Smithsonian Data Management Best Practices

Describing Your Project: Citation Metadata

The overall description for your project could be referred to as project metadata, citation metadata, a data record, a metadata record, or a dataset record. The information supplied in the project description should be sufficient to enable finding and properly citing your data.

An easy way to ensure you have supplied enough information in the citation record is to ask yourself if you have answered the "Who" "What" "Where" "When" and "How" of your project, and that you have included a persistent identifier. Avoid abbreviations and short hand. Remember that the description of your project may be read by someone outside your field of study or even yourself years later.

Always include:

- **Creator**/author(s) -- including complete names, institutional affiliations (including SI unit if depositing into an SI repository) and any ORCIDs
- **Title** -- a meaningful and descriptive title, prefaced with "Dataset:"[1]. Title can include a facility, or title of a larger dataset if the one you are describing is a derivative or subset of that dataset.
- **Publication Date** -- year (and if relevant, month and day) the data was made public, or if under embargo, the date the embargo expires. If data is restricted and not publicly available, use the date it was deposited.
- **Persistent Identifier**/Location -- a DOI is preferred, but a URN, Handle, EzID or ARK are acceptable. If no persistent identifier is available, a working URL/URN for the data is mandatory.

➔ **The SI Libraries can 'mint' a DOI for any dataset deposited into Smithsonian Research Online (SRO) or SIdora. contact research-online@si.edu**

Include when possible:

- **Resource type** –the general format of your data, e.g., tabular data, database, audio files, sensor data, images, etc.
- **Publisher** – usually this will be the hosting location or organization with which you have deposited your data. You can use the institution or project name, or a URL or URN for the repository.
- **Grant** – either the name of the grant, e.g., "CLIR Hidden Collections 2017" or the grant number associated with the dataset
- **Abstract**/Description – an abstract for the dataset that covers who, what, where, when, why in a narrative format.
- **Preferred citation format** – MLA, APA, Chicago, etc.
- **Related publications** – this could be a published article, or related datasets, referenced with a URL or a DOI
- **Rights** – any licenses, intellectual property rights, and/or restrictions that should be applied to the data
- **Version** – a number increased when the data changes, e.g., through addition of data or re-running an analysis or derivation process.

Example:

```
[Dataset:] Templates for Statistical Resample Methods Maximize Accuracy and Efficiency of
Colorimetric Data Collection for Monitoring Biocolonization on Stone. Perets, Ethan A.;
Charola, A. Elena; Liu, Yun; Grissom, Carol; DePriest, Paula T.; Koestler, Robert J.
2016. Repository.si.edu. DOI: https://doi.org/10.5479/data.mci.2016.0629

Abstract:
Non-parametric and semi-parametric statistical approaches were developed to maximize
```

---

[1] Use of "Dataset:" to begin a title is a convention used in SRO.

accuracy of colorimetric data for monitoring biocolonization on stone surfaces, while simultaneously optimizing efficiency of data collection in the field. These approaches were applied to colorimetric data sets collected on three Kasota limestone capstones located at the National Museum of the American Indian in Washington, DC. Data was randomly resampled without replacement (the statistical "jackknife"), producing data subsets of diminishing resample sizes. …. Factors affecting the necessary minimum sample size for achieving pre-selected confidence levels and acceptable measurement error – including the impacts of a biocide treatment and heterogeneity of surface textures – were also investigated. Comparison of results for textured capstones suggests that rougher stones require greater numbers of measurements at identical d and confidence. Corresponding author: Paula DePriest.

References

CrossRef. Required, Recommended and Optional Elements. https://support.crossref.org/hc/en-us/articles/213077846-Required-Recommended-and-Optional-Elements

DataCite Metadata Working Group. (2017). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.1. DataCite e.V. 10.5438/0014.

Ball, A. & Duke, M. (2015). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides