

“Knowledge Was Their Treasure”— Applying KO Approaches to Archaeological Research†

Edmund Lee

Historic England, The Engine House, Fire Fly Avenue, Swindon SN2 2EH, United Kingdom,
<Edmund.lee@HistoricEngland.org.uk>



Edmund Lee, MCIFA, is an archaeologist working for Historic England, a UK government agency. He is a graduate of the Institute of Archaeology, University College London. He has worked in government-funded heritage agencies since the early 1990s, developing interests in cultural heritage databases and the application of knowledge organisation principles to heritage data. Particular experience includes the development of the MIDAS standard for heritage inventories and the associated MIDAS XML schemas and advising on the Arches open source heritage inventory software. His current post as Knowledge Transfer Manager focusses on identifying and meeting the knowledge needs of the heritage sector.

Lee, Edmund. 2017. “Knowledge was their Treasure”: Applying KO Approaches to Archaeological Research. *Knowledge Organization* 44(8): 644-655. 20 references.

Abstract: This paper presents a case study for the application of knowledge organisation techniques, including classification, indexing, metadata definition and ontological modelling, to the production of narrative. It examines research work in archaeology, specifically in England. The paper opens with a review of the role of the archaeologist as protagonist in common narrative structures. The paper then first provides examples of classification schemes as a feature of archaeological analysis and their role in narrative and the development of indexing approaches applied to inventories of archaeological sites in England. It contrasts the approach and purpose of classification and indexing. The second section presents examples of how synthesis of archaeological research in England enabled by good knowledge organisation is providing new narratives. The third section presents current work to apply knowledge organisation approaches to research agendas for archaeology in England, specifically a dataset of research questions. It identifies the issues raised and presents initial results—a draft metadata element set for capturing and managing research questions. Modelling of research questions in a knowledge organisation system is a new area for KO application.

Received: 18 September 2017; Accepted: 18 September 2017

Keywords: archaeological research, research questions, knowledge, knowledge organization

† Paper presented at ISKO-UK: Knowledge Organization, What’s the Story? ISKO-UK Biennial Conference, 11-12 September, Canada Water Library & Culture Space, London, UK

1.0 Introduction

Any discussion of archaeology in the context of narrative, the subject of the 2017 ISKO-UK conference, must at least acknowledge in passing the complex interplay between archaeology and both fictional and real-world narratives. The stories that archaeology can tell have powerful resonance; firstly, because they are the stories of human lives, but also because they can only ever be told, in part, from incomplete evidence, which leaves space for interpretation and imagination.

In his book *The Seven Basic Plots—Why We Tell Stories*, (Booker, 2005) Chris Booker has proposed a classification of narratives in fiction into seven essential stories that are so common, because they appeal so deeply to human listeners. The archaeologist is a common protagonist in at least three of these plots:

The Quest: the archaeologist goes in seek of a long-lost treasure.

Overcoming the monster: a foolish archaeologist opens a tomb and unwittingly releases a curse or monster from the past.

Voyage and return: the archaeologist travels either into the real or to a mythical past imagined from archaeology.

Needless to say, archaeologists have not been immune to the allure of telling a good story from their research. How many factual TV programmes featuring archaeology have started with a presenter inviting the audience to “join them on a journey of discovery,” to “uncover hidden secrets,” and to reveal “previously untold stories of the past?” English Heritage, which runs visitor attractions at some four hundred historic sites around the country,

such as Dover Castle and Stonehenge, is explicit in connecting visitors to a historical narrative; their current marketing campaign invites visitors to "step into England's story."

However, the main purpose of the first part of this paper is to present knowledge organisation as applied in archaeological research in the United Kingdom. The intention is to provide a specific case study of how knowledge organisation within one discipline can be a stimulus for the generation of new factual knowledge and new evidence-based narratives, an example that hopefully can find broader value in the discussion of KO as a discipline. The second part presents some examples of new narratives in heritage, benefiting from the traditions of good knowledge organisation systems. It also notes how this work can feed back into developing the need for improvements in knowledge organisation in the sector. The third part then reviews one particular body of knowledge in use in UK archaeology—the use in England of published research frameworks to co-ordinate archaeological research—as an example of how improved knowledge organisation may benefit society. It reports work in progress to develop a prototype for better organisation of a knowledge base of research questions. This work has the potential to increase the efficiency of research in the sector and so to maximise the public benefit gained—a better understanding, a new and relevant narrative, of their past.

2.0 KO in the history of archaeology

2.1 The three-age system

One of the earliest and still in some ways most influential knowledge organisation systems in archaeological research is the "three age" system. This was developed around two hundred years ago by Christian Jürgensen Thomsen. Thomsen, at the age of twenty-eight in 1816, was appointed the curator of the antiquarian collection of the National Museum of Denmark. He set about reorganising the collection of antiquities. His approach was to organise the material not by the sites from which they had come, or who had donated it, but chronologically. He examined the materials they were made of and the associations between materials which occurred together. He extrapolated from this a broad chronology. His classification of material into three "ages," the Stone Age, Bronze Age, and Iron Age, was widely adopted, and is, though widely discussed and debated, still a prevalent narrative in discussion of prehistory in Europe. In England, these ages roughly cover the first ninety-five percent of human occupation up to the first century CE, ending with the Roman conquest of Britain. The "three age" system is so prevalent that it is worth reminding ourselves that this classification is qualitatively

different from the period labels used in historical discourse, such as "Roman" or "Victorian." Those terms would have had some meaning for people alive at the time. In contrast, no-one knew they were living in the "Bronze Age." But Thomsen's great achievement was to, by means of his classification, create a powerful and enduring narrative, which is only now starting to be seriously challenged by archaeological research.

