EUROPEAN COLLECTED LIBRARY OF ARTISTIC PERFORMANCE

www.ECLAP.eu

Grant Agreement No 250481

DE4.2.1
Content and Metadata, Selection Aggregation and Augmentation

Version: 7.1
Date: 10/05/2011

Project Title: ECLAP Project Number: ICT-PSP-250481 Deliverable Number: DE4.2.1 Accessibility: public Work-Package contributing to the Deliverable: (WP4, WP4.2, WP4.3, WP4.4) Nature of the Deliverable: document Status: final

Contractual Date of Delivery: 30/04/2011 Approve for quality control by: 10-05-2011 Finally approved by coordinator: 10-05-2011 Actual Date of Delivery: 10-05-2011

Document responsible: Natasa Sofou Email address: natasa@image.ntua.gr Affiliation acronym: NTUA
Authors: • Natasa Sofou, Costas Pardalis, Nikos Simou (NTUA) • Pierfrancesco Bellini, Michela Paolucci, Marco Serena, Nicola Mitolo, Paolo Nesi (DSI) • Jaap Bloom (B&G) • Maia Borelli, Irene Scaturro, Luisa Tinti, Ferruccio Marotti (UNIROMA)

Revision History:

Revision	Date	Author	Organization	Description
v.01	**14/03/2011**	**Natasa Sofou**	**NTUA**	**ToC**
v.02	**7/04/2011**	**Natasa Sofou**	**NTUA**	
v.03	**19/04/2011**	**Natasa Sofou**	**NTUA**	**Contributions from all partners added**
v.04	**27/04/2011**	**Natasa Sofou**	**NTUA**	**Finalizing contributions from all partners**
v.6	**06-05-2011**	**Paolo Nesi**	**DSI**	**Revise and closure**

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This deliverable contains original unpublished work except where clearly indicated otherwise.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	**Content and Metadata, Selection, Aggregation and Augmentation**
Identifier.de	DE4.2.1
Identifier.ISBN	
Creators	Natasa Sofou, Costas Pardalis, Nikos Simou (NTUA); Pierfrancesco Bellini, Michela Paolucci, Marco Serena, Nicola Mitolo (DSI); Jaap Bloom (B&G); Ferruccio Marotti, Maia Borelli, Irene Scaturro, Luisa Tinti (UNIROMA)
Subject	Content and Metadata, Selection, Aggregation and Augmentation
Description	Report describing content and metadata selection aggregations and enrichment
Keywords	Content, metadata, metadata mapping, interoperability, content clustering, content aggregation, content augmentation, rich media
Publisher	ECLAP
Date	10-05-2011
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 5

2 INTRODUCTION 6

3 METADATA, DESCRIPTORS AND INTEROPERABILITY 7

3.1 METADATA DESCRIPTORS 7
3.1.1 Europeana Data Model Overview 7
3.1.2 EDM Schema 9
3.1.3 EDM RDF 14
3.2 INTEROPERABILITY MAPPINGS 15
3.2.1 ECLAP Metadata Ingestion 15
3.2.2 The Mapping Module 25
3.2.3 Statistics 31
3.2.4 Transformation Services 32
3.2.5 Implementation Details 34
3.2.6 Functional Analysis 34
3.2.7 Related Work 35
3.3 ECLAP METADATA INGESTION OAI_PMH SERVER 39
3.3.1 Introducing OAI_PMH 39
3.3.2 Analysis of the ECLAP Metadata Ingestion OAI_PMH Server Architecture and Implementation 40

4 CONTENT SELECTION AND AGGREGATION 45

4.1 EUROPEAN POLICY ON BROADENING THE ONLINE ACCESS TO CREATIVE CONTENT AND PERFORMING ARTS DEFINITION 45
4.2 CONTENT SELECTION CRITERIA 45
4.3 AGGREGATION SELECTION CRITERIA 48
4.4 CONTENT CONTEXTUALIZATION CRITERIA 50
4.5 CONTENT ORGANIZATION 53
4.5.1 Organization 53
4.5.2 Clustering 53
4.5.3 ECLAP's Monographs 57
4.5.4 Issues to Be Addressed 57

5 CONTENT AUGMENTATION AND RICH MEDIA 58

5.1 CONTENT AGGREGATION TOOLS AND RICH MEDIA CONTENT 58
5.1.1 ECLAP Playlists 58
5.1.2 ECLAP Collections 63
5.1.3 ECLAP e-learning Courses 65
5.1.4 Rich Media and Intelligent Content for PC & Mobile 72
5.2 CONTENT AUGMENTATION AND ENRICHMENT 76
5.2.1 Votes and comments 76
5.2.2 Free tagging 77
5.2.3 Annotations and non linear media fruition 80
5.2.4 Metadata enrichment, taxonomization and translation 84
5.2.5 Technical metadata production 90

6 BIBLIOGRAPHY 92

7 GLOSSARY 94

ANNEX I 95

1 Executive Summary and Report Scope

WP4 involves the selection aggregation and augmentation 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.

It involves the development of a realisable content selection strategy that appeals to the broadest range of user communities, that most effectively supports and integrates with the Europeana initiative, but that also stays in line with content providers' holdings. The work package includes the mapping of proprietary metadata sets to the common metadata schema. The work package also includes the uploading of content and metadata. The development of both the content selection policy and the common metadata schema necessarily involves a review, revision or enhancement of the available metadata schemas in the cultural domain.

Main WP4 objectives include:

- To collect and create online access to theatrical content catalogues with their metadata and descriptors;
- To collect and create online access to performing art content at item level listed in these open access catalogues across Europe.
- 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.
- To collect and automatically produce rich media content for further enrichment and augmentation integrating descriptors, annotations and digital resources with features and intelligence for their perusal.

This is the second deliverable of WP4. It is structured in three parts.

- Metadata - descriptors and interoperability. Responsible partner: NTUA. This part focuses on the interoperability among metadata and classifications, taking into account Europeana standard.
- Content Selection and Aggregation. Responsible partner: UNIROMA. This part includes the selection criteria for content and their aggregation and contextualization. The aggregated content is going to have explicit relationships to make possible their reciprocal linking into ECLAP and possibly into Europeana. The content would be grouped into semantic clusters to simplify the usage activities for education and training.
- Content Augmentation and Rich Media. Responsible partner: DSI. This part of the deliverable includes the description of the solution used for content augmentation and enrichment, thus reporting the activity carried out towards this direction. It also includes some criteria and models to create rich media content including original sources and the obtained semantic descriptors in a form usable from end users.

2 Introduction

This deliverable focuses on content and metadata selection, aggregation and augmentation. Its contribution is crucial for WP4 since it summarizes the work of all WP4 tasks. For convenience, the deliverable is structured in three parts:

1. Metadata - Descriptors and Interoperability. This part focuses on the interoperability among metadata and classifications, taking into account Europeana standard. In particular it provides an overview of the Europeana Data Model, EDM schema and EDM RDF. The ECLAP aggregation and ingestion services are described in detail with special focus on the mapping module, transformation services, statistics, functional analysis implementation details and related work. The OAI-PMH mechanism for interoperability between the Ingestion Tool and various other modules or platforms is described, with special focus on Europeana. Specifically the Ingestion Tool constitutes the data provider while Europeana or the ECLAP portal are service providers that are able to request and retrieve metadata records via the OAI-PMH verbs or services.
2. Content Selection and Aggregation. This part includes the selection criteria for content and their aggregation and contextualization. The aggregated content is going to have explicit relationships to make possible their reciprocal linking into ECLAP and possibly into Europeana. The content would be grouped into semantic clusters to simplify the usage activities for education and training.
3. Content Augmentation and Rich Media. This part of the deliverable includes the description of the solution used for content augmentation and enrichment, thus reporting the activity carried out towards this direction. It also includes some criteria and models to create rich media content including original sources and the obtained semantic descriptors in a form usable from end users.

In the remainder of this document, each of the aforementioned parts is described in detail.

3 Metadata, Descriptors and Interoperability

3.1 Metadata Descriptors

This section focuses on the metadata standards, interoperability and mappings between ECLAP metadata schema and Europeana standards. In the following sections there is an overview of the Europeana Data Model, EDM schema and EDM RDF. The ECLAP aggregation and ingestion services are described in detail with special focus on the mapping module, transformation services, statistics, functional analysis implementation details and related work.

3.1.1 Europeana Data Model Overview

In this section a short introduction to the Europeana Data Model (EDM) is given. The interested reader is referred to [1] and [2] for a complete and comprehensive reference to EDM. EDM was proposed in order to structure the data that Europeana ingests, manages and publishes and it is a major improvement of Europeana Structural Elements (ESE), which was the initial data model that Europeana began life with. Figure 3.1 graphically summarizes the hierarchy of the EDM classes.

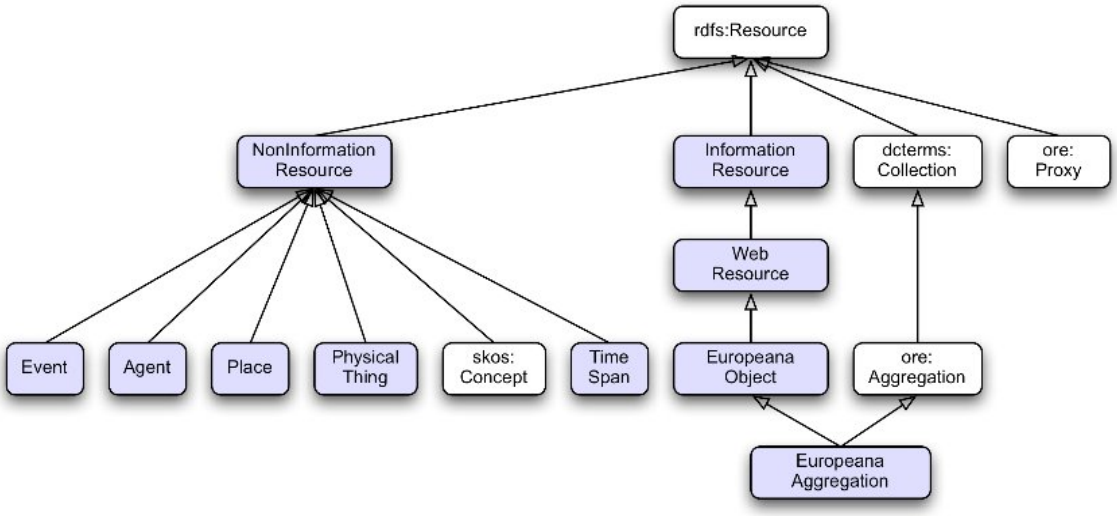


Figure 3.1 *The EDM Class hierarchy. The classes introduced by EDM are shown in light blue rectangles. The classes in the white rectangles are re-used from other schemas; the schema is indicated before the colon.*

As we can observe, the classes that are defined in EDM (blue rectangles) can be divided into NonInformation and Information Resources. Information Resource class represents resources whose essential characteristics can be conveyed in a single message e.g. a text is an Information Resource. Web Resource is a subclass of Information Resource and more specifically it is defined as an Information Resource that has at least one Web Representation and at least a URI. EuropeanaObject and EuropeanaAggregation are subclasses of WebResource, but since these classes are only used from Europeana for managing the data and they are beyond the interest of this document.

On the other hand, all the resources that are not Information Resources are instances of the NonInformation Resource class that has various subclasses. Firstly, class Event represents a change of states in cultural, social or physical systems, regardless of scale, brought about by a series or group of coherent physical, cultural, technological or legal phenomena or a set of coherent phenomena or cultural manifestations bounded in time

and space. The second subclass of NonInformation Resource, that is Agent class, comprises people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible. Class Place represents an extent in space, in particular on the surface of the earth, in the pure sense of physics: independent from temporal phenomena and matter. The next one is PhysicalThing class which represents a persistent physical item such as a painting, a building, a book or a stone. Finally, class TimeSpan is an abstract temporal extent having a beginning, an end and duration.

The following figure illustrates the properties defined in EDM to interconnect the classes. Detailed explanation of their semantics is not given in this section. A short description for some of the properties is provided in the next section.

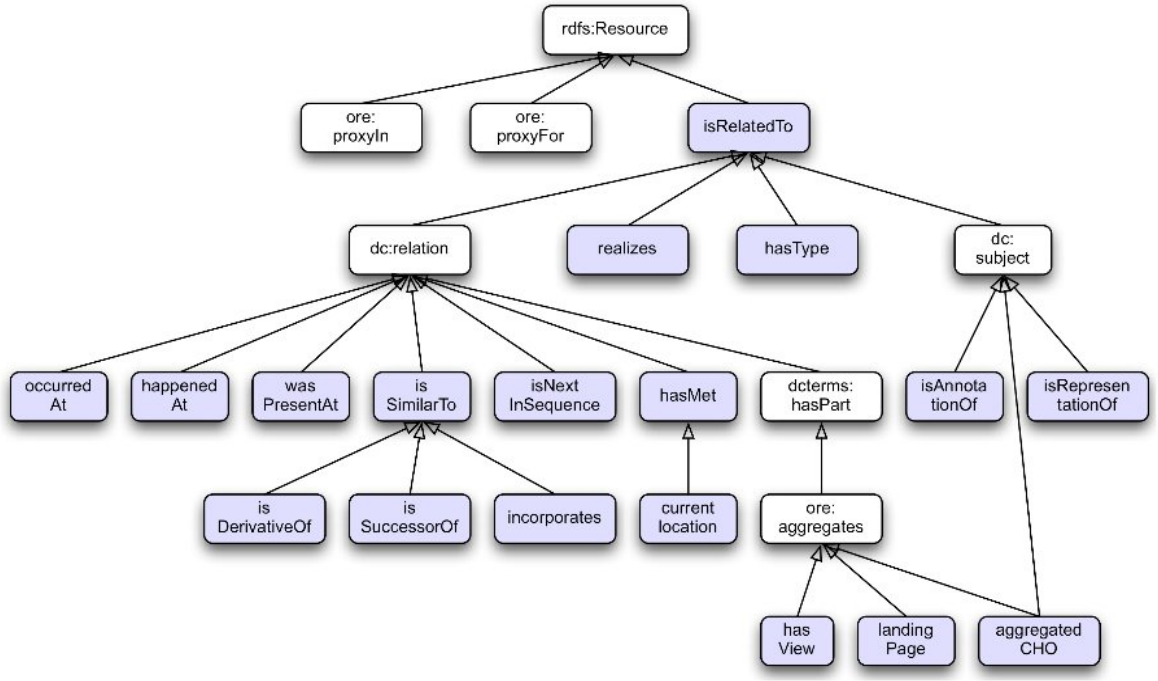


Figure 3.2 *The EDM property hierarchy without the properties included in ESE (for readability). The properties introduced by EDM are shown is light blue rectangles. The properties in the white rectangles are re-used from other schemas.*

The overall data structuring within EDM is based on two classes borrowed from Object Reuse and Exchange (ORE) schema ore:aggregation and ore:proxy. Every cultural heritage object (e.g. a painting) has a digital representation (e.g. a thumbnail of painting's digital picture). Therefore a distinction between the actual works and the digital representation is necessary. EDM manages this distinction by following the ORE specification and using aggregations. Hence, EDM considers that the provided object, together with the digital representations that are contributed by one provider, form an aggregation. Proxies, on the other hand, are used in order to handle the fact that Europeana takes data from many providers and this data may be about the same real world resource, thus giving multiple views on the same resource. In addition, Europeana can add its own data about that resource giving yet another view on the same resource. Therefore, since it very likely that the metadata differ, e.g., different names may be used for the same creator, these views remain distinct by using ore: proxies.

To better understand the way data are represented using EDM, let us assume that we have two records of Mona Lisa, one from the Joconde database and another from the Louvre. Each data submission to Europeana gives rise to a specific instance of the ore:Aggregation class, used to group all the elements related to one resource that come from one provider. Both providers indeed contribute a different set of digital

representations, e.g., different resolutions, different file types and, of course, different locations for the representations. In this way an aggregation is one provider's contribution for an object. Moreover each metadata record provided to Europeana also gives raise to a specific proxy for the object described, modelled using the ore:Proxy resource. This proxy is specific to a given provider, and is used to represent the description of the provided object, as seen from the perspective of that specific provider. With proxies it is possible to represent different, possibly conflicting pieces information on provided objects, while still keeping track of the provenance of this information. A proxy is connected to the resource by using the ore:proxyFor property while it is connected to its provider's aggregation using the ore:proxyIn property. *Figure 3.3* provides a graphical representation of the described example.

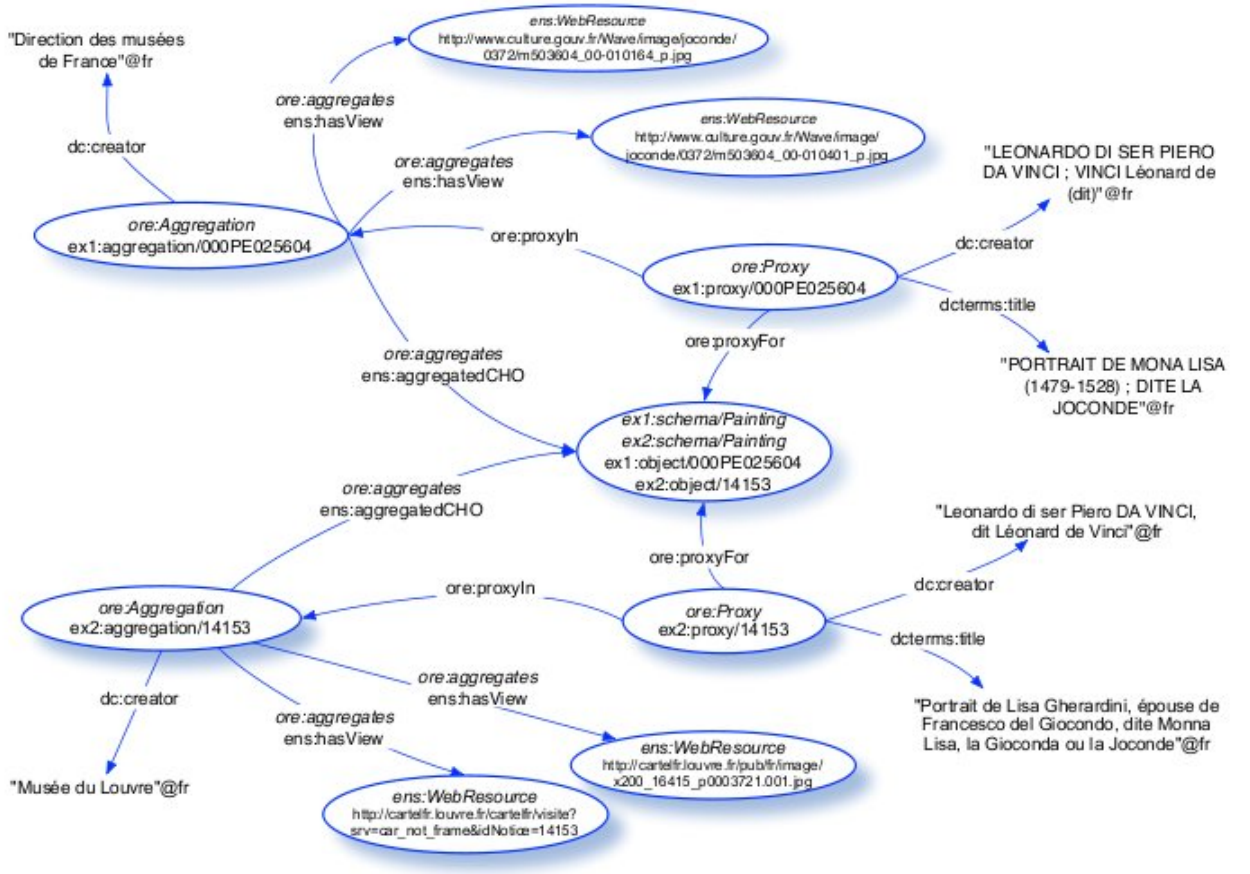


Figure 3.3 *EDM Representation for two providers for an object.*

3.1.2 EDM Schema

An xml schema was implemented based on the EDM in order to mediate between content provider's native data and the data models implemented by ECLAP. More specifically the main requirement for this schema was to act as a harvesting XML schema intended for delivering metadata to the ECLAP service environment about an organisation's online collections and digital objects. Furthermore, this schema had to ensure interoperability between the native metadata held by heritage organisations and the standards and schemas used in ECLAP and Europeana.

Having in mind the EDM, we implemented a schema that is based on EDM's classes. In general the design methodology that was followed for the implementation of EDM XSD was to construct complex types for the EDM classes that are comprised of elements that correspond to the EDM properties. Therefore, since according to EDM for every ingested item given by the provider an aggregation is constructed, we defined a complex type named AggregationType to represent aggregation.

As illustrated in Figure 3.4 AggregationType consists of 4 mandatory elements which are the proxy, the aggregated Cultural Heritage Object (CHO), the webResources and the creator. Among them only the creator is of SimpleLiteral type, while the remaining are complex types in order to represent the corresponding EDM classes.

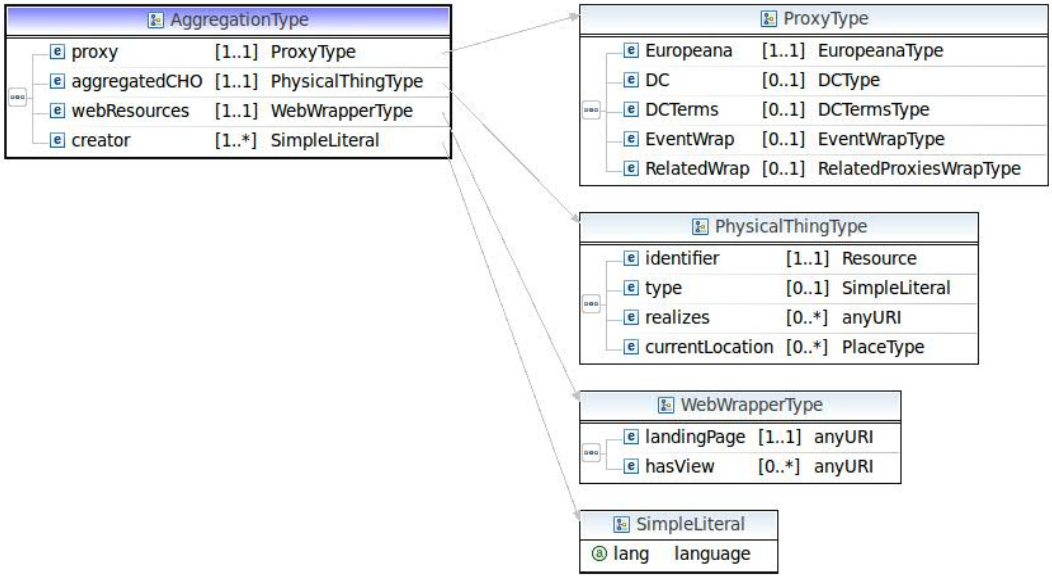


Figure 3.4 *Overview of EDM XSD*

ProxyType is a complex type repsesenting EDM's class proxy. Since proxy class is responsible for holding all the metadata of the CHO, it consists of five other complex types. DC and DCTerms elements, which are illustrated in the following figures, were generated to collect the metadata borrowed from the DC and DCTerms schemas respectively.

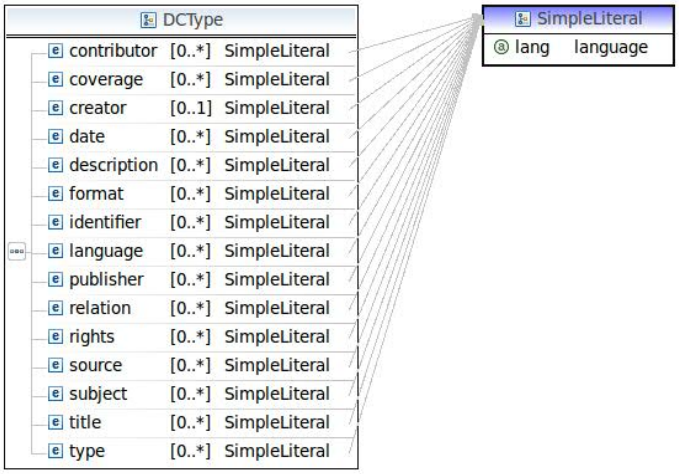


Figure 3.5 *DCType*

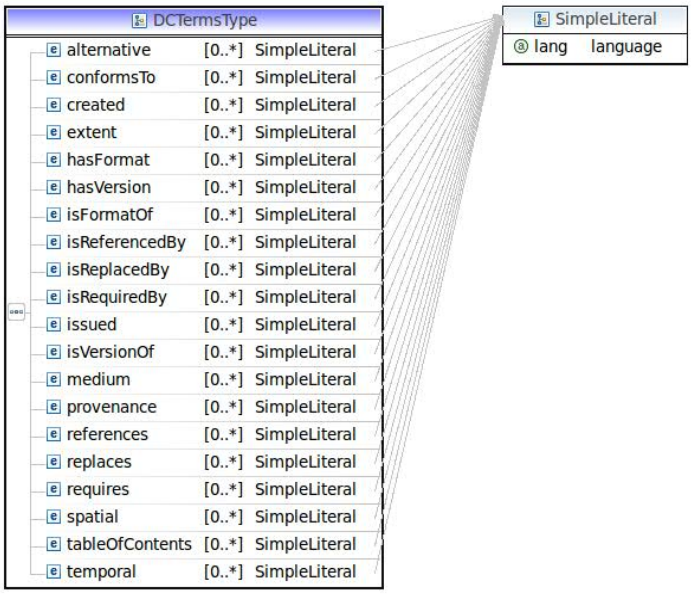


Figure 3.6 *DCTermsType*

None of the elements included is mandatory and their type is SimpleLiteral. A more correct representation, in terms of DC and DCTerms properties semantics, would be to set the type of these elements to SimpleLiteral or anyURI, since these properties are defined as RDF properties in DC and DCTerms respectively. However, this representation would prove rather confusing in our case, since the scope of EDM schema is to act as a harvesting schema and the existing metadata about CHOs are literals and not resources.

The third complex type that is included in ProxyType is the EuropeanaType that is the only mandatory element in proxy and includes the properties defined in EDM (see Figure 3.7). In this case, there exist mandatory elements, and not all elements are of the same type. The mandatory elements were defined according to the EDM. Hence the properties that are specified with cardinality 1 were converted to the mandatory elements. These are the country and the language that are of SimpleLiteral type, the provider that is of string type, since different languages cannot apply to this field, and the type that is of Edmtype type. Edmtype is a string which can only take one value from TEXT, VIDEO, IMAGE and SOUND defined in that way in accordance to the EDM specification for the type property. For the same reason (i.e. because of the EDM specification) the type of rights and uri elements was set to anyURI.

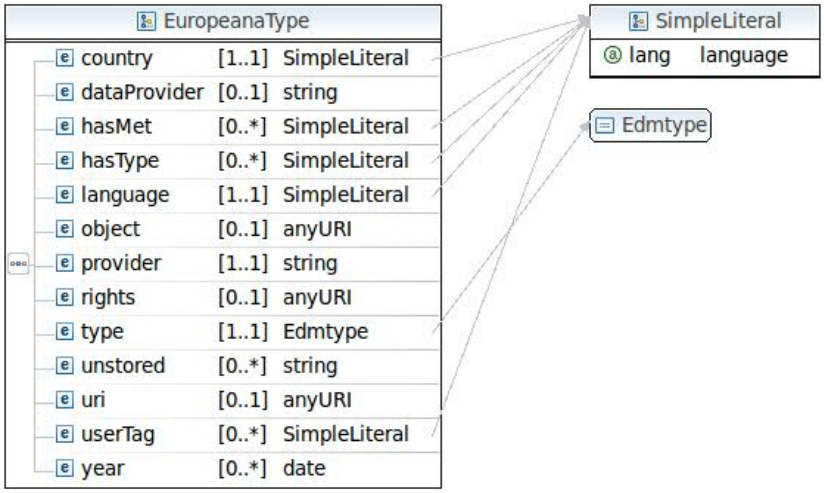


Figure 3.7 *EuropeanaType*

The fourth element that is included in ProxyType is the EventWrapType, which was implemented to in order to fit the EDM class Event. More specifically this type represents the EDM property wasPresentAt which relates a Proxy with an Event. At this point, we should mention that EDM permits both "object centric" and "event centric" approaches for metadata. The first one focus on the object described, in other words information comes in the form of statements that provide a direct linking between the described object and its features. Object centric approach is represented in our schema by DCType and DCTermsType. Event-centric approach, on the other hand, considers that descriptions of objects should focus on characterizing the various events in which objects have been involved. This approach underlies models such CIDOC-CRM and is represented in our schema by EventWrapType and EventType.

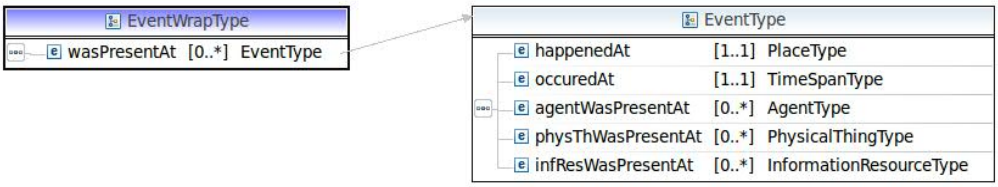


Figure 3.8 *EventWrapType*

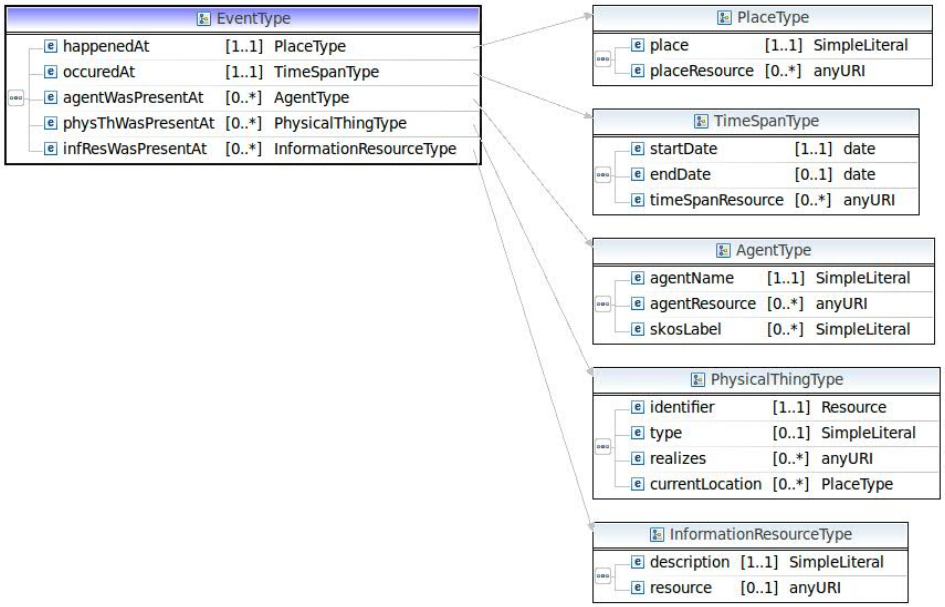


Figure 3.9 *EventType*

EventType includes elements happenedAt and occuredAt that are mandatory for an event and are of PlaceType and TimeSpanType types respectively. Having a Linked on Data (LOD) representation in the back of our minds - especially for the event centric approach - we added elements that are not specified in EDM. Therefore a place is represented by a simple literal while a resource for it, if available, is optional. In a similar manner TimeSpan is represented by the mandatory element start date and the optional end date and resource.

The remaining three optional elements that are included in EventType are special conditions of the wasPresentAt property. More specifically the domain of property wasPresentAt is defined in EDM as the union of class Agent, InformationResource and PhysicalThing. In other words any of them can be present in an event. Therefore each of AgentWasPresentAt, PhThWasPresentAt and InfResWasPresentAt correspond to Agent, InformationResource or PhysicalThing complex types that indicate their presence in an Event. As before, these complex types are represented with a mandatory field, agentName identifier and description repsectivelly, together with other optional elements among them a resource. It is important to mention at this point that since it would be quite rare to have the resource for these complex types (Agent, Place, TimeSpan,

InformationResource, PhysicalThing) and always having in mind a LOD representation, some additional elements may be added in the schema for them. In that way there will be more information about them and therefore a resource discovery operation by searching is specific datasets would be feasible.

The last element that is included in the proxyType is the relatedProxiesType. This compex type was implemented to represent properties defined in EDM that connect two different proxies. Therefore the element RelationType is a string enumeration of these properties that are hasPart, isDerivativeOf, isSuccessorOf, incorporates, isRelatedTo and isNextInSequence while the proxyUri specifies the target-object proxy through its URI.

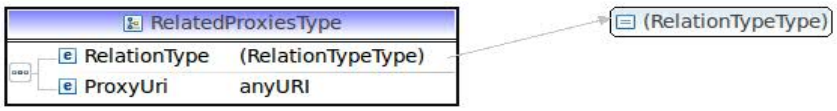


Figure 3.10 *RelatedProxiesType*

Finally, going back to AggregationType (Figure 3.4) from where we started, the two remaining elements are arggregatedCHO that is of PhysicalThingType and webResources of WebWrapperType. AggregatedCHO represents the CHO that is described while the webResources element collects the webResources for the CHO that are the landingpage i.e. the provider's web page about the CHO and the shownAt page that is a web page with the thumbnail of the CHO.

3.1.3 EDM RDF

The purpose of EDM XSD is to collect metadata about CHOs for delivering them to the ECLAP service environment about an organisation's online collections and digital objects. Since it is based on the EDM that is ontology, an XML instance of this XSD can be converted to an RDF instance based on EDM. More specifically as mentioned in the previous section the design methodology, based on which XSD implemented is to create complex types for the classes that include elements which are the properties of EDM. Hence, RDF construction is done by creating an Aggregation resource according to the metadata entered in AggregationType that specify the RDF interconnections. A graphical example of the RDF that is created for MonaLisa from an instance of EDM XSD is illustrated in Figure 3.11.

