Institution: University of South Wales		
Unit of Assessment: B11 Computer Science & Informatics		
Title of case study: Transforming archaeological data practice by enabling semantic interoperability		
Period when the underpinning research was undertaken: 2007 - 2018		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Douglas Tudhope	Professor	1993 – present
Ceri Binding	Research Fellow	2000 - 2005; 2007 - present
Period when the claimed impact occurred: 2014-2020		
Is this case study continued from a case study submitted in 2014? No		

1. Summary of the impact

Research conducted at the University of South Wales (USW) has investigated how knowledge organization systems can help to overcome the semantic interoperability problems within digital heritage that result in a lack of meaningful connection between datasets; searches overlook relevant material or return false matches. Nationally, this research led to a major enhancement to quality of service (improved search and data entry) in the redeveloped OASIS (Online AccesS to the Index of Archaeological Investigations), the online database for UK archaeologists and heritage practitioners. Internationally, USW proposed a central spine vocabulary for European archaeological infrastructure and developed open source mapping tools, templates and guidelines. This resulted in a multilingual concept based semantic search capability, opening access to non-English language resources, for the major European archaeological data centres and the broader archaeological community.

2. Underpinning research

The Hypermedia Research Group (HRG) is situated within the Faculty of Computing, Engineering and Science of the University of South Wales. Since 2007, research in HRG has addressed different aspects of semantic interoperability. Database structure varies and simple differences in table and field format can mislead attempts at cross searching, comparing, reusing, or re-interpreting the data. These problems are compounded by terminology issues arising from synonyms and homonyms, particularly in archaeology, where a variety of methods and many different excavation recording systems are employed with differing schema and vocabularies **(R1)**. HRG conducted research on knowledge organization systems and the complementary use of formal ontologies and vocabularies (e.g. thesauri) in order to overcome these problems **(R2, R3)**. Tudhope was a member of the ISO Working Group (2008-2013) that developed the current international thesaurus standard (ISO 25964).

In 2007, HRG began applying its methodology in the archaeological domain through a series of AHRC funded projects in collaboration with the Archaeology Data Service, Bespoke HER User Group, English Heritage, Royal Commissions on the Ancient and Historical Monuments of Scotland/Wales, Wessex Archaeology. This research investigated possibilities for an underlying semantic framework combining a formal ontology with national archaeological vocabularies from English Heritage, the government's statutory adviser on the historic environment in England, and RCAHMS/W. Tools, such as STELLAR and STELETO, were developed to make it easier for non-computing specialists to publish Linked Data in standard semantic representations. This resulted in Linked Data publication of major excavation datasets by archaeologists **(R3, R6)**. The publication of national archaeological thesauri as Linked Open Data, allowing concepts to be unambiguously referenced in subject metadata was also a major outcome **(R4)**. HRG developed web service APIs to facilitate the programmatic use of these linked data vocabularies in applications.

Between 2013-2017, HRG participated in the ARIADNE FP7, a major European archaeological infrastructure project which sought to overcome the lack of connection between isolated European archaeological data resources. It provided an international context

with 24 project partners (including the major European data centres) and a research focus for HRG on mapping between vocabularies for multilingual retrieval capability. As leaders of the ARIADNE Work Package on Linking Archaeological Data, HRG were aware of the challenges posed by resource metadata in different languages for effective cross search in the ARIADNE Portal, the main point of access for searching and browsing datasets from partner institutions. String matches of a term in English (say) will not return results for other partner languages. Additionally, a term, such as 'coins', has different meanings in English and French. HRG research on use cases for knowledge organization systems and information lifecycle frameworks showed the importance of mapping (alignment) between vocabularies for semantic interoperability **(R2)**. Further research by HRG showed the benefits of a hub vocabulary for multilingual retrieval, rather than the many-to-many mapping of partner vocabularies, and the feasibility of adopting the Getty's Art and Architecture Thesaurus (AAT) as the mapping hub **(R5)**. An HRG presentation at a Research Infrastructures Conference (2014) at the National Library in Rome and prototype demonstration at a major European archaeology conference (CAA 2015, Siena) resulted in the ARIADNE project adopting AAT vocabulary mapping and concept-based search in the Portal. HRG designed and implemented practical approaches for data providers to map their local subject vocabularies to concepts in the AAT **(R4, R6)**. Building on this, HRG conducted an item-level integration case study, including semantic query expansion over thesaurus relationships, to demonstrate how detailed archaeological datasets and reports in different languages could be efficiently integrated and cross-searched **(R6)**.

