

Vocabulary Matching

Ceri Binding & Douglas Tudhope

University of South Wales, Trefforest

[tgn:7029392](#) World

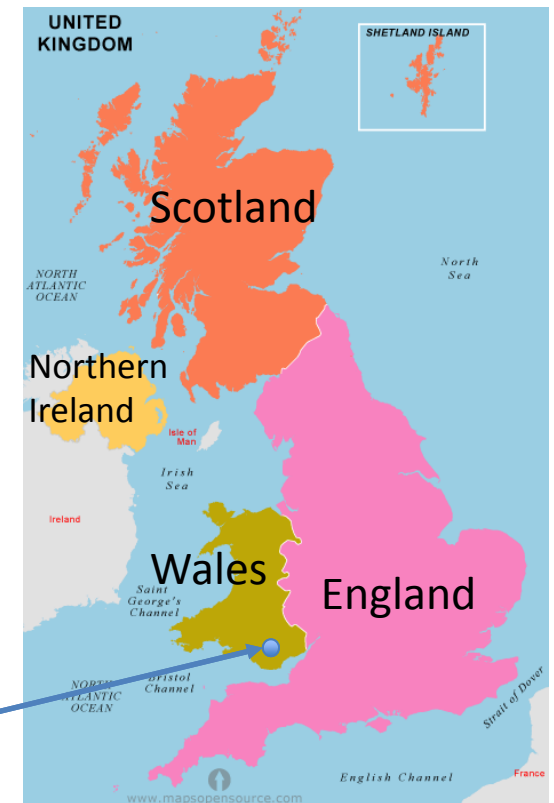
[tgn:1000003](#) Europe

[tgn:7008591](#) United Kingdom

[tgn:7002443](#) Wales

[tgn:7018963](#) Rhondda Cynon Taf

[tgn:7441565](#) Trefforest



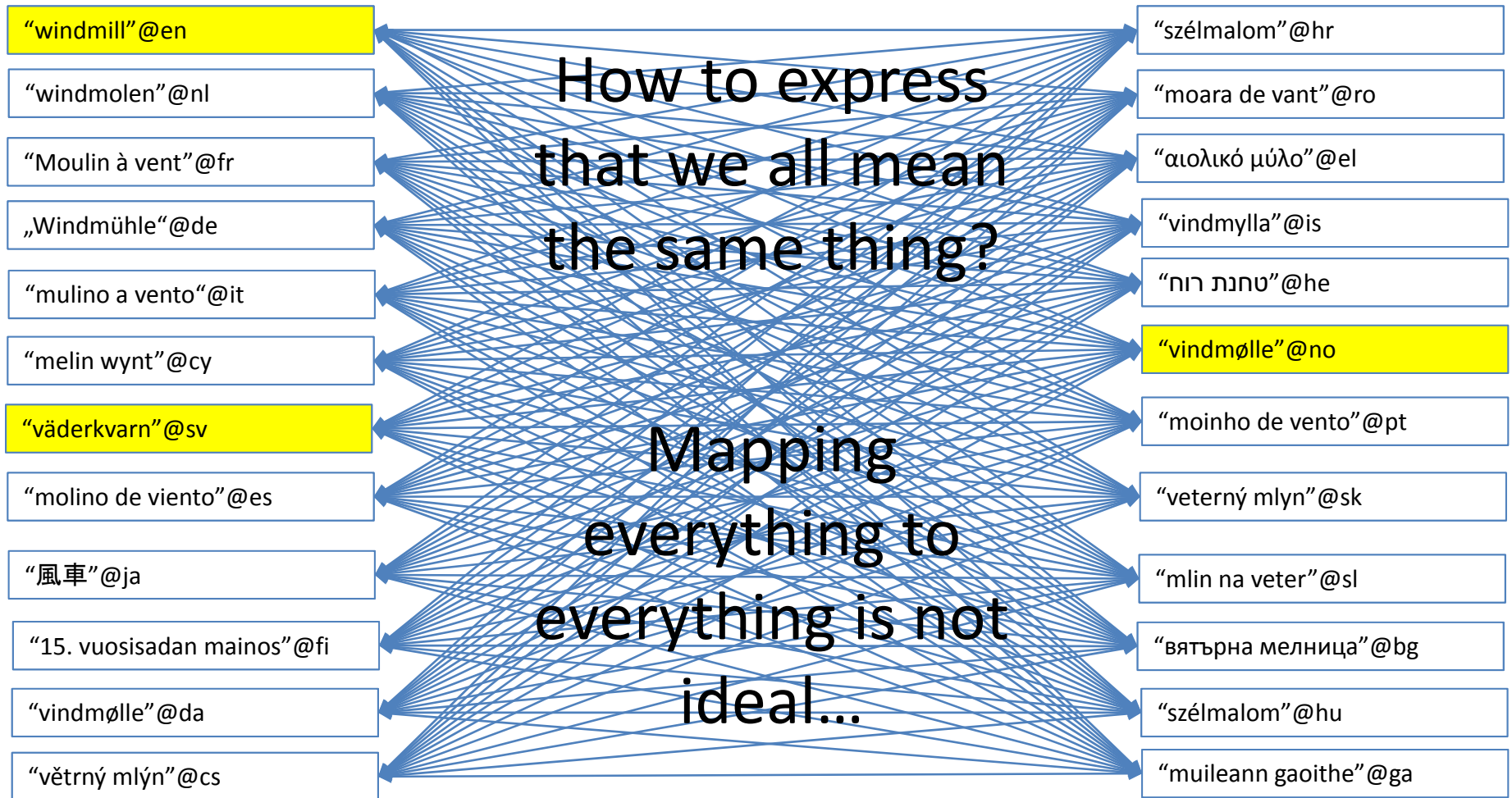
ARIADNE projects

- **ARIADNE I:**
 - 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)
