

Data Policy: Much More than Just the Data Set “USGS Science Data Policies: An Evolving Culture of Data Management”

Mike Frame¹, Vivian Hutchison¹, and Ben Wheeler¹

¹U.S Geological Survey

November 23, 2022

Abstract

In 2013, the Office of Science and Technology Policy (OSTP) issued a memorandum directing Federal agencies with over \$100 million in annual research and development expenditures to develop a plan to support increased access to federally funded research results. In response, the US Geological Survey developed a Public Access Plan and published four new data management policies. The policies focus on review and approval of scientific data supporting scholarly conclusions, requirements for metadata, preservation, and data management planning. The new policies, in conjunction with the Public Access Plan, represent a shift in culture in how the USGS manages and provides access to its science data. The USGS recognizes that successful implementation of these new policies requires multiple pillars of support, from USGS leadership and staff buy-in, to effective tools. Active community engagement in the Bureau is stimulated through the Community for Data Integration (CDI), an open forum for community discussion and engagement, and an important component creating buy-in and contributing to the success of the new policies. Also critical are a suite of tools available to scientists to ensure their ability to implement the policies. Finally, support from leadership that manifests in the Fundamental Science Practices Advisory Council (FSPAC), a committee of representatives from across the Bureau who preside over policies and guidance is a critical component. While far from complete, the USGS has shifted its approach to science data management by engaging the community, offering tools to support policy, and providing leadership support for the quality and scientific integrity of USGS science data.

Data Policy: Much More than Just the Data Set

“USGS Science Data Policies: An Evolving Culture of Data Management”

Data Policy: Much More than Just the Data Set
“USGS Science Data Policies: An Evolving Culture of Data Management”
 Michael Frame, Vivian Hutchison, Ben Wheeler
 Core Science Systems, U.S. Geological Survey, United States, mike_frame@usgs.gov

Long History of USGS
 USGS has been managing and releasing much of its science data for decades. John (Jack) Powell made the first systematic data management of the Rio Grande River of Embury Station from Mexico in 1925. This early, 100-year-old data management process has evolved into what is known today as the USGS National Water Information System (NWIS) data to support Federal

USGS Release Data Example
 The USGS produces and releases scientific data across a broad range of disciplines. Offering data formats, tools and other information present a wide variety of interesting challenges in release and preservation needs for the USGS.

USGS Data Management and Release
 Enterdata, a USGS Trusted Digital Repository for release of scientific data, offers scientists a secure, government-backed platform to release data. Enterdata starts with the Enterdata Data Release application that builds a landing page and creates a digital object identifier.

USGS Science Data Lifecycle
 The USGS follows the USGS Science Data Lifecycle for managing data throughout its lifecycle. From project inception to preservation, data is released only after it has undergone review and approval, is accompanied by a metadata record and a digital object identifier, and the data resides in a Trusted Digital Repository such as Enterdata. The metadata is submitted to the Science Data Catalog for redistribution.

USGS Policies
 In 2015, the Office of Science and Technology Policy (OSTP) issued a memorandum directing Federal agencies with over \$100 million in annual research expenditures to develop a plan to support increased access to Federal United States research results. In response, the U.S. Geological Survey developed a Public Access Plan and published four new data management policies. The policies focus on release and approval of scientific data supporting scientific investigations, requirements for metadata, preservation and data management planning. The new policies in conjunction with the Public Access Plan

USGS Data Management Website
 Enterdata is on the USGS Data Management Website for guidance and best practices in data management and release. The Data Management Website supports Fundamental Science Principles, a set of policies and principles by which USGS maintains scientific values. The site is organized by the components of the data lifecycle, and offers best practices, tools, and training. <https://data.usgs.gov/datamanagement/>

Michael Frame, Vivian Hutchison, Ben Wheeler

Core Science Systems, U.S. Geological Survey, United States, mike_frame@usgs.gov

PRESENTED AT:

LONG HISTORY OF USGS DATA

USGS has been managing and releasing much of its science data for decades. John Wesley Powell made the first systematic flow measurement of the Rio Grande River at Embudo Station New Mexico in 1889. This early, 129-year-old data management process, has evolved into what is known today as the USGS National Water Information System (NWIS). Now, due to current Federal directives on data release, USGS authors are expected to release all data used to support conclusions reached in publications.



