

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



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 license, which means that the text may be used for non-commercial purposes, provided credit is given to the author. For details go to <http://creativecommons.org/licenses/by-nc-nd/3.0/>.

Copyright © 2015 Andrew T. Wilson and Ben Edwards

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 — IX

Benjamin Edwards and Andrew T. Wilson

- 1 Open Archaeology: Definitions, Challenges and Context — 1**
- 1.1 Introduction — 1
- 1.2 ‘Open Source’ Archaeology and ‘Open’ Archaeology — 1
- 1.3 Open Source Archaeology — 1
- 1.4 Open Archaeology — 2
- 1.5 The Public Context of Open Access — 2
- 1.6 Open Ethics — 4
- 1.7 Outline of the Volume — 5
- Bibliography — 5

Jeremy Huggett

- 2 Digital Haystacks: Open Data and the Transformation of Archaeological Knowledge — 6**
- 2.1 Introduction — 6
- 2.2 Openness and Access — 7
- 2.3 Openness and Reuse — 10
- 2.4 Approaches to Open Data — 13
- 2.5 From Data to Knowledge? — 15
- 2.6 From Knowledge to Data? — 17
- 2.7 Putting the ‘Capta’ Back into Data? — 19
- 2.8 Transforming Knowledge? — 21
- 2.9 Open Data is for Sharing — 24
- Bibliography — 26

Ray Moore and Julian Richards

- 3 Here Today, Gone Tomorrow: Open Access, Open Data and Digital Preservation — 30**
- 3.1 Introduction — 30
- 3.2 Sharing Data: The ‘Traditional’ Treatment of Archaeological Data — 31
- 3.3 Accessing Data: The Case of the Archaeology Data Service — 36
- 3.4 Conclusion — 40
- Bibliography — 41

Gareth Beale and Nicole Beale

- 4 Community-Driven Approaches to Open Source Archaeological Imaging — 44**
- 4.1 Introduction — 44
- 4.1.1 Using Case Studies — 45

4.2	Technological Innovation: New Development Models —	46
4.3	Introducing the Methodology —	48
4.3.1	Reflectance Transformation Imaging —	48
4.3.2	Photogrammetry —	48
4.3.3	Assessing the Benefits of Open Source Imaging Methodologies —	49
4.4	Community Methodologies and Technological Uptake —	50
4.4.1	Development Model Alternatives —	50
4.4.2	Ensuring Meaningful Software Access —	51
4.5	Case Study One: Basing House Community, Archaeology and Technology Project —	52
4.5.1	Skills Sharing —	54
4.5.2	Experimental Atmosphere —	54
4.5.3	Basing House Conclusions —	55
4.6	Case Study Two: Re-Reading the British Memorial —	55
4.6.1	Adaptive Methodologies —	57
4.6.2	Decentralised Approach —	59
4.6.3	Project Sustainability —	59
4.6.4	Recognising Skills —	59
4.6.5	Re-Reading the British Memorial Project Conclusions —	61
4.7	Conclusion —	61
	Bibliography —	62

Hector Orengo

5	Open Source GIS Geospatial Software for Archaeology: Towards its Integration into Everyday Archaeological Practice —	64
5.1	Introduction —	64
5.2	What is Open Source Software? —	66
5.3	Why Use Open Source GIS? —	66
5.4	Problems with Open Source GIS —	69
5.5	Common Misconceptions Regarding Commercial and Open-Source Software —	70
5.6	Which Open-Source Desktop GIS is more Convenient? —	73
5.6.1	GRASS (Geographic Resources Analysis Support System) GIS —	73
5.6.2	QGIS (Formerly Known as Quantum GIS) —	75
5.6.3	SAGA (System for Automated Geoscientific Analyses) GIS —	76
5.6.4	gvSIG (Generalitat Valenciana Sistema d'Informació Geogràfica) —	76
5.6.5	Other Open Source Geospatial Software —	78
5.7	Open Geospatial Data? —	79
5.8	Conclusions —	80
	Bibliography —	81

Lolita Nikolova

- 6 What was Published is as Important as How it was Published — 83**
- 6.1 Introduction — 83
- 6.2 Open Access and Global Society — 83
- 6.3 Open Access, Archaeology and Ethics — 87
- 6.4 Conclusions — 88
- 6.5 The Future — 89
- Bibliography — 90

