

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

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.



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

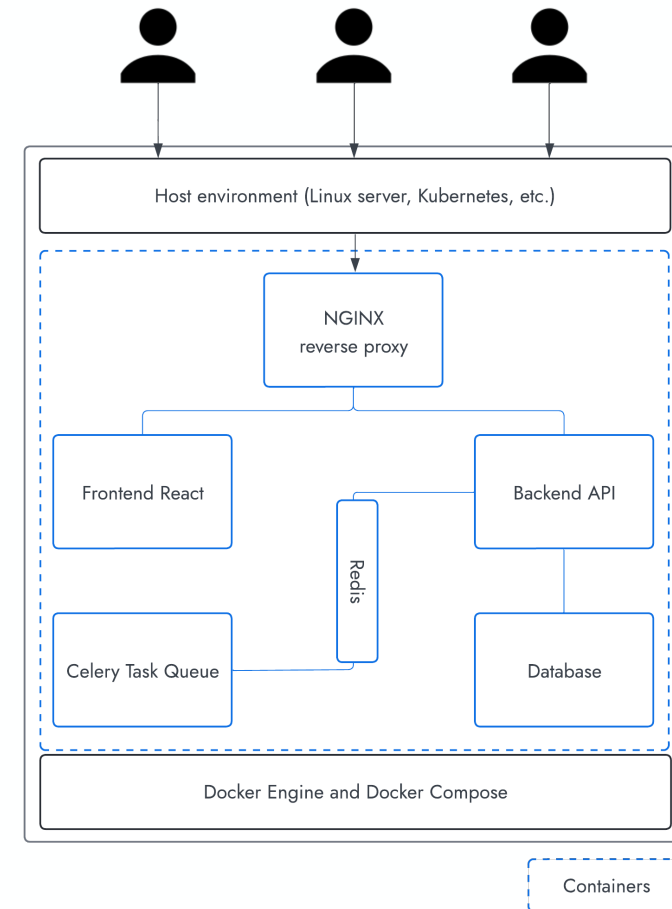
Searchable catalog of published datasets from application instances

- STAC Browser
- STAC FastAPI w/ pgstac

Application instance

Open-source, anyone can run an application instance.

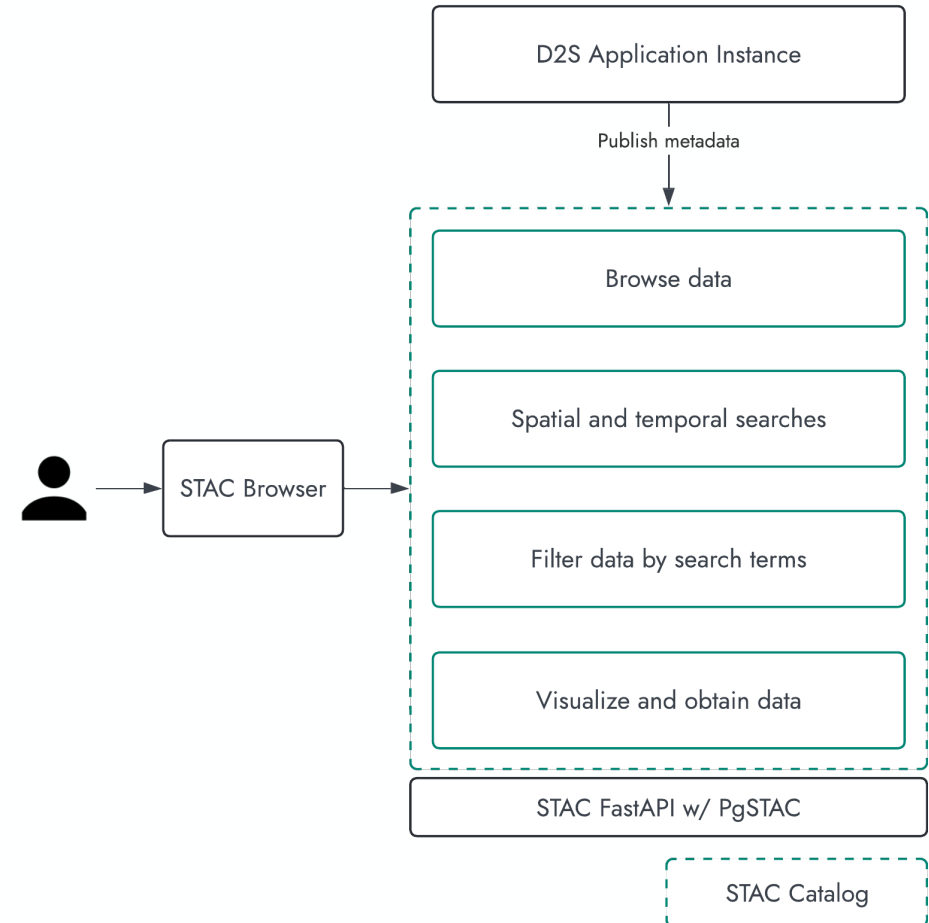
- Multi-container application
 - Each container runs a single service that supports the application
- Easy deployment with Docker Compose
- Can be hosted from a variety of environments
- Data always remains inside application instance



