

MINT – METADATA INTEROPERABILITY SERVICES

Nikolaos Simou – National Technical University of
Athens

What is MINT?

2

- Mint is a herb having hundreds of varieties that is mostly known as a mouth and breath fresher



Basic Concepts

3

- Metadata
 - ▣ Data about data
- Interoperability of metadata
 - ▣ Transformation of heterogeneous metadata into one inter-operable metadata standard
- Europeana
 - ▣ www.europeana.eu is an internet portal that acts as an interface to millions of books, paintings, films, museum objects and archival records that have been digitized throughout Europe

What is MINT?

4

- MINT is an open source, web based platform for Metadata INTeroperability
 - ▣ <http://mint.image.ece.ntua.gr>
 - ▣ It has been successfully used in more than 15 Europeana feeder projects
 - 300 cultural organizations
 - 500 users
 - ▣ More than 6.000.000 metadata records have been produced by it and published on Europeana

Scope

5

- Present MINT and its functionalities
- Familiarize the audience with a typical metadata aggregation workflow
- Highlight the importance of high-quality metadata

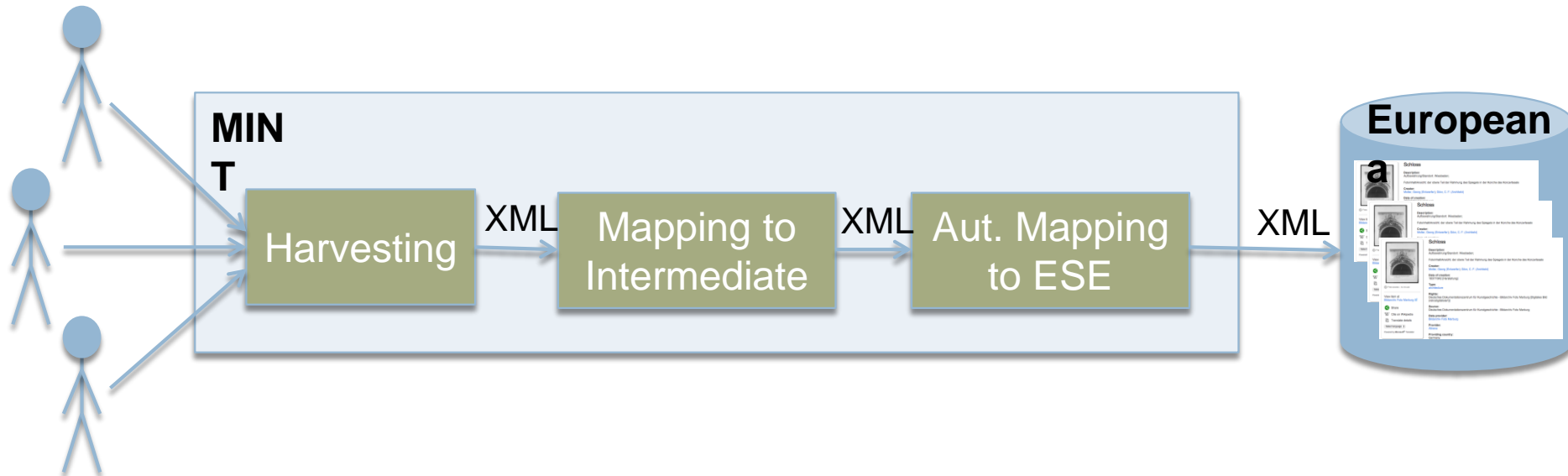
MINT – The early days

6

- MINT development started by the NTUA team during the Athena project back in 2008
- Athena's main objectives were
 - ▣ The delivery of 4.200.000 metadata records coming from 23 different museums and organisations to Europeana
 - ▣ The development of a metadata standard for the museum sector that would act as an intermediate for the project
 - ▣ The development of technical tools for assisting content providers to meet project objectives

MINT – Initial Workflow

7



Harvesting requirements

8

- Different metadata models
 - ▣ Standards e.g. Dublin Core
 - ▣ In house metadata models
- Different File formats
 - ▣ XML
 - ▣ CSV
- Different delivery protocols
 - ▣ File upload (i.e. HTTP)
 - ▣ Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)

Mapping & Transformation req.

