Archaeology as Community Enterprise

Néhémie Strupler

Archaeology as Community Enterprise

Pre-print of the paper published in Campana, S., Scopigno, R., Carpentiero, G. & Cirillo, M. (Eds.) CAA2015 KEEP THE REVOLUTION GOING, Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, 2016, 1015-1018

Néhémie Strupler

Abstract

Computers are commonly used to address practical, methodological, and theoretical issues in archaeology. However little discourse is devoted to the software that is used to perform the analysis, manipulate data, or to how the software workflow should be available. This paper addresses the pressing need to adopt Free Software and transparent research pipeline now when data is becoming easily available online, and tools to make reproducible research are becoming widespread. This configuration challenges current ways to disseminate and evaluate archaeological research.

Introduction[1]

The Open Access movement generates plenty of sympathy among archaeologists. Academia.edu has grown to one of the biggest repositories for 'freely' available archaeological papers, cross-cutting the boundaries of traditional reading circles[2]. Everyone can understand the benefits of this ready availability of papers. On the other hand, the Free Software movement, where the origin of the Open Access movement is to be situated, is mostly ignored since it is not considered as relevant. However, on-going and growing debates about benefits of data release available through new information technologies challenge this position. Why, though, should archaeologists become more aware of the philosophy of Free Software?

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 rarely acknowledged[3]. Free software is in many cases (in the imaginary) assumed to be complicated, untrustable or not as effective as 'industry-standard' software. More generally software is treated as a neutral tool to answer research questions, and emphasis is placed on the results with little time devoted to the reuse of either work-flows or of the collected data.

The aim of this paper is to give a brief overview of the intersection of Free Software with Archaeology. This closer look draws our attention to changes in information technologies that question orthodox models of knowledge dissemination. Open Access is important not only to enable access for a much wider audience to research outcomes, but it falls on us to make the access to the data and the analysis in both transparent and reproducible ways, ideals which are closer to the ethics of scientific research. This will create a stronger research environment into the future.

[1]This paper is a revised version of a talk given during the session 10A "ArcheoFOSS: free/open source software and archaeological research, ten years later" at the CAA 2015 conference "KEEP THE REVOLUTION GOING". 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: 10.5281/zenodo.16596. A video of that lecture was recorded by Doug Rocks-Macqueen and has been published under his project **Recording Archaeology**. I would like to thank Mathias Strupler, Vera Egbers, Toby C. Wilkinson and Joseph Lehner for discussions, contributions, useful comments and suggestions on the language and structure of the manuscript. I also am grateful to two anonymous reviewer for their close reading and perceptive comments.

[2]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. For now, it is just a guess ...

[3]See the definition of the Free Software Foundation [Accessed: 30th June 2015]

Earlier Work

Papers discussing Free Software in archaeology have been published from the end of the 1990s onwards, presenting showcases, new software, or dedicated environments[4]. Benjamin Ducke recently published articles discussing the relation of Free (and Open Source) Software and archaeological software (Ducke 2012, Ducke (2013)). He focusses attention on the problem of 'black boxes' (Morin et al. 2012) and of sustainability for the development of archaeological software. Proprietary software restricts the review of methods and control over the processes, it and hinders the dissemination of the analysis. All those without a license to use specific proprietary software are *de facto* excluded from the understanding and reproduction of the analysis. In this sense, proprietary software acts like 'black boxes', where, at best, only inputs and outputs are (partially) released. As he acknowledges, Ducke does not examine the philosophical or social aspects of Free Software. In the analysis here, I would like to focus more narrowly on these aspects and how these paradigms question methods established in archaeological research without narrowing the questionnment on software.

Scientific community

science is cumulative even though it is often obscure how exactly knowledge accumulates. In any case it is a community endeavour. In this process, members of the scientific community contribute different points of views, different questions, and different methods 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 community allow knowledge to accumulate and become 'stable' faster[5].

The scientific community also creates a motivating force for recognition, respect from peers or academic prestige, and institutes controls over the quality of research by scrutinising the work of others (e.g. peer-review in journals or books review). Both represent a system of checks and balances that assures that claims are not fraudulent. The communication and the open process inside the community make claims more robust, or, contrarily, it allows the rejection of weak claims. Indeed it is not the work of a lone researcher that makes science strong and reliable, it is rather the true scrutiny and critics of peers (Fanelli 2013).

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 review, where open data is subject to validation and replication. There is a strong culture of credit, civility, reputation, and communication.

The motivations to do science or develop 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.