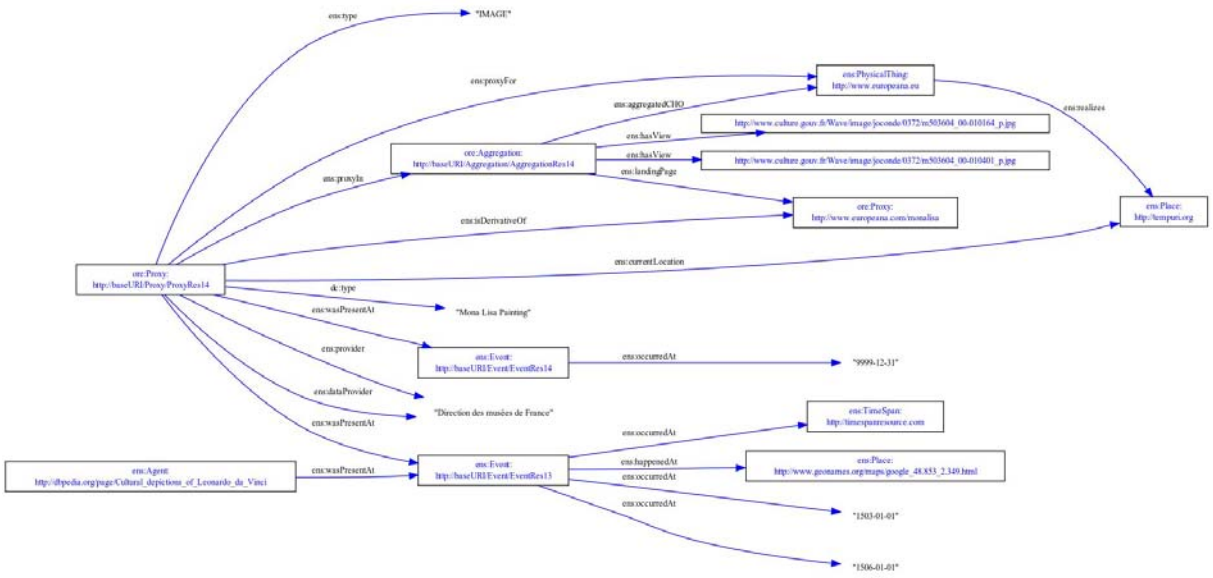


Figure 3.11 *EDM RDF The Mona Lisa example.*

3.2 Interoperability Mappings

3.2.1 ECLAP Metadata Ingestion

ECLAP has setup a platform to offer services for content providers in order to perform ingestion and aggregation of metadata, leveraging the expertise and available resources regarding metadata crosswalks. Due to the nature of a European thematic aggregation there is an expected diversity among participating providers and content. The software tools need to accommodate inexperienced users and legacy data, while taking advantage of the domain experts' knowledge and the project's working groups' results. A tool for visually mapping local metadata schemas to the ECLAP is one that will ensure the success of a large scale aggregation by providing an intuitive, user-friendly approach that reduces the effort to create translation logic for mappings and the turnaround time between human-readable crosswalks and executable code.

3.2.1.1 Content upload

Currently, allowed metadata formats for uploads are:

- XML in any schema.
- zip archives of the above
-

The following methods are supported for uploading content:

- HTTP upload; suggested only for relatively small amounts of data (<2MB)
- Upload to a dedicated FTP server.
- Remote HTTP or FTP browsing.
- OAI-PMH repository harvesting.
- SuperUser uploading from local file system (restricted).

3.2.1.2 User and organisation management

Users belong exclusively to one organization and cannot access data non related to that. Users can be assigned with different levels of access, that grant roles ranging from data browsing, over editing and annotating, to being allowed to edit other users' details (administrators). We have extended user roles to allow parent users, for organizations that might not have expertise or manpower to use the system and thus, delegate the job to an organization which is then their designated "parent" organization. Parent users extend their rights to child organizations and provide the functionality to build the access hierarchy for any given country/thematic category. The current role set can be easily adjusted to allow more freedom in rights management.

The following rights are currently implemented for metadata:

- change/add/delete user
- change/add/delete organization
- edit/upload/delete metadata
- publish / declare finished metadata sets
- read-only browsing rights

These rights have been grouped to the following roles:

- Administrators (all user, data and organization rights for the organizations they manage),
- Annotators (data management rights),
- Publishers (publishing data rights),
- Data Viewers (simple viewing rights).
- The system also contains some hardcoded super-users that have full rights to all organizations and their data in the system.

3.2.1.3 Functionality and user interfaces

Users can join the ECLAP metadata ingestion service using the registration page (Figure 3.12). During registration they are prompted to select the organization they belong to.

Figure 3.12 *Registration Screen of ECLAP Metadata Ingestion Service*

The administrator of that organization is notified by email for the pending user registration and is authorized to grant the appropriate rights and finalize the procedure. In case a user's organization is not present in the list of registered organizations, the user can register in the system without providing one. In this scenario he is given the opportunity to create a new organization and automatically become the administrator for it.

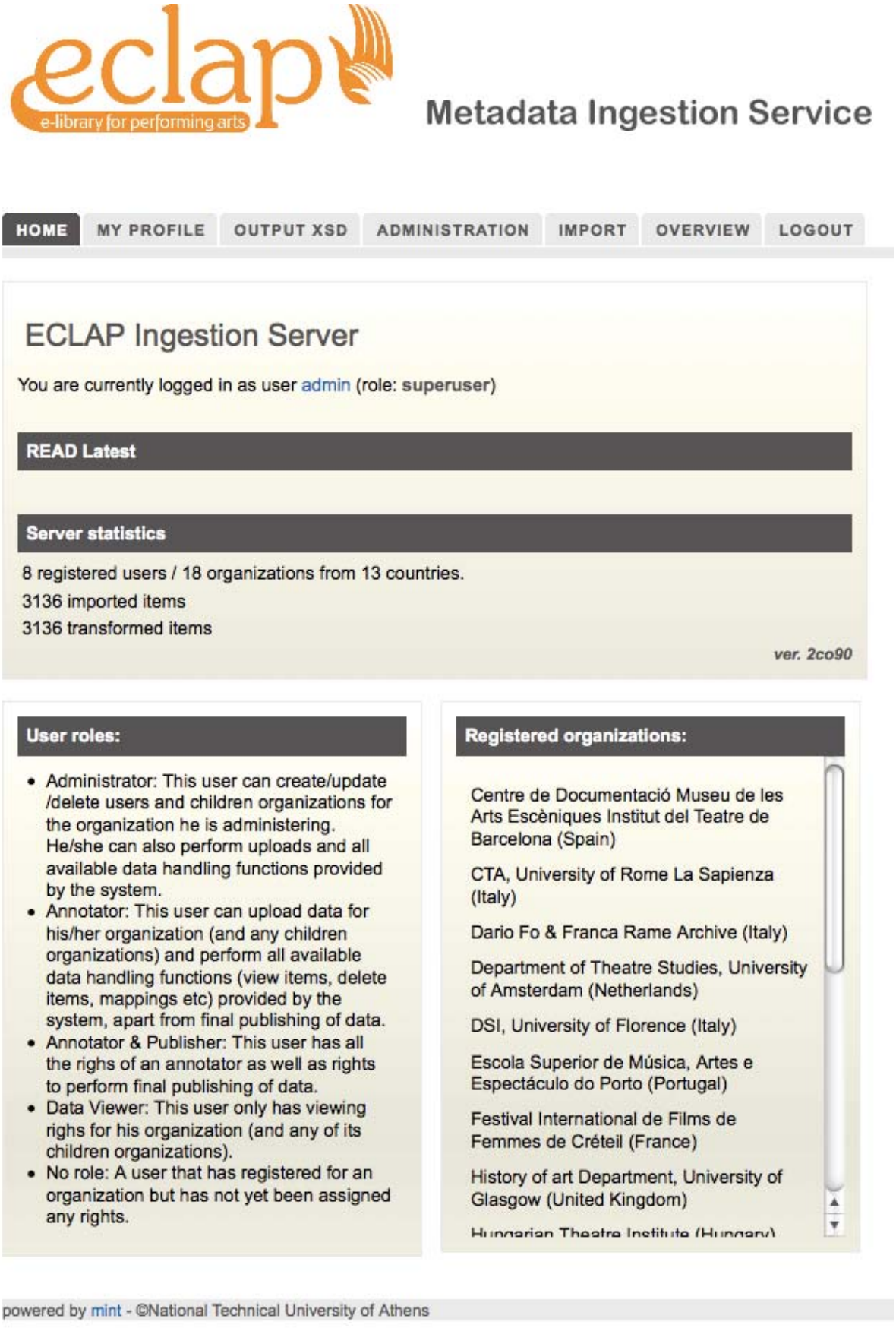


Figure 3.13 *ECLAP Metadata Ingestion Service - Home screen.*

Organizations within the system can have parental organizations. Users of parental organizations extend their rights to the children of the organization (and in turn to grandchildren that may exist, and so on). Every organization can have at most one parent organization. The parent organization has to agree on publishing the data of the child organizations (among other things). This way an aggregator for example can define and manage all the organizations (and their respective data) he is supervising within ECLAP.

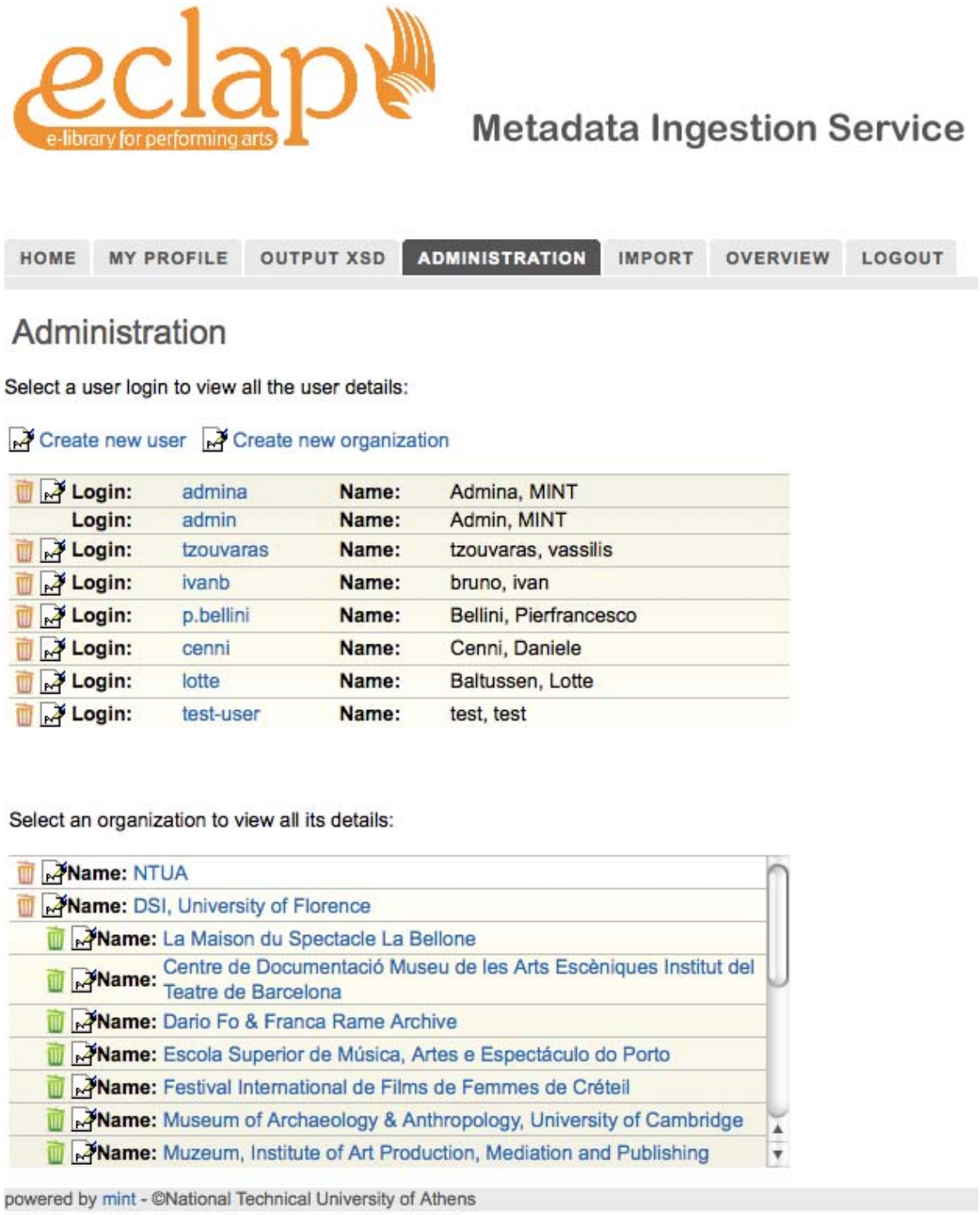


Figure 3.14 *ECLAP Metadata Ingestion Service - User's Administration Screen.*

Every user of the system has the right to see all the other users and data within the same organization (Figure 3.14). S/he can change her own details by using the Profile page (Figure 3.15).

Figure 3.15 *ECLAP Metadata Ingestion Service - User Profile Screen.*

Administrators and Annotators of an organization can upload data using the import page (Figure 3.16).

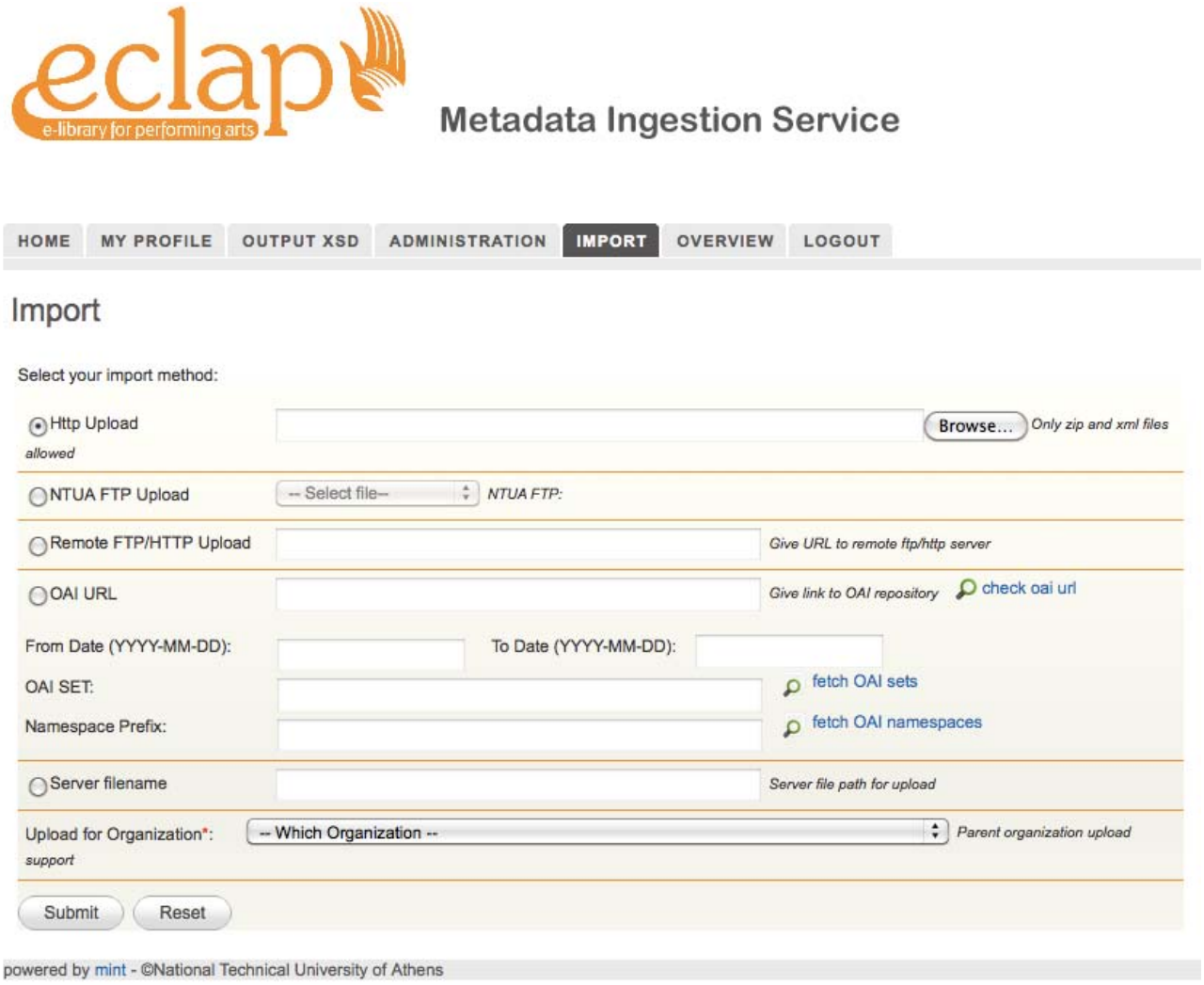


Figure 3.16 *ECLAP Metadata Ingestion Service - Import Interface.*

A history of all uploads for an organization can be browsed using the Overview interface (

Figure **3.17**). Different icons are used to show the current status of an import (hourglass for processing, green arrow for completed, red 'x' for failed).

An import can be deleted, thus deleting all items it contains. After an import process has been completed, mappings have to be defined for the data set in order to see all the items it contains. Every mapping defined for an organization can be saved, edited and reused at a later stage.

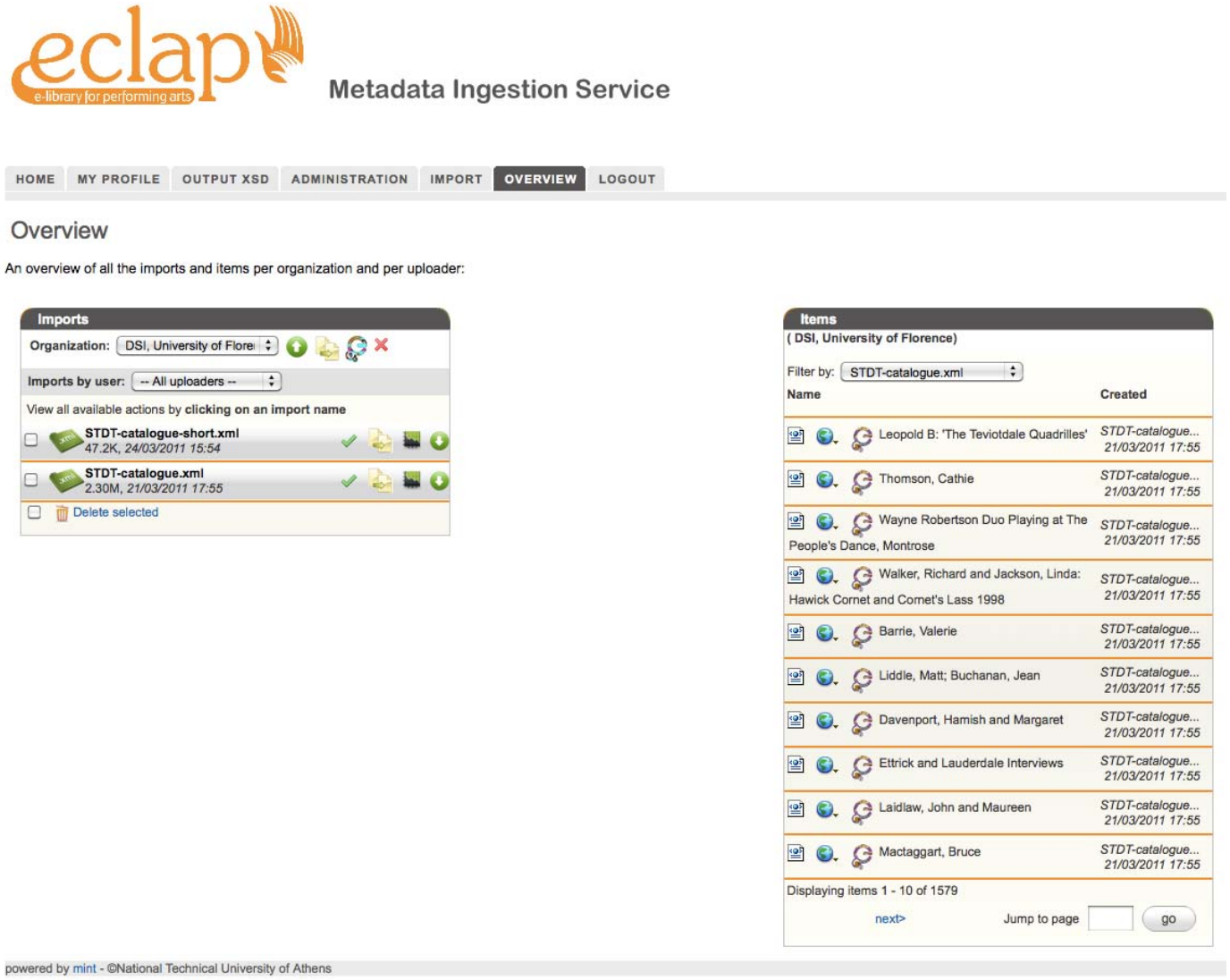


Figure 3.17 *ECLAP Metadata Ingestion Service - Overview Interface.*

In the overview screen the user can browse through items that have been uploaded. Metadata can be uploaded in a single XML containing multiple items or in multiple XML files, each one containing one item. In order to preview the items that belong to each upload, the user has to firstly define the item‘s root element in the XML structure (Figure 3.19) and additionally an element that will serve to label the separated items. The actions that are available to the user as part of the Overview Interface, as presented in

Figure **3.18**, are the following:

- When a “Green” tick appears, near the import name, it indicates that the importing process was successful. If a “Red” cross appears it means that the importing process has failed. In any case if the user hovers the mouse over the icon; various information regarding the process is displayed.
- . This is the Show Items icon. When the user presses this button a new modal window is entered where he/she is able to review the dataset of the import. More details about this functionality in the “Review Original Dataset” chapter.
- . This is the statistics button. When the user presses it a new modal window is rendered where various information regarding the imported dataset is presented. All the different XPaths that were extracted are presented in a tabular form together with statistical information regarding the distribution of the various values of each XPath.

- . The download button. When the user presses it he/she is able to download an archive with all the XML files that are part of the ingested dataset.

Figure 3.18 *How the imports are presented to the user in the "Overview" Tab.*

When the user clicks on the name of the Import an extended view for the current Import is presented as depicted in Figure 3.20. The rest of the mandatory steps that are part of the ingestion tool core workflow are executed from this extended view. More specifically:

- . This button invokes the process for defining the root and label element from the extracted XPaths from the ingested data set.
- . This button executes the transformation of the items.
- . This button invokes the mapping tool in order to perform the semantic mappings between the source and target Schemas and produce an XSLT.

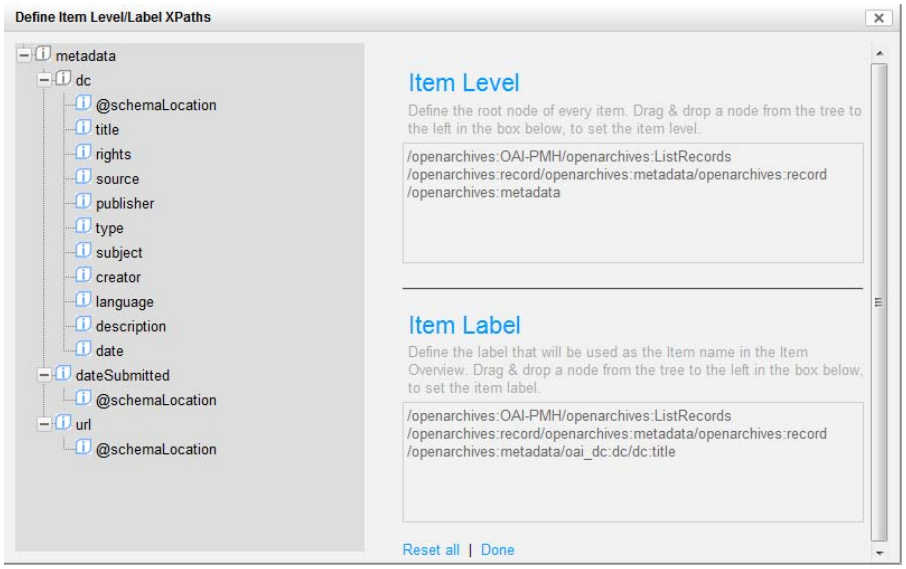


Figure 3.19 *ECLAP Metadata Ingestion Service - Definition of item root element.*

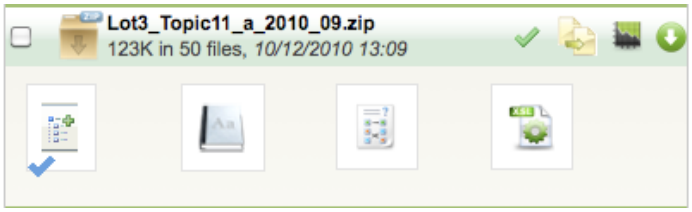


Figure 3.20 *The extended view of an Import in the "Overview" Tab.*

Having defined the item root element, the items that belong to the specific upload can now be previewed in Figure 3.21. Finally, through an icon next to each item, the user views the original XML (Figure 3.22)

Figure 3.21 *ECLAP Metadata Ingestion Service - Items Screen.*

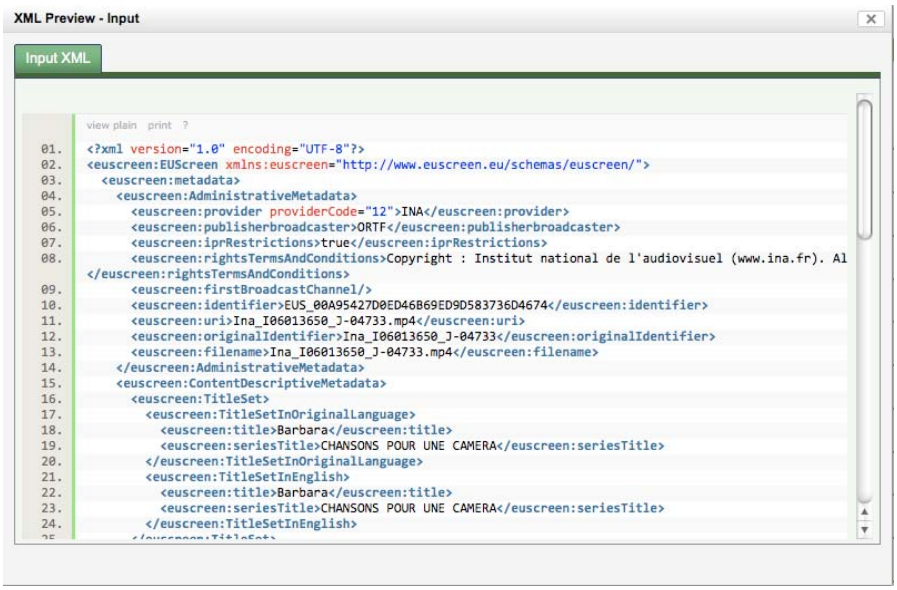


Figure 3.22 *ECLAP Metadata Ingestion Service - Input XML.*

3.2.2 The Mapping Module

The core module of the ECLAP Metadata Ingestion tool is the mapping tool. Although the service shares functionality with many existing metadata repositories, e.g. DSpace and Fedora, one of its main goals it to provide support for a great diversity of metadata schemas or simple data structures, thus widening metadata interoperability. The ECLAP Metadata Ingestion platform aims to be able to store and manipulate metadata that are described using different conceptual models for encoding and decoding information. For this reason both a syntax and semantics have to be defined in order to obtain a complete and expressive model. XML is used as the machine understandable syntax and can be interpreted using different parsers depending on the specified needs. The mapping tool provides the interfaces and mechanisms for identifying and registering through a reference model the semantics of the models used.

Data integration processes comprise of various tasks including data matching, data transformation, and schema/semantic matching. Many solutions have been proposed by the community for each one of those tasks, ranging from applications that rely heavily to the user, to applications that are semi-automatic and in some cases completely automatic depending on the task, thematic category and schema complexity. For the case of schema/semantic matching many techniques and platforms have been developed, enabling the user to complete successfully the task. Notable cases are the schema mapping tool provided by Altova that offers a rich editing environment where the user is able to map any number of arbitrary schemas, but the whole process is totally manual, and the COMA++ platform that offers the user an environment for semi-automatic schema mapping, using state of the art algorithms. Although these approaches attempt to solve the general problem, the case of the ECLAP project has specific characteristics that lead to a more specialized solution to efficiently handle large amounts of diverse data and metadata.

By choosing a semantically rich and well-defined reference schema as the target of the mapping process the user has the opportunity to semantically enrich his data and metadata while quality of the aggregated content is ensured. The mapping process is manual and the tool offers previewing, assisting and validation capabilities in order to ensure the quality of the result. The design principles of the mapping tool ensure the extensibility of the tool itself on a software level and of the system on a data level.

3.2.2.1 Mappings

The metadata mapping tool allows the user to define semantic mappings between the source and target schemas. An XSLT is then generated, based on these mappings that can automatically convert all imported items. An example of the mapping tool is depicted in Figure 3.23.

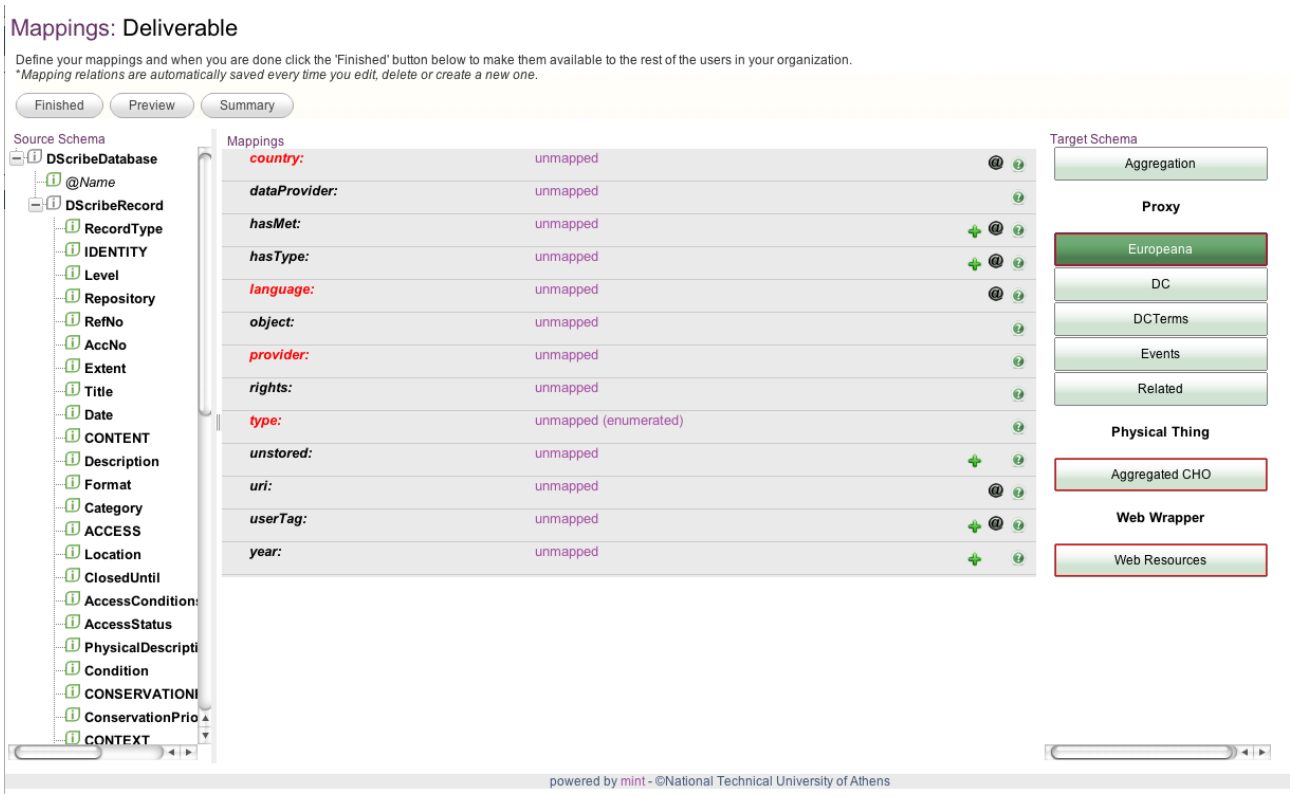


***Figure 3.23** ECLAP Metadata Ingestion Service - Mapping tool.*

Source schema

On the left of the mapping tool is a tree like structure of the source schema similar to the one presented during dataset item level selection. The user can navigate in the schema by clicking the nodes on the left of the tree elements. Elements with a grey information icon (ⓘ) are structural elements and do not contain data values. The rest of the elements are leaf elements with data while elements starting with '@' are attributes of the corresponding father. Elements in blue are elements that have been already used in this mapping. More information about the schema elements is provided by clicking the ⓘ icon. This action invokes a panel as show in the following Figure:

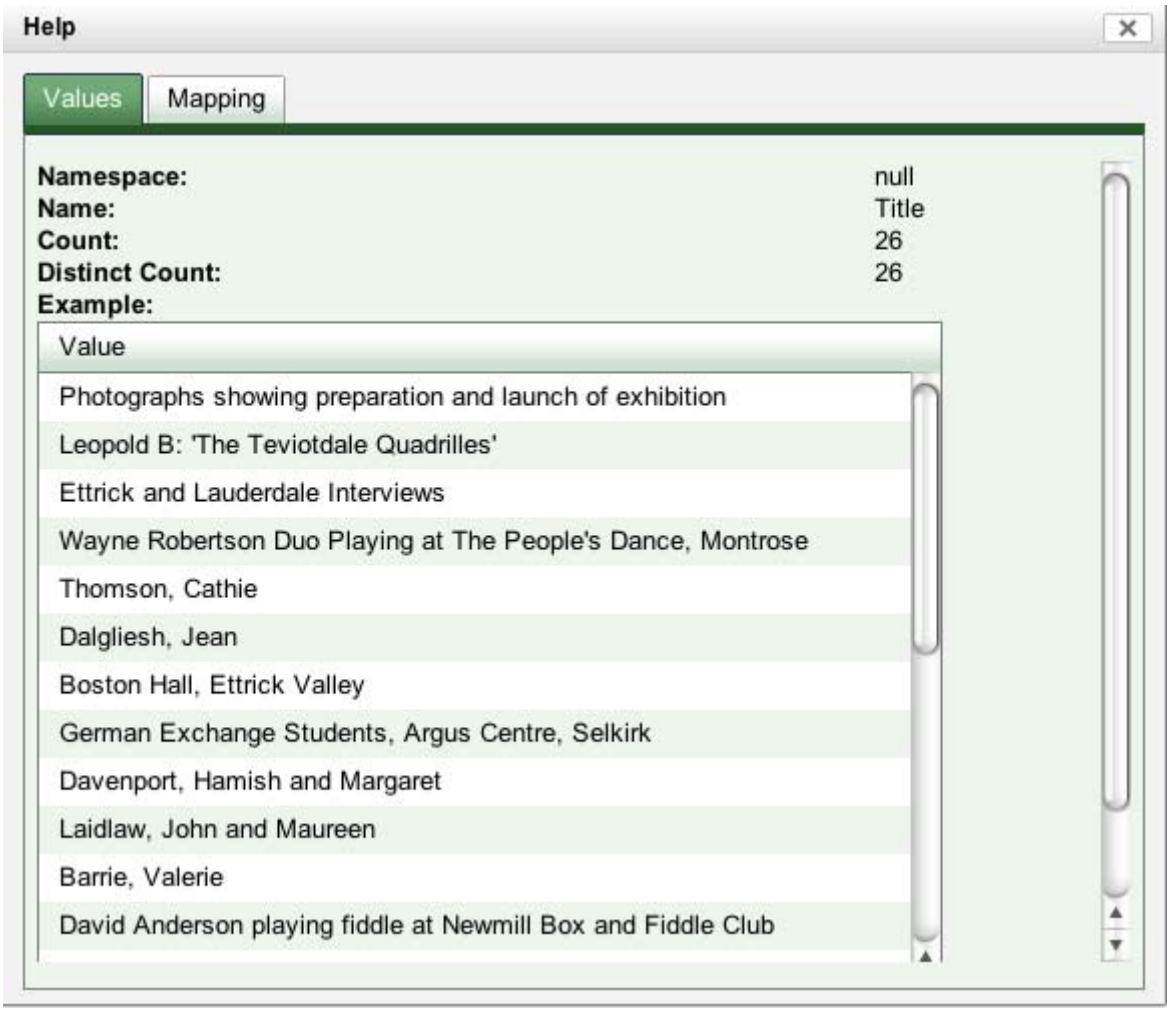


Figure 3.24 *ECLAP Metadata Ingestion Service - Source schema element information panel.*

This panel contains two tabs: Values and Mapping. The Values tab shows the following information about the specific element:

- **Namespace:** The XML namespace to which this element belongs.
- **Name:** The element name.
- **Count:** The number of times the XPath of this element exists in the imported dataset.
- **Distinct Count:** The number of unique values associated with this XPath in the imported dataset.
- **Example:** A sample of these values sorted by their frequency of appearance in the imported dataset.

The Mapping tab shows where and how the specified element is being used in the mapping that is being edited.

Target Schema & Mapping Area

The target XML schema is split into sections that appear as buttons on the right side of the mapping tool. These buttons are used to navigate to specific parts of the target XML schema. By clicking these buttons, the corresponding part is loaded in the middle of the mapping tool along with its specified mappings.

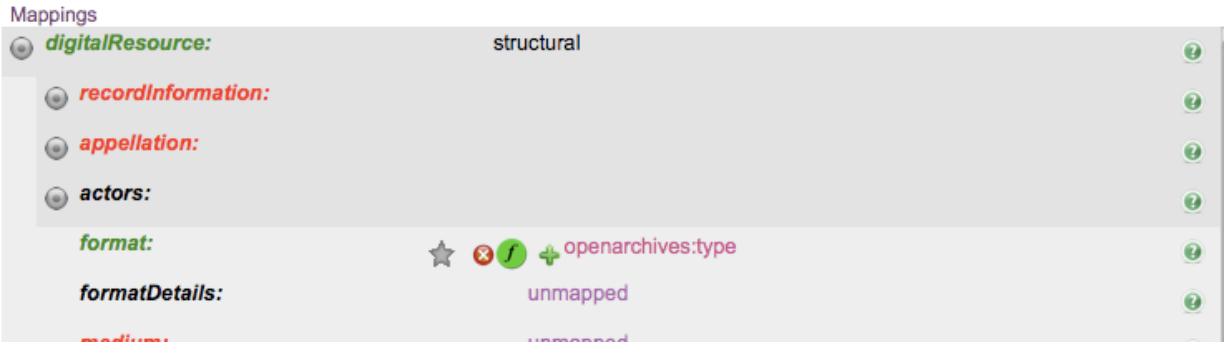


Figure 3.25 *Example of the mapping definitions.*

Each row in the mapping area corresponds to a mapping element of the target xml schema. Rows with grey background are structural elements and contain other elements. They can be expanded and their can be seen by clicking the button to left of their name. The rest of the elements can contain data and have an 'unmapped' area which can be used to defined mappings, as explained later. Elements can also have attributes which can be seen by clicking the @ icon when available. If an element can exist more than one times in the target XML file then a + icon is available on the right of the element row. By clicking this icon an additional row will appear on the mapping area. More information about each element can be provided by clicking the corresponding ? icon on the right of the element row. Elements with green names are elements with mappings or have children with mappings. Elements with red names are mandatory elements that have no defined mappings or have mandatory children with no mappings defined.

Mapping types

Various mapping types are available by using the mapping tool. The most common type of mapping is **XPath mapping**. This type of mapping will copy data from a source element to a target element. To define this kind of mapping, drag n' drop a source element to the 'unmapped' area of a target element. The source element will then appear in the unmapped area of the target element. By clicking the + icon on the left of the source element name, an additional unmapped area will appear for the same target element. More mappings can be performed on this unmapped area and the XML result will contain a concatenation of the provided values. Clicking the corresponding ⊗ icon will remove a defined mapping.

Double clicking on the unmapped area will define a **constant value mapping**. The following panel is invoked:

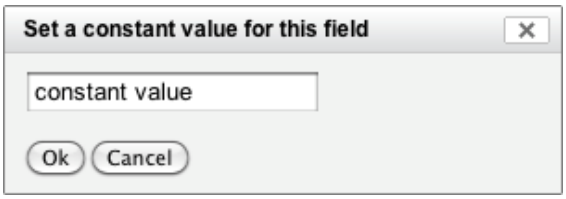


Figure 3.26 *ECLAP Metadata Ingestion Service - Constant Value Panel.*

The user can type a constant value in the provided text field. The value will then appear in the mapping area and in the result XML files. This type of mapping is useful for text that is intended to appear in all transformed items. Constant value mappings can be combined with XPath mappings to construct specific values such as URLs.

Conditions

Mappings can be restricted so that they will apply only under certain conditions. To define these conditions the button is used. This will allow the input of condition as shown in the following Figure:

Figure 3.27 *ECLAP Metadata Ingestion Service - Mapping Condition.*

The conditions supported are in the form of <Source XPath> = <Constant Value>. The corresponding mapping will apply only if the source XPath data for a specific item equals to the constant value provided. The condition source XPath can be set by dragging n' dropping a source element to the condition input area as shown in the previews Figure. The constant value can be set by double clicking on the constant value area.

If a more complex condition is required then the provided condition editor must be used by clicking on the icon next to the condition. An example of the condition editor is shown in the following figure:

Figure 3.28 *ECLAP Metadata Ingestion Service - Condition Editor.*

The condition editor provides the means to define more complex conditions. In addition to each clause's source element and constant value, the relational operator can be set. The logical operator that combines the clauses can also be defined. Additional clauses or subclauses can also be created as needed, by clicking the corresponding icon.

Preview & Mapping Summary

A summary of the defined mappings can be seen by clicking the 'Summary' button on the top of the mapping tool. This will invoke the following panel:

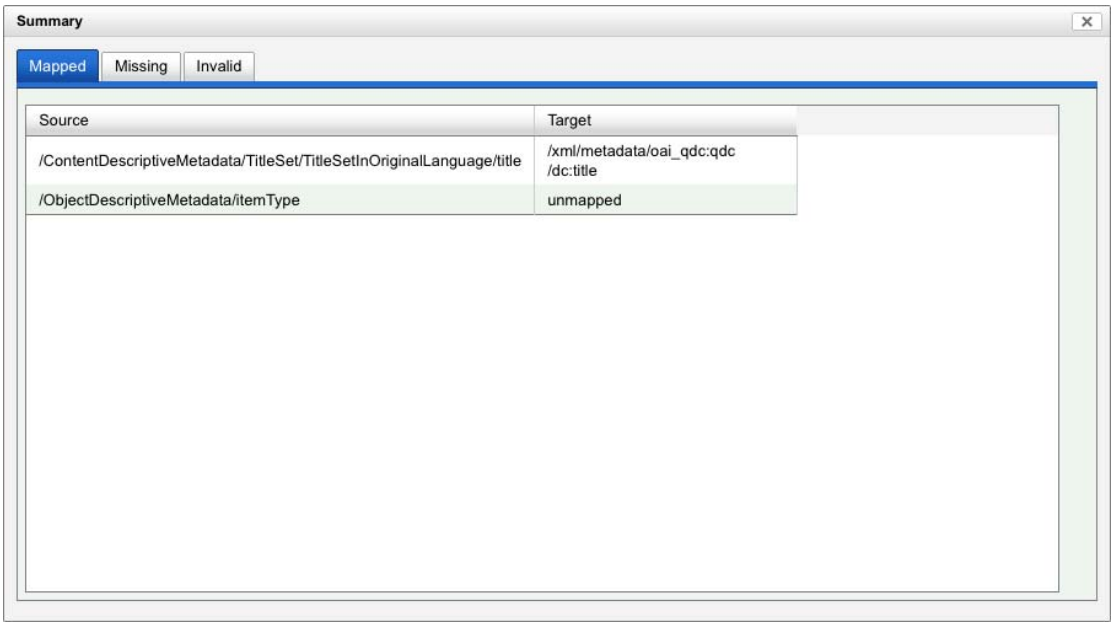


Figure 3.29 *ECLAP Metadata Ingestion Service - Mapping Summary Panel.*

This panel contains the following tabs:

- **Mapped:** All mapped source elements and the corresponding target elements.
- **Missing:** Mandatory target elements that have no mappings.
- **Invalid:** If a mapping definition was loaded based on another dataset, all XPaths form this dataset that do not exist in the current dataset appear in this tab.

The generated XSL can be previewed at any time by clicking the 'Preview' button on the top of the mapping tool. This will invoke the following panel:

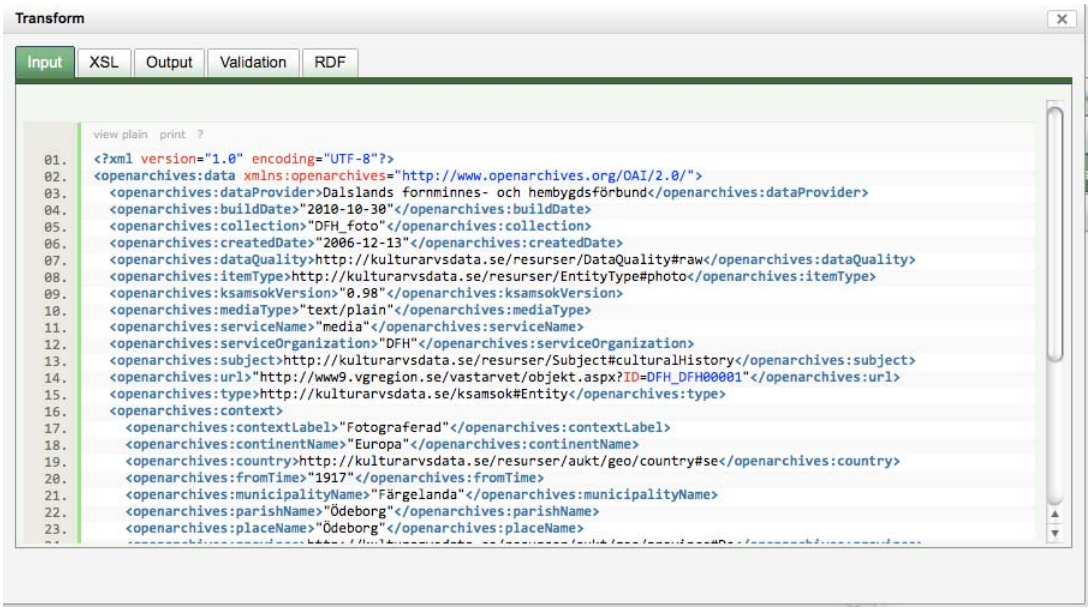


Figure 3.30 *ECLAP Metadata Ingestion Service - Preview Transform Panel.*

This panel contains the following tabs:

- **Input:** The XML that corresponds to the first item from the imported dataset.

- **XSL:** The generated XSL.
- **Output:** The item transformed to the ECLAP schema based on the previous XSL
- **Validation:** Validation information for the transformed item. Any errors from wrong or incomplete mappings will be reported there.
- **RDF:** The output in EDM RDF.

3.2.3 Statistics

The statistics module facilitates the mapping and monitoring procedures. The following figure illustrates the statistics interface.

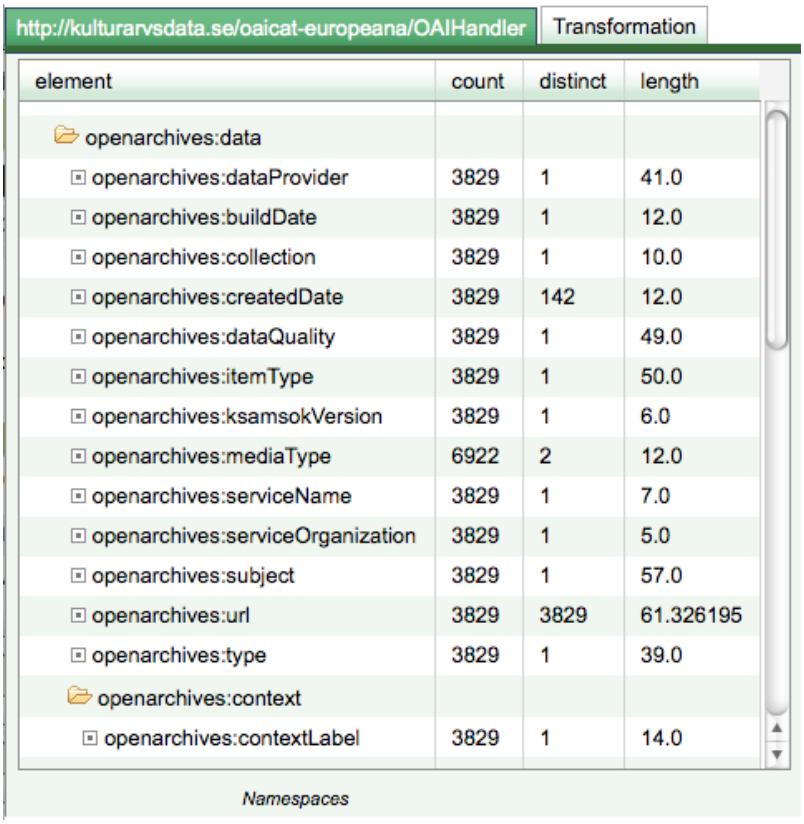


Figure 3.31 *ECLAP Metadata Ingestion Service - Input schema statistics.*

- **Element.** This is the name of the Element or the attribute of an element found in the import and that belongs to a specific XML Schema.
- **Count.** The number of times this element appear in the upload
- **Distinct.** The numbers of distinct/unique values the element or attribute holds.
- **Length.** The average length of the values the element or attribute holds.

Figure 3.32 presents the interface of the values of an element.

- **Value**. The rows under this column contain a distinct value found for a specific element or attribute.
- **Frequency**. The rows under this column contain the frequency of that specific value that appears in the preceding row.

.

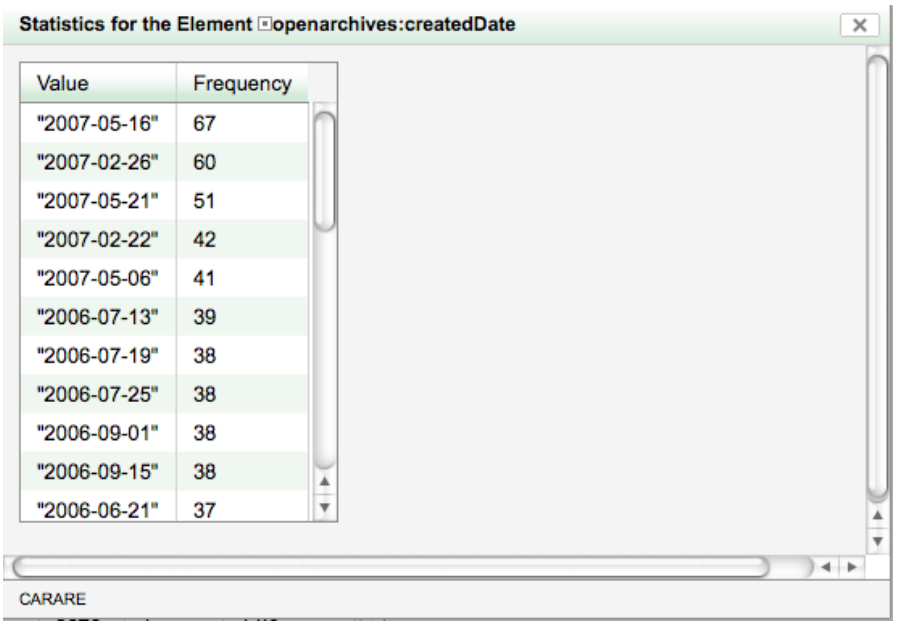


Figure 3.32 *ECLAP Metadata Ingestion Service - Value distribution statistics for a specific element or attribute.*

3.2.4 Transformation Services

3.2.4.1 Transform

When the user has successfully defined the root and label elements for a specific "Import" and the mappings between the extracted source XML Schema and the target Schema of the system, he/she is able to perform the transformation of the data. For this to happen the user has to click the  button which is visible under the extended view of a specific "Import" in the "Overview" tab. When this event is triggered by the user, he/she is presented with the modal window depicted in Figure 3.33. The user is presented with information regarding the different states a mapping might be based on the appropriate Icons. In order for the user to continue the transformation process he/she has to select a mapping from the drop down list and click on the "Submit" button.

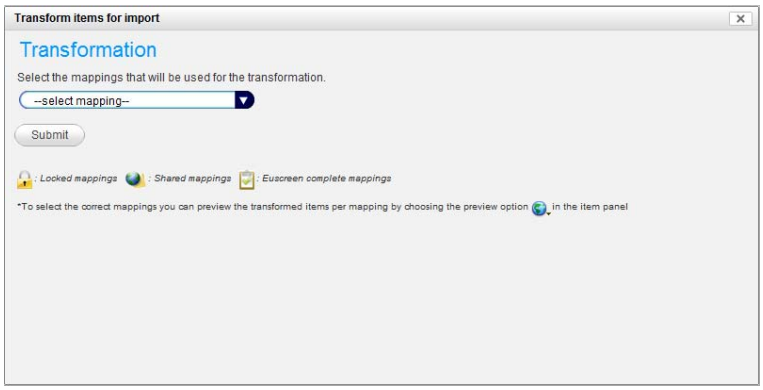


Figure 3.33 *The transformation modal window.*

In the case where the mapping is not correct, for example mandatory mappings are missing, the user is presented with a modal window explaining what the problems are, like the one depicted in Figure 3.34. This modal window has two distinct tabs presenting different kind of information to the user. The first tab

presents any missing mappings to mandatory XPaths of the target Schema while the second one presents XPaths with erroneous mappings. The user is able to review the errors and then he/she has to go back and either select a different mapping or complete/correct the current select one.

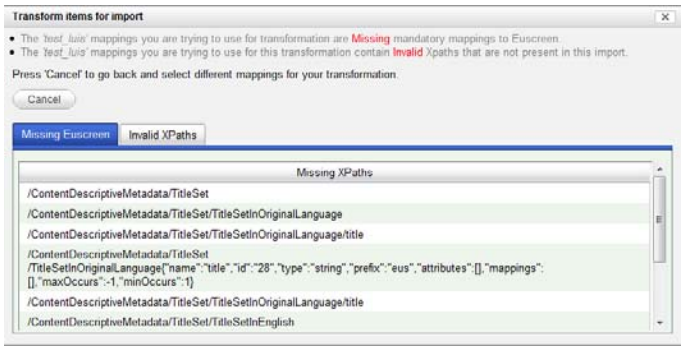


Figure 3.34 *The modal window presenting the errors found in a selected for transformation mapping.*

If everything goes well and the defined mappings do not contain any invalid XPaths or and there are no Missing mandatory mappings the user is redirected back to the "Overview" tab and an animated Icon takes the position of the button. When the user positions the mouse pointer over the animated icon, information regarding the progress of the transformation process is presented to him on a tooltip. Actually, while in the process of transformation, the system extracts each item XML instance based on the root element the user has defined and applies the XSLT transformation that is generated through the process of defining the mappings between the two Schemata, every generated item is then stored to the ingestion tool persistent data layer and is associated with the current Import. In the case where an error occurs in the process of transformation a "Red X" icon appears on top of the icon. When the user positions the mouse pointer on top of the icon he/she is able to review the errors that caused the transformation process to abort. In the case where the transformation ends without errors a "Blue" tick sign appears on top of the "transformation" icon and the user is able if he/she wishes to download the transformed items.

3.2.4.2 Review Transformed Dataset

After the completion of the transformation step in the ingestion tool core workflow, the user is able to review the original data together with the resulted transformed items and the generated XSLT on a per item level through the item browser in the "Review" tab. In order to do that the user has to press the for an individual item in the item browser. When this event occurs the user is prompted with a modal window where he/she is able to review the results of the whole process as depicted in Figure 3.35. The user is able to view the Input XML, the Invalid XPaths if any exist, the XSL generated in the mapping process, the output XML in the target Schema of the system and the output XML of the Publishing Schema.

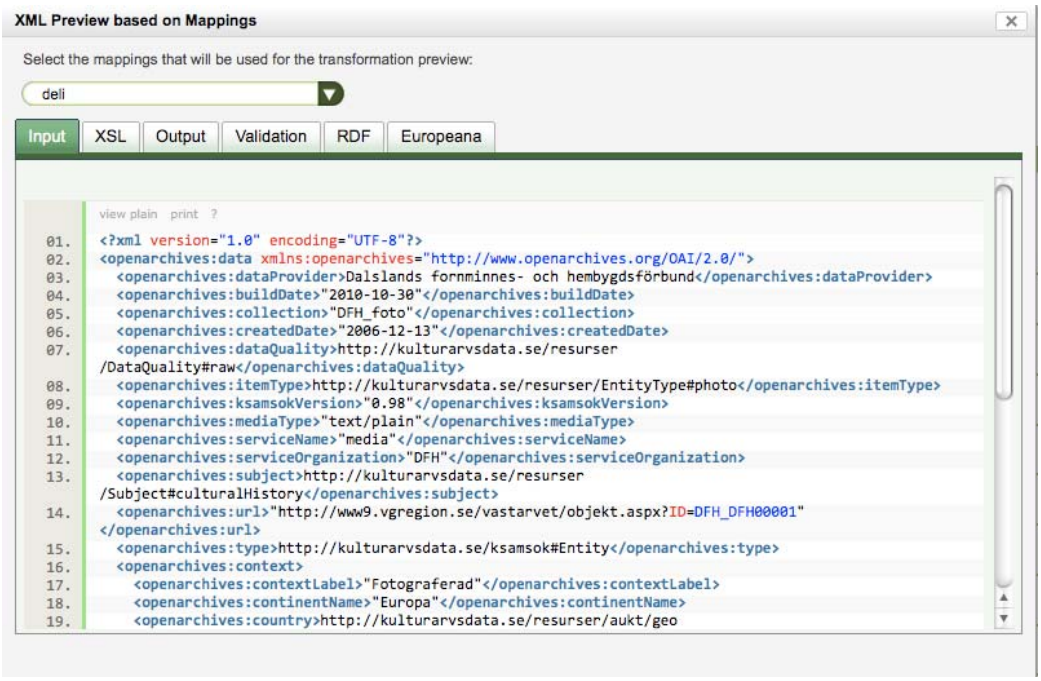


Figure 3.35 *The modal window where the user is able to review the results of a transformation on an individual item.*

3.2.5 Implementation Details

The system is implemented as a web service, where authentication is required to perform a series of tasks that correspond to work flow steps. The service is an application written in the Java programming language and hosted on a web server by the Tomcat servlet engine. Data is imported into a PostgreSQL database in xml format (as BLOB).

Once uploaded, the xml structure is parsed and represented in a relational database table. As this table can grow quite large it is partitioned into one partition per data upload. All data within one upload is treated as having the same structure, so it is not possible to upload different schemas (or more likely updated schemas) in one upload.

Most of the communication between the application and the database is implemented on the Hibernate framework, a high performance object/relational persistence and query service. This allows for powerful, yet simplified, management of housekeeping objects like Users, Organizations and Data Uploads while also providing additional functionalities such as integration with Lucene for indexing and querying data.

Once data is parsed into the relational table, indexes are built to allow quick access to any part or sub-tree of the xml-tree like data. These are currently constructed as PostgreSQL BTREE indexes; when content full text indexing is implemented, it will be based on Hibernate's search architecture. All further data manipulation such as mapping and transformation, normalization, enrichment, etc. is structured through the addition of extra tables annotating but not altering the original data. This allows easier comparison between uploads and facilitates the versioning strategy.

3.2.6 Functional Analysis

The mapping tool is designed using a client – server approach. A subset of the functionalities is implemented as server side services, while the user interface is rendered on the client inside a web browser. The communication between the client and the server is achieved using AJAX calls. One of the core design concepts of the mapping tool is that the user should be able to use all the functionality he might need in order to achieve the best possible result with minimal effort. In order to achieve that, the tool must be intuitive and visual aiding and appealing. Another important design concept is that performance must be ensured because the Ingestion platform must be able to perform computational intense tasks, e.g. metadata transformation and data parsing, without affecting the interaction between the user and the web service. This is achieved in a great degree by separating the interface rendering and the interaction with the user from intense tasks that are

executed on the server side. At the same time the communication overhead between the client and the server was minimized as much as possible.

The XML Schema Parser sub-module is responsible for parsing the target XML Schema and retrieves any valuable information it's stored in its structure, e.g. annotations used for documenting the schema. After parsing the XML schema the sub-module generates an intermediate data structure serialized using the JSON language in order for the user interface to parse and generate the corresponding visual components. The rationale behind choosing JSON as the serialization language for that data structure is the software interoperability the Ingestion platform is attempting to achieve. JSON is a well supported language with interpreters for every major language platform available which reduces the overhead introduced by using XML for exchanging messages both by a reduced memory footprint needed and a simpler structure which makes parsing a much easier and lightweight task. The XML Schema Parser sub-module requests and retrieves the schema needed from the persistent data layer. By using Hibernate for accessing and manipulating the data model, the software's architecture ensures the platform neutrality and separates the maintenance of the sub-modules from that of the data model itself.

The XML Schema parser sub-module is based on the XML Schema Object Model (XSOM) API that is part of the JAXB API for XML data binding. The main design goals of the XSOM API are a) to expose all the defined in the schema spec and b) to provide additional methods that help simplifying client applications. XSOM consists of roughly three parts; the first part is the public interface the entire functionality of XSOM is exposed by this interface to the client. The second part is the actual implementation of these interfaces. Finally the third part is a parser that reads XML representation of XML Schema and builds the XSOM data model accordingly. This part of the code is mainly generated by the RelaxNGCC API.

For the needs of the ECLAP Metadata Ingestion service, an import is not required to include the schema used. This simplifies the actual work for the user and at the same time the set of schema components that have to be mapped is reduced to only those that are used, thus reducing redundancy. The Schema Generator sub-module produces the required simplified version of the schema that corresponds to a specific import by the user. When a user triggers the invocation of the mapping tool for a specific import, this sub-module is also invoked. It communicates with the data layer using the Hibernate persistent API. The next step in the workflow of the Schema Generator sub-module is to parse the data for a specific import and generate a tree like structure using HTML elements that represents the schema used. This tree like structure is then transmitted to the User Interface sub-module and is enhanced using JavaScript in order to create an interactive tree that represents a snapshot of the XML schema that the user is going to use as input for the mapping process.

The User Interface sub-module is responsible for creating and presenting an intuitive and visual appealing environment for the user to define mappings, without sacrificing any of the functionality needed to properly achieve the task of schema mapping. This sub-module is invoked by the user through the Overview interface of the Ingestion platform per import listed there. When the invocation occurs the server retrieves the id of the import and the workflow of the mapping tool is executed; the final step of that workflow is the transmission of all the appropriate structures to the user's browser where the mapping tool is rendered. The User Interface sub-module is implemented in JavaScript using the YUI library from Yahoo. The usage of that library for implementing the visual components also ensures cross-browser compatibility.

3.2.7 Related Work

Next we provide an overview of relevant platforms and tools that deal with ingesting, mapping and transforming metadata records as well as with enabling permanent access to digital works.

The HP-MIT DSpace Repository project

The DSpace project was initiated in July 2000 as part of the HP-MIT alliance. In 2007 the DSpace foundation was formed as a non-profit organization to provide support to the growing community of institutions that use DSpace. The foundation's mission is to lead the collaborative development of open source software to enable permanent access to digital works.

DSpace is a platform that allows you to capture items in any format – in text, video, audio and data in general with the purpose of distributing it over the web. It indexes the data so users can search and retrieve the items that constitute it. Moreover, another major functionality of DSpace is the ability to preserve the data over long term. It is typically used as an institutional repository supporting the following three roles:

- Facilitate **capture** and **ingest** of materials, including any related metadata.
- Facilitate **easy access** to the materials, both by **listing** and **searching**.
- Facilitate the **long term preservation** of the materials.

Finally, DSpace can be used to store any type of digital medium, e.g. videos, images, data sets, journal papers and others. The overall architecture of DSpace is presented in Figure 3.36.

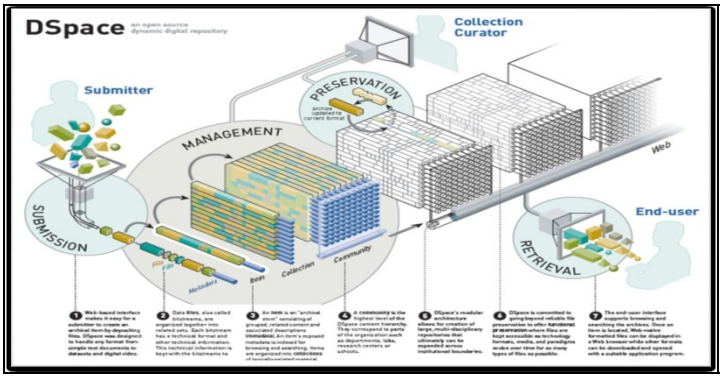


Figure 3.36 *Architecture of the DSpace platform.*

DSpace is designed to work "out of the box" for basic repository needs while still being customizable; the system has been put to a wide variety of uses, and has been entrusted with important intellectual content produced by many institutions. While a certain amount of evolution can take place simply by patches, contributions and reimplementation of specific components the DSpace foundation recognized the necessity of a major review of the core architecture of DSpace, motivated mainly from the widespread adoption of the platform and the technological developments that occurred from the initial design of DSpace back in 2000. For this reason a group of experts and DSpace committers was formed in 2004 that would be responsible of re-evaluating the DSpace platform and propose a set of technical characteristics the DSpace Version 2.0 should have. The outcome of this group can be summarized in the following list of principles. These principles will govern the development of the next version of DSpace.

1. DSpace should be primarily open source software for building digital repositories.
2. DSpace should be usable based purely on free and open source software.
3. DSpace should have a decoupled, stable and application neutral core.
4. DSpace should be usable for a variety of applications but at the same time it will retain useful "out of the box" functionality for common use cases.
5. DSpace should employ and support existing, open standards where possible and practical.
6. DSpace releases should be minimal disruptive.
7. DSpace should support an exit strategy for content.
8. DSpace should evolve.

Based on these design principles the group of experts compiled a list of specific recommendations that would be part of the new version of DSpace. These recommendations attempt to tackle existing issues that appear in the current version of DSpace, e.g. scalability and interoperability among others.

The Fedora Digital Object Repository Management System

The Fedora digital object repository management system is based on the Flexible Extensible Digital Object and Repository Architecture (FEDORA). The system is designed to be a foundation architecture upon which

full featured institutional repositories and other interoperable web based digital libraries can be built. It was jointly developed by the University of Virginia and Cornell University, the system implements the Fedora architecture, adding utilities that facilitate repository management. The current version of the software provides a repository that can handle one million objects efficiently. Subsequent versions of the software will add functionality important for institutional repository implementations, such us policy enforcement, and performance enhancement to support very large repositories. The system's interface comprises three web based services:

1. A management API that defines an interface for administering the repository, including operations necessary for clients to create and maintain digital objects;
2. An access API that facilitates the discovery and dissemination of objects in the repository; and
3. A streamlined version of the access system implemented as an HTTP-enabled web service.

Fedora supports repositories that range in complexity from simple implementations that use the web service's "out of the box" defaults to highly customized and full featured distributed digital repositories.

Another characteristic of Fedora is that since it is a web service, it does not have a standard front end. Nevertheless, many UI applications have been implemented to front-end Fedora by the open source community that supports it.

One of the main strengths of the Fedora framework is that it demonstrates the best scalability among the most used repositories that exist. At the same time it easily supports the storage of multiple types of digital objects and collections particularly well. Another noticeable strength of the platform is that as a foundation architecture that provides powerful API based interoperability features, Fedora is highly flexible and powerful, and has proven itself with large networked repositories similar to those envisaged with the OARINZ project. With no set user interface, Fedora has true separation between the 'backend' and 'frontend'. Fedora provides good interoperability among different systems, with different options allowing for smart and flexible integration methods. Finally, it is supported by a strong development team and development map.

In a sense, a key strength can also be perceived as a weakness. With no user interface, Fedora cannot offer a full repository service 'out of the box' and therefore provides a conceptual complexity which other systems like DSpace do not. The code base of the Fedora platform is probably the largest among the commonly used repositories while at the same time the Fedora development community can be described as closed. These two weaknesses reduce the adoption of the Fedora platform by the repository community.

The EPrints repository platform

The EPrints software has probably the largest and most broadly distributed base of the majority of the repository platforms that exist. It was developed at the University of Southampton and the first version of the system was released in late 2000. The project is supported by JISC, as part of the Open Citation Project and by NSF. EPrints worldwide installed base affords an extensive support network for new implementations. The size of the installed base for EPrints suggests that any institution can get it up and running with minimal effort or technical expertise. Moreover, the number of EPrints installations that have augmented the system's baseline capabilities, for example by integrating advanced search, extended metadata and other features, indicates that the system can be readily modified to meet local requirements.

As already mentioned, the EPrints platform is a good candidate as the repository platform of choice for many institutions because it is one of the least complex systems in existence and hence it has a low skill barrier to implement and maintain. At the same time, because it has one of the widest install bases, it goes a long way to ensure its longevity as a fully supported system. Finally, the code base of EPrints is uniform and well documented making it easier to work on for low level customization.

A major weakness of EPrints lies in the data model used which causes some scalability issues, although these could be addressed with some development effort. Also, its method of adding new digital content type can lead to disparate data models and compatibility issues if maintaining multiple systems. Finally, the development team of EPrints denies any external contribution to the code base of EPrints.

The CERN Document Server Software (CDSware)

The CERN Document Server Software (CDSware) was developed to support the CERN Document Server. The software is maintained and made publicly available by CERN (the European Organization for Nuclear Research) and supports electronic preprint servers, online library catalogs and other web based document repository systems. CERN uses CDSware to manage over 350 collections of data, comprising over 500,000 bibliographic records and 220,000 full text documents, including preprints, journal articles, books and photographs.

CDSware was designed to accommodate the content submission, quality control, and dissemination requirements of multiple research units. Therefore, the system supports multiple workflow processes and multiple collections within a community. The service also includes customization features, including private and public baskets or folders and personalized email alerts.

CDSware was built to handle very large repositories holding disparate types of materials, including multimedia content catalogs, museum object descriptions and confidential and public sets of documents. Each release is tested live under the rigors of the CERN environment before being publicly released.

The CDSware exhibits the following major weaknesses,

- It has extremely complex installation steps.
- CDSware also does not have a good community around it. The mailing list has had very limited traffic since 2002, which indicates that this project may have sustainability issues going forward.
-

DRIVER: Building a sustainable infrastructure of (European) Scientific Repositories

The DRIVER platform which is the outcome of the European funded e-infrastructure project "DRIVER: Building a sustainable infrastructure of (European) Scientific Repositories", does not constitute a repository platform but a framework for creating and managing a network of existing repositories. The main aims and objectives of the Driver platform are the following:

- To organize and build a virtual, European scale network of existing institutional repositories.
- To assess and implement state-of-the-art technology, which manages the physically distributed repositories as one large scale virtual content resource.
- To assess and implement a number of fundamental user services
- To identify, implement and promote a relevant set of standards
- To prepare the future expansion and upgrade of the DR infrastructure across Europe and to ensure the widest possible involvement and exploitation by users.

