



SENESCHAL

<http://www.heritagedata.org/>



Arts & Humanities
Research Council



ENGLISH HERITAGE



Royal
Commission on the
Ancient and
Historical
Monuments of
Scotland

100
CBHC
RCAHMW
1908-2008



COMISIWN BRENHINOL
HENEBION CYMRU
ROYAL COMMISSION
ON THE ANCIENT AND HISTORICAL MONUMENTS OF WALES

ads

ARCHAEOLOGY
DATA SERVICE



The SENESCHAL Project

- **seneschal** n. *Historical*
“The steward or major-domo of a medieval great house”
- 12 month AHRC funded project: March 2013 → February 2014
- University of South Wales (formerly Glamorgan) and ADS
with Project Partners including, RCAHMS, RCAHMW, English Heritage
- Knowledge Exchange based on enhanced vocabulary services
- Make it significantly easier for data providers to index their data with uniquely identified (machine readable) controlled terminology – ie semantically enriched and compatible with Linked Data.
- Make it easier for vocabulary providers to make their vocabularies available as Linked Data. EH Thesauri and RCAHMS/W thesauri as exemplar cases.



The SENESCHAL Project

- Deliverables
 - Controlled vocabularies online
 - Vocabularies from EH, RCAHMS, RCAHMW
 - Conversion to a common standard format (SKOS)
 - Persistent globally unique identifiers for every concept
 - Made available online as Linked Open Data
 - Also downloadable data files and listings
 - Web services
 - Facilitate concept searching, browsing, suggestion, validation
 - Tools to use controlled vocabularies
 - Browser-based 'widget' user interface controls
 - Search, browse, suggest, select concepts
 - Case studies
 - Legacy data to thesaurus alignment
 - Thesaurus to thesaurus alignment
 - Third party use of project outcomes



What is Linked Data?

“The Web enables us to link related documents. Similarly it enables us to link related data. The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. Key technologies that support Linked Data are URIs (a generic means to identify entities or concepts in the world), HTTP (a simple yet universal mechanism for retrieving resources, or descriptions of resources), and RDF (a generic graph-based data model with which to structure and link data that describes things in the world).”

<http://linkeddata.org/faq> and

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.

Also see <http://data.gov.uk/linked-data>



Linked Data from (AHRC) STELLAR Project



ARCHAEOLOGY
DATA SERVICE

http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE0007_1010



1010

Property	Value
?P2_has_type	<http://data.archaeologydataservice.ac.uk/10.5284/1000365/E55_EHE0007_deposit>
?P3_has_note	fill of posthole
?P7i_witnessed	<http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE1001_1010>
?P87_is_identified_by	<http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE0061_1010>
?P89_falls_within	<http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE0003_hcb+03> <http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE0005_146> <http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE0007_1009>
?label	1010
?type	<http://purl.org/crmeh#EHE0007_Context>

Metadata [\[show\]](#)

This page shows information obtained from the SPARQL endpoint at <http://data.archaeologydataservice.ac.uk/sparql/repositories/archives>. You can query the endpoint directly with a SPARQL client or at our [SPARQL query interface](#). This data is also available as [RDF/XML](#) and [Turtle](#).

View the data in the following other browsers:

- [Disco](#)
- [Tabulator](#)

[University of York legal statements](#) | [ADS terms and conditions](#)



ARCHAEOLOGY
DATA SERVICE

THE UNIVERSITY of York

National thesaurus concepts only free text since URIs not then available, making difficulties for linking data - a point of departure for SENESCHAL project.



Ambiguity in metadata indexing

1. Google search on “Wives of Henry VIII”
2. Click on “Images”
3. Spot the problem in the TOP 2 results from the multi billion dollar biggest search company on the entire planet:



- Words are ambiguous, so using them in metadata indexing just invites incorrect search results
- We can (and should) do better than this



A problem: You say potato, I say tomato...

- Multiple datasets, multiple organisations, multiple languages
- Unification of data *structures* may be possible, BUT...
 - Incompatible terminology hinders cross search and prevents greater interoperability
 - Indexing using text is ambiguous, leading to incorrect search results
 - Applications attempting to reuse data must all individually tackle the same problems
- E.g. *Find all the iron age post holes...*
- The problem here is in the use of text to convey meaning – whereas the underlying logical concepts are actually the same
- The solution is the use of concept-based controlled vocabularies

Feature	Period
Post-hole	IRON AGE
Posthole	ron age
POST HOLE	Iron age?
POSTHLOLE	EARLY IRON AGE
POST HOLE (POSSIBLE)	250 BC
POSTHOLES	C 500-200 B.C.



Semi-controlled vocabularies

Deposit Colour			Deposit Texture	Deposit Compaction
(Reddy) Brown	Dark orange/brown	Orangy brown, very light	Firm	Plastic
9Reddy) brow				Sticky
Brown				y (wet)
Brown red				y/firm
Brown/redd				s
Dark brown				
Dark brown				
Dark grey b				
Dark orange				
Dark orange				
darker patch				
Dark orange loam	Or	Yellow brown		
		Yellow/orange brown		

"...another of my examples has something about some flint that is 'snuff coloured' & I don't know if I've ever seen snuff, let alone know what colour it is, or might have been over 150 years ago, and I would think it would make sense to take some kind of integrated approach from the outset,...."
[G. Carver]

We already have plenty of controlled vocabularies in the cultural heritage domain, however tension exists at point of data entry between being descriptive vs. controlled indexing for data retrieval – so data entry is often not restricted to controlled vocabularies. Trying to do two different things at once.

For data entry: Semi-controlled vocabularies represent a useful compromise somewhere between descriptive and controlled vocabularies, ***the best of both worlds!***

For data retrieval: ***The worst of all worlds (Re. find all the iron age post holes)***

This problem arises from trying to do two different things within a single input field. Should do both separately – describe using free text description fields, and index using controlled index fields



“But our data is already all neat and tidy”

- The project is considering issues of wider interoperability
- We have seen that words are just too ambiguous for indexing:
 - TENEMENT (RCAHMS Monument Types Thesaurus): “A large building containing a number of rooms or flats, access to which is usually gained via a common stairway.”
 - TENEMENT (EH Monument Types Thesaurus): “A parcel of land.”
- We cannot automatically disambiguate the search results originating from two separate sources
- Thesaurus **concepts** however are not ambiguous – these are clearly two *different* concepts. With *different identifiers*.
- If we supplement data with **concept identifiers** it becomes clear what is meant by “TENEMENT”, and we can also express search criteria more accurately

3 RESULTS PAGE 1 OF 1		⏪ ⏩ 1 ⏪ ⏩
TITLE	YEAR	
York Archaeological Trust Fascicules York Archaeological Trust	2011	
Anglo-Saxon Stafford. Archaeological Investigations 1954-2004. Field Reports On-line Martin Carver	2010	
Norwich, Castle Mall Norfolk Archaeological Unit	2009	
PAGE 1 OF 1		⏪ ⏩ 1 ⏪ ⏩

(Search on “tenement” - Archaeology Data Service)

<u>NT27SE 1033</u>	Edinburgh, 48, 48a Frederick Street	SHOP (19TH CENTURY), TENEMENT (18TH CENTURY)	EDINBURGH, CITY OF
<u>NT27SE 701</u>	Edinburgh, 108 West Bow	SHOP, TENEMENT (19TH CENTURY)	EDINBURGH, CITY OF
<u>NT27SW 451</u>	Edinburgh, 34, 34a Castle Street	PUBLIC HOUSE, TENEMENT	EDINBURGH, CITY OF

(Search on “tenement” - Canmore)

