USGS
science for a changing world

Prepared in cooperation with the Community for Data Integration

Sharing Our Data—An Overview of Current (2016) USGS Policies and Practices for Publishing Data on ScienceBase and an Example Interactive Mapping Application

Open-File Report 2016–1202

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

Sharing Our Data—An Overview of Current (2016) USGS Policies and Practices for Publishing Data on ScienceBase and an Example Interactive Mapping Application

By Katherine J. Chase, Andrew R. Bock, and Roy Sando

Prepared in cooperation with the Community for Data Integration

Open-File Report 2016–1202

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Director

U.S. Geological Survey, Reston, Virginia: 2017

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/store.usgs.gov/.

Suggested citation:
Chase, K.J., Bock, A.R., and Sando, Roy, 2017, Sharing our data—An overview of current (2016) USGS policies and practices for publishing data on ScienceBase and an example interactive mapping application: U.S. Geological Survey Open-File Report 2016–1202, 10 p., https://doi.org/10.3133/ofr20161202.

ISSN 2331-1258 (online)

Acknowledgments

The data used for the example in this report were generated under a project funded by the Plains and Prairie Potholes Landscape Conservation Cooperative and the Bureau of Land Management. The authors appreciate the support of those agencies.

The authors also thank colleague reviewers Drew Ignizio, Tamar Norkin, and John Risley for their edits and suggestions that greatly improved this report.

Contents

Figures

Abbreviations

API	application program interface
ASCII	American Standard Code for Information Interchange
CDI	Community for Data Integration
CIDA	Center for Integrated Data Analytics
CRAN	Comprehensive R Archive Network
CSAS&L	Core Science Analytics, Synthesis and Libraries
CSS	cascading style sheets
CSV	comma separated value
DOI	digital object identifier
FGDC	Federal Geographic Data Committee
FSP	Fundamental Science Practices
GCM	General Circulation Model
GIS	geographical information system
HTML	hypertext markup language
IM	Instructional Memorandum
IPDS	Information Product Data System
KML	keyhole markup language
NetCDF	network common data form
OGC	Open Geospatial Consortium
OWI	Office of Water Information
PDF	portable document format
REST	representational state transfer
SD	service definition
SDC	Science Data Catalog
URL	uniform resource locator
USGS	U.S. Geological Survey
WFS	Web Feature Services
WMC	Web coverage services
WMS	Web mapping services
XML	extensible markup langua

Sharing Our Data—An Overview of Current (2016) USGS Policies and Practices for Publishing Data on ScienceBase and an Example Interactive Mapping Application

By Katherine J. Chase, Andrew R. Bock, and Roy Sando

Abstract

This report provides an overview of current (2016) U.S. Geological Survey policies and practices related to publishing data on ScienceBase, and an example interactive mapping application to display those data. ScienceBase is an integrated data sharing platform managed by the U.S. Geological Survey. This report describes resources that U.S. Geological Survey Scientists can use for writing data management plans, formatting data, and creating metadata, as well as for data and metadata review, uploading data and metadata to ScienceBase, and sharing metadata through the U.S. Geological Survey Science Data Catalog. Because data publishing policies and practices are evolving, scientists should consult the resources cited in this paper for definitive policy information.

An example is provided where, using the content of a published ScienceBase data release that is associated with an interpretive product, a simple user interface is constructed to demonstrate how the open source capabilities of the R programming language and environment can interact with the properties and objects of the ScienceBase item and be used to generate interactive maps.

Introduction

The ways scientists view and share data are changing (Fristom, 2013; Tenopir and others, 2011; Tenopir and others, 2015). New tools, such as Web map services available through U.S. Geological Survey (USGS) ScienceBase (https://www.sciencebase.gov/about/content/view-data-using-sciencebase-web-map-services-wms), can be used to generate interactive maps. In addition, Federal and state agencies, universities, and other organizations are working together to make water resources data more accessible and interconnected through groups such as the Advisory Committee on Water Information (http://acwi.gov/spatial/owdi/). However, it can be difficult to find guidance about organizing, sharing, and displaying data through interactive maps.

USGS Fundamental Science Practices (FSP) policies were established in 2006 to ensure the quality and integrity of USGS science activities. All USGS data intended for public release are subject to rigorous FSP review, approval, and release requirements (http://www.usgs.gov/fsp/). Fundamental Science Practices requirements have been recently (2015) revised to incorporate release of scientific data and metadata (http://www.usgs.gov/fsp/faqs_datarelease.asp, https://www2.usgs.gov/fsp/guide_to_datareleases.asp).

ScienceBase is a collaborative scientific data and information management platform developed and maintained by the USGS (https://www.sciencebase.gov/about/content/about-sciencebase). ScienceBase is designed to provide scientists and the general public access to scientific data, Bureau resources, and other information through access using a Web browser, an application program interface (API), and through representational state transfer (REST) -based Web services. ScienceBase is one option for USGS researchers to use as a platform for posting finalized data release materials (reviewed data and associated metadata).

This report describes resources that U.S. Geological Survey Scientists can use for writing data management plans, formatting data, and creating metadata, as well as for data and metadata review, uploading data and metadata to ScienceBase, and sharing metadata through the U.S. Geological Survey Science Data Catalog. Because data publishing policies and practices are evolving, scientists should consult the resources cited in this paper for definitive policy information.

For this report, simulated streamflow data published on ScienceBase (Chase and others, 2016b) were used as an example illustrating one way to develop interactive maps. Chase and others (2016a, b) estimated streamflow for baseline and future conditions for select locations in eastern and central Montana, using temperature and precipitation data from a gridded data set representing baseline conditions and from three different General Circulation Models (GCMs). The data were summarized in static plots, maps, and tables in the companion interpretive product (Chase and others, 2016a). However, these large datasets could be better visualized with interactive maps, where a user could select a time period and GCM of interest to display.

