# Following a guiding STAR?

## Latest EH work with, and plans for,

## Semantic Technologies

**Presented by Keith May**

**Based on research work of English Heritage staff especially Paul Cripps & Phil Carlisle (NMR DSU)**

**and**

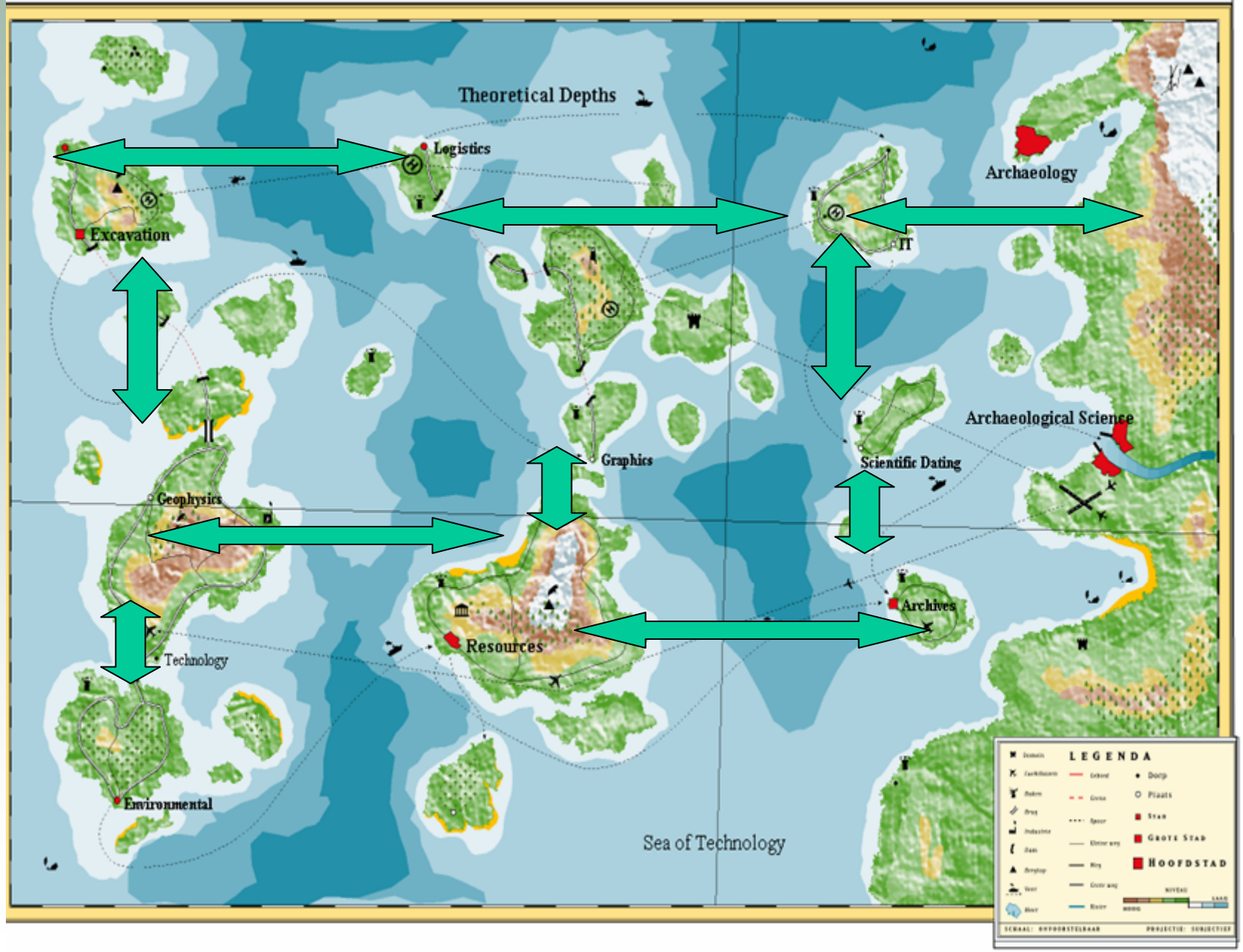
**Doug Tudhope, Ceri Binding and Thanos Zafiriu at Glamorgan University**

# In the beginning - Revelation?

**"Revelation is an English Heritage project to provide a coherent digital information system that will make the capture, analysis and dissemination of our research faster and more effective"** ***Cross et al 2003*****.**

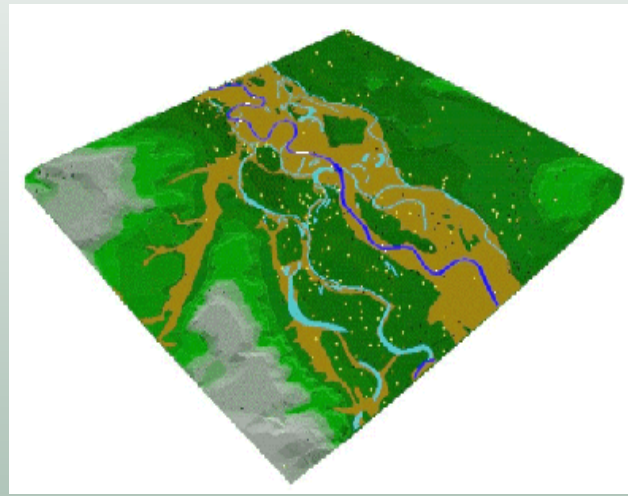
## Integrating archaeological data & information

