REUSE REMIX RECYCLE: REPURPOSING ARCHAEOLOGICAL DIGITAL DATA

Jeremy Huggett Archaeology, School of Humanities, University of Glasgow, Glasgow G12 8QQ, UK (jeremy.huggett@glasgow.ac.uk, corresponding author)

Abstract

Preservation of digital data is predicated on the expectation of its reuse, yet that expectation has never been examined within archaeology. While we have extensive digital archives equipped to share data, evidence of reuse seems paradoxically limited. Most archaeological discussions have focused on data management and preservation, and on disciplinary practices surrounding archiving and sharing data. This paper addresses the reuse side of the data equation through a series of linked questions: what is the evidence for reuse, what constitutes reuse, what are the motivations for reuse, and what makes some data more suitable for reuse than others? It concludes by posing a series of questions aimed at better understanding our digital engagement with archaeological data.

La conservación de los datos existentes en formato digital se basa en la expectativa de su posterior reutilización y aprovechamiento. Sin embargo, esta expectativa, como posibilidad razonable, nunca ha sido examinada dentro del campo de la arqueología. Aunque contamos con grandes repositorios digitales equipados para compartir datos, las evidencias de casos de una reutilización clara de los mismos parecen ser, paradójicamente, limitadas. La mayor parte de los análisis acerca de estas cuestiones desde la arqueología se han centrado en cómo gestionar y conservar los datos, y también en las prácticas de almacenamiento y acceso compartido a los datos llevadas a cabo dentro de la disciplina. Este artículo aborda el aspecto de la reutilización de los datos en este contexto a través de una serie de preguntas vinculadas entre sí: ¿Qué evidencias o casos de reutilización existen? ¿Qué constituye realmente una reutilización de datos?, ¿Cuáles son las motivaciones para su reutilización, y qué hace que algunos datos sean más adecuados para reutilizar que otros? Así mismo, el artículo concluye planteando una serie de preguntas enfocadas a comprender mejor nuestra relación entre el medio digital y los datos arqueológicos.

> "The best thing to do with your data will be thought of by someone else"
> (Rufus Pollock, cited in Wentworth 2008)

The Digital Data Paradox

In 1995, Brian Fagan drew attention to what he called "archaeology's dirty secret" – the failure of archaeologists to publish definitive reports on their fieldwork, pointing to problems in scholarly culture including a focus on funding field research rather than analysis or publication (Fagan 1995:16). He identified the digital as a catalyst for change: "The demands of the electronic forum will make it harder to duck the responsibility of preparing one's data for scholarly use and scrutiny. In many cases, 'publication' will consist of meticulously organized databases, including graphics." (Fagan 1995:17). In the intervening years there has been a significant development of digital

repositories in the UK, USA and elsewhere facilitating this vision by providing for the preservation and distribution of archaeological data. However, rather than resolving archaeology's 'dirty secret', they have in fact unwittingly extended it. Consequently, referencing Fagan's 'dirty secret', Cherry observed that archaeology "remains stubbornly intransigent in the face of digital technologies" (Cherry 2011:12).

In the meantime the preservation of archaeological data within digital repositories has become normative practice (Kansa, Kansa and Arbuckle 2014:58), underlined by the current range of professional archaeology codes of practice. However, evidence of actual reuse of these data remains rare (Kansa, Kansa and Arbuckle 2014:58; Huvila 2016) and, reminiscent of Fagan's earlier criticism, much archaeological research remains focused on the generation and collection of new data. This is despite a number of studies demonstrating archaeological commitment to and support for digital preservation and data reuse (e.g. Faniel, Kansa, Kansa, Barrera-Gomez and Yakel 2013; Frank et al. 2015). There is therefore a paradox: archaeologists deposit and share their data but make comparatively little use of data shared by others.

This has implications for the sustainability of the digital repositories that manage the data, and for the wider discipline. Justifying the considerable investment in time, money, expertise and energy to create and manage archaeological data archives and to make data available for sharing might be open to question if they are seen solely as places for storage, even if those data are among the primary surviving evidence of fieldwork encounters with the past. Data resilience through reuse is seen to support data preservation into the future, and at the same time to meet the ethical principle that research results should be capable of review and refinement. As the Archaeology Data Service (ADS) emphasizes, "reuse of data is the single surest way of maintaining the integrity of data and tracking errors and problems with it" (ADS 2014). However, the presumption of digital reuse contrasts with the evidence from non-digital archives which traditionally show low levels of use of their resources (Merriman and Swain 1999:259-60). Why should we expect digital reuse to be any different?

Absence of Evidence and Evidence of Absence

Digital archaeologists, digital archivists, and digital curators have recently begun to express anxiety about the difficulty of demonstrating reuse of digital resources, although the lack of evidence for data reuse in archaeology is often anecdotal. A survey of the literature reinforces the impression of a lack of reuse since much of it focuses on the mechanics of making the data reusable in the first place – not unreasonable during an era which has seen the case made for the creation of digital repositories, and their subsequent development and growth towards a tipping point at which they become genuinely useful resources. Numerous studies report on creating the means to support data sharing: for example, the FAIR principles (Findability, Accessibility, Interoperability, Reusability) (Wilkinson et al. 2016) have reuse at their core but stop at the point of ensuring that it is feasible. Unless steps are taken to encourage those data to be taken up and reused, the data cycle easily stalls in the absence of motivation or incentive to reuse. Making data shareable and accessible is not the same as actual reuse, but evidencing this gap in the cycle has proved problematic.

Reassuringly this is not just an issue for archaeology. For example, a survey of digital archives across Europe identified reuse of data as a key aspect of sustainability and at the same time pointed to a lack of means to evaluate levels of reuse (Sasse et al. 2017:67-8). Standard web metrics such as page

impressions, click rates and downloads are often used as indicative of reuse (for instance, Richards 2002:359; Green 2016:21; Richards 2017:5-6) although they have a very ambiguous relationship with it; indeed, such metrics primarily capture supply rather than reuse (see Figure 1). Even data downloads are a poor proxy for reuse since the purpose behind the download does not necessarily lead to reuse. For instance, in Figure 2 the Wharram Percy archive at the ADS saw 1614 downloads resulting in 13 citations (including self-citations) according to the Thomson Reuters Data Citation Index (DCI). In comparison, 671 downloads of the Cleatham archive resulted in 0 citations in the DCI although a Google search revealed at least three citations (one PhD and two academic journal papers). Consequently, downloads demonstrate potential reuse at best and an equivocal relationship with subsequent citations.

Nevertheless, data citations are frequently seen as a more robust means of demonstrating reuse (for example, Piwowar and Vision 2013) and the use of persistent Digital Object Identifiers (DOI) enables datasets to be uniquely identified and reliably located. However, analyses of data citation demonstrate that their use is as yet limited: research data commonly goes uncited (Peters et al. 2016:740), explicit data citations in the reference sections of publications are rare, and where reference is made at all, it is to the title of the dataset in the body of the paper (Mooney and Newton 2012:7), or in the acknowledgements or supplementary materials. The lack or informality of data citation makes identification and documentation of reuse difficult (Park and Wolfram 2017:457), and a brief overview of archaeological data citation suggests the situation is much the same. For instance, a survey of Archaeology Data Service datasets indexed in the DCI indicates that 56 of 476 datasets (as of August 2017) have been cited, although only 12 have been cited more than once. In most cases these citations are by the original authors of the datasets themselves, often within the published reports that the data belong to, and such self-citation makes the evaluation of actual data reuse more complex still (Park and Wolfram 2017:457). Citation indexes such as the DCI currently lack contextual nuance capable of distinguishing between the different kinds of formal and informal citation, original use and actual reuse. Despite the problems surrounding data citation it is considered to be a primary means of recognizing and validating the contribution of the original authors through increasing peer visibility and reputation (e.g. Dallmeier-Tiessen et al. 2014:25). That this is poorly practiced and evidence for reuse is inadequately captured remains a significant problem.

