
SEMANTIC AGGREGATION OF CULTURAL CONTENT

STEFANOS KOLLIAS

COMPUTER SCIENCE DEPARTMENT
NATIONAL TECHNICAL UNIVERSITY OF ATHENS

ATHENA Final Conference
Rome, April 28, 2011



Introduction

ATHENA: THE INGESTION TO EUROPEANA MECHANISM

- Aggregating cultural content for the museum sector and uploading this to Europeana has been one of the main targets of the ATHENA project.
- Based on the adoption of the LIDO metadata schema for harmonising content aggregation obtained from the large number of partners & associated content providers, an ingestion tool has been designed, implemented and evaluated during aggregation of more than 4 Million items.
- The tool has been used to map partner metadata schemas to LIDO, as intermediate step for interoperability and then (from the richer LIDO model) to the Europeana ESE model.



Introduction (2)

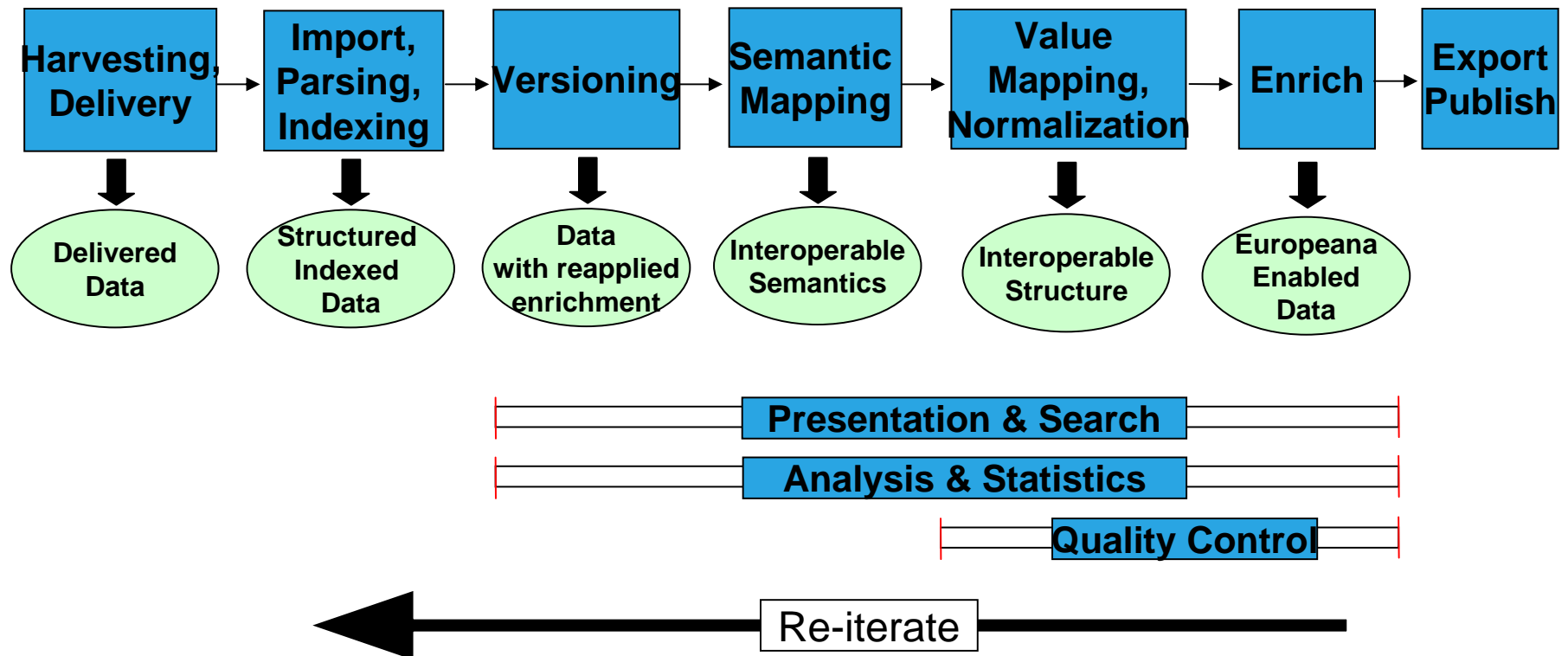
MOVING FURTHER AHEAD:

- Following this procedure it has become possible to start mapping the metadata aggregated in the LIDO format to the evolving Europeana EDM data model, as well as to semantic web RDF representation.
- Based on these representations it is possible to move towards inferencing on networked cultural content, through usage of reasoning, linked open data and query answering methodologies.
- Achieving effectiveness and efficiency in this framework constitutes a great challenge for the next developments in the field.

