Archeologia e Calcolatori
35.2, 2024, 75-84
doi 10.19282/ac.35.2.2024.09

THE GOOD, THE BAD AND THE UGLY: EVALUATING OPEN SCIENCE PRACTICES IN ARCHAEOLOGY

1. Introduction: disentangling Open Access and Open Science

In contemporary discourse, distinguishing between Open Science (OS) and Open Access (OA) can often prove challenging[1]. While OA constitutes just one facet of the broader OS framework, it entails unfettered online access to research findings and data, as stated by seminal documents such as the 2002 Budapest Open Access Initiative (BOAI) (https://www.budapestopenaccessinitiative.org/read/) and the 2003 Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (https://openaccess.mpg.de/Berlin-Declaration). Despite the noble intentions driving the OA movement – such as democratising the production and consumption of scholarly outputs – critical reflections emerged in the wake of its fervent inception.

It is not a case that in 2022 the BOAI, in the occasion of its 20th anniversary, declared that «Open access is not an end in itself, but a means to further ends. Above all, it is a means to the equity, quality, usability, and sustainability of research» (https://www.budapestopenaccessinitiative.org/boai20/). Consequently, the BOAI steering committee recognised that OA practices inadvertently exacerbated the digital divide between well-funded institutions of the Global North and their under-resourced counterparts in the Global South. In response, the committee delineated four main strategies to mitigate this inequity:1) hosting OA research on open infrastructure owned or managed by non-profit organisations; 2) reforming research assessment criteria for funding allocations and academic promotions; 3) fostering inclusive publishing and distribution channels that never exclude authors on economic grounds; and 4) advocating for dissemination models benefitting all regions of the world, under the stewardship of academic-led and non-profit organizations.

In this context, OS has been proposed as a sort of corrective for some of the issues highlighted above because it aims to reform research through more transparent, equitable, inclusive, and collaborative practices (Vicente-Saez, Martinez-Fuentes 2018; UNESCO 2021). However, OS remains

[1] This paper has benefited from the research funding provided by the Italian Ministry of Research through the Rita Levi Montalcini program, as well as by the Gerda Henkel Foundation. The digital archive associated with this paper facilitates the reproducibility of Figs. 1 and 3 through two datasets and a script written in the R statistical computing language. This digital archive is openly accessible via GitHub (https://github.com/apalmisano82/Open_Science) and Zenodo (https://doi.org/10.5281/zenodo.10971326).

a multifaceted and somewhat nebulous concept, encompassing diverse interpretations and it can be perceived as a movement, a goal, a framework of policies, or even a research strategy (Fecher, Friesike 2014).

Furthermore, with the exception of a few archaeologists familiar with quantitative and computational approaches, OS is only slowly reaching mainstream archaeology. In the realm of archaeology, OS predominantly revolves around the principles of open data sharing and open access publications. However, compared to the practices observed in the natural sciences, aspects like open methods and reproducibility have received relatively less attention. This disparity underscores a significant gap between archaeology and scientific disciplines.

The adoption and development of OS practices have exhibited considerable variation across different countries. Notably, the United Kingdom and the United States have emerged as frontrunners in this regard. This prominence is partly attributed to the classification of archaeology as a discipline within the social sciences domain in these countries, as opposed to its categorisation under the humanities in many other countries.

Hence, there exists a pressing need to reverse a prevailing trend within archaeology. In this paper, we endeavour to offer a comprehensive and foundational introduction specifically tailored to researchers with an interest in OS practices within the field of archaeology. Our aim is to critically evaluate the current state of the art concerning open data, methods, and publication practices within archaeology. Additionally, we will provide insightful recommendations for applying OS practices in the archaeological research.