Public Central Catalog

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


- Built on top of the **SpatioTemporal Asset Catalog (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



STAC for Data to Science

carroll-2021-ortho API Source Share Language: English

in STAC for Data to Science Up Collection Browse Search



Collection

Indiana Ortho Imagery
Ortho for Indiana Statewide
1/1/2021, 12:00:00 AM UTC - 12/31/2023, 12:00:00 AM UTC

General

Time of Data 10/11/2023, 5:37:42 PM UTC

Asset


ortho-image

Cloud-Optimized GeoTIFF image

[Download](#) [Copy URL](#)


GOAL

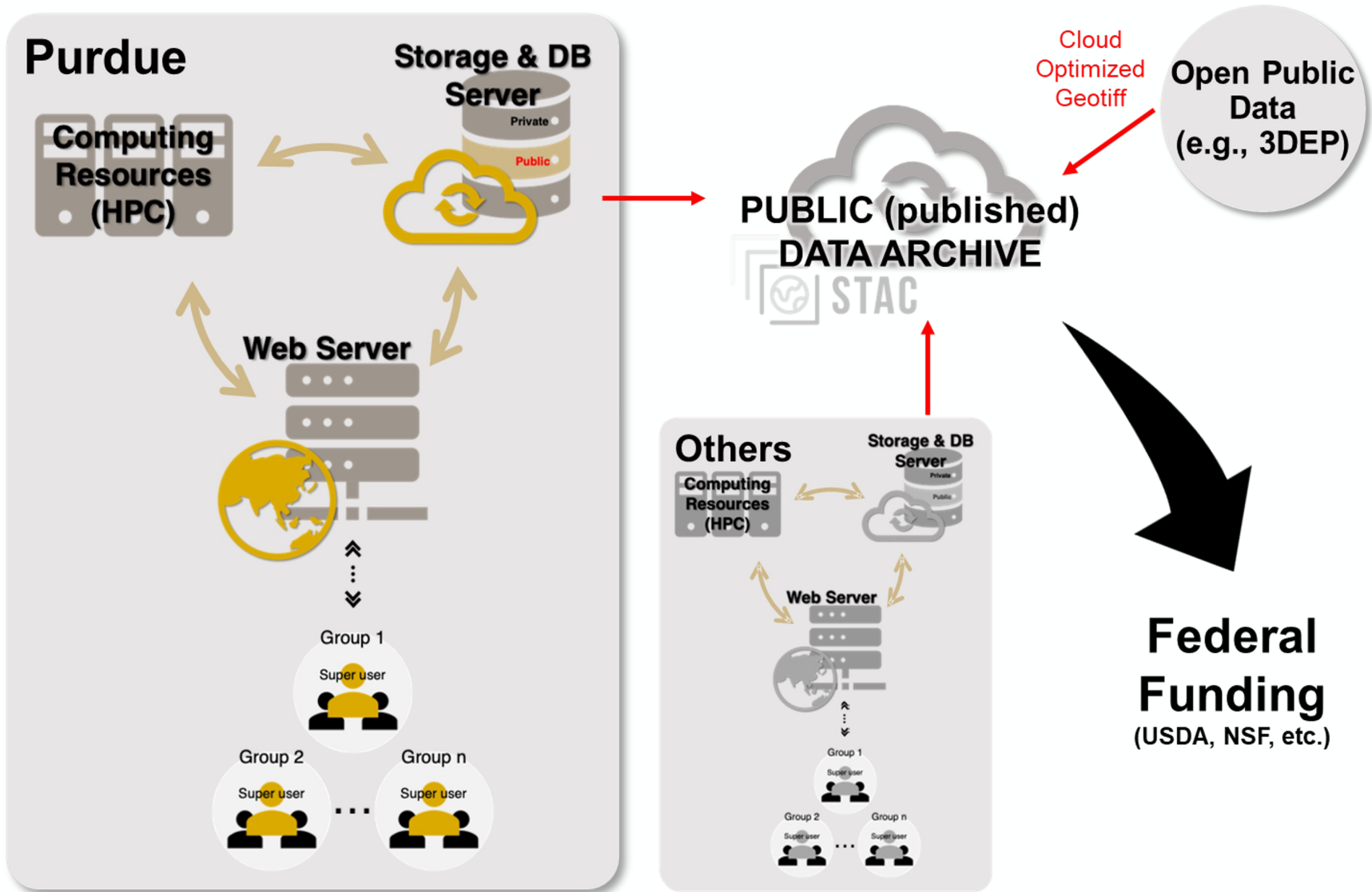
“To make UAS HTP data FAIR and an online platform SCALABLE”

