

CDI Monthly Meeting 20210908

September 8, 2021: Landscape Science, Research to Operations with Energy and Minerals Data, and Data Standards

The Community for Data Integration (CDI) meetings are held the 2nd Wednesday of each month from 11:00 a.m. to 12:30 p.m. Eastern Time.

Connection information

Connection information is sent to the [CDI mailing list](#).

[Click here to join the meeting](#)

Or call in (audio only)
[+1 202-640-1187, 209207438#](#) United States, Washington DC

Phone Conference ID: 209 207 438#
[Find a local number](#)

Meeting Recording and Slides

Recordings and slides are available to CDI Members approximately 24 hours after the completion of the meeting.

These are the publicly available resources. Log in to view all meeting resources. If you would like to become a member of CDI, join at <https://listserv.usgs.gov/mailman/listinfo/cdi-all>.



 During the call, you can ask and up-vote questions at [slido.com](https://www.slido.com), event code #CDISEP.

Agenda (in Eastern time)

- 11:00 am Welcome and Opening Announcements
- 11:15 am Collaboration Area Announcements
- 11:25 am **USGS Landscape Science Strategy** - Zack Bowen, USGS
- 11:45 am **Is there a gap between research and operations in the USGS, and how do we fill it?** - Sky Bristol, USGS
- 12:10 pm **Data Standards information** - Amanda Liford, USGS
- 12:30 pm *Adjourn*

Abstracts

USGS Landscape Science Strategy - Zack Bowen, USGS

Our landscapes and resources are undergoing continual change from a complex and interacting suite of stressors that include traditional and nontraditional land uses, a changing climate, a dynamic economy, and a culturally diverse and ever-changing society. Landscape science seeks to understand how the physical, biological, and social components of ecosystems and landscapes interact with each other and are affected by these stressors across local to global scales.

The USGS has developed a landscape science strategy to focus and strengthen the agency's efforts to inform critical conservation, restoration, and management decisions for American landscapes. The strategy directly supports the overarching 21st-century science strategy of the USGS and expands our perspective and focus on partnerships, including with internal, external, traditional, and nontraditional partners.

Landscape Science Strategy 2020-2030: <https://doi.org/10.3133/cir1484>

Zachary Bowen is the Landscape Program Coordinator in the USGS Ecosystems Mission Area.

Is there a gap between research and operations in the USGS, and how do we fill it? Sky Bristol, USGS

In USGS over the last two decades, we've made massive strides forward in embracing and harnessing the power of data technologies in advancing what we do as an institution across our science disciplines. Thanks to things like the seed funding provided through the CDI, we've engaged a huge interdisciplinary and diverse workforce in figuring out what our science can use out of the myriad technological choices increasing in sophistication every day. One area where we need to continue filling the gap has been the subject of a part time detail I've been on for the last year working with the Energy and Minerals Mission Area. We still sometimes lack a mechanism and a place to put the best ideas over the edge into a fully operational state within our Science Centers or as part of some enterprise-wide service. In this presentation, I'll share experiences from three discrete projects we've envisioned and been working on over the summer where we've been exploring this dynamic – how to bring ideas for technological solutions through an engineering cycle to develop long-lasting capabilities.

Sky Bristol is a biologist and data scientist with the Science Analytics and Synthesis program. He works to advance scientific practices and accelerate the pace of discovery through the application of open technologies and frameworks. Sky is currently on part time detail with the Energy and Minerals Mission Area, working to develop principles and an approach toward a new geoscience data framework in the USGS.

Data Standards Information - Amanda Liford, USGS

Data standards are the guidelines by which data are described and recorded. In order to share, exchange, combine and understand data, we must standardize the format as well as the meaning. The USGS Data Management Website has updated content about data standards.

<https://www.usgs.gov/products/data-and-tools/data-management/data-standards>

Amanda Liford is a science data manager at in the Science Data Management Branch within the USGS Science Analytics and Synthesis program. She is a co-facilitator of the CDI since 2019.

