EUROPEAN COLLECTED LIBRARY OF ARTISTIC PERFORMANCE

www.ECLAP.eu

Grant Agreement No 250481

DE4.3
Metadata Descriptors Interoperability

Version: 0.5
Date: 19/12/2012

Project Title: ECLAP Project Number: ICT-PSP-250481 Deliverable Number: DE4.3 Accessibility: public (PU) Work-Package contributing to the Deliverable: WP4, WP4.1) Nature of the Deliverable: report Status: final

Contractual Date of Delivery: 31/12/2012 Approve for quality control by: 29/12/2012 Finally approved by coordinator: 29/12/2012 Actual Date of Delivery: 29/12/2012

Document responsible: Natasa Sofou, Pierfrancesco Bellini Affiliation acronym: NTUA and DSI
Authors: • Natasa Sofou (NTUA) • Theofilos Malis (NTUA) • Nasos Drosopoulos (NTUA) • Pierfrancesco Bellini (DSI)

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Catalogue:

Title	Metadata Descriptors Interoperability
Identifier.de	DE4.3
Identifier.ISBN	
Creators	Natasa Sofou (NTUA), Theofilos Malis (NTUA), Nasos Drosopoulos (NTUA), Pierfrancesco Bellini (DSI)
Subject	Metadata Descriptors Interoperability
Description	Metadata Descriptors Interoperability
Keywords	Metadata Descriptors Interoperability
Publisher	ECLAP
Date	19/12/2012
Format	Document
Type	PDF or DOC
Language	EN

Citation Guidelines

Author(s) name Surname, Deliverable number, Deliverable title, ECLAP Project, DD/MM/YY, URL: univocally determined on http://bpnet.eclap.eu

ECLAP Copyright Notice

Please note that:

- You can become affiliated with ECLAP. This will give you access to a great amount of knowledge, information related to ECLAP services, content and tools. If you are interested please contact ECLAP coordinator Paolo Nesi at info@eclap.eu. Once affiliated with ECLAP you will have the possibility of using the ECLAP for your organisation.
- You can contribute to the improvement of ECLAP by sending your contribution to ECLAP coordinator Paolo Nesi at info@eclap.eu
- You can attend ECLAP meetings that are open to public, for additional information see www.eclap.eu or contact ECLAP coordinator Paolo Nesi at info@eclap.eu

Table of Contents

1 Executive Summary and Report Scope

From Sophocles, Terence, Shakespeare, Calderon and Molière to Pirandello, Beckett, Brecht and Dario Fo, the richness and value of the European performing arts heritage is unquestionable. Latest technological achievements have made performing arts multimedia content available to a broader audience with the only prerequisite its access to the Internet. While content is now being digitized and published online, access still remains fractured and scattered, for there is no coordination between the digital libraries and theatre domains. ECLAP will improve this status quo by building a network of theatrical institutions, archives and umbrella organizations and by selecting and making accessible via Europeana and ECLAP a substantial body of digitized content.

WP4 involves the selection and delivery of content and metadata for a wide range of user communities as well as the definition of the harvesting metadata schema and its semantic mappings to a spectrum of commonly used standards. Of crucial importance is to ensure semantic interoperability between the ECLAP system and the Europeana system. This is illustrated by the main objectives of WP4 which include:

- To support cataloguing, metadata and programme content with additional contextual information for a range of users to integrate with the Europeana initiative.
- To define metadata and descriptors coming from performing art institutions and suitable for posting on Europeana.
- To define interoperability map among several different models for metadata and descriptors for performing art content with respect to the semantic meaning of Europeana classification model.

This report describes how interoperability between ECLAP and Europeana is accomplished.

Before presenting the methodology in order to achieve interoperability between the two systems, it is crucial to have a wider perspective on the topic of metadata:

1. First of all we must clarify the term metadata, meaning data used to describe other data structured in formats easily understood by machines. Moreover the term “metadata framework” refers to the structural plan that ensures that metadata are formatted, structured, used, managed, and stored in an appropriate way.
2. In order to appropriately handle metadata there have been proposed many different standards that can be divided into: i) Descriptive data structure standards for different kinds of community resource descriptions. ii) Markup languages and schemas for encoding metadata in machine-readable syntaxes. iii) Ontologies for semantic mediation between data standards. iv) Protocols for distributed search and metadata harvesting.
3. Of special importance are the Europeana metadata standards that are designed to provide integrated access to digital objects from the cultural heritage organizations of all the nations of the European Union. Europeana has proposed two data structure standards: Europeana semantic elements specification (ESE) and Europeana data model (EDM).

In order to build the ECLAP metadata schema and the ECLAP ingestion services a survey was performed to collect information about 'individual' collections that a content provider is prepared to submit to ECLAP. Based on the results and the analysis of the provided questionnaires and samples the most commonly used standards among content providers are Dublin Core, MARC, EAD, CDWA and a simplified customization of FRBR. Unfortunately the majority of content providers did not use any standard for the metadata which makes it difficult to ingest this cultural data into ECLAP and Europeana.

The ECLAP metadata schema was structured based on the EDM standard and taking into account the previously mentioned metadata standards. This approach ensures that ECLAP and Europeana have a similar basic structure and common elements while allowing different components to vary in depth and details. The interoperability and exchange of metadata between ECLAP and Europeana is further facilitated by creating a metadata crosswalk between the two schemas. A crosswalk is a mapping of the elements, semantics, and syntax from one metadata schema to those of another. The successfulness of the crosswalk is ensured by the similarity of the two schemes, the granularity of the elements in the Europeana scheme compared to that of the ECLAP, and the compatibility of the content rules used to fill the elements of each scheme. In order to create the crosswalk between initial metadata standard (if any) to ECLAP schema and then to EDM the ECLAP Metadata Ingestion Service portal was employed. The metadata mapping service allows the user to define a mapping between the source) and target schema. An XSLT is then generated, based on this mapping that can convert all existing items initially to ECLAP format and then to EDM.

Subsequently, data are delivered to the Europeana database via the OAI-PMH protocol. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low barrier mechanism for repository interoperability. Data providers are repositories that expose structured metadata via OAI-PMH. Service providers then make OAI-PMH service requests to harvest that metadata. OAI-PMH is a set of six verbs or services that are invoked within HTTP. In the context of this deliverable, ECLAP is the data provider while Europeana is the service provider.

2 Introduction

From Sophocles, Terence, Shakespeare, Calderon, Moliere and Pirandello, to Beckett, Brecht and Dario Fo, the richness and value of the European performing arts heritage is unquestionable. Latest technological achievements have made performing arts multimedia content available to a broader audience with the only prerequisite its access to the Internet. While content is now being digitized and published online, access still remains fractured and scattered, since there is no coordination between the digital libraries and theatre domains. ECLAP will better this status quo by building a network of theatrical institutions, archives and umbrella organizations and by selecting and making accessible via Europeana and ECLAP a substantial body of digitized content.

WP4 (Content Provision and Augmentation) involves the selection and delivery of content and metadata for a wide range of user communities as well as the definition of the harvesting metadata schema and its semantic mappings to a spectrum of commonly used standards. This deliverable reports on the results of Task 2 (Metadata/descriptors interoperability maps) of WP4. Its main purpose is to provide all the necessary mappings between ECLAP and Europeana. In addition, part of this task is to create the mappings between the descriptive and multimedia metadata. This task will ensure the semantic interoperability inside the ECLAP system and more importantly with the Europeana system. The choice of the FRBR (Functional Requirements for Bibliographic Records) standard will facilitate the semantic interoperability. Europeana will adopt a rich standard similar to FRBR in order to cover the requirements of all Cultural Organizations. This will enable the definition of rich mappings between the two schemas. This report focuses on the interoperability of metadata standards and classifications, between ECLAP and Europeana. It also reports on the lessons learnt in the first period and the changes done to adapt the model to real cases.

This report includes the identified metadata and their corresponding semantics in relationship with those of Europeana, taking into account all kind of annotations coming from the several partners' providers. Specifically in Section 2 an introduction is given to the scope of the WP and its role in the project. Sections 3, 4, 5 provide background knowledge. Section 3 presents the basic concepts related to metadata and terminology. Section 4 provides an outline of the most important metadata standards, proposed by different institutions and projects, for the documentation of digital cultural heritage, while Section 5 gives an overview of the Europeana standards, i.e. the Europeana Data Model (EDM) and the Europeana Semantic Elements Specification (ESE). The following section (Section 6) describes the results of the survey on providers' metadata, carried out by the Work Package (WP) 4 leader that set the requirements for the ECLAP schema. Section 7 demonstrates ECLAP crosswalks. Specifically in this section a brief description of the ECLAP schema, which is designed in order to describe performing arts content, is provided The full eclap schema is available as Appendic in section 11. Additionally, the mapping of the elements required for the documentation of performing arts content, as those decided by the providers, to the Europeana Data Model is presented illustrating its expressive power. Furthermore, there is an overview of the procedures and tools that are deployed within ECLAP, in order to establish interoperability between ECLAP's metadata and the Europeana repository. And finally there is also an overview of the changes made to the ECLAP schema to adapt the model in real cases with emphasis on the lessons learned during the first period of the project. In the remaining sections are the conclusions, the bibliography, the glossary and the appendix containing the complete ECLAP Schema.

3 Understanding Metadata and Terminology

3.1 Knowledge representation and metadata

Knowledge Representation is a two sided concept. Knowledge on cultural heritage objects is represented in metadata schemas (mainly in the semantic description of a cultural heritage object, not in the technical or administrative part of a metadata schema). Knowledge on cultural heritage object is also represented in 'controlled vocabularies' or 'knowledge organization systems' of all kinds, therewith controlling the content of several metadata elements or attributes of a metadata schema.

Metadata; many definitions have been provided for the term metadata, e.g. "a cloud of collateral information around a data object" as defined by Clifford Lynch (director of the Coalition for Networked Information). Metadata (Greek: meta- + Latin: data "information") are defined literally as "data about data" or "information about information", but the term is normally understood to mean structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource. A resource may be anything that has identity and a resource may be digital or non-digital. Operations might include, for example, disclosure and discovery, resource management (including right management) and the long-term preservation of resource. For a single resource different metadata may be required to support these different functions. A metadata record is a file of information, compiled (automatically and/or manually) in the format of the metadata schema concerned, which captures the basic characteristics of a data or information resource (e.g. a cultural heritage object). In other words, metadata refers information that describes information sources or objects, e.g. a Dublin Core record or a record from the catalogue of an archive.

The term metadata is used differently in different communities. Some use it to refer to machine understandable information, while others use it only for records that describe electronic resources. In the

library environment, metadata is commonly used for any formal scheme of resource description, applying to any type of object, digital or non-digital. Traditional library cataloging is a form of metadata; MARC 21 and the rule sets used with it, such as AACR2, are metadata standards. Other metadata schemes have been developed to describe various types of textual and non-textual objects including published books, electronic documents, archival finding aids, art objects, educational and training materials, and scientific datasets.

