Community for Data Integration 2015 Annual Report

By Madison L. Langseth, Michelle Y. Chang, Jennifer Carlino, J. Ryan Bellmore, Daniella D. Birch, Joshua Bradley, R. Sky Bristol, Daniel D. Buscombe, Jeffrey J. Duda, A. Lance Everette, Tabitha A. Graves, Michelle M. Greenwood, David L. Govoni, Heather S. Henkel, Vivian B. Hutchison, Brenda K. Jones, Tim Kern, Jennifer Lacey, Rynn M. Lamb, Frances L. Lightsom, John L. Long, Ra'ad A. Saleh, Stan W. Smith, Christopher E. Soulard, Roland J. Viger, Jonathan A. Warrick, Katherine E. Wesenberg, Daniel J. Wieferich, and Luke A. Winslow

Open-File Report 2016–1165

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: 2016

Suggested citation:
Langseth, M.L., Chang, M.Y., Carlino, Jennifer, Bellmore, J.R., Birch, D.D., Bradley, Joshua, Bristol, R.S., Buscombe, D.D., Duda, J.J., Everette, A.L., Graves, T.A., Greenwood, M.M., Govoni, H.S., Henkel, H.S., Hutchison, V.B., Jones, B.K., Kern, Tim, Lacey, Jennifer, Lamb, R.M., Lightsom, F.L., Long, J.L., Saleh, R.A., Smith, S.W., Soulard, C.E., Viger, R.J., Warrick, J.A., Wesenberg, K.E., Wieferich, D.J., and Winslow, L.A., 2016, Community for Data Integration 2015 annual report: U.S. Geological Survey Open-File Report 2016–1165, 57 p., http://dx.doi.org/10.3133/ofr20161165.

ISSN 2331-1258 (online)

Contents

Figures

Tables

Abbreviations

ADIwg	Alaska Data Integration Working Group
CCS	Crowdsourcing and Citizen Science
CDI	Community for Data Integration
CDWG	Connected Devices Working Group
CMGP	Coastal Marine and Geology Program
CSWG	Citizen Science Working Group
DGS	Digital Grain Size
DMWG	Data Management Working Group
DOI	U.S. Department of the Interior
DRIP	Dam Removal Information Portal
EPA	Environmental Protection Agency
EROS	Earth Resources Observation Systems
ETWG	Earth-Science Themes Working Group
FGDC	Federal Geographic Data Committee
FSP	Fundamental Science Practices
FY	Fiscal year
GRAN	Geological Survey R Archive Network
HDDS	Hazards Data Distribution System
HTML	HyperText Markup Language
ICT	Information and Communication Technology
ISO	International Organization for Standardization
JPEG	Joint Photographic Experts Group
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
LTA	Long-Term Archive
MPEG	Motion Pictures Expert Group
NASA	National Aeronautics and Space Administration
NHD	National Hydrography Dataset
NICTA	National ICT Australia
NOAA	National Oceanic and Atmospheric Administration
NPN	National Phenology Network
NWHC	National Wildlife Health Center
NWIS	National Water Information System
OSTP	Office of Science and Technology Policy
PDF	Portable Document Format
PI	Principal investigator
PNAMP	Pacific Northwest Aquatic Monitoring Partnership
RFP	Request for proposals
SOI	Statement of Interest
SSF	Science Support Framework

SSURGO	Soil Survey Geographic Database
STATSGO	State Soil Geographic dataset
TIF	Tagged Information File
TSWG	Technology Stack Working Group
TWC	Tetherless World Constellation
UAS	Unmanned Aircraft System
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
WIWSC	Wisconsin Water Science Center
WQX	Water Quality Exchange

Community for Data Integration 2015 Annual Report

By Madison L. Langseth, Michelle Y. Chang, Jennifer Carlino, J. Ryan Bellmore, Daniella D. Birch, Joshua Bradley, R. Sky Bristol, Daniel D. Buscombe, Jeffrey J. Duda, Anthony L. Everette, Tabitha A. Graves, Michelle M. Greenwood, David L. Govoni, Heather S. Henkel, Vivian B. Hutchison, Brenda K. Jones, Tim Kern, Jennifer Lacey, Rynn M. Lamb, Frances L. Lightsom, John L. Long, Ra'ad A. Saleh, Stan W. Smith, Christopher E. Soulard, Roland J. Viger, Jonathan A. Warrick, Katherine E. Wesenberg, Daniel J. Wieferich, and Luke A. Winslow

Abstract

The Community for Data Integration (CDI) continued to experience success in fiscal year 2015. The CDI community members have been sharing, learning, and collaborating through monthly forums, workshops, working groups, and funded projects. In fiscal year 2015, CDI coordinated 10 monthly forums with 16 different speakers from the U.S. Geological Survey and external partners; funded 11 collaborative projects; and hosted an in-person, four-day workshop, which attracted 168 (134 in-person and 34 remote) data practitioners, data providers, and data consumers from across the USGS, academia, industry, and other government agencies. The Citizen Science, Connected Devices, Data Management, Semantic Web, and Tech Stack Working Groups continued to accomplish great things in fiscal year 2015. These working groups were major stakeholders in planning the 2015 CDI Workshop; they continued developing solutions to pressing challenges, and they brought in speakers throughout the year for more focused presentations and discussions. Additionally, a new working group formed during the 2015 CDI Workshop—the Earth-Science Themes Working Group.

History of the Community for Data Integration

The U.S. Geological Survey (USGS) researches Earth science to help address complex issues affecting society and the environment. In 2006, the USGS held the first Scientific Information Management Workshop to bring together staff from across the organization to discuss the data and information management issues affecting the integration and delivery of Earth science research and investigate the use of "communities of practice" as mechanisms to share expertise about these issues. Out of this effort emerged the Council for Data Integration, which was conceived as an official organizational function that would help guide data integration activities and formalize communities of practice into working groups; however, by 2009 it became evident that many members of the Council for Data Integration had an interest in developing data integration solutions and sharing expertise in a less formal, grassroots manner, which transformed the Council into a Community for Data Integration (CDI). As of 2015, the CDI represents a dynamic community of practice focused on advancing science data and information management and integration capabilities across the USGS and the CDI community.

The CDI fosters an environment for collaboration and sharing by bringing together expertise from external partners and representatives across the USGS who are involved in

research, data management, and information technology. Membership is voluntary and open to USGS employees and other individuals and organizations willing to contribute to the community (if interested, contact cdi@usgs.gov). The purpose of the CDI is to:

- advance understanding of Earth systems through enhanced use of data and information,
- provide a forum for data practitioners to share ideas and learn new skills and techniques, and
- grow USGS data and information capabilities by increasing the visibility of data integration work across the USGS and the CDI.

To achieve these goals, the CDI focuses on activities within four applied areas: monthly forums, annual workshop/webinar series, working groups, and projects. The monthly forums provide an open dialogue to share and learn about data integration efforts or to present problems that invite the community to offer solutions, advice, and support. Since 2010, the CDI has sponsored annual workshops and webinar series to share ideas and increase visibility of current projects and activities. Stemming from common interests, the CDI working groups focus on efforts to address data management and technical challenges including the development of standards and tools, improving interoperability and information infrastructure, and data preservation within USGS and its partners. In 2013, in support of the activities of the CDI working groups, the CDI established its first formal request for proposals (RFP) process to fund projects that produce tangible data integration products. CDI's Executive Sponsors, Kevin Gallagher and Tim Quinn, provide guidance, contribute funding, and advocate for CDI's activities and projects.

Community for Data Integration Science Support Framework

In order to provide an overarching context and vision for CDI goals and activities, the CDI Coordinators, consisting of working group leads and facilitators, developed the Science Support Framework (SSF) in 2012 (U.S. Geological Survey, 2015a). The SSF categorizes the activities and processes through which research data flow and upon which the CDI operates. It is these categories that provide the operational foundation and conceptual architecture that illustrates how CDI activities contribute to Bureau-level data integration efforts.

The vertical elements in the SSF (fig. 1) represent the "how" of the CDI: processes; implementation of standards and best practices; and interactions among people, data, and technology necessary to achieve data integration. Starting from the bottom of the framework, the activities of monitoring, assessing, and researching flow through the science data life cycle (Faundeen and others, 2013) processes with the aid of applications, Web services, and semantics (that is, common frameworks and ontologies for sharing data across applications, communities, enterprises, and more). Moving to the top of the framework, the assets are transformed into information products that increase knowledge and understanding of the Earth's physical and biological systems.

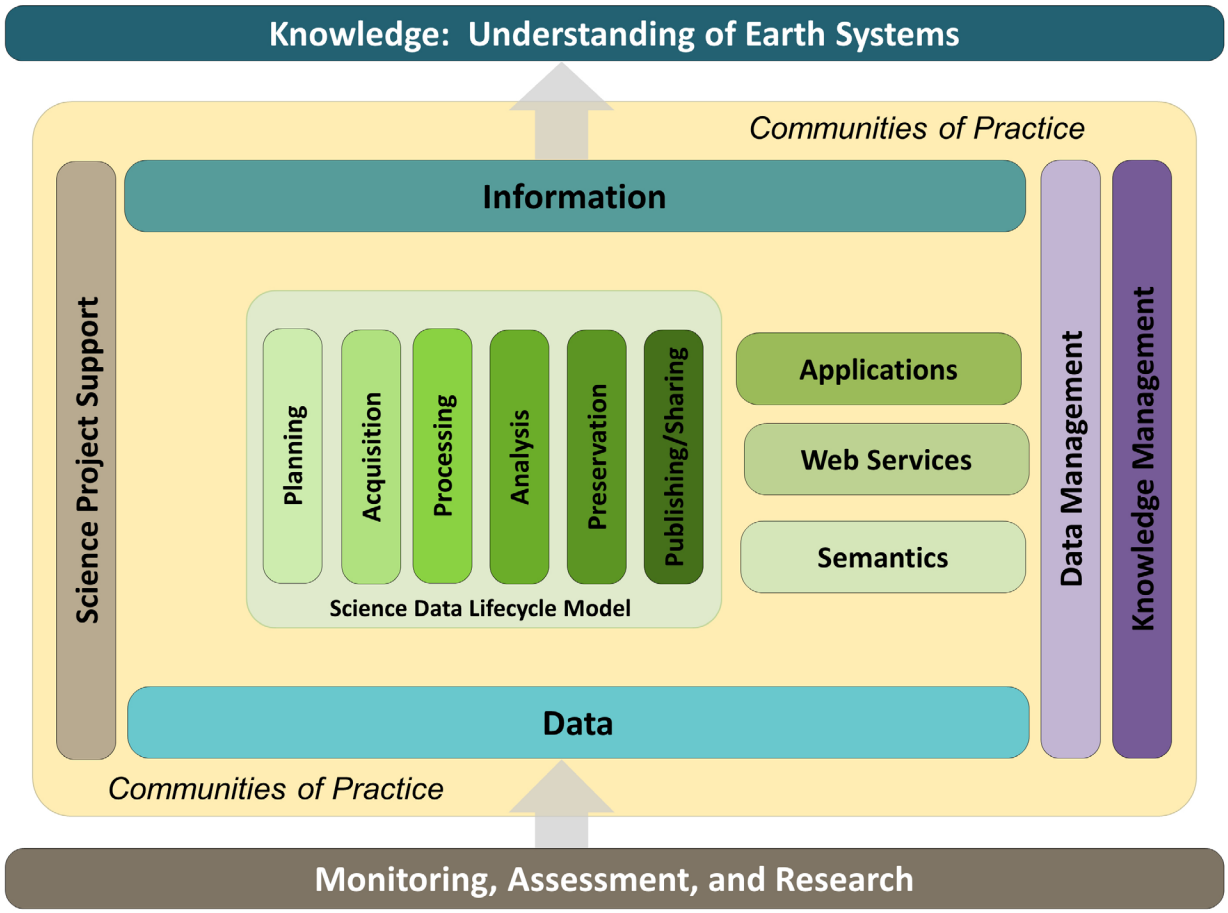


Figure 1. Community for Data Integration Science Support Framework (Faundeen and others, 2013).

The horizontal elements in the SSF (fig. 1) represent the "what" of the CDI: products, tools, and the mechanisms that mediate and contribute to the discovery and effective use of scientific data in systematic research. Data assets are managed within the context of the individual science projects, flowing horizontally (left to right) from science project support to the Science Data Lifecycle processes and applications and ultimately to data and knowledge management.

Monthly Forums

Every month, the CDI gathers for a virtual meeting forum. Monthly forums enable community members to stay up to date on new tools, best practices, standards, and policies within the Earth and biological sciences community. The CDI members and nonmembers alike are invited to give presentations on topics related to data integration. Table 1 lists the presentations from fiscal year (FY) 2015. During these monthly forums, community members are encouraged to ask questions, present challenges, and share solutions to problems. The monthly forums also provide the CDI Executive Sponsors and coordinators with the opportunity to announce upcoming CDI events. Additionally, the CDI working group leads are able to report progress on their activities during these meetings.

Table 1. Monthly forum presentations for fiscal year 2015.
[USGS, U.S. Geological Survey; NASA, National Aeronautics and Space Administration; NOAA, National Oceanic and Atmospheric Administration; NAS, Nonindigenous Aquatic Species]

Date	Presentation title	Speaker(s)
October 8, 2014	Earth Science Information Partnership: Data Preservation and Stewardship Activities Participate in the 2015 ESIP Federation Winter Meeting	Ruth Duerr, National Snow and Ice Data Center Erin Robinson, Earth Science Information Partners Federation
November 12, 2014	Development of Enhanced Feature Recognition Software for the Extraction of Mine Features from USGS Topographic Maps	Greg Fernette, USGS Central Mineral and Environmental Resources Science Center
December 10, 2014	U.S. Group on Earth Observations Data Management Working Group	Sky Bristol, USGS; Curt Tilmes, NASA; Jeff de La Beaujardiere, NOAA
January 14, 2015	Advancing Environmental Modeling and Ecosystem Services Assessments using Semantic Modeling	Kenneth Bagstad, USGS, and Ferdinando Villa, Basque Centre for Climate Change
February 11, 2015	EcoINFORMA and EnviroAtlas: Connecting People, Ecosystems, Health, and the Economy	Anne Neale, EnviroAtlas Project Lead, Environmental Protection Agency
March 11, 2015	NAS Web API: Web Services Access to the Nonindigenous Aquatic Species Database	Matt Neilson, Pam Fuller, Southeast Ecological Science Center
April 8, 2015	Data Challenges + the Crowd	Jeff Chen, Presidential Innovation Fellow, NASA Earth Science
May 13, 2015	2015 CDI Workshop	Various speakers
July 8, 2015	Update on the National Geospatial Platform	Jerry Johnston, Geospatial Information Officer, U.S. Department of the Interior
August 12, 2015	The 3D Elevation Program	Jason Stoker, National Geospatial Program
September 9, 2015	Introduction to Globus Leveraging Globus to Support Access and Delivery of Scientific Data Integrated Ocean Observing System Coastal and Ocean Modeling Testbed Use Case for Globus	Rachana Ananthakrishnan, University of Chicago Tom Cram, National Center for Atmospheric Research Rich Signell, USGS Coastal and Marine Program

2015 Community for Data Integration Workshop

In FY 2015, the CDI convened its first in-person workshop since FY 2011. The workshop was held from May 11–14, 2015, at the Denver Federal Center in Lakewood, Colorado. The workshop theme, "Building Communities to Advance Science: Inspiring, Innovating, Integrating," attracted 168 (134 in-person and 34 remote) data practitioners, data providers, and data consumers from across the USGS, academia, industry, and other government agencies. In support of the theme, the 2015 CDI face-to-face workshop provided an opportunity for this diverse group of people to come together to learn about the latest tools, share success stories, and identify new opportunities for collaboration, innovation, and advancing data integration in the Earth sciences. The purpose of the workshop was to:

- assess CDI's accomplishments from FY 2014;
- identify new, high-value opportunities for advancing data integration in the Earth sciences;

- share successes in data integration, applications, and tools;
- build upon lessons learned and strategize for future activities; and
- provide training on USGS data management, high performance computing, and integration tools.

On Monday, May 11, the workshop program featured an assortment of optional, hands-on training sessions, including the following:

- High-Performance Computing—Using the USGS Yeti and USGS Cloud Tech Stacks
- Adding Geospatial References to Publications
- Developing Metadata using International Organization for Standardization (ISO) Tools
- Tools for USGS Data Release
- Setting Up Applications in the USGS Cloud—Getting Started and Planning for the Long Term

The Workshop Plenary began on Tuesday, May 12, with sessions through Thursday, May 14, and included a variety of presentations and panel discussions about topics such as the new open data policies, data management tools, scientists' perspectives on data integration, applications for linked data and citizen science, and Earth systems informatics (table 2). Workshop participants also had the opportunity to network and showcase their current work at the interactive DataBlast poster session on Tuesday evening.

Table 2. Presentations at the 2015 CDI Workshop.

Date	Presentation Title	Presenter
Tuesday, May 12, 2015	**Welcome and Opening Remarks**	Max Ethridge, Southwest Regional Director, USGS
	Building Communities to Advance Science	Kevin T. Gallagher, Associate Director, Core Science Systems, USGS
	The Critical Zone	Martin Goldhaber, Senior Scientist and Co-Director, John Wesley Powell Center, USGS
	Open Data Policies: Can We Get There from Here?	Moderator: Fran Lightsom, USGS
	Top-Level Policy Drivers for Open Data and Access	David Govoni, USGS
	Scientific Data Management Foundation	Heather Henkel, USGS
	Metadata for Scientific Data, Software, and Other Information Products	Vivian Hutchison, USGS
	Review and Approval of Scientific Data for Release	Keith Kirk, USGS
	Preservation Requirements for Digital Scientific Data	John Faundeen, USGS
	The Research Data Commons	Philip E. Bourne, Associate Director for Data Science, National Institutes of Health
	Science Data Lifecycle Management—Tools for Open Data	Moderators: Heather Henkel, USGS and Viv Hutchison, USGS
	USGS Data Release Workflows	Keith Kirk, USGS
	Data Management Planning Tools	Madison Langseth, USGS
	Metadata Wizard	Drew Ignizio, USGS
	Online Metadata Editor	Lisa Zolly, USGS
	Digital Object Identifier Tool	Lisa Zolly, USGS

Table 2. Presentations at the 2015 CDI Workshop.—Continued

Date	Presentation Title	Presenter
	ScienceBase as a Data Release Platform	Drew Ignizio, USGS
	Science Data Catalog	Ben Wheeler, USGS
	Vocabulary Server Use Cases Skit	Semantic Web Working Group
	DataBlast: Interactive Poster Sessions and Online Demonstrations	
Wednesday, May 13, 2015	**Science Panel: Earth Scientist Perspectives on the Future of Data Integration**	Moderator: Roland Viger, USGS Panel Members: Rich Signell, USGS; Lauren Hay, USGS; Jeffrey Morisette, USGS; Jill Baron, USGS
	Linked Data: What, Who, Why, How?	Moderator: Dalia Varanka, USGS
	Use of Linked Data in the Global Change Information System	Stephan Zednik, Rensselaer Polytechnic Institute
	An Open Geospatial Consortium Project to do Linked Data for the National Hydrography Dataset	Dave Blodgett, USGS
	How and Why the MRData is Served as Linked Data	Peter Schweitzer, USGS
	A Semantic Analysis for National Hydrography Dataset Linked Open Data	Dalia Varanka, USGS
	Citizen Science and Crowdsourcing	Moderator: David Govoni, USGS
	Federal Community of Practice on Crowdsourcing and Citizen Science	Jay Benforado, Environmental Protection Agency and Lea Shanley, National Aeronautics and Space Administration
	USGS Citizen Science Guidebook	Carl Shapiro, USGS and Pierre Glynn, USGS
	Data Quality and the National Map Corps	Elizabeth McCartney, USGS
	The USA National Phenology Network—Data Product Development and Delivery Framework	Jake Weltzin, USGS
	Informatics in Earth Science: Practical Considerations for Integration	Moderator: Daniella Birch, USGS
	Scientific Computing Literacy	Mariela Perignon, Software Carpentry
	Open Cans of Worms…	Kevin McNinch, USGS
	Web Services	Tim Kern, USGS
	Applied Research Computing	Jeff Falgout, USGS
Thursday, May 14, 2015	**Go Code Colorado and Colorado Open Data Initiatives**	Moderator: Matt Tricomi, USGS Panel Members: Mike Hardin, Colorado Secretary of State's Office; Trevor Timmons, Colorado Secretary of State's Office; Andrew Cole, Colorado Secretary of State's Office; Matt Kane, Xentity Corporation
	CDI Projects in Review—Successes from CDI Past Projects	Moderator: Jennifer Carlino, USGS
	Science Data Lifecycle Implementation	John Faundeen, USGS
	Rethinking the Mobile Application Framework	Tim Kern, USGS
	Enhancing Climate Adaptation Research Through the Geo Data Portal Project	Dave Blodgett, USGS
	Evaluation of Downscaled General Circulation Model Output	Lauren Hay, USGS

On Thursday, May 14, 2015, the CDI working groups held concurrent meetings to network, identify new opportunities, plan prospective projects, and develop strategies for confronting major data-related issues. A major outcome from these meetings was the establishment of a new Earth-Science Themes Working Group. This group plans to build collaborations around specific themes, such as water or land cover, to help develop and share methods, data, software, and conceptual models and to ensure that the CDI projects, tools, and best practices are meeting the needs of scientists. See the "Earth-Science Themes Working Group" section for more information on this new Working Group.

