Classical Art Semantics Information Extraction:
CASIE Pilot Project

Dr. Andreas Vlachidis
Hypermedia Research Unit
University of South Wales
http://hypermedia.research.southwales.ac.uk/

University of South Wales
Prifysgol De Cymru

The Beazley Archive – University of Oxford
http://www.beazley.ox.ac.uk

Classical Art Research Online Services CLAROS
http://www.clarosnet.org/

Introduction

University of South Wales
Prifysgol
De Cymru

CASIE Project

Motivation

Apply semantic technologies to make digital humanities material easily discoverable and available for reuse and comparative analysis purposes.

Aims

- Automatic extraction of information about cultural objects from classical art scholarly texts
- Represent information in terms of the CIDOC-CRM (ISO 21127:2006) metadata standard for cultural heritage

Focus

Scholarly Text: Corpus Vasorum Antiquorum (CVA)
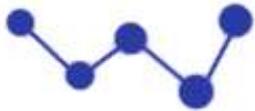
Background

CVA

- The oldest research project of the Union Académique

- CVA Initiated in 1922

- Contains 300 high-quality catalogues (fascicules) of ancient Greek painted pottery

- 100,000 vases Illustrations

- 120 collections

- In 26 different countries.

- 2004: Beazley archive completed the digitisation of the CVA fascicules

- Digitised result available from CVA online (www.cvaonline.org)

- But in bitmap format ☹

University of South Wales
Prifysgol De Cymru

CASIE Project

University of South Wales
Prifysgol
De Cymru

CASIE Project

Background

The CLAROS Project

- Classical Art Research Online Services (CLAROS)
- An international interdisciplinary research initiative
- Focused on semantic integration of world classical art records
    - Beazley Archive, the German Archaeological Institute, the Ashmolean Museum, the Eastern Art, Jameel Collection, the National Archaeological Museum of Greece, and other
- Delivers searchable semantic web interface
- CIDOC-CRM to enable semantic interoperability
- www.clarosnet.org

Method

University of South Wales
Prifysgol
De Cymru

CASIE Project
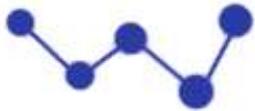
Semantic Annotation

Specific **metadata** which are usually generated with respect to a given **ontology** and are aimed to automate identification of **concepts** and their relationships in documents

Development Approach

Semantic annotation process driven by a **rule-based** Information Extraction (IE) techniques supported by domain-oriented **vocabulary**