Purpose and Scope

This report provides an overview of current (2016) U.S. Geological Survey policies and practices related to publishing data on ScienceBase, and an example interactive mapping application to display data associated with an interpretive product. Because U.S. Geological Survey data publishing policies and practices are evolving, scientists should consult the resources cited in this paper for definitive policy information. In addition, USGS Web sites could change, and the information, uniform resource locators (URLs), and email addresses provided in this report could be superseded.

Publishing Data on Sciencebase

U.S. Geological Survey scientists intending to publish data on ScienceBase are encouraged to read through the ScienceBase data release Web pages (https://www.sciencebase.gov/about/content/data-release) and reference the workflow diagram (fig. 1). Before uploading data to ScienceBase, data should be organized in a format and structure that will best lend itself to meet the needs of a project, and metadata should be created. Data and metadata must undergo USGS-initiated review, and documentation of the data and metadata reviews must be uploaded into the Information Product Data System (IPDS). After the data and metadata reviews are completed and reconciled, data and metadata are uploaded to ScienceBase, and the ScienceBase Team will create a digital object identifier (DOI; a persistent identifier assigned to a digital object—such as a report or published dataset—so that the object can be located even if the object is moved). Finally, after Bureau approval and publication, the metadata file associated with the data can be sent to the USGS Science Data Catalog (discussed later in this section).

ScienceBase and ScienceBase Items

ScienceBase (https://www.sciencebase.gov/about/) is a scientific data and information management platform where USGS scientists can post data and enable permission-controlled shared access; items can be accessible only to a limited number of users or made publicly available, as appropriate. ScienceBase uses the basic informational unit of an "item" to describe resources in the ScienceBase catalog. Items may describe datasets, projects, publications, maps, or other informational assets. All items share a core set of attributes based on a standardized object model. Items can be used to store and share uploaded data, metadata, or links to online resources. Depending on how items are stored and structured within the system, ancestral relationships between "parent items" (items that contain one or more other items) and "child items" (items that are located inside of another parent item) can be used to organize resources and present context. For example, a parent

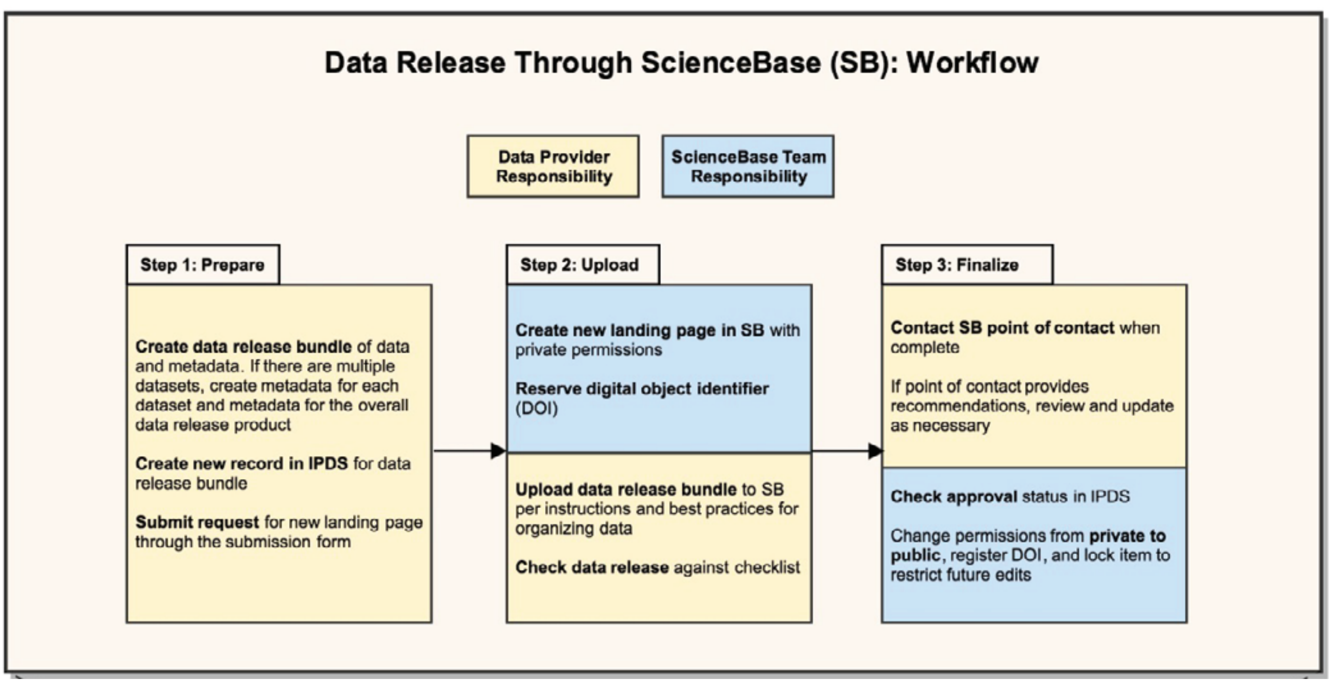

Figure 1. Current (2016) workflow for preparing a ScienceBase data release. From the ScienceBase Data Release Frequently Asked Questions Web page at https://www.sciencebase.gov/about/content/data-release.

folder can contain a child subfolder, and the child subfolder can be a parent item to a child record.

Data Management Plan

Well-documented data are important to USGS scientists and cooperators, can be integrated more readily into projects and datasets, and are more legally and scientifically defensible than undocumented data (USGS Data Management Web site; http://www2.usgs.gov/datamanagement/index.php). Since 2006, USGS Fundamental Science Practices (https://www2.usgs.gov/usgs-manual/500/502-2.html) have required documentation of data collection and research activities. U.S. Geological Survey Instructional Memorandum (IM) OSQI 2015-01 (http://www2.usgs.gov/usgs-manual/im/IM-OSQI-2015-01.html), released February 19, 2015, provides guidance for managing and sharing data and establishes requirements that all projects initiated on or after October 1, 2016, have Data Management Plans, which help ensure that data published by the USGS are well-documented and archived.

