Project

Project Acronym:	AthenaPlus
Grant Agreement number:	325098
Project Title:	Access to cultural heritage networks for Europeana

Deliverable

Deliverable name:	**The MINT ingestion platform**
Deliverable number:	**D3.1**
Delivery date:	August 2013 – Revision June 2014
Dissemination level:	Public
Status	First Draft
Authors (organisation)	Nikolaos Simou (NTUA), Eleni Tsalapati
Contributors (organisation)	Nasos Drosopoulos (NTUA)
Reviewers (organisation)	Erik Buelinckx (KIK-IRPA); Marzia Piccininno (ICCU)

Revision History

Revision	Date	Author	Organisation	Description
V0.1	2013-08-28	Nikolaos Simou	NTUA	First Draft
V0.2	2013-09-15	Eleni Tsalapati	NTUA	Update
V0.3	2013-10-18	Nikolaos Simou	NTUA	Version for review
V0.4	2013-10-31	Nikolaos Simou	NTUA	Final Version
V1	2013-10-31	Nikolaos Simou	NTUA	Final Version – submitted
V1.1	2014-06-13	Nikolaos Simou	NTUA	Update according to reviewer's comments
V1.2	2014-06-13	Maria Teresa Natale	ICCU	Formal check

Table of Contents

1 EXECUTIVE SUMMARY

The main objective of this deliverable is to describe the MINT ingestion platform presenting its architecture and the technologies used for its implementations. Additionally, the MINT workflow is explained and detailed user instructions for every MINT procedure are provided. The role of MINT in AthenaPlus project is to assist content providers in the transformation of their in house metadata to Europeana Data Model and their publication to Europeana. MINT development has started by NTUA from the ATHENA[1] ingestion server and evolved through other Europeana feeder projects like Linked Heritage[2], EuScreen[3] and ECLAP[4]. MINT 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 ingestion platform in the AthenaPlus project is to enable users to

- Provide metadata records in a range of "source" formats
- Convert metadata to LIDO (that is used as an intermediate standard)
- Transform LIDO metadata to the latest version of EDM
- Publish metadata to Europeana
- Monitor the progresses of content provision

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 of their metadata to LIDO through a very user-friendly interface (see figure below).
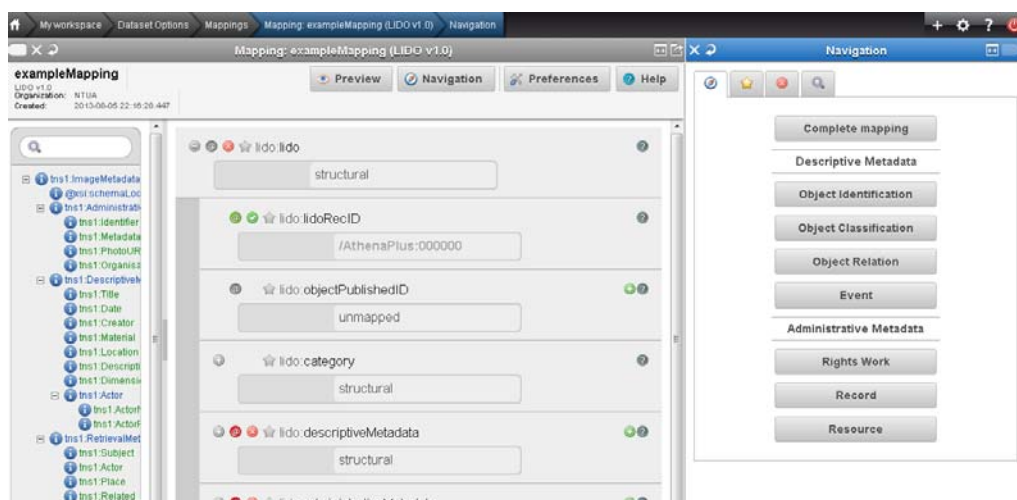
Figure 1:MINT's mapping editor

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. This tree is extracted from the user's import and represents the snapshot of the XML schema that is used as input for the mapping process.

---

[1] http://www.athenaeurope.org/
[2] http://www.linkedheritage.eu/
[3] http://euscreen.eu/
[4] http://www.eclap.eu/drupal/?q=en-US

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 as links for accessing their corresponding sub-elements. These are visualized in 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.

After the creation of a valid mapping, content providers can transform their metadata to the LIDO. This action is followed by another transformation, this time LIDO metadata is transformed to ESE and EDM[5]. This is performed before publishing (transmitting) metadata to NTUA's OAI-PMH server. 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.
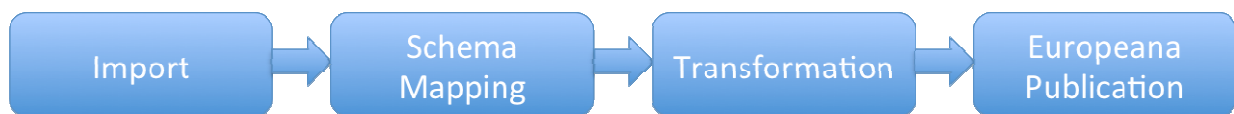

Figure 2: Overall workflow

The MINT ingestion platform has been used in many European projects for delivering metadata to Europeana, however many specializations have been implemented to meet the special needs of each project content providers.

In detail, the MINT instance that was set up for the Athena Plus project differs from MINT instances used in other projects in the following

- It has been customized so as to assist the content providers to easily map their metadata to LIDO with the aid of element's bookmarks.
- The MINT mapping editor has been totally re-constructed taking into account the experience gathered from the usage of MINT in other projects and also the feedback from content providers. The new mapping editor is based on a new JSON library (JsonSmart) that is much faster and responsive than one used in previous versions. It also permits the creation of more complex mappings by the use of structural mappings on any element, and if-else conditional mappings making in that way the implementation of crosswalks from any provider's input schema to LIDO possible. In addition the MINT tool has been appropriately modifies so as to be accessible and fully functional from tablets and smartphones.
- It fully supports the Europeana Data Model (crosswalk from LIDO to EDM and Europeana previews based on their new portal).
- The repository exposes the records in RDF and OAI-DC fully supporting in that way the harvesting from Europeana (RDF) and the OAI-PMH interface.
- Progress reports are supported for download and also through MINT per content provider for monitoring of the metadata production and publication to our (OAI-PMH) repository.
- Progress reports for the metadata production and publication to our (OAI-PMH) repository are supported either as a pdf document for download or as charts that are visible through MINT.

The delivery of the platform and the present report achieve a project milestone, MS7 MINT/Athena Plus ready for delivering content to Europeana, Month 6. This constitutes the starting point of the aggregation of metadata, which will be finalized by Month 30. The delivery of this report was originally planned for month 6, as well, however it has been delayed for two months. This delay was caused by the integration of the latest MINT developments and specializations required for the AthenaPlus project that are described in detail in section 7. Since the present report provides reference to the content providers of AthenaPlus for the use of the MINT ingestion platform, the development of its core functionalities and the project specializations had to be finalized before authoring it.

---

[5] This step appears as 'Prepare for Publish' in the MINT tool.

2   INTRODUCTION

This deliverable describes the MINT ingestion platform that is used in the AthenaPlus project for the large-scale contribution of metadata to Europeana. The development of MINT started in the ATHENA[6] project when the NTUA team integrated all the necessary components for publishing metadata to Europeana into a common technology platform, while it evolved through its use in other Europeana-feeder projects like Linked Heritage[7], EuScreen[8], ECLAP[9], Fashion[10], EuropeanaPhotography[11] and others. The MINT platform provides content holders with the ability to perform the required mapping of their own metadata schemas to LIDO as well as its publication to Europeana. It addresses the ingestion of metadata from multiple sources, the mapping of the imported records to a target 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.

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

The MINT ingestion 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 ;

---

[6] http://www.athenaeurope.org/
[7] http://www.linkedheritage.eu/
[8] http://euscreen.eu/
[9] http://www.eclap.eu/drupal/?q=en-US
[10] http://www.europeanafashion.eu/
[11] http://www.europeana-photography.eu/

- transforming records from the project's intermediate standard to the Europeana Semantic Elements and the Europeana Data Model.

2.2 Role of this Deliverable in the Project

This deliverable corresponds to the MINT ingestion platform that implements tasks:

- 3.1 "Compliance with CH metadata modelling requirements"
- 3.2 "Help desk support system"
- 3.3 "Metadata Ingestion"
- 3.4 "Delivery to Europeana".

The MINT ingestion platform enables users to

- provide metadata records in a range of "source" formats ;
- convert metadata to LIDO that is the AthenaPlus' intermediate metadata standard ;
- submit the records to Europeana.

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 a project milestone (MS7 MINT Ingestion platform available, Month 6) and constitutes the starting point for mapping and delivery of metadata to Europeana.

2.3 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, according to the Europeana 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 ingestion platform 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 Europeana (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 Europeana 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 Europeana.

The EUscreen project also follows the same aggregation workflow for Europeana 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 Europeana (OAI-PMH for ESE).

Similarly, the CARARE, JUDAICA Europeana, ECLAP, DCA, Linked Heritage, and AthenaPlus projects utilize MINT to accommodate their aggregation and remediation requirements for their specific domain and project, and for Europeana. The group is actively participating in several metadata modelling 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 Europeana Data Model harvesting XSD and RDFS ontology, and EBU Tech. It is also involved in the development of Europeana and the Digital Public Library of America. The growing user base (more than 400 cultural heritage organizations and 600 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.

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

2.4 Structure of the document

The rest of the document presents more details on the aggregation workflow that is implemented in the AthenaPlus Ingestion Platform (Section 3), emphasising on the metadata ingestion platform (Section 4). The following section (Section 5) provides the detailed technical specifications for the implementation of MINT. Section 6 presents the latest features of MINT while the section that follows presents the developments and the operations that the NTUA team has performed specifically for the AthenaPlus project. Finally, the user manual is provided in Appendix 2.

3 INGESTION WORKFLOW TO EUROPEANA USING MINT

The developed system facilitates the ingestion of semi-structured data and offers the ability to establish crosswalks to the reference schema (LIDO) 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.
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.
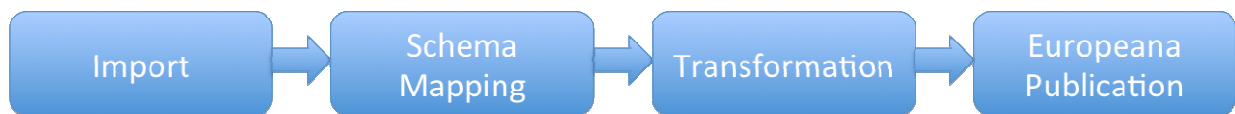
Figure 3: Ingestion workflow

The metadata ingestion workflow, as illustrated in *Figure 3*, consists of four main steps. First is the Import of provider's metadata using common data delivery protocols, such as OAI-PMH, HTTP and FTP. Following is the Schema Mapping procedure, during which the imported metadata are mapped to LIDO, the intermediate metadata standard – between provider's metadata and EDM – used in AthenaPlus. A graphical user interface assists content providers in mapping their metadata to LIDO, using an underlying machine-understandable mapping language. Furthermore, it provides useful statistics about the provider's metadata also supporting the share and reuse of metadata crosswalks and the establishment of template transformations. The third step is the Transformation procedure, during which provider's metadata is transformed to LIDO by using the mapping they made in the previous step. 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.

4   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 centrepiece. 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.
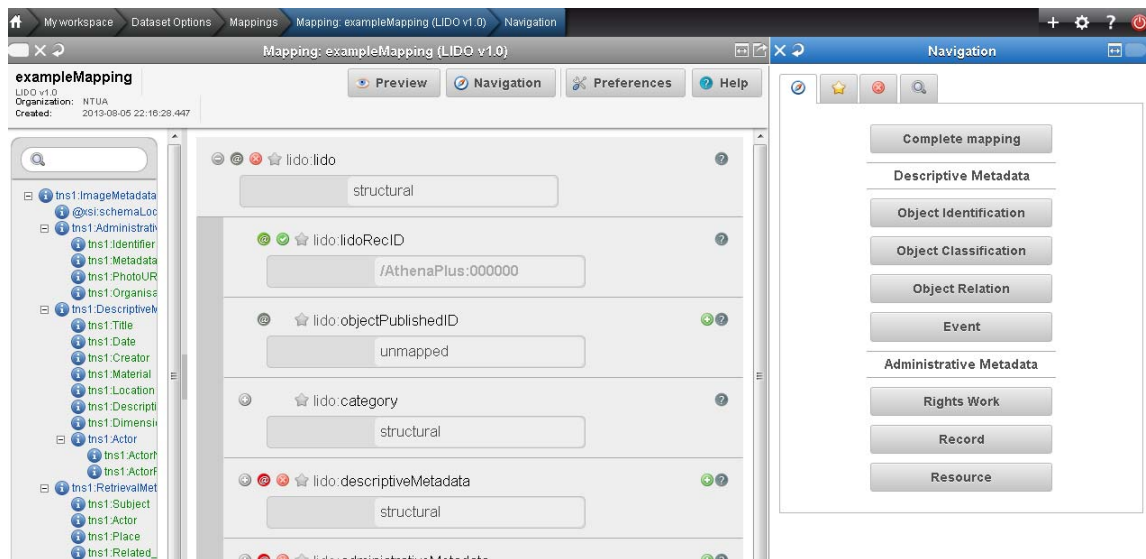
Figure 4: Screenshot of the mapping editor

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. At this point it is important to mention the mapping download in both MINT and XSL format is functionality that can prove very useful for Task 2.2 "Survey and description of mappings suitable to LIDO" 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.
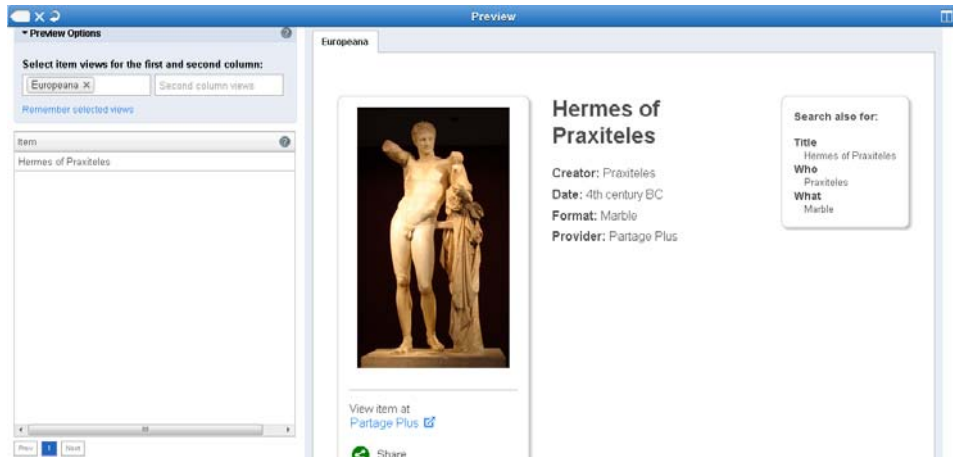
Figure 5: Transformation and HTML rendering preview

5 TECHNICAL SPECIFICATIONS

5.1 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.

