Structuring and Documenting a USGS Public Data Release

Part I - General Considerations
Terms Used in this Presentation

**Dataset**: A structured collection of data. A dataset may consist of a single file or an ordered collection of files that contain observations or measurements (unprocessed or processed) as text, numbers, or multimedia.

**Public Data Release**: A type of USGS information product designed to provide USGS scientists with a channel to publish reviewed and approved data. Unless exempted under provisions addressing sensitive or proprietary data, USGS authors are required to make bureau-funded data publicly available. The data that support scholarly publications must be released prior to, or simultaneously with, the publication. Other project data must be released no later than the end of the project. These requirements may be met with USGS data releases. Data release products contain one or multiple datasets, alongside metadata that describe what the files contain, provide relevant details about the collection or production of the data, and offer guidelines regarding appropriate use of the data.

**Landing Page**: A webpage that provides access to the data and metadata for a data release. The digital object identifier for a data release should link to a landing page.
Part I Overview

Determine how you want users to discover the data

Organize data to support user access and exploration

Choose a method and platform for distributing the data
Determine how you want users to discover the data

- **Single Package**
  - If components require the context of the entire data release

- **Individual Datasets**
  - If components don’t require the context of the entire data release and may have utility independent from the data release
Determine how you want users to discover the data - Metadata Considerations

- **Single Package**
  - Data release metadata

- **Individual Datasets**
  - Data release metadata
  - Dataset-level metadata

https://data.usgs.gov/datalocatlog/
Organize data to support user access and exploration

- **Single package**
  - Downstream users will likely want to acquire all of the data in the data release
  - Data release is relatively small and can be easily downloaded as a single file or resource
    - Aggregated into a single dataset or database
    - Zip file to package individual files

- **Individual datasets**
  - Downstream users will likely want to pick and choose one or more of the datasets within the data release
  - Data release is large and users may have trouble downloading the data release as a single package
Organize data to support user access and exploration - Metadata Considerations

Single Package

Individual Datasets
Distributing Data - Choose a Method(s)

- Data download
- Web services
  - Visualize data
  - Subset data
  - Query or analyze data without downloading
Distributing Data - Metadata Considerations

- Distribute data and metadata together
  - Easy to download together
  - Data shared via web services should also have metadata available
Distributing Data - Choose a Repository

Fundamental Science Practices Advisory Committee’s list of acceptable repositories for USGS data release:
https://www2.usgs.gov/fsp/acceptable_repositories_digital_assets.asp

Some considerations for choosing a repository:

- Is there an appropriate discipline-specific repository for your data?
- Are the file sizes and formats acceptable for the repository?
- Do your data require web services?
- What support is available from the repository? (e.g. DOI minting and management, file format migration, etc.)
- What are your needs for interacting with your data?
Distributing Data via ScienceBase

- Is there an appropriate discipline-specific repository for your data?
  - ScienceBase is not a discipline-specific repository

- Are the file sizes and formats acceptable for the repository?
  - ScienceBase can accept files up to 10 GB

- Do your data require web services?
  - ScienceBase can generate WMS, WFS, WCS, and ArcGIS REST services for certain geospatial file types

- What support is available from the repository (e.g. DOI minting and management, file format migration, etc.)
  - ScienceBase mints and manages DOIs for data producers

- What are your needs for interacting with your data?
  - ScienceBase offers a robust application programming interface (API) for interacting with data
Structuring and Documenting a USGS Public Data Release

Part II - Distributing Data Through ScienceBase
Part II Overview

- The ScienceBase Framework
- Determine how you want users to discover the data in ScienceBase
- Organize data to support user access and exploration in ScienceBase
- Improve accessibility of ScienceBase data
The ScienceBase Framework

Interconnected network of items
The ScienceBase Framework

ScienceBase Item

Landing Page for your Data
Determine how you want users to discover the data in ScienceBase

Single Package

Dataset(s) + 1 XML Metadata

Landing Page

Individual Datasets

1 XML Metadata

Dataset + 1 XML

Landing Page

Dataset + 1 XML

Child Page 1

Dataset + 1 XML

Child Page 2

Dataset + 1 XML

Child Page 3
Missouri Breaks Project, Montana - Digitized aeromagnetic data

Dates
Publication Date: 2019-09-06
Time Period: 1981-01-01

Citation

Summary
From February 12 to March 8, 1981, EOS and G-Geometrics conducted an aeromagnetic survey in Montana for Anaconda Copper Company. A Piper H650 aircraft was used to conduct the survey. The survey was flown along the north-south flightline spaced 600 feet at a nominal height of 400 feet above the terrain. Five uniformly spaced east-west flight lines were also flown. During processing, the International Geophysical Reference Field (IGRF) of 1975 was updated to 1981 and removed from the survey observations. The resultant anomalous magnetic field was contoured at a 4.0 gamma interval and plotted to map sheets. In 2015, the map sheets were scanned, registered, and the contours were digitized. The point data extracted from the digitized contour lines and includes latitude, longitude, and magnetic-field value (gamma).

