

STAR

Semantic Technologies for Archaeological Resources

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



Arts & Humanities
Research Council

DANMARKS 
BIBLIOTEKSSKOLE

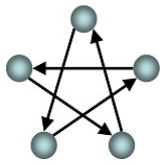


ENGLISH HERITAGE

University of Glamorgan

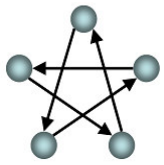
you live, you learn





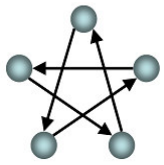
Project Outline

- **3 year AHRC funded project**
 - Started January 2007, finish December 2009
- **Collaborators**
 - English Heritage
 - RSLIS, Denmark
- **Aims**
 - *“To investigate the potential of semantic terminology tools for widening access to digital archaeology resources, including disparate datasets and associated grey literature”*
 - To demonstrate cross search and browsing at detailed, meaningful level

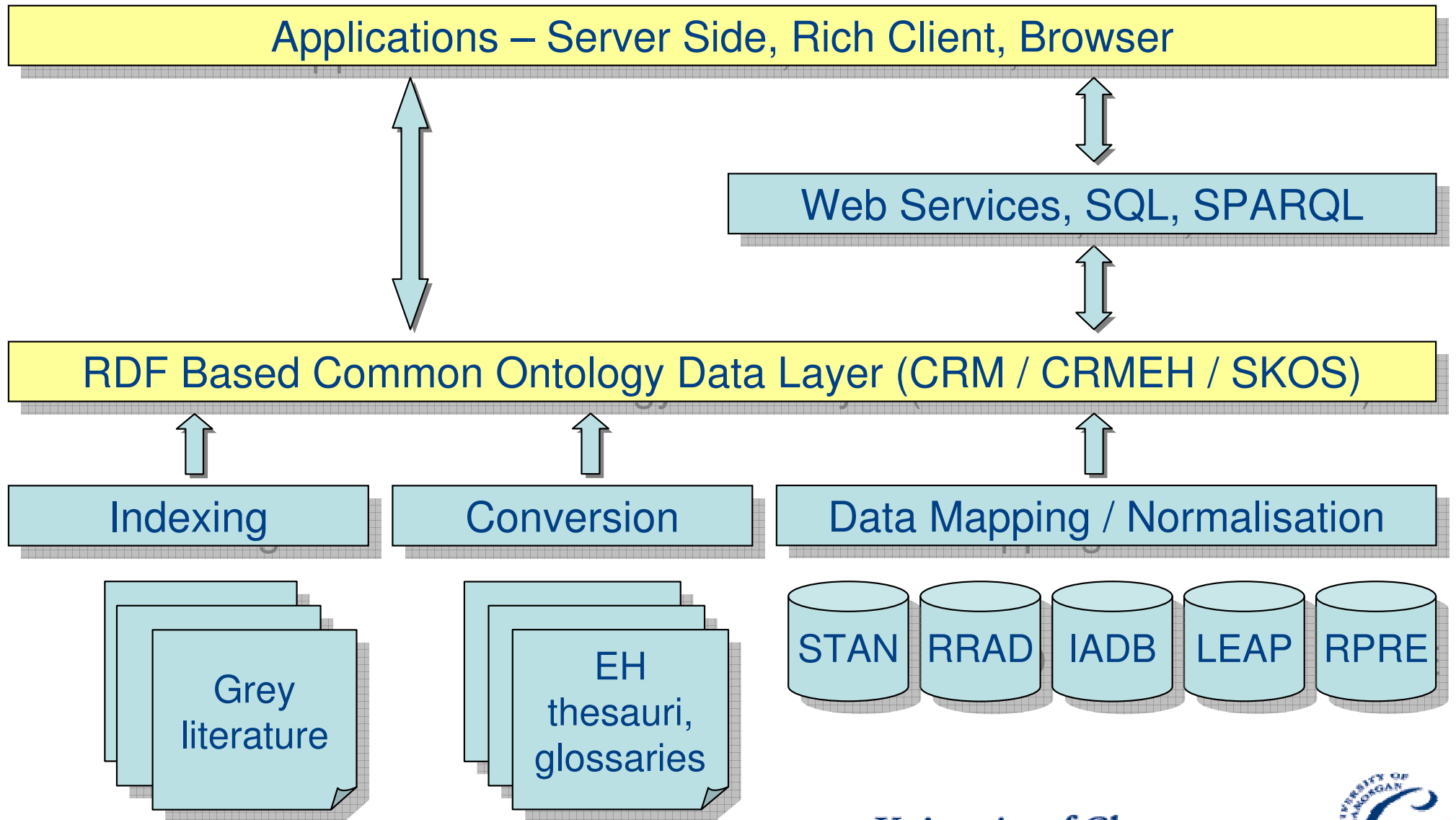


