

THE CHALLENGE OF ARCHIVING AND PRESERVING REMOTELY SENSED DATA

John L. Faundeen

Archivist, U.S. Geological Survey, EROS Data Center, Sioux Falls, SD 57198 USA

Email: faundeen@usgs.gov

ABSTRACT

Few would question the need to archive the scientific and technical (S&T) data generated by researchers. At a minimum, the data are needed for change analysis. Likewise, most people would value efforts to ensure the preservation of the archived S&T data. Future generations will use analysis techniques not even considered today. Until recently, archiving and preserving these data were usually accomplished within existing infrastructures and budgets. As the volume of archived data increases, however, organizations charged with archiving S&T data will be increasingly challenged (U.S. General Accounting Office, 2002). The U.S. Geological Survey has had experience in this area and has developed strategies to deal with the mountain of land remote sensing data currently being managed and the tidal wave of expected new data. The Agency has dealt with archiving issues, such as selection criteria, purging, advisory panels, and data access, and has met with preservation challenges involving photographic and digital media. That experience has allowed the USGS to develop management approaches, which this paper outlines.

Keywords: Remote sensing, Science data, Archiving, Preservation, Selection criteria, Purging, Data access, Records management, U.S. Geological Survey.

1 BACKGROUND

The U.S. Geological Survey's (USGS) scientific use of remotely sensed data spans more than six decades. Since 1939, aerial photographs have been used to interpret features and create map products. In 1966, the Department of the Interior (DOI), which includes the USGS, first proposed to design, build, launch, and manage an Earth-observing satellite. That initiative led to the Landsat series of satellites (U.S. Geological Survey, 1979). This tradition created opportunities to include the Corona declassified data and aerial photographs from other U.S. Federal agencies. Combined, more than 12 million images acquired from satellites or aerial platforms reside with the USGS. Those images serve environmental, geologic, geographic, and human health needs through scientific applications (Draeger, Holm, Lauer & Thompson, 1997). Figures 1-3 illustrate scientific uses for the USGS remotely sensed data. This paper provides an overview of the management approach applied to overcome the challenges of maintaining and preserving these diverse scientific data sets.



Figure 1. Mt. St. Helens Eruption Monitoring (USGS).

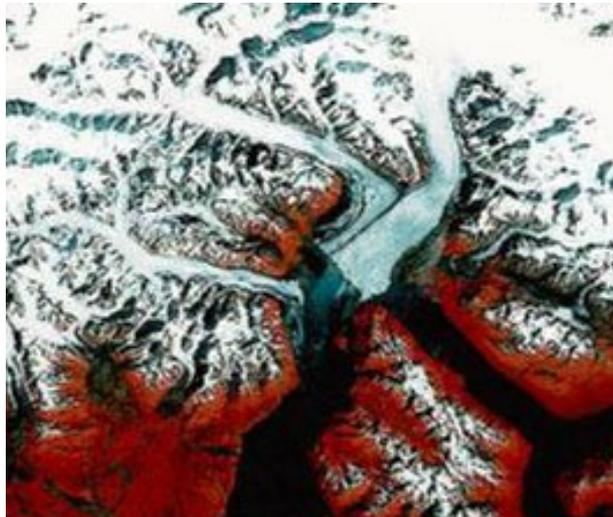


Figure 2. Hubbard Glacier Tracking (USGS).



Figure 3. Brazilian Rainforest Change Analysis 1975, 1986, and 1992 (left to right, USGS).

2. MANAGING AND PRESERVING

To adequately manage and preserve large volumes of remotely sensed scientific data, the USGS has adopted records management techniques (Ham, 1993). The life cycle of records management includes (a) the creation of records, (b) the maintenance and use of records, and (c) the final disposition of records.

The creation stage, in USGS terms, relates to when data are acquired from satellites or delivered after an aerial photography mission. A primary challenge is determining what to accept since the possible public and private sources of data now total more than 50, each having its own attributes. An organization's mission or mandate must be examined to ensure that only the most important data are acquired (Ham, 1993). Advisory committees can guide an organization and must be composed of a balance of backgrounds and contain a large scientific component to address archive issues (15 USC 5601, 1992). An investment of time and money from the archive is needed to work with such committees. Resources must also be available to formulate new dataset evaluation in terms of media, hardware and software, training, and human skills.

