
# GUIDANCE ON THE SELECTION AND APPRAISAL OF GEOSPATIAL CONTENT OF ENDURING VALUE

**Users/Historical Data Working Group**

**Federal Geographic Data Committee**

**May 2015**

# Federal Geographic Data Committee

Established by Office of Management and Budget Circular A-16, the Federal Geographic Data Committee (FGDC) promotes the coordinated development, use, sharing, and dissemination of geographic data.

The FGDC is composed of representatives from the following member agencies: Federal Communications Commission*, General Services Administration, Library of Congress, National Aeronautics and Space Administration, National Archives and Records Administration, National Capital Planning Commission*, National Science Foundation, Office of Management and Budget, Office of Personnel Management, Small Business Administration, Smithsonian Institution, Social Security Administration, Tennessee Valley Authority, U.S. Agency for International Development, U.S. Army Corps of Engineers*, U.S. Department of Agriculture, U.S. Department of Commerce, U.S. Department of Defense, U.S. Department of Education, U.S. Department of Energy, U.S. Department of Health and Human Services, U.S. Department of Homeland Security, U.S. Department of Housing and Urban Development, U.S. Department of the Interior, U.S. Department of Justice, U.S. Department of Labor, U.S. Department of State, U.S. Department of Transportation, U.S. Department of the Treasury, U.S. Department of Veterans Affairs, U.S. Environmental Protection Agency, and the U.S. Nuclear Regulatory Commission.

*Non-voting members

Additional Federal agencies participate on FGDC subcommittees and working groups. The Department of the Interior and the Office of Management and the Budget co-chair the FGDC.

FGDC subcommittees work on issues related to data categories coordinated under OMB Circular A-16. Subcommittees establish and implement standards for data content, quality, and transfer; encourage the exchange of information and the transfer of data; and organize the collection of geographic data to reduce duplication of effort. Working groups are established for issues that transcend data categories.

For more information about the committee, or to be added to the committee's newsletter mailing list, please contact:

Federal Geographic Data Committee Secretariat
c/o U.S. Geological Survey
12201 Sunrise Valley Drive
590 National Center
Reston, Virginia 20192
Facsimile: (703) 648-5755
Internet (electronic mail): fgdc@fgdc.gov
World Wide Web: http://www.fgdc.gov

**ACKNOWLEDGEMENT**

The Federal Geographic Data Committee Users/Historical Data Working Group would like to acknowledge the valuable contributions of the group participants in the authoring of this document.

# Contents

# Executive Summary

This "Guidance on the Selection and Appraisal of Geospatial Content of Enduring Value" document is authored by the Federal Geographic Data Committee's Users/Historical Data Working Group. The U/HDWG prepared this guidance to help Federal agencies and data stewards identify geospatial content of enduring value to the nation. "Enduring" in this context represents a time period beyond the immediate short-term.

The document situates selection and appraisal within the FGDC Geospatial Data Lifecycle and proposes a set of common appraisal and selection elements that guide data creators, data managers, theme leads and others in enumerating and defining activities and functions that support the ongoing accessibility and comprehension of digital geospatial data with enduring value.

This guidance suggests possible priority approaches on how resources might be allocated to support long-term preservation and access through appropriate Selection and Appraisal (S&A) processes in a challenging budget environment.

Geospatial content plays a significant role in a wide range of applications that support planning and decision-making for a broad range of Federal government activities. While many Federal government applications rely on the most current available content, there is increasing demand for older content to support historical and temporal comparative analyses related to change in the earth's natural and human landscape and physical infrastructures. Examples of applications that require historic content include: the study of climate change; disaster planning; environmental impact analysis; industry site location planning; and the resolution of legal challenges.

It is neither possible nor desirable to preserve every bit of geospatial information created by the Federal government. The S&A processes are tools to shape and describe the decisions made as to what geospatial content to keep and what to discard. S&A processes are critical because of the limited resources available across the government to provide for the long-term stewardship of geospatial content.

Federal libraries, archival, and museum institutions, including the National Archives and Records Administration, the Library of Congress, and the Smithsonian Institution, among others, have traditionally been stewards for geospatial content of long-term value to the nation.

However, the rapid pace of change of digital technologies and the exponential increase in digital data volume adds urgency to a call for reevaluation of S&A processes across the government by engaging content creators, aggregators, and other intermediary data stewards as early as possible in the processes of identifying, evaluating, managing, and preserving digital geospatial materials of long-term value.

NARA's disposition instructions state whether individual series of records are "permanent" or "temporary," as well as how long to retain the records. Records with historical value, identified as "permanent," are transferred to NARA, but there are many intermediate S&A actions that both

support agency missions and assist NARA in carrying out its records management responsibilities. This guidance aligns with the existing portfolio management approach used by the FGDC and records management processes developed by NARA (see Appendix 2, "Geospatial Data as Federal Records Subject to Management Requirements," for background on records management actions across the Federal government).

This document suggests S&A guidance that aligns with the lifecycle approach to the management of geospatial content. The document is designed to guide data creators and stewards on positive steps they can take early in the lifecycle of information to identify opportunities to make S&A decisions that will in turn support long-term stewardship processes. It also identifies a range of stewardship concerns that need to be addressed across the lifecycle to ensure that valuable information of importance to the nation remains accessible and usable.

The list of potential data creators and stewards is expansive and may include dataset and metadata managers, theme leads (*themes* are electronic records and coordinates for a topic or subject), NGDA theme lead agencies, data centers such as the National Oceanic and Atmospheric Administration (NOAA) discipline-oriented environmental data centers, NARA affiliated archives or NARA affiliated relationship organizations, such as the U.S. Geological Survey Earth Resources Observation and Science (EROS) Center.

Organizational focus has driven S&A decisions, with data producing agencies, data managing agencies, archives and libraries each making decisions according to their individual needs. Institutional mission will continue to drive behavior, but it is worthwhile to consider the utility of a broad, national, multi-organizational focus in addressing S&A decisions. This guidance suggests options that will ensure the successful stewardship of geospatial content of enduring value to the nation.

# 1. The Users/Historical Data Working Group (U/HDWG)

The Users/Historical Data Working Group is established under the auspices of the Federal Geographic Data Committee (FGDC) to promote and coordinate activities among those Federal agencies that are primarily users of, not generators of, geospatial data.

The U/HDWG promotes awareness among Federal agencies of the historical dimension to geospatial content. It works to facilitate the long-term retention, storage, preservation and accessibility of historic and superseded geospatial content and to establish a mechanism for the coordinated development, use, sharing, and dissemination of historically valuable geospatial content that has been financed in whole or part by Federal funds.