9

- Agnostic to metadata input
- Target schema based on a metadata model
 - ▣ XSD support
- Crosswalks between known schemas
 - ▣ Project's intermediate to ESE
- User friendly interface
 - ▣ Most of the content providers did not have any technical background, they were (in most cases) well aware of their metadata

Mapping Editor

10

The screenshot displays the Mapping Editor interface for 'testLido (LIDO 1.0)'. The interface is divided into three main sections, each highlighted with an orange rounded rectangle:

- Input Schema Area (Left):** Contains a tree view of the input schema elements. The tree is organized as follows:
 - tns:ImageMetadata
 - @xsi:schemaLocation
 - tns:AdministrativeMetadata
 - tns:Identifier
 - tns:MetadataURI
 - tns:PhotoURI
 - tns:Organisation
 - tns:DescriptiveMetadata
 - tns:Title
 - tns:Date
 - tns:Photographer
 - tns:Photographic_Technique
 - tns:Location
 - tns:Photographic_Practice
 - tns:Material
 - tns:Description
 - tns:Copyright
 - tns:Dimensions
 - tns:Actor
 - tns:RetrievalMetadata
 - tns:Subject
 - tns:Actor
 - tns:Place
 - tns:Related_Object
- Mapping Area (Center):** Displays the mapping configuration for the 'lido:lido' root element. It shows a hierarchical structure of elements and their mappings:
 - lido:lido (structural)
 - lido:lidoRecID (mapped to /Photography:000000)
 - lido:objectPublishedID (mapped to unmapped)
 - lido:category (structural)
 - lido:descriptiveMetadata (structural)
 - lido:administrativeMetadata (structural)
- Navigation Area (Right):** A list of elements available for navigation, including:
 - xml:@lang
 - lido:term
 - lido:classification
 - lido:classificationWrap
 - lido:objectClassificationWrap
 - lido:appellationValue
 - lido:titleSet
 - lido:titleWrap
 - lido:objectIdentificationWrap
 - lido:descriptiveMetadata
 - xml:@lang
 - lido:@type
 - lido:recordID
 - lido:appellationValue
 - lido:legalBodyName
 - lido:recordSource
 - lido:recordWrap
 - lido:linkResource
 - lido:resourceRepresentation
 - lido:linkResource
 - lido:resourceRepresentation
 - lido:term

Further requirements...

11

- Better control of metadata and its quality
 - ▣ Metadata cleaning services
 - Many of the providers have realised that they also hold a lot of rubbish they wanted to get rid before publishing to Europeana
 - ▣ Validation according to target schema
 - Mapping to a common target schema itself did not provide high quality metadata
 - ▣ Preview interfaces
 - Providers wanted to preview their metadata for controlling its quality before the actual publication

Mappings – Concatenate mapping

12

- Allows users to combine more than one mappings for producing a new mapping

The screenshot shows a mapping editor interface with two mapping blocks. The top block is for the property `skos:altLabel` and the bottom block is for the property `@xml:lang`.

Top Mapping Block:

- Property: `skos:altLabel`
- Condition: `if`
- Value: `Εθνικό`
- Value: `Μετσόβιο`
- Value: `Πολυτεχνείο`

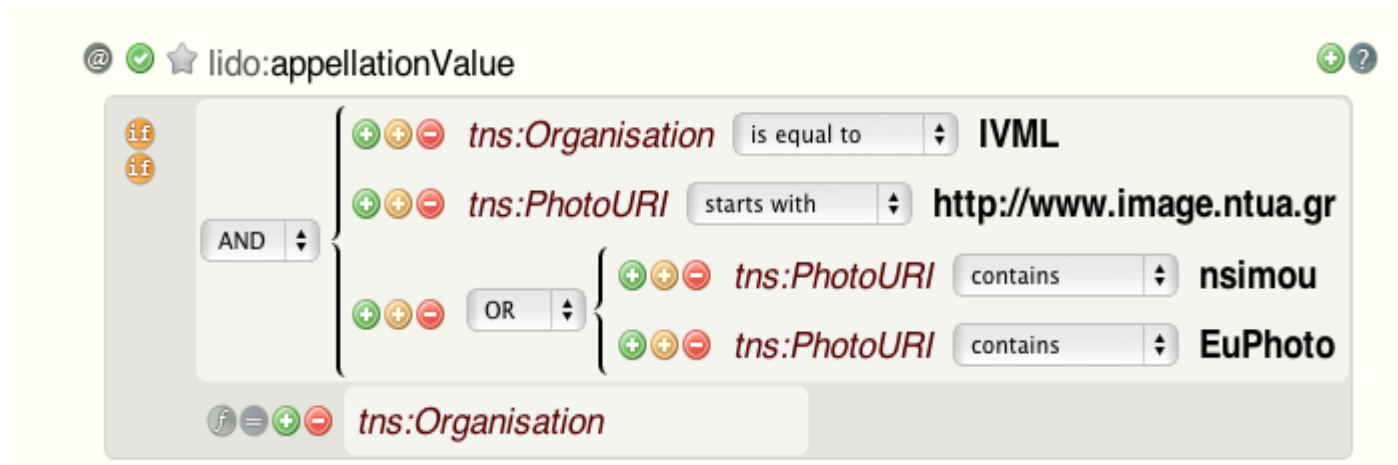
Bottom Mapping Block:

