



europæana
fashion

Ingestion Tool v1 Report

Deliverable 3.1

CIP ICT PSP 2011 – Call 5
Objective 2.1 : Aggregating content for Europeana
Project Number: 297167

Project start - end date (duration)
01/03/2012 - 28/02/2015 (36 months)

Name of Lead Contractor for this Deliverable: National Technical University of Athens
Author: Kostas Pardalis



Co-funded by the European Commission within
the ICT Policy Support Programme

Connected to:



DOCUMENT INFORMATION

Deliverable number: 3.1
Deliverable title: Ingestion Tool v1 Report
Contractual date of deliverable: M10 (December 2012)
Actual date of deliverable: 31st December 2012
Author(s): Kostas Pardalis
Participant(s): NTUA
Workpackage: WP3
Workpackage title: Content Aggregation, Access and Services
Workpackage leader: NTUA
Dissemination Level: PU (Public)
Version: 1.0
Keywords: Content Aggregation, Ingestion Tool, Mint Platform

HISTORY OF VERSIONS

Version	Date	Status	Author (Partner)	Description/Approval Level
0.1	21/12/2012	Draft	Kostas Pardalis (NTUA)	
1.0	31/12/2012	Final	Kostas Pardalis (NTUA)	Revised by Marco Rendina (FRD)

Acknowledgement

"Work partially supported by European Community under the Competitiveness and Innovation Framework Programme as a Best Practice Network with Grant Agreement n°297167. The author is solely responsible for the content of this paper. It does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of data appearing therein"

EXECUTIVE SUMMARY

This document reports on the first version of the Europeana Fashion “Ingestion Tool”, that has been made available online to all the consortium partners for validation, training and actual use in the content provision workflow towards the Europeana Fashion aggregator and Europeana. The technology team led by NTUA integrated all the necessary components into a common technology platform. The Europeana Fashion Ingestion Platform provides content holders with the ability to perform in an efficient way the required mapping of their own metadata schemas to the project's target metadata schema (see Deliverable 2.2 on EDM-fp), as well as their publication to the Europeana Fashion portal and the Europeana portal. It is based on NTUA's metadata interoperability platform MINT, that follows a typical web-based architecture offering an expanding set of services for metadata aggregation and remediation. It addresses the ingestion of metadata from multiple sources, the mapping of the imported records to a well-defined machine-understandable reference model, the transformation and storage of the metadata in a repository, and the provision of services that consume, process and remediate these metadata. Although its deployment is also guided by expediency, the system has been developed using established tools and open standards, embodying best practices in order to integrate familiar content provider procedures in an intuitive and transparent way also for newcomers.

The following document could also be used by content providers as a reference for the functionalities and the use of the MINT ingestion platform.

TABLE OF CONTENTS

Contents

Document Information.....	2
History of Versions.....	2
Executive Summary	3
Table of Contents	4
1 Introduction	6
2 Metadata Aggregation Platform.....	7
3 Mapping editor	8
4 The Mint Ingestion Tool	10
4.1 User & Organization Registration.....	10
4.1.1 User Registration	10
4.1.2 Organization Registration & Management.....	12
4.2 Import of Metadata	13
4.2.1 Dataset Options.....	15
4.2.2 Wrong Import.....	15
4.2.3 Successful Import.....	16
4.3 Mapping	20
4.3.1 Create New Mapping.....	20
4.3.2 Upload Mapping.....	24
4.3.3 Upload XSL.....	24
4.3.4 Mappings Management.....	25
4.3.5 Mappings.....	26
4.3.6 Preview Records	33
4.4 Transformation & Publication.....	34
5 Technical specification	36
5.1 Platform	36

5.2 ingestion	36
5.3 processing	36
5.4 normalization & vocabularies	36
6 Conclusion	37
Appendices.....	38
Appendix A: table of Figures	38
Appendix B: List of tables	39

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

- harvesting and aggregating metadata records that were created using shared community standards or proprietary metadata schemas,
- migrating from providers' models (whether standard or local) to a reference model.

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

Key functionalities include:

- Organization and user level access rights and role assignment.
- Collection and record management (XML serialization).
- 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 1 Ingestion Workflow

The metadata ingestion workflow, as illustrated in Figure 1, consists of four main steps. First is the Harvesting/Delivery procedure, which refers to the collection of metadata from content providers through common data delivery protocols, such as OAI-PMH, HTTP and FTP. Following is the Schema Mapping procedure, during which the harvested metadata are mapped to the common reference model. A graphical user interface assists content providers in mapping their metadata structures and instances to a rich, well defined schema (e.g. LIDO), using an underlying machine-understandable mapping language. Furthermore, it provides useful statistics about the provider's metadata while also supporting the share and reuse of metadata crosswalks and the establishment of template transformations. The third step is the Transformation procedure, which also aims at the transformation of the content provider's list of terms to the vocabularies and terminologies introduced by the reference model. The last step is the Revision/Annotation procedure that enables the addition and correction of annotations, group editing of items in order to assign metadata not available in the original context and, further transformations and quality control checks according to the aggregation guidelines and scope (e.g. for URI/URLs).

Metadata mapping is the crucial step of the ingestion procedure. It formalizes the notion of a metadata crosswalk, hiding the technical details and permitting semantic equivalences to emerge as the centerpiece. It involves a user-friendly graphical environment (Figure 2 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.

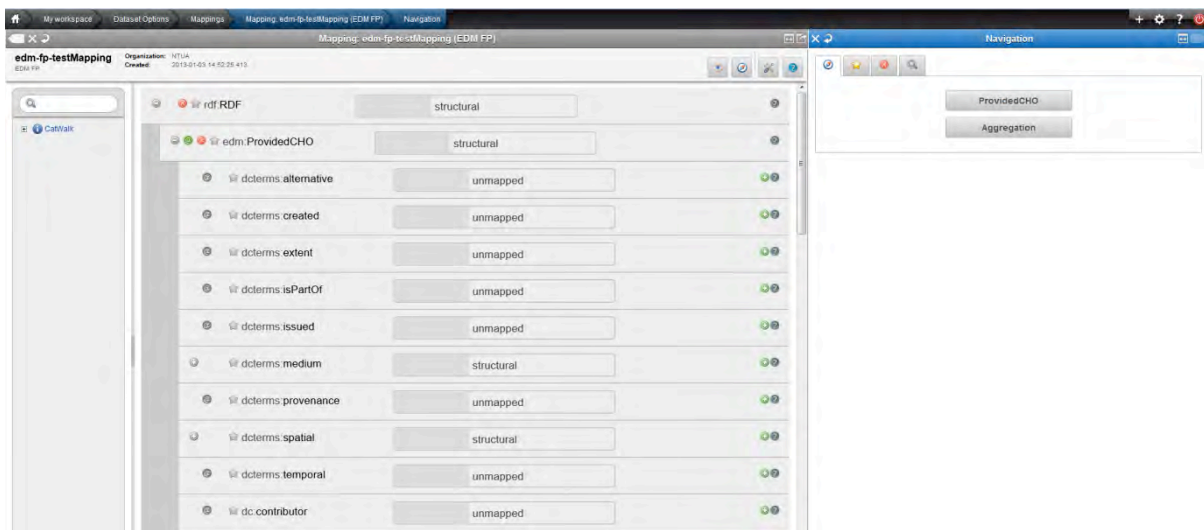


Figure 2 The Mint Mapping Tool

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

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

The user interface of the mapping editor is schema aware regarding the target data model and enables or restricts certain operations accordingly, based on constraints for elements in the target XSD. For example, when an element can be repeated then an appropriate button appears to indicate and implement its duplication. Several advanced mapping features of the language are accessible to the user through actions on the interface, including:

- String manipulation functions for input elements.
- m-1 mappings with the option between concatenation and element repetition.
- Structural element mappings.
- Constant or controlled value assignment.

- Conditional mappings (with a complex condition editor).
- Value mappings editor (for input and target element value lists)

Mappings can be applied to ingested records, edited, downloaded and shared as templates. Preview interfaces 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.

This section presents the MINT Ingestion tool functionality in detail. 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 and the Europeana Fashion Portal.

