



[Additional Resources](#)

[Editing Stats Programs](#)

[Using StatTransfer](#)

Introduction to Data Librarianship

Workshop presented by
[Paul H. Bern](#)
at the
[2006 IASSIST Conference](#)
in Ann Arbor, MI

Acquisition

Finding data

1. One of the best ways to find data is to look in other archives and databases. Other institutions may have more resources (i.e. people) as well as a longer history in finding data, and, therefore, have created archives and/or databases of resources.
 1. [ICPSR](#), of course!
 2. [Data on the Net](#) Links to sites of numeric social science statistical data, data catalogs, data libraries, social science gateways, etc. Provided by University of California, San Diego.
 3. [Research Resources for the Social Sciences](#) Organized into 18 categories including general resources, reference materials, data archives, and various social sciences disciplines.
 4. [SOSIG](#) SOSIG (Social Science Information Gateway). Extensive links to social sciences resources worldwide.
 5. [CESSDA](#) Home page of CESSDA (Council of European Social Science Data Archives). Includes access to catalogs of member organizations and a clickable map of Social Science Data Archives all over the world.
 6. [Social Science Data Archives](#) From the Australian National University.
 7. [Madera](#) (Multilingual Access to Data Infrastructures of the European Research Area) portal provides access to an unprecedented quantity of social sciences quantitative datasets using an easy to use Web interface. It harvests statistical datasets and variables published on the Semantic Web from all the largest European social sciences data archives.
 8. [Resources for Economists on the Internet](#) has lists of web sites that provide data on many topics.

9. [Economagic: Economic Time Series Page](#) A comprehensive site of free, easily available economic time series data useful for economic research, in particular economic forecasting. The majority of the data is USA data.
 10. [Qualidata](#) Site for information on this Archival Resource Centre of ESRC (Economic & Social Research Council) qualitative data.
2. Various "update newsletters"
 1. [The Scout Report](#) The Scout Report is the flagship publication of the Internet Scout Project. Published every Friday both on the web and by email, it provides a fast, convenient way to stay informed of valuable resources on the Internet.
 2. [Current DEmographic Research Reports \(CDERR\)](#) is a weekly email report that helps researchers keep up to date with the latest developments in the field.
 3. Many commercial products will send you updates even if you have not purchased any of their products.
 3. Trade and Industry associations often collect (or contract to collect) data about their constituents. Often, you need to be a member to get this data, but it never hurts to have the patron contact them and simply ask.
 4. Statistics databases are good places to find other compiled data on various topics. Of course, your institution must subscribe to these (and they're NOT cheap!).
 1. Lexis/Nexis Statistical - focus is on US Federal, State and Local data
 2. TableBase - From the RDS Business Reference Suite is a database of tables and charts published in business and industry periodicals.
 3. IPoll - Roper's database of poll questions is available directly from Roper, or through Lexis/Nexis Statistical.

Acquisition Models

1. Reactive - Get what is asked for when it's asked for.
2. Proactive - get data without being asked, subscriptions
3. Partly a decision based on archiving capability. If you have a mechanism for archiving data, then a proactive approach may make your life easier. If not, then you have no place to store the data once you have it!
 1. A lot of data are already archived elsewhere such as ICPSR and do not need to be archived locally.
4. Also, of course, based on financial resources. Subscriptions to online data sources can be very expensive. Often, a particular purchase will require cooperation from other librarians as well as the department making the request.

ARCHIVING

What to Archive?

1. Heavily used data
 1. Can be re-packaged to make it easier to use.
 1. StatTransfer can produce a flat file with SAS, SPSS and Stata code to read it
 2. Also for preservation - ASCII format only! You can provide daily-use copies of the data in proprietary (SAS, SPSS) format, but not back-up or preservation copies.
 1. If you have the Transport (SAS) and/or Portable (SPSS) files, you can archive those with the ASCII data
2. Difficult to use data
 1. Some data do not come with pre-written code to read the data
 2. Some data are very complex and difficult for the average user to manipulate.
3. Rare and difficult to find data
 1. You may never find it again!

How to Archive It

1. Can be put on a server and patrons given web or network access.
 1. Perhaps the best way
 2. Make sure the data and the server are backed up at regular intervals!!
2. Another option might be to put the data on CD.
 1. Would have to decide whether to put one study on a CD or several
 1. One study per CD will lead to a proliferation of CDs
 2. More than one study per CD may mean somebody has to wait because someone else is using the CD
 3. If the CD goes bad (and they do!) then all of the studies will have to be re-generated.
3. Another consideration is cataloging the data.
 1. Will you put the record in the main catalog, a "databases" catalog, or maintain your own?
 1. If you put it in the main catalog, then one study per CD may be necessary.
 2. Can your main catalog handle records for items that are on the server? How will it do so? How will it handle changes on the server?
 1. One option might be for the record to point to your web page which would have further instructions for accessing the data.
 3. Maintaining your own may be the best way to go
 1. You have full control over the records
 2. Can change things when you need

- or want to
 - 3. Can integrate it with whatever archival/access system you have.
 - 2. What information should you catalog?
 - 1. Number of observations, location, dates, type (cross-section, time series, panel, etc)
 - 2. Related studies - particularly for time-series and panel studies
 - 3. Should the data, programs and codebook be cataloged separately or as one record?
 - 1. Depends on how you archive them and provide daily access.
 - 1. All files zipped together, then one record, with information on what files are included.
 - 2. If all files are separate, then you'll obviously need separate records.
 - 4. What to do about codebooks?
 - 1. Most, if not all, are now online
 - 1. Should they be printed out?
 - 1. If they are very large and heavily used.
 - 2. If they are part of a time series or panel study; it's often easier to look at several codebooks at once if they are all laid out side-by-side.
 - 3. Must be sure that they are current! Errors are sometimes discovered and the online versions are changed with no notice.