Figure 1: John Wesley Powell Expedition, Colorado River, 1872, E.O Beamanm USGS, Reference: <https://library.usgs.gov/photo/#/item/51dc898fe4b097e4d3839785>

USGS POLICIES

In 2013, the Office of Science and Technology Policy (OSTP) issued a memorandum directing Federal agencies with over \$100 million in annual research expenditures to develop a plan to support increased access to federally funded research results. In response, the US Geological Survey developed a Public Access Plan and published four new data management policies. The policies focus on review and approval of scientific data supporting scholarly conclusions, requirements for metadata, preservation, and data management planning. The new policies, in conjunction with the Public Access Plan, represent a shift in culture in how the USGS manages and provides access to its science data.

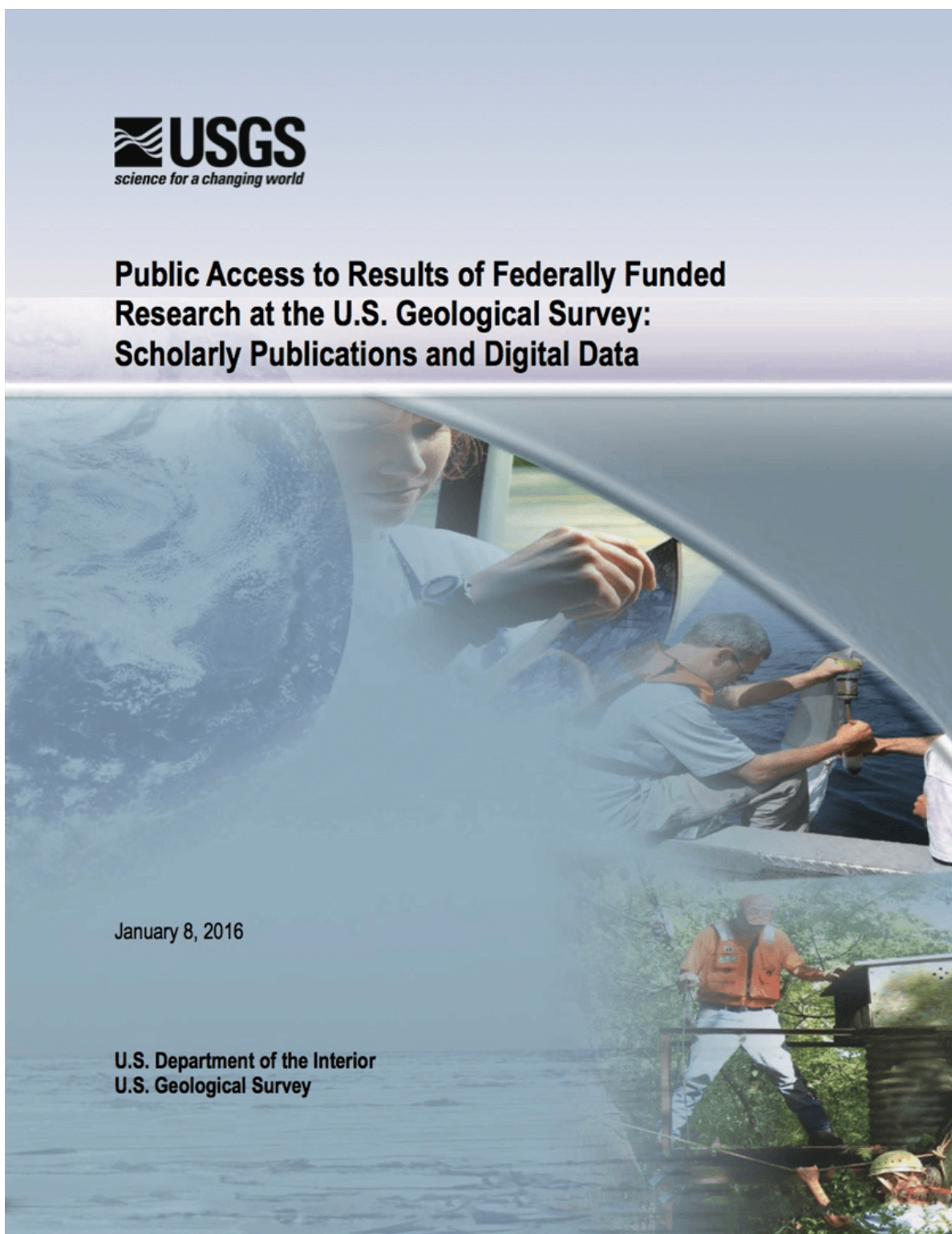


Figure 2: USGS Public Access Plan, USGS

USGS KILAUEA DATA EXAMPLE

The USGS produces and releases scientific data across a broad range of disciplines. Differing data formats, sizes and other characteristics present a wide variety of interesting challenges in release and preservation needs for the Bureau.

In this example, the recent eruption of the Kilauea volcano in Hawaii created demand for a rapid release of data to protect lives in an emergency situation for the island.

[VIDEO] <https://www.youtube.com/embed/jnQPKfqMrE8?feature=oembed&fs=1&modestbranding=1&rel=0&showinfo=0>

Animation 1: 3D model of Kilauea crater following the 2018 eruption

USGS DATA MANAGEMENT AND RELEASE TOOLS

Sciencebase, a USGS Trusted Digital Repository for release of scientific data, offers scientists a secure, permissions-controlled platform to release data hassle-free. Scientists start with the ScienceBase Data Release application that builds a landing page and reserves a digital object identifier.

The screenshot shows the 'ScienceBase Data Release' web application. At the top is the USGS logo with the tagline 'science for a changing world'. To the right are links for 'USGS Home', 'Contact USGS', and 'Search USGS'. Below the header, a breadcrumb trail reads 'CSAS&L / ScienceBase Data Release Home / Create Data Release', and a link for 'Report a technical problem: ScienceBase Data Release Service Desk' is on the right. The main heading is 'Create New Data Release'. A