Networking Session: Global Information Structures for Science & Cultural Heritage - The Interoperability Challenge

«INTEROPERABILITY FROM THE CULTURAL HERITAGE PERSPECTIVE»

STEFANOS KOLLIAS

DC-NET Project

ECE School, National Technical University of Athens

ICT, Brussels, September 29, 2010

Contents

- Introduction – Interoperability in CH
- Semantic Interoperability
- The Athena Ingestion Tool
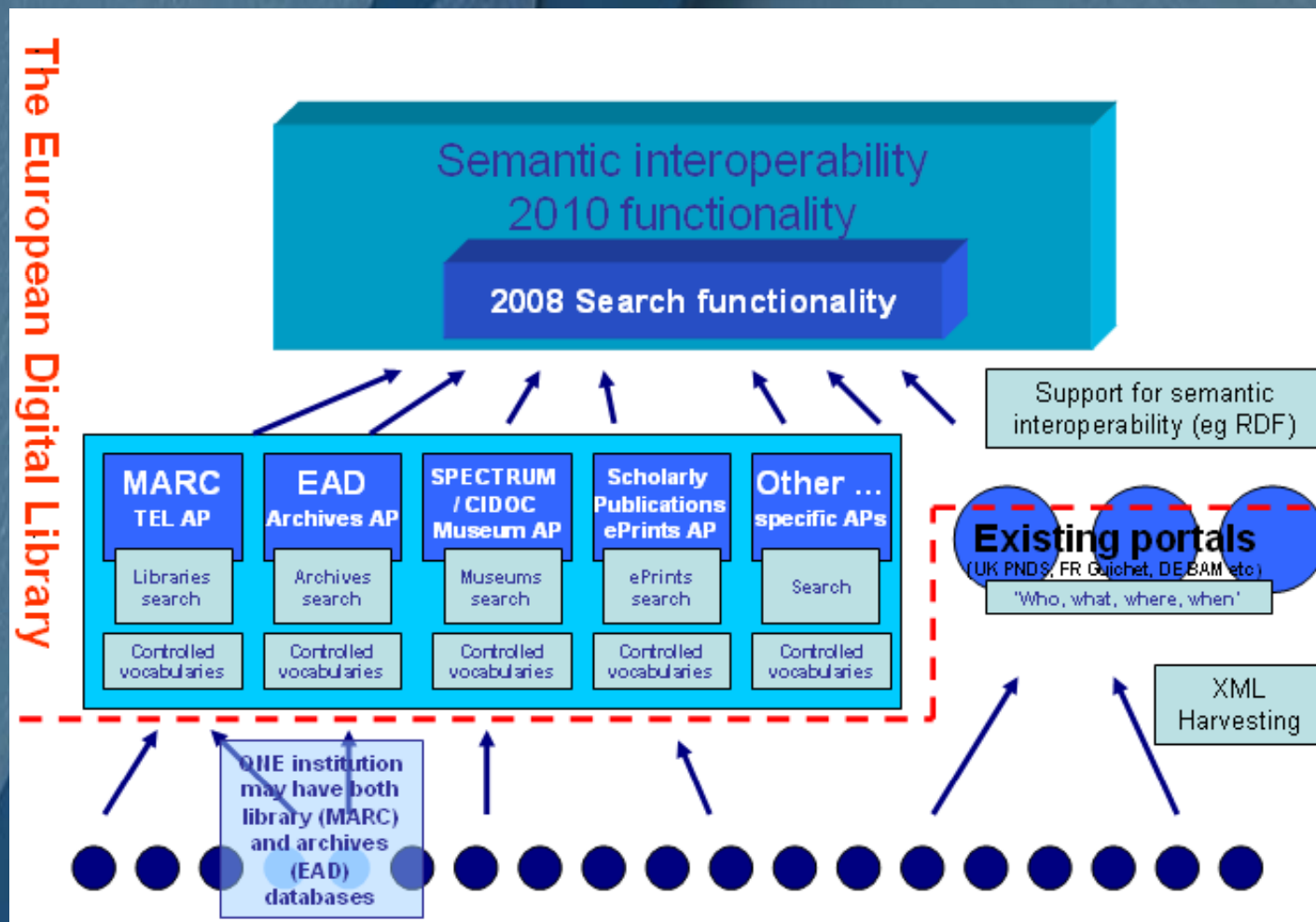- Linked Data in CH
- E-Infrastructures
- Conclusions