- Property: `@xml:lang`
- Condition: `if`
- Value: `Greek, Modern`

Mappings – Conditional Mapping

13

- Allows users to set conditions on their mappings
 - ▣ 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



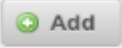
Mappings – Value Mapping


14

- Allows users to normalize their metadata by mapping values of their input to specific values they set

Input:

Output:



Input	Output	
Petros Katsaros	Petros Emmanuel Katsaros	

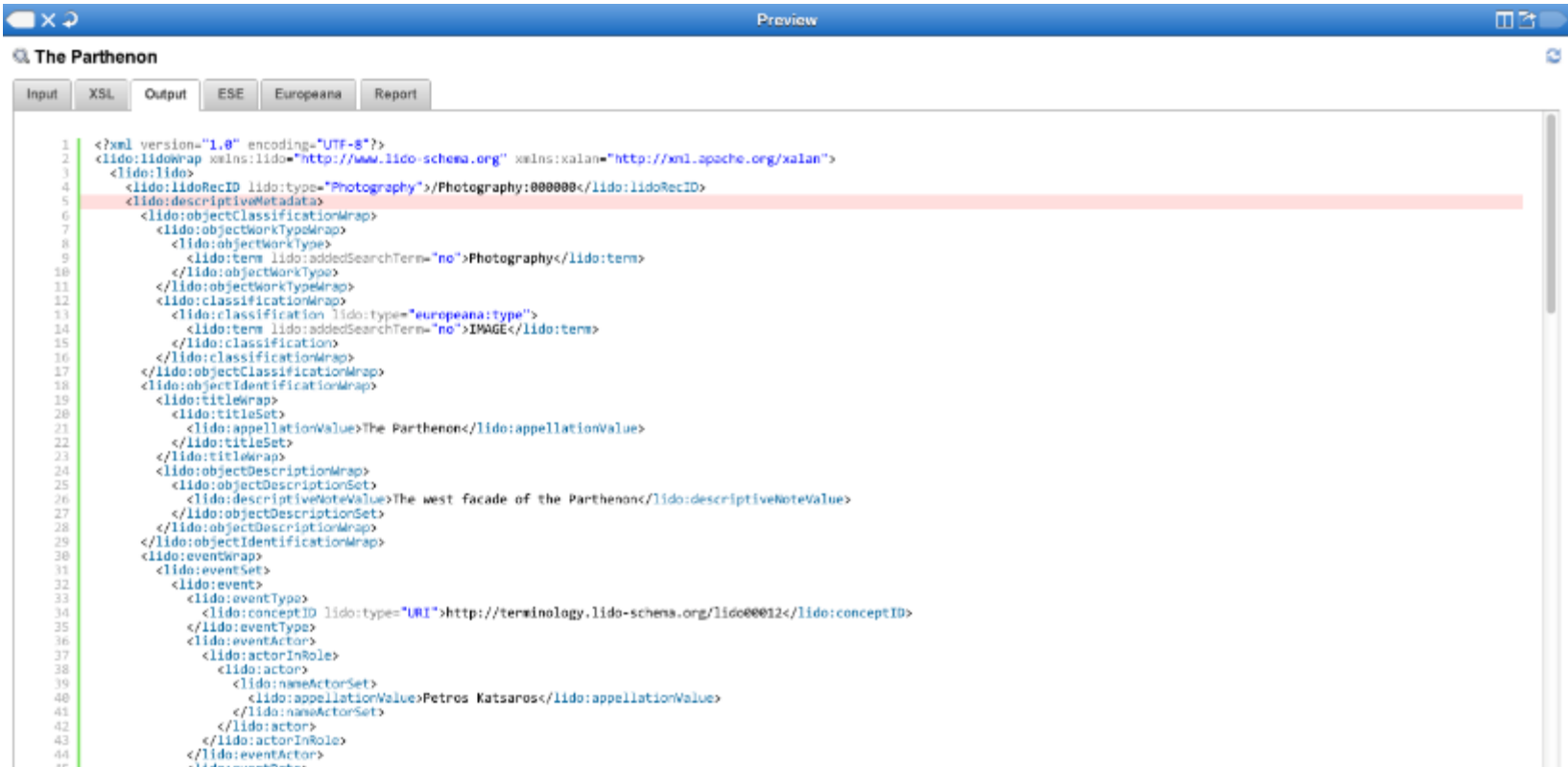
Mappings – Functional Mapping

15

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

Preview & Validation

16




```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <lido:lidoWrap xmlns:lido="http://www.lido-schema.org" xmlns:xalan="http://xml.apache.org/xalan">
3   <lido:lido>
4     <lido:lidoRecID lido:type="Photography"/></lido:lidoRecID>
5     <lido:descriptiveMetadata>
6       <lido:objectClassificationWrap>
7         <lido:objectWorkTypeWrap>
8           <lido:objectWorkType>
9             <lido:term lido:addedSearchTerm="no">Photography</lido:term>
10            </lido:objectWorkType>
11          </lido:objectWorkTypeWrap>
12          <lido:classificationWrap>
13            <lido:classification lido:type="europeana:type">
14              <lido:term lido:addedSearchTerm="no">IMAGE</lido:term>
15            </lido:classification>
16          </lido:classificationWrap>
17        </lido:objectClassificationWrap>
18        <lido:objectIdentificationWrap>
19          <lido:titleWrap>
20            <lido:titleSet>
21              <lido:appellationValue>The Parthenon</lido:appellationValue>
22            </lido:titleSet>
23          </lido:titleWrap>
24          <lido:objectDescriptionWrap>
25            <lido:objectDescriptionSet>