Method

University of South Wales Prifysgol De Cymru

CASIE Project

Java Pattern Engine

Ontology
-CIDOC CRM-EH

General Architecture for Text Engineering

CVA Fascicules

Domain Vocabulary

CASIE Pilot Project

University of South Wales
Prifysgol De Cymru
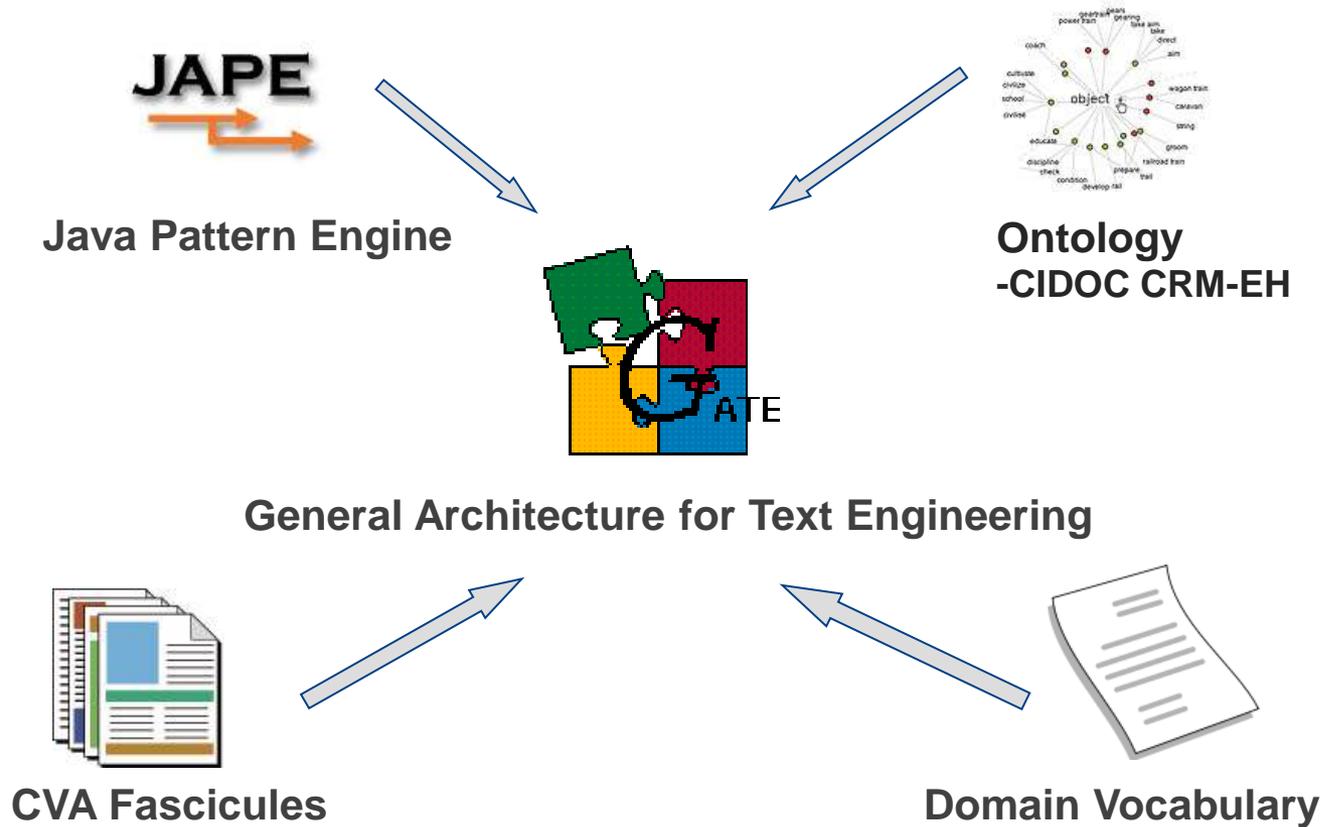
CASIE Project

CVA Fascicules (high quality catalogues)

- **12** Fascicules originating from
    - The British Museum (8)
    - The Ashmolean Museum (3)
    - Thessaloniki Archaeological Museum (1)
- **Published** between 1925 – 1998
- **Structure**: reasonably consistent among fascicules of same origin
- **Inconsistency** in terms of
    - Dimension abbreviations
    - Catalogue reference format
    - Size of descriptive passages of artefacts
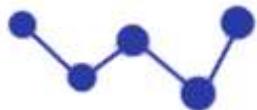
CVA Fascicules

British Museum

1. **Oenochoe.** White, yellow and purple paint, and incised lines. On neck, egg-pattern and parti-coloured wave line ; dots. On body, vine wreath. Under the foot is painted a red cross. Ht. 262. Old Collections.—Bibl. *Vase Cat.* (1870), no. 1713.

Ashmolean Museum

1, 2 (and Plate 10). NECK AMPHORA, 1965.116. Formerly Spencer-Churchill Collection.

J. D. Beazley, *JHS* xlvii (1927) 83, no. 9; 77, figs. 14, 15; *ABV* 273, no. 111 (Antimenes Painter); *Spencer-Churchill* no. 61; *Para* 119; Brommer[3], 349, no. 22.

Height 38·5. Width of lip 16·5, of body 25·6, of foot 13·0. Mended from several large pieces.

Black lip, plain top edge painted within to 6·5. Black triple-reeded handles, unpainted beneath. Below the handles a palmette cross with central dot, small flowers
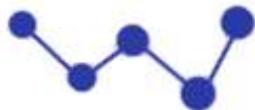
Information Extraction Focus

Sample Text

1. **Oenochoe.** White, yellow and purple paint, and incised lines. On neck, egg-pattern and parti-coloured wave line ; dots. On body, vine wreath. Under the foot is painted a red cross. Ht. 262. Old Collections.—Bibl. *Vase Cat.* (1870), no. 1713.

