

Open Context Data Publishing

**EDITORIAL POLICIES
&
AUTHOR GUIDELINES**



Copyright ©2012, The Alexandria Archive Institute. Some rights reserved. This work is licensed under a Creative Commons Attribution 3.0 License (<http://creativecommons.org/licenses/by/3.0/>).

CONTENTS

1. EDITORIAL POLICIES	2
1.1 Focus & Scope	2
1.2 Open Access	2
1.3 Licensing & Citation	3
1.4 Professional Review	3
1.5 Turnaround Time.....	4
2. AUTHOR GUIDELINES.....	5
2.1 How it Works: Publishing your Data in Open Context.....	5
2.2 Preparing your Dataset for Digital Publication	5
2.2.1 Required Information	5
2.2.2 Additional Tips	6
2.2.3 Scientific Review Criteria by Content Type.....	7
2.3 Data Publication Checklist	9
2.4 Publishing Fees	9
2.4.1 Explanation of Fees	9
2.4.2 Information for Grant Seekers	10

1. EDITORIAL POLICIES

1.1 Focus & Scope

Thank you for considering Open Context to publish your research data.

Open Context is an online, open access publisher of editorially reviewed research data. We welcome contributions from both established and emerging scholars from anywhere in the world. The Open Context editorial team collaborates with contributors to publish high-quality research data with appropriate documentation and alignment to accepted standards. Our publication processes help make datasets easier to discover, use, and relate to data published by initiatives across the Web. We also help contributors archive and preserve their research through collaboration with the California Digital Library, one of the world leaders in digital preservation and curation. Though centered on archaeology, Open Context also welcomes datasets from other related disciplines.

Open Context works to complement manuscript (book, article, report) publication by providing a venue for the dissemination of structured datasets (spreadsheets and databases), media files, journal and notes files, and other documentation. We also publish stand-alone datasets that are not related to conventional publication. Open Context integrates and interrelates project content to enable powerful search, visualization, and data analysis capabilities. Thus, Open Context not only helps share and preserve the rich digital documentation created by researchers, but also magnifies the impact and use of data (whether “born digital” or converted to a digital).

Our editorial policies differ from many journals in that we do not accept or reject datasets based on a perception of “significance.” Even a seemingly dull dataset may contribute to future research, even if only by increasing the statistical power of future meta-analyses. Our philosophy regarding data publication is that researcher-created datasets are potentially useful, either now or in the future, if they are properly documented and reviewed to improve quality and internal consistency.

1.2 Open Access

All content published in Open Context is freely and openly accessible. Open Context requires no login to access and download data. We believe open access and open licensing of research data are powerful tools for encouraging better and more collaborative scientific practice. However, they are not universally appropriate. Open Context's editors ask contributors to carefully consider issues of intellectual property, data management, and data access, and communicate their perspectives with the Open Context editorial staff. We expect users of Open Context to exercise appropriate ethical behavior and respect the integrity of Open Context's contributors,

stakeholders, and content. If you have any concerns about the ethical nature of the information presented in Open Context, please contact the Editor, Sarah Witcher Kansa (skansa@alexandriaarchive.org).

1.3 Licensing & Citation

Open Context publishes open access, editorially reviewed datasets to facilitate future research and instructional opportunities. To achieve this, data contributors must make their content legally usable by others. To ensure legal reuse, we require that all content be released to the public domain, or that contributors use Creative Commons (<http://creativecommons.org>) copyright licenses on their content. We encourage authors to select one of the following two options:

- **CC BY**: We strongly recommend that users select the Creative Commons “Attribution” (CC BY) license. This is the easiest license to interpret and makes data as widely useful as possible. While we allow licenses that restrict commercial uses, we recommend against such restrictions, which are inherently ambiguous and would inhibit important uses (such as inclusion of content in textbooks or even journals distributed through sales).
- **CC 0**: In the United States, copyright applies to expressive works, not compilations of factual information. Therefore, Creative Commons copyright licenses are not appropriate for some datasets, especially those with limited “expressive” content. Datasets that are less expressive and have less “authorial voice” tend toward a more “scientific” and factual nature (i.e., those that mainly include physical measurements and adhere to widely used conventions in nomenclature and recording). These datasets should use the Creative Commons-Zero (public domain) dedication.

Please note that copyright and licensing issues are largely independent of scholarly citation and attribution. Professional standards dictate that all users properly cite data contributors even for public domain content, especially for scholarly uses. This professional norm of conduct works independently of the copyright status of content.

Authors should choose a single license to apply to their entire dataset; however, if needed, we can also assign different license choices to individual items.

