CLAROS: tensions between CIDOC CRM and the real world

Oxford CLAROS project partners
[Donna Kurtz (Beazley Archive), David Shotton (Department of Zoology), Yorick Wilks (Oxford Internet Institute), Greg Parker (Beazley Archive), Graham Klyne (Department of Zoology), and Andrew Zisserman (Engineering Science)]
presented by Sebastian Rahtz (OUCS).

Oxford, March 29th 2010

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

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

CLAROS: data resources

Target:

- data web integrating access to the world's 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.

CLAROS: data resources (2)

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 eXist XML database serving TEI-XML data. XQuery front end.

The 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

System Components
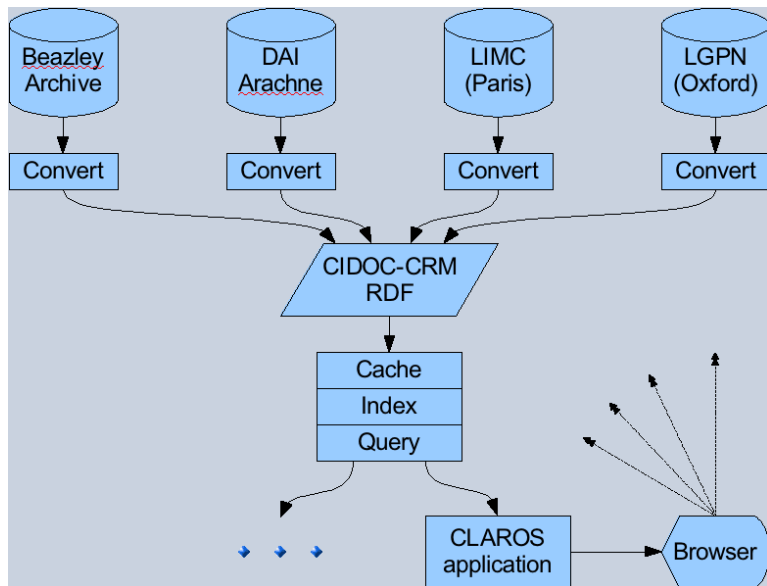
Technicalities

- ad hoc data loaders
- Jena TDB
- Lucene indexing
- SPARQL interface
- The Jena/LARQ (an extension of the ARQ query component in Jena) allows Lucene index queries to be accessed via SPARQL

Example results — look for 'kalos'

Source	Type	*n*	Select
Arachne	Object	4	Arachne / Object
LGPN	Person	2	LGPN / Person
Beazley	Object	23	Beazley / Object

| Source	Type	Description
Arachne	Object	Schale mit kalos-Inschrift, Nike und Jüngling - Schale mit kalos-Inschrift, Nike und Jüngling
LGPN	Person	http://clas-lgpn2.classics.ox.ac.uk/id/V2a-37410 - Kalos
Arachne	Object	Pinax mit kalos-Inschrift - Pinax mit kalos-Inschrift
Arachne	Object	Stele der Psyche, Frau des Kalos - Stele der Psyche, Frau des Kalos
Arachne	Object	fragmentierte Schale mit kalos-Inschrift, zwei Jünglinge - fragmentierte Schale mit kalos-Inschrift, zwei Jünglinge
Beazley	Object	138, Berlin, Antikensammlung, 3228 - Marcade, J., Eros Kalos (Geneva, 1962): 116-17
Beazley	Object	402, Athens, National Museum, XXXX402 - Marcade, J., Eros Kalos (Geneva, 1962): 137
Beazley	Object	8907, Paris, Musee du Louvre, XXXX8907 - Marcade, J., Eros Kalos (Geneva, 1962): 146-47
Beazley	Object	200618, Paris, Cabinet des Medailles, 509 - Marcade, J., Eros Kalos (Geneva, 1962): 111

CLAROS dataweb delivery

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.

The CLAROS interface

- Each partner can integrate CLAROS data from the other partners using their own programming platform.
- As an example Beazley Archive set up a *CLAROS Explorer* show what is possible.
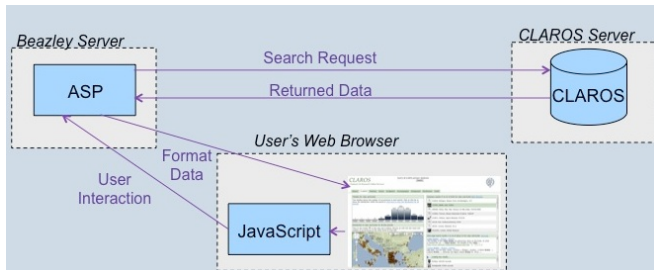
CLAROS faceted browser
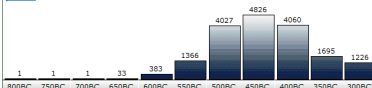
CLAROS
Classical Art Research Online Services

Search all CLAROS partners' databases
[   ] [Search]

Home **CLAROS** **Pottery** **Gems** **Sculpture** **Iconography** **Antiquaria** **Dictionary** **Tools**
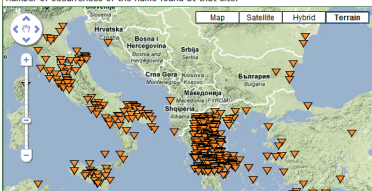
Timeline for calyx and krater

The timeline shows the number of occurrences in each period. Click on the bar to show the distribution within the period or click here to view the distribution for all periods.

