



Structuring and Documenting a USGS Public Data Release

Part I - General Considerations

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

This tutorial is designed to help scientists think about the best way to structure and document their USGS public data releases. The goal is to present data in a logical and organized manner that enables users to quickly understand the data. In the first part of this tutorial, we will discuss general considerations for structuring and documenting a data release, regardless of the platform being used to distribute the data. In the second part of this tutorial, we will discuss how these general considerations can be implemented in ScienceBase.

This tutorial is brought to you by the USGS Core Science Analytics, Synthesis, and Library's Science Data Management Branch.

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.



To ensure that we are using the same vocabulary, let's define a couple of terms that we will use in this presentation: dataset, data release, landing page. A dataset is 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. A public data release is 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 a scholarly publication 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. A landing page is a webpage that describes a data release and provides access to the data and metadata. The digital object identifier for a data release should resolve to a landing page. There are various types of data releases in USGS, including those done for cooperators. For this tutorial, we are focusing on the publicly accessible release of USGS data.

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



If you have a data release with multiple datasets, there are three things that you will want to consider when determining the best approach for structuring and documenting your data release. You'll want to think about how you want users to discover the data once they are publicly available. You'll want to make sure that the data are organized in a manner that will support user access and exploration. And you'll want to consider the method and platform that you will employ for distributing your data. In this first part of the tutorial, we will dive into each of these considerations.

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



Take a moment to think about the different datasets of your data release. Do you think that downstream users will require the context of the entire data release in order to understand the different datasets? If so, you will likely want users to discover your data release as a single package. If you believe that the different datasets could stand on their own and may have utility independent from the data release, you may want to allow users to discover the datasets individually.

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/datacatalog/>

Data discovery is enabled by metadata in the form of an XML file. If you want users to discover your data release as a single package, you will want to have a single metadata record that is made available to be indexed by metadata catalogs like the USGS Science Data Catalog and by search engines. If you want users to be able to discover the datasets of your data release individually, you will need to ensure that each dataset's metadata is available to metadata catalogs and search engines. Additionally, you may want to share a metadata record for the entire data release. For more information on the Science Data Catalog, follow the link shown on the screen.

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



You should also consider how users may want to explore and access your data: as a single package or as individual datasets. You may come to the same conclusion for discovery and access or you may not.

When thinking about exploration and access, consider if downstream users will likely want to acquire all of the datasets in the data release or if users will likely want to pick and choose one or more of the datasets within the data release. You should also consider the size of the data release. If the data release is relatively small and can easily be downloaded as a single file or resource, the single package approach may work. For example, you could consider aggregating the datasets into a single dataset or database or you could use a zip file to package individual files together. If the data release is large and users may have trouble downloading or using the data release as a single package, you will likely want to provide users with access to individual datasets or components.

Organize data to support user access and exploration - Metadata Considerations



Single Package




Individual Datasets

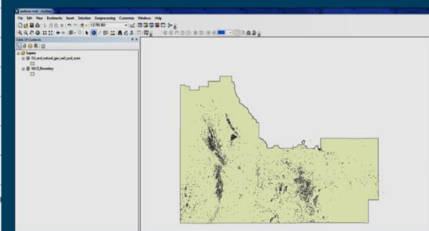
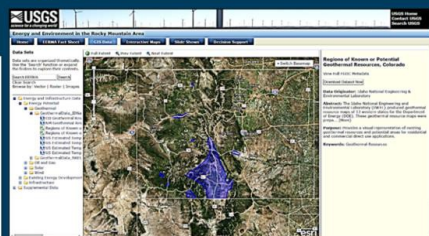
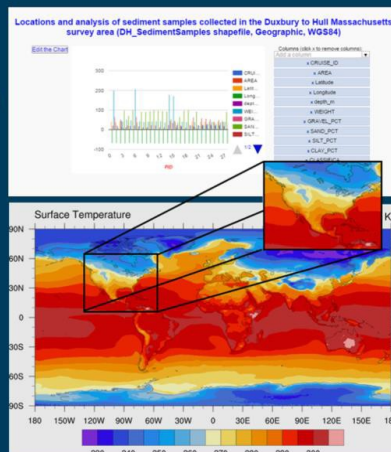


The way that you organize your data may have an impact on the way that you document the data in the form of metadata. You want to ensure that data and metadata for understanding the data are packaged and made accessible together. If your data are being organized as a single package, you may need to create fewer metadata records; however, you should always ensure that all of the data are sufficiently documented. You want to balance the efficiency of creating metadata records for your data release with the ease of understanding and use for downstream users. If your data are being organized as separate resources that are accessible individually, you want to ensure that each dataset is paired with a metadata record to help users understand the data.

Distributing Data - Choose a Method(s)



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



Finally, the method that you choose for distributing your data may also affect the way that you structure your data release. Often the easiest approach for distributing your data release will be to allow users to download the data and metadata files; however, for some data types, data producers may also be able to make use of web services.

Web services can help improve access for downstream users by allowing them

- to visualize the data in interactive maps before downloading or extracting them;
- to subset the data if they only need a portion of the data; or
- to query or analyze the data without downloading them by using a web feature service or web coverage service.

Repositories may have different requirements or recommendations for how you should structure your data to take advantage of web services.



Distributing Data - Metadata Considerations

- o Distribute data and metadata together
 - Easy to download together
 - Data shared via web services should also have metadata available



Again, it is important to ensure that data and metadata are distributed together, regardless of the distribution method that is employed. Metadata should be easy to download with the data, and 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?



Next, you need to choose a repository for your data. The USGS Fundamental Science Practices Advisory Committee (FSPAC) maintains a list of acceptable repositories for USGS data releases. Follow the link on the screen to browse this list.

As you are deciding on the best repository for your data, here are some considerations to keep in mind:

- Is there an appropriate discipline-specific repository for your data? For example, discipline-specific repositories exist for genetics data and seismology data.
- Are the file sizes and formats of your data acceptable for the repository?
- Do your data require web services?
- What support is available from the repository? For example, will the repository mint and manage DOIs for the data? Will the repository perform file format migration in the future?
- What are your needs for interacting with your data? For example, do you need to dynamically pull information about your data into a web page?

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



Let's take a look at these considerations with respect to ScienceBase, one of the acceptable repositories for USGS data releases.

ScienceBase is not a discipline-specific repository, so if an appropriate discipline-specific repository does exist, your data should go into that repository.

ScienceBase accepts open, machine-readable file formats and can accept files up to 10 GB.

ScienceBase can generate web mapping services, web feature services, web coverage services, and ArcGIS REST services for certain geospatial file types.

ScienceBase will mint and manage DOIs for data producers; however, at this time, ScienceBase does not perform file format migration services.

Finally, ScienceBase offers a robust application programming interface or API for interacting with data.



Structuring and Documenting a USGS Public Data Release

Part II - Distributing Data Through ScienceBase

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

Now that you know some of the general considerations for structuring and documenting a USGS public data release, as well as some basic information about the ScienceBase repository, let's take a look at some options for structuring and documenting a data release in the ScienceBase system.

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

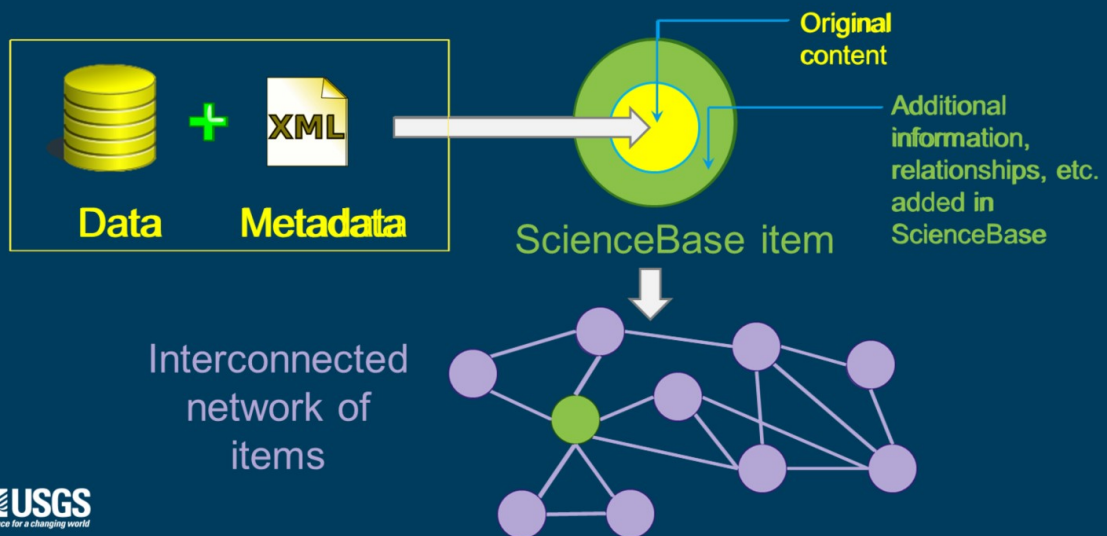


- Improve accessibility of ScienceBase data



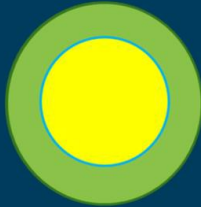
In this section of the tutorial, we will start off by discussing the ScienceBase framework. Then, we will talk about how users might discover your data in ScienceBase, options for organizing data to support user access and exploration, and how to improve accessibility of your data in ScienceBase.

The ScienceBase Framework



First, let's take a look at the ScienceBase framework. ScienceBase is a robust data platform developed to provide shared, permission-controlled access to scientific data products and Bureau resources. Rather than serving as a generic online storage location, ScienceBase is designed to add value to digital data by exposing well-organized, documented datasets and scientific information over the web. Content within ScienceBase is stored within a standardized item model with consistent informational facets such as title, abstract, keywords, etc. Each ScienceBase item consist of data, which can be a single file or a collection of files, as well as an XML metadata record that documents the data. Another layer of information is added to that original content, such as relationships with other items, permissions, and item audit history. Each item exists independently at a URL endpoint and can have a digital object identifier assigned to it. An item can also be related to other ScienceBase resources and connected to other processes and systems. The ScienceBase system is, therefore, an interconnected network of items.

The ScienceBase Framework



ScienceBase
Item



The screenshot shows the ScienceBase Catalog interface. At the top, the USGS logo and 'ScienceBase-Catalog' are visible. The main heading is 'Missouri Breaks Project, Montana - Digitized aeromagnetic data'. Below this, there are sections for 'Dates' (Publication Date: 2016-05-24, Time Period: 1981-01-01), 'Citation' (Anderson, E.D., Parks, H.L., Jenkins, M.C., Nguyen, D.M., Hearn, B.C., Jr., and Zientek, M.L., 2016. Missouri Breaks Project, Montana - Digitized aeromagnetic data. U.S. Geological Survey data release, http://dx.doi.org/10.5066/7F7TH8JTN), 'Summary' (From February 12 to March 8, 1981, EG and G Geometrics conducted an aeromagnetic survey in Montana for Anaconda Copper Company. A Piper Navajo aircraft was used to conduct the survey. The survey was flown along north-south flightlines spaced 500 feet at a nominal height of 450 feet above the terrain. Five uniformly spaced east-west tie-lines were also flown. During processing, the International Geomagnetic 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 2016, the map sheets were scanned, registered, and the contours were digitized. The point data released here were extracted from the digitized contour lines and include: latitude, longitude, and magnetic-field value (gamma).), 'Contacts' (Point of Contact: Eric D Anderson; Originator: Eric D Anderson, Heather L Parks, Michael C Jenkins, Duc M Nguyen, B Carter Hearn Jr., Michael L Zientek; Metadata Contact: Eric D Anderson), 'Map' (a small map showing the survey area), 'Communities' (USGS Data Release Products), and 'Tags' (GIS, aeromagnetic data, aeromagnetic surveying, airborne surveys, field monitoring stations, geophysical surveys, geophysics, geospatial datasets, magnetic, magnetic field (earth), magnetic surveys, residual magnetic field, total field). The location is listed as Blaine County, Dodson Quadrangle, Fergus County, Healy Quadrangle, Lewistown.