# The Archaeological Archipelagos

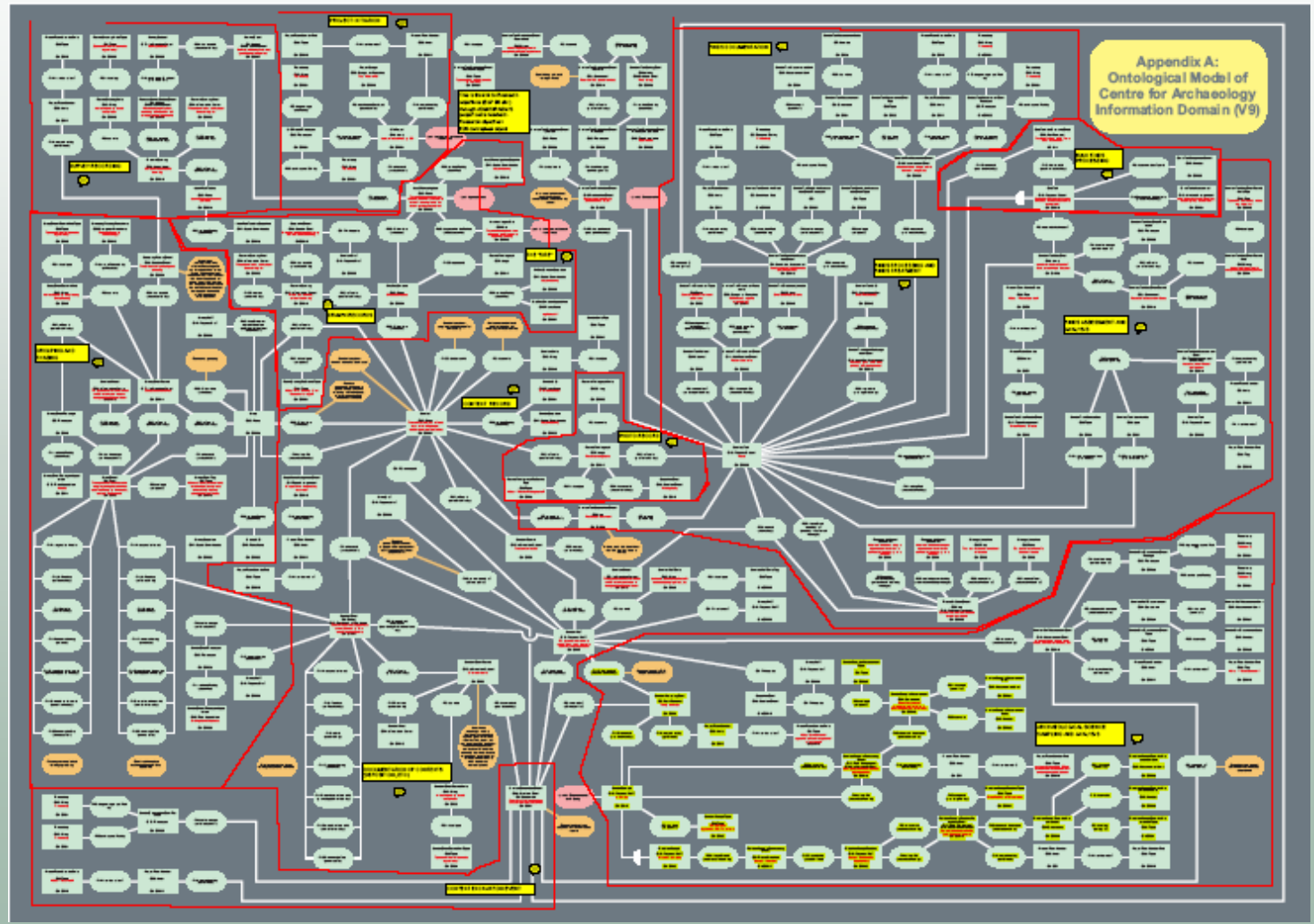


# 4. Modelling versus Mapping

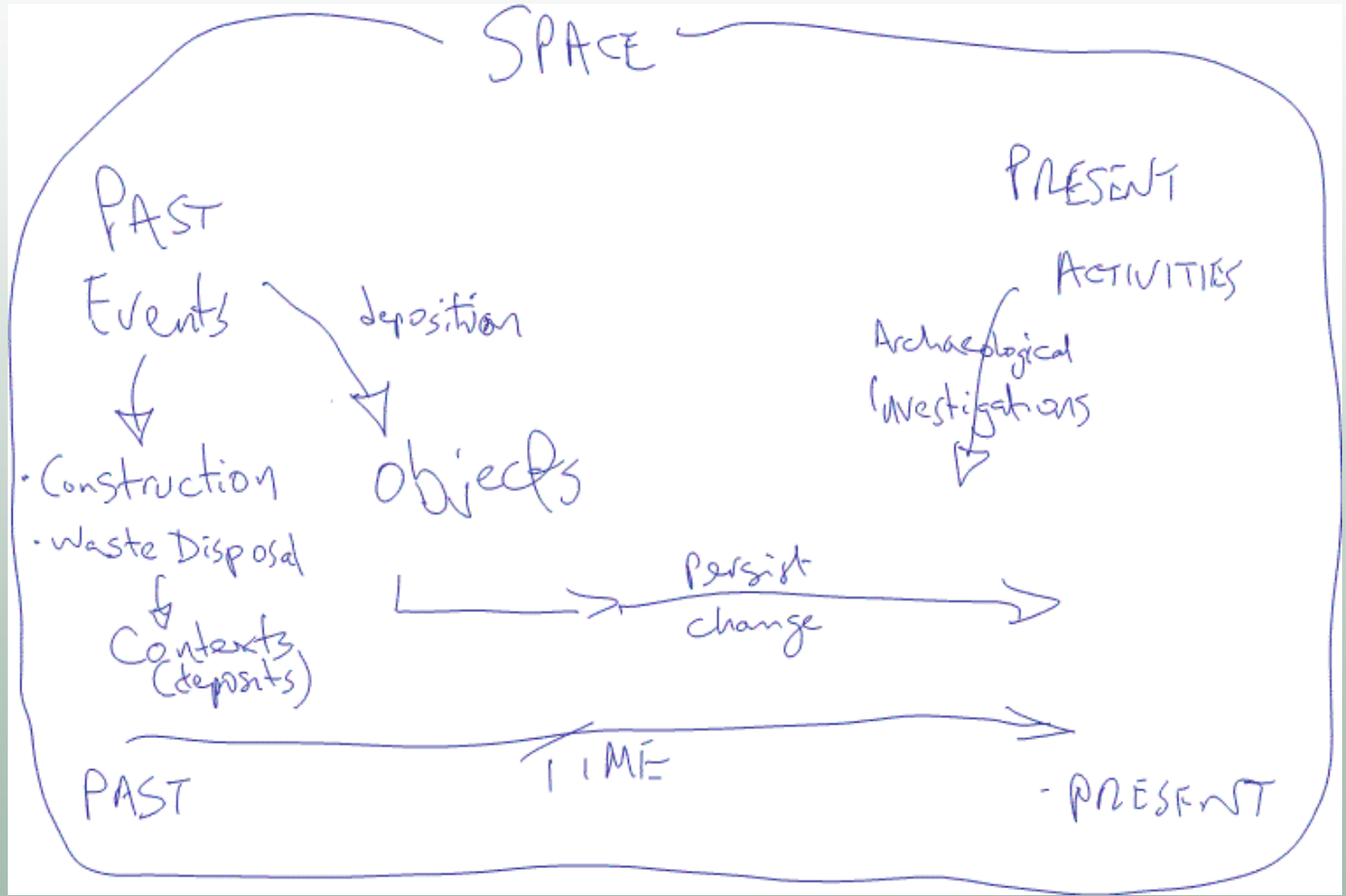
- *Model* new systems requirements
- *Map* to legacy or current data records
- *Mapping* to ‘virtual fields’ in Archaeological recording system
- *Representing* different degrees of Granularity (different levels of detail between conceptual info and actual data fields)



# CRM diagram of Archaeological Information Domain (ref: http://cidoc.ics.forth.gr/technical_papers.html)



# Archaeological Processes









# Events in the present

Excavation // Drawing and Photography
Survey // Sampling
Treatments & Processing
Classification // Grouping, & Phasing
Measuring, including scientific dating
Recording of observations
Dissemination // Interpretation/Analysis

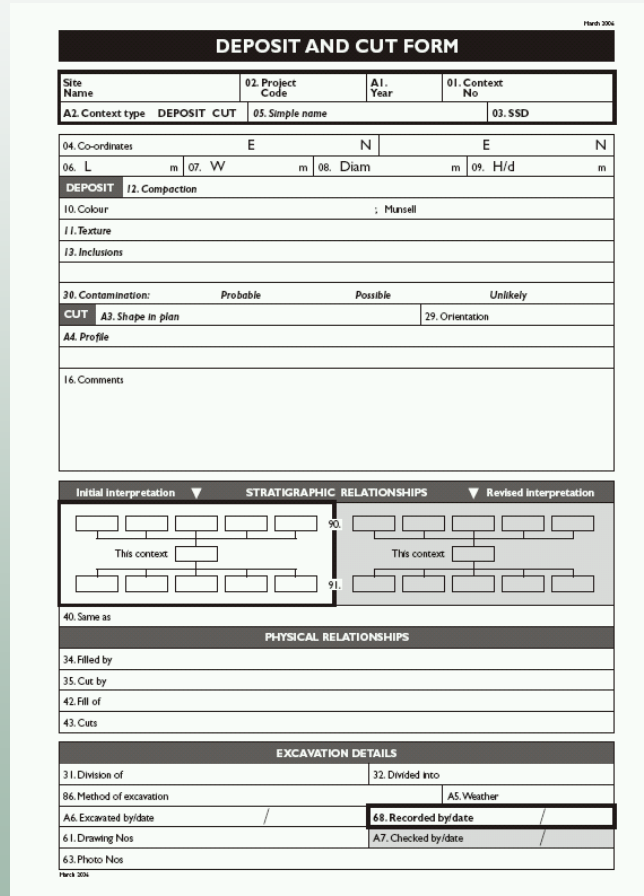


# Events in the past

- Context formation and depositional events (stratigraphy)
- Geochemical, geological, environmental and biological processes
- Object production, disposal or loss (finds deposition)
- Construction, deposition, modification and destruction events relating to layers, features, structures, buildings (taphonomy)
- Events occur at places; spatial operators for reasoning about spatial relationships
- Allen's Temporal Operators for reasoning about the sequence of events and building the site matrix

# Background to Archaeological model

- Limited degree of detail
- Context record sheet modelled as CRM Information Object (E73)
- Model still complex enough - most domain users (archaeologists) find it daunting



DEPOSIT AND CUT FORM

| | | | |
|---|---|---|---|
| Site Name | 02. Project Code | A1. Year | 01. Context No |
| A2. Context type DEPOSIT CUT | 05. Simple name | | 03. SSD |

| | | | |
|---|---|---|---|
| 04. Co-ordinates | E | N | E N |
| 06. L m | 07. W m | 08. Diam m | 09. H/d m |