2. Open data

The field of archaeology has experienced a transformative shift in recent years, spurred by a growing deluge of novel digital data about the human past accessible through online databases and open repositories (Bevan 2015). Within archaeology, a diverse range of data emerges from fieldwork activities and the digitisation of information from published works. As recently pointed out by B. Marwick and S.E. Pilaar Birch (2018), archaeology is a poor-data field affected by restricted data-sharing practices and the lack of standardised methods for sharing and citing data. Particularly concerning is the reluctance of authors to share their data, a phenomenon known as 'data-hoarding', driven by fears of losing opportunities for future publications and the reluctance to invest time in organising data for their reuse (Marwick, Pilaar Birch 2018, 129-131).

There are three common ways to share data in archaeology: databases, supplementary materials and online DOI-issuing repositories. The first approach poses significant challenges in our field due to ambiguity surrounding

what constitutes data sharing among most archaeologists. Often, the visually appealing and multifunctional interfaces of online databases are mistakenly perceived as facilitating data sharing. However, in many cases, the stored information cannot be downloaded and reused, necessitating users to manually digitise and gather data anew. Noteworthy examples are the MedAfriCarbon (https://theia.arch.cam.ac.uk/MedAfriCarbon/) and NEONET (http://shinyserver.cfs.unipi.it:3838/C14/) databases that allow the download of raw data (for detailed descriptions refer to Lucarini *et al.* 2020; Huet *et al.* 2022).

Sharing data as supplementary materials attached to journal articles is another common mistake as files are not accessible to nonsubscribers for paywalled journals. Moreover, supplementary materials are often presented in tables stored in binary file formats like .PDF or Excel, impeding data reuse. Instead, sharing data through publicly accessible online repositories is the recommended practice (Marwick, Pilaar Birch 2018, 133-135). Notable repositories tailored for archaeologists include the Archaeology Data Service (https://archaeologydataservice.ac.uk/) and the Digital Archaeological Record (https://core.tdar.org/). While these repositories may entail reasonable fees for data maintenance and curation, users also have the option of utilizing free repositories such as the archaeology-focused Mappa Open Data (MOD) archive (https://digitallib.unipi.it/it/archivio/MOD-Mappa-Open-Data-archive/), as well as discipline-agnostic repositories such as Zenodo (https://zenodo.org/) and Figshare (https://figshare.com/). Furthermore, institutional repositories of universities serve as valuable resources for disseminating data at no cost, showcasing and providing open access to research outputs.

Bearing in mind the aforementioned issues, here we present a small pilot-study aiming at understanding how archaeologists share research data. We analysed a total of 621 publications from four methodology-based archaeology journals published in 2020-2021: «Journal of Archaeological Science» (JAS, n = 300), «Journal of Archaeological Method and Theory» (JAMT, n = 53), «Journal of Computer Applications in Archaeology» (JCAA, n = 30) and «Plos One» (n = 238). This pilot study was limited in scope, as it focused on a brief time frame due to the need for meticulous examination of numerous publications to determine if they were based on research that could lead to the release of a dataset and workflow. The sample does not include review articles and commentaries, focusing solely on research articles utilizing data and methods. In the case of the multidisciplinary journal «Plos One», only articles within the field of archaeology were considered. The results (Tab. 1) indicate that the 46.7% of articles share data either as supplementary materials (34.6%) or via a public DOI-issuing online repository (12.1%).

It is alarming that the majority of archaeologists inclined to share their data opt not to use online repositories. Upon further examination of each journal, it becomes evident that «Plos One» stands out, with approximately

	Total articles	Sharing data	Supplementary materials	DOI-issuing Online repositories	Reproducible research
Journal of Archaeological Science (JAS)	300	122 (40.6 %)	100 (33.3%)	22 (7.3%)	21 (7%)
Journal of Archaeological Method and Theory (JAMT)	53	19 (35.8 %)	16 (30.2 %)	3 (5.6 %)	7 (13.2 %)
Journal of Computer Applications in Archaeology (JCAA)	30	15 (50 %)	2 (6.6 %)	13 (43.4%)	11 (36.6 %)
Plos One	238	134 (56.3 %)	97 (40.8 %)	37 (15.5 %)	37 (15.5 %)
Total	621	290 (46.7) %	215 (34.6 %)	75 (12.1 %)	76 (12.2 %)

Tab. 1 – Summary of data sharing and reproducibility practices between 2020-2021.

