Lost in information? Ways of knowing and modes of representation in e-archaeology

Jeremy Huggett
Archaeology, School of Humanities, University of Glasgow

Biographical note

Jeremy Huggett is a Senior Lecturer in Archaeology at the University of Glasgow, Scotland. His research has focused on the nature, development, impact and implications of information technologies in relation to archaeology, and their effects on our understanding of the past.

Abstract

This paper is concerned with the development of new cyberinfrastructures which are being constructed on the foundations of Web 2.0 looking forward to Web 3.0, employing semantic techniques for modelling archaeological information in order to enhance the location, retrieval and use of data. It argues that, as these techniques emerge, it is important to consider their origins, their underlying models, and their consequent effects on the subject. Representations of archaeological information are predicated upon new and existing data standards and constitute additional layering or nesting of standards within each other. The paper identifies the need for an ethnographic study of the creation of these new technologies since they present the potential for radical change within the subject but presently lack proper evaluation.

Keywords

cyberinfrastructure; standards; standardisation; ontology; ethnography

Acknowledgements

I should like to thank Kath Baker, the Editor, and an anonymous referee for their helpful comments on earlier drafts of this paper. Errors and misconceptions are, naturally, my own.

> *"... a classified and hierarchically ordered set of pluralities, of variants, has none of the sting of the miscellaneous and uncoordinated plurals of our actual world." (Dewey 1925: 49).*
>
> *"... when you design those protocols you are designing the space in which society will evolve. You are designing, constraining the society which can exist for the next 10-20 years." (Berners-Lee 1997).*

Archaeology in the Information Age

By common consent, we live within the Information Age as members of an Information Society (see Castells 1996, for example). One of the key characteristics of this is the increasing flows of information that we are subjected to, leading to concerns about a perceived information overload.

For example, Koski's 'infoglut' (2001) is seen to give rise to an intellectual and social incapacity to deal with the volume of information available (Burgelman 2000). Weinberger characterises the computer technologies developed over the past fifty years for handling information as "knowing-by-reducing", effectively operating by exclusion at the risk of omitting things that were worthy of consideration, whereas with increased access to and reliance on networked information we are moving to a situation of "knowing-by-including" (Weinberger 2011: 9) with different kinds of filters which are designed to bring relevant information to our attention while pushing everything else into the background. This focus on the issues surrounding how people access, use, interact, and handle information in a wired world is particularly significant now, at a transition point between the two-way interactivity of Web 2.0 and the emergence of the so-called Web 3.0 with its use of semantic technologies, linked and 'big' data, amongst other things (for example, Kroeker 2010).

This paper is concerned with the implications of these developments for archaeological knowledge and argues that as archaeological cyberinfrastructures begin to emerge it is important to consider their origins, underlying models, and the consequent effects on the subject. It reviews the nature of data standards in relation to content standards, documentation standards and ontologies, and argues that their embedded nature requires examination in the context of these new information infrastructures. Amongst other things, the development of archaeological e-Science or cyberinfrastructure tools is predicated upon a much closer association with computer science than has perhaps been the case in the past (see Schroeder 2008) and a resurrection of techniques such as artificial intelligence alongside the new tools for networking and collaboration which in combination may present the possibility of radical innovation and/or change. This has fundamental ramifications for the nature of the subject.

The nature of archaeological data

In part, the significance of these developments will depend on how we approach the nature of archaeological data (for a recent discussion, see Lucas 2012). At one extreme, questioning these issues may seem relatively unimportant from a culture-historical/processual perspective which sees the archaeological record as objective, static, knowable, where facts are given, not made. On the other hand, if information about the past is seen as situated, contingent and incomplete, data as theory-laden, and relationships as constantly changing depending on context (for example, Jones 2002: 3-22), then the construction of archaeological cyberinfrastructures may seem more problematic. That said, whichever view of archaeology is subscribed to is, in some respects, irrelevant - when constructing systems which model and support our ways of knowing the past it is safer to assume that data are contingent than to assume they are not. A topic of debate over years (Patrik 1985, Binford 1987, Barrett 1988, Hodder 1999, amongst many others), this is not peculiar to archaeology: more generally it is understood that knowledge is situated, distributed, and social (for example, Van House 2003: 271):

- The situated nature of data arises from their creation by specific people, under specific conditions, for specific purposes. This leads inevitably to data diversity (Kansa and Bissell 2010: 43) and is often seen as a characteristic of archaeology by outside observers. For example, recent surveys have suggested that archaeology is unusual in comparison with other humanities subjects in terms of the variety and complexity of its data and sources, but note that data collection, management, sharing, and preservation practices are not

especially standardised (Harley *et al*. 2010: 31, Babeu 2011: 66, Khazraee and Khoo 2011: 377).

- The distributed nature of data arises through their creation by multiple individuals and by a general desire to share data - indeed, a responsibility to do so is laid down by archaeological professional organisations. This gives rise to cooperation amongst archaeologists and others who may be separated spatially and temporally.

- The social nature of data emphasises the community of practice which creates, discusses, shares, and debates them. It also agrees what is considered to be representative and trustworthy.

