Journal of open archaeology data

Open-archaeo: A Resource for Documenting Archaeological Software Development Practices

DATA PAPER

ZACHARY BATIST
JOE ROE
*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

Open-archaeo (https://open-archaeo.info) is a comprehensive list of open software and resources created by and for archaeologists. It is a living collection—itself an open project—which as of writing includes 548 entries and associated metadata. Open-archaeo documents what kinds of software and resources archaeologists have produced, enabling further investigation of research software engineering and digital peer-production practices in the discipline, both under-explored aspects of archaeological research practice.

CORRESPONDING AUTHOR:
Zachary Batist
Faculty of Information, University of Toronto, CA
z.batist@mail.utoronto.ca

KEYWORDS:
Archaeological practice; open source software; web-based resources

TO CITE THIS ARTICLE:
Batist Z, Roe J 2023 Open-archaeo: A Resource for Documenting Archaeological Software Development Practices. *Journal of Open Archaeology Data*, 11: 9, pp. 1–7. DOI: https://doi.org/10.5334/joad.111

(1) OVERVIEW

CONTEXT

This dataset documents research software engineering and digital peer-production practices in archaeology. More specifically, it is a comprehensive list of software and other technical resources that archaeologists created to support their work.

The data were compiled from web-based platforms with international reach. We did not trace the specific locations in which software were developed, however the means through which the data were collected introduced a potential bias that emphasized including projects based in Europe and North America.

The earliest project included in the dataset was initiated in 2005. Open-archaeo is a living project and we continue to add new items as archaeologists develop new software that fit our inclusion criteria. The primary source repository for the list is hosted on GitHub (https://github.com/zackbatist/open-archaeo) as an open source project. This is used to build a more user-friendly web interface (https://open-archaeo.info) and is periodically archived in the *Zenodo* long-term repository (https://doi.org/10.5281/zenodo.8299651). The most recently updated entries were still under active development at the time this paper was published, and a snapshot of the database was archived.

(2) METHODS

STEPS

We compiled the dataset by browsing collaborative software development platforms such as GitHub, GitLab, and Codeberg. This entailed manually crawling through users' profiles, particularly those who identify as archaeologists or who have contributed to tools that pertain to archaeological work. We also traced connections to other personal, professional, and institutional websites that describe and host additional archaeological software.

Both authors are most active on GitHub, which potentially means the other code collaboration platforms are under-represented in the dataset. It is also likely that our ad-hoc discovery process favors large, centralized platforms (e.g. GitHub, GitLab, Codeberg) over decentralized alternatives (e.g. self-hosted GitLab or gitea/forgejo). Considering recent calls to reorient online scholarly communication toward federated and community-owned infrastructure [1], we intend to prioritize addressing this bias in the future development of open-archaeo.

SAMPLING STRATEGY

While our initial intention was to only list open source software [2], open-archaeo's scope has expanded to include all software created by and for archaeologists. This is due to the ambiguous meaning of the term "open" [3: 302], and the common practice of sharing code with no license at all.

The main data file (open-archaeo.csv) accounts for several variables whose values were manually copied from other web-based resources, including code-hosting platforms, package registries, social media accounts, and personal and project websites. Our conceptual model (Figure 1) documents associations between information recorded in open-archaeo and in the source material.

QUALITY CONTROL

We annotated each record with tags that describe what aspect of archaeological research each tool or resource was meant to address. We assigned tags based on how developers identified their projects' purpose and scope. The list of tags and their associated definitions is subject to grow or change as archaeologists continue to develop software for additional use cases, but right now there are 56 tags (see Table 1).

We also categorized records based on how each tool or resource is meant to be accessed or used (see Table 2). This provides a sense of the pervasiveness of various development models, and the requisite technical capabilities that developers assume users hold.

CONSTRAINTS

Open-archaeo is a relatively comprehensive list of recently created resources, however it generally lacks software or resources created before the growth and general use of collaborative software development platforms, such as GitHub, among archaeologists. It also lacks references to software and resources that are not available on the internet.

Like any online directory, open-archaeo is susceptible to "link-rot" when resources are changed, moved, or deleted after inclusion. To mitigate this issue, open-archaeo includes persistent identifiers (e.g. DOIs to stable repositories) wherever possible. We also periodically archive snapshots of all git repositories and other resources hosted on the web with the Internet Archive (https://archive.org/details/open-archaeo).