To facilitate citation, Open Context provides a “suggested citation” for each item, data table, and media resource. As in conventional citation, Open Context citations contain author(s) and project name(s). In addition, Open Context provides a stable URL unique to the item(s) being cited and a date/time for the citation.

1.4 Professional Review

Open Context content must pass professional editorial scrutiny before it can be published on the Web. Open Context accepts content from professional / accredited researchers, students, government officials, and museum staff. Contributors must be able to demonstrate adherence to

appropriate legal, ethical, and professional standards of conduct and methodological rigor. An Editorial Board comprised of experts in various fields reviews all content published in Open Context. Unlike printed synthetic publications, data publications in Open Context do not undergo editorial review for the “scholarly merits” of the content, since future merits cannot be known. Rather, Open Context editors review content for clarity and robustness (Is the content clean? Does it make sense? Does it have sufficient metadata documentation?).

Data authors may request professional peer review of their contributions. If peer review is desired, authors must provide names and contact information for possible reviewers, as well as a brief description of the expertise needed to make informed judgments of the content for which peer review is sought. Content that undergoes this additional review will be highlighted as such in Open Context.

1.5 Turnaround Time

We aim to publish datasets as quickly as is consistent with a thorough evaluation of the content. Authors can speed the publication process by ensuring their submitted content follows the *Author Guidelines* in the following section. A well-prepared dataset normally takes 30 to 60 days to undergo editorial review, import, and publication. Very large and complex datasets (multi-year projects with many media files, and contributions from many specialists) may take longer and may require extended collaboration between Open Context’s Editorial Board and contributing researchers.

2. AUTHOR GUIDELINES

2.1 How it Works: Publishing your Data in Open Context

Digital data publication is a similar process, in many ways, to print publication. The first step is to contact the editors by email (publishing@opencontext.org). Provide a brief description of your dataset. Datasets can be small or large, simple or complex, linked to a print publication or stand-alone. The goal is to make publication of quality content as easy as possible so that other scholars can access it, understand it, and employ it in new research.

To ensure speedy publication of your data, please review carefully the following sections and provide all required information.

2.2 Preparing your Dataset for Digital Publication

Just like a print publication, some preparation is required before publishing digital data. Most of this preparation involves documenting a dataset so that it can be understood by Open Context's editors and future users of the content, including other researchers and students. This section describes how to prepare your content to facilitate its publication in Open Context.

2.2.1 Required Information

Please provide the Editor with a document including all of the following items. These form the basis of your project's descriptive information (metadata). Open Context editors may contact you for additional information.

1. **Title:** A short descriptive name of the dataset
2. **Creator:** Name the creator(s) of the dataset (agents for attribution), as well as their affiliation(s) and/or other credentials for performing the work.
3. **Site Name:** Provide the site name(s)
4. **Location:** The exact geographic location of the site (if site security is a problem, you may provide a less precise location. Please contact the editors if you would like to discuss options.
5. **Period:** A cultural period and/or calendar date range for your dataset.
6. **Short Description:** A 140-character (max.) description of the dataset
7. **Keywords:** A few short terms or phrases that describe your dataset
8. **Abstract:** A narrative description of the dataset, sufficiently detailed to guide informed users about the nature of the data. This should include introductory information describing the project goals, key findings, as well as methods and recording systems. For

large projects, contributors can also provide additional supplemental background descriptions of specialist analyses. The abstract should provide future users with sufficient understanding of the aims of the original project and details on the specific methods employed, so that the new user can determine whether the dataset is appropriate for his/her re-analysis.

- 9. Methodological Notes:** A narrative describing data collection procedures and methods that may be significant in the interpretation and use of the dataset.
- 10. Potential Applications of Data:** Briefly describe potential future research applications of your dataset.
- 11. Related Publication(s):** If applicable, provide bibliographic references to related publications and published datasets. Provide links to online publications.
- 12. Table Field Descriptions:** Provide short descriptions of all the column headings in your spreadsheet.
- 13. License Choice:** Indicate your selection of a license for your work: (1) Creative Commons Zero (CC-0) or (2) Attribution (CC-BY). See section 1.3 for more information on these license choices. Alternatively, you may indicate that you would like to negotiate a special license with the editor (such as using different licenses for certain items).

2.2.2 Additional Tips

In addition to providing the above required information, please review the following list of general tips to ensure that your dataset is tidy and well-documented.