These characteristics frequently give rise to competing, contradictory, demands. For example, there is a tension between data diversity and a desire to share that data. Data collection by different individuals tends to lead to diversity unless control is exerted; controls over data collectors may limit what is recorded; what is selected for record will depend on what is agreed as being representative; what is considered to be representative will depend upon the contemporary theoretical environment. A perspective of data as unproblematic and knowable will tend towards control of recording, whereas a view of data as situated and contingent will tend to support data diversity. These and similar assertions are all capable of dispute, but this complex of relationships and the challenges they represent are subsumed within a digital environment through the application of data standards.

Standardised representations

Standardisation has been contested in the past. In an early example of the debate within computer archaeology, Cooper argued for the centralisation of control and standardisation of recording (1985: 86-89) while Lock and Spicer argued that data standardisation should be situated at the point of data exchange, not recording (1985: 109). Reilly recognised the benefits of comparability between recording systems but argued that standardised datasets would constrain future research (1985: 68-69). Richards considered standardisation of recording to be inappropriate and potentially damaging to the development of archaeology (1985: 97-98). The same volume also carried an early critique of the potential of artificial intelligence within archaeology, which concluded that the widespread adoption of such techniques and the level of standardisation they incorporated would constrain the development of the subject (Huggett 1985: 139). Since then, the standardisation debate within computer archaeology has largely ceased with the exception of a handful of interventions (for example Huggett 1995, 2004, Richards 2009), although it has continued elsewhere (for instance, Witmore 2009, Lucas 2012). This has much to do with the ubiquity of standards: they apply from the computers we use to the databases we populate and the internet we inhabit. Without standards, much of what we take for granted would cease to work. As a result, many of these standards are largely invisible, unconsidered at least until they stop working or have to change (for example, the impending switch from IPv4 to IPv6 as available internet addresses run out). Most of us do not know, and do not need to know, about the operation of these kinds of standards: however, standards which impact within archaeology, whether generated externally or internally, ought to be of interest to archaeologists, all the more so if they have become so customary as to be largely invisible.

Leaving aside standards associated with technical issues (file formats, communication protocols, software standards etc. which are largely outwith our control), Richards identifies two other types of standards (2009: 28), to which a further category can be added:

1. Data content standards: the representation of a unit of information, typically associated with a data table or its equivalent. Formal - and frequently informal - standards may apply in relation to the nature of the item in question, whether an artefact, context, site, or a component part of these. Informal standards may simply arise from the realisation through bitter experience of the need for consistency in order to be able to retrieve and analyse data with confidence, but formal standards exist as wordlists which provide approved terms and as thesauri which define terms alongside their alternative forms (for examples, see English Heritage 1999, RCAHMS 2004). Formal and informal standards permeate all aspects of data content, although there is a significant difference between, say, measurements being taken in standard metric units and a standard defining which measurements should be taken. Importantly, data standards list "what information should be recorded and how it should be recorded, *to meet a particular objective*" (Lee 2007: 8, my emphasis). From this we may infer that data which are not required to meet that objective are unlikely to be captured; data are recorded for a specific, not a universal, purpose.

2. Data documentation standards: metadata for resource discovery. Metadata standards define what data are, not the data themselves; they are data about data and consequently operate at a higher level than data content standards (Miller 1996, Wise and Miller 1997). Typically, metadata provides information about aspects such as the source of the data, what it represents, its method of collection, its scale, and its location. It is capable of, but generally not used for, describing the meaning of attributes within tables, their properties and intended use, along with other detailed descriptors of the data themselves. The Dublin Core metadata standard is widely applied and consists of a set of 15 elements which are used to describe a resource at a generic level (DCMI Usage Board 2010). It provides the basis of, for example, the UK Archaeology Data Service catalogue (Richards 2002). Metadata facilitate the location and co-location of datasets, but have no knowledge of interrelationships between the datasets beyond the metadata terms they hold in common. This high-level cataloguing of metadata gives an impression of 'light-touch' standards, which consequently have little impact on the data themselves. However, resource discovery will be on the basis of a restricted set of criteria - typically the 'what', 'where', 'when' triad of questions - and hence impose limits on the kind of research questions that can be directly addressed.

3. Ontology standards: an ontology is "a systematic representation of all categories of objects or concepts that exist in the specified domain and the relationship between them" (Richards 2006: 976). Ontologies provide a shared and common understanding which allows the mapping of concepts to information (Cripps *et al.* 2004: 3). They are designed to produce semantic interoperability between different datasets by representing the information contained within them using a formal standardised, structured syntax for describing and capturing the concepts and contexts of information, and the relationships between them; a methodology akin to that used in artificial intelligence. This interoperability between disparate datasets linked through shared ontologies is what constitutes the 'linked data'

> which underpins the creation of the Semantic Web (see Berners-Lee *et al* 2001) although ontologies are said to be capable of much more: knowledge acquisition, reasoning, and hypothesis generation and verification, for example (Khazraee and Khoo 2011: 376).

All three areas involve creating and managing a complex set of socio-technical relationships - from defining the basic technical requirements through to the relationships between the different individuals and organisations involved. Furthermore, all present user-dependent, partial accounts of the world.

It is difficult to argue against the need for standards: one need only come up against a large chaotic set of data to realise that some degree of standardisation has its place. For example, a recent research project sought to merge data derived from three organisations, each of which was recording the same essential evidence, and each of which operated multiple field teams undertaking the recording (Roberts 2011). The data consisted of large numbers of tables, some redundant, with fields which were inadequately controlled. These allowed multiple entries to be made in single fields and inconsistencies between information in different fields within the same record. There was also duplicated data and data entered in the wrong fields (Roberts 2011, 139). The data cleaning exercise was such that, in the end, it would have been easier to have re-entered the data from scratch had the primary data been available (Roberts 2011, 147). Even where standards exist to a greater or lesser extent, not everyone necessarily interprets them in the same way, especially where those standards are relatively informal. This gives rise to problems in aligning data from derived from different groups or collected across time and space (for example, Bateman and Jeffrey 2011). Standards are therefore a prerequisite for sharing or linking data, and consequently the early debates about standardisation in archaeology died down as facilities for sharing became available and the benefits began to be demonstrated.

There has been considerable investment in data standards within archaeology, as evidenced in the various national thesauri for monument types, object types, and the host of wordlists categorised by INSCRIPTION for use within inventories (FISH 2004), for example. Indeed, this is one of the reasons why archaeology has been seen as suitable as a test-bed for the development of ontologies and associated tools which underlie the Semantic Web (for example, Jeffrey *et al*. 2009: 2515, Richards *et al*. 2011: 42). Taken individually, standards may seem relatively unproblematic, although they are not uncontested: for example, Baines and Brophy (2006) question the loss of subtlety of terminology and meaning within monument thesauri while recognising their value. Likewise, standardisation still sits uncomfortably with a situated and contingent view of archaeological information, and in some respects it recalls the criticism of GIS applications as being rooted in essentially positivist approaches.

However, the three-way characterisation outlined above also implicitly reveals the embedded nature of standards. Formal standards for naming and categorising entities within databases are incorporated into metadata documentation, while both formal data standards and metadata standards are captured in ontologies which also are capable of formalising hitherto informal data according to specified rules. In other words, standards are built using other standards, standards are integrated with other standards, standards are superseded by other standards, and in the process a complex and ill-defined web of relationships and dependencies are created which have the potential to constrain the discovery and subsequent analysis of information resources.

This therefore is not so much an argument that standards and standardisation are good or bad - it is to argue that they are not neutral, and failure to recognise this may give rise to unintended and even undesirable consequences. As a result, it is important to know how the standards were derived, who is/was responsible for them, the issues encountered during their creation, the way those issues were resolved (the arguments behind the scenes, the areas of conflict and their resolution), how the standards were tested and validated, how they are maintained , and so on, in order to begin properly to understand the nature and context of standards at all levels, and the extent to which they shape our knowledge of the past. This kind of study is not evident in the archaeological literature, raising the suspicion of a kind of drift in which initial concerns about standardisation have been increasingly left to one side as the technological tools seduce us with new ways of looking at data, but the operation of these tools is predicated on the use of standards. The development of the Semantic Web or Web 3.0 continues the pattern: critiques of artificial intelligence techniques in the 1980s are left to one side as rule-based knowledge engineering and information extraction come back to the fore. Data mining and natural language processing techniques may facilitate the semantic integration of diverse datasets (for example, May *et al.* 2010, Tudhope *et al.* 2011) and alter the point at which standards need to be applied, but standards remain a prerequisite for mapping the different entities while the use of automated schema introduce additional questions.

The standardisation process

Not standing back and taking a critical view of standards and standardisation is not specific to archaeology: standards more generally are seen to be an underdeveloped area of study (for example, Brunsson and Jacobsson 2002: 7). Standards and standardisation are seen as ubiquitous but underestimated phenomena (Timmermans and Epstein 2010: 70). Star and Lampland (2009: 10-11) suggest a series of reasons why standardisation (whether designed to streamline procedures or regulate behaviours or require specific results) is rarely questioned:

- Standardisation has come to be understood as a valuable and necessary process.
- Standardisation is considered to be necessary to facilitate other tasks which means that standards are encountered by most people as fully developed forms.
- Standards are ahistorical and have become a taken-for-granted, often completely embedded in everyday tools. Categories such as 'none of the above', 'unclassified' and 'miscellaneous' reveal the boundaries of such largely invisible standards.
- Standards are boring: their embeddedness and virtual invisibility mean that even if they are recognised, they are not seen as being of central importance. Studying a standard has been likened to having the entertainment value of reading the phone book (Bowker and Star 1999: 321).

