Data to science: an open-source online platform for managing, visualizing, and publishing UAS data

Presented by Ben Hancock

SPIE Conference 13053
April 22, 2024

PURDUE UNIVERSITY

SPIE.

Motivation

- Researchers today are capturing large volumes of UAS HTP at a high spatial and temporal resolution
- Research can be impeded when we don’t know what data we have or how to access it
- How can we store, process, visualize, and collaborate findings with other researchers across disciplines?

Building an open-source online platform

One-stop shop researchers can use to…

Manage	Process	Visualize	Collaborate	Publish
Organize uploaded datasets and make accessing data easy from the D2S platform.	Convert datasets to cloud optimized formats and provide tools for deriving new data products.	Visualize raster and point cloud datasets on interactive maps and viewers.	Create teams and manage access to projects with simple controls. Share findings with other users on the platform.	Publish data products to a public catalog for anyone to discover.

Major Components

The application instance and public central catalog make up the D2S platform.

Application Instance

Containerized full-stack web application

- React web application
- FastAPI backend
- PostgreSQL database
- NGINX web server
- Celery and Redis

Public Central Catalog

Searchable catalog of published datasets from application instances

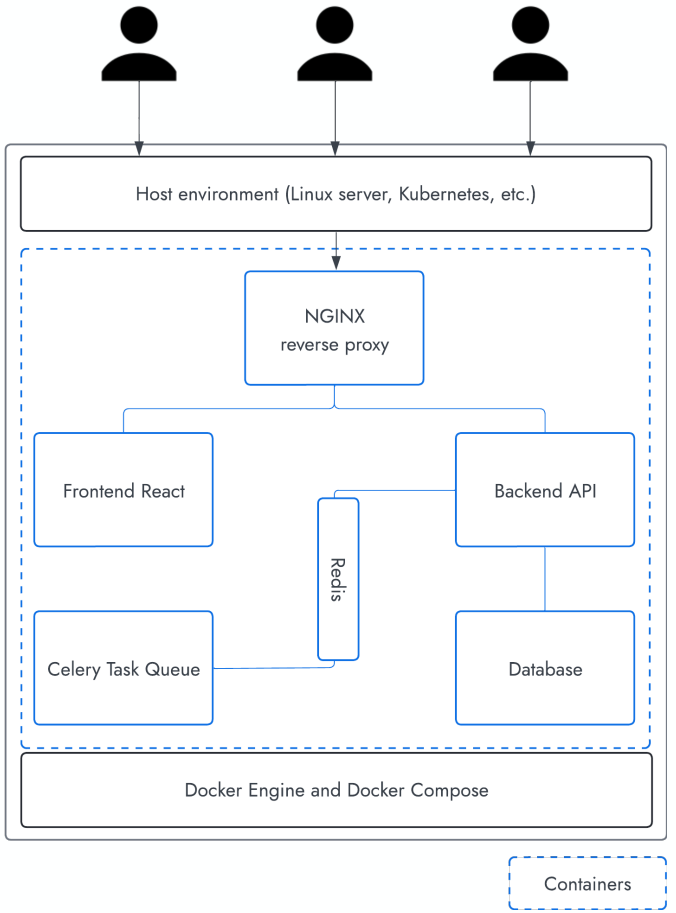
- STAC Browser
- STAC FastAPI w/ pgstac

PURDUE UNIVERSITY

Application instance

Open-source, anyone can run an application instance.

- Multi-container application
  - Each container runs a single service that supports the application
- Easy deployment with Docker Compose
- Can be hosted from a variety of environments
- Data always remains inside application instance