Metadata is sometimes classified according to the functions it is intended to support. In practice, individual metadata schemas often support multiple functions and overlap the categories:

1. **Descriptive** metadata is mainly information to identify and describe the object or information source and what it expresses. These metadata include the author/title cataloguing as well as the subject indexing. In other words, the descriptive metadata include the subgroup of the objective elements that formally describe the object (e.g. identification number, title, creation date, creator name, the language of the object, physical media). And the subgroup of semantic elements (also called analytical metadata) that contain information on the subject of the object to enhance access to the resources contents (e.g. subject keywords, classification codes, and abstract). Note, that the descriptive metadata and especially the semantic elements are the scope of DE4.1. Note also: descriptive metadata can be of a technical character, think of for instance ‘compression Schema’ (this is the algorithm used to compress the audiovisual essence), the number of pages (book), black and white / color (photograph, film) or specific information on the storage medium or carrier.
2. **Structural** metadata describes the logical or physical relationships between the parts of a compound object. For example a physical book consists of sequences of pages to form a chapter.
3. **Technical** metadata describe the technological characteristics of the related object (e.g. data that must be available to be able to use out the material, file locations, authentication and security information, characteristics needed for computer programming and database management)
4. **Administrative** metadata provides information for managing and administering the objects concerned (e.g. content provider name, acquisition information, copyrights, location information, language of record, and record number). There are several subsets of administrative metadata; two that sometimes are listed as separate metadata types are:
    - **Rights management** metadata**,** which deals with intellectual property rights and
    - **Preservation** metadata, which contains information needed to archive and preserve a resource(as it was published in 1988 by Working Group on Preservation Issued of Metadata constituted by the Research Libraries Group -RLG)

3.2 Metadata Framework

A metadata framework can be viewed as having five key components:

1. A *schema* (the categories of information you choose to record)
2. *Vocabulary* (specific ‘words' or ‘values' you enter into those categories)
3. *Conceptual model* - the underlying model that describes how all the information and concepts inherent in a resource are related to one another
4. *Content standard* - practical standards that describe how specific information (e.g. vocabularies) should be entered within metadata schema categories (e.g. Cataloguing Cultural Objects)

5. *Encoding* - which is concerned with the way the metadata is presented (e.g. XML)

Based on the above structure of a "metadata framework", in the rest of this section we attempt to provide some definitions and descriptions of the basic components of a metadata framework along with the description of other key terms related to this framework.

Metadata schema refers to the format and structure of metadata that is often dictated in a set of rules, called metadata schema. It can be defined as:

- A full, logically organized structure of relations between defined (groups) of metadata and the information objects they describe. [1]
- A set of rules for encoding information that supports specific communities of users. A metadata schema consists of several metadata elements. For some elements the input is free (e.g. Title), for other elements the input is guided by syntactical rules or guidelines or even restricted by controlled vocabularies of all kinds (e.g. thesaurus for subject keywords or closed term list for object type).

Metadata element is an item, or an editorial part of metadata. A semantic metadata element is an element from the descriptive metadata that describes the cultural heritage object. A metadata element name is given to a data element in, for example, a data dictionary or metadata schema or registry. In a formal data dictionary, there is often a requirement that no two data elements may have the same name, to allow the data element name to become an identifier, though some data dictionaries may provide ways to qualify the name in some way, for example by the application system or other context in which it occurs. A data element definition is a human readable phrase or sentence associated with a data element within a data dictionary that describes the meaning or semantics of a data element.

Controlled Vocabulary; A limited set of terms that must be used to index | represent | tag the subject matter | content of documents | objects (indexing tools in use to describe a cultural heritage object). Examples: Alphabetic lists of "approved" words or phrases, thesauri, subject heading systems, classification schemes, ontologies, taxonomies. These examples illustrate that controlled vocabularies are largely applied for subject keywords or generic concept identification. However, controlled vocabularies or lists of preferred terms are also applied for other metadata elements, e.g. person names like author or creator, names of historical people and corporate bodies on the cultural heritage object or as its subject of the cultural heritage object, geographic places (actual location of the cultural heritage object / place of creation / place where the cultural heritage object was found / place as subject of the cultural heritage object) and organization names. *See also: Authority files in this section.*

Classification schemes, Taxonomies and Categorization schemes; these terms are often used interchangeably. Although there may be subtle differences from example to example, in general these types of knowledge representation provide ways to separate entities into buckets or relatively broad topic levels. Some examples provide a hierarchical arrangement of numeric or alphabetic notation to represent broad topics. These types of knowledge representation may not follow the strict rules for hierarchy required in the ANSI NISO Thesaurus Standard (Z39.19) (NISO), and they lack the explicit relationships presented in a thesaurus. Examples of classification schemes include the Library of Congress Classification Schedules (an open, expandable system), the Dewey Decimal Classification (a closed system of 10 numeric sections with decimal extensions), and the Universal Decimal Classification (based on Dewey but extended to include facets). Subject categories are often used to group thesaurus terms in broad topic sets, outside the

hierarchical scheme of the thesaurus. Taxonomies are increasingly being used in object oriented design and knowledge management systems to indicate any grouping of objects based on a particular characteristic. "Taxonomy" may also refer to a scheme that presents subject elements in a hierarchical arrangement based on some characteristic. For the definitions of the several types of controlled vocabularies the following sources is used: [2], [3].

Thesauri are knowledge organization systems based on concepts, and they show relationships between terms. Relationships commonly expressed in a thesaurus include hierarchy, equivalence, and associative (or related). These relationships are generally represented by the notation BT (broader term), NT (narrower term), SY (synonym), and RT (associative or related). There are standards for the development of monolingual thesauri (NISO, 1998; ISO, 1986) and multi-lingual thesauri (ISO, 1985). It should be noted that the definition of a thesaurus in these standards is often at variance with schemes that are actually called thesauri. There are many thesauri that do not follow all the rules of the standard, but are still generally thought of as thesauri. Many thesauri are very large (more than 50,000 terms). Most were developed for a specific discipline, or to support a specific product or family of products.

Subject headings; this scheme provides a set of controlled terms to represent the subjects of items in a collection. Subject heading lists can be extensive, covering a broad range of subjects. However, the subject heading lists structure is generally very shallow, with a limited hierarchical structure. In use, subject headings tend to be pre-coordinated, with rules for how subject headings can be joined to provide more specific concepts. Examples include the Medical Subject Headings (MeSH) and the Library of Congress Subject Headings (LCSH).

Authority files are lists of terms that are used to control the variant names for an entity or the domain value for a particular field. Examples include names for countries, individuals, and organizations. Non-preferred terms may be linked to the preferred versions. This type of knowledge organization generally does not include a deep organization or complex structure. The presentation may be alphabetical or organized by a shallow classification scheme. There may be some limited hierarchy applied in order to allow for simple navigation, particularly when the authority file is being accessed manually or is extremely large. Specific examples of authority files include the Library of Congress Name Authority File and the Getty Geographic Authority File.

Semantic Network; with the advent of natural language processing, there have been significant developments in the area of semantic networks. These knowledge organization systems structure concepts and terms not as hierarchies but as a network or a Web. Concepts are thought of as nodes with various relationships branching out from them. The relationships generally go beyond the standard BT, NT and RT. They may include specific whole-part relationships, cause-effect, parent-child, etc. One of the most noted semantic network is Princeton's WordNet, which is now used in a variety of search engines.

An **Ontology** is a data model that represents the existing knowledge within a domain and is used to reason about the objects in that domain and the relations between them. Ontologies are used as a form of knowledge representation about the world or some part of it. Ontologies (as defined in www.wikipedia.org) generally describe:

- Individuals (the basic or "ground level" objects); Classes (sets, collections, or types of objects);
- Attributes (properties, features, characteristics, or parameters that objects can have and share);
- Relations (ways that objects can be related to one another)

Therefore thesauri and classification schemes can be regarded as ontologies with a relatively little number of relationships.

Ontologies can represent complex relationships between objects, and include the rules and axioms missing from semantic networks. Ontologies that describe knowledge in a specific area are often connected with systems for data mining and knowledge management.

Upper Ontology (top-level ontology, or foundation ontology): an ontology that describes very general concepts, applicable across all domains. The aim is to have a large number of ontologies accessible under this upper ontology.

Markup ontology languages; these languages use a markup scheme to encode knowledge, most commonly XML (SHOE, XOL, DAML+OIL, OIL, RDF, RDF Schema, OWL)

The **Semantic Web** provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. The Semantic Web intent is to enhance the usability and usefulness of the Web and its interconnected resources. A Semantic Web-compatible markup guarantees a rich use (mainly in retrieval functionality) of the metadata on cultural heritage objects in combination with several ontologies related to the cultural heritage domain. A domain ontology (or domain-specific ontology) models a specific domain, or part of the world. An ontology on arts can be used to say, for instance that “Picasso” is a “Painter”, and that a “Painter” is an “Artist”. The combination of such ontologies together with indexes automatically provides the end user with several extra ways to navigation through the collection. E.g. this combination can present all cultural heritage objects from museums in Spain, without the need for the content providing partners to manually add extra metadata to the descriptions of their objects.

An **XML schema** is a description of a type of XML document, typically expressed in terms of constraints on the structure and content of documents of that type, above and beyond the basic syntax constraints imposed by XML itself. An XML schema provides a view of the document type at a relatively high level of abstraction. There are languages developed specifically to express XML schemas. The Document Type Definition (DTD) language, which is native to the XML specification, is a schema language.

A **Data Model** is a model that describes in an abstract way how data are represented in a business organization, an information system or a database management system. This term is ambiguously defined to mean

- How data generally are organized, e.g. as described in Database management system. This is sometimes also called "database model" or,
- How data of a specific business function are organized logically (e.g. the data model of some business).

While simple data models consisting of few tables or objects can be created "manually", large applications need a more systematic approach. Within the relational database modeling community, the entity-relationship model method is used to establish a domain-specific data model. In computer science, an entity-relationship model (ERM) is a model providing a high-level description of a conceptual data model. Data modeling provides a graphical notation for representing such data models in the form of entity-relationship diagrams (ERD). A conceptual schema, or high-level data model or conceptual data model, is a map of concepts and their relationships, for example, a conceptual schema for a karate studio would include

abstractions such as student, belt, grading and tournament." A data model, especially the concepts or entities and relationships of the model, dictate the metadata elements that are needed in the metadata schema that goes along with the data model.

Semantic Interoperability: The term refers to the ability of computer systems to transmit data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems.

Semantic interoperability is therefore concerned not just with the packaging of data (syntax), but the simultaneous transmission of the meaning with the data (semantics). This is accomplished by adding data about the data (metadata), linking each data element to a controlled, shared vocabulary. The meaning of the data is transmitted with the data itself, in one self-describing "information package" that is independent of any information system. It is this shared vocabulary, and its associated links to an ontology, which provides the foundation and capability of machine interpretation, inferencing, and logic.

Syntatic interoperability refers to the packaging and transmission mechanisms for data. Syntatic interoperability is a prerequisite for semantic interoperability.

Metadata Crosswalks: The interoperability and exchange of metadata is further facilitated by metadata crosswalks. A crosswalk is a mapping of the elements, semantics, and syntax from one metadata schema to those of another. A crosswalk allows metadata created by one community to be used by another group that employs a different metadata standard. The degree to which these crosswalks are successful at the individual record level depends on the similarity of the two schemes, the granularity of the elements in the target scheme compared to that of the source, and the compatibility of the content rules used to fill the elements of each scheme. Crosswalks are important for virtual collections where resources are drawn from a variety of sources and are expected to act as a whole, perhaps with a single search engine applied. While these crosswalks are key, they are also labor intensive to develop and maintain. The mapping of schemes with fewer elements (less granularity) to those with more elements (more granularity) is problematic.

4 Standards Landscape

As explained earlier, metadata are data used to describe other data structured in formats easily understood by machines. One of the most familiar ways to organize metadata is through ontologies. Metadata standards are ontologies that define the vocabulary that describes the concepts and relations among them in the specified domain of interest. Metadata schema refers to the format and structure of metadata that is often dictated in a set of rules. Many different metadata schemas are being developed in a variety of user environments and disciplines.

