
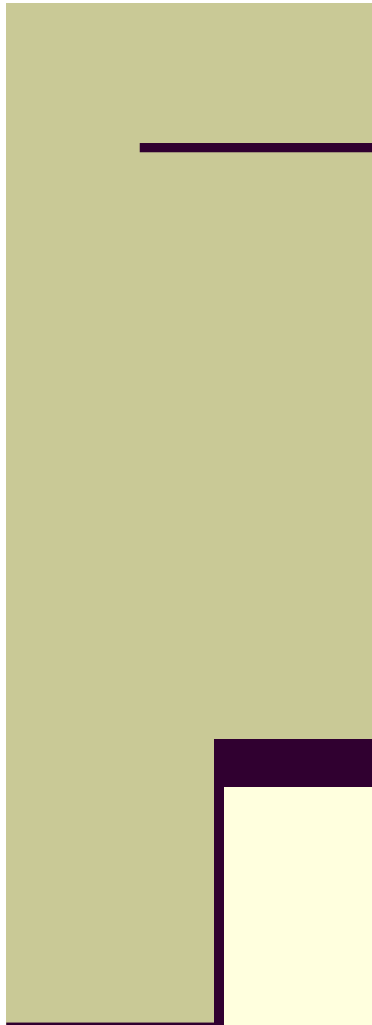




Data Sharing: Benefits and Barriers



Roberta Balstad, Ph.D.
Columbia University
U.S. Board on Research Data and Information

Focus Today

- The barriers and the benefits of scientific data sharing
- Examination of barriers to data sharing in the past and today
- Examination of scientific barriers and institutional barriers to data sharing
- The relationship between data sharing and professional data management

Benefits of Scientific Data Sharing

- Scientists base their research on foundations laid by other scientists
- In the past, this meant building their research on scientific theories proven by other scientists
- Today it can also mean using data that have been collected by other scientists
- Sharing data reduces both the cost of data collection and the overall cost of research

Benefits of Scientific Data Sharing

- New data collection is expensive; data sharing can cut costs
- Data collection requires considerable time and effort from scientists
- The data that can be obtained from other scientists are often better in terms of quantity, quality, and scale than data any single scientist can collect on his own
- Data collected by scientists at different time periods can be combined to permit the analysis of change over time

More Benefits of Data Sharing

- New, synthetic data bases can be built of data combined from many smaller data bases
- Data mining of the new data base can be an effective research strategy
- The use of common data bases allows scientists to test and retest their findings against those of other scientists and promotes progress in science

Early Data Sharing

- Scientists have been advocating data sharing for many decades
- In some fields, scientists have always used data collected by others (examples: economics and demography)
- In others, the idea of using data collected for other purposes was new (example: *Secondary Analysis of Sample Surveys*, 1972)
- Although many scientists agreed that it was a wise strategy, it was not widely adopted

Barriers to Early Data Sharing

- Similar to barriers today
- Barriers to using data from other scientists:
 - Needed to locate and obtain data
 - Needed to evaluate data (concern about multiplication of errors)
 - Needed to validate data (check its accuracy)
- Required appropriate methods for data analysis (examples: Meta analysis and data mining)
- Required technologies capable of handling large data bases

How Does Data Sharing Differ Today?

- New technologies
- New methods
- New analytic requirements
- Data sharing requires that scientific data be managed professionally

Barriers to Providing Data to Other Scientists

- Data possessiveness, particularly if data were difficult to obtain
- Data documentation and management
 - These require professional data support if they are to be done correctly (Example: preparing metadata, checking for data corruption, transferring data to new media over time)
- Common data standards are needed (Example: metadata standards, geospatial data standards)

Institutional Barriers to Scientific Data Sharing

- Access to data
 - Scientists are producing massive amounts of data, only some of which are available through formal data centers
 - Other databases are held by individuals and can be difficult to obtain, particularly years after the data were collected
 - Some institutions restrict access to data or exert ownership over the data
- Research scientists rarely have access to data managers
- If data are to be shared, there must be common standards for the data and the software
 - Example: OGC standards for digital spatial data

International Barriers

- Language differences
- Economic differences that make data sharing expensive
- Legal differences that promote copyrighting rather than sharing data
- In some parts of the world, there are bandwidth limitations that make data access difficult
- But increasingly there are common data documentation standards and growing ties among scientists that encourage data sharing

Thank you!
