USGS
science for a changing world

Community for Data Integration 2017 Annual Report

Open-File Report 2018–1110

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

Community for Data Integration 2017 Annual Report

By Leslie Hsu and Madison L. Langseth

Open-File Report 2018–1110

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2018

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

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

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:
Hsu, L., and Langseth, M.L., 2018, Community for Data Integration 2017 annual report: U.S. Geological Survey Open-File Report 2018–1110, 19 p., https://doi.org/10.3133/ofr20181110.

ISSN 2331-1258 (online)

Contents

Tables

Abbreviations

3D	three-dimensional
API	application programming interface
AWS	Amazon Web Services
CDI	Community for Data Integration
CHS	Cloud Hosting Solutions
DevOps	Software Development and Information Technology Operations
DMWG	Data Management Working Group
EPA	U.S. Environmental Protection Agency
ERDDAP	Environmental Research Division’s Data Access Program
FY	fiscal year
GHSC	Geologic Hazards Science Center
ICEMM	Interagency Collaborative for Environmental Modeling and Monitoring
lidar	light detection and ranging
NGA	National Geospatial-Intelligence Agency
RFP	request for proposals
SfM	Structure from Motion
SOI	statement of interest
SysAd	System Administration
TSWG	Technology Stack Working Group
USGS	U.S. Geological Survey
WMA	Water Mission Area

Community for Data Integration 2017 Annual Report

By Leslie Hsu and Madison L. Langseth

Abstract

The Community for Data Integration (CDI) is a group that helps members grow their expertise on all aspects of working with scientific data. The CDI's activities advance data and information integration capabilities in the U.S. Geological Survey and in the wider Earth and biological sciences. This annual report describes the presentations, activities, collaboration areas, workshop, and other CDI-sponsored events in fiscal year 2017. The report also describes the objectives of the 11 CDI-funded projects in fiscal year 2017. The report shows how the CDI activities fulfill the strategic objective of the U.S. Geological Survey's Core Science Systems Mission Area to develop a workplace model for interdisciplinary science.

Introduction

The Community for Data Integration (CDI) is a group that helps members grow their expertise on all aspects of working with scientific data. When it was originally chartered in 2009, the U.S. Geological Survey (USGS) Council for Data Integration was conceived as an official organizational function that would help guide data integration activities in the USGS. However, it quickly became apparent that many more people had an interest in forming a community of practice to help the data integration effort from a more grassroots perspective. Thus, the council was abandoned, and the organization became the Community for Data Integration. The integration of wide-ranging USGS data is important because it facilitates analysis of scientific data and information for scientists and decision makers to do their work more effectively. The CDI focuses on opportunities to share information across disciplines and organizational structures, invigorating cross-boundary communication. Past accomplishments of the community include helping develop data policy and data management education for the USGS, creating and influencing USGS-wide tools and resources such as ScienceBase (https://www.sciencebase.gov) and the MetadataWizard (Ignizio and others, 2014; Talbert, 2017), and supporting open and reusable software practices through training and resources.

The CDI is funded and led by the USGS, but membership is voluntary and open to USGS employees and other individuals and organizations willing to contribute to the community. Members include data managers, research scientists, information technology professionals, program managers, communication specialists, and others. The CDI had 852 members at the end of fiscal year (FY) 2017, 216 of whom were welcomed during the year.

The goal of the CDI is to advance the understanding of Earth systems by

- creating and supporting a community of people interested in sharing strategies, methods, tools, and infrastructure, and providing a forum where members can grow their expertise;
- advocating for practices that support the integration of science information across disciplines and organizational structures;
- supporting innovative ideas through seed-funded projects; and
- developing and holding training opportunities that support data and science integration activities.

The CDI's activities are closely aligned with the USGS Core Science Systems Strategy (Bristol and others, 2013), in particular, with Objective 3.2—develop a workplace model for interdisciplinary science. The strategic actions under that objective are:

1. *Funding models that transcend boundaries.*—Develop funding models to reward interdisciplinary science proposals that transcend discipline and mission area boundaries, including a competitive grants process.

2. *Reduce barriers to interdisciplinary research.*—Reduce cultural and spatial barriers to interdisciplinary research, such as creating virtual laboratories and collaborative offices, linking field scientists directly to the Modular Science Framework (Bristol and others, 2013), and relocating scientists to high-level collaborative centers, such as the USGS Powell Center (https://powellcenter.usgs.gov/) or the National Water Center (http://www.nws.noaa.gov/oh/nwc/).

3. *Identify approaches to interdisciplinary science.*—Continually study the USGS and conduct comparative analyses with other organizations to identify best practices and approaches to interdisciplinary science.

4. *Create opportunities for collaborative learning and advancement of science.*—Promote and facilitate employee involvement with communities of practice (Wenger, 1998) such as the CDI to create more opportunities for collaborative learning and advancement of science.

5. *Inform decision making for coupled human and natural systems.*—Collaborate with organizations that conduct and support research of social and ecological systems to better inform decision making for coupled human and natural systems.

6. *Develop reimbursable opportunities.*—Embrace the development of more reimbursable opportunities at high levels to facilitate accountability in collaborative partnerships with governmental organizations (for example, move to some degree of soft funding to introduce innovation through competition).

This report describes CDI FY 2017 activities and how they relate back to the Core Science Systems Strategy. Activities include monthly forums, an annual workshop, webinar series, collaboration areas, and funded projects. Table 1 summarizes CDI activities and their correlation to the six specific strategic actions listed above.

Table 1. Mapping of Community for Data Integration activities to the six strategic actions under the U.S. Geological Survey Core Science Systems Strategy Objective 3.2—develop a workplace model for interdisciplinary science (Adapted from Bristol and others, 2013).

[Shading indicates not applicable. S, strategic action]

Strategy	Monthly forums	Annual workshop	Collaboration areas	Funded projects
S1—Funding models that transcend boundaries				X
S2—Reduce barriers to interdisciplinary research	X	X	X	X
S3—Identify approaches to interdisciplinary science		X	X	
S4—Create opportunities for collaborative learning and advancement of science	X	X	X	X
S5—Inform decision making for coupled human and natural systems				
S6—Develop reimbursable opportunities				

Monthly Forums

Every month, the CDI gathers for a virtual meeting. These monthly forums enable community members to stay up-to-date on new tools, best practices, standards, and policies within the Earth and biological sciences community. The monthly forums align with Strategy 2—reduce barriers to interdisciplinary research, and Strategy 4—create opportunities for collaborative learning and advancement of science.

Both CDI members and nonmembers are invited to give presentations on topics related to data integration during the monthly forums. Table 2 lists the presentations given in FY 2017. Community members are encouraged to ask questions, present challenges, and share solutions to data integration problems. The monthly forums also provide the CDI executive sponsors and coordinators with the opportunity to announce upcoming CDI activities and interact directly with the community. Additionally, the CDI collaboration area leads are able to report progress on their activities during these meetings. An average of 81 people attended the monthly meetings in FY 2017 (table 2).

Table 2. Monthly Community for Data Integration (CDI) forum presentations for fiscal year (FY) 2017.

[USGS, U.S. Geological Survey]

Date	Presentation title	Speaker(s)	Number of Attendees
October 12, 2016	Funding and Partnership Opportunities Through the USGS Innovation Center	Jonathan Stock, USGS	76
	The Powell Center for Analysis and Synthesis	Jill Baron, USGS	
	mdEditor—A Modern, Accessible Application For Creating Metadata	Joshua Bradley, U.S. Fish and Wildlife Service	
November 9, 2016	CDI FY 2017 Statements of Interest, close of voting session	Leslie Hsu, USGS	82
	Council of Data Facilities—History and Overview	Danie Kinkade, Woods Hole Oceanographic Institute	
	Trusted Digital Repositories (TDR)—Proposed New Criteria and Process Flow	Keith Kirk, John Faundeen, and Clara Brown, all from USGS	
December 14, 2016	FY 2015 CDI Project—Web-Enabled Visualization and Access of Value-Added Disaster Products	Brenda Jones, USGS	53
	FY 2016 CDI Project—Data at Risk and the Legacy Data Inventory and Reporting System	Lance Everette, USGS and John Faundeen, USGS	
January 11, 2017	FY 2015 CDI Project—ScienceCache	ScienceCache Team, USGS	97
	USGS Mapping Innovation Series Report	Mike Tischler, USGS	
	Boundless Open Geographic Information System Platform	Monty Kickert and Steve Stout, both from Boundless	
February 8, 2017	Interagency Collaborative for Environmental Modeling and Monitoring (ICEMM)	Brenda Rashleigh, U.S. Environmental Protection Agency	75
	Facilitating Reproducibility of Scientific Findings through Access to Data, Code, and Research Objects	Victoria Stodden, University of Illinois	
March 8, 2017	CDI FY 2016 Funded Projects, Part 1		84
	Facilitating the USGS Scientific Data Management Foundation by Integrating the Process Into Current Scientific Workflow Systems	Colin Talbert	
	Integration of Phenological Forecast Maps for Assessment of Biodiversity: An Enterprise Workflow	Jake Weltzin	
	Crowd-Sourced Earthquake Detections Integrated into Seismic Processing	Michelle Guy	
	Evaluating a New Open-Source, Standards-Based Framework for Web Portal Development in the Geosciences	Rich Signell	
	Development of Recommended Practices and Workflow for Publishing Digital Data through ScienceBase for Dynamic Visualization	Kathy Chase	
	Hunting Invasive Species with HTCondor: High Throughput Computing for Big Data and Next Generation Sequencing	S. Grace McCalla (all from USGS)	
April 12, 2017	CDI FY 2016 Funded Projects, Part 2		79
	A Data Management and Visualization Framework for Community Vulnerability to Hazards	Jeanne Jones	
	Birds and the Bakken: Integration of Oil Well, Land Cover, and Species Distribution Data to Inform Conservation in Areas of Energy Development	Todd Preston	
	Integration of National Soil and Wetland Datasets: A Toolkit for Reproducible Calculation and Quality Assessment of Imputed Wetland Soil Properties	Eric Sundquist	
	A web-based application for the management and visualization of land-use scenario data	Jason Sherba	
	Data Management Training Clearinghouse	Tamar Norkin	
	National Stream Summarization: Standardizing Stream-Landscape Summaries	Daniel Wieferich (all from USGS)	

Table 2. Monthly Community for Data Integration (CDI) forum presentations for fiscal year (FY) 2017.—Continued

[USGS, U.S. Geological Survey]

Date	Presentation title	Speaker(s)	Number of Attendees
May 10, 2017	U.S. Topo Maps Production System Modernization In-person and Virtual Participation at the 2017 CDI Workshop—Enabling Integrated Science	Bill Marken, USGS Leslie Hsu, USGS	77
June 14, 2017	2017 CDI Workshop Debrief—Charting a Course toward Integrated Science	Leslie Hsu, USGS	68
July 12, 2017	Augmented reality—A Brief Overview of Indiana-Kentucky Water Science Center's Augmented Reality-Related Activities and Processes Drone Based Terrain Capture and Virtual Reality	Peter Cinotto, USGS Ryan Spicer and David Krum, both from University of Southern California	102
August 9, 2017	CDI Software Development Cluster, code.gov, and Software Metadata User Experience at the USGS and University of Tennessee, Knoxville	Blake Draper, USGS, and Eric Martinez, USGS Rachel Volentine, University of Tennessee, Knoxville	98
September 13, 2017	"Reducing Risk Where Tectonic Plates Collide—A USGS Plan to Advance Subduction Zone Science" GeoMAC Wildfire Application Event-Based Flood Data Collection with the Short-Term Network Database—How The USGS Collects, Manages and Disseminates Critical Flood Data for Science and Emergency Response	Joan Gomberg, USGS Elizabeth Lile, USGS Blake Draper, USGS	75

During the CDI monthly forums, CDI members and others have the opportunity to present the challenges they are facing and crowdsource possible solutions. These presentations take place during a short block of time at the beginning of each meeting in a segment called Scientist's Challenge (table 3). The purpose of the crowdsourcing is to tap into CDI's powerful collective body of knowledge, form connections, and identify possible future collaborations between the USGS and the Earth and biological sciences community. Each Scientist's Challenge is posted to the CDI forum, and community members are able to reach out to the scientists and submit guidance, resources, collaboration opportunities, or further questions. Outcomes and solutions are also documented on the CDI forum at https://my.usgs.gov/confluence/x/xylKI.

Table 3. Scientist's Challenges for fiscal year 2017.

[USGS, U.S. Geological Survey; CDI, Community for Data Integration]

Date	Scientist's Challenge	Presenter(s)
November 11, 2016	What Collaboration Methods and Workflows are Scientific Programmers Using?	Jeremiah Lant, USGS
December 14, 2016	Mobile App Framework for Water and Environmental Field Data Collection	Ian Ferguson, Bureau of Reclamation
January 11, 2017	Open the Subsurface to the Public: Visualizing Subsurface Data in a Virtual Globe	Geoff Phelps, USGS
February 8, 2017	Frontiers in Collection and Delivery of Lakes Ecosystem Data	Peter Esselman, USGS
April 12, 2017	Cooperative Distributed Spatial Search for Scientific Data	Peter Schweitzer, USGS
May 10, 2017	Learning Where Our Members Learn	CDI Coordinators
June 14, 2017	Social Media, Breakfast with Bill, and a Multi-Beam Community	CDI Coordinators
July 12, 2017	Data-Driven Web Design with A/B Testing and Experimentation	Jordan Read and Lindsay Carr, USGS
September 13, 2017	Getting Started with Jupyter Notebooks and R Shiny Apps	CDI Coordinators

2017 Community for Data Integration Workshop

The 2017 CDI Workshop, with the theme Enabling Integrated Science, was held in Denver, Colorado, from May 16–19, 2017, at the Denver Federal Center. The purpose of the workshop was to bring together interested parties to discuss current topics, shared challenges, and steps forward to advance integrated science at the USGS. There were 183 in-person attendees and 35 virtual attendees over the 4 days.

This CDI workshop provided a forum for scientists, technologists, data and resource managers, program managers, and other interested parties to convene face-to-face meetings to discuss common methods, interests, challenges, and solutions related to scientific data and technologies. This opportunity for interaction allowed connections to be made across disciplines, backgrounds, and geographical locations for future activities and collaborations. All attendees were encouraged to share their ideas using the mobile application sli.do, which allowed real-time questions and feedback from the audience to be collected.

The primary outcomes of the workshop are recommendations for further action decided during the breakout sessions, which are published in the workshop proceedings in the two sections "Roadmap Discussions on Enabling Integrated Science," and Topical Sessions (Hsu and others, 2018). These sessions, as well as the plenary discussions, identified new areas for collaboration and learning that the CDI plans to facilitate, such as data science, software development, scientific modeling practices, and user needs and user experience. The CDI will build on the results of the workshop to guide its future topics, events, and funding opportunities that build an integrated science capacity for the USGS. A full description of the workshop agenda and outcomes is available in Hsu and others (2018).