DEPOSIT 12. Compaction

10. Colour ; Munsell

11. Texture

13. Inclusions

30. Contamination: Probable Possible Unlikely

CUT A3. Shape in plan 29. Orientation

A4. Profile

16. Comments

Initial interpretation ▼ STRATIGRAPHIC RELATIONSHIPS ▼ Revised interpretation



40. Same as

PHYSICAL RELATIONSHIPS

34. Filled by

35. Cut by

42. Fill of

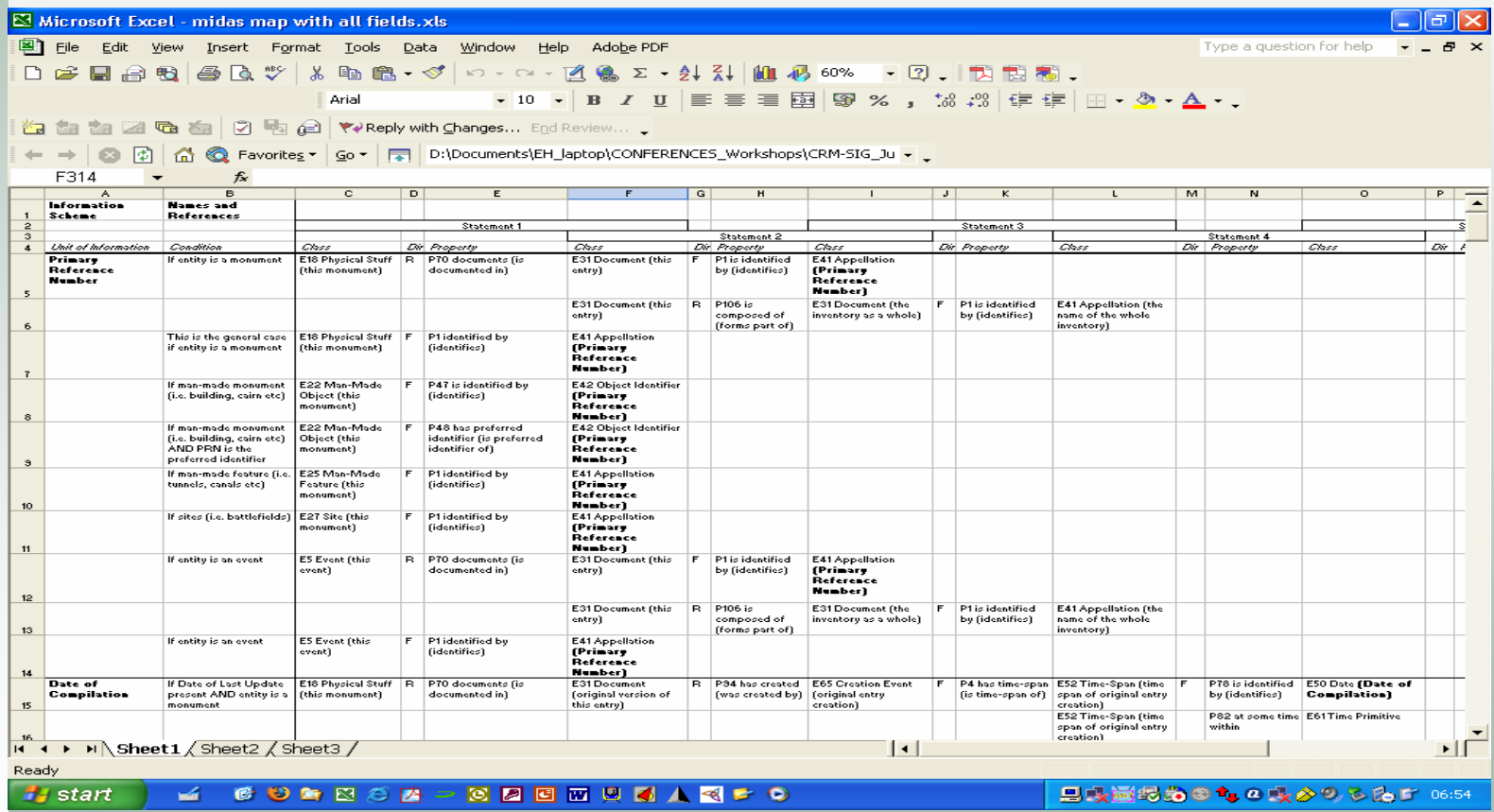
43. Cuts

EXCAVATION DETAILS

| | |
|---|---|
| 31. Division of | 32. Divided into |
| 86. Method of excavation | A5. Weather |
| A6. Excavated by/date / | 68. Recorded by/date / |
| 61. Drawing Nos | A7. Checked by/date / |
| 63. Photo Nos | |

