model is based on three main fields: OA research, teaching, and publishing. In the research area, the OAPW is carrying on studies and experiments in ‘digital communication’ and ‘publishing’ looking for new models of scholarly publishing where digital literary books can be supported and improved by reading and interpretation tools. Additionally, the OAPW works on the identification of good practices in copyright management in Open Access volumes, journals, and websites, in the Humanities Area of Cultural Heritage, with a particular focus on the academic authors’ attitudes and viewpoints. Good practices should contribute to increase awareness and understanding of copyright issues related to open access, among stakeholders in the scholarly communication area, especially among academic authors. The aforementioned research activities mainly follow two different routes: a) a theoretical perspective which conceives the cultural web as focused on literary and linguistic knowledge, as a ‘work in progress’ to be fully defined, organized, managed and turned to the scholar’s needs; b) a practical perspective which ensures its systematic planning, experimentation, and application of OA publishing as a political, cultural and legal entity, as well as an economically sustainable activity [17]. From 2005 to 2010 this double perspective has been consolidated.

With regard to teaching, the OAPW holds cultural-editorial training courses for BA, MA and PhD students, in collaboration with the Faculty of Humanities (by means of an Agreement of Professional Training). Within the Workshop, students attend a training period during which they are introduced to scientific and cultural concepts and activities related to the digital book (e-books), as well as to the concepts and activities related to the publishing workflow as performed in publishing houses. The training program consists of two phases. The first phase, called ‘initial training’, involves three types of activities: a theoretical introduction to the publishing world and system; a subsequent approach to writing programs and publishing softwares; a startup to the editorial workflow. During this phase, participants can study the Florence University Press different stylesheets for journals and volumes, in order to learn

how to apply these guidelines to the Department editorial volumes and journals, published within ‘Biblioteca di Studi di Filologia Moderna: Series, Journals and Workshop’ (BSFM, <www.collana-filmod.unifi.it>). They learn how to edit and proofread a paper, to hear the paper as well as to see it on the page. Students are invited to identify all the problems emerging during the different stages of the editorial workflow and to handle them. They learn how to work individually and as part of a team. During the second phase, called the ‘followup training’, students can test the knowledge they have previously gained, by editing one of the volumes forthcoming in Biblioteca di Studi di Filologia Moderna. They will work in different publishing teams according to their linguistic and cultural competencies. The followup program is constantly monitored by the OAPW Advisory Board and supported by its Editorial Board.

As to publishing, the OAPW is engaged in producing new and feasible digital editorial models and products, both experimental and traditional, accomplished and *work in progress*, for individual and/or collective use. All the volumes and journals edited so far by the OAPW within “Biblioteca di Studi di Filologia Moderna” and published by the Florence University Press (thanks to publishing agreements) are fully golden open access as to their content accessibility, and in accord with the Platinum Road since Department authors are never charged. Actually, Department authors, by signing an editorial agreement, have the chance to enter a specific body called “Collettivo di Autori”. By signing this contract, they accept to publish their scientific work under open access Creative Commons Public Licenses (<www.creativecommons.org>) and to undertake the Department peer-review process and procedures; they delegate the editorial management of their scientific work to the Advisory and Editorial Boards of the Department and are allowed to profit by the OAPW editorial services for free. It is worth noticing that the editorial infrastructure expenses are covered by the University of Florence, the digital publishing costs are absorbed by the Department of Comparative Languages, Literatures and Cultures, while printing on demand expenses are covered by scholars’ own research funds.

The OAPW is currently aiming to develop an institutional repository for the Department open access green products, built upon an open source software, in order to preserve the entire intellectual output of the institution. Furthermore, by means of a collaboration network of Open Access Publishing Workshops, which the Florence OAPW is attempting to promote and establish, it will be possible to develop a new web of institutional repositories implying that each institution will no longer have to work entirely on its own. The OAPWs consortium will provide an appropriate context for implementing oa editorial products and institutional repositories thanks to a mutual, joint, growth. Actually, the network could make economies, helping the different institutions

taking part in the project and running the White Road to OA, to share expensive technological systems, infrastructures and editorial softwares, as well as to distribute open contents and improve scholarly communication. A new survey on a shared OA publishing model able to enhance each workshop/institution own fields of research might be started, as well as a research project on the improvement of the aggregation between those cultural realities with fully or partly common objectives.

At present, the Florence OAPW is collaborating with weblearning system ‘Federica’, implemented by ‘Federico II’ University of Naples; with the Co-Lab of the University of Parma; with the Open Society Archive of Budapest and with Open Edition programme for open access to scientific journals in Paris. The Florence OAPW collaboration agreements are currently aimed to host, sponsor, and participate to a wide-range of European research, educational and cultural projects, and literature, art and media exhibitions and seminars; to promote the adoption of standards and best practices; to rely on proven open source technologies to create a unified digital content repository service; to make metadata and content available through Europeana. Such cooperation network of OAPWs on the model of the Florence OAPW, might prove to be a good road both to the creation of interrelated institutional repositories as well as to the promotion of golden and green products.

D. The OAPW Open Challenges. A Survey

“The existence of the means to create significant change does not mean that change will occur. The fact that electronic media exist has implications for the market. It is up to the players in the market to decide how they will use the means at their disposal. The dominance of the commercial publishers will be challenged only if the other players use the opportunities available to them” [18].

In its attempt to take scholarly communication and publishing in the Humanities towards innovation, the Florence OAPW is still facing a number of open challenges and obstacles while running the open access roads in its research, teaching and publishing services.

The Workshop is based on an open-source-like business model, in which universities, departments and research funds cover all the costs, at the moment. It seems that this model can prove to be successful, above all in the terms of the aforementioned network of workshops, but there is a number of restrictive conditions to consider.

First of all, Open Access Workshops need to be acknowledged by universities and OA bodies at national and international level. Such formal and concrete acknowledgement could catch the interest of various stakeholders, giving them greater confidence to experi-

ment OAPWs alternative publishing models, thus supporting its own research in fields like mobile publishing, e-publishing, and radio-publishing; e-books and e-journals; improved online reading; on-demand communication and printing; advertising profit sharing and database marketing. The type of business model *currently* used is suitable to academic and scientific environments with a moderate number of submissions, published papers, and volumes, and with limited requests for graphical artworks and special layouts, which would require higher levels of professionalization in terms of IT skills. What has still to be proved is whether this business model could equally be applied outside this environment, that is to say in environments with different business dynamics. A research in this field would be of great importance, since it would point out other possible environments that may be able to support this model and its potential needs for customizations.

On a more technical level, our experience has taught us that academic authors in general and academic authors in the Humanities, in particular, are very likely to rely more on 'traditional publishing'. They seem to be fully satisfied only with printed versions of papers. The challenge here is to promote and increase the value associated to electronic editions in comparison to the legacy printed one. In fact, the latter are usually considered as more trustable and complete than the former due to the presence of well-oiled processes and workflows, as well as to copyright laws. Actually, current copyright laws need to be revised and adapted to digital scholarly communication and publishing, supporting its legitimacy and validity. Another helping task would be promoting author-oriented publishing models where scholars are allowed to be actively participants in the publishing process, as well as in the look-and-feel of their volumes.

Additionally, open access journals and volumes in general have often been criticised for low quality standards of reviewing, low numbers of published papers and poor editing processes. The improvement in the practices and the standardization in both governance and processes promoted by the Florence OAPW is particularly aimed to overcome the aforementioned critics and will be more and more effective once joint to promotional initiatives like research projects, workshops, congresses and meetings focused on the oa topic and its peculiarities. As well, PhD students, post-docs and scholars *tout court*, being an integral part of the Florence OAPW, have to be supported in their editorial learning process and constantly updated in terms of their competencies. At the moment, the aforementioned need still remains partially unsolved due to a severe turnover of BA, MA and PhD students, who are allowed (both for institutional policies as well as for private economical reasons) to study and work within the OAPW only for a limited period.

Concerning research projects, the Florence OAPW has elaborated a number of projects focused on the OA

issue. Among these, it is worth mentioning the following projects whose unsuccessful outcome represents for us an open challenge:

1) Monte dei Paschi Foundation, Call for Proposals n. 12 (2009). Title of project: *The Open Access Publishing Workshop at the Department of Modern Philology: Research and Innovation in the Humanities Higher Education: didactic Technologies, multi-language and interactive Publishing, Master*. The project was submitted to Prof. Marinelli, Rector of the University of Florence by the BSFM Head-in-Chief as part of a call for proposal issued by the Foundation Monte dei Paschi di Siena.

2) Rectors Conference: Invitation for submission of university-based cultural projects on copyright (2009). Title of project: *Copyright and the Internet*. The project sees the collaboration of Firenze University Press (Drs. P. Cotoneschi, E. Brillì) and the Open Access Publishing Workshop at the Department of Comparative Languages, Literatures, and Cultures (Prof. B. Tottosy, Dr. A. Antonielli) as project coordinators working on "Copyright culture and the new publishing", also with the support of MICC (Media Integration and Communication Center, <<http://www.micc.unifi.it>>).

3) FIRB 2010 - "Futuro in ricerca", Action 2. Title of project: *Web and Open Access: Editorial Innovation and Research, and Teaching Humanities* (2010). The project aimed to develop an Integrated Open Access Infrastructure in the three key fields of academic discipline (research, teaching, and training) as well as a series of editorial workshops in cooperation with the Italian Digital Library. Partnerships: Casalini Digital Division, Cilea, ICCU, Univ. "Federico II" of Naples, Universities of Florence, Padua, Parma, and Pisa.

The unsuccessful outcome of the aforementioned projects was mostly determined by their undersized profile with regard to information and technology resources (both of personnel and equipment). Whereas many open access publishing platforms already exist (i.e. the Open Journal System) and are currently used by the Florence OAPW for its new journals (<<http://www.fupress.net/index.php/bsfm-sijis>>; <<http://www.fupress.net/index.php/bsfm-jems>>; <<http://www.fupress.net/index.php/bsfm-lea>>), we require a good work of implementation together with a strong investment in ICT. The greatest number of components of the Florence OAPW are experts in the Humanities and too few ones in ICT Technologies. Our challenge is therefore to overcome this inadequacy of researchers with specific ICT skills, which appears to represent a serious obstacle in achieving our major goals, and to improve ICT technologies and infrastructure. Such attempt to face the current problems by opening the Florence OAPW to ICT resources and scholars also includes another aspect: i.e. to make scholarly communication in the humanities free from its narrow academic circuit and able to achieve and stand

true and fruitful comparison with other disciplines. Such comparison would certainly prove to be very helpful in improving self-evaluation processes.

REFERENCES

- [1] Horizon 2020, Proposal for a COUNCIL REGULATION on the Research and Training Programme of the European Atomic Energy Community (2014-2018) complementing the Horizon 2020, Brussels: The Framework Programme for Research and Innovation, European Commission, 2011, p. 21.
- [2] See G. A. Chressanthis and J. D. Chressanthis, "Publisher monopoly power and third-degree price discrimination of scholarly journals," *Technical Services Quarterly*, 11, 2, 1993, pp. 13-36; M. A. Stoller, R. Christopherson, and M. Miranda, "The Economics of Professional Journal Pricing," *College & Research Libraries*, 57, 1996, pp. 9-21; D. P. Carrigan, "Commercial Journal Publishers and University Libraries: Retrospect and Prospect", *Journal of Scholarly Publishing*, 27, 1996, pp. 208-221; C. T. Bergstrom and T. C. Bergstrom, *The Economics of Scholarly Journal Publishing*, Washington: University of Washington, 2001, <<http://octavia.zoology.washington.edu/publishing/>>; R. E. and D. Shulenburger, "The High Cost of Scholarly Journals (And What to Do About It)," *Change*, 35, 6, 2003, pp. 10-19; L. C. Van Orsdel and K. Born, "Periodical Price Survey 2005: Choosing Sides," *Library Journal*, 130, 7, 2005, pp. 43-48.
- [3] J. Guédon, "Locating the Information Society within Civil Society: The Case of Scientific and Scholarly Publications," in *Communicating in the information society*, B. Gerard and S. Seochru, Eds. Geneva: United Nations, pp. 165-194 (here pp. 170-171).
- [4] CRUI-Commissione Biblioteche, Gruppo Open Access, *Riviste ad accesso aperto. Linee guida*, p. 6. My translation. Accessible online at: <<http://www.cru.it/HomePage.aspx?ref=1789#>>.
- [5] A. Okerson, "The Missing Model: 'A Circle of Gifts,'" *Serials Review*, 18, 1-2, 1992, pp. 92-96.
- [6] For further information on library consortia, see the Italian National Forum Electronic Resources (INFER) website, edited by Paola Gargiulo and Domenico Bogliolo: <<http://www.infer.it>>. Concerning consortia is also available an Italian and international Bibliography at the webpage <<http://www.infer.it/strumenti.html>>, titled "Gli strumenti: Bibliografie, discussioni, repertori, riviste, terminologia."
- [7] P. Suber, *Open Access Overview*, <<http://www.earlham.edu/peters/fos/overview.htm>>, 2004; S. Harnad, "Post-Gutenberg galaxy: the fourth revolution in the means of production of knowledge," *Public-Access Computer Systems Review*, 2 (1), 1991, pp. 39-53. Retrieved 20 July, 2007 from <ftp://ftp.princeton.edu/pub/harnad/Harnad/HTML/harnad9/post_gutenberg.html>.
- [8] Budapest Open Access Initiative, <<http://www.boai.org>>, 2004.
- [9] E. Hoorn, M. van der Graaf, *Towards good practices of copyright in Open Access Journals. A study among authors of articles in Open Access journals*, Pleiade Management and Consultancy BV, Keizersgracht, Amsterdam, p. 12. Retrieved 10 February, 2011 from: <http://micro010.ensp.fiocruz.br/~ensp/acesoaberto/attachments/104_Towards%20Good%20Practices%20of%20Copyright%20in%20Open%20Access%20Journals.pdf>.
- [10] The Objectives of the Zwolle Group 7 Principles can be consulted at: <http://copyright.surf.nl/copyright/zwolle_principles.php>.
- [11] C. Tenopir, "Online Scholarly Journals: How Many?," *Library Journal*, 1 February 2004. Available at <<http://www.libraryjournal.com/article/CA374956.html>>.
- [12] See "Economic implications of alternative scholarly publishing models: Exploring the costs and benefits", JISC EI-ASPM Project. A report to the Joint Information Systems Committee, 2007, p. 143. Available online at: <<http://www.jisc.ac.uk/publications/reports/2009/economicpublishingmodelsfinalreport.aspx>>
- [13] For other in-depth analysis and reports on the Platinum Route, see: Bo-C. Björk, Ž. Turk, "The Electronic Journal of Information Technology in Construction (ITcon): an open access journal using an un-paid, volunteer-based organization," *Information Research*, 11, 3, 2006, paper 255. Available at <<http://informationr.net/ir/11-3/paper255.html>>. J. Willinsky and R. Mendis, "Open access on a zero budget: a case study of Postcolonial Text", *Information Research*, 12, 3, 2007, paper 308. Available at <<http://InformationR.net/ir/12-3/paper308.html>>.
- [14] Available at: <<http://info-research.blogspot.com/2007/11/green-brass-and-platinum-three-routes.html>>. Here are a few examples of commercial authorpays models: Elsevier Sponsored Articles, Wiley Interscience Online Open, Oxford Open Journals, SAGE Open, Springer Open Choice, Taylor & Francis Journals: iOpenAccess. On the SHERPA/RoMEO website can be found a list of publishers using the hybrid models.
- [15] For an in-depth overview of the Open Access Movement in Italy, see: S. Mornati, *Open Access e repositories istituzionali*, corso Cilea, 2008. P. Gargiulo and M. Cassella, "Open access in Italy," *Open access in Southern European countries*. Lluís Als Anglada, Ernest Abadal, Eds. Madrid: FECYT, 2010, pp. 63-82. M. Guerrini, *Gli archivi istituzionali: open access, valutazione della ricerca e diritto d'autore*, a cura di A. Capaccioni. Milano, Editrice Bibliografica, 2010. M. T. Miconi, "L'Open Access in Italia: progressi e prospettive," *Rapporto sulle biblioteche italiane*, Associazione Italiana Biblioteche: Roma, 2011, pp. 149-158. G. Vitiello, "Open access, biblioteche e strategie italiane per i commons della conoscenza," *Biblioteche Oggi*, 28, 2, 2010, pp. 62-77. Available at: <<http://www.bibliotecheoggi.it/content/20100206201.pdf>>.
- [16] P. Galimberti, *Il progetto "Libri gratuiti in Ateneo" della casa editrice Polimetrica*, AIB Notizie, 6-7, 2008. Available at: <<http://www.aib.it/aib/editoria/n20/0616.htm3>>.
- [17] On this topic, see J. Houghton *et al.*, *Economic Implications of Alternative Scholarly Publishing Models. Exploring the Costs and Benefits*. A report to the Joint Information Systems Committee (Jisc). Available at: <<http://www.jisc.ac.uk/media/documents/publications/rpteconomicoapublishing.pdf>>.
- [18] W. Trust, *Economic Analysis of Scientific Research Publishing: A Report Commissioned by the Wellcome Trust*, Histon, Cambridgeshire: SQW Limited, 2003, p. V.

