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

What are object-based scientific collections?

"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?

SIDEBAR 2-10
Examples of Inaccessible Geoscience Data and Collections

1) The paleontological collection at USGS Denver Federal Center is probably the largest such collection in the United States for which there is no funding for curation. It is also one of the largest with no standardized, computerized catalog. Knowledge about the collection resides with only a few people, many of whom are retired. As large and scientifically important as the USGS fossil collection is, there is no budget for collections management (committee survey response, 2001). Staff paleontologists have direct responsibility for curating their own collections (each cataloging specimens in their own style), yet official allocated time for curation is zero. The result is a variety of catalogs in handwritten ledgers, typed index cards, or computer database systems with no standardized format or medium of storage. Individual collections are commonly accessible only when the investigator is present. When a scientist retires or leaves, much of the institutional memory about the collection also departs. (The committee visited the paleontological collection in June 2001.)
2) An independent oil company, HS Resources, acquired Amoco's interests in an oil field in 1997. By June 2001, after being stored outdoors for 2 years, the cores were on unorganized pallets in a warehouse with random equipment laid on top of them. HS Resources merged with Kerr McGee in September 2001. The cores were still in the same location in February 2002 (John Ladd, Kerr McGee Rocky Mountain Corporation, personal communication, 2002).
3) DOE cores stored at Oak Ridge National Laboratory (ORNL) in Tennessee are stacked outside buildings in the open air and are overgrown with weeds (see photograph below). If they are not curated soon, these cores will be useless. Even if the rock should survive, the boxes and annotations on the samples will be lost, thus rendering them nearly valueless. The cores in the photograph in this sidebar were obtained by the Tennessee Valley Authority on the Clinch River Breeder Reactor site. Cores at DOE's Hanford, Washington, site also are exposed to the elements, although they are not maintained as poorly as those at Oak Ridge. In addition, the operating contractor in charge of ORNL, University of Tennessee–Battelle, has indicated that it may dispose of all but a few thousand feet of the 35,000 feet of rock core for the Oak Ridge Reservation—samples that indicate the fractured rock characteristics and basic subsurface geology for the ORNL site. The replacement cost for these cores is estimated at $5 million to $10 million (Richard Ketelle, Bechtel Jacobs Company LLC, personal communication, 2002).

Cores stored outside at Oakridge National Laboratory, Tennessee, in Spring 2001. These cores are from the Tennessee Valley Authority's Clinch River Breeder Reactor site. Rescue of cores in this state of degradation is unlikely given the probable loss of documentation associated with them. SOURCE: Richard Pawlowicz, Bechtel National, Inc., San Diego, California.

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

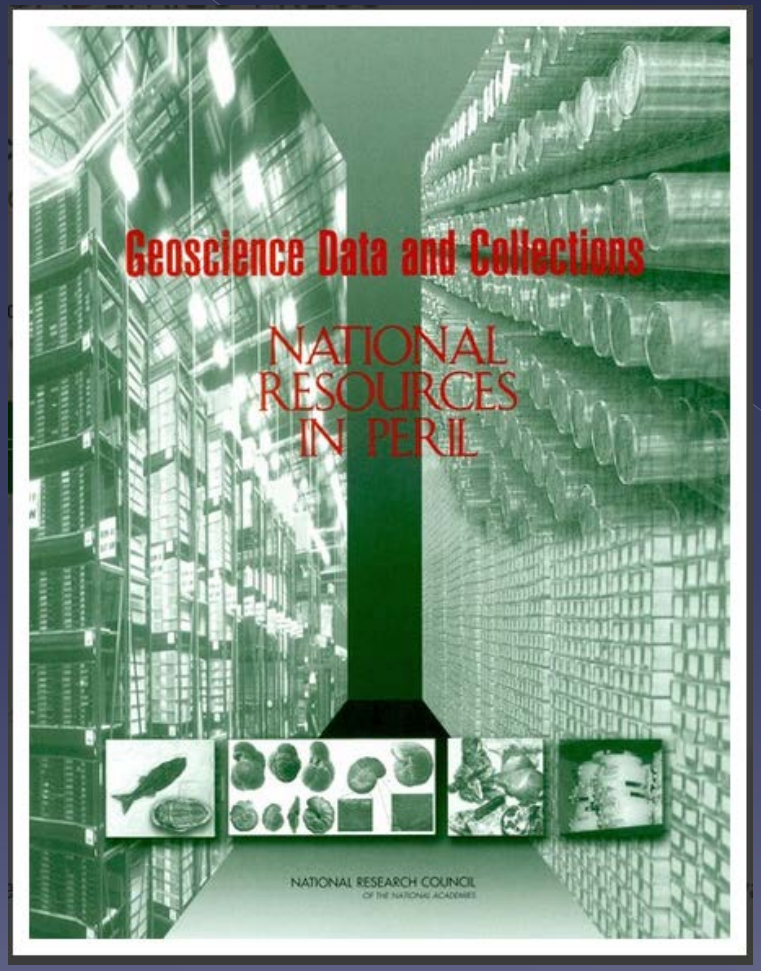
Why the USG?

