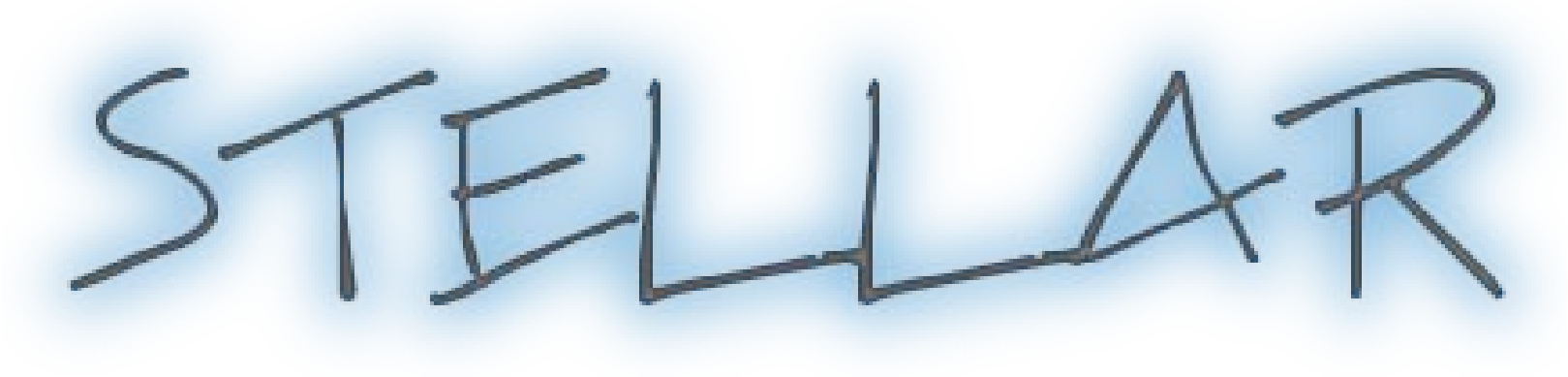


STAR/STELLAR Case Study: Complementary use of ontologies and (other) KOS

Douglas Tudhope

Hypermedia Research Unit, University of Glamorgan

Arts & Humanities Research Council

ENGLISH HERITAGE

Presentation

- Differences between thesauri and ontologies?

- STAR/STELLAR project case study
  Linked data project with archaeological excavation datasets
  Produced tools for non-experts to map/extract RDF

- Possibilities for interoperability
- How to consider typical purpose of a KOS?

Factors governing types of KOS **(from NKOS 2006)**
(AI) Ontology – revisited on *purpose*

Entities
Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)
Types / expressivity of relationships:
low (core set) – medium – high (definable)
concept-concept, concept-term, term-term
monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest
Metadata element: subject, various elements, general
Granularity of application objects: un/structured; discrete individual / general (document)
Relationship applying concepts to objects in domain
Extent of Interactive/automatic use -- information seeking tactics/logical inferencing
about (fuzzy), instance
Exhaustivity: low - high
Specificity: low - high
Coordination: low – high - expressivity and formality of relationships in coordination

STELLAR

- 12 month AHRC funded project
  - Hypermedia Research Unit, University of Glamorgan
  - Archaeology Data Service, University of York
  - English Heritage Centre for Archaeology, Portsmouth

- Builds on previous 3 year AHRC funded STAR Project

- Acknowledgments
  Ceri Binding (University of Glamorgan)
  Andreas Vlachidis (University of Glamorgan)
  Keith May, English Heritage (EH)
  Stuart Jeffrey, Julian Richards,
  Archaeology Data Service (ADS)
  Archaeology Department, University of York

STAR – Aims and background

- *Investigate semantic technologies for integrating and cross searching datasets and associated grey literature*

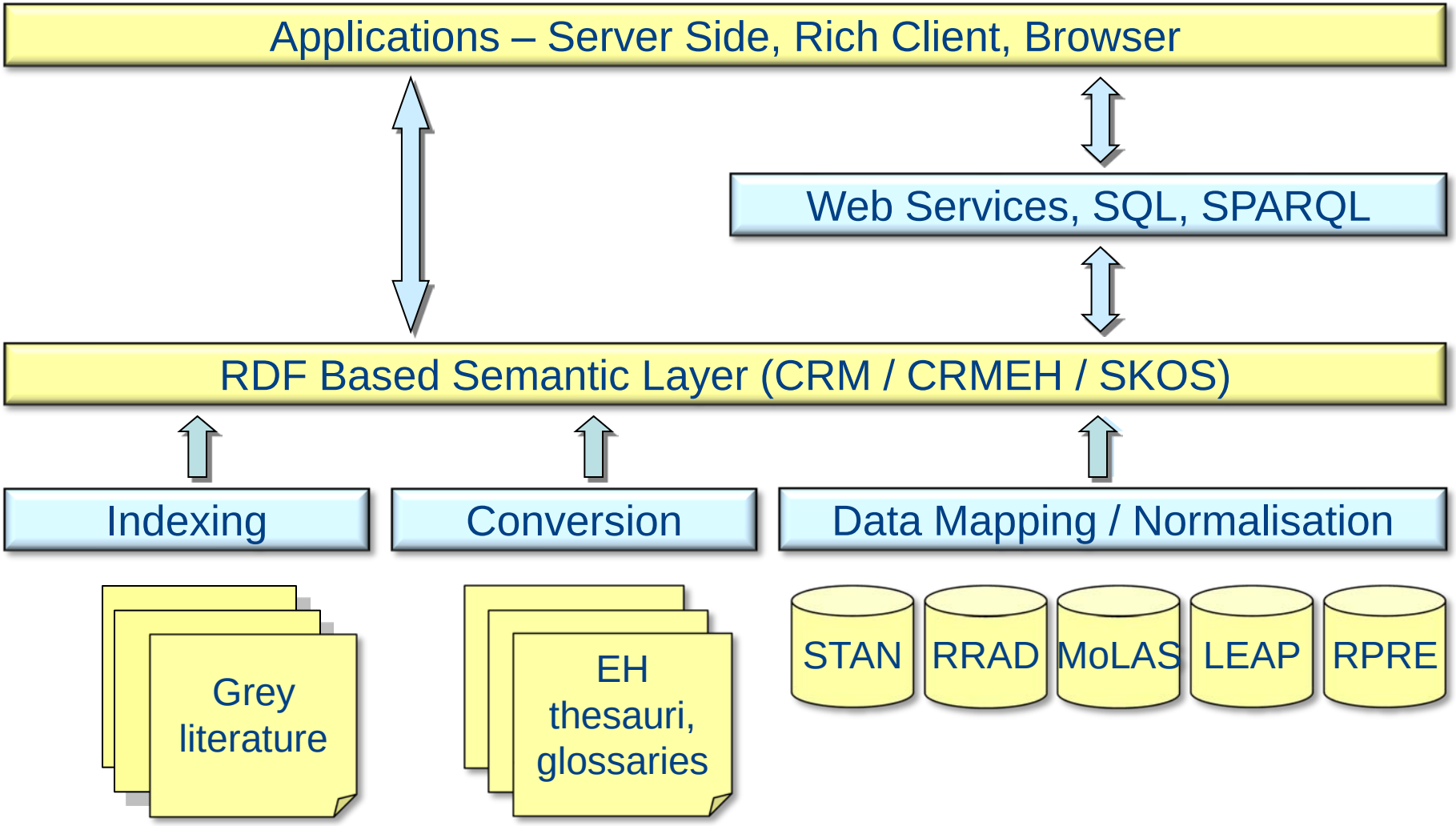
- Current situation - fragmented datasets with different terminology