# 2. Definitions

a. Appraisal: a procedure typically associated with archival and records management processes and is defined as the evaluation of government information to determine its ongoing value and its merits for long-term or permanent retention.

b. Framework Layer: Initially referred to the most significant Federal government data, but the concept of "framework layer" has now been expanded to include the NGDA themes.
c. Geospatial Content: information and/or data that has a geospatial component, including resources such as geographic information system (GIS) data sets, digitized maps, associated metadata, remote sensing data resources and tabular data that are tied to specific locations on the surface of the earth.
d. Sample/Select: Used by NARA in instances where it is desirable to choose only certain files of value from a records series, rendering the remaining files as disposable.
e. Selection: a procedure typically initiated by libraries and other collecting institutions, and provides a comprehensive method to evaluate and document the materials that make up an organization's collection and the choices that go into acquiring materials of long-term value
f. Stewardship: the series of managed activities, policies, strategies and actions to ensure the accurate rendering of digital content for as long as necessary, regardless of the challenges of media and technological change to provide business users with high quality data that is easily accessible in a consistent manner.

# 3. Selection and Appraisal in the FGDC Data Lifecycle Model

The S&A processes enumerated in Section 4 below do not exist in a vacuum. They directly address the "archiving" component documented in Stage 7 (pdf, pg. 7) of the FGDC Geospatial Data Lifecycle developed by the FGDC Lifecycle Working Group. The lifecycle approach to the management of geospatial data is referenced in OMB Circular A-16.

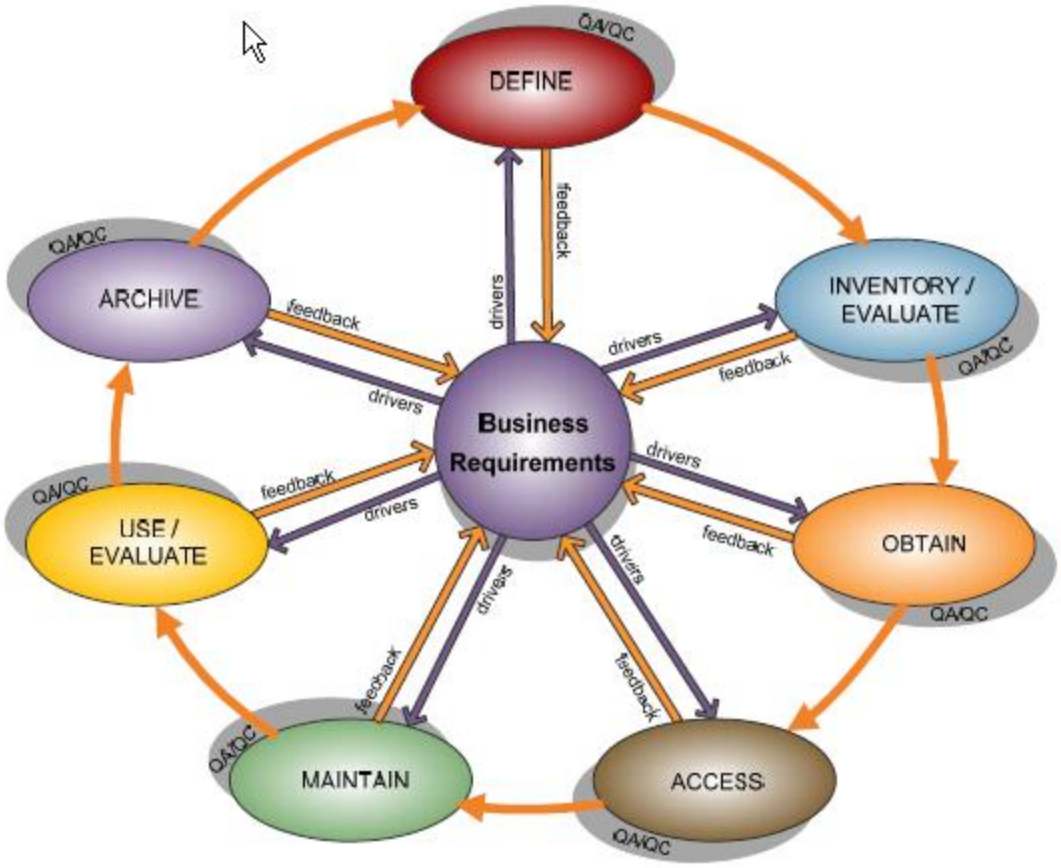


**Figure 1. The Geospatial Data Lifecycle: Image from FGDC document "Stages of the Geospatial Data Lifecycle pursuant to OMB Circular A–16, sections 8(e)(d), 8(e)(f), and 8(e)(g)."**

OMB Circular A-16 provides direction for federal agencies that produce, maintain or use spatial data, either directly or indirectly in the fulfillment of their mission. When published in 1990, it established a coordinated approach to developing the National Spatial Data Infrastructure (NSDI) and established the Federal Geographic Data Committee (FGDC), an interagency committee chaired by the Secretary of the Interior. OMB Circular A-16 was revised in 2002 to reflect changes in technology and further describe the components of the NSDI and assign agency roles and responsibilities for developing it.

OMB Circular A-16 "Supplemental Guidance" (pdf), released in November 2010, further defines and clarifies selected elements of OMB Circular A-16 to facilitate the adoption and implementation of a coordinated and effective Federal geospatial asset management capability that will improve support of mission-critical business requirements of the Federal Government and its stakeholders. Its primary focus is on geospatial data as a "capital asset" and "refers to its acquisition and management in terms analogous to financial assets to be managed as a National Geospatial Data Asset Portfolio" (from CRS Report Issues and Challenges for Federal Geospatial Information (pdf)).

The Supplemental Guidance provides the foundation for a portfolio management approach to a National Geospatial Data Asset (NGDA) Portfolio that comprises NGDA Themes and their associated NGDA Datasets (see list below in the section on *Data Inventory*). An NGDA Dataset is defined as a geospatial dataset that has been designated as such by the FGDC Steering Committee and meets at least one of the following criteria: supports mission goals of multiple federal agencies; statutorily mandated; supports Presidential priorities as expressed by Executive Order or by OMB.

The Supplemental Guidance defines a "Geospatial Data Lifecycle" that includes an "Archive" function, which is defined as "required retention of data and the data's retirement into long-term storage." The enshrinement of an "archive" function in lifecycle planning provides an impetus for implementation in creating agencies of the long-term processes, functions, actors and initiatives that represent S&A activities.

The "archive" function does not happen exclusively at the end of a dataset's useful life; it includes the S&A actions described throughout this document that can happen at different stages across the entire lifecycle. Additionally, true S&A approaches involve more than just redundant backup storage. From the stewardship community perspective, S&A implies managed storage at a government repository with management processes in place to ensure the long-term preservation and appropriate access to the data.

The need to address the entire lifecycle of digital geospatial data is even more important when taking into account recent federal data memorandum, including the release of the "OMB Memorandum M-13-13, Open Data Policy—Managing Information as an Asset" (pdf) from May 9, 2013 and the Office of Science and Technology Policy "Increasing Access to the Results of Federally Funded Scientific Research" memorandum of February 22, 2013.

