IPR Centered Institutional Service and Tools for Content and Metadata Management

Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Michela Paolucci
DISIT, Distributed Systems and Internet Technology Lab,
DINFO, Dipartimento di Ingegneria dell'Informazione,
Università degli Studi di Firenze, Firenze, Italy, http://www.disit.dinfo.unifi.it
{pierfrancesco.bellini, ivan.bruno, paolo.nesi, michela.paolucci}@unifi.it

***Abstract* —** Multimedia services of cultural institutions need to be supported by content, metadata and workflow management systems to efficiently manage huge amount of content items and metadata production. Online digital libraries and cultural heritage institutions, as well as portals of publishers need an integrated multimedia back office in order to aggregate content collection and provide them to national and international aggregators, in the respect of Intellectual Property Rights, IPR. The aim of this paper is to formalize and discuss about requirements, modeling, design and validation of an institutional aggregator for metadata and content, coping with IPR Models for conditional access and providing content towards Europeana, the European international aggregator. The paper presents the identification of the Content Aggregator requirements for content management and IPR, and thus the definition and realization of a corresponding distributed architecture and workflow solution satisfying them. The main contributions of this paper consist of the formalization of IPR Model that enable the shortening of the activities for the IPR resolution, and avoid the assignment of conflicting rights/permissions during IPR model formalization and thus of licensing. The proposed solution, models and tools have been validated in the case of the ECLAP service and results are reported in the paper. ECLAP Content Aggregator has been established by the European Commission to serve Europeana for the thematic area of Performing Arts institutions.

Keywords: institutional archive; content aggregator; IPR Model; IPR management; workflow; grid computing; metadata validation; semantic computing.

I. INTRODUCTION

With the introduction of web 2.0/3.0, and thus of data mining and semantic computing, including social media and mobile technologies most of the digital libraries and museum services became rapidly obsolete and were constrained to rapidly change. Thus the online digital libraries and cultural heritage institutions such as ACM, PubMed and IEEE and the other portals of publishers needed an integrated multimedia back office. In most cases, the cultural institutions see their content ingested, promoted, distributed and exploited by final users via online commercial partners (e.g., YouTube, Vimeo), that may take benefits to commercial resell and/or via advertising. Also some of these social services collect user generated content requesting at the users to select a limited number of permissions they would like to grant at the other users. For example, allowing the public access or not, permitting the downloads, etc.

In Europe and in US, most of the cultural heritage institutions aggregate content and provide them to national and international aggregators such as Europeana in Europe [Europeana], Library of Congress in US [Congress]. Europeana has more than 24 million of content items coming from more than 2200 different content providers, and about 100 Content Aggregators (such as: thematic aggregators and/or national aggregators). Therefore, in order to cope with these needs and participating in the aggregation process towards large collectors and archives, the institutional services need to be supported by **I**ntellectual **P**roperty **R**ights, IPR, centered solutions for content, metadata and workflow management to efficiently manage huge amount of content items and metadata coming

from different sources (local archives, museums) that have to be harmonized and managed for internal purposes and public distribution and aggregation.

In more detail, large Content Aggregators have to cope with the complexity of collecting and aggregating heterogeneous content collections coming from several archives and providing different IPR requirements, different metadata models, different languages, different databases, etc. Thus they have to regularize and augment metadata with specific information needed for the international libraries, such as the (i) clearance and the formalization of intellectual property aspects (i.e., IPR aspects and licensing aspects), (ii) regularization of references to standard dictionaries and taxonomies, (iii) resolution of references to well-known person names, locations, dates, etc., and finally also (iv) adaptation of content format to suitable formats for internet distribution and streaming. In most cases, the Content Aggregators have to put at disposal of their personnel and content providers collaborative tools for IPR formalization, management and resolution; and for content management supporting massive content ingestion and processing. Therefore, specific workflow and metadata processing and enrichment (name resolving, date analysis, linking with open data, creation of relationships, etc.) are needed to cope with the content aggregation and publication duties, at reasonable costs. To this end, specific solutions and architectures are needed to cope with the above mentioned aspects taking into account IPR model assignment, IPR management, content production workflow integrated with the content management.

A. Contributions of this article

In this paper, the main aspects identified to model the IPR centered solution for cultural heritage thematic content aggregators are discussed and reported. The solution has been applied to ECLAP that one of the European Aggregators. ECLAP (European Collected Library of Artistic Performance, http://www.eclap.eu [ECLAP]) has been set up by the European Commission ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme, Theme CIP-ICT-PSP.2009.2.2, Grant Agreement N°250481. ECLAP aggregates content and performs metadata semantic enrichment for Europeana [Europeana]. Europeana (namely the European digital library) does not collect content item files, but only collects classification information (i.e., metadata, descriptors) including the original URLs to the content. The collected data and URLs refer to the original content owner and/or to the Content Aggregator who provided the content metadata. ECLAP is a living lab in which several new research results, technologies and solutions in the area of semantic computing and social media have been developed and put under trial of the final users and institutions. ECLAP has been open for both content and research result experimentations, and presently comprises more than 35 prestigious international institutions coming from 13 European countries plus: Russia, Chile and South Africa (see the full list on ECLAP portal [ECLAP Partners]).

The contributions to the advancements of the state of the art reported in this article correspond to:

- The formalization of IPR Model, algorithms and tools for (i) simplifying and shortening the activities of IPR resolution in the application of permissions, and (ii) avoiding the assignment of conflicting rights/permissions that may come from the combination of user profiles, content kinds, and many other constraints, partially coming from contracts and partially from the technological limitations. Where for rights we may intend actions for example of: play, distribute, streaming, embed. These aspects are not presently supported by standards and research grounded solutions at the state of the art, as discussed later in the paper.
    - Point (i) also refers to IPR Model interpretation that addresses key factors of complexity of real time computation of permissions that can be granted to the users accessing to the protected content in the respect of rights according to the provided platform and context. Even if the metadata are licensed as public domain, the content items may be protected. In this case, the IPR and protection model can be used to enforce conditional access rules by defining permissions and rights for content access. They can be different for different categories of people, and for different kinds of content, and taking into account certain conditions that can only be estimated in real time at the access moment. The IPR Model logical interpretation has to be performed in real time when created to show the implied permissions and when is processed at the content access.

- For point (ii), a specific IPR Model compliant tool named IPR Wizard have been realized to assist the content providers in assigning the licensing and conditional access rules to content collections, avoiding the creation of non-consistent or incomplete permission sets and rules (IPR Models), that do not have to conflict with right expression language standards and/or legal aspects (the logical computation has to be performed in real time as in (i)). Thus guarantees that only content with a formalized, consistent and coherent set of IPR conditions are passed to Europeana to reach the large audience and repositories, and put in publication on ECLAP portal. In the formalization of a set of rights (IPR Model) to be assigned to a digital content inconsistencies may be provoked among: (i) rights/permissions according to their semantics and subsumption relationships; (ii) assigned Europeana.rights models (compliant with the Europeana standard [Europeana]); and (iii) permissions granted to the different kind of users in different conditions.

- The identification of the Content Aggregator requirements and needs, and thus the definition and realization of an IPR centered distributed architecture and workflow solution satisfying them. The aggregators have to cope with huge amount of content and metadata, and perform their massive processing for IPR assignment, semantic enrichment and finally for publication. To this end, a workflow model for content aggregation has to integrate human and automated back office activities for IPR management, semantic computing, content adaptations, metadata quality estimation, validation, translation, etc.

The proposed solutions have been developed and validated in ECLAP infrastructure and service. Until now, ECLAP has ingested and processed about 180.000 object files (images, document, video, audio, 3D, braille, e-books, etc.) with more than 1.1 million items (pages, images, video, audio), and billions of information managed, and thus millions of accesses with IPR processed in the last year of work.

The paper is organized as follows. In Section II, related work is presented to put in evidence the state of the art in the research sectors related to IPR modeling, management and conditional access for cultural heritage content, and about the solutions for automated content management for Content Aggregators. Section III reports the general requirements and the ECLAP architecture for institutional services and tools with workflow details. In Section IV, the detailed formal aspects related to the IPR Model, IPR Logical Model and the corresponding IPR Wizard tool are presented. The exposition highlights the research results provided with respect to the state of the art and the related advantages described above. Section V presents the validation of the solution in terms of evidence and results obtained in the massive usage of the solution proposed, providing statistical analysis related to the IPR Wizard tool usage on ECLAP portal and content ingestion and distribution services, and about the advantages of the IPR Wizard and IPR Model usage. Conclusions are drawn in Section VI.

II. RELATED WORK

This section reviews the state of the art related to the IPR modeling and management for cultural heritage content, and about the automation of content management for Content Aggregators, in which the IPR techniques have to be enforced. The analysis is putting in evidence a number of unsolved problems for creating an IPR centered content management solution for Content Aggregator including (a) IPR modeling, execution and management, (b) automated content management and semantic computing. Thus, the solutions presented in this paper have been designed and developed for ECLAP and can be of value for many institutional archives and content aggregators.

A. Initial requirements on IPR Management

The management of the intellectual property on an online cultural heritage archive and portal may present a number of problems and phases, such as:

A. **rights clearance** which corresponds to the activity of the identification of the rights that effectively are owned by the content provider and may be passed to the archive for distribution and aggregation. This corresponds to the phase in which the Content Provider, CP, identifies on paper contracts rights/permissions that could be potentially grated/exploited. This activity is typically performed by legal and administrative personnel as desktop work. For example, the CP may get from a contract to have the right for distributing on Internet a content collection. On the other hand, any owned right also implies to get the subsumed rights according to a

standard meaning of rights [Delgado et al., 2003]. For example, the owned right also implies the rights to perform mobile distribution, in any resolution and towards applications. On the contrary, the CP may realize to have the need of distributing content on mobile, and may have difficulties in realizing it from the contract, may even provide higher level rights. Moreover, they may realize to do not have the correct right for distribution, thus to have the need to clear a right by requesting the signature of some new contracts with right owners. For these activities, a clear view of rights to be provided to their final user and the relationships among them can be of help.

B. **formalization of the IPR profile** (set of rights / permissions that can be granted) for each single content or collection. Each right may have a set of associated conditions constraining its applicability and may be applied to content files and/or to the corresponding *metadata*. The applicability can be constrained to user category, device, and contexts; for example, play on mobile in Italy. The provided rights of the set (profile) have to be un-ambiguously defined and have to be coherent according to the subsumptions. For example it has not sense to provide the right/possibility of change a video and not the right of streaming. Legally, to grant a certain right also imply to grant a set of subsumed granted rights. In most cases, the relationships among rights are legally defined and can be formalized according to definitions [WIPO]. The activity of IPR formalization can be performed according to the identified rights extracted from phase (A). The formalization has to be performed in the workflow of content management, taking into account that some of the activities performed on content can only be performed by exploiting specific rights, and thus having a grant for doing that. For example, a file can be adapted (for example, scaled up for mobile streaming) only having the legal permission for doing it..