Even if data citation capture were more comprehensive, its reliance on a traditional academic publication model restricts its value since there are modes of reuse that are not captured or warranted by academic-style citations, such as data used for teaching purposes, student projects, experimentation unrelated to the origins of the data, and so on (Piwowar and Vision 2013:21). Altmetrics may provide an alternative mechanism for gaging levels of a wider range of reuse practices through capturing mentions on social media, social bookmarking, social recommending, blogging and micro-blogging sites, for instance. However, altmetrics tools bring their own set of problems (e.g. Haustein 2016). Most rely on the use of DOIs which assumes these are widely used, but this does not yet appear to be the case in archaeology: for instance, Google searches for the DOIs of random ADS archives lead back to the original archive rather than to references to the archive, even in cases where the archives were known to have been reused and cited. Event services such as those being developed by DataCite and Crossref are intended to capture the broader range of reuses, but are dependent on the use of persistent identifiers (Dappert et al. 2017, Robinson-Garcia et al. 2017). Until archaeologists consistently use DOIs or other such persistent identifiers for their data , the capture of altmetrics for archaeological data citation seems equally problematic and equally unreliable as standard data citations, as indeed is the experience elsewhere (Peters et al. 2016).

What is Reuse?

What this highlights is that reuse is in itself not straightforward and can mean different things in different circumstances; indeed, definitions can frequently seem contradictory. A useful distinction to start with is between 'use' and 'reuse', where 'use' is essentially related to the purposes of the originator of those data. Hence,

> In the simplest situation, data are collected by one individual, for a specific research project, and the first 'use' is by that individual to ask a specific research question. If that same individual returns to that same dataset later, whether for the same or a later project, that usually would be considered a 'use' [Pasquetto et al. 2017:3].

By this definition, therefore, the level of data self-citations highlighted above is actually a measure of 'use' rather than 'reuse'. 'Use' occurs after data collection and storage although some definitions might suggest that all data processing subsequent to initial collection technically constitute reuse rather than use (Custers and Uršič 2016:7-8). However, since data collection is a prerequisite for any kind of subsequent use, it seems more reasonable to see data 'use' in terms of the actions of the primary data author(s) up to and including any of their analyses employing the data. Whether or not the primary author(s) returning to the data subsequent to the data's deposition in a repository and retrieving them for further analysis constitutes use or reuse remains an open question (Pasquetto et al. 2017:3) although one might imagine in most cases the original investigator(s) would work with copies that they retained prior to deposition which would logically imply use rather than reuse by this definition.

A common characterization of 'reuse' is as secondary use for purposes other than that for which the data were originally collected (e.g. Faniel, Kriesberg and Yakel 2016:1404; Zimmerman 2008:634). On the face of it, this would mean that that reproducibility, one of the most common reasons lying behind data sharing (Pasquetto et al. 2017:4), cannot be seen as reuse since it entails the reanalysis of a prior study using the same data and methods as the original investigator in order to confirm or deny previous research (e.g. Marwick 2017:427). Other investigators, separated in time and space, use the data for the purpose for which they were originally collected, and to deny that this constitutes 'reuse' would seem perverse. The simplest approach to this use/reuse dichotomy is to see the key distinction between use and reuse as being less to do with the purpose of data (re)usage and more closely linked to the authorship of the data themselves: hence, 'use' is associated with actions of the primary data producer(s) whereas 'reuse' is any secondary use by those other than the primary producer(s). This aligns with Borgman's characterization of 'sources' and 'resources', where sources are data originating with the investigator and resources are existing data reused for a project (Borgman 2015, 64-5).

Beyond these general definitions of 'use' and 'reuse', it is also clear that reuse itself can entail a variety of different practices. Reuse may go beyond simply reusing data over again: for example, it may involve taking the data and combining them with other data to create a new dataset for a different and/or extended analysis (a 'remix'), or it may reprocess ('recycle') the data to create something new through altering the focus and/or discarding data. An alternative characterization would be to see data as being reused within their original purpose (such as in reproducibility) or reused for a different purpose altogether (a different research focus, such as a cemetery excavation dataset used in a pottery study); equally data may be reused in the same context (the cemetery

dataset used in a comparative study with other contemporary datasets, for example), or reused in an entirely different context (the cemetery dataset contributing to a regional identity study, for instance). In this sense, reuse may entail repurposing or recontextualization of the data (c.f. Custers and Uršič 2016:9). In the process, reuse may result in new derived data products which themselves may be used and reused in subsequent analyses, creating a network of interdependencies and authorships.

'Reuse' is therefore a more slippery concept than it might at first seem, but whether reused, remixed, recycled, repurposed, or recontextualized, two key questions lie behind potential reuse: what are the motivations behind reuse? and how reusable are the data themselves?

Why Reuse It?

Reasons behind reuse are rarely discussed; more common are reasons why data should be shared. For example, Borgman suggests four rationales for sharing data: reproducibility/verification, ensuring publicly funded research is widely available, enabling new questions to be asked of the data, and advancing research and innovation (Borgman 2012:1067). The first and third of these are equally appropriate as motivations for reuse, and the fourth might reasonably be hoped to be a consequence of them. At the same time, however, they reinforce an association between reuse and research which may be over-emphasized. For example, a recent study of social scientists' use of data archives shows that the proportion of teaching (15%) and learning (64%) significantly outweighed research-related downloads of data from the UK Data Service, with similar results from the Finnish Data Service (Bishop and Kuula-Luumi 2017:4-7). Conversely, an analysis of the ADS registered user demographic indicated that academic research was the most common reported reason for using the ADS (Green 2016:21-2), although this is not limited to downloads.

The reproducibility or verification of research is most commonly cited as the motivation for sharing data (Borgman 2012:1067; Pasquetto et al. 2017:4). Reproducibility is a key aspect of scientific method, but an inability to reproduce results across a wide range of disciplines in the sciences and social sciences has led to claims of a reproducibility 'crisis'. For example, a survey by *Nature* found that more than 70% of researchers had tried and failed to reproduce the results of others, while more than 50% had failed to reproduce their own experiments; overall, 52% believed there was a significant reproducibility crisis (Baker 2016:452). While data reuse can contribute to the ability to test archaeological arguments and explanations (e.g. Smith 2015), such applications are relatively rare. In part, this is due to limited provision made for reproducibility, linked to issues with the availability of the data and associated processing methods (Marwick 2017), but it can also be related to an approach to reproducibility which employs comparative data rather than in-depth re-evaluation of the original data. In this respect, little seems to have changed since Barrett's non-digital reanalysis of the Iron Age Glastonbury Lake Village which entailed a deep reading of the original excavation report rather than relying on evidence from analogous sites and archives (Barrett 1987:410). Such practice remains rare, and suggests that simply making data digitally available for reuse will not address reproducibility: there remains a strong tradition within archaeology of conducting new research by collecting new data rather than reusing old data collected by others (Huvila 2016).

