Int J Digit Libr (2016) 17:5–21
DOI 10.1007/s00799-015-0166-y

Improving interoperability using vocabulary linked data

Ceri Binding[1] · **Douglas Tudhope**[1]

Received: 12 January 2015 / Revised: 22 June 2015 / Accepted: 11 August 2015 / Published online: 27 August 2015

Abstract The concept of Linked Data has been an emerging theme within the computing and digital heritage areas in recent years. The growth and scale of Linked Data has underlined the need for greater commonality in concept referencing, to avoid local redefinition and duplication of reference resources. Achieving domain-wide agreement on common vocabularies would be an unreasonable expectation; however, datasets often already have local vocabulary resources defined, and so the prospects for large-scale interoperability can be substantially improved by creating alignment links from these local vocabularies out to common external reference resources. The ARIADNE project is undertaking large-scale integration of archaeology dataset metadata records, to create a cross-searchable research repository resource. Key to enabling this cross search will be the 'subject' metadata originating from multiple data providers, containing terms from multiple multilingual controlled vocabularies. This paper discusses various aspects of vocabulary mapping. Experience from the previous SENESCHAL project in the publication of controlled vocabularies as Linked Open Data is discussed, emphasizing the importance of unique URI identifiers for vocabulary concepts. There is a need to align legacy indexing data to the uniquely defined concepts and examples are discussed of SENESCHAL data alignment work. A case study for the ARIADNE project presents work on mapping between vocabularies, based on the Getty Art and Architecture Thesaurus as a central hub and employing an interactive vocabulary mapping tool developed for the project, which generates SKOS mapping relationships in JSON and other formats. The potential use of such vocabulary mappings to assist cross search over archaeological datasets from different countries is illustrated in a pilot experiment. The results demonstrate the enhanced opportunities for interoperability and cross searching that the approach offers.

Keywords Linked Data · Controlled vocabularies · Thesauri · Alignment · Mapping · Getty Art and Architecture Thesaurus · SKOS · LOD

✉ Ceri Binding
ceri.binding@southwales.ac.uk

Douglas Tudhope
douglas.tudhope@southwales.ac.uk

[1] Hypermedia Research Group, Faculty of Computing Engineering and Science, University of South Wales, Llantwit Road, Pontypridd CF37 1DL, United Kingdom

1 Introduction

1.1 Knowledge organization

Knowledge organization draws on techniques originating from the field of Library and Information Science, encompassing conceptual modelling, descriptive indexing, classification and search. The combination of these techniques with modern semantic technologies using flexible and extensible modelling approaches can enable the integration of previously disparate data, improve search facilities and produce platform neutral data resources and services for collaboration and dissemination. Traditionally, subject thesauri were often associated with one particular dataset. In today's world, an online thesaurus might be associated with a range of different datasets, many unknown to the creators of the thesaurus. In large-scale aggregation projects such as Europeana [1], multiple thesauri are involved in many different languages. Thus publication of thesauri and other vocabularies in standard representations and the mapping between different vocabularies is becoming a key component of interoperability. See

[2] for a review of registries of vocabularies and a discussion of their attributes and functionality. Zeng and Chan [3] provide an extensive review of vocabulary mapping. They identify various types of mapping—the ARIADNE project's aim in the case study reported here is to employ a *switching language* (using the AAT as a hub vocabulary) for multilingual search.

This paper presents a case study, exploring various aspects of vocabulary mapping that builds on work over the last 2 years in collaboration with the Archaeology Data Service (ADS) and vocabulary providers, Historic England—formerly English Heritage (EH), the Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS) and Wales (RCHAMW). It draws on work from two connected research projects (SENESCHAL followed by ARIADNE).[1]

The case study builds on previous research in collaboration with EH and ADS undertaken as part of two University of South Wales research projects, STAR [6] and STELLAR [7]. STAR provided semantic interoperability between diverse archaeological datasets from different organizations and excavation reports. This was achieved via a unifying upper level conceptual framework, the CIDOC Conceptual Reference Model (CRM—ISO 21127:2014) [8]. Previously cross searching was not possible; the STAR Demonstrator cross searches five excavation datasets and an extract from the ADS OASIS Grey Literature digital library [9]. The CRM does not supply a vocabulary of concepts so various EH thesauri and glossaries provided the controlled terminology required for a semantic link to be made between different terms for the same concept, for example "*post-hole*" and "*posthole*". These vocabularies were made available as web services for the purposes of the STAR project [10].

The subsequent STELLAR project generalized the data extraction tools produced by STAR to facilitate their adoption by third-party data providers, who are not ontology specialists. The data were represented in standard Resource Description Framework (RDF) [11] formats and published by the ADS as Linked Data [12] using the tools developed during the STELLAR project that make it easier for non-specialists to generate linked data from their application data. Using templates, the STELLAR tools convert archaeological data to RDF without requiring detailed knowledge of the underlying ontology [13]. In addition to CRM-based templates, there is a template allowing a glossary or thesaurus to be expressed in SKOS.

[1] *Semantic Enrichment Enabling Sustainability of Archaeological Links* (SENESCHAL) [4] was funded by the UK Arts and Humanities Research Council and coordinated by the Hypermedia Research Group at the University of South Wales (formerly University of Glamorgan). *Advanced Research Infrastructure for Archaeological Dataset Networking in Europe* (ARIADNE) [5] is an ongoing EC FP7 Project.

The key aim of both projects was to achieve semantic interoperability through the use of a common ontological model and domain controlled vocabularies. While the research objectives were met, overcoming a lack of vocabulary control (and lack of unique identifiers where vocabulary did exist) consumed more resources than anticipated [9]. Although the STELLAR tools can generate controlled types with unique (URI) identifiers, the linked data produced by the immediate project employed free-text strings, since standard unique identifiers for the domain thesauri were not available. This means that the resulting Linked Data are not connected to other data via the thesaurus. While the need for standards for the representation of vocabularies in the archaeology domain is widely recognized [14], this situation has acted as a break on the impact of semantic technologies in the sector. The SENESCHAL project addressed these issues.

It was initially believed that the degree of vocabulary control practiced in archaeology would be high given the available terminology resources together with advisory guidelines encouraging the use of controlled vocabulary. However, empirical evidence gathered from datasets previously encountered during the STELLAR project [7] suggested this is not always the case (as shown in the examples in Sects. 2.1 and 2.2). This was due in part to data entry forms allowing uncontrolled free-text data entry, a contributing factor could also be that open licensed controlled vocabularies were not readily available in standard semantic formats. It became clear that methods to provide controlled indexing within the data entry workflow could be improved. Some data entry systems employ pick lists based on major thesauri, but the output is still text rather than any standard common identifier that other systems might also employ. Links to online thesauri exist within some web-based data entry systems but allowing free-text entry will typically introduce some errors. Clearly, there is a need for standard unique identifiers for vocabularies and their concepts, allowing unambiguous references to them.

1.2 Linked Data

The concept of Linked Data [15] has been an emerging theme within the computing and digital heritage areas in recent years. It is anticipated that it will facilitate an organic and evolutionary approach to semantic technologies and semantic web ambitions. Linked Data are characterized as going beyond the linking of web documents by affording the linking of 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 mech-

anism 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)

Linked Data rest upon layers of technological standards. Within archaeology, vocabulary standards have been envisaged as a potential solution to the current fieldwork situation where isolated silos of data impede sharing, cross search, comparison and reinterpretation of archaeological information. Interoperable standards for encoding fieldwork data and reports will afford a step change in archaeological practice with respect to digital publication and dissemination of data and also results. This will enable meta-research explorations that ask new questions of existing dispersed datasets. The ARIADNE FP7 project on archaeological infrastructure [5] promotes best practices for publishing and interlinking datasets for sharing, integration and reuse of archaeological data. Publication and reuse of Linked Data are seen as important innovative practices in this regard.