The workshop addresses Core Science Systems Strategies 2, 3, and 4 (table 1). Strategy 2—reduce barriers to interdisciplinary research—was achieved by bringing together a diverse group of people from different USGS program areas, geographical regions, and positions to share knowledge and learning. Strategy 3—identify approaches to interdisciplinary science—was addressed by keynote speaker Bruce Caron from Earth Science Information Partners, in his presentation "Beyond the Fourth Paradigm—Integrative Science is also about People." Strategy 4—create opportunities for collaborative learning and advancement of science—was accomplished by the poster and demo session (called the DataBlast) and topical sessions proposed by participants.

Collaboration Areas

The CDI is organized into groups, or collaboration areas, based on common interests in specific topics related to data integration (table 4). Collaboration areas have various names (working groups, clusters, or communities of practice) that reflect their goals and activities and sometimes reflect the naming conventions in effect at the time they were formed. However, all collaboration areas provide a platform for sharing resources and knowledge, discussing challenges, and identifying solutions that will help advance data integration in the Earth and biological sciences.

Each group has one or more leaders to coordinate meetings, projects, and information sharing, as well as to report current activities to the larger CDI community. Collaboration area membership is voluntary and open to anyone interested in participating. In FY 2017, eight new collaboration areas were proposed: Bioinformatics, Data Science, DevOps, Interagency Collaborative for Environmental Modeling and Monitoring, Metadata Reviewers, Open Source Coffee Talks, Software Development, and Structure from Motion. In addition, work continued in FY 2017 on the Communication, Connected Devices, Data Management, Earth-Science Themes, Semantic Web, and Technology Stack collaboration areas. A brief description of each collaboration area and its activities in FY 2017 is provided in the following sections.

These collaboration areas address several of the Core Science Systems strategic actions under Objective 3.2—develop a workplace model for interdisciplinary science. The CDI meets CSS Strategy 2—reduce barriers to interdisciplinary research—by bringing together (virtually) a diverse group of scientists and data professionals from different USGS program areas, geographical regions, and positions to share knowledge and learning opportunities. Collaboration areas meet CSS Strategy 3—identify approaches to interdisciplinary science—by inviting speakers from outside the USGS to the collaboration area meetings for discussions on common topics. CSS Strategy 4—create opportunities for collaborative learning and advancement of science—is accomplished by convening speaker series on focused topics, hosting informal work sessions, and surveying members for topics of greatest interest.

Table 4. Community for Data Integration collaboration areas with activity in fiscal year 2017 and contacts.

Collaboration area topic	Group contact(s)
Bioinformatics	Scott Cornman—rcornman@usgs.gov Christina Kellogg—ckellogg@usgs.gov Denise Akob—dakob@usgs.gov
Communication	John C. Nelson—jcnelson@usgs.gov Marcia McNiff—mmcniff@usgs.gov
Connected Devices	Tim Kern—kernt@usgs.gov Lance Everette—everettel@usgs.gov
Data Management	Viv Hutchison—vhutchison@usgs.gov Cassandra Ladino—ccladino@usgs.gov
Data Science	Lindsay Carr—lcarr@usgs.gov
DevOps	Brian Fox—bfox@usgs.gov
Earth-Science Themes	Roland Viger—rviger@usgs.gov
Interagency Collaborative for Environmental Modeling and Monitoring	Brenda Rashleigh—rashleigh.brenda@epa.gov
Metadata Reviewers	Fran Lightsom—flightsom@usgs.gov
Open Source Coffee Talks	Cassandra Ladino—ccladino@usgs.gov
Semantic Web	Fran Lightsom—flightsom@usgs.gov
Software Development	Blake Draper—bdraper@usgs.gov Michelle Guy—mguy@usgs.gov
Structure from Motion	Pete Chirico—pchirico@usgs.gov
Technology Stack	Richard Signell—rsignell@usgs.gov

Bioinformatics Community of Practice

The Bioinformatics Community of Practice meets monthly to discuss bioinformatics tools, methods, and resources, and data handling techniques (table 5). The Bioinformatics Community of Practice was started under the Earth-Science Themes Working Group in January 2017.

Table 5. Bioinformatics Community of Practice meetings and presentations for fiscal year 2017.

[USGS, U.S. Geological Survey; eDNA, environmental deoxyribonucleic acid; RNA-seq, ribonucleic acid sequencing]

Date	Meeting/presentation title	Speaker(s)
January 24, 2017	Inaugural call	Scott Cornman, Christina Kellogg, and Denise Akob, all from USGS
February 28, 2017	Alces Flight GeoHackathons and eDNA and Invasive Species Work	Courtney Owens, USGS Sophia Liu, USGS
March 21, 2017	Data Release for Bioinformatics Data	JC Nelson, USGS
April 18, 2017	Yeti Resources	Janice Gordon, USGS
May 16, 2017	No meeting because of overlap with 2017 Community for Data Integration Workshop in Denver, Colorado	None
June 20, 2017	Review of Bioinformatics Platform Options and Getting Started with CLC Genomics Workbench	Scott Cornman, USGS, and Janice Gordon, USGS
July 20, 2017	RNA-seq—Measuring Gene Expression with High-Throughput Sequencing	Scott Cornman, USGS
August 15, 2017	KBase—A Software and Data Platform Designed to Meet the Grand Challenge of Systems Biology	Ben Allen, Oak Ridge National Laboratory

Communication Working Group

The goal of the Communication Working Group, which was started in FY 2016, is to create lines of communication between the CDI, science centers, regional offices, and mission areas of the USGS. During FY 2017, the Communication Working Group met every few months to discuss various topics including finalization of the CDI communication plan, feedback from a CDI interactive session, improvements to the CDI wiki site, how to follow up on the CDI Annual Meeting, and helping to recruit CDI member stories. Member stories are brief profile pages that describe a member's interests, how they got involved in CDI, a CDI event or topic that was exciting to them, and what they hope the CDI network will help them achieve. Toward the end of the year, the group agreed to meet only when communication needs arise, such as to help organize training and resources for CDI members, or to utilize the expertise of trained USGS communicators.

Connected Devices Working Group

The Connected Devices Working Group explores application development and the use of mobile tools, frameworks, and thingbots to support scientists. In December 2016, the group discussed the Nonindigenous Aquatic Species Mobile Data Collector (https://nas.er.usgs.gov/mobilesightingreport.aspx) and the U.S. Department of the Interior Mobile Privacy Policy (https://www.doi.gov/sites/doi.gov/files/uploads/ocio_directive_2016-003_doi_mobile_applications_privacy_policy.pdf). The group also considered the next steps for their Mobile App Development Checklists in light of the U.S. Department of Interior Mobile Privacy Policy and the October 31, 2016 USGS Instructional Memorandum, Review and Approval of Software for Release (https://www2.usgs.gov/usgs-manual/im/IM-OSQI-2016-01.html). In mid-FY 2017 the working group was reformed as a USGS Slack channel (#mobile under https://usgs.slack.com). The focus in FY 2017 was helping new developers navigate the mobile application release process.

Data Management Working Group

The Data Management Working Group fosters best practices and collaborative approaches for incorporating data management into USGS science and educating scientists about the value of data management. The group seeks to elevate the practice of data management such that it is seen as a critical part of the pursuit of science in the USGS. In FY 2017, the Data Management Working Group hosted a series of presentations to provide updates and information on data management tools in the USGS and beyond (table 6).

Table 6. Data Management Working Group (DMWG) webinar series and monthly meeting presentations for fiscal year 2017.

