

Breaking Down Barriers to Interoperability



Ceri Binding & Douglas Tudhope
Hypermedia Research Unit, University of Glamorgan
incorporating the work of
Keith May, *English Heritage*

University of Glamorgan

you live, you learn



STAR Project

- Semantic Technologies for Archaeological Resources
 - 3 year AHRC funded project
 - In collaboration with English Heritage
 - Aim: *“To investigate the potential of semantic terminology tools for widening access to digital archaeology resources, including disparate data sets and associated grey literature”*



STAR Project

Focus on *specific* research questions to excavation data
- demonstrate cross search and browsing at detailed,
meaningful level

Eg Roman corn drying ovens where archaeobotanical
analysis has taken place

Eg Charred plant remains and charcoal from four-post
structures

Eg Post-holes which contain burnt grain

Prototype Information Extraction

(Andreas Vlachidis)

An archaeological evaluation was carried out by ECC FAU on behalf of Essex Police on the site of a proposed new police station at Smiths Farm, on the southeastern outskirts of Great Dunmow, Essex. The site was formerly rough pasture. The Chelmsford Road, which is thought to be the line of a Roman road, runs immediately to the east of the site. Five 30m x 2m trenches were excavated within the footprint of the proposed building and the area of associated carpark. Only one archaeological feature was revealed, a ditch containing prehistoric pottery dating to the Late Bronze Age or Early Iron Age along with burnt flints and flint flakes. No other archaeological features were identified, although a number of prehistoric pottery sherds and flint flakes were discovered on the surface of the natural geology. Although the results of the evaluation do not suggest intensive landscape use during the Late Bronze/ Early Iron Ages it is clear from this and other nearby investigations that a focus for the low level activity seen may well lie in the general vicinity. The absence of Roman or medieval remains indicates that this site was well outside the settlements of these periods. The low quantity and quality of the remains encountered on the site suggests that there is only a minor archaeological implication for the location of the proposed police

LATE BRONZE AGE OR EARLY IRON AGE	<table border="1"> <thead> <tr> <th>Term</th> <th>skos</th> </tr> </thead> <tbody> <tr> <td>LATE BRONZE AGE</td> <td>134734</td> </tr> <tr> <td>EARLY IRON AGE</td> <td>134735</td> </tr> </tbody> </table>	Term	skos	LATE BRONZE AGE	134734	EARLY IRON AGE	134735	E49_Time_Appellation #text 5			
Term	skos										
LATE BRONZE AGE	134734										
EARLY IRON AGE	134735										
ROMAN OR MEDIEVAL	<table border="1"> <thead> <tr> <th>Term</th> <th>skos</th> </tr> </thead> <tbody> <tr> <td>ROMAN</td> <td>134738</td> </tr> <tr> <td>MEDIEVAL</td> <td>134745</td> </tr> </tbody> </table>	Term	skos	ROMAN	134738	MEDIEVAL	134745	<table border="1"> <tbody> <tr> <td>EARLY IRON AGE</td> </tr> <tr> <td>Broad Term: IRON AGE</td> </tr> <tr> <td>Top Term: CULTURAL PERIOD</td> </tr> </tbody> </table>	EARLY IRON AGE	Broad Term: IRON AGE	Top Term: CULTURAL PERIOD
Term	skos										
ROMAN	134738										
MEDIEVAL	134745										
EARLY IRON AGE											
Broad Term: IRON AGE											
Top Term: CULTURAL PERIOD											
PREHISTORIC PERIOD	<table border="1"> <thead> <tr> <th>Term</th> <th>skos</th> </tr> </thead> <tbody> <tr> <td>PREHISTORIC</td> <td>134718</td> </tr> </tbody> </table>	Term	skos	PREHISTORIC	134718	#text 2					
Term	skos										
PREHISTORIC	134718										

- Looking to extract CRMEH period, context, find, sample entities
- Aim to cross search with data

