Setting up a CIDOC CRM Adoption and Use Strategy

CIDOC CRM: Success Stories, Challenges and New Perspective

George Bruseker

CIDOC 2017

Tbilisi, Georgia

27/09/2017
Goal: A Semantic Knowledge Ecosystem

**Layer of “Latest Knowledge”**
- Burial of Qin Shi Huang
- Discovery
- Qin Shi Huang
- Terracotta Warrior Museum
- Xiangyang
- Xian
- Terracotta Warriors

**Evidence layer**
- Things Sources Corpora

**Publications**
- Stories
- Exhibitions

**Refer**
- Interpret
- Present

**Search, correlate, integrate**
- Discover
- Collect
- Aggregate
- Update

**Researchers, interpreters, curators, conservators, excavators**
- Ford

**FORTH INSTITUTE OF COMPUTER SCIENCE**
- Information systems laboratory
- Centre for cultural informatics
Components View of a CRM / Semantic Adoption Strategy
CRM Adoption Strategy Overview

Planning
- Need
- Source
- Conceptual Modelling
- Tools Decision

Mapping
- Mapping
- Testing
- Normalization
- Harmonization

Deployment
- Interface deployment
- Interface loading
- Interface testing

Use
- User adoption
- User feedback
Planning Overview

1) Identify Need

2) Identify Sources

3) Conceptual Modelling

4) Identify Tools
Need Assessment

**IF**
- Common research or presentation goals

**AND**
- Heterogeneous Data

**AND**
- One System not possible/desired

**THEN**
- You need semantics!
## Source Analysis

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need</td>
<td>Sources</td>
<td>Modelling</td>
<td>Tools</td>
</tr>
</tbody>
</table>

- List Sources
- Classify Sources (encoding/purpose)
- Identify Maintainers of Resources
- Identify who talks to who, who uses who’s information?
- Study Semantic Content of Sources
- Identify and List Research Questions
- Choose most likely appropriate target ontologies
- Learn ontology

**Tip:** Data is in constant state of change. Use at least a spreadsheet to track source information over time.
## Conceptual Modelling

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need</td>
<td>Sources</td>
<td>Modelling</td>
<td>Tools</td>
</tr>
</tbody>
</table>

- Test target ontologies by attempting mappings
- Look for potential gaps in ontology constructs
- Use list of questions identified in source analysis to critique mapping
- Identify potential gaps in model and propose extension classes/relations suitable to your end

**Tip:** CRM is supported by a community of researchers and professionals. Ask the CRM list.
**Tool Selection**

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need</td>
<td>Sources</td>
<td>Modelling</td>
<td>Tools</td>
</tr>
</tbody>
</table>

- You will need at least
  - Mapping tool
  - Triple Store or Graph DB
- You will probably want
  - Graphical Environment/platform for exploration and/or creation and/or visualization and reporting of semantic data
- Your decision will be affected by
  - Budget
  - Available expertise

**Tip:** One of the foundational points of CRM data is to unbind users from proprietary formats and commitment to closed data models. Consider always having an open transform of your data to RDF or OWL.
<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need</td>
<td>Sources</td>
<td>Modelling</td>
<td>Tools</td>
</tr>
</tbody>
</table>

**Sources Register**

- Excel
- PARTHENOP

**Mapping Tools**

- 3M Mapping Memory Manager
- Karma

**Triple Stores / Graph DBs**

- OrientDB
- neo4j
- blazegraph
- GraphDB

**Semantic Data Management Platforms**

- WissKI
- ResearchSpace
<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mapping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Normalizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Harmonizing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Workflow of Mapping Based on X3ML Suite

Heterogeneous, Related Datasets → DBs

Standard, Ontology → X3ML Mapping Files

Data Extraction and analysis → Cleaning and Enrichment → Learn target schema → Create mappings → Implements generator → Transform Data → Explore Harmonized Data

- a. Initial Setup
  - Data Engineer
  - Domain Specialist

- b. Occasional Review
  - Data Engineer

- c. Scheduled Ingests and Updates
Mapping

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping</td>
<td>Testing</td>
<td>Normalizing</td>
<td>Harmonizing</td>
</tr>
</tbody>
</table>