Version 1.0 of the D-NET Software: Driver network-Evolution-Toolkit is already released under the Apache open source license to the public including the following modules:

- Repository network administration software (such as the Repository Network Manager, Resource Monitoring and others).
- End User services (search, browse, profiling).
- Support service to local repository managers and aggregators (Validation Tool).

The current Driver infrastructure supports three groups of users, A) the repository manager, B) the service provider and C) the researcher, reader, public. Apart from only providing the appropriate technological tools to support the creation and maintenance of a repository network, Driver also defines and supports the concept of the European Community of repository networks. "Community" means that the members agree to some fundamental principles and that the Driver community is wider than the Driver consortium and has no legal restrictions and is open to new members. Some of these fundamental principles are listed below:

1. Make research publications open to the public.
2. Become partner in a repository service network.
3. Follow "guidelines" to make data and services interoperable.
4. Ensure long term access to an institution's research publications.

REPOX – A Metadata Space Manager

Repox is a framework to manage metadata spaces. It comprises several channels to import metadata from data providers, services to transform metadata between different schemas according to user's specified rules, and services to expose the results to the exterior. This tailored version of Repox aims to provide to the TEL partners a simple solution to import, convert and expose their bibliographic data via OAI-PMH, by the following means:

- Cross platform. Repox is developed in Java so it can be deployed in any operating system that has an available Java Virtual Machine.
- Easy deployment. Repox is available with an easy installer, which includes all the required software and libraries.
- Support for several metadata formats. Repox currently supports MARC21, UNIMARC, MarcXchange and MARCXML schemas out of the box and encodings in ISO 2709 (including several variants).
- Metadata crosswalks. It offers crosswalks for converting MARC21 and UNIMARC records to simple Dublin core as also to TEL-AP. A simple user interface makes it possible to customize these crosswalks and create new ones for other formats.

Repox is not a complete repository platform, although it imports metadata and stores them in a custom format for easy access providing at the same time a way of exposing these metadata to the web using an implementation of the OAI-PMH protocol for exchanging metadata over the web. It also includes a mapping tool capable of mapping various input metadata schemas to the TEL format. For this reason, in its current state Repox is limited to support only the exposure of metadata transformed in the format defined and supported by the TEL project.

3.3 ECLAP Metadata Ingestion OAI_PMH Server

As it was described in detail in DE3.1, the following hierarchy exists within ECLAP regarding metadata ingestion, harvesting and mapping:

1. the ECLAP Metadata Ingestion OAI_PMH Server that is used to make the ingested and mapped metadata accessible for the harvesting of the ECLAP Server in the ECLAP Ingestion Schema format.
2. then the ingested metadata in the ECLAP Ingestion Schema format are harvested by the ECLAP Server. Once ingested then they are enriched, aggregated and augmented on the ECLAP Server portal, either automatically or manually by users.
3. then the ECLAP Server uses its own ECLAP OAI_PMH Server to make accessible ECLAP metadata to Europeana.

The following sections refer to ECLAP Metadata Ingestion OAI_PMH Server that makes accessible the ingested and mapped metadata for the harvesting of the ECLAP Server in the ECLAP Ingestion Schema format.

3.3.1 Introducing OAI_PMH

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 [3]. In the context of the ECLAP project, OAI-PMH provides a mechanism for interoperability between the Ingestion Tool and various other modules or platforms. E.g. using the OAI-PMH terminology, the Ingestion Tool constitutes the data provider while Europeana or the ECLAP portal are service providers that are able to request and retrieve metadata records via the OAI-PMH verbs or services.

The ECLAP Metadata Ingestion Tool is capable for managing heterogeneous collections of metadata records while exposing services for mapping and transforming from one metadata schema to another. In order to

extend the functionalities of the ECLAP Metadata Ingestion Tool with the OAI-PMH protocol and thus to expose Metadata through an interoperable mechanism, someone has to implement the defined OAI-PMH verbs on top of the underlying and domain specific data layer. An issue that arise for the case of the ECLAP Metadata Ingestion Tool is that while being able to manage collections of metadata records, the OAI-PMH verbs operate on an item level, something that makes the implementation of the appropriate verbs, directly on top of the Ingestion Tool data layer a difficult task. For this reason and also because it is desired to also include a set of other functionalities that are not directly associated with both the Ingestion Tool and the OAI-PMH protocol, it was decided to follow a technical approach in which an exporting mechanism exist between the Ingestion Tool platform and another data repository more suitable for the needs of an OAI-PMH data repository. The proposed architecture is depicted in Figure 3.37.

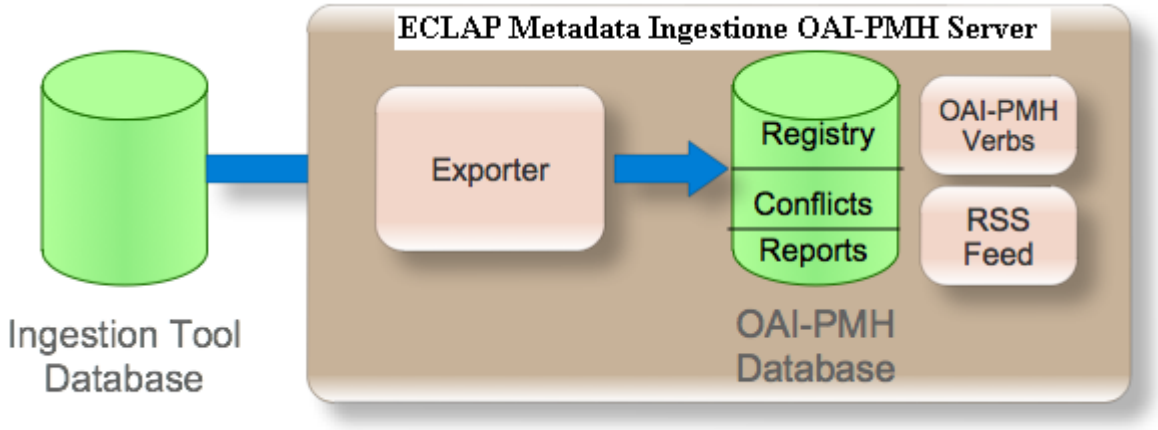


Figure 3.37 *The ECLAP Metadata Ingestion OAI-PMH Server*

Based on the architecture presented in Figure 3.37 the ECLAP OAI-PMH Platform consists of an appropriate OAI-PMH database together with a set of services, these are the following:

- Exporter: a mechanism that is responsible for iterating the imports that are stored in the Ingestion Tool database, transform them in the appropriate format of the OAI-PMH database and store them.
- OAI-PMH database: constitutes the data layer of the ECLAP Metadata Ingestion OAI_PMH Server where the metadata records that will be exposed via the OAI-PMH verbs are stored.
- OAI-PMH Verbs: the various verbs that are defined by the OAI-PMH v2.0 protocol for exposing metadata records through HTTP are implement in such a way that expose directly and without transformation the metadata records that are stored in the OAI-PMH database.
- RSS Feed: an implementation of the Atom protocol for exposing automatically information to subscribers regarding the activity of the ECLAP Metadata Ingestion OAI_PMH Server. E.g. every time a new data set is imported a notification is sent through the RSS Feed to all the subscribers.

In the following chapter, a more detailed technical description of the various modules of the ECLAP Metadata Ingestion OAI_PMH Server implementation will be presented.

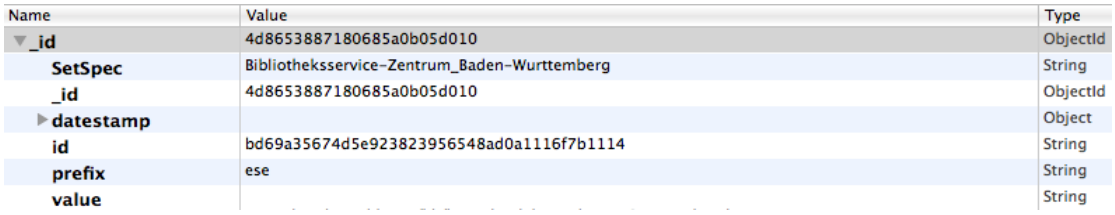
3.3.2 Analysis of the ECLAP Metadata Ingestion OAI_PMH Server Architecture and Implementation

A core characteristic of the ECLAP Metadata Ingestion Tool is that of being agnostic in regard to the Schema of an imported dataset, a characteristic that is also inherited to the ECLAP Metadata Ingestion OAI_PMH Server. For this to be achieved the data layer of the platform has to be able to handle heterogeneous metadata Schemata. Based on this assumption, it was decided to implement the underlying data layer using a NoSQL solution that does not enforce any particular schema and thus it will be able to

adapt to metadata records that conform to different Schemata. The NoSQL solution that was used for the ECLAP Metadata Ingestion OAI_PMH Server is the MongoDB [4] document database. MongoDB is designed around the concept of documents that are internally implemented as JSON [5] document and internally stored using the BSON [6] format which stands for a binary serialization of the JSON format. It allows the existence of JSON documents in the same database or even collection having different fields and thus it does not enforce any specific data model schema. Finally it provides a rich set of native implementations of drivers [7] for communicating with the database while the JSON format provides added value in the development of web application because the stored data do not have to be transformed to a different format in order to be consumed by the applications.

The ECLAP Metadata Ingestion OAI_PMH Server Data Layer is designed around three distinct collections that exist inside a MongoDB data base, collections can be perceived as tables of typical SQL databases, although it is not required that each document conforms to a specific datamodel and set of fields. These collections are the following:

1. Registry: in this collection the actual metadata records are stored and accessed by the implemented OAI-PMH verbs.
2. Conflicts: every time a new dataset is imported in the OAI-PMH repository, it is checked for the existence of duplicate records, if any exist, they are reported and stored in a different collection while they are also associated with a specific Report document.
3. Reports: every time an operation occurs on the OAI-PMH Repository, a report is created which includes any usefull information regarding the operation. For example, in the case of import if any conflict is identified between the items it is reported for further reference.
4. The Registry collection constitutes the core collection of documents for the OAI-PMH repository. It contains all the records that it is desired to be exposed via the OAI-PMH repositories. Each record is stored inside a JSON document which also contains various other information that might be useful to the platform. More specifically, the record document contains information regarding the organization to which the item belongs to, a unique hash key that is generated by calculating the SHA1 hash of the string representation of the item, a datestamp which represents the date and time the record was inserted and finally a namespace “prefix” value which is required by the OAI-PMH specification. An example of a Registry Document instance is depicted in Figure 3.38.



Name	Value	Type
▼ _id	4d8653887180685a0b05d010	ObjectId
SetSpec	Bibliotheksservice-Zentrum_Baden-Wurttemberg	String
_id	4d8653887180685a0b05d010	ObjectId
▶ datestamp		Object
id	bd69a35674d5e923823956548ad0a1116f7b1114	String
prefix	ese	String
value		String

Figure 3.38 *Structure of the Registry Document.*

As it was mentioned earlier, another important collection of documents that is stored as part of the OAI-PMH repository is the Reports collection. The documents that are stored in this collection represent a set of valuable information that corresponds to specific actions of the repository platform. These actions are stored as values in the type attribute of the document and take one of the following values:

1. Add: this type represent an addition action in which records are added in the registry.
2. Update: this type represents an update action in which a set for a specific import already exist and it is updated by adding new metadata records.
3. Delete: this type represents the action of deleting a number of records from a specific import that already exists in the registry.

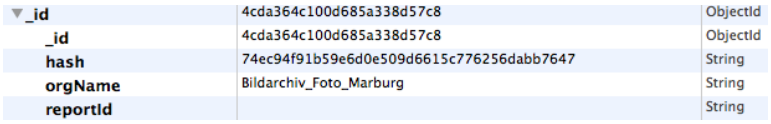
Apart from the action type a number of other values are also stored as part of the Report Document. More specifically, a set of valuable statics are stored; the number of conflicted items that were identified, the total number of the inserted records and the total number of items which corresponds to sum of the inserted records plus the conflicts. Two datestamps are also stored as part of the Report document, one datestamp

corresponds to the time of creation of the document and the other to the time of closing of the document, in this way it is possible for the repository to calculate the time it took to import a whole data set into the database. Finally the date the import was published to the Ingestion Tool is stored together with the name of the organization it belongs to. A visual representation of the Report document structure is depicted in Figure 3.39.

ConflictsNumber	0	Int
InsertedNumber	9576	Int
TotalItems	9576	Int
_id	4cda3279100d685a312b47c8	ObjectId
▼closed		Object
date	2010/11/10	String
time	07:49:53	String
▼created		Object
date	2010/11/10	String
time	07:49:45	String
orgName	Rybinsk_State_History,_Architecture_and_Art_Museum-reserve	String
▼publicationDate		Object
date	2010-07-11	String
time	18:05:30.303	String
type	add	String

Figure 3.39 *Structure of the Report document.*

The last collection of the OAI-PMH repository database is the conflicts collection. The documents stored in this contain any metadata records that at the time of the exporting of an import from the Ingestion Tool to the OAI-PMH repository were found to be conflicted. These documents are quite simple in their structure. They contain an SHA1 hash of the conflicted item that was found together with the record and a reference to the Report document that it belongs to. In this way it is possible for someone to browse the actions that were made on the repository (e.g. additions, deletions and updates) and directly view the items that were found as conflicted for the cases of additions and updates. An example of a conflict document is depicted in Figure 3.40. It has to be noted that the whole procedure of creating unique hash codes and identifying conflicted items is an important functionality of the ECLAP OAI-PMH repository platform because it provides a mechanism for creating unique ids for the metadata records and also a mechanism for identifying duplicate.



▼_id	4cda364c100d685a338d57c8	ObjectId
_id	4cda364c100d685a338d57c8	ObjectId
hash	74ec94f91b59e6d0e509d6615c776256dabb7647	String
orgName	Bildarchiv_Foto_Marburg	String
reportId		String

Figure 3.40 *Structure of the Conflict document*

On top of the data layer that was described so far a set of functionalities is built and more specifically an RSS Feed based on Atom and of course the implementation of the actual OAI-PMH verbs. The implementation of the verbs is based on the customization of the oaicat [8] API which provides an abstract implementation of the OAI-PMH v2.0 specified verbs that can be customized in order to operate on top of different data layer technologies (e.g. flat XML files, relational databases etc.). The verbs that were implemented in order to work with the current OAI-PMH Repository implementation are the following:

- Identify: this verb provides basic information regarding the running instance of the OAI-PMH v2.0 data repository such as, contact details of the admin of the repository, the base url that can be used by a harvester, a sample of an identifier among others. For a complete list of the information provided by the Identify verb someone can refer to the OAI-PMH specification. The information served by this verb does not have to be stored in the underlying data layer but is part of the configuration files of the verbs implementation.
- GetRecord: given the identifier of a record and the desired namespace prefix, this verb fetches from the MongoDB database the corresponsive Metadata record and delivers it as a response to the harvesting client. In the current implementation a query is executed based on the prefix which is part of the Registry Document and the unique id which is generated by the SHA1 hashing of the initial Metadata Record, this query corresponds to an exact match on the database.

- ListIdentifiers: given a set name and a namespace prefix, this verb responds with a list of identifiers of items that correspond to these criteria. The set name is identical to the name of the organization. In this way it is possible to organize the records of the repository around organizations/providers. Again, the namespace prefix is matched with the prefix field of the Registry Document.
- ListRecords: this verb operates in a similar way to the ListIdentifiers with the main difference being that instead of returning only the identifiers, the complete Metadata Record is served.
- ListMetadataFormats: this verb is implemented by aggregating all the unique prefix values that are stored in the data layer by executing an aggregation for uniqueness query on the Registry collection. The resulting response that is served by this verb contains all the unique namespace prefixes that exist in the OAI-PMH repository and can be used for accessing the Metadata Records.
- ListSets: this verb returns a list of all the sets that exist in the OAI-PMH repository. Sets are named after the organizations that provide metadata records to the OAI-PMH repository, in this way it is possible for someone to retrieve only the records that are associated with a specific organization. The way the values are extracted is similar to the ListMetadataFormats verb.

In every case that is needed, the verbs are implemented in such a way that they support paging through the mechanism of resumption tokens as it is defined by the OAI-PMH specification. The number of the returned items is specified through the configuration files of the oaicat running instance. Finally, by being a servlet implementation, the oaicat specific instance can be served through any of the available servlet contains that exist, e.g. tomcat, jboss, jetty etc. Currently it is served via a running Apache tomcat instance.

The RSS feed is implemented following the Atom Syndication Format [9] which is an XML language used for web feeds while the Atom Publishing Protocol (APP) is a simple HTTP based protocol for creating and updating web resources. The purpose of an RSS feed in the current OAI-PMH repository implementation is to provide a mechanism for notifying metadata consumers for the occurrence of specific actions, for example when new items are added or updated to the repository in an automatic way. This is achieved by creating the RSS Feed on top of the Reports collections that was described earlier, in this way everytime a new report is generated the feed is automatically updated and the subscribers are informed for the associated action. The implementation of the RSS Feed service is based on the Apache Abdera project [10]. The goal of the Apache Abdera project is to build a functionally-complete, high performance implementation of the IETF Atom Syndication Format (RFC 4287) and Atom Publishing Protocol (RFC 5023) [11] specifications. The current implementation of the RSS Feed which is based on the Apache Abdera project, is built by communicating directly with the MongoDB based data layer of the OAI-PMH repository, everytime a new Report document is inserted, it is also transformed into the Atom protocol XML representation and published on the Atom Feed that is maintained through the API of Apache abdera. Finally, an Apache Abdera instance can be served via any of the available servlet containers, but in the current implementation it is contained in a Jetty servlet container for performance reasons.

The last module of the OAI-PMH Repository is the exporting API which is responsible for fetching data from the Ingestion Tool, performing any needed processing on the metadata records and storing them into the OAI-PMH Repository data layer. As it was also mentioned in the previous chapter, the requirements of these two platforms are different in the sense that the OAI-PMH Repository operates on a record level while the Ingestion Tool is designed and implemented to operate on a record collection level. For this reason it is necessary to equip the repository with a mechanism capable of dealing with the differences between the two requirements sets. The exporting API is capable of iterating the list of published data sets on the Ingestion Tool and imports them into the OAI-PMH Repository. The workflow that is executed is the following. Initially, the Exporting API iterates all the records that exist on the publication table of the Ingestion Tool. For each publication it checks whether an import to the OAI-PMH Repository exist and if yes then it checks if it needs an update. In any case after it is decided that items from the publication will be imported in the OAI-PMH Repository, it start iterates the records of the publication, processes them and stored them in the MongoDB database and more specifically in the Registry collection. Also, at the initial step of execution of the Exporting API, a new Report Document is created and its fields are updated as long as the processing of the current publication is executed. For each record that is parsed from the publication an SHA1 hash is generated and by querying that data layer of the repository, it is possible to identify if it conflicts with another record that is already stored. In this case, a conflict document is generated and stored in the conflict

Collection. By doing that, it is guaranteed that at the end the OAI-PMH Repository will contain only unique items that will served through the OAI-PMH verbs that were described earlier. Finally, the execution of the Exporting API can be scheduled by using the Quartz scheduler [12] API.

4 Content Selection and Aggregation

4.1 European policy on broadening the online access to creative content and Performing Arts definition

Activities of conservation of our cultural heritage on Performing Arts are already being consolidated through the centuries in libraries and archives repositories all over European countries (National or specialized libraries, private archives, Performing Arts Academies and schools, archives and universities). Nevertheless the European Performing Arts content - dating back from the origin of western culture itself, from Greek theatre onward – considered not at a national scale but within a broader transnational European level, still needs to be contextualised and analysed. Therefore the establishing of an European Digital Library on Performing Arts aggregating content in Europeana - as the one of ECLAP - will be of great help in the ongoing process of renewal of European cultural heritage that the European Commission is undertaking with the so called Digital Agenda for Europe "2015.eu". In this version of the Digital Agenda for Europe, Action 4 "Wide stakeholder debate on further measures to stimulate a European online content market" - a Green Paper on the distribution of audiovisual content and other content is to be expected in 2011 - and Action 5 "Simplifying the distribution of creative content", are of particular interest in order to foster and give a larger horizon to our ECLAP activity on content selection that is now taking its first steps.

One of the European Parliament resolutions on the matter states that: «whereas Europe's cultural and creative industries not only play an essential role in promoting cultural diversity, pluralism of the media and participative democracy in Europe, but also constitute a major engine of sustainable growth and economic recovery in the European Union; whereas particular attention must be paid to cultural and language specificities in the debate on the establishment of a single market in the creative content sector» the European Parliament «Underlines that greater attention must be paid in the new Digital Agenda to the digitisation of, and improving citizens' access to, Europe's unique cultural heritage».

While the European Parliament stresses the importance of the ongoing process of accessing creative content to European citizens, the digitisation of the European collections on Performing Arts - as the one provided by the Content partners of the ECLAP project - is quite a recent phenomena. Consequently content aggregation selection criteria are not yet established in a broader scale, neither officially published by academics. Bibliography related to assessment and clustering of items offered online in digital libraries is lacking or fragmentary and disorganic.

The ECLAP Content partners are very experienced actors in the field of the Performing Arts heritage conservation. Their single repositories are of very different types, collecting from photos to audiovisual to flyers and posters, documenting widely on many different kinds of Performing Arts. Subject is focused on masters and roots related to theatre, dance, music and live art, showing people and performances played mainly from the beginning of the XX Century to nowadays.

Performing Art Institutions they represent range from Libraries and Institutes (ITB); Universities and Schools (UNIROMA, UG, UCAM, UCLM, UVA, IKP and ESMAE); Authors' catalogues (CTFR); Documentation Centers (La Bellone), Content Providers (B&G); Audiovisual/ Multimedia Festival Archives (FIFF), Theatre Archives and Library Musems (OSZMI, Muzeum). Many of them are involved in the Performing Arts teaching process (i.e. Universities and Schools from Italy, Portugal, Holland, United Kingdom, Spain, Poland) therefore their content is already selected and contextualised according to their teaching needs. Some of them are also production centres (i.e. UNIROMA-CTA and CTFR for Italy, Muzeum for Slovenia) or public archives and libraries (ITB for Spain, OZMI for Hungary) whose documentation activities witness the importance of their work in their country and beyond.

4.2 Content Selection Criteria

If we refer to entry found in dictionaries, we usually define Performing Arts as "forms of creative activity that are performed in front of an audience, such as drama, music, and dance." (Oxford Pocket Dictionary of Current English, 2009) and therefore related mainly to theatre and dance shows or to music concerts.

The collections of our content partners, described in our DoW at pages 25-45 (Table of underlying content), beside several sets of music or dramaturgy's masterpiece, have a consistent and original set of items grouped in genres (like Live Art, Sci-Art, etc.) that broaden the above traditional Performing Arts definition (see examples in Table with Genres at 4.5.2).
The complete set of genres we adopted for our content selection is listed in the Content Survey Results available at http://bpnet.eclap.eu/drupal/?q=en-US/node/3609.
Here is an example of the richness of ECLAP content selection including all dramaturges resulting from grouping that follows the people criteria.

***Table 4.1** Table People- dramaturges*

Dramaturge	**No. of items**	**Geography**	**Content Partner**
William Shakespeare	4234+21+35 +22	UK	UG+ CTA+B&G+UVA
Brendan Behan (1923-1964)	18	Irish	B&G
Personal photos and performance photos of Ferenc Molnár and Jenő Heltai	70	Hungary	OSZMI
Denis Cannan (1919-)	35	UK	B&G
Dario Fo	80.000	Italy	CTFR
Letters of Edward Gordon Craig written to Sándor Heves	30	UK and Hungary	OSZMI
Henrik Ibsen (1828-1906)		Norwegian	UVA and IKP
Anton Chekhov		Russian	UVA and IKP
Aischylos		Greek	UVA
Sophocles		Greek	UVA
Euripides		Greek	UVA
Georg Buchner		German	UVA
Heine Muller		German	UVA
Racine		French	UVA
Herman Heijermans (1864-1924)		Dutch	UVA
Gerardjan Rijnders (1949-)		Dutch	UVA
Adrià Gual Queralt (1872-1943)	362	Spain	ITB
Adolphe Appia (1862-1926)	437	Switzerland	UG
Bertolt Brecht		German	IKP
Anton Chekhov		Russian	IKP
Jean Genet		French	IKP
Witold Gombrowicz		Polish	IKP
Jerzy Grotowski		Polish	IKP
Konrad Swinarski		Polish	IKP
Jerzy Grzegorzewski		Polish	IKP
Krystian Lupa		Polish	IKP
Adam Mickiewicz		Polish	IKP
Sławomir Mrożek (1930-)		Polish	IKP
Tadeusz Różewicz (1921-)		Polish	IKP
Józef Szajna		Polish	IKP
Leon Schiller (1887-1954)		Polish	IKP + UVA
Heiner Mueller		German	IKP
Zygmunt Krasiński (1812-1859)		Polish	IKP
Juliusz Słowacki		Polish	IKP

Witkacy		Polish	IKP
Stanisław Wyspiański		Polish	IKP
Tadeusz Kantor		Polish	IKP
Andrej Rozman Roza (1955-)		Slovenian	MUZEUM
Blaž Lukan (1955-)		Slovenian	MUZEUM
Angélica Liddell (1966-)		Spanish	UCLM
Rodrigo García (1964):		Argentinian	UCLM

This table is just a partial representation of the data gathered, for a complete description; please see the Content Survey results available at Available at http://bpnet.eclap.eu/drupal/?q=en-US/node/3609.

Our content selection includes a series of audiovisual items documenting digital and technological-assisted forms of performance: including body art, video art, sci-art, etc. This kind of content has been selected as it better represents research and avant-garde in the Performing Art developed in Europe over the last decades.

Regarding the selection of contemporary content related to Performing Arts, in our recent content survey (submitted to partners on the 28th of March 2011, please see Annex I), we asked content partners to group their collections following, among the others, a chronological criteria, by time spans of decades (question number 2.3). The aggregation of their answers shows that the majority of our partners' collections date back to the second part of the XX century, as predictable by the audiovisual nature of most of the items. Nevertheless, the oldest ECLAP collections are dated from the last decade of XIX Century, as the UK photos of the University of Cambridge collection about Southern and Eastern roots of the Performing Arts:

Table 4.2 *People - anthropologist and photographers from UCAM collection*

People and time span	**Number of photos**	**Location**
G.I. Jones (1904-1995)	700	South Africa
I.H. N. Evans (1886-1957)	100	Malaysia, Sarawak
William Ridgeway (1853-1926)	50	India, Burma, N. Africa
Margaret Mildward and Sarat Chandra Roy	30	India
A. C. Haddon (1855-1940)	100	Torres Strait, Australian
John Roscoe (1861-missing year of death)	100	Uganda
David Buxton (1910-2003)	20	Ethiopia
Frederick and Margaret Williamson	150	Bhutan, Sikkim (Malaysia Dance, Malaysia Music, Sarawak Dance, Sarawak Music)

This precious content selection of more than 300 old photos on Indian Dance and rituals and other 150 old photos on Indonesian dance is giving to Performing Arts historians the important task and chance to contextualize items in relation to historical cultural events occurring during the same time span or, on the contrary, the possibility to operate cross-cultural comparison through similar events being performed in different time span. For the list of time span in ECLAP collections please see Table 4.7. For example the CTA-Uniroma and University of Glasgow collections offer to ECLAP 125 videos on similar subjects (India and Indonesian dance and rituals), being filmed almost one century later.

Table 4.3 *Eastern dance*

Title and time span	**No. of videos**	**Location**	**Content Partner**
Surendranath Jena (Odissi dance)	4	India	UG

Terukuttu street theatre	3	India Tamil Nadu	UNIROMA
Kathakali dance drama, Kolyiattam dance drama	34	India Kerala	UNIROMA
Nagamandalam snake trance ritual, Chakkyarkuttu dance drama	24	India Kerala	UNIROMA
Karnataka Butha Kola trance ritual	9	India Karnat	UNIROMA
Bengal Charak trance ritual	12	India Bengal	UNIROMA
Trance rituals, Barong trance drama, Rangda trance drama, Jauk dance, kris dance, wayang lemah ritual, kechak dance, children kris dance ritual,	24	Indonesia, Bali	UNIROMA
Jayaprana, Sita	5	Indonesian	UG

This table is just a partial representation of the data gathered. For a complete description, please Content Survey Results available at Available at http://bpnet.eclap.eu/drupal/?q=en-US/node/3609.

4.3 Aggregation Selection Criteria

Aggregation and clustering criteria depends from precise and detailed identification of end users needs and requirements, as established in ECLAP previous deliverables and guidelines. The selection of content in ECLAP is intended as a tool for the identification of content to be aggregated, enriched and valorized according to the DOW and not with the aim of filtering the content access to the service. The selected content would be those that can be better be used as example for the best practice networks and related activities.

The activity of identifying users' requirements is performed by WP2 Continuous Requirements and Scenarios Analysis, coordinated by Beeld and Geluid, with an ongoing monitoring of our user requirements (see DE2.1.1 on User requirements and use cases, released at the end of September 2010 at M3). See also ECLAP WG guidelines, released by Uniroma, and DE2.2.1 on User group set up and maintenance, released at the end of December at M6 by Rinascimento Digitale, listing our potential user group (all documents are available on public access at ECLAP BPNET).

ECLAP users belong to the most various environments: even if at the beginning of the project we stated that our ECLAP activity of content aggregation and clustering would be dedicated mainly to the educational field - i.e. e-learning for Performing Arts students, university professors or scholars, high school education teachers and teaching staff in institutions for professional performing arts training – nevertheless other targets emerged within the world of professional artists, while another important target might be identified among Performing Arts lovers and in the general public of the Internet users: the Performing arts amateurs who would be interested in exploring the ECLAP's content as a leisure activity and entertainment, even if they do not possess any performing arts expertise. Last but not least, one of our end users group is of course related to the community working in the field of the European projects, who will have the possibility to share experience, expertise and knowledge with their peers (ECLAP members and other ICT projects' partners).

Table 4.4 *Target users and their needs for content selection activity*

Target User	**Content needs**	**Content use**
Performing arts professionals	Rare content items and fully expressed metadata to use all over the world, disseminating European creative culture	- To access to pan-European Performing Arts funds and collaborative material, previously unavailable online. -To develop creativity through shared interest and interoperability.
Researchers and scholars	Rare content items and fully expressed metadata to use all	- To get content resource for research and scholarship enabling comparative pan European

	over the world	strategies.
Performing Arts curators	Rare content items and fully expressed metadata to use all over the world	- To get content resource for organizing and curating cultural events. - To enable comparative pan European strategies.
Teachers	Content focusing on Performing Art history and Masters	- To aggregate content coming from several different educational institutions in order to provide them as training aggregated material.
Students	Content focusing on Performing Art history and Masters	- To improve their Performing arts studies through access at moving image, graphic and text resources in order to enhance their learning process. - To access a significant amount of relevant materials and referring to that directly.
Content providers, ECLAP members and ICT projects'partners	Quantity of content to be shared online	- To access at collaborative material. – To use aesthetic and formal knowledge bank for their business. -To develop creativity through shared interest and interoperability.
Performing arts lovers	Popular content for leisure and entertainement	- To access a large quantity of quality materials, including audiovisual contributes, currently very difficult to find. - To access at large collection of material from a unique thematic portal and a simple manner.

Our first concern, in order to perform any content and selection aggregation activity, will be to select from our Content Partners' collections the items which could be of interest for our different kind of viewers, organizing them in different clusters and offering our cultural heritage content beside the border of our content partner local repository.
Our Content Partners are sharing their important repository of Performing Arts of the following countries: Italy, United Kingdom, Spain, Germany, Portugal, The Netherlands, Denmark, France, Slovenia, Belgium and Hungary. Their collections contain also a considerable number of extra European items (See geographical grouping of the Content Survey Results, Table 4.8)
All of the items aggregated in ECLAP will pass through a process of serious semantic filtering and analysis. This process of disaggregation – that is going to be performed for each national collection set - will be essential to organize and re-aggregate a new series of content clusters, each of them expressing the maximum of the collections' potentiality, each of them focusing on the different viewers' needs. To fulfil the different user needs and possibilities of browsing among our content, items grouping criteria will be: people, rarity, geographical location and historical period.
Each cluster will be related to the proposed aggregation criteria and offered online in a way already adapted to users needs, but of course open to further users' annotations and linking (folksonomy, etc.).
Therefore each of the proposed clusters will overcome the single and national collections source and will contain linked data grouping different content partner's set. A table with minimum aggregation criteria for the valorization of our content can be found below.

Table 4.5 *Aggregation criteria*

User	**Criteria**	**Description of criteria**	**Name**
General user for leisure and entertainment	Criteria of the broader interest or popularity	An aggregation criteria could be the popularity of the person related to the item, i.e. Roberto Benigni, Tadeusz Kantor, Jerzy Grotowski, William Shakespeare or Dario Fo,	Masters of the Performing Arts

		etc.	
Academics, professional and artistic world.	Criteria of the rarity and preciousness of the items for professional and creative use.	This criteria aims to valorise the best of our different collection's set, to attract professionals and Performing Arts experts.	The best of ECLAP collections
Students and amateurs	Geographical criteria.	These criteria will emphasize the quantity of items derived by European cultural heritage, comparing them with Performing Arts performed in the rest of the world.	European dance, European music, European theatre, Eastern rituals, American film directors, etc.
Students and amateurs	Historical or chronological criteria.	These criteria will help to build a truly European Performing Arts digital library (transnational).	XX Century decades of European Performing Arts

4.4 Content Contextualization Criteria

The slogan of the Europeana logo "Think Culture" can be read as a programmatic statement meaning that Europeana mission is not just about gathering a great quantity of content, but also about enabling the knowledge pertaining to cultural artefact. As a matter of fact, quantity is not enough: to transform mere information in knowledge, each resource should be contextualised thus becoming a useful part of a specific context. This process might be based on social relations – using the information provided by a person or from a group of people – or semantically based – by relating content to contextual information through common properties in order to build a semantic class of information. Through this process it is possible to derive new information from combined existing knowledge applying a form of interpolative reasoning similar to the formal logic of artificial intelligence applications [23]. In order to align ECLAP's aims with Europeana's ones, a contextualisation process is needed also for our content. We will then proceed on three fronts: organizing content inside the portal, allowing users to add information to the content, linking our metadata to other data already available on the web.

Firstly we will organize our items within the portal both by linking each item showed on ECLAP to the portal of the institution it belongs to (following Europeana's example) - so that users can relate the item to its original context and retrieve additional information - and by relating items between them through hyperlinks in order to show their reciprocal connections and engage the user's curiosity.

Moreover we are going to contextualise ECLAP's content relating each item to similar ones in conceptual clusters (see section 4.5.2 of this deliverable) and placing the collections provided by each partner in time and space. During an ECLAP mailing list discussions (15/09/2010) interesting solutions were suggested by partners on this latter issue: Lotte Belice Balthussen (B&G) proposed a timeline integration showing items thumbnails, such as the Europeana one (see below Figure 4.1) to be combined with faceted search; while Erik Lint (UVA) noticed that a "Google maps strategy" where items could be linked to a map would give a comprehensive overview of "location based performance pictures" (see Figure 4.2).

