Andrew T. Wilson and Ben Edwards (Eds.)

Open Source Archaeology

Ethics and Practice

Andrew T. Wilson and Ben Edwards (Eds.)

Open Source Archaeology: Ethics and Practice

Managing Editor: Katarzyna Michalak

Associate Editor: Łukasz Połczyński

DE GRUYTER
OPEN

Published by De Gruyter Open Ltd, Warsaw/Berlin
Part of Walter de Gruyter GmbH, Berlin/Munich/Boston

ISBN 978-3-11-044016-4
e-ISBN 978-3-11-044017-1

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

www.degruyteropen.com

Cover illustration: © Łukasz Połczyński

Contents

Acknowledgements

We would like to use this section to acknowledge the help and support received during the course of this volume. Firstly we would like to thank each of the authors, without your continuing research and contributions to the field of archaeology and open access this volume would have been very empty. Secondly we would like to thank Dr Seren Griffiths and Dr Helen Miles for all the help during the production of this volume, without you this volume would have never been completed within the timely manner with which it was. Finally we would like to thank you the reader for taking the time to look at the volume and the snapshot into Open Archaeology this volume provides.

Andrew

Benjamin Edwards and Andrew T. Wilson

1 Open Archaeology: Definitions, Challenges and Context

1.1 Introduction

Over the last two decades archaeology has slowly been going through an information revolution, affecting the ways in which it is researched and published. These changes have come about as a result of an idea: being ‘open’. Open source software, open access to archaeological data and open ethics. ‘Open’ has become an increasingly attractive thing to be; from research, to corporations and governments. Openness gives an air of transparency, ideas of public accountability and scientific repeatability, and as such provides a buzzword for perceived public good (Costa et al., 2014; Lake, 2012). In this volume, the term ‘open’ is given a specific definition:

> “A piece of content or data is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and/or share-alike.” (Open Definition)

1.2 ‘Open Source’ Archaeology and ‘Open’ Archaeology

Although based on the same ideas of openness, open source archaeology and open archaeology have come to mean very different things. Open source archaeology comes from the open source software example of the computer sciences; whereas open archaeology emerges from the concepts of open publishing and free access to archaeological datasets.

1.3 Open Source Archaeology

‘Open-source software’ is a term used to describe computer programs that are distributed as readable program source code - statements written in a (high-level) programming language. This availability of the source codes allows the end user to not only run the final program but manipulate, change, redevelop and understand how the underlying functionality of the program works. FOSS, free and open-source software, is not just the software itself but also a repository of knowledge for the tool (Ducke 2012 and see Ducke this volume: Chapter 7).

Benjamin Edwards: Manchester Metropolitan University, Manchester, UK
Andrew T. Wilson: Bangor University, Bangor, UK

1.4 Open Archaeology

Open archaeology, in contrast, is focused on ensuring datasets and publications are freely available for use by the wider academic community and the public. Data and publication are two related but distinct strands in this movement, though both are now coming to wider attention. Open publication (see below) has been actively placed on the agenda by both national and European governments in recent years, with public policy now backing the idea that publically funded research should be freely available to the public, although this is currently limited to journal articles and conference proceedings. Open data is at the same time a very old fashioned but also radical idea. It has long been a principle of the natural sciences that experiments should be reproducible, and that datasets should therefore be available to other researchers. The radical element here, and that which is contributing to the 'open' movement, is the nature of this access. It has been a slow start, but databases are now becoming available online in raw and unprocessed form, be these statistical, excavation archive, GIS-based survey or image/3D data archives.

As examples, services such as the UK Archaeology Data Service, Open Context and the Digital Archaeological Record have been pioneering the sharing of archaeological data via the internet, with licenses that encourage re-use. Private bodies such as Oxford Archaeology and Wessex Archaeology have started to also make their grey literature available as an open archive (Costa et al., 2014). This breaks away from the model of data storage, where it is nominally accessible and in reality subject to controlled release by individuals or research organisations. Now the data is becoming accessible. However, whilst the movement is gathering pace, it is still rare to see published datasets alongside finished articles, and also rare to be given a dataset required to reproduce an analysis and produce results. Thus it is possible to define two distinction models of data sharing: dynamic datasets that continue to be updated, versus static datasets that are released once as a finished resource.

1.5 The Public Context of Open Access

The move towards open software and open archaeology is not occurring in a vacuum. Open-access is a trend that cuts across disciplinary boundaries, and is also finding support in political and policy-making spheres, reflected in the priorities of research funding bodies. In the UK, HEFCE (the Higher Education Funding Council for England), the AHRC (Arts and Humanities Research Council), and the ESRC (Economic and Social Research Council) have recently published a new policy on open access to scholarly research. This states, amongst other things, that the content of all peer-reviewed journal articles and conference proceedings must be made available as open access through institutional repositories once an embargo period has elapsed, with implications for further funding eligibility if this requirement is not met (Higher Education Funding Council for England, 2014). Whilst this policy does not yet apply to monographs or data, it surely illuminates a trend that will only continue to gather momentum in coming years, and one for which the academic community must be prepared.

This output-based priority for open access is also reflected in project-design phase in the form of requirements placed on new projects in the arts and social sciences. It is increasingly common to find funding calls from major European research bodies placing an emphasis on public engagement and access to data. Ring-fenced funding calls, such as the AHRC's 'Connected Communities' programme or the European 'Horizon 2020' scheme, stressing the connections between academic research and public access, are aimed at bridging the gap between the production of academic knowledge and its impact in society - open-access to both data and interpretation are seen as key to success in these programmes. Similarly, the move toward measuring the 'impact' of academic research is relevant in this regard. Whilst open access is not the only tool to ensure 'impact' (defined as the extent to which the results of research make a difference to society, culture or policy-making outside of the academy), it is seen as an important part of any strategy that attempts to engage the wider world in the practice or results of research. How open access contributes to measures of impact differs across the sector, but can include the direct public participation in data collection through to the public availability of research outputs in a manner that his engaging and aimed at a non-specialist audience.

This movement toward open access has not been entirely philanthropically motivated however. Whilst individual researchers are clearly committed to the ideals of open research and open access, as the content of this volume testifies, it is certainly true that the public mood toward academic research is also changing. This has found its expression in the UK recently, with the government's response to Finch Group report (BIS 2012), a report into open access in UK academia by Dame Janet Finch, recommended the removal of paywalls surrounding published academic research that was funded by the taxpayer through the UK research councils. Universities will now be expected to pay the costs of open access up front. Unsurprisingly, this move was illustrative of wider international trends, with 2012 seeing the European Research Council setting out a new policy on open access to research. Research funded by the ERC must now be made available as open access within six months of its publications date (European Research Council, 2012).

On a broader socio-cultural level, it is possible that these policy-based movements in open access are reflecting trends with other roots. In the UK, recent economic difficulties have either prompted, or been used as an excuse for (the choice here is left to the discretion of the reader), changes to the way in which University courses are funded. Heritage and archaeology have suffered alongside other social sciences because they do not fit the 'STEM' agenda of science, technology, engineering and maths', losing government subsidy as a result. The humanities and social sciences find themselves in the (some would say) ridiculous position of being forced to justify their own existence, with the additional difficulty that the 'debate' is framed in a utilitarian language that presupposes the greater importance of the hard sciences. 'Impact' in this context becomes one of the measures, either pernicious or otherwise, of this justification of existence. Whether one agrees with these changes in policy and attitude toward the social sciences is not strictly relevant for this volume. What is clear, however, is that in the new academic, financial and socio-cultural environment that researchers find themselves, open access to research is now not only a moral imperative - it is increasingly vital for the survival of meaningfully funded research.

1.6 Open Ethics

With the development of 'open' access within archaeology, a new set of open ethical issues surrounding the use and distribution of the data have come to light. These focus on the types of data that are made 'open' and, critically, its quality. Unpublished research is a key area where open data may have a transformative impact, but also an area where ethical considerations become of relevance. There are various reasons for non-publication, but the sheer cost of bringing archaeological research (especially fieldwork) to formal publication is often a key issue, and there is a significant backlog of mid-twentieth century excavation unpublished in the UK alone. In such cases, we know that some data exists, even if in a raw and unprocessed form (most likely a paper archive). While some might argue that publishing such data without offering a synthetic overview alongside would make for a very limited resource, it is undoubtedly better to have access to data than to have nothing at all. It seems that the greater evil would be to allow such data to remain utterly unpublished, given the relatively low cost of photographically digitising paper archives.

Opening up of grey data (i.e. that collected and published as part of commercial archaeology, usually associated with the planning process, see Huggett, this volume: Chapter 2) is in the view of many, as a way to meet the minimal requirements to publish research (Costa et al., 2014). If these publications, as part of these requirements, had to include the raw research data this would be major step forward for open data. The current view is that a publication is more of a report on the research rather than the total outcome from archaeological activity, as such presented with limited interpretation and, lacking all but the presented data. Reasons for this disjunction are varied, and depend on national context, but a major problem is the cost of making data available (both via publication and permanent online resources) for commercial companies involved in producing grey literature, who work on tight budgets and cannot justify non-statutory expenses. We need to transform our understanding of what constitutes full and satisfactory publication, but accept that this will come with an attached cost. This is one of the great challenges of the open data movement - how to involve commercial data producers. What is certain, is that this involvement should be an ethical imperative (Costa et al., 2014).

1.7 Outline of the Volume

This volume came about because of a perceived lack of published works on the ideas surrounding ‘open’ archaeology. The transformation to open access in archaeological data has not been examined in sufficient detail. Through a series of papers this volume sets out to examine, not only the open archaeological software currently being employed and open access to archaeological information, but also the emerging change in culture and ethics this ‘open’ revolution is producing. As such, this volume has three mains themes throughout; open source software, open archaeological data, and open ethics. Each paper touches on at least two or all of these themes, and as such they are hard to pigeonhole. What this volume also demonstrates is the breadth of work being undertaken with an open ethic, and the commitment of individuals and teams of researchers driven by personal belief to be at the forefront of an emerging field, actively creating and shaping a vision of open archaeology that will be an important legacy in the future.

Bibliography

Costa, S., Beck, A., Bevan, A. and Ogden, J. (2014), Defining and advocating open data in archaeology, *in* ‘Archaeology in the Digital Era: Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology’.

Ducke, B. (2012), ‘Natives of a connected world: free and open source software in archaeology’, *World Archaeology* **44**(4), 571–579.

European Research Council (2012), ‘European research council open access guidelines for researchers’.
URL: http://erc.europa.eu/sites/default/files/document/file/open_access_policy_researchers_funded_ERC.pdf

Higher Education Funding Council for England (2014), ‘STEM teaching capital funding allocation’.
URL: http://www.hefce.ac.uk/pubs/year/2014/cl022014/name,85551,en.html

Lake, M. (2012), ‘Open archaeology’, *World Archaeology* **44**(4), 471–478.

Jeremy Huggett

2 Digital Haystacks: Open Data and the Transformation of Archaeological Knowledge

> "There is a great need for theorization precisely when emerging configurations of data might seem to make concepts superfluous to underscore that there is no Archimedean point of pure data outside conceptual worlds. Data always has theoretical enframings that are its condition of making ..."(Boellstorff, 2013).

2.1 Introduction

Since the mid-1990s the development of online access to archaeological information has been revolutionary. Easy availability of data has changed the starting point for archaeological enquiry and the openness, quantity, range and scope of online digital data has long since passed a tipping point when online access became useful, even essential. However, this transformative access to archaeological data has not itself been examined in a critical manner. Access is good, exploitation is an essential component of preservation, openness is desirable, comparability is a requirement, but what are the implications for archaeological research of this flow – some would say deluge – of information? Lucas has recently pointed to the way archaeological reality can change as a consequence of intervention: as archaeologists change their mode of intervention so reality shifts and interpretations change (Lucas, 2012, p. 216). If this is true of archaeological practice, to what extent might the change in our relationship to data – the move from traditional modes of creation and access to digitally-enhanced methods – represent a potential paradigm shift in our archaeological reality, or place limits on future changes? As more data are 'born digital' with access to them open to an increasingly wide audience, is it realistic to assume that archaeological knowledge itself remains unchanged in the process? How does our relationship with archaeological data change as the observations, measurements, uncertainties, ambiguities, interpretations and values encapsulated within our datasets are increasingly subject to scrutiny, comparison, and re-use? What are the implications of increasing access to increasing quantities of data drawn from different sources which are more or less open, more or less standardised, and increasingly reliant on search tools with greater degrees of automation and linkage? Given the fundamental – and frequently contested – nature of archaeological data, it is surprising that the implications of open access to those data remain largely uncontested. Instead, archaeology's digital haystack repre-

Jeremy Huggett: University of Glasgow, Glasgow, UK

sents a largely unexplored set of practices mixing old and new in the creation of new infrastructures which transform the packaging, presentation, and analysis of the past. Examining this entails revisiting the notion of the 'archaeological record' within the context of the new technological frameworks, and considering the consequences of this digital data intervention.

2.2 Openness and Access

Open archaeology has been a concept receiving increasing attention in recent years, most evidently in an issue of World Archaeology which sought to extend awareness of the implications of open approaches to a wider archaeological audience (Lake, 2012, p. 471). As Lake observes, and as reflected in that issue and this volume, openness can cover the use and reuse of software, publications, creative works, and data, although within the archaeological debate attention has until recently focussed extensively, though not exclusively, on publication.

The most common starting point for considering 'openness' is the Open Definition: "A piece of data or content is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike." (Open Definition, 2014). Archaeology may seem to be well-served with free access to archaeological data via organisations such as the Archaeology Data Service in the UK, tDAR and Open Context (USA), DANS (Netherlands), as well as national heritage organisations (for example, Royal Commission on the Ancient and Historical Monuments of Scotland, English Heritage) and regional Historic Environment Records. However, with some exceptions, much of this data is only partially 'open', leaving Kansa to suggest that openness remains largely at the margins of archaeological practice (2012, p. 499). In part, this is a consequence of distinctions between different levels of 'open access data' and 'open data'. For example, a hierarchy can be defined in increasing order of 'openness':

1. Open access data which provides online access to view datasets, limited only by a presumption of Internet access and the requirement for a modern web browser. Use of the data beyond viewing and searching online is restricted (commonly seen with most Historic Environment Records, National Monuments data and including commercial organisations such as CyArk etc.). A variant of this approach enables a map to be created on demand within desktop GIS software. This generally entails access to Web Mapping Services (WMS) which provide a graphical image as output, with limited functionality beyond the image itself. These are typically available for National Monuments data accessed via open government websites such as data.gov.uk.
2. Open access data which returns summary geographical information as a downloadable output of a search query or via Web Feature Services (WFS). This can then be further analysed using GIS software as if the data were held locally. For

example, the Archaeology Data Service's ArchSearch has download functionality for registered users, and Historic Scotland/RCAHMS's PastMap similarly enables summary location data to be accessed via downloadable comma-separated values files. Currently most WFS feeds in archaeology are used internally within organisations, or to create interoperable services from multiple feeds (resources such as PastMap itself, and Scotland's Places) but are not accessible more widely (for example, McKeague et al. (2012)). Leaving technical issues aside, in part this seems to arise out of concern to limit bulk downloads of data: hence downloads from ArchSearch or PastMap are restricted to one or two hundred records at a time, for example.

3. Open access data consisting of entire datasets which can be downloaded but where restrictions apply to the use and reuse of data and hence is not truly open data in the technical sense. For example, the Archaeology Data Service Common Access Agreement (Archaeological Data Service, n.d.) specifies that the data should only be used for teaching, learning, and research purposes, although the definition of 'research' is drawn very broadly such that it includes commercial funding, and the primary condition is that the results are placed in the public domain. In other cases, the restriction is more of a 'health-warning': for instance, the PastMap terms and conditions specify that the data provided is intended for information only and that professional advice should be sought to properly interpret it, emphasising the need to understand its limitations (PastMap, 2013). On the other hand, English Heritage's Heritage Gateway applies strict copyright restrictions to data accessed and downloaded from the site (Heritage Gateway, 2007).
4. Open data which has no exclusions or restrictions on use, and conforms to the Open Definition or the most permissive Creative Commons licenses. In general these datasets relate to specific projects, sites, or collections. For example, in the United States both Open Context and tDAR organisations use the Creative Commons CC-BY licence which enables the data to be shared and reworked, simply requiring attribution or citation of the original work. As Kansa points out, certain datasets within the Archaeology Data Service collections are now also governed by the CC-BY license rather than the standard terms and conditions (Kansa, 2012, p. 507).

Much archaeological data therefore is not truly 'open', and recent papers on open data in archaeology tend to focus on the desirability of increasing openness and the restrictions and impediments to achieving it (for example, Beale 2012; Beck and Neylon 2012; Bevan 2012*b*; Kansa 2012). These are not new issues: for example, in a discussion of copyright and archaeological data in 1997 Carson asked: "Who owns the right to reproduce raw data? Who owns the right to publish a manipulated version of that data? And who owns the right to produce second-generation items, such as models, from that data?" (Carson, 1996, p. 291). The ethical responsibility of archaeologists to make their data available is frequently cited: for example, Carson argues that:

> “Archaeologists, like other scientists, have an ethical obligation to publish, and to allow others to critique, their findings. Publishing data sets in machine-readable form is the ultimate expression of this obligation, in that others are free to analyze the basis of an archaeologist’s findings and come to their own conclusions.“ (Carson, 1996, p. 316).

Kansa puts the case more strongly, arguing that “the discipline should not continue to tolerate the personal, self-aggrandizing appropriation of cultural heritage that comes with data hoarding” (2012, p. 507) and goes on to say:

> “Failure to incentivize greater data transparency would demonstrate an egregious failure of leadership and utter dysfunction in a discipline supposedly devoted toward building and preserving knowledge of the past.” (2012, p. 507).

Most professional archaeology codes of practice emphasise this link between the stewardship of the past and the requirement to report and publish and to preserve the records made, including computer data. For example, the Institute for Archaeologists in the UK specifies that the results of archaeological work should be made available with reasonable dispatch (Institute for Archaeologists, 2013, Principle 4) and establishes that this includes the analysis and publication of data (Institute for Archaeologists, 2013, 4.4). In the light of this it would be tempting to ask why more open data is not available. One reason may be that the ethical codes emphasise that rights of primacy exist: in the case of both the IfA and the European Association of Archaeologists this persists for up to ten years (Institute for Archaeologists 2013, 4.4; European Association of Archaeologists (1997, 2.7)), although the Archaeological Institute of America, the Society for American Archaeology, and the Canadian Archaeological Association, for example, only specify the need to make results available in a timely fashion and to make evidence available to others within a reasonable time (of America 2008, I.4; Society for American Archaeology (1996, 5); Canadian Archaeological Association (n.d.)). Consequently rights of primacy may restrict access to data and, without enforcement, the timescales specified may be stretched: indeed, there is a long and unfortunate history of archaeological archive data being retained by an individual for a lifetime. In such a context, Kansa’s expostulation is understandable.

One issue regularly raised in relation to open archaeological data is that they frequently include spatial information which might facilitate looting (for example, Bevan 2012*b*, p. 7–8; Kansa 2012, p. 508–509). Degrading the quality of spatial data and making full resolution data available only to ‘approved’ users are approaches that have been adopted, but restricting access like this flies in the face of open data requirements. Other common arguments about the limits to open data relate to authority and the risk of reducing confidence as a consequence of revealing discrepancies and errors in the data. With datasets consisting of millions of records in some cases, it would be surprising if errors did not creep in, especially as the data are increasingly manipulated by automated means. Whether this damages the authority of the data is open to question: arguably issues with the data such as different levels of precision of locational information are likely to be more problematic for would-be users than the occasional rogue item.

2.3 Openness and Reuse

In the light of the pressures for access to open data it is perhaps worth emphasising that there has been no empirical study of the demand for open data in archaeology. This means that, to a large extent, the level of demand remains undemonstrated and unquantified. However, a recent study of the Archaeology Data Service sought to evaluate and quantify the 'value' of online access to data (Beagrie and Houghton, 2013). It employs a range of approaches to assessing value: for example, investment value (amount invested in the services), use value (amount spent by users to access the service), contingent value (for instance, how much people would be willing to pay). In combination these give rise to the net economic value (the difference between the willingness to pay and the cost of obtaining the service minus the investment value) (Beagrie and Houghton, 2013, figure 4.1). On this basis, the investment value of the Archaeology Data Service was calculated to be about £1.2m per annum, made up of £698,000 from funders or sponsors and around £465,000 indirectly contributed by depositors (Beagrie and Houghton, 2013, p. 35). Direct use value to the user community was estimated to be about £1.4m per annum (Beagrie and Houghton, 2013, p. 35) but the efficiency impacts were estimated to be anywhere between £13m and £58m per annum (Beagrie and Houghton, 2013, p. 40). Research efficiency gains were equivalent to around 7 hours per week as a consequence of access to ADS data (Beagrie and Houghton, 2013, p. 39). Interestingly, there were objections to the survey's use of questions about willingness to pay for the service and how much people would be willing to accept in return for giving up the service, and 6-9% of respondents refused to estimate this, arguing that access and data should be free (Beagrie and Houghton, 2013, p. 36–37). The results show that the value of access to data is considerable – however, as with everything 'open', the challenge is to make openness sustainable financially.

The extent to which open access data is actually used also remains largely unquantified. Ironically, access to data about access to open archaeological data is often not directly accessible; however the Archaeology Data Service website provides statistics for a variety of metrics and, as one of the longest-established providers of a broad range of archaeological data, could reasonably be viewed as representative. Web metrics are notoriously difficult to disentangle and interpret, but the evidence suggests a surprisingly high number of downloads relative to visits to the site (Figure 2.1). Much of this relates to downloads of PDF files from the large collections of unpublished grey literature reports and back-issues of journals and other volumes (Green pers comm – Figure 2.2), rather than downloads of specific datasets.

The Archaeology Data Service download statistics do not differentiate between PDF and other file types, so estimating usage of datasets is not straightforward. How-

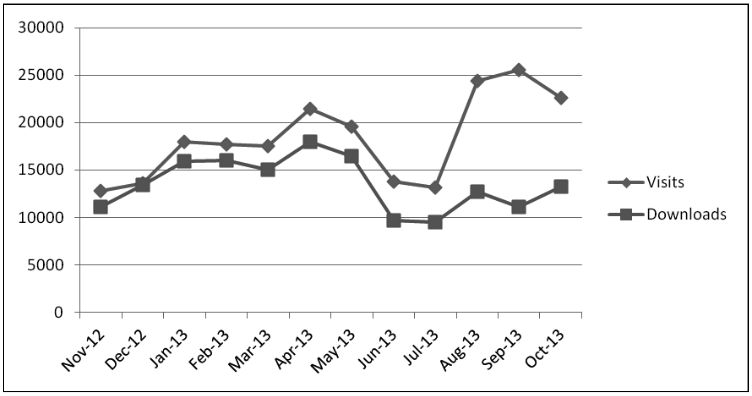


Figure 2.1: Archaeology Data Service access statistics

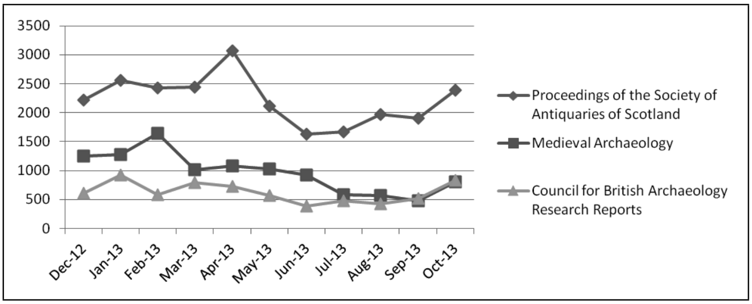


Figure 2.2: ADS access statistics - examples of PDF downloads

ever, Figure 2.3 provides an approximate comparison to Figure 2.2 based on examples of field projects which include downloadable data, simply to demonstrate the order of magnitude difference between field data downloads and PDF downloads. The reason for this difference may be simply that the majority of PDFs relate to free access to back issues of journals and volumes that would otherwise require subscription or purchase, or access to grey literature about excavated sites that would be costly in time and effort to acquire otherwise (for example, Bradley 2006, 7–8), while the field datasets require a very specific level of interest and, to some extent, expertise. Clearly, there is much more to be gained from a deeper and more nuanced analysis of these kinds of access data.

Issues with open data (and non-open data, for that matter) really come to the fore only when those data are put to analytical use. Detailed accounts of data reuse are as yet rare, and those reports there are tend to stress the positive outcomes and minimise the efforts entailed in achieving them. For example, Bevan (2012*a*) demonstrates the potential benefits from the examination of several large scale georeferenced inventories and how built-in data biases might be overcome, but apart from reference to an

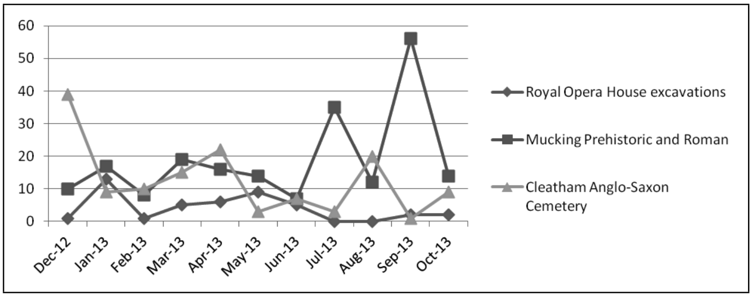


Figure 2.3: ADS access statistics - examples of fieldwork data downloads

"intensive effort of cross-checking and problem-flagging" (2012*a*, p. 493) there is no information provided about any data-cleansing and manipulation that may have been required in advance of analysis. An earlier study using some of the same data provides a clearer indication of the kind of work that can be required to make data usable. The Viking and Anglo-Saxon Landscape and Economy of England (VASLE) project combined data from the Portable Antiquities Scheme database with the Early Medieval Corpus of Coin Finds and an extensive data cleansing exercise was required (Naylor and Richards, 2005; Richards et al., 2008, 2009) to resolve issues of comparability, compatibility, and standardisation of classifications across the two datasets. For example, many dates in the Portable Antiquities Scheme database were only recorded at a generic level, while different recorders classified the same kinds of artefacts under different headings. The level of effort entailed underlines not so much the complexity of the data but the complexity of the task. Unsurprisingly, the researchers concluded that

> "Re-use of data requires a close understanding of the context of data collection and of the vocabulary used to describe the observations. The archaeologist of tomorrow needs training not so much in methods of data collection, but in data analysis and re-use." (Naylor and Richards, 2005, p. 90).

Similar conclusions are reached in a recent study which interviewed a sample of archaeologists about their experience of reusing data and reported that the lack of context was a persistent problem (Faniel et al., 2013). This arose for a variety of reasons, including the variability of archaeologists and their recording procedures which ranged from the meticulous to the careless (Faniel et al., 2013, p. 298). As a consequence, they identify a series of gaps in current archaeological data standards such as the need to capture the range of methodological procedures undertaken during excavation or survey, including specifications of instruments, information about how the data were collected, the strategy adopted, etc. (Faniel et al., 2013, p. 302). However, despite the problems encountered, they note that archaeologists still reused data, sometimes finding alternative means of recovering context – or, presumably, either making assumptions about context or ignoring it altogether. Archaeologists are not unique in this respect. For example, an examination of the reuse of open government data drew attention to the lack of contextual metadata, poor documentation, variability of data quality, conflicting data definitions, and a host of other practical impediments to reuse (Zuiderwijk et al., 2012, p. 162–164). What this highlights is one of the key paradoxes that lies behind open data: increasing access to increasing amounts of data has to be set against greater distance from that data and a growing disconnect between the data and knowledge about that data (Huggett, Forthcoming).

2.4 Approaches to Open Data

One of the problems of open data is that archaeologists are only just starting to consider the issues surrounding open access to archaeological data. Most discussions focus on the desirability of openness, the ethical responsibility to be open, and what benefits might accrue from open access to data for both archaeology, archaeologists, and the wider community. The very diversity of archaeology – its coverage, scope, quantity and range of data sources, multiplicity of practices, and limited standardisation – is often seen as an attraction for e-science studies (e.g. Faniel et al. 2013, p. 295–296; Jeffrey et al. 2009, p. 2515; Richards et al. 2011, p. 42), but the technological responses to this diversity tend to focus on deconstructing archaeological information into semantic structures as a means of managing and controlling the data, a process which itself is not without issues (Huggett, 2012). However, this diversity of archaeological data is not what makes them really distinctive: what is particularly characteristic about archaeological data is their time dimension (what Arbesman 2013) has termed 'long data' in contrast to 'big data') and the peculiarly destructive nature of much of their collection methodology. Individually, neither is especially unique – geology deals with especially 'long data', for instance – but in combination, it makes for an especially challenging prospect for open data. This is because of the conceptual approach to open data and its subsequent reuse.

For example, in a recent definition of what constitutes archaeological open data, Anichini and Gattiglia (2012, p. 54) follow the Italian Association for Open Government (Belisario et al., 2011, p. 11–12) in defining archaeological digital open data as being complete (capable of being exported, used, integrated and aggregated with other data, and including information about their creation), primary ('raw' data capable of integration with other data), timely (available), accessible (free, subject only to costs associated with Internet access), machine-readable (capable of automated processing), non-proprietary (free from licenses that limit their access, use or reuse), reusable, searchable (through catalogues and search engines), and permanent. Unsurprisingly, these do not greatly differ from other open data definitions such as that provided by Open Definition (n.d.).

Although such a characterisation may seem fairly uncontroversial, the concept of the completeness and primacy of the data is problematic from an archaeological perspective since it loses sight of what these data actually are. Completeness and availability may imply that the data are finished and ready for reuse. Beale's reminder that "data must first be prepared and then care taken to identify the moment when we are no longer preparing them for release, but are in fact working on them" (2012, p. 623) distinguishes between data preparation and subsequent analysis but in the process implies the existence of a form of basic un-worked data – 'raw' data in Anichini and Gattiglia's characterisation – which is seen as providing the building blocks for archaeological knowledge. Bevan (2012*b*, p. 6) is suspicious of the use of 'raw data' as a term for something which has a clear interpretative component, but sees the actual problems for open data lying at a higher level with its spatial content (2012*b*, p. 7).

Kansa et al. observe that primary archaeological data have received little theoretical attention while recognising their importance in the production of archaeological knowledge (2010, p. 303). Although it is true that primary data have not been critically discussed within an archaeological informatics context, the nature of archaeological data has been a focus of much debate over the years, recognising that these data are situated, contingent, incomplete, and theory-laden (for example, Patrik 1985; Binford 1987; Barrett 1988; Hodder 1999; Chippindale 2000; Lucas 2001; Lucas 2012). An exception in this regard is Llobera's discussion of data within the context of defining the basis of an Archaeological Information Science (Llobera, 2011), although much of his concern lies with data representation and data structures:

> "...the topic of data representation within archaeology has not received as much attention as it should, especially in the light of the pivotal role it has in the production of archaeological knowledge and its potential to precipitate different interpretations. The consequences of this oversight become deeper and more far-reaching the moment information systems are adopted. It is all too easy for the user to forget that he/she is subscribing to a particular form of data representation." (Llobera, 2011, p. 213–214).

Llobera suggests that archaeologists have generally been concerned with the choice of which data to collect based on prior research questions, rather than the form in which the data are collected (Llobera, 2011, p. 214). Although he recognises that recording data is subject to the theoretical orientation and the goals of the researcher, he argues that the data structures used to contain these archaeological observations are not themselves interpretative and hence:

> "The fact that they organize observations explicitly and that their manipulation is done via a set of operations defined a priori provides transparency and flexibility. Indeed, it is the marriage between data and purpose that make them so powerful and appealing." (Llobera, 2011, p. 215).

Although Llobera's focus on the significance of data structures and their ability to support new forms of archaeological investigation is important, it largely sidelines the origins and nature of data themselves: they become 'reasoning artefacts' that contribute to analysis and interpretation (Llobera, 2011, p. 214). How the data are structured is without doubt crucial to their analysis and any subsequent reuse. However, the significance of the data themselves is equally profound, if not more so.

2.5 From Data to Knowledge?

Within the field of archaeological informatics, responses to issues raised concerning data tend to emphasise structural and organisational approaches and solutions – while the data may be recognised as essentially interpretive, the implications of this are generally left for others to deal with. Consequently a term like 'raw data' is frequently used without reflection and indeed, the term 'data' itself is often open to confusion. In the last things seemed much simpler. For example, Trigger (1998, p. 3) identifies Glyn Daniel, Stuart Piggott and Christopher Hawkes as drawing a clear distinction between facts and interpretations – archaeological data were facts and constituted the core of the discipline, while interpretations were transient and changing. Accordingly the archaeological record was seen to become 'better' as a result of the collection of more data and the development of better techniques for interpreting these data (Trigger, 1998, p. 22). A similar view is held within Information Systems studies, where data are often seen as facts – the raw materials captured within data structures for creating information (for example, Räsänen and Nyce 2013, p. 656), and in the context of 'big data' large datasets are seen increasingly as providing significant opportunities to create new knowledge. In much the same way, the knowledge management industry is predicated on refining data into knowledge (for example, Tuomi 1999, p. 103; Weinberger 2011, p. 2–3).

Superficially, data are not complex. For example, the Royal Society recently defined data as "Numbers, characters or images that designate an attribute of a phenomenon", and as

> "Qualitative or quantitative statements or numbers that are (or assumed to be) factual. Data may be raw or primary data (e.g. direct from measurement), or derivative of primary data, but are not yet the product of analysis or interpretation other than calculation." (Royal Society, 2012, p. 12).

However, this immediately introduces two types of data – 'raw' and 'derived' – and a corresponding contradiction: on the one hand derived data are calculated from other data (for example, average rainfall); on the other hand, calculated data are seen as information (for example, the numbers generated by a survey instrument are data used to calculate the height of a feature which is classified as information) (Royal Society, 2012, p. 14). Not surprisingly, the Report admits that there is sometimes confusion, with data, information, and knowledge being used as overlapping concepts.

One outcome of this more-or-less commonsense technical approach to data is a view of data as sitting at the bottom of a hierarchy which moves from data through information to knowledge (and in some models, to wisdom beyond that). This data-information-knowledge (-wisdom) pyramid (for example, Weinberger 2011, p. 1–5) essentially sees the acquisition of knowledge (or wisdom) as constructed from a series of building blocks: data are used to create information, information combined to generate knowledge. For instance,

> "Data are seen as raw materials for information. Data become information when it is structured and arranged in a particular context or relations set. Information is talked about as though it has a meaning, but no (appended) judgments. It is commonly thought that knowledge contains meaning and judgment and beliefs and commitment regarding a particular action." (Räsänen and Nyce, 2013, p. 659).

From an archaeological perspective, Darvill has expressed concern that such a structure is destabilised by the generation of vast amounts of archaeological data which remain to be turned into information or knowledge (Darvill, 2007, p. 445): an archaeological digital data mountain which increasingly we struggle to deal with (Huggett, 2000, p. 15–16) but which in a world of 'big data' appears much more amenable. When presented with access to these large quantities of data, the data-information-knowledge approach seems self-evident: we are faced with data which we seek to make sense of and ultimately use to draw conclusions about aspects of the past. This is one of the key benefits identified for open data – the provision of access to fundamental building blocks which will enable us to create new knowledge which would otherwise be much harder – or impossible – to do.

However, this outwardly logical approach disguises a hidden technological agenda: as Weinberger observes, this image of knowledge creation as a pyramid with increasingly fine filters being applied at each level is associated with an Information Age "which has been all about filtering noise, reducing the flow to what is clean, clear and manageable. Knowledge is more creative, messier, harder won, and far more discontinuous" (Weinberger, 2010). One might equally add that information and data are just as messy and creative in nature.

The issue lies with the fundamental nature of data. For example, Borgman points out that data carry very little information in and of themselves: "Data are subject to interpretation; their status as facts or evidence is determined by the people who produce, manage, and use those data." (Borgman, 2007, p. 121). Data have no value – indeed, data do not exist – without some degree of interpretation. In archaeological terms, data are contemporary observations about attributes we consider to have some value in understanding past activities – they are the result of the archaeologist's judgements at the time as to what might be worthy of recording: "all archaeological data are generated by us in our terms" (Binford, 1987, p. 393). The kind of data collected from a given assemblage will vary between individuals depending on a variety of factors including recovery methods and research questions (for example, Atici et al. 2013, p. 665). A perspective of data as 'raw' in the sense of being uncontaminated by methodological and theoretical biases and therefore more likely to result in an accurate outcome (Carson, 1996, p. 316) is therefore a simplistic view of what constitutes data. Indeed, some would claim that 'data' is a misleading word to use in the first place. Both Chippindale (2000) and Drucker (2011) have independently argued that 'capta' is a more appropriate term. Drucker emphasises that 'capta' are taken actively, whereas data are assumed to be a given that can be recorded and observed (Drucker, 2011, p. 3) . Chippindale proposes that data as 'capta' are "things we have ventured forth in search of and captured", with all the associated connotations of hunting and gathering, danger, uncertainty and risk (Chippindale, 2000, p. 605). Both emphasise the creative aspect of data (or capta), that they are partial, selective, and change through time. Data/capta rely on prior knowledge and experience: to capture data requires recognition, identification, and classification in order to be recorded in the first place. Additionally, data may not be easily described and hence receive a decreasing amount of attention, they may not break up into natural units so are highly dependent on the level of analysis applied at the time, and they may not be considered worthy of recognition or capable of capture (Bowker, 2005, p. 141–144). As a result,

> "...we are producing a set of models of the world that – despite its avowed historicity – is constraining us generally to converge on descriptions of the world in terms of repeatable entities, not because the world is so, but because this is the nature of our manipulable data structures" (Bowker, 2005, p. 146).

Data and datasets are therefore of their place and time: they are constrained by the conditions of their creation, all the more so if the question of when data become data is considered. As Borgman points out, in some circumstances data may not be considered to be data until they are cleaned and verified – and how much cleaning and verification is required before they are considered usable data is a question of judgement (Borgman, 2007, p. 183). This is a constant issue for digital archives: the distinction between processed and unprocessed data, and how much processing is 'enough'. So what one person considers data might not be recognised as such by another, in terms both of what is captured and what is not, as well as the extent to which it has been processed. As an example of the problem, Chippendale cites the case of recording rock art where effort went into removing the natural elements from the data, overlooking that the natural features may have been an integral aspect of the art which subsequently required the works to be re-recorded (Chippindale, 2000, p. 608). Of course, the rock art was still there to be re-recorded, which cannot be said for the objects of much archaeological data.

2.6 From Knowledge to Data?

The simple perception of data as the base constituents for the construction of information and knowledge may seem attractive and logical when faced with a technological infrastructure consisting of large quantities of data, but it misrepresents the situation and as a result reuse risks misuse. Making sense of data in computer systems is not a straightforward process:

> "Someone has articulated knowledge using languages and conceptual systems available and – in the case of a computer database – represented the articulated knowledge using a predefined conceptual schema. Someone else then accesses these data and tries to recover their potential meaning." (Tuomi, 1999, p. 111).

In order to make sense, the data-information-knowledge model should actually be reversed (for example, Knox 2007; Tuomi 1999) such that data are seen to emerge only as a consequence of knowledge and information; in other words, data come into existence in the first place through human engagement. This is all the more true in the context of digital data: "Data can emerge only if a meaning structure, or semantics, is first fixed and then used to represent information" (Tuomi, 1999, p. 107). Tuomi argues that knowledge has to be articulated in order to become information which can be represented; in order for it to be represented in a digital environment, information needs to be broken down into atomic elements, or data (Tuomi, 1999, p. 107) – a situation familiar to anyone who has constructed a database from scratch. The problem here is that the knowledge that is articulated and atomised is by definition explicit and more easily represented and communicated than contextual tacit knowledge. Tacit knowledge is more easily displayed or exemplified as practice rather than transmitted (Duguid, 2005, p. 113) and therefore tends to be more or less invisible in a digital data environment. 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" (2007, p. 167). The data are therefore accessed in a largely de-contextualised state, and the increasing development of automated processing techniques associated with 'big data' exacerbates this situation still further.

As far as the data user is concerned, making sense of the data relies to a considerable extent on their own tacit knowledge and – as Tuomi emphasises – ultimately requires trust in the data originator, since the data-information-knowledge of the end user only emerges as a consequence of their understanding of the knowledge-information-data disarticulation by the original creator ((Tuomi, 1999, p. 112). If, as Gramsch argues, we also need to be able to scrutinise what the data might reveal beyond the originator's intentions (Gramsch, 2011, p. 62), the significance of knowledge about the whole data lifecycle, including the original knowledge-information-data process, the circumstances of collection, and the contextual and tacit information associated with it, becomes greater still. The alternative risks data being wrenched from context, arguments separated from evidence, interpretations transformed into 'facts', explicit knowledge separated from tacit knowledge, and, in the context of digital data processing, push-button solutions substituted for knowledgeable actions (Huggett, 2004*a*,*b*).

