

The World-Historical Dataverse: Design and problem-solving for a large-scale, heterogeneous, historical dataset

Patrick Manning, Director, World History Center

April 22, 2010

3:30 - 5:00 p.m.

IS501 (IS Building on Bellefield)

ABSTRACT

The World-Historical Dataverse project (“the universe of world-historical data” – www.dataverse.pitt.edu) is intended to lay the groundwork for creating consistent historical data for localities worldwide so that they may be aggregated into global totals. Data are to address economic, social, health, and environmental issues for about the last four centuries. The existence of such a dataset ought to get beyond current historical knowledge, at local and regional levels, to reveal historical patterns and dynamics at the global level; such information would help policy-makers set plans for the global future that account for past patterns rather than ignoring or speculating about the global past. The two co-directors – Patrick Manning of the World History Center and by Siddharth Chandra of Michigan State University – are both economic historians by training and have worked on this project since 2007.

The opening section of the presentation provides an overview of the main projected steps in the project, along with problems likely to arise at each step. Design and creation of a universal dataset must proceed along iterative steps, both to address the many problems entailed in such a task and to gain the confidence of funders. Project directors have so far identified six “departments” for carrying out the succeeding steps of project work: a portal with links to a large number of online data portals and datasets; a blog to discuss issues in creating large-scale historical datasets; a campaign to collect relevant datasets from scholars, agencies, and publications; an archive to store and display original and transformed versions of historical datasets; federation of archived datasets to create larger and more extensive datasets; and design and ultimate construction of a universal dataset.

The second and more detailed portion of the talk centers on the interplay between building the archive and federating archived datasets. The federation of datasets through an online interface will be the project’s first substantial step, as it goes beyond mere collection of data and into creation of enlarged datasets. Preparation of datasets for federation will require preparation of extensive metadata; full specification of spatial, temporal, and topical coordinates; and conversion of weights and measures into consistent dimensions. The project has just received funding from the Pitt School of Arts & Sciences for a three-year Postdoctoral Associate for this work. The appointee is to have advanced IT skills in design and programming, to focus on the work of federating archived datasets. Once a candidate is selected, work will begin in September 2010.