paragraph explains the process: 'Please fill out and submit this form to start a data release in ScienceBase. A new landing page with a digital object identifier (DOI) will be created. You will receive an email with a link to the new landing page and further instructions on how to complete the data release.' A red asterisk indicates required fields. The first section is '* Data Release Title', with a note: 'Please choose a title that is descriptive. This title will be used in the data release citation and will be discoverable in ScienceBase search results.' Below this is a text input field containing the example: 'e.g. Boat-based water-surface elevation surveys along the upper Willamette River, Oregon, in March 2015'. The second section is '* IPDS number', with a note: 'To get an IPDS number for your data release go to <https://ipds.usgs.gov>, create a new record, and select "Data Release" as the product type. e.g. IP-012345'. Below this is a text input field containing 'IP-'. The third section is '* USGS Mission Area', with a note: 'Please select the Mission Area with which this data release is associated'. Below this is a dropdown menu with a downward arrow.

Figure 3: Screenshot of ScienceBase Data Release Tool, USGS

A data release in ScienceBase includes an overview of the data, citation information generated from the metadata upload, data and metadata files, the digital object identifier, and a reference to the associated publication.

ScienceBase Catalog → USGS Data Release Products → Airborne Geophysical Surveys → Magnetic Data of Airborne G...

Magnetic Data of Airborne Geophysical Surveys over the Eastern Adirondacks, New York State

View

Dates

Publication Date : 2016

Start Date : 2015-12-07

End Date : 2015-12-21

Citation

Shah, A.K., 2016, Airborne Geophysical Surveys over the Eastern Adirondacks, New York State: U.S. Geological Survey data release, <https://dx.doi.org/10.5066/F72R3PT0>.

Summary

Airborne geophysical surveys were conducted in the eastern Adirondacks from Dec. 7, 2015 - Dec. 21, 2015, by Goldak Airborne Surveys. The area was flown along a draped surface with a nominal survey height above ground of 200 meters. The flight line spacing was 250 meters for traverse lines and 2500 meters for control lines. Here we present downloadable magnetic data from those surveys as flight line data (csv format) or Geotiff images.