Some existing examples of prominent large-scale published Linked Open Data (LOD) reference resources include the Library of Congress Subject Headings (LCSH) [http://id.loc.gov/authorities/subjects.html], the Dewey Decimal Classification (DDC) [http://dewey.info/] and the Getty Art and Architecture Thesaurus (AAT) [http://vocab.getty.edu/].

With regard to the case study presented here, the research aims for SENESCHAL included providing access to key vocabulary resources and the semantic enrichment of existing metadata to align legacy datasets with the LOD controlled vocabularies. As a result, a set of 14 prominent national UK archaeological thesauri and vocabularies originating from EH, RCAHMS, RCAHMW (ranging in size from just 16 terms to over 7900 terms and including Event types, Materials, Monument types, Object types, and Periods) are now freely available as LOD—together with open source web services and user interface controls (widgets) using the Linked Data vocabularies.[2] Online documentation is available on the operation of the services and widgets and how to apply them in the context of browser-based applications [16, see].

The LOD vocabularies published as a result of the SENESCHAL project are made available as downloadable files ('raw' data files provided in Simple Knowledge Organization System (SKOS) [17] RDF format, alphabetical and hierarchical listings provided in Portable Document Format (PDF). There is a SPARQL [18] endpoint exposed for formulation of queries directly on the data, and also a set of web services and 'widget' user interface controls encapsulating some pre-defined functionality for embedding in applications.

2 Alignment

2.1 Problems with uncontrolled text entry fields

Data providers lack an easy means to provide uniquely identified controlled indexing of data that is compatible with semantic technologies and standards, such as Linked Data and SKOS. Currently, thesauri are not fully part of the workflow for user indexing and search. During the SENESCHAL project it was frequently observed that free-text data entry had taken place in legacy archaeological data, often evidenced by simple syntactic anomalies being present in the archived dataset. These minor differences in spelling or punctuation can hinder subsequent attempts at alignment of data, preventing local or wider interoperability. Spell check validation can help to an extent, although some errors may form valid words in their own right. A pilot analysis conducted on two existing datasets uncovered a wide range of problems that would act as a barrier for any cross search. This empirical review identified a number of potential problems with existing fields that were intended to hold (or could hold) controlled values. Many of the issues encountered (and possible suitable solutions) may be a familiar theme to those from a library science/knowledge organization background:

- Various spelling errors (e.g. "*POSTHLOLE*", "*CESS PITT*", "*FURRROWS*", "*CAIRNN*", "*NEOTLITHIC*")
- Alternate word forms (e.g. *"BOUNDARY" / "BOUNDARIES", "GULLEY" / "GULLIES"*)
- Reordering of words in phrases (e.g. "*PIT, CESS*", "*TRENCH, ROBBER*")
- Additional term prefixes or suffixes (e.g. "*RED HILL (POSSIBLE)*", "*TRACKWAY (COBBLED)*", "*CROFT?*", "*PORTAL DOLMEN (RE-ERECTED)*")
- Terms not intended for indexing but nonetheless required to indicate a negative finding (e.g. "*NONE*", "*UNIDENTIFIED*", "*N/A*", "*INCOHERENT*")
- Attempts at providing additional structure within a single field (e.g. nested delimiters: "*POTTERY;CERAMIC TILE;IRON OBJECTS;GLASS*")
- Very specific longer compound phrases (e.g. "*SIDE WALL OF POT WITH LUG*", "*BRICK LINED INDUSTRIAL WELL OR MINE SHAFT*", "*ALIGNMENT OF PLATFORMS AND STONES*")

Thus in practice both a lack of truly controlled indexing and errors with existing indexing were frequently encountered. This widespread problem also affects metadata for non-text datasets and metadata for grey literature repositories. It

[2] Following completion of the project operational management of HeritageData.org and governance of the vocabularies was transferred to EH, RCAHMS, RCAHMW collectively, under the rubric of the FISH Terminology Working Group [19].

Table 1 Some examples of matches between original legacy archaeological data values and Historic England monument types thesaurus terms using a string similarity algorithm

Original data values [sic]	Highest scoring controlled concept term matches		
	Concept ID	Term matched	Score (%)
AXE FACOTRY	69115	AXE FACTORY	90
BOUNDARIES	70323	BOUNDARY	77
BOUNDARY	70323	BOUNDARY	100
BUIED SOIL HORIZON	140223	BURIED SOIL HORIZON	97
CAIRN	68612	CAIRN	100
CAIRN (POSSIBLE)	68612	CAIRN	100
CAIRNN	68612	CAIRN	90
CESS PITT	70434	CESS PIT	94
CHAMBERED TOM	70064	CHAMBERED TOMB	96
COMERCIAL	68777	COMMERCIAL	94
CROFT?	68617	CROFT	90
CUP-MARKED STONE	69996	CUP MARKED STONE	93
DICTH	70351	DITCH	80
ENCLSOURE	70354	ENCLOSURE	88
EXTRACTION PIT	69101	EXTRACTIVE PIT	85
EXTRACTIVE PIT	69101	EXTRACTIVE PIT	100

affects both legacy metadata and metadata for newly created datasets coming online. It is postulated that one reason behind the use of free text is a tension between conflicting ideas on the purpose of the activity—a desire to be as descriptive as possible during data entry versus conforming to recommendations on use of controlled indexing with consequent limited expressivity for later retrieval purposes. Unfortunately some of the resultant data values observed achieve neither of these things—they are not descriptive enough, whilst confounding efficient data retrieval due to a lack of control. The solution of course is for data entry systems to provide separate fields to enable both free-text description *and* restricted controlled indexing.

2.2 Alignment of legacy datasets to controlled vocabularies

An experimental set of alignment mappings were produced between the metadata from two datasets and controlled vocabulary concepts. Some examples of the previously described issues observed are evident in the data value column shown in Table 1. An edit distance string similarity algorithm (*Levenshtein*) was employed to identify candidate matches by simply comparing legacy data values to all thesaurus concept labels and returning the concept identifier of the best scoring matches. This introduced some flexibility in matching by measuring the optimal number of character edits required to change one string into another, so accommodating small spelling differences or errors. Some pre-processing was applied to assist the similarity algorithm—both data values and terms were trimmed of extraneous whitespace and converted to uppercase characters prior to matching, and any bracketed qualifier suffixes present in terms were removed, e.g. "*CAIRN (POSSIBLE)*" would become "*CAIRN*" for the purposes of term comparison.

The result of the matching algorithm was converted to a percentage score value for display purposes, with a 100 % value representing an exact match. It is important to acknowledge that a 100 % syntactic match should not be regarded as a semantic match, and further contextual evidence should be used to determine whether the suggested concept match is actually correct (see Sect. 2.4).

2.3 Mapping between vocabularies

Another issue that became apparent following the creation and aggregation of these online resources is that while there is fairly rich intra-thesaurus concept linkage (hierarchical and associative links), there are currently no inter-thesaurus links present, despite the fact that a number of thesauri converted in the SENESCHAL project have quite significant semantic overlap. Some share a common origin—RCAHMS and RCAHMW each maintain a separate Monument Types thesaurus, both historically derived from the original Historic England Monument Types thesaurus. Historic England and RCAHMS also have separate Archaeological Object Types thesauri, derived from a thesaurus originally developed by the Archaeological Objects Working Party. Clearly there is great scope for some fairly straightforward inter-thesaurus linking of concepts.

A related issue is that there are currently only minimal links out to *external* Linked Data resources. Tim Berners-

★	Data made *openly* available on the web in any format
★★	As above, but in a machine readable structured data format (e.g. Excel)
★★★	As above, but in a non-proprietary structured data format (e.g. XML)
★★★★	As above, but using W3C open standards (e.g. URIs, RDF & SPARQL)
★★★★★	As above, and also **linking out** to other external LOD

Fig. 1 The 5 star deployment scheme for linked open data

Lee devised a 5 star deployment scheme [20] with which to grade LOD[3], indicating that the thesauri made available as a result of the SENESCHAL project currently (at the time of writing) achieve 4 stars.