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

4.1 USER & ORGANIZATION REGISTRATION

4.1.1 USER REGISTRATION

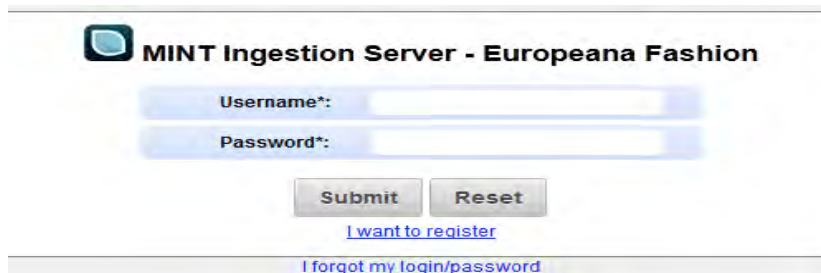
The very first step that each provider has to make is to register himself/herself and also his/her institution to the MINT mapping tool. The managers or the technical experts of the organizations that participate in Europeana Fashion are suggested to follow the next steps.

- User registration without organization (see Section 4.1.1.2) – to register themselves
- Organization Registration & Management (see Section 4.1.2) – to register their organization(s)
- User management (& User creation under an organization by its administrator) (see Section 4.1.1.3) – to create accounts for the users that work for their organization(s).

On the other hand the annotators of the organizations that participate in Europeana Fashion are suggested to register themselves under an existing organization (see Section 2.1.1.2). If the organization they work for is not yet registered, then they are suggested to contact the manager or the technical expert of their organization.

4.1.1.2 Self-Registration from the Mint Ingestion Tool

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



The screenshot shows the MINT Ingestion Server - Europeana Fashion landing page. It has a header with the title and a logo. Below the header is a login form with two input fields labeled 'Username*' and 'Password*', followed by 'Submit' and 'Reset' buttons. At the bottom of the form are two blue links: 'I want to register' and 'I forgot my login/password'.

Figure 3 Mint Ingestion Server Landing Page

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

Europeanana Fashion Ingestion Server - Registration

Username*: cpard

Password*: ••••••••

Password Confirmation*: ••••••••

First Name*: Kostas

Last Name*: Pardalis

Email*: cpard@image.ntua.gr

Contact phone num:

Job role:

☐ Join default organization for test purposes

or Select -- Please Select --

Organization:

If you can't find your organisation in the list, leave blank and press submit. You can then create an organisation in the administration tab.

Submit Reset

[Back to Login](#)

Figure 4 User Registration Form

At this point the user 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 Europeanana Fashion project and they just want to check the functionality of the MINT mapping tool.
- Select one of the existing organizations – This option is recommended to users that work for an organization that participates in the Europeanana Fashion 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 Europeanana Fashion 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.

After registering under an existing organization the user is able to log in to MINT. When this is done the user can see on the left, under the pane MINT Home -that is the central functionality pane of MINT mapping tool- the following options:

- 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.

4.1.1.3 User Management (User creation under an organization by its administrator)

For creating a new user or activating a user registered for the organization you are administrating log in to MINT mapping tool and select “Administration” 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 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 5 The create new user form

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

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

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

In addition an organization follows 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 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.

The screenshot shows a web form titled "Create Organization". The form contains the following fields and controls:

- Country*:** A dropdown menu with the selected value "--no country--".
- English name*:** A text input field.
- Name*:** A text input field.
- Organization acronym:** A text input field.
- Type*:** A dropdown menu with the selected value "--not specified--".
- Address:** A text input field.
- Organization url:** A text input field.
- Organization description:** A larger text input area.
- Select parent organization:** A dropdown menu with the selected value "-- No parent --".
- Primary contact user*:** A dropdown menu with the selected value "-- Please Select --".
- ☐ **Organization can publish**
- Submit** and **Reset** buttons at the bottom.

Figure 6 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.

4.2 IMPORT OF METADATA

To import your metadata to the MINT mapping tool, you have to log in and to select “My workspace” from the MINT Home pane. Note that your user role must have been assigned to annotator or higher, 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.

The screenshot shows a web application window titled "Start Import". It contains several sections for different upload methods:

- Local upload:** Includes a radio button, a text input field, an "Upload a file" button, and a note "(.csv, .txt, .zip, .xml files allowed)". There is also a checkbox labeled "This is a CSV upload".
- Remote FTP/HTTP Upload:** Includes a radio button and a text input field.
- OAI URL:** Includes a radio button, a text input field, and a link icon.
- From Date (YYYY-MM-DD):** Includes a text input field.
- To Date (YYYY-MM-DD):** Includes a text input field.
- OAI SET:** Includes a radio button, a text input field, and a link icon.
- Namespace Prefix:** Includes a radio button, a text input field, and a link icon.
- Server filename:** Includes a radio button and a text input field.
- Server file path for upload:** Includes a text input field.
- Upload for Organization:** Includes a dropdown menu with "NTUA" selected.
- Parent organization upload support:** Includes a text input field.
- This import conforms:** Includes a checkbox, a dropdown menu with "-- Select schema --", and a note "Select this option in addition to your import method in case your upload already conforms to the selected schema and no mapping is necessary."

At the bottom, there are "Submit" and "Reset" buttons.

Figure 7 Upload Form

HTTP Upload

It Allows the 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.

The screenshot shows a web application window titled "Import". It contains several sections for different upload methods:

- Local upload:** Includes a radio button, a text input field, an "Upload a file" button, and a note "(.csv, .txt, .zip, .xml files allowed)". There is also a checkbox labeled "This is a CSV upload".
- This is a CSV upload:** Includes a checked checkbox, a red warning message "The file must be encoded using UTF-8 (Unicode). Other encodings are not supported.", and a checkbox labeled "Contains header".
- Contains header:** Includes a checked checkbox and a note "If checked first line of CSV will be used as header".
- Define the field separator:** Includes a dropdown menu with "comma (,)" selected.
- Define the escape character:** Includes a dropdown menu with "no escape" selected.


Figure 8 CSV Upload Options

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



Remote FTP/HTTP Upload

This option is used to upload 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.

4.2.1 DATASET OPTIONS

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

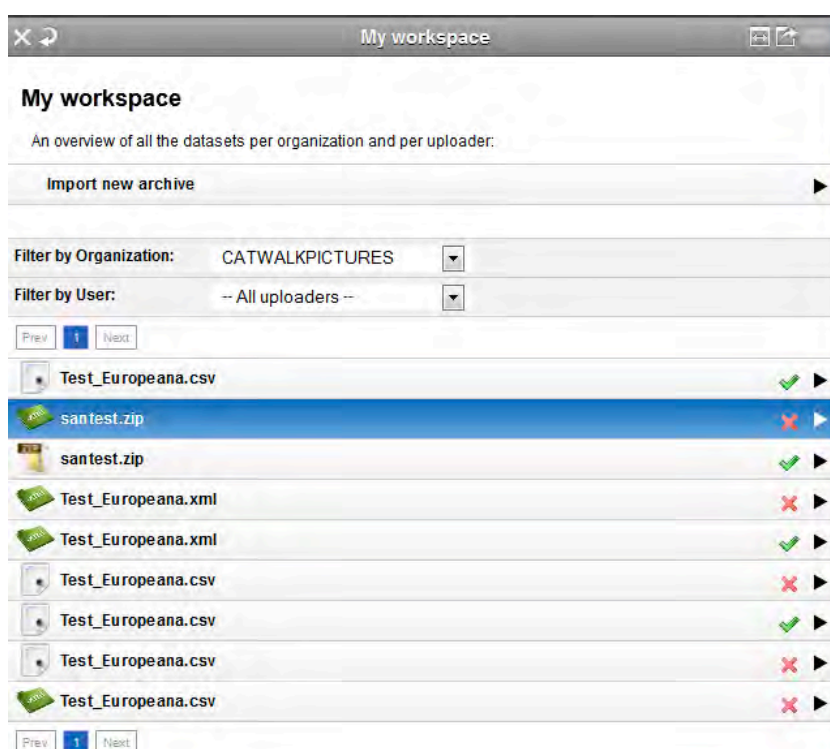


Figure 9 Workspace with successful and unsuccessful imports

4.2.2 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.

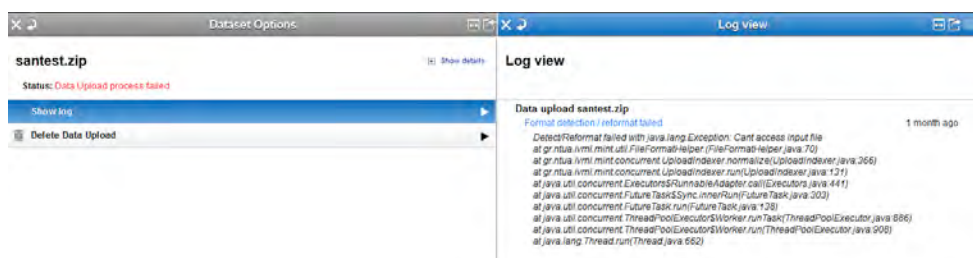


Figure 10 Error found in Import

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.

