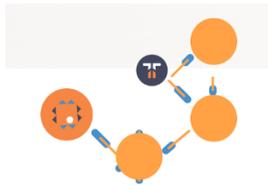
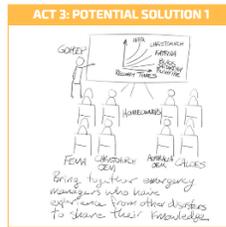


# April 2020 Collaboration Area Activity

The CDI Collaboration Areas are keeping me busy. You can get to all of these groups and sign up for mailing lists on the [CDI Collaboration Area wiki page](#).



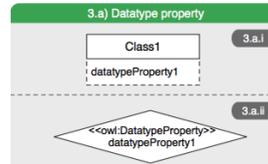
What is managed open source?



Training Data



Fast Track



From upper left corner, clockwise: DevOps: image from Tidelift website; SoftwareDev: logo for uvicorn; Risk: Impact360 worksheet; AI/ML: image from AI /ML DELTA presentation; Semantic Web: image from Garillo and Poveda-Villalon; Open Innovation: image from OI wiki page; Tech Stack: image from Unidata gateway webpage; Usability: image from Sayer's Paperwork Reduction Act presentation

- 4/6 Metadata Reviewers - revision or release information in titles
- 4/7 DevOps - managed open source with Tidelift
- 4/9 Semantic Web - implementing FAIR vocabularies and ontologies
- 4/9 Tech Stack - Unidata Science Gateway
- 4/13 CDI Data Management - changes to the USGS Science Data Catalog
- 4/14 Artificial Intelligence / Machine Learning - fine scale mapping of water features at the national scale
- 4/15 Usability - how the Paperwork Reduction Act affects usability studies
- 4/16 Risk - Product evaluation/testing and integrating solutions into strategy
- 4/17 Ignite Open Innovation - Open Innovation and COVID-19
- 4/21 Fire Science - stakeholder input on USGS Fire Science
- 4/23 Software Dev - FastAPI

## 4/6 Metadata Reviewers - revision or release information in titles

In April the Metadata Reviewers group dove into a question about including the date of a revision or release in the title of the data release. Doing so would help to distinguish between different versions of a data release. After much discussion the group concluded that **two metadata records should not have the same title in their citation elements**.

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

## 4/7 DevOps - managed open source with Tidelift

The DevOps group heard a presentation from [Tidelift](#). Tidelift partners with open source maintainers in order to support application development teams. This **saves time and reduces risk when using open source packages to build applications**.

See the recording and slides on the [DevOps Meeting page](#). If you are interested in using Tidelift for a USGS application, get in touch with Derek Masaki at [dmasaki@usgs.gov](mailto:dmasaki@usgs.gov). If you'd like a presentation from Tidelift, contact Melanie Gonglach at [melanie@tidelift.com](mailto:melanie@tidelift.com).

## 4/9 Semantic Web - implementing FAIR vocabularies and ontologies

The group discussed "[Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web](#)" by Daniel Garijo and Mara Poveda-Villalon. The discussion focused on sections 2 and 3 of the paper, URIs (uniform resource identifiers) and Documentation. The group recognized that **implementation of the best practices in the paper (for example, stable, permanent identifiers) would depend not only on semantic specialists, but also those who set policy for the USGS network**. This point was communicated to the group that is working on enabling FAIR practices in the USGS.

See more at the [Semantic Web meetings page](#).

## 4/9 Tech Stack - Unidata Science Gateway

Julien Chastang presented on the Unidata Science Gateway (<https://science-gateway.unidata.ucar.edu/>) Unidata is exploring **cloud computing technologies in the context of accessing, analyzing, and visualizing geoscience data**. From the abstract: "With the aid of open-source cloud computing projects such as OpenStack, Docker, and JupyterHub, we deploy a variety of scientific computing resources on Jetstream for our scientific community. These systems can be leveraged with data-proximate Jupyter notebooks, and remote visualization clients such as the Unidata Integrated Data Viewer (IDV) and AWIPS CAVE."

Slides and recording on the joint CDI Tech Stack and [ESIP IT&I webinars on the ESIP page](#).

## 4/13 CDI Data Management - changes to the USGS Science Data Catalog

Lisa Zolly presented on changes coming with the USGS Science Data Catalog version 3. Today, the Science Data Catalog (<https://data.usgs.gov/>) has more than 21,000 metadata records. In order to serve its human and machine stakeholders, **a number of changes are planned in order to address the changing landscape of federal data policy, substantial growth of the catalog, improvement of workflows, improvement of usability, and more robust reporting and metrics**.

Slides and recording are posted at the [meeting wiki page](#).

## 4/14 Artificial Intelligence / Machine Learning - fine scale mapping of water features at the national scale

Jack Eggleston (USGS), John Stock (USGS), and Michael Furlong (NASA) presented on "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." The **availability of high-resolution satellite imagery, combined with machine learning analysis to rapidly process the satellite imagery**, provides the USGS with a new capability to map natural resources at the national scale.