Tim Quinn, Chief of the USGS Office of Enterprise Information, was in attendance at the 2015 CDI Workshop. During the closing discussion, he made the following remarks:

> "CDI is the future of the USGS. [We] cannot provide 21st century science without this group… [CDI represents] exposure to a way of thinking and problem solving that is pure genius."

Shortly after the workshop, CDI's Executive Sponsor, Kevin Gallagher, enlisted Tim Quinn as Co-Executive Sponsor for the community.

After the workshop, participants were asked to complete a workshop evaluation questionnaire. Fifty-nine people responded to the survey for a response rate of 35 percent. The evaluation was overwhelmingly positive with 98 percent of respondents agreeing that the workshop provided information and (or) resources that will help them on the job and 93 percent of respondents agreeing that the workshop helped them professionally. Fifty-three percent of respondents reported that their main reason for attending the workshop was for networking purposes. The top two greatest benefits experienced by respondents at the workshop were the opportunity to connect with other CDI community members and learn about tools and products that are relevant to their work.

Working Groups and Focus Groups

The CDI is organized into working groups that form around common interests in specific topics related to data integration (table 3). These working groups provide a platform for sharing resources and knowledge, discussing challenges, and identifying solutions that will help to advance data integration in the Earth and biological sciences. Some working groups meet on a regular basis, whereas others meet when the need arises. Each working group has one or more leaders to coordinate meetings, projects, and information sharing as well as to report current activities up to the larger CDI community. Working group membership is voluntary and open to all interested CDI members.

Table 3. Community for Data Integration Working Groups and Contacts

Working Group Name	Working Group Contact(s)
Citizen Science Working Group	Dave Govoni – dgovoni@usgs.gov (2015) Sophia Liu - sophialiu@usgs.gov (2016)
Connected Devices Working Group	Tim Kern – kernt@usgs.gov Lance Everette - everettel@usgs.gov
Data Management Working Group	Viv Hutchison – vhutchison@usgs.gov Heather Henkel - hhenkel@usgs.gov
Earth-Science Themes Working Group	Roland Viger – rviger@usgs.gov
Semantic Web Working Group	Fran Lightsom – flightsom@usgs.gov
Technology Stack Working Group	Daniella Birch - dbirch@usgs.gov (2015) Richard Signell - rsignell@usgs.gov (2016)

Citizen Science Working Group

The purpose of CDI Citizen Science Working Group (CSWG) has, since its inception, been to do the following:

- Promote an understanding of the role and potential benefits of citizen science and citizen scientists in the conduct of USGS research.
- Facilitate and enhance connections between the USGS and the larger Federal and public citizen science communities.
- Provide access to information and tools to support the proper, effective, and creative use of citizen science-derived data within the USGS.
- Engage the public in USGS and USGS-partner science.
- Foster and improve the public's scientific literacy.

The Working Group was, for the most part, inactive during FY 2015; however, several members broadly engaged with the larger Federal citizen science and crowdsourcing community, the President's Office of Science and Technology Policy (OSTP), and the U.S. Congress to help develop tools and policies to encourage and ease the use of citizen science and crowdsourcing in the Federal sphere.

Citizen Science Working Group Accomplishments

Because of the CSWG's general inactivity in FY 2015, there are no accomplishments specific to the CDI and the CDI Working Group to report; however, as part of the USGS's robust engagement with the Federal citizen science and crowdsourcing community, the OSTP, U.S. Congress, and others within the USGS research community, CSWG members significantly contributed to several key initiatives and outcomes in FY 2015, which are described in the next section.

Federal Community of Practice on Crowdsourcing and Citizen Science

The Federal community of practice on Crowdsourcing and Citizen Science (CCS) works across the government to share lessons learned and to develop best practices for designing, implementing, and evaluating crowdsourcing and citizen science initiatives. Several CSWG members continued to represent the USGS in this community and directly supported key activities in FY 2015, including the following:

- The CSWG members provided substantial background and procedural information and advice to the author of "Crowdsourcing, Citizen Science, and the Law: Legal Issues

Affecting Federal Agencies," a study published in April 2015 by the Woodrow Wilson International Center for Scholars, Washington, D.C. (Gellman, 2015).

- The CSWG members participated as planners, co-conveners, moderators, and speakers in the September 2015 workshop, "Implementation and Metrics of Success of Federal Citizen Science Projects," co-sponsored by the CCS and National Institutes of Health. The Federal-only event was convened to informally discuss on-the-ground challenges in implementation and measurement of citizen science and crowdsourcing projects and to share solutions to difficult challenges.

Federal Crowdsourcing and Citizen Science Toolkit

The Obama Administration's 2013 Second Open Government National Action Plan called on Federal agencies to encourage and accelerate the use of open innovation methods, such as citizen science and crowdsourcing, to help address a wide range of scientific and societal problems. To enable effective and appropriate use of these approaches, the National Action Plan specifically committed the Federal government to "convene an interagency group to develop an Open Innovation Toolkit for Federal agencies that will include best practices, training, policies, and guidance on authorities related to open innovation, including approaches such as incentive prizes, crowdsourcing, and citizen science." As part of the first phase of this effort, the OSTP, in close collaboration with the CCS, created the Web-based "Federal Crowdsourcing and Citizen Science Toolkit" (2015). The toolkit walks users through the process of establishing and executing citizen science projects. Although it is aimed specifically at the Federal community, the toolkit is also available to the public. The CSWG members were heavily involved in all aspects of toolkit and Web site design and development, including the information architecture design; functional design; content research, selection, and creation (including identification and selection of six USGS case studies); coordination of legal and other content reviews; user testing; and coordination of USGS public outreach.

Addressing Societal and Scientific Challenges through Citizen Science and Crowdsourcing Memorandum

Dr. John Holdren, Director of the OSTP, released a memorandum to Federal agencies that encourages agencies to properly utilize citizen science and crowdsourcing projects as appropriate to advance their missions (Holdren, 2015). The memorandum addresses important principles that Federal agencies should consider when designing citizen science and crowdsourcing projects, such as data quality, openness, and public participation, and provided steps that agencies should take in order to facilitate the use of citizen science and crowdsourcing. The memorandum also includes several USGS citizen science projects as exemplars. The CSWG members were invited to review and comment on the draft text of the memorandum prior to release.

Crowdsourcing and Citizen Science Act of 2015

On September 30, 2015, Senator Chris Coons and Senator Steve Daines introduced the Crowdsourcing and Citizen Science Act of 2015 (S.2113) to provide clarification to government agencies on citizen science and crowdsourcing, removing any ambiguity about whether an agency can use these techniques. The bill specifically authorizes each Federal agency, or multiple Federal agencies working cooperatively, to use crowdsourcing and citizen science approaches to conduct activities designed to advance the agency's mission or the joint mission of

the group of agencies. Following earlier reviews and an interview with Senate staff, CSWG members participated with other Federal CCS members in a round of reviews to refine the language of the bill as finally submitted. The event on September 30 also featured closed-door stakeholder discussion sessions in which CSWG members took part as participants and in one case as a focus area moderator.

U.S. Fish and Wildlife Service Citizen Science Framework

The U.S. Fish and Wildlife Service is developing a comprehensive policy with a procedural and technical framework to guide and support the design of citizen science programs on national wildlife refuges. The CSWG members served as invited subject matter experts representing the USGS at a workshop held in May 2015. These CSWG participants contributed substantively to the structure and content of the framework.

USGS Citizen Science Guidebook

Several members of the CSWG, working under the auspices of the USGS Science and Decisions Center, began work on a document intended to provide guidance to USGS scientists on when, where, and how public participation should be used in performing our scientific activities. Outline development and content research is in-progress as of November 2015.

Connected Devices Working Group

To help information technology professionals, software developers, and scientific researchers come to grips with the possibilities, opportunities, and challenges of the "Internet of Things", the Community for Data Integration established the Connected Devices Working Group (CDWG). This group explores the use of cutting edge mobile tools, frameworks, and security to support scientists. Some of the technologies scheduled for evaluation include wearable computing devices, cellular-based tracking devices, and mobile science support tools.

This Working Group supplants the legacy CDI Mobile Applications Working Group but maintains support in the areas of responsive design and mobile development frameworks. Although the CDWG continues to explore mobile application frameworks and technical advances, it also delves into new topics associated with all things Internet accessible.

Connected Devices Working Group Accomplishments

The CDWG works on multiple fronts including organizing group calls, coordinating with DigitalGov, working with DOI and USGS policy contacts, and developing one-on-one contact with USGS staff members that are developing mobile software projects. Through these FY 2015 efforts, the USGS released three mobile applications for Apple operating system devices: iPlover (application for collecting information on coastal beach habitats), CO2Calc (application for calculating CO_2-system parameters), and pHPhotometer (application that links with a pH photometer to aid in the collection of pH data). iPlover was also released for Android mobile devices. The CDWG redeveloped the USGS mobile framework and updated the Mobile Application Release Checklist (Kern, 2015a). This activity, in turn, led to a more comprehensive software release checklist (Kern, 2015b) that the group hopes will be used in conjunction with the upcoming USGS Software Release Instructional Memos.

Finally, Working Group leads also worked with CDI at large to establish a CDI-specific idea collection application known as the IdeaLab. This is designed to collect suggestions for

speakers and new projects as well as gather and evaluate feedback on current speakers and initiatives. The CDWG-specific forum is building ideas for FY 2016 (U.S. Geological Survey, 2015b).

Connected Devices Working Group Meetings and Presentations

The CDWG met periodically in FY 2015. From February to April of 2015, the group held a number of working sessions on USGS software release review and redevelopment. In May of 2015, the group held a breakout session at the 2015 CDI Workshop in Denver.

Data Management Working Group

The Data Management Working Group (DMWG) seeks to answer the following question: How can the Community for Data Integration promote the practice of data management throughout the USGS, elevate the value and accessibility of USGS data, and make data integration possible?

The USGS produces a vast number of valuable datasets every year that are used to advance science. USGS scientists, across all missions and programs, work to develop, analyze, and publish papers on data collected by the USGS; however, the lifecycle of a dataset does not end with a given scientist or project. The ability to integrate multiple datasets for analysis and reuse increases the utility and value of the original data. In fact, data integration is necessary for answering more complicated questions in science; however, before data integration can be undertaken, it requires the data to be managed properly. To enable better and easier integration of USGS data, the DMWG develops mechanisms for incorporating data management into USGS science and educating scientists of its value. The group seeks to elevate the practice of data management as a critical component in the pursuit of science at the USGS.

Data Management Working Group Accomplishments

The DMWG accomplishments in FY 2015 focused on policy, data management best practices, and helping science centers and scientists understand policy implementation importance and options.

USGS Data Policy Team

Members of the DMWG were selected to participate in a Data Policy team, sponsored by Alan Thornhill (Office of Science Quality and Integrity) and Kevin Gallagher (Core Science Systems). The following policies were written, reviewed by multiple groups across the Survey, and released as USGS Instructional Memoranda in February 2015:

- Scientific Data Management Foundation (U.S. Geological Survey, 2015c)
- Metadata for Scientific Data, Software, and Other Information Products (U.S. Geological Survey, 2015d)
- Review and Approval of Scientific Data for Release (U.S. Geological Survey, 2015e)
- Preservation Requirements for Digital Scientific Data (U.S. Geological Survey, 2015f)

With the rising cost of collecting, analyzing, and publishing scientific results, the underlying USGS data, which are critically important to understanding Earth's systems, are incredibly valuable. The adoption of science data management at the USGS in a consistent,

systematic manner supports data access, understanding, and reuse. The Instructional Memoranda are one result of this important need for data management within the USGS.

Data Management Webinar Series

The DMWG engaged in a year-long partnership with the Pacific Northwest Aquatic Monitoring Partnership (PNAMP) to implement a Data Management Webinar Series. (More information is available in the "Meetings and Presentations" section below.) The series used the USGS Science Data Lifecycle Model (Faundeen and others, 2013) to focus each month's presentation. Each of the presentations was recorded, and the slides and recordings are available on the Confluence wiki for the series (Norkin, 2015).

The webinar series offered scientists and data managers an educational opportunity to gain an understanding of the components of the science data lifecycle. The webinar series focused on teaching scientists about best practices and tools to help them meet the new open data requirements within the USGS. On average, approximately 40 to 130 participants joined the webinar series each month, for an average monthly attendance of 78. The successful turnout each month demonstrates the need for this type of knowledge sharing.

Data Management Web Site

The DMWG supported the creation of the USGS Data Management Web site in 2012 and continues to provide support for the site (U.S. Geological Survey, 2015g). During 2015, the Web site added new information and resources including the following:

- policy language from the new instructional memoranda
- standard data citation formats for USGS
- metadata section that follows a sequential, process-step approach
- examples of science center data management plans
- new USGS Exit Survey (developed by the DMWG)
- information about conducting official USGS data release

This Web site is a critical, central resource for scientists and data managers to obtain information on best practices in science data management. Additionally, links to the Web site are embedded in the Instructional Memoranda, providing practical and beneficial information to scientists, data managers, and others about how to implement the policy. The Web site had a substantial increase in usage in 2015 (fig. 2).

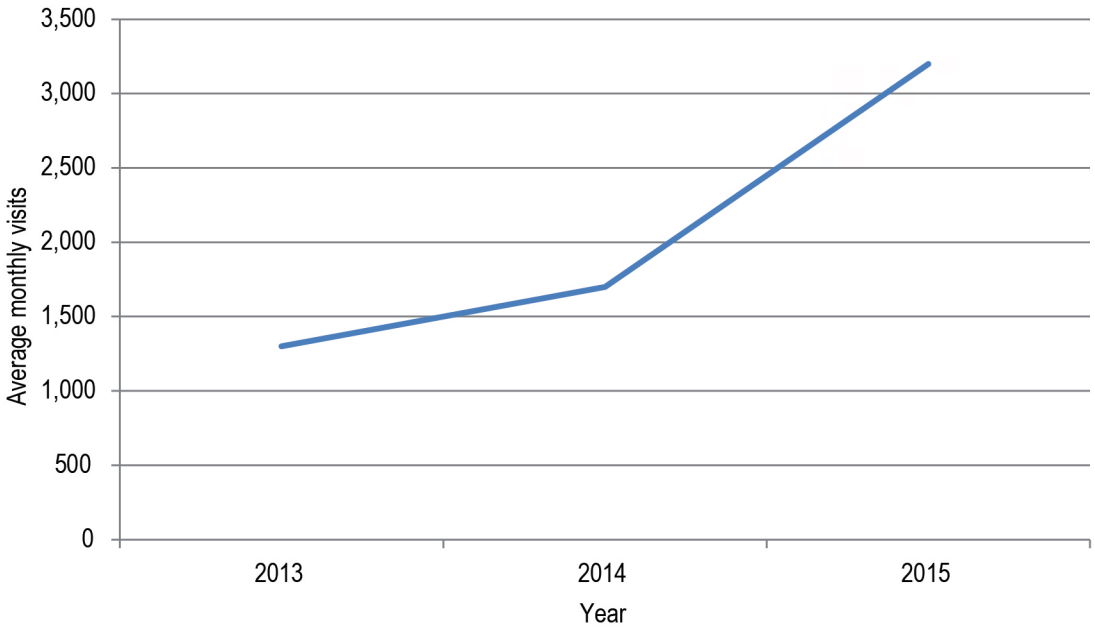


Figure 2. Average monthly visits to the Data Management Web site for 2013, 2014, and 2015.

Supplemental Data Management Training Modules

An alternate learning method for topics on the data lifecycle is important to support a variety of learning styles; therefore, the DMWG collaborated with the Office of Organizational and Employee Development to obtain funding to develop the following three additional training modules to supplement the Web site:

- USGS Science Data Lifecycle
- Metadata for Research Data
- Planning for Data Management

The modules were funded by the USGS Office of Employee Development and will be posted to the Data Management Web site in early FY 2016. Learning modules such as these are useful for teaching environments in universities and libraries.

Data Management Working Group Meetings and Presentations

The DMWG, in partnership with the PNAMP, sponsored a Data Management Webinar Series throughout FY 2015 and into early FY 2016 (described in the "Data Management Working Group Accomplishments" section). The series' topics followed the components of the science data lifecycle. Table 4 shows the date, the title of the presentation, the speakers, and their institutional affiliation.

Table 4. Data Management Webinar Series presentations

Date (number of participants)	Title	Speaker
January 2015 (128)	Data Management Overview: Where Is All This Leading Us?	Sky Bristol, USGS
February 2015 (89)	Data Management Planning Part 1: Overview and a USGS Program Experience	Emily Fort, USGS
March 2015 (98)	Data Management Planning Part 2: Theory and Practice in Research Data Management	Steve Tessler and Stan Smith, USGS
April 2015 (119)	Data Collection Part 1: How to Avoid a Spreadsheet Mess—Lessons Learned from an Ecologist	Stephanie Hampton, Washington State University
May 2015 (66)	Data Collection Part 2: Relational Databases—Getting the Foundation Right	Keith Hurley, Nebraska Game and Parks Commission
June 2015 (44)	Data Sharing and Management within a Large-Scale, Heterogeneous Sensor Network Using the Consortium of Universities for the Advancement of Hydrologic Science, Inc. Hydrologic Information System	Jeff Horsburgh, Utah State University
July 2015 (83)	Metadata: Standards, Tools, and Recommended Techniques	Lisa Zolly, USGS
September 2015 (69)	Monitoring Resources: Web Tools Promoting Documentation, Data Discovery and Collaboration	Becca Scully and Katie Pierson, PNAMP

Earth-Science Themes Working Group

The new Earth-Science Themes Working Group (ETWG) was established at the 2015 CDI Workshop. Working Group members are developing wiki pages within the CDI Confluence Web site in the ETWG branch (Viger, 2015). This group is unique among CDI working groups in that it focuses on building a federation of smaller communities that focus on specific Earth science research and analysis topics. The ETWG leverages the many contributions coming from the other working groups while exploring Earth science questions. Although the ETWG's subordinate focus groups, such as ETWG-Elevation, ETWG-Water, and ETWG-Land Cover, correspond to the thematic mapping programs of the National Map, these groups are not intended to be exclusively oriented towards these National Map programs. They are intended to be places for project- and program-level scientists to develop Earth science knowledge by bringing their own questions, which may range from theory to digital representation to methods and techniques. Data producers are strongly encouraged to participate, not only to improve the use, and therefore the impact, of their product with an active user community, but also to more aggressively seek out feedback to improve their own products.

Although still developing, there has been valuable activity in the ETWG-Soils and ETWG-Water Focus Groups. The former includes members from the USGS, U.S. Department of Agriculture (USDA) Natural Resources Conservation Service, USDA Agricultural Research Service, Environmental Protection Agency (EPA), and a number of universities. An interagency research paper comparing the State Soil Geographic dataset (STATSGO) and the Soil Survey Geographic Database (SSURGO) across scales is currently being developed by members of the ETWG-Soils focus group. The ETWG-Water Focus Group has a specialization in the analysis of hydrographic networks, such as the National Hydrography Dataset (NHD) and NHDPlus. Members are drawn from the EPA, and several USGS programs, including StreamStats, USGS

National Water-Quality Assessment Program, and the Water Mission National Research Program. A spinoff of this work has resulted in discussions with members of several Canadian Federal and Provincial ministries, via the International Joint Commission, regarding better integration of the United States NHD with the Canadian National Hydrographic Network for use in hydrologic modeling.

The ETWG-Vertical Integration Focus Group specializes in topics that require knowledge, data, and techniques drawn from more than one of the other ETWG focus groups. A good example is detection, mapping, and analysis of agricultural tile drainage systems. After members from ETWG-Soils initiated this, several new members interested in hydrological modeling joined. It has also attracted interest from several scientists using different remote sensing platforms, from thermal infrared to airborne prototypes of the National Aeronautics and Space Administration (NASA) Surface Water and Ocean Topography sensor. Coordinators hope to guide the group to submit proposals for funding in the next fiscal year. One of the hopes of this focus group, and the ETWG as a whole, is to begin building consensus on practical connections between these fundamental Earth science themes in the spirit of increasingly popular concepts of holistic "Earth" frameworks, like the EarthCube, the Digital Crust, or the Consortium of Universities for the Advancement of Hydrologic Science, Inc. Hydrologic Information System.