5.2 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 ingestion platform and subsequent services.

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

5.3 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.

5.4 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.

6 MINT2 Latest features

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 other deliverables authored by the NTUA team for other European aggregation projects like AthenaPlus 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",
- EuropeanaPhotography D 5.2 "The MINT Mapping tool"

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 a common metadata standard for the project and then its delivery to Europeana according to the project specific requirements – remains the same, many new features have been implemented to meet the special needs of the content providers. In detail MINT has evolved through its use in many aggregation projects to a very effective and multi-functional tool. The instance that has been set up for the AthenaPlus projects is a new version of MINT called MINT 2 and this section presents the latest features of it.

6.1 User Interface

A major development in the MINT2 is its new user interface. The main objective was to redesign the user interface of MINT 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 updated user interface.

6.1.1 *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 YUI[12]. The interface approach followed there was tab-oriented as shown below

Figure 6: MINT's old user interface

---

[12] http://yuilibrary.com/

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 main problem with that approach however was that actions were not related and the users did not have any guidance thought 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. For that purpose in the new user interface of the MINT mapping tool, a different approach has been followed and the jQuery library Kaiten[13] 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.
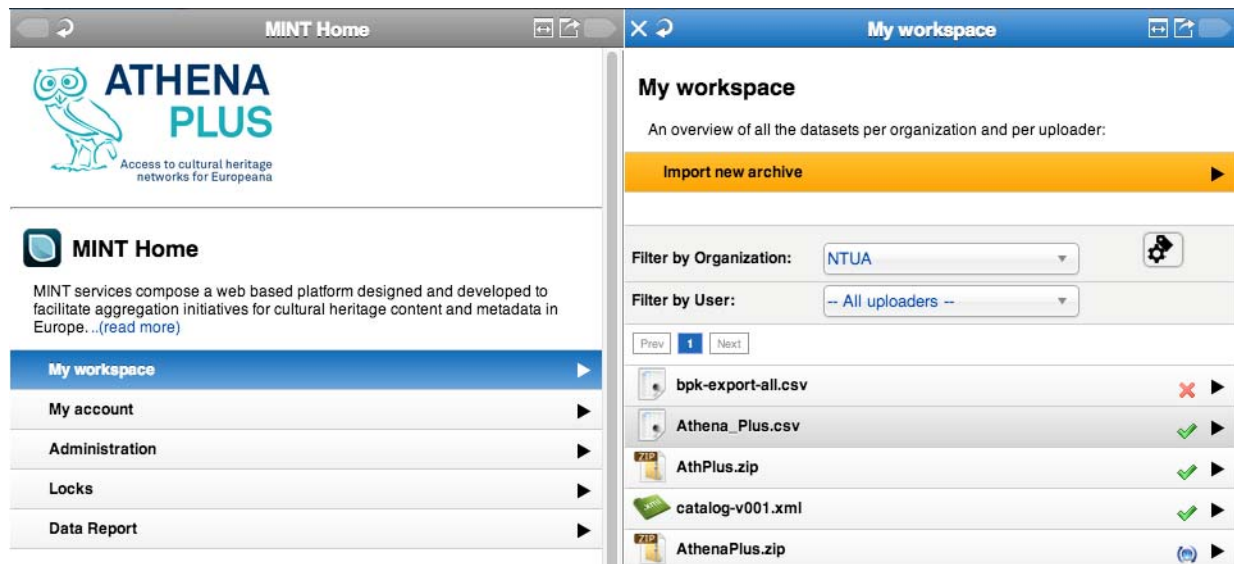
Figure 7: MINT's new interface (My Workspace)

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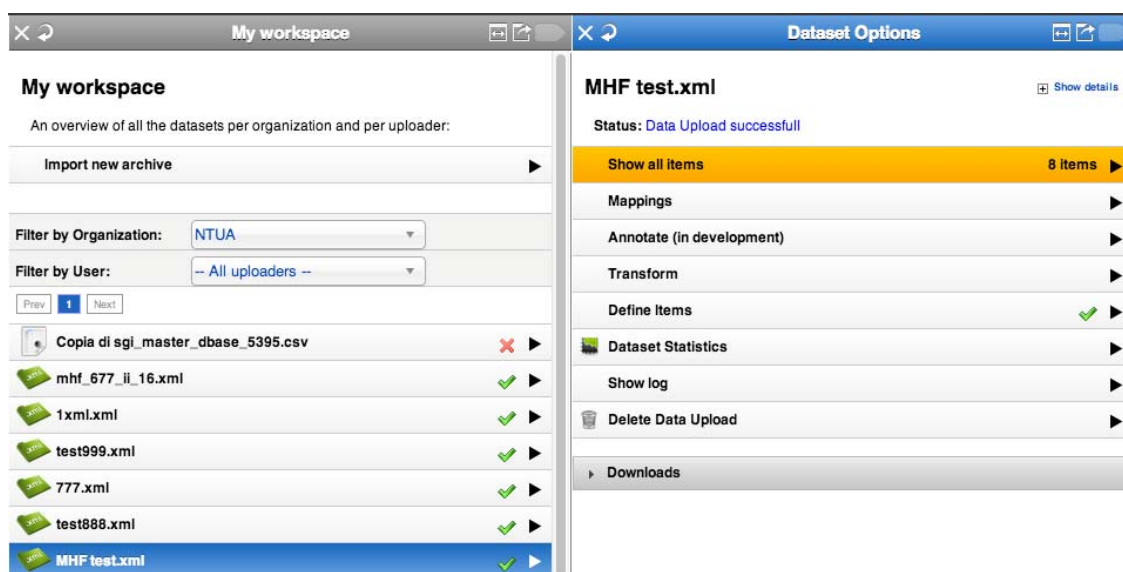
Figure 8: MINT's new user interface (Dataset Options)

---

[13] http://www.officity.com/kaiten/

6.1.2 *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 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 AthenaPlus recommended elements.

Figure 10: MINT's new mapping editor

6.1.3   *Improved item preview interface*

Another update of the MINT release used for AthenaPlus project is the brand new item preview interface. In the previous version of the tool the user could only see the preview of one item while working on a mapping. This was not always very convenient, since different items of an import may vary on the information (elements of input metadata) they had. For example the first item of an import could have element 'Author' while the rest of the imports may have missed that. Cases like were resulting in transformations with many errors simply because the user did only have feedback of his mapping for the first record.
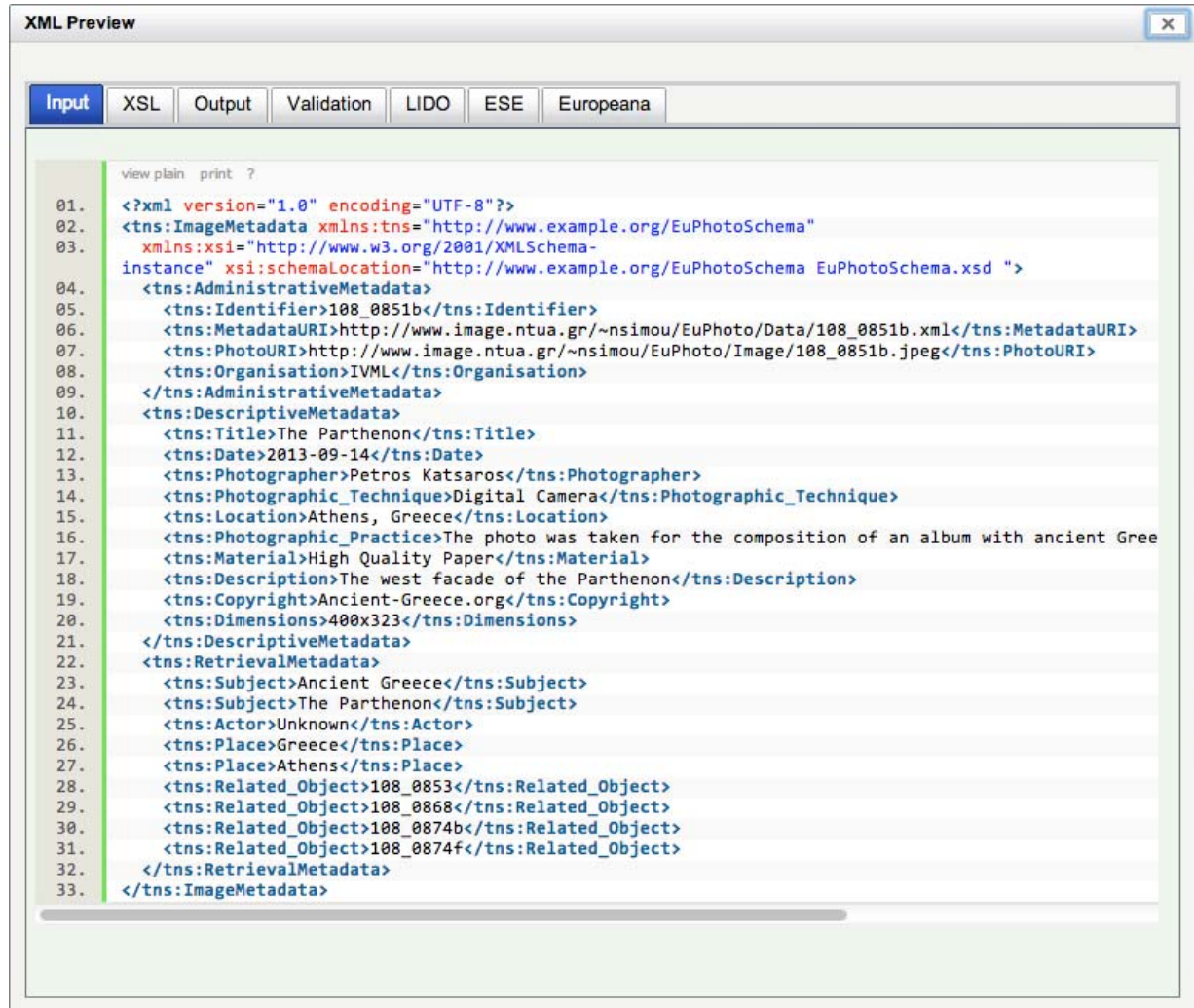
Figure 11: MINT's old item preview interface

The new Item Preview interface does not suffer from these limitations, in detail the user can:

- preview all his items while working on a mapping;
- select the previews he prefers to see in the two preview panes;
- get inline validation for each preview according to its' related XML schema;
- instantly select a different mapping and preview its result for the same item;
- search the items he wishes to preview (this feature is still under development).

AthenaPlus D 3.1 The MINT ingestion platform

Figure 12: MINT's new item preview interface

6.2 MAPPING FUNCTIONALITY

MINT's main functionality is the mapping and transformation of the providers' in-house metadata to the AthenaPlus reference schema and its publication to Europeana. Content providers perform this process by using MINT's mapping editor, which offers a wide range of mapping functionalities. Therefore the development of new mapping functionalities always comprises part of NTUA's on-going work. More specifically the mapping editor has been totally reconstructed using a new JSON library (JsonSmart) that is much faster than the one used in previous versions. In some tests we performed comparing the two libraries, we observed that JsonSmart only required a couple of seconds for the parsing of a mapping while for the same mapping the previous library may have required a minute of parsing. The following sections describe in detail the latest mapping functionalities.

6.2.1 *Negative closure for functional mappings*

An important mapping functionality that has been implemented in the new version of MINT is the negation for all the conditions used in conditional mappings. 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.

6.2.2 *If else mapping*

In addition to the negative closure for all the MINT functions a user can also make mappings using if else statements. This functionality can be better understood by the example shown below.

The table on the right displays all the values of element 'tns:title'. This XPath is first mapped to element 'dcterms:title'. After that a conditional mapping is made by clicking on the 'if'-button. Then by clicking on the '+'-button located below the 'if'-button additional conditions can be set. The last condition set is the closure of all the previous. In simple words the condition set in the example below is the following.

- if the value of element 'tns:Title' contains "East" then map the value of the element 'tns:Title' to 'dcterms:Title'
- else if the value of element 'tns:Title' contains "South" then map the value of the element 'tns:Title' to 'dcterms:Title'
- else map the value "Unknown" to 'dcterms:Title'

So for the items having 'tns:Title' values 'Caryatids', 'Parthenon' and 'The Parthenon' the value 'Unknown' will be mapped[14].

Figure 13: If else mapping

6.2.3 *Structural on any element*

Another very important mapping functionality available in the MINT release for AthenaPlus is the "structural on any element". The need for that functionality occurred from the complexity of some metadata standards. Some of the providers have metadata exports in schemas that had very extensive hierarchical structure. When attempting to map the very well defined information to a schema that was less expressive they could not exploit their in house structure in their mapping.

The "structural on any element" mapping functionality solves that problem by permitting to content providers of such schemas to threat each target schema simple element (i.e. 'lido:term') as a complex type taking in that way advantage of their in house metadata hierarchy and of the mappings that can be set to a structural element through MINT. The structural mapping area appears by clicking on the elements' name.

Figure 14: Structural on any element

6.2.4 *SKOS Vocabularies Support*

MINT's functionality extended to support SKOS vocabularies by the development of an additional module. More specifically a semantic repository 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 specification. 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).

---

[14] The use of the "and" button in Figure 13 can be a bit confusing. The "AND" button refers to the condition set on the conditional mapping. The reader is reffered to section 10.3.5.6 for a better understanding of conditional mappings.

6.3 MINT OAI-PMH Server

A core characteristic of the MINT ingestion platform is that of being agnostic with respect to the schema of an imported dataset, a characteristic that is also inherited by the OAI-PMH service. For this to be achieved the data layer of the platform has to be able to handle heterogeneous metadata schemata. Based on this assumption, the metadata layer is implemented using a NoSQL solution that does not enforce any particular schema, thus being able to adapt to varying metadata models. The NoSQL solution that is used for the MINT OAI-PMH Platform is the MongoDB[15] document database. MongoDB is designed around the concept of documents that are internally implemented as JSON document and internally stored using the BSON format, which stands for a binary serialization of the JSON model. It allows the existence of JSON documents in the same database or even collection having different fields and thus it does not enforce any specific data model schema. Finally it provides a rich set of native implementations of drivers for communicating with the database while the JSON format provides added value in the development of web application because the stored data do not have to be transformed to a different format in order to be consumed by the applications.