3. References to the research

R1. Tudhope D, May K, Binding C, Vlachidis A. 2011. Connecting Archaeological Data and Grey Literature via Semantic Cross Search. Internet Archaeology, 30, (AHRC funded)
R2. Golub K, Tudhope D, Zeng M, Žumer M. 2014. Terminology Registries for Knowledge Organization Systems: Functionality, Use, and Attributes. Journal of the Association for Information Science and Technology, 65(9), 1901-1916. Wiley. (JISC funded)
R3. Binding C, Charno M, Jeffrey S, May K, Tudhope D. 2015. Template Based Semantic Integration: From Legacy Archaeological Datasets to Linked Data. International Journal on Semantic Web and Information Systems, 11(1), 1-29. IGI Global. (AHRC funded)
R4. Binding C, Tudhope D. 2016. Improving Interoperability using Vocabulary Linked Data. International Journal on Digital Libraries, 17(1), 5-21. Springer. (AHRC/FP7 funded)
R5. Meghini C, et al. 2017. ARIADNE: A Research Infrastructure for Archaeology. ACM Journal on Computing and Cultural Heritage, 10(3), Article 18, (FP7 funded)
R6. Binding C, Tudhope D, Vlachidis A. 2018. A study of semantic integration across archaeological data and reports in different languages. Journal of Information Science, 45(3), 364-386. (FP7 funded)

Selection of Relevant Research Funding (amount awarded to USW)

FP7 Infrastructures Grant. Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (ARIADNE), Tudhope WP Leader, 2013 – 2017 (£170k)
AHRC Grant. (Follow on Fund): SENESCHAL (Semantic ENrichment Enabling Sustainability of arCHAeological Links).Tudhope PI, 2013 – 2014 (£50k)
AHRC Grant. Semantic Technologies Enhancing Links and Linked data for Archaeological Resources (STELLAR), Tudhope PI, 2010 – 2011 (£73k)
AHRC Grant. (AH/D001528/1): Semantic Tools for Archaeological Resources (STAR) in collaboration with English Heritage, Tudhope PI, 2007 - 2010 (£220k)

4. Details of the impact

Enabling accurate and unambiguous use of key archaeological information
OASIS (Online AccesS to the Index of Archaeological Investigations) is an online database for UK archaeologists and heritage practitioners to provide information on their investigations to local Historic Environment Records (HERs) and the national bodies (Historic England,

Historic Environment Scotland, and RCAHMW). It began in 2001 and is maintained by the Archaeology Data Service (ADS), the only accredited UK digital repository for long term preservation of heritage data and a mandated digital archive for UK research councils and Historic England. Its key aim is to open up access to the under-utilised, 'grey literature' (not commercially published) investigation reports produced by commercial developer funded (and community) fieldwork. This is to facilitate reuse or re-interpretation of the results by local and national record managers, heritage professionals and community archaeology groups. OASIS constitutes one of the largest collections of Open Access unpublished heritage reports in the world – 8208 reports were released 2018/19 with 118 new users **(S1)**.

ADS used the Linked Data vocabularies and web services resulting from HRG research **(R4)** to improve data accuracy in the latest iteration of OASIS. The OASIS redevelopment by ADS (2018-2020) was funded by Historic England as part of the Heritage Information Access Strategy. The Deputy Director of ADS confirmed that:
"Programmatic access to national Linked Open Data archaeological vocabularies is a major enhancement to the quality of service in the redeveloped OASIS Index of archaeological investigations; in the new OASIS system, data providers unambiguously specify (as Uniform Resource Identifiers) key information, such as monument and finds types, via vocabulary web services resulting from USW's AHRC funded SENESCHAL research" **(S2)**.

Previously, OASIS data entry did not support accurate entry of controlled fields, such as monument and finds types, resulting in problems with the accuracy and consistency of the data and search results; a 2019 ADS investigation of monument subject terms found only 2042 mapped cleanly to the controlled vocabulary out of 11,851 (unique) entries. **(S2)**.

OASIS has wide reach. Completing an OASIS heritage report is compulsory in the vast majority of English and Scottish Local Authorities for development work undertaken through the Planning process (including Built Heritage). It is also a recommendation of the Chartered Institute for Archaeologists (the leading professional body representing archaeologists working in the UK) that an OASIS form is filled out. In 2019 there were 108,000+ requests for pages **(S2)**. Local Authorities use it to sign off on archaeological conditions as part of development control. OASIS is used by 1386 organisations: mainly commercial archaeological contractors, but also data managers in Historic Environment Records (HERs) and National Monument Records (NMRs), Archaeology Scotland **(S2)**.

Additionally, during 2019-2020, Historic England, the government's statutory adviser on the historic environment made use of HRG linked data research outcomes.
Historic England *"used HRG's STELETO tool to produce Linked Open Data for the suite of heritage reference vocabularies Historic England makes available and promotes as part of its role in setting standards for recording built and buried heritage"* **(S3)**.

Connecting European archaeological via vocabulary mapping
In 2015, the FP7 ARIADNE project employed HRG's methodology - the AAT vocabulary mapping and concept-based search **(R4, R6)** - in its infrastructure. This resulted in the introduction of a multilingual concept based semantic search capability in the (public) Portal **(R5)**, improving the visibility and connectivity of European archaeological resources. This augmented search functionality was available to ARIADNE users, around 10,500 participants from 84 international institutions, including 19 organisations outside Europe **(S4).** In the final project year (Jan 2016 - Jan 2017), the portal had 10,819 visitors from 15,400 sessions **(S4)**. The European Commission recognised the impact of this work and in 2019 funded the follow-on Horizon 2020 ARIADNEplus four year infrastructure project, where HRG are extending the mapping work to new partners from further countries.

