A Selection and Archiving Strategy for Science Records

John L. Faundeen; U.S. Geological Survey Earth Resources Observation and Science Center; Sioux Falls, SD/USA

Abstract

The U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center archives electronic science collections that total over two petabytes in size and over 100,000 rolls of aerial and satellite film. Limited resources, the evolution of missions, and recommendations from advisory committees have led to the development of a scientific records appraisal process as a means for determining long-term archiving priorities. The process was formed through extensive literature searches describing approaches used to appraise administrative, physical artifacts, and science records. Less information was available that specifically addressed science records; therefore, relevant portions from each records appraisal process was assembled. In addition, involvement with the appropriate stakeholders was deemed critical and led to the active participation of scientists, records managers, and senior managers in the process. As part of the documentation portion of the process, an extensive online tool was developed to capture information describing each collection and detail preservation or access challenges that may be part of a collection. The U.S. National Archives and Records Administration recommends the tool as a best practice for U.S. federal agencies. To date, over 30 science collections have been appraised. This paper will detail the process used to appraise science collections for long-term archiving, the composition and rationale for the tool elements, and the results the USGS has attained.

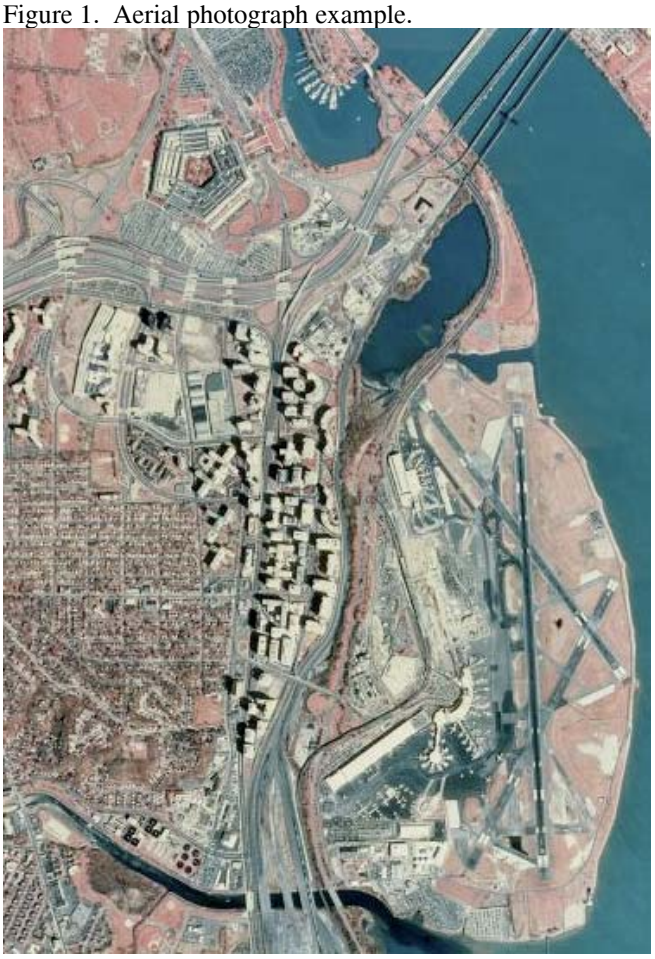
Context

The USGS EROS Center was established in 1972 to receive, archive, and distribute satellite images from the Landsat series of satellites. For the past 37 years, the Landsat satellites have provided continuous data of the Earth's land mass, coastal boundaries, and coral reefs -- creating an unprecedented comprehensive record of landscape dynamics. Landsat-5 and -7 continue to capture hundreds of images of the Earth's surface each day.

The imagery provided from the satellites was recorded on digital media. Historically, most researchers were more comfortable analyzing film images using photo interpretation techniques developed extensively in World War II. Large corporations, government research labs, and academic programs were among the first to use computers for their analysis work. To meet the needs for analog and digital users, both electronic and images written to film were created and archived. Landsat digital images total over 2.3 million and occupy nearly 650 terabytes. Landsat film images total nearly 3 million. (These numbers do include duplicate images.)

