The U.S. Government Role in Preserving Geoscience Sample and Data Collections

Ester Sztein, PhD
National Academies of Sciences, Engineering, and Medicine

American Geophysical Union, 2016

This presentation represents the author’s views and not necessarily those of the National Academies of Sciences, Engineering, and Medicine
“Sets of physical objects, living or inanimate, and their supporting records and documentation, which are used in science and resource management and serve as long-term research assets that are preserved, catalogued, and managed by or supported by Federal agencies for research, resource management, education, and other uses.”

OSTP March 2014 memo
Why preserve physical collections?

- Vouchers from earlier critical findings
- Standards
- Repositories for rare samples too expensive or impossible to recollect
- Samples for future analysis/experimental use
- Scientific collections for purposes originally unforeseen
- Old samples can be reprocessed with new technology: new knowledge (reusability)
Preserving a collection?

Geoscience Data and Collections: National Resources in Peril (NRC, 2002)
A good physical collection...

- Well documented - metadata attached
  - field number
  - geographic location,
  - collector,
  - date collected,
  - sample type,
  - reason for collection,
  - project name,
  - other important data

- Analysis and derivative samples

- Well preserved and curated
Why the USG?

- Federally-supported science collections are public assets
- Support mission-critical science
- Excellent ROI for U.S. taxpayers
- Historic bipartisan interest

**BUT:** No one-size-fits-all approach

- Each Agency develops its own policies and procedures to ensure access, storage, transfer, and disposal of geologic samples and data
Important reports: NRC 2002, IWGSC 2009
Interagency Working Group on Scientific Collections (IWGSC)

- In 2006 by White House National Science and Technology Council: focus attention and planning for Federal/Federally-funded collections management
- Run by OSTP; USDA & Smithsonian co-chairs
- 15 USG Agencies with scientific collections and/or granting programs
- Most GEO collections are in 8 Agencies
- 21% of their collections are geoscience
- 2009 Green Book includes 7 recommendations
IWGSC’s first 3 recommendations

- **Budgeting for Collections**
  - Collect/share budgeting information
  - Assess/project costs

- **Policies and Best Practices**
  - Identify/disseminate: organization, management, access (physical and online), long-term preservation

- **Data and Metadata Accessibility**
  - Document physical objects; make collection information available online
  - Develop online clearinghouse for information on contents and access to Federal scientific collections
OSTP 2014 memo asks each agency to develop plans to manage their physical scientific collections

“To improve management of and access to scientific collections”

“An essential base for developing scientific evidence and ... resource for scientific research, education, and resource management”

End goal: “systematic improvement of the development, management, accessibility, and preservation of scientific collections owned and/or funded by Federal agencies”

Only for long-term institutional, archival collections, not for short-term, project collections
2014 Memo on Scientific Collections

Each Agency’s policy to include:

- Role to advance its mission
- Legislative/regulatory requirements
- Who has responsibility to carry out policies
- How to project costs of developing, preserving, managing scientific collections
- Agency requirements and standards for long-term preservation, maintenance, accessibility for public use
- Strategy to provide online information about physical collection contents and access to objects and digital files, unless limited by law or to protect national interests
- Process to de-access, transfer, dispose of collections
- Resources within each Agency to implement policy
- Consistency with 2013 Open, Machine-Readable Data OSTP memo
- Agencies to work together, coordinate through IWG SC
Curated source of information about object-based science collections owned or managed by US Federal departments and Agencies (usfsc.grscicoll.org)

Supported by IWG SC, managed by SI
Of the 15 USG Agencies, 8 have significant geoscience collections:

- **National Aeronautics and Space Administration**
  - NASA Policies
- **National Science Foundation**
  - Policies: In Development
- **Smithsonian Institution**
  - SI Collections Management Policy; SI Reports and Other Policies; SI Collections Space Framework Plan
- **U.S. Department of Agriculture**
  - USDA ARS Collections Policy
  - USDA ARS Reports
  - FSIS Policies: In Development
  - APHIS, NIFA, NRCS, USFS Policies: Expected December 2016
- **U.S. Department of Commerce (NIST, NOAA)**
  - NIST Policies: Expected December 2016
  - NOAA Policies: In Development
- **U.S. Department of Energy**
  - Policies & Reports: Awaiting content submission
- **U.S. Department of Interior (BIA, BLM, NPS, Reclamation, FWS, USGS)**
  - Interior-wide Museum Collection Policies and Program Reports
- **U.S. Environmental Protection Agency**
  - EPA Scientific Collections Policy
The U.S. Geological Survey Geologic Collections Management System (GCMS)

A Master Catalog and Collections Management Plan for U.S. Geological Survey Geologic Samples and Sample Collections

