



ICT-PSP Project no. 297158
EUROPEANAPHOTOGRAPHY
EUROPEAN Ancient PHOTOgraphic vintaGe repositoRies of digitAized Pictures of
Historical qualitY

Starting date: 1st February 2012

Ending date: 31st January 2015

Deliverable Number:	D 5.2
Title of the Deliverable:	The MINT Mapping Tool
Dissemination Level:	Restricted

Contractual Date of Delivery to EC:	Month 09
Actual Date of Delivery to EC:	October 2012
Second Release:	19 April 2013

Project Coordinator

Company name : KU Leuven
Name of representative : Fred Truyen
Address : Blijde-Inkomststraat 21 B-3000 Leuven PB 3301
Phone number : +32 16 325005
E-mail : fred.truyen@kuleuven.be
Project WEB site address : <http://www.europeana-photography.eu>

Context

WP 5	Metadata Transformation and Ingestion
WP Leader	NTUA – National Technical University of Athens
Task 5.3, 5.4	Source to Intermediate, Ingestion to EDM
Task Leader	NTUA – National Technical University of Athens
Dependencies	

Author(s)	Nikolaos Simou (NTUA)
Contributor(s)	Nasos Drosopoulos (NTUA),
Reviewers	Andrea de Polo, Sam Minelli, Leonardo Lemmi (Alinari), David Iglesias (CRDI)
Reviewers of second release	David Iglésias Franch (CRDI); Evgeni Dimitrov (NALIS)
Approved by:	

History

Version	Date	Author	Comments
0.1	24 Oct. 12	Nikolaos Simou, Nasos Drosopoulos	First Draft
0.2	27 Oct. 12	Andrea de Polo, Sam Minelli, Leonardo Lemmi	First Review
0.3	27 Oct. 12	David Iglesias	Second Review
1	30 Oct. 12	Nikolaos Simou	Final Version
1.1 (second release)	4 Apr. 13	Nikolaos Simou	Updated according to EC reviewers' comments
1.2 (second release)	18 Apr. 2013	Nikolaos Simou	Updated according to peer reviewers' comments
2.0 (second release)	18 Apr. 2013	Valentina Bachi, all consortium	Final reading

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	6
1 INTRODUCTION	8
1.1 BACKGROUND	8
1.2 ROLE OF THIS DELIVERABLE IN THE PROJECT	9
1.3 APPROACH	9
1.4 STRUCTURE OF THE DOCUMENT	10
2 MINT MAPPING TOOL.....	11
3 MAPPING EDITOR.....	12
4 MINT SPECIALIZATIONS FOR EUROPEANA PHOTOGRAPHY.....	14
4.1 BACKEND RECONSTRUCTION.....	14
4.2 USER INTERFACE	14
4.2.1 <i>Aggregation workflow guided interface.....</i>	<i>14</i>
4.2.2 <i>Improved Schema browsing.....</i>	<i>16</i>
4.3 MAPPING FUNCTIONALITY	17
4.3.1 <i>Enhanced mapping functionalities.....</i>	<i>17</i>
4.3.2 <i>SKOS Vocabularies Support.....</i>	<i>17</i>
5 USER MANUAL.....	18
5.1 USER & ORGANIZATION REGISTRATION	18
5.1.1 <i>User Registration</i>	<i>18</i>
5.1.1.1 Self-registration from MINT Mapping Tool.....	18
5.1.1.1.1 User registration without an organization.....	20
5.1.1.1.2 User registration under an existing organization.....	20
5.1.1.2 User management (& User creation under an organization by its administrator).....	20
5.1.2 <i>Organization Registration & Management</i>	<i>21</i>
5.2 IMPORT OF METADATA	23
5.2.1 <i>Dataset Options</i>	<i>25</i>
5.2.1.1 Wrong Import	25
5.2.1.2 Successful import.....	26
5.2.1.2.1 Define Items	26
5.2.1.2.2 Show Items.....	29
5.2.1.2.3 Dataset statistics	30
5.3 MAPPING	30
5.3.1 <i>Create new mapping</i>	<i>31</i>
5.3.2 <i>Upload mapping</i>	<i>35</i>
5.3.3 <i>Upload XSL.....</i>	<i>35</i>
5.3.4 <i>Mappings Management.....</i>	<i>36</i>
5.3.5 <i>Mappings.....</i>	<i>37</i>
5.3.5.1 Mapping environment.....	37
5.3.5.2 Constant value mapping.....	39

5.3.5.3	Concatenate mapping	39
5.3.5.4	Value mapping	40
5.3.5.5	Functional mapping	40
5.3.5.6	Conditional mapping.....	41
5.3.5.7	Structural mapping	42
5.3.5.8	Mapping Using Terminologies	42
5.3.6	<i>Preview items</i>	44
5.4	TRANSFORMATION - PUBLICATION TO EUROPEANA.....	46
6	MINT SUPPORT	49
7	TECHNICAL SPECIFICATIONS	50
7.1	PLATFORM	50
7.2	INGESTION	50
7.3	PROCESSING	50
7.4	NORMALIZATION & VOCABULARIES	50
8	CONCLUSION	51
8.1	RESULTS.....	51
8.2	IMPACT.....	51
	APPENDIX I: DEFINITIONS OF TERMS AND ABBREVIATIONS	52
	APPENDIX II: MAPPING OF EUROPEANA PHOTOGRAPHY MANDATORY FIELDS TO LIDO	53

EXECUTIVE SUMMARY

This deliverable describes the MINT mapping tool that is used in the Europeana Photography project to assist content providers in the transformation of their in house metadata to the intermediate standard selected for the project (see D 5.1) and its publication to Europeana. MINT development has started by NTUA from the ATHENA¹ ingestion server and evolved through other projects like Linked Heritage², EuScreen³ and ECLAP⁴. It follows a typical web-based architecture offering an expanding set of services for metadata aggregation. It addresses the ingestion of metadata from multiple sources, the mapping of the imported records to the intermediate metadata schema and the transformation and storage of the metadata in a repository.

The main role of the MINT mapping tool in the Europeana Photography project is to enable users to

- Provide metadata records in a range of “source” formats
- Convert metadata to the Europeana Photography’s intermediate standard
- Map local terminologies to the adopted reference terminologies
- Submit the records to Europeana

while its key functionalities include:

- Organization and user level access rights and role assignment.
- Collection and record management (XML serialisation).
- Direct import and validation according to registered schemas (XSD).
- OAI-PMH based harvesting and publishing.
- Visual mapping editor for the XSLT language.
- Transformation and previewing (XML and HTML).
- Repository deployment and remediation interfaces.

MINT allows providers to perform mappings from their schemas to the intermediate schema of Europeana Photography through a very user-friendly interface (see figure below).

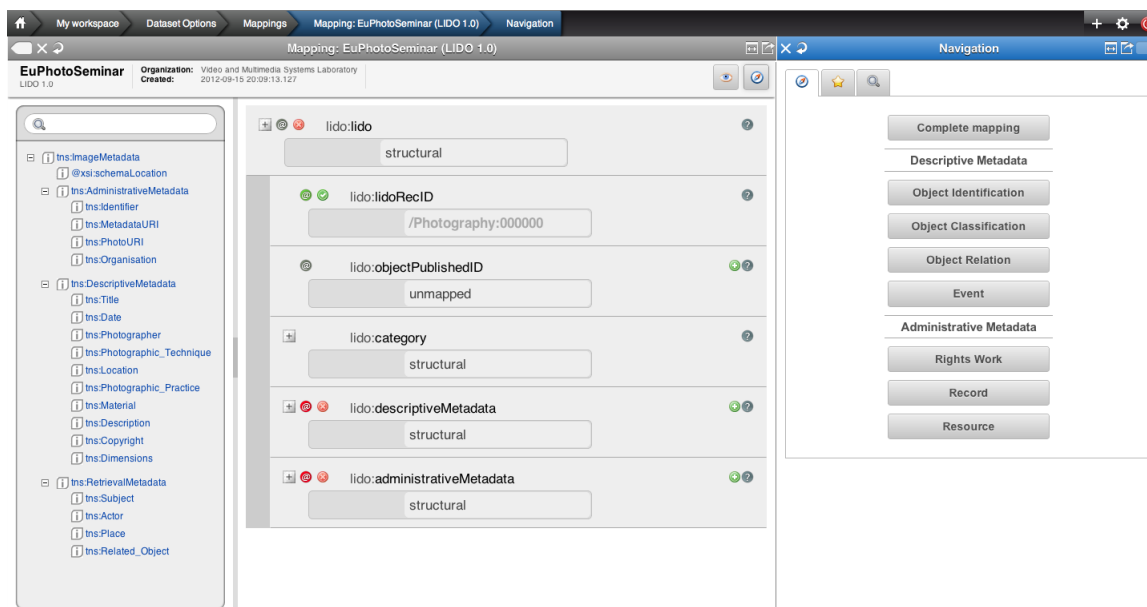


Figure 1: MINT's mapping editor

¹ <http://www.athenaeurope.org/>

² <http://www.linkedheritage.eu/>

³ <http://euscreen.eu/>

⁴ <http://www.eclap.eu/drupal/?q=en-US>

The structure that corresponds to a user's specific import is visualized in the mapping interface as an interactive tree that appears on the left hand side of the editor. The tree represents the snapshot of the XML schema that is used as input for the mapping process. The user is able to navigate and access element statistics and also to search the tree by using the text field on the top.

On the right hand side, buttons correspond to high-level elements of the target schema and are used to access their corresponding sub-elements. These are visualized on the middle part of the screen as a tree structure of embedded boxes, representing the internal structure of the complex element. The user is able to interact with this structure by clicking to collapse and expand every embedded box that represents an element, along with all relevant information (attributes, annotations) defined in the XML schema document. To perform an actual (one to one) mapping between the input and the target schema, a user has to simply drag a source element from the left and drop it on the respective target in the middle.

The resulting repository offers an OAI-PMH interface currently exposing the records in the Europeana Semantic Elements schema model that will be changed to EDM -according to the project requirements- by the end of M15. After the creation of a valid mapping, content providers can transform their metadata to the intermediate schema. This action is followed by the publication to NTUA's OAI-PMH server during which metadata are transformed to EDM (currently ESE, because the implementation of a crosswalk from LIDO to EDM is in progress). Finally, publication to Europeana is performed by informing Europeana's Ingestion office to harvest metadata from the NTUA's server. The overall workflow that a provider should follow for the publication of his metadata to Europeana is shown in the figure below.



Figure 2: Overall workflow

The MINT-mapping tool has been used in many European projects aggregating content to Europeana, however many specializations have been implemented to meet the special needs of the project's content providers. More specifically a new module for processing the XML imports has been implemented considering the experience gathered from the usage of MINT in other projects and also the feedback from content providers. This new metadata-processing module is responsible for the itemization of the imported metadata, functionality that was not available in MINT releases used for other projects and improves the overall scalability. Another major development in the MINT release used for the Europeana Photography is its new user interface. Our main objective was to redesign the MINT's mapping tool user interface in such a way that would permit to the various content providers to easily use it and to better understand the overall workflow towards Europeana. Finally, new mapping functionalities have been developed allowing providers to perform mappings using SKOS vocabularies and also to have better control of their metadata mappings based on their values.

The delivery of the platform and the present report achieve a project milestone, MS13 MINT Mapping Tool available, Month 9. This enabled providers to start processing metadata and towards Indicator No 5 "1,000 photographs successfully delivered to Europeana".

1 INTRODUCTION

This document reports on deliverable D5.2 “The MINT mapping tool”, that is made available online for validation and for the large-scale contribution of content from EuropeanaPhotography project to Europeana and for dissemination & training (WP7, WP5). The NTUA team integrated all the necessary components into a common technology platform, starting from the basis of the ATHENA⁵ ingestion server and evolving through other projects like Linked Heritage⁶, EuScreen⁷ and ECLAP⁸. MINT mapping tool provides content holders with the ability to perform in an efficient way the required mapping of their own metadata schemas to the project’s intermediate metadata schema (see D 5.1), as well as its publication to Europeana. It follows a typical web-based architecture offering an expanding set of services for metadata aggregation. It addresses the ingestion of metadata from multiple sources, the mapping of the imported records to the intermediate metadata schema and the transformation and storage of the metadata in a repository. Although its deployment is also guided by expediency, the system has been developed using established tools and standards, embodying best practices in order to animate familiar content provider procedures in an intuitive and transparent way also for newcomers.

1.1 BACKGROUND

Metadata records are critical to the documentation and maintenance of interrelationships between information resources and are being used to find, gather, and maintain resources over long periods of time. The consistent application of a descriptive metadata standard improves the user’s search experience and makes information retrieval within a single collection or across multiple datasets more reliable. Descriptive, administrative, technical, and preservation metadata contribute to the management of information resources and help to ensure their intellectual integrity both now and in the future. In parallel with other domains, many researchers in the digital cultural heritage community recognized the need to lower the barriers for the management and aggregation of digital resources, by implementing some measure of interoperability among metadata standards and then with proprietary data structures. There is a wide range of proposed solutions, including crosswalks, translation algorithms, metadata registries, and specialized data dictionaries.

A crosswalk provides a mapping of metadata elements from one metadata schema to another. The prerequisite to a meaningful mapping requires a clear and precise definition of the elements in each schema. The primary difficulty is to identify the common elements in different metadata schemas and put this information to use in systems that resolve differences between incompatible records. Crosswalks are typically presented as tables of equivalent elements in two schemas and, even though the equivalences may be inexact, they represent an expert’s judgment that the conceptual differences are immaterial to the successful operation of a software process that involves records encoded in the two models. A crosswalk supports the ability of a retrieval mechanism to query fields with the same or similar content in different data sources; in other words, it supports semantic interoperability.

Crosswalks are not only important for supporting the demand for single point of access or cross-domain searching; they are also instrumental for converting data from one format to another. However, aggregating metadata records from different repositories may create confusing display results, especially if some of the metadata was automatically generated or created by institutions or individuals that did not follow best practices or standard thesauri and controlled vocabularies. Mapping metadata elements from different schemas is only one level of cross walking. Another level of semantic interoperability addresses datatype registration and formatting of the values that populate the metadata elements, e.g. rules for

⁵ <http://www.athenaeurope.org/>

⁶ <http://www.linkedheritage.eu/>

⁷ <http://euscreen.eu/>

⁸ <http://www.eclap.eu/drupal/?q=en-US>

recording personal names or encoding standards for dates, and the alignment between local authority files and adopted terminologies.

The MINT-mapping tool implements an aggregation infrastructure offering a crosswalk mechanism to support subsequent critical activities:

- harvesting and aggregating metadata records that were created using shared community standards or proprietary metadata schemas,
- migrating from providers' models (whether standard or local) to a reference model,
- transforming records from the project's intermediate standard to the Europeanana Semantic Elements and the Europeanana Data Model.

1.2 ROLE OF THIS DELIVERABLE IN THE PROJECT

This deliverable corresponds to the MINT mapping tool that implements Tasks 5.3 "Source to intermediate" and 5.4 "Intermediate to EDM and Ingestion". The software services are employed for the aggregation of metadata in order to realize the content delivery to Europeanana as it is planned and in the DoW. MINT mapping tool enables users to

- Provide metadata records in a range of "source" formats
- Convert metadata to the Europeanana Photography's intermediate standard
- Map local terminologies to the adopted reference terminologies
- Submit the records to Europeanana

Specifications are informed by the adoption of terminologies and development of the terminology system designed in WP4. The delivery of the system is related to Task 6.2 "Sustainability Planning" while it is a project milestone (MS13 MINT Mapping Tool available, Month 9) and constitutes the starting point for mapping and delivery of metadata records to Europeanana.

1.3 APPROACH

Work for this deliverable was carried out by the NTUA team, that integrated all the necessary components into a common technology platform starting from the basis of the ATHENA ingestion server, using the Europeanana metadata aggregator requirements. MINT services compose a web based platform that was designed to facilitate aggregation initiatives for cultural heritage content and metadata in Europe. It is employed from the first steps of such workflows, corresponding to the ingestion, semantic alignment and aggregation of metadata records, and proceeds to implement a variety of remediation approaches.

The MINT-mapping tool has been deployed for a variety of aggregation workflows corresponding to the whole or parts of the backend services. Specifically, it has served the aggregator of museum content for Europeanana (and one of the largest in volume and significance), the ATHENA project, that has ingested from 135 organizations over 4 million items, which users aligned to the LIDO format. The resulting repository offers an OAI-PMH interface exposing the records in the Europeanana Semantic Elements schema. The use of a reference model allowed the rapid support of updated ESE versions that were introduced in the duration of the project (2008-2011), with minimal input from providers. User efforts to align their data to an adopted domain model motivated them to update their collection management systems and improve the quality of their annotations in order to take advantage of a well defined, machine understandable model and, subsequently, control and enrich their organization's contribution and visibility through Europeanana.

The EUscreen project also follows the same aggregation workflow for Europeanana while, in addition, it provides a portal for Europe's television heritage where both the video content and metadata records are offered to users. The metadata records for the portal are based on the selected reference models (EUscreen and EBUcore) for which an item annotator guide was introduced. MINT serves the

aggregation and remediation of records both for the portal (also offering the Lucene indexes for the search engine) and Europeanana (OAI-PMH for ESE).

Similarly, the CARARE, JUDAICA Europeanana, ECLAP, DCA, LinkedHeritage and Europeanana Photography projects utilize MINT to accommodate their aggregation and remediation requirements for their specific domain and project, and for Europeanana. The group is actively participating in several metadata modeling activities such as LIDO, where the tool is used for the presentation and revision of the LIDO schema by the corresponding CIDOC working group, EDM, for the prototyping of the Europeanana Data Model harvesting XSD and RDFS ontology, and EBU Tech. It is also involved in the development of Europeanana and the Digital Public Library of America. The growing user base (more than 350 cultural heritage organizations and 500 users) contributes to its ongoing development, improvement and support, while the first version, MINT-Athena, was released under an open source license in July 2011.

During the first nine months of the project the intermediate schema of the project was defined (LIDO). In addition the workflow for the project's content delivery and modeling requirements concerning LIDO and related terminologies were put in place by the respective work packages. WP5 implemented these specifications in the technology platform.

The platform is hosted and administered by NTUA, users can register and access the services at <http://mint-projects.image.ntua.gr/photography>

1.4 STRUCTURE OF THE DOCUMENT

The rest of the document presents more details on the aggregation workflow that is implemented in the Europeanana Photography Technology Platform (Section 2), emphasising on the metadata mapping tool (Section 3). The following section (Section 4) presents the MINT specializations that were implemented for meeting the requirements of the Europeanana Photography project. In Section 5 a user manual of the MINT-mapping tool is presented, followed by a section that explains to providers how to get support on MINT-related issues, while detailed technical specifications for the implementation are given in Section 7.