Workshop on the *implementation* of CIDOC-CRM

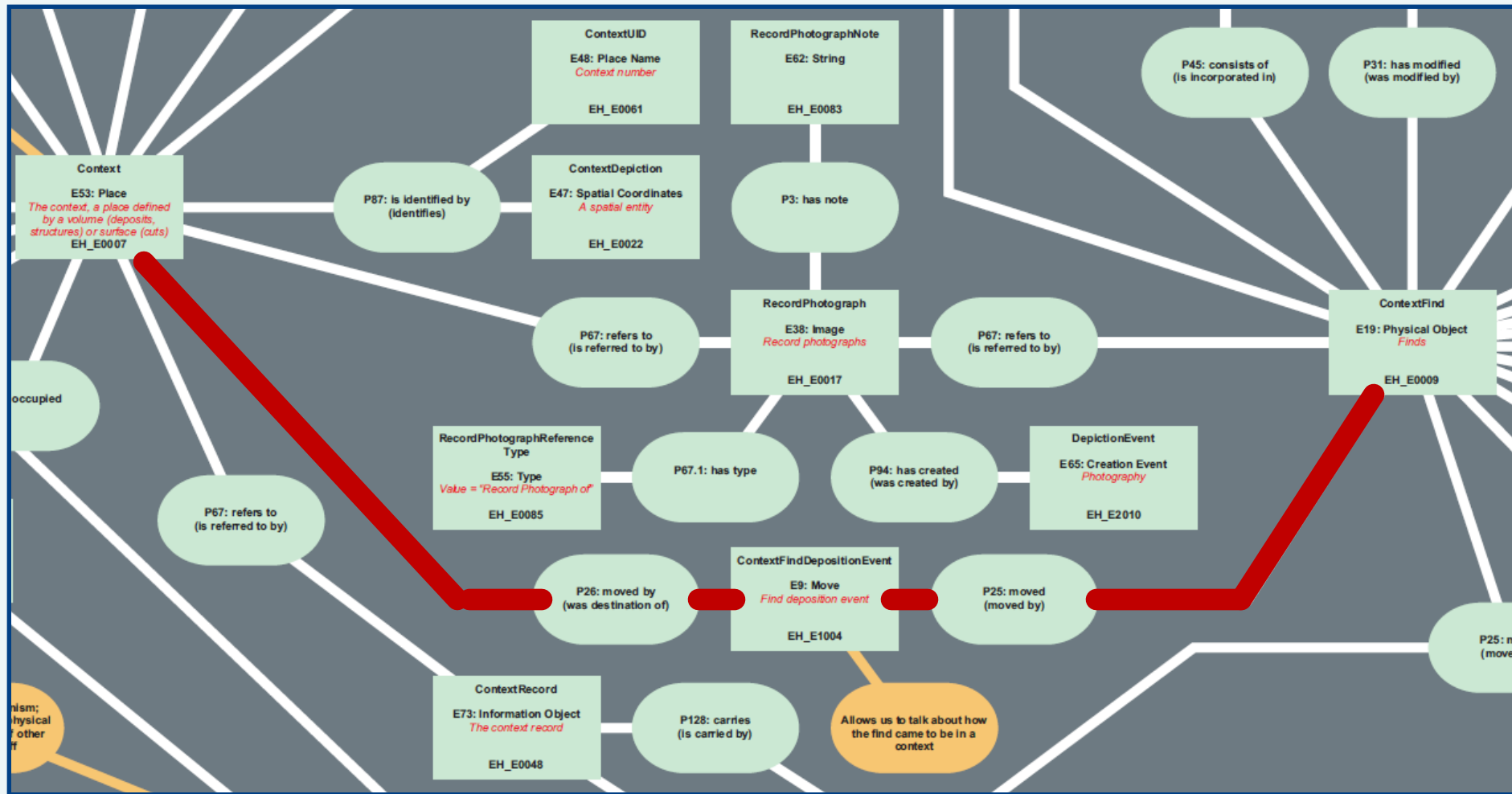
- CRM is a *high level conceptual* model
“of the intellectual structure of cultural documentation” ...
- “Users of the CRM should be aware that the definition of data entry systems requires support of community specific terminology, guidance to what should be documented and in which sequence, and application-specific consistency controls. The CRM does not provide such notions.
- By its very structure and formalism, the CRM is extensible and users are encouraged to create extensions for the needs of more specialized communities and applications.”

Definition of the CIDOC Conceptual Reference Model v4.2

Implementation issues from experience in STAR

- Need specification of implementation representations of various primitives
 - For application interoperability may need agreement on various implementation representations
- Need provision of vocabulary (terminology)
 - Our approach to employ SKOS to model vocabulary elements and link to CRM
- CRM can be extended for domain specificity in search/discovery of information
 - CRMEH allows access at a more specific level of generality for our goals
 - Still permits semantic interoperability at the original higher level
- CRM is event-based and therefore
 - Mapping a data element to CRM typically results in a chain of CRM relationships
 - Directly representing the model results in complex views for user interfaces
 - Need for 'short cuts' and simplified views for particular purposes
- Data can be mapped to multiple CRM elements
 - depending on what is considered relevant and important
 - Need for guidelines as to focus/purpose of a mapping exercise

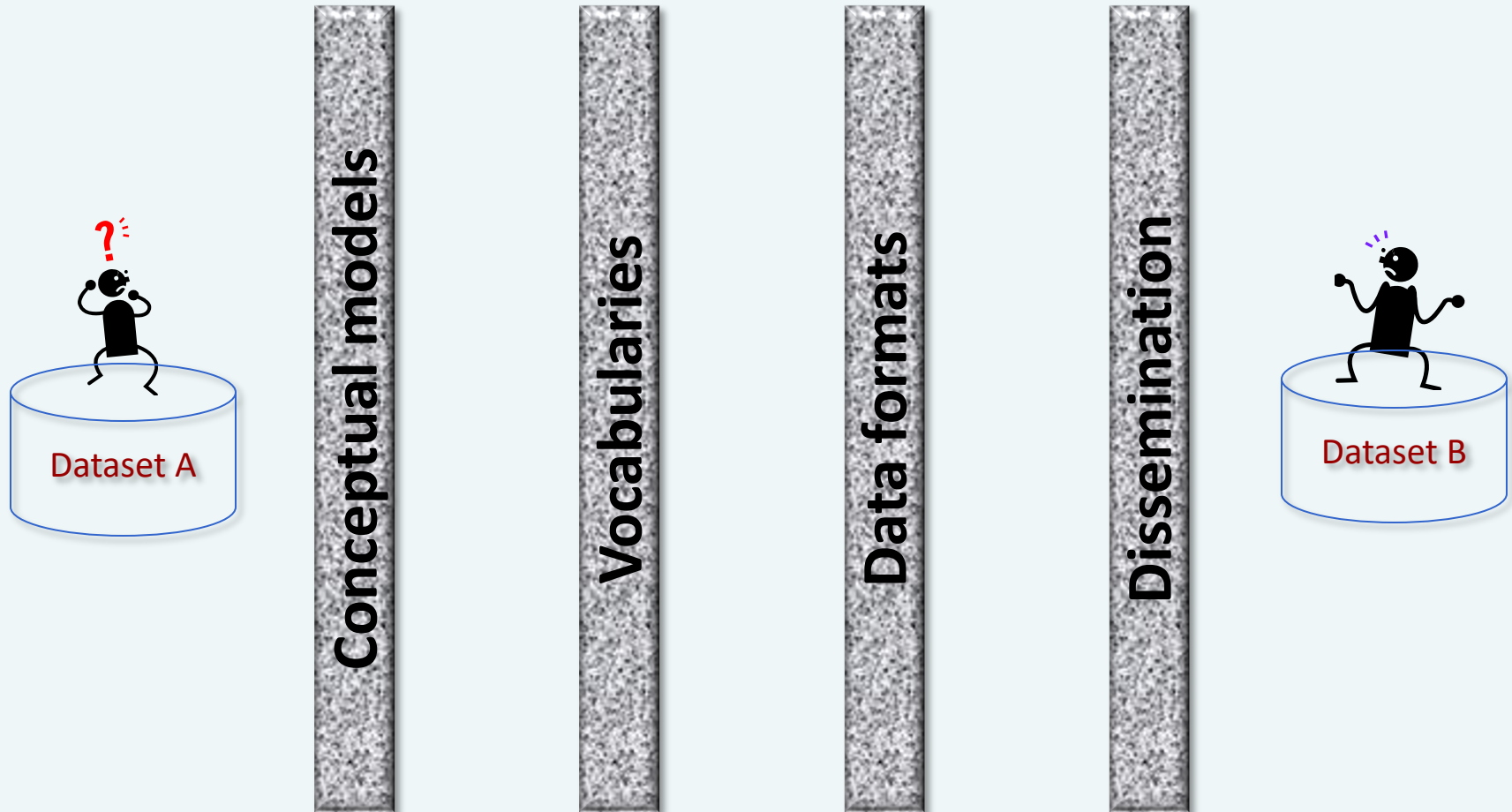
CRM is event-based and chains of relationships connect major entities



