

STELLAR

# Semantic Technologies and Linked Data

Ceri Binding

Hypermedia Research Unit, University of Glamorgan, Wales, UK

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

[cbinding@glam.ac.uk](mailto:cbinding@glam.ac.uk)

University of  
Glamorgan

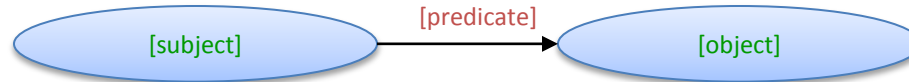
CARDIFF • PONTYPRIDD • WALES • UK

# Introduction

- **STELLAR** – Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources
- 12 month AHRC funded project, March 2010-February 2011
- Tools to assist in the production of **Linked Data** from archaeological datasets
- Linked Data implemented using Resource Description Framework (RDF)
- Made available via Archaeology Data Service

# Resource Description Framework (RDF) – quick primer

RDF triples consist of **subject** → **predicate** → **object**:



The **object** of one triple may be the **subject** of another, forming a **chain**:

Digital Past 2013 → took place at → Shire Hall → is in → Monmouth

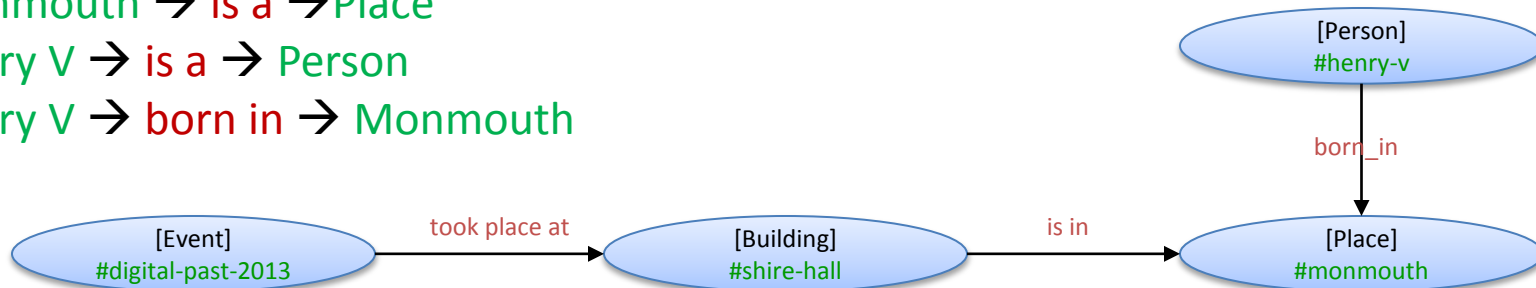
Can then append additional information, forming a **graph**:

Shire Hall → is a → Building

Monmouth → is a → Place

Henry V → is a → Person

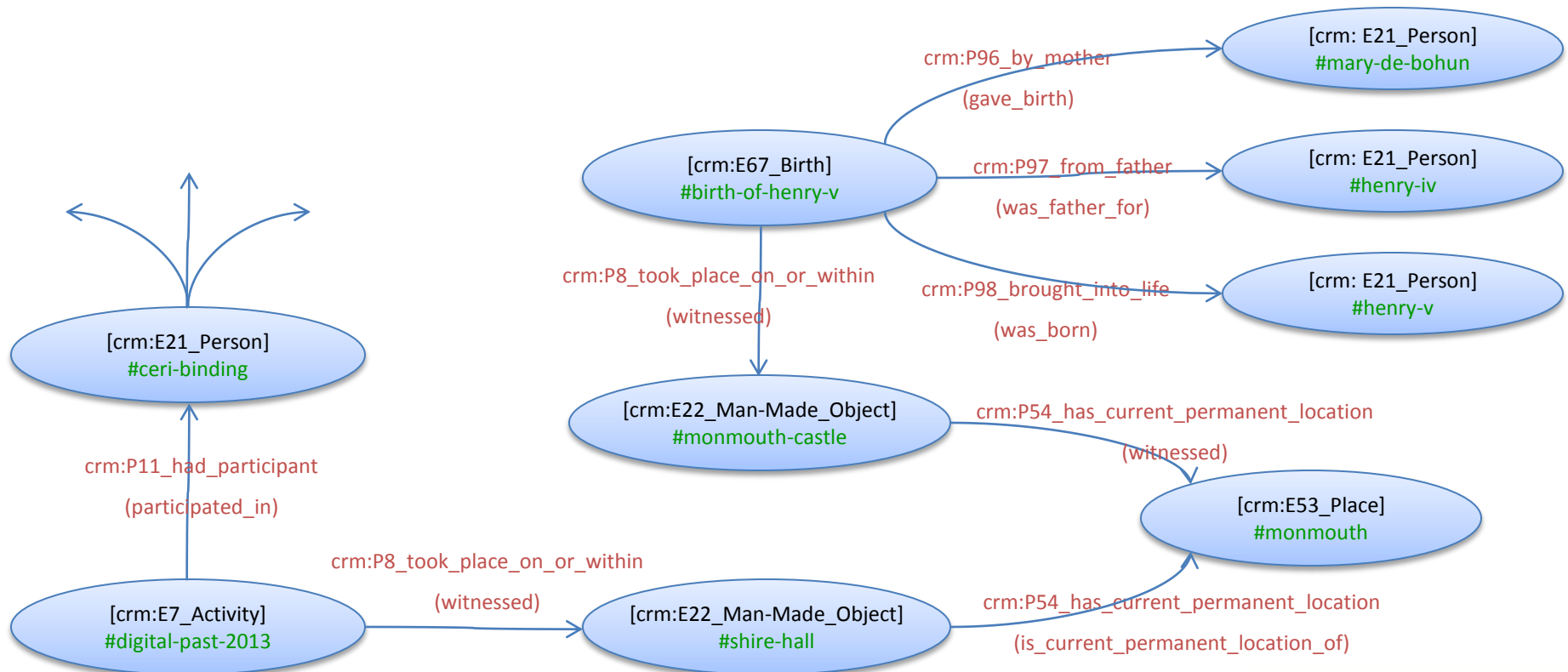
Henry V → born in → Monmouth



# RDF – using an ontological model (CIDOC CRM)

BUT just a local idiosyncratic data model (Event, Place, born\_in etc.)



- How do we append information from elsewhere in a scalable manner?
- How do we agree on meaning and structure for wider interoperability?
- Conform to a common ontological model for structuring the known facts...
  - e.g. CIDOC Conceptual Reference Model (CRM) <http://www.cidoc-crm.org/>



# Linked Data

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

# Archaeology Data Service (ADS) Linked Data

ARCHAEOLOGY  
DATA SERVICE  
[http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE0007\\_1010](http://data.archaeologydataservice.ac.uk/10.5284/1000365/EHE0007_1010)

## 1010

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



### 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](#)



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

# STELLAR outputs

- Linked Data is a simple concept, implementation may be complicated by...
  - Conceptual modelling issues
    - Identifiers, co-reference, entities, relationships, inheritance, transitivity, versioning, controlled vocabularies
  - Initial data formats
    - Data cleansing, data mapping, interpretation, conversion
  - RDF/XML syntax
    - Brackets, tags, attributes, character encoding, namespaces, URIs
- STELLAR produced tools to assist in managing complexity & maintaining consistency
- ‘Templates’ containing placeholders, to generate predefined data patterns

# STELLAR applications


**STELLAR**

An application for converting delimited (CSV) format data to valid RDF data conforming to a chosen 'template'.

Delimited Data File  test\_crmeH\_contexts\_strat\_lower\_id.csv

Template name

Namespace prefix

Validator 

**Results**

[ria33qvl.rdf](#)  
(the download link will remain available on the server for 30 minutes, after which it will be automatically deleted)

Statistics

- 5 unique subject URIs
- 8 unique object URIs
- 1 unique literals using 1 languages
- 3 unique class URIs

```
<http://purl.org/crmeH/EHE0007_Context> [1]
<http://purl.org/crmeH/EHE0061_ContextUID> [1]
<http://purl.org/crmeH/EHE1001_ContextEvent> [2]
14 statements, using 9 predicate URIs:
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> [4]
<http://www.w3.org/2000/01/rdf-schema#label> [1]
<http://www.w3.org/1999/02/22-rdf-syntax-ns#value> [1]
<http://purl.org/NET/cmw-owl#P87_is_identified_by> [1]
<http://purl.org/NET/cmw-owl#P871_identifies> [1]
<http://purl.org/NET/cmw-owl#P7_took_place_at> [2]
<http://purl.org/NET/cmw-owl#P71_witnessed> [2]
<http://purl.org/NET/cmw-owl#P120_occurs_before> [1]
<http://purl.org/NET/cmw-owl#P120i_occurs_after> [1]
```

## STELLAR.Web

```
STELLAR.Console v1.0
<type HELP for instructions>

STELLAR.Console>help
For information on a particular command type HELP command
DBNAMES      List databases in a directory
DBTABLES     List tables in a database
DIRCOLUMNS  List columns in a database table
DBROWCOUNT  Count rows in a database table
TAB2DB       Import tab delimited file to database table
CSU2DB       Import comma delimited file to database table
SQL2CSU      Run SQL, export result to CSU file
SQL2TAB      Run SQL, export result to tab delimited file
CSU2RDF      Convert comma delimited file to RDF file (via template)
TAB2RDF      Convert tab delimited file to RDF file (via template)
SQL2RDF      Run SQL, export result to RDF file (via template)
TEMPLATES    Display list of possible templates to use
CSUSTATS     Display statistics for comma delimited file
RDFSTATS     Display statistics for RDF file
EXIT         Exit the application

STELLAR.Console>
```

## STELLAR.Console

Stellar.Win v1.0 C:\Projects\IntrasisLinkedData\data\phase\_testing\stellar

Conversion description: Test conversion from Intrasis PHASE data to RDF

Delimited data file: C:\Projects\IntrasisLinkedData\data\Phase\_te... Choose... Edit...

Text delimited by:  Tab  Semicolon  Comma  Space  Other

First row contains field names?

Delimited data preview (up to first 50 records)

phase_id	phase_name	phase_note	prelim
10001489	Natural Geology		
10001490	Peinadical leach res		

Template: C:\Program Files\STELLAR\Templates\CRME... Choose... Edit...

Options file: C:\Projects\IntrasisLinkedData\data\OPTIONS Choose... Edit...