2.2 Typology and classification

The development of typological systems and the classification of material has become, in the two hundred years since Thomsen, a common, almost a defining, activity for archaeological research. Every excavation is followed by a protracted "post-excavation analysis" of the material recovered. The incomplete and partial nature of the evidence pushes archaeologists to attempt to wring the maximum knowledge from their material by developing typologies, classifying and cataloguing, comparing and contrasting artefacts of all types (Orton 1982). Lacking a coherent and central organising paradigm as found in, for example, biological classification using Linnaean taxonomy, archaeologists have created a vast array of different, overlapping, and conflicting typologies and classification schemes for their material.

2.3 Classification to narrative—seriation

There is a purpose to this activity. Archaeologists use classification to build from data to information to knowledge and so to narrative. An example is the technique of seriation. Archaeologists typically excavate and record a sequence of superimposed layers—a stratigraphy. Layers deeper in the ground are assumed to have been deposited earlier and thus be older than layers nearer the surface which overlie them. A common problem, however, is to establish a chronological sequence for archaeological layers (or more correctly "contexts") that have no direct stratigraphic association between them—no common overlying layer. Seriation has been developed as a way to address this problem by providing a relative dating between different contexts, based on the artefacts found within them. The British archaeologist Flinders Petrie, while working in Egypt in 1898, first addressed this problem when he excavated some nine hundred graves of the nineteenth to eighteenth centuries BCE at Diospolis Parva in upper Egypt. He wanted to tell the story of the development of the cemetery, by identifying which graves were earlier and which later. To do this, he developed a classification scheme for stylistic elements of the pottery vessels that were found in the graves; for example, the shape of a rim or base or handle and catalogued all the pottery using that

scheme. He then developed a hypothesis that the different stylistic elements he had identified would have come into fashion, been popular for a while, then declined and gone out of use—in statistical terms forming a bell-curve or normal distribution graph of numbers of pots displaying each element in use against time. The pots included with the burials would represent a “snapshot” of the overlapping bell-curves representing each style type, capturing what was in use at the time of the burial of each individual. Then, by coding what was present or absent in each grave on a strip of paper and reordering the strips to best reflect the assumed bell-curve of changing frequency over time, he could come up with a relative dating of the graves that best fit the evidence—a narrative derived from knowledge organisation.

The technique of seriation and related statistical analyses, as used in modern archaeological practice is not straightforward and can be challenged on several grounds. Most obviously, the frequency of use of a particular pottery style use might not follow a classic bell-curve over time, or its variation might relate to another factor than the passing of time (for example a difference between pottery buried with individual of different gender or status or simply the work of different potters). However, applied with care, it remains a good example of the ways in which classification of material is used in archaeological research to generate narrative knowledge from information. Indeed, the statistical analysis of archaeological material that classification makes possible has developed as a distinct sub-discipline within the field, greatly aided by the introduction of affordable computing power from the 1970s onwards. An international Computer Applications in Archaeology conference has run annually since 1973.

2.4 Inventories—recording the archaeology and historic buildings of England

A parallel development to the use of computing in the detailed examination of individual archaeological contexts and artefacts has been, within the UK, the computerisation of national and local inventories of known or suspected archaeological sites, historic buildings, and other historic assets (Carlisle and Lee 2016). Typically, these are maintained by public authorities and used to map landmarks, to maintain statutory registers of sites protected under law, to prepare a response to development proposals, or to provide resources for education and community involvement. In England, the government agency Historic England maintains several national inventories, most significantly the National Heritage List for England, including some 400,000 entries, which is the definitive statutory list of sites protected by law. Some eighty local councils in England maintain inventories for

their area, known as Historic Environment Records (HERs). Other public agencies, such as the Ministry of Defence, the Church of England, the National Trust, or many National Parks, maintain similar inventories of property in their care. Together, these inventories hold well in excess of one million records of individual sites for England.

2.5 Overlapping and inconsistent records

The different purposes of these inventories shapes the design of the information they contain. The purposes of current inventories are far more varied than, for example, the focussed set of tasks that determine the content of records of collections in a museum or the books in a library. As a result, inconsistencies will occur in the content of these inventories, particularly in three areas:

- The unit of record: That is, what feature of the landscape is separated out by those creating an entry to form a conceptually separate entry in the inventory. A complex site, such as a large settlement may at one extreme be divided into many individual entries to capture and present the information about the different elements—buildings, structures, boundaries, streets, open spaces—or, at the other, may simply be treated as one entry. These are often subjective decisions.
- The sphere of interest: Not all inventories have the same definition of what is of interest. A database just of the buildings in the care of one organisation, such as the Church of England, will not contain entries for, for example, contemporary sites that are not under their management. Similarly, some inventories will have a “cut-off date” and not record sites or buildings that are later, whereas others will record sites right up to the present, if they are of particular interest.
- The recording standard: The different functional requirements and anticipated uses of the inventory mean that the fields of information recorded may well vary from one to another.