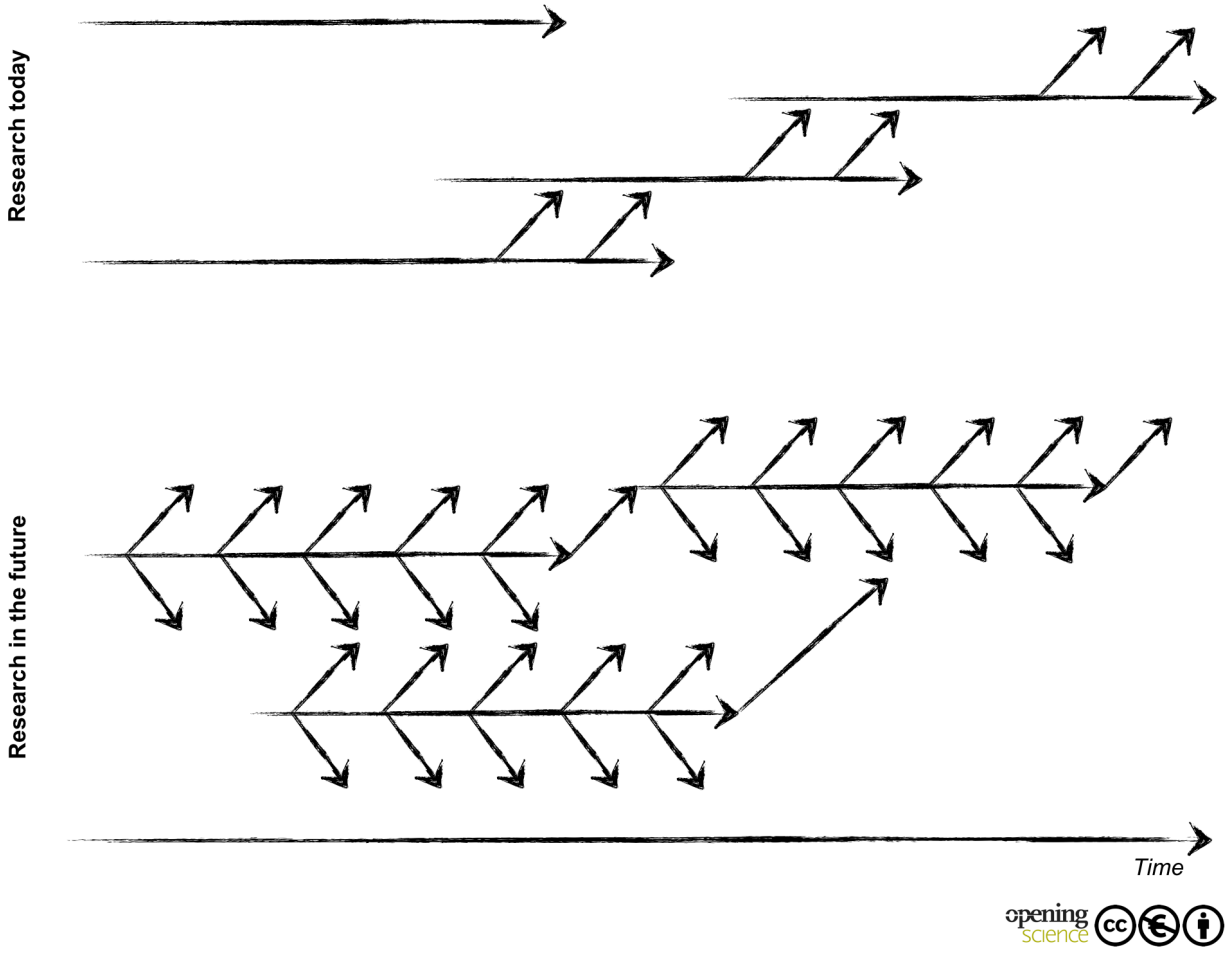
But similarities between Free Software and science should not hide differences. Dissimilarities, specifically Free Software's assets, challenge what is taken for granted in the process of accumulating knowledge.

[4]It is not the purpose to review earlier literature in this paper but the proccedings of ArcheoFOSS (Serlorenzi 2013 for the latest published) and papers published in the proceedings of the CAA are a good starting point to gain an overview about Free Software and Archaeology.

[5]"Stable knowledge" refers to the process of establishing a fact, among others Latour et al. (1979).