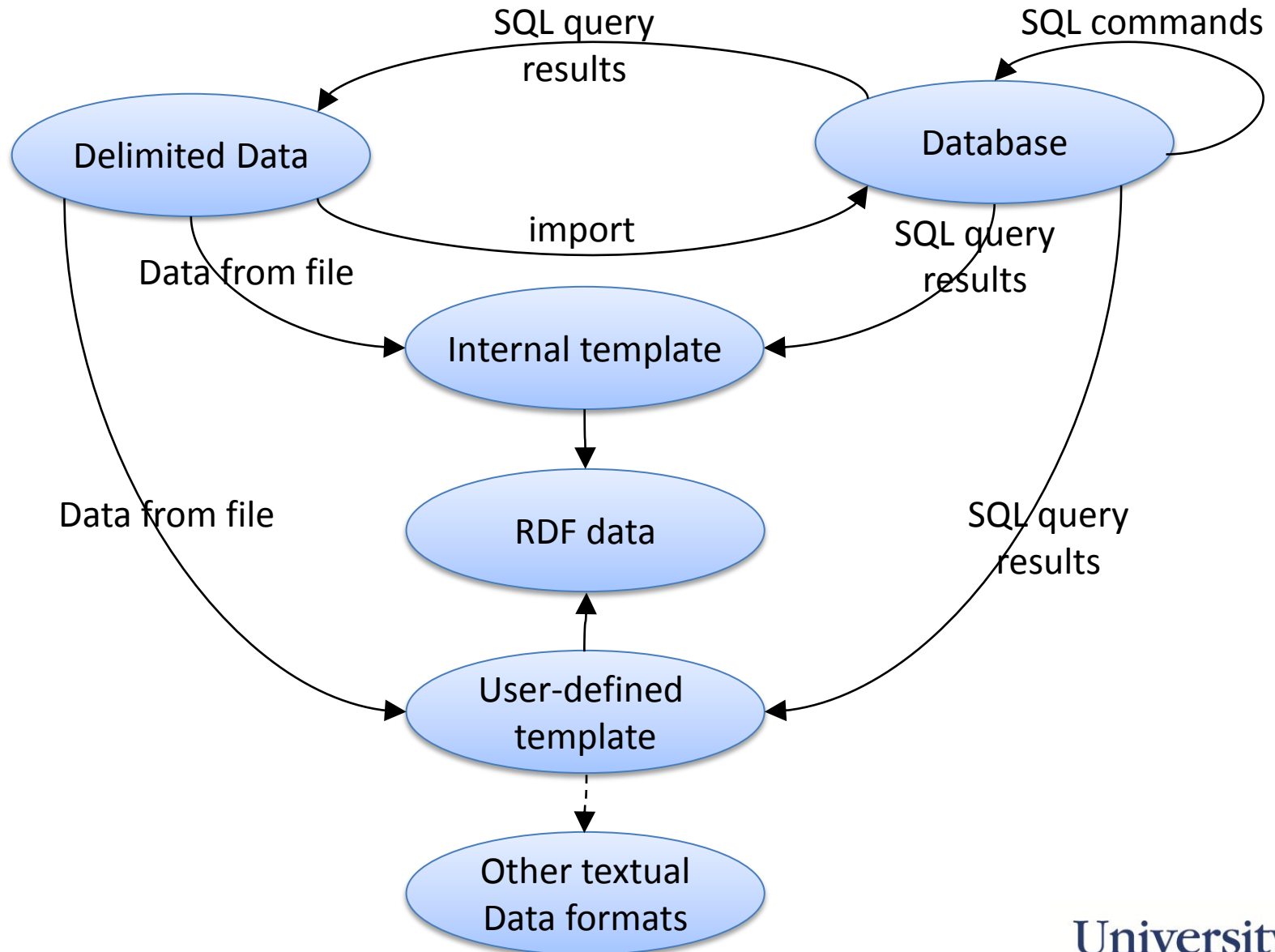
Equivalent STELLAR Console command: DELIMIT2TG /data "C:\Projects\IntrasisLinkedData\data\Phase\_testing.csv" /delete "" /sg "C:\Program Files\STELLAR\Templates\CRMEH\_PHASES.stg"

Converted output: <?xml version="1.0" encoding="UTF-8"?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:skoscl="http://www.w3.org/2008/05/skos-cl#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:cmw="http://stanger-cmw.org/010001/" xmlns:cmwH="http://purl.org/cmwH#" /> <cmwH:EHE001\_GroupEventRecord rdf:about="http://purl.org/crmeH/EHE001\_PHASE\_10001489#" <dc:label xml:lang="en">10001489</dc:label /></cmwH:EHE001\_GroupEventRecord> <cmwH:EHE1003\_GroupEventRecord rdf:about="http://purl.org/crmeH/EHE1003\_PHASE\_10001489#" <dc:label xml:lang="en">10001489</dc:label /></cmwH:EHE1003\_GroupEventRecord> <cmwH:EHE0070\_GroupEventTimeSpanRecord rdf:about="http://purl.org/crmeH/EHE0070\_PHASE\_10001489#" <dc:label xml:lang="en">10001489</dc:label /></cmwH:EHE0070\_GroupEventTimeSpanRecord>

## STELLAR.Win



# STELLAR data conversions



# Using STELLAR templates to produce RDF

```
// STELLAR template to write RDF header
```

```
HEADER(options) ::= <<
```

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<rdf:RDF
```

```
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:crm="http://www.cidoc-crm.org/rdfs/cidoc-crm#">
```