Semantic Web Working Group

The Semantic Web Working Group is a small group of data managers who are working together to explore Semantic Web technologies for use in our day jobs and also to improve the discovery, access, use, and integration of USGS data. In the past, the Working Group hosted presentations and seminar-style discussions and sponsored two CDI-funded projects. At the face-to-face meeting during the 2015 CDI Workshop, the group embarked on a practical learning project that was independent of the CDI funding process. This project involves combining the group's expertise and institutional resources to release foundational USGS data holdings as open, linked data that can be used as a semantic framework for interdisciplinary data integration.

Semantic Web Working Group Accomplishments

The Working Group sponsored a 2014–15 CDI-funded project named "Use of Controlled Vocabularies in USGS Information Applications" to produce a plan for implementing controlled vocabularies to improve USGS metadata and the discovery of data in the USGS Science Data Catalog. For more information, see the project description in the "Community for Data Integration Projects" section of this report.

At the 2015 CDI Workshop, the Working Group presented a skit, "Keywords for Better Metadata: A Morality Play in One Act," to share what we have learned about Semantic Web technologies with other CDI members (fig. 3).

Figure 3. Opening scene from "Keywords for Better Metadata: A Morality Play in One Act," a skit presented at the 2015 USGS Community for Data Integration Workshop, May 11–14, 2015, Denver, Colorado. Photo by Daniel Wieferich, U.S. Geological Survey.

Semantic Web Working Group Meetings and Presentations

The Semantic Web Working Group met on a monthly basis throughout FY 2015. Table 5 provides a brief description of each meeting.

Table 5. Semantic Web Working Group fiscal year 2015 meetings.

Meeting date	Meeting description
November 2014	New Working Group member, Daniella Birch, from the Illinois Water Science Center, introduced herself and her projects.
December 2014	Working Group member Stephan Zednik from Rensselaer Polytechnic Institute gave a preview of his American Geophysical Union talk, "Linked Vocabulary API for the Earth Sciences Community" (Zednik and others, 2014).
January 2015	Working Group member Ken Bagstad and Ferdinando Villa from the Basque Centre for Climate Change discussed Semantic Web details of the presentation to the Community for Data Integration (CDI) monthly meeting on January 14, 2015, "Advancing environmental modeling and ecosystem services assessments using semantic modeling."

Table 5. Semantic Web Working Group fiscal year 2015 meetings.—Continued

Meeting date	Meeting description
February 2015	The Working Group discussed how it can contribute to the 2015 CDI Workshop and developed the following ideas: • presenting a skit about the group's vocabulary services project • offering a Data Blast event that uses a vocabulary service to produce buzzword bingo cards • focusing the Semantic Web Working Group meeting on talking about the next big project that the Working Group would like to tackle
March 2015	The Working Group developed a plan for a plenary session at the 2015 CDI Workshop titled, "Linked Data: What, Who, Why, How?"
April 2015	Working Group members previewed the skit for the 2015 CDI Workshop and received feedback for improvements from the rest of the Working Group.
May 2015	The Working Group met during a breakout session at the 2015 CDI Workshop and decided that the next Working Group project would be to develop a reusable linked data technical implementation that can be used by science centers for georeferenced USGS data, such as the locations of the National Water Information System sites and coastal and undersea features.
June 2015	The Working Group evaluated its contributions to the 2015 CDI Workshop and reaffirmed its intention to work on a linked data implementation for the USGS. Alan Allwardt shared a start at writing triples for coastal features and Stephan Zednik recommended that we use GeoSPARQL vocabulary for expressing the spatial extents. The group assessed the GeoSPARQL information recommended by Stephan and agreed that it is a good approach. Stephan agreed to take the next step by expressing the Geographic Names Information System feature classes as an ontology.
July 2015	Maxwell Taylor from the Great Basin Landscape Conservation Cooperative presented the potential of using defined vocabularies for tags in ScienceBase. The group discussed the current status of controlled vocabulary use in ScienceBase and how the group could help ScienceBase achieve the benefits of controlled vocabularies.
August 2015	Working Group member Stephan Zednik demonstrated two browsers he has made use Elastic Search. Some Working Group members decided to develop metadata records using controlled vocabularies to give to Stephan to see how they can be used in browsers similar to his own. Stephan's goal is to develop a faceted search over records in https://www.data.gov using the Data Categories for Marine Planning and related controlled vocabularies, including the topic categories of the Global Change Master Directory.
September 2015	The Working Group investigated the GeoSPARQL test site at ArcGIS Online (Esri, [n.d.]), and concluded that it is only designed to demonstrate the concept and is not available for use as a USGS endpoint. Working Group member Dave Blodgett offered to set up an experimental internal implementation of the Parliament endpoint. The Working Group identified its next step as agreeing on a model for triple data and identified the first use cases that will be explored with the new endpoint.

Technology Stack Working Group

The purpose of the Technology Stack Working Group (TSWG) is to build a community of scientists, data managers, and technology practitioners who share and evaluate tools and best practices related to scientific computing, data creation, and data publication. The TSWG has three main goals. The first goal is to use technology to improve scientific workflows and data integration. The second goal is to provide a unifying umbrella under which many member-defined focus groups can be hosted. The overarching Working Group seeks to provide coordination across the focus groups and develop a more holistic understanding of the many technologies, standards, and protocols that are constantly evolving. The final goal is to help

inform thinking in other working groups, such as the Data Management Working Group, by providing expertise on how concepts, protocols, Fundamental Science Practices (FSP), and technical memos could be implemented.

Technology Stack Working Group Accomplishments

Leading up to the 2015 CDI Workshop, Roland Viger and Daniella Birch planned and organized several informative talks for the workshop featuring work within and outside of the USGS. The topics covered high-performance computing, Software Carpentry principles, and open source tools for Web management. A major outcome of the 2015 CDI Workshop was a recognized need for the TSWG to undertake future presentations and activities on sharing code resulting in follow-up meetings focused on how to share code via Git. Sharing software code was recommended as a requirement of the FY 2016 CDI Request for Proposals, emphasizing the importance of TSWG meetings.

Technology Stack Working Group Meetings and Presentations

Daniella Birch, with the assistance of Roland Viger, committed to coordinating a monthly schedule of informative presentations for FY 2015. Table 6 provides the date, presentation title, speaker(s), and presentation abstract for each meeting.

Table 6. Presentations given at Technology Stack Working Group meetings.

Date	Title	Speaker	Abstract
January 2015	Cesium—A 3D Web Mapping Application Programming Interface	Matt Amato, CesiumJS	Cesium is a JavaScript library for creating 3D globes and 2D maps in a Web browser without a plugin. It uses WebGL for hardware-accelerated graphics, and is cross-platform, cross-browser, and tuned for dynamic-data visualization. Cesium is open source under the Apache 2.0 license. It is free for commercial and noncommercial use.
February 2015	Version Control	Jordan Walker, Center for Integrated Data Analytics	This presentation discussed version control and its many uses within the USGS environment. A major focus was on differentiating among Git, GitHub, the public GitHub site, the internal USGS one, and the USGS Stash.
March 2015	GPlates Web Portal	Mark Turner, CalTech	GPlates enables the interactive manipulation of plate-tectonic reconstructions and the visualization of geodata through geological time, and it facilitates interoperability of plate tectonic data and models with geodynamic computing services for applied and fundamental research purposes. The GPlates Portal is the gateway to a series of Web pages for the interactive visualization of cutting-edge geoscience datasets shown in Cesium.
June 2015	NICTA and Cesium	Kevin Ring, Bill Simpson-Young, and Peter Leihn, NICTA	The National ICT (Information and Communication Technology) Australia (NICTA) is Australia's largest Information and Communication Technology Centre of Excellence. NICTA is using Cesium on a growing number of initiatives involving Web-based 3D geospatial visualization.

Table 6. Presentations given at Technology Stack Working Group meetings.—Continued

Date	Title	Speaker	Abstract
July 2015	Stash and Confluence	Dell Long, Fort Collins Science Center, and Roland Viger, USGS National Research Program	Stash is an internal USGS Git tool. During this presentation, Dell explained how to access Stash and make use of it. Roland demonstrated how to post code samples on Confluence.
August 2015	GitLab	Eric Martinez, USGS	GitLab is a Git tool made available to the USGS internally. Eric explained how to access GitLab and make use of it.

Annual Community for Data Integration Request for Proposals

The CDI seeks to build and share knowledge about topics such as data integration, data stewardship, scientific computing, and approaches for knowledge delivery. The main goal of CDI funding is to improve our collective knowledge about how to create better, longer lasting, and more accessible science products by leveraging the tools, methods and datasets available to the Earth and biological science communities. The CDI places high value on innovative projects that, in the near term, produce new and reusable ideas, methods or tools that have an impact beyond a single program, center, region, or mission area. CDI project proposals are evaluated based on their alignment with the CDI Science Support Framework, the evaluation criteria laid out in the RFP guidance document (scope, technical approach, project experience and collaboration, sustainability, budget justification, and timeline), and the following goals:

- Focus on targeted efforts that yield near-term benefits to science.
- Leverage existing capabilities and data in new domains.
- Identify and demonstrate innovative solutions, methodologies, and tools that can be replicated and (or) scaled.
- Ensure sustainability of products and services.
- Seek efficiencies or substantial return on investment.
- Expose USGS data.
- Organize science models and outputs (for example, conceptual frameworks for understanding processes and process relationships).
- Preserve and easily access project data.
- Develop, organize, and share knowledge and best practices.

In 2014, the CDI established a two-phased RFP process. This approach provides more transparency and community participation in the selection process by inviting community members to vote on two-page statements of interest (SOIs) submitted by project principal investigators (PIs). The SOIs receiving the most votes from the community, as well as SOIs identified by the Executive Sponsor as addressing an emerging priority, are asked to submit a full proposal. Formal guidance for the FY 2015 RFP was released on September 10, 2014, outlining this two-phased approach for selecting CDI FY 2015 projects.

Phase I—Statements of Interest

Two-page SOIs were due on October 10, 2014. A total of 39 SOIs were submitted representing 9 SSF Elements (table 7). The lead PIs on the SOIs represented three USGS mission areas and six USGS regions.

Table 7. Number of statements of interest addressing each Science Support Framework element.

Science Support Framework element	Number of proposals
Applications	19
Web services	19
Data	17
Data management	15
Science data lifecycle	14
Communities of practice	10
Knowledge management	10
Information	8
Semantics	1

The CDI community members were asked to read all 39 SOIs and vote on them based on the CDI Science Support Framework, the evaluation criteria, and the goals described above. The voting period began on October 20, 2014, and closed on November 5, 2014. Each community member was allowed 15 votes to use across all SOIs. Each SOI could receive a maximum of three votes per person. A total of 115 community members elected to vote in FY 2015. A closing session was held on November 5 to allow the community to agree on the number of SOIs that would be recommended to move forward to the full proposals phase of the RFP. Following the closing session, the CDI Coordinators also reviewed the SOIs and provided feedback for the final recommendations to the Executive Sponsor. In the end, 19 SOIs were approved by the Executive Sponsor to be invited to submit full proposals.

Phase II—Full Proposals

Full proposals were due on January 30, 2015. One author chose not to submit a full proposal; therefore, 18 full proposals were submitted for the second phase of the RFP process. The CDI convened a formal, six-person review panel to evaluate the 18 full proposals. The reviewers were all USGS Federal employees and volunteered their time to the review panel. The reviewers represented a wide range of mission areas, regions, and programs and brought with them a variety of scientific and technical expertise. The review panel consisted of both CDI and non-CDI members. The reviewers were responsible for disclosing any potential conflicts of interest and recusing themselves from discussions involving proposals in question. The reviewers were also asked not to divulge the identity of the other reviewers.

Each reviewer was assigned to lead the discussion of three proposals. The discussion leader was responsible for having in-depth knowledge of these three proposals; however, the reviewers were responsible for reading all 18 full proposals and providing a cursory evaluation of each proposal's strengths and weaknesses. Reviewers scored each proposal based on the following weighted evaluation criteria:

- Scope (25 percent)
- Technical approach (25 percent)

- Project experience and collaboration (25 percent)
- Sustainability (15 percent)
- Budget justification (5 percent)
- Timeline (5 percent)

Each proposal was discussed in turn over the course of two review sessions, and reviewers were allowed to modify their scores based on the feedback of the other reviewers. Scores for each proposal were averaged to obtain the final score for the proposal. During the final review session, the panel collectively discussed the proposal rankings and agreed upon a final recommendation for each proposal. Eight projects were recommended for funding by the review panel (table 8).

Bureau-Wide Applications

In FY 2014, three projects were identified as having Bureau-wide applications and were eligible to receive two years of funding. The PIs for these projects were required to submit a status report, which outlined their first-year accomplishments and a revised plan for their second year, including a new budget, to receive their second year of funding.

Recommendations

The recommendations of the CDI Review Panel, along with the three Bureau-wide application projects, were submitted to the CDI Executive Sponsor, Kevin Gallagher (Associate Director, Core Science Systems mission area) for final approval. On April 8, 2015, Kevin announced funding for the eight new projects and the three second-year, Bureau-wide applications projects (table 8). The "Community for Data Integration Projects" section describes the projects and their accomplishments in more detail.

Table 8. Overview of the Community for Data Integration request for proposals projects funded in fiscal year 2015 (in alphabetical order).

Title	Principal investigator(s)	Program
Geocaching Natural Features—Applying Game Mechanics to Citizen Science Data Collection	Tabitha Graves and Dell Long	Northern Rocky Mountain Science Center and Fort Collins Science Center
Geographic Searches for USGS Publications[1]	Katherine Wesenberg	National Wildlife Health Center
Integration of Land Cover Trends Field Photography with an Online Map Service	Christopher Soulard	Western Geographic Science Center
Making Unmanned Aircraft System (UAS) Data Available to USGS Scientists and the Public	Jennifer Lacey and Raad Saleh	Earth Resources Observation Systems Data Center
National Dam Removal Database: A Living Database for Information on Dying Dams	Jeff Duda*, James Bellmore†, Jonathan Warrick‡, Sky Bristol§, Vivian Hutchison§, and Daniel Wieferich§	*Western Fisheries Research Center, †Forest and Rangeland Ecosystem Science Center, ‡Pacific Coastal and Marine Science Center, §Core Science Analytics, Synthesis & Libraries
Portable ISO 19115-2 Open Source Developer's Toolkit[1]	Stan Smith and Joshua Bradley	Alaska Science Center and U.S. Fish and Wildlife Service

Table 8. Overview of the Community for Data Integration request for proposals projects funded in fiscal year 2015 (in alphabetical order).—Continued

Title	Principal investigator(s)	Program
sbtools: An R Package for ScienceBase	Luke Winslow and Scott Chamberlain	Center for Integrated Data Analytics and Museum of Paleontology
Standards-based Integration and Delivery of USGS and EPA STORET Biomonitoring Data via the Water Quality Data Portal	Peter Ruhl	National Water Quality Assessment Program
The "Digital Grain Size" Web and Mobile-Computing Application	Daniel Buscombe	Grand Canyon Monitoring Research Center
Use of Controlled Vocabularies in USGS Information Applications—Requirements Analysis for Automated Processes and Services[1]	Frances Lightsom	Woods Hole Science Center
Web-Enabled Visualization and Access of Value-Added Disaster Products	Rynn Lamb and Brenda Jones	Earth Resources Observation Systems Data Center

[1]Bureau-wide application project

Community for Data Integration Projects

The FY 2015 projects represented many elements of the SSF including data, communities of practice, applications, semantics, science data lifecycle, data management, and knowledge management. Many of the projects in FY 2015 focused on developing or enhancing applications to increase the exposure, discoverability, and accessibility of USGS data. These projects helped to make it easier and more efficient for scientists to document and share their data. Each of the FY 2015 projects is described in detail below with references to completed products and deliverables. Many project teams continued working on deliverables after the end of the fiscal year. For example, journal articles and open-file reports associated with projects may take 6 to 12 months after the completion of the project to be published. Updates and additions to project accomplishments and deliverables will be made to the projects' records in ScienceBase (Community for Data Integration, 2015).

Geocaching Natural Features—Applying Game Mechanics to Citizen Science Data Collection

ScienceCache is a scientific geocaching mobile application framework that targets two user groups for citizen science data collection: youth and geocachers. By melding training and games into the hunt for place-based data collection sites and incorporating photo uploads as data and authentication, new volunteers can collaborate in robust data collection. Scientists build a project on the administrative Web site app, specifying locations or goals for new data collection sites, clues for established sites, questions to answer, measurements, or other activities for the site based on their individual data needs. The project builds on the success of the USA National Phenology Network (NPN) and the ScienceBase project to develop an enabling technology for citizen science and Federal data collection efforts. Two reference implementations, assessing phenology of bear foods in Glacier National Park and evaluating tree invasion into alpine

meadows using repeat photography, will allow the project team to apply lessons-learned to future efforts. The project also seeks to develop new workflows allowing for more rapid project approval and data acquisition for government citizen science efforts.

Since the beginning of the project, the project team has encountered many scientists with an interest in using ScienceCache to collect data as part of their work. Potential data collection includes blueberry phenology, pollinator ecology, and fungi data collection.

Technical Approach and Methods

The technology behind ScienceCache builds on a service-oriented architecture. Power users (scientists and resource specialists) design a Trip, which is a data collection exercise that includes both waypoints or points of interest and one or more target areas for data collection. That Trip gets incorporated into a JavaScript Object Notation (JSON)-based configuration file. The designers can use any number of technologies to build this configuration file; the published design model allows developers to design their own JSON-builder.

All Trips are hosted on a ScienceCache gateway Web application, which will be publicly available at https://www.sciencebase.gov/sciencecache by October 2016, and the ScienceCache mobile application looks at the ScienceCache gateway to see what Trips are available to the user (the data collector) based on location. The user selects one of the Trips, and the ScienceCache mobile application downloads that configuration file from the gateway. The ScienceCache mobile application reads the JSON-based configuration file and converts it into a mobile-ready data entry form. The user follows the Trip designed by the power user, finding interesting sights along the way (fig. 4). At the target area(s), the user performs some Trip-specific data collection using the designed data entry form and image capture options (fig. 5). The user sends their collected data back to the ScienceCache gateway when he or she is in cell or wireless range. If the Trip involves phenology data, those observations are sent to the NPN via the NPN Web service. Data from the collection, available as Web services, can then feed project-specific portals.

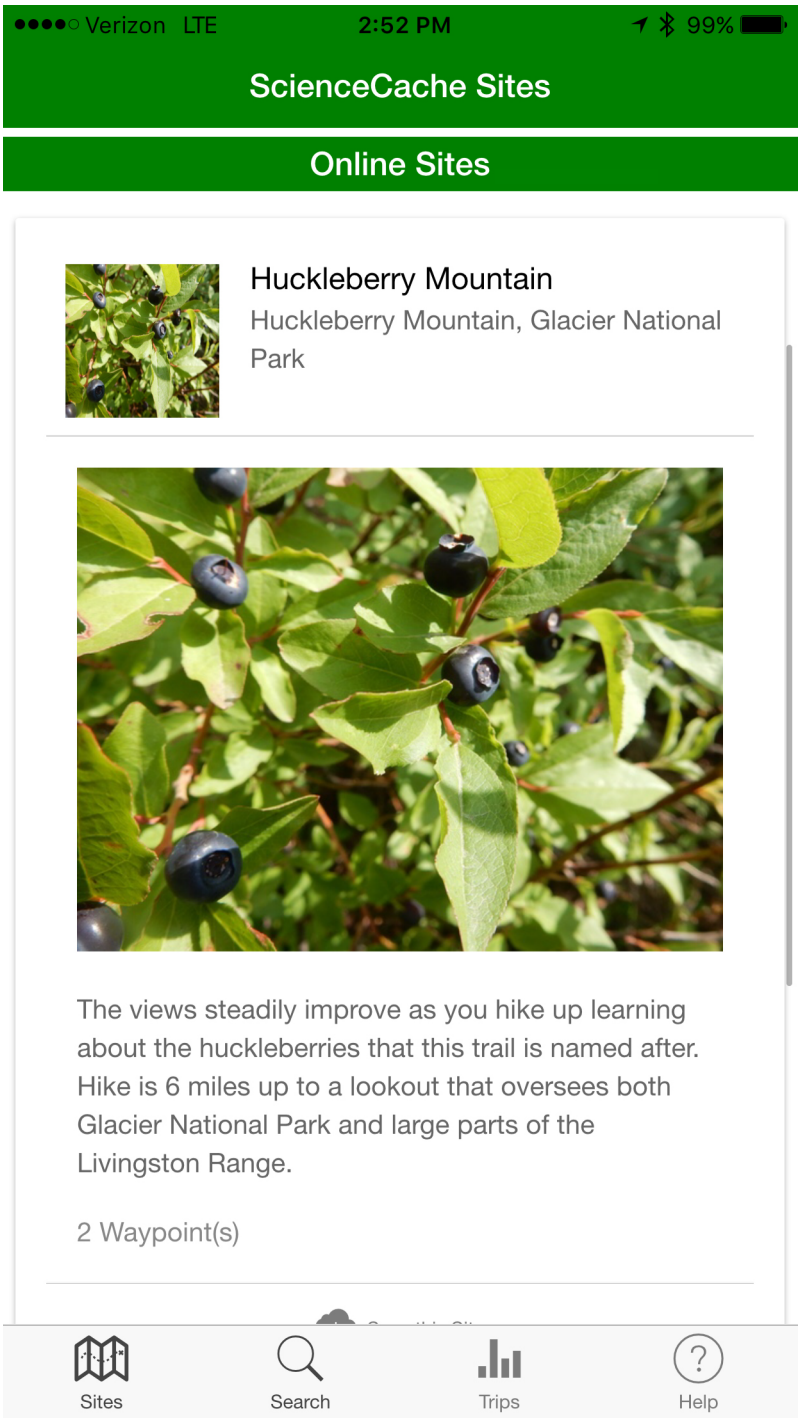


Figure 4. Introduction screen for a ScienceCache mobile application Trip.

