

STAR - Semantic Technologies for Archaeological Resources



Arts & Humanities
Research Council

DANMARKS 
BIBLIOTEKSSKOLE

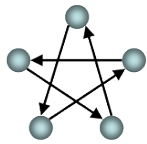


ENGLISH HERITAGE

University of Glamorgan

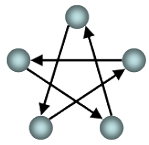
you live, you learn



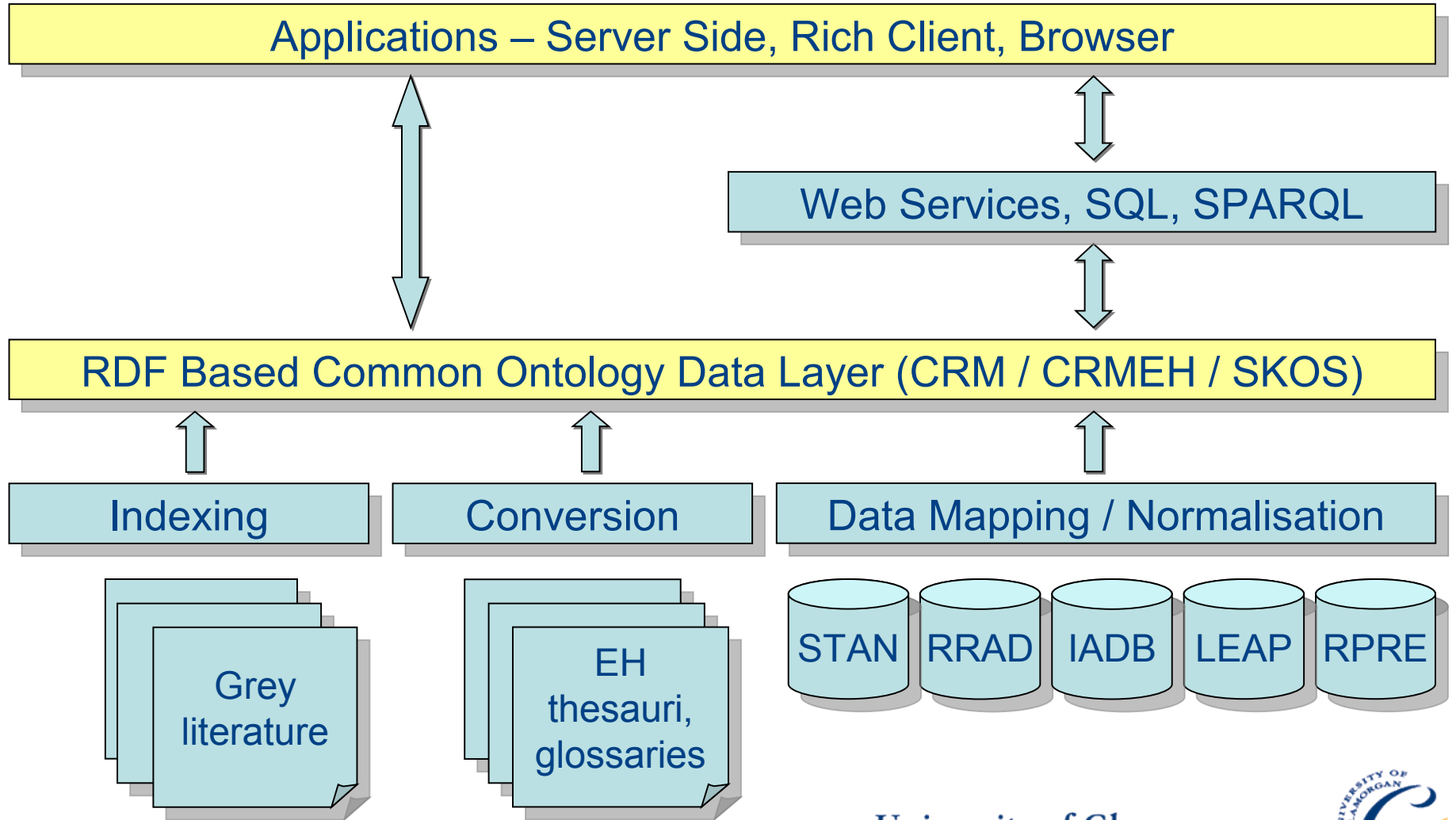


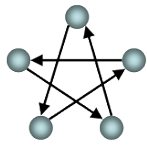
Introduction

- **Data Modelling (RDF)**
 - English Heritage Ontological Model (CRMEH)
 - English Heritage thesauri & glossaries (SKOS)
- **Data Mapping**
 - Mapping issues – domain specific vs. general model
 - Granularity, coverage
- **Data Extraction**
 - Custom extraction utility
 - Consolidation to RDF ‘triple store’ database
 - 5 archaeological datasets, 6 thesauri, 27 glossaries
- **Pilot Applications**
 - SKOS Web Service / Client applications
 - CRMEH Web Service / Client applications



