

Community for Data Integration 2020 Annual Report



COMMUNITY for DATA INTEGRATION

Open-File Report 2022–1034

Community for Data Integration 2020 Annual Report

By Leslie Hsu, Amanda N. Liford, and Grace C. Donovan

Open-File Report 2022–1034

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

U.S. Geological Survey, Reston, Virginia: 2022

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., Liford, A.N., and Donovan, G.C., 2022, Community for data integration 2020 annual report: U.S. Geological Survey Open-File Report 2022–1034, 16 p., <https://doi.org/10.3133/ofr20221034>.

ISSN 2331-1258 (online)

Contents

Abstract.....	1
Introduction.....	1
Monthly Forums.....	2
Collaboration Areas.....	2
Data Visualization	4
Inland Riverine Oil Spill (IROS)	4
Usability.....	4
Special Events and Training.....	4
Annual Community for Data Integration Request for Proposals.....	7
Additional Community for Data Integration Publications and References	7
Summary.....	7
Acknowledgments.....	7
References Cited.....	9
Appendix 1. Presentations and Speakers	10

Figures

1. Pie chart showing membership affiliations of the Community for Data Integration as of September 20201
2. Community for Data Integration monthly meeting attendance growth, fiscal year 2016–202
3. Community for Data Integration collaboration area names and what they do6

Tables

1. Monthly Community for Data Integration (CDI) forum presentations from fiscal year 20203
2. Active Community for Data Integration collaboration areas during fiscal year 2020 and contact information for group leads5
3. Overview of the Community for Data Integration projects funded in fiscal year 20208

Abbreviations

CDI	Community for Data Integration
eDNA	environmental deoxyribonucleic acid
FAIR	findable, accessible, interoperable, reusable
FY	fiscal year
IROS	Inland Riverine Oil Spill
USGS	U.S. Geological Survey

Community for Data Integration 2020 Annual Report

By Leslie Hsu, Amanda N. Liford, and Grace C. Donovan

Abstract

The Community for Data Integration is a community of practice whose purpose is to advance the data integration capabilities of the U.S. Geological Survey. In fiscal year 2020, the Community for Data Integration held 11 monthly forums, facilitated 13 collaboration areas, and supported 13 projects. The activities supported the broad U.S. Geological Survey priority of producing building blocks for doing integrated predictive science. Specifically, the activities supported tools and methods for findable, accessible, interoperable, and reusable (FAIR) data and wildland fire and water prediction. Through these efforts, community members were informed of new and emerging technologies and data topics that helped them accomplish their professional responsibilities.

Introduction

“The Community for Data Integration (CDI) community of practice builds the U.S. Geological Survey’s (USGS) knowledge base for data integration in the Earth and biological sciences. The CDI accomplishes this objective by creating an environment to help members of the community increase their expertise in all aspects of working with scientific data. The CDI focuses on providing a space to share information across disciplines and organizational structures, invigorating cross-boundary communication. Through these efforts, community members are informed of new and emerging technologies and topics that may help them in their varied professional and personal responsibilities.” (Hsu and Liford, 2021, p. 1)

At the end of fiscal year (FY) 2020, the CDI had 1,686 members, an increase of 25 percent in 1 year. Members included USGS employees, USGS contractors, and members from academia, non-profits, commercial companies, and other State, Federal, and international organizations (fig. 1). The roles that members serve in their organizations include data managers, data scientists, research scientists, software developers, managers, and other practitioners.

The CDI acts as a connection point between grassroots practitioner efforts and USGS executive leadership priorities. The major types of CDI activities described in this report include

- Monthly virtual forums for all the CDI,
- Smaller, more targeted collaboration area meetings,
- Special events and training, and
- An annual proposals process.

While CDI activities and topics are driven by member needs, yearly themes for the proposal process are chosen by CDI’s executive sponsors (members of the USGS Executive Leadership Team) to align with current USGS science goals. The themes requested for FY 2020 proposals include (1) tools and methods supporting wildland fire and water prediction, (2) producing findable, accessible, interoperable, and reusable (FAIR) data and tools, and (3) reusing or repurposing modular tools such as those that were developed by previous CDI projects.

A key component of the CDI community is the emphasis on dynamic and engaging activities aimed to build the capacity of members of all career stages. Programming is focused on innovation and communication between CDI members who may not otherwise cross paths in their professional duties. To plan and support valuable CDI activities, community facilitators are specifically tasked with collecting input and ideas from the community and implementing those ideas. In FY 2020, facilitation was done by 3 partially-supported USGS employees (about 1.25 full-time employees), as well as a broader group of about 30 individuals that volunteered their time to run collaboration area activities. The CDI seeks out new, interactive methods for skill-building, not only in members’ scientific methods, but in how they learn, share, and collaborate. This is how the CDI fills a need in the practitioners’ quest for exemplary Earth science data collection, management, analysis, and publication. The following sections explain in more detail the events, outputs, and impact of the CDI community of practice during FY 2020.

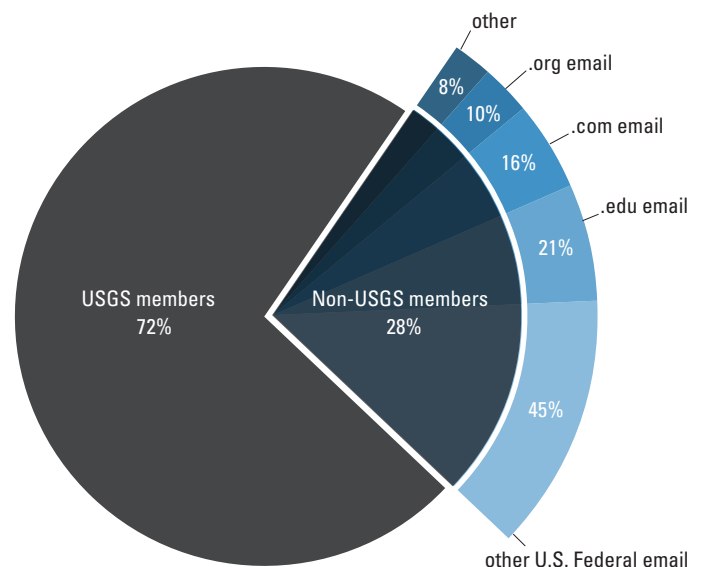


Figure 1. Pie chart showing membership affiliations of the Community for Data Integration as of September 2020.

Monthly Forums

The CDI monthly virtual forums are 90-minute online meetings that inform members about new tools, best practices, standards, and policies within the Earth and biological sciences community. The monthly forums foster interdisciplinary interaction, create opportunities for collaborative learning, and promote communication between all CDI members, including USGS leadership sponsors.