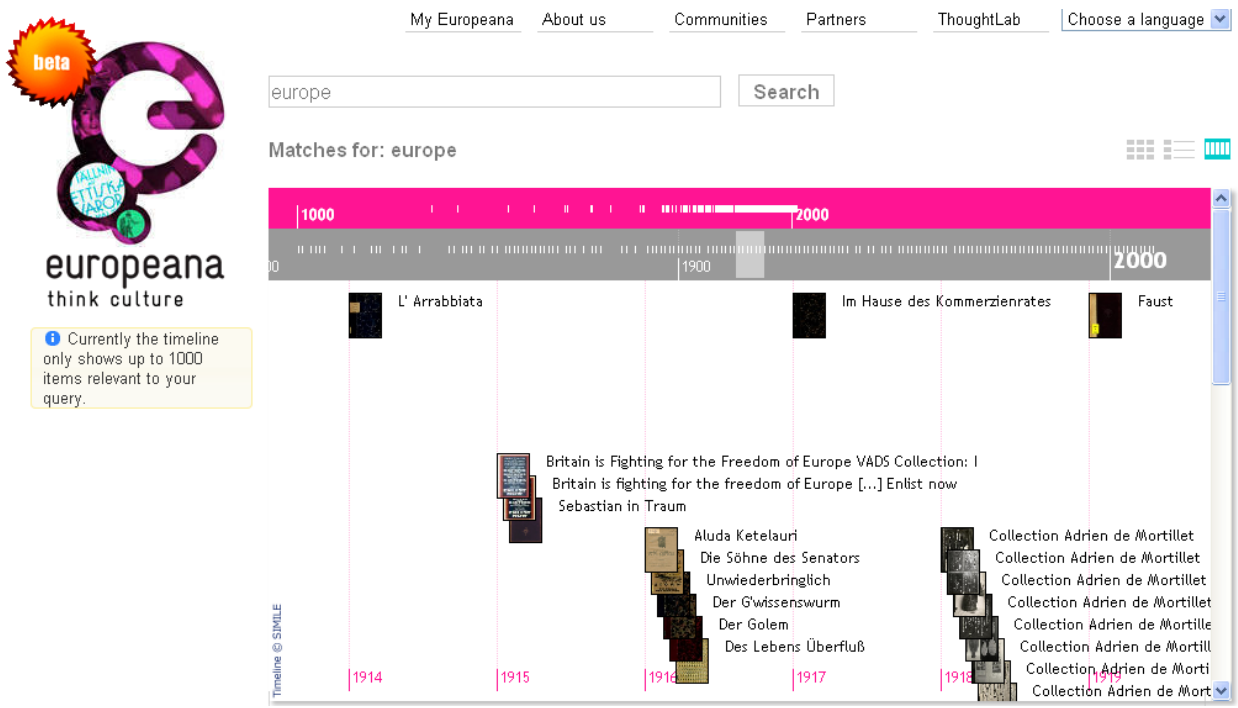


Figure 4.1 *Europeana's timeline from http://www.europeana.eu/portal/timeline.html*

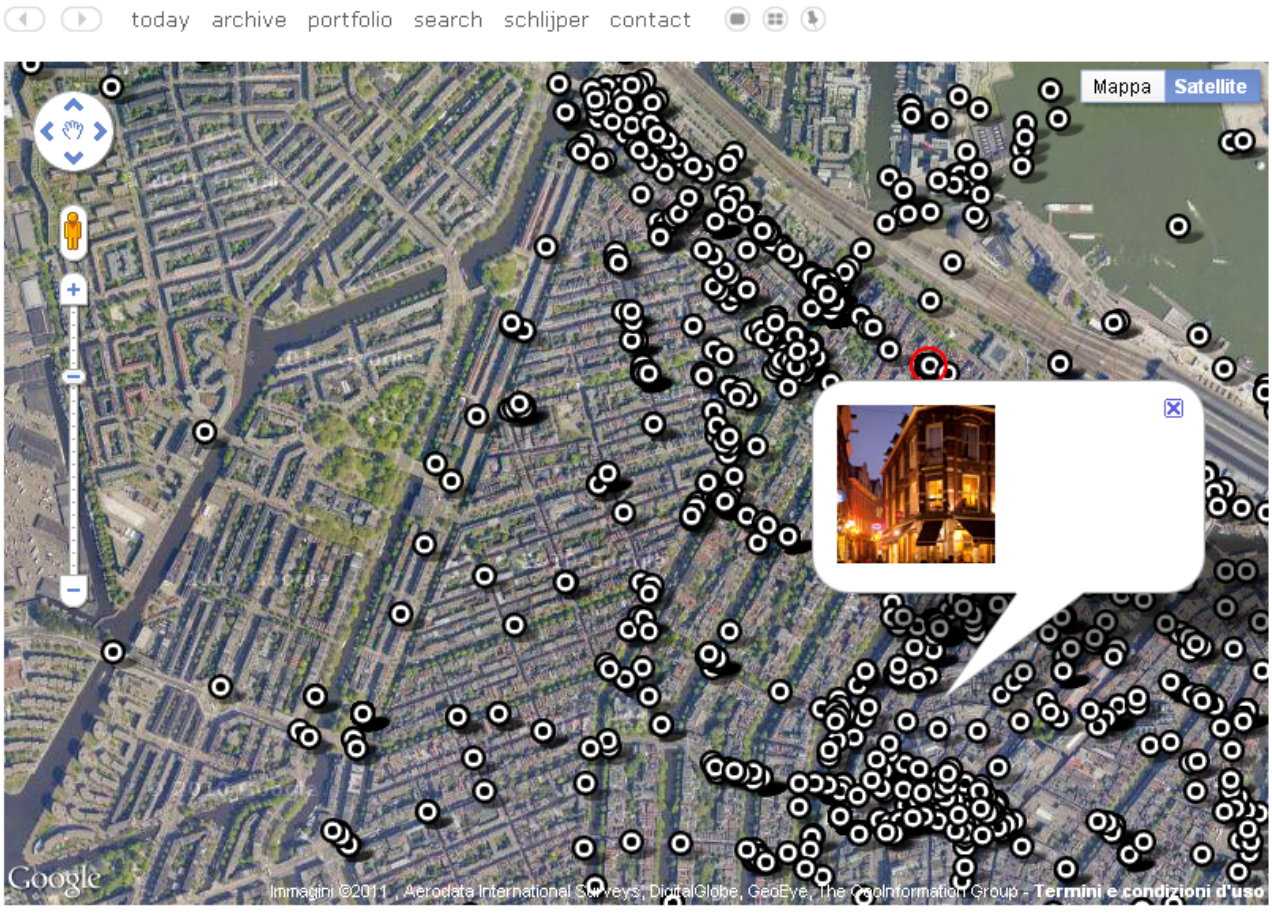


Figure 4.2 *http://schlijper.nl/110415-28-haarlemmerstraat.map*

Secondly, it will be possible for some specific kinds of users, such as the Registered Users who are entitled to play the role of content enricher, to add information to the content by performing different actions such as leaving a set of tags, attaching textual and semantic annotations to them or editing metadata via the ECLAP Metadata Editor. The navigation performed can also be saved/recorded as a personal navigation experience and shared among other users (for more details about actions users may perform to enrich contextualisation please see ECLAP portal at http://bpnet.eclap.eu/drupal/?q=en-US/node/3738, also, please see section 5 of this deliverable);

Finally, the metadata related to the digital objects will include embedded links to contextualisation resources. At this third level, items' metadata could be linked to Linked Open Data (LOD) [1] on the Web or to authority files used within the data supplier's production environment [23]. The figure below shows the data sets that have been published and interlinked by the W3C SWEO Linking Open Data community project so far (September 2010).

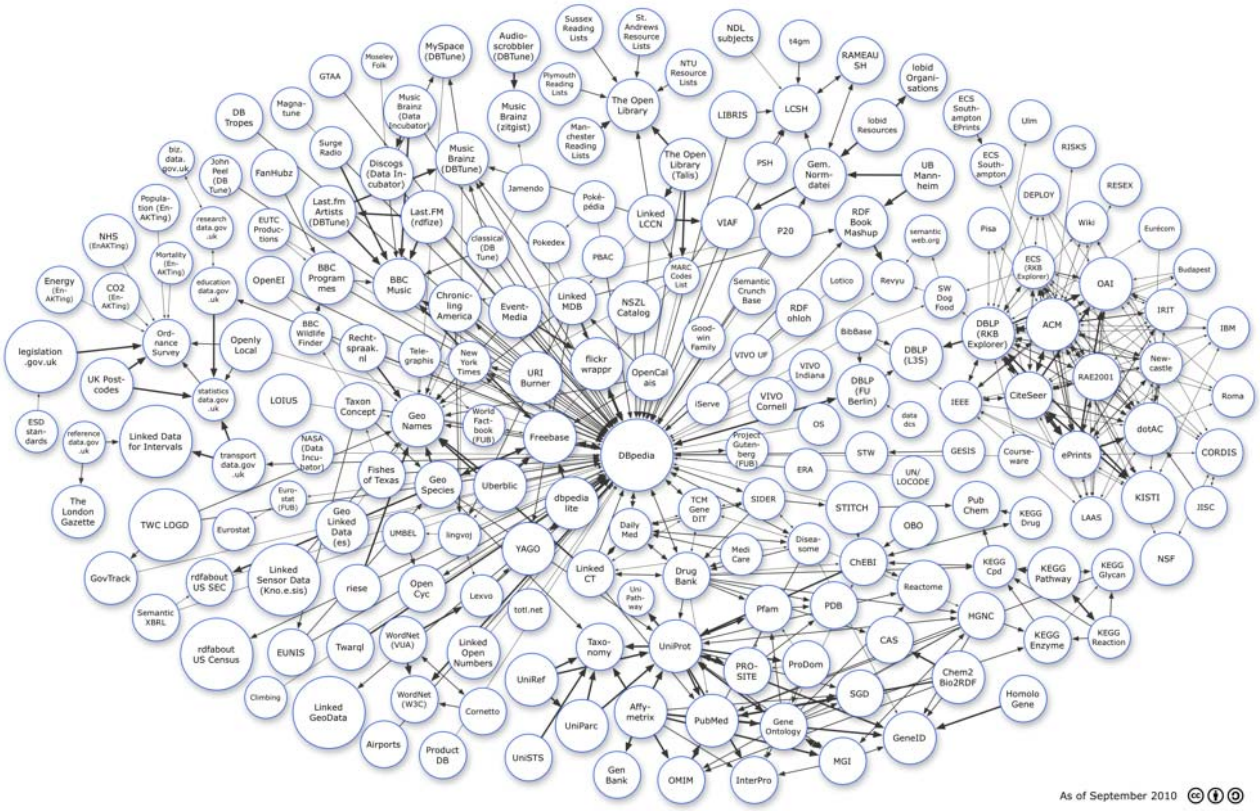


Figure 4.3 *Linked Open Data Cloud*
http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

There are already various interesting open data sets available on the Web, such as Geonames, MusicBrainz, Mo, etc. which is published under Creative Commons or Talis licenses. The figure below, offered by the UMBEL (Upper Mapping and Binding Exchange Layer) project - designed to help content interoperate on the Web - shows some of the class-level interlinking of the data dictionaries (shared vocabularies, schemas, ontologies) associated with the data sets shown above.

[1] These features are a "plus" which may be added towards the end of the projects after fulfilling all the other requirements

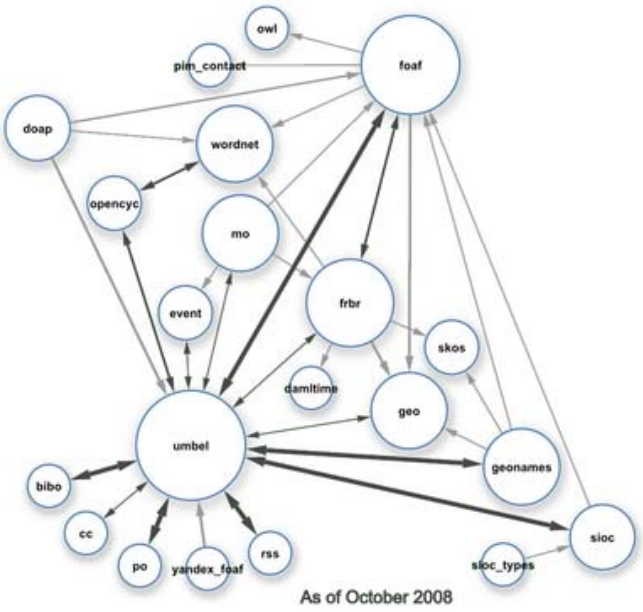


Figure 4.4 *http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData*

As it is stated in the DoW, at a latter stage, we will take into account the content already available trough Europeana in order to organize ECLAP's collections also in relation to it.

4.5 Content Organization

4.5.1 Organization

From the survey results of the Content Survey (see Content Survey results) it appears that ECLAP's content can be divided into five main "Performing Arts Type" framing the complete range of the "Performing Arts" covered by ECLAP:

- Cinema (feature films, short films, animated film, variety show, videoart, new media);
- Dance (including: folkdance, ballet, ethnic);
- Music;
- Theatre (including: opera, amateur performance, puppet theatre, poetry, happenings, rituals);
- Performance Art.

In fact, ECLAP's partners are providing such a great quantity and variety of items that it is fundamental to organize them in order to enhance and exploit the potential of the cultural objects provided while allowing the user to easily and fruitfully explore the ECLAP portal. In order to facilitate the access to content, a strategy that proved to be useful when applied by other websites - which are specialized in cultural heritage and which have been developed by top heritage institutions around the globe [17], [18] and [24] (See Case Studies of Annex III in DE 2.1.1) - is to realize some "overviews" that a user should be able to visualize to get an idea of the portal's content at a glance. To this aim we asked partners to cluster their items according to different categories so that we could investigate the possibilities of content overviews showing the relationships between pieces of information gathered trough the metadata.

4.5.2 Clustering

The goal of clustering is to determine the intrinsic grouping in a set of data through the identification of the possible relationships between data sets and it deals with finding a structure in a collection of unlabeled data [13]. The first step made to organize ECLAP's content is to create conceptual clusters, so that objects are grouped in the same cluster defining a concept that is common to all of them [25]. To this aim we also considered that the criteria adopted in building these conceptual clusters could not be independent from the final aim of the clustering: to suit the user needs. We examined again the results of the Users' Survey (see

DE 2.1.1, Annex I): the answers showed that most target users are interested in Performing Arts genres, followed by historical periods, people and themes and that in order to have a proper choice, the portal should offer all kind of content types. The following table is an example of how we intend to shape ECLAP's content in different homogeneous groups reflecting the users' searching criteria.

Table 4.6 *Genre*

GENRE	PARTNERS' COLLECTION	NUMBER OF ITEMS	INSTITUTION
Ballet			
	Dutch Ballet	37	B&G
	Slovene Modern and Contemporary Ballet	100	MUZEUM
Drama			
	Hungarian drama theatre photos (1860-2010)	4350	OZMI
	Videos of Hungarian drama theatre and opera (1950-2010)	700	OZMI
	Dutch Drama	1108	B&G
	Dutch Modern and Contemporary Drama and Tragedy	100	UVA
	In-yer-face drama in Polish theatre		IKP
	Slovene drama	100	MUZEUM
Puppetry			
	Hungarian puppet theatre photos (1900-2010)	200	OZMI
	Videos of Hungarian puppet theatre	770	OZMI
	Theatre playbills and small prints of Hungarian puppet theatre	300	OZMI
	Reviews of puppet theatre in Hungary	1000	OZMI
	Photographs of Hungarian puppets from the collection of OSZMI (1900-2000)	150	OZMI
	Dutch Puppetry	199	B&G
	Slovene Puppetry	75	MUZEUM
	Catalan Puppetry	400	ITB
Performance Art			
	Programs, leaflets, press kits about Theatre and performance art	1000	BELLONE
	Videos of Theatre and Performance art	530	BELLONE
	Polish performance art		IKP
	Live Art (Performance Art)	40	UCL

This table is just a partial representation of the data gathered, for a complete description; please refer to Content Survey results.

To allow a better contextualisation of the content within the portal, the type of clusters adopted are non-exclusive, overlapping clusters so that an item belonging to a definite cluster could also be included into a another one. In the Content Survey (see Annex I), partners were asked to group their content following different criteria, but they were allowed to put the same item into more than a cluster. For example, the table 4.2.1 shows some of the Content Survey's results for "People" (see Content Survey results) grouped by "Dramaturge" (a category of "Performing Arts Professionals" individuated in DE 2.1.1 [14]), but, even if

different kinds of dramaturges from different countries and institutions merged in this first cluster, they can be found again in the second table which shows some clustering results for "Time Span".

***Table 4.7** Time spans and Dramaturges*

TIME SPAN	DRAMATURGE	INSTITUTION
1870-1890		
	Anton Chekhov	IKP
	Herman Heijermans	UVA
1890-1910		
	Adolphe Appia	UG
	Anton Chekhov	IKP
	Adrià Gual Queralt	ITB
	Herman Heijermans	UVA
	Leon Schiller	IKP + UVA
1910-1930		
	Adolphe Appia	IKP
	Herman Heijermans	UVA
	Adrià Gual Queralt	ITB
	Leon Schiller	IKP + UVA
1930-1950		
	Brendan Behan	B & G
	Denis Cannan	B & G
	Jean Genet	IKP
	Witold Gombrowicz	IKP
	Adrià Gual Queralt	ITB
	Leon Schiller	IKP + UVA
1950-1970		
	Denis Cannan	B & G
	Jean Genet	IKP
	Jerzy Grotowski	IKP
	Tadeusz Kantor,	IKP
1970-1990		
	Jerzy Grotowski, Jerzy Grzegorzewski, Józef Szajna, Tadeusz Kantor, Konrad Swinarski, Sławomir Mrożek, Tadeusz Różewicz,	IKP
	Gerardjan Rijnders	UVA
1990-2010		
	Gerardjan Rijnders	UVA
	Krystian Lupa	IKP

This table is just a partial representation of the data gathered, for a complete description, please refer to Content Survey results.

Where possible, clusters obtained might be hierarchically organized as in the table below where items that are grouped firstly by continent, then by country and finally by genre. The table below show a glimpse of the ECLAP's items, the one related to Performing Arts anthropology. For a general overview, also showing a wide selection of European items, please refer to Content Survey results.

***Table 4.8** Geography and Genre*

INSTITUTION	CONTINENT	COUNTRY	GENRE	ITEM
	Oceania			
UCAM		**Torres Strait Papua New Guinea**	Torres Strait Dance (ethnic), Papua New Guinea Magician Show, Oceania Music (ethnic)	250
UCAM		**New Zealand Torres Strait**	New Zealand multi genre, Torres Strait muti genre	70
	Asia			
		India		
UG			India: Choreography	4
UNIROMA			Tamil Nadu: Terukuttu street theatre	3
UNIROMA			Kathakali dance drama, Kolyiattam dance drama	34
UNIROMA			Kerala: Nagamandalam snake trance ritual, Chakkyarkuttu dance drama	24
UNIROMA			Karnataka Butha Kola trance ritual	9
UNIROMA			Bengal Charak trance ritual,	12
UCAM			India Dance, India Ritual, Burma Dance, Bhutan Dance, Sikkim Dance	400
		Indonesia, Bali		
UNIROMA			trance rituals, Barong trance drama, Rangda trance drama, Jauk dance, kris dance, wayang lemah ritual, kechak dance, children kris dance ritual,	24
UCAM		**Mongolia; Japan**		
			Japan - Butoh dance, interwiews	27
UCAM			Mongolia Ritual, Mongolia Music, Japanese Noh	30
UCAM		**Malaysia**		
UCAM			Malaysia Dance, Malaysia Music, Sarawak Dance, Sarawak Music	100
	Africa			
UCAM		**Nigeria; Uganda; Ethiopia**		
UCAM			Nigeria Dance, Nigeria multi media, Uganda Dance, Ethiopia Dance,	900
UNIROMA		**Somalia:**	Kabebe dance, Training on "Midsumer night's dream", African Magic, Trance rituals, interwiews	44

This table is just a partial representation of the data gathered, for a complete description, please refer to Annex II.

4.5.3 ECLAP's Monographs

As already stated during the ECLAP meeting of 8 September 2010 in Rome, on ECLAP portal elements will be grouped by common attributes and "this collection could be even promoted inside a portal channel, creating subpages, topics, categories that are also indexed" (see the User Requirements Meetings minutes of DE 2.1.1, p. 105). From the questionnaire results it emerged that institutions hold some collections consisting in a relevant number of homogeneous items, but that there are also items, which have been grouped together because of their common properties, that cannot be considered proper collections. These small groups of items, belonging to different institutions, can merge together in the ECLAP portal forming new collections or can be incorporated to other collections to integrate the pre-existing content. This procedure allows us to foresee the creation of possible ECLAP's collections obtainable by drawing together items held by different institutions. For example, (see table below), different institutions hold items about William Shakespeare so that gathering them together means we can build a significant ECLAP's Shakespeare Collection composed by different object types. A possible cluster could therefore consist of videos on Shakespeare performances, texts of his plays and poetry, audio of interviews, photos of costumes, etc. The same could be done for "Dario Fo" or "Jerzy Grotowski" or "Tadeusz Kantor".

Table 4.9 *Shakespeare grouping*

INSTITUTION	PEOPLE	NUMBER OF ITEMS
CTA	William Shakespeare	21
B&G	William Shakespeare	35
UVA	Shakespeare	
UG	William Shakespeare	4234
UG	King Lear (Shakespeare)	22
IKP	William Shakespeare	

This table is just a partial representation of the data gathered, for a complete description, please see Content Survey Results.

4.5.4 Issues to Be Addressed

Since some of the partners need more time to group their content by a specific genre, the "Genre" clusters of Table 4.6 are biased. For example, many grouped their items simply as "dance", which is not a genre but a performing art type, thus "hiding" items that could be inserted in the cluster "ballet" that might contain a major quantity of items than what it seems to contain now. Another problem we had is with partners putting two different genres (for example "drama and tragedy" or "drama and opera") in the same group, thus making unpredictable the quantity of items they will provide per each genre. Nevertheless this problem could be solved by an accurate filling of the metadata schema where appropriate keyword are to be used in order to make it possible to automatically cluster items.

Moreover, the Content Survey results show that the way items are classified by genre changes according to the nationality, because every country (and sometimes every institution) has its own classification: even if a list of genres was already developed in DE 2.1.1 [14], we still need to define a shared vocabulary within ECLAP; while a list of thematic categories or topics (such a "Training", "History of Theatre", "Stage Design") should be developed taking into account the educational context. This issue pertains to the area of classifying and contextualizing performances and should be discussed within WG A – Performing Arts Education and Training in the next Working Groups session (see points 5.3 and 5.4 of "Performing Arts Education and Training: Starting Points" on the WGA ECLAP web page [22]).

5 Content Augmentation and Rich Media

In this section it is presented how the ECLAP portal users can augment the ECLAP content aggregating content (creating playlist, collections and e-learning courses, see section 5.1) and how they can enrich the metadata associated with content providing new information (e.g. comments, free text tagging, taxonomization, translation, votes) or corrections (e.g. correction of metadata translations) (see section 5.2).

ECLAP portal also support submission and playback of interactive rich media content (see section 5.1.4).

5.1 Content Aggregation Tools and Rich Media Content

The ECLAP portal provides tools allowing users to aggregate content: users can create playlists of audio/visual content (see section 5.1.1) to be shared with all the ECLAP users, users can create content collections (see section 5.1.2) allowing them to aggregate content on some perspective (e.g. content related to the same performance, related to the same director, etc.), collections can be shared with other ECLAP users or can be kept personal. Moreover content can be aggregated into e-learning courses (see section 5.1.3) to be used by students, courses can be to be augmented with verification tests, some specific texts.

The ECLAP portal supports the submission and fruition of interactive rich media content (see section 5.1.4) that can be used on PCs and smart phones.

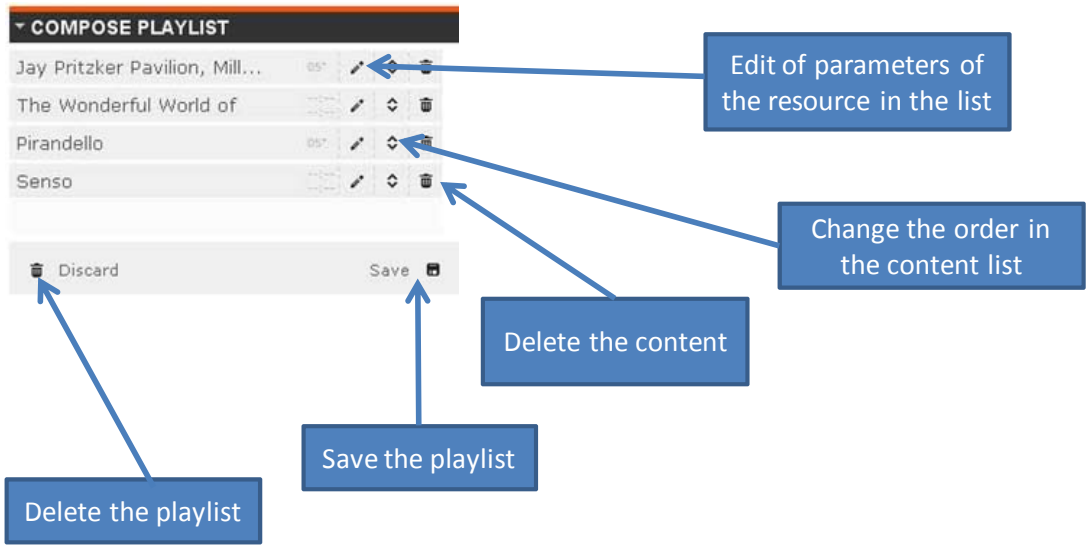
5.1.1 ECLAP Playlists

ECLAP BPNET allows an advanced management of content playlists. By using this functionality a user is able to define a sequence of audio, video and images or select excerpts of audio and video resources arranged into a chained list to be associated to a playlist for a much more effective content consultation, for example to prepare a sequence of content excerpts to be used for educational and promotional purposes. Also, when an image is added to a playlist, the user can set the time duration for the visualization of that image.

5.1.1.1 Creation of a playlist

A user, to create a new playlist has to be registered and logged in the portal.

Below each preview icon of an audio, video or image resource, a button “Add to Playlist” is present. By clicking it, a new box named **Compose playlist** is shown on the right column and the selected title will be added in the list, as shown in the image below.



For each resource added in the playlist, three buttons are present on the right: the button visualizes the resource and allows to set some parameters for the play of the resource; the button allows to modify the content order on the list, and finally the waste bin button to delete the resource from the playlist.

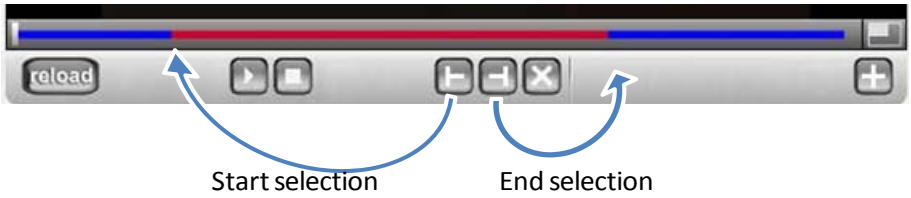
By clicking the button e on the right of a content title in the **Compose playlist** box, the content is played and a set of new command buttons are shown in the play toolbar. These buttons are different according to the resource type played.

In case of a play of an audio or video resource, commands in the play toolbar will be as shown below:

In addition to the typical *reload*, *play* e *stop* buttons, the play toolbar contains also the following:

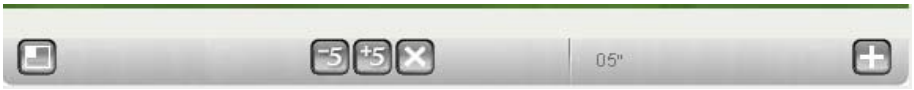
- Set the start of the audio or video excerpt to be included in the playlist (in the form mm:ss)
- Set the end of the audio or video excerpt to be included in the playlist (in the form mm:ss)
- Delete the defined audio or video excerpts
- Add additional excerpts of the same video or audio resources to the playlist.

The audio or video excerpts defined by the user will be highlighted in red in the play toolbar.



In the playback window, if an excerpt of audio or video content has been added, on the right of the content title the start time and the end time are shown, as shown in the following figure.

In the case of images, the commends available in the play toolbar are the following:



The first button on the left allows visualizing the image in full screen. The other buttons are the following:


Decrease of 5 seconds the time of visualization of the image

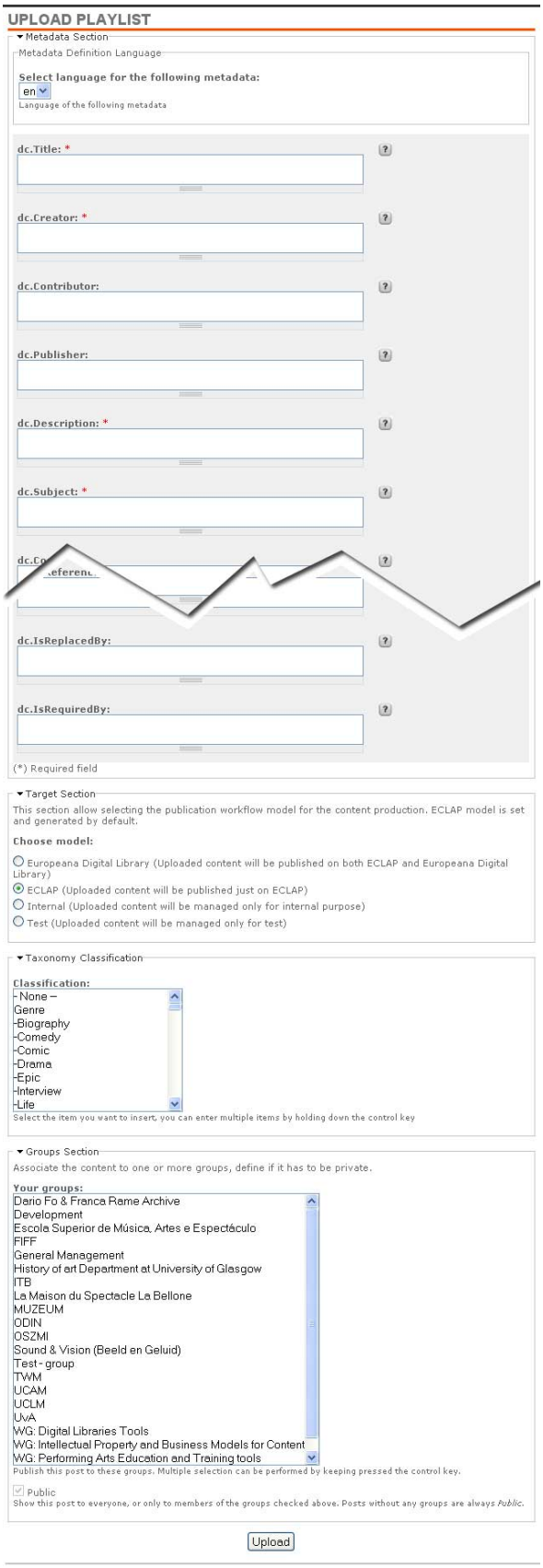

Increase of 5 seconds the time of visualization of the image


Delete the time


Add an new visualization time as new resource in the playlist (it is useful if another visualization of the same image has to be added in the playlist)

Also in the play toolbar, the actual visualization time is shown.

When the procedure of content selection to be included in the playlist has been completed, by clicking the "Save" button a new web page name **Upload Playlist** is shown, that allows specifying additional parameters related to the playlist, as shown in the next figure.



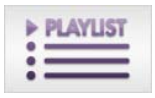
For each *playlist* it is possible to specify mandatory Dublin Core metadata, such as: language, title, creator, description, subject, taxonomy and discussion groups (for multiple selections it is possible to hold the CTRL key down).
Many other not mandatory metadata are available, such as:

- dc.Contributor
- dc.Publisher
- dc.Coverage
- dc.Relation
- dc.Rights
- dc.Source
- dc.Type
- dc.Format
- dc.Identifier
- dc.Date
- dc.Alternative
- dc.Audience
- dc.Created
- dc.DateAccepted
- dc.IsPartOf
- dc.IsReferencedBy
- dc.IsReplacedBy
- dc.IsRequiredBy

For each field an Info button is available that shows a tooltip text when the mouse is explaining the significance of the field according to the Dublin Core standard.
Also it is possible to specify if the playlist is public or accessible only to registered users.
When all information have been completed, by pressing the Upload button the new playlist will be available on the ECLAP BPNET.
A user cannot in any case delete or modify a playlist. To delete a playlist it is necessary to contact the ECLAP Administrator.

5.1.1.2 Visualization of a playlist

Playlists are identified in the ECLAP BPNET by the following preview icon:

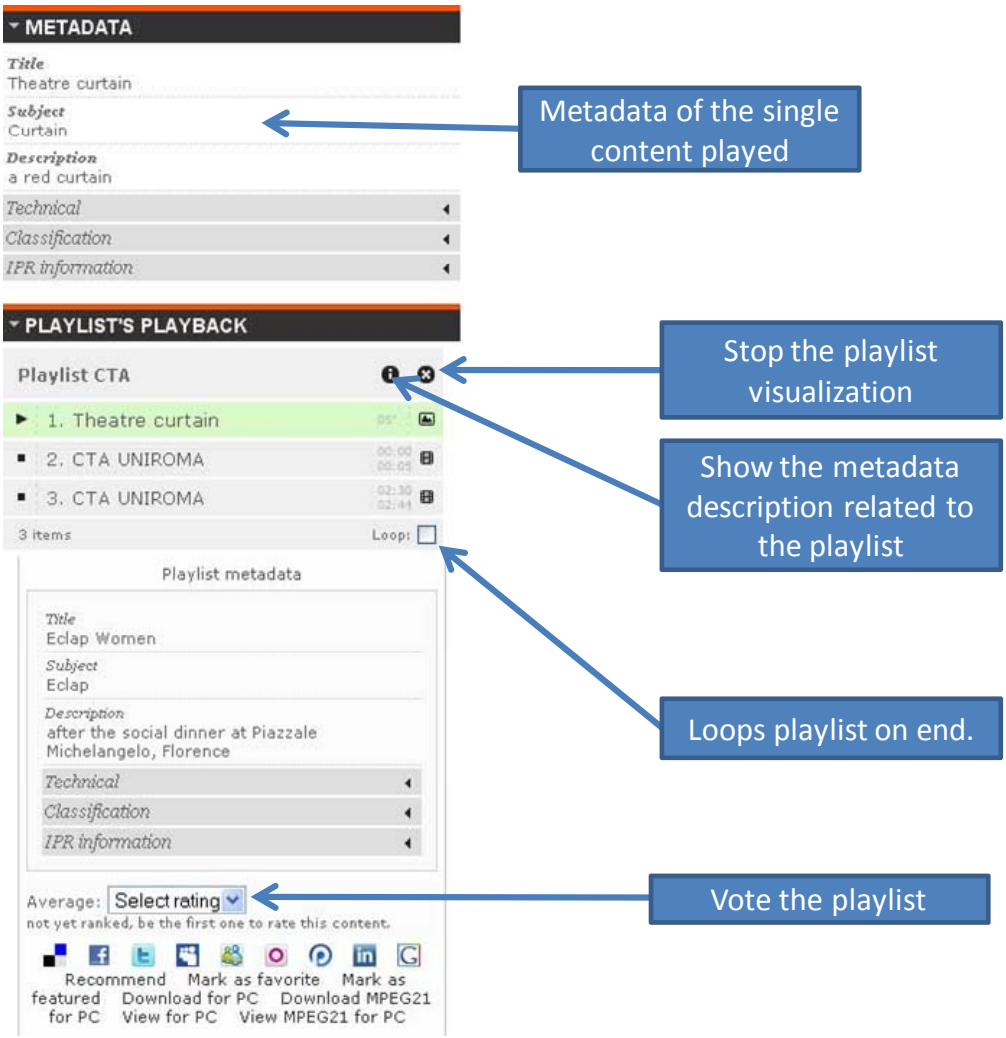


By clicking on a playlist, the **Playlist's Playback** window is shown on the right column of the web page, with the list of associated content. For each content item in the list the following information are available: start time and end time in case of audio or video excerpts, visualization time in case of image. A small icon helps to identify if the content type in the playlist is a video, an audio or an image.