The concern, therefore, is that the combination of access to data and distance from understanding the nature of those data in many respects reinforces Postman’s prediction, that:

> “...the tie between information and human purpose has been severed, i.e., information appears indiscriminately, directed at no one in particular, in enormous volume and at high speeds, and disconnected from theory, meaning, or purpose.” (Postman, 1993, p. 70).

This is all the more prescient given the development of ‘big data’ and Chris Anderson’s famous claim that the new ‘Petabyte Age’ :

> “...calls for an entirely different approach, one that requires us to lose the tether of data as something that can be visualized in its totality. It forces us to view data mathematically first and establish a context for it later ... We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.” (Anderson, 2008).

Delivering data in increasingly large amounts but without accompanying awareness about the theories, purposes and processes which lie behind those data means that the data arrive at the would-be user contextless and consequently open to misunderstanding, misconception, misapplication, and misinterpretation.

2.7 Putting the ‘Capta’ Back into Data?

The expansion in access to increasing volumes of open archaeological data in many respects presages the arrival of a new archaeological ‘record’. In 2005, for example, Naylor and Richards predicted that researchers will be increasingly expected to use existing data and will need to justify primary data collection in future (2005, p. 90). More recently, Beck and Neylon suggested that access to dynamic open archaeology data may question the orthodoxy of excavation (2012, p. 494). The risk identified here is that we may get caught up in this brave new technological world of data and lose sight of the underlying issues in the thrill of enhanced access. For instance, Gitelman and Jackson warn that a shared sense of starting with the data

> “...often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself. If we’re not careful, in other words, our zeal for more and more data can become a faith in their neutrality and autonomy, their objectivity.” (Gitelman, 2013, p. 2–3).

Archaeological debates about open data may not fall into this trap and certainly cannot be characterised as excessively utopian in outlook. However, focussing on structures and organisation rather than the data, emphasising their access and delivery, pays relatively less attention to what the access is to, what the delivery is of, and what the consequences of such access and delivery might be.

Lucas characterises archaeological intervention and the consequent creation of a record as a combination of re-materialisation and de-materialisation: re-materialisation in the sense of creating new interpretative objects from the old (sherds, flakes etc.) and the new (photographs, drawings, descriptions etc.), and de-materialisation in the conversion of the physical (excavation trench) into the paper records, photographs, finds and so on (Lucas, 2012, p. 258–259). This is reminiscent of the classic view of information technology as bringing about a shift from atoms (the material world) to bits (the digital world) (Negroponte, 1996, p. 11ff). The introduction of a digital dimension to the archaeological record can be seen as an additional level of de-materialisation, further removing the original objects of record from the interpretative traditional record. The digital record is therefore distanced from the objects lying behind those data just as access to digital data is distanced from the conditions of creation of those data.

Solutions to this distancing are available; however they entail adding new data structures which attempt to capture missing contextual information in the form of elaborated metadata and ontologies. As this effectively applies more technology to a problem created by the technology in the first place, it is not necessarily a robust methodology (Tuomi 1999, p. 110, Bowker 2005, p. 126), even assuming the information can be adequately captured and represented in the first place. For example, the London Charter is frequently cited as an example of the attempt to document computer-based visualisation of cultural heritage by incorporating information about the interpretative decisions made in the course of creating a visualisation. Hence:

> "Documentation of the evaluative, analytical, deductive, interpretative and creative decisions made in the course of computer-based visualisation should be disseminated in such a way that the relationship between research sources, implicit knowledge, explicit reasoning, and visualisation-based outcomes can be understood." (Charter, 2009, p. 8–9).

This is undertaken through the capture of provenance metadata (or paradata) (for example, Baker 2012; Mudge 2012), which contrasts with the more typical metadata currently used by organisations such as the Archaeology Data Service which focuses on issues of authorship, rights, and sources, and carries only limited descriptive information and nothing relating to process or derivation. To a large extent this provenance metadata remains vapourware, with little or no implementation to date. That said, there are technically-derived provenance metadata available which are captured automatically: for instance, ESRI's ArcGIS system captures information about derivable properties of the data and some information about geoprocessing techniques applied to the data without user intervention. Similarly, the EXIF metadata automatically captured by many digital cameras includes information about settings used in the creation of the photograph. If it were feasible, the availability of this kind of contextual metadata would offer the prospect of providing a better understanding of some of the collection processes and circumstances that lie behind the data themselves, as well as potentially improving appreciation of the authority and reliability of the data.

Provenance metadata can, therefore, be seen as a means of addressing the lack of contextual information typically associated with digital data, the absence of which ought to present significant issues when those data are situated, contingent, and incomplete. On the other hand, provenance metadata also increases the data load associated with any given dataset, especially since it cannot necessarily be assumed to exist simply at the collection level. For example, individual records or sets of records within an excavation database will be created by different people and individual contexts will be excavated using different methods; likewise a single individual might be associated with the creation of a GIS dataset but that dataset itself consists of multiple layers which have been created using various data sources and algorithms. It could be argued therefore that provenance metadata would be required at all levels of a given dataset, with significant implications for capturing and representing this information.

The creation of metadata – both supporting resource discovery and providing provenance or contextual information about data – essentially creates more data about data in a structured web of dependencies and relationships. Issues of identification, classification, atomisation, and standardisation are compounded in an environment which adds new data definitions to old. If the original data are perceived in some respects to have been squeezed into pre-defined pigeonholes in order to capture them, this is equally the case with metadata. In this way the technological solution offered by metadata can be seen to reinforce the issues it is intended to resolve. Additionally, it remains to be demonstrated that such contextual metadata would be either useful or used. While the metadata in common use currently is understood to have value as a finding aid, there is little evidence of provenance metadata use as yet or indeed a clear demonstration of how it would work. Provenance metadata may be theoretically valuable, but data users are more accustomed to resorting to textual documentation in order to understand the meaning of a particular data field or its contents, assuming such documentation exists in the first place (c.f. Faniel et al. 2013). Indeed, metadata is in many respects of more significance to computational tools than to the human agents themselves who simply receive the results of the computations as a consequence of a query. We commonly perceive knowledge as passing from one knowledge worker to another with data as the intermediary, whereas increasingly knowledge is handled via a program–data–program or data–program–data cycle with a minimum of human intervention (Bowler, n.d., p. 169–170).

2.8 Transforming Knowledge?

Computer software can be seen as protecting the human user by disguising the underlying complexities of a problem or task through inserting layers of opacity (for example, Huggett 2004*a*, p. 83–84). Similarly, in a kind of utopian determinism, the expectation is that computer systems will resolve current limitations and remove restrictions in terms of access, processing, and storage of data. However, access to these data and the immediateness of their delivery can both overwhelm and isolate the data user from the moment of discovery and capture, with the de-contextualised knowledge-information-data process inserting distance between originator and user. Recognition of this is key to knowledgeable action: for example, Turkle has characterised computer systems as creating a seduction of simulation, in which we become accustomed to manipulating a system whose core assumptions we do not understand, hence leading to the abdication of authority to the simulation (Turkle, 1997, p. 36–42). Equally we may become accustomed to manipulating data whose core assumptions we no longer understand, abdicating authority and responsibility to the system which delivered those data in response to our query. This becomes all the more important when those data are removed from their original context and repurposed – in other words, the data may be purpose-laden, collected not so much with research in mind but resource-management (Huggett, 2004*a*) which brings different priorities and concerns to the fore. Indeed, in addition to being theory-laden and purpose-laden, data may also be process-laden, with aspects of their creation and subsequent modification embedded, often invisibly, within them. The operationalisation of data within a computer environment strips out the context of creation – or at the very least, increases the distance from it (and provenance metadata seems likely to provide a poor proxy at best).

Digital data structures can be seen as constraining subsequent action and analysis, an argument which goes back to the near prehistory of computer archaeology (for example, contributions to Cooper and Richards 1985) but has seen relatively little attention since. These largely unseen and potentially unrealised aspects of digital data are not dissimilar to discussions about the way that traditional context sheets work "to make the objects of archaeology comparable . . . by making the actions of the people that use them comparable" (Yarrow 2008, p. 123; Lucas 2001, p. 9). Although Yarrow suggests that context sheets are actually less restrictive than they might appear (2008, 130-2), it is not clear that the same can be said for data structures. The database is not neutral: data have to be structured in order to be represented, and the choice of representation carries different implications for the data. For example, hierarchical databases, where each item has a single parent, impose a detailed line of authority which required to be followed to retrieve any information (Bowker 2005, p. 130–131, Bowler n.d., p. 169). Relational databases separate the physical organisation of data in the computer and the representation of the data: each entity is identified by a record number, and – in theory – at any point the user can specify a set of relationships to produce a view that reflected those relationships, though in reality the range of relationships is more limited (Bowker, 2005, p. 131). The structure of object-oriented and object-relational databases means that basic operations can be redefined and reconfigured: "any structure can be evanescent providing we know the inputs or outputs of any objects within it" (Bowler, n.d., p. 169), but these are not yet the source of most archaeological data, and even if they were, it remains to be seen how much control the data user is actually allowed. A database is therefore not an 'empty vessel' into which data can be poured – and if it were, the computer would be invisibly organising and making sense of the data which would make the process still less transparent than the traditional structures currently in use. However, there has still been relatively little attention paid within archaeology to the effects of structuring data for a database on the way that we think about that data, on the way we go about recording that data, the way in which we retrieve that data, and the way in which we subsequently analyse that data (Huggett, 2004*a*).

This become more important, not less, as we move into the disruptive realms of what Anderson has described as the "end of theory", in which he claims "'Correlation is enough.' ... We can analyze the data without hypotheses about what it might show." (Anderson, 2008). Such proponents of 'big data' frequently adopt a fetishistic approach to the power of systems to overcome the limits of 'small data'. The sheer quantity of data is argued to make quality less significant, so that the size of the datasets will offset any problems associated with errors and inaccuracies in the data to the extent that "It isn't just that 'more trumps some', but that, in fact, sometimes 'more trumps better'." (Mayer-Schönberger and Cukier, 2013, p. 33). However, just because a dataset is large does not mean it is representative or unbiased, and methodological issues are even more important with large and disparate datasets (Boyd and Crawford, 2012, p. 669). Indeed, boyd and Crawford highlight the mythological aspects of 'big data': specifically that large datasets somehow offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy (2012, p. 663). 'Big data' explicitly adopts the data-information-knowledge model, with the 'bigness' of data seen as requiring it to be collected prior to interpretation (Boellstorff, 2013), and in the process presumes that knowledge can be generated in a theoretical vacuum. This may be true in the sense of data automatically captured through instruments, sensors, click-throughs, and the like, but even the creation of a device (whether hardware or software) has knowledge fixed into it, since what it records is designed into the system (Tuomi, 1999, p. 108–109). In reality, 'big data' always entails 'big theory', whether or not this is recognised (Boellstorff, 2013). Losing sight of these issues risks what Carr (2013) has identified as automation complacency and automation bias, lulling the user into a false sense of security and certainty such that we fail to recognise errors and shortcomings as computers increasingly mediate our understanding (Carr, 2010).

Archaeology may not yet be dealing in 'big data', but the foundations are being laid for doing so. Open data are implicated in this, as is the construction of new data infrastructures (for example, Niccolucci and Richards 2013), the creation of automated processes to align data of different types drawn from different sources (for instance, Jeffrey et al. 2009; May et al. 2010), and processes to automatically extract information from online publications and datasets (Byrne and Klein 2010; Vlachidis et al. 2010, for example). These, and projects like them, are challenging, innovative, and exciting; however, all are based on automatic extraction, processing, and transformation of archaeological data and their results typically become the basis of the tools we use to access archaeological data in the future. One of the clearest examples of this is the faceted classification system developed by the Archaeotools project, which now sits beneath the ArchSearch browser used by the Archaeology Data Service as a primary means of accessing its data and resources (Jeffrey et al., 2009; Richards et al., 2011). The ARIADNE project seeks to integrate existing archaeological data infrastructures across Europe, and while there is no doubting that digital data across Europe are scattered amongst different silos and access is constrained by a lack of common standards and agreed metadata (Niccolucci and Richards, 2013, p. 85), the level of manipulation of data in order to achieve integration across these disparate datasets is likely to be considerable, and the data users potentially removed still further from the data as originated.

2.9 Open Data is for Sharing

None of this should deny the value, importance, and potential of open data in archaeology. When access to the Archaeology Data Service has reduced the time required for data acquisition and data processing for 79% of archaeologists surveyed, has improved the efficiency of archaeological research in the UK (JISC/Research Information Network, 2011, p. 34), and those efficiency impacts are valued at between £13m and £58m per year (Beagrie and Houghton, 2013, p. 40), the benefits seem unarguable. Instead, the concern is to recognise the implications of increasing access to data for users separated by space, and inevitably and increasingly time, from the data originators, and the effects of the ways in which the tools used seek to capture the consequences of interpretation, classification, and identification which remain largely tacit. The benefits for archaeology, in terms of an enhanced ability to access and use data, are predicated upon a clear understanding of those data as well of the level of control and authority implicit in their delivery. Indeed, as the tools formalising the information for delivery are increasingly automated, the status of the data user can become little more than a powerless consumer. Given the way that classification standards, information infrastructures, and the data themselves shape future practice, it is all the more important to reveal the forms, decisions and assumptions which underpin them rather than allow them to remain invisible. These classifications and standards are the means by which data from one time and place are linked to data from another, since they provide for the regularisation of the data, allowing them to be communicated between different contexts (for example, Bowker and Star 1999, p. 290; Huggett 2012).

The ease with which data are communicated within a technological environment is in marked contrast to earlier generations where data were held in notebooks and card indexes and presented in the form of published reports. The benefits of this seem clear:

> "…sharing primary data allows us to better confront some of the biases in the data collection and analysis process, and to do more informed research, rather than simply taking the interpretive publication at face value." (Atici et al., 2013, p. 666).

Making individual datasets available for reuse is largely a matter of providing access and adequate documentation to provide the necessary theoretical and methodological background and explanation (for example, Atici et al. 2013, p. 677–679). In certain respects reusing such data presents similar challenges to reinterpreting traditional non-digital archives. This is not the case where the data have been made interoperable for the purposes of comparison and combination into large datasets. Linking data with other datasets is not a simple process: although semantic tools such as ontologies are used to provide mappings between the different datasets, these are in no sense absolute (Huggett, 2012, p. 543–545). These mappings may be carried out manually or increasingly automatically but "their methods require potentially contestable judgement calls" (Atici et al., 2013, p. 674), and these methods and judgement calls are not made explicit nor are they widely appreciated. As argued elsewhere (Huggett, 2012), little attention has been paid to the means by which data standards have been developed and implemented in order to achieve interoperability and communication – or at least, such as there has been is not in the public domain. In the process, the implications of the methods by which data become interoperable become lost in the face of engagement with these unified datasets which are, by definition, no longer primary and yet may be treated as if they are. Where these mappings are undertaken automatically, the data themselves are no more than tokens shunted around in a manner which reshapes and reformulates them within a technical environment. This is far removed from the eventual human agents who remain largely oblivious to the actions that have been undertaken in order to deliver the data to them.

For example, in the context of the thousands of mostly small-scale archaeological interventions undertaken across the UK and only available as grey literature, Fowler estimated that he was able to take account of less than five percent of the information gained over the past 20 years in attempting to write a work of archaeological synthesis (Fowler, 2001, p. 607). Similarly, Bradley's synthesis of British and Irish prehistory entailed four years of professorial research leave, plus the salary of a research assistant for three years (Bradley 2006, 10) in order to travel the country to seek out grey literature reports accumulated over 20 years. Now, however, there are over 22,000 grey literature reports in the Archaeology Data Service digital library, and more are added each month through the OASIS project in England and Scotland. Access to this resource clearly changes the nature of the task of synthesis, but if natural language techniques are applied to these reports in the search to gain comparability and interoperability of the data and the information codified within them (for example, Richards et al. 2011), what would then be the nature of a synthesis that might be derived as a consequence of such technical intervention?

As Bevan (2012*a*, p. 493) has recently pointed out, the availability of large-scale datasets should shift our goalposts and enlarge our interpretative ambitions, an observation that can be widened to incorporate open data in general. However, as he also points out, access brings with it issues associated with recovery and recording biases – and, as is argued here, potentially a lot more besides. The challenge is to recognise these issues when the emphasis surrounding openness is instead, perhaps inevitably, focussed on facilitating the availability, interoperability, and ease of delivery of the data.

Acknowledgements

I should like to thank Kath Baker for her comments on drafts of this chapter, and also Katie Green and Michael Charno for their assistance with the Archaeology Data Service access figures. As ever, all errors and misconceptions remain my own.

Bibliography

Anderson, C. (2008), ‘The end of theory’, *Wired magazine* **16**(7), 16–07.

Anichini, F. and Gattiglia, G. (2012), #mappaopendata. from web to society. archaeological open data testing, *in* ‘Opening the Past: Archaeological Open Data’, Vol. 3, Metodologie Applicate alla Predittvita del Potenziale Archeologico, pp. 54–56.

Arbesman, S. (2013), ‘Stop hyping big data and start paying attention to ‘long data”, *Wired magazine* **1**(21).

Archaeological Data Service (n.d.), ‘The terms of use and access to ads resources’. URL: http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess

Atici, L., Kansa, S. W., Lev-Tov, J. and Kansa, E. (2013), ‘Other people’s data: A demonstration of the imperative of publishing primary data’, *Journal of Archaeological Method and Theory* **20**(4), 663–681.

Baker, D. (2012), Defining paradata in heritage visualisation, *in* ‘Paradata and Transparency in Virtual Heritage’, Ashgate Publishing, Ltd., pp. 163–175.

Barrett, J. C. (1988), ‘Fields of discourse reconstituting a social archaeology’, *Critique of Anthropology* **7**(3), 5–16.

Beagrie, N. and Houghton, J. (2013), ‘The value and impact of the archaeology data service: A study and methods for enhancing sustainability’. URL: http://archaeologydataservice.ac.uk/research/impact

Beale, N. (2012), ‘How community archaeology can make use of open data to achieve further its objectives’, *World Archaeology* **44**(4), 612–633.

Beck, A. and Neylon, C. (2012), ‘A vision for open archaeology’, *World Archaeology* **44**(4), 479–497.

Belisario, E., Cogo, G., Epifani, S. and Forghieri, C. (2011), *Come si fa Open Data? Istruzioni per luso per Entie Amministrazioni Pubbliche Version 2*, Associazione Italiana per l’Open Government.

Bevan, A. (2012*a*), ‘Spatial methods for analysing large-scale artefact inventories’, *Antiquity: a quarterly review of archaeology* **86**(332), 492–506.

Bevan, A. (2012*b*), Value, authority and the open society. some implications for digital and online archaeology, *in* 'Archaeology and Digital Communication: Towards Strategies of Public Engagement', Archetype, pp. 1–14.

Binford, L. R. (1987), 'Data, relativism and archaeological science', *Man* pp. 391–404.

Boellstorff, T. (2013), 'Making big data, in theory', *First Monday* **18**(10).

Borgman, C. L. (2007), *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*, MIT press.

Bowker, G. C. (2005), *Memory practices in the sciences*, MIT Press Cambridge, MA.

Bowker, G. and Star, S. (1999), 'Sorting things out: classification and its consequences', *Inside technology* .

Bowler, G. (n.d.), Data flakes: an afterword to "raw data" is an oxymoron, *in* "'Raw" data is an oxymoron', Cambridge, MA: MIT Press, pp. 167–171.

Boyd, D. and Crawford, K. (2012), 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society* **15**(5), 662–679.

Bradley, R. (2006), 'Bridging the two cultures–commercial archaeology and the study of prehistoric britain', *The Antiquaries Journal* **86**, 1–13.

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', Oxford: BAR International Series, pp. 48–56.

Canadian Archaeological Association (n.d.), 'Canadian archaeological association principles of ethical conduct'.
URL: http://canadianarchaeology.com/caa/about/ethics/principles-ethical-conduct

Carr, N. (2010), *The shallows: How the internet is changing the way we think, read and remember*, Atlantic Books Ltd.

Carr, N. (2013), 'All can be lost: The risk of putting our knowledge in the hands of machines', *The Atlantic (November 2013* **November 2013**.
URL: http://www.theatlantic.com/magazine/archive/2013/11/the-great-forgetting/309516

Carson, C. A. (1996), 'Laser bones: Copyright issues raised by the use of information technology in archaeology', *Harv. JL & Tech.* **10**, 281.

Charter, L. (2009), 'The london charter for the computer-based visualisation of cultural heritage (version 2.1)'.
URL: http://www.londoncharter.org

Chippindale, C. (2000), 'Capta and data: On the true nature of archaeological information', *American antiquity* **65**(4), 605–612.

Darvill, T. (2007), 'Research frameworks for world heritage sites and the conceptualization of archaeological knowledge', *World Archaeology* **39**(3), 436–457.

Drucker, J. (2011), 'Humanities approaches to graphical display', *Digital Humanities Quarterly* **5**(1).

Duguid, P. (2005), "the art of knowing': Social and tacit dimensions of knowledge and the limits of the community of practice', *The information society* **21**(2), 109–118.

European Association of Archaeologists (1997), 'European association of archaeologists code of practice'.
URL: http://e-a-a.org/codes.htm

Faniel, I., Kansa, E., Whitcher Kansa, S., Barrera-Gomez, J. and Yakel, E. (2013), The challenges of digging data: a study of context in archaeological data reuse, *in* 'Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries', ACM, pp. 295–304.

Fowler, P. (2001), 'Time for a last quick one?', *Antiquity* **75**(289), 606–608.

Gitelman, L. (2013), *Raw data is an oxymoron*, MIT Press.

Gramsch, A. (2011), Theory in central european archaeology: dead or alive?, *in* 'The Death of Archaeological Theory?', Oxbow Books, pp. 48–71.

Heritage Gateway (2007), 'Terms and conditions'.
URL: http://www.heritagegateway.org.uk/gateway/termsandcondition
Hodder, I. (1999), *The archaeological process: An introduction*, Blackwell Oxford.
Huggett, J. (2000), *Computers and archaeological culture change*, Oxford University Committee for Archaeology Monograph 51, pp. 5–22.
Huggett, J. (2004*a*), 'Archaeology and the new technological fetishism', *Archeologia e calcolatori* (15), 81–92.
Huggett, J. (2004*b*), 'The past in bits: towards an archaeology of information technology', *Internet Archaeology* **15**.
Huggett, J. (2012), 'Lost in information? ways of knowing and modes of representation in e-archaeology', *World Archaeology* **44**(4), 538–552.
Huggett, J. (Forthcoming), 'Promise and paradox: accessing open data in archaeology'.
Institute for Archaeologists (2013), 'Ifa 2013 bylaws of the institute for archaeologists: Code of conduct, reading'.
URL: 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: Mathematical, Physical and Engineering Sciences* **367**(1897), 2507–2519.
JISC/Research Information Network (2011), 'Jisc 2011 data centres: their use, value and impact'.
URL: http://www.jisc.ac.uk/publications/generalpublications/2011/09/datacentres.aspx
Kansa, E. (2012), 'Openness and archaeology's information ecosystem', *World Archaeology* **44**(4), 498–520.
Kansa, E., Kansa, S. W., Burton, M. M. and Stankowski, C. (2010), 'Googling the grey: Open data, web services, and semantics', *Archaeologies* **6**(2), 301–326.
Knox, K. T. (2007), 'The various and conflicting notions of information', *Issues in Informing Science and Information Technology* **4**(1), 675–89.
Lake, M. (2012), 'Open archaeology', *World Archaeology* **44**(4), 471–478.
Llobera, M. (2011), 'Archaeological visualization: Towards an archaeological information science (aisc)', *Journal of Archaeological Method and Theory* **18**(3), 193–223.
Lucas, G. (2001), *Critical approaches to fieldwork: contemporary and historical archaeological practice*, Routledge.
Lucas, G. (2012), *Understanding the archaeological record*, 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', Oxford: BAR International Series, pp. 227–233.
Mayer-Schönberger, V. and Cukier, K. (2013), *Big data: A revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt.
McKeague, P., Corns, A. and Shaw, R. (2012), 'Developing a spatial data infrastructure for archaeological and built heritage', *International Journal of Spatial Data Infrastructure Research* **7**, 38–65.
Mudge, M. (2012), *Transparency for empirical data*, Farnham, Surrey: Ashgate, pp. 177–188.
Naylor, J. and Richards, J. (2005), 'Third-party data for first class research', *Archeologia e Calcolatori* (XVI), 83–91.
Negroponte, N. (1996), *Being digital*, Random House LLC.
Niccolucci, F. and Richards, J. (2013), 'Ariadne: Advanced research infrastructures for archaeological dataset networking in europe', *International Journal of Humanities and Arts Computing* **7**(1-2), 70–88.
of America, A. I. (2008), 'Archaeological institute of america code of professional standards'.
URL: http://www.archaeological.org/pdfs/AIA_Code_of_Professional_StandardsA5S.pdf

Open Definition (2014), ‘Open definition version 1.1’.
URL: http://opendefinition.org/okd/

PastMap (2013), ‘Terms and conditions’.
URL: http://pastmap.org.uk/

Patrik, L. E. (1985), ‘Is there an archaeological record?’, *Advances in archaeological method and theory* pp. 27–62.

Postman, N. (1993), ‘Technopoly: the surrender of culture to technology’.

Räsänen, M. and Nyce, J. M. (2013), ‘The raw is cooked data in intelligence practice’, *Science, Technology & Human Values* **38**(5), 655–677.

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*, Los Angeles: UCLA Cotsen Institute of Archaeology Press, pp. 31–56.
URL: http://escholarship.org/uc/item/1r6137tb

Richards, J., Naylor, J. and Holas-Clark, C. (2008), ‘The viking and anglo-saxon landscape and economy (vasle) project’.
URL: http://archaeologydataservice.ac.uk/archives/view/vasle_ahrc_2008/index.cfm

Richards, J., Naylor, J. and Holas-Clark, C. (2009), ‘Anglo-saxon landscape and economy: using portable antiquities to study anglo-saxon and viking age england’, *Internet Archaeology* .

Royal Society (2012), ‘Science as an open enterprise: open data for open science’.
URL: http://royalsociety.org/policy/projects/science-public-enterprise/report/

Society for American Archaeology (1996), ‘Society for american archaeology principles of archaeological ethics’.
URL: http://www.saa.org/AbouttheSociety/PrinciplesofArchaeologicalEthics/tabid/203/Default.aspx

Trigger, B. G. (1998), ‘Archaeology and epistemology: dialoguing across the darwinian chasm’, *American Journal of Archaeology* pp. 1–34.

Tuomi, I. (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.

Turkle, S. (1997), ‘Life on the screen: Identity in the age of the internet’, *Literature And History* **6**, 117–118.

Vlachidis, A., Binding, C., Tudhope, D. and May, K. (2010), Excavating grey literature: A case study on the rich indexing of archaeological documents via natural language-processing techniques and knowledge-based resources, *in* ‘Aslib Proceedings’, Vol. 62, Emerald Group Publishing Limited, pp. 466–475.

Weinberger, D. (2010), ‘The problem with the data–information–knowledge–wisdom hierarchy’, *Harvard Business Review Blog Network* **2**.

Weinberger, D. (2011), ‘Too big to know’.

Yarrow, T. (2008), In context: meaning, materiality and agency in the process of archaeological recording, *in* ‘Material Agency’, Springer, pp. 121–137.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R. and Sheikh_Alibaks, R. (2012), ‘Socio-technical impediments of open data’, *Electronic Journal of e-Government* **10**(2), 156–172.

Ray Moore and Julian Richards

3 Here Today, Gone Tomorrow: Open Access, Open Data and Digital Preservation

3.1 Introduction

The increasing popularity and pervasiveness of open access and open data approaches within contemporary society continues to have a significant impact on the archaeological profession. A primary concern within these discussions has been the movement towards providing unrestricted access to the peer-reviewed textual content produced in the aftermath of archaeological research, particularly content published in scholarly journals, although other forms of written output (monographs, thesis, books, etc.) have become increasingly drawn into the discussion. A more recent refocusing of this debate generally, and increasingly within archaeological discourse, has seen a return to the issue of accessibility of the primary data produced during research in the hope that openness will promote wider discussion and revitalise understanding. Certainly, the increased and unrestricted access promised by a more open approach to archaeological data is likely to change the nature of archaeological discourse and to facilitate new interpretations of the past. At the same time, the effects of access to the grey data produced during fieldwork within commercial archaeology, although less well understood, could have huge benefits both intellectually and economically. Discussions have suggested that in order to deal with the increasing quantities of open data generated during fieldwork and research the profession will need to develop infrastructures to deal with both the dissemination and preservation of this data (Kintigh, 2006; Snow et al., 2006). It is our intention here to suggest that these two outcomes need not be mutually exclusive; that digital archives and repositories can take a leading role in both the maintenance of access and in the curation of datasets. The experiences of the UK based Archaeology Data Service (ADS) are brought into focus as an instance where both these outcomes have been successfully achieved. Discussions of the work of the ADS have often focused on its role in the preservation of data, but it also taken a leading role as a data broker, aggregator and distributor. It is hoped that a better understanding of the sharing of archaeological data over the longue durèe will help us understand contemporary concerns. In focusing on the work of the ADS we contend that when promoting open data a hybrid approach to dissemination and preservation has the greatest potential to succeed.

Ray Moore: Archaeological Data Service, York, UK
Julian Richards: Archaeological Data Service, York, UK

3.2 Sharing Data: The 'Traditional' Treatment of Archaeological Data

The destructive nature of many forms of archaeological investigation has compelled researchers to make considerable efforts, both legally and ethically, to preserve and disseminate their results. The principal outcome of this research continues to be text-based publication; whether the monograph, book, journal or grey literature report. Yet at the same time archaeological investigations generate significant quantities of primary data which are not easily reproduced or disseminated using traditional media. Data sharing has often been difficult for technical reasons and while paper has proved convenient for the distribution of the textual outputs of archaeological research, it has never been an efficient medium for the dissemination of complex datasets (Jeffrey, 2012). As early as 1975, a review by the Ancient Monuments Board for England concluded that "publication in printed form of all the details of a large modern excavation is no longer practicable" (Frere, 1975, p. 2). This unsustainability led to a 'publication crisis' within British archaeology as the profession struggled to deal with the quantity and scale of the outputs produced by archaeological fieldwork (Richards, 2002). Solutions typically involved limiting the print-based publication, with increased emphasis placed on the archiving of associated data (Frere, 1975; Cunliffe, 1983; Carver et al., 1992). In providing an answer to the crisis in publication, solutions often did so at the expense of accessibility of the data produced during archaeological fieldwork. A working group, created by the Council for British Archaeology and the Department the Environment, attempted to address the accessibility issue by promoting microfiche as an alternative to print (Cunliffe, 1983), but this solution never gained popular acceptance (Richards, 2002; Jones et al., 2001). Unfortunately, "technology lagged behind and lacked the means of providing access to an archive with links between it and the summary publication" (Richards, 2002, p. 356). The increased pervasiveness of digital technology, and the growing popularity of the web, marked a significant sea-change in the landscape opening new avenues for the sharing, collaboration, and analysis of archaeological data. The Publication of Archaeological Projects (PUNS) report, commissioned by the Council for British Archaeology, took a user-driven perspective in examining the use of publications and the data within the profession (Jones et al., 2001). It concluded that:

> "While print remains favoured, it is clearly no longer the only or even main medium for dissemination. The point has been reached, indeed, at which 'publication' and 'dissemination' must be seen as different things. As a means of giving access to archives or disseminating material that would otherwise be relegated to grey literature, the advantages of the Internet are immense, and increasingly accepted" (Jones et al., 2001, p. 69–70).

The PUNS report promoted a 'layered' approach to the publication and dissemination of archaeological research; an approach that takes advantage of the benefits of textual and digital technologies and where the traditional narrative can give contextual information and meaning to the archaeological data[1]. Since PUNS there have been no further reviews of publication policy, and its recommendations are still valid, although implementation has been slow. There have been a number of experiments in alternative forms of online and mulit-layered publication, including Scottish Archaeological Internet Reports (Society of Antiquaries of Scotland, 2013) and the LEAP project (Richards et al., 2011). Similarly, Takeda et al. (2013) promote a 'rich interactive framework' which incorporates supplementary information to support journal based publication.

This volume itself demonstrates a growing awareness of the issues associated with openness within archaeological discourse, whilst the recent dedication of an entire volume of World Archaeology, one of the discipline's premier and mainstream journals, to the subject of 'open archaeology' attests to its pervasiveness and entrance into the mainstream (Lake, 2012). At the same time the broad subject matter of these works attests to the innate complexity of the open movement within archaeology; with concepts like open access, open source, open software, open standards, open archaeology etc. already firmly entrenched in the vernacular and increasingly implicated in archaeological practice[2].

Within archaeology the debate on openness has typically focused on 'open access' publication, and have been particularly focused on its impacts on the 'traditional' outputs of research and grey literature (Lake, 2012). Yet, as the benefits of openness within archaeological publication have been recognised, its expansion to the structured data produced during archaeological research and fieldwork seems logical. The development of so-called 'open data' has, and will continue to have a significant impact on the development of the profession. However, what do we mean by open data? Open data can be broadly defined as, "data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike" (Open Knowledge Foundation, n.d.). More specifically it can be defined according to three concepts:

1. Technical openness: data should be made available in widely used, non-proprietary formats that can be used across multiple computing and software platforms.
2. Legal openness: data must be free of encumbering intellectual property restrictions.
3. Access: datasets must be made available freely and, unless there are overriding

1 The implications of this 'linked' approach to the textual and digital outputs of archaeological research will be discussed below.

2 It would be redundant to rehearse the discussions articulated by others in defining the nuances of open archaeology consequently we would refer those seeking a wider understanding of the concept to the other papers in this work and the World Archaeology volume on the subject (Lake, 2012).

privacy or security needs, data releases need to be both comprehensive and sufficiently documented to enable reuse (Kansa, 2012, p. 506).

For many Open Data is often equated with Linked Open Data and attempts to develop a linked data cloud of open data sets, in which key concepts are each linked to other online sources, in fulfilment of Berners Lee’s original vision of a semantic web of machine-readable data (Binding, 2010; Wright, 2011; Isaksen, 2011; Tudhope, Binding, Jeffrey, May and Vlachidis, 2011; Tudhope, May, Binding and Vlachidis, 2011). However, in this paper we are concerned with open data more broadly. In fact the concept of open access to scientific data is not a new one, and long pre-dates the Internet. Indeed, it was first institutionally established in preparation for the International Geophysical Year of 1957-8. The International Council of Scientific Unions established several World Data Centers to minimize the risk of data loss and to maximize data accessibility, further recommending in 1955 that data be made available in machine-readable form. In 2004, the OECD (Organisation for Economic Co-operation and Development) Science Ministers ruled that all publicly funded archive data should be made publicly available.

The European Commission has outlined a ‘digital agenda for Europe’ which seeks to promote open data for publicly funded research (2011). Similarly the UK Government has advocated ‘a culture of openness’ which contends that “access to data is fundamental if researchers are to reproduce and thereby verify results that are reported in the literature” (House of Commons, Department for Business Innovation and Skills, 2012). Endorsing the findings of the Finch report (2012), the UK Government has promoted greater accessibility for research data and grey literature through subject and institutional repositories (House of Commons, Department for Business Innovation and Skills 2012, p. 4; Finch 2012). The government ‘Open Data White Paper’ “sets out clearly how the UK will continue to unlock and seize the benefits of data sharing” by enhancing access to data and safeguarding it from potential misuse (UK Government Cabinet Office, 2013). In light of these developments research councils, funding agencies and higher education institutions have outlined commitments to open data (Research Councils UK, 2013). The implications of these statements are currently being worked out through the policies and procedures of individual councils, with the Engineering and Physical Sciences Research Council (EPSRC) taking one of the strongest positions to date, namely that research organisations are expected to publish online appropriately structured metadata describing the research data they hold, normally within 12 months of the data being generated, and for the data themselves to be made available without restriction for a minimum of 10 years. Although no additional funding has been made available to support data archives or institutional repositories, research organisations in receipt of EPSRC funding are expected to have a roadmap in place by May 2012 for compliance with the EPSRC policy framework on research data by May 2015.

The international Open Data Movement has recently received two further boosts. On 13 June 2013 the European Parliament ratified new rules on Open Data, and specifically included cultural heritage data held by public archives museums and galleries. Less than a week later, on 18 June 2013, the Open Data Charter was unveiled at the G8 Summit at Loch Erne, in Northern Ireland. It recognises "a new era in which people can use open data to generate insights, ideas, and services to create a better world for all" (UK Government Cabinet Office, 2013). The G8 Charter establishes 5 principles: (1) that data should be open by default; (2) that steps should be taken to increase the quality, quantity and reuse of data that is released; (3) that it should be usable by all; (4) that releasing data should improve governance; and (5) that releasing data should increase innovation.

Within archaeology we have long recognised the benefits and potential impact that the sharing and reuse of data can bring. Yet, as Kansa observes

> "... these barriers show growing cracks as current norms of closed access and data withholding research in archaeology become increasingly untenable and new modes of understanding and communicating the past take root" (Kansa, 2012, p. 499).

Nonetheless, the benefits of increased accessibility, and the messages of open access and open data, are especially poignant for archaeology, given the primary and unrepeatable status of most data sets. Indeed, within

> "... a discipline that relies upon destructive research methods, lack of information sharing not only inhibits scholarship, but also represents a tragic loss of irreplaceable cultural and historical knowledge. The discipline urgently requires a more professional approach if researchers are to make credible and replicable knowledge claims and act as better stewards of cultural heritage" (Kansa and Kansa, 2013, p. 88).

As a profession archaeologists have sometimes been reluctant to share their primary research data with others. For some this is attributed to the technical barriers associated with providing access to data (Condron et al., 1999; Kansa and Kansa, 2013) or more practical restrictions on the dissemination of data imposed by publishers or data providers. Yet by far the greatest hurdle to overcome is conceptual; while Pratt has observed that "archaeologists are eager to find ways to publish these data sets" (Pratt, 2013, p. 101), some remain unconvinced about the benefits that open data promotes. Others may be reluctant to expose perceived deficiencies in primary data recording to the critical scrutiny of their peers, or may believe that there is a risk that their data will be published by others before they have the opportunity to do it themselves. An awareness of the academic, symbolic and economic 'capital' of archaeological data streams has hindered the sharing of data (Porter, 2013); whilst potential misuse and misappropriation of data have always been concerns. For Kansa "the discipline should not continue to tolerate the personal, self-aggrandizing appropriation of cultural heritage that comes with data hoarding”, indeed data withholding “represents a clear threat to preserving the archaeological record” (Kansa, 2012, p. 507).

Such cultural reluctance is not new to archaeology; yet these issues have not precluded the sharing of data in the past, but have simply constrained the scale of dissemination. Within the current climate with disparate groups and communities conducting related research; where the scale of research and the data produced has increased exponentially, such an approach is unsustainable. Open data offers researchers a mechanism to improve disciplinary interaction and, as a consequence, enhance research. The unrestricted accessibility presented by open data also presents archaeology with opportunities to use, and reuse data. Offering the potential for the ‘remixing’ of archaeological data and its application in new and innovate ways that will enhance understanding the past. The use of text mining and natural language processing within the Archaeotools project, for example illustrates how the application of new analytical techniques to archaeological data can lead to enhanced understandings (Richards et al., 2011). Such re-use may also provide unexpected dividends in the form of re-use of data of research questions that were not envisaged at the outset. In the case of Archaeotools for instance, it became apparent that the application of techniques of information extraction to historic journal runs (in this case the Proceedings of the Society of Antiquaries of Scotland) not only provided a means of automated indexing, but also allowed us to trace the development of controlled vocabularies in archaeology (Bateman and Jeffrey, 2011).

Archaeologists have always been mindful of the need for transparency and repeatability within our negotiations with the past; driven, in part, by a concern over the historical misappropriation of previous generations Trigger (1989); Jones et al. (2001). Increased accessibility has the potential to allow others to test the validity of our interpretations; allowing them to examine and reanalyse the original data. As Lake contends these “[o]pen approaches to knowledge have the potential to bolster scientific rigour by increasing transparency” (Kansa, 2012, p. 473). At the same time this transparency can serve to illustrate the professionalism of data creators by highlighting good research practice (Kansa, 2012).

As ever increasing quantities of open data are released onto the web, concerns have been expressed over the quality of the data. While there are data creators producing well-formed data accompanied by the appropriate metadata, there are large quantities accompanied with only minimal or no documentation. The development of data content, documentation and ontology standards within archaeology has facilitated the creation of ‘good’, well documented data; data with the highest potential for use and reuse (Richards, 2009; Mitcham et al., 2010). Yet, much is still down to individuals, communities and those institutions hosting data to ensure that these, or similar, standards are adhered to and enforced. Many of these standards are already deeply engrained in archaeological practice. The Guides to Good Practice, created by the ADS and Digital Antiquity have become pivotal to the profession, providing assistance with the ever increasing diversity of data types and formats (Mitcham et al., 2010)[3]. At the same time quality assurance of digital resources has become necessary, encouraging data creators to document data appropriately. The Journal of Open Archaeology Data (JOAD)[4] and Internet Archaeology (reference forthcoming) each promote good practice through the production of peer-reviewed data papers which "ensure that the associated data are professionally archived, preserved, and openly available. Equally importantly, the data and the papers are citable, and reuse is tracked" (Journal of Open Archaeology Data, n.d.). Whilst there is certainly a place for such formal appraisals of data, we should not underrate the abilities of data users themselves to make assessments of data quality. The simple fact of the matter is that good data will continue to be used, whilst poor data will not.

3.3 Accessing Data: The Case of the Archaeology Data Service

Founded in 1996 the Archaeology Data Service (ADS) was established as one of five disciplinary data centres, under the auspices of Arts and Humanities Data Service (AHDS), to provide specialist advice and expertise during the lifecycle of digital data from its creation, through to its preservation, and onward to its potential reuse. At the same time an awareness of the need for subject specific expertise to assess the research value and successfully validate digital assets and documentation was recognised. From the outset the

> "…specific brief of the ADS [was] to collect, describe, catalogue, preserve, and provide user support for the re-use of digital data generated in the course of archaeological research by British archaeologists, wherever they are working" (Richards, 1997, p. 1058).

In doing so it provides support for research, learning and teaching within the archaeological sector, through the provision of freely available, high quality and dependable digital resources. This is achieved through the preservation and dissemination of digital data over the long term; an action that has allowed data creators and users to plan not only for preservation, but also use and reuse of digital assets. Throughout its existence the ADS has maintained a broad collections policy that covers all the archaeology of the British Isles, and all areas of the globe where British archaeologists undertake research. It maintains data resources from chronologically, thematically and geographically disparate areas, so much so that now maintains over 1100 digital collections created by individuals, projects and organisations working within both academic and commercial sectors. Of course this is not to say that all digital data should be preserved. As one of the authors has previously suggested the costs of archiving

3 The Guides to Good Practice, developed by the Archaeology Data Service and Digital Antiquity, http://guides.archaeologydataservice.ac.uk/, accessed 30 August 2013.

4 JOAD - http://openarchaeologydata.metajnl.com/, accessed 30 August 2013.

mean that “it is important to establish the reuse potential of any data set before expending resources on its preservation” (Richards, 2002, p. 347); a belief that continues to underpin the ADS philosophy (Richards, 1997). As a result the focus of the ADS has always been in preserving quality, well documented datasets that show the greatest potential for reuse.

Throughout its history a principal focus of the ADS has been the preservation of data created during archaeological fieldwork and research. As noted above the majority of UK funding bodies now recommend, and increasingly require, that digital data produced during research should be preserved. At the same time the landscape has changed significantly in recent years as museums and ‘traditional’ physical archives, many of which lack the necessary digital expertise, progressively compel those working in commercial archaeology to deposit the digital outputs of fieldwork into a mandated digital archive. In both instances the ADS has taken a role in the preservation of archaeological data. Many will be familiar with its work in traditional academic research environments, but it also works with partners within the commercial sector. Acting as a data broker, it has assisted in the creation of OASIS, an online form used throughout the profession to record the outcomes of archaeological fieldwork (Hardman, 2009)[5]. The OASIS system has been enhanced by a facility that allows users to upload the reports produced as a consequence of these activities. These outputs are preserved and, perhaps more significantly, disseminated through the ADS’ Grey Literature Library[6]. This library now provides direct access to some 20,000 unpublished fieldwork reports, produced by over 140 contracting units working within Britain; and has become an important research tool in its own right (Fulford and Holbrook, 2011).

While much discussion has focused on the work of the ADS in preserving the outputs of archaeological research its role as a data disseminator has received much less attention. It has offered free access to its collections and the data therein, a policy developed long before the concepts of open access had been rigorously defined. The terms of use developed by the ADS provide access to data through a “non-exclusive, non-transferable licence” with the depositor (Archaeological Data Service, n.d.); which means:

> “Anyone is permitted to use data held by the ADS so long as it is for research or educational purposes, and these are defined quite broadly as purposes intended to develop knowledge and where the research output is itself destined for the public domain. Therefore reuse of data held by ADS by commercial contractors is not prevented so long as publication of their work is not limited by issues of client confidentiality” (Richards, 2002, p. 349).

5 OASIS standing for Online AccesS to the Index of archaeological investigationS - http://oasis.ac.uk/, accessed 21 September 2014.

6 The Grey Literature Library - http://archaeologydataservice.ac.uk/archives/view/greylit/, accessed 21 September 2014.

In seeking to protect the rights of the depositor the ADS conditions of use are therefore broadly equivalent to a CC-BY-NC licence. Indeed, the terms of use state that the ADS "seeks to protect the intellectual property rights and copyright of the originators of data where that can reasonably be achieved" through a common access agreement (Archaeological Data Service, n.d.). Encouraging them to be "fair and reasonable in their use of the data supplied" (Archaeological Data Service, n.d.). At the same time those depositing data are expected to sign a deposit licence that declares their copyright and ownership to all the data within the collection (Richards, 2002); an action that serves to project the rights of other data creators. This policy has obvious drawbacks, indeed the ADS has even refused deposits where the ownership is obscure or where it is derived from data streams of other individuals and organisations (Mitcham, n.d.). Therefore the approach taken by the ADS places a much greater emphasis on securing the intellectual property rights of data creators, whereas the onus within a full open data environment is firmly placed on the user. For many these 'restrictions' may seem prohibitive, yet experience suggests this is not the case (Heath, 2010). This is not to say that the ADS insists on a single rights management framework, and when requested data can be disseminated under another form of licence. The Antikythera Survey Project, for example, is disseminated under an open data compliant Creative Commons licence (CC-BY 3.0) (Bevan and Conolly, 2012).

While increased accessibility and reuse has done much to raise awareness of the intrinsic value of research data, official recognition of its importance has served to encourage data creators to share these outcomes. The UK Government, for example, has stated that:

> "The work of researchers who expend time and effort adding value to their data, to make it usable by others, should be acknowledged as a valuable part of their role. Research funders and publishers should explore how researchers could be encouraged to add this value" (UK Government, 2011).

Despite this change in mind-set the data outputs of archaeological research can still be treated with some diffidence; an incongruent outcome of less significance than the final interpretation or synthesis. Costa, et al. propose that in order to overcome this mind-set archaeological data needs to treated as "a more relevant part of the archaeological publication, research, management, curation and policy process, and not merely an afterthought" (Costa et al., Forthcoming; Atici et al., 2013; Pratt, 2013). The solution advocated by many is treat the dissemination of data as a form of publication; one which should employ established practice found within text-based publishing, included citation and editorial control (Kansa et al., 2010; Kansa and Kansa, 2011). This it is believed will instil a sense of familiarity to process of disseminating and citing digital resources. This movement towards, what is termed 'data sharing as publication', is intended make the dissemination of data "a more regular and integral part of professional practice" (Bevan and Conolly, 2012, p. 161).

To a large extent such ‘publication of data’ is already part of the ADS workflow. From the outset has endeavoured to promote links between the traditional outputs of research and supporting datasets. The ADS and the e-journal Internet Archaeology have co-published peer-reviewed articles and associated data (*Internet Archaeology*, n.d.). The award-winning Linking Electronic Archives and Publications (LEAP) project set out explicitly to provide a series of exemplars of linked publications and archives, including the projects of Merv, Silchester, Troodos, and Whittlewood (Richards et al., 2011). Of course this relationship is not exclusive and the ADS has always disseminated data on behalf of other digital and paper based reports and articles. Working with the Council for British Archaeology the ADS distributes digital versions of its research reports and occasional papers, including additional supporting data and other material(for British Archaeology, 2007). The ADSnow has an agreement with Elsevier to provide access to supplementary data supporting articles in the Journal of Archaeological Science. Our ejournal Internet Archaeology has also published articles linked to data sets held in other data archives, including tDAR in the United States (Holmberg, 2010). This linking of content is not restricted to the research environment; within commercial archaeology the ADS disseminate data derived from large-scale infrastructural developments, such as Channel Tunnel Rail Link(Framework Archaeology, 2011*a*) and Heathrow Terminal 5 (Framework Archaeology, 2011*b*);bridging the gap between the traditional fieldwork monograph and the supporting digital data. Working with Southampton Arts and Heritage the ADS has also published some 12 discrete excavation archives from the Southampton area, each of which is linked to the grey literature report lodged in the Grey Literature Library[7].

A more open archaeology and the dissemination of increasing quantities of data will necessitate the development of new techniques and tools to deal with the proper referencing and citation of digital resources; indeed without this there is a very real possibility of becoming ‘lost in information’ (Huggett, 2012). At the same time a common concern amongst data creators is the lack of accreditation for data. Both concerns could be addressed through improved citation. Traditionally digital resources have utilised the URL to reference digital resources, however, the durability of this method of citation has begun to be questioned (Jeffrey, 2012). A number of schemes have attempted to address this issue; one of these is the DOI system which “allows collections of data or individual data files to be allocated a URL that will not change irrespective of changes to the physical location of the files in question” (Jeffrey, 2012, p. 564). The ‘minting’ and subsequent management of DOI’s is handled by a conglomerate of organisations, working as part of the International DOI Foundation, who guarantee the sustainability of the citation system (*Datacite*, n.d.). As an adopter of the DOI system the ADS creates persistent identifiers that consistently and accurately reference dig-

7 Part of the Southampton’s Designated Archaeology Collections Programme - http://archaeologydataservice.ac.uk/archives/view/southampton/, accessed 23 August 2013.

ital objects and collections. This serves to address one the principal concerns of the PUNS report (Jones et al., 2001) by formalising associations between digital resources and printed outputs. An important outcome of the DOI system is that it also allows citations to be tracked, meaning that data creators, users and repositories can track the use and impact of specific data sets or publications (Hole, 2012).

3.4 Conclusion

The movement towards open data, and the associated dissemination of ever increasing quantities of data, has the potential to transform archaeology and our understandings of the past. While some contend that "we face the great challenge of bridging two realities – moving from currently entrenched practices to a future of more open and diverse scholarly outputs" (Kansa and Kansa, 2013, p. 103) others have recognised that "many of the wider social, cultural, political and economic issues raised by the various planks of the 'Open' movement are not in themselves new" (Lake, 2012, p. 476). Within the UK the work of the ADS in facilitating access, promoting good practice, endorsing proper citation and encouraging the reuse of research data is making an important contribution towards the Open Data movement. The growing propensity for open data within archaeological discourse will continue to necessitate change and the development of new archiving techniques and workflows, but the experience of the ADS confirms the important role of discipline-based data archives in supporting open access and suggests that the current infrastructure has the innate flexibility to deal with the new demands of a more open archaeology. The pressure from funders requiring research institutions to ensure open access and preservation of data generated by their employees has led to a rapid development of institutional repositories, but whilst the majority provide excellent self-archiving and pre-print repositories in support of open access publication, they also recognise that the long term curation of primary and specialist research data requires them to work with discipline-based data archives. Indeed, the exchange of metadata allows institutions to maintain an institutional view of datasets produced by their employees, and to satisfy audit purposes, whilst sub-contracting long term curation of specialist data sets to other facilities (Rumsey and Jefferies, 2013). A 2011 report commissioned by the Research Information Network and the JISC from Talis concluded that national data centres have an important role to play in terms of providing a focus for data access, overcoming the potential fragmentation of multiple institutional repositories with a focus on short term curation rather than access, and that there was considerable additional value-added from a discipline-based view (JISC, 2011). In 2013, another JISC-funded survey was undertaken by Neil Beagrie and John Houghton into the Impact and Value of the ADS (Beagrie and Houghton, 2013). Adopting techniques for measuring the economic value of non-costed services, Beagrie and Houghton concluded that over a period of 30 years, every £1 invested in ADS, would yield an economic return to the UK economy of up to £8.30. Whilst the value of Open Data should not be seen in economic terms alone, and we would argue that there are strong societal benefits from providing open public access to our shared cultural heritage, the economic argument in favour of Open Data is a useful one. Hopefully it will ensure the current political pressures in favour of Open Data will continue, and that the archaeological profession will continue to benefit from them.

Bibliography

Archaeological Data Service (n.d.).
URL: http://archaeologydataservice.ac.uk

Atici, L., Kansa, S. W., Lev-Tov, J. and Kansa, E. C. (2013), 'Other people's data: A demonstration of the imperative of publishing primary data', *Journal of Archaeological Method and Theory* **20**(4), 663–681.

Bateman, J. and Jeffrey, S. (2011), 'What matters about the monument: reconstructing historical classification', *Internet Archaeology* **29**.

Beagrie, N. and Houghton, J. (2013), 'The value and impact of the archaeology data service. a study and methods for enhancing sustainability', *Archaeology Data Service* .

Bevan, A. and Conolly, J. (2012), 'The antikythera survey project [data-set]'.
URL: http://archaeologydataservice.ac.uk/archives/view/antikythera_ahrc_2012/

Binding, C. (2010), Implementing archaeological time periods using CIDOC CRM and SKOS, *in* 'The Semantic Web: Research and Applications', Springer, pp. 273–287.

Carver, M., Chapman, H., Cunliffe, B., Hassall, T., Hebditch, M., Lawson, A., Longworth, I., Morris, R., Phillipson, D., Schofield, J. and Others (1992), 'Archaeological publication, archives and collections. towards a national policy', *London: Society of Antiquaries. Printed as a supplement to British Archaeological News* **7**(2).

Condron, F., Richards, J., Robinson, D. and Wise, A. (1999), *Strategies for digital data: findings and recommendations from digital data in archaeology: a survey of user needs*, Archaeology Data Service, University of York.
URL: http://ads.ahds.ac.uk/project/strategies/

Costa, S., Beck, A., Bevan, A. and Ogden, J. (Forthcoming), 'Defining and advocating open data in archaeology'.

Cunliffe, B. (1983), *The publication of archaeological excavations: The report of a joint working party of the Council for British Archaeology and the Department of the Environment*, Joint Working Party.

Datacite (n.d.).
URL: http://www.datacite.org/

Finch, J. (2012), 'Accessibility, sustainability, excellence: how to expand access to research publications', *Report of the Working Group on Expanding Access to Published Research Findings* .
URL: http://www.researchinfonet.org/publish/finch/

for British Archaeology, C. (2007), 'Cba research reports'.

Framework Archaeology (2011*a*), 'Channel tunnel rail link section 1 [data-set]'.
URL: http://archaeologydataservice.ac.uk/archives/view/t5_framework_2011/

Framework Archaeology (2011*b*), 'Framework archaeology heathrow terminal [data-set]'.
URL: http://archaeologydataservice.ac.uk/archives/view/ctrl/

Frere, S. (1975), *Principles of Publication in Rescue Archaeology: Report by a Working Party of the*

Ancient Monuments Board for England, Committee for Rescue Archaeology, Department of the Environment.

Fulford, M. and Holbrook, N. (2011), 'Assessing the contribution of commercial archaeology to the study of the roman period in england, 1990–2004', *The Antiquaries Journal* **91**, 323–345.

Hardman, C. (2009), 'The online access to the index of archaeological investigations (OASIS) project: facilitating access to archaeological grey literature in england and scotland', *The Grey Journal* **5**(2), 76–82.

Heath, S. (2010), *Diversity and reuse of digital resources for ancient Mediterranean material culture*, Farnham, UK: Ashgate, pp. 35–52.

Hole, B. (2012), *A Call for Open Scholarship in Archaeology*, Archetype.

Holmberg, K. (2010), 'Placing immateriality: Situating the material of highland chiriquí', *Internet Archaeology* **28**.

House of Commons, Department for Business Innovation and Skills (2012), *Letter to Dame Janet Finch on the Government Response to the Finch Group Report: "Accessibility, sustainability, excellence: how to expand access to research publications"*, London.

Huggett, J. (2012), 'Lost in information? ways of knowing and modes of representation in e-archaeology', *World Archaeology* **44**(4), 538–552.

Internet Archaeology (n.d.).
URL: http://intarch.ac.uk/

Isaksen, L. (2011), Archaeology and the semantic web, PhD thesis, University of Southampton.

Jeffrey, S. (2012), 'A new digital dark age? collaborative web tools, social media and long-term preservation', *World Archaeology* **44**(4), 553–570.

JISC, R. I. N. . (2011), *Data centres: their use, value and impact*, London: The Research Information Network.

Jones, S., MacSween, A., Jeffrey, S., Morris, R. and Heyworth, M. (2001), 'From the ground up: the publication of archaeological projects, a user needs survey'.

Journal of Open Archaeology Data (n.d.), 'JOAD'.
URL: http://openarchaeologydata.metajnl.com/

Kansa, E. (2012), 'Openness and archaeology's information ecosystem', *World Archaeology* **44**(4), 498–520.

Kansa, E. C. and Kansa, S. (2011), *Toward a do-it-yourself cyberinfrastructure: open data, incentives, and reducing costs and complexities of data sharing*, Cotsen Institute of Archaeology Press Los Angeles, CA, pp. 57–91.

Kansa, E. C., Kansa, S., Burton, M. M. and Stankowski, C. (2010), 'Googling the grey: Open data, web services, and semantics', *Archaeologies* **6**(2), 301–326.

Kansa, E. C. and Kansa, S. W. (2013), 'Additional thoughts on sustaining and promoting open data in archaeology', *Journal of Eastern Mediterranean Archaeology and Heritage Studies* **1**(1), 102–103.

Kintigh, K. (2006), 'The promise and challenge of archaeological data integration', *American Antiquity* pp. 567–578.

Lake, M. (2012), 'Open archaeology', *World Archaeology* **44**(4), 471–478.

Mitcham, J. (n.d.), *Preservation of Digital Objects at the Archaeology Data Service*, Portsmouth: The University of Portsmouth, pp. 76–85.

Mitcham, J., Niven, K. and Richards, J. (2010), *Archiving archaeology: Introducing the guides to good practice*.

Open Knowledge Foundation (n.d.), 'Open data - an introduction'.
URL: http://okfn.org/opendata/

Porter, B. W. (2013), 'Sharing data is hard!–but worth it', *Journal of Eastern Mediterranean Archaeology and Heritage Studies* **1**(1), 100–101.

Pratt, D. (2013), 'Not an either/or proposition: combining interpretive and data publication', *Journal* *of Eastern Mediterranean Archaeology and Heritage Studies* **1**(1), 101–102.

Research Councils UK (2013), *RCUK Policy on Open Access and Supporting Guidance*, London.

Richards, J. D. (1997), 'Preservation and re-use of digital data: the role of the archaeology data service', *Antiquity* **71**(274), 1057–1059.

Richards, J. D. (2002), 'Digital preservation and access', *European journal of archaeology* **5**(3), 343–366.

Richards, J. D. (2009), 'From anarchy to good practice: the evolution of standards in archaeological computing', *Archeologia e Calcolatori* pp. 27–35.

Richards, J., Jeffrey, S., Waller, S., Ciravegna, F., Chapman, S. and Zhang, Z. (2011), *The Archaeology Data Service and the Archaeotools project: faceted classification and natural language processing*, Cotsen Institute of Archaeology, Los Angeles, CA, pp. 31–56.

Rumsey, S. and Jefferies, N. (2013), 'Datafinder: A research data catalogue for oxford', *Ariadne* **71**.

Snow, D., Gahegan, M., Giles, C., Hirth, K., Milner, G., Mitra, P. and Wang, J. (2006), 'Information science. cybertools and archaeology.', *Science* **311**(5763), 958.

Takeda, K., Earl, G., Frey, J., Keay, S. and Wade, A. (2013), 'Enhancing research publications using rich interactive narratives', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**(1983).

Trigger, B. G. (1989), *A history of archaeological thought*, Cambridge University Press.

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

Tudhope, D., May, K., Binding, C. and Vlachidis, A. (2011), 'Connecting archaeological data and grey literature via semantic cross search', *Internet Archaeology* **30**.

UK Government (2011), 'Science and technology committee – eighth report'.
URL: http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/85602.htm

UK Government Cabinet Office (2013), 'G8 open data charter and technical annex'.
URL: https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

Wright, H. (2011), Seeing Triple: Archaeology, Field Drawing and the Semantic Web, PhD thesis.

Gareth Beale and Nicole Beale

4 Community-Driven Approaches to Open Source Archaeological Imaging

4.1 Introduction

During the previous decade the use of two-dimensional (2D) and three-dimensional (3D) imaging techniques in Archaeology has developed at great speed, from being virtually unknown to being a conventional part of the archaeological toolkit. Imaging techniques are frequently used for the documentation and analysis of archaeological material. This trend has been facilitated by the falling costs of devices and software which allow imaging data to be captured, processed and analysed. However, in spite of these rapid declines in cost, these tools remain inaccessible to large parts of the archaeological research community. Digital imaging techniques often remain costly in comparison to traditional forms of archaeological recording and analysis. The acquisition of hardware and software can represent an unrealistic level of investment for less well funded groups working within research, voluntary or commercial sectors. These financial barriers are further compounded by limitations on expert knowledge required to capture and effectively use these data within research methodologies.

The advent of free and open source imaging software has the capacity to disrupt this pattern. Techniques such as Reflectance Transformation Imaging (RTI), photogrammetry, and multi-spectral imaging can now be undertaken with little or no investment in additional equipment (Downing et al., 2012; Jordan and Angelopoulou, 2010). However despite the availability of these tools there remain considerable barriers to uptake and consistent use. Methodologies are often written for audiences with a degree of expert knowledge which effectively precludes use by those with little or no specialist knowledge of imaging. Even where this is not the case, as groups often remain unaware of imaging technologies or the potential impact which they might have for their work.

This paper will argue that the proliferation of open and inexpensive technology represents a unique opportunity to expand and to normalise the use of imaging techniques which have typically been seen as inaccessible and requiring expert knowledge. Furthermore it will argue that the use of open source software within archaeology can, if properly supported, lead to the development of tools which more effectively

Gareth Beale: University of York, York, UK
Nicole Beale: University of Southampton, Southampton, UK

meet the needs of archaeological researchers. We will argue that in order to ensure that these techniques become accessible as well as available archaeologists must develop methodologies and strategies for distribution which engage technical non-specialists within the archaeological research community including those from voluntary, commercial and academic research backgrounds.

Limited uptake of digital imaging techniques by technical non-specialists has not only inhibited the spread of these tools but has also prevented their potential versatility from being realised. The paper will draw upon two case studies which have sought to expand the use of low-cost and free imaging techniques amongst research communities with little or no experience of archaeological computing methodologies. It will argue that in addition to expanding the community of researchers with the skills necessary to use these techniques, these efforts have also altered and extended existing perception of the value and possible applications of these tools. In part, this has been due to the development of methodologies which allow non-specialist users to influence the development of open source software.

4.1.1 Using Case Studies

The case studies detailed in the paper are the Re-reading the British Memorial Project (OuRTI) and the Basing House Community Archaeology and Technology (CAT) Project. The OuRTI project is a community documentation project which aims to spread the use of RTI amongst amateur and professional researchers already engaged in the documentation of British cemeteries and graveyards. The emphasis of the project has been on the development of accessible methodologies and sustainable organisational structures which can be used to spread use of the technique without the need for ongoing expert support. The project builds upon existing networks of expertise and aims to normalise the use of RTI as a conventional documentation and analysis tool.

The Basing House CAT Project is a research excavation at Basing House in Hampshire which incorporates researchers from a range of organisations including Hampshire County Council, the University of Southampton, Winchester School of Art, Basingstoke Archaeological and Historical Society and the University of York. In conjunction with students, volunteers and professional archaeologists the project aims to develop innovative and experimental applications for low cost and open source imaging technologies. The project places an emphasis on skills sharing and encourages researchers from all backgrounds to learn how to use digital imaging technologies as well as developing conventional archaeological skills.

Each of these projects involves the development of distinctive methodologies which are enriched through the addition of imaging technology presented in an accessible way. Pre-existing expertise and diverse perspectives on the use of technology contributes towards the development of methodologies which maximise the effectiveness of technologies and often leads to them being used in unanticipated and creative ways. Before discussing the specific methodological approaches adopted in each of these projects the following section will discuss the impact which different models of software development have upon the effective production, distribution and use of digital imaging tools within archaeology.

4.2 Technological Innovation: New Development Models

The open source and the free software movements have presented a challenge to conventional modes of technological development. Models of software development can be divided into four broad categories. Proprietary software (of which most commercial software is a part) can be bought for individual or group use, the software is made available under license to the user, and the developer maintains a certain level of control over the program. Shareware is generally available for free initially, but with restrictions on usage so that the software must be purchased to gain access to all functionality, or to continue to use it at the end of a period of trialling. Freeware is released freely but the source code is not made available, there are no development opportunities and the software is used 'as is'. Open source is available freely at source code level and can be developed and built on as well as reused for other programs. Free software must remain free no matter the re-use and re-development activities.

For software to be open source much more is required than the release of source code openly. Free software presents an added level of complexity as open source software licenses do not necessarily guarantee that software will be free. There are criteria as defined by the Open Source Initiative (OSI) from whom the open source license trademark is received. In addition to the source code being made freely available, licenses for open source software must not restrict redistribution and cannot charge for reuse. If software is produced which is derived from an earlier product then the elements of the derivative work which are re-used retain the original license. Even through redistribution, no additional licenses are required. Modification of source code is allowed, but only if permitted by the license, however, the license cannot prevent the use of a program away from the original product release, this means that all parties involved in redistribution have the same rights, the license also cannot include restrictions on other software that might be included in future distributions, and all releases must be technology-neutral, without restrictions on technology type or style (Open Source Initiative, n.d.).

Open source development would seem to represent a great opportunity for archaeology; allowing for collaborative software development which is often less resource intensive than building a product from scratch. There are alternative models for software design and release, such as freeware and shareware which offer alternative modes of acquiring low cost software but open source is unique in its ability to provide the user community with a genuine way to contribute to the development of the program.

However, in order to understand the impact of these development models upon archaeology it is important to consider the nature of the development communities and the nature of the products which are produced. The rhetoric surrounding open source software development and release tends to emphasise the altruistic nature of releasing software in this way (Hars and Ou, 2001). There is also a regular emphasis on the democratising effects of source code release and the potential for on-going collaborative development (Von Hippel and Von Krogh, 2003; Huizingh, 2011). However, it would be a mistake to assume that the use of open source software necessarily represents an inexpensive solution. Open source development increases the connection between the systems engineering, process and management (Sage and Rouse, 2009, p. 4). But there is still much that can be done to improve the link between the end user and the development of the technology. This disconnection often means that additional costs are incurred where open source tools are implemented within archaeological research methodologies. These costs can consist of time and resources spent familiarising a research teams with software which has not been subject to the same degree of user testing that a commercially produced product might have been. Alternatively, it may be necessary to modify software in order to prepare it for use in an archaeological setting.

Despite the seemingly public facing nature of an 'open' release of software, using platforms such as GitHub and sourceforge, developers of open source software are not necessarily better equipped to provide explanatory notes to accompany software releases. Open source software can often be even less accessible to non-computer specialist users than software developed using other development models. A model of software development and release that results in a product marketed at paying customers will have a greater emphasis on usability and on clear statements of purpose for users. The financial sustainability of developers using this model is often dependent upon the mass uptake of software within a targeted community. The key to the adoption of new software by archaeology is the availability of use case instances. Software which is shown to be useful is far more likely to achieve widespread use than software which is functional but largely unknown beyond technical specialists. This can even be the case with software and indeed technologies that develop a negative reputation for complicated user interfaces or expensive hardware requirements.

Open source development has huge potential to deliver tools which are built for use within an archaeological setting and developed with the requirements of archaeologists in mind. However, the fact that development takes place within an open framework does not necessarily ensure that the process of development or the end product is any more accessible to non-specialist users than the proprietary alternatives.

4.3 Introducing the Methodology

The case studies covered in the second part of this chapter make use of a variety of low cost and free technology solutions for computational imaging. Highlight RTI and photogrammetry are two methods that make up a part of the suite that was used.

4.3.1 Reflectance Transformation Imaging

Highlight RTI is a form of computational photography which uses polynomial texture mapping (PTM) in order to create an interactive file within which an object can be viewed with a moveable light source. The process is relatively straightforward and requires only a digital camera for which the aperture and focus can be set manually, a moveable light source, such as a flash with a remote trigger or a torch, and a reflective sphere. The lowest cost option for this is a black or red snooker ball. A series of photographs are taken with a static camera which has been set to allow very little ambient light into the image, relying instead on the remote light source. The images are taken one after another with the light source in a different location each time, in order to have a collection of photographs with the light projecting onto the object from a wide range of angles. The reflective sphere must also be in the photographs, but static, remaining in the same location throughout. In each photograph, the reflective sphere will show a highlight from the light source.

The open source software, RTIBuilder uses the location of this highlight to create the output file. RTIViewer software opens the RTI file and allows the user to interact with the results, producing a variety of visualisations from the data. These include the production of a normals visualisation and also the use of specular enhancement or diffuse gain to provide images with views of edges or surface details not apparent in a conventional raking light photographs.

4.3.2 Photogrammetry

Photogrammetry is a means of calculating 3D geometry based upon a series of still images of an object or scene. Recording for a photogrammetric model is relatively straightforward. Some parameters must be set on the digital camera to ensure calibration before collecting photographs, these include the focal length and format aspect ratio, and planning to compensate for lens distortion is important. Photogrammetric models can be created by taking a series of photographs of an object or scene where the distance of the camera from the item is maintained, but the camera is moved around to cover as much of the surface area as possible. The photographs must have a substantial overlap. The chosen software compiles the images together to automatically create a 3D mesh from those images. There are various options for photogrammetry processing, both free and commercial.

4.3.3 Assessing the Benefits of Open Source Imaging Methodologies

The open source software solutions are often cheaper in cost than their proprietary counterparts, but this is not necessarily the only cost involved. If we consider the example of three-dimensional data capture and processing, in one of the case studies to be discussed in this chapter, a combination of Highlight RTI with photogrammetry is used in replacement of a laser scanner. There are several open source solutions for the compiling of photographs to create 3D meshes for photogrammetry and remote sensing of small and large objects (from coins to building facades), and so results are often very varied. Despite hardware and software solutions often being complicated to install and to use, these methods are being used increasingly by archaeologists, often in lieu of the availability of a laser scanner, as the data capture requires equipment that is generally available to the archaeologist, such as a digital SLR camera.

There are benefits and disadvantages to using both approaches. Laser scanners are accurate and the results are reliable, however recording even the smallest object is a time consuming exercise. Photogrammetry on the other hand is a method which produces a less detailed 3D mesh than laser scanner results and is often unreliable, but recording using a camera is much faster than using a laser scanner. In addition to this, when used in conjunction with Highlight RTI, which produces highly accurate surface imaging, although not currently linked to 3D meshes, this is less problematic for many projects' requirements. The two approaches have differing suitability depending on the situation, but one selling point of photogrammetry combined with Highlight RTI over laser scanners is the price tag. A terrestrial laser scanner is a costly piece of equipment, even when rented for short time periods there is a substantial cost that must be considered. Photogrammetry combined with Highlight RTI, because of the open source and also freeware options for data processing, is far cheaper. The methods require only a digital camera for data capture. However, the cost for photogrammetry with highlight RTI is high in other ways. Open source software options are often less user friendly than costly proprietary alternatives, and the software for Highlight RTI is an example of this.

RTIBuilder is the open source software used for Highlight RTIs. The software is released under GNU General Public License, which means that it is free software. This is slightly different to open source software. The OSI has accepted licenses, such as the Reciprocal Public License 1.5 (RPL-1.5) which allow for non-free software. This license would not qualify under the definition of free software from the Free Software Foundation (Free Software Foundation, 2012). The license used for RTIBuilder is a free software license, and so there is more freedom for development. RTIBuilder was funded and developed by a team at the Universidade do Minho in Braga, Portugal, with additional funding and input from Cultural Heritage Imaging. The current release of the software (2.0.2) incorporates HSHFitter (GNU General Public License version 3, copyright University of California, Santa Cruz and Cultural Heritage Imaging 2007 – 2009). RTIBuilder requires either HSHFitter or PTMFitter (Binary Code License Agreement, by Hewlett Packard) to work. PTMFitter is not compatible with the GNU license and so must be downloaded separately.

As with many open source software, RTIBuilder has been amended and improved over time since its inception more than five years ago, and consequently there are numerous steps and requirements for the processing of an RTI. The user manual currently available to download on the Cultural Heritage Imaging website is over 20 pages in length (Schroer and Bogart, 2011). The instructions are excellent and the steps in the manual are easy to follow. However, a manual of this size and complexity for one part of the process can be a daunting discovery for an enthusiastic computational photography beginner.

4.4 Community Methodologies and Technological Uptake

The availability of a technology and the fact that it has the technical capability to fulfil a specific methodological role does not ensure its adoption by members of the archaeological research community. It is necessary also to ensure that available tools are accessible to the community in question and that they are perceived to be relevant. In terms of the development of computational tools the concepts of availability, accessibility and relevance are intimately connected.

4.4.1 Development Model Alternatives

A key failure in the uptake of technology is matching the need of the user community to the technology itself, even in projects using participatory design for agile software development with grounding in complex adaptive systems theory, end-users can become a secondary consideration (Kautz, 2011). This can lead to a disconnect between users and developers, particularly in the open source community where users tend to have a higher level of computational knowledge than those using a more 'out of the box' type of software or hardware solution.

Archaeology has traditionally embraced new technologies, and is an area within which much innovation has occurred driven by the user community rather than the developer community (Sillar and Tite 2000; Wheatley et al. 2002, p. 2). In a world where these definitions are blurring, we find ourselves at a crossroads between continuing as consumers of technology developed within other sectors, or as producers of technology within our own sector. Open source makes the latter possible, but there are still substantial hurdles that must be overcome in order for the full potential of open source software and hardware development in archaeological imaging to be realised.

There are additional requirements to the technology for a change to occur. Up until now, there has been a focus on computational photography experts within archaeology collaborating with or contributing to computational photography experts outside of the sector to develop the technology. Archaeology has long had a tradition of involvement from amateurs (Kelley, 1963; Levine, 2003; Taylor, 1995; Moshenska, 2010). There is potential user group and development community that has as of yet not been mobilised for the improvement of archaeological imaging. Community methodologies can engender new forms of technological practice.

A common frustration between researchers and developers has been the failure of transfer of ideas from one to the other (Suchman, 1993, p. 23). The open source software model offers a way to surmount this problem by placing the emphasis on the research, development, testing and redevelopment of software by all of the community. In this way it is possible to ensure that the software under development does not only have the necessary functionality but that it also reflects the desires and requirements of the user community. This can include factors such as the use of appropriate vocabulary, the layout of the interface or the ability to annotate data. These factors are what ensure that a piece of software is not just functional and available but also accessible and relevant to the community for which it is being developed. GitHub, one of the major platforms for the releasing of open source software code is popular as a space for the development of software from an idea to a product as it facilitates collaboration. Used appropriately these collaborative platforms and ways of working can help to ensure the relevance and usability of source code once it is released.

Community based methodologies that allow non-experts to contribute more readily to product development offers a solution to improving the gap between research and product. The open source development model provides a scaffold upon which to build this methodology, but additional components are needed to ensure successful uptake of open source solutions both within and outside of professional archaeology.

4.4.2 Ensuring Meaningful Software Access

What is needed for successful uptake of any new technology is the possibility for development not just of the source code, but of the methodology surrounding the adoption and adaption of the technological solution. An example of this is PTMFitter; the software which underpins RTI. The uptake of the technique of PTM has been driven by the cultural heritage sector, in particular by Cultural Heritage Imaging (CHI), which has forged a new path for the use of PTM as part of the RTIBuilder and RTIViewer software (Mudge et al., 2010).

Not only have the organisation worked with other communities to develop the software, but they have also written handbooks for data capture and processing, and have published papers encouraging the academic and commercial community to use the method. CHI has created numerous 'How To' videos and delivered many training workshops to archaeologists, historians, archivists and others. PTMFitter as a standalone open source option for computational imaging is not enough to lead to the uptake of RTI as a key part of computational imaging for the area of archaeology. RTIBuilder represents a substantially more user friendly solution to the use of polynomial texture mapping for the recording of objects. The development of RTI builder in conjunction with the anticipated user community has helped to ensure that the software is appropriate for use within the context of cultural heritage. However it has been through the development of an RTI user community within the cultural heritage sector that the technique has truly developed into a popular research tool.

There are also open source and free software solutions for photogrammetry. ArchTeam's ArcheOS (an archaeological operating system using Debian Squeeze) which incorporates Python Photogrammetry Toolbox, or VisualSFM from the University of Washington offer open source solutions for the creation of photogrammetric models from images, but the installation and use of these options is complicated and requires computational experience. There are alternative options, such as the Arc3D web-service produced by the Epoch Network of excellence, but results from this are varied and tend to need substantial reworking. What is needed now is a pipeline for the use of a suite of options to record objects, which will open up the potential of computational imaging to outside of the small group of experts who currently make use of the technique.

Non-professional archaeology communities can allow us to question conventionalised narratives relating to the function, value and purpose of technology. As we shall see, it is important that we develop methodologies which allow us to take these groups into account. A flexible methodology is required to allow for a more inclusive user community to develop.

4.5 Case Study One: Basing House Community, Archaeology and Technology Project

The Basing House CAT project is a multi-organisational research excavation carried out by paid staff, archaeology and history student, and volunteers. The project consists of a summer fieldwork season during which excavations are carried out on the remains of the multi-phase site of Basing House, a Tudor period fortified house built on top of a Medieval motte and bailey castle with keep. The house was partially destroyed after years of sieges during the English Civil War (Allen et al., 2013). Material has emerged from the site which suggests that occupation pre-dates the earliest Norman phase evident in the standing remains.

The Basing House CAT project has also offered a unique environment within which to develop and to refine the use of imaging technology. Unlike the OuRTI project, activity at Basing House is confined to a brief four week field season each summer. During this time fifty staff, students and volunteers can be on site at any time. During the 2013 season at least half of these team members were trained in the use of Highlight RTI and photogrammetry (Figure 4.1).

One of the core objectives of the Basing House CAT project is to promote methodological innovation within field archaeology, particularly in the use of digital imaging. Project members are encouraged throughout their time onsite to consider how imaging techniques might be usefully employed within their work. The excavation is run on a rota system with all participants having the opportunity to excavate but also to contribute to other elements of the project including building survey, the documentation of standing remains, public outreach and the documentation of the entire project as it proceeds. The latter activity takes many forms including a project blog which is updated daily during the field season, a photographic record of the excavations and work undertaken in collaboration with an artist in residence.

The project represents an ideal setting for the intensive development of experimental imaging methodologies. As well as the variety of activities within which imaging techniques can be applied the project also benefits from the fact that the onsite team have an enormous variety of experience and specialisms. Team members conventionally include archaeologists, historians and artists but volunteers with professional careers outside of academia or commercial archaeology contribute a vast range of additional skills and experience.

Figure 4.1: Basing House CAT project team members use a GoPro camera, along with the remote controlling iPad app, to photograph the trenches at the end of the excavation season. The images were used to create a photogrammetric model of the excavation

4.5.1 Skills Sharing

The working environment of the fieldwork project coupled with the diversity of skills and experience present within the excavation team mean that the sharing of skills is inevitable. While engaged in other activities team members are encouraged to think creatively about how they could apply the skills learnt elsewhere on the project to the situation within they find themselves. As a consequence methodological innovation is not confined to the use of imaging technologies.

The mixed skill sets coupled with this approach mean that unexpected approaches to the use of technology are a regular occurrence (Beale et al., 2013). The situation helped to ensure that where innovations occur they are rapidly refined and developed in response to expert advice from other team members. This helped to encourage relatively inexperienced team members to suggest new approaches and helped to ensure that these suggestions were often developed into effective methodological strategies. One example of this followed a group visit to the village church of Old Basing very close to the excavation led by a church archaeologist and team member. An undergraduate student proposed that the techniques being applied to digitally image the site might also be used to document signs of English Civil War activity in the church including iconoclastic damage to statues. This suggestion of this project (which should be completed during the coming field season) came about as a result of the students' knowledge of the imaging techniques used onsite but also as a result of expert advice from an experienced specialist.

4.5.2 Experimental Atmosphere

The encouraging of experimental approaches to technology, while not always successful has regularly led to innovative and highly successful trials. These have included the use of Highlight RTI to document sections in addition to the use of conventional section drawings, the use of Highlight RTI to assess the difference between original Tudor bricks and 19t̂h Century replacements. RTI as a means of documenting sections was championed by an experienced volunteer excavator who saw the potential benefits of this recording technique after having seen demonstrations of RTIs of small archaeological objects. Highlight RTI images of sections enabled the detailed documentation and subsequent analysis of sections in addition to the use of conventional drawings and photographs. Under certain circumstances, especially situations within which objects and complex features were evident in the sections, this technique proved to be a valuable aid to presentation and subsequent explanation.

The use of Highlight RTI to differentiate between 19t̂h century bricks and Tudor originals was suggested by an undergraduate student who had particularly enjoyed learning to use the technique and was tasked with finding innovative applications. The project is still underway and so the effectiveness of the technique in this context is not yet known. However, the fact that comparatively inexperienced students were addressing complex conservation issues and considering how new technologies might be employed in this setting is noteworthy in its own right.

Finally and perhaps most significantly was the use of RTI to document Roman Coins found on the Basing House excavation. The desire to identify these coins, several of which were badly corroded led to multiple attempts by different project members to produce effective visualisations using Highlight RTI and photogrammetry. The activity quickly became playfully competitive leading to extremely high quality visualisations and identification and of the coins. The major benefit though came from the communication between project members as they sought to produce high quality visualisations. An emphasis on problem solving rather than abstract acquisition of another new skill meant that participants very quickly refined their technique and familiarise themselves with the nuances and challenges of 3D and 2D imaging techniques in a fieldwork scenario.

4.5.3 Basing House Conclusions

Projects like Basing House CAT provide an opportunity to develop imaging methodologies within the context of a broader field archaeology project. This helps to ensure that methodological innovations are driven by the demands of the research process and not solely by the capabilities of the technology. Within an atmosphere of diverse skills and experience it would quickly become apparent if a new approach was ineffective or unnecessary allowing the quality of new approaches to be rapidly honed. Subsequent to the field season lessons learned were integrated into new methodologies which can be used in future seasons and distributed to other field archaeology teams. As well as allowing the development of methodological approaches, the experimental use of imaging techniques within this setting also produced a list of weaknesses in software which have subsequently been incorporated into plans for on-going software development.

4.6 Case Study Two: Re-Reading the British Memorial

The OuRTI project came about in 2012 following the design and delivery of a Lifelong Learning module, Urban Archaeology, by Gareth Beale and Nicole Beale at the University of Southampton (Figure 4.2). This module aimed to introduce adult learners to the potential of using archaeological techniques to learn more about their local heritage environment. Over twelve weeks, the course introduced and reviewed numerous technologies which could be used to find out about a series of themes relating to urban archaeology, such as the built environment, industrial heritage, cultural diversity, and family links. Throughout the module, the focus was on providing realistic options for learners to adopt for their own research. Learners came from all over the county of Hampshire and were from varied backgrounds, the course coordinators therefore had to ensure that the technologies and methodologies being taught were available to individuals with no connection with academic, third sector, or commercial archaeological organisations, and with little or no access to expertise in the technology being introduced. The majority of adult learners at the University of Southampton do return over a period of years to modules within the same discipline, but continued access to University resources is not a given. Urban Archaeology therefore relied upon low cost or free technology solutions. This often meant that software was open source or free software and that hardware was not the most recent incarnation of that particular technology. Part-way through the module, the lecturers introduced computational imaging for archaeology. The concept captured the imagination of the class, and ideas and technologies were returned to over the following weeks. There seemed to be a significant interest in the ways that computational imaging could be used by amateur archaeologists not just for dissemination of findings but for recording and interpretation.

Figure 4.2: OuRTI team members giving a Highlight RTI demonstration at Royal Garrison Church, Portsmouth

The module had an emphasis on practical demonstration and on hands-on learning, and so as part of the computational imaging tutorial a workshop was held at a local church. Churches are rich case studies for an introduction to the principles of computational imaging for recording. They are sites full with artefacts and architecture, and these bring with them challenges that are present across the discipline. The church had been identified earlier by the module team as being a useful test bed for a suite of technologies being used by the Archaeological Computing Research Group and had been visited a number of times by the module conveyors. A workshop had been organised for the history group aligned with the church to introduce RTI as a method for interpreting badly eroded and damaged headstones in the church graveyard. The workshop had been very popular and had led to a surveying project with the group. Following the Urban Archaeology and the training that had been delivered at the church to the local history group, a number of challenges and research questions had been identified.

A team was formed to begin a project that would test the usefulness of RTI and other open source or free software solutions for the recording of graveyards, cemeteries and church memorials by special interest groups through a series of workshops (Figure 4.3). The project won a small amount of Digital Humanities funding from the University of Southampton to facilitate a pilot of case study identification and workshop delivery (Beale and Beale, 2012).

The OuRTI project aims to facilitate the use of computational imaging techniques for the recording and interpretation of church memorials. Since the project's inception, the team have worked with numerous local history groups and special interest groups to record graveyards and memorials inside churches using technologies such as RTI and photogrammetry. The project has also incorporated additional community requirements as have been requested by specific groups, such as surveying techniques and data management.

4.6.1 Adaptive Methodologies

OuRTI involves diverse communities of expertise, and this is a key strength of the approach established early on in the project. Generally, a case study will begin with contact between the project team and a group of individuals working alongside the church staff to carry out a specific research project at the church. This can range from attempting to identify particular individuals' resting place within a graveyard to carrying out a full-scale recording survey of a church's interior. A key element of the project is that there is no control over pre-existing structure of organisations within each case study. With each case study, the hierarchy must be respected, and the objectives for that particular case study must be negotiated. All case studies agree to allow the data collected as part of the project be copied and archived by the OuRTI team. This ensures that there is a comprehensive record of the data, but also that there is a collection of RTIs that showcase the potential of the technology and can be used as resource for the study of church memorials, as well as being useful for the teaching of processing archaeological imaging data.

For every case, the project team meets with church representatives and gives a demonstration of the technologies available. The consultation period is essential to the working of the project, as each case study has a different need and the techno-

(a) No treatment
(b) Diffuse gain
(c) Normal unsharp masking
(d) Image unsharp masking
(e) Specular enhancement
(f) Normals visualisation

Figure 4.3: Results of an RTI from St. Andrew's Church, Holcombe

logical solution is tailored to their requirements. The most commonly requested technology is RTI, as the visually impressive nature of the results are easy to engage with and to interpret. Interestingly, the project team assumed that RTI and photogrammetry would be the most popular of the technologies because the capture process, when seen as part of a demonstration, seems to be the most straightforward and require the least amount of specialised equipment. In fact, in the majority of cases, the group representatives do not enquire as to the nature or details of the capturing process, and the simplicity of this part of the methodology is not witnessed until the training is delivered. This is representative of the significance of the outputs of technological solutions being easily readable, as these seem to be the factors that impact on the decision for take-up, rather than the ease of recording being a deciding factor.

4.6.2 Decentralised Approach

From the outset, the OuRTI team encourages a decentralised approach to organisation of the sub-project. The importance of knowledge of photography is emphasised so that each case study is aware from the outset that the involvement of someone with photography experience will be useful. Whilst not essential, this involvement is emphasised at the first conversations between the OuRTI team and the church representatives. The project team acknowledge from the beginning of the project that they are merely acting as facilitators between the technological solutions and the church representatives. Recognising the specialisms already present within the sub-project is an essential part of the project. This motivation is two-fold.

4.6.3 Project Sustainability

Firstly, the project methodology must be sustainable. As OuRTI is managed by a small team of archaeologists who are employed elsewhere, the project is unable attend in order to train and then support every church-based group in the UK to use computational photography techniques. But a major aim of the project is that every church-based project is able to do this. This means that alternative approaches needed to be identified in order to make this possible.

4.6.4 Recognising Skills

Secondly, there is a latent resource present in the form of special interest groups in the UK which have skills that could contribute in a very real way to the facilitation of computational photography for church recording. These groups are a resource that is currently under used, and the OuRTI project identified very early on that these groups could provide technical support that was not possible from the initiating project team. Models such as the online crowdsourcing of data interpretation by organisations such as Ancestry.com and large national cultural heritage institutions such as the Victoria and Albert Museum, illustrate the potential of harnessing the energy and specialist knowledge base of people working at an amateur level (Oomen and Aroyo, 2011; Smith-Yoshimura et al., 2012). The BBC and Public Catalogue Foundation (PCF) managed Your Paintings project provides an excellent example of this as it allows for advanced annotations to the crowdsourced metadata being collected around images in the collection. Your Paintings is digitising and sharing online images of oil paintings held in collections across the UK (Greg, 2011). The Your Paintings Tagger tool is part of the project and aims to improve the discovery possibilities of the data being collected by crowdsourcing metadata from online contributors. The tags being collected range from the identification of basic attributes such as colour and easily identifiable components, to allowing users with specialisms such as art historians, or with specific knowledge about aspects contained within the images, like architectural details or period costume details, to contribute more advanced data relating to the images.

The Your Paintings project recognises that there are skills that are not directly related to the analysis of oil paintings but that can contribute to improving the data around the images. In the same way, OuRTI aims to harness the skills of community members who have not necessarily identified in the past with the archaeology of churches, but who have an interest in a particular technology that could support work being done. The two key groups for the Re-reading the British Memorial project are photography enthusiasts and those involved in careers or hobbies that include the use of computers. The OuRTI team make use of the fact that technology can attract new people to projects for which they do not immediately identify an interest. In the instance of the OuRTI project, the case studies' use of computational photography techniques to carry out the tasks that were necessary for the recording, interpretation or dissemination projects that were being undertaken drew in additional participants. At one church, a husband of a local history group came to learn how to install and use the RTIBuilder and RTIViewer software. In another location, younger friends of a project participant visited the church to find out how to expand their knowledge of digital photography to incorporate the techniques being taught. Frequently, locally based individuals have visited recording projects whilst they have been occurring in order to find out more about a technology that they had heard was being demonstrated. In many cases, these individuals remain a part of the recording project.

The project also identified that there are instances in which some aspects of a person's knowledge base that may be thought of as exclusionary, could in fact turn out to be an essential component in leading to the community at a church thriving and expanding. For example, at one location two individuals in the group were reluctant to participate as they had no experience in using digital cameras. There was a concern from the group as a whole that this may slow down the learning process for the creation of RTIs. It was not long into the morning's workshop that it became clear to all in the room that the two individuals had been keen photographers in their youth, and that this had led to a good grounding in the principles of analogue photography as they had used film SLR cameras for many years. These two group members were much quicker to understand the methods being employed in the setting up of the camera to record a highlight RTI and showed other members how to use the manual settings on the digital SLR cameras, such as adjusting exposure, aperture and shutter speeds.

Including new user groups within the case studies for the OuRTI has led to the technologies being used in unanticipated ways. Local history groups, digital photography groups and software specialists have all engaged with the technologies in different ways. Combining these skills in each case study and supporting the use of the technologies for research questions as defined by those groups rather than by the OuRTI team, has allowed for extensive and detailed user testing to occur. The project draws upon each individual involved in each case study in different ways, and works to allow for a flexible methodology design for every instance of the project at each church. No two case studies are the same, the requirements of the community are different and it is therefore taken as a given that the requirements of the technology will also be different. The team avoids making assumptions about the ways in which the technologies will be used, and in so doing avoids taking a technologically deterministic approach to applying technologies to the problem.

4.6.5 Re-Reading the British Memorial Project Conclusions

There is a tendency in software development to focus on the importance of the relevance, the usability, and therefore identification of the value of software. The appropriateness is indeed important, and we see in the instance of the RTI for cultural heritage community that there is a bias towards making the software ‘better’, with incrementally improving user interfaces with each software release. The OuRTI project has found that an alternative approach is to identify potential new community users, who have an advanced understanding of the nuances of their own specialist area and its requirements. If the appropriate technology and the appropriate support mechanisms are put in place to allow for use of the technology, then these individuals will modify their own practice to make use of the technology that is available, whilst also contributing in a real way to the development of the software. This is a further form of user testing and one that pushes the technology to it limits but from all directions, resulting in new and exciting uses of the software or hardware that could not have been predicted by the originating community.

4.7 Conclusion

The availability of open source imaging technologies has the potential to have a significant impact on conventional archaeological practice. The low costs associated with the use of these techniques mean that they are available to virtually all archaeological researchers who have access to a camera and a computer. Beyond issues of cost, the open source development model also creates the possibility of tools and methodologies which are developed by archaeological researchers with archaeological research in mind. This overturns the traditional model of technical adoption in archaeology which has required archaeologists to adopt technologies developed for other applications. Not only then are these tools available, they have the capacity to be highly relevant and useable.

However, as outlined above it is not enough simply to make tools available and to assume that people will locate them, acquire them and learn to use them. Skill levels, restrictions on resources and perceptions regarding the capabilities of technology can all prevent the full potential of these tools from being realised. In order to capitalise on the availability of free or cheap digital imaging methods it is essential that methodologies are developed which take into account the requirements and perceptions of specific user communities.

This paper has demonstrated, through the presentation of two case studies, the flexibility and diversity of approach which is required in order to assure adoption by two very different communities of researchers. The projects described here demonstrate the degree of methodological innovation and creativity which can occur when digital imaging technologies are used in conjunction with existing techniques and in dialogue with communities of specialists. In both cases these researchers not only enthusiastically embraced these technologies and new ways of working in the area of archaeological imaging, they also contributed to the development of methodologies and the training of other researchers.

In order to ensure that digital imaging techniques are widely adopted it is important to make sure that they are not just theoretically available to groups who might benefit from their use but that they are also accessible. The development of methodological approaches which specifically enable particular groups to engage with and use digital imaging technology has been highly successful both as a means of promoting the use of these tools and also further refining and developing methodologies which improve the user experience of future users. The involvement of different communities of practice has, as detailed in the case studies above, served to ensure that the limits and capabilities of the techniques are explored.

Bibliography

Allen, D., Beale, N., Elmer, C., Jones, J., Strutt, K., Beale, G., Allen, C. and Jones, D. (2013), 'Basing house spring-summer 2013: Interim report', Hampshire County Council and University of Southampton.
URL: http://basinghouseproject.org/2014/02/21/interim-2013-intro/

Beale, N. and Beale, G. (2012), 'The potential of open models for public archaeology', *Digital Futures 2012 (DE2012)* .

Beale, N., Beale, G., Dawson, I. and Minkin, L. (2013), 'Making digital: Visual approaches to the digital humanities', *Electronic Visualisation and the Arts (EVA 2013)* pp. 240–247.

Downing, J., Murray, A. A. and Harvey, A. R. (2012), Low-cost multi-spectral imaging camera array, *in* 'Computational Optical Sensing and Imaging', Optical Society of America, pp. JW1A–6.

Free Software Foundation (2012), 'Various licenses and comments about them', GNU Operating System.
URL: http://www.gnu.org/licenses/license-list.html

Greg, A. (2011), 'Your paintings: public access and public tagging', *Journal of the Scottish Society for Art History* **16**, 48–52.

Hars, A. and Ou, S. (2001), Working for free? motivations of participating in open source projects, *in* 'System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on', IEEE, pp. 9–pp.

Huizingh, E. K. (2011), 'Open innovation: State of the art and future perspectives', *Technovation* **31**(1), 2–9.

Jordan, J. and Angelopoulou, E. (2010), Gerbil–a novel software framework for visualization and analysis in the multispectral domain, *in* 'Vision, Modeling, and Visualization (2010)', The Eurographics Association, pp. 259–266.

Kautz, K. (2011), 'Investigating the design process: participatory design in agile software development', *Information Technology & People* **24**(3), 217–235.

Kelley, J. H. (1963), 'Some thoughts on amateur archaeology', *American Antiquity* pp. 394–396.

Levine, P. (2003), *The Amateur and the Professional: Antiquarians, Historians and Archaeologists in Victorian England 1838-1886*, Cambridge University Press.

Moshenska, G. (2010), 'What is public archaeology?', *Present Pasts* **1**.

Mudge, M., Schroer, C., Earl, G., Martinez, K., Pagi, H., Toler-Franklin, C., Rusinkiewicz, S., Palma, G., Wachowiak, M., Ashey, M., Matthews, N., Noble, T. and Dellepiane, M. (2010), Principles and practices of robust, photography-based digital imaging techniques for museums, *in* 'The 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, VAST 2010'.

Oomen, J. and Aroyo, L. (2011), Crowdsourcing in the cultural heritage domain: opportunities and challenges, *in* 'Proceedings of the 5th International Conference on Communities and Technologies', ACM, pp. 138–149.

Open Source Initiative (n.d.), 'The open source definition'.
URL: http://opensource.org/osd

Sage, A. P. and Rouse, W. B. (2009), *Handbook of systems engineering and management*, John Wiley & Sons.

Schroer, C. and Bogart, J. (2011), 'Reflective transformation imaging: Guide to highlight image processing v1.4', Cultural Heritage Imaging.
URL: http://culturalheritageimaging.org/What_We_Offer/Downloads/rtibuilder/RTI_hlt_Processing_Guide_v14_beta.pdf

Sillar, B. and Tite, M. S. (2000), 'The challenge of 'technological choices' for materials science approaches in archaeology', *Archaeometry* **42**(1), 2–20.

Smith-Yoshimura, K., Cronin, C., Godby, C. J., Lovins, D. and Raine, H. (2012), 'Social metadata for libraries, archives, and museums: executive summary'.

Suchman, L. (1993), 'Working relations of technology production and use', *Computer Supported Cooperative Work* **2**(1-2), 21–39.

Taylor, B. (1995), 'Amateurs, professionals and the knowledge of archaeology', *British Journal of Sociology* pp. 499–508.

Von Hippel, E. and Von Krogh, G. (2003), 'Open source software and the "private-collective" innovation model: Issues for organization science', *Organization science* **14**(2), 209–223.

Wheatley, D., Earl, G. and Poppy, S. (2002), *Contemporary themes in archaeological computing*, Oxbow Books.

Hector Orengo

5 Open Source GIS Geospatial Software for Archaeology: Towards its Integration into Everyday Archaeological Practice

5.1 Introduction

Geospatial software refers to all software employed to display, manipulate, analyse and produce geospatial information. Although this term is normally used for software classified as geographic information systems (GIS) it also includes others that can work with and produce geographically referenced information such as CAD, Remote Sensing or photogrammetric software. This paper focuses on GIS but references will be also made to other types of geospatial software that can work in conjunction or complement GIS.

During the last 20 years GIS have played an increasingly important role in archaeology. Their capacity to work on a multi-layered and multi-scale spatial frame renders them a very flexible tool to handle all spatially referenced information. Nowadays, they are routinely employed in archaeological research and practice with more researchers applying them to address not only landscape but also on site issues. However, their incorporation into the archaeologist's toolbox is relatively recent. The first GIS software (if they could have been called as such back then) were developed in the 60s and they were restricted to very basic functions, such as the creation of 3D messes or the extraction of contours. During the 70s the first commercial software was developed by the Environmental Systems Research Institute (ESRI), the most successful vendor of proprietary software today. The first applications, however, of GIS to archaeological research were only later developed under the influence of New Geography by the American pioneers of New Archaeology. The positivist approach of New Archaeology was an ideal framework for GIS. Operations such as the delineation of Thiessen polygons, site catchment areas or the development of predictive models of the location of archaeological sites were simple and straightforward operations for GIS, and therefore, had a prolific application during the late 80's and early 90's.

These first GIS were aimed to work with either vector or raster data but nowadays, most available GIS software packages have evolved to include an increasing number of functions that have been usually regarded the realm of other types of software (Wheatley and Gillings, 2002, p. 9). Nowadays it is quite common using GIS as a CAD program, to develop 3D reconstructions and virtual environments, to perform digital

Hector Orengo: University of Sheffield, Sheffield, UK

image analysis operations and remote sensing analyses, database management and query, and spatial statistics.

Archaeological research is one of the most active fields in the application of GIS today. This is perhaps a consequence of the often-subtle nature of archaeological evidence. Archaeologists are concerned with human behaviour in the past and, it has to be admitted, the traces left by it are mostly elusive and difficult to discern. While most GIS users are concerned with the physical environment as it is today, landscape archaeologists (the group most involved in the use of GIS) have to employ a much wider variety of GIS-based techniques to be able to discover, analyse and compare traces of past human activities. It is therefore common in archaeological research to integrate GIS with remote sensing and other geospatial techniques. Another limiting factor is that, in many instances, current environmental datasets can rarely be applied to solve archaeological questions since large-scale landscape modifications during the last centuries could have been important and these need to be taken into account when studying ancient environments. Also, the interest of the landscape archaeologist lies in the study of societies in the past, in how they interacted with their environment and in the consequences of these interactions. This is a particularly difficult topic to explore since it does not only take into account environmental variables but also social, cultural, temporal and economic inputs for which information is often lacking. Research in archaeology and the humanities also incorporates subjective and imprecise data that can be difficult to map (for a more thorough discussion see Jessop 2008). In order to explore all these aspects archaeologists use most techniques available in the GIS toolbox, such as topographic analyses, including several types of visibility analysis, and least cost route modelling; predictive site location modelling, hydrographic analysis, network analysis, and so on.

Archaeologists focussing on site-oriented research or excavations have also stretched GIS resources to their limit due to the very high spatial resolution and exactitude required in the recording of the excavation process, its finds and features. Nowadays archaeological recording procedures include the use of complex spatially aware object-oriented databases, 3D reconstruction techniques, and so on. Perhaps this is further illustrated by the multiple scripts developed specifically for archaeological purposes or the adaptation of open source GIS software to archaeological research.

The number of techniques currently put into archaeological enquiry is abundant and chances are that a single project will require the use of many of those. However, it should be noted that no single GIS includes all possible functionalities. As in most other types of software the selection of the software package is an essential one and some familiarity with GIS is needed in order to choose the most appropriate software suite for each research question/project. Nevertheless, most GIS are modular in nature, that is, they are composed of a collection of modules directed to do specific functions. The majority of GIS can easily increase their capabilities by integrating new modules (also called extensions, scripts, plugins or add-ons) to add specific functionalities. As GIS users become aware of the particular functionalities and modules available in different packages, it is more common to employ several software packages, performing specific tasks or analyses using the most adequate or convenient module to the task at hand.

5.2 What is Open Source Software?

Open source software (OSS) is any computer software the source code of which is made available through a license according to which the copyright holder provides the right to download, analyse, modify and distribute the software to anyone and for any purpose (including commercial). "Free and open source software" (FOSS) is another commonly employed term that includes both open source and free software. As stated by the Free Software Foundation, the term "free" refers to freedom more than to "gratis" although, FOSS, at least as described by the Open Source Initiative and the Free Software Foundation, is by definition no-cost. Indeed, as stated by the Free Software Foundation:

> "The term "open source" software is used by some people to mean more or less the same category as free software. It is not exactly the same class of software: they accept some licenses that we consider too restrictive, and there are free software licenses they have not accepted. However, the differences in extension of the category are small: nearly all free software is open source, and nearly all open source software is free."

Also, there is plenty of closed source software (often proprietary) distributed at no cost through freeware, shareware or other types of licenses. That must be kept in mind since, as it will be indicated in later sections, this causes strong restrictions in the development of an online community of contributors and the long-term improvement and maintenance of the software. Proprietary software is also commonly referred to as commercial although, as Ducke has stated (2012, p. 572), this might be a somewhat misleading term because most open source software is developed through hired programmers and therefore involves commercial processes.

In this chapter the terms OSS and FOSS will be used interchangeably. All the software discussed in the text will refer to open source unless stated otherwise.

5.3 Why Use Open Source GIS?

There are some evident advantages in the use of open source GIS for archaeological research and indeed much literature has been dedicated to it (see, for example, Serlorenzi 2013). One of these is the transparency of open source software where the source code can be studied and changed. Although most archaeologists do not have programming skills and therefore cannot take direct advantage of this feature, the availability of source code makes it open to public scrutiny and improvement (Chambers, 2008, p. 6–7) and, as a consequence, most (or at least those widespread) open source GIS software is today very reliable and rapidly evolving. As pointed out by several authors (e.g. Hafer and Kirkpatrick 2009, p. 126; Ducke 2012; Ince et al. 2012), open source software should be in fact a requirement in research since the principles of scientific enquiry demand research to be open to scrutiny. Proprietary software's inaccessible source code, not being open to public scrutiny and reproducibility, prevents understanding of the processes involved in the production of outputs and thus their critical evaluation. As Morin et al. have stated (2012, p. 159), close source programmes act as 'black boxes', effectively hindering good scientific practice.

Another very important advantage of open source software, and one with many consequences (see below), is that it is distributed at no cost and it is publicly available. The very high price of proprietary GIS software often prohibits its acquisition by small enterprises or research centres, which would be obliged to spend tens of thousands of pounds for a single license (not including extensions) that needs to be renewed constantly. Neither archaeological research nor commercial archaeological practice can be included in those disciplines with higher returns on investment, and therefore, the use of proprietary geospatial software in archaeology is particularly taxing.

A consequence of the free availability of open source programs is that they are portable, that is, the user can have as many copies installed in as many computers as necessary. In other words, the user does not need to go to a specific workstation in a lab in order to access a GIS-enabled computer. Providing the user has the necessary permissions to install software, open-source software can be installed in any computer. Due to the transferability of open source GIS, having experience in the use of open source GIS programs is an increasingly important factor in the geospatial job market. Open source GIS users are able to move to different working environments and to continue being productive without the need to adapt to the specific software choice available at their new location, since they can always continue using their preferred open source GIS at no additional cost for the employer. This is in line with the tendency towards a growing market for open source GIS users. Enterprises with geospatial analysis needs can save significant amounts of money in adopting FOSS or changing their proprietary software for FOSS and accordingly they might be interested in their personnel to have experience with open source GIS.

Free software not only means a substantial saving of money but also implies the existence of a vibrant online community of users and multiple learning resources, which include tutorials and other learning resources, such as datasets. The most widespread open source GIS like GRASS GIS or QGIS offer plenty of manuals, tutorials and video tutorials, online help and multiple busy forums maintained by thriving communities where the user can learn and discuss applications. Following the open approach to software distribution, manuals and tutorials are available to download for free. This renders FOSS GIS an excellent choice for both the amateur and the experienced user.

Also, as a consequence of their large online community, the most senior GIS projects have been running for a long time and therefore have evolved into very secure and competent packages. They continuously release updated versions in which 'bugs' or programming errors are solved and new functionalities implemented. As most of these are to some degree dependent on public participation and online communities of users and developers they are also very sensitive to user demands and feedback and usually incorporate improvements and innovations based on these (Brandtzæg et al., 2010).

The public nature of open source GIS has also resulted in the extensive availability of plug-ins that, given the modular nature of GIS, accounts for an important increase in open source GIS functionality with respect to that available in commercial software in which extensions are relatively minimal, usually more complex, and extremely expensive if the user requires several of them. Having access to many different modules is very important since they can save the user significant effort and time. While using proprietary software one is limited to the modules available within the software or those extensions one can buy, most extensions or plugins in open source software are small scripts directed to accomplish a simple function written by other users with the same needs. Software suits such as QGIS offer plugin managers where extensions can be easily queried, downloaded and installed without the need to restart the program. Open source add-ons have another important advantage: many of them are cross-platform. That is, they can be loaded on more than one open-source software, making life easier for those users who perform repetitive tasks in more than one program. This is important since, as any GIS experienced user would agree, it is very difficult to work with a single GIS package. The modular nature of GIS accounts for a large number of available modules with different performances and capabilities. It is therefore important to identify which modules are the most appropriate for the task at hand.

Many open source GIS programs can run in windows, Mac or Linux. This is a substantial difference with commercial software, such as ArcGIS, IDRISI or ERDAS IMAGINE (some of the most widespread geospatial analysis packages) that can only be installed on Microsoft Windows, forcing users of other operating systems to install windows in virtual environments (which can be a source of multiple problems) or in a partition of their hard drive. Many of them like QGIS or GVSIG provide versions that can run on tablets or, even mobiles, turning a portable device into a GPS-integrated mobile GIS, a truly useful feature for such common fieldwork as archaeological survey.

FOSS projects tend to join efforts thanks to the willingness of their developing teams, users and the fact that their modules' source code is publicly available. That is also a substantial difference with proprietary software where competence for market shares would, to say the least, restrict collaborations between competing enterprises. This tendency results in FOSS being developed more rapidly and incorporating more and more advanced functionalities than its commercial counterparts.

5.4 Problems with Open Source GIS

The growing use of open source GIS applications can be easily explained by the several advantages outlined above but it still remains a long way back from that of commercial software (Figure 5.1), particularly from ESRI's ArcGIS, which keeps a large portion of the market shares. This might be due to several reasons. Some of them can be labelled as 'psychological', such as the feeling most users have that paying for something is justified by receiving a higher quality product, and therefore, free software must be of lower quality. Commercial software have also a much longer history of use. Many users find it difficult to adapt to new software packages, while the new versions of their traditional software incorporate similar (if not the same) commands, terminology, graphic interface, etc., which softens the transition process. Related to this issue is the widespread use of proprietary software in the classrooms and GIS laboratories of universities and research centres where most GIS users are trained. Most of this teaching is still based on the repetition of practical processes directed at solving specific case-studies and not on the learning of the concepts, functions, applications and basic structure of GIS, which renders the adaption to new GIS software difficult.

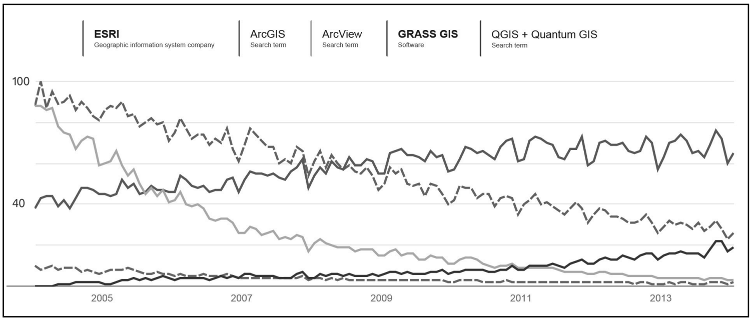


Figure 5.1: Google Trends search comparison between different GIS-related search terms

However, part of this still marginal use of open source GIS can also be related to some intrinsic problems of this type of software:

– While the strong focus of many GIS projects on the development of a community of users and developers is key to their success, it also brings some problems: Outdated scripts as a consequence of the lack of a central organisation arranging the distribution of new releases and updating the extensions, many of which are made by individual programmers which might not be willing to keep updating it for latter releases. However, the efforts made by OSGeo (the Open Source Geospatial Foundation) have made a significant step towards the interoperability and maintenance of geospatial open source code, significantly reducing the risk of outdated modules in the projects it manages.

- Abandoned open source GIS development projects are also common, mainly in archaeology where projects are linked to relatively short term funding, and thus, long-term maintenance, updating and support are rarely contemplated. However, a tendency towards embedding single projects within larger community-lead projects such as OSGeo can improve this situation (Ducke, 2012). To avoid this, the novice open source GIS user could direct his/her GIS interests towards large, well-developed projects, such as the ones described in the section on open source GIS programs.
- Long-term support and the availability or consistency of documentation is also very dependent on the vitality of the online community of users and/or developing organism (Steiniger and Bocher, 2009). As in the previous point this can be easily solved by employing well-established open source packages.

5.5 Common Misconceptions Regarding Commercial and Open-Source Software

Many of the misconceptions dealt with here might have been painfully true ten years ago. Many long-term open source users will still remember difficulties in installation, frequent crashing or bugs preventing certain scripts to run properly. These current misconceptions therefore include, as usually said about legends, an element of truth. However, updating and clearing old beliefs is necessary, particularly when dealing with fast developing computer software.

- 'Complex graphic user interfaces (GUIs) or the use of command lines makes difficult the use of open source GIS'. This might be true for some open-source packages, for example GRASS GIS, which although much improved from previous versions, still employs a GUI that can be considered strange for those used to ESRI's products. However, most established open source GIS suits available today offer similar environments to those available in ESRI's software and some of them were, in fact, designed to serve as an easy-to-used graphic interface for more complex packages.
- 'Lack of compatibility with proprietary formats'. Although some open source GIS, such as GRASS, still use their own native format to which imported files need to be transformed before being able to work with them, most open source packages accept proprietary or other common file formats. In fact they employ very sophisticated geospatial libraries, such as GDAL/OGR, that are able to import most file formats. It is also common to find plugins, that allow importing and working with unusual and specific file formats. All in all, the truth is that open source software has a much wider compatibility than commercial software.
- 'Commercial software is of higher quality, more secure and less-prone to crashing'. As with other of the listed misconceptions this might have been true some

time ago but not anymore. In fact, my own personal experience as a higher education academic who is forced to use both open source and proprietary software is that the former, particularly ArcGIS, has a tendency to crash more often and more destructively. In my experience, this seems to be a pretty common conviction among academics and archaeologists involved in both types of GIS distribution strategies.

– ‘Open source GIS is useful for computer geeks or programmers but not adequate for the non-technical GIS user: a lot of the open source GIS software expects the user to develop his/her own modules and tools’. Modern open source GIS is developed with the non-technical user in mind and one of its priorities is promoting the widespread use of open source tools. However, contrary to commercial software suits, open source promotes user-developed innovations and active participation in its development. It is curious that such a promotion might be considered in detriment to its usability since it is this openness to public design that renders open source GIS such a complete and advanced set of tools.
– ‘Open source software is geared towards the Linux community’. Similar to the previous point, this idea arises from the first open source program available, GRASS, which was developed in UNIX/Linux and until its version 6.4 in 2010 there was not a native Windows version available (although previous versions could be run on Windows using Cygwin). Still today the Windows native GRASS has reduced functionality compared to that distributed for Linux. Nowadays GRASS and other open source software packages, such as gvSIG and QGIS (which were multi-platform since their first releases), are distributed for all major operating systems including Mac OS. This is something than many non-Windows users are demanding from their proprietary GIS software developers/vendors for many years now.
– ‘Commercial software offers more help and support’. Related to this idea, is the conception that it is difficult to learn how to use open source GIS. For the most established open source suits, such as GRASS (Neteler & Mitasova 2008), QGIS (Thiede et al., 2013; Graser, 2013) and gvSIG (Arnalich and Ton-That, 2010), there are available manuals that can be bought at any major book seller or online bookstore and are as expensive as any handbook dedicated to commercial software. These also offer extensive documentation, help, online forums and, as commented before, an active online community always prone to help with doubts and problems. As in the case of commercial software there are plenty of enterprises that offer training and support services for open source software (Steiniger and Bocher, 2009). QGIS, for example, has listed in its website 27 international organisations that offer commercial training, support and programming for QGIS, GRASS and other open source software. Learning how to use open source GIS should not be expected to be easier or cheaper than learning how to use commercial software.
– ‘Proprietary software offer long-term support’. While, as mentioned previously, it is true that many open source GIS projects, particularly those specifically directed

to archaeological purposes (Ducke, 2012), never had continuity after their first development, the largest open source GIS projects are running for decades and they can be considered very effective in offering long-term support. It is, in fact, commercial software the one posing real risks to its long-term use. Many will still remember with yearning all those old scripts written in Avenue lost when migrating from ArcView to ArcGIS. Some others will still keep their old ArcView so they can use these scripts that incorporated functionalities that ArcGIS still does not have or requires an expensive extension to perform. Although the non-technical open source GIS user can also lose the scripts by updating their software, any user with programming skills can update a script or contract a company to do so thanks to the public availability of source code (Steiniger and Bocher, 2009). This would not be possible with close source software. More importantly, proprietary software licenses pose specific problems for long-term use. Steiniger and Bocher (2009, p. 1364) clearly illustrate this with several examples:

> "Other problems may occur due to vendor license changes or changes of license distribution rules. For example, the developers of Forestry GIS were hit by the first case; subsequently, as of November 2005 they were not allowed to freely distribute updates of their software because the license terms for the used proprietary GIS kernel were changed (ForestPal, 2008). The second case happened to Oxford Archaeology, an educational charity, which faced problems when ESRI Inc. was announcing to change their criteria for the awarding of educational licenses (see Cook 2008)."

– 'Commercial software is technically more advanced'. Just having a look to the list of modules and operations offered by GRASS or SAGA it becomes transparent that this does not hold true. Many of the newer open source projects offer a limited quantity of operations, but, the strong tendency that open source software has to join resources (and modules) in open repositories that are accessible to open source desktop GIS programs, promises great analytical possibilities even to the smallest project. Several authors (Camara et al., 2000; Steiniger and Bocher, 2009) have already commented the slow implementation rate of innovations in proprietary software, many of which have been available on open source GIS for years. Could it possibly be that this deceleration in innovation is due to a conscious effort to sell specific extensions or updates at a conveniently paced rate for market purposes?

Several of those misconceptions, such as complex GUIs and lack of compatibility with common file formats, seem to have their origins in older versions of GRASS as this has been the forefront and most visible representative of open source GIS software, but it is now time to re-evaluate them and acknowledge the huge progress made by FOSS during the last years.

5.6 Which Open-Source Desktop GIS is more Convenient?

Although many remote packages exist, this paper focuses on desktop applications that are software installed and run on personal computers, are not executed on a server and are not remotely accessed or controlled from or by a different computer. The following list is based on the author's experience, expertise and research interests and it is therefore not exhaustive, the reader is, therefore, directed to more specialised publications for comparison between different software choices (Sherman, 2008; Steiniger and Bocher, 2009; Hengl et al., 2009; Cattari and Clutterbuck, 2011; Steiniger and Hunter, 2013) and is encouraged to try and experience other packages to find the one that better adapt to his/her experience and needs.

5.6.1 GRASS (Geographic Resources Analysis Support System) GIS

This is the big name in open source GIS, with thirty years since its original development by the U.S. Army Corps of Engineers - Construction Engineering Research Laboratories (USA-CERL). GRASS was developed as a raster-analysis program and this is perhaps why its raster analysis capacities are its main analytical feature but its latest versions confidently support vector creation editing and analysis, database management systems, network analysis, LiDAR data processing, image processing and analysis (including multispectral data analysis), spatial modelling with capacity to deal with multi-temporal data, 3D raster (voxel) development, analysis and advanced visualization and graphics, and maps production. The last release of GRASS contains over 400 modules (Neteler et al., 2012) that can be increased with more than 200 user-developed extensions, some of them created specifically for archaeological applications that can be found in the OSGeo AddOns webpage.

GRASS offers an excellent option for advanced GIS users and it has commonly been applied to archaeological research. Some of its capabilities are unique in both commercial and open source desktop GIS. It is the only GIS desktop software that allows creating and analysing 3D surfaces or voxels, and NVIZ, its 3D visualization engine (Figure 5.2), allows single and multi-map 2D perspective support but also multi-dimensional visualisation, voxel visualisation as isosurfaces or cutting plane rendering, animations, full resolution exports, and 3D queries (Hofierka et al., 2002; Neteler et al., 2012). Apart from these fancy tools, GRASS can import almost all types of existing geospatial file types, which makes it extremely useful if only to load, see and export one's files in other formats. Although still experimental, two new features in the latest stable release (GRASS 6.4.3) provide important new functionalities. The new graphical modeller offers a tool for the implementation of complex or repetitive workflows, saving a substantial amount of time to the user. The graphic composer solves some of the issues with GRASS map composition and graphic export capacities.

GRASS can also be linked to other open source software, such as QGIS and the statisti-

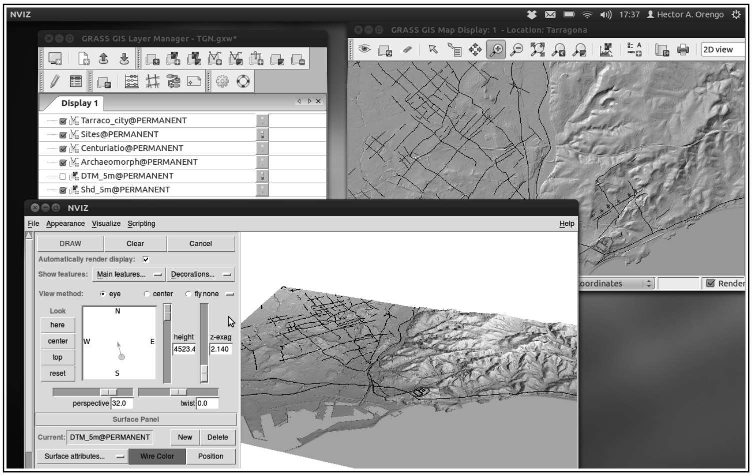


Figure 5.2: GRASS GIS running on Linux Ubuntu showing the Layer Manager, Map Display and the NVIZ module

cal package R, to increase and/or complement its capabilities. In archaeology GRASS has been employed in analysis of past landscapes (e.g. Barton, Ullah and Mitasova 2010; Barton, Ullah and Bergin 2010; Ullah 2011), predictive modelling of routes and archaeological sites (Indruszewski and Barton, 2006; Orengo and Miró, 2011), visibility analyses (Lake et al., 1998; Cooper, 2010), for the volumetric reconstruction of the excavation process (Lieberwirth, 2008; Orengo, 2013), and many other archaeology-related tasks.

GRASS GIS has, however, some problems that the unexperienced user should take into account when firstly approaching it. GRASS's 'floating window' graphical user interface (Figure 5.2), although much improved from earlier versions, can be a bit awkward to users accustomed to ESRI products. It is also difficult to start experimenting with GRASS functionalities without a basic knowledge of GIS and geographical concepts: for the program to start running you need to specify your project location, projection and coordinate reference system. Also before you can start using or even visualising your data you need to import it into GRASS's own native format. In this regard, its quality and coherence can be one of the problems hindering the widespread adoption of GRASS. Equally, the complexity of the analyses performed by some of the modules available in GRASS makes some of them difficult to apply without a good knowledge of the principles or theory on which these were based.