C. **right enforcement**, is the set of technical implementations (encryption, authentication, certification, etc.) that assure in a certain device/player that a certain right (of a given IPR profile defined in phase (B)), is respected. For example, the enforcement of right for play video in a media player has to guarantee that a user would be capable to play only the videos for which the user has received the grant authorization. The enforcement may implies to adopt one or more specific technological protection solutions, such as special players, encryption model, certification and authentication, etc. [Bellini, Nesi, Pazzaglia, 2013]. The activity of right enforcement is typically performed by the tools which are used for content access, i.e., media player, browser, decoder, media player on mobile, etc.

Moreover, when the content is delivered to third parties, such as the posting of metadata to Europeana (central aggregator or distributor), specific guidelines and formulation of related IPR have to be followed; thus requesting additional information that has to be formalized [EuropeanaRights] and not conflicting with the legal aspects identified in (A) and the formal definition of the IPR profile/model of (B) phase. For example, if the aggregator would like to have metadata as public domain information, the phase (A) could accept in ingestion only metadata and content with that kind of IPR Model associated to metadata, and IPR Model from phase (B) has to be coherently created.

B. IPR Management and Conditional Access for Cultural Heritage Content

Before the distribution of content the above mentioned phases, from (A) to (C), have to be covered with a coherent solution.

The formalization of an IPR profile (as stated in (B)) for a given content can be performed by producing a set of license models. A license model aims to regulate a possible use of a digital content (e.g., play on a specific device, play for a specific time period, play on a specific territory, print for educational purposes, etc.). In order to allow interoperability, several standards have been developed for the formalization of licenses such as the MPEG-21 REL (Right Expression Language) [Wang, 2004], ODRL [Iannella et al., 2012], OASIS XAMCL [Moses 2005]. For the definition of the relationships among rights, possible rights ontologies or vocabularies could be used such as MPEG-21 RDD (Right data dictionary) [Wang et al., 2004], the ODRL vocabulary [Iannella and Guth, 2012] or the Access Management Ontology. In [Chang et al., 2003], a visual tool for defining authorization workflow models for e-commerce application has been proposed. The mentioned models allow the formalization of licenses authorizing a given user to exploit certain rights on a specific identified content, under certain conditions. On the

other hand, the concept of IPR Model should allow the formalization of the whole domain of licenses that can be associated with a given content or collection, under different conditions and user profiles. A partial solution (only modeling the set of possible rights for a given object) has been adopted to formalize the Potential Available Rights, PAR of a given content in AXMEDIS [Bellini, Bruno, Nesi, 2011]. Moreover, the Creative Commons licensing framework [CC], [Abelson et al., 2008], allows users to formalize the usage of content that the owner would like to share, applying some legally formalized restrictions on its reuse (e.g., no commercial use, attribution, no derivative). This means that the CC models address only partially the problem of phase (B) of right formalization since CC models assume that the content can be distributed, and nothing about phase (C) since the right enforcement is not performed.

For the technical aspects of phase (C) of rights enforcement, the IPR management systems can be classified in (i) Digital Rights Management (DRM) [Ku and Chi-Hung, 2004] systems allowing to control 'all' the possible different uses/rights of a digital content (e.g., distribution, enhancement, play, adaptation), or (ii) Conditional Access Systems (CAS) allowing to control only the access to the digital resource that is provided with some encryption until the access is completed. This means that the main difference from DRM and CAS mainly resides on the enforcement of rights/permissions and on the constraints that increase the complexity of technological protection solutions adopted for the grant authorization and for the enforcement. For example, the DRM aims to guarantee that once content is accessed and downloaded by a granted user it will not be usable by some other, while most of the CAS protect only the content before the access is performed. Currently, the media industry is much more oriented on CAS solutions. In any way, solutions are present for both CAS and DRM, addressing the protection and limited IPR management for audio/video content access (see for example solutions of Microsoft, Apple and Adobe for PCs/mobile/tablets and by Irdeto, Nagravision and others for satellite, IPTV, cable TV digital transmissions).

On the other hand, in the case of cultural heritage content, a limited number of rights can be owned by the archive, while the content distribution may be performed towards specific protected user categories such as: impaired, students, researchers, professors, for which some of the right constraints have to be relaxed and these aspects have to be formalized into the right profile identified in phase (B).

The above mentioned standards and solutions are unsuitable to be used as IPR Model for defining the whole set of IPR rules to be applied to cultural heritage content according to the different kind of users, devices, content type, etc. Most of them are only focused on formalizing the single license to be attributed to specific users, the license model of the above mentioned standards request a unique user identification. That approach is too expensive to be used for cultural heritage content, where the users accessing to a certain content are not known in advance, and the real time creation of licenses would lead to the computation of trillions of licenses, that is an unacceptable cost. These facts motivated the study and development of the proposed IPR Model and solution presented in this paper and applied into ECLAP IPR Model and tools.

C. Automation of Content Management for Content Aggregators

The Content Aggregators have to cope with a large variety of content collections coming from several Content Providers, CP, with the aim of processing content and metadata, targeting the final goal of enriching and publishing the content towards several different directions and exporting them to other services. The main problems to be solved are related to the IPR addressing and to the massive management of content processing and the automated application of semantic computing algorithms. The IPR aspects have been addressed in previous sections.

To cope with the aspect of content management, media or semantic grid tools are used. Among the several solutions: MmGrid supports interactive applications with graphics, rendering, streaming, and tele-immersion [Basu et al., 2003]; AXCP provides an integrated approach to perform both content management and semantic computing [Bellini et al., 20011]; GRISINO combines web services with intelligent content and grid computing, where its integration with workflow management is performed via web service [Toma et al., 2006]. GridCast is a service-oriented architecture for broadcasting media via IP [Harmer et al., 2005]. GridCast provides some content management capabilities, mainly for broadcasters as controlling playout servers. In [Volckaert et al., 2008], the usage of grid solutions for media production and distribution have been introduced, focused the application of the technology on the content delivering and management. In [Messina et al., 2011], HMNews system for automated

aggregation and consumption of information streams from digital television and the Internet has been presented. These solutions provided support for video segmentation, annotation and indexing, while many other aspects have not been considered, such as text processing, semantic computing, etc. In the management of News, it is quite common to perform natural language processing analysis to extract information. In the usage of news, NewsML as well as other formats provides a limited number of metadata [Bellini, Bruno, Nesi, 2011]. In Google News tool [Dolan and Pedersen, 2010] automatically aggregates information extracted from newspapers and magazines to offer personalization service displaying news articles tailored to the user's interests. Most of the high-level activities may be passed from grid to a workflow management system, which also plays the role of integrating other tools via programming [Yu and Buyya, 2005], [VanDerAalst and Van Hee, 2002]. In general, the above mentioned solutions and tools provide limited IPR and semantic computing capabilities for processing content metadata for cultural heritage content; where metadata are typically provided in several different metadata formats, standards and dialects.

On the other hand, the metadata enrichment has to be addressed by specifically algorithms and tools. For example, processing metadata and exploiting data mining and natural language processing for extracting and disambiguate VIP names and linking them on dbPedia, to extract and disambiguate geographical names, to understand and disambiguate dates, assess metadata quality, etc. Also these aspects are specific of cultural heritage content aggregation and have to be integrated into the workflow and semantic computing tools by programming. If the workflow management tool can support the harmonization of the work to be done, by who and when, the media grid has to provide support for parallelizing the process and executing these complex algorithms on multiple computer machine nodes.

Among these tools, only few of them address the problem of massive IPR management and licensing productions in the management language, one of the is the AXCP which provide an integrated java script language that may be used for generating MPEG-21 licenses and semantic computing integrated into content workflow [Bellini et al., 20011].

III. General Requirements, ECLAP Architecture and Content Management

Before to present the detailed aspects related to the formal IPR Model and the corresponding tools; this section reports the general requirements and the ECLAP architecture for IPR centered institutional services and tools.

A. General Requirements for Content Management

Content Providers, CP, provide to the ECLAP Content Aggregator, CA, both content files and metadata. CP can be archives, museum, theatres, educational institutions, foundations, etc. According to ECLAP aim, the metadata have to be provided to Europeana only after to have enriched and linked them to a corresponding reachable digital resource (via an URL) and when the IPR issues have been correctly defined for each content file. In this context, ECLAP acts as an intermediary between several CPs and Europeana, by aggregating and maintaining accessible content files, via the URLs to the content file maintained accessible by ECLAP on internet. Moreover, ECLAP supports the CPs in their work: starting from the ingestion phase, passing through the definition and management of IPR permissions and licenses on contents and metadata, managing the real content and providing to ECLAP users the typical services of a Social Network. The main requirements for the typical CA back office for cultural heritage are reported as follows. The full set of detailed requirements is reported in [Baltussen et al., 2012], [ECLAPSupportPage].

The infrastructure for a cultural heritage CA has to be capable to **ingest** a wide range of metadata formats (such as: Dublin Core, METS, MPEG-21, etc.) coming from different channels (http, ftp, oai-pmh, LOD, etc.), and to map them on a single unified model without losing any semantic meaning of the original metadata model [Bellini and Nesi, 2014]. The ingestion process should acquire both metadata and content files, link them together, collect them on a suitable storage, and ingest IPR information if available at the ingestion time. The solution has to provide support to perform **manual content enrichment,** such as metadata language translations and validation; addition of comments/annotations and tags; social media promotion; voting/rating; publication to other portals and social media; editing and correcting metadata; quality assessment, etc. This implies that each single expert may be authorized via IPR control to perform several different content activities in the workflow and life cycle. The system

architecture has to be capable to perform **automated activities,** such as: estimation of technical parameters (duration, size, etc.), extraction of content descriptors, indexing, automated translations, searching for VIP names, geonames resolutions, linking with LOD, metadata assessment (completeness and consistency [Jung-Ran, 2009], [EBellini and Nesi, 2013]), IPR verification for consistency and completeness, and content adaptation. Among these activities, the content adaptation aims at producing different file formats to satisfy a range of distribution channels and final users' devices, for example several file formats of the same video. The adaptation process can be performed only on content for which the CP has granted the rights to CA to perform the content adaptation. The **back office architecture has to be scalable** to cope with large number of transactions on metadata information and files per day. Therefore, the execution of a massive and distributed processing on content resources is needed, and can be provided by using some media grid solution.

The approach has to be capable to assign a specific **workflow life-cycle for each content item in production.** For example, a different life cycle can be associated for producing content for different international aggregators, for different content collections, IPR models, etc.

The solution has to **harmonize and coordinate the activities of both human and automated processes**. For example, identifying when the human actions are needed, taking trace of the performed manual activities, blocking the humans when the automated elaboration has locked the resource and vice versa. Among these activities of the workflow the CA has also the needs of (i) logging and keeping trace of metadata versioning, (ii) formalizing and managing different roles/capabilities of users (i.e., enricher, publisher and validator), (iii) providing tools for: multilingual metadata editing, IPR managements, content editing, etc.

