

## DATA PAPER

# Data from "Assessing Open Science Practices in Phytolith Research"

Emma Karoune

Independent researcher, GB  
ekaroune@googlemail.com

This is a dataset gathered to assess the state of open science practices in phytolith research. All articles presenting primary phytolith data were extracted from 16 prominent archaeological and palaeoecological journals between 2009 and 2018. In total, the dataset contains information on 341 articles. This included archaeological (n = 214), palaeoenvironmental (n = 53) and methodological (n = 74) studies. Information was recorded regarding the data location and what type of data was included in the text and as supplementary files. There was also data recorded in relation to open access, picture inclusion, use of the International code for Phytolith Nomenclature (ICPN) and the inclusion of a full method.

**Keywords:** Phytolith; Archaeobotany; Open data; Open Science; Open access

## (1) Overview

### Context

Open science practices are an increasingly important element of scientific research that are likely to become a requirement of every journal, funding body and employer. It is based on four pillars: open data, open methods (protocols and code), open papers and open reviews [9]. This movement seeks to open up science to the wider community, both academic and the public, by making all scientific outputs available to enable greater transparency. The adoption of these approaches will facilitate more efficient transfer of skills and knowledge throughout the research community, and so improve the diversity of the discipline and equity of access. A discussion of open science in Archaeology by Marwick *et al.* [7] identifies specific practices that will benefit researchers and the wider scientific community: i) the use of pre-prints; ii) depositing data in repositories; and iii) transparent and reproducible workflows. However, in many scientific fields the current extent of these practices is unknown and therefore reviews need to be conducted to assess the current situation.

As much of the output of scientific research is publications in journal articles, there have been several journal article reviews of data sharing, and other aspects of open science, conducted in archaeology but none concerning phytolith research [4, 5, 8]. Recent reviews of archaeological science [8] and macro-botanical remains [5] found low levels of data sharing, 53% and 56% respectively. It was therefore important to assess where phytolith research was in relation to other sub-disciplines of environmental archaeology and archaeological science in general.

The use of phytolith analysis for archaeological and palaeoecological studies has been increasing in recent year in terms of its methods and their applications [1, 3] Phytoliths are silica bodies that are formed within plant cells during the lifetime of the plant and can be used to identify plant taxa to different taxonomic levels [11]. Phytolith analysis is not only used to examine the floral component of past sediments from archaeological sites and their wider environment but is now increasingly being used for radiocarbon dating and isotope analysis [2, 12, 13]. Extraction of phytoliths from artefacts and ecofacts, such as grinding stones, tooth calculus and pottery, are innovations that are addressing new archaeological questions [1, 10].

It is therefore important that this increase in research and particularly the upsurge in publications with associated data is made as useful to other colleagues as possible. Embedding open science practices in a research project will allow for the greatest transparency and therefore the ability for other researchers to validate findings and build on research. Knowing the current extent of open science practices in phytolith research highlights areas in need of improvement and enables guidelines to be established to bring researchers in closer alignment to good open science practices.

### Spatial coverage

Articles in this study cover a global range.

### Temporal coverage

Articles in this study are not restricted to one particular time period. They range from studies of the palaeoenvironments of early hominids to historical archaeological sites. It also includes articles focused on methodological studies of modern environments.

**Table 1:** List of journals and categories used by [5] and also in this dataset.

Archaeobotanical journals	Archaeological science journals	General Archaeology	Cross-disciplinary Journals
Vegetation History and Archaeobotany	Archaeological and anthropological science	Antiquity	PLOS One
	Environmental Archaeology	Journal of Field Archaeology	Proceedings of the National Academy of Sciences
	The Holocene	Oxford Journal of Archaeology	
	Journal of Archaeological Science	Journal of Anthropological Archaeology	
	Journal of Archaeological Science reports	Journal of World Prehistory	
	Journal of Ethnobiology		
	Quaternary International		
	Journal of Wetland Archaeology		

## (2) Methods

This dataset was designed to complement the dataset produced by Lodwick [5]. Therefore, the same journals (see **Table 1** for the list of journals sampled) and same period (2009–2018) were sampled. This enabled the two datasets to be compared as they both concern sub-disciplines of archaeobotany.

