

June 2020 Collaboration Area Activity

Artificial Intelligence/Machine Learning, 6/9 - Tallgrass Supercomputer for AI/ML

Natalya Rapstine presented "USGS Tallgrass Supercomputer 101 for AI/ML," an overview of the new USGS Tallgrass supercomputer designed to support machine learning and deep learning workflows at scale, and deep learning software and tools for data science workflows. Natalya's slides covered the software stack that supports Deep Learning, including PyTorch, Keras, and TensorFlow. She then illustrated the capabilities with the "Hello World!" example of Deep Learning - the MNIST Database of Handwritten Digits.

See many more links to resources in the Slides and recording available at [AI/ML Meeting Notes](#)



ASCII art for the Tallgrass supercomputer

Data Management, 6/8 - Data Curation Network - extending the research data toolkit

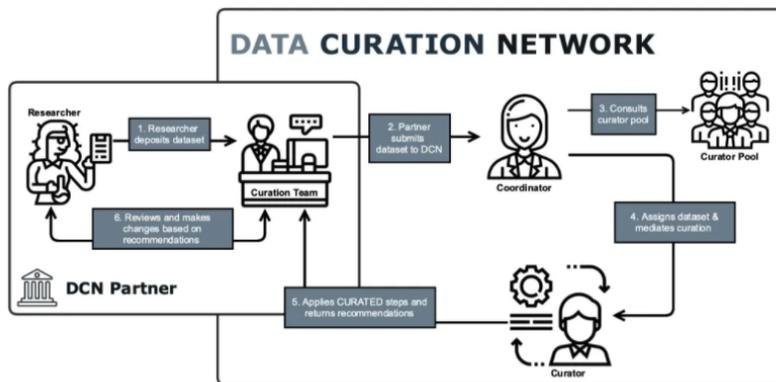
Guests from the Data Curation Network, Lisa Johnston, Wendy Kozlowski, Hannah Hadley, and Liza Coburn presented on their recent work.

CURATED stands for: **C**heck files/code; **U**nderstand the data; **R**quest missing info or changes; **A**ugment metadata; **T**ransform file formats, **E**valuate for FAIRness; **D**ocument curation activities.

Checklists and primers related to these topics for specific file formats are available at: <https://datacurationnetwork.org/resources/>

Also of interest is an Excel Archival Tool, which programmatically converts Microsoft Excel files into open-source formats suitable for long-term archival, including .csv, .png, .txt, and .html: <https://github.com/mcgrory/ExcelArchivalTool>

The DCN Curation Workflow



Data Curation Network infographic at <https://datacurationnetwork.org/resources/>

DevOps, 6/2 - Elevation data processing at scale

Josh Trahern, Project Manager of the NGTOC Elevation Systems Team led a discussion titled "Elevation Data Processing At Scale - Deploying Open Source GeoTools Using Docker, Kubernetes and Jenkins CI/CD Pipelines"

The presentation highlighted the Lev8 (pronounced as "elevate" and doing petabyte scale processing of DEMs) & QCR Web Applications, produced by the Elevation team. These tools are used by Production Ops to generate the National Elevation Dataset (NED). The NED dataset is a compilation of data from a variety of existing high-precision datasets such as LiDAR data, contour maps, USGS DEM collection, SRTM, and other sources which are combined into a seamless dataset, designed to cover all the United States territory in its continuity.

Moving away from proprietary software and owning the code base – to prevent trying to fit a square peg into a round hole. Working toward 100% automation, 100% documentation and moving to Linux environment. Making all of these changes while the system was operational.

See the recording on the [DevOps Meetings page](#).

Fire Science, 6/16 - NLCD Rangeland Fractional Component Time-Series: Development and Applications

Fire Update: Paul Steblein gave a Fire Update and Matthew Rigge, (EROS) – presented on "NLCD Rangeland Fractional Component Time-Series: Development and Applications."

