A Case Statement to Support a Social Sciences Collection
California Digital Library

Draft 3/11/98

Written by Sherry DeDecker, Lucia Snowhill and Mary Larsgaard, UC Santa Barbara
Edited by Cecily Johns

There are three major areas of emphasis in this case statement supporting social sciences as the next digital collection: numerical data, government documents and geospatial data.

USER BASE AND ACADEMIC PROGRAMS:

The user base for a social sciences digital collection includes academic departments such as economics, business, political science, sociology, education, social psychology, public policy, environmental studies, urban planning, geography, geology, anthropology, archaeology, and history.

CONTENT AVAILABLE:

To support social sciences research, the most substantial data available is statistical and numerical data, primarily from governmental agencies at all levels, including international governmental organizations (World Bank, European Communities), the federal government, and California. In addition to statistical information, government agencies also produce research reports, patent information, legislative, regulatory and policy information applicable to a broad spectrum of academic research. A UC/Stanford committee, Electronic Government Information Initiatives Group (EGIIG) has written a white paper about a number of possible projects to begin to integrate government information into CDL collections (see http://gort.ucsd.edu/docs/final.html).

A great deal of electronic information is already available through library websites (see GPO Gate at http://www.gpo.ucop.edu), however the data is scattered and often difficult to locate. A social sciences digital collection would provide a user-friendly front end to the data, and work with government agencies and other groups to archive the data in greatest demand, e.g. census data.

Other organizations, both commercial and academic, provide social sciences data for research. A major source of numerical data is available through the ICPSR - Inter-University Consortium for Political and Social Research. Many UC data archive centers and academic computing facilities hold memberships in this consortium, which provides data sets of various research studies and selected governmental sources. CDL could collaborate with these centers to provide a single user-friendly interface to ICPSR data and that of other similar organizations. There are also a number of important commercial resources, such as DRI, Datastream and EIU Online which provide high level digital information to the corporate world. Two other databases of stock market and financial data, CRSP and Compustat, are important to finance and accounting faculty.
The Alexandria Project, funded by NSF and based at UC Santa Barbara, is in the process of building a library of geospatial data for California that should become an integral part of a social sciences digital collection. Alexandria is also designed to be a gateway to geospatial data at other sites. Another source of geospatial content would be the metadata of faculty-generated GIS data available on UC campuses.

In addition, archival and other special collections, such as University Archives, present opportunities for creating digital content for the CDL. Most of these collections are now in paper form, but the CDL might take the lead to seek grant funds to digitize these collections. Oral histories, photographs, manuscripts and archival collections support research in history, public policy and environmental studies, among other disciplines.

**INTELLECTUAL RESOURCES NEEDED AND AVAILABLE:**

Suggested categories of information that might be included in a social sciences digital collection to support research and teaching at UC includes:

- economic and demographic data
- legislative and policy information
- regulatory law
- historical statistical data, e.g., U.S. Census
- social research
- scientific and technical reports
- patent information
- maps
- digitized air photos
- satellite images
- ethnic/racial group information
- public opinion data

**TECHNICAL AND DEVELOPMENT ISSUES (E.G. PRESERVATION AND STANDARDS):**

Computer hardware and software capable of dealing with large datasets, including servers with sufficient RAM and ROM, speedy processors, and scanners appropriate to the format, e.g., photogrammetric scanners for air photos and large format scanners for maps, will be needed to implement a social sciences digital collection of numerical data and geospatial data.

Expertise of UC faculty, librarians, geospatial specialists and technical staff will be needed to develop a social sciences collection and determine the system resources needed. For many of the social sciences digital databases now available sophisticated programming skills, including C, Fortran and SAS, are needed to access and utilize them.

Archiving social sciences and particularly geospatial data will require high end computer hardware, software, and large amounts of disk space. Computer specialists will be needed to maintain and refresh the data. Both federal and state agencies are looking for partners in academia to help archive their data for preservation and for future access. The California State
Library, for example, has not been able to provide an electronic archive of California state publications.

Archiving and providing improved access to social sciences data could be achieve by partnering with the following organizations:

- OCLC for storage and archiving data
- Census to become Research Data Center: [http://www.census.gov/ces/rdc.announce.html](http://www.census.gov/ces/rdc.announce.html), allowing access to economic and demographic microdata
- NTIS [http://ccsftp.ucdavis.edu/webpgs/ntis.html](http://ccsftp.ucdavis.edu/webpgs/ntis.html) to deliver full-text federal technical report literature through the Internet
- California State Assembly and Senate legislation, and specific California agencies such as the California Resources Agency
- Environmental Protection Agency and Air Resources Board statistics on air pollution and emissions
- EPA technical reports
- U. S. Geological Survey and California Department of Water Resources data on water quality

**POLITICAL APPEAL:**

A social sciences digital collection would include a considerable amount of statistical and geospatial data that corporations, schools and research organizations, as well as members of the general public, would find useful. Some of the organizations, that would benefit from ready access to a social sciences digital collection include State and local government agencies, the California legislature, law firms, environmental consultants, public school systems, and large and small businesses that use economic and demographic data for market analysis, long range planning and a variety of other purposes.

**OTHER CELLS IN THE COLLECTION MATRIX:**

Other types of information that should be considered/explored:

- digitizing newspaper projects (California Newspaper Project)
- digitizing UC academic department working papers
- subscriptions to electronic journals supporting social science research
- professional association proceedings and journals
- online databases that index journals and government documents, such as PAIS, ERIC, Disclosure and ABIInform
- atlases already in digital form
- encyclopedias of cartography
- gazetteers, such as the one available at the Alexandria site