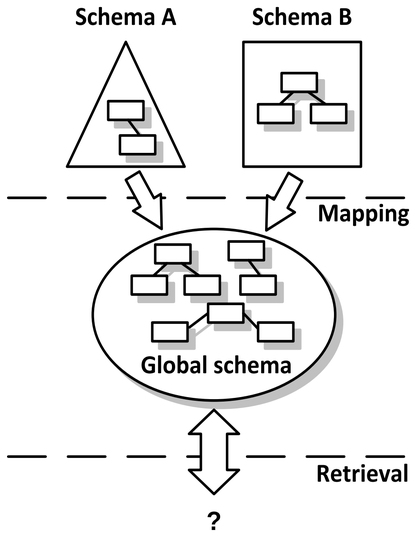
CIDOC CRM in Action
– Experiences and Challenges –

Philipp Nußbaumer – Research Studios, Studio Digital Memory Engineering, Vienna, Austria
Bernhard Haslhofer – University of Vienna, Departement of Distributed and Multimedia Systems, Vienna, Austria

1 INTRODUCTION

Integration of metadata from heterogeneous sources is a major issue when connecting cultural institutions to digital library networks. Uniform access to metadata is impeded by the structural and semantic heterogeneities of the metadata and metadata schemes used in the source systems.

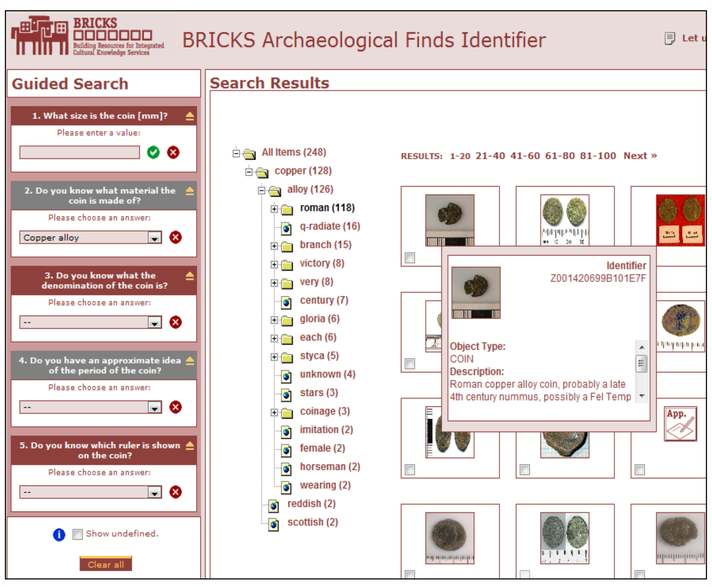


Within the context of the BRICKS Project [1] we have integrated metadata and content from a number of archaeological institutions. To provide interoperability, we have chosen to use the CIDOC Conceptual Reference Model (CIDOC CRM) [2]. Its central idea is to map each proprietary metadata scheme to a global ontology which is tailored to the cultural heritage domain.

2 APPLICATION OVERVIEW

The **Archaeological Sites Finds Identifier** application is a tool for expert users and non-professionals to identify findings made all over Europe. In our prototype implementation, developed in course of the BRICKS project, the integrated findings are restricted to coins found in the United Kingdom.
A user may explore the reference collections in different ways to identify a finding:

- *Browse*: helps a user to get an understanding of the reference items' properties
- *Simple Search*: full-text search
- *Guided Search*: guides the user by posing several questions about prominent features of the finding.

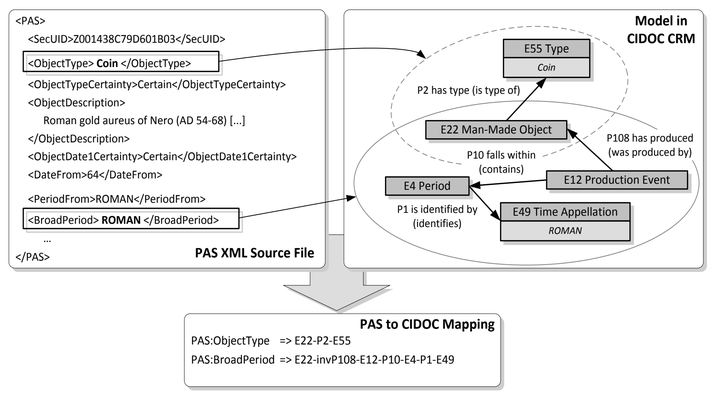


3 METHODOLOGY

Integration of metadata from different sources using the CIDOC CRM involves **three main steps**:

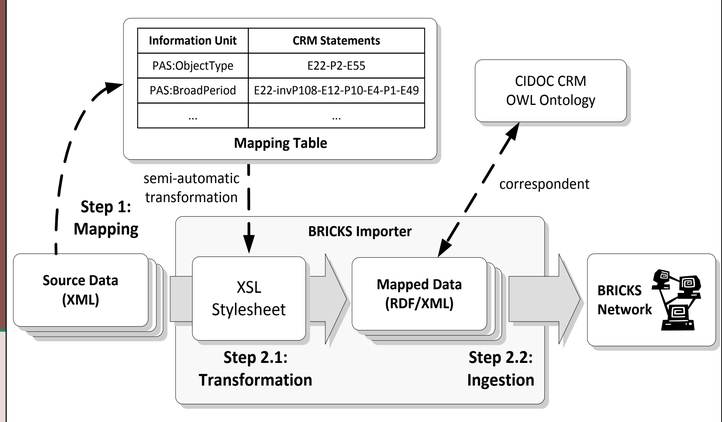
I Mapping

As an initial step, the source schemes have to be mapped to the ontology by experts of the source and target schema.
In our integration scenario, mappings are defined using spreadsheets which are then semi-automatically transformed to XSL stylesheets used to transform the source data to their target representation.
The CIDOC CRM does not present a methodology or guidance to what and how to document metadata. Therefore mapping inconsistencies – such as different mappings for equivalent metadata or equal mappings for semantically different metadata – may easily occur, particularly in scenarios where several source schemes are mapped independently.



II Lifting and Normalisation

The instance data must be made available to the application, thereby lifted and normalized into a common representation.
The BRICKS framework has been designed to import data from source systems.
The process of lifting and normalisation involves two individual steps, namely (i) the *data transformation* according to the mapping specification created in Step I and (ii) the actual *data ingestion* whereby the (transformed) data is stored in the system.



III Data Processing

This step is concerned with providing means of searching, retrieving and rendering the integrated metadata.
The actual structure of the integrated metadata depends on the initial mapping, so querying for specific aspects requires incorporation of mapping information, i.e. the different classes and properties and vocabularies used.
Our configurable faceted-style search (*guided search*) involves creation of SPARQL queries from the mapping chains known by means of the mapping specification created in Step I. This allows taking into account both the semantics of the integrated data and the possibility of different mapping structures of different source schemes.

4 CONCLUSIONS

In a data integration scenario, a global ontology provides the concepts against which the data source specific schema elements are mapped and over which user requests are formulated.

There are two main issues that could impede the actual goal of metadata interoperability:

The first issue is the abstractness of the concepts (e.g. Time Appellation, Man-Made Object) defined by the global ontology, which makes them ambiguous to any human user. Even expert users have produced ambiguous mappings and have required several iterations to produce consistent mapping definitions.
If several experts specify mappings independently from each other, it is very likely that they will produce incompatible mappings and fail the goal of enabling interoperability.
Another point directly connected to the abstractness of the concepts, is the presentation to the user. Basically a graphical user interface is required which hides the complexity of the global ontology and allows the user to formulate queries over more concrete concepts.

The second issue is the lack of technical specifications in global ontologies such as the CIDOC CRM. Without any detailed instructions of how to implement the mappings, represent instances, and process data during run-time, it is likely that each institution applies its own interpretation on a standardised global ontology. This again causes heterogeneities in scenarios that initially have aimed at providing interoperability.

5 FURTHER INFORMATION

Try the Archaeological Sites Finds Identifier at http://finds.brickscommunity.org:8091/findsidentifier.

For more detailed information on how the CIDOC CRM has been applied in our application context, see our technical report [3].

If you have any questions, please contact:
philipp.nussbaumer@researchstudio.at
bernhard.haslhofer@univie.ac.at

REFERENCES

1. EU-FP6: BRICKS – Building Resources for Integrated Cultural Knowledge Services (IST 507457) (2007) http://www.brickscommunity.org
2. CIDOC Documentation Standards Group: CIDOC Conceptual Reference Model (CRM) – ISO 21127:2006) (December 2006)
3. Nussbaumer, P., Haslhofer B.: Putting the CIDOC CRM into Practice – Experiences and Challenges. Technical Report, University of Vienna (September 2007) http://www.cs.univie.ac.at/publication.php?pid=2965

Project BRICKS, Building Resources for Integrated Cultural Knowledge Services
European Commission • Sixth Framework Programme