4.2.3 SUCCESSFUL IMPORT

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

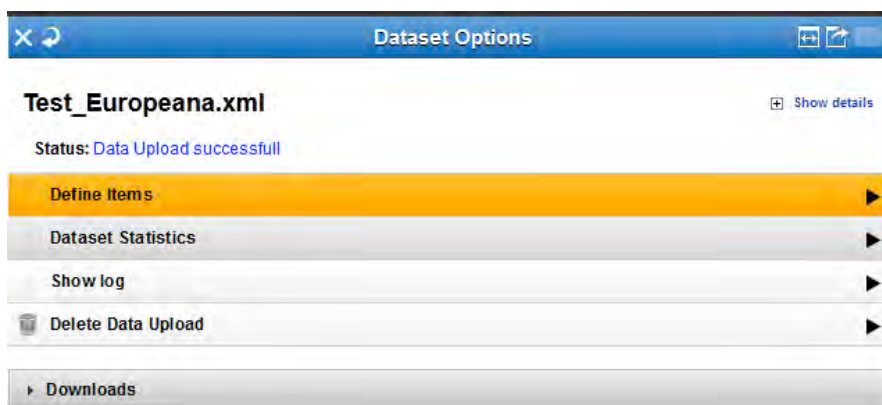


Figure 11 Available options for a successfully uploaded dataset

4.2.3.1 Define Items

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

Value	Count
10	1
100	1
1000	1
1001	1
1002	1
1003	1
1004	1
1005	1
1006	1
1007	1
1008	1
1009	1
101	1
1010	1

Figure 12 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. In green letters you see the elements that have unique values. You have to set the following and then to click on the “Done” Button.

- 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.

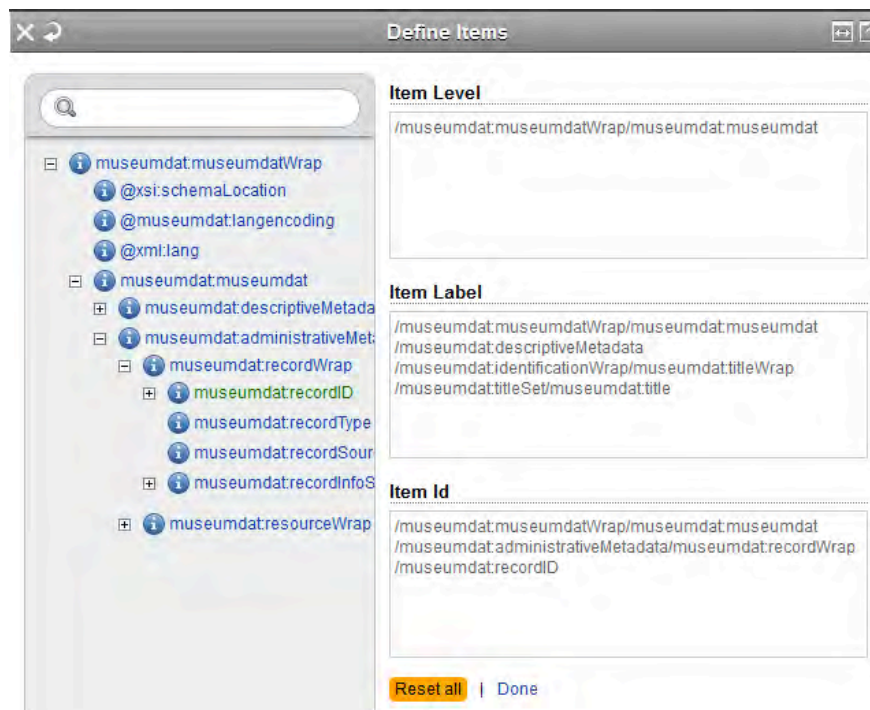


Figure 13 Define Items

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.

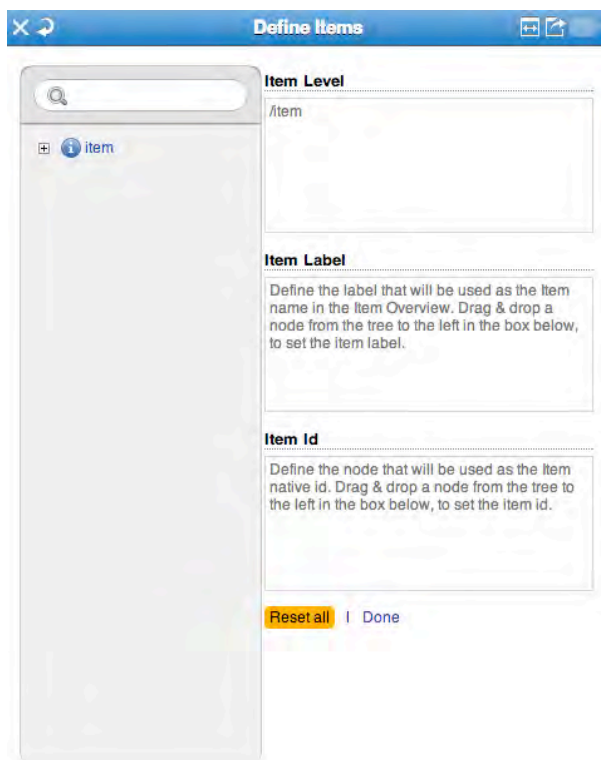


Figure 14 Define items for a CSV import

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

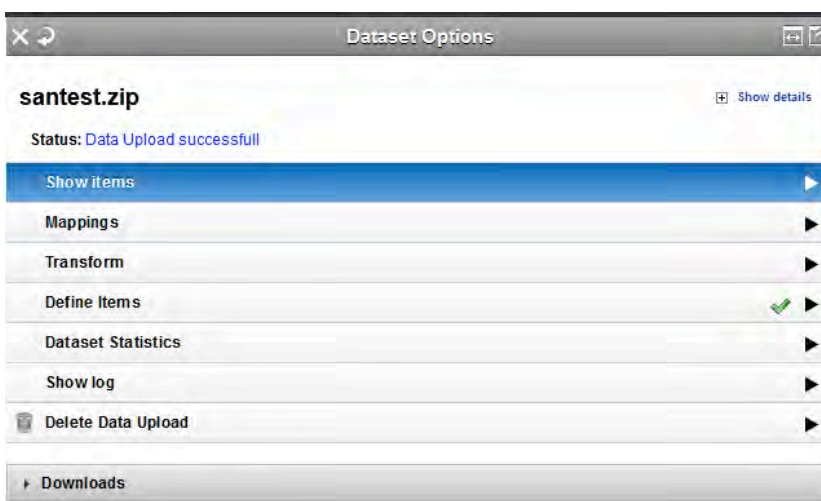


Figure 15 Dataset Options after having defined the items

4.2.3.2 Show Items

After having defined the items, 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).