ATHENA and the Ingestion Mechanism

- The aim of the ingestion mechanism has been to collect and prepare providers' metadata for uploading to EUROPEANA.
- The involvement of content providers has been crucial in achieving this objective.
- Content providers are using their own proprietary metadata standards.
- OAI-PMH technology is preferred to harvest the providers' metadata.
- Syntactic and Semantic interoperability are achieved by using the LIDO schema, which is rich enough to cover the needs of the museum sector.

The Workflow



Semantic Mapping

- Manually map Providers fields to the LIDO Athena schema.
- Mapping transformations are stored and the provider can edit them later.
- Support value concatenation (many-to-one mappings), event type schema mapping, functions for data transformation, data preview.
- Visual design so as to enable mapping of large schemas.

Enrichment

- 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)

Search, Statistics & Europeana Preview

- This module provides the ability to perform full-text search in provider's data.
- Automatically checks and reports on Content Provider's data (i.e. missing values, malformed data).
- Error reports & warnings can be used to edit mappings until providers' data successfully pass Quality control checks.

Europeana Preview



[My Europeana](#) [About us](#) [Communities](#) [Partners](#) [Thought lab](#) [Choose a language](#)

[Advanced search](#)

Matches for: acropolis

Item details

[Return to results](#)

[SHARE](#)



 **Acropolis**

Title: Acropolis

Date: 0000-00-00 00:00:00; BC

Creator: Robert Gordon University

Description: Ancient Greece, Athens, Acropolis

Language: en

Format: image/jpeg

Rights: Robert Gordon University

Provider: Scran ; Uk

[More](#)

[View in original context](#)
Opens in a new window

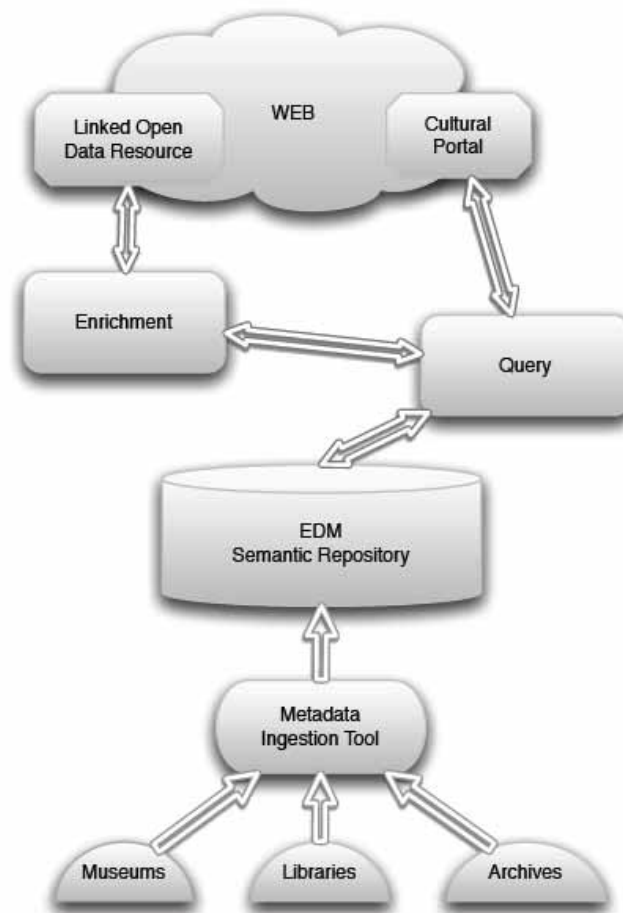
[Using Europeana](#) [Accessibility](#) [Sitemap](#) [Terms and conditions](#) [Privacy](#) [Language policy](#) [Contacts](#) | [Send us feedback](#)

co-funded by the European Union 



ATHENA
Access to cultural heritage
networks across Europe

Semantic Cultural Aggregation & Inferencing



Semantic Cultural Content Access

- Cultural content descriptions are usually incomplete.
- Perform content metadata enrichment by letting content providers and users execute specific searches and queries.
- Expressing sophisticated users' queries (*e.g., provide me with all items in Europeana that have been used for storage of olive oil*) in terms of the metadata schema is a major restriction for data integration
- Query answering should scale to huge amounts of data.

Semantic Cultural Content Access (2)

How to proceed:

- Cultural content metadata schemas are mapped to axiomatically expressed terminological knowledge
- Semantic queries are expressed in terms of the terminology
- Semantic query answering is based on reasoning (implicit information should be captured and used for this).