The Data Management Plan should be produced during project scoping and should be included in the proposal. The plan should describe data handling throughout the data lifecycle, including data acquisition, data processing and analyzing, data preservation, and data publication and sharing. The USGS Data Management Web site (http://www2.usgs.gov/datamanagement/index.php) offers detailed information, templates, and checklists to help project managers create and maintain data management plans.

Data Formats

As described in the 2013 Open Data Policy memorandum from the U.S. Office of Management and Budget (https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf) and the Data and File Formats page on the USGS Data Management Web site (https://www2.usgs.gov/datamanagement/plan/dataformats.php), non-spatial data should be in open, machine-readable formats, such as comma separated value (CSV), extensible markup language (XML), text, or American Standard Code for Information Interchange (ASCII). Spatial data should be formatted as shapefiles, geodatabase feature classes, rasters, or network common data form (NetCDF) files. Recommended formats for spatial data are discussed further in the section “Creating Interactive Maps.”

Metadata

Metadata (data about data) are required for all USGS scientific data products. Metadata files include documentation that describes a data resource, how it was produced, and other relevant information related to understanding its appropriate use. Policy regarding USGS metadata requirements is provided in USGS IM OSQI 2015-02 (http://www2.usgs.gov/usgs-manual/im/IM-OSQI-2015-02.html). For ScienceBase data releases, metadata should be in XML format. Tools for creating metadata are available on the USGS Data Management Metadata Web page (http://www2.usgs.gov/datamanagement/describe/metadata.php). If more than one dataset will be published, it is often advantageous to have a project-level XML metadata record that describes the overall data release (including abstract information, keywords, and a general overview of the published data to provide context), as well as individual XML metadata records for each dataset or ordered assemblage of files that describe column values, analytical methods, and details about appropriate use.

For creation of metadata for the example used in this report, we used the Online Metadata Editor (https://www1.usgs.gov/csas/ome/), which lists simple questions about a dataset and then produces a metadata record in the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (http://www.fgdc.gov/csdgm-graphical/index.html), based on the user input. In addition, the metadata reviewer for the example used in this report (see the section “The Information Product Data System and Data and Metadata Review” below) used the USGS Geospatial Metadata Validation Service (http://geo-nsdi.er.usgs.gov/validation/) to check the metadata. An open-source word processor, such as Notepad++ (https://notepad-plus-plus.org), with an XML formatting plug-in, such as XML Tools (http://docs.notepad-plus-plus.org/index.php?title=Plugin_Development), can be used to edit metadata files after they have been created. XML Notepad is another free editing tool that can be used to check, view, and edit metadata (https://www2.usgs.gov/datamanagement/documents/InstructionsXMLNotepad.pdf).

The Information Product Data System and Data and Metadata Review

Per USGS IM OSQI 2015-03 (http://www2.usgs.gov/usgs-manual/im/IM-OSQI-2015-03.html), which “provides requirements and procedures for review and approval of scientific data prior to release or dissemination,” at least two USGS-initiated reviews are required for a data release; one review of the data and one review of the metadata. Both reviews can be by the same person, or one person can review the data and another can review the metadata. However, reviewing the data and metadata together is best, as this helps ensure that the documentation adequately and accurately describes the data and that there are no discrepancies between the two. Guidance for checking metadata with data can be found on the USGS data management Web site: https://www2.usgs.gov/datamanagement/documents/CheckingMetadataWithData.pdf. The metadata and data reviews must be placed in the USGS Information Product Data System (IPDS; https://ipds.usgs.gov/WPP/Home.aspx), under the product type “Data Release.”

Data and Metadata to ScienceBase

Scientists wishing to publish data on ScienceBase can reference the current (2016) workflow (fig. 1) and initiate contact with the ScienceBase Team by filling out a Data Release Submission Form (link available on the Data Release Instructions Web page https://www.sciencebase.gov/about/content/data-release; note that the ScienceBase Team is currently (2016) working on a Web application to replace the Data Release Submission Form). The ScienceBase Team can assist authors with establishing an item in the system to store the data (fig. 2), and acquiring a digital object identifier (DOI) that will resolve to the ScienceBase item and serve as permanent way to reference the data resource. The data and metadata will not be publicly accessible until the data release is approved and published. During the period when the data release is in progress, the author can set up permissions for USGS individuals (and authorized individuals outside the USGS) to access and edit the ScienceBase items. In addition, an anonymous link can be created for unauthorized individuals outside of the USGS (such as colleague reviewers) to access the data. Current (2016) information about these options and much more can be found on the ScienceBase Data Release Instructions Web page (https://www.sciencebase.gov/about/content/data-release).

Science Data Catalog

The USGS Science Data Catalog (SDC; http://data.usgs.gov/datacatalog/), developed by USGS Core Science Analytics, Synthesis and Libraries (CSAS&L; https://www2.usgs.gov/core_science_systems/csas/), is a searchable clearinghouse for XML metadata records that describe USGS data resources and specify where they can be obtained. The SDC is separate from ScienceBase, and data do not reside in the SDC.

Figure 2. Example of a ScienceBase item (Chase and others, 2016b).

The XML metadata file associated with data on ScienceBase can be sent to the SDC from ScienceBase. This metadata record can then be searchable in the SDC and provide a direct link back to the data in ScienceBase. Questions about the SDC or about submitting XML metadata records can be emailed to sciencedatacatalog@usgs.gov.

Creating Interactive Maps

Interactive maps can be created with data from ScienceBase using ScienceBase geospatial services and relatively simple programs with open-source capabilities such as R. Other options for serving interactive maps include Web portals, interactive portable document format (PDF) files, and Esri Story Maps.