Topics from FY 2020 included visualization tools, the open-source Pangeo framework, and a project evaluating the portfolio of USGS Water Mission Area web applications. There were 13 presentations on the outputs from FY 2019 funded CDI projects. A new segment called “CDI Pop-Up Lab” gave CDI members the opportunity to present questions, challenges, and solutions to the community for discussion. [Table 1](#) is a comprehensive list of the presentation dates, titles, speakers and total number of attendees. CDI monthly forum participation increased by 31 percent from the previous year to an average of 147 online participants per virtual meeting. The steady increase in monthly meeting attendees illustrates the growth of the community attending these calls to learn about information and tools to help them work with their data ([fig. 2](#)).

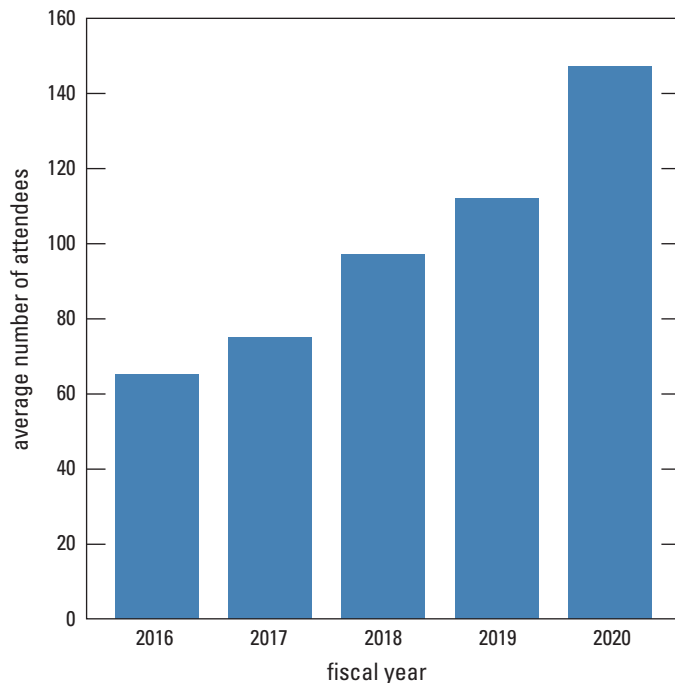


Figure 2. Community for Data Integration monthly meeting attendance growth, fiscal year 2016–20.

Collaboration Areas

The CDI is organized into collaboration areas that form around common interests in specific topics related to data integration ([table 2](#), [fig. 3](#)). In FY 2020, three new collaboration areas were formed or started holding their activities within the CDI: Data Visualization, Inland Riverine Oil Spill, and Usability. In addition, the Citizen-Science collaboration area was rebranded to Open Innovation, to better reflect its focus. Ten other collaboration areas—Artificial Intelligence and Machine Learning, Data Management, DevOps, eDNA, Fire Science, Metadata Reviewers, Risk Research and Applications, Semantic Web, Software Development, and Technology Stack, continued their activities. All collaboration area activities are documented in appendix 1.

Collaboration area leads completed an exercise to document their group purpose and activities to describe the value and benefits of their groups. Representative answers are shown below.

What is the purpose of the collaboration area?

- Data Management—elevate the practice of data management such that it is seen as a critical part of the pursuit of science in the USGS.
- Risk—ensure USGS information is incorporated and used by decision makers to reduce loss.
- Usability—help USGS tools and services be useful.
- eDNA—improve collaboration within the community

Who should join?

- Risk—anyone who wants to contribute to and learn from the community.
- eDNA and Semantic Web—anyone with an interest in the topic.
- Software Development—anyone who develops, reviews, or publicly releases code in the USGS.

How best can the membership collaborate?

- Data Management—share resources and give presentations.
- Risk—attend monthly meetings and participate in the Request For Proposals.
- Usability—provide feedback on resources shared.
- Metadata Reviewers—ask questions and discuss the questions others ask.

Table 1. Monthly Community for Data Integration (CDI) forum presentations from fiscal year 2020.

[DevOps, Software Development and Information Technology Operations; NCAR, National Center for Atmospheric Research; USGS, U.S. Geological Survey; eDNA, environmental deoxyribonucleic acid; FAIR, findable, accessible, interoperable and reusable; SAS, Science, Analytics, and Synthesis; NGA, National Geospatial-Intelligence Agency; GEOINT, Geospatial Intelligence; CDI, Community for Data Integration]