 Findable

 Accessible

 Interoperable

 Reusable





Welcome to **Data to Science**

A Data-driven Open Science Community for Sustained Innovation

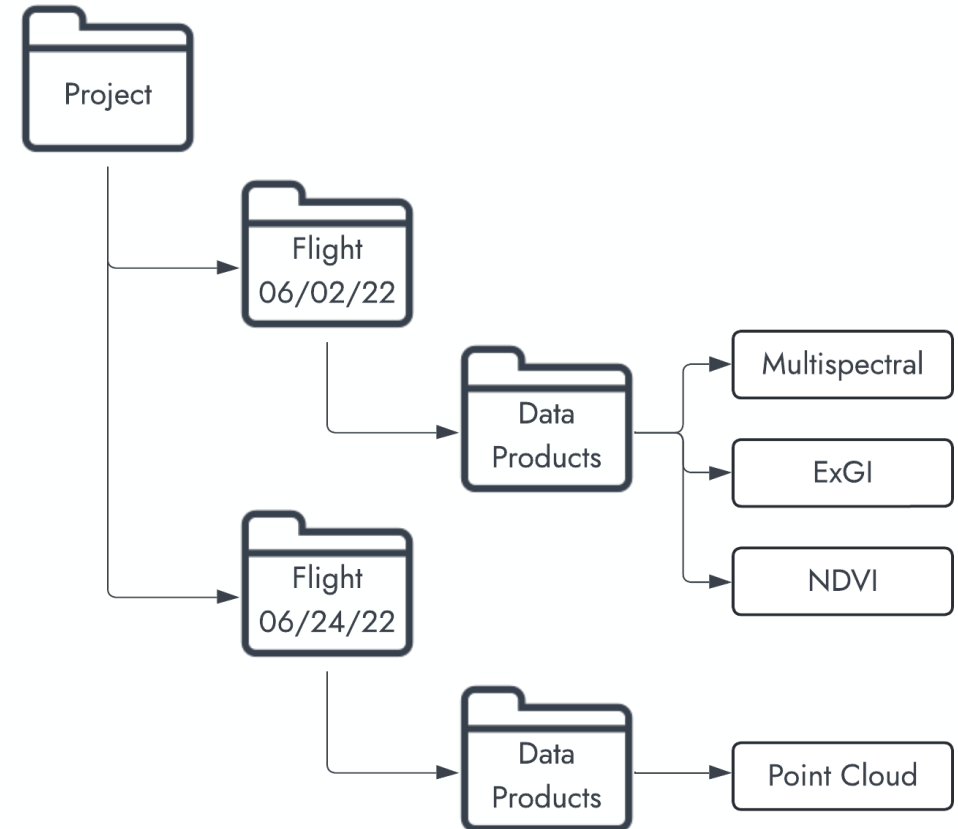
[Sign up](#)

[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.



D2S Homepage

View project locations and individual data products on map.

The screenshot shows the D2S homepage interface. At the top, there is a navigation bar with 'DATA TO SCIENCE', 'HOMEPAGE', 'WORKSPACE', and 'MY TEAMS'. A user profile for 'Ben Hancock' is visible in the top right. On the left, a 'Projects' sidebar contains a search bar and a list of five projects, each with a location pin icon and a flight count:

- ACRE PROJECT**: Acre analysis, 7 Flights
- Crop Residue-MorehouseChurch**: Field Morehouse, 2 Flights
- Crop residue-Chris40**: Fields Chris 40, 2 Flights
- Crop residue-County Line**: field County Line, 2 Flights
- Gujranwala Model Town**: Collaboration with UN PULSE project team, 1 Flight

The main area features a world map with several project locations marked by colored pins (yellow, green, blue). A scale bar at the bottom right indicates 3000 km and 2000 mi.

This screenshot shows a detailed view of a data product. The top navigation bar is identical to the previous screenshot. The main content area is titled 'Saturday, Aug 20, 2022' and includes the following information:

- Platform: M300, Sensor: RGB
- Altitude (m): 40
- 2 Data Products

The 'DSM' (Digital Surface Model) section is expanded, showing:

- Band Info**: b1 (Gray)
- Statistics: Mean: 188.69, Min: 186.13, Max: 190.50, Std. Dev: 0.35
- Color Properties**: Color Ramp* is set to 'rainbow'.
- Min / Max Value Settings**: 'Min/Max' is selected, with Min* at 186.128 and Max* at 190.502. An 'Apply Changes' button is at the bottom.

The map on the right displays a color-coded DSM of a field. A white box highlights a specific area. A vertical color scale legend on the left of the map ranges from 186.5 (blue) to 190.5 (red). A scale bar at the bottom right shows 20 m and 50 ft.

D2S Workspace

Manage projects, flights and data.

The screenshot shows the D2S Workspace interface for a project named "2024 Spring Workshop > RGB | M300 | 2022-06-04". The top navigation bar includes "DATA TO SCIENCE", "HOMEPAGE", "WORKSPACE", "MY TEAMS", and a user profile for "Ben Hancock". The main content area is divided into two sections: "Raw Data" and "Data Products". The "Raw Data" section indicates that no raw data has been uploaded. The "Data Products" section displays two data product thumbnails: "DSM" and "ORTHO", both dated "2022-07-25" and associated with "M300". Each thumbnail includes a "Click to Copy URL" button, a "View" button, and a "Toolbox" button. A "Upload Data" button is located at the bottom right of the interface.