2 MINT MAPPING TOOL

The key concept behind the aggregation part of the system has been that, although 'low-barrier' standards such as Dublin Core were used in the first stages of Europeana (ESE data model) to reduce the respective effort and cost, to also support richer and better-defined models. Moreover, since the technological evolution of consuming services for cultural heritage is greater than that of most individual organizations, a richer schema like LIDO allows harvesting and registering of all annotation data regardless of the current technological state of the repositories or their intended (re-) use.

The developed system facilitates the ingestion of semi-structured data and offers the ability to establish crosswalks to the reference schema in order to take advantage of a well-defined, machine understandable model. The underlying data serialization is in XML, while the user's mapping actions are registered as XSL transformations. The common model functions as an anchor, to which various data providers can be attached and become, at least partly, interoperable.

Key functionalities include:

- Organization and user level access rights and role assignment.
- Collection and record management (XML serialisation).
- Direct import and validation according to registered schemas (XSD).
- OAI-PMH based harvesting and publishing.
- Visual mapping editor for the XSLT language.
- Transformation and previewing (XML and HTML).
- Repository deployment and remediation interfaces.



Figure 3: Ingestion workflow

The metadata ingestion workflow, as illustrated in *Figure 3*, consists of four main steps. First is the Harvesting/Delivery procedure, which refers to the collection of metadata from content providers through common data delivery protocols, such as OAI-PMH, HTTP and FTP. Following is the Schema Mapping procedure, during which the harvested metadata are mapped to the common reference model. A graphical user interface assists content providers in mapping their metadata structures and instances to a rich, well defined intermediate schema (LIDO), using an underlying machine-understandable mapping language. Furthermore, it provides useful statistics about the provider's metadata while also supporting the share and reuse of metadata crosswalks and the establishment of template transformations. The third step is the Transformation procedure, which also aims at the transformation of the content provider's list of terms to the vocabularies and terminologies introduced by the reference model. The last step is the Europeana publication procedure, during which metadata are transformed from the intermediate schema to EDM -according to the project requirements- and stored to the NTUA's OAI-PMH server. Currently records are stored in ESE because the implementation of a crosswalk from LIDO to EDM is done as part of a collaboration of projects Linked Heritage, Europeana Photography and Partage Plus and is in progress (it will be finalized by M15). Publication to Europeana is then performed by informing Europeana's Ingestion office to harvest metadata from the NTUA's server.

3 MAPPING EDITOR

Metadata mapping is the crucial step of the ingestion procedure. It formalizes the notion of a metadata crosswalk, hiding the technical details and permitting semantic equivalences to emerge as the centerpiece. It involves a user-friendly graphical environment (*Figure 4* shows an example mapping opened in the editor) where interoperability is achieved by guiding users in the creation of mappings between input and target elements. User imports are not required to include the respective schema declaration, while the records can be uploaded as XML or CSV files. User's mapping actions are expressed through XSLT style sheets, i.e. a well-formed XML document conforming to the namespaces in XML recommendation. XSLT style sheets are stored and can be applied to any user data, exported and published as a well-defined, machine understandable crosswalk and, shared with other users to act as template for their mapping needs.

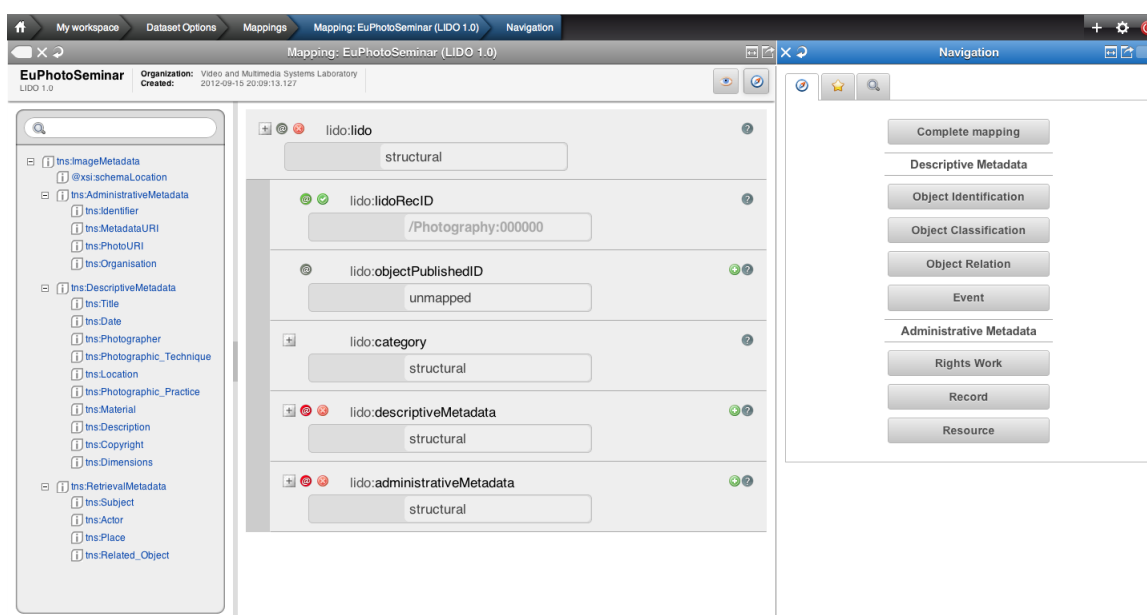


Figure 4: Screenshot of the mapping editor

The structure that corresponds to a user's specific import is visualized in the mapping interface as an interactive tree that appears on the left hand side of the editor. The tree represents the snapshot of the XML schema that is used as input for the mapping process. The user is able to navigate and access element statistics for the specific import while the set of elements that have to be mapped can be limited to those that are actually populated. The aim is to accelerate the actual work, especially for the non-expert user, and to help overcome expected inconsistencies between schema declaration and actual usage.

On the right hand side, buttons correspond to high-level elements of the target schema and are used to access their corresponding sub-elements. These are visualized on the middle part of the screen as a tree structure of embedded boxes, representing the internal structure of the complex element. The user is able to interact with this structure by clicking to collapse and expand every embedded box that represents an element, along with all relevant information (attributes, annotations) defined in the XML schema document. To perform an actual (one to one) mapping between the input and the target schema, a user has to simply drag a source element from the left and drop it on the respective target in the middle.

The user interface of the mapping editor is schema aware regarding the target data model and enables or restricts certain operations accordingly, based on constraints for elements in the target XSD. For example, when an element can be repeated then an appropriate button appears to indicate and

implement its duplication. Several advanced mapping features of the language are accessible to the user through actions on the interface, including:

- String manipulation functions for input elements.
- m-1 mappings with the option between concatenation and element repetition.
- Structural element mappings.
- Constant or controlled value assignment.
- Conditional mappings (with a complex condition editor).
- Value mappings editor (for input and target element value lists)

Mappings can be applied to ingested records, edited, downloaded and shared as templates. Preview interfaces (Figure 5) present the steps of the aggregation such as the current input xml record, the XSLT code of mappings, and the transformed record in the target schema, subsequent transformations from the target schema to other models of interest, and available html renderings of each xml record. Users can transform their selected collections using complete and validated mappings in order to publish them in available target schemas for the required aggregation and remediation steps.

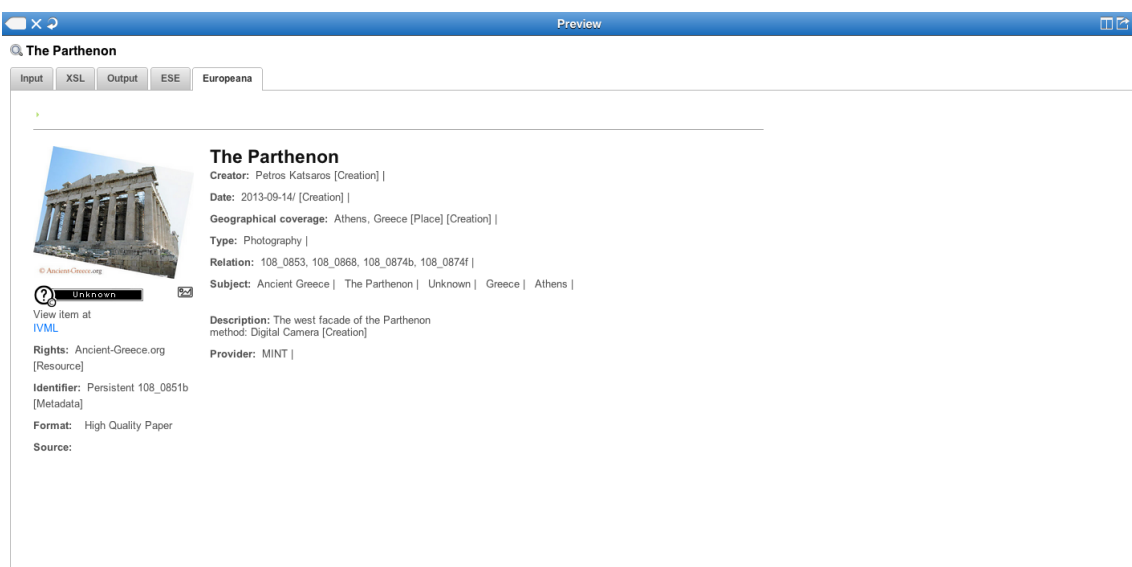


Figure 5: Transformation and HTML rendering preview

4 MINT SPECIALIZATIONS FOR EUROPEANA PHOTOGRAPHY

The previous sections presented the overall workflow as well as the main functionality of the MINT mapping tool. This description of MINT is also available in deliverables authored by the NTUA team for other European aggregation projects like Europeana Photography such as

- Athena D 7.1 “First version of the semantic interoperability plug-ins”,
- ECLAP D 4.1 “Metadata descriptors Identification and Definition”,
- Linked Heritage D 5.1 “Linked Heritage Technology Platform”,

but it still remains vital for the content providers of this project.

However even if the main functionality of the MINT-mapping tool -that is the transformation of the metadata extracted from the provider’s metadata management systems in various standards to an intermediate metadata standard and then its delivery to Europeana according to the project specific requirements- remains the same, many specializations are implemented to meet the special needs of the content providers for each project. In this section we present the specializations that were implemented for the Europeana Photography project.

4.1 BACKEND RECONSTRUCTION

One of the most important developments that are implemented for the Europeana Photography project is the major reconstruction of the MINT’s backend platform. More specifically a new module for processing the XML imports has been implemented considering the experience gathered from the usage of MINT in other projects and also the feedback from content providers. The main problem that content providers experienced in MINT versions used in other projects was its non-scalable behavior, especially in operations like item previewing and dataset statistics.

This new metadata-processing module is responsible for the itemization of the imported metadata, functionality that was not available in MINT releases used for other projects. In detail this module improves the overall backend scalability because

- Items of an import are pre-calculated and computational time is saved by preventing the dynamic fetch of items.
- Items statistics are calculated while import again saving computational time by preventing its dynamic calculation.
- No indexing of the imported metadata is performed – as it was in the previous MINT versions – cause it proved to be time costly and without any useful functionality.

4.2 USER INTERFACE

Another major development in the MINT release used for the Europeana Photography version is its new user interface. Our main objective was to redesign the MINT’s mapping tool user interface in a way that would permit to the various content providers to easily use it and to better understand the overall workflow towards Europeana. The following subsections present in more detail the new interface and how this is guided by the aggregation workflow and also new ways of browsing the input and target schema.

4.2.1 Aggregation workflow guided interface

The user interface of the MINT release used in previous Europeana-feeder projects like Athena, Linked Heritage, ECLAP, EuScreen and others was implemented using the Javascript library YUI⁹. The interface approach followed there was tab based as shown below. In simple words the content provider could navigate to MINT’s functionalities through the different tabs that appeared on the top of the window. The

⁹ <http://yuilib.com/>

main problem with that approach however was that actions were not related and the users did not have any guidance through the aggregation process. Therefore the fact that a provider first had to import his metadata by using the import preview, then to click on the Overview preview to locate its import and after that to perform mappings and transformations was not clear, and even after training sessions many of the providers had problems in following the right steps.

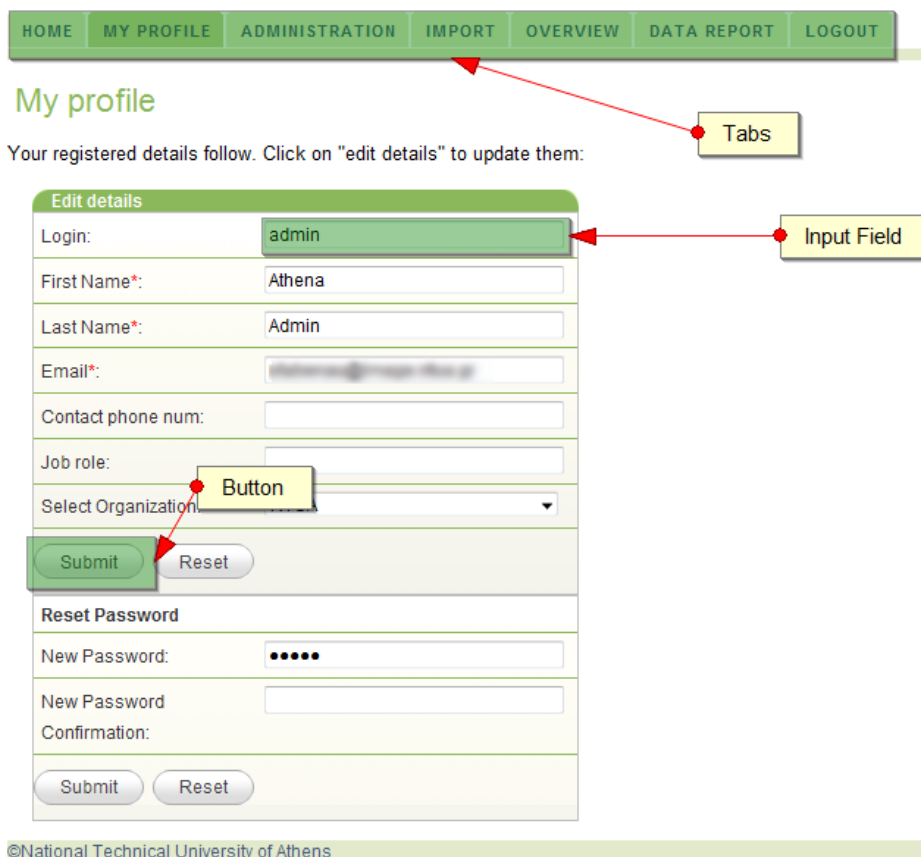


Figure 6: MINT's old user interface

For that purpose in the new user interface of the MINT-mapping tool a different approach has been followed and the jQuery library Kaiten¹⁰ has been used. The important difference in the new approach is that interactions between the user and the application are a stack of contiguous screens where each screen is presented in columns as shown below.

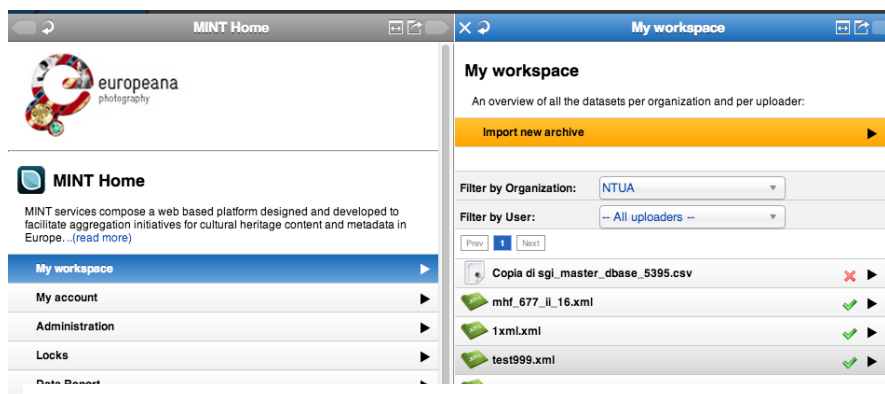


Figure 7: MINT's new user interface (workspace)

¹⁰ <http://www.officity.com/kaiten/>

In that way the content provider can only perform specific actions depending on his starting point. For example when clicking on the workspace button from MINT's start page he can either make an import (that is highlighted) or select one of the existing imports. In a similar manner after selecting an import the following screen appears illustrating to the provider his possible options (It is important to mention at that point that the highlighted option varies depending on the state of the import see 5.2.1.2).

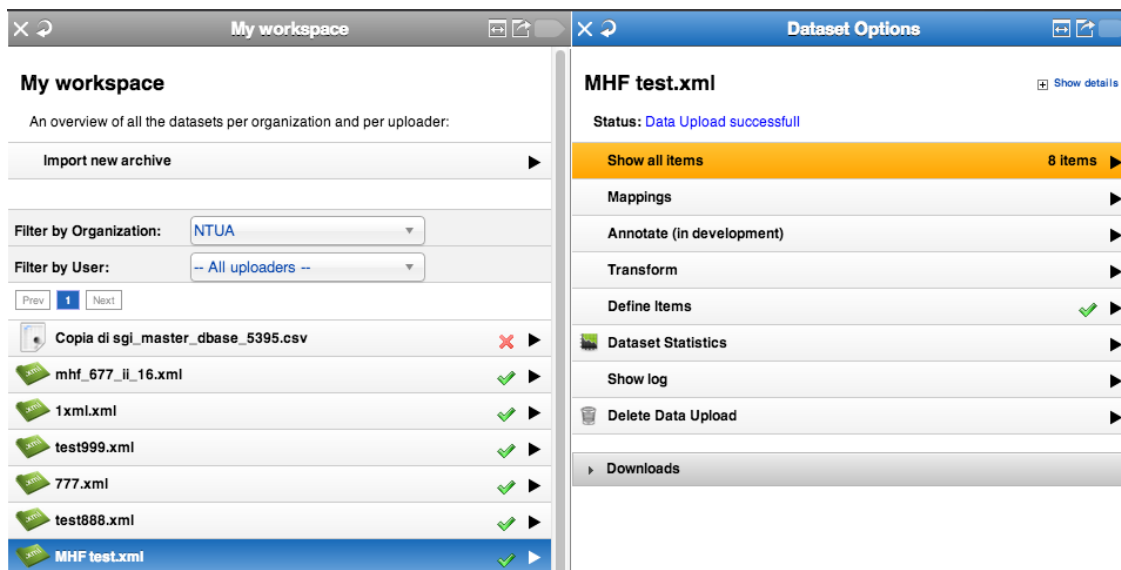


Figure 8: MINT's new interface (Dataset Options)

4.2.2 Improved Schema browsing

One of the major difficulties that content providers experienced when producing mappings using the MINT-mapping tool was in browsing the (their) input and the target schema. The MINT release used in previous projects had limited functionality for browsing schemas. It was only permitting providers to explore the input schema by using its hierarchy and browse the target schema by clicking on some predefined buttons that categorized elements as shown in the figure below.