Date	Title(s)	Speaker(s)	Number of attendees
October 9, 2019	“Packaging Data and Software: Jupyter as a Publication Platform”	Keith Maull and Matthew Mayernik, NCAR	112
	“DevOps's Role in Data integration and Delivery”	David Hughes and Rob Djurasaj, USGS	
November 13, 2019	“Visualization Tools at the USGS”	Lindsay Platt, Laura DeCicco, Rich Signell, Ben Letcher, and Dionne Zoanni, USGS	190
January 8, 2020	“Application Evaluation: How to get to a Portfolio of Mission Effective Applications”	Nicole Herman-Mercer, USGS	105
	“USGS Open Innovation Strategy for Crowdsourcing, Citizen Science, and Competitions”	Sophia B. Liu, USGS	
February 2, 2020	“Pangeo: A Flexible Open-source Framework for Scalable, Data-proximate Analysis and Visualization”	Rich Signell and Renee Pieschke, USGS	155
March 11, 2020	“Subsidence Susceptibility Map for the Conterminous United States”	Jeanne Jones, USGS	109
	“High-Resolution, Interagency Biosurveillance of Threatened Surface Waters in the United States”	Elliott Barnhart, USGS	
	“National Public Screening Tool for Invasive and Non-native Aquatic Species Data”	Wesley Daniel, USGS	
April 8, 2020	“Open-source and Open-workflow Climate Scenarios Toolbox for Adaptation Planning”	Aparna Bamzai, USGS	165
	“Develop Cloud Computing Capability at Streamgages using Amazon Web Services GreenGrass Internet of Things Framework for Camera Image Velocity Gaging”	Frank Engel, USGS	
	“Establishing Standards and Integrating environmental DNA (eDNA) Data into the USGS Nonindigenous Aquatic Species Database”	Jason Ferrante, USGS	
May 13, 2020	“Building a Roadmap for Making Data FAIR in the U.S. Geological Survey”	Fran Lightsom, USGS	127
	“Implementing a Grassland Productivity Forecast Tool for the United States Southwest”	Sasha Reed, USGS	
	“A Generic Web Application to Visualize and Understand Movements of Tagged Animals”	Ben Letcher, USGS	
June 10, 2020	“Extending ScienceBase for Disaster Risk Reduction”	Joe Bard, USGS	172
	“Coupling Hydrologic Models with Data Services in an Interoperable Modeling Framework”	Rich McDonald, USGS	
	“Transforming Biosurveillance by Standardizing and Serving 40 Years of Wildlife Disease Data”	Neil Baertlein, USGS	
July 8, 2020	“It's July 2020: Do You Know What Your Data are Doing? SAS Science Data Management: Contributing to USGS Progress in Management of Scientific Data”	Viv Hutchison, USGS	176
	“Field Data Collection using NGA's free and open source Mobile Awareness GEOINT Environment (MAGE) and MapCache Mobile Applications”	Justin Connelly and Ben Foster, National Geospatial-Intelligence Agency	
August 12, 2020	“Integrating Short-term Climate Forecast into a Restoration Management Support Tool”	Caitlin Andrews, USGS	140
	CDI Pop-Up Lab:		
	“Cloud Optimized Files and New Transfer Options”	Theodore Barnhart, USGS	
	“Irregular Meshes for Data Operations—Quadrees”	Thomas Rapstine, USGS	
	“Streamstats”	Katharine Kolb, USGS	
	“Speaking Git and code.usgs.gov ”	Leslie Hsu, USGS	
September 9, 2020	“Who is in the CDI? Staff Profiles”	Leslie Hsu, USGS	168
	“Structured Data on the Web for Earth Science: Emerging Best Practices and Remaining Challenges”	Dave Blodgett, USGS	
	CDI Pop-Up Lab:		
	“Meeting the CDI Community—Mining Expertise Keywords”	Sky Bristol, USGS	
	“Highlights from the Southeast Region Science Meeting”	Greg Steyer, USGS	
“Machine Learning, Satellite Imagery, and Tile Drains”	Tanja Williamson, USGS		

How will members benefit from joining?

- Data Management—Stay up to date on data management best practices, resources, and policies in the USGS.
- Software Development—improve skillsets and become part of a collaborative network.
- Metadata Reviewers—learn how others are dealing with challenges and contribute to standard interpretations of metadata standards.
- Inland Riverine Oil Spill—Expand the technical understanding and use of their oil spill fate and behavior tools to the response community and DOI land managers.

What USGS or Earth science needs are being addressed?

- Data Management—sharing data management practices across domains.
- Risk—reducing risk from hazards.
- Usability—enabling useful scientific services and tools.
- Metadata Reviewers—improving FAIRness of USGS data.

Descriptions of the new collaboration areas formed in FY 2020 are in the following sections, for more information about the other groups, see Hsu and Liford (2021) and [tables 1.1–1.13](#).

Data Visualization

The Data Visualization collaboration area serves as a professional hub and creative home to promote open data visualization trainings and discussions. The top focus of the group is to foster a creative, collaborative community of people who come to data visualization from a variety of backgrounds with any set of design tools. The Data Visualization collaboration area engages with the CDI community mainly through quarterly meetings. Planned activities are to:

- Host events where members can connect and learn from each other,
- Invest in communication platforms to help share skills, knowledge, ideas, and meaningful critique, and
- Encourage new and underrepresented voices as the USGS grows its data visualization community.

Inland Riverine Oil Spill (IROS)

The goals of the Inland Riverine Oil Spill (IROS) collaboration area are to:

- Facilitate communication regarding existing and future national rapid tools for understanding inland riverine spill fate and behavior that are available for spill planning and response.
- Describe common geospatial and hydrologic data uses and needs among the tools.
- Identify a future pathway for more communication and integration across tools and platforms.

The group is cross-agency and the focus of FY20 included compiling tools and models that support the group's work, and hearing presentations on some of the tools.

Usability

The Usability collaboration area brings together people who want to help increase the success of users of their tool or service. This is achieved by learning how to make the tools or services more useful. The top focus in FY20 was to facilitate the understanding of foundational usability concepts such as utility, usability, usefulness, user experience, and user interfaces. Group interactions and knowledge exchange occurred through town hall meetings and resource reviews, which allowed for both virtual discussions and self-paced learning.

Special Events and Training

A virtual group-learning event in 2020 included topics that will help members in their daily work. The topics were usability, the NetCDF data format, unit testing in code development, and Microsoft Power Automate and Power Apps. The objective of the group-learning event was to create accountability for CDI members to spend time on their own to learn about a resource. Facilitated on the online wiki space, interested members were instructed to set a goal, gather as a group, make a deadline, and share learning progress with each other. Fifteen people participated in the group learning event.

Table 2. Active Community for Data Integration collaboration areas during fiscal year 2020 and contact information for group leads.

[DevOps, Software Development and Information Technology Operations; eDNA, environmental deoxyribonucleic acid]

Collaboration area	Group contact(s)
Artificial Intelligence/Machine Learning	Pete Doucette—pdoucette@usgs.gov JC Nelson—jcnelson@usgs.gov
Data Management	Viv Hutchison—vhutchison@usgs.gov Madison Langseth—mlangseth@usgs.gov
Data Visualization	Colleen Nell—cnell@usgs.gov Dionne Zoanni—dzoanni@usgs.gov Amy Puls—apuls@usgs.gov Sophie Hou—chungyihou@contractor.usgs.gov Ellen Bechtel—ebechtel@contractor.usgs.gov Katherine Trickey—ktrickey@usgs.gov
DevOps	David Hughes—drhughes@usgs.gov Derek Masaki—dmasaki@usgs.gov
eDNA	Jason Ferrante—jferrante@usgs.gov Damian Menning—dmanning@usgs.gov Chris Merkes—cmerkes@usgs.gov
Fire Science	Paul Steblein—psteblein@usgs.gov
Inland Riverine Oil Spill	Faith Fitzpatrick—fafitzpa@usgs.gov Jo Ellen Hinck—jhinck@usgs.gov Peter McCarthy—pmccarth@usgs.gov Daniel Sullivan—djsulliv@usgs.gov
Metadata Reviewers	Fran Lightsom—flightsom@usgs.gov
Risk Research and Applications	Kris Ludwig—kaludwig@usgs.gov Emily Brooks—ebrooks@usgs.gov David Ramsey—dramsey@usgs.gov
Open Innovation	Sophia Liu—sophialiu@usgs.gov
Semantic Web	Fran Lightsom—flightsom@usgs.gov
Software Development	Michelle Guy—mguy@usgs.gov Cassandra Ladino—ccladino@usgs.gov Jeremy Newson—jknewson@usgs.gov
Technology Stack	David Blodgett—dblodgett@usgs.gov
Usability	Sophie Hou—chungyihou@contractor.usgs.gov



Figure 3. Community for Data Integration (CDI) collaboration area names and what they do. Column 1: group started activity in fiscal year (FY) 2009–15, column 2: FY 2016–18, column 3: FY 2019, column 4: FY 2020 (DevOps, Software Development and Information Technology Operations; USGS, U.S. Geological Survey; eDNA, environmental deoxyribonucleic acid).

