MINT – METADATA INTEROPERABILITY SERVICES

Nikolaos Simou – National Technical University of Athens
What is MINT?

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

- **Metadata**
  - Data about data

- **Interoperability of metadata**
  - Transformation of heterogeneous metadata into one inter-operable metadata standard

- **Europeana**
  - [www.europeana.eu](http://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?

- MINT is an open source, web based platform for Metadata INTeroperability
  - [http://mint.image.ece.ntua.gr](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

- 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

- 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
Harvesting requirements

- 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.

- 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
Further requirements...

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

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petros Katsaros</td>
<td>Petros Emmanuel Katsaros</td>
</tr>
</tbody>
</table>
Mappings – Functional Mapping

- **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
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 Photography

Format: High Quality Paper 400x323
Identifier: 0851b
Rights: Ancient-Greece.org
Data provider: IVML - Image Video and Multimedia Systems Laboratory
Provider: Europeana Photography
Metadata Quality

- Link to record

**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 Ερευνής της Ιστορίας του Νεωτέρου Ελληνισμού

**Data provider:**
Academy of Athens - Research Center for the Study of Modern Greek History / Ακαδημία Αθηνών - Κέντρο Ερευνής 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

- 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
Target schema elements can take values from SKOS vocabularies.
Mappings – Thesaurus mapping

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

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

- Annotator
  - Create, delete and modify records
Metadata Annotation

- Group Edit
  - Tag records using elements of the schema
...and more requirements...

- Publication
  - Support of multiple publications
    - Europeana
    - Thematic portals
    - Linked Open Data
MINT – Current Workflow

MINT

Statistics → Import → Mapping XSL → Output

Harvesting → Mapping → Transformation Annotation → Publication

A → B → ...

NoSQL → OAI → Europeana
RDBMS → LOD → SPARQL
Th. Portal → Triple store
Lesson learnt

- 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

- 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

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

Nikolaos Simou

Email: nsimou@image.ntua.gr
LinkedIn: http://www.linkedin.com/pub/nikolaos-simou/68/31a/9aa

National Technical University of Athens