These may not all be equally characteristic of archaeological standards, and archaeology introduces at least one other reason why standards may avoid scrutiny: the adoption in some quarters of an approach to the subject which finds the formalisation of archaeological knowledge to be rational and desirable. For example, Barceló refers to "absurd prejudices and the weight of individual authority" (2009: 103) in a rejection of the critique of artificial intelligence in archaeology by archaeologists who had accused "the most promising computational techniques ... of excessive simplification, of

forcing knowledge, or distorting it, and of failing to exploit fully the knowledge of the expert" (Barceló 2009: 103). This underlines the importance of considering the background, creation, development, and application of standards - and at the same time establishing whether and where prejudice, individual authority, or other factors might play a role.

To reinforce this argument, we could consider the challenges for the construction of classification schemas posed by Bowker and Star, which they characterise in terms of comparability, visibility, and control (1999: 231-232).

- Comparability refers to the extent to which terms can be defined consistently and shared unproblematically. This may be less of an issue on a local scale but becomes challenging when the objective is to construct a standard which is capable of catering for multiple communities of practice across both space and time. Standards, at whatever level they refer, represent a level of abstraction and consequently may not be equally appropriate for all audiences and purposes.

- Visibility is concerned with what is and is not included within a standard. Since wordlists, thesauri, metadata standards, and ontologies are abstractions, they have to omit terms, categories and, in the case of ontologies, concepts. This fixes their level of conceptualisation and anything which does not fit within that abstraction becomes essentially invisible, unclassifiable by definition. In terms of resource discovery or ontological interoperability, these invisible aspects disappear and may as well not exist. Many of these invisible elements may relate to embedded, tacit, or local knowledge which are less susceptible to codification and hence transfer within a digital environment, but may nevertheless retain archaeological significance. Bowker and Star also point out that invisibility need not only be through erasure: it may arise through intimacy with the domain such that what should be explicit remains implicit because of the familiarity of those involved (1999: 232).

- Control relates to the degree of prescription that a scheme imposes on its users. At one level, archaeology is traditionally permissive in terms of the application of standards: for example, the Institute for Archaeologists in the UK simply require professionals to comply with specified data standards (for example, see IfA 2008a: 3.37, 2008b: 3.38), while the Archaeological Institute of America makes no direct reference to data standards (AIA 2008). Although specific data standards are not prescribed they are, in a sense, surreptitious: if standards are not conformed with appropriately, data may not be as discoverable or reusable as they might otherwise be. The reverse is equally true: if the database, catalogue, search tool, or reasoning engine relies on a particular terminology in relation to, say, artefact types, then an enquiry using terms which do not conform to that standard (even if they are derived from another) may not generate reliable results, if they generate any at all.

These considerations represent a complex series of decisions taken during the process of defining and establishing any standard. However, reconstructing that series of decisions after the effect is not straightforward: standards cannot easily be reverse-engineered since the codification of practice and knowledge within a standard is not necessarily captured within the standard itself. For example, Thévenot warns that

> "the new invested form is established by sacrificing other possible forms of equivalency. Yet, once invested in and immobilized, a form cannot conserve the traces of its own past if it is to operate effectively as the new principle of equivalence. Should they appear, such traces would insinuate doubt. Any recollection of the processes through which the convention was established would most certainly reopen anxieties about its initial arbitrariness." (2009: 795).

Consequently, he argues that doubts internal to the standardisation process will be forgotten because they would undermine the standard itself (2009: 797). This emphasises the need for the process of creation to be captured in order to allow a subsequent deconstruction of the decisions behind the definitions and structures which have been employed in stabilising knowledge, freezing actions, deleting outliers and residuals and preparing it for use by non-experts (Star and Lampland 2009: 13). In the archaeological context, it is certainly the case that the majority of publications regarding the creation and application of standards tend towards technical descriptions with little reference to the immediate ethical, political, and professional values and decisions along with other socio-cultural issues operating on the actors at the time. Standards may not appear fully-formed, but their birth pains are frequently glossed over. This means that the whole lifecycle of a standard needs to be investigated.

An ethnography of e-ontology

Relatively little attention has been paid to the detailed consideration of process in the creation of archaeological standards, and, where detailed discussions with user-groups, communities, and others have taken place, this is poorly reflected within the literature. For example, a workshop was held in 1997 by the Archaeology Data Service which sought to investigate issues relating to the potential of the Dublin Core metadata schema. The subsequent report is a highly detailed account but its dry academic language disguises the nature of the debates which took place at the time - issues are reported but the specifics of the discussions are lost (Miller and Wise 1997). The draft report remains available which allows comparison with the final report, but a detailed examination of the considerations and concerns expressed by individuals at the workshop remains out of reach. This is not unusual. Discussion of the Archaeotools facetted classification and natural language information extraction remains at a high level with little reference to detailed decision-making processes underlying the development of the tools (see Jeffrey *et al*. 2009, Richards *et al*. 2011). Similarly, the plethora of publications concerning applications of the CIDOC Conceptual Reference Model from initial development through to ISO Standard focus on the technical and rarely reveal aspects of the range of issues and negotiations behind the standard (although working drafts, reports and minutes of meetings are available on the CIDOC website). Occasionally some of the underlying human issues do emerge from the published technical discussions. For example, the team working with the CIDOC CRM within the English Heritage Centre for Archaeology noted the difficulty experienced by archaeologists in discussing their domain of expertise in CRM terms, the solution being to hide the complexity of the ontology and discuss it in natural everyday language (Cripps *et al*. 2004: 25-26). Subsequent papers on the CRM-EH are essentially technical, but also raise similar issues: for example, the steep learning curve required for the domain experts to engage with the ontology meant that there was a strong preference for the ontology experts to map the archaeological data rather than the archaeologists who provided the data (May *et al*. 2010: 231).