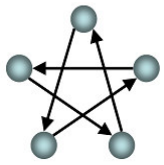
Introduction

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



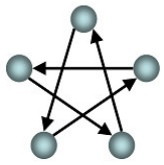
General Architecture





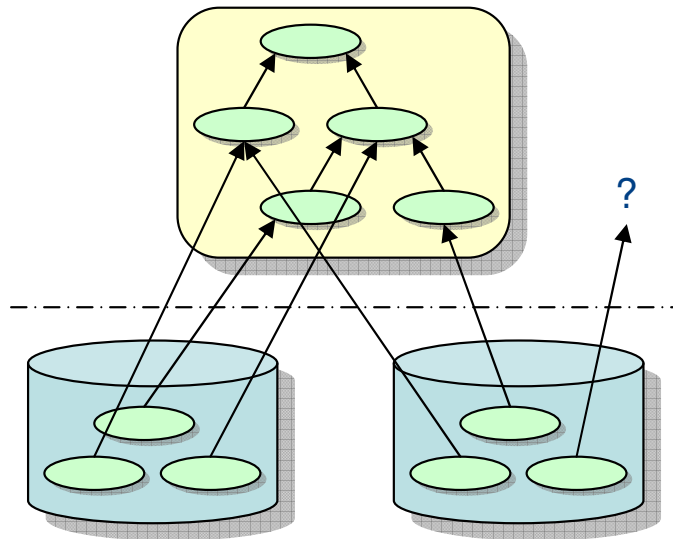
Data Modelling - RDF

- **CRM** [<http://cidoc.ics.forth.gr/>]
 - CIDOC Conceptual Reference Model
 - International standard ISO 21127:2006
- **CRMEH** [<http://hypermedia.research.glam.ac.uk/kos/CRM/>]
 - English Heritage Ontological Model
 - Extends CIDOC CRM for archaeological domain
- **SKOS** [<http://www.w3.org/2004/02/skos/>]
 - Simple Knowledge Organisation System
 - RDF representation of thesauri, glossaries, taxonomies, classification schemes etc.



Potential Data Mapping Problems

General schema (CRM)



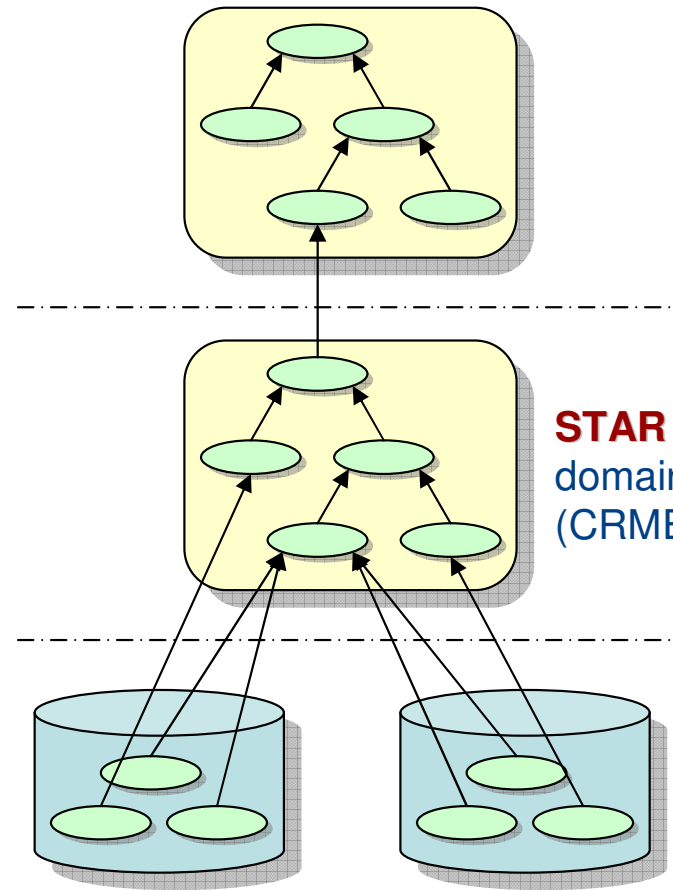
Domain specific datasets

"...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."