Two sides of the same coin

Access to archival collections, a case study on HOPE

Gabriella Ivacs⁽¹⁾, Kathryn Mathe⁽²⁾

(1) Central European University, Open Society Archives (OSA), Budapest, Hungary, ivacs@ceu.hu

(2) Central European University, Open Society Archives (OSA), Budapest, Hungary, mathek@ceu.hu

Abstract—Short and long-term access to archival records is socially and culturally significant in the digital environment. New licensing frameworks and austere policies can often make conditions for the re-use of material unmanageable for archival curators. Legal uncertainty and restrictive regulations may jeopardize the knowledge ecosystem by limiting access to information; a thorough analysis of the new environment has become increasingly imperative. The challenges to developing and implementing policies with appropriate levels of control must be identified and debated by an array of stakeholders. Institutional and national settings differ significantly across the archival domain and so do the challenges and barriers that have emerged. These processes are complex and, in parallel with the advancement of technical expertise and investment in digitisation, require serious rethinking as they affect the role of archives and their relationships with citizens.

long-term access; short-term access; copyright; rights management

I. OPEN SOCIETY ARCHIVES AND THE HOPE BEST PRACTICE NETWORK

The Open Society Archives (OSA) at Central European University functions as a research institute, an archival laboratory, and a historical repository in the traditional sense. While actively collecting, preserving, and making openly accessible documents related to recent history and human rights, OSA continues to experiment with new ways to contextualize primary sources, developing innovative tools to explore, represent, or bridge traditional archival collections in a digital environment. Our approach to acquisition is increasingly proactive and inclusive, and we actively seek out non-traditional material, material previously marginalized based on its content, social origin, or form. Through all of these endeavors, OSA advocates: open access and transparency; equal rights to information; the ethical use of private data; open formats and open standards; and broad access to cultural heritage.

OSA is currently participating as a research partner and content provider in the three-year European Union project entitled the Heritage of the People's Europe (HOPE). HOPE is a Best Practice Network of archives, libraries, and museums across Europe that aims to improve access to highly significant but scattered digital collections on social history. It will achieve this goal by promoting the adoption of standards and best practices for digital libraries amongst its partners, by ensuring that

the metadata and the content become available through Europeana and LabourHistory.net, and by implementing a full scale discovery-to-delivery model. HOPE meets Objective 2.2 of the European Digital Library ICT-PSP programme—i.e. to increase the quantity of quality content available through Europeana.

II. ARCHIVES IN THE DIGITAL LANDSCAPE

It has been more than decade now that Open Access (OA) has been successfully dismantling the restricted domain of scientific publication in order to enable free, online access to scientific knowledge and make it available to a wider audience. (Here, the word “free” does not mean gratis, it suggests rather liberty; as Richard Stallman defined Free Software: “free” as in “free speech”, not as in “free beer”.) The OA movement has rightly highlighted technological, legal, and administrative constraints on the accessibility of research literature and has defined copyright as the primary barrier to distribution and reproduction. The movement has reduced radically the role of copyright in an attempt to give authors control back over their works, the right to be acknowledged, cited, and republished. As we mark the 10th anniversary of the Budapest Initiative, it is time to consider the limitations of the model: primarily why the model fails in the case of unpublished primary source material. Archival collections, manuscripts, public or private records on recent history—often digitized but still locked in institutional repositories—are not being researched and reused in the broad sense of the term.

In some sense, we are all prisoners of a paradigm in which archives are confined to the role of trusted custodians of proprietary records, guardians of history, curators of provenance, whose interests lie in preserving authenticity and integrity in a paper environment. Access, with or without the internet, is not an imperative. It is true that, with a few notable exceptions, archives have been reluctant to take advantage of new digital technologies, creating comparatively little digital content and few tools to provide meaningful resource discovery within their collections. It would be easy to characterize archives as “secret chambers of government files” and dismiss archivists as “gatekeepers of information”, but this belies the very real efforts made by archives over the past decades to respond to changing needs and expectations [1]. The rising demand for alternative stories, private histories, and documents of everyday life, saw archives broaden collecting policies to include ephemeral artifacts, grey literature, sociological data, and

ordinary objects. Many, like OSA, expanded their mission, moving away from a strict focus on provenance, instead seeking to proactively document social phenomena.

So, archives have shown a general readiness to expand their mission to fit changing circumstances. Then what are the obstacles they currently face when endeavoring to provide access to their collections in the new context? Does lack of use signal a broad lack of interest in content increasingly seen as irrelevant or have archivists simply been unable to manage access when confronted with the networked models of work, transnational technologies, and obtuse licensing frameworks?

III. HOPE AND ACCESS: THE FEDERATED MODEL

In the terms of the OAIS Reference Model [2][3], HOPE is a Federated Archives. In the Federated Archives model, “several OAIS archives provide access to their holdings via one or more common finding aids.” Archives provide dissemination information packages to a Common Catalog. (The Common Catalog may serve solely as a finding aid or may also provide common dissemination of locally stored Products—i.e. including digital content.) In more prosaic terms, this means that each of the thirteen institutional content providers, including OSA, has agreed to submit descriptions (with or without digital content) in a standard format to a common system, which in HOPE is called the Aggregator. In HOPE the model is more complex than the standard OAIS federated model in that the Aggregator cannot be accessed directly through a search portal by end users. Instead, the Aggregator, based on D-NET [4] technology, pushes descriptive information and content to the various discovery services and portals listed above.

The OAIS Federated Archives model is at its essence a model for increasing and enhancing access to dispersed collections by increasing interoperability among disparate systems and data sets. In HOPE this is accomplished through: a data model which maps and integrates several domain standards (library/MARCXML, archival/EAD, and visual/LIDO) into a common descriptive schema; the normalization of key value types, e.g. country, language, date, content type, and digital rights; shared authority files for names, places, and topics; and the support and management of multilingual data (a content provider may provide the same metadata in several languages to enhance access to multilingual material). Access is further enhanced by a content policy that calls for the creation of HOPE Themes, a common set of social history thematic terms that can be assigned to aggregated collections, as well as of collection descriptions based on DCCAP, which provide a uniform entry point to items described according to various, and often quite opaque [5], domain standards. Finally, long-term access and interoperability are both facilitated by the mandatory use of PIDs for all digital content and descriptive records. Through the above measures, the project has confronted many of the access issues inherent in cross-domain and cross-

language thematic collections, such as that proposed by the HOPE consortium.

The actual dissemination of content to discovery services is controlled through general dissemination profiles for each discovery service. Dissemination profiles allow the Aggregator to filter content to selected discovery services based on data already present in submitted descriptions, e.g. content type, available formats, and access and use rights. As envisioned, content providers will eventually be able to fine tune the dissemination of their collection metadata by overriding default dissemination values at institution-, collection-, or item-level. For example, an institution could decide against providing any of its content to YouTube; or to provide thumbnails but not access copies of a certain collection to Flickr; or to supply metadata only to the LabourHistory.net portal for a selection of items. In the context of HOPE, the dissemination profiles have the potential to express refined institutional dissemination policies—but this is not equal to regulated access.

How then does the network propose to administer the rights to access and use submitted material through its service? In the case of HOPE, the situation would seem relatively straightforward. The current content policy stipulates that digital collections submitted to HOPE should already be “freely accessible” (in practice, this means that the institutions must make available a so-called access copy, suitable for viewing or listening but not for high-quality reproduction) and copyright cleared (content should be either in the public domain or cleared via licensing). The Aggregator simply accepts all submitted descriptive content, including links to access copies and thumbnails of digital content, under a policy of open access. Copyrights and other use rights over digital content are expressed by content providers through an array of European controlled rights values [6]. These include creative commons (CC) licenses and other rights statements that can be interpreted by the Aggregator and/or exported and expressed directly by discovery services such as Europeana and the LabourHistory.net portal. As is clear, the current content policy was developed primarily to meet the requirements of the Europeana discovery portal. This was a necessary starting point that allowed the project to focus on relatively simple aims: the creation of default policies on freely accessible content with relatively clear sets of rights held and granted.

IV. HOPE AND ACCESS: LOCAL PRACTICE

From a local standpoint, the situation as it currently stands would also seem simple. HOPE institutions need only pre-select freely accessible content for inclusion into HOPE and assign Europeana rights values. Users wishing to see content will follow links back to local sites, where institutions may regulate access to digital content and deliver as they do for local users. The scenario is true to a point. The seed content submitted to HOPE under the terms of the three-year project is by its very definition of limited quantity and clear of complex rights-related issues. Satisfying HOPE requirements is simply a matter of hand-picking and supplying the

promised content with appropriate digital rights statements—added most likely at the point of export to avoid the additional complexity of storing and managing these values internally.

This scenario, however, belies the underlying realities which archives and other repositories now confront—realities that will also encroach upon content providers as they set up their systems for HOPE compliance over the long term. While such informal procedures may be perfectly appropriate for ten or so digital collections, how would these procedures scale as submission of content to HOPE is integrated into an institution's internal workflow? How could such procedures be adapted if the institution opts to join other federated archives? As awareness about data protection and use rights grows among creators, donors, service providers, researchers, and the general public, how could informal procedures help archives make sense of and balance the needs and rights of various stakeholders—including but not limited to HOPE and Europeana—in their collections?

In fact, HOPE (and other such federated archives) may prove to be one of many forces compelling archives to regulate access and use in a manner that is at once more granular and less “concrete”, less based on the physical control of analog originals and more in tune with new realities where content must be managed in several formats, migrated, and made available over the long term. Yet several obstacles still remain.

Number of stakeholders: In contrast to academic and public libraries, which participate in fairly routine publishing, licensing, and distribution chains, the sheer number and varied types of stakeholders for a given archival *fonds* or document is an added burden for archives attempting to assess and express rights over their content.

At the heart of the matter is the fact that many institutions are hesitant to stir up trouble—suspecting, probably rightly, that most stakeholders are either uninformed, uninterested, or simply impractical to negotiate with. The donors themselves are often unaware of full range of stakeholders and fail to provide adequate legal provisions in donation/deposit agreements. This is particularly true for content that was created when copyright was less monolithic and proscriptive. And the issue is further complicated by concerns, both legal and ethical, for the privacy of individuals mentioned in non-published works. The digitization and online publication of content has only brought these problems to light.

Lack of legal provisions: Related to the above, many archives function under outdated donation/deposit agreements, which lack the provisions needed to support the archival mission to preserve and provide access over the long term. As archives struggle to update older agreements and develop new ones, they feel the lack of standard legal clauses empowering them to curate data under the new technological norms—migrating formats, creating derivatives tailored to use, and actively disseminating, rather than passively “giving access to”, content.

Archival workflow: Covering a wide range of material of varying type, quality, and provenance, archival description must strike a delicate balance between the whole and the parts. Donation/deposit agreements generally apply to a whole set of donated materials and can be difficult to apply in a granular manner. This is exacerbated by the accumulative nature of archival accession; agreements may cover large sets of ill-defined and heterogeneous material that will not be in the possession of the repository for many years.

The need to maintain integrity between the various levels of hierarchical description proves an added obstacle. And inheritance rules, which pull data to higher levels, have worked against a high level of specificity—even for data related to access and rights. By shifting the focus to individual documents, digitization is beginning to upset established practice.

Domain standards: The widespread adoption of domain-specific descriptive standards (ISAD(G), MARC, VRACore, etc.) has been one of the success stories of recent years and has been viewed as a boon for access to archival and other cultural heritage content. Yet archival descriptive practice is still motivated by the physical control and ordering of the analog “carriers” of content. In the past, the formidable physical and financial barriers to accessing original analog content *in situ* made robust access and use management unnecessary. As a result, domain standards require limited structured metadata on access and use rights; even today machine-readable data is scarce.

Here too, digitization and online publication have presented a challenge to the *status quo*. In recent years, the digital repository community has made moves to redress the situation. Digital content standards such as METS and PREMIS have been developed with special sections dedicated to digital rights management, filling the void left by descriptive standards. This has unfortunately created a false dichotomy between digital and non-digital formats of the same work that only complicates attempts at full-scale rights and access management.

Licensing options: The current options for licensing content for sharing and re-use are primarily designed to serve traditional publishing and scholarly communication networks. Thus far, there has been very little work to develop standard licenses to fit the needs of archives and other custodians of unpublished non-proprietary content. As more material is presented online, the need for such licenses will surely increase.

When taken together these obstacles, several rooted deeply in archival practice, have hindered efforts at full-scale rights and access management—particularly in small- and medium-sized archives and repositories. Federated archives like HOPE offer new possibilities for integrating and connecting collections, broadening the user base, disseminating and exploiting content, and delivering materials in a range of formats and quality through a suite of fully automated procedures. But in order to fully realize these possibilities, local content

providers must first tackle the challenges at their doorstep.

The HOPE project has skirted such issues thus far by accepting content that is clear of legal complication, suggesting that beyond this content providers are responsible for their own access policies. But as the initial phase of development comes to a close, seed collections are harvested, and basic project requirements have been met, the HOPE content policy is sure to expand. HOPE should be prepared to guide content providers through best practices which support robust rights and access management. These may include: model legal clauses empowering institutions to carry out preservation, dissemination, and delivery activities; a broad-based “opt out” policy enabling institutions to disseminate “orphan” works (including those where stakeholders are numerous or unclear); guidance on the full-scale application of Europeana rights metadata in local systems; and an analysis of existing rights and access schema as they apply to analog, digital, and multi-format collections.

From their side, HOPE’s content providers should move to introduce robust access and rights management into their repositories. The issue should be addressed through the entire archival workflow. First, donation agreements and other licenses should be updated to ensure that archives can actively curate and disseminate the content in their care over the long term. Legal clauses should focus on function rather than form to support unforeseen technological innovation. Local copyright and privacy legislation should be clarified and an attempt should be made to pin down stakeholders in all collections. Internal policies on access and content sharing and re-use should be drafted and collections with exceptional status identified. Institutional “opt out” policies should be considered and, if necessary, applied. Second, the rights thus established should be captured, ideally, through the use of existing rights standards such as METS RightsDeclarationMD or PREMIS Rights—though these would need to be extended to cover the full range of analog and digital content and derivatives. Permissions along with embargoes and use constraints (e.g. non-commercial, attribution, etc.) should be captured at the highest level possible but stored at item-level granularity. Finally, such permissions should regulate the transformation of “archival information packages” to “dissemination information packages” (in OAIS terms), ensuring that content is disseminated in an appropriate form to different systems and services. Standard licenses and rights statements with attribution clauses, such as the Europeana Rights values, should be exported for display near descriptive metadata and thumbnails. Only by undertaking such full-scale rights and use management, can the HOPE mission to create a full-scale discovery-2-delivery model be realized.