However, fundamentally, the different inventories are records of the same archaeological sites and historic buildings, the same landscape or “historic environment.” To gain a full insight into the recorded information about any particular site, it is often necessary to consult multiple inventories.

2.6 Knowledge organisation in UK heritage inventories

The need to improve access to this mass of information, generally collected at public expense, and so improve pub-

lic services, has provided a strong motivation for many knowledge organisation initiatives relating to inventories maintained by public authorities in the UK since the late 1980s. The aim has been to improve interoperability and cross-searching between these inventories (Lee 2004). Three areas of knowledge organisation in particular illustrate this.

Metadata element or database field standards were developed from manual inventories systems such as those maintained by the Ordnance Survey up until 1985, then subsequently by the Royal Commission on the Historic Monuments of England (RCHME). The RCHME published database field definitions in "Recording England's Past" in 1993. This developed subsequently into a more flexible approach, influenced by comparable standards in the museums documentation world, as MIDAS—the Monument Inventory Data Standard in 1997. The association between monument inventories and museum records was then further developed in the work to create an ISO standard ontology for cultural heritage documentation under the auspices of the International Council of Museums (ICOM) documentation committee CIDOC. This was first published in 2006 as ISO 21127:2006—the CIDOC Conceptual Reference Model (CIDOC-CRM). This has subsequently been extended to cover archaeological excavation records, creating the possibility of an overarching ontology to bridge the gap between inventory level and site level data. MIDAS remains the local application of the CIDOC-CRM for inventory type records in England, and is now available as a W3C XML schema definition to guide the design of interoperability tools.

The availability of semantic analysis toolkits for knowledge organisation in the sector provides an infrastructure on which further analytical tools have been developed. The STELLAR project has examined the ways in which semantic analysis of the texts of archaeological reports can be used to mine new standards-compliant datasets from existing digital context records (Hypermedia Research Unit website 2017).

From the late 1980s, the RCHME developed standardised thesauri for the indexing of first architectural and then archaeological site types. These were combined in the 1990s and parallel thesauri developed to cover more specialist records, such as historic watercraft and aircraft wreck sites as these became more common as the subject of investigation and legal protection. As international standards for thesauri were published, these thesauri were deployed in different formats and, in their current versions, are now available as linked data and in RDF and SKOS versions to support implementation in inventory systems (Terminology FISH n.d.).

2.7 Classification and indexing contrasted

It is worth contrasting the knowledge organisation work in common use in the two sub-disciplines of archaeological work reviewed above, the study of archaeological artefacts, and the management of heritage inventories. The nature of the domain, with incomplete material typically open to challenge and re-interpretation makes archaeology a useful case study of the distinction between classification and indexing. Formal classification of archaeological material from individual site archives is undertaken to support analysis and develop new propositions about the past. A "good" classification will generally mean that each item in the set of material to be analysed can be assigned to one, and only one, class, and new material can be accommodated within the existing classification. Indexing of site-level inventory entries should reflect all the valid propositions made about a site to support the widest possible discovery and retrieval. A "good" index will provide many indexing terms to guide search, with new "candidate" terms added in response to changing indexing requirements. Both are necessary. You need to use indexing to find the information to include in your analysis. You need to use classification to analyse information to develop and propose new knowledge and present new narratives.

3.0 Developing narrative from archaeological information

3.1 Development-led archaeology

It is not, perhaps, commonly known that the majority of archaeological research in England takes place in the commercial, not the academic, sector. Current UK government planning policy for England and Wales places a duty on property developers and the construction industry to pay for the costs of archaeological investigation in advance of development. For some twenty-five years, this has been the major funding source for archaeological work, and has yielded an enormous body of detailed evidence, to complement the work of the academic and public sector research effort. The scale of work, combined with improved access to the results arising from improved knowledge organisation systems, has enabled synthesis of new narratives from this corpus (Historic England 2015). For example, the prehistoric population of England was larger, more mobile, more sophisticated technologically, and more connected to the continent than we previously had known.

3.2 Neolithic enclosures project

One particular project illustrates the potential. A class of monuments familiar in most areas of Britain are the

"causewayed enclosures" of the early Neolithic—the "New Stone Age" period when settled farming practices appear in England—around 4,800 BCE. The Monument Type thesaurus term "CAUSEWAYED ENCLOSURE" we apply to index these sites in inventories is simply descriptive—the scope note reads "A Neolithic monument comprising an irregularly circular enclosing ditch, interrupted by frequent causeways, and often accompanied by an internal bank, also causewayed." We know little, with certainty, about their function and significance to their builders. However, they are widely distributed within southern Britain and Ireland, which gives us some sense that whatever function they served was similarly widespread.

The date of construction of many of these sites had previously been measured using radiocarbon dating, which estimates the year of death of an organism based on the decay of the carbon-14 radioactive isotope contained in a sample—a seed, or a piece of bone, wood or other material. These dates are expressed as probability curves with typically wide error margins and low confidence ratings for any particular date. The Gathering Time project (Whittle 2011) reviewed the dates and applied Bayesian modelling to refine the results for the dates of construction of nearly forty of these sites.