In considering mapping between thesauri it is necessary to first decide on a suitable architecture for the mappings. Figure 2 illustrates that the maximum number of links between equivalent concepts in a many-to-many (M2M) architecture is $n^2 - n$ (where n is the number of datasets containing potentially matching concepts); for a hub or star architecture the maximum is $2n$. For a small project interlinking just two or three datasets the M2M architecture would be satisfactory, for anything above three datasets the hub architecture becomes more appropriate. These issues are discussed in ISO25964-2:2013 section 6 "Structural models for mapping across vocabularies" [21]. Mapping between any more than three vocabularies would be more efficient and scalable using the hub architecture, an intermediate structure of nodes onto which the concepts from each local vocabulary may be mapped. A search on a concept originating from any one vocabulary can then utilize this mediating structure to route through to concepts originating from any of the other vocabularies, possibly expressed in other languages.

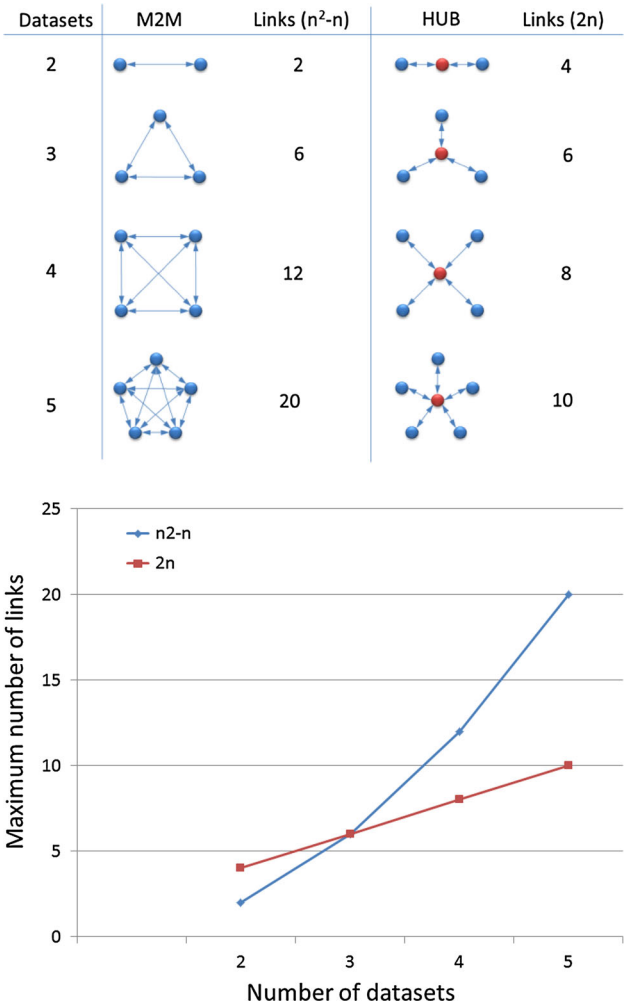

Fig. 2 Maximum potential number of inter-thesaurus links between equivalent concepts using M2M versus HUB architecture

2.4 Use of contextual evidence in creation of mappings

Automated tools can assist in the creation of mappings to an extent; however, these tools should be used in conjunction with domain expert mediation to ensure a consistent quality of mappings. Results still require manual oversight using other contextual data associated with the concept, as even an exact match between preferred terms is still only a syntactic match not a semantic match. An illustration of this issue is shown in the selected data fields compared in Table 2.

Here the RCAHMS and EH 'Monument Types' thesauri each define a concept with a preferred term of *TENEMENT*. However, the scope notes and related concepts show that these are actually fundamentally different concepts and should not be mapped together. The requirement is concept alignment not term alignment and so additional contextual evidence should be exposed and used by both tools and humans to further qualify a match. For example:

- *Syntactic matching on concept labels*—may be inexact matching, employing stemming, string matching algorithms (e.g. using the *Levenshtein* edit distance approach as described previously). There may be a need to strip bracketed term 'qualifiers', and to consider also strip-

[3] It should be noted that the criteria of the "5 star" scheme as described in Fig. 1 do not measure data *quality* OR *quantity*, the scheme only grades LOD in terms of data formats used, licensing conditions and the presence of (an unspecified quantity of) external links.

Table 2 100 % exact textual match on preferred labels, but further contextual evidence shows that the two concepts are fundamentally different

Thesaurus	Monument type thesaurus (RCAHMS)	Monument type thesaurus (Historic England)
Identifier	http://purl.org/heritagedata/schemes/1/concepts/467	http://purl.org/heritagedata/schemes/eh_tmt2/concepts/68997
Preferred label	TENEMENT	TENEMENT
Scope note	A large building containing a number of rooms or flats, access to which is usually gained via a common stairway	A parcel of land
Broader concept(s)	MULTIPLE DWELLING	SETTLEMENT
Related concept(s)	LODGING HOUSE, FLAT, FLATS	DWELLING

ping white space and punctuation. Case sensitivity may also need to be considered, depending on the similarity algorithm adopted. Terms may require prior translation in the case of mapping multilingual terminology resources.

- *Scope note evidence*—there may be full, partial or no overlap in scope between concepts, realistically this contextual evidence requires human oversight. Scope notes may require prior translation in the case of mapping multilingual terminology.
- *Synonyms*—the assessment of groups of alternate synonymous terms may help to reinforce the case for a match between two concepts.
- *Hierarchical context* —ancestors and descendants. If a top-down approach is employed there may be existing mappings higher up in the structure produced during the mapping process that can give additional contextual evidence to a potential match under consideration. The Ontology Alignment Evaluation Initiative (OAEI) [22] 2013 Library Test Case in matching two real-world thesauri [23] noted that "matchers still rely too much on the character string of the labels […] incorrect matches could be prevented […] by taking these higher levels of the hierarchy into account […] We believe that further exploiting this context knowledge could be worthwhile".

It is also important to record additional metadata about the mappings being produced, as a new set of mappings constitutes a new dataset in its own right and so requires appropriate authorship and licensing information. One approach to this is the use of the VoID vocabulary [24], which may be used to describe linked RDF datasets using the *Linkset* element. The SENESCHAL vocabularies have each been described using VoID metadata which is documented together with links to example resources (see http://datahub.io/dataset?q=heritagedata). RDF and Linked Data are intended to be machine readable and self-describing—allowing automated exploration without reference to external supplementary documentation. The production and publishing of external static supplementary metadata in this way introduces issues of data currency and a maintenance legacy; particularly where the metadata contains data quantity summaries (e.g. total counts of concepts)—so some thought will need to be devoted to keeping any external supplementary metadata updated in line with any updates made to the actual underlying data.

3 Tools and techniques

The alignment issues described in Sect. 2 require suitable techniques, methodologies and practical tools to devise mappings between thesaurus concepts. ISO 25964-2:2013 [21] provides an overview of vocabulary mapping and describes approaches for creating mapping relationships between concepts in different vocabularies. It notes the need for accuracy, stating *"…it is better to have no mapping at all than to establish a misleading one"*. Section 14 of the standard discusses some techniques for identifying candidate mappings.