Benjamin Ducke

- 7 Free and Open Source Software in Commercial and Academic Archaeology — 92**
- 7.1 Introduction — 92
- 7.2 Selected Aspects of F/OSS — 94
- 7.2.1 Open Source Economics — 94
- 7.2.2 Social Dynamics of F/OSS — 96
- 7.3 F/OSS in Research — 97
- 7.3.1 Publish (Your Source Code) or Perish! — 98
- 7.3.2 Reproducible Research — 100
- 7.3.3 Data-Centric Research — 100
- 7.4 Case Studies — 102
- 7.4.1 Oxford Archaeology Digital: F/OSS Migration in the Workplace — 102
- 7.4.2 gvSIG and gvSIG CE: The Role of Social Capital in F/OSS — 104
- 7.4.3 Survey Tools: F/OSS for Field Archaeology — 106
- 7.5 Conclusions — 107
- Bibliography — 109

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

- 8 Building the Bazaar: Enhancing Archaeological Field Recording Through an Open Source Approach — 111**
- 8.1 Introduction — 111
- 8.2 FAIMS: Overview and History of the Project — 111
- 8.3 The State of Play: Sharing in the World of Archaeology — 114
- 8.3.1 Archaeologists and Open Source Software — 114
- 8.3.2 The Ethos of Sharing in the Archaeological Community — 116
- 8.3.3 Creating and Sharing Repurposable Digital Data — 117
- 8.4 Open Source Beyond Software — 120
- 8.4.1 Free-as-in-beer and Free-as in Speech: Open Source Paradigms for Scholarship — 120
- 8.4.2 The GitHub Revolution — 121
- 8.5 New Applications of Open Source Techniques — 123

- 8.5.1 Open Source Approaches to the Development of Recording Systems — **123**
- 8.5.2 Improving Sustainability through Reuse and Redeployment — **125**
- 8.5.3 Improving Archaeological Practice through Dataset Interoperability — **126**
- 8.6 Conclusion — **127**
Bibliography — **128**

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 — 130

- 9.1 Introduction — **130**
- 9.2 What Geographic Information Systems and Free and Open Source Software are not — **131**
- 9.3 What does Open Source mean in GIS — **133**
 - 9.3.1 Pros and Cons — **133**
 - 9.3.2 Software Application Summaries — **133**
- 9.4 Use Case One: FOSS GIS with Heritage Management Data — **135**
- 9.5 Use Case Two: FOSS GIS and Archival Management at the VAARNG Curation Facility — **137**
- 9.6 Use Case Three: FOSS GIS in the University — **140**
 - 9.6.1 GIS Classes — **141**
 - 9.6.2 Introductory and Advanced Archaeology Classes — **142**
 - 9.6.3 Archaeological Field Schools and Other Research — **142**
- 9.7 Conclusion — **143**
Bibliography — **144**

List of Figures — 147

List of Tables — 148

Index — 149

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

 © 2015 Benjamin Edwards and Andrew T. Wilson

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License.

Brought to you by | Macquarie University

Authenticated

Download Date | 11/16/15 12:11 AM

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 Ed-

ucation 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 is 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 main 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.

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 years¹. 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 overseas². 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 (SpatialLite). 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 long-standing diversity of archaeological research agendas, methods, and field procedures. All data produced using the platform benefits from a robust but flexible underlying datastore⁴, 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 inducements⁵.”

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 cumulative, 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 2000a, 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 the Digital Data Survey asked about specific attitudes towards the sharing of primary data⁶. 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 all⁷. 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 Kiritzi 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') tables⁸. 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 attempts 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 GIS software 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 Access and 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 code⁹.

“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 contribu-

tor. 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/Nl20B0>) 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, interpreta-

tions, 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 sys-

tems - 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 (MEM-SAP), and the Boncuklu Höyük Project (a Neolithic tell in Turkey). Each adaptation was different, responding to 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-

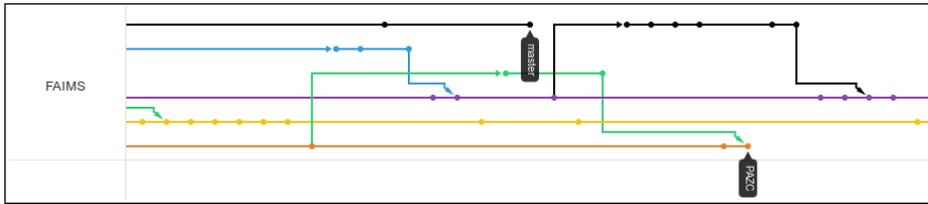


Figure 8.1: Network graph of the FAIMS ‘Excavation’ repository, with four branches: PAZC (orange), Boncuklu (purple), MEMSAP (yellow), and the original, ‘Master’ branch (black)

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 fur-

ther 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 Android 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 datasets¹⁰.

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 2000a, 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', 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 Kiriati, 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. (2000a), 'The cathedral and the bazaar'.
URL: <http://www.catb.org/esr/writings/cathedral-bazaar/cathedral-bazaar>
- Raymond, E. (2000b), '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. Tragwillia, 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>