Interoperability – Practical Issues for Consideration

- Semantic / Conceptual compatibility
 - Establishing global identifiers
 - Use of controlled vocabularies (SKOS)
 - Conceptual data models and extensions (CRM/EH)
- Syntactic / Data compatibility
 - Character encodings
 - Date/time formats (modelling of periods)
 - Coordinate formats
 - Measurement units
 - Languages
- Dissemination strategy - making the data available for (re)use
 - Permissions
 - Dataset serialization formats (XML, RDF, JSON)
 - Web service access
 - Linked data access

What Barriers?...



Semantic Compatibility

- Conforming to a common conceptual data model (CRM/EH)
- Establishing unique global identifiers for known entities and concepts (URIs)
- Use of controlled vocabularies with a common data model (SKOS)

Syntactic Compatibility

- “The CRM relies on *existing* syntactic interoperability and is concerned only with adding semantic interoperability”
 - Character encodings
 - EBCDIC, ASCII, UTF-8
 - Measurement units
 - metric, imperial, antiquated(!)
 - Date / time / period formats
 - Gregorian years, 3 age system, monarchs
 - Coordinate systems
 - WGS84, NAD27, OSGB, UTM
 - Languages!

Formats representing dates and periods

- Centuries
- BC/AD years
- 3 age system
- Monarchs / emperors
- Geological periods
- Prefixes: pre, post, mid etc.
- Combinations of these

Time periods encountered

MLC2-C3

AD 341-6

Iron Age

First half 1st century?