Line data (EAdirondacksMag.csv) are presented as comma separated values with the following definitions (see technical report for details):

LINE: Flight line number
DATE: YYYY/MM/DD
TIME: Time (UTM, 24-hour) hh:mm:ss
LONG: Longitude, WGS84 (degrees)
LAT: Latitude WGS84 (degrees)
x_UTM18N: Easting, UTM Zone 18N, WGS84 (meters)
y_UTM18N: Northing, UTM Zone 18N, WGS84 (meters)
RALT: Radar altimeter height (meters)
BALT: Barometric altimeter height (meters)

... show more ...

Contacts

Point of Contact : Anjana K Shah

Originator : Anjana K Shah

Metadata Contact : Anjana K Shah

Publisher : U.S. Geological Survey

Distributor : U.S. Geological Survey - ScienceBase

Attached Files

Click on title to download individual files attached to this item or [download all files](#) listed below as a compressed file.

EAdirondacks_MagRes.tif		6.9 MB
Magres_colorbar.jpg		17.88 KB
DataDefinitions_Mag.csv		1.13 KB
EAdirondacksMag.csv		319.29 MB
MagAnom_thumbnail.jpg		94.61 KB
EAdirondacksMag_metadata.xml <i>Original FGDC Metadata</i>		12.7 KB

Related External Resources

Type: Online Link

Sciencebase catalog parent item	https://www.sciencebase.gov/catalog/item/582b9457e4b0c253be072a40
---------------------------------	---

Map

Communities

USGS Data Release Products

Tags

Categories : Data

Theme : aeromagnetic maps, aeromagnetic surveying, aeroradiometric surveying, airborne magnetics, economy, geologic mapping, geologic maps, geoscientific information, magnetic anomaly maps, magnetic field (earth), magnetics, mineral resources, mineral resources, subsurface imaging, subsurface maps

Place : Adirondack Mountains, Essex County, New York, United States

Provenance

Data source : Input directly

Figure 4: Example of a data release landing page in USGS ScienceBase, USGS, Reference: <https://www.sciencebase.gov/catalog/item/582b9457e4b0c253be072a40>

<https://agu2018fallmeeting-agu.ipostersessions.com/Default.aspx?s=37-0F-6F-2E-78-6B-BB-AF-3E-6B-52-25-7F-75-5A-92&pdfprint=true&guestview=true>

8/12

USGS SCIENCE DATA LIFECYCLE IMPLEMENTATION

The Bureau follows the USGS Science Data Lifecycle for managing data throughout its lifespan, from project inception to preservation. Data is released only after it has undergone review and approval, is accompanied by a metadata record and a digital object identifier, and the data resides in a Trusted Digital Repository such as ScienceBase. The metadata is submitted to the Science Data Catalog for re-distribution.

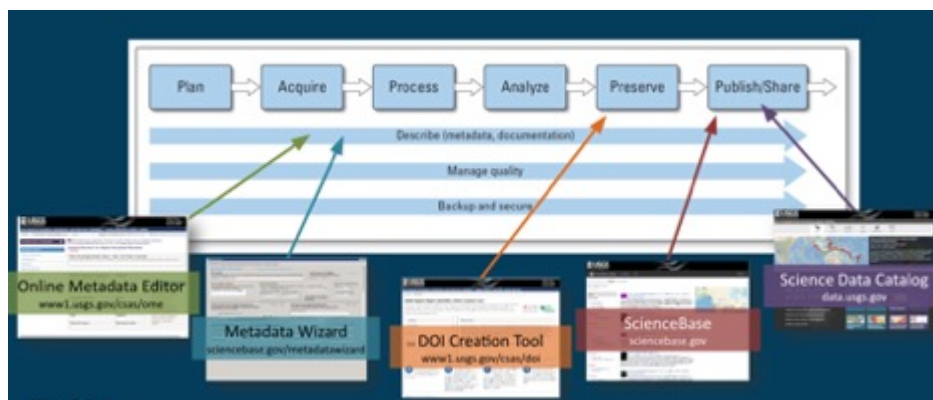


Figure 5: Example of the USGS Science Data Lifecycle, and tools that assist USGS researchers and data managers in implementing it, USGS, Reference:

https://docs.google.com/presentation/d/1h-plcQloMxP53iFTQe_PH1ceZA__7MCIF8QgMN-KhmM/edit#slide=id.p21

USGS DATA MANAGEMENT WEBSITE

Scientists rely on the USGS Data Management Website for guidance and best practices in data management and release. The Data Management Website supports Fundamental Science Practices, a set of policies and principles by which USGS researchers conduct science. The site is organized by the components of the data lifecycle, and offer best practices, tools, and training.