It should be noted here that “schemas" is used in a broad sense, to describe a set of categories (i.e. "elements" or "units") of information used to describe resource. Metadata schemas can be differentiated in many different ways, for example:

- Their size and scope (e.g. comprehensive or 'core'; emphasis on description, administration, preservation; concern with single items or collections or both)
- Things they describe (e.g. art images, audio, video, objects, books, places)
- Communities they serve (e.g. libraries, museums, educators)

Furthermore distinctions between schemas, conceptual models, content standards, and encoding standards are often not fixed or discreet. Several metadata schemas describe their underlying conceptual models, provide guidance on what data might to be entered within their categories, or indicate how the metadata should be encoded. Dublin Core, for example, provides all of these.

Since the ECLAP project deals with cultural heritage content, this section lists some of the most important metadata standards and schemas used within the cultural heritage domain. It should be noted that this document does not attempt to categorise the schemas and standards available. Instead it provides a list of all the necessary standards and technology components to facilitate intracommunity knowledge sharing most related to the ECLAP project:

4.1 Descriptive Data Structure Standards

The technology of digital archives offers great potential for inter-institutional collaboration: not only can multiple collections be rendered cross-searchable in the style of a union catalogue, but the objects that constitute these collections can themselves readily be integrated into inter-institutional virtual repositories. To do so effectively, however, requires standard approaches to metadata. Without these, problems rapidly arise when digital library collections reach any substantial size. A useful classification for digital metadata indicates the range of information that must be included:

- Descriptive metadata: Analogous to the tradition catalogue record, descriptive metadata contain information on the item's intellectual contents which allows it to be retrieved and its value to the user assessed.
- Administrative metadata: The information necessary to curate the digital item, which includes technical, rights management and digital provenance metadata.
- Structural metadata: Information necessary to record the internal structure of an item so that it can be rendered to the user in a sensible form. This type of metadata is necessary as an item may often be comprised of multiple of files - for example, the images of individual pages that make up a digitized book.

Some of the most commonly adopted descriptive metadata structure standards include:

- MPEG Multimedia Metadata: The ISO/IEC Moving Picture Experts Group (MPEG) has developed a suite of standards for coded representation of digital audio and video. MPEG-7 [4] defines the metadata elements, structure, and relationships that are used to describe audiovisual objects including still pictures, graphics, 3D models, music, audio, speech, video, or multimedia collections. MPEG-21 [4] was developed to address the need for an overarching framework to ensure interoperability of digital multimedia objects.
- Dublin Core: Dublin Core [6] is a standard for cross-domain information resource description. It provides a simple and standardised set of conventions for describing things online in a machine understandable way making them easier to find. Dublin Core is a metadata standard used mainly to describe content of multimedia essence, such as video, sound, image, text and composite media.
- CDWA: CDWA (Categories for the Description of Works of Art) is a framework for describing and accessing information of cultural heritage resources [7]. It provides access to art databases for describing and collecting information about works of art, architecture, other material culture, groups and collections of works, and related images.

- LIDO: LIDO (Lightweight Information Describing Objects) was developed with the purpose of contributing content to cultural heritage repositories [8]. LIDO satisfies the need for a convenient common instrument for providing core data from different collections, data structures or software systems.
- SPECTRUM: The SPECTRUM documentation standard (Standard ProcEdures for CollecTions Recording Used in Museums) [9] was created as a guide to good practice for museum documentation, established in partnership with the museum community. It contains procedures for documenting objects and the processes they undergo, as well as identifying and describing the information which needs to be recorded to support the procedures.
- IMS: The IMS (Instructional Management Systems) Learning Resource Metadata Information Model [10] identifies a subset of IEEE LOM metadata elements to be used to describe learning materials in various types of learning systems. It provides open technical specifications for interoperable learning technology and standards for delivering learning products and services.
- AMICO: The AMICO (Art Museum Image Consortium) metadata vocabulary [11] is mainly used in the collection of art museum images. The AMICO metadata vocabulary using the DC and CDWA vocabularies provides a framework for the specification of images and multimedia files.
- MARC: MARC (Machine-Readable Cataloguing) [12], [13] is a family of metadata standards for representing library resources. Although chiefly used by libraries to describe bibliographic material as well as non book material such as images or archival collections. MARC is a very extensive and formalised standard, with hundreds of potential categories and a rigid way of encoding its data.
- MODS: MODS (Metadata Object Description Schema) [14] was developed in 2002 as a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. As an XML schema it is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records.
- METS: METS (Metadata Encoding and Transmission Standard) [15] is a standard for encoding metadata within an XML format. Although it contains descriptive and administrative elements of its own, a key function of the METS standard is to structure or "package" other metadata or data for exchange or delivery.
- EAD: The EAD (Encoded Archival Description) [16] Document Type Definition (DTD) is a data structure standard for encoding archival finding aids. It defines the structural elements and designates the content of descriptive guides to archival and manuscript holdings following the syntax of the Standard Generalized Markup Language (SGML – ISO8879).

4.2 Markup Languages and Ontologies

Markup languages and ontology languages are adopted for encoding metadata in machine-readable syntaxes. A markup language is a modern system for annotating a document in a way that is syntactically distinguishable from the text. Each markup language, in order to be human and machine readable, has its own syntax. Syntax is concerned with the structure of the language and refers to the logical or grammatical form of sentences, rather than what they refer to or mean. Ontology languages are formal languages used to construct ontologies. Ontologies formally represent knowledge as a set of concepts within a domain, and the

relationships between pairs of concepts. Apart from their syntax ontology languages also have specific semantics that are concerned with the meaning of words and sentences.

Most popular markup and ontology languages include:

- XML: XML (Extensible Markup Language) [17] is a set of rules for encoding documents in machine-readable form. It is defined in the XML 1.0 Specification produced by the W3C, and several other related specifications, all gratis open standards. XML's design goals emphasize simplicity, generality, and usability over the Internet.
- RDF: RDF (Resource Description Framework) [18] is a general-purpose language for representing information in the web. RDF's main elements are resources, properties and property values. A resource represents an object in an ontology which is connected through a property to some value which is either a literal or another resource.
- RDFS: RDFS (RDF Schema) [19] is an extension of RDF that is more expressible by allowing classes, as well as class and property subsumption. It provides mechanisms for describing groups of related resources and the relationships between these resources as well as other characteristic of resources, such as domain and range.
- OWL: OWL [20] is a Web Ontology Language with rich vocabulary for describing properties and classes. It allows to express relations between classes (e.g. disjointness), cardinality, equality, richer typing of properties and characteristics of properties (e.g. symmetry), and enumerated classes. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

4.3 Ontologies for Semantic Mediation between Data Standards

Ontologies have been used for semantic mediation between data standards. These ontologies provide definitions and a structure for describing implicit and explicit concepts and relationships used in various domains. More specifically:

- SKOS: SKOS (Simple Knowledge Organization System) [21] is an RDF vocabulary for representing semi-formal knowledge organization systems (KOSs), such as thesauri, taxonomies, classification schemes and subject heading lists.
- CRM: CRM (CIDOC Conceptual Reference Model) [22] is a formal ontology that provides definitions and a structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.
- FRBR: FRBR [23] is a conceptual model for describing information resources within a library context.
- FRBRoo: FRBRoo [24] is a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information.

4.4 Distributed Search and Metadata Harvesting Protocols

These protocols are used for searching, retrieving, sorting and browsing information from remote computer databases. They include:

- Z39.50: Z39.50 [25] is a client–server protocol for searching and retrieving information from remote computer databases. It is widely used in library environments and is often incorporated into

integrated library systems and personal bibliographic reference software. It supports a number of actions, including search, retrieval, sort, and browse.

- OAI-PMH: OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) [26] provides an application-independent interoperability framework for metadata harvesting. It is used to harvest (or collect) the metadata descriptions of the records in an archive so that services can be built using metadata from many archives.

5 Metadata Modelling in Europeana Standards

5.1 ESE - Europeana Semantic Elements Specification

Europeana provides integrated access to digital objects from the cultural heritage organisations of all the nations of the European Union. It encompasses material from museums, libraries, archives and audio-visual archives with the aim of making Europe's multicultural and multilingual riches discoverable together in a common on-line environment. To do this Europeana harvests and indexes the descriptive metadata associated with the digital objects. As there is no one universal metadata standard applied across the participating domains, a set of metadata elements has been developed that will allow a common set of information to be supplied to support the functionality desired by the user and needed for the operation of the underlying system. The Europeana Semantic Elements V3.3 (ESE) is an updated version of the metadata set used in the Europeana prototype in November 2008. It has been amended to include additional elements for the Rhine release of the portal in July 2010. It is a Dublin Core-based application profile providing a generic set of terms that can be applied to heterogeneous materials thereby providing a baseline to allow contributors to take advantage of their existing rich descriptions.

To provide metadata in the ESE format, it is necessary for contributors to map elements from their own metadata format to ESE. In addition to the mapping it is necessary for a normalisation process to be carried out on some values to enable machine readability. In the initial implementation of the Europeana prototype much of the mapping and normalisation was carried out centrally in the Europeana Office. This work is increasingly being passed to data providers or aggregators. An XML Schema has also been produced as a further tool to assist providers in ensuring compliance with ESE. ESE v3.3 is a sub-set of the metadata initially defined in the Europeana Metadata Requirements described in the EDLnet deliverable D2.5 "Europeana Outline Functional Specification".

The ESE v3.3 XML Schema (http://www.europeana.eu/schemas/ese/ESE-V3.3.xsd) is the XML representation of the Europeana Semantic Elements (ESE) specifications v3.3 (http://version1.europeana.eu/web/guest/technical-requirements/). This schema can be used to validate XML instances of Data Sets to be submitted to Europeana. The ESE v3.3 XML Schema extends the DC XML Schema with the addition of elements belonging to the Europeana namespace. The Europeana Semantic Elements (the ESE), consist of the 15 original Dublin Core (DC) metadata elements, a subset of the DC terms and a set of thirteen elements which were created to meet Europeana's needs. The ingestion process currently ignores the xml:lang attribute although it is present in data from some providers. It is anticipated that functionality will soon be in place to take advantage of these attributes in the display of metadata values, in particular where they are provided in one or more languages. Providers are encouraged to include them in all appropriate metadata elements.

5.2 EDM – Europeana Data Model

The Europeana Data Model (EDM) is a new proposal, still under development, aimed at being an integration medium for collecting, connecting and enriching the descriptions provided by Europeana content provider [27]. The purpose of the open structure of EDM is to enable the linking of data, placing it in the vanguard of semantic web developments.

Outline of EDM: The initial development of Europeana was based on Europeana Semantic Elements (ESE) data model which is evolved into EDM. Particularly, ESE was developed in order to constitute the lowest common denominator of the different data standards used for each one of the heritage sectors. EDM reverses this reductive approach and attempts to transcend the respective information perspectives of the sectors that are represented in Europeana.

In addition, EDM has upgraded ESE with respect to its content. In terms of a digitized book, the individual chapters, illustrations and index can be understood both individually and collectively. The same holds for an archival finding aid or fonds with respect to the constituent letters, deeds, manuscripts or other items. Finally, in contrary to ESE, EDM supports the preservation of original data while still allowing interoperability.

The strength of EDM lies on the fact that its development is not based on a specific standard but rather adopts an open, cross-domain Semantic Web based framework. It can accommodate several rich standards like LIDO for museums, EAD for archives or METS for digital libraries.

Apart from its ability to support standards of high richness, it also enables data enrichment from a range of third party sources. In this way, a particular digital object from a specific provider can be enriched by metadata from another provider and at the same time by additional data held from a third party. EDM enables this interoperability while clearly providing the provenance of all the data linking to the digital object.

One of the crucial purposes of EDM is to answer the basic queries “Who?”, “What?”, “When?” and “Where?” for every digital object and to make connections between the networks that will animate Europeana’s content.