ScienceBase Geospatial Services

ScienceBase offers users several choices of geospatial services that can be used to display the shapefiles or spatial content of data releases. Users can upload shapefiles, rasters, and geodatabases. In addition, ScienceBase supports ArcGIS mapping services through Service Definition (SD) files. Methods for using these different formats can be found in the online ScienceBase documentation (https://www.sciencebase.gov/about/content/sciencebase-geospatial-services).

ScienceBase can generate both Web mapping services (WMS) and Web feature services (WFS) for an uploaded shapefile containing vector data. A WMS allows a user to view the data in clients such as ArcGIS desktop, or a browser application (Open Geospatial Consortium, Inc., 2006). It is a ‘picture’ of the data, but the actual data are not included. On the other hand, a WFS provides feature level access to geospatial data, and once properly connected with an appropriate client, the data can be retrieved, updated, and modified directly in GIS software (Open Geospatial Consortium, Inc., 2005). Clients that can access the WFS format include ArcGIS desktop, and various open source programs such as QGIS (an open source geographic information system, http://www.qgis.org/en/site/) and R (R Core Team, 2016). For the example used in this report, Web Feature Services (WFS) were chosen because the data being mapped were vector data, and because use of open source programs and USGS-supported tools (see discussion under “R, sbtools, Shiny, and the Leaflet R Package” below) best fit the project needs.

Additional ScienceBase capabilities (not used in this example or further discussed in this report) include WMS for raster data, Web Coverage Services (WCS) for raster data, and generation of keyhole markup language (KML) files (for both raster and vector data). Web Coverage Services for raster files are analogous to Web Feature Services (WFS) for vector data. The KML is an xml-based file format used to represent geographic features (Open Geospatial Consortium Inc., 2009). Applications that use the KML format include Google Earth and ArcGIS Explorer.

If ScienceBase geospatial or map services are needed, shapefiles must include a projection (PRJ) file and rasters must be in the GeoTiff (TIF) file format (tagged image file format). For faster map service operability the projection should be in a geographic coordinate system (as opposed to a projected coordinate system). The Web Mercator WGS84 coordinate system (http://spatialreference.org/ref/epsg/wgs-84/) is used for many Web mapping applications (including Google and ArcGIS Server). Ultimately, the coordinate system should be appropriate for the data, supported in the uses or applications envisioned for the data, and well documented in the metadata, with a projection (.prj) file included in the data.

Using R, Sbtools, Shiny, and the Leaflet R Package to Interact with ScienceBase Resources

The ScienceBase application program interface (API) allows software or computer programs to interact programmatically with data and geospatial services on ScienceBase. R, sbtools, Shiny, and Leaflet are utilities that can be used to leverage the ScienceBase API for interactive mapping.

R is a computer language and a free software environment for statistical computing and graphics (R Core Team, 2016). For the interactive maps used as an example for this report, R was used to access and map the ScienceBase item (as described in more detail below). R is supported by a number of USGS communities for data access and statistical analysis. No licenses are required to run R.

R can be extended via packages, or sets of code, which are freely available at the Comprehensive R Archive Network, or CRAN (https://cran.r-project.org/mirrors.html). Scientists with little programming experience can download packages and documentation for performing statistical analyses, plotting, and mapping, and connect with other users through USGS-R, a community of support for R users (http://owi.usgs.gov/R/). U.S. Geological Survey communities such as the Office of Water Information (OWI; http://owi.usgs.gov/) and the Community for Data Integration (CDI; https://www2.usgs.gov/cdi/) have helped to develop a large number of tools and packages using R for accessing USGS data (https://owi.usgs.gov/R/index.html; https://my.usgs.gov/bitbucket/projects).

For this example, the USGS-R package sbtools (https://github.com/USGS-R/sbtools), which contains code/libraries that enable access to online data on ScienceBase, was used to access the data in ScienceBase. The sbtools package, developed and supported by the USGS OWI, allows access into the ScienceBase Web service API, including attached data files, metadata, links to parent and child items and Web feature services (when available), as well as capabilities to manage or edit ScienceBase items.

Shiny is a relatively lightweight (easier to learn and use than Hyper Text Markup Language [HTML] or JavaScript) Web application framework for R for creating interactive Web applications (http://shiny.rstudio.com/). Shiny can be used with a basemap obtained through the Leaflet R package to build an interactive mapping application that can be downloaded to a user's computer. Leaflet (http://rstudio.github.io/leaflet/) is an open-source JavaScript library for interactive maps. Scientists without Web development skills, but with some familiarity with R, can create maps using the Leaflet R package and Shiny applications (https://rstudio.github.io/leaflet/shiny.html).

Example—Maps Showing Simulated Changes in Streamflow for Seven Montana Watersheds

Data Release Used for this Example

For our example, an interactive map application was developed to display data from a previously published ScienceBase data release (Chase and others, 2016b). Chase and others (2016b) includes simulated monthly streamflow data for baseline and future conditions for seven watersheds in eastern and central Montana. The data accompany a journal article that documents the model and results (Chase and others, 2016a).

Preparation of Shapefiles, Data, and Metadata

The ScienceBase data release help materials (https://www.sciencebase.gov/about/content/data-release) include recommendations for structuring data releases as well as helpful Frequently Asked Questions lists and answers. Scientists can consult with the USGS ScienceBase Team if they have any questions regarding the recommendations.

The general instructions for loading shapefiles to ScienceBase are available on the ScienceBase Geospatial services Web page (https://www.sciencebase.gov/about/content/sciencebase-geospatial-services). Upon loading a zipped shapefile to the ScienceBase page and then choosing the "unzip" option, ScienceBase automatically recognizes the shapefile extension, and creates spatial services in the Open Geospatial Consortium (OGC) standard.