Antonine

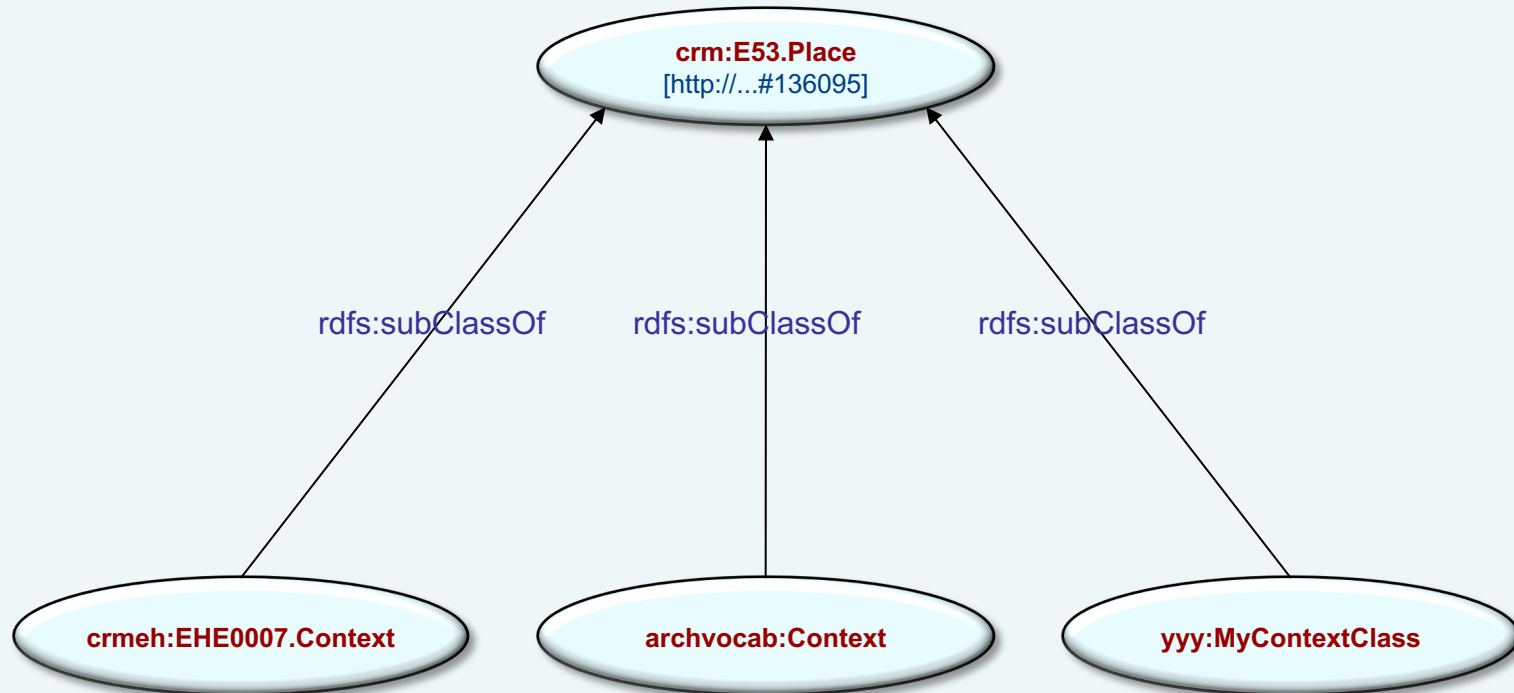
Early C3

Time period alignment – STAR.TIMELINE application

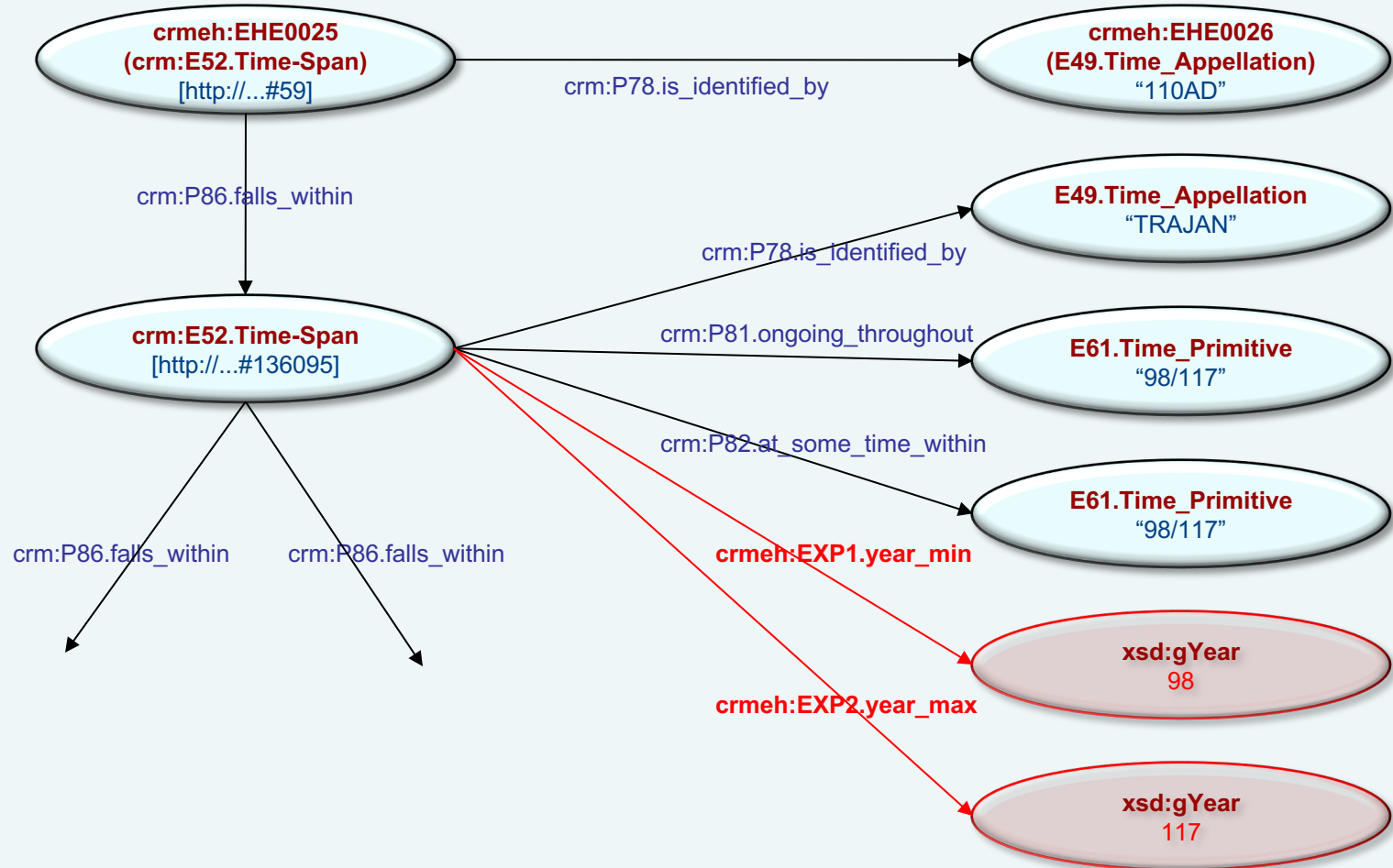
- Aligns data with closest ‘known’ periods

