CDI Monthly Meeting 20200513

May 13, 2020: CDI Funded Projects - FAIR data, grassland productivity forecast, animal movement visualization

Meeting Recording and Slides

Recordings and slides are available to CDI Members approximately 24 hours after the completion of the meeting. These are the public slides. Log in as a CDI member to view ALL of the meeting resources, including recording.

If you would like to become a member of CDI, join at https://listserv.usgs.gov/mailman/listinfo/cdi-all.

Agenda (in Eastern time)

11:00 am Welcome and Opening Announcements
11:15 am Working Group Announcements
11:25 am Building a Roadmap for Making Data FAIR in the U.S. Geological Survey, Fran Lightsom, USGS
11:45 am Implementing a Grassland Productivity Forecast Tool for the U.S. Southwest, Sasha Reed, USGS
12:05 pm A generic web application to visualize and understand movements of tagged animals, Ben Letcher, USGS
12:30 pm Adjourn

Highlights

1. See blog post on this meeting here.
2. Questions and comments for follow-up with the CDI community
   a. Ben Letcher (bletcher@usgs.gov): who has experience with podcasts for telling science stories?
      i. Also DOI Creative Comms, a group that meets periodically to learn about new ways to communicate scientific information has an archive of presentations on their website at www.doi.gov/CreativeComms
   b. I'd like to further explore some of the tools that have been presented in the past few months together with other CDI members in a coordinated way. Possible?
   c. What are links between TAME and USGS data publication? (Kelsey, Emily (Emma))
4. In the future, there may be opportunities to help implement the FAIR roadmap for USGS data, if you are interested, contact cdi@usgs.gov.
5. The current GrassCast tool is at https://grasscast.unl.edu/, check out the introductory video that explains the project in less than 5 minutes. There are many opportunities for collaboration - including upcoming workshops and seminar series. Sasha Reed: screed@usgs.gov.

Notes

1. Welcome and Opening Announcements
   a. Leslie shared some tips for using Microsoft Teams
   b. Kevin Gallagher also shared an EarthMAP update. EarthMAP is launching two new teams that will help advance their efforts to develop the EarthMAP concept - the Capacity Assessment Team (CAT) and the Use-Case Development Team (UCDT). USGS employees can see more on the corresponding blog post, and the latest Need to Know from Monday, May 11.
c. Tim Quinn spoke about EarthMap and Integrated Modeling. Starting in two weeks, a series of webinars designed to introduce the USGS community to the EarthMAP concept will begin. The first webinar will have a focus on the role of integrated monitoring and modeling, along with a summary of the current initiatives going on in USGS.
d. More on EarthMAP for USGS employees:
   i. Blog post
   ii. Intranet webpage
   iii. EarthMAP Microsoft Team

2. Working Group Announcements

a. For more detail on all collaboration areas, see https://my.usgs.gov/confluence/x/yhv1I.
b. Tech Stack/ESIP IT&I
   i. Latest calls have focused on idea of a science gateway.
   ii. Next event, May 14 (tomorrow): CUASHI HydroShare Update
   iii. June webinar will be a quarantine addition: tools that you’ve found and been using while at home working remotely.
c. Usability
   i. This group alternates between town hall meetings and resource reviews (odd months = resource review; even = town hall meetings)
   ii. May’s resource review is on how usability and interface influence user experience, including credibility and use.
   iii. Last meeting was on the Paperwork Reduction Act and how to meet its requirements.
   iv. Link to Usability Collaboration Area’s Resource Reviews: https://my.usgs.gov/confluence/display/cdi/CDIU%3Ausability%3DResource%3DReviews
   v. Link to Usability Collaboration Area: https://my.usgs.gov/confluence/display/cdi/Usability+Collaboration+Area
d. Data Management Working Group
   i. Next event, June 8: Data Curation Network will talk about their data curation activities and data curation primers
   ii. Last meeting covered electronic records management
e. Semantic Web Working Group
   i. Next event, May 14 (tomorrow): The ESIP Agriculture and Climate cluster experiment to use concept maps to document workflows
   ii. Past events: discussion on implementing FAIR vocabularies and ontologies
f. Metadata
   i. Ongoing discussions on metadata for software and code
   ii. Past event: metadata for data publication vs for research publication
g. Risk
   i. Next event, May 21: Panel discussion on communicating hazard and risk science
   ii. Our annual face-to-face meeting will be virtual, August 11-13.
h. eDNA cop
   i. Revived group
   ii. Next event: Re-starting group in the next month or so (June or early July)
   iii. eDNA page and link to spreadsheet for membership: https://my.usgs.gov/confluence/display/cdi/ETWG%3DeDNA%3DCommunity%3Dof%3DPractice
i. Software Development
   i. Last event: covered fastAPI and Flask
   ii. Next event: data warehouse basics and ETL pipelines
j. Open Innovation
   i. Next events: Paperwork reduction act and indigenous observation network
   ii. Open Innovation (OI) Community Resources:
      - USGS Open Innovation Teams
      - Open Innovation Community Wiki
      - Ignite OI Forums Stream Channel
      - April 2020 OI Newsletter