Beginning in the late 1970s, the USGS EROS Center began to accept and solicit observations taken from aerial platforms. These images were originally collected by U.S. federal agencies such as the National Aeronautics and Space Administration (1.8 million images), the Department of Defense (650,000 images), the Bureau of Land Management (80,000 images), the Bureau of Reclamation (60,000 images), and the National Park Service (35,000 images). Combined with the USGS' own 5.7 million, over 8.6 million aerial images supplement the Landsat imagery. Adding to that mix are numerous smaller collections containing satellite or aerial digital or film files. The collections date from the late 1930s up to the present and geographically occur primarily over the United States for the aerial images and throughout the world for the satellite images.

Figure 1. Aerial photograph example.



Color-Infrared aerial photo over the Crystal City, VA area, March 17, 1994.

Reasons to Appraise

Documentation as to how, when, or most importantly, why all of these images came to be part of our holdings was often lacking. We have speculated that because the EROS Center was established

as an archive for observational holdings, federal agencies may have felt their similar records would be well placed here. Perhaps some agencies did not have the capability or desire to serve researchers requesting copies of the images. It may also have been a financial driver such that agencies strategized their archiving costs could be reduced if another entity bore that responsibility. The need to know the reasons science records were transferred to the USGS has existed for some time.

Recent years have also been challenging from a budget standpoint. When resources are limited, priorities often must be more clearly determined and followed. Such an atmosphere lends itself to taking stock of what is being archived and to more closely scrutinize the collections that are offered regularly to us.

It is not uncommon for an institution's mission to evolve over time. When that occurs, reevaluations of decisions made earlier are often done. How a collection that was obtained decades earlier aligns to a current mission is a legitimate question and should be periodically asked.

The EROS Center has benefited from several federal advisory committees [1] related to our archiving role of observational records. One outcome from these committees has been a set of recommendations concerning general categories of observational records we should be seeking or specific collections to acquire and, in some cases, collections that we should dispose of. While this guidance is priceless, acting upon the recommendations was often difficult because of organizational structure or confusing oversight roles. Clearly, we needed to establish a means to take advantage of this guidance, address the lack of transfer documentation, ensure that our shrinking budgets supported the right science collections, and programmatically align the collections we should archive and dispose of those that are now incompatible with our mission.

The Appraisal Process

Around 2001, the EROS Center began to adopt the principles of records management as a means to better address its archiving role. The principle of appraisal was thought to hold promise as a way to adequately judge collections offered to us and as a means to determine if we are currently archiving the right collections. We began by conducting literature searches for "science records appraisals" and soon discovered few had been documented. We expanded our scope to include the appraisal of physical artifacts, administrative files, and any other type of record appraised that appeared in archival or records management literature. From the multitude of papers examined, several general points were identified as well as specific ways organizations had evaluated records. As a result of these searches, we assembled an initial list of over 70 questions. The questions were Web–enabled, and the information entered is maintained and provided to the submitter as a text file. The online questions are referred to as the scientific records appraisal tool [2].

The questions fall into the sections of mission relevancy, policy, physical, metadata, cost/benefit, and a general category. Additional questions cover film collections. Under the mission relevancy section, questions relating to how the collection fits within our collection policy and how it complements our current holdings are addressed. Creating and using a collection policy was repeatedly mentioned in the literature. Craig [3] refers often to an organization's mandate, which can be represented through an institutional collection policy.

The policy section documents how the collection stands up to the four International Standards Organization [4] records management standards of authenticity, reliability, integrity, and usability. Any legal ramifications, such as copyright or legal rights of citizens, are also noted here.

The physical section records the current and best preservation levels or media for the collection. Electronic records may have processing histories available, or they may have had compression applied to the records. Regardless of media, the file naming convention is also noted.

