# CLAROS — bringing classical art to a global public

Donna Kurtz (Beazley Archive, Classics), David Shotton
(Department of Zoology), Florian Schroff (Engineering Science),
Yorick Wilks (Oxford Internet Institute), Greg Parker (Beazley
Archive), Graham Klyne (Department of Zoology), and Andrew
Zisserman (Engineering Science).

Presented by Sebastian Rahtz
(Oxford University Computing Services).

Oxford, December 12th 2009



#### 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







Pottery



Gems



# The CLAROS programme

- Development of a humanities dataweb combining leading classical art history and related databases
- Demonstration interfaces to explore classical art
- Innovative searching based on shape analysis
- Large-scale RDF database providing a testbed for performance research
- 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 datasets (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.

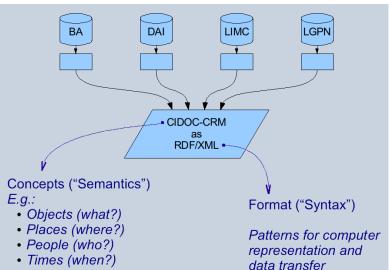


# The CLAROS data web approach

- 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

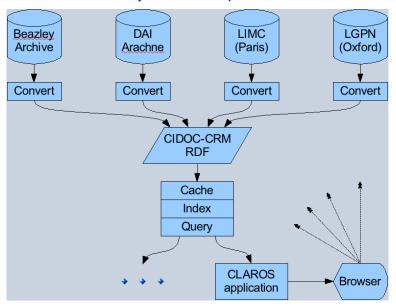


# Combining data



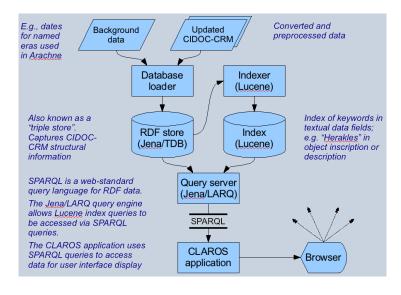


## **System Components**





#### **Technicalities**





#### **CLAROS and CIDOC CRM**

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

- 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



# RDF example

```
<F21.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>
 </Pl31.is identified by>
 <P131.is identified by xml:lang="el-grc-x-lgpn">
   <E82.Actor Appellation>
     <value>Paramonos</value>
   </E82.Actor Appellation>
 </Pl31.is identified by>
 <P98.was born>
   <F67. Rirth>
     <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>
```



#### **CLAROS** extensions to CIDOC CRM

A very few extensions were needed. 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 primtive value objects



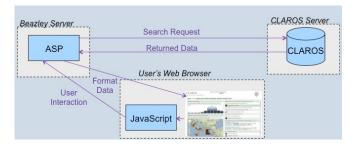
# **CLAROS** dataweb delivery

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 his own programming platform.
- As an example Beazley Archive set up a CLAROS Explorer show what is possible.





# The CLAROS homepage



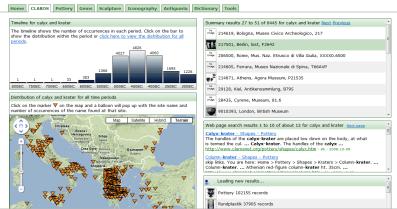


#### **CLAROS** faceted browser



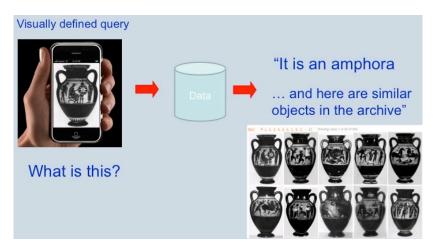
Search all CLAROS partners' databases





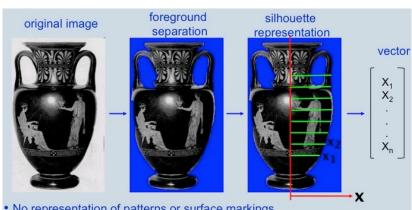


# Image recognition





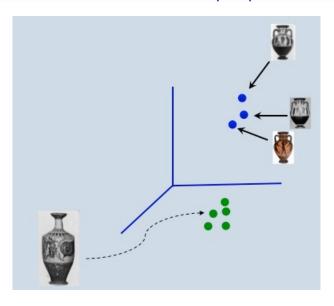
# Shape representation



- No representation of patterns or surface markings
- 100-dimensional "vase shape space"

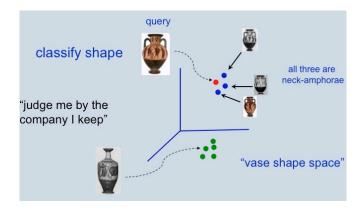


# Vase shape space





## Compute three nearest neighbours for each vase





# **Companions**

- Take the interface up a level from text searching
- Provide a personalized avatar: an agent or 'presence' that stays with the user for long periods of time, developing a relationship and 'knowing' its owners preferences and wishes. It will communicate with the user primarily by using and understanding speech.

Part of larger EU Companions project: http://www.companions-project.org/



#### 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