V. CONCLUSION

The term *access* has been heavily inflated over the past few decades; its connotation varies through disciplines, domains, expertise, culture, and institutional

settings. Archives have a rich tradition in managing access within their domain [7], but institutional practices nowadays tend to be measured against external baselines: legal regulations, digital technical standards, and transatlantic professional community practices, which do not respect geographical borders or domains. Archives are interdisciplinary by nature, and are at becoming more and more explicitly so.

Access to cultural heritage collections across the spectrum has been hampered by the complexity of copyright licensing, a lack of legal certainty about educational and other non-commercial use, tedious procedures regarding works of unknown origin, and the prevalence of deep-seated cultural and linguistic barriers. Donor agreements, contractual provisions, statutory frameworks, and ethical concerns can overshadow research interests even in case of public domain materials [8]. The appearance of digital technology has brought these matters to the surface by threatening to dismantle the formal and operational methods hitherto used by archives to regulate access to and use of collections.

In this paper, we have endeavored to link high level policy concerns about access to and reuse of digital archival collections with concrete access management issues. Current economic realities and the drive for long-term sustainability compel medium and small archival institutions to use common cyber infrastructures in order to reduce technical costs. Access to archival material increasingly rests upon the active dissemination of content to the services frequented by target users. In both respects, HOPE is an exemplary model. However, as HOPE pushes archives beyond the state of the art, it also exposes underlying problems related to access and rights management. Archives themselves must respond.

REFERENCES

- [1] T. Cook, “Viewing the World Upside Down: reflections on the theoretical underpinning of archival public programming,” in *Archivaria* 31 (1990/91), pp.123-134.
- [2] Consultative Committee for Space Data Systems, “Reference Model for an Open Archival Information System (OAIS)”, CCSDS 650.0-B-1 Blue Book, January 2002.
- [3] B. Lavoie, “Meeting the challenges of digital preservation: the OAIS reference model,” in *OCLC Newsletter*, No. 243:26-30 (January/February 2000).
- [4] P. Manghi, M. Mikulicic, L. Candela, D. Castelli, P. Pagano, “Realizing and Maintaining Aggregative Digital Library Systems: D-NET Software Toolkit and OAISter System”, in *D-Lib Magazine* 16:3/4 (March/April 2010).
- [5] E. Yakel, “Encoded Archival Description: are finding aids boundary spanners or barriers for users?” in *Journal of Archival Organization* 2:1–2 (2004), pp. 63–77.
- [6] Europeana, “Guidelines for the rights in objects submitted to Europeana: selecting a value for the europeana:rights and edm:rights metadata element”, 9 February 2012.
- [7] International Council on Archives Committee on Best Practices and Standards, Working Group on Access, “Principles of Access to Archives”, 2011, unpublished.
- [8] I. Hargreaves, “Digital Opportunity: A review of Intellectual Property and Growth”, 2111, available at www.ipa.gov.uk/ipreview-finalreport.pdf

The Co-Laboratory: a tool for research and education at the University of Parma

Anna Maria Tammaro⁽¹⁾, Sara Valla⁽²⁾ Elisabetta Longhi⁽³⁾

(1) University of Parma. Department of Language, Parma, email: annamaria.tammaro@unipr.it

(2) University of Parma. Co-Lab, Parma, email: sara.valla@unipr.it

(3) University of Parma. Department of Language, Parma, email: elisabetta.longhi@unipr.it

Abstract— Co-Lab proposal at the University of Parma is meant to create a virtual organization, starting from Humanities and Social Science Faculties, to foster quality concerning education and research through a collaborative approach. The problem is not about IT tools, already available to use, but the creation of a mindset and stimulation of the existing infrastructure to improve results. Collaboration and the development of a project ground for everybody might upgrade learning performances inside the University. Educational frameworks such as Masters and courses have been reinterpreted as collaboration experiences and methods to gather actors have been designed through an experimental environment. Interviews to teachers and students are used to tune up the type of service needed. Organization, technology and knowledge are considered as entangled and all necessary to Co-Lab development.

Collaboration laboratory; educational resources; problem solving; experience sharing

I. BACKGROUND AND PURPOSE. THE PROBLEM.

In present knowledge society learning and educational methods and ways of thinking are basic and important elements, together with techniques for knowledge creation.

We are facing a situation that shows a spontaneous and widespread use by citizens of network and technologies such as dialogue and social environment, for confrontation and opinion exchange.

It is necessary to consider at the same time technologies, knowledge and culture for learning [1].

The University of Parma invested human and financial resources in the education- and research-supporting infrastructure following the current national trends, and made services and tools available for professors and researchers to innovate or improve education and research.

Such an infrastructure facilitates the creation of digital contents and the access through user interfaces.

The University of Parma was one of the first ones to adopt distance education by adhering to the Progetto Nettuno since its beginning. Nowadays there are two organizational structures devoted to the technological support to the University professors: SITA and CEDI. SITA, formerly Centro di Calcolo, includes a Service for the Support to Education and E-Learning that maintains Moodle (LEA) and webinar software, and manages

YOUUnipr video server for video files produced by University personnel and organizational structures.

The CEDI, Centro Didattico di Ingegneria, was born from a previously existing Laboratory and aims to satisfy specific needs concerning services for Faculty professors.

The University of Parma also offers an Open Archive performing on a web server hosted by CILEA and uses DSpace, with policies defined by the University Senate that make the insertion of the results of research compulsory as far as doctoral thesis are concerned.

Educational experiences have been already carried out in the last years at University of Parma to create educational resources exploiting technologies in order to support and enrich lectures with innovative methods for individual study, review and knowledge verification.

Literature on e-learning and open access underlines that some obstacles for online or blended learning diffusion can be teachers' lack of technological expertise, sustainability and shortage of resources, including sometimes the quality of products, lack of awareness of the infrastructure, copyright issues and persistence of a conservative approach. Such problems are also present at the University of Parma; in spite of the availability of some updated IT applications and tools, resources are often not enough to supply an organizational support to teaching.

II. THE POSSIBLE SOLUTION

The proposal of Co-Lab rose up inside the Faculty of Arts and Philosophy among a group of teachers and experts, sensitive enough to information technology innovation and willing to solve precise teaching problems: it was to give birth to a sort of informal co-operation which took the name of the digital Co-Lab Unipr (Co-Laboratorio Digitale dell'Università di Parma).

The project inspiration lies upon Licklider, Engelbart and Borgman's work, and aims to explore all possible opportunities to increase the quality of research and University educational activities by using computer technologies for professor driven activities.

Literally, the Co-Lab is a laboratory created to foster collaboration, in order to experiment the qualitative extension that technological tools allow, and increase student learning and the quality of research.

The Co-Lab focus is therefore on students, who are part of the Laboratory as real actors and not only passive characters; the range of activities also includes support for digital publishing, starting from an investigation on the evaluation of the results of research.

Data updated at July 2011 show that professors and experts from the Faculty of Arts and Philosophy are those, together with Faculty of Engineering ones, who are mostly using learning environment at the University of Parma (35,70% of the total online courses).

The team assumes that listening to needs and the evaluation of educational experiences made online by students and experts can bring significant improvements of teaching and research performance without necessity for particular investments.

Digital Co-Lab has been founded upon 3 basic principles (3C): COLLABORAZIONE (Collaboration and Co-Operation), CONDIVISIONE (sharing of techniques and methods, environments, software and contents - Open Access), CREATIVITA' (creativity and creation) [2].

Our purpose is using e-collaboration style to discover and exploit opportunities offered by IT tools, the Internet and network to reach goals together with others [3].

III. METHODS, MATERIALS AND PROCEDURES

Co-Lab activities are graphically portrayed in Fig. 1; the Agenda lies upon an experimental methodology. We start from the problem to be solved, or the activity to be fostered through a sensible use of the available technical tools. We then agree on purposes to be reached, that are definite and measurable, in order to carry out a step-by-step monitoring during experimentations.

For each activity professors and technicians keep an updated diary.

At the end of each activity the evaluation of the results leads to a collective reflection on all aspects to be improved.

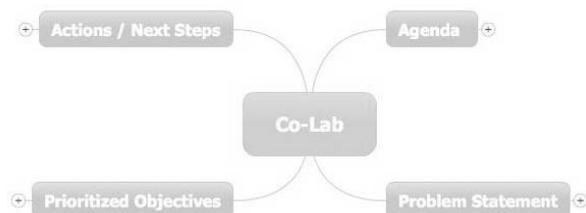


Figure 1. Co-Lab methodology

Our team started laboratories and seminars to involve students, professors, teachers, experts and researchers.

The activities carried out until now include: two international Masters (the Master DILL, the Master

METAV), the Seminari di Informatica Umanistica, the Workshop Futuro del Libro.

We observed activities and events during their sequence, trying to give advice and support in real time. Observation was carried out both through tutor presence and the use of different methods of communication (Facebook groups and pages, use of Moodle forum activities).

As an example of the CoLab activities, we can describe the Co-Lab support given to the METAV Masters. We carried out qualitative interviews asking for the feedback of professors who were involved in METAV Master, concerning pros and cons of the experience, their idea of course structure, possible improvements and a creative use of activities and resources they would suggest.

We made proposals for a creative use of tools and systems in order to give professors the opportunity to choose some activities and methods that could be used.

The proposals included a creative usage of user roles and features inside Moodle Learning Management System, in order to create an ideal environment for interactions and the birth of a community.

Interviews to Master course professors.

Structured interviews were very useful as a qualitative methodology, to determine the satisfaction degree of professors concerning the currently installed Moodle Platform, and their ideas concerning best practices and future creative educational activities.

Most professors asked for additional tools to show things to students and interact with them; that shows that they were not aware of webinar services at University.

Many of them asked for applications that are in fact already available on the platform; this means that they are not conscious of all the possibilities that available platforms offer and that it will be necessary and useful to pursue a slightly different point of view concerning service level and approach.

All interviewed experts underlined the importance and effectiveness of the social format for Moodle courses, and an active participation in forum activities. This can give evidence that a social approach inside courses can help to reach student better involvement.

The majority of professors asked for a more functional and flexible version of the platform, as concerns file management and tools to increase interactivity and co-operation.

The opinions we gathered from interviews convinced and led us to the installation of an alternative version of the LMS platform for Co-Lab team, in order to test, together with teachers and students, the advantages of Moodle 2.x releases.

Another example of the Co-Lab support activities can be evidenced by the openEyA laboratory.

We started a laboratory for the creation of online educational competences and the development of specific methodologies and contexts, as the use of multimedia contents is concerned, and a new way of cooperation with Science Dissemination Unit of International Centre for Theoretical Physics; we are testing openEyA (www.openeya.org), that is their open source solution for lecture and events recording [4]. Co-Lab team organized some tutorials for experts, teachers and students, in order to spread the voice and get users' opinions. Tutorial sessions were also practical so that participants could join and test the solution immediately.

Both teachers and students showed great will to learn how to use such tools and students started recording lessons and seminars together with teachers, and interviews to professors.

Next steps will involve experimentation on mobile devices (Apple iPad), and the investigation to find alternative suitable solutions to experiment, in order to offer open resources obtained by recordings carried out with EyA system.

An other important activity of Co-Lab is a continuous training support: MIXMeS laboratory, an course to co-operate on real projects

Considering the unconference model, where meetings are driven by participants, we proposed a project driven learning event, a learning by doing and cooperative learning experience, where voluntary experts, teachers and researchers can join and decide together the projects they want to work upon; after a survey to investigate attitude and concepts on IT tools and their usage level among future participants, a brainstorming session to identify interesting topics and projects and a training period on the Co-Lab online Learning Platform, participants start working on educational and research projects, acquiring competences and skills concerning the right choice for the specific context and needs. We decided to adopt a plurality of methods and tools, to be coherent with project purposes: in order to choose the best tools and methods participants will have to test and verify them in specific situations and contexts. Participants are going to use and test webinar and chat tools, as Adobe Connect, Google+ hangouts and Skype. They will use social and e-collaboration environments, such as video and image sharing platforms (iTunes-U, YouTube, YOUNipr, Moodle, Vimeo, Flickr), online scheduling tools (Doodle, Google Calendar), Social Bookmarking and Social Reading activities to pursue social learning and will explore the opportunities offered by mobile learning in order to understand if students and teachers are ready for mobile learning and teaching [5] [6].

The laboratory has started in January 2012 and is on the way. The pilot project is involving participants from the Department of Foreign Languages of Faculty of Arts and Philosophy.

IV. FINDINGS AND RESULTS AT THE PRESENT MOMENT

The starting idea sprung up last year and nowadays, after a one-year activity, the Digital Co-Lab managed to make the most of some existing initiatives and aggregate professors by spreading criteria and teaching innovative methods; the Co-Lab team gave all those who were interested in experimenting in a specific education and research field, an effective support to enhance teaching. We are now gathering further data concerning experimenting in educational frameworks as Master degrees and proficiency courses, as the Digital Librarianship Learning Program and the Master degree in audio-visual media translation.

The Co-Lab is showing efficiency in the solution of the starting problem the idea arose from, that is stimulating the use of the existing infrastructure to improve results concerning education and research carried out at the University, starting from Humanities and Social Science Faculties.

The challenge is starting from where the professors are, taking as a base for activities daily problems to be solved, and stimulating collaborative work and the sharing of experiences as a method.

The methodology chosen by the Co-Lab team, that starts from problems to reach defined and evaluable objectives, is the one that can keep the promise of technologies and make it real, that is an extension of human possibilities to foster an overall improvement of the quality of the University as a whole.

REFERENCES

- [1] C. Hutchison, Cultural constructivism: the confluence of cognition, knowledge creation, multiculturalism, and teaching. *Intercultural Education*, Volume 17, Issue 3, 2006, pp 301 – 310
- [2] E. Longhi, F. Monaco, A.M. Tammara, M. Valero, S. Valla, "Un Co-Laboratorio presso l'Università degli Studi di Parma per la creazione e l'accesso di contenuti didattici multimediali", in *Minerva T., Colazzo L. (a cura di) Atti VIII° Congresso Sie-L (Società Italiana di E-Learning) 2011, Connessi! Scenari di Innovazione nella Formazione e nella Comunicazione. Reggio Emilia, 14-16 settembre 2011, Edizioni Ledizioni LediPublishing*, pp. 561-566
- [3] S. Sancassani, F. Brambilla, P. Menon, P. Marengi, e-Collaboration - Il senso della Rete - Metodi e strumenti per la collaborazione online. 2011, Milano: Apogeo s.r.l.
- [4] E. Canessa, C. Fonda, M. Zennaro, Webcasting of Traditional Chalkboard Lectures: The EyA System, *EURODL European Journal of Open, Distance and E-Learning*, 2007, Retrieved November 25, 2011, from http://sdu.ictp.it/eya/papers/Canessa_Fonda_Zennaro.html
- [5] S. Valla, I. Comelli, YOUNipr: a 2.0 video server. A video distribution experience at Parma University, *Research, Reflections and Innovations in Integrating ICT in Education - Vol 3*, 2009, pp. 1508-1512
- [6] J.R. Corbeil, M.E. Valdes-Corbeil, Are You Ready for Mobile Learning?, *EDUCAUSE QUARTERLY*, Number 2, 2007, pp 51-58

Opening up culture may not be sufficient: looking for a policy

Rosanna De Rosa ⁽¹⁾

(1) Department of Sociology, University of Naples Federico II, Naples, Italy, rderosa@unina.it

Abstract— Starting from the experience of Federica - the weblearning platform of University Federico II - this paper aims at promoting a general debate around the open education practices moving from some critical points: the lack of a strong political and academic commitment, the limiting effects of the copyrights laws, the still difficult access to scientific and cultural resources. Last, we will show – for the first time – the 3D Federico II Campus with its historical architectures and information, cultural resources and educational material. A new immersive environment for resource discovering which will be soon available for the whole public.