# Progress, Tools and Possible futures

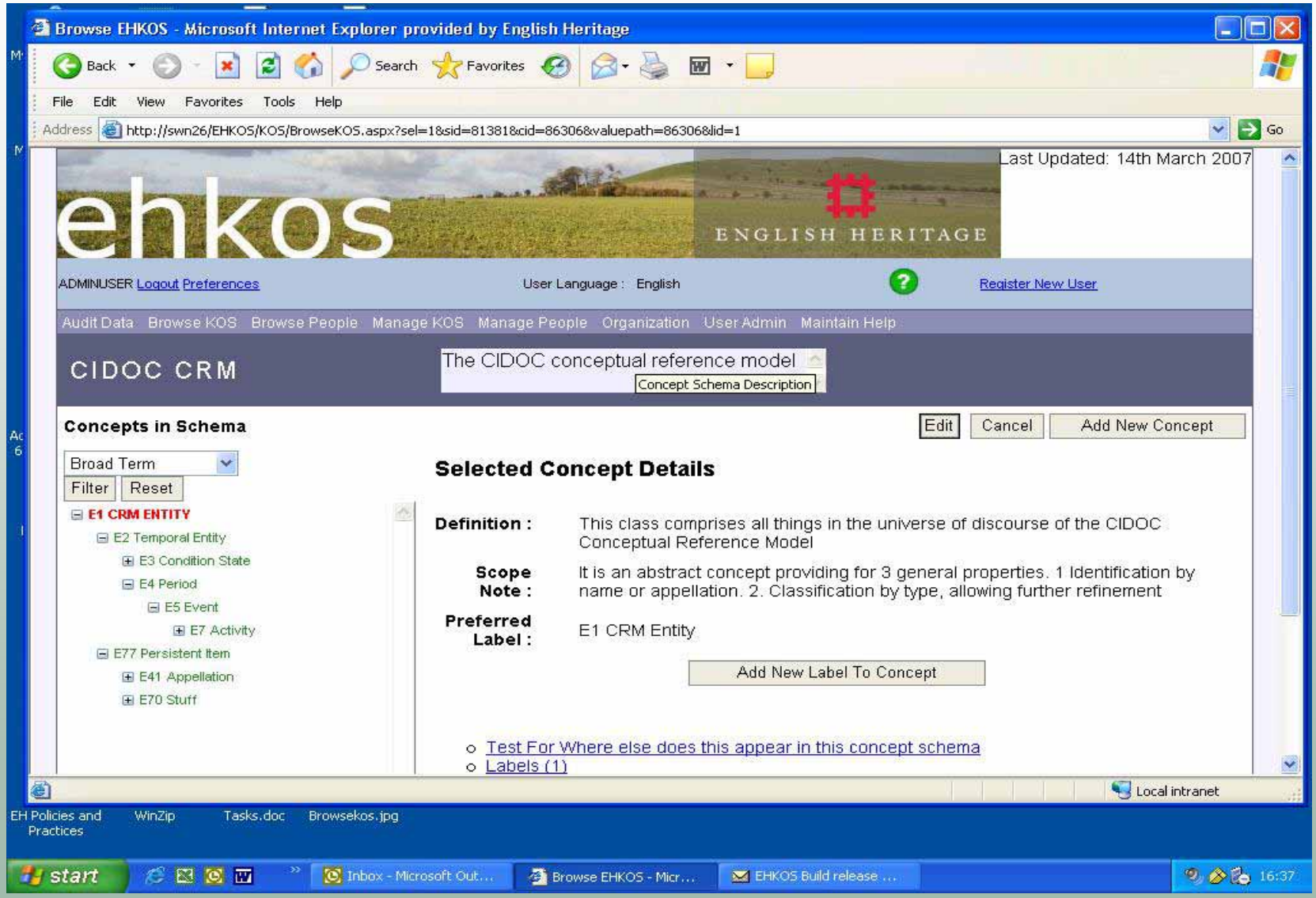

# MIDAS mapping to CRM



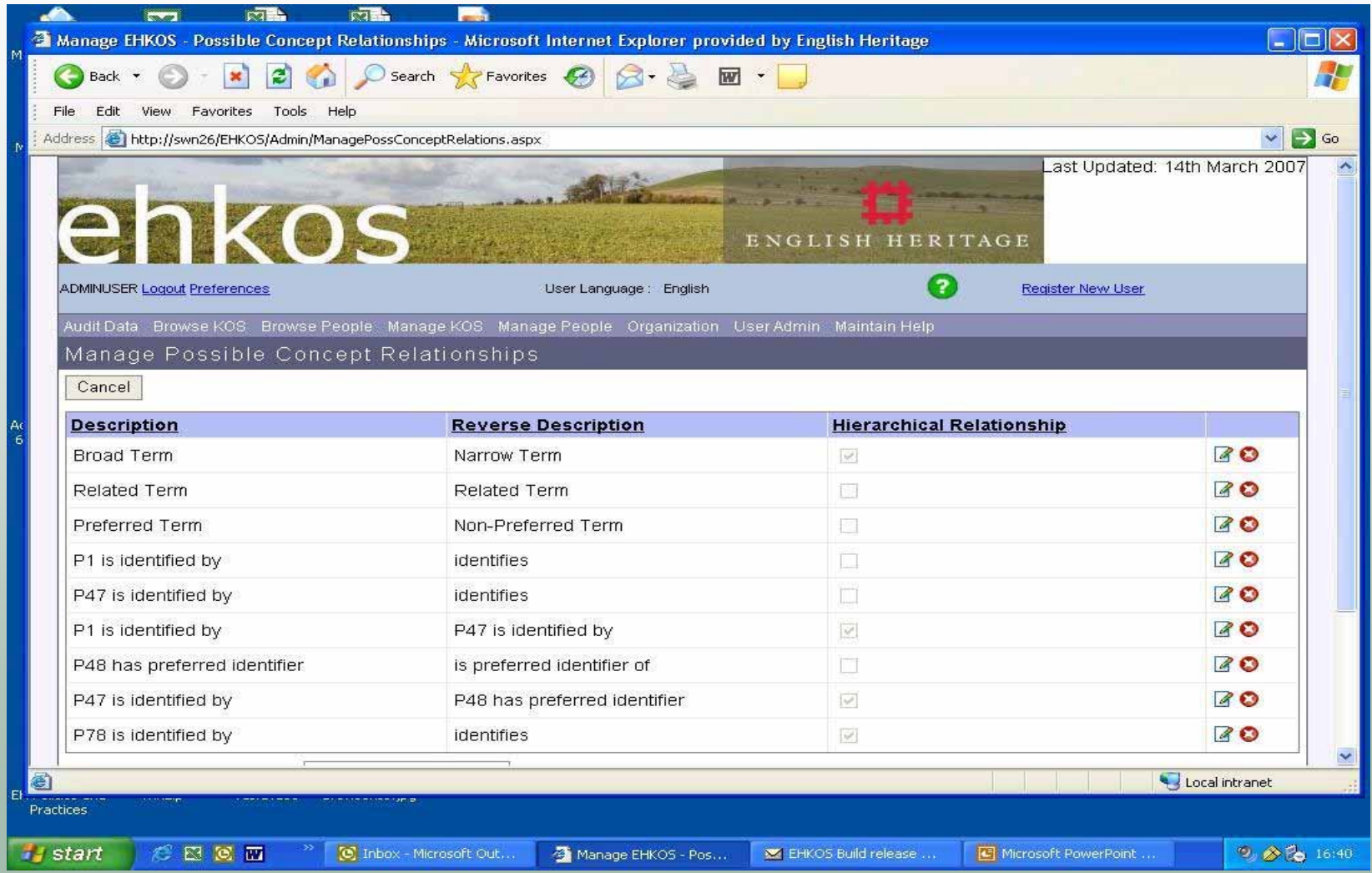

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **Information Scheme** | **Names and References** | | | | | | | | | | | | | | |
| 2 | | | Statement 1 | | | | | | Statement 3 | | | | | | | |
| 3 | | | | | | Statement 2 | | | | | | Statement 4 | | | | |
| 4 | *Unit of Information* | *Condition* | *Class* | *Dir* | *Property* | *Class* | *Dir* | *Property* | *Class* | *Dir* | *Property* | *Class* | *Dir* | *Property* | *Class* | *Dir* |
| 5 | **Primary Reference Number** | If entity is a monument | E18 Physical Stuff (this monument) | R | P70 documents (is documented in) | E31 Document (this entry) | F | P1 is identified by (identifies) | E41 Appellation **(Primary Reference Number)** | | | | | | | |
| 6 | | | | | | E31 Document (this entry) | R | P106 is composed of (forms part of) | E31 Document (the inventory as a whole) | F | P1 is identified by (identifies) | E41 Appellation (the name of the whole inventory) | | | | |
| 7 | | This is the general case if entity is a monument | E18 Physical Stuff (this monument) | F | P1 identified by (identifies) | E41 Appellation **(Primary Reference Number)** | | | | | | | | | | |
| 8 | | If man-made monument (i.e. building, cairn etc) | E22 Man-Made Object (this monument) | F | P47 is identified by (identifies) | E42 Object Identifier **(Primary Reference Number)** | | | | | | | | | | |
| 9 | | If man-made monument (i.e. building, cairn etc) AND PRN is the preferred identifier | E22 Man-Made Object (this monument) | F | P48 has preferred identifier (is preferred identifier of) | E42 Object Identifier **(Primary Reference Number)** | | | | | | | | | | |
| 10 | | If man-made feature (i.e. tunnels, canals etc) | E25 Man-Made Feature (this monument) | F | P1 identified by (identifies) | E41 Appellation **(Primary Reference Number)** | | | | | | | | | | |
| 11 | | If sites (i.e. battlefields) | E27 Site (this monument) | F | P1 identified by (identifies) | E41 Appellation **(Primary Reference Number)** | | | | | | | | | | |
| 12 | | If entity is an event | E5 Event (this event) | R | P70 documents (is documented in) | E31 Document (this entry) | F | P1 is identified by (identifies) | E41 Appellation **(Primary Reference Number)** | | | | | | | |
| 13 | | | | | | E31 Document (this entry) | R | P106 is composed of (forms part of) | E31 Document (the inventory as a whole) | F | P1 is identified by (identifies) | E41 Appellation (the name of the whole inventory) | | | | |
| 14 | | If entity is an event | E5 Event (this event) | F | P1 identified by (identifies) | E41 Appellation **(Primary Reference Number)** | | | | | | | | | | |
| 15 | **Date of Compilation** | If Date of Last Update present AND entity is a monument | E18 Physical Stuff (this monument) | R | P70 documents (is documented in) | E31 Document (original version of this entry) | R | P94 has created (was created by) | E65 Creation Event (original entry creation) | F | P4 has time-span (is time-span of) | E52 Time-Span (time span of original entry creation) | F | P78 is identified by (identifies) | E50 Date **(Date of Compilation)** | |
| 16 | | | | | | | | | | | | E52 Time-Span (time span of original entry creation) | | P82 at some time within | E61 Time Primitive | |

# New Corporate Reference Data Module

# Includes CIDOC-CRM

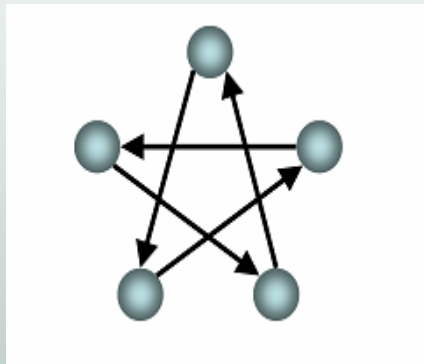


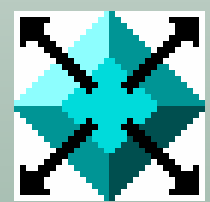
# Concept relations – CRM Properties



# STAR - Semantic Technologies for Archaeological Resources

- Building upon EH Ontological Modelling work and…
- FACET ***a collaborative project investigating the potential of faceted thesauri for retrieval from multimedia collections***
  - **Faceted Access to Cultural hEritage Terminology**
  - D. Tudhope & Ceri Binding
    Glamorgan University
    Faculty of Advanced Technology





# Project Outline

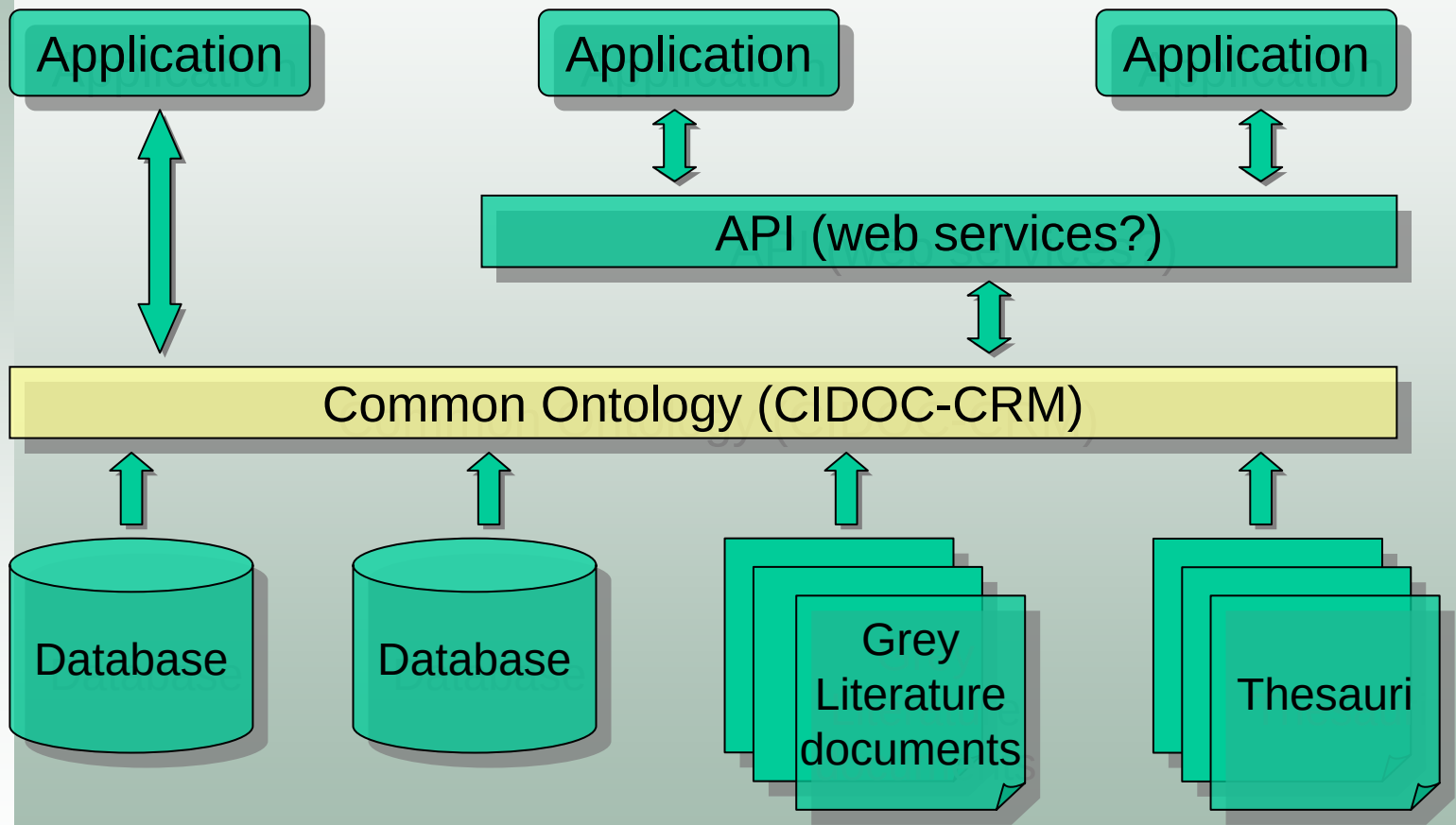
- 3 year AHRC funded project
- Started January 2007, finish December 2010
- Collaborators
  - University of Glamorgan
  - English Heritage
  - RSLIS Denmark
- Aim – "*To investigate the potential of semantic terminology tools for widening access to digital archaeological resources, including disparate datasets and associated grey literature*"