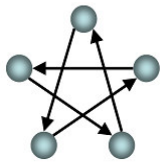
[from **BRICKS** FP6 IP Poster, ECDL2007]

General schema (CRM)



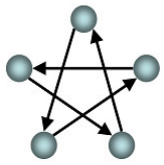
STAR approach -
domain specific schema
(CRMEH)

Domain specific datasets



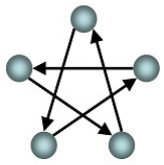
Data Mapping Exercise

- 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



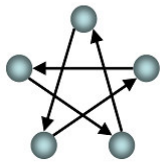
Data Normalisation Issues

- **Identifiers – URI References**
 - [Prefix]entity.database.table.column.rowid
 - <http://...#EHE0007.rrad.context.contextno.24845>
- **Coordinates – EH Recording Manual**
 - Easting/northing/AOD (in metres)
 - 110575/190865/120.48
- **Dates/Periods – W3C/ISO8601 ‘big-endian’ format**
 - CCYY-MM-DDThh:mm:ss
 - 2007-05-03T16:19:23
- **Controlled fields - Data cleansing / mapping**
 - EH Thesauri from EH DSU, Swindon
 - EH Glossaries from Recording Manual



Data Extraction - Scope

- Extraction of data to RDF triples
 - 5 archaeological datasets
 - Custom data extraction application
- Conversion of controlled terminology
 - 6 thesauri converted to SKOS
 - 27 glossaries created in SKOS
 - Created based on recording manuals
 - MultiTes XSL transformation to SKOS