Keywords: *weblearning; Open culture; Federica web learning*

I. INTRODUCTION

Distance learning programs have taken on a more significant role in the way educational institutions deliver their courses, and have become an integral part of university planning strategy.

After an initial experimental phase, the introduction of new technology into educational processes gathered momentum thanks to three main factors: the development of broadband, the evolution of the consumer market with low-cost high technology products, the expansion of the digital contents market. In 2011, 77% of the 1000 American universities and colleges were already offering distance learning courses. Online courses are offered not only to make learning more accessible to certain types of students, but also as an integral part of degree courses. And of these three-quarters of American campuses offering online courses, over 58% offer ones which are completely distance learning. 71% of these are profit-making institutions. Growth forecasts in the sector are very positive, based on the exponential development of electronic publishing (tablets, ebooks, digital libraries) [1]. In Europe in 2011, 32% of the population consulted the web for self-study. Which is ten per cent more than in 2007. 23% used the web to get information and guidance on courses and study opportunities, which is 4% more than in 2007. The use of Internet to actually follow online courses is still in its infancy, though even this group presented an increase of 2% compared to 2007. Finland, Spain, Lithuania, Sweden and Great Britain lead the way with the highest percentages with 14% in Finland and 7% in Great Britain [2]. In Italy, online self-study stands at 35.5%, study orientation at 26 % and following online courses at 4%. The number of people who use the web

"to look for information about learning opportunities and courses in general" is 36% overall, and 51% in the 18 – 24 age range [3].

These figures say a lot about how young people use the web to aid their learning. And their approach clearly indicates that where they learn is now more important than how. OER are central to this scenario.

II. OER: SO MUCH POTENTIAL, SOME LIMITATIONS

Accepting the principle of OER means implies a complex idea of an integrated media and information ecosystem that is created when ICTs are developed as part of knowledge transfer. This ecosystem was designed as a seamless environment for access to open content and, as such, should encourage and guarantee the use of OER for learning and teaching. Nonetheless, despite huge efforts on the part of institutions, the scientific community and professional organisations to open up the bedrocks of knowledge and make them available to everyone, there is no empirical evidence from the world of learning to show how these resources are actually used.

One report from the *Open eLearning Content Observatory (Olcos)* [4] suggests that fundamental weaknesses mean that few people actually make use of available OER. Three factors in particular are thought to have a negative effect on the creation, development and use of OER.

The most significant factor is the lack of formal recognition for the educator. In more general terms this can be ascribed to the absence of any real political or institutional commitment to OER and the absence of a real incentives system. In other words, Academic Authorities do not see Open Educational Resources as a priority where investment is concerned.

A second important factor, which may or may not be connected to the first, is the lack of funding for OER projects and more generally speaking, the lack of any long-term plan that sees the development of OER as a strategic move for improving any university or cultural institution's reputation and position on the international stage. OER's lack of a business model and its poor institutional uptake is something everyone is aware of.

The last crucial, critical factor identified in the report is the failure to concentrate on organisational aspects, especially where support (logistic, technological and financial) for the community of practice is concerned.

The creation of educational repositories is not an end in itself. In fact it only serves to highlight the difficult relationship that exists between communities of practice – educators – and making use of available cultural and organisational resources. It also opens up debate about the interaction between OER and the educational context they belong to. One of the most important things, therefore, is the relative competence of educators, and how willing they are to include materials written by others in their own learning pathways, motivated by the simple desire to help people learn. This is not something that can be taken for granted. It requires new types of competencies and a special kind of mindset. [5]. As the Olcos report points out: «open access to resources is an important element in educational innovation, but not the only solution per se. The decisive factor is that open educational practices are fostered by the appropriate institutional culture and mindset and supportive environment, including easily accessible and shareable tools, services and content» [5].

In other words, OER has enormous potential to change the way things are, but this remains largely unexpressed. They could contribute to making educational processes more transparent and improving the quality of education, as they lead to a real re-definition of the role of the teacher and their ability-opportunity to co-construct knowledge, within a policy framework designed to encourage their use and institutionalisation. The role of the institutions is crucial in this regard, as is illustrated in the recent Unesco-Col *Guidelines for Open Educational Resources in Higher Education 2011* [6]. The guidelines suggest supporting the use of OER through revision of policies regulating higher education, aiming to increase awareness of key issues in OER, and then reshaping connection strategies to enable schools and their operators to successfully access online resources before finally adapting national legislation to fit in with open licensing. Adopting open standards, as is already the case in public administration, in conjunction with specific investment to make sure that people have the right training to produce learning support, can lead to the creation of institutional repositories at all levels. The Italian path towards Open Educational Resources continues to be blocked by an over-cautious legislation which, while allowing for the use of “downgraded” images and videos from the net, then gives the groups representing the publishers the power to grant permission to use their images or not, thereby nullifying the principle of openness inherent in the law.

III. THE FEDERICA MODEL

The most common approach in university e-Learning projects is to focus on the development of software packages and ICT/Web platforms, leaving out

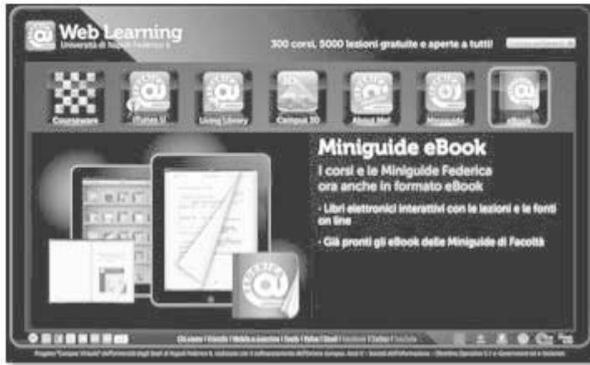
of count the importance of re-engineering human resources and organisational processes. As a consequence there was a proliferation of very expensive ICT/Web platforms at the forefront of innovation, which neglected to take into account several cardinal points, such as: didactic organisation, human resources and the overall educational offer. As a result, many projects remain at a prototypical stage, simply because they have failed to develop working models that are easy for the main target audience to manage: professors and students. As an initiative systematically and organically structured brought online in 2007 - supported by FESR (ERDF) funds of the European Commission and managed by University of Naples Federico II – Federica (www.federica.unina.it) is the only web-learning platform that, in the framework of the Italian university system, is completely open access¹.

The Federica project arises as a response to a systemic vision of pathways for change undertaken by cultural, educational and research institutions. A general reduction in available resources, coupled with the possibility of accessing enormous cultural wealth through today’s web infrastructure has led to a reduction in the academic publishing supply chain, inciting many stakeholders to experiment with sustainable models and original publishing initiatives. «Academic publishing is undergoing a major transformation, with authorship (the sanctuary of scientific communication) facing the challenges of open access and open culture (Lessig, 2004; Willinsky, 2005). While the death of the book, as the bulwark of traditional knowledge, may be one of those apocalyptic forecasts that never takes place, the circulation of ideas is nonetheless taking unprecedented forms and channels» [7]

As a result, e-learning and distance-learning models which based their success on in-house content have been outclassed by commercial platforms which were able to harvest and distribute content packages on a larger scale. The scenario changes even further when educational reforms led public schools to compete with those in the private sector, which are always more aggressive and more receptive to innovation in teaching and vocational training.

This is the scenario which made us want to invest in open knowledge as a way of promoting the educational processes within a large public university and making them more transparent.

¹ The project is managed by Softel (Servizio per l’Orientamento, la Teledidattica e il Weblearning). The scientific direction of the project is held by Mauro Calise, while Monica Zuccarini is in charge for the executive management and Tania Melchionna for the communication planning. The team is composed mainly of young experts in digital content development, accessibility and web design that I wish to thank.



The Federica project takes its inspiration from the weblearning paradigm. This means that it uses the web itself – along with its infrastructure – as a decentralised platform which allows for the creation of an efficient relationship between the scientific web, the teaching web and the cultural web. This kind of model requires an environment where the research materials and educational resources are open access and – more importantly – can be included as part of highly-personalised learner pathways.

Openness, therefore, is one of the important conditions for web learning. Federica has created an innovative learning and organisational project out of awareness of the intimate relationship between the different forces in a knowledge society (convergence, integration, interconnection) and the way they radically change our perspective.

The decision to offer academic content in open source without using a special learning management platform may appear strange at first. Actually, this choice was made on the basis that separate platforms with restricted access intimidate students who consider these initiatives a limit to their freedom; and teachers, who have no opportunity to compare the quality and quantity of the content across courses, nor to learn from other teachers. By contrast, an open access model empowers both students and teachers, improving internal cooperation and collaboration, and enabling the creation of joint projects and integrated courses etc. Finally, open access provides countless functions supporting the operation of this very large university, such as: orientation for students; lifelong learning and education for professionals unable to attend the university; learning support for foreign students; internal communication to create a common university culture; and external institutional communication to improve the overall image of the university (corporate identity).

For Federica, therefore, open access has meant that many of the technical and organisational problems that acted as obstacles to the development of e-learning in Italy have been overcome. For example, content that is accessible without a password has enabled many more students to enter a world of knowledge which was formerly restricted to academic lecture halls or at best to

complicated databases. It has also meant that quality materials are included in search engine selection procedures. Reflecting on the *Amazoogle* phenomenon, i.e. students' tendency to use very few sources to find the information they need, it appears that the web has already become a knowledge management platform and that any new initiative should be based on the increasing power of search engines on the one hand and on the growing desire of cultural institutions on the other – especially libraries – to make the incredible resources embedded in the deep web come to the surface. «In the beginning, everything was “deep” – information was secluded within each proprietary archive. [...] Then the internet arrived, and all sorts of material started being made accessible through hypertext pages and links crawling all over cyberspace. At first it looked like chaos, with information overload the curse of early cybernauts. Soon, however, some kind of order was (re)introduced thanks to all-powerful search engines. The amount of resources available at a mouse click was such an unprecedented bonanza that it seemed as if any information anywhere now belonged to our desktop. [...] Yet, we were (and are) only starting to scratch the surface of the web» [8].

The project for developing the platform thus formed part of an overall vision of the way electronic environments should be used for teaching and research, of the way information retrieval strategies were changing, and of cognitive frameworks for accessing and organising knowledge.

At the same time, the decision to opt for content structure based on web language and Dublin Core metadata has allowed for previously unthinkable developments in terms of interoperability and conversion of the content into other formats (for example, eBooks). Last but not least, in-depth study of the way user interaction design organises the graphical interface, combined with sociological awareness of how young people's consumer habits are changing, formed the background to the creation of the Federica format. A syncretic mix of power point presentation technology and a touch interface within a frame that looks like a popular smartphone. In other words, a format which is self-consistent, has good internal coherence, is user-friendly and recognisable, and is capable of bringing the idea of a public university within everyone's grasp. The close connection that was established between the user space, the visible space and the logical space [9] translates into semantic iconography with well signposted routes enabling users to navigate freely between lesson content and knowledge resource links. Completeness and depth were guaranteed by a post-production team with a background in the humanities and a methodology that many would not hesitate to define as *agile*, i.e. one that adapts the structure of the project in response to external stimuli, especially where teachers needs and developments in e-publishing are concerned. The main asset of the project is that Federica is not its content (or at least not only) but its format. It is

based on the idea that the contents are inextricably linked to the way they are organised and presented, as five centuries of newspaper history has taught us [10]. At the same time we feel that University content cannot be separated from its authorship, as it is the author who guarantees the quality of the work and its survival over time. The author's choice of concept and linguistic devices position the work in a specific disciplinary context, as do the references to scientific sources. In this sense «Federica still has a few reservations about interweaving content from different educational, social and technological environments to take full advantage of the learning environment as a complex, comprehensive information ecosystem, where formal and informal learning, cultural and social background, and the potential of the knowledge web and academic resources are all organically interwoven. This is why the project, despite being covered by the Creative Commons licence, does not encourage the commercial or "derivative" use of the OER available on Federica» [11].

IV. NUMBERS ON FEDERICA

Three years into its working life and Federica supports over 300,000 visits per month, an average of 10,000 unique visitors per day from 198 different countries. These numbers are clearly related to the fact that Federica ensures immediate control of all stages of content production and navigation, which is crucial for users with low knowledge of network technologies. The methods of iconographic representation of the academic content and cognitive organization scheme of the teaching materials, designed on the basis of students' common level skills, allow easy and immediate interaction with the lessons. In addition to these developments Federica presents some specific functions (e.g. Living Library, Campus3D, eBooks) connecting the university courseware units with cultural offerings available on the web, and effecting integration into the educational paths of both primary and secondary scientific sources, experimental research data, and academic production.

Federica is a modular environment comprising four main subsections:

Courseware: Federica hosts the Federico II courses, presented in the same format, through a flexible and user-friendly interface: syllabuses, lessons, research materials, images, audio and video files, as well as links to web resources. (Currently it is made up of 6,000 lessons, 600 podcasts, 6,000 links, 40,000 images, 700 videos and 3,000 documents.) The contents are protected under Creative Commons License. In the next release of Federica, English translations of some technical courses will be available, responding to



demand for scientific knowledge from Mediterranean countries.

Podstudio: Federica's course lessons are also available as podcast files, easy to use on latest-generation multimedia devices, to browse and read the study materials anywhere and anytime. Federico II was the first Italian University to launch an iTunes U channel, with more than 700,000 visitors and 125,000 downloads in only seven months.

Living Library: is the Federica digital library. A gateway to electronic resources, selected and reviewed to facilitate free access to learning materials. An authoritative guide to hundreds of online libraries and archives, journals, e-books, encyclopedias, and databases aimed at improving information literacy and awareness.

Campus 3D: Federica is a 3D interactive environment bringing all university buildings together in a virtual square. It's a scenographic representation that reproduces the strong analogy between the virtual and the real world. A multimedia pyramid helps students find their way through all Federica's resources for open access to higher learning. Thus, it is an ambitious pilot project that will provide access to specialised multimedia resources, but it already represents a strong element of recognition for the University of Naples Federico II and the educational opportunities available on Federica. Federica also offers a range of orientation support functions: something which is crucial when the pace of change is so fast and information to explain it is not always up to speed. The miniguides – which are also available in eBook format – provide useful, straightforward information, designed for the average student who has little knowledge of the complex academic world. The online courses and miniguides enable students to have hands-on experience of the subjects they intend to study, and thus make a much more informed choice about which faculty and degree to go for. This is definitely one of the most interesting aspects of open access academic content. It had never been possible before to offer such a deeply structured kind of orientation, one which is able to publicise and

clarify the range of learning opportunities available, that even people in the field are not always aware of.

V. CAMPUS3D



It was originally known as *Piazza Federica* and started life as part of a regional project entitled *Modem. Elearning pathways for distance learning*, with FSE funding. It formed the nucleus of what would later develop into *Campus3D*.