```
  >>
```

```
  >>
```

```
// Template writes RDF entities and properties based on each data row;
```

```
// $placeholder.value$ is replaced with the named field data at runtime
```

```
RECORD(options, data) ::= <<
```

```
  <crm:E53 rdf:about="#E53_$(data.id)$">
```

```
    <crm:P87F rdf:resource="#E44_$(data.id)$"/>
```

```
  </crm:E53>
```

```
  <crm:E44 rdf:about="#E44_$(data.id)$">
```

```
    <rdfs:label xml:lang="it">$(data.name)$</rdfs:label>
```

```
    <crm:P87B rdf:resource="#E53_$(data.id)$"/>
```

```
  </rdf:Description>
```

```
>>
```

```
// STELLAR template to write RDF footer – closure of header elements
```

```
FOOTER(options) ::= "</rdf:RDF>"
```

- Templates are just text files. May be copied, edited, exchanged, disseminated.

- XML/RDF syntax and namespace details are handled within the template.

- User input is simple tabular delimited textual data with named fields, e.g.:

id, name

1, Bergamo

2, Milano Centrale

3, Bologna Centrale

4, Prato Centrale

- Predefined patterns of entities, properties and inverse properties are created by the template, data populates placeholders at runtime.

- Output is consistent and repeatable.

# Templates hide complexity

context_id	strat_lower_id
123	456

## Template input

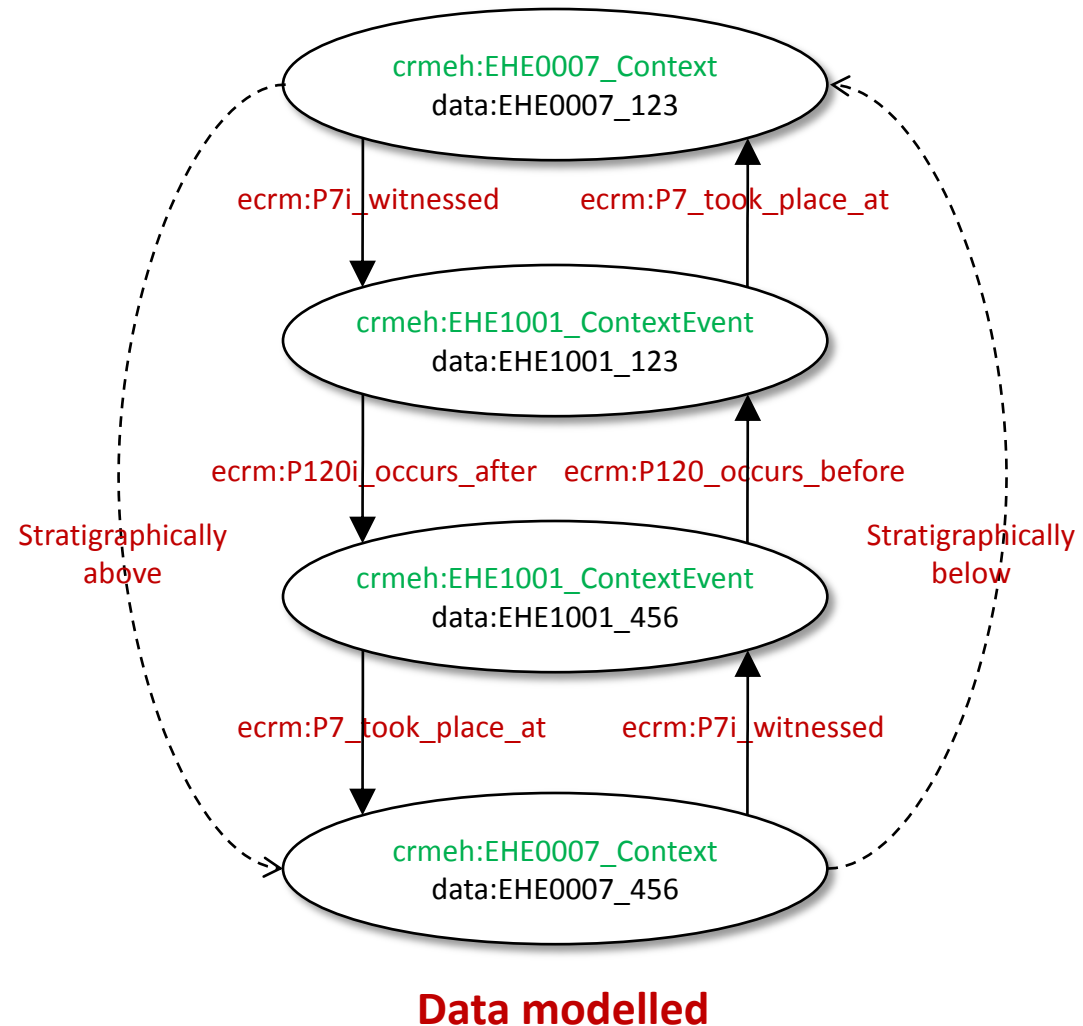
```

