Claros - towards a unified semantic database for the world of ancient art

Sebastian Rahtz, Information and Support Group Manager, Oxford University Computing Services

[for Research databases in the humanities - where next?]

Oxford, January 21st 2011
CLAROS Vision

CLAROS (Classical Art Research Online Services) is a collaborative initiative led by the University of Oxford, working in two areas of multi-disciplinary research:

- Classical antiquity
- Information and Communication Technologies

and with two aims:

- To enhance and disseminate the highest level of scholarship to the broadest global public
- To use datasets in Classics and Classical Art to exploit the potential of ICT for public service
Classical art is the starting point

Sculpture

Pottery

Gems
The CLAROS programme

1. Development of a humanities dataweb combining leading classical art history and related databases
2. Demonstration interfaces to explore classical art
3. Innovative searching based on shape analysis
4. Large-scale RDF database providing a testbed for performance research
5. Changing the approach to data discovery by development of a conversational Companion
A database is not just for Christmas

The oldest approach  Collect data, put it on index cards, write a book. END

The old approach  Collect data, manipulate it using Access or Excel, write a book. END

The newer approach  Collect the data, put it in a MySQL database, write a web front end with search boxes and browsing. END.

The best approach  Collect data, share it with others, look at their data, write a book, leave the data available for the next generation. NO END
Our target is to produce **linked open data**.
Web sites are exemplars not sole gateways.
We follow the semantic approach of RDF and a public ontology.

We aim at **standardization** to assist **long-term preservation** and **extensibility**.
The semantic web mantra, 5 star data

| ★ | Available on the web (whatever format), but with an open licence |
| ★★ | Available as machine-readable structured data (e.g. Excel instead of image scan of a table) |
| ★★★ | As (2) plus non-proprietary format (e.g. CSV instead of Excel) |
| ★★★★ | All the above, plus: Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff |
| ★★★★★ | All the above, plus: Link your data to other people’s data to provide context |

http://www.w3.org/DesignIssues/LinkedData.html
CLAROS: data resources

Target:

- data web integrating access to scholarly information on classical art
- semantic integration of the distributed, heterogeneous and non-interoperable digital resources held by CLAROS partners

University of Oxford – Beazley Archive:

- Electronic documentation started 1979
- 150,000 Pottery records and 130,000 images
- 50,000 Engraved gem and cameo records and 30,000 images
- 900 Plaster casts records (classical sculpture) and 1000 images
- 900 Antiquarian photographs

University of Oxford - Lexicon of Greek Personal Names:

- Electronic documentation started 1975.
- 400,000 recorded individuals. Over 35,000 unique personal names.
University of Cologne – Research Sculpture Archive:

- Electronic documentation started 1972
- 250,000 Sculpture records, 490,000 images.

German Archaeological Institute:

- 1,500,000 photographs

University of Paris X - Lexicon Iconographicum Mythologiae Classicae:

- Created 1972.
- 100,000 records, 180,000 images of mythological and religious iconography from 2,000 museums and collections.

A total of 2 million records and images
Disparate technologies

**Beazley Archive**  ‘XDB’ – XML data, SQL Server Database, ASP front end.

**Cologne Research Archive and German Archaeological Institute**  
‘Arachne’ - MySQL database, PHP front end.

**LIMC**  MySQL database, PHP front end.

**LGPN**  Ingres relational database, also available as an XML database serving TEI XML data
Implementation of CLAROS data web approach

- **No changes** to the databases of the individual sources
- **Semantic differences** between data sources are resolved by mapping selected metadata from each source to CIDOC-CRM
- **Syntactic differences** between data sources are resolved by converting the selected metadata to RDF, accessed from a single triple store using SPARQL
- **The co-reference problem**, where the same entity is known by different names in different databases, is solved by creating a co-reference service to disambiguate synonyms

CLAROS is simply a resource discovery service using minimal metadata — the user is ultimately directed back to the original data publisher's site for full information about an event, object, place or person of interest.
System Components

- Beazley Archive
- DAI Arachne
- LIMC (Paris)
- LGPN (Oxford)

Converting to CIDOC-CRM RDF

- Cache
- Index
- Query

CLAROS application

Browser
What is CIDOC?