### Steps

To find the articles needed for this research on open science practices in phytolith research, the following steps were taken:

1. The author accessed the journal website and searched the term 'Phytolith'.
2. This was then refined to the 10-year period required (2009–2018).
3. Once the list of articles was found, each article in the list was examined for primary phytolith data. Only articles that provided primary data were selected for the dataset. The articles could be archaeological, palaeoenvironmental or methodological. This was determined from the main focus of the article and the research questions being addressed. There is often overlap between palaeoenvironmental and archaeological studies and therefore some articles could have fallen into either category. In these cases, the author put the articles into the archaeological category as they were focused on samples from an archaeological site.

A full list of the categories (column headings) recorded in the dataset can be found in **Table 2**. This also sets out the key to the codes used. The categories recorded from each article were selected to gain the most information concerning open science practices therefore they included open access, data sharing and other information provided with the articles such as methods, pictures and use of the International Code for Phytolith Nomenclature (ICPN) [6]. These later categories could also be termed as the metadata. Both the raw data and metadata should be

made available with all articles to allow thorough peer review and to give other researchers the opportunity to build on previous research.

The author decided to simplify the collection of data from the selected articles by taking a presence/absence approach to most of the categories in the dataset. Therefore, several categories need further clarification as to how they were recorded as Yes or No answers:

- *Raw count data in re-useable format* – there was a wide variety of data presentation methods and types of data found in the articles and recording all of these would not add anything to the argument of data reusability (it was recorded in the other details section of the dataset). Therefore, the author determined that to enable other researchers to reuse phytolith data, the raw counts and the weights taken during processing need to be provided. This is the actual raw data created in phytolith analysis and making this available will allow other researchers to conduct any form of analyses on the data. This data also needs to be in a format easily accessible, therefore, to get a Yes for this category the raw count data needed to also be in an excel spreadsheet, csv file or in an open repository as an excel or csv file.
- *Picture of phytolith morphotypes* – this could be photographs or diagrams of the morphotypes identified in the study. These are important for validation of identification.
- *Other access found* – many of the journal articles were not Gold open access. However, the author felt that it was important to record how available these articles were online in other locations. Many researchers post their articles on academic social media sites such as Research gate and Academia.edu or make them available through open repositories such as their own university repositories. This meant that a true sense of the openness of the publications in this study was determined.
- *Use of ICPN 1.0* – an effort has been made by the phytolith community to standardise the nomenclature used in research. The author wanted to determine

**Table 2:** A key to categories (column headings) and codes used in the dataset.

Category name	Codes and details recorded in category
Journal name	Full name of the journal
Journal type	Archaeobotany Archaeological science General Archaeology Cross-disciplinary
Year	Range between 2009 and 2018
Title	Full title of the article
DOI	DOI code
Region	Geographic country of study –geonames used to standardise ( <a href="https://www.geonames.org/">https://www.geonames.org/</a> ).
Period	Archaeological period – dates or name of period used in the study – given in the introduction of each article.
Type of study	Archaeological Palaeoenvironmental Methodological
Sub-type	For methodological papers only: MRS = Modern reference soil/habitat MRP = Modern reference plant M = Morphometrics T = Taphonomy R = Radiocarbon dating MD = Method development
Data location	N = no data given Table Graph .docx = word document pdf .xlsx = excel spreadsheet .csv Repository
Raw count data in re-useable format (supp material as excel or csv, or in repository)	Y = Yes N = No
Pictures/photos provided of phytolith morphotypes	Y = Yes N = No
Open access article	Y = Yes N = No
Open journal	Y = Yes N = No
Other access found	N/A – open access articles already. Research gate Academia Repository – other than social media ones above 6 months – article made available after 6 months by journal N – had to request N – could not request
Used ICPN 1.0	Y = Yes N = No
Full methods given	Y/N Y = Written in full in the text/supplementary file or referenced the use of one method article.
Other details – details of data given.	More detail of the types of data were recorded – what form the data was provided in – raw counts/absolute counts/percentage/types of graphs, etc.

to what extent this was being used, as the standardisation and use of specific codes for morphotypes is important for the reuse of data.