Annual Community for Data Integration Request for Proposals

Through an annual proposal process, the CDI supports projects that focus on data integration for interdisciplinary research, innovative data management, and demonstration of new technology within the USGS computing systems (Hsu and others, 2019). [Table 3](#) provides a list of funded projects. In FY 2020, the CDI themes for its annual proposals were:

- Projects that address one of the major components of the USGS Director’s vision.
- Integrate USGS data into a comprehensive data lake.
- Develop, test, and apply an integrated predictive science capability (EarthMAP) that incorporates data, interpretations, and knowledge that spans discipline boundaries, geographies, and sectors.
- Provide actionable intelligence that can be used via dashboards and applications to enhance situational awareness, provide new operational capabilities, and inform decision making using technologies such as artificial intelligence, machine learning, and high performance computing.
- Tools and methods supporting wildland fire and water prediction, aligned with the Earth Monitoring, Analyses, and Projections (EarthMAP) vision (Jenni and others, 2017).
- Producing FAIR data and tools for integrated predictive science capacity.
- Reusing or repurposing modular tools such as those that were developed by previous CDI projects, including the CDI risk map (Wood and others, 2018).

Additional Community for Data Integration Publications and References

The CDI facilitation team published a study analyzing the outcomes and outputs from the 60 projects funded by the CDI from 2010–16 (Hsu and others, 2019). The projects were evaluated for seven relevant sustainability influences drawn from other studies in other fields: outputs modified, code repository used, champion present, workforce stability, support

from other organizations, collaboration and partnerships, and integration with policy. The study produced a list of lessons for those applying for short term grants and guidance for those who lead programs that provide seed-funding for Earth science informatics projects.

The Community for Data Integration 2018 funded project report (Hsu and others, 2020) summarized the goals, activities, and accomplishments from the 10 projects supported in FY 2018. The FY 2018 Request for Proposals theme was risk assessment and hazard exposure.

The USGS report, “Ecological Forecasting—21st Century Science for 21st Century Management” (Bradford and others, 2020), noted that the CDI community of practice model would be useful for facilitating a community based on forecasted phenology and potentially focused on the concept of species migration networks.

Summary

Through the monthly forums, collaboration areas, events, projects, and listening to the community’s needs, the Community for Data Integration (CDI) provides content that keeps members informed of information and tools to work with their data. The community continues to grow in participants and scope, with new collaboration areas in fiscal year 2020 focused on Usability, Inland Riverine Oil Spill, and Data Visualization.

CDI’s executive sponsors continue to challenge the community to use its diverse expertise to help advance U.S. Geological Survey data integration and management capabilities. To help reach this goal, CDI’s activities in fiscal year 2020 focused on tools and methods supporting the USGS Director’s vision of modernized data capabilities and findable, accessible, interoperable, and reusable (FAIR) data and wildland fire and water prediction. As the CDI increases its visibility at the U.S. Geological Survey and beyond, the CDI team will continue to facilitate activities supporting data and science integration for the Earth and biological sciences.

Acknowledgments

Funding for the Community for Data Integration in fiscal year 2020 came from the U.S. Geological Survey (USGS) Core Science Systems Mission Area and U.S. Geological Survey Office of Enterprise Information (now the Office of the Associate Chief Information Officer). The authors would like to thank Cian Dawson and Jessica LeRoy, both with the USGS, for providing comments that improved this report.

8 Community for Data Integration 2020 Annual Report

Table 3. Overview of the Community for Data Integration projects funded in fiscal year 2020 (in alphabetical order by principal investigator last name).

[USGS, U.S. Geological Survey; AI, artificial intelligence; FAIR, findable, accessible, interoperable, reusable; eDNA, environmental deoxyribonucleic acid; HDF5, Hierarchical Data Format 5]

Project title	Lead principal investigator	U.S. Geological Survey organization
“Building a Framework to Compute Continuous Grids of Basin Characteristics for the Conterminous United States”	Theodore Barnhart	Wyoming-Montana Water Science Center
“So, You Want to Build a Decision Support Tool? Assessing Successes, Pitfalls, and Lessons Learned for Tool Design and Development”	Amanda Cravens	Fort Collins Science Center
“Using Jupyter Notebooks to Tell Data Stories and Create Reproducible Workflows”	Richard Erickson	Upper Midwest Environmental Sciences Center
“USGS Cloud Environment Cookbook”	Aaron Fox	Upper Midwest Environmental Science Center
“Enabling AI for Citizen Science in Fish Ecology”	Nathaniel Hitt	Leetown Science Center
“Implementing FAIR Practices: Storing and Displaying eDNA Data in the USGS Nonindigenous Aquatic Species Database”	Margaret Hunter	Wyoming-Montana Water Science Center
“Developing a “Fire-aware” Stream Gage Network by Integrating USGS Enterprise Databases”	Katharine Kolb	South Atlantic Water Science Center
“Waterbody Rapid Assessment Tool (WaterRAT): 3-Dimensional Visualization of High-Resolution Spatial Data”	Andrea Medenblik	South Atlantic Water Science Center
“Development of a Flexible Multi-Channel Spatiotemporal Geophysical HDF5 Data Format Supporting FAIR”	Jared Peacock	Geology, Energy, Minerals, and Geophysics
“Real-time Coastal Salinity Index for Monitoring Coastal Drought and Ecological Response to Changing Salinity Values”	Matthew Petkewich	South Atlantic Water Science Center
“GrassCast: A Multi-agency Tool Using Remote Sensing, Modeling, and On-the-ground Science to Forecast Grassland Productivity in the Southwest”	Sasha Reed	Southwest Biological Science Center
“Moving Towards EarthMAP: Establishing Linkages among USGS Land Use, Water Use, Runoff, and Recharge Models”	Terry Sohl	Earth Resources Observation and Science
“Using Machine Learning to Map Topographic-soil and Densely-patterned Sub-surface Agricultural Drainage (Tile Drains) from Satellite Imagery”	Tanja Williamson	Ohio-Kentucky-Indiana Water Science Center