Nor is it necessarily the case that data are reused as a means of driving forward new research. For example, in a study of researchers at the Center for Embedded Network Sensing it was found that

the purpose behind data reuse was typically as a baseline or context to provide a background to research, rather than the external data being foregrounded as a primary means of research (Wallis et al. 2013:14).

A study of data reuse behaviors and the factors associated with data reuse practice amongst social scientists (Curty 2016, Curty and Qin 2014) helps clarify potential motivations behind reuse or reasons for non-reuse within archaeology. Six core factors influencing reuse were identified (see Figure 3), with each subdivided into additional factors which to varying degrees affect attitudes towards and feasibility of data reuse (Curty 2016:100; Curty and Qin 2014:2).

Perceived benefits are associated with the extent to which new knowledge might result from reuse, a cost/benefit analysis of using secondary data rather than collecting new primary data, and the extent to which the data available are seen as reliable (Curty 2016:100-1). A balance of positive outcomes across these factors would support data reuse.

Perceived risks include the perception that reuse is less valued than original data in research, that there may be ethical issues associated with secondary data (for instance, copyright and confidentiality), that reuse might be perceived as misuse or misinterpretation, and that there might be hidden errors within the secondary data that were not picked up (Curty 2016:101-3). A balance of negative outcomes – high risk – across these factors would limit likelihood of data reuse.

Perceived effort relates to the work entailed in handling secondary data that were not created by the researcher. Issues include the accessibility of the data, the ease of discovery, the extent to which the data are adaptable to the current research objectives, the extent to which the data have to be reworked for reuse in the new context, and an understanding and appreciation of the complexities of the original research that created the data in the first place (Curty 2016:103-5). A balance of negative outcomes across these factors, suggesting high friction, would limit the likelihood of data reuse.

Perceived reusability covers the extent and quality of data documentation, the fitness of purpose, the trustworthiness and credibility of the data creator(s), the quality of the data in terms of its consistency (accuracy and reliability) and completeness (minimal missing data), and the rigor (methods, procedures etc.) of the original research that created the data (Curty 2016:105-7). A balance of positive outcomes across these factors, indicating confidence in the data, would support subsequent reuse.

Enabling factors facilitate reuse, and relate to the availability of documentation associated with previous applications and contexts of use rather than data reusability (see above), availability of data repositories, the potential accessibility of the data creators themselves, the level of training in data analysis and reuse, and the availability of support and assistance for reuse (Curty 2016:107-9). A balance of positive outcomes across these factors would support reuse.

Social factors concern the attitudes of peers and the wider discipline: the extent to which data reuse is accepted as legitimate research, and the levels of encouragement and incentivization for reuse (Curty 2016:109-10). A balance of positive outcomes across these factors would support reuse.

The data reuse behavioral model in Figure 3 highlights a range of factors which contribute to a determination of the feasibility of reuse in a complex web of interrelated decisions. Some factors (such as data accessibility and data quality) are likely to be more significant than others (for example, availability of primary investigators) in terms of archaeological reuse, emphasizing that the balance of factors will vary between different potential reuse applications.

How Reusable is It?

If reuse is a slippery concept, so too are data. What constitute data is a matter of negotiation between the potentially conflicting perspectives of data producers, data curators, and data users, changing through time and through (re)use so that what was not considered to be data at the time may be of interest at a later stage, and conversely, what was understood to be data may no longer be seen to have value. Consequently, we are dependent on the tacit knowledge and research agendas of those who precede us, just as those who follow us will be limited by ours (e.g. Wylie 2017:207). This has implications for the collection, curation, and subsequent reuse of data. For instance, recognizing that the future use of data cannot be fully anticipated reinforces the tension between the concept of "analytical destiny" (Carver 1985:50) in which material/data have a known value and are collected where there is an identified purpose, versus data which might conceivably have a future, as yet unforeseen, value which would otherwise be lost but yet there is a cost to its capture and retention. The FAIR Guiding Principles define a series of characteristics required for data to be reusable: rich description, detailed provenance, use of domain standards, and clear licensing arrangements (Wilkinson et al. 2016: 4), but meeting even these basic requirements is not straightforward.

Data Wrangling

A considerable challenge to reuse is the need to rework data in terms of managing different recording conventions, data formats and data models (e.g. Kandel et al. 2011). Numerous types of errors can occur within the data themselves: for instance, syntactical data anomalies (irregularities in format or value), semantic anomalies (inconsistencies within or across data such as duplicates, contradictory data), and coverage anomalies (missing or incomplete data) (Müller and Freytag 2003:6-7). Such issues only increase as datasets are increasingly aggregated into meta-analyses (Kansa 2015:225) and incorporated into 'big data'-style analyses (Gattiglia 2015). Recognizing and dealing with such anomalies can be extremely time-consuming, assuming they are spotted in the first place, although it can be argued that this is outweighed by the cost saving in not having to collect the data again (e.g. Heidorn 2008:290). Of course, in archaeology the data collection process is frequently unrepeatable, and instead new analogous data may be collected as an alternative to reuse.

Digital tools make it relatively easy to manipulate single datasets, to aggregate datasets that are notionally similar, or to integrate datasets that are seemingly unrelated, and in the process create 'new' data from old. This is assumed to be beneficial, but equally such consolidated datasets may be little more than what Clarke characterizes as a "melange": collections of mixed or uncertain quality, with uncertain associations, and combined and modified in ways that are not transparent (Clarke 2016:83). Clarity regarding the means by which reused datasets are arrived at prior to analysis can be variable, complicated by the range of possible issues that may be encountered beyond technical issues with the data. Commonplace are contextual discrepancies, methodological biases, and problems with associated documentation. Again, these are just as much a feature of non-digital data but are compounded by technical issues linked to comparability, compatibility and standardization of the digital data. Where little or no information is provided about any data-cleansing and manipulation in advance of analysis, confidence in the derived data and its subsequent use can be limited at best, especially when data 'cleaning' may more often be concerned with the ease with

which the analytical tools can be applied to the data rather than the cleanliness of the data themselves (Clarke 2016:83).

Data Contexts

A key factor that a number of authors have identified is that archaeological data requires associated contextual information in order to be capable of considered reuse (e.g. Faniel, Kansa, Kansa, Barrera-Gomez and Yakel 2013:297; Atici et al. 2013:667; Faniel, Barrera-Gomez, Kriesberg and Yakel 2013:798; Kansa, Kansa and Arbuckle 2014:59; Frank, Yakel and Faniel 2015:144). This is typically addressed through the provision of metadata – data about the data in question – although the levels of detail and coverage can vary widely. The most common metadata focuses primarily on the needs of discovery (authorship, rights, sources, etc.) rather than reuse of the data which requires paradata – data about provenance, process, and derivation (e.g. Baker 2012, Mudge 2012). Indeed, depending on the kind of reuse, even paradata may not be sufficient as it does not include computer code or scripts used in the original analysis and manipulation of the data, or even the computer environment itself (Marwick 2017). Consequently different kinds of reuse may require different levels of contextual detail to support it (e.g. Bechhofer et al. 2012:603), and by implication, the levels of contextual information available will place limits on the kinds of reuse that are possible for a specific dataset.