26              <lido:descriptiveNoteValue>The west facade of the Parthenon</lido:descriptiveNoteValue>
27            </lido:objectDescriptionSet>
28          </lido:objectDescriptionWrap>
29        </lido:objectIdentificationWrap>
30        <lido:eventWrap>
31          <lido:eventSet>
32            <lido:event>
33              <lido:eventType>
34                <lido:conceptID lido:type="URI">http://terminology.lido-schema.org/lido00012</lido:conceptID>
35              </lido:eventType>
36              <lido:eventActor>
37                <lido:actorInRole>
38                  <lido:actor>
39                    <lido:nameActorSet>
40                      <lido:appellationValue>Petros Katsaros</lido:appellationValue>
41                    </lido:nameActorSet>
42                  </lido:actor>
43                </lido:actorInRole>
44              </lido:eventActor>
45            </lido:event>
46          </lido:eventSet>
47        </lido:eventWrap>
48      </lido:descriptiveMetadata>
49    </lido:lido>
50  </lido:lidoWrap>
```


Europeana Preview

17


EDM **Europeana** [Show all](#)





© Ancient-Greece.org

© Rights Reserved - Free access

View item at
[IVML - Image Video and Multimedia Systems Laboratory](#)

 Share

 [Cite on Wikipedia](#)

 Translate details

The Parthenon by Petros Katsaros

Description: The west facade of the Parthenon

Subject:
<http://bib.arts.kuleuven.be/photoVocabulary/30808>

Creator: Petros Katsaros

Place: Athens, Greece

Date: 2013-09-14/

Type: Ancient PhotographyPhotography Ancient Photography
<http://bib.arts.kuleuven.be/photoVocabulary/22000>

Format: High Quality Paper 400x323

Identifier: 0851b

Rights: Ancient-Greece.org

Data provider: IVML - Image Video and Multimedia Systems Laboratory

Provider: Europeana Photography

Search also for:

Title
[The Parthenon by Petros Katsaros](#)

Who
[Petros Katsaros](#)

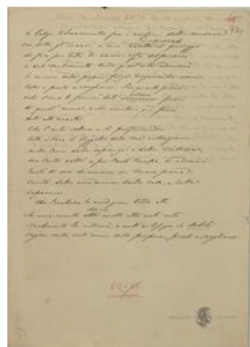
What
[Ancient PhotographyPhotography](#)
[Ancient Photography](#)

Provider
[IVML - Image Video and Multimedia Systems Laboratory](#)
[Europeana Photography](#)

Metadata Quality

18

□ Link to record



View item at
[Academy of Athens - Research Center for the Study of Modern Greek History / Ακαδημία Αθηνών - Κέντρο Ερεύνης της Ιστορίας του Νεωτέρου Ελληνισμού](#)



Share



Cite on Wikipedia



Translate details

Select language

Powered by Microsoft® Translator

Αυτόγραφα Σολωμού - Σατιρικά

Identifier:
local 100005 [Metadata]

Rights:
Academy of Athens; Ακαδημία Αθηνών

Source:
Academy of Athens - Research Center for the Study of Modern Greek History; Ακαδημία Αθηνών - Κέντρο Ερεύνης της Ιστορίας του Νεωτέρου Ελληνισμού

Data provider:
[Academy of Athens - Research Center for the Study of Modern Greek History / Ακαδημία Αθηνών - Κέντρο Ερεύνης της Ιστορίας του Νεωτέρου Ελληνισμού](#)

Provider:
Athena

Providing country:
Greece

Other items you may be interested in:



Αυτόγραφα Σολωμού -
Σατιρικά



Αυτόγραφα Σολωμού -
Σατιρικά διάφορα.



Αυτόγραφα Σολωμού -
Σατιρικά διάφορα

[Show all 156 items](#)

The shift from ESE to EDM

19

□ Metadata Quality

▣ Metadata reconciliation

- Connecting provider's collection-specific vocabularies to controlled vocabularies on the Web

▣ Metadata Annotation

- Some of the providers had metadata of limited expressivity for publishing on Europeana

