USGS
science for a changing world

Community for Data Integration 2013 Annual Report

By Michelle Y. Chang, Jennifer Carlino, Christopher Barnes, David L. Blodgett, Andy Bock, Anthony L. Everette, Gregory L. Fernette, Lorraine E. Flint, Janice Gordon, David L. Govoni, Lauren E. Hay, Heather S. Henkel, Megan K. Hines, Sally L. Holl, Collin Homer, Vivian B. Hutchison, Drew A. Ignizio, Tim Kern, Frances L. Lightsom, Steven L. Markstrom, Michael O'Donnell, Jacquelyn L. Schei, Lorna A. Schmid, Kathryn M. Schoephoester, Peter N. Schweitzer, Susan K. Skagen, Daniel J. Sullivan, Colin Talbert, and Meredith Pavlick Warren

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, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015

Contents

Figures

Tables

Acronyms and Abbreviations

AESIR	Applied Earth Systems Informatics Research
BCM	Great Basin Characterization Model
BCO-DMO	Biological and Chemical Oceanography Data Management Office
CCC	California Climate Commons
CDI	Community for Data Integration
CONUS	Conterminous United States
CSDGM	Content Standard for Digital Geospatial Metadata
CSV	Comma Separated Value
CSWG	Citizen Science Working Group
DMWG	Data Management Working Group
DOIs	Digital Object Identifiers
DOI	Department of Interior
EPA	U.S. Environmental Protection Agency
FGDC	Federal Geographic Data Committee
FORT	Fort Collins Science Center
FSP	USGS Fundamental Science Practices
GDP	Geo Data Portal
GIS	Geographic Information Systems
IMLGS	Index to Marine and Lacustrine Geological Samples
LEO	Local Environmental Observer Network
MODIS	Moderate Resolution Imaging Spectroradiometer
MRP	USGS Mineral Resource Program
MWBM	Monthly Water Balance Model
NEMI	National Environmental Methods Index
NGP	National Geospatial Program
NLCD EVA	National Land Cover Database: Evaluation, Visualization, and Analysis
NLCD	National Land Cover Database
NOAA	National Oceanic and Atmospheric Administration
NSF	U.S. National Science Foundation
OSTP	Office of Science and Technology Policy
PPSR	Public Participation in Scientific Research
RFP	Request for Proposals
SDCN	Science Data Coordinator Network
SSF	Science Support Framework
SWWG	Semantic Web Working Group
TSWG	Tech Stack Working Group
USGS	U.S. Geological Survey
USMIN	Mineral Deposit Database Project
XLSX	Mircosoft Excel File Format

Community for Data Integration 2013 Annual Report

By Michelle Y. Chang, Jennifer Carlino, Christopher Barnes, David L. Blodgett, Andy Bock, Anthony L. Everette, Gregory L. Fernette, Lorraine E. Flint, Janice Gordon, David L. Govoni, Lauren E. Hay, Heather S. Henkel, Megan K. Hines, Sally L. Holl, Collin Homer, Vivian B. Hutchison, Drew A. Ignizio, Tim Kern, Frances L. Lightsom, Steven L. Markstrom, Michael O'Donnell, Jacquelyn L. Schei, Lorna A. Schmid, Kathryn M. Schoephoester, Peter N. Schweitzer, Susan K. Skagen, Daniel J. Sullivan, Colin Talbert, and Meredith Pavlick Warren

Overview

The U.S. Geological Survey (USGS) conducts 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 apparent that many members of the council had an interest in developing data integration solutions and sharing expertise in a less formal grassroots perspective, thus transforming the "Council" into a "Community" for Data Integration (CDI). Today, 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 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 including associated tools and techniques,
- provide a forum for people doing work with data integration to come together to share ideas as well as learn new skills and techniques, and
- grow overall USGS capabilities with data and information by increasing visibility of the work of many people throughout the USGS and the CDI community.

To achieve these goals, the CDI operates within four applied areas: monthly forums, annual workshop/webinar series, working groups, and projects. The monthly forums, also known as the Opportunity/Challenge of the Month, 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 also sponsored annual workshops/webinar series to encourage the exchange of ideas, sharing of activities, presentations of current projects, and networking among members. Stemming from common interests, the working groups are focused 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. The growing support for the activities of the working groups led to the CDI's first formal request for proposals (RFP) process in 2013 to fund projects that produced tangible products. Today the CDI continues to hold an annual RFP that create data management tools and practices, collaboration tools, and training in support of data integration and delivery.

CDI 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. 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. The activities of monitoring, assessment, and research flow through the Science Data Lifecycle (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 so forth). The assets are transformed into information products that increase knowledge and understanding of the Earth's physical and biological systems.

The horizontal elements in the SSF (fig. 1) represent the "what" of the CDI: products and 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 from science project support through the Science Data Lifecycle processes, applications, and ultimately to data and knowledge management. Continue to explore the CDI Science Support Framework at http://www.usgs.gov/cdi.
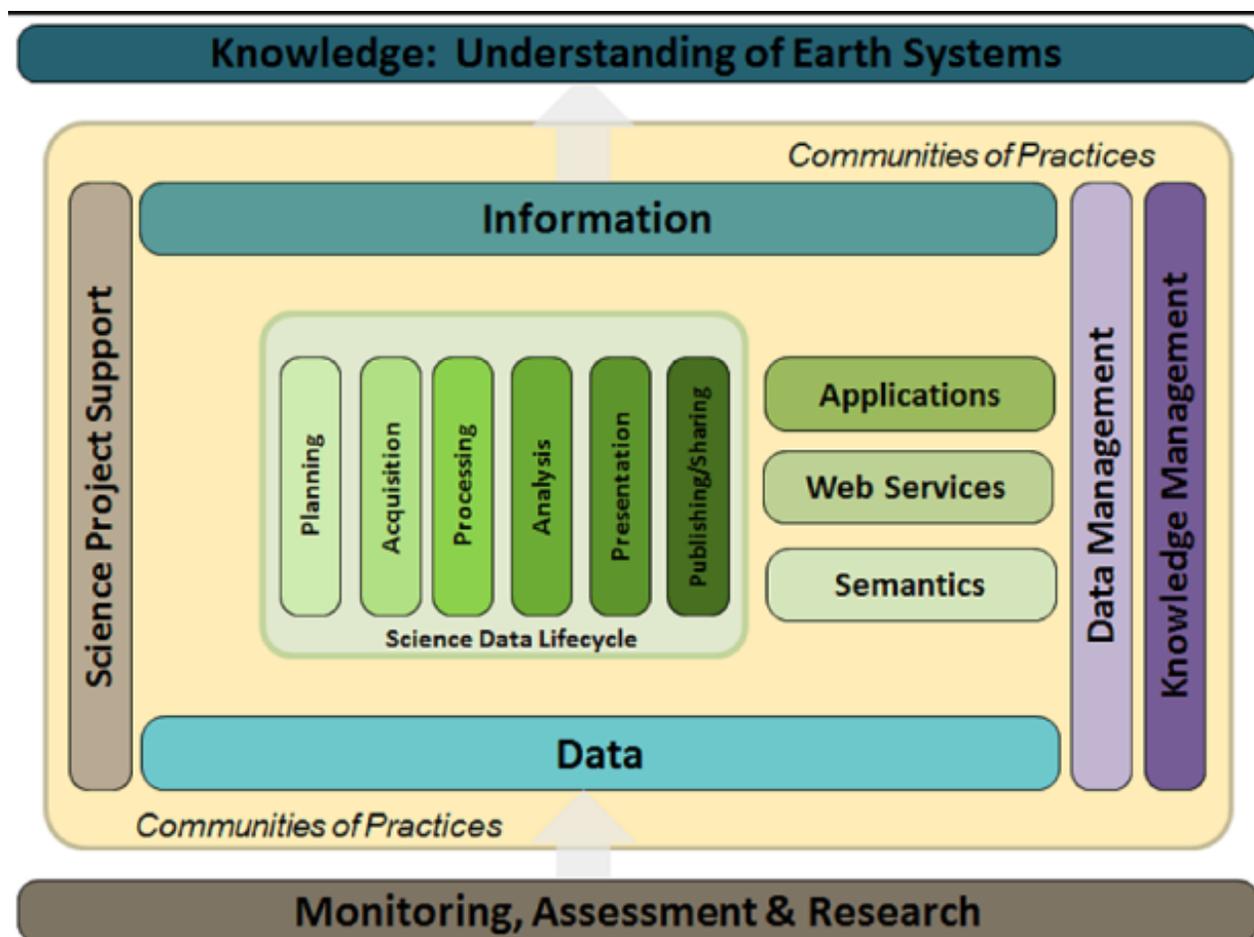
Figure 1. Community for Data Integration Science Support Framework.

Monthly Forums and Annual Workshop/Webinar Series

The CDI monthly forums bring together the community to encourage discussion through presentations from USGS programs and science centers as well as outside organizations to share application development and new technologies, discuss data initiatives, and address data and metadata management challenges. Furthermore, the CDI hosts an annual meeting that brings together scientists, managers, data managers, information technologists, and data users to leverage existing tools in support of mission-critical science and to stimulate future thinking for emerging "big data" and other pertinent questions. Participants share activities and accomplishments through presentations, workshops, papers, poster sessions, and informal gatherings. In recent years, the annual workshop has become a webinar series where information is shared via WebEx and teleconference. These annual meetings and webinars have also become an outlet to share the results and progress of CDI-awarded projects. Together the monthly forums and annual workshop/webinar series provide a key component in the networking and sharing among community members.

CDI Monthly Forums

In fiscal year 2013, the CDI held monthly forums that covered a range of topics from USGS initiatives to the activities of external organizations (table 1).

Table 1. List of 2013 Community for Data (CDI) monthly forums.

Date	Presentation title	Speaker(s)
October 10, 2012	iPlover: A Smartphone Application to Characterize Piping Plover Nest Locations	Rob Thieler, USGS Woods Hole Science Center
November 14, 2012	US Global Change Research Program Interagency Crosscutting Group on Climate Change and Human Health and The Metadata Access Tool for Climate and Health: a new resource for assessing health impacts of climate change	Juli Trtanj and Jacqueline Mize, National Oceanic and Atmospheric Administration
November 28, 2012	The USGS 3D/4D Modeling and Visualization Working Group (a new group associated with CDI)	Linda Jacobsen, National Cooperative Geologic Mapping Program
February 13, 2013	BioData: Overview of the USGS Aquatic Bioassessment Database	Mitchell Harris, USGS Illinois Water Science Center
March 13, 2013	The DataUp Tool: helping researchers document, organize, preserve, and share their scientific data	Carly Strasser, University of California Curation Center at the California Digital Library
April 10, 2013	data.gov 2.0 and CKAN for Geospatial Resource Management	Doug Nebert, Federal Geographic Data Committee
May 8, 2013	You could only find it if you knew it was there: Opening the USGS Publications Warehouse to Google and the world	Jim Kreft, USGS Center for Integrated Data Analytics and Jenna Nolt, USGS Libraries
June 12, 2013	USGS CSAS Announces Winners for the App-lifying Earth Science Data Challenge	rOpenSci, California, and Kimberly Sparks, North Carolina
July 10, 2013	The U.S. Fish and Wildlife Service Catalog	Todd Sutherland, U.S. Fish and Wildlife Service
August–November 2013	CDI Webinar Series 2013	See CDI Webinar Series 2013 section

CDI Webinar Series 2013

The CDI collaborated with The National Map Users Conference to hold a concurrent conference in May 2013. However, the 2013 National Map Users Conference and CDI Workshop was postponed indefinitely. In lieu of the postponed meeting, CDI held a Webinar Series, showcasing a variety of data-integration-related activities from the USGS and other organizations, each with a different session theme. A session was held every month from August through November 2013 in place of CDI monthly forums, which are described below.