[USGS, U.S. Geological Survey; CDI, Community for Data Integration]

Date	Title	Speaker
November 14, 2016	Earth Science Information Partners Data Management Training Data Management Plans Page Tiger Team Update	Tamar Norkin, USGS Michelle Chang, USGS
December 12, 2016	Open Source Metadata Tools—Standards, Translator, Editor	Joshua Bradley, U.S. Fish and Wildlife Service
January 9, 2017	Metadata Implementation Guidance Metadata Reviewers Group Update	Ray Obuch, USGS Peter Schweitzer and Fran Lightsom, USGS
February 13, 2017	Experiences in Coordinating the Ecosystem Mission Area Science Centers for Data Management and Release	JC Nelson, USGS
March 13, 2017	ORCiDs Pubs Warehouse Updates Data Management Website Updates	Clara Brown, USGS Jim Kreft, USGS and Clara Brown, USGS Michelle Chang, USGS
April 10, 2017	Legacy Data Inventory Evaluation and Prioritization Updates to the Online Metadata Editor	Lance Everette, USGS Lisa Zolly, USGS
May 8, 2017	CDI Workshop theme—Enabling Integrated Science Face-to-Face Meeting and Hot Topics for the Coming Year	Leslie Hsu, USGS Cassandra Ladino, USGS
June 12, 2017	Review results of the CDI DMWG in-person meeting	Viv Hutchison, USGS
July 10, 2017	USGS Science Data Catalog	Lisa Zolly, USGS
September 11, 2017	DMWG Updates and Introduction to Data Management for Integrated Science Overview of DAMA International Data Management Body of Knowledge (DMBoK) v.2	Cassandra Ladino, USGS Lowell Fryman (DAMA International, the Data Management Association, Rocky Mountain Chapter/Collibra)

Data Science Community of Practice

The purpose of the Data Science Community of Practice is to share content related to data science at the USGS. For purposes of the CDI, data science is defined as the application of computer science, machine learning, data visualization, and other emerging technical approaches to enhance more traditional USGS science. The group was initiated at the 2017 CDI Workshop when a directory for data science enthusiasts was started on the CDI wiki space. The Data Science Community of Practice does not have regular meetings. Instead, they communicate through forums on GitHub (https://github.com/usgs/best-practices) and on USGS Slack (#data-science).

DevOps Working Group

The purpose of the DevOps Working Group is to share new techniques and lessons learned using DevOps tools and methods. DevOps is short for Software Development and Information Technology Operations and the group aims to improve efficiency by unifying software development and software operation, which have traditionally been separate tasks in organizations. The DevOps Working Group existed as a separate group in USGS, but came under the umbrella of CDI in June 2017 to increase awareness and participation in the group. The DevOps Working Group has two focus groups: (1) Project Management Sync and (2) System Administrator (SysAd) and Developer Sync.

Both focus groups facilitate communication across organizational, regional, and managerial boundaries, allowing USGS project managers, information technology, and development staff to share how DevOps-related methods, techniques, and tools are enabling their local activity. These focus groups provide feedback on current cloud capabilities and performance to USGS representatives. The groups also allow technology managers and staff from throughout the USGS to discuss policy recommendations, and provide a venue for senior bureau leadership to hear about opportunities to eliminate barriers related to technology, policy, or process (tables 7 and 8).

Table 7. Software Development and Information Technology Operations (DevOps) Working Group Project Management Sync topics for fiscal year 2017.

[USGS, U.S. Geological Survey; GHSC, Geologic Hazards Science Center; EPA, U.S. Environmental Protection Agency; WMA, Water Mission Area; AWS, Amazon Web Services; NGA, National Geospatial-Intelligence Agency; CHS, Cloud Hosting Solutions]

Date	Title	Speaker
February 1, 2017	DevOps “saved” the new USGS Stream Gage Data Management System	Scott Lewein, USGS
	DevOps/Information Technology Operations at the USGS Astrogeology Science Center	Rian Bogle, USGS
	USGS Cloud Hosting Solutions	Tim Quinn, USGS
March 1, 2017	Quick Overview of the USGS Software Release Instructional Memoranda	Michelle Guy, USGS
	Overview of DevOps Process and Various Tools	Brian Paulsmeyer, Centric Consulting
	Minimal Viable Products, Evolution over Revolution	Lynda Lastowka, USGS
April 4, 2017	EPA’s implementation of DevOps	Robin Gonzalez, EPA
	How WMA Automates Deployments In AWS	Ivan Suftin, USGS
May 2, 2017	Geospatial Intelligence Services DevOps	Mike Finnessy, NGA
	WMA Provisioning of AQUARIUS Time-Series Servers within AWS/CHS	Joel Dudley, USGS
	Demo of GHSC Cloud Foundry	Eric Martinez, USGS
June 6, 2017	How Tasktop Achieves Traceability Across the Value Stream	Laura Horner, Tasktop
	DevOps—What Does a High Performing Team Look Like?	Richard Seroter, Pivotal
	Discussion on enterprise tools	Brian Fox, USGS
July 11, 2017	Walkthrough of Software Release Policy	Michelle Guy, USGS
	Redhat OpenShift	Atif Chaughtai, Red Hat
August 1, 2017	Pivotal and the U.S. Air Force	Jeff Howard, Pivotal
	WMA—“Optimizing the Whole”	Scott Lewein, USGS
September 12, 2017	CHS Docker Managed Service	Jonathan Russo, USGS
	CHS Overview and Road to a Test/Dev Environment	Courtney Owens, Eric Larson, and Emma Sirr, all from USGS

Table 8. Software Development and Information Technology Operations (DevOps) System Administrator (SysAd) and Developer Sync topics for fiscal year 2017.

[USGS, U.S. Geological Survey; WMA, Water Mission Area]

Date	Title	Speaker
July 13, 2017	Use of Pivotal Cloud Foundry Versus a Do-It-Yourself Approach Demo of System Monitor Tool	Tim Kern, USGS Robert Djurasaj, USGS
August 1, 2017	Use of Pivotal Cloud Foundry: Demo and question and answer session WMA Monitoring within Amazon Web Services/Cloud Hosting Solutions (CHS)	Eric Martinez, USGS Jim Morris, USGS
September 12, 2017	Terraform Automating Secure Sockets Layer (SSL) Certificate Creation	Ivan Fetch, USGS Shawn Noble, USGS

Earth-Science Themes Working Group

The goal of the Earth-Science Themes Working Group is to provide a forum for applied Earth science within the CDI. An additional goal of the group is to bring fundamental Earth science data producers, such as the USGS National Hydrography Dataset, 3D Elevation, and Multi-Resolution Land Characteristics Programs, into more direct and regular contact with scientists who work to integrate the sometimes independent data sources developed by these programs. In FY 2017, the Earth-Science Themes Working Group provided an umbrella for distinct themes of bioinformatics, elevation, water, soils, and land cover.

Interagency Collaborative for Environmental Modeling and Monitoring

The Interagency Collaborative for Environmental Modeling and Monitoring (ICEMM; https://my.usgs.gov/confluence/x/0K5tI) is a U.S. Federal government group chartered through a Memorandum of Understanding. The group includes six Federal agencies: (1) U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research; (2) U.S. Department of Defense, Army Corps of Engineers, Engineer Research and Development Center; (3) U.S. Department of Energy, Office of Environmental Management; (4) U.S. Department of the Interior, U.S. Geological Survey; (5) U.S. Environmental Protection Agency, Office of Research and Development; and (6) National Science Foundation, Geoscience Directorate.

