



University
of Glasgow

Marwick, B. et al. (2017) Open science in archaeology. *SAA Archaeological Record*, 17(4), pp. 8-14.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/148887/>

Deposited on: 29 September 2017

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Open Science in Archaeology

Ben Marwick*, Jade d'Alpoim Guedes, C. Michael Barton, Lynsey A. Bates, Michael Baxter, Andrew Bevan, Elizabeth A. Bollwerk, R. Kyle Bocinsky, Tom Brughmans, Alison K. Carter, Cyler Conrad, Daniel A. Contreras, Stefano Costa, Enrico R. Crema, Adrienne Daggett, Benjamin Davies, B. Lee Drake, Thomas S. Dye, Phoebe France, Richard Fullagar, Domenico Giusti, Shawn Graham, Matthew D. Harris, John Hawks, Sebastian Heath, Damien Huffer, Eric C. Kansa, Sarah Witcher Kansa, Mark E. Madsen, Jennifer Melcher, Joan Negre, Fraser D. Neiman, Rachel Opitz, David C. Orton, Paulina Przystupa, Maria Raviele, Julien Riel-Salvatore, Philip Riris, Iza Romanowska, Néhémie Strupler, Isaac I. Ullah, Hannah G. Van Vlack, Ethan C. Watrall, Chris Webster, Joshua Wells, Judith Winters, Colin D. Wren

* corresponding author, bmarwick@uw.edu

Introduction

In archaeology, we are accustomed to investing great effort into collecting data from fieldwork, museum collections, and other sources, followed by detailed description, rigorous analysis, and in many cases ending with publication of our findings in short, highly concentrated reports or journal articles. Very often, these publications are all that is visible of this lengthy process, and even then, most of our journal articles are only accessible to scholars at institutions paying subscription fees to the journal publishers. While this traditional model of the archaeological research process has long been effective at generating new knowledge about our past, it is increasingly at odds with current norms of practice in other sciences. Often described as 'open science', these new norms include data stewardship instead of data ownership, transparency in the analysis process instead of secrecy, and public involvement instead of exclusion. While the concept of open science is not new in archaeology (e.g., see Lake 2012 and other papers in that volume), a less transparent model often prevails, unfortunately. We believe that there is much to be gained, both for individual researchers and for the discipline, from broader application of open science practices. In this article, we very briefly describe these practices and their benefits to researchers. We introduce the Society of American Archaeology's Open Science Interest Group (OSIG) as a community to help archaeologists engage in and benefit from open science practices, and describe how it will facilitate the adoption of open science in archaeology.

What is Open Science?

Openness in science is significant in that it both defines the origins of modern science, and imagines the future of science (Fecher and Friesike 2014). In their review of discussions of open science, Fecher and Friesike (2014) identified five themes: infrastructure (i.e., creating tools and services to improve research efficiency); the public (i.e., making science accessible for non-scientists); measurement (i.e., developing alternative metrics to measure the impact of research); democracy (i.e., making knowledge freely accessible to all); and pragmatics (i.e.,

making collaborative research more efficient). The broader public benefits of advancing open science has been widely discussed (OECD 2015), and we will not expand on those here. Instead, we take a researcher-centric approach, drawing on our experience as practicing archaeologists to focus on specific examples of openness that can offer maximum benefit for researchers. From this perspective, we have identified three elements of open science that cross-cut Fecher's and Friesike's themes: open access, open data, and open methods.

Open access

Open access typically refers to the permanent online access to the full text of scholarly work, especially publications, without charge to readers or libraries. There are many ways to accomplish this: for example, 'Gold Open Access' refers to the author paying a fee for publication (typically referred to as an article processing charge or APC). The fee is intended to defray the cost of publication that the publisher would recoup through institutional subscriptions. However, these APCs can be quite expensive and often deter researchers from granting access to their publication. This particularly affects researchers in developing countries, authors from traditionally underrepresented groups, early-career researchers and those in disciplines, such as archaeology, where article subvention fees are not commonly awarded in research grants (although some journals offer waivers). An alternative approach, referred to as 'Green Open Access', is for authors to make their manuscripts freely available online as preprints prior to journal publication (Figure 1). An advantage of Green Open Access is that it is free for authors to submit and free for readers to access (the preprint of this paper containing additional citations is online at doi.org/10.17605/OSF.IO/3D6XX).