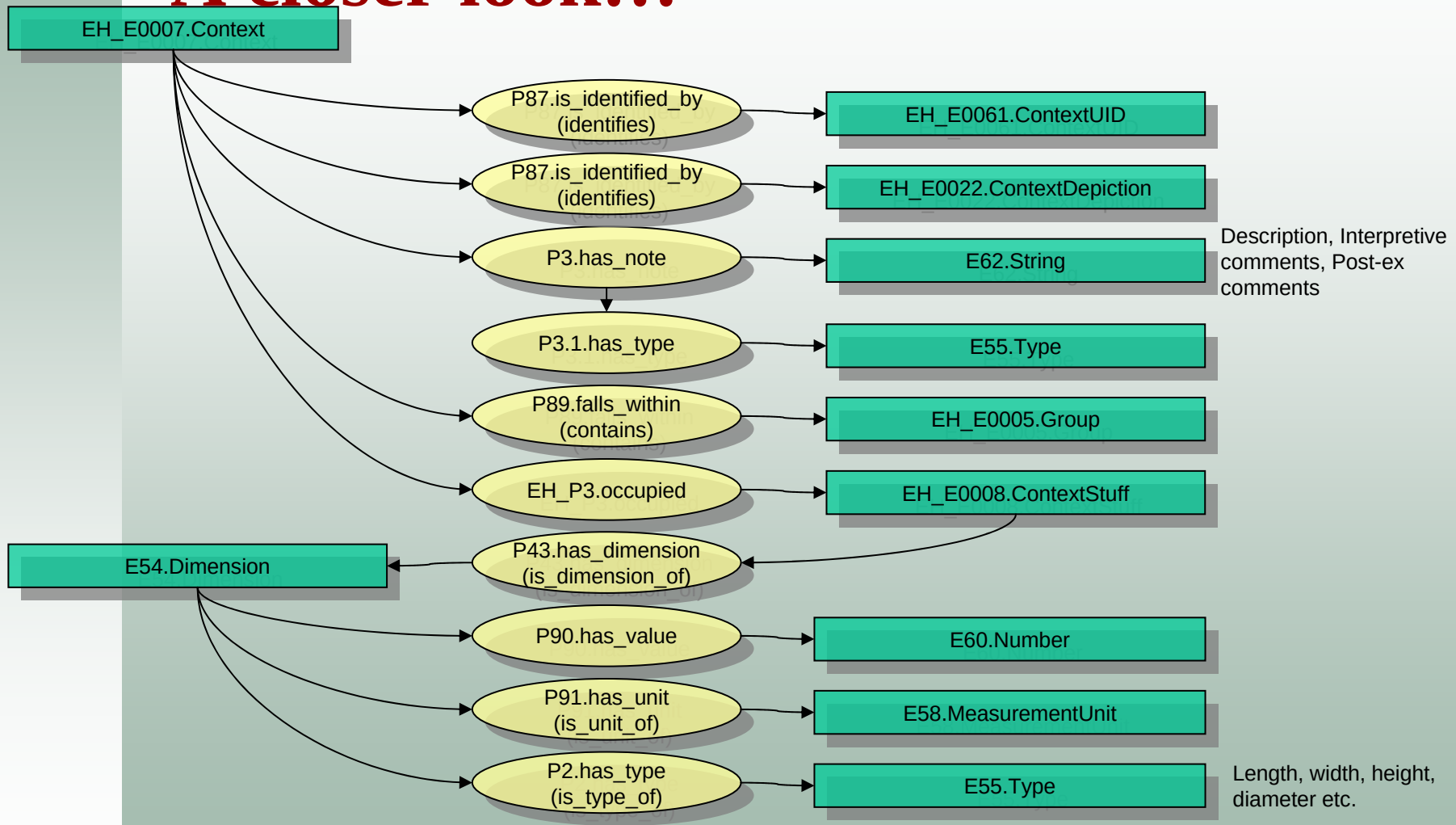

# Archaeological Resources - to explore

- Raunds Data & Grey Literature reports
- Online AccesS to the Index of archaeological excavationS (OASIS) [http://ads.ahds.ac.uk/project/oasis/]
  - Library of unpublished fieldwork reports
- Keyword Extraction Algorithm (KEA) [http://www.nzdl.org/Kea/]

# General architecture

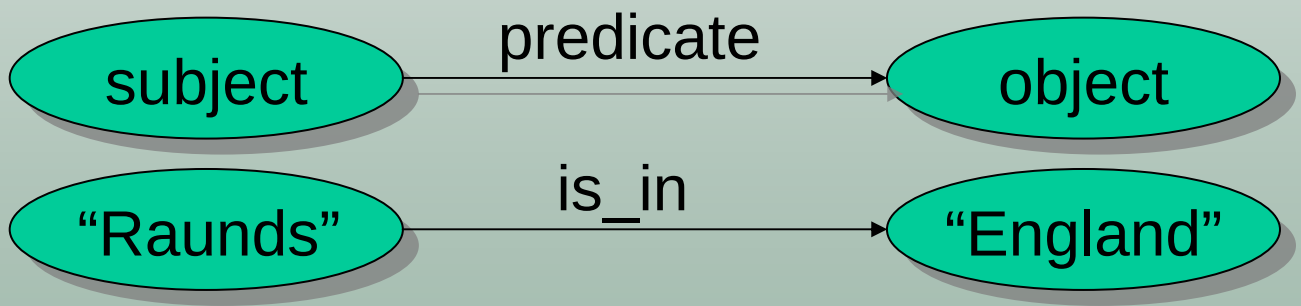


# A closer look…



# Resource Description Framework (RDF)

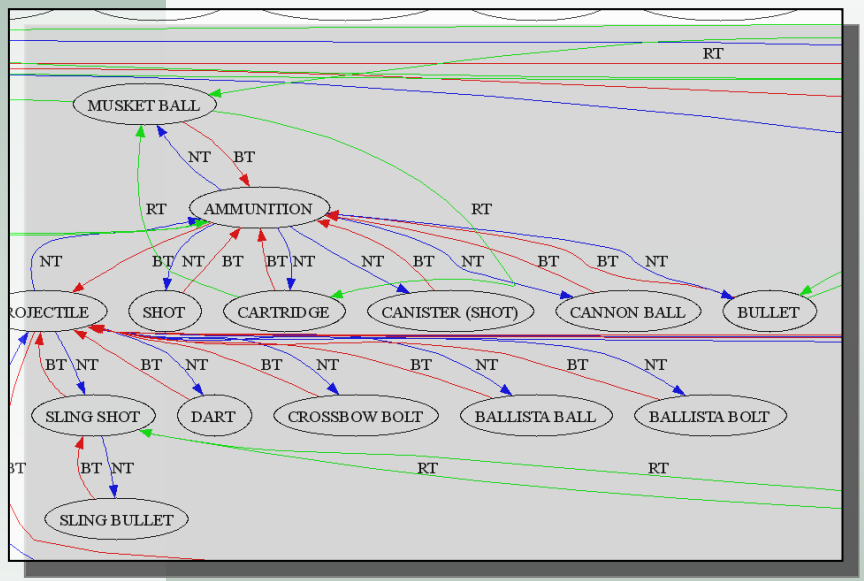
- http://www.w3.org/RDF/
- XML / URI based format
- RDF triples:




# Simple Knowledge Organisation Systems (SKOS)

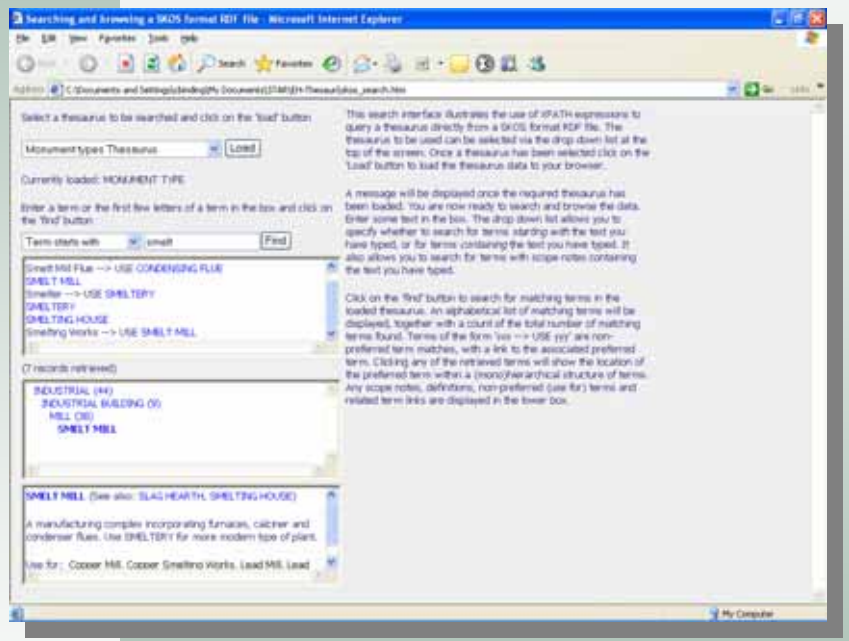
- http://www.w3.org/2004/02/skos/
- Representation of thesauri, taxonomies, classification schemes etc. in RDF
- Looser semantics than e.g. OWL