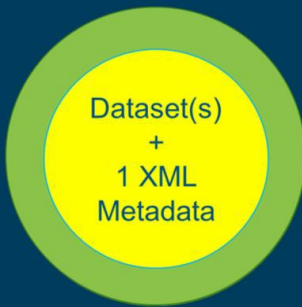
Landing Page for your Data

When thinking about data releases in ScienceBase, an item is essentially a landing page that provides users access to your data. The landing page will include information from your metadata such as the title, citation, abstract, purpose, contacts, dates, and keywords. From here on out, we will refer to ScienceBase items as pages.

Determine how you want users to discover the data in ScienceBase

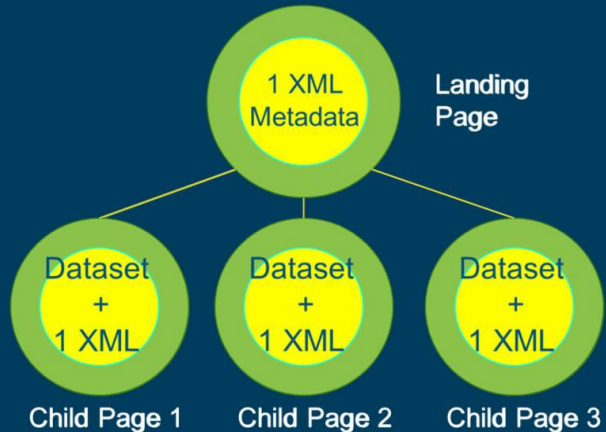


Single Package



Landing Page

Individual Datasets



Each ScienceBase page is individually discoverable and, therefore, each page should have one metadata record describing the available resource. ScienceBase pages are indexed by Google, and ScienceBase will send one metadata record per page to the USGS Science Data Catalog.

If you believe that your data release should be discovered as a single package, you will likely want to have a single landing page, with a metadata record describing the entire data release and all of the data available on that page.

If you believe that the components of your data release should be discovered individually, then you will likely want to have a single landing describing the entire data release and a nested child page for each component dataset and its metadata.

Next, we'll take a look at some examples of ScienceBase data releases and discuss how they were structured and why.

Example 2

One ScienceBase Page



Expert opinions of demographic rates of Argentine black and white tegus in South Florida

Metadata:

- Identification Information
- Data Quality Information
- Entity and Attribute Information
- Identification Information
- Metadata Information

Identification Information:

Citation:

Citation Information:

Originator: Fred A. Johnson
Publication Year: 2017
Title: Expert opinions of demographic rates of Argentine black and white tegus in South Florida
Geospatial Data Presentation Format: Tabular digital data (CSV)
Publication Information:
Publication Place: Gainesville, FL
Publisher: U.S. Geological Survey
Online Linkage: <https://doi.org/10.5066/F7ZC810J>

Description:

Abstract:

We illustrate the utility of expert elicitation, explicit recognition of uncertainty, and the value of information for directing management and research efforts for invasive species, using tegus (*Ninia diademata*) in southern Florida as a case study. We posited a post-birth pulse...

Attached Files

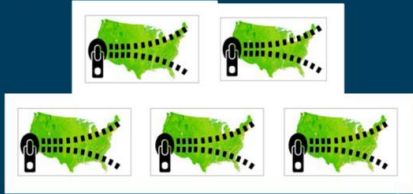
- ExpertFinalAge.csv 168 Bytes
- ExpertFinalParams.csv 14.5 KB

<https://doi.org/10.5066/F7ZC810J>

Example number two is a data release with two comma-separated values or CSV files with one metadata record describing them on a single landing page. The metadata record adequately captures information for both of the CSV files. For instance, the metadata has two entity and attribute sections. The data producer wants these two datasets to be discovered and accessed together. Click on the DOI at the bottom of the page to see this example in ScienceBase.

