From Primitive Numbers to Knowledge: how technology has enhanced the dissemination of social science data

Chiu-chuang (Lu) Chou
University of Wisconsin-Madison
Institutional Organization
Meet the DACCers
SSML

- Social Science Microcomputing Lab, http://www.dacc.wisc.edu/ssml/index.html
- An Instructional Lab with 42 PCs
- One 50% lab manager
- One 50% statistical consultant
- Annual budget of $52,081 (hardware, software, student hourly, facilities and supplies)
DPLS

- Staff
- Operation Budget
- Technical setup
- Collections
- Services
- Outreach
Staff

- Total of 2.2 full time employees without students
  - Cindy Severt, Director (full time)
  - Chiu-chuang (Lu) Chou, Senior Special Librarian
  - Joanne Juhnke, Special Librarian
Operation Budget in 2006

- $26,976 annual budget
- $13,110 for ICPSR membership
- $3,000 for Roper Center
- $1,000 for Datastream (Joint subscription with Business School Library)
- $1,000 for IFS
- $8,866 for purchasing other datasets and expenses
Technical Setup

- Web server
- FTP server
- 5 Staff PCs, 1 Scanner and two Printers
- Public-use PCs
  - Preparatory CD-Roms (Census CDs, CPS utilities, etc.) and CD-Roms with user’s interface, e.g. NCES studies, CBP and etc.)
  - Online subscriptions (Datastream, ICPSR Direct, Roper Express, IFS, WDI and Source OECD)
Collections

- DPLS Catalog (3,304 studies)
- DPLS Archive (64 studies)
- BADGIR (20 studies)
- Internet Crossroads in the Social Sciences (839 Internet sites)
- ICPSR
- Roper Center
Services

- Data Search
- Data Acquisition
- Online Data Provision
- Technical and Programming Assistance
- Archiving Data Collections for Public Distribution
- User Instruction
- Statistical Referrals
Outreach

- Quarterly Newsletter
- Data Literacy for Health Science librarians on campus
- IASSIST
  - IASSIST Quarterly readers
  - Hosting 2004 IASSIST conference
  - Representing IASSIST at a Statistical Disclosure Control workshop, 2004
- Special Purpose Libraries Group
  - “How Many Animals?” Story Time for kindergarteners and 1st graders
What has happened?

- Human Research Protections and Institutional Review Board
- DDI and Metadata standards
- Trend of mainstreaming social science data
- New web technology and tremendous information growth on the Internet
- Data and data extracting tools available from various web sites
- Higher user expectation
Human Research Protections

- Obtained IRB exempt status for public-use datasets available from DPLS, ICPSR and Roper
- Work with UW Purchasing services for license agreement (e.g. KLD STATS)
- Certain datasets requires users to sign an agreement first (e.g. Add Health)
- Refer users to Sensitive Data Facility at the Social Science Computing Cooperative for secured storage if they have restricted/sensitive data
  - Cold Room
  - Frozen Tundra
UW-Madison Social and Behavioral Sciences IRB Exempt datasets/repositories

- Data and Program Library Service (DPLS)
- Center for Demography and Ecology (CDE)
- Better Access to Data for Global Interdisciplinary Research (BADGIR)
- Inter-University Consortium for Political and Social Research (ICPSR)
- Roper Center for Public Opinion Research
- National Election Studies
- National Center for Health Statistics
- National Center for Education Statistics
- U.S. Bureau of the Census
DDI and Metadata

- 2001: Survey Documentation Analysis (SDA) a trial project with two studies, [http://sda.berkeley.edu:7502/data/wisconsin/](http://sda.berkeley.edu:7502/data/wisconsin/)
- 2003: Participated in UW system Digital Repository project (2 small datasets were evaluated in DSpace)
DDI and Metadata continued

- 2004 to present: Better Access to Data for Global Interdisciplinary Research (BADGIR),
  [http://nesstar.ssc.wisc.edu/index.html](http://nesstar.ssc.wisc.edu/index.html)
  - DDI complied
  - Nesstar Suite
  - 20 DPLS archival studies
  - National Survey of Families and Households (NSFH)
  - Survey of Health, Well-being and Aging in Latin America (SABE)
Mainstream Social Science Data

Datasets/Repositories available to the whole campus via broadband

- ICPSR Direct (11/2001)
- Proxy Server for ICPSR Direct (11/2002)
- SourceOECD Online (9/2001)
- IFS Online (11/2003)
- WDI Online (4/2003)
- Roper Express (11/2005)
Usage Statistics

ICPSP Direct Files Download

<table>
<thead>
<tr>
<th>Year</th>
<th>Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y2001</td>
<td>1000</td>
</tr>
<tr>
<td>Y2002</td>
<td>6000</td>
</tr>
<tr>
<td>Y2003</td>
<td>7500</td>
</tr>
<tr>
<td>Y2004</td>
<td>9000</td>
</tr>
<tr>
<td>Y2005</td>
<td>23000</td>
</tr>
<tr>
<td>Y2006</td>
<td>3000</td>
</tr>
</tbody>
</table>
Usage Statistics

- Before Roper Express
  - 50 studies per year

- Roper Express
  - 2005 4th quarter
    - 45 data files and 58 document files
  - 2006 1st quarter
    - 165 data files and 160 document files
Web Technology

- 2002 Used Active Server Page (ASP) technology to build dynamic applications
  - Catalog: http://dpls.dacc.wisc.edu/newcatalog/index.asp
  - Crossroads: http://dpls.dacc.wisc.edu/newcrossroads/index.asp
  - Archive downloads: http://dpls.dacc.wisc.edu/archive/login.asp
Data Web Sites

- How are our collection development and technical service affected?
  - ICPSR and Roper studies are not longer added to our holdings database
  - Instead of maintaining user interface programs on various CD-Roms, we direct users to interface programs on the web

- How effectively are our users utilizing these sites?
  - Are the contents on a site well-organized?
  - Is a site easy to navigate?
  - Is its data extracting tool intuitive?
  - Are users aware of the limits of data posted on a site?
DAS for NCES Studies

- Respondents’ confidentiality is protected
- The new DAS is not a table aggregator any more
- Users can manipulate the data without direct access to the data
- It can modify up to several thousand variables on the fly
- Users can request new variables be added to their analysis
What have we done?

- Provide accelerating access to data sources (ICPSR Direct, Roper Express, IFS, SourceOECD, WDI and etc.)
- Constant contents update on our web site
- Collaborate with other campus libraries in resource sharing and referrals
- Network with research units
New Projects

- Promote data literacy
- Improve data delivery
  - Add a direct data download link from our online holdings catalog with php
  - Store current CD-Roms as ISO images on a network drive and allow users to access them without loading/unloading CD-Roms in the drive
- Collaborate with GIS unit on campus
  - Link social Science data with GIS data
More New Projects

- Expand our archival holdings to include qualitative studies
- Develop teaching modules using studies in BADGIR
- Advocate metadata standards to local data producers
- Participate in the UW system-wide digital repository
Our Focus

- Facilitating teaching, learning and research in social sciences
  - Identify, locate and obtain high-quality datasets
  - Organize, preserve and distribute these datasets
  - Collaborate with PIs, data vendors, IRB, tech supports, statistical consultants, and users in providing the best support in secondary research
We measure our success in helping one user at a time.