To expect publications such as those cited above to include the level of detail suggested here is perhaps unfair and certainly unrealistic. What is needed, therefore, is an ethnographic study of archaeologists undertaking the construction of ontologies, given the present emphasis on these for the next generation of web tools. Ethnography has been successfully applied in science and technology studies in general and has been attempted in one or two other disciplines in relation to ontology: for example, cell biology (Randall *et al*. 2011), ecology (Baker and Millerand 2010, Millerand and Bowker 2009), and the geosciences (Ribes and Bowker 2009, Ribes and Finholt 2009). An archaeological study has been undertaken but this largely entails a study of archaeological reasoning within an institution rather than following the construction of a specific ontology (Khazraee and Khoo 2011:380). Some preliminary findings and emerging themes are reported: for example, the heterogeneity of datasets, the way archaeological understanding is arrived at, the tension between codification and interpretation of data, and the narrative nature of archaeology (Khazraee and Khoo 2011:381-383). They identify a need for what they call 'practice-based ontologies' which allow archaeologists to create their own ontologies in the process of their work based on their requirements and perspectives, and which can sit alongside the more general ontologies primarily used for information organisation and interoperability (Khazraee and Khoo 2011:384). An ethnography of an ontology seeks to follow its development from initial concept, through its development and implementation, tracing the issues which arise during the process of creation and use, and the reactions and responses of those collaborating on the construction and those subsequently putting the results into use. The outcome of such a study would be a greater understanding and appreciation of the decisions, policies and strategies that were incorporated during the creation of the ontologies, and the consequent implications for their application and use.

The kinds of questions which might be addressed would include:

- How is the scope of the work established? Who is the ontology intended for, and what is its purpose? Who determines these? Is it designed to fit a gap in existing ontologies, to replace existing ontologies, or to add to existing ontologies by modelling information in a different way? Who sets the agenda (Kansa 2011: 21-22)? Who funds it? Who facilitates it?

- Who is involved in the construction and at what point? The Dublin Core workshop (Miller and Wise 1997) and the discussion of the ontological modelling of the work of English Heritage's Centre for Archaeology (Cripps *et al*. 2004: 5) are rare examples where the membership of the domain expert group is identified. In other cases, no domain experts other than the project personnel are involved. A reluctance of people to become involved because of the boring nature of the work is frequently a characteristic (Randall *et al* 2011: 223), and indeed, some resistance to the idea of an ontology was reported at the Centre for Archaeology (Cripps *et al*. 2004: 28). Limp goes so far as to propose that "the reward structure in archaeological scholarship provides a powerful disincentive for participation in the development of semantic interoperability and, instead, privileges the individual to develop and defend individual terms/structures and categories" (2010: 278), a claim that would be worth pursuing. When people should be involved is also a potential issue in terms of the expectations of both the builders and the communities: in the case of the CRM-EH, the domain experts were brought into the process once the project members had prepared the basic model for discussion, but even then there were difficulties experienced with terminology and presentation (Cripps *et al*. 2004: 25-26).

- How are the formal characterisations of the attributes of entities arrived at, and how are these related to their archaeological meaning? Indeed, is it desirable or even possible to seek to include 'everything' or is the focus most appropriately placed on a restricted domain? What is included, what is left out, and why? What are the strengths and weaknesses of the representations used, and are these imposed by the model or already inherent in the data?

- How is the balance between a desire for simplicity and a need for complexity arrived at? There are different strategies for constructing ontologies, different software tools to assist in the process, and there are trade-offs to be had between the extremes of keeping it simple and producing it quickly, or seeking sufficient complexity to provide a greater sense of faithfulness to the original. Different approaches are already evident within archaeology, ranging from crowd-sourced tagging of images in museums (for example, Boast and Biehl 2011) through to formal ontological modelling of excavation data (for example, May *et al*. 2010) and automated information extraction using ontological tools (see Jeffrey *et al*. 2009, Richards *et al*. 2011, Byrne and Klein 2010). In what circumstances is one or other methodology seen as being more appropriate?

- How well does the model represent 'reality'? This will be related to the simplicity/complexity trade-off, but also concerns the flexibility of the model - whose reality is represented and can multiple representations be catered for? Some aspects of modelling can be quantified to an extent: for example, the Archaeotools project reports on the proportion of grey literature reports in their sample for which 'what/when/where' terms could not be extracted, although they note that this does not guarantee accuracy of those that were (Richards *et al*. 2011: 48). Similarly, the levels of accuracy of automated extraction of various event relations from the National Monuments Record of Scotland are reported (Byrne and Klein 2010: 53-55). Where numeric measures are not available, tag clouds have been used to represent the terms that have been employed in the automatic classification of a report (Richards *et al*. 2011: 48-49, fig 1.4) and hence provide a visual evaluation of their validity. However, these approaches focus on the validity of an end product which is by definition a restricted representation of the original data themselves.