Example 3

5 Zipped GeoTIFFs with Metadata



One Top Level Metadata Record

Mapping enhanced grazing potential based on the NAWQA Wall-to-wall Anthropogenic Land-use Trends (NWALT) product, 1974-2012

Example:
<http://dx.doi.org/10.5066/F7C R5RFF>

Metadata:

- Identification Information
- Data Quality Information
- Spatial Data Organization Information
- Spatial Reference Information
- Temporal and Annotative Information
- Distribution Information
- Metadata Reference Information

Identification Information:

Citation:
Citation Information:
Originator: James A. Falcione
Publication Date: 2016
Title:
Mapping enhanced grazing potential based on the NAWQA Wall-to-wall Anthropogenic Land-use Trends (NWALT) product, 1974-2012
Geospatial Data Presentation Form: Raster Digital Data Set
Publication Information:
Publication Place: Reston, Virginia
Publisher: U.S. Geological Survey
Online Linkage: <http://dx.doi.org/10.5066/F7C R5RFF>



One ScienceBase Page



USGS
ScienceBase Catalog

System -> USGS Data Release Products -> Mapping enhanced grazing p...

Mapping enhanced grazing potential based on the NAWQA Wall-to-wall Anthropogenic Land-use Trends (NWALT) product, 1974-2012

Dates
Publication Date: 2016
Start Date: 1974
End Date: 2012

Citation
Falcione, J.A., 2016. Mapping enhanced grazing potential based on the NAWQA Wall-to-wall Anthropogenic Land-use Trends (NWALT) product, 1974-2012. U.S. Geological Survey data release. <http://dx.doi.org/10.5066/F7C R5RFF>

Summary
This dataset provides an additional "Grazing Potential" land use class to the previously published U.S. Geological Survey (USGS) National Water-Quality Program (NAWQA) Wall-to-wall Anthropogenic Land-use Trends (NWALT) product (Falcione, 2015; USGS Data Series 145). As with the NWALT, the dataset consists of five national 60-m land use grids, for the years 1974, 1982, 1992, 2002, 2012. The only change to the dataset is, for every year, some pixels which are coded "0" (no wall) in the NWALT, are reclassified to a new class 46 "Grazing Potential Expected". The purpose of the reclassification is to identify areas which are likely to have had at least some grazing activity based on agreement of historical land cover/use datasets, and not already captured in another land use class by the original NWALT. The reclassification occurred as follows: pixel would otherwise be in class 00 (Low Use), 10 (an Agriculture or Grassland class in Macomber and Anderson (1982)), 16 (an Agriculture or Rangeland class in 1970s-era data), and 18 (a Grassland/Herbaceous class (7)) in the NLCD 2011, without restrictions to proximity to water or slope.

Falcione, J.A., 2015. U.S. continental wall-to-wall anthropogenic land use trends (NWALT), 1974-2012. U.S. Geological Survey Data Series 145, 33 p., plus appendices 3-4 as separate files. <http://dx.doi.org/10.3133/D145>

Macomber, F.J. and Anderson, J.R., 1987. Major land uses in the United States. U.S. Geological Survey. <http://water.usgs.gov/GIS/computerprograms/MajorUS/index.html>

Contacts
Point of Contact: James A. Falcione
Distributor: U.S. Geological Survey - ScienceBase
Metadata Contact: U.S. Geological Survey
Originator: James A. Falcione
Publisher: U.S. Geological Survey

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

| | |
|--|-----------|
| class_descriptions_expanded_grazing_april_2016.xlsx | 13.45 KB |
| class_descriptions_expanded_grazing_april_2016.xlsx | 13.45 KB |
| nwalt-grazing-nwalt-grazing-potential.zip | 378.79 KB |
| nwalt1974_v105015.zip | 346.33 MB |
| nwalt1982_v105015.zip | 317.72 MB |
| nwalt1992_v105015.zip | 317.72 MB |
| nwalt2002_v105015.zip | 317.68 MB |
| nwalt2012_v105015.zip | 317.63 MB |
| Mapping_enhanced_grazing_potential_NWALT_v105015.xml | 18.81 KB |

Origin: Unknown

Map >

Communities

Tags

Categories: Data
Theme: Grazing, Land cover, Land use, Land-use trends, Raster data
Place: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Connecticut US, Delaware, District of Columbia, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming
Types: Citation
Provenance
Data source: Input directly

Example number three is a data release with five geotiffs, each with embedded metadata, uploaded as separate zip files and a top level metadata record describing the whole data release. The data producer wants all of these datasets to be discovered together so he included them on the same landing page; however, the geotiffs are relatively large so he uploaded them as five separate zip files so that users could access them separately. The author uploaded an image file to provide a visual representation of the data. Visit the DOI shown on the bottom of the screen to explore this data release in ScienceBase.