Of course, mapping between vocabularies predates Linked Data. For example, Liang and Sini [25] describe a method for mapping the multilingual AGROVOC thesaurus to the bilingual Chinese Agricultural Thesaurus, employing exact, broader and narrower mapping relationships. Candidate mappings are to be derived programmatically by string comparison of terms and subsequently validated by intellectual review, taking into account the hierarchical and associative relationships of the corresponding concepts in the two thesauri. General tools exist for generating mappings between Linked Data items [a useful list can be found at [26]. A full evaluation of the many tools available is beyond the scope of this paper; however, informal experimentation explored a number of tools including OpenRefine [27], an extended version LODRefine [28], SAIM [29] Instance Matching application (a browser-based interface for creating LIMES [30] Link Discovery Framework link specifications), and the Silk Link Discovery Framework [31]. The preliminary results with the Silk application were encouraging; a 'link specification' was configured fairly quickly to compare preferred labels from two separate thesauri using the *Levenshtein* distance algorithm. Such tools do require some installation and configuration and there is an inevitable learning curve involved in correctly setting up linkage rules, property paths, transformations, aggregations, etc., and in interpreting resul-

tant match scores. While the Silk application did facilitate some manual intervention and interactive comparison, the focus of such tools is typically on automatic link generation functionality and bulk automation; they do not necessarily present the user with sufficient contextual thesaurus-specific data up front in a convenient form to make an informed decision on potential mappings. For example, techniques involving syntactic matching of terms will not identify a connection between terms regarded as synonymous if they do not co-occur in the entry vocabularies of the thesauri, e.g. "*CURRENCY*" and "*MONEY*", another reason why human oversight is important. In the case of thesaurus-to-thesaurus mapping it can be useful for instance to compare hierarchical structures side by side, displaying any existing confirmed mapping links between these structures as well as any candidate links. In our view, user-centred tools tailored for the specific task of thesaurus-to-thesaurus mapping, together with documented methodologies, techniques and approaches can improve the accuracy of the overall process. Section 5 describes the creation and use of such a dedicated tool in mapping between LOD vocabulary concepts. As an illustration of the value of vocabulary mapping, an initial experimental exercise in applying vocabulary mapping techniques in the context of the multilingual ARIADNE project is described in Sect. 4.

4 Exploratory case study—cross search of ARIADNE resources by subject

4.1 Describing resources by subject

The ARIADNE project [5] does not aim to replace existing repositories, but to consolidate their metadata to facilitate cross search. The metadata describing repository resources is modelled using the ARIADNE Catalog Data Model (ACDM) [32], an extension of the Data Catalog Vocabulary (DCAT) [33]. ACDM represents resources as subclasses of the *ArchaeologicalResource* class. Among other properties, this class uses the *dct:subject* property to associate the resource with one or more items from an existing controlled vocabulary. Where practicable it is considered good practice to use identifiers representing vocabulary items, rather than potentially ambiguous literal text values. Figure 3 compares the two approaches.

Despite the ARIADNE repository items being described by subjects originating from controlled vocabularies, cross search would still remain a problem as there are multiple overlapping local vocabularies in use with no formal semantic links or mappings currently existing between them.

4.2 Mapping local vocabulary resources to a mediating hub structure

Leading up to and contributing to ARIADNE, major archaeological vocabularies have been published online, allowing them to be reused in a wide variety of applications. These include the following:

- Data Archiving and Networked Services (DANS) provide a list of monument types (*Archeologische complextypen*) (http://rce.rnaviewer.net/nl/structures). The data are also available in a custom-structured XML format, containing embedded SKOS concepts with unique identifiers.
- FASTI Online (FASTI) uses a flat list of monument types in the advanced search interface (see http://www.fastionline.org/data_view.php).
- The Italian Ministry for Heritage and Cultural Activities: Central Institute for Cataloguing and Documentation (Istituto Centrale per il Catalogo e la Documentazione, ICCD) publishes terminology for types of archaeolog-

```
@prefix data: <http://example.org/data/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix aat: <http://vocab.getty.edu/aat/> .

# Subject as literal value - potentially ambiguous,
# has undefined meaning and scope; no further data.
# Other refs to "vessel" may mean something else.
data:resource1 dct:subject "vessel"@en .

# Subject as unique identifier - unambiguous,
# meaning and scope can be defined with further
# data attached to the subject identifier. Other
# refs to aat:300198642 always mean the same concept.
data:resource1 dct:subject aat:300198642 .
```

Fig. 3 Examples describing the subject of a resource (Turtle RDF format)

Fig. 4 SPARQL 1.1 query to extract the AAT poly-hierarchical structure

```
# SPARQL 1.1 - extracting the AAT hierarchical structure
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX gvp: <http://vocab.getty.edu/ontology#>

CONSTRUCT {?s gvp:broader ?o}
WHERE {
    ?s skos:inScheme <http://vocab.getty.edu/aat/>;
      (gvp:broaderGeneric | gvp:broaderPartitive) ?o .
    MINUS {?s a gvp:ObsoleteSubject} # not required
    MINUS {?o a gvp:ObsoleteSubject} # not required
}
```

ical sites (see http://www.iccd.beniculturali.it/getFile.php?id=182). The ICCD terminology is intended to be made available in SKOS RDF format for use in ARIADNE.

- The SENESCHAL project resulted in publication of 14 thesauri as SKOS RDF LOD including the Historic England Monument Types Thesaurus (see http://www.heritagedata.org/blog/vocabularies-provided/).
- Deutsches Archäologisches Institut (DAI) have produced a multilingual archaeological dictionary (http://archwort.dainst.org/thesaurus/en/). A faceted hierarchical structure is apparent when viewing terms in the German language.

Mapping between these local vocabularies would provide a useful mediation platform for cross search; however, as discussed in Sect. 2.3 the creation of links directly between concepts in multiple different vocabularies can quickly become unmanageable as the number of vocabularies increases. Using the mapping hub architecture, a search on a concept originating from any one vocabulary can utilize this mediating structure to route through to concepts originating from other vocabularies, possibly expressed in different languages.

The Getty Research Institute has published online as LOD a number of significant structured vocabulary resources intended for use in the cultural heritage domain in an ongoing project [34]:

- AAT—Art and architecture thesaurus
- TGN—Thesaurus of geographic names
- ULAN—Union list of artist names
- CONA—Cultural objects name authority

The AAT has a faceted poly-hierarchical structure of concepts related to cultural heritage, with labels and notes in multiple languages. It has a good breadth of domain coverage; together with clear scope notes defining the scope of usage for each concept so has the potential to act as a hub for vocabulary mapping. Using the Getty Vocabularies LOD SPARQL interface (http://vocab.getty.edu/sparql) the poly-hierarchical structure (only) of the AAT was extracted using the SPARQL 1.1 query as shown in Fig. 4, to be used as a mediating structure in an experimental mapping and querying exercise.

Mappings from local vocabulary resources to the concept identifiers of the AAT can be expressed in RDF using SKOS mapping relationships [35]. The example FASTI to AAT mappings shown in Fig. 5 were determined manually by consulting the Getty Vocabularies LOD search facility (http://vocab.getty.edu/). The looser *skos:closeMatch* relationship has been used in each instance rather than *skos:exactMatch,* in light of the absence of scope notes for the FASTI terms.

Further example SKOS mappings were made for other local vocabularies for the purpose of the pilot exercise. In the case of vocabularies in languages other than English, Google Translate (https://translate.google.com/) was used to determine English translations of the terminology in order to assess potential matches. The terms were then manually mapped to AAT concepts (Fig. 5). Example extracts of the experimental mappings produced are listed in "Appendix 1". Once the local vocabularies were mapped to AAT concepts this aggregated data formed the basis of a *semantic framework* for subsequent mediation of queries.