Despite this, meta/paradata associated with data are most commonly of value to the computational tools used to locate and manage the data than to the human agents seeking to make use of it. Indeed, although the lack of contextual information is frequently cited as a shortcoming of archaeological data, nevertheless data are reused as ways are found of circumventing it (Faniel, Kansa, Kansa, Barrera-Gomez and Yakel 2013:298; Atici et al. 2013:667). As Borgman observes, "The effort required to explain one's research records adequately increases as a function of the distance between data originators and users" (Borgman 2007:167), and contextual information accompanying data is seen as a means of overcoming the 'knowledge distance' between those reusing and the original data producers (Markus 2001:88). Consequently, its absence should be a significant barrier to reuse.

However, the significance of absent or limited contextual information is difficult to assess given a relative lack of studies explicitly examining reuse of archaeological data. An investigation of the reuse of faunal data which lacked background contextual and methodological information (provenance, relationships, standards etc.) demonstrated that, as expected, different analysts arrive at different conclusions from the same data (Atici et al. 2013:670ff). Individual analysts responded differently to the lack of contextual information in relation to issues of aggregation of data categories, and assessment of data reliability, consistency and comparability, for instance (Atici et al. 2013:667). However, whether the absence of contextual information primarily governed some of these differences is debatable: for example, while the lack of information regarding sampling bias is clearly a limiting issue, choosing to lump (or split) data categories is an analytical decision likely to be equally associated with the personal experience and preferences of the analyst as with a lack of contextual information.

This suggests that different kinds of data and different kinds of reuse require different levels of contextual information. Qualitative data may be more sensitive to these requirements than quantitative data, and this might explain why common reuse applications tend to employ either primarily quantitative data (for example, zoological (e.g. Jones and Gabe 2015; Arbuckle et al. 2014; Conolly et al. 2011), dating (e.g. Armit et al. 2013), burial (e.g. Bradbury et al. 2016) or else locational data from large scale regional databases (e.g. Cooper and Green 2016; Bevan 2012; Richards et al.

2009). Large scale artifact and monument datasets in particular are widely used primarily because their focus on type and location enables their use as a backdrop to original research (c.f Wallis et al. 2013:11) often with little or no contextual knowledge required. The greater frequency of reuse of quantitative as opposed to qualitative datasets is also noted in other disciplines (Park and Wolfram 2017:459), suggested in part to be simply due to the greater amount of quantitative data available and the relative ease of reuse of such data compared to more narrative forms. Indeed, textual data is frequently converted into quantitative data via coding to make it more 'computable'.

Data Epistemologies

How we conceive of data clearly has implications for the practice of reuse; at the same time data analytics frequently operate within different epistemological frameworks which can easily be set aside or go unrecognized in the pursuit of reuse. A traditional view of data as empirical 'facts' implies a straightforward approach to data reuse and contrasts with a perspective of data as observations about attributes perceived as having some contemporary value in understanding the past, which suggests a more complex, ambiguous approach to data reuse. One outcome of this may again be a preference for collecting new data rather than reusing existing data, not only to avoid data wrangling and contextual issues, but also as a means of sidestepping theoretical or philosophical tensions such as these.

Many examples of data reuse fall into the category of 'meta-analysis' or data mining, an approach which only becomes feasible with the availability of large quantities of data held in open repositories. Incorporating data collected by others, frequently unknown to the reuser, with the data being used for very different purposes to its original intent, can be seen as an extreme form of reuse in which contextual information is useful but frequently absent (Markus 2001:71-72). This aligns with an increasing interest in 'Big Data', an approach to data analysis where quantity of data is often seen to overcome limitations in quality or sampling (e.g. Gattiglia 2015:114, following Mayer-Shönberger and Cukier 2013:33). Such approaches tend to adopt (or implicitly assume) a traditional view of data, one in which data are conceived as raw, unprocessed, unworked, typically acquired using rigorous scientific methods and distinct from subjective interpretation or other contextual matters (Costa et al. 2013:450; Huggett 2015:14ff). This is often characterized as a data-information-knowledge pyramid (see Figure 4) – a view of data at the bottom of a hierarchy which moves through information to knowledge, and, ultimately, wisdom. In this DIKW model, knowledge is created from a hierarchy of building blocks – data are transformed into information, information is processed into knowledge, and we travel up this hierarchy to arrive at wisdom (e.g. Kitchin 2014:9-12; see also Gattiglia 2015:115). This has its attraction as an apparently common-sense perspective, mapping conveniently onto archaeological perceptions such as the image of a data mountain that is waiting to be transformed into archaeological knowledge. However, it is a perspective that that is embedded in an essentially positivist approach to archaeology, in which data are seen as "separate from the subjectivities that generate them, and independent of the relational and intersubjective contexts that give rise to them" (Mauthner and Parry 2009:292; see also Huggett 2015:15-17). In contrast, data are a consequence of cultural processes, and hence are theory-laden, process-laden, and purpose-laden (Huggett 2014:5), and not raw in any sense; they emerge as the outcome of the application of knowledge and information in a reversal of the DIKW model (e.g. Tuomi 1999) (see Figure 5). Hence, based on their experience, research objectives etc., the data creator articulates their knowledge to identify and categorize information, and that information, in order to be represented in a digital environment, is atomized to create data. Essentially, the data-information-knowledge of the reuser only emerges through an understanding of the knowledge-information-data disarticulation of the original creator (Huggett 2015:18). In the process, explicit, representable knowledge is prioritized, and what tends to be lost is the tacit, contextual knowledge.

While most archaeological data does not conform to the characteristics of 'big' data, nevertheless big data analytical approaches are increasingly being applied to 'small' data, aided by the growth of archaeological data repositories and the access they provide to a wider range of data and the aggregation and integration of datasets that become possible as a result. 'Big' or 'small', these analytical approaches come with new epistemologies attached (see Kitchin 2014:133ff). For example, data analytics are (in)famously associated with an 'end of theory' (Anderson 2008) in which the quantity of data enables it to be analyzed without hypotheses, allowing algorithms to find patterns and identify correlations. Associated with this is the idea that data can speak for themselves free from theory: that data trump theory in a truly objective neutral fashion, stripped of context and bias. Related to this is the concept that data can be context-free – that they do not require contextual or domain-specific knowledge to carry meaning. Alternatively, the quantities of data available facilitate data-driven approaches in which information emerges from the data, replacing traditional knowledge-based approaches based on hypothesis. Such approaches to reuse present a challenge to traditional ways of doing archaeology, and their consequences are as yet not understood.

Challenges for reuse

We are not yet at the stage where archaeologists are increasingly expected to reuse existing data and need to justify primary data collection (Naylor and Richards 2005:90), or where the orthodoxy of excavation is challenged by access to reusable archaeological data (Beck and Neylon 2012:494), although obligations to funders may cause this to change in future. There is broad agreement with the view that digital archives should not be seen as the final resting places for digital data (Green 2016:17) or as the residue of research (Kansa and Kansa 2014:223) but are instead seen as part of a larger data lifecycle. However, while this data cycle is reasonably well-established in terms of capturing, documenting, and packaging the data for reuse, the missing element in the cycle is all too often the actual reuse of data. At the same time, most discussions concerning digital data focus on these established areas of the lifecycle, and pay significantly less attention to issues surrounding reuse (e.g. Borgman 2007; 2015) which is itself a rich and complex cycle of interconnections, interactions, and interrelationships (see Figure 6). As a consequence, we can develop an appreciation of why we can benefit from reusing data, but the actual approaches and methods of reuse and their limitations are less well understood. This then makes evaluating the documentation and data that are available for sharing in terms of their worth and robusticity a primarily theoretical exercise, presenting a significant risk to the sustainability of digital data archiving in archaeology. The purpose of digital curation is ultimately to ensure future access and reuse, but until reuse becomes part of mainstream practice alongside archiving and sharing this cannot be reliably confirmed.