Example 4

One Top Level Landing Page...



...with Two
Child Pages

Example:
<http://dx.doi.org/10.506/6/F7W9577Q>



System -- USGS Data Release Products -- Ozark Plateaus seepage run ...

Ozark Plateaus seepage run dataset, southern Missouri and northern Arkansas, 1982-2006

Go to: View

Point of Contact: U.S. Geological Survey Lower Mississippi-Gulf Water Science Center

Dates
Publication Date: 2015

Citation
Kotze, K.J., Wagner, D.M., Roland, V.L., and Nottmeier, A.M., 2015, Ozark Plateaus seepage run dataset, southern Missouri and northern Arkansas, 1982-2006. U.S. Geological Survey data release. <http://dx.doi.org/10.506/6/F7W9577Q>.

Summary
This dataset compiled 24 years of U.S. Geological Survey (USGS) seepage-run data from 15 studies covering southern Missouri and northern Arkansas in the Ozark Plateaus Physiographic Province (Ozarks). Previously these data were only available separately from the original USGS studies. The dataset can be used in surface-water and groundwater investigations assessing water quantity, quality, and availability in the Ozarks. The dataset includes "Ozark Plateaus seepage flow-line dataset, southern Missouri and northern Arkansas, 1982-2006" and "Ozark Plateaus seepage point dataset, southern Missouri and northern Arkansas, 1982-2006".

Seepage-run data were collected by measuring discharge at points along a stream, as well as at tributaries and springs contributing water to that stream. Any differences between the upstream and downstream discharge measurements (greater than the measurement error and taking tributaries/springs into account) were interpreted as gaining streams (if downstream discharge was greater than upstream discharge) or losing streams (if downstream discharge was less than upstream discharge).

The discharge data were originally collected as part of numerous USGS surface-water and groundwater studies to quantify the amount of water gained or lost by streams to aquifers; in some cases the data were collected as part of [show more](#) ...

Map >>

Spatial Services
ScienceBase WMS: <https://www.sciencebase.gov/catalog/>

Communities
USGS Data Release Products @ USGS Lower Mississippi-Gulf Water Science Center
USGS Water/Groundwater

Tags
Types: Map Service, OGC WFS Layer, OGC WMS Layer, OGC WMS Service

Provenance
Data source: Input directly

Child Items (2)

- Ozark Plateaus seepage flow-line dataset, southern Missouri and northern Arkansas, 1982-2006
- Ozark Plateaus seepage point dataset, southern Missouri and northern Arkansas, 1982-2006

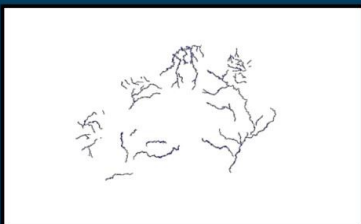
Contacts
Point of Contact: U.S. Geological Survey Lower Mississippi-Gulf Water Science Center, Katherine J. Kotze

Examples one through three were all data releases that the authors want to be discovered as a single package and, therefore, were all included on one landing page.

Example number four is a data release with a top level landing page describing the whole data release and a child item for each of the component datasets. This example is similar to Example number 3 in that there is a metadata record for each dataset; however, in this example, the data producer wants users to be able to discover and access the two datasets individually. The data producer enables this type of independent discovery by having a separate page and metadata record for each dataset.

Example 4

One Shapefile with Metadata



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

Metadata:

- [Line Features Information](#)
- [Data Dictionary Information](#)
- [Spatial Data Organization Information](#)
- [Spatial Reference Information](#)
- [Units and Coordinate System Information](#)
- [Data Dictionary Information](#)
- [Metadata Information](#)

Identification Information:

Citation:

Original Author: Katherine J. Korte
 Original Author: David M. Wagner
 Original Author: Victor L. Sobotik
 Original Author: Anna M. Nordin
 Publication Date: 03/05/2013

Title:
 Ozark Plateaus seepage flow-line dataset, southern Missouri and northern Arkansas, 1982-2006

Conceptual Data Processing/Production Name: Vector Digital Data Set (VDD-DS)
Order Number: <http://dx.doi.org/10.5066/F7W9577Q>
Large File Citation:

Client Information:

Original Author: Katherine J. Korte
 Original Author: Victor L. Sobotik



ScienceBase Child Page #1



USGS
 ScienceBase Catalog

System - USGS Data Release Products - Ozark Plateaus seepage flow - Ozark Plateaus seepage flow...