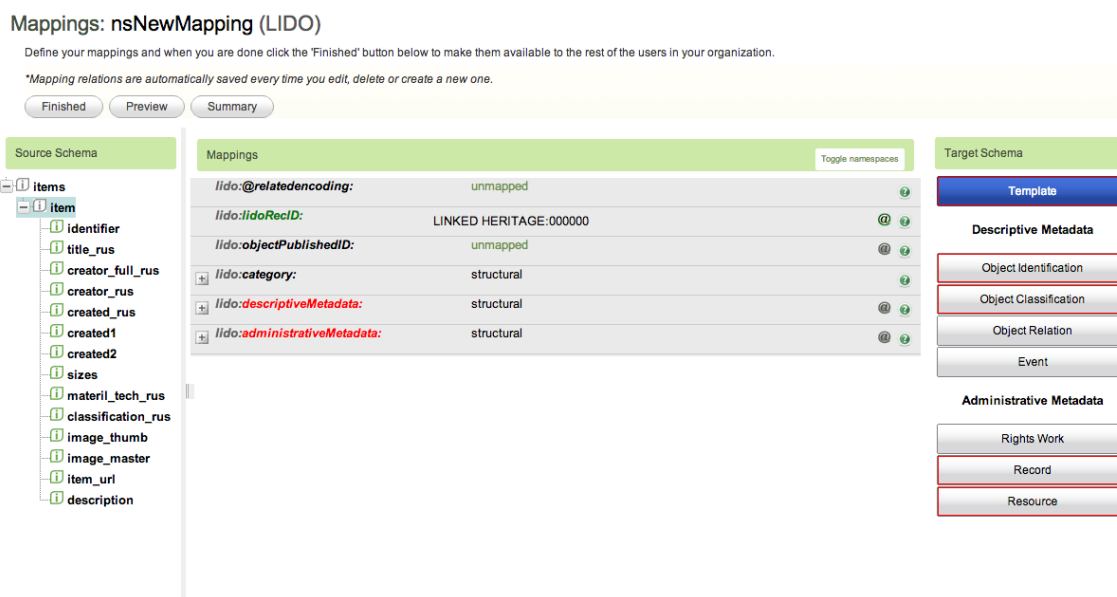


Figure 9: MINT's old mapping editor

In order to make the discovery of schema's elements easier and for facilitating the mapping process the new MINT interface allows to providers to search elements in both input and target schema. In addition for further assisting providers to meet the project requirements and to get familiarized with the target schema bookmarks have been set up for the Europeana Photography recommended elements. (more details about input and target schema browsing can be found at Section 5.3.1)

4.3 MAPPING FUNCTIONALITY

MINT's main functionality is the mapping and transformation of the providers' in-house metadata to the Europeana Photography intermediate schema and its publication to Europeana. Content providers perform this process by using MITN's mapping editor that offers a wide range of mapping functionalities. Therefore the development of new mapping functionalities always comprises part of NTUA's ongoing work.

4.3.1 Enhanced mapping functionalities

In the MINT version developed for the Europeana Photography project various new mapping functionalities have been developed. These include the implementation of the negation for all the conditions used in conditional mappings (see Section 5.3.5.6). By using this functionality a content provider can perform mappings in which his dataset is separated into two sets according to its values (e.g. those that start with, and those that don't start with) and each one of them is treated in a different way. In addition current developments (that will be finalized by month 15 of Europeana Photography) include the implementation of the else statement for further completing the conditional mapping functionality of the tool.

Additionally another very important mapping functionality that is under development is the "group by". The need for that functionality occurred after the Europeana Photography training session that was held in Athens. In this training providers have been asked to bring their own metadata, and by using them to get familiarized with the MINT mapping tool environment. Some of the providers had metadata exports in schemas that did not have any hierarchical structure like MARC. In detail the different types of information that described their collections (e.g. photographer, location etc.) appeared as different occurrences of the same element but with different attributes. The problem with content providers that use input schemas like MARC through MINT is that only one element appears on the mapping editor – and therefore providers cannot distinguish between their input values and the desired mappings. The "group by" mapping functionality will solve that problem by permitting to content providers of such schemas to treat each element occurrence as an independently considering its different attributes.

4.3.2 SKOS Vocabularies Support

One of the most important outcomes for the Europeana Photography project is the development of SKOS vocabularies specific to the photographic content (WP 4) and their use for the metadata production and delivery to Europeana by the content providers. Since the content providers of Europeana Photography use the MINT-mapping tool for that purpose, the integration of SKOS vocabularies to metadata through it was vital. In order to extend MINT's functionality to support SKOS vocabularies an additional module has been developed. More specifically a semantic repository¹¹ has been set up to which the SKOS vocabularies are stored. The communication of the MINT-mapping tool with it is established by using SPAQRL 1.1 to retrieve the vocabularies' terms based on the SKOS specification¹². Additional semantic properties can be added - if necessary - to the vocabularies for controlling selectable and non-selectable terms through the mapping tool (skos:member, skos:Collection) and also for selecting to display only subcategories of them(skos:inScheme, skos:ConceptScheme). (see Section 5.3.5.8)

¹¹ https://jena.apache.org/documentation/serving_data/

¹² <http://www.w3.org/TR/skos-reference/>

5 USER MANUAL

This section presents the MINT mapping tool functionality, part of it can be found at <http://mint.image.ece.ntua.gr/> that is the MINT web page. However this section is the first documentation of the MINT-mapping tool that has been implemented within the Europeana Photography project including MINT's latest developments (<http://mint.image.ece.ntua.gr/> describes the MINT-Athena version). It is divided in 4 sub-sections that correspond to the main actions that a content provider has to perform before submitting his/her content to Europeana.

- User organization and registration.
- Import of metadata.
- Mapping.
- Transformation – Publication to Europeana.

5.1 USER & ORGANIZATION REGISTRATION

The very first step that each provider has to make is to register himself/herself and also his/her institution – organization to the MINT mapping tool. The managers or the technical experts of the organizations that participate in the Europeana Photography are suggested to follow the next steps.

1. User registration without and organization (see Section 5.1.1.1.1) – to register themselves
2. Organization Registration & Management (see Section 5.1.2) – to register their organization(s)
3. User management (& User creation under an organization by its administrator) (see Section 5.1.1.2) – to create accounts for the users that work for their organization(s).

On the other hand the annotators of the organizations that participate in the Europeana Photography are suggested to register themselves under an existing organization (see Section 5.1.1.1.2). If the organization they work for is not yet registered, then they are suggested to contact the manager or the technical expert of their organization.

5.1.1 User Registration

User registration can be done in two ways.

- *Self-registration from MINT Mapping Tool* – The user registers himself/herself directly (see Section 5.1.1.1).
- *User creation under an organization by its administrator* – The organization's administrator creates an account for a user (see Section 5.1.1.2).

5.1.1.1 Self-registration from MINT Mapping Tool

A user can register himself/herself by clicking on the blue link “I want to register” that appears in the start page of MINT <http://mint-projects.image.ntua.gr/photography>



Figure 10: MINT Ingestion Server

This link directs to the following form that has to be filled in for registration.

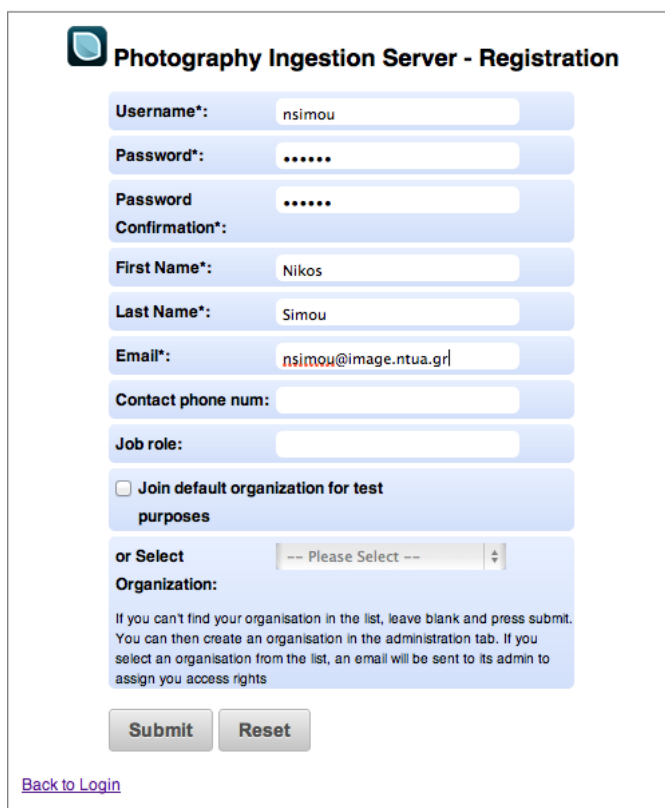


Figure 11: Registration form

At this point you can

- *Join the default organization (NTUA) for test purposes* – This option was implemented for dissemination purposes and it is recommended to people that do not participate in the EuropeanaPhotography project and they just want to check the functionality of the MINT mapping tool.
- *Select one of the existing organizations* – This option is recommended to users that work for an organization that participates in the EuropeanaPhotography project. If you select an organization from the drop down list, an email will be sent to its administrator to assign you access rights.

- *The organization has not been registered yet* – This option is recommended to users that work for an organization that participates in the EuropeanaPhotography project, but it has not been registered. By leaving the selected organization blank (i.e. set to “-- Please Select --”) you can register and then create an organization.

5.1.1.1.1 User registration without an organization

After registering without an organization you are logged in to MINT. On the left, under the pane MINT Home -that is the central functionality pane of MINT mapping tool- the following options appear

- *My workspace* – By selecting it you can manage your imports. At this point it is not activated because you are registered under no organization, and therefore the creation of an organization is first required.
- *My account* – By selecting it you can edit your account details.
- *Administration* – By selecting it you can create an organization and users that work for it.
- *Locks* – By selecting it you can unlock mappings locked by other users. Not available at this point
- *Mint Documentation* – By selecting it you are directed to the MINT documentation page.

At this point you have to create an organization by clicking on the Administration button and then selecting “Create new organization” as shown below (see Section 5.1.2).

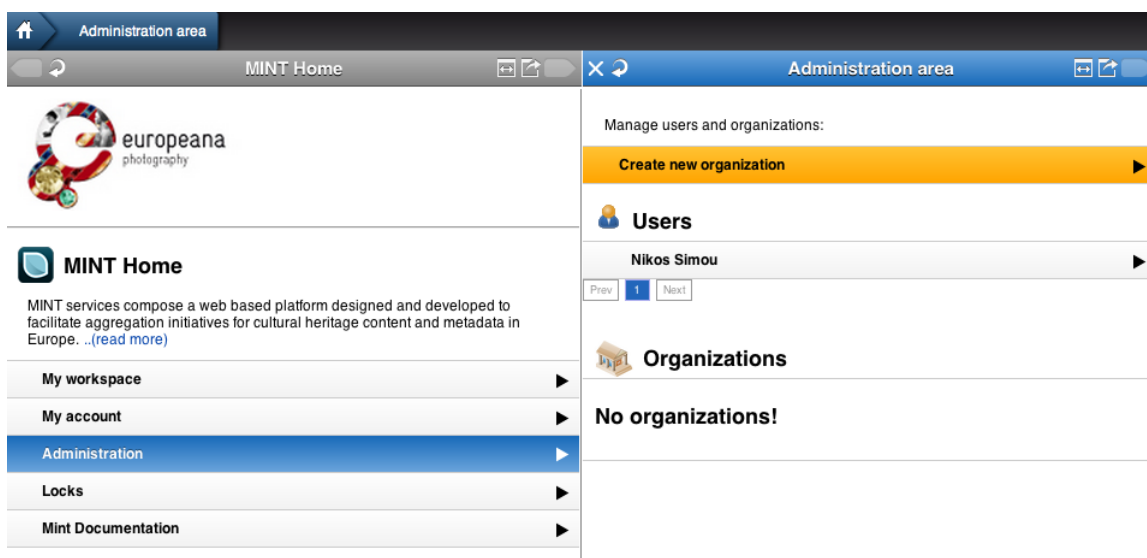


Figure 12: Create organization

5.1.1.1.2 User registration under an existing organization

After registering under an existing organization you are logged in to MINT. On the left, under the pane MINT Home -that is the central functionality pane of MINT mapping tool- the following options appear.

- *My account* – By selecting it you can edit your account details.
- *Mint Documentation* – By selecting it you are directed to the MINT documentation page.

Note that at this point only these options appear because the administrator of the organization you registered for has not assigned access rights to you.

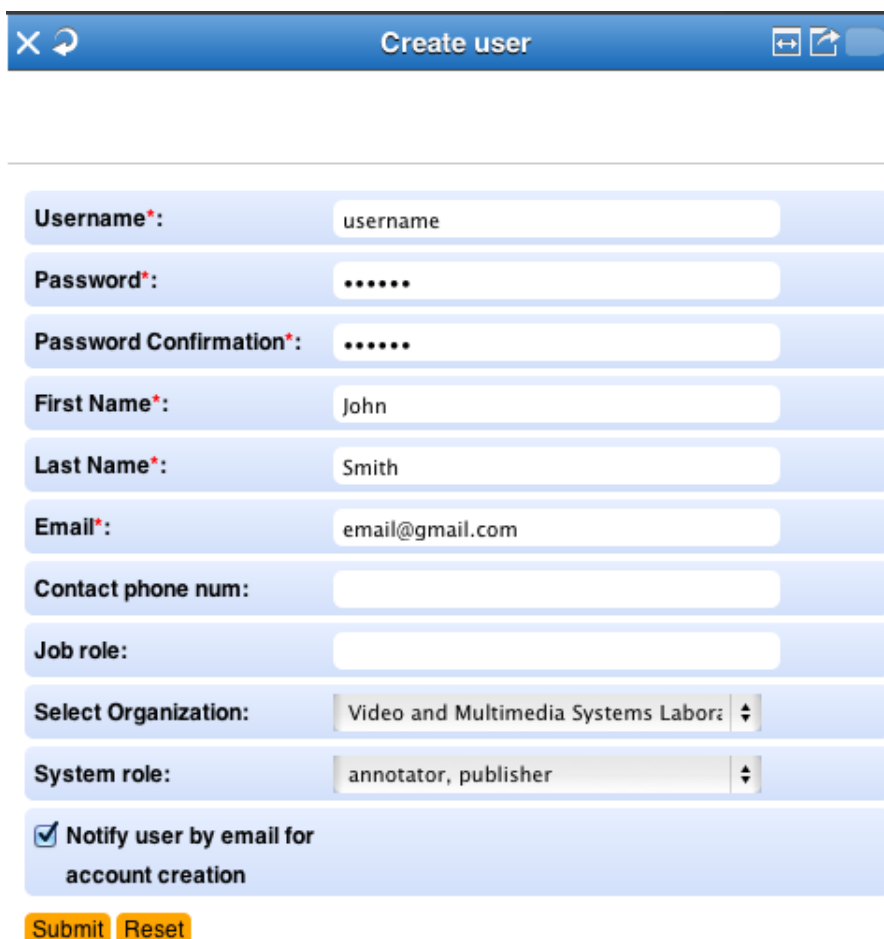
5.1.1.2 User management (& User creation under an organization by its administrator)

For creating a new user or activating a user registered for the organization you are administrating log in to MINT mapping tool and select “Administration” for the MINT Home pane. This opens the Administration

Area in which you can create new users, new organizations and also edit or delete the existing users and organizations.

For creating a new user select “Create new user” and then fill in the following form that appears. A user can have one of the following roles.

- **Administrator:** This user can create/update/delete users and children organizations for the organization he/she is administering. He/She can also perform uploads and all available data handling functions provided by the system.
- **Annotator:** This user can upload data for his/her organization (and any children organizations) and perform all available data handling functions (view items, delete items, mappings etc) provided by the system, apart from final publishing of data.
- **Annotator & Publisher:** This user has all the rights of an annotator as well as rights to perform final publishing of data.
- **Data Viewer:** This user only has viewing rights for his/her organization.
- **No role:** A user that has registered for an organization but has not yet been assigned any rights.



Create user

Username*: username

Password*:

Password Confirmation*:

First Name*: John

Last Name*: Smith

Email*: email@gmail.com

Contact phone num:

Job role:

Select Organization: Video and Multimedia Systems Laborz

System role: annotator, publisher

Notify user by email for account creation

Submit **Reset**

Figure 13: Create user form

You can delete or edit -to assign a role to- an existing user by selecting it and pressing the “Delete” or “Edit” button respectively.

5.1.2 Organization Registration & Management

For creating a new organization or the children organizations of an existing organization you have to register yourself without an existing organization (see Section 5.1.1.1) or to be the administrator of the

organization you want to edit. Log in to MINT mapping tool and select “Administration” from the MINT Home pane. This opens the Administration Area in which you can create new users, new organizations and also edit or delete the existing users and organizations.

For creating a new organization select “Create new organization” and then fill in the following form that appears. It is important to note at this point that every organization must have a primary contact user (i.e. an administrator). For the selection of the primary user a drop down list appears with all the registered users in the organization.

In addition an organizations follow a hierarchical structure. Each organization may be divided in sub-organizations, while each sub-organization can be divided in further sub-organizations, forming a tree-like structure. The users registered in parent organizations can view the imports and mappings of users registered in their children organizations (see Section 5.2.1) while administrators of a parent organization have also administrative rights to the children organization. For the selection of the parent organization a drop down list appears with the organizations administrated by you.



Create Organization

Country*: Greece

English name*: Video and Multimedia Systems Laboratory

Name*: Εργαστήριο Επεξεργασίας Εικόνας και Βίν

Organization acronym: IVML

Type*: Other

Address: Iroon Polytexneiou 9

Organization url: http://www.image.ece.ntua.gr/

Organization description: of digital image, video and multimedia processing.

Select parent organization: -- No parent --

Primary contact user*: nsimou

Submit **Reset**

Figure 14: Create organization form

You can delete or edit an existing organization by selecting it and pressing the “Delete” or “Edit” button respectively. The following figure illustrates the edit form for an organization.



Country*: Greece

English name*: Image Laboratory

Name*: Εργαστήριο Εικόνας

Organization acronym: IL

Type*: Other

Address:

Organization url:

Organization description:

Parent organization: Video and Multimedia Systems Labor

Primary contact user*: -- Please Select --

Submit **Reset**

Figure 15: Edit organization

5.2 IMPORT OF METADATA

To import your metadata to the MINT mapping tool, you have to log in and to select “My workspace” from the MINT Home pane. Note that your user role must have been assigned to annotator or higher (see Section 5.1.1.2) otherwise this option will not appear there.