 - Opportunities to feed back terms & mappings to Getty?
-

Vocabulary matching - why?

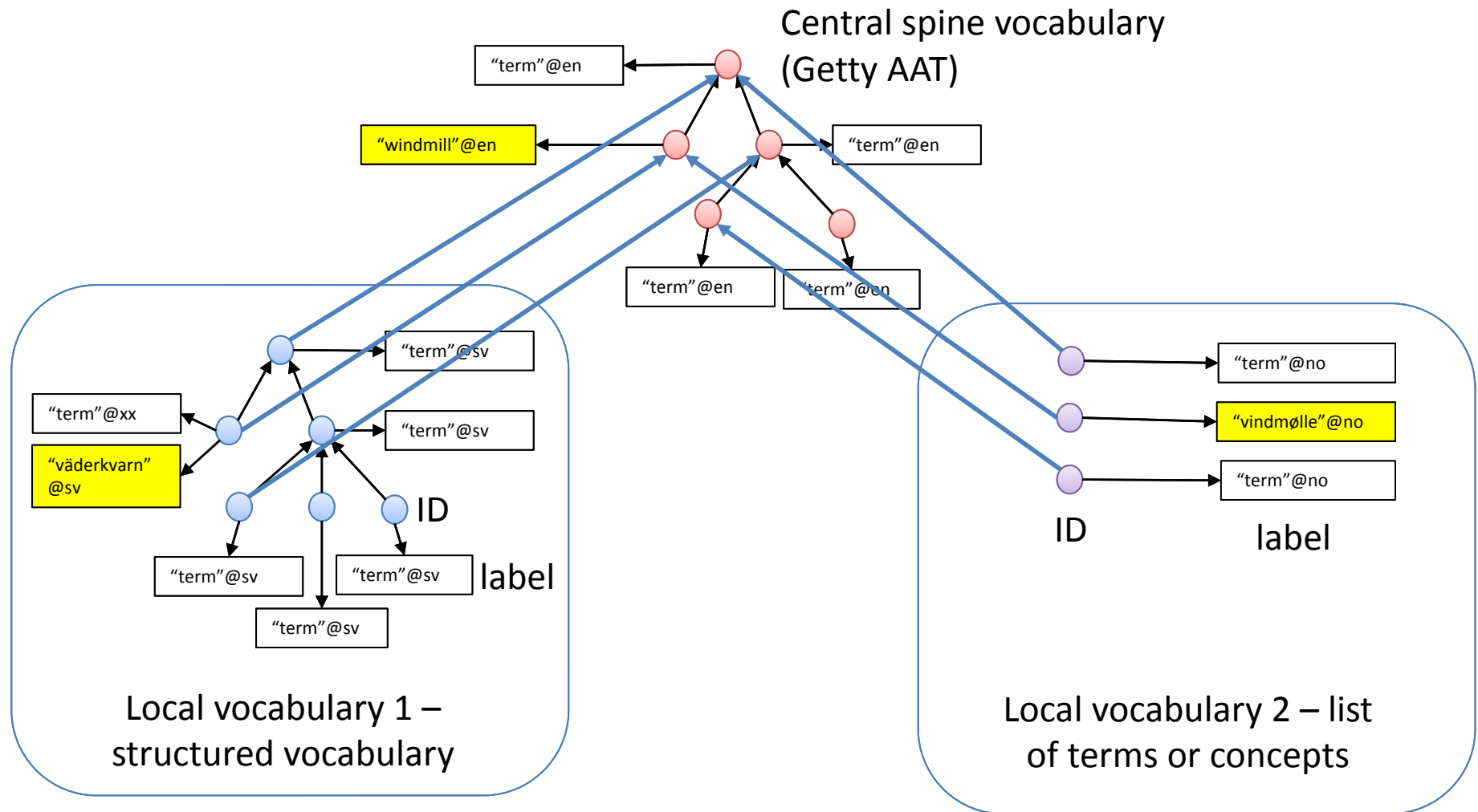
- Source datasets not necessarily produced with aggregation, consolidation, cross-search and reuse in mind
 - I say “*potato*”, you say “*pomme de terre*”, she says “*maris piper*” – multiple barriers to cross-searching subject metadata: language, punctuation, spelling, homonyms, synonyms, level of specificity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