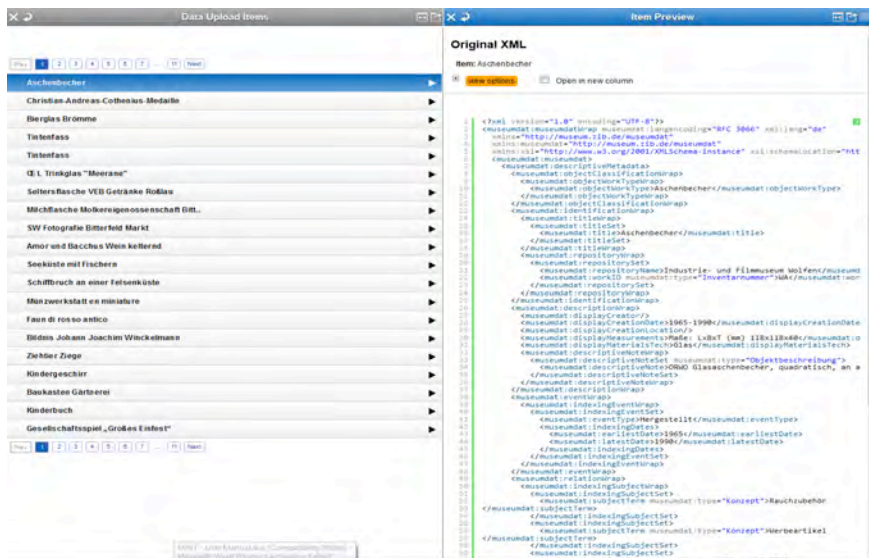


Figure 16 Show Items

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

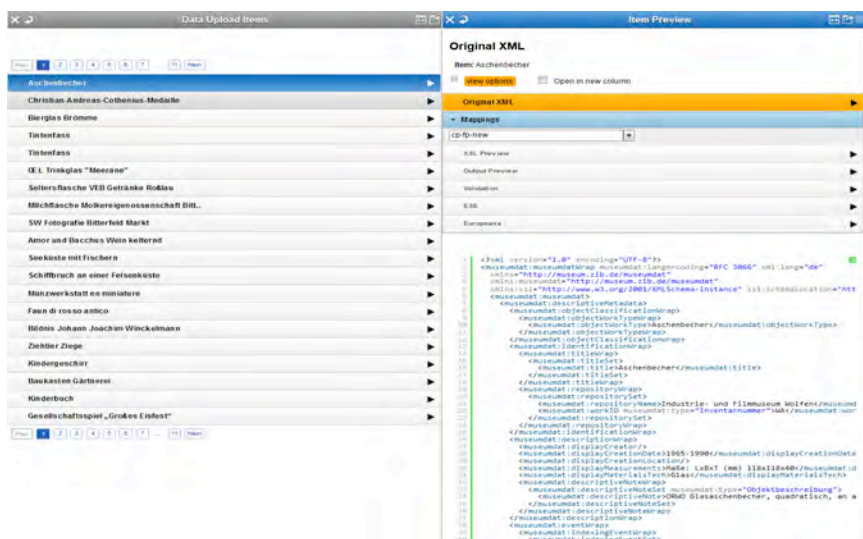


Figure 17 Show items after having defined a mapping

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

4.2.3.3 Dataset statistics

After having defined the items, 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.

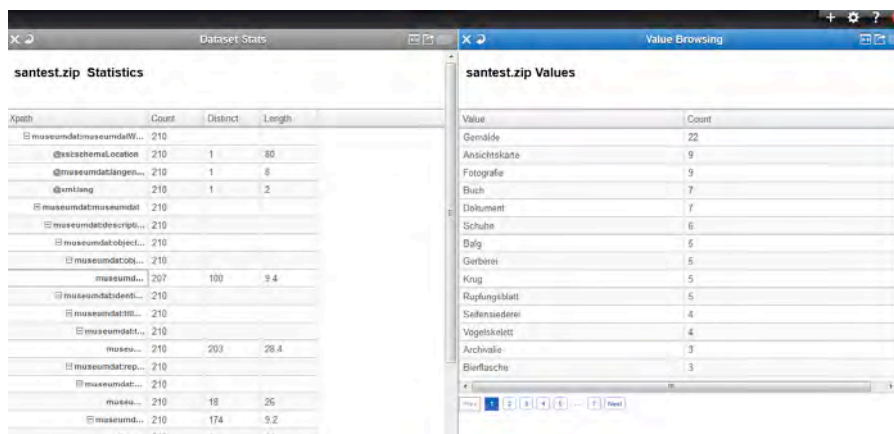


Figure 18 Dataset statistics

4.3 MAPPING

After having defined the items, the user can select the “Mappings” from 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.

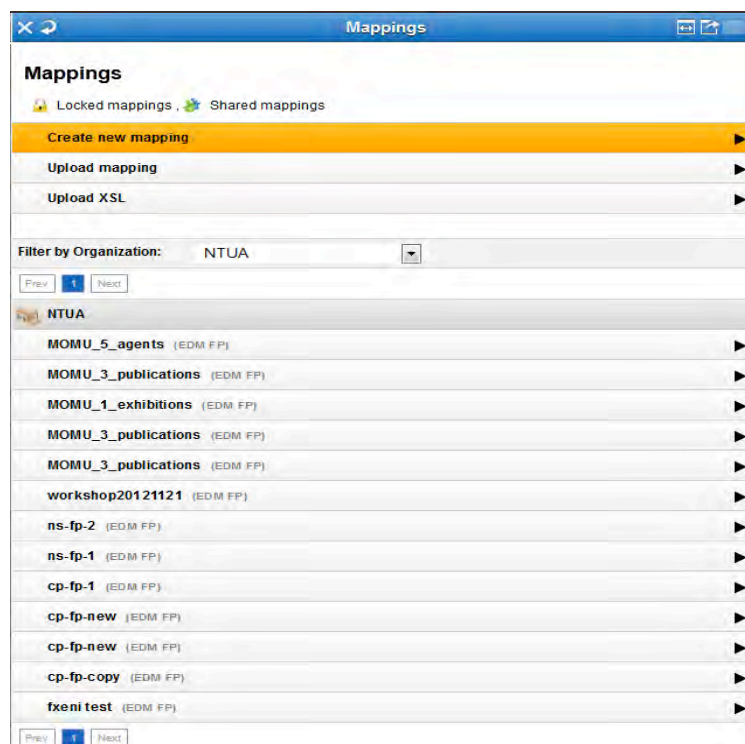


Figure 19 Mapping options

4.3.1 CREATE NEW MAPPING

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

Figure 20 New mapping creation options

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

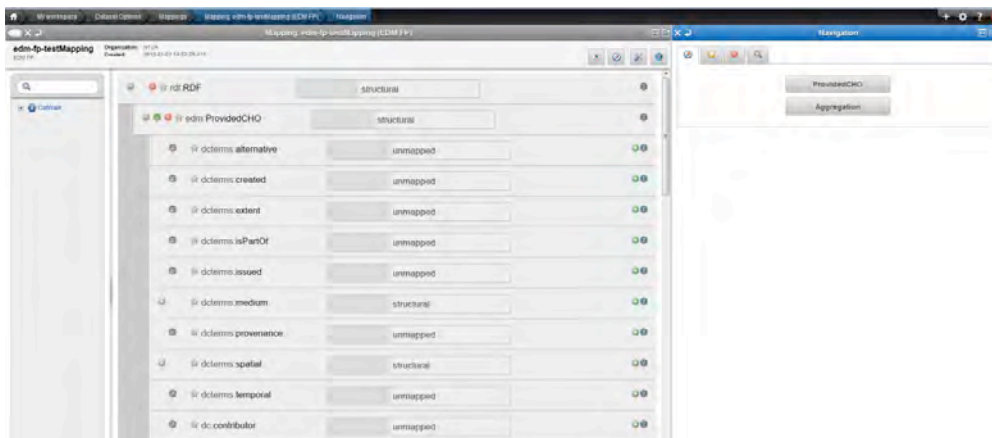


Figure 21 The mapping tool

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

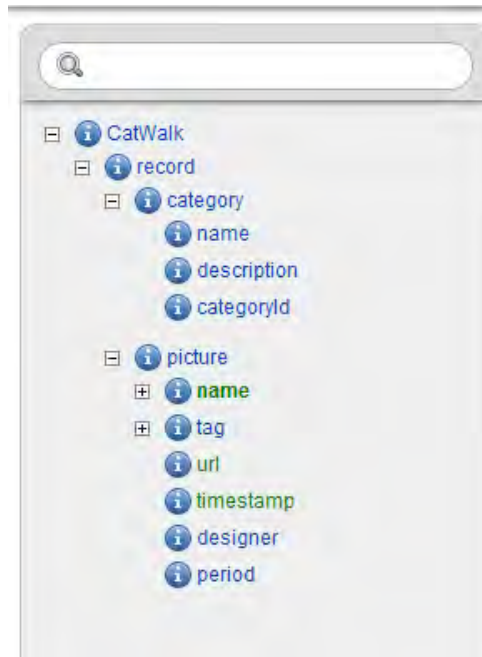



Figure 22 Input schema area

On the right hand side, buttons correspond to high-level elements of the target schema (see Figure 24 - if not visible click on the  button on the top right of the mapping tool) and are used to access their corresponding sub-elements. These are visualized on the middle part of the screen as a tree structure of embedded boxes, representing the internal structure of the complex element. The user is able to interact with this structure by clicking to collapse and expand every embedded box that represents an element, along with all relevant information (attributes, annotations) defined in the XML schema document. To perform an actual (one to one) mapping between the input and the target schema, a user has to simply drag a source element from the left and drop it on the respective target in the middle.