The purpose of ICEMM is to continue and strengthen a framework for facilitating cooperation and coordination among Federal agencies in research and development of multimedia environmental models, software, and related databases. Multimedia model development and simulation supports interagency investigations into risk assessment, uncertainty analyses, water supply issues, and contaminant transport. ICEMM was started in 2014 and was brought under the CDI umbrella in January 2017 to increase awareness and participation of the group. ICEMM consists of four workgroups: Watershed Modeling, Data Assimilation, Integrated Modeling and Monitoring, and Ecosystem Functions and Services. ICEMM hosts annual in-person public meetings to discuss the work that is taking place across various Federal agencies.

Metadata Reviewers Community of Practice

The purpose of the Metadata Reviewers Community of Practice is to provide a forum for people who review metadata so that consistent standards can be used throughout the USGS. This group enables people new to this role to learn from experienced metadata reviewers. The group met monthly to discuss various topics related to metadata review, listen to presentations, and to improve resources for USGS metadata reviewers (table 9).

The outcomes of the group discussions were recorded on the CDI wiki. Questions discussed on the CDI wiki include

- Can we help science fields that don’t mesh with Federal Geographic Data Committee metadata standard in, for example, the field of genomics and others that contribute to big integrated databases?
- How do we deal with the suggestion that some data are not worth the time and trouble it takes to write complete metadata records? What is our response as individual reviewers, and as a community?
- How much information is enough for data quality information? Are there good examples for different situations?

Table 9. Metadata Reviewers Community of Practice topics for fiscal year 2017.

[USGS, U.S. Geological Survey; CDI, Community for Data Integration]

Date	Title	Speaker
October 3, 2016	Keywords in Metadata	USGS Thesaurus Team
November 7, 2016	Reviewing the Data and Metadata Review Checklists The CDI Workshop	Group discussion
December 5, 2016	Discussion of updates to the USGS Data Management Website	Group discussion
February 6, 2017	Geospatial Metadata Validation Service	Peter Schweitzer, USGS
March 6, 2017	Proposed Energy Program Standards for Metadata Quality	Ray Obuch, USGS
April 3, 2017	How do we deal with the suggestion that some data are not worth the time and trouble it takes to write complete metadata records? What is our response as individual reviewers, and as a community?	Group discussion
May 1, 2017	New Developments with the Metadata Wizard	Colin Talbert, USGS
June 5, 2017	Reviewing the USGS Data Management Website and Data and Metadata Review Checklists	Group discussion
July 3, 2017	Reviewing the USGS Data Management Website and Data and Metadata Review Checklists	Group discussion
August 7, 2017	Metadata Tips for Better Discoverability of Data in the USGS Science Data Catalog	Lisa Zolly, USGS
September 5, 2017	Reviewing the USGS Data Management Website and Data and Metadata Review Checklists	Group discussion

Open Source Coffee Talks

The Open Source Coffee Talks are held by a group of web development and communications specialists interested in building community and learning how industry-leading open source packages can help the USGS provide scientific information on the web. This group is currently being run in a casual coffee talk format to facilitate knowledge and culture building among participants. The group's goal is to pose new questions, try new technologies, and create an interactive learning experience. A topic is proposed by a different member each month and the group investigates and discusses it during the 1-hour-long meeting. Topics and technologies covered included libraries, application programming interfaces, and graphical user interfaces for charting and graphing JavaScript Object Notation data; trello; and GitLab.

Semantic Web Working Group

The Semantic Web Working Group is a group of data practitioners who are working together to explore semantic web technologies to improve the discovery, access, use, and integration of USGS data. Topics discussed in the FY 2017 monthly meetings included vocabulary server governance; Integrated Taxonomic Information System vocabulary services; persistent identifiers/locators for linked data components; user stories about the use of controlled vocabularies; possible future activities that would provide semantic web techniques to enhance USGS capabilities for integrated science; semantic metamodeling (Villa and others, 2017); a CDI Knowledge Base; and user stories for at least two potential future projects, a permanent USGS triple store (a database built for the storage and retrieval of triples [a type of data entity] through semantic queries) and a USGS database of data dictionary elements.

Software Development Cluster

The Software Development Cluster is a community for USGS software developers and other interested parties to discuss software release protocols and policies; development best practices; software metadata; and software libraries, packages, and tools. The Software Development Cluster was initiated in August 2017. The cluster held web conference discussions on credit and citation for code, and software repository requirements and recommendations. The group also has discussions on the USGS Slack channel #software-dev.

Structure from Motion Community of Practice

The Structure from Motion (SfM) Community of Practice was initiated in November 2016 to facilitate the sharing of information and tips on SfM methods and tools. SfM is a photogrammetric technique for estimating three-dimensional structures from two-dimensional images. The method often employs very large volumes of image data and new software technologies. The group started a directory for those who are interested in SfM on the CDI wiki space. The SfM Community of Practice communicates and shares resources using the SfM Forum at https://my.usgs.gov/confluence/x/co86IQ.

Technology Stack Working Group

The goal of the Technology Stack Working Group (TSWG) is to explore and share technologies that aid data discovery, access, and interoperability. The TSWG informs USGS providers and users about tools and techniques to improve efficiency when working with scientific data. TSWG continued its partnership with Earth Science Information Partners for the Tech Dive webinar series; information on the monthly webinars is provided at http://wiki.esipfed.org/index.php/Interoperability_and_Technology. The TSWG piloted some new presentation formats such as the Environmental Research Division's Data Access Program (ERDDAP) lightning talks, which allowed participants to learn about a wide range of the program's implementations and examples in one meeting. Table 10 presents information on the group's meetings and presentations for FY 2017.

Table 10. Technology Stack Working Group meetings and presentations for fiscal year 2017.

[3D, three-dimensional; USGS, U.S. Geological Survey; ERDDAP, Environmental Research Division's Data Access Program]

Date	Title	Speaker(s)
October 13, 2016	EarthCube Integration and Test Environment (ECITE)	Phil Yang, George Mason University
November 10, 2016	Introducing 3D Tiles	Todd Smith, Analytical Graphics, Inc.
December 8, 2016	Vector Tile Maps	Sam Matthews, Mapbox
January 19, 2017	Introduction to Google Earth Engine	Jess Walker, USGS
February 9, 2017	Web AppBuilder for ArcGIS	Derek Law, Esri
March 9, 2017	Introduction to Esri Story Maps	Christine White, Esri
April 13, 2017	Processing Planetary-Scale Data in the Cloud	Drew Bollinger, Development Seed
May 11, 2017	TerriaJS—A Free, Open-Source Library for Building Web-Based Geospatial Data Explorers	Kevin Ring, CSIRO (Commonwealth Scientific and Industrial Research Organisation)/Data61, Australia
June 6, 2017	Installing JupyterHub in the Cloud Using Kubernetes Helm	Yuvi Panda, Berkeley Institute for Data Science
July 13, 2017	GeoServer Developments	Jody Garnett and Kevin Smith, Boundless
August 10, 2017	ERDDAP—Easier Access to Scientific Data	Bob Simons, National Oceanic and Atmospheric Administration
August 31, 2017	[Bonus] ERDDAP 5-minute lightning talks	Jenn Sevadjian, National Oceanic and Atmospheric Administration Jim Potemra, University of Hawai'i Conor Delaney, Irish Center for High-End Computing Kevin O'Brien, University of Washington John Kerfoot, Rutgers University Stephanie Petillo, Woods Hole Oceanographic Institution Charles Carleton, National Oceanic and Atmospheric Administration Eli Hunter, Rutgers University
September 14, 2017	JupyterHub and JupyterLab Developments	Brian Granger, California Polytechnic State University