The results were groundbreaking. We can now say that these sites were a late introduction into the Neolithic culture, appearing some three hundred years after the first signs of Neolithic culture. Moreover, construction dates of these types of site form a wave front, which moves from the south east of England, spreading out over the course of less than two hundred years, between the thirty-eighth century BCE to the mid to late thirty-sixth century, throughout the country. Some sites were only in use for decades. We can see, in what has until now been "prehistory," a real historical event—a narrative extracted from the information. The methodology used for this project is applicable to any large-scale synthesis of archaeological research data that uses probabilistic dating techniques to build chronology. We are starting to be able to move our narrative on from Thomsen's "three ages" to a situation where we can study the deep past in the same way that we discuss historical periods; we can talk of centuries and decades—human lifespans—rather than arbitrary period labels. The report notes how the new approach could change the existing perception of what is possible: "many archaeologists have been resigned to imprecise timescales, which affects the kinds of pasts they regularly construct" (Whittle 2011, ix).

3.3 Roman rural settlements project

The re-evaluation of existing information to generate narrative, such as that applied to the Neolithic enclosures archives, is now becoming a mainstream area of archaeological study. The research agenda published by Historic England includes a research topic on new narratives from big data (Historic England 2017). UK archaeology is starting to gain genuine benefit from the decades-long tradition of collecting and sharing information structured and managed with standardised knowledge organisation tools. Collaborations between the university sector and the data managers in local authorities and national agencies are developing as universities seek to demonstrate genuine research impact—real benefit to society, and local and national agencies look to make best use of declining public sector funding. A recent example is the Rural Settlement of Roman Britain project, funded by the Leverhulme Trust, and led by Reading University, 2012-2015 (Reading University 2014). This project has re-examined the reports from many years of developer funded archaeological investigations of Roman settlements to balance our understanding of the better studied urban centres. The information from some 3,600 reports available up to the end of 2014 has been consistently analysed to produce new datasets and is now yielding new insights into the first to fifth centuries CE. Significantly, the project has also produced methodological reviews, funded by Historic England, to identify lessons learned and make recommendations for future data collection and publishing from investigations. This will provide a road map for improvements in methodology, which include specific improvements in knowledge organisation and sharing in the sector—suggesting, for example, the need for a more consistent data format and coding of artefact analysis databases to support interoperability between the archives of different investigations.

4.0 Research frameworks and research questions

4.1 Discipline boundaries within archaeology

The potential for a significant increase in our knowledge of England's historical narrative arising from projects such as this is enormous. There are, however, significant hurdles to overcome to maximise co-ordination of the research into the archaeology of England. Within the sector, there are different sub-sectors with different motivations for their involvement in research. These differences can limit opportunities for co-ordination of knowledge and the works of synthesis that can be undertaken. In simple terms, the sector divides into four sub-domains:

- Archaeology.co.uk, the archaeological contractors. Companies need to make a profit to survive and need to price contract bids in line with that. This limits the opportunities for them to invest in co-ordination of research across commercial boundaries.
- Archaeology.ac.uk, the academic sector. University based research needs to score highly in research excellence ratings that influence research funding.
- Archaeology.gov.uk, the central and local government sector. Has a duty of care towards protecting archaeological sites. Needs to be able to demonstrate public benefit and operate within an environment of austerity.
- Archaeology.org.uk, the community or volunteer sector. Disconnected from the networks and processes available to the profession, supporting the application of good practice and standards.

4.2 Research frameworks

One practical tool used to improve co-ordination of the research effort is the suite of published Research Frameworks for archaeology, developed with national agency funding since the mid-1990s. Each publication covers the research interests relevant to either a particular region (for example the north west of England), or a particular time period (for example the Mesolithic period) or, in some cases, a particular artefact type (for example Medieval pottery). Less common examples relate to a specific industry (for example mining sites). The format varies, but typically consists of three principle sections, following a model first proposed in the mid-1990s (Olivier 1996):

- A synthesis—what is currently known
- An agenda—what we want to know
- A strategy or plan—how we are going to create that knowledge.

These research frameworks have been developed by specially commissioned teams of experts in the relevant subject matter, drawn from academia, the experience of commercial archaeologists, amenity societies and special interest groups. A survey by Historic England lists more than fifty different research frameworks in use in England. Most are available as hard-copy documents with some available in digital formats (Historic England 2015).

Their purpose is to guide the research element of new archaeological investigations, co-ordinating that new work with work that has happened previously to enhance public understanding. For example, under current government policy (the National Planning Policy Framework, section 12), if a plan is made to redevelop a site, the archaeological advisor to the local council has the power to require the

developer to undertake an investigation of the site as a condition of the grant of planning permission. The advisor will typically draw on information from their Historic Environment Record to assess what is already known about the archaeology of the site affected by the development and consult relevant research frameworks to see what the research priorities or objectives might be for the investigation. That will then form a brief to guide the developer in commissioning the investigation from an archaeological contractor, who sets out their methodology in a "Written Scheme of Investigation" or project design. The new information from that investigation should then feed back into the Historic Environment Record and possibly into the research framework. This is the research cycle that seeks to maximise the public benefit—the public knowledge gained—arising from the loss to the public of the physical remains of the archaeological site through development (Figure 1).