Construction Principles: EDM complies with the modelling principles that underpin the approach of the Semantic Web. Therefore, there is no fixed schema that dictates a particular way to represent the data. Instead, the common model of EDM functions as an anchor to which various finer-grained models can be attached. In this way, they become partly interoperable at the semantic level, while the data retain their original expressivity and richness.

One of the main features of EDM is that via the digital representations submitted to Europeana it enables the representation and accessing of the provided objects. It is also able to ingest the descriptive metadata supplied by various providers and at the same time to represent new information added by Europeana. In addition to this, not only it accommodates various description paradigms of the ingested objects, but also enables further enrichment of the objects by connecting the to semantically enriched resources. At the same time, it still allows for different levels of granularity in the descriptions by taking advantage of special features of semantic mapping.

The requirements and principles that EDM follows according to Europeana [28] are:

- Distinct the provided object (book, painting, sculpture), which is the focus of the users’ interest, from its digital representations which are the elements manipulated by information systems like Europeana

- Distinct the provided object from the metadata record describing the object
- Allow for multiple records for the same object, even if they contain contradictory statements with each other
- Support objects that are composed of other objects
- Standard metadata format that can be specialized
- Standard vocabulary format that can be specialized
- Should be based on existing standards

Conceptually, four are the main concepts used in EDM and these are: ore:Aggregation, ore:Proxy, ore: EuropeanaAggregation and ens:WebResource. Following the Object Reuse and Exchange (ORE) model, EDM considers that the provided object, along with its digital representations contributed by any provider, form an aggregation that is represented as the ore: Aggregation class. Each instance of ore: Aggregation relates through the property ore:aggregates to one resource that represents the provided object and through the property ens:hasView to one or more resources (ens:WebResource) that are digital representations of the object. Each provider contributes a different set of digital representations and a new aggregation connected to the web resources.

Inspired again by ORE model, EDM leverages the proxy mechanism to enable the representation of different views on the same resource. Each provider contributes a separate metadata record using the ore: Proxy resource, in order to represent the description of the provided object as seen from the perspective of the specific provider. A proxy is related to the resource using the ore: proxyFor property and to the provider's aggregation through the ore: proxyIn property.

Finally, Europeana creates its own aggregation, the ens:EuropeanaAggreagation, and proxy in order to be able to add new information to the original object description and representation while keeping a clear distinction from the contributed information.

6 Metadata Standards Used Among Content Providers - Results Obtained from ECLAP Survey

The aim of the ECLAP survey is to collect information about 'individual' collections that a content provider is prepared to submit to ECLAP. The collected information will be used by technical partners to develop the IPR Wizard and the metadata and content collection system.

The ECLAP survey was divided in two main sections. The first one was to collect general information related to content partners. The second section's aim was to collect additional information related to collections provided by Content Partners to ECLAP from a technical point of view. Each content partner had to fill in a technical survey form (in the form of Excel file) providing information for each content set. The ECLAP General survey had 2 sections:

- Content Provider Information
- Metadata and terminology

Each partner has to answer the same number of question in general questionnaire part. The technical questionnaire can be found in Appendix A.

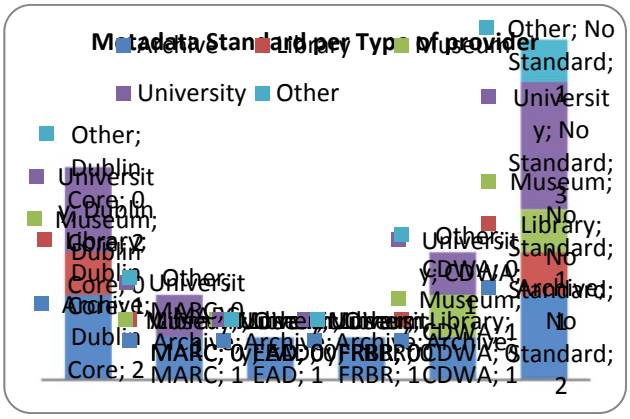
In the technical questionnaire each partner had to provide technical information regarding content type, number of items, IPR, metadata etc. A table providing a summarized view of results with emphasis to metadata is available in section 6.3.

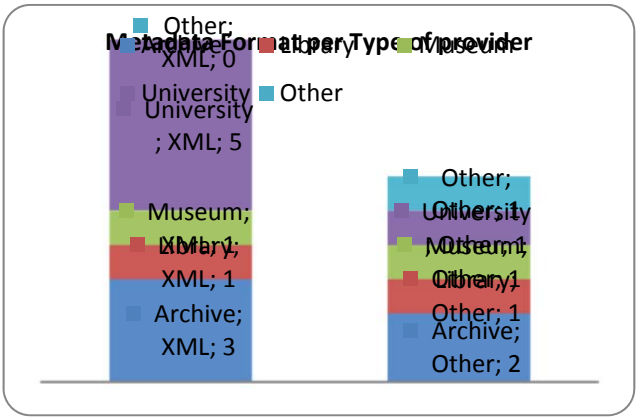
Along with the completed questionnaire each partner had to provide a sample file of metadata in xml. These files would provide a more complete view of how each content provider organizes the available metadata and serve as test data for ECLAP mapping tool.

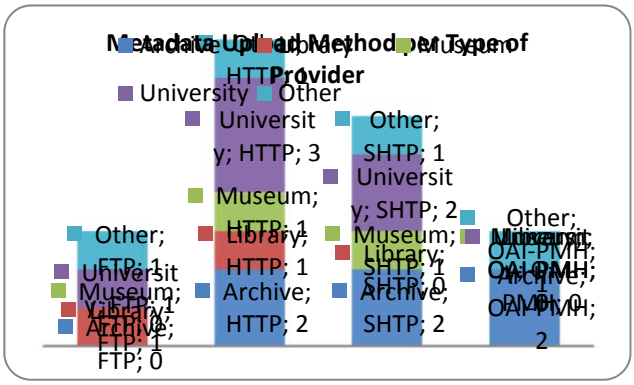
6.1 Information Schemes (Metadata)

6.1.1 Metadata Types

Charts below show how many provider organisations use which information schemes (metadata). Note that where a scheme does not appear in answers to the survey it is not included in the chart. Additionally, where there are multiple collections from the same organisation, using the same schemes, they only contribute once to the chart. Organizations with more than one type (e.g. library, museum, archive, etc) appear with more than one entry in the chart.







6.1.2 Metadata Conclusions

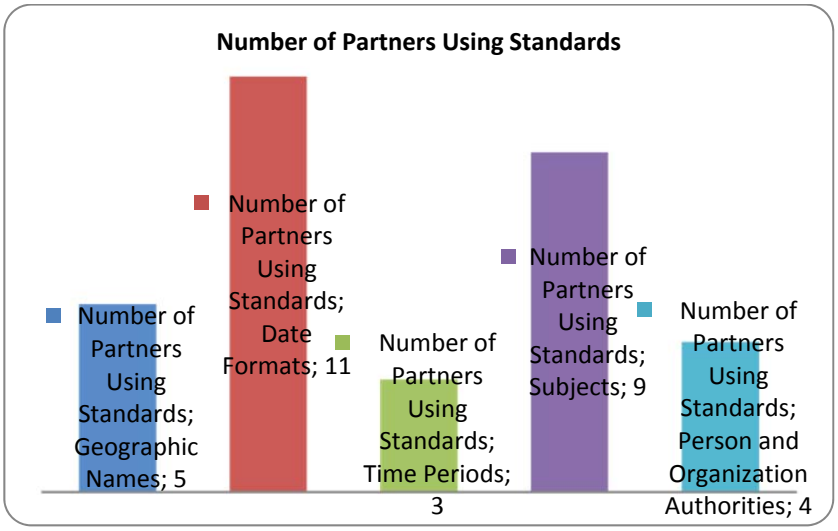
Based on the results and the analysis of the provided questionnaires and samples the most commonly used standards among content providers are Dublin Core, MARC, EAD, CDWA and a simplified customization of FRBR. The majority of content providers do not use any standard for the metadata. The first conclusion reached so far is that Dublin Core is a popular metadata scheme. This scheme has been fashionable over the

last few years for public access to cultural material. The challenge for a Dublin Core-based system is whether it can support the rich nature of performing arts data as exemplified by the cultural domain standards. Secondly, a significant number of providers use no standard for their metadata schemes, on inhouse customized schemes. This is difficulty for automated ingestion of cultural data by ECLAP as well as Europeana.

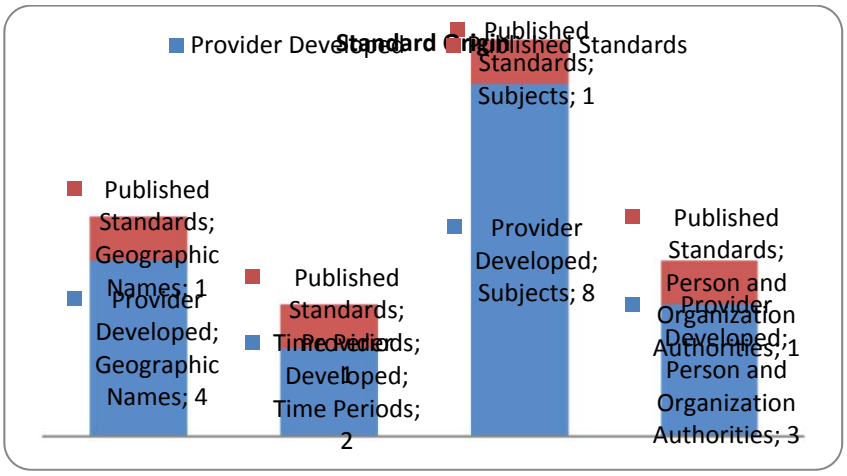
6.2 Metadata Terminology

6.2.1 Overview

Regarding metadata terminology the following chart summarizes the number of partners that use standards for different areas of terminology.