Finally, the CA infrastructure has to cope with the (i) **IPR Model** formalization avoiding incoherencies, (ii) assignment the model to single content and/or collection and (iii) perform the enforcement at run time by verification of user profiles and conditions. The adopted IPR Model has to be executable to regulate uses/accesses to the digital content, and the content manipulation and reuse. A different IPR Model can be applied by the CP on content collection or sub collection. The final collector of metadata, such as Europeana, has the need to be authorized of providing the metadata for several different purposes, for creativity, cultural promotion, education, etc. Moreover, the process of enriching metadata is facilitated if the IPR Model for metadata is public domain (e.g., CC0 1.0 Universal Public Domain License) [CC].

B. Overview of ECLAP Architecture for IPR Enabled Content Management

In this section, an overview of the ECLAP architecture is reported with the aim of presenting and highlighting the components involved by the IPR management and content workflow. The ECLAP architecture has been designed to satisfy the above presented requirements and consists of three main layers: ECLAP back office, a distributed storage shared with the several servers and processes, and the ECLAP Front End (**see Figure 1**). The **ECLAP Back Office** is implemented by a media grid to implement a parallel and scalable architecture for scheduling and managing processes to automatically perform back office activities, exploiting AXCP [Bellini et al., 2011]. The back office activities can be grouped in the categories of **Content Processing** and **Semantic Reasoning.** Most of them are activated when content is updated, deleted, exported, etc., other are periodically executed.

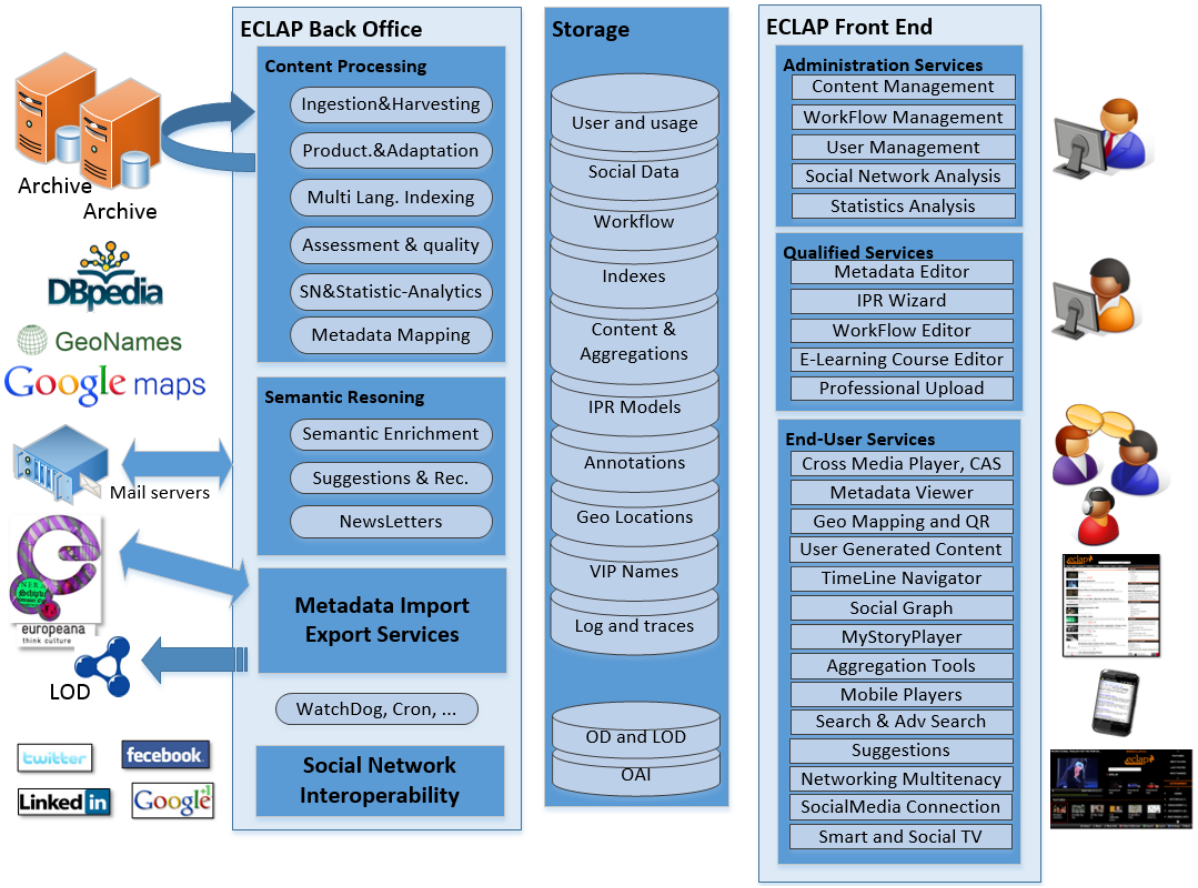

Figure 1 - ECLAP Architecture Overview

As regards the **Content Processing**, it includes the activities of:

- **content ingestion & harvesting:** metadata ingestion from the several CPs performed by means processes of content harvesting and acquisition. Metadata are typically retrieved from the CMSs and databases of CPs by using several kind of protocols (e.g., web services, rest, OAI-PMH), while digital resources are downloaded from the provided URL or FTPs.
- **Metadata Mapping:** the ingested metadata sets are provided in different schemas and have to be mapped according to the ECLAP ingestion semantic model [Bellini and Nesi, 2014].
- **Production and adaptation** consist of converting the digital files (audio, video, images, documents, etc.) into the several formats needed to play them on different devices and services; and to extract thumbnails for the front end. These kinds of processes are time consuming and thus may be performed in reasonable time only on a media grid infrastructure.
- **Multilingual indexing** of cross media content to provide search service on the ECLAP Front End. In this case, the complexity is mainly due to the heterogeneity of metadata for the different kind of resources, and the optimization indexing model to maximize the precision and recall with respect to the user intention [Bellini, Cenni and Nesi, 2012].
- **assessment and quality assessment.** This process consists in the general quality assessment of the ingested content. For example, to identify content that does not present the minimum level of completeness and consistency of metadata with respect to the acceptance level imposed.

The activities of **Semantic Reasoning** include the Semantic Enrichment to enrich the content descriptors by processing metadata with natural language processing algorithms to extract person names, geographical names and dates; with the aim of finding synonyms, disambiguation of names, geographical names and dates; connecting person names with dbPedia and geographical names with geonames; regularizing dates for temporal ordering and navigation. These activities are time consuming and imply the connection with external services such as dpPedia,

Geonames.org, etc. Another activity of semantic computing consists in computing similarities among users and content, to produce suggestions and recommendations.
The back office activities regarded as **Metadata Import Export Services** include processes and services for: producing metadata in the EDM (Europeana Data Model) mapping a part of the ECLAP semantic model. Once produced, the metadata descriptors are exported via an OAI-PMH to Europeana. Moreover, also LOD of the ECLAP semantic model are made accessible [Bellini and Nesi, 2013], see for example the access of Linked Open Data published by ECLAP with Linked Open Graph [Bellini et al. 2014]. The **Social Network Interoperability** processes include activities for the integration of ECLAP infrastructure with third party social networks and services.
In order to complete the IPR management a set of tools has to be provided. The **ECLAP Front End** services can be classified in Administration Services, Qualified Services and End-users Services. The **Administration Services** are related to the management of: content, users, workflow, social networking and statistical & analytics tools. The **Qualified Services** at disposal of group administrators and trusted users, are: Metadata Editor (for editing and validating metadata); IPR Wizard to define IPR Models and associate IPR Model to content and collection (discussed in the sequel of the article); and workflow management. The **End-User Services** are oriented to support ECLAP final users' activities on web and mobile, and also in this case, the tools have to take into account the IPR aspects. Among the most important users' activities tools influenced by the IPR constraints we have the C**ross Media Player** which is capable to produce and play on browser and mobiles: documents, video, audio, e-pub, 3D, animations, slides, etc., according to the device and browser used. The Cross Media Player also includes a CAS **(Conditional Access System)** to implement content protection enforcing the IPR Model to avoid the access of un-authorized users, provides video streaming in progressive download adapting the video format and bitrate that can be sustained by the established connection and device, and presents the links to perform content download and/or embedding according to the IPR Model associated with each specific content element. Moreover, the **Metadata Viewer** provides metadata in multiple languages: DCMI metadata consisting of a set of Dublin Core fields; Taxonomy as multiple classification terms selection; Groups assignment (ECLAP Groups); Resource data by selecting one or multiple files; Workflow type associated with the content life-cycle. Among the valuable metadata for the user a description of the declared rights is also reported in terms of IPR Model. This clarifies the content access with respect to the different kinds of users and rights. The Metadata are also enriched by links to VIP names of dpPedia, geonames, and to ECLAP users mentioned in them. This enrichment has to be performed without modifying the metadata provided by the CP, since the metadata modification would implies to have the IPR to performed metadata adaptation.
Others services are briefly described as follows. Geo Mapping and QR: Geo mapping technology is used to provide content access via geographical coordinates and may exploit the Google Map service for rendering and access. User Generated Content allows final users to upload content on the portal. The TimeLine Navigator provides a browseable model for navigating among dates (date of production, first performance, last update, etc.); Social Graph provides at the users a browseable model for playing and navigating among the several relationships established by implicitly citing users, resources, names, geonames, and open data into metadata; MyStoryPlayer provides access for producing and accessing audiovisual annotations and synchronous playing of multiple videos (for example for multicamera view); Aggregation Tools support playlists, collections, courses; Mobile Player provides support for smart phone e pad/tab to access at the right player and content format; Search and Advanced Search module allows the search and retrieval of content with faceted support, by using full text as well as advanced operators and interface for querying. It also performs query expansion by translating and expanding keywords and propagating queries to other service portals such as Europeana; thus exposing results to the users. Suggestion module provides suggestions about other users and content in different contexts; it also provides ready to access recommendations such as: most played content, recently uploaded, most voted, content of colleagues, keywords clouds, query cloud, etc.; Networking Multitenacy provides support for creating custom branded working on groups, pages, events, forums, and blog; Social Media Connection provides support for social network interoperability exploiting back office processes for: migration of contacts, promotion and publication towards other portals and social networks, etc.; Smart and Social TV module provides connection and data towards ECLAP application on Samsung Smart TV smart hub.

C. Automated Content Management and Workflow

In order to better understand the content and metadata management, it is useful to describe the scenario in terms of content status, workflow roles, and procedures, etc. All the back office processes may work concurrently in accessing and updating information on the shared and distributed storage. Media grid back-office services access and process content on multiple and parallel computational nodes concurrently with the users' activities performed on the ECLAP Front-End. Activities of metadata enrichment, content adaptation, validation, IPR Model definition and assessment have to produce consistent results. Each content item or collection managed in ECLAP is associated with a specific workflow and lifecycle; see for example in **Figure 2**, the workflow for Europeana is reported. In this case, content can be: proposed; uploaded/ingested; automatically enriched by the processing rules; manually enriched via the metadata editor; manually or automatically associated with an IPR Model; accessed and validated; automatically assessed to verify the availability of the minimum quality requirements by the automated tools; and finally proposed to be published to Europeana.