Highlights

1. CDI will be hosting a **final Carpentries training covering Python, Shell, and Git on October 26-27**. More information available here: <https://my.usgs.gov/confluence/x/tpVsKQ>
2. The **CDI FY2022 Request for Proposals** is now open!
 - a. The themes are **1) enabling data connection, data readiness, and data comprehension** in support of USGS and Administration research and development priorities such as climate change, renewable energy, and the "Conserving and Restoring America the Beautiful" report, and **2) advancing equity in data**, including approaching data collection with the goal of unbiased representation, and considering the audience for data delivery and reducing barriers to broad access and understanding, especially for underserved communities.
 - b. Schedule:
 - i. Wednesday, September 15, 1PM ET: **RFP Information Session**
 - ii. Friday, October 15, 5PM ET: **Submission Deadline for Statements of Interest**
 - iii. Wednesday, October 20, 1-2:30PM ET: **SOI Lightning Presentation**
3. Zack Bowen shared more about the **Landscape Science Strategy report**, available here: <https://doi.org/10.3133/cir1484>
4. Sky Bristol delved into his **R2E20 Model** and the impetus and thinking behind it, notes below.

Notes

Welcome and Opening Announcements

1. The Future of Work
 - a. Leslie presented **audience responses on predictions on the future of work**, like less travel and more video conferencing, decentralized organization of functional groups that collaborate across domains and places, and virtual-first collaboration and professional development approaches.
2. Intro to Python, Shell, and Git, Oct 26-27
 - a. We will be hosting our final Carpentries training on October 26-27.
 - b. The course will cover:
 - i. Data **analysis and visualization in Python**
 - ii. Working with core data structures
 - iii. Using **conditionals and loops**
 - iv. Writing custom functions
 - v. **Creating customized plots**
 - vi. More information: <https://my.usgs.gov/confluence/x/tpVsKQ>
3. FY2022 CDI Request for Proposals
 - a. **The FY22 RFP is now open!**

- i. The purpose of the CDI FRP is to build USGS capabilities in data integration and management
 - ii. The CDI RFP increases communication across boundaries and creates opportunities to work with people outside of your normal program
 - b. Two-phase RFP
 - i. **Phase 1: statement of interest**
 - 1. Short, 1 page description
 - 2. Lightning presentation
 - 3. Commenting and voting by CDI members
 - 4. Top ranked SOIs are invited to submit a full proposal
 - ii. **Phase 2: invited full proposal**
 - 1. Proposal narrative
 - 2. Review panel
 - 3. Final review by executive leaders
 - c. 2022 Themes
 - i. **Enabling data connection, data readiness, and data comprehension** in support of USGS and Administration research and development priorities such as climate change, renewable energy, and the “Conserving and Restoring America the Beautiful” report.
 - ii. **Advancing equity in data**
 - 1. Equity in data includes
 - a. Approaching data collection with the goal of unbiased representation
 - b. Considering the audience for data delivery and reducing barriers to broad access and understanding, especially for underserved communities
 - d. **CDI RFP Schedule**
 - i. Wed, Sept 15, 1PM ET: RFP Information Session
 - ii. Fri, Oct 15, 5PM ET: Submission Deadline for SOI
 - iii. Wed, Oct 20, 1-2:30PM ET: SOI Lightning Presentations
- 4. Get started
 - a. **View RFP Guidance on CDI SharePoint**
 - b. Discuss ideas on Microsoft Teams channel
 - c. Send questions to the Teams channel or to gs_cdi@usgs.gov

Collaboration Area Announcements

- 1. eDNA
 - a. 5th Annual eDNA Technical Exchange workshop, **Oct 19-21**
 - b. NOAA ‘omics seminar series: **3rd Wed of the month at 12ET**
 - c. **Feb 1-4**; National Workshop on Marine eDNA
- 2. Tech Stack
 - a. **September 9**, TerrainR – generating 3D Landscape Visualizations Using R and Unity
- 3. DevOps
 - a. Next event: **October 5**, ESRI ArcGIS Enterprise Kubernetes Platform – Deploying on CHS
- 4. Inland/Coastal Bathymetry Collaboration Area
 - a. Next event: **September 23**, 10:30AM CT, Monthly Bathymetry Research Coordination Meeting
- 5. DataViz
 - a. Next event: **September 16th**, 1PMET, Beyond Bars and Box Plots – Chart alternatives with ggplot2
- 6. Imagery
 - a. Next event: **October 21**, 2PM ET, Imagery Data Management Workflows
 - b. Cian Dawson: Check out the [beta interactive hydrograph announcement](#) I mentioned in today’s call, using webcams that are part of the Storm Summary Timelapse network
- 7. Usability
 - a. Next event: **October 6 and 20**, 3PM ET, User Centered Design Process – Evaluation: presentation and demo
- 8. Data Management
 - a. Next event: **September 13**, 2PM ET, Data Reports & other Companion Publications: How data and publishing work together

USGS Landscape Science Strategy - Zack Bowen, USGS

- 1. **Landscape Science Strategy 2020-2030**: <https://doi.org/10.3133/cir1484>
- 2. Defining terms
 - a. Landscape science – **aiming to look at ecosystems and interacting components**; use this understanding to help people make policy and management decisions (the what)
 - b. Coproduction – working closely with research partners, managers, and stakeholders to make decisions together (the how)
 - c. Actionable science – **things that people can take and put into use with management actions or policy decisions** (the outcome)
- 3. Why do we need this strategy?
 - a. DOI **priorities of clean energy, tribal nations, equity and environmental justice** demand an integrated approach with our science efforts
 - b. We have a complex landscape of resources.
- 4. Goals and Actions
 - a. Goal A: deliver science and insights that are used in landscape management
 - b. Goal B: understand and communicate how landscapes have changed and will change
 - c. Goal C: engage in science that transcends disciplinary bounds
 - d. Action 1: build approaches that support our partners
 - e. See slides for more.
- 5. Example products
 - a. Greater Sage Grouse science and tech
 - i. Project goal: **compile recently published GRSG science, summarize key information relevant to managers for each article, produce a synthesis for managers**

- ii. Need: inform decisions and potential updates to RMPs within the greater sage grouse range
 - iii. Products: 2 USGS open-file reports, online searchable tool
- 6. Alignment – USGS Science Strategy
 - a. Data Acquisition and Management
 - b. Modeling and Prediction
 - c. Delivery of Actionable Intelligence
 - i. A lot of overlap in the Landscape Science Strategy
 - d. Science to support strong decision-making in the BLM: quality data, relevant science, standard analysis, best practices
 - i. Also overlaps with USGS Science Strategy and Landscape Science Strategy
- 7. **Conserving and Restoring America the Beautiful contains DOI priorities that directs us to produce data and products to implement the initiative to conserve 30% of U.S. lands and water**
- 8. Invitation
 - a. Identify **shared management information needs for data, science, and tools to support decisions**
 - b. Expand on existing projects and relationships across DOI
 - c. Develop an **implementation plan**

Is there a gap between research and operations in the USGS, and how do we fill it? - Sky Bristol, USGS

1. Wanted to **explore the notion that USGS is behind other contemporaries in terms of data capabilities** (in comparison to GeoScience Australia, etc.)
 - a. Just focusing on the 'portal' idea that GeoScience Australia has implemented for some of their data: there are two parts to the portals. Portal includes catalog of all their products, can create a persona to bring back specific data/products. In USGS, we have different catalogs for publications, data, models, etc.
 - b. GeoScience Australia also have nimble, responsive tools that can produce visualizations.
2. Spoke to people in USGS to try to **get to the core of what the problems might be**
 - a. Hero Development Environments – one hero that takes on a database, capability, etc., maintains it. Some things remain sustainable, while others are "one Oracle patch away from disaster". We have older technologies performing critical jobs, but these have not moved forward technologically.
3. **Research to Operations conundrum**
 - a. See slides for Research Lifecycle at University of Central FL lifecycle, ITIL Service Lifecycle, and Engineering life cycles
 - b. See slide for Sky's R2E20 Model
4. Principles-based Architecture
 - a. **Set priorities based on the mission now with mission-next in mind**
 - b. Traceability, transparency, and buildability (ability to have everything that we put out to be able to be built upon)
 - c. Loose coupling, shared services, and data linkability – idea that everything should be coupled with some other parts
 - d. Have a very good reason to rebuild something already built
 - e. New investments **pay down technical debt**
 - f. Work within constraints of operational core competence
 - g. Always be prepared to fail and throw stuff away; if not now, pretty soon
5. Testing principles in action
 - a. GeoArchive
 - i. Trying to **move from a big collection of documents to actual data that we can work with**
 - ii. Toward a solution of reference management and collaborative annotation
 - iii. Ended up going towards use of a reference management framework in Zotero to manage these resources
6. LIMS and NMIC Commodity Stats
 - a. Toward a solution for automated data transformation and distribution through schema documentation
 - b. **Both projects have pointed towards a need for more nimble, linkable, and versioned documentation of methods and protocols**