(3) DATASET DESCRIPTION

OBJECT NAME

open-archaeo.csv – The main data file. We describe each column's meaning in Table 3.

tags.md – Descriptions of tags and categories.

open-archaeo.graphml – Graph documenting the conceptual model.

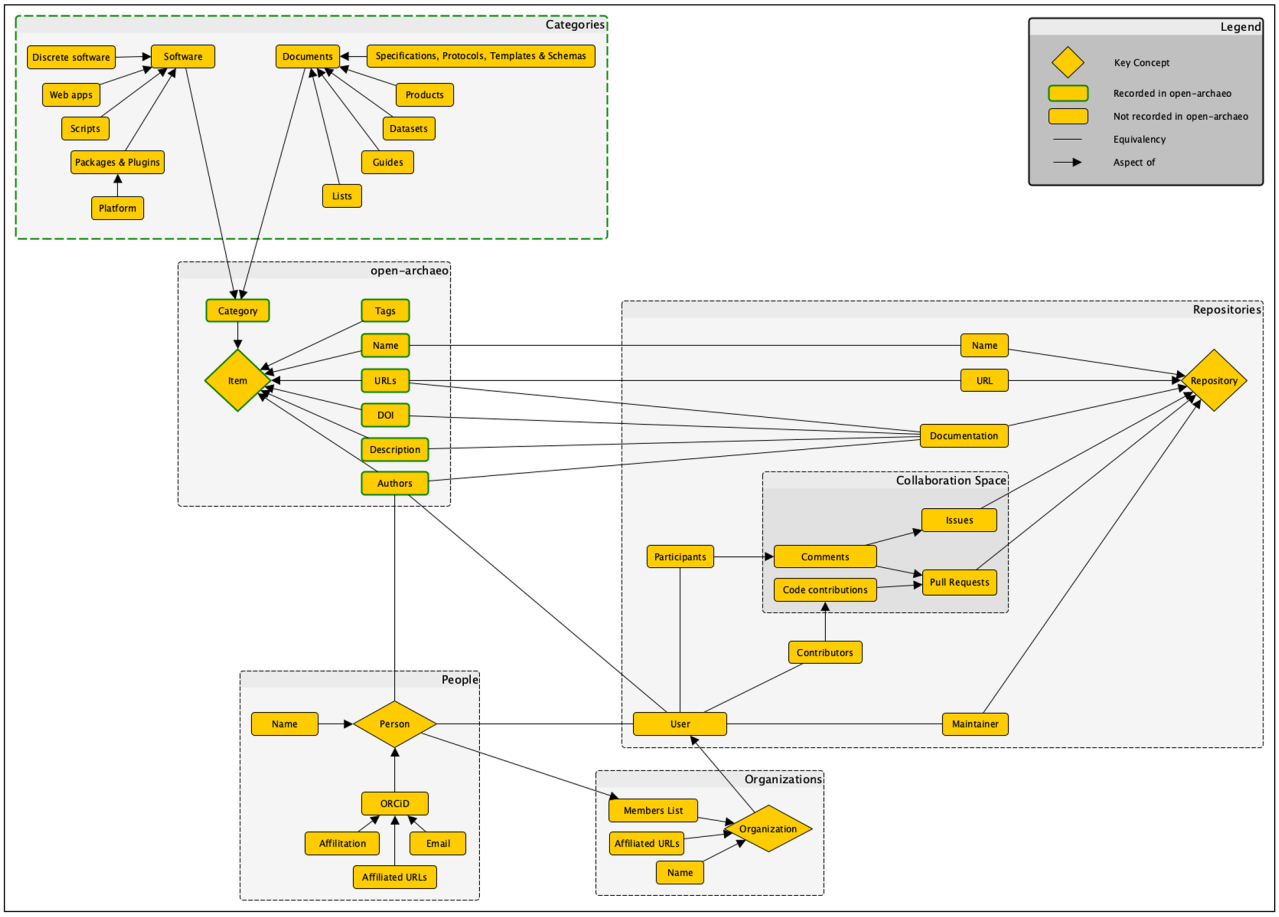

Figure 1 Conceptual model documenting relationships between data recorded in open-archaeo and other relevant information in the source material and elsewhere on the web.