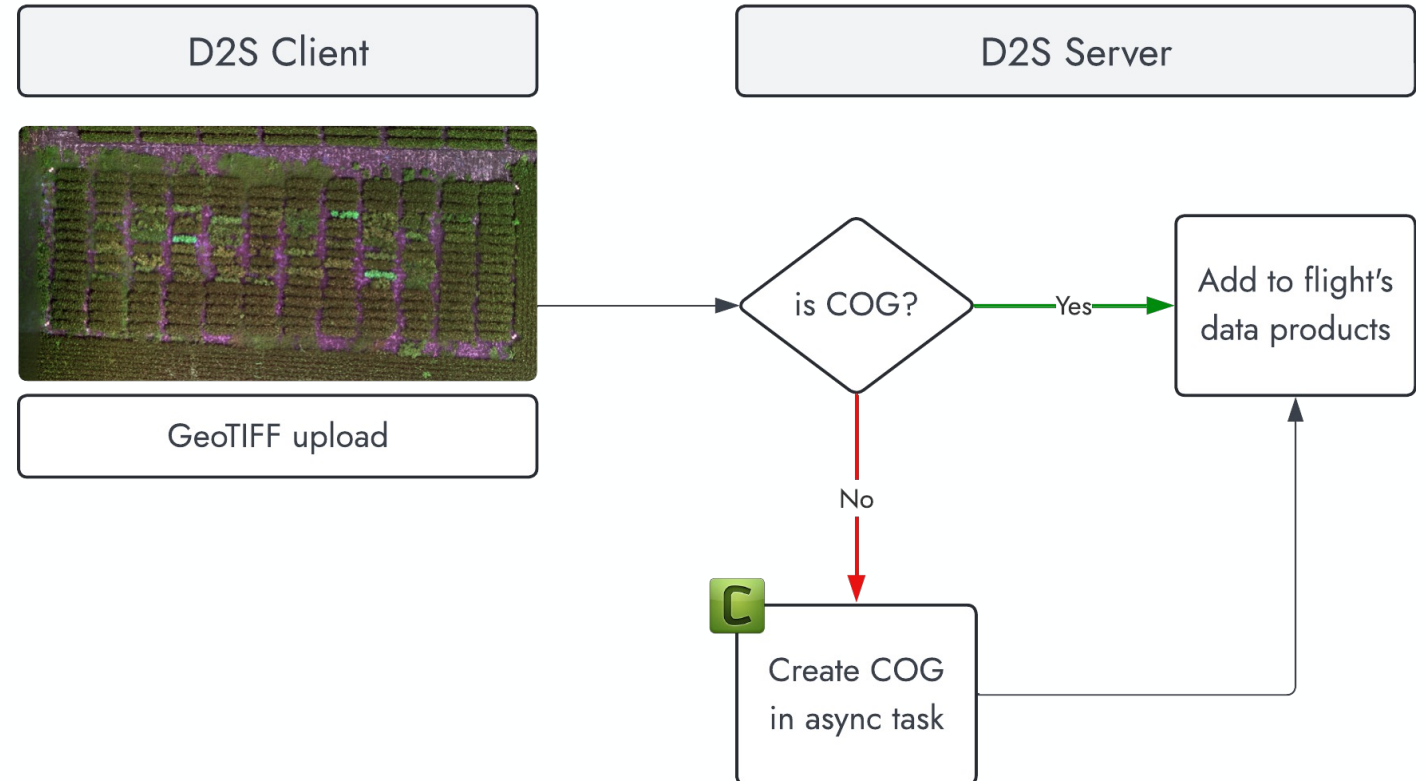
The screenshot shows the D2S Workspace interface for a project named "2024 Spring Workshop > Multispectral | M300 | 2022-07-25". The top navigation bar is identical to the previous screenshot. The main content area shows the "Raw Data" section with a "No raw data has been uploaded" message. The "Data Products" section displays a list of data products: "RGB" (2022-07-25, M300), "Multispectral" (2022-07-25, M300), "LiDAR" (2022-07-25, M300), and "LiDAR" (2022-06-23, M300). The "Multispectral" product is highlighted with a red border. A "Toolbox" modal is open over the "Multispectral" product, allowing the user to select data products to be generated. The modal includes a "Run" button. The "Toolbox" modal contains the following options:

- Select data products to be generated:
 - Excess Green Vegetation Index (ExG)
 - Normalized Difference Vegetation Index (NDVI)
- Select Red and Near-Infrared Bands:
 - Red Band*:
 - NIR Band*:

A "Run" button is located at the bottom of the modal. The "Upload Data" button is visible at the bottom right of the interface.