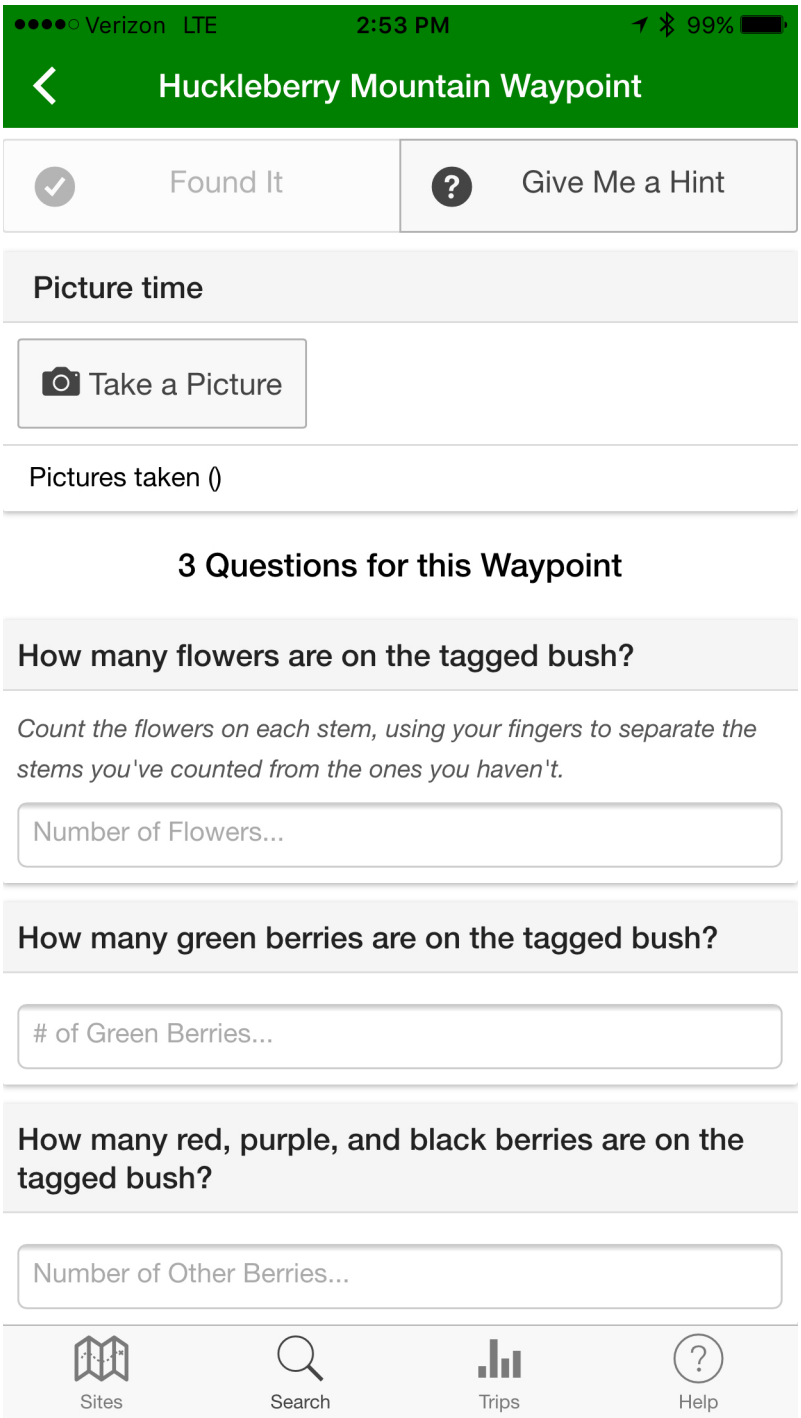


Figure 5. Data entry form for a target area in the ScienceCache mobile application.

This project contributes innovative mobile application design patterns to the USGS body of work. The framework as a whole will enable increased spatial and temporal data collection and increase engagement of the public in science across divisions. User privacy is a primary concern for this or any other data collection application. This project simplified the user workflow, eliminating any data that could be considered personal information.

This project supports the applications (mobile and Web), Web services (Web application and connections to ScienceBase), acquisition (through facilitation of citizen science data collection), and knowledge management (by enabling engagement of sectors of the public that may not otherwise be easily reached) elements of the CDI SSF.

Accomplishments

The accomplishments for the Geocaching Natural Features project were as follows:

- "Monitoring Responses of Bear Foods to Climate Change through Citizen Science" poster was presented at the 53rd Annual Conference of the Montana Chapter of The Wildlife Society (Graves, Belt, and Boyd, 2015).
- "ScienceCache—Engaging Citizen Scientists in Data Collection through Geocaching" poster was presented at the 2015 CDI Workshop and the Roundtable on the Crown of the Continent on September 14, 2015 (Graves, Long, and others, 2015).
- Federal toolkit launch discussion appeared on #CitSciChat Tweet-up on October 7, 2015.
- ScienceCache was referenced in the Third Open Government National Action Plan, "The USGS will roll out Science Cache, a Web and mobile-based application for engaging the public in citizen science projects, such as finding huckleberry plants in Glacier National Park and taking pictures and recording data to inform research on climate change impacts" (Office of Science and Technology Policy, 2015).

Expected Future Accomplishments

The project team anticipates completing the following future accomplishments:

- *Huckleberry phenology data via NPN*.—Estimated availability mid-summer 2016
- *Tree invasion repeat photography on ScienceBase*.—Estimated availability mid-summer 2016
- *ScienceCache code available at USGS Stash Repository*.—Estimated availability spring 2016
- *ScienceCache administrative web application available at USGS Stash Repository*.—Estimated availability spring 2016
- *ScienceCache mobile application available at USGS Stash Repository*.—Estimated availability spring 2016
- *Techniques and methods report*.—Designing an open-source, reusable mobile data collection framework by Dell Long, Tim Kern, and Tabitha Graves, with estimated availability in February 2016
- *Fact sheet*.—*ScienceCache*.—Geocaching natural features to support citizen science data collection, location to be determined, with estimated availability fall 2016
- *Journal article*.—*ScienceCache*.—Geocaching natural features to support citizen science data collection, Citizen Science in Theory and Practice Journal, fall 2016
- *Web site—ScienceCache*.—Engaging new citizen scientists, spring 2016
- *Journal article*.—Temporal susceptibility of food resources for bears to climate change, Biological Conservation, December 2017

Geographic Searches for USGS Publications

The purpose of this project is to improve the USGS Publications Warehouse (Pubs Warehouse) so that a person can search for USGS publications by geographic region in addition to existing search criteria; for example, one could search using map zooms or congressional districts. The addition of geographic searches allows users to narrow their search results to specific areas of interest, which reduces the time required to sift through all results outside the area of interest.

In FY 2014, the project team determined that the ScienceBase Footprint Studio would be an appropriate tool for creating the footprints for USGS publications and decided on the technical implementation for information exchange between ScienceBase and Pubs Warehouse. The project team also updated the Pubs Warehouse interface to allow users to search for publications within a geographic area. Additionally, the team updated the backend capabilities of the Pubs Warehouse to accept geospatial footprints related to publications.

In FY 2015, the project team developed workflows for creating and adding polygons in ScienceBase and in the MyPubs Publication Warehouse interface. The Coastal Marine and Geology Program (CMGP) tested the ScienceBase workflow and the National Wildlife Health Center (NWHC) and the Wisconsin Water Science Center (WIWSC) tested the MyPubs workflow. The workflows were documented in a training module for using the footprint tools.

In FY 2016, the lessons learned from this project will be used to update and migrate the footprint project to an approach that will allow for long-term sustainability. New development in the Pubs Warehouse will more tightly integrate the Pubs Warehouse and the ScienceBase geospatial dataset, giving users at cost centers who want to add new footprints an easy and effective way to do so.

Accomplishments

The accomplishments for the Geographic Searches for USGS Publications project were as follows:

- The project team tested and completed implementation of information exchange from ScienceBase Footprint Studio to Pubs Warehouse.
- The CMGP added polygons for San Francisco Delta Bay publications through the ScienceBase Footprint Studio and identified older CMGP publications that were not displaying in the Pubs Warehouse.
- The NWHC examined 967 of their previous publications and added polygons for all publications with geospatial extents through the MyPubs interface.
- The WIWSC examined 509 of their previous publications, added polygons for all publications with geospatial extents through the MyPubs interface, and identified older WIWSC publications that were not displaying in the Pubs Warehouse.
- As of December 14, 2015, approximately 25,000 polygons have been created for USGS publications in Pubs Warehouse.
- Training module PowerPoint presentation and supplemental handouts and exercises were developed and made available on the project's ScienceBase Web page (fig. 6) (Wesenberg and Allwardt, 2015).
- A three-hour training session, using the materials listed above, was held at the 2015 CDI Workshop.

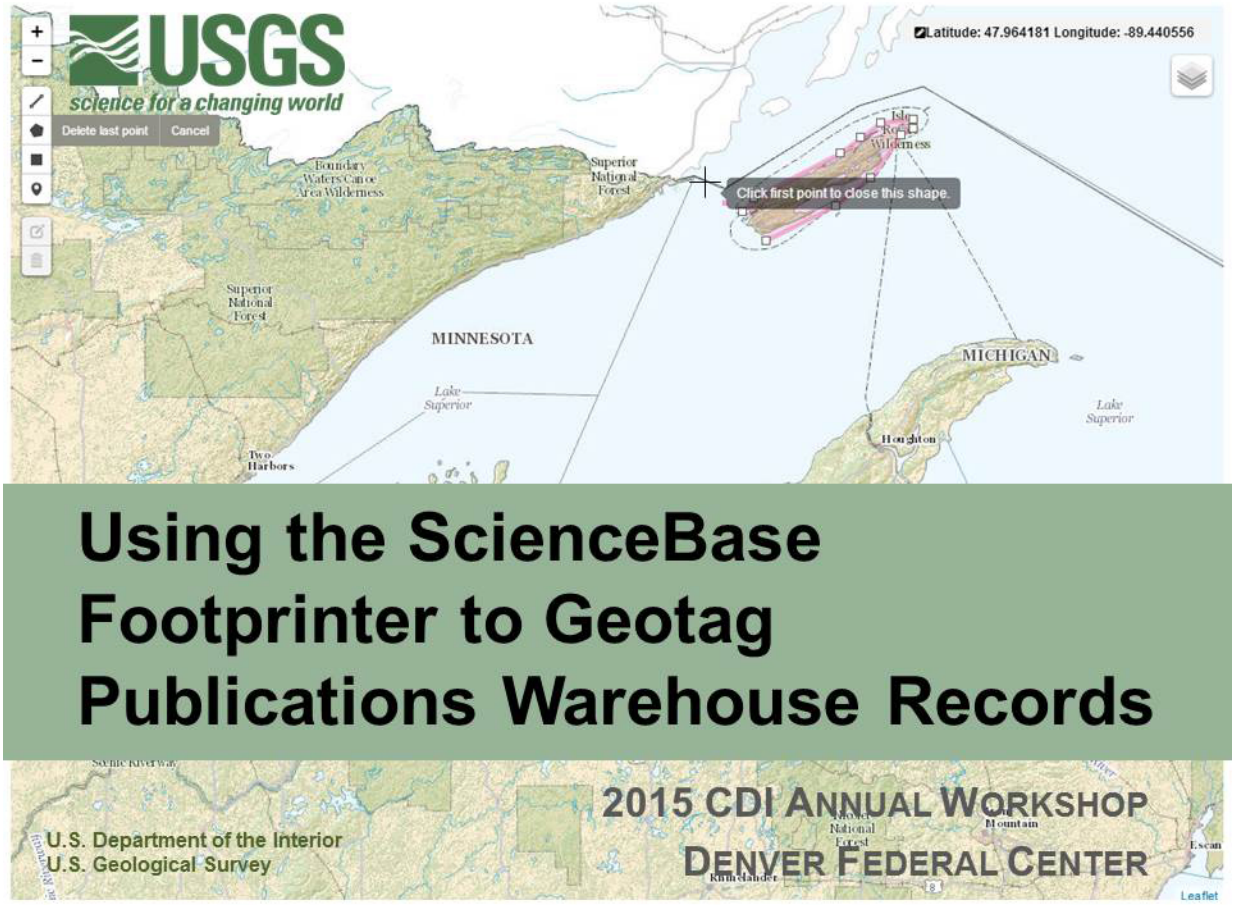


Figure 6. Introductory slide for the footprinting training module.

Integration of Land Cover Trends Field Photography with an Online Map Service

The USGS National Land Cover Trends Project has the largest repository of field photos at the USGS (over 33,000 photos). Prior to CDI funding, Land Cover Trends had limited funding to make the national collection of photos available online for researchers, land managers, and citizens. The goal of this CDI project was to add geotags and keywords to the digital copies of each field photo and make the collection searchable and downloadable via the Internet. By funding the effort to integrate Land Cover Trends field photography and online mapping technology, CDI has helped provide access to geographic data needed to conduct science and support policy decisions. Sharing georeferenced photography distributed across the conterminous United States creates an excellent avenue for the scientific community to provide access to scientific results. Additionally, this type of sharing allows the community to develop future research opportunities, such as future repeat photography research, or applications in which photos may serve as training or test site data for other remote sensing classifications. The integration of online map services and Land Cover Trends field photography also provides public access to government research and increases the visibility of such research. Serving data

to the public is directly in line with the USGS's mission and provides opportunities for future scientific collaboration by communicating USGS research to the scientific community.

Accomplishments

As of October 27, 2015, the accomplishments for the Land Cover Trends project were as follows:

- The project plan was presented as part of the 2015 CDI Workshop in Denver, Colo.
- The project team embedded geographic coordinates and keywords from field spreadsheets into exchangeable image file format metadata for over 15,000 field photos distributed across 44 of 84 Level III ecoregions (Commission for Environmental Cooperation, 1997).
- Field photos with personally identifiable information were removed from the geotagged dataset.
- A clean database of field photos was delivered to Web development teams at Earth Resources Observation Systems (EROS) Data Center and Western Geographic Science Center.
- A Web platform, titled the "USGS Trends Photo Explorer," was developed by the Western Geographic Science Center using Leaflet to present photos in a land-use/land-cover context (U.S. Geological Survey, 2015h).
- The USGS Trends Photo Explorer was approved in the Information Product Data System on October 26, 2015, for public dissemination. The USGS Trends Photo Explorer uses a map interface to present field photos online. The Web site employs the Leaflet map interface to present over 13,000 photos in 44 Level III ecoregions (fig. 7). The user can select 2011 National Land Cover Dataset, Level III ecoregions, and Land Cover Trends sample blocks to see photos in a land-use/land-cover context (fig. 8). The user can zoom in to see the spatial distribution of field photos in a specific region or search by metadata attributes, such as keywords and photo date (fig. 9). Finally, the user may browse or download the field photos of interest (fig. 10).
- The Web development team at EROS Data Center has also placed photos within the USGS Earth Explorer (U.S. Geological Survey, 2015i). A second FSP review at EROS Data Center is underway and will be completed shortly.

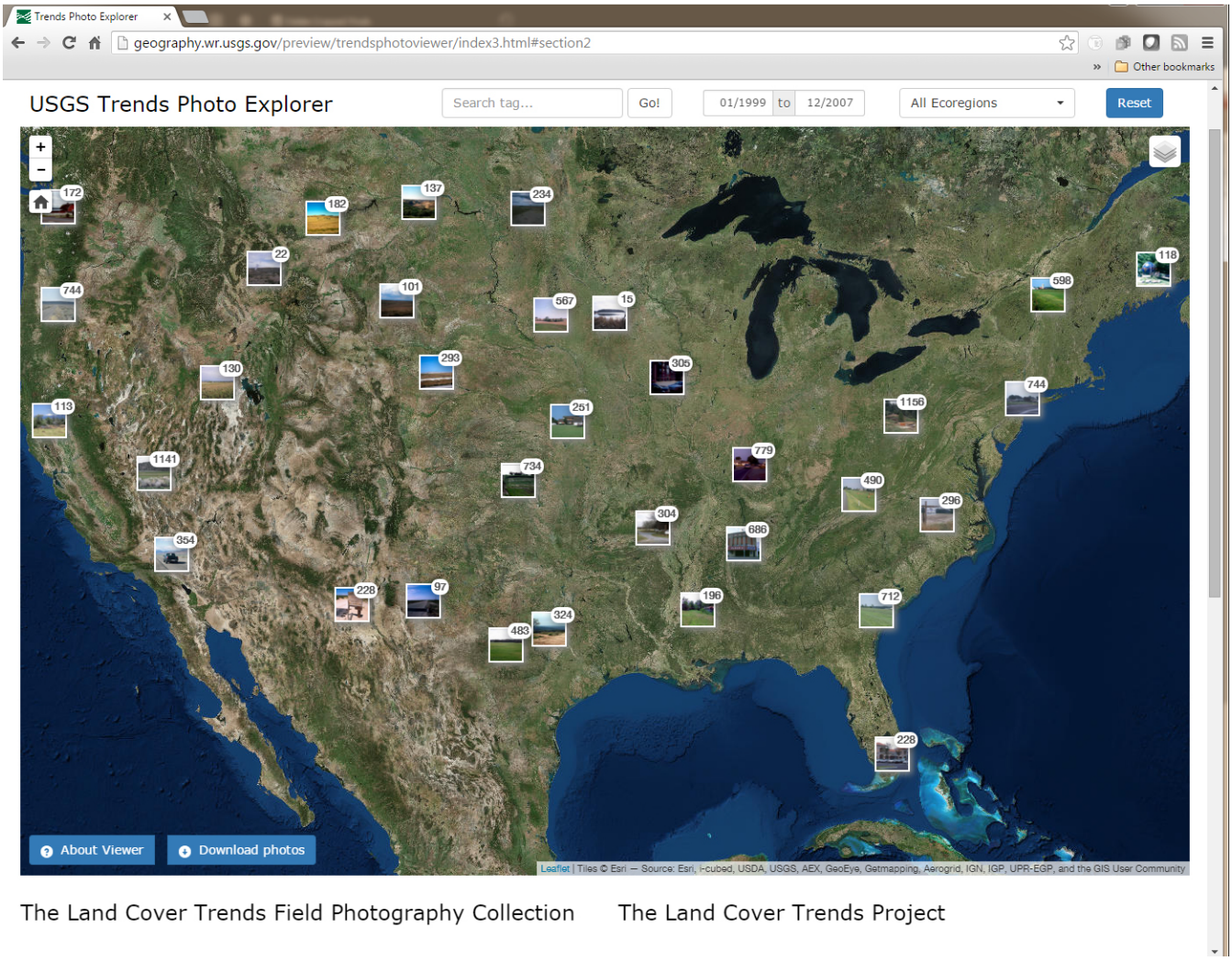


Figure 7. Screenshot of the Leaflet map interface of the USGS Trends Photo Explorer.

Figure 8. Screenshot of the USGS Trends Photo Explorer displaying 2011 National Land Cover Dataset and Land Cover Trends sample blocks to provide land-use/land-cover context to the photos.