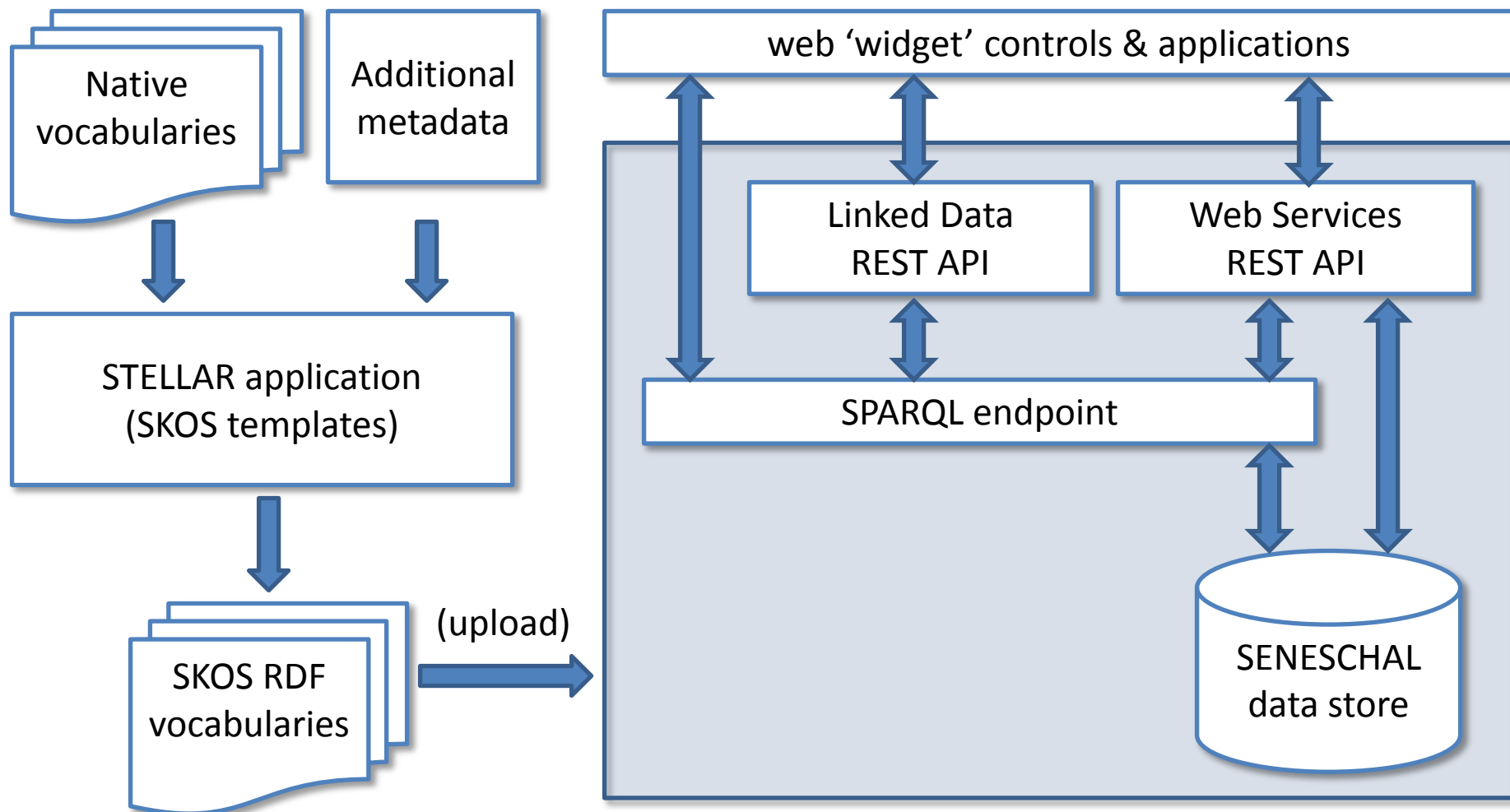


Towards a solution - SENESCHAL

- Controlled vocabularies (revisited)
 - Commonly agreed concepts, terminology and identifiers
 - Common standard format, open licensed data
 - Availability - web services, bulk downloads
- Alignment of existing data
 - Data cleansing
 - Alignment techniques
- Alignment of new data
 - Interactive embedded data entry and lookup tools
 - Vocabularies and validation at point of data entry
 - Rather than trying to solve this familiar vocabulary problem, help to prevent it from happening in the first place



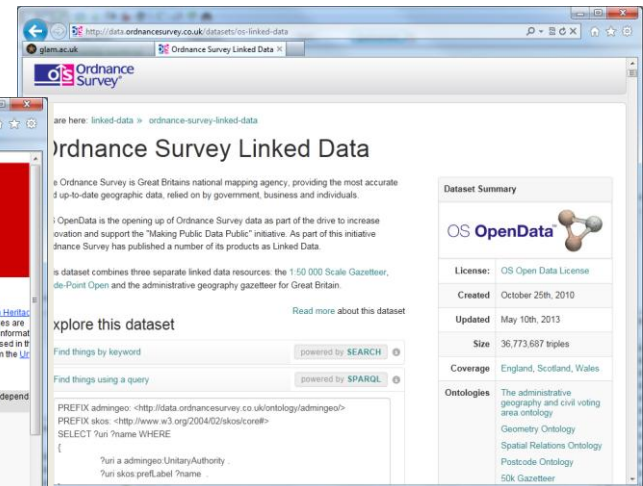
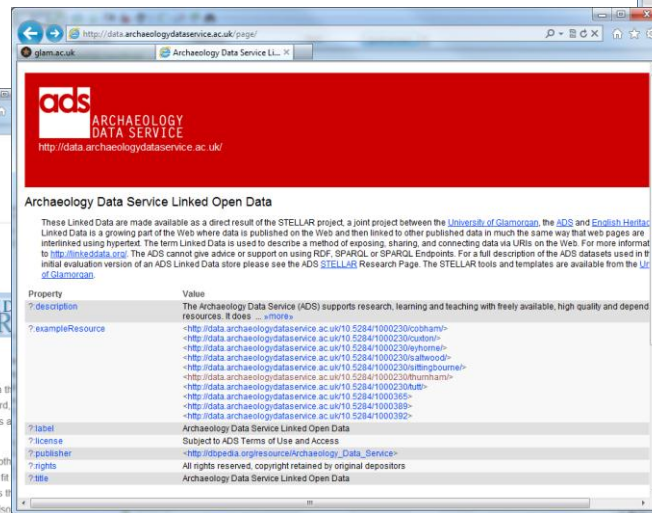
General System Architecture





Linked Data

- Making RDF format data available via the web
- Data expressed in RDF
- Using (HTTP) URIs as identifiers for things
- When someone looks up a URI, provide useful information (including links to other things)
- Will it work for cultural heritage...? Yes
 - <http://data.ordnancesurvey.co.uk/>
 - <http://collection.britishmuseum.org/>
 - <http://data.archaeologydataservice.ac.uk/>





