Collaboration in Building a Global Archive:
Vision, practice, methodology, and analysis

We welcome readers to the initial issue of JWHI. The journal is part of a broader mission to promote the collection, publication, review, and analysis of world-historical data. It focuses on collaborative discussion across numerous disciplines, in order to bring together the best scholarly minds in the creation of Big Data in history. Its audience is to include scholars and practitioners in every field of history and social sciences, in information science, in the humanities and natural sciences.

The collection and analysis of Big Data—corpora too large and complex to process using traditional database management techniques—has transformed the study of topics ranging from astronomy to genomics. Even scholars of literature are mining vast digital datasets. The promise of Big Data is its capacity to reveal patterns that would not be visible in smaller datasets. For world history, Big Data opens the door to investigating topics like the impact of war, trade, colonialism or climate on social development at a global scale.

This approach reflects a new stage in the collection and analysis of data on human society. In the nineteenth and twentieth centuries, national states sponsored massive efforts to collect data and to aggregate it from local levels to national levels to serve their expanding needs. They produced national censuses, national data on shipping, and ultimately national income accounts and national social surveys. What of the world as a whole? Of course it existed throughout the period of nation-building, as it had before in the eras of ancient empires, expansion of great religious traditions, and in the stages of expanding oceanic interconnection. But societies engaged with one another only at certain moments, such as the shock of two world wars in the first half of the twentieth century and the recognition of global patterns in environment, culture, and economy at the end of the twentieth century. Even as discussion of globalization became widespread, analysts generally relied on comparative national data rather than exerting the effort to create global data. The United Nations and UNESCO, founded to facilitate international collaboration and a global perspective, saw their membership rise from just over 50 to nearly 200, thus contributing both to nationhood and to global outlooks. Many discussions of the global past have transpired without much access to data at all, and have remained qualitative and speculative.

At last it is feasible to develop systematic data on the human social system at all geographic scales and at times extending to the past several centuries and perhaps more. Both the human skills and technical infrastructure

now exist for pulling together Big Data in history. The task is to assemble and organize the expertise, the technology, and the data.

Preparatory work has been going on in several venues. Certain of these efforts have been able to combine and bring about multi-institutional and transnational collaboration. In the case of those who came together to create this journal, groups from seven different institutions in Europe and North America wrote collaborative proposals in 2010 and 2011, and then in 2012 succeeded in gaining funding. (The institutions are Boston University, Harvard University, the International Institute of Social History, Michigan State University, University of California – Merced, University of Pittsburgh, and University of Portsmouth). Other collaborative groups in the social sciences include CLIO-INFRA (economic historians based the University of Utrecht and the International Institute of Social History), the Interuniversity Consortium for Political and Social Research (founded in 1962 at the University of Michigan), Minnesota Population Center (founded at the University of Minnesota in 2000), the Electronic Cultural Atlas Initiative (ECAI, founded at the University of California – Berkeley in 1997), Dyn-Coop Net (collaboration of Idaho State University and University of Porto well established by 2008), and several additional groups focusing on GIS analysis.

JWHI was first proposed at the University of Pittsburgh in 2011. The University Library System at Pitt provided rapid and effective support, and the proposal for the journal was approved in November 2011. The journal is to be an arena of exchange for scholars of all fields. It is to be cross-institutional, cross-disciplinary, and is to publish work written and read by scholars in information science, history, geography, GIS, social sciences, health, and natural sciences. The journal publishes six types of article: editorials such as this, reviews of published datasets, and four categories of articles: "vision" articles on the vision of large-scale historical datasets, "practice" articles on the practice of historical datasets published recently or immediately forthcoming, "methodology" essays on theory and analytical technique, and "analysis" articles developing interpretive results out of datasets.

In reviewing datasets, we open up an area of publication of which we are particularly proud. The reviews in this issue represent a start of that tradition, and JWHI will publish reviews in every issue. The purpose is dual—to develop critical standards for datasets and to provide formal academic recognition to the authors and compilers of datasets as scholars who have assembled and documented historical data.

Articles on **vision** are to articulate the direction of historical studies relying on Big Data, identifying major challenges and proposing paths for overcoming obstacles in the development of historical data. Articles on **practice** are to showcase recently published datasets (or datasets nearing publication), demonstrating the techniques of data collection, data definition, archiving, analysis, and visualization. **Methodological** articles are expected to range widely in addressing challenges concerning the creation, integration, archiving, analysis, and visualization of Big Data in world history. **Analytical** articles are to draw on existing data to locate advances in scientific knowledge that can come from expanding the temporal and spatial scope of historical studies. Current trends in large-scale historical analysis focus on quantitative data, but it is our expectation that qualitative data, notably texts and images, will also grow steadily as an aspect of Big Data in history, and that historical studies will become an important arena for linking quantitative and qualitative analysis of human society through electronic archiving and analysis.

It is our hope and belief that a focus on the practical work of creating and analyzing Big Data in history will nurture fruitful debate and collaboration across disciplinary lines. We also believe that the review and critique of datasets—of all sizes—will develop higher standards of analysis and a clearer language for description and assessment of datasets. The successful construction of Big Data in history will require that scholars of all sorts share the academic workplace and that they work collaboratively and in mutual understanding to fit the products of their labors together into a smoothly functioning electronic edifice. We wish this journal to be another electronic space where the problems and results of that work undergo discussion by a broad and committed community.

The Editors

This journal is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press.