References Cited

- Bradford, J.B., Weltzin, J.F., McCormick, M., Baron, J., Bowen, Z., Bristol, S., Carlisle, D., Crimmins, T., Cross, P., DeVivo, J., Dietze, M., Freeman, M., Goldberg, J., Hooten, M., Hsu, L., Jenni, K., Keisman, J., Kennen, J., Lee, K., Lesmes, D., Loftin, K., Miller, B.W., Murdoch, P., Newman, J., Prentice, K.L., Rangwala, I., Read, J., Sieracki, J., Sofaer, H., Thur, S., Toevs, G., Werner, F., White, C.L., White, T., and Wiltermuth, M., 2020, Ecological forecasting—21st century science for 21st century management: U.S. Geological Survey Open-File Report 2020–1073, 54 p., accessed September 29, 2021, at <https://doi.org/10.3133/ofr20201073>.
- Hsu, L., Hutchison, V.B., and Langseth, M.L., 2019, Measuring sustainability of seed-funded earth science informatics projects: PLoS One, v. 14, no. 10, p. e0222807, accessed September 29, 2021 <https://doi.org/10.1371/journal.pone.0222807>.
- Hsu, L., Andrews, C.M., Bradford, J.B., Buscombe, D.D., Chase, K.J., Daniel, W.M., Jones, J.M., Fuller, P., Mirus, B.B., Neilson, M.E., Vraga, H.W., Walker, J.J., Walworth, D.H., Warrick, J., Weltzin, J., Wieferich, D.J., and Wood, N.J., 2020, Community for Data Integration 2018 funded project report: U.S. Geological Survey Open-File Report 1062, 9 p., accessed September 29, 2021, at <https://doi.org/10.3133/ofr20201062>.
- Hsu, L., and Liford, A.N., 2021, Community for Data Integration 2019 Annual Report: U.S. Geological Survey Open-File Report 2021–1016, 19 p., accessed September 29, 2021, at <https://doi.org/10.3133/ofr20211016>.
- Jenni, K.E., Goldhaber, M.B., Betancourt, J.L., Baron, J.S., Bristol, R.S., Cantrill, M., Exter, P.E., Focazio, M.J., Haines, J.W., Hay, L.E., Hsu, L., Labson, V.F., Lafferty, K.D., Ludwig, K.A., Milly, P.C., Morelli, T.L., Morman, S.A., Nassar, N.T., Newman, T.R., Ostroff, A.C., Read, J.S., Reed, S.C., Shapiro, C.D., Smith, R.A., Sanford, W.E., Sohl, T.L., Stets, E.G., Terando, A.J., Tillitt, D.E., Tischler, M.A., Toccalino, P.L., Wald, D.J., Waldrop, M.P., Wein, A., Weltzin, J.F., and Zimmerman, C.E., 2017, Grand challenges for integrated USGS science—A workshop report: U.S. Geological Survey Open-File Report 2017–1076, 94 p., accessed September 29, 2021, at <https://doi.org/10.3133/ofr20171076>.
- Martel, A., Spitzen-van der Sluijs, A., Blooi, M., Bert, W., Ducatelle, R., Fisher, M.C., Woeltjes, A., Bosman, W., Chiers, K., Bossuyt, F., and Pasmans, F., 2013, *Batrachochytrium salamandrivorans* sp. nov. causes lethal chytridiomycosis in amphibians: Proceedings of the National Academy of Sciences of the United States of America, v. 110, no. 38, p. 15325–15329 <https://doi.org/10.1073/pnas.1307356110>.
- Wood, N.J., Jones, J.M., Henry, K.D., Sherba, J.T., and Ng, P., 2018, CDI Risk Map: U.S. Geological Survey ScienceBase web page, accessed August 31, 2020, at <https://www.sciencebase.gov/catalog/item/5b91a0c2e4b0702d0e808bb2>

Appendix 1. Presentations and Speakers

This appendix lists presentation topics and speakers from the different collaboration area meetings for fiscal year 2020.

Table 1.1. Artificial Intelligence and Machine Learning meetings and presentations from fiscal year 2020.

[MLHub: Machine Learning Hub; AI/ML, artificial intelligence and machine learning; USGS, U.S. Geological Survey; DELTA, Deep Learning Language Technology Platform; NASA, National Aeronautics and Space Administration; AWS, Amazon Web Services; AI, artificial intelligence; CHS, Cloud Hosting Services]

Date	Title(s)	Speaker(s)
October 8, 2019	“Radiant MLHub: A Repository for Machine Learning Ready Geospatial Training Data”	Hamed Alemohammad, Radiant Earth Foundation
November 12, 2019	“AI/ML in USGS Enabled by Tallgrass: Classifying Golden Eagle Behavior Using Telemetry”	Natalya Rapstine, USGS
April 14, 2020	“Fine-scale Mapping of Water Features at the National-scale Using Machine Learning Analysis of High-resolution Satellite Images: Application of the New AI/ML Natural Resource Software—DELTA”	Michael Furlong, NASA-Ames Intelligent Robotics Group, Jack Eggleston and John Stock, USGS
May 12, 2020	“Amazon SageMaker”	Inseok Heo, Envision Engineering AWS, Amogh Gaikwad, AWS, Phillip Dawson, USGS
June 9, 2020	“USGS Tallgrass Supercomputer 101 for AI/ML”	Natalya Rapstine, USGS
July 14, 2020	“New York Water Science Center AI and Gage-Cam”	Daniel Beckman, USGS
September 8, 2020	“Advancing AI/ML for the USGS in the CHS Environment”	Matt Kuckuk, USGS

Table 1.2. Data Management monthly meeting presentations from fiscal year 2020.