Next to the playlist title, on the top of the **Playlist's Playback** box, two buttons are present. They can be activated during the playlist visualization and respectively allow to see the metadata description and stop the playlist visualization.

If the user presses the button to stop the playlist, the content actually played, will continue to be visualized, but at the end, the play will be stopped.

The content actually played is highlighted in light green in the list, as shown in the following image. During the playlist play, the user can in any time recommend the playlist to colleagues, mark the playlist as favorite and add a vote to the playlist.



5.1.2 ECLAP Collections

ECLAP provides to Registered Users a set of content aggregation facilities. By creating a **Collections** the user may decide to publish the personal collection to the ECLAP community for sharing it with other colleagues, this implies the creation and publication of the Collection. A collection may include any kind of content items, including play lists. The published collections can be updated and used as the primary container and sources for creating ECLAP Courses via the e-learning facility of ECLAP.

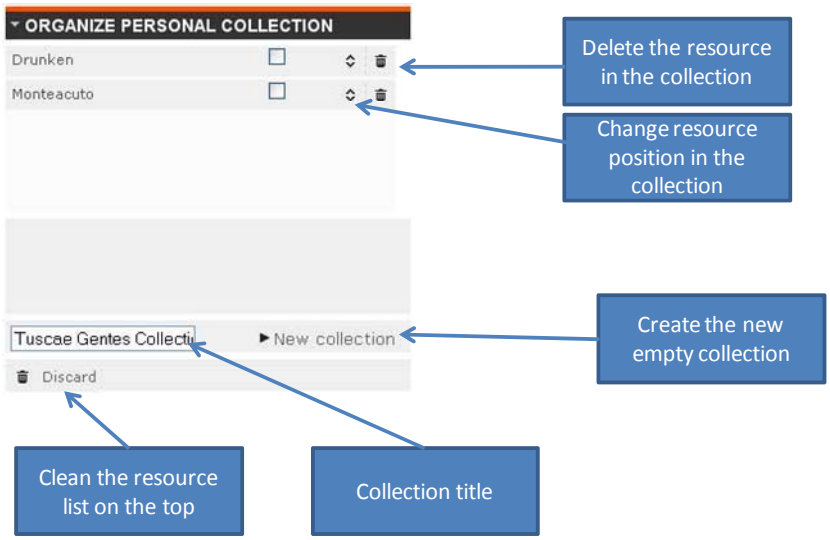
The Collection facility allows users to organize their collections as follow:

- create a new collection, specifying the title of the collection and its metadata;
- delete a published collection;
- add/remove cross-media resources in/from the collection;
- publicly publish the collection on the ECLAP portal;
- update the published collection;
- view the list of collections (published and unpublished) and the resources aggregated in each collection.
-

5.1.2.1 Create a new collection

To create a new collection, the user has to be registered and logged in the portal. When the user decides to create a new collection by aggregating different kind of resources published in the portal, he has to select the "**Add to personal collection**" item available in the menu **Actions** on the bottom right of each resource preview box.

Alternatively, a resource can be added to a collection during the play, by clicking on "**Add to personal collection**" button available on the **ACTIONS** block on the right column of the webpage (see the figure on the left).



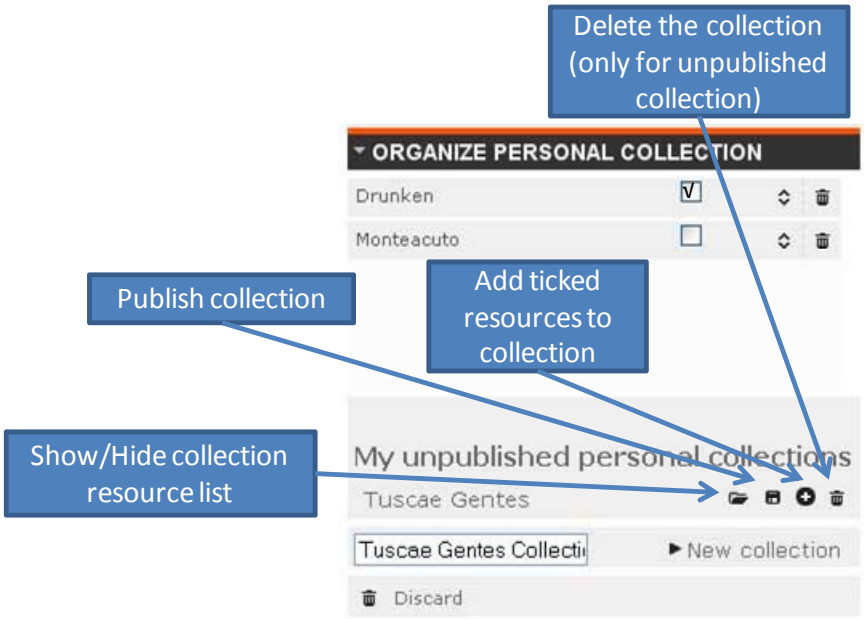
Immediately a new collection is created and a block named **ORGANIZE PERSONAL COLLECTIONS** is shown on the right column.

Each time a new resource is played and the "**Add to personal collection**" button is clicked, the resource is added in the top list in the **ORGANIZE PERSONAL COLLECTIONS** block. This is the first step, since only the resources listed in the top of the block can be added in the collection.

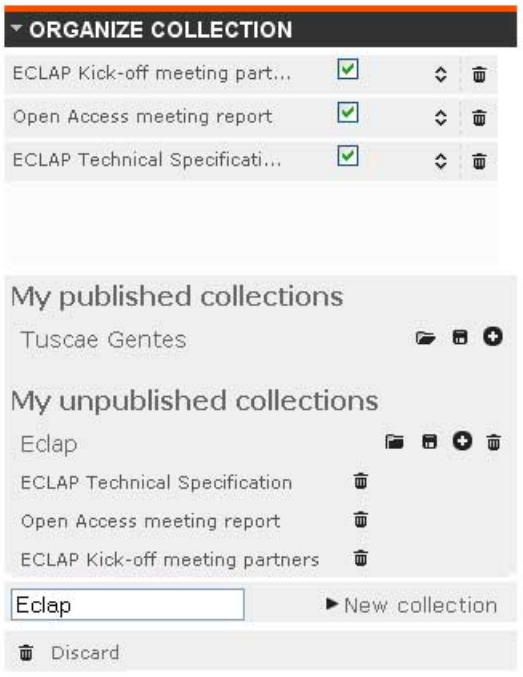
As shown in the figure on the left, in this first phase the **ORGANIZE PERSONAL COLLECTION** block allows to:

- **Delete** resources from the list;
- **Change the position** of the single resource in list;
- Specify the **collection title**;
- **Delete** the collection (if previously created and listed as "unpublished. See later);
- **Add the new collection** in the Personal Collection when the collection is defined.

When the "**New collection**" button is clicked, a new empty collection is added in the personal unpublished collection of the user, and the **ORGANIZE PERSONAL COLLECTION** block will show the new collection name in the list, as shown in the following figure. Since in this phase the collection is not completed, it is visible only for the "creator" and it is not published for the others users, so that the new collection is listed in the "**My unpublished personal collections**" list.



To proceed with the creation of the collection, the user has to select the resources to be added, by selecting them from the list of resources on the top. On the right of the unpublished collection title, the button Add resource(s) is available. by clicking it, the ticked resources are added to the collection and listed below the unpublished collection title.



At the right of each resource listed in the unpublished collection, the button allows to delete the single resource.

When the collection is complete, the user can decide to publish on the portal it by clicking on the **Publish collection** button .

The new published collection is added in the **My published collections** list, as shown in figure on the left.

After this, the webpage to add metadata related to the collection is shown. The upload page is similar to the content upload page.

5.1.2.2 Update a collection

Content inside a collection can be updated and modified for both published and unpublished collections by adding or deleting resources in it.

The collection update can be managed in the **ORGANIZE PERSONAL COLLECTION** block. On the right of each content title, a set of buttons are available.

To add a new resource in a collection, select the resource to be added from the top list and press the corresponding **Add resource** button of the collection to be updated.

To delete a resource from a collection, click on the delete button of the resource listed in the collection.

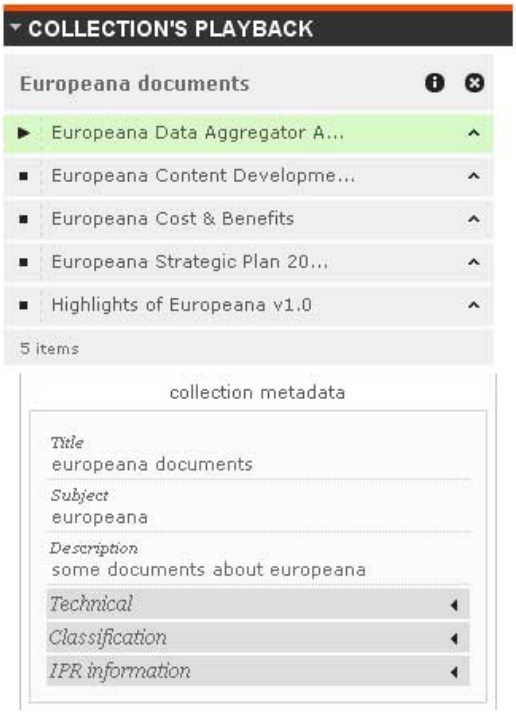
If only the collection title is visible in the collection list, click on the **view/collapse resources** button to open/hide the list of resources available in the collection.

When the resources in the collection have been modified, to update the collection definitively click on the **update collection** button .

5.1.2.3 Collection playback

The user can access to their collection from the top menu **CONTENT – MY COLLECTIONS** or by clicking on the link **My Published Collections** in the **CONTENT** block on the right column.

By clicking on a collection the first resource in the collection is played and a the block **COLLECTION'S PLAYBACK** is opened on the right column.



The **COLLECTION'S PLAYBACK** block shows:

- the collection title
- the list of resources included in the collection. The resource actually played is highlighted in light green;
- the total number of resources included in the collection;
- the collection metadata (available by clicking in the button on the top right corner).

By clicking on another resource listed in the block, the resource is shown. Collections can also be used as e-learning course content (see paragraph 5.1.3.4).

5.1.3 ECLAP e-learning Courses

The ECLAP Portal provides e-learning activities to the registered users. In order to do this, an integration with one of the the most diffused Learning Content Management System has been realised. The LCMS used is Moodle [http://moodle.org/] and this integration allows users of the ECLAP social network to publish educational content and to exploit all the advantages provided by the use of a complete and effective tool for creation and management of on-line courses in an integrated manner. This integration system allows the teachers to create the courses using the Moodle interface which is diffused in the educational communities, with the additional possibility of enriching the courses also using the multimedia content available on the ECLAP Portal.

All ECLAP users can view the list of the available courses in order to have the possibility to subscribe to the course to which they are interested in, thus becoming students for that course and having the possibility to access to the content related to it and performing the lesson, tests or homework that the teacher assigns.

When an ECLAP user is also a student for a course he/she can access to that course and follow the activities inside it provided by the teachers.

When the ECLAP user is a teacher, he/she can both access to the course and also create and organise it through the Moodle Interface. For example dividing the course into weekly lessons and putting links to ECLAP multimedia content, writing the tests that the students have to pass, etc.

Here after, the main e-learning activities and what type of users can access to them:

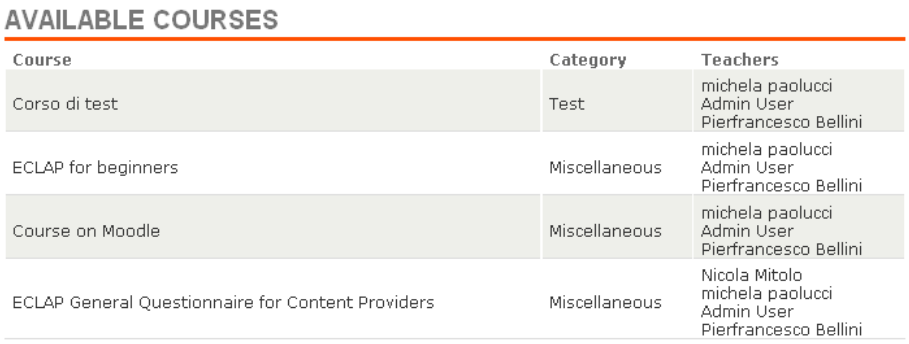
- Available course list: all ECLAP users
- Registration of users to courses and new course creation request: ECLAP registered users
- Course Access: ECLAP registered users that are also student or teachers for a specific course
- Course editing and management: ECLAP registered users that are also teachers for a specific course

5.1.3.1 Available course list: all ECLAP users

When a user wants to see the ECLAP courses list, he/she has to look at the lateral block called "e-Learning courses" and choose one of the following links:

- "All courses": a page containing a list of all courses present in the site will be visible (the 'Available Course' page). Clicking on one course link, will be visible the 'Course program' page in which:
  - the users not enrolled to courses: can see the course program (or a short textual description) and the instructions for their subscription
  - the users enrolled at the course: the course program and a link to access to the course activities (see section 5.1.3.3)
- "All course categories": a page containing all course categories will be visible
- "All teachers": a page containing all teachers will be visible, and the related ECLAP profiles will be directly accessible (clicking on teacher names)
- "My e-Learning activity":
  - For users not enrolled to courses: a message advising the user that he/she is not subscribed to any course and the instructions for the subscription will be detailed
  - For users enrolled at least to one course: his/her courses (the list of the courses of which he is a student or a teacher) and the related course topics and activities will be accessible.

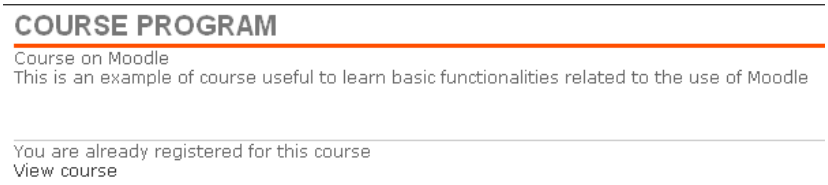
***Figure 5.1** - Lateral e-learning block*



AVAILABLE COURSES

Course	Category	Teachers
Corso di test	Test	michela paolucci Admin User Pierfrancesco Bellini
ECLAP for beginners	Miscellaneous	michela paolucci Admin User Pierfrancesco Bellini
Course on Moodle	Miscellaneous	michela paolucci Admin User Pierfrancesco Bellini
ECLAP General Questionnaire for Content Providers	Miscellaneous	Nicola Mitolo michela paolucci Admin User Pierfrancesco Bellini

Figure 5.2 *- Available Courses page*



COURSE PROGRAM

Course on Moodle
This is an example of course useful to learn basic functionalities related to the use of Moodle

You are already registered for this course
View course

Figure 5.3 *- Course program visible only to the users (users already enrolled to the course)*

COURSE PROGRAM

Corso di test
Descrizione del corso di Test....

You are not currently enrolled in this course
More info about e-learning activities

Figure 5.4 *- Course program visible only to the users (users not enrolled to the course)*

COURSE CATEGORIES

Category	Description
Miscellaneous	
Eclap	this course is about the....

Figure 5.5 *- All Course Categories*

LIST OF TEACHERS

Name	Courses
Admin User	Corso di test ECLAP for beginners Course on Moodle ECLAP General Questionnaire for Content Providers
Pierfrancesco Bellini	Corso di test ECLAP for beginners Course on Moodle ECLAP General Questionnaire for Content Providers
Nicola Mitolo	ECLAP General Questionnaire for Content Providers
michela paolucci	Corso di test ECLAP for beginners Course on Moodle ECLAP General Questionnaire for Content Providers

Figure 5.6 *- All teachers*

Figure 5.7*- My e-Learning activity*

5.1.3.2 Registration of users to courses and new course creation request

Each user that is not enrolled to the e-learning courses available on the ECLAP Portal but to which he/she wants to access and view the lessons as a student, can click on the course link and follow the instructions.

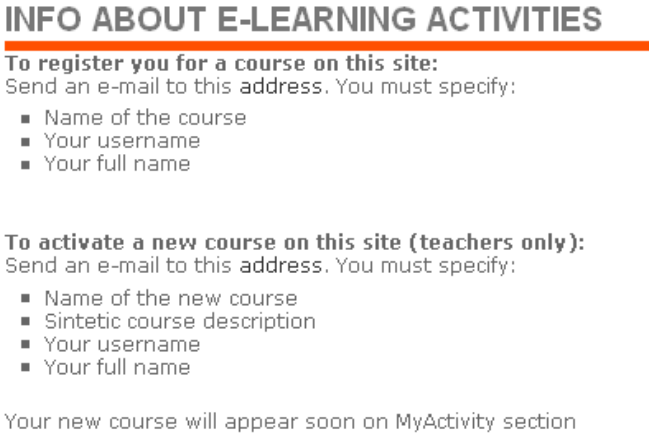


Figure 5.8*- Instruction to enrol to a Course*

5.1.3.3 Course Access

When a user is already enrolled to a course and clicks on the course title, he/she can view the course program and all the lesson units: view the resources, read the text pages and in general done the lesson activities provided by the teacher

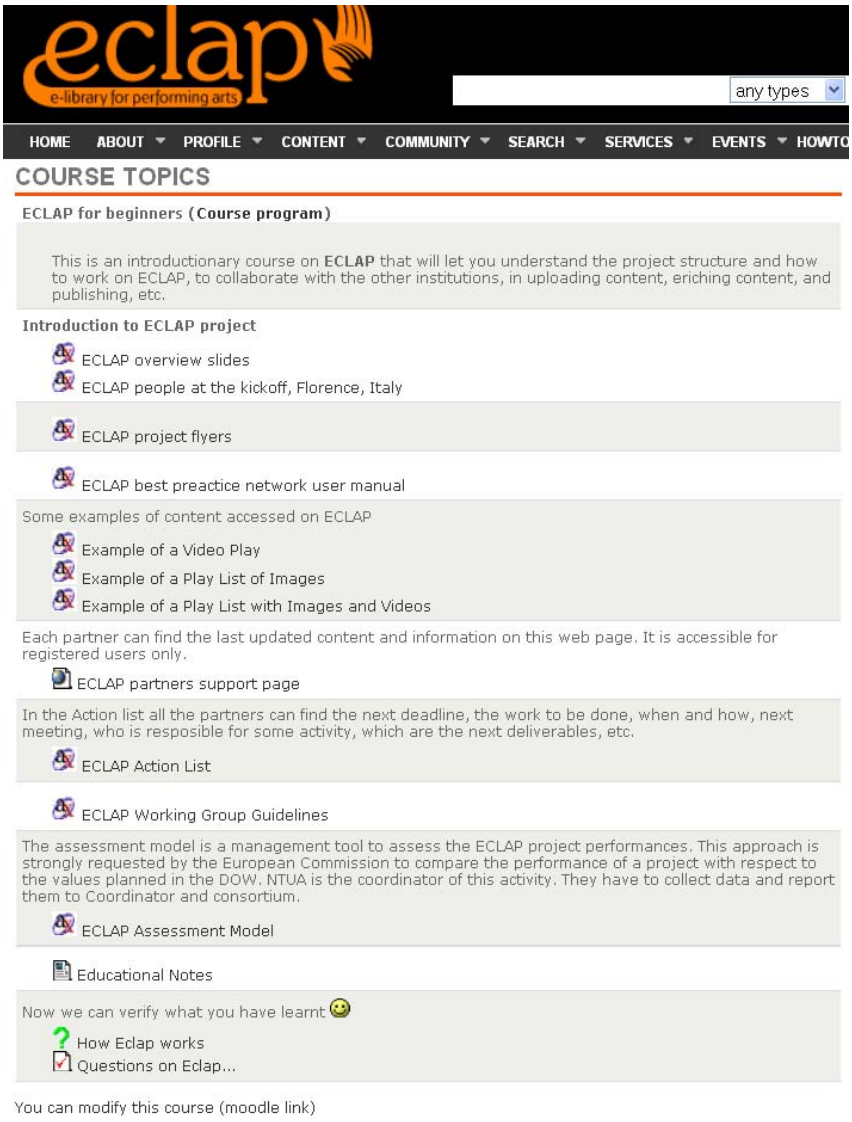
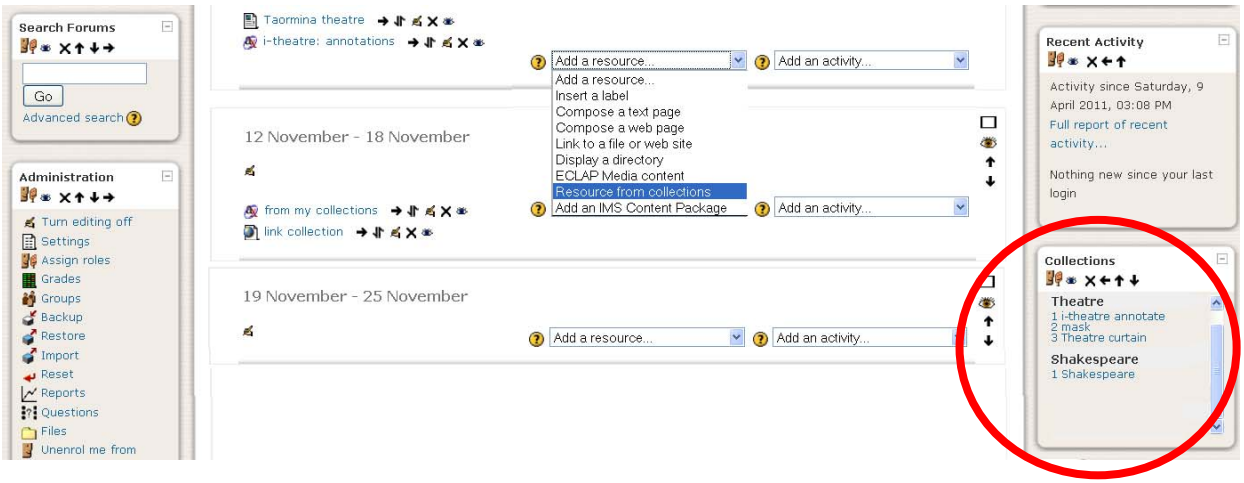


Figure 5.9*- Course Access*

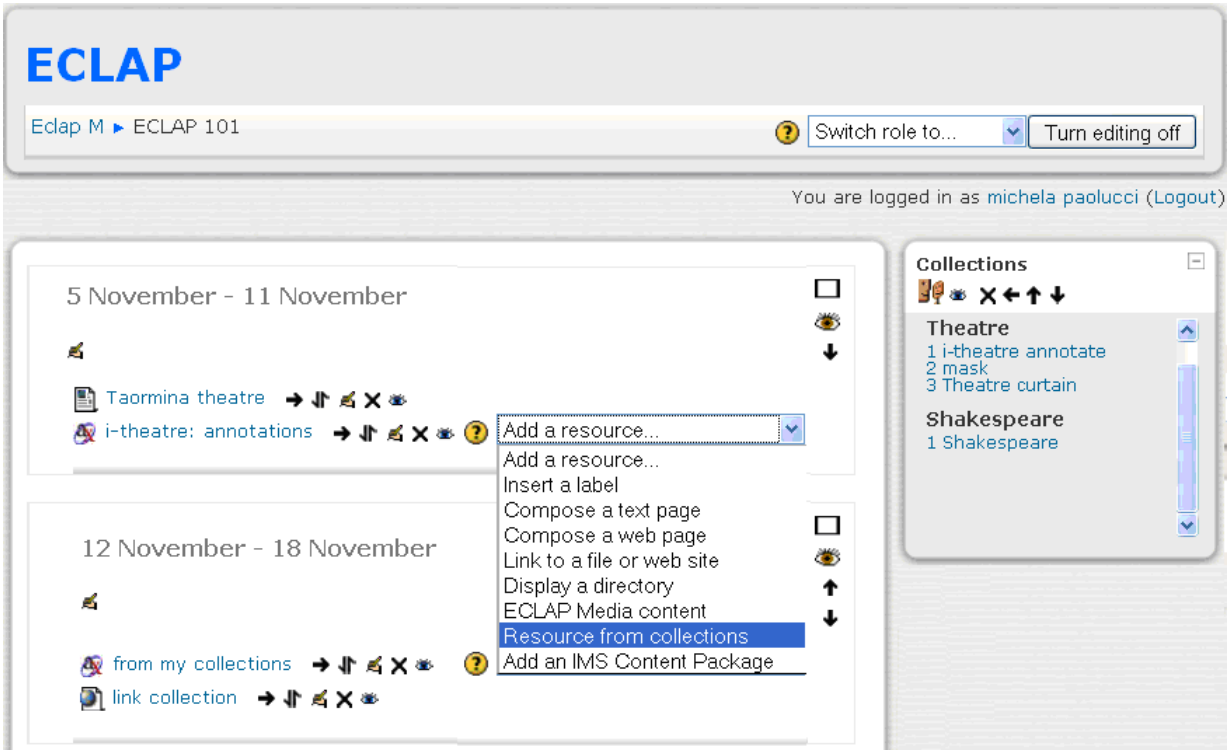
5.1.3.4 Course editing and management: ECLAP registered users that are also teachers for a specific course

Resources that have been previously included in a collection published (visible to the ECLAP registered users) or unpublished (visible only to its creator), in the ECLAP Portal can be added in a course. This procedure is different with respect to the previous one (related to the adding of single ECLAP resources). While to add in a course a single ECLAP resource it is necessary to copy and paste its link (with the AXOID), to include a resource contained in a collection it is sufficient to select it from a list.

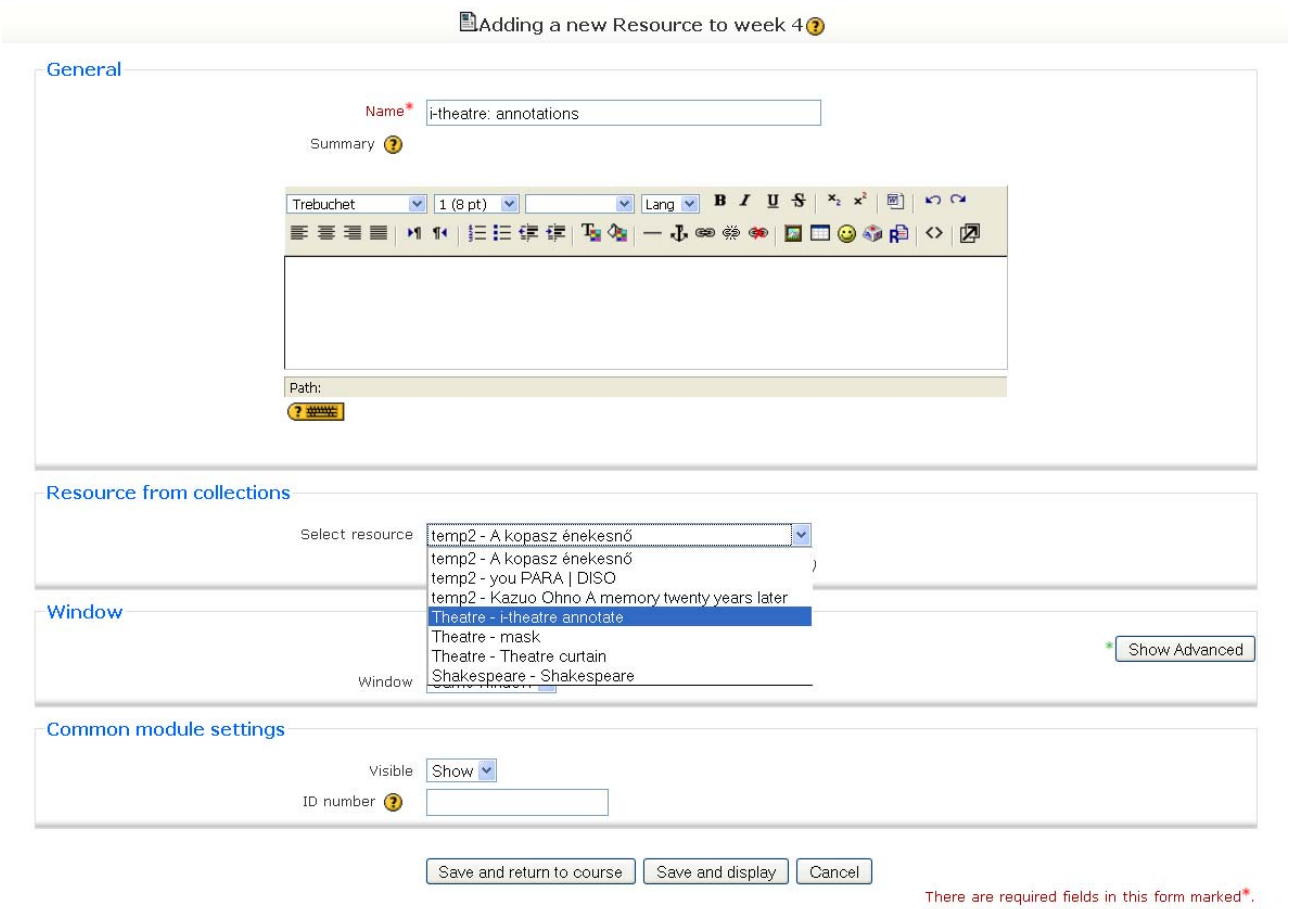
Collections published in the ECLAP portal are also listed in the Collections block available in the **Moodle** front-end interface.



To add a resource from a collection, click on the menu "Add a resource…" and select the item "Resource from collections" (see next image).



A new page is opened (see the next image) in which the teacher has to: insert the title of the ECLAP resource, insert the description, select the resource title from the list of the field "Resource from collections", and finally the click the "Save and return to course" button.



After this, the resource will be included in the Course program: the teacher can view the resource presence from the Moodle Platform and the student can see the added resource from the ECLAP Portal course page.

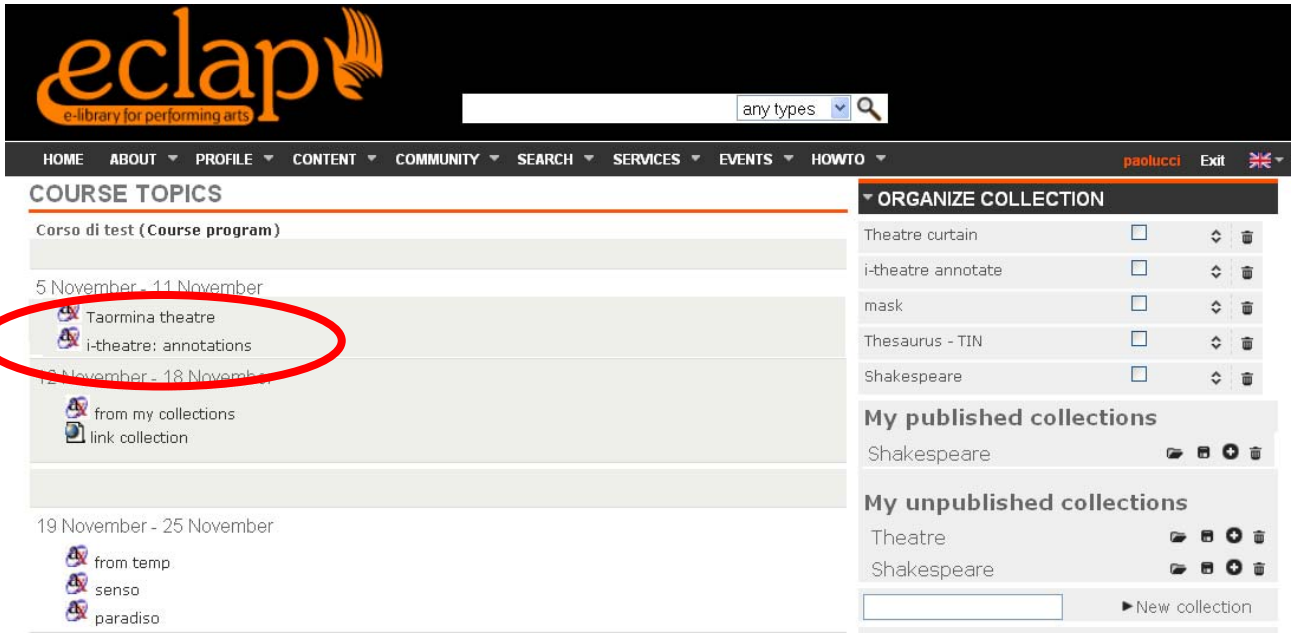


Figure 5.10*- ECLAP Portal course page (visible to Eclap user enrolled to the course)*

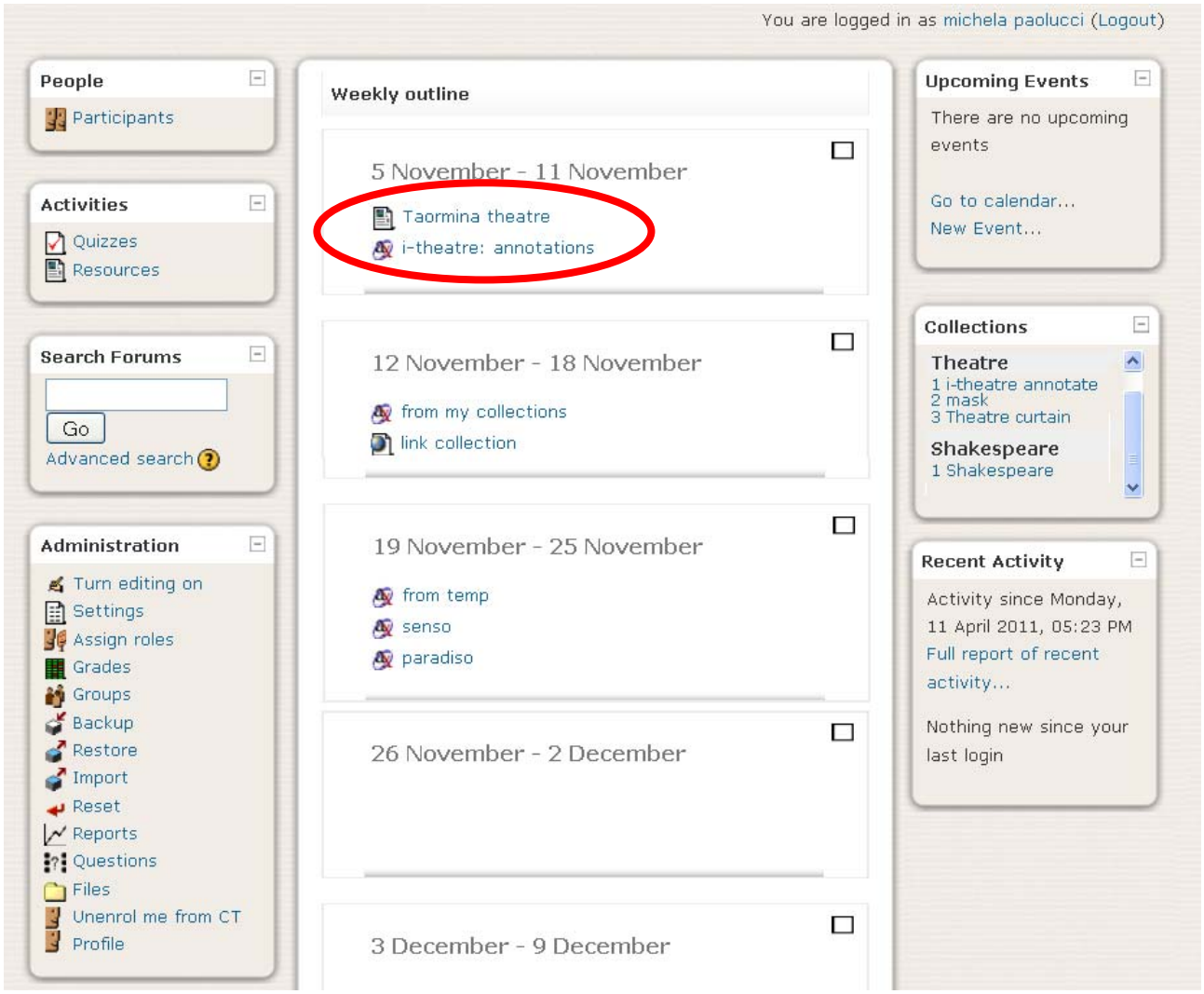


Figure 5.11*- Moodle course page (accessible only for the teachers)*

5.1.4 Rich Media and Intelligent Content for PC & Mobile

The ECLAP portal besides the standard media formats as images, audio, video and documents supports interactive rich media content formats like Flash and/or HTML (with Javascript) based content. These interactive formats allow the realization of a broad range of "applications" that can be downloaded and used offline. The rich media interactive content can be used for the realization of:

- Quizzes and tests;
- Interactive games;
- Calculators (e.g. medical calculators);
- Presentation of e-learning material;
- Etc.