Mappings – Thesaurus mapping

20

- Target schema elements can take values from SKOS vocabularies

Up to lido:subject

lido:subjectConcept

structural

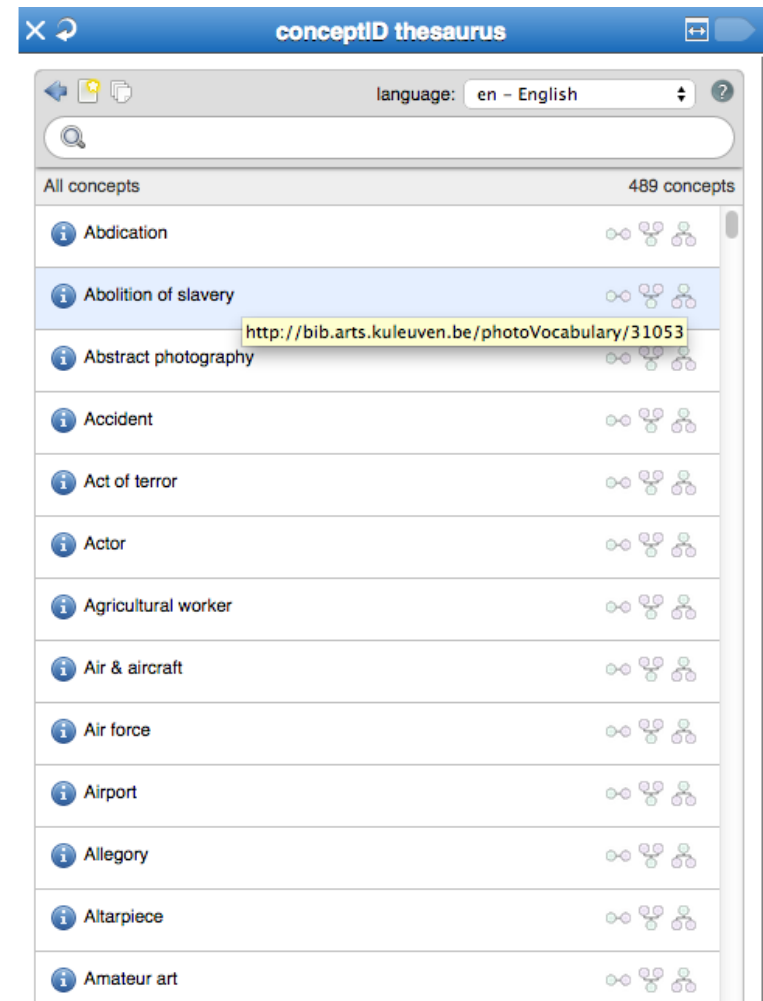
lido:conceptID

unmapped (thesaurus)

Mappings – Thesaurus mapping

21

- The user can
 - select a term from the thesaurus that will be applied to all items



Mappings – Thesaurus mapping

22

- The user can
 - align its in-house values to terms selected from the thesaurus

The screenshot shows a web application titled "conceptID value mappings". It features an input field, a search icon, and an "Add" button. Below the input field, it displays "Output: 13000" with a magnifying glass icon. A language dropdown menu is set to "en - English". A search bar contains the text "Photographic techniques". Below the search bar, a section titled "Narrower concepts of Photographic techniques" shows "3 concepts": "general", "negative", and "positive". Each concept has an information icon and a small diagram. At the bottom, a table shows the mapping of "Digital Camera" to a URL.

Input	Output
Digital Camera	http://bib.arts.kuleuven.be/pho...

Metadata Annotation

23

- Annotator
 - ▣ Create, delete and modify records

The screenshot shows the 'Annotator demo' web application. The interface includes a top navigation bar with tabs for 'My workspace', 'Dataset Options', and 'Annotator demo'. Below the navigation bar, there are buttons for 'Switch to Group Item Annotator', 'Create new item', 'Delete current item', 'Switch to complex', and 'Preview'. The main content area is divided into three tabs: 'Object Descriptive Metadata', 'Content Descriptive Metadata', and 'Administrative Metadata'. The 'Object Descriptive Metadata' tab is active, showing a list of metadata fields with their corresponding values and status indicators (red minus and green plus icons). The left sidebar contains a search bar and a list of items, with 'Intertoto Congress in Budapest' selected. The bottom of the sidebar shows pagination controls and a 'Show: 15 100 1000' option.

Field	Value	Status
Title	Intertoto kongresszust rendeznek Budapesten	✖
Series/Collection title	Retró híradók	✖
Title in english	Intertoto Congress in Budapest	✖
Series/Collection title in english	Retrospective newsreels	✖
Original Language	Hungarian	✖ ✔
Local keyword	kommunikáció	✖ ✔
Local keyword	piackutatás	✖ ✔
Local keyword	sportfogadás	✖ ✔

Metadata Annotation

24

- Group Edit
 - Tag records using elements of the schema

The screenshot displays the Metadata Annotation interface, which is divided into three main sections:

- Left Panel (Item List):** A table listing items with IDs (e.g., 00054, 00057, 00058, 00060, 00133, 00227, 00230, 00365, 00457, 00641, 00648, 00686, 00709, 00856, 00904). Each item has a checkbox and a 'Show' button. At the bottom, there are controls for 'Deselect all', 'Show: 15 100 1000', and a 'Next' button.
- Center Panel (Schema Tree):** A tree view showing the metadata schema. The root is 'rdf:RDF', which contains 'edm:ProvidedCHO'. Under 'edm:ProvidedCHO', there are several properties: 'dc:contributor', 'dc:creator', 'dc:description', 'dc:identifier', 'dc:type', 'edm:type', 'edm:localType', 'gr:color', 'skos:Concept', and 'skos:Concept'. The 'skos:Concept' property is selected, and its 'Apply Function' is set to 'Set constant value'. The 'New value' field contains 'http://thesaurus.europeana'. An 'Apply' button is visible below the 'New value' field.
- Right Panel (Group Actions Preview):** A section titled 'Group Annotator: Actions log'. It contains a list of actions performed, such as 'search for all in dataset / select 1000 items' and 'search for red in everything / select 132 items'. Below the list, there are 'Clear all' and 'Commit all' buttons.

...and more requirements...