STAR - General Architecture





Some Added Extras..

```

<rdf:RDF xml:lang="fr" xmlns:crm="http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#" x
http://www.w3.org/2000/01/rdf-schema#"
  <rdf:Description rdf:about="&crm;E1.CRM_Entity">
    <rdfs:label>Entité CRM</rdfs:label>
    <rdfs:comment>Il s'agit du concept abstrait de l'ensemble des entités qui cons
susceptibles d'être classées par « type », ce qui permet d'affiner encore la sous-c
des informations qu'il n'est pas possible de faire entrer dans le cadre formel des lie
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E2.Temporal_Entity">
    <rdfs:label>Entité Temporelle</rdfs:label>
    <rdfs:comment>Il s'agit d'une entité abstraite dont il n'est pas possible de donn
phénomènes qui sont limités dans le temps. Elle se subdivise en deux sous-classes
objet donné sur une durée plus ou moins longue.</rdfs:comment>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E3.Condition_State">
    <rdfs:label>État de Conservation</rdfs:label>
    <rdfs:comment>L'état d'un objet, sous l'angle de sa conservation matérielle su
« en ruine » constitue le type d'état matériel.</rdfs:comment>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E4.Period">
    <rdfs:label>Période</rdfs:label>
    <rdfs:comment>Une période se caractérise par une conjonction (intentionnelle
temps déterminés.
Les avis divergent sur la question de savoir si un « style » se définit par des caract
Exemples: Période glaciaire, l'Âge de bronze, la dynastie Ming, l'Impressionnisme, le
  </rdfs:comment>

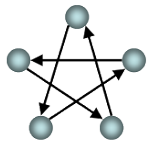
```

```

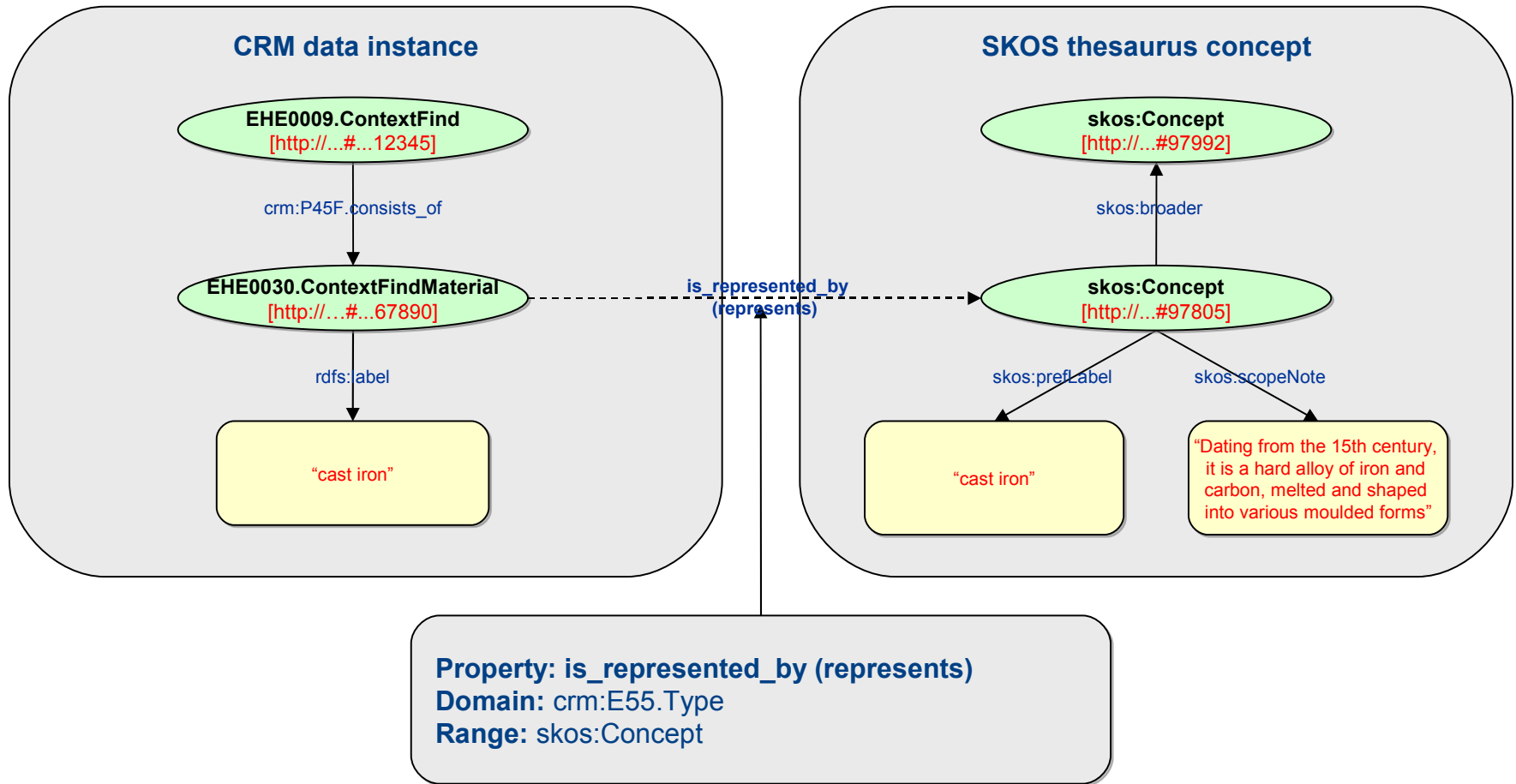
<rdf:RDF xml:lang="el" xmlns:crm="http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs
http://www.w3.org/2000/01/rdf-schema#"
  <rdf:Description rdf:about="&crm;E1.CRM_Entity">
    <rdfs:label>Οντότητα CIDOC CRM</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E2.Temporal_Entity">
    <rdfs:label>Έγχορηγη Οντότητα</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E3.Condition_State">
    <rdfs:label>Κατάσταση</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E4.Period">
    <rdfs:label>Περίοδος</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E5.Event">
    <rdfs:label>Συμβάν</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E6.Destruction">
    <rdfs:label>Καταστροφή</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E7.Activity">
    <rdfs:label>Δράση</rdfs:label>
  </rdf:Description>
  <rdf:Description rdf:about="&crm;E8.Acquisition">
    <rdfs:label>Απόκτηση</rdfs:label>
  </rdf:Description>

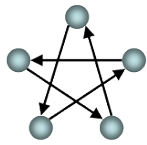
```

Multilingual labels and comments for CRM (RDFS version)

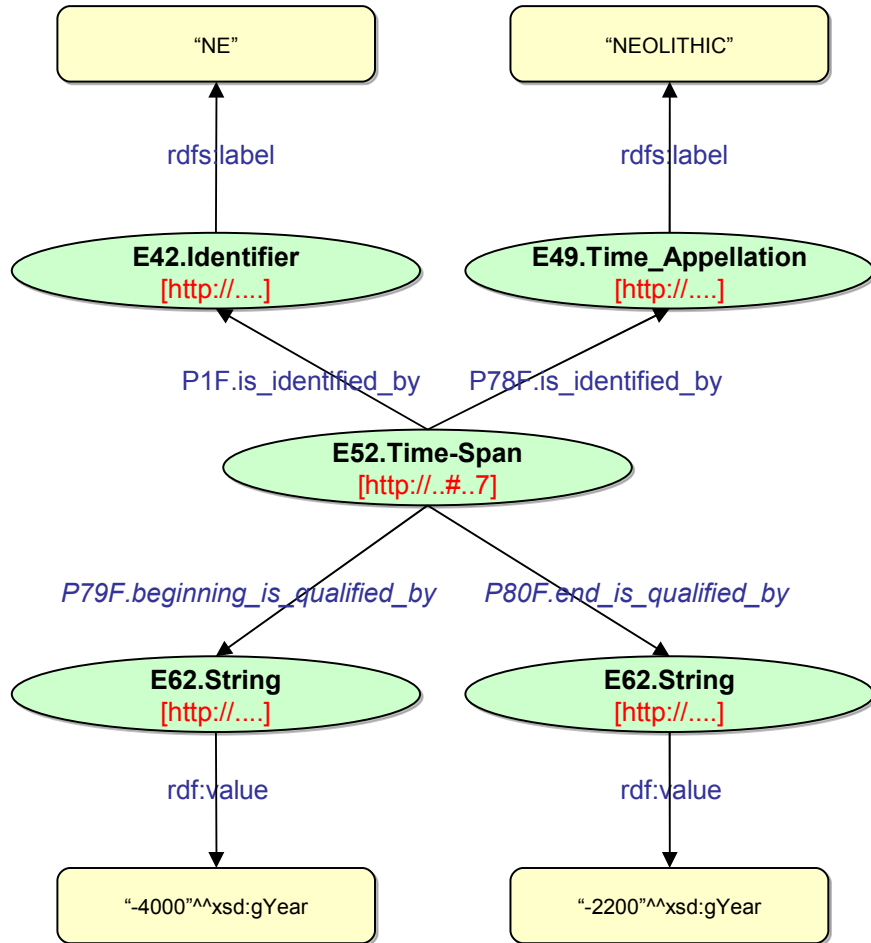


Linking CRM types to SKOS concepts



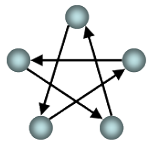


RCHME Archaeological Periods in CRM



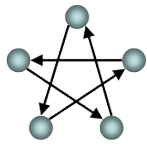
```
<crm:E52.Time-Span rdf:about="&base;E52.rchme.periods.termuid.7">
  <crm:P1F.is_identified_by>
    <crm:E42.Identifier rdf:about="&base;E49.rchme.periods.code.7">
      <rdfs:label>NE</rdfs:label>
    </crm:E42.Identifier>
  </crm:P1F.is_identified_by>
  <crm:P78F.is_identified_by>
    <crm:E49.Time_Appellation rdf:about="&base;E49.rchme.periods.term.7">
      <rdfs:label>NEOLITHIC</rdfs:label>
    </crm:E49.Time_Appellation>
  </crm:P78F.is_identified_by>
  <crm:P79F.beginning_is_qualified_by>
    <crm:E62.String rdf:about="&base;E62.rchme.periods.mindate.7">
      <rdf:value type="xsd:gYear">-4000</rdf:value>
    </crm:P79F.beginning_is_qualified_by>
  <crm:P80F.end_is_qualified_by>
    <crm:E62.String rdf:about="&base;E62.rchme.periods.maxdate.7">
      <rdf:value type="xsd:gYear">-2200</rdf:value>
    </crm:P80F.end_is_qualified_by>
</crm:E52.Time-Span>
```





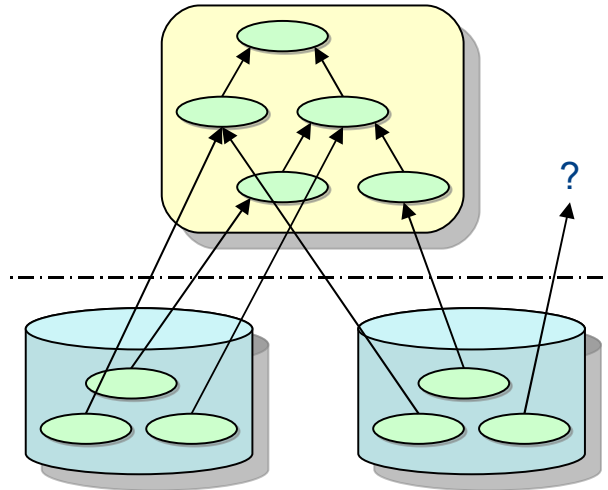
Data Mapping Exercise

- Mapped DB Field → CRMEH Entity
 - manual process, requires expert domain knowledge
- Needed to Consider
 - Granularity & completeness of mapping
 - Model coverage vs. dataset coverage
 - Event based vs. relational model
 - Events only implicit in mappings and datasets



Potential Data Mapping Problems

General schema (CRM)



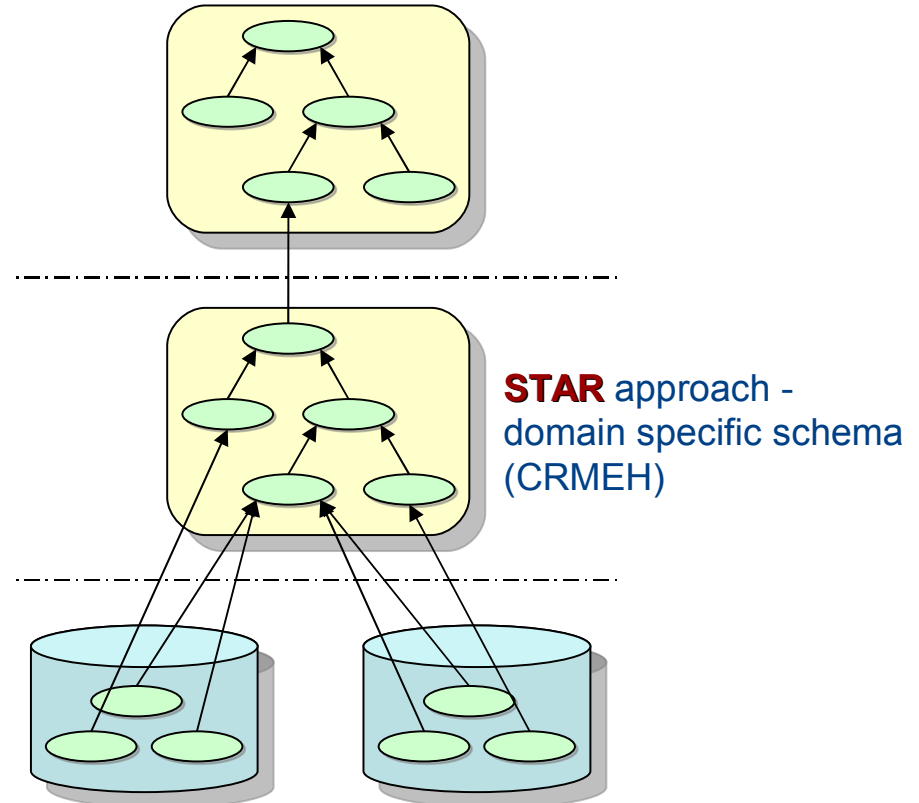
Domain specific datasets

“...the abstractness of the [CRM] concepts...makes them ambiguous to any human user.”

“...If several experts specify mappings independently...they will produce incompatible mappings and fail the goal of enabling interoperability.”

[from **BRICKS** FP6 IP Poster, ECDL2007]

General schema (CRM)



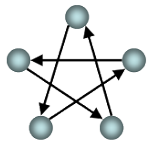
STAR approach -
domain specific schema
(CRMEH)

Domain specific datasets

University of Glamorgan



you live, you learn



Custom Data Extraction Application

SQLBuilder v1.0

File

Database: RRAD

Subject Predicate Object

```

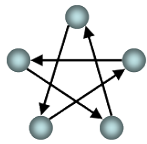
<?xml version="1.0"?>
<rdf:RDF xml:base="http://tempuri/star/base#"
xmlns:crm="http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#"
xmlns:crmeh="http://tempuri/star/crmeh#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<crmeh:EHE0007.Context rdf:about="http://tempuri/star/base#EHE0007.rrad.context.contextno.1">
<crm:P3F.has_note>
<crmeh:EHE0046.ContextNote rdf:about="http://tempuri/star/base#EHE0046.rrad.context.description.1">
<rdf:value>Upper ploughsoil over whole site no Sub-division for the convenience of finds processing '1' contains finds
contexts '3759', '3760' and '3763'.</rdf:value>
</crmeh:EHE0046.ContextNote>
</crm:P3F.has_note>
</crmeh:EHE0007.Context>
<crmeh:EHE0007.Context rdf:about="http://tempuri/star/base#EHE0007.rrad.context.contextno.10">
<crm:P3F.has_note>
<crmeh:EHE0046.ContextNote rdf:about="http://tempuri/star/base#EHE0046.rrad.context.description.10">
<rdf:value>Original recorded coordinates: 0980/0960</rdf:value>
</crmeh:EHE0046.ContextNote>
</crm:P3F.has_note>
</crmeh:EHE0007.Context>
Etc.
  
```

Generated RDF Data

Test SQL

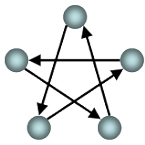
Write RDF...





Data Extraction and Consolidation

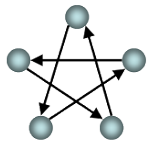
- **Data Extraction** – resultant files
 - 6 EH thesauri → 6 RDF files (SKOS)
 - 27 EH glossaries → 27 RDF files (SKOS)
 - 5 archaeological datasets → 305 RDF files (CRMEH)
- **Data Consolidation** - RDF “triple store” database
 - 1,148,882 entities, 2,998,005 triples
 - Diverse data held within single DB schema
 - CRM, CRMEH, SKOS, OWL, DC etc.
 - Platform independent import/export
 - RDF-XML, NTriples, Turtle
 - Custom extension for full text search
 - Inbuilt support for SPARQL querying



CRMEH querying - SPARQL

```
# Get contexts with associated samples containing notes mentioning charcoal
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX crm: <http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#>
PREFIX crmeh: <http://tempuri/star/crmeh#>
SELECT ?context ?sample ?notevalue
WHERE
{
  ?context a crmeh:EHE0007.Context .
  ?contextstuff crmeh:EHP3F.occupied ?context .
  ?samplingevent crm:P112F.diminished ?contextstuff; crm:P113F.removed ?sample .
  ?sample crm:P3F.has_note ?note .
  ?note rdf:value ?notevalue .
  FILTER (REGEX(?notevalue,"(?i)charcoal")).
}

<result>
  <binding name="sample">
    <uri>http://tempuri/star/base#ehe0018.rrad.sample.sampleno.10560</uri>
  </binding>
  <binding name="context">
    <uri>http://tempuri/star/base#ehe0007.rrad.context.contextno.4900</uri>
  </binding>
  <binding name="notevalue">
    <literal>Sample taken during excavation of pit (?) '4899', which contained some burnt material (burnt
    clay(?), flecks charcoal), sample was taken from middle of pit(?) 5 cms from bottom</literal>
  </binding>
</result>
<result>
  <binding name="sample">
    <uri>http://tempuri/star/base#ehe0018.rrad.sample.sampleno.10613</uri>
  </binding>
  <binding name="context">
    <uri>http://tempuri/star/base#ehe0007.rrad.context.contextno.3117</uri>
  </binding>
  <binding name="notevalue">
    <literal>Comparison with other samples from '3101' - charcoal</literal>
  </binding>
</result>
Etc.
```



CRMEH Client Application

STAR.CRM.WSClient v1.4

Suggest terms
brooch [Go]

Suggested terms

- CRUCIFORM BROOCH
- DISC BROOCH
- DOLPHIN BROOCH
- DRAGONESQUE BROOCH
- HEADSTUD BROOCH
- HOD HILL BROOCH
- Hook Norton Brooch
- KNEE BROOCH
- Lamberton Moor Brooch
- LANGTON DOWN BROOCH
- LONG BROOCH
- NAUHEIM DERIVATE BROOCH
- PENANNULAR BROOCH
- PLATE BROOCH
- POLDEN HILL BROOCH
- Rosette Brooch
- SAUCER BROOCH
- Simple One-Piece Brooch
- SMALL LONG BROOCH

Expanded query for selected term
+("PENANNULAR BROOCH" "BROOCH" "Fibula" "Brooch Spring")

Search Favourites
+brooch +(pennanular fantail >nauheim) [Go]

Legend

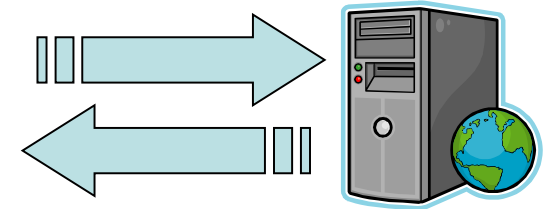
- Raunds Roman
- Raunds Prehistoric
- LEAP Silchester
- Show Colours

Property	Text
value	Nauheim derivative, the spring formed in one piece with the rest of the brooch. Two tur...

Nauheim derivative, the spring formed in one piece with the rest of the brooch. Two turns remain of a four turn spring made in one piece with a plain bow of sub-rectangular section which tapers to a narrow foot bearing an unperforated catch-plate. Type common in South Britain in the mid first century AD.

http://tempuri/star/base#e62.rrad.object.descriptivetext.1567

CRMEH Client Application



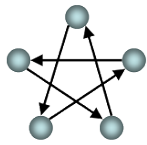
CRMEH Web Service

Search and browse seamlessly across multiple archaeological datasets

University of Glamorgan

you live, you learn





CRMEH Client Application

★ STAR.CRM.WSClient v1.4

Suggest terms

Go

Suggested terms

Search Favourites

plough| Go

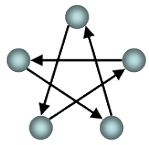
Legend

- Raunds Roman
- Raunds Prehistoric
- LEAP Silchester
- Show Colours

Expanded query for selected term

Property	Text



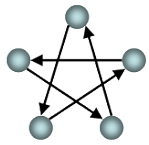


Browser Based Client Application

Semantic Technologies for Archaeological Resources

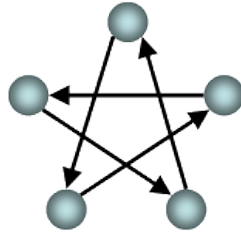
The screenshot displays a web application interface with two main panels:

- Term Suggester:** Contains a text input field with the word "axe" and a "Find" button. Below the input is a list of suggested terms: "STONE AXE FACTORY", "Axe Factory", "Axe Working Floor", "AXE MOULD", "AXE TRIMMING FLAKE", "POLE AXE", "THROWING AXE", "AXE (TOOL)", and "AXE (WEAPON)". A "Use" button is located at the bottom right of this panel.
- CRM Data Search:** Contains a text input field with the word "axe" and a "Search" button. Below the input is a large empty area, likely intended for search results.



STAR - Next Steps...

- Expose support for more complex query patterns
- Closer integration of thesauri into overall search process
- Browser-based application components (AJAX)
- Include further archaeological datasets
- Ontology driven information extraction from grey literature



STAR - Semantic Technologies for Archaeological Resources

<http://hypermedia.research.glam.ac.uk/kos/star/>

dstudhope@glam.ac.uk
cbinding@glam.ac.uk



Arts & Humanities
Research Council



ENGLISH HERITAGE

University of Glamorgan

you live, you learn

