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, RCAHMS/W, 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).”

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-hole</td>
<td>IRON AGE</td>
</tr>
<tr>
<td>Posthole</td>
<td>Iron age</td>
</tr>
<tr>
<td>POST HOLE</td>
<td>Iron age?</td>
</tr>
<tr>
<td>POSTHOLE</td>
<td>EARLY IRON AGE</td>
</tr>
<tr>
<td>POST HOLE (POSSIBLE)</td>
<td>250 BC</td>
</tr>
<tr>
<td>POSTHOLES</td>
<td>C 500-200 B.C.</td>
</tr>
</tbody>
</table>
Semi-controlled vocabularies

<table>
<thead>
<tr>
<th>Deposit Colour</th>
<th>Deposit Texture</th>
<th>Deposit Compaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Reddy) Brown</td>
<td>Dark orange/brown</td>
<td>Firm</td>
</tr>
<tr>
<td>Reddy) brown</td>
<td>Dark orange/brown</td>
<td>Firm</td>
</tr>
<tr>
<td>Brown</td>
<td>Dark orange/brown</td>
<td>Firm</td>
</tr>
<tr>
<td>Brown red</td>
<td>Dark orange/brown</td>
<td>Firm</td>
</tr>
<tr>
<td>Brown/red</td>
<td>Dark orange/brown</td>
<td>Firm</td>
</tr>
<tr>
<td>Dark brown</td>
<td>Orange/brown, very light</td>
<td>Firm</td>
</tr>
<tr>
<td>Dark brown/orange</td>
<td>Orange/brown, very light</td>
<td>Firm</td>
</tr>
<tr>
<td>Dark orange brown</td>
<td>Orange/brown, very light</td>
<td>Firm</td>
</tr>
<tr>
<td>Dark orange brown with</td>
<td>Orange/brown, very light</td>
<td>Firm</td>
</tr>
<tr>
<td>Dark orange brown</td>
<td>Orange/brown, very light</td>
<td>Firm</td>
</tr>
<tr>
<td>Dark orange loam</td>
<td>Orange/brown, very light</td>
<td>Firm</td>
</tr>
<tr>
<td>Dark orange</td>
<td>Yellow brown</td>
<td>Firm</td>
</tr>
<tr>
<td>Orangy brown</td>
<td>Yellow/orange brown</td>
<td>Firm</td>
</tr>
</tbody>
</table>

…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

(Search on “tenement” - Archaeology Data Service)

(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

- Native vocabularies
- Additional metadata

- STELLAR application (SKOS templates)
  - (upload)
  - SKOS RDF vocabularies

- Linked Data REST API
- Web Services REST API
  - SPARQL endpoint

- web ‘widget’ controls & applications
  - SENESCHAL data store
  - (upload)
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
Multilingual labels & scope notes attached to concepts

Possible to search in one language, retrieve in another
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, .rdf, .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, .rdf, .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:
Thesaurus searching and browsing
Web services and user interface widgets
Web Services

- URI service calls returning JSON data
  - [http://www.heritagedata.org/blog/services/](http://www.heritagedata.org/blog/services/)

- All services available as Open Source
  - Creative Commons Attribution (CC-BY) license
  - [https://github.com/cbinding/SENESCHAL](https://github.com/cbinding/SENESCHAL)

- Service calls
  - /getSchemes
  - /getTopConceptsForScheme
  - /getConceptsForScheme
  - /getConceptRelations
  - /getConceptLabels
  - /getConceptLabelMatch
    - General term search facilities
  - /getConceptExists
    - For term validation

```json
[
  {
    "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."
  },
  {
    "uri": "http://purl.org/heritagedata/schemes/ch_tmc",
    "label": "MARITIME CRAFT TYPE (EH)",
    "label lang": "en",
    "description": "A thesaurus of craft types which survive as wrecks in English Heritage\ufffd's 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."
  }]
```
User interface controls - widgets

- Javascript widgets. Function in any modern browser, on PC, Mac, smartphone, tablet, console etc.
  - [http://www.heritagedata.org/blog/widgets/](http://www.heritagedata.org/blog/widgets/)

- All widgets are available Open Source
  - Creative Commons Attribution (CC-BY) license
  - [https://github.com/cbinding/SENESCHAL](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
- Composite control

(scheme list)

(scheme details)

(top concepts)

(composite control)
Concept widgets

- Concept details
- Concept relations
- 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
Early adoption (continued)

- Archaeology Data Service - use of SENESCHAL web services and concept URIs in CMS, exploratory use of widgets
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
Typical alignment issues encountered

- Simple spelling errors
  - POSTHLOLE”, “CESS PITT”, “FURRRROWS”, FLINT SCRAPPER”

