

Data Management Plan for *Lexos* and Associated “In the Margins” Documentation

Data to be generated

Type of data	When shared?	Under what conditions?
Open Source computer code associated with <i>Lexos</i> tool, interface, and server-side component development.	At end of yearly milestones and at the conclusion of the start-up project, when testing has been completed for each phase of the work.	All source code and associated documentation are and will be freely available on GitHub (http://wheatoncs.github.io/Lexos/).
Graphical User Interface (GUI) layouts, including tool-tip translations.	At end of the start-up project.	All GUI layouts will be freely available on GitHub (along with main <i>Lexos</i> source code or in a separate GitHub repository, to be determined).
User-generated data consisting of modified versions of texts uploaded to the <i>Lexos</i> website, as well statistical tables, reports, and visualizations based on these texts.	Data are for use in individual research experiments and are <i>not</i> made public by the online <i>Lexos</i> tool. Data are stored in session-keyed folders (uniquely private to the user), and <i>Lexos</i> ’ Python Flask framework contains mechanisms to ensure that access to session data is secure.	Data generated by <i>Lexos</i> from user-uploaded materials may be published only at the user’s discretion. The <i>Lexos</i> website does not provide public access to user’s uploaded texts. The project does not check the copyright status of user-submitted materials.
“In the Margins” documentation and “Talking head” (3-5 minute) videos	Available as embedded content and links from within the <i>Lexos</i> tool and from a separate Scalar website.	Content will be open access from the Scalar website and the <i>Lexos</i> tool. Videos may be licensed for use outside of these platforms at the Lexomics Group’s discretion.
Assessment data generated during the testing phase.	Aggregated data will be shared via the final report to NEH.	No information will be shared that could identify individuals participating in the assessment process.
A final (white paper) report to NEH.	At the conclusion of the project.	Dissemination of the final report will be the responsibility of NEH.

Period of data retention

The *Lexos* project does not permanently cache files uploaded to the *Lexos* website or associated data files. Users’ individual session folders (including all uploaded, pre-processed texts, and subsequent data analyses and visualizations) are regularly deleted from the *Lexos* server by server-side, time-based (“cron”) jobs. *Lexos* provides functionality for users to download their results; it does not provide access to this data beyond the life of the user’s session.

Data formats and dissemination

Computer code and relevant documentation will be available as open source in a publicly accessible code repository (GitHub). “In the Margins” content will be made available as Scalar “pages,” which may be accessed from within Scalar or using Scalar URLs. Additional PDF materials may be disseminated via the project website. “Talking Head” videos will be accessible through the *Lexos* tool, as well as the Scalar website. Data generated by *Lexos* takes the form of text files (.txt), comma- or tab-separated value data files (.csv, .tsv), and image files (which may be in .pdf, .jpg, .png, or SVG format). Data and conclusions based on our testing will be disseminated in the form conference papers and print publications. Data produced by other users will be disseminated at their discretion.

Data management and maintenance

All source code for the project will be stored in Github. An existing code base is currently housed at: <http://wheatoncs.github.io/Lexos/>. Mark LeBlanc (PI) will be responsible for managing data collection and removal on Wheaton College servers and for managing updates at Github. Should LeBlanc leave the project or server space at Wheaton is no longer available, the GitHub repository contains instructions for how other users can install the *Lexos* tools locally on their own server and auto-installation tools will be produced as part of the grant project.

“Talking Head” videos will be archived on Wheaton’s local DSpace repository for faculty and researchers. Scalar is a stable environment for hosting “In the Margins” content. The Alliance for Networking Visual Culture, which produces Scalar, has secured a long-term server environment at the University of Southern California. This includes a redundant server cluster with a dedicated high speed port, which will seamlessly “failover” to the secondary server if the primary server fails. It performs daily back-ups of all Scalar content both locally to the Disaster Recovery Center at Arizona State University in Phoenix. All Scalar content is exportable, and Scalar is open source software, which may be cloned and installed to another server, in the event that USC should discontinue its free hosting service.