AG2PI COCONUT GRANT - PROJECT FINAL REPORT

PROJECT NAME	Open-Source Online Platform for UAS High Throughput Phenotyping Data Managment

PROJECT PRINCIPAL INVESTIGATOR	TODAY'S DATE	PROJECT START DATE	DATE OF COMPLETION
Jinha Jung (Associate Professor, Lyles School of Civil Engineering, Purdue University)	April 28, 2024	March 1, 2023	February 29, 2024

TEAM MEMBERS (co-PI, co-I, personnel)	COLLABORATORS
Mitch Tuinstra (Professor, Dep. Agronomy, Purdue University), **Diane Wang** (Assistant Professor, Dep. Agronomy, Purdue University), **Carol Song** (Chief Scientist, Rosen Center for Advanced Computing, Purdue University), **Jeffrey Gillan** (Data Scientist, BIO5 Institute, University of Arizona), **Mahendra Bhandari** (Assistant Professor, Texas A&M AgriLife Research), **Amir Ibrahim** (Professor, Department of Soil and Crop Sciences, Texas A&M University)	**Lan Zhao** (Research Scientist, Rosen Center for Advanced Computing, Purdue University), **Tyson Swetnam** (Research Assistant Professor, Geoinformatics/BIO5, University of Arizona), **Bryan Barker** (Bryan Barker, Senior Research Advocate, Oracle for Research)

NOTE: A shortened version of this report will be made available on the AG2PI website after any sensitive items have been removed. You will have final approval of the website version.

ACCOMPLISHMENTS

Please provide a short summary of the conclusions (both successes and failures) made from your project. Include a description of how this project will provide benefits to the agricultural genome to phenome community and, possibly, to a broader audience. You should include both qualitative and quantitative details, as necessary, to support your conclusions. Include a short accomplishment statement in non-technical language and do not include names.

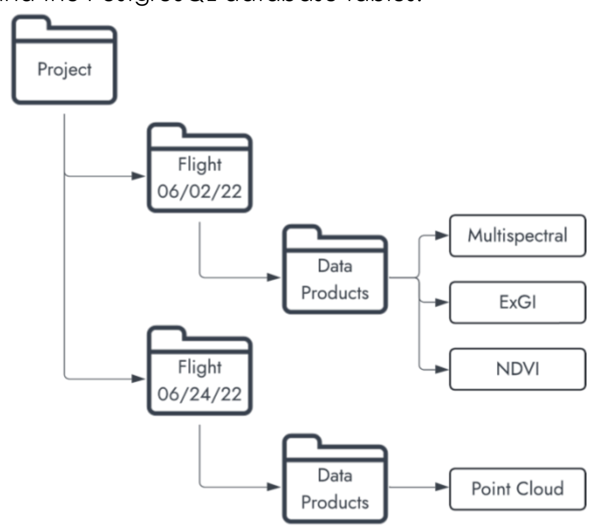
This is not a technical report. Please keep to no more than 6-8 sentences (e.g., 1-2 sentences per point, above).

Although UAS (unoccupied aerial system) HTP (high throughput phenotyping) data has demonstrated its usefulness in various agricultural applications, individual research groups have practical difficulties in managing, visualizing, analyzing, and sharing agricultural UAS big data. This project successfully developed an open-source online platform for the UAS HTP data management, 'Data to Science Engine (D2SE)', and published the platform's entire codes to the public. D2SE enables individual research groups to manage their UAS data efficiently by providing user-convenient UI/UX design and easy installation by containerization technique. Furthermore, D2SE has been deployed on federally funded cloud computing resources, such as Cyverse and Anvil, and these websites are open to agricultural researchers. D2SE has been introduced to diverse agricultural communities through multiple presentations and training workshops with an online user manual and video instruction with sample data. Given the active usage of UAS data in different research fields, D2SE is not limited to the phenotyping communities but will broadly impact any research communities using UAS data.

