



Open Library of Humanities

Challenges in Open Data: A TRAJ Perspective

Katherine A. Crawford, Science & Technology in Archaeology & Culture Research Center, The Cyprus Institute, CY, k.crawford@cyi.ac.cy

Francesca Mazzilli, University of Bergen, NO, mazzillifra@gmail.com

The editorial of the fifth issue of the *Theoretical Roman Archaeology Journal* focuses on the topic of open data. The editors consider both the challenges of using and disseminating open data within the field of Roman archaeology. We also reflect upon how *TRAJ* can improve on the inclusion of open-access data in the future.



In this issue we have a broad range of articles, all of which are connected by the topic of data. Lawrence tackles big data in Roman archaeology from a historiographical perspective, drawing out four examples of how big data has been used in various sub-fields. He focuses on how large data has been used in the past as well as its potential of driving future narratives and frameworks. Stansbie's article looks specifically at big data and the use of legacy datasets. He proposes the usefulness of applying a boundary object approach as one method of drawing out new patterns and understanding from existing large datasets. Shaw examines the possible function of foot-handled jugs based on their assemblage contexts, providing a detailed symbolic analysis of these artefacts. Using a range of different archaeological data, Campanaro looks at whether or not *atirum* would have been roofed, applying an IBE (inference to the best explanation) approach. Baryshnikov introduces us to the topic of theoretical roman archaeology from a Russian academic tradition. While not dealing with specific datasets, this article highlights how access to academic data in the form of scholarship can shape approaches to Roman archaeology in a national context.

A recurrent issue is whether or not the data in which an article is grounded is provided. One issue, especially in the use of large data sources, is accessibility to other scholars. Stansbie, while granted permission to use data from the English Landscapes and Identities (EngLaId) Project, was not able to publicly share the data, thereby making any analysis non-reproducible. This raises the question of the usefulness of big data projects if only a select few researchers have access to the data, and if the results and methods cannot be checked or even challenged.

Another challenge is fully accessible data access. Often datasets are drawn from a researcher's previous projects. Shaw's article is one example where the primary and secondary data sources are cited throughout the text, but the actual corpus of material is not provided. Until relatively recently it has been a hallmark of Roman archaeology to discuss results and provide analysis without overtly providing access to all (or often any) data, instead, data is obscured by references.

Within Roman archaeology, and archaeology more widely, there is a growing positive trend towards open data. Numerous data repositories are now widely accessible to scholars such as tDAR (The Digital Archaeology Record) and the ADS (Archaeology Data Service). For projects that may be limited by funds, there are opportunities like hosting smaller datasets on Zenodo which still provide datasets with a dedicated DOI, ensuring that data follows FAIR (Findable, Accessible, Interoperable, and Reusable) principles. New journals, like the Journal of Open Archaeology Data, provide another venue for publishing data while requiring that data are also hosted in a fully open-access data repository. The drive towards open data can be most prominently seen at the funding

level, with many local and regional funding schemes now requiring the inclusion of a data management plan to ensure that data follows FAIR principles (Geser et al. 2022).

Increased access to big data was predicted by Kintigh et al. (2014) to be one of the major future challenges of archaeology. Recent years have seen a substantial increase in the number of large data projects, intent not only on collecting and creating databases, but ensuring that they are fully accessible and open access. Data projects range from collecting settlement geo-spatial data at both local and regional levels (e.g. Hanson 2016) to compiling datasets of specific forms of material culture (e.g. Colley and Evans 2018). Numerous questions remain about both the usefulness and the ethics of generating such 'big' data projects (Cooper and Green 2016; Fisher et al. 2021), as well as considering the methodological approaches that should be taken in generating these datasets (Huggett 2020). In order for such datasets to be applicable or useful to future scholars we additionally need to consider how comprehensive these data are, can or will these data be updated in the future, and for how long will the data repositories be secure?

Promoting greater data accessibility within TRAJ

From its foundation, TRAJ has required the inclusion of open data or links to primary data used within articles. Part of the TRAJ data policy states, 'all underlying data which support tables and figures must be cited'.¹ This primarily can be seen through the inclusion of supplementary files, or more commonly, references to where the data is accessible and which may or may not be open access. A total of 19 articles have provided downloadable supplementary data without a major variance over the years, roughly half of the articles published per year, with exception of the current volume (**Figure 1**). This overarching figure can be interpreted to be quite positive, however, it makes the TRAJ editors question and reflect on how TRAJ can take further steps to further promote the accessibility of data within the articles we publish.