E22.Man-Made_object

E42.Intenifier

E54.Dimension

P3_has_note

CASIE Project

University of South Wales
Prifysgol De Cymru

Development Phases

University of South Wales
Prifysgol De Cymru

**CASIE Project**

Pre-processing

- Prepared images for OCR (Photoshop)
- Performed OCR (Abby Fine Reader 9)

Main Information Extraction Phase

- Developed the main IE pipeline
    - GATE Gazetteer development
    - IE rules development
    - Iterative process
    - Necessary adjustments (fascicules oriented)

Conversion of Semantic Annotations to RDF triples

- Bespoke PHP script using DOM
- Delivery of RDF expressions consistent with CLAROS (CIDOC- CRM) format
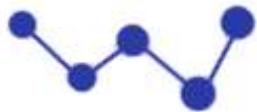
Pre-processing Phase

Adjustment of Image Levels (Photoshop)

Improve contrast between white background and black text to minimise OCR errors

OCR Result

2a and 2b. Amphora, (a) Athena ; on 1. Hermes ; on r. bearded man with staff, perhaps Zeus, (b) Winner of horse-race ; procession of youth bearing wreath and tripod^ mounted youth and bearded herald announcing A V N EI -KETV:HIPOZ:NIKAI, Av(o>eu?jT(°)" ttr(ii)os wkS. Ht. 447. From Vulci ; 1849.—Bibl. Cat. B 144

Main Information Extraction Phase

University of South Wales
Prifysgol De Cymru

CASIE Project
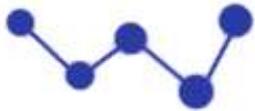
GATE Gazetteer

- Gazetteer support the IE with domain vocabulary
- Vase form listing originates from CVA online
- Supportive project specific list also created to support
    - Extraction of dimension
    - Extraction of Catalogue Reference

Beazley Archive Pottery Search

Search by Shape

To search the database click on the shape below, or c

A B C D E F G H I J K L M N O P

Alabastron
Amphora
Amphoriskos
Aryballos
Askos
Astragalos

Linear Definition

New Load Save Save as...

vase_shape_name.lst:shape
plate_strings.lst:plate
height.lst:height

New Load Save Save as...

ALABASTRON
ALABASTRON FRAGMENT
ALABASTRON FRAGMENTS
ALABASTRON FRAGMENT
ALABASTRON FRAGMENTS
AMPHORA
AMPHORA
AMPHORA FRAGMENT
AMPHORA FRAGMENTS
AMPHORA A
AMPHORA A

Main Information Extraction Phase

CASIE Pipeline

- A cascading mechanism of NLP components including GATE modules and bespoke JAPE rules

!	Name	Type
	Document Reset	Document Reset PR
	Tokenizer	ANNIE English Tokeniser
	Sentence Splitter	ANNIE Sentence Splitter
	POS	ANNIE POS Tagger
	CVA	ANNIE Gazetteer
	Plates_Objects	Jape Transducer
	Last_Word	Jape Transducer
	Plate_Section	Jape Transducer
	Object_Section	Jape Transducer
	Height	Jape Transducer
	Exporter	Flexible Exporter

Selected Processing resources

Type: Jape Transducer

CASIE Project

Main Information Extraction Phase

JAPE Rules

Advanced finite state transducer implementing elaborate regular expressions

```
{Lookup.majorType==shape}|
{Token contains Lookup.majorType==shape}|
((({Token.category==RB}|
{Token.category==NNP}|
{Token.category==JJ})
({SpaceToken.kind==space})?)[1,3]
({SpaceToken.kind==space})*
{Lookup.majorType==shape})
```

The above rule will match cases such as, "**Amphora**", "**Neck-Amphora**" and "**Fragment of belly of amphora**"

CASIE Project

Main Information Extraction Phase

Semantic Annotation Result in GATE

CASIE Project

2. Askos with neck-spout and bird-tail projection; two suspension-handles at sides. Drab clay and slip ; dull black bands and wave line. Ht. 103. Ohnefalsch-Richter, 1884.Bibl. Cat. C 809.

3. Lekythos. Buff clay and polished slip ; grey-black bands and, on shoulder, concentric circles. Ht. 70. Pierides, 1868.Bibl. Cat. C 791, pi. IV.