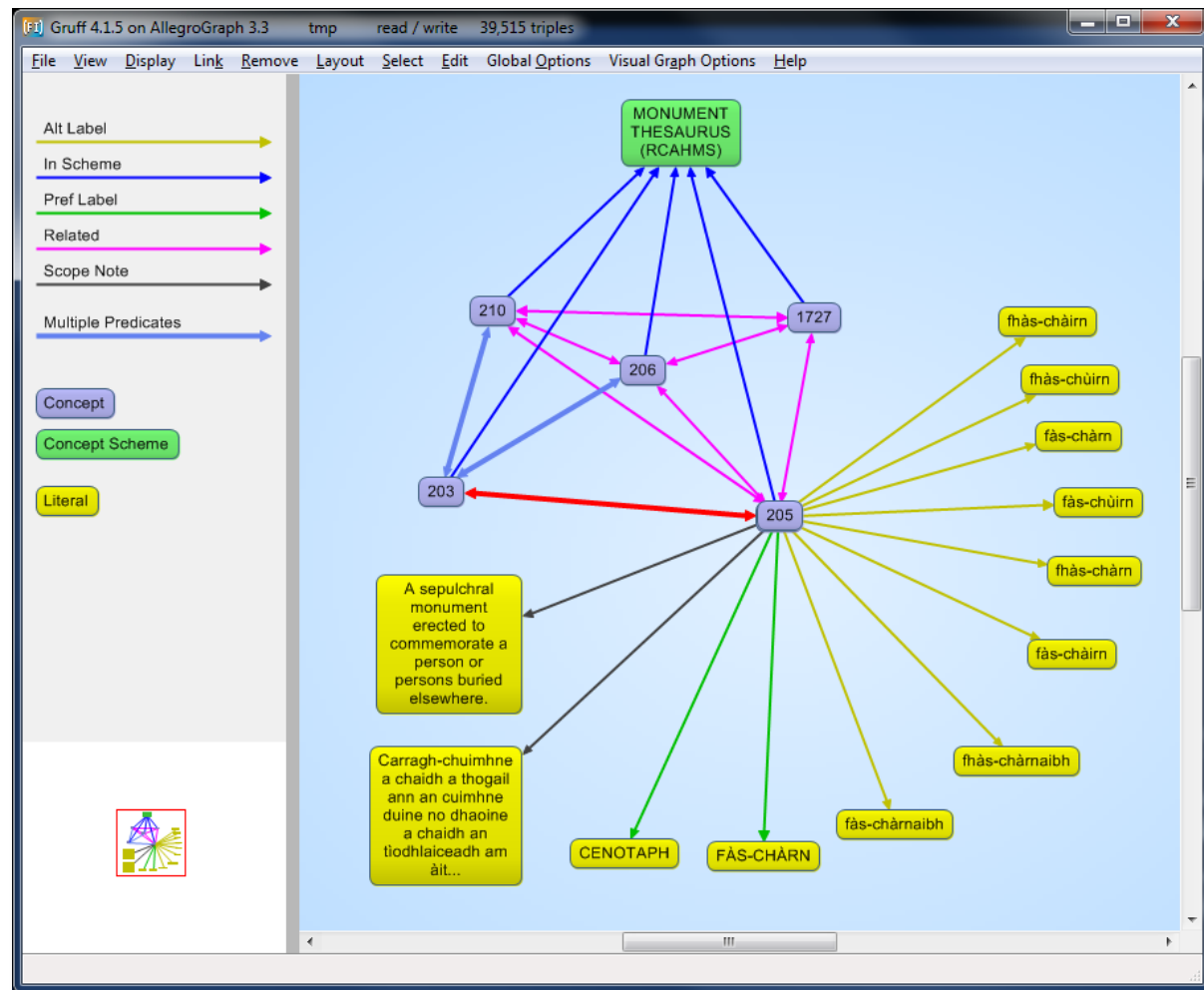
Vocabularies online as SKOS RDF Linked Data

- Vocabularies from English Heritage
 - Archaeological Sciences
 - Building Materials
 - Components
 - Event Type
 - Evidence
 - FISH Archaeological Objects
 - Maritime Craft Type
 - Monument Type
 - Periods
- Vocabularies from RCAHMS
 - Archaeological Objects Thesaurus (Adapted version of the FISH Archaeological Objects Thesaurus)
 - Maritime Craft Thesaurus
 - Monument Type Thesaurus (Multilingual - includes Scottish Gaelic translations)
- Vocabularies from RCAHMW
 - Monument Type Thesaurus
 - Period
- Moving from term based towards concept based indexing
 - Start to create links between concepts... between vocabularies... between datasets... between sites... between countries
 - Alignment from legacy data to persistent concept identifiers
 - Alignment between thesauri
 - True interoperability of (multilingual) cultural heritage resources



Concepts: Accommodating multilingual terms

- Multilingual labels & scope notes attached to concepts
- Possible to search in one language, retrieve in another





Concepts: Accommodating colloquial terms

Dr. Johnson: (proudly) *“Here it is sir, the very cornerstone of English scholarship. This book contains every word in our beloved language.”*

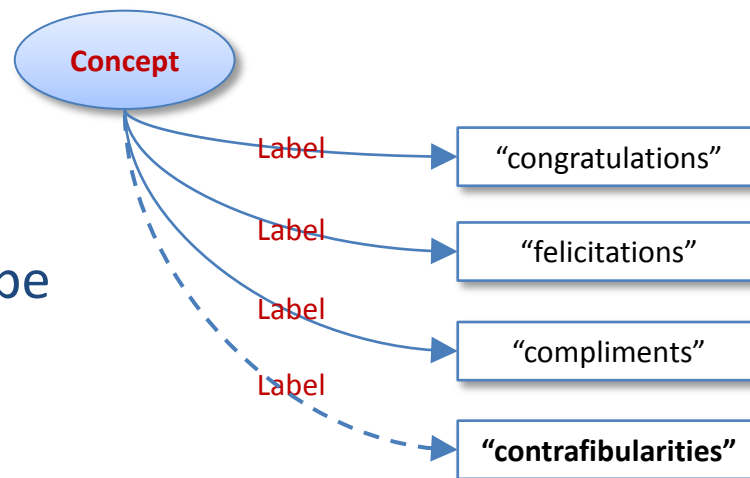
Blackadder: *“every single one sir? [...] In that case I hope you will not object if I also offer my most enthusiastic ... **contrafibularities**”.*

Dr. Johnson: *“What?”*

Blackadder: *“**contrafibularities** sir – it is a common word down our way.”*

Dr. Johnson: (flustered and scribbling) *“Damn...”*

Blackadder’s mischievous suggestion may be a new *term*, but it is not a new *concept*. It fits into the existing concept structure, further enriching the entry vocabulary.





Linked Data – Identifiers

- The project implemented an organization-neutral base URI:
 - <http://purl.org/heritagedata/>
- This is the base URI for all scheme and concept identifiers:
 - <http://purl.org/heritagedata/schemes>
 - return list of all SKOS concept schemes held
 - <http://purl.org/heritagedata/schemes/{id}>
 - return details of specified SKOS concept scheme
 - <http://purl.org/heritagedata/schemes/{id}.html>, [.n3](http://purl.org/heritagedata/schemes/{id}.n3), [.rdf](http://purl.org/heritagedata/schemes/{id}.rdf), [.json](http://purl.org/heritagedata/schemes/{id}.json)
 - return different serializations of the data, obtained either by content negotiation or by direct request including extension
 - <http://purl.org/heritagedata/schemes/{id}/concepts/{id}>
 - return details of specified SKOS concept
 - <http://purl.org/heritagedata/schemes/{id}/concepts/{id}.html>, [.n3](http://purl.org/heritagedata/schemes/{id}/concepts/{id}.n3), [.rdf](http://purl.org/heritagedata/schemes/{id}/concepts/{id}.rdf), [.json](http://purl.org/heritagedata/schemes/{id}/concepts/{id}.json)
 - return different serializations of the data, obtained either by content negotiation or by direct request including extension



Linked Data - Implementation

- <http://purl.org/heritagedata/schemes/2>
 - Redirects to <http://heritagedata.org/live/schemes/2.html> (HTML representation of scheme data)
- <http://purl.org/heritagedata/schemes/2/concepts/501497>
 - Redirects to <http://heritagedata.org/live/schemes/2/concepts/501497.html> (HTML representation of concept data)
- <http://purl.org/heritagedata/schemes/2/concepts/501497.rdf>
- <http://purl.org/heritagedata/schemes/2/concepts/501497.ttl>
- <http://purl.org/heritagedata/schemes/2/concepts/501497.json>
 - Each will return the *same* concept data, in the requested RDF serialisation format:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:ns0="http://creativecommons.org/ns#" xmlns:ns1="http://www.w3.org/2004/02/skos/core#" xmlns:ns2="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  <rdf:Description rdf:about="http://purl.org/heritagedata/schemes/2/concepts/501497">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <ns0:license rdf:resource="http://reference.data.gov.uk/id/open-government-licence"/>
    <ns0:attributionURL rdf:resource="http://www.rcahms.gov.uk"/>
    <ns0:attributionName xml:lang="en">RCAHMS</ns0:attributionName>
    <ns1:topConceptOf rdf:resource="http://purl.org/heritagedata/schemes/2"/>
    <ns1:inScheme rdf:resource="http://purl.org/heritagedata/schemes/2"/>
    <ns1:prefLabel xml:lang="en">RELIGION OR RITUAL</ns1:prefLabel>
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5017">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5012">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5017">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5016">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5012">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5015">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5015">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5014">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5013">
    <ns1:narrower rdf:resource="http://purl.org/heritagedata/schemes/2/concepts/5012">
    <ns1:scopeNote xml:lang="en">This is the top term for the class covering objects us
  See RELIGION OR RITUAL Class List for narrow terms.</ns1:scopeNote>
    <ns2:publisher rdf:resource="http://www.rcahms.gov.uk"/>
    <ns2:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2007-07-19T07:40:45</ns2:modified>
    <ns2:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-07-17T10:08:34</ns2:issued>
  </rdf:Description>
</rdf:RDF>

@prefix ns1: <http://creativecommons.org/ns#> .
@prefix ns2: <http://purl.org/dc/terms/> .