Annual Community for Data Integration Request for Proposals

The CDI seeks to build and share knowledge about topics such as data integration, data handling and stewardship, scientific computing, and knowledge delivery. The main goal of CDI is to improve our collective knowledge about how to create better, longer lasting, and more accessible science products by leveraging the tools, methods, and datasets available to the Earth and biological science communities. The CDI places high value on innovative projects that, in the near future, produce new and reusable ideas, methods, or tools that have an impact beyond a single USGS program, center, region, or mission area. The CDI provides up to $50,000 per project. Project proposals are evaluated based on (1) their alignment with the CDI Science Support Framework (USGS, 2015); (2) the evaluation criteria laid out in the request for proposals (RFP) guidance document (USGS, 2016), including scope, technical approach, project experience and collaboration, sustainability, budget justification, and timeline; and (3) how the proposal supports the following CDI guiding principles (USGS, 2016):

- Focus on targeted efforts that yield near-term benefits to Earth and biological science;
- Leverage existing capabilities and data;
- Implement and demonstrate innovative solutions, such as methodologies, tools, or integration concepts, that could be used or replicated by others at scales from project to enterprise;
- Preserve, expose, and improve access to Earth and biological science data, models, and other outputs; and
- Develop, organize, and share knowledge and best practices in data integration.

Formal guidance for the FY 2017 RFP (USGS, 2016) was released on September 14, 2016. The guidance document outlined the two-phased approach that would be used for selecting the CDI FY 2017 projects.

The annual CDI proposals process addresses CSS strategies 1, 2, and 4:

- Strategy 1—funding models that transcend boundaries—is addressed with a competitive grants process that rewards interdisciplinary science proposals that transcend discipline and mission area boundaries.
- Strategy 2—reduce barriers to interdisciplinary research—is addressed by providing funding that allows interdisciplinary project teams to meet during their project.
- Strategy 4—create opportunities for collaborative learning and advancement of science—is addressed by promoting employee involvement with communities of practice. When principle investigators that are not yet members of the CDI submit statements of interest (SOIs) during the proposals process, they are included in the CDI community and given information about our other activities.

Phase I—Statements of Interest

Two-page SOIs were due October 14, 2016. Thirty SOIs were submitted that focused on 14 CDI Science Support Framework elements (table 11). The lead principal investigators and collaborators on the SOIs represented six USGS mission areas (table 12) and all seven USGS regions (table 13).

Table 11. Number of statements of interest addressing each Science Support Framework element for fiscal year 2017.

[Statements of interest could relate to up to three Science Support Framework elements.]

Science Support Framework element	Number of proposals
Publishing/sharing	14
Applications	9
Communities of practice	9
Data	8
Data management	8
Analysis	8
Processing	6
Web services	5
Information	4
Science project support	4
Preservation	4
Knowledge management	2
Planning	2
Acquisition	1

Table 12. Number of statements of interest with representation from each U.S. Geological Survey mission area for fiscal year 2017.

[The lead principal investigators and collaborators for each statement of interest could come from more than one mission area.]

Mission area	Number of statements of interest
Ecosystems	18
Water	7
Core Science Systems	5
Climate and Land-Use Change	4
Natural Hazards	4
Energy and Minerals	2

Table 13. Number of statements of interest with representation from each U.S. Geological Survey region for fiscal year 2017.

[The lead principal investigators and collaborators for each statement of interest could come from more than one region.]

Region	Number of proposals
Midwest	10
Southwest	10
Alaska	6
Northeast	5
Northwest	5
Pacific	5
Southeast	4
Headquarters	4

The CDI community members were asked to review all 30 SOIs and vote on them based on the CDI Science Support Framework, the evaluation criteria, and the guiding principles previously described. The voting period began on October 19, 2016, and closed on November 9, 2016. Each community member was allowed 15 votes to use across all SOIs, and each SOI could receive a maximum of 3 votes per person. During the closing session, the community agreed that the top 19 SOIs should be recommended to the Executive Sponsors for the full proposal phase of the RFP. Following the closing session, the CDI Coordinators also reviewed the SOIs and recommended that an additional two proposals move on to the next phase. In the end, 21 SOIs were approved by the Executive Sponsors to be invited to submit full proposals.

Phase II—Full Proposals

Nineteen full proposals were submitted for the second phase of the RFP process. The CDI convened a formal, 6-person review panel to evaluate the 19 full proposals. The reviewers were USGS employees, both members of the CDI and nonmembers, who volunteered their time to participate on the review panel. The reviewers represented a wide range of USGS mission areas, regions, and programs and brought with them a variety of scientific and technical expertise. The review panel agreed on an order of priority for the full proposals to be funded based on the availability of funds.

Recommendations

The prioritized list from the CDI review panel was presented to the CDI Executive Sponsors, Kevin Gallagher and Tim Quinn, for final selection and approval. Funding came from the USGS Core Science Systems Mission Area and the USGS Office of Enterprise Information. The "Community for Data Integration Projects" section provides a summary for each of the 11 projects funded in FY 2017 (table 14). A description of the accomplishments for each of the projects will be provided in a separate report.

Table 14. Overview of the Community for Data Integration projects funded in fiscal year 2017 (in alphabetical order by principal investigator last name). Project title hyperlinks resolve to a ScienceBase web page describing the project and linking to external resources such as publications, code repositories, and related websites.

[USGS, U.S. Geological Survey; API, application programming interface; lidar, light detection and ranging]

Title	Lead Principal Investigator(s)	USGS Lead Program
An Interactive Web-Based Application for Earthquake-Triggered Ground Failure Inventories	Kate Allstadt	Geologic Hazards Science Center
Automating the Use of Citizen Scientists' Biodiversity Surveys in iNaturalist to Facilitate Early Detection of Species' Responses to Climate Change	Erin Boydston	San Diego Field Station, Western Ecological Research Center
Flocks of a Feather Dock Together—Using Docker and HTCondor to Link High-Throughput Computing Across the USGS	Richard Erickson	Upper Midwest Environmental Sciences Center
USGS Data at Risk—Expanding Legacy Data Inventory and Preservation Strategies	Anthony Everette	Fort Collins Science Center
Exploring the USGS Science Data Life Cycle in the Cloud	Nadine Golden	Pacific Coastal and Marine Science Center
Empowering Decision Makers—A Dynamic Web Interface for Running Bayesian Networks	Erika Lentz	Woods Hole Coastal and Marine Science Center
Web Mapping Application for a Historical Geologic Field Photo Collection	Sarah Nagorsen	Science Publishing Network
Visualizing Community Exposure and Evacuation Potential to Tsunami Hazards Using an Interactive Tableau Dashboard	Jeff Peters	Western Geographic Science Center
Developing APIs to Support Enterprise Level Monitoring Using Existing Tools	Brian Reichert	Fort Collins Science Center
Extending ScienceCache Mobile Application for Data Collection to Accommodate Broader Use Within USGS	Mark Wiltermuth	Northern Prairie Wildlife Research Center
Evaluation and Testing of Standardized Forest Vegetation Metrics Derived from Lidar Data	John Young	Aquatic Ecology Branch, Leetown Science Center

Community for Data Integration Projects

The CDI coordinators decided to postpone the publication of the full project reports for FY 2017 until FY 2018 to give project teams the time necessary to write up their accomplishments and complete their document deliverables. Project titles in table 14 are hyperlinked to the ScienceBase record for each project, which provides links to current deliverables and related external resources such as publications, code repositories, and websites. The sections below provide summaries of each FY 2017 CDI project.

An Interactive Web-Based Application for Earthquake-Triggered Ground Failure Inventories

Earthquakes often trigger landslides and can liquefy loose, wet soils—a process known as liquefaction—increasing the potential for damage to buildings and infrastructure. Tools that predict where landslides or liquefaction may occur after an earthquake can help local and Federal planners decrease earthquake impacts. These tools also can assist emergency responders in planning their response efforts. Ground failure inventories are historical data that document the landsliding and liquefaction triggered by past earthquakes. These inventories are a vital component in developing these much needed tools. Many of these inventories have been created during the past decades but there is no centralized location where a scientist can access the information in a consolidated format. The aim of this project is to use ScienceBase and ArcGIS Online tools to create a web application where researchers and the general public can interactively browse the inventories, perform basic analyses, and download data and metadata. Our project team plans to provide a template for combining existing user-friendly tools that are free to USGS users to create a custom interactive database and web application. This approach could be replicated by other persons who seek custom solutions to sharing their data and results but lack access to web development resources.

Contact: Kate Allstadt, USGS Geologic Hazards Science Center, (303) 273-8570, kallstadt@usgs.gov

Automating the Use of Citizen Scientists' Biodiversity Surveys in iNaturalist to Facilitate Early Detection of Species' Responses to Climate Change

A BioBlitz is a field survey method that finds and documents as many species as possible in a specific area over a short period of time. The National Park Service hosted the largest BioBlitz ever held in 2016, with citizen scientists at more than 120 national parks using the iNaturalist app on their mobile devices to document the species they observed. The resulting data are spatially accurate because global positioning systems were used and biologically accurate because the data were checked by naturalists. As a result, the data provide an unprecedented resource for surveying biodiversity. Additional processing, integration, and analysis would make these data available to inform conservation and management decisions. This project plans to develop a process to rapidly integrate iNaturalist citizen science data with existing species lists; this integration may help detect range shifts of native species or new occurrences of nonnative species. The process can also serve as the basis for incorporating other online databases of citizen science input and to increase engagement of the public in biodiversity stewardship.

Contact: Erin Boydston, USGS Western Ecological Research Center, (805) 370-2362, eboydston@usgs.gov

Flocks of a Feather Dock Together—Using Docker and HTCondor to Link High-Throughput Computing Across the U.S. Geological Survey

USGS scientists often face computationally intensive tasks that require high-throughput computing capabilities. Several USGS facilities use HTCondor to run their computational pools, but these pools may not be connected to the larger USGS pool. This project plans to document how to connect HTCondor pools by "flocking," or coordinating, within the USGS. We also plan to develop tutorials on how to "sandbox" code using Docker within the USGS environment for use with high-throughput computing. The results from this project would not only help the USGS to operate more efficiently by sharing computational resources, but can be adapted by other organizations utilizing HTCondor to run their computational pools.

Contact: Richard Erickson, USGS Upper Midwest Environmental Science Center, (608) 781-6353, rerickson@usgs.gov

U.S. Geological Survey Data at Risk—Expanding Legacy Data Inventory and Preservation Strategies

For more than 135 years, the USGS has collected diverse information about the natural world and how it interacts with society. Much of this legacy information is one-of-a-kind and in danger of being lost forever through decay of materials, obsolete technology, or staff changes. This project plans to produce a systematic way for the USGS to continue efforts to meet the challenge of preserving and making accessible the enormous amount of information that is currently in inaccessible formats. The project plans to develop a formal method to submit, document, and evaluate legacy data known to be in need of preservation. This tool could assist the USGS and other data collection organizations in identifying and prioritizing significant historical legacy data for archiving and release, thereby preserving the information for current and future generations to further scientific discovery, public policies, or decisions.

Contact: Lance Everette, USGS Fort Collins Science Center, (970) 226-9225, everettel@usgs.gov

Exploring the U.S. Geological Survey Science Data Life Cycle in the Cloud

USGS scientists run surface water, groundwater, ocean, and geophysical simulations; transform thousands of photographs into topography on hundreds of individual computers; and generate datasets as large as 10 terabytes. These activities may require the movement of large volumes of data from servers to local computers using state-of-the-art hardware, but the computers may be limited in processing and sharing capabilities. The USGS supports Cloud Hosting to make it more efficient for scientists to acquire, analyze, preserve, and share these large datasets, but specific workflows have not been established. This project plans to assess the benefits, costs, and any issues associated with transitioning two workflows, coastal ocean modeling and groundwater modeling, into the Cloud Hosting infrastructure. The results would be useful to USGS scientists looking to transition their workflow into the cloud environment.

Contact: Nadine Golden, USGS Pacific and Coastal Marine Science Center, (831) 460-7530, ngolden@usgs.gov

Empowering Decision Makers—A Dynamic Web Interface for Running Bayesian Networks

Many groups of people need information on sea-level rise and its effect on coastal landscapes, including prospective home buyers, community planners, and natural resource managers. USGS scientists have expertise in developing probabilistic models (Bayesian Networks) to predict potential beach erosion, sea level rise impacts, habitat change, and groundwater availability. Currently, to use these models, technical software and statistical knowledge are needed. As a result, much of the information contained in the models is largely inaccessible by the general public. To improve access to the models and the scenarios used by the USGS to drive them, this project plans to use freely available and open software to create a user-friendly, interactive web interface. The end product would allow a user to explore a variety of coastal hazard scenarios generated by Bayesian Networks and improve communication of USGS models and their outcomes.

Contact: Erika Lentz, USGS Coastal and Marine Science Center, (508) 457-2238, elentz@usgs.gov

Web Mapping Application for a Historical Geologic Field Photo Collection

Presently, photos are easier to take, are of higher quality, and capture much more information than in the past. Geospatial information recorded by digital cameras could be incorporated into geographic information system mapping tools to easily explore and interact with field photo collections. Many studies could benefit from the ability to share and display photos by position within a study area. This project plans to repurpose the Land Cover Trends Field Photo Map application (CDI FY 2015 project) to more effectively display photos from a 43-year Grand Canyon geologic mapping project. Open source tools and instructions would be developed and published allowing others to geotag photos and create photo map applications. These tools could also be used to streamline and improve methods for sharing USGS and other Federal photo collections with the general public.

Contact: Jason Sherba, USGS Western Geographic Science Center, (650) 329-4248, jsheba@usgs.gov

Visualizing Community Exposure and Evacuation Potential to Tsunami Hazards Using an Interactive Tableau Dashboard

Risk reduction planning organizations across the United States rely on USGS science to determine community exposure to and evacuation potential for natural hazards. Currently, USGS science is shared in published reports and journal articles that contain static maps, figures, and tables. Interactive graphics to visualize this science would allow interested parties to tailor the content, form, and appearance of a vulnerability analysis to best suit their specific planning needs. This project plans to create a new model for disseminating hazard-exposure data using the third-party software Tableau to provide interactive interpretation of maps and results. The project also plans to provide a comparison of labor and maintenance costs and interactive functionality of using licensed software versus using in-house programmers to develop and publish interactive graphic interfaces. The project plans to use community exposure to and pedestrian evacuation for tsunami hazards on the island of O‘ahu, Hawai‘i, as the case study for this project. Results would lay the foundation for a new way to better communicate community vulnerability for all hazards.