The primary challenges for digital data reuse therefore are not so much related to questions of suitable means of storage and migration of data, or the provision of adequate documentation to facilitate reuse, or issues of trust and credibility of the data, or the appropriate recognition or citation of reuse, or even the need for training and reward structures to support reuse, important though these are. Instead – or in addition – what is needed is a deeper understanding and appreciation of reuse in terms of what it entails: issues associated with the re-presentation and interpretation of old data, the nature and purpose of reuse, and the opportunities and risks presented by reuse. Such questions are not specific to digital data and have seen some debate over the years; however, digital data change the terms of engagement with their near-instant access,

volume, and flexibility, and their potentially transformative effects on the practice of archaeology now and in the future.

These questions include:

- Recognizing that archaeological data do not create themselves but are performative in nature, how are they created and subsequently used, and how can this be represented for future reuse? How can the presumptions and 'preunderstandings' (Wylie 2017:204) be incorporated in data past and present?
- Given that data have potential that we cannot anticipate, how should data be defined, captured, recorded and curated in order that limits on future reuse are minimized?
- Since archaeological data are by their nature incomplete and imperfectly captured, how can this incompleteness and inconsistency best be represented and handled during their reuse? What are the implications of data being made to speak for data that are not captured and hence not available?
- What are the effects of reusing, repurposing, remixing, recycling, recontextualizing data on archaeological knowledge? For instance, what are the implications of applying alternative frames of analysis, different research agendas etc. to data collected under different theoretical regimes?
- What reprocessing and extraction procedures are required to extract and operationalize data from underlying sources? What are the consequences of methods such as standardizing, reducing, and mapping data into schema as part of a pre-analysis cleansing process?
- What are the implications of the different epistemologies associated with data analytics applied to archaeological data?
- How are steps in the data reuse cycle (Figure 6) shaping archaeological practice? What is being privileged, and what is being left behind?
- What mechanisms need to be put in place to support appropriate reuse of archaeological data?

Addressing these questions in the context of digital data reuse will address not only the immediate concerns and technical issues associated with data reuse, but will also have implications for a deeper understanding of the nature of archaeology's digital turn. Borrowing from Ash et al. (2016), we are increasingly witness to archaeologies *through* the digital, archaeologies *produced* by the digital, and archaeologies *of* the digital, but it is as yet unclear how these are reshaping archaeological practice. A closer appreciation of how we work with our digital data is a fundamental part of this, and key to addressing archaeology's 'dirty secret'.

Acknowledgements

References Cited

Archaeology Data Service (ADS)
    2014   Data Management and Sharing Plans. Electronic document, https://archaeologydataservice.ac.uk/advice/DataManagementPlans.xhtml, accessed August 9, 2017.

Anderson, Chris
    2008   The End of Theory: the Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine* 16 (7). Electronic document, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory/, accessed August 9, 2017.

Arbuckle, Benjamin S., Sarah Whitcher Kansa, Eric Kansa, David Orton, Canan Çakırlar, Lionel Gourichon, Levent Atici, Alfred Galik, Arkadiusz Marciniak, Jacqui Mulville, Hijlke Buitenhuis, Denise Carruthers, Bea De Cupere, Arzu Demirergi, Sheelagh Frame, Daniel Helmer, Louise Martin, Joris Peters, Nadja Pöllath, Kamilla Pawłowska, Nerissa Russell, Katheryn Twiss, and Doris Würtenberger
    2014   Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey. *PLoS ONE* 9(6):e99845. DOI: 10.1371/journal.pone.0099845

Armit, Ian, Graeme T. Swindles and Katharina Becker
    2013   From Dates to Demography in Later Prehistoric Ireland? Experimental Approaches to the Meta-Analysis of Large 14C Data-Sets. *Journal of Archaeological Science* 40:433-438. DOI: 10.1016/j.jas.2012.08.039

Ash, James, Rob Kitchen, and Agnieszka Leszczynski
    2016   Digital Turn, Digital Geographies? *Progress in Human Geography*, in press. DOI: 10.1177/0309132516664800

Atici, Levent, Sarah Whitcher Kansa, Justin Lev-Tov, and Eric Kansa
    2013   Other People's Data: A Demonstration of the Imperative of Publishing Primary Data. *Journal of Archaeological Method and Theory* 20, 663-681. DOI 10.1007/s10816-012-9132-9

Baker, Drew
    2012   Defining Paradata in Heritage Visualisation'. In *Paradata and Transparency in Virtual Heritage*, edited by Anna Bentkowska-Kafel, Hugh Denard, and Drew Baker, pp.163-175. Ashgate: Farnham, Surrey.

Baker, Monya
    2016   Is there a reproducibility crisis? *Nature* 533:452-454

Barrett, John C.
    1987   The Glastonbury Lake Village: Models and Source Criticism. *Archaeological Journal* 144:409-423.

Bechhofer, Sean, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Danius Michaelides, Stuart Owen, David Newman, Shoaib Sufi, and Carole Goble
    2012 Why Linked Data Is Not Enough For Scientists. *Future Generation Computer Systems* 29:599-611. DOI: 10.1016/j.future.2011.08.004

Beck, Anthony, and Cameron Neylon
    2012   A Vision for Open Archaeology. *World Archaeology* 44:479-497. DOI: 10.1080/00438243.2012.737581

Bevan, Andrew
    2012   Spatial Methods for Analysing Large-Scale Artefact Inventories. *Antiquity* 86:492-506. DOI: 10.1017/S0003598X0006289X

Bishop, Libby, and Arja Kuula-Luumi

    2017   Revisiting Qualitative Data Reuse: A Decade On. *SAGE Open* 7:1-15. DOI: 10.1177/2158244016685136

Borgman, Christine L.

    2007   *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. MIT Press: Cambridge, MA.

    2012   The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology* 63:1059-1078. DOI: 10.1002/asi.22634

    2015   *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press: Cambridge, MA.

Bradbury, Jennie, Douglas Davies, Mandy Jay, Graham Philip, Charlotte Roberts, and Chris Scarre

    2016   Making the Dead Visible: Problems and Solutions for "Big" Picture Approaches to the Past, and Dealing with "Large" Mortuary Datasets. *Journal of Archaeological Method and Theory* 23:561-591. DOI: 10.1007/s10816-015-9251-1

Carver, Martin

    1985 The Friendly User. In *Current Issues in Archaeological Computing*, edited by Malcolm. A. Cooper and Julian D. Richards, pp. 47- 61. British Archaeological Reports International Series: Oxford.

Cherry, John F.

    2011   Still Not Digging, Much. *Archaeological Dialogues* 18:10-17. DOI: 10.1017/S1380203811000055

Clarke, Roger

    2016   Big Data, Big Risks. *Information Systems Journal* 26:77-90. DOI: 10.1111/isj.12088

Costa, Stefano, Anthony Beck, Andrew Bevan, and Jessica Ogden

    2013   Defining and Advocating Open Data in Archaeology. In *Archaeology in the Digital Era: Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*, edited by Graeme Earl, Tim Sly, Angeliki Crystanthi, Patricia Murietta-Flores, Constantinos Papadopoulos, Iza Romanowska, and David Wheatley, pp. 449-456. Amsterdam University Press: Amsterdam.