<http://purl.org/heritagedata/schemes/2/concepts/501497>
  rdf:type ns0:Concept ;
  ns1:license <http://reference.data.gov.uk/id/open-government-licence> ;
  ns1:attributionURL <http://www.rcahms.gov.uk> ;
  ns1:attributionName "RCAHMS"@en ;
  ns1:topConceptOf <http://purl.org/heritagedata/schemes/2> ;
  ns0:inScheme <http://purl.org/heritagedata/schemes/2> ;
  ns0:prefLabel "RELIGION OR RITUAL"@en ;
  ns0:narrower <http://purl.org/heritagedata/schemes/2/concepts/501768> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501253> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501742> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501635> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501251> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501509> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501500> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501499> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501380> ,
    <http://purl.org/heritagedata/schemes/2/concepts/501274> ;
  ns0:scopeNote "This is the top term for the class covering objects used or associated
    with religious, ritual or funerary activities. See RELIGION OR RITUAL Class List for
    narrow terms."@en ;
  ns2:publisher <http://www.rcahms.gov.uk> ;
  ns2:modified "2007-07-19T07:40:45"^^<http://www.w3.org/2001/XMLSchema#dateTime> ;
  ns2:identifier "http://purl.org/heritagedata/schemes/2/concepts/501497" ;
  ns2:issued "2013-07-17T10:08:34"^^<http://www.w3.org/2001/XMLSchema#dateTime> .

<http://purl.org/heritagedata/schemes/2/concepts/501497> : {
  "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" : [
    { "value" : "http://www.w3.org/2004/02/skos/core#Concept", "type" : "uri" }
  ],
  "http://purl.org/dc/terms/identifier" : [
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501497", "type" : "literal" }
  ],
  "http://creativecommons.org/ns#license" : [
    { "value" : "http://reference.data.gov.uk/id/open-government-licence", "type" : "uri" }
  ],
  "http://www.rcahms.gov.uk" : [
    { "value" : "http://www.rcahms.gov.uk", "type" : "uri" }
  ],
  "http://purl.org/dc/terms/publisher" : [
    { "value" : "http://www.rcahms.gov.uk", "type" : "uri" }
  ],
  "http://purl.org/dc/terms/issued" : [
    { "value" : "2013-07-17T10:08:34", "type" : "literal", "datatype" : "http://www.w3.org/2001/XMLSchema#dateTime" }
  ],
  "http://purl.org/2004/02/skos/core#topConceptOf" : [
    { "value" : "http://purl.org/heritagedata/schemes/2", "type" : "uri" }
  ],
  "http://www.w3.org/2004/02/skos/core#inScheme" : [
    { "value" : "http://purl.org/heritagedata/schemes/2", "type" : "uri" }
  ],
  "http://purl.org/2004/02/skos/core#prefLabel" : [
    { "value" : "RELIGION OR RITUAL", "type" : "literal", "lang" : "en" }
  ],
  "http://purl.org/2004/02/skos/core#narrower" : [
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501274", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501380", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501499", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501500", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501509", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501251", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501635", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501742", "type" : "uri" },
    { "value" : "http://purl.org/heritagedata/schemes/2/concepts/501768", "type" : "uri" }
  ],
  "http://purl.org/2004/02/skos/core#scopeNote" : [
    { "value" : "This is the top term for the class covering objects used or associated with
      religious, ritual or funerary activities. See RELIGION OR RITUAL Class List for narrow terms.",
      "type" : "literal", "lang" : "en" }
  ],
  "http://purl.org/dc/terms/modified" : [
    { "value" : "2007-07-19T07:40:45", "type" : "literal", "datatype" : "http://www.w3.org/2001/XMLSchema#dateTime" }
  ]
}
```



Thesaurus searching and browsing

Heritage Data: Linked Data ... x

Heritage Data

Linked Data Vocabularies for Cultural Heritage

Scheme List | Concept Search | SPARQL Query | About The Project

Search

Scheme: **FISH Archaeological Objects Thesaurus** Label: **brooch**

37 matches returned for search on 'brooch'

- AESICA BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141439
- ANIMAL BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141228
- ANNULAR BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/97114
- AUCISSA BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141229
- Ansate Brooch**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141231
- BOW AND FANTAIL BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141230
- BOW BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/96662
- BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/96665
- BROOCH PIN**
http://purl.org/heritagedata/schemes/mda_obj/concepts/96666
- Brooch Spring**
http://purl.org/heritagedata/schemes/mda_obj/concepts/96665
- CATERPILLAR BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141231
- COLCHESTER BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141232
- CROSSBOW BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141233
- CRUCIFORM BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141235
- DISC BROOCH**
http://purl.org/heritagedata/schemes/mda_obj/concepts/141227

Heritage Data: Linked Data ... x

Heritage Data

Linked Data Vocabularies for Cultural Heritage

Scheme List | Concept Search | SPARQL Query | About The Project

<http://purl.org/heritagedata/schemes/1/concepts/447>

Property	Value
rdf:type	skos:Concept
cc:license	http://reference.data.gov.uk/id/open-government-licence
cc:attributionURL	http://www.rcahms.gov.uk
cc:attributionName	RCAHMS
skos:inScheme	Monument Type Thesaurus (Scotland)
skos:prefLabel	SCHOOLHOUSE
skos:prefLabel	TAIGH-SGOILE [gd]
skos:broader	HOUSE
skos:broader	EDUCATION
skos:scopeNote	A dwelling attached to a school, usually occupied by a school teacher.
skos:scopeNote	Àite-còmhnaidh a tha co-cheangailte ri sgoil, mar is trice bhiodh tidsear na sgoile a' fuireach ann. [gd]
skos:altLabel	thighean-sgoile [gd]
skos:altLabel	thaighean-sgoile [gd]
skos:altLabel	tighean-sgoile [gd]
skos:altLabel	taighean-sgoile [gd]
skos:altLabel	thaighe-sgoile [gd]
skos:altLabel	taighe-sgoile [gd]
skos:altLabel	thigh-sgoile [gd]
skos:altLabel	thaigh-sgoile [gd]
skos:altLabel	tigh-sgoile [gd]
skos:altLabel	taigh-sgoile [gd]
skos:related	SCHOOL
dct:publisher	http://www.rcahms.gov.uk
dct:created	2006-07-03T14:26:54
dct:identifier	http://purl.org/heritagedata/schemes/1/concepts/447
dct:issued	2013-07-17T10:06:54

RDF downloads ([N-Triples](#) [Turtle](#) [JSON](#) [XML](#))



Web services and user interface widgets



Web Services

- URI service calls returning JSON data
 - <http://www.heritagedata.org/blog/services/>
- All services available as Open Source
 - Creative Commons Attribution (CC-BY) license
 - <https://github.com/cbinding/SENESCHAL>
- Service calls
 - /getSchemes
 - E.g. <http://heritagedata.org/live/services/getSchemes?limit=3>
 - /getTopConceptsForScheme
 - /getConceptsForScheme
 - /getConceptRelations
 - /getConceptLabels
 - /getConceptLabelMatch
 - General term search facilities
 - /getConceptExists
 - For term validation