4.3 Cross searching and expanding the mapped vocabularies

Figure 6 illustrates an extract of the existing AAT hierarchical structure for concept 300266755 ("*cemeteries*"). Using a free-text approach to search on literal subject indexing tags to find instances of "*cemeteries*" would not match on the term "*cemetery*"; it would not locate any multilingual indexing terms, and would not find items indexed using other semantically related terms (e.g. "*graveyards*", "*catacombs*", etc.). Controlled vocabularies such as AAT often include multiple (sometimes multilingual) alternate concept labels, as in the example shown in Table 3. When searching a collection

Fig. 5 Example SKOS mappings from FASTI concepts to AAT concepts (Turtle RDF format)

```
@prefix fasti: <http://www.fastionline.org/monuments/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix aat: <http://vocab.getty.edu/aat/> .

# Note: FASTI identifiers were invented for this example
fasti:burial skos:prefLabel "Burial"@en;
  skos:closeMatch aat:300387004 .
fasti:catacomb skos:prefLabel "Catacomb"@en;
  skos:closeMatch aat:300000367 .
fasti:cemetery skos:prefLabel "Cemetery"@en;
  skos:closeMatch aat:300266755 .
fasti:columbarium skos:prefLabel "Columbarium"@en;
  skos:closeMatch aat:300000370 .
fasti:mausoleum skos:prefLabel "Mausoleum"@en;
  skos:closeMatch aat:300005891, aat:300263068 .
```

Table 3 Multilingual preferred and alternate labels for AAT concept 300266755

Concept ID	http://vocab.getty.edu/aat/300266755
Preferred labels (language)	"Cemeteries" (en), "campos santos" (es), "campi santi" (es), "cimetières" (fr), "begraafplaatsen" (nl)
Alternate labels (language)	"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)

Fig. 6 Extract of the AAT hierarchical structure for the concept of "*cemeteries*"

```
• cemeteries [aat:300266755]
  – <cemeteries by form> [aat:300000366]
    o graveyards [aat:300000360]
    o catacombs [aat:300000367]
    o columbaria (cemeteries) [aat:300000370]
    o necropolises [aat:300000372]
    o memorial parks [aat:300266756]
    o lawn cemeteries [aat:300266757]
  – <cemeteries by function> [aat:300000373]
    o cineraria (cemeteries) [aat:300000368]
    o national cemeteries [aat:300000375]
    o pet cemeteries [aat:300000376]
    o potter's fields [aat:300000378]
    o war cemeteries [aat:300000380]
    o churchyards [aat:300008170]
    o military cemeteries (veteran cemeteries) [aat:300266758]
```

without controlled indexing, one recall enhancing information retrieval technique is to expand textual search criteria to include these additional alternate terms. However, whilst improving recall this could have a detrimental effect on precision, as the alternate terms may be ambiguous.

A hierarchically structured controlled vocabulary facilitates the implementation of another form of semantic expansion to encompass all narrower descendants of a concept. If concept identifiers have been used for subject indexing rather than just textual terms then this semantic query expansion has the advantage of (potentially) improving the recall measure for query results *without* sacrificing precision.

As a practical demonstration of such a hierarchical semantic expansion technique, the AAT data previously

Table 4 Results of running the query from Fig. 7 against the combined AAT structure and manual mappings

Concept identifier	Concept label
tmt:70054	Barrow cemetery
tmt:70055	Cairn cemetery
fasti:catacomb	Catacomb
tmt:91386	Catacomb (funerary)
iccd:catacomba	Catacomba
tmt:70053	Cemetery
fasti:cemetery	Cemetery
dans:be95a643-da30-40b9-b509-eadfb00610c4	Christelijk/joodse begraafplaats
iccd:cimitero	Cimitero
iccd:colombario	Colombario
fasti:columbarium	Columbarium
tmt:70056	Cremation cemetery
dai:1819	Friedhof
dai:1947	Gräberfeld
tmt:70060	Inhumation cemetery
dans:6a7482e5-2fd5-48fb-baf4-66ad3d4ed95e	Kerkhof
dai:3736	Kolumbarium
tmt:92672	Mixed cemetery
iccd:necropoli	Necropoli
tmt:70053	Necropolis
dai:2485	Nekropole
dans:abb41cf1-30dc-4d55-8c18-d599ebba1bc2	Rijengrafveld
dans:b935f9a9-7456-4669-91d0-2e9c0ff7d664	Vlakgrafveld
tmt:100531	Walled cemetery

Fig. 7 SPARQL 1.1 query on AAT semantic framework plus local vocabulary mappings to locate concepts related to FASTI "*cemetery*" concept

```
# SPARQL 1.1 query to locate concepts related via AAT
# mediating structure to FASTI "cemetery" concept
PREFIX gvp: <http://vocab.getty.edu/ontology#>
PREFIX fasti: <http://fastionline.org/monumenttype/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT DISTINCT ?concept ?label WHERE {
fasti:cemetery (skos:exactMatch | skos:broadMatch |
skos:closeMatch) ?aatconcept .
 ?aatdescendant gvp:broader+ ?aatconcept .
 {
   {?concept (skos:exactMatch | skos:broadMatch |
skos:closeMatch) ?aatdescendant}
   UNION
   {?concept (skos:exactMatch | skos:broadMatch |
skos:closeMatch) ?aatconcept}
}
 OPTIONAL {?concept skos:prefLabel ?label}
}
```

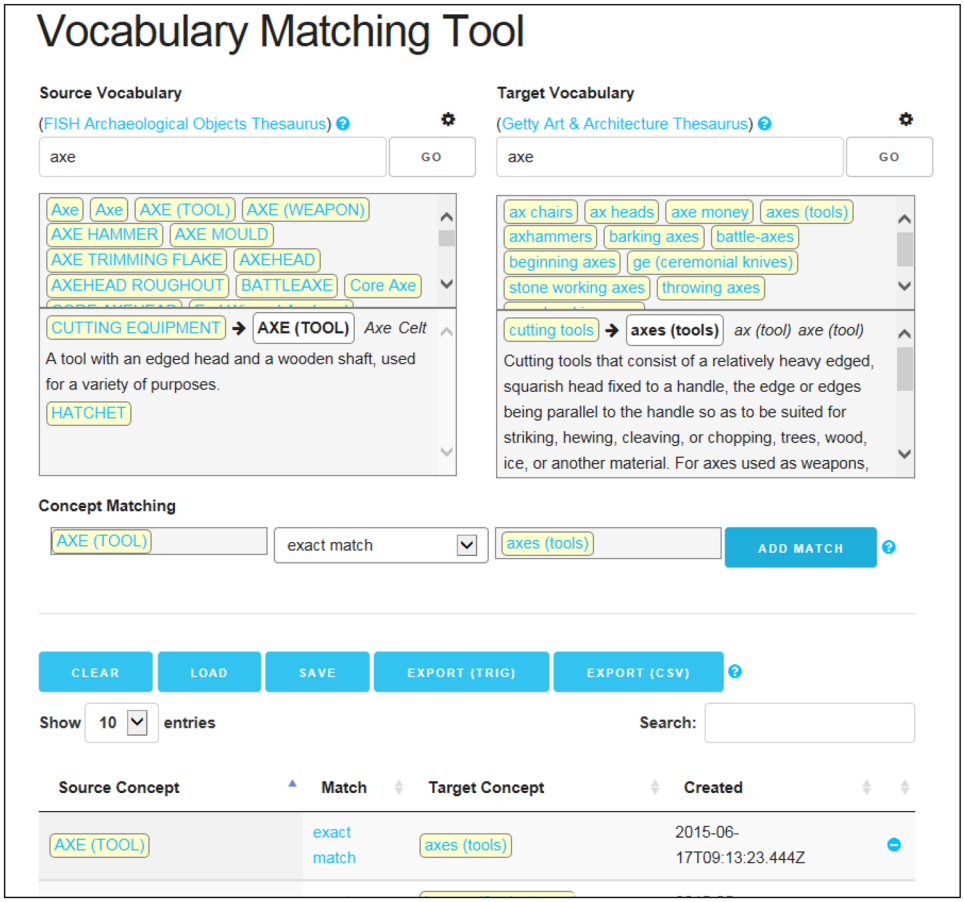

Fig. 8 Vocabulary matching tool

extracted together with all the manual mappings produced were imported to a desktop-based RDF querying application SPARQL GUI (part of the *dotnetredf toolkit*, http://dotnetrdf.org/) and queried using the SPARQL 1.1 query as shown in Fig. 7. This query uses the concept identifier *fasti:cemetery* as a search entry point to locate vocabulary concepts from any other vocabulary mapped into the AAT structure at that hierarchical level *or below*. Note that this query uses concept identifiers rather than textual terms; it is assumed that the initial search term has already been translated to an appropriate concept URI in a previous disambiguation stage by the search interface.