(HINT: You can expand sections as necessary by pulling down on the square in the lower right corner of each box)

Products

Please list any products from this project. This may include (but not limited to) publication, concept/white paper, workshop, conference presentation, website, publicly available data or pipelines, etc. Reminder: you are required to make your prodcuts available to the broader stakeholder community using standard USDA practices, open source, FAIR, or other models. Metrics may include number of participants or times accessed, etc. Include links to recordings, DOI, etc. when possible. For presentations and posters, provide authors, date, location and presentation title.

ACTIVITY / PRODUCT	**DESCRIPTION (include URL, if applicable)**	**OUTCOME / METRICS**
Establish-ment of geospatial DB schema	A DB scheme for the naturally unstructured UAS data and other component of the online platform was designed and established upon data FAIR (Findable, Accessible, Interoperable, and Reusable) principles by collaborating PIs.	**A geospatial DB scheme for UAS HTP data in a PostgreSQL server:** A database schema was designed to establish the relationships among the platform's objects (e.g., users, teams, and geospatial data). The designed current schema focused on UAS flight data but can include field-measured phenotyping data in the future. To systematically archive UAS data, the DB scheme was designed to have a hierarchy of project, flight, and UAS data products, as shown in Figure 1. The entire DB diagram of every component in the platform was implemented in a PostgreSQL relational database with PostGIS extension to enable spatial queries and storage of spatial data. SQLAlchemy Object-Relational Mapping (ORM) was used to construct the relationship between the server application's data models and the PostgreSQL database tables.  *Figure 1. Hierarchy of UAS Data Product* Existing tables in the schema were modified as features were improved based on feedback from internal testing and new tables were added to cover new features (Figure 2). For example, the project members table was revised in the database schema to provide more granular control over the permissions an individual user has while interacting with a project. New tables were added to track the progress of long-running jobs, such as file conversions and generating new data products, and to provide access controls for publicly sharing datasets. Throughout the implementation of the database diagram, a test-driven approach was taken when implementing the create-read-update-delete (CRUD) operations to verify each operation returned the expected data or properly handled invalid inputs. Currently, the database schema is undergoing further changes as we are in active development of a new feature for pre-processing field data.