- 1. Data Formats and Structures:** Data for import should be in Microsoft Excel tables. The first row (“row 1”) of the table should contain data field names (columns). The other rows should have the data records in the table, with each data record listed in a separate row. If you do not have Excel or cannot produce Excel spreadsheets from your database, we can also handle Filemaker, Access, and Open Office, as well as comma separated value (.csv) files. Please note, however, that you must first extract images and other media from a database (if stored in “binary fields”) and store them as individual files. The project abstract/background should be in Microsoft Word or a similar format. In addition, you may also provide as much supporting or related documentation as you like, such as PDFs of related publications, extended bibliographies in Word format, and links to related web resources (such as descriptive project web sites, profiles of project participants on their institutional websites or links to self-archived publications related to the dataset).
- 2. Clean-up and Edits:** Nominal values (terms used over and over again), such as those used to describe artifacts in a small finds database (“lamp,” “coin,” “spindle-whorl”), should be consistent in terms of plurals, terminology, and spelling to aid search and understanding. Identifiers for objects or contexts (such as “catalog #,” “locus #”) should also be free of errors.
- 3. Decoding:** To speed up data entry, many people use coding systems as a convenient way to record data. However, these coding systems may be unintelligible without explanation. To facilitate understanding and reuse of datasets, we request that data contributors replace code with intelligible text when they submit the data for publication.

4. **Provide “keys”:** Related to item #3, if your data was originally in a relational database, please provide the primary and secondary (foreign) “keys” in each table to aid the editor in relating the data tables in Open Context.
5. **Images and Media:** Please clearly link each media file to one or more specific records in your dataset (such as a find number, a locus, a person, a daily journal entry, etc.). Prepare a new Excel table that lists each image file name, an image description (if desired), the number / identifier of the object or place the image describes, and the folder in which the image is stored. All images should be kept in a separate folder, each labeled with a unique name. That is, images should not be embedded in the spreadsheet. If you have images embedded in a relational database, please save them to a separate folder and prepare an Excel file listing all images, as described here.
6. **People and Attribution:** For citation purposes, every record in Open Context must be attributed to one or more specific person(s). If any individuals other than the data contributor deserve attribution for the data being published, please provide their name and affiliation, and clearly indicate which data should be associated with their name.

2.2.3 Scientific Review Criteria by Content Type

In addition to supplying the Required Information (section 2.2.1) and reviewing the Additional Tips (section 2.2.2), data authors should ensure they have provided information specific to their content type. That is, datasets should be described in as much detail as needed for other researchers to use appropriately. Datasets should generally adhere to conventions in a given specialization, region, and time period. Where datasets diverge from conventions (say where a researcher used new / innovative methods and recording strategies), the divergences need to be clearly documented and described.

These requirements will play out in different ways, depending on the discipline; however, the main aim of data authors should be to make their published data intelligible and usable to informed specialists in a given area. Some examples are:

Faunal data should include a description of methodologies such as quantification, sampling strategies, identification practices, access to reference materials, methods for aging and sexing, and description of measurements taken.

Microarchaeological data should include taxonomies and typologies that follow those used in other disciplines (zooarchaeology, paleoethnobotany, etc.). All measurements should be in metric (liters, millimeters, grams). Count and weight ratios are raw count or weight (in grams) divided by the volume of soil in the sample before processing via flotation or screening. The size of a “microartifact” varies in size by the researcher, but most are 1cm/10mm or smaller.

Information that should be included in microarchaeology data publications are:

- Identification or category (fish scale, coarse ceramic, etc.)
- Count/weight densities or percentages for each size category (>5.6mm, >4mm etc.) or for the size class under investigation
- Volume of sample (pre-flotation)

- Context (e.g., pit, oven, surface)
- Size (in mm)
- Retrieval method (flotation, wet screening, dry sieving, etc.)
- Equipment used (stereomicroscope, binocular visor, etc.)
- Site, area, level
- Information on field and lab-based sampling strategies
- Information on identification strategies: type of microscope used, reference collections, specialist colleagues consulted (such as zooarchaeologists, paleoethnobotanists, ceramic specialists)
- Images to support identifications

Ancient architecture data should include information about the project methodology, technology / software used, and data collection / survey techniques. Methods and types of equipment used will vary depending on the material, the structure, and the research aims.