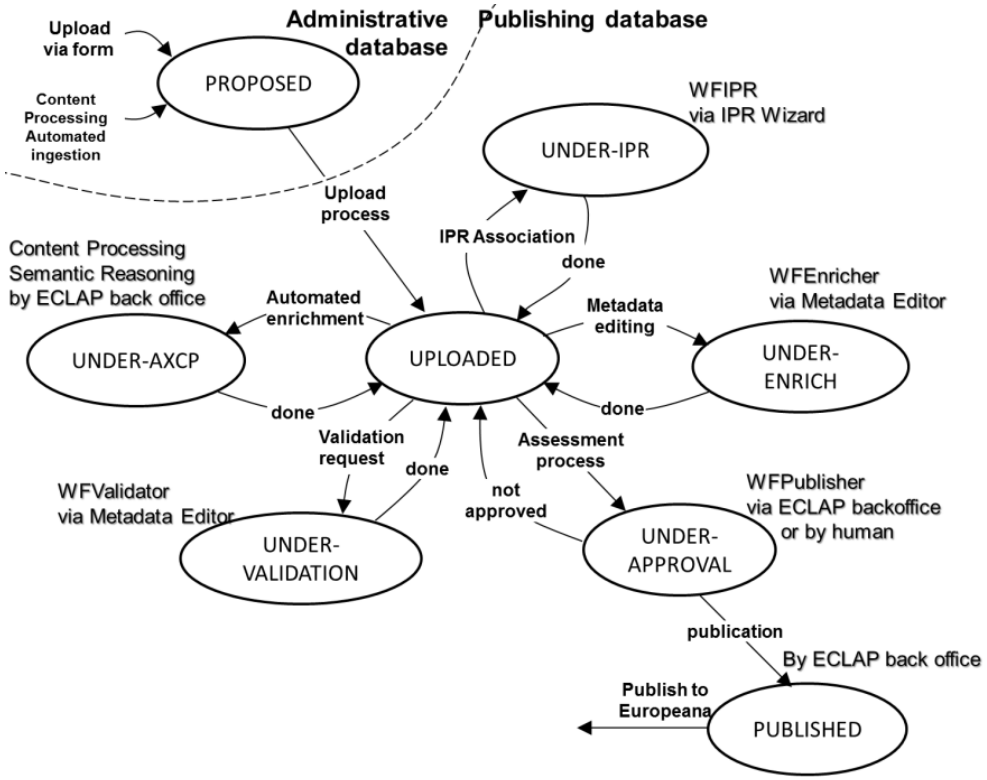

Figure 2 – Example of ECLAP Workflow diagram for Europeana content life cycle. The activities which can be performed without looking the content are not reported.

In order to implement a high quality content enrichment process, each specific activity on a given content has to be granted according to the corresponding IPR Model and according to the skill and role of the involved personnel. Qualified users may be delegated to perform specific activities on the content by using corresponding tools. The control for right exploitation has to be applied to final users and to the qualified users that are collaboratively participate to the content production in the workflow. Metadata Editor is the tool for editing, enriching and validating metadata, while each action has to be authorized and granted by the computation of the IPR Model. To this end, changes made on metadata have to be tracked to maintain their history including who made the change and when to have the evidence of the work performed and eventually recover wrong situations. According to the user role and workflow phase, the Metadata Editor is activated for enrichment or for validation. Therefore, for the content management, specific user roles have been assigned by each CP at qualified users. To this end, formal roles have been defined with parameters as the Content Provider ID, CPID the set of language assigned. In fact, some of the roles can be assigned for a subset of languages only.

Therefore, the identified user roles related to the content management and workflow are:

- **WFIPR (CPID):** definition and validation of IPR Models, and IPR assignment to the content of the CPID; by using the **IPR Wizard** and during the Upload for the IPR Model Assignment, also called **IPR Manager**.
- **WFEnricher (CPID, {Languages}):** metadata enrichment and changes in the specified languages (add, edit metadata), via Metadata Editor.
- **WFValidator (CPID, {Languages}):** validation of the metadata for the identified language. The metadata fields can be singularly validated until the object may pass the whole approval phase, via Metadata Editor.
- **WFPublisher (CPID):** to take the final decision for publishing content from ECLAP to Europeana. The publishing of single or groups of content can be also performed by using the above mentioned **Content Management** and/or by means of specific procedures on the ECLAP Back Office.

Please note that, the many back-office activities are not associated/authorized with/by any specific user's role since they are executed as processes running at administrator level in the ECLAP Back Office according to the IPR.

IV. Modeling the IPR Model and Wizard for Cultural Institutions

The general requirements related to IPR presented in Section II.A and those for content management reported on Section III.A have to be refined to cope with the details and complexity of the IPR management. The final goal is to associate at each content element (or collection) an IPR Model to define its accessibility with respect to ECLAP users' categories and for the publication of complete information towards the general aggregator, e.g., Europeana.

Once the CPs have understood, from the legal point of view, their position with respect to the content (the so called phase (A) of Section II.A), in phase (B) of Section II.A some assistive model and tool have to guide them on defining and assigning the IPR Model to content (with related permissions and conditions). This activity has to be performed in according to the above described content life cycle. The aim of phase (B) (Section II.A) consists in supporting and guiding the CPs while creating a correct IPR Model based on a set of coherent rights that can be enforced into the players and tools of the service (phase (C). As discussed in Section II.B, the mentioned standards and solutions are unsuitable to be used to formalize IPR Model for defining the whole set of IPR rights and rules to be applied to cultural heritage content according to the different kind of users, devices, content type, etc. Formal licensing solution are focused on formalizing the single licenses to be attributed to an users.

The main detailed requirements of the IPR Model and tools have to:

- allow the creation of a set of IPR Models for each CP that can be applied to single content or collection.
- avoid to the assignment rights/permissions which are different with respect to those really owned by the CP (those cleared during the phase (A) of Section II.A). This can be performed helping the CP to understand the meaning about the right they are selecting. This action can be performed by providing the evidence about the rights implied by each right selection.
- avoid the production of an IPR Model which may assign inconsistent rights/permissions on the same model, for example, allowing to exploit a right and blocking the permission to exploit a subsumed right.
- give the possibility to apply and modify an IPR Model to a collection of multiple content without constraining to edit a single IPR model for each content element.
- allow to associate the IPR Model to single content or collections during the content ingestion, or at any phase of the workflow. The assignment of the IPR Model at the ingestion may allow optimizing the activities of content processing; for example, the production of model content formats if the permission for mobile distribution are not intended to be granted.
- allow to insert in the IPR Model additional conditions according to MPEG-21, ODRL, OASIS XAMCL, such as nationality, times of plays, temporal usage (from date to date).

Inconsistency among rights/permissions can be due to (i) the definition of limitations that cannot be enforced (phase (C)) in a certain device and context; for example, they may request to avoid images to be copied if they are visualized on a computer (i.e., a snapshot is always possible); (ii) inconsistencies among the rights; in fact according to MPEG-21 RDD some right is subsumed by other rights. For example, the grant of right of adaptation implies that of copy, the grant of download implies that of play and thus of streaming down (in fact, it has no

sense to allow the download of unprotected content and forbid the content play). These ontological relationships among rights have to be taken into account in formalizing the IPR Model, in the tool for the IPR Model formalization, and in the execution of the IPR Model on the enforcement phase in the players.

A. IPR Models Definition

Given the variety of CPs and of related needs on their content, a general and flexible IPR Model formalization has been defined. The IPR Model is grounded on the conceptual relationships among rights defined in MPEG-21 RDD, while the defined IPR Model presents relevant innovations that simplify the IPR management of cultural heritage content for the Content Aggregators and archives, satisfying the above presented requirements. Classical and standard licensing models such as MPEG-21 REL and ODRL are unsuitable to satisfy the above mentioned requirements and IPR workflow, since they are focused on producing licenses, as previously discussed. The proposed IPR Model allows formalizing what can be done on the content by each user category, and a range of devices, and which are the permissions. All the non-authorized rights are forbidden for default.

The IPR Model has to allow to define a set of IPR permissions for each different geographical area (localization or domains, for example the University of Florence), for each different devices, for each different content kind and type of user. In fact, it is quite common to have different distribution contracts and rights for each country. A distributor could be entitled to distribute content in a country via streaming and may be disallowed to make the same in other countries. Another important aspect consists in the specification of different rights for different resolutions of the same content. Typically, the CPs allow to download low resolution versions of audiovisual content while high resolution versions are reserved for trusted users and/or for commercial purposes.

Therefore, according to our analysis a formal IPR Model has been defined and formalized as follows. The management of an IPR Model addressing location, users, devices, etc., put in evidence that a single IPR Model can formalize the whole set of permissions that could be granted in millions of licenses of classical DRM solutions, thus reducing the computational complexity.

The set *IPRModels* of possible IPR Models can be defined as follows:

$$IPRModels \subset ModelID \times CPID \times CPLicensePageUrl \times Description \times Europeana.rights \times IPRLocalization$$

Where:

- *ModelID* is the set of possible IPR Model IDs, each model ID is the identification of the IPR Model, which can be also used in the ingestion phase for shortening the assignment time;
- *CPID* := $\{CP_1, \ldots CP_n\}$ is the set of possible content provider IDs;
- *CPLicensePageUrl* is the set of possible URLs, each of them links to a web page in which eventual conditions to acquire additional rights;
- *Description* is the set of possible model text descriptions;
- *Europeana.rights* is the set of possible license URLs used by Europeana as explained in the following;
- *IPRLocalization* := $\wp(Location \times \wp(Permission))$ is the set of possible association of country codes with a set of permissions;
- *Location* := {other, it, en, de, fr, …, or web domain} is the set of country codes with the additional 'other' country representing any other country, or is an IP domain mask;
- *Permission* := $\{ p \in ContentKind \times Right \times UserKind \times \wp(Condition) \mid right(p) \in possibleRights(contentKind(p)) \}$ is the set of possible permissions considering that not all rights can be associated with all content kinds;
- *ContentKind* := {Video, Audio, … } is the set of possible content kinds;
- *Right* := { Play, Play-MD/LD-PC, Play-HD-PC, Download-HD-PC, DownloadHD-Mobile, Embed …} is the set of rights, where HD is for High Definition, MD for Mid Definition, etc.;
- *possibleRights*: *ContentKind* → *Right* is a function associating for each kind of content the rights that can be used on it;
- *UserKind* := {PublicUser, CPGroupUser, EduResUser, TrustedUser} is the set of possible kinds of users;

- *Condition* is the set of possible conditions defined as additional constraints (such as those that can be associated to rights in standard MPEG-21, ODRL, OASIS XAMCL) for example: the expiration date, the duration of the validity, etc.;
- $\wp(X)$ is the power set of X, $\wp(\{x,y,z\}) = \{\{\},\{x\},\{y\},\{z\},\{x,y\},\{y,z\},\{x,z\},\{x,y,z\}\}$.

The IPR Model definition has to be completed with the elements presented in the following sections, defining the logic and the relationships coherent with Europeana.rights values, the relationships with the user profile, and the relationships among possible rights, etc.

B. Europeana.rights vs Permissions and Restrictions in the IPR Model

In the above presented IPR Model, the **Europeana.rights** variable can assume a limited number of possible values according to Europeana EDM model [EuropeanaRights]. The document and guideline for the attribution of Europeana.rights does not report explicitly that not all values can be applied for non-fully accessible content, so that inconsistencies are easily be created from the permissions imposed in the IPR Model and Europeana.rights declared on the content metadata. To avoid them, a formal and legal analysis has been performed to identify the Europeana.rights values that can be associated with different sets of *Permissions* adopted to formalize the IPR Model. As a conclusion of this analysis, the **IPR Logic Model** proposed in this paper divided Europeana.rights in two categories that can be applied to different conditions of IPR *Permissions*. This approach helps the IPR Managers to take the correct decisions and avoids the creation of inconsistencies among the selected *Permissions* in the IPR Model and the *Europeana.rights* formalization. The two identified cases are discussed in the following as: public accessible content, and content that is accessible with some restrictions.

For publicly accessible content, it is intended content which is freely accessible on Internet without access restriction (full access permission for public on the portal and outside). Thus, the associated possible values of Europeana.rights mainly refer to CC (Creative Commons [CC]), plus others that have been identified from public domain, that have been specifically formalized for Europeana (such as the first and the latter of the following list). Each CP has to associate with its public accessible content a legal statement, among the possible values ordered by the less restrictive to the most (see guidelines for their meaning [EuropeanaRights]):

- Europeana: Unknown copyright status (Unknown)
- Public Domain Mark (Public Domain Mark)
- Public Domain Dedication (CC – Zero universal)
- Attribution-ShareAlike (CC BY-SA)
- Attribution-NonCommercial (CC BY-NC)
- Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)
- Attribution-NoDerivs (CC BY-ND)
- Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)
- Europeana: Rights Reserved - Free Access (Rights Reserved - Free Access)

Please note that, the CC is a set of legal (licenses) models to formalize what the users, accessing to the content, have to legally respect. CC models are very different from DRM and CAS solutions since the CC are not supported by technological solutions to perform the right enforcement (see phase (C) of Section II.A). The CC licensing models are formalized only at legal level and may be associated with a content file by placing a corresponding mark / logo. On the other hand, the content owner associating a CC license does not have any confidence that those legal licenses will be respected by the final user.

Certain content may be distributed only with some restriction to its access, and thus the CP has to be capable to impose a limited number of permissions to users (for example, defining that the ECLAP public users can play a content, but not download it). In these cases, the above mentioned CC licenses cannot be used, since all of them give for granted the full access to content and impose only restriction on the final user behavior. For example, the assignment of CC BY-NC-ND to a content that is distributed with some restriction (for example only in stream), would legally violate the CC license that grants to the users the content access and thus free download. In this

case, we have content access with an IPR Model providing only some permission (and thus restricting the access). As a consequence, the CP may only select the Europeana.rights among the following possible values:

- Europeana: Unknown copyright status (Unknown)
- Europeana: Rights Reserved - Restricted Access (Rights Reserved - Restricted Access)
- Europeana: Paid Access-Rights Reserved

C. Classification of Users vs Permissions

In the IPR Model formalization users' profile plays an important role since each set of permissions is associated with a specific category of users. The users involved in the IPR management can be users registered to the ECLAP or not (e.g., *public users*). Each registered user may have additional roles: each role can have a set of IPR permissions associated to it. The main users' categories are: **TrustedUser** (a restricted number of users authorized to have a high level of control, they are the administrators), **CPGroupUser** (registered users on the portal which are also subscribed to a specific group by which a certain content is distributed, associated with), **EduResUser**, Education and Research users of the CPGroup (users that present on their profile and have been recognized to be students, professors, researchers, etc.), **PublicUser** (users that are not logged into the portal, and may be not registered), etc. These users' categories provide increasing restrictions in terms of content access.
It should be noted that the *IPR Managers* are *TrustedUsers* with the specific role of IPR management. For example: if an *IPR Manager* assigns a permission to a *Public User*, the system has to automatically associate the same permissions to all the registered users (*Group Users and Group and Educational Users*). Note that the *Trusted Users* always have all the permissions.

To complete the view, an example of IPR Model for DISIT CP for distributing Videos and Audio in Italy and Germany (with a temporal condition) is formalized as follows:

(Distrib1, DISIT, http://www.eclap.eu/3010, "for distributing in Italy and Germany",
http://www.europeana.eu/rights/rr-r/, {
(*it*, { (*Video*, *Play-LD/MD-PC*, *EduResUser*, {}),
(*Video*, *Play-HD-PC*, *CPGroupUser*, {})),
(*Audio*, *Play*, *EduResUser*, {}) }),
(*de*, { (*Video*, *Download-HD-PC*, *CPGroupUser*, { (FROMTO, 2012, 2013) }) ,
(*Audio*, *Download-PC*, *CPGroupUser*, {}) })
}
)

As a result, an IPR Model can formalize permissions for different kind of resources, users and countries. It is a change of focus with respect to the licensing models for DRM proposed in MPEG-21 and ODRL, which are based on formalizing for each content, single rights for each specific person [MPEG21-REL]. In those models, the main generalization is the concept of domain, by which the specific right for a specific content is associated to a group of users or devices. The proposed model presents a clear advantage for cultural heritage collections and business users, since they have to mimic for each collection the contract they have. In fact, in content collections as well as on artistic works, it is quite frequent to have the need to formalize a unique contract with rights covering a large set of different resources and media, or even treating all the media as works/products of the creative activities. Thus a unique IPR Model includes the formalization of the whole set of permissions for the content collection, addressing in a single model all content kinds, resolutions, users kinds, etc.; thus opening the space to add more content and data in the same collection as well.

D. Relationships among right permissions and users' kinds

According to the MPEG-21 ontology (as well as to other ontologies among rights), to grant a certain right/permission to a user also imply to grant eventual specialized rights of the granted right. For example, providing the rights of video editing also implies the right of creating copy of the same video. This also means that has no sense to provide at the WFIPR Manager a tool to formalize all possible combination of right, users kind, content type, etc. It would be a heavy and un-useful work with the high risk of selecting an inconsistent set of

rights. To avoid this problem, the IPR Model and Wizard provide the automatic computation of permissions in real time according to the reference ontology. Moreover, not all the rights reported in the standards can be enforced and thus controlled into a web and mobile distribution. This fact has to be clearly shown to the IPR Manager during the creation of IPR Model. For example, on web browsers for PC it is almost impossible to protect the copy of an Image. The users may use a screen capture and or simply use the browser facilities to download the image. So that, for images the rights under control can be limited and specialized.

In order to solve the above mentioned problems, a reference **IPR Logical Model** has been formalized and adopted as reported in this paper. IPR Logical Model is based on the following main aspects:

- relationships among rights permissions (relationships of specialization and implication);
- relationships among user kinds (relationships of specialization);
- possibilities of the enforcement provided by the players adopted (web, mobile, applications, smart TV)

D.1 - IPR Logical Model: Relationships among right permissions

The relations among possible permissions are represented as different kind of relationships, for example:

- **Direct implication,** for example between ***Download-HD-PC → Play-HD-PC***: if a CP allows to a user to download a content, the CP implicitly allows also to play it (play via streaming and/or progressive download). This because, from a technical point of view: if someone downloads a content (without encryptions or protection) from the web he can play/view it on his PC whenever he wants.
- **Reciprocal implication,** for example between ***Download-LD/MD-PC ←→ Download-mobile-browser*** if a Content Partner allows an ECLAP user to download a content from PC, implicitly allows him also to download the same content from a mobile device. This because the users can download a content via a browser in their PC, then the user may transfer the content into its mobile device, so that the application of a restriction to avoid the download via mobile can be easily circumvented and thus it has no sense to be applied. It is also true the vice-versa.

According to the above model, the set of relationships among video permissions are formalized as follows and hold for any kind of users (where LD/MD/HD are for Low, Mid and High Definition):

- Download-HD-PC → Play-HD-PC
- Download-HD-PC → Download-LD/MD-PC
- Play-HD-PC → Play-LD/MD-PC
- Download-LD/MD-PC → Play-LD/MD-PC
- *Download-LD/MD-PC ←→ Download-mobile-browser*
- *Play-LD/MD-PC ←→ Play-mobile-browser*
- Download-mobile-browser→ Play-mobile-browser
- Download-mobile-browser→ Download-mobile-app
- Play-mobile-browser → Play-mobile-app
- Download-mobile-app → Play-mobile-app

Similar sets of relationships (with different rights set) have been identified and formalized for Audio, Documents (PDF, Slides, PPT, etc.), Images, crossmedia (MPEG-21, SMIL, HTML, ePub, etc.) and Animations. The permission of Embed implies the usage (access and play) of a provided content in a web page of third party. This technique implies that once provided the permission of content Embedding, the content may be accessed by public users on third party portals. Therefore, this permission can be formalized as:

(Video, Embed, u)→ (Video, Play-LD/MD-PC, Public user)

This implies that, granting the permission of embed for a given user, a number of implied permissions have also to be granted as described **in Figure 3** and according to the above presented logical rule of the IPR Logical Model.

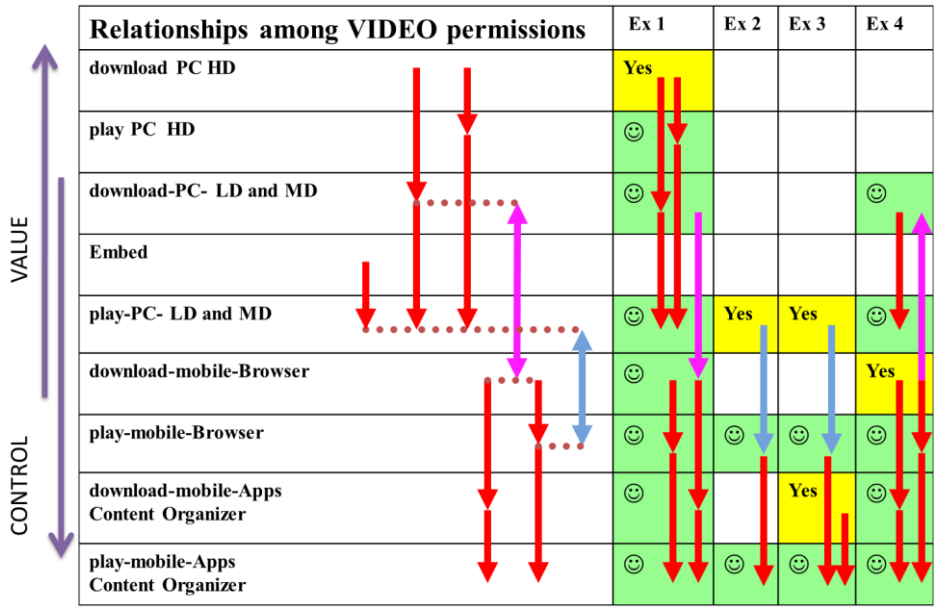

Figure 3 – Examples of how few clicks (Yes) can be used to define the IPR Model exploiting IPR relationships. In Example 1, a single Yes (grant of permission to download on PC in HD) has automatically granted all the other lower values permissions; Example 2: the grant of Play on PC at LD and MD implied also the play on mobile, but not the download of the same digital resource.