- Lack of semantic interoperability and cross search

- Need for integrative metadata framework
  CIDOC CRM (ISO standard) as high level, core ontology
  together with the CRM-EH archaeological extension of the CRM

  along with relevant EH thesauri and glossaries

STAR Project - General Architecture



Knowledge Organization Systems

- English Heritage thesauri
- Reengineering to ontology would be significant effort (and would change them)
- STAR holds thesauri in SKOS

  Applies concepts to instances of CRM classes via E55 Type

EH Monuments Type Thesaurus

- WORKHOUSE
- RUBBISH PIT
- SADDLERY
- SERGEANTS MESS
- SERVANTS HALL
- SERVICE WING
- SETTLEMENT
  - CONSTRUCTION CAMP
  - CRANNOG
  - DESERTED SETTLEMENT
  - DISPLACED PERSONS CAMP
  - ENCLOSED SETTLEMENT
    - BURH
    - CLOTHES LINE ENCLOSURE
    - ENCLOSED HUT CIRCLE SETTLEMENT
    - ENCLOSED OPPIDUM
    - HILLFORT
      - BIVALLATE HILLFORT
      - MULTIPLE ENCLOSURE FORT
      - MULTIVALLATE HILLFORT
      - UNIVALLATE HILLFORT
    - HILLTOP ENCLOSURE
      - PALISADED HILLTOP ENCLOSURE
    - OPPIDUM
      - ENCLOSED OPPIDUM
    - PALISADED ENCLOSURE
    - PALISADED HOMESTEAD
    - PALISADED SETTLEMENT
    - PROMONTORY FORT
      - CLIFF CASTLE
    - ROUND
  - EXTRA MURAL SUBURB
  - HAMLET
  - HOMESTEAD
  - HOUSING ESTATE

Natural Language Processing (NLP)

of archaeological grey literature

Extract key concepts in same semantic representation as for data.

Allows unified searching of different datasets and grey literature

in terms of same underlying conceptual structure

“ditch containing prehistoric pottery dating to the Late Bronze Age”

EHE1002.ContextFindProductionEvent	
prehistoric pottery dating to the Late Bronze Age	
EHE0009.ContextFind	EHE0039.TimeSpanAppellation
pottery [#ehg027.2]	Late Bronze Age [#134734]

EHE1004.ContextFindDepositionEvent	
ditch containing prehistoric pottery	
EHE0007.Context	EHE0009.ContextFind
ditch [#ehg003.20]	pottery [#ehg027.2]

Information extraction is context dependent

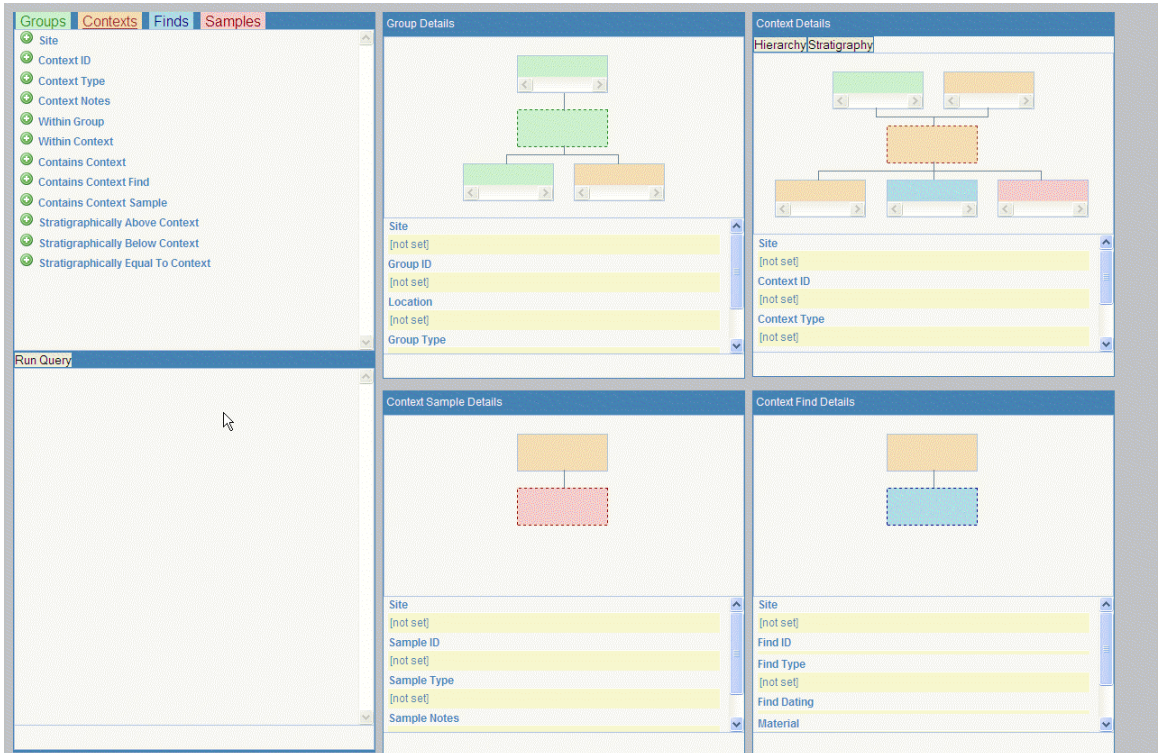
Annotation terms – ontology not an instance relationship
but a less certain relationship