a. Opportunity: using FAIR principles to improve value and usefulness of USGS products
   i. FAIR - findable, accessible, interoperable, reusable
b. Project focused on creating a roadmap to enable FAIR
c. Project activities
   i. Conducted a survey of data producers
   ii. Collected use cases of projects that integrate data
   iii. Hosted a workshop September 9-11, 2019
   iv. Drafted report & list of recommendations
d. Challenges
   i. Difficulty scheduling meeting
   ii. Difficulty collaborating in online documents in group not exclusively DOI
   iii. An abundance of good ideas
e. Recommendations
   i. 100 discrete recommendations from workshop
   ii. Open science requires extension of FAIR beyond data to samples, methods, software, and tools
   iii. USGS can support research teams in creating FAIR data and metadata, as well as enterprise systems that maintain those products' fairness
f. Recommendations
   i. 14 essential
   ii. 38 important
   iii. 44 useful
g. Implementing recommendations would be responsibility of many groups
   i. CDI: may be able to help by creating a coordination council that includes representatives from other groups
h. Further objectives:
   i. Increasing use of globally unique/persistent identifiers - there is more we can do with physical samples and software
   ii. Policy development
   iii. Researching best practices
iii. Soil brightness
iv. Evergreen plants

Tea et al. found that SIF was noisy and was not a better metric; however, NiRv and SATVI did a good job. There is some promise in SIF for capturing the timing of the growing season - could be a good proxy to include.

j. Plan now to incorporate these data into the current Grass-Cast. Southwest button will be available soon.

k. Next Steps
i. Ultimately, team wants to integrate across these different methods; going beyond Arizona and New Mexico
ii. There is lots of room for collaboration; stay tuned for upcoming workshops and seminars

5. A generic web application to visualize and understand movements of tagged animals, Ben Letcher, USGS

a. Problem
i. Tons of tracking/tagging data
   1. The Movebank database has 2 billion observations
ii. Data collection is expensive and takes lots of time & effort
iii. Project team wants to maximize the value of this data and make it easier for people interact with these complex types of data

b. What we did
i. Made the TAME web application, a data exploration tool.
ii. Goals: easy to use; interactive, explorable
iii. Based everything on open source libraries
iv. Excited about developing movement stories - plans to develop a podcast on this

c. Features of TAME
i. 4 intro videos
ii. User account system - can upload own data
   1. Not visible publicly, with an option to publish
iii. Map observations to: color, size, or outline
iv. Can select individuals or select by area
v. Multiple area selections also possible
vi. Cross filters: filter any one variable; can make a movie/time series of the data

d. Demo
i. See recording for demonstration.

Questions and Answers

1. Comment from sli.do:
If a registry is created I would encourage Darwin Core be listed as one of the standards - specifically for biological data. I can help with drafting text.

2. Question from sli.do:
Are the recommendations generated based on existing recommendations created by the community or are they specific to USGS?
   a. Fran: We started with the community's concept of FAIR. Found ourselves moving toward how can we enable this in USGS, so that I think i would say that the goals and methods that would get us there for making things FAIR are very much international in the scientific community but the specific recommendations are focused on USGS organizational structure.
3. Question from Sli.do:
   Nice project! The DayCent model produces productivity estimates on its own. Are the satellite data used to constrain the forecast DayCent model parameters?
   a. Sasha: Yes. ... Daycent platform brings in different data streams.
4. Question from Sli.do: Are there differences in the stakeholders in the Great Plains vs Southwest that you need to take into account? (you mentioned environmental differences)
   a. Sasha: Big difference in stakeholders. Great plains - lots of private lands and ranching; in Southwest, stakeholders are mainly public land managers.
5. Have you looked at how productivity estimates match with dust on snow loading in the San Juan Mountains? Could have interesting implications for water managers in the area.
   a. Sasha: AWESOME question! We haven't looked at links yet between productivity and dust on snow, but that would be such a cool thing to do and could, as you suggest, inform water management options!
6. Question from Sli.do: Sorry if I missed it, what format does the data need to be in?
   a. Data format for TAME is CSV. Details are in the upload data section or in the upload video.
7. Ben, I'm wondering if you have had any interactions with the biologging community about this? I know Sarah Davidson with Movebank and if it made sense I could connect you. I also have contacts in Ocean Tracking Network (https://oceantrackingnetwork.org/) and the Animal Telemetry Network (https://ioos.noaa.gov/project/atn/) and Australia’s IMOS.
   a. Ben: If anyone has good ideas with how to interact with them, let me know. The hope is that we can get the attention of these people and have them use this application.
8. Question from Sli.do:
   I'd like to further explore some of the tools that have been presented in the past few months together with other CDI members in a coordinated way. Possible?
9. Thank you Ben! This is a really cool app. Can the TAME app be used as a way to publish USGS tracking data? My team and I at WERC struggle with how best to publish our data (per USGS requirements) but in a way that is accessible to the public.
   a. Kelsey, Emily (Emma) I love that idea. CDI is the perfect place to get feedback on requirements for publishing data. Right now, users include a citation, either to a paper or database, but I suppose this could be expanded. Happy to talk with anyone more who's interested.
10. Is TAME open source or in the USGS GitLab?
    a. Difficulty figuring out approval system. Learning how to do an official software release.
11. Can you link to the Github repo?
    a. https://github.com/walkerjeffd/sheds-tame
12. Any interest/ability in being able to direct upload from MoveBank itself?
    a. Yeah, I skipped that bullet point in the last slide. I think it would be easy to set up. Do you have a contact who would be interested?