- Public Participation in Scientific Research (August 14, 2013)—Sponsored and organized by the CDI Citizen Science Working Group, the session discussed models of citizen participation, described successful use cases, and addressed challenges facing citizen science in support of data collection and USGS research. Special guest speakers included Colleen Flanagan from the National Park Service, Sally Holl and Jake Weltzin from USGS, and Greg Newman from Colorado State University.

- Data Management (September 11, 2013)—Sponsored and organized by the CDI Data Management Working Group, the session introduced an exploration into the nature of scientific data management, why it is important, and how the Science Data Lifecycle Model can be applied to existing projects, planning, and data-integration efforts. Presenters from USGS included Heather Henkel, Viv Hutchison, John Faundeen, Karen Morgan, Emily Fort, Keith Kirk, Fran Lightsom, Alan Allwardt, Cassandra Ladino, and Rex Sanders.
- Earth System Informatics—The session due to occur in October 8, 2014 was postponed to February 12, 2014. The session addressed the Open Data Initiatives and the USGS response in a series of talks from representatives including Nancy Sternberg from the Office of Administration and Enterprise Information, Tim Woods and Karen Armstrong from the Office of Communications and Publishing, Carolyn Reid from the Office of Science Quality and Integrity, and Mike Frame from the Core Science Analytics Synthesis and Libraries.
- Mobile Science (November 13, 2013)—A session focused on the development and use of mobile technologies and platforms in the USGS and other organizations' research programs. Presenters Lorna Schmid, Tim Kern, and Mark Hamill from the USGS, as well as Alyssa Rosemartin and Lee Marsh of USA National Phenology Network, provided demonstrations of the innovative integration of data collection, analysis, visualization, and other functions in the mobile environment.

With an average of more than 70 participants attending each session, the CDI 2013 Webinar Series provided an opportunity for increasing awareness of ongoing activities, enhancing communications across the Bureau, and promoting collaboration opportunities across the USGS and the broader science community.

Working Groups and Focus Groups

As part of networking and sharing knowledge and experience, the CDI has expanded its community approach through working groups that form around common interests, help address challenges, and identify solutions that enable data integration efforts. Working groups have one or more leads, and membership is open to all employees, other government agencies, and external organizations. Some working groups have also further subdivided into smaller focus groups that address more specific issues and report to their respective working group (table 2).

Table 2.   Community for Data Integration 2013 working groups and focus groups.

Working groups	Focus groups	Leads
3D/4D Modeling and Visualization		(vacant)
Citizen Science		Dave Govoni, Megan Hines, and Barbara Poore
Data Management	Data Release Use Case Team Digital Object Identifiers Group Data Management Policy Team Science Data Lifecycle Model Team	Heather Henkel and Viv Hutchison Viv Hutchison and Fran Lightsom Viv Hutchison Fran Lightsom, Viv Hutchison, Carolyn Reid, and John Faundeen
Semantic Web		Janice Gordon and Fran Lightsom
Technology Stack		Dave Blodgett (vacant in 2014)
Mobile Applications Development		Tim Kern and Lorna Schmid

3D/4D Modeling and Visualization Working Group

After joining CDI at the end of 2012, the goal of the 3D/4D Modeling and Visualization Working Group is to help USGS stay abreast and even ahead of rapidly evolving multidimensional technologies. The working group also provides a forum for USGS staff to share information on 3D/4D modeling, visualization, research, and integration efforts and to enable the sharing of expensive software licenses. More information is available at 3D/4D Modeling and Visualization Working Group. The group is currently seeking new leadership. If interested, please contact cdi@usgs.gov.

Citizen Science Working Group

The goals of the Citizen Science Working Group (CSWG) are to (1) promote an understanding of the role and potential benefits of citizen science and scientists in the conduct of USGS research; (2) facilitate and enhance connections between the USGS and the citizen science community; (3) provide access to information and tools to support the proper, effective, and creative use of citizen science-derived data in the USGS; and (4) engage the public in USGS and partner science to improve scientific literacy.

Citizen Science Working Group Accomplishments

The CSWG finalized and published USGS Open-File Report, "Partnering for Science: Proceedings of the U.S. Geological Survey Workshop on Citizen Science" (Hines and others, 2013). The workshop, held in September 2012, brought together 50 attendees representing a diverse array of USGS science, technology, and program staff as well as partners, practitioners, and developers from academia, nongovernmental organizations, and State and local government engaged in various aspects of citizen science. This activity supported the community building and innovation goals of the CDI by (1) raising awareness of the breadth and success of current citizen science activities within the USGS; (2) identifying potential opportunities for incorporating citizen science into the broader fabric of USGS research; (3) highlighting scientific, institutional, and other challenges relating to the effective and

reliable engagement of citizen scientists; and (4) sharing ideas and best practices within USGS and the wider citizen science community.

The CSWG sponsored and organized several meetings and presentations to foster ongoing knowledge exchange and community building efforts in citizen science:

- Local Environmental Observer (LEO) Network by Michael Brubaker, Alaska Native Tribal Health Consortium, Center for Climate and Health, and Eric Wood, USGS Tribal Liaison for the Climate and Land Use Change Mission Area. LEO is a citizen science project that relies on tribal environmental professionals within native Alaskan communities who apply traditional knowledge, western science, and technology to document unusual plants and wildlife, extreme weather, erosion, flooding, droughts, wildfire, and other events that can threaten food security, water security, and community health.
- Citizen Science and the Management of Natural Resources and Environments: A New Approach by Pierre Glynn, USGS National Research Program. A systems approach to citizen science was described for the improved management of natural resources and environments. The approach seeks a much greater level of engagement, not only from volunteers but also from paid professionals, than is present in existing citizen science activities. The approach provides new opportunities for the lay public but also solicits community and policy interactions by scientific experts and stakeholders.
- Can public participation in scientific research (PPSR) better connect natural resource science with decision making? by Duncan C. McKinley, U.S. Forest Service. The potential use of highly participatory forms of PPSR by Federal land management agencies to broaden the reach of science and its benefits to society by creating mechanisms that engage the public, scientists, and decision makers through cogeneration of scientific knowledge, potentially bridging the divide between knowledge creation and implementation/public policy, was explored.

Citizen Science Working Group Benefits

The benefits of a Citizen Science Working Group include

- increased exposure and recognition for USGS citizen science-enabled projects both within the USGS and in the larger Federal science community;
- identification of opportunities and methods to expand the data collection reach of the USGS by harnessing the enthusiastic support of citizen scientists;
- support for the Administration's goal of "increasing open innovation by encouraging expanded use of challenges, incentive prizes, citizen science, and crowdsourcing to harness American ingenuity," as articulated in the Second Open Government National Action Plan for the United States of America;
- USGS science is more approachable and understandable to students, teachers, and the general public; and
- expanded science knowledge and scientific literacy among citizen participants.

Working Group Highlight: Interagency Community Building

CSWG members Megan Hines, Dave Govoni, Sally Holl, and others established a working relationship with the Woodrow Wilson International Center (Wilson Center) for Scholars' Science and Technology Innovation Program Commons Lab to share information and best practices relating to USGS citizen science and crowdsourcing projects. The working group participated in the initial phase of the development of a planned Wilson Center-sponsored Federal citizen science projects inventory,

including general scoping, content and services definition, database design, metadata crosswalks, and identification of opportunities for interoperability with the USGS Citizen Science Project Inventory underpinning the USGS myScience—USGS Citizen Science Project Discovery and Public Engagement Web Application.

Beginning in 2013, the Wilson Center Commons Lab began formal support for the creation of an open Federal crowdsourcing/citizen science community of practice called the Federal Community of Practice on Crowdsourcing and Citizen Science. Current membership includes representatives from Department of Interior (DOI), U.S. Geological Survey (USGS), U.S. Fish and Wildlife Service (USFWS), U.S. Environmental Protection Agency (EPA), Federal Communications Commission, General Services Administration (GSA) (Data.gov), Intelligence Advanced Research Projects Activity, National Aeronautics and Space Administration (NASA), National Institutes of Health (NIH), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), Smithsonian Institution (SI), United States Global Change Research Program (US GCRP), National Archives and Records Administration (NARA), and U.S. Forest Service (USFS).

Dave Govoni serves as a member of the joint Wilson Center/Federal Agency Planning Team established in fiscal year 2013, which is tasked with (1) overseeing establishment of the governance, technical, and knowledge management framework for the community and (2) identification of the team's initial priorities and activities. Dave Govoni also served as a member of the planning team responsible for the Wilson Center-sponsored New Visions in Citizen Science meeting and workshop, held November 20, 2013. The event served as the kickoff for the Federal Community of Practice and featured 17 case studies covering a range of Federally sponsored citizen science and open innovation projects, from in-the-field data collection to online games for collective problem solving. Included in the case studies were two by the USGS—"Did You Feel It?" and "Twitter Earthquake Detection"— and the USA Phenology Network's Nature's Notebook.

Data Management Working Group

The Data Management Working Group (DMWG) was established in 2010 to identify mechanisms for incorporating data management into USGS science activities and develop ways to train scientists about the value of data management. Data management has a clear connection with the CDI Science Support Framework, in part because the "data management" element cross cuts several other elements such as the Science Data Lifecycle, which follows information from initial planning through preservation and sharing and illustrates how data management activities relate to project workflows, assist with understanding the expectations of proper data management, and help to identify how policies can be implemented. Since its inception, the DMWG has continued to grow each year and now has a membership of approximately 100 people from across the USGS representing each mission area and region.

Data Management Working Group Accomplishments

The 2013 year saw the development of new data-management-related teams. A small group of members has contributed to creating new or updating existing Survey Manual chapters that introduce or reemphasize the policy around data management for the USGS. Additionally, the "Data Release Use Case Team" continued to meet weekly, developing workflows for approval and release of data via publications and on the Web. Finally, a Digital Object Identifiers (DOIs) Team arose to develop a "Frequently Asked Questions" document concerning the use of DOIs in the USGS and the potential needs and requirements associated with acquiring DOIs.