From the workspace pane you can view all the datasets per organization and per user. You can import a new archive by selecting the respective button and filling the following form depending on the upload type. Note that if you have access to more than one organization, you need to specify for which organization you are performing the upload by selecting it from the drop down list. Finally, a user can select the “This import conforms to” together with the appropriate selection of a schema in the case that his/her upload already conforms to the selected schema and no mapping is necessary.

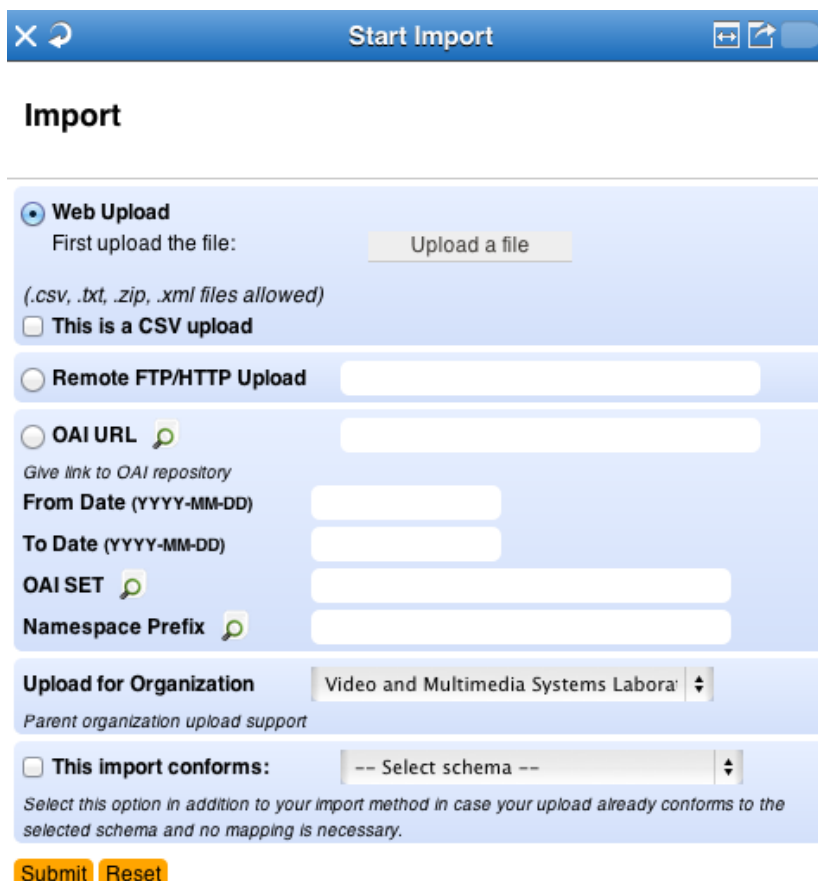


Figure 16: Upload form

HTTP Upload

Allows uploading your data directly from your web browser. Simply click "Upload a file" button and select the file containing your data. You can upload XML, CSV files or zip files containing the XML and CSV files.

Note that if you upload a CSV file, you have to tick on "This is a CSV upload" check box and specify the delimiter as shown in the figure below. It is also important to note at this point that only encoding UTF-8 is supported.¹³

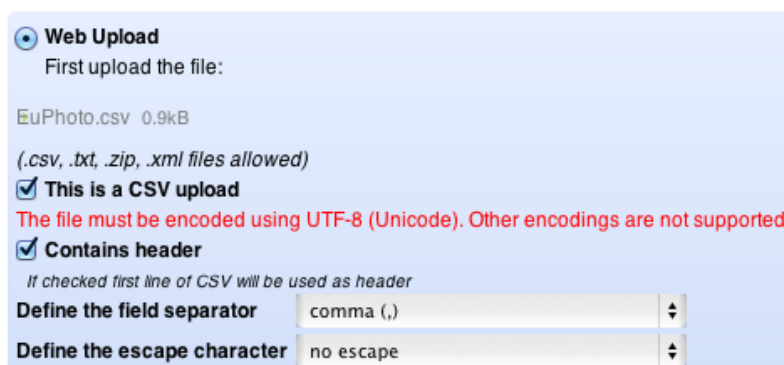


Figure 17: CSV upload


¹³ Content providers can check the encoding of their exports either by using software like <http://encodingchecker.codeplex.com/>, a unix command line (file -bi) or a text editor. However since the validity of the above is questionable providers are recommended to set the encoding to UTF-8 during the export of their metadata from their management systems.

IMPORTANT NOTE: The MINT ingestion tool operates mainly on the "upload" level of the metadata files. Every import that has at least one file is considered as a new upload of metadata and handled in the same way as if it had more than one. For this reason it is encouraged to use zip archives with many XML or CSV files compressed and limit the ingestion of single XML or CSV files only for testing purposes.



Remote FTP/HTTP Upload

Uploads a file from a remote FTP/HTTP location. Useful when your files are available already on a remote server.

OAI URL

The MINT ingestion tool has an integrated OAI-PMH V2 harvester which can be used to access data stored in a remote server that supports this protocol. You have to fill in the base URL of the OAI-PMH repository and then click on the  button next to it in order to check its validity. You can also fetch data based on a specific date interval as it is defined by the OAI-PMH protocol. Finally, you can fetch the OAI-PMH sets and namespaces that the remote repository supports and choose from them the set and type of metadata you wish to import.

5.2.1 Dataset Options

After importing a file to the MINT mapping tool it can be viewed in "My Workspace pane" either with a green tick () that indicates that the import was successful or with a red x () that indicates a problem in the upload as shown in the following figure.

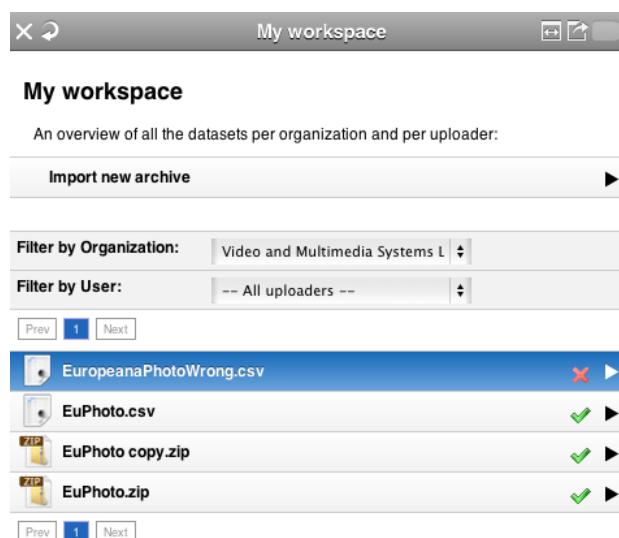


Figure 18: Workspace with successful and unsuccessful imports

5.2.1.1 Wrong Import

By selecting an unsuccessful import you can get feedback on the error occurred by clicking on the "Show Log" as illustrated in the following figure.

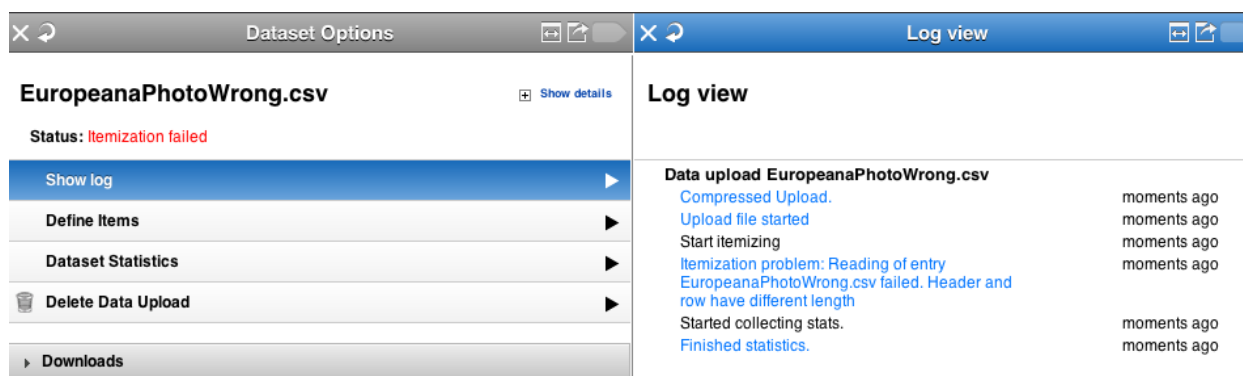


Figure 19: Error in import

The Define Items and Dataset Statistics options appear there but they don't work since the import was not successful and the user that encounters this situation is suggested to click on "Delete Data Upload" to delete the upload.

5.2.1.2 Successful import

When a dataset is imported successfully the following options appear in the "Dataset Options" pane.

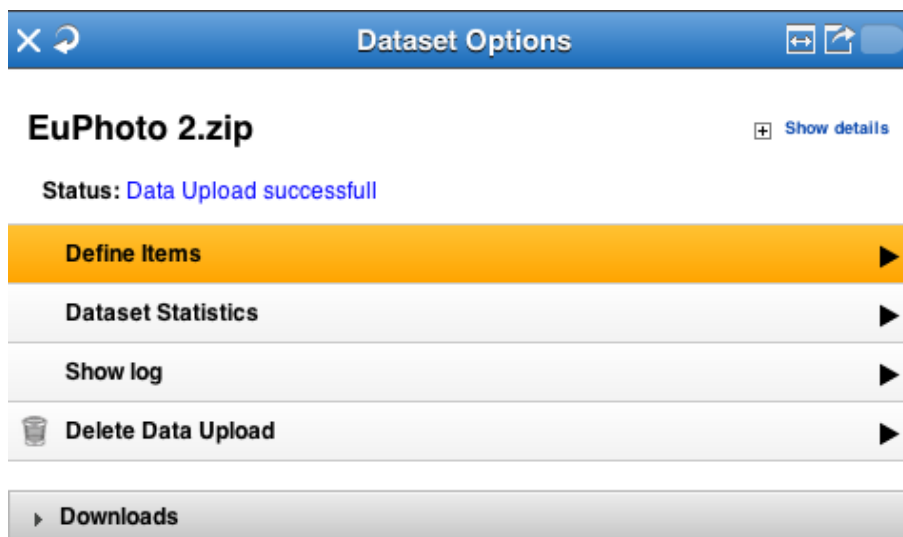
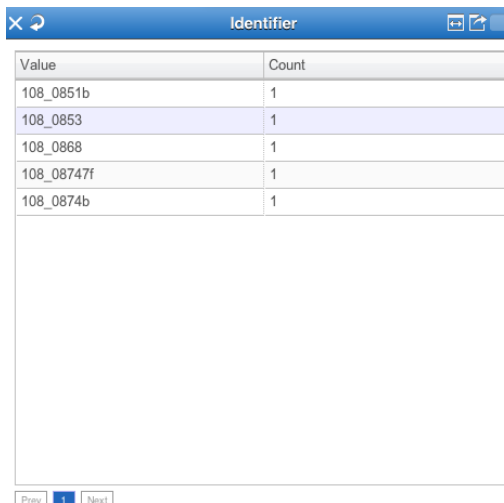


Figure 20: Successful upload dataset options

5.2.1.2.1 Define Items

By clicking on the "Define Items" the following pane appears. On the left part of the panel there is a tree representing the structure of the XML file(s) you have imported. Items with a "+" on the left can be expanded, while items with a "-" can be collapsed. By clicking on the information icon (i) on the left of each node you get information and statistics about the values of this node.

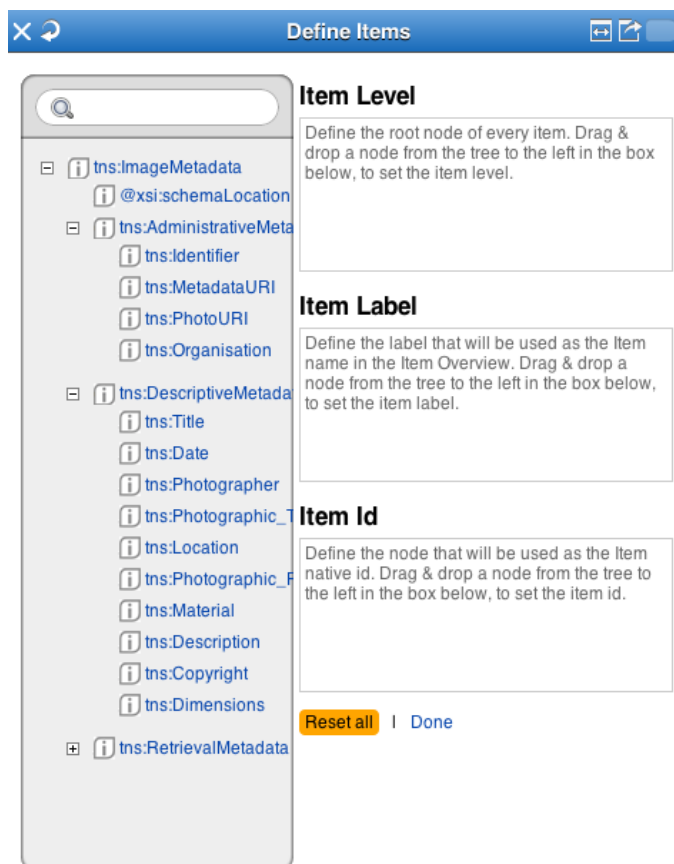


Value	Count
108_0851b	1
108_0853	1
108_0868	1
108_08747f	1
108_0874b	1

Figure 21: Information about the values of a node

Note that on the top of the tree there is a text field that can be used to search the tree structure of the XML file(s) you have imported. You have to set the following.

- **Item Level** - Define the root node of every item. Drag & drop a node from the tree to the left in the box below, to set the item level.
- **Item Label** - Define the label that will be used as the Item name in the Item Overview. Drag & drop a node from the tree to the left in the box below, to set the item label.
- **Item Id** - Define the node that will be used as the Item native id. Drag & drop a node from the tree to the left in the box below, to set the item id.



✕ ↻
Define Items
↔ ↗

- [-] i tns:ImageMetadata
 - i @xsi:schemaLocation
- [-] i tns:AdministrativeMetadata
 - i tns:Identifier
 - i tns:MetadataURI
 - i tns:PhotoURI
 - i tns:Organisation
- [-] i tns:DescriptiveMetadata
 - i tns>Title
 - i tns>Date
 - i tns:Photographer
 - i tns:Photographic_T
 - i tns:Location
 - i tns:Photographic_F
 - i tns:Material
 - i tns:Description
 - i tns:Copyright
 - i tns:Dimensions
- [-] i tns:RetrievalMetadata

Item Level

Define the root node of every item. Drag & drop a node from the tree to the left in the box below, to set the item level.

Item Label

Define the label that will be used as the Item name in the Item Overview. Drag & drop a node from the tree to the left in the box below, to set the item label.

Item Id

Define the node that will be used as the Item native id. Drag & drop a node from the tree to the left in the box below, to set the item id.

Reset all | Done

Figure 22: Define Items

After setting the above click on Done.

It is important to note at this point that while the “Define Items” action is mandatory for an XML import and Item level and at least one of item label or item id must be set in order to proceed this is not the case for CSV imports. The reason why is because a CSV file is converted to an XML file during the import and therefore the root item is set during the conversion. Hence the Define Items pane looks like the figure below for a CSV import. Users that import CSV files, however are highly encouraged to also define the Item label and Item Id of their collections because that will allow them to have better control of their metadata.

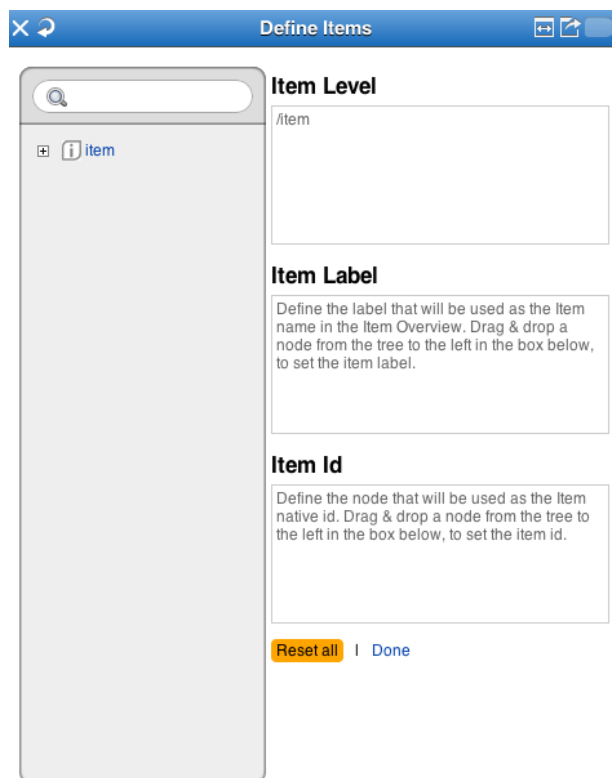


Figure 23: Define items for a CSV import

After having defined the items the following options appear in the “Dataset Options” pane.

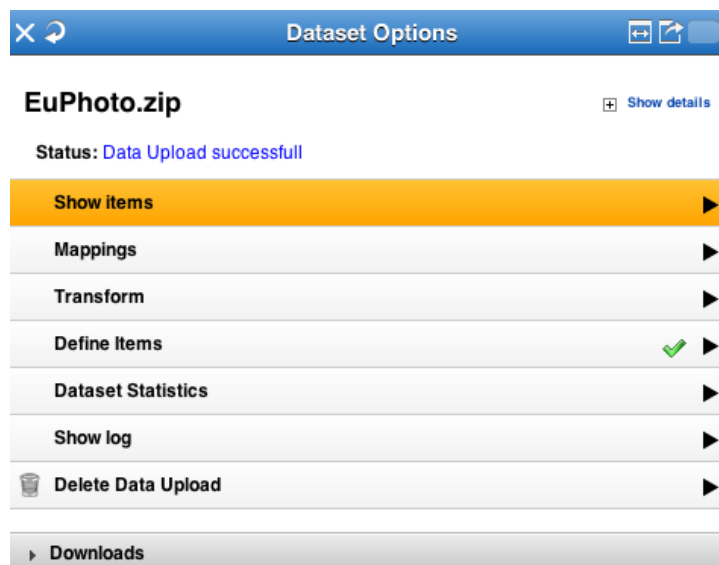


Figure 24: Dataset Options after having defined the items

5.2.1.2.2 Show Items

After having defined the items (see Section 5.2.1.2.1) by selecting the “Show Items” you can view the items you have uploaded as shown below (click on the view options to show or hide the view options menu).