Data record – dates deduced from labels				Closest controlled match based on dates			
ID	Label	From	To	ID	Label	From	To
1315	AD 228-31	228	231	136122	ALEXANDER SEVERUS	222	235
1316	AD 364-78	364	378	900014	3RD QUARTER 4TH CENTURY AD	351	375
1317	AD 69-79	69	79	136087	VESPASIAN	69	79
1318	AD 270-4	270	274	136164	TETRICUS I	270	274
1319	AD 275-402	275	402	134825	4TH CENTURY AD	300	399
1320	AD 341-6	341	346	900013	2ND QUARTER 4TH CENTURY AD	326	350
1321	AD 268-70	268	270	136154	CLAUDIUS II GOTHICUS	268	270
1322	AD 367-75	367	375	900014	3RD QUARTER 4TH CENTURY AD	351	375
1324	AD 270-84	270	284	135952	LATE 3RD CENTURY	266	299
1325	AD 270-84	270	284	135952	LATE 3RD CENTURY	266	299
1326	AD 367-75	367	375	900014	3RD QUARTER 4TH CENTURY AD	351	375
1327	AD 383-8	383	388	900015	4TH QUARTER 4TH CENTURY AD	376	399
1328	AD 330-40	330	340	900013	2ND QUARTER 4TH CENTURY AD	326	350
1337	Post-medieval	1540	1901	134746	POST MEDIEVAL	1540	1901
1370	Medieval	1066	1540	134745	MEDIEVAL	1066	1540
1371	AD 1943	1943	1943	134848	SECOND WORLD WAR	1939	1945

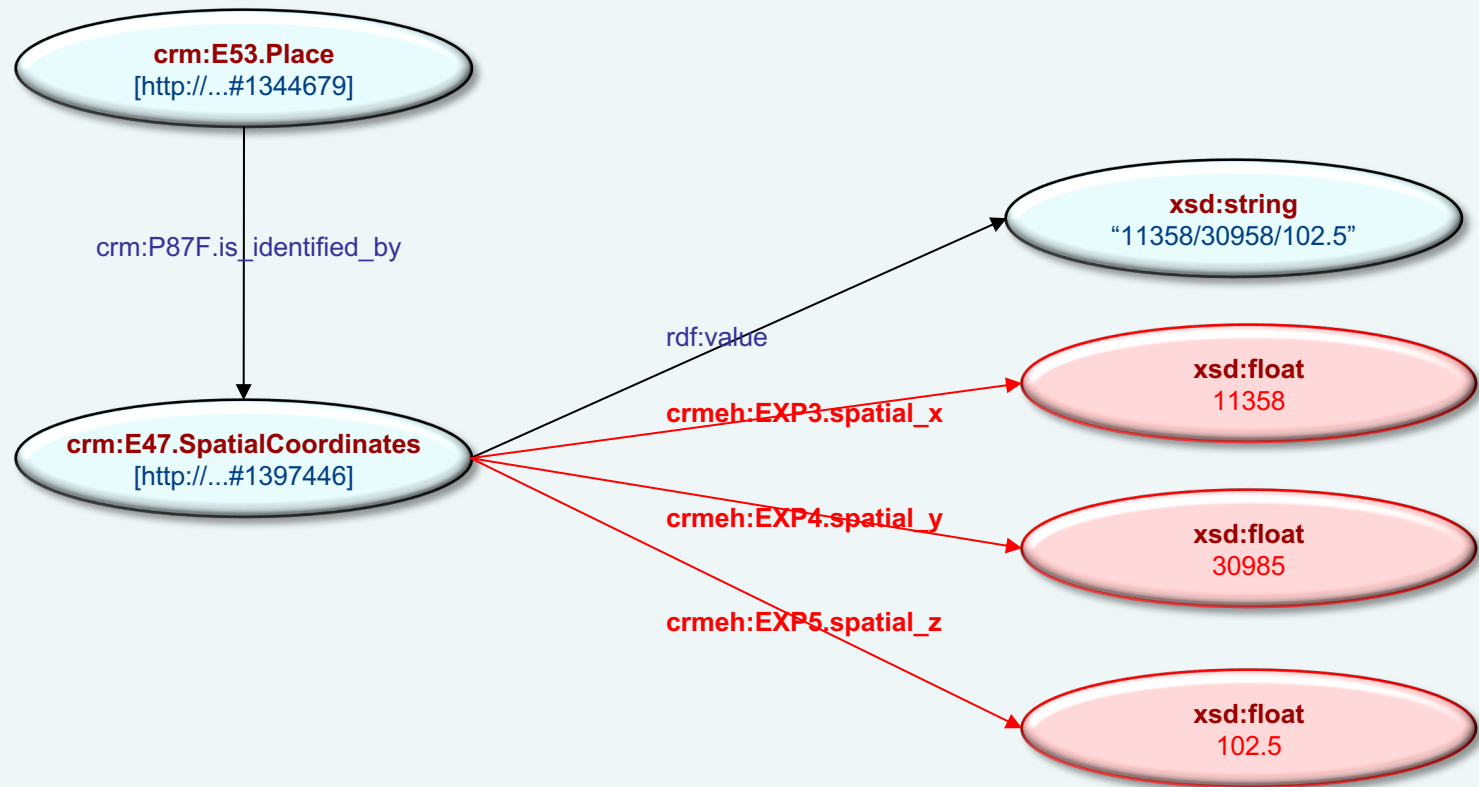
Shared common model dictates degree of interoperability