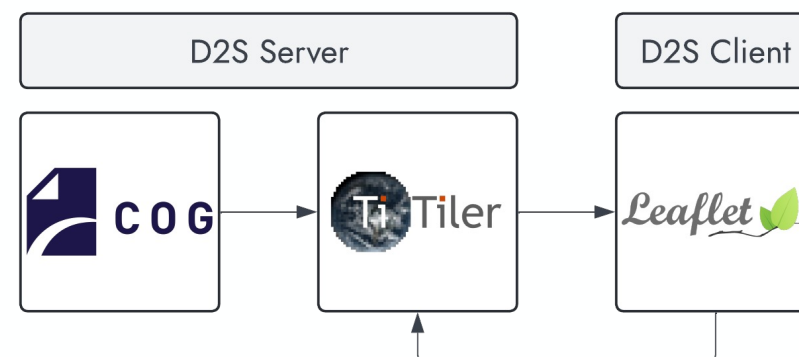
Cloud Optimized File Formats

- 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

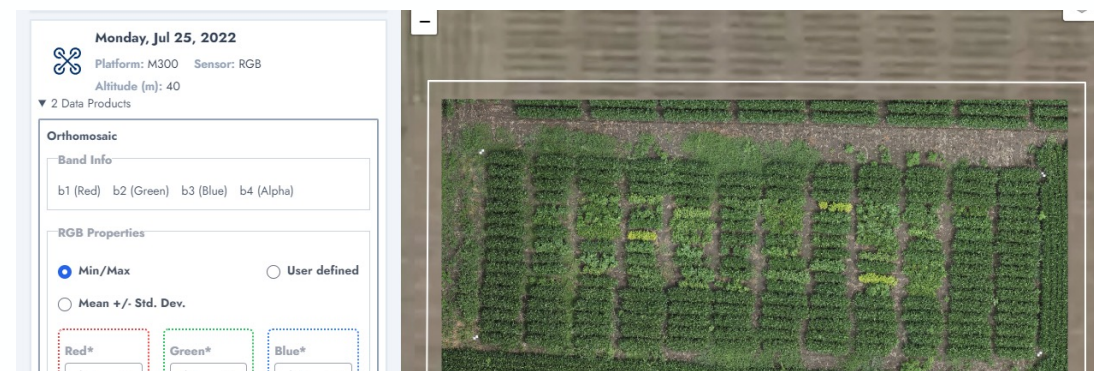


Visualizing COGs on D2S

- 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

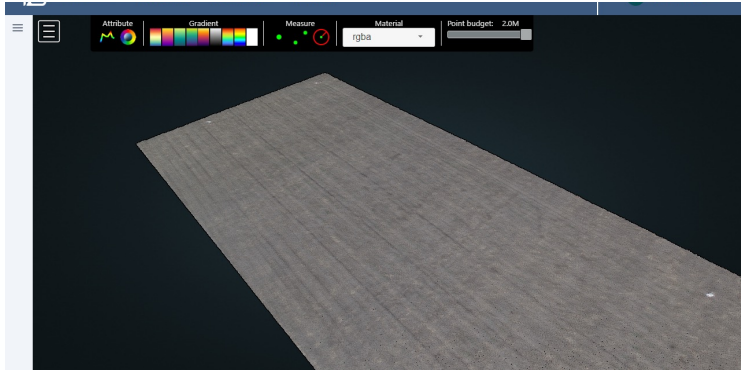


Process for visualizing COG. Update to scale or color map

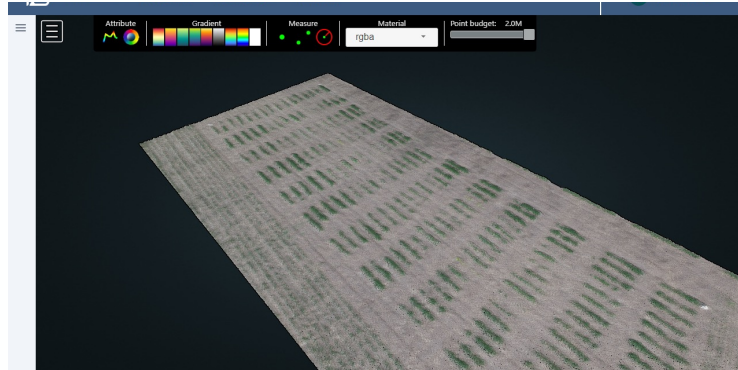


Screenshot of a Cloud Optimized GeoTiff in D2S.

