9. Data management plan

Documentation and Metadata

As a companion endeavor to the Cuneiform Digital Library Initiative, cdliwiki[1] documents all aspects of the CDLI in a range of articles on history, specific inscribed artifacts, and genres, as well as discussions of processes and data acquisition. On the CDLI website itself, there are also articles discussing the museum collections holding the physical artifacts[2] as well as the terms of use of the data[3]. These tools will be used to help document the project and its outcomes. Awaiting the results of a French research group[4] working on the alignment of our texts' metadata with the CIDOC-CRM ontology, we will prepare an XML output for each search result view. It will be machine readable and easily modifiable to include CIDOC-CRM attributes. The software created through the project will be thoroughly commented, and a GitHub Jekyll website will serve as a code documentation hub directly in the same GitHub location as the code.

Textual Data

The transcriptions and some text structural information of cuneiform inscriptions of artifacts are preserved using the "Canonical ASCII Transliteration Format" (C-ATF).[5] The text itself is encoded in UTF8 and the original language transliterations are restricted to simple ASCII characters. This notation system has been in use for 15 years and because of its simplicity and high level of standardization, many research projects base their work on the CDLI or will use a derivative of C-ATF. The ATF notation created by the CDLI is the widest-used standard in the field. With this project, we will continue to offer the input and output of transcriptions in this format but we will also store and offer to view and download data in an updated format where the transliterations lines of the text will also be UTF-8, and appropriate IPA characters will be used instead of derived ASCII transcription. This use of two standards makes the data usable by more people and thus will enhance its preservation. These textual data are currently viewable online and downloadable in text format. Because the CDLI is a long-lasting initiative, there are already quality checks and versioning systems in place. These checkers will be enhanced as part of the work plan. Each time a change is saved in one of the texts, a backup copy of the previous version is saved in the database.

Licensing

New software and documentation generated by the project will be released to the public domain by using the Creative Commons license "Public Domain Dedication" (CC01.0).[6]

Storage and Backup

During the research, GitHub will be used as a versioning system for the code base of the project. The Center for Digital Humanities (CDH) at the University of California, Los Angeles gives us technical support and external backups that increase the security and recoverability of the data. We also have a mirrors of the servers at the Max Planck Institute for the History of Science, Berlin (MPIWG); and through them at the Max Planck Society's persistent storage hub in Göttingen) and at the University of Oxford. Additionally, we expect news shortly from

[1] <http://cdli.ox.ac.uk/wiki/>
[2] See, for example, the page of the British Museum <http://cdli.ucla.edu/collections/bm/bm.html>
[3] <http://cdli.ucla.edu/?q=terms-of-use>
[4] <http://triplestore.modyco.fr:8080/ModRef/>
[5] <http://oracc.museum.upenn.edu/doc/help/editinginatf/cdliatf/index.html>
[6] <https://creativecommons.org/publicdomain/zero/1.0/>

Compute Canada concerning an application for resources allocation in order to set up a Canadian web site mirror and backups. These services come at no cost when allocated.

Preservation

By renewing periodically our agreements with the CDH, the MPIWG-Berlin and the University of Oxford, we are convinced that the CDLI offers optimal storage security and web server longevity; CDLI is in fact a model of data persistence—the longest lived digital humanities project in the field of Assyriology, with its predecessor the Uruk Project at the Free University of Berlin now 26 years in existence. Since the framework update project will increase the access of the data and interface of the CDLI, its usage will increase. For any eventual risk to the preservation of the software or the data, we will put copies of our work in official repositories to maximize their preservation.

Data Sharing

The code produced by this project will be released in the public domain and we will encourage anyone to use, modify and reuse any of its components. It will be available on GitHub with its accompanying full documentation.

Responsibilities and Resources

Because the CDLI has been running for many years, our lab is equipped with the needed material to undertake the framework update. Since some of our operations will be transferred to virtual servers at the Center for Digital Humanities at UCLA, costs will be impacted negatively where we will be able to repurpose two of our physical servers and retire another one, and we will no longer be required to maintain the retired physical server, or the material update of the servers now virtual. Our mirrors and backups are hosted for free at the Oriental Institute in Oxford and at the MPIWG. Each of these services are responsible for the maintenance and backup of their own servers.