```
[
  {
    "uri": "http://purl.org/heritagedata/schemes/3",
    "label": "Maritime Craft Thesaurus (Scotland)",
    "label lang": "en",
    "description": "Types of craft that survive as wrecks, or are documented as losses, in Scottish maritime waters.",
    "attribution": "RCAHMS"
  },
  {
    "uri": "http://purl.org/heritagedata/schemes/eh_tmc",
    "label": "MARITIME CRAFT TYPE (EH)",
    "label lang": "en",
    "description": "A thesaurus of craft types which survive as wrecks in English Heritage\\ufffds maritime record",
    "attribution": "English Heritage"
  },
  {
    "uri": "http://purl.org/heritagedata/schemes/eh_tbm",
    "label": "BUILDING MATERIALS (EH)",
    "label lang": "en",
    "description": "Thesaurus of main constructional material types (eg. the walls) for indexing of monuments.",
    "attribution": "English Heritage"
  }
]
```

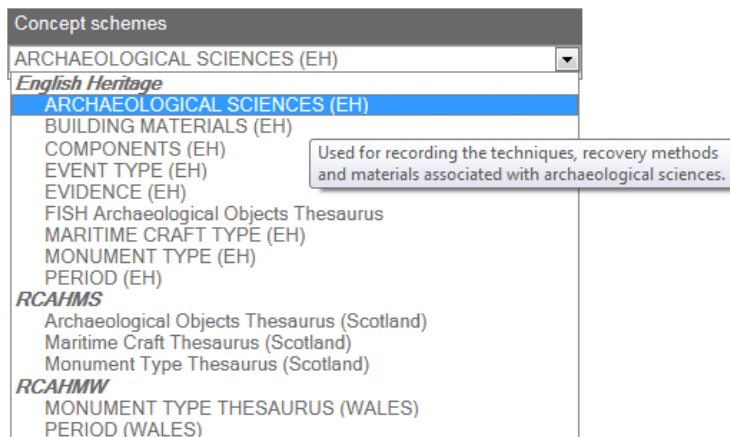


User interface controls - widgets

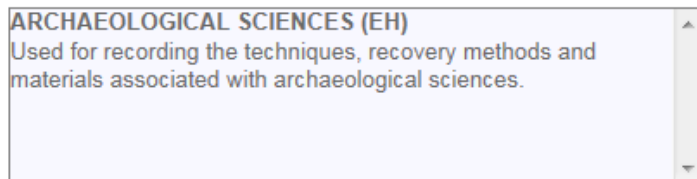
- Javascript widgets. Function in any modern browser, on PC, Mac, smartphone, tablet, console etc.
 - <http://www.heritagedata.org/blog/widgets/>
- All widgets are available Open Source
 - Creative Commons Attribution (CC-BY) license
 - <https://github.com/cbinding/SENESCHAL>
- Using only the web services as their data source (eat your own dog food)
- Configurable to work with specific concept schemes; can be easily embedded into web applications and data entry forms, see online working example pages



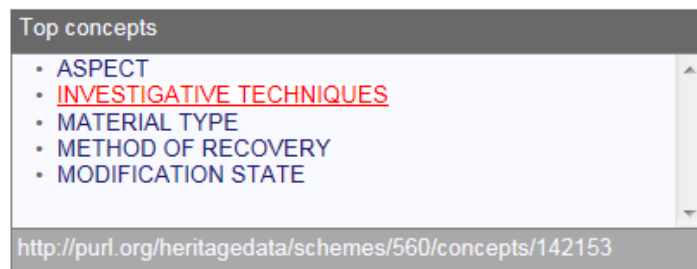
Scheme widgets



(scheme list)

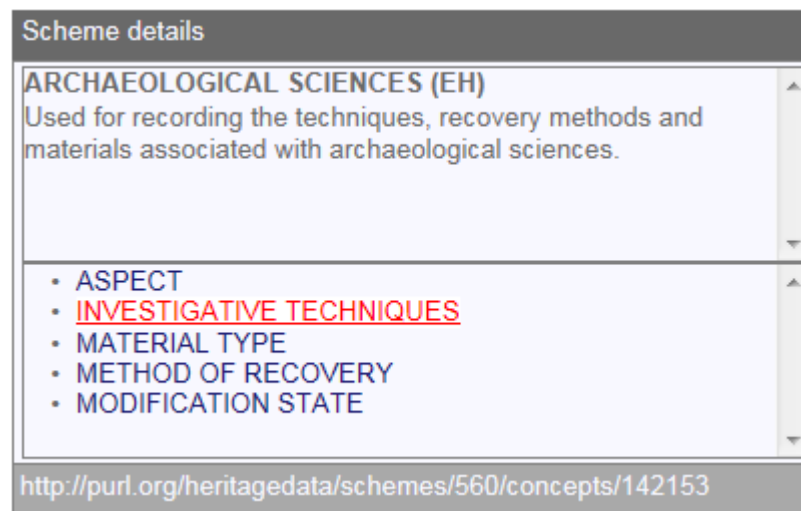


(scheme details)



(top concepts)

- Scheme list
- Scheme details
- Top concepts
- Composite control



(composite control)



Concept widgets

- Concept details
- Concept relations
- Composite control

SCHOOLHOUSE
TAIGH-SGOILE [gd]
thighean-sgoile [gd]
thaighean-sgoile [gd]
tighean-sgoile [gd]
taighean-sgoile [gd]
thaighe-sgoile [gd]
taighe-sgoile [gd]
thigh-sgoile [gd]
thaigh-sgoile [gd]
tigh-sgoile [gd]
taigh-sgoile [gd]
A dwelling attached to a school, usually occupied by a school teacher.
Àite-còmhnaidh a tha co-cheangailte ri sgoil, mar is trice bhiodh tidsear na sgoile a' fuireach ann. [gd]

(concept details)

BT DWELLING
BT MONUMENT (BY FORM)
NT AISLED BUILDING
NT BLACKHOUSE
NT BUNGALOW
NT BYRE DWELLING
NT CABETAKERS HOUSE

(concept relations)



Concept details

BOW BROOCH
Any brooch with the pin sprung or hinged at one end and the catchplate at the other end of a curved or flat bow. Can be one or two piece construction.

BT BROOCH
NT **AESICA BROOCH**
NT AUCISSA BROOCH
NT BOW AND FANTAIL BROOCH
NT CATERPILLAR BROOCH
NT COLCHESTER BROOCH
NT CROSSBOW BROOCH
NT DOLPHIN BROOCH
NT HEADSTUD BROOCH
NT HORN BOW BROOCH

http://purl.org/heritagedata/schemes/mda_obj/concepts/141439

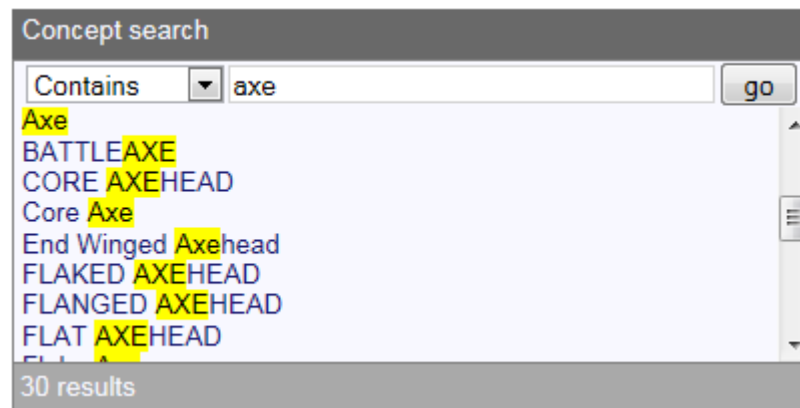
(composite control)



Term widgets

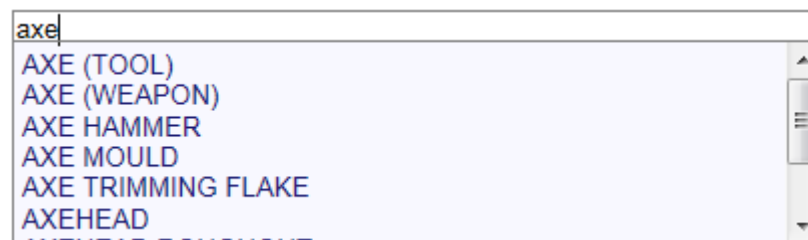
■ Term search

