CDI Monthly Meeting 20180613

CDI Monthly Meeting - 20180613

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.

Meeting Recording

Meeting recordings are available to CDI Members approximately 24 hours after the completion of the meeting. Please log in to view the recording. If you would like to become a member of CDI, email cdi@usgs.gov.

During the call, you can ask and up-vote questions at slido.com, event code #7988.

Agenda (in Eastern time)

11:00a **Scientist's Challenge** - **Data Sharing Agreements**, JC Nelson, USGS

11:10a Welcome - Cheryl Morris, Director of Core Science Analytics, Synthesis, and Library

11:15a Working Group Announcements [PDF]

11:25a **Amplifying USGS science with timely and digestible data visualizations** - Jordan Read, USGS

11:50a **Data visualization for science: comparing 3 dashboard building software packages -** Kevin Henry, Jason Sherba, Jeff Peters, USGS

12:30p Adjourn

Abstracts

Amplifying USGS science with timely and digestible data visualizations - Jordan Read, USGS

Our Nation's rapidly growing store of environmental data makes new demands on researchers: to take on increasingly broad-scale, societally-relevant analyses and to rapidly communicate findings to the public. Interactive web-based data visualizations now commonly supplement or form the core of mainstream journalism, and science journalism has followed suit. USGS can employ a similar approach to connect with new audiences, highlight key findings, and integrate multiple datasets and research projects into a story with broad public interest. This presentation will discuss collaboration and data challenges for creating visualizations consistent with USGS science standards, highlight several environmental data visualizations and their underlying technologies, and share social media outreach strategies designed to increase impact. Building greater capacity in data visualization will help serve the USGS's need to deliver timely relevant science to the public.

Jordan Read is committed to ensuring that the USGS takes advantage of recent technological advances and the exponential growth in data. Our agency's ability to extract insights from the modern-day deluge of data has become a requirement for remaining competitive. As the chief of the Water Mission Area's Data Science Branch, Jordan oversees a team that focuses on building reproducible data workflows, refining approaches to distill large/complex data into meaningful information (including data visualization), and teaching others data science techniques.

Data visualization for science: comparing 3 dashboard building software packages - Kevin Henry, Jason Sherba, Jeff Peters, USGS

Interactive dashboards are becoming an effective and popular format to display results of scientific analyses that can be difficult to interpret through static graphics. While dashboards were once costly and time consuming to develop, a set of new data visualization platforms offer the ability to quickly create interactive, shareable dashboards without extensive programming knowledge. In this presentation we discuss three dashboard platforms: Tableau, ArcGIS Online, and PowerBI, and their use in hazard exposure analysis. Each platform's pros and cons will be discussed, resulting in a clearer picture of where these different, but similar, platforms can fit into USGS science.

Presentations

Presentation: Slides are available to CDI Members. Please log in to download the slides. If you would like to become a member of CDI, email cdi@usgs.gov.

Highlights

Links from meeting announcements:

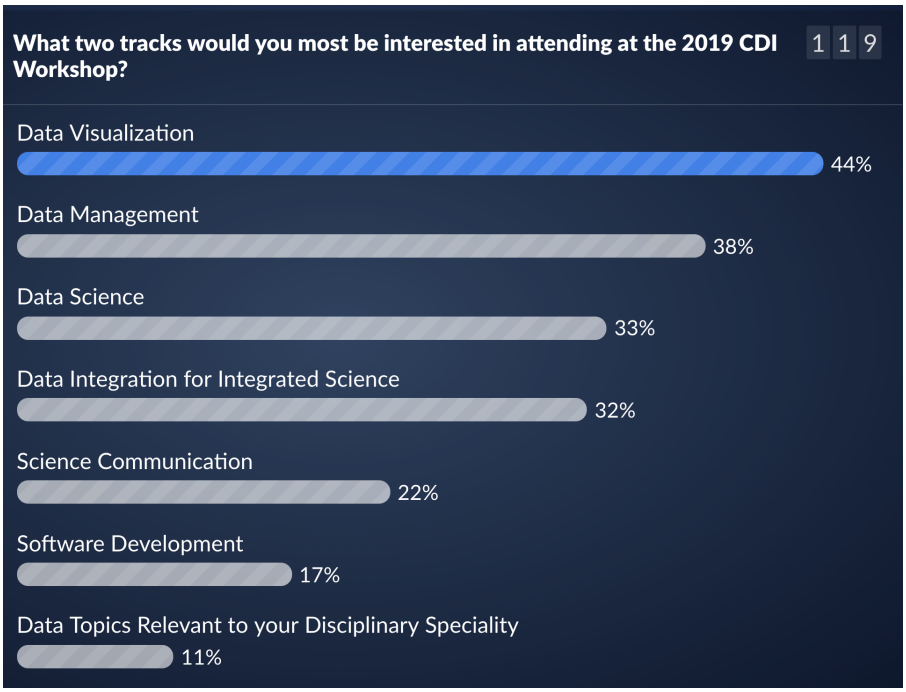
- Data Sharing Agreements Google Form: http://goo.gl/VtJC4S (intended for USGS users). Others can comment on the CDI form at https://my.usgs.gov/confluence/x/WYi8Iw

- Get your science used: https://pubs.er.usgs.gov/publication/cir1419

Q&A

Sli.do Polls

What two tracks would you most be interested in attending at the 2019 CDI Workshop?