Multilingual subject index terms



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

Map local concepts to a central spine



Vocabulary Matching Tool

- 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

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

Vocabulary Matching Tool English

| Source Concept | | Match Type | Target Concept | Suggest | Delete Row |
|-----------------------------------------------------------------------------------------------|-----------------------------|-------------|--------------------------|---------|------------|
| Identifier | Label | | Filter column... | | |
| http://purl.org/heritagedata/schemes... | Abbey Church | Close Match | abbey churches | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | ABBEY | Exact Match | abbeys (monasteries) | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | AGRICULTURAL BUILDING | Exact Match | agricultural buildings | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | AGRICULTURAL DWELLING | Broad Match | agricultural buildings | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | AGRICULTURAL HALL | Broad Match | agricultural buildings | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | FARM BUILDING | Close Match | agricultural buildings | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | FIELD SYSTEM | Broad Match | agricultural land | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | FIELD SYSTEM | Broad Match | agricultural land | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | LAND USE SITE | Broad Match | agricultural land | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | LYNCHET | Broad Match | agricultural land | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | CURVILINEAR ENCLOSURE | Broad Match | agricultural settlements | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | DITCHED ENCLOSURE | Broad Match | agricultural settlements | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | DOUBLE DITCHED ENCLOSURE | Broad Match | agricultural settlements | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | ENCLOSED SETTLEMENT | Broad Match | agricultural settlements | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | ENCLOSURE | Broad Match | agricultural settlements | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | AGRICULTURE AND SUBSISTENCE | Broad Match | agriculture | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | AIR RAID SHELTER | Exact Match | air raid shelters | Q | ⊖ |
| http://purl.org/heritagedata/schemes... | AIRCRAFT | Close Match | aircraft | Q | ⊖ |

390 rows

[IMPORT JSON](#)
[EXPORT JSON](#)
[EXPORT CSV](#)
[+ ADD NEW ROW](#)
[CLEAR ROWS](#)
[SHOW HELP](#)



Created by University of South Wales

ARIADNEplus is a Horizon 2020 project funded by the European Commission (Grant Agreement No 823914)

This application retrieves some information originating from Getty Art & Architecture Thesaurus (AAT)® which is made available under the ODC Attribution License. See <http://vocab.getty.edu/> for further details.