Contact: Jeff Peters, USGS Western Geographic Science Center, (650) 329-4221, jpeters@usgs.gov

Developing Application Programming Interfaces to Support Enterprise-Level Monitoring Using Existing Tools

From the individual researcher to the institutional level, there is a growing demand for better and more consistent documentation of monitoring and evaluation protocols. Monitoring Resources (https://www.monitoringresources.org/) offers resources that promote better documentation and enable more efficiency in collaboration and data sharing between programs. To demonstrate an example of a project using these shared community resources, this project will connect the North American Bat Monitoring Program web application and database to the web application MonitoringResources.org using application programing interfaces. This project would enhance existing web applications, data discovery tools, and metadata documentation to support aspects of the data management process so that data from different projects will become more compatible for analysis. Results would illustrate a process that individual research and monitoring projects that operate at different scales can use to select standard monitoring site locations and coordinate monitoring protocols associated with those sites.

Contact: Brian Reichert, USGS Fort Collins Science Center, (970) 226-9245, breichert@usgs.gov

Extending ScienceCache Mobile Application for Data Collection to Accommodate Broader Use Within the U.S. Geological Survey

ScienceCache is a mobile device application originally developed for a citizen science project to do place-based data collection. There is great potential to extend the technology behind ScienceCache to be more useful and customizable for researchers and citizen scientists collecting data on a mobile device. The primary goal is to develop a system where researchers can create a survey, deploy that survey to mobile devices, and manage the resulting data in an online database. Software upgrades include integration of mobile device sensors to record data such as location and images, real-time or near real-time upload of information into a centralized database, and data validation at time of observation. These upgrades would reduce the time needed for research scientists to collect, enter, validate, and manage large amounts of field data.

Contact: Mark Wiltermuth, USGS Northern Prairie Wildlife Research Center, (701) 253-5567, mwiltermuth@usgs.gov

Evaluation and Testing of Standardized Forest Vegetation Metrics Derived from Lidar Data

Light detection and ranging (lidar) data contain a wealth of information that is currently being underutilized. Generally, the product of interest has been high-resolution digital elevation models, but characterizing the three-dimensional nature of vegetation with lidar data enables mapping of vegetation height, structure, and volume over large areas. These mapped attributes have proven to be extremely useful for habitat studies, vegetation biomass and biomass change studies, and wildfire behavior models. This project plans to formalize procedures for the automated generation of vegetation attributes from lidar data using data collected under the USGS 3D Elevation Program. It would also produce a standardized set of vegetation products that would be stored in the cloud and could be processed for individualized products. The project would make available large sets of vegetation products not currently available and allow others using similar lidar technology to produce their own vegetation products.

Contact: John Young, USGS Leetown Science Center, (304) 724-4469, jyoung@usgs.gov

Developing a Workplace Model for Interdisciplinary Science

As discussed in this report and shown in table 1, the community activities, monthly meetings, collaboration areas, annual workshop, and funded projects correlate with the Core Science Systems Strategy actions under Objective 3.2—develop a workplace model for interdisciplinary science. All four CDI categories correlate with the two strategies of reducing cultural and spatial barriers to interdisciplinary research and creating opportunities for collaborative learning and advancement of science.

The CDI facilitators undertake additional activities that also are aligned with the objective of developing a workplace model for interdisciplinary science. The CDI facilitators attend external meetings and conferences in order to support Strategy 3—continually study the USGS and conduct comparative analyses with other organizations to identify best practices and approaches to interdisciplinary science. For example, in 2017, the USGS presented its report "The Community for Data Integration (CDI)—Connection and Collaboration with the Research Data Alliance" at the Research Data Alliance Plenary 10.

The CDI facilitators, in partnership with the community's sponsors, also look for opportunities to support Strategy 5—collaborate with organizations that conduct and support research of social and ecological systems to better inform decision making for coupled human and natural systems. The CDI members support Strategy 5 by attending USGS and external meetings focused on these topics and looking for collaboration opportunities. For example, CDI members attended the Community for Surface Dynamics Modeling System 2017 meeting Modeling Coupled Earth and Human Systems—The Dynamic Duo.

Strategy 6 is embrace the development of more reimbursable opportunities at high levels (that is, the USGS program level) to facilitate accountability in collaborative partnerships with governmental organizations. Although Strategy 6 is, for the most part, out of the scope of the grassroots CDI, the cross-organization relationships forged in the CDI activities play a role in identifying and pursuing additional collaborative partnerships.

Summary

Through monthly forums, workshops, working groups, projects, and constant surveying of the community's needs, the Community for Data Integration (CDI) has provided valuable content that keeps current members engaged and attracts new members. In fiscal year 2017, the CDI experienced increased membership and a sharp increase in the number of proposed collaboration areas. We also increased the opportunities community members have to let others know of their work in collaboration areas or in their own research, with new segments in the monthly meetings and efforts like the CDI member stories.

As shown in this report, the CDI activities strongly correlate to the strategic actions under the Core Science Systems Strategy Objective 3.2—develop a workplace model for interdisciplinary science, with the intent of not only advancing the capabilities of the Core Science Systems Mission Area, but all areas of the U.S. Geological Survey. The CDI is able to achieve these actions with its unique position as a forum for cross-U.S. Geological Survey, cross-region, and cross-discipline communication. As the CDI increases in visibility at the U.S. Geological Survey and beyond, we will continue to facilitate activities to support data and science integration activities for the Earth and biological sciences.

Acknowledgments

The authors would like to thank all of the members of the Community for Data Integration, especially the CDI coordinators, for their input into this annual report. We would also like to thank the two USGS reviewers, Leah Colasuonno and Mona Khalil, for comments that improved the text.

References Cited

Bristol, R.S., Euliss, N.H., Jr., Booth, N.L., Burkardt, N., Diffendorfer, J.E., Gesch, D.B., McCallum, B.E., Miller, D.M., Morman, S.A., Poore, B.S., Signell, R.P., and Viger, R.J., 2013, U.S. Geological Survey Core Science Systems Strategy—Characterizing, synthesizing, and understanding the critical zone through a modular science framework: U.S. Geological Survey Circular 1383–B, 33 p.

Hsu, L., Hutchison, V.B., Langseth, M.L., and Wheeler, B., 2018, U.S. Geological Survey Community for Data Integration 2017 Workshop Proceedings: U.S. Geological Survey Open-File Report 2018–1081, 56 p., https://doi.org/10.3133/ofr20181081.

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., accessed May 31, 2018, at https://doi.org/10.3133/ofr20141132.

Talbert, C., 2017, MetadataWizard: U.S. Geological Survey, accessed May 31, 2018, at https://doi.org/10.5066/f7v9870d.

U.S. Geological Survey [USGS], 2015, U.S. Geological Survey Community for Data Integration (CDI) Science Support Framework (SSF): U.S. Geological Survey, 3 p., accessed March 29, 2018, at http://www.usgs.gov/cdi/cdi-ssf/cdi-ssf-components.pdf.

U.S. Geological Survey [USGS], 2016, U.S. Geological Survey (USGS) Community for Data Integration (CDI) request for proposals (RFP): U.S. Geological Survey, 14 p., accessed May 31, 2018, at https://my.usgs.gov/confluence/display/cdi/2017+Proposals?preview=/549946297/555648157/CDI%20FY17%20Request%20for%20Proposals_final.pdf.

Villa F., Balbi S., Athanasiadis, I.N., and Caracciolo, C., 2017, Semantics for interoperability of distributed data and models—Foundations for better-connected information [version 1; referees: 2 approved with reservations]: F1000Research, 6:686, accessed May 31, 2018, at https://doi.org/10.12688/f1000research.11638.1.

Wenger, E., 1998, Communities of practice—Learning, meaning, and identity: Cambridge, United Kingdom, Cambridge University Press, 318 p.

Hsu and Langseth—**Community for Data Integration 2017 Annual Report**—Open-File Report 2018–1110

ISSN 2331-1258 (online)
https://doi.org/10.3133/ofr20181110