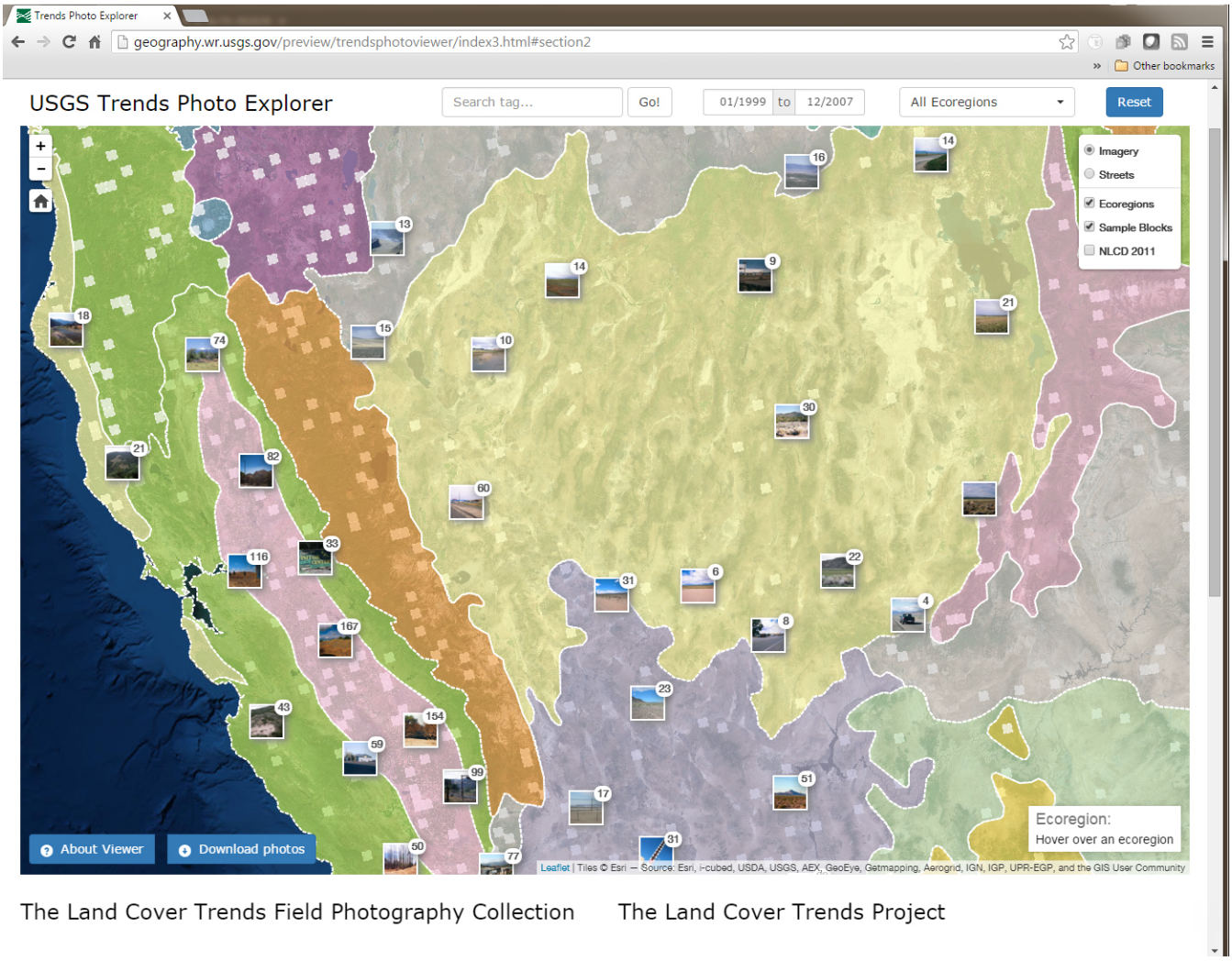


Figure 9. Screenshot of the USGS Trends Photo Explorer zoomed in and displaying Level III ecoregions and Land Cover Trends samples blocks.

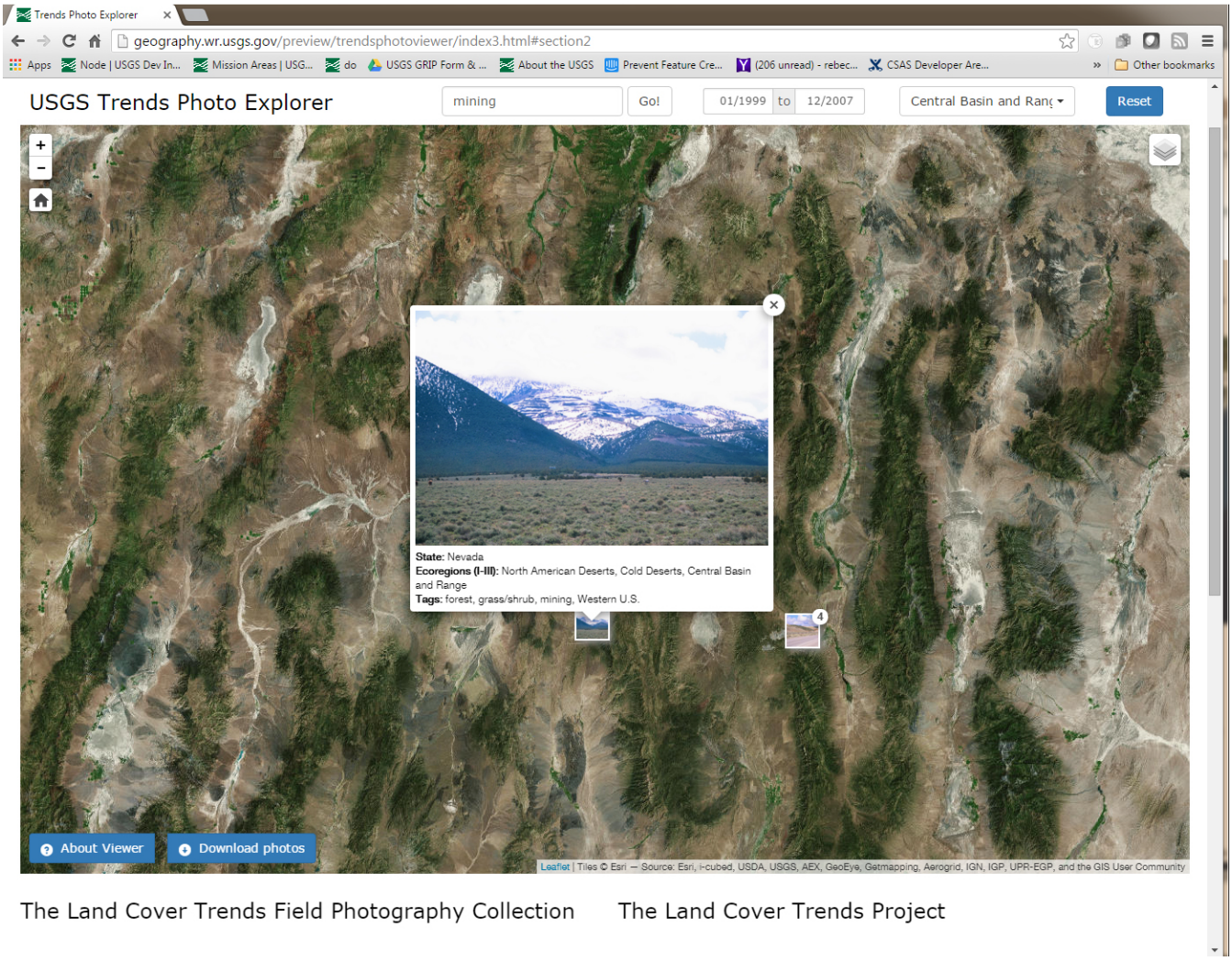


Figure 10. Screenshot of the USGS Trends Photo Explorer displaying a single photo and its metadata.

The team expects the field photos to be available via Earth Explorer early in FY 2016. The team also intends to generate a press release and possibly a USGS fact sheet to highlight the public release of the USGS Trends Photo Explorer Web page when the peer review process is complete. The team continues to geotag field photos, which will be made public as they become available.

Making Unmanned Aircraft System Data Available to USGS Scientists and the Public

Prior to this project, data acquired from the USGS Unmanned Aircraft Systems (UAS) have been provided to requesting scientists but have not been made available to the broader USGS community, the U.S. Department of the Interior (DOI) bureaus, or the public at large. This project performed a pilot study and developed a strategy that is scalable to evolve into a permanent UAS data management capability. The goal is to make UAS datasets available over the Internet to the USGS, DOI, and public science communities by establishing robust data management strategies and integrating these data with other geospatial datasets in the existing infrastructure located at the USGS EROS Data Center.

Accomplishments

The accomplishments for this project were as follows:

- The Long-Term Archive (LTA) project developed a data management plan that encompasses the definitions and structure needed for the various product types within UAS projects and metadata requirements as well as for providing services for electronic and (or) media deliveries (fig. 11) (Gacke, 2015a).
- The LTA designed a database, developed a data ingest process, and released five UAS product types through EarthExplorer (fig. 12). The data ingest capabilities created for this USGS CDI grant establish the framework for future UAS product deliverables with a pathway to end users that need access to the UAS datasets in a timely manner. The LTA UAS capability released in September 2015 supports the following data products: orthophotography, photos, point cloud, videos, and supporting documents.

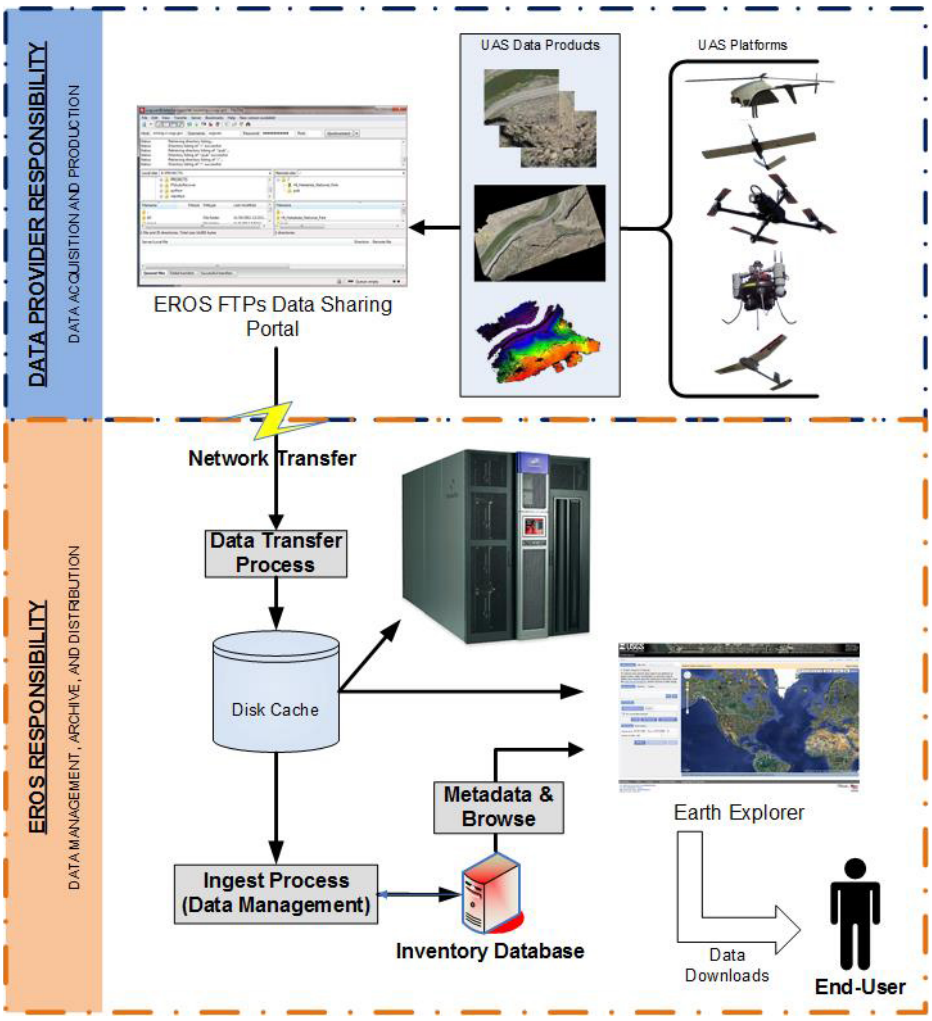


Figure 11. Unmanned aircraft system (UAS) data flow overview illustrating the end-to-end data flow from the initial acquisition of data to the final archival and distribution of products to the end-user community.

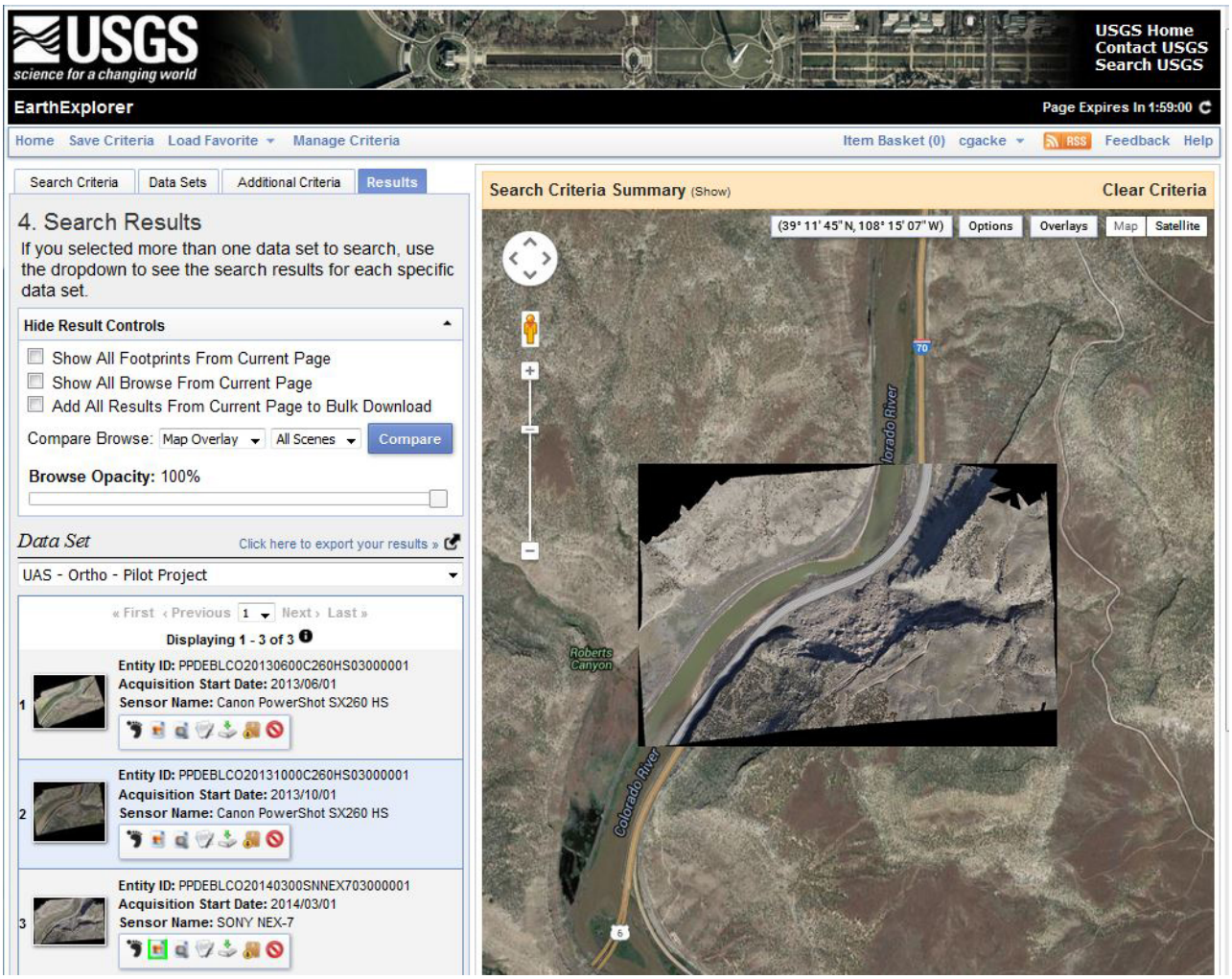


Figure 12. Unmanned aircraft system orthophotography product type example in EarthExplorer.

National Dam Removal Database—A Living Database for Information on Dying Dams

The United States has over 2 million dams on rivers and streams (Graf, 1999), and more than 84,000 of the larger dams are documented in the congressionally mandated National Inventory of Dams (U.S. Army Corps of Engineers, 2015). The average age of these National Inventory of Dams is 52 years; by the year 2030, over 80 percent will be at least 50 years old (American Society of Civil Engineers, 2015). As a result of this aging infrastructure, dam removal has increased during recent decades with the total number of removed dams estimated at around 1,200 (American Rivers, 2014).

Many factors drive downstream physical and biological responses following dam removal, with most rivers changing rapidly and demonstrating ecosystem resiliency (O'Connor and others, 2015). An emerging need for scientists, land managers, and communities facing decisions about dams is access to relevant scientific information about the physical, biological, and ecological responses of rivers and reservoirs to dam removal.

In response to this need, this project team used CDI funding to create a dynamic database of dam removals and the scientific studies associated with them. This project expands upon the work from a recently completed dam removal synthesis project at the John Wesley Powell Center for Analysis and Synthesis, where we created a relational database of scientific studies associated with dam removal responses (Bellmore and others, 2015). The current CDI project sought to transform Bellmore and colleagues static relational database into a dynamic data system that is accessible through an online interface, connected to existing USGS and partner databases, and interactive with viewing, searching, and accessibility functions. The nonprofit group American Rivers is an important partner to the project, with its comprehensive database on dam removals (American Rivers, 2014). Our goal is to increase the accessibility of dam removal science to researchers and managers by providing a centralized location of information related to dam removal studies.

As a first phase of the project, the project team created the Dam Removal Information Portal (DRIP; fig. 13) (U.S. Geological Survey, 2015j). This online Web site, powered by CartoDB and using data and services from USGS ScienceBase, currently contains tools for visualization and analysis of georeferenced dam removals and associated scientific studies contained in the American Rivers (2014) and Bellmore and others (2015) databases, respectively. From these databases, the team created a dam removal science registry in ScienceBase that contains properties about individual dams that have been removed (for example, location and size) as well as the associated scientific studies (both peer-reviewed literature and other report sources). The particular attributes of each study are cataloged and include features such as study design, study duration, and metrics studied as well as demographic characteristics about the studied dam(s). As such, each dam that has been removed and has associated scientific studies is represented in a ScienceBase collection, where additional information (for example, before and after imagery), digital object identifiers, and links to available studies are contained. The project team refers to this underlying architecture as the National Dam Removal Science Database (fig. 14).

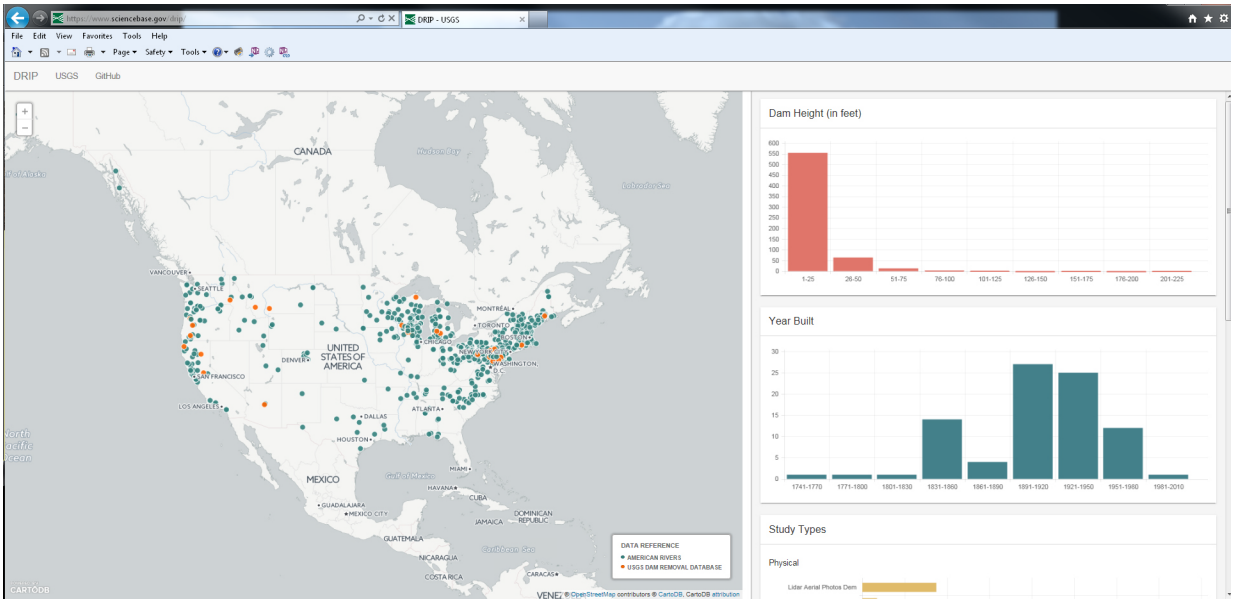


Figure 13. A screen shot of the USGS Dam Removal Information Portal (DRIP), a metaknowledge Web site providing information and associated scientific studies on U.S. dam removals.