Introduction

- Organisation of Content to ensure effective access to it is of major importance to European stakeholders.
- In 2007 the European Interoperability Group was set up by the EC, where a roadmap for the relevant working issues of the cultural heritage sector was defined.
- In particular, a framework for metadata & semantic interoperability (including thesauri transformation to SKOS, metadata enrichment and knowledge alignment) has been adopted by the Group for short and long term CH development.
- Specific architectures and implementations of semantic interoperability are devised in the framework of Europeana v1.0/v2.0 and Europeana Connect.
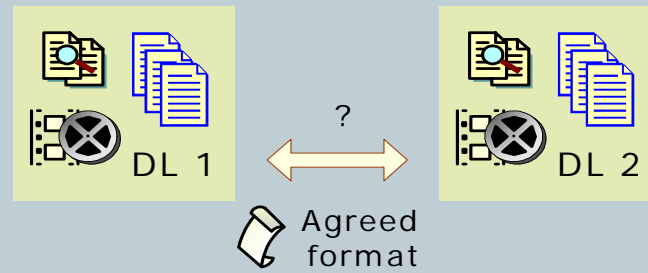
Interoperability in Cultural Heritage

- Aggregation and Effective Access to European Content is one of the main goals of CH developments today .
- In the 2010 Rhine and 2011 Danube Milestones more than 12.000.000 cultural objects from all over Europe will be accessible through the EUROPEANA access point.
- Aggregation and Interoperability of content metadata through appropriate data models (ESE, EDM) is the means to achieve this goal.
- In the 2011-2013 period semantic access and use of linked open data will be key issues in the developments.
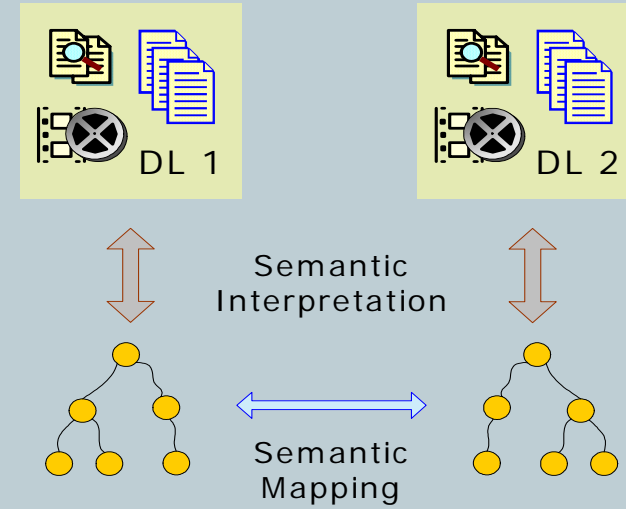
The Interoperability Framework
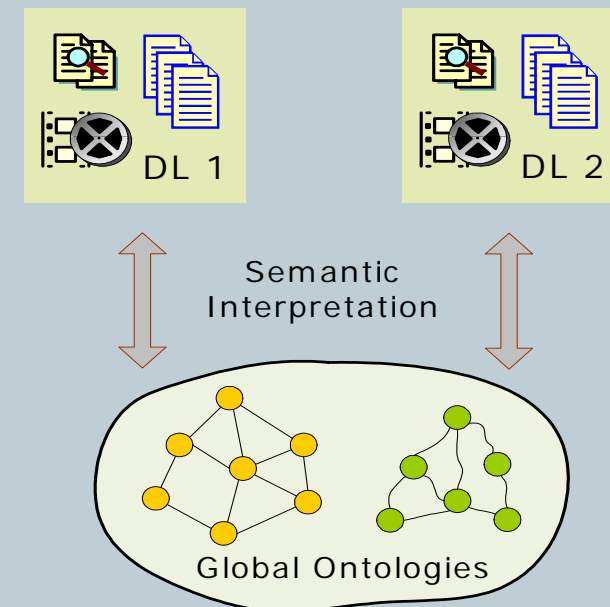
Figure 1. Achieving Semantic Interoperability

Principle 1: Semantic Annotation

- Description of objects with concepts of a shared vocabulary

Original Abstract
Interview of the painter Salvador Da
Owner

Date created
1962-01-19
Languages
French
...
...
...

Principle 2: Semantic Linking

- Search for objects which are linked via concepts (semantic link)
- Use the type of semantic link to provide meaningful presentation of the search results.
    - Create a link between the 2 thesauri
- Thesauri/Vocabularies
    - TGN for Geographic Names
    - ULAN for famous artists
    - IPTC for News, Sports etc.
    - Other thesauri and/or vocabulary

Semantic Organisation of Content

- Content Aggregation & Enrichment are major technical issues. Content belongs not to independent sources, but to linked knowledge that can be semantically accessed in its whole.

- Users can share & re-use content aggregations in the own context and different scenarios.

- Content at source or mediator (e.g. aggregator) level includes metadata, but may also include, after settling IPR issues, multimedia information.

Interoperability: The Ingestion Tool

- NTUA has developed an Ingestion Tool (in the framework of ATHENA), with which content of the museum area is mapped, under user control – to the LIDO (and then to the ESE) data model.

- This is used by ATHENA for content ingestion to Europeana. It is being adapted and used also in the CARRARE (archaeology) and EUSCREEN (TV, audiovisual) projects. Projects APENet, (Archives), Judaica have tested its usage for their content ingestion as well.