Notable examples of disciplinary-oriented preprint repositories are [arXiv.org](https://arxiv.org), a repository for physics, mathematics, computer science, astronomy, and related papers and [bioarXiv.org](https://bioarxiv.org) for biomedical and life sciences. In fact, some biology funding sources require preprints to be deposited prior to publication. Preprint repositories commonly used by archaeologists include socarxiv.org and papers.ssrn.com, both of which specialize in the social sciences. We note that academia.edu and researchgate.net are popular for sharing articles online, however these are private, for-profit companies that do not own the rights to host most of their content (and so vulnerable to legal action) and require registration to access. These should not be considered substitutes for a preprint repository. Most research-intensive universities have their own open access repositories to enable their researchers to disseminate their work as preprints. Many journals allow researchers to post preprints of their published articles, giving researchers a wider choice of journals in which to publish (compared to the small number of Gold Open Access journals), while still enabling open access. The individual policies of specific journals can be checked online at the [SHERPA/RoMEO](https://www.sherpa.ac.uk/romeo/) database. Open access publications benefit researchers because they typically achieve increased impact by being cited more frequently and receiving more media coverage (see McKiernan, et al. 2016 for a summary of empirical work on this topic). Researchers may also benefit from their publications being easily accessible to prospective students and non-academic collaborators, such as local and indigenous communities.

Open data

Open data means open access to datasets. Data can take many forms, here we refer to items such as a spreadsheet of artifact measurements or a GIS layer of site locations and attributes — the information used to make the summary tables and plots that typically appear in reports and publications. Traditionally, archaeologists have viewed datasets as their proprietary products, and having paid a high upfront cost to collect the data, they hope to recover that cost through publications based on exclusive access to those data. In many fields, this data-ownership mindset is viewed as obsolete, and has been replaced by the idea of data stewardship (Hampton, et al. 2015). Data stewardship advocates that researchers collect and share data on behalf of the scientific community and society, rather than for their individual career ambitions. These norms can be seen in the policies of funding agencies (e.g., Wellcome Trust, Bill and Melinda Gates Foundation, National Endowment for the Humanities, and the National Science Foundation) and journals (e.g., *PLOS*, *Evolution*, *Scientific Data*, Royal Society journals), that require researchers to share their data with other investigators by depositing the data in a public repository. Substantial technology and infrastructure has appeared to accommodate the data availability requirements of these funding agencies and journals. A comprehensive list of repositories (many of which are free to use) for various fields is available at www.nature.com/sdata/policies/repositories. Examples of repositories specifically for archaeological data include opencontext.org, tdar.org, and archaeologydataservice.ac.uk, among others. The attributes of trustworthy data repositories include: having an explicit mission to provide access to and preserve data, offering appropriate licenses covering data access and use (e.g. CC-0), having a continuity plan to ensure ongoing access and preservation of its holdings, guaranteeing the integrity and authenticity of the data (e.g. by using version control), and enabling users to discover the data and refer to them in a persistent way through proper citation (e.g. with a DataCite DOI). Using a trustworthy data repository is important for ensuring ongoing availability of data because direct requests to researchers for their privately-held data often fail (Vines, et al. 2014).

Providing open access to data is more challenging than opening access to publications because of the potential for harm to people and cultural heritage that can result from misuse of the data or the release of sensitive information (such as personally identifiable data or detailed site locations). Opening data also requires consideration of intellectual property ownership, especially for archaeologists working in large teams, in commercial and government sectors, and/or with indigenous/descendant communities. Many of these ethical issues can be addressed by negotiation, legal instruments (such as Creative Commons licenses) or technical solutions, for example, redacting portions of data, limiting spatial precision (an approach used successfully in projects such as The Digital Index of North American Archaeology), restricting access, or imposing embargos. Of course, researchers must be vigilant in comprehensively addressing any negative impacts prior to opening their data to public access. Nevertheless, our experience is that for most archaeologists it will not be burdensome to share the minimal data

behind the tables and figures in their journal articles, or even the more detailed original and unaggregated records. Indeed, many archaeologists already do this routinely via supplementary online material for their journal articles. Similar to open access, there are citation advantages and an increase in the impact of their work for researchers who share the data behind their publications (see McKiernan, et al. 2016 for discussion of the empirical research). There are also benefits to other archaeologists from opening access to data. For example, researchers can find their past research data more easily when it is publicly available at a reputable repository, and in our experience with our own research, data are likely to be better documented and easier to reuse when it is prepared for public access.