For the application described in this report, we used a subset of 185 stream segments from Hydrologic Region 10U available from the Geospatial Fabric for National Hydrologic Modeling (Viger and Bock, 2014). These segments were used as the modeling units in Chase and others (2016a, b). The shapefiles for the 185 stream segments are hosted as a child item of Chase and others (2016b; https://www.sciencebase.gov/catalog/item/5522bdf7e4b027f0aee3d039).

Creating an Interactive Map Using R

This section describes the components of the interactive map developed for this application using the R scientific programming language (R Core Team, 2016). The user can download the R scripts to their computer and run the scripts to create the interactive map used for this example. R was chosen for the map because the streamflow dataset to be mapped (Chase and others, 2016b) is relatively simple, and because the USGS OWI develops and supports R tools. The size of the intended audience of the application (probably less than about 50 users total, with only a few users accessing the application at the same time) also contributed to the decision to use R. For tools and interactive maps that incorporate more complex datasets and are intended for much larger distribution, other approaches might be more appropriate. For examples of other approaches see Blodgett (2013) and the Wyoming Landscape Conservation Initiative Web site (https://www.wlci.gov/wlciMapviewer/).

The R scripts for the application have four specific components (fig. 3). First, access points to the ScienceBase item and the WFS associated with the stream segment shapefiles from the geospatial fabric (Viger and Bock, 2014) are created through the R interface called sbtools (Winslow and others, 2016). Second, the CSV files in appendix 2 from the Chase and others (2016b) ScienceBase data release are retrieved and summarized as data objects within R. Third, a basic interactive map using Leaflet (Cheng and Xie, 2016) is created to connect the data objects with the WFS. And fourth, Shiny (Chang and others, 2016), a java-based Web framework for R, is used to create an interactive interface.

Accessing the Data Using Sbtools

The sbtools package in R uses the ScienceBase item ID to access the properties of a ScienceBase item. For our example we used sbtools to reference the ScienceBase ID (5522bdf7e4b027f0aee3d039) of Chase and others (2016b). Then, we used sbtools to retrieve the model output files in appendix 2 of Chase and others (2016b), to call up the WFS of the stream segments subset from the Upper Missouri Region (Hydrologic Region 10U) of the Geospatial Fabric for National Hydrologic Modeling (Viger and Bock, 2014), and to retrieve watershed boundaries of the seven watersheds from Chase and others (2016a) derived from hydrologic response units of the Geospatial Fabric for National Hydrologic Modeling within the same region (Viger and Bock, 2014). A key file that holds additional identification information for each stream segment is accessed locally by sbtools in the Shiny application directory (*Streamsegments_Qchange_Buffer.csv*).

Manipulating the Data

Each model output file is read directly from the source into an R data object using the RCurl R package (Temple Lang, 2016; https://cran.r-project.org/web/packages/RCurl/

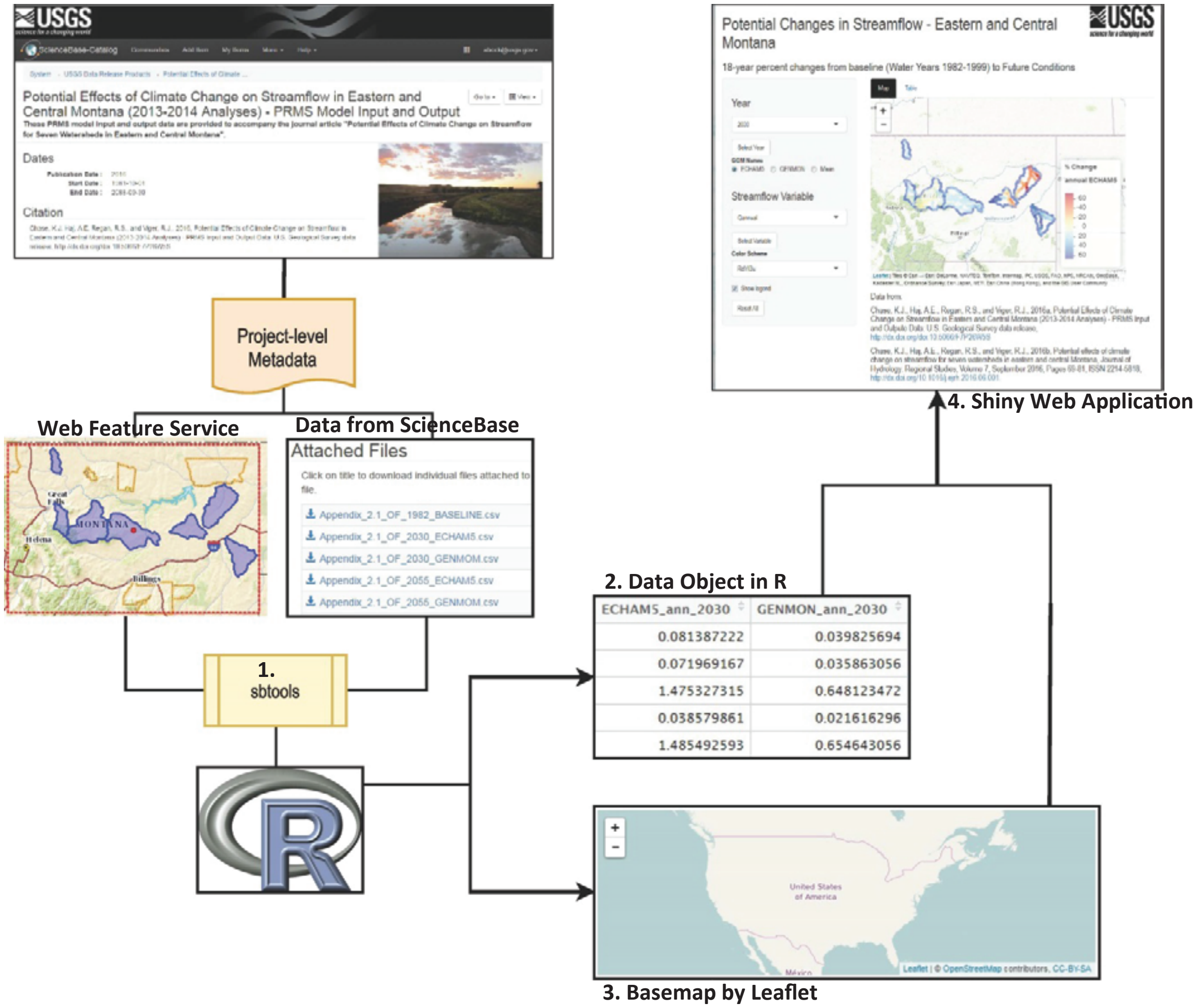

Figure 3. Flowchart illustrating creation of an interactive map using a ScienceBase data release, R, Leaflet, and Shiny. From the ScienceBase data release, the Web feature service (WFS) containing the geographic entities, and data indexed to the geographic entities are brought into R environment through the use of the sbtools package (1). The data can be summarized and directly read as R data objects (2), and the WFS components can be incorporated into open source basemaps (3) utilizing the leaflet library through R. The Shiny Web application framework (4) is then utilized to build a user interface that connects the geographic entities to the underlying data objects summarized in R to allow a simple level of user interaction.

RCurl.pdf). Each file includes simulated monthly streamflow for baseline conditions and for future conditions for each stream segment. The simulated data include four 18-year periods; one baseline period (WY 1982 – 1999) and three future periods (WYs 2021–38, 2046–63, and 2071–88) for multiple GCMs (Chase and others, 2016a, b). For this application, the authors summarized the time series for annual, seasonal, and mean monthly streamflow at each segment for each time period and for each GCM.

After the incorporation of the WFS stream segments and watershed boundaries in R, the features are then re-projected to the WGS84 geographic coordinate system (EPSG:4326) (http://spatialreference.org/ref/epsg/wgs-84/). The WGS84 is used because several of the major map service providers (such as ESRI and GoogleMaps) use slightly different definitions of the WGS84 Web Mercator Auxiliary Sphere GCS (EPSG:3857), and many commonly used map tile services (including *Leaflet*) expect coordinates of vector features to be in latitude-longitude format.

Mapping the Data Using Leaflet

The Leaflet widget used for this example displays polylines representing the Geospatial Fabric stream segments and polygons representing the watershed boundaries (Viger and Bock, 2014) on a background basemap of open source map tiles. Leaflet functionality was combined with interactive capabilities of Shiny (see “Building the User Interface Application

Using Shiny") to symbolize stream segments based on variable combinations selected by the user. The Leaflet map can be modified or updated based on these user requests. The style, color, and representation of the features within the Leaflet widget can be easily modified.

Building the User Interface Application Using Shiny

Shiny is an interactive Web application framework for R that uses HTML, cascading style sheets (CSS), and JavaScript capabilities (Chang and others, 2016). The use of R within the application allows for on-the-fly integration of R statistical or graphical capabilities into a desktop application. The application is composed of two components: the user interface, which contains the widgets and tools that allow user interaction; and the server, which launches the application and executes or dynamically reacts to the user-requests behind it. The Leaflet mapping widget can be easily integrated within a Shiny application.

In this example, we have built an application using the Leaflet/Shiny integration to display data hosted in ScienceBase. This application can interactively and dynamically map simulated changes in streamflow for three future periods for seven watersheds in eastern and central Montana (fig. 4, 185 stream segments in the study area), and generate a table summarizing these changes in streamflow for a specific stream segment.

Documenting the R code using R Markdown and Hosting the R Code in Bitbucket

The R scripts used to produce interactive maps for this project, along with documentation in R Markdown files, are archived in the USGS Bitbucket repository (https://my.usgs.gov/bitbucket/projects/CDI/repos/interactive-mapping/browse). Users with some knowledge of R or with assistance from an R programmer can download and run the R code on their computer to create an interactive map displaying data from Chase and others (2016b) similar to the one shown in fig. 4.

Other Mapping Options

Many other options exist for creating interactive maps, with or without accessing online data. USGS Streamstats is one example of interactive maps served through a Web portal (http://streamstatsags.cr.usgs.gov/streamstats/). Developing a Web portal requires expertise in Web design and programming, as well as an organization to host and maintain the portal. Interactive PDF maps have been published by USGS scientists; see appendix 1 in Thamke and others (2014) for one example (http://pubs.usgs.gov/sir/2014/5047/pdf/sir2014-5047.pdf). Interactive PDF maps only allow the user to switch displays among static, premade maps, and might be useful for displaying a few sets of data. However, creation and storage of several premade maps can be more difficult than writing a program to quickly retrieve and map data. Additionally, users cannot easily extract data from interactive pdf maps. Finally,