<usgs.gov/datamanagement>

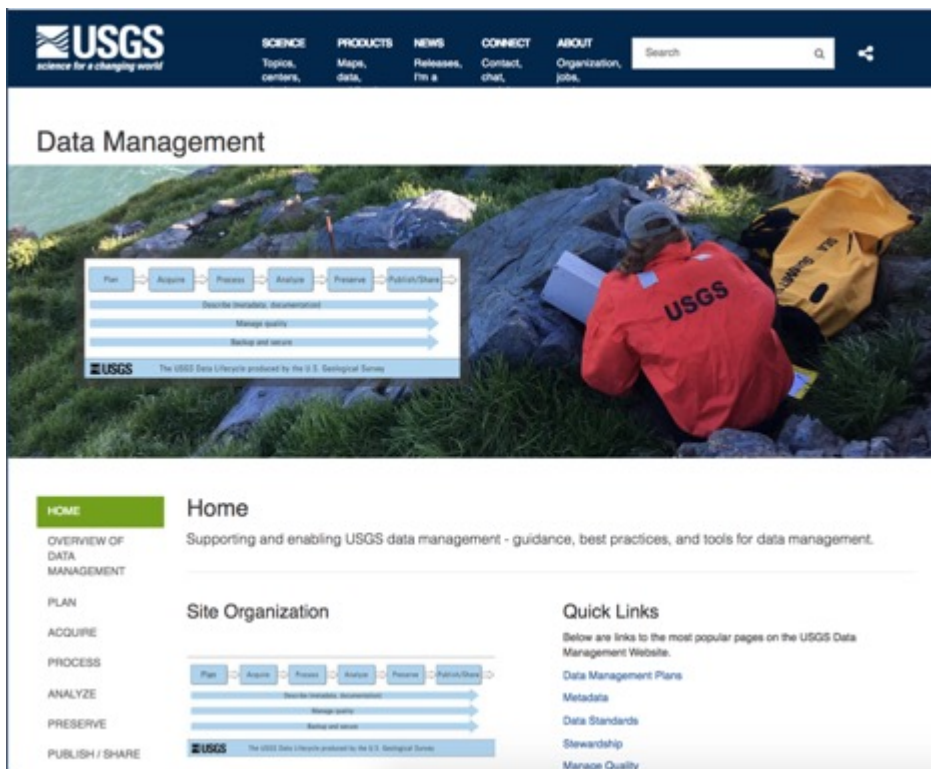


Figure 6: Screenshot of USGS Data Mangement Website, USGS

LINK:

ABSTRACT

In 2013, the Office of Science and Technology Policy (OSTP) issued a memorandum directing Federal agencies with over \$100 million in annual research and development expenditures to develop a plan to support increased access to federally funded research results. In response, the US Geological Survey developed a Public Access Plan and published four new data management policies. The policies focus on review and approval of scientific data supporting scholarly conclusions, requirements for metadata, preservation, and data management planning. The new policies, in conjunction with the Public Access Plan, represent a shift in culture in how the USGS manages and provides access to its science data.

The USGS recognizes that successful implementation of these new policies requires multiple pillars of support, from USGS leadership and staff buy-in, to effective tools. Active community engagement in the Bureau is stimulated through the Community for Data Integration (CDI), an open forum for community discussion and engagement, and an important component creating buy-in and contributing to the success of the new policies. Also critical are a suite of tools available to scientists to ensure their ability to implement the policies. Finally, support from leadership that manifests in the Fundamental Science Practices Advisory Council (FSPAC), a committee of representatives from across the Bureau who preside over policies and guidance is a critical component. While far from complete, the USGS has shifted its approach to science data management by engaging the community, offering tools to support policy, and providing leadership support for the quality and scientific integrity of USGS science data.