Open methods

Open methods are methods of data collection, analysis, and visualization that are available for inspection, and reuse by the public. This approach can include empirical methods (e.g., the details of chemicals used to prepare samples) and computational and statistical methods (e.g., the details of taking raw data and producing statistical tests, models, and visualizations). Open methods are important for improving the reproducibility of research, that is, the ability to redo a study, with the same materials and methods, and get the same result, which is a cornerstone of science (Stodden et al. 2014). This is because the complexity of most current research, especially computational and statistical methods, means that a typical journal article is too short to communicate enough details to enable reproducibility. Open methods have emerged in other fields in response to highly-publicized failures to reproduce the results of notable studies in biomedicine, psychology, genomics, political science, and economics.

This has resulted in extensive discussion of how to improve reproducibility across many fields (e.g., Goodman et al. 2016; Munafò et al. 2017; Sandve et al. 2013; Stodden et al. 2016; Stodden and Miguez 2014; Wilson, et al. 2014), including archaeology (Marwick 2016). These discussions have converged on a few frequently recommended practices (Figure 2), including: using a transparent software environment that enables reproducibility (such as R or Python, Figure 3) to analyze data rather than software whose analytical algorithms are proprietary black boxes (e.g., Excel, SPSS, PAST); using a version control system that can efficiently track and log changes and simplify collaboration (such as Git, similar to ‘track changes’ in Microsoft Word, or ‘revision history’ in Google Docs); using open source licenses to make the code maximally available for reuse while ensuring recognition of effort (such as Apache, MIT, or GPL licenses); and archiving these methods at trustworthy repositories where they are freely accessible (e.g. R or Python script files deposited at osf.io, zenodo.org or figshare.com). These recommendations simplify the task of making our scientific workflows available at the time of publication, and so streamline the task of making the methods available for public inspection and reuse.

How is open science relevant to archaeologists?

These three practices — open access, open data, and open methods — are relevant to three of the goals of the Society of American Archaeology (SAA).

First, the SAA ‘advances archaeological research and disseminates archaeological knowledge.’ Open science practices support this goal by encouraging archaeologists to conduct research that is transparent, reusable, and easily accessible (open data and open methods) without financial or copyright barriers (open access). The Open Science Interest Group will help to educate archaeologists about options for using software that enables reproducibility, generating scripted workflows, using environments for version control and collaborative analysis, making data and preprints available through public repositories, and publishing research in open access journals.

Second, the SAA ‘improves the practice of archaeology and promotes archaeological ethics.’ Open science practices improve archaeology by increasing transparency and reproducibility in archaeological research. This approach enables archaeologists to more readily and responsibly build on the work of their colleagues, advancing archaeological practice and accelerating discovery. Transparency and reproducibility also enhance the credibility of archaeological research by allowing more complete independent assessment of research findings than is possible with traditional peer review of only research results. Open science practices promote ethical research by enabling researchers to efficiently demonstrate the chain of reasoning behind their data analysis, and expose more of their research workflow to the research community and the public. The OSIG will help to educate archaeologists on how to improve their research, and the field of archaeology more broadly, with open science tools and methods.

Third, the SAA ‘serves as a bond among archaeologists worldwide in all segments of the archaeological community.’ Community best practices for open science in archaeology facilitate the sharing of methods, data, and results by encouraging researchers to deposit them in trustworthy online repositories. Standardizing research sharing practices enhances engagement between archaeologists, our collaborators, and the communities we work with, including policymakers and project managers. Open science practices promote inclusiveness because they remove financial, institutional and other barriers from researchers engaging with each other, and with methods and data.

