
Europeana en de digitale ontsluiting van cultureel erfgoed
Europeana et l'accessibilité numérique du patrimoine culturel
2009-12-16



Workshop 3 Aggregation



Mel Collier – Jef Malliet



What is aggregation?

- General: (e.g. geology, engineering)
 - Assemblage constitutes a new unity
 - Components/materials that do not react with each other
- In ICT:
 - Assemblage of documents that forms a new document; e.g. for copyright purposes
 - Collection of articles from various sources, presented together on a website
 - E.g.:

From en.wikipedia.org

Aggregator: In general internet terms, a news aggregation website is a website where headlines are collected, usually manually, by the website owner.

Aggregation in Europeana (1)

Definitions

- Definition: (from Europeana Content Strategy)

An Aggregator is an organization that collects metadata from its group of content providers and transmits them to Europeana, helps content providers with guidance on conformance with Europeana norms and converts metadata if necessary. The aggregator also supports the content providers with administration, operations and training.

A Content Provider is any organization that provides digital content for access via Europeana and the metadata that enables the access.

- CCPA: Council of Content Providers and Aggregators
 - Content providers and aggregators participate in Europeana decision making

Aggregation in Europeana (2)

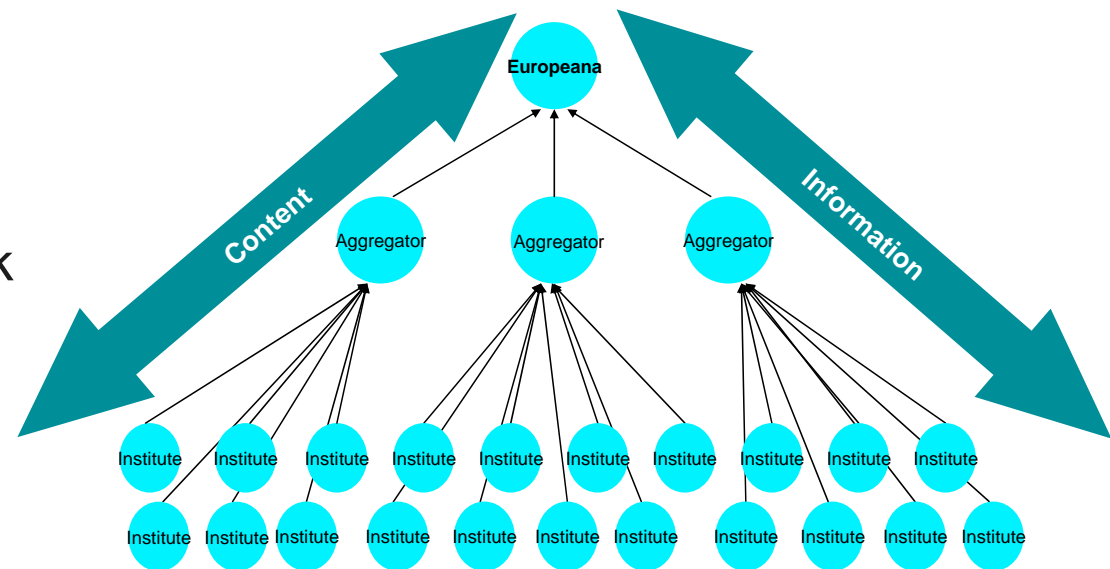
Options

- Repository:
 - Including digital objects
 - Only metadata and indexes
- Portal:
 - With public interface
 - Only database - 'dark portal'
- Type of aggregation:
 - Vertical *single domain, accross administrative/geographic borders*
 - Horizontal *cross domain, within administrative/geographic borders*
 - Thematic *cross domain, no borders*

Aggregation in Europeana (3)

Role of aggregators

- Disseminate vision and objectives of Europeana through their network
- Provide feedback to Europeana from their network
- Promotion and implementation of standards
- Provide domain specific expertise and skills to institutions and Europeana



Content for Europeana (1)

Organisational issues

- Europeana prefers collecting content through aggregators
- Each content provider to contribute through one aggregator only
- Flowchart describes decision path for choosing the best way to contribute for new content providers
- Content provider/aggregator responsible for delivering data in accepted format (currently ESE 3.2.1 specifications)
- Content provider/aggregator to make data available for harvesting by Europeana through OAI-PMH protocol

Content for Europeana (2)