The Science Data Lifecycle subteam started in 2011 to develop a Scientific Lifecycle Model that accurately captures how USGS scientists conduct their projects, reflecting the stages that science data go through from initial selection through collection, preparation, use, dissemination, and final disposition. The team designed a visual representation of the lifecycle to articulate its key components, relations, and workflow. In 2013 the team published "USGS Science Data Lifecycle Model" (Faundeen and others, 2013).

The DMWG sponsored and organized several meetings and presentations, listed below, to foster ongoing knowledge exchange and community building efforts in data management:

- Science Data Coordinator Network (SDCN) Task Report by Heather Henkel, SDCN Chair
- Silent Data Corruption (aka Bit Rot) by Rex Sanders, USGS Pacific Coastal and Marine Science Center. Rex Sanders presented "What is silent data corruption (aka Bit Rot), why should you care, and what can be done about it?" Also results were described from studies by Google, CERN (European Council for Nuclear Research), Sun/Oracle, and others who handle many petabytes of data, as well as some personal experiences on a much smaller scale.
- Midwest Regional Efforts: Datageddon by Annie Stoehr, USGS Great Lakes Science Center. An overview of the Science Information Management project commissioned by the USGS Midwest Region to inventory and assess the science data management practices.
- Increasing Access to the Results of Federally Funded Scientific Research by Dave Govoni, USGS Office of Science Quality and Integrity. An overview and discussion of the memorandum from the Office of Science and Technology Policy (OSTP).
- Geology Lab Portal by Jenifer Bracewell, USGS Energy and Minerals Mission Area. An overview of the Geology Lab Portal, which was developed to provide information on geology laboratory facilities, equipment, and resources.
- Open Data Policy Managing Information as an Asset by Dave Govoni, USGS Office of Science Quality and Integrity. A continued discussion of the Open Data policies from OSTP.
- Data Release Use Case Team by Viv Hutchison, USGS Core Science Analytics Synthesis and Libraries. An update on the activities and progress of the DMWG Data Release Use Case Team.
- USGS Data Management Assessment and USGS Science Data Catalog by Lisa Zolly and Mike Frame, USGS Core Science Analytics Synthesis and Libraries. An overview of the Data Management Assessment effort to gather information across the Bureau about data management practices, perspectives, and areas for improvement. The USGS Science Data Catalog is a new project created in response to the recent Open Data policies and requirements from Office of Management and Budget and OSTP.

Data Management Working Group Benefits

The benefits of a Data Management Working Group include
- participation in 2014 CDI proposal process development, which helps to ensure future existence of CDI and the DMWG;
- members share their expertise and participated in several high-visibility CDI working groups including Data Release Use Case Team, Data Management Policy Team, Digital Object Identifiers Focus Group, and the Science Data Lifecycle Model Team;
- communication and collaboration with a variety of teams and projects across the USGS, including Coastal and Marine Spatial Planning, Midwest Region, Science Data Catalog, Science Data Coordinator Network; and

- team members have kept the DMWG updated on past, current, and future activities related to the Open Data Policy; without this, many members may not have heard about this wide-reaching effort.

Working Group Highlight: Building Web Resources for Data Management

The DMWG released the USGS Data Management Web site in November 2012. The Web site continues to be well-received and has shown increased usage since its launch. The USGS Data Management Web site tied as a winner for the 2013 USGS Shoemaker Communication Awards under the External: Internet Category.

> *"In a nutshell, the DM Web site is great and I direct USGS staff to it*
>
> *constantly!!! I think it fits perfectly with where most Centers are in terms of*
>
> *information management."*
>
> —Annie Stoer, Great Lakes Science Center

> *"It's not a trivial task to boil a complex topic into easily digestible areas... I am*
>
> *looking forward to sharing this with scientists who are considering embarking*
>
> *on a data management project."*
>
> —Mark Wimer, USGS Patuxent Wildlife Research Center Science Center

Figure 2. USGS Data Management Web site home page.

Semantic Web Working Group

The Semantic Web Working Group (SWWG) is focused on exploring semantic web technologies as possible solutions for integrating and exposing related but disparate datasets.

Semantic Web Working Group Accomplishments

The SWWG sponsored and organized several meetings and presentations, listed below, to foster ongoing knowledge exchange and community building efforts in Semantic Web (2013 SWWG Meetings):

- SemantEco by Josh Dein, USGS; Evan Patton, Rensselaer Polytechnic Institute; and Patrice Seyed, DataOne. Overview of the architecture behind the Semantic Ecology and Environmental Portal developed at Rensselaer Polytechnic Institute.
- What is ISO 25964? by Alan Allwardt, USGS. The working group discussed the ISO 25964 standard for thesauri and discussed the potential uses within USGS. The group also proposed a project to create a registry of Semantic Web work at USGS.
- Vocabulary, SKOS, etc. The working group learned about vocabulary services and activities through a series of short presentations given by working group members. Presentations included
  - Alan Allwardt and Fran Lightsom: Simple Knowledge Organization System (SKOS) conversion of Marine Realms Information Bank controlled vocabularies,
  - Janice Gordon: Making thesauri play nice with each other,
  - Fran Lightsom: Expressing Coastal Marine Ecological Classification Standard as an ontology,
  - Peter Schweitzer: What's happening with the USGS thesaurus,
  - Dalia Varanka: National Hydrography Dataset (NHD) ontology, and
  - Katherine Yee: Finding vocabularies for ocean planning data discovery.
- Linked data and collaboration between R2R and the IMLGS by Bob Arko, Lamont-Doherty Earth Observatory. The mission of the Rolling Deck to Repository (R2R) project is to provide uniform stewardship for routinely collected environmental sensor data from the U.S. academic fleet of research ships. But what if your research requires the seabed samples and cores collected on the same cruises? Those are being cataloged separately in the Index to Marine and Lacustrine Geological Samples (IMLGS). Arko told how he is connecting these two data services using Open Geospatial Consortium (OGC) Web Services, standard vocabularies, a Simple Protocol and Resource Description Framework Query Language (SPARQL) endpoint, and the World Wide Web Consortium (W3C) "Linked Data" approach.
- Vocabulary Management at BCO-DMO by Dicky Allison, Woods Hole Oceanographic Institution. The Biological and Chemical Oceanography Data Management Office (BCO-DMO) is funded to serve the data management requirements of investigators funded NSF Biological and Chemical Oceanography Sections and the Division of Polar Programs Antarctic Sciences Organisms & Ecosystems Program. Some of the 6,500 datasets served by BCO-DMO include location, time, depth, and physical and chemical parameters of the water column. Scientists supply names for the various types of data they are submitting such as "T" (temperature), "lat" (latitude), "time of day," "salinity," and "water depth in meters." These variable names are then matched to our internal standard names, which are in turn mapped, if possible, to the terms used by the SeaDataNet community, served by the Natural Environment Research Council Vocabulary Server at the British Oceanographic Data Centre. BCO-DMO is currently exploring ways to also incorporate or map other controlled vocabularies, including the Darwin Core terms,

into the dataset metadata and our evolving ontology. BCO-DMO makes use of the dataset metadata to determine how best to visualize the data in order to facilitate data discovery and support data reuse. This presentation illustrates the vocabulary mapping effort and shows some examples of data visualization using metadata terms.

Technology Stack Working Group

The Technology Stack Working Group (TSWG) concerns a broadly defined and broadly applicable set of technology (the "stack of technology") that can be used to promote the ability of project-level workers to manage and use Web-based data and processing services that are readily found and used by others. Part of this effort is to promote interoperability between individual services in order to help scientists pursue more integrated science. The working group is currently looking for new leadership. If interested, please contact cdi@usgs.gov.

Technology Stack Working Group Accomplishments

The TSWG sponsored and organized several meetings and presentations, listed below, to foster ongoing knowledge exchange and community building efforts in innovative technology methods.

- National Geospatial Program OGC Interoperability Assessment by Glenn Guempel, USGS,
- Future National Map staged downloads and services, outcomes of moving to The Cloud by Rob Dollison, USGS,
- Transitioning a VisTrails based habitat model from local data to Web services by Colin Talbert, USGS,
- Real-time data analysis and rendering with HTML5 Canvas by Tom Kunicki, USGS,
- Automated feature-based simulations by Jordan Read, USGS, and
- IPython & IPython Notebook by Roland Viger, USGS.

Mobile Applications Development Working Group

The purpose of Mobile Applications Development Working Group is to establish and promote best practices regarding mobile applications development within USGS. The group's goals are to (1) focus on being a learning environment and a forum for members to exchange ideas, promote coding and design best practices, and discuss technical and design issues; (2) facilitate collaboration among different groups and try to promote joint projects; and (3) focus on working with vendors to gain insight into new technologies being developed.

Mobile Applications Development Accomplishments

Starting in 2012 the working group began designing and implementing a mobile applications development framework to provide a life-cycle management approach to the development of application products and make available development resources and best practices. The working group participated in several coordination activities within USGS and the Department of Interior to foster enterprise-level support for mobile applications. For example the Mobile Framework Team participated in a responsive design coordination effort that got staff involved in the GSA Mobile Government project, arranged webinars for Web site developers on best practices for Content Management System (CMS) responsive design, and released USGS-compliant Drupal responsive themes and support modules. Furthermore, the team acquired development and release licenses for Apple/iOS and Android stores. The initial efforts of the Mobile Applications Development Group have gained considerable momentum and interest at the Bureau and Department level. In 2013, the USGS Administration and Enterprise Information assumed

the lead role on core Web site and application deployment support allowing the group to refocus attention on the next generation of mobile efforts.

Working Group Highlight: A Transition from Mobile Applications to Connected Devices

Core phone and tablet mobile applications development work has morphed from native to hybrid approaches. In addition, Enterprise Information and the Web Reengineering Team now provide support for these efforts. There is less need for a CDI group to specifically address these topics. Instead, technology advances with wearable computing devices, cellular-based tracking devices, and other mobile field-science support tools provide CDI with a fresh opportunity to support a new generation of technologists and scientists. In fiscal year 2013, members of the Mobile Apps Development Group proposed to start the Connected Devices Working Group to supplant the legacy CDI Mobile Applications Development Group. This group would explore the use of cutting-edge mobile tools and frameworks to support field science. New leads will be Tim Kern, Lance Everette, and Bob Waltermire.

Annual CDI Request for Proposals

For the first time, a formal Request for Proposals (RFP) was announced on September 5, 2012, to the broader USGS community for projects that support one or more of the Science Support Framework (SSF) categories:
- Category 1—Management, Policy, and Standards
- Category 2—Computational Tools and Services
- Category 3—Data and Information Assets
- Category 4—Community Innovation

The 2013 RFP projects and products are available at http://www.usgs.gov/cdi/products-publications.html.