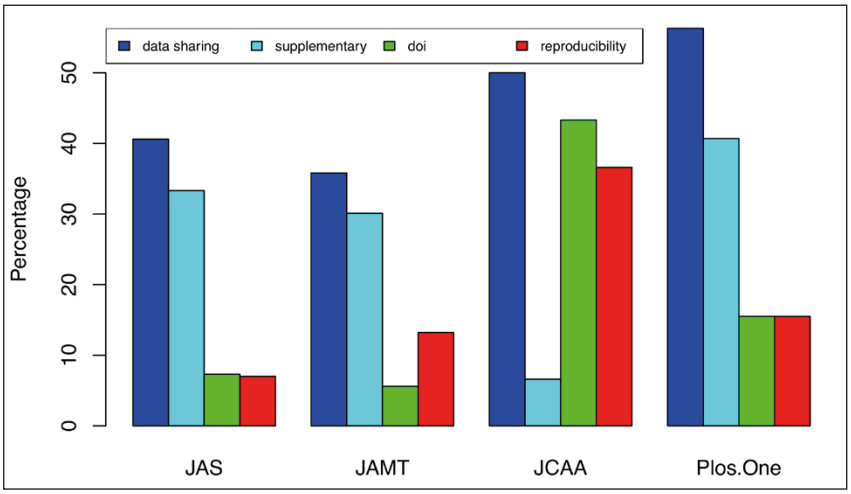

Fig. 1 – Graph showing data sharing and reproducibility practices between 2020-2021.

60% of its published archaeology articles sharing data, largely due to the journal's policy mandating authors to make their data publicly available (Fig. 1). However, it is disappointing to note that across all journals, except for «JCAA», authors tend to share data primarily as supplementary material. «JCAA» counters this trend, likely due to the authors' strong computational skills and familiarity with open science practices.

Another recommendable practice gaining traction among archaeologists is the dissemination of datasets through the publication of data papers. These papers offer comprehensive insights into the creation, structure, methodology, and potential applications of a dataset. Such publications are invaluable as they enable researchers to share their datasets within their scientific communities and to receive recognition for their efforts in dataset creation and dissemination (Hole

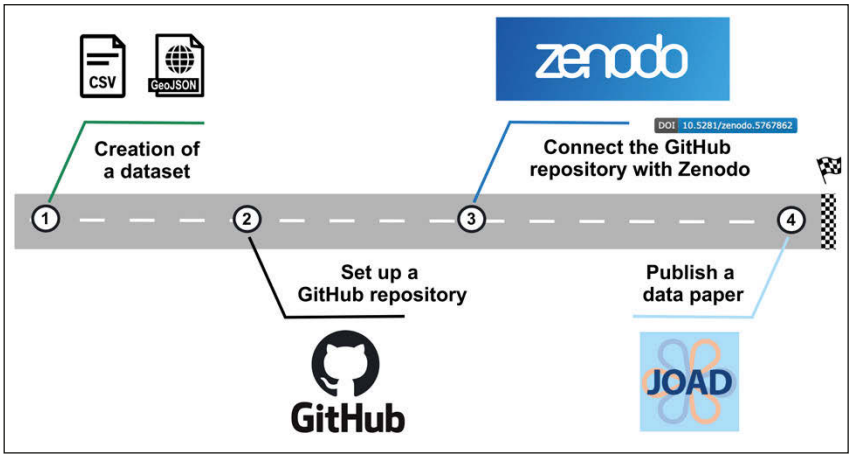

Fig. 2 – The roadmap for disseminating openly a dataset.