Linked Open Data

- Linked data is the prevalent initiative towards semantic resource linking and discovery, fully endorsed by Europeana and a growing number of Cultural Heritage aggregators and providers worldwide.
- There is a vast amount of published linked data sources that can be exploited for the annotation, enrichment and reconciliation of the cultural content.

Query Answering technologies (1)

Technological issues and restrictions:

- Description Logics, underpinning the W3C Web Ontology Language (OWL), is the standard framework for representing formal terminologies (ontologies) on the web.
- Conjunctive queries (conjunctions of atoms), formalised in SPARQL or SQL, is the standard framework for accessing web data.

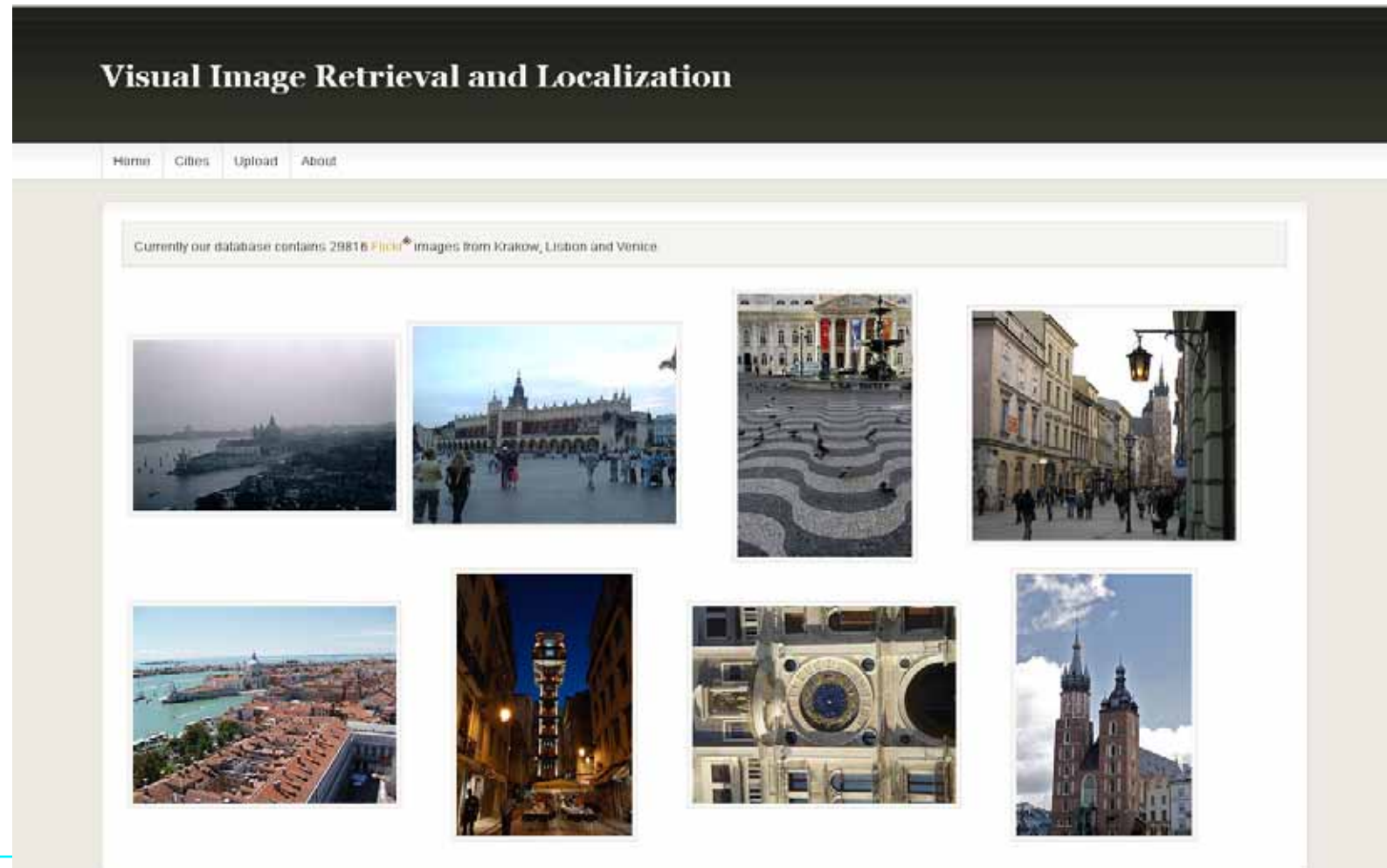
Query Answering technologies (2)

- The complexity of conjunctive query answering over ontologies strongly depends on the expressivity of the ontology representation language (scalable only for the tractable fragments of OWL).
- Applying query answering methods in association with Linked Open Data facilitates scalability.
- Query answering based on the Query Rewriting methodology constitutes a promising approach for achieving this goal.