Circular 1410
U.S. Department of the Interior
U.S. Geological Survey

http://pubs.usgs.gov/circ/1410/
USGS Geologic Collections Management System (GCMS)

- Virtual network connecting USGS sample collections housed throughout the US to be used by US and foreign scientists.
- Does NOT specifically address the preservation of digital data (seismic, magnetic, electric) or library materials.
- ScienceBase (cyberinfrastructure for samples and data) will integrate GCMS and geologic data repositories.
ScienceBase is a cyberinfrastructure developed by the U.S. Geological Survey (USGS) that organizes, catalogs, and integrates scientific data for USGS and partner programs and projects (https://www.sciencebase.gov/catalog/). As a part of ScienceBase, data collected by the Geologic Collections Management System will be accessible in context with complementary USGS geoscience data.
Some definitions

- **Active collection (aka Working Collection)**: Contains material from ongoing research and is actively used by the project scientists. New samples are added as research continues. When completed, materials will be evaluated for permanent retention in a GCMS repository.

- **Legacy Collection**: Contains samples from research scientists who are no longer with the USGS. With proper documentation, these collections will be incorporated into the GCMS and treated as resource collections.

- **Orphan Collection**: Consists of poorly documented samples with little foreseeable research value and do not warrant inclusion in the GCMS. If still have education/outreach value, may be transferred to another entity.

- **Reference Collection**: Contains samples of a distinct nature that provide an objective standard against which other samples are compared. This type of collection may be augmented with new samples.

- **Resource Collection**: Contains materials from completed research that remain significant as research assets and are made available for current and future research and preserved for an indefinite period of time.
Universally Unique Identifiers (UUID)

Sample identification needs to be standardized. As each active repository contributes the metadata for its sample inventories through the GCMS to the centralized registry, each sample will be assigned an UUID. International Geo Sample Number (IGSN): Internationally-recognized ID system.

http://www.geosamples.org/
Think early and often...

about your data management plan

AND

your physical collection management plan
4-Point Standard

1. What?
   - Do the samples have IDs, or can they be assigned IDs?

2. Where?
   - Do the samples come with locality information, or can it be recovered?

3. Who?
   - Do you know who collected the samples?

4. When?
   - Do you know when the samples were collected?

Develop descriptive information if not already available.

GCMS Collection Determination Process

- Do the samples have iconic, historic, or scientific value?
  - Yes
  - These samples meet the requirements for the USGS Geologic Collections Management System and should be retained by the USGS or transferred to the Smithsonian Institution’s National Museum of Natural History.
  - No
  - Are some samples useful to someone else within the USGS or another Federal or State agency?
    - Yes
    - Transfer to the individual, with appropriate documentation.
    - No
    - Can the samples provide value to an educational institution?
      - Yes
      - Transfer to the educational institution, with appropriate documentation.
      - No
      - Do the samples serve any outreach utility?
        - Yes
        - Provide to outreach function, with appropriate documentation.
        - No
        - DISCARD SAMPLES, with appropriate documentation.

- No
  - Are some samples useful to someone else within the USGS or another Federal or State agency?
    - Yes
    - Transfer to the individual, with appropriate documentation.
    - No
    - Can the samples provide value to an educational institution?
      - Yes
      - Transfer to the educational institution, with appropriate documentation.
      - No
      - Do the samples serve any outreach utility?
        - Yes
        - Provide to outreach function, with appropriate documentation.
        - No
        - DISCARD SAMPLES, with appropriate documentation.
The future of a USGS collection

- Staff scientists > initial recommendation to local collections’ manager/curator
- Joint request to local collections governing board (Science Center level)
- Recommendation to USGS Collections Steering Committee: advice on vetting collection by others within/outside USGS
- Recommendations then made by local governing board with Steering Committee approval to Science Center Director for decision
- USGS provides some funds for intramural collection management and grants to State Geological Surveys and other DOI agencies
- National Geological and Geophysical Data Preservation Program: datapreservation.usgs.gov
Digital Data

- Characteristics of good digital data: FAIR (Findable, Accessible, Interoperable, and Reusable)
- FAIR also applicable to physical collections
- Similar issues: appropriate metadata, what to keep and what to discard, and cost of managing collections
Conclusions

- OSTP issued specific directive to the Agencies
- Agency plans and strategies unevenly established and implemented
- Budget needed for collection storage, management, accessibility, and reusability
- Researchers must consider data and physical collection management strategies and cost early in the project
- Individual scientists’ initiative very important
- Disciplinary and data scientists work together to ensure FAIR
Thank you for your attention

esztein@nas.edu
Partial Bibliography