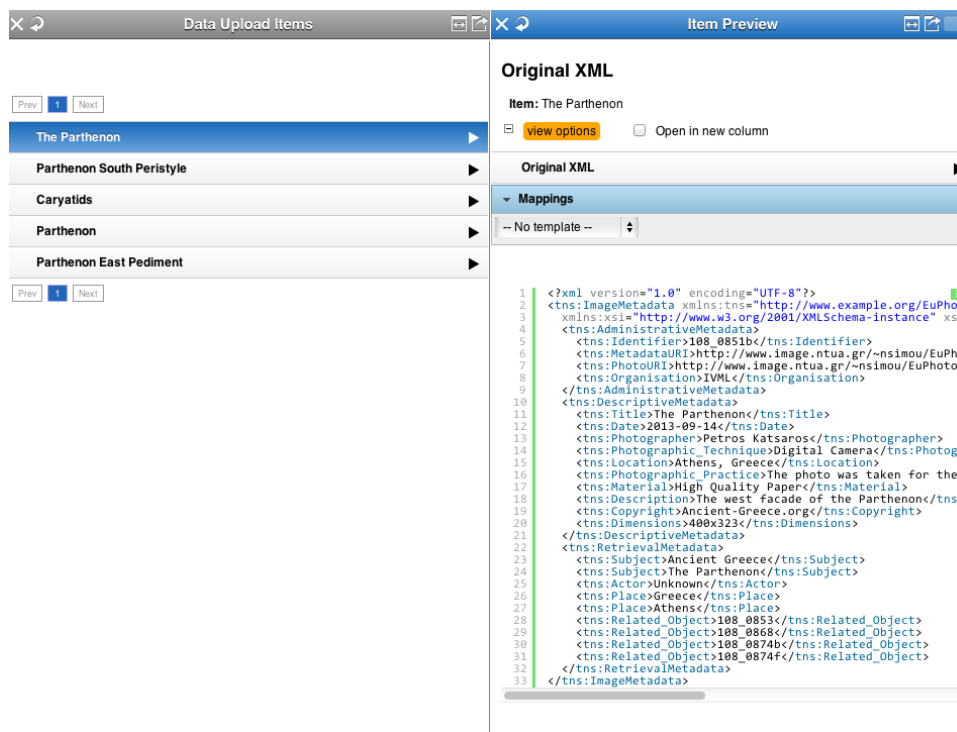


Figure 25: Show Items

If you select “Show Items” after you have implemented a mapping (see 5.3) then you can select it and the following options appear.

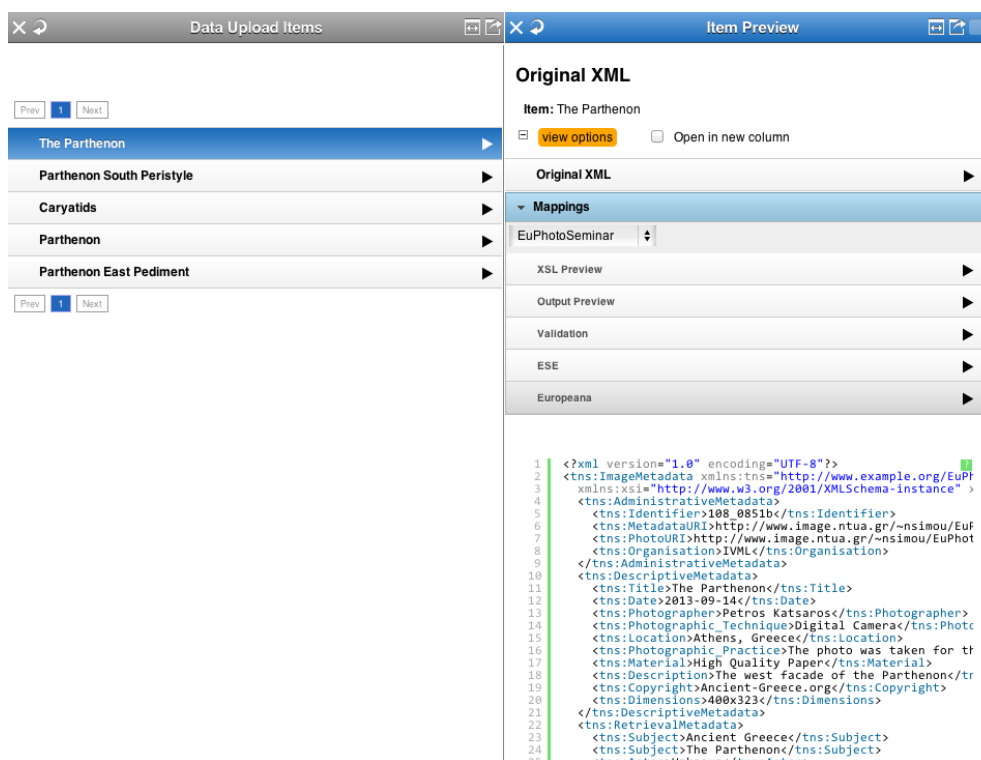


Figure 26: Show items after having implemented a mapping

- **XSL Preview** – Shows the XSL generated that transforms the import metadata to intermediate schema.
- **Output Preview** – Shows the metadata transformed to the intermediate schema.
- **Validator** – Shows the validation output for the metadata transformed to the intermediate schema.
- **ESE** – Shows the metadata transformed to ESE.
- **Europeana** – Shows the items as it will appear in the Europeana portal after its publication.

5.2.1.2.3 Dataset statistics

After having defined the items (see Section 5.2.1.2.1) by selecting the “Dataset statistics” you can view statistics about your import. In detail you can see all the xpaths of the imported dataset together with their distinct values count and the average length of their values, while by clicking on an element you can browse its values.

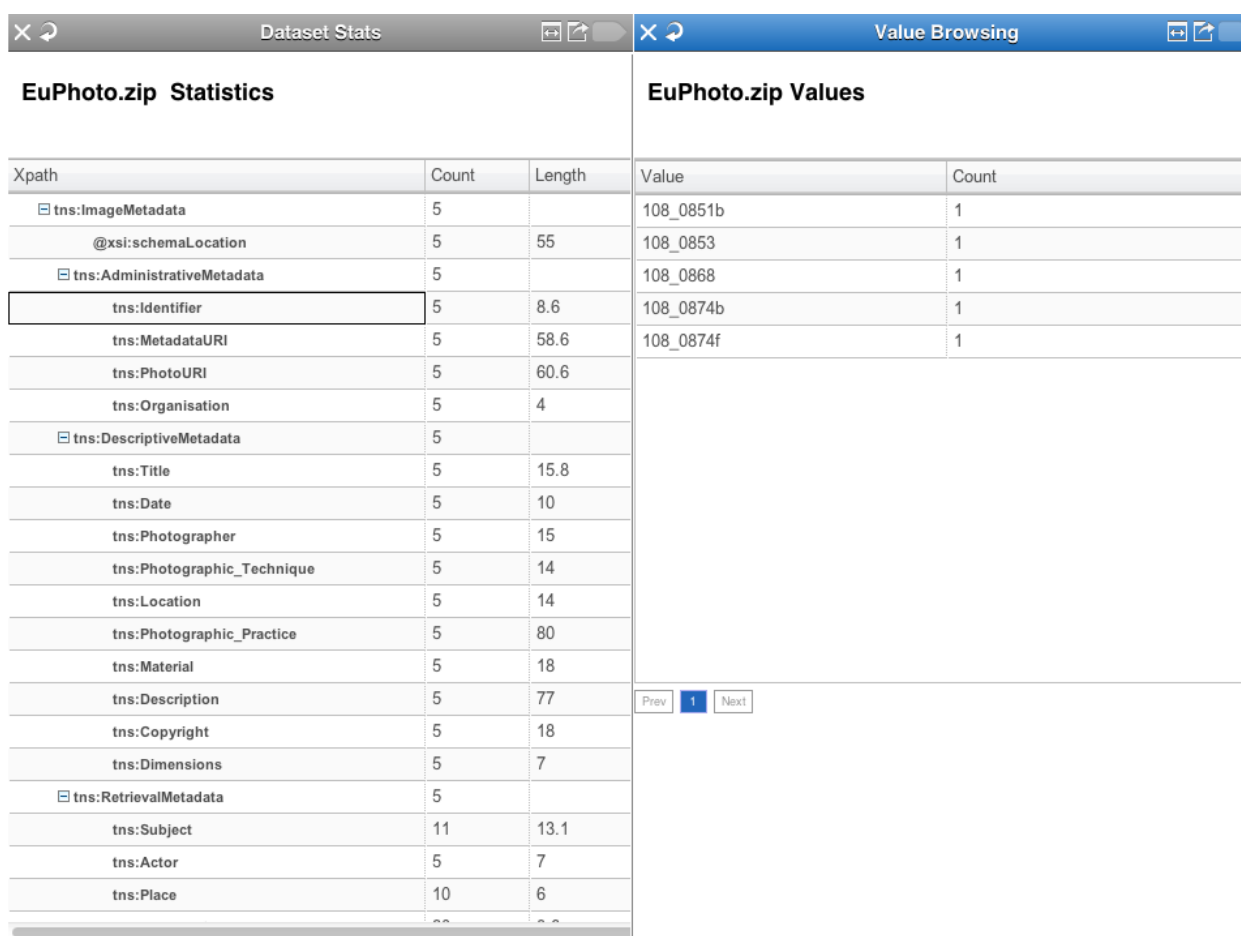


Figure 27: Dataset statistics

5.3 MAPPING

After having defined the items (see Section 5.2.1.2.1) select the “Mappings” for the “Dataset Options”. This opens the Mappings pane in which you can create a new mapping, upload a mapping or an XSL and manage the existing mappings.



Figure 28: Mapping options

5.3.1 Create new mapping

By selecting the Create new mapping option the “New Mapping” pane appears that is shown below.

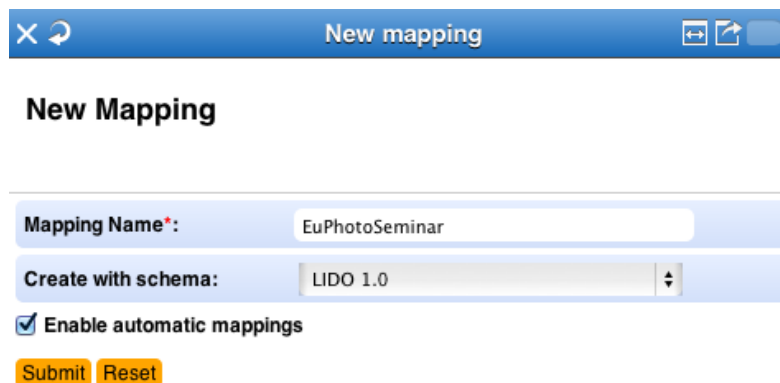


Figure 29: New mapping

There you set the mapping name and whether you want to enable or not automatic mappings and you press submit to be directed in the mapping tool shown in the following figure. The automatic mappings feature automatically maps the xpaths of the input schema to those of the target schema (i.e. LIDO that is the Europeana Photography intermediate schema) that are exactly the same. So users are suggested to use this feature only if their input metadata are already in LIDO.

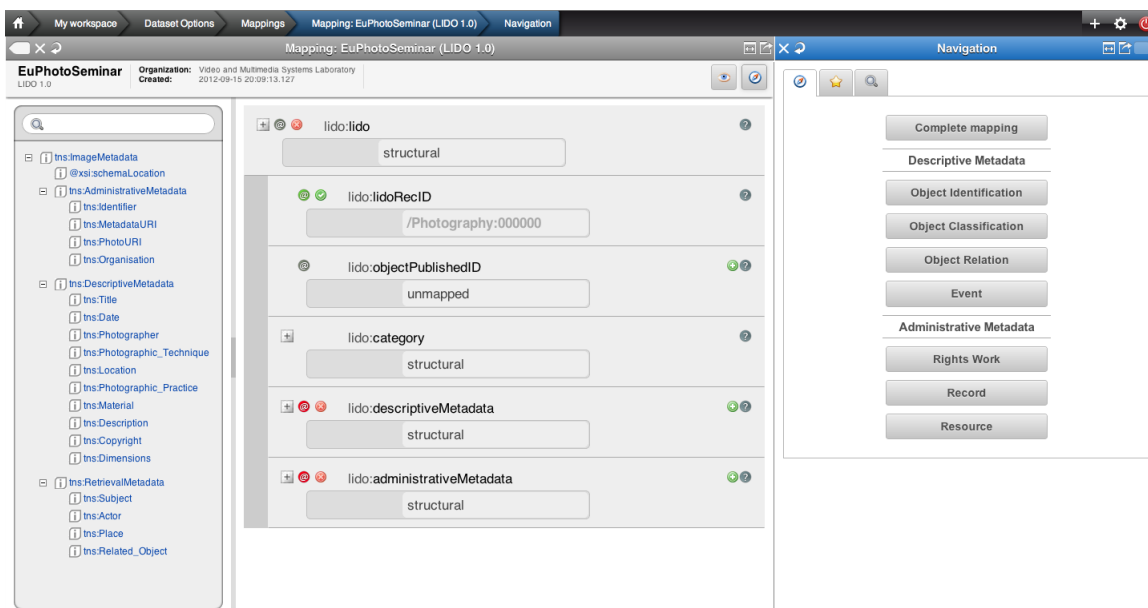


Figure 30: The mapping tool

The structure that corresponds to a user's specific import is visualized in the mapping interface as an interactive tree that appears on the left hand side of the editor (see figure below). The tree represents the snapshot of the XML schema that is used as input for the mapping process. The user is able to navigate and access element statistics and also to search the tree by using the text field on the top.

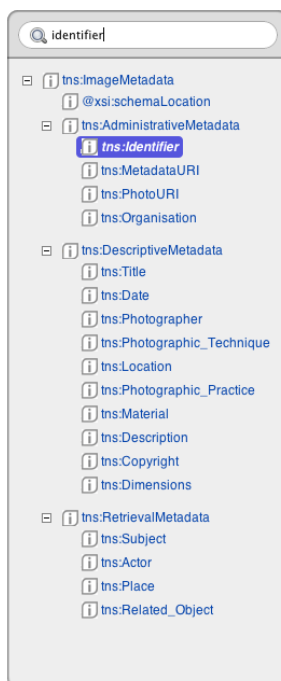



Figure 31: Input schema area

On the right hand side, buttons correspond to high-level elements of the target schema (see Figure 33 - if not visible click on  button on the top right of the mapping tool) and are used to access their corresponding sub-elements. These are visualized on the middle part of the screen as a tree structure of embedded boxes, representing the internal structure of the complex element. The user is able to interact with this structure by clicking to collapse and expand every embedded box that represents an element, along with all relevant information (attributes, annotations) defined in the XML schema document. To perform an actual (one to one) mapping between the input and the target schema, a user has to simply

drag a source element from the left and drop it on the respective target in the middle (see 5.3.5 for details).

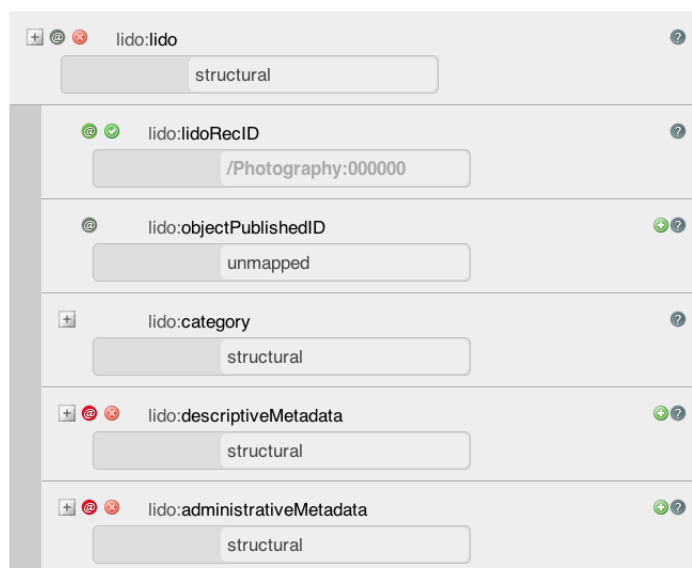


Figure 32: Mapping area

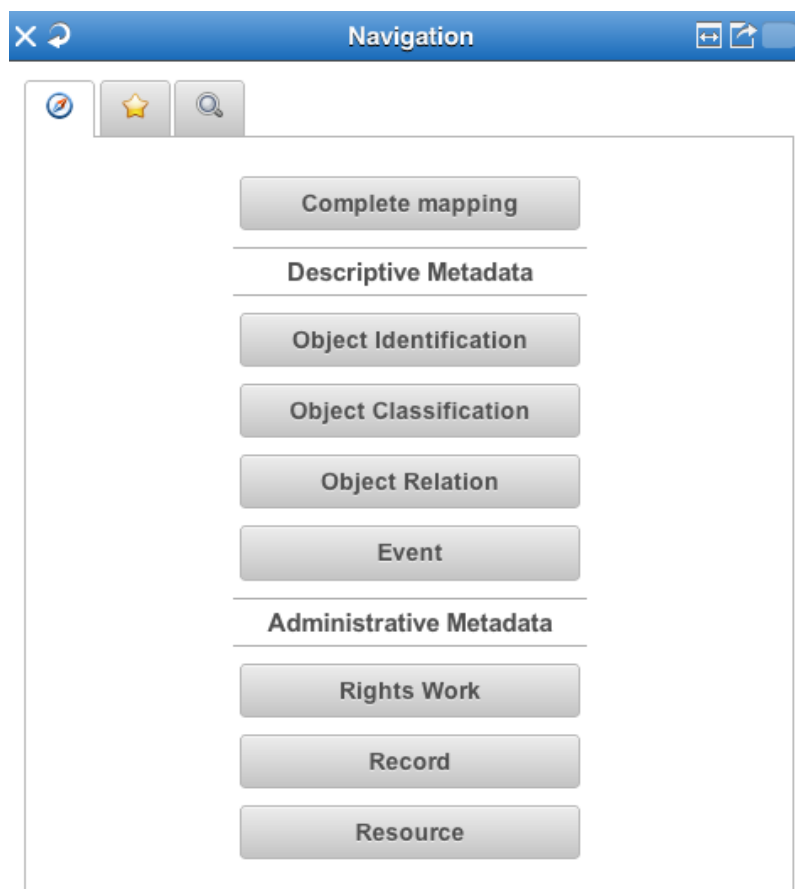


Figure 33: Navigation area - Schema navigation

For the needs of the Europeana Photography project and for assisting providers that are not familiar with LIDO, bookmarks have been created to the LIDO elements that will accommodate the Europeana Photography mandatory fields as decided in the content seminar (see Appendix I). In that way the provider by clicking on the star button (★) on the “Navigation” pane can view the bookmarks. Then he or she can click on the bookmark, named after the Europeana Photography mandatory fields, to see the respective LIDO element and easily map a value from his/her metadata.