D.2 - IPR Logical Model: Relationships among users' kinds

The **IPR Logical Model** also includes a set of specialization relationships among users. In particular, any permission granted to a kind of user is also granted to less restrictive user categories. This approach is valid for all the permissions except for the permission of Embed. Therefore, the following rules is valid:

$\forall\, k \in ContentKind, \forall\, r \in possibleRights(k)\ and\ r \neq Embed\ |$

(k, r, TrustedUser)

(k, r, CPGroupUser) → *(k, r, EduResUser),*

(k, r, PublicUser) → *(k, r, CPGroupUser)*

The above formalized IPR Logical Model with all relationships has to be applied (i) during the definition of each IPR Model, Phases (A) and (B) (the license) to avoid the production of inconsistencies and incompleteness among the rights, (ii) at the play time by the conditional access module (Phase (C)), to verify which kind of rights can be granted according to context, condition and user kind (by the Cross Media Player CAS of Figure 1). These two aspects are described in the following two sections.

E. Exploiting IPR Logical Model during license model definition

In order to guarantee that all the produced IPR Models are consistent and complete and present the needed quality according to the Europeana model, the **IPR Wizard** tool has been realized. This tool is grounded on the IPR Logical Model that takes into account implications and relationships among user kind and permission, and the relationships among permissions and Europeana.rights modeling. This is a strong simplification with respect to the other solutions adopted by Content Aggregators on which they have to be very careful in selecting coherent rights and Europeana.rights values before content delivering.

According to the IPR Logical Model, only a few of the rights are set by the IPR Manager, all the others are automatically computed according to the IPR Logical Model. This approach allows to avoid incompleteness (for example, creating IPR Models with only those set by the IPR Manager and not including those that are logically and legally implied from the former and thus due to the final users as well). In the formalization of each IPR Model, the system save only the minimum set of rights imposed by the IPR Manager while the others are computed according

to the closure of the permission set associated with an IPR model considering the logical dependencies among permissions (see **Figure 4).** This activity is performed by both the IPR Wizard at the IPR Model editing time, and during the content access in the Cross Media Player CAS (Conditional Access System) (see **Figure 1**). This approach allows to strongly reduce the time and effort to produce IPR Models as it is shown later in the paper.

```
function permissionsClosure( P ⊂ Permission ) {
    // add to P the permissions allowing trusted users to do anything on all kinds of content
    P' = P ∪ { (k, r, TrustedUser, {}) | k ∈ ContentKind ∧ r ∈ possibleRights(k) };
    P_new = P';
    while (P_new ≠ ∅ ) {                    // repeat while there is some new permission to be processed
        P'' = ∅;                            // P'' is the set of new permissions implied from P_new
        for ( (k, r, u, c) ∈ P_new ) {      // for each new permission to be processed
            D = implied( k, r, u );         // compute the set of permissions directly implied from (k,r,u)
            for ( ( k', r', u' ) ∈ D ) {    // for each permission implied
                if  ( k', r', u', c ) ∉ P' ∪ P'' {   // check if it is a new permission and add it to P''
                    P'' = P'' ∪ { (k', r', u', c) };
                }
            }
        }
        P_new = P'';      // update P_new to the new permissions to be processed at the next round
        P' = P' ∪ P'';// add the new permissions to P'
    }
    return P';
}
```

Where, function *implied(k, r, u)* returns for a content kind *k*, a right *r* and user type *u*, the set of permissions that are logically directly implied, so for example:

implied(*Video, Play-LD/MD-PC, PublicUser*) = {
(Video, Play-LD/MD-PC, CPGroupUser),
(Video, Play-mobile-browser, PublicUser) }

Figure 4 – Algorithm for the computation of the permission closure during the IPR Model editing and production. It is activated at each change to the IPR Model.

Once the IPR Model is defined, it can be associated to a single content or collection manually from the interface of the ECLAP workflow or automatically. If the legal framework adopted by the community of CPs change, changing for example the relationships among rights, a change at level of IPR Logical Model is performed and the automated update to all content is applied.

F. IPR Logical Model: Checking algorithm during Conditional Access

At the time of content access, the IPR Logical Model has to be executed to computer which kind of rights can be granted or not according to context and conditions (content kind, device, date, locale, etc.) and user kind. In order to check if an user can do a certain action (play, download, access, etc.) on a content managed with a specific IPR Model the algorithm reported in **Figure 5** has to be used to compute the closure of the permission set associated with an IPR model considering the logical dependencies among permissions. To this end, function *checkIPR* is used to check if a user of kind *u* can exercise the right *r* on a content of kind *k* in a location *l* on a content that is managed by IPR Model *m.* This approach allows managing a limited number of IPR Models to manage a large number of content items and users. In the case of ECLAP, only about 67 models allowed to cope with thousands of users and hundreds thousands of content.

```
function checkIPR(k ∈ ContentKind, r ∈Right, u ∈UserKind, l ∈ Location, m ∈IPRModel) {
    P = iprPermissions(m, l); // retrieves from the model the permissions available for location  l
    // check if exists a permission in the closure of P matching k, r, u and satisfying all the connected conditions
    if  ∃ C. (k, r, u, C) ∈ permissionsClosure(P) ∧ ∀c∈ C. validCondition(c) {
        return true;
    }
    return false;
}
```

Figure 5 – Algorithm for the computation of the permission closure during the IPR Model application during content access.

G. Anticipating IPR Model Association

In most cases, IPR information is typically not formally associated with content since the beginning. Therefore, the metadata does not contain any reference to a finalized IPR Model. Metadata standards, such as DC, METS, MPEG-7, EDM, FRBRoo, etc., do not provide a formal model to represent IPR Models for content access licensing. In the best cases, metadata models provide some string field in which one could store some custom information and reference to eventual IPR Model. On the other hand, Europeana requests to fill mandatory EDM metadata Europeana.rights by using one of the predefined admitted values [EuropeanaRights].

Content and metadata are typically ingested into the content management infrastructure by following different paths. Metadata may be accessible by taking exports from databases and/or archives. Then, metadata are ingested into a metadata mapper for producing data that can be easily ingested in the ECLAP Back Office and storage. Content files/resources are provided via ftp services or uploaded from web, or provided as hard disk.

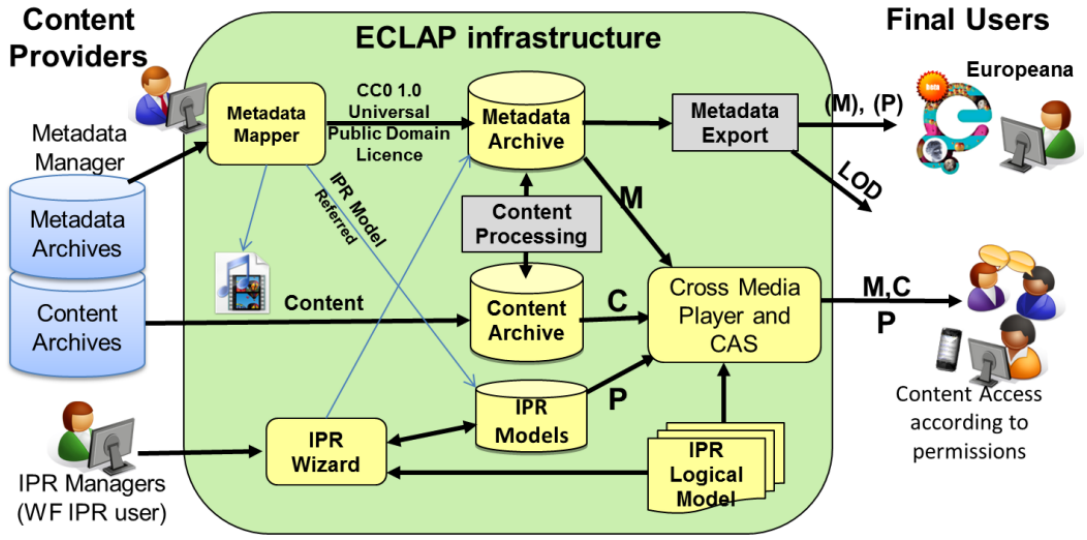

Figure 6 -- ECLAP IPR Work Flow, a simplified view, where M is for Metadata, C for Content and P for IPR Model and Permission**. Please note that Europeana receive only a part of the Metadata and of the IPR Model (that is the permission), for this reason we reported (M) and (P).**

According to **Figure 6**, the IPR Managers use the IPR Wizard to formalize the IPR Models. Once the IPR Model is defined, it can be associated with content and/or content collection (even if collection contains content of different kind). Thus, the association of the IPR Model to content is considered an integral part of the metadata set, while the IPR Model itself is stored in a separated database. This allows tuning the IPR Model to a large number of content elements limiting the activities on the metadata database.

Once the CP has already produced some IPR Models, the CP can associate one of the produced IPR Model directly during the ingestion phase. Thus avoiding the manual association of the IPR Model to content via the IPR Wizard and Content Management tool. When content is ingested without any reference to a specific IPR Model, a

default IPR Model is assigned providing *maximum restriction (no permissions to non trusted users).* The only granted permissions are those related to the managed of content at administrator level and by the trusted personnel of the CP owner. This avoids creating situations in which the content cannot be managed for the lacks of rights.

V. REPORT ON TOOLS USAGE AS ITS VALIDATION

In this section, results about the ECLAP Back Office activities performed on the content, metadata and IPR until April 2013 are reported (this period can be considered the validation phase of the IPR Model and tools). The ECLAP service allowed users and content management for enrichment processes to interoperate securely and to increase the quality and accessibility of content and metadata, avoiding conflicts and each other. The data reported have been obtained by the system usage by 31 international institutions. The number of state transitions and their distribution in the time period put in evidence the whole activity of the portal on content and metadata and allow analyzing singularly both the back-office and the user activities. Some results are reported in the temporal domain considering the "month" as a time period unit.

The validation have been performed by the real usage by 29 qualified users. Each user had single or multiple user roles (grant authorization) according to the IPR aspects. The user roles were distributed as: 24 enrichers (WFENRICHER), 6 validators (WFVALIDATOR), 23 IPR Manager users (WFIPR) and 9 publishers (WFPUBLISHER).

In the validation period, 1,036,406 state transitions have been handled for 173,562 content items with an average of 6 transitions per content and a maximum of 139 transitions for a single content. These transitions were performed in 652 days with an average of 1,589 transitions per day and a maximum of 15,786 transitions in one day, with a maximum of 14 different virtual nodes in the AXCP grid.