- Carried out by Domain Expert
- Carried out by someone who knows the source data
- Mapping takes place relative to research problematic documented in planning
- Mapping process should be saved, repeatable
- Mapping can be linked to dialogue with other partners in project

Screenshot X3ML Mapping Suite
Semantic mappings must ultimately ‘make sense’ to a human user

Carried out by Domain Expert with Developer

Check RDF or use visualization tool to verify transformed data communicates intended meaning

Return to mapping where necessary

Repeat

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping</td>
<td>Testing</td>
<td>Normalizing</td>
<td>Harmonizing</td>
</tr>
</tbody>
</table>
Data Normalizing

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping</td>
<td>Testing</td>
<td>Normalizing</td>
<td>Harmonizing</td>
</tr>
</tbody>
</table>

- Schema to Schema matching from a source to CRM leaves data values unchanged.
- Normalization of data values according to international standard thesauri ensures better interoperability.
- Normalization Operation should be documented.
Ontology is like a pidgin language
Like any language there are multiple ways to say the same thing
After mapping individual sources, resulting maps must be semantically harmonized.
Having a database of mappings allows you to manage this process.

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping</td>
<td>Testing</td>
<td>Normalizing</td>
<td>Harmonizing</td>
</tr>
</tbody>
</table>

Screenshot from X3ML tool suite; Db of mappings helps organize harmonization
### Deployment Overview

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
</table>

- **Deploy Platform**
- **Load**
  - Ensure context
- **Test**
  - Formulate questions as queries
- **Return to Mapping if Necessary**
## Use Overview

<table>
<thead>
<tr>
<th>Planning</th>
<th>Mapping</th>
<th>Deploying</th>
<th>Use</th>
</tr>
</thead>
</table>

- **User Adoption**
  - Query, Data Entry, Visualization, Navigation

- **User Feedback**
  - Researcher satisfaction
  - Return of data required
  - Input for platform improvement
  - Input for data improvement, expansion, Enrichment
  - Input for Query Improvement
  - Return to Original Planning Phase

![ResearchSpace](image1.png)

*e.g. ResearchSpace Visual Query Mechanism*

![Wisski](image2.png)

*e.g. Wisski Semantic Data Display*
Semantic Data and CRM, Why Again?

Research

• Ask and answer questions across ‘knowledge silos’ (in and across institutions and disciplines)
• Ask and answer more complex questions
• Greater knowledge discovery facility
• Rediscover and reanalyze ‘lost’ data

Information Management

• Improve data awareness / interoperability
• Support sustainability and reusability of data
• Software neutral exchange format
## Semantic Data and CRM Adoption

Learn More

### CIDOC CRM Reference Materials

<table>
<thead>
<tr>
<th>CIDOC CRM Specification</th>
<th>Visual Charts</th>
<th>Tutorials</th>
<th>Mailing list</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.cidoc-crm.org/releases_table">http://www.cidoc-crm.org/releases_table</a></td>
<td><a href="http://old.cidoc-crm.org/cidoc_graphical_representation_v_5_1/graphical_representation_5_0_1.html">http://old.cidoc-crm.org/cidoc_graphical_representation_v_5_1/graphical_representation_5_0_1.html</a></td>
<td>One video</td>
<td></td>
</tr>
</tbody>
</table>

### Research Data Management Model

<table>
<thead>
<tr>
<th>Parthenos Project</th>
<th>(in progress)</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.parthenos-project.eu/">http://www.parthenos-project.eu/</a></td>
<td></td>
</tr>
</tbody>
</table>

### X3ML Mapping Suite Tool

<table>
<thead>
<tr>
<th>Open Access Service</th>
<th>Semantic Platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://www.ics.forth.gr/isl/3M">https://www.ics.forth.gr/isl/3M</a></td>
<td>Wiss-ki Project</td>
</tr>
<tr>
<td>Source Code</td>
<td>Research Space Project</td>
</tr>
<tr>
<td><a href="https://github.com/isl/3MEditor">https://github.com/isl/3MEditor</a></td>
<td><a href="http://www.researchspace.org/">http://www.researchspace.org/</a></td>
</tr>
<tr>
<td><a href="https://github.com/isl/x3ml">https://github.com/isl/x3ml</a></td>
<td></td>
</tr>
</tbody>
</table>
End

Questions?

Dr. George Bruseker
ICS-FORTH
bruseker@ics.forth.gr