The MINT's OAI-PMH repository is designed around three distinct collections that exist inside a MongoDB database. Collections can be perceived as tables of typical SQL databases, although it is not required that each document conforms to a specific data model and set of fields. These collections are the following:

- Registry: in this collection the actual metadata records are stored and accessed by the implemented OAI-PMH verbs.
- Conflicts: every time a new dataset is imported in the OAI-PMH repository, it is checked for the existence of duplicate records, and if any exist, they are reported and stored in a different collection while they are also associated with a specific Report document.
- Reports: every time an operation occurs on the OAI-PMH Repository, a report is created which includes any useful information regarding the operation. For example, if any conflict is identified between the items it is logged and reported for further reference.

The Registry collection constitutes the core collection of documents for the OAI-PMH repository. It contains all the records the user wishes to be exposed via the OAI-PMH repositories. Each record is stored inside a JSON document, which also contains other information that might be useful to the platform. More specifically, the record document contains information regarding the organization to which the item belongs to, a unique hash key that is generated by calculating the SHA1 hash of the string representation of the item, a date stamp which represents the date and time the record was inserted, and finally a namespace "prefix" value which is required by the OAI-PMH specification. An example of a Registry Document instance is depicted in Figure 15.

Name	Value	Type
▽ _id	4d8653887180685a0b05d010	ObjectId
SetSpec	Bibliotheksservice–Zentrum_Baden–Wurttemberg	String
_id	4d8653887180685a0b05d010	ObjectId
▷ datestamp		Object
id	bd69a35674d5e923823956548ad0a1116f7b1114	String
prefix	ese	String
value		String

Figure 15: Structure of the registry document

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

- Add: this type represents an action in which records are added in the registry.
- Update: this type represents an update action in which a set for a specific import already exists and it is updated by adding new metadata records.
- Delete: deletion of records from a specific import that already exists in the registry.

---

15   http://www.mongodb.org/

Apart from the action type, a number of other values are also stored as part of the Report Document. More specifically, a set of valuable statistics are stored; the number of conflicted items that were identified, the total number of the inserted records and the total number of items which corresponds to sum of the inserted records plus the conflicts. Two date stamps are also stored as part of the Report document, one corresponding to the time of creation of the document and the other to the time of closing of the document, in this way it is possible for the repository to calculate the time it took to import a whole data set into the database. Finally the date the import was published in the ingestion platform is stored together with the name of the organization it belongs to. A visual representation of the Report document structure is depicted in Figure 16.

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

Figure 16: Structure of the report document

The last collection of the OAI-PMH repository database contains conflict logs. The documents stored contain any metadata records that at the time of the exporting of an import from the ingestion platform to the OAI-PMH repository were found to be conflicted. These documents are quite simple in their structure. They contain an SHA1 hash of the conflicted item that was found together with the record and a reference to the Report document that it belongs to. In this way it is possible for someone to browse the actions that were made on the repository (e.g. additions, deletions and updates) and directly view the items that were found as conflicted for the cases of additions and updates. An example of a conflict document is depicted in Figure 17. It has to be noted that the whole procedure of creating unique hash codes and identifying conflicted items is an important functionality of the OAI-PMH repository platform because it provides a mechanism for creating unique ids for the metadata records and also a mechanism for identifying duplicates.

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

Figure 17: Structure of the conflict document

On top of the data layer that was described so far a set of functionalities is built and more specifically an RSS Feed based on Atom and of course the implementation of the actual OAI-PMH verbs. The implementation of the verbs is based on the customization of the OAICat[16] web application, which provides an abstract implementation of the OAI-PMH v2.0 specified verbs that can be customized in order to operate on top of different data layer technologies (e.g. flat XML files, relational databases etc.). The verbs that were implemented in order to work with the current OAI-PMH Repository implementation are the following:

- ⚒ **Identify**: this verb provides basic information regarding the running instance of the OAI-PMH v2.0 data repository such as, contact details of the admin of the repository, the base url that can be used by a harvester and, a sample of an identifier among others. For a complete list of the information provided by the Identify verb someone can refer to the OAI-PMH specification. The information served by this verb does not have to be stored in the underlying data layer but is part of the configuration files of the verbs implementation.

- ⚒ **GetRecord**: given the identifier of a record and the desired namespace prefix, this verb fetches from the MongoDB database the corresponding metadata record and delivers it as a response to the harvesting client. In the current implementation a query is executed based on the prefix,

---

16   www.oclc.org/research/activities/oaicat/default.htm

which is part of the Registry Document and the unique ID that is generated by the SHA1 hashing of the initial Metadata Record, this query corresponds to an exact match on the database.

- ⚔ **ListIdentifiers**: given a set name and a namespace prefix, this verb responds with a list of identifiers of items that correspond to these criteria. The set name is identical to the name of the organization. In this way it is possible to organize the records of the repository around organizations/providers. Again, the namespace prefix is matched with the prefix field of the Registry Document.

- ⚔ **ListRecords**: this verb operates in a similar way to the ListIdentifiers with the main difference being that instead of returning only the identifiers, the complete Metadata Record is served.

- ⚔ **ListMetadataFormats**: this verb is implemented by aggregating all the unique prefix values that are stored in the data layer by executing an aggregation for uniqueness query on the Registry collection. The resulting response that is served by this verb contains all the unique namespace prefixes that exist in the OAI-PMH repository and can be used for accessing the Metadata Records.

- ⚔ **ListSets**: this verb returns a list of all the sets that exist in the OAI-PMH repository. Sets are named after the organizations that provide metadata records to the OAI-PMH repository, in this way it is possible for someone to retrieve only the records that are associated with a specific organization. The way the values are extracted is similar to the ListMetadataFormats verb.

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

An RSS feed is implemented following the "Atom Syndication Format" which is an XML language used for web feeds, while the "Atom Publishing Protocol" (APP) is a simple HTTP based protocol for creating and updating web resources. The purpose of this RSS feed in the current OAI-PMH repository implementation is to provide a mechanism for notifying metadata consumers for the occurrence of specific actions, for example when new items are added or updated to the repository in an automatic way. This is achieved by creating the RSS Feed on top of the Reports collections that was described earlier, in this way every time a new report is generated the feed is automatically updated and the subscribers are informed for the associated action. The implementation of the RSS Feed service is based on the Apache Abdera[17] project, a functionally complete, high performance implementation of the IETF Atom Syndication Format (RFC 4287) and Atom Publishing Protocol (RFC 5023) specifications. The current implementation of the RSS Feed communicates directly with the MongoDB based data layer of the OAI-PMH repository, every time a new Report document is inserted, it is also transformed into the Atom protocol XML representation and published on the Atom Feed that is maintained through the API of Apache Abdera. The RSS Feed can be served via any of the available servlet containers, for performance reasons it was decided to use a Jetty servlet container in the current implementation.

6.4 Metadata progress reports

In order to improve the monitoring of the metadata production and publication to Europeana a new module has been implemented that allows content providers to download reports with their overall progress. In detail short-term metadata production targets will be set – in coordination with WP2 – for each content provider of AthenaPlus. After that the content provider will be able to download a progress report with the following tables.

The Transformations table shows the imports that have been transformed and the resulting valid Items according to LIDO.

---

17   http://abdera.apache.org/

Transformations

Transformation Name	Date Modified	Parent Import	Mapping Used	Valid	Invalid
LIDO 1.0 Transformation	10/26/12 12:47 AM	EuPhoto 2.zip	EuPhoto	5	0
LIDO 1.0 Transformation	7/21/13 11:51 PM	EuPhoto-Extended.tgz	MINT Demo - v2	5	0
LIDO 1.0 Transformation	7/23/13 1:42 PM	EuPhoto-Extended2.tgz	Peccioli Sofie	5	0
LIDO 1.0 Transformation	7/21/13 10:56 PM	EuPhoto.zip	MINT Demo	5	0
			* Estimated items ready for publication :	20	

* Number of items ready for publication is estimated roughly by the number of valid transformed items without checking for duplicates .

Figure 18: MINT progress report (Transformations)

The publications table presents the datasets that have been published through MINT,

Publications

Publication	User	Date Created	Date Modified	Items
EuPhoto.zip	Nikos Simou	10/26/12 6:11 PM	7/21/13 11:49 PM	5
EuPhoto-Extended2.tgz	Photography Admin	7/23/13 11:17 AM	7/23/13 1:49 PM	5
EuPhoto-Extended.tgz	Photography Admin	7/21/13 11:51 PM	7/21/13 11:55 PM	5

Figure 19: MINT progress report (Publications)

The OAI current status table shows that number of unique records that are on NTUA's OAI-PMH server and are ready to be harvested by Europeana (in this case the same records have been published three times resulting to 5 unique records in OAI-PMH)

Oai Current Status

Namespace	Unique Items Commited
rdf	5

Figure 20: MINT progress report (OAI-PMH)

Finally, the progress status table shows the total number of transformed valid items, published valid items, OAI published items and also the current and the overall target assigned to content providers.

Progress status

Status / Target	Date	Items
Transformed Valid Items	7/26/13 1:48 PM	20
Published Valid Items	7/26/13 1:48 PM	15
OAI Published Items	7/26/13 1:48 PM	5
Current Target Number of Items	11/30/13 12:00	30
Final Target Number of Items	11/30/14 12:00	90
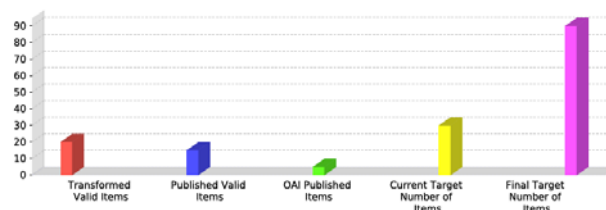
Figure 21: MINT progress report (Progress status)

6.5 EDM support

The main objective of the AthenaPlus project is to deliver quality metadata to Europeana. For this purpose the support of the Europeana Data Model, that is the harvesting schema of Europeana, is of major importance. Therefore for fully supporting EDM an XSLT has been implemented for transforming LIDO records to EDM. The current version is fully functional for the AthenaPlus core cataloguing elements (i.e. bookmarks) but it is still under development and it will be documented in deliverable D 3.2 "Description of the LIDO to EDM mapping".

Figure 22: Europeana preview on MINT

In addition to this crosswalk the Europeana preview on MINT has been updated based on Europeana's new portal. Again an XSLT has been implemented – according to Europeana's specifications – for transforming EDM metadata to an html very similar to the one shown on Europeana (see Figure 22).

It is very important to mention at this point that the reasons why the Europeana preview on MINT is not exactly the same as the one shown at Europeana are:

- Europeana is still working on their portal
- Europeana makes a kind of enrichment on the harvested metadata producing auto-generated tags (Figure 23).

Figure 23: Europeana's preview with enrichments (link)

7 DEVELOPMENTS AND OPERATIONS FOR ATHENA PLUS

This section presents the developments and the operations performed by the NTUA team for the AthenaPlus project. It consists of three subsections, the first one focuses on the setup of the MINT2 instance for the AthenaPlus project, the second one presents the technical support provided to the users of the tool while the third one outlines the new developments that have been specifically performed for the AthenaPlus MINT instance together with the main bugs resolved.

7.1 MINT2 Setup for AthenaPlus

For setting up the AthenaPlus MINT instance, the NTUA team has closely examined the needs of the project in terms of users and connections and proceeded with the following actions:

1. A Postgres 9.2 database has been setup on an NTUA server with the following specs
   a. **CPU**: Intel(R) Core(TM) i7 CPU 960 @ 3.20GHz 8 cores,
   b. **RAM**: 12G RAM,
   c. **Storage**: 2x1TB SATA3 7.2K RPM in RAID1, 4x1TB SATA3 7.2K RPM in RAID10.
2. An Apache Tomcat (6.037) web container has been setup on an NTUA server with the following specs.
   a. **CPU**: Intel(R) Xeon(R) CPU E5640 @ 2.67GHz (8 cores)
   b. **RAM**: 16G
   c. **Storage**: 300G Virtual Disk on RAID5 SAS 6GBPS 15K RPM DISKS
   It is important to mention at this point that it has been decided to use different servers for the database and the web application for achieving resource utilization and the best possible performance.
3. A new CMS has been implemented for MINT based on the AthenaPlus theme.

The AthenaPlus project focuses on the museum cultural community and for that reason it uses LIDO i.e. a metadata model widely used by this community - as an intermediate schema towards its way to Europeana. The LIDO schema has been set up as the target schema considering the project requirements. In detail the following customizations have been performed.

1. Some of the LIDO elements and attributes are automatically filled in. In detail

   a. **lido:recordID is set to /AthenaPlus:000000:** This is a fixed value i.e. it cannot change by the provider. This decision was made for having a way – if needed in the future – for automatically assigning identifiers to records
   b. **attribute xml:lang of lido:administrativeMetadata and lido:descriptiveMetadata have been set mandatory:** In addition these attributes are controlled by a SKOS vocabulary that includes the ISO languages codes. In that way normalized values for languages and better quality metadata are ensured.
   c. **lido:descriptiveMetadata/lido:objectClassificationWrap/lido:classificationWrap/lido :classification has been triplicated:** The first instance of it does not have any lido:type attribute and it is intended for general use. The lido:type attribute of the second one is set to europeana:type and it can only take values IMAGE, VIDEO, TEXT, 3D, AUDIO (edm:type) while the lido:type attribute of the third one is set to Europeana:project and the lido:term values is fixed to "Athena Plus"
   d. **lido:recordInfoSet/lido:recordInfoLink has been set mandatory:** This has been done because this element is used to hold the link to provider's metadata (edm:isShownAt).
   e. **lido:administrativeMetadata/lido:recordWrap/lido:recordSource/lido:legalBodyNam e has been set madatory:** This has been done because this element is used to hold the provider's name (edm:provider)
   f. **lido:administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepr esentation has been triplicated:** The first instance of it does not have any lido:type attribute and it is intended for general use. The lido:type attribute of the second one is set to image_master and it is used to hold the link to the high resolution image (edm:isShonwBy) while the lido:type attribute of the third is set to image_thumb and it is used to hold the link to the low resolution image (edm:object)

      g. **lido:administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:rightsResour
        celido:rightsType has enumerated values according to the Europeana rights
        statements**

      h. **lido:descriptiveMetadata/lido:eventWrap/lido:eventSet has been duplicated:** A
        production event is made by default in the mapping. The user can make new events of
        which type is enumerated according to LIDO event specification.

2. The following table presents the bookmarks that have been setup in the MINT mapping editor
for assisting providers during the mapping process. These bookmarks have been selected
according to the project requirements for metadata quality.