Type of match between concepts

Exact Match



BUT: don't rely on label matches;
consider full context – meaning and
scope of concepts

Close Match



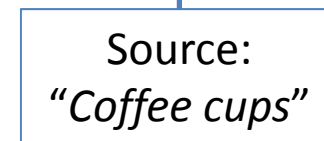
Where scope or context of concepts
suggests slight conceptual differences

[Note: skos:narrowMatch also exists]

"Some/all" rule for generic
hierarchical relationships:



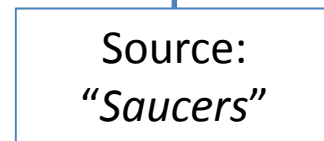
Broad Match



Some cups are
coffee cups; all
coffee cups are
cups



Related Match



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

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?
 - local:12345 "CEMETERY" skos:exactMatch [aat:300266755](#) "*cemeteries*"
 - 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*", "*PLAGUE 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, "cimetières"@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, "cœmeterium (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](https://www.wikidata.org/wiki/Q39614) is already directly mapped to [aat:300266755](https://www.getty.edu/research/conceptservices/aat/aat300266755) (“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, Cimiterio, Fonsal, 墳場, cmentarz, cemitério, pokopališče, гробље 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

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Example mappings expressed in RDF/XML serialization format (-->
<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">bengömma</skos:prefLabel>
    <skos:closeMatch rdf:resource="http://vocab.getty.edu/aat/300265420"/><!--remains-->
  </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"/><!--mine structures-->
  </skos:Concept>

  <skos:Concept rdf:about="http://snd.gu.se/sv/catalogue/keyword/bildristning">
    <skos:prefLabel xml:lang="sv">bildristning</skos:prefLabel>
    <skos:broadMatch rdf:resource="http://vocab.getty.edu/aat/300080131"/>
  </skos:Concept>

  <skos:Concept rdf:about="http://snd.gu.se/sv/catalogue/keyword/bjorngrav">
    <skos:prefLabel xml:lang="sv">björngrav</skos:prefLabel>
    <skos:broadMatch rdf:resource="http://vocab.getty.edu/aat/300005907"/>
  </skos:Concept>
</rdf:RDF>
```

```
# Mappings expressed in Turtle RDF serialization format
@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:broadMatch aat:300080131 . # rock carvings