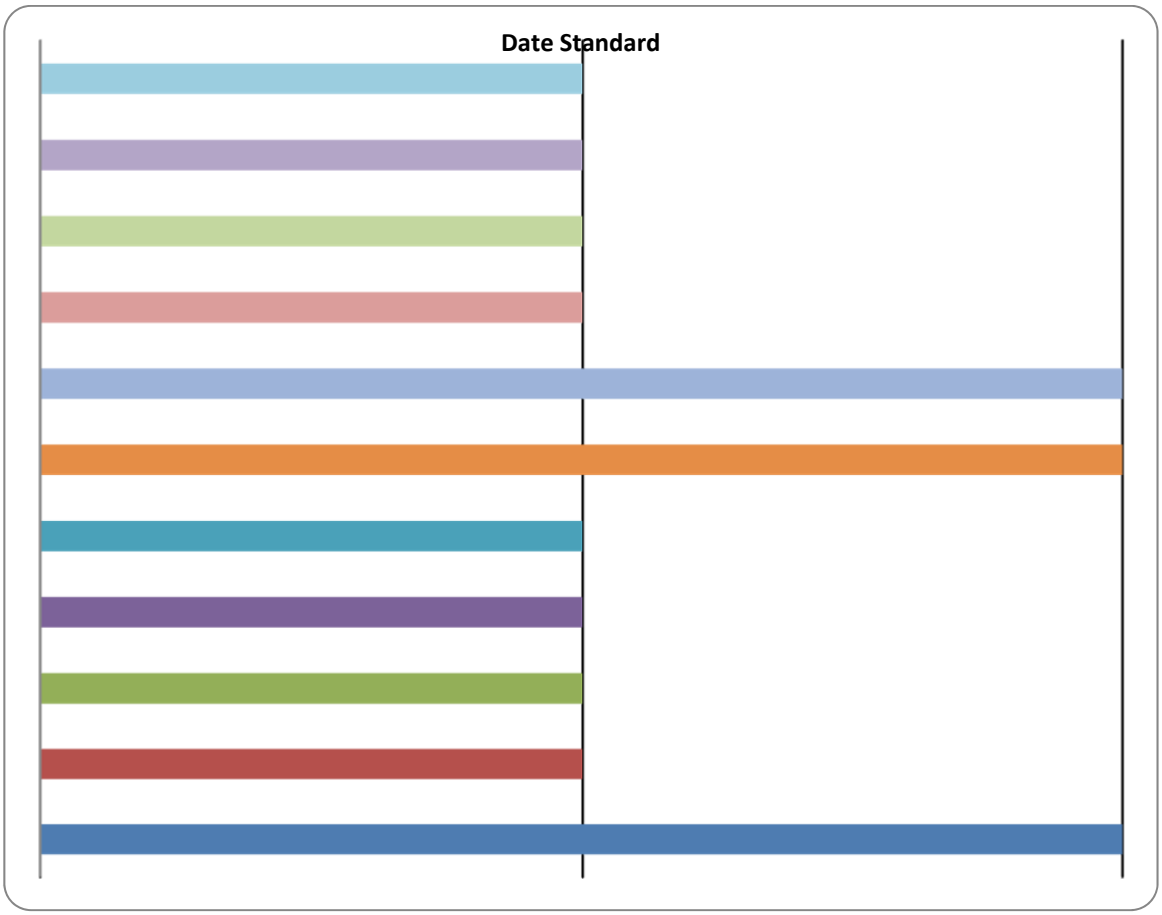


Regarding standards for metadata terminology, some of them are published standards whereas other are developed by the provider, as illustrated in the following chart:



6.2.2 Date Format Standards

Regarding date format standards, 11 out of 14, partners, that is 78,5%, answered that they used a standard for date format of their items. Among the date format standards they use there is almost no overlap. The date format standards used by content providers can be found in the following chart.



6.2.3 Summarized results of technical questionnaire

A table of summarized results for each partner with respect to metadata obtained by the technical questionnaire can be found in the following table.

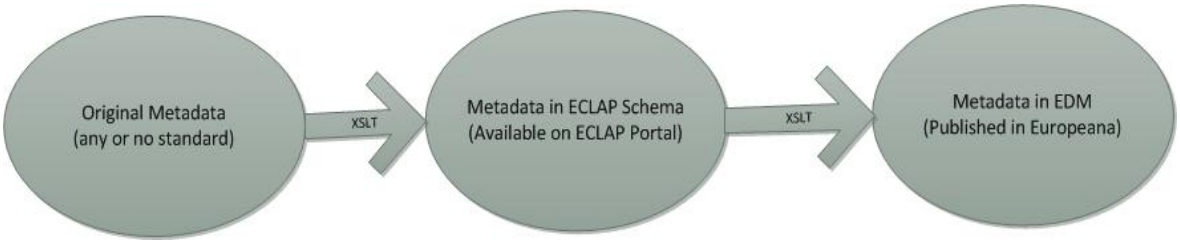
	CONTENT TYPE	METADATA UPLOAD METHOD	METADATA STANDARD	METADATA FORMAT	METADATA LANGUAGE	METADATA SAMPLE
UNIROMA	video	HTTP	no standard	xml	Italian	✓
CTFR	image	HTTP	Dublin Core	xml	Italian	
		SFTP	EAD			
	text	OAI-PMH	MARC			
B&G	video	OAI-PMH	based on FRBR but customized	xml	Dutch	✓
	image					
ITB	image	FTP	Dublin Core	xml	Catalan	✓
UvA	video	HTTP	Dublin Core	xml	Dutch or English	✓
ESMAE	video	HTTP	MARC	xml	Portuguese	✓
	audio					
	text		no standard			
	image					
UCLM						
FIFF	video	no existing metadata				
	audio					
	text					
	image					
OSZMI	video	HTTP	no standard	other formats	Hungarian	✓
	audio					

	text					
	image					
Bellone	video	FTP	no standard	other formats	French	
	audio					
	text					
	image					
UCAM	image	SFTP	CDWA	xml	Dutch	
	video					
MUZEUM	missing					
IKP	missing					
UG	Video	FTP, SFTP	Dublin Core	xml	English	✓
	Audio					
	Text					
	Image		no standard	other formats		
	Animation					
	Html					

7 Metadata Crosswalks in ECLAP (DSI)

This section has been produced by DSI. The interoperability and exchange of metadata is facilitated by metadata crosswalks. As defined in previous section, a crosswalk is a mapping of the elements, semantics, and syntax from one metadata schema to those of another. Within ECLAP, a *crosswalk* can be considered as follows:

Content providers use their preferable standard for their metadata representation. When content is contributed to ECLAP project, the corresponding metadata need to be transformed into a new form under the predefined ECLAP schema. The available metadata (regardless the standard they conform to) are in xml format. In order to transform their metadata, content providers use the ECLAP Metadata Ingestion Service portal, and map their metadata to the ECLAP schema by defining an XSLT that is used in the mapping phase. Although, all metadata are uniformly stored in ECLAP under ECLAP schema, in order to be available through Europeana portal they first need to be transformed from ECLAP schema to EDM and then published to Europeana. Thus a new mapping takes places, transforming metadata from ECLAP schema to EDM. For more details on the aforementioned hierarchy and process the reader can refer to *D3.1-Infrastructure: ingestion and processing content and metadata* and *D4.2.1 Content and Metadata, Selection Aggregation and Augmentation.*



In the following sections we provide a detailed description of the ECLAP schema, the mapping from ECLAP schema to EDM, the delivery to the Europeana office as well as the changes applied to ECLAP schema in order to a adapt the model in real cases .

7.1 ECLAP Mapping Schema

In this section a brief description of the current ECLAP metadata schema is reported. The complete description is reported in appendix.

The metadata schema is divided in the following parts:

- ***General information about the content***, it reports the internal identifiers like the axoid, the drupal nid, the provider name and id and the url to the content on ECLAP portal
- ***Information about the digital resource*****,** it reports information on the resource format (audio, video, image, document, 3d, crossmedia, etc.), in case of video content if it is available in MD and HD resolutions, if the resource is available on smartphones, the url to the preview gif (animated for video), etc.
- ***Information on IPR***, on intellectual property rights for the digital resource, it contains the Europeana rights url, if the content is public or private, the title and description of the IPR model associated with the content.
- ***GeoSpatial information about the resource***, reports the GPS position associated with the content as well as the extent of the area surrounding the position.
- ***Dublin Core metadata (DC and DCTERMS)***, reports Dublin Core metadata describing the content.
- ***Performing Arts metadata***, reports the metadata specific for performing arts:
    - *FirstPerformance Date, Location, City and Country*, indicates where and when the premiere of the piece depicted in the digital resource was performed.
    - *Performance Date, Location, City and Country****,*** indicates where and when the performance depicted in the digital resource was performed.
    - *PerformingArtType*, for the type of performing art (e.g. theatre, dance, etc.)
    - *PlotSummary*, summary of the plot
    - *PerformingArtsGroup*, name of the theatre or dance company or musical group (e.g. Momix)
    - *Cast*, name/names of cast member
    - *PerformersAndCrew*, name/names of performers and crews of the performance
    - *Professional*, people involved in the performance indicating which role each person had in the performance (eg. Actor, director, set designer etc.).
    - *Object*, object used in the performance
    - *PieceRecord,* script of the play
    - *Genre*
    - *HistoricalPeriod*
    - *ArtisticMovementAndActingStyle,* Artistic movement and acting styles in which the work can be categorized (e.g. Classicism, Dada, Epic, Expressionism, etc.)
    - *RecordingDate*, date when the recording was made
- ***Taxonomy based classifications***, reports information about the taxonomy terms associated with the content, for each term is reported the label in every language, the term id, the id of the top term for the hierarchy and the path from the term to the top term.
- **ECLAP Groups**, reports the ECLAP Groups to which the content is associated with.
- ***ECLAP Aggregations (Collections & Playlists)***, for collections and playlists reports the identifiers (axoids) of the content in the collection/playlist, for playlists items it is also eventually present the startTime/endTime/duration indications.

7.2 Mapping from ECLAP Schema to EDM

In this section is reported how the ECLAP metadata is mapped to EDM using an Object centric perspective (the only one that now europeana ingestion supports), it have to be noted that in the material to be provided to europeana in many cases it does not represent strictly an Object (like a book, a painting, a sculpture, …) while often it represents an event happened in the past, the performance.

In the following is reported how the ECLAP metadata are mapped to EDM elements. The Dublin core elements (dc and dcterms) are mapped directly to the ProvidedCHO elements while the PerformingArts metadata are mapped to DublinCore elements where possible, also the taxonomy associations are mapped to DublinCore depending on the top hierarchy element (Subject is mapped to dc:subject, PerformingArtType to dc:type, HistoricalPeriod to dcterm:temporal, etc.).
For each ECLAP content is provided: one *ProvidedCHO* element, one *WebResource* element representing the ECLAP portal page showing the content and one *Aggregation* element aggregating the two preceding ones. The @ sign indicate an attribute of the element.

edm:ProvidedCHO

@rdf:about	*axoid*
dcterms:*	All ECLAP dcterms fields
dcterms:issued	"<PerfArts.FirstPerformance.Date> (first performance)"
dcterms:issued	PerfArts.Performance.Date
dcterms:spatial	PerfArts.Performance.Place
dcterms:spatial	PerfArts.Performance.City
dcterms:spatial	PerfArts.Performance.Country
dcterms:temporal	PerfArts.HistoricalPeriod
dcterms:temporal	
@rdf:resource	"http://www.eclap.eu/Classification/HistoricalPeriod/<*histPeriodId*>"
dcterms:references	PerfArts.PieceRecord
dc:*	All ECLAP dc fields
dc:date	PerfArts.RecordingDate
dc:creator	PerfArts.PerformingArtsGroup
dc:contributor	"<*PerfArts.Professional*> (<*role*>)"
dc:contributor	PerfArts.Cast
dc:contributor	PerfArts.PerformersAndCrew
dc:contributor	PerfArts.PersonRecord
dc:contributor	PerfArts.ProductionRecord
dc:description	PerfArts.PlotSummary
dc:description	PerfArts.Object
dc:subject	
@rdf:resource	"http://www.eclap.eu/Classification/Subject/<*subjId*>"
dc:subject	PerfArts.Genre
dc:subject	
@rdf:resource	"http://www.eclap.eu/Classification/Genre/<*genreId*>"
dc:type	PerfArts.PerformingArtsType
dc:type	
@rdf:resource	"http://www.eclap.eu/Classification/PerformingArtType/<*paTypeId*>"
dc:type	PerfArts.ArtisticMovementAndActingStyle
dc:type	
@rdf:resource	"http://www.eclap.eu/Classification/ArtisticMovementAndActingStyle/<*amasId*>"
edm:type	based on Resource.Format (video, audio, image, document)

The *histPerdiodId*, *subjId*, *genreId*, *paTypeId*, *amasId* are the ids of the terms in the ECLAP taxonomy to which the content is associated with. The SKOS taxonomy defining the concepts used are provided to europeana using a specific file.

edm:WebResource

@rdf:about	"http://www.eclap.eu/europeana/<*axoid*>"
edm:rights	IPR.EuropeanaRightsUrl

ore:Aggregation

@rdf:about	"<*axoid*>:aggregation"
edm:aggregatedCHO	*axoid*
edm:dataProvider	eclap:ProviderName
edm:provider	"ECLAP, e-library of Performing Arts"
edm:rights	IPR.EuropeanaRightsUrl
edm:isShownAt	"http://www.eclap.eu/europeana/<axoid>"
edm:object	eclap:Preview

This mapping should be enhanced by enriching the metadata with associations with Places, TimeSpans, Agents thus integrating the text metadata with an association with an rdf resource coming from linked open data initiatives or well known authority files as VIAF for person names, GeoNames for places etc.