form of ring ditches or ploughed out barrows, to the south of the study area.
1.15 A number of stray finds of Mesolithic, Neolithic and Bronze Age stone and flint implements have been recorded within 1km to the north and east of the present site (SMR nos 6531, 6591, 6595 and 6598). No evidence of settlements of these dates has been recovered from the vicinity of the study area.
1 Roman
1.16 Evidence for Roman activity in the immediate vicinity of the study area is relatively sparse. No settlement of this period has yet been identified nearby, but an archaeological evaluation on land at Tunbridge Lane, approximately 700m to the north-east, revealed a number of ditched enclosures, possibly agricultural in purpose (Seddon, 2000). The fact that modern field boundaries follow the alignment of the Roman enclosure ditches suggests that the basic layout of the landscape may have changed very little since then. The quantity of pottery sherds and other finds recovered from the ditches suggest occupation in the vicinity. Pottery sherds were also recovered by a metal detectorist on a building site, approximately 550m to the north of the study area (SMR no 6586).
Land off Bell Road, Bottisham, Cambs Evaluation Report
HN310\dba.sam Page 5
1.17 The remains of a large building were discovered in Swaffham Bulbeck at NGR TL 559 613, approximately 2.5km to the north-east of the present site. The evidence included ceramic roofing tile and opus signinum.
1.18 The lodes, or canals, running south-east from the river Cam to Swaffham Bulbeck and Lode were originally thought to date to the Roman period. The lack of major Roman occupation sites close to the lodes makes this unlikely. David Hall (Hall, 1996, p112) has demonstrated that they are much more likely to be Saxon or early medieval in date. The southern end of the canal into Lode is approximately 3km north-west of the present site.
1 Anglo-Saxon
1.19 Although no finds of Saxon date have been recovered from the vicinity of the present

STAR Demonstrator – search for a conceptual pattern

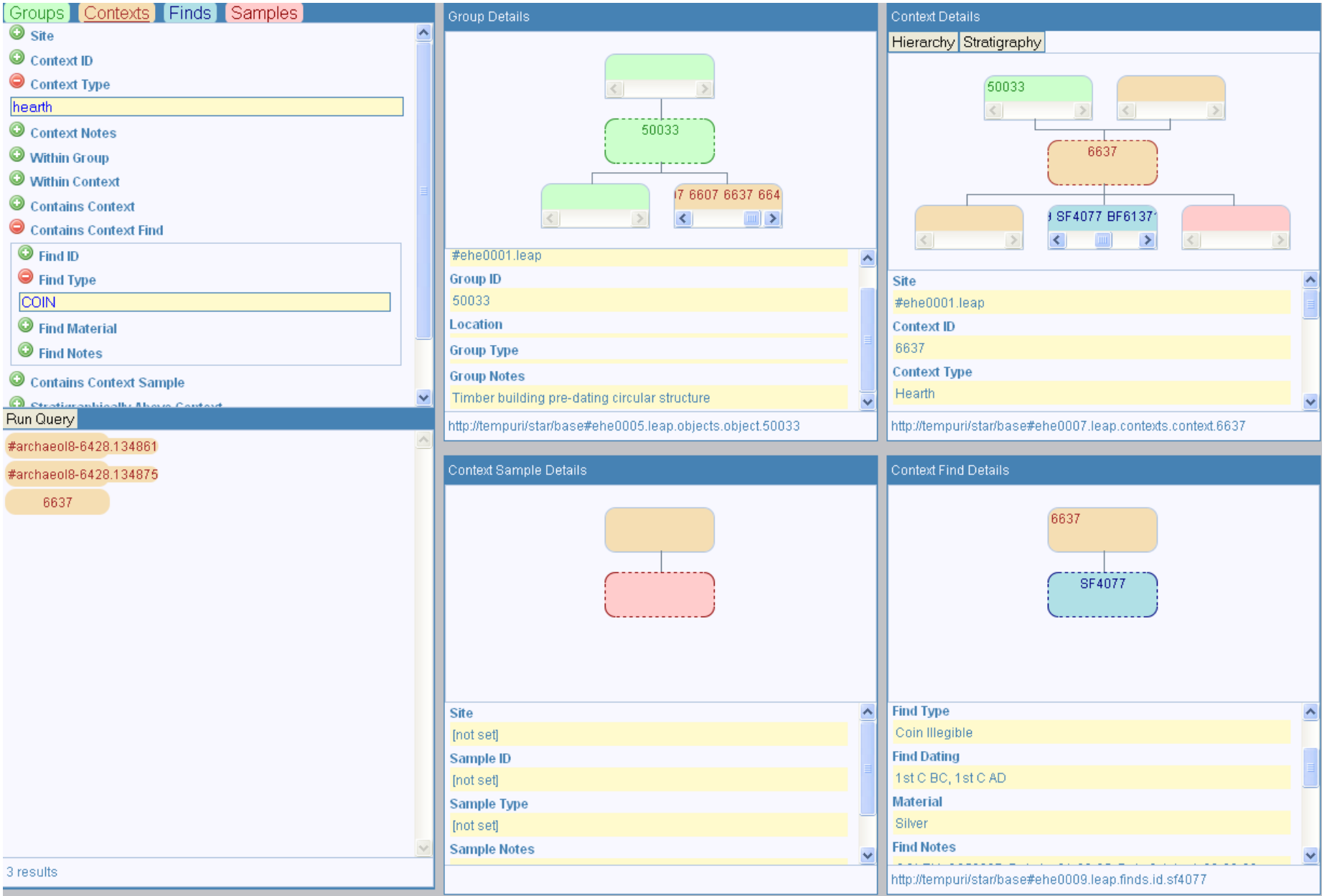
A research publication on one of the (Silchester Roman) datasets used in STAR discusses the finding of a *coin* within a *hearth*.
-- does the same thing occur in any of the grey literature reports?

Requires comparison of extracted data with NLP indexing
in terms of the ontology and the vocabularies.
http://hypermedia.research.glam.ac.uk/resources/star-demonstrator/



STAR Demonstrator – search for a conceptual pattern

Research paper reports finding a *coin in hearth* – exist elsewhere?