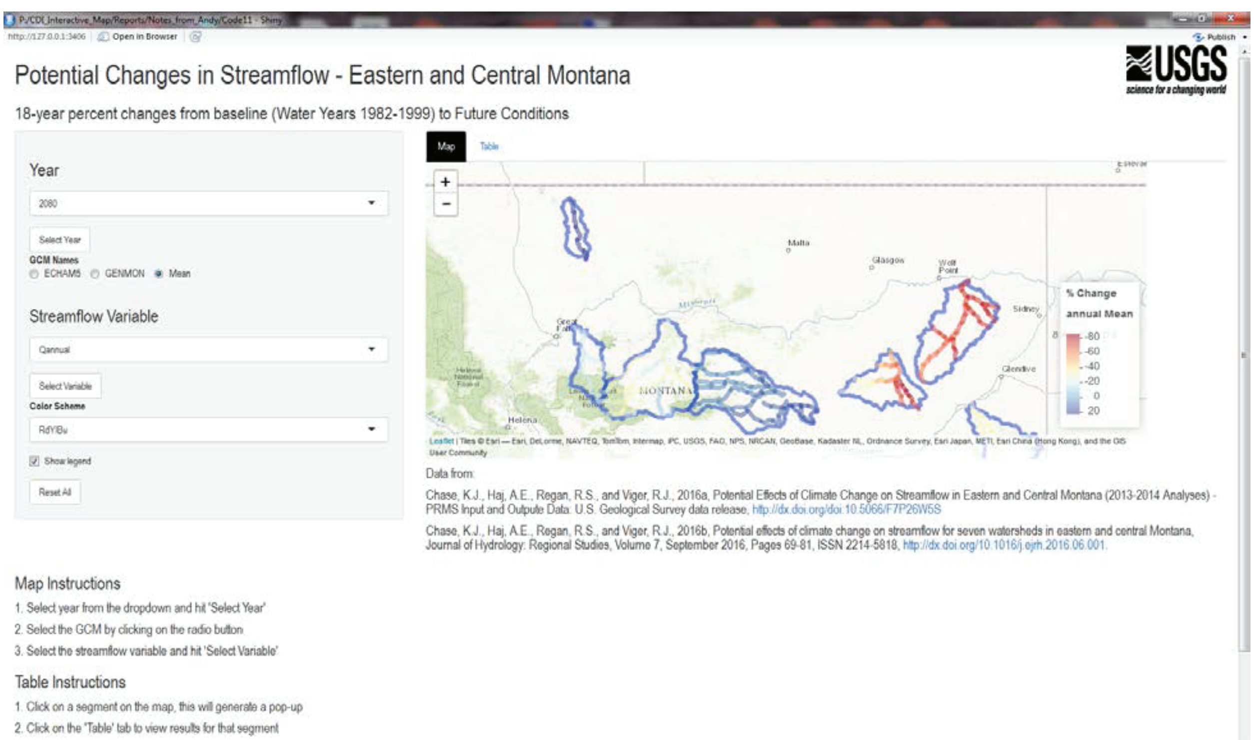

Figure 4. Example of an interactive map showing simulated changes in streamflow for seven watersheds in Montana generated using data released on ScienceBase.

Esri Story Maps let users combine maps, text, images, and multimedia to share data (https://storymaps.arcgis.com/en/). One example is the "100 Years of the National Park Service" storymap (http://story.maps.arcgis.com/apps/MapJournal/index.html?appid=370f50e999534f50922bfb98fbac645f).

Lessons Learned, Limitations and other Considerations

As discussed in the "Purpose and Scope" section, USGS data publishing policies and practices are evolving, and scientists should consult the resources cited in this paper for definitive policy information. In addition, USGS Web sites could change, and the information, uniform resource locators (URLs) and email addresses provided in this report could be superseded.

For the example used in this report, data from an existing data release (Chase and others, 2016b) were mapped. For future mapping efforts and mapping applications, ideally the data would be formatted and structured with input from the map creators and map users.

Both WFS and WMS appear to work more easily with shapefiles with a minimal number of attribute fields (for example, only mandatory Esri fields such as "FID" (feature identifier) and "Shape" fields). If this approach is used, these fields can be linked to data stored in CSV files on ScienceBase. The CSV files should contain the same ID key as the shapefile in the ScienceBase mapping service. All columns and the values they contain should be documented in the associated metadata. Column headings for xyz data should conform to standard coordinate names or abbreviations (latitude/longitude or lat/long). Scientists and programmers should be aware of blank fields and symbols or values that represent "No Data" in ScienceBase items as well as program defaults for "No Data" in software or programming languages used.

Optimizing WFS and WMS requires scientists and programmers to consider the characteristics of their spatial data, including the number of features, the complexity of the features (number of vertices), the projection, symbology (the simpler the better), and the file size. All Web-based services (in ScienceBase or other GIS server platforms) will be limited by network bandwidth and considerations like the above. For optimal performance, users should carefully consider their project goals and ways to make the services more efficient.

The example mapping application created for this project is for distribution to a limited number of map users (probably less than about 50 users total, with only a few users accessing the application at the same time). For more complex datasets, or for applications to be used by hundreds or thousands of people, a Web portal or other enterprise-grade solution might be more appropriate. The USGS OWI Data Science Team (http://cida.usgs.gov/datascience.html) developed the sbtools R package used for our example. They can assist with more complex and expansive mapping projects. Scientists at the USGS Fort Collins Science Center also develop Web mapping applications (https://www.fort.usgs.gov/science-tasks/2326). These teams might be good resources or collaborators for large and/or complex interactive mapping applications.

When planning interactive map projects, scientists might consider the following questions:

1. How many users are expected? How many users might be accessing the data at the same time? High concurrent user volumes can affect performance of the maps and user experience.

2. How long will the application need to be supported? Can the volume of hosted data or the overall user base be expected to grow through time?

3. Who will maintain the map applications? Will user support be available by telephone and email? Who will fix any errors discovered after distribution?

Answers to those questions can help guide selection of methods, tools, and programs used to create and maintain the maps.

Summary

This report provides an overview of current (2016) U.S. Geological Survey (USGS) policies and practices related to organizing and sharing data using ScienceBase. An example interactive mapping application to display data that are associated with an interpretive product is provided.

Before uploading data to ScienceBase, data should be organized in a format and structure that will best lend itself to meet the needs of a project, and metadata should be created. Data and metadata must undergo USGS-initiated review, and documentation of the data and metadata reviews must be uploaded into the Information Product Data System (IPDS). After the data and metadata reviews are completed and reconciled, data and metadata are uploaded to ScienceBase, and the ScienceBase Team will create a digital object identifier. Finally, after Bureau approval and publication, the metadata file (in extensible markup language format) associated with the data can be sent to the USGS Science Data Catalog.