LIDO Bookmark	LIDO Xpaths
Identifier	/lido/administrativeMetadata/recordWrap/recordID
Descriptive metadata language	/lido/descriptiveMetadata/@ lang
Administrative metadata language	/lido/administrativeMetadata/@ lang
Europeana type	/lido/descriptiveMetadata/objectClassificationWrap/ classificationWrap/classification@lido:type='europeana:type'/term
Object/Work Type	/lido/descriptiveMetadata/objectClassificationWrap/ objectWorkTypeWrap/lido:objectWorkType/term
Object Title/Name	/lido/descriptiveMetadata/objectIdentificationWrap/titleWrap/ titleSet/appellationValue@lido:pref='preferred'
	/lido/descriptiveMetadata/objectIdentificationWrap/titleWrap/ titleSet/appellationValue@lido:pref='alternative'
Object Description	/lido/descriptiveMetadata/objectIdentificationWrap/objectDescriptionWrap/ objectDescriptionSet/descriptiveNoteValue
Dimension	/lido/descriptiveMetadata/objectIdentificationWrap/objectMeasurementsWrap/ objectMeasurementsSet/displayObjectMeasurements
Producer	/lido/descriptiveMetadata/eventWrap/eventSet/event/ eventActor/actorInRole/actor/nameActorSet/appellationValue & /lido/descriptiveMetadata/eventWrap/eventSet/event/eventType/conceptID = http://terminology.lido-schema.org/lido00007
Producer's role	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventActor/

	actorInRole/roleActor & /lido/descriptiveMetadata/eventWrap/eventSet/event/eventType/conceptID = http://terminology.lido-schema.org/lido00007
Production Date	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventDate/displayDate & /lido/descriptiveMetadata/eventWrap/eventSet/event/eventType/conceptID = http://terminology.lido-schema.org/lido00007
Production Place Name	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventPlace/displayPlace & /lido/descriptiveMetadata/eventWrap/eventSet/event/eventType/conceptID = http://terminology.lido-schema.org/lido00007
Material (Production Event)	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventMaterialsTech/ materialsTech/termMaterialsTech@lido:type = 'material'/conceptID & /lido/descriptiveMetadata/eventWrap/eventSet/event/eventType/conceptID = http://terminology.lido-schema.org/lido00007
Technique (Production Event)	/lido/descriptiveMetadata/eventWrap/eventSet/event/eventMaterialsTech/ materialsTech/termMaterialsTech@lido:type = 'technique'/term & /lido/descriptiveMetadata/eventWrap/eventSet/event/eventType/conceptID = http://terminology.lido-schema.org/lido00007
Subject / Theme (Concept)	/lido/descriptiveMetadata/objectRelationWrap/subjectWrap/subjectSet/ subject/subjectConcept/term
Repository	/lido/descriptiveMetadata/objectIdentificationWrap/repositoryWrap/ repositorySet@lido:type ='current'/repositoryName/ legalBodyName/appellationValue
Rights Information for Work	/lido/administrativeMetadata/rightsWorkWrap/rightsWorkSet
Record Type	/lido/administrativeMetadata/recordWrap/recordType/term
Record Source (Name)	/lido/administrativeMetadata/recordWrap/recordSource@lido:type='europeana:dataProvi

	der'/
	legalBodyName/appellationValue
Record Link (Link to metadata)	/lido/administrativeMetadata/recordWrap/recordInfoSet/recordInfoLink
Link to DCHO (Master)	/lido/administrativeMetadata/resourceWrap/resourceSet/
	resourceRepresentation@lido:type='image_master'/linkResource[1] (the first occurrence)
	/lido/administrativeMetadata/resourceWrap/resourceSet/
	resourceRepresentation@lido:type='image_master'/linkResource[>1] (the remaining occurrences)
Link to DCHO (Thumbnail)	/lido/administrativeMetadata/resourceWrap/resourceSet/
	resourceRepresentation@lido:type='image_thumb/linkResource
Resource Rights Type	/lido/administrativeMetadata/resourceWrap/resourceSet/
	rightsResource/rightsType/term@lido:pref='preferred'

3. A crosswalk that is used for the automatic transformation of valid LIDO records to valid and semantically aligned ESE records has been set up. This crosswalk is the XSL file that has been developed in the Linked Heritage project.

4. A crosswalk that is used for the automatic transformation of valid LIDO records to valid and semantically aligned EDM records has been set up. This crosswalk is currently an XSL file that the main outcome of Task 3.1 and is described in detail in D 3.2. Its final version will be described in D 3.3.

5. The following SKOS vocabularies have been integrated into MINT

Bookmark Name	SKOS Vocabulary
Object/Work Type (PP vocabulary)	Partage Plus Objects
Producer's role	LIDO Actor's roles
Material (Production Event - PP Vocab)	Partage Plus Objects
Technique (Production Event - PP Vocab)	Partage Plus Techniques
Subject / Theme (Concept) (EuPhoto vocabulary)	EuropeanaPhotography keywords

7.2 Technical Support

The transformation and the publication of providers' metadata to LIDO is a process during which providers may encounter difficulties, even if they are well aware of the MINT 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 functionalities to reach the desired result.

For that reason a help desk support system has been set up that assists the Content Providers during all the phases of the AthenaPlus project. More specifically the Content Providers are able to get

instructions on how to map their in-house metadata to LIDO for exploiting its full expressiveness, as well as technical information about MINT functionalities that will permit them to fit their metadata perfectly according to the AthenaPlus requirements. In addition, a set of explanatory documents concerning the various aspects of importing metadata to MINT, mapping to LIDO and publishing metadata to Europeana will be carried out within the task 3.2. The documents that will provide guidelines to Content Providers together with the on-going support during the project will ensure seamless transformation and delivery processes. This task is coordinated by ICCU (see also Task 2.4)

7.3 MINT Developments for AthenaPlus

In section 6 the main features of MINT 2 have been presented. The implementation of these features came though provider's requests in the various aggregation projects MINT has been used. Specifically for the AthenaPlus the following features have been implemented

- **MINT UI tablet/smartphone compatible**: Some of the AthenaPlus providers have requested the fully functional use of MINT 2 through tablets and smartphones. The main problem MINT had when accessed through mobile devices was the drag and drop operations on which most of its functionally is based. This has been resolved by the usage of appropriate UI libraries in MINT that permit drag and drop operations through tablets and smartphones.

- **Search of items**: A problem content providers had encountered was the search of a specific item among their imports. This was important for them since they wanted to locate and preview - using the EDM and Europeana previews - their masterpieces. For resolving this issue the MINT team has setup a SOLR indexing that runs through all imports and also through the derived transformations of an import creating indexes on their free text (i.e. LIDO, ESE and EDM transformation). This permits MINT users to search (see Figure 24) using

    o free text to match the labels
    o free text to match an except of the record
    o a solr query
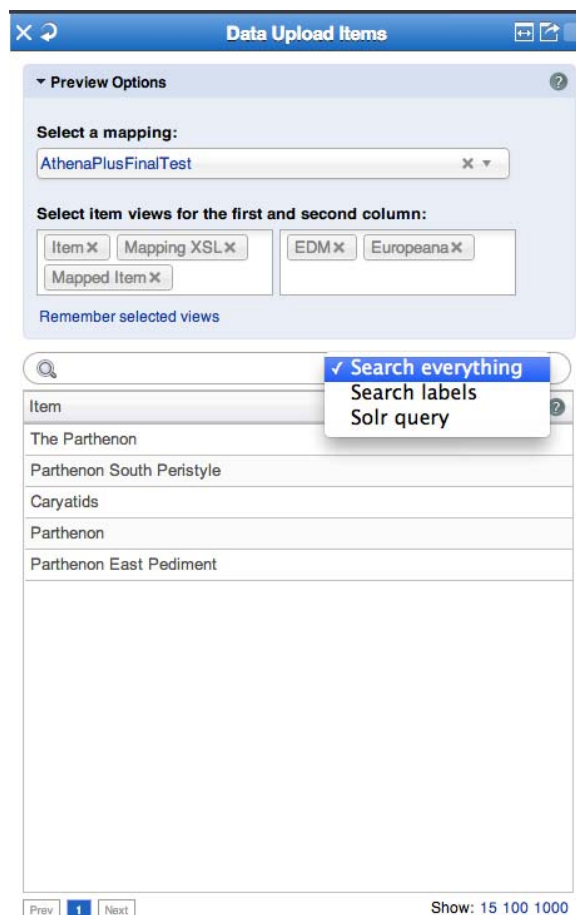
Figure 24: Search functionality in items preview

28

- **OAI new strategy for EDM:** Section 6.3 presents the OAI-PMH repository and the data structures it uses. A custom insertion strategy has been implemented for the AthenaPlus project that is used for the insertion of EDM records to the OAI-PMH repository. In detail this strategy processes the EDM record extracting the dc and dcterms elements out of it for creating the OAI-DC record. In addition this strategy performs a different duplication detection check – based on the way that Europeana checks for duplicates – than the one used in previous MINT versions. In the strategy used in previous versions which was based on ESE records an SHA1 hash was created from the whole document. Before the insertion of a record into the OAI repository a check was performed and if this hash was not found then the record was unique and it was inserted into the repository. The current approach is using the @rdf:about of the edm:ProvidedCHO, since EDM is an RDF based metadata model and as such the class identifiers it uses have to be unique.

- **Structural on any element/ if-else mapping:** The mapping editor in MINT 2 has been totally reconstructed using a new JSON library (JsonSmart) that is much faster than the one used in previous versions. In addition the functionality described in sections 6.2.2 and 6.3.3 has been implemented for the AthenaPlus project.

- **Progress reports integrated into MINT interface:** Section 6.4 describes the reporting functionality of the MINT mapping tool. The AthenaPlus providers have requested this functionality integrated into the MINT user interface. For that purpose the Data Statistics button has been added under the Home panel, and by using it providers can directly preview their import, transformation and publication history as well as the overall project statistics (Figure 25-26).
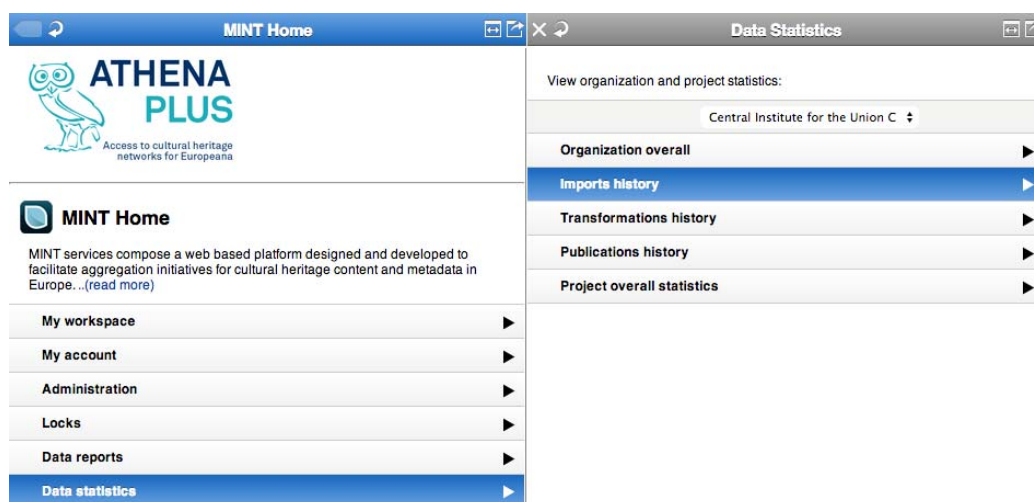
Figure 25: MINT reporting functionality

Central Institute for the Union Catalogue of the Italian Libraries

Transformed items per schema
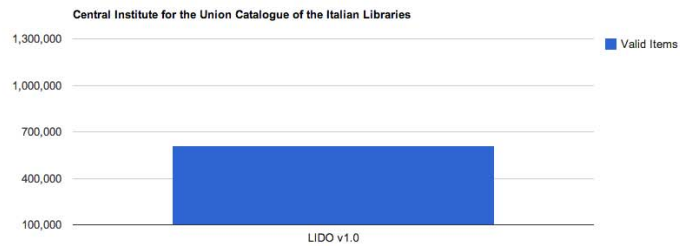
Transformations

Parent Dataset Name	Date	Total Items	Valid Items	Invalid Items	Target Schema	Mapping
http://www.culturaitalia.it/oaiProviderCI/OAIHandler - amshistorica	Apr 14, 2014, 5:03:38 PM	181	117	64	LIDO v1.0	historica
http://www.culturaitalia.it/oaiProviderCI/OAIHandler - biblio_braidense_milano	Jan 23, 2014, 3:36:19 PM	265713	265713	0	LIDO v1.0	internet_culturale_04
http://www.culturaitalia.it/oaiProviderCI/OAIHandler - biblio_centrale_firenze	Jan 22, 2014, 11:48:57 PM	134757	134624	133	LIDO v1.0	internet_culturale_03
http://www.culturaitalia.it/oaiProviderCI/OAIHandler - internet_culturale_audiovisivi_roma	Jan 22, 2014, 7:52:10 PM	129483	129483	0	LIDO v1.0	internet_culturale_suoni
http://www.culturaitalia.it/oaiProviderCI/OAIHandler - casa_musica_cirpem_parma	Jan 22, 2014, 3:06:36 PM	24256	24256	0	LIDO v1.0	internet_culturale_manoscritti_01
http://www.culturaitalia.it/oaiProviderCI/OAIHandler - biblio_nazionale_marciana	Jan 22, 2014, 2:56:17 PM	1149	1149	0	LIDO v1.0	internet_culturale_02
http://www.culturaitalia.it/oaiProviderCI/OAIHandler - opere_museid	Jan 22, 2014, 12:24:20 PM	66549	56342	10207	LIDO v1.0	musei_d_italia

Figure 26: ICCU's transformation history

It is also important to mention at this point that during the development of these features various bugs have been reported by the providers. The NTUA team uses a project management web application (Redmin) to track them and resolve them. A short but characteristic list of them is shown below

- EDM direct import was not working.
- Itemization was of imports was not working.
- Status of mapped elements was not working.
- SKOS search was case sensitive.
- Import of CSV documents was not working.
- Status of mapped elements on the input schema tree was not working.
- Replace function was not working properly.
- Value mapping allowed the mapping of the same input to different output resulting in loss of items in transformation.

8 CONCLUSION

The present document constitutes the report of deliverable D 3.1 "The MINT Ingestion Platform", that is made available online for validation and for the large-scale contribution of content to Europeana and for dissemination & training. The MINT ingestion platform 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 AthenaPlus 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 ingestion platform (Task 3.1), that is available online to the project partners at http://mint-projects.image.ntua.gr/athenaplus

8.2 Impact

The delivery of the platform and the present report achieve a project milestone, MS7 MINT Ingestion platform available, Month 6. This constitutes the starting point for mapping and delivery of metadata records to Europeana.