- What are the potential implications of the eventual outcome for the subject, and are these considered at all? What might be constrained? What might be enhanced? And at what cost?

Conclusion

Star and Lampland observe that "Perhaps the most intriguing aspect of standards is their always already incomplete and inadequate (compared to some ideal) character ... The push to standardize presumes the ability to constrain a phenomenon within a particular set of dimensions, as well as the ability to dictate behavior to achieve the narrowly defined dimensions that stipulate its outcome" (2009: 14). The narrowing of dimensions is compounded by the nesting of standards within standards, and consequently a detailed examination of how these are constructed and used is needed to appreciate their impacts on archaeological knowledge, while recognising the important role that standards play in facilitating the interoperability and accessibility of data which are already transforming the subject.

This paper has argued for the need for research into the structures which are being created for representing, locating, and retrieving archaeological data. Although these methods are based on standards at various levels and of various kinds, they remain largely uninvestigated and unchallenged. Since we increasingly construct our images of the past through a technological lens, an archaeology of the cyberinfrastructures we are constructing, including the range of interrelated and interdependent technical, organisational, and social aspects, is of vital importance. In the process it will support the creation of an understanding of the extent to which new ways of knowing within a wired world affect archaeological processes and outcomes.

To paraphrase Berners-Lee (1997) quoted at the outset of this paper, we are currently engaged in designing the spaces in which archaeology will evolve, and which will shape the subject for years to come. We should do so in full knowledge of the consequences of the strategies and decisions being taken by us and for us. To do otherwise would fail to experience the full strength of Dewey's sting (1925: 49).

References

AIA 2008. *Archaeological Institute of America Code of Professional Standards*. Boston: AIA. http://www.archaeological.org/pdfs/AIA_Code_of_Professional_StandardsA5S.pdf.

Babeu, A. 2011. *"Rome Wasn't Digitized in a Day": Building a Cyberinfrastructure for Digital Classicists*. Washington, DC.: Council on Library and Information Resources Publication No. 150. http://www.clir.org/pubs/abstract/pub150abst.html.

Baker, K. and Millerand, F. 2010. Infrastructuring Ecology: challenges in achieving data sharing. In *Collaboration in the New Life Sciences* (eds. J. Parker, N. Vermeulen and B. Penders). Farnham: Ashgate Publishing Ltd., pp. 111-138.

Baines, A. and Brophy, K. 2006. What's another word for thesaurus? Data standards and classifying the past. In *Digital Archaeology: Bridging Method and Theory* (eds. T. Evans and P. Daly). Abingdon: Routledge, pp.236-250.

Barceló, J. 2009. The birth and historical development of computational intelligence applications in archaeology. *Archeologia e Calcolatori* 20: 95-109.

Barrett, J. 1988. Fields of Discourse: reconstituting a social archaeology. *Critique of Anthropology* 7(3): 5-16.

Bateman, J. and Jeffrey, S. 2011 What matters about the monument: reconstructing historical classification. *Internet Archaeology* 29. http://intarch.ac.uk/journal/issue29/bateman_index.html.

Berners-Lee, T. 1997. The World Wide Web - Past, Present and Future. *Journal of Digital Information* 1(1). http://journals.tdl.org/jodi/article/view/3/3.

Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The Semantic Web. *Scientific American* May 2001, 34-43.

Binford, L. 1987. Data, Relativism and Archaeological Science. *Man (New Series)* 22(3): 391-404.

Boast, R. and Biehl, P. 2011 Archaeological knowledge production and dissemination in the digital age. In *Archaeology 2.0: New Approaches to Communication and Collaboration* (eds. E. Kansa, S. Kansa and E. Watrall). Cotsen Digital Archaeology Series. Los Angeles: UCLA Cotsen Institute of Archaeology Press, pp. 119-155. http://escholarship.org/uc/item/1r6137tb.

Bowker, G. and Star, S.L. 1999. *Sorting Things Out: Classification and its Consequences*. Cambridge, Mass.: MIT Press.

Brunsson, N. and Jacobsson, B. 2002. The contemporary expansion of standardisation. In *A World of Standards* (eds. N. Brunsson and B. Jacobsson). Oxford: Oxford University Press, pp 1-18.

Burgelman, J. 2000. Travelling with communication technologies in space, time, and everyday life: an exploration of their impact. *First Monday* 5(3). http://www.firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/733/642.

Byrne, K. and Klein, E. 2010. Automatic extraction of archaeological events from text. In *Making History Interactive: Computer Applications and Quantitative Methods in Archaeology 2009* (eds. B. Frischer, J. Crawford & D. Koller). Oxford: British Archaeological Reports International Series 2079, pp. 48-56.