TAG	SCOPE
3D modelling	Creation and manipulation of computer-generated three-dimensional models.
Aerial and satellite imagery	Techniques for capturing images of the earth from aerial perspectives.
API interfaces and web scrapers	Tools that facilitate retrieval of data from web resources.
Archaeoastronomy	Concerned with the investigation of the astronomical knowledge of prehistoric cultures.
Archaeogenetics	Analysis of degraded DNA from archaeological or anthropological sources, including skeletons of humans, domesticate animal or plants, artefacts, or cave sediments.
Artefact morphology	Identification and analysis of artefacts' physical dimensions and shape.
Augmented reality	Creating a composite view comprising computer-generated images superimposed on real world environments.
Bibliography	A list of scholarly work relevant to a particular topic.
Biological anthropology	Concerned with the biological and behavioral aspects of human beings, their extinct hominin ancestors, and related non-human primates.
Bits and bobs	Tools that do not necessarily pertain to a specific archaeological activity, but which may be particularly useful in practical matters commonly faced by archaeologists.
Chronological modelling	Identification and analysis of patterns relating to the distribution of events or timespans.
Cultural evolution	Concerned with evolutionary theory of social change.
Data collection	Tools used to facilitate the orderly data collection and input into digital data management systems.
Data management	Tools used to organize and give structure to information.
Datasets	Series of records, collated in a consistent manner, motivated by a need, desire or warrant to render them comparable.

(Contd.)

TAG	SCOPE
Dendrochronology	Techniques for dating events and environmental change by using the characteristic patterns of annual growth rings in timber and tree trunks.
Diagrams and visualizations	Tools for summarizing and presenting in abtracted graphical form.
Drivers and IO	Tools that facilitate the transmission of information across digital devices.
Drones	Semi-automated machines used to capture data in environments that are unsuitable or unsafe for direct human presence.
Educational resources and practical guides	Documents that convey instructions regarding how to achieve particular goals.
Ethics and professional development	Concerned with the social, moral and ethical frameworks that archaeologists adhere to or are impacted by as they go about their work.
Games	A structured form of play typically undertaken for entertainment or fun.
Geoarchaeology	Application of geology and earth science methods.
Geophysical survey	Systematic collection and analysis of geophysical signals, such as magnetic and gravitational fields emanating from the earth.
Harrix matrix	Representation of archaeological sites as series of connected abstract entities.
Iconography	Techniques for examining ancient symbols.
Instrumental Neutron activation analysis	Technique for identifying the elemental composition of physical objects by bombarding samples with neutrons.
LiDAR	A method for determining variable distance by targeting an object with a laser and measuring the time for the reflected light to return to the receiver.
Lists	A set of items, not necessarily recorded with systematic intent.
Literary analysis and epigraphy	Techniques for examining ancient texts.
Lithic analysis	Identification and analysis of stone tool artefacts.
Luminescence dating	Absolute dating techniques that rely on measuring doses of ionizing radiation.
Machine learning	Use of data and algorithms to imitate the way that humans learn, gradually improving computer performance on some set of tasks.
Museums	Institutions responsible for curating, conserving or exhibiting collections of artefacts or other objects.
Network analysis	Process of investigating social structures through the use of networks and graph theory.
Palaeobotany	Techniques used to identify and analyze plant remains.
Photogrammetry	Techniques for capturing and processing spatial information based on comparison of images taken from various perspectives.
Photography	The practice of photographing archeological entities to create a lasting record of that work.
Platforms and publications	Venues that host, validate, review or contextualize published work.
Public archaeology	The practice of presenting archaeological findings to the general public.
Public policy and civic action	The practice of creating or responding to principles that govern how archaeology should be done and its role in society.
Radiocarbon dating, calibration and sequencing	Methods for determining and comparing the ages of objects using the properties of radiocarbon.
Schemas and ontologies	A pattern of thought or behavior that organizes categories of information and the relationships among them.
Seriation	Technique for identifying chronological sequences based on the ordered properties of archaeological assemblages.
Shape recognition	Techniques for automated classification of objects based on input representations of their physical shapes or morphological properties.
Simulation	Concerned with modelling archaeological processes in abstract terms.
Site mapping	Techniques for illustrating archaeological sites.
Spatial analysis	Identification and analysis of patterns relating to spatial distributions or relationships.

(Contd.)

TAG	SCOPE
Stable isotope analysis	The identification of isotopic signatures relative abundance of chemical elements within organic and inorganic compounds to understand the flow of energy through a food web and to reconstruct past environmental and climatic conditions.
Statistical analysis	Identification and analysis of quantitative distribution patterns.
Templates	Sample documents that already have some details in place.
Viewshed analysis	Techniques for considering localized spatial perspectives of landscapes.
Virtual reality	Simulating an environment in a 3D interactive virtual space.
Writing	Communicating ideas through written language.
X-Ray Fluorescence	Technique for identifying the elemental composition of physical objects by bombarding samples with high-energy radiation.
Zooarchaeology	Techniques used to identify and analyze faunal remains.