data:bjorngrav a skos:Concept;
  skos:prefLabel "björngrav"@sv ;
  skos:broadMatch aat:300005907 . # graves
```

ARIADNE type-ahead suggestions

The screenshot displays the ARIADNE search interface. At the top, the ARIADNE logo is visible. Below it, a search bar contains the text 'axe|'. A dropdown menu shows the following suggestions:

- axes / axes / ax / axe ⓘ
- axhammers / axhammers / axehammers / axe-hammer / axes, masons' / masons' axes / axhammer/ ... ⓘ
- socketed axes / socketed axes / axes, socketed / socketed axe ⓘ
- flanged axes / flanged axes / flanged axe / axes, flanged ⓘ
- cleavers / cleavers / meat-ax / cleavers, butchers' / cleavers, meat / meat cleavers / meat axes/ ... ⓘ
- ax heads / ax heads / axe heads / ax head / heads, ax / axe-heads / ax-heads ⓘ
- axes / axes / ax ⓘ

Below the search bar, there are three main sections:

- Welcome**: Explore the digital resources and learning and teaching.
- Browse the Catalog**: This section contains three interactive panels:
 - Where**: A map of Europe with country labels.
 - When**: A bar chart showing the frequency of terms over time, with a peak around 1500.
 - What**: A word cloud of terms including 'barns', 'houses', 'churches (buildings)', 'pits (earthworks)', 'farms', 'cist graves', 'hearth', and 'cairns'.

Getty AAT subject term type-ahead suggestions during search

Records supplemented with AAT

Start a new search...
Catalog
Services
About

Lopération se situe au sein dun secteur archéologiquement sensible.

Les tranchées réalisées, ont permis de reconnaître des indices d'occupations humaines se rapportant à deux périodes chronologiques : le Néolithique et la période Médiévale à Moderne.

Les paléoenvironnements locaux ont été restitués grâce à l'observation des séquences sédimentaires (nature et géométrie des dépôts).

Pour le Néolithique, quatre fosses (F6 à F9) montrent une occupation du secteur au cours de la première moitié du 4e mill. av. J.-C. (Néolithique moyen 2) qui peut être corrélée à celle, plus dense, fouillée en 1992 légèrement plus au Sud-Ouest (site de PEER 2).

Pour la période Médiévale à Moderne, des creusements successifs (F1 à F5) du fossé méridional bordant l'axe de communication Riom/Varennes-sur-Morge/Thuret, matérialisé actuellement par la D 211, ont été observés dans le sondage S1. Il est vraisemblable que cet axe soit relativement ancien U+003B le mobilier du fossé F5 tend à montrer que ce dernier est en place au moins depuis la période médiévale.

[routes](#)
[modern European ceramics styles](#)
[Medieval \(European\)](#)
[industry \(object groupings\)](#)
[fauna](#)
[teeth \(animal components\)](#)
[quartz \(mineral\)](#)
[geomorphology](#)
[Pottery Neolithic](#)
[pits \(earthworks\)](#)

[Temps Modernes](#)
[Moyen Age](#)
[Néolithique moyen](#)
[Néolithique](#)
[Néolithique moyen](#)
[Néolithique](#)

📍 rue Ludwig von Beethoven, RIOM (score geo: 65)

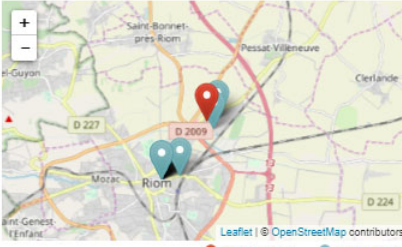
Metadata

| | |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ARIADNE ID | 25108064 |
| Original ID | 29407 |
| Language | French |
| Audience | Scientific |
| Resource type | Event/Intervention resource |
| Subject | routes modern European ceramics styles Medieval (European) industry (object groupings) fauna teeth (animal components) quartz (mineral) geomorphology Pottery Neolithic pits (earthworks) |
| Original Subject | voie céramique moderne céramique médiévale industrie lithique faune dent quartz géomorphologie céramique néolithique fosse |
| Dating | 1492 – 1815 , Temps Modernes 500 – 1500 , Moyen Age -5300 – -4501 , Néolithique moyen |

Resource is part of

Dolia

Geographically similar



Thematically similar

- Valdivienne (86), La Pétusière à Saint-Martin-la-Rivière - Phase 2 : rapport de diagnostic
- Villeneuve-la-Garenne (Hauts-de-Seine), 40-42 quai Alfred Sisley : rapport de diagnostic
- Paris Xie, 127-131a rue du Chemin Vert : rapport de diagnostic
- Tonnay-Charente, 7 rue Pierre Berné : rapport de diagnostic
- Saint-Pierre-du-Perray (Essonne), Plaine des Clés de Saint-Pierre, zones 2, 3 et 4 : lieu-dit La Garenne : occupations préhistoriques et occupations diachroniques agricoles et domestiques du Bronze final et du Hallstatt final : rapport de fouille
- Triel-sur-Seine [et] Vermouillet (Yvelines), Nouveau pont de Triel : liaisons RD1-RD55 et RD1-RD154 : rapport de diagnostic
- Mittelschaeffolsheim, Bas-Rhin, Beim Berstetter Weg, construction de la LGV Est Européenne, tronçon H-Site 10-5 : indices d'une occupation du Néolithique récent : rapport de fouille

Subjects derived from local vocabulary mappings to AAT

Selected references and links

References

- Binding, C, Tudhope, D & Vlachidis, A 2018, 'A study of semantic integration across archaeological data and reports in different languages' Journal of Information Science, vol 45, no. 3, pp. 364-386. [doi:10.1177/0165551518789874](https://doi.org/10.1177/0165551518789874)
- Binding, C & Tudhope, D 2016, 'Improving interoperability using vocabulary linked data' International Journal on Digital Libraries, vol 17, no. 1, pp. 5-21. [doi:10.1007/s00799-015-0166-y](https://doi.org/10.1007/s00799-015-0166-y)

Links

- ARIADNEplus project: <http://www.ariadne-infrastructure.eu/>
- ARIADNE portal: <https://ariadne-infrastructure.eu/portal/>
- Vocabulary Matching Tool (VMT): <https://vmt.ariadne.d4science.org/vmt/>
- 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
-