STELLAR outcomes

- Make it easier to map and extract archaeological excavation datasets to CIDOC CRM and CRM-EH archaeological extension of the ontology
- Generalise the data extraction tools produced by STAR so third party data providers can use them
- Develop methods for mapping and extraction of archaeological datasets into RDF/XML conforming to CIDOC CRM and CRM-EH ontology with unique global identifiers for entities and concepts (http URIs) for publication as linked data
- Resulting linked data available from ADS website http://data.archaeologydataservice.ac.uk
- Freely available tools and guidelines/tutorials http://hypermedia.research.glam.ac.uk/kos/STELLAR/

STELLAR outcomes

- In practice mapping to CRM has tended to require specialist knowledge of the ontology and been resource intensive

- Given the wide scope of the CRM, it is possible to make multiple valid mappings depending on the intended purpose and focus of the mappings

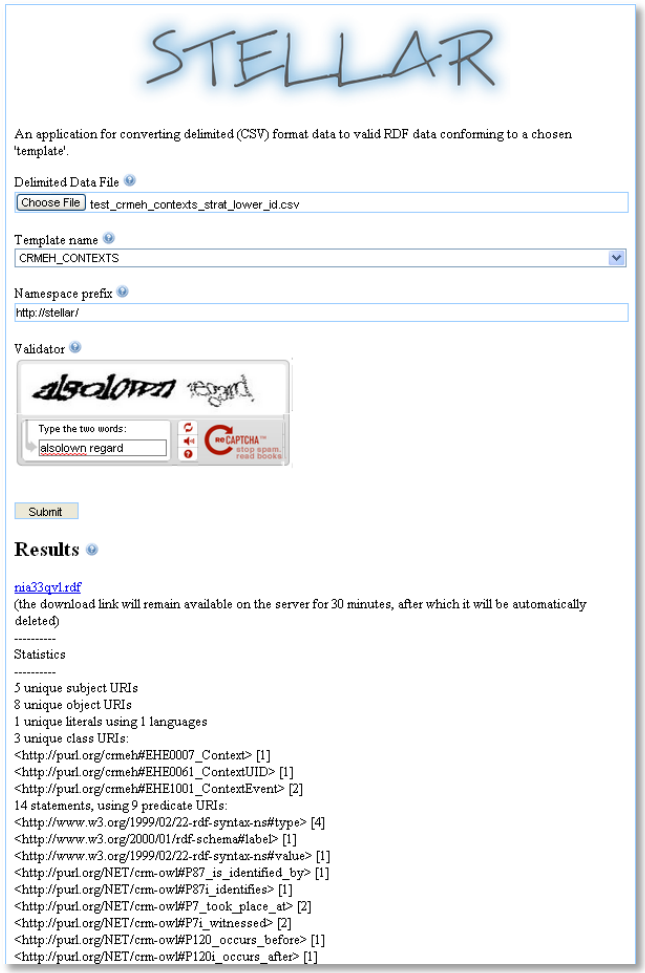
- STELLAR tools convert archaeological data to CRM/RDF in a consistent manner, without requiring detailed knowledge of the underlying ontology

- User chooses a template for a particular data pattern and supplies the corresponding input from their database (combination of optional elements with a mandatory ID)

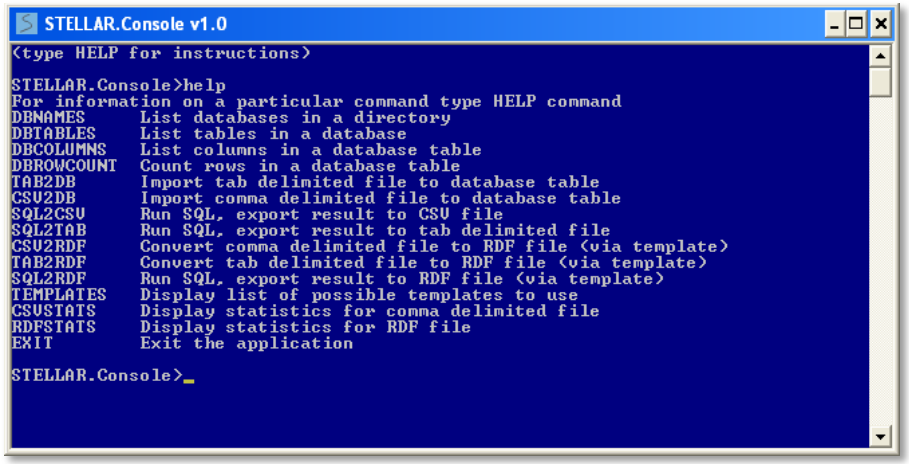
- STELLAR templates for
  - CRM-EH archaeological extension to the CIDOC CRM
  - Some more general CIDOC CRM templates conforming to the CLAROS Project format
  - SKOSifying a glossary/thesaurus connected with the dataset