5.6.2 QGIS (Formerly Known as Quantum GIS)

QGIS started developing in 2002 as a response to the expert-oriented design of GRASS. It aimed to act as a data viewer and, at a later stage, provided a graphical user interface for GRASS, increasing in this way its analytical capacities. Since then, however, QGIS has become the popular open source desktop GIS choice (see the increase in QGIS-related web searches during the last years in figure 5.1). This might be due to its smooth and clear design(Figure 5.3), its ease of use and its restricted needs in terms of disc space, RAM and processing power. Its simple approach to the editing of vector features contrasts with most other commercial and open source suits. Similarly to many other open source packages it is cross-platform, running on Linux, Mac OSX, Windows and Android.

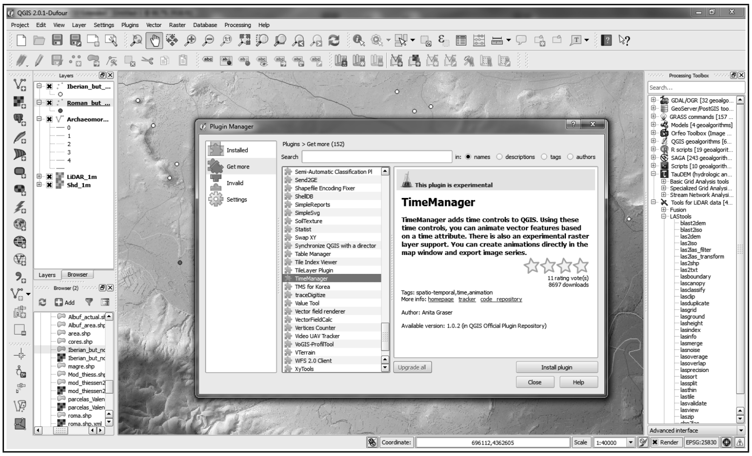


Figure 5.3: QGIS running natively on Windows 7 showing its attractive GUI with the Processing Toolbox (at the right side) and the Plugin Manager (in the centre)

The latest version of QGIS (Dufour 2.0.1) provides access through its toolbox (Figure 5.3) to modules not only developed for GRASS but also for Sextante (a series of raster tools initially developed for gvSIG), SAGA GIS, Orfeo Toolbox, R package for statistical computing, etc. However, not all modules from the original software are accessible through the QGIS toolbox. For instance, only 157 modules of GRASS modules and 243 from the almost 600 available in SAGA can be accessed. A great advantage of using the last QGIS version is that GRASS modules run from QGIS will not require the files to be transformed into the GRASS native format. QGIS plugin manager (Figure 5.3) provides a really simple way to query and install extensions without the need to exit the program. The 174 extensions available, some of them preloaded in the program, include all kinds of vector and raster analysis, including simple network capabilities and multispectral image analysis.

QGIS can load and edit ESRI's shapefiles and other common vector formats without the need to transform them. It can also load and work with the most common raster file types. Its many extensions provide compatibility with not so common geospatial file formats. Also, QGIS allows on-the-fly coordinate reference system transformation, which enormously simplifies the creation of projects joining data in different projections or coordinate systems, saving the need to transform the files for their integration.

Just as GRASS does, QGIS incorporates a graphical model builder to simplify and automate workflows and a print composer, which allows a smooth production of attractive maps.

Therefore, QGIS is a very complete desktop GIS with important capabilities for visualisation, analysis and cartographic production. Its ease of use and complete documentation, help files, manuals and on-line communities make QGIS the obvious choice for those starting to use open source GIS.

5.6.3 SAGA (System for Automated Geoscientific Analyses) GIS

SAGA GIS has been under development since 2001 and in 2004 its source code was publicly released. The latest release (2.1.0) runs under Windows and Linux and contains 586 modules distributed in 67 libraries, which renders it one of the most complete GIS packages in the market. It is a raster-oriented GIS but it has capacity to work with and edit vector data although, compared with the other packages described here, these are complex and hard to use. Its 3D render engine is efficient and fast but limited in the range of file types it can represent. SAGA is not a good choice for GIS starters or even archaeologists working with vector data. Its graphic export capabilities are very limited and the GUI (Figure 5.4) still has some bugs that limit its use. However, SAGA GIS is very conceptual in its development and users with a good knowledge of raster-based GIS concepts will feel comfortable with it. It was designed for a simple and effective implementation of spatial algorithms. Its set of tools is particularly efficient for Digital Terrain Models analysis, hydrologic analysis and environmental modelling and incorporates multiple tools for multispectral image analysis.

5.6.4 gvSIG (Generalitat Valenciana Sistema d'Informació Geogràfica)

This open access GIS project is unusual for being developed by a public administration. In a few years since it was publicly released it has become well known and a number of archaeologists employ it for their daily GIS work. It is developed in Java, which makes it flexible and easy to adapt to different operative systems. Its amiable graphic interface (Figure 5.5), ease of use (its vector editing tools are very simple and

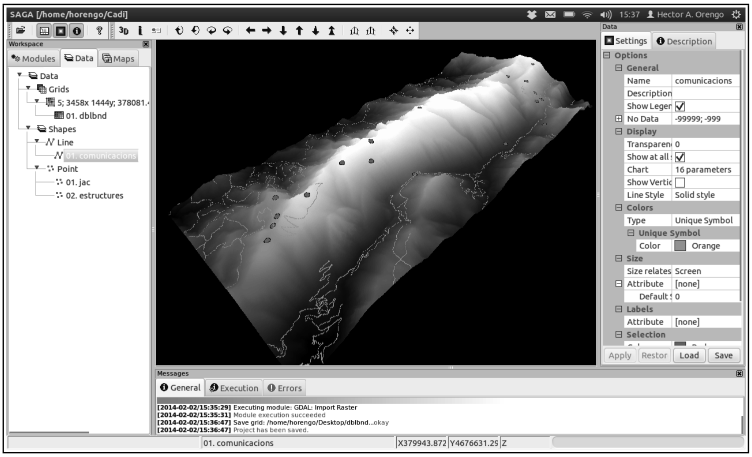


Figure 5.4: SAGA GIS running on Linux Ubuntu with the 3D window integrating raster and vector data

efficient) together with its capacity to work natively with the most common file types, justify its expanding use during the last few years. Several add-ons, such as Sextante, incorporate advanced analytical capabilities to work with raster data. Other useful extensions include 3D analysis, network analysis, remote sensing and LiDAR data analysis, although this last is still restricted in its capacities. It should be noted, however, that most of these extensions are still not available for the latest stable version (2.0) and some of these may not ever be. GvSIG add-ons Manager (Figure 5.5) provides a simple and useful way to manage and get new plugins.

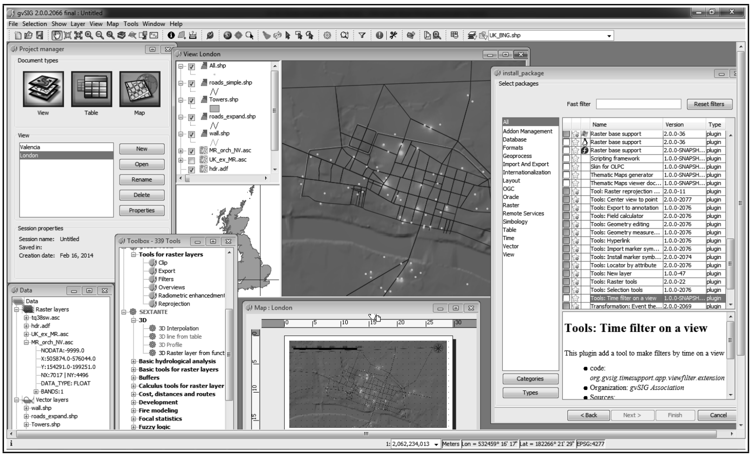


Figure 5.5: gvSIG running on Windows 7 displaying the Project Manager (top left), the View Window (top centre), the Map Window (bottom centre), and the Add-ons Manager (right)

This, therefore, is an excellent software package for both starters and advanced users. Its current desktop stable version is very secure and the existence of a mobile option makes it particularly useful for those working in the field. gvSIG is currently in the OSGeo's "incubation" process, that is, is being evaluated to be one of OSGeo's projects. Being part of OSGeo will ensure or at least facilitate the long-term maintenance and the growth of resources available for the communities of users and developers of gvSIG.

It is interesting to note, in regard to the applicability of gvSIG to archaeology-related tasks, that Oxford Archaeology has developed a version of this software: gvSIG OA Digital Edition. gvSIG OADE is fully portable, has a simpler menu and installation process and incorporates GRASS and Sextante modules. However, gvSIG OADE is based on gvSIG v.1.10 and does not include the latest updates and improvements present in gvSIG 2.0.

5.6.5 Other Open Source Geospatial Software

GIS is rarely used in isolation and, consequently, it is worth briefly reviewing other geospatial applications that can be used in conjunction with GIS to solve all the archaeologist's geospatial needs. Some FOSS packages, although providing some GIS capabilities, are oriented towards specific types of analysis. Programmes such as the public domain HEC-RAS and the open source Calypso, provide advance hydrologic modelling, which include flood analysis. Some other software packages that can also be used in combination with GIS are more directed towards Remote Sensing, such as ILWIS, the BEAM Toolbox (with VISAT GUI), Orfeo Toolbox and Opticks (which is currently being considered as an OSGeo project).

Other remote sensing-related group of programs are those directed to the treatment of LiDAR data. Although some of the open source GIS software explored before offer the possibility of importing, treating and transforming LiDAR data, only a few packages, such as rapidlasso's LAStools (note that not all tools are open source but all are free for non-commercial use) and MeshLab, offer the possibility to deal with LiDAR-generated point clouds without the need to import them into specific GIS software.

Another category of software able to generate georeferenced point cloud data is the photogrammetric 3D reconstruction programs. During the last few years several open source projects have been aimed at the development of software capable of automated image processing for 3D reconstruction. Most of these, being relatively recent, are still difficult to use with few manuals or online help and, when present, unattractive GUIs. However, some of these provide excellent photogrammetric capabilities, able to match their most common commercial counterparts (Remondino et al., 2012). Perhaps the best known among these is Bundler (Snavely et al., 2008), which is now somewhat outdated and complex to use for those starting with photogrammetric 3D modelling. VisualSfM, free for personal, non-profit, or academic use, was developed by Changchang Wu and offers a similar solution to that of Bundler but with a much easier to follow workflow and a simple GUI. Finally, Apero and MicMac, a set of programs developed by the Matis laboratory of the French IGN, provide advanced and accurate capabilities for the production of photogrammetric 3D models. An advantage of MicMac over other packages is that it can be used to produce georeferenced orthoimages, without the need to employ a GIS program. MicMac is distributed under the French CeCILL-B license (fully compatible with BSD-like licenses).

Both photogrammetric and LiDAR-derived point clouds can be loaded, treated and exported with open source programs, such as Meshlab or ParaView, which, with the PCL plugin, allow users to access algorithms from the Point Cloud Processing Library (PCL). ParaView also offers other 3D advanced visualisation capabilities and is able to load and explore different types of volumetric data.

5.7 Open Geospatial Data?

After having reviewed some of the different open source GIS options available to the archaeologist it is necessary to consider the availability of geospatial data needed to perform GIS analysis. This is a critical issue since without data it is not possible to perform any GIS analysis (Cattari and Clutterbuck, 2011, p. 24). Although most of the previously described packages offer datasets for training purposes, these are very restricted in spatial terms and will not provide data on a project's specific area of research. Of course the archaeologist is expected to be able to generate his/her own data from field survey, excavation, site and heritage records, document and map analysis, and so on. Archaeological geospatial data, however, are difficult to come by since, as Jessop has pointed out (2008, p. 44), 'the capture of spatial data is a major task in many projects and thus the scale of effort and cost of creating digital datasets make some groups reluctant to share them without charge'. Institutions such as English Heritage and initiatives such as the Portable Antiquities Scheme offer free access to their spatially enabled databases after subscription. However, for those working on a regional scale or in landscape projects, data such as digital terrain models, geological maps or aerial imagery are necessary and few projects' budget can afford the resources necessary to generate them.

International efforts like the INSPIRE Directive, which aims to produce an EU spatial data infrastructure to facilitate public access and share environmental spatial information across Europe, are incomplete and institutions like the Ordnance Survey in England or the Institut Géographique National in France still charge very expensive prices for digital geographic data. Some other geographic data providers, such as Edina Digimap and Mimas Landmap for the UK's higher education and academic research community, charge expensive subscription rates than can only be afforded by big enterprises.

Fortunately, following the philosophy developed in the European INSPIRE Directive, more countries, such as Spain, provide vast amounts of reliable geographical information at no cost following a registration process. Although the availability of open access geographical data is highly variable depending on the country some international bodies such as the European Environment Agency, which provides environmental data for Europe, deliver quality open access data in common GIS formats with good associated metadata, that is, information on how the information was gathered and the data treated and produced.

The US NASA and USGS provide worldwide satellite imagery and elevation datasets, in many cases free of charge upon registration through USGS EarthExplorer, USGS GloVis, Global Data Explorer. The European Space Agency also provides access to satellite imagery following registration and, depending on the data sought, a lengthy process involving the setting up of a project. Global elevation datasets can also be found at CGIAR-CSI GeoPortal (SRTM data at 90m/cell resolution) and at the ASTER GDEM site (at 30m/cell resolution).

There are also repositories of free digital geospatial data, often contributed by a community that can be downloaded freely from the Internet. Nonetheless, many of these datasets do not usually incorporate metadata that would allow the user to have a good idea of the processes that were followed to create, transform and distribute them. This is an essential requirement for geospatial information and the use of these data may be risky when highly accurate information is required.

5.8 Conclusions

The advantages of employing open source GIS for archaeological research and practice are evident and have been thoroughly described here. Open source GIS software have undergone dramatic changes during the last few years and are now complete and secure packages that can successfully address all of the archaeologist's needs.

The coming years will probably see an increase in the capacities, efficiency and ease of use of these programmes, with more resources being available for training and specialised use. This will be, in part, a consequence of the development of foundations and associations, such as OSGeo, which support the collaborative development, maintenance and collaboration of open source geospatial software. This is very significant since the nurturing of cross-project collaboration and integration of open source resources generated by multiple projects will, no doubt, produce the most advanced and secure geospatial software available. This tendency toward the integration of open source projects is already a reality in GIS suits like QGIS through extensions and plugins. An increase in the integration of capabilities in open source GIS it is thus expected to materialise in the next few years that will, hopefully, result in complete geospatial information systems.

The innovation generated by joined open source initiatives may also force commercial software to increase the capacity and quality of their products and hopefully reduce their prices.

Bibliography

Arnalich, S. and Ton-That, T. (2010), *gvSIG y Cooperación: CÃşmo construir e incorporar un Sistema de InformaciÃşn GeogrÃąfica a tu proyecto*, Arnalich.

Barton, C. M., Ullah, I. I. and Bergin, S. (2010), 'Land use, water and mediterranean landscapes: modelling long-term dynamics of complex socio-ecological systems', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **368**(1931), 5275–5297.

Barton, C. M., Ullah, I. and Mitasova, H. (2010), 'Computational modeling and neolithic socioecological dynamics: a case study from southwest asia', *American Antiquity* **75**(2), 364–386.

Brandtzæg, P. B., Følstad, A., Obrist, M., Geerts, D. and Berg, R. (2010), Innovation in online communities–towards community-centric design, *in* 'User centric media', Springer, pp. 50–57.

Camara, G., Souza, R., Pedrosa, B., Vinhas, L., Monteiro, A. M. V., Paiva, J., Carvalho, M. d. and Gattass, M. (2000), Terralib: technology in support of gis innovation, *in* 'II Workshop Brasileiro de Geoinformática, GeoInfo2000', Vol. 2, pp. 1–8.

Cattari, G. and Clutterbuck, R. (2011), 'Free open-source geographical information systems for irish archaeology', *Archaeology Ireland* pp. 21–25.

Chambers, J. (2008), *Software for data analysis: programming with R*, Springer.

Cook, J. (2008), 'Dear ESRI, it's not me, it's you', Open Source Computing and GIS in the UK.
URL: http://www.archaeogeek.com/blog/2008/03/12/dear-esri-its-not-me-its-you/

Cooper, J. (2010), 'Modelling mobility and exchange in pre-columbian Cuba: GIS led approaches to identifying pathways and reconstructing journeys from the archaeological record', *Journal of Caribbean Archaeology* pp. 122–137.

Ducke, B. (2012), 'Natives of a connected world: free and open source software in archaeology', *World Archaeology* **44**(4), 571–579.

ForestPal (2008), 'Forestry GIS'.
URL: http://www.forestpal.com/fgis.html

Graser, A. (2013), *Learning QGIS 2.0*, Packt Publishing Ltd.

Hafer, L. and Kirkpatrick, A. E. (2009), 'Assessing open source software as a scholarly contribution', *Communications of the ACM* **52**(12), 126–129.

Hengl, T., Grohmann, C., Bivand, R., Conrad, O. and Lobo, A. (2009), 'SAGA vs GRASS: a comparative analysis of the two open source desktop GIS for the automated analysis of elevation data', *Proceedings of Geomorphometry* .

Hofierka, J., Parajka, J., Mitasova, H. and Mitas, L. (2002), 'Multivariate interpolation of precipitation using regularized spline with tension', *Transactions in GIS* **6**(2), 135–150.

Ince, D. C., Hatton, L. and Graham-Cumming, J. (2012), 'The case for open computer programs', *Nature* **482**(7386), 485–488.

Indruszewski, G. and Barton, C. (2006), Simulating sea surfaces for modeling viking age seafaring in the baltic sea, *in* 'Digital discovery: exploring new frontiers in human heritage. CAA 2006. Computer applications and quantitative methods in archaeology. Proceedings of the 34th conference, Fargo, United States, April 2006', pp. 616–30.

Jessop, M. (2008), 'The inhibition of geographical information in digital humanities scholarship', *Literary and Linguistic Computing* **23**(1), 39–50.

Lake, M., Woodman, P. and Mithen, S. (1998), 'Tailoring GIS software for archaeological applications: an example concerning viewshed analysis', *Journal of Archaeological Science* **25**(1), 27–38.

Lieberwirth, U. (2008), 'Voxel-based 3D GIS: Modelling and Analysis of Archaeological Stratigraphy', *Beyond Illustration: 2D and 3D Digital Technologies as Tools for Discovery in Archaeology. Oxford: Archaeopress* pp. 85–94.

Morin, A., Urban, J., Adams, P., Foster, I., Sali, A., Baker, D. and Sliz, P. (2012), 'Shining light into black boxes', *Science* **336**(6078), 159–160.

Neteler, M., Bowman, M. H., Landa, M. and Metz, M. (2012), 'GRASS GIS: A multi-purpose open source GIS', *Environmental Modelling & Software* **31**, 124–130.

Orengo, H. A. (2013), 'Combining terrestrial stereophotogrammetry, DGPS and GIS-based 3D voxel modelling in the volumetric recording of archaeological features', *ISPRS Journal of Photogrammetry and Remote Sensing* **76**, 49–55.

Orengo, H. A. and Miró, C. (2011), Following Roman waterways from a computer screen: GIS-based approaches to the analysis of Barcino's aqueducts, *in* P. Verhagen, A. Posluschny and A. Danielisová, eds, 'Go Your Own Least Cost Path: Spatial Technology and Archaeological Interpretation: Proceedings of the GIS Session at EAA 2009, Riva Del Garda', British Archaeological Reports, International Series. Oxford, Archaeopress, pp. 47–53.

Remondino, F., Del Pizzo, S., Kersten, T. P. and Troisi, S. (2012), Low-cost and open-source solutions for automated image orientation–a critical overview, *in* 'Progress in Cultural Heritage Preservation', Springer, pp. 40–54.

Serlorenzi, M., ed. (2013), *ARCHEOFOSS Free, Libre and Open Source Software e Open Format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11-13 giugno 2012)*.

Sherman, G. (2008), *Desktop GIS: Mapping the Planet with Open Source Tools*, Pragmatic Bookshelf.

Snavely, N., Seitz, S. M. and Szeliski, R. (2008), 'Modeling the world from internet photo collections', *International Journal of Computer Vision* **80**(2), 189–210.

Steiniger, S. and Bocher, E. (2009), 'An overview on current free and open source desktop GIS developments', *International Journal of Geographical Information Science* **23**(10), 1345–1370.

Steiniger, S. and Hunter, A. J. (2013), 'The 2012 free and open source GIS software map–a guide to facilitate research, development, and adoption', *Computers, Environment and Urban Systems* **39**, 136–150.

Thiede, R., Sutton, T., Düster, H. and Sutton, M. (2013), *The Quantum GIS Training Manual: A Comprehensive Introduction to Quantum GIS*, Locate Press.

Ullah, I. I. (2011), 'A GIS method for assessing the zone of human-environmental impact around archaeological sites: a test case from the Late Neolithic of Wadi Ziqlâb, Jordan', *Journal of Archaeological Science* **38**(3), 623–632.

Wheatley, D. and Gillings, M. (2002), *Spatial Technology and Archaeology*, London, Taylor & Francis.

Lolita Nikolova

6 What was Published is as Important as How it was Published

6.1 Introduction

Archaeology is a complex discipline which studies the past based on material evidence. There are three main fields of research archaeology: prehistoric, historic and contemporary. Beyond research archaeology, which includes field excavations, systematisation and interpretation of data, there is also theoretical archaeology; other classifications may also include academic archaeology and public archaeology. All fields and subdisciplines cross over each other, and it is therefore, useful to think about archaeology as a complex discipline of the material past (Nikolova, 2013).

It is essential to understand that archaeology is a sensitive social discipline since people have different approaches to the past, and the social, political and cultural context of archaeology in different countries is vastly diverse. Archaeologists have social twins - treasure-hunters - while government museums compete with private museums and collectors. These two realities, although in many cases analysed separately, complicate the function of archaeology as a subculture in the contemporary global world.

The global value of open access (OA), which is recognised by support at the highest levels of government (Lenzer, 2008), turns the goals of the author of this chapter toward searching for perspectives of increasing of the role of open access in archaeology as a discipline which produces and depends on enormous amounts of reports and publications.

6.2 Open Access and Global Society

OA in 21st century science refers to providing unrestricted online access to scholarly material, such as journal articles, theses, monographs, book chapters and other online materials. OA innovation in the scholarly world is a consequence of the development of the internet and globalisation (Suber, 2012). As a theoretical category, OA speeds the advance of the science through the dissemination of information about the newest discoveries. Another argument to widening the access of scholarly publications, is that the taxpaying public deserves access to the outputs of the research they

Lolita Nikolova: International Institute of Anthropology, Utah, USA

fund (Neylon, 2013, p. 1). Re-use – beyond the initial reading and financial platform – may include re-analysis and re-distribution of archaeological reports and data (see also Beck and Neylon 2012).

In the early 21st century, OA databases have had the same revolutionary role in human society as, for instance, the Bauhaus Period (1919–1923) advances of architecture and design. The internet itself is a sort of tool that can be used in a variety of ways. The internet includes elements of high culture, communication innovation, spam, danger, crime, etc. How science uses the internet is crucial for the future of culture, since science is the father of culture, in the way that art is its mother.

OA is misleading from the perspectives of the publishers of scientific papers, since OA, in most cases, means that the publishers should be paid by the institutions or by the authors. Even if an author would like to publish in a popular OA journal, it may be impossible because of publishing fees. OA can be seen as a coin with different values – while authors volunteer or pay, publishers have benefit from the profit of the publications. McCabe et al. (2013) created an economic model - based on the platform market theory - to attempt to clarify in which cases the publishers would prefer open access.

Beyond the traditional classification ("green" and "gold" open access, and gratis and libre) (Open access, online), we can offer the following classification:

1. Completely OA journals
2. Journals with OA selected articles of whole issues.
3. Websites with links or .pdf files of OA publications.
4. Books with "look inside" (books.google.com, amazon.com).
5. Academic profiles at academia.edu
6. OA librarian metadata
7. OA Museum database
8. OA virtual museums
9. OA data base of excavations
10. Reports of excavations published online (e.g. Catahüyük's reports of the team directed by Ian Hodder).
11. Abstracts published online.
12. Complimented materials to hard-copies published online as OA.
13. Blogs with texts which have values of scientific publications because of originality.
14. Open access videos (e.g. youtube.com, Vimeo.com), etc.
15. Open access through academic hosts (for students and instructors/professors) (e.g. EBSCO e-books and EBSCO articles), etc.

Key meetings and initiatives which stimulated the OA publications include the Budapest Open Access Initiative (BOAI) in 2002 and Global Open Access weeks. The European Union plans for up to 100% of funded Horizon Research Programs to be OA by 2020 (Macilwain, 2013, p. 7). PubMed Central, which is operated by the National Center for Biotechnology Information (NCBI), a division of the National Library of Medicine (NLM) at the U.S. National Institutes of Health (NIH) has a very active positive role within the USA (Kurata et al., 2013). Leading universities (Harvard, Massachusetts, Stanford, Kansas) have passed an OA policy granting a license to share their scholarly journal articles openly (Emmett et al., 2011). According to some authors, by 2010 about 20% of the scientific research was available as OA (Björk et al., 2010). However, for some areas of the field – such as Africa Open Access – there is limited support by local governments and OA has a fragmented status (Nwagwu, 2013). Personally, the author has been experiencing the positivity of Open Publishing (OP) through the popularity at Academia.edu of her first OA published article (Nwagwu, 2013). In the context of global interest in OA, archaeologists have been expanding their contributions by increasing the number of open publications (OA archaeology online), discussing the theory of the subject (e.g. Lake 2012), and creating OA communities (Open Access Archaeology on Facebook).

The motifs outlined below in this chapter include an artistic reflection on OA in archaeology and open science.

"O like Open" (Fig 6.1a) expresses the limitation of open - it requires direction and usually the success depends on narrowing of paradigms. There is an interconnection between openness in science and ethics. Ethics is a framework and its application means limitation. The difference between open and moral is that open is the right to freely communicate scholarly mind, while moral is mostly about what one should not do. People who oppose OA from moral perspectives (e.g. authorship) in fact belong to those segments of society which always look at human culture as a sum of often not interacting and not related subculture. The supporters of open success look at society as a whole system and believe that this system has humanistic parameters which need to be nourished intellectually.

"A like Access" (Fig 6.1b) represent the non-linear reality of the current OA in science - its dependence on individuals, institutions and policies. This context may change and the visual language would also change.

"A like Archaeology" (Figure 6.1c) includes as abstract expressions some of the hallmarks of archaeology - layering, popularity of circle structures (pits, houses, fortresses, etc.), the black layers of burnt villages; the chess as one of popular ornamental motifs, as well as a symbol of the world, etc. This communicates the complex character of archaeology that may also mean a hierarchy of values, which is still not well revealed in the scholarly and popular literature. The truth is that many intelligent minds depart from archaeology since neither the governments nor the related subcultures can resolve the actual question of treasure-hunters and collectors. Archaeologists are placed in a difficult situation, since by increasing the values of material culture, they stimulate an increasing of treasure-hunting. The most curious case studies come from Bulgaria. Beyond Vasil Bozhkov and Dimitur Ivanov, in the last decades, and in the context of excavations of Thracian tombs work by Georgi Kitov, some treasure-hunters have become among the richest Bulgarians. There are also archaeologists who have developed the psyche of treasure-hunters instead of being attached to the paradigm of preservation of cultural heritage. Accordingly, what archaeologists publish as OA is as important as how they look at culture, society and the material past. The way they accept the material past may have enormous impact on human culture as repeating circles.

“S like Science” (Fig 6.1d) is based on the solid view on science as structure. Science connects the world as a whole, and the role of the scientists is to make new discoveries and new structures to actualise the wholeness of the complex world (nature and humans).

(a) O like Open

(b) A like Access

(c) A like Archaeology

(d) S like Science

Figure 6.1: Four motifs demonstrate that the problem which combines open, access, archaeology and science

The four motifs demonstrate that the problem which combines open, access, archaeology and science deals with different quality values, even the reflection towards possible conflicts because of this difference in the quality values.

6.3 Open Access, Archaeology and Ethics

OA archaeology has developed not only as a social practice, but also as a sub-branch of theoretical archaeology (Carver 2007; Lake 2012).

From a theoretical point of view, one of the key questions is: has archaeology become more influential in science and in society with development of OA and globalisation? Open Archaeology Access online has listed more than 200 journals which provide archaeological information from different parts of the world. This access theoretically should be able to change archaeological methodology and make archaeology much more influential in the field of science, especially in evolutionary research. The term evolution is not very clear, since the human culture knows both progress and devolution. Evolution may mean new quality changes in human society, which are usually non-invertible, but such changes may have a positive or negative value, then, change can be a progress or regress. To research evolution then means to prove a progression. Accordingly, in many cases evolution has not been used in the right way, since the so-called evolution could be a change, which has a negative impact on society.

Openness is expressed as open availability, or publicly available. It refers to:

1. specifications of a file format that are publicly available and accessible
2. the autonomy of a file format, which relies on several factors (self-contained document, structure, format, metadata information; (Park and Oh 2012, p. 48, see also Lake 2012; Beck and Neylon 2012).

The Directory of Open Access Journals (DOAJ) currently lists nearly 10,000 fully OA journals (*Directory for Open Access Journals*, n.d.).

A newly published OA article on “evolution of human family” (Smaldino et al., 2013) is a very good example how OA itself does not stimulate innovative research. While OA provides easy accessibility to the most recent literature, some scholars continue to believe that replicating methods based on very limited theoretical and methodological frameworks can be considered science and even be published by paid OA journals; that means additional expenses possibly applied to the institutions where the authors work. In many research disciplines scholars are locked in their own small debates, while the science of human civilisation needs a revolution in the understanding of science as a human interest, not as a field of human occupation alone. Smaldino et al. (2013) have occupied themselves in criticising the models of some of their colleagues, although their methodology of making models based on small samples looks itself , could be questioned.

There are many examples in archaeology where the design of publications is a reproduction of local methodology which completely ignores OA publications that may change that methodology. Such publications keep archaeology at a lower level which is a crisis of context, since the traditionalists are very active - as in this case where the authors reportedly worked in an institute. In contrast, OA may stimulate high standards of reporting better serving the needs of 21st century science (see e.g. Sun et al. 2013).

OA itself does not mean democratisation of archaeology – it may create a more rigorous environment of censure, limiting access for people who have the power in the social practices and who in fact prevent the progress in archaeology.

As Olivieri recalls, "…most intellectual historians date the birth of modern science to the foundation of the Royal Society in Britain" (2006, p. 176). However, ethics and knowledge have become a problem since the earliest stage of human civilisation. The distribution of innovations, ceramic styles and exchange of objects were among the earliest contexts of the building of rules and the creation of norms of communication between individuals and communities.

One of the key issues of ethics is the fact that ethics as a cultural category has many ambiguous components, and is based often on fragmented behaviour. For instance, an author in science in fact can be a psychotronic terrorist who has been using science to masking antisocial behaviour. In the same way OA journals may in fact represent an attempt of a group of scholars to use science for non-ethical social practices (censure in publication, activity which in fact replace most actual for society critics, etc.). Global society is an integrated community. It is multicultural society with mixed view on traditions, science and ethics.

On the whole, the personalities of people have changed since they have been armed with the internet and after the end of Cold War (1989), the value of people today is first of all their personality. Negative (e.g. authoritarian or visible or invisible violent) personalities devalue their own cultural products. The Global Age (the period after 1989) is an Age of moral revolution. The existence of human civilisation depends on the results of this moral rebuilding of global society. OA is a powerful positive step in making society accessible, transparent and communicating based on valid actual data. However, OA has had a diverse response in archaeology, with doubt thrown upon the ethics of certain researchers (e.g.Bartman 2012; Kansa 2012). Copyright and peer-review is an issue of ethical debate. According to Harnad (2007), peer-to-peer access is far more important than direct public access and OA needs to be mandated for all research, by researchers' institutions and funders.

6.4 Conclusions

The following conclusions summarize the main points in this chapter:

1. Archaeology as a discipline is still at a stage of emerging OA discipline, since the

open success publications are insufficient in compare of the traditionally published archaeological works.
2. OA theoretically would provide a stage of revolution in archaeology since having easy accessible publications online would stimulate research, as well as information from archaeology which may have high positive values of advance of culture and the relation of people to human culture. However, as in the motifs outlined above, there is a communication of different quality values which complicate the problem of faster integration of archaeology in the global OA community.
3. Development of OA in archaeology may stimulate the advance of the other sciences with direct or indirect links to archaeology. However, the research in the different archaeological fields as branches of complex social science in many cases is traditional, and archaeological publications often do not meet the criteria to be useful for research from other scientific branches.
4. Development of OA may increase the culture of archaeologists and invite in archaeology intellectuals with deep interest in culture.
5. Development of OA in archaeology may increase the numbers of authors who do not work in the field of archaeology but use archaeology for their research.
6. Advance in the OA of science of archaeology would stimulate increasing of value of archaeological publications and the value of archaeologists as people who make and not destroy and consume culture.
7. OA may increase the competition and corruption in archaeology since some may use OA journals to drawing a specific picture of national archaeologies serving political and ideological goals (especially in small countries like Bulgaria).

On the whole, archaeology is a field of social human activities, in which different people work, and OA is an opportunity to change of the types of personalities who chose archaeology as their field of human interest. OA increases the opportunity for critical approach, then, as authors will rely mostly on professionalism and not on power and positions. Many professors who use their power in a corruptive way will lose the main weapon of reproduction in archaeology – choosing specific people with multipersonalities with the only goal to reproduce their own heritage non-critically.

OA if it develops as a positive social practice, will probably make a revolution in ethics in archaeology and will make archaeology a stage of dominated positive values.

6.5 The Future

Some authors are afraid to predict the future of digital publishing (Arms, 2000, p. 263). However, others like Neylon state that the "...difference between the utopian and dystopian futures ...is public engagement in science. My suspicion is that if we can't bring interested members of the public into the process of research then we won't be looking at a happy future in terms of funding." (Neylon, 2013, p. 4).

It is very easy to make analogies between genealogy and archaeology as disciplines. Both are integrated with society. However, while genealogy has been blooming on the internet, archaeology is one of the least represented disciplines although a considerable part relates to visual culture. Involving those interested with archaeological education in archaeological research - from the field to the theory - is not a utopia; it is a very beneficial and possible to realize this process. Such integration will increase the quality of personalities of archaeologists and will make sense of OA not as a field of academic competition, but a field of complex research of archaeology as a segment of complex science.

Expanding of the meaning of the OA archaeology together with enriching of the existed branches of OA (journals, e-books, metadata, "green" files, etc.) is a real step in making OA the dominant mean of archaeological publications and a way to integrate archaeology as a real essential segment of the complex science.

Bibliography

Arms, W. Y. (2000), *Digital libraries*, MIT press.

Bartman, E. (2012), 'From the president: Open access', *Archaeology* **65**(3).

Beck, A. and Neylon, C. (2012), 'A vision for open archaeology', *World Archaeology* **44**(4), 479–497.

Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T. and Gunason, G. (2010), 'Open access to the scientific journal literature: situation 2009', *PloS one* **5**(6), e11273.

Carver, M. (2007), 'Archaeology journals, academics and open access', *European Journal of Archaeology* **10**(2-3), 135–148.

Directory for Open Access Journals (n.d.).
URL: http://doaj.org/

Emmett, A., Stratton, J., Peterson, A. T., Church-Duran, J. and Haricombe, L. J. (2011), 'Toward open access: It takes a "village"', *Journal of Library Administration* **51**(5-6), 557–579.

Harnad, S. (2007), 'Ethics of open access to biomedical research: just a special case of ethics of open access to research', *Philosophy, Ethics, and Humanities in Medicine* **2**(1), 31.

Kansa, E. (2012), 'Openness and archaeology's information ecosystem', *World Archaeology* **44**(4), 498–520.

Kurata, K., Morioka, T., Yokoi, K. and Matsubayashi, M. (2013), 'Remarkable growth of open access in the biomedical field: analysis of PubMed articles from 2006 to 2010', *PloS one* **8**(5), e60925.

Lake, M., ed. (2012), *Open archaeology*, Vol. 44, Taylor & Francis.

Lenzer, J. (2008), 'Us congress and european research council insist on open access to research results', *BMJ* **336**(7637), 176–177.

Macilwain, C. (2013), 'Is open access finally on the ascendancy? this may be the year for reality to catch up with rhetoric', *BioScience* **63**(1), 7–11.

McCabe, M. J., Snyder, C. M. and Fagin, A. (2013), 'Open access versus traditional journal pricing: Using a simple "platform market" model to understand which will win (and which should)', *The Journal of Academic Librarianship* **39**(1), 11–19.

Neylon, C. (2013), 'Architecting the future of research communication: building the models and analytics for an open access future', *PLoS biology* **11**(10), 1–5.

Nikolova, L. (2013), 'Open access archaeology and open science', examiner.com.
URL: http://www.examiner.com/article/open-access-archaeology-and-open-science

Nwagwu, W. E. (2013), 'Open access initiatives in africa–structure, incentives and disincentives', *The Journal of Academic Librarianship* **39**(1), 3–10.

Olivieri, R. (2006), 'Business, science, and the common good', *Journal of library administration* **44**(3-4), 173–185.

Park, E. G. and Oh, S. (2012), 'Examining attributes of open standard file formats for long-term preservation and open access', *Information Technology and Libraries* **31**(4), 46–67.

Smaldino, P. E., Newson, L., Schank, J. C. and Richerson, P. J. (2013), 'Simulating the evolution of the human family: cooperative breeding increases in harsh environments', *PloS one* **8**(11), 1–12.

Suber, P. (2012), *Open Access*, MIT Press.
URL: http://cyber.law.harvard.edu/hoap/Open_Access_%28the_book%29

Sun, G. H., Aliu, O. and Hayward, R. A. (2013), 'Open-access electronic case report journals: the rationale for case report guidelines.', *Journal of clinical epidemiology* **66**(10), 1065–1070.

Benjamin Ducke

7 Free and Open Source Software in Commercial and Academic Archaeology: Sustainable Investments and Reproducible Research

7.1 Introduction

While computing and software-based research pervade every aspect of modern archaeology, the pool of software specifically created for archaeological applications remains small. This is no small paradox, given that we live in a world of digital plenty; a world where software for just about any purpose has been available gratis, in the form of "Free and Open Source Software" (F/OSS), for decades. In key areas such as GIS, database management and statistical computing, there are free alternatives that are every bit as capable as even the most expensive paid-for offerings (see the expositions of F/OSS GIS by Neteler and Mitasova 2008; Sherman 2008; Rey 2009). And if some open source software doesn't quite do the trick, then it can easily be modified and enhanced, thanks to the full disclosure of its source code. After all, the effort required to learn a programming language and start creating customised software is minute, compared to what archaeologists commonly invest in learning spoken languages, Latin, Old Greek, or even basic statistics and calculus. In the Age of the Internet, all that it takes is to go online. The tools, the knowledge and the support are all there, freely available to everyone around the globe.

Why then, do university departments, research projects and individuals still pay the considerable license fees asked for by vendors of proprietary, closed-source software? Why do they not take that same money and invest it into the digital assets of their own discipline, into staff training, transferable programming skills and tailor-made software that's free to share with colleagues? And why, after all that is known and published about the prevalence of errors in complex programs (e.g. Hatton 1997; Merali 2010), do they still entrust their valuable research data to software that allows no access to its source code, that cannot be peer-reviewed and thus effectively acts as a black box in data processing (Morin et al., 2012)? There are many factors that must be taken into consideration when trying to find answers to these questions. But a lack of "quality" or other presumed shortcomings of F/OSS itself can be safely excluded. Rather, what seems to be holding back wide-spread endorsement of F/OSS in archaeology is a mixture of die-hard habits, as well as a lack of awareness of the intricate links

Benjamin Ducke: Freien Universität, Berlin, Germany

between economic and academic aspects of software, and between technological and social capital in the open source community.

The present chapter addresses these issues and discusses pathways to successful investments in archaeological F/OSS; although due to space restrictions, this will have to be done in a brief and somewhat superficial manner (see Fogel 2009 for an excellent, much more detailed account of how to run a successful F/OSS project). It will open with a discussion of the economic aspects of F/OSS, before it will go on to discuss social and academic aspects. This order reflects the author's personal experience, that sustainable research, software-based or otherwise, requires a sound funding model. At the same time, good academic software practice cannot be sustained without a strong focus on transparency and reproducibility of research. Given the complexity of modern software applications, the ability to track data through complex processing chains is key to successful collaborative research, as is the ability to peer-review the software (e.g. Barnes 2010; Morin et al. 2012). Therefore, this chapter will discuss in some detail how the academic quality and usefulness of data-based research can benefit from the use of F/OSS.