Custom Data Extraction Application

SQL Builder v1.0

Database: RRAD

	Subject	Predicate	Object
Type			
Pre			
Column			
FROM clause			
WHERE clause			

```

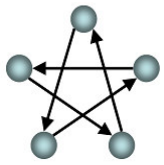
<?xml version="1.0"?>
<rdf:RDF xml:base="http://tempuri/star/base#"
xmlns:crm="http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#"
xmlns:crmeh="http://tempuri/star/crmeh#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<crmeh:EHE0007.Context rdf:about="http://tempuri/star/base#EHE0007.rrad.context.contextno.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

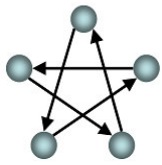
Write RDF...



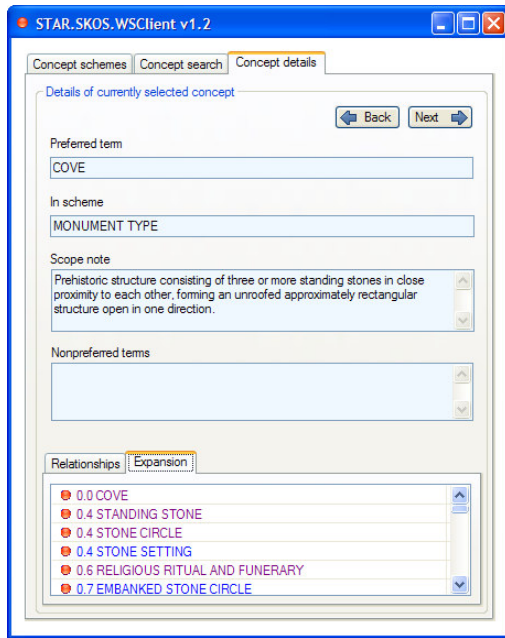


Data Extraction and Consolidation

- **Data Extraction** – resultant files
 - 6 EH thesauri → 6 RDF files
 - 27 EH glossaries → 27 RDF files
 - 5 archaeological datasets → 305 RDF files
- **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
 - Inbuilt support for SPARQL queries
 - Custom extension for full text search



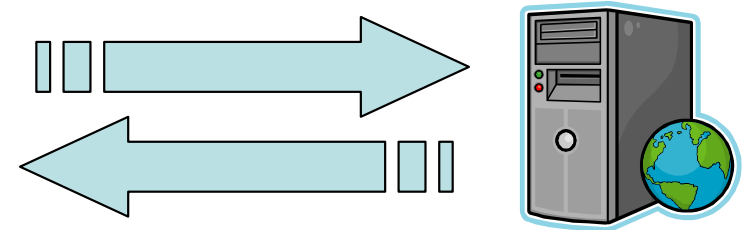
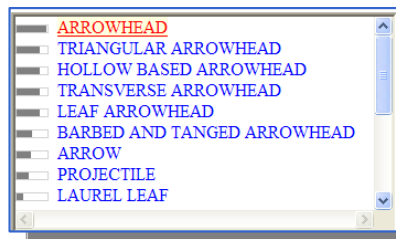