Current Status

- ESE specifications valid for Rhine release (Summer 2010)
- No specifications yet for Danube release (2011)
- No specifications yet for persistent identifiers
- No decisions yet concerning re-harvesting for data updates
- Many unsolved IPR issues
- Very few existing aggregators
- Handbook with guidelines is being prepared

Businessmodels for aggregators

- Collect data for Europeana
- Give access to national/regional/local heritage information
- Promote national/regional/local profile or identity
- Provide resource for educational/tourist services
- Reinforce relevance of heritage institutions
- Assist and support heritage managers for digitization
- Keep digitized cultural assets in the public domain
- Increase access to knowledge about cultural heritage
- Underpin the knowledge economy



Interoperability

Interoperability is the key in the essence, purpose and construction of an aggregator:

- Technical interoperability – Exchange Protocols
- Structural interoperability – Datastructures
- Content interoperability – Semantics, Languages

Technical interoperability Protocols - Harvesting

Aggregator must be able to acquire (ingest) and interpret the source data

- Internet
- XML
- OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)

Structural interoperability

Data Structures - Mapping & Normalisation

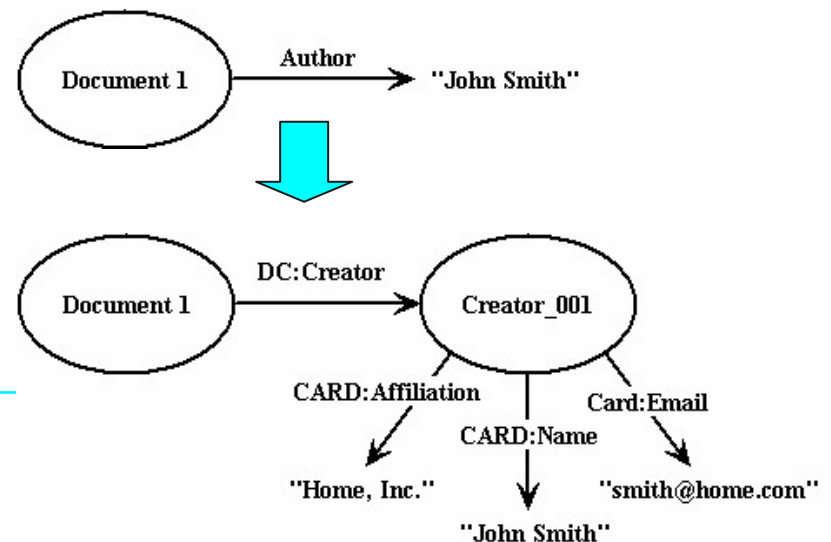
- Aggregator uses its own uniform datastructure, designed according to the purpose of the aggregator
 - Simple: e.g. Dublin Core, ESE (cross-domain aggregators)
 - Detailed: e.g. MARC, Spectrum (specific domain, vertical aggregators)
 - Semantic (for Semantic Web)
- Source data must be converted to aggregator datastructure
 - Mapping fields
 - Normalization
 - THE condition: internal **Consistency**

Content interoperability (1)

Semantics - Enrichment

Many aggregators aim at a semantic datamodel

- New information created from combination with other sources
 - Relations between objects/concepts
 - New, broader context emerges
- Requires Semantic Web technology:
 - Resources
 - Identification through URI
 - Relations with RDF



Content Interoperability (2)

Thesauri

SKOS

- Evolution from ISO standards for thesauri (ISO2788 & 5964)
- W3C specifications
- Semantic Web technology (RDF), object oriented
- New concept approach to thesaurus: taxonomy of concepts rather than terms
- Terms are identifiers for the concept
- Concepts from one thesaurus can be connected to concepts from another thesaurus: tool for merging thesauri

Content Interoperability (3)