Curty, Renata Gonçalves

    2016   Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study. *International Journal of Digital Curation* 11:96-117. DOI: 10.2218/ijdc.v11i1.401

Curty, Renata Gonçalves, and Jian Qin

    2014   Towards a Model for Research Data Reuse Behavior. *Proceedings of the American Society for Information Science and Technology* 51:1–4. DOI: 10.1002/meet.2014.14505101072

Custers, Bart, and Helena Uršič

    2016   Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection. *International Data Privacy Law* 6:4-15. DOI: 10.1093/idpl/ipv028

Conolly, James, Sue Colledge, Keith Dobney, Jean-Denis Vigne, Joris Peters, Barbara Stopp, Katie Manning, and Stephen Shennan

    2011   Meta-Analysis of Zooarchaeological Data from SW Asia and SE Europe Provides Insight Into the Origins and Spread of Animal Husbandry. *Journal of Archaeological Science* 38:538-545. DOI: 10.1016/j.jas.2010.10.008

Cooper, Anwen, and Chris Green

    2016   Embracing the Complexities of 'Big Data' in Archaeology: the Case of the English Landscape and Identities Project. *Journal of Archaeological Method and Theory* 23:271-304. DOI: 10.1007/s10816-015-9240-4

Dallmeier-Tiessen, Suenje, Robert Darby, Kathrin Gitmans, Simon Lambert, Brian Matthews, Salvatore Mele, Jari Suhonen, and Michael Wilson

    2014   Enabling Sharing and Reuse of Scientific Data. *New Review of Information Networking* 19:16-43. DOI: 10.1080/13614576.2014.883936

Dappert, Angela, Adam Farquhar, Rachael Kotarski and Kirstie Hewlett

    2017   Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research. *Data Science Journal*, 16: 28, pp. 1–16. DOI: 10.5334/dsj-2017-028

Fagan, Brian

    1995   Timelines: Archaeology's Dirty Secret. *Archaeology* 48 (4):14-17.

Faniel, Ixchel, Adam Kriesberg, and Elizabeth Yakel

    2016 Social Scientists' Satisfaction With Data Reuse. *Journal of the Association for Information Science and Technology* 67:1404-1416. DOI: 10.1002/asi.23480

Faniel, lxchel, Julianna Barrera-Gomez, Adam Kriesberg, and Elizabeth Yakel

    2013 A Comparative Study of Data Reuse Among Quantitative Social Scientists and Archaeologists. *iConference 2013 Proceedings*, pp797-800. Electronic document, http://hdl.handle.net/2142/42099, accessed August 9, 2017.

Faniel, Ixchel, Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel

    2013   The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries (JCDL '13), pp.* 295-304. ACM: New York, NY. DOI: 10.1145/2467696.2467712

Frank, Rebecca D., Elizabeth Yakel, and Ixchel Faniel

    2015   Deconstruction/Reconstruction: Preservation of Archaeological and Zoological Research Data. *Archival Science* 15(2):141-167. DOI: 10.1007/s10502-014-9238-9

Gattiglia, Gabriele

    2015   Think Big About Data: Archaeology and the Big Data Challenge. *Archäologische Informationen* 38:113-124.

Green, Katie

    2016   Digital Data in Archaeology: Preservation, Access and Re-Use – The Work of the Archaeology Data Service. In *Proceedings of the 2nd Data Management Workshop - Kölner Geographische Arbeiten 96*, edited by Constanze Curdt and Christian Willmes, pp17-23. Geographisches Institut der Universität zu Köln: Cologne. DOI: 10.5880/TR32DB.KGA96.4

Haustein, Stefanie

    2016   Grand Challenges in Altmetrics: Heterogeneity, Data Quality and Dependencies. *Scientometrics* 108:413-423. DOI 10.1007/s11192-016-1910-9

Heidorn, P. Bryan

    2008   Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends* 57:280-299 DOI: 10.1353/lib.0.0036

Huggett, Jeremy

    2014   Promise and Paradox: Accessing Open Data in Archaeology. In *Proceedings of the Digital Humanities Congress 2012*, edited by Clare Mills, Michael Pidd, and Esther Ward. HRI Online Publications: Sheffield. Electronic document, https://www.hrionline.ac.uk/openbook/chapter/dhc2012-huggett, accessed August 9, 2017.

    2015   Digital Haystacks: Open Data and the Transformation of Archaeological Knowledge. In *Open Source Archaeology: Ethics and Practice*, edited by Andrew T. Wilson and Ben Edwards, pp. 6-29 (Berlin: De Gruyter Open). DOI: 10.1515/9783110440171-003

Huvila, Isto

    2016   If We Just Knew Who Should Do It, Or The Social Organization of the Archiving of Archaeology in Sweden. *Information Research* 21(2). http://www.informationr.net/ir/21-2/paper713.html, accessed August 9. 2017.

Jagadish, H.V.

    2015 Big Data and Science: Myths and Reality. *Big Data Research* 2(2):40-52. DOI: 10.1016/j.bdr.2015.01.005

Jones, Emily Lena, and Caroline Gabe

    2015   The Promise and Peril of Older Collections: Meta-Analyses and the Zooarchaeology of Late Prehistoric/Early Historic New Mexico. *Open Quaternary* 1:1–13. DOI: 10.5334/oq.ag

Kandel, Sean, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono

    2011 Research Directions in Data Wrangling: Visualizations and Transformations for Usable and Credible Data. *Information Visualisation* 10:271-288. DOI: 10.1177/1473871611415994

Kansa, Sarah Whitcher

    2015   Using Linked Open Data to Improve Data Reuse in Zooarchaeology. *Ethnobiology Letters* 6:244-231. DOI: 10.14237/ebl.6.2.2015.467

Kansa, Sarah Whitcher, and Eric Kansa

    2014   Data Publishing and Archaeology's Information Ecosystem. *Near Eastern Archaeology* 77:223-227.

Kansa, Eric, Sarah Whitcher Kansa, and Benjamin Arbuckle

    2014   Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology. *International Journal of Digital Curation* 9:57-70. DOI: 10.2218/ijdc.v9i1.301

Kitchin, Rob

    2014   *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications: London.

Markus, Lynne M.

    2001   Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. *Journal of Management Information Systems* 18:57-93. DOI: 10.1080/07421222.2001.11045671

Marwick, Ben

    2017   Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory* 24:424-450. DOI: 10.1007/s10816-015-9272-9

Mauthner, Natasha S., and Odette Parry

    2009   Qualitative Data Preservation and Sharing in the Social Sciences: On Whose Philosophical Terms? *Australian Journal of Social Issues* 44:289-305.

Mayer-Shönberger, Viktor, and Kenneth Cukier

    2013   *Big Data: A Revolution That Will Transform How We Live, Work and Think*. John Murray: London.