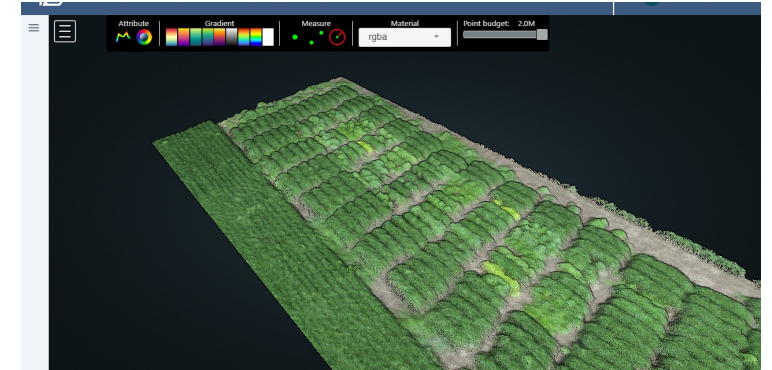
Visualizing COPCs with Potree on D2S



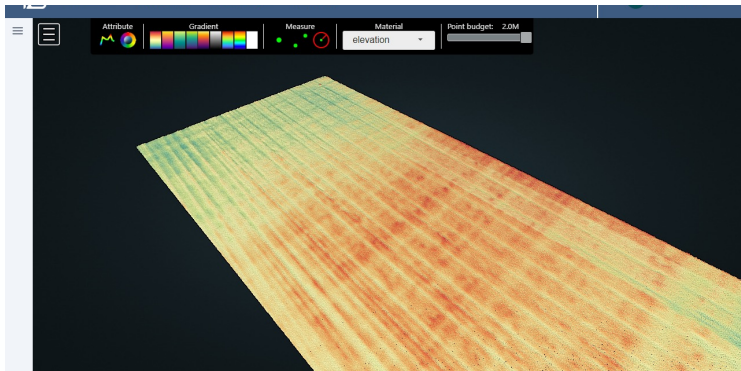
June 4th, 2022 - RGB



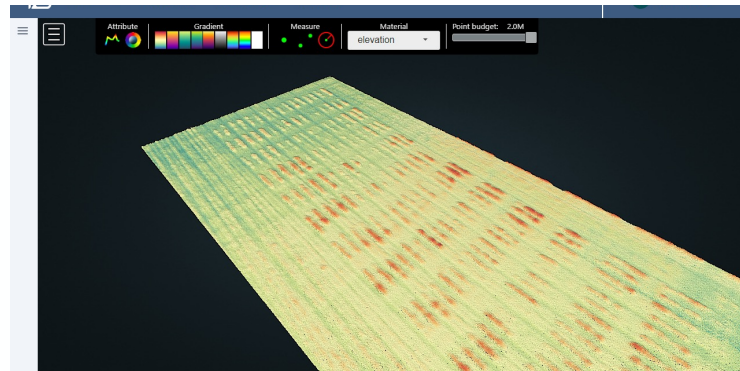
June 23rd, 2022 - RGB