Verifiable information on the successes or failures of F/OSS in different fields of application remains difficult to obtain. The detailed report by Wheeler (2005) provides some numbers on the key issues of market share, reliability, performance, scalability, security, and total cost of ownership. In addition, this chapter cites a number of other sources that do exist in the form of peer-reviewed, printed publications in English, originating from diverse academic disciplines. Three real-world case studies serve to supply some empirical evidence for the validity of the statements and the usefulness of the recommendations made in this chapter. The author has been directly involved in all three projects selected as case studies. Although this adds substantial prejudice to the discussion, it is hoped that the reader will nonetheless benefit from the insights and experience available to someone who has been a long-time, active participant in several F/OSS projects. The innovation cycles of archaeological computing are naturally connected with the much bigger dynamics of the bigger IT industry and business world. Therefore, this chapter will occasionally broaden its perspective to a more global view, whenever that helps to understand the current role of F/OSS in archaeology. However, for all practical details and as far as the real-world case studies are concerned, it will focus on applications within archaeology.

Advocates of Free and Open Source Software like to emphasise that the "Free" stands for much more than gratis computer programs, namely everyone's liberty (the term "libre software" is being used with increasing frequency) to use, modify and also redistribute software. This property of F/OSS clearly sets it apart from "freeware" (which is a more limited give-away) and "shareware (which is usually an even more limited give-away that serves as an incentive to purchase the fully functional version). Neither of these terms overlaps completely with what is known as the "public domain". Works in the public domain do not have any form of copyright protection; F/OSS, however, does, and its authors do not give up their copyright privileges by default. Instead, they distribute their software under a modified version of traditional copyright (cf. McGowan 2005), often referred to as “copyleft” (Stallman, 2002, p. 127–134) and specified in detail by a license agreement such as the popular GNU General Public License (GPL; https://gnu.org/licenses/gpl.html; see also Stallman 2002, p. 165–203; Fogel 2009, p. 162–175). Typical open source license terms are very liberal, but some limits remain (see Rosen 2005 for comprehensive coverage of different open source licenses). Some software, for example, may only be used freely for academic and non-commercial purposes.

In fact, this kind of dual-licensing, that attempts to generate revenue from those licensees that “can afford paying”, is not uncommon; and it shows that F/OSS can support various business models. It is therefore somewhat misleading to speak of “commercial software” as the antithesis of F/OSS. Better terms for non-F/OSS are “closed source” or “proprietary” software. Avoiding to think of F/OSS as non-commercial also helps avoiding another misconception, namely that F/OSS is primarily created within an altruistic sphere of students, hobbyists and enthusiasts. In fact, most F/OSS programmers are professionals that contribute open source code, because they get paid, or because they have other, “selfish” reasons (see Klemens 2005, p. 96; Ghosh 2005). These clarifications and distinctions are critical for understanding both the economic and academic properties of F/OSS.

7.2 Selected Aspects of F/OSS

The following discussion points will sometimes make strong presumptions. Adequate evidence for their validity will only be given later, in the context of the real-world case studies. At this point, the intention is merely to provide some background to the most important aspects stressed within this chapter. The selection of aspects exposed is very reductive and does not attempt to do justice to the full complexity of the open source phenomenon. In particular, nothing will be said about the technological merits (or shortcomings) of specific F/OSS versus specific proprietary software. This would be a futile undertaking, given the pace of software evolution, and the widely diverging user skills and expectations. More extensive coverage, also on the roots of the open source movement, is provided by the well-known works of Raymond (2001) and Stallman (2002), with the latter striking a more ideological tone. For those capable of reading German, the book by Grassmuck (2004) provides another free and detailed source of information.

7.2.1 Open Source Economics

No academic software culture can prosper in the absence of continuous investment. The mere fact that a thriving F/OSS community exists in this world proves that the idea of free software is fully compatible with a competitive capitalist economy. The availability of free, highly effective solutions for code sharing and online team work allows global collaboration with unprecedented intensity. That a modern business model does not necessarily involve proprietary, in-house programming and selling of licenses, is indeed one of the most important lessons to be learned from successful F/OSS projects (cf. Fogel 2009, p. 75–88; Lerner and Tirole 2004; von Hippel 2005, p. 265–346). Presumably, the underlying causes for the continuing lack of tailor-made, archaeological software solutions are to be found in a combination of the discipline's little economic significance, short-lived and project-based funding cycles, and a habitual acceptance of the "pay-per-license" model, rather than a lack of interest among its scholars and practitioners. F/OSS, on the other hand, lends itself to demand-driven and pooled funding models that offer clear and real advantages in an environment where individual decision makers have little financial resources at their disposal.

After all, the costs for a fully equipped workstation with licensesfor CAD, GIS, DTP, and perhaps some software for statistical analysis, can easily approach the price of a new luxury car. Although this may not be an issue for members of the academic community, who are often fenced against such exorbitant cost by "campus license" agreements (for which the general tax payer more often than not picks up the bill instead), it is a critical problem for the much larger group of archaeological practitioners "out there", who work for public services or for commercial companies, or are self-employed. It comes as no big surprise then, that many software users are attracted to open source solutions first and foremost because they view them as a way to save money. However, this strong "zero cost" attractor can lead to a certain conundrum in the context of a long-term economic strategy. Firstly, while it may be true that many open source alternatives to proprietary software exist, not all of them can be considered direct, "drop-in" replacements. Complex software such as GIS requires extensive user training and often employs undisclosed data formats or patented technology that cannot be used by open source competitors. This can mean that users will not be able to migrate away from proprietary software without a very costly overhead for data conversion; an effect that is known as "vendor lock-in", and that will play a prominent role in the case studies to be discussed. Secondly, the development of F/OSS requires resources, just like any other form of software development. For larger projects, the resource requirements can be immense. As an example, the open source GIS gvSIG CE (to be discussed in detail later) consists of roughly 1.5 million lines of program source code. This represents programming work worth many millions of Euros. Clearly, endeavours of this magnitude require continuous investment (be it in the form of money or dedicated work time), carried by a broad base of users and supporters.

Sooner or later, any ambitious F/OSS project must therefore look for funding to sustain itself. Save for the discussion of selected aspects exposed via the case studies, this chapter will not get into the details of different F/OSS-based revenue models, as this would require a separate treatise (such as Krishnamurthy 2005). Generally speaking, the fact that common sources of income, mostly license fees, are not available to non-proprietary software projects means that traditional business models also do not apply. This can be a challenge, since it requires F/OSS advocates to craft tailor-made revenue strategies and argue for their feasibility when looking for financial support from traditional-minded investors and public agencies. On the other hand, this need for creativity has opened up a plethora of new opportunities. At at time, when traditional software vendors are struggling to convince their clients of the "benefits" of ever more restrictive, top-down "cloud" and subscription-based models, F/OSS offers attractive alternatives, based on bottom-up, community-driven software development, that promise drastically increased flexibility and return on investment.

Besides the technological output in the form of software, the greatest return on investment in F/OSS comes in the form of social capital. Direct involvement in the decision making process, design and programming of complex software allows for transfer and sharing of skills and knowledge in a way that surpasses what can be gained from buying and using proprietary, closed-source software. As will be discussed later, this is of critical importance when it comes to the use of software in academic and research environments. At this point, the important thing to note is that there is an intimate link between the social and technological health of any F/OSS project: Due to the fact that social capital plays such a central role, a thriving and content user community is of utmost importance. More often than not, open source contributors are faced with the challenge of not just contributing money or program code, but also making sure that their contributions fit into the social fabric of the project.

7.2.2 Social Dynamics of F/OSS

An open source project is, by its very nature, averse to clearly marked ownership or predefined and rigid hierarchies. This makes the social dynamics, especially those of larger projects, all the more important (Ghosh, 2005; Lakhani and Wolf, 2005). Human emotions are constantly at work behind the scenes of F/OSS development, and negative ones are vented, often immediately and unfiltered, through public communication channels, such as mailing lists. Although this specific form of openness does have the advantage of preventing conflicts from simmering, it also stresses the need for good communication skills (Fogel, 2009, p. 98–117). In the absence of professional mediators, a small misunderstanding that may be quickly forgotten in "offline" communication, may stay on record and cause friction for a very long time on the Internet.

Bad social climate will result in the loss of precious social capital. And thus an open source project that does not achieve a feeling of equality and belonging among the members of its community, is destined to fail in the long run, no matter how generous the funding or how highly developed the contributors' technological skills may be. On the other hand, a project that can build and sustain a loyal community will be able to weather even the toughest of economic times. As an example, the open source GRASS GIS (http://grass.osgeo.org; Neteler and Mitasova 2008) has been in the public domain for three decades. This is a tremendous achievement that is on par with the longest running proprietary enterprises and far exceeds the meagre average of three to five years for a funded research project. It is all the more impressive, given that the project has had to find new backers and investors multiple times throughout its long history.

To achieve such longevity, an open source project must assign the highest priority to accessibility, transparency and inclusiveness in its decision making processes. Intriguingly, these same attributes are also highly desirable in academic research. They are the reason why, as will be elaborated later, F/OSS is so conducive to good scientific practice. What all this actually means for managing an F/OSS project, is something that needs to be learned by doing. Much of today's web technology is geared towards flat-hierarchy interaction, and any aspiring open source project should make ample use of collaborative tools, such as mailing lists, online forums and wikis.

One important thing that needs to be kept in mind,however, is that a web presence, as immaterial as it may be, still amounts to a form of "territorial claim" in social terms. In order to attract external collaborators, such a claim should therefore be made on behalf of the entire community, not a single project partner. If, for example, the "University of A" initiates an F/OSS projected hosted at "http://university-of-a/our-great-project", then potential collaborators from the "University of B" will instinctively be much less inclined to join up, no matter how benevolent the initial idea might have been. This should serve as an example, learned through painful experience, of just how fleeting social capital in a highly interactive environment such as the Internet can be.

As a consequence, it is generally easier and more effective to join an existing project and share its resources, than to start from scratch. At the time of writing this, the North American service provider Sourceforge.net alone was hosting more than 200,000 active open source projects (http://sourceforge.net). There is a good chance that one of them will at least provide part of whatever software solution an archaeologist might be looking for. Attaching oneself to a larger, long-lived project does not only allow immediate access to greater technological and social capital. It will also ensure that investments will remain effective for much longer than the lifetime of the average academic funding cycle or budget plan.

7.3 F/OSS in Research

Archaeological research is increasingly based on computational methods and software. In fact, it would be difficult to name any aspect of the discipline that remains completely devoid of computer technology today. These developments have profound economical and epistemological implications that must be critically reviewed. This is particularly true for financial barriers to reproducible research and data processing opacity, both of which are unavoidable effects of the use of proprietary software.

Therefore, whereas in a commercial context open source may be one suitable option among several, in an academic context it is really without alternative (see also Ince et al. 2012). Proprietary software acts like a "black box" in research (Morin et al., 2012, p. 159) and thus stands at odds with good scientific practice. Firstly, it prevents researchers and students from ever fully understanding every detail of the data processing. Secondly, it transgresses against what is arguably the most fundamental requirement of science, that of reproducibility (see below). Given the same data and methods, any number of independent researchers should be able to recreate each other's studies, verify or falsify them and improve upon them (see also Chambers 2008, p. 6–7). This is especially critical in archaeology, a discipline that has no established concept of proof, by which to assess the validity of specific methods, hypotheses or conclusions.

F/OSS presents itself as an obvious alternative; one that is better aligned with both the economic constraints of project-based research and the demands of good scientific practice. So far, however, little independent research has been published that speaks clearly about these aspects and their relevance to archaeological research. There is some urgency in rectifying this situation. As long as the rapid evolution of software is not being accompanied by increased scrutiny regarding the transparency of computational methods and data processing, the opacity of software-based research will increase and become ever more problematic (see Ducke 2012 for a current, albeit partial view on this problem).

7.3.1 Publish (Your Source Code) or Perish!

> "Academic computer science has an odd relationship with software: Publishing papers about software is considered a distinctly stronger contribution than publishing the software. The historical reasons for this paradox no longer apply, but their legacy remains." (Hafer and Kirkpatrick, 2009, p. 126)

Those archaeologists with an interest in the theory and application of computational methods will have been confronted with the statement that computers are "nothing more than tools". Tools, of course, are designed to fit some relatively primitive purpose, and are thus simply not important enough to bother with such mundane concepts as computational transparency. This opinion, still commonly held in academic archaeology, is a crass underestimation of the role that mathematics and computing play in modern research. After all, computers are the technological manifestations of mathematical reasoning, an advanced aspect of human intellect, and computing is applied mathematics. Thus, the connection between software and the mathematical methods used to model, understand and ultimately solve real-world research problems is very explicit. Computer programs, that frequently run into the millions of lines of program code, are perhaps already humanity's most comprehensive, certainly its fastest growing repository of formalised knowledge.

An illustrative example from archaeology is the publication on Bronze Age trade networks by Knappett et al. (2008). The part of the this study that has been published in paper form amounts to a mere introduction to the research theme. The associated Internet resources (http://theory.ic.ac.uk/ time/networks/arch) are somewhat more informative. But for those interested in the details of the mathematical models and formal reasoning behind the study, the source code of the software is the only place that can provide detailed insights. As another example, numerical simulation using agent-based modelling (ABM) has made a considerable impact in archaeology in recent years (e.g. Kandler et al. 2012). ABM is representative of a modern computing approach that is entirely unsuitable for traditional publication. There is simply no way in which printed static screen images or textual descriptions could adequately convey the impressions gained by interactively modifying and observing a dynamic ABM simulation. And it is not possible to understand what an ABM is doing, unless there is full disclosure of the program code behind the simulation. It is therefore no coincidence that proprietary solutions are all but absent from ABM research (http://www.openabm.org; see also Janssen et al. 2008). For similar reasons, the F/OSS project R (http://r-project.org) has grown into an extensive repository of applied statistical research, marginalising the scientific role of proprietary offerings in areas such as spatial analysis, that are of central importance to archaeological research (Bivand et al. 2013; Chambers 2008).

As software-based research becomes ever more pertinent in archaeology, the effectiveness of traditional forms of publication must be doubted wherever mathematical models and computing are concerned. In this respect, the prevailing modus operandi (not just in archaeology) leaves much to be desired. Indeed, withholding the source code from the academic community amounts to withholding the very means which enable others to peer review and learn from publicly funded research (Morin et al., 2012):

> "Despite increasing reliance on computing in every domain of scientific endeavor, the computer source code critical to understanding and evaluating computer programs is commonly withheld, effectively rendering these programs 'black boxes' in the research work flow. "

However, the scope of this "black-box problem" is not limited to the aspect of withheld knowledge. The use of proprietary software in academic research, even for the most routine tasks, is an inherently flawed approach to scientific practice and scrutiny. This will become more obvious when thinking about the critical concept of reproducibility of software-based research.

7.3.2 Reproducible Research

> "[..] the results of scientific calculations involving significant amounts of software should be treated with the same measure of disbelief as an unconfirmed physical experiment." (Hatton, 1997)

Modern archaeological research employs complex software, most notably GIS, that provides hundreds of data processing functions. Bearing in mind the real limits of software testing and quality assurance, it cannot be assumed that each one of them is free of errors and will always produce the expected results. In addition, even the most basic numerical methods (algorithms) exist in a variety of implementations; be it because closed source code forces programmers to re-invent the same algorithm multiple times, or because perceived shortcomings call for modified and improved versions. As a consequence, not even the most basic operations, such as a simple line-of-sight analysis in GIS, can always be accurately reproduced across different software platforms (Ducke, 2012, fig. 1).

However, if it is impossible to read the actual program code that works on the data, it is also impossible to account for unexpected results. In addition, as everybody knows all too well from personal experience, any given software has a significant number of errors (cf.Merali 2010), and any problems encountered in the use of software may be due to one of these "bugs" as much as to flawed input data or other factors. Therefore, full publication of the source code has been a demand in science for some time (e.g. Buckheit and Donoho 1995; Donoho et al. 2009; Barnes 2010).

There are few means to assess the quality of archaeological research objectively. Reproducibility is one obvious criterion, as it seems indisputable that only reproducible results can be built on and verified or (more importantly) falsified by independent peers. It seems therefore imperative to uphold the ideal of fully reproducible research and to not sacrifice it needlessly, certainly not for the sake of convenience or for the mere habit of using proprietary software. After all, even if closed source software could square its own circle and function in a flawless and fully transparent manner, its cost would impose another, equally significant, limitation on the reproducibility of research.

7.3.3 Data-Centric Research

Nothing has been said so far about the relations between F/OSS and the second main resource of computing, the data. For obvious reasons, denying others access to research data ultimately leads to the same problems regarding transparency and reproducibility, as denying others access to documentation or program source code. This has been recognised in archaeology, as the Journal of Open Archaeology Data (http://openarchaeologydata.metajnl.com/) and related endeavours demonstrate (see also Kansa 2012). Although this chapter cannot discuss the interrelated, and without a doubt important, aspects of "open data" and "open access", it should be noted that F/OSS provides the best technological solutions for making data open and accessible, particularly in the long term, to the general public or at least the academic world. The fact that the Internet basically runs on open source software speaks volumes in this respect.

Good research requires discipline, self-regulation, and constant questioning of one's chosen methods and preconceptions. In the context of software-based research, one of the most dangerous traps into which one might fall is that of "application-centric" thinking. In this all-too-seductive mode of thinking, the researcher allows the available capabilities of specific, reassuringly familiar, software to determine the analytical methods. The best safeguard against letting the software impose limits on the potential outcome of the research, is to acquire a habit of "data-centric" thinking instead.

As the name suggests, a data-centric approach firmly places the data into the centre of the research workflow. This approach recognises that the data represents the most valuable and irreplaceable investment, and that the choice of application software is a concern of secondary importance. After all, given the diversity of today's software offerings, there is always an alternative. Modern data management technology supports this perspective by allowing users to create shared data infrastructures, accessible through standardised interfaces and protocols. End-user software can then be attached to a central data repository, allowing access to the data on different levels, via multiple user interfaces. With regard to GIS and spatial data infrastructures, a robust technical criterion for the inclusion or exclusion of software is how well it supports the standard protocols and data formats specified by the independent Open Geospatial Consortium (http://www.opengeospatial.org/).

Besides facilitating the ideas of data sharing and transparent processing, open and standards-based data infrastructures also alleviate the risk of data getting locked into undisclosed proprietary formats and thereby enable long-term data storage, archiving and accessibility. From an educational and academic point of view, the data-centric approach is attractive because it favours broader, transferable skills over narrower, application-specific skills. From an economical point of view, it opens up a broader range of investment options and more finely grained control over software spending. It is generally more cost-efficient to modify or even create software that integrates well into an existing, open infrastructure, than to license all the components required for a complete proprietary infrastructure. This is certainly the case for spatial data infrastructures and GIS, for which complete open source solutions exist (cf. Sherman 2008).

7.4 Case Studies

Having discussed some economic, social and academic aspects in a rather abstract manner, it is now time to let the proverbial rubber hit the road and take a look at the potentials and challenges of F/OSS in the real world. The following three case studies provide inside looks at attempts to include F/OSS as a central component in the development and usage of software to support archaeological applications and workflows. All of them demonstrate an involvement in the theory and practice of open source that goes far beyond the simple "gratis software" approach.

The discussion of each case study includes some background information (most importantly answering the question why F/OSS was chosen to play a central role), a description of the main challenges and the means employed to address them, and a short analysis of why critical goals were achieved or not achieved. Note that these short exposes cannot provide complete accounts (in fact, all of the cases discussed here are ongoing projects), but they may still provide valuable lessons that reinforce the advise given so far and help the reader devise good F/OSS strategies.

7.4.1 Oxford Archaeology Digital: F/OSS Migration in the Workplace

Oxford Archaeology (OA) is one of the world's largest providers of archaeological services (http://thehumanjourney.net). Commercial archaeological practices such as OA have become indispensable pillars of heritage management in the UK, providing employment for thousands of archaeologists across the country and evolving in the contact zone of archaeology, landscape conservation and the construction industry. This environment places the highest demands on accountability, cost-efficiency and flexibility. It was therefore no small decision for OA to initiate a migration away from proprietary solutions and towards F/OSS. After all, having invested into proprietary databases, CAD and GIS for decades, the threat of encountering costly vendor lock-ins was omnipresent.

One important driver behind the decision to switch to F/OSS was the unpredictability of licensing costs in the long term. Although not all proprietary licenses have an official time limit after which they must be renewed, many do have a factual one. Since software vendors frequently change ("update") the file formats used by their software, and since these formats are generally undisclosed, users who do not renew licenses frequently will soon be unable to exchange data with their clients and contractors. However, the vendor is free to modify the terms of the license agreement with each new deal, exposing the licensee to the risk of rising prices or other changes for the worse, each time through the cycle. When the latest such issue hit OA in the form of significantly increased licensing cost for their proprietary desktop GIS, a decision was made to prioritize finding an F/OSS replacement for the proprietary GIS licenses.

Another, not less important, driver, was the realisation that the potential for software-based innovation in commercial archaeology and the establishment of new digital revenue models could only be unlocked on the basis of F/OSS. Given the economic climate of commercial archaeology, attempting to earn money through customised software development, consulting, and paid-for technical support and training was simply not viable, as long as it involved prohibitively high licensing cost on the side of either the service provider or the client. The result of this insight was the founding of Oxford Archaeology Digital (OAD), a division within OA with the objective to create and promote open source software for archaeology and to find new F/OSS-related business opportunities (http://oadigital.net). More than five years on, OAD still stands out as one of the most concerted efforts to realise these aims. Among its greatest successes is the release of gvSIG OA Digital Edition, the first F/OSS GIS to provide a complete drop-in replacement for proprietary desktop GIS. The involvement of OAD in the development of the open source desktop GIS gvSIG is a particularly instructive effort that will be elaborated on in the case study on gvSIG CE.

In addition to tangible output in the form of free software, OAD also contributed research in the field of digital archaeological site documentation. It was among the first to systematically assess and publish the feasibility of using open source software to generate highly detailed, three-dimensional models from overlapping digital images (Ducke et al., 2011) and the use of ultra-portable communication devices ("smartphones") to replace paper forms in the field. Not all of these efforts have led to fruition, but that was not to be expected. Even Silicon Valley produces more short-lived, technological failures than long-lived successes.

Indeed, in the context of this chapter, the more pressing question is whether OA's other aim, the migration of its in-house IT to F/OSS, was a success. After all, a full migration of such a large and complex operation to open source software could serve as a benchmark case for the successful switch of a critical production environment. Unfortunately, the answer in this case cannot be clearly positive; but the lessons learned by OA are still valuable.

First of all, it quickly emerged that social inertia was one of the greatest obstacles (see also Stallman 2002, p. 245–246). Human beings can be surprisingly attached to technology, to a point where they establish an emotional connection to an inanimate piece of hardware or software. Such attachment is not entirely irrational. After all, learning to effectively use complex technology requires a huge personal investment. Not only would that investment be partly lost after the switch to another technology, but more importantly, an individual's competitive "edge", marked by e.g. "mastery" of a certain piece of software, would also become blunted. Such cases of interest of conflict call for clear company policies, proactive communication, intense staff training and the central deployment of new technologies, all of which consume considerable resources.

Unfortunately, the drain on OA's resources was such that the F/OSS transition could only be a partial success. The organically grown, bottom-up structure of OA proved to be a hindrance for establishing a new central IT infrastructure based on open standards and technology. The case-by-case deployment of F/OSS solutions, although successful in the initial, testing stage, proved problematic in the context of larger operations. The technological blame for this lies squarely with the proprietary file formats and interfaces. As long as it remains legal for software vendors to deny their customers the details of their data formats, there will always be a risk that the cost for switching to an alternative solution will be overwhelming. Even if an in-house F/OSS transition can be completed, there remains the issue of clients handing in or expecting delivery of proprietary data formats to fit into their own workflows. Therefore, a complete phasing-out of proprietary software seems almost impossible to achieve for an operation such as OA's within the current market conditions. The next case study will tell the much-related story of a large public institution's F/OSS migration. But this time, the resources available are several magnitudes greater.

7.4.2 gvSIG and gvSIG CE: The Role of Social Capital in F/OSS

Without a doubt, GIS is one of today's most important software platforms. Of critical importance not only to science and research, but also to businesses, consumers, and public agencies, accessible (in terms of cost and ease-of-use) GIS is a corner stone of modern information technology. Until recently, however, a "drop-in" F/OSS replacement for proprietary desktop GIS, i.e. a system that would not require its users to completely rethink their approach to GIS, that would allow them to continue working without having to convert their data to another format first, and that would cover the entire workflow, from data editing and processing to map publication, was simply not available. The fact that this situation has changed dramatically, and that archaeologists, among others, no longer need to pay for expensive proprietary GIS, is in no small part thanks to gvSIG.

The history of gvSIG (Generalitat Valencia Sistema de Información Geográfica) goes back to 2003, in which year the Spanish software house Iver was awarded a contract by the Regional Ministry of Infrastructure and Transport of Valencia (CIT) to develop a new, open source GIS. The aim of the development, endowed with generous funding, was clearly defined: to replace proprietary solutions for spatial database access, CAD and GIS with one integrated software, functional and stable enough to be used in cadastral works, spatial planning and the management of public infrastructure. However, despite these promising initial conditions, gvSIG never managed to match its main F/OSS competitor, Quantum GIS, in popularity and has instead remained largely confined to a smaller, Spanish-speaking community. Today, in times of austerity, the CIT's ambitious project has lost much of its initial momentum and development activity has slowed significantly (see Boga et al. 2011). The case of this software, therefore, provides an illustrative example of the consequences that can arise from relying on internal strength alone, and from failing to capture external social capital.

To understand what happened, one must look back at 2009. That year saw the release of gvSIG 1.9. Initially conceived as the last incarnation of the 1.n code base, which was to be succeeded by a completely reworked version 2.0, gvSIG 1.9 was in fact the first F/OSS desktop GIS solution that could be considered a fully functional replacement for established proprietary GIS. Unfortunately, however, the quality of the software did not meet the general public's expectations. It soon became obvious that error testing had been conducted entirely within the narrowly defined workflows at the CIT, and that, when faced with different types of data and use cases, obvious malfunctions ensued. Other problem areas were the incomplete English translation of the user interface, which sometimes left users without Spanish reading skills clueless about the content of on-screen messages, and a prevalence of rough edges in the graphical user interface that hampered productive workflows.

None of these problems would have been fatal within a regular open source project. In such a case, internal developers and external contributors would file error reports and add corrections to the code, until the software would become usable again and eventually another, better release could be made. In the case of gvSIG, however, it quickly became obvious that these mechanisms could not take hold. Despite being publicised as an open source project by CIT and its contractors, gvSIG appeared far from open as regards its organisation and management. When the project surfaced on the Internet, it materialised as an opaque entity, with a strict hierarchy and an internal decision making process that favoured controlled communication and blocked outside influence on important technical decisions. This was accompanied by rigorous routines and rules for code contribution that were a far cry from the low-key, fast turn-around practice, called "agile development", so popular with open source programmers (http://www.agilealliance.org; see also Robbins 2005).

To make matters worse, after the release of gvSIG 1.9,the project's technical steering committee decided to focus all future development efforts on the release of version 2.0, which was to be written from scratch and had no set release date. Reverting to an older version of gvSIG was also not a feasible option for most users, as the next-older version, gvSIG 1.1.2, was far inferior in terms of its functionality and lacked many capabilities required of a professional GIS solution. Thus, users were trapped between an outdated version, an error-prone version and one that existed largely as a prospect.

In hindsight, what happened next was to be expected. Oxford Archaeology (see preceding case study), an archaeological service provider that had invested into gvSIG and depended on it as the core element of its F/OSS GIS migration, used its own resources to improve gvSIG 1.9 to the point where it could be used for productive work. The result, named gvSIG OADE 2010 remains one of the best, most comprehensive options for free desktop GIS available today. Its success eventually moved the CIT's project team to follow up on the 1.9 release with further improved versions, in time leading up to the current 1.12 release.

While OA's actions certainly solved a technological problem, they also proved to be the final nail in the coffin for friendly relations with the "official" gvSIG project. Realising that it would not be able to maintain a project the size of gvSIG OADE 2010 on its own, OA started to look for collaborators. It found them in a number of users and developers that had been equally estranged by the CIT project team's practice, and a new project, called gvSIG Community Edition (CE) was started. In technical terms, the CE project is a "fork": Since its inception, two different development teams have been contributing code to the two different versions of the software. It is the natural fate of forks to drift apart. Currently, code written for both versions is still largely interchangeable, but a point of no return will eventually be reached.

As opposed to the OADE version, which was largely welcomed by gvSIG users, the CE fork caused more controversy, not only between the different camps, but also within them. However, looking back at the original reasons that led to the fork, it becomes clear that there was no better option. Since direct collaboration had become impossible, the only other choices for the "breakaways" would have been to turn towards a competing project, such as Quantum GIS, to start a new project from scratch, or to go back to proprietary solutions. Thus, as opposed to an often voiced opinion, forks are not the worst way to resolve such conflicts, but rather a common occurrence that in many situations constitutes the least harmful path, as they preserve at least some potential for collaboration. In addition, the investments of those partnering in the CE fork have been successfully preserved.

7.4.3 Survey Tools: F/OSS for Field Archaeology

The case studies discussed so far have provided evidence that financial and social resources must both be sufficient and managed with equal diligence if an F/OSS project is to be sustained. This, final case study is an example of a project that intends to put such insights into practice. Because Survey Tools (http://www.survey-tools.org), a project dedicated to creating light-weight F/OSS tools to be used in field documentation and surveying, is still in its initial phase, it is too early for a verdict on whether its approach to F/OSS in archaeology will ultimately be sustainable. However, its technological focus should be of greatest interest to the archaeological reader.

The primary motivation for Survey Tools lies in gaining flexibility and technological independence. Faced with an initial situation similar to that of Oxford Archaeology and the CIT, the State Heritage Management (SHM) of the German state of Baden-Württemberg needed to find a way of mitigating its dependence on costly specialist software for field documentation. This led to an internal review of actual user needs. Under the direction of the SHM's Digital Archaeology unit, current field workflows were analysed and individual F/OSS solutions were considered. A central element in the SHM's strategy was the transition of topographic survey activities from proprietary

CAD to F/OSS GIS (more specifically, gvSIG CE). To make this possible, a new software had to be devised to act as the link between the surveying hardware and the GIS.

The result was the development of survey2gis, a flexible and user-friendly open source tool, capable of processing raw survey records from devices such as total stations and GPS and converting them into topologically cleaned GIS datasets. After a prolonged phase of testing and refinement, the software was made available to the public as the first component of the Survey Tools. After the completion of the initial funding phase, the project must now look for sustainable funding outside of the SHM. This is done through a collaborative platform on the Internet, paid-for support and subscription models, actively advertised on specialist meetings and conventions.

One of the most intriguing aspect of the Survey Tools project is its ability to show how F/OSS can unlock innovation potential. Prior to the inception of survey2gis, the SHM's field workflows had oriented themselves along the lines defined by user interfaces and functionalities of proprietary software, such asCAD.With the freedom to create new, customised software, however, also came the freedom to inspect and modify existing workflows in order to make them more efficient. As a result, survey2gis is highly customisable and includes a number of features designed to boost productivity in the field. This is a significant type of return-on-investment that is often overlooked when comparing the license fee savings against the cost of open source software development and staff training.

7.5 Conclusions

This chapter was not written as a condemnation of either proprietary software or traditional business models. F/OSS and proprietary software coexist and will continue to do so for the foreseeable future, as the diversity of user demands and expectations calls for an equal diversity of approaches. From a financial point of view, there are certainly scenarios in which proprietary offerings are worth their money, provided that they can solve a clearly specified problem or make a specific workflow more cost-efficient. If such an off-the-shelf product suits the user's needs, then it may be the most readily available solution. And as long as software is really just used a tool, withheld source code might not be an issue. However, excessive or fluctuating license fees, the risk of vendor lock-in, a lack of shared investment options, and not least the serious limitations of closed-source programs in research and education all speak in favour of considering alternatives.

The fact that intra-disciplinary software development in archaeology remains confined to sporadic investment and small-scale developments, suggests that alternatives are indeed required. At present, however, opportunities for long-term funding of archaeological F/OSS remain rare and mostly restricted to the commercial sphere. In this respect, the role of universities and the academic funding system need to be reviewed. For an outsider, it can be hard to understand why public research money is more often spent in a way that benefits software corporations, then in a way that benefits open research and the general public. It should also be noted that technological developments in the commercial and academic spheres are ultimately linked. Companies attempting to break free of vendor lock-ins are not the least struggling, because university departments keep teaching their students application-centric thinking instead of transferable skills. The fact that such curricular alignment with the proprietary software industry is often done in the name of “the job market”, must seem ironic to employers like Oxford Archaeology, and to all those interested in investing into archaeological F/OSS.

Strictly speaking, as far as education and research are concerned, there seems no justifiable role for closed source software, except to serve a tool-like purpose for the most menial and routine tasks. This might sound harsh, but the fundamental ideals of good scientific practice, in particular that of reproducible research, are simply incompatible with trade secrets and the many barriers that proprietary software imposes on the free flow of information (Stallman, 2002, p. 57–58). Software-based research can be expensive, but there is no reason why it should be prohibitively expensive to reproduce such research. Licensing costs have become considerable obstacles for public institutions and small research projects. At the same time, the side-effects of proprietary software business, in the form of excessive “intellectual property” enforcement and software patents, all of which go far beyond the original, fair-use intent of copyright law, are threatening free science (see Klemens 2005 for a detailed account; also Stallman 2002, p. 89–92 & 105–134).

Indeed, the problems of closed source software become strikingly obvious in the context of its academic use. In the scientific domain, peer review of software should be as mandatory as that of text, and source code should be considered a part of the academic output and published accordingly. It would be curious indeed for an academic discipline to encourage peer review for philosophical treatises, where it matters least, but not for scientific software, where it matters most! The “missing functionality” argument against the exclusive use of F/OSS, at least, is no longer valid (provided that it ever was). On the contrary, projects such as GRASS GIS and the R language for statistical computing are immense repositories of scientific methods.

The limits of the usefulness of software are ultimately set by the paradigm under which it operates. In archaeology, software has traditionally been viewed as a tool that just serves a well-defined purpose, but this view is too narrow. Complex programs represent the result of countless hours spent on brainstorming sessions, elaborate project designs, fundamental and applied research, creativity and problem-solving skills. Software must therefore be published in its source code form, so that it can undergo the collaborative cycle of peer review, exchange and refinement that is commonly called “research”.

Finally, it should be noted that technology, like everything human-made, is a social phenomenon. F/OSS tends to bring this fact into the foreground. Getting involved in an open source project exposes all collaborators to social dynamics that must be managed well, if an open source investment is to bear fruits. While this can be challenging at times, building a loyal open source community will result in insightful, careful and sustainable development, and in the growth of technological infrastructures that are open, diversified and innovative.

Bibliography

Barnes, N. (2010), 'Publish your computer code: it is good enough', *Nature* **467**(7317), 753–753.

Bivand, R. S., Pebesma, E. and Gómez-Rubio, V. (2013), 'Applied spatial data analysis with r'.

Boga, A., Puga, F., Eiries, A. and Varela GarcÃŋa, F. (2011), 'Analysis on free software communities (I): a quantitative study on GRASS, gvSIG and QGIS'.
URL: http://nosolosoftware.com/analysis-on-free-software-communities-i

Buckheit, J. B. and Donoho, D. L. (1995), *Wavelab and reproducible research*, Springer.

Chambers, J. (2008), *Software for data analysis: programming with R*, Springer.

Donoho, D. L., Maleki, A., Rahman, I. U., Shahram, M. and Stodden, V. (2009), 'Reproducible research in computational harmonic analysis', *Computing in Science & Engineering* **11**(1), 8–18.

Ducke, B. (2012), 'Natives of a connected world: free and open source software in archaeology', *World Archaeology* **44**(4), 571–579.

Ducke, B., Score, D. and Reeves, J. (2011), 'Multiview 3d reconstruction of the archaeological site at weymouth from image series', *Computers & Graphics* **35**(2), 375–382.

Fogel, K. (2009), *Producing open source software: How to run a successful free software project*.
URL: http://producingoss.com

Ghosh, R. A. (2005), 'Understanding free software developers: Findings from the FLOSS study', *Perspectives on free and open source software* pp. 23–46.

Grassmuck, V. (2004), 'Freie software. zwischen privat-und gemeineigentum. 2., korr. aufl., red'.

Hafer, L. and Kirkpatrick, A. E. (2009), 'Assessing open source software as a scholarly contribution', *Communications of the ACM* **52**(12), 126–129.

Hatton, L. (1997), 'The t experiments: errors in scientific software', *Computing in Science and Engineering* **4**(2), 27–38.

Ince, D. C., Hatton, L. and Graham-Cumming, J. (2012), 'The case for open computer programs', *Nature* **482**(7386), 485–488.

Janssen, M. A., Alessa, L. N., Barton, M., Bergin, S. and Lee, A. (2008), 'Towards a community framework for agent-based modelling', *Journal of Artificial Societies and Social Simulation* **11**(2), 6.

Kandler, A., Perreault, C. and Steele, J. (2012), 'Cultural evolution in spatially structured populations: A review of alternative modeling frameworks', *Advances in Complex Systems* **15**(2).

Kansa, E. (2012), 'Openness and archaeology's information ecosystem', *World Archaeology* **44**(4), 498–520.

Klemens, B. (2005), *Math you can't use: Patents, copyright, and software*, Brookings Institution Press.

Knappett, C., Evans, T. and Rivers, R. (2008), 'Modelling maritime interaction in the Aegean Bronze Age', *Antiquity* **82**(318), 1009–1024.

Krishnamurthy, S. (2005), *An analysis of open source business models*, MIT Press, pp. 279–296.

Lakhani, K. R. and Wolf, R. G. (2005), *Why hackers do what they do: Understanding motivation and effort in free/open source software projects*, MIT Press, pp. 3–22.

Lerner, J. and Tirole, J. (2004), 'Economic perspectives on open source', *Advances in the Study of Entrepreneurship, Innovation & Economic Growth* **15**, 33–69.

McGowan, D. (2005), *Legal Aspects of Free and Open Source Software*, MIT Press, pp. 361–391.

Merali, Z. (2010), 'Computational science: Error, why scientific programming does not compute', *Nature* **467**(7317), 775–777.

Morin, A., Urban, J., Adams, P., Foster, I., Sali, A., Baker, D. and Sliz, P. (2012), 'Shining light into black boxes', *Science* **336**(6078), 159–160.

Neteler, M. and Mitasova, H. (2008), *Open source GIS: a GRASS GIS approach*, Vol. 2, Springer.

Raymond, E. S. (2001), *The Cathedral & the Bazaar: Musings on linux and open source by an accidental revolutionary*, " O'Reilly Media, Inc.".

Rey, S. J. (2009), 'Show me the code: spatial analysis and open source', *Journal of Geographical Systems* **11**(2), 191–207.

Robbins, J. (2005), 'Adopting open source software engineering (OSSE) practices by adopting OSSE tools', *Perspectives on free and open source software* pp. 245–264.

Rosen, L. (2005), *Open source licensing*, Prentice Hall.

Sherman, G. (2008), *Desktop GIS: Mapping the Planet with Open Source Tools*, Pragmatic Bookshelf.

Stallman, R. (2002), 'Free software, free society: Selected essays of Richard M. Stallman'.

von Hippel, E. (2005), *Open source software projects as user innovation networks*, Cambridge, Massachusetts: The MIT Press, pp. 267–278.

Wheeler, D. A. (2005), 'Why open source software/free software (OSS/FS, FLOSS, or FOSS)? look at the numbers!'.

Shawn Ross, Brian Ballsun-Stanton, Adela Sobotkova and Penny Crook

8 Building the Bazaar: Enhancing Archaeological Field Recording Through an Open Source Approach

8.1 Introduction

This chapter summarises the experience acquired by the Federated Archaeological Information Management Systems (FAIMS) project over the course of developing open-source software for archaeologists. open-source software development, which excels at coordinating discrete contributions from many people and organisations, offers the best hope for producing complex and expensive tools in a discipline where resources are limited. Over the course of this project, we have come to realise that open-source approaches have applications in archaeological research beyond the development of software itself. The development of redeployable field recording systems, which must be flexible and robust in order to accommodate the diversity of archaeological data, represent one such application. FAIMS project software facilitates this type of development by separating the (large and complicated) application code from the (relatively simple and largely human-readable) document files that customise the application for use by a particular project. Distributed version control systems like GitHub, which are already being used for texts and documents beyond code, provide a capable platform for coordinating peer production of these definition documents. FAIMS has used GitHub successfully for its internal development of early-adopter field projects over the last year, demonstrating its potential. Just as open-source approaches have improved software by bringing the insights of an entire community to bear on difficult problems, field recording systems - as well as the methods and approaches they embody - also benefit from the transparency provided by wide distribution and collaboration facilitated by version control systems.