While the defined set of "archive" processes is in a very early stage of development, the guidance provided in this S&A guidance document is a first step towards enumerating and defining S&A activities and functions by concentrating on the initial steps of S&A that can be tackled by creating agencies.

# 4. Common Elements of Selection and Appraisal Processes for Digital Geospatial Information

The common elements of an S&A are listed in decreasing order of importance. For example, agencies cannot successfully appraise or select digital geospatial information unless it has been inventoried and a determination made that it fits the organizational mission and legal mandate. .

The Common Elements:

A. Data Inventory
B. Alignment with Organizational Mission
C. Legal Rights, Restrictions and Mandates
D. Spatial Reference Information, Spatial Extent, and Temporal Information
E. Current Scientific or Cultural Heritage Value
F. Technology and Obsolescence Risks
G. Cost-Benefit Analysis

H. Tangible Media and Physical Condition
I. Metadata Quality, Completeness and Usability Uniqueness
J. Uniqueness
K. Provenance
L. Future Value Determination

## *A. Data Inventory*

Data inventory should be one of the first steps in S&A processes and part of regular, ongoing data management. The basic guidelines for inventory of Federal government geospatial assets are grounded in the OMB Circular A-16 "Supplemental Guidance" (see above), especially the section on "Elements of the National Geospatial Data Asset Portfolio," which provides for an accurate and accountable inventory of Federal Geospatial Portfolio assets. NGDA Themes and Datasets provide the structure around which inventories of geospatial content will be built.

The FGDC has identified 16 NGDA Themes (as of February 2013):

- Biota
- Cadastre
- Climate and Weather
- Cultural Resources
- Elevation
- Geodetic Control
- Geology
- Governmental Units, and Administrative and Statistical Boundaries
- Imagery
- Land Use-Land Cover
- Real Property
- Soils
- Transportation
- Utilities
- Water – Inland
- Water – Oceans & Coasts

This is in contrast to the original 34 NSDI data themes established in OMB Circular A-16. Seven of the original 34 NSDI data themes were identified as especially critical to the National Spatial Data Infrastructure and were thus identified as "framework layers":

- Cadastral
- Digital Orthoimagery
- Elevation
- Geodetic Control
- Governmental Unit Boundaries
- Hydrography
- Transportation

The Supplemental Guidance has established that NGDA datasets will be routinely inventoried and recommended for inclusion in the NGDA Portfolio when merited. Data inventories can provide a comprehensive view of what is available and what may be at risk in order to support acquisition priorities. Data catalogs are intended to support data discovery and sharing by end users, and may be populated by data inventories. Both inventories and catalogs may be configured to provide a comprehensive view of available data.

The initial inventory process is ongoing and is considered as a collaborative responsibility of NGDA Theme Leads, Thematic Committees, and the FGDC "NGDA Datasets, Themes and Theme Lead List" (xlsx) and is governed by the March 2014 "National Geospatial Data Asset Management Plan" (pdf). For approval by the FGDC Steering Committee as an NGDA Dataset, a geospatial dataset shall meet at least one of the following criteria:

- Used by multiple agencies or with agency partners such as State, Tribal and local governments;
- Applied to achieve Presidential priorities as expressed by OMB;
- Required to meet shared mission goals of multiple Federal agencies; or
- Expressly required by statutory mandate.

The NGDA Dataset Manager shall annually submit an NGDA Dataset Report to the relevant NGDA Theme Lead and will assist with incorporation of that information into a comprehensive annual NGDA Theme Report. National Geospatial Dataset Asset Management Plan Lifecycle Maturity Assessment Tools are also under development.

Data.gov is the most comprehensive Federal data catalog. When Data.gov was launched in May 2009, the Geospatial One-Stop portal and catalog content were migrated into a new site, eventually at the URL http://catalog.data.gov. Under the terms of the 2013 Federal Open Data Policy (pdf), newly-generated government data is required to be made available in open, machine-readable formats. Data.gov follows the Project Open Data Metadata Schema v1.1, and metadata fields are also listed in the data.gov Glossary of Terms.

The Data.gov infrastructure provides an authoritative process for identifying geospatial content of determined value. These vetted inventories are a source of enduring value for the subsequent S&A of geospatial content.

Outside of the Data.gov infrastructure, there is an ad hoc network of other inventory tools. For example, the U.S. Fish and Wildlife Service maintains the National Wetlands Inventory, with tools that allow data managers to contribute wetlands data to the wetlands geospatial data layer maintained by the Fish and Wildlife Service.

First step guidance on accessing digital materials stored on physical media can be found in the 2012 OCLC white paper "You've Got To Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media" (pdf).

While outside the scope of Federal agency concern, the GIS Inventory System maintained by the National States Geographic Information Council can be used to track the availability of data

resources within a specific geographic or thematic domain and provides an opportunity to assess the quantities of existing data, current formats, stewarding responsibility, creation date and data origin as well as the status of geographic information system implementations in state and local governments to aid planning and building of statewide spatial data infrastructures (SSDI). The Random Access Metadata for Online Nationwide Assessment (RAMONA) database is a critical component of the GIS Inventory.

The GIS Inventory automatically generates metadata that is minimally compliant with the Content Standard for Digital Geospatial Metadata published by the FGDC. It posts the metadata to a web folder that is harvested by the Geospatial Platform.

GIS Inventory metadata on digital elevation data and orthoimagery is shared with the Federal Emergency Management Agency (FEMA) for their Flood Map Modernization program, the National Digital Orthophoto Programs Committee (NDOP) and the National Digital Elevation Program Committee (NDEP). This is done as a service to the users to decrease the number of data inventories conducted by the Federal Government.

## *B. Alignment with Organizational Mission*

Proposals for the acquisition, design and development, production, operations and maintenance, or continuing stewardship of geospatial data should justify how the datasets are aligned with the mission of the organization. These proposals should describe how the proposed data are relevant to the objectives in the organizational mission statement, how the data will help to attain the long-term goals described in strategic plans, how the data will meet the needs of the designated community that the organization serves, and how the data will contribute to or complement current or planned collections to meet the mission and objectives of the organization over time. An example of a detailed agency mission statement is that of the National Oceanic and Atmospheric Administration (NOAA) Fisheries.

Mission alignment and relevance can be determined by reference to agency strategic plans. The FGDC has actively supported the development of strategic plans at the state level through its Cooperative Agreements Program (CAP). Numerous examples of completed state government geospatial strategic plans can be found at the 50 States Initiative program.

Additionally, stewarding organizations often have "collection development policies." Examples include policies from the National Geospatial Digital Archive (doc) and the Cornell University Geospatial Information Repository (CUGIR) (pdf).