Contacts
Point of Contact: Eric D. Anderson
Metadata Contact: Eric D. Anderson
Publisher: United States Geological Survey
Distributor: U.S. Geological Survey - ScienceBase

Attached Files
Click on File to view/download individual files attached to this item or download all files listed below as a compressed file.

Purpose
Aeromagnetic surveys are used for geophysical prospection. Various anomalies in magnetic measurements are caused by rocks that carry a significant amount of magnetic minerals (the most common). These anomalies reflect variations in the amount and type of magnetic material, and the shape and depth of the body of rock. The features and patterns of aeromagnetic anomalies can also be used to delineate details of subsurface geology including the locations of buried faults and the thicknesses of surficial sedimentary rocks.

Example: http://dx.doi.org/10.5066/F7TH8JTN
Example 2

One ScienceBase Page

Example:
https://doi.org/10.5066/F7ZC810J
Example 3

5 Zipped GeoTIFFs with Metadata

Example:
http://dx.doi.org/10.5066/F7CRR5RF
Example 4

One Top Level Landing Page...

...with Two Child Pages

Example:
http://dx.doi.org/10.5066/F7W9577Q
**Example 4**

**One Shapefile with Metadata**

---

**Ozark Plateau seepage flow-line dataset, southern Missouri and northern Arkansas, 1982-2006**

**Metadata:**
- Identification Information
- Spatial Attribute Information
- Spatial Temporal Information
- Quality and Metadata Information
- Metadata Reference Information

**Identification Information:**

**Citation:**

- **Source:** USGS National Streamgauging Program
- **Date:** 2006
- **DOI:** http://dx.doi.org/10.5066/F7W9577Q

---

**ScienceBase Child Page #1**

**Ozark Plateau seepage flow-line dataset, southern Missouri and northern Arkansas, 1982-2006**

**Dates**
- **Publication Date:** 2015
- **End Date:** 2006-09-20

**Citation**


**Summary**

This dataset was compiled to examine discharge measurements from several published groundwater and surface-water studies in the Ozarks of southern Missouri and northern Arkansas. The discharge measurements were part of the USGS National Streamgauging Program and conducted at gauging stations identified in the Ozarks data layer. The discharge was measured using a variety of techniques, including direct measurement of flow at the surface and indirect measurement using a variety of methods, such as dye tracing and hydraulic modeling. The data are available for download through the USGS EarthExplorer Data Release.

---

**Contacts**

- **Point of Contact:** Knecht, K.J., USGS National Streamgauging Program
- **Distribution:** U.S. Geological Survey - ScienceBase
- **Metadata Contact:** Knecht, K.J., USGS National Streamgauging Program

**Attached Files**

- **Dataset:** OzarkPlateauSeepageFlowLines.zip
  - **Shapefile:** OzarkPlateauSeepageFlowLines.shp
    - **Data:** OzarkPlateauSeepageFlowLines.shp
      - **Size:** 2.09 MB
    - **NetCDF:** OzarkPlateauSeepageFlowLines.nc
      - **Size:** 1.29 MB
  - **Projection:** OzarkPlateauSeepageFlowLines.prj
    - **Shapefile:** OzarkPlateauSeepageFlowLines.shx
      - **Size:** 6.12 MB
    - **XML:** OzarkPlateauSeepageFlowLines.xml
      - **Size:** 245 Bytes
  - **Metadata:** OzarkPlateauSeepageFlowLines.xml
    - **Shapefile:** OzarkPlateauSeepageFlowLines.shp
      - **Size:** 334.56 MB
    - **XML:** OzarkPlateauSeepageFlowLines.xml
      - **Size:** 7.08 MB
    - **XML:** OzarkPlateauSeepageFlowLines.xml
      - **Size:** 884 Bytes
  - **Data:** OzarkPlateauSeepageFlowLines.shp
    - **Shapefile:** OzarkPlateauSeepageFlowLines.shp
      - **Size:** 5 Bytes

**Provenance**

- **Data Source:** Input data

---

Example: http://dx.doi.org/10.5066/F7W9577Q
Improve accessibility of ScienceBase data

ScienceBase supports:

- Basic data downloads
- Visualization of data via web mapping services
- Creation of web feature services and web coverage services
- API access to content
Summary

Determine how you want users to discover the data

Organize data to support user access and exploration

Choose a method and platform for distributing the data

https://www.sciencebase.gov/about/content/data-release
Still have questions?

Contact the ScienceBase team at sciencebase@usgs.gov