PROJECT TITLE

Building confidence in an era of doubt: Bringing the 'credibility revolution' to archaeological field research

PROJECT QUALITY AND INNOVATION

Over the past two decades, researchers have recognised a 'reproducibility crisis' in the natural and social sciences, with indications that perhaps one-third to two-thirds of published research is false (Baker 2016; Ioannidis 2005). This crisis undermines public trust in research (Pew Research Center 2019; Vazire 2017) and sabotages genuine discovery (Smaldino and McElreath 2016). A 'credibility revolution' has arisen in response (Vazire 2018), which has produced reproducibility studies (Errington et al. 2021; Open Science Collaboration 2015) and introduced approaches like preregistration and Findable, Accessible, Interoperable, and Reusable (FAIR) data to improve reliability (Nosek et al. 2018; Wilkinson et al. 2016). It has also sought to change research culture and incentives (Center for Open Science 2014; Cuevas Shaw et al. 2022). More recently, the movement has inspired national and international standards (OECD 2021; UNESCO 2021; NHMRC et al. 2019). Adoption of open science principles, approaches, and practices, however, varies significantly by domain.

While the credibility revolution has focused on experimental sciences, it extends to field research disciplines. Amongst these, ecology and geoscience are leaders. Ecologists have assessed the reproducibility crisis in their discipline and proposed responses (Fidler et al. 2017), explored variability in data analysis (Gould et al. 2023), examined computational reproducibility (Powers and Hampton 2019), explored technical and sociotechnical aspects of data sharing (Wallis et al. 2008), and proposed minimum standards for data and code (Jenkins et al. 2023). Geoscientists have studied computational reproducibility (Konkol et al. 2019), assessed the availability of data (Boone et al. 2022; Chamberlain et al. 2021), articulated a model for reliable research publications (Gil et al. 2016), started implementing this vision through journal mandates (Springer-Nature 2019), and built infrastructure like persistent identifiers for physical samples (**Klump** et al. 2021).

Archaeology, however, has been slower to engage the credibility revolution. What few systematic evaluations of openness or reliability exist tend to focus on data: D'Gluyas and Gibbs (2022) found only 15% of 20 New South Wales (Australia) field projects published reusable datasets; Lodwick (2019) reviewed 239 archaeobotany articles and found that while half share data, few datasets are readily reusable; **Karoune** (2022) found that 4% of 341 phytolith articles contained reusable data. Proposals for improvements, likewise, have tended to emphasise one aspect of openness, like preregistration (**Ross** and **Ballsun-Stanton** 2022), FAIR data (**Kansa** et al. 2020; **Ross** et al. 2022), computational reproducibility (**Marwick** 2017), or sociotechnical aspects of adopting reproducibility practices (**Karoune** and **Plomp** 2022). Overall, most archaeological research does not use open-research approaches, leaving a gap that undermines research credibility. In this regard, archaeological practice fails to align with national and international standards, and archaeologists remain unprepared for funder and publisher mandates. Without well-documented and machine-readable data, furthermore, archaeologists cannot exploit existing and emerging tools like traditional machine learning (ML) or large language models (LLMs), which are transforming the discipline.

Now is the time to address this unsustainable gap. An open-science ecosystem exists outside archaeology, giving archaeologists the opportunity to learn from adjacent disciplines like ecology and geoscience (**CI Hose; PIs Klump and Wyborn**). In this context, this project proposes the first large-scale, comprehensive evaluation of research credibility in archaeology and subsequent articulation of approaches and practices to improve the openness and reliability of archaeological research. To achieve these aims, it includes three work packages (WP):

WP1 will develop a framework for evaluating archaeological research, including excavation, surface survey, and associated specialised studies. It is not yet clear how research should be assessed in the discipline, since archaeology is 'abductive' (cycling between deductive and inductive approaches), uses both nomothetic and idiographic approaches, often cannot be repeated (e.g., excavation destroying what it studies), employs diverse approaches and methods, has few disciplinary standards, and lacks robust information infrastructure across much of the research lifecycle (see Borgman 2015 for these 'small data' challenges). This WP interrogates concepts like reproducibility, replication, and transparency to develop appropriate approaches to credibility in the discipline. It will begin by systematically reviewing methodological literature and model outputs from recent field projects to assess the current state of open-science approaches and practices, benchmarking them against international guidelines and standards as well as good practice in other disciplines (especially ecology and geoscience). Social surveys designed with input from the Center for Open Science (COS) will contribute sociotechnical context. This review will then be developed into a framework of feasible good practice for evaluating the reliability of archaeology field projects in WP2.