The portal uses the MPEG21format [37] to represent this kind of content, the MPEG21 is used as a container for the various files.

The portal allows to upload MP21 content produced using the AXMEDIS editor [http://www.axmedis.org], [40], [41] (see Figure 5.12 or using specific tools created for the production of specific interactive content

(e.g. the Algorithm Designer used to produce interactive medical procedures to be followed using a smart phone during an emergency, Figure 5.13).

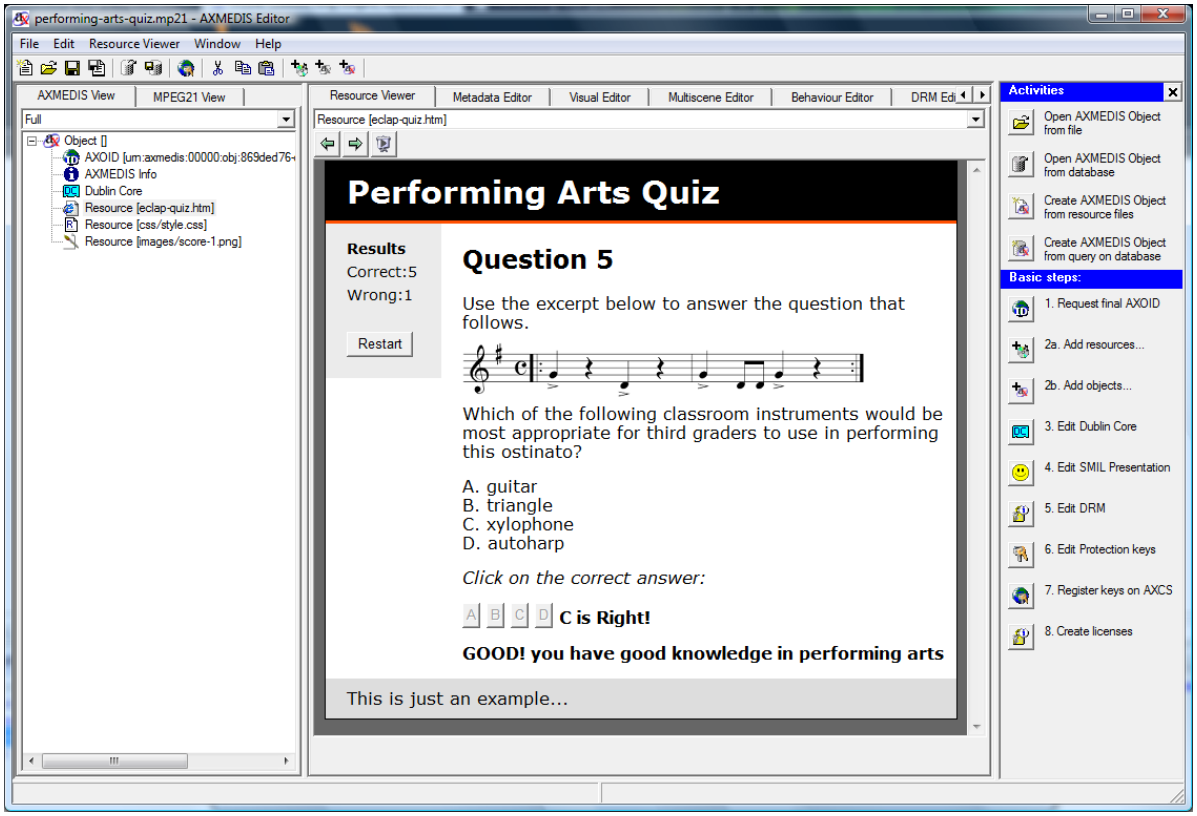


Figure 5.12 *– AXMEDIS Editor showing an HTML based content*

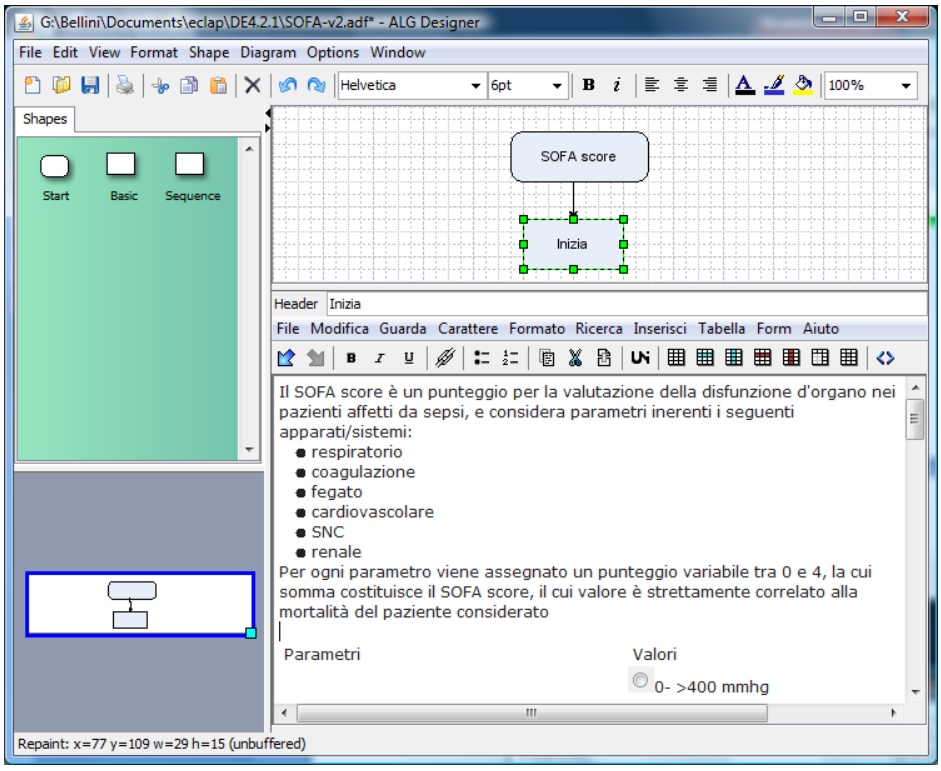


Figure 5.13*– Algorithm Designer for the production of procedures*

When the MP21 file is uploaded on the portal if its content is HTML/flash based it is extracted to allow the use of this content even online (see Figure 5.14).

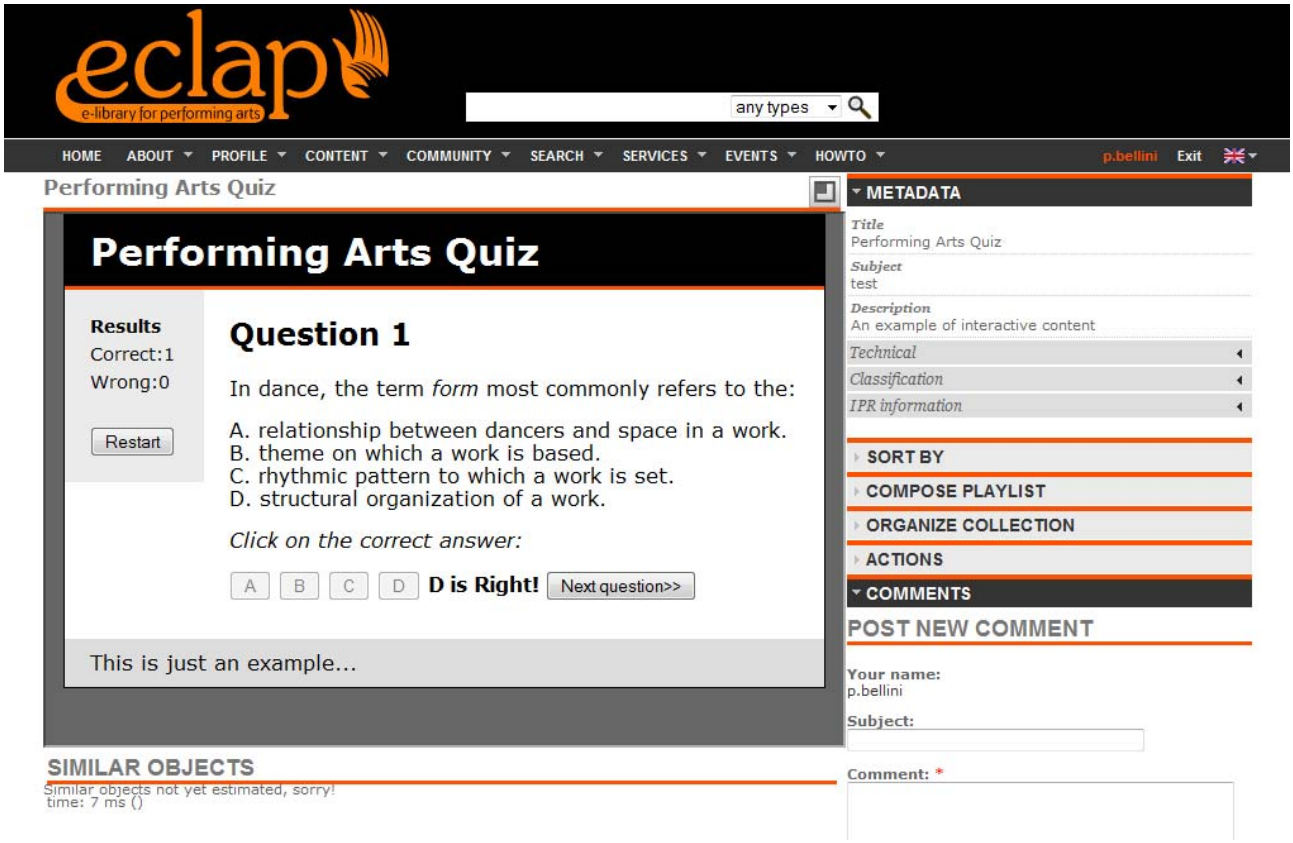


Figure 5.14– *The ECLAP portal showing an interactive content*

However when the user requests to download the content the MP21 file is provided for PCs and for the Windows Mobile 6.5. The MP21 file downloaded from the portal can be used on a Windows PC using the AXMEDIS Player (see Figure 5.15) and on Windows Mobile 6.5 phones using the AXMEDIS PDA Player.

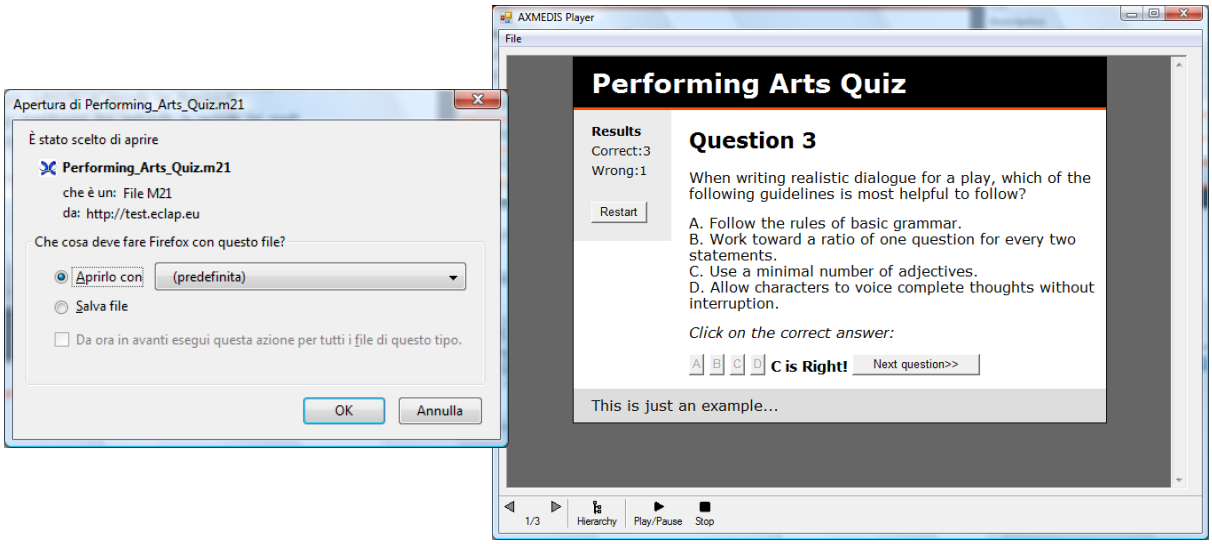


Figure 5.15– *Download and view on a PC with AXMEDIS Player*

For other smart phones as iPad/iPhone/iPod touch and for Windows Phone 7 and Android based phones the content is directly downloaded using a specific application which stores the content locally and allows to use it offline (see Figure 5.16). In this applications with the content is also downloaded the metadata that are used to locally index the content allowing the user to search for it on their phone.

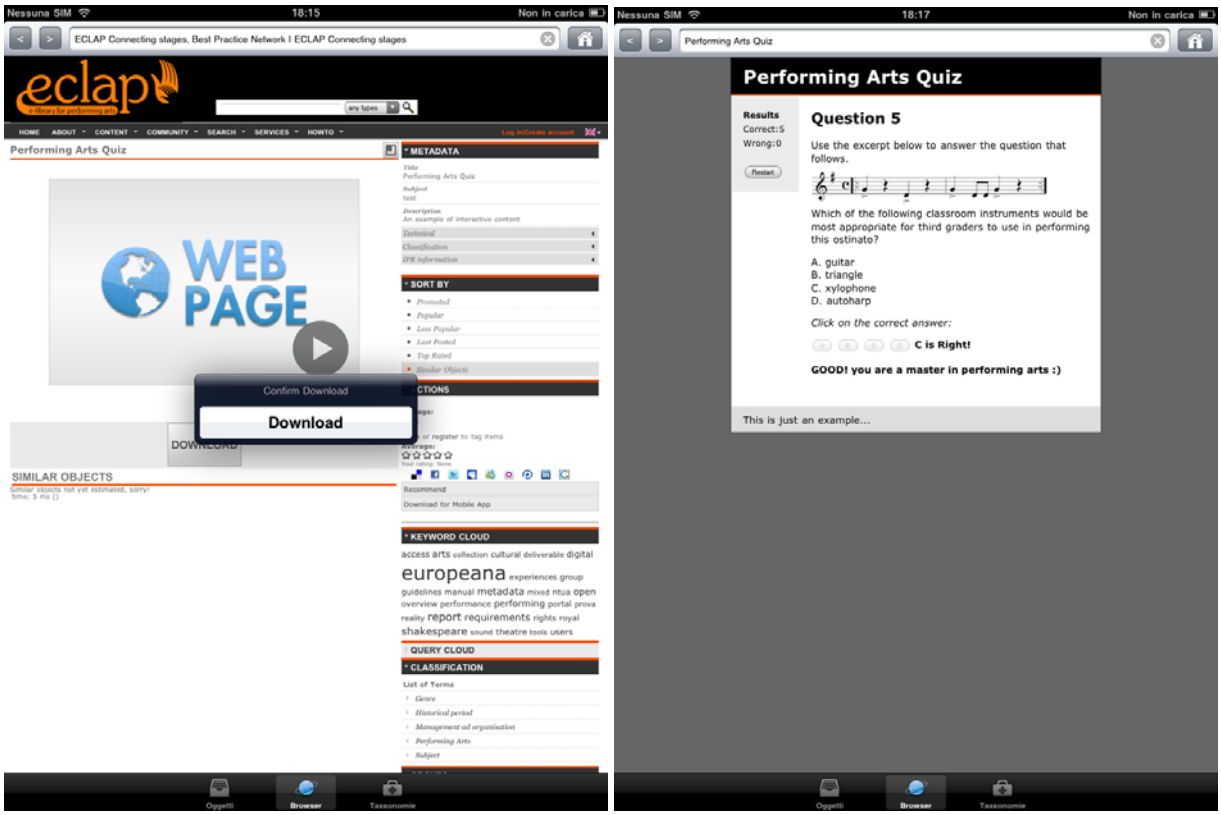


Figure 5.16– *Download and view content on an iPad using the* Content Organizer *App*

In the future versions of the portal also epub format for e-books will be supported allowing to upload and download content as epub

Another usage of the right content is the adoption of training courses with the video of the speaker synchronized with the slides as in the following figure.

5.2 Content Augmentation and Enrichment

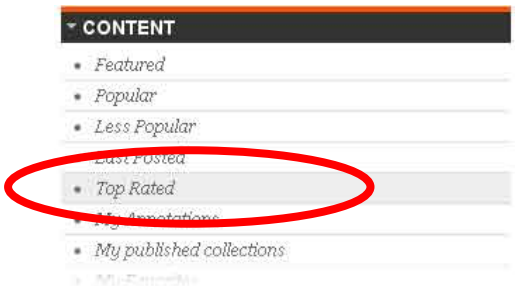
ECLAP portal allows users to enrich information associated with content adding comments and votes (see section 5.2.1), adding free text tagging (see section 5.2.2), adding annotations and allow non linear content fruition (see section 5.2.3), enriching metadata, associating content with taxonomy and translating metadata in different languages (see section 5.2.4) .

5.2.1 Votes and comments

During the play of a resource, unregistered users can see in the **ACTIONS** block in the right column, the average rating expressed by the users and referred to the resource they are viewing. The average of all ratings an item has received shows the perceived quality of a content item.

Registered users can decide to rate the content or to leave a comment on it. The rating process is available in the **ACTIONS** block on the right column. ECLAP provides a mechanism for users to rate an item on a numeric scale from 1-5, where a rating of one is worse than a rating of 5. The rating is expressed graphically in the form of five starts. By passing the mouse on the rating stars, the explanation text of the rating is shown below the little stars.

The rating mechanism is very useful also for the ECLAP community, since users can see the list of top rated contents available by selecting the "**Top Rated**" item in the **CONTENT** block.



Also registered users can decide to post a comment related to a content item.

When a resource is played, comments provided by registered users are listed on the **COMMENTS** block on the right column. Each comment contains also the date of the comment and the name of the user who leaves it. The **COMMENTS** block allows registered user to post a comment by filling in a simple form by including a Subject and the comment. Before definitively posting the comment, users can view a preview of it by clicking on the **Preview** button.

Finally, below each comment, a **Reply** button is available to allow registered users to reply to a comment.

5.2.2 Free tagging

This chapter describes the work done in respect to free tagging and the ideas and research it is based on. In a later stage of the project, when the first user evaluations will be completed, this document will be amended with gained experiences as well as more elaborate and precise recommendations.

5.2.2.1 What is free tagging

In terms of content enrichment, free tagging allows users to label (tag) online objects present with keywords of their choice. Tags in this respect can be entered in the form of free text and are usually single words that describe or categorize the content (see more on user motivations for tagging in the next section).

A folksonomy [27] is the combined collection of user tags of a system. In other words: a folksonomy emerges through large-scale collective tagging efforts. Every time a user adds a tag, it is added to a database, indexed and added to the folksonomy.

In contrast to a taxonomy [28] which is usually made and curated by a small group of experts, a folksonomy is built up from the collective efforts of different types of users. Because of this difference in authority and

the variety of users that add the tags, a folksonomy can offer interesting additional views on and categorizations of collections. However, interpreting a folksonomy is very hard. First of all it is unknown what the user refers to when he adds a tag, for instance when tagging a photo and entering the tag *blue*, the user could refer to the main color in the picture or the feeling it represents (*feeling blue*). Moreover as a folksonomy can possibly contain tags from multiple languages, interpreting the semantic relations between the tags and the referred content becomes even harder, [29], [30], [31].

In ECLAP an effort is made to tackle one of these problems by trying to automatically assign the most likely language to the tags users enter. See chapter 5.2.2.4 for an overview of the work done on implementing this functionality.

5.2.2.2 Motivations for tagging

Research on the motivations of people that tag online materials [32] have shown that the reasons are:

- Motivations related to indexing (in order to make content easier to find for all users)
- Motivations related to socializing (using specific tags for fun/socializing)
- Motivations related to communication (friends using tags to group content for their own convenience)

In ECLAP the user group most likely to be motivated for indexing is the group of professional content providers responsible for uploading performing arts materials. Besides having the option to annotate their content by using terms from ECLAP's taxonomy, these users are likely to use the option of free tagging in order to be able to provide a more accurate and additional index for their material.

Another group likely to be motivated to use tagging for improving the indexing of content are the members of the Educational user group. Teachers and students might want to use specific tags for specific assignments making it more convenient to point out relevant items of content for those specific assignments or lessons.

Motivations related to socializing and communications are most likely to apply to Leisure users. In some cases, this kind of tagging can be similar to leaving a comment. (e.g. users tagging with words like: *nice, great, worthless* etc)

5.2.2.3 ECLAP requirements

Besides the requirement of allowing users to tags content, there are the following additional tagging requirements in ECLAP (see D3.1: *Infrastructure: ingestion and processing content and metadata* for these requirements in full detail):

- Detect the language of tags entered
- Detect the user role of users that enter the tags
- Generate tag clouds [33] that show tags in the language of the user that views them

In the next chapter the fulfilment of these requirements is described in detail

5.2.2.4 Work done in ECLAP on free tagging / social tagging

Regarding the implementation of free tagging in ECLAP, NISV has developed a Drupal [34] module that supports functionalities for adding tags and the generation of tag clouds.

Adding tags

In the current implementation, the tagging functions can be used in both the front-end and back-end of the portal.

In the front-end, all registered users can use free tagging on pages that show the detailed information of a digital object.

In the back-end, only users logged in as content providers can enter tags in the pages where the metadata of digital objects can be edited. In both cases, there is no limit to the number of tags that can be added.

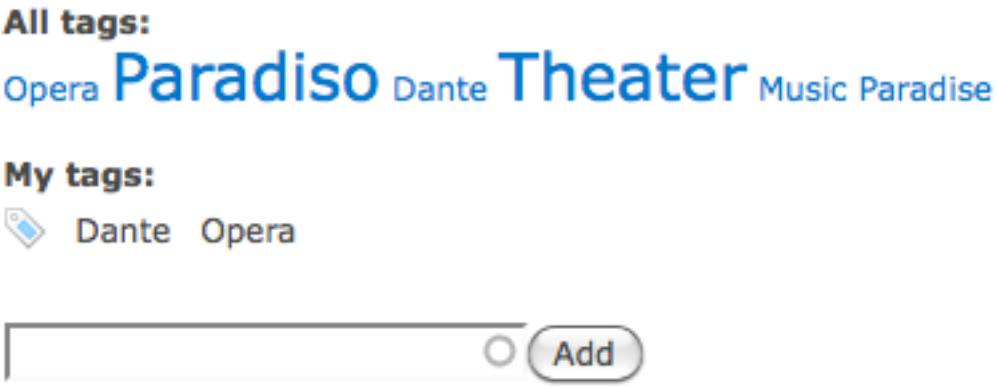


Figure 5.17*- Form for adding tags*

Whenever a user adds a tag, using the above form, the system automatically detects the most likely language of the added tag. The reason for automating this, rather than asking the user to provide the language manually is that the success of free tagging relies on its simplicity. By adding an additional language field to the tagging form, users might be discouraged to enter tags.

The algorithm for the language detection first collects the following information:

- the user's current preferred language for viewing content
- the current language of the content being viewed
- the language setting of the user's browser

Using this information the following is checked:

- if the browser's language setting and the language of the content are the same, this language will be attributed to the tag. If this is not the case
- the user's preferred language setting in his browser is attributed to the tag.

Tag clouds

Tag clouds generated for folksonomies usually show the most popular tags that were entered by users – the more times a tag is added, the larger it appears in the tag cloud. In ECLAP these tag clouds are enhanced with the following features:

- distinction between tags of different user roles
- clouds show tags in the preferred language of the user (and supplemented with English terms)

The following figure shows the tag cloud as it is displayed in ECLAP:

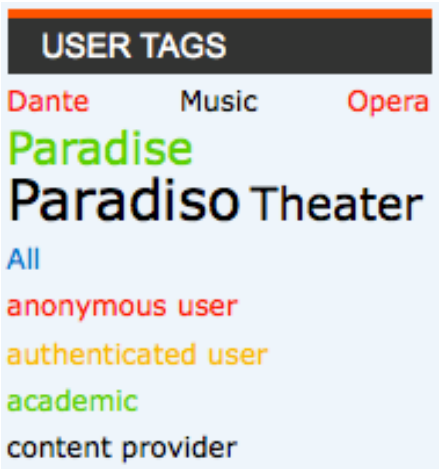


Figure 5.18*- ECLAP Tag cloud*

As shown in this figure the tags can appear in different colours of which each colour represents the user group that added this tag the most times. The cloud can also be filtered by user group, by clicking on one of the displayed user roles (*anonymous user, authenticated user, academic, content provider*).

5.2.2.5 Recommendations

When defining the requirements for tagging functionalities for any system, it is advised to first make sure to reflect on the motivations as described in 5.2.2.2. If the tagging is primarily for leisure or socializing, analyzing the semantics of the folksonomy might not be a priority. If there is however a demand to use tags to optimize for instance:

- suggestions for related content (within the same system)
- suggestions for related external sources, such as Wikipedia or Europeana

it is important to analyze the semantics of the tags in the folksonomy as much as possible.

For instance, if a user tagged a photo of a group of people posing and smiling for a group picture with the tag *cheese*, most English speakers will understand why this tag was added. A recommendation system however might relate this tag with the concept of cheese in the sense of *food*, rather than the sense of *posing and smiling people in front of a camera*. In this case, for example, the system would offer viewers of the group picture related content in the form content related to cheese.

Fortunately there are several research projects [29], [30] and [31] that deal with the topic of interpreting and disambiguating the semantics of folksonomies. Several solutions are proposed as an outcome of these projects, which should be addressed depending on the user requirements and available resources in a project.

5.2.2.6 Future work

Statistics

With respect to the aforementioned difficulties of interpreting folksonomies and automatically assigning a language to a tag (see 5.2.2.1 and 5.2.2.5) it is useful to study ways of generating statistics from the user tags entered. For instance, in order to validate whether the automatic assignment of language is correct, it would be useful to calculate the number of tags clicked (in a tag cloud) in the preferred language of user as opposed to the number of English tags clicked. If it shows that users generally click tags in their preferred language (rather than English tags), it could mean that the system of automatically assigning the most likely language to tags is useful.

Tag suggestions

Tagging suggestion functionality shows the user possibly relevant tags as he is entering a tag. Implementing such a feature encourages the reuse of tags, making them possibly more relevant for searching purposes. Having this feature should also minimize the amount of different tags that are intended for the same purpose.

5.2.3 Annotations and non linear media fruition

ECLAP portal allows users to annotate media objects present in the portal.

The annotations are performed by a tool called MyStoryPlayer. The innovative part of MyStoryPlayer lies in the fact that no difference between media and the user's annotation exists, because both categories are referred to multimedia objects and they are temporally connected.

This aspect could be of strength in educational environment, for example when a teacher wants to prepare a lesson for his students, synchronizing slides and video in the same environment, or adding some pictures or audios examples that can explain better the concepts of his lesson.

Before to explain in detail all the necessary steps to make an annotation on ECLAP, it is better to focus on what a user is going to do when he/she annotates a media through MyStoryPlayer.

An annotation on MyStoryPlayer is composed mainly by two elements:

- *A text description, as in other annotation tools,*
- *A link between two media, related through a time relation.*

For example a user can annotate a part of a video that lasts one minute, with a slice of another video that lasts 30 seconds. In the MyStoryPlayer interface, these two media will be reproduced synchronously according to their temporal lines.

This kind of annotation allows the user to connect two media that in his opinion are related by topics or events, characters in the scene, music, moviemaker, or they are a different view of the same event and he want to see them synchronously.

The functionality provided to users is:

- Add an annotation
- Load an annotation
- Navigate through the annotation structure

5.2.3.1 Add an annotation

Adding an annotation on a media means to select a part of it through a proper player interface, add a text description on it, and select another part of media to relate it with the first one.

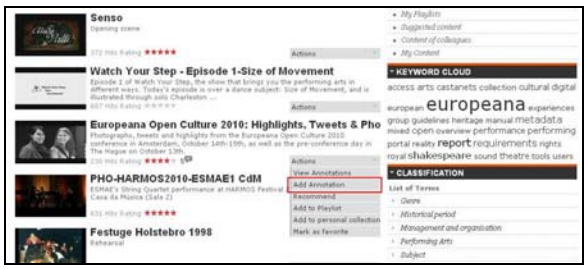
User can add an annotation typically when he/she is watching a media (so the user can do it on the action box), or he/she is in the homepage and, searching for a specific media, he wants to add an annotation on a result (in this case the user has to click on link below the preview).

The main steps to be done in order to create a new annotation are:

1. Select a media to annotate and click to "*Add Annotation*" link,
2. Decide how long should be the new annotation (i.e. select a start and end of annotation through a proper player) according to the timeline of the media,
3. Add a text description to this annotation,
4. Choose another media to be related to this annotation (the media reference that will synchronously appear aside to the main media in the MyStoryPlayer). The choice of the second media could be done by performing a new search on the system, or viewing more contents on the portal and choose one of them, or select it in the list of similar objects proposed by the system.
5. Once the user has chosen the second media, he can select which part to relate to the annotation on the first media in the same manner done for the step 2.

In the end push the Save button.

Step 1: Choose a media



The *Add Annotation* link appears in the Action panel below the preview of multimedia contents or in the Action box during the playing. Clicking on it, the user can add an annotation to the chosen media.

A block will appear on the right side of the page. Clicking on the pen icon, or on the title, a proper player will be loaded in order to allow the user to select a start and an end of the annotation.

Step 2: choose the annotation length

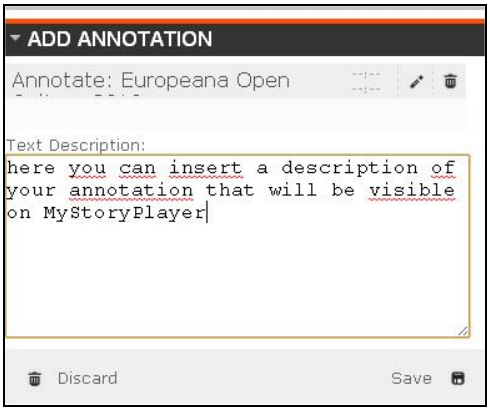
Once the user clicks on the pen icon, or on the title in the box, a player will be loaded. In this player there will be three control buttons to use in order to choose the length of annotation. This can be done as for the playlist (see section 5.1.1.1).

NB: if no piece is selected, the default length of the annotation will be the video length.

For images, the default length is 20 seconds.

Step 3: add a text description to annotation

In the box Text *Description*, just like in the other annotation tools, the user can add a description about the annotating scene.



Step 4 : choose another media as media reference

Once the user has edited the first media, he may choose another media as annotation.

This could be done in many ways: picking one media among the list of similar objects provided by the system, or performing a query in the search field, or using the "*Contents*" box on the right and selecting the list of objects (*last posted, featured, popular, my Contents, suggested contents, etc.*).

The user using the "Add annotation" link can choose the media to relate with the first allowing users to view both synchronously on MyStoryPlayer.

This could be done for many reasons, for example to connect media by events, or for didactical purpose, in which a teacher can make a video of his lesson and put slides that go in parallel with it.

Another case could be represented by a user, owner of many video about the same event that can put them together in MyStoyPlayer, synchronizing them.

Step 5: select a part of second item as media reference and save.

The piece of the second media can be chosen in the same manner as the first. The two lengths could be different. MyStoryPlayer will have different behaviours depending on the length of the second piece respect to the first one.