The recording is posted at the [meeting wiki page](#).

## 4/15 Usability - how the Paperwork Reduction Act affects usability studies

James Sayer presented on the Paperwork Reduction Act (PRA) and Usability Testing. The PRA is designed to protect the public from inappropriate data collection. **All agencies have their own PRA procedures, so implementation in other agencies won't necessarily translate to USGS implementation**. James reviewed Fast Track procedures and exclusions. His advice included to start early in thinking about PRA in your usability work, and to talk to your ICCO (Information Collection Clearance Officer) if you have any questions.

The slides, notes, and recording are posted on the [meeting wiki page](#). Do you have more questions? Contact James at [jsayer@usgs.gov](mailto:jsayer@usgs.gov).

## 4/16 Risk - Product evaluation/testing and integrating solutions into strategy

The Risk Community of Practice April meeting was part 3 of a series of training webinars provided by [Impact360 Alliance](#) on human-centered design thinking and inclusive problem solving. Emphasis was given to the tools for **product evaluation/testing ("[Re]Solve") and integrating solutions into strategy ("[Re]Integrate")**. Worksheets were provided to "Create and Test a Solution in Three Acts." A follow-up session on April 23 discussed examples of the worksheets.

Access the slides and recording, and handouts at the [Risk Meetings page](#) (must log in as a CDI member, [join here](#) if you're not a member yet).

## 4/17 Ignite Open Innovation - Open Innovation and COVID-19

April was Citizen Science Month! At the Open Innovation meeting, Sophia B Liu (USGS Open Innovation Lead) provided an **overview of the various open innovation efforts inside and outside of government that have emerged in response to COVID-19**. She also discussed The Opportunity Project Earth Sprint and proposed Problem Statements.

See more information and list of COVID-19 sites at the [meeting wiki page](#).

## 4/21 Fire Science - stakeholder input on USGS Fire Science

James Meldrum and Ned Molder of the USGS Fort Collins Science Center presented on **Analysis of stakeholder input on USGS fire science communication and outreach, science priorities, and critical science needs**. The group also heard updates on the USGS Fire Science strategy, recent fire activity, and held a discussion on "How is Covid 19 affecting your fire science"?

Contact Paul Steblein ([psteblein@usgs.gov](mailto:psteblein@usgs.gov)) or Rachel Loehman ([rloehman@usgs.gov](mailto:rloehman@usgs.gov)) for more information.

## 4/23 Software Dev - FastAPI

The Software Dev cluster had Brandon Serna and Jeremy Fee present about their **work using FastAPI with some comparisons to Flask**. I am not a developer so I will summarize by pasting some links, tag lines, and interesting things I heard.

Recommended resources.

- <https://github.com/tiangolo/fastapi>

- <https://github.com/tiangolo/full-stack-fastapi-couchbase>

I'm going to take a little bit of space to list some of the things I Googled while listening to this call, because to me these descriptions (and some of the logos) are fascinating. It would be fun to do a tagline-logo-name matching game.

1. FastAPI, <https://fastapi.tiangolo.com/>: FastAPI framework, high performance, **easy to learn, fast to code, ready for production**
2. Flask: <https://flask.palletsprojects.com/en/1.1.x/>: **web development, one drop at a time**
3. Hot reloading <- this sounds very exciting, and according to the internet it is "The idea behind hot reloading is to keep the app running and to inject new versions of the files that you edited at runtime. This way, you don't lose any of your state which is especially useful if you are tweaking the UI"
4. Uvicorn: <https://www.uvicorn.org/>: **The lightning-fast ASGI server**
5. Cookiecutter <https://cookiecutter.readthedocs.io/en/1.7.2/>: **Better Project Templates**
6. Gunicorn: <https://gunicorn.org/>: Gunicorn 'Green Unicorn' is a Python WSGI HTTP Server for UNIX. It's a pre-fork worker model. The Gunicorn server is broadly compatible with various web frameworks, **simply implemented, light on server resources, and fairly speedy**
7. Pyenv: <https://github.com/pyenv/pyenv>: pyenv lets you easily switch between multiple versions of Python. It's **simple, unobtrusive, and follows the UNIX tradition of single-purpose tools that do one thing well**
8. Pipenv: <https://pipenv-fork.readthedocs.io/en/latest/>: Pipenv is a tool that aims to bring the best of all packaging worlds (bundler, composer, npm, cargo, yarn, etc.) to the Python world. **Windows is a first-class citizen, in our world**
9. Hypercorn: <https://pgjones.gitlab.io/hypercorn/>: Hypercorn is an **ASGI** web server based on the sans-io hyper, **h11**, **h2**, and **wsproto** libraries and **inspired by Gunicorn**

See more at the Software Dev [wiki meetings page](#).

--

[More CDI Blog Posts](#)