

7. Data Management Plan

PeriodO will generate a variety of data needing professional management and curation. These data include software, structured period assertion and instance of use data, and associated documentation and dissemination materials. To maximize the wide public availability and preservation of all of these forms of data, this project will use multiple digital repositories.

Short-term storage

For short-term storage, we will use GitHub, a commercial software repository service. Software will be developed publicly through GitHub, and thus the code will be made available on an ongoing basis as it is created. GitHub has well-developed version control and issue and bug tracking features which will greatly facilitate project collaboration. Its service is free for publicly visible projects such as PeriodO. GitHub's utility is not limited to software source code, and it can be used to facilitate version control and issue tracking of the project's period assertion and instance of use data. These data will be archived in GitHub as JSON (JavaScript Object Notation) and CSV (Comma separated values) formatted files. Open Context makes similar use of GitHub to further disseminate its data. Project documentation will be made available as HTML and PDF files via GitHub and via the project's website.

Long-term storage

While GitHub will greatly facilitate short-term dissemination and collaboration, it is not a long-term digital archive. For the long-term preservation of data, PeriodO will use digital repository services provided by multiple university-based and disciplinary digital repositories. In keeping with the archival principle of LOCKSS ("Lots of Copies Keeps Stuff Safe"), project data (time period attestations) will be accessioned into both the University of California's California Digital Library (CDL) and the Digital Repository of the University of Texas library system. Documentation files and source code will also be accessioned into these repositories.

Persistent identifiers

Persistent identifiers will be assigned to all PeriodO records using the EZID system of the California Digital Library, through existing arrangements with Open Context. A server provided by the Liberal Arts Information Technology Services at UT Austin will mediate the creation of a Document Object Identifier (DOI) via EZID. This DOI will provide a persistent identifier for long-term citation of PeriodO records, even if the repository is eventually moved from GitHub. A DOI will be assigned for the entire dataset, enabling citation in scholarly literature according to DataCite.org specifications.

File formats and public domain dedication

To facilitate data interoperability and longevity, all data will be made publicly available and archived in open, non-proprietary formats including JSON and CSV resources. The data will be released into the public domain under the Creative Commons Zero (CC-Zero) declaration, thus conforming to Science Commons and Open Knowledge Foundation licensing requirements for open data. Use of CC-Zero allows PeriodO to express copyright permissions using standard RDF expressed metadata.