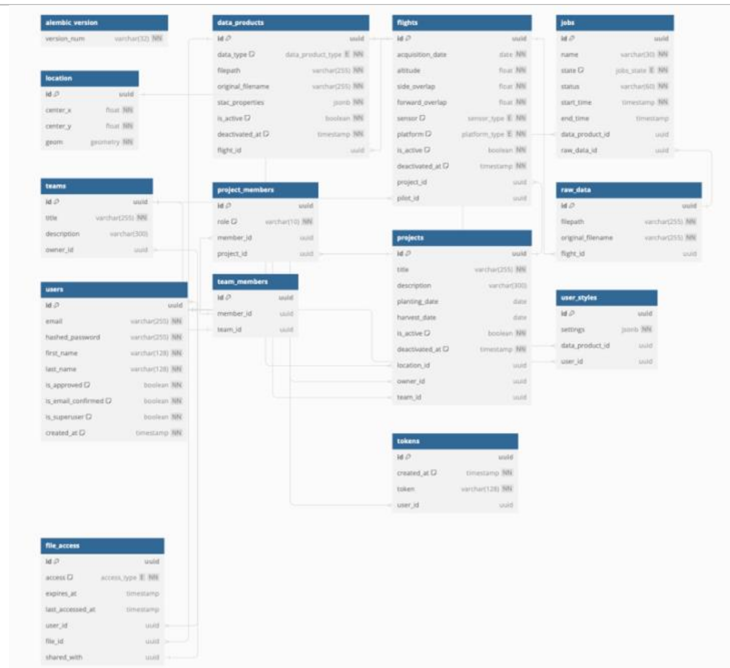
<table>
<tr>
<td></td>
<td></td>
<td>

Figure 2. Entire DB Scheme</td>
</tr>
<tr>
<td>Develop-ment of core tools</td>
<td>The core tools (management, visualization, analysis, sharing) were developed to demonstrate the potential use of the online platform, and their entire codes have been open to the public through GitHub Repo.: https://github.com/gdslab/data-to-science</td>
<td>UAS Data Management by user/team:
User authentication and authorization have been implemented in the backend server API. The server API identifies clients and determines the level of access clients have to requested data. Access to data is determined by the relationships established in the database schema. Per the schema, data access is available to 1) the data originator, 2) team members if the data is associated by the originator with a team, and 3) project members.
User and data management has been fully implemented on the platform. By default, users will not have access to any data on the site until explicit permission has been granted by the data owner. Data is added to the site by project owners, and project owners are provided controls to grant team members access to a project and its data or add individual users with accounts on the platform to a project. Project owners can designate project members with three role types: owner, manager, and viewer. Each role provides the member with a certain set of capabilities, from adding/deleting data in a project to only being able to view existing data. One of the platform's major goals is to enable the sharing of data with the public. To achieve this, file permissions at a data product level were implemented, allowing appropriate user accounts to toggle a data product from private to public. Once made public, maps of a data product can be shared with anyone, and the data product can be streamed to external applications such as QGIS. Users with the appropriate permissions can also internally share maps on the platform.

UAS Data Visualization:
(1) Outline of Projects: The platform features a map where users will be able to visually explore their projects and data products. On initial viewing of the map, users will see markers at the position of each project. Marker clustering was implemented to group together nearby projects and help users intuitively understand the spatial distribution of their UAS products. Users have the option to choose between four freely available basemaps: Open Street Map, USGS Topo, USGS Orthoimagery, and USGS Orthoimagery with Topo. An optional environment variable can be provided to the platform to enable Mapbox's Satellite Streets basemap. The map was</td>
</tr>
</table>

implemented using the popular, open-source JavaScript library Leaflet.

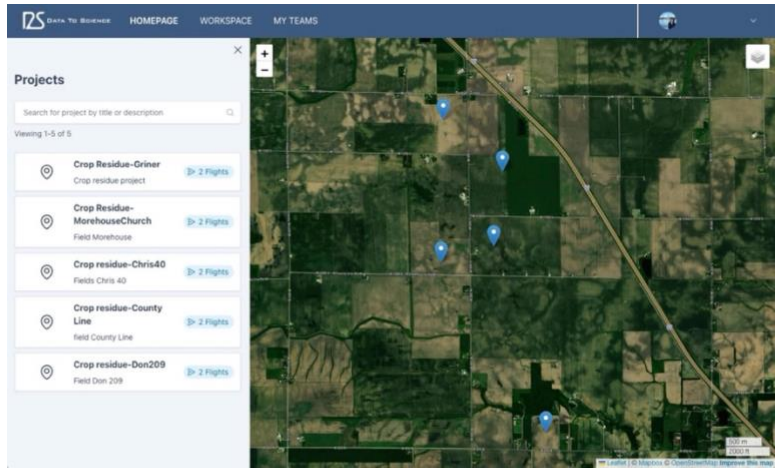


Figure 3. Outline of Projects

(2) Raster Data Visualization: After users have selected a specific project on the map, they are able to visualize the project's uploaded UAS products (e.g., ortho and DSM, Digital Surface Model) and any derived data products (e.g., Normalized Difference Vegetation Index, NDVI). Multiple symbology controls are provided, allowing the user to switch between different color ramps, band compositions, and min/max value settings on the fly. To make this possible, GeoTIFF UAS data products are converted to Cloud Optimized GeoTIFFs (COG) during the initial upload process. The COG format internally contains tiles and overviews, and it enables HTTP Get Range requests. The former eliminates the need to generate and maintain static tiles and overviews at different zoom levels, while the latter makes it possible to request only the portion of the GeoTIFF that is required for visualizing it at the map's current extent and zoom level. This can reduce the amount of data sent to the client when viewing a data product on their map.

To further reduce the load on the client when viewing large data products and to make on-the-fly color schemes, band composition, and scale changes, we self-host an open-source dynamic tile server named TiTiler. TiTiler ingests the COG along with the desired symbology properties. The resulting tiles it dynamically generates are displayed on the map using Leaflet's standard TileLayer object.

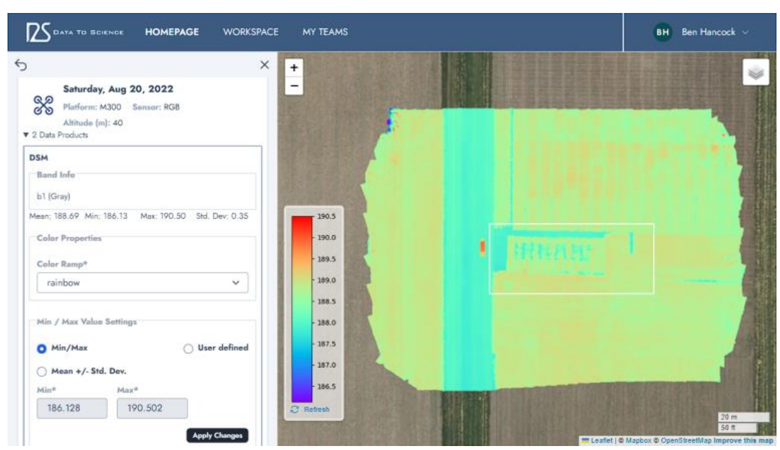


Figure 4. Raster Data Visualization

(3) Point Cloud Visualization: In addition to GeoTIFF data products, the platform can visualize point clouds using the open-source library Potree. When users upload a .las or .laz point cloud data product, it is converted to a Cloud Optimized Point Cloud (COPC). Similar to the COG format, this enables the data product to be streamed from a single source. When a user selects a point cloud data product for visualization, an embedded document containing the Potree application opens. Potree ingests the COPC data product and renders a 3D view of the point cloud using WebGL. The Potree interface provides the user with tools for performing a variety of measurements and manipulating the view of the point cloud.

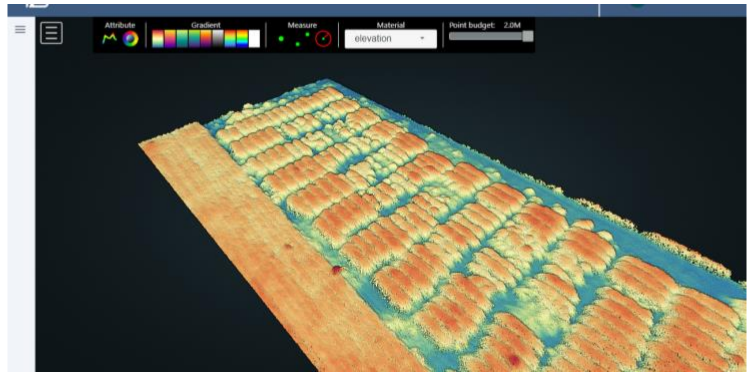


Figure 5. Point Cloud Visualization

UAS Data Analysis:

(1) Map tools: A map tool is available for performing a bi-temporal comparison of raster data products from the same location but collected on different dates. When activated, the map is split vertically into two sections with a slider control in the center of the map. Each half of the map has input controls for selecting a data product from any flight within the currently active project. The vertical slider control is used to reveal or hide one of the overlaying data products.

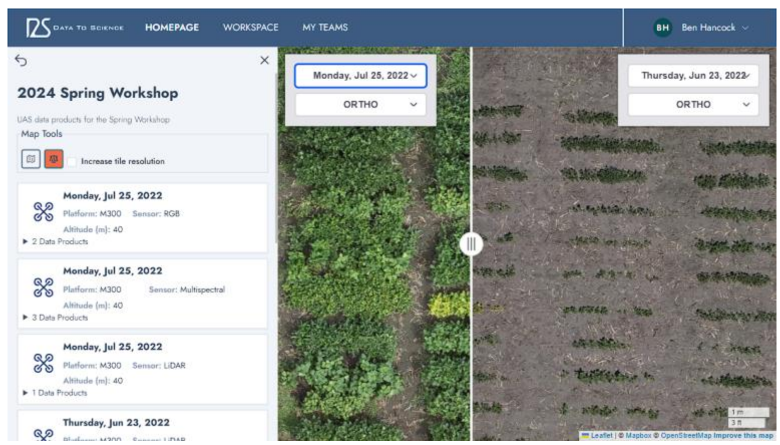


Figure 6. D2SE Map Tool

(2) Toolboxes: Users can derive new data products for analysis from their uploaded ortho data products. If the ortho data product was collected with an RGB sensor, the user would have access to a toolbox with an Excess Green Index (ExG) tool. If a multi-spectral sensor was used, the toolbox would contain the ExG tool and an NDVI tool. Depending on the size of the input data product, the tools can take from seconds to minutes to run. These tasks are offloaded to a background worker process on the platform's server, and the status of each task is tracked in the database. Users are free to leave the site after starting one of the tools. Once a tool has finished running, the derived data product will be made available for visualization on the map.

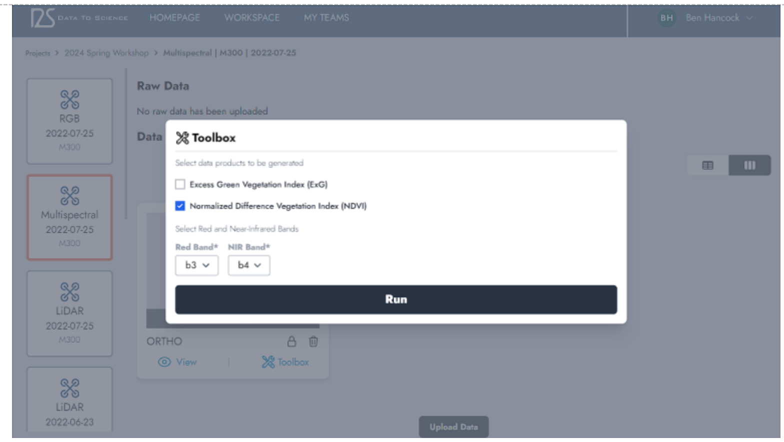
<table>
<tr>
<td></td>
<td></td>
<td>


Figure 7. D2SE Toolbox

UAS Data Sharing:

Users may share a data product internally with other users or they can choose to make a data product public, enabling anyone inside or outside of the platform to access it. By default, a data product will only be accessible internally to users that have the correct permissions through their project member status. A user can take a "snapshot" of their map in the form of a shareable link that contains information about the current symbology settings and share it with another project member. The project member receiving the shared link will need to be signed in to the platform to view the shared map.

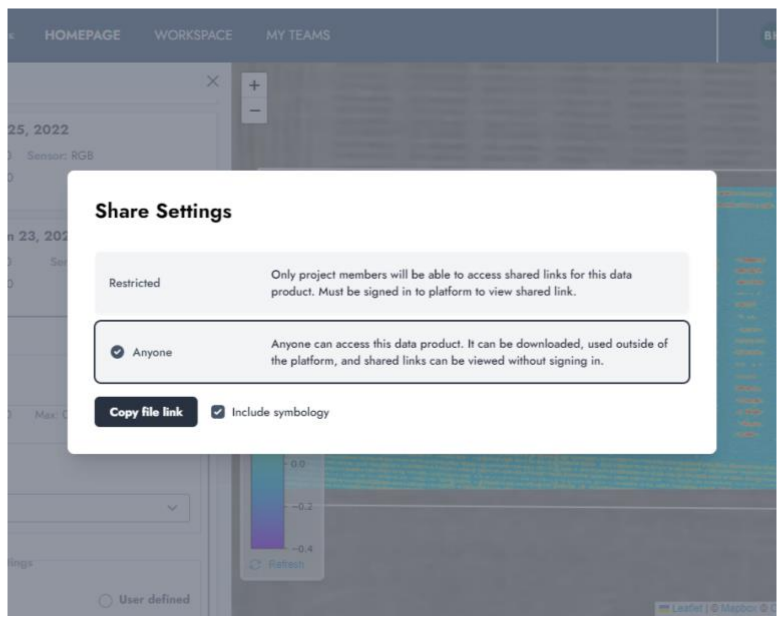
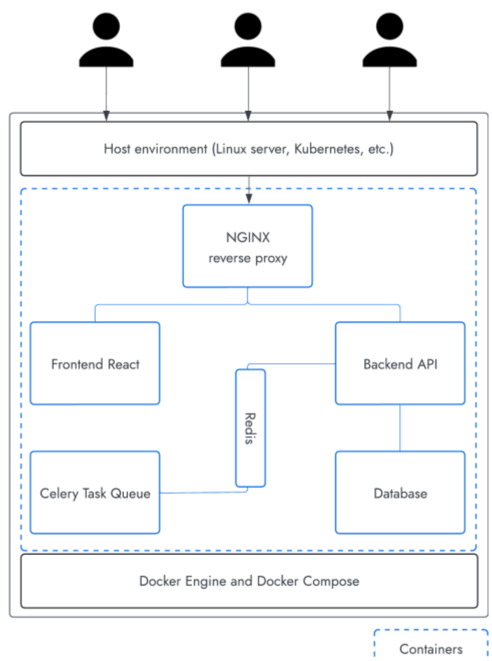