Merriman, Nick, and Hedley Swain

    1999   Archaeological Archives: Serving the Public Interest? *European Journal of Archaeology* 2:246-267. DOI: 10.1179/eja.1999.2.2.249

Mooney, Hailey, and Mark P. Newton

    2012   The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication* 1(1):eP1035. DOI: 10.7710/2162-3309.1035

Mudge, Mark

2012   Transparency for Empirical Data. In *Paradata and Transparency in Virtual Heritage*, edited by Anna Bentkowska-Kafel, Hugh Denard, and Drew Baker, pp. 177-188. Ashgate: Farnham, Surrey.

Müller, Heiko, and Johann-Christoph Freytag

2003 *Problems, methods and challenges in comprehensive data cleansing*. Technical Report HUB-IB-164, Humboldt-Universität zu Berlin (Berlin: Institut für Informatik). Electronic document, http://www.dbis.informatik.hu-berlin.de/fileadmin/research/papers/techreports/2003-hub_ib_164-mueller.pdf, accessed August 9, 2017.

Naylor, John, and Julian D. Richards

2005   Third-Party Data for First Class Research. *Archeologia e Calcolatori* 16:83-91. Electronic document, http://www.archcalc.cnr.it/indice/PDF16/NAYLOR83-92.pdf, accessed August 9, 2017.

Park, Hyoungjoo, and Dietmar Wolfram

2017   An Examination of Research Data Sharing and Re-use: Implications for Data Citation Practice. *Scientometrics* 111:433-461. DOI 10.1007/s11192-017-2240-2

Pasquetto, Irene V., Bernadette M. Randles, and Christine L. Borgman

2017   On the Reuse of Scientific Data. *Data Science Journal* 16:art8. DOI: 10.5334/dsj2017-008

Peters, Isabella, Peter Kraker, Elisabeth Lex, Christian Gumpenberger, and Juan Gorraiz

2016   Research Data Explored: An Extended Analysis of Citations and Altmetrics. *Scientometrics* 107:723-744. DOI 10.1007/s11192-016-1887-4

Piwowar, Heather A., and Todd J. Vision

2013   Data Reuse and the Open Data Citation Advantage. *PeerJ* 1:e175. DOI 10.7717/peerj.175

Richards, Julian D.

2002   Digital Preservation and Access. *European Journal of Archaeology* 5:343-366. DOI: 10.1179/eja.2002.5.3.343

2017 Twenty Years Preserving Data: A View from the United Kingdom. *Advances in Archaeological Practice*, in press. DOI: 10.1017/aap.2017.11

Richards, Julian D., John Naylor, and Caroline Holas-Clark

2009   Anglo-Saxon Landscape and Economy: Using Portable Antiquities to Study Anglo-Saxon and Viking Age England. *Internet Archaeology* 5. DOI: 10.11141/ia.25.2

Robinson-Garcia, Nicolas, Philippe Mongeon, Wei Jeng, and Rodrigo Costas

2017   DataCite as a novel bibliometric source: Coverage, strengths and limitations. *Journal of Informetrics* 11: 841-854. DOI: 10.1016/j.joi.2017.07.003

Sasse, Tom, Amanda Smith, Ellen Broad, Jeni Tennison, Peter Wells, and Ulrich Atz

2017   *Recommendations for Open Data Portals: from setup to sustainability*. Electronic document, https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf, accessed August 9, 2017.

Smith, Michael E.

2015   How Can Archaeologists Make Better Arguments?. *The SAA Archaeological Record* 15(4):18-23.

Tuomi, Ilkka
    1999   Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. *Journal of Management Information Systems* 16(3):103-117.

Wallis, Jillian, Elizabeth Rolando, and Christine L. Borgman
    2013   If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE* 8(7):e67332. DOI:10.1371/journal.pone.0067332

Wentworth, Donna
    2008   Voices from the future of science: Rufus Pollock of the Open Knowledge Foundation. *Science Commons* (blog), August 18th, 2008. Electronic document, http://sciencecommons.org/weblog/archives/2008/08/18/voices-from-the-future-of-science-rufus-pollock-of-the-open-knowledge-foundation/, accessed August 9, 2017.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, Barend Mons
    2016   Principles for scientific data management and stewardship. *Scientific Data* 3: article no. 160018. DOI: 10.1038/sdata.2016.18

Wylie, Alison
    2017   How Archaeological Evidence Bites Back: Strategies for Putting Old Data to Work in New Ways. *Science, Technology, & Human Values* 42:203-225. DOI: 10.1177/0162243916671200

Zimmerman, Ann S.
    2008   New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology, & Human Values* 33:631-652. DOI: 10.1177/0162243907306704
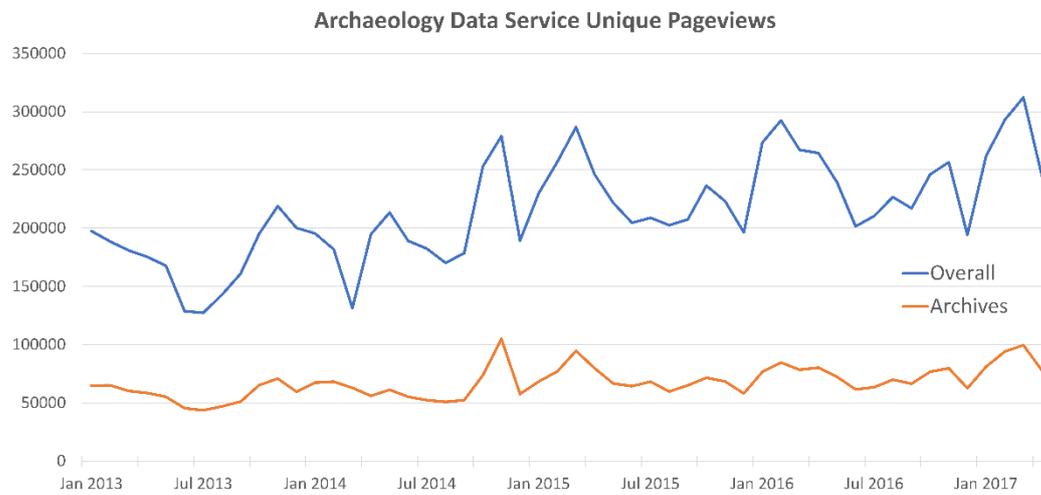
Figures

Figure 1. Archaeology Data Service unique pageviews, comparing overall views with archive views. (Data provided by the Archaeology Data Service)
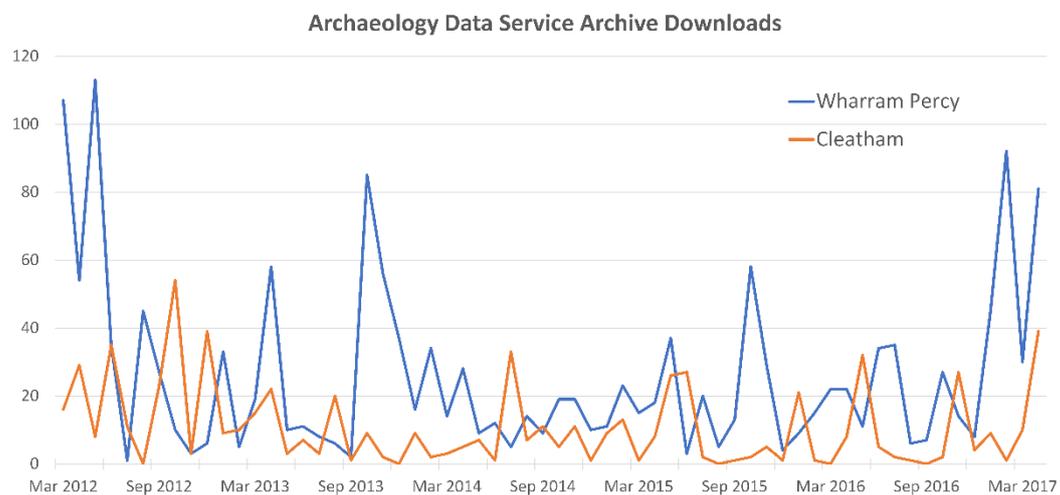

