

2024 AI in Ag Conf.

Data to Science Engine (D2SE)

A Data-Driven Open Science
Ecosystem for Sustained Innovation

by **Jinha Jung**
Associate Professor @ Lyles School of Civil Engineering, Purdue University

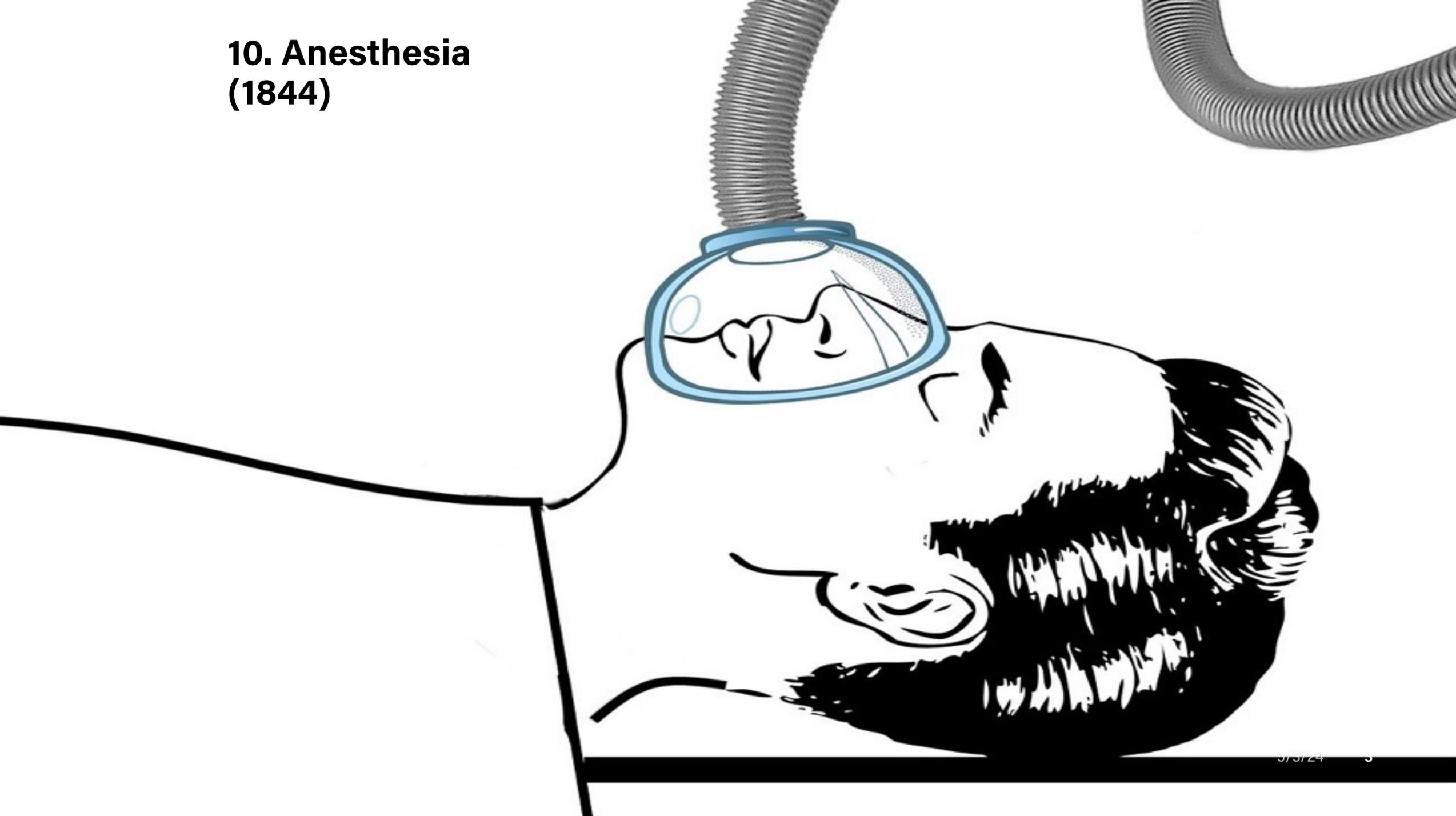


10

The Greatest Inventions

In the past 1,000 years

10. Anesthesia (1844)



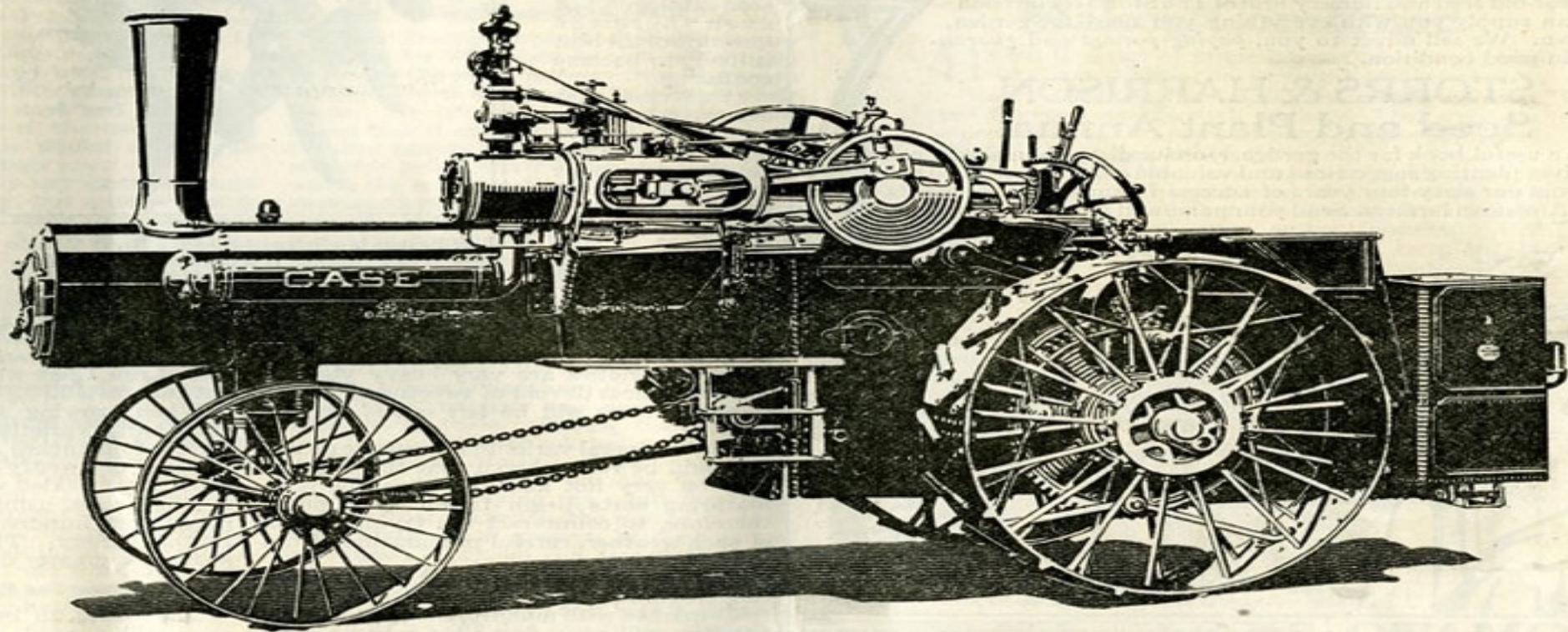
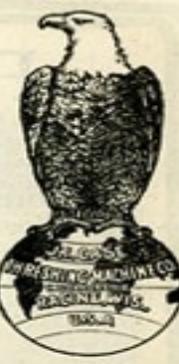