In addition to advancing the goals of the SAA, the OSIG will help archaeology contribute to the open science movement that has become part of normal scientific practice in many fields. For example, members of the Ecological Society of America, the European Geosciences Union, and the Organization for Human Brain Mapping, have organized open science sections to help researchers benefit from openness. Similarly, formal open science policies have been developed by the [Association for Psychological Science](#), and the [American Heart Association](#). Some scholarly journals and conferences require submissions to be subjected to reproducibility reviews before they are accepted for publication (e.g., Association for Computing Machinery's Special Interest Group on Management of Data, American Journal of Political Science, Quarterly Journal of Political Science). Many journals require data to be openly accessible as a condition of publication, and some journals reward authors for following open access practices (e.g., *Biostatistics*, *Psychological Science*). In archaeology we have a small number of similar

initiatives, such as the open badges policy at [Internet Archaeology](#), to certify when open practices were followed by the authors of a journal article, and the [Journal of Open Archaeology Data](#), which publishes descriptions of datasets hosted on trusted repositories.

What is the Open Science Interest Group doing? What can you do?

The mission of the OSIG is twofold: (1) advance transparency and accessibility in the ways archaeologists and institutions manage data, methods, and research outputs; and, (2) share information with individuals and institutions on how to develop open practices that enable reproducible research. We will endeavor to fulfil this mission through our individual research practices and through activities as an SAA Interest Group. As researchers, we will strive to make our research more reproducible, and to influence others to do the same through the following practices:

- Generating and making accessible explicit or scripted, reproducible workflows for our data analysis. To the extent possible, we will employ transparent and accessible analytical tools and software (such as R, Python, and other programming languages) so that our research can be easily evaluated by others (Marwick, 2016).
- Requesting data and code when we review manuscripts, and when in editorial positions, advocating for data and code review as a part of standard peer review practices at a journal (Stodden, et al. 2013)
- Including—and following through on—comprehensive data management plans in all research designs
- Teaching our students and mentees to work reproducibly and openly (Marwick, 2016)
- Archiving our papers or preprints in open access repositories (McKiernan, et al. 2016)
- Archiving our research data and code in trustworthy repositories and citing these archives in our published work using DOIs (McKiernan, et al. 2016)

These actions align with recent recommendations for increasing openness and reproducibility in science generally (Miguel, et al. 2014; Nosek, et al. 2015; Stodden, et al. 2016). We recognize that there are different degrees and dimensions of openness that are available to researchers, depending on their circumstances and skills. Thus, not all of us can take these actions all the time, but through the aggregate of our individual actions we can improve archaeological research practice toward the norms of open science.

As an SAA Interest Group, we have identified two initial activities relevant to our mission. Our first activity is to incentivize open practices, by issuing Center for Open Science (COS) badges for Open Data and Open Materials (osf.io/tvyxz), for display on qualifying posters and slide presentations at the SAA Annual Meetings and other professional venues. These badges are used in many disciplines, and have been shown to increase data sharing (Kidwell, et al. 2016). We also will work with the SAA publication committee and other archaeological journals to explore how COS badges can be applied to journal articles. The second activity is to conduct workshops using [Software Carpentry/Data Carpentry](#) pedagogy and materials. These

workshops aim to train researchers to use open science tools so that they can work more efficiently, reproducibly, and openly. We will offer these workshops in-person at SAA meetings, online via the SAA webinar series, and elsewhere. We will also host and sponsor traditional SAA meeting events to foster the exchange of ideas and community interaction, as well as collaborate with related communities, such as the Digital Data Interest Group and Public Archaeology Interest Group.

We invite all archaeologists to join us in becoming more responsible researchers by following the individual best practices for open science listed above to the benefit of all members of the archaeological community, other scientists, and the public more broadly. To support open, transparent, and reproducible science in archaeology as a member of the SAA Open Science Interest Group, please subscribe to our email list at <https://groups.google.com/group/saa-osig/subscribe> for updates. The OSIG web site, at: osf.io/2dfhz/, contains further information about the group, resources, and details of news and updates.

Summary

In this article we have briefly surveyed the goals and best practices of current open science initiatives, and identified specific practices that have been shown to benefit individual researchers, as well as science more broadly. These are: (1) increasing open access publication by depositing preprints; (2) accompanying published articles with open datasets deposited in trustworthy repositories; and (3) creating and making available transparent and reproducible scientific workflows, including relevant code, along with published research. We have outlined how these practices are relevant to archaeologists and how they advance the goals of the Society of American Archaeology. In addition, we have described some of the activities of the SAA Open Science Interest Group and explained how they will help to make open science more a part of normal archaeological practice.