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

Dates

Publication Date: 2013
 Start Date: 1982-03-05
 End Date: 2006-11-14

Citation

Korte, K.J., Wagner, D.M., Sobotik, V.L., and Nordin, A.M. 2013. Ozark Plateaus seepage flow dataset, southern Missouri and northern Arkansas, 1982-2006. U.S. Geological Survey data release. <http://dx.doi.org/10.5066/F7W9577Q>.

Summary

The dataset was compiled to summarize 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 numerous USGS studies to assess interactions between streams and groundwater aquifers. A spring stream is described as a surface-water stream that gains water from the groundwater aquifer and a losing stream is described as a surface-water stream that loses water to the groundwater aquifer. This project is expected to be used in surface-water and groundwater investigations assessing water quantity, quality, and availability. The product includes flow-line data organized into a feature class for each stream reach. Each reach is a line representing discharge measurements along the stream reach. The flow-line data includes a coordinate system, a projection, and units. The flow-line data is organized into a feature class for each stream reach. The flow-line data is organized into a feature class for each stream reach. The flow-line data is organized into a feature class for each stream reach.

Spatial Services

ScienceBase WMS:
[WMS \(non-arcgis.com\)](#)
 ScienceBase WFS:
[WFS \(non-arcgis.com\)](#)

Communities

USGS Data Release Products

Tags

Categories: Data
 Theme: Groundwater Resources Program, discharge, gaining streams, groundwater, hydrogeology, hydrographs, IADL, losing streams, seepage flow, surface water
 Place: Arkansas, Missouri, Ozark Aquifer System, Ozark Physiographic Province, Ozarks, Ozark region
 Type: Data Dictionary, Map Service, OGC WFS Client, OGC WMS Layer, Shapefile

Provenance

Data source: input directly

Attached Files

| | |
|---|-----------|
| Ozark_plateaus_seepage_flow_lines.zip | 2.09 MB |
| Ozark_plateaus_seepage_flow_lines.apr | 1.28 MB |
| Ozark_plateaus_seepage_flow_lines.shp | 6.12 MB |
| Ozark_plateaus_seepage_flow_lines.prj | 400 Bytes |
| Ozark_plateaus_seepage_flow_lines.apr | 234.36 KB |
| Ozark_plateaus_seepage_flow_lines.apr | 7.09 MB |
| Ozark_plateaus_seepage_flow_lines.apr | 500 Bytes |
| Ozark_plateaus_seepage_flow_lines.apr | 8 Bytes |

Example: <http://dx.doi.org/10.5066/F7W9577Q>

The data producer also wants to distribute these two shapefile datasets as separate web services in addition to simple downloads. To create separate web services for each dataset, the data producer needed to upload the different datasets to separate child pages. On this slide, you can see one of the child pages hosting one of the shapefiles. Web mapping and web feature services have been generated from the uploaded shapefile.

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



As we saw in some of the previous examples, ScienceBase has some functionality that can help to improve the accessibility and distribution of your data. ScienceBase supports

- Basic data downloads;
- Creation of web mapping services to enable visualization of data both in ScienceBase as well as in external, interactive mapping applications;
- Creation of web feature services, web coverage services, and ArcGIS REST services to enable users to query and analyze the data without downloading them; and
- API access to ScienceBase 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>



To recap, in this tutorial we discussed considerations for structuring and documenting a public USGS data release, including considerations around discovery, access and exploration, and distribution. We also reviewed the ScienceBase system and some examples of data releases in ScienceBase. If you are ready to get started with a data release in ScienceBase, click the link on the bottom of your screen to be directed to the ScienceBase data release instructions page.

Still have questions?

Contact the ScienceBase team at sciencebase@usgs.gov



Do you still have questions about how to structure and document your data release?

The ScienceBase team is happy to discuss your options with you! Just send an email to sciencebase@usgs.gov.