Title: Archaeology as Community Enterprise

Author: Néhémie Strupler, email: nehemie.strupler@etu.unistra.fr

Affiliations:

1. « Archéologie et Histoire ancienne : Méditerranée-Europe », ARCHIMÈDE (UMR 7044), Université de Strasbourg
2. Institut für Altorientalische Philologie und Vorderasiatische Altertumskunde, Westfälische Wilhelms-Universität Münster
3. Deutsches Archäologisches Institut, Abteilung Istanbul

Keywords:

Free Software, Collaborative Authoring, Reproducibility, Computational Archaeology

Abstract:

To address methodological and theoretical issues in archaeology computational (and quantitative) archaeology is commonly used. However few debates occur on the software that perpetrate the analysis, manipulate data or how this should be available. This paper addresses the need to use Free Software at a moment, where data is becoming easy available online, and tools to make reproducible research become widespread. This configuration challenges current ways of publishing quantitative research.

Text:

1. Introduction[[1]](#footnote-1)

The Open Access movement raises a lot of sympathy among archaeologists. Academia.edu has grown to one of the biggest repositories for archaeological papers « freely » available, cross cutting boundaries of traditional reading circles[[2]](#footnote-2). Everyone understands some of the benefits of this availability. On the other side, the Free Software movement, where the origin of the Open Access movement is to be situated, is not considered as relevant. However, on-going and growing debates about benefits of dynamism in publication available through new information technologies challenge this position. Why should archaeologists consider the philosophy of Free Software with more awareness?

Even if Free Software is well established inside and outside academia, the advantages of Free software (that is "the 'free' as in 'free speech', and not the 'free' as in 'free beer' ") are not acknowledged[[3]](#footnote-3). Free software is in many cases (in the imaginary) understood as complicated, untrustable or no as effective as proprietary software. Generally software is considered as a tool to answer questions and the emphasis is put on the results and not on the reuse neither of work-flows nor of the collected data.

The aim of this paper is to give a brief overview of the intersection of Free Software with computational Archaeology. This closer look draws the attention on changes in information technologies that question knowledge dissemination. Not only Open Access is important to sustain for large access to research outcome but the access to the data and the analysis should be done in transparent and reproducible ways, which are closer to the ethic of scientific research. This will sustain a stronger research environment.

2. Precedent Work

The paper of Benjamin Ducke is, to my knowledge, the only article about the relation of Free (and Open Source) Software and Archaeology (Ducke, 2012). He focuses the attention on the problem of 'black boxes' (Morin, 2012) and of sustainability for the development of archaeological software. Proprietary software restricts the assessment, the control of the processes and hinders the dissemination of the analysis. All non-owner of the specific proprietary software are de facto excluded of the reproduction of the analysis. However as he states, his article does not focus on philosophical or social aspects of Free Software. In my take up, I would like to focus narrower on these problems and how access to data questions methods established in archaeological research.

3. Scientific community

Science is cumulative even if it is not clear how exactly scientific knowledge accumulates. In any case it is a community endeavour. In this process, the scientific community brings different points of views, different questions and different ways to resolve same problems. The community is a pool of inspiration: interactions encourage innovation and spark ideas about new lines of evidence, new applications, new questions, and alternate explanations. Lively interactions of the scientific community make the scientific knowledge accumulating and becoming "stable" faster[[4]](#footnote-4).

The scientific community is a motivating force for recognition and respect from peers and it controls the quality of scientific research by scrutinising the work of others. It's a system of checks and balances that assures that claims are not fraudulent. The communication and the open process inside the community make claim robust. It's not the claim of a unique researcher that makes science strong and reliable but true scrutiny and critics of peers (Fanelli, 2013).

4. Free Software and Science

Free Software shares a lot of characteristics with science. Both have scrutiny and cumulative knowledge, which can be seen in process like peer reviewing, open data subject to validation and replication. There is a strong culture of credit, civility, reputation and communication

The motivations to do Science or Free Software are in many aspects similar: it's based on the reputation earned with published work and there is in both cases an ethic to attribute the work (Kelty, 2001). Researchers make their work available to others and citations accredit reuse of ideas, concepts or code. In Free Software or in Science differences are made according to the status of the contribution (author, contributor, maintainer, etc.). Contributions are based on same principles of cumulative knowledge, its reuse, mixing and modulation.