- Configure to search within specific concept scheme
- Search on preferred and alternate terms
- Search on multiple languages
- Starts with / contains
- Highlight matches
- Returns selected concept identifier



■ Term suggestion

- Configure to suggest from specific concept scheme
- interactive drop-down list as you type
- Returns selected concept identifier






Early adoption




Early adoption

- Portable Antiquities Scheme - mapping of PAS terminology to SENESCHAL Concept URIs




Portable
Antiquities
Scheme
www.finds.org.uk

[Log in](#) | [Register](#)

[All artefacts
& coins](#) 

[All images](#)

[Search
database](#) 

[Home](#) » [Finds database](#) » [Terminology](#) » Period

Details for the **EARLY MEDIEVAL** period

Date range: AD 410 to AD 1066

Notes:

This period is diverse and must be qualified wherever possible with Early, Middle or Late. The currently accepted chronology (spring 2013) is given below, but this is likely to change in the near future. English Heritage do not specify the boundaries of these sub-periods.

Early, 410 to 700

Middle, 701 to 850

Late, 851-1066

Linked data

- SENESCHAL URI: http://purl.org/heritagedata/schemes/eh_period/concepts/EM
- British Museum URI: <http://collection.britishmuseum.org/id/thesauri/x41022>

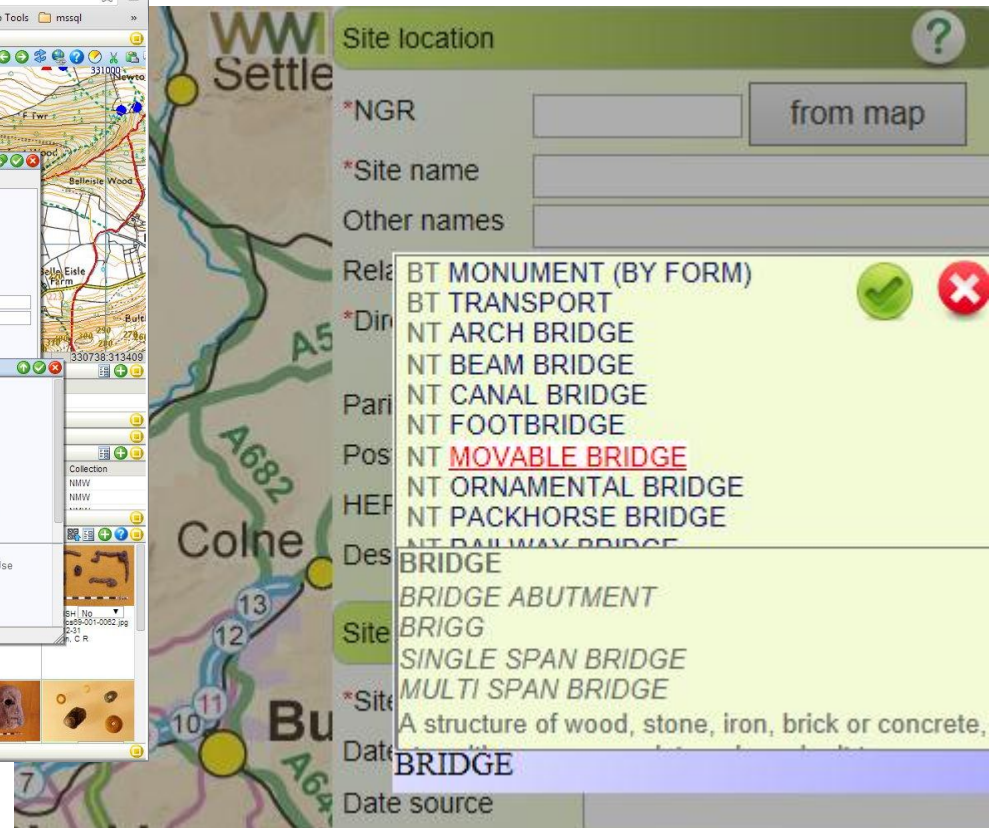
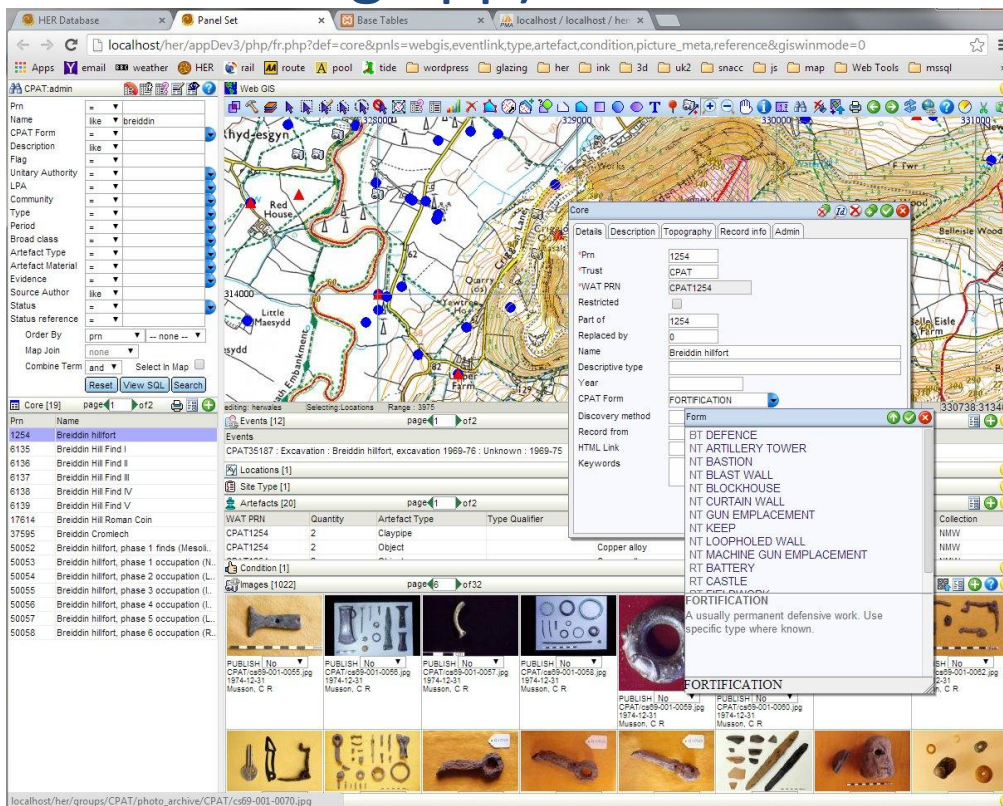


-
- The figure consists of two screenshots of the ADS Collection Management System. The left screenshot shows a modal dialog titled 'Event Type Lookup'. It contains a text input field with the word 'excavation' and a 'Lookup' button. Below the input is a 'Results:' section with a table. The table has three columns: 'Info', 'Term', and 'Select'. It lists six excavation types, each with an information icon, the term, and a 'select' button. The right screenshot shows the main application interface. At the top, it says 'ADS Collection Management System'. Below is a navigation bar with links like 'Home', 'Tracking', 'Collections', 'People', 'Admin', 'Project Docs', 'Advice', and 'Advanced Search'. The main content area shows details for a collection named 'Burdale: an Anglian settlement in the Yorkshire Wolds'. It includes a 'View this Collection' section with tabs for 'General', 'Coverage', and 'Relationships'. A dropdown menu is open, showing a list of subjects. The 'Subjects' section contains a list of subjects with search icons, including 'ENCLOSED SETTLEMENT', 'FIELD SYSTEM', 'INDUSTRIAL SITE', 'INDUSTRIAL', 'DEFENCE', and 'AGRICULTURE AND SUBSISTENCE'. There are also fields for 'Coordinates' and 'Bounding Box'.



Early adoption (continued)

- Clwyd-Powys Archaeological Trust (SENESCHAL widgets embedded into HER application and mobile field recording app)





Data alignment



Bulk data alignment exercise

- Bulk metadata alignment process
 - ADS OASIS
 - ADS ImageBank
- Alignment of specific fields against 3 controlled vocabularies
 - Monument types
 - Object types
 - Periods