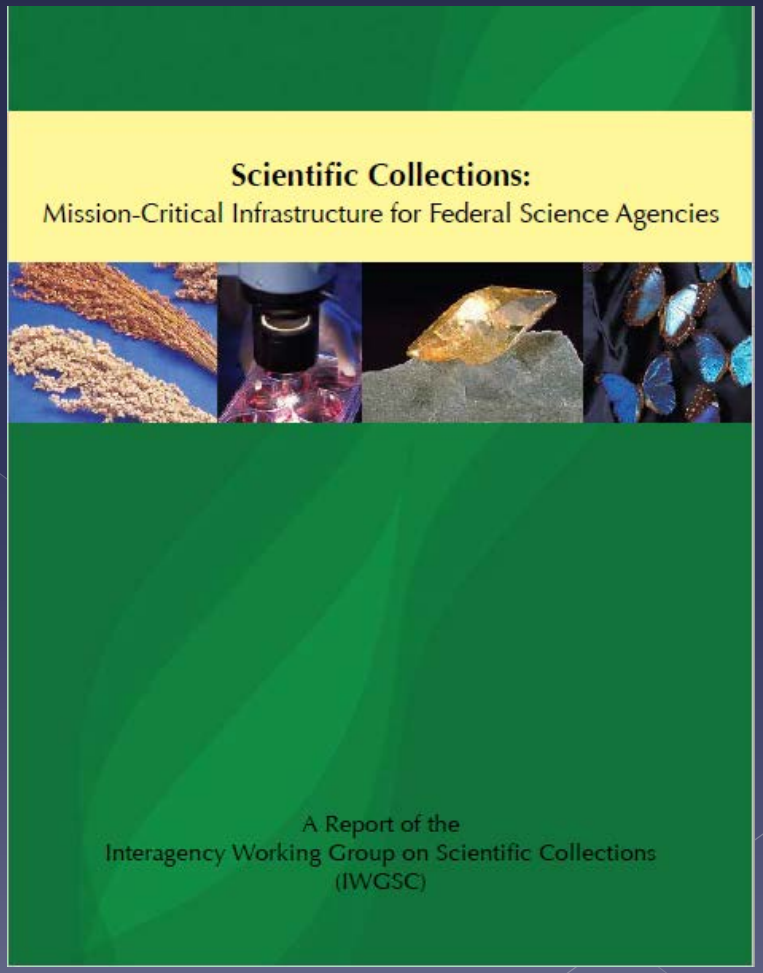
- Federally-supported science collections are public assets
- Support mission-critical science
- Excellent ROI for U.S. taxpayers
- Historic bipartisan interest