- 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

<table>
<thead>
<tr>
<th>Data value</th>
<th>Highest scoring match</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBEY FOUNDATIONS</td>
<td>Foundation</td>
<td>74%</td>
</tr>
<tr>
<td>AXE FACTORY</td>
<td>Axe Factory</td>
<td>90%</td>
</tr>
<tr>
<td>BOUNDARIES</td>
<td>BOUNDARY</td>
<td>77%</td>
</tr>
<tr>
<td>BOUNDARY</td>
<td>BOUNDARY</td>
<td>100%</td>
</tr>
<tr>
<td>BURIED SOIL HORIZON</td>
<td>BURIED SOIL HORIZON</td>
<td>97%</td>
</tr>
<tr>
<td>CAIRN</td>
<td>CAIRN</td>
<td>100%</td>
</tr>
<tr>
<td>CAIRN (POSSIBLE)</td>
<td>CAIRN</td>
<td>100%</td>
</tr>
<tr>
<td>CAINN</td>
<td>CAIRN</td>
<td>90%</td>
</tr>
<tr>
<td>CESS PIT</td>
<td>CESS PIT</td>
<td>94%</td>
</tr>
<tr>
<td>CHAMBERED TOM</td>
<td>CHAMBERED TOMB</td>
<td>96%</td>
</tr>
<tr>
<td>COMERCIAL</td>
<td>COMMERCIAL</td>
<td>94%</td>
</tr>
<tr>
<td>CROFT?</td>
<td>CROFT</td>
<td>90%</td>
</tr>
<tr>
<td>CUP-MARKED STONE</td>
<td>CUP MARKED STONE</td>
<td>93%</td>
</tr>
<tr>
<td>DITCH</td>
<td>DITCH</td>
<td>80%</td>
</tr>
<tr>
<td>ENCLOSEURE</td>
<td>ENCLOSURE</td>
<td>88%</td>
</tr>
<tr>
<td>EXTRACTION PIT</td>
<td>EXTRACTIVE PIT</td>
<td>85%</td>
</tr>
<tr>
<td>EXTRACTIVE PIT</td>
<td>EXTRACTIVE PIT</td>
<td>100%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Data value</th>
<th>Highest scoring match</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRICK</td>
<td>PICK</td>
<td>66%</td>
</tr>
<tr>
<td>FE NAILS</td>
<td>NAIL</td>
<td>66%</td>
</tr>
<tr>
<td>FLINT SCRapper</td>
<td>SCRAPER (TOOL)</td>
<td>66%</td>
</tr>
<tr>
<td>INDUSTRIAL RSSIDUE</td>
<td>INDUSTRIAL BY PRODUCT</td>
<td>71%</td>
</tr>
<tr>
<td>LOOM WEIGHT</td>
<td>LOOMWEIGHT</td>
<td>95%</td>
</tr>
<tr>
<td>POTTEY</td>
<td>POTTERY</td>
<td>92%</td>
</tr>
<tr>
<td>SAMIEN SHERD</td>
<td>RIM SHERD</td>
<td>66%</td>
</tr>
<tr>
<td>UNIDENTIFIED OBJECT</td>
<td>UNIDENTIFIED OBJECT</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data value</th>
<th>Highest scoring match</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEOLOTHIC</td>
<td>NEOLITHIC</td>
<td>88%</td>
</tr>
<tr>
<td>NEOLITHIC</td>
<td>NEOLITHIC</td>
<td>94%</td>
</tr>
<tr>
<td>POST-MEDIEVAL</td>
<td>POST MEDIEVAL</td>
<td>92%</td>
</tr>
<tr>
<td>IRON AGE</td>
<td>IRON AGE</td>
<td>87%</td>
</tr>
</tbody>
</table>
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

