Vocabulary mapping tool for archaeology in ARIADNEplus

Holly Wright, Archaeology Data Service
Ceri Binding & Douglas Tudhope
University of South Wales, Trefforest

tgn:7029392 World
tgn:1000003 Europe
tgn:7008591 United Kingdom
tgn:7002443 Wales
tgn:7018963 Rhondda Cynon Taf
tgn:7441565 Trefforest
• **ARIADNE**
  – 24 partners, 13 countries, 9 languages, 27 subject vocabularies
  – 1.9 million data records aggregated/integrated
  – Subject vocabularies coordinated via mapping to Getty AAT – total 6416 mappings produced

• **ARIADNEplus**
  – 41 partners, 29 countries, 22 languages, ?? subject vocabularies
  – Data aggregation/integration work currently in progress
  – Reusing, revising and supplementing previous mappings
  – Adding vocabulary mappings from new data partners
  – Adding Wikidata mappings (multilingual entry vocabulary)
Why do we need vocabulary matching in ARIADNE?

- Source datasets not necessarily produced with aggregation, consolidation, cross-search and reuse in mind
- I say “potato”, you say “pomme de terre”, she says “maris piper” – multiple barriers to cross-searching subject metadata: language, punctuation, spelling, homonyms, synonyms, level of specificity
- Text-based search is limited by any/all of these
- Need to establish mutually agreed meaning...
Multilingual subject index terms

How to express that we all mean the same thing?

Mapping everything to everything is not ideal...

Ideally we want to include any/all of these variants in a single query

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782
Map local terms to a central concept

The **words** may be different, but the **concept** is (more or less) the same...

Now we can include any/all of these variants in a single query.
Map local concepts to a central spine
Vocabulary Matching Tool

https://vmt.ariadne.d4science.org/vmt/

- For matching local subject terms / concepts to Getty AAT concepts
- Search & browse Getty AAT
- No auto match: examine scope and context of source / target concepts
Type of match between concepts

**Exact Match**

- **Source:** "Cups"
- **Target:** "Cups"

But: don’t rely on label matches; consider full context – meaning and scope of concepts.

**Close Match**

- **Source:** "Cups"
- **Target:** "Cups"

Where scope or context of concepts suggests slight conceptual differences.

[Note: skos:narrowMatch also exists]

**Some/all” rule for generic hierarchical relationships:**

- **Target:** "Cups"
- **Source:** "Coffee cups"

Some cups are coffee cups; all coffee cups are cups.

- **Target:** "Cups"
- **Related Match**
  - **Source:** "Cups"
  - **Source:** "Saucers"

Some other association exists between the concepts. Where possible prefer one of the other match types though.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782.
Why? Using vocabulary mappings

• Search term = "CEMETERY"
  – may retrieve some results. Expand with plural “cemeteries”. May retrieve a few more...

• What do the mappings give us?

• What other local vocabulary terms are mapped to [aat:300266755]?
  – "BARROW CEMETERY", "CHOLERA BURIAL GROUND", "FRIENDS BURIAL GROUND", "INHUMATION CEMETERY", "JEWISH CEMETERY", "MUSLIM CEMETERY", "NONCONFORMIST CEMETERY", "PLAQUE CEMETERY", "ROMAN CATHOLIC CEMETERY", "WALLED CEMETERY"

• We now have an expanded search, and have uncovered potential links between records indexed using any of these terms.
  – However they are all in one language...
Why? Using vocabulary mappings

- Multilingual terms associated with concept aat:300266755:
  -Preferred labels: "cemeteries"@en, "campos santos"@es, "campi santi"@it, "cimetiere"@fr, "begraafplaatsen"@nl, "Friedhof"@de
  -Alternate labels: "cemetery"@en, "campos santos (cemeteries)"@en, "campo santo (cemetery)"@en, "campo santo"@es, "campo santo"@it, "cimetière"@fr, "coemeterium (cemeteries)"@la, "camposanto (cemetery)"@en, "camposanto"@it, "begraafplaats"@nl, "Friedhöfe"@de

- Can we utilize AAT (poly)hierarchical structure? (Yes!)
  -AAT concepts narrower (more specific) than cemeteries:
    -catacombs, columbaria (cemeteries), graveyards, lawn cemeteries, memorial parks, necropolises, Reihengräberfelder, churchyards, cineraria (cemeteries), military cemeteries (veteran cemeteries), national cemeteries, pet cemeteries, potter's fields, war cemeteries
    -Plus each of these concepts has multilingual preferred / alternate terms - we now have a semantically expanded multilingual search
Why? Using vocabulary mappings

• And finally...
  – Wikidata contains mappings to AAT concepts
  – `wikidata:Q39614` is already directly mapped to `aat:300266755` ("cemeteries") and has many more multilingual labels:
    • Cemetery, graveyard, burial ground, cemeteries, churchyard, cimetière, champ de repos, boulevard des allongés, champ du repos, Friedhof, Totenacker, Begräbnisplatz, Gottesacker, Kirchhof, Leichenhof, Begraafplaas, Asie, Fosal, Fosar, Zimenterio, Corralón, Fusal, Sagrero, Fosal d'os moros, Cimenterio, Fonsal, 墳場, cmentarz, cemitério, покопалишце, гробъе etc.

• One mapping brings in many alternative terms/concepts to improve multilingual query experience and to expand potential results.

• Use of semantic links can improve recall without necessarily sacrificing precision
RDF serialisations of mappings

Example mappings expressed in RDF-XML serialization format:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://snd.gu.se/sv/catalogue/keyword/bengomma">
    <skos:prefLabel xml:lang="sv">bengomma</skos:prefLabel>
    <skos:closeMatch rdf:resource="http://vocab.getty.edu/aat/300265420"/>
  </skos:Concept>

  <skos:Concept rdf:about="http://snd.gu.se/sv/catalogue/keyword/bergshistorisk-lamning-ovrig">
    <skos:prefLabel xml:lang="sv">bergshistorisk lämning övrig</skos:prefLabel>
    <skos:closeMatch rdf:resource="http://vocab.getty.edu/aat/300006423"/>
  </skos:Concept>

</rdf:RDF>
```

Example mappings expressed in Turtle RDF serialization format:

```turtle
@prefix data: <http://snd.gu.se/sv/catalogue/keyword/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix aat: <http://vocab.getty.edu/aat/> .

data:bengomma a skos:Concept;
  skos:prefLabel "bengömma"@sv ;
  skos:closeMatch aat:300265420 . # remains

data:bergshistorisk-lamning-ovrig a skos:Concept ;
  skos:prefLabel "bergshistorisk lämning övrig"@sv ;
  skos:closeMatch aat:300006423 . # mine structures

data:bildristning a skos:Concept;
  skos:prefLabel "bildristning"@sv ;
  skos:closeMatch aat:300000131 . # rock carvings

data:björngrav a skos:Concept;
  skos:prefLabel "björngrav"@sv ;
  skos:closeMatch aat:300005907 . # graves
```
ARIADNE type-ahead suggestions

Welcome
Explore the digital resources and learning and teaching.

Browse the Catalog

 Getty AAT subject term type-ahead suggestions during search

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782
Records supplemented with AAT
Selected references and links

References


Links

- ARIADNEplus project: http://www.ariadne-infrastructure.eu/
- ARIADNE portal: https://ariadne-infrastructure.eu/portal/
- USW Hypermedia Research Group: https://hypermedia.research.southwales.ac.uk/

Contact

- ceri.binding@southwales.ac.uk ORCID: 0000-0002-6376-9613
- douglas.tudhope@southwales.ac.uk