Proposal Submissions

A total of 43 proposals were submitted under the four SSF categories. In order to reduce the time commitment of the reviewers, categories 2 and 3, which received the highest submissions, were assigned multiple review panels (table 3). In total, seven review panels were established with three members per panel.

Table 3.   Number of submitted proposals and review panels by category.

[SSF, Science Support Framework]

Category number	SSF category name	Number of proposals submitted	Number of review panels
Category 1	Management, Policy, and Standards	6	1
Category 2	Computational Tools and Services	19	3
Category 3	Data and Information Assets	11	2
Category 4	Community Innovation	7	1
	Total:	**43**	**7**

Reviewers

All panel reviewers were USGS Federal employees who volunteered their time. The 21 reviewers encompassed a wide variety of expertise and represented a broad range of Mission Areas, Regions, and Programs. Nearly half of the reviewers were affiliated with the CDI, and the remainder participated out of interest in the topic of data integration.

Review Panel Objectivity

Specific actions were taken to minimize potential conflicts of interest by review panel members. First, the review panels consisted of a mix of CDI members and non-CDI members and were assigned to proposals outside their program or center. Reviewers that were a part of submitted proposals were not allowed to evaluate proposals in that category. Furthermore, reviewers disclosed potential conflicts of interest by signing a conflict of interest form. Lastly, reviewer identity remained anonymous outside their panels to ensure no discourse between review panels would occur.

Review panels worked independently of each other reviewing only their assigned proposals. As a result, every proposal submitted was reviewed by three individuals and collectively by the review panel.

Review Panel Process

Reviewers were asked to document a summary of the strengths and weaknesses of each proposal based on guidance from the RFP Document. Reviewers also scored each proposal based on the required sections as follows:
- Scope (25 points)
- Technical Approach (25 points)
- Project Experience (25 points)
- Commitment to Effort (15 points)
- Budget (5 points)
- Timeline (5 points)

Having read and evaluated the proposals individually, the reviewers then participated in a group review panel meeting to discuss and rank proposals within their category. Reviewers discussed the strengths and weaknesses of each proposal and gave their recommended scores. Scores for each proposal were averaged to obtain an initial score for the proposal.

Based on initial scoring and reviewer discussion, the panel collectively agreed on a final recommendation for each proposal. Three types of recommendations could be made for each proposal:
- proposal has merit and aligns with CDI goals and should be supported by the CDI,
- proposal has merit but should be supported through a program or region, and
- proposal does not meet merit standards and should not be supported.

Proposals recommended by the panel to be awarded funds by the CDI were then ranked in order of first choice, second choice, third choice, and so forth.

Recommendations

The recommendations of the review panel were submitted to CDI Executive Sponsor, Kevin Gallagher (Associate Director of Core Science Systems Mission Area) for final recommendation. On February 13, 2013, Kevin announced funding for the seven projects ranked number 1. During the CDI Webinar Series on August 14, 2013, Kevin Gallagher announced an additional 3 projects that would

also be supported for a total of 10 CDI projects in 2013 (table 4). The next section describes the projects and their accomplishments in more detail.

Table 4. Overview of 2013 CDI projects (in alphabetical order).

Title	Principal investigator(s)	Program
Data Management Workshops for USGS: Let the Culture Change Begin	Viv Hutchison	Core Science Analytics Synthesis
Development of Enhanced Feature Recognition Software for the Extraction of Mine Features from USGS Topographic Maps	Greg Fernette and Peter Schweitzer	Mineral and Environmental Resources Science Center
Establishing the Land Use Land Cover Geo Data Portal and Supporting Data Services	David Blodgett and Meredith Warren	Center for Integrated Data Analytics
Evaluation of Downscaled GCM Output for Current and Future Conditions and Associated Error in Simulated Runoff for CONUS	Andy Bock and Lauren Hay	Colorado Water Science Center, National Research Program
Metadata Wizard: An Easy-to-Use Tool for Creating FGDC-CSDGM Metadata for Geospatial Datasets in ESRI ArcGIS Desktop	Drew Ignizio, Colin Talbert, and Michael O'Donnell	Fort Collins Science Center
Mining USGS Data Landscape	Anthony Everette and Tim Kern	Fort Collins Science Center
myScience: USGS Citizen Science Project Discovery & Public Engagement Web Application	Sally L. Holl and Megan K. Hines	Texas Water Science Center and University of Wisconsin
National Land Cover Database Visualization and Information Tool	Collin Homer and Christopher Barnes	Earth Resources Observation Systems
Networking the California Climate Commons with the USGS GeoData Portal	Lorraine Flint and David Blodgett	California Water Science Center, Center for Integrated Data Analytics
USGS Protocol Library: An Implementation Based on the National Environmental Methods Index	Daniel Sullivan, Jacquelyn Schei, and Kathryn Schoephoester	Wisconsin Water Science Center, Northwest Region

CDI Projects

The following provides details of the 2013 CDI projects listed in table 4 and their accomplishments.

Data Management Workshops for USGS: Let the Culture Change Begin

Purpose and Objectives

USGS data are one of the most valuable assets of the organization. It is critical that we ensure our scientists and staff produce and manage data in such a way that, at the completion of a project, the data continues to be accessible in useable formats and documented so they can be understood and preserved properly for future uses.

The goals of this project include (1) produce three online training modules (fig. 3) that relay the importance of data management, best practices for planning, and guidance for preparing science data to share; (2) target audiences of researchers, data managers, and the public; and (3) format the modules to consist of automated PowerPoint slides, voice-over narration, and video segments featuring scientists talking about lessons learned in the practice of data management.

This project contributes to the CDI Science Support Framework in multiple areas including Data Management, Knowledge Management, and the Scientific Data Lifecycle. It addresses key elements in data management principles, supports knowledge sharing, and promotes the USGS Science Data Lifecycle.
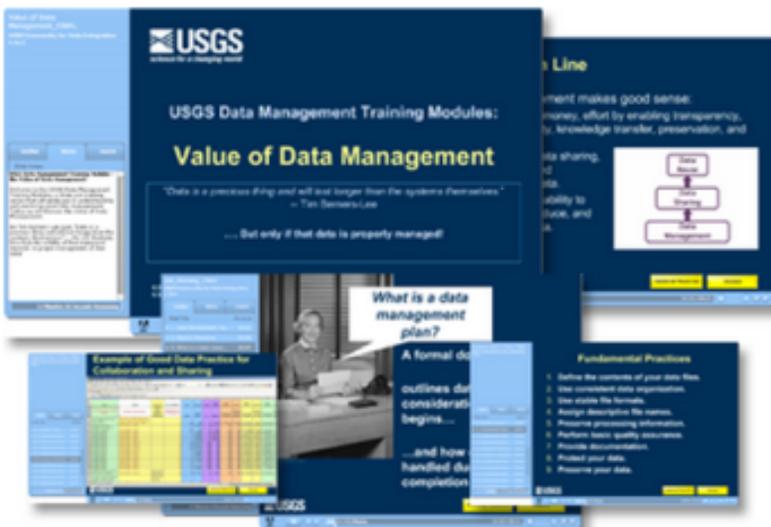
Figure 3.   USGS data management training modules.

The modules were created in cooperation with members from the CDI Data Management Working Group, Bureau of Land Management, California Digital Library, and the Oak Ridge National Laboratory. The modules include video commentary from USGS scientists, links for users to download slides, a glossary of terms used in each module, and additional resources. Special thanks to Jeffrey Morisette of the Department of Interior North Central Climate Science Center and Janice Gordon of the USGS Core Science Analytics Synthesis and Libraries for presenting their user stories. Also thanks to T.J. Lane of the USGS Office of Organizational and Employee Development for development of the training modules and the National Indian Programs Training Center for providing voice talent.

Benefits

- The modules support USGS policies and Fundamental Science Practices (FSP) and provide highlights to key points from the USGS Data Management Web site. Policies tell scientists and managers what needs to be done but not how to do it; the Data Management modules provide one resource for understanding how to implement data management at USGS. Coupled with the Web site, the modules reinforce the principles and allow for alternative ways of learning.

- By posting these three modules to the Data Management Web site, several important benefits were seen: it helped promote a better understanding of data management, it assisted in helping scientists and data managers learn the fundamentals of managing data, and the slides are available at any time.
- The modules leveraged existing and newly created resources on the Data Management Web site and other existing (non-USGS) Data Management Training materials to communicate good data management practices.

> *"I believe that your data management modules are the single best*
>
> *contributions to DOI data stewardship that have been produced thus far. In my*
>
> *view, DOI needs to work from the ground up to change agency culture and*
>
> *make data management a permanent part of the scientific method. Your*
>
> *modules do just that."*
>
> —Douglas Clark, Bureau of Reclamation

Deliverables

- USGS Data Management Training Modules (Hutchison and others, 2013) are available at http://www.usgs.gov/datamanagement/training/modules.php.
  - Module 1: "Value of Data Management"
  - Module 2: "Planning for Data Management"
  - Module 3: "Best Practices for Preparing Science Data to Share"

Development of Enhanced Feature Recognition Software for the Extraction of Mine Features from USGS Topographic Maps

Purpose and Objectives

The Mineral Deposit Database Project (USMIN) is a multiyear project of the USGS Mineral Resource Program (MRP) whose objective is to develop a comprehensive geospatial database of the mines, mineral deposits, and mineral districts of the United States. This database builds upon MRP projects, which date back to the late 1960s, and will provide data that will be of value to other parts of the USGS, other Federal and State agencies, and the general public.

Mine Features, which are a major component of the database, are defined as "a single human-made object or disturbance associated with mining, such as a shaft or adit (vertical or horizontal opening), tailings, machinery and facilities, and so forth." A mine can be comprised of one or more features. Mine features have historical data, including mining and processing methods, and production data associated with them. These features are compiled from a number of data sources of which USGS topographic maps are one of the most important. Capture of these features by manual digitizing is extremely time consuming. In other words, those data are not available for integration and therefore have limited use to the scientific community. The process of manually digitizing those data is, for large

areas, exceptionally time consuming and extremely costly. Moreover, there are historical (but nonintegrable) data that could be used to compare temporally separate, but spatially overlapping features, but the effort required to capture those data is formidable if not prohibitive.

The objective of the CDI supported subproject is to develop an automated feature recognition tool to transform nonintegrable raster-based information—mine feature symbols on digital topographic maps—into data that can be effectively integrated with other spatial data layers.