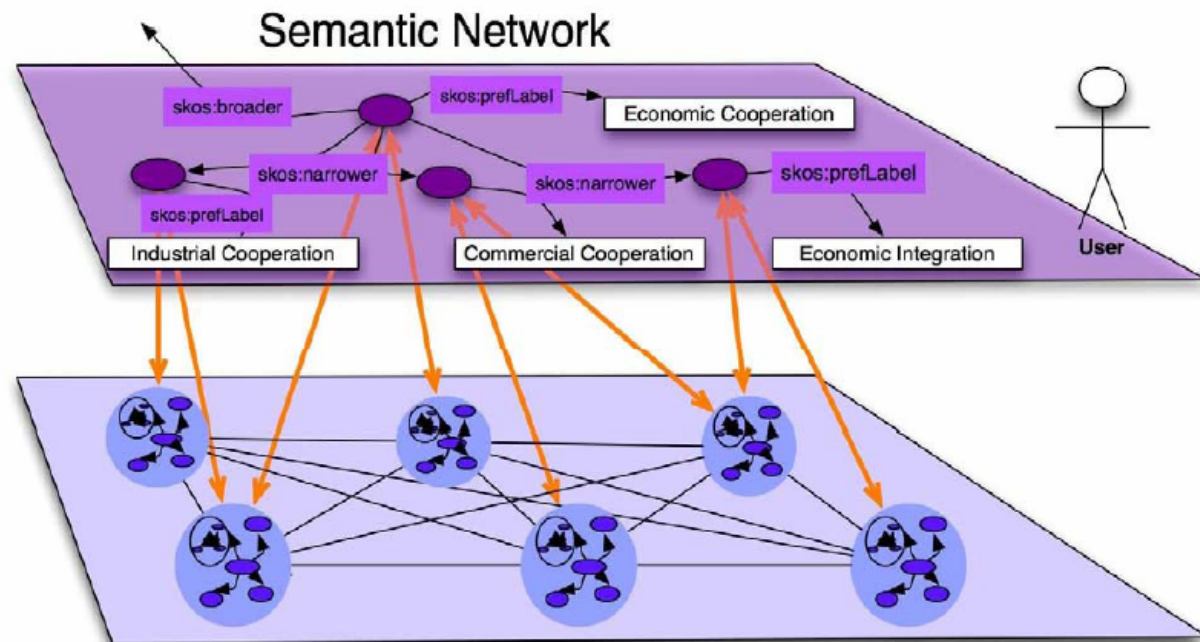
Thesaurus e.g. AAT

- AAT is concept based, perfect fit with SKOS datamodel
- Concepts are well defined by Scope Notes
- Multilingual: English, Spanish, Dutch, (French), (German), (Chinese)
- Must become more dynamic
- Content is responsibility of the users
- Improving/maintaining AAT must be done by the heritage sector

Content Interoperability (4)

Semantics in Europeana

- Collecting thesauri used in source databases
- SKOS format
- Multilingual
- Building blocks for semantic layer in the data model
- For Danube release (model and actions to be finalized)



Existing Aggregators

- National aggregators:
 - Austria: www.kulturpool.at
 - France: www.culture.fr
 - Germany: www.bam-portal.de
 - Italy: www.culturaitalia.it
- Aggregators in Belgium:
 - Vlaamse Kunstcollectie: www.vlaamsekunstcollectie.be
 - Religieus Erfgoed, CRKC: www.religieuserfgoed.be
 - MovE, Oost-Vlaanderen: www.museuminzicht.be
 - Erfgoedplus.be, Limburg&Vlaams-Brabant:
www.erfgoedplus.be

Participating in Europeana Conditions

- Content suitable for Europeana
 - Metadata about digital objects (text, image, audio, video)
 - Sufficient metadata
 - Metadata convertible to ESE
- Copyright cleared
- Digital objects can be accessed directly through URL
- Preferably contribute through a suitable aggregator
 - Museums: Athena
 - Oost-Vlaanderen & Limburg: EuropeanaLocal through MovE or Erfgoedplus.be
 - Other Europeana cluster projects: see www.group.europeana.eu

Preparing for Europeana (1)

Checklist

- Systems
 - Can the data be exported in XML format?
 - Can authority files be used for controlling content of relevant fields?
- Datastructures
 - What standard was used?
 - How consistently has the standard been applied?
 - Were departures from the rules documented?
 - Can the data be understood outside the original context?

Preparing for Europeana (2)

Checklist

- Thesauri

- Which thesauri were used?
- Were additions/modifications made?
- Were the deviations properly documented?
- Are the thesaurus terms understandable outside the original context?

- Images

- Are they stored in a clear file structure?
- Which rules were used and were they followed consistently?
- Which file formats have been used?

Thank you @

- Mel Collier – mel.collier@bib.kuleuven.be
- Jef Malliet – jmalliet@limburg.be