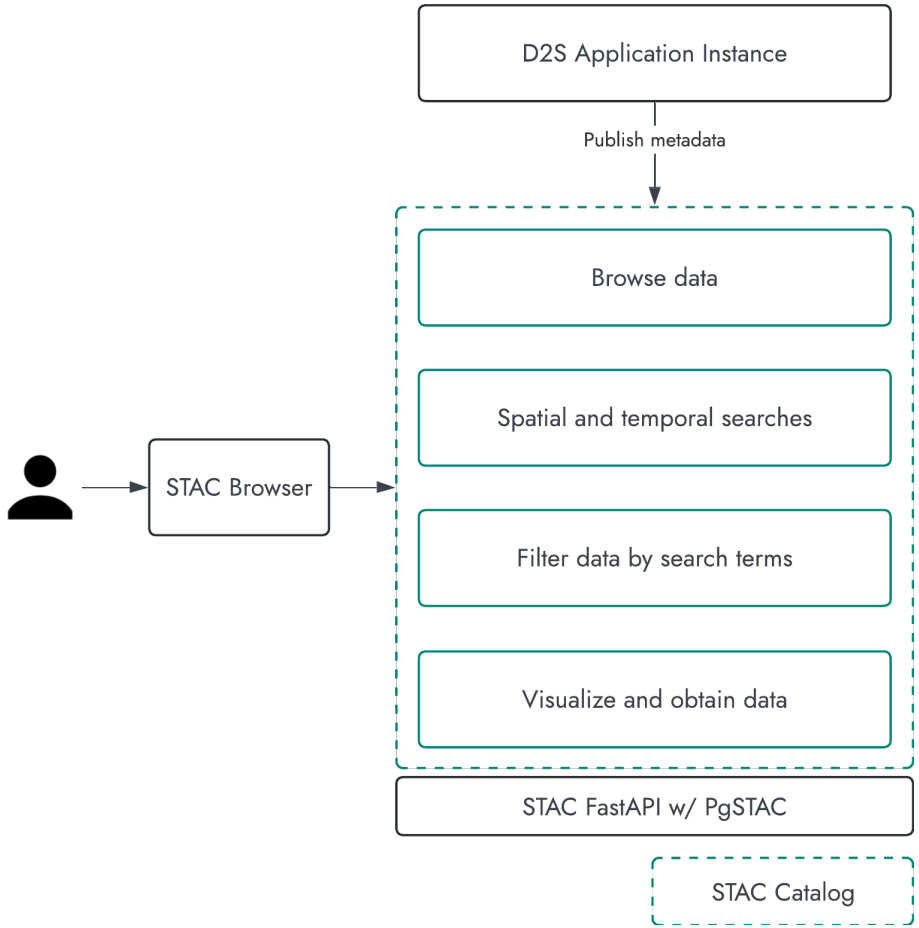


PURDUE UNIVERSITY

Public Central Catalog

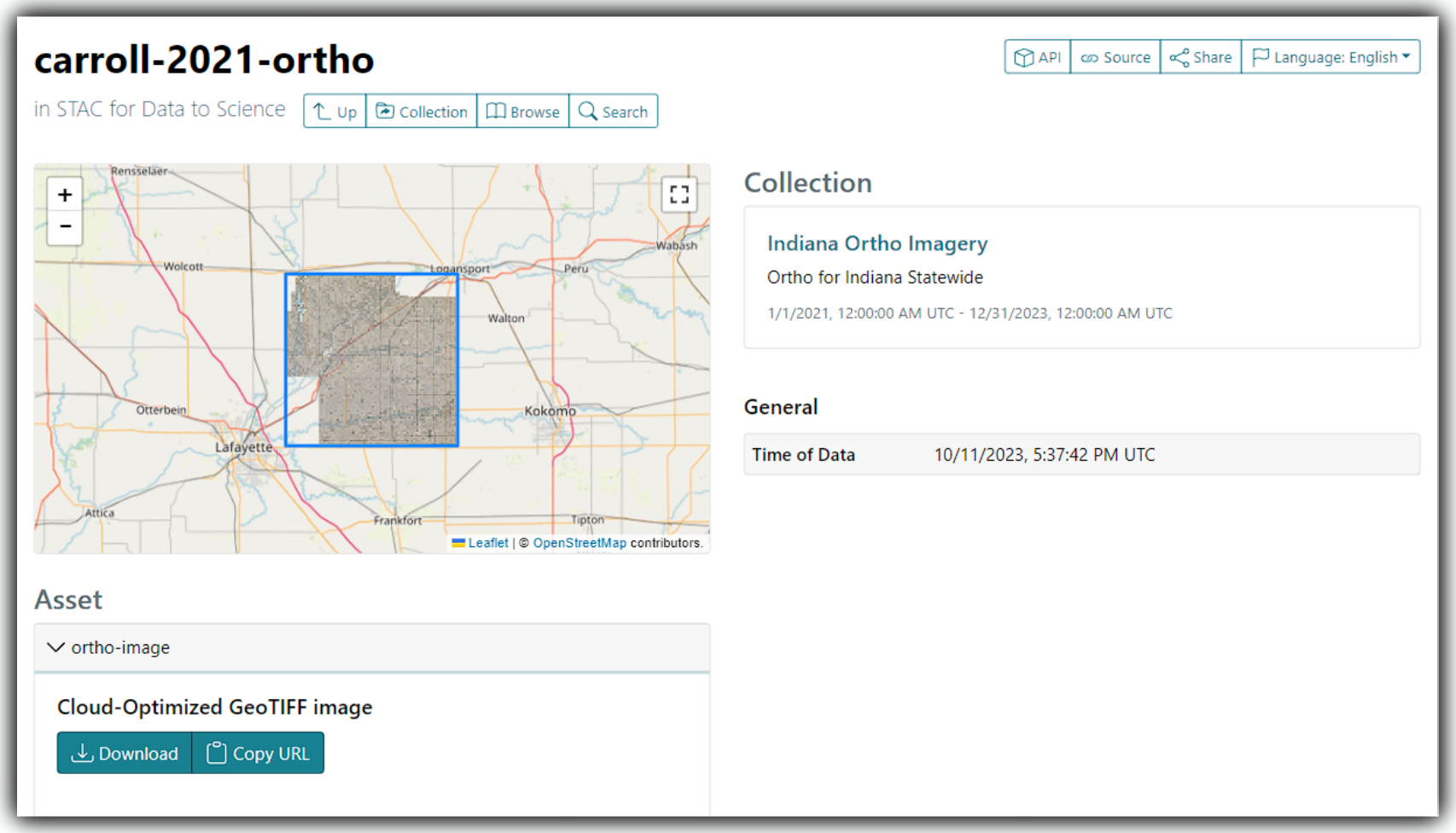
Publish metadata to public catalog to share with a broader audience.

- Built on top of the **S**patio**T**emporal **A**sset **C**atalog (STAC) specification
  - Common set of metadata fields for geospatial data
  - Easy to extend with custom metadata fields
  - Interoperable with other geospatial standards
  - Large collection of community-built tools and extensions
- **Only metadata** will be stored in the catalog
- Physical dataset remains in application instance



PURDUE UNIVERSITY

STAC for Data to Science

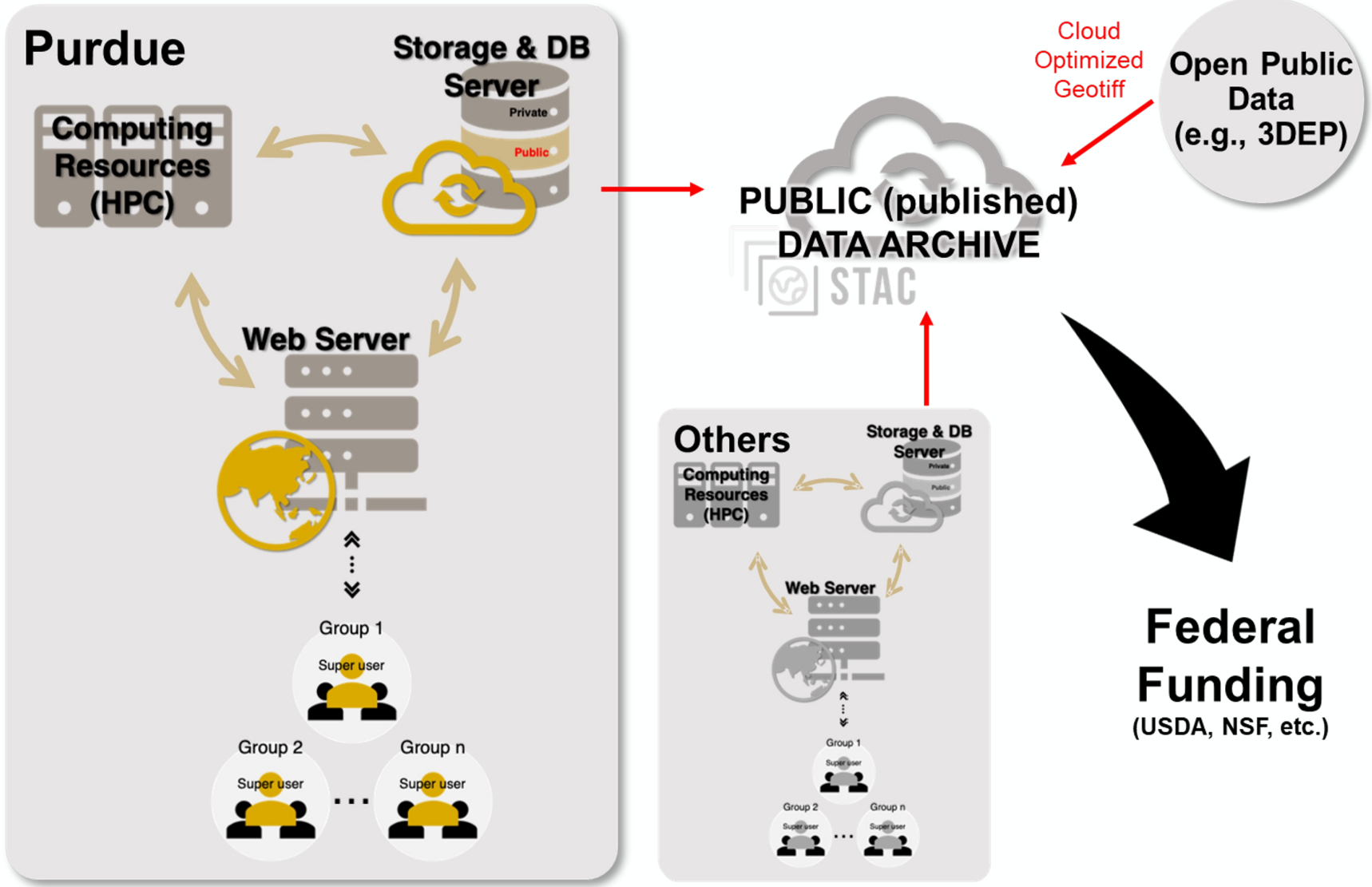


PURDUE UNIVERSITY

GOAL

“To make UAS HTP data FAIR and an online platform SCALABLE”