```
<record>
  <header>
    <identifier>oai:archaeologydataservice.ac.uk:GLL_18031</identifier>
    <timestamp>2012-07-02</timestamp>
    <setSpec>GLL</setSpec>
  </header>
  <metadata>
    <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:gll="http://archaeologydataservice.ac.uk/archives/view/greylit/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <dc:title>Archaeological excavation on a development at Centurion House, Old Chester Road, Derby</dc:title>
      <gll:author>Kinsley, G.</gll:author>
      <dc:description>A4, comb binding</dc:description>
      <dc:publisher>Trent & Peak Archaeology</dc:publisher>
      <gll:published>2004</gll:published>
      <dc:date>2012-07-02 00:00:00.0</dc:date>
      <dc:format>1 pdf</dc:format>
      <dc:identifier>18031</dc:identifier>
      <gll:link>http://archaeologydataservice.ac.uk/archives/view/greylit/details.cfm?id=18031</gll:link>
      <gll:source>OASIS</gll:source>
      <dc:coverage>ENGLAND, DERBYSHIRE, DERBY, DERBY</dc:coverage>
      <gll:place>Nottingham</gll:place>
      <gll:site>Centurion House, Old Chester Road, Derby</gll:site>
      <dc:coverage_type="spatial">SK 355 376</dc:coverage>
      <dc:subjectMonuments>MODERN-LAYER</dc:subjectMonuments>
      <dc:subjectMonuments>UNCERTAIN-LAYER</dc:subjectMonuments>
      <dc:subjectFinds>MODERN-BRICK</dc:subjectFinds>
      <dc:subjectFinds>POST MEDIEVAL-BRICK</dc:subjectFinds>
      <dc:subjectFinds>MODERN-POTTERY</dc:subjectFinds>
      <dc:coverage_type="temporal">MODERN</dc:coverage>
      <dc:coverage_type="temporal">UNCERTAIN</dc:coverage>
      <dc:coverage_type="temporal">POST MEDIEVAL</dc:coverage>
    </oai_dc:dc>
  </metadata>
</record>
```



Typical alignment issues encountered

- Simple spelling errors
 - POSTHOLE", "CESS PITT", "FURROWS", FLINT SCRAPER"
- Alternate word forms
 - "BOUNDARY"/"BOUNDARIES", "GULLEY"/"GULLIES"
- Prefixes / suffixes
 - "RED HILL (POSSIBLE)", "TRACKWAY (COBBLED)", "CROFT?", "CAIRN (POSSIBLE)", "PORTAL DOLMEN (RE-ERECTED)"
- Nested delimiters
 - "POTTERY, CERAMIC TILE, IRON OBJECTS, GLASS"
- Terms not intended for indexing
 - "NONE", "UNIDENTIFIED OBJECT", "N/A", "NA", "INCOHERENT"
- Terms that would not be in (any) thesauri
 - "WOTSITS PACKET", "CHARLES 2ND COIN", "ROMAN STRUCTURE POSSIBLY A VILLA", "ST GUTHLACS BENEDICTINE PRIORY", "WORCESTER-BIRMINGHAM CANAL", "KUNGLIGA SLOTTET", "SUB-FOSSIL BEETLES"
- More specific phrases
 - "SIDE WALL OF POT WITH LUG", "BRICK-LINED INDUSTRIAL WELL OR MINE SHAFT", "ALIGNMENT OF PLATFORMS AND STONES"



Data alignment approach

- *Levenshtein* edit distance algorithm
 - Measures optimal number of character edits required to change one string into another
 - Accommodates small spelling differences/errors
- Bulk alignment process
 - Compares each value to all terms from specified thesaurus – obtain best textual match
 - Similarity threshold introduced to suppress low scoring matches. Levenshtein algorithm will always produce a match, even if it is a bad one!
 - Periods require an additional approach due to mixed formats (named periods, numeric ranges etc.)



Data Alignment Results – Monument Types

Data value	Highest scoring match	Score
ABBEY FOUNDATIONS	Foundation	74%
AXE FACOTRY	Axe Factory	90%
BOUNDARIES	BOUNDARY	77%
BOUNDARY	BOUNDARY	100%
BUIED SOIL HORIZON	BURIED SOIL HORIZON	97%
CAIRN	CAIRN	100%
CAIRN (POSSIBLE)	CAIRN	100%
CAIRNN	CAIRN	90%
CESS PITT	CESS PIT	94%
CHAMBERED TOM	CHAMBERED TOMB	96%
COMERCIAL	COMMERCIAL	94%
CROFT?	CROFT	90%
CUP-MARKED STONE	CUP MARKED STONE	93%
DICTH	DITCH	80%
ENCLSOURE	ENCLOSURE	88%
EXTRACTION PIT	EXTRACTIVE PIT	85%
EXTRACTIVE PIT	EXTRACTIVE PIT	100%

Data value	Highest scoring match	Score
FEATURE – COBBLED SURFACE	Cobbled Surface	75%
GULLEY	GULLY	90%
GULLIES	GULLY	66%
HILL FORT	HILLFORT	94%
HILLFORT	HILLFORT	100%
IINEAR SYSTEM	LINEAR SYSTEM	92%
MEDIEVAL CASTLE / FORTIFIED MANOR RUINS	FORTIFIED MANOR HOUSE	60%
PARIS CHURCH	PARISH CHURCH	96%
PASSAGE GRACE	PASSAGE GRAVE	92%
PORTAL DOLMEN (RE-ERECTED)	PORTAL DOLMEN	100%
POSTHOLE	POST HOLE	88%
PRIORY? WALL	Priory Wall	95%
RED HILL (POSSIBLE)	RED HILL	100%
ROMAN STRUCTURE POSSIBLY A VILLA	TRAINING STRUCTURE	52%
SOIL FILLED PIT	RIFLE PIT	66%
ST GUTHLACS BENEDICTINE PRIORY	Benedictine Priory	75%
STONE ALIGMENT	STONE ALIGNMENT	96%
TRACKWAY (COBBLED)	TRACKWAY	100%
WORCESTER-BIRMINGHAM CANAL	ORNAMENTAL CANAL	52%



Data Alignment Results (Objects and Periods)

Data value	Highest scoring match	Score
BRICK	PICK	66%
FE NAILS	NAIL	66%
FLINT SCRAPPER	SCRAPER (TOOL)	66%
INDUSTRIAL RSSIDUE	INDUSTRIAL BY PRODUCT	71%
LOOM WEIGHT	LOOMWEIGHT	95%
POTTEY	POTTERY	92%
SAMIEN SHERD	RIM SHERD	66%
UNIDENTIFIED OBJECT	UNIDENTIFIED OBJECT	100%

Data value	Highest scoring match	Score
NEOLOTHIC	NEOLITHIC	88%
NEOTLITHIC	NEOLITHIC	94%
POST-MEDIEVAL	POST MEDIEVAL	92%
IRON AGE	IRON AGE	87%



Data alignment results - categorised

- **Correct matches** – may not be 100% textual match