## *C. Legal Rights, Restrictions and Mandates*

In addition to the legal statutes that mandate the retention of records, S&A of geospatial data should consider any limitations, restrictions or mandates that have been placed upon the data and rights or constraints for dissemination that have been specified in licenses or legal documents. Furthermore, security and confidentiality concerns (such as the existence of Personally

Identifiable Information) also may apply to protect individuals, property, wildlife, locations, or inhabitants. Legal rights, restrictions and mandates should be documented in metadata.

The jurisdiction of such laws also must be considered. Violating any limitations that have been imposed upon the data could result in criminal or civil penalties. As part of the review, evidence of rights or restrictions should be attained, examined, and retained to justify any decisions that are based on the review. Furthermore, the constraints and rights associated with the data should determine how the data are accessed, used, or distributed.

Rights or restrictions could also apply to any products or services that have been developed based on the data. This aspect of the review should also consider the entities or individuals who are authorized to grant rights for the data or materials being appraised. The length of time that is associated with such rights and limitations also should be considered and documented so that the time constraints can be included in the S&A decision process and in any plans to allow or enable future access, use, or dissemination.

## *D. Spatial Reference Information, Spatial Extent and Temporal Information*

The spatial extent of a data set can be defined in several ways. The FGDC Content Standard for Digital Geospatial Metadata (FGDC-STD-001-1998) (pdf, pg. 19) calls it "the description of the reference frame for, and the means to encode, coordinates in the data set," while the associated CSDGM Workbook (pdf, pg. 36) talks about the geographic "footprint" or "areal domain" of the data set. An S&A for extent and temporal range ascertains that the location and time periods represented by the data fit clearly into the organizational mission and under its legal mandate.

Descriptions of a data set's reference frame include physical information in terms of horizontal and vertical datum, coordinates, latitude and longitude, resolutions, or geographic or planar projections. The spatial reference information serves as a point of orientation for the data set's location and provides information about the physical measurements of the spatial framework of the data set. Accompanying information about the physical aspects of the data set should be collected for preservation, in order to provide an accurate description of the data set's geographic extent, and to establish its quality and uniqueness.

The spatial domain of a data set defines the areal extent bounding the geography of the data. The spatial domain can be described in terms of the corner coordinates of a polygonal geographic area of the data set ("bounding boxes"), or by various descriptions of geographies of scale, such as states, countries or continents. The spatial domain of data sets can be used in the preservation selection process to determine the extent of coverage and/or the overlap of data, to ensure completeness of the data collection.

Temporal range is described in the CSDGM Workbook (pdf, pgs. 35-36) as the time period when the data was collected. In certain cases, the time period refers only to the publication date of the data set, or in other cases, it may be unknown. Reporting of the temporal range for a data set can vary from the most detailed information, including dates in terms of measured periods (calendar,

single date), or as measured time (time of day, first hour, minutes), to generalized descriptions (multiple years, range of dates, event). It is important to document the time the data set was created. In addition, recorded information should include the frequency of changes or additions to the data set following its initial completion.

## *E. Current Scientific or Cultural Heritage Value*

The current *scientific* value of data is based on the concept that data are used to communicate the results of research studies and are required in order to continue research, create new science, or augment current research in other disciplines.

The current *cultural heritage* value of data is the importance of any particular set of digital information as determined by the aggregate of values attributed to it. According to the Assessing the Values of Cultural Heritage research report (pdf) from the Getty Conservation Institute, Los Angeles, the values considered in this process should include those held by experts (historians, archaeologists, architects, and others) and those brought forth by new stakeholders or constituents.

Organizations other than the originator of the data could achieve and share potentially significant benefits by stewarding and providing access to records and data. In a current example, historic records created by a federal agency that contain information on Marcellus Shale are of significant contemporary interest and use to Federal, private and non-profit sectors. The current availability of this historic data provides key inputs in determining policy and community responses.

Not all data is of equal value and is worthy of long-term preservation. The authorities best positioned to provide an appraisal of the value of current data are the creators or responsible agencies of the data. There are several criteria that responsible agencies can consider as they make early assessments of the value of current data:

a) Identification as potentially valuable by project lead or lead scientist

The initial responsibility for identifying data with current scientific or historic value comes from the project lead or lead scientist, acting on behalf of the responsible agency. All project leaders must take into consideration the potential value of project data, including those outside of the discipline of the current project focus. The opposite is also true; not all data has long-term value. Project leaders must attempt a reasonable valuation of their work at creation, and make good-faith efforts to provide supporting materials to any internal groups tasked with stewarding the materials so that those groups can reassess value as necessary.

b) Association with a scientific report or publication

Data that is referenced in a scientific publication or technical report should be considered of current and historic benefit to science.

It is now possible to store complete data sets and create Digital Object Identifiers (DOIs), Archival Resource Keys (ARKs) or other resources to point to data located in the networked environment. DOIs are persistent links that are represented by unique alphanumeric strings assigned by a registration agency. ARKs are URLs that support long-term access to information objects and are not permanent. ARKs can be modified or updated as needed. DOIs and ARKs make it possible to access information on an ongoing basis to continually reap its value.

c) Association with a federally funded project with distribution/preservation requirements

Many federal agencies require projects that include a data component to include a data management plan. These plans often include a section that addresses the long-term preservation of the project data after project completion. Final project data, especially data that is highlighted or included in a final report or scientific publication, should be considered to have current or historic value and thus preserved. Preservation of source data allows current and future researchers to access the data and compare results or augment ongoing research.

d) Association with a natural disaster, current event, or other significant occurrence

Data associated with ongoing or current events of social significance should be preserved for ongoing and future research. For example, data related to Hurricane Katrina that was gathered by government agencies and researchers in real time can be used not only to assess the event but also to prepare for and potentially predict the next major event. In the book "Preservation in Digital Cartography: Archiving Aspects," a FEMA official stated that, "If we do not preserve this data and use it for research purposes, then we have wasted time and energy and done a great disservice to those who will be affected by the next major hurricane"

Questions about cost effectiveness of preserving data should be answered in part by both the returns on the original investment as well as the potential costs of redeveloping or acquiring the data.

## *F. Technology and Obsolescence Risks*

In order to adequately preserve geospatial data, especially in digital form, proactive steps must be taken to prevent or mitigate the effects of technology obsolescence. Determining when to put these steps into action can be difficult, but a definition from the 2012 Reference Model for an Open Archival Information System (OAIS) Magenta Book (pdf) suggests that "long-term" stewardship may best be affected by dividing actions into shorter, punctuated durations with more regular monitoring. The report defines "long-term" as:

"A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future."

The LC21 report from 2000 suggests that digital materials become unreadable and inaccessible if the playback devices necessary to retrieve information from the media become obsolete or if the software that translates digital information from machine- to human-readable form is no longer available.