2012). Associating a citable publication with a specific dataset could incentivise archaeologists to share their data and alleviate data hoarding tendencies. Within the field of archaeology, there are two prominent data journals: the «Journal of Open Archaeology Data» (https://openarchaeologydata.metajnl.com/) and «Archeologica Data» (https://www.mappalab.eu/archeologica-data-2/). Moreover, the publication of a data paper should be just the final step in a workflow aiming at making a dataset FAIR (Fig. 2): 1) creation of the dataset; 2) publishing it via GitHub; 3) linking the GitHub repository to Zenodo to archive the dataset by assigning a DOI and ensuring that all versions of the dataset are tracked in both repositories; 4) publishing a data paper.

3. Open methods

Open methods are important to guarantee the reproducibility of research, which is the practice of producing the same results of a published work by using the same data and methods. While this practice is well-established in the natural sciences, its application in archaeology it is still at its infancy. Enabling reproducibility is one of the keystones of OS and it is a crucial step to ensure the validity and transparency of a specific research. Regrettably, within the field of archaeology, typical journal articles describing the results of a study seldom provide the raw data, workflows and methods to reproduce its graphs and figures (Karoune, Plomp 2022). Consequently, when a novel methodological paper is published without any accompanying software and data, the utility of such developments becomes markedly arduous for fellow researchers to harness.

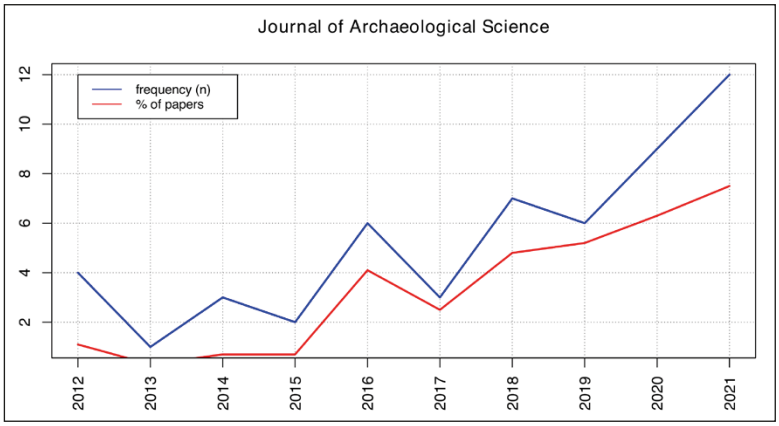

Fig. 3 – Change over time of reproducible research in «JAS» from 2012 to 2021.

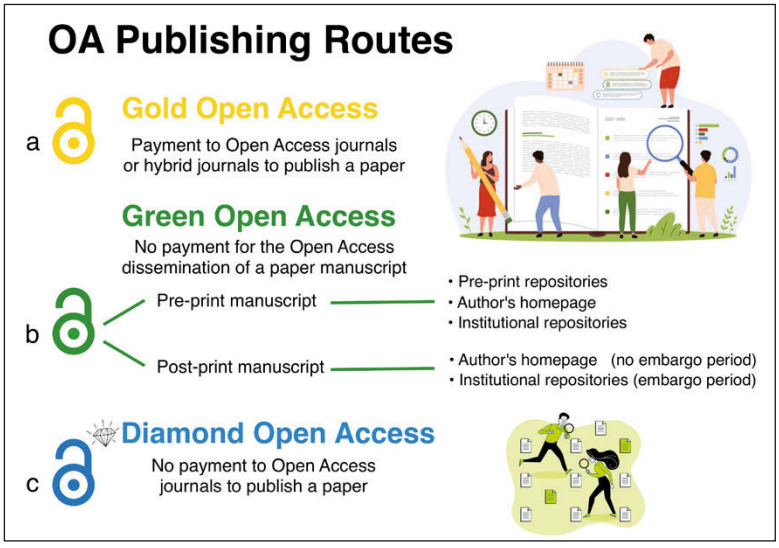

Fig. 4 – Routes to Open Access (OA) publications (created by using images from http://www.freepik.com/).