[USGS, U.S. Geological Survey; IPDS, Information Product Data System; CMS, Content Management System; SGCI, Science Gateways Community Institute]

Date	Title(s)	Speaker(s)
November 18, 2019	“Data Management Challenge—Updating Thousands of Metadata Records”	Greg Gunther, USGS
	“Bulk Updating Metadata Records in Practice”	VeeAnn Cross and Peter Schweitzer, USGS
December 9, 2019	“How Data Flow from IPDS to CMS”	Lance Everette, USGS
	“Connecting USGS Systems”	Drew Ignizio, USGS
January 13, 2020	“Compass”	Heather Schreppel and VeeAnn Cross, USGS
	“DataDash”	Colin Talbert, USGS
February 10, 2020	“Developing Data Management Value Propositions”	Claire Stirm and Juliana Casavan, SGCI
February 13, 2020	“Developing Data Management Value Propositions—Working Session”	Claire Stirm and Juliana Casavan, SGCI
March 9, 2020	“Communicating Data Management Value Propositions to Scientists”	Claire Stirm and Juliana Casavan, SGCI
April 13, 2020	“Upcoming Updates to the USGS Science Data Catalog”	Lisa Zolly, USGS
May 11, 2020	“Records Management—Turn and Face the Change(s)”	Chris Bartlet, USGS
June 8, 2020	“The Data Curation Network: Extending the Research Data Toolkit”	Liza Coburn, Lisa Johnston, Wendy Kozlowski, and Hannah Hadley, Data Curation Network
July 13, 2020	“Collections Management Instructional Memorandum and Center-level Collection Management Plans”	Lindsay Powers and Brian Buczkowski, USGS
August 10, 2020	“Department of the Interior Records Management Repository”	Lynda Speck and Jim Nagode, Bureau of Reclamation
	“Data Exit Story Time”	Tara Bell, Robin Tillitt, and Sue Kemp, USGS
September 14, 2020	“Digitizing Maps, Photos, and Field Records”	Kelly Haberstroh and Jenny Stevens, USGS
	“Data Management”	Janela Biagas, USGS

Table 1.3. Data Visualization meetings and presentations from fiscal year 2020.

[USGS, U.S. Geological Survey; COVID-19, Coronavirus disease 2019]

Date	Title(s)	Speaker(s)
July 2, 2020	“Exploring the Design Process and Reflections on the USGS COVID-19 Case Finder Platform Built in Tableau”	Chuck Hansen and Brendan Wakefield, USGS

Table 1.4. Software Development and Information Technology Operations (DevOps) topics from fiscal year 2020.

[CDI, Community for Data Integration; USGS, U.S. Geological Survey; O365, Office 365; API, Application Programming Interface; CI/CD, Continuous Integration/ Continuous Delivery and Deployment; AWS, Amazon Web Services; AI, Artificial Intelligence]

Date	Title(s)	Speaker(s)
October 24, 2019	“Software Release and Inventories” “CDI Statements of Interest Related to Software”	Eric Martinez, USGS Aaron Fox and Steve Fick, USGS
January 24, 2020	“Software Policies You Should Know” “O365 Tips and Tricks”	Cassandra Ladino, USGS Group Discussion
February 28, 2020	“Swagger and Micronaut” “Water Quality Data Portal Example”	Jeremy Fee, USGS
March 26, 2020	“Getting Started on the Cloud: Example Workflow for Running USGS Slab2 Models”	Kirstie Hayne, USGS
April 23, 2020	“Fast API and You” “Flask versus Fast API”	Brandon Serna, USGS Jeremy Fee, USGS
May 28, 2020	“Data Warehouse with Redshift (Configuration Management Committee Example)” “ETL (Extract, Transform, Load) with Glue”	Cassandra Ladino, USGS Jeremy Newson, USGS
June 25, 2020	“Using Serverless and GitLab CI/CD to Continuously Deliver AWS Step Functions”	Carl Schroedl, USGS
August 27, 2020	“Synthetic Data and Build Process for AI Imagery and Deep Learning Methods”	Daniel Beckman, USGS

Table 1.5. Inland Riverine Oil Spill meetings and presentations from fiscal year 2020.

[USGS, U.S. Geological Survey; RPS, Rural Planning Services; USACE, United States Army Corps of Engineers; NOAA, National Oceanic and Atmospheric Administration; EPA, Environmental Protection Agency]

Date	Title(s)	Speaker(s)
June 16, 2020	“Overview of Rapid Response Tools, National Models, and Data Requirements”	Faith Fitzpatrick, Jo Ellen Hinck, Peter McCarthy, Al Rea, Roland Viger, David Soong, USGS, Bill Samuels and Rakesh Bahadur, Leidos, Matt Horn, RPS, Chandra Patahck, USACE, and Chris Barker, NOAA
June 17, 2020	“Coordination of Building Blocks for Spill Planning and Rapid Response”	Faith Fitzpatrick, Jo Ellen Hinck, Karl Winters, USGS, Jon Gulch, EPA, and Adam Davis, NOAA

Table 1.6. Fire Science community of practice meetings and presentations from fiscal year 2020.

[lidar, light detection and ranging; USGS, U.S. Geological Survey; NLCD, National Land Cover Database; CASC, Climate Adaptation Science Centers; DOI-OWF; Department of the Interior Office of Wildland Fire]

Date	Title(s)	Speaker(s)
October 15, 2019	“New tools to Understand Post-Fire Hazards: Modeling, lidar Observations, and Data Mining”	Francis Rengers, USGS
December 10, 2019	“Aquatic Ecosystem Vulnerability to Fire and Climate Change in Alaskan Boreal Forests”	Jeff Falke, USGS
March 2, 2020	“Diverse, Long-term, Place-based USGS Fire Ecology Research in northern New Mexico Mountains: Past, Present, and Future”	Craig Allen, USGS
April 21, 2020	“Analysis of Stakeholder Input on USGS Fire Science Communication and Outreach, Science Priorities, and Critical Science Needs”	James Meldrum and Ned Molder, USGS
May 19, 2020	“Scaling up Tree-ring Fire History: From Trees to the Continent and Seasons to Centuries”	Ellis Margolis, USGS
June 16, 2020	“NLCD Rangeland Fractional Component Time-Series: Development and Applications”	Matthew Rigge, USGS
July 21, 2020	“CASC—Climate-Fire Synthesis”	Madeleine Rubenstein, USGS
August 18, 2020	“Wildland Fire Information and Technology (WFIT)—WFIT Strategy”	Roshelle Pederson, DOI-OWF
September 15, 2020	“Grasslands, Grazing, Fire, Invasives”	Amy Symstad, USGS

Table 1.7. Metadata Reviewers meetings and presentations from fiscal year 2020.