Benefits

- USGS topographic maps represent a massive store of high-quality, time-stamped geospatial data, which in its present form cannot be easily brought into geographic information systems (GIS). The mine feature recognition software will not only meet a need of the USMIN Project but will also provide a foundation for the extraction of other types of data such as wetlands and springs from raster versions of USGS topographic maps.
- The features extracted using the feature recognition tool will be used in the MRP database project and can also be used by the National Geospatial Program (NGP) for the National Map. The value of this is shown by the fact that NGP has provided half the funding for the project.
- The USMIN database is also of value to other Federal agencies such as the BLM, USFS, and EPA. The Central Mineral and Environmental Science Center has an Interagency Agreement with BLM under which BLM has provided $500,000 in funding to date.
- The value of the mine feature data is also shown by the fact that two State agencies, Nevada Department of Minerals (http://www.nbmg.unr.edu/dox/dox.htm) and California Department of Conservation (http://www.consrv.ca.gov/omr/abandoned_mine_lands/Documents/Statewide%20Abandoned%20Mine%20Features%20Jan-2012.pdf), captured these data by manual digitization.

Deliverables

The deliverable for the project will be the enhanced feature recognition software. The software will be developed by Dr. James Burt of the University of Wisconsin-Madison. Dr. Burt developed the Quad-G software, which was used to georeference scanned USGS topographic maps for the Historic Topographic Map Project.

Establishing the Land-Use Land-Cover Geo Data Portal and Supporting Data Services

Purpose and Objectives

This project sought to incorporate land-use land-cover time-series data into the Geo Data Portal Web service infrastructure. Until now, the Geo Data Portal has been used for downscaled climate data. Data from EPA Integrated Climate and Land Use Scenarios were incorporated into a Web service compatible with the Geo Data Portal and is now hosted in the Geo Data Portal catalog. Great progress was made in establishing connections between the Land Processes-Distributed Active Archive Center's holdings of Moderate Resolution Imaging Spectroradiometer (MODIS) satellite-derived land-cover data. At this time, the data is only available on development servers, but many issues were fixed and close partners are using the integration to access MODIS data at greatly reduced project cost. The Geo Data Portal Web-user interface was reworked to ease discovery and use of the higher resolution land-use data. This was completed and has been used successfully on the production Geo Data Portal (GDP) (http://cida.usgs.gov/gdp/). Services for land-use land-cover projections from a model run at the

Southeast Climate Science Center were prototyped and will also be available in production from the Geo Data Portal soon. This project enhances core aspects of the CDI Science Support Framework; it integrates data from partner agencies into a system that enhances discoverability and accessibility for all users of these important data.

Benefits

- New data delivery mechanisms for land-use land-cover data decreasing the burden on scientists who need these data for their work
- Enhanced discoverability and usability of the Geo Data Portal web interface
- Greatly contributed to the body of research that is the Geo Data Portal and related big-data web service technologies

Deliverables

- Web service access to EPA Integrated Climate and Land Use Scenarios via the THREDDS data server available at http://cida.usgs.gov/thredds/catalog.html
  - A new and improved Geo Data Portal user interface to support the needs of land-use land-cover data available at http://cida.usgs.gov/gdp/

Evaluation of Downscaled General Circulation Model Output for Current and Future Conditions and Associated Error in Simulated Runoff for Conterminous United States

Purpose and Objectives

This project will assess the accuracy of climate drivers (precipitation and temperature) from different sources for current and future conditions. The impact of these drivers on hydrologic response will be using the monthly water balance model (MWBM). The methodology for processing and analysis of these datasets will be automated for new climate datasets when they become available on the USGS Geo Data Portal (http://cida.usgs.gov/climate/gdp/). This will ensure continued relevancy of project results, future opportunities for research and assessment of potential climate change impacts on hydrologic resources, and comparison between generations of climate data. To share and distribute the results with scientists and environmental resource managers, project-related data will be hosted with collaborators using a customized Web portal (http://www.fort.usgs.gov/WebApps/SciBase.asp) developed by USGS ScienceBase (http://www.sciencebase.gov).

Benefits

- The accuracy of climate data made available from the USGS Geo Data Portal, and subsequent MWBM output, for historical and future conditions, will be evaluated on modeling units we refer to as the GeoSpatial Fabric. The GeoSpatial Fabric is a spatial discretization of the conterminous United States based on the USGS National Hydrography Dataset Plus feature set (http://www.horizon-systems.com/nhdplus/).This allows results to be summarized both locally and for specific real-world hydrographic features such as USGS streamgages, major waterbody outlets, and river confluences.
  - The Web portal will provide scientists and resource managers with needed information to choose the appropriate dataset with associated uncertainties for their region of application.

- Deliverables can be customized for specific applications or research targets defined by research partners such as Department of Interior Climate Science Centers and Regional Landscape Cooperatives.

Deliverables

- Data Products and Web Services
    - Climate data is summarized for the GeoSpatial Fabric for CONUS including PRISM, DAYMET, GSD (Maurer), and BCSD for CMIP3 and CMIP5 (>200 GCMs).
- Metrics describing uncertainty associated with climate drivers are currently in progress and will be developed by Powell Center post-doctoral fellow Kirsti Hakala.
- Web portal uses an interactive Web mapping service to display and distribute results.
    - A working prototype has been developed for the DOI South Central Climate Science Center.
    - The USGS Fort Collins Science Center Web Applications Team is currently working on a prototype based for the nation.
- Publications and Presentations
    - Earth Interactions Journal article published (Hay and others, 2014)
    - Journal paper for MWBM parameter regionalization used with CONUS application (in progress)
    - USGS fact sheet for using the Web portal (in progress)
    - CDI monthly meeting presentation on September 10, 2014
    - AGU presentation on MWBM portal scheduled for fall 2014

Metadata Wizard: An Easy-to-Use Tool for Creating FGDC-CSDGM Metadata for Geospatial Datasets in Esri ArcGIS Desktop

Purpose and Objectives

The Metadata Wizard is a fully functional ArcGIS Toolbox that allows for streamlined Federal Geographic Data Committee (FGDC) endorsed Content Standard for Digital Geospatial Metadata (CSDGM) creation and editing for geospatial professionals working in the Esri software environment. The tool overcomes the difficulties of working with metadata that were introduced in ArcCatalog 10.x by properly handling metadata conversion and exportation between the Arc10 and FGDC metadata formats, automating much of the metadata creation process, employing default values whenever possible, and allowing for easy metadata template importation to save time and effort.
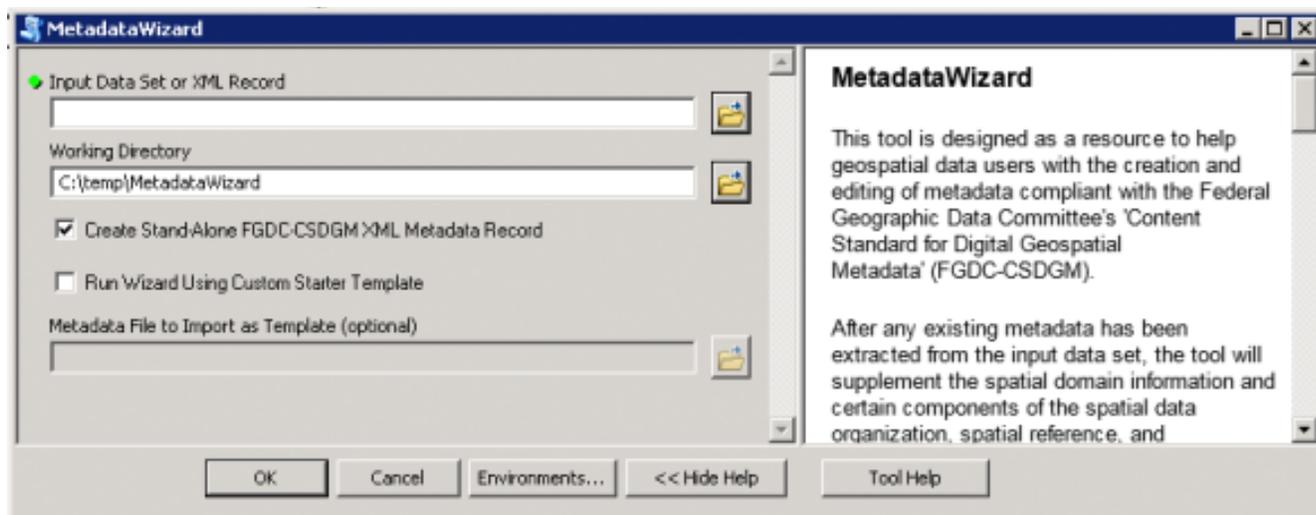
Figure 4. Screenshot of USGS Metadata Wizard Tool.

The Metadata Wizard constitutes a resource for USGS professionals and other researchers in support of the CDI Operational Context. In general, the tool can be considered a Computational Tool or Service, but one that specifically integrates and formalizes the Management, Policy, and Standards focus of the CDI as a way to add value to and better preserve data and information assets. The tool facilitates the preservation of project information and ensures data integrity by leveraging a widely used existing resource (for example, standard Esri software).

Benefits

- The tool is fully functional and complete. The positive feedback we have already received for the Metadata Wizard is encouraging and indicates the extent of the need for this type of tool in the scientific community.
- The bottom line is that easier metadata creation will dramatically improve the ability of scientists to adhere to and integrate the guidelines for data management outlined by the CDI.
- The following are additional comments in support of the Metadata Wizard:

> *"I was just running the metadata wizard v10.2 on a raster dataset. It works*
>
> *like a charm…My previous attempt at metadata years ago (and avoided since)*
>
> *was very painful. So great job! You guys helped conquer my fear of metadata."*
>
> —Carol S. Mladinich, USGS Geosciences and Environmental Change
>
> Science Center

"You guys get me pumped regarding metadata, and at least for a while I am motivated to improve our metadata here...until some big project pushes me off of that and onto something else again. But at least for a period of time metadata finds itself at the top of the stack. Thanks guys!" —Danny Lewis, Texas Parks & Wildlife Department, Texas

"I like the interface a lot! Should be a big help for users who only write an occasional record. Thanks a million for all your work on making a user-friendly tool. The easy install, which does not require a separate windows application install, is a big help for our various users whose organizations may not be very GIS friendly." —Edith Konopka, New Jersey Office of Geographic Information Systems

"I came across the Metadata Wizard tool that you and others at the USGS have worked on and I just want to thank you for sharing it. It truly is amazing and I can see that it will help us very much!" —Neomi Mustain, ManTech International Corporation

"Thanks again for all of your help with this. I can't tell you how much I appreciate it. I think I need all the metadata help I can get!" —Tara Chesley-Preston, USGS Northern Rocky Mountain Science Center

Deliverables

- The tool is complete and freely available. It can be downloaded at http://www.sciencebase.gov/metadatawizard.