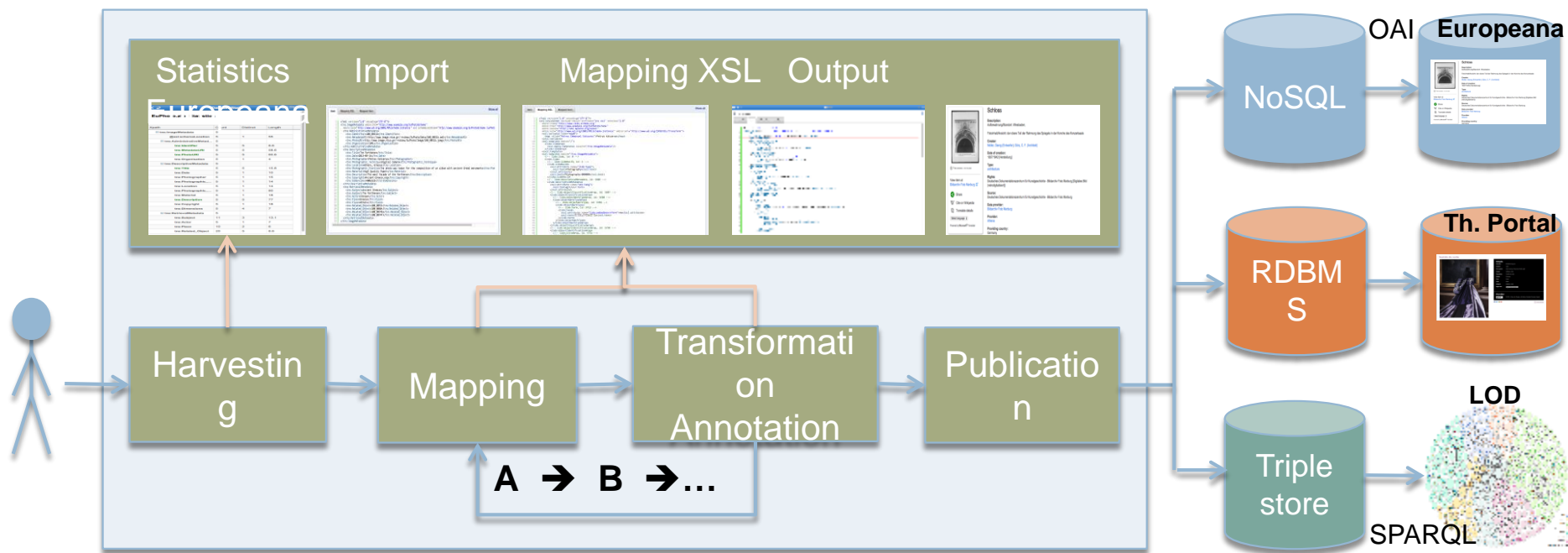
26

- Publication
 - ▣ Support of multiple publications
 - Europeana
 - Thematic portals
 - Linked Open Data

MINT – Current Workflow

27

MINT



Lesson learnt

28

- Metadata
 - ▣ The production of high quality metadata will always depend on the user. A tool can only provide useful services
- Never-ending requirements
 - ▣ Be one of them, find out what they'll need before they do
 - ▣ Be flexible by modularizing everything as much as possible

Future developments

29

- Cleaning and Linking
 - ▣ Developing functionalities based on semantic technologies for cleaning and linking metadata to existing Linked Open Data sources
- APIs for creative applications
 - ▣ Simplify access to MINT manipulated metadata for building applications, websites and mash-ups

Useful links

30

- MINT User Documentation
 - ▣ <http://mint-wordpress.image.ntua.gr/mint-end-user-documentation/>
- MINT on github
 - ▣ <https://github.com/mint-ntua>
 - ▣ http://mint.image.ntua.gr/redmine/projects/mint/wiki/Mint_Installation_instructions
- MINT Instance for AthenaPlus Museum
 - ▣ <http://mint-projects.image.ntua.gr/athenaplusmuseum>
- Europeana Fashion Thematic Portal
 - ▣ <http://www.europeanafashion.eu/>

Contact

31

Nikolaos Simou

National Technical University of Athens

Email: nsimou@image.ntua.gr

LinkedIn: <http://www.linkedin.com/pub/nikolaos-simou/68/31a/9aa>