In practice, there are obstacles that interrupt this cycle. An Historic England review of the Research Frameworks (Pye Tait 2014) has confirmed their value as a research tool but also identified some specific issues with the application of the current suite of research frameworks. They:

- may not be known about
- may not be up-to-date
- may not be accessible.

These issues hamper the application of the research frameworks at the appropriate moment in the research cycle when a new project is designed. There is a marked contrast, therefore, between the traditionally structured and managed, hard-copy publication format of Research Frameworks and the highly visible, updateable and searchable Historic Environment Record databases.

4.3 Next generation research agendas: the use of research questions

To address these issues, and to lead development in England of a new generation of frameworks to co-ordinate research, Historic England has commissioned consultants Landward Research Ltd to undertake the research and development of a new interactive online platform to host new and revised research frameworks for the sector. We are drawing on parallel experience in the development of national archaeological research frameworks online in Scotland (the ScARF project, www.scottishheritagehub.com) and the Netherlands (Lauwerier 2017). Three projects to revise existing Research Frameworks, covering the north west, north east, and east regions of England will pilot this new platform. By making them available in one system, the intention is to gain for these documents some of

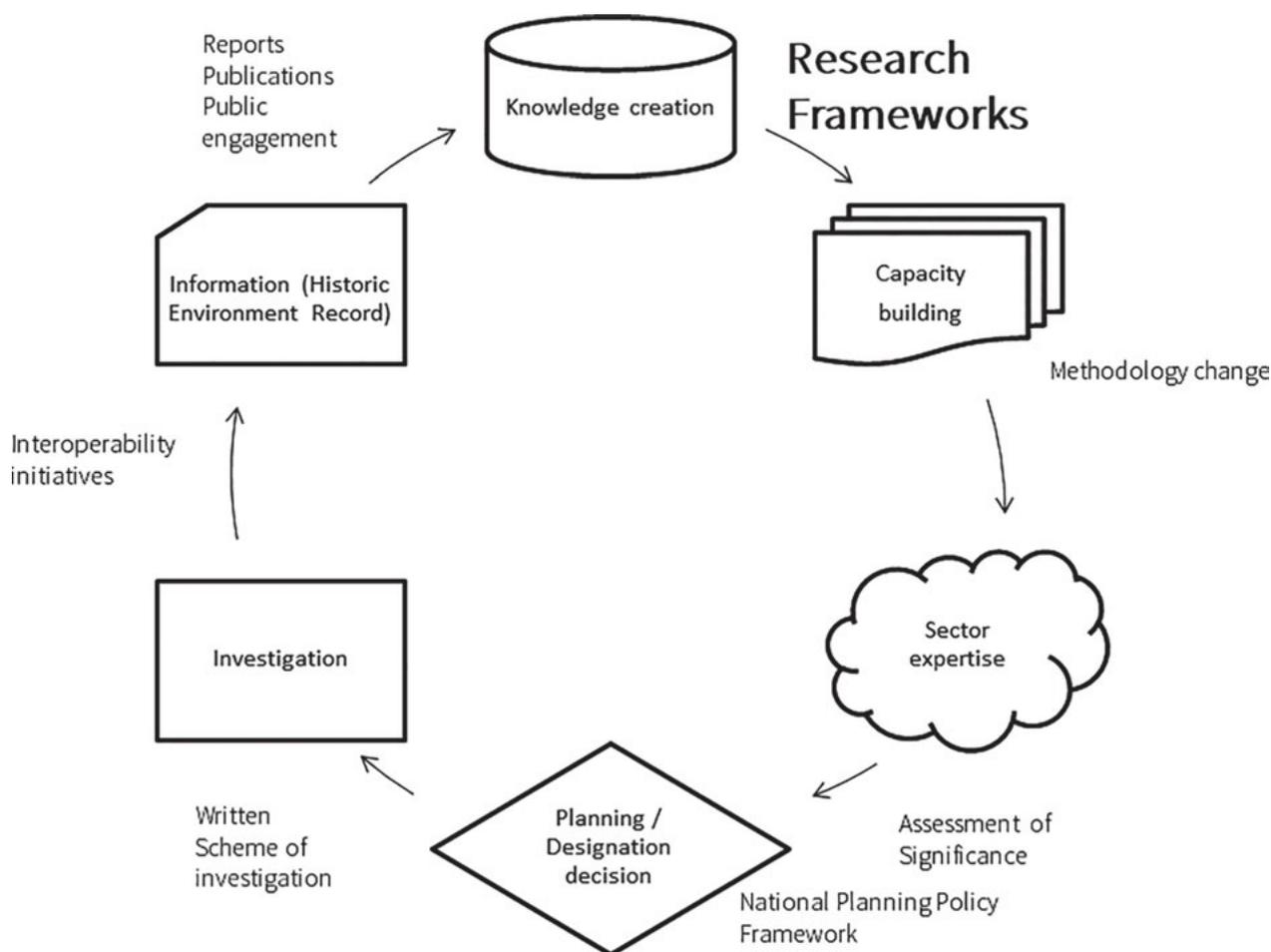


Figure 1. The role of Research Frameworks in the research cycle.

the added value of standardisation that has worked well for the inventory databases in the sector. If successful, their content will be more visible, more easily updated with new discoveries, and more easily searchable to assist in the design of new research. This paper is part of that research project.