So far there has been no homogeneity in the format and the choice of supplementary data, which has been down to the author's discretion. In future, the TRAJ data policy needs to implement a clear outline of how the data can be presented in order to optimise its reuse potential. Files that are provided as word docs or pdfs, for instance, are not conducive to data reuse and require substantial data formatting in order to be used by subsequent scholars. One step is to make it mandatory that data be published with a DOI, either in a data repository or as a csv file in TRAJ supplementary material. We also need to implement a policy of requiring underlying map data—which show the distribution of finds, sites, etc.—also to be made available while adhering to any ethical constraints in making geospatial data accessible. Making data as accessible as possible is one of the challenges that TRAJ aims to tackle in forthcoming years to ensure that researchers can reuse data to push forward cutting-edge research that TRAJ already provides.

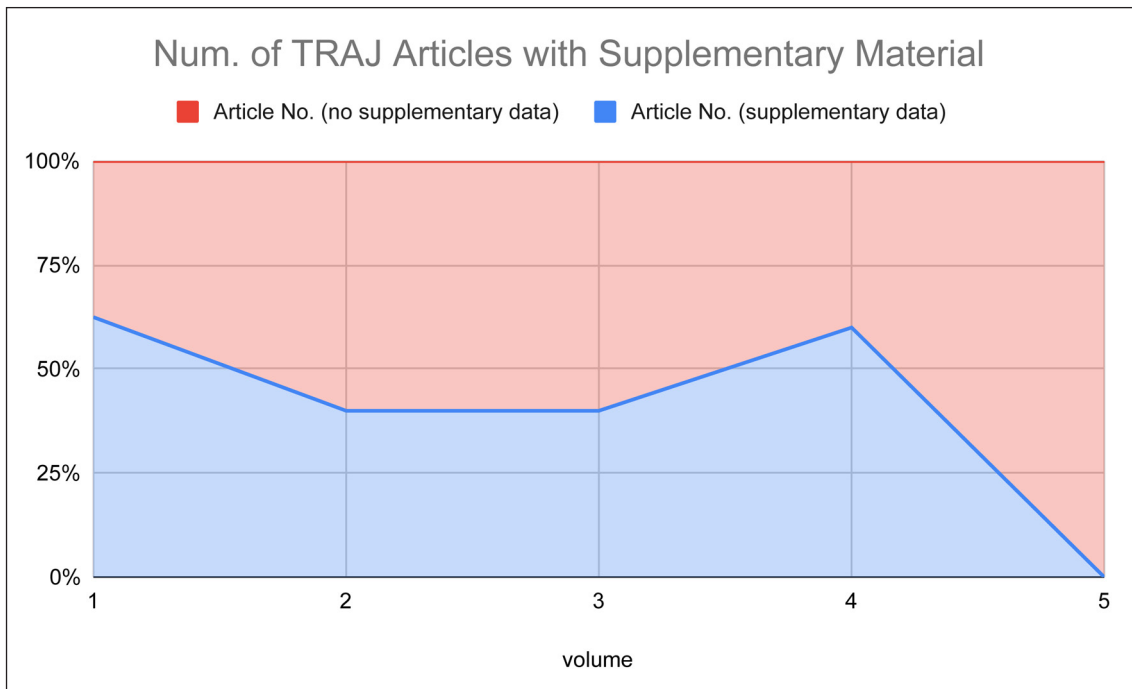


Figure 1: Graph showing the number of TRAJ articles that include supplementary material.

Note

¹ <https://traj.openlibhums.org/submissions/>.

Competing Interests

The authors have no competing interests to declare.

References

Colley, S. and Evans, J. 2018. Big data analyses of Roman tableware: information standards, digital technologies and research collaboration. *Internet Archaeology* 50. DOI: <https://doi.org/10.11141/ia.50.19>

Cooper, A. and Green, C. 2016. Embracing the complexities of 'big data' in archaeology: the case of the English Landscape and Identities Project. *Journal of Archaeological Method and Theory* 23: 271–304. DOI: <https://doi.org/10.1007/s10816-015-9240-4>

Fisher, M., Fradley, M., Flohr, P., Rouhani, B., and Simi, F. 2021. Ethical considerations for remote sensing and open data in relation to the endangered archaeology in the Middle East and North Africa project. *Archaeological Prospection* 28(3): 279–292. DOI: <https://doi.org/10.1002/arp.1816>

Geser, G., Richards, J.D., Massara, F. and Wright, H. 2022. Data management policies and practices of digital archaeological repositories. *Internet Archaeology* 59. DOI: <https://doi.org/10.11141/ia.59.2>

Hanson, J.W. 2016. *An Urban Geography of the Roman World, 100 BC to AD 300*. Oxford: Archaeopress. DOI: <https://doi.org/10.2307/j.ctv17db2z4>

Huggett, J. 2020. Is big digital data different? Towards a new archaeological paradigm. *Journal of Field Archaeology* 45: S8–S17. DOI: <https://doi.org/10.1080/00934690.2020.1713281>

Kintigh, K.W. et al. 2014. Grand challenges for archaeology. *American Antiquity* 79(1): 5–24. DOI: <https://doi.org/10.7183/0002-7316.79.1.5>