The maintenance and use stage is generally broken down this way:

- Organize into groups
- Maintain
- Manage access
- Facilitate retrieval
- Preserve
- Audit or oversight

The USGS organizes its records into groups on the basis of the media the data reside on and then by project or mission. Analog photographs are separated from digital data, as they require different management approaches. Segregating by project or mission allows the USGS to manage entire collections more easily

and facilitates easier access. As large projects or missions are created, integrating the data into existing archives can be difficult.

The *maintain* element represents environmental protections (Van Bogart, 1995). Proper temperature, humidity, and security controls ensure physical maintenance (American National Standards Institute, 1993). Locations with seasonal temperature and humidity swings present challenges to controlling the environment.

Managing the physical access to remotely sensed data requires strict enforcement. Many times the bigger issues concern managers who do not possess the programmatic need for access, but who nonetheless demand complete access.

Facilitating retrieval is accomplished indirectly because researchers are not allowed to physically handle the original or master copies of the scientific data. Generating a copy print or film product accommodates requests for aerial photographs. Requests for data on digital media are met by reading the archive data into a production machine for copying. Researchers use metadata systems to determine what data are available and to verify geographic coverages of interest. Maintaining the metadata has become more complex with new access means that are supported by these databases.

Preservation is a big part of any records management role. Regarding scientific data, the goal is to ensure that the data are available for future researchers. Technological obsolescence is one of the biggest digital data challenges to a records manager (U.S. General Accounting Office, 2002). Media, hardware, and software all become obsolete. An earlier rule of thumb was that data residing on a media needed to be migrated to a next-generation media every 10 years. Although general in nature, the 10 years usually provided enough lead-time to accomplish data migrations. Today, technology is moving faster, causing the 10-year window to shrink. Five years has been discussed in various archiving circles as a more realistic period for digital data. The USGS has designed, built, and operated three data migration systems. These complex systems migrated the digital scientific data from aging to newer media and cost millions of dollars while spanning many years. The successful operation of these migrations ensure that the data will be available for researchers. A fourth system is being built to handle previously migrated data to yet another media. The cycle continues. Planning for these preservation activities is extensive and must be done before data are threatened.

The *audit or oversight* element is intended to ensure that data being acquired are aligned with an organization's mission or statutory code. For some of the satellite data managed by the USGS, a public law, cited as P.L. 102-555, states that the DOI/USGS must maintain, preserve, and make available those data (15 USC 5601, 1992). Without this element, archives can add to their burden of volume by taking on data that rightfully should go to other organizations. Often, this step is not undertaken, resulting in the misalignment of data in regard to organizations held responsible for their care.

The final disposition stage occurs when data have exceeded their scientific value to the originating organization. At that point, a determination must be made as to the continuing value of the data to the hosting organization (Schellenberg, 1956). If continuing scientific value for the data exists, the data could be transferred to an organization's or a nation's central archive. If the data have been completely superseded by better data, or if copies of the data exist with other organizations that intend to maintain them, destruction of the records may be in order. Determining this is often time consuming, and for good reason. Destruction is final.

3. SUMMARY

Challenges abound for archiving and preserving remotely sensed data. The USGS uses records management techniques and life cycle management concepts to help it deal with the volume, breadth, and depth of the remotely sensed data available today. Those same skills are used to preserve the decades-old data that help us better understand today through the eyes of yesterday.

4. REFERENCES

15 USC 5601 (16), The Land Remote Sensing Policy Act of 1992.

American National Standards Institute (1998) *American National Standard for Imaging Materials: Processed Safety Photographic Films: Storage. ANSI/PIMA IT9.11-1998*. New York, NY: Author

Draeger, W. C., Holm, T. M., Lauer, D.T., & Thompson, R. J., (1997) The Availability of Landsat Data: Past, Present, and Future, *Photogrammetric Engineering & Remote Sensing*, 63(7), 869-875.

Ham, F. G. (1993) *Selecting and Appraising Archives and Manuscripts*. Chicago, IL: The Society of American Archivists.

Schellenberg, T. R. (1956) *Modern Archives: Principles and Techniques*. Chicago, IL: The University of Chicago Press.

U.S. General Accounting Office (2002) *Information Management: Challenges in Managing and Preserving Electronic Records, GAO-02-586*. Washington, DC: Author.

U.S. Geological Survey (1979) *Landsat Data Users Handbook* (rev. ed.), Arlington, VA: Author.

Van Bogart, J. W. C. (1995) *Magnetic Tape Storage and Handling, A Guide for Libraries and Archives*. Washington, DC and St. Paul, MN: Commission on Preservation and Access and the National Media Laboratory.