9   APPENDIX 1: DEFINITION 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
JSON   JavaScript Object Notation

10 APPENDIX 2: USER MANUAL

This section presents the MINT ingestion platform 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 ingestion platform that has been implemented within the AthenaPlus 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.

10.1 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 ingestion platform. The managers or the technical experts of the organizations that participate in the AthenaPlus are suggested to follow the next steps:

- User registration (see Section **Errore. L'origine riferimento non è stata trovata.**) – to register themselves.
- Organization Registration & Management (see Section **Errore. L'origine riferimento non è stata trovata.**) – to register their organization(s).
- User management (& User creation under an organization by its administrator) (see Section **Errore. L'origine riferimento non è stata trovata.**) – 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 AthenaPlus are suggested to register themselves under an existing organization (see Section **Errore. L'origine riferimento non è stata trovata.**). 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.
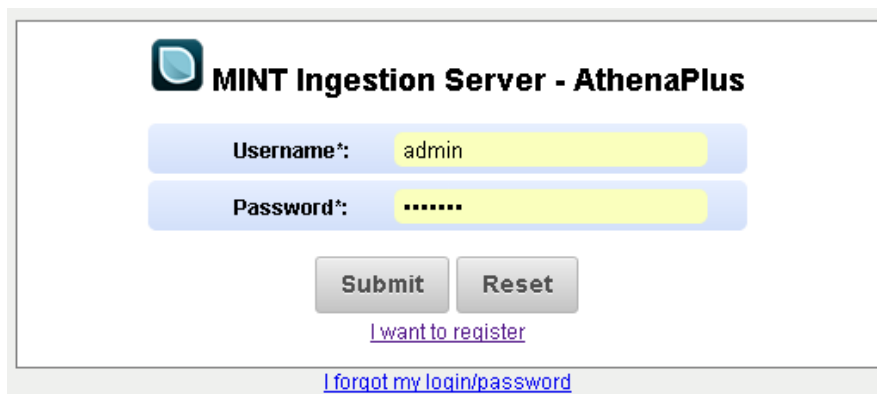
10.1.1 *User Registration*

User registration can be done in two ways:

- *Self-registration from MINT Ingestion platform* – The user registers himself/herself directly (see Section **Errore. L'origine riferimento non è stata trovata.**).
- *User creation under an organization by its administrator* – The organization's administrator creates an account for a user (see Section **Errore. L'origine riferimento non è stata trovata.**).

10.1.1.1 Self-registration from MINT Ingestion platform
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/athenaplus/
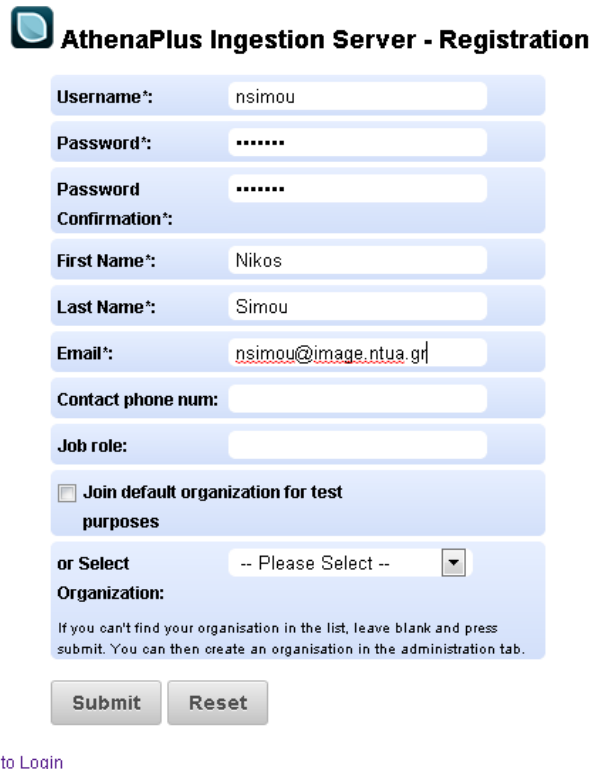
Figure 27: MINT Ingestion Server

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

Figure 28: 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 AthenaPlus project and they just want to check the functionality of the MINT ingestion platform.
- *Select one of the existing organizations* – This option is recommended to users that work for an organization that participates in the AthenaPlus 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 AthenaPlus 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.

10.1.1.1.1 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 ingestion platform- 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 **Errore. L'origine riferimento non è stata trovata.**).
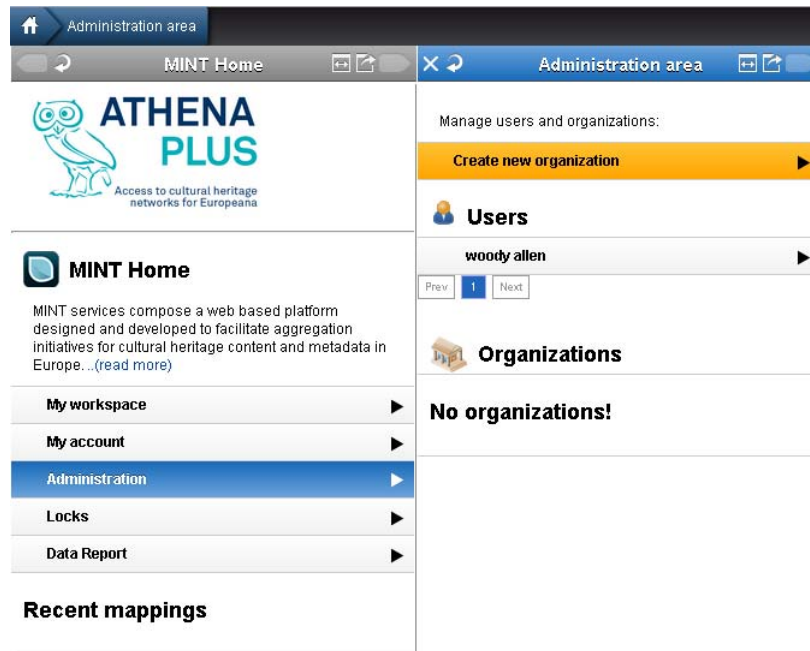
Figure 29: Create organization

10.1.1.1.2 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 ingestion platform - 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.

10.1.1.2 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 ingestion platform 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.

Figure 30: 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.

10.1.2 *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 **Errore. L'origine riferimento non è stata trovata.**) or to be the administrator of the organization you want to edit. Log in to MINT ingestion platform 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 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 **Errore. L'origine riferimento non è stata trovata.**) 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.

Figure 31: 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.

Figure 32: Edit organization

10.2 Import of Metadata

To import your metadata to the MINT ingestion platform, 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 **Errore. L'origine riferimento non è stata trovata.**) 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 33: 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.[18]

Figure 34: CSV upload

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

---

[18] 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.

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.

10.2.1 *Dataset Options*

After importing a file to the MINT ingestion platform 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.
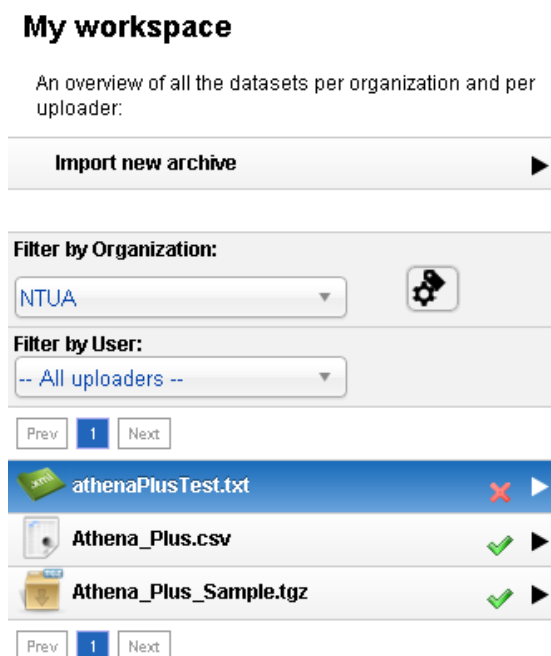
Figure 35: Workspace with successful and unsuccessful imports

10.2.1.1 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.
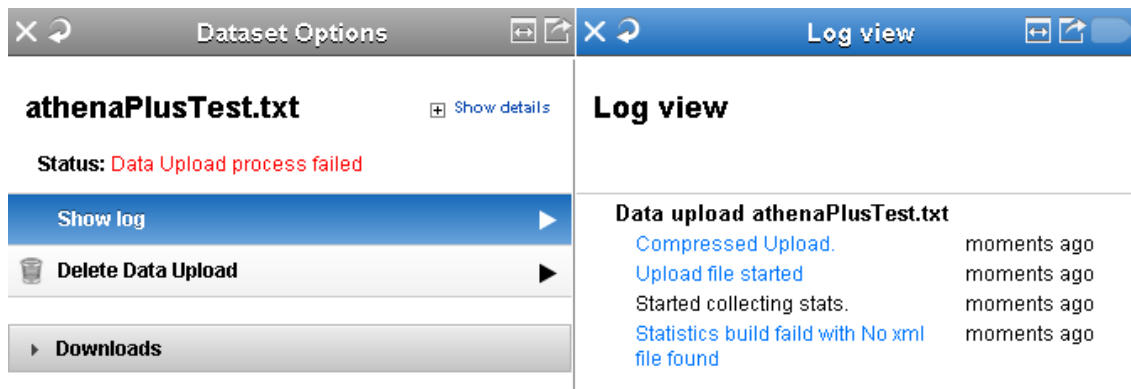
Figure 36: Error in import

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.

10.2.1.2 Successful import
When a dataset is imported successfully the following options appear in the "Dataset Options" pane.
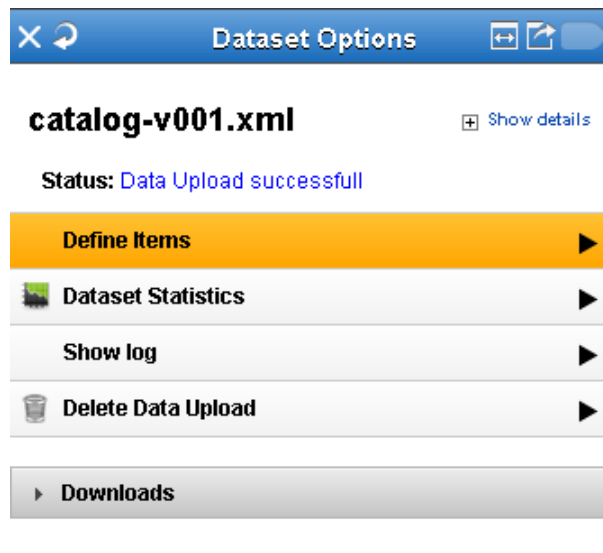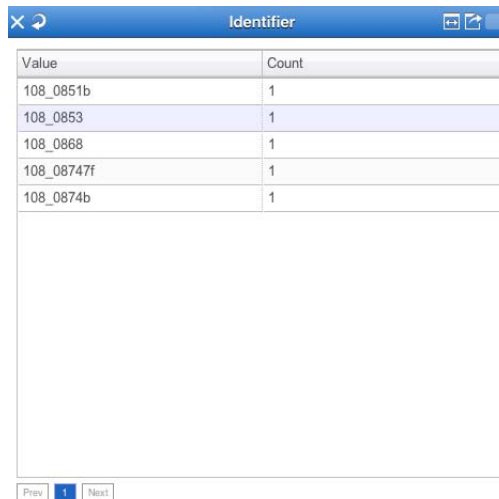
Figure 37: Successful upload dataset options

10.2.1.2.1 Define Items

By clicking on the "Define Items" the next pane appears (see Figure 36). 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 (ⓘ) on the left of each node you get information and statistics about the values of this node (see Figure 35).

Figure 38: 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.
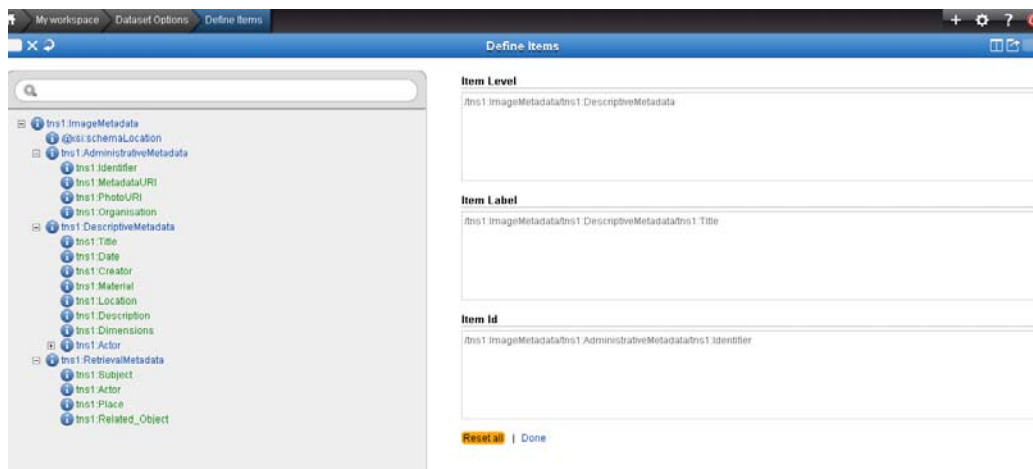
Figure 39: Define Items

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.
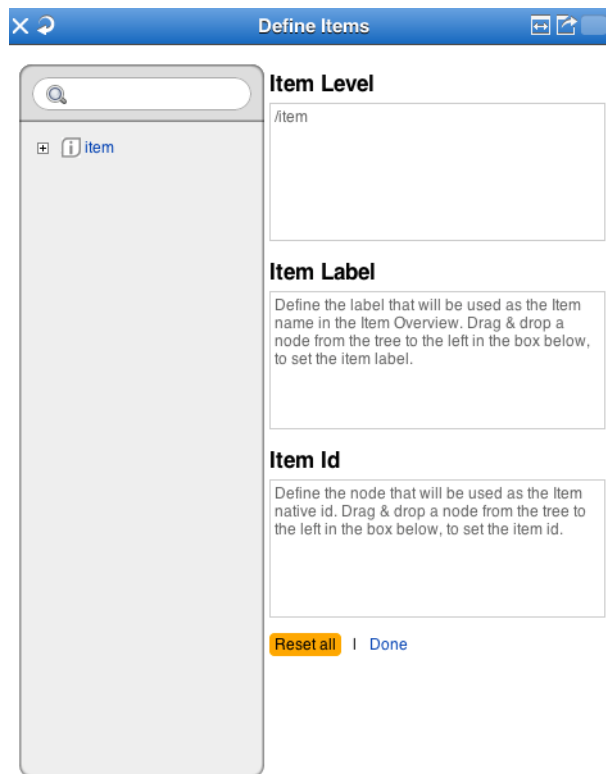
Figure 40: Define items for a CSV import