GIS data should include extensive metadata, the documentation on data collection methods. These metadata should include (as applicable): the layer-naming system, specific definition of coordinate systems, geographic coordinates of survey datum points, precise declination of grid or survey lines to a cardinal direction (geographic or magnetic), relation of vertical measurements to arbitrary or geographic standards, precision of measurements (larger projects may have multiple categories), use of 3D, set viewpoints, definition of hardware and software used to create the data, and any other descriptive information necessary for users to reconstruct spatial arrangements from the supplied data. Open Context does not demand any particular metadata standard, but authors should be as descriptive as possible to the extent that geospatial data are important to the project (see <http://www.fgdc.gov/metadata>). Final data models should be visually inspected by the authors from various angles, in order to check for any geometric irregularities.

Open Context prefers GIS data to be submitted in the [WGS 84 standard](#) for geographic measurements. The preferred coordinate system is [UTM](#); UTM zones and the accuracy and precision of metric measurements should be communicated in the metadata. Larger survey projects spanning multiple UTM zones may prefer to communicate coordinate data in [decimal degrees](#), also with accuracy and precision communicated in the metadata (e.g., 0.0000001 DD precision \approx 1 cm at the equator). Our staff can provide guidance to translation assistance, and we are willing to discuss reasons to maintain geographic data in other standard or legacy formats.

GIS data compilation and translation can be hindered by the availability of popular proprietary software packages (especially the ESRI ArcGIS products). Authors submitting GIS data to Open Context may wish to obtain lower-cost or free GIS research software or data management utilities to complete their work. Recommended options include:

- [Manifold GIS](#) - low-cost, proprietary GIS software
- [Quantum GIS \(QGIS\)](#) - free and open source GIS software
- User-friendly Desktop Internet GIS ([uDig](#)) - free and open source GIS software
- [US National Geodetic Survey](#) - free data utilities created by the NGS community

Authors are encouraged to investigate these and other options to find the solution that best suits their needs. Our staff can provide guidance on choices, all those listed above are research-quality and have robust user guides and support communities.

Open Context requires that all data authors provide professional assurance that GIS data to be published are clear of legal restrictions regarding site security (as defined by legal statutes and government regulatory agencies), or privacy rights of landowners (as defined by legal statutes or local standards). If data sensitivity is a concern, our staff can provide guidance on translation of data to an acceptable level of granularity.

Human osteological data should include un-coded data. Coded data may accompany un-coded data; however, the author must provide clear descriptions of codes that match accepted norms in the field. Any exceptions to standard methodologies should include a logical explanation for doing so. The data author should describe how the data were collected, by whom, and under what circumstances (e.g. was the research part of a larger project, how many people had responsibility for the data and over what period of time?). The author should also include information on how the collections were accessed and their current disposition. Data fields should not be left blank (that is, the code for unobservable should be entered in all cases so as not to cause confusion regarding incomplete data collection).

Lithic data should include a short description of methodology including sampling and collection strategy, typological considerations, identification and nomenclature, and description of any measurements taken of both debitage and tools. If possible, the final disposition of the assemblage should also be included.

Don't see your content type here? Please [contact us](#).

2.3 Data Publication Checklist

Be sure to include all of the following in your email to Open Context. If you are sending large numbers of files, [contact us](#) to make arrangements for file transfers.

- Dataset (preferably in Excel or similar format)
- Other content (such as images, with a worksheet listing each image name and its associated item in the spreadsheet, such as a unique catalog number or a locus number).
Note: images should be separate from the dataset, i.e. Excel spreadsheet, not embedded.
- Completed required information (from section 2.2.1).

2.4 Publishing Fees

2.4.1 Explanation of Fees

To support open access publication and archiving, Open Context has developed a pricing structure based on a one-time “contributor-pays” model. Publication fees generally range

between \$250 and \$6000 depending on the complexity and size of the contributed database and related content. For example, a single spreadsheet of faunal data, with no related images would cost on the low end of this spectrum. In contrast, a complex project with several databases, specialist analyses, and thousands of media files, would be on the high-end of this scale. Open Context developers can provide additional fee-based services for implementations based on Open Context's Web-services (API) or other customizations. For assistance in budgeting, interested contributors should [contact the editors](#) to establish a fee for their specific project.

2.4.2 Information for Grant Seekers

An increasing number of public and private granting agencies are requiring applicants to submit a Data Management Plan. The expectation then is that fees for data access and archiving will be included in the grant budget. Open Context has a form to help data authors plan and budget appropriately for data sharing. Upon submission of the estimation form, users receive an email with a budget estimate and language to add to their Data Management Plan. This language includes a description of the access, interoperability, and archiving issues. This form is available here: <http://opencontext.org/about/estimate>. Additional guidance on developing a Data Management Plan can be found here: <https://dmp.cdlib.org/> (the DMPTool, developed by the California Digital Library and other institutions).

Last updated: April 30, 2012