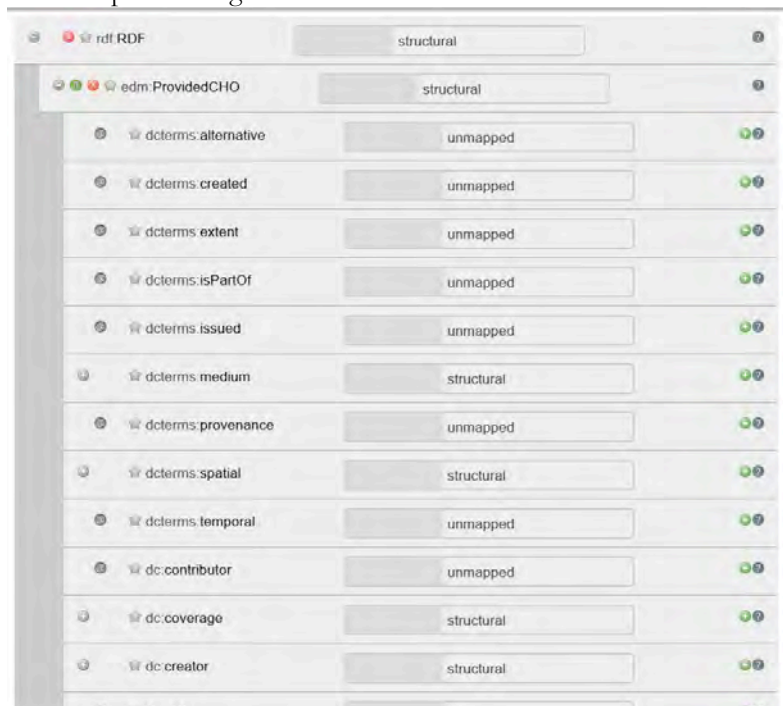


Figure 23 The mapping area



Figure 24 The navigation Area

For the needs of the Europeana Fashion project and for assisting providers that are not familiar with EDM-fp, bookmarks will be created to the EDM-fp elements that will accommodate the Europeana Fashion 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 Europeana Fashion mandatory fields, to see the respective EDM-fp element and easily map a value from his/her metadata.

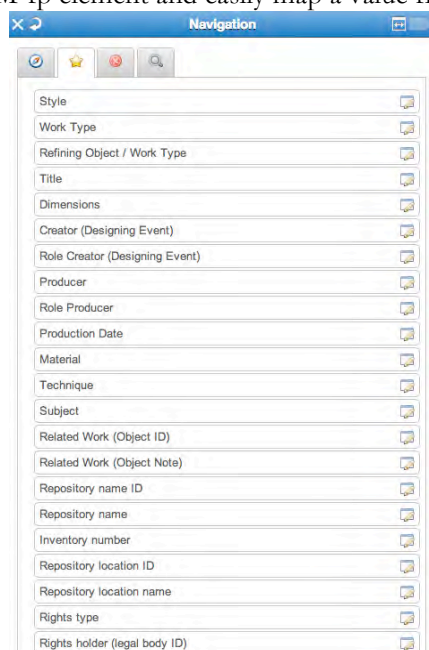


Figure 25 Navigation area, an example of bookmarks navigation

Finally, a third way of exploring the target schema is available and this is by searching the xpaths. The user can click on the magnifying glass icon 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.

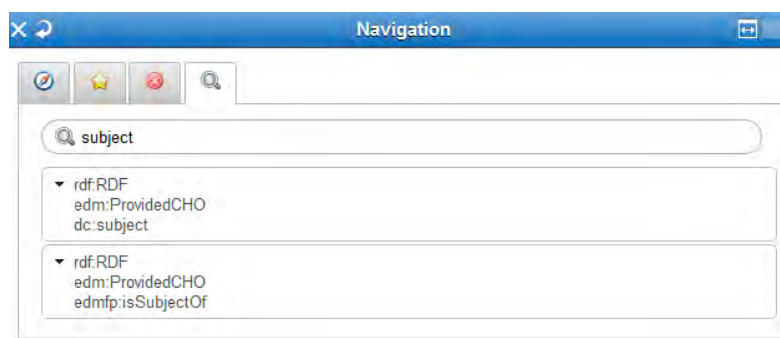


Figure 26 Navigation area - search target schema

4.3.2 UPLOAD MAPPING

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

 A screenshot of the 'Upload mapping' pane. It has a blue header with 'Upload mapping' and icons. Below the header, there's a section titled 'New Mapping'. Inside this section, there's a text input field labeled 'First upload the mapping file:' with a button 'Upload a file' next to it. Below that is a text input field labeled 'Mapping Name*:' which is empty. Underneath is a dropdown menu labeled 'Create with schema:' with 'EDM FP' selected. At the bottom of the form are two buttons: 'Submit' (orange) and 'Reset' (blue).

Figure 27 Upload mapping

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

4.3.3 UPLOAD XSL

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

 A screenshot of the 'Upload XSL' pane. It has a blue header with 'Upload XSL' and icons. Below the header, there's a section titled 'New XSL'. Inside this section, there's a text input field labeled 'First upload the mapping file:' with a button 'Upload a file' next to it. Below that is a text input field labeled 'Mapping Name*:' with the text 'TestXSL' entered. Underneath is a dropdown menu labeled 'Create with schema:' with 'EDM FP' selected. At the bottom of the form are two buttons: 'Submit' (orange) and 'Reset' (blue).

Figure 28 Upload XSL

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



Figure 29 The XSL Editor

4.3.4 MAPPINGS MANAGEMENT

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

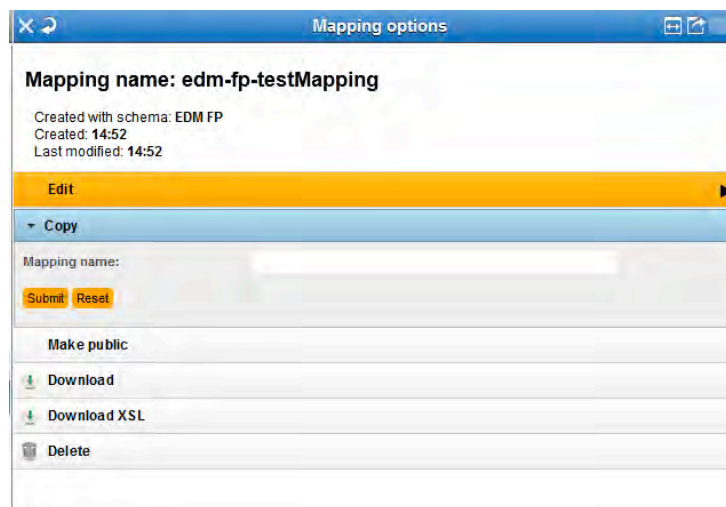


Figure 30 Mapping management

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

4.3.5 MAPPINGS

4.3.5.1 Mapping Environment

As mentioned previously, 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.

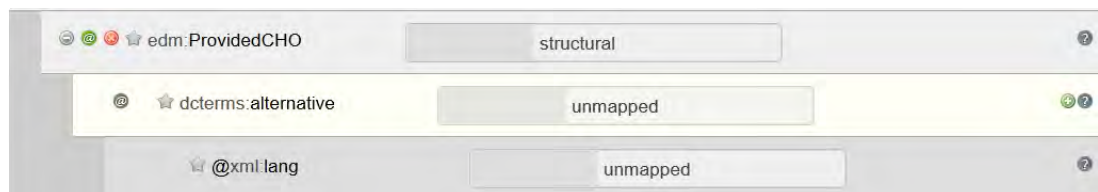


Figure 31 Unmapped element in the mapping area

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

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

Table 1 Mapping tool informative icons

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



Figure 32 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
	Appears on the left of the mapping to indicate that no conditional mapping is used. By clicking on it conditional mapping is activated.
	Appears on the left of the mapping to indicate that a conditional mapping is used. By clicking on it the conditional mapping is de-activated.
	Appears on the left of the mapping to indicate that no functional mapping is used. By clicking on it the functional mapping is activated.
	Appears on the left of the mapping to indicate that a functional mapping is used. By clicking on it the functional mapping is de-activated.
	Appears on the left of the mapping to indicate that no value mapping is used. By clicking on it the value mapping is activated.
	Appears on the left of the mapping to indicate that a value mapping is used. By clicking on it the value mapping is de-activated.
	Appears on the top left of a mapping. By clicking on it concatenate mapping is activated.
	Appears on the top left of a mapping. By clicking on it you remove the mapping.

Table 2 Mapping tool functional icons

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

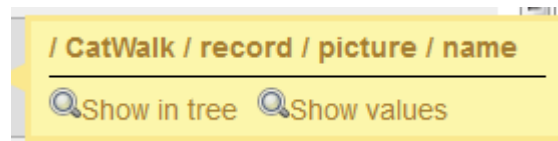


Figure 33 Shortcuts to input tree and values from mapping

4.3.5.2 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.