Founded
1842

9. Gas Powered Tractor
(1892)

CASE

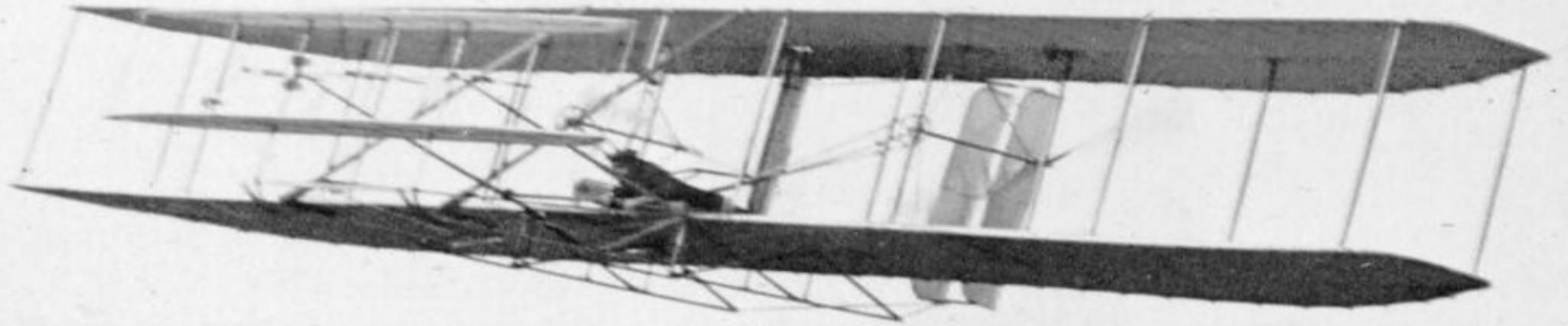
Famous
the
World
Over



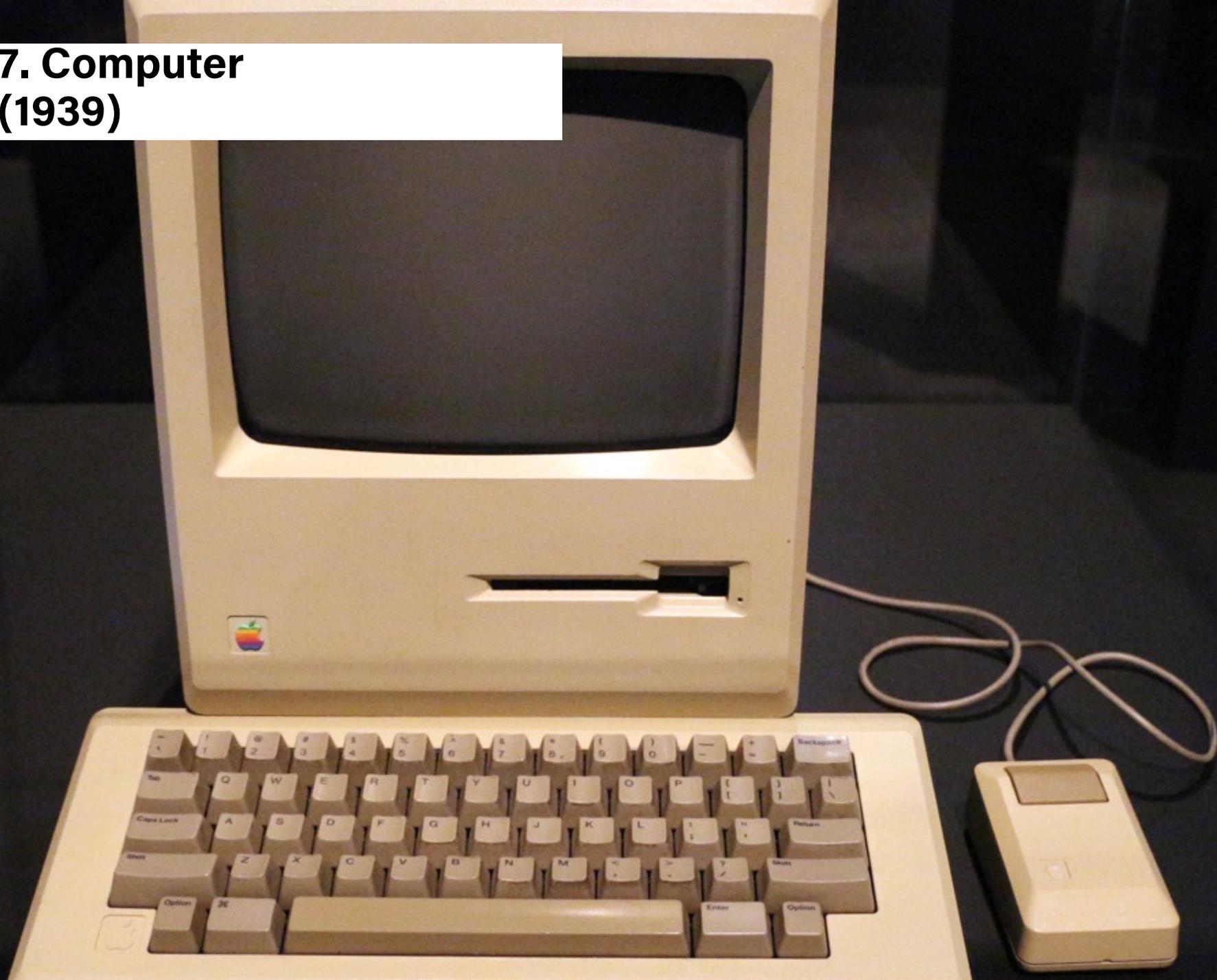
Case Steam Engines

Formed for Power, Simplicity, Durability, Economy

8. Airplane (1903)



**7. Computer
(1939)**



6. Vaccination (1796)



5. Radio and Television (1895, 1926)



4. Telephone (1876)



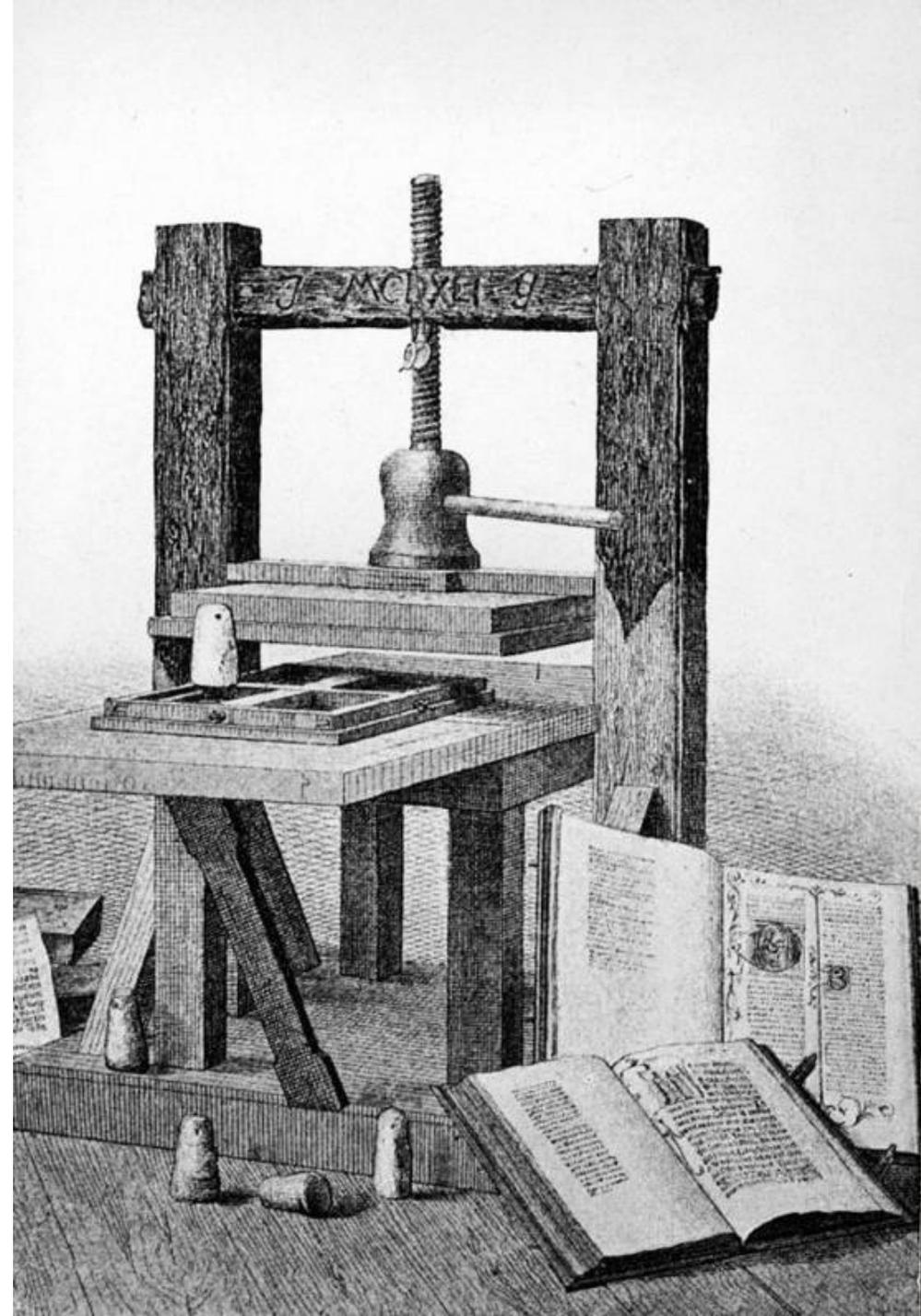
3. Automobile (1885)



2. Electric Light (1879)



1. Printing Press (1450)



Mass Media



Independent Media



● dvd
Search term

● youtube
Search term

+ Add comparison

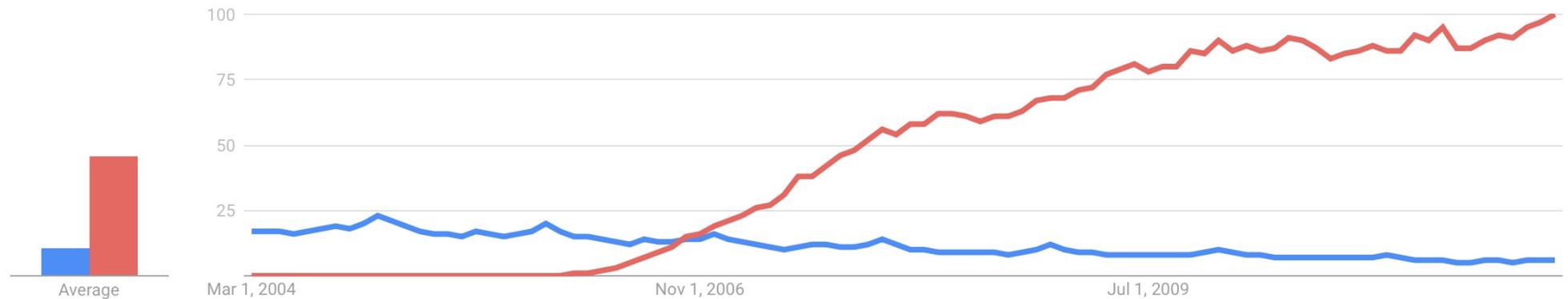
Worldwide ▼

3/1/04 - 12/4/11 ▼

All categories ▼

Web Search ▼

Interest over time ?



M

Market Share



Media Players and Streaming Platform

98%



Indian record label **T-Series** is the most-subscribed channel, with over 263 million subscribers as of April 2024.

American YouTube personality **MrBeast** is the most-subscribed individual and second most-subscribed channel overall, with 252 million subscribers as of April 2024.

The complex block contains two images at the top: the T-Series logo (a white 'T' in a circle on a red background with the word 'SERIES' below it) and a portrait of MrBeast. Below the images are two columns of text providing context for each.

263 M
Subscriber

<https://6sense.com/tech/media-players-and-streaming-platforms/youtube-market-share>

Traditional Remote Sensing Platforms

Spaceborne

Satellite
Optical Sensor/SAR
700-900km



Space Shuttle

185-575km



Airborne

Aerial Photography



Airborne SAR



Modern Remote Sensing Platforms





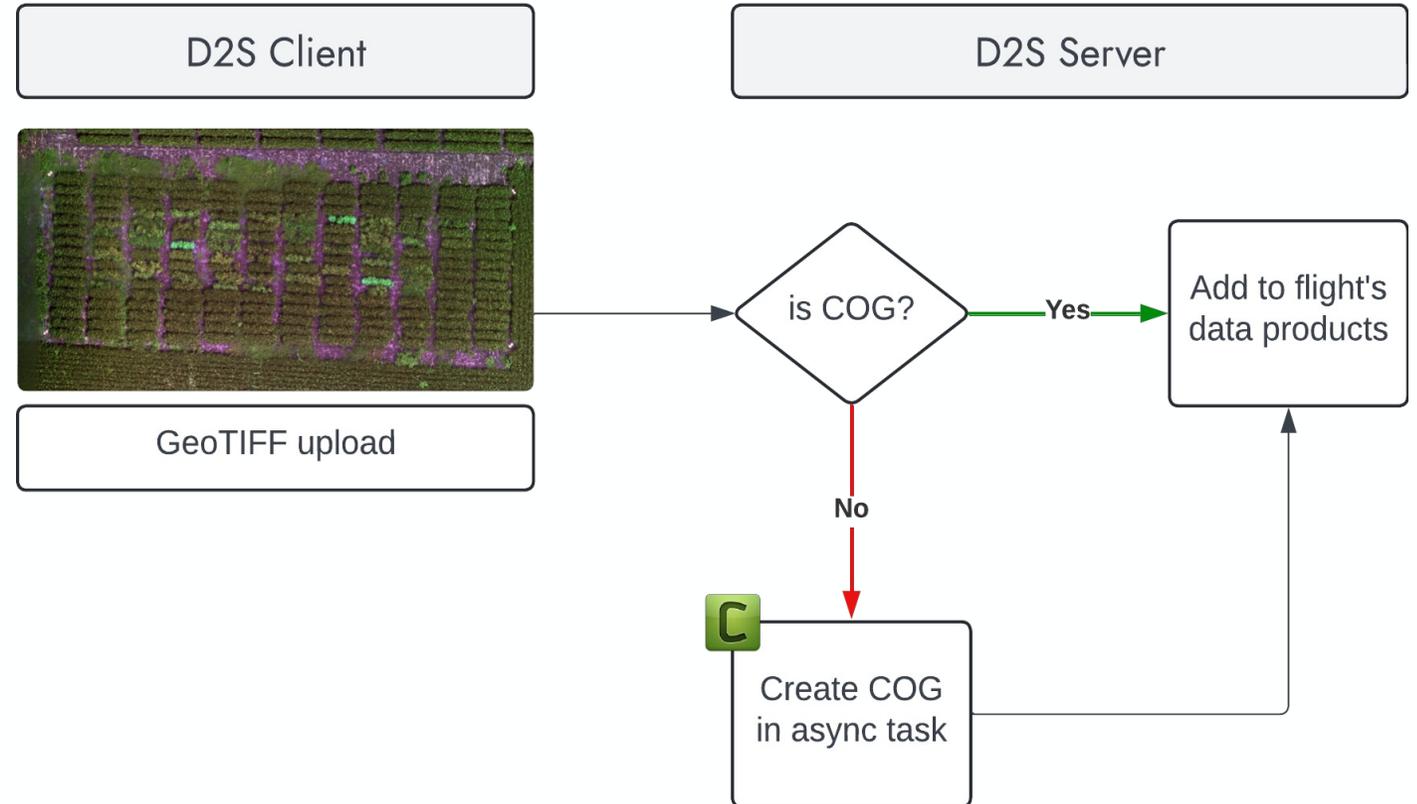
Building an open-source online platform

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.

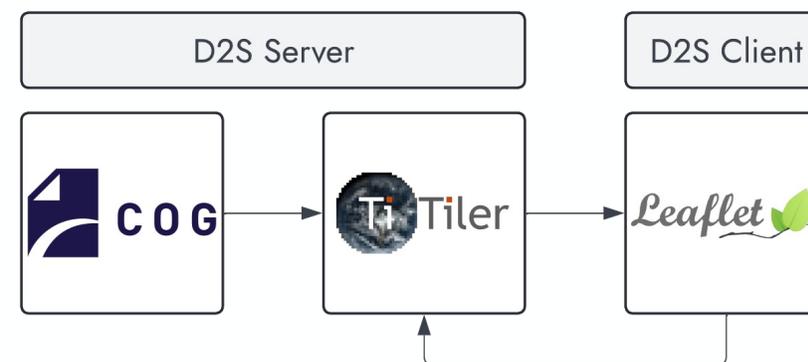
Cloud Optimized File Formats

- 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



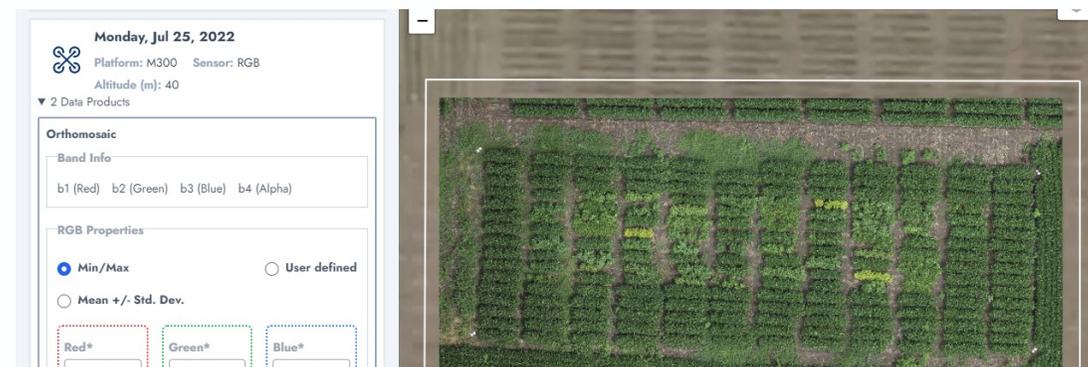
Visualizing COGs on D2S

- 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



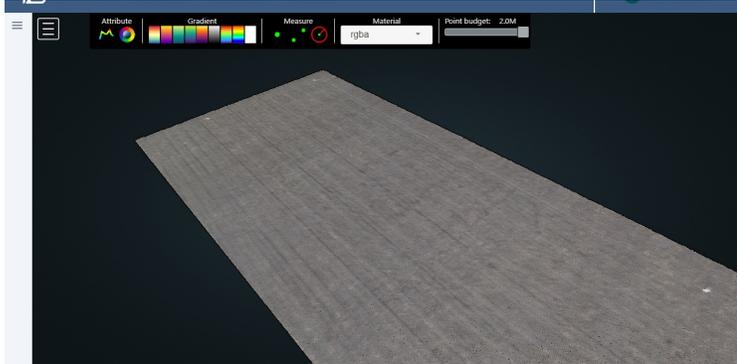
Process for visualizing COG.

Update to scale or color map

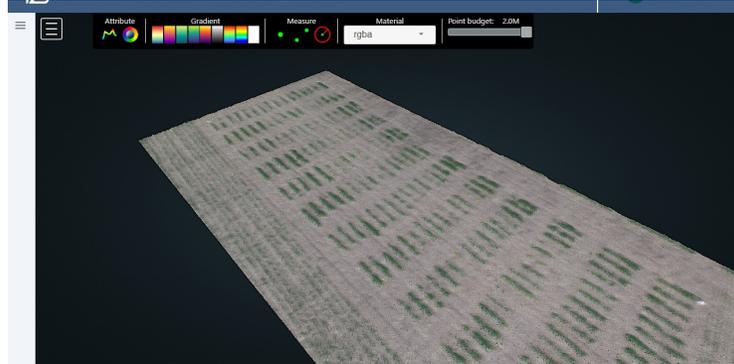


Screenshot of a Cloud Optimized GeoTiff in D2S.

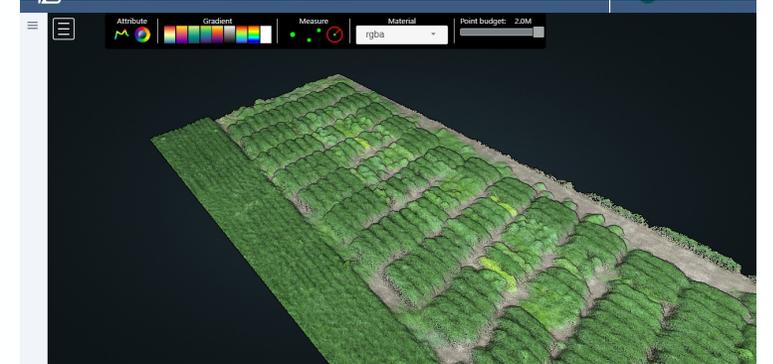
Visualizing COPCs with Potree on D2S



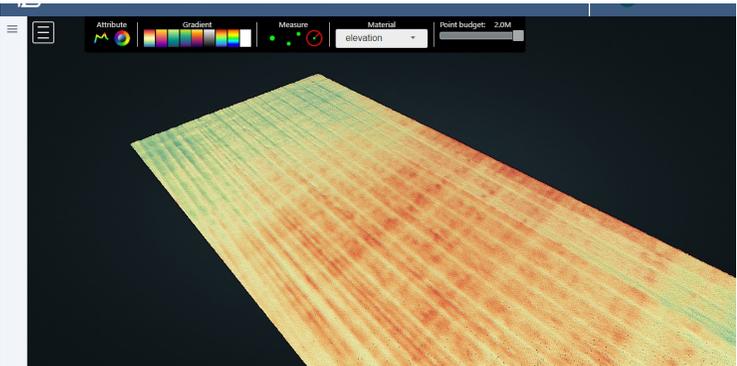
June 4th, 2022 - RGB



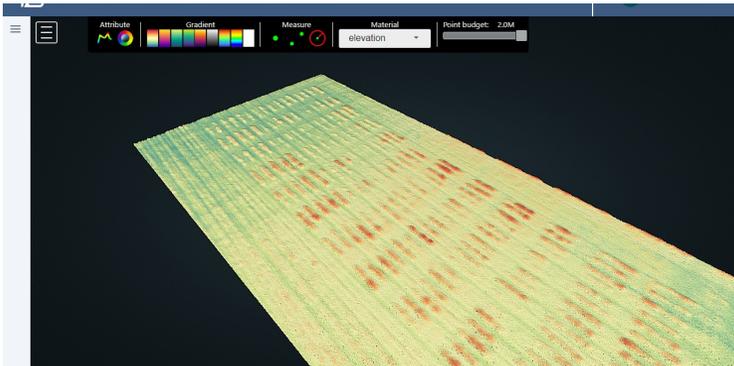
June 23rd, 2022 - RGB



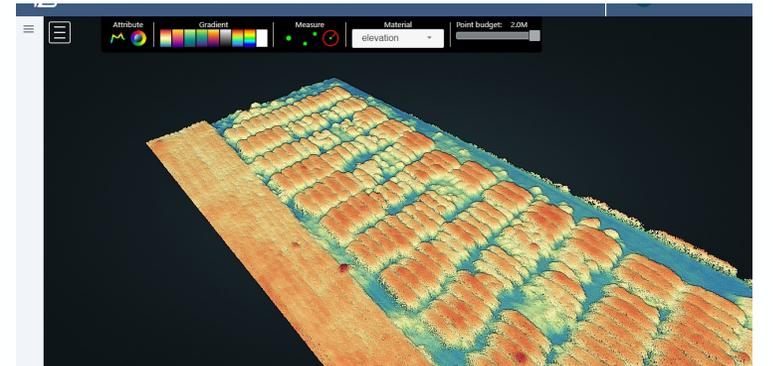
July 25th, 2022 - RGB



June 4th, 2022 - Elevation



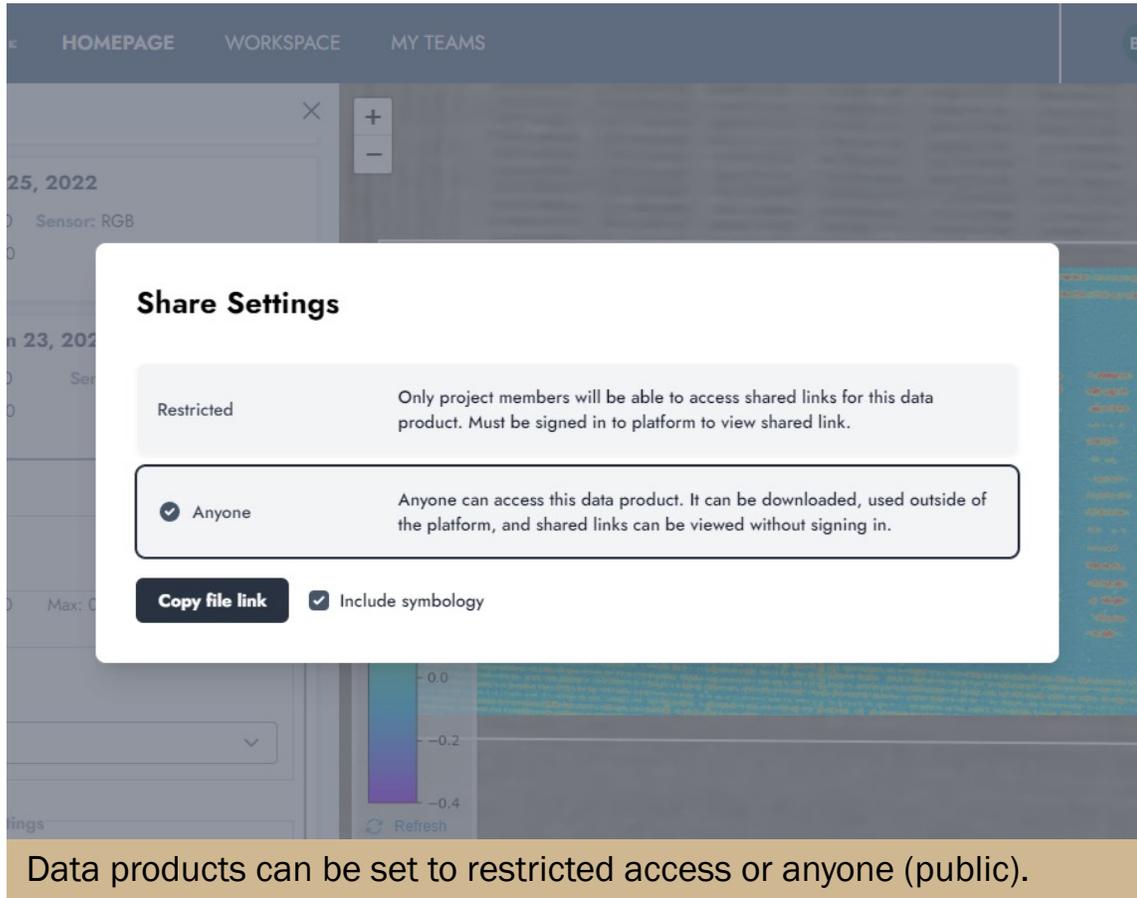
June 23rd, 2022 - Elevation



July 25th, 2022 - Elevation

Share maps

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



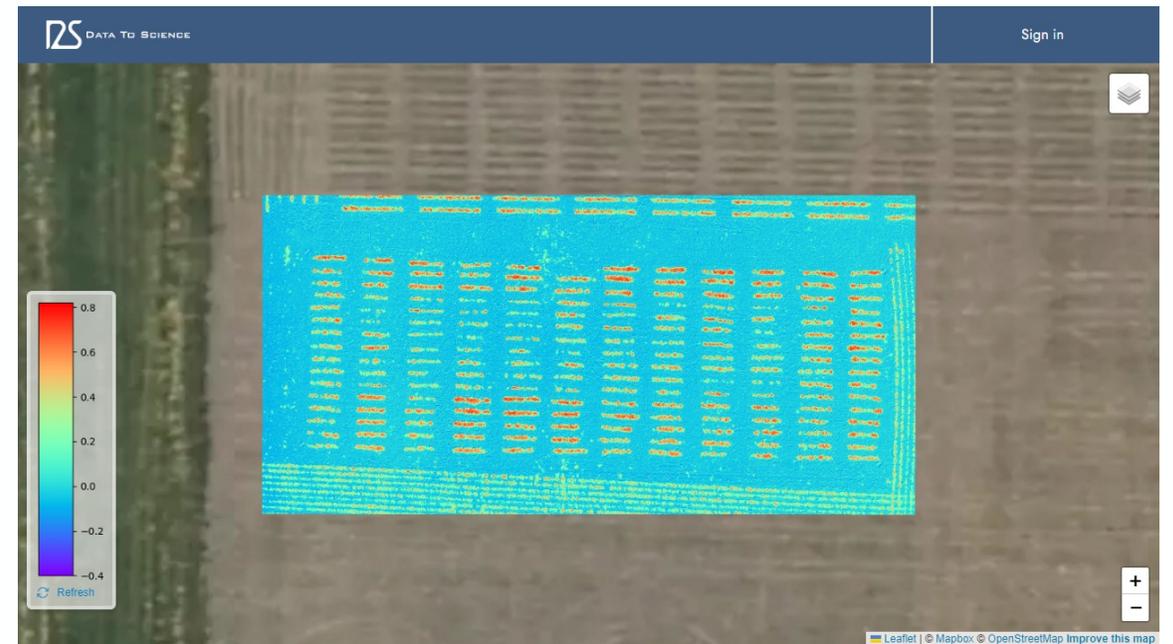
Share Settings

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).