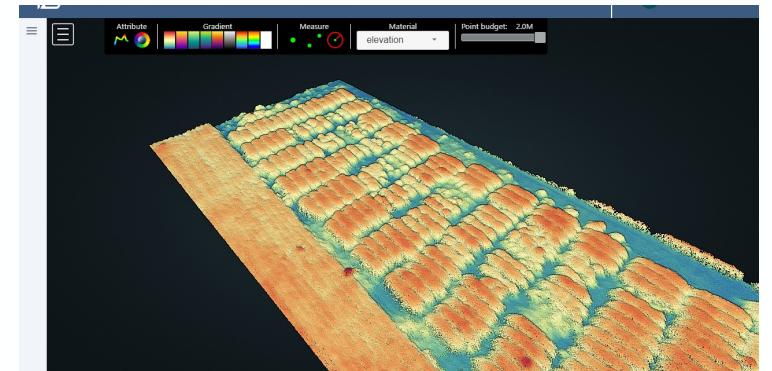
July 25th, 2022 - RGB



June 4th, 2022 - Elevation



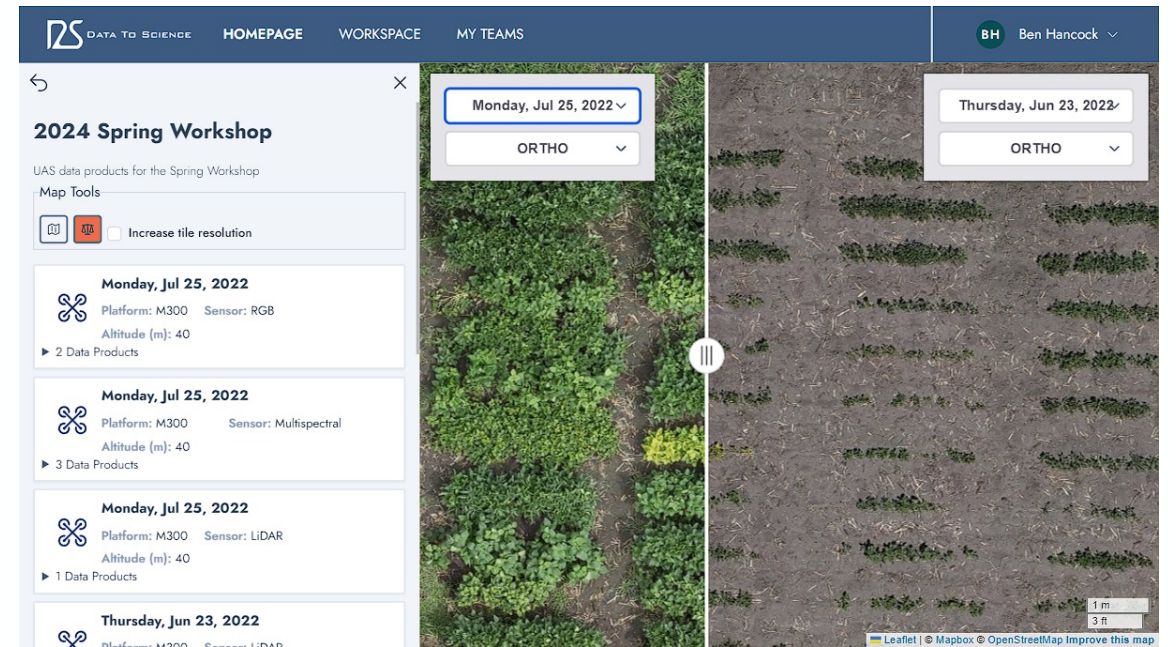
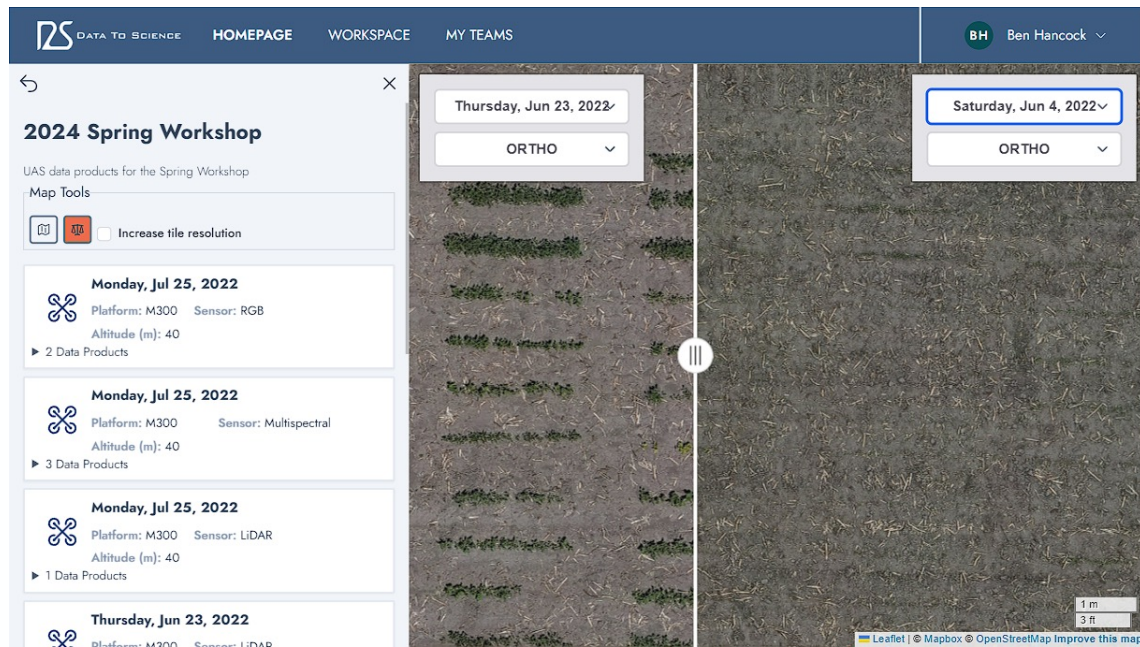
June 23rd, 2022 - Elevation



