Preservation of Geoscience Data and Collections

By Linda R. Musser

Fletcher L. Byrom Earth and Mineral Sciences Library
Pennsylvania State University
105 Deike Building
University Park PA 16802
Telephone: (814) 863-7073
Fax: (814) 865-1379
Email: Lrm4@psu.edu

DATA AND COLLECTIONS IN PERIL

Every day, geoscience data and collections are in peril of being lost– through deterioration, lack of space, loss of equipment to read the data, or lack or loss of documentation or metadata. The sheer volume of geoscience data and collections are daunting (table 1). A 1997 survey by the American Geological Institute (AGI) identified those data and collections available to be transferred to a repository if one were available (table 2). Unfortunately, few state geological surveys have space available to accept these materials.

In 2001, the National Research Council formed a committee to investigate this issue and recommend solutions. Specifically, the Committee on the Preservation of Geoscience Data and Collections was asked to:

- Develop a strategy for determining which geoscience, paleontological, petrophysical, and engineering data to preserve.
- Examine options for the long-term archiving of and provision of access to these data.
- Examine three to five access and repository case studies as examples of successes and failures.
- Distinguish the roles of public and private sectors in data preservation.

In addition to documenting the extent and nature of the problem, the committee identified factors that led to the loss of geoscience data and collections. These included:

- Loss or lack of space.
- Inadequate supporting information (metadata).
- Retirement or departure of staff without recording their knowledge.
- Deterioration of materials or metadata over time.
- Failure to migrate data to newer, more stable media or to operate with current technology/hardware.
- Lack of understanding by decision-makers of the value of data and collections.

Table 1. Minimum estimates of the volume of geoscience data and collections in the United States (National Research Council, 2002).

Collections	Units	Total
Core (ice)	Tubes	14,500
Core (rock/sediment)	Boxes	8,015,715
Cuttings	Boxes	10,402,000
Thin sections	Slides	647,000
Washed residues		180,000
Other well record		2,045,000
Fossils	Specimens	122,935,000
Minerals/rocks	Specimens	828,000
Data		
Seismic (2-d)	Line-miles	357,020,300
Seismic (3-d)	Square miles	249,849
Velocity surveys	Paper/digital	87,500
Well logs	Paper/film/ tape/digital	46,021,700
Scout tickets	Paper//film	21,960,350
Geochemical analyses	Paper	1,750,000

The committee recommended that priority for preservation should be placed on geoscience data and collections that are in danger of being lost (National Research Council, 2002). The highest priority should be given to those data and collections that are well documented and impossible or extremely difficult to replace. In addition to establishing priorities, the committee developed recommendations regarding storage, curation, cataloging and indexing, access, and discovery and outreach of geoscience data and collections.

The Committee's report reinforced the need for adequate space and funding to preserve geoscience data and collections through a combination of new space, support for existing repositories, and the creation of partnerships and consortia among repositories. Regarding curation,

Table 2. Volume of geoscience data and collections available to be transferred to a repository (National Research Council, 2002).

Data source	Volume available
Cores	10,000,000 feet
Cuttings	2,500,000 boxes
Thin sections	30,000 slides
Seismic data (paper, film and digital)	102,500,000 line-miles or films
Related data	25,000 velocity surveys
Well logs (paper, fiche and digital)	7,100,000 logs, cards, or tapes
Scout tickets	2,500,000 paper and fiche
Geochemical analyses	50,000 paper

cataloging, and indexing, the Committee emphasized the need for more support for these value-added activities and recommended that methods be developed within the scientific community to recognize outstanding contributions to curation, cataloging, and indexing of geoscience data and collections. Recommendations related to access, discovery, and outreach activities included more funding to make indexes available via the Internet and promoting recognition of the value of geoscience data and collections via citation.

Since publication of the committee’s report there has been progress on several of the recommendations. A task force has been formed to promote the citation of geoscience data and collections by the geoscience community. The energy bill, currently before Congress, requests $30 million per year for 5 years for the USGS to distribute to state surveys for preservation efforts. Efforts are ongoing in the private sector and via the AGI Foundation to raise funds for preservation, and a group has been formed to act as a national advisory board on preservation of geoscience data and collections.

GOOD PRESERVATION PRACTICES

Hopefully, these initiatives will prove successful and more funding and space will become available to support preservation and archiving activities. In the meantime, it is important to take steps now to preserve geoscience data and collections. Be aware of existing guidelines, or staff in your organization who may assist you in your efforts.

For example, are there existing records management guidelines that cover your data? Have you consulted with the state archives staff regarding assistance they may be able to provide to your agency? Are there policies or guidelines developed by others that would be useful? The resources and guides developed by the International Council on Archives, the Archaeology Data Service , and the Council on Library and Information Resources offer good advice and examples. Some tips for handling digital collections include:

- **Document the data (create metadata).**
  The committee noted the lack of documentation as a persistent and widespread problem. If data lacks adequate documentation, it is of limited use. At a minimum, keep some documentation about the data on paper or simply written on the storage medium itself. For example, it is helpful to have something readable by the human eye that tells you what the data is, what format it is stored in, and what equipment or operating system is required to read it.
- **Transform data into a common format.**
  Don’t risk the data by leaving it in a proprietary format unless absolutely necessary. There is a reason that text format is still widely used.
- **Store the data on long-term storage media.**
  Examples include gold CDs, DVD, and, yes, paper. Also, find a safe place, preferably offsite, to store back-up copies of data in case of catastrophe.
- **Establish a schedule for refreshing and migrating the data.**
  Long-term storage medium doesn’t mean it lasts forever. Every storage medium needs to be periodically checked and refreshed. Technology changes may require migration to an alternate medium to accommodate loss of hardware. (Example – how many people today own turntables that can play a 78-speed LP?)
- **Assign responsibility for preservation.**
  Make certain it is part of someone’s job to monitor preservation.
- **Make knowledgeable decisions.**
  It is important to recognize that not everything is worth saving. This is why records management guidelines are useful.

SUGGESTED REFERENCES

Archaeology Data Service, 2003, Guides to Good Practice. <http://ads.ahds.ac.uk/project/goodguides/g2gp.html>.

Council on Library and Information Resources, 2004, CLIR Reports. <http://www.clir.org/pubs/reports/reports.html>.

International Council on Archives, 1997, Guide for Managing Electronic Records from an Archival Perspective. <http://www.ica.org/biblio/guide_eng.html>.

National Research Council, Committee on the Preservation of Geoscience Data and Collections, 2002, Geoscience Data and Collections - National Resources in Peril: Washington, DC, National Academy Press, 107 p., <http://www.nap.edu/catalog/10348.html>