TABLE I. DISTRIBUTION OF STATE TRANSITIONS

From	To	Number of Transitions
'Uploaded'	'Under-AXCP'	259,086
'Under-AXCP'	'Uploaded'	259,086
'Proposed (creation)'	'Uploaded'	173,631
'Uploaded'	'Under-Approval'	170,609
'Under-Approval'	'Published'	170,317
'Uploaded'	'Under-IPR'	945
'Under-IPR'	'Uploaded'	945
'Uploaded'	'Under-Enrichment'	681
'Under-Enrichment'	'Uploaded'	681
'Under-Approval'	'Uploaded'	224
'Uploaded'	'Under-Validation'	43
'Under-Validation'	'Uploaded'	43
'Published'	'Uploaded'	7

A. Back-Office services

As previously described, the back-office services consist of a set of grid processes that run periodically automated workflow processes both on a single and on multiple contents:

1. Content and Metadata Ingestion. The number of content ingested and processed by the back-office has been 173,631corresponding to the UPLOADED workflow status of content.
2. Metadata Analysis Metadata analysis for assessment or automated translation performs a transition to the UNDER-AXCP status in order to lock the content and avoid that a user could be access to it for manual editing or validation. In total, 259,086 of these transitions were performed.
3. Metadata Validation Every time content passed the metadata analysis the back-office performs a transition to the UNDER-APPROVAL. In total, 170,609 of these transitions were performed.

4. Content Publication Every time the back-office performs the publication of content in the UNDER-APPROVAL workflow state it performs a new transition to the final state of PUBLISHED. In total, 170,317 of these transitions were performed.

B. Front-Office tools

In this section, the analysis of the activity performed by users via front-office IPR tools is reported. In the considered period, 11336 content elements have been manually uploaded by users via the upload service. Once uploaded the process has been passed to the back-office for indexing and semantic enrichment. Only the 0,2% of processes failed to conclude due to incompatible file formats. In order to evaluate the usage of Metadata Editor for the enrichment and validation activities, both the number of workflow transitions from UPLOADED to UNDER-ENRICH state and from UPLOADED to UNDER-VALIDATION status have been considered. The former transition gives a measure of manual enrichment activity, while the latter of the manual validation activity. The transitions related to enrichment were 681, and 43 for validation.

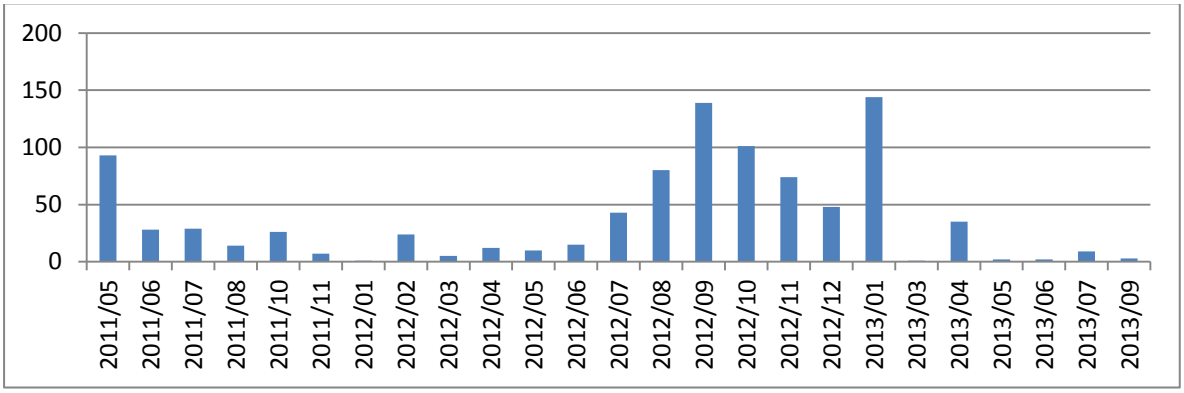

Figure 7 – Distribution of the IPR Wizard activities along the time period

C. IPR Model and Tools validation

The *IPR Wizard* has been largely used by more than 35 partners in Europe. To keep under control and evaluate the usage of IPR Wizard, the number of workflow transitions from UPLOADED to UNDER-IPR states over time have been tracked. The transitions were 945, and their distribution is reported in the **Figure 7**.

For the IPR aspects, 67 different IPR Models have been defined and used by the 35 partners, using the 23 IPR Managers. In **Table I,** the 945 changes to IPR have been performed to implement/update the IPR Models. For example, imposing initially restricted model and then releasing other rights as soon as the clearance of them have been possible by the Content Provider; thus, obtaining an average of 14 changes per IPR Model. Please note that, the solution proposed allowed to avoid the regeneration of licenses for the 3000 users of ECLAP since the IPR Model is executed for computing the rights at the access time, and thus a newly updated IPR Model is automatically applied at the next content access without relicensing.

Over the 67 IPR Models, 40 of them are restrictive not public models, while 27 are public models allowing the full content access. Most of CPs used only 1, 2 or 3 different IPR Models for covering their content collections, while a few partners used 4, 8 or 12 different IPR Models. **Figure 8** reports the number of content associated to IPR Model. The most commonly used has been applied to more than 40.000 content files. It is evident the huge advantage of coping with a unique model for unlimited number of content and unlimited number of users. This is a radical cost reduction with respect to DRM computational costs of licensing [Bellini, Nesi, Pazzaglia, 2013]. According to the analysis, it can be seen that the most diffuse IPR Models covered more than the ½ of the whole content collected on ECLAP. On the other hand, the semantic flexibility of the IPR Model proposed allowed to cope with the many different needs of the content owners in the cultural heritage domain that impose specific IPR Models according to legal rights they can really provide, and they identify during phase of right clearance (Section II, phase (A)).

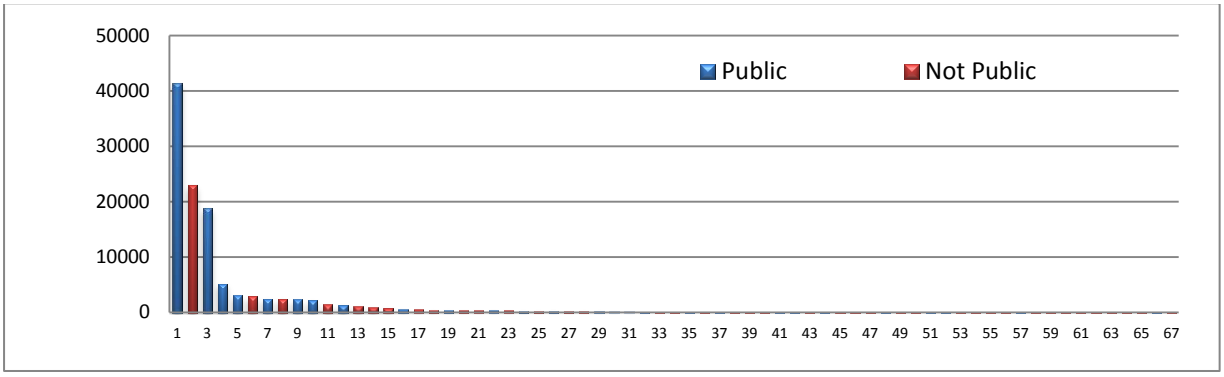

Figure 8 – Distribution of IPR Models (ID with respect to the number of content their manage).

The 68% of content is associated with a public IPR model. Regarding the 30 IPR Models with restrictive permissions, in 11 cases the IPR Model granted the access only to educational group users and in 6 cases only to group users (both educational and not educational), exploiting in this case the specialization among user kind. Moreover, 23 IPR Models have been used to grant play steaming without enabling the download of the digital resource for some kind of resource type (regardless of the user type).

TABLE II. IPR MODELS ALLOWING PERMISSIONS BY USER TYPE

Permission	***UserKind***		
	Public	***CPGroup***	***EduRes***
only play/access	11	13	19
download & play	3	8	11
no permission	19	12	4

Table II reports, for three user types (public, group and educational), how many IPR Models allow only play/access of the digital resource or allow the download and play of it or no permission are provided. It can be seen that in most cases the models are used to restrict access from the public users only (19 over 30) and to limit the download of the resource, leaving the permission of play/streaming. The produced IPR Models addressed all the content IPR and the 7% of them have been revised at least once without the need of regenerating licenses, and without the need to reprocessing content. This reduced the management costs with respect to the traditional DRM and CAS solutions that would be constrained to regenerate the licenses.

Moreover, in **Table III**, the evidence of the validation of the proposed approach in avoiding incompleteness and inconsistencies is reported (phase (B)). The whole set of IPR Models include 67 models. As explained above, and how it can be tested on the ECLAP service, the IPR Wizard allowed the IPR Managers to create and edit IPR Models. In the process, the IPR Managers can select some rights while the implied others are computed according to the above presented IPR Logical Model (see **Figure 4**). This approach allowed to set up 5452 rights only by using 570 ticks on the models, thus saving the 89% of time and highly avoiding incompleteness since the simple missing of a single derived right of the computer 4882 would create a non-complete IPR Model also legally inconsistent, and thus creating a large problem for thousands of content items. Moreover, this effect is also evident by the facts that, some IPR Managers selected rights, that later have been overwritten by stronger rights imposed. In whole set of ECLAP IPR Models, 66 **rights overwritten** by the IPR Logical Model have been found (1.16% of not useful ticks). They were not strictly useful for the formalization of the IPR Model since they have been also computed. In substance, these are "*mistakes" (unuseful ticks)* of the IPR Managers in setting the models. Similarly, the IPR Wizard dis-allowed the imposition of inconsistent rights, and automatically computed and completed the un-completed IPR Models in terms of rights and Europeana.rights values.

Some millions of content accesses have been registered in the validation time period without any king of copyright violation, thus demonstrating the validity of the IPR Logical Model and of the real time computation of the enforcement phase (C) of IPR access via procedure reported **in Figure 5.**

TABLE III. IPR LOGICAL MODEL EXPLOITATION AND ADVANTAGES OVER THE FULL SET OF IPR MODELS DEFINED, IN THE VALIDATION

Measure	Measure Value	% with respect to total rights
Number of IPR Models	67	
Total number of rights set by IPR Managers	570	10,4%
Total number of **rights overwritten** by the IPR Logical Model and thus not strictly useful for the formalization of the IPR Model since they can be computed (these rights are "*mistakes"* of the IPR Managers in setting the models).	66	
Total number of strictly useful rights set by IPR Managers	504	9,24%
Total number of rights computed into the IPR Models production	4882	89,54%
Total number of rights enforced by the full set of models	5452	100%

D. Applicability of the proposed solution

The proposed solution including the IPR Model and tools consists of a new approach for implementing conditional access systems by providing access to content according to user profile, device kind, access right, location, conditions, etc., without addressing the effort of producing the precise grant authorization for each specific case. Therefore, the approach is much different with respect to DRM and CAS solutions, and seems to be an evolution of the approaches used by YouTube, Vimeo, etc., to grant the access authorization according to some IPR defined by the content providers. The approach proposed in ECLAP is more focused on culture heritage and professional content providers as museums, archives, intuitions, etc. On the other hand, the approach proposed could be extended to be used for managing rights of content providers in social networks, in collaborative work portals, and for managing the access to data in smart city and big data applications (for example for regulating the access and exploitation of data related to users behavior only in given context according to a limited number of accepted rules proposed to the user). On this regard, an early example is the solution proposed for modeling content access filtering from the consumer point of view in addition to the IPR Model of the content provider [Li and Chang, 2014]. That proposal is grounded on the IPR Model presented in this paper.