Developed in LUA code, *Piazza Federica* included some of Federico II's most representative buildings, like the facade, the reception hall, the "Minerva staircase" and the facades of all the historic buildings in every Faculty, all brought together in the one striking Virtual Piazza. *Piazza Federica* is not only innovative. In the way it uses iconographic but photorealistic representation it provides a response to Second Life and offers an overall view of one of the oldest public universities in Europe.

One of the main ideas behind the project – and which later became its leading asset – was that of returning Federico II to its old Campus status, integrating teaching and other services in a single location. Development of the University over time has led to the forced relocation/dislocation of the different faculties and services over the whole city and, in some cases, to other towns. The photorealistic reproduction of some of the most beautiful of Federico II's buildings within a fairly restrained, recognizable and esthetically quality space could help reinforce the University's identity while providing an entry point to functions of the site: orientation; academic content in an innovative format; integration of University's cultural, teaching and knowledge resources.

The continuation of the project termed "Virtual Campus" with FESR 2007/2013 funding from the Regione Campania, (Operational Objective 5.1) involved the development and integration of additional architecture so that further services relating to the University's teaching operations could be developed. Interactive three-dimensional views of the interiors of the thirteen University faculties were thus created, comprising three-dimensional polygonal models complete with textured surfaces and organised into scenes within a hierarchical framework. For the moment, Campus 3D represents a visual strategy for

resource discovering. The holograph panels positioned inside the University buildings (lecture halls, cloisters, corridors and gardens) enable students to discover Federico II's cultural and scientific resources, many of which are costly and rarely used, browse around the living libraries, take advantage of online courses, read the main news headlines from around the world or access the University's net libraries. They can also enjoy some of the events at *The Court of Federico*, a special collection of popular science lectures given by well-known personalities in the field, from both Italy and abroad. A format which has proved very appealing but which – for reasons of time and space – is accessible only to a limited audience.

Information on the historical architecture, personalities, interiors and on Neapolitan and international knowledge and culture is presented in a new way, using hypertext as a strategy for visual discovering, leading to the creation of an immersive learning environment.

Once it is fully operational, Campus3D will have its own astronomy station connected to a telescope positioned near the Faculty of Science (Ruggiero De Ritis Public Observatory) which can be moved using a virtual three-dimensional console. This offers students an amazing opportunity to learn more about science through observation of the sky and through simulation, in line with what Antinucci pointed out in 1999 when he wrote that schools «are based on a particular type of learning that can be termed "reconstructional-symbolic" which is upheld by a specific type of technology, that of the printed word. Knowledge is formulated as text, a totalising and self-sufficient extended form of language that typically takes the form of a book. The text is made up of linguistic symbols which need to be decoded before the objects and situations they refer to can be understood. This kind of reconstruction takes place solely within the mind, and it is within the mind that people work on these constructs to further elaborate them. Schools, in fact, do not take other forms of learning that we are capable of into consideration. There is no room for what is termed "perceptive-motor" learning, whereby students work with the real world rather than with symbols, and do not elaborate knowledge inside their own heads but outside, through feeling and doing. [...] Knowledge is gained through experience» [12].

Campus 3D represents a first step towards the development of immersive environments in which students are able to act and interact with a recreational and challenging space to acquire new experiences. This is why we hope to be able to increase the number of interactive simulation activities on Campus3D, successfully bringing together printing literacy and digital culture.

CONCLUSION

Open Educational Practice (OEP) can therefore be defined as the use of Open Educational Resources to improve the quality of the educational experience. While OER focus on content and resources, OEP concentrates on the way an educational method can be employed to create an environment where OER can be used or created as learning resources [13]. Having said this, however, it is time to start reflecting in a non-ideological way about Open Educational Practices and questioning, in a pragmatic way, how we should go about increasing the use of this kind of innovation in educational processes, starting with a concrete approach which recognises the importance of technological and organisational support. In other words, if the creation of OER is closely linked to the politics of an information society in terms of big numbers and long-term projects, so the use of OER is linked to their reception by the teaching world in general and by the organisational culture of educational institutions.

In its current state, the impact of OER would seem to be fairly poorly defined and there would seem to be little empirical evidence regarding people's perception and use of open resources in University learning. Research in the field refers only to rather small samples or exploratory studies. Larger-scale projects, and more of them, would help workers in the field, providing them with a useful tool for analysing and measuring the impact of OER on educational processes.

REFERENCES

- [1] K. Parker, A. Lenhart, K. Moore, *The Digital Revolution and Higher Education*, Pew Internet Survey 28 aug. 2011.
- [2] Eurostat, *Eurostat Community Survey on ICT usage in households and by individuals*, Brussel, 2010
- [3] Istat, *Cittadini e nuove tecnologie*, 2011, http://www.istat.it/salastampa/comunicati/in_calendario/nuovete c/20101223_00/testointegrale20101223.pdf
- [4] G. Geser (ed.), *Open eLearning Content Observatory Services, Open Educational Practice and Resources*, OLCOS Roadmap Salzburg, 2012, 17.
- [5] L. Masterman, J. Wild, *JISC Open Educational Resources Programme: Phase 2 OER Impact Study: Research Report*, london, JISC 2011.
- [6] Unesco-Col, *Guidelines for Open Educational Resources in Higher Education 2011*, draft report.
- [7] M. Calise M., R. De Rosa, "Electronic publishing, knowledge sharing and open access: a new environment for political science", *European Political Science*, 9 (3), p. 598, 2010
- [8] M. Calise M., R. De Rosa, "Electronic publishing, knowledge sharing and open access: a new environment for political science", *European Political Science*, 9 (3), p. 596, 2010.
- [9] G. Bettetini, B. Gasparini, N. Vittadini, *Gli spazi dell'ipertesto*, Milano, Bompiani, 1999.
- [10] J. D. Bolter, R. Grusin, *Remediation. Competizione e integrazione tra media vecchi e nuovi*, Milano, Guerini e Associati, 2002.
- [11] R. De Rosa, M. Zuccarini, "Federica: la via italiana alle risorse educative aperte". *TD Tecnologie Didattiche*, 19 (2), pp. 96-101, 2011.
- [12] F. Antinucci, "Con il computer nelle scuole simulando (e giocando) s'impara", *Quaderni di Telèma*, Primavera 1999, <http://baldo.fub.it/telema>
- [13] U. Ehlers, "Open Educational Practice - approaching a definition for a new concept", *Discussion paper*, URL: <http://www.icde.org/?module=Files&action=File.getFile;ID=610>, 2010.

Aggregation as a Model for a Sustainable Economy?

The Case of Cultural Journals

Author Rosario Garra⁽¹⁾

(1) General Secretary of the Coordinamento Riviste Italiane di Cultura (Cric) –
Coordinating Committee for Italian Cultural Journals, Roma, EMAIL rosario.garra@gmail.com

Abstract—Cric was created to represent interdisciplinary Italian Journals. It is meant to create agreements among the participating journals in order to develop cultural projects with common goals, as well as to promote dissemination and reading of the cultural journals and to develop the relationships with all the other media and libraries. Condition of possibilities are investigated, also in agreement with the Monti Government. Aggregation projects such as Cric should be economically supported in order to be able to develop new communication and distribution models.

Keywords—component; CRIC; aggregation; sustainable economy; cultural journals

I. CRIC: WHY THIS PROJECT?

The “Coordinamento delle riviste italiane di cultura” (Cric; Coordinating Committee for Italian Cultural Journals) was founded in Rome in April 2003, on the initiative of the directors of a group of journals and thanks to the help of the “Consorzio Baicr Sistema Cultura” and of the “Associazione delle Istituzioni di Cultura Italiane” (Aic), who followed the idea of Federico Coen – director of the journal *Lettera Internazionale* who later became Cric’s first president – of realising in Italy an experience similar to those which were successfully realised by Arce (“Asociación de Revistas Culturales de España”) in Spain and by “Ent’Revue” in France. The project of aggregation and coordination among different cultural journals was developed after two important meetings: “Le riviste culturali oggi” (Cultural Journals today), during Bibliocom meeting on October 17th, 2002, and “Idee in cerca di lettori” (Ideas in search of readers), during the National Fair of small and medium size publishers on November 30th, 2002. The project involved other associations which already existed and which shared common interests and activities in the fields of research and cultural production such as the Consorzio Baicr and Aici (both represent the leading Italian cultural foundations) and the Writer National Corporation. However, since its creation, Cric always led its project in full autonomy, organizing and financing its activities and gradually extending its offer to new journals and publishers.

Cric’s profile, as described in its statute, was originally meant to represent an area of Italian journals characterized by an interdisciplinary approach to contents and by its academic independence within the

publishing market. Then, it included literary and cultural general-interest journals, which do not refer to specialized boards and which are mostly sold by subscription or can be bought in bookshops or newsstands and are more and more published on the web. I will come back to this aspect later on. I am insisting on the association’s original profile because it played and still plays an important role in the evolution and in the future outlooks of the association’s action and coordination in the field of cultural journals. That field, although it represents a ‘niche’ in the publishing sector, is very heterogeneous.

I would like to talk about two initiatives among those which were organized by universities and by research centres because they can be considered the most similar and interesting experiences of an aggregation of cultural journals aiming at promoting their role in the reflection on contemporary themes, debating their problems and reflecting on their future. The “Biennale europea delle riviste culturali” (Berc; European Biennial of cultural journals) was organized by the cultural association “Passaggi” in Genoa for two succeeding years in 1999 and in 2001. From the 9th to the 11th November 2001, the Bianciardi Foundation organised the exhibition-meeting “Riviste di cultura e industria della comunicazione” (Cultural journals and communication industry) [1], starting a series of seminars on some literary and cultural journals, many of which are published in Tuscany. Both initiatives encouraged the contact and dialogue among small publishing houses and cultural projects, which otherwise can hardly aggregate because of their fragmentation. As a matter of fact, they were (and still are) limited by the smallness of their distribution chain, even if they would be able to have a European dimension.

The main aims of the new coordinating committee (Cric) were immediately clear: catching an area of readers potentially bigger than the one journals can reach by using “levers” which, until that moment, had seldom been used by publishers and by journals’ editorial offices. The programme provides for reaching agreements among journals and among other subjects in order to develop common interest cultural projects, to promote the diffusion and reading of cultural journals, especially in education courses and universities, to develop relations with other media and libraries, to improve the knowledge of national and foreign publications through the participation to book fairs and

using the web's communication potentialities, and to improve cultural journals' distribution.

II. THE RELATIONSHIPS WITH THE INSTITUTIONS: AN UNSOLVED PROBLEM

When I was involved in the foundation of Cric and I was named general secretary of the association, I had been working for some years for the "Istituto per il Libro e la Lettura" (Institute for the Book and Reading of the Ministry for Art and Culture), after being a functionary of the "Associazione Italiana Editori" (Italian Association of Publishers) for 15 years. My cooperation with the Institute for the Book represented a new opportunity for combining, among the Institute's activities of public interest, some projects and contents of cultural journals. The Institute was born under the impulse of the "Progetto Libro – Linee d'intervento per lo sviluppo del libro e della lettura" (Book project – guidelines for the development of the book and of reading) [2], presented in Turin in 1997 by Walter Veltroni, who was Minister for Art and Culture and Vice-President of the Cabinet. It was a project for a framework-law planned by the National Commission which included some experts of known repute such as Cesare Garboli and Luigi Malerba, the representatives of all the associations of the book and reading supply chain (authors, publishers, booksellers, librarians), and other Ministries which might have supported that project (for instance the Ministry for Foreign Affairs through the Italian Cultural Institutes abroad).

Although the project never became law, during the following years it boosted many initiatives of the Institute for the Book. I remember two of them, of which I personally took care: the creation of the website <<http://www.ilpianetalibro.it>> and the survey which brought to the "Primo Rapporto sui periodici culturali in Italia" (First Report on cultural journals in Italy) [3].

The website of the Institute for the Book "Italia Pianeta Libro" was created in May 2005 with the objective of increasing the knowledge and synergies in the book's world and was meant to become a national crucial reference point for the whole community of staff and readers. The services dedicated to the publishing production are gathered in the portal of the Biblioteca Digitale Italiana (the Italian Digital Library) and Network Turistico Culturale (Cultural Tourism Network) called "Internet Culturale" – opened during the same year by the General Direction of the book of the MiBAC – that provides access to the archives of libraries and cultural institutions and to the national library service search engine Opac. The website "Il pianeta libro" was active in that same version until 2008 [4], allowing access to important information on the world of books and of publishing: the data bases "Case Editrici" (Publishing Houses), monthly updated inventory of more than 7.000 publishers working in Italy; "Libri del Mese" (Books of the Month), a survey of the 4.000 new books and journals published every month in Italian bookshops; the service "Aiuto agli operatori" (Help to staff), to apply for subsidies given by the Ministry; the online version of

the journals "Libri e Riviste d'Italia" (Books and Journals of Italy) and "Accademie & Biblioteche d'Italia" (Academies and Libraries of Italy), on professional and service culture in their respective fields. In the website there are also "L'Italia del libro", a guide to the institutions and associations operating for promotion; "Editoria e Regioni", an inventory of the laws on culture and libraries, and the guide "Premi letterari". The site hosted the data base of "Ottobre, piovano libri" (October, it's raining books), a campaign for reading promotion which took place for five succeeding years until 2010 and which involved the peninsula's territory for the duration of a whole month; the site had links to hundreds of happenings organised by different Italian Town Councils, such as meetings with the authors, book launches, readings for adults and children, music or theatre shows, guided visits to libraries, exhibitions and cultural itineraries.

The site opened specific communication spaces about journals' contents: the guide "L'Italia delle riviste di cultura" (Cultural journals in Italy) presents about one hundred journals and a survey of this sector; the "Laboratori di lettura" (Reading workshops), that is an hyper-textual investigation on the themes examined in journals; the blog <<http://www.lamialletteralmondo.com>> is dedicated to students and secondary school teachers who participate to the project "Dalla carta al web" (From paper to the web). With the "Laboratori online" (Online workshops), we experimented a first series of reading courses dedicated to the main themes of the cultural debate [5]; those workshops were organized around some articles given by the journals' editing staff, other research and information resources (book reviews, multimedia, web resources), documents, authors' biographical and bibliographical profiles. The workshops are available on the site "Il Pianeta libro" (Planet book) and represent a gathering of knowledge, experiences and cultural contents that can be freely increased and shared by the readers' community and used in projects for schools and libraries.

Thanks to those and other initiatives, a condition of possibilities for a sustainable economy was being outlined. The importance of cultural journals and their synergy with the big institutional projects for cultural promotion and digitizing of book legacy and national publishing was more and more taken into account. The Institute for the Book, born within the general direction of the book Legacy, aimed especially at supporting creativity and high quality in already existing publishing productions, exactly as other organizations were doing in France, Greece and other advanced European states. It was then possible to start a collaboration between the Cric and the Institute for the Book very similar to that between Ent'Revue and the Centre du Livre in France and between Arce and the Ministry for Culture in Spain. That kind of relationship with institutions, which involved libraries and reading promotion, was interrupted in 2009 when the Institute for the Book was replaced by the Centro per il Libro e la Lettura (Cepell; Centre for the Book and Reading) which radically cut the funding for these projects already in progress. Cepell

changed its organogram and its agenda and got a totally new profile. Its new director was Gian Arturo Ferrari, a very renowned figure in the publishing industry who had guided Mondadori Book Division for years. Strategies changed and, as to the aims outlined by the decree regulating the Centre's services, it is mostly growth and book consumption, which were increased. So Cepell looks like a sort of 'hircocervus' because of the way it operated in the administration of culture during the last years. When, in 2011, contributions for high cultural value journals were stopped, we had another discontinuity element with the tasks assigned to the MiBAC and foreseen by law n. 416/1981 on book trade. Cric has tried to make up for the absence of an institutional reference point for Italian cultural journals. During the session at the Cabinet's VII Commission for Culture on the 18th of July 2011, Cric's president Valdo Spini clearly denounced: "now we need a strong relationship with a precise national reference point. That could be the Centre for the Book and Reading or, if it is believed that that one still ought to be the Ministry for Culture, then there should be a true Journal Project at the Ministry agreed with associations like ours".