# Representing Thesauri in SKOS RDF

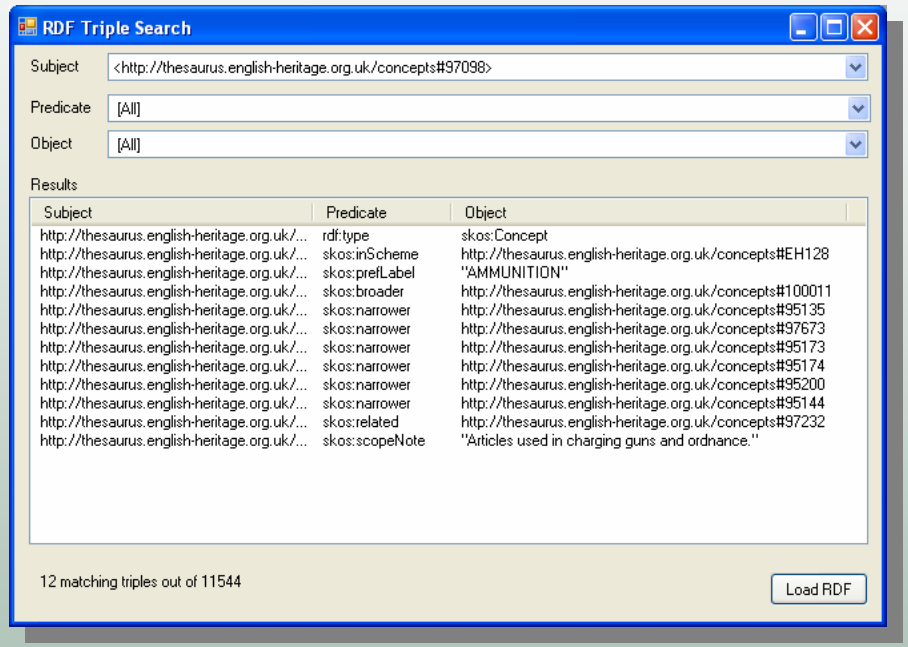



```
<skos:Concept rdf:about="#100062">
    <skos:inScheme rdf:resource="#EH128"/>
    <skos:prefLabel>MISERICORD (DAGGER)</skos:prefLabel>
    <skos:broader rdf:resource="#95138"/>
    <skos:scopeNote>A form of dagger.</skos:scopeNote>
</skos:Concept>
<skos:Concept rdf:about="#100063">
    <skos:inScheme rdf:resource="#EH128"/>
    <skos:prefLabel>HANGER</skos:prefLabel>
    <skos:broader rdf:resource="#95221"/>
    <skos:scopeNote>A type of sword often used by infantry.</skos:scopeNote>
</skos:Concept>
<skos:Concept rdf:about="#100064">
    <skos:inScheme rdf:resource="#EH128"/>
    <skos:prefLabel>SABRE</skos:prefLabel>
    <skos:broader rdf:resource="#95221"/>
    <skos:scopeNote>A curved sword designed to cut with used by cavalry.</skos:scopeNote>
</skos:Concept>
<skos:Concept rdf:about="#97098">
    <skos:inScheme rdf:resource="#EH128"/>
    <skos:prefLabel>AMMUNITION</skos:prefLabel>
    <skos:broader rdf:resource="#100011"/>
    <skos:narrower rdf:resource="#95135"/>
    <skos:narrower rdf:resource="#97673"/>
    <skos:narrower rdf:resource="#95173"/>
    <skos:narrower rdf:resource="#95174"/>
    <skos:narrower rdf:resource="#95200"/>
    <skos:narrower rdf:resource="#95144"/>
    <skos:related rdf:resource="#97232"/>
    <skos:scopeNote>Articles used in charging guns and ordnance.</skos:scopeNote>
</skos:Concept>
```

# Applications using the RDF data

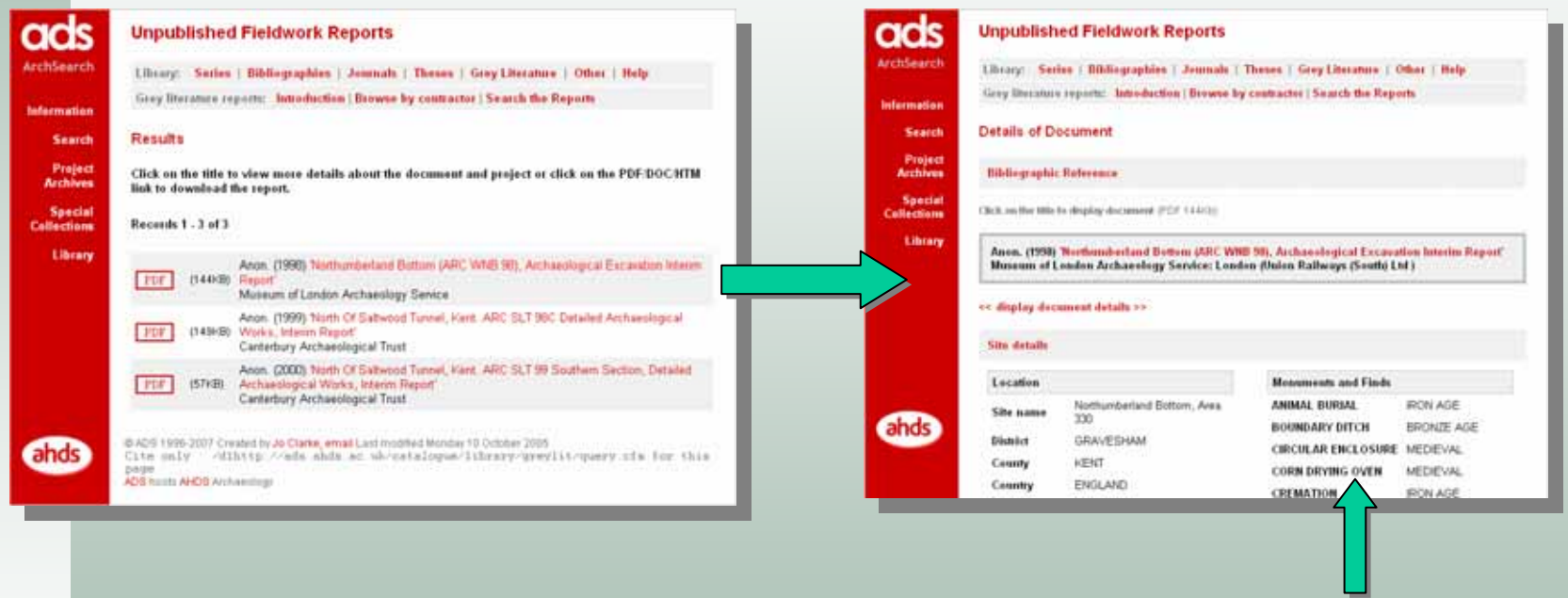


SKOS thesaurus browser




RDF triple search

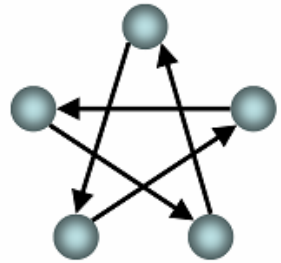
# OASIS – Archaeology Data Service
# Grey Literature online library