 - “AXE FACOTRY” → [AXE FACTORY](#)
 - “CAIRNN” → [CAIRN](#)
 - “PASSAGE GRACE” → [PASSAGE GRAVE](#)
 - “STONE ALIGMENT” → [STONE ALIGNMENT](#)
- **Unsure matches** – illustrate the need for expert oversight of results
 - “ARCHITECTURAL FEATURE” → [ARCHITECTURAL FRAGMENT](#)
 - “AXIAL-STONE CIRCLE” → [SMALL STONE CIRCLE](#)
 - “RADIAL CAIRN” → [TRI RADIAL CAIRN](#)
- **Incorrect matches** – may be reduced by raising the score threshold
 - “CLAY STRUCTURE” → [COAL GAS STRUCTURE](#)
 - “CONCENTRATION CAMP” → [CONSTRUCTION CAMP](#)
 - “RAIN MAKING SITE” → [PAINTBALLING SITE](#)
- **Non matches** – score exceeding threshold was not achieved
 - “ARCHAEOLOGY”
 - “CLAVA CAIRN COMPLEX”
 - “DOMKYRKAN”
 - “WEDGE TOMB”



Alignment: identifying periods

Data value	Identified Start year	Identified End year
250-400	250	400
500 BC	-500	-500
600-300 BC	-600	-300
AD 375-8	375	378
AD400-600	400	600
C2-C3	101	300
C6	501	600
EARLY 3RD CENTURY	201	240
EARLY FOURTH CENTURY BC	-400	-360
LATE 3RD CENTURY	260	300
LATE FOURTH CENTURY BC	-340	-301
MID 4TH CENTURY BC	-370	-330
MID THIRD CENTURY	230	270

- Achieved by matching predefined textual patterns (plus a bit of processing - e.g. "AD 375-8")
- AD:
 - Centuries starting at year 1 and finishing at year 100
 - "EARLY" = 1 → 40
 - "MID" = 30 → 70
 - "LATE" = 60 → 100
- BC:
 - Centuries starting at year -100 and finishing at year -1
 - "EARLY" = -100 → -60
 - "MID" = -70 → -30
 - "LATE" = -40 → -1
- There is no year zero...
- Once delimiting years are identified, can be aligned with known periods framework



Alignment: Thesaurus to thesaurus

Concept	Best Match	Score
GALVANIZING WORKS	GALVANIZING WORKSHOP	85%
PENSTOCKS	PENSTOCK	88%
FLAX KILN	FLARE KILN	80%
CUP AND RING MARKED ROCK	CUP AND RING MARKED STONE	84%
GUNCOTTON STORE	GUNCOTTON STOVE	93%
GOOD STATION	GOODS STATION	92%
STAITH	STAITHE	85%
TEXTILE PRINT WORKS	TEXTILE PRINTING WORKS	86%
GRAVE	GRAVE	100%
CIST	CIST	100%
ENCLOSED CREMATION CEMETERY	ENCLOSED CREMATION CEMETERY	100%
HOFFMAN KILN	HOFFMANN KILN	92%
ROAD BLOCK	ROADBLOCK	90%
ANTI AIRCRAFT DEFENCES	ANTI AIRCRAFT DEFENCE SITE	84%
TAKEAWAY	TAKE-AWAY	88%
SETTLING POND	RETTING POND	84%
SUSPENSION FOOTBRIDGE	SUSPENSION BRIDGE	80%
SESSION HOUSE	SESSIONS HOUSE	92%
ALUMINA WORKS	ALUMINIUM WORKS	80%
SHIP BREAKING YARD	SHIP BREAKERS YARD	83%

RCAHMS monuments to EH monuments

Concept	Best Match	Score
CANDLEHOLDER	CANDLE HOLDER	92%
MANUFACTURING AND PROCESSING	MANUFACTURE AND PROCESSING	89%
CRUSIE	CRUSE	83%
INORGANIC MATERIAL	ORGANIC MATERIAL	88%
PERSONAL ADORNMENT	PERSONAL ORNAMENT	83%
BALANCE	BALANCE	100%

RCAHMS objects to FISH objects

Concept	Best Match	Score
MOTOR GUN BOAT	MOTOR GUNBOAT	92%
HOUSEBOAT	HOUSE BOAT	90%
CONTAINER SHIP	CONTAINER SHIP	100%
LIBERTY SHIP	LIBERTY SHIP	100%
COLLIER	COLLIER	100%
DUMB HOPPER BARGE	(no match above threshold)	

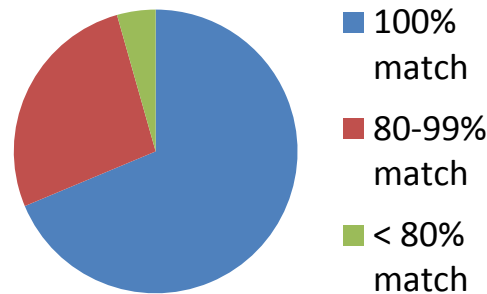
RCAHMS maritime to EH maritime



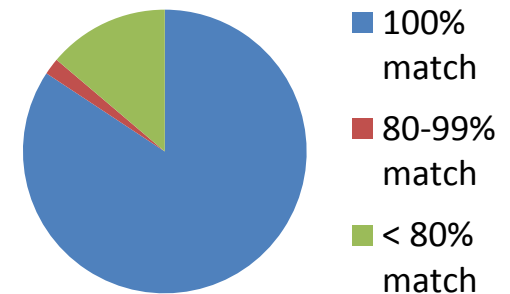
Alignment: inter-thesaurus concept linking

- Given the origins of the thesauri, 100% matches on preferred terms are concepts that likely could be linked (note: still requires caution, recall *TENEMENT*)
- 80-99% matches need a bit more expert overview
- < 80% matches are definitely unreliable

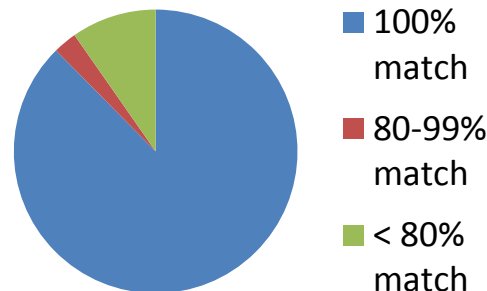
RCAHMS monuments
to EH monuments



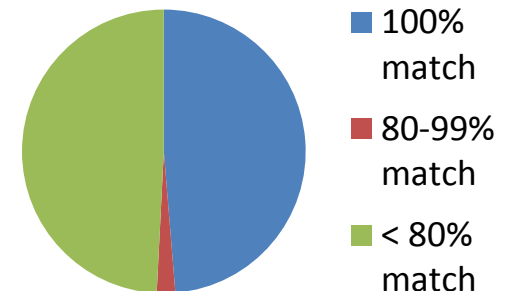
RCAHMS
monuments to EH
monuments



RCAHMS objects to
FISH objects



RCAHMS maritime to
EH maritime





Summary

- Controlled vocabularies online
 - Linked Open Data (SKOS)
 - Downloadable data files
- Vocabulary data available via web services
 - Concept search, suggestion, validation
- Using the vocabulary data
 - Data alignment exercises
 - Thesaurus alignment exercises
 - Browser-based 'widget' user interface controls for embedding into web applications



Next steps...

- Identifying potential applications and uses across the domain
- Incorporating other vocabularies, linking between vocabularies, combining vocabularies?
- Creating useful links to other LOD resources
- More widgets – ideas??



SENESCHAL

<http://www.heritagedata.org/>



Arts & Humanities
Research Council



ENGLISH HERITAGE



Royal
Commission on the
Ancient and
Historical
Monuments of
Scotland

100
CBHC
RCAHMW
1908-2008



COMISIWN BRENHINOL
HENEBION CYMRU
ROYAL COMMISSION
ON THE ANCIENT AND HISTORICAL MONUMENTS OF WALES

ads

ARCHAEOLOGY
DATA SERVICE



Contact information

Ceri Binding, Douglas Tudhope
Hypermedia Research Unit
Faculty of Computing, Engineering and Science
University of South Wales
Pontypridd CF37 1DL

ceri.binding@southwales.ac.uk

douglas.tudhope@southwales.ac.uk

<http://hypermedia.research.southwales.ac.uk/kos/>

<http://www.heritagedata.org/>

http://intarch.ac.uk/journal/issue30/tudhope_index.html

STAR Internet Archaeology paper (open access)

<http://data.archaeologydataservice.ac.uk>

STELLAR linked data