Findable Accessible Interoperable Reusable




Welcome to Data to Science
A Data-driven Open Science Community for Sustained Innovation
Sign up
Sign in

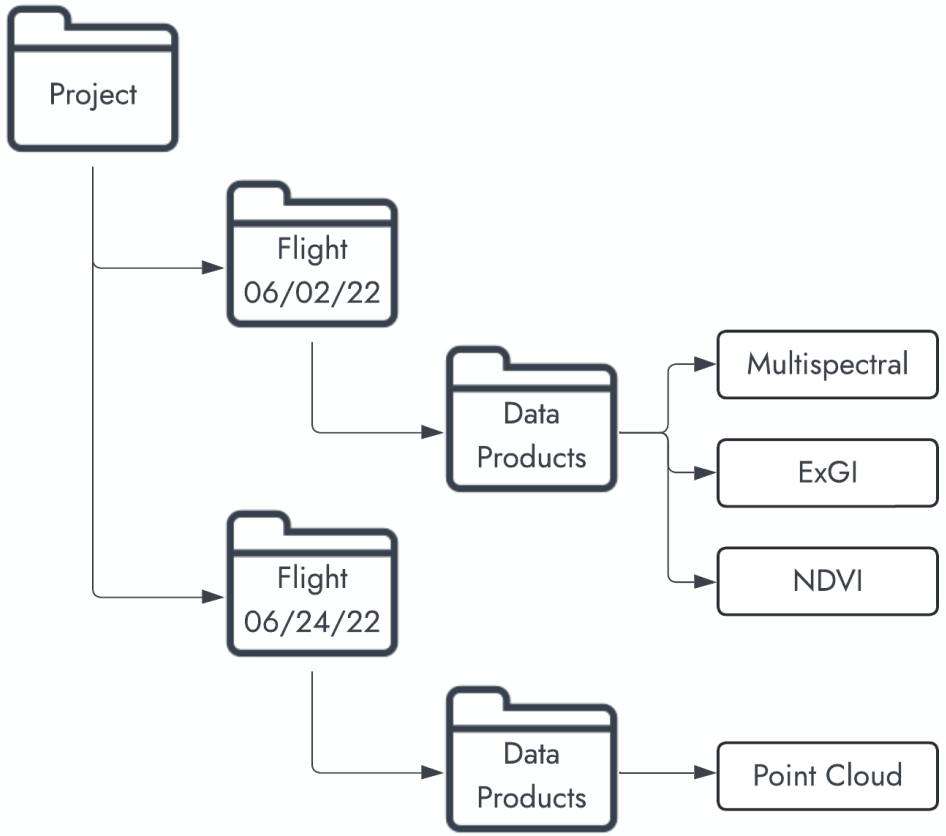
Data to Science Overview

Overview of main pages and how D2S organizes data.

- Main pages: Homepage (Map), My Teams, and Workspace (Projects)
  - Homepage – Visualize projects and data products
  - My Teams – Create, join, manage teams
  - Workspace – Create and view projects

- Project is the base level for organizing data
  - A single project may contain multiple flights
    - Data products are associated with a single flight

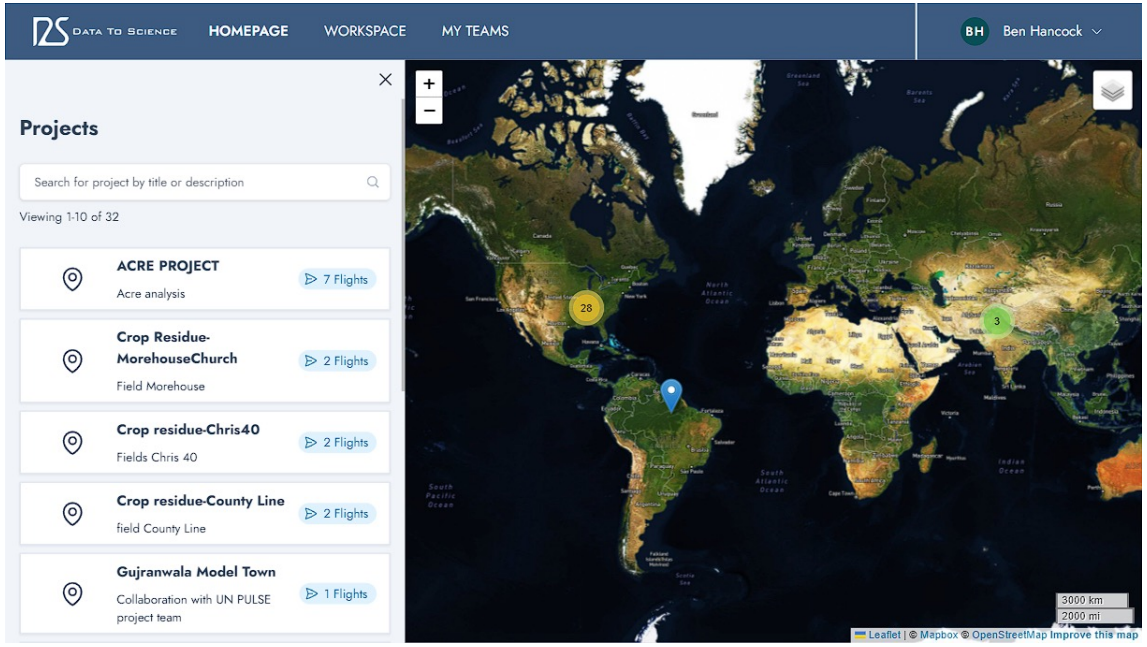
- Flights might be for different dates, different sensors, etc.