- All source code for the project is also freely available in a GitHub repository (https://github.com). Please contact Drew Ignizio at dignizio@usgs.gov for more information.
- March 2013: WebEx presentation of tool to FGDC metadata working group. Presentation and video available at http://www.fgdc.gov/participation/working-groups-subcommittees/mwg/index_html.
- July 2013: WebEx presentation of tool to Office of Fish and Wildlife Information Managers.
- December 2013: WebEx presentation of tool to CDI for 2013 Webinar Series.
- May 2014: USGS Enterprise GIS Brown Bag Presentation, "Metadata Wizard: An Easy-to-Use Tool for Creating FGDC-CSDGM Metadata for Geospatial Datasets in ESRI ArcGIS Desktop."
- August 2014: USGS Publication "Metadata Wizard: An Easy-to-Use Tool for Creating FGDC-CSDGM Metadata for Geospatial Datasets in ESRI ArcGIS Desktop" (Ignizio and others, 2014).

Mining the USGS Data Landscape

Purpose and Objectives

The scientific legacy of the USGS is the data and the scientific knowledge gathered over 130 years of research. However, it is widely assumed, and in some cases known, that high-quality data, particularly legacy data critical for large time-scale analyses such as climate change and habitat change, is hidden away in case files, file cabinets, and hard drives housed in USGS science centers and field stations (both hereafter "science centers"). Many USGS science centers, such as the Fort Collins Science Center, Fort Collins, Colo. (FORT), have long established research histories, are known repositories of datasets, and conduct periodic "file room clean-out" days that establish and enforce some minimal data lifecycle management and maintains a cursory inventory of maintainable data—data that are of high enough interest or impact that they should be maintained at a minimum readable format for future access and use. But many science centers currently lack a clear understanding of data lifecycle management best practices and simple inventory tools to manage their data through its lifecycle. The CDI developed their data management lifecycle framework to provide that clear understanding for managing data throughout their lifecycle. Through this project we are testing and validating the CDI lifecycle framework as well as using those workflows to create a data inventory and management tool that enables USGS science centers to properly apply the CDI data lifecycle framework.

Objective 1. Validate and document the application of the CDI Data Management Lifecycle framework.

Through cooperation with the USGS Core Science Analytics, Synthesis and Libraries Species Occurrence Records and Data Transformation Processes project, we were able to expand the scope of this objective to include additional bat and white-tailed kite data to the project, increasing our sample size for estimating resource requirements from three datasets to nine. Chapter 1 of the project's completion report is in draft form. Details of the progress for each dataset follows.

- The southeastern Arizona riparian bird and habitat data (completed). Data has been quality controlled and documented and is being prepared for submission as a USGS digital data series product.
- Texas, Kansas, Oklahoma, South Dakota, and North Dakota wetlands and shorebird data (processing). The original format for these data requires Quattro Pro 7 software licensing, which we have acquired and converted files to comma separated value (CSV) and Mircosoft Excel (XLSX) format for further processing.

- Eastern Colorado prairie bird and habitat data (processing). The original format for these data requires Quattro Pro 7 software licensing, which we have acquired and converted files to CSV and XLSX format for further processing.
- Bats of the Rocky Mountain Arsenal Mist Net Data (completed). Data and metadata have been reviewed and archived in the Fort Collins Science Center ScienceBase community and is ready for FSP approval.
- Bat Inventory of Ouray National Wildlife Refuge Mist Net Data (completed). Data and metadata have been reviewed and archived in the Fort Collins Science Center Sciencebase community and is ready for FSP approval.
- Bats of Mesa Verde National Park Mist Net Data (processing). Data has been reviewed and metadata is in the process of being completed. The data and metadata have been archived in the Fort Collins Science Center Sciencebase community.
- White-tailed Kite Historic Data (processing). We are working with the original USGS principal investigator and the museums who contributed data to this dataset to verify data agreements and to develop metadata. The data itself has been converted from its original QuatroPro 6 format to XLSX and CSV for final data processing.
- White-tailed Kite Physiological Data (processing). Metadata for this dataset is complete, and the data has been converted from its original QuatroPro 6 format to XLSX and CSV for processing.
- White-tailed Kite Morphological Data (processing). Metadata for this dataset is complete, and the original capture data has been converted from its original QuatroPro format to XLSX and CSV for processing.

Objective 2. Inventory, prioritize, and estimate the cost of integrating the USGS Data Mine.
- Conducted an inventory of the Fort Collins Science Center datasets based on metadata produced from 1994–2011. We have identified 440 potential datasets to date. All inventoried records have been migrated to the Fort Collins Science Center ScienceBase community as part of a Data Mine space, which is restricted to internal FORT data stewards and principal investigators. Once datasets are completely processed and approved for public distribution, those ScienceBase dataset items will be moved to the FORT community's public datasets space for distribution.
- FORT case files were also inventoried where dataset metadata was associated with active FORT staff. Fewer than a dozen additional datasets have been identified through this review of case files; however, this process is ongoing.
- Initial estimates for datasets similar to those completing Objective 1 processing have been assigned. These are updated as dataset processing times are analyzed. The current average is 24 hours per dataset to process.
- Chapter 2 of the project's completion report is in draft form.

Objective 3. Develop a USGS Data Mine Web application.
- Based on our experience inventorying the FORT's metadata and case files, we are currently documenting our suggested inventory workflows and designing wireframes for the Data Mine data management application. These make up chapter 3 of the project completion report.
- We are evaluating other data and project lifecycle management applications being developed within the USGS to partner with; the objective being that the Data Mine application becomes a legacy data portal that unifies and (or) expands upon the strengths of each of those related applications.

Benefits

- Deliver an immediate benefit to solve an existing data integration challenge
- Demonstrate a methodology and (or) solutions architecture that can be replicated for other data or research projects
- Create an environment that allows future innovative applications to access USGS data
- Provide a benefit to scientists
- Promote standards and best practices for data management
- Complement other SSF categories, elements, and CDI projects

Deliverables

- Nine datasets documented and archived
- USGS open-file report (in progress)
- Open-source Data Mine inventory code and database (in progress)

myScience: USGS Citizen Science Project Discovery and Public Engagement Web Application

Purpose and Objectives

In 2012 there was a collaborative investment led by the CDI Citizen Science Working Group to build a database of USGS citizen science project metadata (CDI Citizen Science Working Group, 2012). Investment was concurrently made in research and publications by Applied Earth Systems Informatics Research (AESIR) and the USGS Office of Communications and Publications, respectively. The database leveraged and integrated these efforts to create a citizen science project discovery application called myScience.

The myScience application is built on a robust, interoperable database of project metadata that is updated by project information providers through an improved Web interface. The public can view and search the project information through the myScience project discovery Web interface. Viewable project information in the Citizen Science Project Inventory currently includes the project name and lead USGS personnel, USGS program name, data and protocols, project partners and affiliations, references, and geographic coverage. The database will be strengthened and expanded to include highlights of specific science outcomes and achievements of the projects and additional fields as needed to improve interoperability with project inventories of USGS partners.

The myScience application adds value and dynamic content to existing Federal and USGS education and (or) recreation Web sites. It enables open access to citizen science project metadata in a format that will help engage public participants in these projects, raise awareness about them, and stimulate discussion and partnerships within both the scientific community and the general public.

Providing unified information access through an engaging Web application linked to the USGS Education Web site is envisioned as a collaborative effort among the CDI Citizen Science Working Group, AESIR, USGS Office of Communications and Publications, USGS Office of Education, USGS Texas Water Science Center, USGS Fort Collins Science Center, and University of Wisconsin-Madison. Improving data access and collaboration across disciplines by leveraging partnerships and existing assets is closely aligned with the CDI mission.
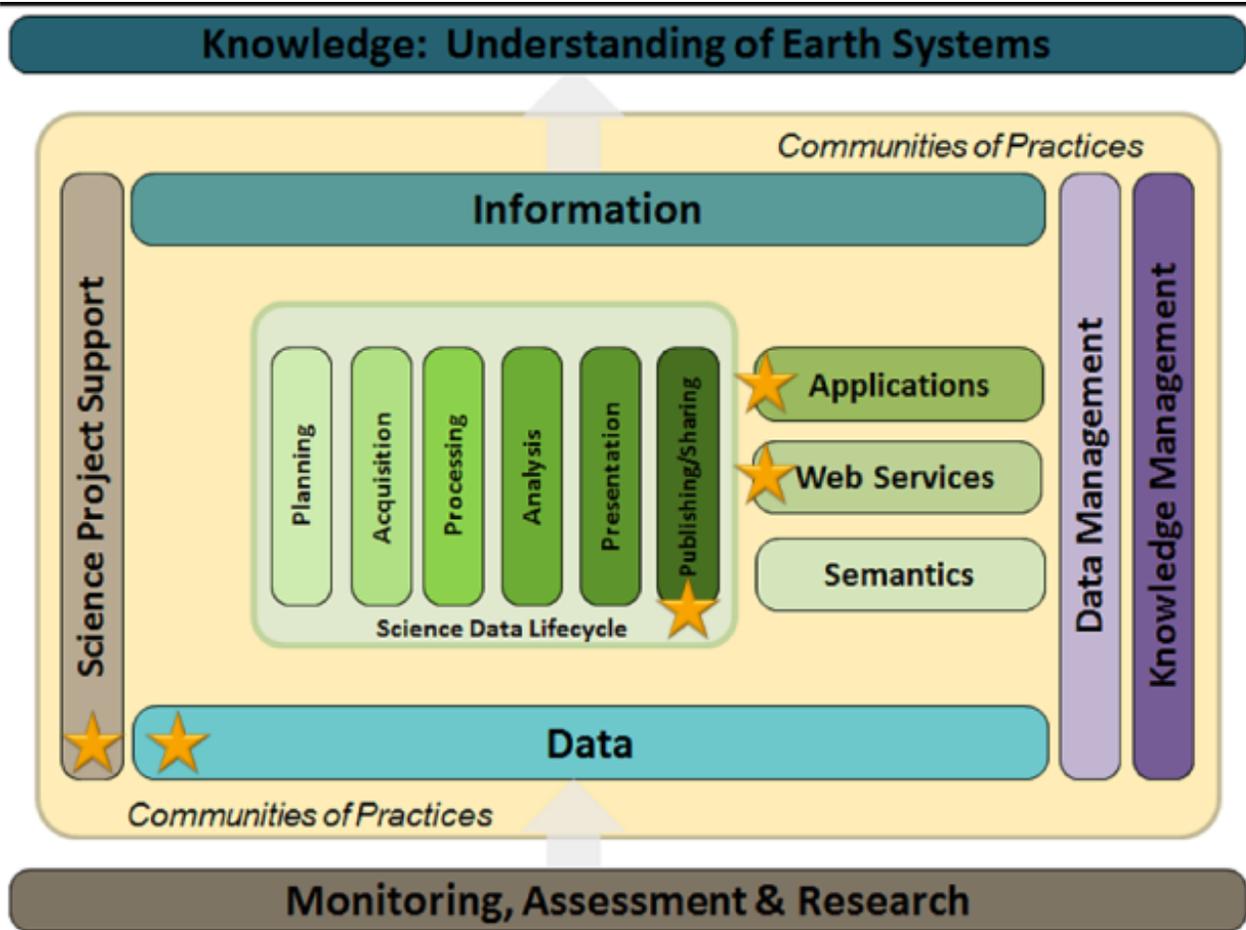
Figure 5. Stars indicate Science Support Framework elements addressed by the myScience project.