8.2 FAIMS: Overview and History of the Project

Our perspective on the fitness of open-source approaches to archaeology reflects the authors' experience leading the Federated Archaeological Information Management

Shawn Ross, Brian Ballsun-Stanton, Adela Sobotkova: The University of New South Wales, Sydney, Australia
Penny Crook: La Trobe University, Melbourne, Australia

Systems (FAIMS) project for the past two years[1]. The purpose of the FAIMS project has been to develop discrete, federated mobile and web applications for the creation, refinement, archiving, and dissemination of digital data. To date, FAIMS has been led by the University of New South Wales, Sydney, in collaboration with participants from 40 organisations, including universities, archaeological consultancies, and heritage agencies in Australia and overseas[2]. During 2012 and 2013, the project was funded by the National eResearch Collaboration Tools and Resources (NeCTAR) initiative - an Australian government grant program tasked with building digital infrastructure for Australian researchers (http://goo.gl/eq4FhU). NeCTAR eResearch Tools provide sector-wide, collaborative, and accessible research software; all NeCTAR-funded projects were encouraged to reuse existing tools where possible and develop new tools as open-source software. Consequently, we joined existing open-source projects for a data refinement web application (Heurist, developed at the University of Sydney) and an online repository (the Digital Archaeological Record, administered by Digital Antiquity). Since no software for field data collection on modern mobile devices existed that met the needs of our stakeholders, we also initiated our own development of an AndroidLinux mobile data collection platform. Development has continued in 2014 thanks to funding from the Australian Research Council's Linkage Infrastructure, Equipment and Facilities (LIEF) scheme (project number LE140100151), which supports cooperative initiatives to develop expensive infrastructure for higher education researchers (http://goo.gl/8v1Iv2). LIEF funding has continued earlier activities. It will also allow us to extend interoperability to additional online data services (Open Context at UC Berkeley and OCHRE at the University of Chicago) and support construction of a portal for research access to Australian state heritage registers through a partnership with the University of Queensland.

The mobile data collection platform was the only component that we decided to build from scratch, and is the focus of this paper. Recognising the challenges of producing such a system, the FAIMS project undertook extensive stocktaking from June to August 2012, which included online surveys and a three-day workshop attended by as many as 80 archaeologists and developers. Subsequently, from September to December 2012, we undertook an extended technical elaboration with our development partners. The elaboration phase sought to determine the technical feasibility and preferred approach to the requirements generated during stocktaking. The stocktaking and elaboration process demonstrated that a static data logger was unlikely to be widely adopted, even if it could be customised and extended to a degree (Agreed Standards Report 2012, 7) - a conclusion supported by 15 years of precedent

1 A brief introduction to the FAIMS project is provided here; for more information about the project's history, including stocktaking, elaboration, and a discussion of component tools and services, please see Ross et al. 2013, p. 107–119, and http://www.fedarch.org/.

2 In January 2015 the FAIMS project will relocate to Macquarie University, Sydney, Australia.

in archaeological mobile software development (Ross et al., 2013, p. 108–109). Instead, we opted to solve the general problem of collecting idiosyncratic data using variable workflows during fieldwork (cf. Kansa et al. 2010, p. 308). NeCTAR-funded development produced the first public release of the mobile platform (v1.3; October 2013). Subsequently, LIEF-supported development in 2014 has prioritised improving the mobile platform, informed by deployments at archaeology and geoscience research projects and field schools in Australia and overseas (v2.0 is scheduled for release in late November 2014).

The mobile platform consists of a Linux server and native Android 4.1+ mobile application built around a generic database management system (SQLite) with geospatial extensions (SpatiaLite). It also incorporates other open software, standards, and protocols where possible (e.g., XML, OSGeo libraries, GNU tools, GeoJSON-LD)[3]. Designed as an archaeology-specific tool for the collection of well-structured digital data in the field and laboratory, the platform incorporates many of the features requested during stocktaking: offline capability, mapping and GIS functionality, multimedia integration, versioning, synchronisation, backup, and sophisticated data validation and automation, some of which are not supported in generic mobile databases or GIS packages. We also use well-established approaches to localisation borrowed from the IT industry to promote semantic, as well as syntactic, data interoperability. Most importantly, the software developed by FAIMS is community-driven, and can grow and adapt in response to the needs of archaeologists in the future (cf. Ross et al. 2013, p. 111–116).

The mobile platform is flexible enough to accommodate archaeologists' idiosyncratic needs and practices. The heart of the system is an interpreter that parses a set of XML documents and a beanshell file (together constituting a 'definition packet') to build fully customised data schemata, user interfaces, local vocabularies, and operational logic on Android devices. This packet defines what data need to be collected, in which format, and with which interface. Customising it to fit different research agendas and workflows requires about as much effort as creating a web-enabled database. Although it is not as easy to deploy as a static data logger, it accommodates the longstanding diversity of archaeological research agendas, methods, and field procedures. All data produced using the platform benefits from a robust but flexible underlying datastore[4], while supporting a wide range of recording systems. It is a tool that helps archaeologists build their own data collection tools.

3 Peer-to-peer wireless networking on Android proved unreliable, requiring a server for project creation and synchronisation.

4 In an append-only entity-key-value datastore modeled after google's protobufs. For technical information about the FAIMS mobile platform, see Sobotkova et al. 2014

8.3 The State of Play: Sharing in the World of Archaeology

8.3.1 Archaeologists and Open Source Software

From the beginning, the FAIMS project has been committed to developing open-source software and introducing open-source approaches to the archaeological community - an unfamiliar subject often met with indifference. As part of the stocktaking exercise, FAIMS circulated a Digital Data Survey amongst 150 members of the FAIMS community (Sobotkova 2013; Ross et al. 2013, p. 111–112). The survey was aimed primarily at Australian archaeologists and focused on their information management practices and attitudes. The professional background of participants was divided between academia (41%) and the private sector (37%). Given the survey's IT focus, the pool of 79 Australian respondents was likely self-selected from the IT-friendly or IT-savvy population. In the survey we asked about preferences for commercial or open-source software, and the most common response (45%) was: 'I don't care'. Almost the same percentage of respondents (42%), however, expressed the desire for open-source tools, while only 13% asked for a commercial product. The number of 'don't care' responses to the open-source vs commercial survey question may indicate that a large number of archaeologists - including the tech-savvy - do not appreciate, or do not understand, the characteristics and potential advantages of open-source approaches to software development.

Over the course of FAIMS project, we have continued to encounter this unfamiliarity. When promoting the benefits of open-source software at the Computer Applications in Archaeology 2013 Conference in Perth, we received some apprehensive reactions to the effect of: 'I don't want to use open source because then I would have to share all my data'! 'Open source' had been conflated with 'open access'; both were interpreted as signifying the imperative to share data without restriction. FAIMS does encourage open licensing of data (CC-0 or CC-BY-SA), because open data is likely to be more valuable and consequential (cf. Kansa and Bissell 2010, p. 42). Individual users, however, fully control the accessibility and licensing of data collected, processed, or archived using FAIMS software. In the FAIMS online repository, for example, they can openly license their data, or keep it entirely private. Data can be embargoed for a specific length of time, or access can be restricted to a specific group of users. This distinction between FAIMS software (distributed free and open-source under a GPLv3 license), and data created, processed, or stored using FAIMS applications (availability and licensing determined by user) requires frequent reiteration in our outreach programs.

Few archaeologists are programmers, and IT literacy in the discipline lags behind many other social-science and science disciplines. Archaeologists' experiences with data-collection software comes mostly through the use of commercial products like MS Office, ESRI ArcGIS, or FileMaker Pro (see 'Commonly Used Programmes' in Sobotkova 2013). The majority of academic archaeologists, as well as those at larger consulting firms, have access to institutional licenses for this software. To such users, most software is 'free', so they may be less concerned by the cost of commercial software, as well as unaware of the non-monetary advantages of open source.

Open source approaches, nevertheless, should be accessible to archaeologists. They have many parallels to the academic pursuits. As Lerner and Tirole (2005, p. 31) observe:

> "The most obvious parallel relates to motivation. As in open source, the direct financial returns from writing academic articles are typically nonexistent, but career concerns and the desire for peer recognition provide powerful inducements[5]."

Not only are incentive structures similar between academia and the open-source software world, but in practical terms academics are often well positioned to make small contributions to culumative, distributed projects. A number of open-source applications with roots in academia have matched or surpassed commercial software, especially in the sphere of analytical tools like qGIS (a geographic information system) and R (a statistical software package).

Niche tools are even beginning to emerge in archaeology, such as an archaeology-flavored Linux (http://goo.gl/UqbVpQ) preloaded with useful applications. Perhaps more importantly there are now web applications like Heurist, Open Context, and tDAR that have not only been developed using open-source software (MySQL for Heurist and Open context; PostgreSQL for tDAR), but are themselves distributed under open-source licenses. Each of these applications, furthermore, strive to exemplify the core open-source idea of a single tool doing one thing well. Heurist excels at data refinement, Open Context at sharing and dissemination of data, tDAR at long-term archiving of legacy data. The FAIMS project continues this approach, by contributing to the development of existing tools like Heurist and tDAR while building additional discrete tools such as mobile applications for data collection.

Open-source approaches to software have great potential for relatively small fields like archaeology and cultural heritage management. Where resources are limited and distributed, community-driven development may provide the only viable route to the production of robust and resilient software tailored to our discipline. Especially since the emergence of online software collaboration tools, peer-based development can coordinate many smaller efforts distributed across organisations and individuals to achieve a particular outcome, often by building single-purpose or narrowly-focused tools that work together through shared standards.

5 For an earlier examination of the parallels between academia and open-source culture, see Raymond 2000*a*, esp. 'Acculturation Mechanisms and the Link to Academia' and 'Gift Outcompetes Exchange'.

8.3.2 The Ethos of Sharing in the Archaeological Community

Although a plurality of Australian archaeologists are ambivalent towards open-source software, the archaeological community, generally, is 'open' to sharing not only data, but also the means of collecting it. Data exchange, however, is currently hampered by inefficient practices, many of which could be improved using the tools of the open-source world.

One question in theDigital Data Survey asked about specific attitudes towards the sharing of primary data[6]. The majority of survey respondents (90%) were open to sharing primary data, with some strings attached. While 20% were willing to share data without restriction (even before their own publication of that data was complete), 46% were willing to share only after they had finished their own publication, and another 24% wanted to restrict sharing to selected persons or groups. Only a very small fraction of respondents (5%) was averse to data sharing at all, while an additional 5% noted that they were prohibited from sharing by their employer. Overall, archaeologists' principal concerns centred on the ability to embargo data until after its originators have published their own interpretation, but indicate a generally positive attitude towards sharing primary archaeological data.

A second survey investigated the origins and transmission of core archaeological concepts (Softley, 2013). This survey included a question about the production of data collection forms, responses to which indicate that a great deal of sharing is already taking place. Some 44% of respondents 'borrowed' or 'adapted' existing forms, while 38% created their own recording forms from scratch and the remaining 18% were not involved in form production at all[7]. More contract archaeologists (40%) than academics (15%) reported that they borrowed or adapted existing forms, while 30% of both contract and academic archaeologists reported creating their forms from scratch (see Table 8.1). The limited degree of reuse in academia was somewhat surprising. We expected that sharing forms - either informally, online, or through publication - would be commonplace, since projects (especially surface surveys) often publish their recording forms in print or online as appendices to their reports and discuss their methodologies at length as part of publication (see, for example, Broodbank and Kiriatzi 2003).

This lower-than-expected degree of sharing amongst academic archaeologists supports Fred Limp's contention that '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

6 Question 30: 'Which best captures your attitude to sharing your primary dataset pending ethical clearance?'

7 Forty-seven contract and 40 academic archaeologists replied to this question, with 14 students, nine government employees and seven others.

Table 8.1: Do you design or manage recording systems? If yes, consider your key fields and attributes. Which of the following statements best describes your situation?

	Number of respondents	Yes: I use or borrow from systems developed elsewhere	Yes: I create new recording systems	No: I am not involved in recording systems design
Total	117	38	33	46
Contract	47	19	14	14
Academic	40	6	12	22
Government Employees	9	3	2	4
Students	14	5	5	4
Other	7	5	0	2

categories' (2011, p. 277). Only a small minority of academics involved in the design of recording systems reuse or adapt existing forms, an observation that likely carries over to recording methodologies more broadly.

While sociocultural factors like those identified by Limp contribute to continuous reinvention of recording methodologies and forms, the lack of a useful (and widely-used) platform for exchange is also a hindrance. As with archaeological data, print publication or personal communication remain the principal means of exchanging the tools for data collection. Since most archaeological recording takes place on paper or using customised spreadsheets, geographic information systems, or databases (or, usually, some combination of these tools), sharing of data collection methodologies is a hit-or-miss, ad hoc affair.

8.3.3 Creating and Sharing Repurposable Digital Data

The production of clean, well-formed data is a prerequisite for effective data sharing and efficient data analysis. Well-formed relational data, like that described by Codd (1982), is granular, avoids both redundancy and sparseness, protects data integrity, and better accommodates unexpected data by systematically dividing data into linked ('related') tables[8]. Most relational database management systems (DBMS) also separate the database's 'back end' (the tables containing the data) from its 'front end' (the forms and reports through which users see and manipulate data), helping to preserve the integrity of data and avoid accidental changes to the data and other errors.

Well-formed relational data produced by a DBMS that separates data from a interface is fine-grained, regular, and compact. It is also robust, in that data can be ma-

8 Other 'NoSQL' approaches like graph or native XML databases may also yield well-formed data; relational data is used here as an example due to its relative familiarity, and because the FAIMS mobile platform employs a highly normalised relational datastore on account of technical constraints inherent to mobile and GIS development.

nipulated, analysed, and presented variably and repeatedly without damaging it. Not only is such data more intrinsically valuable because of its granularity, consistency, and integrity, but computers can reliably parse it - factors that greatly facilitate effective data sharing and reuse. Granularity and machine-readability facilitate 'loose coupling' approaches to data sharing (Kansa and Bissell, 2010) and are required for more ambitious attemps at syntactic and semantic interoperability.Like researchers in other 'small sciences', however, many archaeologists are accustomed to asking only their own research questions of their data, and fail to consider how data might be re-purposed by others in the future (Kansa and Bissell, 2010, p. 42). Producing well-formed relational data requires time, resources, and expertise. It involves data modelling, the instantiation of the model as an effective database, and at least basic programming ability for form behaviours and validation, all of which are specialised skills that relatively few archaeologists have acquired (cf. Sobotkova 2013, table 1). Often, the increased initial cost and effort to produce well-formed data are not considered worthwhile.

Instead, archaeologists tend to use office productivity software like spreadsheets that are quick and easy to deploy, or they build bespoke systems using more sophisticated desktop database or GISsoftware familiar to them from other aspects of their work. In the FAIMS survey, 98% of respondents reported using spreadsheets (mostly MS Excel) and 81% reported using GIS software (mostly ESRI ArcGIS). While 87% reported using relational database software (most commonly MS Access), only 30.4% reported an ability to build databases. Frequently, archaeologists combine several of these tools with extensive paper recording (Sobotkova 2013, p. 6-7; cf. Kansa and Bissell 2010, p. 42–44).

Use of these familiar software packages, however, often impedes the reuse of data. As noted above, almost all archaeologists use spreadsheets, but 'flat' datastores have a number of drawbacks. Human-readable spreadsheets are often difficult to manipulate programmatically (a requirement for genuinely repurposable datasets), since they commonly lack basic data standards: cells often contain more than one value, more than one data-type is stored per column, data is duplicated in multiple columns, data becomes sparse as a spreadsheet expands to accommodate rare multiple instances of some phenomenon, or records spill across rows in unpredictable ways (intuitive to people but opaque to machines). Some of these problems can be mitigated through good spreadsheet design, but difficulty and likelihood of failure increase as data becomes more extensive and complex. Relational databases, in contrast, address these and similar problems structurally and systematically.

To take another example, most archaeologists also use ESRI ArcGIS (52% of the FAIMS Survey sample; cf. Sobotkova 2013, p. 7), primarily for mapping and spatial analysis, but with data collection performed using its mobile component, ArcPad. The problem is that even though ArcGIS is built around a powerful relational DBMS (MS SQL Server),in its default configuration it stores data in a single large table rather than as relational data. It is difficult and time consuming to design and implement a properly-structured SQL Server relational database that also performs well within ArcGIS, since doing so requires mastery of two complex software packages as well as their interactions with one another. As a result, most archaeological ArcGIS geodatabases are not relational, suffering from the same limitations as spreadsheets. Other mobile GIS packages and data collectors used by archaeologists, including GIS Pro and Open Data Kit, also produce flat data.

Properly structured, robust databases can of course be built using commercial or open-source DBMS products, ranging from MS Accessand FileMaker Pro to MS SQL Server, MySQL, PostgreSQL, SQLite, or even Oracle. Most of these products can also be used as data sources for commercial or open-source GIS or statistical software. Bespoke databases, however, face significant challenges. Desktop DBMSes are generic in nature, developed without regard for the particular needs of archaeology as a discipline or fieldwork as a practice. As a result, they require the ex nihilo construction of properly designed data structures, interfaces, and logic (form behaviours, validation, etc.). Individual deployments may or may not be well designed and executed. Much effort is also duplicated. Many projects re-create databases for common archaeological activities from scratch. Even when projects share databases, they often painstakingly rebuild them to address small variations in what are otherwise similar workflows, a costly undertaking considering that desktop databases like MS Access are not designed with coordinated redeployment in mind (e.g., once a database has been 'cloned' and populated with data, any improvement in the new database will likely have to be recreated by hand in the original). In all cases, bespoke databases require money, time, and expertise to build well, test thoroughly (a step omitted by many), and maintain - usually more than was initially thought.

Even if the necessary resources are expended and a project's database is well constructed, further knowledge and planning regarding online distribution and interoperability is required to avoid trapping data on a researcher's hard drive or a destination website in a form that is hard to locate, strictly human-readable, and ill-suited for automated reuse (Kansa and Bissell 2010, p. 43–45; cf. Blanke and Hedges 2010). Efforts are underway to improve the quality of archaeologists' databases in this regard, promoting the production of syntactically interoperable data (e.g., through XML export; cf. http://www.codifi.info/, and Ashley et al. 2011). Semantic interoperability, however, remains difficult to attain in bespoke systems, especially in light of the fact that archaeology lacks widely shared data standards, conceptual vocabularies, or ontologies.

In short, many distinct challenges face the archaeologist who wants to produce reusable and repurposable archaeological datasets that can be deployed to answer new and unanticipated questions. Most archaeological data is only partly digital. When it is digital, it comes in a variety of formats, most of them unstructured. Even structured data is often not well-formed for computerised reuse. If it is well-formed, then too frequently data is housed in a silo, making it difficult to discover and extract in a machine-readable form. Overall, the data generated by most projects is of limited utility; it cannot be easily discovered, retrieved, re-analysed, and repurposed.

8.4 Open Source Beyond Software

8.4.1 Free-as-in-beer and Free-as in Speech: Open Source Paradigms for Scholarship

The initial allure of open-source software is that it is 'free'. Stallman (2012) differentiates two types of free when it comes to software: 'free as in beer' and 'free as in speech'. open-source software is not necessarily free as in beer, but it should always be free as in speech. Lessig (2000) illustrates the matter with an analogy between code and law:

> "Ours is the age of cyberspace. It, too, has a regulator. This regulator, too, threatens liberty … this regulator is code - the software and hardware that make cyberspace as it is. This code, or architecture, sets the terms on which life in cyberspace is experienced."

Raymond 2012 elaborates two particular dangers of closed-source software. The first is 'agency harm': 'closed-source software puts you in an asymmetrical power relationship with the people who are privileged to see inside it and modify it. They can use this asymmetry to restrict your choices, control your data, and extract rent from you'. The second is 'lock-in harm': 'Closed source increases your transition costs to get out of using the software in various ways, making escape from the other harms more difficult'. Proprietary and open-source development paradigms embed particular social and philosophical outlooks into software production, producing divergent results that are more far-reaching than the monetary cost of the software itself.

A revolutionary idea motivates open source: we have the right to see and alter that which controls our lives. In the first instance, that right extends to software; the code that regulates cyberspace should be free-as-in-speech - open, available, and alterable. This principle, however, can be extended beyond software. Many people reduce 'technology' to its products: ever more dazzling gadgets, or perhaps the online services that are becoming more and more ubiquitous. Technology, however, is better thought of as the tools and techniques people use to manipulate the environment, all operating within the constraints of implicit or explicit 'regulators' analogous to Lessig's code[9].

> "To the extent that scholarship is the creation and curation of human knowledge, scholarship is an open-source endeavor. The end product - human knowledge - is not a fixed product, it is

9 For a fuller discussion of the social construction of technology, see Ballsun-Stanton and Carruthers (2010).

> distributed, has diverse manifestations, and belongs to no individual or entity. Some scholarship involves the creation of new theories, systems, or tools. Some involves the repurposing of existing theories, systems, or tools for another domain. Some scholarship involves synthesis. Some involves critique. It always involves accessing the work of others in order to (re)build something that will enter public discourse (in other words, 'publish'). And no matter how isolated the work, no matter how selfish the motivations, no matter how ignored the results, ultimately scholarship belongs to the human community."

Scholarship (theories, methods, and practices) is, in this sense, code; our results and interpretations (knowledge) are its output. open-source approaches, moreover, declare that we should and can share and modify our methods and approaches collaboratively, as if they were code. Such peer-based production - continuous sharing, borrowing, changing and adapting - is analogous to traditional academic practice in many ways, but when realised systematically using open-source approaches and tools, it marks a revolutionary shift that improves research by making assumptions explicit and interrogating authority.

8.4.2 The GitHub Revolution

GitHub (http://github.com/) is emerging as one of the most important tools for peer-based production. It is a web-based hosting system for code (and other text) that emerged from the 'distributed version control system' (DVCS) known as Git. In Git (http://git-scm.com/) and its contemporaries, Mercurial and Darcs, do not recognise a single, true code 'repository' (a project container). Instead, every copy of a repository is equally valid. Repositories can interact with one another. If you want to work on code from another repository, you can 'fork' that repository - copy its code at a particular point in time. Copied ('cloned') code becomes your own; you can then modify the code in your 'downstream' repository as you wish. If another repository makes incremental changes that you want to incorporate into your work, you can 'pull' them into your own repository. If you want to share your own changes with another repository (usually the 'upstream' one), you can file a 'pull request' with them, which they may or may not 'commit' (if the upstream repository does not commit your code, you simply continue to host a divergent fork of that repository). Each repository evolves independently, but code may be shared at will. GitHub's innovation lies in providing a technical platform for easily sharing and tracking code changes online.

Instead of requiring a central authority's approval for each change to source-of-truth master repository of code, a distributed, spontaneously ordered community replicates, modifies, and shares code. The code becomes more free-as-in-speech. Apparent anarchy is resolved not through a leviathan of centralised authority, but through a democratic process of use. Hosted repositories that make good decisions become popular; they are cloned widely, used frequently, and accrue some authority. Pull requests accepted into these repositories bring particular status to the contributor. If a repository declines to commit your changes and you continue to host your fork, you still contribute to the community by offering choice - and may attract a following. The failure to commit valuable pull requests invites popular rebellion, where users defect from one repository to another. Repositories that are inactive or unresponsive, or serve only limited needs, are left in obscurity.

Individual contributions are recorded and reputation matters; all changes are 'owned' - carefully tracked and attributed. For example, in our own 'faims-android' application repository, GitHub automatically cites Eric Frohnhoefer as the creator of the spatialite-android codebase that we use as part of our mobile GIS, simply because we pulled that code into our repository and committed it (http://goo.gl/2At0Q9). Despite the fact that he may not even know of our project's existence, he is credited with 20 commits. GitHub produces a social community where standing is established when your code is pulled and committed in this manner, whether you are aware of it or not. The process is analogous to academic citation, but more automated and nuanced.

GitHub Beyond Software

Although GitHub was initially developed as a collaboration platform for software, it has become a leader in peer-based production of all sorts (Rogers, 2013). The City of Chicago, for example, has posted street location, building footprint, bike route, pedestrian route, and bike rack locations on GitHub and encouraged users to improve it (Chicago Digital, 2013). Lawyers are now using GitHub to distribute and improve legal documents (e.g., McMillan 2013; *Series Seed* 2013; *SeriesSeed / Equity* 2013). Any information amenable to a cycle of publication, distributed improvement, and re-publication can benefit from GitHub's peer-based production model. It has even been applied to university courses (http://goo.gl/Nl2OB0) and PhD dissertations (http://goo.gl/d1UyWJ).

Shaffer (2013) argues that GitHub has great potential for scholarship and research:

> "Though not designed specifically for academic use, GitHub is designed with text, sharing, collaborating, and freedom in mind. For those looking to 'hack' existing work, to offer their own materials for others to hack, to collaborate with others, and particularly to do so with websites, software, or complicated text resources, GitHub is an amazing resource. And due to its social, collaborative nature, it is a resource that is consistent with the ideology of liberal education, and will grow in utility the more our academic communities make use of it."

This potential lies in overcoming barriers to collaborative development. Forking, pushing, and pulling processes work to disaggregate and re-aggregate ideas; all that is useful in an upstream text can be retained, while specific improvements can be made. Related repositories can incorporate those 'improvements' (or not) and make their own incremental changes. Instead of bundling hundreds of ideas in a journal article, or being forced to run an entire study to suggest a change to some small element of a methodology, academics can now treat their 'texts' (methods, approaches, interpretations, etc.) as code, incorporating and contributing incremental improvements to the overall body of ideas - with, of course, full credit for the originator, since recognition and reputation work as powerful incentives in both the academic and open-source worlds (cf. Raymond 2000*b*).

8.5 New Applications of Open Source Techniques: Building, Sharing, and Improving Field Recording Systems

8.5.1 Open Source Approaches to the Development of Recording Systems

The remainder of this article explores how open-source principles inform our approach to implementation of mobile data recording software at individual archaeological projects. In particular, our mobile data collection software lends itself to sharing and improving field recording methods and practices themselves - not just the underlying software - using distributed, peer-based production.

As discussed above, static data loggers are ill-suited to the needs of the archaeological community, while existing software used by archaeologists does not foster the production of reusable and repurposable datasets. Instead, our mobile device platform is built around an Android interpreter that can instantiate a wide range of data models and workflows and still produce well-structured data.

This approach, however, leaves some problems of implementing data management systems unsolved. The danger of wasteful duplication remains, data and workflows at particular projects must still be modeled, and the production of data schemata and UIs based on these models still requires time and expertise. Despite the fact that FAIMS is comparatively well-funded, we lacked the resources to develop a GUI for module design (which would have doubled mobile development costs). Instead, implementation is accomplished through definition packets. The use of definition packets allows deep customisation of recording systems, but is far less costly to implement (and does not preclude later development of a GUI). By separating the data schema from UI and logic scripts, moreover, we can deliver different interfaces atop the same data models. Finally, the use of definition packets allowed us to explore open-source solutions to implementation problems such as obstacles to sharing and consequent duplication of effort.

The fact that the FAIMS interpreter and renderer are themselves open source is of limited utility for archaeologists, as the software is complex and few will have the expertise to contribute to its development. Of greater importance is the fact that archaeologists can develop the definition packets, which are much simpler, using open-source tools and approaches. The architecture of the platform separates the underlying software from the description of the recording system more completely than is the case with generic database management systems. As a result, FAIMS implementations (customised schemata and UIs) are more portable. Our approach allows recording systems - as well as the methods and approaches that underlie them - to be shared and modified like code. The use of definition documents for customisation combines with peer-production tools like GitHub to allow efficient, distributed, and cooperative development of redeployable archaeological recording systems.

FAIMS definition packets are placed on GitHub in an open repository under a GPLv3 license. They are free to download, adapt, and deploy, so long as the resulting modified packets are distributed in the same way. Over time, a growing range of definition packets can emerge, each building on the others using GitHub's ability to fork code and pull changes. To start this process, FAIMS has established a library of definition packets. Over the past year, we built and refined modules for excavation, survey, and geosampling, based on our experience supporting a range of field projects.

As an example, in 2013 we created the FAIMS Excavation module for single-context, multi-trench excavation (http://goo.gl/z7Cq3O). It is informed by a detailed comparison of 11 excavation recording sheets submitted by FAIMS partners, using core definitions derived from the Museum of London Archaeological Site Manual (1994). Over the course of 2014 FAIMS field deployments, this module has been adapted for three major research excavations: Proyecto Arqueológico Zaña Colonial (PAZC; an early Colonial project in Peru), the Malawi Earlier-Middle Stone Age Project (MEMSAP), and the Boncuklu Höyük Project (a Neolithic tell in Turkey). Each adaptation was different, responding the needs of the project. PAZC required a full translation into Spanish, with some minor alterations to attributes. Boncuklu required significant localisation of the recording schema and UI to mirror existing paper forms to enable continuing for a long-standing project. MEMSAP also included significant adaptation to accommodate project idiosyncrasies, but also stripped the module of its multiple context types and introduced complex validation to ensure quality control. These significant feature improvements would have been more difficult and expensive without the common basis and a version control system. When we developed features of common interest, we merged them back into the 'Master' excavation module.

Figure 8.1 shows a network graph of our 'Excavation' repository, with four branches corresponding to the three projects plus a 'Master' (live at: http://goo.gl/0Z4zUh). The lines diverge and converge as changes are made to each branch of development, with desirable changes shared across branches and re-committed to the 'Master'. Although this figure represents branches within a single repository, interaction across repositories works similarly. Three versions of Excavation result, for use in different contexts, plus an evolving Master Excavation incorporating shared characteristics. This internal development is seeding the ecosystem with a variety of definition packets. As time goes on, it will become more likely that any given project will find a packet closely suited to its needs.

Conceptually, this approach should be familiar, since archaeologists already borrow and adapt paper forms. Compared to haphazard sharing of paper forms, however, the infrastructure of open-source software improves discoverability, reduces duplication, and facilitates the mechanics of sharing. Creating and publishing new record-

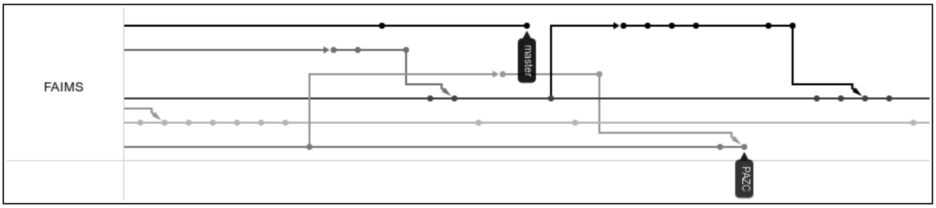


Figure 8.1: Network graph of the FAIMS 'Excavation' repository, with four branches: PAZC (orange), Boncuklu (purple), MEMSAP (yellow), and the original, 'Master' branch (black)

ing systems, editing existing versions, tracking changes, and importing desirable improvements made by others all become more systematic and transparent. The incentives common to open source and academia also come into play: since the entire GitHub process is monitored and displayed, it is easy to see who is using whose packets, with the most popular packets reflecting well on their creators. It is even possible to fit this model more directly into academic settings, as packets that reach a certain level of adoption may be subjected to expert peer review using an approach analogous to the Journal of Open Archaeology Data (http://goo.gl/MunVjc).

A platform for sharing can build a community. As they create and modify packets, archaeologists can help one another by including detailed metadata or annotations, making it easier to determine a packet's applicability to another project. Such metadata might include a theoretical or methodological considerations that influenced the recording system, data models and schemata, representations of workflows, UI screenshots, and other useful information. If the production of such metadata can be systematised, it would foster rigorous practice. Unlike trading paper forms, the metadata and annotations attached to definition packets in a GitHub repository capture research design: field recording workflows and data models become transparent, revealing much about the methods and practices used on any given project - an outcome that would contextualise the data and interpretations produced by that project, further encouraging reuse, repurposing, and reinterpretation (Huggett, 2012, p. 541–542). Such a process has a great potential to improve the self-awareness of archaeologists and the rigour of archaeological practice.

8.5.2 Improving Sustainability through Reuse and Redeployment

Evolving definition packets will facilitate and systematise the informal practices of archaeologists. Currently, archaeologists often base their own hard-copy recording forms on published models, like those presented in reference works or the appendices of reports (Snow et al., 2006). With proper citation, the same sources could also inspire definition packets, but improvements or customisation would not be limited to a single project. Instead, changes would immediately become available for reuse and further development elsewhere. GitHub supports both ‘bug tracking’ and ‘code review’; in this context the former would allow errors or omissions to be flagged while the latter provides an ongoing discussion about contentious aspects of a recording system and suggestions for improvement. Analogous processes go on now with the sharing and improvement of forms, but they could be automated and opened, sharing benefits and reducing duplication.

As with open-source development of core software, this approach to the design of definition packets fosters the sustainability and uptake of the FAIMS ecosystem. Peer-based production through GitHub spreads the burdens of development, encouraging the improvement of existing implementations and facilitating the production of new ones while avoiding duplication. GPLv3 licensing allows reuse, but requires that modifications be distributed under the same license, so that improvements remain available to the community. Modifications of existing packets can be undertaken within the scope and budget of even small projects. As the library grows, the likelihood of finding a close match to any particular project’s needs will increase, reducing the time and cost of deployment - a development critical to the sustainability of the ecosystem. Increased uptake, and the associated generation of ever more nuanced variations of definition packets, perpetuates the cycle. The declining costs of deployment associated with a growing library of definition packets differentiates re-deployable systems like FAIMS using open-source infrastructure like GitHub from bespoke production of databases. While desktop databases can always be copied, no technique comparable to use of a distributed version control system like GitHub exists for managing varied adaptations and re-incorporating useful improvements made by others.

8.5.3 Improving Archaeological Practice through Dataset Interoperability

To this point, we have discussed the application of open-source approaches and the GitHub platform to the development of definition files used to instantiate an archaeological project on FAIMS field recording infrastructure. One of those files is worth special attention: the ‘localisation’ document for mapping project-based terminology to a core vocabulary of concepts (and other acts of translation).

Facilitating the creation of interoperable datasets constitutes the overriding goal of the FAIMS project. Such datasets are required for reproducibility, reinterpretation, and comparative and large-scale study in archaeology, yet mechanisms for producing them have been slow to emerge. Considering the diversity of archaeological data, and the idiosyncrasies of archaeological practice, no widely-shared core data standards are likely to be adopted by the archaeological community in the near future. Preliminary research conducted for the FAIMS project by P. Crook, however, indicates that within archaeological sub-disciplines, some 70% of project-specific terms should be mappable to a core concept vocabulary for field excavation. Mapping of data to master ontologies facilitates production of compatible datasets, and ’ontology mappers’ have been built into repositories like tDAR. Mapping at the end of a project, however, when data is ingested into a repository, is expensive and time-consuming. Furthermore, there is some risk in mapping terms after recording, if the definitions used during data creation are misconstrued during the mapping process.

FAIMS has sought to address these problems by building concept mapping into data creation, using techniques borrowed from software localisation, a process by which (for example) a web site's menu or a product's UI is automatically displayed in a local language (the FAIMS localisation document has, in fact, been used to translate the Andoid application's UI between English and Spanish). The FAIMS mobile platform can map a 'local' archaeological term to a 'global' or 'core' concept, with the user always seeing the local term but the data automatically associated with the core concept. The terms 'context', 'locus', 'spit', and 'unit' could all, for instance, be mapped to a core concept of 'stratigraphic unit' (this core concept, can also be annotated with open linked data URIs in the data schema). Concept mapping is encapsulated in a human-readable, plain-text document within the definition packet.

Peer-based production could contribute to improving the core-concept lists (and, eventually, ontologies) embedded within the localisation document, fostering the production of compatible datasets. The localisation document can be developed using GitHub in the same manner as the other files of the definition packet. Direct community engagement with ontology production may increase buy-in and increase the likelihood of wider adoption of shared ontologies, thereby advancing the overall goal of producing interoperable archaeological datasets[10].

8.6 Conclusion

Fred Limp identified the problem of 'polemical differentiation' as a disciplinary incentive in archaeology (2011, p. 277). In many ways, the world of proprietary software with its operating system and browser wars has faced analogous problems. open-source models offers an alternative built around a paradigm of peer production that esteems collaboration and openness over the isolated cultivation of hidden, protected ideas and techniques. Under this model, competition is redirected away from battles between closely guarded, rival products. Instead individuals strive for the prestige of contributing to a community that benefits from, and values, helpful participation.

In addition to making software more free-as-in-speech, the open-source approach has, perhaps counterintuitively, increased the quality of software. The apparent an-

10 The scenario here involves using GitHub to share and modify local ontologies embedded in FAIMS definition packets used by particular projects. Eric Kansa (per. comm.) has suggested the use of GitHub to distribute and evolve proposed 'core' ontologies unrelated to particular systems, in order to advance data compatibility still further.

archy - perhaps better considered spontaneous order - of open-source development reduces complexity, corrects errors, and finds new solutions (cf., Raymond 2000*a*, esp. 'How Many Eyeballs Tame Complexity'). By exposing code and removing barriers to collaboration, many experienced eyes can take a fresh look at software. Individual contributors can make small, incremental, coordinated improvements that chip away at large and complicated problems - with appropriate credit given to every pair of hands wielding an axe.

Applied to archaeological research, open-source approaches can distribute development of costly and complex software amongst many organisations and individuals, each of which has limited resources but also particular strengths. Such approaches can also expose field recording systems - along with their embedded theories, methods, and practices - in order to improve both the systems and the underlying methodologies cooperatively. As such, open-source approaches enhance not only the data management software or field recording tools, but also the rigour of archaeology as a discipline.

Bibliography

Ashley, M., Tringham, R. and Perlingieri, C. (2011), 'Last house on the hill: digitally remediating data and media for preservation and access', *Journal on Computing and Cultural Heritage (JOCCH)* **4**(4), 13.

Ballsun-Stanton, B. and Carruthers, K. (2010), #c3t the command & control of Twitter: On a socially constructed Twitter & applications of the philosophy of data, *in* 'Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference on', IEEE, pp. 161–165.

Blanke, T. and Hedges, M. (2010), A data research infrastructure for the arts and humanities, *in* S. Lin and E. Yen, eds, 'Managed Grids and Cloud Systems in the Asia-Pacific Research Community', Springer, pp. 179–191.

Broodbank, C. and Kiriatzi, E. (2003), 'Archaeological survey: Methods and preliminary results'. URL: http://www.ucl.ac.uk/kip/survey.php

Chicago Digital (2013), 'Chicago on GitHub'. URL: http://digital.cityofchicago.org/index.php/chicago-on-github/

Codd, E. F. (1982), 'Relational database: a practical foundation for productivity', *Communications of the ACM* **25**(2), 109–117.

Huggett, J. (2012), 'Lost in information? Ways of knowing and modes of representation in e-archaeology', *World Archaeology* **44**(4), 538–552.

Kansa, E. C. and Bissell, A. (2010), 'Web syndication approaches for sharing primary data in" small science" domains', *Data Science Journal* **9**, 42–53.

Kansa, E. C., Kansa, S. W., Burton, M. M. and Stankowski, C. (2010), 'Googling the grey: Open data, web services, and semantics', *Archaeologies* **6**(2), 301–326.

Lerner, J. and Tirole, J. (2005), 'The economics of technology sharing: Open source and beyond', *Journal of Economic Perspectives* **19**(2), 99–120.