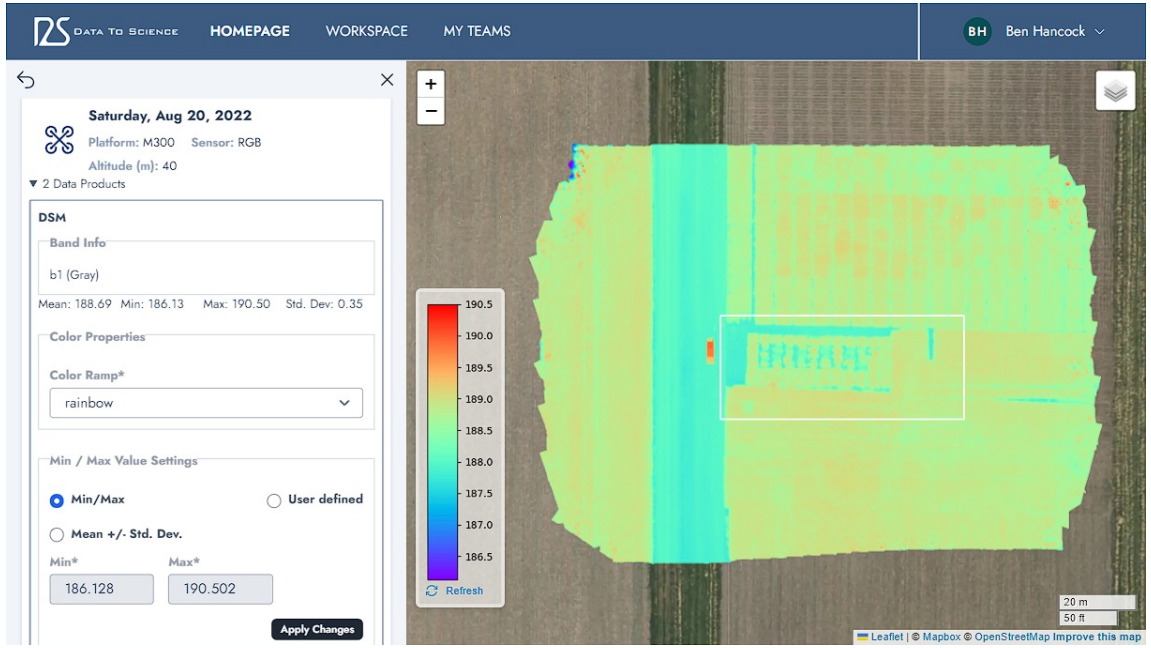


PURDUE UNIVERSITY

D2S Homepage

View project locations and individual data products on map.

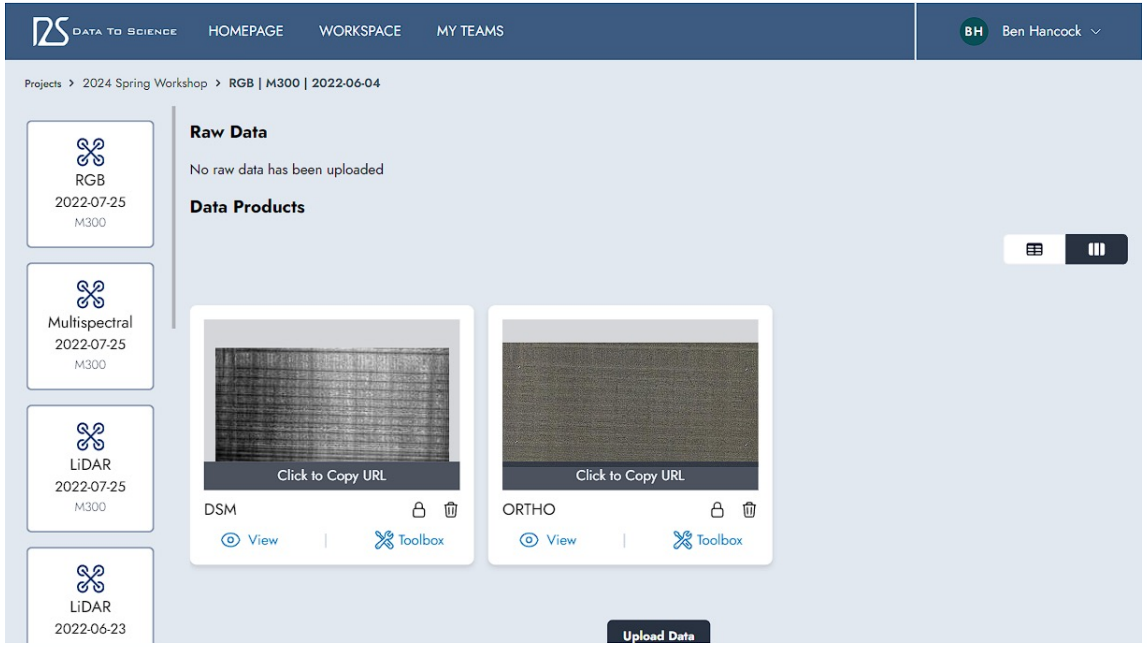


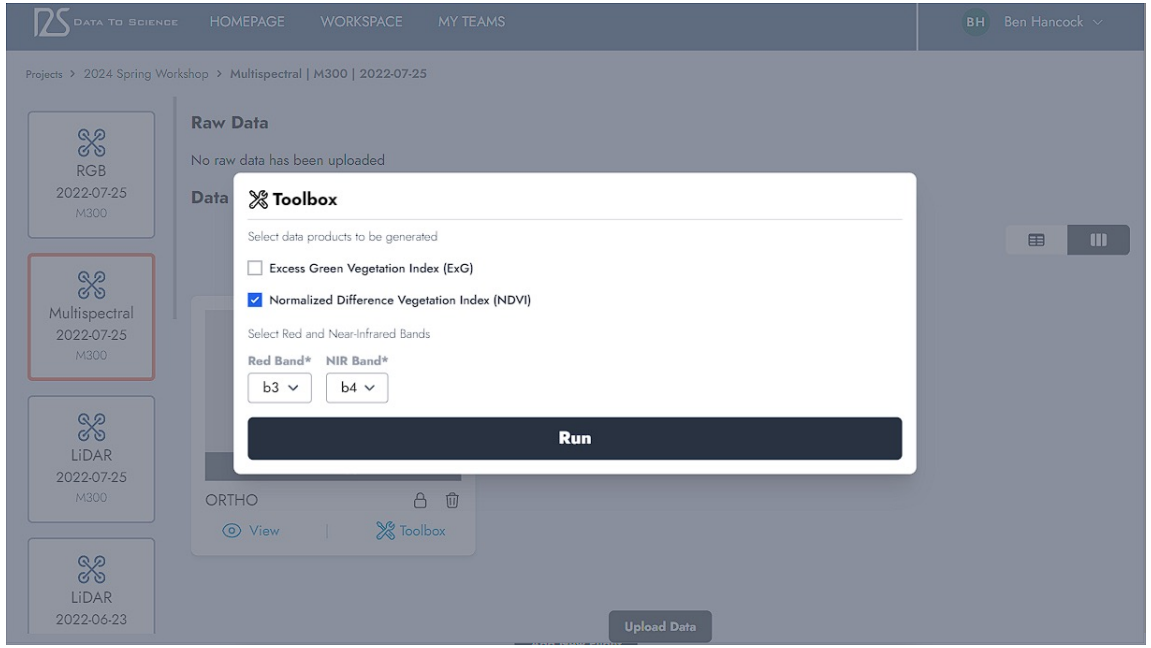


PURDUE UNIVERSITY

D2S Workspace

Manage projects, flights and data.