What specific session topics would you like to see at the 2019 CDI Workshop? (multiple answers allowed)

- Sharing a list and how to use all the "free" resources available. IE: USGS AGOL...
- Dashboard technologies with greater interactivity and programming required
- Data Archive Strategies
- How to use data/science topics
- CDI Hackathon or Data Jam to put our discussions into practice and develop quick prototypes
- archiving our science (data, models, videos, presentations, other media)
- GeMS - new geologic map standard and trusted repositories
- Creating non-conventional visual representations of data (InfoGraphics)
- open source data science tools
- science collaboration tools (for collaborations inside and outside USGS/DOI)
- Strategies for designing visualizations and organizing data so it can be presented effectively
- presentations from the CMC on connecting USGS tools to understand more about our Bureau science activities
- delivering data in the wret environment
- cloud storage, data lakes/ data warehouses, API's, data governance for large data collections in the cloud, software IM, software management website.
- real time capabilities in dashboards, visualizations, and similar communications and information sharing solutions
- Strategies for working with and visualizing very large data using high performance computing and/or web services.
- Geoplatform vs. ScienceBase vs. Data.gov vs. Others... WCS Services vs. OpeNDAP services
- model/data fusion
- R and Shiny for beginner/intermediate - how to develop and serve interactive apps/visualizations Best practices for R, Data Management
- Communicating the robust variety of data sets that are available for use by researchers worldwide. Promotes reuse and citation.
- Data Science for Decision Makers
- Tools for integrated modeling and monitoring
- How do different types of datasets need different data management supports (including tools, workflow, documentation types, etc.)?
- real time data applications. GeoEvents Server.
- using yeti
- Hands on training for beginning coders
- GIS
- strategies for large data release
- slack or slack like tools
- QGIS
- ArcGIS Online
- Advancing science via the Cloud
- Working with other Federal agencies

- machine learning
- strategies for legacy data
- Cloud, HPC
- open source gis visualization
- Data and metadata standards
- Workflows
- Data munging principles and practices

Audience Questions from Zoom and Sli.do

(in progress)

1. How did you include the animated images in your Google slides?
    a. Jordan: GIFs are an image format that can be inserted into a file in the same way you put in a jpeg or png. I'm on a mac, and used quicktime player to do a screen recording, which generates a .mov file. From that .mov file, I converted to a bunch of .png files for each frame (and adjusted the image size) using the following resource: https://gist.github.com/tskaggs/6394639 which also has instructions for turning that series of images into a GIF. R and Python also have modules that make gifs from a series of plotting commands. Alternatively, the .mov or other video files can be uploaded to https://ezgif.com/video-to-gif to create a gif output, which is in many ways easier and doesn't need new libraries installed.
2. This is a comment rather than a question, but I think it's great that reproducibility is a priority. This is helpful for education purposes as well.
    a. Jordan: Thank you! Reproducibility is *the* top priority of our team, as it cuts across all of the work we do. If you haven't seen Alison Appling's talk on tools for reproducibility and are interested in hearing more, content is here for a DMWG talk: https://my.usgs.gov/confluence/display/cdi/DMWG+Meeting+2018_03_12
3. how r u coordinating social media for pub push? I co-manage the oregon acct & have not seen any requests 2 help on campaign
    a. On the water-use GIFs, the national water-use team will be emailing information on the gifs and the release schedule this week. Link will be provided next week.
4. Are there USGS resources/people for software with greater interactivity and programming? (plotly, Python, etc.)
5. Was R Shiny considered for data viz?
6. Marty Smith: Great Job guys, I'm wondering what the funding model is for something like the Hurrican Maria viz. I'm assuming the whole group was working on other projects and put them aside for this.
    a. Jordan S Read: Marty, yes, you are correct that we had to pause other projects to hop onto the hurricane data visualizations. That way of executing the work is fairly disruptive, but seems to be a necessary way of operating in the real-time extreme event space. Hurricanes are the only viz product we have done (to date) that we don't have a longer term planning arc and iteration for. Funding presently includes project-specific work, hosting, and re-usable template development (which make additional hurricane/storm visualizations cheaper).
7. Jordan, this is really awesome work and I'm glad to see your helping to innovative how we deliver USGS science in new ways. How do we spread this capability across the bureau? Do we need to do more training with existing scientists or hire more people that can do this?
    a. Thanks - I agree that we need to spread these skills more than we are currently doing. I think a combination of hiring and training is the best way to do that. The job market is favorable in this space if we have positions that have sustainable work needs. Training is something that I think would be best to coordinate across different mission areas.
8. Which version of Power BI was tested? And, is Power BI mobile friendly?

    a. Jason: I tested the Pro version.

Attendees

A Participant Report is available to CDI Members. Please log in to download the report. If you would like to become a member of CDI, email cdi@usgs.gov.