Castells, M. 1996. *The Information Age: Economy, Society and Culture. Volume 1: The Rise of the Network Society.* Oxford: Blackwell.

Cooper, M. 1985. Computers in British archaeology: the need for a national strategy. In *Current Issues in Archaeological Computing* (eds. M. Cooper and J.D. Richards). Oxford: British Archaeological Reports International Series 271, pp. 79-92.

Cripps, P., Greenhalgh, A., Fellows, D., May, K. and Robinson, D. 2004. *Ontological Modelling of the work of the Centre for Archaeology*. CIDOC CRM Technical Paper. Paris: ICOM. http://cidoc.ics.forth.gr/technical_papers.html.

Dewey, J. 1925. *Experience and Nature.* Chicago: Open Court Publishing.

DCMI Usage Board 2010. Dublin Core Metadata Initiative Metadata Terms. http://dublincore.org/documents/dcmi-terms/.

English Heritage 1999. *National Monuments Record Thesauri*. Swindon: English Heritage. http://thesaurus.english-heritage.org.uk/

FISH 2004. *Forum on Information Standards in Heritage: INSCRIPTION*. York: FISH. http://www.fish-forum.info/inscript.htm.

Harley, D., Acord, S., Earl-Novell, S., Lawrence, S. and King, J. 2010. *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Berkeley: Center for Studies in Higher Education. http://escholarship.org/uc/item/15x7385g.

Hodder, I. 1999. *The Archaeological Process: An Introduction*. Oxford: Blackwell.

Huggett, J. 1985. Expert systems in archaeology. In *Current Issues in Archaeological Computing* (eds M. Cooper and J.D. Richards). Oxford: British Archaeological Reports International Series 271, pp. 123-142.

Huggett, J. 1995. Democracy, data and archaeological knowledge. In *Computer Applications and Quantitative Methods in Archaeology 1994* (eds. J. Huggett and N. Ryan). Oxford: British Archaeological Reports International Series 600, pp. 23-26.

Huggett, J. 2004. The Past in Bits: Towards an Archaeology of Information Technology? *Internet Archaeology* 15. http://intarch.ac.uk/journal/issue15/huggett_index.html.

IfA 2008a. *Institute for Archaeologists Standards and Guidance for Excavation.* Reading: IfA. http://www.archaeologists.net/codes/ifa

IfA 2008b. *Institute for Archaeologists Standards and Guidance for the collection, documentation, conservation and research of archaeological materials*. Reading: IfA. http://www.archaeologists.net/codes/ifa

Jeffrey, S., Richards, J., Ciravegna, F., Waller, S., Chapman, S. and Zhang, Z. 2009. The Archaeotools project: faceted classification and natural language processing in an archaeological context. *Philosophical Transactions of the Royal Society A* 367: 2507-2519.

Jones, A. 2002. *Archaeological Theory and Scientific Practice*. Cambridge: Cambridge University Press.

Kansa, E. 2011. New directions for the digital past. In *Archaeology 2.0: New Approaches to Communication and Collaboration* (eds. E. Kansa, S. Kansa and E. Watrall). Cotsen Digital Archaeology Series. Los Angeles: UCLA Cotsen Institute of Archaeology Press, pp. 1-26. http://escholarship.org/uc/item/1r6137tb.

Kansa, E. and Bissell, A. 2010. Web syndication approaches for sharing primary data in "small science" domains. *Data Science Journal* 9: 42-53.

Khazraee, E. and Khoo, M. 2011. Practice-Based Ontologies: a new approach to address the challenges of ontology and knowledge representation in history and archaeology. In *Metadata and Semantic Research* (eds. E. García-Barriocanal, Z. Cebeci, M. Okur and A. Öztürk). Communications in Computer and Information Science vol. 240. Berlin: Springer, pp 375-386.

Koski, J. 2001. Reflections on information glut and other issues in knowledge productivity. *Futures* 33: 483-495.

Kroeker, K. 2010. Engineering the Web's third decade. *Communications of the ACM* 53(3): 16-18.

Lee, E. (ed) 2007. *MIDAS Heritage: the UK Historic Environment Data Standard*. London: English Heritage/Forum for Information Standards in Heritage (FISH). http://www.english-heritage.org.uk/publications/midas-heritage/.

Limp, W.F. 2011. Web 2.0 and beyond, or on the Web nobody knows you're an archaeologist. In *Archaeology 2.0: New Approaches to Communication and Collaboration* (eds. E. Kansa, S. Kansa and

E. Watrall). Cotsen Digital Archaeology Series. Los Angeles: UCLA Cotsen Institute of Archaeology Press, pp. 265-280. http://escholarship.org/uc/item/1r6137tb.

Lock, G. and Spicer, R. 1985. Real and imaginary limitations of micro-computers in archaeology. In *Current Issues in Archaeological Computing* (eds. M. Cooper and J.D. Richards). Oxford: British Archaeological Reports International Series 271, pp.102-122.

Lucas, G. 2012. *Understanding the Archaeological Record*. New York: Cambridge University Press.