The application addresses several elements of the CDI Science Support Framework (fig. 5). Public engagement tools are key components of citizen science efforts and one need identified for USGS projects. As a high-visibility Web product with the potential to increase public engagement through linkage with the USGS Education Web site, the citizen science discovery application will provide overall science project support for USGS citizen science projects. The application Web interface highlights key science outcomes of projects and links to project Web sites and datasets, and in this way, addresses the data sharing component of the scientific data lifecycle within the SSF Data element. The feasibility of sharing the Citizen Science Project Inventory data via Web services will be evaluated as well as addressing the SSF Web services element.

Figure 6. USGS Education featuring myScience Web application.

The goal of this project is to create a unified, publicly accessible Web hub for USGS citizen science project metadata, including project-result highlights when available. The following questions are addressed by a project discovery Web application:

- How may the public discover opportunities for participation in USGS scientific research?
- What citizen science projects are currently active within the USGS?
- How may PIs increase public engagement in, and awareness of, their citizen science projects and outcomes?
- How may scientists have access to information about projects and results within the USGS citizen science community of practice for the purpose of improving communication, creating and strengthening partnerships, improving workflows and resulting science outcomes, and increasing results-sharing?

Benefits

myScience provides transparent, open access to USGS citizen science activities. It meets the need to connect the public with experiential science education opportunities that are proven and highly effective in teaching science process skills (Brotherton and Preece, 1996). For example, the program coordinator at the Boeing Scholars Academy in Chicago can now use myScience to search by zip code, topic of interest, or a variety of other relevant criteria to connect his students with projects. The myScience project inventory records can be accessed at any time to accurately respond to frequent

requests for lists of these activities such as from the USGS Citizen Engagement Officer, Karen Armstrong, USGS Office of Communications and Publishing.

The benefits of this product to USGS science include increasing public engagement in existing citizen science projects; connecting scientists' leading projects with each other; and raising awareness among Bureau scientists about the nature, methods, and results of public participation in USGS scientific research. Linking the project discovery application to the long-standing education.usgs.gov domain has the advantage of leveraging that site's pre-existing and well-established public audience base, resulting in maximum visibility and integration with existing USGS Education resources. The resulting benefit to the USGS Education Web site is enhanced relevance, currency, and visibility.

The benefits to USGS projects include raising awareness about projects that will increase public engagement and lead to better research results. Projects will also generate partnerships and begin dialogues as a result of making project metadata, including contact persons, openly available. For example, a DataONE participant and expert on citizen science cyberinfrastructure suggested that it would be useful if USGS provided access to information about project leads and practices related to citizen participation in research. Transparent access to this information improves communication, creates and strengthens partnerships, improves workflows and results, and increases results-sharing.

There are also benefits to the public citizen science community of practice. For example, the myScience Project Team has shared information and provided support to developers of a Federal Citizen Science Project Inventory at the Wilson Center Commons Lab Science and Technology Innovation. The team has also provided the myScience Inventory Crosswalk product to interagency groups convened by the Wilson Center to collaborate on inventory development and to respond to requests from interested parties such as the North Lakeland Discovery Center in Manitowash Waters, Wisconsin.

> *"I enjoyed your talk, and I am really interested in getting my students involved in Citizen Science in Chicago. I will test out the new portal and check out the projects…"* —Ross Ludwig, Boeing Scholars Academy

> *"Thank you for sharing more information about the database that USGS is developing for their citizen science projects…I found your presentation very interesting and am wondering if you would be willing to share with me the spreadsheet to catalog all of the citizen science project databases..."*
> —Heather Lumpkin, North Lakeland Discovery Center

Deliverables

- Application-related:
  - myScience—Search USGS citizen science projects
  - Citizen Science Internal—View all USGS citizen science projects and people, then enter your project. Contact Sally Holl (sholl@usgs.gov) for more information.
  - Inventory Crosswalk—Excel table linking myScience fields with partner inventory fields in anticipation of future Web Services creation. Contact Sally Holl (sholl@usgs.gov) for more information.
  - Database Schema—PDF document outlining database structure and fields
- Outreach-related:
  - Established Partnerships Letters of Support (see appendix B) with Wilson Center and DOI Office of Youth, Partnerships, and Service
  - Attended DataONE PPSR Working Group meeting in Ithaca, N.Y., May 6–10, 2013
  - CDI Webinar—2013 CDI Webinar Series Session 1: Public Participation in USGS Scientific Research, August 14, 2014
  - Global StemX Education Conference Presentation—Recording (mp4 and mp3 options)
  - myScience featured in the Wilson Center blog, available at http://wilsoncommonslab.org/2014/05/08/video-usgs-citizen-science-efforts-on-the-grand-canyon/
  - CDI fact sheet (in progress)

Access the full report on the myScience CDI Confluence Page, available at myScience Summary Report.

National Land Cover Database Visualization and Information Tool

Purpose and Objectives

The National Land Cover Database (NLCD) serves as the authoritative Landsat-based, 30-meter resolution, land-cover database for the Nation. NLCD supports a wide variety of Federal, State, local, and nongovernmental applications that seek to assess ecosystem status and health, understand the spatial patterns of biodiversity, predict effects of climate change, and develop land management policy. However, access to NLCD products for the USGS community and the public is a concern due to large file sizes, limited download options, and the expectation that users must download and analyze multiple land-cover products in order to answer even basic land-cover-change questions. Therefore, the goal of the NLCD Evaluation, Visualization, and Analysis (NLCD EVA) Tool was to facilitate unprecedented online access to wall-to-wall national land-cover and land-change data for USGS scientists and managers across all mission areas, as well as the public (fig. 7).
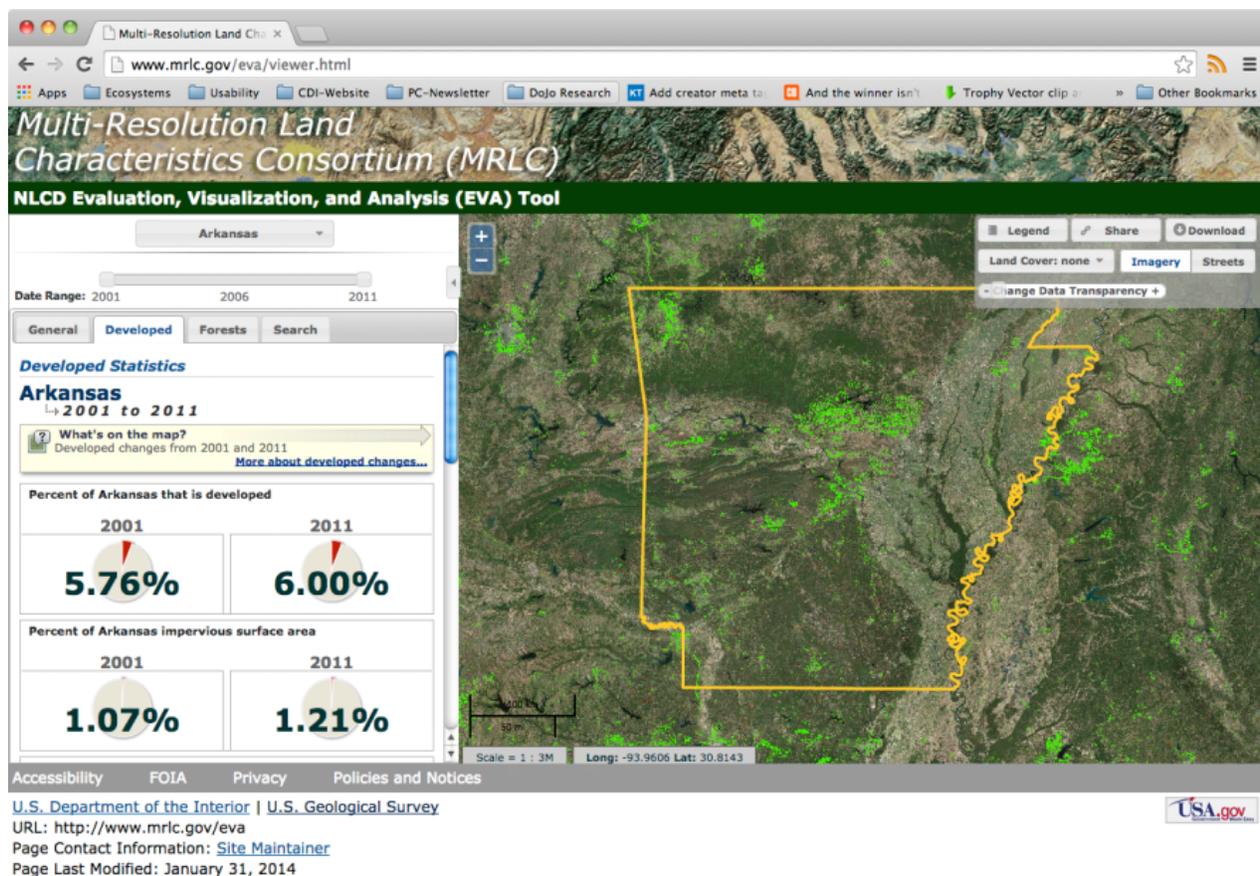
Figure 7. The National Land Cover Database Evaluation, Visualization, and Analysis Tool provides the U.S. Geological Survey community and the public access to a nationally standardized database of land-cover and land-change information across all lower 48 States.

The NLCD EVA Tool provides a solution that transfers land-cover data from the Scientific Data Life Cycle Processes to the Information and Knowledge: Understanding of Earth Systems horizontal element in the CDI Science Support Framework. Furthermore, this solution supports goal three of the USGS Global Change Science Strategy: A Framework for Understanding and Responding to Climate and Land-Use Change, which aims to "improve understanding of land-use and land-cover changes: rates, causes, and consequences." The shifting emphasis of the NLCD program from mapping to a monitoring initiative provides the opportunity for future NLCD data products (such as, 2016 and later) to be added to the tool that will result in continued innovative access to USGS land-cover data.

Success of the NLCD EVA Tool would not have been possible without NOAA agency personnel who provided valuable technical resources learned from their existing Coastal Change Analysis Program Land Cover Atlas viewer.

The NLCD EVA Tool project addressed the Data and Information Elements of the CDI Science Support Framework, which is aimed at creating or improving USGS data and information products. NLCD EVA facilitates unprecedented online access to wall-to-wall national land-cover and land-change data for USGS scientists and managers across all mission areas, as well as the public, enhancing communication and understanding of Earth's physical and biological systems.