Lessig, L. (2000), 'Code is law: On liberty in cyberspace', Harvard Magazine. URL: http://harvardmagazine.com/2000/01/code-is-law-html

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

pp. 265–80.
McMillan, R. (2013), ‘Your startup’s legal docs: Now on GitHub’, Wired.
URL: http://www.wired.com/2013/03/series-seed/
Raymond, E. (2000*a*), ‘The cathedral and the bazaar’.
URL: http://www.catb.org/esr/writings/cathedral-bazaar/cathedral-bazaar
Raymond, E. (2000*b*), ‘Thomesteading the noosphere’.
URL: http://www.catb.org/esr/writings/cathedral-bazaar/homesteading/index.html
Raymond, E. (2012), ‘Evaluating the harm from closed source’, Armed and Dangerous.
URL: http://esr.ibiblio.org/?p=4371
Rogers, M. (2013), ‘The GitHub revolution: Why we’re all in open source now’, Wired.
URL: http://www.wired.com/2013/03/github/
Ross, S., Sobotkova, A., Ballsun-Stanton, B. and Crook, P. (2013), ‘Creating eresearch tools for archaeologists: The federated archaeological information management systems project’, *Australian Archaeology* (77), 107.
Series Seed (2013).
URL: http://www.seriesseed.com/
SeriesSeed / Equity (2013), GitHub.
URL: https://github.com/seriesseed/equity
Shaffer, K. (2013), ‘Push, pull, fork: GitHub for academics’, Hybrid Pedagogy.
URL: http://www.hybridpedagogy.com/Journal/push-pull-fork-github-for-academics/
Snow, D. R., Gahegan, M., Giles, C. L., Hirth, K. G., Milner, G. R., Mitra, P. and Wang, J. Z. (2006), ‘Cybertools and archaeology’, *Science* **311**(5763), 958.
Sobotkova, A. (2013), ‘The use of information technology in australian archaeology: the FAIMS digital data survey report’.
URL: https://fedarch.org/documents/DigitalDataSurveyReport.pdf
Sobotkova, A., Ballsun-Stanton, B., Ross, S. and P, C. (2014), Arbitrary offline data capture on all of your androids: The FAIMS mobile platform., *in* A. Tragvillia, ed., ‘Across Space and Time: Selected Papers from the 41st Computer Applications and Quantitative Methods in Archaeology Conference’.
Softley, C. (2013), A Culture of Inertia: The realities of content standards and the archaeologists who need them, PhD thesis.
Stallman, R. (2012), ‘Why open source misses the point of free software’, Philosophy of the GNU Project.
URL: http://www.gnu.org/philosophy/open-source-misses-the-point.html

Joshua Wells, Christopher Parr and Stephen Yerka

9 Archaeological Experiences with Free and Open Source Geographic Information Systems and Geospatial Freeware: Implementation and Usage Examples in the Compliance, Education, and Research Sectors

9.1 Introduction

Geographic information systems (GIS) have today become a recognizably standard series of tools within the archaeologist's kit. The continued proliferation of relatively-cheaper computing power worldwide has meant that the number of hardware, operating system, and software limitations to GIS utilization have been on a steady downward trajectory for the last quarter century. As these personal computer barriers to GIS usage crumbled, the general trend toward improved Internet connectivity (wired and wireless, and certainly not globally achieved) has permitted archaeological geospatial investigators to better share their finished products and base data within the professional community and with interested publics; also to allow archaeologists to avail themselves of growing stores of public and private geospatial data related to the earth and environmental sciences. The remaining and significant barriers to GIS usage in archaeology, however, are the costs and availability of GIS software packages themselves since the most popular GIS applications are proprietary, expensive, and bogged-down with license-based user restrictions that serve to limit the practical implementation of the software for legal reasons. These last two barriers are particularly pernicious for archaeologists, whose laboratories frequently do not have the highest funding priorities in even the best of economic times, and thus creates functional limits to the scope of GIS deployment among the archaeological community. These fee-and-license restrictions in the workplace have served to effectively hobble more thorough development of archaeological GIS and database training among universities and professional development outlets, which has helped to maintain the current climate in which archaeological GIS training and sensibilities remain somewhat esoteric even in a time of rampant technological changes throughout the sciences and humanities.

Fortunately, a potential solution to the fee-and-license barriers exists. This paper

Joshua Wells: Indiana University South Bend, South Bend, USA
Christopher Parr: Virginia Department of Military Affairs,Virginia, USA
Stephen Yerka: University of Tennessee, Tennessee, USA

demonstrates and evaluates the use of free and open-source (FOSS) GIS software in the pursuit of several projects related to heritage management and archaeological reconnaissance. The strengths and weaknesses of Quantum GIS (QGIS) as it pertains to several archaeological projects will be discussed in detail, along with briefer assessments of other FOSS and free-ish applications such as the User-Friendly Desktop GIS (uDIG), and Google Earth (which is definitively not FOSS, but freeware).

The projects to be discussed include the interoperation of landscape-scale spatial databases from heritage management offices in several US states, the management of archaeological projects on an American military base, and the use of software applications to promote GIS and general geospatial training in undergraduate education and research. This is not an exploration of GIS methods in archaeology, we assume most readers today are basically familiar with what GIS packages do (database mapping, spatial statistical analyses, layering of data, etc.); nor is this a "how-to" guide for GIS methods with FOSS softwares, although we do include references to a number of good training materials. Instead, particular emphasis will be given to the ways in which the different FOSS GIS software discussed can be integrated with a professional workflow in archaeological research, educational, and outreach settings. Through these examples, interested archaeologists with GIS experience will better understand the differences between their fee-and-license products and FOSS products, and archaeologists interested in GIS experience will have a reference base from which to enter the world of FOSS GIS training materials.

9.2 What Geographic Information Systems and Free and Open Source Software are not

Although we assume some basic familiarity with the products of GIS and the existence of some software packages, it seems prudent to briefly clarify what a GIS system generally is not and is so that the FOSS distinctions are more readily apparent in the examples below. A GIS is a software package that combines database behaviors with spatial representations in order to create, manage, and analyse spatial data. Conversely, and simply put, a GIS is not an ESRI product. Although ArcGIS (ArcInfo, ArcView, ArcMap, etc.) is a GIS, there is much more diversity in geographic information systems as a population of software than is generally recognized. To many people, even learned professionals, the term GIS is often (unfortunately) a shorthand for ESRI products (or some other license-restricted software such as Erdas Imagine, Manifold, etc.). This terminological imprecision about GIS is simply a feature of marketplace dominance, in the same way that many residents of the United States refer to cellophane tape as "Scotch" or disposable tissues as "Kleenex" the use of "GIS" is often considered ArcGIS use. Fortunately, there are a number of resources available with extensive listing of many license-restricted and open-source GIS softwares, past and present (Dempsey, 2012; OSGeo (Open Source Geospatial Foundation), 2013*b*; Sherman, 2012; Steiniger, 2013; Steiniger and Bocher, 2009; Steiniger and Hay, 2009; Steiniger and Hunter, 2013; Wikipedia, 2013).

The influence of ESRI over discussion of GIS is so pervasive that even the basic language of data types is beholden to ESRI: the shapefile, actually a set of three or more files that function together for the display and management of vector data (points, lines, and polygons), was developed by ESRI in the early 1990s for use with its ArcView product, but is used in practice as an open standard that is regulated by ESRI (ESRI, 1998); the similarly open ESRI Geodatabase standard (ESRI, 2013), with several different formats for spatial databases, is often referenced colloquially as simpler name for a spatial database.

Some important caveats, regarding the professional environment for GIS in which archaeologists participate, are worth stating: (1) although ESRI is often criticized (lovingly or otherwise) for its market dominance, it is worth mentioning that each author on this paper is also a highly-trained user and customer of ESRI and other proprietary products; and (2) market dominance is an important feature of the social landscape to which all archaeological practitioners must adapt. There are important functional differences in work processes that use FOSS GIS to complete finished products for dissemination, and those that use FOSS GIS to share processed data with distant offices that use other proprietary software. Each reader must analyse their own needs, resources, and choices in both the technical and social landscapes in order to determine the best course of action for the professional or institutional adoption of any new software (e.g., Sherman 2012, cf. Kling et al. 2005).

As discussed elsewhere in this volume, free and open-source software is also a frequently misunderstood concept among the general public, including the scientific and humanities communities (outside of the computer sciences, where they are then otherwise debated). The word “open” generally refers to the availability of intelligible source code and the rights of users to review and modify it. The “open” designation is frequently mistaken to mean “zero cost”. A popular adage in the open source community clarifies different senses of the term “free”, where “free as in speech” refers to human rights and liberties for expression, while “free as in beer” simply refers to goods or services available without a monetary charge. The free-software movement, fully cognizant of the labor-costs invested in creating software, emphasizes the “free as in speech” civil liberty aspects of open source code. In her ethnographic account of the open source movement, anthropologist Gabriella Coleman (2013, p. 200–205) demonstrates how open source programmers engage in a particularly effective form of cultural critique, particularly a critique of the commoditization of intellectual labor and expression through normative, highly-restrictive (crippling “free as in speech”) intellectual property practices. Differing definitions for the exact meanings of the terms free and open, hearkening to debates that began in the earliest days of free and open-source movements, continue to slowly proliferate today (e.g. Creative Commons 2013; Open Source Initiative 2013; Raymond 2000; Sherman 2012; Williams 2002). Throughout this discussion, we will clarify exactly the modes of "free" and "open" made available by the development communities for each software presented; this is done for the sake of clarity, and in recognition of the fact that because of often limited research budgets and personal penchant, archaeologists may be also interested in free beer.

9.3 What does Open Source mean in GIS

9.3.1 Pros and Cons

The particular software applications discussed here are by no means a complete representation of all available FOSS GIS offerings, but are the applications represented in the use case example sections below. Descriptions are provided for completeness, and we encourage potential users to investigate the documentation for any software named here. Any potential user of FOSS software (GIS or otherwise) should explore their work cycle and consider the strengths and weaknesses of each application in context. A tremendous benefit of FOSS software is that you can experiment with it to learn these details. It may also be useful to have a frank discussion with institutional IT personnel (if available) to ask what advice and support they can provide for laboratory and broader usage within the institutional framework (we have found our IT staffs to be both very interested in new FOSS software and helpful in solving conflicts before they arise).

9.3.2 Software Application Summaries

Quantum GIS (QGIS). This is a full-featured, graphical user interface, desktop GIS application. QGIS installs on Linux, Mac, or Windows operating systems (there is also a mobile Android application) with dozens of language translations for installation. In terms of use and feature prowess, the best analogy is that QGIS is to the ArcGIS-like proprietary GIS software world what OpenOffice and LibreOffice are to the proprietary MS-Office-like office software world. Begun in 2002, the Quantum GIS project since 2007 has been a project of the Open Source Geospatial Foundation (OSGeo), throughout its lifespan it has a fairly frequent update cycle averaging about four months between releases.

QGIS users can employ ESRI shapefiles, personal geodatabases, and coverages, also KML, and numerous other filetypes for vector data. Numerous raster data types are also supported, including ESRI grids, GeoTIFF, and Erdas Imagine; complex raster manipulation is accomplished with QGIS tools that interface with GRASS (another FOSS GIS software that installs automatically with QGIS). Supported databases include aforementioned ESRI personal geodatabases, PostGIS, and Spatialite, among others. The most recent version of QGIS (2.0) supports a nuanced map composer for high-quality cartographic output (Ross et al., 2013).

Two high-quality QGIS instruction books have recently been published by Graser (2013) and by Thiede et al. (2013). There is a robust user community for support, facilitating the development of numerous official and third-party plugins, help forums, blogs, and multimedia instructional materials. As a FOSS application, QGIS is made available through the terms of the GNU General Public License which guarantees the rights of users to use, copy, and modify the software (the source code is available). Official QGIS documentation is all covered by a Creative Commons Attribution-Share-Alike license that allows users to share, adapt, and use those materials for commercial purposes with the conditions that the original source is attributed and derivative products will be shareable as well (Creative Commons, 2009).

The User-Friendly Desktop GIS (uDIG). This is a full-featured, graphical user interface, desktop GIS application that runs on Linux, Mac, or Windows operating systems through the use of a Java virtual machine (this is a common design strategy for cross-platform software, Java is easily downloaded for free or may already exist on your computer). In terms of use and features, uDIG is a viewer as much as an editor. It gives users the useful ability to drag and drop supported filetypes into the application, including URLs for geospatial web services. Although uDIG supports many vector geospatial formats, inluding ESRI shapefiles and KML; the raster functionality is mainly focused on image overlays, which in the case of aerial imagery or scanned maps may be of important interest to archaeologists. Map creation through a specialized editor is straightforward and uncomplicated. The user community for uDIG seems smaller, but fairly active; the developers of uDIG provide a great deal of documentation and also a YouTube channel with numerous how-to videos. uDIG is based on the Eclipse Rich Client Platform and is made available through the terms of the Eclipse Distribution License which allows redistribution and modification of the software (the source code is available) but without the endorsement or liability of its creator (Eclipse Foundation, 2007).

Google Earth. This is neither a true GIS, nor open source (although it is zero cost at the entry level), but is a powerful virtual globe application, with data creation and editing capabilities, that installs on Linux, Mac, or Windows operating systems (with less functional versions available for iOS and Android). In terms of use and features, Google Earth is a viewer as much as an editor that utilizes drag-and-drop functionality to overlay vector and raster data on top of Google's proprietary delivery of detailed aerial imagery and topographic relief. The basic datatype for storing Google Earth information is KML (an XML variant that can also contain linked data like hyperlinks and image links); numerous FOSS and proprietary tools exist to convert various geospatial data types to KML (QGIS and uDIG do this), also fee-based Google Earth Pro has conversion functions. There is a highly robust user community for Google Earth that produces and shares KML data sets; the public accessibility and visibility of this resource is a definite strength for archaeologists interested in sharing their data professionally or publicly among users who may not have definitive GIS skills or access (e.g. Hochstetter et al. 2011; Beale 2012; Harris 2012). As a proprietary product and service that is made available freely (as in free beer) there are significant restrictions on the reuse of Google's content, however Google generally allows for fair use research products like print media (Google 2012, 2013).

Other FOSS GIS applications Archaeologists may find uses for other software such as GRASS (aforementioned), gvSIG (similar to uDIG), Marble (similar to Google Earth), OpenJump (similar to uDIG), NASA World Wind (similar to Google Earth but geared for software developers). There are good compilations of numerous FOSS GIS applications for the GIS planning and natural science communities (Steiniger, 2013; Steiniger and Hunter, 2013) that may be useful for comparative purposes.

People who wish to experiment with a wide variety of FOSS GIS applications with little difficulty or investment may wish to try the OSGeo Live Disk (OSGeo (Open Source Geospatial Foundation), 2013*b*,*a*), a Xubuntu Linux system that is designed to be loaded on a flash drive or burned to a DVD. A test user may boot their computer with the disk and use the software without changing the operating system or the persistent memory of the computer. The annual release of the OSGGeo Live Disk is celebrated by users at the international Free and Open Source for Geospatial (FOSS4G) conference as an important tool to help develop geospatial competencies in all manner of user communities.

9.4 Use Case One: FOSS GIS with Heritage Management Data

For professionals who use legacy archaeological heritage management data, and create new data sets - whether for compliance, research, outreach, or all - a FOSS GIS application is an excellent tool for conducting work. In the example presented here, Quantum GIS was used to edit and visualize archaeological site data from several State Historic Preservation Offices (SHPOs) in the Midwestern United States. An important duty of a SHPO is to maintain a database that describes the preservation status and potential information value of all known historic and prehistoric cultural resources within a state's jurisdiction (Neumann and Sanford, 2001). These databases are mainly used for heritage management and coordination with other government services, but aspects of them can have excellent potential for research and modeling. Governmental practice in the American federal system has resulted in the creation of many unique state site database file structures that are not necessarily interoperable. Although they all contain comparable archaeological descriptive data with scientific value, these are recorded in various formats to suit local software requirements and the shorthand of local government paperwork. Importantly, however, the functions of geospatial data management do provide some cross-cutting file types which can be used to begin to evaluate the intersections of the data within them. The relationships of similar archaeological sites within two different SHPO databases is somewhat analogous to that of electronic patient records in the medical sector. Two different clinics may treat patients with similar conditions but keep distinctly differently organized databases. Each clinic necessarily uses official jargon for qualitative descriptions and recognized standards for metrics, but an expert is required to interoperate both sets of records into a contiguous and analytical set (Andersson et al., 2003; Kaplan et al., 2003; Thiru et al., 2003). QGIS as a full-featured GIS is an good choice for the assessment, query, and translation of these data sets. The archaeologist can immediately grasp the spatial aspects of the complete sets and derived queries, and begin to consider ontological plans for interoperation.

The goals of this exercise (full details available in Wells 2011) were to create one interoperable data set from four SHPO archaeological databases in order to shed light upon the potential of these cybertools (Kintigh, 2006; Snow et al., 2006) to be used in analytical functions regarding landscape distributions, site functions, or other anthropological questions. The exercise used available GIS structures which were provided in original ESRI shapefile, personal geodatabase, and delimited text formats from their governmental sources. The site records under investigation describe components from the archaeological cultural tradition defined as Mississippian. These include farmsteads, villages, towns, and special use sites relating to numerous prehistoric agricultural polities with various hierarchical organizations throughout the American Midwest and Southeast, in the date range AD 1000-1500. Figure 9.1 illustrates the geographic scope of data sets involved within QGIS and its 3D globe plugin.

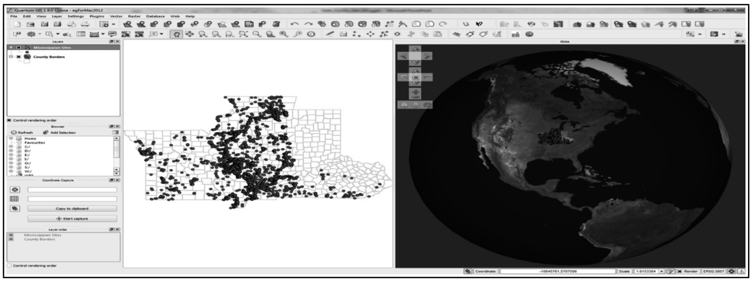


Figure 9.1: Mississippian sites from separate governmental databases in the Midwestern United States visualized through QGIS

This exercise demonstrates just one use of a FOSS GIS application to deal with public archaeological data available from government offices to qualified professionals. There is also a growing trend in publication of more open archaeological data resources related to heritage management on the Web that provide geospatial data for archaeological resources in ways that are both scientifically important and compliant with legal statutes and ethical imperatives for resource protections. Organizations such as the Archaeology Data Service, Open Context, the Paleoindian Database of the

Americas (PIDBA), and the recently founded Digital Index of North American Archaeology (DINAA) project all make such geospatial data available for public use, and are excellent resources for exploring the potential for FOSS GIS.

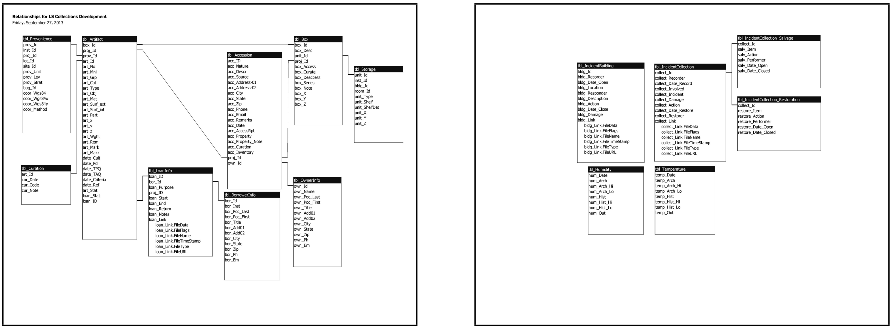
9.5 Use Case Two: FOSS GIS and Archival Management at the VAARNG Curation Facility

The Virginia Army National Guard (VAARNG) Curation Facility at Fort Pickett (Blackstone, Virginia), which operates as a proponent of the Department of Military Affairs Facilities Management Office's Cultural Resources Management (CRM) Program, has served as the repository for the VAARNG Archaeology Collection since 2003. The collection consists of over 20,000 artifacts and records associated with more than 150 investigations at 14 training facilities and readiness centers throughout the Commonwealth in accordance with the National Historic Preservation Act (NHPA) and the National Environmental Policy Act (NEPA). Most of these, however, were conducted at Fort Pickett, which is a 41,000 acre Maneuver Training Center in the Southside Virginia Piedmont. In this respect, the curation facility has proven an invaluable resource for understanding the history of a region largely overlooked in Virginia's history.

From 2003 to 2012, the VAARNG Curation Facility had been ably administered by the Conservation Management Institute (CMI) of Virginia Polytechnic Institute and State University through a contract with the CRM Program. Now that the Department of Military Affairs has assigned a permanent Collections Manager to the facility, the CRM Program will continue to meet its statutory obligations regarding the curation of state and federal archaeological collections. However, the CRM Program will also leverage data management technologies and GIS to improve both the administration of the VAARNG Archaeology Collection and the overall effectiveness of the CRM Program in its support of the National Guard's mission.

One such innovation is the development of a purpose-built, MS Access-based collections management system, Lil'Sorrel (named for the locally famous horse of a Virginia Confederate general). This relationsal database system incorporates the Sonoma Historic Artifact Research Database (SHARD), a freely available database scheme promoted by the Society for Historical Archaeology (SHA) and the Anthropological Studies Center at Sonoma State University; although the structure of SHARD is considered public with attribution credit the database itself is distributed in a native MS-Access format (SHA (Society for Historical Archaeology), 2011). Lil'Sorrel has enabled the Collections Manager to account for every item in the VAARNG Archaeology Collection in real time at the artifact-, box-, and project- (or accession-) level. Lil'Sorrel was preceded by an earlier system, Lil'Benny 2.0, for a previous project detailed by Parr (2011); the earlier Lil'Benny system served as a working prototype for the relational model governing Lil'Sorrel, although certain modifications were required to meet the specific needs of the VAARNG Curation Facility and the CRM Program. In an attempt to standardize the general functional classification system employed by Lil'Sorrel, the SHARD coding system was adopted but modified to accommodate prehistoric sites. The classification system and its associated lexicon are enforced by linking the necessary fields with (sometimes cascading) drop-down menus. These link to various "Options" tables, which can be accessed separately to allow new selections to be added as needed.

A database model for Lil'Sorrel is shown in Figures 9.2a and 9.2b. These inform the more than 80 queries, forms, and reports that have been developed to operate the system. Its most basic function is to account for the artifacts as they enter (e.g., accession), move through (e.g., exhibit, inspection, study), and leave (e.g., deaccession, loan) the VAARNG Curation Facility. Additionally, Lil'Sorrel has automated several routine tasks performed by the Collections Manager: it estimates the remaining "free space" available in the storage rooms; it generates loan agreements and provides notification of approaching due dates; it records and reports the daily environmental (e.g., temperature, relative humidity) readings inside the facility; and it documents any adverse incidents involving the building or the collection. Most importantly, Lil'Sorrel allows the CRM Program to track its artifacts "from screen to shelf" by georeferencing each to its original findspot. When combined with flexible search options, Lil'Sorrel enables the CRM Program (and by extension the contract archaeology firms it regularly employs) to plot, identify, and model artifact distribution patterns as never before.



(a) Database relationships for Lil'Sorrel collections management

(b) Database relationships for Lil'Sorrel collections incident reporting

Figure 9.2: Database design

The example shown in Figure 9.3 illustrates some of the FOSS capacity for mapping archaeological collections through QGIS. Geological occurances of diabase dikes - shallow, intrusive igneous deposits of fine-grained and sometimes glassy rock - have been noted throughout Fort Pickett, and the CRM Program is investigating how this local resource may have been utilized in the past. A simple search in Lil'Sorrel for "diabase" materials yielded 61 specimens recovered from 32 contexts ranging from surface finds to shovel test units (this number will grow, as the database migration is still in progress). These can be exported (as a Microsoft Office Excel table), converted to a comma separated values (CSV) format, and imported into a GIS (i.e., QGIS 2.0.1) to allow for low-usage-barrier geospatial analysis by either the Collections Manager or another subject matter expert. The search-export feature, combined with FOSS GIS, permits more efficient data sharing between the CRM Program and the scientific community.

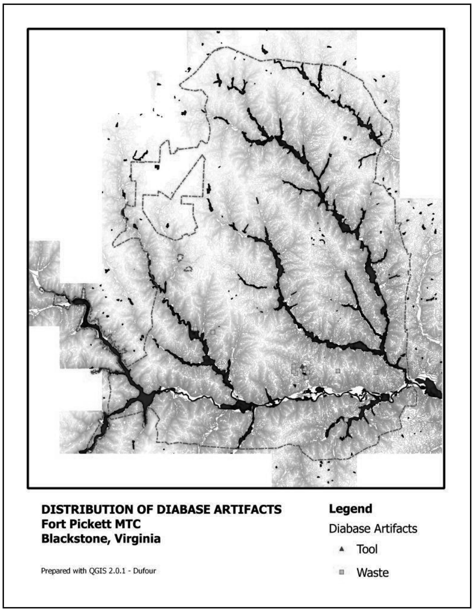


Figure 9.3: Example QGIS mapping output from Lil'Sorrel queries of diabase artifacts at VAARNG

An MS Access-based system has allowed the CRM Program to rapidly develop and deploy Lil'Sorrel, and thereby demonstrate its potential and encourage support from both inside and outside the CRM Program. It has also afforded an opportunity to integrate data from the curation program with other databases that have been developed to manage the CRM Program's projects and geospatial information. Unfortunately, Lil'Sorrel faces serious challenges with MS Access: it limits its size to 100,000 entries and 2.0 GB; the backend database resides on the VAARNG's network and can only be accessed internally by CRM Program personnel; and there is no guaranteed funding source available to research, develop, or deploy a replacement.

There are numerous opportunities to overcome these challenges with FOSS. While it is a common misconception that the Department of Defense forbids Open Source Software (OSS, as FOSS is termed by the DoD), its departments have made widespread use of such applications including web browsers (e.g., Mozilla Firefox), scripting languages (e.g., Python, PHP) and relational database systems (e.g., MySQL, PostgreSQL). According to a recent white paper sponsored by the Office of the Secretary of Defense, OSS provides the military with increased flexibility, faster delivery, more innovation, and information assurance at lower costs than proprietary "off the shelf" resources (Scott et al. 2011). The DoD has issued guidance through its Chief Information Officer to encourage the adoption and use of OSS throughout the Department and its affiliates (Wennergren, 2009).

This presents Lil'Sorrel with two available paths. The first is to replace Lil'Sorrel outright with an actual museum management program such as CollectiveAccess or CollectionSpace, two freely distributed open source systems developed specifically for managing collections held by museums, archives, and historical societies of all sizes worldwide. The other alternative is to adapt Lil'Sorrel to an FOSS equivalent of MS Access such as Glom, Kexi, or Wavemaker (in conjunction with either MySQL or PostgreSQL for the backend database). Although these will not deploy as rapidly as the "ready out-of-the-box" options mentioned above, FOSS applications will afford the CRM Program the ability more readily retain its ability to integrate the various databases that govern its operations. All of these are web-based applications, which hold the possibility for increased access to the VAARNG Archaeology Collection through a secure server. As the mission of the Department of Military Affairs is to administer the National Guard and its facilities in Virginia, it is not a research institution and therefore cannot function as such. Still, the CRM Program will continue to evaluate these and similar options in the hope of extending access of its archaeological resources to those institutions that would best benefit.

9.6 Use Case Three: FOSS GIS in the University

FOSS GIS is a boon to educator-researchers in a university system, in that it provides an exceedingly useful and accessible toolkit with which students at all levels can develop geospatial competencies and utilize geospatial data for a number of course-related projects. Furthermore, the accessibility of these geospatial tools lowers a number of barriers and creates a very open and fluid dynamic for entry of student research assistants into the productive process for creation and analysis of geospatial data. The examples in this section are from coursework, and research activities with student involvement, at Indiana University South Bend. These activities were part of the undergraduate curriculum in anthropological archaeology that includes courses developed on the Massachusetts Institute of Technology's model of "technology-enabled active learning" (TEAL) that includes training in software applications to help facilitate deeper and more functional comprehension of scientific concepts in a university setting (e.g. Barak and Dori 2005; Dori and Belcher 2005). The examples to be discussed in this section include: (1) dedicated courses in GIS methods, based completely or partially in Quantum GIS; (2) the uses of Google Earth to support geospatial comprehension in introductory and advanced courses in archaeological science; and (3) the uses of QGIS to facilitate student participation in geospatial data collection, analysis, and presentation related to archaeological field schools and other research projects in support of and collaboration with faculty.

9.6.1 GIS Classes

A first question to be answered of FOSS GIS is simply, "Can you actually teach a meaningful GIS class with it?" The answer is, "Yes!" A dedicated course in GIS using FOSS applications can be a rewarding experience for the instructor as well as the students since all involved can utilize the software on any available computer, not just simply from a dedicated lab or other locations where proprietary licenses are installed (or a limited number are swapped between available installations). The generally lower system requirements for FOSS GIS also makes student use of their personal computer for coursework an almost certain success.

The introductory course described here used Quantum GIS (v.1.6 CopiapÃş) as the primary software for instruction. The class was primarily about vector-based data collection, management, analyses, and map production. Training on raster data usage was limited to manipulations that added interpretive value to map production, such as clipping, and instructions that translated raster projections to correspond with vector projections (QGIS 1.6 only had on-the-fly projection for vector data, later versions have this feature for all data). University technology support staff were requested to install QGIS on all general campus computers (running Windows 7), and students were provided instructions, in-class examples, and online guidance to install QGIS through their personal Windows, Mac, and Linux operating system computers. The ease and elegance of multiple operating system installation created an immediate environment of competency, camaraderie, and ownership on the part of the students.

As a fairly standard introduction to mostly vector-based data analyses, the course progressed over the course of a semester through a series of lessons and exercises that explained data management mainly through the examples of shapefiles and KML. Cartographic, general GIS, and software specific references were required textbooks, including Making Maps: A Visual Guide to Map Design for GIS (Krygier and Wood 2005), How to Lie with Maps (Monmonier 1996), and the Quantum GIS User Guide (QGIS Development Team 2011 [2013]). The instructor provided students with useful example web resources for each lesson topic. Students were encouraged to explore user community forums on their own outside of class. Students also created collaborative "quick use guides" in class-shared Google Docs, based on their own trials and forum searches. In total, the collaborative-yet-self-reliant practices of digital literacy that are a part of good practice in the FOSS GIS community helped students to develop competencies and skills across a wide range of associated technologies which would serve them well in future GIS classes or in GIS workplaces.

9.6.2 Introductory and Advanced Archaeology Classes

Freely available mapping applications create many new affordances and opportunities for students to learn core course topics and geospatial concepts in tandem. For courses where there is neither available time nor prerequisite preparation for a full-featured GIS application, the free use of Google Earth is a pragmatic and effective choice. In a freshman-level class on human evolution and archaeology, Google Earth provides students with the capacity to create, share, combine, and review (as part of an overall TEAL process Wells and VanderVeen n.d.) point and polygon data sets such as modern and extinct primate distributions; geolocational data from the UNESCO World Heritage List database, and other heritage organizations can be downloaded and manipulated in order to learn about traits and distributions of international heritage properties. At more senior levels, Google Earth is suitable for mapping individual and group research on heritage sites at regional or global scales. Students can annotate their KML data with image links and hypertext descriptions that build a functional public outreach project from their original research that is easily published through Google Maps or any other KML viewer. An archaeological methods class can engage in a systematic surface survey based on a geolocated datum and visualize their data quickly and easily using either (1) algebraic coordinate calculation and plotting manually or through CSV to KML transformation using numerous third party FOSS applications or websites, or (2) radial point plotting using Google Earth's ruler tool.

9.6.3 Archaeological Field Schools and Other Research

A field school is an arena the demonstrates the immense pedagogical and research potential of FOSS GIS applications. Quantum GIS is the perfect application for geospatial data management and analysis during an ongoing field project, to integrate students into the process. The lightweight, multiple platform capability, and of course freely shareable installation of QGIS makes GIS work a possibility on every project computer, and on available student computers as well.

QGIS has fairly low and readily attainable system requirements, although these are not specifically defined by the developers due to the wide variety of operating systems they support. On recent archaeological field schools, QGIS 1.6, 1.7, and 1.8 all ran well on bargain-priced, ruggedized netbooks from 2GoPC. These systems each had a

1.6 GHz processor, 1GB of RAM, and a 1024 by 600 pixels (compare to ArcGIS minimum system requirements of a 2.2 GHz processor and 2GB of RAM).

The distribution of FOSS GIS-ready laptops throughout the project made on-site data entry a viable possibility that kept the notebook-to-digital stream flowing in near to real time. Through structured file sharing among project participants, remaining data entry and rudimentary analyses could be conducted during evenings by interested students on their personal computers. In sum, QGIS afforded the project the opportunity to develop a faster workflow, and students the opportunity to learn and actualize upon training with geospatial data concepts in a low-overhead and license-free (no swapping, no dongles!) scenario. For some students this training was directly applicable to later projects in a related archaeological laboratory course.

9.7 Conclusion

This chapter has provided an introduction to ways in which free and open-source geographic information systems and other free geospatial software can be integrated into a professional workflow in several arenas of archaeological activity. Archaeologists are encouraged to begin accessing these resources and participating in user communities, beginning with the guideposts described here. The ability to use GIS and geospatial applications outside of specialized labs, now on any available and moderately-aged computer, should help promote more rigorous training, project-level, and disciplinary activities in data sharing and reuse. The “open” concept is not just for software, but for archaeological data in general (cf. Kansa et al. 2011; Anderson et al. 2010)!

At the university, GIS and geospatial training and concepts, at every level in archaeology, should no longer be relegated to a select few who have obtained the necessary equipment and software, which have themselves often been expensive and with significant use limitations related to license restrictions. FOSS GIS and other free geospatial software applications should be a part of every archaeological degree program from the undergraduate through the doctoral, and given a priority equivalent to database management and statistics in curriculum design. Indeed, GIS and geospatial training need no longer be separate from the other data-driven portions of a curriculum; maps should be considered as another mode of freely accessible visualization (“carto graphs” should be taught alongside line graphs and bar graphs) that facilitates comprehension of a dataset for analytical and outreach purposes.

In the compliance and research workplaces, GIS and geospatial data production, visualisation, and dissemination, should no longer be constrained by limitations of cost and accessibility. Many smaller laboratories, public and private, run on a tight budget and proprietary GIS licensing costs and restrictions may literally price some professionals out of the marketplace. Within the professional archaeological community, the broader employment of FOSS GIS should promote conversations with heritage management agencies, funding agencies, and curatorial facilities to augment existing frameworks for best practices in data management and publication (e.g. Archaeology Data Service / Digital Antiquity 2011; Open Context 2012; SAA (Society for American Archaeology) 2013; WLP (Shelby White and Leon Levy Program for Archaeological Publications) 2013) for local needs in a way that maintains data interoperability and reuse for the long term. The availability of FOSS GIS and geospatial freeware should also promote greater technological emphases in public outreach with spatial information, allowing the stakeholding public to engage in a reciprocal cycle of data production and consumption with the archaeological community (e.g. Beale 2012; Harris 2012; Hochstetter et al. 2011; Wells and McCullough 2009).

Bibliography

Anderson, D. G., Miller, D. S., Yerka, S. J., Gillam, J. C., Johanson, E. N., Anderson, D. T., Goodyear, A. C. and Smallwood, A. M. (2010), 'PIDBA (Paleoindian Database of the Americas) 2010: Current Status and Findings', *Archaeology of Eastern North America* **38**, 63–90.

Andersson, A., Hallberg, N. and Timpka, T. (2003), 'A model for interpreting work and information management in process-oriented healthcare organisations', *International Journal of Medical Informatics* **72**(1), 47–56.

Archaeology Data Service / Digital Antiquity (2011), 'Guides to good practice'. URL: http://guides.archaeologydataservice.ac.uk/g2gp/Main

Beale, N. (2012), 'How community archaeology can make use of open data to achieve further its objectives', *World Archaeology* **44**(4), 612–633.

Coleman, E. G. (2013), *Coding freedom: The ethics and aesthetics of hacking*, Princeton University Press.

Creative Commons (2009), 'Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)'. URL: http://creativecommons.org/licenses/by-sa/3.0/

Creative Commons (2013), 'History - creative commons'. URL: http://creativecommons.org/about/history

Dempsey, C. (2012), 'Open Source GIS and Freeware GIS Applications', GIS Lounge. URL: http://www.gislounge.com/open-source-gis-applications/

Eclipse Foundation (2007), 'Eclipse Distribution License - v 1.0'. URL: http://www.eclipse.org/org/documents/edl-v10.php

ESRI (1998), 'ESRI Shapefile Technical Description'. URL: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

ESRI (2013), 'Geodatabase Overview'. URL: http://www.esri.com/software/arcgis/geodatabase

Graser, A. (2013), *Learning QGIS 2.0*, Packt Publishing Ltd.

Harris, T. M. (2012), 'Interfacing archaeology and the world of citizen sensors: exploring the impact of neogeography and volunteered geographic information on an authenticated archaeology', *World Archaeology* **44**(4), 580–591.

Hochstetter, F. T., Haoa, S. R., Lipo, C. P. and Hunt, T. L. (2011), 'A public database of archaeological resources on easter island (rapa nui) using google earth', *Latin American Antiquity* **22**(3), 385–397.

Kansa, E. C., Kansa, S. W. and Watrall, E. (2011), 'Archaeology 2.0: new approaches to communication and collaboration', *Cotsen Digital Archaeology series* .

Kaplan, B., Schold, J. and Meier-Kriesche, H.-U. (2003), 'Overview of large database analysis in renal transplantation', *American Journal of Transplantation* **3**(9), 1052–1056.

Kintigh, K. (2006), 'The promise and challenge of archaeological data integration', *American Antiquity* pp. 567–578.

Kling, R., Rosenbaum, H. and Sawyer, S. (2005), *Understanding and communicating social informatics: A framework for studying and teaching the human contexts of information and communication technologies*, Information Today, Inc.

Neumann, T. W. and Sanford, R. M. (2001), 'Cultural resources archaeology: an introduction'.

Open Context (2012), 'Open Context Data Publishing: Editorial Policies and Author Guidelines'.
URL: http://alexandriaarchive.org/wp-content/uploads/2012/04/OpenContext_DataPublishing_Policies&Guidelines.pdf

Open Source Initiative (2013), 'The open source definition'.
URL: http://opensource.org/osd

OSGeo (Open Source Geospatial Foundation) (2013*a*), 'Open Source GIS History - OSGeo Wiki'.
URL: http://wiki.osgeo.org/wiki/Open_Source_GIS_History

OSGeo (Open Source Geospatial Foundation) (2013*b*), 'Welcome to OSGeo-Live 7.0–OSGeo-Live 7.0 Documentation'.
URL: http://live.osgeo.org/en/index.html

Parr, C. (2011), 'From the ground up: Best practices for balancing usability with theoretical utility in archaeological databases', *Society for American Archaeology Annual Meeting* .

Raymond, E. (2000), 'The cathedral and the bazaar'.
URL: http://www.catb.org/esr/writings/cathedral-bazaar/

Ross, S., Sobotkova, A., Ballsun-Stanton, B. and Crook, P. (2013), 'Creating eresearch tools for archaeologists: The federated archaeological information management systems project', *Australian Archaeology* (77), 107.

SAA (Society for American Archaeology) (2013), 'Society for american archaeology guidance regarding national science foundation archaeological data management plans'.
URL: http://www.nsf.gov/sbe/bcs/arch/SAA_Data_Management_Plan_Guidance.pdf

SHA (Society for Historical Archaeology) (2011), 'Sha artifact cataloging system'.
URL: http://www.sha.org/research/artifact_cataloging_system.cfm

Sherman, G. (2012), *The Geospatial Desktop: Open Source GIS & Mapping*, Locate Press; 2nd edition.

Snow, D. R., Gahegan, M., Giles, C. L., Hirth, K. G., Milner, G. R., Mitra, P. and Wang, J. Z. (2006), 'Cybertools and archaeology', *Science* **311**(5763), 958–959.

Steiniger, S. (2013), 'SourceForge.net: Open Source GIS Software'.
URL: http://sourceforge.net/userapps/mediawiki/mentaer/index.php?title=Open_Source_GIS_Software

Steiniger, S. and Bocher, E. (2009), 'An overview on current free and open source desktop GIS developments', *International Journal of Geographical Information Science* **23**(10), 1345–1370.

Steiniger, S. and Hay, G. J. (2009), 'Free and open source geographic information tools for landscape ecology', *Ecological Informatics* **4**(4), 183–195.

Steiniger, S. and Hunter, A. J. (2013), 'The 2012 free and open source GIS software map–a guide to facilitate research, development, and adoption', *Computers, Environment and Urban Systems* **39**, 136–150.

Thiede, R., Sutton, T., Düster, H. and Sutton, M. (2013), *The Quantum GIS Training Manual: A Comprehensive Introduction to Quantum GIS*, Locate Press.

Thiru, K., Hassey, A. and Sullivan, F. (2003), 'Systematic review of scope and quality of electronic patient record data in primary care', *BMJ* **326**(7398), 1070.

Wells, J. (2011), 'Four States of Mississippian Data: Best Practices at Work Integrating Information from Four SHPO Databases in a GIS-Structured Archaeological Atlas'.

URL: http://visiblepast.net/see/americas/four-states-of-mississippian-data-best-practices-at-work-integrating-information-from-four-shpo-databases-in-a-gis-structured-archaeological-atlas/
Wells, J. and McCullough, R. (2009), 'Multiple scales of data on falls mississippian settlement practices', *Indiana Archaeology* p. 56.
Wells, J. and VanderVeen, J. (n.d.), 'Anthropological pedagogy through technology enabled active learning in undergraduate education'.
Wennergren, D. (2009), 'Clarifying guidance regarding open source software (OSS)'.
URL: http://dodcio.defense.gov/Portals/0/Documents/FOSS/2009OSS.pdf
Wikipedia (2013), 'List of geographic information systems software'.
URL: http://en.wikipedia.org/w/index.php?title=List_of_geographic_information_systems_software&oldid=569623888
Williams, S. (2002), 'Free as in freedom: Richard Stallman's crusade for free software'.
URL: http://oreilly.com/openbook/freedom/index.html
WLP (Shelby White and Leon Levy Program for Archaeological Publications) (2013), 'The Shelby White-Leon Levy Program: Data Management Plan'.
URL: http://www.fas.harvard.edu/~semitic/wl/dataplan.html

List of Figures

List of Tables

Index