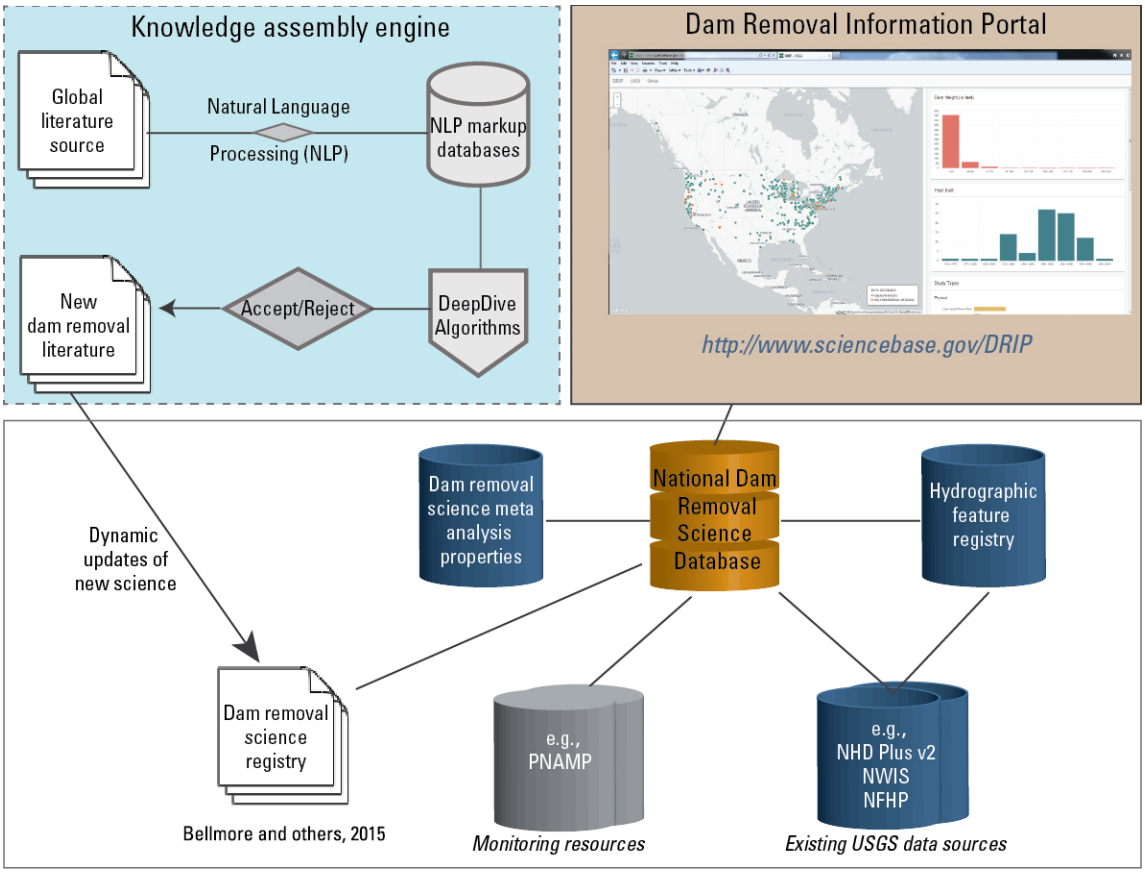


Figure 14. A conceptual diagram showing the architecture of the National dam removal science database (bottom panel), the Dam Removal Information Portal (DRIP; upper right), and a knowledge assembly engine (upper left) that are currently under development. Properties of scientific studies associated with georeferenced dam removal projects are linked with existing USGS cyberinfrastructure (for example, National Hydrography Dataset [NHD], National Water Information System [NWIS], the Pacific Northwest Aquatic Monitoring Partnership's [PNAMP] monitoring resources, and the National Fish Habitat Partnership [NFHP]) through a hydrographic feature registry to build the underlying database, which is viewed online through the DRIP interface. Based on a static relational database of all scientific dam removal studies through 2014 (Bellmore and others, 2015), future versions will incorporate a knowledge assembly engine, which will dynamically update new literature and associated properties into the database.

Additional functionality is being developed for the National Dam Removal Science Database and DRIP. For example, the locations of all dam removals are being associated with stream segments in the National Hydrography Dataset Plus version 2. The goal is to allow users to query existing hydrographic features (for example, basin area, river miles upstream of the dam, land use, and land cover) associated with individual projects or collections of dam removals. Linking to other existing USGS data sources (for example, NWIS) is also envisioned. The project team has taken steps to help ensure long-term viability of the National Dam Removal Science Database by developing the capability within the Biogeographic Information System, a larger overall platform maintained within ScienceBase by the Core Science Analytics, Synthesis and Libraries organization.

The next phase of the project, using supplementary, non-CDI funding acquired for FY 2016, will incorporate additional functionality into the database and its user interface. An important element of this work will be to continually update the dam removal science registry with new information as it becomes available in the scientific literature. We are exploring possibilities of automating locating, processing, and extracting data from new dam removal studies from literature sources by creating a Knowledge Assembly Engine. This feature is based on DeepDive technology (Niu, 2012), which is a machine learning system for extracting structured information from text (for example, online PDF versions of literature sources). Another possible feature that the team is exploring is implementing an interface so that users have the ability to upload basic information about new dam removals and dam removal studies via the DRIP interface.

Providing a comprehensive and accessible knowledge base about the science of data removal is one of the anticipated overarching benefits of this project. Dam removal is likely to continue to be an important part of national and international responses to aging infrastructure and an important method for river restoration. Natural resource managers, practitioners, and scientists need to use the best available information to plan and conduct dam removal projects as well as to conduct effectiveness research and monitoring to track and understand important ecosystem responses.

Accomplishments

The accomplishments for this project were as follows:

- *USGS DRIP*.— Web site available at http://www.sciencebase.gov/drip
- *USGS open-file report*.—On the structure of the National Dam Removal Science Database and the DRIP (in preparation)

Portable ISO 19115-2 Open Source Developer's Toolkit

Over the last few years, the ISO 19115 family of metadata standards has become the predominantly accepted worldwide standard for sharing information about the availability and usability of scientific datasets among researchers. The U.S. interests in the ISO standard have also been growing as global-scale science demands participation with the broader international community; however, adoption has been slow because of the complexity and rigor of the ISO metadata standards. In addition, support for the standard in current implementations has been minimal.

In 2009, the Alaska Data Integration Working Group members (ADIwg) mobilized to jointly address common data integration efforts. Beginning in 2012, ADIwg started to focus on difficulties associated with generating and exchanging metadata using the ISO family of standards. Like a microcosm of the larger world, ADIwg partners vary in size from small nongovernmental organization research groups to state offices, universities, and large Federal bureaus. This differential in size and technical capabilities among its membership brings with it the predictable diversity in metadata requirements and ability to provide the necessary technical support to its researchers.

The ISO 19115 metadata standard is not a single standard but a family of ISO and Geography Markup Language Encoding standards. Understanding this amalgamation of standards to a degree sufficient to generate a valid metadata record requires an investment of time well beyond what can reasonably be expected of any PI or end user. These issues are not

specific to ADIwg and its members; they are shared by many organizations striving to transition to the ISO metadata standard.

After much discussion, the ADIwg membership decided to co-author an ISO software toolkit with an architecture that has the flexibility to achieve the following goals:

- Isolate the complexity, rigor, formatting, and terminology of the ISO standards from the PI and (or) data steward. Users should not be required to be experts on the ISO standards to generate ISO metadata.
- Provide clear developer and user documentation.
- Design a layered architecture and provide developers with access to each layer. Developers should be able to enter the architecture at whatever layer best fits their needs. For example, they should be able to download the code library to do custom programming or post JSON metadata to a hosted JSON-to-ISO translator.
- Design the architecture to accommodate incremental development cycles and add features without disrupting previous implementations.
- Implement all code and services as open-source software projects using a GitHub repository (Alaska Data Integration Working Group, 2015a).
- Use simple JSON for storing and transferring user metadata.
- Provide flexibility to write to multiple metadata standards, including ISO 19115-2:2009, 19115-1:2014, 19110, and possibly the Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata.
- Provide an online, interactive, client-side, end-user metadata editor written in JavaScript to gather and edit a user's metadata and format it in the mdJSON format.
- Host a publicly available version of the online metadata editor for end users to enter and edit metadata and then request a valid ISO standard metadata record to be returned.
- Write all code in popular, well-supported computer languages that are platform independent, royalty-free, and available through an open-source repository to encourage participation from and benefit to the widest possible user base.

Accomplishments

The FY 2015 accomplishments for this project were as follows:

- ADIwg renamed mdJSON-schema-viewer to mdTools to present a friendlier name and reflect extended capabilities added to the tool (Alaska Data Integration Working Group, [n.d. a]).
- The project team finalized and published the nonbeta release of mdTools 1.0.0 and updates (1.1.0) with extensions for the following features:
  - notification of deprecated portions of the mdJSON schema
  - support for permalinks, search, and HyperText Markup Language (HTML) writer
  - additional tab for viewing of code lists maintained in mdCodes
  - support for mdJSON schema versions 1.0.0 and 1.1.0
- The project team presented a full afternoon workshop at the 2015 CDI Workshop in Denver on May 11, 2015 (Alaska Data Integration Working Group, 2015b).
- The project team gave a presentation to the ISO Geographic information/Geomatics Technical Committee at the Standards in Action Workshop in London in June 2015.
- The project team finalized and published nonbeta release of mdTranslator 1.0.0 and issued numerous point releases improving code architecture and adding the following

features in versions 1.1.0, 1.1.1, 1.2.0, 1.2.1, and 1.3.0 (Alaska Data Integration Working Group, [n.d. b]):

- support for coverage information, image information, sensor information, classified data, and grids
- replaced language and character set with local to position for 19115-1
- ability for user provided Cascading Style Sheets in HTML writer
- code refactoring to remove all global variables
- HTML writer for easy online viewing of metadata records
- an allowance for multiple data dictionaries

- The project team finalized and published nonbeta release of mdCodes 1.0.0 and issued numerous point releases improving code architecture and adding the following features in versions 1.0.1, 1.1.0, and 1.2.1:
  - ISO 3166-1 alpha-3 country code list
  - ISO 639 Part 2 language code list
  - Internet Assigned Numbers Authority character set code list
  - code lists for cell geometry, dimension name, image condition, and coverage content
- ADIwg selected Ember.js (Katz and others, 2015), Ember-CLI (Ember-CLI, [n.d.]), and Bootstrap (Otto and Thornton, [n.d.]) as development architecture for mdEditor and developed the following elements:
  - "proof-of-concept" prototype using Ember.js framework
  - screenshots and navigation strategy using Balsamiq Mockup in Google Docs (fig. 15)
  - "live" HTML layout based on mockups to test user experience and Bootstrap components (fig. 16) (Alaska Data Integration Working Group, 2015c)
- The project team documented many of the fundamentals of the ISO Developer's Toolkit as mdBook using the GitBook publishing service (Alaska Data Integration Working Group, 2015d).

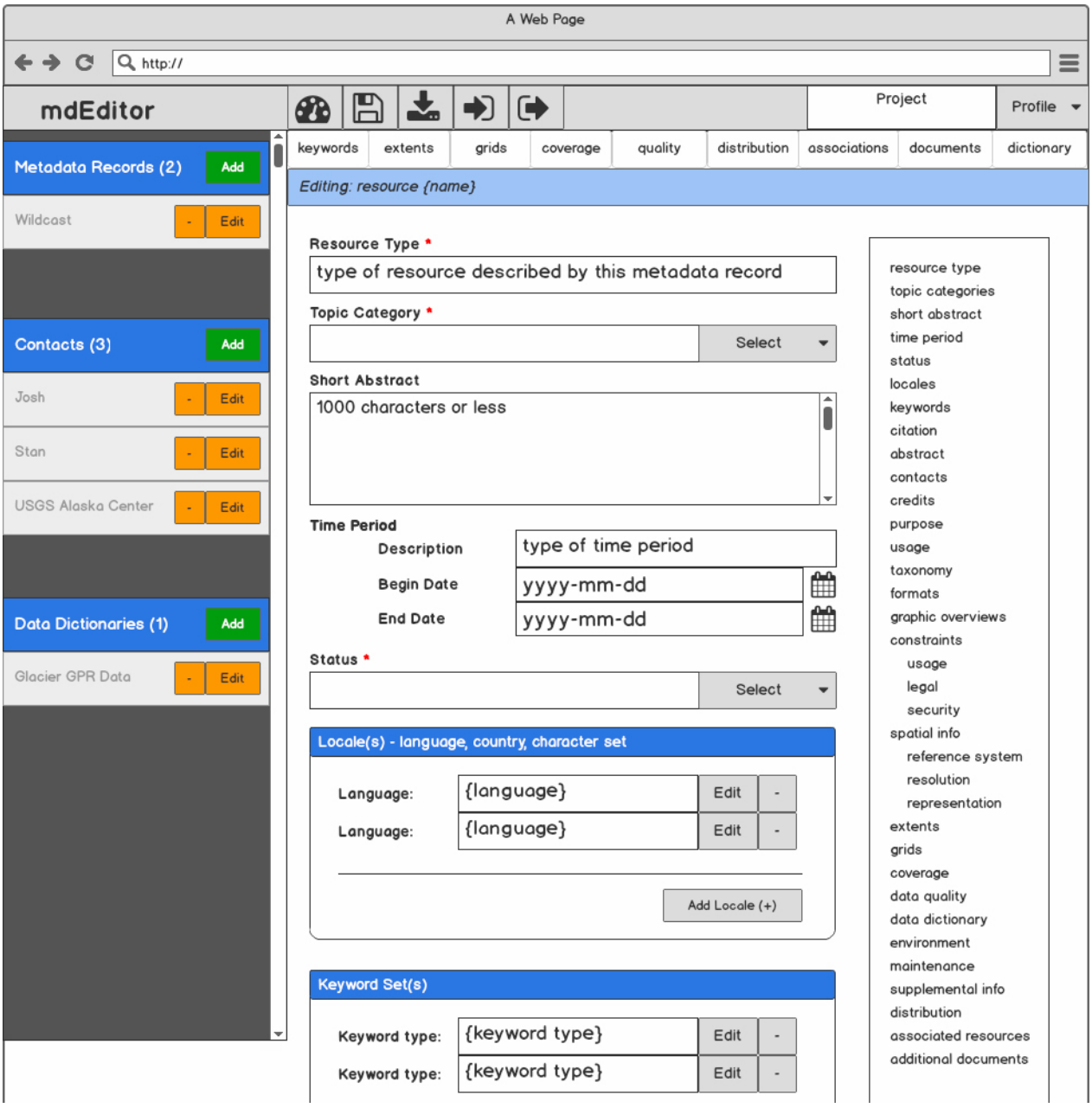


Figure 15. Balsamiq Mockup of mdEditor user interface.

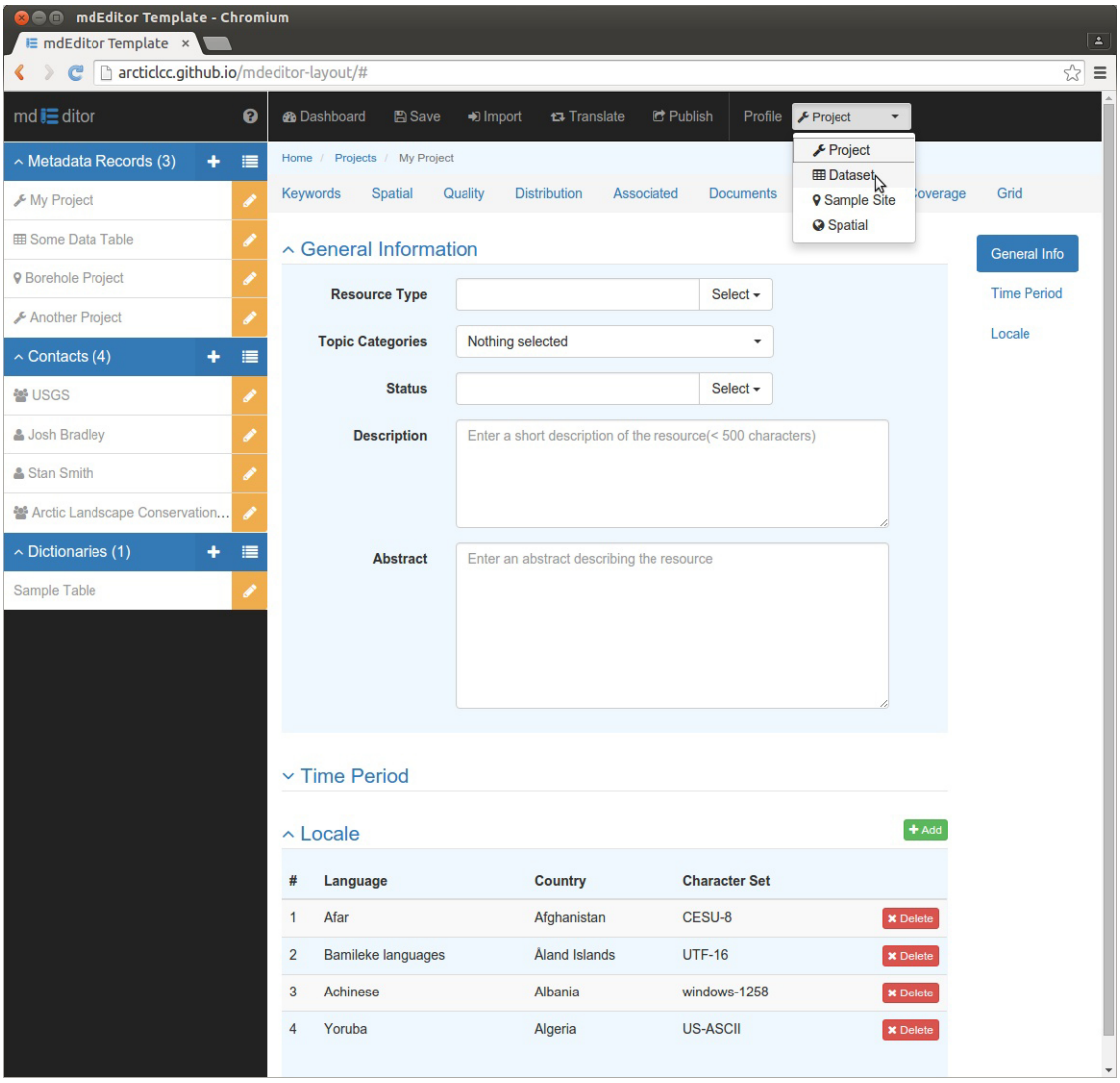


Figure 16. Hypertext markup language (HTML) version of mdEditor interface using Bootstrap framework.

The Ruby Gem (Ruby language code library) for mdTranslator has been downloaded more than 9,000 times as of October 2015. To fulfill its promised role, the Toolkit will require upgrades, extensions, and support for years to come.

The Toolkit has been, or is planned to be, implemented in multiple environments relating to USGS operations. USGS ScienceBase has planned to implement the Toolkit as one of its primary engines for generating ISO metadata. The Landscape Conservation Cooperatives and Climate Science Centers anticipate using the mdEditor and mdTranslator as a tool for scientists to document and format their metadata content. The Arctic Landscape Conservation Cooperatives, University of Alaska, and Alaska Science Center are currently at various stages of integrating the Toolkit into existing systems.

There is also interest in the Toolkit and mdEditor from other national and international science organizations desiring to participate in the ISO metadata community. Among these are the Interagency Arctic Research Policy Committee members, Sustaining Arctic Observing Networks, the Arctic Data Committee, and the Polar Data Forum.