Benefits

- Delivers consistent user-friendly access to wall-to-wall NLCD data
- Helps users of all abilities to visually analyze and explore NLCD data
- Allows users to query specific types of land-cover changes between 2001, 2006, and 2011
- Generates summary reports and data tables of land-cover changes to enhance communication and advance understanding of U.S. land changes and trends

Deliverables

- An online tool (http://www.mrlc.gov/eva/) that provides the first ever, spatially explicit, nationally standardized database of land-cover and land-change information, delivering an immediate solution to an existing USGS data integration and access challenge.

Networking the California Climate Commons with the USGS GeoData Portal

Purpose and Objectives

The California Climate Commons (CCC) and USGS Geo Data Portal Teams have collaborated to curate and host California and Great Basin Characterization Model (BCM) results. The CCC has successfully set up a Web-server and installed needed software to serve these model results using data and Web service standards that are compatible with the GDP. All raw monthly data has been transferred to the Geo Data Portal Team for processing and metadata development for hosting on the GDP. The GDP and CCC Teams have made significant progress in converting raw BCM model data to archive formats and are moving forward as planned. The project experienced delays in transferring funds to the Point Blue Conservation Science Team responsible for implementing CCC Web capabilities and in receiving required raw data. Despite these setbacks, the project is on track to deliver Web-service data access and processed summaries via Web or script interfaces. These deliverables fit in the CDI Science Support Framework as a machine-to-machine integration between USGS and a partner organization, providing greatly improved access to raw and summarized version of a valuable dataset.

Benefits

- Improve machine and human access to an important USGS data product
- Demonstrate the ability to integrate a partner organization's Web service data into the USGS Geo Data Portal data-sharing framework
- Establish a partnership of shared data and services between two climate change data systems and user communities

Deliverables

- California Climate Commons data server accessible to the Geo Data Portal and other client software
- California Climate Commons application leveraging Geo Data Portal processing to construct summary visualizations

USGS Protocol Library: An Implementation Based on the National Environmental Methods Index

Purpose and Objectives

A number of monitoring method and protocol libraries are currently in existence. Although these systems have been tailored to certain disciplines or research foci, the underlying principles, mechanisms, and processes have commonalities that could facilitate synthesizing content and information.

The Protocol Library project consists of modifying and thus extending the capabilities of the existing National Environmental Methods Index (NEMI) methods compendium. To accomplish the task of incorporating a broad array of protocols, NEMI developers have conducted user-requirement queries of protocol owners. In addition, protocol owners have been asked how they would prefer to access the data. Input forms have been created to accommodate the desires of protocol owners as much as possible.

This project represents another step toward a common framework for storage and retrieval of methods and protocol information. Use of this existing and proven database will greatly enhance the accessibility of this information and will provide researchers the opportunity to track versions of protocols as they evolve over time.

It was determined after consultation with stakeholders that the most relevant information for scientists was the methods information contained within protocols. Therefore the approach has been to add methods information that is pertinent to the development of the BioDATA database, including USGS-NAWQA and EPA-NRSA methods. The NEMI database will also accept metadata on protocols that will then be available and mated to methods contained within them.

Benefits

- NEMI (https://www.nemi.gov/home/) is already fully functional and available to the public.
- NEMI is built on a data model that was designed to be flexible.
- By leveraging the NEMI database, we are building on a solid, proven, and long-lasting solution and applying it to the storage and retrieval of protocol information.

Deliverables

- Conducting user requirements review (completed)
- Modification Data model modified: Ongoing, further modifications as needed
- Continuation of creating or modifying data entry forms
- Selected NAWQA and NRSA methods entered and reviewed
- Matching NAWQA and NRSA protocols entered and reviewed (in progress)

Conclusion

The 2013 fiscal year marked CDI's ongoing commitment to enhancing USGS data and information by developing useful products and tools, while providing an open forum for members to collaborate, leverage existing capabilities, and share expertise. The CDI Monthly Forums covered a variety of topics with guest speakers from USGS and outside organizations drawing regular participation from the community. The CDI Working Groups have also featured many focused presentations and made exciting progress in the realm of policy, guidance, and workflows. The CDI Webinar Series, held in lieu of the 2013 The National Map Users Conference and CDI Workshop, was a success with consistently high participation and provided a forum to share existing efforts in citizen

science, data management, open data initiatives, and mobile applications. The CDI also completed its first formalized RFP process, with more than 43 submissions proposing new ideas in management and policy, computational tools, data and information assets, and community innovation. The overall RFP process received positive feedback from the community and reviewers with suggestions for improvements, which will be implemented in the 2014 RFP. Overall the CDI has had a highly successful year and continues to bring researchers, data managers, and information technologists together to sustain the CDI Science Support Framework in support of USGS earth science research.

Membership is voluntary and open to USGS employees and other individuals and organizations willing to contribute to the community. Interested parties are invited to visit the CDI Web site at http://www.usgs.gov/cdi/. Contact *cdi@usgs.gov* for more information about joining CDI.

References Cited

Brotherton, P.N., and Preece, P.F.W., 1996, Teaching science process skills: International Journal of Science Education, v. 18, no. 1, p. 65–74, http://dx.doi.org/10.1080/0950069960180106.

CDI Citizen Science Working Group, 2012, USGS Citizen Science Project Inventory: CDI Citizen Science Working Group Web site, accessed November 07, 2012, at https://my.usgs.gov/confluence/display/cdi/USGS+Citizen+Science+Project+Inventory.

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, E., 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., http://dx.doi.org/10.3133/ofr20131265.

Hay, L.E., LaFontaine, J.H., and Markstrom, S.L., 2014, Evaluation of statistically downscaled GCM output as input for hydrological and stream temperature simulation in the Apalachicola-Chattahoochee-Flint River Basin (1961–1999): Earth Interactions, v. 18, no. 9, 1 p., http://dx.doi.org/10.1175/2013EI000554.1.

Hines, Megan, Benson, Abigail, Govoni, David, Masaki, Derek, Poore, Barbara, Simpson, Annie, and Tessler, Steven, 2013, Partnering for science, Proceedings of the USGS Workshop on Citizen Science: U.S. Geological Survey Open-File Report 2013–1234, 51 p., http://pubs.usgs.gov/of/2013/1234/.

Hutchison, V., Burley, T., Chang, M., Chatfield, T., Cook, R., Henkel, H., Strasser, C., Zolly, L., 2013, USGS data management training modules: U.S. Geological Survey Community for Data Integration Web site, http://dx.doi.org/10.5066/F7RJ4GGJ.

Ignizio, D.A., O'Donnell, M.S., and Talbert, C.B., 2014, Metadata wizard—An easy-to-use tool for creating FGDC–CSDGM metadata for geospatial datasets in ESRI ArcDesktop: U.S. Geological Survey Open-File Report 2014–1132, 14 p*.,* http://dx.doi.org/10.3133/ofr20141132.

Appendix A: CDI Publications

Citizen Science Workshop Proceedings—Partnering for Science: Proceedings of the U.S. Geological Survey Workshop on Citizen Science (U.S. Geological Survey Open-File Report 2013–1234)

Semantic Web Technical Documentation—In preparation with USGS Science Publishing Network

USGS Science Data Lifecycle, at http://pubs.usgs.gov/of/2013/1265/

Appendix B: myScience: Building Partnerships Letters of Support

United States Department of the Interior

OFFICE OF THE SECRETARY
Washington, D.C. 20240

JAN 1 3 2014

Sally Holl
U.S. Geological Survey
Texas Water Science Center
1501 Ferguson Lane
Austin, Texas 78754

Dear Ms. Holl,

The Office of Youth, Partnerships, and Service (YPS) is excited about your Community for Data Integration (CDI) proposal and happy to support your efforts to engage other Department bureaus in the work to further develop and expand myScience.

In 2013, Interior released the U.S. Department of the Interior STEM Education and Employment Pathways Strategic Plan. This vision for the plan is to support a science-literate public and prepare students and professionals from all backgrounds to understand and value the role of science and science inquiry in the stewardship of America's natural resources and cultural heritage. This document provides guidance for Interior, including a series of objectives

The myScience project advances the objectives under "Key Strategic Area 1--Facilitate Access to Coordinated DOI Resources," where the intent is to facilitate access to our resources to increase public use of our STEM-rich assets. This proposal to enhance myScience efforts to date aligns well with the plan's three objectives under "Citizen Science."

I'm happy to assign YPS Education Program Analyst, Maria Arnold, to serve as a collaborator for the proposed myScience project—working with myScience developers, the USGS citizen science community of practice, and others to better connect USGS' work with other bureaus and partners. All will benefit from learning more about each other's citizen science efforts, challenges, and solutions. And, ultimately, partners will better engage and serve the very public they wish to engage in the scientific process.

Best Regards,

Michael Gale
Director
Office of Youth, Partnerships, and Service

Dr. Barbara Poore
US Geological Survey
600 4th Street South
Saint Petersburg FL 33701

January 10, 2013

Dear Dr. Poore,

The Wilson Center's Commons Lab advances research and non-partisan policy analysis on emerging technologies and approaches that facilitate collaborative, science-based and citizen-driven decision-making, with an emphasis on their social, legal, and ethical implications. We are pleased to support a proposal—myScience: USGS Citizen Science Resources, Services, and Partnerships—to the Community for Data Integration of the U.S. Geological Society. Commons Lab collaborated with the 2013 project, which established the myScience web application. We see mutual benefit to both organizations in extending this collaboration in 2014.

Over the last three years, the Commons Lab has published on many of the challenges of citizen science and open government innovation, including: "New Visions in Citizen Science," "Connecting Grassroots and Government," "Tweeting Up A Storm: The Promise and Perils of Crisis Mapping," and "Transforming Earthquake Detection and Science through Citizen Seismology." The latter was written jointly with the USGS. These case studies can serve as resources to the myScience project in constructing a "Project Starter Cookbook." The Wilson Center also can contribute guidance materials for citizen science projects dealing with Federal regulations, such as the Paperwork Reduction Act, the Anti-Deficiency Act, the Information Quality Act, and the Privacy Act. In return, the myScience project will provide guidance to the Wilson Center on citizen science data models for project inventories, on harvesting project inventory metadata through web services, and on the issues surrounding the deployment of citizen science applications on mobile devices. myScience project leaders will participate in the Commons Lab Roundtable on Open Innovation and Science to help establish a community of practice among Federal agencies in the citizen science arena.

The USGS has been a long-standing leader in citizen science and open government innovation within the federal government. We strongly urge the USGS to support this valuable project, an important step forward that will advance open innovation and public participation in government.

Best regards,

Lea Shanley

Lea Shanley
Senior Associate and Director of the Commons Lab
Science and Technology Innovation Program
Woodrow Wilson International Center for Scholars