Modifiability

Modifiability is without any doubt the most fascinating concept among the dissimilarities between science and Free Software (Kelty 2008: 12). Books and articles have bolstered science by making stable knowledge. Yet the impression often given by books or articles is that of a final or definitive version of expertise. Free Software questions this concept with practices like forking, new versioning, cloning, and the constant evolution of source code. Every 'newly available' operating system, even proprietary, contains older pieces of software.



[Figure1] [Open and wide communication of negative as well as positive findings. From Bartling et al. 2014, Opening Science (fig. 3, 10)](http://dx.doi.org/10.1007/978-3-319-00026-8_1) | CC BY-NC 3.0

The top of this diagram [Figure1], shows a representation of the current way of doing research with stable knowledge: results are only used when they are stable. The bottom part represents how software development works and how science could be made to work. The process of sharing and reusing readily available knowledge becomes 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: 300)? In this sense Free Software can be said to challenge the 'Power of Knowledge'[6].

[6] I refer to the expression of C. Kelty in his book, Two Bits (2008)

Free Software challenges the Power of Knowledge

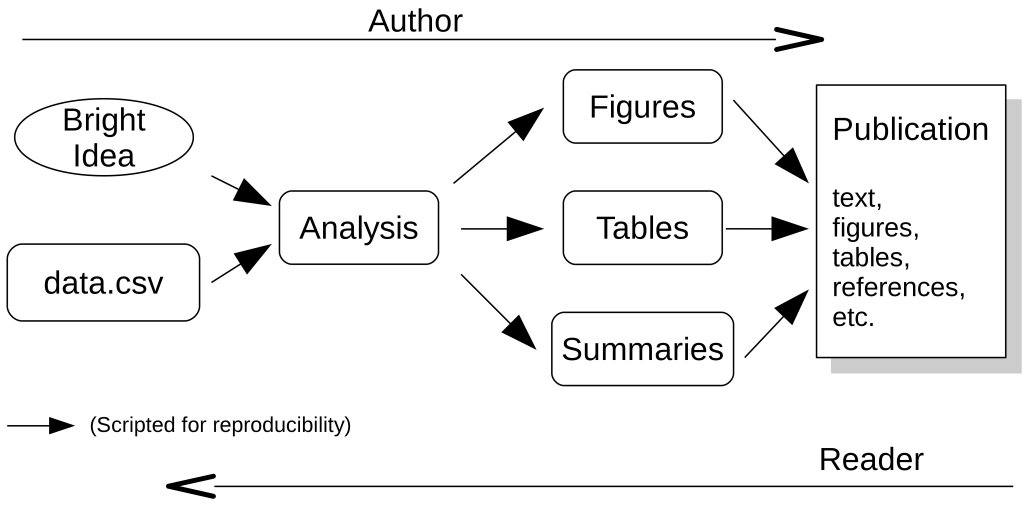
New practices of publication associated with the appearance of new information technologies render knowledge more dynamic (Vinck et al. 2014). Wikipedia is the most famous example: its content can be updated, changed, deleted, copied, or forked at any time by anyone. The last 10 years have seen a complete revaluation of Wikipedia across the academy. Ten years ago, as I saw it in France, it was mostly rejected as 'untrustable source'. Now academics explain to their students how to use it in scholarly context (Garrison 2015).

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 – *Push and Publish* – will certainly have a bright future (Kansa et al. 2014). 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 publication. 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 results are reached. This is the only way to reconcile the ethic of science and the establishment of (computational) facts. It is with this sense in mind that I will turn to the problem of reproducibility for computational archaeology.

Reproducibility

Reproducibility allows anyone to start from the same data and check all the processes undertaken, in order to verify the results. The complete process, from ideas and data to the publication has been termed as a 'research pipeline' [figure 2]. In archaeology there has been no culture of creating code for reproducibility until very recently[7]. If we turn back to the definition of science, however, reproducibility and sharing of code are even *sine qua non* for scholarly communication, permitting the community to scrutinise the work. It is 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

[7]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.

It is unrealistic to describe all the steps of an archaeological project by means of programming language. But several sub-steps - where computer plays a central role - can be made more explicit with scripts. A script is a plain text file of code composed in programming languages that instruct a computer to accomplish a set of tasks. Scripts may be augmented by comments that work like annotations. They help readers to understand the logic of the tasks. Scripts are firstly meaningful for creators to remember themselves how they obtained the results, and they record workflows that can be inspected, reused, or modified by others. Making scripts public allows reader to fully understand how results have been obtained.