- *Full method* – it was determined that a full method was supplied if a full description of the phytolith extraction process (from sediment or plant material) was given in the text of the article or as a supplementary file, or there was a clear reference to one methodological paper.

### Quality Control

Once the data collection stage was completed, the dataset was checked for spelling mistakes and consistency of terms. All entries were checked for inconsistencies using **Table 2** to confirm that the codes used and the criteria for presence/absence categories were applied correctly. Location (Region) data was standardised to countries using geonames (<https://www.geonames.org/>).

### Constraints

The period category was entered for archaeological articles only, however, standardising these entries proved difficult due to the global nature of the dataset. Named periods often have different meanings in different geographic regions. Therefore, this data was not standardised and was collected as either a period name, date or date range given in the introduction of the article.

The decision to use a presence/absence category to record the sharing of raw data was determined partly by problems with labelling tables and graphs in the published articles. Some of the data was not labelled adequately to allow the type of data to be determined.

Another factor that did not constrain but hindered the collection of data was the poor labelling of supplementary data files. Often the files were labelled as supplementary file 1, with no other explanation of the contents of the file in the title. It was also found that these files were not adequately referred to in the text. Therefore, to determine what the file contained, it had to be downloaded. If a researcher was collating a large amount of data for meta-analysis, this lack of labelling would add considerably to their workload. All supplementary files in this dataset were downloaded and examined for the completion of the dataset.

## (3) Dataset description

### Object name

Raw data table for Karoune 2020 Assessing Open Science Practices in Phytolith Research – information extracted from 341 articles of primary phytolith data from 16 archaeology and palaeoecology journals between 2009–2018.

Key to codes for Karoune 2020 – description of the data and codes used in each column heading of the dataset.

### Data type

Primary and secondary data.

### Format Names and Versions

Raw data table for Karoune 2020 Assessing Open Science Practices in Phytolith Research – CSV file.

Key to codes for Karoune 2020 – CSV file.

### Creation Dates

Dataset created between October 2019 to June 2020.

### Dataset Creators

The primary researcher responsible for the data collection was Emma Karoune.

### Language

English

### License

CC0 1.0 Universal

### Repository Location

Open Science Framework: <https://osf.io/8p3bn/>

### Publication Date

23/07/2020

## (4) Reuse Potential

There are several potential ways that this data could be reused. Firstly, this dataset adds to the growing review of open science practice in Archaeology and more specifically Environmental Archaeology. It could therefore be collated with other evidence or built upon further to draw together an overview throughout the discipline.

It could serve as an aid to teaching open science practices. The dataset could be used for a simple data analysis task for students in teaching modules, along with other such evidence from Archaeology or Science in general.

The dataset could also be used in teaching environmental archaeology, particularly phytoliths or archaeobotany modules. It could aid the teaching of data analysis and discussions concerning academic publishing and the application of open science practices.

Within phytolith research, this dataset can be used by the phytolith community to examine the way forward in terms of drawing up guidelines for data sharing and open science practices in this field.

In terms of academic publishing, this data could be used by journal editors to draw up guidelines for journal data availability policies.

### Acknowledgements

This project was supported by the phytolith community through those that have made their articles open access or deposited them in an online repository. I offer my thanks to those colleagues that answered my emails for articles and supplementary file requests. I am also grateful to colleagues that sent me articles that I was unable to access through other means – Dr Ruth Pelling and Professor Dorian Fuller. Special thanks go to Rand Evett, who works to maintain of the International Phytolith Society bibliography and offered much help to enable me to access supplementary files to finish the dataset. I also offer thanks to the three reviewers that wrote very fair and constructive reviews that helped to improve this article and the dataset.