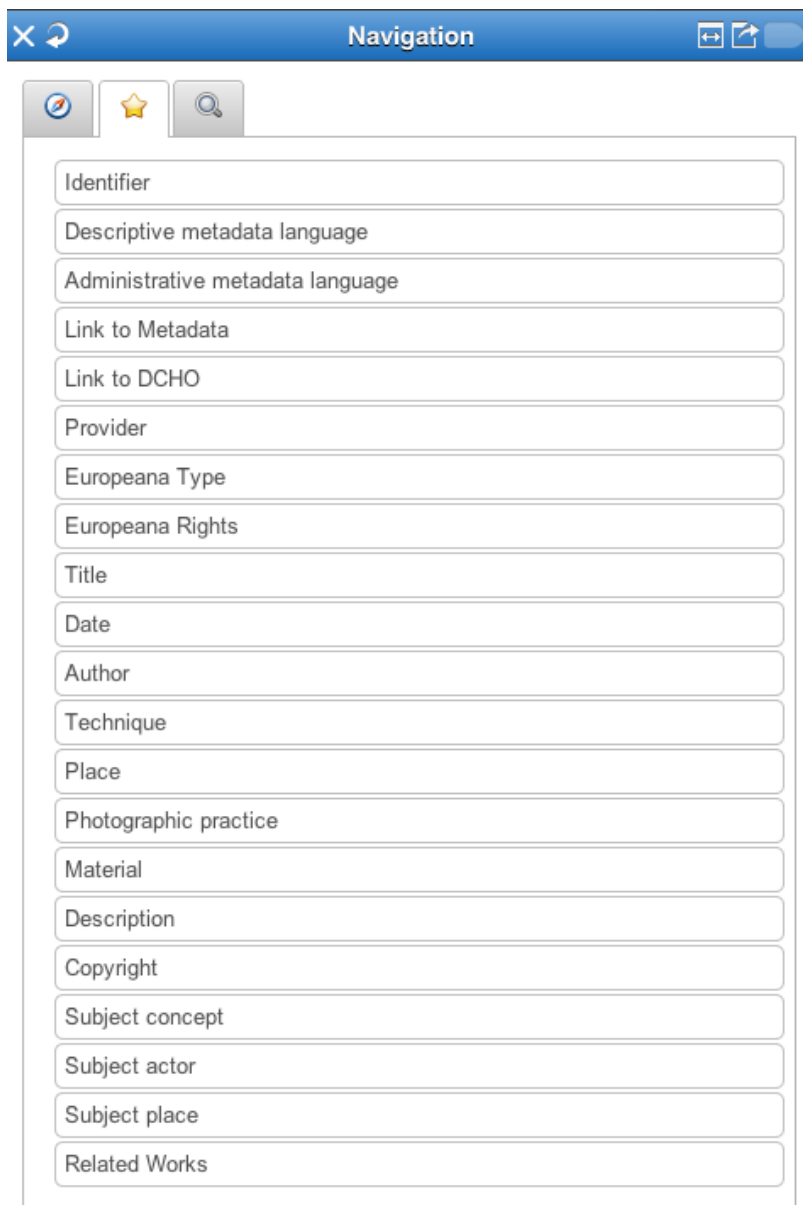


Figure 34: Navigation area - Bookmarks navigation

Finally, a third way of exploring the target schema is available and this is by searching the xpaths. The provider can click on the 🔍 button on the Navigation pane and then by entering a string in the text field he/she can perform a search in the target schema xpaths.

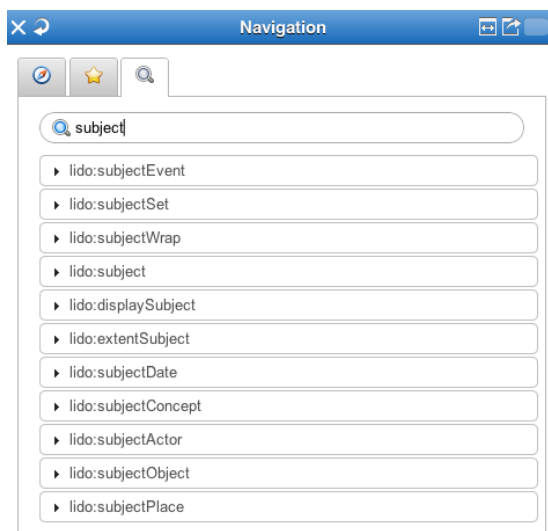


Figure 35: Navigation area - Search target schema

5.3.2 Upload mapping

By selecting the Upload mapping from the “Mappings” pane the following pane appears where the user can select a mapping to upload.

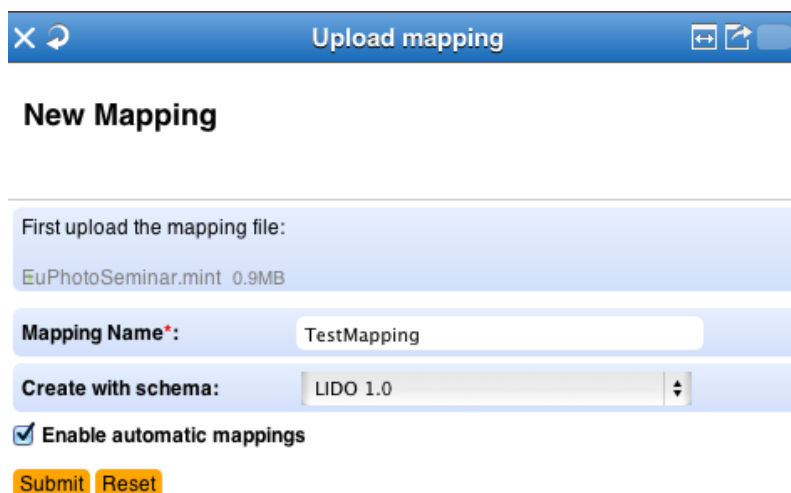
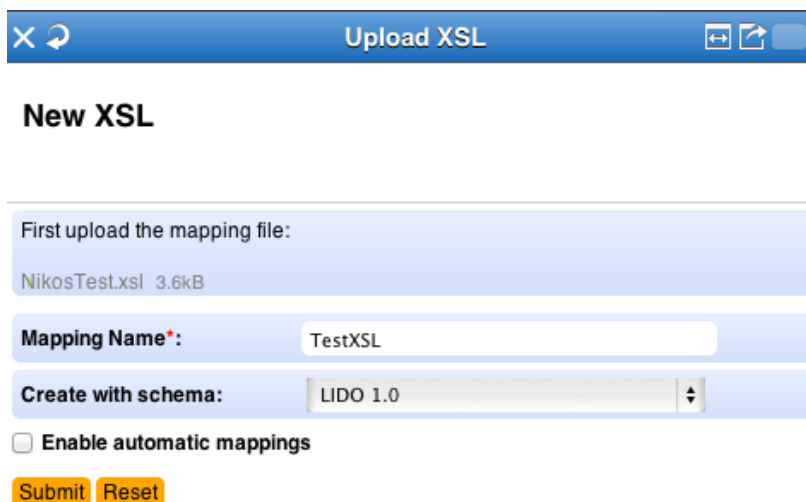


Figure 36: Upload mapping

After pressing the submit button the mapping with the name you specified appears in the “Mappings” pane and the mapping tool opens.

5.3.3 Upload XSL

By selecting the Upload XSL from the “Mappings” pane the following pane appears where the user can select an XSL to upload.



The screenshot shows a web interface for uploading an XSL file. At the top is a blue header with a close button, a refresh icon, the text 'Upload XSL', and navigation icons. Below the header is a section titled 'New XSL'. The main content area has a light blue background and contains the following elements: a text box with the message 'First upload the mapping file:' and a file name 'NikosTest.xsl 3.6kB'; a 'Mapping Name*' field with the value 'TestXSL'; a 'Create with schema:' dropdown menu set to 'LIDO 1.0'; an unchecked checkbox labeled 'Enable automatic mappings'; and two orange buttons, 'Submit' and 'Reset'.

Figure 37: Upload XSL

After pressing the submit button the XSL with the name you specified appears in the “Mappings” pane and the XSL editor appears as shown below. It is important to mention at this point that an XSL cannot be edited visually as a mint mapping. XSLs are edited through the text editor shown below.