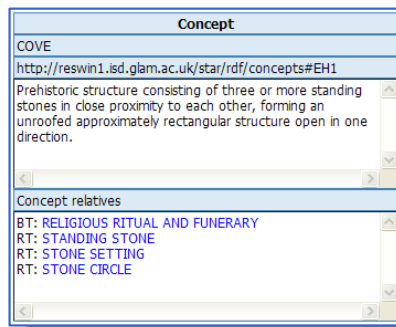
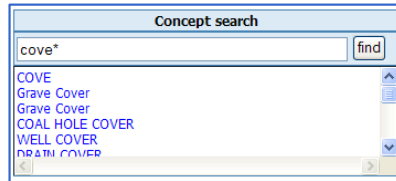
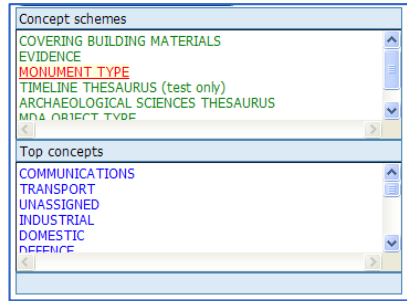
SKOS Web Service and Client Applications



Windows based client application

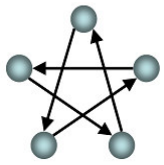
Web browser based components ('widgets')

SKOS Client Applications

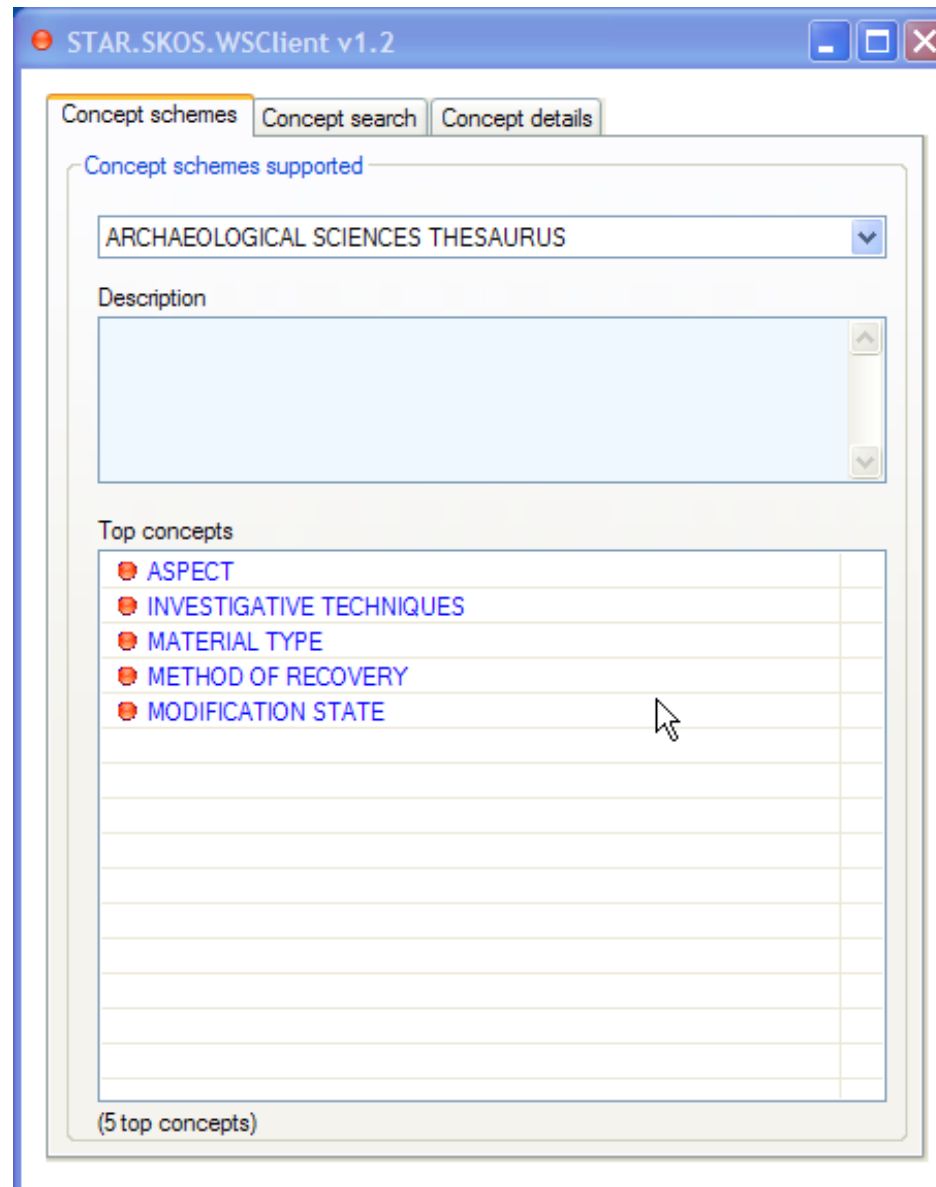


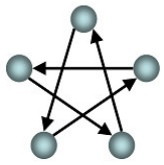
SKOS Web Service





SKOS Client – Windows Application





SKOS Client - Widgets

Concept search

Concept Search

Concept

BROOCH

<http://reswin1.isd.glam.ac.uk/star/rdf/concepts#EH128>

Ornament often with a hinged pin and catch, worn fastened to clothing.

Concept relatives

BT: JEWELLERY
NT: **ANNULAR BROOCH**
NT: BOW BROOCH
NT: DRAGONESQUE BROOCH
NT: LONG BROOCH
NT: DENANNIII AR BROOCH

Concept Details

Concept schemes

- EVIDENCE
- MONUMENT TYPE
- TIMELINE THESAURUS (test only)
- ARCHAEOLOGICAL SCIENCES THESAURUS
- MDA OBJECT TYPE
- MAIN BUILDING MATERIALS

Top concepts

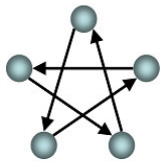
- ANIMAL EQUIPMENT