This is made possible by the use of technological data standards that have been followed in ARIADNE and which underpin controlled vocabulary LOD (RDF [11], SKOS [17] and SPARQL [18]). The results (Table 4) show that a query on a concept from one partner vocabulary has located concepts originating from five separate multilingual controlled vocabularies, all related via mappings to the AAT structure. The query has also performed hierarchical semantic expansion to include more specific concepts. A straightforward extension of this technique would find all ARIADNE registry collection resources subject indexed using any of these multi-vocabulary concept identifiers.

5 Vocabulary matching activities

Subsequent to the exploratory pilot case study described in Sect. 4, a more comprehensive vocabulary matching exercise was undertaken by staff at ADS, as part of a larger scale mapping exercise for the ARIADNE project, intended to support cross search in the forthcoming portal. In support of the exercise a vocabulary matching tool was developed by the first author. This section gives details of the tool produced and the mapping exercise undertaken employing it. Retrospective incorporation of the resulting mappings would achieve the 5th interoperability "star" for the LOD published as a result of the SENESCHAL project, as described in Sect. 2.3.

5.1 Vocabulary matching tool

The vocabulary matching tool shown in Fig. 8 was produced to assist in the creation of mappings from source LOD vocabulary concepts published as part of the SENESCHAL project, to Getty AAT LOD target concepts. The tool presents concepts from the chosen source and target vocabularies side by side, exposing additional contextual evidence (as suggested in Sect. 2.4) to allow the user to make a more informed choice when deciding on potential mappings.

The tool is an Open Source lightweight browser-based application working directly with live LOD—querying external SPARQL endpoints rather than storing any local copies of complete vocabularies. The set of mappings developed may be saved locally, reloaded and exported to a number of different output formats.

5.2 Vocabulary matching exercise

The vocabulary matching exercise undertaken by ADS used the vocabulary matching tool, described in Sect. 5.1, to create mappings from source concept terms used in the ADS ArchSearch system to target Getty AAT LOD concepts [34]. Listings of all the unique subject terms used in ArchSearch were compiled in separate files according to the originating thesaurus. As part of previous ADS alignment work, these terms had already been pre-processed to supplement them with URIs of the HeritageData concepts they represented. The listings contained concept terms originating from 5 separate source thesauri (see Table 5 for the counts of mappings produced as a result of the mapping exercise).

Instructions on usage of the vocabulary matching tool together with mapping guidelines were developed to inform and assist the exercise. Results from a pilot mapping exercise were discussed with ADS prior to the full exercise. This proved a useful preliminary stage as it highlighted some issues potentially useful for illustrating general mapping principles. It resulted in some additional recommendations being included in the mapping guidelines to further clarify the requirements and also minor revisions to the mapping tool.

Some initial mappings were made to 'guide term' target concepts—these are intended for structuring vocabulary hierarchies, but not intended for use in indexing. This was considered undesirable for ARIADNE purposes and the mapping guidelines were revised to suggest alternatives in this situation, for example using a *skos:broadMatch* relationship to map to a concept with broader scope where an appropriate concept at the same level of specificity did not exist in the target vocabulary (If desired the vocabulary matching tool could be modified to prevent mappings being made to guide term concepts.)

Discussion with ADS also took place on when multiple mappings from the same concept should be asserted and on the use of *skos:relatedMatch* for ARIADNE purposes. The original guidelines permitted multiple mappings, as in the example shown in Fig. 9. In this case, the *related* mappings were influenced by the existence of narrower concepts (with identical preferred terms to the related AAT target concepts) of the source HeritageData concept, *hospitals*. While not incorrect, the source and target thesauri had varying degrees of specificity in different subject areas and fine-grained mapping seemed inappropriate where an appropriate exact match had been identified. It was also considered to pose unnecessary complications for the intended cross search use case for the mappings. Accordingly, the guidelines were revised to clarify the conditions for multiple mappings and to discourage the use of the *skos:relatedMatch* mapping relationship. In the final exercise, the pilot mappings were revised to retain the *skos:exactMatch* relationship and remove the *skos:relatedMatch* mappings. The revised guidelines suggested multiple mappings from the same concept only in cases where the source concept has two genuinely different expressions in the AAT (that are not immediate parent or child concepts). Since the ARIADNE vocabularies tend to contain some archaeological concepts at a more specific level than covered in the AAT, the use of *skos:broadMatch* was encouraged where appropriate in the guidelines, while *skos:narrowMatch* was considered less likely.

Vocabulary matching exercise results

The final exercise conducted by an ADS staff member with review by a senior archaeologist produced a total of 844 target thesaurus mappings for terms originating from five separate source thesauri. The total numbers and proportions of mappings produced are shown in Table 5, broken down by match type. Some example mappings from the exercise (output by the mapping tool in JSON format) are listed in "Appendix 2". These include examples of mappings that could not

Table 5 Number of mappings produced

Originating thesaurus for subject indexing terms	Exact match	Close match	Broad match	Related match	Totals
EH building materials	4 (33.3 %)	8 (66.7 %)	0 (0 %)	0 (0 %)	12
EH components	7 (77.8 %)	1 (11.1 %)	1 (11.1 %)	0 (0 %)	9
FISH archaeological objects	197 (47.9 %)	96 (23.4 %)	118 (28.7 %)	0 (0 %)	411
EH maritime craft types	13 (54.2 %)	8 (33.3 %)	3 (12.5 %)	0 (0 %)	24
EH monument types	139 (35.8 %)	107 (27.6 %)	141 (36.3 %)	1 (0.3 %)	388
Totals	360 (42.7 %)	220 (26.1 %)	263 (31.2 %)	1 (< 0.1 %)	844

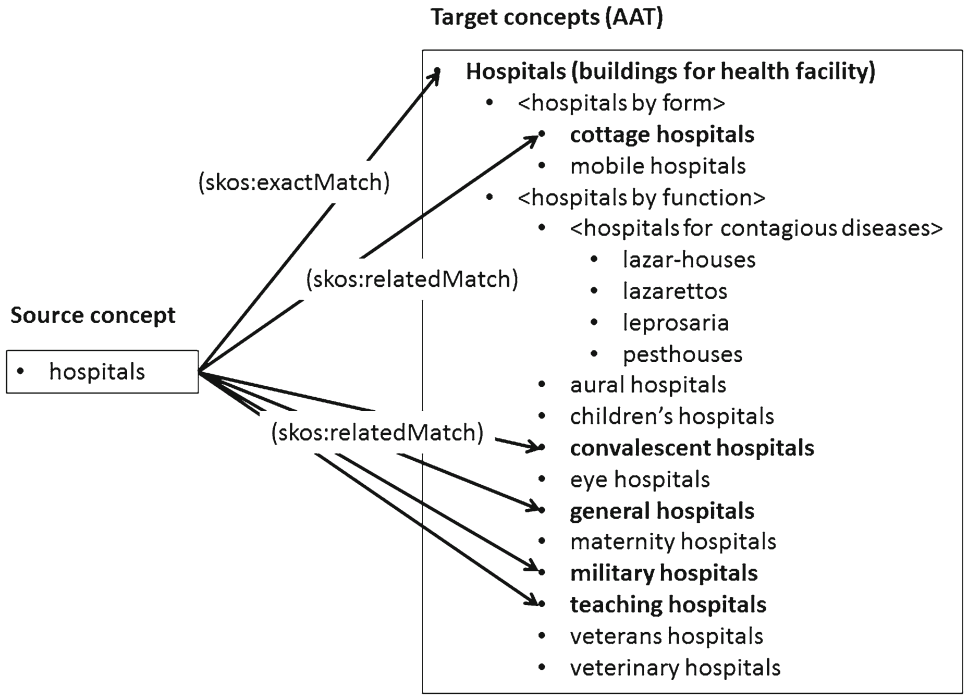

Fig. 9 Multiple target mappings produced from single source concept

be made by string similarity matching alone (e.g. *LYNCHET => AGRICULTURAL LAND*).

The relative proportion of *skos:exactMatch* mappings indicate areas where the target vocabulary exhibited good domain coverage. On the other hand, *skos:closeMatch* is more approximate, indicating similar concepts at the same general level of specificity but where the scope may be slightly different. The *skos:broadMatch* mappings (e.g. in both EH Monument Types and FISH Archaeological Objects) indicate areas where the source concepts were more domain specific than those available in the target vocabulary and where equivalent target concepts at the same level of specificity did not exist. A secondary review of the mappings produced suggested that the presence of a single *skos:relatedMatch* between EH monument type "*CHURCH*" and AAT concept "*churches (buildings)*" was possibly due to an incorrectly selected match type.

As a consideration for further work, the proportions of each type of match present could be used as the basis for developing a similarity metric expressing the overall degree of domain subject coverage overlap between different controlled vocabularies. Currently, the mappings produced by the vocabulary matching tool are unidirectional; reciprocal mappings are not generated by the tool. Reciprocal mappings to supplement those manually entered could be a useful future enhancement. The "Mapping Properties" section of the SKOS reference documentation [35] declares *skos:exactMatch*, *skos:closeMatch* and *skos:relatedMatch* as being instances of *owl:SymmetricProperty*. So $<x>$ *skos:closeMatch* $<y>$ also implies that $<y>$ *skos:close Match* $<x>$. A collaborative environment for encouraging crowdsourcing and developing consensus on the resulting mappings could also form part of some future development.

6 Conclusions

This paper discusses various aspects of vocabulary mapping. Experience from the SENESCHAL project in the publication of controlled vocabularies as LOD is discussed, emphasizing the importance of unique URI identifiers for vocabulary concepts. There is a need to align legacy indexing data to the uniquely defined concepts and examples are discussed of SENESCHAL data alignment work. A case study for the ARIADNE project presents mapping work between vocabularies, based on the Getty AAT as a central hub and employing an interactive vocabulary mapping tool developed for the project, which generates SKOS mapping relationships in JSON and other formats. The potential use of such vocabulary mappings to assist cross search over archaeological datasets from different countries is illustrated in a pilot experiment (see Sect. 4).

The growth and scale of Linked Data has underlined the need for greater commonality in referencing, to avoid local redefinition and duplication of reference resources. Whilst there is little chance of obtaining domain-wide agreement on common vocabularies, the prospects for large-scale interoperability can be substantially improved by creating alignment

links from datasets out to external reference resources. However, this alignment cannot rest upon a simple string match between data elements and the entry vocabulary of a Linked Data thesaurus.

The problems of expressing high-quality mappings between data and a reference vocabulary are related to the issues involved in mapping between vocabularies. The publication of cultural heritage vocabularies as Linked Data as described in Sect. 1 has brought to the fore the issue of expressing mappings between vocabularies as Linked Data. This need is particularly acute in large-scale multilingual projects such as ARIADNE.

There are multiple overlapping local vocabularies in use by ARIADNE data providers with no formal semantic links currently existing between them. An intermediate hub or spine structure is a scalable approach for mapping between concepts in any more than three vocabularies. The Getty AAT provides one example of an appropriate mediating structure, and the exploratory case study demonstrates some advantages of this approach by performing mediated cross search with semantic expansion across multiple multilingual vocabularies.

Currently, completely automatic solutions to mappings appear unlikely to deliver absolute accuracy and some form of assistance should be offered to support intellectual review. Thus operational mapping tools should include the option of a review phase and provide contextual data so that vocabulary providers can make an informed decision on proposed mappings. Mappings published should be accompanied by provenance metadata that includes information on the methods employed. True interoperability within Linked Data can only rest upon quality mappings. The case study discussed in this paper helps to advance this goal by illustrating the type of interactive tools that are required.

Looking ahead, ARIADNE plans to support the provision, management and use of Linked Data in its integrated infrastructure. While some specific linking by hand may be possible directly between individual data elements in closely associated datasets, this is not a scalable approach. Critical for this vision are interlinked concepts from major Linked Open Data vocabularies that can act as hubs in the evolving web of archaeological data.

Acknowledgments The SENESCHAL project was supported by the UK Arts and Humanities Research Council [grant number AH/K002112/1]. The ARIADNE project is funded by the European Commission's 7th Framework Programme (FP7-INFRASTRUCTURES-2012-1-313193). An early version of this work was presented at the 13th European Networked Knowledge Organization Systems (NKOS) Workshop in association with the Digital Libraries 2014 conference. Thanks are due to the Archaeology Data Service for their work with the vocabulary mapping tool reported in this paper. Thanks are also due to ARIADNE and SENESCHAL project partners and the participants of the SENESCHAL workshops.

Appendix 1

Extracts of the concept mappings used for the exploratory case study described in Sect. 4 (expressed in TURTLE RDF format):

```
# declare namespace prefixes
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix aat: <http://vocab.getty.edu/aat/> .
@prefix fasti: <http://fastionline.org/monumenttype/> .
@prefix iccd: <http://www.iccd.beniculturali.it/monuments/> .
@prefix dans: <http://www.rnaproject.org/data/> .
@prefix tmt: <http://purl.org/heritagedata/schemes/eh_tmt2/concepts/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix gvp: <http://vocab.getty.edu/ontology#> .
@prefix dai: <http://archwort.dainst.org/thesaurus/de/vocab/?tema=> .

# Extract of ICCD concepts
iccd:catacomba skos:prefLabel "catacomba"@it .
iccd:cenotafio skos:prefLabel "cenotafio"@it .
iccd:cimitero skos:prefLabel "cimitero"@it .

# Extract of ICCD->AAT mappings
iccd:catacomba skos:closeMatch aat:300000367 .
```

```
iccd:cenotafio skos:closeMatch aat:300007027 .
iccd:cimitero skos:closeMatch aat:300266755 .

# Extract of DANS concepts
dans:8f14ae7e-3d66-4e85-b77c-454a261150e9 skos:prefLabel "begraving"@nl .
dans:e98c8cf0-aa0d-4fcd-99a2-db76cd1d827d skos:prefLabel "begraving, onbepaald"@nl .
dans:87a2f9e9-8e40-4c97-b17b-82275d54c78d skos:prefLabel "brandheuvelveld"@nl .

# Extract of DANS->AAT mappings
dans:8f14ae7e-3d66-4e85-b77c-454a261150e9 skos:closeMatch aat:300387004 .
dans:e98c8cf0-aa0d-4fcd-99a2-db76cd1d827d skos:closeMatch aat:300387004 .
dans:be95a643-da30-40b9-b509-eadfb00610c4 skos:broadMatch aat:300266755 .

# Extract of EH Monument Type concepts
tmt:70053 skos:prefLabel "cemetery"@en .
tmt:100531 skos:prefLabel "walled cemetery"@en .
tmt:92672 skos:prefLabel "mixed cemetery"@en .

# Extract of EH Monument Type->AAT mappings
tmt:70053 skos:closeMatch aat:300266755 .
tmt:100531 skos:broadMatch aat:300266755 .
tmt:92672 skos:broadMatch aat:300266755 .

# Extract of FASTI concepts
fasti:burial skos:prefLabel "Burial"@en .
fasti:catacomb skos:prefLabel "Catacomb"@en .
fasti:cemetery skos:prefLabel "Cemetery"@en .

# Extract of FASTI->AAT mappings
fasti:burial skos:closeMatch aat:300387004 .
fasti:catacomb skos:closeMatch aat:300000367 .
fasti:cemetery skos:closeMatch aat:300266755 .

# Extract of DAI concepts
dai:1819 skos:prefLabel "Friedhof"@de . #cemetery
dai:1947 skos:prefLabel "Gr?berfeld"@de . #graveyard
dai:3736 skos:prefLabel "Kolumbarium"@de . #columbarium

# Extract of DAI->AAT mappings
dai:1819 skos:closeMatch aat:300266755 .
dai:1947 skos:closeMatch aat:300000360 .
dai:3736 skos:closeMatch aat:300000370 .
```