PURDUE UNIVERSITY

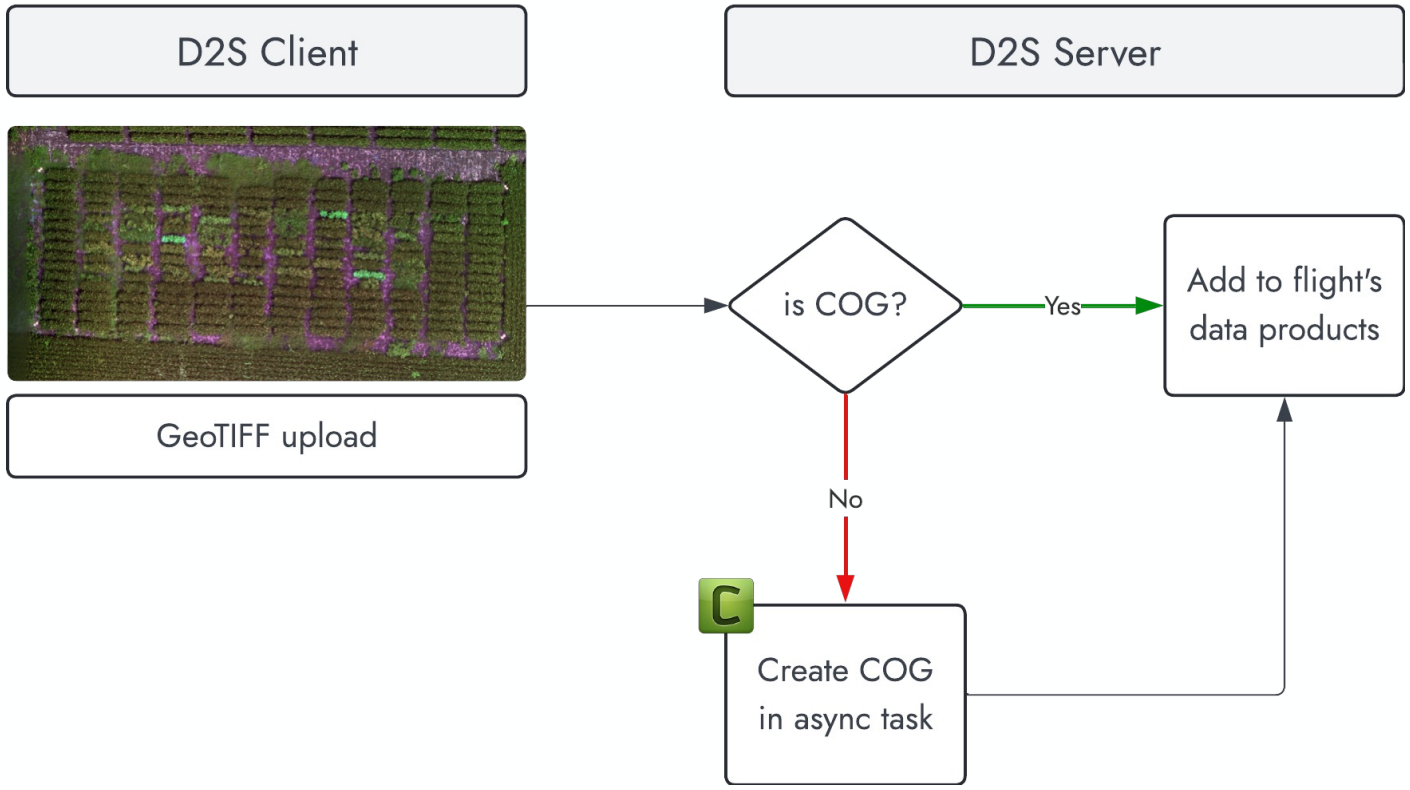
Cloud Optimized File Formats

- Automatically convert uploaded GeoTIFFs to **Cloud Optimized GeoTIFF** (COG) format and LAS point clouds to **Cloud Optimized Point Cloud** (COPC) format

- GDAL used to create COG in asynchronous background task
  - Prevents blocking user from interacting with site while waiting for conversion to complete

- Untwine used to create COPC in background

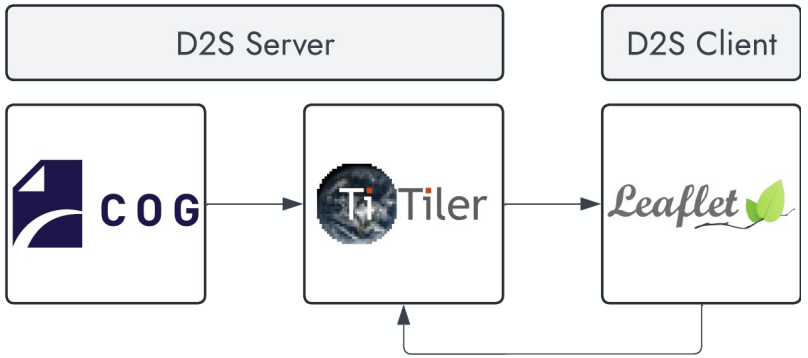
- Stream only required portions of COG and COPC data instead of downloading entire dataset



PURDUE UNIVERSITY

Visualizing COGs on D2S

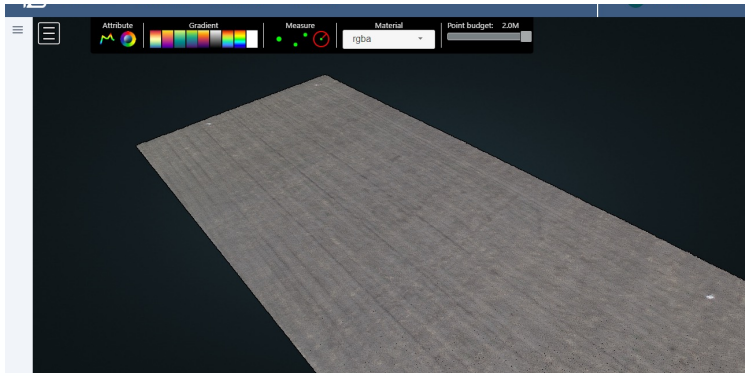
- Dynamic tile server, TiTiler, accesses COG stream and generates map tiles
- Leaflet's TileLayer seamlessly interacts with dynamic and static map tiles
- TiTiler generates new map tiles on-the-fly when symbology properties updated in UI
- Successfully used on datasets exceeding 50 GB



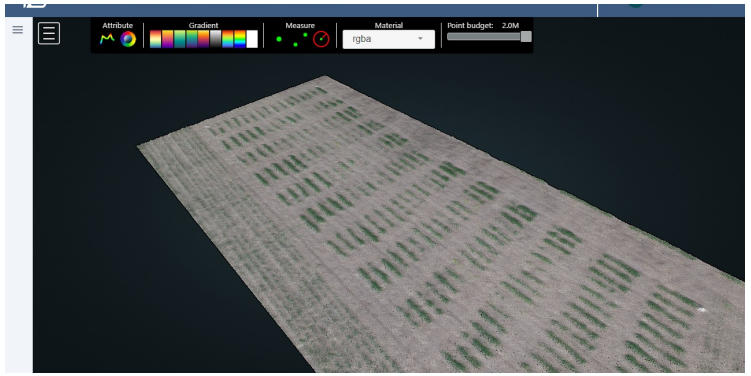
Process for visualizing COG.