VI. CONCLUSIONS

The paper discussed about requirements, design and validation of ECLAP which is an institutional aggregator for metadata and content coping with IPR models for providing metadata towards Europeana, the European international aggregator. The proposed solution takes into account issues connected to the cultural heritage cross media content and integrates front office tools and an automated back-office based on a grid. The solution allows coping with high quality and provides large scale multimedia services to manage huge amount of content and metadata, coming in turn from several national and local institutions: museum, archives, content providers.

The first contribution presented in this article corresponds to the formalization of IPR Model and solution that allow shortening the activities of IPR Model formalization avoiding the assignment of conflicting rights/permissions. The proposed IPR Model presents computational advantages (in the editing and access phases) with respect to classical licensing model of DRM and CAS. In addition, the flexibility of the IPR Model proposed allowed to cope with the many different needs of the content owners in the cultural heritage domain that impose specific IPR Models according to legal rights they can really provide, and they identify during the phase of right clearance. A second contribution consists in the identification of the Content Aggregator requirements, and thus in the definition and realization of a distributed architecture and solution satisfying them and at the same time addressing the IPR aspects. The aggregators have to cope with huge amount of content and metadata covered by many IPR restrictions, and performing their massive processing for semantic enrichment and publication. To this end, a specific model for content aggregation and enrichment management is needed.

Finally, the usage analysis puts in evidence the whole activities of ECLAP on content, metadata and IPR until July 2013. It underlines that the activity on content and metadata aggregation, analysis and validation to match the Europeana requirements has been mainly automated and performed by the back-office, thus allowing to keep

content processing cheap and sustainable. Regarding the front office side, the most used tools by content providers have been associated with IPR, namely IPR Wizard (that saved more than the 89% of time in producing IPR Models) and the Content Management since they allow users to finalize the rights and to provide a connection of the content versus Europeana. Most of the metadata provided were already in a good shape and less than the 1% of content has been corrected from that point of view. On the other hand, the IPR details requested by Europeana constrained the content provider to associate to the 100% of the content a new IPR Model. This huge effort has been kept under control by exploiting the IPR Model, and applying only 67 models to the whole set of more than 170.000 different content coming from more than 35 different collections and institutions.

ACKNOWLEDGMENT

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

REFERENCES

[Abelson et al., 2008] H. Abelson, et al. "ccREL: the Creative Commons Rights ExpressionLanguage," wiki.creativecommons.org/images/d/d6/Ccrel-1.0.pdf, 2008.

[Baltussen et al., 2012] Lotte Belice Baltussen, et al., ECLAP DE2.1.2, Revised User Requirements and Use Cases, http://www.eclap.eu/73166

[Basu et al., 2003] S. Basu et al., ''mmGrid: Distributed Resource Management Infrastructure for Multimedia Applications,'' Proc. 17th Int'l Parallel and Distributed Processing Symp., IEEE CS Press, 2003, p.88.

[Bellini and Nesi, 2013] Pierfrancesco Bellini, P. Nesi, "A Linked Open Data service for performing arts", Proc. of the ECLAP 2013 conference, 2nd International Conference on Information Technologies for Performing Arts, Media Access and Entertainment, Springer Verlag LNCS, 2013.

[Bellini and Nesi, 2014] P. Bellini, P. Nesi, "Modeling Performing Arts Metadata and Relationships in Content Service for Institutions", International Multimedia Systems Journal, Springer, presently online at http://link.springer.com/article/10.1007/s00530-014-0366-0, 2014

[Bellini et al., 2011] Bellini P., Bruno I., Cenni D., Nesi P., "*Micro grids for scalable media computing and intelligence on distributed scenarious*", IEEE Multimedia, 18 Aug. 2011, IEEE Computer Society Digital Library, ISSN: 1070-986X.

[Bellini et al., 2014] P. Bellini, P. Nesi, A. Venturi, "Linked Open Graph: browsing multiple SPARQL entry points to build your own LOD views", International Journal of Visual Language and Computing, Elsevier, 2014, DOI information: http://dx.doi.org/10.1016/j.jvlc.2014.10.003

[Bellini, Bruno, Nesi, 2011] P. Bellini, I. Bruno, P. Nesi, "EXPLOITING INTELLIGENT CONTENT VIA AXMEDIS/MPEG-21 FOR MODELLING AND DISTRIBUTING NEWS", International Journal of Software Engineering and Knowledge Engineering, World Scientific Publishing Company, 2011, Vol.21, n.1, pp.3-32.

[Bellini, Cenni and Nesi, 2012] P. Bellini, D. Cenni, P. Nesi, "On the Effectiveness and Optimization of Information Retrieval for Cross Media Content", Proc. of the KDIR 2012, Part of IC3K 2012, Int. Joint Conf. on Knowledge Discovery, Knowledge Engineering and Knowledge Management. Barcelona, Spain, 4-7 October 2012.

[Bellini, Nesi, Pazzaglia, 2013] P. Bellini, P. Nesi, F. Pazzaglia, "Exploiting P2P Scalability for Grant Authorization in Digital Rights Management Solutions", International Journal Multimedia Tools and Applications, Springer press, April 2013.

[CC] Creative Commons, http://creativecommons.org

[Chang et al., 2003] Chang, Shi Kuo; Polese, Giuseppe; Cibelli, Maurizio, "Visual Authorization Modeling with Applications to Electronic Commerce", IEEE Multimedia, Vol.10, n.1, pp.44-54, 2003.

[Congress] The Library of Congress, http://www.loc.gov

[Delgado et al., 2003] J. Delgado, et al. "IPROnto: An Ontology for Digital Rights Management," Proc. Conf. Legal Knowledge and Information Systems, 2003.

[Dolan and Pedersen, 2010] Liu J, Dolan P, Pedersen E, "Personalized news recommendation based on click behaviour". In: IUI '10: Proc. of the 14th intl. conf. on intelligent user interfaces, pp.31-40, 2010.

[EBellini and Nesi, 2013] Bellini E., Nesi P., "Metadata Quality assessment tool for Open Access Cultural Heritage institutional repositories", Proc. of the ECLAP 2013 conference, 2nd Int. Conf. on Information Technologies for Performing Arts, Media Access and Entertainment, Springer Verlag LNCS, 2013.

[ECLAP Partners] ECLAP Partners. List and information on ECLAP Partners available at: http://www.eclap.eu/3578

[ECLAP] European Library of Artistic Performance, ECLAP, http://www.eclap.eu/

[ECLAPSupportPage] ECLAP support page: http://www.eclap.eu/3609

[Europeana] Europeana, http://www.europeana.eu

[EuropeanaRights] 'Guidelines for the europeana:rights metadata element', v4.0 – 20, 9 February 2012, http://pro.europeana.eu/documents/900548/0d423921-23e0-45fa-82a4-2ac72b3e6f38

[Harmer et al., 2005] T.J. Harmer, J. McCabe, P. Donachy, R.H. Perrott, C. Chambers, S. Craig, R. Lewis, B. Mallon, L. Sluman, ''Gridcast: A Grid and Web Service Broadcast Infrastructure,'' Proc. IEEE Int'l Conf. Services Computing, Vol.1, pp.35-42, IEEE CS Press, 2005.

[Iannella and Guth, 2012] R. Iannella and S. Guth. ODRL version 2.0 common vocabulary. Specification, W3C ODRL Community Group, 04 2012. http://www.w3.org/community/odrl/two/vocab

[Iannella et al., 2012] R. Iannella, S. Guth, D. Pähler, and Andreas Kasten. ODRL version 2.0 core model. Specification, W3C ODRL Community Group, 04 2012. http://www.w3.org/community/odrl/two/model/

[Jung-Ran, 2009] Park Jung-Ran, Metadata quality in digital repositories: A survey of the current state of the art, "Cataloging & classification quarterly", vol.47, no.3-4 (April 2009), pp.213-228.

[Ku and Chi-Hung, 2004] W. Ku, and C. Chi-Hung, "Survey on the Technological Aspects of Digital Rights Management," Proc. Int'l Conf. Information Security, LNCS 3225, 2004.

[Li and Chang, 2014] Mao-Lin Li, Shi-Kuo Chang, "A Formal Model for Intellectual Relationships among Knowledge Workers and Knowledge Organizations", International Conference on Distributed Multimedia Systems 2014, pag.204-2010.

[Messina et al., 2011] Messina, Alberto; Montagnuolo, Maurizio; Di Massa, Riccardo; Borgotallo, Roberto; Hyper Media News: a fully automated platform for large scale analysis, production and distribution of multimodal news content, Multimedia Tools and Applications, Vol.63, n.2, pp.427-460, 2011.

[Moses 2005] T. Moses, "Privacy policy profile of XACML v2.0," Oasis standard, OASIS, 02 2005. http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-privacy_profile-spec-os.pdf

[MPEG21-REL] MPEG-21 Part 5: Rights Expression Language, ISO/IEC 21000-5:2004.

[Toma et al., 2006] Toma, Ioan; Burger, Tobias; Shafiq, Omair; Doegl, Daniel; Behrendt, Wernher; Fensel, Dieter, GRISINO: Combining Semantic Web Services, Intelligent Content Objects and Grid Computing, 2006 Second IEEE International Conference on e-Science and Grid Computing (e-Science'06), pp.39, 2006.

[VanDerAalst and Van Hee, 2002] W.M.P. van der Aalst and K.M. van Hee. Workflow Management: Models, Methods, and Systems. MIT Press, Cambridge, MA, USA, 2002.

[Volckaert et al., 2008] B. Volckaert, T. Wauters, M. De Leenheer, P. Thysebaert, F. De Turck, B. Dhoedt, P. Demeester, "Gridification of collaborative audiovisual organizations through the MediaGrid framework", Future Generation Computer Systems, vol. 24, pp.371–389, Elsevier, 2008.

[Wang et al., 2004] Xin Wang, Thomas DeMartini, Barney Wragg, M. Paramasivam and Chris Barlas, "The MPEG-21 rights expression language and rights data dictionary," IEEE Transactions on Multimedia, vol.7, no.3, pp.408–417, Jun. 2005.

[Wang, 2004] Wang X., "MPEG-21 rights expression language: Enabling interoperable digital rights management," IEEE Multimedia, vol.11, n.4, pp.84–87, 2004

[WIPO] WIPO, World Intellectual Property Organization, http://www.wipo.int/portal/en/.

[Yu and Buyya, 2005] Yu, J., & Buyya, R.. "A taxonomy of workflow management systems for grid computing". Journal of Grid Computing, 3(3-4), pp.171-200, 2005