This not only contravenes the very essence of disseminating novel methodologies but also inevitably hampers any advancement in the discipline. A robust approach to enhance reproducibility involves making both raw data and analytical workflows readily accessible for the reproducibility of analyses and figures presented in published works (Marwick 2017). While one initial step

towards sharing methodologies may include the creation of tutorials or video recordings detailing the mouse-driven processes of interface-based software commonly utilized by researchers, a more effective solution would be writing a script (Marwick 2017, 431-432). A script is a plain text written in a specific programming language (e.g., Java, Python, R) which provides a comprehensive set of executable instructions for a computer. This latter approach ensures a meticulous reconstruction of the research workflow, enabling thorough scrutiny of each analysis step by any interested party. Notably, R stands out as the predominant programming language employed by archaeologists (R Core Team 2024). B. Marwick (2022) provides a list of R packages for data manipulation, and analyses and of reproducible studies making use of scripts written in R.

Transparency and reproducibility not only bolster the credibility of research but also facilitate independent evaluations of research findings by peers. In this context, particularly emblematic are two independent studies on palaeodemography focussing on a similar region and timespan published in the same issue of «Journal of World Prehistory» (Palmisano *et al.* 2021; Parkinson *et al.* 2021). As both studies openly shared their raw data and R scripts, the respective authors were invited by the journal's editor to review each other's data and workflows, subsequently reflecting on the similarities and disparities in methodology and findings between the two studies in two distinct commentaries published openly. This was an excellent example of transparency and a rare opportunity to reflect on methods and results to foster advancement in the field.

As shown in the small pilot study above, if data sharing is quite common in archaeology (at least in the methods-based journals), we cannot say the same for reproducibility practices. Merely 12.2% of articles are accompanied by workflows and raw data, thereby guaranteeing reproducibility (Tab. 1, Fig. 1), with «JCAA» exhibiting a notably higher percentage at 36.6%. This observation suggests that authors within this journal are more acquainted with computational and quantitative methodologies. We have also assessed how trends in reproducibility practices have been developed in the past decade and to do so we have scrutinised all publications in «JAS» from 2012 to 2021 (Fig. 3). The findings indicate a rising trend in efforts to make analyses reproducible, yet the peak reached in 2021 reveals that only 8% of published studies achieve reproducibility. Considering «JAS» standing as one of the foremost method-based journals, this statistic may reflect a more dismal scenario in traditional archaeology journals.

4. Open publishing

There are different options to publish OA online texts of scholarly work: Gold, Green and Diamond OA (Fig. 4) (Caravale, Piergrossi 2012). The article processing charge (APC) to publish Gold OA often requires a

substantial cost serving as a barrier for researchers affiliated with poor institutions as well as for early-career researchers lacking significant research funding (Fig. 4a). Furthermore, it is not advisable publishing via Gold Open Access in hybrid journals because they operate under a model where articles are available to users having access through institutional subscriptions. This model is thus subject to considerable controversy in terms of sustainability, as it gains revenue from both subscriptions and APCs.

Alternatively, researchers have the option to publish their work through Green OA enabling authors to share their manuscripts freely online without incurring any costs (Fig. 4b). A manuscript, distinct from the publisher's finalised version, lacks pagination and full formatting for publication and exists either as a pre-print or a post-print. A post-print is subject to certain restrictions and must meet specific requirements set by academic journals, such as embargo periods, copyright rules, and self-archiving policies. These details can be consulted through the online resource SHERPA/RoMEO (https://www.sherpa.ac.uk/romeo/).