<crneh:EHE0007_Context rdf:about="http://stellar/EHE0007_123"/>
<crneh:EHE0007_Context rdf:about="http://stellar/EHE0007_456"/>
<crneh:EHE1001_ContextEvent rdf:about="http://stellar/EHE1001_123"/>
<crneh:EHE1001_ContextEvent rdf:about="http://stellar/EHE1001_456"/>
<rdf:Description rdf:about="http://stellar/EHE1001_123">
  <ecrm:P7_took_place_at rdf:resource="http://stellar/EHE0007_123"/>
</rdf:Description>
<rdf:Description rdf:about="http://stellar/EHE0007_123">
  <ecrm:P7i_witnessed rdf:resource="http://stellar/EHE1001_123"/>
</rdf:Description>
<rdf:Description rdf:about="http://stellar/EHE1001_456">
  <ecrm:P7_took_place_at rdf:resource="http://stellar/EHE0007_456"/>
</rdf:Description>
<rdf:Description rdf:about="http://stellar/EHE0007_456">
  <ecrm:P7i_witnessed rdf:resource="http://stellar/EHE1001_456"/>
</rdf:Description>
<rdf:Description rdf:about="http://stellar/EHE1001_456">
  <ecrm:P120i_witnessed rdf:resource="http://stellar/EHE1001_123"/>
</rdf:Description>
<rdf:Description rdf:about="http://stellar/EHE1001_123">
  <ecrm:P120i_witnessed rdf:resource="http://stellar/EHE1001_456"/>
</rdf:Description>

```

## Template output

Intermediate entities,  
inverse properties and  
shortcuts all generated  
by the template

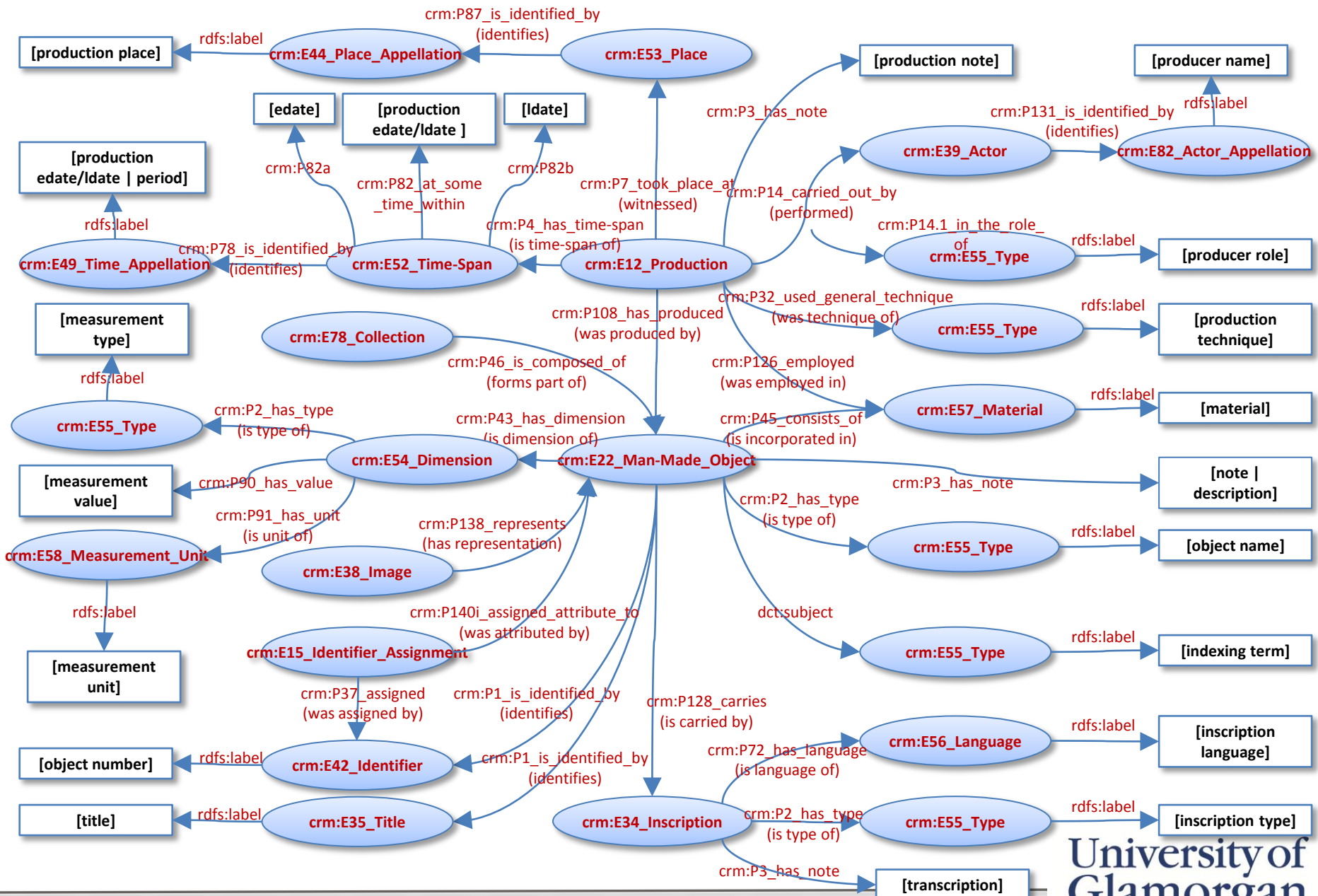


## Data modelled

# Pilot study – numismatic data

- Data originating from 4 different collections:
  - National Museum of Wales (NMW) – Roman, Civil War & Tudor numismatic collections
  - National Museum of Science & Industry (NMSI) – “Coins, Medals & Tokens” collection
- Mapping data to CIDOC CRM
- Extraction & conversion using STELLAR tools
- Querying & visualisation

# Data fields mapped to CIDOC CRM Entities / Properties



# Example of National Museum of Wales data in RDF, conforming to CIDOC CRM ontological model