800BC	750BC	700BC	650BC	600BC	550BC	500BC	450BC	400BC	350BC	300BC
1	1	1	33	383	1366	4027	4826	4060	1695	1226

Distribution of calyx and krater for all time periods

Click on the marker ▼ on the map and a balloon will pop up with the site name and number of occurrences of the name found at that site.

[Map] [Satellite] [Hybrid] [Terrain]

Summary results 27 to 51 of 6445 for calyx and krater Next Previous

- 214619, Bologna, Museo Civico Archeologico, 217
- 217501, Berlin, lost, F2642
- 206500, Rome, Mus. Naz. Etrusco di Villa Giulia, XXXX0.6500
- 214605, Ferrara, Museo Nazionale di Spina, T66AVP
- 214871, Athens, Agora Museum, P21535
- 29128, Kiel, Antikensammlung, B795
- 28435, Cyrene, Museum, 81.6
- 9010393, London, British Museum

Web page search results 1 to 10 of about 13 for calyx and krater Next page

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.clarosnet.org/pottery/shapes/calyx.htm - 8k - 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. ...

Loading new results...

- Pottery 102155 records
- Rundplastik 37905 records

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 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

The future of CLAROS

- Enhance performance to a level acceptable to the general public
- Bring in new partners with more classical art data
- Develop visual search and analysis interfaces
- Research dynamic aggregation
- Expand the humanities dataweb concepts chronologically and geographically

Musing on CIDOC tensions

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. eg

- participate in
- refer to
- have location
- within

CIDOC CRM components

- Acquisition Information
- Appellation Information
- Attribute Assignement
- 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
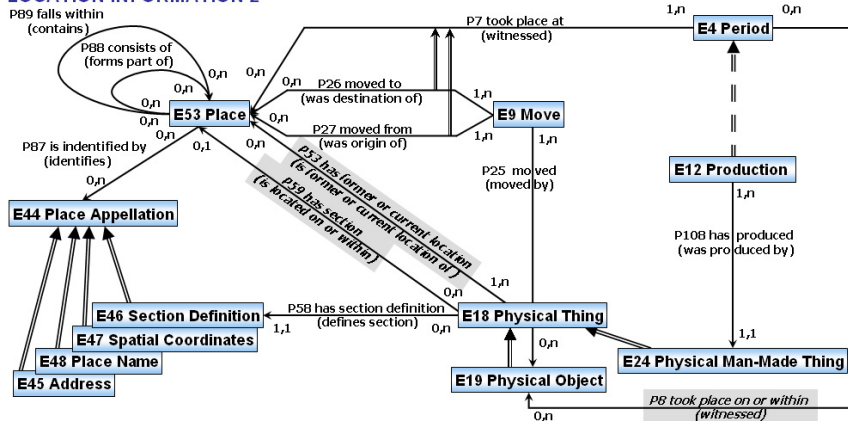
(from http://cidoc.ics.forth.gr/cidoc_graphical_
representation_v_5_1)

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

Example: a Greek

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:

```
    id     |    name    |   floruit   | sex |  status   |    settlement
-----------+------------+-------------+-----+-----------+------------------
 V1-43005  | Παράμονος  | hell.-imp.  |  1  |           | Mytilene
 V1-47408  | Παράμονος  | m.iii BC    |  1  | paroikos  |
 V1-76555  | Παράμονος  | ii-i BC     |  1  |           | Eretria Vathia
 V1-76557  | Παράμονος  | c.100BC     |  1  |           | Histiaia-Oreos
 V1-78877  | Παράμονος  | iv/iii BC   |  1  |           | Eretria
 V1-85238  | Παράμονος  | iii/ii BC   |  1  |           |
```

or XML:

```xml
<person  n="1-7"  xml:id="V1-85238">
<sex  value="1"/>
<persName  type="main"  nymRef="#nPara1monos">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>
```
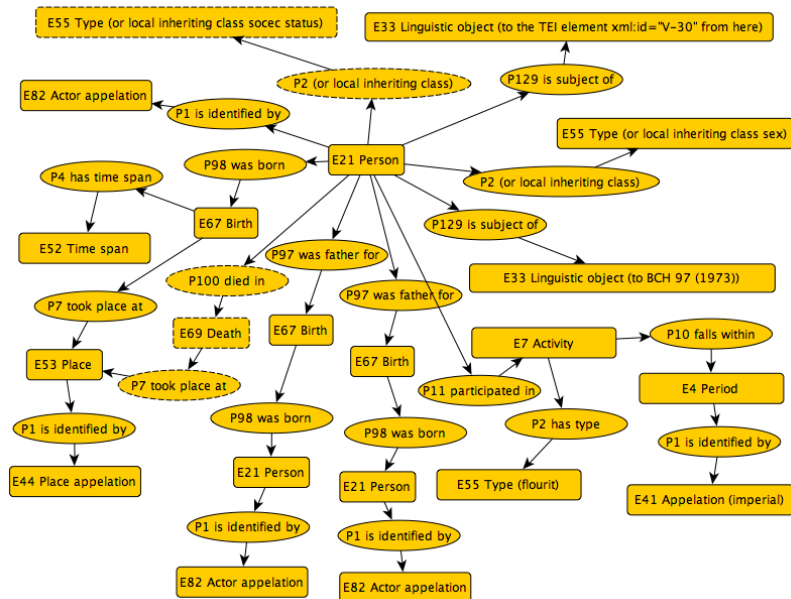
A CIDOC structure for a person like this

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>
```

Problem type 1: vague dates

? iii/ii BC

- At the least, we have to map this to a year range for sorting
- We probably have to map between relative calendars 'in the 3rd archonship of XXX' vs 'the 163rd year of the glorious rule of the YYY dynasty'
- We have periods which are well understood between scholars, clearly relative (to other similar periods), but not dated: 'Hellenistic', 'late Minoan'.

Problem type 2: names and persons

Heracles, Ηρακλής, Hercules, *Hraklh3s, Ηρακλής

- we have to map between alternative modern *transliterations*
- we have to allow for modern *translations*
- we have to distinguish between the *name* Ηρακλής, the *mythical person* Ηρακλής, and the *historically-attested person* named Ηρακλής
- Is our target to map to the name or the person?

Problem type 3: references

IG XI (4) 1256

A good bibliographical record, but not a linkable resource yet.

Problem type 4: undifferentiated human-parseable information

*A: EROTIC (COURTING), YOUTHS B: DRAPED YOUTHS AND
WOMEN UH: DOG AND LION SKIN I: PELEUS AND THETIS*

The information is all there, but only amenable to brute-force
textual recovery of the *name*> 'Peleus' and the *feature* 'dog and lion
skin'

Problem type 5: language

143135: fragmentierte Schale mit kalos-Inschrift, zwei Jünglinge Athen, Kunsthandel

CLAROS ingests data in three languages already — we are lucky here that the term 'kalos' is universal. 'Schale'? (bowl, cup; skin; peel; husk; shell, scallop; (hunting) hoof)

Problem type 5: place names

Paris (Texas); Paris (France); Athens; Athen; Athènes; Athina; Αθήνα

The spelling/transliteration of placenames is relatively easy to sort out. But:

- Which Paris is it? — no magic bullet
- Where is Paris? — ok if its a modern name, but geonames.org's 8 million placenames do not include all the classical places (where is 'Chef el Chamis' in Cyrenaica?)
- As with time, how to cope with uncertainty? 'near Abdera'
- How to express vagueness? 'Attica' vs 'Athens'?
- Like periods, what about undefined, yet precise, areas? 'they live on the other side of the river XXX'

Answering one part of the place puzzle: the Barrington Atlas

What if we simply mark all our places with the name, page number and grid reference in the Atlas?

Problem type 6: missing granularity in CIDOC

Coordinates are a specific form of E44 Place Appellation, that is, a means of referring to a particular E53 Place. Coordinates are not restricted to longitude, latitude and altitude. Any regular system of reference that maps onto an E19 Physical Object can be used to generate coordinates.

```
<P87.is_identified_by>
  <E48.Place_Name>
    <value>Aigiale</value>
  </E48.Place_Name>
</P87.is_identified_by>
<P87.is_identified_by>
  <E47.Place_Spatial_Coordinates>
    <claros:has_geoObject>
      <geo:Point>
        <geo:lat>25.516389</geo:lat>
        <geo:long>36.916667</geo:long>
      </geo:Point>
    </claros:has_geoObject>
  </E47.Place_Spatial_Coordinates>
</P87.is_identified_by>
```

CLAROS extensions to CIDOC CRM

A very few extensions have been made so far — we will likely need more in the area of uncertainty e.g.

- some additional RDF vocabulary for time metadata relating to imprecise periods and eras i.e. claros:not_before and claros:not_after, applied to a crm:E61.Time_Primitive object

This allows us to capture partial or imprecise quantitative information that is not expressed by a crm:has_PrimitiveTime property

New properties have been introduced as OWL datatype properties on the primitive value objects.

Conclusions

There are four critical features of the CLAROS use of CIDOC:

1. We are 'only' using the RDF/CIDOC database as an explorer — we map and expose from each partner just those elements which *can* be mapped

2. CIDOC has both provision for extensions, and hooks for separate typologies

3. We will be mediating our ingest and searching through extensive co-reference databases / thesauri

4. The CIDOC choices are documented at
   http://www.clarosnet.org/wiki/index.php