[FSP, Fundamental Science Practices; USGS, U.S. Geological Survey]

Date	Title(s)	Speaker(s)
October 7, 2019	“FSP Guidance on Documenting Revisions to USGS Scientific Digital Data Releases”	Group discussion
November 4, 2019	“Identifying Different Types of Metadata”	Group discussion
December 2, 2019	“Review Questions from Community Forums”	Group discussion
January 6, 2020	“Digital Object Identifier (DOI) Tool: New Features and Use with Dataset Revisions”	Lisa Zolly, USGS
February 3, 2020	“Need for Persistent Unique Identifiers in Metadata Records”	Lisa Zolly, USGS
March 2, 2020	“At Your Office, How Much Do You Create a Single Metadata Record For?”	Group discussion
April 6, 2020	“Would it Work to Include, in Metadata Titles, the Date of the Revision or Release of the Data?”	Group discussion
May 4, 2020	“What Type Information (in the Metadata) is Necessary for a Data Publication versus Research Publication?”	Group discussion
June 1, 2020	“Why Does Each of us Review Metadata?”	Group discussion
July 6, 2020	“Discussion of Metadata for Software and Code”	Eric Martinez, USGS
August 3, 2020	“Metadata for Public Release of Legacy Data, for which Full Documentation is Not Available”	Tara Bell, Matt Arsenault, and Sofia Dabrowski, USGS

Table 1.8. Open Innovation meetings and presentations from fiscal year 2020.

[USGS, United States Geological Survey; STEM, Science, Technology, Engineering, and Math; COVID-19, Coronavirus disease 2019; USA, United States of America; NGA, National Geospatial Intelligence Agency; GMU, George Mason University; DOS, Department of State; FEMA, Federal Emergency Management Agency; CAP, Civil Air Patrol; FedCCS, Federal Crowdsourcing and Citizen Science; COASST, Coastal Observation and Seabird Survey Team; DOI, Department of the Interior; USFWS, U.S. Fish and Wildlife Service; NOAA, National Oceanic and Atmospheric Administration; OSTP, Office of Science and Technology Policy; IDA, Institute for Defense Analyses; STPI, Science and Technology Policy Institute]

Date	Title(s)	Speaker(s)
February 26, 2020	“Using Open Innovation to Engage Youth and Underserved Communities in the Pacific Islands”	Sheree Watson, USGS
	“Why Open Innovation Matters and Participating in the USGS Open Innovation Strategy”	Sophia Liu, USGS
March 16, 2020	“Using Volcanic Hazards in Hawai‘i as a STEM Platform in Problem-based Learning”	Jefferson Chang, USGS
April 17, 2020	“All Hands on Deck: COVID-19 Open Innovation Efforts and the Opportunity Project Earth Sprint”	Sophia Liu, USGS
April 23, 2020	“Let’s Celebrate Citizen Science Month, Earth Day, and Take Your Kids to Work Day! Take Your Kids Teleworking with Citizen Science to Understand Our Earth”	Erin Posthumus, USA National Phenology Network, Eleanor Snow, Sally (Sarah) Cook and Erin Korris, USGS
April 30, 2020	“Crowdsourcing All Things Geo”	Elizabeth McCartney, USGS, Chris Clasen, NGA/GMU, Erika Nunez, DOS, Will Mortenson, NGA, Katie Picchione, FEMA, and Scott Kaplan, CAP
May 28, 2020	“Data Quality of the Indigenous Observation Network”	Nicole Herman-Mercer, USGS
	“A Resource of Resources for Citizen Science Data Quality”	Hilary Burgess, COASST
	“The Landscape of COVID-19 Citizen Science and Open Science”	Lea Shanley, University of Wisconsin-Madison
June 18, 2020	“Tackling the Paperwork Reduction Act (PRA) in the Age of Social Media and Web-based Interactive Technology”	Jeff Parrillo, DOI, Madonna Baucum, USFWS, and James Sayer, USGS
June 19, 2020	“Indigenous Observation Network (ION): Community-Based Water Quality Monitoring Project”	Ryan Toohey and Nicole Herman-Mercer, USGS
June 25, 2020	“FedCCS June 2020 Meeting: NOAA Citizen Science on Data Quality and Eterna OpenVaccine for COVID-19”	Laura Oremland, NOAA, and Do Soon Kim, Stanford
July 1, 2020	“Open Innovation for Floods Crowdsourcing, Citizen Science, and Competitions as Force Multipliers for Flood Science and Management”	Sophia Liu, USGS
July 30, 2020	“FedCCS July 2020 Meeting: Online Citizen Science, Update on OSTP Data Call, and COVID-Economics Wiki”	Maria Aristeidou, The Open University, Giff Wong, IDA STPI, and Peter Meyer, Bureau of Labor Statistics
August 27, 2020	“Citizen Science for Social Justice”	Abby Kinchy, Rensselaer Polytechnic Institute
	“Discussion on Digital.gov Webinar on Federal Citizen Science Report”	Group discussion

Table 1.9. Risk meetings and presentations from fiscal year 2020.

[USGS, U.S. Geological Survey; FY, Fiscal Year; RFP, Request for Proposals, HERA; Hazard Exposure and Analytics Tool; DOI, Department of the Interior; USFWS, U.S. Fish and Wildlife Service]

Date	Title(s)	Speaker(s)
October 17, 2019	“Process and Results from a Recent Usability Study of the USGS Volcano Hazards Notification System (HANS)”	Carolyn Driedger and Tina Neal, USGS
October 31, 2019	“Special Meeting—FY20 Risk RFP Informational Webinar”	Risk team
November 21, 2019	“Understanding the Value of Usability and User Experience”	Mike Frame and Sophie Hou, USGS, Rachel Volentine, University of Tennessee, Knoxville
December 19, 2019	“The Burned Area Emergency Response (BAER) Team”	Steve Sobieszczyk, USGS
	“What We Learned at the 2019 Society for Risk Analysis Meeting”	Kris Ludwig, USGS
	“Ecological Risk Assessment: Demystifying Our Science to Inform Regulatory Decisions”	Jeff Steevens, USGS
January 16, 2020	“Applying Usability to Web-Tool Design and Development—Using HERA as a Use Case”	Sophie Hou, USGS
	“Update on the Strategic Hazard Identification and Risk Assessment (SHIRA) Project: Engaging Stakeholders in the Pacific Northwest”	Emily Brooks and Alice Pennaz, USGS
February 20, 2020	“Training: Introducing Impact360 and Toolkit360” (1 of 3)	Scott Miles, Impact360 Alliance
March 19, 2020	“Training: Introducing Impact360 and Toolkit360” (2 of 3)	Scott Miles, Impact360 Alliance
April 16, 2020	“Training: Introducing Impact360 and Toolkit360” (3 of 3)	Scott Miles, Impact360 Alliance
May 21, 2020	“Panel Discussion on Communicating Hazard and Risk Science”	Sara McBride, USGS, Kerry Milch, Temple University, Nanciann Regalado, DOI/USFWS
June 18, 2020	“Quantifying Rock Fall Hazard and Risk to Roadways in National Parks: Yosemite National Park Pilot Project”	Brian Collins, USGS
	“The State of Our Coasts: Coastal Change Hazards Stakeholder Engagement and User Need Assessment”	Juliette Finzi-Hart, USGS
	“Re-visiting Bsal Risk: How 3 Years of Pathogen Surveillance, Research, and Regulatory Action Change Our Understanding of Invasion Risk of the Exotic Amphibian Pathogen Batrachochytrium salamandrivorans (Martel and others, 2013)”	Dan Grear, USGS
	“Communications of Risk—Uranium in Groundwater in northeastern Washington State”	Sue Kahle, USGS
July 16, 2020	“Assessing the Risk of Global Copper Supply Disruption from Earthquakes”	Kishor Jaiswal, USGS
	“Cascading Communication: Exploring How Scientific Research Affects Policy and Earthquake Preparedness Communication in Areas of Induced Seismicity”	Sara McBride, USGS
	“Hazard Exposure Analyst Tool (HEAT)”	Jason Sherba, USGS
	“Ecological Forecasts for Risk Management”	Jake Weltzin, Alyssa Rosemartin, U.S. National Phenology Network
August 11–13, 2020	“Risk Community of Practice Annual Meeting”	Various presenters
September 17, 2020	“Review of 2020 Risk Community of Practice Annual Meeting and Post-Meeting Survey Results”	Risk team
	“Simplified Hazard Maps of Cascade Volcanoes: Lessons Learned and Future Directions + Discussion”	Dave Ramsey, USGS
	“Open Innovation Playbook for Risk Project Update”	Sophia Liu, USGS