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

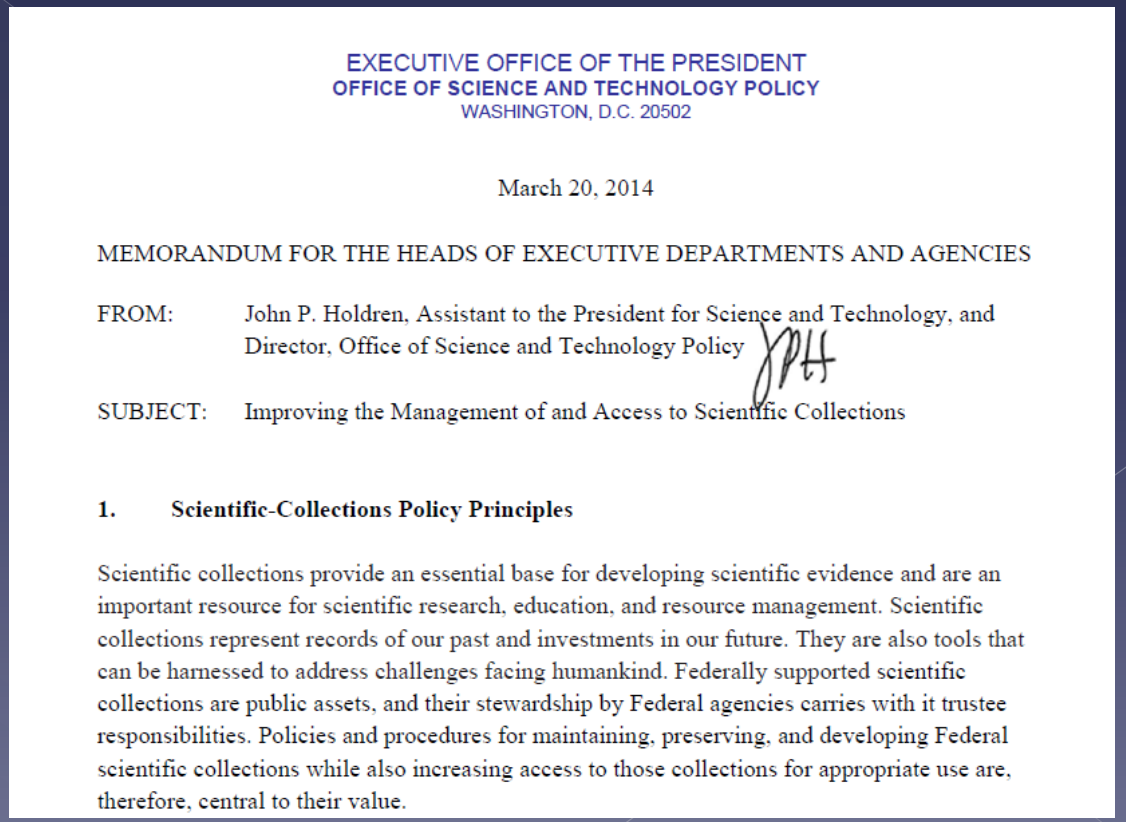
IWGSC's first 3 recommendations

- 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



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

March 20, 2014

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren, Assistant to the President for Science and Technology, and Director, Office of Science and Technology Policy

SUBJECT: Improving the Management of and Access to Scientific Collections

1. Scientific-Collections Policy Principles

Scientific collections provide an essential base for developing scientific evidence and are an important resource for scientific research, education, and resource management. Scientific collections represent records of our past and investments in our future. They are also tools that can be harnessed to address challenges facing humankind. Federally supported scientific collections are public assets, and their stewardship by Federal agencies carries with it trustee responsibilities. Policies and procedures for maintaining, preserving, and developing Federal scientific collections while also increasing access to those collections for appropriate use are, therefore, central to their value.