Screenshot of a Cloud Optimized GeoTiff in D2S.

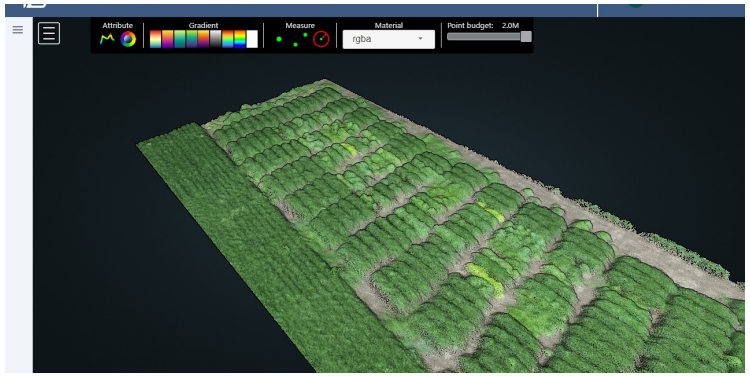
Visualizing COPCs with Potree on D2S



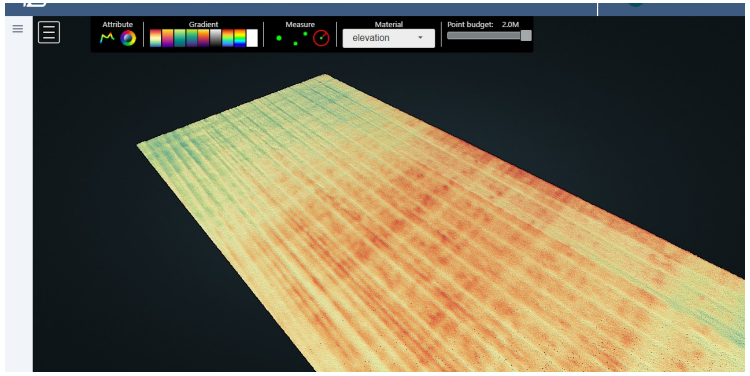
June 4th, 2022 - RGB



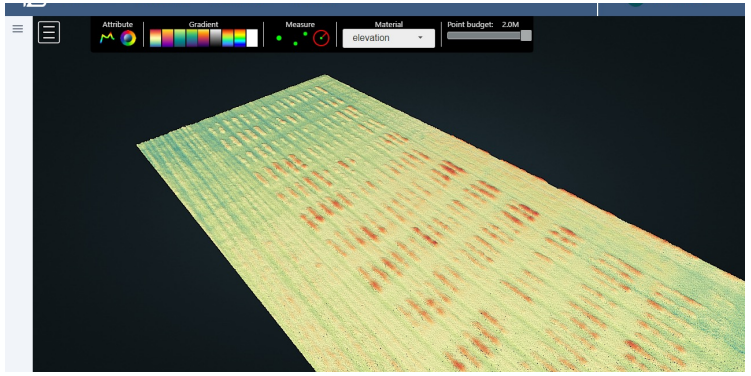
June 23rd, 2022 - RGB



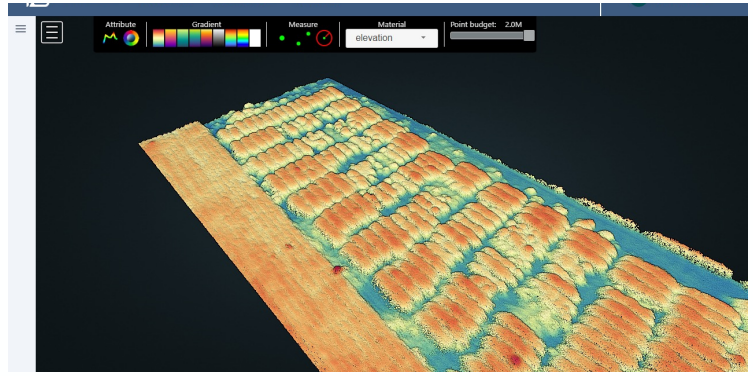
July 25th, 2022 - RGB



June 4th, 2022 - Elevation



June 23rd, 2022 - Elevation

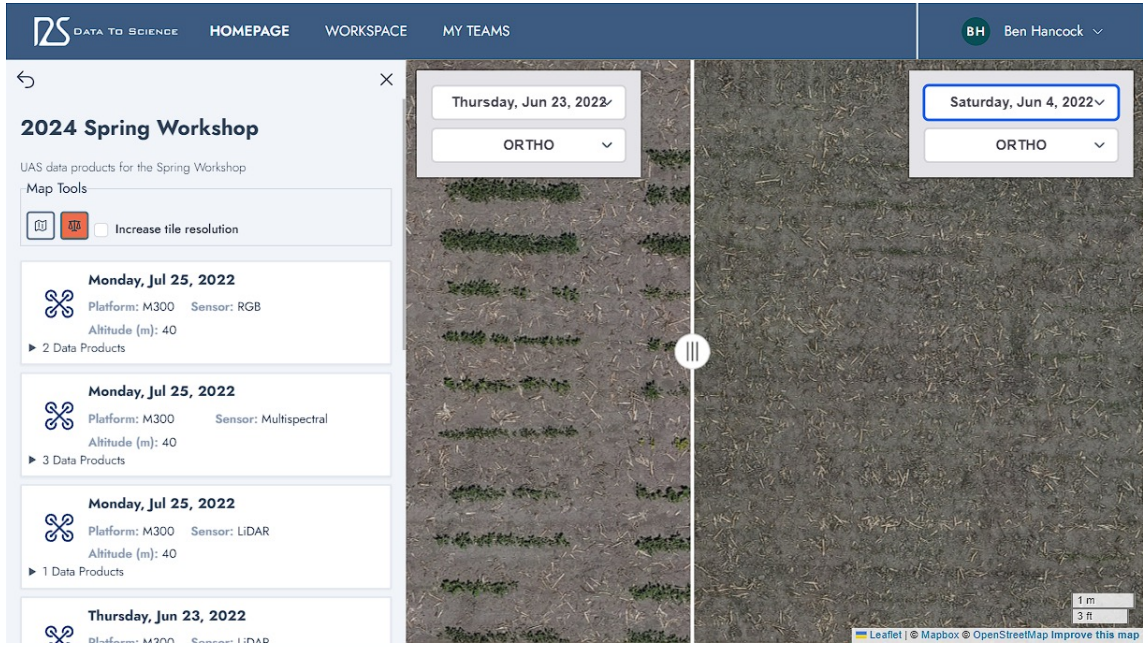


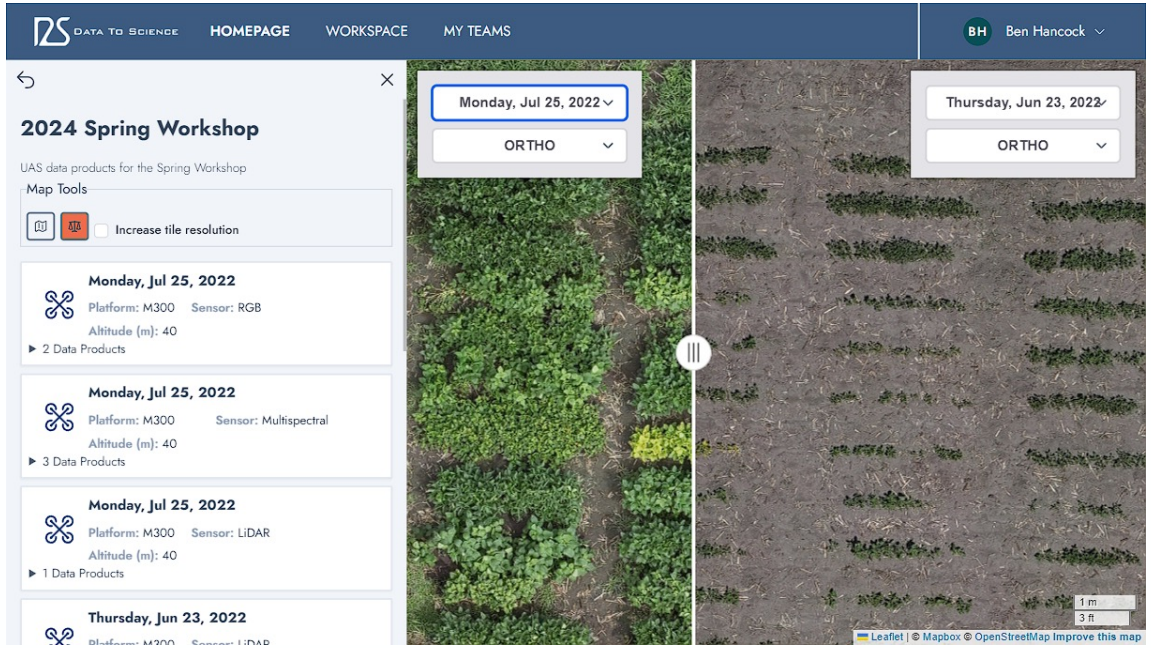
July 25th, 2022 - Elevation

PURDUE UNIVERSITY

Map tools

Use slider to compare two data products from different dates.