Metadata is a critical section and we try to determine the amount, quality, level, and availability for the collection. Increasingly, it is important for the metadata to be electronic itself even if the collection is not. Many sources included the requirement to understand the metadata associated with a collection. The U.S. National Archives and Records Administration [5] (NARA) and Lyon [6] specifically illustrate the importance of metadata to a review.

Our cost/benefit section attempts to capture information related to acquiring, preserving, and making a collection accessible. Ham [7] emphasizes the need for doing cost/benefit analysis on collections, even stating that archivists must "attach a price tag" to their decisions. This is by far the hardest section for us to complete because often not all aspects of a collection are understood well enough to quantify the cost elements. Doing formal rough order of magnitude estimates would be the preferred approach, but we simply cannot afford the cost of doing all of the estimates. Our alternative is to apply relative terms like "high," "medium," and "low" to at least provide reviewers some sense of the cost associated with retaining or accepting a collection. Even disposing a collection has costs involved that should be documented.

The general section covers a wide range of questions including the temporal and spatial characteristics, how complete the collection is, provenance or chain of custody documentation, whether another organization may possess the same collection, and any unique preservation or access challenges the collection may entail.

If the records are photographic, additional questions such as the film base, the generation of the film, camera specifications, manufacturer of the film, and the overall condition of the film are noted. Knowing the film base, for example, is critical to us as older types of film can be flammable or deteriorate quickly.

The next step in developing our process was to gain management support to implement such a procedure. Because of our science focus, we first approached the science staff and briefed them on the concept and goals of establishing such a process. When their full support was secured, senior management was approached. With the science staff backing, senior managers quickly endorsed the concept and established the appraisal process as a required practice for existing and all offered collections.

Following science and management concurrence, we needed to determine who was actually going to be part of the appraisal process. The Archivist was given oversight of the process to ensure that the proper stakeholders were involved, that complete information was assembled, that consistent approaches were used, and that documentation was captured to defend our decisions.

Because we deal with observational science records, the inclusion of scientists in the review process was deemed critical. This soon became our first real challenge. While every scientist we approached supported the concept of reviewing our own collections and those offered to us, gaining their time to actually review a collection proved to be difficult. Their own research is time-consuming and they are often under time constraints to complete projects. When the process was first used, months could slip away while the scientists found time to review and compile their opinions. Eventually, we settled upon a rotating group of five scientists; one or two with the most experience related to the collection being reviewed would respond to a request. This approach was decided upon after discussions with the senior science manager. Their time spent reviewing collections also had to be accounted for, and an overhead budget was established for the science reviewers to charge their time.

Decisions made from the appraisal process will have financial implications involving long-term archiving, preservation, and access functions. To address these financial aspects, Project Leads were included because they have budgetary oversight for our science collections. If this area were left unrepresented, any recommendations would not be acted upon. By including them in the process, they become vested in the outcomes and could plan for financial impacts identified in the process.

The last group of stakeholders is the senior managers who ensure that the recommendations align to our agency's programmatic goals and objectives. They determine whether a recommendation will be endorsed.

The actual process begins with the Archivist who maintains the list of collections being appraised and the order that they are addressed. Because each appraisal has its own unique background and characteristics, the time to complete an appraisal varies significantly. Thus, several appraisals are generally occurring at the same time. Records management staff assemble all of the known characteristics about the collection using the developed scientific records appraisal tool to document the information.

The science reviewers are then engaged and provided the information gathered by the records management staff as background. The scientists are free to make any comments, but at a minimum, they must address the three following questions:

- Is there another organization within the scientific community that might benefit from or have an interest in these records?
- What were the original scientific uses for these records?
- What may be future scientific uses of these records?

The scientists have readily answered the first two questions but have been quite hesitant to address the last. Our experience shows the science staff to be very conservative when asked to make value judgments about science collections. While projections about future science use is understandably hard, these are the exact people to make such projections. The Archivist often has to reiterate this point with the scientists.

Once the scientist's comments and the information collection steps are completed, the records management staff provides a briefing to the Archivist and the relevant Project Lead. The Project Lead and the Archivist discuss all of the information provided, determine the merits of the collection, and seek concurrence in their recommendation.