July 25th, 2022 - Elevation

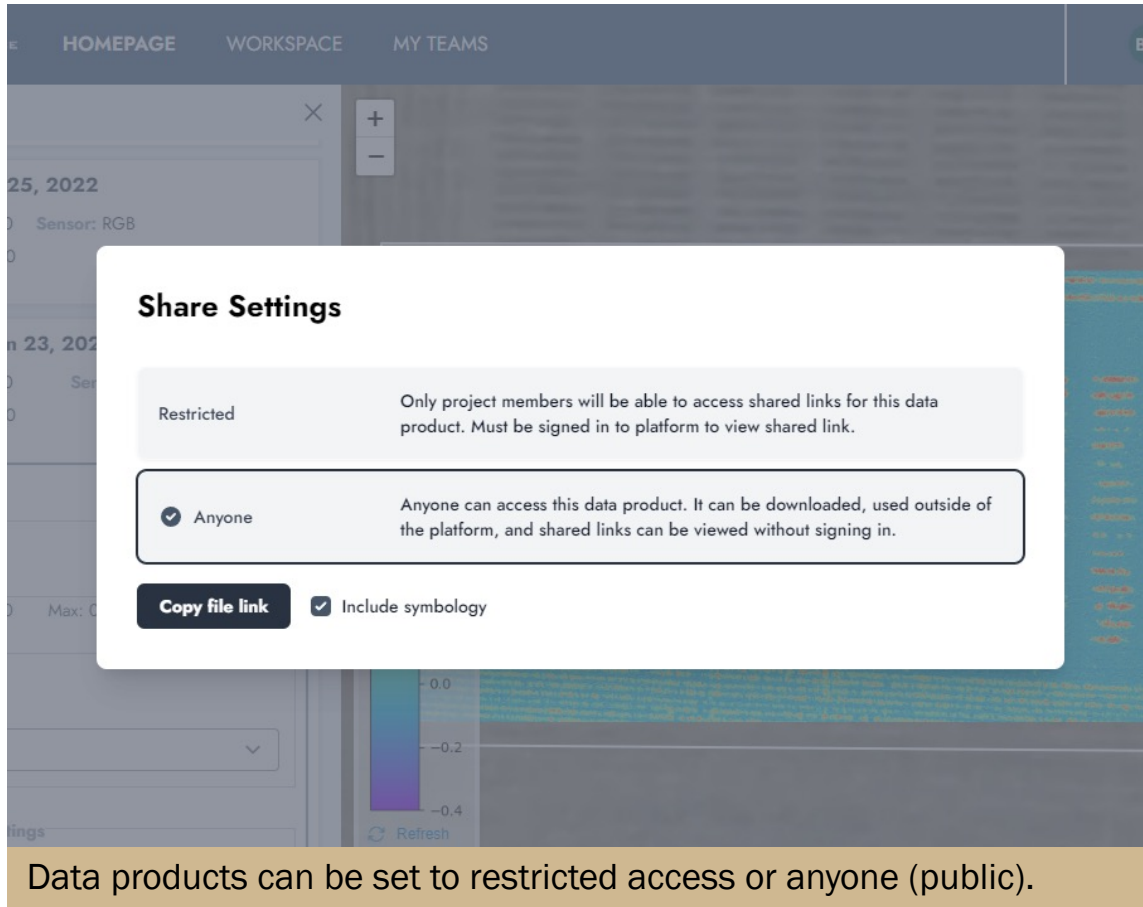
Map tools

Use slider to compare two data products from different dates.



Sharing maps

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



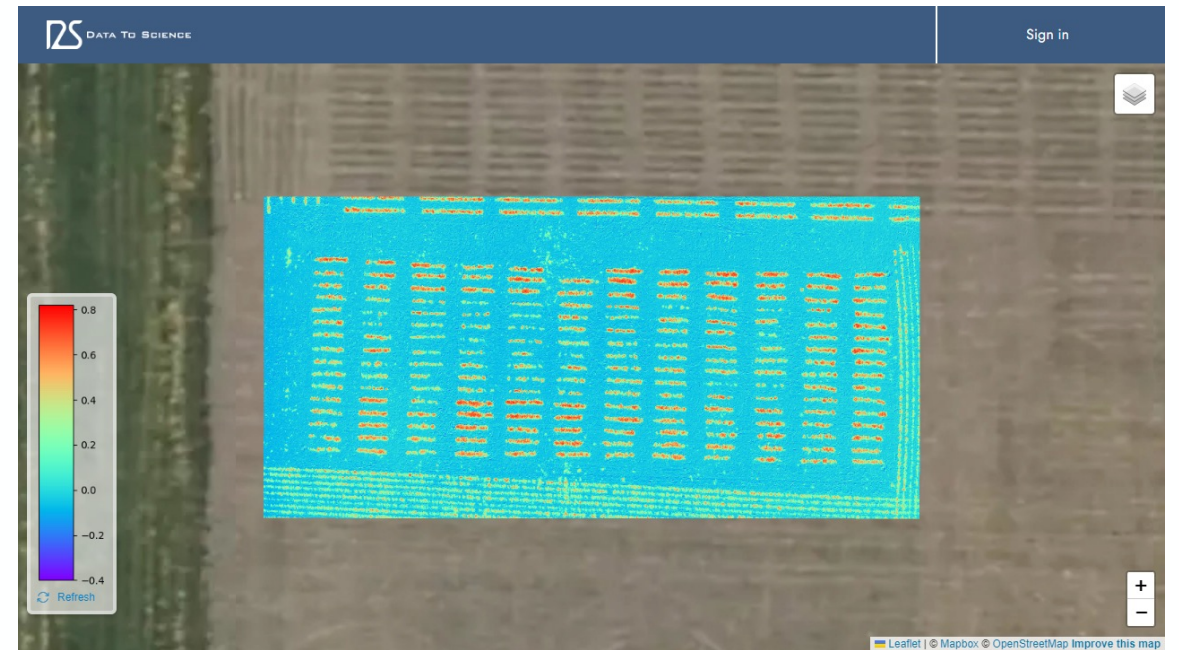
Share Settings

Restricted Only project members will be able to access shared links for this data product. Must be signed in to platform to view shared link.

Anyone Anyone can access this data product. It can be downloaded, used outside of the platform, and shared links can be viewed without signing in.

[Copy file link](#) Include symbology

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



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

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