The focus of research for this paper has been the research agenda component of the new generation Research Frameworks—what we want to know. The format for these agendas varies across the different existing Research Frameworks. Typically, they will consist of organised lists of topics that the expert panels of authors have proposed as in need further research. The format of these topics will be either phrased either as propositions or objectives for research (e.g., “We need to excavate more Roman pig stys”), or as research questions (e.g., “What was the scale of pig-production in the Roman period?”). The literature on which approach constitutes good practice in the design of research agendas is very limited. At the time of writing, the approach under investigation is to adopt the research question style for the future development of research agendas.

The reasons for selecting research questions as an approach are:

- Questions specifically identify a knowledge gap. This has benefits both for the authors, as a guide to constructing the agenda, and for potential researchers.
- Questions focus less on the method of research and more on the knowledge outcome.
- Questions can more easily lead to development of specific research objectives for a written scheme of investigation or project design (i.e., the filling of the knowledge gap) and so assist in assessing the eventual impact of that research (e.g., the practical application of that knowledge).
- Questions stimulate dialogue between those asking the question and those who may be in a position to answer it as research opportunities arise.
- Questions provide for a more open research culture and so may stimulate greater interest in contributing to and using the research agendas for example by community- or volunteer-led research.

- Questions are seen as a recognised starting point for good practice in research design in the other social sciences (e.g., Wyse 2016), offering the possibility of sharing learning in research processes.
- Questions have been the subject of automated text analysis research into "question and answer" systems (e.g., Tomuro and Lytinen 2001), opening up the possibility of further technical development of the platform.

The total dataset of research questions from the three pilot Research Frameworks in development is not as yet clear, but from the extent of the published documents can be estimated to be in the hundreds.

4.4 Towards a KOS for research questions

The development of an approach to the systematic recording of research questions from research frameworks in an indexable and retrievable format is a work in progress.

4.4.1 Organisation: an online book or a "pot of questions?"

The issue here is how to organise and present a large list of research questions. As traditional monograph publications, the existing Research Frameworks are structured as a monohierarchical classification of chapters, sections, sub-sections, and in some cases sub-sub-sections. The regionally based frameworks we are currently piloting tend to organise into period-based chapters with functional or site type themes at the next level down. For example, chapter three on the Roman period, section 3.1 on "military sites," sub-section 3.1.1 on research questions. Given that this is the structure of a regionally based framework, a national overview would look as follows:

Region
 Period
 Site Type
 Question

This approach has the advantages that authors are already familiar with such a structure and the ease of translating an existing monograph structure into a similar online book-like structure of pages.

The alternative is a looser approach—what Dan Miles from the project team has called a "pot of questions" from which relevant questions can be filtered. This approach is more similar to an inventory as discussed in part one. The embedded contextual meaning given by the position of a question in a monograph-like hierarchy could be replaced by metadata associated with the "question," and defined re-

lationships between the "question" and other entities in the Research Framework as follows:

Question
 Has relevance in "region" =
 Relates to a "class" of heritage assets
 Which "has date range" = AND "has site type" =
 Etc

This approach has more flexibility and makes it easier to add new questions as they arise. Questions could easily be shared between different research agendas by allowing multiple indexing with different values. The "regional research agenda" would not be a separate fixed information structure but would instead simply consist of a search result generated as needed and including those questions currently tagged with that "region name." This approach also has the advantage that Research Frameworks structured in other ways than around a "region" (for example by period or by site type) could more easily be included in a flexible structure.

4.4.2 Metadata and vocabulary standard to describe research questions

If the "pot of questions" approach is adopted, then the project will need to adopt and apply an appropriate metadata element set to support consistent coding for discovery and management. For example, "what is the question about" (i.e., what type of site and archaeological period or date range), "who is asking it," "who has approved its inclusion in the 'agenda,'" "when was it included" (to help manage the agenda), or "in what circumstances is it relevant" (for example, where in the country or what sort of research techniques might be necessary)? As discussed above, for some of these elements at least, there has been considerable investment in the modeling of the databases in the sector, so we anticipate that this issue will mainly consist of adopting relevant modeling and standards from existing sources such as the CIDOC-CRM. Specifically, a research question may be modeled as part of the existing CIDOC-CRM E29 design or procedure and Property P68 foresees use of (see <http://www.cidoc-crm.org/Entity/e29-design-or-procedure/version-6.1>), though this has yet to be tested. The use of existing vocabulary standards for relevant metadata elements has already been mandated by the project. Table 1 sets out a provisional metadata element set for capturing key information about a research question.

4.4.3 Geospatial indexing of questions— tying questions to places

A special case of metadata associated with a research question is the use of geospatial information to tie a research question or set of questions to a place—either at a small scale (say an individual building or site) or at a larger scale (an area, settlement, administrative region, etc.). This would have particular value in archaeological research, as the agenda could then have an associated GIS layer. In practice, this layer could form part of the map-base for an Historic Environment Record, such that known information about sites could be displayed alongside research questions that might be relevant should sites in the HER become the subject of planning applications. Clearly, this potentially has a large overhead in terms of preparing relevant geospatial indexing and maintaining its currency, but the potential is exciting.