Work is anticipated to continue into 2016 and beyond depending on funding availability. The following items are planned to be delivered in FY 2016–2017:

- release of initial public version of mdEditor in March 2016
- a writer application for FGDC compliant metadata
- a writer application for ISO 19115-1 compliant metadata
- a possible reader application for FGDC metadata

sbtools—An R Package for ScienceBase

Science is an increasingly collaborative endeavor. In an era of Web-enabled research, new tools reduce barriers to collaboration across traditional geographic and disciplinary divides and improve the quality and efficiency of science. Collaborative online code management has moved project collaboration from a manual process of email and thumb drives into a traceable, streamlined system where code can move directly from the command-line onto the Web for discussion, sharing, and open contributions. Within the USGS, however, data have no such analogous system. To bring data collaboration and sharing within the USGS to the next level, we are missing crucial components.

The sbtools project team built sbtools, an R interface to ScienceBase (fig. 17). To build this package, the team organized a diverse group representing experts in data management, science, and research-software development domains. This interface provides scripted R access to ScienceBase to manage metadata and data files, to search the catalog of datasets, and to view and modify data in formats familiar to R users. The package is designed to keep most interactions simple and includes internal handling of complex Web service interactions (for example, authentication) to simplify use for end-users. With its release, sbtools enables rapid and reproducible access to one of the single largest repositories of Earth-science data and an advanced cloud-based data collaboration platform. sbtools will add value to ScienceBase for scientists, developers, and users of R in the USGS. The project team will release and distribute the code via the Comprehensive R Archive Network (CRAN) and USGS-R repositories.

Figure 17. A Twitter announcement of the sbtools package development collaboration was well received by the wider community (link to tweet: https://twitter.com/USGS_R/status/636606584253079553).

Accomplishments

The following accomplishments have been made or are in progress as of September 2015:

- sbtools R package available in the following locations:
  - Geological Survey R Archive Network (GRAN) (U.S. Geological Survey, 2015k)
  - GitHub (Winslow, 2015)
  - CRAN (pending FSP review and approval of sbtools publication)
- sbtools publication (submitted for review October 2015)

Standards-Based Integration and Delivery of USGS and EPA STORET Biomonitoring Data via the Water Quality Data Portal

The purpose of this project was to test and develop first-generation biological data integration and retrieval capabilities for the Water Quality Portal (National Water Quality Monitoring Council, [n.d.]) using the Water Quality Exchange (WQX) data exchange standard (Environmental Information eXchange Network, 2016). The Water Quality Portal (Portal) is a significant national water data distribution node that is aligned with the vision of the Open Water Data Initiative (Advisory Committee on Water Information, [n.d.]). The Portal is sponsored by the USGS, the EPA, and the National Water Quality Monitoring Council. The WQX data exchange standard is a mature standard widely adopted within the water quality monitoring community and is used to support data flows into EPA's STORET data warehouse. The WQX standard also provides the data integration framework for physical-chemical water quality data

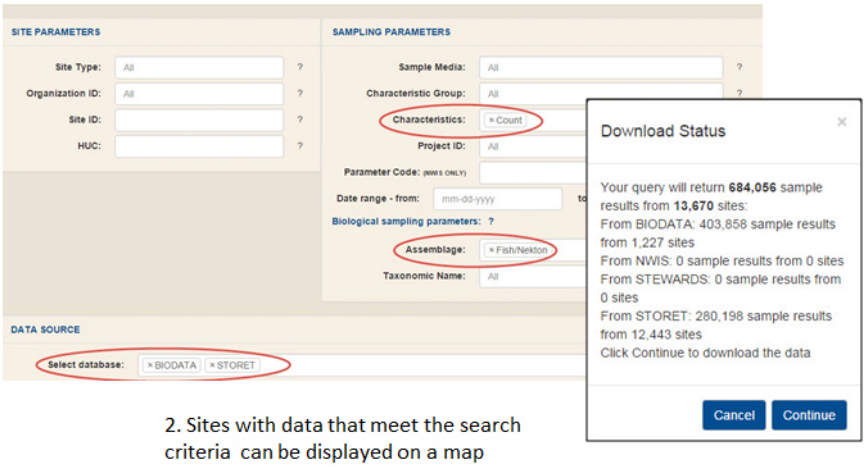
currently served by the Water Quality Portal. The standard supports taxon abundance data including population census, frequency class, group summaries, individual results, biological index scores, and biological scoring metrics, but prior to the inception of this project, the Portal did not deliver those types of data.

The specific objective of this project was to integrate and deliver fish community data from two large Federal data sources: (1) the USGS BioData system and (2) USEPA STORET. The processes, tooling, and experience gained can then be applied to incorporate additional biological data from these and other sources.

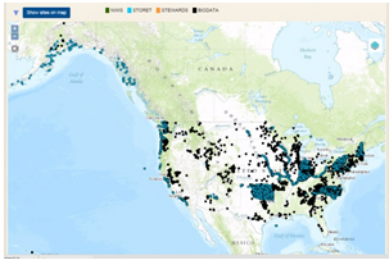
Accomplishments

This project team successfully developed the ability for the public to retrieve BioData and STORET fish community data from the Water Quality Portal Web site (fig. 18). This was accomplished in July 2015 and provides access to over 680,000 BioData and STORET fish abundance records for over 13,000 sites. As of February 2016, over 11,000 download requests have included BioData fish community data.

1. Users set search criteria to find the data they are interested in.



2. Sites with data that meet the search criteria can be displayed on a map



3. Users download the data in one of three formats (Microsoft Excel, csv, or tab-delimited)

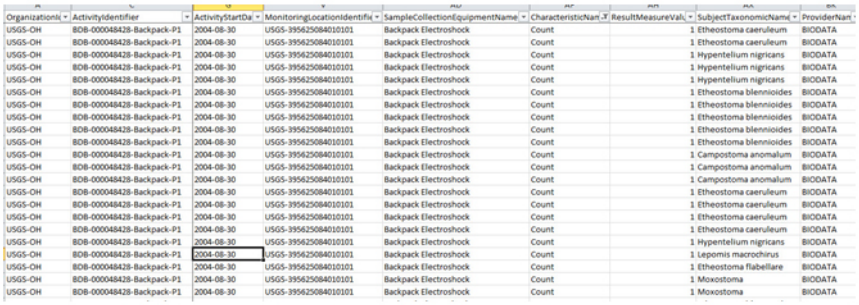


Figure 18. Workflow for retrieving BioData and STORET fish community data from the Water Quality Portal Web site (http://waterqualitydata.us/portal/).

The Digital Grain Size Web and Mobile-Computing Application

This project team developed a Web-hosted application (that can also be used on mobile platforms) for automatic analysis of images of sediment for grain-size distribution, using the "Digital Grain Size" (DGS) algorithm of Buscombe (2013) ("DGS-Online," 2015). This is a free, browser-based application for accurately estimating the grain-size distribution of sediment in digital images without any manual intervention or even calibration. It uses the statistical algorithm of Buscombe (2013) that estimates particle size directly from the spatial distribution of light intensity within the image. The application is designed to batch-process tens to thousands of

images, utilizing cloud computing storage and processing technologies. Typical processing times are 1–60 seconds per image, depending on the size of file and the user inputs.

At the Web site, the user can (1) login to their personal dashboard using a Google account (gmail address); (2) create a "job" which involves uploading sediment imagery and assign "tags" and processing options to each image separately, to groups of images, or to all images at once; (3) launch a Web application to upload and process imagery on a cloud server; (4) and download grain-size distributions and other statistics in three formats (csv, xml, and json); and (5) create graphs of their results. The user will be able to store their imagery on the server for reprocessing images and redownloading data. The user can create as many separate jobs as they wish, privately and securely on their own personal dashboard, from which they can share results through their browser, securely, with whomever they wish.

Accomplishments

- *pyDGS Program*.—A streamlined and faster version of the pyDGS program (version 3.0.0). This version of the code is running on the Web application, and is also publicly available through GitHub (Buscombe, 2015a) and python package index (Buscombe, 2015b).
- *DGS-Online*.—A scalable Web application running node.js and python 2.7 on the Heroku platform, utilizing Amazon Web services simple queue service, simple storage service data storage, DynamoDB NoSQL cloud database, and Amazon Elastic compute cloud technologies (fig. 19). The application is scalable if the tool proves to be very popular.
- The *DGS-Online Web site* (https://www.digitalgrainsize.org).—This is where users launch and interact with the program and read Web pages related to all aspects of the project, including (1) rationale, (2) history of algorithm development, (3) program user guide, (4) detailed information on how the algorithm works, (5) comprehensive FAQs, and (6) advice and best practices on image collection (fig. 20). As of October 30, 2015, the documentation is almost complete, and the Web site functionality is almost complete, with finishing work still in progress on graphics, design, logo, and domain name..

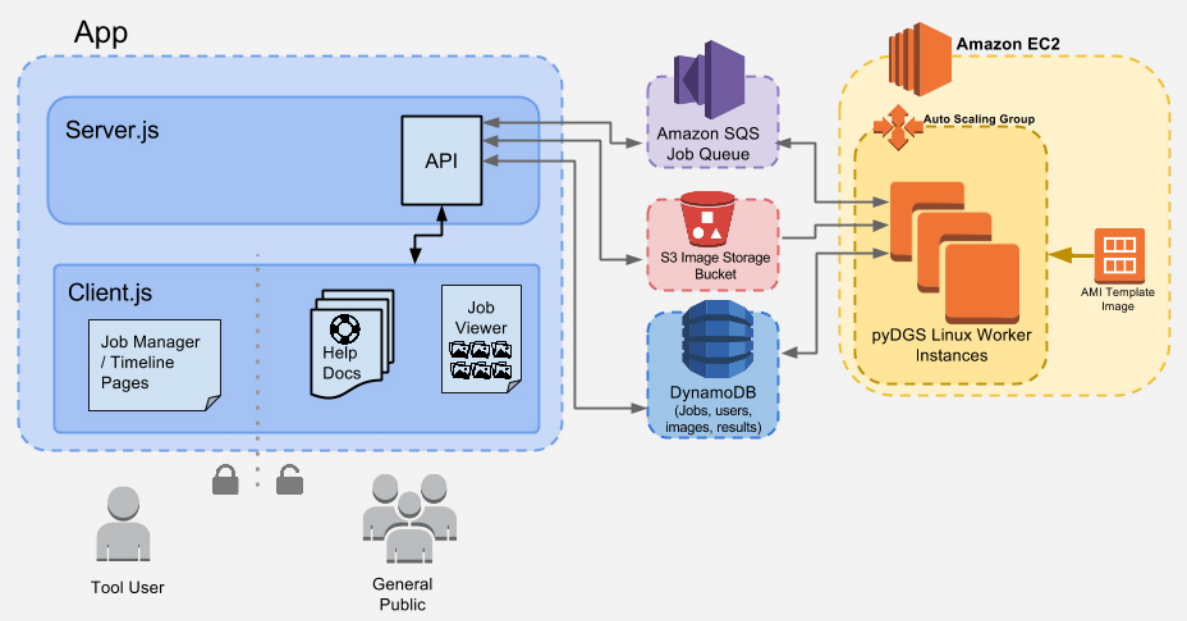


Figure 19. Schematic of the general computing workflow employed by DGS-Online, the "Digital Grain Size" Web Computing Application. API, application programming interface; EC2, Elastic compute cloud; S3, Simple Storage Service; AMI, Amazon Machine Images.

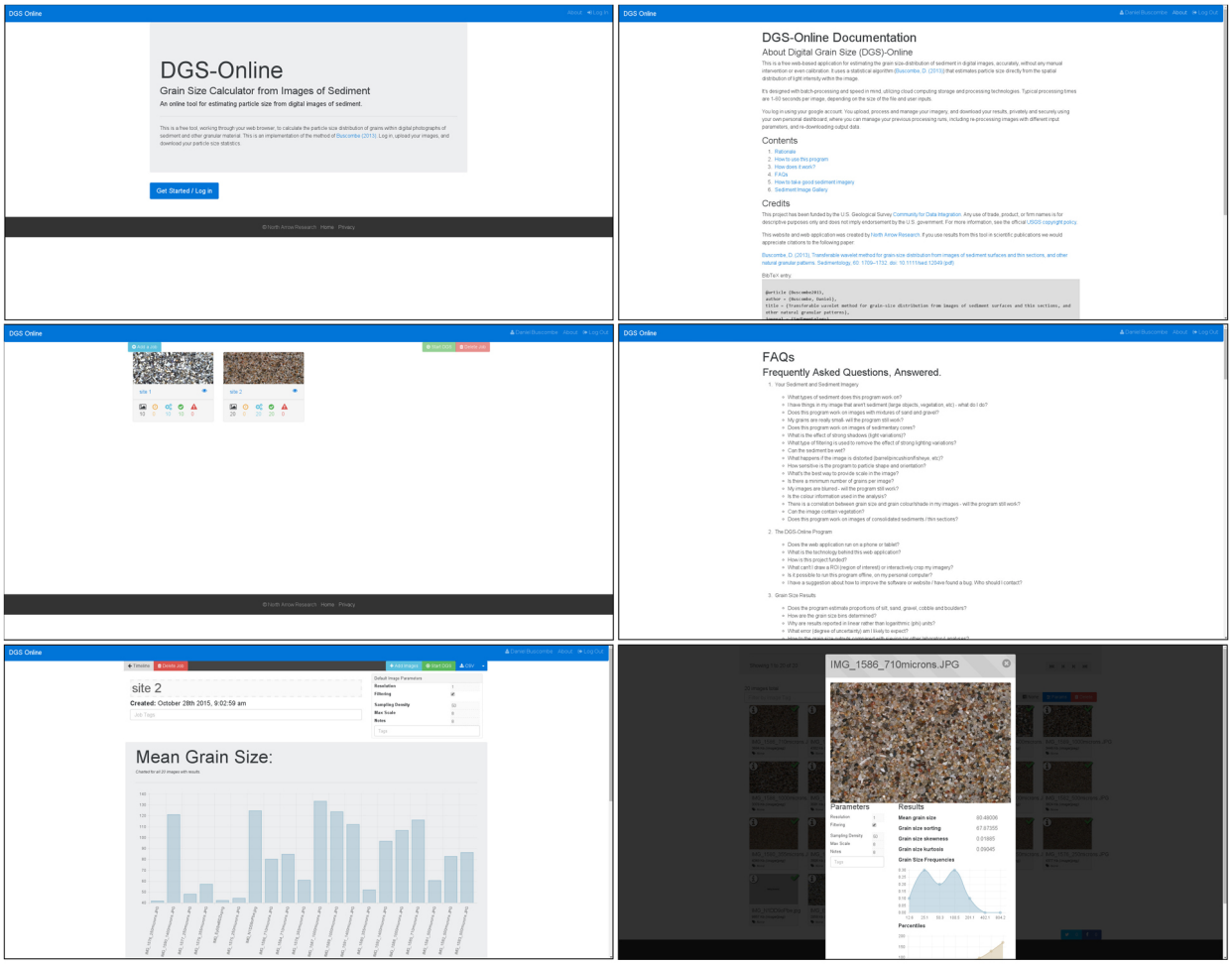


Figure 20. Screenshots of the fully-functioning prototype DGS-Online Web application, as of October 30, 2015, running on Google's Chrome Web browser.

Use of Controlled Vocabularies in USGS Information Applications—Requirements Analysis for Automated Processes and Services

Large online data catalogs use controlled vocabularies to categorize datasets in ways that allow end users to sort and select data matching their needs. The eventual goal of this project is to build functional services so that the USGS Thesaurus and other USGS-controlled vocabularies will be available to the English-speaking scientific community, especially within the USGS where they can be used to improve metadata quality and data discovery.

The project team used the Tetherless World Constellation (TWC) Semantic Web Methodology, which is designed to examine use cases and determine both functional and nonfunctional system requirements without prejudicial commitments to meeting those requirements by utilizing particular technologies, platforms, hardware, or software. This iterative process was developed in 2008 by Peter Fox and Deborah McGuinness of the TWC at Rensselaer Polytechnic Institute and has been taught by Fox and his collaborators to members of the project team (Fox and McGuinness, 2008). During the first year of the project, the team

developed a set of use cases and a conceptual model and engaged a panel of expert reviewers to evaluate them (fig. 21). Afterward, the resulting system requirements were tested by developing prototype vocabulary services and by modifying an existing USGS metadata tool, the Metadata Wizard, to make use of the vocabularies offered by the new services. The proposal also included modifying the Science Data Catalog to make use of the controlled vocabularies. The project team planned the catalog modifications; however, the plans were not implemented because of a shortage of metadata that included controlled terms. Finally, and largely as a result of input from the expert review panel, the project team drafted a "Controlled Vocabulary Manifesto" that proposes a strategy for full implementation of controlled vocabularies in USGS in order to enable people using USGS data catalogs to be confident that their search results are both comprehensive and focused, with good *recall* (nothing relevant missed) and good *precision* (nothing irrelevant included).

The project has taken the first steps toward development of Web services and applications that will enable researchers and data managers to use community-standard vocabularies so that USGS data can be found as easily as possible, especially when people do not already know that those data exist. In the CDI SSF, the project is integrating semantics into Web services and applications principally to support the "Describe" (Metadata) and "Publish/Share" components of the Science Data Lifecycle Model.

Accomplishments

- Use case documentation revised. See link at https://my.usgs.gov/confluence/display/cdi/Use+Cases+for+Vocabulary+Web+Services.
- Review panel report completed. See link at https://my.usgs.gov/confluence/display/cdi/Use+Cases+for+Vocabulary+Web+Services.
- Prototype vocabulary services documented at http://www.usgs.gov/science/services.html.
- Metadata Wizard, version 1.7 with documentation and file download available at http://www.sciencebase.gov/metadatawizard.
- Controlled Vocabulary Manifesto finished. See link at https://my.usgs.gov/confluence/display/cdi/SWWG+2014+Proposal+Project.

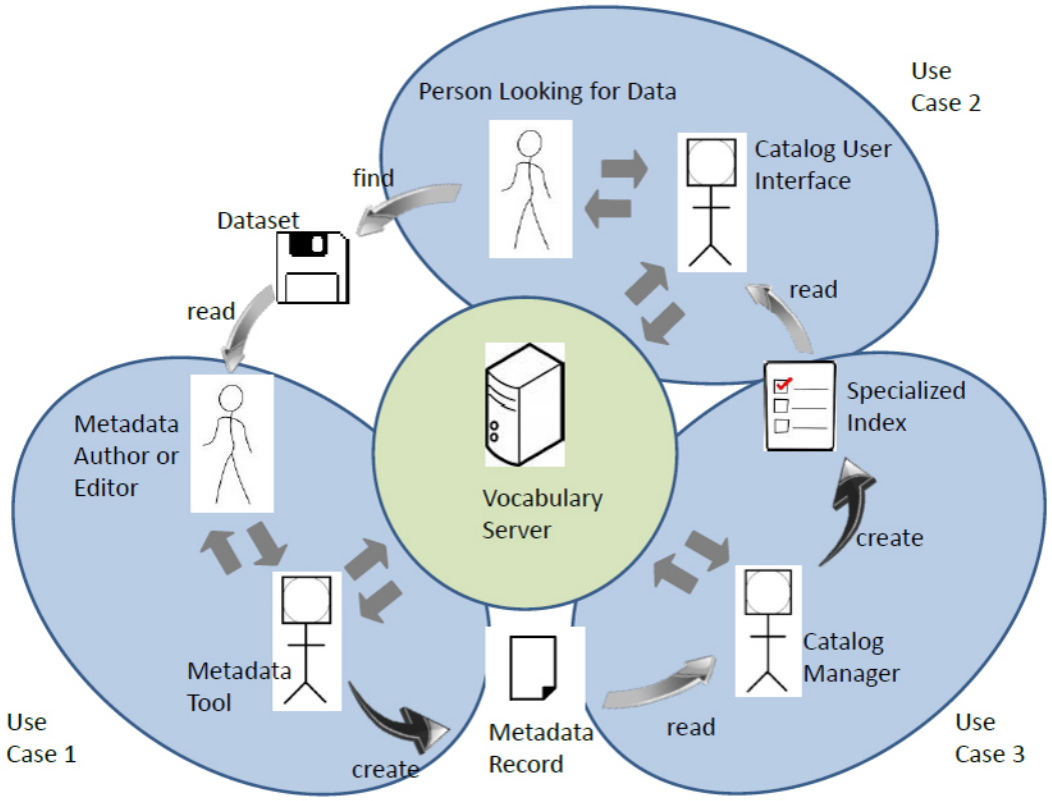


Figure 21. Diagram showing the three use cases and their relationships to each other and to the vocabulary server (center) that provides the services that "drive" the use cases.