```
<crm:E22_Man-Made_Object rdf:about="http://tmp/nmw/E22_1000">
  <rdfs:label xml:lang="en">81.79H/1.1</rdfs:label>
  <crm:P1_is_identified_by rdf:resource="http://tmp/nmw/E42_1000" />
  <crm:P140i_was_attributed_by rdf:resource="http://tmp/nmw/E15_1000" />
  <crm:P108i_was_produced_by rdf:resource="http://tmp/nmw/E12_1000" />
  <crm:P2_has_type rdf:resource="http://tmp/nmw/E55_denarius" />
  <crm:P45_consists_of rdf:resource="http://tmp/nmw/E57_silver" />
  <crm:P70i_is_documented_in rdf:resource="http://tmp/nmw/E31_crawford" />
  <crm:P70i_is_documented_in rdf:resource="http://tmp/nmw/E31_crawford_222%2f1" />
  <crm:P43_has_dimension rdf:resource="http://tmp/nmw/E54_1000_weight" />
  <crm:P46i_forms_part_of rdf:resource="http://tmp/nmw/E78_nmw+roman" />
  <crm:P128_carries rdf:resource="http://tmp/nmw/E34_1000_reverse" />
</crm:E22_Man-Made_Object>
<crm:E15_Identifier_Assignment rdf:about="http://tmp/nmw/E15_1000">
  <rdfs:label xml:lang="en">81.79H/1.1</rdfs:label>
  <crm:P140_assigned_attribute_to rdf:resource="http://tmp/nmw/E22_1000" />
  <crm:P37_assigned rdf:resource="http://tmp/nmw/E42_1000" />
</crm:E15_Identifier_Assignment>
<crm:E42_Identifier rdf:about="http://tmp/nmw/E42_1000">
  <rdfs:label xml:lang="en">81.79H/1.1</rdfs:label>
  <crm:P1i_identifies rdf:resource="http://tmp/nmw/E22_1000" />
  <crm:P37i_was_assigned_by rdf:resource="http://tmp/nmw/E15_1000" />
</crm:E42_Identifier>
<crm:E12_Production rdf:about="http://tmp/nmw/E12_1000">
  <rdfs:label xml:lang="en">81.79H/1.1</rdfs:label>
  <crm:P108_has_produced rdf:resource="http://tmp/nmw/E22_1000" />
  <crm:P126_employed rdf:resource="http://tmp/nmw/E57_silver" />
  <crm:P14_carried_out_by rdf:resource="http://tmp/nmw/E39_rome+mint" />
  <crm:P4_has_time-span rdf:resource="http://tmp/nmw/E52_-143%2f-143" />
</crm:E12_Production>
```

Data converted using STELLAR templates

# Querying CRM properties with SPARQL

## Example cross-searching multiple datasets

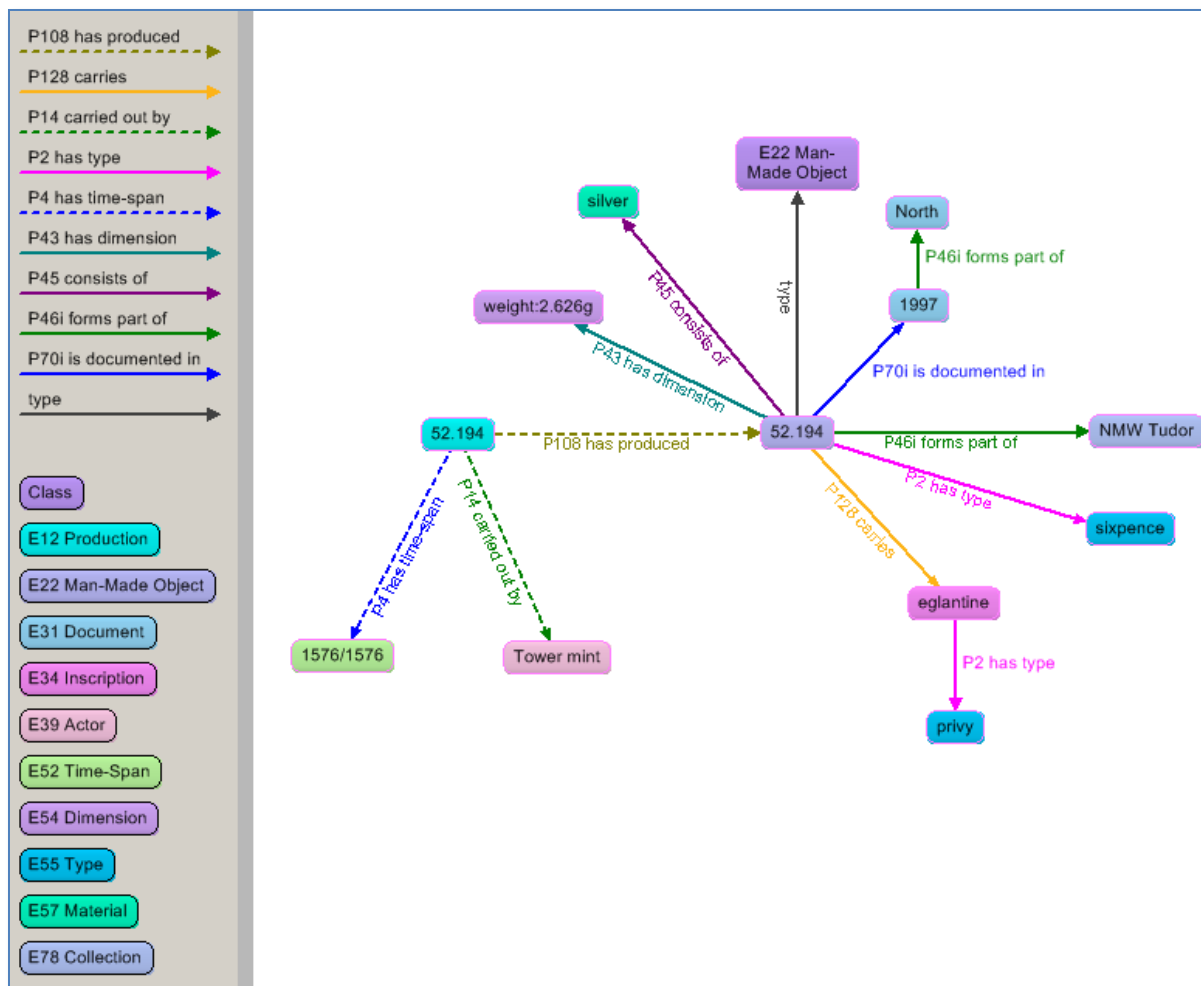
### Types of objects made of copper, produced between 1600 and 1800

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX crm: <http://erlangen-crm.org/101001/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT DISTINCT ?type ?collection WHERE {
  ?x crm:P2_has_type [rdfs:label ?type];
  crm:P46i_forms_part_of ?collection;
  crm:P45_consists_of [rdfs:label "copper"@en];
  crm:P108i_was_produced_by [crm:P4_has_time-span
    [crm:P82a ?min_date ; crm:P82b ?max_date ] ] .
  FILTER(xsd:dateTime(?min_date) >= xsd:dateTime("1600-01-01T00:00:00") &&
    xsd:dateTime(?max_date) < xsd:dateTime("1800-01-01T00:00:00")) .
}
```