STELLAR applications

http://hypermedia.research.glam.ac.uk/resources/STELLAR-applications/

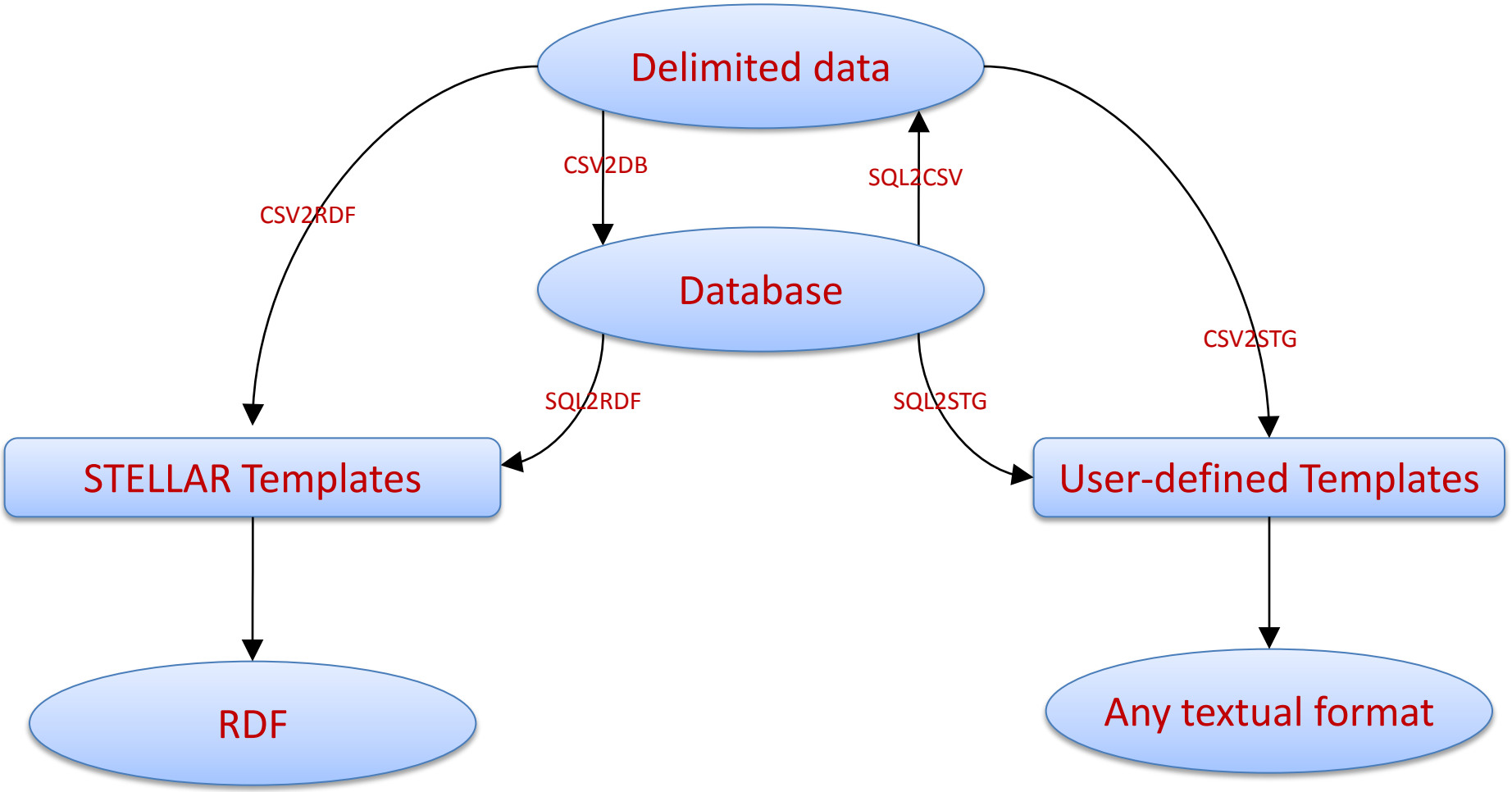


STELLAR.Web



STELLAR.Console

Data conversions



User defined templates

CSV2STG

```
isbn,title,author
1234567890, Winnie the Pooh, A.A. Milne
2345678901, Alice in Wonderland, Lewis Carrol
3456789012, The Cat in the Hat, Dr. Seuss
```

Delimited (CSV) data file

CSV2STG

```
group books_to_html;

HEADER() ::= <<
      <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
      <html>
      <head>
            <title>List of books</title>
      </head>
      <body>
>>

FOOTER() ::= "</body></html>"

MAIN(data) ::= <<
$HEADER()$

<h1>List of books:</h1>
<table border=1>
$data:{ record |
      <tr>
      <td>$record.isbn$</td>
      <td>$record.title$</td>
      <td>$record.author$</td>
      </tr>
}$
</table>

$FOOTER()$
 >>
```

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<html>
<head>
   <title>List of books</title>
</head>
<body>

<h1>List of books:</h1>
<table border=1>

   <tr>
   <td>1234567890</td>
   <td>Winnie the Pooh</td>
   <td>A.A. Milne</td>
   </tr>

   <tr>
   <td>2345678901</td>
   <td>Alice in Wonderland</td>
   <td>Lewis Carrol</td>
   </tr>

   <tr>
   <td>3456
   <td>The C
   <td>Dr. Se
   </tr>

</table>

</body></htr
```

List of books:

1234567890	Winnie the Pooh	A.A. Milne
2345678901	Alice in Wonderland	Lewis Carrol
3456789012	The Cat in the Hat	Dr. Seuss

Template to convert the data to HTML

```
group books_to_rdf;

HEADER() ::= <<
<?xml version="1.0" encoding="UTF-8"?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
   xmlnd:dc="http://purl.org/dc/elements/1.1/">
>>

FOOTER() ::= "</rdf:RDF>"

MAIN(data) ::= <<
$HEADER()$

$data:{ record |
      <rdf:Description rdf:about="http://tmp
            <dc:identifier>ISBN $record.isbn$
            <dc:title>$record.title$</dc:title>
            <dc:creator>$record.author$</d
            <dc:format>Book</dc:format>
      </rdf:Description>
}$

$FOOTER()$
>>
```

```
<?xml version="1.0" encoding="UTF-8"?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
   xmlnd:dc="http://purl.org/dc/elements/1.1/">

<rdf:Description rdf:about="http://tmp/1234567890">
      <dc:identifier>ISBN 1234567890</dc:identifier>
      <dc:title>Winnie the Pooh</dc:title>
      <dc:creator>A.A. Milne</dc:creator>
      <dc:format>Book</dc:format>
</rdf:Description>

<rdf:Description rdf:about="http://tmp/2345678901">
      <dc:identifier>ISBN 2345678901</dc:identifier>
      <dc:title>Alice in Wonderland</dc:title>
      <dc:creator>Lewis Carrol</dc:creator>
      <dc:format>Book</dc:format>
</rdf:Description>

<rdf:Description rdf:about="http://tmp/3456789012">
      <dc:identifier>ISBN 3456789012</dc:identifier>
      <dc:title>The Cat in the Hat</dc:title>
      <dc:creator>Dr. Seuss</dc:creator>
      <dc:format>Book</dc:format>
</rdf:Description>

</rdf:RDF>
```

Template to convert the data to RDF

Using the RDF data

```
<crmeh:EHE0007_Context rdf:about="http://stellar/silchester/EHE0007_1017">
     <rdfschema:label>1017</rdfschema:label>
     <ns:P87_is_identified_by>
          <crmeh:EHE0061_ContextUID rdf:about="http://stellar/silchester/EHE0061_1017">
               <rdf:value>1017</rdf:value>
               <ns:P87i_identifies rdf:resource="http://stellar/silchester/EHE0007_1017" />
          </crmeh:EHE0061_ContextUID>
     </ns:P87_is_identified_by>
     <ns:P7i_witnessed rdf:resource="http://stellar/silchester/EHE1001_1017" />
</crmeh:EHE0007_Context>
<crmeh:EHE1001_ContextEvent rdf:about="http://stellar/silchester/EHE1001_1017">
     <ns:P7_took_place_at rdf:resource="http://stellar/silchester/EHE0007_1017" />
     <ns:P120i_occurs_after rdf:resource="http://stellar/silchester/EHE1001_1015" />
     <ns:P120i_occurs_after rdf:resource="http://stellar/silchester/EHE1001_1302" />
</crmeh:EHE1001_ContextEvent>
```

RDF data output from STELLAR

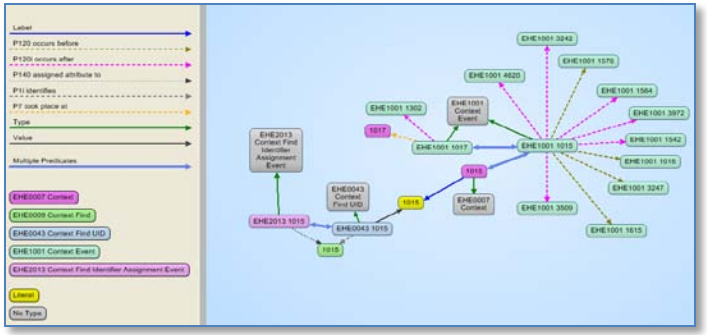
```
PREFIX ecrm: <http://erlangen-crm.org/101001/>
PREFIX crmeh: <http://purl.org/crmeh#>