The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

Key elements:
- Actors (people)
- Conceptual objects
- Physical things
- Events
- Time spans
- Places

and relationships between them, e.g.:
- participate in
- refer to
- have location
- within
# CIDOC CRM components

- Acquisition Information
- Appellation Information
- Attribute Assignment
- Changing Thing
- Collection Information
- Condition Information
- Deaccession and Disposal Information
- Description Information
- Documentation and References
- Existence Information
- Group Dynamics
- Image Information
- Institution Information
- Location Information
- Mark and Inscription Information
- Material and Technique Information
- Measurement Information
- Object Association Information
- Object Collection Information
- Object Entry Information
- Object Name and Classification Information
- Object Number Information
- Object Production Information
- Object Title Information
- Part and Component Information
- Person Nationality Information
- Planned Activities (design
- Recorder Information
- Reference Information
- Reproduction Rights Information
- Spatial - Temporal Relationship
- Subject Depicted Information
- Taxonomic Discourse
- Time-Span Information
Details of a typical subset of CIDOC

(from http://cidoc.ics.forth.gr/cidoc_graphical_representation_v_5_1)
CLAROS and CIDOC CRM

We have found CIDOC CRM to be well suited for CLAROS data

- There is an OWL implementation of CIDOC CRM by Erlangen University
- We focused initially on the CIDOC CRM Core terms, and employed additional terms as necessary
- CIDOC CRM Core can describe the complex provenance of artefacts and their relationships with key events, people, places and times
- The necessary complexity of the resulting RDF/XML is mostly invisible to developers, and entirely hidden from users
- The CIDOC CRM "E55.Type" system is particularly useful to permit faceted/drill-down queries, e.g. restricting results by the shape of a pot
- The CIDOC choices are documented at http://www.clarosnet.org/wiki/index.php
An inscription published in *Inscriptiones Graecae* volume XI (4), p. 1256 documents a man called Παράμονος, attested at Delos in the 3rd or 2nd century BC. He is noted as being the father of someone called Δημήτριος.
The Greek in data source

Relational DB:

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>floruit</th>
<th>sex</th>
<th>status</th>
<th>settlement</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1-43005</td>
<td>Παράμονος</td>
<td>hell.-imp.</td>
<td>1</td>
<td></td>
<td>Mytilene</td>
</tr>
<tr>
<td>V1-47408</td>
<td>Παράμονος</td>
<td>m.iii BC</td>
<td>1</td>
<td>paroikos</td>
<td></td>
</tr>
<tr>
<td>V1-76555</td>
<td>Παράμονος</td>
<td>ii-i BC</td>
<td>1</td>
<td></td>
<td>Eretria Vathia</td>
</tr>
<tr>
<td>V1-76557</td>
<td>Παράμονος</td>
<td>c.100BC</td>
<td>1</td>
<td></td>
<td>Histiaia-Oreos</td>
</tr>
<tr>
<td>V1-78877</td>
<td>Παράμονος</td>
<td>iv/iii BC</td>
<td>1</td>
<td></td>
<td>Eretria</td>
</tr>
<tr>
<td>V1-85238</td>
<td>Παράμονος</td>
<td>iii/ii BC</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

or XML:

```xml
<person n="1-7" xml:id="V1-85238">
  <sex value="1"/>
  <persName type="main" nymRef="#nParalmonos">Παράμονος</persName>
  <birth notAfter="-0175" notBefore="-0225">
    <placeName key="LGPN_11270" evidence="attested">Delos</placeName>
  </birth>
  <floruit>iii/ii BC</floruit>
  <state key="#relationship">
    <label>f. <persName type="relationship" xml:lang="el-grc" nymRef="#nDhmh1trios">Δημήτριος</persName>
  </label>
  </state>
  <bibl>
    <title>IG</title> XI (4) 1256
  </bibl>
</person>```
A CIDOC structure for a person like this

- **E55 Type (or local inheriting class socsec status)**
- **E33 Linguistic object (to the TEI element xml:id="V-30" from here)**
- **E82 Actor appellation**
- **P2 (or local inheriting class)**
- **P129 is subject of**
- **E55 Type (or local inheriting class sex)**
- **P4 has time span**
- **P98 was born**
- **E67 Birth**
- **P97 was father for**
- **E33 Linguistic object (to BCH 97 (1973))**
- **E52 Time span**
- **P100 died in**
- **P129 is subject of**
- **E47 Activity**
- **P11 participated in**
- **E4 Period**
- **E53 Place**
- **P7 took place at**
- **E41 Appellation (imperial)**
- **P1 is identified by**
- **E44 Place appellation**
- **E21 Person**
- **E7 Activity**
- **P2 has type**
- **E55 Type (flourit)**
- **P1 is identified by**
- **E82 Actor appellation**
- **E21 Person**
- **E82 Actor appellation**
The Greek expressed in RDF XML