After having defined the items the following options appear in the "Dataset Options" pane.
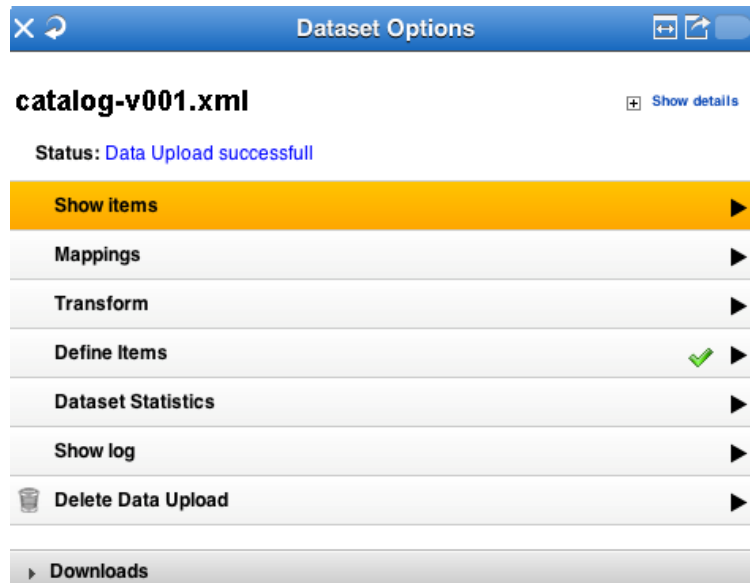
Figure 41: Dataset Options after having defined the items

10.2.1.2.2 Show Items

After having defined the items (see Section **Errore. L'origine riferimento non è stata trovata.**) 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).
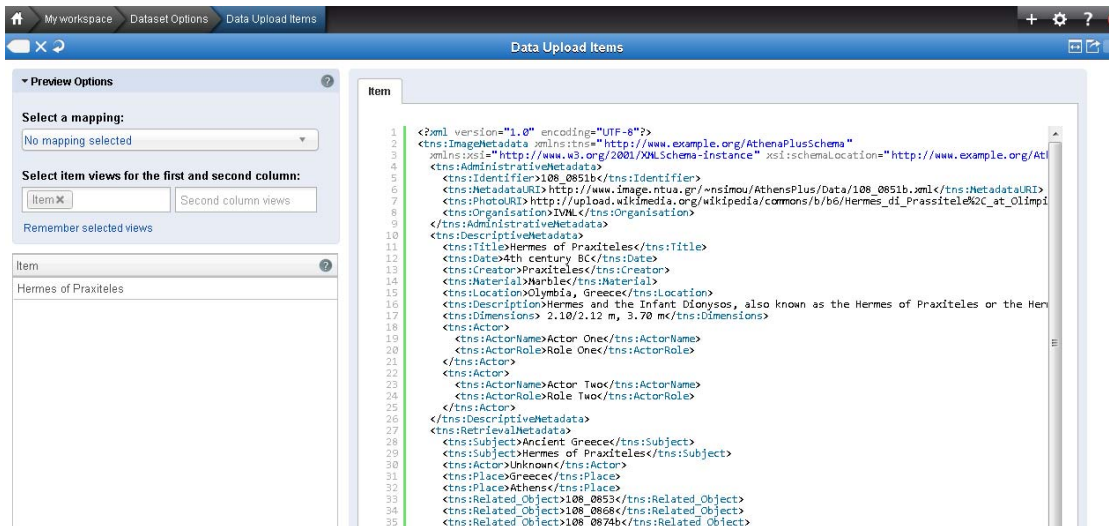
Figure 42: Show Items

If you select "Show Items" after you have implemented a mapping (see **Errore. L'origine riferimento non è stata trovata.**) then you can select it and the following options appear.
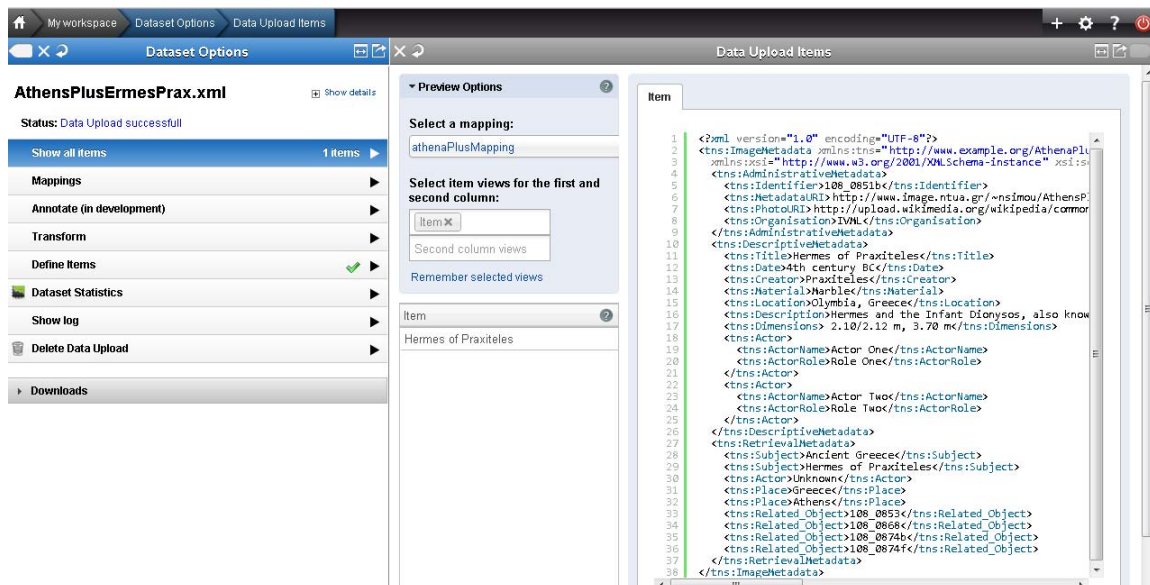
Figure 43: Show items after having implemented a mapping

- **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 item as it will appear in the Europeana portal after its publication.

10.2.1.2.3 Dataset statistics

After having defined the items (see Section **Errore. L'origine riferimento non è stata trovata.**) 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.
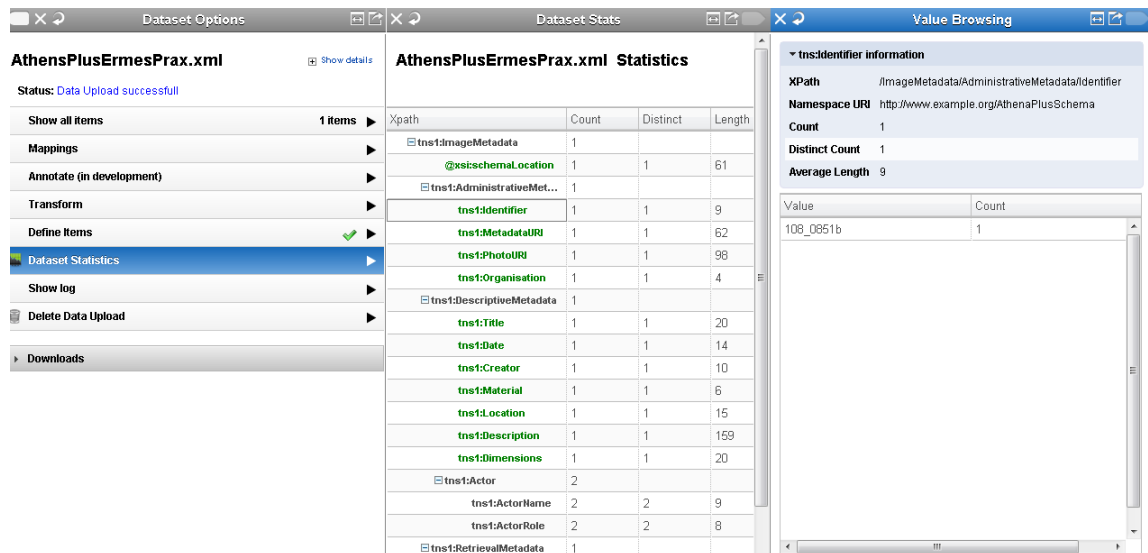
Figure 44: Dataset statistics

10.3 Mapping

After having defined the items (see Section **Errore. L'origine riferimento non è stata trovata.**) 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.
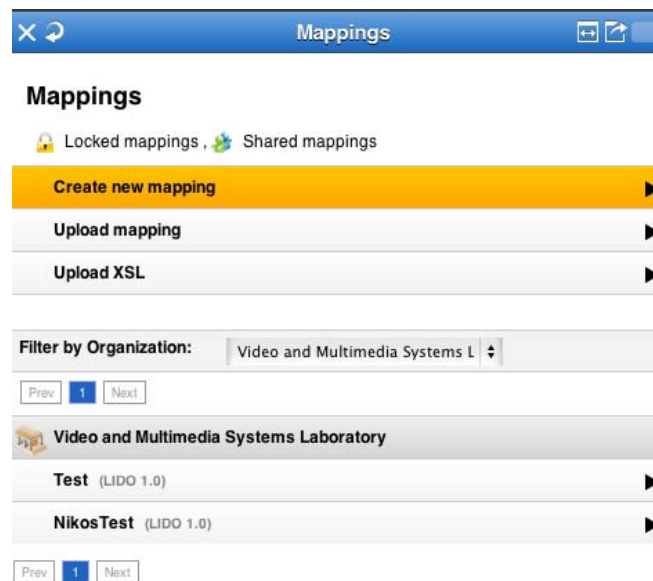
Figure 45: Mapping options

10.3.1 *Create new mapping*

By selecting the Create new mapping option the "New Mapping" pane appears that is shown below.
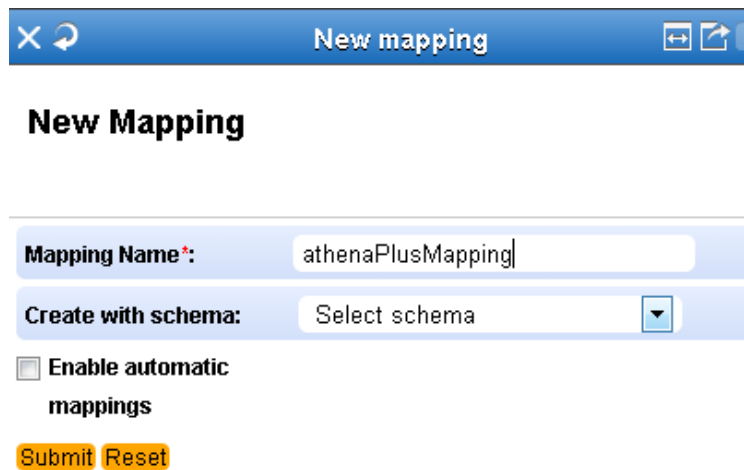
Figure 46: New mapping

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 ingestion platform 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 AthenaPlus intermediate schema) that are exactly the same. So users are suggested to use this feature only if their input metadata are already in LIDO.
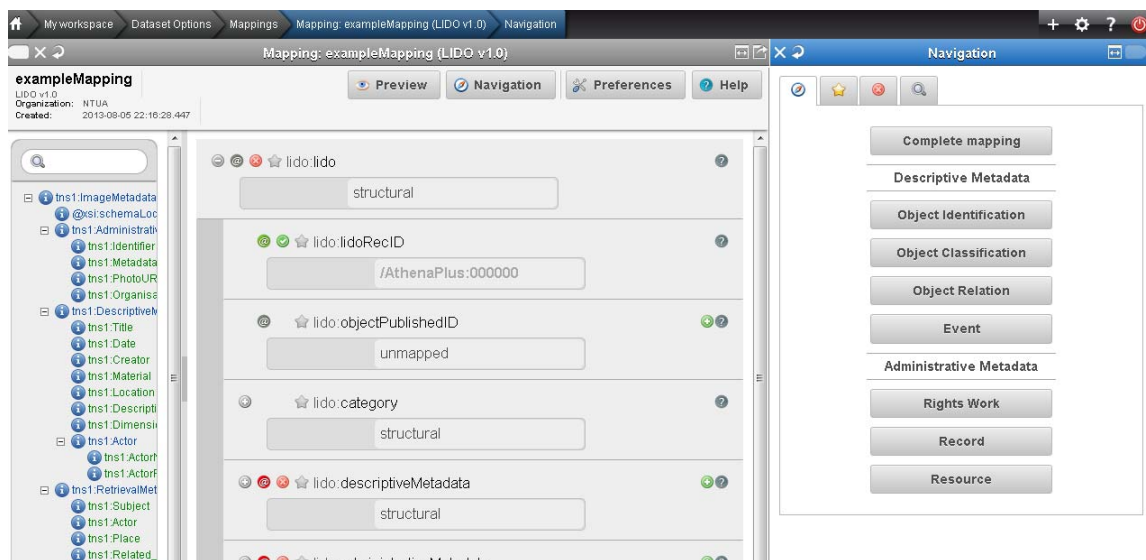
Figure 47: The ingestion platform

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.
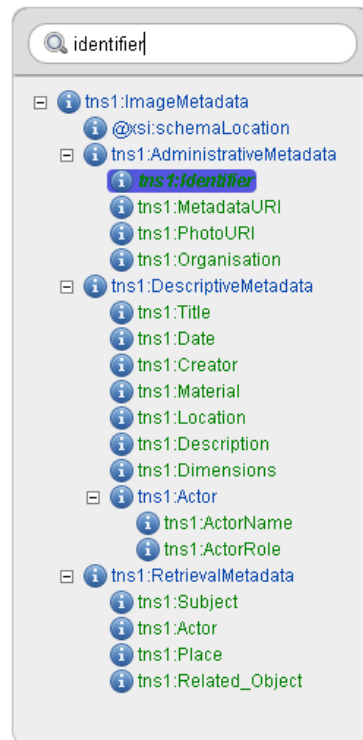
Figure 48: Input schema area