We recognize that many archaeologists may be unfamiliar with open science practices, and could initially imagine that incorporating these practices into their normal work might entail additional investment of time, effort, and other resources. While depositing preprints can be a quick and simple action, learning a new program for data analysis requires considerably more effort (although our experience has been that learning an open source program like R or GRASS is little or no more difficult than initially learning any other complex software like SPSS or ArcGIS). To address this, the OSIG plans to offer training workshops to speed the adoption of open methods. These workshops will initially include the open source statistical programming language R, the version control system git, and the use of data repositories. In the long run, we believe that use of scripted workflows in environments like R and Python actually *improves* researcher efficiency considerably, while using open source software significantly reduces licensing costs. Similarly, some archaeologists may fear the limitations to publication potential that could result from others using their open data and code, the possibility that their materials may be used without citation, and the risk that competitors may gain an advantage. Our view is

that these risks have always been present in the traditional research practices of scholarly communication and peer review, and that open science licensing and citation practices effectively mitigate them. Moreover, because sharing of data and code enables and encourages collaborative research, more open science practices can even increase the potential for new research (and publications) with extant data -- an important benefit to junior researchers in particular. Overall, we believe any costs for the practice of open science are well worth the many substantial benefits it brings to archaeologists and the archaeological community.

Acknowledgements

Thanks to Tobi Brimsek for her advice on the initial steps of organising this group. BM conceived and wrote the paper and figures, the other authors (listed in alphabetical order) edited the text, and supported and endorsed the formation of the OSIG. A version of this document containing extensive citations, hyperlinked text and marginal discussions among the authors is online at <http://bit.ly/TSAR-OSIG>

References cited

Fecher, Benedikt and Sascha Friesike

2014 Open science: one term, five schools of thought. In *Opening science*, pp. 17-47. Springer.

Goodman, Steven N., Daniele Fanelli and John P. A. Ioannidis

2016 What does research reproducibility mean? *Science Translational Medicine* 8(341):341ps312.

Hampton, Stephanie E, Sean S Anderson, Sarah C Bagby, Corinna Gries, Xueying Han, Edmund M Hart, Matthew B Jones, W Christopher Lenhardt, Andrew MacDonald and William K Michener

2015 The tao of open science for ecology. *Ecosphere* 6(7):1-13.

Kidwell, Mallory C., Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, Agnieszka Slowik, Carina Sonnleitner, Chelsey Hess-Holden, Timothy M. Errington, Susann Fiedler and Brian A. Nosek

2016 Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology* 14(5):e1002456.

Lake, Mark

2012 Open archaeology. *World Archaeology* 44(4):471-478.

Marwick, Ben

2016 Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. *Journal of Archaeological Method and Theory*:1-27.

McKiernan, Erin C., Philip E. Bourne, C. Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, Brian A. Nosek, Karthik Ram, Courtney K. Soderberg, Jeffrey R. Spies, Kaitlin Thaney, Andrew Updegrave, Kara H. Woo and Tal Yarkoni

2016 How open science helps researchers succeed. *eLife* 5:e16800.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn and M. Van der Laan
2014 Promoting Transparency in Social Science Research. *Science* 343(6166):30-31.

Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware and John P. A. Ioannidis
2017 A manifesto for reproducible science. *Nature Human Behaviour* 1:0021.

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson and T. Yarkoni
2015 Promoting an open research culture. *Science* 348(6242):1422.

OECD

2015 *Making Open Science a Reality*. OECD Science, Technology and Industry Policy Papers, No. 25. OECD Publishing, Paris.

Peng, R. D.

2011 Reproducible Research in Computational Science. *Science* 334(6060):1226-1227.

Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor and Eivind Hovig

2013 Ten Simple Rules for Reproducible Computational Research. *PLoS Computational Biology* 9(10).

Stodden, Victoria, Peixuan Guo and Zhaokun Ma

2013 Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. *PloS one* 8(6):e67111.