May, K., Binding, C. and Tudhope, D. 2010. Following a STAR? Shedding more light on semantic technologies for archaeological resources. In *Making History Interactive: Computer Applications and Quantitative Methods in Archaeology 2009* (eds. B. Frischer, J. Crawford & D. Koller). Oxford: British Archaeological Reports International Series 2079, pp.227-233.

Miller, P. 1996. Metadata for the masses. *Ariadne* 5. http://www.ariadne.ac.uk/issue5/metadata-masses/.

Miller, P. and Wise, A. 1997. *Resource Discovery Workshops: Final report from the Archaeology Data Service*. York: ADS. http://ads.ahds.ac.uk/project/metadata/workshop1_final_report.html.

Millerand, F. and Bowker, G. 2009. Metadata standards: trajectories and enactment in the life of an ontology. In *Standards and their Stories: how quantifying, classifying, and formalizing practices shape everyday life* (eds. M. Lampland and S.L. Star). Ithaca: Cornell University Press, pp. 149-165.

Patrik, L. 1985. Is there an archaeological record?. *Advances in Archaeological Method and Theory* 8: 27-62.

Randall, D., Procter, R., Lin, Y., Poschen, M., Sharrock, W. and Stevens, R. 2011. Distributed ontology building as practical work. *International Journal of Human-Computer Studies* 69: 220-233.

RCAHMS 2004. *Royal Commission on the Ancient and Historical Monuments of Scotland Thesauri*. Edinburgh: RCAHMS. http://lmid1a.rcahms.gov.uk/rcahms/apps/f?p=210:1:4825932420168742

Reilly, P. 1985. Computers in field archaeology: agents of change? In *Current Issues in Archaeological Computing* (eds. M. Cooper and J.D. Richards). Oxford: British Archaeological Reports International Series 271, pp.63-78.

Ribes, D. and Bowker, G. 2009. Between meaning and machine: learning to represent the knowledge of communities. *Information and Organization* 19: 199-217.

Ribes, D. and Finholt, T. 2009. The Long Now of Technology Infrastructure: articulating tensions in development. *Journal of the Association for Information Systems* 10: 375-398.

Richards, J.D. 1985. Standardising the record. In *Current Issues in Archaeological Computing* (eds. M. Cooper and J.D. Richards). Oxford: British Archaeological Reports International Series 271, pp. 93-102.

Richards, J.D. 2002. Digital preservation and access. *European Journal of Archaeology* 5: 343-366.

Richards, J.D. 2006 Archaeology, e-publication and the Semantic Web. *Antiquity* 80: 970-979.

Richards, J.D. 2009. From anarchy to good practice: the evolution of standards in archaeological computing. *Archeologia e Calcolatori* 20: 27-35.

Richards, J., Jeffrey, S., Waller, S., Ciravegna, G., Chapman, S. and Zhang, Z. 2011. The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing. In *Archaeology 2.0: New Approaches to Communication and Collaboration* (eds. E. Kansa, S. Kansa and E. Watrall). Cotsen Digital Archaeology Series. Los Angeles: UCLA Cotsen Institute of Archaeology Press, pp. 31-56. http://escholarship.org/uc/item/1r6137tb.

Roberts, J. 2011 An anthropological study of war crimes against children in Kosovo and Bosnia-Herzegovina in the 1990s. Glasgow University: unpublished PhD thesis. http://theses.gla.ac.uk/2562/.

Schroeder, R. 2008. e-Sciences as research technologies: reconfiguring disciplines, globalizing knowledge. *Social Science Information* 47(2): 131-157.

Star, S.L. and Lampland, M. 2009. Reckoning with standards. In *Standards and their Stories: how quantifying, classifying, and formalizing practices shape everyday life* (eds. M. Lampland and S.L. Star). Ithaca: Cornell University Press, pp. 3-24.

Thévenot, L. 2009. Governing Life by Standards: A View from Engagements. *Social Science Information* 39 (5): 793-813.

Timmermans, S. and Epstein, S. 2010. A World of Standards but not a Standard World: toward a sociology of standards and standardization. *Annual Review of Sociology* 36: 69-89.

Tudhope, D., Binding, C., Jeffrey, S., May, K. and Vlachidis, A. 2011. A STELLAR role for knowledge organisation systems in digital archaeology. *Bulletin of the American Society for Information Science and Technology* 37(4): 15-18.

Van House, N. 2003. Digital Libraries and Collaborative Knowledge Construction. In *Digital Library Use: Social Practice in Design and Evaluation* (eds. A. Bishop, N. Van House and B. Buttenfield). Cambridge, Mass.: MIT Press, pp. 271-295.

Wise, A. and Miller, P. 1997. Why metadata matters in archaeology. *Internet Archaeology* 2. http://intarch.ac.uk/journal/issue2/wise_index.html.

Witmore, C. 2009. Prolegomena to Open Pasts: On Archaeological Memory Practices. *Archaeologies: Journal of the World Archaeological Congress* 5(3): 511-545.

Weinberger, D. 2011. *Too Big to Know*. New York: Basic Books.