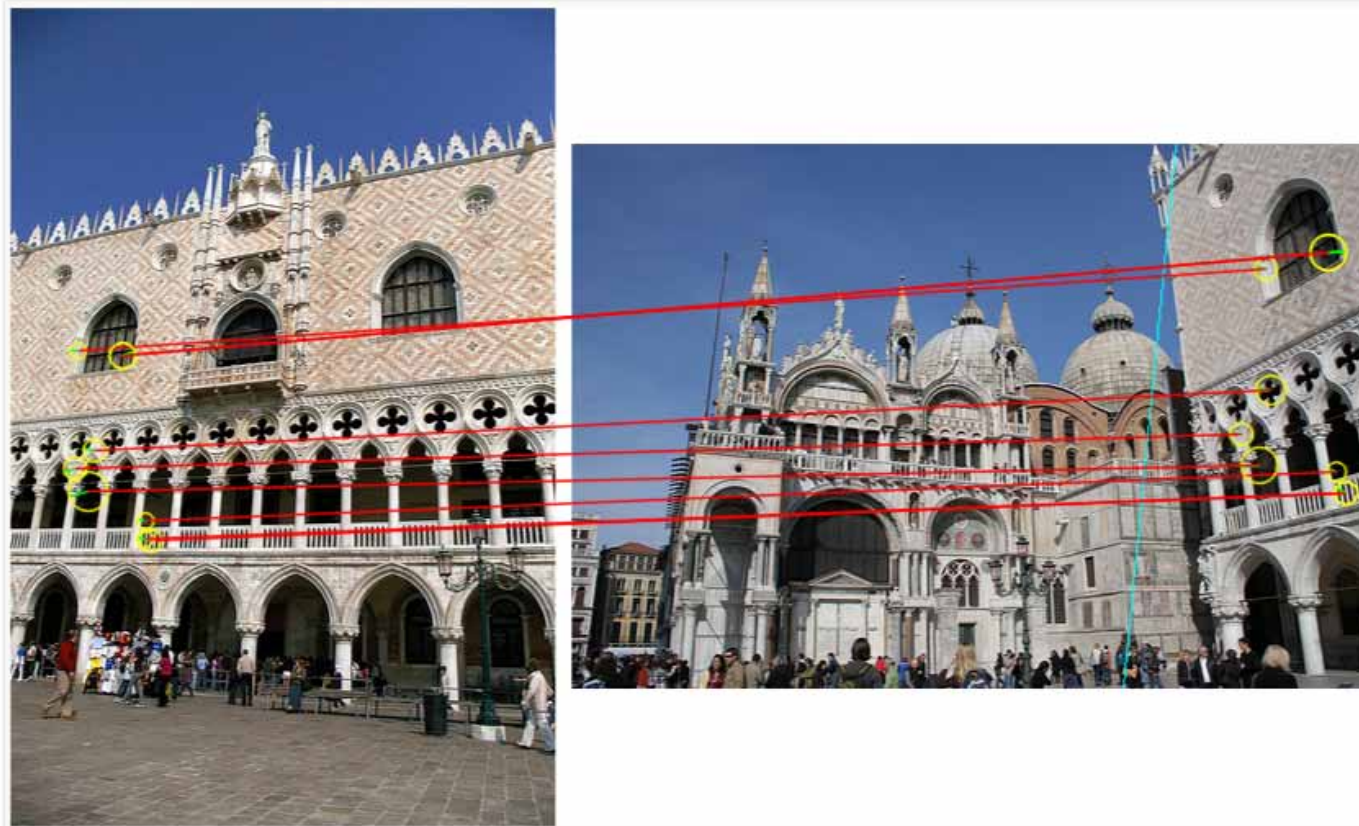
Visual Metadata (Icons) Search

- Enhance Content Enrichment through Visual Similarity Matching.
- Operate either on Europeana Items' Icons, or on Aggregators' Items' Images.
- Correlate metadata search with visual matching and indexing.
- Provide affine invariant image search, matching and retrieval.

Visual Image Retrieval & Analysis (VIRAL)



Similarity Content Selection (VIRAL tool)



Thank You!

stefanos@cs.ntua.gr

