

# WORLD-HISTORICAL DATAVERSE

## Previous and Competing Data Resources

Xi Zhang



### PURPOSE

Review previous and competing data resources to understand their objectives, achievements, technical contributions, collaborations, boundaries and limits, processes for update.

### PROCEDURE

Critical review, for each relevant program, of its structure, history, collection system, archiving, reporting and visualization.

### LEADING DATA RESOURCES

Teradata<sup>1</sup>

Human Relations Area Files (HRAF)<sup>2</sup>

China Historical Geographic Information System (CHGIS)<sup>3</sup>

Inter-University Consortium for Political and Social Research (ICPSR)<sup>4</sup>

Integrated Public Use Microdata Series (IPUMS) USA and International<sup>5</sup>

United Nations Development Program (UNDP) Human Development Report<sup>6</sup>

### INTERIM RESULTS OF THIS STUDY

For the existing databases that have been reviewed, their data structures were established in isolated disciplines (e.g., HRAF), within national or ethnic limits (e.g., CHGIS), set up in the technology of several decades ago (e.g., HRAF), or without significant emphasis on long-term or historical analysis (e.g., UNDP).

### TERADATA



- Teradata is a software company, founded in 1979, that develops and sells a relational database management system (RDBMS) with the same name. Its products are commonly used by companies to manage data warehouses for analytics and business intelligence purposes. Teradata was formerly part of NCR Corporation, with the spinoff from NCR in 2007.
- Teradata acts as a single data store that can accept large numbers of concurrent requests from multiple client applications and it has several significant features: unconditional parallelism, complex ad hoc queries, parallel efficiency, and scalability.
- Teradata sets up the Teradata University Network (TUN), which serves as a web portal and educational resource for database, data warehousing, business intelligence and decision support systems. The goal of TUN is to build a collaborative IT education community at both undergraduate and graduate levels so that instructors and students can use it free of charge.
- ❖ *Teradata has useful features that could shed light on building databases for academic purpose, but its own data are mostly industrial oriented.*

### HUMAN RELATIONS AREA FILES (HRAF)

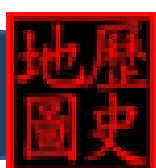


- Founded in 1949 at Yale University as a financially autonomous research agency, HRAF aims to encourage and facilitate comparative studies of human behavior, society, and culture.
- The data include thousands of pages of text from books, articles, and unpublished manuscripts as well as English translations of foreign texts, covering contemporary and prehistoric cultures all over the world and fostering both research on a particular culture and cross-cultural comparative research on humans in all their variety.
- HRAF provides two full-text electronic databases: eHRAF World Cultures (Ethnography) and eHRAF Archaeology, covering worldwide cultures and major archaeological traditions respectively.
- The full texts are indexed with a classification system called the Outline of Cultural Materials (OCM) to facilitate across cultures searches.
- ❖ *The HRAF dataset and its OCM coding system were established in the 1940s. With the continuously updated and expanded collections, the OCM can no longer adequately serve to index and classify all the relevant information for current global studies. Moreover, HRAF was conceptualized in specific and isolated disciplines (i.e., anthropology, archaeology) which provides no basis for interdisciplinary analysis.*

### REFERENCES

- <sup>1</sup> <http://www.teradata.com/>.
- <sup>2</sup> <http://www.yale.edu/hraf/>.
- <sup>3</sup> <http://www.fas.harvard.edu/~chgis/>
- <sup>4</sup> <http://www.icpsr.umich.edu/ICPSR/access/subject.html>
- <sup>5</sup> <http://usa.ipums.org/usa/> (USA site)  
<https://international.ipums.org/international/> (International site)
- <sup>6</sup> <http://hdr.undp.org/en/>

### CHINA HISTORICAL GEOGRAPHIC INFORMATION SYSTEM (CHGIS)



- CHGIS project was launched in 2001 to establish a database of populated places and historical administrative units for the period of Chinese history between 221 BCE and 1911 CE. CHGIS provides a base GIS platform for researchers to use in spatial analysis, temporal statistical modeling, and representation of selected historical units as digital maps.
- The main task is to create unique records for all of the administrative units down to the county (*xian*) level from the time of unification (221 BCE) to the end of the dynastic period (1911 CE), and to provide documentation of the sources used to create each record. The purpose is to create a basic database to contain all the administrative units which can be queried and linked to digital geographic objects.
- Sources of data include contributions from Fudan University, Harvard University, University of California, Davis, and Russian Academy of Science.
- ❖ *The data are modified according to historical name change to maintain consistency. Not all provinces of China are included in the current version of GIS data. New versions covering more provinces need to be presented.*

### INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR)



- Based at the University of Michigan, ICPSR began collecting and distributing social science datasets during the 1960s.
- Most of the datasets in ICPSR archive are raw numerical data from surveys, censuses, and administrative records. They were originally gathered in research projects and for administrative purposes. ICPSR maintains a data archive of over 500,000 files of social science research and hosts 16 specialized data collections in education, aging, criminal justice, substance abuse, terrorism, and other fields.
- Most of ICPSR data holdings are open to public. There are two types of restricted data: data that are available to vetted researchers on removable media and data that are accessible only within a secure data enclave.
- ❖ *ICPSR data structures were conceptualized several decades ago and the data are not transformed and displayed in a systematic way, which makes it hard for researchers to conduct global studies.*

### Integrated Public Use Microdata Series (IPUMS) USA and International



- The IPUMS project is a collaboration of the Minnesota Population Center, national statistical offices, and international data archives, dedicated to collecting and distributing census data from around the world.
- The population data are coded and documented consistently across countries and over time to facilitate comparative research.
- The data series include information on fertility, nuptiality, life-course transitions, migration, labor-force participation, occupational structure, education, ethnicity, and household composition. Use of the IPUMS data is restricted to scholarly and educational purposes.
- The IPUMS is microdata, making it possible for researchers to create tabulations tailored to their particular questions.
- *Microdata are subject to strict confidentiality measures that limit their usefulness for some applications. IPUMS is inappropriate for research that requires the identification of specific small geographic areas in those census years. In addition, IPUMS data are limited to territories and times with published censuses.*

### UNITED NATIONS DEVELOPMENT PROGRAM (UNDP)



- In 1990, UNDP started annual reports for human development around the world based on various official sources of social and economic data, aiming to analyze country experience to distill practical insights and make relevant experience available to all policymakers.
- The website contains the data on general economic and social indicators for most countries from 1990 to present, all open to public.
- Raw data are analyzed and constructed into indices to report each country's advantage and disadvantage for human development, to compare across countries and years, and to suggest policy implications. Users can also use the website to build their own development indices based on different research interests. The website provides Stat Planet world map and thematic maps which are linked to reports on specific countries, years, and indicators.
- ❖ *UNDP data only covers the years since 1989. Historical data with a systematic structure are needed to facilitate meaningful world historical studies. Moreover, UNDP data are reported only at the national/aggregate level, making cross-national comparisons difficult.*