4. Small stamnos. Drab clay and slip ; grey-black bands. Ht. 67.Inv. 1908, 4-11, 40.

5. Jug with spout. Similar technique ; bands and con-centric hooks (on this ornament see Myres, op. cit., p. 81). Ht. 97. A. W. Franks, 1879.Bibl. Cat. C 772.

6. Mastos with spout and two small suspension-handles. Similar technique ; lattice triangles. Ht. 91. Ohnefalsch-Richter, 1884.Bibl. Cat. C 801, pi. IV.
II c c2.

7. Flask. Similar technique ; concentric rings. Ht. 110. From Kouklia, 1899.Bibl. Cat. C 799.

8. Jug. Similar technique ; bands ; on base, Maltese cross within rings. Ht. 60. From Curium, 1895, tomb 56.Bibl. Cat. C 775 ; Excav., p..74, fig. 129.

9. Jug. Buff clay and polished slip ; grey-brown paint. On lip, triangles ; herring-bone down handle ; on shoulder, concentric hooks ; below, bands. Ht. 81.Inv. 1908, 4-11, 37.

10. Jug with spout. Similar; dull black concentric circles on neck and shoulder; below, bands. Ht. 91. Pierides, 1868.Bibl. Cat. C 771.

11. False-necked jar. Drab clay and greenish-drab slip ; dull black paint, covering all lower part. On shoulder, triangles. Ht. 123. From Kouklia, 1899.Bibl. Cat. C 695, fig. 259.

Main Information Extraction Phase

Semantic Annotation Result in GATE

CASIE Project

Askos with neck-spout and bird-tail projection; two suspension-handles at sides. Drab clay and slip ; dull black nds and wave line. Ht. 103. Ohnefalsch-Richter, 1884.Bibl. Cat. C 809.
Lekythos. Buff clay and polished slip ; grey-black bands and, on shoulder, concentric circles. Ht. 70. Pierides, 68.Bibl. Cat. C 791, pi. IV.
Small stamnos. Drab clay and slip ; grey-black bands. Ht. 67.Inv. 1908, 4-11, 40.
Jug with spout. Similar technique ; bands and con-centric hooks (on this ornament see Myres, op. cit., p. 81). Ht. 97.
. W. Franks, 1879.Bibl. Cat. C 772.
Mastos with spout and two small suspension-handles. Similar technique ; lattice triangles. Ht. 91.
nefalsch-Richter, 1884.Bibl. Cat. C 801, pi. IV.
; c2.
Flask. Similar technique ; concentric rings. Ht. 110. From Kouklia, 1899.Bibl. Cat. C 799.
Jug. Similar technique ; bands ; on base, Maltese cross within rings. Ht. 60. From Curium, 1895, tomb 56.Bibl.
t. C 775 ; Excav., p..74, fig. 129.
Jug. Buff clay and polished slip ; grey-brown paint. On lip, triangles ; herring-bone down handle ; on shoulder,
ncentric hooks ; below, bands. Ht. 81.Inv. 1908, 4-11, 37.

RDF Conversion Phase

PHP Script

Converted the semantic annotations to RDF expressions
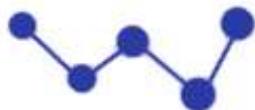
consistent with CLAROS (CIDOC-CRM) format
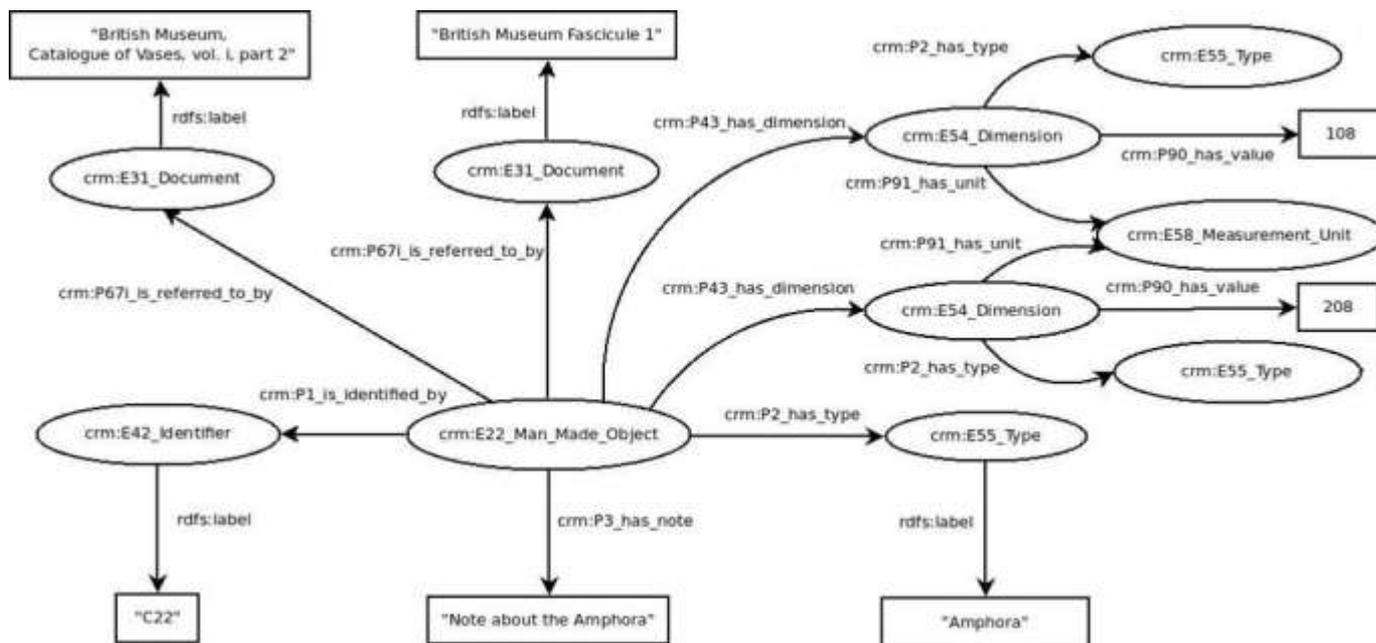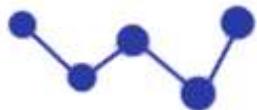
RDF Conversion Phase
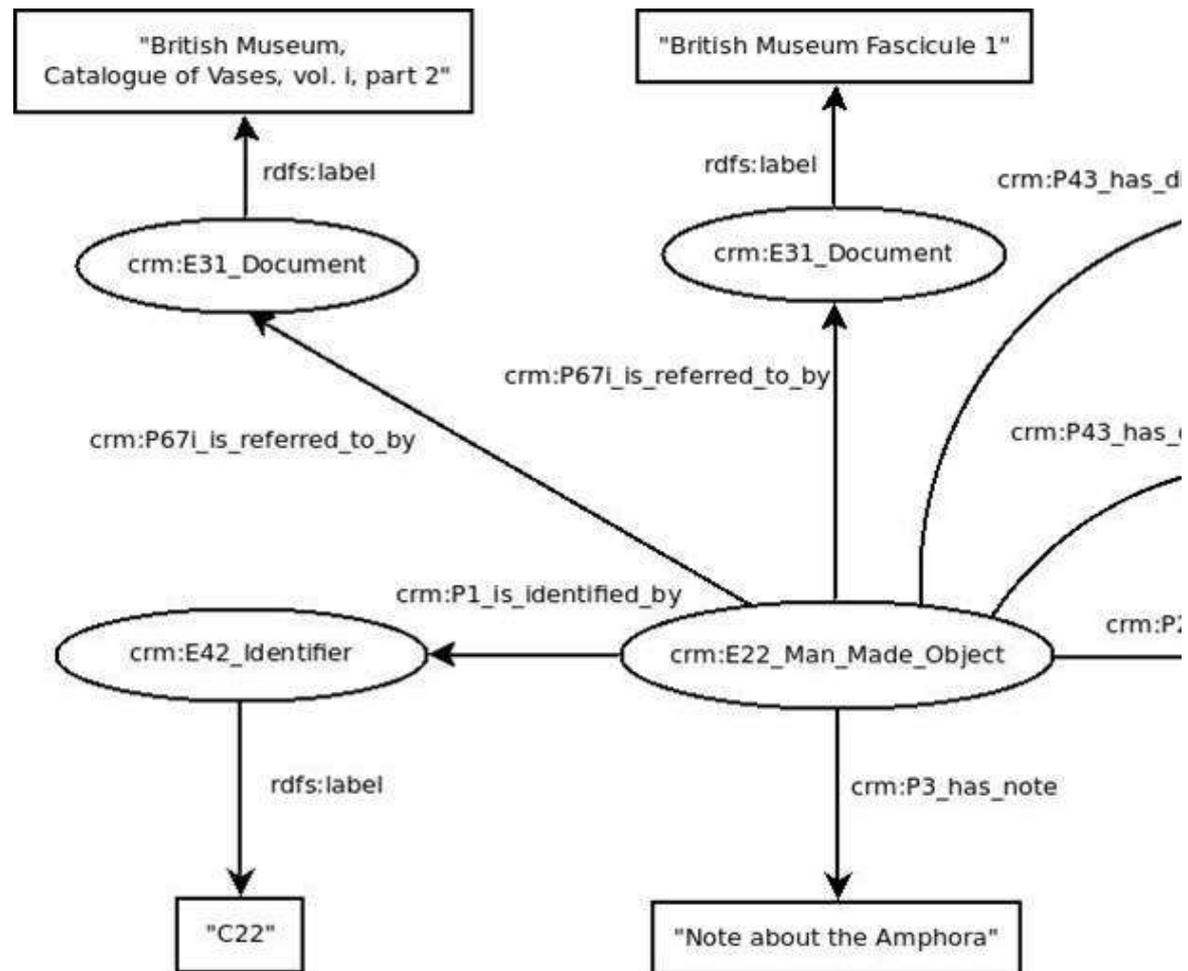
RDF Conversion Phase
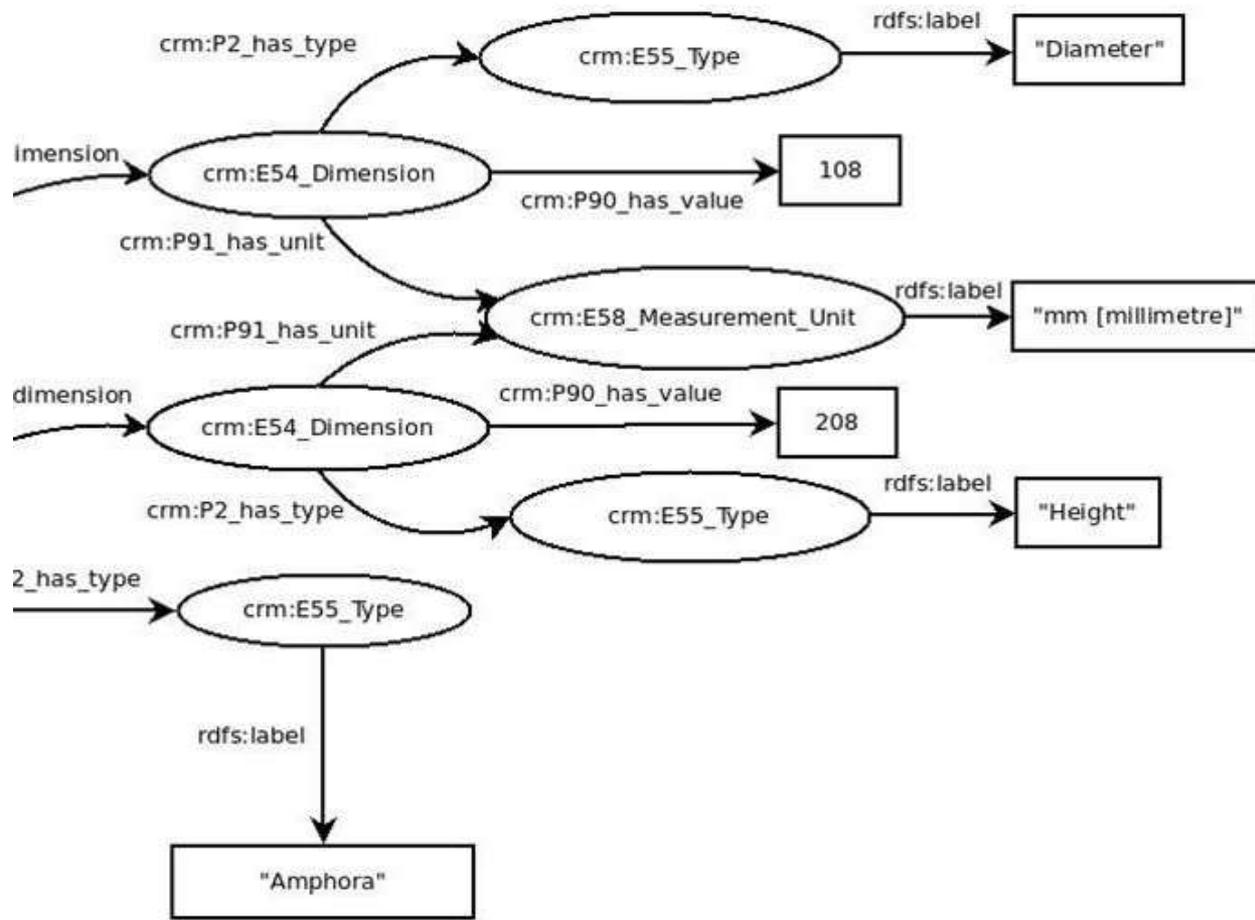
RDF Conversion Phase