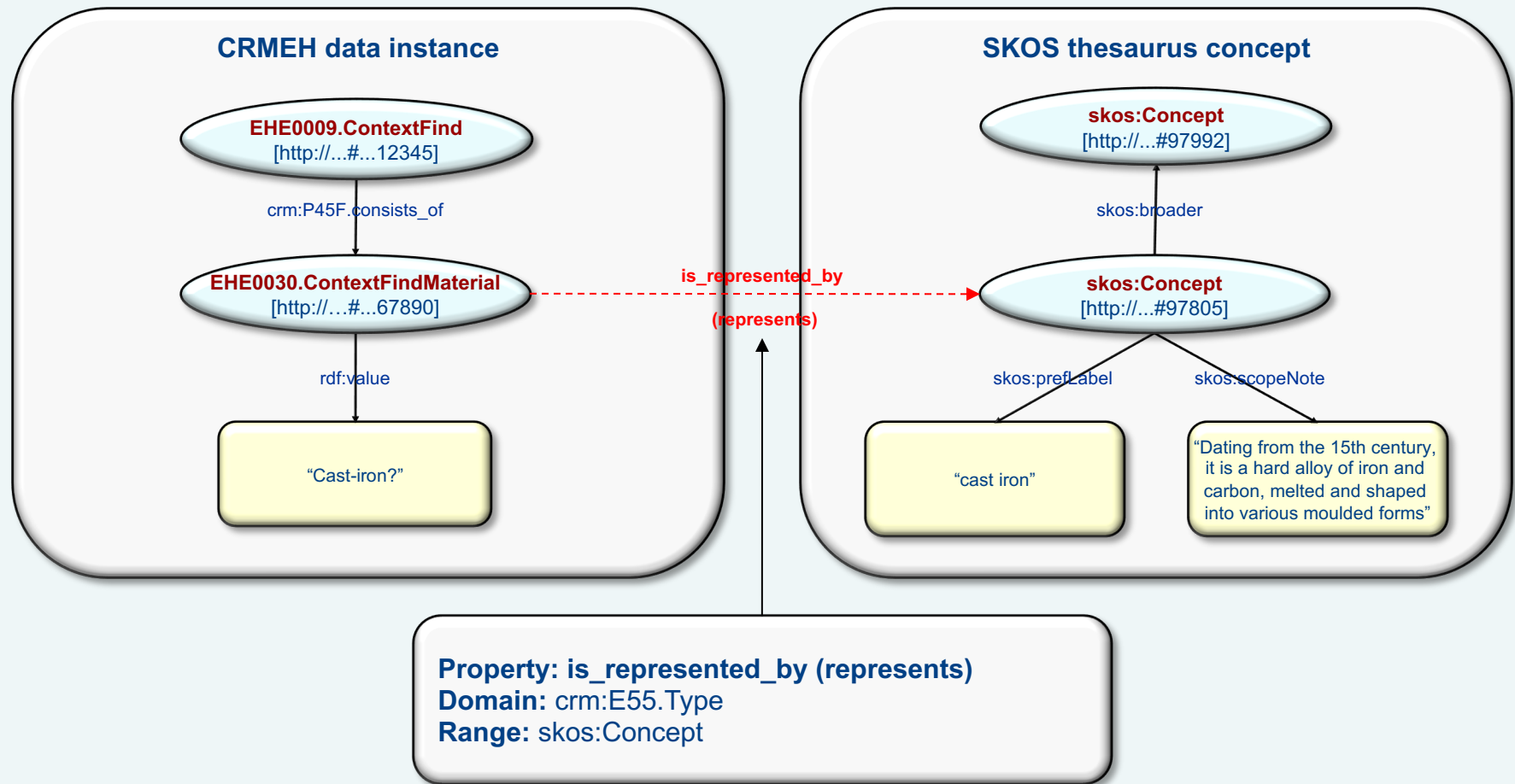
STAR implementation – extension properties for time periods



STAR implementation - extensions properties for coordinates



STAR implementation - linking CRM instances to SKOS concepts



STAR RDF Data Extraction Tool

SQLBuilder v1.0

File

Database: RRAD

Subject Predicate Object

Type Prefix Column

Literal value:

FROM clause

WHERE clause

Generated SQL:

```
SELECT DISTINCT
" AS [SUBJECTTYPE],
'http://tempuri/star/base#.rad.' & AS [SUBJECT],
" AS [PREDICATE],
" AS [OBJECTTYPE],
'http://tempuri/star/base#.rad.' & AS [OBJECT],
" AS [LITERAL]
FROM
WHERE 1 = 1
```

Resultant Data:

Test SQL

Write RDF...

Resultant extracted data (RDF/XML)

```

<?xml version="1.0"?>
<rdf:RDF xmlns:crneh="http://tempuri/star/crneh#" xmlns:crm="http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#" xmlns:rdf="http://www.w3.org/1999/02/22-
rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://tempuri/star/base#">
  <crneh:EHE0007.Context rdf:about="http://tempuri/star/base#ehe0007.rrad.context.contextno.1">
    <crm:P3F.has_note>
      <crneh: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>
      </crneh:EHE0046.ContextNote>
    </crm:P3F.has_note>
  </crneh:EHE0007.Context>
  <crneh:EHE0007.Context rdf:about="http://tempuri/star/base#ehe0007.rrad.context.contextno.2">
    <crm:P3F.has_note>
      <crneh:EHE0046.ContextNote rdf:about="http://tempuri/star/base#ehe0046.rrad.context.description.2">
        <rdf:value>Sub-division of '60' in villa area.</rdf:value>
      </crneh:EHE0046.ContextNote>
    </crm:P3F.has_note>
  </crneh:EHE0007.Context>
  <crneh:EHE0007.Context rdf:about="http://tempuri/star/base#ehe0007.rrad.context.contextno.3">
    <crm:P3F.has_note>
      <crneh:EHE0046.ContextNote rdf:about="http://tempuri/star/base#ehe0046.rrad.context.description.3">
        <rdf:value>Destruction layer of building material over main villa structure. Consists of much wallstone and roof slate (stone). Also mortar;
some tile; op sig, wall plaster and tesserae. Lying immediately beneath the lower ploughsoil (2) it mounds up over the middle of the building, thinning out
towards the edges. It is cut by a series of deep plough marks, remnant of medieval ridge and furrow which does not show at ground level. To the east of
corridor 24, the destruction material contains more mortar than on the west - also some box flue tile.</rdf:value>
      </crneh:EHE0046.ContextNote>
    </crm:P3F.has_note>
  </crneh:EHE0007.Context>
  <crneh:EHE0007.Context rdf:about="http://tempuri/star/base#ehe0007.rrad.context.contextno.4">
    <crm:P3F.has_note>
      <crneh:EHE0046.ContextNote rdf:about="http://tempuri/star/base#ehe0046.rrad.context.description.4">
        <rdf:value>A shallow linear depression orientated on an east-west axis running across the width of excavation. Original recorded
coordinates: 0980/0980</rdf:value>
      </crneh:EHE0046.ContextNote>
    </crm:P3F.has_note>
  </crneh:EHE0007.Context>