Web-Enabled Visualization and Access of Value-Added Disaster Products

The purpose of this project was to support the enhanced search, access, and visualization capability for disaster maps and other contributed products on the public USGS Hazards Data Distribution System (HDDS) (U.S. Geological Survey, 2015). These products are often provided to USGS by collaborators for sharing across the response community during the course of an emergency event response; however, in the past, they were not easy for users to discover or access. This project involved the design, testing, and delivery of a new capability for HDDS to ingest, catalog, and display informational or value-added products when provided in a variety of formats. As a result of this work, the user community will be able to interactively search, preview, and access these products alongside the remotely sensed imagery (satellite and aerial photography) already available through the HDDS interface.

The HDDS is a public USGS-hosted Web portal that provides a consolidated point-of-entry and distribution system for remotely sensed imagery and other geospatial datasets related to emergency response. When disasters occur, the system provides a critical source of satellite and aerial imagery for the emergency response community. The imagery and datasets on HDDS include imagery collected by USGS as well as contributed datasets from many other government agencies and collaborators. After ingest, the HDDS-hosted imagery is accessed by end users from all levels of government (Federal, state, local, tribal, and international) as well as many

other organizations and communities engaged in emergency event support. This project supported the expansion HDDS ingest capabilities to include user-contributed maps and other value-added products, allowing them to be more easily shared, discovered, visualized, and accessed by the user community.

With system and software developments supported by CDI, the HDDS now includes the capability to ingest, catalog, and display maps and other information products provided by data contributors. The newly supported product types include image-based maps in the form of Portable Document Format (PDF), Joint Photographic Experts Group (JPEG), Motion Pictures Expert Group (MPEG), Tagged Information File (TIF), and Microsoft Word files as well as vector products such as Keyhole Markup Language (KML) and shapefiles .

Along with the HDDS system development, a supporting document ("HDDS Product Specifications and Services") was created to provide technical guidance for data contributors and collaborators seeking to share their products via HDDS (Gacke, 2015b). This document provides the detailed product and metadata specifications (format, file naming, and more) required to support HDDS ingest along with data transfer instructions and requirements.

As a result of these CDI-funded activities, the contributed maps and other products can be more easily shared across the emergency response community. For any ingested product, the user community is now able to interactively search, preview, and access the map products directly through the existing map-based HDDS interface (fig. 22).

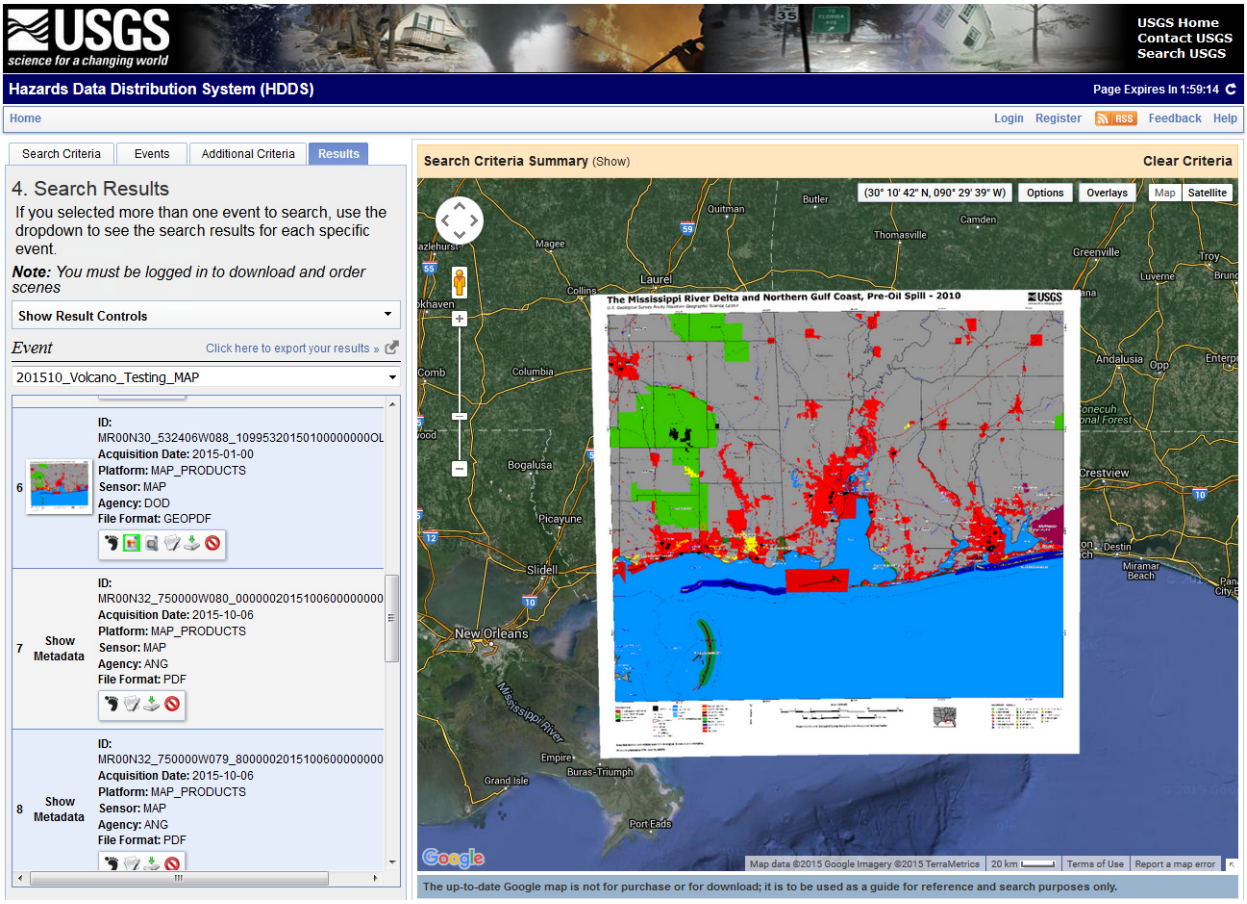


Figure 22. Screen example showing the Hazards Data Distribution System (HDDS) interface with the search results and "browse overlay" view now available and integrated for ingested map products.

Accomplishments

- *Technical Document*.—The "HDDS Product Specifications and Services Document" was developed to provide detailed information and guidance on acceptable product format(s), metadata requirements, and other supporting information for contributors of incoming products. The document also provides detailed instructions for product transfer to the USGS and HDDS systems. As of October 30, 2015, the specification document has been finalized and is now available for distribution to cooperators and contributors.
- *HDDS Ingest and Archive Software*.—The internal HDDS system capability was modified to support map product ingest, including metadata extraction and browse/thumbnail image creation. The metadata, browse, and ingested products are placed into the existing HDDS system catalog for potential access by end users.
- *Web-Based Search/Query and Display*.—User discovery and visualization of disaster maps and other products are now supported by the HDDS Web interface at http://hddsexplorer.usgs.gov/ and presented to the end user in a fully integrated manner alongside other available image datasets. The system now allows user-defined search queries, metadata display, graphical visualization on a map background, and full download options according to image license/access provisions.

Summary

The grassroots nature of the Community for Data Integration (CDI) has enabled the community to accomplish a tremendous amount of work in fiscal year (FY) 2015 through its monthly forums, workshops, working groups, and projects. Through these activities, the community has done a great deal of sharing, learning, and collaborating this year. The CDI hosted 13 presentations at monthly forums from both USGS participants and external partners, in addition to dozens of presentations hosted by CDI's working groups. These presenters shared outcomes from previous CDI projects, new tools, and applications from the USGS and other Earth science organizations, and activities taking place in the broader Earth science community, such as the U.S. Department of the Interior , the Federation for Earth Science Information Partners, United States Group on Earth Observations, and the National Aeronautics and Space Administration (NASA).

The face-to-face 2015 CDI Workshop, which took place in Denver, Colorado, was an excellent occasion for CDI community members to forge even stronger relationships and collaborations across mission areas and regions. It also provided a great opportunity to learn about the new USGS data management policies and tools to help us comply with these policies as well as update the community on new and emerging technologies and current USGS projects. Two outcomes of the workshop were the development of the new Earth-Science Themes Working Group and the recruitment of Tim Quinn as Co-Executive sponsor for CDI.

All of the CDI working groups were able to host breakout sessions during the CDI 2015 Workshop. During these breakout sessions, the working groups discussed their many accomplishments during the year, from the Data Management Working Group's Data Management Webinar Series, which brought in 40 to 80 participants every month, to the work of the Citizen Science Working Group in the Federal sphere. All of the working group activities helped the USGS cover new ground and provide superb learning opportunities for each of their members.

The CDI 2015 Workshop was also an excellent stage for FY 2015 principle investigators and collaborators to share their upcoming projects with the community, solicit feedback on the proposed work, and build excitement about the end products. These projects were, in part, selected by the community through the Request for Proposal's Statement of Interest voting process, and it was beneficial for project teams to interact with and get feedback from the community once again at the DataBlast poster session during the initial stages of their projects. Now that the FY 2015 projects have been completed, it is clear that they have increased the accessibility of USGS data by documenting, integrating, and exposing existing USGS datasets through new workflows, applications, and Web services.

The FY 2015 work described in this report demonstrates that the CDI community is meeting the goals that it has established for itself. The community is focusing on targeted efforts that yield near-term benefits to science, have substantial return on investment, leverage existing capabilities, and ensure sustainability of products and services. The CDI working groups and projects continue to identify and demonstrate innovative solutions and best practices for handling major challenges and sharing them with the broader Earth-science community.

References Cited

Advisory Committee on Water Information, [n.d.], Open water data initiative overview: Water Information Coordination Program, accessed on March 29, 2016, at http://acwi.gov/spatial/owdi/.

Alaska Data Integration Working Group, [n.d. a], mdTools: Alaska Data Integration Working Group, accessed October 29, 2015, at http://mdtools.adiwg.org/.

Alaska Data Integration Working Group, [n.d. b], Welcome to the mdTranslator: Alaska Data Integration Working Group, accessed October 29, 2015, at http://mdtranslator.adiwg.org/.

Alaska Data Integration Working Group, 2015a, Alaska Data Integration Working Group—Working on methods to efficiently integrate and share data: Alaska Data Integration Working Group, GitHub repository, accessed October 29, 2015, at https://github.com/adiwg.

Alaska Data Integration Working Group, 2015b, mdWorkshop: GitHub, accessed October 29, 2015, at https://github.com/adiwg/mdWorkshop.

Alaska Data Integration Working Group, 2015c, mdEditor Prototype: GitHub, accessed October 29, 2015, at https://github.com/adiwg/mdEditor/wiki/Prototype.

Alaska Data Integration Working Group, 2015d, mdBook—Documentation for ADIwg open source metadata tools: Alaska Data Integration Working Group, GitBook, accessed October 29, 2015, at https://www.gitbook.com/book/adiwg/mdbook/details.

American Rivers, 2014, American rivers dam removal database: American Rivers, accessed May 20, 2015, at http://www.americanrivers.org/initiative/dams/projects/2014-dam-removals.

American Society of Civil Engineers, 2015, 2013 report card for America's infrastructure: American Society of Civil Engineers, accessed October 30, 2015, at http://www.infrastructurereportcard.org.

Bellmore, J.R., Vittum, K.M., Duda, J.J., and Greene, S., 2015, USGS dam removal science database: U.S. Geological Survey data release, accessed November 24, 2015, at http://doi.org/10.5066/F7K935KT.

Buscombe, D., 2013, Transferable wavelet method for grain-size distribution from images of sediment surfaces and thin sections, and other natural granular patterns: Sedimentology 60, p. 1709–1732.

Buscombe, D., 2015a, pyDGS—Wavelet-based digital grain size analysis: GitHub, accessed December 2, 2015, at https://github.com/dbuscombe-usgs/pyDGS.
Buscombe, D., 2015b, pyDGS 3.0.0—Wavelet-based digital grain size analysis: Python, accessed December 2, 2015, at https://pypi.python.org/pypi/pyDGS/3.0.0.
Commission for Environmental Cooperation, 1997, Ecological regions of North America—Toward a common perspective: Commission for Environmental Cooperation, accessed March 25, 2016, at ftp://ftp.epa.gov/wed/ecoregions/cec_na/CEC_NAeco.pdf.
Community for Data Integration, 2015, CDI Projects Fiscal Year 2015: U.S. Geological Survey, accessed March 29, 2016, at https://www.sciencebase.gov/catalog/item/550c9621e4b02e76d759d7a2.
DGS-Online—Grain size calculator from images of sediment, 2015, accessed March 25, 2016, at https://www.digitalgrainsize.org/
Ember-CLI, [n.d.], Ember-CLI version 1.13.8: Ember-CLI, accessed October 29, 2015, at http://www.ember-cli.com/.
Environmental Information eXchange Network, 2016, Water Quality Exchange—WQX: Environmental Information eXchange Network, accessed February 29, 2016, at http://www.exchangenetwork.net/data-exchange/wqx/.
Esri, [n.d.], Linked data portal for ArcGIS Online: Esri, accessed October 29, 2015, at http://maps.esri.com/SP_DEMOS/linkeddata/.
Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, and Zolly, L.S., 2013, The United States Geological Survey science data lifecycle model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., accessed November 23, 2015, at http://dx.doi.org/10.3133/ofr20131265.
Federal crowdsourcing and citizen science toolkit, 2015, accessed November 23, 2015, at https://crowdsourcing-toolkit.sites.usa.gov/.
Fox, P., and McGuinness, D.L., 2008, TWC semantic Web methodology: Rensselaer Polytechnic Institute, accessed November 25, 2015, at http://tw.rpi.edu/web/doc/TWC_SemanticWebMethodology.
Gacke, C., 2015a, USGS Unmanned aircraft systems data management plan: U.S. Geological Survey, accessed November 24, 2015, at https://www.sciencebase.gov/catalog/item/5524509be4b027f0aee3d3f4.
Gacke, C., 2015b, USGS Emergency Operations Hazards Data and Distribution System Product Specifications and Service: U.S. Geological Survey, accessed January 14, 2016, at http://hdds.usgs.gov/sites/default/files/DigitalLibrary/HDDS_Product_Specification_and_Services.pdf.
Gellman, R., 2015, Crowdsourcing, citizen science, and the law—Legal issues affecting Federal agencies: Woodrow Wilson International Center for Scholars, accessed November 23, 2015, at https://www.wilsoncenter.org/publication/crowdsourcing-citizen-science-and-the-law-legal-issues-affecting-federal-agencies.
Graf, W.L., 1999, Dam nation—A geographic census of American dams and their large-scale hydrologic impacts: Water Resources Research, v. 35, no. 4, p. 1305–1311.
Graves, T.A, Belt, J., and Boyd, K., 2015, Monitoring responses of bear foods to climate change through citizen science, *in* Montana Chapter of The Wildlife Society Annual Conference, 53rd, Helena, Mont., February 24–27, 2015, Proceedings: The Wildlife Society, Montana Chapter, p. 39, accessed November 23, 2015, at http://wildlife.org/wp-content/uploads/2015/05/2015-conference-program.pdf.

Graves, T.A., Long J., Kern, T., and Weltzin, J., 2015, ScienceCache—Engaging citizen scientists in data collection through geocaching: U.S. Geological Survey, accessed March 29, 2016, at https://my.usgs.gov/confluence/x/oIzyHw.

Holdren, J.P., 2015, Addressing societal and scientific challenges through citizen science and crowdsourcing: Office of Science and Technology Policy, accessed November 23, 2015, at https://www.whitehouse.gov/sites/default/files/microsites/ostp/holdren_citizen_science_memo_092915_0.pdf.

Katz, Y., Dale, T., Glowacki, T., Bryn, E., Selden, K., Penner, S., Silber, L., Matchneer, A., Jackson, R., Terzic, I., Beale, M., Faulkner, E., and Munoz, M., 2015, A framework for creating ambitious Web applications: Ember, accessed October 28, 2015, at http://emberjs.com/.

Kern, T., 2015a, Mobile application release checklist: U.S. Geological Survey, accessed November 23, 2015, at https://my.usgs.gov/confluence/x/JInQHw.

Kern, T., 2015b, Software project checklist: U.S. Geological Survey, accessed November 23, 2015, at https://my.usgs.gov/confluence/x/jQAAI.

National Water Quality Monitoring Council, [n.d.], Water quality portal: U.S. Geological Survey and Environmental Protection Agency, accessed February 29, 2016, at http://waterqualitydata.us/.

Niu, Feng, Zhang, Ce, Re, Christopher, and Shavlik, J.W., 2012, DeepDive—Web-scale knowledge-base construction using statistical learning and inference, *in* Brambilla, Marco, Ceri, Stefano, Furche, and Georg Gottlob, eds., Proceedings of the Second International Workshop on Searching and Integrating New Web Data Sources, Istanbul, Turkey, August 31, 2012: Very Large Data Bases Endowment, p. 25–28.

Norkin, T., 2015, 2015 Data management webinar series: U.S. Geological Survey, accessed November 23, 2015, at https://my.usgs.gov/confluence/x/4gb2Hg.

O'Connor, J.E., Duda, J.J., Grant, G.E., 2015, 1000 dams down and counting: Science, v. 348, no. 6234, p. 496–497.

Office of Science and Technology Policy, 2015, The open government partnership—Third open government national action plan for the United States of America: Office of Science and Technology Policy, p. 12, accessed November 23, 2015, at https://www.whitehouse.gov/sites/default/files/microsites/ostp/final_us_open_government_national_action_plan_3_0.pdf.

Otto, M.D., and Thornton, J., [n.d.], Bootstrap version 2.3.2: Bootstrap, accessed October 29, 2015, at http://getbootstrap.com/2.3.2/index.html.

U.S. Army Corps of Engineers, 2015, National inventory of dams: U.S. Army Corps of Engineers accessed on November 4, 2015 at http://nid.usace.army.mil.

U.S. Geological Survey, 2015a, U.S. Geological Survey Community for Data Integration (CDI) science support framework (SSF): U.S. Geological Survey, 3 p., accessed October 27, 2015, at http://www.usgs.gov/cdi/cdi-ssf/cdi-ssf-components.pdf.

U.S. Geological Survey, 2015b, CDI IdeaLab: U.S. Geological Survey, accessed November 23, 2015, at https://usgs1.uservoice.com/forums/222985-cdi-ideas-for-fy16.

U.S. Geological Survey, 2015c, Scientific data management foundation—IM OSQI 2015-01: U.S. Geological Survey, accessed November 23, 2015, at http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-01.html.

U.S. Geological Survey, 2015d, Metadata for scientific data, software, and other information products—IM OSQI 2015-02: U.S. Geological Survey, accessed November 23, 2015, at http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-02.html.
U.S. Geological Survey, 2015e, Review and approval of scientific data for release—IM OSQI 2015-03: U.S. Geological Survey, accessed November 23, 2015, at http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-03.html.
U.S. Geological Survey, 2015f, Preservation requirements for digital scientific data—IM OSQI 2015-04: U.S. Geological Survey, accessed November 23, 2015, at http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-04.html.
U.S. Geological Survey, 2015g, Data management Web site: U.S. Geological Survey, accessed November 23, 2015, at http://www.usgs.gov/datamanagement.
U.S. Geological Survey, 2015h, USGS trends photo explorer: U.S. Geological Survey, accessed October 9, 2015, at http://landcovertrends.usgs.gov/fieldphotomap/.
U.S. Geological Survey, 2015i, USGS Earth Explorer: U.S. Geological Survey, accessed October 29, 2015, at http://earthexplorer.usgs.gov/.
U.S. Geological Survey, 2015j, DRIP: U.S. Geological Survey, accessed December 17, 2015, at https://www.sciencebase.gov/drip/.
U.S. Geological Survey, 2015k, GRAN: U.S. Geological Survey, accessed October 9, 2015, at http://owi.usgs.gov/R/gran.html.
U.S. Geological Survey, 2015l, Hazards Data Distribution System (HDDS) explorer: U.S. Geological Survey, accessed December 2, 2015, at http://hddsexplorer.usgs.gov/.
Viger, R., 2015, Earth-science themes working group: U.S. Geological Survey, accessed March 25, 2016, at https://my.usgs.gov/confluence/x/AgHOAw.
Wesenberg, K.E., and Allwardt, A.O., 2015, Using the ScienceBase Footprinter to geotag publications warehouse records: U.S. Geological Survey, accessed March 29, 2016, at https://www.sciencebase.gov/catalog/item/56faad72e4b0a6037df129a3.
Winslow, L., 2015, sbtools: GitHub, accessed October 9, 2015, at https://github.com/USGS-R/sbtools.
Zednik, S., Fox, P., Fu, L., West, P., and Ma, X., 2014, Linked Vocabulary API for the Earth Sciences Community: Rensselaer Polytechnic Institute, accessed October 9, 2015, at http://tw.rpi.edu/web/doc/IN32A-07.

ISSN 2331-1258 (online)
http://dx.doi.org/10.3133/ofr20161165.