- http://ads.ahds.ac.uk/catalogue/

Some controlled vocabulary indexing

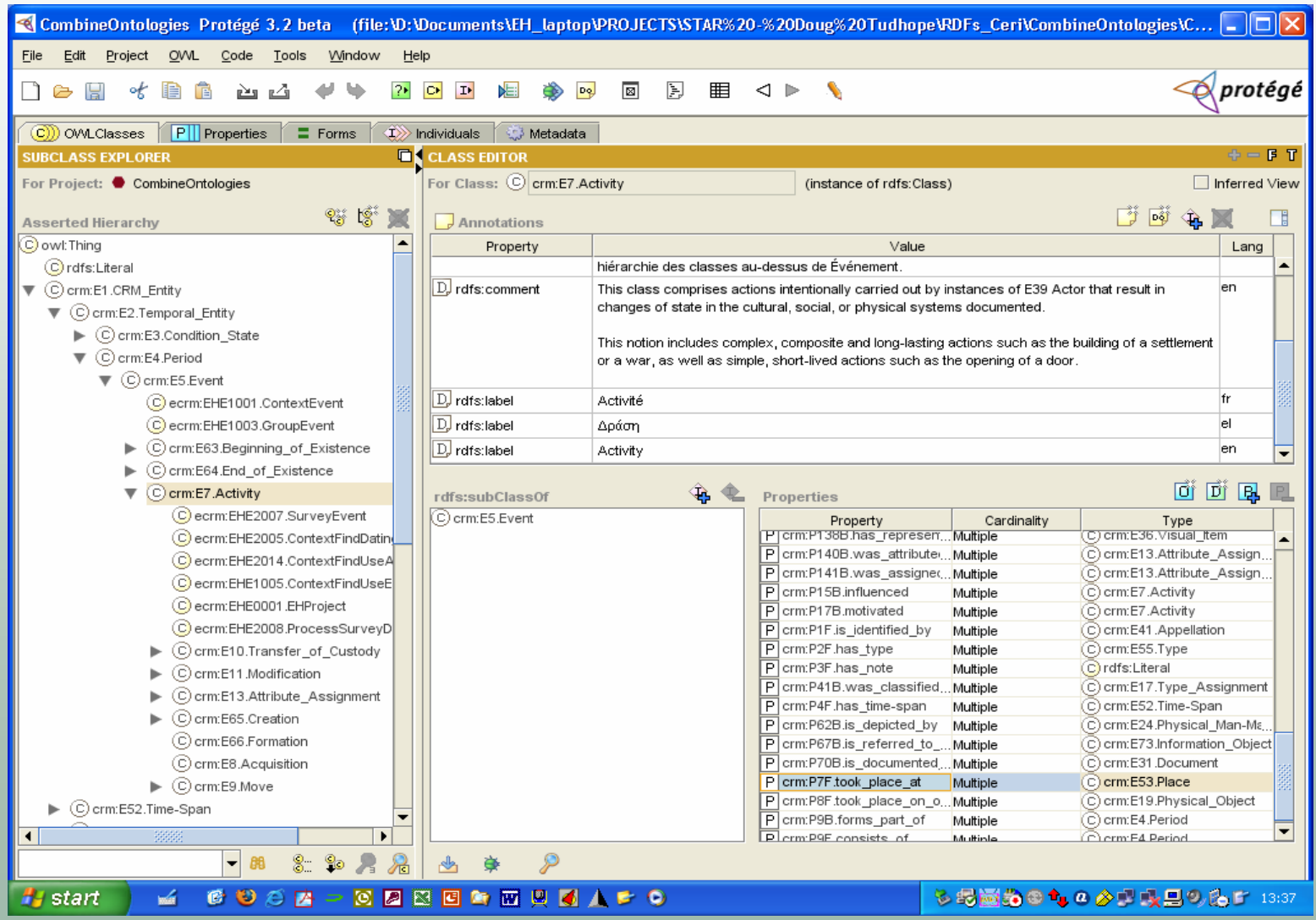
# Summary



CIDOC-CRM as overarching logical structure

- Thesauri
  - English Heritage Thesauri (SKOS)
- Grey Literature
  - Raunds grey literature reports
  - OASIS / ADS ArchSearch
  - Indexing using KEA
- Datasets
  - Raunds RRAD MS-Access database
  - Other EH archaeological projects – legacy data
  - Silchester Roman town – IADB MySQL database

# Protégé – Dynamic CRM-EH



# Protégé modelling – pros and cons

+ Pros: modelling much more updateable – dynamic
+ easy to disseminate in RDF formats for developers
+ Protégé is open source

- Cons: Not good dissemination tool for EH domain users
- Not much use to the wider Archaeological or Heritage community for agreeing a standard ontology?
- Protégé graphing tools are unwieldy for complex modelling
- Protégé is open source – difficult to maintain as a standard tool for dissemination?
- Protégé networking – a whole further project at EH

# CRM-EH extensions in RDF

- "CRM-EH" – need a 'published' version (Where? – EH, CIDOC, currently Glamorgan)
- First need to complete RDF descriptions (90+)
- CRM-EH RDFs online
  - – currently on Glamorgan server
  - May need agreed protocols for how these are used?
- Standards evolve - What mechanisms for updating & keeping current on different servers?
- E.g. How best to incorporate MIDAS changes?

# Semantic Web - interface examples

1. Oracle Technology Network
   - Beta test site
   - http://otnsemanticweb.oracle.com/
2. iGoogle
3. IkeWiki

# Results showing Faceted Navigation interface

Oracle Technology Network Semantic Web - Microsoft Internet Explorer provided by English Heritage (Admin)

File Edit View Favorites Tools Help

Back Search

Address http://otnsemanticweb.oracle.com/index.jsp?tn=0Identity&tv=service+oriented&ss=1&

Go del.icio.us TAG

ORACLE TECHNOLOGY NETWORK

Service Oriented

Home Headlines Downloads Forums Articles Podcasts Blogs More Sources

Collapse all | Expand all

**NAVIGATE BY**

**DATABASE**
- Oracle Database 10g (24)
- Database Options (19)
- Zend Core for Oracle (7)
- Embedded Database (5)
- Warehouse Builder (3)
- Secure Backup (1)
- Rdb (1)
- Gateways (1)

**MIDDLEWARE**
- Oracle Developer Tools (32)
- Oracle Fusion Middleware (10)
- SOA Suite (10)

**APPLICATIONS**
- On Demand (3)
- Oracle E-Business Suite (4)
- PeopleSoft Enterprise (14)
- Siebel (1)

**TECHNOLOGIES**
- Business Integration (15)
- Business Intelligence (13)

## Welcome to OTN Semantic Web (Beta)

Get a *Semantic Web* view of OTN content.

Read the FAQ | Provide Feedback | Visit "classic" OTN

**Filter**

All Fields
**service oriented**

See All Results in an Integrated View

**RECENTLY POSTED**

Generate Functional JSF-ADF Applications using Oracle JHeadstart
posted 2007-05-25

Guide to Advanced Linux Command Mastery, Part 2
posted 2007-02-21

Building Event-Driven Architecture with an Enterprise Service Bus
posted 2007-02-08

Download Oracle Secure Enterprise Search 10g (10.1.8)
posted 2007-02-05

Download Oracle WebCenter Suite
posted 2007-02-05

**ARTICLES**

Guide to Advanced Linux Command Mastery, Part 2
posted 2007-02-21

Building Event-Driven Architecture with an Enterprise Service Bus
posted 2007-02-08

Accessorize Oracle Database with Ruby
posted 2007-01-31

Creating an Ajax Process Using PHP and Oracle
posted 2007-01-24

Drupal + Oracle: Inside the OraDrup Project
posted 2007-01-18

**DOWNLOADS**

Oracle Containers for J2EE 10g (OC4J) Version 10.1.3.2.0
posted on 2007-03-01

Oracle Rdb 7.2.1.1.x Alpha and Itanium
posted on 2007-03-01

Enterprise Manager 10g Grid Control Release 3 (10.2.0.3) for Solaris (SPARC), HP-UX, AIX5L
posted on 2007-03-01

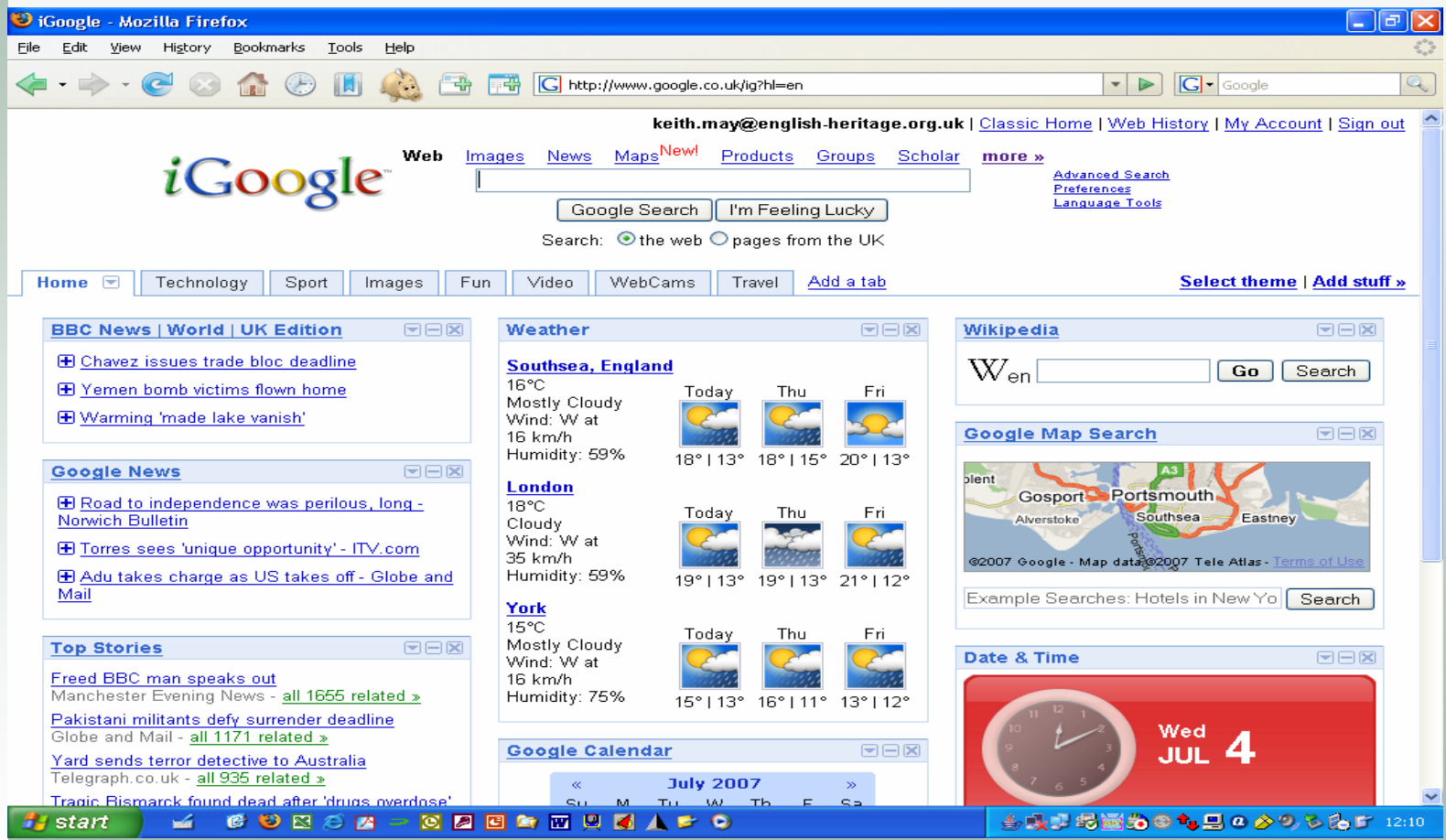
TimesTen In-Memory Database v6.0.8
posted on 2007-02-01