```

Data extraction - property chains

- CRM events – not explicit in datasets OR mappings
 - Additional work required to satisfy logical mappings
- E.g. **Sample** taken from **Context**:

crmeh:EHE0018.Sample [*crm:E18.PhysicalStuff*]

→ *crm:P113B.was_removed_by*

→ **crmeh:EHE2006.ContextSamplingEvent** [*crm:E80.PartRemoval*]

→ *crm:P112F.diminished*

→ **crmeh:EHE0008.ContextStuff** [*crm:E18.PhysicalStuff*]

→ **crmeh:EHP3.occupied**

→ **crmeh:EHE0007.Context** [*crm:E53.Place*]

Querying CRMEH data with SPARQL

```
# Get contexts having associated samples, where the samples have 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>
```

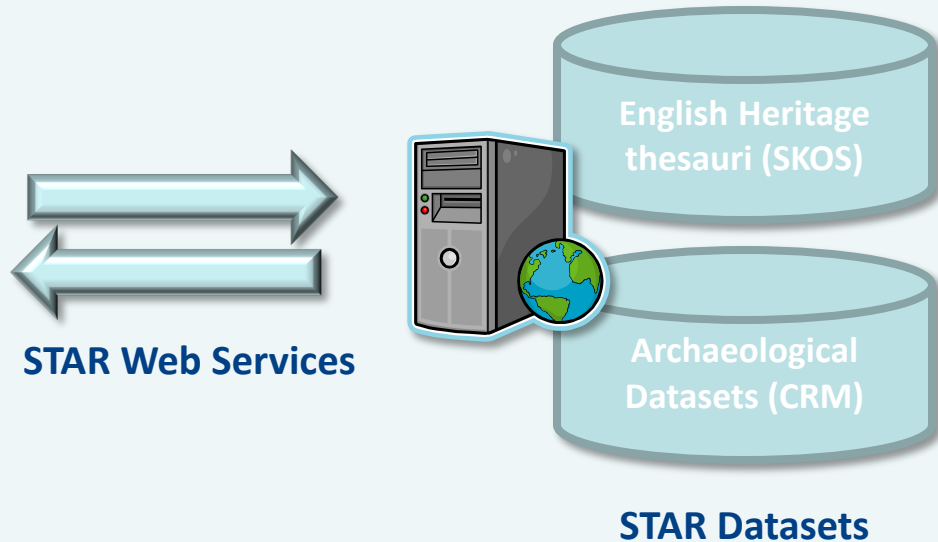
```
  <binding
name="sample"><uri>http://tempuri/star/base#ehe0018.rrad.sample.sampleno.10560</uri></binding>
  <binding
name="context"><uri>http://tempuri/star/base#ehe0007.rrad.context.contextno.4900</uri></binding>
  <binding name="notevalue"><literal>Sample taken during excavation of pit (?) '4899', 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.

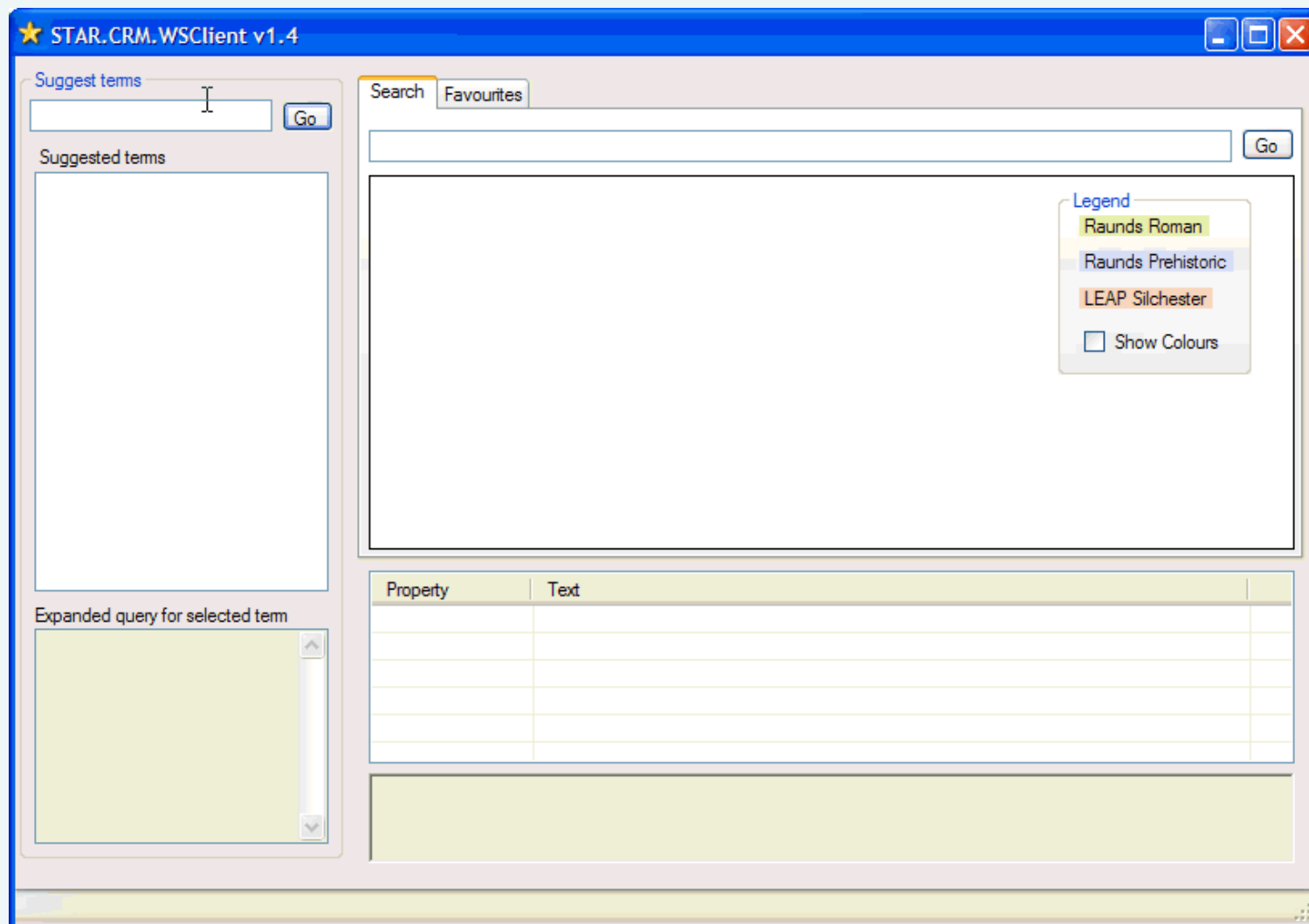
STAR – Web Services and Client Applications

- Windows applications
- Browser components
- Full text search
- Browse concept space
- Navigate via expansion
- Cross search archaeological datasets

STAR Client Applications



STAR Windows Client Application



University of Glamorgan



Query Builder

Query type: Group Context Find Sample

Context

Show all fields

- ID: (any)
- Type: (any)
- Note: (any)
- Dated: (any)
- Location: (any)
- Within group: (any)
- Within context: (any)
- Borders context: (any)
- Contains context: (any)
- Contains find: (any)
- Contains sample: (any)
- Stratigraphically above: (any)
- Stratigraphically below: (any)
- Stratigraphically same as: (any)

(JSON)