type	collection
twopence	NMW Civil War
token	SCM – Coins, Medals & Tokens
turner/bodle	NMW Civil War
stirling	NMW Civil War
coin	SCM – Coins, Medals & Tokens
penny	NMW Tudor

# Visualising RDF – entities and properties



(Object 52.194 from the NMW Tudor numismatics collection)



# Summary

- Overall process is complex, needs tools to improve consistency and repeatability

Templates handle low level syntax and implement predefined patterns of data

- improving consistency
- reducing complexity
- if only we can agree on the data patterns to use!

# Next steps... The SENESCHAL Project

- **seneschal** n. *Historical* the steward or major-domo of a medieval great house
- **S**emantic **EN**richment **E**nabling **S**ustainability of ar**CHA**eological **L**inks
- 12 month AHRC funded project. March 2013 - February 2014
- English Heritage controlled vocabularies online as (SKOS) Linked Data
  - Monument Types Thesaurus
  - Object Types Thesaurus



# Interoperability

- *“The terminology of a subject is the key to interoperability”* (John F. Sowa)
- Interoperability requires more than just a common data model
- Data compatibility occurs on 2 levels – semantic and syntactic. Ontologies / data structures deal with the semantic but not necessarily the syntactic.
  - *“The CRM relies on existing syntactic interoperability and is concerned only with adding semantic interoperability”* (CIDOC CRM documentation)

# You say potato, I say tomato...

- Multiple datasets, multiple organisations
- Unification of data structures is possible, BUT...
- Lack of interoperability – incompatible terminology hinders cross search
- E.g. Get all the iron age post holes:

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

**Solution: data cleansing and controlled vocabularies?**

# Semi-controlled vocabularies

	Deposit Colour	Deposit Texture	Deposit Compaction
(Reddy) Brown	Dark orange/brown	Orangy brown, very	Firm
9Reddy) brown	Dark red brown	light brown on edges	Eriable
Brown			Plastic
Brown r			Sticky
Brown/			
Dark br			
Dark br			
Dark gre			
Dark ora			
Dark ora			
with dar			
patches			
Dark orange loam	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 do already have controlled vocabularies, however tension exists between being descriptive indexing vs. controlled indexing at point of data entry

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

For data retrieval: ***The worst of all worlds?***

# Typical interoperability issues encountered

- Simple spelling errors
  - POSTHOLE”, “CESS PITT”
- 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”

# Solutions - SENESCHAL



- Controlled vocabularies (again)
  - Commonly agreed concepts and terminology
  - Existing / new thesauri – community contributions?
- Openness and availability
  - Licensing, web services, downloads, data formats
- Alignment of existing data
  - Data cleansing tools
  - Alignment techniques
- Alignment of new data
  - Interactive data entry tools
  - Validation at point of data entry
  - Rather than trying to solve the vocabulary problem, prevent it from happening in the first place

STELLAR

# Semantic Technologies and Linked Data

Ceri Binding

Hypermedia Research Unit, University of Glamorgan, Wales, UK

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

[cbinding@glam.ac.uk](mailto:cbinding@glam.ac.uk)

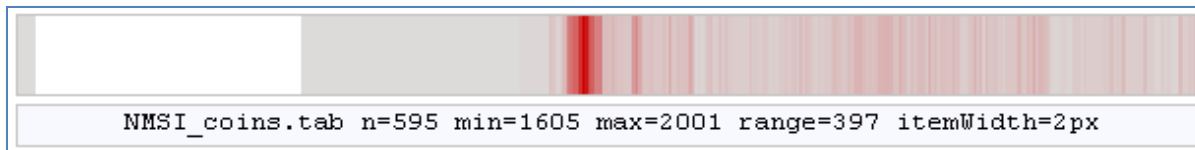
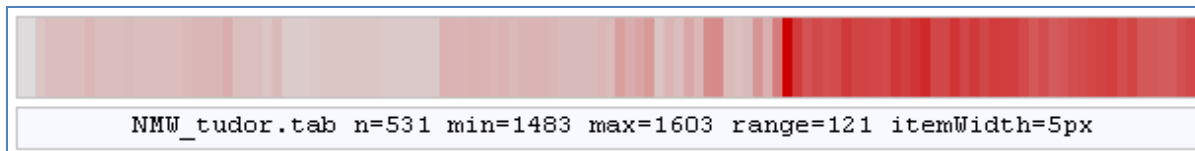
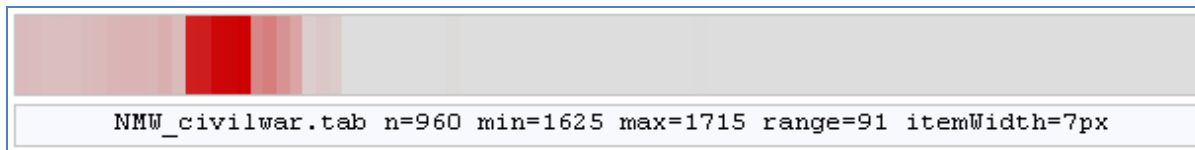
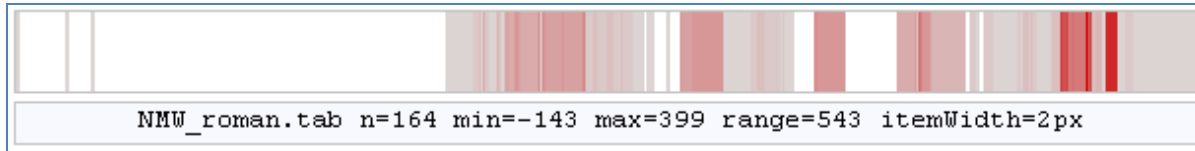
University of  
Glamorgan

CARDIFF • PONTYPRIDD • WALES • UK

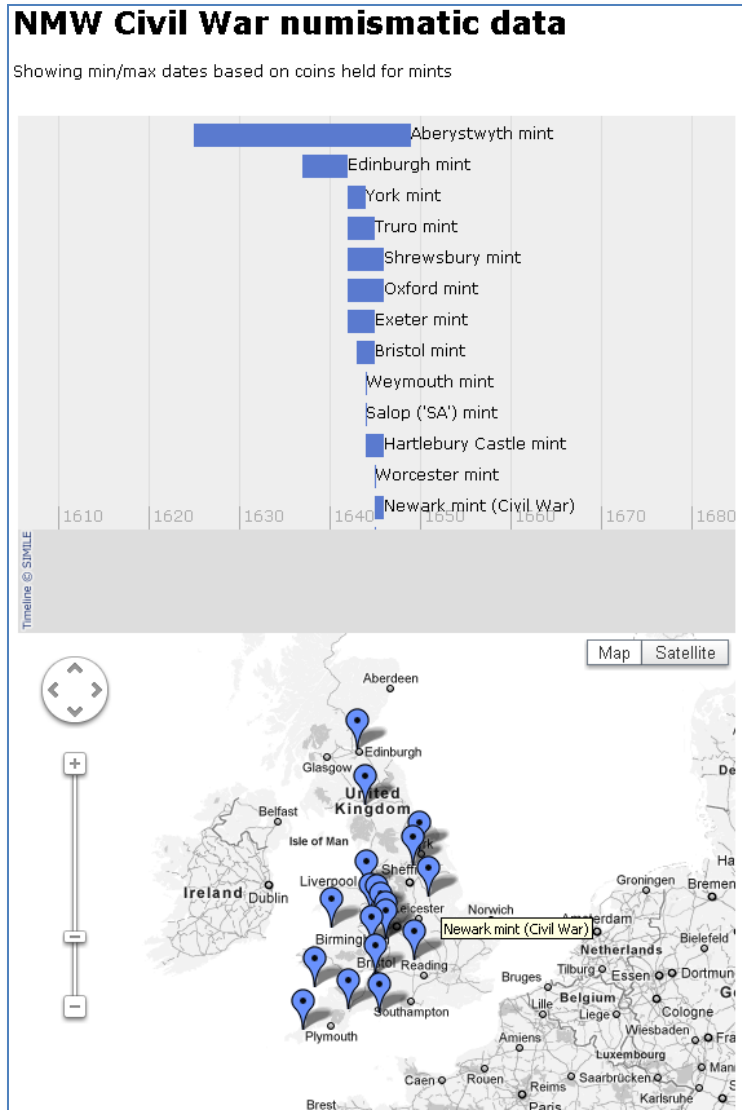




# Visualisation - data distribution



# Visualisation - TimeMap



- TimeMap - Interactive temporal / geographical display
- Combines Google Map and Simile Timeline
- Displaying apparent mint activity based on coins from NMW Civil War collection