The screenshot shows the XSL editor interface. It has a blue header with a close button, a refresh icon, the text 'XSL', and navigation icons. The main area is a text editor with a light gray background, displaying XML code. The code is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet exclude-result-prefixes="tns xsi" version="
3   xmlns:lido="http://www.lido-schema.org"
4   xmlns:tns="http://www.example.org/EuPhotoSchema"
5   xmlns:xalan="http://xml.apache.org/xalan"
6   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xm
7 <xsl:template match="/">
8   <lido:lidoWrap>
9     <xsl:apply-templates select="/tns:ImageMetadata"/>
10  </lido:lidoWrap>
11 </xsl:template>
12 <xsl:template match="/tns:ImageMetadata">
13   <lido:lido>
14     <lido:lidoRecID>
15       <xsl:attribute name="lido:type">Photography</xsl:a
16     <lido:descriptiveMetadata>
17       <lido:objectClassificationWrap>
18         <lido:objectWorkTypeWrap>
19           <lido:objectWorkType>
20             <lido:term>
21               <xsl:attribute name="lido:addedSearchTerm":
22             </lido:objectWorkType>
23           </lido:objectWorkTypeWrap>
24         <lido:classificationWrap>
25           <lido:classification>
26             <xsl:attribute name="lido:type">europeana:ty
27           </lido:classification>
28         </lido:classificationWrap>
29       </lido:objectClassificationWrap>
30     <lido:eventWrap>
31     ...
```

At the bottom of the editor is a gray 'Submit' button.

Figure 38: The XSL editor

5.3.4 Mappings Management

By selecting a mapping from the Mapping pane the following pane appears that allows you to

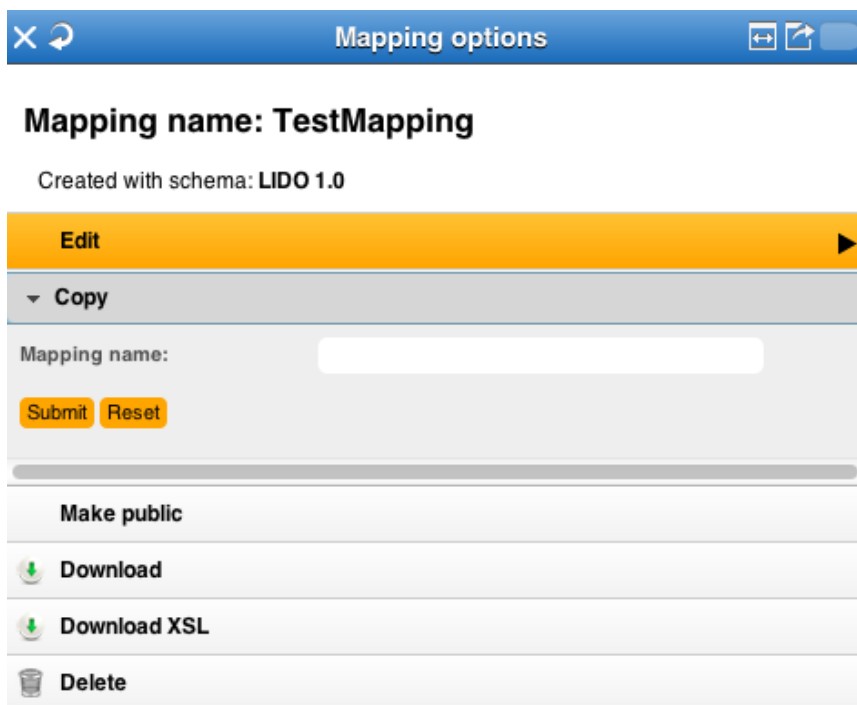



Figure 39: Mappings management

- **Edit a mapping** – This action opens the mapping tool to edit the mapping.
- **Copy a mapping** – This action creates a copy with the name specified of the mapping.
- **Make public/Make private** – This action makes the mapping public  (i.e. visible to all the users registered in all the organizations) or private (i.e. visible to all the users registered for parent organizations).
- **Download** – This action downloads the MINT mapping.
- **Download XSL** – This action downloads the XSL only.
- **Delete** – This action deletes the mapping.

5.3.5 Mappings

5.3.5.1 Mapping environment

As mentioned in Section 5.3.1 one way of mapping an element from your input schema to one of the target schema is the Xpath mapping that is performed simply by dragging the xpath from the input tree and dropping it to the desired element (within the area named unmapped – see figure below) of the mapping area.



Figure 40: Unmapped element in the mapping area

As we can see on the top of each element in the mapping area – the schema prefix and the element name are shown- there are some indicators that are shown in the following table.











Icon	Description
	Appears on the top left of an element to indicate that it is complex and thus by clicking on it you can view its sub-elements.
	Appears on the top left of an element to indicate that it has attributes. By clicking on it the attributes are displayed.
	Appears on the top left of an element to indicate that it has mandatory attributes that have not been assigned yet. By clicking on it the attributes are displayed.
	Appears on the top left of an element to indicate that attributes have been assigned to it.
	Appears on the top left of an element to indicate that it is mandatory and a value has to be assigned to it.
	Appears on the top left of an element to indicate that a value has been assigned to it.
	Appears on the top left of an element to indicate that it is not in the bookmarks. By clicking on it the star turns yellow and the element is added in the bookmarks.
	Appears on the top left of an element to indicate that it is in the bookmarks. By clicking on it the star turns grey and the element is removed from the bookmarks.
	Appears on the top right of an element to indicate that its cardinality can be greater than 1. By clicking on it a new element is added.
	Appears on the top right of all elements. By clicking on it you get the schema's documentation about that element.

Table 1: Mapping tool informative icons

After mapping an xpath from the input schema the mapping area turns as shown below and the xpath selected from the input schema is highlighted in bold.

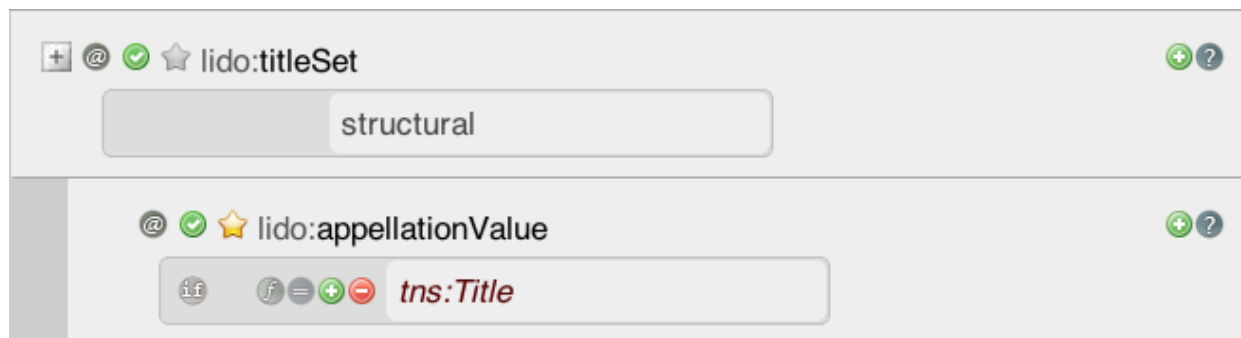





Figure 41: An element the has been mapped in the mapping area

As it can be observed additional buttons appear that are shown in the following table and discussed in the following sections.

Icon	Description
	Appears on the left of the mapping to indicate that no conditional mapping is used. By clicking on it conditional mapping is activated.
	Appears on the left of the mapping to indicate that a conditional mapping is used. By clicking on it the conditional mapping is de-activated.
	Appears on the left of the mapping to indicate that no functional mapping is used. By clicking on it the functional mapping is activated.






Icon	Description
	Appears on the left of the mapping to indicate that a functional mapping is used. By clicking on it the functional mapping is de-activated.
	Appears on the left of the mapping to indicate that no value mapping is used. By clicking on it the value mapping is activated.
	Appears on the left of the mapping to indicate that a value mapping is used. By clicking on it the value mapping is de-activated.
	Appears on the top left of a mapping. By clicking on it concatenate mapping is activated.
	Appears on the top left of a mapping. By clicking on it you remove the mapping.

Table 2: Mapping tool functional icons

If you leave the cursor above the mapping the following shortcuts to the input tree and xpath values appear.

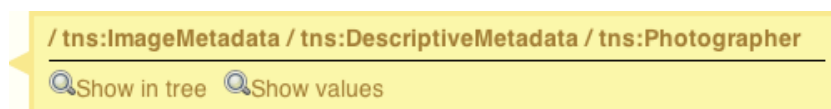


Figure 42: Shortcuts to input tree and values from mapping

5.3.5.2 Constant value mapping

By double clicking on the unmapped area you can define a constant value mapping. You can type a constant value in the provided text field. The value appears in the mapping area and in the resulting 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.

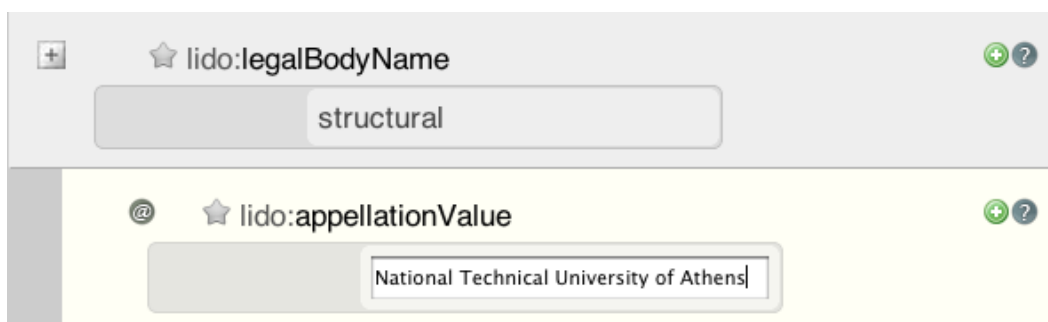



Figure 43: Constant value mapping

5.3.5.3 Concatenate mapping

By clicking on the  icon you can perform a concatenate mapping i.e. to combine more than one mapping for producing a new mapping. The resulting value that appears in the XML files is the concatenation of the xpath's values. In the following example the value from the xpath tns:Title is appended to the constant mapping "Collection for Ancient Greece".

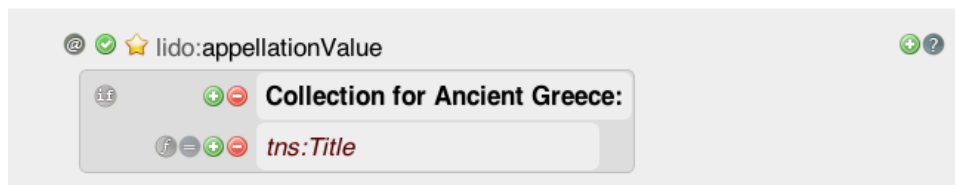





Figure 44: Concatenate mapping

5.3.5.4 Value mapping

By clicking on the  icon you can perform a value mapping i.e. to map specific values of your input metadata to specific values that you set. On the top of the pane you specify the input value of the selected element or you can browse values by clicking on  button, while in the field below you specify the desired output value. After that you add the value mapping link and the mapping you've entered appears in the text area below. You can remove a value mapping by clicking on the  button on the right of it. Value mapping is very useful when you want to normalize your data.

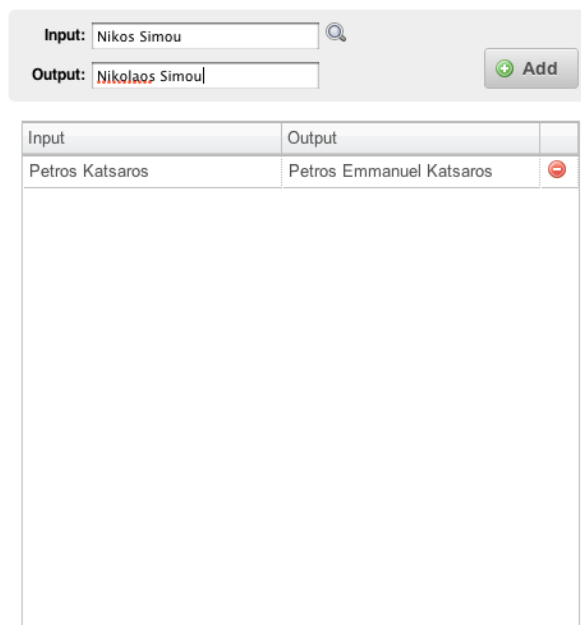

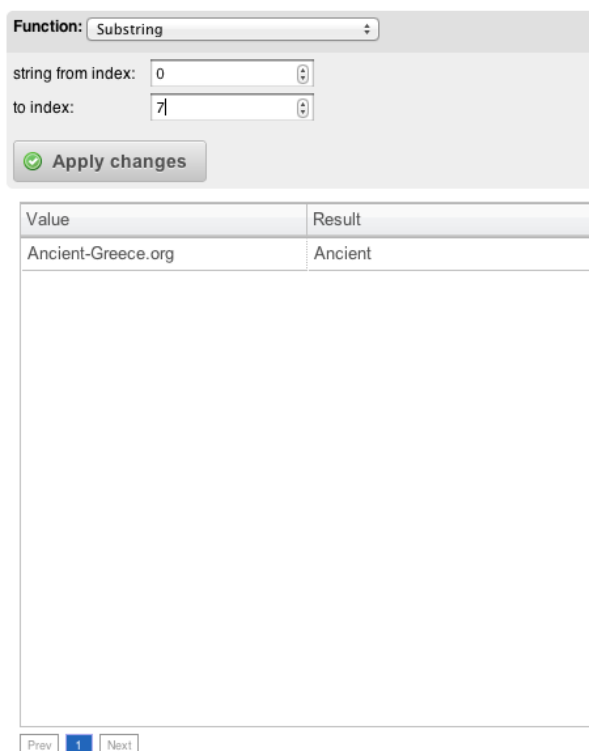


Figure 45: Value mapping editor

5.3.5.5 Functional mapping

By clicking on the  button you can perform a functional mapping i.e. to transform the value from an input xpath by applying a string manipulation function to it. On the top of the pane you select the function to apply, below it you can set its parameters, and in the table below a preview of the results is illustrated. Currently the following functions are supported


- **Substring** – You set the start and the end index.
- **Substring after** – You set the substring of the original string after which the value is taken.
- **Substring before** – You set the substring of the original string before which the value is taken.
- **Substring between** – You set the substrings of the original string after and before which the value is taken.
- **Split** – You set the delimiter for tokenization and the start index of the original string.
- **Tokenize content and generate an element per content** – You set the delimiter for tokenization.



Value	Result
Ancient-Greece.org	Ancient

Figure 46: Functional mapping editor

5.3.5.6 Conditional mapping

By clicking on the  button you can perform a conditional mapping i.e. to transform the value from an input xpath by using conditions. On the left the drop down menu with the type of condition appears that can be AND or OR. A condition is set using one of the following functions. (Note the xpath used in the condition can be different that the one that is used to the target element value – see the following figure)

- **Is equal to** – sets a condition that is satisfied when the given xpath is equal to the given value.
- **Is not equal to** – sets a condition that is satisfied when the given xpath is not equal to the given value.
- **Exists** – sets a condition that is satisfied if the given xpath exists. It is important to note at this point that the fact the xpath of an element exists in the input tree does not mean that it exists for all the data in the imported collection. (In other words the input tree shown on the left aggregates all the possible xpaths found in the input data).
- **Does not exist** – sets a condition that is satisfied if the given xpath does not exist. It is important to note at this point that the fact the xpath of an element exists in the input tree does not mean that it exists for all the data in the imported collection. (In other words the input tree shown on the left aggregates all the possible xpaths found in the input data).
- **Contains** – sets a condition that is satisfied if the given xpath contains the given value.
- **Does not contain** – sets a condition that is satisfied if the given xpath does not contain the given value.
- **Starts with** – sets a condition that is satisfied if the given xpath starts with the given value.
- **Does not start with** – sets a condition that is satisfied if the given xpath does not start with the given value.
- **Ends with** – sets a condition that is satisfied if the given xpath ends with the given value.
- **Does not end with** – sets a condition that is satisfied if the given xpath does not end with the given value.

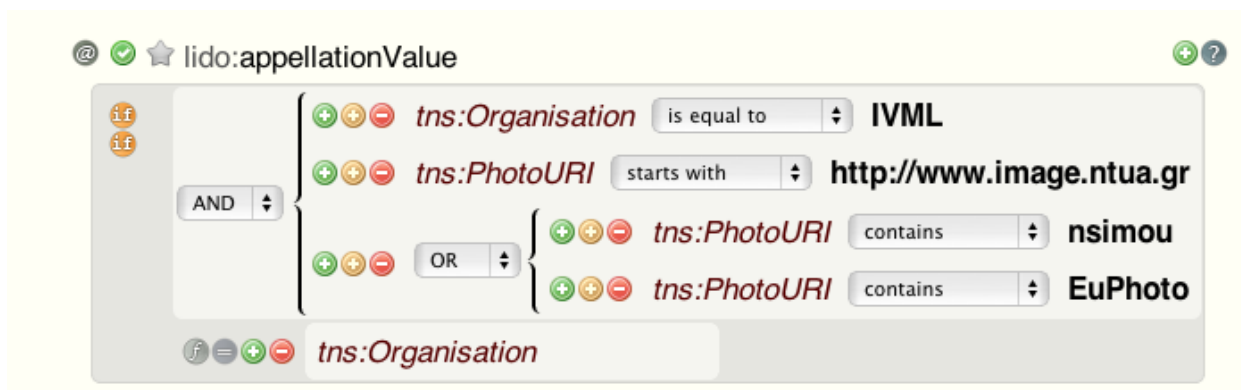


Figure 47: Conditional mapping

In the figure above we see a conditional mapping with many clauses, that is interpreted in the following way. If the value of xpath `tns:Organisation` **is equal to** “IVML” and the value of xpath `tns:PhotoURI` starts with “`http://www.image.ntua.gr`” and either the value of xpath `tns:PhotoURI` contains “`nsimou`” or “`EuPhoto`” then the value of the xpath `tns:Organisation` will be mapped to the xpath `lido:appellationValue` of the target schema.

5.3.5.7 Structural mapping

Structural mappings are a special category of mapping. As you may have noticed in the complex types of the target schema illustrated in the mapping area text structural appears instead of unmapped. At this area you can map complex elements of your input metadata. Let’s assume you have a complex element named “B” in your metadata – having “C” as a child – and B appears 3 times in only one record. Also assume complex element “BTarget” – having “CTarget” as a child – in the target schema. If you make a structural mapping of B (by dragging it and dropping it) to “BTarget” and then you map “C” to “CTarget” 3 “BTarget” complex elements will appear in the output XML having “CTarget” as child and having the values of “C”.

5.3.5.8 Mapping Using Terminologies

As written in the DoW the main objective of WP 4 is the creation of terminologies that will be used in specific LIDO elements (see Euroepana Photography Mapping Guidelines) for the enrichment of your metadata. There are two ways of using terminologies for the enrichment of your metadata through MINT. The first is to assign a term to your complete dataset (similarly to constant mapping). To do that double click on an element that takes values from a terminology (e.g. click on bookmark Technique and then double click on the `lido:conceptID`) and the following window appears from where you can select a term by clicking on it. (Only terms written in black can be used for mappings, the grey terms are used for the completeness of the terminology, see figures below)

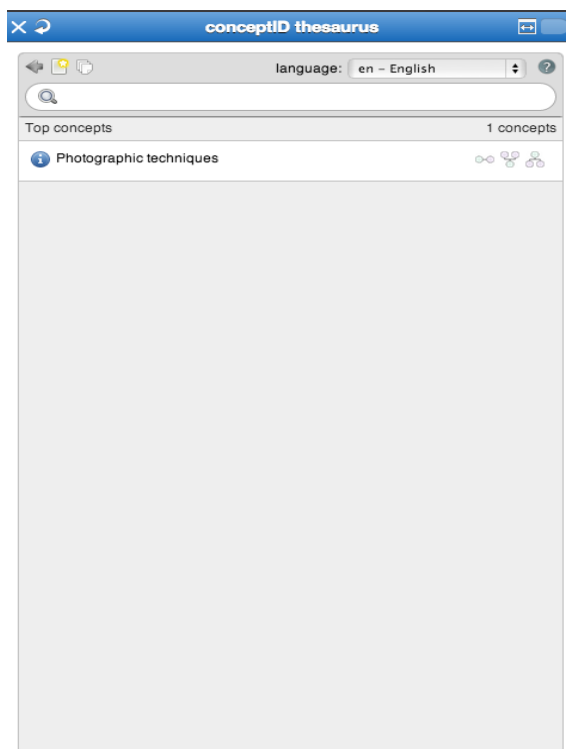


Figure 48: Terminology navigation pane

In this pane the top concepts of the terminology appear (in this case it is only one). On the top right you can select the language you wish from the drop down menu (note however that the selection of a language other than English may result in the appearance of fewer terms because not all the terms are yet translated to all the languages) while on the left there buttons that allow you to see all the concepts of the terminology (📄), the top concepts (📄) or to go back to his/her previous view (🏠). In addition you can use the search pane on the top to search for a specific concept or to explore the terminology’s hierarchy by selecting the related (🔗), the broader (👉) or the narrower (👈) terms.

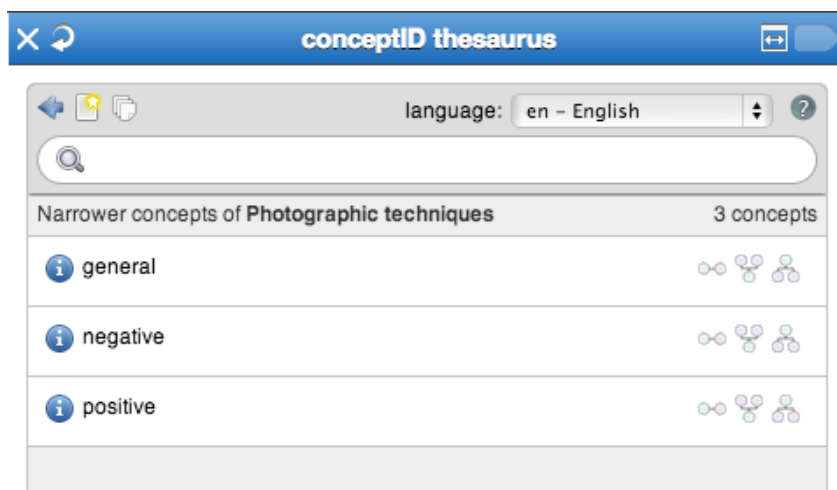




Figure 49: Narrower terms of term Photographic Techniques

The second way for using the terminologies through MINT is based on the value mapping. First you perform an xpath mapping using the xpath of your input schema that contains your in-house terminology. After that by clicking on the 🍽️ icon you can perform a value mapping i.e. to map your in house