But similarities between Free Software and Science should not hide differences. Dissimilarities, specifically Free Software's assets challenge in return, what is taken for granted in Science.

5. Modifiability

The concept of modifiability is without any doubt the most fascinating concept among the dissimilarities between Science and Free Software (Kelty, 2008, p. 12). Books and articles have acclimatised science making with stable knowledge. The impression given by books or articles is that of a definitive version, tangible under the fingers. But Free Software with practices like forking, new versioning, cloning, constant evolution of source code questions this concept. Every "newly available" operating system, even proprietary, contains older pieces of software.

[Figure1] Open and wide communication of negative as well as positive findings (Bartling et Friesike 2014, fig. 3, p. 10) | CC BY-NC 3.0

On top of the drawing [Figure1], a representation of the current way of doing research with stable knowledge: results are only used, when they are stable. At the bottom a prospect on how software development looks like and how science could be made. The process of sharing and reusing knowledge is becoming more dynamic. However this poses a question. If the content does not need any more to be stable and it is used like Free Software, "how should the authority, stability and reliability of knowledge be assessed" (Kelty, 2008, p. 300)? In this sense Free Software challenges the "Power of Knowledge"[[5]](#footnote-5).

6. Free Software challenges the Power of Knowledge

New practices of publication with the appearance of new information technologies render the knowledge more dynamic (Vinck, 2014). Wikipedia is the most famous example with content that can be updated, changed, deleted, copied or forked at any time by anyone. The last 10 years have seen a complete revaluation of Wikipedia at the University. 10 years ago, as I saw it in France, it was mostly rejected as "untrustable source". Now academics explain how to use it scientifically.

Even if there is a growing discussion on publication of dynamic content there are still few examples of work-flows in academia. The model proposed by Kansa and others (2014) – Push and Publish – will certainly have a bright future. In this model, data are published on-line and may be updated, completed, corrected, cloned after the first online release. Repositories and digital archiving enable new dissemination methods.

By looking at the practice of Free Software, it is easier to ask how authority is established and question the finality of publications. Free Software focuses not only on the results (the software) but unifies operations and the possibility of control, improvement and reuse. Free Software allows computational archaeology not only to concentrate on the results but broaden the attention consequently on how evidences are reached. This is the only way to reconcile the ethic of science and the establishing of (computational) facts. It's with this sense in mind that I will turn to the problem of reproducibility for computational archaeology.

7. Reproducibility

Reproducibility allows anyone to start from the same data, check all the processes in order to verify the results. The complete process, from ideas and data to the publication has been termed as research pipeline [figure 2]. In Archaeology there is no culture of exchanging code for reproducibility until now, even if it is changing[[6]](#footnote-6). If we turn back to the definition of science, however, reproducibility and sharing of code are even sine qua non for science, to permit the community to scrutinise the work. It's the only way to stick with a robust exposition of facts. "We often forget that scientific knowledge is reliable not because scientists are more clever, objective or honest than other people, but because their claims are exposed to criticism and replication." (Fanelli, 2013)

[Figure2] Research Pipeline (after ideas from Roger Peng and Kieran Healy) | N. Strupler CC BY-NC 4.0

For computational research, best practices permit to knit together data, code, and results through scripting (Marwick, 2015). Reproducibility must be attained to adhere to the ethic of science. Moreover reproducibility has a huge potential for empirical research, like archaeology. Reproducibility is important in order to reuse and maintain data up to date. If, as shown in the Published and Push model, data are updated, then the analysis can be quickly updated, and therefore results, do not need to stay in the (outdated) state as published.

Reproducibility helps to cast the research in modules. Bits of research based on code can be transformed and reused for similar projects or build upon for new projects. The modularity of code impacts on two levels. Firstly, open and reproducible archaeology makes resources easy available as paradigmatic pedagogical object. Hands-on workshops, modulations of research, testing of new hypothesis can be done in teaching environments with a direct, do it yourself approach. Secondly, the modularity of Free Software eases the blurring of academic borders in unpredictable ways. By providing a common language, Free Software creates new transversal communities that make science stronger.