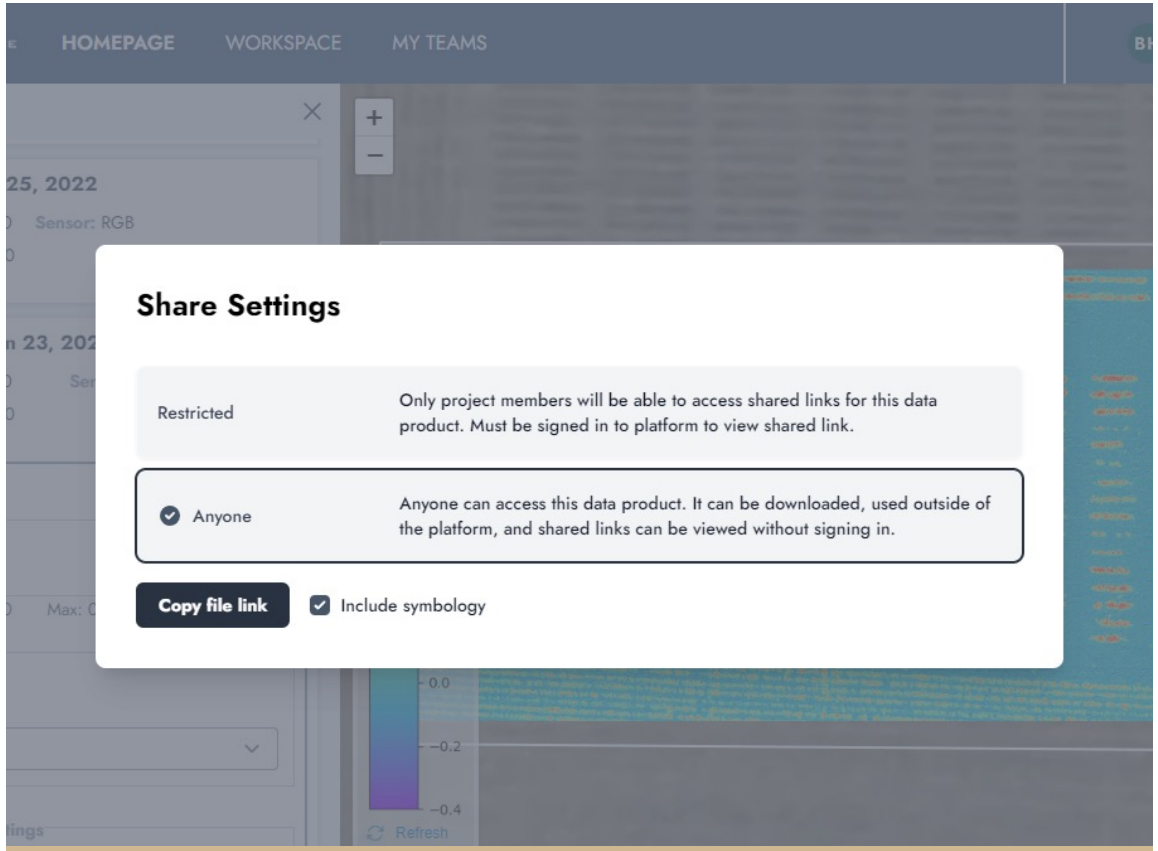




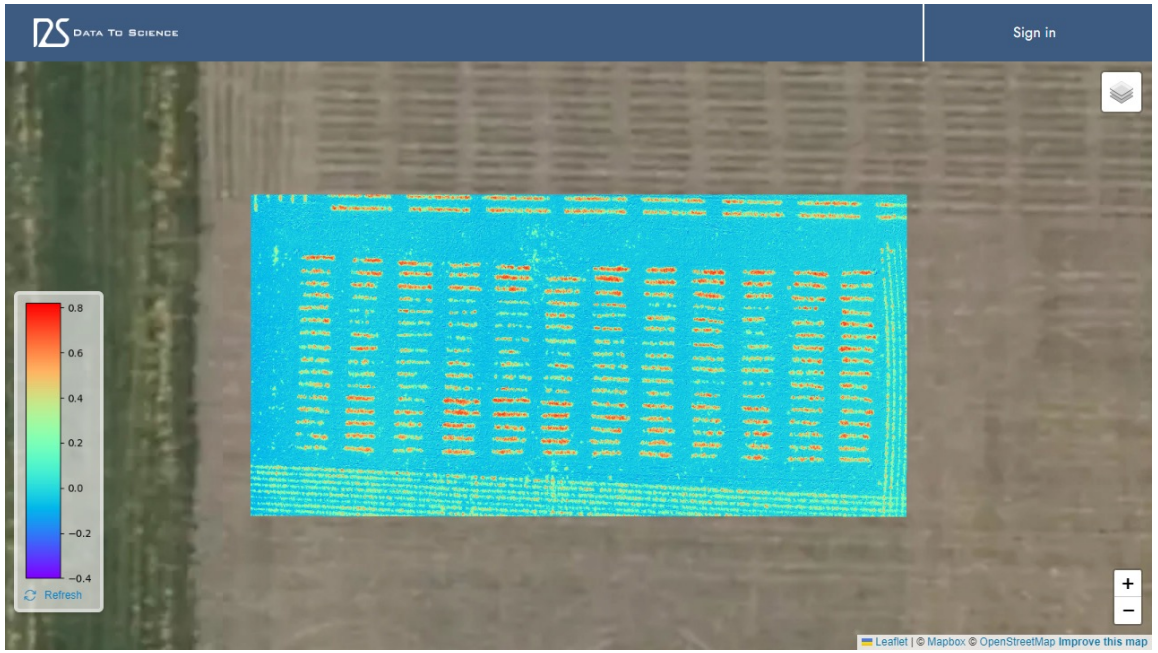
PURDUE UNIVERSITY

Sharing maps

Share maps with the public or restricted to only D2S users with appropriate permissions.



Data products can be set to restricted access or anyone (public).

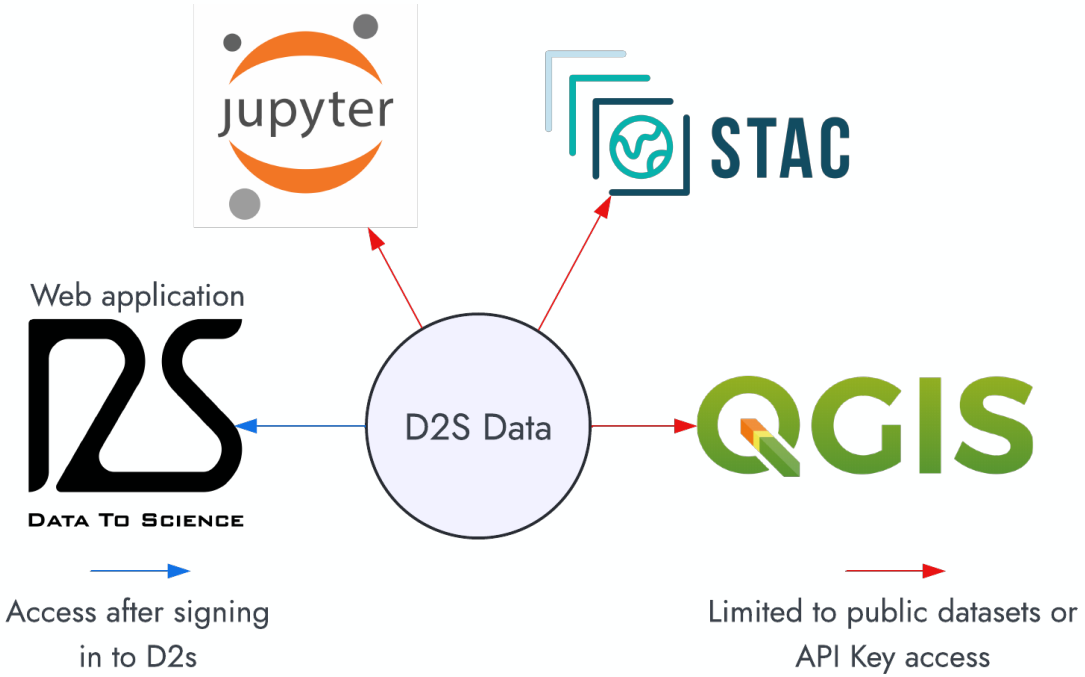


Maps for publicly shared data products can be viewed by anyone.

PURDUE UNIVERSITY

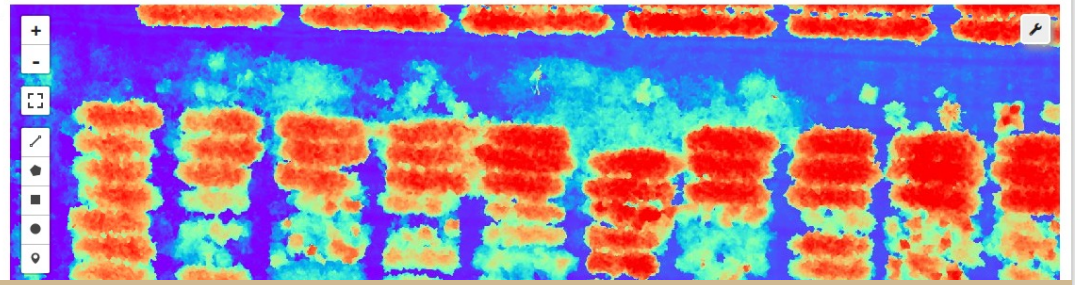
Accessing your data

Data stored on D2S can be accessed from external applications that support COG and COPC formats.




```
[2]: # Use the TiTiler demo endpoint. Replace this if needed.
     os.environ["TITILER_ENDPOINT"] = "
[3]: # initial map
     Map = leafmap.Map()

     # construct url with api key for D2S hosted cog
     cog_url = "
     api_key = "
     cog_url_final = f"{cog_url}?API_KEY={api_key}"
[4]: # bounding box for cog
     bounds = leafmap.cog_bounds(cog_url_final)
[5]: # add cog layer to map and zoom to its extent
     Map.add_cog_layer(cog_url_final, name="DSM", colormap_name="rainbow")
     Map.zoom_to_bounds(bounds)
[6]: # create colorbar for cog layer using min/max from D2S statistics
     vmin = 187.93
     vmax = 188.99
     Map.add_colormap(
         "rainbow",
         width=8.0,
         height=0.4,
         orientation="horizontal",
         vmin=vmin,
         vmax=vmax,
     )
[7]: Map
[7]:
```



Example of accessing a D2S data product in a Jupyter Notebook.

PURDUE UNIVERSITY

Acknowledgements

- This research was supported [in part] by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture, Grant (USDA-NIFA award 2022-70412-38454).

  The findings and conclusions in this preliminary presentation have not been formally disseminated by the U. S. Department of Agriculture and Should not be construed to represent any agency determination or policy.

- Purdue Plant Sciences 2.0 Initiative

- Purdue Digital Forestry

PURDUE UNIVERSITY

Thank You

Development Team Contact

Jinha Jung (jinha@purdue.edu)
Project Lead

Minyoung Jung (jung411@purdue.edu)
Project Manager

Cheryl Qian
UI/UX Lead

Ben Hancock (hancocb@purdue.edu)
Web Developer

Na Zhou
UI/UX Designer

PURDUE UNIVERSITY®