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 IWGSC

US Federal Scientific Collection Registry

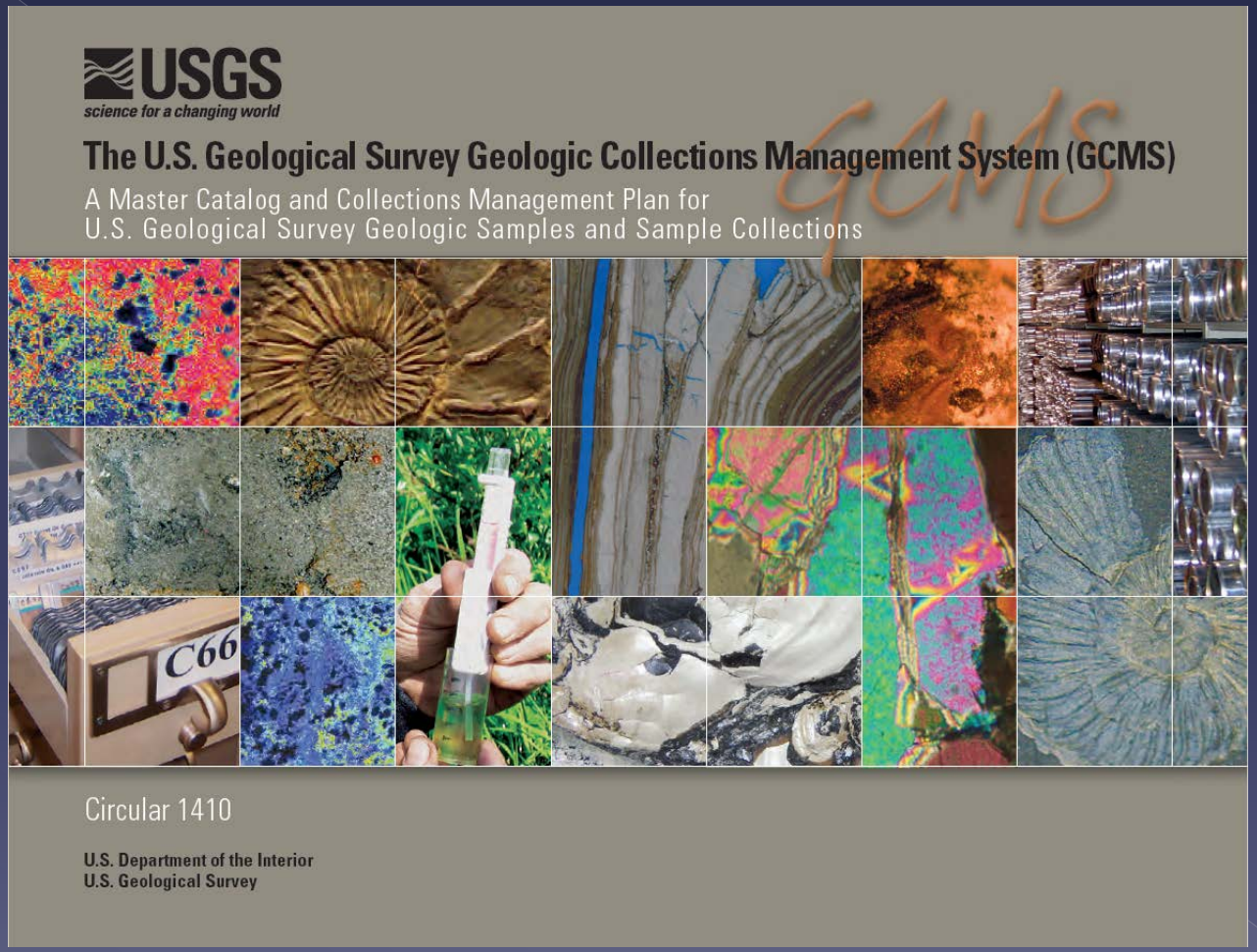
Curated source of information about object-based science collections owned or managed by US Federal departments and Agencies (usfsc.grscicoll.org)