GOAL

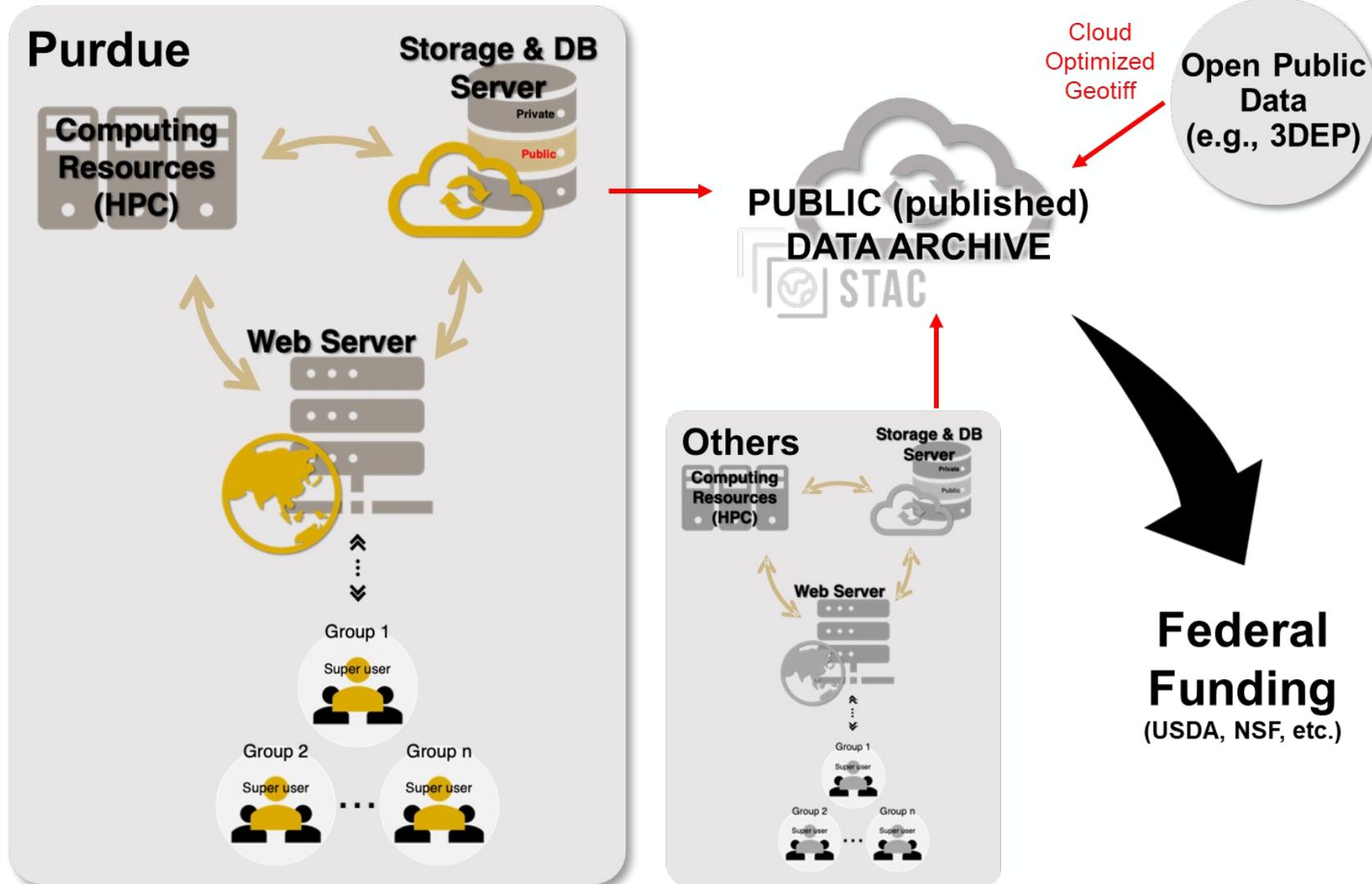
“To make UAS HTP data FAIR and an online platform SCALABLE”

Findable

Accessible

Interoperable

Reusable



Vide

o



Geospati
al Data



<https://docs.gdsl.org/data-to-science-engine-workshop/>



AG2PI

Agricultural Genome to
Phenome Initiative



PURDUE
UNIVERSITY®

College of Agriculture

backend.dockerfile - d2s - Code - OSS

File Edit Selection View Go Run Terminal Help

backend.dockerfile × docker-compose.yml db.example.env

```
backend > backend.dockerfile > ...
40 FROM python-base
47
48 WORKDIR /app/
49
50 # create d2s user
51 RUN useradd d2s
52
53 # copy over virtual environment
54 COPY --from=conda-env-base --chown=d2s:d2s $CONDA_ENV_PATH $CONDA_ENV_PATH
55
56 # update path to include venv bin
57 ENV PATH="$CONDA_ENV_PATH/bin:$PATH"
58
59 # entwine and proj libs
60 ENV LD_LIBRARY_PATH=/usr/local/lib
61 ENV PROJ_LIB="$CONDA_ENV_PATH/share/proj"
62
63 # install curl and gdal
64 RUN apt-get update && apt-get install -y curl gdal-bin && rm -rf /var/
65
```



Ln 58, Col 1 Spaces: 4 UTF-8 LF

1:11:02 / 2:54:31

Big UAS high throughput phenotyping data management and analysis using D2S online platform



Analytics

Edit video

0



Share

Promote



All From GDSL Kubernetes Computer pro >



Machine Learning for Everybody - Full Course

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



Agricultural Genome to
Phenome Initiative



College of Agriculture

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