The following is an example of mapping the metadata of an Image from the Dario Fo and Franca Rame Archive.

The source metadata is:

```
<eclap:Content axoid="urn:axmedis:00000:obj:36b2407e-0ca0-4f44-892b-ebf254118a2d">
  <eclap:url>http://www.eclap.eu/drupal?q=home&amp;axoid=urn:axmedis:00000:obj:36b2407e-0ca0-4f44-892b-ebf254118a2d</eclap:url>
  <eclap:nid>35923</eclap:nid>
  <eclap:Version>2</eclap:Version>
  <eclap:InsertUpdateTime>2011-08-06T00:36:26</eclap:InsertUpdateTime>
  <eclap:ProviderId>CTFR</eclap:ProviderId>
  <eclap:ProviderName>Dario Fo &amp;amp; Franca Rame Archive</eclap:ProviderName>
  <eclap:DefaultMetadataLanguage>it</eclap:DefaultMetadataLanguage>
  <eclap:Resource>
    <eclap:Format>image</eclap:Format>
    <eclap:Type>image</eclap:Type>
    <eclap:Width>597</eclap:Width>
    <eclap:Height>800</eclap:Height>
    <eclap:Extension>.jpg</eclap:Extension>
  </eclap:Resource>
  <eclap:Platforms>
    <eclap:AvlForPDA>yes</eclap:AvlForPDA>
    <eclap:AvlForIPhone>yes</eclap:AvlForIPhone>
    <eclap:AvlForPC>yes</eclap:AvlForPC>
  </eclap:Platforms>
  <eclap:IPR>
    <eclap:IsPublic>yes</eclap:IsPublic>
    <eclap:IPRTitle>CTFR IPR</eclap:IPRTitle>
    <eclap:IPRDescription>CTFR model</eclap:IPRDescription>
    <eclap:EuropeanaRightsUrl>http://www.europeana.eu/rights/rr-f/</eclap:EuropeanaRightsUrl>
    <eclap:LicenseUrl>http://bpnet.eclap.eu/drupal/?q=node/2862</eclap:LicenseUrl>
  </eclap:IPR>
  <eclap:Preview>http://www.eclap.eu/gif/urn_axmedis_00000_obj_36b2407e-0ca0-4f44-892b-ebf254118a2d.gif</eclap:Preview>
  <eclap:DublinCoreMetadata>
    <dc:description xml:lang="it">Tavole a colori di Dario Fo.</dc:description>
    <dcterms:extent xml:lang="it">1 pagine</dcterms:extent>
    <dc:rights xml:lang="it">Archivio Dario Fo e Franca Rame, alcuni diritti riservati</dc:rights>
    <dc:title xml:lang="it">Sant'Ambrogio - 2005 Testo teatrale di Dario Fo. 32/116</dc:title>
    <dc:type xml:lang="it">Disegni</dc:type>
    <dc:date xml:lang="it">2005</dc:date>
    <dc:language xml:lang="it">it</dc:language>
  </eclap:DublinCoreMetadata>
  <eclap:PerformingArtsMetadata xml:lang="it">
    <eclap:Performance>
      <eclap:Country>Italia</eclap:Country>
    </eclap:Performance>
  </eclap:PerformingArtsMetadata>
  <eclap:Classification>
    <eclap:term id="504" vid="5" root="664" path="664">
      <eclap:label xml:lang="it">Teatro</eclap:label>
      <eclap:label xml:lang="en">Theatre</eclap:label>
      <eclap:label xml:lang="da">Teater</eclap:label>
      <eclap:label xml:lang="nl">Theater</eclap:label>
      <eclap:label xml:lang="fr">Théâtre</eclap:label>
      <eclap:label xml:lang="de">Theater</eclap:label>
      <eclap:label xml:lang="el">Θέατρο</eclap:label>
      <eclap:label xml:lang="hu">Színház</eclap:label>
      <eclap:label xml:lang="pl">Teatr</eclap:label>
      <eclap:label xml:lang="pt">Teatro</eclap:label>
      <eclap:label xml:lang="es">Teatro</eclap:label>
      <eclap:label xml:lang="ca">Teatre</eclap:label>
      <eclap:label xml:lang="sl">Gledališče</eclap:label>
    </eclap:term>
  </eclap:Classification>
  <eclap:Group id="2862">
    <eclap:label xml:lang="en">Dario Fo &amp;amp; Franca Rame Archive</eclap:label>
```

```
    </eclap:Group>
</eclap:Content>
```

That is mapped to EDM as:

```
<rdf:RDF …>
   <edm:ProvidedCHO rdf:about="urn:axmedis:00000:obj:36b2407e-0ca0-4f44-892b-ebf254118a2d">
      <dcterms:extent xml:lang="it">1 pagine</dcterms:extent>
      <dcterms:spatial>Italia</dcterms:spatial>
      <dc:description xml:lang="it">Tavole a colori di Dario Fo.</dc:description>
      <dc:rights xml:lang="it">Archivio Dario Fo e Franca Rame, alcuni diritti riservati</dc:rights>
      <dc:title xml:lang="it">Sant'Ambrogio - 2005 Testo teatrale di Dario Fo. 32/116</dc:title>
      <dc:type xml:lang="it">Disegni</dc:type>
      <dc:date xml:lang="it">2005</dc:date>
      <dc:language xml:lang="it">it</dc:language>
      <dc:type rdf:resource="http://www.eclap.eu/Classification/PerformingArtsType/504"/>
      <edm:type>IMAGE</edm:type>
   </edm:ProvidedCHO>

   <edm:WebResource rdf:about="http://www.eclap.eu/europeana/urn:axmedis:00000:obj:36b2407e-0ca0-4f44-892b-ebf254118a2d">
      <edm:rights>http://www.europeana.eu/rights/rr-f/</edm:rights>
   </edm:WebResource>

   <ore:Aggregation rdf:about="urn:axmedis:00000:obj:36b2407e-0ca0-4f44-892b-ebf254118a2d:aggregation">
      <edm:aggregatedCHO rdf:resource="urn:axmedis:00000:obj:36b2407e-0ca0-4f44-892b-ebf254118a2d"/>
      <edm:dataProvider>Dario Fo &amp;amp; Franca Rame Archive</edm:dataProvider>
      <edm:provider>ECLAP, e-library for Performing Arts</edm:provider>
      <edm:rights>http://www.europeana.eu/rights/rr-f/</edm:rights>
      <edm:isShownAt rdf:resource="http://www.eclap.eu/euroepana/ urn:axmedis:00000:obj:36b2407e-0ca0-4f44-892b-ebf254118a2d"/>
     <edm:object rdf:resource="http://www.eclap.eu/gif/urn_axmedis_00000_obj_36b2407e-0ca0-4f44-892b-ebf254118a2d.gif"/>
   </ore:Aggregation>
   <skos:Concept rdf:about="http://www.eclap.eu/Classification/PerformingArtsType/504">
      <skos:prefLabel xml:lang="it">Teatro</skos:prefLabel>
      <skos:prefLabel xml:lang="en">Theatre</skos:prefLabel>
      <skos:prefLabel xml:lang="da">Teater</skos:prefLabel>
      <skos:prefLabel xml:lang="nl">Theater</skos:prefLabel>
      <skos:prefLabel xml:lang="fr">Théâtre</skos:prefLabel>
      <skos:prefLabel xml:lang="de">Theater</skos:prefLabel>
      <skos:prefLabel xml:lang="el">Θέατρο</skos:prefLabel>
      <skos:prefLabel xml:lang="hu">Színház</skos:prefLabel>
      <skos:prefLabel xml:lang="pl">Teatr</skos:prefLabel>
      <skos:prefLabel xml:lang="pt">Teatro</skos:prefLabel>
      <skos:prefLabel xml:lang="es">Teatro</skos:prefLabel>
      <skos:prefLabel xml:lang="ca">Teatre</skos:prefLabel>
      <skos:prefLabel xml:lang="sl">Gledališče</skos:prefLabel>
   </skos:Concept>
</rdf:RDF>
```

7.3 Delivery Method to Europeana Office

The delivery of metadata to the Europeana Office it is done using an OAI-PMH server that publish the metadata using the ECLAP schema. An XSLT has been developed to transform the source XML to EDM compliant schema, this XSLT is used at Europeana to transform records to EDM.
For each partner a specific OAI set has been created allowing to update only the relevant records.

The ingestion and production of records for Europeana follows the following steps:

1. Partners upload XML with metadata to the Metadata Mapping Portal that is mapped and then published for ECLAP portal
2. When a publish notification is received the metadata records are ingested in the ECLAP portal
3. If content is set to be present on the eclap FTP, it is checked if it is present and files that are not present are reported to the user.
4. If no errors are found the ingestion of content starts.
5. After 1 week from the ingestion/content upload the content is checked for the minimal metadata needed by Europeana. In case it is ok the content workflow state transit from Uploaded to Under-Approval otherwise an email is sent to the user.

6. When content is Under-approval it can be manually published or automatically published for Europeana (depends on the choice of the partner)
7. Regularly content that is marked published for Europeana is published on the OAI-PMH server. In this phase it is also checked if the preview icon with the minimum requirements for Europeana is present.

7.4 List of Changes Done to ECLAP Schema to Adapt the Model in Real Cases

The changes done to the initial ECLAP schema were:

1. Addition of PerformanceAndCrew to be distinct from Cast
2. Addition of Professional role: Concept_originator

Other changes were done for technical problems when validating the xml.

8 Conclusions

In this deliverable we mainly focus on the basic objectives of WP4 which are:

- to define metadata and descriptors coming from performing art institutions and suitable for posting on Europeana.
- To define interoperability map among several different models for metadata and descriptors for performing art content with respect to the semantic meaning of Europeana classification model.

To do so we first clarified the term metadata, meaning data used to describe other data structured in formats easily understood by machines and provided a general overview of the metadata framework which is the structural plan that ensures that metadata are formatted, structured, used, managed, and stored in an appropriate way (clarifying concepts such as schema, vocabulary conceptual model, content standard and encoding). We also had a brief look at the standards landscape and in order to appropriately handle metadata we provided a categorization of metadata standards that can be seen as: i) Descriptive data structure standards for different kinds of community resource descriptions. ii) Markup languages and schemas for encoding metadata in machine-readable syntaxes. iii) Ontologies for semantic mediation between data standards. iv) Protocols for distributed search and metadata harvesting. We emphasized on the Europeana metadata standards that are designed to provide integrated access to digital objects from the cultural heritage organizations of all the nations of the European Union: Europeana semantic elements specification (ESE) and Europeana data model (EDM).