The Fire Science coordinators and CDI staff are working on syncing content on the internal OneDrive and the [CDI wiki](#), contact Paul at psteblein@usgs.gov if you have any questions about the group.

Metadata Reviewers, 6/1 - What is important to metadata reviewers

The Metadata Reviewers group had a discussion about what matters to them when reviewing metadata. Some themes were making USGS data as findable and reusable as possible, avoiding unnecessary complexity, and making metadata easier to write.

See more notes on the discussion at their [Meetings wiki page](#).

Open Innovation, 6/18 and 6/19 - Paperwork reduction and Community-based water quality monitoring

On June 18, the topic was "Tackling the Paperwork Reduction Act (PRA) in the Age of Social Media and Web-based Interactive Technology." Three Information Collection Clearance Officers from DOI (Jeff Parrillo), USGS (James Sayer), and FWS (Madonna Baucum) explained the basics of the Paperwork Reduction Act (PRA), discussed how the PRA applies to crowdsourcing, citizen science, and prize competition activities, and participated in a Q&A discussion with the audience. More information on the [Open Innovation wiki](#).

On June 19, Ryan Toohey and Nicole Herman-Mercer presented on "Indigenous Observation Network (ION): Community-Based Water Quality Monitoring Project." ION, a community-based project, was initiated by the Yukon River Inter-Tribal Watershed Council (YRITWC) and USGS. Capitalizing on existing USGS monitoring and research infrastructure and supplementing USGS collected data, ION investigates changes in surface water geochemistry and active layer dynamics throughout the Yukon River Basin. More information on the [Open Innovation wiki](#).

Risk, 6/18 - Funded Project Reports

This was "round 1" of final project presentations from the [FY19 Risk RFP awardees](#). Please see the list below for presenters - each one is about 10-12 minutes in length. Pls from each project provided a project overview, a description of their team, accomplishments, deliverables, and lessons learned.

- *Quantifying Rock Fall Hazard and Risk to Roadways in National Parks: Yosemite National Park Pilot Project*, Brian Collins, Geology, Minerals, Energy, and Geophysics Science Center
- *The State of Our Coasts: Coastal Change Hazards Stakeholder Engagement & User Need Assessment*, Juliette Finzi-Hart, Pacific Coastal and Marine Science Center
- *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**, Dan Grear, National Wildlife Health Center
- *Communications of risk - uranium in groundwater in northeastern Washington state*, Sue Kahle, Washington Water Science Center

See more at the [Risk community of practice wiki page](#).

Tech Stack, 6/11 - ESIP Collaboration Infrastructure 2.0

In June the joint Tech Stack and ESIP IT&I meeting hosted three presentations

[Ike Hecht](#), [WikiWorks](#) on the ESIP Wiki. Mediawiki upgrade from v1.19 to 1.34 of the ESIP wiki

[Lucas Cioffi](#), [QiQoChat](#) lead developer on the technical side of QiQoChat. Utilizing QiQoChat to bring together our asynchronous workspaces with our virtual conferences and meetings

[Sheila Rabun](#), ORCID US Community Specialist on the ORCID API. Becoming an ORCID member to gain access to ORCID API keys to integrate ORCID authentication into the wiki.

[See more at the IT&I meetings page](#)

Software Dev, 6/25 - Serverless!

Carl Schroedl presented on "Using Serverless and GitLab CI/CD to Continuously Deliver AWS Step Functions." See: <https://aws.amazon.com/lambda>

Notes and more links:

- Available as of 6/18: <https://aws.amazon.com/blogs/compute/using-amazon-efs-for-aws-lambda-in-your-serverless-applications/>

- Check out saml2aws on [github.com](https://github.com/Versent/saml2aws) by Versent to authenticate and be able to run commands that tools like AWS CLI would enable: <https://github.com/Versent/saml2aws>
- Using GitLab CI (continuous integration) runners, you can use shared runners from CHS or you can set up your own · AWS best practices <https://github.com/awslabs/aws-solutions-constructs>

--

[All CDI Blog Posts](#)