Stodden, Victoria, Marcia McNutt, David H. Bailey, Ewa Deelman, Yolanda Gil, Brooks Hanson, Michael A. Heroux, John P. A. Ioannidis and Michela Taufer

2016 Enhancing reproducibility for computational methods. *Science* 354(6317):1240.

Stodden, Victoria and Sheila Miguez

2014 Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. *Journal of Open Research Software* 2(1).

Wilson, Greg, D. A. Aruliah, C. Titus Brown, Neil P. Chue Hong, Matt Davis, Richard T. Guy, Steven H. D. Haddock, Kathryn D. Huff, Ian M. Mitchell, Mark D. Plumbley, Ben Waugh, Ethan P. White and Paul Wilson
2014 Best Practices for Scientific Computing. *PLoS Biol* 12(1).

Figures and captions:

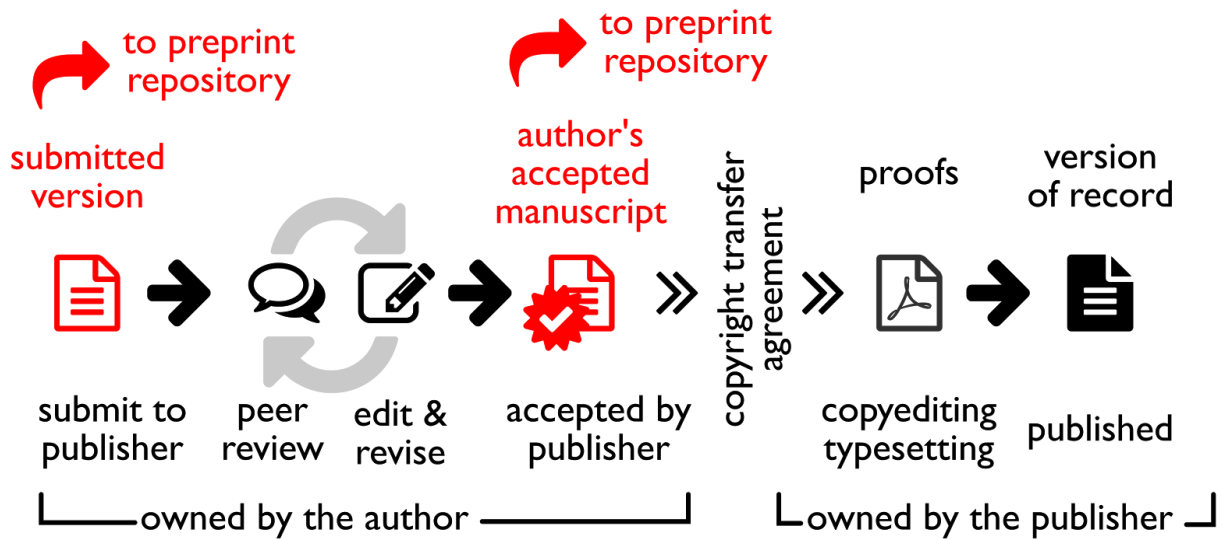
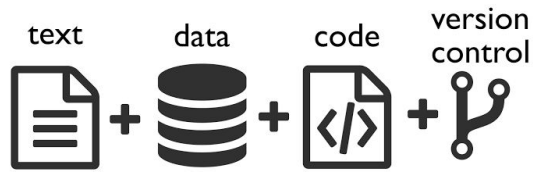


Figure 1. Preprints and the typical cycle of scholarly journal article publication. Typically, manuscripts are submitted to a preprint repository at the same time they are submitted to a journal for peer review or after they have been accepted for publication (but before proofs are prepared). The preprint can be updated by the author, with versions tracked by the repository. The submitted version and the author's accepted manuscript are owned by the author, so these can be posted to a preprint repository without copyright infringement. After the author signs the copyright transfer agreement, the versions of the paper produced by the publisher are not owned by the author. For example, the proofs and published version are owned by the publisher, so in most cases the author is not legally permitted to make these publicly available in preprint repositories or elsewhere (e.g. academia.edu and researchgate.net).

advertising:
text & final
results only



science:
text, code &
data available,
linked & licensed



Adapted with permission from Rodriguez Escobar S, Pérez-López AJ, Bermejo Utrera S (2014) Codes reproducibility: why you should care. *Communications* 25(2): 83-91. <http://dx.doi.org/10.7918/ECOL2014.25.2.11>. See also Merriën S (2018) Computational Reproducibility in Archaeological Research: Best Practices and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory* 25(1): 1-27. <https://doi.org/10.1007/s10816-018-9127-9>. This figure is CC-BY.