Figure 2. A comparison of data archive downloads from the Archaeology Data Service. (Data downloaded from individual Archaeology Data Service archive pages)
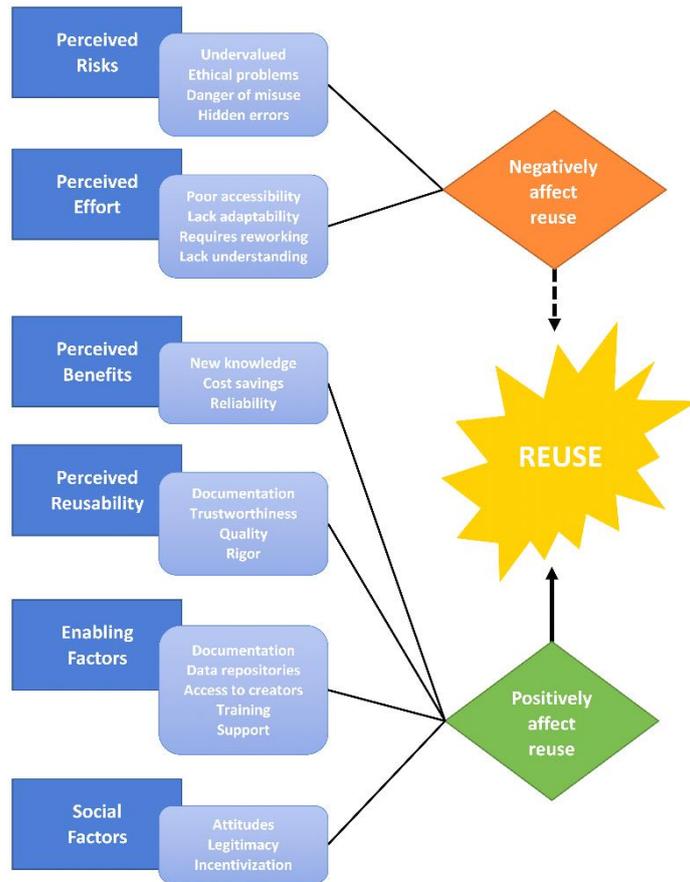
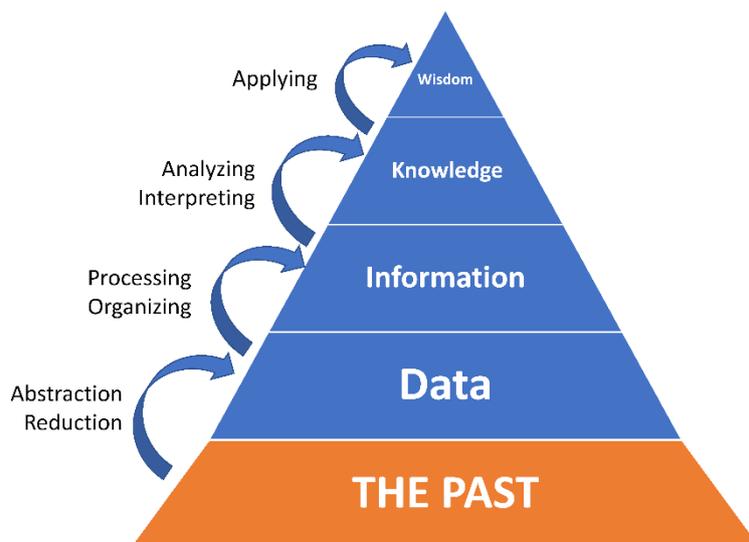Figure 3. A data reuse behavioral model (adapted from Curty and Qin 2014, figure 1)

Figure 4. An archaeological variant of the Data-Information-Knowledge-Wisdom pyramid (adapted from Kitchin 2014, figure 1.1)
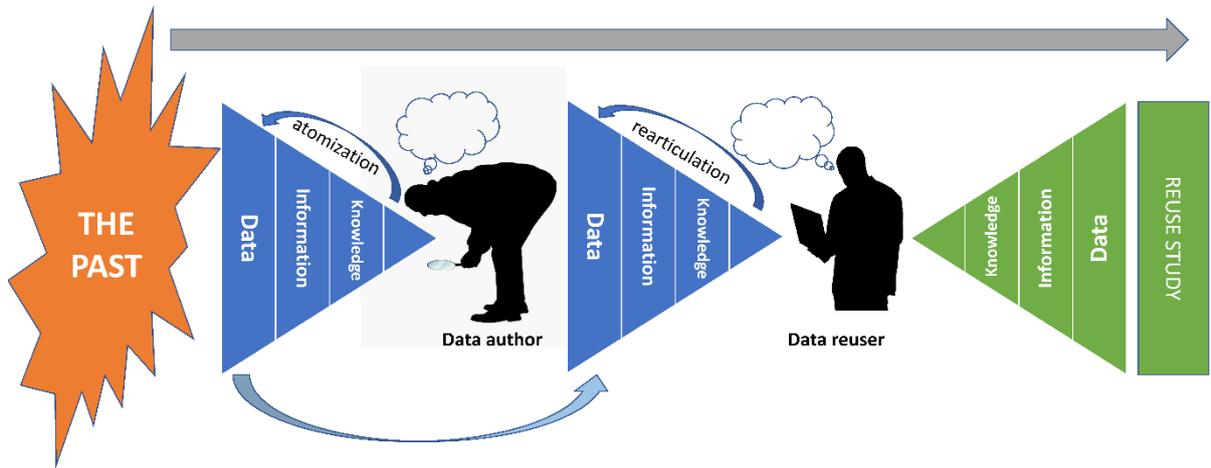
Figure 5. A modified Data-Information-Knowledge model of archaeological data reuse, emphasizing the relationship between the reuser and the creator.
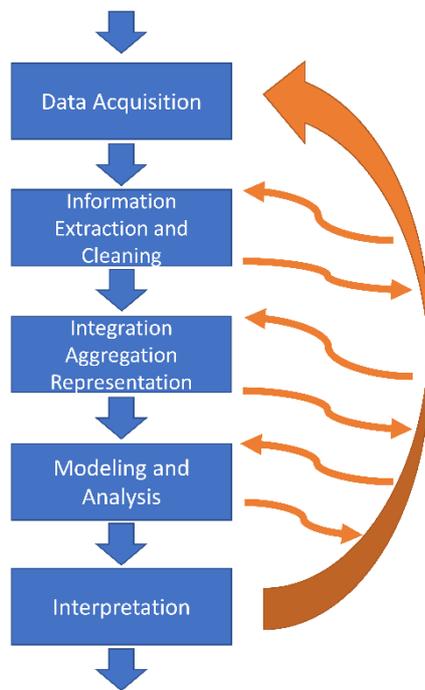

Figure 6. Data reuse cycles and feedbacks (adapted from Jagadish 2015, figure 1)