Table 1.10. Semantic Web meetings and presentations from fiscal year 2020.

[USGS, U.S. Geological Survey; FAIR, findable, accessible, interoperable, and reusable; ESIP, Earth Science Information Partners]

Date	Title(s)	Speaker(s)
January 9, 2020	“Look Together at the Use of Semantic Technology in SemantEco: a Semantically Powered Modular Architecture for Integrating Distributed Environmental and Ecological Data”	Group discussion
March 12, 2020	“A Practical Example of Semantic Technology in Action: Assessing the Status of Biodiversity in the World’s Oceans”	Sky Bristol, USGS
April 9, 2020	“Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web”	Daniel Garijo, University of Southern California and Maria Poveda-Villalon, Universidad Politecnica de Madrid
May 14, 2020	“ESIP Project to Use Concept Maps”	Brian Wee, University of Colorado, Denver
June 11, 2020	“The Semantic Zoo – Smart Data Hubs, Knowledge Graphs and Data Catalogs”	Group Discussion

Table 1.11. Software Development topics from fiscal year 2020.

[USGS, U.S. Geological Survey; O365, Office 365; API, Application Programming Interface; CI/CD, Continuous Integration/Continuous Delivery and Deployment; AWS, Amazon Web Services; AI, Artificial Intelligence]

Date	Title(s)	Speaker(s)
October 24, 2019	“Software Release and Inventories”	Eric Martinez, USGS
	“CDI Statements of Interest Related to Software”	Aaron Fox and Steve Fick, USGS
January 24, 2020	“Software Policies You Should Know”	Cassandra Ladino, USGS
	“O365 Tips and Tricks”	Group Discussion
February 28, 2020	“Swagger and Micronaut”	Jeremy Fee, USGS
	“Water Quality Data Portal Example”	
March 26, 2020	“Getting Started on the Cloud: Example Workflow for Running USGS Slab2 Models”	Kirstie Haynie, USGS
April 23, 2020	“FastAPI and You”	Brandon Serna, USGS
	“Flask versus FastAPI”	Jeremy Fee, USGS
May 28, 2020	“Data Warehouse with Redshift (Configuration Management Committee Example)”	Cassandra Ladino, USGS
	“ETL (Extract, Transform, Load) with Glue”	Jeremy Newson, USGS
June 25, 2020	“Using Serverless and GitLab CI/CD to Continuously Deliver AWS Step Functions”	Carl Schroedl, USGS
August 27, 2020	“Synthetic Data and Build Process for AI Imagery and Deep Learning Methods”	Daniel Beckman, USGS

Table 1.12. Technology Stack meetings and presentations from fiscal year 2020.

[USGS, United States Geological Survey; DGGS, Discrete Global Grid System; UCAR, University Corporation for Atmospheric Research; CUAHSI, Consortium of Universities for the Advancement of Hydrologic Science; ESIP, Earth Science Information Partners; ORCID, Open Researcher and Contributor Identifier; SELFIE, Second Environmental Linked Features Interoperability Experiment]

Date	Title(s)	Speaker(s)
October 10, 2019	“ELFIE: Environmental Linked Features Interoperability Experiment”	David Blodgett, USGS
November 14, 2019	“Location Integration Project”	Shane Crossman and Matthew Purss, Geoscience Australia
February 12, 2020	“Urban Flooding Open Knowledge Network”	Mike Johnson, University of California, Santa Barbara
March 12, 2020	“DGGS in Action: Provision of Rapid Response During Australian Bushfires and Other Applications”	Shane Crossman and Irina Bastrakova, Geoscience Australia
April 9, 2020	“Unidata Science Gateway”	Julien Chastang, UCAR
May 14, 2020	“CUAHSI HydroShare Update”	Jerad Bales and Anthony Castronova, CUAHSI, Jeff Horsburgh, Utah State University
June 11, 2020	“ESIP Collaboration Infrastructure 2.0”	Erin Robinson, ESIP, Ike Hecht, WikiWorks, Lucas Cioffi, QiQo Chat, Sheila Rabun, ORCID
August 13, 2020	“ESIP Summer Meeting Highlights”	Group discussion
September 10, 2020	“SELFIE and geoconnex.us update”	David Blodgett, USGS

Table 1.13. Usability meetings and presentations from fiscal year 2020.

[CDI, Community for Data Integration; USGS, U.S. Geological Survey; ESIP, Earth Science Information Partners]

Date	Title(s)	Speaker(s)
December 18, 2019	“Introduction to the Collaboration Area Overview of Usability Concepts Application of Usability Concepts to CDI wiki”	Sophie Hou, USGS
February 19, 2020	“Choosing Usability Techniques”	Sophie Hou, USGS
April 15, 2020	“Meeting the Paperwork Reduction Act”	James Sayer, USGS
June 24, 2020	“ESIP—Putting Data to Work—Webinar Series, Usability: How Usability Techniques Can Support Your Scientific Communities”	Sophie Hou, USGS
August 19, 2020	“Human-centered Approach and Usability”	Jamie Albrecht, Impact360 Alliance