4.4.4 Capturing non-specialist subject indexing

In addition to the use of specialist thesauri for site type, for example, there is a need to capture more general research subject areas (for example, "disease"). The selection of a relevant KOS will be important for integration with other online sources. An exciting possibility, developed for the BBC programme archive, is the use of Wikipedia subject categories (Kent 2016).

4.4.5 Modelling the character of the question asked

Beyond metadata to describe what the question is about, and the places it may be relevant, there are also intrinsic details of the question that may need to be captured to make the agenda as useful as possible. For example:

- Does the question have a broad or narrow focus? What "level" would it be in a hierarchy of questions? Broad questions may be too general to have useful application and may simply not be answerable. But they may serve to group and connect other questions. Specific questions may be too narrowly focused and never get asked if a relevant research opportunity does not arise. But they might be easier to apply if a relevant research opportunity does present itself.
- What category of question is it?—open ended or closed, qualitative in focus or quantitative. Tomuro and Lytinen have provided (Tomuro and Lytinen 2001) a categorization of question types which may be helpful. What, when, where, and how questions may be more susceptible to archaeological study, but should we exclude "why" questions?

4.4.6 Modelling the research question development process

Beyond simple capture and storage of research questions, the need to maintain the research agendas as living documents suggests the need for the design of the online platform to have embedded functionality to model aspects of the research process. Elements in this might include:

- The submission of candidate research questions and relevant metadata
- The adoption and issue of a candidate research question by a recognised authority
- The appointment or other formation that creates an authority recognised to manage the research questions
- The addition of comments on an adopted question, e.g., to flag new research results
- The modification, enhancement or replacement of an adopted research question by a recognised authority
- The recording of the application of a research question to the design of a specific research project or research event
- The proposition that a research question has been answered (and is now no longer needed)
- The decision by a recognised authority that that proposition is true (and the research question can be removed from the agenda)

Again, existing modeling for the research sector may well be applicable. The CRM-Inf extension to the CIDOC CRM ontology has been developed to model the process of argumentation in academic discourse with the purpose of "facilitating the management, integration, mediation, interchange and access to data about reasoning by a description of the semantic relationships between the premises, conclusions and activities of reasoning" (Stead et al. 2015).

4.4.7 Capturing metrics on usage of the new research agendas

Work on the connection of Research Frameworks to the research projects that address them is already in hand through a separate project to develop OASIS, a central register for archaeological investigations in the UK (Archaeology Data Service forthcoming). When completed, the latest version of OASIS will allow researchers to make a manual connection between their research project and the relevant Research Framework. In future, this might extend to the relevant research question. Either approach could provide an alert to the Research Frame-

	Question attribute	Examples	DCMI metadata element /term*
Identity	Question text	What is the distribution of "new forest" slipped ware?	Description
	Identifier	A uri or similar globally unique identifier for this question	Identifier
	Type of question	Quantitative	Type
Subject matter the question relates to	Paradigm (epistemological stance)	Processualist	
	What class or type of site does the question relate to?	Domestic	Subject (use of monument type thesaurus)**
	What class or type of artefact does the question relate to?	Pottery	Subject (use archaeological objects thesaurus)**
	What date of site does the question relate to?	Roman; 3 rd century; 4 th century; 200 - 400	Temporal (use periods list)**
	In what geographic areas is this question relevant?	North east; north west	Spatial (use of o.s. Linked data)
	What broad topic includes this question?	Economy; manufacture; trade; transport	Ispartof
Status	Who has asked the question?	Roman pottery research group	Creator
	Where has it been asked?	Http://potsherd.net/atlas/ware/nfcc.html	Source
	What is its status?	Open	
	Who has authority to change that status?	Roman pottery research group	Publisher
Impact	What contribution would answering the question make?	A better understanding of the localised distribution of this key high-quality fine ware at domestic (non-military) sites, a distance from its manufacturing site, would give insight into the adoption of roman material culture by the indigenous population.	
Method	Methodology—how should this be addressed?	All finds of this ware should be quantified using "estimated vessel equivalents" as a proportion of contemporary pottery finds and notified to the rprg.	Cidoc-crm e29 design or procedure ***
Management	Date accepted	25-sep-2017	Dateaccepted
	Next review date	01-jan-2020	Valid

Table 1. A draft metadata element set for research questions in archaeology.

*See <http://dublincore.org/documents/dcmi-terms/> for definition.

** See <http://heritage-standards.org.uk/fish-vocabularies/>.

*** See <http://www.cidoc-crm.org/Entity/e29-design-or-procedure/version-6.1>.

work managers and others with an interest in a particular research topic.

4.5 Future research questions

The development of an online interactive research question dataset for archaeological research, as a part of the new generation of Research Frameworks, raises further research questions for the future. These are listed here as a contribution to a future research agenda for knowledge

organisation in archaeology and perhaps in other knowledge organisation contexts. Can we extract new research questions from an automated analysis of the text of published research reports or other information sources? Tomuro's work suggests that there is much that can be done here. Can a dataset of research questions be used to improve the results of search algorithms to help us find the answers in new research? Several thousand new reports of archaeological discoveries in the UK are issued each year (Bournemouth University 2010). To keep pace,

and to build a research infrastructure that can respond at scale and speed to that level of work, we need to make the best use of knowledge organisation that we can.