On the right hand side, buttons correspond to high-level elements of the target schema (see **Errore. L'origine riferimento non è stata trovata.**: if not visible click on 🧭 button on the top right of the ingestion platform) 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 **Errore. L'origine riferimento non è stata trovata.** for details).
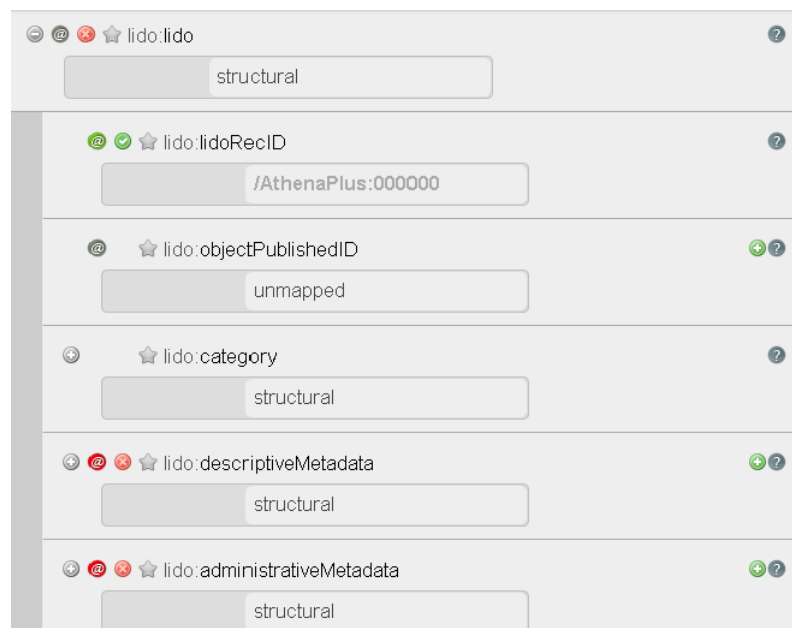
Figure 49: Mapping area
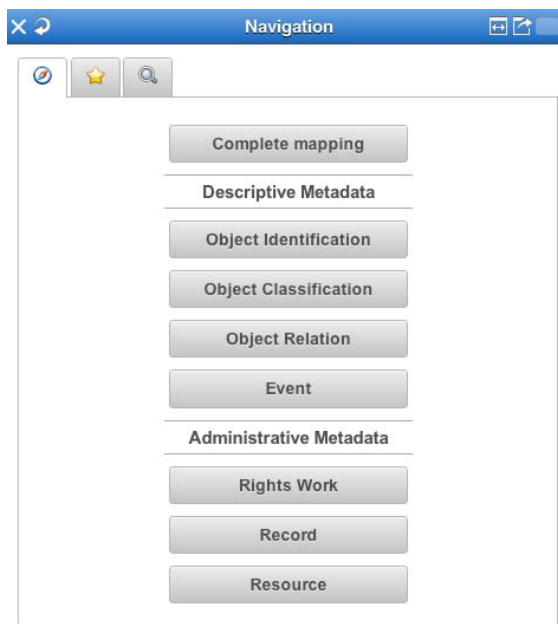
Figure 50: Navigation area - Schema navigation

For the needs of the AthenaPlus project and for assisting providers that are not familiar with LIDO, bookmarks have been created to the LIDO elements that will accommodate the AthenaPlus mandatory

fields. 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 AthenaPlus mandatory fields, to see the respective LIDO element and easily map a value from his/her metadata.
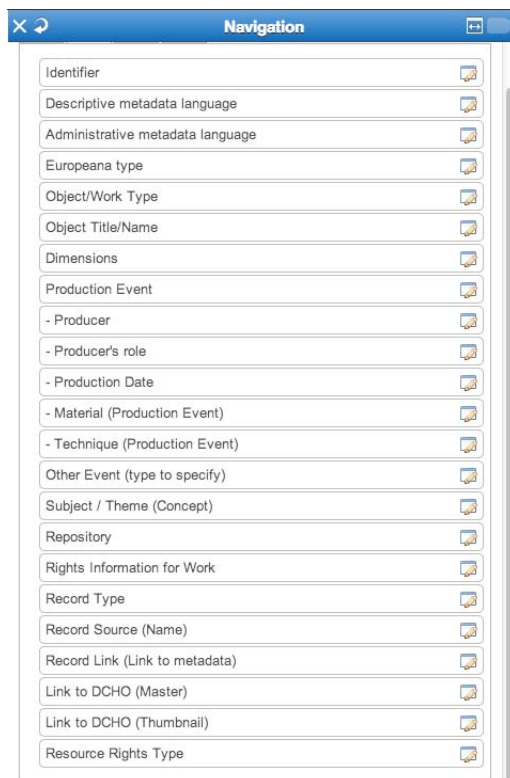
Figure 51: Navigation area - Bookmarks navigation

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.
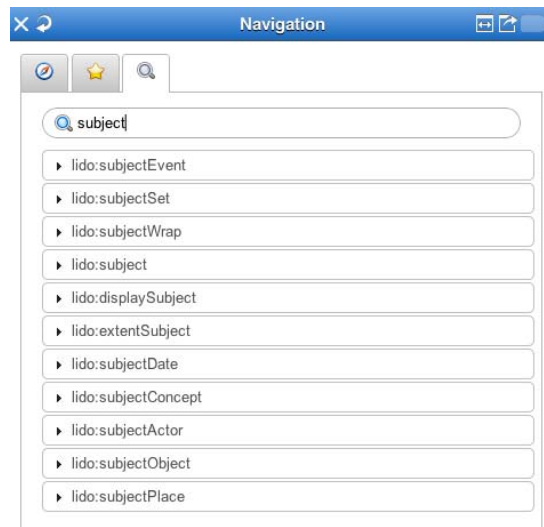
Figure 52: Navigation area - Search target schema

10.3.2 *Upload mapping*

By selecting the Upload mapping from the "Mappings" pane the following pane appears where the user can select a mapping to upload.
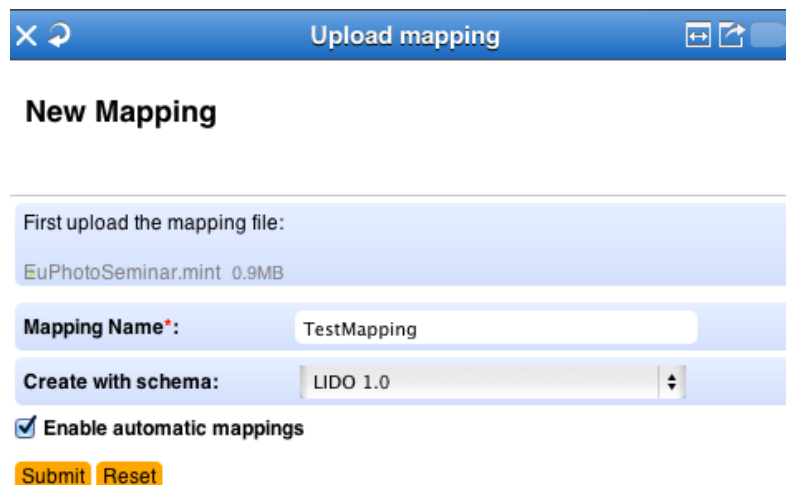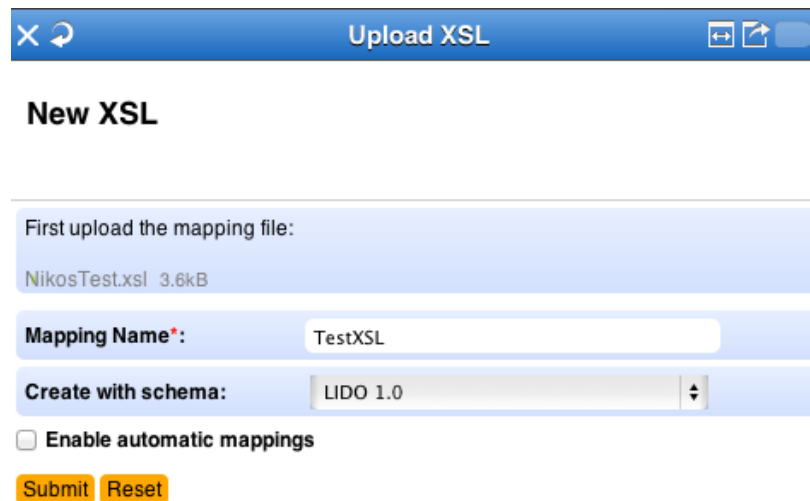
Figure 53: Upload mapping

After pressing the submit button the mapping with the name you specified appears in the "Mappings" pane and the ingestion platform opens.

10.3.3 *Upload XSL*

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

Figure 54: 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.

Figure 55: The XSL editor

10.3.4 *Mappings Management*

By selecting a mapping from the Mapping pane the following pane appears that allows you to:
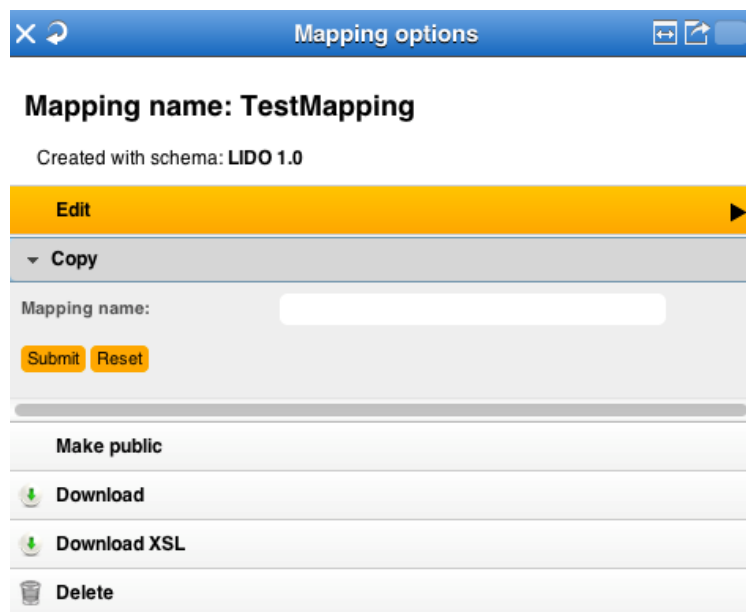
Figure 56: Mappings management

- **Edit a mapping** – This action opens the ingestion platform 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 ![icon](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.

10.3.5 *Mappings*

10.3.5.1 Mapping environment

As mentioned in Section **Errore. L'origine riferimento non è stata trovata.** 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 57: 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: Ingestion platform informative icons

After mapping an Xpath from the input schema the mapping area changes as shown below and the XPath selected from the input schema is highlighted in bold.
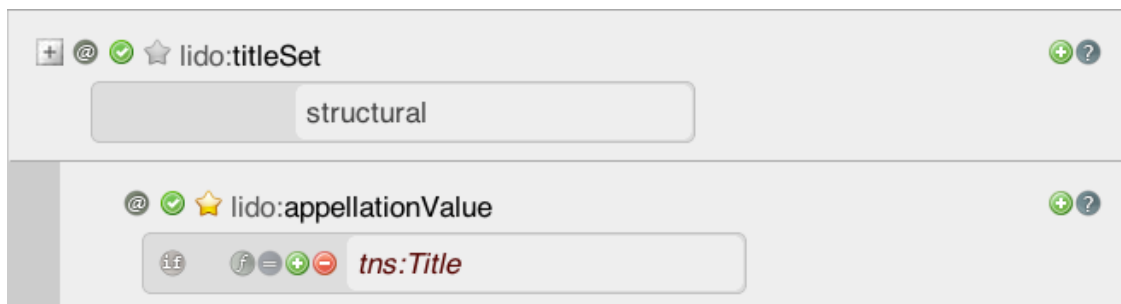
Figure 58: An element that has been mapped in the mapping area

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

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

Icon	Description
	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: Ingestion platform functional icons

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

Figure 59: Shortcuts to input tree and values from mapping

10.3.5.2 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.
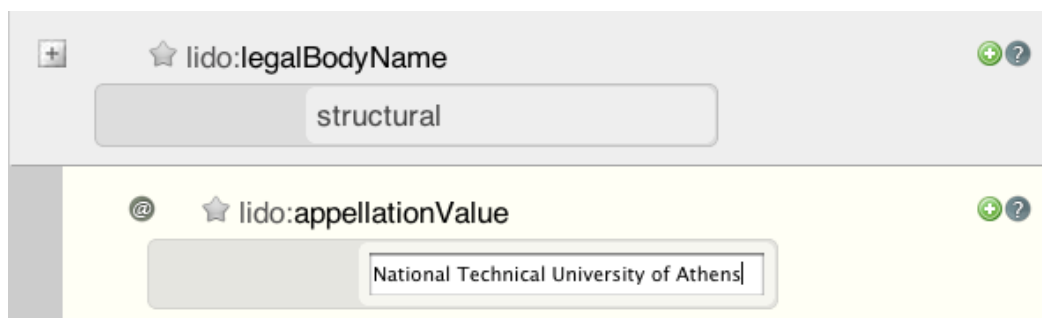
Figure 60: Constant value mapping
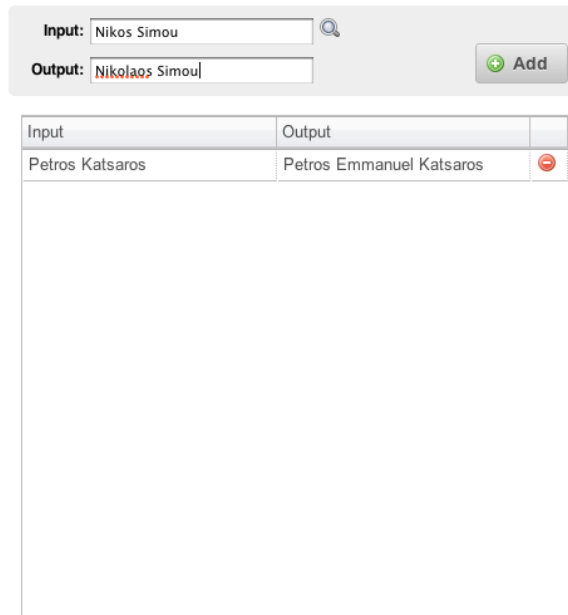
10.3.5.3 Concatenate mapping

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".
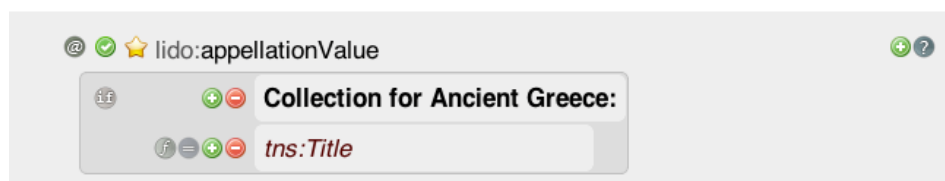
Figure 61: Concatenate mapping

10.3.5.4 *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 62: Value mapping editor

10.3.5.5 *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.