RDF Expressions
Man-made Object, Note

```
-<rdf:RDF>
  -<crm:E22_Man-Made_Object rdf:about="http://purl.org/NET/Claros/#British_Museum_Fascicule_1.566707">
    -<crm:P2_has_type>
      -<crm:E55_Type>
        <rdfs:label>Cup on three legs </rdfs:label>
        <crm:P127_has_broader_term rdf:resource="http://purl.org/NET/Claros/vocab#Shape"/>
      </crm:E55_Type>
    </crm:P2_has_type>
    -<crm:P3_has_note>
      Cup on three legs ; elementary spout. Height 179 millimetres. From Phoenikiais.Bibl. Cat. C 13.
    </crm:P3_has_note>
```

RDF Conversion Phase

CASIE Project

RDF Expressions
Fascicule (Document), Catalogue Reference

```
-<crm:P67i_is_referred_to_by>
 -<crm:E31_Document>
    <rdfs:label>British_Museum_Fascicule_1</rdfs:label>
  </crm:E31_Document>
 </crm:P67i_is_referred_to_by>
-<crm:P67i_is_referred_to_by>
 -<crm:E31_Document>
    <rdfs:label>British Museum, Catalogue of Vases, vol. i, part 2</rdfs:label>
  </crm:E31_Document>
 </crm:P67i_is_referred_to_by>
-<crm:P1_is_identified_by>
 -<crm:E42_Identifier>
    <rdfs:label>C 13</rdfs:label>
  </crm:E42_Identifier>
 </crm:P1_is_identified_by>
```
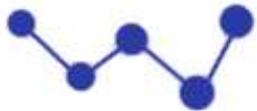
RDF Conversion Phase

RDF Expressions

Dimension

```
-<crm:P43_has_dimension>
 -<crm:E54_Dimension>
  -<crm:P91_has_unit>
   -<crm:E58_Measurement_Unit>
      <rdfs:label>mm [millimetre]</rdfs:label>
    </crm:E58_Measurement_Unit>
  </crm:P91_has_unit>
  -<crm:P2_has_type>
   -<crm:E55_Type>
      <rdfs:label>Height</rdfs:label>
    </crm:E55_Type>
  </crm:P2_has_type>
  <crm:P90_has_value>179</crm:P90_has_value>
 </crm:E54_Dimension>
</crm:P43_has_dimension>
```

University of South Wales Prifysgol De Cymru

CASIE Project

Conclusion

CASIE Pilot Project

- Delivered reasonably good results
- Rule-base IE can support the task
- Availability of Domain vocabulary (English)
- Pilot investigation paved the way for a potential large scale project
- Future development should address
    - Multilingual characteristics
    - Writing style of individual fascicules

Classical Art Semantics Information Extraction:
CASIE Pilot Project

Dr. Andreas Vlachidis
Hypermedia Research Unit
University of South Wales
http://hypermedia.research.southwales.ac.uk/

The Beazley Archive – University of Oxford
http://www.beazley.ox.ac.uk

Classical Art Research Online Services
CLAROS
http://www.clarosnet.org/