WP2 will apply the framework developed in WP1, evaluating the practice of 20-40 archaeological projects from fieldwork through specialist analyses to publication. These projects will be selected from 'flagship' excavation and survey projects that (a) have won major research grants in Australia, the USA, the UK, and Europe and (b) have few inherent barriers to openness such as sensitive data. This unencumbered and well-resourced research will provide a baseline for potential future evaluation of commercial research and projects with more sensitive data. The COS will facilitate design and execution of the study. WP2 will produce a comprehensive gap analysis comparing how flagship projects align with good practice as defined in WP1. The analysis is not intended to criticise particular projects, but instead to identify omissions, inefficiencies, and opportunities that can be addressed using proven approaches already in use in archaeology and adjacent disciplines or with emerging approaches viable in the near-term.

WP3 combines the feasible good practice developed in WP1 with the evaluation from WP2 to produce a suite of practical approaches, protocols, and tools that can be deployed to improve the credibility of archaeological research. This toolkit will include guides to research design; templates for data management plans, (pre)registrations, FAIR implementation profiles, and similar apparatus; recommendations for producing machine-actionable FAIR data and reproducible analytical workflows; a roadmap to current and emerging technical tools and infrastructure; benchmarks for adequate resourcing of open, accessible research; and approaches to deploying ML and LLMs for data validation, analysis, and pattern discovery - all tied together by an end-to-end overview of practical open research in archaeology. These templates will also be tested across disciplines, as they should be adoptable by ecologists, geochemists, and other fieldwork-based researchers with modest adaptation. Throughout, emphasis will be placed on approaches that do the most to close the gap with good practice for the least investment of time and resources. This work package will produce the toolkit itself, as well as key recommendations for researchers, funding agencies, and publishers regarding reasonable standards and resourcing needs for open and credible research in archaeology.

To gain wider input and amplify impact, WP1 and WP2 will be coordinated with the proposal and operation of a Research Data Alliance (RDA) Interest Group; WP3 will be the focus of a RDA Working Group. Afterwards, an ongoing RDA Community of Practice will be established. (Several Investigators have experience with RDA.)

Cultural and economic benefits will include actionable guidelines and useful materials for improving the quality and credibility of archaeological and other fieldwork-based research in Australia and conducted by Australians overseas, benchmarks to help funding agencies such as the ARC evaluate research and infrastructure grant proposals, greater sustainability and cost-savings in field research via reuse of data and methods (substituted, to an extent, for new fieldwork). It will also help future-proof data in archaeology and other field disciplines by enabling more consistent production of comprehensive, well-documented, and machine-actionable datasets that can yield predictable results when used with emerging technologies like ML and LLMs.