Figure 8. Data Sharing

A project owner can make data products within their project publicly accessible. In this scenario, when a map link is shared with someone, they will not need to sign in to the platform before viewing it. Anyone with the link will be able to view the shared map for this specific data product. Public data products can also be streamed to external applications, such as QGIS, where further analysis can then be performed. Every request made for a data product goes through an authorization check on the server before allowing any data through. Once a public data product has been switched back to private, all access outside of the normal permissions structure will be denied.
</td>
</tr>
<tr>
<td>Develop-
ment of
an online
platform</td>
<td>An online platform was developed by integrating the DB scheme and the core tools, and it's entire code are open to the</td>
<td>

The source codes of the online platform with UI/UX design:

The developed tools were integrated systematically into one comprehensive online platform (Data to Science Engine, D2SE) based on the designed DB scheme. The user interface for D2SE was
</td>
</tr>
</table>

	public through GitHub Repo.: https://github.com/gdslab/data-to-science Furthermore, an online platform (http://ps2.d2s.org) is open for the any researchers.	developed with the open-source JavaScript library React. The mapping capabilities on the site come from the Leaflet JavaScript library and its community-developed plugins. Data communication between the D2SE client application and the backend application occurs entirely over API requests. FastAPI, a Python framework, was used to develop the backend application. The developed API adheres to the OpenAPI standard making it interoperable with tools such as Swagger UI, a tool for generating API documentation. For long-running tasks, such as converting an uploaded point cloud to the Cloud Optimized Point Cloud (COPC) format, the tasks are queued with Celery and sent to worker processes. The database schema was implemented in a PostgreSQL database with the PostGIS extension. Networking between the frontend and backend application is handled by an NGINX reverse proxy service. To ease the deployment process for those unfamiliar with these individual products, the entire platform was designed to run inside Docker containers. The containers have the exact environment and configurations required by each application component. Individuals can either build the Docker containers locally or use pre-built Docker images hosted on DockerHub. Another Docker tool, Docker Compose, is used to orchestrate building, starting, and stopping the containers.  *Figure 9. The online platform, D2SE, Structure* **The platforms operated for individual research communities**: Considering data security, D2SE can be operated within individual research communities. As of the report submission, two online platforms (websites) are operated. One is for an agricultural research community at Purdue University (http://ps2.d2s.org), and the other is for a research community (Departments of Agricultural Science and Engineering, Environmental Engineering, and Food and Animal Science) at Tennessee State University (http://tsu.d2s.org). Particularly, the former is currently opened to not only Purdue University's researchers but also any researchers to foster a broad user base open-source ecosystem.
Deploy-ment on cloud platforms	Deploy the containerized online platform through federally-funded computing platforms	**D2SE has been successfully deployed** on Tombstone VM part of the **CyVerse** ecosystem (https://cyverse.d2s.org/) and on **Anvil** Composable Subsystem with Kubernetes (https://d2s.anvilcloud.rcac.purdue.edu/).