Using the structure and data content of a ScienceBase data release, a simple user interface was constructed to demonstrate how the open source capabilities of the R programming language and environment can interact with the properties and objects of the ScienceBase data release and be used to generate interactive maps. Shiny, a relatively lightweight application framework for R, was used along with a basemap obtained through the Leaflet R package to build an interactive mapping application that can be downloaded to a user's computer.

U.S. Geological Survey data publishing policies and practices are evolving, and scientists should consult the resources cited in this paper for definitive policy information. In addition, USGS Web sites could change, and the information, uniform resource locators (URLs) and email addresses provided in this report could be superseded. Finally, the example application created for this project is for distribution to a limited number of map users (probably less than about 50 users total, with only a few users accessing the application at the same time). For more complex datasets, or for applications to be used by hundreds or thousands of people, a Web portal or other enterprise-grade solution might be more appropriate.

References Cited

Blodgett, D.L., 2013, The U.S. Geological Survey Climate Geo Data Portal—An integrated broker for climate and geospatial data: U.S. Geological Survey Fact Sheet 2013–3019, 2 p.

Chase, K.J., Haj, A.E., Regan, R.S., and Viger, R.J., 2016a, Potential effects of climate change on streamflow in Eastern and Central Montana: Journal of Hydrology Regional Studies, v. 7, September 2016, p. 69–81, accessed December 2016 at http://dx.doi.org/10.1016/j.ejrh.2016.06.001.

Chase, K.J., Haj, A.E., Regan, R.S., and Viger, R.J., 2016b, Potential effects of climate change on streamflow in Eastern and Central Montana (2013–2014 analyses)—PRMS input and output data: U.S. Geological Survey data release, accessed December 2016 at http://dx.doi.org/doi:10.5066/F7P26W5S.

Chang, Winston, Cheng, Joe, Allaire, J.J., Xie, Yihui, and McPherson, Jonathan, 2016, shiny—Web application framework for R: R package version 0.13.2 Web page, accessed September 2016 at https://cran.r-project.org/web/packages/shiny/.

Cheng, Joe, and Xie, Yihui, 2016, Leaflet—Create interactive web maps with the Javascript 'Leaflet' library: R package version 1.0.1 Web page, accessed September 2016 at https://CRAN.R-project.org/package=leaflet.

Fristom, David, 2013, Why science should be (even more) open: Boston, Mass., BU Today, Boston University, accessed September 2016 at https://www.bu.edu/today/2013/pov-why-science-should-be-even-more-open/.

Open Geospatial Consortium, Inc., 2005, OpenGIS Web feature service implementation specification: Open Geospatial Consortium, Inc., accessed August 8, 2016, at http://www.opengeospatial.org/standards/wfs.

Open Geospatial Consortium, Inc., 2006, OpenGIS Web map service implementation specification: Open Geospatial Consortium, Inc., accessed August 8, 2016, at http://www.opengeospatial.org/standards/wms.

Open Geospatial Consortium, Inc., 2009, OGC KML standard development best practices: Open Geospatial Consortium, Inc., accessed August 8, 2016, at http://www.opengeospatial.org/standards/kml.

R Core Team, 2016, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing Web page, accessed September 2016 at www.R-project.org.

Tenopir, Carol, Allard, Suzie, Douglass, Kimberly, Aydinoglu, A.U., Wu, Lei, Read, Eleanor, Manoff, Maribeth, and Frame, Mike, 2011, Data sharing by scientists—Practices and perceptions: PLOS ONE, v. 6, no. 6, e21101, accessed September 2016 at http://dx.doi.org/10.1371/journal.pone.0021101.

Tenopir, Carol, Dalton, E.D., Allard, Suzie, Frame, Miek, Pjesivac, Ivanka, Birch, Ben, Pollock, Danielle, and Dorsett, Kristina, 2015, Changes in data sharing and data reuse practices and perceptions among scientists worldwide: PLOS ONE, v. 10, no. 8, e0134826, accessed September 2016 at http://dx.doi.org/10.1371/journal.pone.0134826.

Temple Lang, Duncan, 2016, RCurl—General network (HTTP/FTP/…) client interface for R: R package version 1.95-4.8, accessed September 2016 at https://CRAN.R-project.org/package=RCurl.

Thamke, J.N., LeCain, G.D., Ryter, D.W., Sando, Roy, and Long, A.J., 2014, Hydrogeologic framework of the uppermost principal aquifer systems in the Williston and Powder River structural basins, United States and Canada (ver. 1.1, December 2014): U.S. Geological Survey Scientific Investigations Report 2014–5047, 38 p. [Also available at http://dx.doi.org/10.3133/sir20145047.]

Viger, R.J., and Bock, A.R., 2014, GIS features of the geospatial fabric for national hydrologic modeling: U.S. Geological Survey data release, accessed May 2013, at http://dx.doi.org/10.5066/F7542KMD.

Winslow, Luke, Chamberlain, S., Appling, A., and Read, J., 2016, sbtools—USGS ScienceBase tools: R package version 1.0.2 Web page, accessed August 2016 at https://libraries.io/cran/sbtools/1.0.2.

Publishing support provided by:

Rolla Publishing Service Center

For more information concerning this publication, contact:

Director, Wyoming-Montana Water Science Center
U.S. Geological Survey
3162 Bozeman Ave
Helena, MT 59601
(406) 457–5900

Or visit the Wyoming-Montana Water Science Center Web site at:

http://wy-mt.water.usgs.gov/

Chase and others—**Sharing Our Data—An Overview of Current (2016) USGS Policies and Practices for Publishing Data on ScienceBase**—Open-File Report 2016–1202

ISSN 2331-1258 (online)
https://doi.org/10.3133/ofr20161202