The definitions above do not provide specific time periods for when media should be refreshed, but they do make it clear that stewards must monitor continual advances in computer hardware, software, firmware, and storage media. While life cycles vary widely depending on the materials under consideration, stewards should evaluate their technology refreshment lifecycle within five-year periods. Further information on the technology refreshment lifecycle can be found in the 2001 publication, Technology Refreshment Within DoD (pdf)

Software migration is often tied to operating system evolutions. When software is migrated, a key point to investigate is whether or not the new software can read media containing geospatial data created under previous versions. If not, a migration of the legacy geospatial data may be required.

In the context of technology obsolescence, "data at risk" are data that is not in a format that permits full electronic access. Such data may be inherently non-digital (e.g. handwritten or photographic), on near-obsolete digital media (such as magnetic tapes), or insufficiently described (lacking meta-data). Some digital data can also be considered "at risk" if they cannot be ingested into managed databases because they lack adequate formatting or metadata. Data that are regarded as unusable tend to be regarded as useless, and thus risk being destroyed.

The concept of format sustainability also comes into play when addressing technology and obsolescence risks. Formats that are sustainable are accessible both throughout their lifecycle and as technology evolves. A sustainable format is one that increases the likelihood of a record being accessible in the future. Both NARA, with its "Frequently Asked Questions (FAQs) About Selecting Sustainable Formats for Electronic Records" and the Library of Congress, with its "Sustainability of Digital Formats Planning for Library of Congress Collections" site have addressed criteria for selecting formats based on their sustainability.

In considering the suitability of particular digital formats for the purposes of preserving digital information as an authentic resource for future generations, it is useful to articulate important factors that affect choices. The Library of Congress does this by listing and defining seven "sustainability factors," each of which is explored in greater detail on the site:

- Disclosure
- Adoption
- Transparency
- Self-documentation
- External dependencies
- Impact of patents
- Technical protection mechanisms

The 2000 publication, Risk Management of Digital Information: A File Format Investigation, provides an excellent introduction to the various risks facing digital information. Table 1 on pg. 7 of that publication epitomizes the risks facing the appraiser of digital content for long-term value (the list below has been edited to represent the chiefly technology-oriented risks):

- Content fixity (bit configuration, including bit stream, form, and structure)
  - Bits/bit streams are corrupted by software bugs or mishandling of storage media, mechanical failure of devices, etc.
  - File format is accompanied by new compression that alters the bit configuration.
  - File header information does not migrate or is partially or incorrectly migrated.
  - Image quality (e.g., resolution, dynamic range, color spaces) is affected by alterations to the bit configuration.
  - New file format specifications change byte order.
- Security
  - Format migration affects watermark, digital stamp, or other cryptographic techniques for "fixity."
- Context and integrity (the relationship and interaction with other files or other elements of the digital environment, including hardware/software dependencies)
  - Reading and processing the new file format require a new configuration because of different hardware and software dependencies.
  - Linkages to other files (e.g., metadata files, scripts, derivatives such as marked-up or text versions or on-the-fly conversion programs) are altered during migration.
  - New file format reduces the file size (because of file format organization or new compression) and causes denser storage and potential directory-structuring problems if one tries to consolidate files to use extra storage space.
  - Media become denser, affecting labels and file structuring. (This might also be caused by file organization protocols of the new storage medium or operating system.)
- References (the ability to locate data definitively and reliably over time among other digital objects)
  - File extensions change because of file format upgrade and its effect on URLs.
  - Migration activity is not well documented, causing provenance information to be incomplete or inaccurate (a potential problem for future migration activities).
- Functionality
  - Features introduced by the new file format may affect derivative creation, such as printing.
  - If the master copy is used for access, changes may cause decreased or increased functionality and require interface modifications (for example, static vs. multi-resolution image or inability of the Web to support the new format).
  - Unique features that are not supported in other file formats may be lost (for example, the progressive display functionality when Graphics Interchange Format [GIF] files are migrated to another format).
  - The artifactual value (original use context) may be lost because of changes introduced during migration; as a result, the "experience" may not be preserved.

NARA released its "Revised Format Guidance for the Transfer of Permanent Electronic Records" in April 2014 that greatly expands the number of digital formats they accept for transfer. The transfer guidance format tables are organized by categories of electronic records, and for each category the tables identify *preferred*, *acceptable*, and in some cases, *acceptable for imminent transfer* formats. Many file formats, especially those used with digital audio, video and geospatial information, are composed of multiple parts that might include multiple embedded encoding streams or codecs and another *wrapping* component. In these cases, the format category table includes a column that specifies the codec or codecs that may be used with each format. Agencies must submit electronic records in files that are valid according to both the wrapper and any specified codec standards. (Further information is available in Section 5, "Geospatial formats," of the transfer guidance format tables).

In addition to the above, data compression is a technology risk that comes into play in the appraisal process. Generally speaking, an archive will want to preserve the highest resolution form of any particular digital file and compressed data has the potential to be at greater risk. The UK Joint Information Systems Committee has an excellent series of web pages on "File Formats and Compression" that looks at the theory of file formats and the common methods of data compression.

Data volume is an ever-increasing technology-oriented risk. The consulting firm IDC has published a series of reports on data volume and its challenges. The chief appraisal risk that comes with volume is that the organization may not have the capacity to archive all the digital information for which they have responsibility. Depending on the type of data under consideration, it may be possible to take representative samples of the data to get a reasonable degree of coverage. For example, if a data set changes infrequently it may not be necessary to take daily samples.

Finally, some data-at-risk may be inherently non-digital (e.g. handwritten or photographic), on near-obsolete digital media (such as magnetic tapes) or insufficiently described (lacking meta-data). Some born-digital data may also be considered "at risk" if they cannot be ingested into managed databases because they lack adequate formatting or metadata. Data that are regarded as unusable tend to be regarded as useless, and then risk being destroyed.

While outside the scope of this S&A document, stewards should be aware of the ability to emulate computing environments in order to gain access to otherwise inaccessible resources. A computer emulator is a hardware or software tool (or both) that mimics (emulates) the functions and environment of one computer operating system in another computer system so that the emulated behavior closely resembles the behavior of the original system. The use of emulation allows the user to view a software environment in a close approximation of the original experience, thus providing a realistic replica information representation in the absence of the original hardware and software.

## *G. Cost-Benefit Analysis*

Costs associated with archiving electronic geospatial records can be a considerable obstacle for agencies and institutions, especially those with moderate and limited budgets. Therefore, geospatial records considered for ingestion into long-term or permanent archives may be subject to a cost-benefit analysis as one component of an overall records S&A . When appraising existing collections, institutional policy determines the specific nature of data to be acquired and identifies any gaps in the collections that require filling. Ensuring that repositories have the right to reject data sets that fall outside their scope of collecting can help avoid acquiring data that may be too costly to maintain, both financially and in terms of staff resources.