After reviewing the most prominent metadata standards we presented the results of a survey that was performed in order to collect information about 'individual' collections that content providers are prepared to submit to ECLAP, what kind of metadata standards and in what way they are used among content providers in ECLAP. Based on the results and the analysis of the provided questionnaires and samples the most commonly used standards among content providers proved to be Dublin Core, MARC, EAD, CDWA and a simplified customization of FRBR. Unfortunately the majority of content providers did not use any standard for the metadata which makes it difficult to ingest this cultural data into ECLAP and Europeana

The ECLAP metadata schema was structured based on the EDM standard and taking into account the previously mentioned metadata standards. This approach ensured that ECLAP and Europeana have a similar basic structure and common elements while allowing different components to vary in depth and details. The interoperability and exchange of metadata between ECLAP and Europeana was further facilitated by creating a metadata crosswalk between the two schemas. The successfulness of the crosswalk was ensured by the similarity of the two schemes, the granularity of the elements in the Europeana scheme compared to that of

the ECLAP, and the compatibility of the content rules used to fill the elements of each scheme. In order to create the crosswalk between ECLAP and Europeana the ECLAP Metadata Ingestion Service portal was employed. The mapping service allows the user to define a mapping between the initial (if any) metadata schema and the target (ECLAP) schema. An XSLT is then generated, based on this mapping that can convert all existing items to the ECLAP format. In turn, the same procedure applies in order to perform the mapping from ECLAP schema to EDM and then deliver items (data and metadata) to Europeana office for publishing via the OAI-PMH protocol.

Thus, providing information about the aforementioned procedures (either in detail where necessary or briefly and then referencing the reader to other more detailed ECLAP reports) we make a comprehensive report about how the main WP4 objectives regarding definition of metadata descriptors and interoperability maps are reached within ECLAP.

9 Bibliography

[1] Metadata in the audiovisual production environment: an introduction / Annemieke de Jong. – Hilversum: Nederlands Instituut voor Beeld en Geluid, 2003

[2] Taxonomy of Knowledge Organization Sources/Systems (1). - Draft June 7, 2000 (revised July 31, 2000)

[3] http://nkos.slis.kent.edu/KOS_taxonomy.htm Last viewed 2006-09-14.

[4] José M. Martínez. MPEG-7 Overview, http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm.

[5] Jan Bormans, Keith Hill. MPEG-21 Overview, http://www.chiariglione.org/mpeg/standards/mpe g-21/mpeg-21.htm.

[6] Dublin Core Metadata Element Set, http://dublincore.org/documents/dcmi-terms/.

[7] Categories for the Description of Works of Art (CDWA) Metadata Standard, http://www.getty.edu/research/conducting_research/standards/cdwa/.

[8] LIDO v0.9 Specification Document: http://www.lido-schema.org/schema/v0.9/lido-v0.9-specification.pdf

[9] SPECTRUM Metadata Standard, http://www.mda.org.uk/spectrum.htm.

[10] IMS Metadata Standard, http://www.imsglobal.org/metadata/index.html.

[11] Art Museum Image Consortium (AMICO) Metadata Standard, http://www.amico.org/ AMICOlibrary/dataspec.html.

[12] http://www.loc.gov/marc/

[13] MAchine-Readable Cataloguing 21 (MARC21) Metadata Standard, http://www.bl.uk/ services/bibliographic/marc21move.html.

[14] MODS Metadata Object Description Schema http://www.loc.gov/standards/mods/

[15] METS http://www.loc.gov/standards/mets/

[16] EAD official website http://www.loc.gov/ead/

[17] Extensible Markup Language (XML), http://www.w3.org/XML/

[18] Resource Description Framework (RDF) http://www.w3.org/RDF/.

[19] RDF Vocabulary Description Language 1.0: RDF Schema http://www.w3.org/TR/rdf-schema/.

[20] Web Ontology Language (OWL) http://www.w3.org/2004/OWL/.

[21] Miles, Alistair, and Sean Bechhofer. "Skos simple knowledge organization system." World Wide Web Consortium (2009).

[22] CIDOC-Conceptual Reference Model (CRM), http://cidoc.ics.forth.gr/.

[23] FRBR http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

[24] FRBRoo http://www.cidoc-crm.org/frbr_inro.html

[25] Payette, Sandra D., and Oya Y. Rieger. "Z39. 50: The User"s Perspective." (1997).

[26] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm#Introduction.

[27] EDM Data Model Primer (http://version1.europeana.eu/c/document_library/get_file?uuid=718a3828-6468-4e94-a9e7-7945c55eec65&groupId=10605)

[28] Europeana Data Model presentation (for v5.2) (http://version1.europeana.eu/c/document_library/get_file?uuid=76eff9ae-5a70-409c-87b0-baf46ede7bd9&groupId=10602)

10 Glossary

AMICO	Art Museum Image Consortium
CDWA	Categories for the Description of Works of Art
CRM	Conceptual Reference Model
DC	Dublin Core
EAD	Encoded Archival Description
EDM	Europeana Data Model
ESE	Europeana Semantic Elements Specification
FRBR	Functional Requirements for Bibliographic Records
IMS	Instructional Management Systems
IPTC	International Press Telecommunications Council
LIDO	Lightweight Information Describing Objects
MARC21	Machine-Readable Cataloguing
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
MPEG	Moving Pictures Expert Group
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OWL	Ontology Web Language

OWL DL	Ontology Web Language Description Logics
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SKOS	Simple Knowledge Organisation System
SPECTRUM	Standard ProcEdures for CollecTions Recording Used in Museums
VRA	Visual Resources Association
XML	Extensible Markup Language

11 Appendix – Complete ECLAP Schema (DSI)

11.1 General information

axoid
unique identifier identifying the content on the portal, the id is based on UUID identification

url
The url on the portal where the content is available it is of the form:
http://www.eclap.eu/drupal?q=home&axoid=<axoid>

nid
drupal node id identifier, id associated by drupal to the content.

Version
Version number of the content, it is incremented when an update is performed

InsertUpdateTime
Date and time when the content was uploaded or when it was updated. The date is in the format YYYY-MM-DDThh:mm:ss.

ProviderId
ECLAP provider acronym used to identify the provider

ProviderName
Complete name of the provider

DefaultMetadataLanguage
default language used for the description of the resource, it should be a 2 letter ISO language code.

11.2 Digital resource information

Format
The resource format it can be “audio, video, document, image, crossmedia, 3d, archive, tool, playlist, collection”

Type
Specifies better the resource format for crossmedia (html, flash) and document (document, epub, pdf, excel, slide, braille music)

Width
Width of the frame for image or video

Height
Height of the frame for image or video

Duration

Duration of audio or video resource in the form “hh:mm:ss.mm”

AvlMDVideo
Is “yes” if a medium definition video resource is available

AvlHDVideo
Is “yes” if a high definition video resource is available

Extension
The file extension for the digital resource.

Preview
Url used for the content preview, for video it is an animated gif.

AvlForPDA
Is “yes” if a version for WindowsMobile 6.5 has been produced

AvlForIPhone
Is “yes” if the content can be used on iPhone or in general on Smart Phones (iOS, Android, WindowsPhone7)

AvlForPC
Is “yes” if the content can be used on a Personal Computer (Windows/MAC/Linux)

11.3 IPR information

IsPublic
Is “yes” if the IPR model associated with the content is public

IPRTitle
The title given to the IPR model associated with the content

IPRDescription
The description given to the IPR model associated with the content

EuropeanaRightsUrl
The Europeana Url given to the IPR model associated with the content

LicenseUrl
The license url given to the IPR model associated with the content

11.4 GeoSpatial Information

This section can provide a set of GPS coordinates
Latitude
The decimal representation of the latitude

Longitude
The decimal representation of the longitude

Radius
The radius in meters of the area where the content is “active”

11.5 Performing arts metadata

In this section are reported the metadata specific for performing arts.

FirstPerformance Place

Name of the theatre or venue where the performance taken place for the first time.

Examples	"Théâtre des Bouffes du Nord"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "place", might not correspond with the place in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held at: "Théâtre des Bouffes du Nord", but what we are looking at on the ECLAP portal might be a video of the performance held months later - while the show was touring – at "The Globe Theatre"
Refinement of	dcterms:spatial

FirstPerformance City

Name of the city where the first performance taken place.

Examples	"Paris"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "City", might not correspond with the city in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held in: "Paris", but what we are looking at on the ECLAP portal might be a video of the performance held months later - while the show was touring – in "London"
Refinement of	dcterms:spatial

FirstPerformance Country

Name of the country where the first performance taken place

Examples	"France"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "Country", might not correspond with the country in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held in: "France", but what we are looking at on the ECLAP portal might be a video of the performance held months later - while the show was touring – in "England".
Refinement of	dcterms:spatial

FirstPerformance Date

Date of the first performance

Examples	"2000-11-20"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "date", might not correspond with

	the date in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held in: "2000-11-20", but what we are looking at on the ECLAP portal might be a video of the performance held months later, in "2001-04-05"
Refinement of	DCTerms.issued

Performance Place
Name of the theatre or venue where the shown performance taken place

Examples	"The Globe Theatre"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "place", might not correspond with the place in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held at: "Théâtre des Bouffes du Nord", but what we are looking at on the ECLAP portal might be a video of the performance held months later - while the show was touring – at "The Globe Theatre"
Refinement of	dcterms:spatial

Performance City
Name of the city where the shown performance taken place.

Examples	"London"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "City", might not correspond with the city in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held in: "Paris", but what we are looking at on the ECLAP portal might be a video of the performance held months later - while the show was touring – in "London"
Refinement of	dcterms:spatial

Performance Country
Name of the country where the shown performance taken place

Examples	"England"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "Country", might not correspond with the country in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held in: "France", but what we are looking at on the ECLAP portal might be a video of the performance held months later - while the show was touring – in "England".
Refinement of	dcterms:spatial

Performance Date

Date of the shown performance

Examples	"2001-04-05"
Count (1, 0..1, 0..many, 1..many)	0..1
Notes	the first performance is the première, therefore its "date", might not correspond with the date in which the show was recorded. For example: the opening night of "The Tragedy of Hamlet" directed by P. Brook might be held in: "2000-11-20", but what we are looking at on the ECLAP portal might be a video of the performance held months later, in "2001-04-05"
Refinement of	dcterms:issued

PerformingArtsGroup

Name of the theatre or dance company or musical group (if present)

Examples	"Momix"
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	None
Refinement of	dc:creator

PlotSummary

Summary of the plot

Examples	"Prince Hamlet mourns both his father's death and his mother, Queen Gertrude's remarriage to Claudius. The ghost of Hamlet's father appears to him and tells him that Claudius has poisoned him: Hamlet swears revenge, etc."
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	None
Refinement of	dc:description

Cast

Name/Names of a member of the cast.

Examples	"Ryszard Cieślak, Rena Mirecka, Antoni Jahołkowski, Mieczysław Janowski, Maja Komorowska, Stanislaw Scierski"
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	Use this element only if the Professional elements cannot be used, as the case of a cast written in a single text that cannot be easily split in all the different professional people
Refinement of	dc:contributor

PerformersAndCrew

Name/Names of a performers and crew of a performance.

Examples	...
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	Use this element only if the Professional elements cannot be used, as the case of a performers written in a single text that cannot be easily split in all the different

	professional people
Refinement of	dc:contributor

Professional
A list of the people involved in the performance indicating which role each person had in the performance (eg. Actor, director, set designer etc.). It includes all the information listed in a playbill, such as the artistic cast of the show and the technicians, but also the names of the troupe which recorded the performance (eg. Cameraman, Director of Photography, etc.). Possible roles are:

- Acrobat
- Actor
- Adaptator
- Architect
- Assistant_director
- Casting
- Choreographer
- Clown
- Composer
- Concept_originator
- Costume_designer
- Critic
- Dancer
- Director
- Dramaturge
- Hairdresser
- Light_designer
- Make-up_artist
- Marketing_manager
- Mask_designer
- Mime
- Musician
- Patron
- Performer
- Playwright
- Producer
- Puppet_designer
- Scenographer
- Seamster
- Set_builder
- Set_designer
- Singer
- Sound_designer
- Stage_manager
- Technician
- Theatre_manager
- Theoretician
- Translator
- Other

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	None

Refinement of	dc:contributor

Object
Objects used in the performance, (i.e. Sets, Costumes, Props, Programs, Prints, Drawings,...)