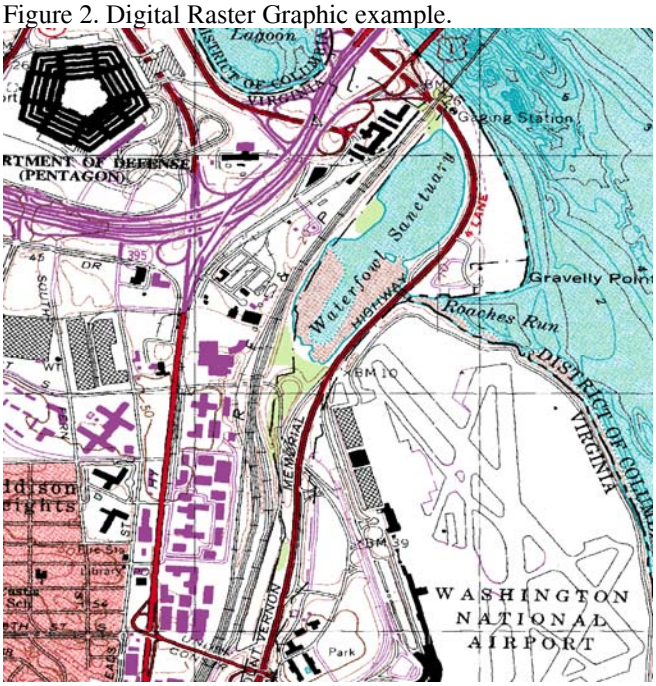
The next step is for the Archivist to compile a formal, written memo summarizing what has been learned about a collection. The memo includes background, evaluation, and recommendation sections. The recommendation states either to retain (or accept) or dispose (reject) a collection. Generally, the Project Lead and the Archivist will be in agreement regarding the recommendation, but there has been one occasion where complete opposite views were held. That situation was handled by clearly indicating in the memo that the Archivist and the Project Lead differed in their opinion. The memo is provided to our senior managers for review.

The last step also includes a formal, written memo from the head of our senior management indicating acceptance or rejection of the recommendation.

Results

The process began slowly and has evolved over time. It became more streamlined and relied upon formal documentation as part of the outcomes. Currently, a minimum of four documents are required to complete an appraisal: 1) the scientific records appraisal tool text file, 2) the appraisal briefing materials, 3) the recommendation memo from the Archivist to senior management, and 4) the memo from the head of our senior management indicating acceptance or rejection of the recommendation memo. Thirty-five collections have been formally appraised spanning a five-year period. To date, the senior managers have accepted each of the Archivist recommendations. Eleven collections were recommended for disposal and two offers from outside organizations were rejected. The remaining 22 collections recommended for retention were all existing collections.

Figure 2. Digital Raster Graphic example.



Scanned paper map of the Crystal City, VA area, 1983.

Historically, it was a common practice for us to accept almost any offer and to retain collections without any reviews. The

recommended disposal/not accept rate of 37% illustrates a large institutional change of behavior. The science collections appraised and the associated recommendations are listed in Table 1.

Table 1. Completed Appraisals

Collection	Recommendation
Apollo	Dispose
Shuttle Hand-Held	Dispose
Gemini Photography	Retain
Large Format Camera	Retain
ERIM Helicopter Data	Retain
AVHRR 10-Day Global	Retain
Shuttle Imaging Radar-A	Dispose
Slant Range Radar	Retain
Side-Looking Airborne Radar	Retain
Digital Orthophoto Quads	Retain
Digital Elevation Model Tiles	Dispose
AVHRR Level 0 Stitched Orbits	Retain
Shoreline Mask	Retain
Land-Sea Mask	Retain
AVHRR Single Passes	Retain
AVHRR 7- & 14-Day US	Retain
AVHRR 7- & 14-Day Alaska	Retain
Northern Great Plains	Retain
1990 US Land Cover	Retain
US & Alaska Companion CD	Dispose
1990 US Land Cover Prototype	Dispose
AVHRR 10-Day North America	Dispose
AVHRR Level 1B Stitched Orbits	Dispose
AVHRR NDVI June 1992	Dispose
Fish & Wildlife Service Photo Packs	Not Accept
Global Land Cover Test Sites	Dispose
Radar APQ-97	Retain
Landsat Return Beam Vidicon Film	Retain
Skylab Photography	Retain
Bureau of Reclamation Photography	Not Accept
Shuttle Imaging Radar-C	Retain
AVHRR US & Alaska CD	Dispose
North America Land Characteristics	Retain
Paper Maps and Imagery	Retain
Digital Raster Graphics	Retain