Oracle SQL Developer 1.1 Patch 1 (1.1.1.25.14)
posted on 2007-01-31

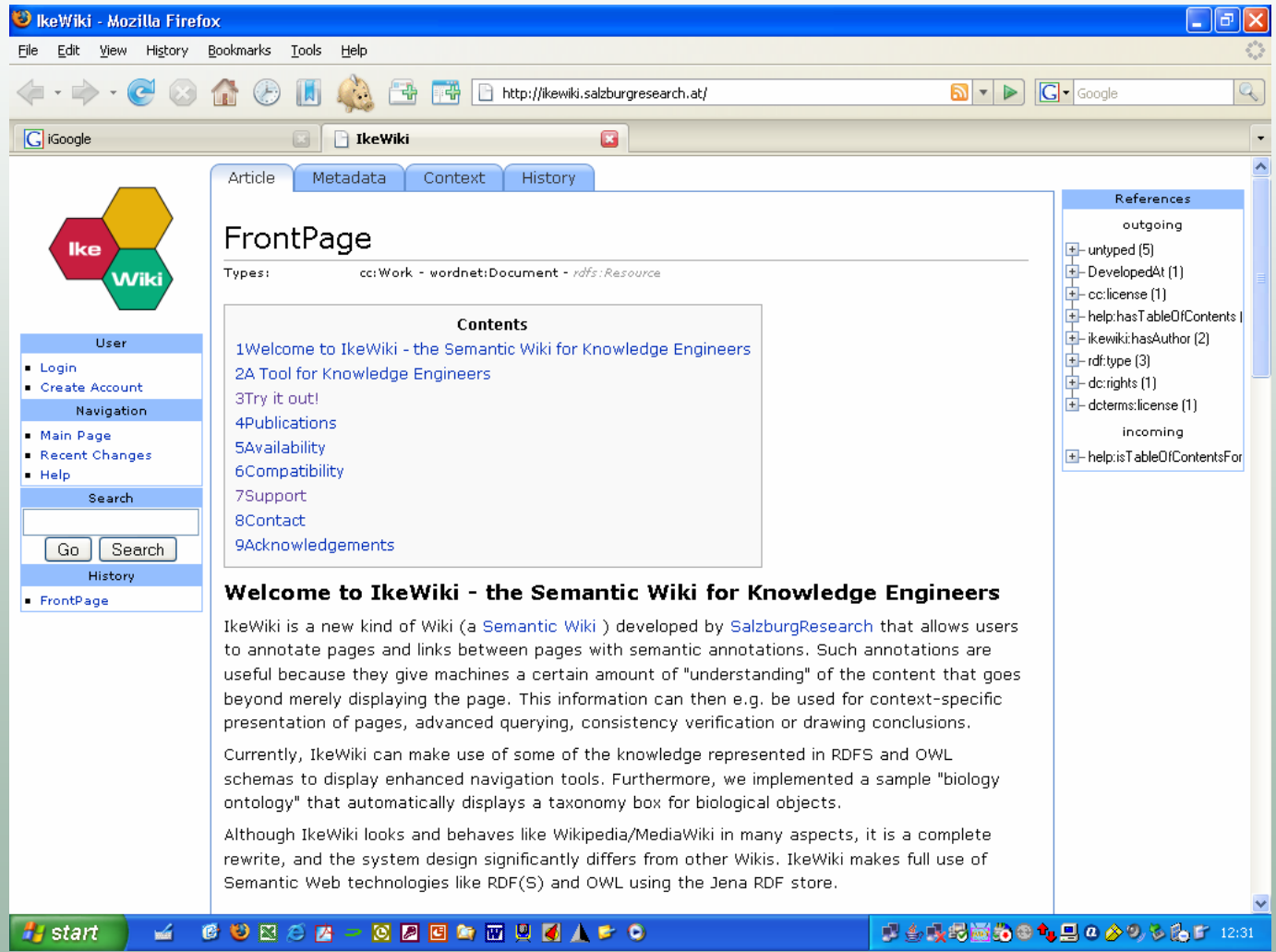
Internet

start 04:57

# iGoogle web service interface?



# IkeWiki – Semantic Wiki



# 6. Conclusions and considerations for further work

# Need to be semantically explicit about the scope of your information domain

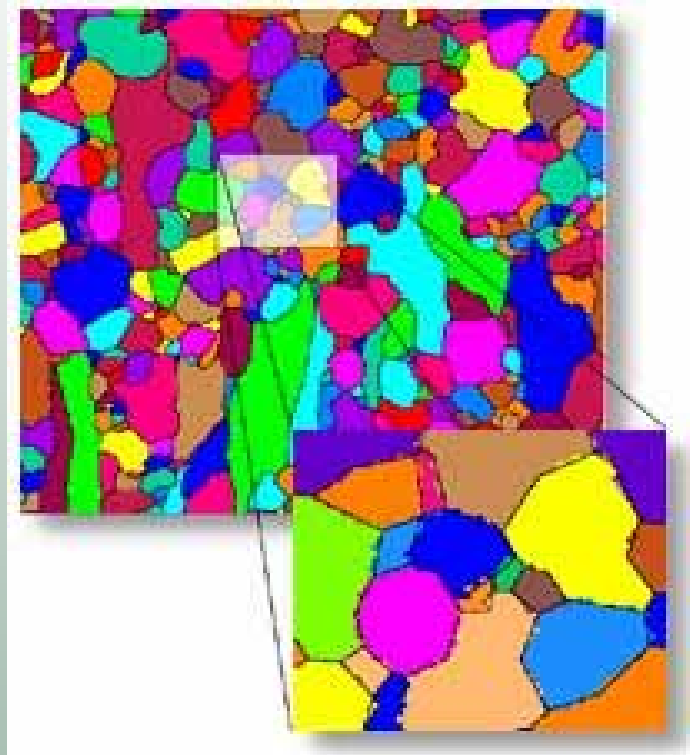
- Identifying Boundaries & how best to work with them & between them
- Important to define differences of Scale or Granularity of info.

  E.g. ? Local

  Regional

  National

  International
- Geo-political Cross-cutting could conflict with some Cultural heritage requirements



# Granularity issues

- Being explicit about the levels of entities within a model or mapping
- Is there a measure to explicitly express the current granularity of the model
  - How? – A Faceted user interface?
- Can/should we define the ‘granularity’ of the dataset as part of the contextual metadata? CRM-Core?

# Version control on CRM extensions?

- Can we cope with interoperable systems using slightly different versions of a standard?
- How well will EH extensions still work with CRM “Vanilla”?
- May need explicit ‘rules of engagement’

for interoperable data sets (web services)

# Dissemination tools to better enable user endorsement of CIDOC-CRM

- Need for wider Heritage engagement with CRM
- EH need other UK archaeologists to adopt
- Need to identify “cost-benefits” for sector
- Dissemination issues with size of model
  - and further mapping is considerably more
- Need better graphical modelling outputs
- Protégé helps – but not for archaeologists
  - (Domain end-users still need convincing)

# Verification & Dissemination

- Dissemination Review at end of project
- Verification by CRM-SIG & FISH, etc?
- Publish updated model & RDFs online
  - CRM-SIG and/or EH websites?
- Other publication & dissemination routes ?
  - Interpreting Stratigraphy conference May 2008
  - Internet Archaeology?
  - Semantic Wiki
    - on CIDOC-CRM website depending upon existing Wiki tools? http://139.91.183.17:81/tiki/tiki-index.php

# Bibliography / References


- Revelation assessment report – Cross et al, EH 2003.
- Denny, M. 2002. *Ontology Building: A Survey of Editing Tools.* http://www.xml.com/pub/a/2002/11/06/ontologies.html
- CIDOC CRM v3.4.9 - http://cidoc.ics.forth.gr/
- CRM-EH Ontological model - Cripps, et al 2005 http://cidoc.ics.forth.gr/technical_papers.html
- STAR - http://hypermedia.research.glam.ac.uk/kos/star/
- FACET Project http://www.glam.ac.uk/soc/research/hypermedia/facet_proj/select_public.php
- Oracle Semantic network http://otnsemanticweb.oracle.com/
- IkeWiki - http://ikewiki.salzburgresearch.at/



- **Keith.may@english-heritage.org.uk**