5.0 Conclusion

This paper has been prepared at an early stage in the development of a new generation of archaeological research frameworks. However, it is clear that there is a long tradition of effective knowledge organisation in the archaeology domain which may be a useful case study for knowledge organisation in other domains. There are clear social benefits to be gained in terms of new narratives of our shared past from addressing the barriers to collaborative research in this domain, and the knowledge organisation tools that already exist provide a sound basis for those collaborations, such as the shared approach to development of a research agenda outlined here. We are at the start of our journey.

References

- Archaeology Data Service. 2018. *Functional Specification for the OASIS System*. York: University of York, forthcoming.
- Booker, Christopher. 2005. *The Seven Basic Plots: Why We Tell Stories*. London: Bloomsbury.
- Bournemouth University. 2010. "Archaeological Investigations Project." Last modified March 4. <https://research.bournemouth.ac.uk/2010/03/archaeology-investigations-project/>
- Carlisle, Philip, and Edmund Lee. 2016 "Recording the Past: Heritage Inventories in England." *Journal of Cultural Heritage Management and Sustainable Development* 6: 128-37.
- Historic England. 2015a. "Building the Future, Transforming our Past: Celebrating Development-Led Archaeology in England, 1990-2015." London: Historic England. <https://content.historicengland.org.uk/images-books/publications/building-the-future-transforming-our-past/building-future-transforming-past.pdf/>
- Historic England. 2015b. "Research Frameworks." <https://historicengland.org.uk/research/support-and-collaboration/research-resources/research-frameworks/>
- Historic England. 2017. Our Research Strategy and Agenda. Website for Historic England. <http://www.historicengland.org.uk/research/agenda>
- Kent Jo. 2016. "ADA: Automated Data Architecture; Creating User Journeys Through Content Using Linked Data." In *The Semantic Web: ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29–June 2, 2016; Revised Selected Papers*, ed. Harald Sack, Giuseppe Rizzo, Nadine Steinmetz, Dunja Mladenić, Sören Auer and Christoph Lange. Lecture Notes in Computer Science 9989. Cham: Springer, 118-22. doi:10.1007/978-3-319-47602-5_24
- Lauwerier, R.C.G.M., M.C. Eerden, B.J. Groenewoudt, M.A. Lascaris, E. Rensink, B.I. Smit, B.P. Speleers and J. van Doesburg, eds. 2017. *Knowledge for Informed Choices: Tools for More Effective and Efficient Selection of Valuable Archaeology in The Netherlands*. Nederlandse Archeologische Rapporten 55. Amersfoort: Rijksdienst voor het Cultureel Erfgoed. <https://archeologiein-nederland.nl/publicaties/knowledge-for-informed-choices>
- Lee, E. 2004. "Building Interoperability for United Kingdom Heritage Inventories." In *Research and Advanced Technology for Digital Libraries: 8th European Conference, ECDL 2004, Bath, England, September 12-17, 2004; Proceedings*, ed. Rachel Heery and Liz Lyon. Lecture Notes in Computer Science 3232. Berlin: Springer, 179-185. doi:10.1108/JCHMSD-02-2016-0013
- Olivier, Adrian. 1996. *Frameworks for Our Past: A Review of Research Frameworks, Strategies and Perceptions*. London: English Heritage.
- Orton, Clive. 1982. *Mathematics in Archaeology*. Cambridge: Cambridge University Press.
- Pye Tait Consulting. 2014. "Review of Research Frameworks." London: Historic England. <https://content.historicengland.org.uk/content/docs/research/review-research-frameworks-historic-environment-sector-england.pdf>
- Reading University. 2014 "The Rural Settlement of Roman Britain." <http://www.reading.ac.uk/archaeology/research/roman-rural-settlement/>
- Stead, Stephen, Martin Doerr, and others. 2015 CRM_{inf}: The Argumentation Model; An Extension of CIDOC-CRM to Support Argumentation. Version 0.7. <http://www.ics.forth.gr/isl/CRMext/CRMinf/docs/CRMinf-0.7.pdf>
- Terminology FISH. n.d. *Introduction to Controlled Vocabulary: The FISH Thesauri and Candidate Terms*. Forum on Information Standards in Heritage. <http://heritage-standards.org.uk/terminology/>
- Tomuro, Norio and Steven L. Lytinen. 2001. "Selecting Features for Paraphrasing Question Sentences." In *Proceedings of the workshop on Automatic Paraphrasing at the Natural Language Processing Pacific Rim Symposium (NLP RS 2001), Tokyo, Japan*. <http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.5.7963>
- University of South Wales, Hypermedia Research Unit. "Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources." University of South Wales. <http://hypermedia.research.southwales.ac.uk/kos/STELLAR/>

Whittle, Alasdair, Frances Healy and Alex Bayliss. 2011. *Gathering Time: Dating the Early Neolithic Enclosures of Southern Britain and Ireland*. Oxford: Oxbow Books.

Wyse, Dominic, Neil Selwyn, Emma Smith and Larry E. Suter, eds. 2016. *The BERA/SAGE Handbook of Educational Research*. Los Angeles: Sage.