

## **Data Management Plan**

**Types of Data.** The project manages and produces the following types of data:

- a. Unstructured text data for “One Book” literary works,
- b. Circulation and holdings records from the CPL ILS system,
- c. Chicago-area demographic data extracted from census records,
- d. Software for text analysis, predictive modeling, and interactive tools,
- e. Features, models, visualizations, and other outputs from analysis and modeling activities,
- f. Social media data (from Twitter, Facebook, and other sites) as accessed through the public APIs of these services,
- g. Textual project updates, meeting notes, and final report(s),
- h. Physical materials from Chicago Public Library (loose papers, CDs, still picture files) and reproductions of the same.

**Data Sharing/Retention/Dissemination.** All applicable electronic materials produced will be made available to the public through DePaul University’s open access repository system, Via Sapiientiae. (Via Sapiientiae is a secure, permanent, read-only repository freely available to all at <http://via.library.depaul.edu>.) Appropriate metadata for project data will be created according to international cataloging standards and loaded into shared catalogs such as OCLC’s WorldCat (a shared catalog containing the holdings of over 10,000 libraries), in order to facilitate discovery of the data sets by researchers worldwide. Data will be made available only after appropriate steps have been taken to protect intellectual property. Confidential material will be handled according to policies and protocols for human subjects, FERPA, and any other applicable regulations and restrictions.

The source code for our analyses, models and tools will be made available as open-source software and distributed using GitHub, thus making it available to the broader research community.

Our use of in-copyright works is confined to non-consumptive processing of the full text of each work within the secure HathiTrust capsule. The project does not have to manage these copyright-protected works. Descriptive features extracted from each text within the computing capsule can be released to our project team after they are vetted by HathiTrust to ensure that they meet fair-use conditions.

When necessary, the project team intends to secure permission (free or paid) to reproduce OBOC-associated texts or images for use in published articles, lectures, presentations at professional conferences, and classroom instruction. However, depending on the content, its owner, and the agreement governing each work, the rights to use content may not extend to other researchers or to follow-on or derivative research by others. Materials will be disseminated only to the extent that content owners will allow, and reproduction or reuse may be restricted.

**Data Formats.** It is anticipated that data will be in text and numerical form (.txt or other common machine-readable formats). Specific file formats and encoding standards are continually determined in relation to the selection of tools and ongoing needs of DePaul and Chicago Public Library.