- AGRICULTURE AND SUBSISTENCE
- ARCHITECTURE
- FURNISHINGS AND FURNITURE
- TRANSPORT
- FOOD PREPARATION AND CONSUMPTION

Concept Schemes

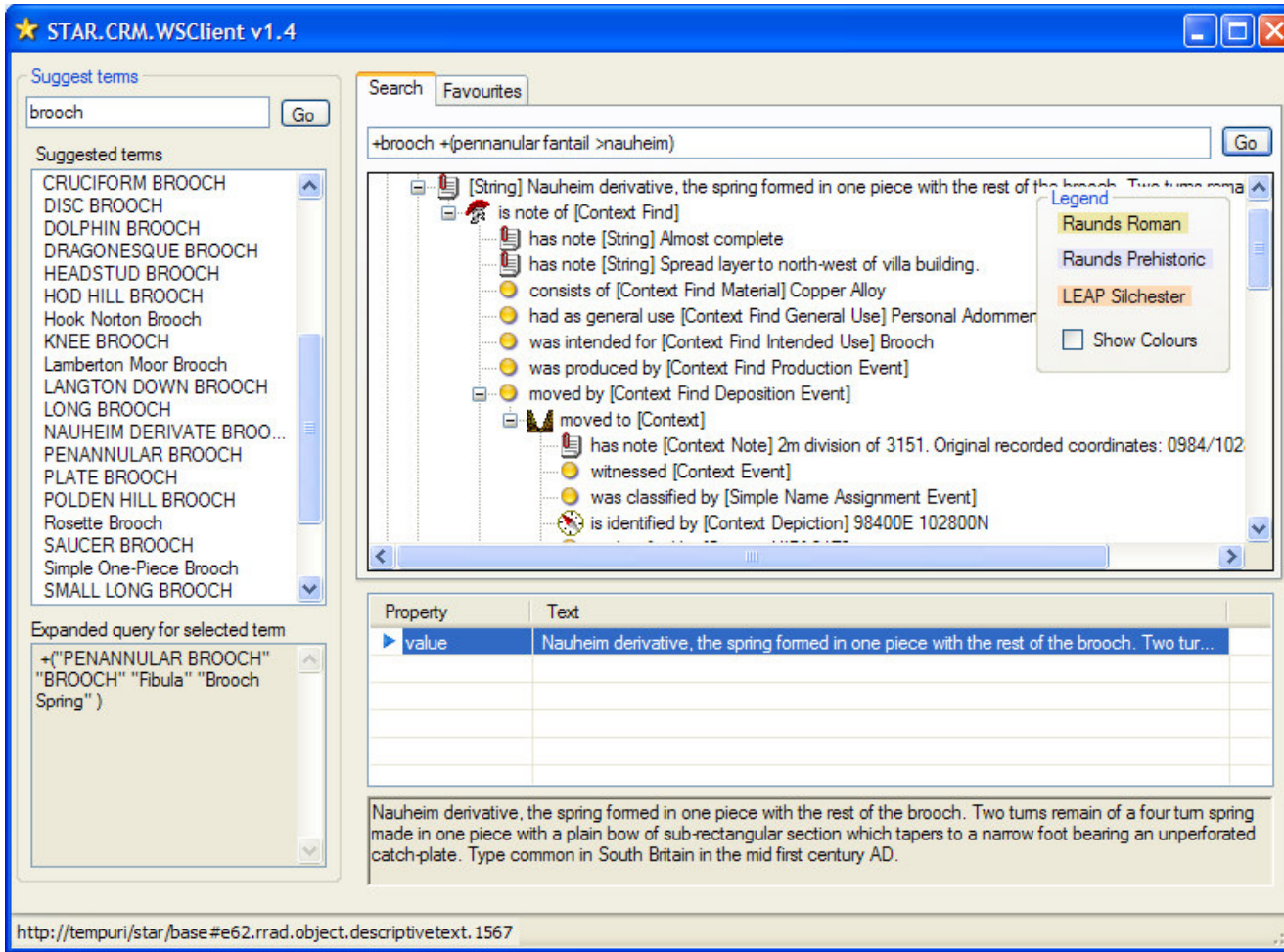
- BOW (WEAPON)
- LONGBOW
- ARROW
- PROJECTILE WEAPON
- HUNTING OBJECT
- PROJECTILE
- BLOWPIPE (WEAPON)
- CROSSBOW
- SPEAR

Concept Expansion

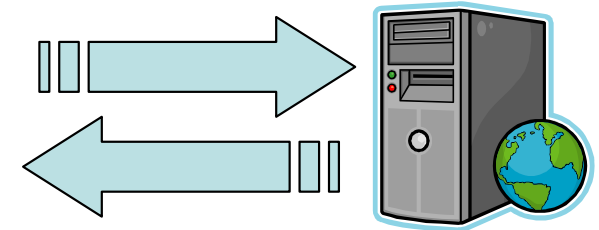




CRMEH Web Service and Client Application



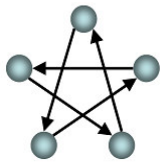
CRMEH Client Application



CRMEH Web Service

Search and browse seamlessly across multiple archaeological datasets

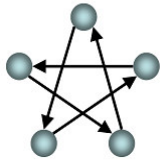




CRMEH Client Application

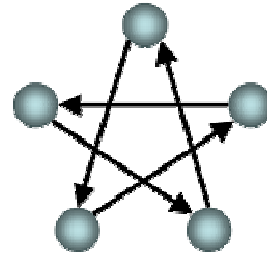
Property	Text





Next Steps...

- Query builder – expose support for complex query patterns (SPARQL)
- Closer integration of thesauri into overall search process
- Browser-based application components (AJAX)
- Include further archaeological datasets



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

DANMARKS
BIBLIOTEKSSKOLE



ENGLISH HERITAGE

University of Glamorgan

you live, you learn