REFERENCES

Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016;533(7604):452–4. **Boone SC,** Dalton H, Prent A, et al. AusGeochem. Geostand. Geoanalytical Res. 2022;46(2):245–59. **Borgman CL**. Big data, little data, no data. MIT press; 2015. **Center for Open Science**. TOP Guidelines. 2014. **Chamberlain KJ**, Lehnert KA, McIntosh IM, et. al. Time to change the data culture in geochemistry. Nat. Rev. Earth Environ. 2021 29;2(11):737–9. **Cuevas Shaw L**, Errington TM, Mellor DT. Toward Open Science: Contributing to Research Culture Change. Sci. Ed. 2022 15;14–7. **D'Gluyas C**, Gibbs M. Future use or no future at all? An examination of post-excavation historical archaeological repositories in NSW. Aust. Archaeol. 2022 4;88(2):129–43. **Errington TM**, Mathur M, Soderberg CK, et al. Investigating the replicability of preclinical cancer biology. eLife. 2021 10;10:e71601. **Fidler F**, Chee YE, Wintle BC, et al. Metaresearch for Evaluating Reproducibility in Ecology and Evolution. BioScience. 2017 25;biw159. **Gil Y**, David CH, Demir I, et al. Toward the Geoscience Paper of the Future. Earth Space Sci. 2016;3(10):388–415. **Gould E**, Fraser H, Parker T, et al. Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. Eco/EvoRxiv 2023 [preprint; https://doi.org/10.32942/X2GG62/ (v3) accessed 26/2/24]. **Ioannidis JPA**. Why Most Published Research Findings Are False. PLoS Med. 2005 30;2(8):e124. **Jenkins GB**, Beckerman AP, Bellard C, et al. Reproducibility in ecology and evolution: Minimum standards for data and code. Ecol. Evol. 2023;13(5):e9961. **Kansa SW,** Atici L, Kansa EC, Meadow RH. Archaeological Analysis in the Information Age. Adv. Archaeol. Pract. 2020;8(1):40–52. **Karoune E**. Assessing Open Science Practices in Phytolith Research. Open Quat. 2022 10;8:3. **Karoune E**, Plomp E. Removing Barriers to Reproducible Research in Archaeology. Peer Community Archaeol.. Zenodo; 2022. **Klump J**, Lehnert K, Ulbricht D, et al. Towards Globally Unique Identification of Physical Samples. Data Sci. J. 2021 28;20:33. **Konkol M**, Kray C, Pfeiffer M. Computational reproducibility in geoscientific papers. Int. J. Geogr. Inf. Sci. 2019 1;33(2):408–29. **Lodwick L**. Sowing the Seeds of Future Research: Data Sharing, Citation and Reuse in Archaeobotany. Open Quat. 2019 15;5:7. **Marwick B**. Computational Reproducibility in Archaeological Research. J. Archaeol. Method Theory. 2017;24(2):424–50. **NHMRC**, ARC, Universities Australia. Management of Data and Information in Research. 2019 Report No.: R41B. **Nosek BA**, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc. Natl. Acad. Sci. 2018 13;115(11):2600–6. **OECD**. Recommendation of the Council concerning Access to Research Data from Public Funding. 2021 Report No.: OECD/LEGAL/0347. **Open Science Collaboration**. Estimating the reproducibility of psychological science. Science. 2015 28;349(6251):aac4716. **Pew Research Center**. Trust and Mistrust in Americans' Views of Scientific Experts. 2019. **Powers SM**, Hampton SE. Open science, reproducibility, and transparency in ecology. Ecol. Appl. 2019;29(1):e01822. **Ross SA**, Ballsun-Stanton B, Cassidy S, et al. FAIRer Data through Digital Recording. J. Comput. Appl. Archaeol. 2022 10;5(1):271. **Ross SA**, Ballsun-Stanton B. Introducing Preregistration of Research Design to Archaeology. In: Watrall E, Goldstein L, editors. Gainesville, FL: University Press of Florida; 2022. **Smaldino PE**, McElreath R. The natural selection of bad science. R. Soc. Open Sci. 2016;3(9):160384. **Springer-Nature**. Announcement: FAIR data in Earth science. Nature. 2019;565(7738):134–134. **UNESCO**. UNESCO Recommendation on Open Science; 2021. **Vazire S**. Quality Uncertainty Erodes Trust in Science. Collabra Psychol. 2017 1;3(1):1. **Vazire S**. Implications of the Credibility Revolution for Productivity, Creativity, and Progress. Perspect. Psychol. Sci. 2018;13(4):411–7. **Wallis JC**, Borgman CL, Mayernik MS, Pepe A. Moving Archival Practices Upstream. Int. J. Digit. Curation. 2008 2;3(1):114–26. **Wilkinson MD**, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data. 2016 15;3(1):160018.