Figure 2. The reproducible research spectrum. Reproducibility is not a binary quality but a spectrum (Peng 2011). Scientific articles that contain only the final text, results and figures (e.g., in a single pdf document) are advertising a finding, and these are the least reproducible -- it is often impossible to reconstruct the whole analytical process from data to results. Publication of the data and/or code used for the analysis greatly improves reproducibility. Similarly, using a version control system (such as git) permits navigating through the complete history of the project. Finally, the most reproducible, and thus scientific, studies are those using dynamic reports (e.g., Rmarkdown notebooks) that integrate text, code and data into an executable environment.

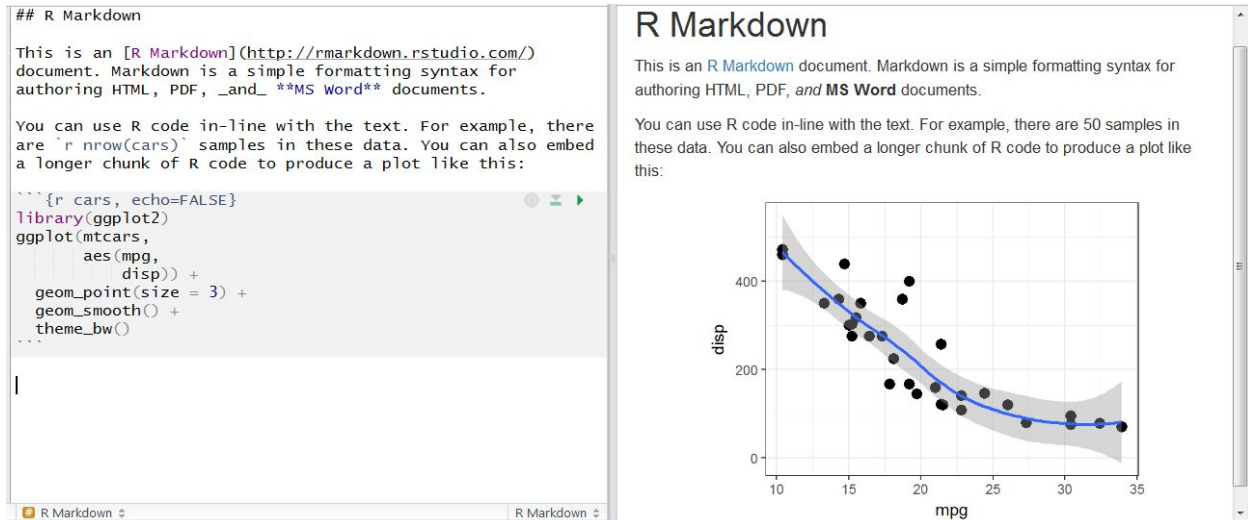


Figure 3. A screenshot from the RStudio program showing how R can be used for reproducible research. In the left panel is a text editor, where we write plain text and code in an R markdown file (known as a Rmd file). In the right panel is the output that is produced when the Rmd file is 'knit', or rendered, into a document. In this example, the Rmd has been knitted to produce a HTML file, but we could also produce a PDF or Microsoft Word document from the same Rmd file. The first paragraph of the text in the example demonstrates how to use markdown for basic text formatting (e.g. a heading, a URL, bold and italic text). The second paragraph shows how R code can be embedded inline in the text. The rendering process automatically runs the code and inserts the result in the text, here it computes the number of rows in the 'cars' dataset, and inserts the result (50) in the rendered document. The text in the grey region on the left is a chunk of R code that produces the plot in the HTML file on the right. We use `echo=FALSE` in the code chunk to specify that the code chunk is not displayed in the HTML file; we see only the plot that the code generates. This method of writing text and code in the same document enhances reproducibility because the methods of data analysis (i.e. the R code) are explicitly included in the same document as the text, and the code can be easily and repeatedly run to generate results. This removes the need to copy and paste tables and plots from other software into the text, eliminating transcription errors and confusion about where a particular result came from.