5.2.3.2 Load an annotation

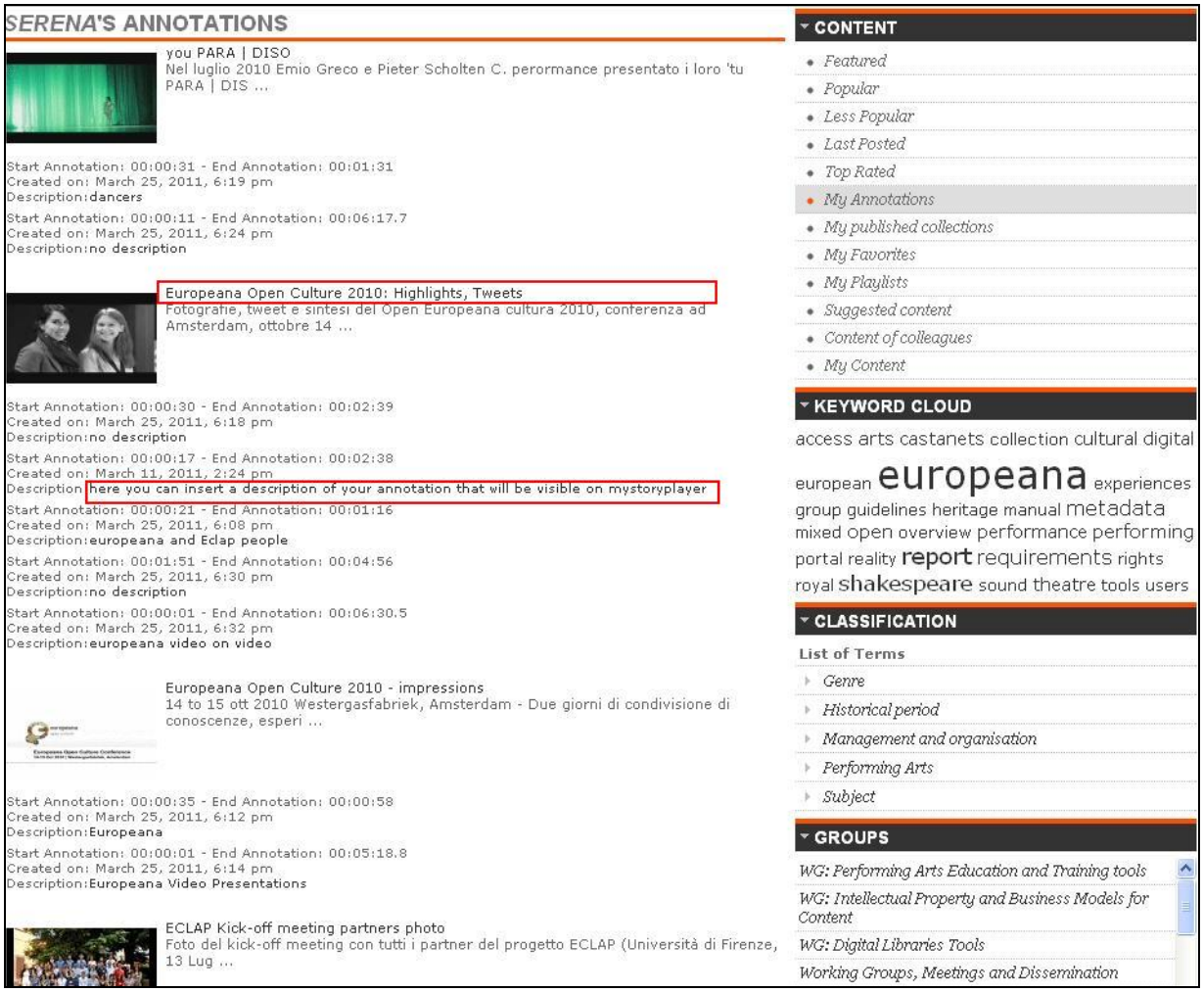
Once the user has saved the annotation, he/she can view it on MyStoryPlayer.

This can be done in two ways:

- If a media is annotated, in the action panel just below the preview, a link *view Annotation* and an icon with the number of annotations related to that media is shown. Clicking on one of them, it is possible to load MyStoryPlayer
- Clicking on *my Annotations* menu in the block on the left, or in the user profile

Choosing the first option, MyStoryPlayer external page will be loaded and the media annotated begin to play from the beginning, and the player will load all the annotations related to it.

Choosing the second option, a list of all user's annotations will be displayed.



In this case, the user can choose among his annotations and click on the title of media or in the text description (in this case the video will be loaded from the point in which the annotation starts).

5.2.3.3 Navigate on MyStoryPlayer

When the main media is running, another video or more media (which can be video, audio or images) starts to play by level with a rectangle aside from the main video.

These media are connected with the main video for the time represented by the length of the rectangles, afterwards they disappear. During the time they are active, the user is granted with many choices: he can keep on viewing the main video or he can click on the other video represented by a side annotation.

This latter action introduces a swap between the two videos, while new contents are loaded in association with the new main media.

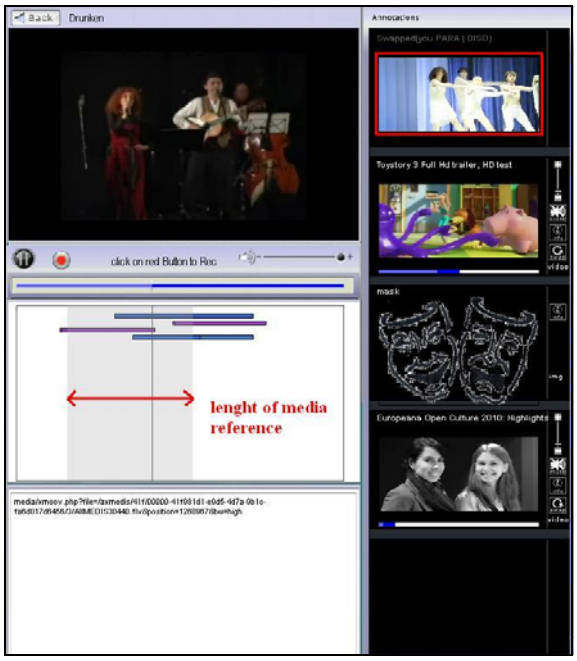
Every time a swap action is executed, user can go back to previous step simply by clicking on the *Back* button, just like it occurs with any web browser.

Let's take a look on what happens when from the ECLAP portal, user loads a media on MyStoryPlayer.

MyStoryPlayer front end will be loaded in order to allow the user to watch the annotations related to this media and navigate among them.

With MyStoyPlayer, user can:

- Have multiple contemporary views and multiple choices on what to watch according to the temporal line
- Swap videos, loading a new scenario with more annotations,
- Navigate among them clicking on the active side annotations and through a back button



5.2.4 Metadata enrichment, taxonomization and translation

Metadata enrichment, taxonomization and translation allow increasing the metadata associated with a digital resource. Some of these activities are performed either automatically by the back-office or manually by users.

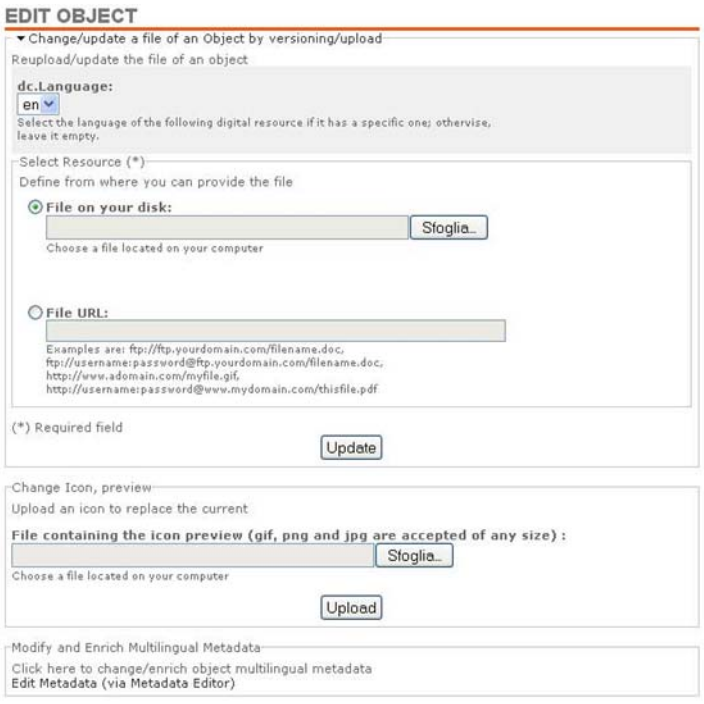
The first metadata translation is performed automatically by using an external service and then will be validated (and modified) by users using the Metadata editor.

Metadata editor is the web page application under ECLAP to be used for metadata enrichment, taxonimization and translation. Since metadata editing is governed by ECLAP Workwlow, the Metadata Editor is a workflow-drive tool.

A registered user that uploaded a content or workflow user can decide to update it by changing the metadata associated to the resource.

To edit an object the authorized user has to be logged in the portal. When the resource is visualized, in the right column of the web page a box named **Actions** is present. Inside the box the option **Edit object** is present (only if the logged user is also the one that uploaded the object in the ECLAP BPNET for the first time).

A new **Edit Object form** is displayed.



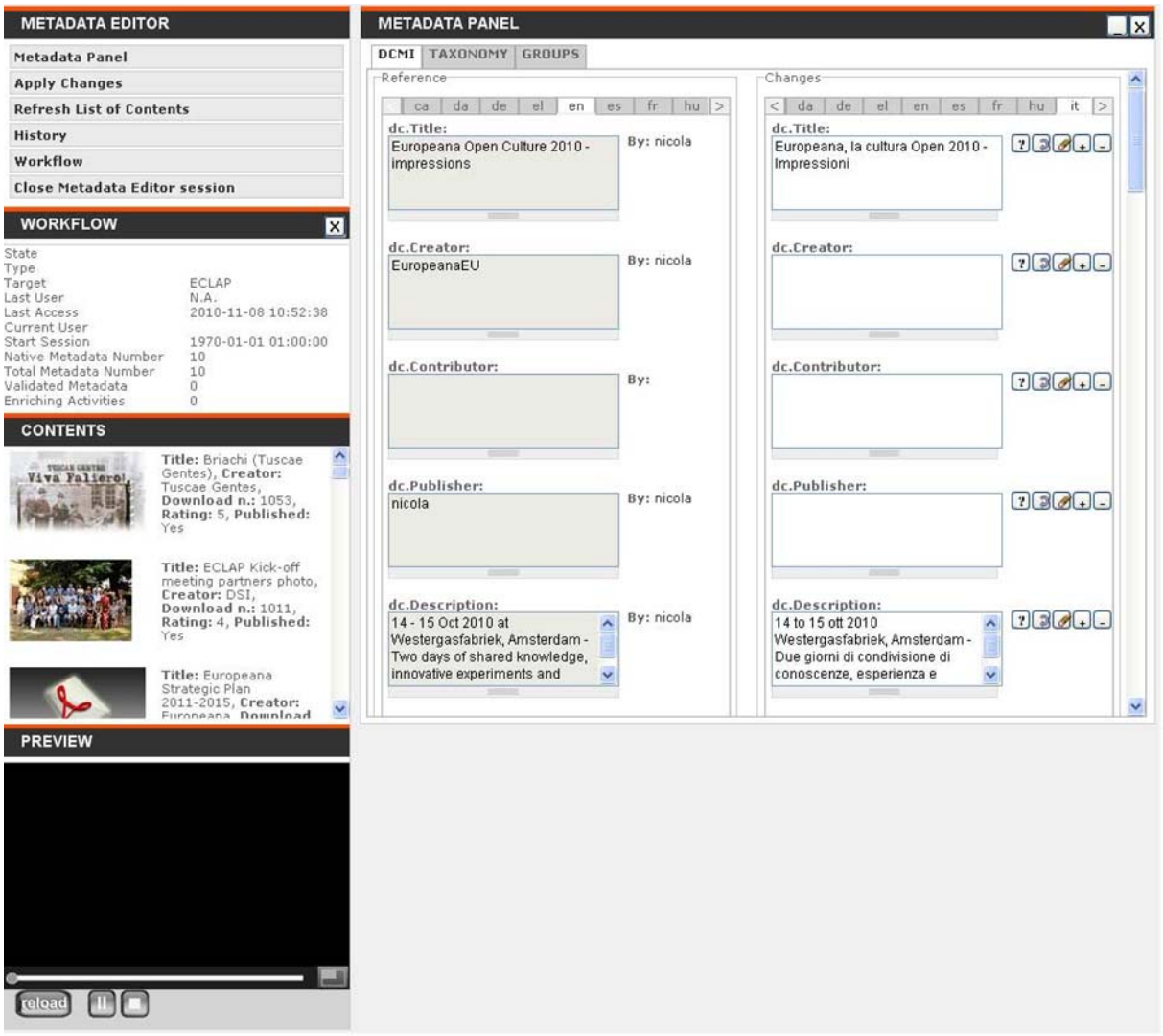
By clicking the **Edit Metadata** link present in the Edit Object form shown before, Metadata editor will be opened.

The metadata editor web page is structured as a 2 columns page.

The right column displays a set of blocks for:

- **Menu (Metadata Editor)**: it provides the access to main functionalities:
  - **Search** (available only for the administrator) will open/close a panel in the middle area with a search form for content searching
  - **Metadata Panel** opens/closes a panel in the middle area with all metadata divided into tabs
  - **Apply changes** sends to the server all made changes on metadata both update and modifications.
  - **Refresh List of Contents** opens/closes the block with all contents assigned by the workflow
  - **Workflow** opens/closes the block of workflow properties and info
  - **Close Metadata Editor Session** sends to server the end of activity to close the workflow session
- **Workflow**: this block displays workflow information.
- **Preview**: this block embeds the content player
- **Contents**: this block displays the list of content assigned by workflow

The following picture shows the layout of metadata editor page:



Metadata Panel

The panel shows the metadata edit area and is organized as tabs to organize metadata sets. Tabs that are used to browse among are:

- **DCMI**: Dublin core set
- **Taxonomy**: classification terms
- **Groups**: List of Groups associated with content and/or Public flag
- **Properties**: Technical set of metadata accessible obly by administrator.

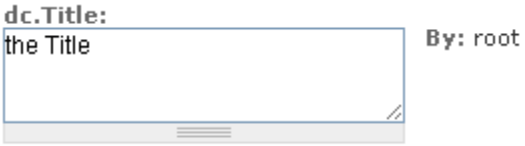
Each tab shows the list of metadata and iconized controls. To help the user each control has a tip with the name of action.

DCMI panel

The DCMI panel is divided into two sub-areas: Reference and Changes. Each of them displays a number of metadata panels according to user language capabilities. Panels are organized by means of language tabs that allow browsing on available languages in both views:

5.2.4.1

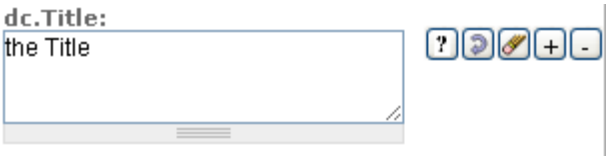
The Reference panels shows the current metadata values in read-only mode according to the selected language and for each of them the user name who made the last change/edit (By).



The Changes Panels allow the user to manually edit/translate/adding new metadata in the panel corresponding to the selected language.

Each metadata is displayed in a text box with a label reporting the metadata field name. Each metadata textbox has also a set of iconized buttons that allows performing:

- **Help**: It opens a popup dialog displaying the help/meaning for the metadata field
- **Undo**: Reset the value to the fist time displayed string
- **Clear**: Empty the text box
- **Add more**: it adds a new text box with own button.
- **Remove**: it removes the text box/metadata value

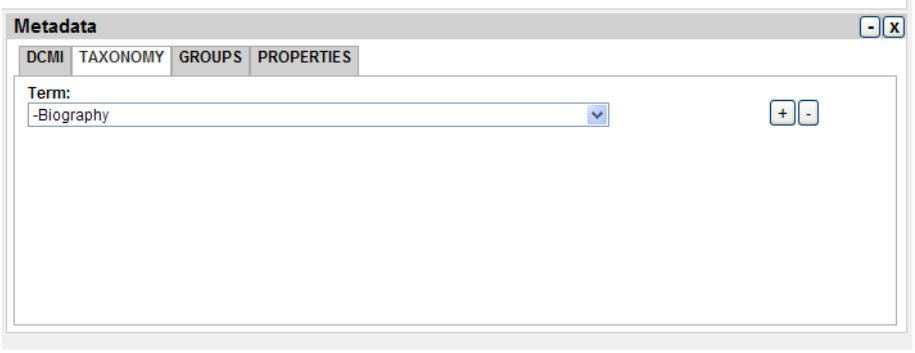


The add and remove buttons are available with multiple metadata.

During the enrichment every time the user makes a change or edits a new metadata, the text box background color becomes red to trace and easily recognize made changes.

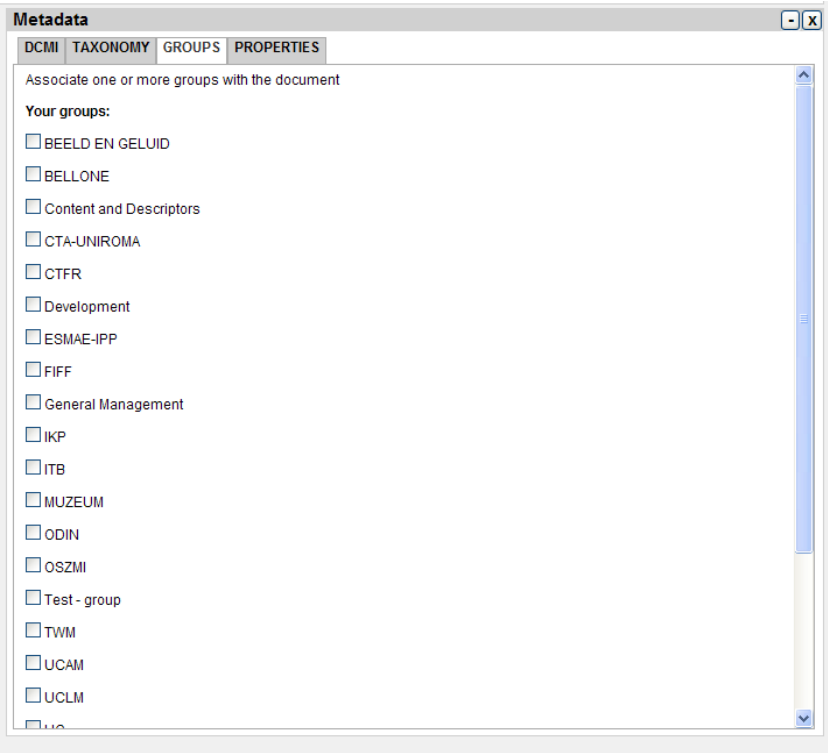
Taxonomy Metadata Panel

The Taxonomy panel allows the use to add/remove classification terms by means a text selector (single selection) and add/remove buttons. Every time an existing value will be changed by selecting a different one, the text selector will be red highlighted to mark that it has been changed/removed.



Groups Metadata Panel

The Groups panel displays the list of available groups. The selection/deselection is possible by enabling/disabling checkboxes. When a checked checkbox is disabled, it will be red highlighted to highlight the fact that it has been changed/removed.



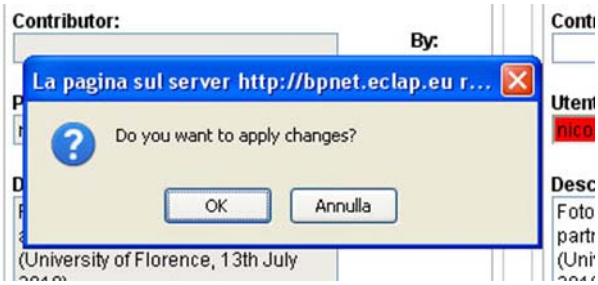
Workflow Block

This block displays a workflow report showing:

- **State**: the current state of content according the workflow status
- **Type**: the workflow model under usage (EclapWorkflow)
- **Target**: the workflow target of content (ECLAP, Europeana & ECLAP, Test, Internal)
- **Last access**: date and time of last editing
- **Last user**: who made the last editing session
- **Current user**: name of user who is currently editing the object
- **Start Session**: date and time of current editing session
- **Native Metadata Number**: the number of native metadata
- **Total Metadata Number**: the current number of metadata
- **Validated Metadata**: the current number of validated metadata
- **Enriching Activities**: how many time the content has been modified

Apply changes & Close Metadata Editor Session

When the process to update the metadata has been completed, the user has to click on the Apply changes item shown on the top left Menu. A confirmation window will be shown. Click OK to confirm changes.



The confirmation dialog will be shown in event of the user decides to end the session with some pending changes or by leaving the page (clicking the back button of browser) or clicking on Close Metadata Editor Session to end the session.

5.2.5 Technical metadata production

Technical metadata are collected during the upload phase and content production by the back-office environment. The digital resource is analysed to extract some technical data according to the metadata reported below:

Raw file: extension of file

Pcview: 1 or 0. When 1 is set, the content can be played by a pc,

Resolution: L or H depending by the original resource,

dx: width of video/image

dy: height of video/image

duration: duration of the fluid resource such as video, audio,

iphone: 1 or 0 if the resource has been generated for Iphone i.e. video

format: an eclap descriptor for format

type: an eclap descriptor for type

source: the URL where the original is stored

publisher: the partner that provided the content

Format and Type table

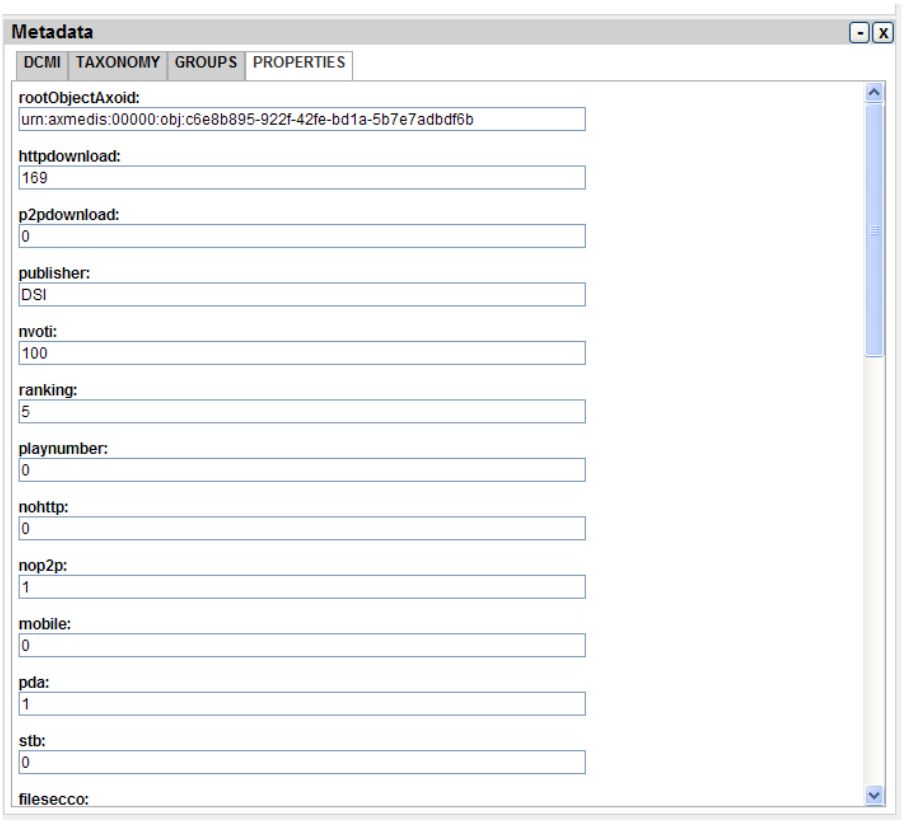
Format and Type descriptors are defined as following:

Format	Type
Audio	Audio
Video	video
Document	• document • html • pdf • slide • excel
Archive	Archive
Image	image
Crossmedia	• audio

	• crossmedia • document • flash • html • image • pdf • slide • smil • video
Tool	Tool
Playlist	Playlist
Collection	Collection

Meatadata Editor - Properties Metadata Panel

The ECLAP administrator can edit technical metadata by means the Metadata Editor through the **Properties** panel that displays all technical metadata as shown below. As for common metadata, the administrator after changes has to apply by clicking on the Apply Changes link.



6 Bibliography

[1] *Europeana Data Model presentation (for v5.2).* (http://version1.europeana.eu/c/document_library/get_file?uuid=76eff9ae-5a70-409c-87b0-baf46ede7bd9&groupId=10602).
[2] *EDM Data Model Primer.* (http://version1.europeana.eu/c/document_library/get_file?uuid=718a3828-6468-4e94-a9e7-7945c55eec65&groupId=10605).
[3] http://www.openarchives.org/pmh/
[4] http://www.mongodb.org/
[5] http://www.json.org/
[6] http://bsonspec.org/#/specification
[7] http://www.mongodb.org/display/DOCS/Drivers
[8] http://www.oclc.org/research/activities/oaicat/default.htm
[9] http://tools.ietf.org/html/rfc4287
[10] http://projects.apache.org/projects/abdera.html
[11] http://www.ietf.org/rfc/rfc5023.txt
[12] http://www.quartz-scheduler.org/
[13] Bagiwa, A.M. & Dishing, S.I. 2011. A Conceptual Framework For Extending Distance Measure Algorithm For Data Clustering. In «International Journal of Computer Trends and Technology», March-April issue. Available at: http://www.ijcttjournal.org/volume-1/issue-1/ijcttjournal-v1i1p12.pdf
[14] Belice Baltussen, L. e.a. 2010. ECLAP DE 2.1.1 – User Requirements and Use Cases.
[15] http://bpnet.eclap.eu/drupal/?q=en-US/home&axoid=urn:axmedis:00000:obj:87e9632e-ad1f-4ead-8303-997a3edfc28f&section=pagesupport
[16] Berners-Lee, T. & Fischetti, M. 1999. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor. Britain: Orion Business.
[17] Borelli, M. 2004. Le teste pensanti. In Borelli, M. & Savarese, N. 2004. Te@tri nella rete, Roma: Carocci, pp. 202-215.
[18] Brown, S. e.a. 2006. RePAH: Research Portals in the Arts and Humanities. A user analysis project. References Cited and Resources Consulted. Available at: http://repah.dmu.ac.uk/report/pdfs/RePAHReport-References.pdf.
[19] Brown, S. & Pickford, R. 2006. Assessing Skills and Practice. Routledge.
[20] European Commission 2010. A new Digital Agenda for Europe: 2015.eu. Bruxelles, European Parliament. text available at link http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2010-0133+0+DOC+XML+V0//EN&language=EN
[21] European Commission. Digital Agenda for Europe "2015.eu". Available at http://ec.europa.eu/information_society/newsroom/cf/pillar.cfm?pillar_id=43&pillar=Digital%20Single%20Market
[22] Eversmann, P. 2011. Performing Arts Education and Training: Starting Points. Available at http://bpnet.eclap.eu/drupal/?q=en-US/node/3161
[23] Gradmann, S. 2010. Knowledge = Information in Context on the Importance of Semantic Contextualisation in Europeana. White Paper 1. Available at http://www.version1.europeana.eu/c/document_library/get_file?uuid=cb417911-1ee0-473b-8840-bd7c6e9c93ae&groupId=10602
[24] Grindley, N. 2007. Digital Tools for Performance. London: AHRC ICT Methods Network. Available at: http://www.methodsnetwork.ac.uk/resources/wkp07.html.
[25] Thangadurai, K. e.a. 2010. A Study On Rough Clustering. In «Global Journal of Computer Science and Technology», Vol. 10, Issue 5.
[26] A Tutorial on Clustering Algorithms. Available at http://www.elet.polimi.it/upload/matteucc/Clustering/tutorial_html/

[27] Weinberger, David. Everything is Miscellaneous: The Power of the New Digital Disorder. New York: Henry Holt and Company, 2008.
[28] http://en.wikipedia.org/wiki/Taxonomy
[29] Hak Lae Kim e.a. 2008. The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies
[30] Ching-man Au Yeung e.a. Understanding the Semantics of Ambiguous Tags in Folksonomies
[31] Glen Farrelly, 2009. On Tags and Signs: A Semiotic Analysis of Folksonomies. Available at: https://fiq.ischool.utoronto.ca/index.php/fiq/article/view/20/64
[32] Velsen, Lex van & Mark Melenhorst. "Incorporating user motivations to design for video tagging." *Interacting with Computers*, vol.21 no.3 (juli 2009), p. 221 – 232.
[33] http://en.wikipedia.org/wiki/Tag_cloud
[34] http://drupal.org/
[35] P. Bellini, I. Bruno, D. Cenni, P. Nesi, "Micro grids for scalable media computing and intelligence on distributed scenarious", IEEE Multimedia, in press, IEEE Computer Soc. Press.
[36] P. Bellini, I. Bruno, D. Cenni, A. Fuzier, P. Nesi, M. Paolucci, "Mobile Medicine: Semantic Computing Management for Health Care Applications on Desktop and Mobile Devices", in press on Multimedia Tools and Applications, Springer.
[37] P. Bellini, I. Bruno, P. Nesi, "EXPLOITING INTELLIGENT CONTENT VIA AXMEDIS/MPEG-21 FOR MODELLING AND DISTRIBUTING NEWS", International Journal of Software Engineering and Knowledge Engineering, World Scientific Publishing Company, in press.
[38] P. Bellini, D. Cenni, A. Fuzier, P. Nesi, G. Pantaleo, M. Paolucci, "Metrics for best practice networks analysis", proc. of DMS 2010: Globalization and Personalization, The 16th International Conference on Distributed Multimedia Systems, Hyatt Lodge at McDonald's Campus, Oak Brook, Illinois, USA, October 14 to 16, 2010.
[39] E. Bellini, M. A. Deussom, P. Nesi, "Assessing Open Archive OAI-PMH implementations", proc. of DMS 2010: Globalization and Personalization, The 16th International Conference on Distributed Multimedia Systems, Hyatt Lodge at McDonald's Campus, Oak Brook, Illinois, USA, October 14 to 16, 2010.
[40] P. Bellini, I. Bruno, P. Nesi, D. Rogai, ''MPEG-21 for the multi-format and multi-channel content production, protection and distribution'', Proc. Of the International Broadcasting Conference, IBC 2007, 2007, Amsterdam, September 2007.
[41] P. Bellini, P. Nesi, D. Rogai, ''Exploiting MPEG-21 File Format for cross media content'', Proc. Of the International Conference on Distributed Multimedia Systems, DMS 2007, September 6-8, 2007, San Francisco Bay, USA, Organised by Knowledge Systems Institute.

7 Glossary

Assessment	The process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs. Assessments can be classified in many different ways. The most important distinctions are: - formative, generally carried out throughout a course or project; - summative, generally carried out at the end of a course or project; - referencing (criterion-referenced, norm-referenced, and ipsative); - informal, usually occurs in a more casual manner and may include observation, inventories, checklists, rating scales, rubrics, performance assessments, participation, peer and self evaluation, and discussion. - formal, usually implicates a written document.
Artistic movements and acting styles	An artistic movement has a common philosophy or goal regarding performing arts. Within these philosophies, specific acting styles are developed.
Clustering	A technique to group together a set of items having similar characteristic
Content type	The content type does not specifically indicate what the content is about, but rather indicates the form of the content. For instance, a video about the genre burlesque can be a documentary about burlesque or a registration of a burlesque performance. All content types are a sub-category of one or more digital object types (Book is a subcategory of the digital object type Text, Interview can both be a sub- category of Video, but also of Audio).
Performing Arts	Forms of creative activity that are performed in front of an audience, such as drama, music, and dance. (Oxford Dictionary of Current English, 2009).

Annex I

Content Survey released to the ECLAP mailing list on March 28th.

2. Content grouping

In the previous general survey each partner already divided its own content in a number of content sets, but the kinds of parameters used to group the content were not homogeneous. We are therefore asking you to rearrange your content sets according, from time to time, to the following criteria: "geography and genre" or "people" (remember that you should describe only the specific collections or that parts of the collections that you will aggregate in ECLAP and not all the collections your institution holds). After individuating each

2.1 Content sets named after the geographical location (where the event took place) and genre/artistic style. Please group your items in different sets dividing them by geographical location and genres or artistic styles, choosing the genre or style among the terms listed below, if none of them is applicable to your collection, please indicate a new term.

Genre/artistic style				
Acrobatic	Environmental theatre	Kagura	Opera, European	Spectacle
Agit-Prop	Epic (Brechtian)	Kathakali	Operetta	Spiritual
Attic comedy	Existentialism	Kolyiattam	Orchestral Music	Sport
Austruckstanz	Experimental theatre	Kyogen	Pantomime	Street dance
Avant-garde	Expressionism	Liturgical drama	Performance art	Street
Baroque dance	Fairytale	Modern dance	Pop music	Surrealism
Baroque theatre	Farce	Musical theatre	Postdramatic	Tango
Baroque opera	Feast	Madrigals	Postmodern	Tanztheater
Ballet	Flamenco	Magician show	Punk music	Terukuttu
Ball room	Fluxus	Melodrama	Puppetry	Theatre of
Body art	Folk	Minstrelsy	Ragtime music	Theatre of the
Bugaku	Futurism	Miracle play	Rakugo	Theatre of the
Burlesque	Greek tragedy revival	Mistery play	Rap music	Toshi Kyogen
Butho	Grotesque	Model opera	Realism	Tragedy,
Chakkiarkuttu	Happenings	Modern	Reggae music	Training
Chamber music	Heikyoku	Multimedia	Ritual	Trance
Chaoju	Impressionism	Naturalism	Rock and Roll	Trauerspiel
Circus arts	Intermediality	New age music	Romanticism	Vaudeville
Classical music	Jazz dance	New Cinema	Satyrical	Videoart
Classicism	Jazz music	New Media	Science-art	Video Dance
Comedy	Jesters	New wave music	Symbolism	Water
Commedia dell'Arte	Jingju	Nouvelle Vague	Sideshow	Xiqu (Chinese
Contemporary	Juggling	Novel	Silent film	Yakshagana
Crossover	Kabuki Buyo	No wave	Soul music	Yueju
Digital Performance		Noh theatre	Special event, multi-genre	
Drama		Nuo		

Please fill in the grid below with the name you created for your content set (geographic indication+genre). Also indicate how many items each set consists of. If you need more rows, please add them.

Examples: not "Puppets" but "Catalan Puppetry"; not "Students concerts" but "Portuguese classical music"; not "Dance, theatre, drama, puppetry, music, comedy" but "Dutch contemporary dance", "Dutch drama",

Content set's name (geographic indication+genre)	Number of items

2.2 Content sets named after the main author, master or "protagonist" (the person the collection is about). Please group your items in different content sets naming them after the person the collection is about (an actor, a composer, a director, a writer, an author, etc.).

Examples: Wolfgang Amadeus Mozart, Samuel Beckett, Jerzy Grotowski, Ariane Mnouchine, Dario Fo, Franca Rame, Jerzy Grotowsky, Sergej Prokoviev, Gian Maria Volonté, Maya Dehren, Manuel De Olivera, Richard Burton, Pina Bausch, Eleonora Duse, Sarah Bernhadt, etc.

Please also indicate how many items each set consists of. If you need more rows, please add them.

Content set's name (person's name)	Number of items

2.3 Please insert each of your content sets in the following grid according to time span. If you need more rows for content sets dated prior to 1850 please add rows.

Time span	Collections names
.....	
1850-1870	
1870-1890	
1890-1910	
1910-1930	
1930-1950	
1950-1970	
1970-1990	
1990-2010	
2010-2011	

3. Types

3.1 Performing arts types. Please indicate the number of items you will provide related to the following performing arts: dance, music, theatre, cinema.

Performing arts types	Number of items
Dance (including: folkdance, ballet, ethnic)	
Theatre (including: opera, amateur performance, puppet theatre, poetry, happenings, rituals)	
Music (classical, rock, castanets, chamber music, orchestral, scores)	
Cinema (only feature films, short films, animated film, variety show, videoart, new media)	

4.2 Main characteristics. Please provide in few lines an overall description for all the content you will provide to ECLAP highlighting the peculiarities of your collections.

4.3 Remarks and comments on your items (others from what already indicated in the Technical Questionnaire) if you have any.