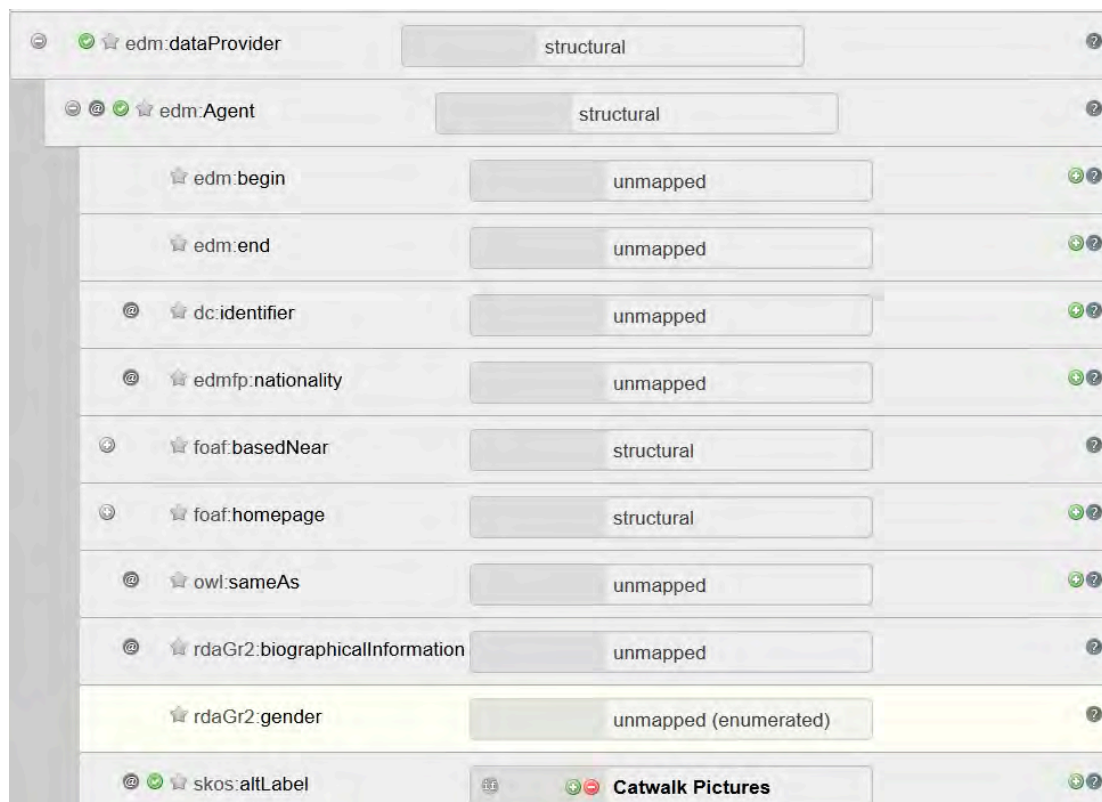



Figure 34 Constant value mapping

4.3.5.3 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 "Catwalk" is concatenated with the value "Pictures" inside the Agent that represents the data provider in EDM-fp.

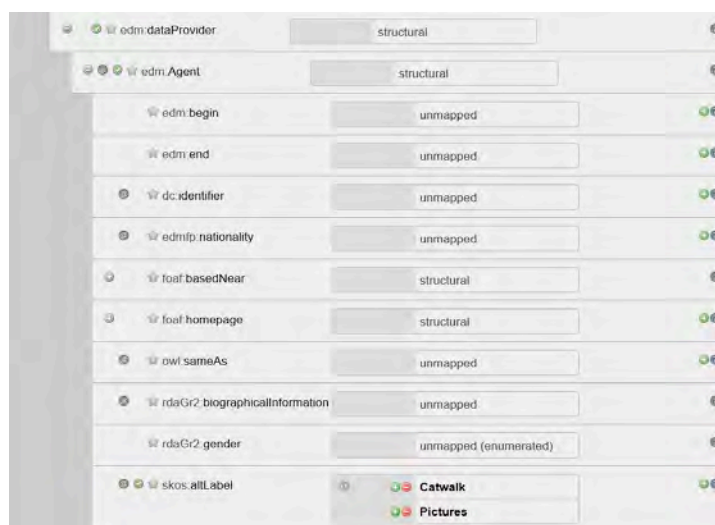





Figure 35 Concatenate mapping

4.3.5.4 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.

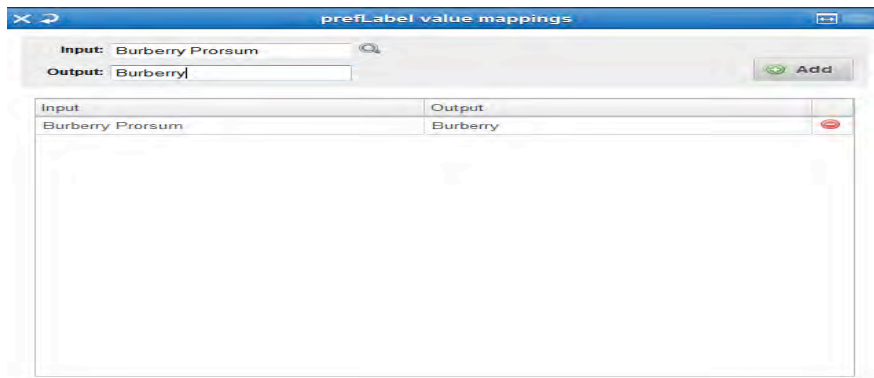



Figure 36 Value mapping editor

4.3.5.5 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.

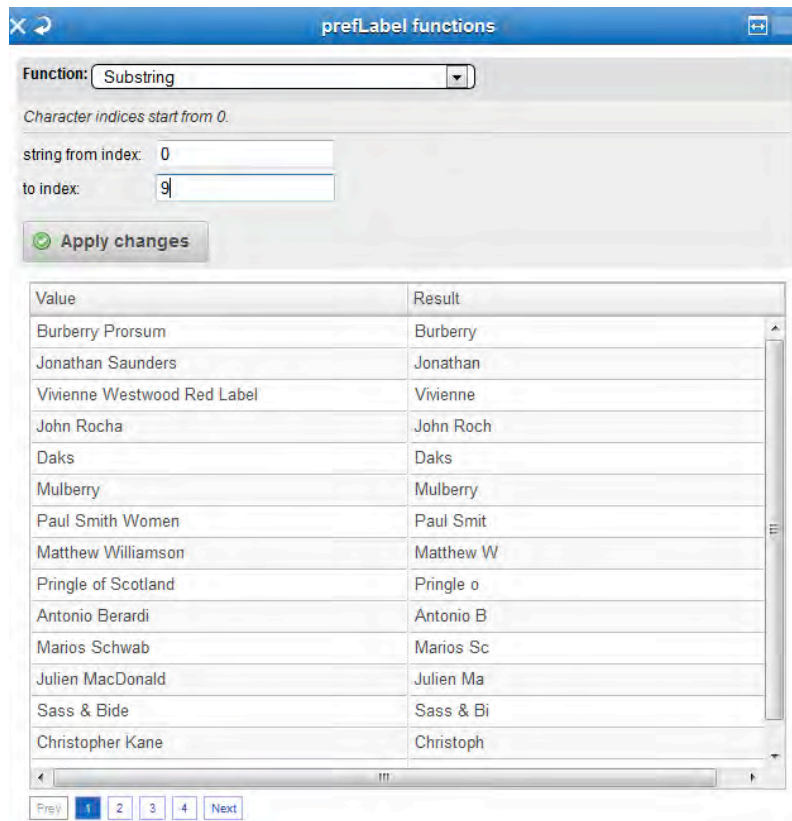



Figure 37 Functional mapping editor

4.3.5.6 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.
- **Starts with** – sets a condition that is satisfied if the given xpath starts with the given value.
- **Ends with** – sets a condition that is satisfied if the given xpath ends with the given value.



Figure 38 Conditional mapping

In the figure above we see an example of a conditional mapping with more than one clause. The interpretation of the mapping is the following, if the designer element contains a value that equals to Burberry and the period element contains a value of “aw11” then the title will have the value “Autumn Winter 2011 Collection for Burberry”.