The "Journal Project" should aim at the solution of some structural problems in the sector (i.e. access to bookshops and libraries, digital divide, distribution, etc.). So, we have two possibilities: either establishing a dialogue with the new Government headed by Mario Monti, whose profile is strongly characterised by innovation, to endeavour the survival of the existing contributions, or focusing on a radical reform of the set of rules in order to achieve a series of interventions more in accordance with our times (that is activating a sort of spending review). For instance, we could follow the model of the "Support to journals" of the Centre national du Livre [6], and also get some useful ideas from a very advanced law of the Tuscan District (n. 21/2010, art. 48) [7], which is presently being enacted. Both models provide for a distinction in financial aids. The attitude is: "I give you a contribution for something that you are going to do, in order to help you reach your objective". A significant financial support could still be given to a small group of journals whose cultural contribution is considered unique, that is a much more selective attitude than before should be applied (as I'll show later, the problem of evaluation criteria of cultural journals is still very debated). Other and new journals should be given the necessary economic support helping them to renew and experiment communication and distribution models in order to reach new readers and allow self-financing.

III. A VERY FRAGMENTED ARCHIPELAGO

The field of culture periodicals is composed by many networks – of an institutional, associational, editorial, academic nature – in which different corpora characterizing Italian culture and society join up. An approximate estimate could include: about 400 publications issued by the 10 major publishing companies; about a hundred journals issued by foundations and cultural institutions; 250 journals

related to the voluntary sector [8]; 681 humanities and social sciences journals [9]; 58 theology and religious culture publications [10]; and an indeterminable number of literary reviews, including several hundreds printed and on-line publications.

It is difficult to fully understand the system in its complexity because of the significant lack of data, information and analysis. Cric has started a research – which would need further elaborating, extending and updating – on a sample of 100 publications chosen in accordance with the identity of the association [11]. Some of the results of the research are meant to identify the specific and critical traits characterizing the economic structure of the sector. Cultural journals show critical aspects that are typical of smaller publishing, such as the difficulties in distribution and the troubles concerning organization and management, that add up to some problems having to do with finding a place for the review as product within the publishing panorama. The field of cultural journals includes different professional figures and legal entities (graphic 1): on one hand the founders/owners and on the other hand the publishers of publications. 72% of publications are issued by publishing companies proper, that have a catalog which includes, in addition to book series, a section with one or more journals; while in 28% of the cases, journals are published directly by the director and the editorial staff or by the institution to which the publication is linked, which can be organized as many different kinds of legal entities (association, society, cooperative).

By analyzing the information regarding the ownership of publications, one can understand in further detail the economical-financial and organizational structure of the sector (graphic 2). The owners of publications – foundations, associations, natural people or institutions, that are often also the founders and the curators of the publications – frequently turn to publishers for the management of the publishing and marketing processes. So, even if the publications issued by official publishing companies are prevalent (72%), it is also necessary to specify that, in the sample at hand, only 42% of the publications are owned by the publisher, while in 30% of the cases the publisher and the owner are two different figures. 58% of the publications are indeed owned by associations and foundations (34,6%), natural persons (17,3%), public institutions (5,7%). The separation between the editorial staff and the publishing company, which is often also physical and geographical, does not in many cases leave time to the staff to effectively develop the product (communication, marketing, Internet, promotion, European projects, etc.), if these aspects are not taken care of by the publisher.

The third indicator to consider is the distribution channel for publications (graphic 3).

The choice of the marketing channel is a crucial factor to reach and select the audience of a publication, and is therefore an unavoidable variable in the editorial line and in the cultural project of the publication. In the organizational chain of a cultural journal, the distribution and self-financing nowadays depend almost entirely on

subscriptions. These represent the main, and often almost exclusive, marketing channel for periodicals (77,7% for monthlies; 67,5% for other periodicals). Companies publishing political and cultural interdisciplinary periodicals capable of the necessary investments to actually sell their product in newsstands do in fact exist, but there are few of them (for instance, you can buy Micromega and Limes, from the L'Espresso publishing group, and Aspenia from Il Sole 24 Ore group on newsstands). The difficulty in distributing the product in libraries is, as known, a further problem, because of the shape of the book market and of the increasingly serial and selective dynamics regulating publishing distribution. The publications that cannot make themselves available in other circuits also have, as a consequence, less subscriptions.

The large and varied constellation of cultural projects, publishing initiatives and networks makes it almost impossible to identify a standard profile for a cultural journal, a single model of economic and productive management, a consolidated and shared circuit for the distribution and diffusion of journals. The diversity in strategies in the current system of information production – according to Yochai Benkler's [12] analysis – allows the dialectic coexistence of management models followed both by commercial and non-commercial participants, both based on the assumption that information is a “public good”.

IV. DIGITAL CONTENTS: PRODUCTS AND COMMON GOODS

The question that must be answered now is if, in digital networks, the necessary provisions are forming for the opening of new spaces to help journals strengthen both their communication with the readers, and their offer of digital contents, evaluating how this is related to the issues of Open Access. The other aspect to investigate is how the aggregation of periodicals – also through, and by initiative of, Cric – may evolve and offer an effective contribution to this perspective.

In the past years the number of on-line reviews – both those migrating content from paper to digital form and those stemming from original realities in the Internet – has increased. A significant part of the journals, especially when made or fueled by young editorial staffs, have already embraced the participative philosophy of the Internet, and have used the interactive and multimedia tools of social networks and web 2.0. In particular, book-related websites and literary blogs (such as Nazione Indiana, Lulu.com) and on-line political journals (like Tamtam democratico, LaVoce.info) have followed the Open Access model, prompted by their spontaneous and innate vocation to build, together with the readers, a strong community on the basis of common interests.

In the field of humanities and social sciences journals, that were once referred to as “militant” and nowadays still intend to keep the cultural debate alive and foster new ideas, there is a persistent attachment to

printed paper, in other words to writing as intended in the modern age. Such attachment is strong and understandable, and needs not be explained, because it is full of practical and symbolic implications. However, such sensibility is also capable of adapting to the general context and to the changes that crucially affect scientific communication and functional communication. It is harder to understand and to accept the fear, and perhaps the prejudice too, that the widespread availability of digital information, and the digitalization of articles and essays from journals, will inevitably cannibalize the printed version and even end it. Both the protection of traditional marketing channels and the safeguard of copyright in the digital realm and on the Internet cannot be narrowed down to a merely defensive strategy. The matters concerning intellectual property and the identification of new possible models for a sustainable economy are of primary importance in the publishing world and in the field related to all of the contradictory transformations investing the complex sphere of authorship. I am obviously referring to the Google phenomenon, which is still relentlessly growing; to the risks deriving from the formation of international publishing concentrations, at the hands of global publishing groups such as Thomson, Pearson, Elsevier [13]; to the stances of the Open Access movement, which finds widespread consensus in sectors of the scientific community, of universities and libraries. There is a variety of models and solutions, also for open access publishing, that can be experimented in the author-publisher-reader chain, and that can be traced back to the so called gold road and green road.

This past September, the Italian newspaper “Repubblica” [14] hosted a debate on the new criteria for the evaluation of academic research (passed by the Government in July 2011), bringing the subject to the attention of the public, outside of the academia. The debate focused in particular on the issue of the internationalization of the results of research in the fields of humanities and social sciences, and more precisely on the Government's choice to assign a different value to articles whether they are published in English or in Italian, on international or Italian periodicals.

The idea of science as reliable and well-grounded expertise which characterizes the various fields of knowledge is, in the field of Humanities and Social Sciences (Shs), structured into a larger domain than that of the so called “hard sciences” community, for which the assumptions and the results of research must necessarily be validated. The cultural role of journals in the field of humanities and social sciences is emphasized by the interpretative and interdisciplinary nature of knowledge, by the ability to create judgments and to assign a meaning to their own object. In this field, “cultural paradigms” are created and renewed by pursuing consensus in the public sphere and in the value sphere, and are grounded in the technical and practical realm and in “narrations”, that are determined within society in its historical development, and that have special links with the national tradition and language.

In the sphere of humanities and social sciences, the situation is considerably varied among the different disciplinary areas. According to comparisons made by the French publisher NecPlus on 2,500 international journals, both in and out of the Isi impact factor list, the impact-factor classified Ssh journals are 49% of the total, and in particular: Sociology 47%; History 42%; Philosophy and Religion 43%; Art 36%; Literature 11%; Political Sciences 50%; Psychology and Pedagogy 65%; Economy 67% [15].

The national agency for university evaluation and research (Anvur) has stressed the need to prompt the best Italian journals to become part of the most important national and international databases (Isi-Thomson, Scopus Elsevier), by matching the requirements of Peer review and Impact factor.

The lower significance of bibliometric indicators in the validation of humanistic and social studies has however been acknowledged by Anvur, that has adopted, at least temporarily, the number of publications as the main parameter in the evaluation of research, while still taking into account the different diffusion of the works published in Italy and those published abroad. SHS publishers and journals are required to submit each product to fixed ex-ante evaluation processes, through Peer review, editorial committees and such. However, the Anvur document specifies that “no consolidated, validated indexes are available for Italian language journals and monographs, which would allow a precise assessment of the respective scientific value”.

V. TOWARDS A NET COOPERATION

There are several elements interconnecting issues of digitalization and the evaluation of the quality of publications, and these are now affecting the circulation of journal networks. To begin with, publications in the fields of the Humanities and Social Science need to be acknowledged in terms of identity, respectability, and the conditions of dissemination. But there is also an area within cultural publishing with no clear distinction among scientific and professional, didactic, documentary, and popular texts. These publications literally weave a productive and critical relationship between science and society, as they also circulate outside the confines of the academy, or else, they enter the academy from the outside world, when they are produced through cooperative cultural and study projects or editorial initiatives by scholars and academics alike.

To sum up, a consideration can be drawn as regards the potential of the Net in order to strengthen the dissemination of published texts and to foster the exchange between academics along clearly defined editorial guidelines, copyright systems, models of economic sustainability, communicative strategies, with their differences and similarities.

Recently, some leading publishers in the Humanities have joined Cric, with a full catalogue of refereed journals that in most cases have been conceived of as a result of cooperations between universities, foundations,

and academic institutions. Within this framework, some editorial directors and editors have started their own association following the indications contained in the Statute. Cric now consists of the following members/partners: Fabrizio Serra Editore, Le Lettere, Olschki, Storia e Letteratura e Casalini Libri. All partners involved in the project are also playing a fundamental role for the promotion of Italian culture abroad, as well as functioning as leading expert at a national level. Through “Online Italian Editing/Publishing”, that is, the main digital archive in Italy, Casalini Libri works to develop complementary elements in line with the scope and objectives of Cric; it functions as a *trait d’union* for the publishing of online journals subscribing to the association and archived in the monographs and articles database, which in turn is made available to libraries and academic institutions in Italy and abroad.

Cric can also benefit from the support of other associate members, namely, leading publishers of journals and relevant publishers in the interested areas. A broader spectrum of this particular field of the publishing sector at large is thus offered for the different institutions, at many levels and in spite of the current economic crisis. Recent cutbacks in public expenditure have worsened the neglect suffered by cultural journals in the publishing industry and the dissemination of such works. On the other hand, the impoverishment of library collections is putting the survival of national heritage at high risk.

A cooperation aimed at strengthening the infrastructure of knowledge by the different parties – mainly schools, universities, Research, cultural institutions, archives and libraries – seems a cogent issue at present. This should entail the adoption of a point of view able to grasp and to valorize the interconnectedness of cultural and research politics and socio-economic development politics.

Within this perspective, some lines of contact emerge which may lead to a possible alliance of the “academy” (meaning, the Humanities sector) as it moves towards the self-management of its own scientific heritage in Open Access systems and cultural professionals (both in the public and the private sector), in keeping with the scopes of the Open Access 2002-2012 programme.

Anvur has provided some relevant comments, adding that: “great effort shall be undergone in order to catalogue journals published in Italy ... the scientific communities have to be involved in the classification of scientific products, so as to improve their comparability and transparency” (issue 25th July 2011). Were the cataloguing process realized through a cooperation between all partners involved, it would allow for complete and correct archived information, as well as conforming to requirements of objectivity and “thirdness/terzità” of the different selecting criteria in their rapport to the features of each relevant category. Cric would then play a fundamental role for the creating of a collection of library information, using the Net and its

own relationships with publishers and editorial offices of each single journal unit.

The role played by private publishers in the editing of journal is much wider in Italy than Spain or France. The French model, often taken as example, is based on the cooperation between the private and the public sectors for the development of a navigation space that may include the whole set of scientific documents in the fields of the Humanities and Social Sciences. The building of partnerships consisting of different actors working in the commercial field, in universities and Research has made it possible to take their digitalization and archiving at a really advanced stage. The case of journals in France clearly demonstrates that “there is great concern on the part of private institutions to publish online both current and back issues of their journals, since university publishers and, to a lesser extent, public institutions, seem more likely to use the Internet to propose the archive of the materials they publish” 16. The time has come for our own National Agenda that may give way to a series of reflections on the future of cultural journals, one in which different stakeholders of the digital offer may discuss and interact.

FIGURE

Home Page del sito <<http://www.ilpianetalibro.it>>

Graphic 1 – Publishing structure

Graphic 2 – Journals ownership

Graphic 3 – Selling channel

REFERENCES

- [1] W. Lorenzoni (ed.), *Riviste di cultura e industria della comunicazione*, Firenze: Società editrice fiorentina, 2005.
- [2] “Progetto Libro. Atti della Conferenza Nazionale del Libro”, *Quaderni di Libri e Riviste d’Italia*, n. 40, 1998.
- [3] “Guida all’Italia delle Riviste di cultura”, a cura del Cric, *Quaderni di Libri e Riviste d’Italia*, n. 55, 2005.
- [4] Important parts of the project were used in versions of the site <<http://www.cepell.it>>, after 2009.
- [5] Two online workshops were realised in 2006: the journal “Lettera internazionale” edited “Noi, cittadini d’Europa”, a journey within themes and readings about contemporary Europe “in search of a new and authentic common culture of the European Union member countries”; the journal “Confronti” created “Percorsi fra culture: le religioni”, which presented reflections and studies on the role of religion in contemporary societies “for an authentic dialogue between different cultures and faiths”.
- [6] <<http://www.centrenationaldulivre>>.
- [7] <<http://www.regione.toscana.it/cultura/editoria/index.html>>.
- [8] “Catalogo delle Riviste specializzate italiane sui temi dell’emarginazione, politiche sociali, volontariato, terzo settore”, in *Rassegna stampa Informazione e marginalità*, n. 15-16, Regione Emilia Romagna, 1998.
- [9] Giuseppe Vitiello, *Il mercato delle riviste in Scienze umane e sociali in Italia*, “Biblioteche oggi”, gennaio-febbraio 2005.
- [10] <<http://www.associazionesantanselmo.org>>.
- [11] “Rapporto sui periodici culturali in Italia”, in *Guida all’Italia delle Riviste di cultura*, quoted.
- [12] Y. Benkler, *La ricchezza della rete*, Milano: Università Bocconi Editore, 2007.
- [13] Thousands of scientists and scholars have joined, since the beginning of 2012, the platform *The Cost of Knowledge*, <<http://thecostofknowledge.com>>, created by the American mathematician Timothy Gowers, which aims to boycott the major publishing company Elsevier. Elsevier is accused of charging exorbitant prices for subscriptions to its journals; of only selling journals in large packs, so that libraries are forced to buy also titles they are not interested in; of supporting laws that aim to limit on-line access to the data from publicly funded researches and studies.
- [14] The contributions published in “La Repubblica” are by Carlo Galli (20.09.11), Andrea Graziosi (24.09.11), Michele Ciliberto (29.09.11) e Roberto Esposito (7.10.11). The matter was also tackled by Adriano Prosperi in his brilliant report at the convention “Le riviste italiane di cultura e il loro ruolo nel XXI secolo” (Italian cultural journals and their role in the 21st century), organized by Cric during Florens 2010 (Palazzo Vecchio, Florence, 29 October 2010), now on “Quaderni del Circolo Rosselli”, n. 1-2/2011. Please also see my essay “Le riviste politico-culturali o SHS e la “coda lunga” della comunicazione”, published on “Queste istituzioni”, n. 157-157/2010; that contribution was based on comparative research carried out in France and little known in Italy, which was made available thanks to Cric’s participation in the Paris Salon de la Revue, which takes place each year in October.
- [15] In 2009 a study was published in France, titled *L’édition scientifique française en sciences humaines et sociales*. This study was carried out by GFII, an association of information industries, for CNRS as part of the ADONIS project “Accès unique aux données et aux documents numériques des SHS”. The results of this study show quite evidently the level of internationalization of electronic journals in the SHS field. The data are gathered by the French publisher NecPlus, which makes them available on the website <<http://necplus.eu>>.
- [16] “Etat des lieux comparatif de l’offre de revues SHS. France – Espagne – Italie” (2005).