Some economic characteristics of records management that may be considered for S&A include:

- The sponsoring program or funding associated with acquiring, preserving, and making the records accessible;
- The identification of cost-sharing opportunities for capital investment and/or recurring expenses. Upkeep of hardware and equipment and ensuring that appropriate security measures are in place should be factored into the overall costs of the long-term preservation of digital data;
- An estimate of the expense to reproduce the collection and how the scientific, operational or secondary value of the collection exceeds the costs to preserve and make the records accessible;
- The approximate costs of identifying, appraising, accessioning and processing the collection to make it accessible;
- The identification of the resources needed for required preservation functions;
- The approximate annual costs of housing the original records. If sampling is appropriate, is there a significant cost savings?
- The identification of special equipment required to read or process the records;
- An estimate of the cost to de-accession, purge or dispose of the collection;
- An estimate of the Non-Replicability (replacement cost) of the candidate resource. Is it feasible or excessively costly or prohibitive to replicate the data or record?

In addition to the S&A of records resulting from processed data, data sets are candidates for long-term preservation if there is no realistic chance of repeating the experiment, or if the cost and intellectual effort required to collect and validate the data are so great that long-term retention is clearly justified. Funding streams for data-generating activities may wish to build-in adequate resources from the start to support end-to-end data management, including long-term stewardship if required, while understanding that the costs of capturing and storing data can and will fluctuate over time.

The engagement of funding bodies to consider the inclusion of data curation costs in the financing of any scientific project producing digital data is one possible solution that has been pursued by stewarding organizations. However, since some data are used again and again while other data are never accessed, linking long-term value to initial funding could be problematic and it may be more cost-effective to regenerate certain kinds of environmental data on demand. The "trust" engendered in curated digital data helps secure maximum economic and social benefits from public investments in the preservation of scientific data through a chain of custody and authority. Trust is essential for encouraging the reuse of data.

S&A decisions should take into account that the uses of data vary according to the level of processing. Processed records are more likely to have long-term value if they would be costly to recreate from the raw data. It may be warranted to appraise both a raw version and one or more processed versions of certain records. With each higher level of processing, records generally become easier to use but less subject to reanalysis. To facilitate future reanalysis, it is usually appropriate to preserve processed records at the lowest level of processing compatible with effective use.

Since S&A criteria are specifically designed to determine if a record is permanent based on the quality of a record's content and its context, cost considerations should not overwhelm the decision making process. NOAA recommends that the cost of long-term maintenance should be considered "only after all appraisal criteria are met," and NARA considers costs only in marginal cases in its "Strategic Directions: Appraisal Policy" and advises:

"… [Cost] should play a significant role only in marginal cases. In such cases, an appraisal should balance the anticipated research potential of the records with the resource implications of retaining them permanently. Other things being equal, records with low long-term cost implications are more likely to warrant permanent retention than those records that carry high long-term costs."

Therefore, a record or collection of records that is appraised as having permanent value is impacted by cost considerations only when the appraised value is questionable. When the record is clearly appraised as permanent, cost considerations should not override that S&A assessment.

Occasionally a re-appraisal of archived geospatial records and data to remove less significant collections is warranted. Prime candidates for re-appraisal include data that are obsolete or redundant, that could be regenerated on demand, or clearly have only short-term uses. This includes older versions of reprocessed data and model output. When re-appraisal does occur, disposing of records does not automatically mean destroying, which is considered to be a last resort.

At the USGS, the EROS scientific records appraisal process recommends that records are to be retained or disposed based upon many factors including mission alignment, accessibility costs, and projected science utility. Collections to be disposed are advertised using the CEOS Purge Alert bulletin board system, with the goal of finding responsible organizations whose missions match the type of records being disposed, and who are willing to take on the collection. Archiving and access decisions are closely related. In general, when resources are limited, access to older or less commonly used data should be scaled back rather than removing data from the archive.

## *H. Tangible Media and Physical Condition*

Tangible media, often called "physical media," is the generic name for external digital storage media, including 8, 5.25 and 3.5 inch "floppy" discs, CD-ROMs, digital video, Blu-ray and other optical discs, memory cards, USB "flash" drives and external hard drives.

These devices may contain important digital files but should first be appraised in their physical form. These items present an elevated preservation risk, in that the tangible media itself is fragile and that fragility endangers the digital materials housed on it.

Detailed guidance on managing digital materials stored on physical media in preparation for transfer is found in the 2012 OCLC white paper "You've Got To Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media" (pdf). Appraisal should include these steps:

- Count and describe all identified media. Retain the order (if one exists) of the original digital media and accompanying items.
- Count the number of each media type, indicate the maximum capacity of each media type, calculate the total maximum amount of data stored in each medium, and then calculate the overall total for the collection. This will enable you to estimate storage needs, though keep in mind that the media are rarely full, so the estimate will likely be far in excess of the actual storage needed.
- Detail the physical condition and overall quality of the tangible media.
- Record anything that is known about the hardware, operating systems, and software used to create the files. Leverage associated documentation if it exists.
- Prioritize appraisal decisions for the tangible media collection by estimating the value, importance, and needs of the collection as a whole, the level of use (or anticipated use) of the collection and potential danger of loss of content because of potential media degradation due to age or condition.

## *I. Metadata Quality, Completeness and Usability*

Metadata is critical to S&A. It comprises administrative, descriptive, preservation, rights management, structural and technical information that provides context to data and helps users comprehend and understand it. It addresses several S&A elements already outlined in this guidance document: legal rights, restrictions, and mandates; spatial reference information, spatial extent, and temporal information; source/lineage; and data and media format.

The report "Utilizing Geospatial Metadata to Support Data Preservation Practices" (pdf) from the GeoMAPP project describes the two primary geospatial metadata standards utilized by the large majority of practitioners: the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) - FGDC-STD-001-1998 and the International Organization for Standardization (ISO) - 19115:2003 Standard for Geographic Information Metadata. The report offers a checklist of important CSDGM fields that facilitate long-term preservation of the geospatial datasets, though individual agencies will need to develop their own metrics on metadata completeness. Though not cited by GeoMAPP, the current version of the ISO standard is "ISO 19115-1:2014, Geographic information - Metadata - Part 1:

Fundamentals," which revises ISO 19115:2003. Recent government open data initiatives also impose non-geospatial metadata requirements on agencies, such as the use of Project Open Data metadata to list agency datasets and application programming interfaces.

Legacy geospatial data often need additional work to provide reasonable or useable metadata files. Beyond a particular dataset's conformance to metadata standards, it is useful to include additional information that was created along with the geospatial data. Items such as libraries of documentation, guides, data information files, fact sheets, FAQs, instrument documentation, design reviews, lessons learned, hardware documentation, engineering models, computer models, platform documentation, algorithm documentation, URLs, principle investigator contacts and algorithm theoretical basis documents may be included. All of these represent valuable information about the data, and the more that are available, the better.