For archaeological analysis, best practices permit the combination of data, code, and results through scripting and shared computational environments (Marwick 2015, (**???**)). This chain must be attained to adhere to the ethic of science. Moreover reproducibility has a huge potential for empirical research, like archaeology. Reproducibility is important 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 remain fossilised in the (outdated) state they were published.

Reproducibility helps to cast research into modules. Bits of research based on code can be transformed and reused for similar projects or built 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.

Discussion

There are undoubtly barriers inhibiting the spreading of this paradigm. Modern archaeology is largely a low-tech field and archaeologists acquire minimal computer training at the university which is not refreshed. Analysis methods and presentation of data did not dramaticaly improve for over a century. This situation, where computers are mainly used as a 'writing machine,' does not favor debates about the limitations and advantages of software or strategies to develop a research program that uses Free Software and open format as standard.

Debates about licensing, black boxes, and reproducibitliy have emerged and are primarily questioned in the computational sciences before occuring in archaeology. Furthermore, an intense advocacy for data publication and reproducibility emerged from multiple scandals about the falsification of results and after the retraction of publications. As far as I know, there is no such (known) case in archaeology and consequently these questions are not widely discussed among archaeologist. The structure of archaeological research does not favor the challenging of the current paradigm because there is only a vague distinction between the acteurs producing, analysing, publishing, and using a specific archaeological data set. More critical evaluation of the careers of individual researchers are focussed on, in particular their papers published in prestious journals, rather than also taking in account the data sets that are easily reusable by others.

Yet archaeology is increasingly dependant on software to generate results, and funding bodies and journals increasingly request publication of data. Current and future projects based on the reuse of data from other projects have to develop a workflow to aggregate, clean, analyse, and visualise a growing body of data, always gaining in resolution. By creating different modules of (reusable) scripts, the use of Free Software appears to be the best solution in the long term. Starting and maintaining a project with Free Software does not cost more time compared to proprietary software; however, the transition for propietary software to Free Software is time consuming. Therefore, it seems now critical to adopt Free Software and develop a transparent research pipeline, otherwise the current aspiration to do 'big data analysis' may turn from dream to nightmare.

Conclusion

Free Software challenges the authority of science and provides a means to transform computational archaeology and, indeed, the practice of archaeological research as a whole by questioning transmission and reuse of research. Free Software coupled with reproducibility by means of scripting makes the entire research pipeline available to the scrutiny of the community and reconciles data, processes, and results. Making the complete research available for free online has the potential to change archaeology into a research environment more robust and open to everyone.

References

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

Ducke, B. 2013. Reproducible data analysis and the open source paradigm in archaeology. In A. Bevan & M. Lake (eds) *Computational approaches to archaeological spaces*, 307–318. Walnut Creek, Calif.: Left Coast Press

Fanelli, D. 2013. Redefine misconduct as distorted reporting. *Nature* 494: 149. Available at: http://dx.doi.org/10.1038/494149a.

Garrison, J.C. 2015. Getting a 'quick fix': First-year college students' use of wikipedia. *First Monday* 20(10). Available at: http://dx.doi.org/10.5210/fm.v20i10.5401.

Heller, L., The, R., & Bartling, S. 2014. Dynamic publication formats and collaborative authoring. In S. Bartling & S. Friesike (eds) *Opening science*, 191–211. Springer Available at: 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: 57–70. Available at: http://dx.doi.org/10.2218/ijdc.v9i1.301.

Kelty, C.M. 2001. Free software/free science. *First Monday* 6(12). Available at: http://pear.accc.uic.edu/ojs/index.php/fm/article/view/902.

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

Latour, B., & 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. Available at: http://www.software.ac.uk/blog/2015-04-24-archaeology-open-source-software-its-getting-easier.

Morin, A. et al. 2012. Shining light into black boxes. *Science* 336(6078): 159–160. Available at: http://dx.doi.org/10.1126/science.1218263.

Serlorenzi, M. ed. 2013. *ARCHEOFOSS free, libre and open source software e open format nei processi di ricerca archeologica. atti del vii workshop (roma, 11-13 giugno 2012)*. All'Insegna del Giglio. Available at: http://www.progettocaere.rm.cnr.it/databasegestione/open_block_pages_sup.asp?IDyear=2013-01-01.

Vinck, D., & Clivaz, C. 2014. Les humanités délivrées. savoir et culture réinventés hors du livre. *Revue d'anthropologie des connaissances* 8.4: 681–704. Available at: http://dx.doi.org/10.3917/rac.025.0681.