Special Data

- 1. Archiving of faculty data - issues: who gets access, cleaning, withdrawal?
 - 1. Mostly done by research units
 - 1. Either required by funding agency or the unit itself.
 - 2. Several issues to consider
 - 1. Who gets access? Some professors want to restrict access to specific individuals.
 - 2. Who and how do you ensure data integrity? Have the data been cleaned, privacy protected, codebook matches the data and questionnaire?
 - 3. What do you do if you find a problem later on?
 - 4. What if the professor decides sometime later on that he/she does not want the data in the archive?
- 2. What about restricted data?
 - 1. Avoid it if at all possible!
 - 1. Most restricted licenses restrict the use of the data to specific individuals and new users must apply for access. Applications can take as long as a month to process and often need to be made in writing.
 - 2. The data cannot leave the building
 - 3. Must be kept in a locked cabinet in a

- locked office when not in use.
4. You WILL be audited - without warning - to ensure compliance. If you are found to be in violation, then the data can be confiscated as well as other penalties.

ACCESS

1. Now that you have all this wonderful data archived and cataloged, how are you going to let people get access to it?
 1. CD - In addition to everything above, you also have to have someplace to store them.
 1. Will you let them circulate?
 2. One option is to get a CD tower or virtual drive and share them across the campus network
 2. Server - definitely the better option overall, but you will need the IT support and other things mentioned above.
 1. Can allow web access to the data
 2. Can allow web extraction as well
 3. Easier to make corrections to files when necessary

ASSISTANCE

1. How much assistance do you give your patrons?
 1. Depends on what your supervisor expects and what other resources are available.
 1. Personally, I believe you should at least know how to get a data file into SAS, Stata or SPSS, as well as be able to use StatTransfer.
 2. To be really effective, it would be advisable to at least sit in on a basic social research methods class and/or basic statistics class.
 2. A full-service data center requires more than one person
 1. Four types of data professionals: computer, statisticians, librarians, producers.
 1. You won't necessarily have to produce the data yourself, but at least have some knowledge of what goes into producing it.
 2. Computer support can be provided by the IT department.
 3. Statistical support can be provided by grad students.
 3. Software packages of which you should be at least aware
 1. [SAS](#) - SAS is the biggest of all statistical packages (as well as being the largest privately-owned software company). SAS can do just about anything you will ever need to do. SAS also has a pretty steep learning curve. There is a fill-in-the-blank interface (SAS/ASSIST) available, but it is not as well-developed as Stata's or SPSS's. To really make the best use

- of SAS, you must write a program.
2. [SPSS](#) - SPSS is another very popular statistical package. It has probably the best GUI interface of the three packages, as well as the ability to write programs. Like SAS, you can probably do everything you will ever need to in SPSS. You can do most of your work in the GUI, but not all, so you may need to learn how to program in SPSS. Like SAS, programming in SPSS has a pretty steep learning curve.
 3. [Stata](#) - Stata is a relatively (compared to SAS and SPSS) easy to learn package which give you a choice among a command-line interface, syntax or program file (called a "do-file" in Stata), and pull-down, fill-in-the-blank GUI interface. Stata is very good with time-series data and has many survival analysis routines. Stata also gives you the ability to program your own commands. One drawback to Stata is that it loads the entire dataset into memory, so if your dataset is very large, you may not be able to use Stata. This is a relatively rare occurrence, however. Generally, if you have little or no experience with any statistical package, Stata is probably your best choice.
 4. [StatTransfer](#) - Used to convert files among many different packages (like SAS, Stata and SPSS), StatTransfer can be the Data Librarian's best friend!
4. Faculty/student outreach
1. Faculty who are heavy users can be helpful
 2. New faculty often need to know what resources are available
 3. Department chairs and deans are not always the best people, but always worth a try!
 4. Those in charge of grad programs
 5. Grad student associations

© 2005 Syracuse University Library, 222 Waverly Avenue, Syracuse, NY 13244 (315) 443-2093

[Departments](#) | [Employment](#) | [Facilities](#) | [Hours](#) | [Locations](#) | [News](#) | [Policies](#) | [Publications](#) | [Staff](#)
[Ask Us](#) | [Circulation](#) | [Digital Imaging](#) | [Instruction](#) | [InterLibrary Loan](#) | [ISDP Services](#) | [Reserves](#) | [Service Request Forms](#)
[Access My Account](#) | [Find Articles](#) | [Find Books](#) | [Remote Access](#) | [Reference Desks](#) | [Visitors Guide](#) | [Workshops & Tours](#)
[SUMMIT Catalog](#) | [Databases](#) | [E-Journals](#) | [MyLibrary](#) | [Subject Guides](#) | [Reference Shelf](#) | [Course Materials](#) | [Other Catalogs](#)

[contact us](#) | [search](#) | [home](#) | [site map](#)
[Send Comments to Web Administrator](#)

Last modified: May 19, 2006 11:11 AM

<http://library.syr.edu/information/mgi/nds/iassist2006/index.html>