SELECT DISTINCT ?g
WHERE
   {
   ?f a crmeh:EHE0009_ContextFind;
      ecrm:P45_consists_of [rdfs:label 'Animal Bone'];
      ecrm:P53_has_former_or_current_location ?c .
   ?c a crmeh:EHE0007_Context;
      ecrm:P89_falls_within ?g .
   ?g a crmeh:EHE0005_Group .
   }
LIMIT 5
```

SPARQL queries

RDF application / triple store

Linked data browsers



RDF enabled applications

Machine readable vs machine understandable

What we say to the machine:

```
<h1>The Cat in the Hat</h1>
<ul>
    <li>ISBN: 0007158440</li>
    <li>Author: Dr. Seuss</li>
    <li>Publisher: Collins</li>
</ul>
```

What the machine understands:

```
<h1>ασδ πλυ βγ ιτη μψσ</h1>
<ul>
    <li>φωφρ: 0007158440</li>
    <li>ωυψροκ: Δρ. Σευσσ</li>
    <li>Πυβλισηερ: Χολλινσ</li>
</ul>
```

(more) machine understandable

What we say to the machine:

```
<h1>Title:The Cat in the Hat</h1>
<ul>
    <li>ISBN: 0007158440</li>
    <li>Author: Dr. Seuss</li>
    <li>Publisher: Collins</li>
</ul>
```

What the machine understands:

```
<h1>ασδ πλυ βγ ιτη μψσ</h1>
<ul>
    <li>φωφρ: 0007158440</li>
    <li>ωυψροκ: Δρ. Σευσσ</li>
    <li>Πυβλισηερ: Χολλινσ</li>
</ul>
```

(getting more) machine understandable

What we say to the machine:

```
<h1>Title:The Cat in the Hat</h1>
<ul>
    <li>ISBN: 0007158440</li>
    <li>Author: Dr. Seuss</li>
    <li>Publisher: Collins</li>
</ul>
```

What the machine understands:

```
<h1>ασδ πλυ βγ ιτη μψσ</h1>
<ul>
    <li>φωφρ: 0007158440</li>
    <li>ωυψροκ: Δρ. Σευσσ</li>
    <li>Πυβλισηερ: Χολλινσ</li>
</ul>
```

(getting more) machine understandable

What we say to the machine:

```
<h1>Title:The Cat in the Hat</h1>
<ul>
    <li>ISBN: 0007158440</li>
    <li>Author: Dr. Seuss</li>
    <li>Publisher: Collins</li>
</ul>
```

Theodor Geisel

What the machine understands:

```
<h1>ασδ πλυ βγ ιτη μψο</h1>
<ul>
    <li>φωφρ: 0007158440</li>
    <li>ωυψροκ: Δρ. Σευσσ</li>
    <li>Πυβλισηερ: Χολλινσ</li>
</ul>
```

Book ***ID***

Author ***Publisher***

metadata structure (ontology)

vocabularies for terminology and knowledge organization

(getting more) machine understandable

Complementary use?

What we say to the machine:

```
<h1>Title:The Cat in the Hat</h1>
<ul>
    <li>ISBN: 0007158440</li>
    <li>Author: Dr. Seuss</li>
    <li>Publisher: Collins</li>
</ul>
```

What the machine understands:

```
<h1>ασδ πλυ βγ ιτη μψσ</h1>
<ul>
    <li>φϖϕρ: 0007158440</li>
    <li>ϖυψροκ: Δρ. Σευσσ</li>
    <li>Πυβλισηερ: Χολλινσ</li>
</ul>
```

Book *ID*

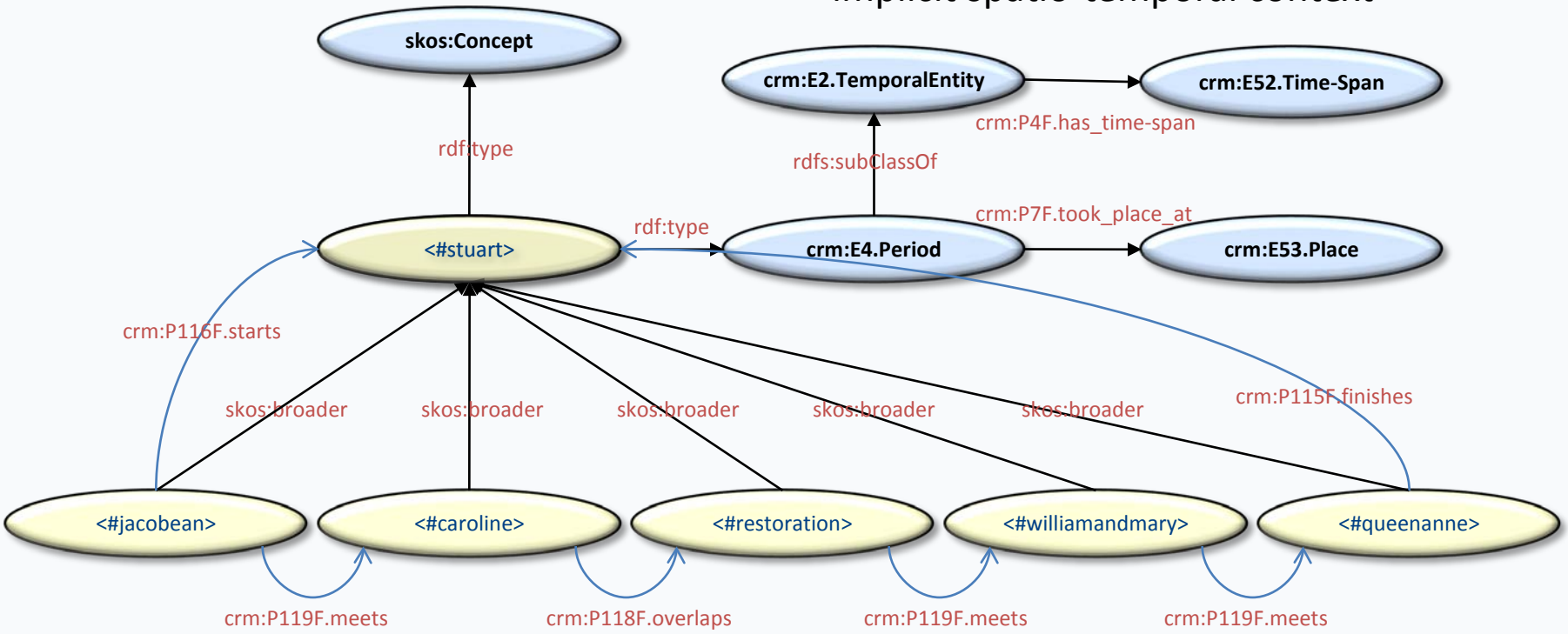
Author *Publisher*

metadata element sets (ontology)

Value vocabularies (KOS)

Combining SKOS Concepts + CRM Classes

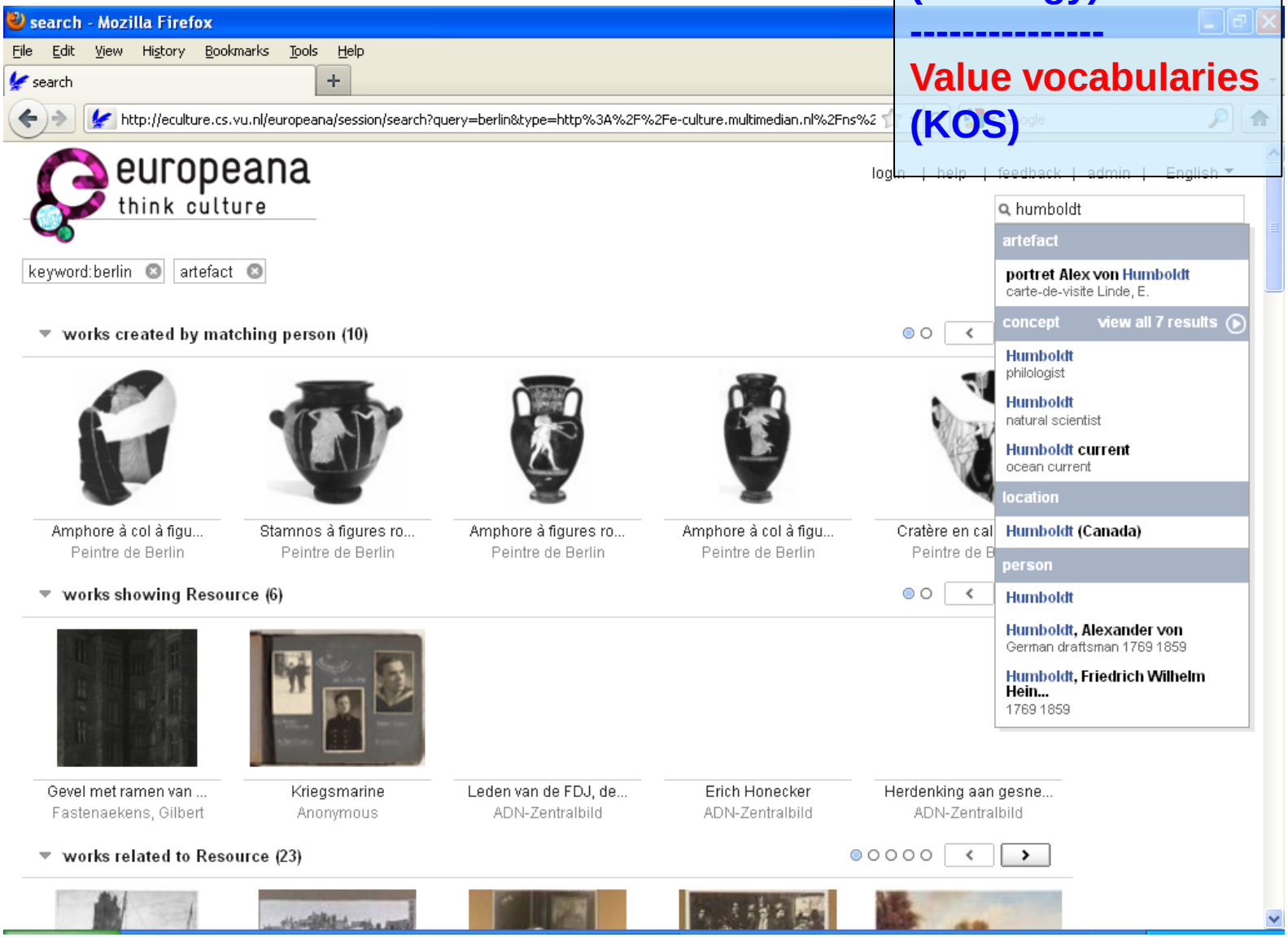
Time period concepts have
implicit spatio-temporal context



Complementary use?

metadata element sets (ontology)

Value vocabularies (KOS)



Complementary use?

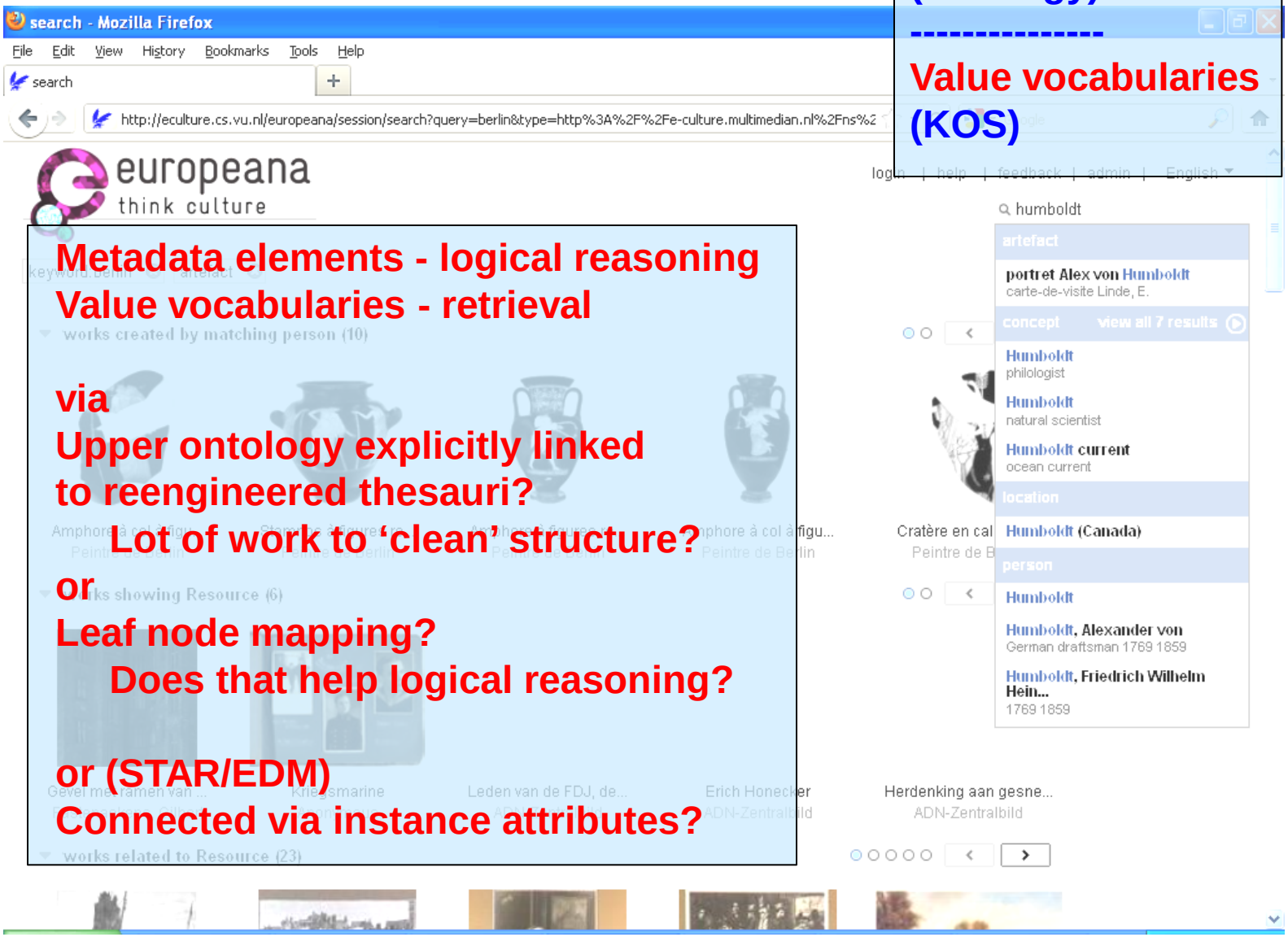
metadata element sets (ontology)

Value vocabularies (KOS)

Metadata elements - logical reasoning
Value vocabularies - retrieval

via
Upper ontology explicitly linked to reengineered thesauri?
Lot of work to 'clean' structure?
or
Leaf node mapping?