4.3.5.7 Structural Mapping

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

4.3.5.8 Mapping Using Terminologies

An important outcome of the Europeana Fashion portal, is the creation of a thesaurus or a set of terminologies describing the fashion domain, the definition and creation of such terminologies are outcomes of WP 2. These terminologies will be used to control the values of specific EDM-fp elements, e.g. the color and technique elements, for the enrichment of metadata. There are two ways of using terminologies for the enrichment of your metadata through MINT. The first is to assign a term to your complete dataset (similarly to constant mapping). To do that double click on an element that takes values from a terminology 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)

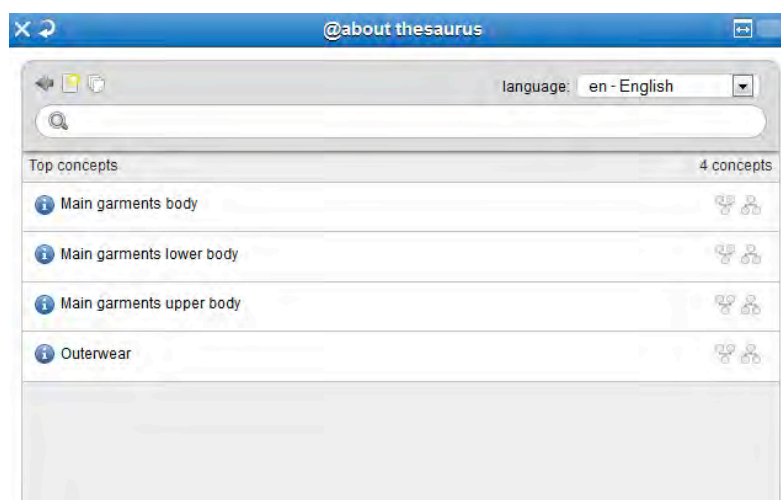


Figure 39 Terminology navigation panel

In this pane the top concepts of the terminology appear. 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.

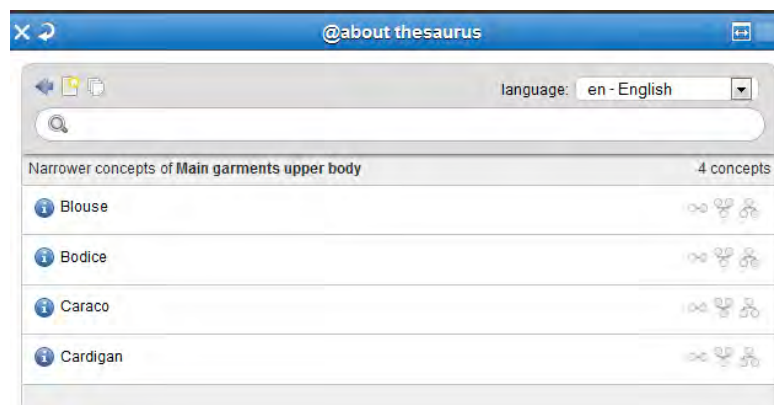


Figure 40 Viewing the narrow terms of a concept

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 🔍.

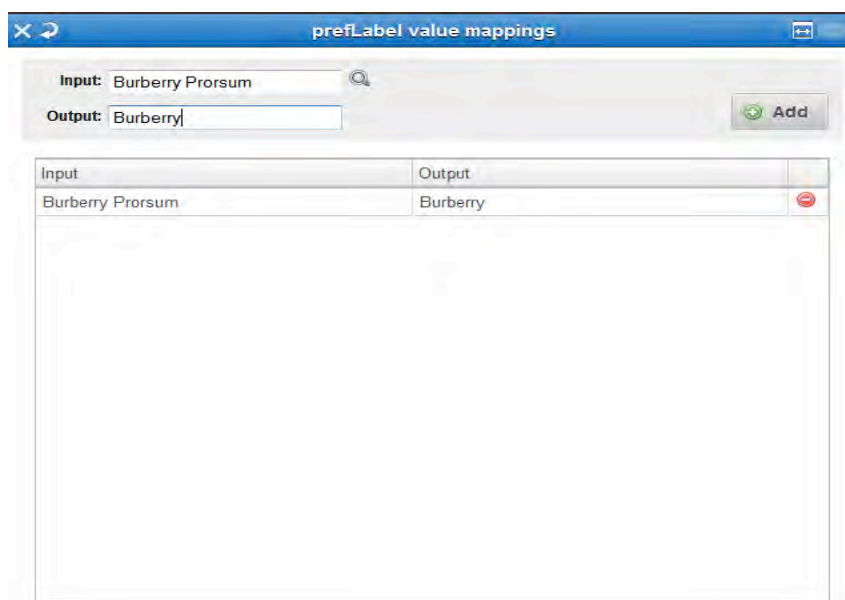



Figure 41 Value mapping using thesaurus terms

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.

4.3.6 PREVIEW RECORDS

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 Partage Plus intermediate format), the ESE files, the Europeana preview and a report of the unmapped elements or mistakes by clicking on the  button on the top right of the mapping tool.