Another option is represented by Diamond Open Access, which refers to a publishing model where the authors pay no APCs (Fig. 4c). In the field of archaeology, Diamond Open Access is guaranteed by journals such as «Archeologia e Calcolatori» (https://www.archcalc.cnr.it) (Rossi, Paraciani 2021) and «Peer Community Journal» (https://peercommunityjournal.org/). The latter publishes papers endorsed by PCI Archaeology (https://archaeo.peercommunityin.org), a community of archaeologists that recommends papers based on peer reviews, with all editorial correspondence (reviews, editors' decisions, authors' replies) being publicly accessible and transparent (Quefellec *et al.* 2023).

5. Conclusions: the roadmap to Open Science in Archaeology

This paper has provided an introduction and review of the most common OS practices in archaeology. The highlighted practices show diverse approaches and have been applied for various purposes. It is worth noting that there is no single optimal solution, and researchers should carefully consider their options based on the nature of the data and analyses utilised in their research endeavours. In this section, we will outline a decalogue, delineating a series of straightforward recommendations and guidelines that may prove helpful for scholars seeking to disseminate their research findings openly:

1. Publish scholarly work electronically (e.g., books, archaeological excavation and surveys reports, research articles).
2. Ensure that your data are deposited in an online, freely accessible repository that assigns a persistent DOI.
3. Store data in its most unprocessed state possible to facilitate reusability for subsequent analyses and future research endeavours.

4. Utilize open formats for data storage to ensure accessibility and independence from proprietary software. Open formats are tailored to foster interoperability, accessibility, and long-term data reuse. Examples include CSV for tabular data, PNG for raster images, SVG for vector graphics, and GeoJSON for spatial data.
5. Accompany the corresponding dataset with relevant metadata containing information on its structure, organization, and description.
6. Disseminate data under an open license to specify the conditions under which others can use the dataset. Recommended Creative Commons (CC) licenses include CC0 and CC-BY.
7. Enhance research reproducibility by providing raw data and a detailed workflow for reproducing all analyses, graphs, and figures. Ideally, scripts written in a specific programming language and executable via open-source software (e.g., Java, Python, R) ensure a comprehensive record of the research workflow.
8. Publish a data paper to ensure that the academic community is informed of the availability and potential reuse of a specific dataset.
9. Whenever feasible, consider publishing your article through Diamond Open Access. Alternatively, opt for Green Open Access to ensure broader dissemination of your work.
10. Consider Gold Open Access publication only in fully open access journals and avoid hybrid journals.

Alessio Palmisano*, Andrea Titolo
Dipartimento di Studi Storici
Università degli Studi di Torino
alessio.palmisano@unito.it, andrea.titolo@unito.it

*Corresponding Author

REFERENCES

Bevan A. 2015, *The data deluge*, «Antiquity», 89, 348, 1473-1484 (https://doi.org/10.15184/aqy.2015.102).

Caravale A., Piergrossi A. 2012, *Archeologia in rete. Le riviste open access: risorse e prospettive*, «Archeologia e Calcolatori», 23, 187-207 (https://www.archcalc.cnr.it/journal/id.php?id=616).

Fecher B., Friesike S. 2014, *Open science: One term, five schools of thought*, in S. Bartling, S. Friesike (eds.), *Opening Science*, Cham, Springer International Publishing, 17-47 (https://doi.org/10.1007/978-3-319-00026-8_2).

Hole B. 2012, *A call for open scholarship in archaeology*, in. C. Bonacchi (ed.), *Archaeologists and the Digital. Towards Strategies of Engagement*, London, Archetype, 114-126 (https://discovery.ucl.ac.uk/id/eprint/1326267/).

Huet T., Cubas M., Gibaja J.F., Oms F.X., Mazzucco N. 2022, *NeoNet Dataset. Radiocarbon dates for the Late Mesolithic/Early Neolithic transition in the North Central-Western Mediterranean basin*, «Journal of Open Archaeology Data», 10, 3, 1-8 (https://doi.org/10.5334/joad.87).