- Current extensions include its use in Europeana EDM Data Model, Semantic Enrichment with Linked Data (LINKED HERITAGE project).

Architecture & Phases

```
┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
│ Harvesting,  │→  │   Import,    │→  │  Versioning  │→  │  Semantic    │→  │    Value     │→  │    Enrich    │→  │   Export     │
│  Delivery    │   │  Parsing,    │   │              │   │   Mapping    │   │   Mapping,   │   │              │   │   Publish    │
│              │   │  Indexing    │   │              │   │              │   │ Normalization│   │              │   │              │
└──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘
       ↓                  ↓                  ↓                  ↓                  ↓                  ↓
 ( Delivered )     ( Structured )      (    Data    )    ( Interoperable )  ( Interoperable )   ( Europeana  )
 (   Data    )     (  Indexed   )      (with reapplied)  (   Semantics   )  (   Structure   )   (  Enabled   )
                   (    Data    )      (  enrichment  )                                          (    Data    )
```

Presentation & Search

Analysis & Statistics

Quality Control
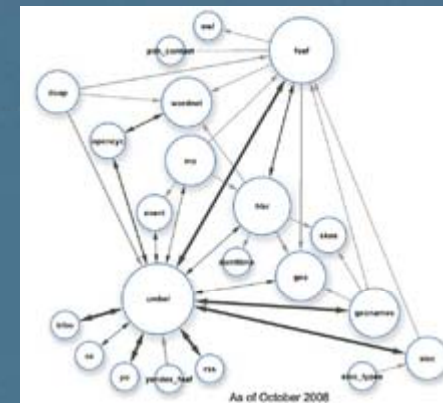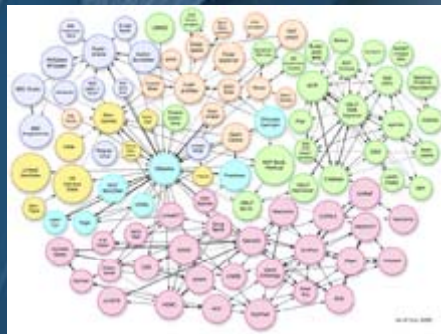
← **Re-iterate**

Semantic Mapping & Enrichment

- Providers manually map fields to Athena schema (i.e. LIDO). The mapping transformations will be stored and the provider will be able to edit them at any point if necessary.

- Support event type schema mapping.

- Designed visually to enable mapping of large schemas.

- Enrichment can be manual or automatic.

- Manual: will enable the addition of data that is not in the original metadata (e.g empty fields, multilingual information, fields that take values from controlled vocabularies).

- Automatic: will enrich and align the existing metadata and vocabularies with external web resources and thesauri (e.g. Dbpedia, GeoNames)

Linking Open Data

- The goal of the W3C Semantic Web Education and Outreach group's Linking Open Data community project is to extend the Web with a data commons by publishing various open datasets as RDF on the Web and by setting RDF links between data items from different data sources. In 2007, datasets consisted of over two billion RDF triples, which were interlinked by over two million RDF links. In 2009 this had grown to 4.2 billion RDF triples, interlinked by around 142 million RDF links.

Linked Data in Cultural Heritage

1. Data is rich, and includes Data created by Users

Cultural data includes all forms of digital asset, including user generated content. Openness must be promoted as widely as possible.

2. Linked Data connects the culture sector to a collaborative effort across the Public Sector

Linked Data also enables the sector to benefit from investment and innovation in other industry sectors.

3. Linked Data can help achieve more efficient practice

Providing and consuming Linked Data can reduce inefficiency, and add value to existing content creation.

Linked Data in Cultural Heritage (2)

4. Linked Data is the next phase in adaptation to the Web

Next logical step in the evolution of cultural services from curation to digitisation to delivery of rich, integrated & personal online services.

5. Commit to commissioning Open Data, not Open Source

Open Source systems do not guarantee that the data they contain will be openly available. The Cultural Sector must ensure that all funding programmes, commissioning and procurement include a requirement to make publicly-funded data openly-available using open and established industry standards.

CH e-Infrastructures

Projects DC-NET, INDICATE :

- Investigate and prepare the migration of existing cultural heritage applications to e-infrastructures platforms
- Evaluate and report on requirements, results and subsequent open issues
- Design and prototype the e-culture science gateway (eCSG)

Digital Cultural Heritage in the future

- Distributed data sets and services (high performance, reliability and scalability)
- Semantic web repositories and services (data expressivity & interoperability)
- Automatic data and service discovery (transparency and agility)
- Security (Web of trust)

In this framework

➢ Semantic metadata and interoperability

➢Metadata Enrichment with shared knowledge (linked open data)

➢Multimedia content analysis and linking

➢Collective intelligence integration

are major issues under examination and implementation in the current/future development of the Cultural Heritage Field

THANK YOU