```
["context":[]]
```

(SPARQL)

```
#STAR SPARQL query: [Wed Nov 18 15:24:03 2009]
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX orm: <http://cidoc.ics.forth.gr/rdfs/cidoc_v4.2.rdfs#>
PREFIX orme: <http://tempuri/star/orme#>
```

Run query

Query Results

Query Result Item (raw)

< Back Next >

The STAR Project - University of Glamorgan, 2009



Issues to take forward or agree to differ?

- Implementation representations
 - Agreement on implementation details (eg primitives) possible/necessary?
- Vocabulary (terminology)
 - Agreement on archaeological vocabulary approaches possible/necessary?
- Extensions for domain specificity
 - Agreement on archaeological CRM extensions possible/necessary?
- Data mapping complexity
 - Agreement on mapping guidelines possible/necessary?



Ceri Binding & Douglas Tudhope
Hypermedia Research Unit, University of Glamorgan

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

University of Glamorgan



References

1. Binding C., Tudhope D., May K. (2008). Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction via the CIDOC CRM. Proceedings (ECDL 2008) 12th European Conference on Research and Advanced Technology for Digital Libraries, Aarhus, 280–290. Lecture Notes in Computer Science, 5173, Berlin: Springer. http://hypermedia.research.glam.ac.uk/media/files/documents/2008-07-05/binding_ECDL2008.pdf (preprint)
2. Binding C., Tudhope D. (2008). SKOS-based semantic web services: experiences from the STAR project. ISKO-UK KOnnecting KOMmunities Seminar: Sharing Vocabularies on the Web via SKOS, University College London. http://www.iskouk.org/SKOS_July2008.htm
3. STAR Project. <http://hypermedia.research.glam.ac.uk/kos/star/>
4. Tudhope D., Binding C., May K. 2008. Semantic interoperability issues from a case study in archaeology. In: Stefanos Kollias & Jill Cousins (eds.), Semantic Interoperability in the European Digital Library, Proceedings of the First International Workshop SIEDL 2008, 88–99, associated with 5th European Semantic Web Conference, Tenerife. <http://hypermedia.research.glam.ac.uk/media/files/documents/2008-07-05/SIEDL08-Tudhope-v3.pdf> (preprint)
5. Vlachidis A, Binding C, May K, Tudhope D. 2009. Excavating grey literature: A case study on rich indexing of archaeological documents by the use of Natural Language Processing techniques and knowledge based resources. Proceedings British Chapter of the International Society for Knowledge Organization (ISKO UK) Conference. http://www.iskouk.org/conf2009/presentations/vlachidis_ISKOUK2009_presentation.pdf