## *J. Uniqueness*

Uniqueness, as defined in the NOAA Procedure for Scientific Records Appraisal and Archive Approval: Guide For Data Managers, describes data that is the only or sole example of its type.

The NARA Strategic Directions: Appraisal Policy (excerpted from the internal NARA Directive 1441) from 2007 states that appraisals shall be conducted in context with other records. The S&A shall determine whether the records under consideration are the only or most complete source for significant information. Records that contain information not available in other records (including other Federal records and files accumulated by state and local governments) are more likely to warrant permanent retention than records containing data that is duplicated in other sources.

The 2007 Environmental Data Management at NOAA : Archiving, Stewardship, and Access report from the National Academy Press offers guidance on uniqueness from the opposite direction, stating that the most obvious candidates for *reduced* archiving requirements are data that are obsolete or redundant, that could be regenerated on demand, or clearly have only short-term uses.

The 1995 report Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources from the National Academy Press states the value of uniqueness as an imperative for preserving data on the physical universe:

"Many observations about the natural world are a record of events that will never be repeated exactly. Examples include observations of an atmospheric storm, a deep ocean current, a volcanic eruption, and the energy emitted by a supernova. Once lost, such records can never be replaced."

The How to Appraise and Select Research Data for Curation document from the UK Digital Curation Centre defines uniqueness as:

> "The extent to which the resource is the only or most complete source of the information that can be derived from it, and whether it is at risk of loss if not accepted, or may be preserved elsewhere."

It poses these questions regarding S&A for uniqueness:

- Is the dataset the only source of its content and will it be preserved elsewhere?
- Does the dataset duplicate existing work?
- Do other copies of the data exist that are accessible and useable?
- If other copies exist, where is the most comprehensive or up-to-date version?
- Are any other copies at risk of loss? And if so, will they be preserved by their holding organization?

A related concept to uniqueness is *intrinsic value*. In the NARA publication cited above, the agency provides a definition of intrinsic value:

> "Records with intrinsic value are rare and possess one or more specific qualities or characteristics as defined by NARA. These include but are not limited to records in an original form that document an early media type (e.g., glass plate negatives, wax cylinder recordings, etc. – Note that only a representative sample would have intrinsic value and not the entire collection), aesthetic or artistic quality (e.g., manuscripts; photographs; pencil, ink, or watercolor sketches; maps, etc.), age (e.g., generally, records of earlier date are of more significance than records of later date)."

## *K. Provenance*

Provenance is an understanding of the context from which a set of geospatial data was created Data provenance documents the inputs, entities, systems, and processes that influence data of interest, in effect providing a historical record of the data and its origins. Captured provenance information helps shed light on the original creation purpose of data and the history of organizational control of data over time. Provenance information can provide significant assistance in determining long-term ownership or engagement with any particular set of data.

Provenance is a fundamental principle of archives and has two components: (1) records of the same provenance should not be mixed with those of a different provenance and (2) stewards should maintain the original order in which the records were created and kept to the greatest extent possible. The significance of archival materials is heavily dependent on the context of their creation, and the arrangement and description of these materials should be directly related to their original purpose and function.

In appraising for provenance, stewards should examine the degree to which contextual information about the origin and ownership of the data in question is available. Provenance information should be documented in metadata.

## *L. Future Value Determination*

A challenging S&A point is the determination of the scientific or public policy value of records to be archived in terms of anticipated future benefits or secondary uses geospatial data.—and the levels of service required to achieve these benefits. The NOAA Procedure for Scientific Records Appraisal and Archive Approval: Guide For Data Managers (pdf) explores these types of valuations and provides pointers to possible guidelines.

The document Appraisal Policy of the National Archives and Records Administration (pdf) is cited in the NOAA document mentioned above. It introduces the concept that data may have value to the agency, the Government, or to the public for unanticipated uses long after they have served their original purpose. The document suggests that future research potential of records is the most difficult variable to determine. What is of relatively low research use today may become of great research use in the future. The most challenging variables to predict are the issues and topics that will be considered of significance in the future. It is necessary to consider the kinds and extent of current research use and make inferences about anticipated use both by the public and by the Government.

The 2010 document, How to Appraise and Select Research Data for Curation, by Angus Whyte of the UK Digital Curation Centre and Andrew Wilson of the Queensland (Australia) State Archives offers a series of questions to ask in relation to the determination of scientific or historical value:

- Is the data scientifically, socially, or culturally significant? Assessing this involves inferring anticipated future use, from evidence of current research and educational value.
- Does the dataset reflect the interests of contemporary society?
- Is the set the only source of its content and will it be preserved elsewhere? Does the data support trends in research awards by national funding bodies, and based on criteria such as the number of projects funded or the amount provided for the relevant research topic?

In addition to future value, there is potential informational value and secondary use of archived geospatial data. While these uses are unknown today, the ability to provide usable, historic geospatial data should be recognized as benefit to the community. Secondary users may interpret, assess and evaluate the data in new and different ways.

For government agencies considering secondary uses in their S&A decisions, a key component is to collect and provide information about the archived data. The National Oceanographic Data Center (NODC), the designated archive center for oceanographic data in the U.S., requires that data be documented to enable secondary use and ensure data posterity. The NODC collects and provides access documentation or metadata pertinent to digital data in the archives.

One area of promise in determining value and secondary use is analysis of citations to publications the data has been used in, or to other authoritative sources such as research assessments. It may be possible to apply value to data retained as part of the research record by considering the findings based on them. For example, the 2003 Bridging data lifecycles: Tracking data use via data citations workshop report identified:

"…a number of common themes, ranging from conceptual debates about data publication to the practical challenges of tracking data use. Data citation initiatives are often tied to the idea that data sets should be published just like other kinds of scholarly products. The idea of publishing data sets, however, becomes problematic when looking at the similarities and differences between traditional scholarly publications and digital data sets."

The document, "Selection and Appraisal of SEDAC Resources for Accession into the SEDAC Long-Term Archive," from the Center for International Earth Science Information Network (CIESIN), Columbia University gives the following guidance:

"Scientific or Historical Value: Verify the scientific or historical value of the candidate resource by examining current evidence of citation, research, and educational use as published in refereed scientific publications or reports received from a recognized committee of scientists representing the discipline of the data."

"Potential Usability and Use: Present evidence of usability, usefulness, and sufficient usage of the resource by the community of users interested in human dimensions of the environment. Adequate evidence should be presented to indicate whether the potential for future use of the resource justifies the costs of long-term archiving."

At a further extreme, the 2007 National Research Council report, Environmental Data Management at NOAA: Archiving, Stewardship, and Access, notes that not all data sets are of equal value and observes that practical constraints prevent organizations from archiving all data. The report suggests that it is extremely difficult to assess the current value of any particular environmental data stream and virtually impossible to anticipate its potential future uses.

The solution is a decision-making process that is iterative and ongoing, with data managers and stewards continually reviewing the data holdings under their purview to determine the appropriate level of service for each data set, given legal and mission requirements, user needs, cost-effectiveness, and available resources.