**Competing Interests**

The author has no competing interests to declare.

**References**

1. **Hart, TC** 2016 Issues and direction in phytolith analysis. *Journal of Archaeological Science*, 68: 24–31. DOI: <https://doi.org/10.1016/j.jas.2016.03.001>
2. **Hodson, MJ, Parker, AG, Leng, MJ and Sloane, HJ** 2008 Silicon, oxygen and carbon isotope composition of wheat (*Triticum aestivum* L.) phytoliths: implications for palaeoecology and archaeology. *Journal of Quaternary Science*, 23: 331–339. DOI: <https://doi.org/10.1002/jqs.1176>
3. **Hodson, MJ, Song, Z, Ball, TB, Elbaum, R and Stuyt, E** 2020 Editorial: Frontiers in Phytolith Research. *Frontiers in Plant Science*, 11(454): 5–7. DOI: <https://doi.org/10.3389/fpls.2020.00454>
4. **Kansa, SW, Atici, L, Kansa, EC and Meadows, RH** 2020 Archaeological analysis in the information age: guidelines for maximising the reach, comprehensiveness and longevity of data. *Advances in Archaeological Practice*, 8(1): 40–52. DOI: <https://doi.org/10.1017/aap.2019.36>
5. **Lodwick, L** 2019 Sowing the seeds of future research: data sharing, citation and reuse in Archaeobotany. *Open Quaternary*, 5(7): 1–15. DOI: <https://doi.org/10.5334/oq.62>
6. **Madella, M, Alexandre, A and Ball, T** 2005 International code for phytolith nomenclature 1.0. *Annals of Botany*, 96: 253–260. DOI: <https://doi.org/10.1093/aob/mci172>
7. **Marwick, B, et al.** 2017 Open Science in Archaeology. *The SAA Archaeological Record*, 17(4): 8–14. DOI: <https://doi.org/10.31235/osf.io/72n8g>
8. **Marwick, B and Pilaar Birch, SE** 2018 A standard for the scholarly citation of archaeological data as an incentive to data sharing. *Advances in Archaeological Practice*, 6(2): 125–143. DOI: <https://doi.org/10.1017/aap.2018.3>
9. **Masuzzo, P and Martens, L** 2017 Do you speak open science? Resources and tips to learn the language. *PeerJ Preprints*, 5: e2689v1. DOI: <https://doi.org/10.7287/peerj.preprints.2689v1>
10. **Pearsall, DM** 2019 *Case studies in Paleoethnobotany*. New York: Routledge. DOI: <https://doi.org/10.4324/9781351009683>
11. **Piperno, DR** 2006 *Phytoliths: a comprehensive guide for archaeologists and palaeoecologists*. Oxford: Altamira Press.
12. **Yang, S, Hao, Q, Wang, H, Van Zwieten, L, Yu, C, Liu, T, Yang, X, Zhang, X and Song, Z** 2020 A review of carbon isotopes of phytoliths: implications for phytolith-occluded carbon sources. *Journal of Soils and Sediments*, 20: 1811–1823. DOI: <https://doi.org/10.1007/s11368-019-02548-4>
13. **Zuo, X and Lu, H** 2019 Phytolith Radiocarbon Dating: A Review of Previous Studies in China and the Current State of the Debate. *Frontiers in Plant Science*, 10, 1302. DOI: <https://doi.org/10.3389/fpls.2019.01302>

**How to cite this article:** Karoune, E 2020 Data from "Assessing Open Science Practices in Phytolith Research". *Journal of Open Archaeology Data* 8: 6. DOI: <https://doi.org/10.5334/joad.67>

**Published:** 18 September 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.