Figure 63: Functional mapping editor

10.3.5.6 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.
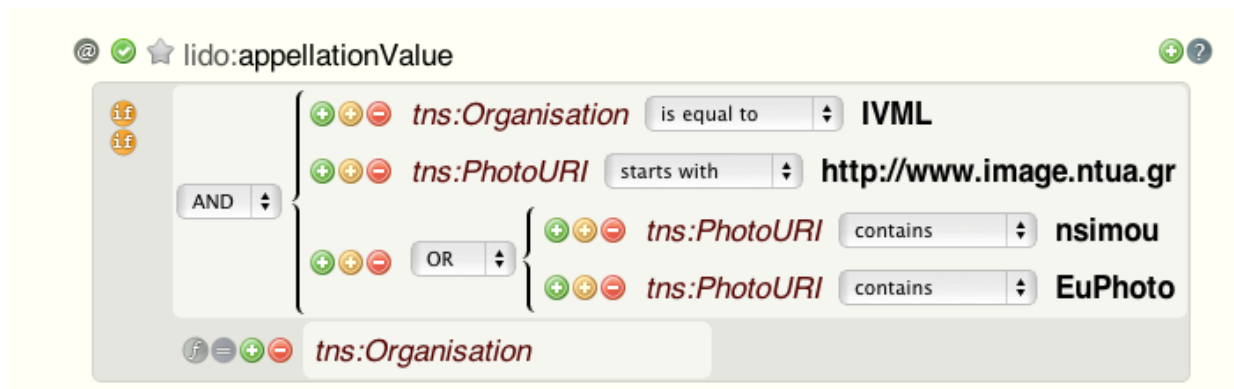
Figure 64: Conditional mapping

In the figure above we see a conditional mapping with many clauses, which 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" of EuPhoto then the value of the XPath tns:Organisation will be mapped to the XPath lido:appellationValue of the target schema.

10.3.5.7 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. In 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".

10.3.5.8 Mapping Using Terminologies

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 Material 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)
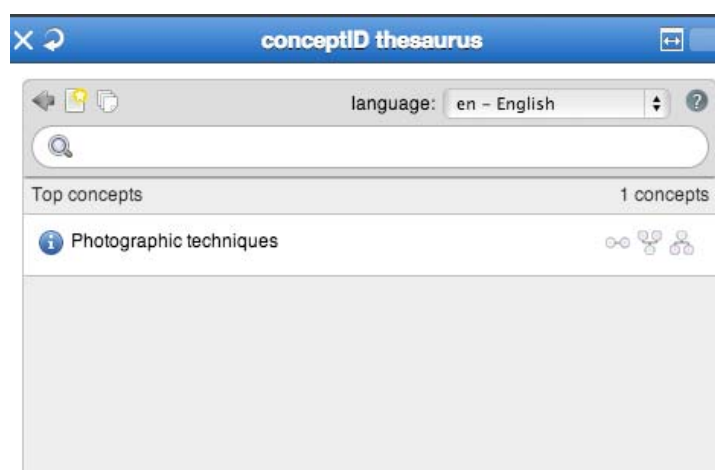

Figure 65: Terminology navigation pane

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 66: Narrower terms of term Materials

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 67: 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.

10.3.6 *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 AthenaPlus intermediate format), the ESE files, the Europeana preview and a report of the unmapped elements or mistakes by clicking on the [icon] button on the top right of the ingestion platform.
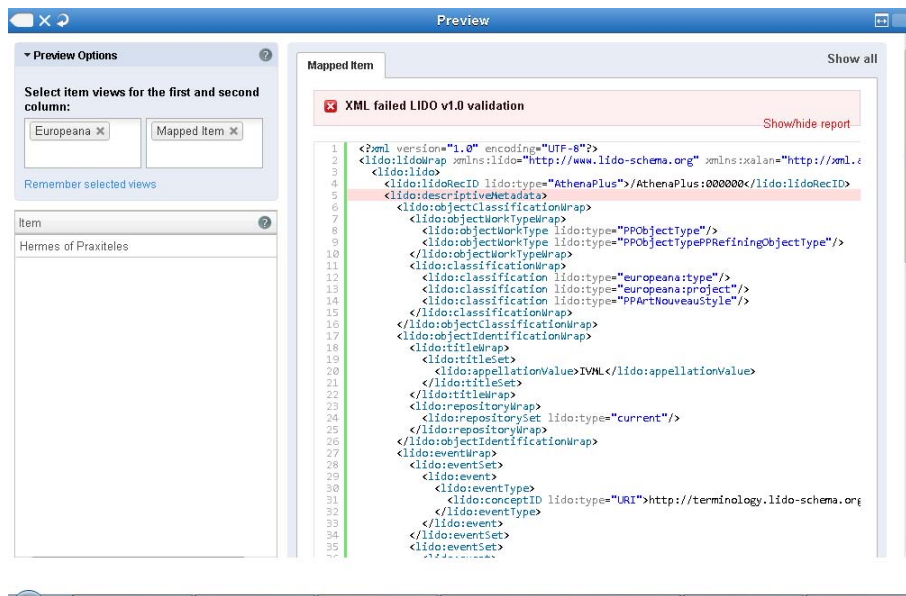

Figure 68: The output preview with an error highlighted

The above figure shows the output preview with an error highlighted. More information on the errors is shown in the "Report" pane.
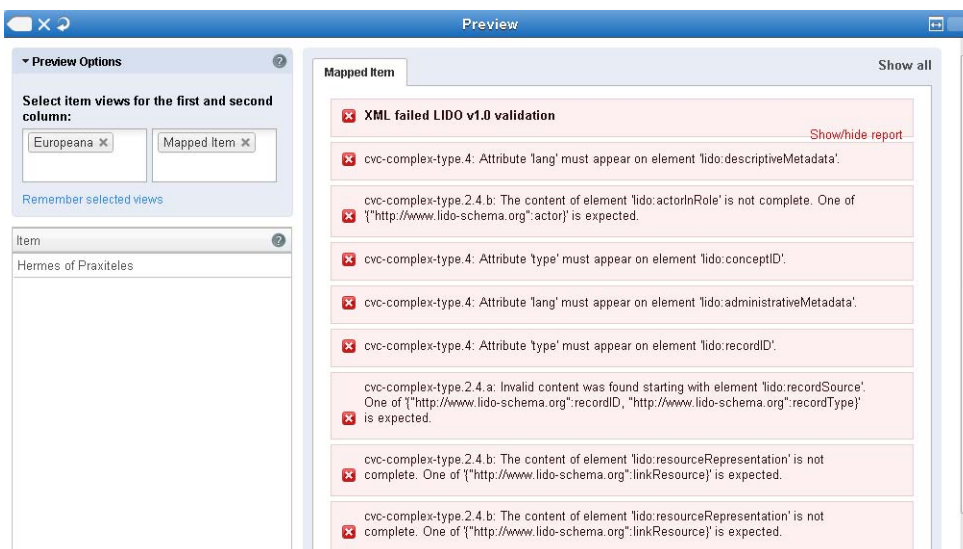
Figure 69: The report pane

The following figure shows the Europeana preview that is very useful since it simulates the view of your items on the Europeana portal.
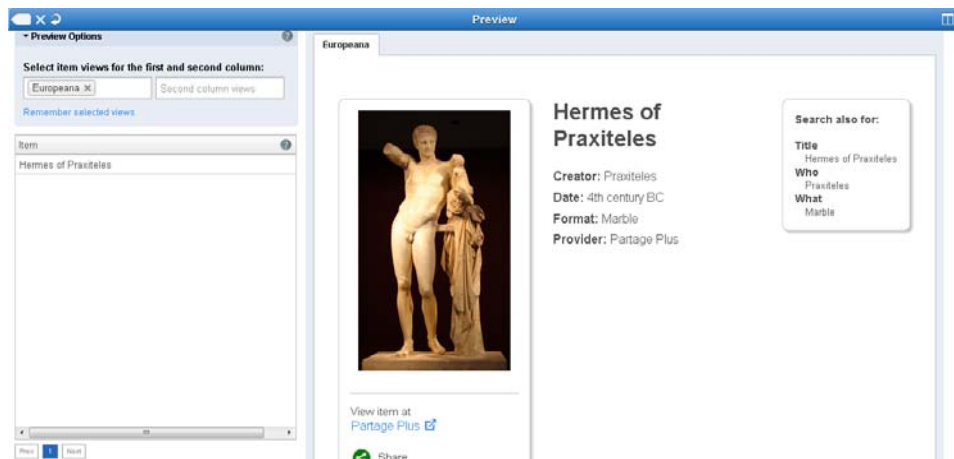
Figure 70: The Europeana preview

10.4 Transformation - publication to Europeana

After having performed mappings (see Section **Errore. L'origine riferimento non è stata trovata.**) select the "Transform" for the "Dataset Options". The "Transform" pane appears from which you can select a mapping for doing the transformation.
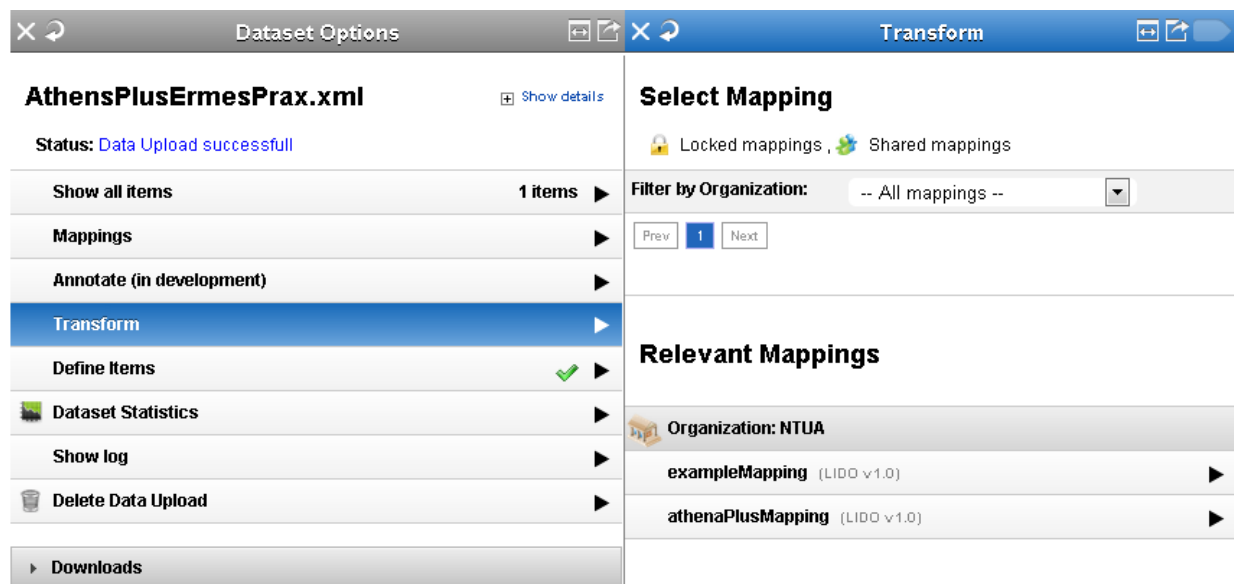

Figure 71: Transform dataset

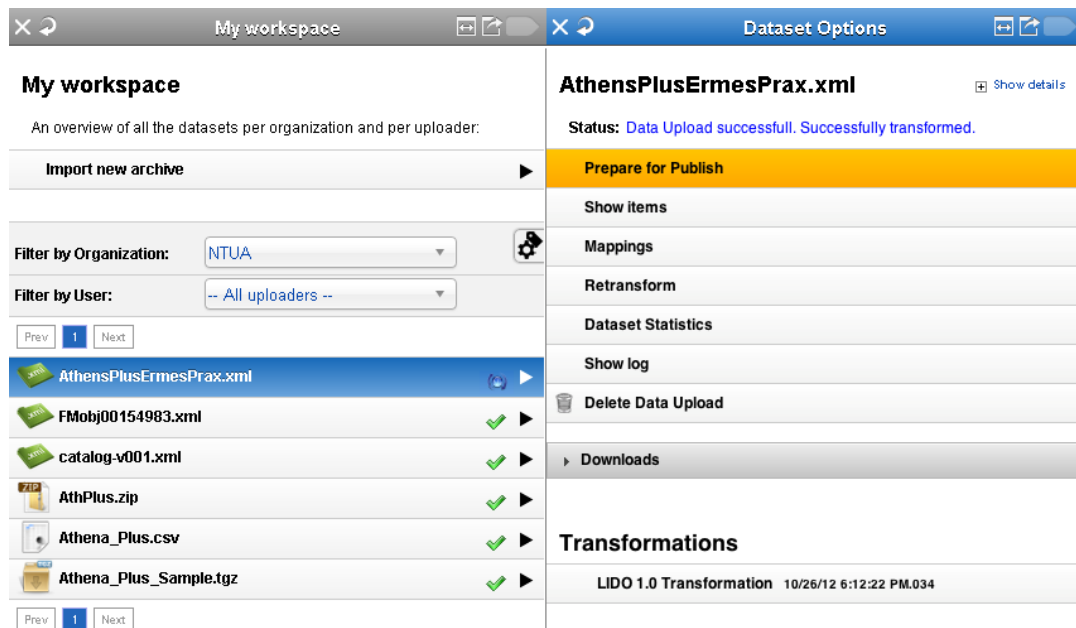After that the transformation the icon appears next to the dataset and prepare for publish option appears to the "Dataset Options" pane.

Figure 72: Transformed dataset

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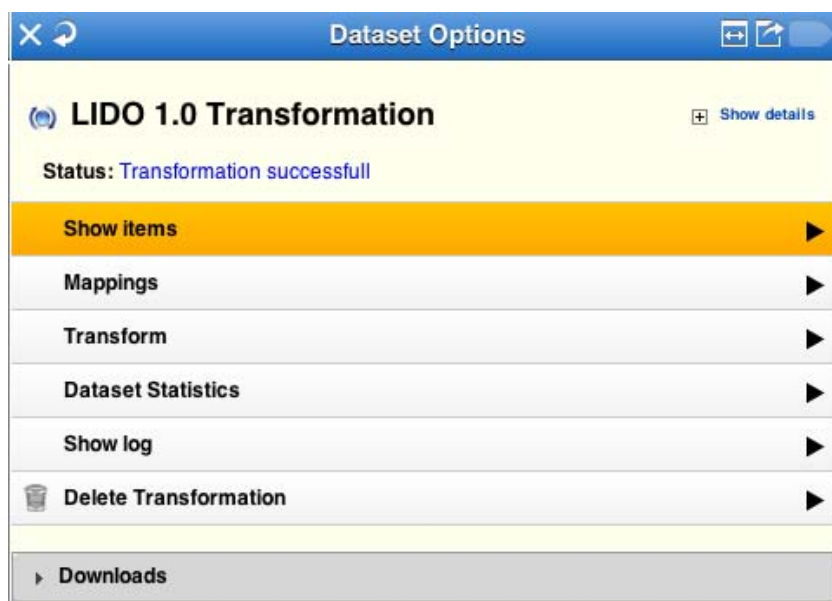

Figure 73: Transformation options

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.

Figure 57: Prepare for publication to Europeana

And now your data is ready for harvesting by Europeana.