Supported by IWGSC, managed by SI

Scientific Collection Policy Status by Agency

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



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

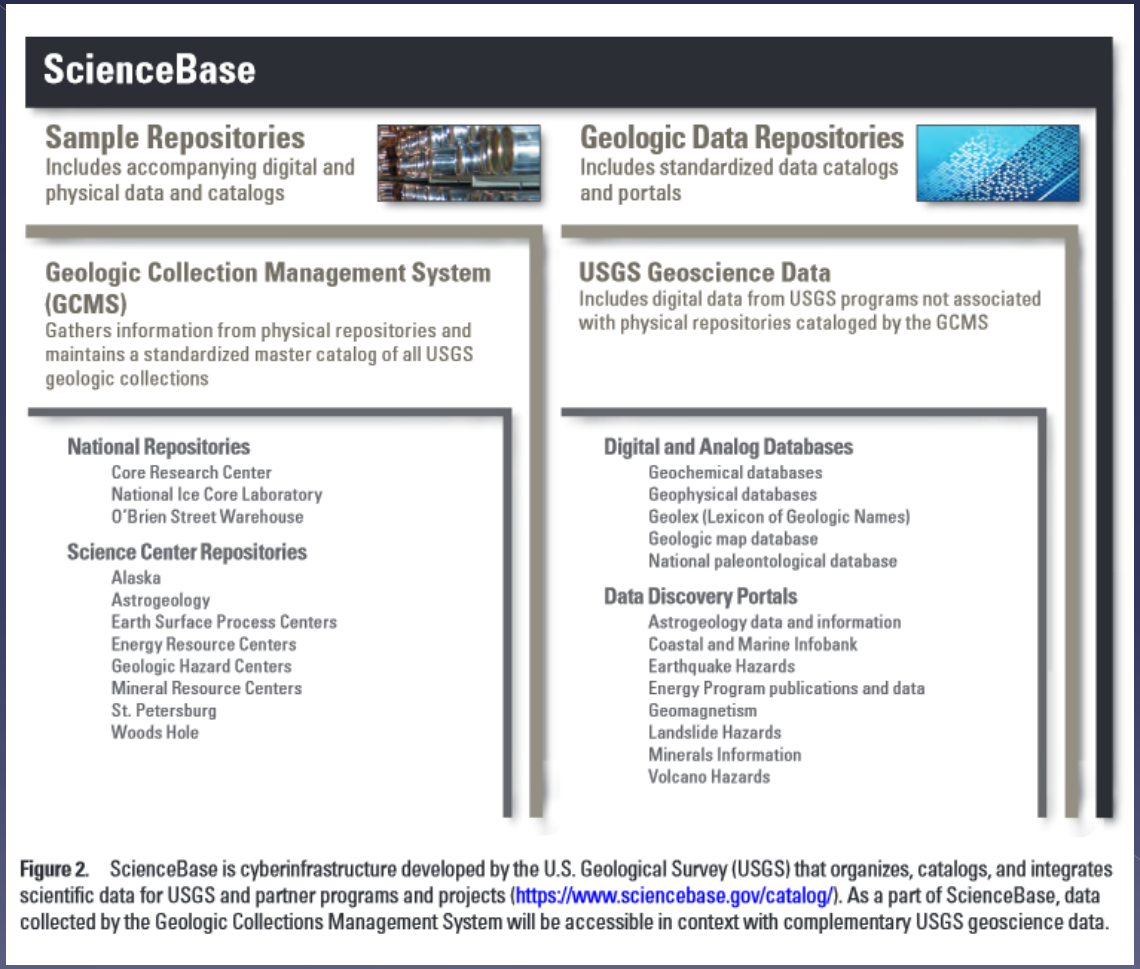


Figure 2. ScienceBase is cyberinfrastructure developed by the U.S. Geological Survey (USGS) that organizes, catalogs, and integrates scientific data for USGS and partner programs and projects (https://www.sciencebase.gov/catalog/). As a part of ScienceBase, data collected by the Geologic Collections Management System will be accessible in context with complementary USGS geoscience data.

Some definitions

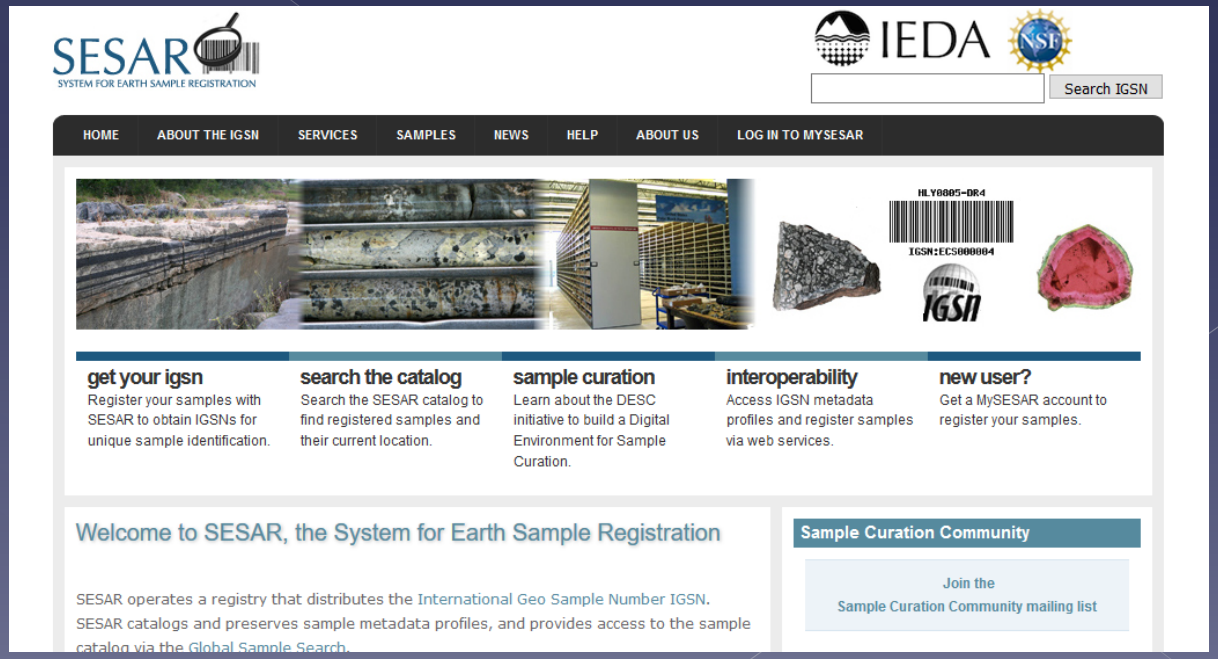
- An **active collection** (aka **working collection**) contains material from ongoing research and is actively used by the project scientists. New samples are added as research continues. When completed, materials will be evaluated for permanent retention in a GCMS repository.
- A **legacy collection** 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.
- An **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.
- A **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.
- A **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.