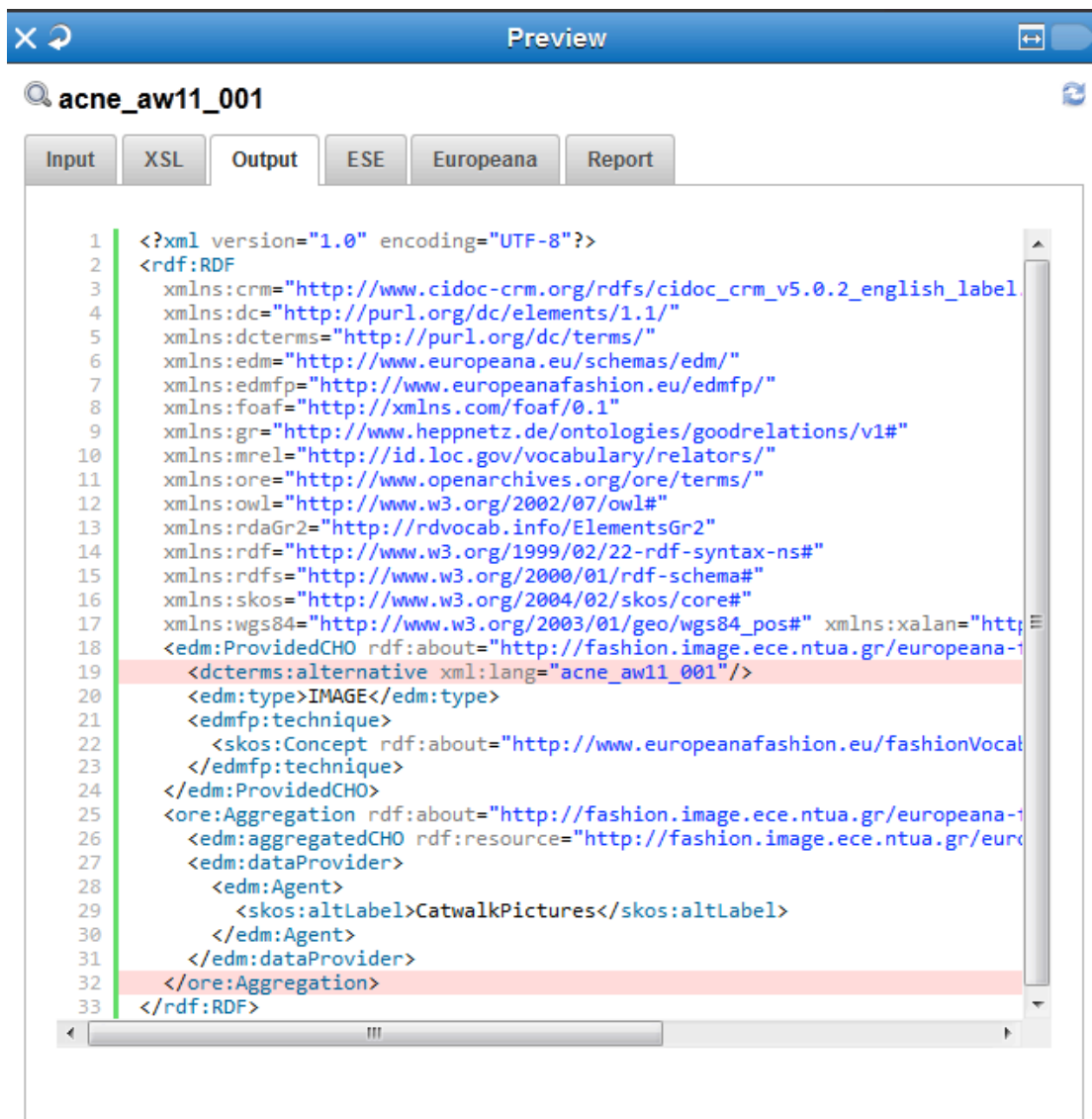


Figure 42 Previewing the result of a mapping with errors

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

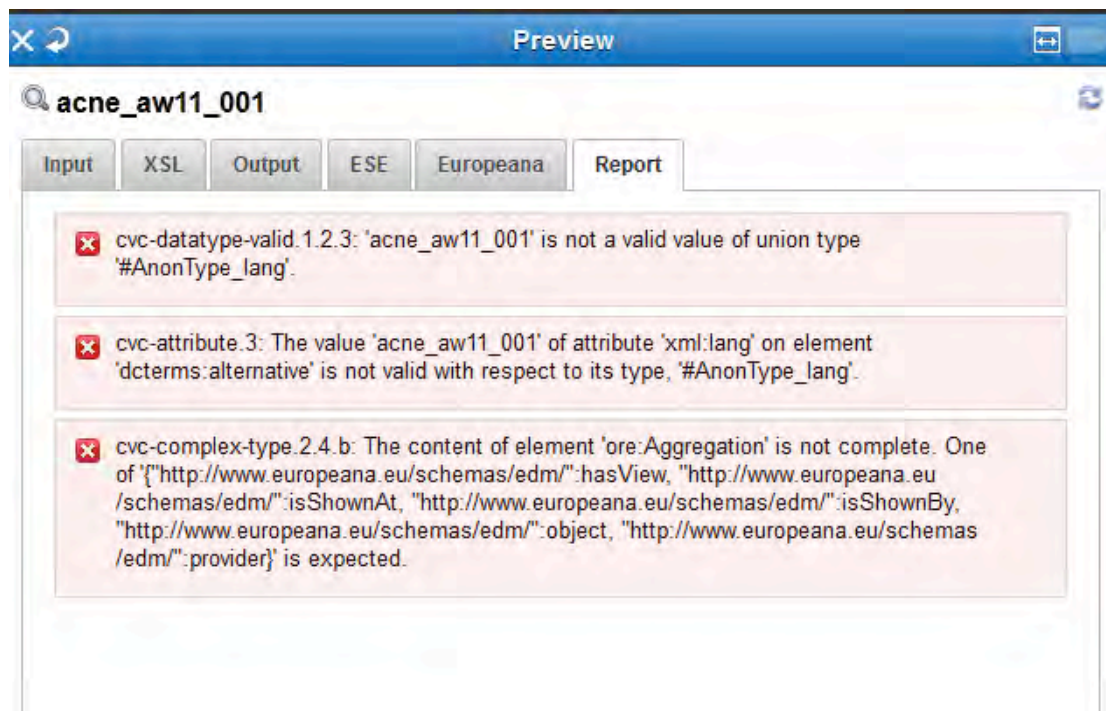



Figure 43 Mapping report

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

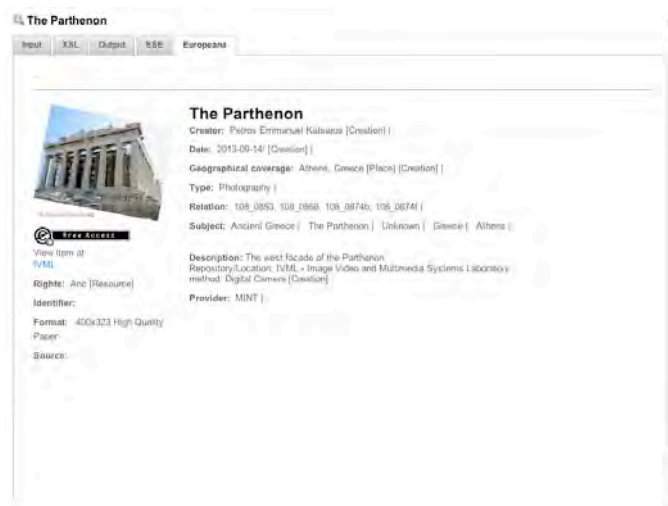


Figure 44 An example of the Europeana preview window

4.4 TRANSFORMATION & PUBLICATION

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

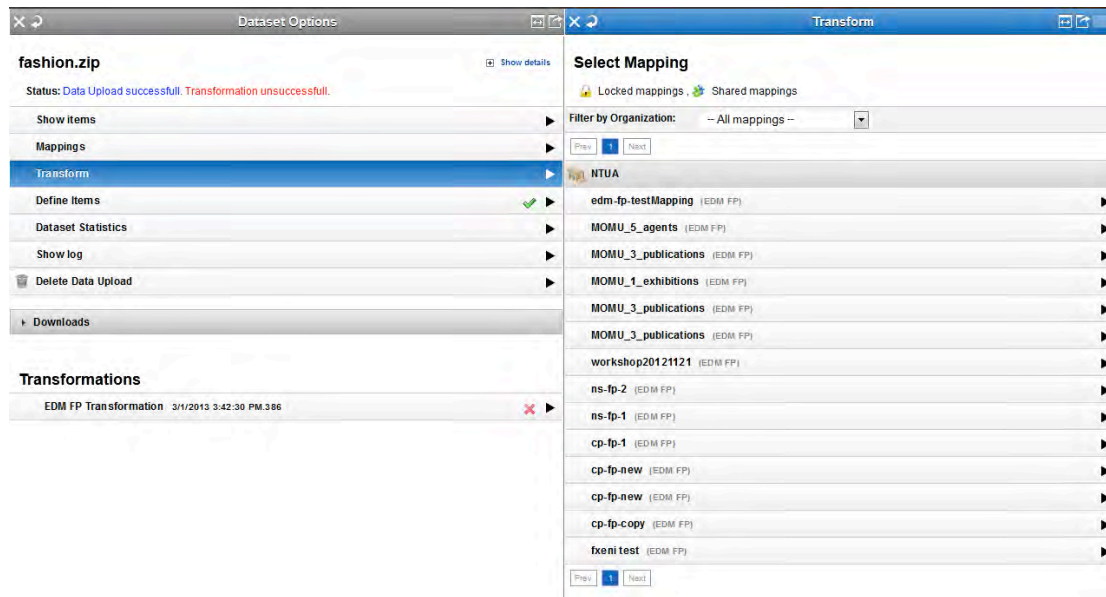



Figure 45 Transform a dataset

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

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 mapping tool 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, with most notable being the results of T2.3 "Creation of a multilingual Fashion Thesaurus" that will be directly integrated to the Mint platform in order to be used during the mapping process by the content providers. 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.

The present document, which constitutes the deliverable D3.1 “Ingestion Tool v1 Report”, is a user reference of the first release of the Europeana Fashion Ingestion Platform, that is made available online for validation, for the large-scale contribution of content to Europeana. NTUA hosts the tool and maintains the service.

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.

The deliverable describes the Mint platform that is deployed for the needs of Europeana Fashion, as it is required by T3.1, and it is available at the following address: <http://oreo.image.ntua.gr:9990/fashion>.

This platform enables providers to start processing metadata for their successful delivery to Europeana and the Europeana Fashion portal.

APPENDICES

APPENDIX A: TABLE OF FIGURES

Figure 1 Ingestion Workflow	7
Figure 2 The Mint Mapping Tool.....	8
Figure 3 Mint Ingestion Server Landing Page	10
Figure 4 User Registration Form	11
Figure 5 The create new user form	12
Figure 6 Create organization form	13
Figure 7 Upload Form.....	14
Figure 8 CSV Upload Options.....	14
Figure 9 Workspace with successful and unsuccessful imports	15
Figure 10 Error found in Import.....	16
Figure 11 Available options for a successfully uploaded dataset.....	16
Figure 12 Information about the values of a node	17
Figure 13 Define Items	17
Figure 14 Define items for a CSV import.....	18
Figure 15 Dataset Options after having defined the items.....	18
Figure 16 Show Items.....	19
Figure 17 Show items after having defined a mapping.....	19
Figure 18 Dataset statistics	20
Figure 19 Mapping options	20
Figure 20 New mapping creation options.....	21
Figure 21 The mapping tool.....	21
Figure 22 Input schema area	22
Figure 23 The mapping area.....	22
Figure 24 The navigation Area.....	23

Figure 25 Navigation area, an example of bookmarks navigation	23
Figure 26 Navigation area - search target schema	24
Figure 27 Upload mapping.....	24
Figure 28 Upload XSL	24
Figure 29 The XSL Editor.....	25
Figure 30 Mapping management	25
Figure 31 Unmapped element in the mapping area.....	26
Figure 32 An element that has been mapped in the mapping area	26
Figure 33 Shortcuts to input tree and values from mapping.....	27
Figure 34 Constant value mapping.....	28
Figure 35 Concatenate mapping	28
Figure 36 Value mapping editor	29
Figure 37 Functional mapping editor	30
Figure 38 Conditional mapping.....	31
Figure 39 Terminology navigation pane.....	31
Figure 40 Viewing the narrow terms of a concept	32
Figure 41 Value mapping using thesaurus terms	32
Figure 42 Previewing the result of a mapping with errors	33
Figure 43 Mapping report.....	34
Figure 44 An example of the Europeana preview window	34
Figure 45 Transform a dataset	35

APPENDIX B: LIST OF TABLES

Table 1 Mapping tool informative icons	26
Table 2 Mapping tool functional icons.....	27