Table 1 List of tags used to describe open-archaeo records.

CATEGORY	KIND	SCOPE
Packages and libraries	Software	Sets of functions assembled with clear purpose and made accessible using standards established by an underlying platform.
Scripts	Software	Sets of pragmatically assembled mutable functions, often lacking complete documentation or adherence to protocols that would otherwise facilitate secondary use outside their original contexts of creation.
Standalone software	Software	Software that may be operated without needing to first access an underlying platform.
Guides	Documents	An educational resource or documented protocol meant to instruct readers how to apply relevant tools or techniques.
Lists and datasets	Documents	A series of consistently organized observations assembled with purpose.
Products	Documents	Stable outcomes of creative work.
Specifications, protocols and schemas	Documents	A formal data structure or framework intended to be used as a model.

Table 2 List of categories that indicate how each tool or resource is meant to be accessed or used.

COLUMN	SCOPE
item_name	Name of the software or resource.
description	Brief description of an item, adapted from the maintainers' own words.
github	URL of a project hosted on the GitHub collaborative coding platform.
gist	URL of a project hosted on the Gist collaborative coding platform.
gitlab	URL of a project hosted on the GitLab collaborative coding platform.
bitbucket	URL of a project hosted on the Bitbucket collaborative coding platform.
launchpad	URL of a project hosted on the Launchpad collaborative coding platform.
twitter	URL of a Twitter account.
youtube	URL of a YouTube channel, playlist, or video.
blogpost	URL of a specific blog post.
cran	URL of a record on the Comprehensive R Archive Network (CRAN).
pypi	URL of a record on the Python Package Index (PyPI).
codeberg	URL of a project hosted on the Codeberg collaborative coding platform.
website	URL of a website other than that which is hosted on a collaborative coding platform.
publication	DOI pertaining to associated scholarly publications.
DOI	DOI pertaining to the software or resource.
author1_name	Name of an item's primary author or maintainer. Contains username on collaborative coding platform, if available.
author2_name	Name of an item's additional author or maintainer. Contains username on collaborative coding platform, if available.

(Contd.)

COLUMN	SCOPE
author3_name	Name of an item's additional author or maintainer. Contains username on collaborative coding platform, if available.
author4_name	Name of an item's additional author or maintainer. Contains username on collaborative coding platform, if available.
author5_name	Name of an item's additional author or maintainer. Contains username on collaborative coding platform, if available.
category	Term that describes how an item is meant to be accessed or used.
platform	Names of the platforms that packages and libraries are built upon and extend from,
tag1	Term that describes what aspect of archaeological research each item is meant to address.
tag2	Term that describes what aspect of archaeological research each item is meant to address.
tag3	Term that describes what aspect of archaeological research each item is meant to address.
tag4	Term that describes what aspect of archaeological research each item is meant to address.
tag5	Term that describes what aspect of archaeological research each item is meant to address.
notes	Notes documenting issues or concerns regarding the construction or representation of an open-archaeo record.
internetarchive	URL of a snapshot hosted in the open-archaeo Internet Archive collection.

Table 3 Descriptions of column headers appearing in open-archaeo.csv.

DATA TYPE
Primary data derived from user-created content and the authors' classificatory schemes.

FORMAT NAMES AND VERSIONS
.csv, .md, .graphml

CREATION DATES
Open-archaeo began in December 2018. It is a live project, and we continue to update it as archaeologists develop new software and resources that match the criteria for inclusion.

DATASET CREATORS
Zachary Batist is open-archaeo's founder and primary maintainer. He is responsible for finding new items and adding them to the dataset. He is also the main point of contact for many people who recommend items to be included in the dataset.

Joe Roe systematized the schema and developed a framework that parses and transforms the dataset so that it may be presented using front-end web environments. This is crucial for making open-archaeo useful as an index that archaeologists may use to find resources that may support their work. This framework also enables us to query the APIs pertaining to the platforms on which projects are hosted, which supports more comprehensive research on how archaeologists develop software.

Numerous other people contributed to open-archaeo's growth and development. Contributions ranged from collecting, suggesting, and organizing content, and supporting front-end web rendering. Early discussions with community members in GitHub Issues also helped define open-archaeo's scope and helped give it purpose. The GitHub repository contains more information on who contributed and in what capacity.

LANGUAGE
English.

LICENSE
CC0 1.0 Universal

REPOSITORY LOCATION
https://doi.org/10.5281/zenodo.8299651

PUBLICATION DATE
August 29, 2023

(4) REUSE POTENTIAL

Open-archaeo is a comprehensive index of software and other resources created by archaeologists. It therefore serves as a potential dataset from which we might examine an under-explored kind of research practice, namely, research software engineering. Our indexical approach, which involves assembling unique identifiers for each record, enables the base dataset to be extended through a series of semantic linkages spanning the open web [4]. Open-archaeo therefore serves as an initial point of departure, supporting a variety of analytical applications, according to their individual needs and research strategies.

For example, in a forthcoming study we used open-archaeo as an index to programmatically retrieve data on collaborative software development in archaeology from the GitHub API. The data retrieved was extremely rich, including detailed histories of contributions to each project (commits, issues, and comments), which users contribute to and maintain projects, and automated classifications of the repositories' contents. Alternatively, open-archaeo may serve as a jumping-off point to conduct qualitative analyses of commit messages, issues,

comments, file naming and organization conventions, community management guidelines, and other textual or visual media. Furthermore, the extensive application of tags and categories facilitates comparison between archaeologists working in various fields or detailed investigations of work performed within specific areas of interest.

Moreover, we put significant effort to ensure that the dataset is compiled systematically. This invites the possibility of linking the terms we use in our conceptual model with standard ontologies pertaining to digital and archaeological research practices (e.g., CIDOC-CRM, NeMO). This would facilitate comparison with other similar indexes maintained for the Digital Humanities and other related disciplines.

Finally, open-archaeo is a central hub in the digital archaeology community and showcases archaeologists' immense creative capabilities. It is a valuable resource for discovering reliable and cost-effective software that is relevant to their work. Many archaeologists do not know about prior work that could support their research, and open-archaeo helps bring these tools to their attention so that they can reuse or adapt existing code for their own purposes, rather than create redundant software from scratch. Additionally, open-archaeo maintains a separate list of tools or resources in demand, which serves as a place where community members may draw attention to software that do not yet exist but would be nice to have, allowing others to step in and collaborate on a new tool (https://github.com/zackbatist/open-archaeo/blob/master/ToDo.md).

ACKNOWLEDGEMENTS

Open-archaeo was supported by the enthusiasm of members of the Scientific Scripting Languages in Archaeology special interest group of Computer Applications in Archaeology (CAA) International. Direct contributors to the project on GitHub include, as of writing, Martin Hinz, James Fellows Yates, Dirk Seidensticker, Clemens Schmid, Bjørn Peare Bartholdy, Simon Hohl, and Nika Shilobod. We are also grateful to Stefano Costa for the suggestion to archive resources with the Internet Archive.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Zachary Batist orcid.org/0000-0003-0435-508X
Faculty of Information, University of Toronto, CA
Joe Roe orcid.org/0000-0002-1011-1244
Institute of Archaeological Science, University of Bern, CH

REFERENCES

1. **Brembs B, Lenardic A, Murray-Rust P, Chan L, Irawan DE.** Mastodon over Mammon: Towards Publicly Owned Scholarly Knowledge. *Royal Society Open Science*. 2023; 10(7): 230207. DOI: https://doi.org/10.1098/rsos.230207
2. **Open Source Initiative.** 2007 The Open Source Definition. https://opensource.org/osd/.
3. **Kelty CM.** Two Bits: The Cultural Significance of Free Software; 2008. Duke University Press. DOI: https://doi.org/10.1215/9780822389002
4. **Wells JJ, Kansa EC, Kansa SW, Yerka SJ, Anderson DG, Bissett TG, Myers KN, DeMuth RC.** Web-Based Discovery and Integration of Archaeological Historic Properties Inventory Data: The Digital Index of North American Archaeology (DINAA). *Literary and Linguistic Computing*. 2014; 29(3): 349–60. DOI: https://doi.org/10.1093/llc/fqu028

TO CITE THIS ARTICLE:
Batist Z, Roe J 2023 Open-archaeo: A Resource for Documenting Archaeological Software Development Practices. *Journal of Open Archaeology Data,* 11: 9, pp. 1–7. DOI: https://doi.org/10.5334/joad.111

Published: 09 September 2023