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	
Refinement of	DC.description

Genre
The genre in which the work can be categorized (i.e. Ballet, Butho, Commedia dell'Arte, Drama, Feast Flamenco, etc)

Examples	"Tragedy"
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	we will work on a shared vocabulary for this
Refinement of	dc:subject

PerformingArtType
Type of performing art present in the content.

Examples	"theatre"
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	identified in WP4 as cinema, dance, music, theatre, performance art
Refinement of	dc:type

HistoricalPeriod
Historical period the topic of the resource refers to.

Examples	"XV century"
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	None
Refinement of	dcterms:temporal

ArtisticMovementAndActingStyle
Artistic movement and acting styles in which the work can be categorized (e.g. Classicism, Dada, Epic, Expressionism, etc.)

Examples	"Futurism"
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	we will work on a shared vocabulary for this
Refinement of	dcterms:type

ManagementAndOrganization
management and organization…

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	Deprecated

RecordingDate
Date of creation of the digital object,

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	Use this element in case what it is recorded is not a public performance (e.g. an interview) otherwise use the Performance Date
Refinement of	dc:date

PersonRecord
Credits for the audio or video recording

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	Deprecated, use Professional elements
Refinement of	dc:contributor

PieceRecord
Credits for the text or image. The meaning of this field is a bit complex.... The text we are dealing with in this field is the script of the play. We intend this field to be filled out with the original title of the performance (eg. Medea) - which might differ from the title of the item (eg. Photo of Medea_2) - and with the name of the person who wrote the script. The records pertaining to the novel or the literary work which inspired the script should be mapped in the field "reference" instead; the field “reference” should also include the title of the novel and its author(s).

Examples	Title: Il Principe Costante; scenario: Jerzy Grotoski; adaptation: Julius Slowacki
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	None
Refinement of	dcterms:references

ProductionRecord
Credits of the production team. the name of the producer(s) and of other people involved in the organization.

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Notes	Deprectaed, use the Professional element with the appropriate role
Refinement of	dc:contributor

11.6 DC – dublin core metadata

This section contains information about the dublin core metadata to be associated with the content to be ingested:

title
The name given to the resource. Typically, a Title will be a name by which the resource is formally known. The title of the original analog or born digital object. The title should be significant.

Examples	“Romeo and Juliet”
Count (1, 0..1, 0..many, 1..many)	1..many
Language	Mandatory
Notes	None

creator
An entity primarily responsible for making the content of the resource. Examples of a Creator include a person, an organization, or a service. Typically the name of the Creator should be used to indicate the entity. In ECLAP, the name of Partner uploading is kept automatically in a separate field. This is the name of the creator of the original analog or born digital object. . This field should be used only to indicate the creator of the work of art (usually the director for a performance, the author if we are dealing with a book, the composer if we are uploading a script and so on). Often, in devised work, the creator might be the whole company or the actors might collaborate with the director. Nevertheless I guess we need to set a rule to be applied to every situation, so that I would consider actors and other artistic figures as contributors and eventually explain in the field "description" if their role as creator of the performance was capital.

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Language	Optional
Notes	None

subject
The topic of the content of the resource. Typically, a Subject will be expressed as keywords or key phrases or classification codes that describe the topic of the resource. Recommended best practice is to select a value from your own classification scheme. This is the subject of the original analog or born digital object.

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Language	Mandatory
Notes	None

description
An account of the content of the resource. Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content. A description of the original analog or born digital object.

Examples	…
Count (1, 0..1, 0..many, 1..many)	0..many
Language	Mandatory
Notes	None

publisher
The entity responsible for making the resource available. Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity. In ECLAP, the name of Partner that has provided the content is automatically tracked and stored in a different field. The name of the publisher of the original analog or born digital object.

Examples	In case of a performance review the name of the newspaper where the review was published
Count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
Notes	None

contributor

An entity responsible for making contributions to the content of the resource. Examples of a Contributor include a person, an organization or a service. Typically, the name of a Contributor should be used to indicate the entity. In most cases, the authors of a document are listed here. The name of contributors to the original analog or born digital object. This could be a person, an organisation or a service.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

date

A date associated with an event in the life cycle of the resource. Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [Date and Time Formats, W3C Note, http://www.w3.org/TR/NOTE-datetime] and follows the YYYY-MM-DD format. If the full date is unknown, month and year (YYYY-MM) or just year (YYYY) may be used. Many other schemes are possible, but if used, they may not be easily interpreted by users or software. Use for a significant date in the life of the original analog or born digital object. Use dcterms:temporal (or dc:coverage) if the date is associated with the topic of the resource.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

type

The nature or genre of the content of the resource. Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMIType vocabulary http://dublincore.org/documents/dcmi-type-vocabulary/). To describe the physical or digital manifestation of the resource, use the FORMAT element. The type of the original analog or born digital object as recorded by the content holder, this element typically includes values such as photograph, painting, sculpture etc.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Mandatory
notes	None

format

The physical or digital manifestation of the resource. Typically, Format may include the media-type or dimensions of the resource. Examples of dimensions include size and duration. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [http://www.iana.org/ assignments/media-types/] defining computer media formats). The unqualified element includes file format, physical medium or dimensions of the original and/or digital object. Use this element for the file format of the digital object or born digital originals. Internet Media Types [MIME] are highly recommended (http://www.iana.org/assignments/media-types/). Use of the more specific elements dcterms:extent (dimensions) and dcterms:medium (physical medium) is preferred where appropriate.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional

notes	None

identifier
An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN). This is the identifier for the original analog or born digital object.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

source
A Reference to a resource from which the present resource is derived. The present resource may be derived from the Source resource in whole or part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system. In general, include in this area information about a resource that is related intellectually to the described resource but does not fit easily into a Relation element. In ECLAP, this value should be the URL or the filename of the original resource. The file uploaded and the URL provided in the upload form are tracked automatically in different fields. This element can be used for several different types of source that are related to the object (such as reference sources). The name of the content holder should no longer be recorded here as a new element.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

language
A language of the resource. Use ISO 639 two letter language tags (it, en, fr, de, el, …) Use this element for the language of textual objects and also where there is a language aspect to other objects e.g. sound recordings, posters, newspapers etc). If there is no language aspect to the digital object (e.g. a photograph), please ignore this element. This element is not for the language of the metadata of a resource, which may be described in xml:lang attribute. In case the digital object presents more languages, use more language elements, one for each language.

examples	en, it, fr, de, el, hu, es, ca
count (1, 0..1, 0..many, 1..many)	0..many
language	No
notes	None

relation
A reference to a related resource. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system. This is information about resources that are related to the original analog or born digital object.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

coverage
The extent or scope of the content of the resource. Coverage will typically include spatial location (a place name or geographic co-ordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [Getty Thesaurus of Geographic Names, http://www. getty.edu/research/tools/vocabulary/tgn/]). Where appropriate, named places or time periods should be used in preference to numeric identifiers such as sets of co-ordinates or date ranges. Coverage is the unqualified spatial or temporal coverage of the original analog or born digital object. Use of the more specific dcterms:spatial and dcterms:temporal elements is preferred where possible.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

rights
Information about rights held in and over the resource. Typically a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource. This is a free text element and should be used for information about intellectual property rights or access arrangements for the digital object that is additional to the controlled value provided in europeana:rights.

examples	"All rights reserved"
count (1, 0..1, 0..many, 1..many)	0..many
language	Mandatory
notes	None

11.7 DCTERMS – dublin core terms metadata

In this section are reported the dcterms elements that are supported, that are the ones supported by Europeana.

alternative
An alternative name given to the resource. Typically, an Alternative title will be a name by which the resource is alternatively referred and it is different from the formal Title. Any alternative title by which the original analog or born digital object is known. This can include abbreviations or translations of the title.

Examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Mandatory
notes	None

tableOfContents
A list of subunits of the resource. A list of the units within the original analog or born digital resource object.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Mandatory
notes	None

created
Date of creation of the resource. This is the date when the original analog or born digital object was created.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

issued
Date of formal issuance (e.g., publication) of the resource. The date when the original analog or born digital object was issued or published.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

extent
The size or duration of the resource. Refinement of format. Size or duration of the digital object and the original object may be recorded.

examples	“30 pages”, “01:15:20”
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

medium
The material or physical carrier of the resource. Refinement of dc:format.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

isVersionOf
A related resource of which the described resource is a version, edition, or adaptation. Changes in version imply substantive changes in content rather than differences in format. Refinement of dc:relation. See also dcterms:hasVersion.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

hasVersion
A related resource that is a version, edition, or adaptation of the described resource. Changes in version imply substantive changes in content rather than differences in format. Refinement of dc:relation. See also dcterms:isVersionOf. Use dcterms:hasFormat for differences in format.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many

language	Optional
notes	None

isReplacedBy
A related resource that supplants, displaces, or supersedes the described resource.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

replaces
A related resource that is supplanted, displaced, or superseded by the described resource.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

isRequiredBy
A related resource that requires the described resource to support its function, delivery, or coherence.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

requires
A related resource that is required by the described resource to support its function, delivery, or coherence.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

isPartOf
Is Part Of - A related resource in which the described resource is physically or logically included. Use for the name of the collection which the digital object is part of.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

hasPart
A related resource that is included either physically or logically in the described resource. Refinement of dc:relation. See also dcterms:isPartOf.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional

notes	None

isReferencedBy
Is Referenced By: A related resource that references, cites, or otherwise points to the described resource.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

references
A related resource that is referenced, cited, or otherwise pointed to by the described resource.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

isFormatOf
A related resource that is substantially the same as the described resource, but in another format. Refinement of dc:relation. See also dcterms:hasFormat.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

hasFormat
A related resource that is substantially the same as the pre-existing described resource, but in another format. Refinement of dc:relation. See also dcterms:isFormatOf. Use dcterms:hasVersion for differences in version.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

conformsTo
An established standard to which the described resource conforms. Refinement of dc:relation. The names of standards that the digital object (digitized or born digital) complies with and which are useful for the use of the object.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

spatial
Spatial characteristics of the resource. Information about the spatial characteristics of the original analog or born digital object, i.e. what the resource represents or depicts in terms of space. This may be a named place, a location, a spatial coordinate or a named administrative entity.

examples	…

count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

temporal
Temporal characteristics of the resource. The temporal characteristics of the original analog or born digital object i.e. what the resource is about or depicts in terms of time. This may be a period, date or date range.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

provenance
A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation. The statement may include a description of any changes successive custodians made to the resource. This relates to the ownership and custody of the original analog or born digital object.

examples	…
count (1, 0..1, 0..many, 1..many)	0..many
language	Optional
notes	None

11.8 Taxonomy Classification

For each drupal taxonomy term associated with the content it is reported:
label
The label of the term in each available language

id
attribute with the drupal id for the term

root
attribute with the id of the root term where the term is a descendent.

vid
attribute with the id of the vocabulary of the term

path
attribute with the term ids separated by spaces from the root to the term (e.g. “664 668”)

11.9 ECLAP Groups

For each drupal og group associated with the content it is reported:
label
the label of the group

id
attribute with the id of the group

11.10 ECLAP Aggregations

If the content is a playlist or a collection are reported the set of axoids that belong to the playlist/collection.

For playlists more information is provided for each content in the playlist:

For audio and video:
startTime
attribute with the time instant in seconds from the audio/video start representing the time in the resource to start resource Audio/Visual rendering, if omitted the resource start time is intended

endTime
attribute with the time instant in seconds from the audio/video start representing the time in the resource to end the resource Audio/Visual rendering, if omitted the resource end time is intended

For images:
duration
the duration in seconds of the image display