Poster Session

Nuova Consonanza Archive and short films of the Sixties and Seventies: an example of low cost preservation

Cecilia Montanaro and Laura Nicoletta Colabianchi

Music Theatre Research Institute, Irtem, Rome, Italy, irtem@mclink.it

Abstract— Between the Sixties and the Nineties Rome was one of the major centers for experimenting in all fields of art especially in that of music. The fermenting creativity of those years, much helped by public financing, has left a conspicuous patrimony of analog recordings, which because of their perishable nature, presented an immediate problem for their conservation.

IRTEM – Music and Theatre Research Institute dedicates a consistent part of its activity recuperating and saving the sound archives and the audiovisual documents from the risk of extinction. In fact, since 1994 has been working on restoring and conserving 54 experimental documentaries of the Sixties and Seventies and 456 Ampex tapes from the Nuova Consonanza Festival, an important reference for Italian contemporary music.

This report will illustrate how a small Institute with very limited resources has managed to recuperate the above-described patrimony.

I. NUOVA CONSONANZA ARCHIVE

The Association was created in the 60s by a group of composers who aimed at promoting the circulation of contemporary music in Rome. At the time, major concert organizations regarded the musical avant-garde with indifference, therefore the creation of a specialized association allowed for the production of concert seasons dedicated to new music. The Association became a meeting point where composers exchange their various views on music. Their unity derived from the desire to make contemporary music available to listeners, and thus to promote its circulation.

To this purpose they established projects in cooperation with foreign cultural institutes, thus favoring the circulation of 20th-century historical music as well as promoting musical activity in schools. One of the Association's peak projects was the Gruppo di Improvvisazione.

Its uniqueness resides in the fact that, for the first time, it was the composers themselves who performed their music. The traditional relationship between performer and composer was overcome. The Nuova Consonanza Association – together with the Roman artistic avant-garde for the graphic

aspects of programmes and catalogues organized an annual Festival featuring distinguished personalities: Boulez, Stockhausen, Nono; as well as performers: the Kontarsky, the Lasalle and Arditti Quartets, and so on. Their concerts and seminars have been recorded since 1976 (353 1/4 inch tapes 32 VHS and a growing number of DAT tapes).

This Sound Archive, besides documenting the Association's history, is also a unique token of Italian contemporary musical development.

Starting in 1996, Irtem – Music Theatre Research Institute was able to finance and carry out the digitization and cataloguing of their collection.

A. Preservation

The project was divided into phases whereby more or less 30 74-minute CD-Rs are digitized yearly, beginning with the oldest. The archive was neither orderly nor catalogued in any way; there was no preservation strategy for tapes and, furthermore, some recordings were lost in time because composers claimed them as personal property. In the best of cases, some were found in the musicians' private archives. Fortunately, the fact that tapes were stored in their original cases allowed us to reconstruct what kind of tapes they were (Ampex 406 and 407 until 1982 and Ampex 456 since 1982). Moreover, the original cases bear technical information such as the date, place, performers and programme, and often they also point out eventual programme changes. In cases where information was insufficient, we were able to integrate data by using concert booklets, which were taken into account in the process of cataloguing. All the relevant data was subsequently copied onto the CD-R master covers, with annotations regarding eventual program changes announced in the course of the recording (as well as encores).

For VHS tapes the starting point was different: good-quality tapes were used, but brands differed. Original cases only carry a progressive number, recording date and a 16 bit pulse code modulation which indicates the digital technique

used. In this case, as well, it was necessary to integrate information available on the support with the use of concert programmes.

B. Restoration technique

- *1/4 Inch tapes*

Following is a brief illustration of the procedure we follow.

1. *Inspection*: the support is accurately inspected.
2. *Preliminary cleaning*: the tape is taken out of its protective plastic bag, eventual remainders of sellotape are removed and, if mould is detected, it is swept off with a paintbrush.
3. *Restoration*: splices are checked and substituted if no longer efficient; if paper leaders were used, they are substituted with new acid-free paper leaders.
4. *Baking*: the more recent Ampex 456 tapes are oven-baked at 60 degrees centigrade for an hour in groups of 4. This procedure is necessary when we encounter what has been defined as the "sticky recording tape" problem, deterioration due to moisture, which can be caused by the binder used to glue tape to its plastic base or to incorrect preservation. The purpose of baking is to dry out moisture, which is what causes tapes to become sticky; this procedure allows for tapes to remain stable for some time, just enough to digitize them.
5. *Determination of tape velocity*: in the case of Nuova Consonanza tapes, velocity is not constant and ranges between 3 and 15 inches per second.
6. *Equalization*: this procedure was necessary in order to correct the signal's frequency response and restore its original characteristics. Tapes were equalized according to one of the European standards.

Digitization is carried out in 3 phases:

1. The material is reproduced and recorded on high definition-recording Pro-Tools at 24 bit, 48 KHz.
2. Minor editing is performed consisting in the elimination of excessive space between pieces, applause, etc.
3. Once the procedure is completed, the resulting sound document is recorded onto a so-called master CD-R, which is subsequently copied. Eventual errors or other information concerning missing sections is included in the master CD-R case.

- *VHS tapes*

Digitization of VHS tapes was much simpler: the excellent-quality tapes used were not deteriorated, and the recorded material was first-class. VHS tapes had been recorded by the same firm which subsequently took

care of digitization, therefore they used the converters originally used for taping.

C. Cataloguing

As the material was being digitized, it was also catalogued in 2 manners.

In the first case the document described is the original support, therefore entries include the following information: analytic description of support, contents, format and state of preservation.

In the second case entries refer to the digitized document, and this caters to quick online consultation.

For accession number, we used the abbreviation CDNC, followed by year of recording, dash, progressive number.

At this point, there were two needs to cater for:

1. Protect the preserved document from eventual deterioration.
2. Devise a way to consult our archive in a quick and flexible manner.

These two needs are closely related to the fact that Irem's archive suffers from limited funding and, furthermore, it is understaffed. We thus had to find an economic solution we could cope with.

In order to protect our archives from eventual deterioration, we have begun to transfer all documents onto hard disk using RAID system.

This mass storage system yields non-compressed, high-quality wav audio files. Wav files are extremely flexible to export, so as to simplify periodical transfers onto the changing supports that new technology constantly offers.

Archive material can be listened to from a computer, without having to physically find the disc, and it is easy to skip from one piece to another; all audio material is correlated with the information mentioned earlier.

II. DOCUMENTARIES

During the Sixties the majority of the Italian documentary films were made thanks to a law, which gave a State contribution for every short film of cultural interest produced. This law gave incentives to small productive units and young directors and composers for experimenting new film/ musical expressions: e.g. Cecilia Mangini, Lino Del Fra, Vittorio Armentano, Luigi Di Gianni etc; among the composers special attention must be made to Egisto Macchi one of the Irem founders, also to Daniele Paris, Domenico Guaccero, Massimo Pradella etc. The majority of the

documentaries have social and cultural subjects, which reflect the care (also political) with which some of the directors of that time worked and sometimes realized in fiction films

The disappearance of the productive units resulted in a dispersion of the films and a still unresolved querelle as to the authors' rights. The first working period consisted in a long search to find the films, as they were almost all in possession of the directors and private collectors. The copies (mostly positive and only in a very few cases negative + soundtrack were in good condition when housed but in very bad condition when stored in cellars, attics etc.

Strategies and restoring operations

The work involved can be ascribed to the category of conserving restoration. Which means conserving and restoring –considering the short film as a document and respecting the form and contents both technically speaking and narratively, reducing to the minimum the corrective intrusions into image and sound.

The best results were obtained when it was possible to involve the same directors or who had closely collaborated in the realization of the short film.

In the following scheme there are the various phases used for preservation

1. Choice of the film material according to the quality and state of deterioration;
2. Supporting restoration, i.e. cleaning the films and revising the joinings;
3. Telecine with corrections, colouring and eventual resynchronization of the soundtrack. Digital Betacam has been chosen as a support for the master;
4. Copies on Laser Videodisc CRV Sony for conservation, DVD (initially VHS) and Betacam SP for consultation of IRTEM's video archives. At present, owing to the obsolete Laserdiscs and the liability not on a long term of Digital Betacam (and of magnetic supports generally) the procedure is to reverse on to a hard disc in uncompressed AVI format.

For the realization of the various working steps the best places for telecine in Rome have been selected (since 2000 in HD) for cleaning, restoring and for copies.

The work has proceeded on 16 and 35mm film, black/white and colour, copies in positive in most cases, and only in a couple of cases in negative. The soundtrack of the various formats is always monophonic.

On explicit request of the directors there has been a regeneration of the film during the restoration phase of the support so as not to alter the color of the film. The soundtrack (always mono) has been transferred in digital format 16bit/44.1kHz without added equalization or denoiser.

The choice of Laserdisc Sony for conservation was dictated by the fact that in 1994 this group of audiovisual supports was considered the most reliable as it did not compress the video or audio signals, insomuch at that time it was used in medical and military ambits for data conservation. Here are the principal characteristics:

- WORM (Write Once Record Memory) optical disc
- uncompressed video recording
- high quality stereo soundtrack (PCM 16bit/44.1kHz)
- registration capacity 48 minutes per side

Not having sufficient funding for the reprinting of negatives, it seemed that the best choice was for quality and durability.

CONCLUSIONS

We've been working at IRTEM for the past 20 years; in this lapse of time we've seen the rise and fall of several audio and audiovisual supports.

From one day to the next, one of our reproduction devices would be considered outdated and then, the next week or month, it would become an invaluable asset because it allowed us to listen to obsolete supports.

In these years, our ability in using a computer has reached an acceptable average level but, more often than not, we find ourselves anxiously chasing after the latest technological innovation, like a hamster on a wheel.

Our experience, therefore, is that of a small archive where the tasks of computer technician, sound engineer, archivist and musicologist are not necessarily performed by different people. Often, one of us alone has to take care of everything, with all the advantages and disadvantages this entails.

To this purpose, establishing relationship and cooperating with similarly small archives is of the utmost importance.

It is also essential to have a chance to harmonize national and international projects aimed at preserving the audio and audiovisual heritage, perhaps one of our culture's most perishable assets.

REFERENCES

- Brandi, C. (1977). *Teoria del restauro*, Torino, Einaudi.
- Burtin, J. (2000). *A Timeless introduction in Motion Picture Sound Recording* <<http://www.agfa.com/motionpictures>>
- Conti, A. (1996). *Manuale di Restauro*, Torino, Einaudi.
- Home Film Preservation Guide, The <<http://www.filmforever.com>>
- Tortora, D. (1991), *Nuova Consonanza. Trent'anni di musica contemporanea*, Lucca, LIM (Musicalia)
- Schamber, L. (1988), *Optical Disk Formats: A Briefing*. <<http://cool.conservation-us.org/bytopic/electronic-records/electronic-storage-media/ed303176.html>>

The Documentation Centre and Museum of the Performing Arts

A century-old institution working for the digital future of theatrical heritage

Mireia Roselló Gargallo⁽¹⁾, *Roger Guasch Arambudo*⁽²⁾

(1) Institut del Teatre. Centre de Documentació i Museu de les Arts Escèniques, Barcelona, Spain, rosellogm@institutdelteatre.cat

(2) Institut del Teatre. Centre de Documentació i Museu de les Arts Escèniques, Barcelona, Spain, guaschar@institutdelteatre.cat

Abstract— The Documentation Centre and Museum of the Performing Arts (MAE) do not have a physical building for the museum and for this reason the digital strategy is crucial for us, in order to meet the objectives of preservation and promotion of heritage.

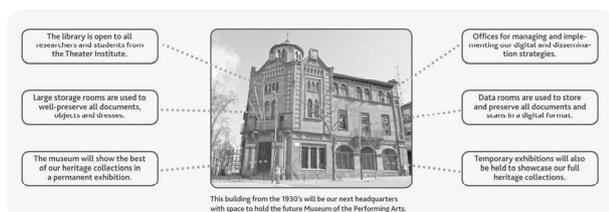
preservation; digital images; digital strategy; museum; library

I. THE INSTITUTION

The Documentation Centre and Museum of the Performing Arts (MAE) has two main objectives:

- support teaching and research of Theatre Institute of Barcelona and
- preserve the memory of performing arts in Catalonia and spread them.

Today it is one of the reference centres in this field in Spain. The MAE does not have a physical building and for this reason the digital strategy is crucial for us in order to meet the objectives of preservation and promotion of heritage.



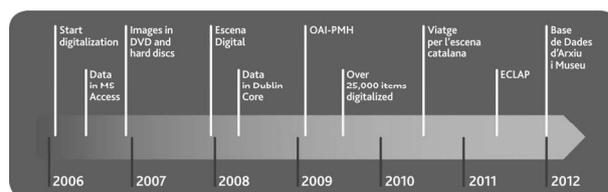
A. Heritage collections of performing arts

The MAE wants to collect everything that is related to professional shows and also those that have been premiered in Catalonia. When we talk about preservation of memory, we refer to anything that can evoke a particular show: posters, music, costumes, photographs, hand programs...



B. The digitalization Integrity of the Specifications

Obtaining a big amount of digital images has given rise to the question of conservation. Initially, we have solved this problem with DVD copies and external hard discs. The documents were described across scattered Access databases not available on Internet. However, this decision has created big difficulties for recovering these images. To solve this problem it was decided to migrate databases to a searchable internet platform.



II. THE DIGITAL HERITAGE COLLECTIONS

The dissemination of our collections follows three preferential paths: 1) present all documents with the most accurate description and according to the international cataloguing standards; 2) present the documents in a context, making a museographic narration that could enrich the visualization of masterpieces; 3) present our content to the world with the internationalization of our collections.