terminology values to the terminology's terms. On the top of the pane you browse input values by clicking on  button, while below you select the desired term from the terminology again by clicking on .

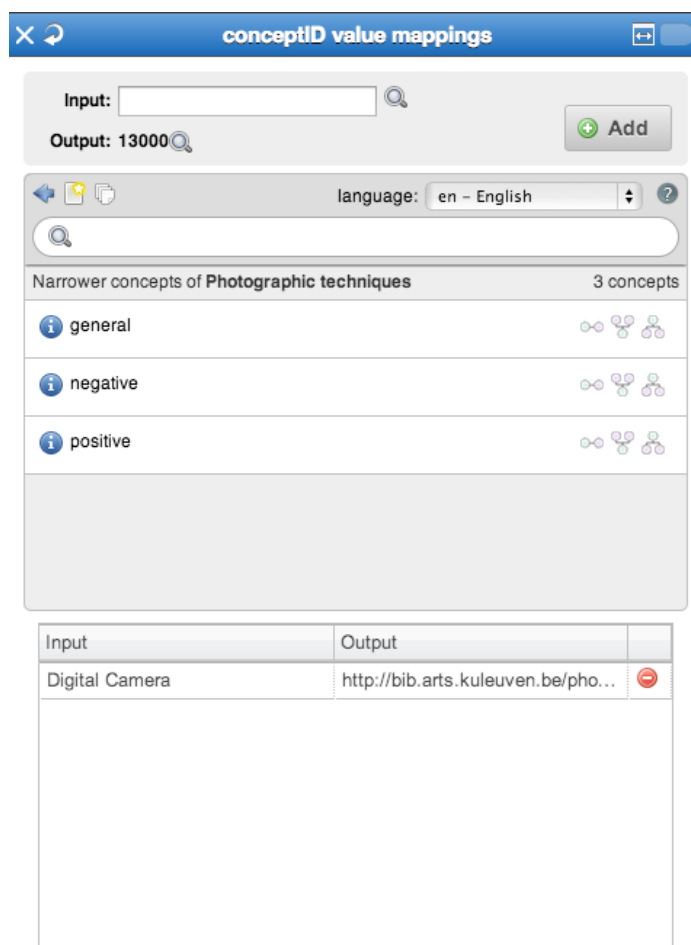




Figure 50: Value mapping using a terminology

After that you add the value mapping link and the mapping you've entered appears in the text area below. You can remove a value mapping by clicking on the  button on the right of it.

5.3.6 Preview items

After having performed a mapping you can preview the input files, the XSL created so far, the output files (i.e. the metadata formed in the Europeana Photography intermediate format), the ESE files, the Europeana preview and a report of the unmapped elements or mistakes by clicking on the  button on the top right of the mapping tool.

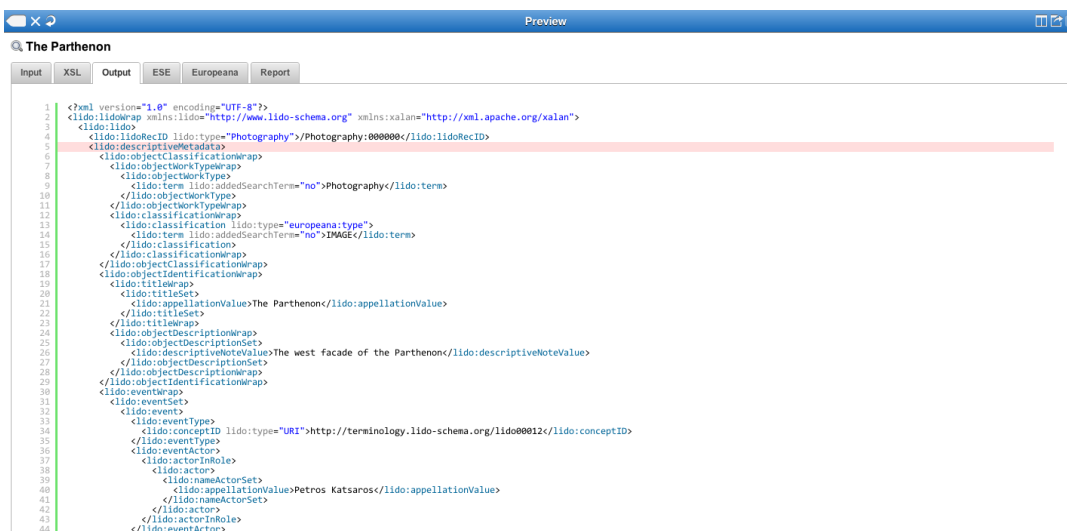


Figure 51: The output preview with an error highlighted

The above figure shows the output preview with an error highlighted. More information on the errors is shown in the “Report” pane.

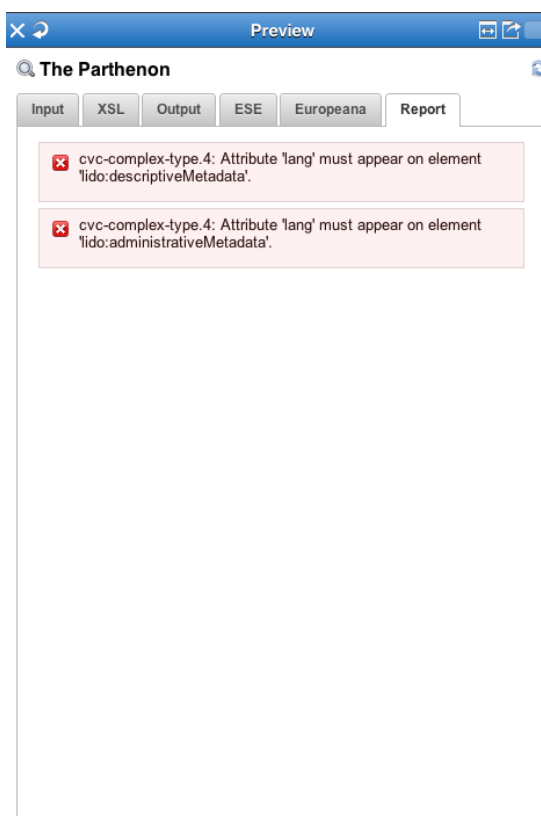



Figure 52: The report pane

The following figure shows the Europeana preview that is very useful since it simulates the view of your items on the Europeana portal. Note that by clicking on the  button next to the item title on the top you can switch the preview to other items.

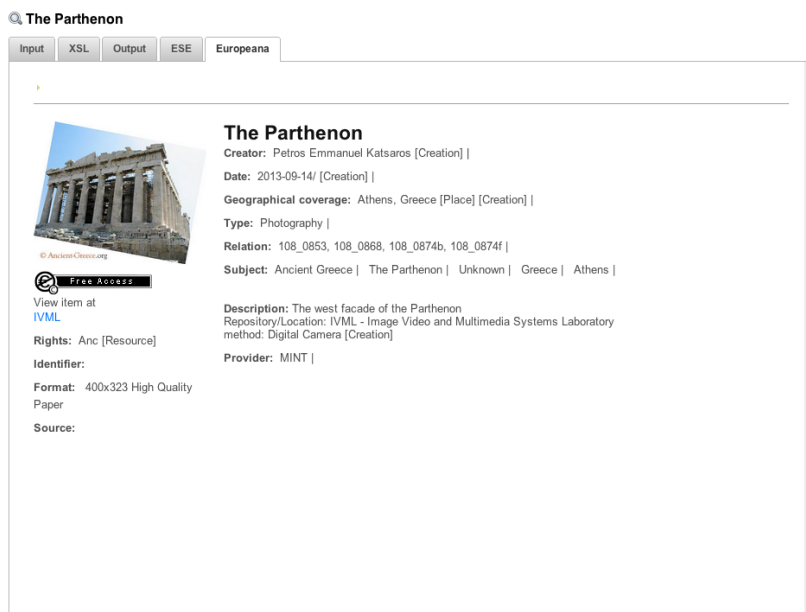


Figure 53: The Europeana preview

5.4 TRANSFORMATION - PUBLICATION TO EUROPEANA

After having performed mappings (see Section 5.3.5) select the “Transform” for the “Dataset Options”. The “Transform” pane appears from which you can select a mapping for doing the transformation.

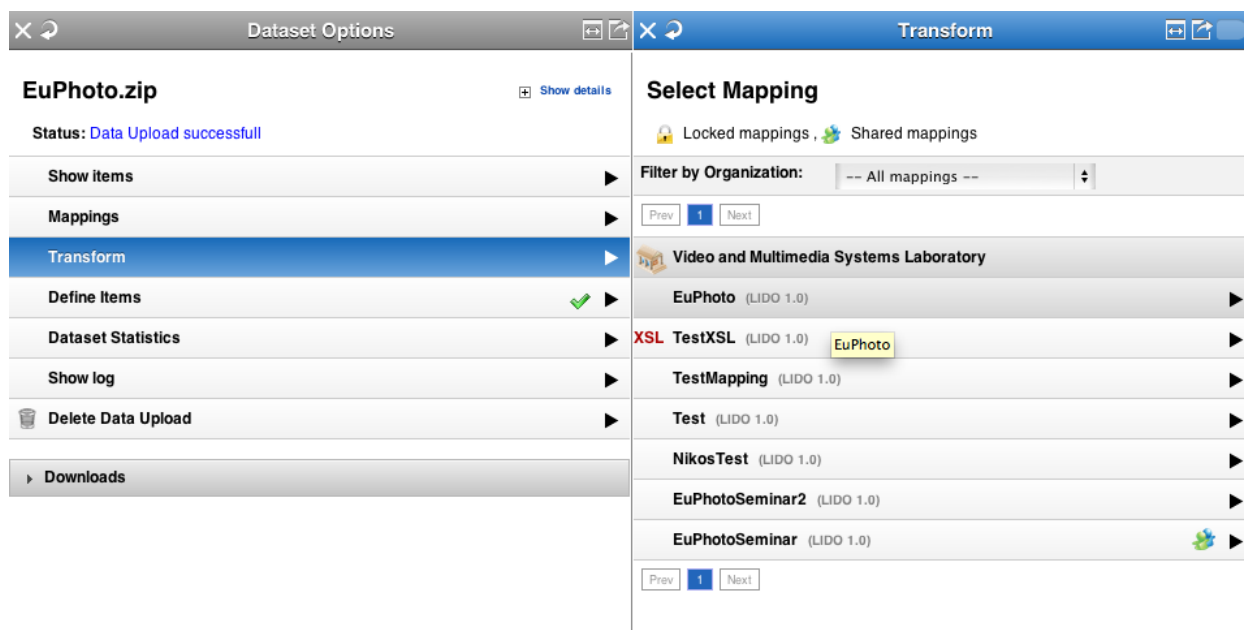



Figure 54: Transform dataset

After that the transformation the icon  appears next to the dataset and prepare for publish option appears to the “Dataset Options” pane.

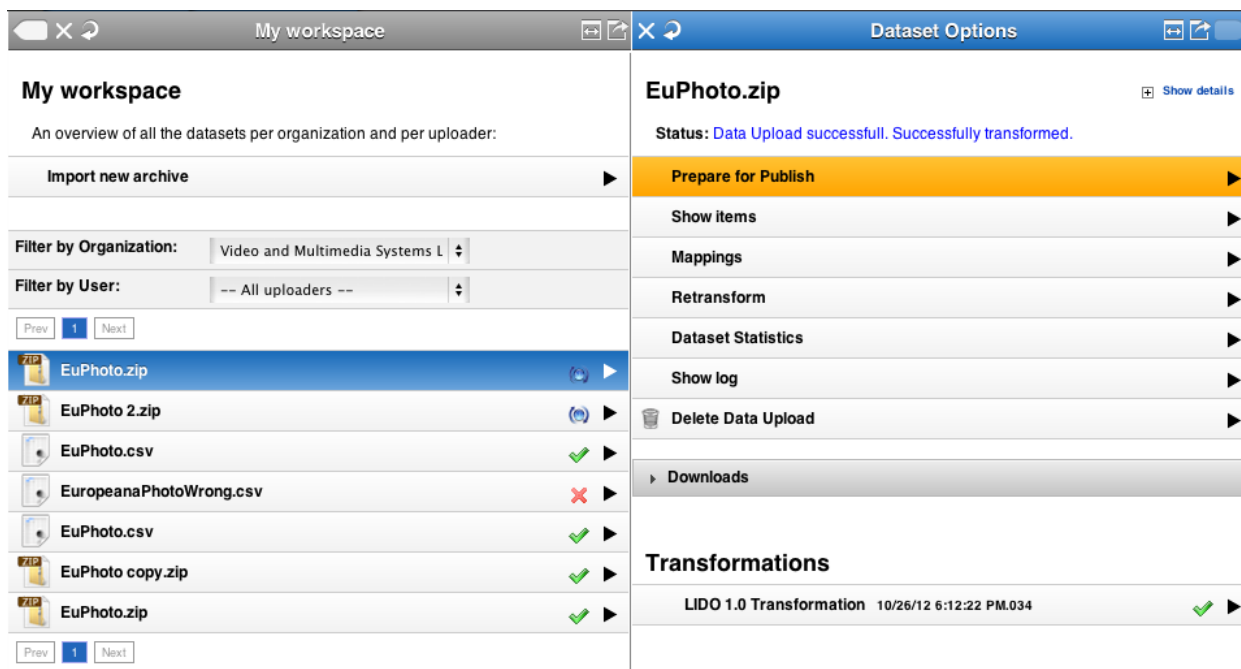


Figure 55: Transformed dataset

In addition the transformation appear on the bottom and by clicking on it the following window appears that allows you to view and download the transformed items.

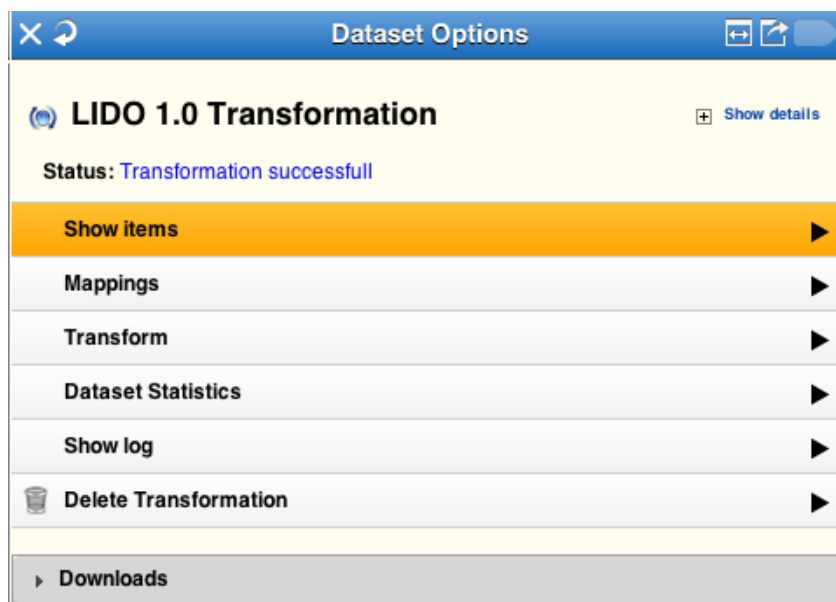


Figure 56: Transformation options

By clicking the prepare for publish option the following pane appears that asks you whether you agree to the Europeana Data Exchange Agreement which you can see by clicking on the link. After clicking on the agree check box you are asked to select the Europeana rights. Finally you can also select the metadata set to publish to Europeana. The first option (CC0) is the default and if selected the complete dataset is published (recommended). However for cases where this is not possible two more options exist one for not publishing the descriptions of the transformed metadata and one for publishing only the required fields by Europeana. After making your choices press the submit button to publish.

Europeana Agreement

Publication options

⚠ To publish you must agree with the terms and conditions of the [Europeana Data Exchange Agreement](#) .

I agree

Select the Europeana rights.

- Europeana: Rights Reserved - Free Access
- Europeana: Rights Reserved - Paid Access
- Europeana: Rights Reserved - Restricted Access
- Europeana: Unknown copyright status

By default the **full metadata** set will be published. For publication options [click here](#) !

Select the metadata set to publish to Europeana.

- Full (CC0)
- Intermediate (CC0 - No descriptions)
- Minimal (CC0 - Mandatory Only)

Submit

Figure 57: Prepare for publish to Europeana

6 MINT SUPPORT

The transformation and the publication of providers' metadata to the intermediate schema of Europeana Photography is a process during which providers may encounter difficulties, even if they are well aware of the MINT's functionality. This is because the majority of the providers do not have a strong technical background and it may be hard for them to combine MINT's functionality to reach the desired result.

For that reason a mailing list has been created (EPmintsupport@promoter.it) to support the content providers overcome their problems. The list is maintained by Promoter and is supported by NTUA with the assistance of KMKG in order to assure feedback to each kind of potential problem (mapping, use of MINT, software bugs, vocabularies, etc.). More specifically the content providers are able to get instructions on how to map their in-house metadata to LIDO as well as technical information about MINT functionalities that permit them to fit their metadata perfectly according to the Europeana Photography requirements.

In addition, hands on training sessions (have been and) will be held together with the plenary meetings of Europeana Photography during which providers could directly ask for assistance on their metadata - specific matters.

7 TECHNICAL SPECIFICATIONS

7.1 PLATFORM

It is written in JAVA, JSP, HTML and Javascript. It uses PostgreSQL as an object-relational database with Hibernate as the data persistence framework, and mongoDB as a document-oriented database. MINT is also reusing other open source development frameworks and libraries according to specific deployments and customizations. Mint source code versions are released under a free software license (GNU Affero GPL).

The platform offers a user and organisation management system that allows the deployment and operation of different aggregation schemes with corresponding user roles and access rights. A Restful web service is available for user management and authentication.

7.2 INGESTION

Registered users can upload their metadata records in XML or CSV serialization, using the HTTP, FTP and OAI-PMH protocols. Users can also directly upload and validate records in a range of supported metadata standards (XSD). XML records are stored and indexed for statistics, previews, access from the mapping tool and subsequent services.

Current developments aim to support relational database schemata and OWL/RDFS ontologies as input.

7.3 PROCESSING

Handling of metadata records includes indexing, retrieval, update and transformation of XML files and records. XML processors (Apache Xerces, SAXON, Nux) are used for validation and transformation tasks as well as for the visualization of XML and XSLT. For issues of scalability with respect to the amount of data and concurrent heavy processing tasks, parts of the services are multi-threaded or use specific queue processing mechanisms.

7.4 NORMALIZATION & VOCABULARIES

Various additional resources such as terminologies, vocabularies, authority files and dictionaries are used to reinforce an aggregation's homogeneity and interoperability with external data sources. A typical usage scenario is the connection of a local (server) or online resource with a metadata element in order to be used during mapping/normalization. The vocabularies have to be represented in SKOS.

8 CONCLUSION

The present document constitutes the report of deliverable D5.2 “The MINT mapping tool”, that is made available online for validation and for the large-scale contribution of content to Europeana and for dissemination & training (WP5, WP7). In addition the delivery of MINT mapping tool is related to WP6 “Sustainability and IPR”. In detail MINT mapping tool uses once-only mappings and simple re-use of local “source” metadata and in that way takes full advantage of the funded project to make a very low-cost continuation possible. NTUA hosts the tool and the provision of these services has minimal additional cost.

The platform implements an aggregation infrastructure offering a crosswalk mechanism to support subsequent critical activities:

- harvesting and aggregating metadata records that were created using shared community standards or proprietary metadata schemas,
- migrating from providers’ models (whether standard or local) to a reference model,
- transforming records from the Europeana Photography model to the Europeana Semantic Elements and the Europeana Data Model.

8.1 RESULTS

The objective of the deliverable is the deployment of the MINT mapping Tool (Tasks 5.3, 5.4), that is available online to the project partners at <http://mint-projects.image.ntua.gr/photography>

8.2 IMPACT

The delivery of the platform and the present report achieve a project milestone, MS13 MINT Mapping Tool available, Month 9. This enables providers to start processing metadata and towards Indicator No 5 “1,000 photographs successfully delivered to Europeana”. In detail 12.300 images have been successfully transformed to LIDO and are ready to be ingested by Europeana.

APPENDIX I: DEFINITIONS OF TERMS AND ABBREVIATIONS

CSV	Comma Separated Values
DC	Dublin Core
EDM	Europeana Data Model
ESE	Europeana Semantic Elements
LIDO	Lightweight Information Describing Objects
XML	Extensible Markup Language
XSD	XML Schema

APPENDIX II: MAPPING OF EUROPEANA PHOTOGRAPHY MANDATORY FIELDS TO LIDO

ESE mandatory	Minimal EU Photo	EU Photo Intermediate	Remarks	LIDO Element
		dc:Identifier	Identifier of the record (local or URI)	/lido/administrativeMetadata/recordWrap/recordID
edm:country		edm:country	country of the data provider	Added by Europeana Ingestion team after providers' publication
edm:isShownAt		edm:isShownAt	link to metadata on your website	/lido/administrativeMetadata/recordWrap/recordInfoSet/recordInfoLink
edm:isShownBy		edm:isShownBy	link to photo on your website	
edm:provider		edm:provider	name of the aggregator, if applicable	"Europeana Photography"
edm:dataProvider		edm:dataProvider	name of the metadata provider	/lido/administrativeMetadata/recordWrap/recordSource/legalBodyName/appellationValue
edm:type		edm:type	type of content partner delivers to Europeana	/lido/descriptiveMetadata/objectClassificationWrap/classificationWrap/classification/term@type=europeana:type
edm:rights		edm:rights	link to the Europeana rights documentation	/lido/administrativeMetadata/recordWrap/recordRights/rightsType/conceptID = enumerated
edm:uri		edm:uri	permanent link where the metadata is shown	Added by Europeana Ingestion team after providers' publication /lido/administrativeMetadata/recordWrap/recordInfoSet/recordInfoLink
edm:language		edm:language	language of the metadata provider	Added by Europeana Ingestion team after providers' publication /lido/descriptiveMetadata/@ lang
dc: type		dc: type	Always "Photography", as agreed with partners @ content seminar	Always "Photography", as agreed with partners @ content seminar /lido/descriptiveMetadata/objectClassificationWrap/objectWorkTypeWrap/objectWorkType
dc:title	dc:title	dc:title	title of your work, can be the caption	/lido/descriptiveMetadata/objectIdentificationWrap/titleWrap/titleSet/appellationValue

ESE mandatory	Minimal EU Photo	EU Photo Intermediate	Remarks	LIDO Element
	date	dcterms:temporal dc:date	date when the photo was taken, so between 1838-1939.	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventDate/date/earliestDate
	author	Author Dc:actor	name of the photographer, or studio, who made the photo	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventActor/actorInRole/actor/nameActorSet/appellationValue
	technique	Technique Dc:description	photographic technique used to make the photo, <u>to be linked to EuPhoto vocabulary</u>	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventMethod/conceptID
	location	dcterms:spatial	place where the photo was taken, <u>to be linked to Geonames or other</u>	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventPlace/displayPlace /lido/descriptiveMetadata/eventWrap/eventSet/event/eventPlace/place/namePlaceSet/appellationValue
	description OR keywords	dc:subject and dc:description dc:description	since not everyone has keywords, a description can be enough. Preferably keywords however	/lido/descriptiveMetadata/objectRelationWrap/subjectWrap/subjectSet/subject/subjectConcept/conceptID or term description: lido:objectIdentificationWrap/lido:objectDescriptionWrap/lido:objectDescriptionSet[lido:descriptiveNoteValue/string-length(.)>0]/
	copyright	Copyright Dc:right	text string with the name of the copyright holder. If empty, the institution's name will be shown	/lido/administrativeMetadata/recordWrap/recordRights/rightsHolder/legalBodyName/appellationValue
		photographic practice	Why was a photo taken? Preferably in keywords, not always available in partners' metadata (or not always useful)	/lido/descriptiveMetadata/objectRelationWrap/subjectWrap/subjectSet/subject/subjectConcept/conceptID or term
		subjectConcept dc:subject	keywords	/lido/descriptiveMetadata/objectRelationWrap/subjectWrap/subjectSet/subject/subjectConcept/conceptID or term

ESE mandatory	Minimal EU Photo	EU Photo Intermediate	Remarks	LIDO Element
		subjectActor dc:subject	person IN the photo, or who made the object in the photo	/lido/descriptiveMetadata/objectRelationWrap/subjectWrap/subjectSet/subject/subjectConcept/conceptID or term subjectActor
		subjectPlace dc:subject	place IN the photo, usually the same as production place, except for paintings	/lido/descriptiveMetadata/objectRelationWrap/subjectWrap/subjectSet/subject/subjectConcept/conceptID or term subjectPlace
		Dimensions Dc:decription	measurements of the original	/lido/descriptiveMetadata/objectMeasurementsWrap/objectMeasurementsSet/displayObjectMeasurements
		Material Dcterms:medium	material of which the original is made	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventMethod/conceptID eventMaterialsTech
		related works term=parOf	links between several records, from the same album, different pictures from the same object, ..	/lido/descriptiveMetadata/objectRelationWrap/relatedWorksWrap/relatedWorkSet/relatedWork/object/objectWebResource or objectID objectNote

Table 3: Mapping of Europeana Photography mandatory fields to LIDO