Data managers should try to envision the needs of the future when making a decision regarding archiving a dataset. It may be useful to research and document the current uses of the data in creating a rationale for preservation. However, this is only a part of the picture, and a sense of vision and imagination may be required in order to make the correct decision.

# Appendices

## *Appendix 1: References*

The Geopreservation.org website offers a rich selection of freely available web-based resources about the preservation and stewardship of geospatial information. Topics include appraisal and selection; citation; content standards; geographic information systems; preservation formats; satellite imagery; software dependencies; virtual environments; and many others.

The Geospatial Data Preservation Resource Center is a project of the National Digital Information Infrastructure and Preservation Program (NDIIPP) at the Library of Congress, which is working with a national network of partners on a strategy for preserving digital information for use in the future.

## *Appendix 2: Geospatial Data as Federal Records Subject to Management Requirements*

Geospatial data may be selected for long-term preservation solely for its value, but there are also purely statutory reasons to steward geospatial data based on government archival and records management processes and legal requirements. This section outlines some of the authorities to take under consideration when making S&A decisions on any particular set of geospatial content.

The ISO 15489-1: 2001 standard defines records management as "[the] field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including the processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records." This international standard serves the geospatial world well by providing principles that can be applied in the evaluation of geospatial data.

ISO 15489-1:2001 defines records as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business." This definition easily applies to geospatial data.

While there are many purposes of and benefits to records management, a key feature of records is their ability to serve as evidence of an event. *Authenticity*, *reliability*, *integrity* and *usability* are aspects of evidence, and each aspect contributes to the overall quality of the electronic records as evidence of an activity.

An *authentic* record is one that can be proven

- To be what it purports to be,
- To have been created or sent by the person [system] purported to have created or sent it, and

- To have been created or sent at the time purported.

To ensure the authenticity of records, organizations should implement and document policies and procedures that control creation, receipt, transmission, maintenance and disposition of records to ensure that records creators are authorized and identified and that records are protected against unauthorized addition, deletion, alteration, use and concealment.

A *reliable* record is one whose contents can be trusted as a full and accurate representation of the transactions, activities or facts to which they attest and can be depended upon in the course of subsequent transactions or activities. Records should be created at the time of the transaction or incident to which they relate, or soon afterwards, by individuals who have direct knowledge of the facts or by instruments routinely used within the business to conduct the transaction.

The *integrity* of a record refers to its being complete and unaltered. A record should be protected against unauthorized alteration. Records management policies and procedures should specify what additions or annotations may be made to a record after it is created, under what circumstances additions or annotations may be authorized, and who is authorized to make them. Any authorized annotation, addition or deletion to a record should be explicitly indicated and traceable.

A *usable* record is one that can be located, retrieved, presented and interpreted. It should be capable of subsequent presentation as directly connected to the business activity or transaction that produced it. The contextual linkages of records should carry the information needed for an understanding of the transactions that created and used them. It should be possible to identify a record within the context of broader business activities and functions. The links between records that document a sequence of activities should be maintained.

The 2003 FGDC document, "Managing Historical Geospatial Data Records" (pdf) provides a brief overview of records management responsibilities as they relate to geospatial records, whether digital or non-digital.

The "Strategic Directions: Appraisal Policy" document sets out the strategic framework, objectives, and guidelines that NARA uses to determine whether Federal records have archival value. The high-level criteria for the permanence of geospatial data would fall generally under Section 7.3, "Records documenting the national experience" and more specifically under the "Observational Data in the Physical Sciences" section of Appendix 2, "Special Considerations for Selected Types of Records." They would also be covered to a lesser degree under Appendix 2's "Environmental Health and Safety Records" and "Research and Development (R&D) Records" sections.

There are a number of sections of the U.S. Code and the Code of Federal Regulations that deal with data dissemination and preservation that provide the rationale for Federal Agency S&A decisions. Title 44 of the U.S. Code deals with Public Printing and Documents and includes chapters covering records management, disposal and agency coordination.

- Section 2107 on the "Acceptance of records for historical preservation" gives the Archivist of the U.S. the authority to "accept for deposit with the National Archives of the United States the records of a Federal agency, the Congress, the Architect of the Capitol, or the Supreme Court determined by the Archivist of the United States to have sufficient historical or other value to warrant their continued preservation by the United States Government."

- "Chapter 29-Records Management by the Archivist of the United States and by the Administrator of General Services" states that Archivist will provide guidance and assistance to Federal agencies to ensure that policies and transactions of the Agency are documented.

- "Chapter 31-Records Management by Federal Agencies" states that each Federal agency will make and preserve records that document the organization and its functions, policies, decisions, procedures, and transactions.

- "Chapter 33-Disposal of Records" covers the lists and schedules of records that the heads of each Government agency shall submit to the Archivist.

- "Chapter 35-Coordination of Federal Information Policy" covers the ways that agencies can: minimize paperwork burdens on people and businesses; ensure the public benefit from and use of Agency information; coordinate, integrate and develop information management policies and practices to improve delivery of services to the public; improve the quality and use of agency information for decision making; minimize costs of creating, collecting, maintaining, use, dissemination and disposition of information; strengthen partnerships between Federal, State, local and Tribal governments; provide for the dissemination of public information; and ensure that the creation, collection, maintenance, use, dissemination and disposition of information is consistent with law.

- Title 36 of the CFR deals with "Parks, Forests, and Public Property." Part 1235 of Title 36 deals with "Transfer of Records to the National Archives of the United States" with authority situated in sections 2107 and 2108 of the U.S. Code. Part 1235.50, "What specifications and standards for transfer apply to electronic records?" covers the general guidelines for the transfer of electronic records, while Part 1235.48 covers the documentation required to be included with an electronic records transfer.

## *Appendix 3: Example Model on Establishing a Selection and Appraisal Process*

Model: USGS / EROS Scientific Records Appraisal Process

- USGS Program Coordinator, Project Manager, or outside entity proposes to the EROS Archivist a collection for review.

- Appraisal Team is assembled that includes:

    - Science Staff
    - Project Manager
    - Archivist

- Archivist documents what is known about the collection by utilizing a question set.
    - http://eros.usgs.gov/government/RAT/tool.php

- Science team members review the documentation and provide their comments and opinions. At a minimum, the three questions below should be addressed:
    - 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?

- Archivist briefs the relevant Project Manager.

- Archivist sends recommendation memo to Center Senior Staff for review.
    - Archivist memo recommends, with justification:
        - Retain / Accept or
        - Dispose / Reject

- The Center Senior Staff pass their comments to the EROS Director.

- EROS Director accepts, rejects, or modifies the recommendation.
    - EROS Director informs Archivist and Project Manager of his decision via memo.
    - Purge recommendations result in a search for a new home. Destruction is the last resort.