



# Cultural Heritage Aggregation in Europe

*Nasos Drosopoulos*  
*National Technical University of Athens*

# Europeana Content Provider



- ❑ Any organisation that provides digital content that is accessible via Europeana. Europeana only ingests and indexes the institution's metadata while the digital object remains at the original institution or aggregator.
- ❑ The amount and type of content, the technical infrastructure, the output formats and the documentation available can vary significantly among existing and potential content providers.

# An Aggregator in the Europeana context



collects metadata from a group of content providers and transmits them to Europeana.

## Aggregators

- gather material from individual organisations,
- standardise the file formats and metadata,
- channel the latter into Europeana according to the Europeana guidelines and procedures and,
- support the content providers with administration, operations and training.

*source: Europeana Aggregators' Handbook*

# Scope



- ❑ An aggregator may or may not have a portal where the content is made accessible to the public. If the aggregator's portal is not accessible to the public, it is referred to as a 'dark portal' or a 'dark aggregator.'
- ❑ An aggregator can act as a repository, storing the digital items. Alternatively, it can act as an intermediary, only collecting metadata with a link to the digital item.

# Council of Content Providers and Aggregators - CCPA



CCPA communicates the value of providing content:

- Improving users' experience by integrating all types of content
- Enriching the content by displaying it alongside related material from other countries, other domains
- Knowledge transfer between different institutions, domains and countries
- Sharing best practice and common standards
- Being part of an award-winning, highly visible portal that is the focus of political attention
- Demonstrating the relevance of cultural and scientific heritage institutions to new generations of users

# Aggregator Types



- ❑ Single (sector, i.e. museums, libraries, archives etc.)
- ❑ Cross-domain (e.g. national portals, Europeana)
- ❑ Thematic (e.g. Judaica)
- ❑ Geographic (e.g. for Europe, The European Library)

Additional benefits depend on strategy

# Technical aspects of aggregation



- ❑ Interoperability is achieved by using a common data model, rich enough to meet the requirements.
- ❑ Providers are asked either to export metadata in the agreed format or use the provided tools to transform/map their metadata.
- ❑ Quality control procedures monitor and validate the aggregation

# Aggregation

## 1. Ingestion



*Gather material from individual organisations.*

- ❑ Typically refers to uploading metadata (and sometimes the content or previews) through the http and ftp protocols.
- ❑ XML is the prevalent metadata serialization. In practice, spreadsheets and tabular formats (e.g. csv) are still in use.
- ❑ Harvesting is a term associated with the OAI-PMH protocol.



# Linked Heritage Aggregator

http://mint-projects.image.ntua.gr/linkedheritage

A screenshot of the "Import" page in the Linked Heritage Aggregator. The page has a dark navigation bar with "Home", "My Profile", "Import", "Overview", and "Logout". The "Import" section is titled "Import" and contains several options for data import. The "Http Upload" option is selected, showing a "Choose File" button and "No file chosen" text. Below it is a checkbox for "This is a CSV upload". Other options include "NTUA FTP Upload" with a file selection dropdown, "Remote FTP/HTTP Upload" with a URL input field, "OAI URL" with a URL input field and a "check oai url" button, and "From Date" and "To Date" input fields. There are also input fields for "OAI SET:" and "Namespace Prefix:" with "fetch OAI sets" and "fetch OAI namespaces" buttons. At the bottom, there is a checkbox for "This import conforms to" with a "Select schema" dropdown and a "schema" label. A note below states: "Select this option in addition to your import method in case your upload already conforms to the selected schema and no mapping is necessary." The page ends with "Submit" and "Reset" buttons.

# LH Aggregator Overview






Home My Profile Import Overview Logout

## Overview






An overview of all the imports and items per organization and per uploader:


### Imports

Organization:    

Imports by user:

View all available actions by **clicking on an import name**



 <b>phaistos.xml</b> 854, 26/09/2011 16:17	   
--	--

 Delete selected

### Items

( NTUA )

Filter by:

Name	Created
  Phaistos Disk	<i>phaistos.xml</i> 26/09/2011 16:17

Displaying items 1 - 1 of 1

Jump to page

# Aggregation

## 2. Mapping



Standardise the file formats and metadata.

- ❑ Metadata modeling varies in aggregators, mostly depending on domain and type of content.
- ❑ Content providers offer metadata records in a variety of models and/or data structures.
- ❑ Mapping refers to establishing crosswalks between, typically, a provider's metadata records and the aggregation's metadata model.
- ❑ Mapping has to be implemented in order to allow the transformation of records

# LH Aggregator Mapping Editor



The screenshot shows the LH Aggregator Mapping Editor interface. It is divided into three main vertical sections: Source Schema, Mappings, and Target Schema. The Source Schema on the left lists a 'record' with properties: title, date, date2, desc, src, rights, and img. The Mappings section in the center shows a tree structure starting with 'titleWrap:' containing 'titleSet:' (structural), which has 'appellationValue:' (mapped to 'title') and 'sourceAppellation:' (unmapped). Below these are 'inscriptionsWrap:', 'repositoryWrap:', 'displayStateEditionWrap:', 'objectDescriptionWrap:', and 'objectMeasurementsWrap:'. The Target Schema on the right shows a 'Template' button and two categories of metadata: 'Descriptive Metadata' (with 'Object Identification' selected) and 'Administrative Metadata' (with 'Record' selected).

# Aggregation

## 3. Publish



Channel metadata into Europeana according to the Europeana guidelines and procedures

- ❑ Metadata (descriptive, administrative) describing a digital object. The metadata must be mapped to the ESE v3.4 (soon to be replaced by EDM).
- ❑ An online preview or thumbnail of the described object.
- ❑ Persistent identifiers - active and stable links to the described digital object on the provider's site or the portal's site.

# Aggregation

## 3. Publish



- Transform metadata records from the aggregation's data model to Europeana's
- Setup an OAI-PMH repository and publish records
- Support harvesting by Europeana
- If not 'dark', publication refers to the aggregator's portal and even to the content itself (e.g. national aggregators, a/v domain etc)

# Aggregation Support



Support the content providers with administration, operations and training

- ❑ The 3 steps mentioned are rarely a one-off procedure
- ❑ It is important to update the aggregation steps according to evolutions in modeling or introduction of new services
- ❑ Depending on the domain and nature, an aggregator is meant to be sustainable

# Aggregation Support



- ❑ New provider contributions reuse existing mappings.
- ❑ New or extended provider data model introduces maintenance and update of mappings.
- ❑ Aggregation data model evolution can be supported by updating mappings
- ❑ Europeana or aggregator portal developments can be tackled by maintaining the mapping between the aggregation and respective services



# Reinforcing aggregation



The adoption of a domain data model is the first step towards the desired level of homogeneity, which is affected by factors such as

- Fullness,
- Language,
- Formatting,
- Naming conventions, etc.

Controlled vocabularies, or even ad-hoc terminologies, can play an important role for the aggregation's quality

# Aggregation Challenges



Providers continue working on existing and new resources towards:

- ❑ Persistent and unique identification; de-duplication, link maintenance
- ❑ Semantic data modelling (EDM and refinements)
- ❑ Resource identification and management
- ❑ User validated reconciliation with linked vocabularies

# Aggregation Challenges



Providers continue working on existing and new resources towards:

- Term/entity extraction
- Moderated resource linking within the repository & Europeana
- Linked Data consumption
- Linked Data deployment



Thank you for your attention  
[ndroso@image.ntua.gr](mailto:ndroso@image.ntua.gr)