

World-Historical Gazetteer Data Management Plan

This Level I Start-Up proposal is primarily intended for working-group discussion rather than for data development. However, we will also build a testbed environment based on the 1.4 million georeferenced entities in English Wikipedia, potentially supplemented by some thousands of entities from other sources. The testbed is the data that we are describing in this plan.

The gazetteer testbed will be integrated with the existing Archive of the NSF-funded Collaborative for Historical Information and Analysis, which already exists in preliminary form. All decisions on gazetteer content and metadata management will be reviewed during the 2014 workshop, and the resulting plan will be implemented thereafter. As a result, this Data Management Plan includes elements of both the current set of actual data management, and the issues to be decided for the gazetteer to be created through this project.

Existing Data

Three categories of existing data stand out at present. 1) Data of the World-Historical Dataverse archive (www.dataverse.pitt.edu), housed at the Dataverse Network at Harvard University. 2) Data now incorporated into the CHIA Archive through the Col*Fusion procedure. These include the “Tycho” dataset on disease surveillance in the U.S., the Colonial Legacies dataset developed by John Gerring of Boston University, and the “20th-century global prototype” dataset developed by Patrick Manning and Chelsea Mafra. 3) Other existing data for contemporary times, notably WikiData and GeoNames.

Gazetteer Data and Gazetteer Development

The objective of this project is creation of a CHIA gazetteer system, which will be permanent and may be either centralized or distributed. The gazetteer data are expected to be permanent, as they are necessary for documentation of the content in the historical archive. Periodic updates of gazetteer data will result as knowledge advances.

The NEH Digital Start-up project itself, in contrast, is a testbed system. This is the project-level work of developing standards for the permanent system. It may evolve into something different as an element of a persistent, long-term, stable CHIA architecture.

Archiving, Storage and Backup

All of the data stored in the CHIA archive will benefit from a program of Digital Stewardship to ensure preservation, documentation, careful identification of sources, and easy access to the materials for public users. Two repositories will hold the data for the gazetteer. First is a server located in and administered by the School of Information Science at the University of Pittsburgh, which is dedicated to the Collaborative for Historical Information and Analysis. Second, to the degree that the volume of data and the intensity of calculations require, data will be stored within the facilities of the Pittsburgh Supercomputer Center, with which CHIA has a contract through XSEDE. Backup procedures in each of these institutions are according to the standard procedures for each.

Responsibility

Dr. Kai Cao, postdoctoral associate at the World History Center, has responsibility for administering the gazetteer data stored and utilized.

Security

For all computers within the firewall of the University of Pittsburgh, the office of Computer Services and Systems Development (CSSD) takes responsibility for maintaining security and monitoring all campus internet installations.

Audience and Access

Users of CHIA and therefore of the gazetteer documenting its data are expected to include a wide community of researchers, teachers, and students. It will be accessible through the CHIA website (www.chia.pitt.edu).

Ethics

The issues of ethics in handling of the World-Historical Gazetteer data is on the agenda for the August 2014 workshop. The decisions adopted at that time will be implemented.

Intellectual Property

The five universities collaborating in the NSF-supported award to CHIA have signed a Memorandum of Understanding governing issues of intellectual property created through work of the collaborative. This MOA (completed in April 2013) can be extended to the participants in the World-Historical Gazetteer.