Preparing your Dataset for Publication

Just like print publication, some preparation is necessary before publishing digital data. Most of this work involves documentation so that data can be understood by Open Context's editors and future users of the content. This section describes how to prepare your content to facilitate its publication in Open Context. To ensure speedy publication of your data, please review carefully the following sections and provide all required information.

Required Information

Please provide the editors with the following information, which forms the basis of your project's descriptive information (metadata). We 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. Feel free to send a small photo of yourself, which will be included in your profile.
3. **Site Name**: Provide the site name(s)
4. **Location**: The exact geographic location of the site(s) (if site security is a problem, you may provide a less precise location. Please contact the editors if you would like to discuss options.) If a given site has a record in an appropriate gazetteer (such as GeoNames, or Pleiades, or even the Wikipedia), please include URIs (Web identifier and address) to the gazetteer record(s).
5. **Period**: A cultural period AND calendar date range that applies to different parts of your dataset (in cases where the data document different discernible phases) or the whole dataset (in cases where only one identified phase is documented). [Please note, if you are using C14 dates, you must include information about calibration, labs, etc.]
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. **Support:** Include a statement of support (to acknowledge grant funding, institutional support, etc.)

12. **Related Publication(s)**: If applicable, provide bibliographic references to related publications and published datasets. Provide links to online publications.
13. **Table Field Descriptions**: Short descriptions of all your column headings (including units of measurement and notes about methodology or recording protocols that would impact reuse).
14. **Current Disposition of the Physical Collection:** A statement about the current location of any physical collections included in your data publication.
15. **License Choice**: Indicate your selection of a license for your work: (1) Creative Commons Zero (CC-0) or (2) Attribution (CC-BY). Alternatively, you may indicate that you would like to negotiate a special license with the editor (such as using different licenses for certain items).
16. **Banner Image:** Your project page will have a unique banner. If you have an image you would like to use (and permission to use it), please send it to us. Please indicate any image credits. If you do not send a banner, we will work with you to create one or we will use the default Open Context banner.

Additional Information

Clean-up and Edits: Because datasets are often fairly "raw," one should not expect perfect spelling, grammar, or compositional excellence in daily logs, database comment fields, etc. Spelling problems in these fields will probably have little impact on the overall usability of contributed data. However, some errors have greater impact. For instance, nominal values (terms used over and over again), such as terms 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. Numeric fields should contain only numbers. If a qualifier is necessary (such as a "?"), please put that in a separate field. ***Hint:*** For an extremely helpful and easy-to-use tool to rapidly clean data, check out OpenRefine.

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.

Images and Media: Images and other media are important components of archaeological documentation. Each individual media file must be clearly and unambiguously linked to one or more specific records in the dataset (such as records of excavation contexts, people, excavation log records, artifact records, etc). The data contributor should prepare a separate table listing each image file name, an image description (if desired) and the number / identifier of the object or place the image describes.

Data Formats and Structures: Data for import should be in tabular format (such as Excel, CSV, OpenOffice, etc.). (Note: Please be aware of character-encoding issues if you're using non-Latin characters or diacriticals. If you're using CSV, please check to make sure that the table values output properly because sometimes conversion to CSV has "escape character" issues.) 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).

Provide “Keys”: Related to #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.

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.

Controlled Vocabularies (“Standards”): Whenever feasible, Open Context strives to reference authoritative controlled vocabularies curated by other expert communities. For instance, Open Context would use URI-identified concepts published by the Library of Congress for the “keyword” subjects of projects. In another example, Open Context would link a researcher’s own description of biological taxonomy with closely matching equivalent URIs in the Encyclopedia of Life. We ask contributing researchers to review and verify such annotations to outside controlled vocabularies. We also ask researchers to help identify which controlled vocabularies and ontologies may be relevant to reference. Not all researcher-defined terminologies will have clear equivalences in existing standard vocabularies. If you define your own terms in a typology, please provide a detailed definition/description.