```xml
<E21.Person
   rdf:about="http://clas-lgpn2.classics.ox.ac.uk/id/V1-85238">
   <P131.is_identified_by xml:lang="el-grc">
      <E82.Actor_Appellation>
         <value>Παράμονος</value>
      </E82.Actor_Appellation>
   </P131.is_identified_by>
   <P131.is_identified_by xml:lang="el-grc-x-lgpn">
      <E82.Actor_Appellation>
         <value>Paramonos</value>
      </E82.Actor_Appellation>
   </P131.is_identified_by>
   <P98.was_born>
      <E67.Birth>
         <P4.has_time-span>
            <E52.Time-Span>
               <P79.at_some_time_within>
                  <E61.Time_Primitive>
                     <claros:not_before rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">-0225</claros:not_before>
                     <claros:not_after rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">-0175</claros:not_after>
                  </E61.Time_Primitive>
               </P79.at_some_time_within>
            </E52.Time-Span>
         </P4.has_time-span>
         <P7.took_place_at rdf:resource="http://clas-lgpn2.classics.ox.ac.uk/placeid/LGPN_11270"/>
      </E67.Birth>
   </P98.was_born>
</E21.Person>
```
Example results — look for ‘kalos’

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>n</th>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArachneObject</td>
<td>4</td>
<td></td>
<td>Arachne / Object</td>
</tr>
<tr>
<td>LGPN Person</td>
<td>2</td>
<td></td>
<td>LGPN / Person</td>
</tr>
<tr>
<td>Beazley Object</td>
<td>23</td>
<td></td>
<td>Beazley / Object</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArachneObject</td>
<td>Schale mit kalos-Inschrift, Nike und Jüngling - Schale mit kalos-Inschrift, Nike und Jüngling</td>
<td></td>
</tr>
<tr>
<td>LGPN Person</td>
<td><a href="http://clas-lgpn2.classics.ox.ac.uk/id/V2a-37410">http://clas-lgpn2.classics.ox.ac.uk/id/V2a-37410</a> - Kalos</td>
<td></td>
</tr>
<tr>
<td>ArachneObject</td>
<td>Pinax mit kalos-Inschrift - Pinax mit kalos-Inschrift</td>
<td></td>
</tr>
<tr>
<td>ArachneObject</td>
<td>Stele der Psyche, Frau des Kalos - Stele der Psyche, Frau des Kalos</td>
<td></td>
</tr>
<tr>
<td>ArachneObject</td>
<td>fragmentierte Schale mit kalos-Inschrift, zwei Jünglinge - fragmentierte Schale mit kalos-Inschrift, zwei Jünglinge</td>
<td></td>
</tr>
</tbody>
</table>
The CLAROS interface

- Each partner can integrate CLAROS data from the other partners using their own programming platform.
- As an example the Beazley Archive set up a *CLAROS Explorer* to show what is possible.
CLAROS Vision

CLAROS: data resources

The CLAROS interface

CLAROS faceted browser

Search all CLAROS partners' databases

Summary results 27 to 51 of 6445 for calyx and krater

Next Previous

Calyx-krater - Shapes - Pottery

The handles of the calyx-krater are placed low down on the body, at what is termed the cul. Calyx-krater. The handles of the calyx...
http://www.clarasnet.org/pottery/shapes/calyx.htm - ak - 2008-10-09

Column-krater - Shapes - Pottery

skip links. You are here: Home > Pottery > Shapes > Kraters > Column-krater. ... Column-krater. ... Athenian red-figure column-krater ht. 35cm. ...
Image-based search

Image Search Results

Your original image

Listed below are pottery images with similar decorations and shape to the image you uploaded. Click on the links or images to open details about each record.

Or click here to view the distribution of pottery of shape AMPHORA, NECK in all the CLAROS partners' databases.

AMPHORA, NECK
302250, Munich, Loeb, SL458
Score: 26
Confidence: High

AMPHORA, NECK
302250, Munich, Loeb, SL458
Score: 22
Confidence: High

AMPHORA, NECK
320306, San Francisco (CA), M.H. de Young Memerorial Museum, 1925.368
Score: 9
Confidence: High
The relevance of CLAROS

We believe that CLAROS is important because

- It is not designed to be a demonstrator, but to deliver real, complete, data
- We will be providing real research routes into the data, used by real researchers
- We are testing an ontology, CIDOC CRM, on a proper scale
- There is a genuine collaboration between humanities, engineering and computer science
- We show the rest of humanities that the dataweb approach can work technically
Who leads this? Professor Donna Kurtz, Beazley Archive, Oxford

Who does the tech? The distributed partners, coordinated at OERC, Oxford

Is it available? The public launch is in spring 2011

How can I take this semantic web approach myself? The InfoDev at OUCS (http://www.oucs.ox.ac.uk/infodev/) are well placed to advise and assist research projects in this area