Summary

The USGS EROS Center set out to address problems associated with prioritizing the use of our shrinking budgets, subtle changes over time in our organizational mission, and following through on advice received from advisory committees regarding what science collections to archive. The ensuing scientific records appraisal process that was created has become a required USGS EROS Center practice to help ensure that our resources are used on the collections we should be preserving and providing access to.

The process evolved over time as experience led to refinement of the procedures. The incorporation of the scientific records appraisal tool became an important documentation piece of the process and was recognized by NARA as a best practice [8] for U.S. federal agencies.

Engaging the right stakeholders was the critical piece in getting the process accepted. Records management staff, scientists, Project Leads, and senior managers all must be involved for the recommendations to carry weight and be implemented.

Outside comments on the scientific records appraisal tool are encouraged. The questions used are available online with several organizations, including the Library and Archives of Canada and the Centre National d'Etudes Spatiales in France currently examining them. An annual review of the entire process and the scientific records appraisal tool are conducted to further refine the process.

Schellenberg [9] is correct that "Scientific records present special problems of evaluation to the archivist." The appraisal of these records, however, should still occur so that resources for preservation and access are applied to the right collections. And while it may not be comforting to mull over Rapport's [10] comment that appraisers are "apt to know nights of troubled soul searching," this may turn out to be one of the most important aspects an archivist or records manager undertakes.

References

[1] Department of the Interior, "National Satellite Land Remote Sensing Data Archive Advisory Committee," <http://eros.usgs.gov/archive/nslrsda/advisory/index.html> (25 February 2009).

[2] U.S. Geological Survey, "Scientific Records Appraisal Tool," <http://eros.usgs.gov/government/ratool/> (25 February 2009).

[3] Barbara Craig, Archival Appraisal: Theory and Practice (K.G. Saur GmbH, Munich, Germany, 2004) pg. 46.

[4] International Standards Organization, "ISO 15489-1:2001 (E) Information and Documentation – Records Management," Geneva (2001).

[5] National Archives and Records Administration, "Strategic Directions: Appraisal Policy," <http://www.archives.gov/records-mgmt/initiatives/appraisal.html> (2 November 2007).

[6] L. Lyon, "Dealing with Data: Roles, Rights, Responsibilities and Relationships," Consultancy Report. University of Bath, United Kingdom Office for Library and Information Networking 28 (2007).

[7] F. Gerald Ham, Selecting and Appraising: Archives and Manuscripts (The Society of American Archivists, Chicago, IL, 1993) pg. 58.

[8] National Archives and Records Administration, "Toolkit for Managing Electronic Records," <http://toolkit.archives.gov/pls/htmldb/f?p=102:1:4105603930723804921> (25 February 2009).

[9] T.R. Schellenberg, The Appraisal of Modern Public Records (General Services Administration, Washington, DC, 1956), pg. 274.

[10] L. Rapport, "No Grandfather Clause: Reappraising Accessioned Records," The American Archivist, 44, 149 (1981).

Author Biography

John Faundeen received his BA in geography from St. Cloud State University (1981) and his MS in geography from South Dakota State University (1983). Since then he has worked as the Archivist at the U.S. Geological Survey's Earth Resources Observation and Science Center in Sioux Falls, SD. He has published over 30 papers and in 2007 he received a research fellowship at the UK Digital Curation Centre for work involving the appraisal of scientific records.