Appendix 2

Example concept mappings produced by ADS in the vocabulary matching exercise as described in Sect. 5 (Vocabulary Matching Tool output in JSON format):


```
{
    "created": "2015-05-18T09:05:45.517Z",
    "sourceURI": "http://purl.org/heritagedata/schemes/mda_obj/concepts/95896",
    "sourceLabel": "AMULET",
    "targetURI": "http://vocab.getty.edu/aat/300266585",
    "targetLabel": "amulets",
    "matchURI": "http://www.w3.org/2004/02/skos/core#exactMatch",
    "matchLabel": "exact match
  },
  {
    "created": "2015-05-18T09:06:25.689Z",
    "sourceURI": "http://purl.org/heritagedata/schemes/mda_obj/concepts/97102",
    "sourceLabel": "ANGON",
    "targetURI": "http://vocab.getty.edu/aat/300036975",
    "targetLabel": "angons",
    "matchURI": "http://www.w3.org/2004/02/skos/core#exactMatch",
    "matchLabel": "exact match
  },
  {
    "created": "2015-03-26T11:50:16.662Z",
    "sourceURI": "http://purl.org/heritagedata/schemes/mda_obj/concepts/95074",
    "sourceLabel": "Animal Bone",
    "targetURI": "http://vocab.getty.edu/aat/300191781",
    "targetLabel": "skeleton components (animal components)",
    "matchURI": "http://www.w3.org/2004/02/skos/core#exactMatch",
    "matchLabel": "exact match
  }
```


References

1. EUROPEANA project. http://www.europeana.eu/
2. Golub, K., Tudhope, D., Zeng, M., Žumer, M.: Terminology registries for knowledge organization systems: functionality, use, and attributes. J. Assoc. Inf. Sci. Technol. **65**(9), 1901–1916 (2014)
3. Zeng, M., Chan, L.: Trends and issues in establishing interoperability among knowledge organization systems. J. Am. Soc. Inf. Sci. Technol. **55**(5), 377–395 (2004)
4. SENESCHAL project: semantic enrichment enabling sustainability of archaeological links. University of South Wales, Hypermedia Research Group. http://hypermedia.research.southwales.ac.uk/kos/seneschal/
5. ARIADNE FP7 project: advanced research infrastructure for archaeological dataset networking in Europe. http://www.ariadne-infrastructure.eu/
6. STAR project: semantic technologies for archaeological resources. University of South Wales: Hypermedia Research Group. http://hypermedia.research.southwales.ac.uk/kos/star/
7. STELLAR project: semantic technologies enhancing links and linked data for archaeological resources. University of South Wales: Hypermedia Research Group. http://hypermedia.research.southwales.ac.uk/kos/STELLAR/
8. ISO standard 21127:2014—the CIDOC conceptual reference model (CRM). http://www.cidoc-crm.org/
9. Tudhope, D., May, K., Binding, C., Vlachidis, A.: Connecting archaeological data and grey literature via semantic cross search. Internet archaeology 30 (2011). doi:10.11141/ia.30.5
10. Binding, C., Tudhope, D.: Terminology web services. Knowl. Organ. **37**(4), 287–298 (2010)
11. Resource description framework (RDF). W3C. http://www.w3.org/RDF/
12. Archaeology Data Service: Linked Data, http://data.archaeologydataservice.ac.uk/
13. Binding, C., Charno, M., Jeffrey, S., May, K., Tudhope, D.: Template based semantic integration: from legacy archaeological datasets to linked data. Int. J. Semant. Web Inf. Syst. 11(1) (2015) (**in press**)
14. Richards, J.D., Hardman, C.S.: Stepping back from the trench edge: an archaeological perspective on the development of standards for recording and publication. In: Greengrass, M. & Hughes, L. (eds.) The Virtual Representation of the Past. Ashgate, pp. 101–112 (2008). http://eprints.whiterose.ac.uk/7795/
15. Bizer, C., Heath, T., Berners-Lee, T.: Linked data—the story so far. Int. J. Semant. Web Inf. Syst. **5**(3), 1–22 (2009). doi:10.4018/jswis.2009081901

16. HeritageData.org—documentation of web services and widget controls produced as part of the SENESCHAL project. http://www.heritagedata.org/blog/
17. Simple Knowledge Organization System (SKOS). W3C. http://www.w3.org/2004/02/skos/
18. SPARQL 1.1 query language. W3C (2013). [http://www.w3.org/TR/sparql11-query/]
19. Forum on Information Standards in Heritage (FISH) Terminology Working Group. http://www.heritagedata.org/blog/about-heritage-data/fish/
20. Berners-Lee, T.: Linked data—design issues (5 star linked data deployment scheme). http://www.w3.org/DesignIssues/LinkedData.html
21. ISO 25964-2:2013 Information and documentation—thesauri and interoperability with other vocabularies—part 2: interoperability with other vocabularies (2013). http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53658
22. Ontology alignment evaluation initiative. http://oaei.ontologymatching.org
23. Cuenca Grau, B., Dragisic, Z., Eckert, K., Euzenat, J., Ferrara, A., Granada, R., Ivanova, V., Jiménez-Ruiz, E., Kempf, A.O., Lambrix, P., Nikolov, A., Paulheim, H., Ritze, D., Scharffe, F., Shvaiko, P., Trojahn, C., Zamazal, O.: Results of the ontology alignment evaluation initiative 2013, pp. 29–31 (2013). http://disi.unitn.it/~p2p/OM-2013/oaei13_paper0.pdf
24. Alexander, K., Cyganiak, R., Hausenblad, M., Zhao, J.: Describing linked datasets with the voID vocabulary. W3C interest group note (2011). http://www.w3.org/TR/void/
25. Liang, A., Sini, M.: Mapping AGROVOC and the Chinese agricultural thesaurus: definitions, tools, procedures. New Rev. Hypermedia Multimed. **12**(1), 51–62 (2006)
26. References to tools and papers about link generation techniques. http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/EquivalenceMining
27. OpenRefine data cleansing and transformation tool. http://openrefine.org/
28. LODRefine (an extension of OpenRefine). https://github.com/sparkica/LODRefine
29. SAIM instance matching application. http://saim.aksw.org/
30. LIMES link discovery framework for metric spaces. http://aksw.org/Projects/LIMES.html
31. Silk link discovery framework. http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/
32. ARIADNE, D12.2 infrastructure design—annex II—ACDM catalogue model, pp. 47–56 (2015). http://www.ariadne-infrastructure.eu/Resources/D12.2-Infrastructure-Design
33. Data Catalog Vocabulary (DCAT) http://www.w3.org/TR/vocab/dcat/
34. Getty vocabularies as linked open data. http://www.getty.edu/research/tools/vocabularies/lod/
35. SKOS mapping relationships. http://www.w3.org/TR/skos-reference/L4138