Data Standards information - Amanda Liford, USGS

1. The Data Management website team **recently revamped the Data Standards page**: [Data Standards \(usgs.gov\)](https://www.usgs.gov/data-standards)
2. Data standards are important for being able to integrate data into other datasets or systems.
3. The site has sections on who produces and ratifies data standards, who is responsible for following data standards, why we need data standards, and includes examples of dataset-level standards, parameter-level standards, and data encoding and interface standards.
4. The DM website also recently put together **an A-Z Index** to make topic browsing easier: [A-Z Index \(usgs.gov\)](https://www.usgs.gov/data-standards/a-z-index)
5. Reach out to GS_Data_Management@usgs.gov with suggestions for the Data Standards page or questions.

Questions

1. USGS Landscape Science Strategy
 - a. From your experience, what are some points in the data life cycle that could use improvement in the landscape science community?
 - i. Biggest area we've spent energy on is effectively using other people's data. There are technical aspects, especially solutions, but also a trust component. We'd like to integrate state data, but it's a challenge overall to be able to effectively work with, integrate, and use other data.
 - b. Is the USGS strategy related to the USFWS' landscape conservation cooperatives?
 - i. The landscape conservation cooperatives have been dissolved/are not funded in the same way. A lot of the ideas in those cooperatives are still important. We are working with USFWS on some of these ideas.
 - c. When we work with our partner agencies, I'm curious what fraction of our contributions are official releases (public) vs. "working data /documents"
 - i. We are right now figuring out ways to better capture when partners are using our products, and what the real-world impact of the products are. There are often informal summaries, analyses, etc., but it is difficult to track those.
 - d. regarding the conservation atlas. can you talk a bit about the process for agreeing on the categories of working/ conserved lands?

- i. There is an existing set of classifications used to look at designating whether something is preserved/protected – IUCN, GAP. For America the Beautiful, we are trying to provide science on informing decisions, but those classifications are not yet established.
- e. Is the idea to develop and implement training that teaches us to implement the strategy? Should we reach out to you for guidance on implementing these goals in our own projects?
 - i. USGS already does this strategy really well. We're trying to get a shoulder behind it and push the strategy more and help people see the value in doing it. One way is reading and thinking about the strategy, also an implementation plan. There aren't official courses offered by us. But do reach out if you have suggestions.
- f. How do we stay up to date on this effort?
 - i. We are trying to improve our web presence. If you have an interest, please contact Zachary Bowen. We aim to have the strategy and body of work represented much better online.
- g. Are there lessons from the public lands-dominated Greater Sage Grouse effort that pertain to the private lands-dominated environment of the eastern U.S.?
 - i. A key component for success was involvement of State and local leaders in the design and implementation of our research and technical development work. Another related key was figuring out how to address state and other folk's concerns about data sharing and use. And a third was dedicated coordination, leadership, and support for the work from all the participating people and places involved across USGS.
- h. Is there a gap between research and operations in the USGS, and how do we fill it?
 - i. Have you seen good models of "research software engineering" in other orgs to try to follow?
 - 1. Absolutely. Software carpentry with the way it is teaching the basics of programming, packaging code, sharing importance of how code is composed. We are 'forced' as scientists to learn engineering practices and implement them in our work.
 - ii. What will it take to add the "E" engineering to R2O? In your opinion is this training, new staff, ??
 - 1. We have a convergence of things going on. The further development and movement of the EDG platform in USGS, for example. The EDG program is similar to the RGE program is an evaluation for staff – you're graded based on products produced (software, hardware).
 - iii. How will we be able to build on the lessons and initial solutions that you are trying out?
 - 1. In the midst of working on this. With some of these solutions, we're looking at third party platforms, just looking to clear them for use. Also working with science center staff on how centers could share developed components, etc.