Does that help logical reasoning?

or (STAR/EDM)
Connected via instance attributes?



Complementary use?

metadata element sets (ontology)

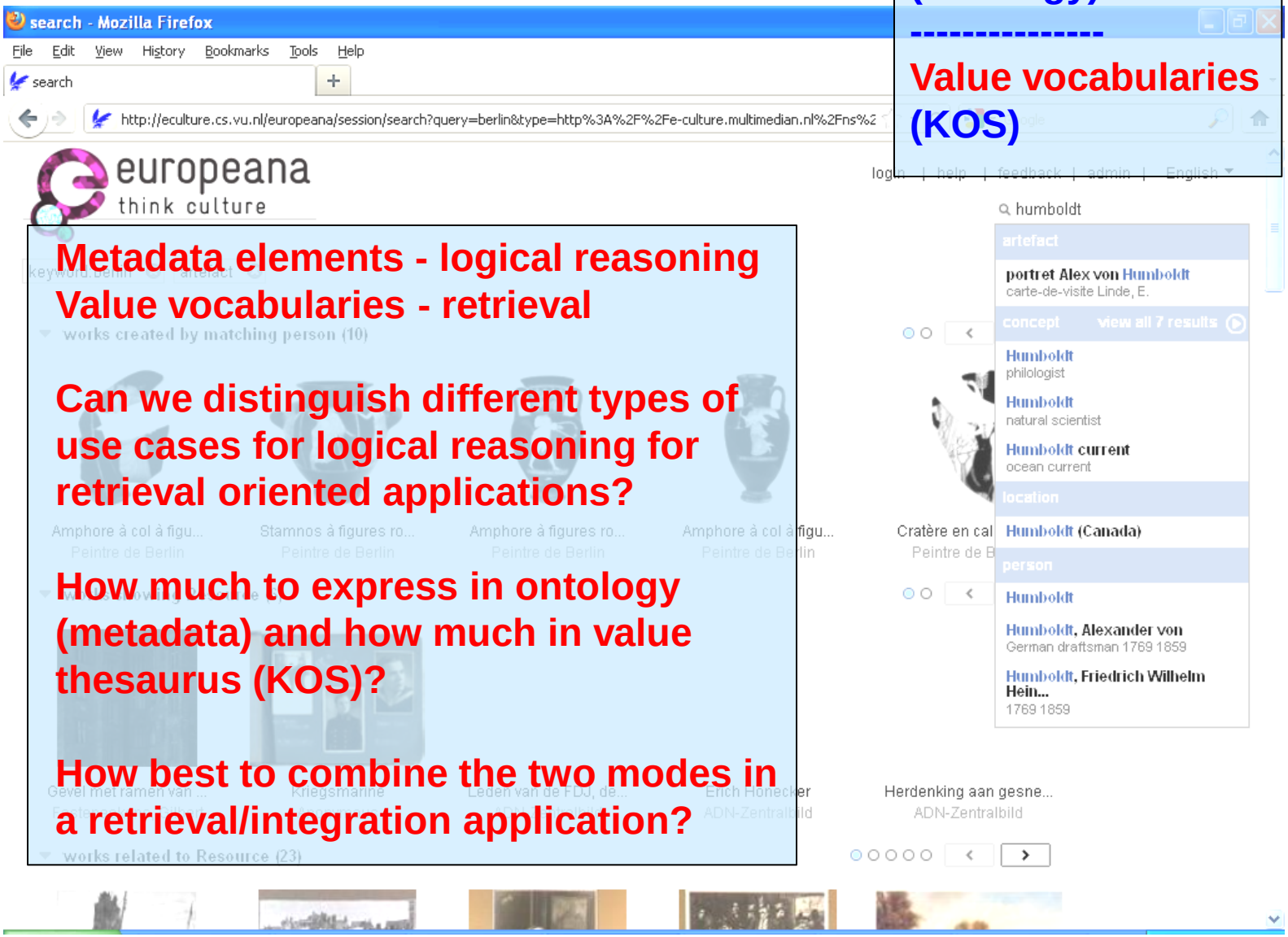
Value vocabularies (KOS)

Metadata elements - logical reasoning
Value vocabularies - retrieval

Can we distinguish different types of use cases for logical reasoning for retrieval oriented applications?

How much to express in ontology (metadata) and how much in value thesaurus (KOS)?

How best to combine the two modes in a retrieval/integration application?



Contact Information

Douglas Tudhope
Faculty of Advanced Technology
University of Glamorgan
Pontypridd CF37 1DL
Wales, UK

dstudhope@glam.ac.uk

http://hypermedia.research.glam.ac.uk/kos/STAR/
http://hypermedia.research.glam.ac.uk/kos/STELLAR/
http://data.archaeologydataservice.ac.uk