The ARIADNE infrastructure combined nearly 2 million metadata records representing diverse datasets and reports from multiple countries and in 23 different languages. However, while integration was achieved at a general schema level, subject indexing in partner languages posed significant challenges for cross searching. HRG developed open source mapping tools,

templates and guidelines for partners to map their source vocabularies to the AAT **(S4)**. Overall, 6,400 local terms from 27 vocabularies employed by 12 project partners were mapped successfully to AAT **(S4)**.

One example of an organisation that benefited from this impact is **The Centre for Archaeology and Innovation Ireland** (Discovery Programme Company Ltd., funded in part by the Heritage Council which provides policy advice for Irish government on heritage issues). It *"made use of the Vocabulary Matching Tool and mapping guidelines developed by the Hypermedia Research group in the Irish Monuments Mapping exercise. This mapped the subject classifications in the Irish National Monuments Service (NMS) and the National Inventory of Architectural Heritage (NIAH) classification scheme to the Getty Art and Architecture Thesaurus (AAT)* **(S5)***."*

The Discovery Programme Technology Manager further observed that HRG tools and guidelines *"…had a significant impact towards the goals of the Discovery Programme* **(S5)**."

Another example is **The Italian government Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information (ICCU)**. They manage the National Library Service, the union catalogue of over 5,200 Italian libraries, and are responsible for providing cataloguing rules and regulations in the country. ICCU contributed Italian datasets and information resources, indexed by the Italian language ICCD-RA Thesaurus. The Director of the Ministry for Cultural Heritage and Activities said the
"ICCU mapped key concepts from its thesaurus to the AAT. The template and guidelines provided by USW facilitated the expression of the ICCU mappings of their thesaurus in terms of the ARIADNE framework. … These mappings had a major impact in improving the search ability … This has significantly enhanced the visibility of Italian ICCU resources within the ARIADNE project and European archaeology more broadly. … ICCU intends to apply this principle in future work and projects **(S6)***."*

Without the AAT mappings, many ICCU resources would only be returned for Italian language searches but can now also be returned for English language searches via AAT concept search **(S6)**.

HRG vocabulary mapping research was adopted by a major Horizon 2020 project.
The mapping work from the ARIADNE FP7 project was adopted by the ArchAIDE H2020 project. One key focus was pottery identification, which is a key component of archaeological investigations including professional contractors and community groups **(S7)**. ArchAIDE (2016-2019) designed an App to support archaeologists in identifying pottery (fragments), using AI to speed up the process. The consortium included two European companies with expertise in archaeological ceramics and the project was awarded 'Best App of 2019' at the Heritage in Motion New Multimedia Competition on European Heritage **(S7, S8)**. The work involved the classification of specific types of pottery, requiring a comparative reference collection with typical pottery types and characteristics applicable to different European languages. International pottery specialists used HRG's mapping methodology and tools to create multilingual vocabularies and AAT mappings for pottery vocabulary (sherd type, vessel form, decoration type, decoration colour) in 7 languages. These multilingual mappings form the spine of the ArchAIDE reference database **(S9)**. An ArchAIDE project report explains the details of the impact:

"Learning from the methodology and using the tools developed … by the Hypermedia Research Group … all partners agreed on a subset of AAT terms that would be used for this neutral spine … it is important to highlight the benefit of aligning the terms used by ArchAIDE with existing vocabularies, in which they offer short-term benefits of concordance and definition, but also longer-term by allowing the archive to be re-used with an inherent and unambiguous understanding of the terminologies used within. The mapping of 'native' to neutral terms was undertaken by each project partner using the mapping tool created by the ARIADNE project" **(S9)**.

5. Sources to corroborate the impact

S1. Wide reach of OASIS. ADS Annual Report 2018/19. https://archaeologydataservice.ac.uk/about/annualReports

S2. OASIS uses HRG vocabulary web services. 2020. Testimonial, Deputy Director, ADS

S3. HRG tool used by Historic England for vocabulary production (2019-2020). Testimonial, Knowledge Organization Specialist, Historic England.

S4. ARIADNE outcomes and significance. 2017. Final Report on Project Impact. http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/ARIADNE_D2-5_Final-Report-on-Project-Impact.pdf

S5. ARIADNE Irish impact. Testimonial Discovery Programme, Technology Manager

S6. ARIADNE Italian impact. Testimonial Ministry for Cultural Heritage and Activities, Director

S7. ArchAIDE context and significance. 2016-2019. http://www.archaide.eu/home

S8. ArchAIDE wins 'Best App' 2019. https://heritageinmotion.eu/himedition/submissions-2019

S9. ArchAIDE Use of HRG methodology and tools. 2019. D10.4. http://www.archaide.eu/documents