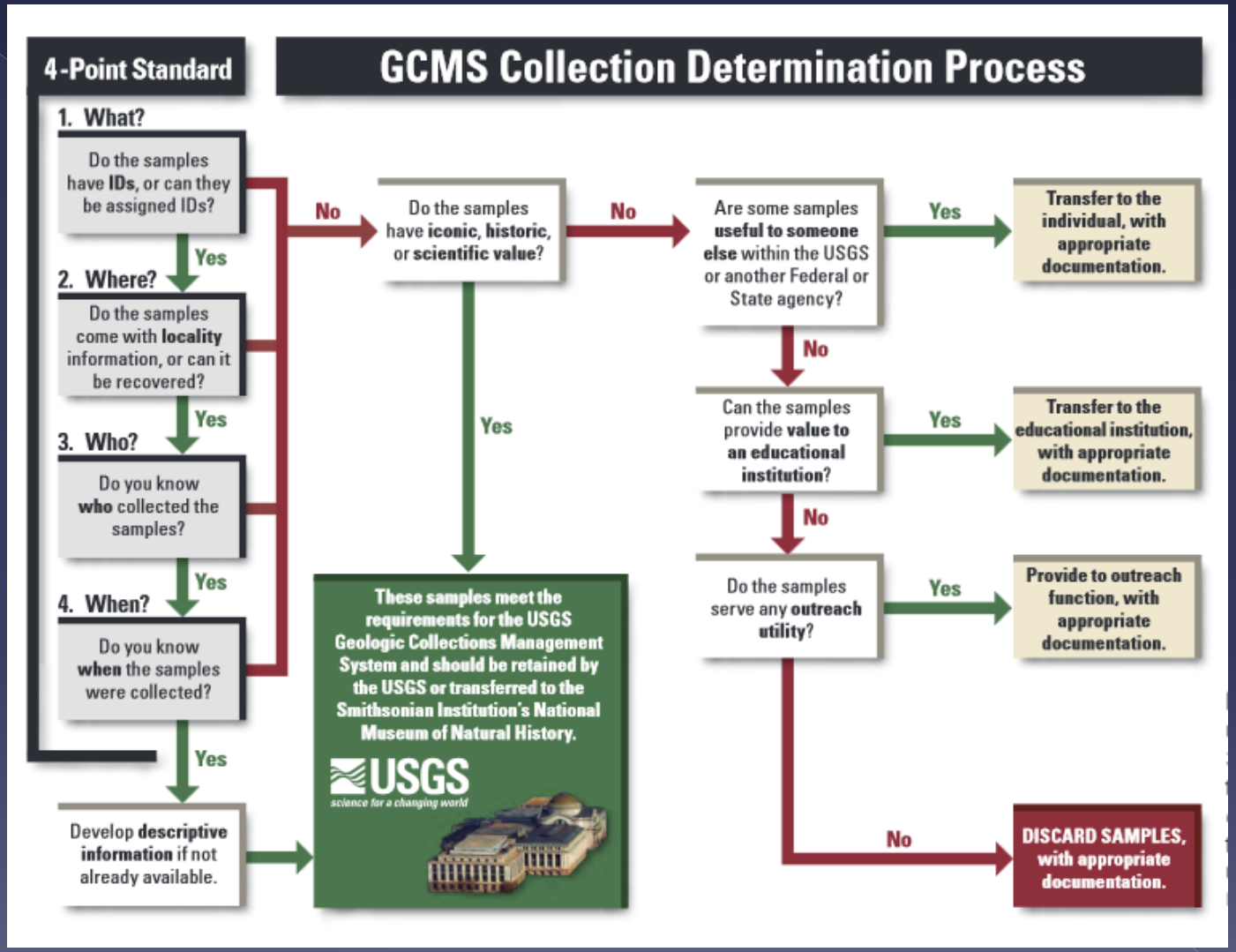
http://www.geosamples.org/

Think early and often…

about your data management plan

AND

your physical collection management plan



http://pubs.usgs.gov/circ/1410/

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

- Interagency Working Group on Scientific Collections (IWGSC): https://usfsc.nal.usda.gov/about-iwgsc
- Geoscience Data and Collections: National Resources in Peril (National Research Council, 2002) https://www.nap.edu/catalog/10348/geoscience-data-and-collections-national-resources-in-peril
- USGS Geological Collection Management System: https://pubs.usgs.gov/circ/1410/
- System for Earth Sample Registration (SESAR): http://www.geosamples.org/