- 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

<table>
<thead>
<tr>
<th>Data value</th>
<th>Identified Start year</th>
<th>Identified End year</th>
</tr>
</thead>
<tbody>
<tr>
<td>250-400</td>
<td>250</td>
<td>400</td>
</tr>
<tr>
<td>500 BC</td>
<td>-500</td>
<td>-500</td>
</tr>
<tr>
<td>600-300 BC</td>
<td>-600</td>
<td>-300</td>
</tr>
<tr>
<td>AD 375-8</td>
<td>375</td>
<td>378</td>
</tr>
<tr>
<td>AD400-600</td>
<td>400</td>
<td>600</td>
</tr>
<tr>
<td>C2-C3</td>
<td>101</td>
<td>300</td>
</tr>
<tr>
<td>C6</td>
<td>501</td>
<td>600</td>
</tr>
<tr>
<td>EARLY 3RD CENTURY</td>
<td>201</td>
<td>240</td>
</tr>
<tr>
<td>EARLY FOURTH CENTURY BC</td>
<td>-400</td>
<td>-360</td>
</tr>
<tr>
<td>LATE 3RD CENTURY</td>
<td>260</td>
<td>300</td>
</tr>
<tr>
<td>LATE FOURTH CENTURY BC</td>
<td>-340</td>
<td>-301</td>
</tr>
<tr>
<td>MID 4TH CENTURY BC</td>
<td>-370</td>
<td>-330</td>
</tr>
<tr>
<td>MID THIRD CENTURY</td>
<td>230</td>
<td>270</td>
</tr>
<tr>
<td>Concept</td>
<td>Best Match</td>
<td>Score</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>GALVANIZING WORKS</td>
<td>GALVANIZING WORKSHOP</td>
<td>85%</td>
</tr>
<tr>
<td>PENSTOCKS</td>
<td>PENSTOCK</td>
<td>88%</td>
</tr>
<tr>
<td>FLAX KILN</td>
<td>FLARE KILN</td>
<td>80%</td>
</tr>
<tr>
<td>CUP AND RING MARKED ROCK</td>
<td>CUP AND RING MARKED STONE</td>
<td>84%</td>
</tr>
<tr>
<td>GUNCOTTON STORE</td>
<td>GUNCOTTON STOVE</td>
<td>93%</td>
</tr>
<tr>
<td>GOOD STATION</td>
<td>GOODS STATION</td>
<td>92%</td>
</tr>
<tr>
<td>STAITH</td>
<td>STAITE</td>
<td>85%</td>
</tr>
<tr>
<td>TEXTILE PRINT WORKS</td>
<td>TEXTILE PRINTING WORKS</td>
<td>86%</td>
</tr>
<tr>
<td>GRAVE</td>
<td>GRAVE</td>
<td>100%</td>
</tr>
<tr>
<td>CIST</td>
<td>CIST</td>
<td>100%</td>
</tr>
<tr>
<td>ENCLOSED CREMATION</td>
<td>ENCLOSED CREMATION</td>
<td>100%</td>
</tr>
<tr>
<td>CEMETERY</td>
<td>CEMETERY</td>
<td>100%</td>
</tr>
<tr>
<td>HOFFMAN KILN</td>
<td>HOFFMANN KILN</td>
<td>92%</td>
</tr>
<tr>
<td>ROAD BLOCK</td>
<td>ROADBLOCK</td>
<td>90%</td>
</tr>
<tr>
<td>ANTI AIRCRAFT DEFENCES</td>
<td>ANTI AIRCRAFT DEFENCE SITE</td>
<td>84%</td>
</tr>
<tr>
<td>TAKEAWAY</td>
<td>TAKE-AWAY</td>
<td>88%</td>
</tr>
<tr>
<td>SETTLING POND</td>
<td>RETTING POND</td>
<td>84%</td>
</tr>
<tr>
<td>SUSPENSION FOOTBRIDGE</td>
<td>SUSPENSION BRIDGE</td>
<td>80%</td>
</tr>
<tr>
<td>SESSION HOUSE</td>
<td>SESSIONS HOUSE</td>
<td>92%</td>
</tr>
<tr>
<td>ALUMINA WORKS</td>
<td>ALUMINIUM WORKS</td>
<td>80%</td>
</tr>
<tr>
<td>SHIP BREAKING YARD</td>
<td>SHIP BREAKERS YARD</td>
<td>83%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
<th>Best Match</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANDLEHOLDER</td>
<td>CANDLE HOLDER</td>
<td>92%</td>
</tr>
<tr>
<td>MANUFACTURING AND PROCESSING</td>
<td>MANUFACTURE AND PROCESSING</td>
<td>89%</td>
</tr>
<tr>
<td>CRUSIE</td>
<td>CRUSE</td>
<td>83%</td>
</tr>
<tr>
<td>INORGANIC MATERIAL</td>
<td>ORGANIC MATERIAL</td>
<td>88%</td>
</tr>
<tr>
<td>PERSONAL ADORNMENT</td>
<td>PERSONAL ORNAMENT</td>
<td>83%</td>
</tr>
<tr>
<td>BALANCE</td>
<td>BALANCE</td>
<td>100%</td>
</tr>
</tbody>
</table>

**RCAHMS objects to FISH objects**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Best Match</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOTOR GUN BOAT</td>
<td>MOTOR GUNBOAT</td>
<td>92%</td>
</tr>
<tr>
<td>HOUSEBOAT</td>
<td>HOUSE BOAT</td>
<td>90%</td>
</tr>
<tr>
<td>CONTAINER SHIP</td>
<td>CONTAINER SHIP</td>
<td>100%</td>
</tr>
<tr>
<td>LIBERTY SHIP</td>
<td>LIBERTY SHIP</td>
<td>100%</td>
</tr>
<tr>
<td>COLLIER</td>
<td>COLLIER</td>
<td>100%</td>
</tr>
<tr>
<td>DUMB HOPPER BARGE</td>
<td>(no match above threshold)</td>
<td></td>
</tr>
</tbody>
</table>

**RCAHMS maritime to EH maritime**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Best Match</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOTOR GUN BOAT</td>
<td>MOTOR GUNBOAT</td>
<td>92%</td>
</tr>
<tr>
<td>HOUSEBOAT</td>
<td>HOUSE BOAT</td>
<td>90%</td>
</tr>
<tr>
<td>CONTAINER SHIP</td>
<td>CONTAINER SHIP</td>
<td>100%</td>
</tr>
<tr>
<td>LIBERTY SHIP</td>
<td>LIBERTY SHIP</td>
<td>100%</td>
</tr>
<tr>
<td>COLLIER</td>
<td>COLLIER</td>
<td>100%</td>
</tr>
</tbody>
</table>

**RCAHMS monuments to EH monuments**
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
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??
http://www.heritagedata.org/
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