These three strategies are materialized in three websites: the repository "Escena Digital" (replaced shortly by the BDAM project), the "Viatge per l'Escena Catalana" and the ECLAP platform.



a) Escena Digital

Escena Digital was created to offer the image and the information of all museum documents of MAE. To implement the repository we used the program CONTENTdm created by OCLC specifically for collections of images and reproductions of objects.

b) *BDAM (Archive and Museum Data Base)*

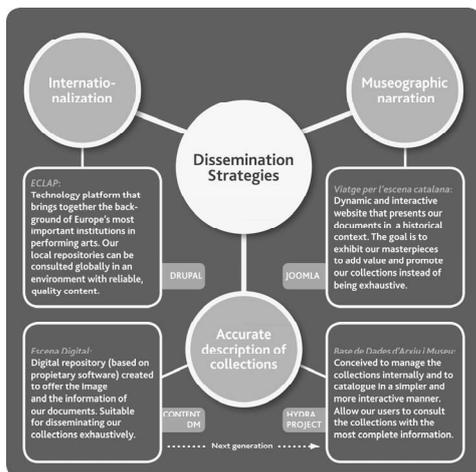
BDAM is conceived for managing collections internally. To make this repository we choose the free software Hydra Project. The main idea of the repository is to link the documents to the shows. This will allow us to catalogue in a simpler and more interactive manner.

c) *ECLAP*

In our digital strategy we consider internationalization as a major goal. For this reason, we are participating in the ECLAP project, a technology platform that brings together the background of Europe's most important institutions in the field of performing arts.

d) *Viatge a l'escena catalana*

Viatge a l'Escena Catalana is a website that aims to present our documents in a historical context. This site includes several ways to approach to our collections, either through concrete exhibitions such as "the stage" as well as the series of the greatest Catalan stars (actors, singers, opera singers etc.).



III. THE DIGITAL STRATEGY

5 years ago we defined a strategic plan to digitalize all the heritage collections. This road map should give us a framework to face the challenges of preservation and dissemination and dissemination of our collections.

A. The standards

In 1998 a new standard¹ that was easier to use was created for cataloguing digital objects through metadata². This standard facilitates the exchange of information between institutions and is called Dublin Core (DC). This has allowed us:

- to standardize the data we collect for each registration,
- simplifying the process of cataloguing,

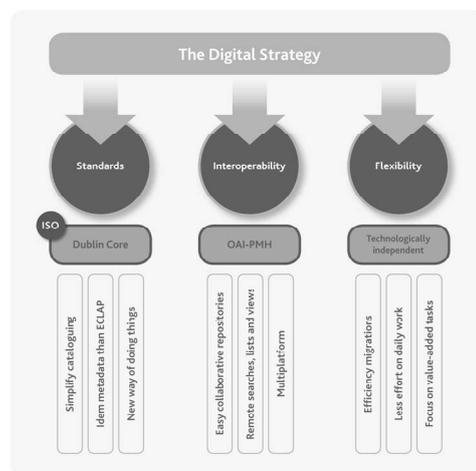
A clear example has been our participation in the ECLAP project: the first metadata ECLAP model was exactly the same as those used by DC.

B. The interoperability (OAI-PMH)

Internet represents a great opportunity for our collections to reach a wider audience. We have decided to collaborate with other organizations to display our collections on their platforms. The OAI-PMH is a protocol that allows the exchange of DC metadata using XML messages between two independent platforms. So, we just have to enable the harvesting and provide the URL of our collections to the institutions that we wish

C. Flexibility

The adoption of standards is long and complex but it has many benefits. After this process we achieved a high degree of freedom: using a single set of metadata, we are technologically independent and it allows us to establish alliances with other institutions. In summary, allows us to spend less effort on a daily work and encourages us to focus on value-added tasks.



IV. CONCLUSION

The MAE opted for a strong Internet presence. To face a challenge like this we define a digital strategic plan detailing the main lines of action to publish tens of thousands of objects on the network:

- Use of metadata standards,
- Introduction of digital repositories easy to use, flexible and powerful
- Promote collaboration with other institutions.

REFERENCES

- [1] Anna Valls, "Fons patrimonials d'arts escèniques" Item. pp. 18–20, 2009.

¹ In 2003 became ISO

² Metadata is data that provide information about other data objects (e.g. the title of a book)

Rediscovery of the shape of artistic artefacts from moulds: tools and methods for 3D archivation

Sergio di Tondo⁽¹⁾, Giulia Adembri⁽²⁾, Monica Gherardelli⁽²⁾

(1) Architecture Department, University of Florence, Florence, Italy, sergioditondo@ndas3a.it

(2) Department of Electronics and Telecommunications, University of Florence, Florence, Italy, giulia.adembri@unifi.it, monica.gherardelli@unifi.it

Abstract— In this paper we illustrate three-dimensional laser scanning recognition of models from plaster moulds. The activity arises from the need to categorize and classify moulds according to sculptural structures made from them. In this perspective 3D laser scanning acquisition, reconstruction and subsequent reassembly of complete models, is an optimal solution for the immediate creation of a digital archive which can be organized and managed efficiently in a Database Management System. A simple identification of moulds referring to the same model is possible provided that these moulds are labelled with an automatic readable identification code (Quick Response Code).

Keywords—3D laser scanning; mould; digital cast; 3D archive; artistic heritage database; reverse engineering.

I. INTRODUCTION

In this paper we illustrate the possibilities offered by 3D laser scanning techniques: recognition of artistic artefacts from moulds, which in our case are “impressions” in plaster, belonging to the collection of the Manifattura Richard-Ginori 1735 S.p.A., located in Sesto Fiorentino (Florence), Italian manufacturer of artistic porcelain. The collection consists of groups made up of piece-moulds, the number of pieces is not exactly known; from each group a three-dimensional model is obtainable through porcelain casting. The complexity of the obtainable figure is obviously a function of the number of pieces needed for its reconstruction. From the foregoing the need to categorize and classify moulds according to sculptural structures made from them arises, taking into account the following aspects: the large number of moulds would require excessive space occupation to accommodate the models obtained by pouring the porcelain; obtaining porcelain models implies an effort in terms of time and cost of materials and skilled workforce; since the hollows have a life expectancy of 30-40 castings, their deterioration precludes the proper preservation of the collection.

II. METHOD

A digital mould archive, produced starting from a laser scanner survey, immediately seemed like a possible solution to the issues listed in the introduction to the research team. In order to verify the method, a test was made on six piece-moulds belonging to the figure called “The Giant”.

A. Acquisition

Scan operations, carried out using laser stripe technology (useful for the survey of small objects

characterized by a high level of detail, as artistic artefacts)[3][4][5], have produced four different colorized range map sets (one for each piece-mould). The main shape features and details of each piece-mould were documented using a sample grid never less than 0.25 mm. For each set we took the range map registration in order to obtain a complete 3D digital model of each piece-mould. (Figure 1.).

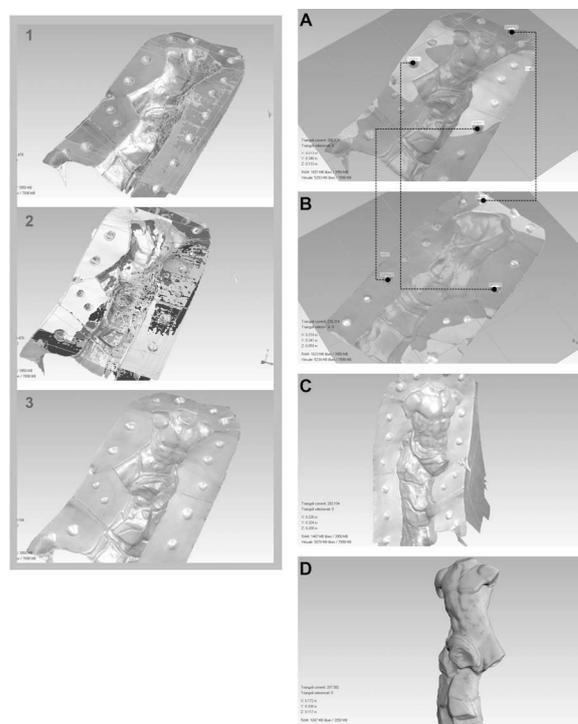


Figure 1. On the left: 1: initial alignment based on homologous points directly identified on the surface. 2: ine registration based on an I.C.P. (interactive closet point) algorithm. 3: Complete 3D digital model of a one piece-mould. On the right: A-B: tie points used to align the two piece-mould related to Giant's bust. C: concave side of the piece-mould. D: Complete 3D digital model of the Giant's bust.

B. Reconstruction

The polygonal surface [4][5] of each piece-mould describes, in its convex side, the shape of the artefact; in this way it is possible to immediately recognize the artefact starting from its piece-mould survey. The piece-mould's contact surfaces are characterized by the presence of joints, which are useful for identifying the right correspondence between the different part of the mould. The accurate survey of these contact surfaces has made it possible to detect joints and, once tie points were placed in each one, they were used to

constrain the alignment between semi-moulds.

These operations resulted in the complete model of the bust and the arms of the “Giant”. According to the contact surfaces and the shape features of the pieces, some solutions have been formulated for the right position of the arms in the bust. Moreover, especially in the absence of references, scholars have the possibility to immediately verify their reconstruction hypothesis (Figure 2.) simply by handling these three digital models, which could be used to visualize solutions in order to share the knowledge with other scholars.

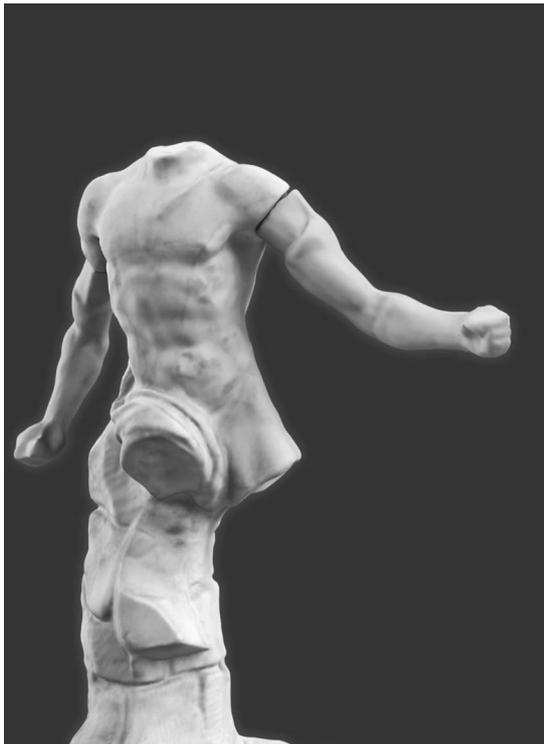


Figure 2. Rendering of the Giant’s bust. One of the reconstruction hypothesis.

C. Storage

A dedicated server is needed for archiving different types of 3D images: those obtained by scanning moulds and others by reconstructing casts and comprehensive models. Moulds, while handled for the scan, are labelled with an automatic readable identification code, which make it possible to identify piece-moulds of the same model. The QR classification data must be entered into a server database. Listed in TABLE I. is the basic information for each piece-mould and for each model, which will be stored in the database. The number of IDs of different piece-moulds varies from a minimum of $2 \times 1 = 2$ to a maximum of about $2 \times 60 = 120$, but on average it is $2 \times 6 = 12$ or $2 \times 7 = 14$. The information about the model can then be extended with XML classification files recording specific features or with other fields deemed useful for research and consultation in the database. In order to carry out the

consultation of the pieces we have to design an ad hoc application that makes it possible to perform specific research queries in the database, presents the resulting information and invokes a program for 3D image viewing and navigating according to their storage format.

TABLE I. DATABASE INFORMATION CONCERNING SEMI-MOULD AND MODEL

Semi-mould	Model
Matching piece-mould ID	Comprehensive model ID
Comprehensive model ID	IDs of different piece-moulds that make up the model
Author	Author
Year	Year
Sizes	Sizes
Shape type	Shape type
Path and filename of original 3D scan	Path and filename of 3D image of the model
Path and filename of 3D image of digital cast	Descriptive sheet

III. CONCLUSIONS

The presented system makes it possible to solve all the problems related to the recovery of historical and artistic heritage represented by the Richard-Ginori collection: reconstruction of the artistic artefacts from moulds in less time than that required by pouring technique; capacity of getting actual mould copies through the use of 3D printers; realization of new moulds from 3D digital models; conservation of the artistic heritage of plaster moulds. As a consequence digital database implementation opens the field to the usability of the models on a global level, both for scientific purposes and for commercial developments. Interesting developments for extending the database are also possible; in addition to the basic information which determines the mould’s identity card, many special features, which create a more sophisticated typological classification, can be associated to individual moulds and especially to models in their entirety. This information can be structured in specific XML files, which are suitable for searches based on ontological features, thus opening the door to powerful modern scientific support tools.

REFERENCES

- [1] M. Levoy (Computer Science Department Stanford University) “Digitizing the Forma Urbis Romae” Proceedings of SIGGRAPH 2000, Computer Graphics, Annual Conference Series.
- [2] Marc Levoy (Computer Science Department Stanford University) “The Digital Michelangelo project: creating a 3D archive of his sculpture using laser scanning” Proceedings of proceedings of Electronic Imaging and the Visual Arts (EVA) ’99, Florence, Italy, March 26, 1999
- [3] G. Guidi, M. Russo, J. A. Beraldin “Acquisizione 3D e modellazione poligonale” McGraw-Hill, Milano, 2010.
- [4] D. Foley; A. van Dam; S. K. Feiner; J. F. Hughes. “Computer Graphics. Principles and Practice”, Addison-Wesley, 1995.
- [5] A. G. Requicha, “GEOMETRIC MODELING: A First Course text for the geometric modelling courses taught at USC” (University of Southern California) Copyright © 1995-1999.

Author Index

- Ács, P., 149
Adembri, G., 189
Aloia, N., 126
Allasia, W., 21
Antonielli, A., 155
Armitage, J., 107
Atkinson, S., 95
- Bachi, V., 58
Bakanas, P., 107
Balmer, J., 107
Bartolini, I., 46
Bellini, P., 52
Bene, A., 101
Bergamasco, M., 113
Bernini, D., 101
Boch, L., 89
Brondi, R., 113
- Ćalić, J., 95
Carrozzino, M., 113
Casey, J., 75
Cenni, D., 52
Chipa, S., 35
Ciapetti, A., 137
Colabianchi, L.N., 183
Concordia, C., 126
- D'Andrea, A., 131
De Michelis, G., 101
De Rosa, R., 169
De Vizia Guerriero, M., 137
Di Carlo, A., 89
di Tondo, S., 189
Donnini, A., 137
- Eversmann, P.G.F., 81
- Felicetti, A., 131
Fresa, A., 58
- Gallo, F., 89
Garra, R., 175
Gherardelli, M., 189
Greller, W., 75
Guasch Aramburu, R., 187
Guha, S., 41
Guidazzoli, A., 29
- Halpin, P., 107
Hudspeth, K., 107
- Iaconesi, S., 62
Ivacs, G., 162
- Liguori, M.C., 29
Longhi, E., 166
Lorenzini, M., 137
Luzzi, C., 46
- Marsella, M., 3
Masci, M.E., 137
Masi, T., 58
Mathe, K., 162
Meghini, C., 126
Merlitti, D., 137
Messmer, G., 143
Montanaro, C., 183
- Nathans, H., 8
Nesi, P., 52
Ng, K., 107
Niccolucci, F., 131
- Patella, M., 46
Persico, O., 62
Piro, F., 137
Pompilio, A., 46
Prandoni, C., 58
- Roselló Gargallo, M., 187
Ruini, C., 46
- Shapira, S., 121
Signore, O., 137
Simonetta, G., 35
Sosnowska, D., 71
Spigarolo, M., 29
Szabó, A., 149
- Tamaro, A.M., 166
Tecchia, F., 113
Timmermans, N., 12
Tisato, F., 101
Toschi, L., 35
Tottossy, B., 155
- Valla, S., 166
- Waller, G., 75
Whatley, S., 4