8. Conclusion

Free Software challenges the authority of science and provides a mean to transform computational archaeology by questioning transmission and reuse of research. Free Software coupled with reproducibility by means of scripting makes all the research pipeline available to the scrutiny of the community and reconciles together data, processes and results. Making the complete research available has the potential to change computational archaeology into a more open and robust research environment.

9. References

BARTLING, S. and FRIESIKE S. (2014) Towards Another Scientific Revolution. In: S. Bartling and S. Friesike (eds). Opening science, Springer, p. 3-15. Available from: http://dx.doi.org/10.1007/978-3-319-00026-8\_1.

DUCKE, B. (2012) Natives of a connected world: Free and open source software in archaeology. World Archaeology. 44 (4). p. 571-579. Available from: http://dx.doi.org/10.1080/00438243.2012.743259.

FANELLI, D. (2013) Redefine misconduct as distorted reporting. Nature. 494. p. 149. Available from: http://dx.doi.org/10.1038/494149a.

HELLER, L., THE, R., and BARTLING, S. (2014) Dynamic Publication Formats and Collaborative Authoring. In: S. Bartling and S. Friesike (eds). Opening science, Springer, p. 191-211. Available from: http://dx.doi.org/10.1007/978-3-319-00026-8\_13.

KANSA, E.C., KANSA, S.W., & ARBUCKLE, B. (2014) Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology. International Journal of Digital Curation. 9 (1). p 57–70. Available from: http://dx.doi.org/10.2218/ijdc.v9i1.301.

KELTY, C.M. (2001) Free software/free science. First Monday. 6 (12). Available from : http://pear.accc.uic.edu/ojs/index.php/fm/article/view/902 [Accessed: 30th June 2015].

KELTY, C.M. (2008) Two bits: The cultural significance of free software. Durham: Duke University Press. Available from: http://twobits.net/pub/Kelty-TwoBits.pdf.

LATOUR, B. and WOOLGAR, S. (1979) Laboratory life: The Social Construction of Scientific Facts. Beverly Hills, Calif.: Sage.

MARWICK, B. (2015) Doing Quantitative Archaeology with Open Source Software. [Online] April 24th 2015. Available from: software.ac.uk. http://www.software.ac.uk/blog/2015-04-24-archaeology-open-source-software-its-getting-easier [Accessed: 30th June 2015].

MORIN, A. et al. (2012) Shining Light into Black Boxes. Science. 336 (6078). p. 159-160. Available from: http://dx.doi.org/10.1126/science.1218263.

VINCK, D. and CLIVAZ C. (2014) Les humanités délivrées. Savoir et culture réinventés hors du livre. Revue d’anthropologie des connaissances. 8 (4). p. 681–704. Available from: http://dx.doi.org/10.3917/rac.025.0681.

1. This paper is the output of a talk given during the session 10A "ArcheoFOSS: free/open source software and archaeological research, ten years later". I am grateful to the organisers and to the participants for the comments in a nice and motivating atmosphere. The original paper, the slide-show and extra material are accessible online. DOI: http://dx.doi.org/10.5281/zenodo.16596. A video of the lecture was recorded by Doug Rocks-Macqueen and has been published under his project *Recording Archaeology*. [↑](#footnote-ref-1)
2. [academia.edu](http://academia.edu/) is « free to use » but it is not free. There is no API, which would have enabled a query of the growth of papers tagged with #Archaeology during the last years and consequently asserts the affirmation from the main text. Until now, it's just a guess ... [↑](#footnote-ref-2)
3. See the definition of the Free Software Foundation. Available from https://fsf.org/ [↑](#footnote-ref-3)
4. "Stable knowledge" refers to the process of establishing a fact, among others (Latour, 1979). [↑](#footnote-ref-4)
5. I refer to the expression of Kelty (2008) in his book, Two Bits. [↑](#footnote-ref-5)
6. For example, see the meeting during the SAA 2015 in San Francisco, "Electronic Symposium Open Methods in Archaeology: How to Encourage Reproducible Research as the Default Practice?" organised by Ben Marwick, Mark Lake and Andrew Bevan. [↑](#footnote-ref-6)