Karoune E., Plomp E. 2022, *Removing Barriers to Reproducible Research in Archaeology*, Zenodo, ver. 5 peer-reviewed and recommended by Peer Community in Archaeology (https://doi.org/10.5281/zenodo.7320029).

Lucarini G., Wilkinson T., Crema E.R., Palombini A., Bevan A., Broodbank C. 2020, *The MedAfriCarbon radiocarbon database and web application. Archaeological dynamics in Mediterranean Africa, ca. 9600-700 BC*, «Journal of Open Archaeology Data», 8, 1, 1-6 (https://doi.org/10.5334/joad.60).

Marvick B. 2017, *Computational reproducibility in archaeological research: Basic principles and a case study of their implementation*, «Journal of Archaeological Method and Theory», 24, 424-450.

Marvick B. 2022, *CRAN task view: Archaeological science*, Zenodo (https://doi.org/10.5281/zenodo.7306328).

Marvick B., Pilaar Birch S.E. 2018, A *standard for the scholarly citation of archaeological data as an incentive to data sharing*, «Advances in Archaeological Practice», 6, 2, 125-143 (https://doi.org/10.1017/aap.2018.3).

Palmisano A., Bevan A., Kabelindde A., Roberts N., Shennan S. 2021, *Long-term demographic trends in prehistoric Italy: Climate impacts and regionalised socio-ecological trajectories*, «Journal of World Prehistory», 34, 381-432 (https://doi.org/10.1007/s10963-021-09159-3).

Parkinson E.W., Mclaughlin T.R., Esposito C., Stoddart S., Malone C. 2021, *Radiocarbon dated trends and central Mediterranean prehistory*, «Journal of World Prehistory», 34, 317-379 (https://doi.org/10.1007/s10963-021-09158-4).

R Core Team 2024, *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Wien (https://www.R-project.org/).

Rossi I., Paraciani N. 2021, *IT applications to archaeology and the OA diamond journals' challenge. Enhancing access and reuse of textual and visual resources,* «Archeologia e Calcolatori», 32.1, 325-347 (https://doi.org/10.19282/ac.32.1.2021.18).

Quefellec A., Maureille B., Arzarello M., Blasco R., Crandell O., Doyon L., Halcrow S., Karoune E., Ruiz-Redondo A., Van Peer P. 2023, *Peer Community In Archaeology: A community-driven free and transparent system for preprints peer-reviewing*, in J. Bogdani, S. Costa (eds.), *ArcheoFOSS 2022. Proceedings of the 16th International Conference on Open Software, Hardware, Processes, Data and Formats in Archaeological Research (Rome 2022)*, «Archeologia e Calcolatori», 34.1, 125-134 (https://doi.org/10.19282/ac.34.1.2023.14).

UNESCO 2021, *UNESCO Recommendation on Open Science*, Paris, United Nations Educational, Scientific and Cultural Organization (https://doi.org/10.54677/MNMH8546).

Vicente-Saez R., Martinez-Fuentes C. 2018, *Open Science now: A systematic literature review for an integrated definition*, «Journal of Business Research», 88, 428-436 (https://www.sciencedirect.com/science/article/abs/pii/S0148296317305441?via%3Dihub).

ABSTRACT

In the past decade, archaeology has witnessed a surge in digital data and methodologies made accessible through online repositories, databases, collaborative platforms like GitHub, and library collections. While this 'digital revolution' has lead to notable advancements in data accessibility and dissemination, it has also presented several challenges. These include developing effective data management strategies, defining scientific publication modalities, addressing ethical concerns regarding the protection of cultural heritage, and bridging the gap between rich and poor research institutions. Taking these challenges into account, this paper aims to outline best practices for data sharing, dissemination, and reproducibility in archaeology, underscoring their benefits to researchers and the wider public. Through a range of approaches and practical examples drawn from everyday research scenarios, we will demonstrate how these practices promote transparency and foster open science.