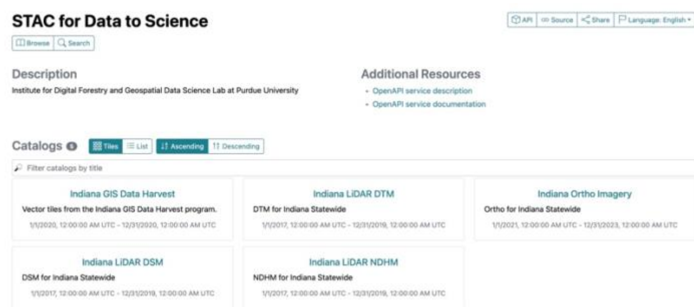
Publicly available data	The STAC specification for Indiana statewide geospatial datasets in CoG formats has been published so that researchers can easily search and download Indiana data.	**STAC catalog for Indiana statewide geospatial data**, including high-resolution ortho imagery, lidar DTM, DSM, and NDHM, and harvest data are available through the below link: https://stac.digitalforestry.org/  *Figure 10. Published STAC data*
Presenta-tions	The ideas and implementation of the online platform have been consistently presented to the wide range of professional groups through different conference and events. Additionally, the online platform has been internally used to conduct research using UAS data during its development. The research outcomes have been presented through podium and poster presentations.	(1) Hancock, B., Qian, C., & Jung, J. Digital Forestry Online Platform Development. Digital Forestry Retreat, West Lafayette, IN., Aug. 1, 2023. (2) Hancock, B., Jung, J., Fei, S., Tuinstra, M., Yang, Y., Wang, D., Song, C., Gillan, J., Bhandari, M., Ibrahim, A., Zhao, L., Swetnam, T., & Barker, B. Open-Source Online Platform for UAS HTP Data Management. ASA, CSSA, and SSSA International Annual Meeting, St. Louis, MO., Oct. 30, 2023. (3) Jung, M., Carpenter, J., Fei, S., & Jung, J. Effects of Location Error, Point Density and Acquisition Period on Aboveground Biomass Estimation Using UAS Lidar Point Clouds in Hardwood Forest. ASA, CSSA, SSSA International Annual Meeting, St. Louis, MO., Oct. 31, 2023. (6) Hancock, B. & Jung, J. Data to Science Engine: Online collaborative platform to integrate geospatial data with science. Purdue GIS Day, West Lafayette, IN., Nov. 9, 2023. (4) Hancock, B. & Jung, J. Open-Source Online Platform for UAS HTP Data Management. Indiana GIS Day, Indianapolis, IN., Nov. 15, 2023. (5) Jung, M., Carpenter, J., Fei, S., & Jung, J. Aboveground biomass estimation of norther hardwood forest using UAS lidar. North American Plant Phenotyping Network (NAPPN) Annual Conference, West Lafayette, IN., Feb. 13-15, 2023. (6) Jung, M., Hancock, B., & Jung, J. Data to Science: UAS data management system for agricultural scientists. Joint Poster Session of College of Agriculture and College of Engineering, Purdue University, West Lafayette, IN., Mar. 29, 2024. (7) Jung, J., Fei, S., Tuinstra, M., Yang, Y., Diane Wang, Carol Song, Purdue Univ. (United States); Jeffrey Gillan, The Univ. of Arizona (United States); Mahendra Bhandari, Amir Ibrahim, Texas A&M Univ. (United States); Lan Zhao, Purdue Univ. (United States); Tyson Swetnam, The Univ. of Arizona (United States); Bryan Barker, Oracle (United States); Minyoung Jung, Ben Hancock. Data to science: an open-source online platform for managing, visualizing, and publishing UAS data. SPIE Defense and Commercial Sensing, Washington D.C., Apr. 22-23, 2024.
Workshop and Seminar	To encourage the phenotyping research community to use D2SE, one training workshop has been held withing the project excution period and additional training workshops/seminars have been held as of the report submission.	**Workshop at NAPPN Annual Conference:** A four-hour training workshop was held at NAPPN Annual Conference (Feb. 12, 2024) to demonstrate how to deploy and use D2SE. **Additional workshops/seminars for professional communities**: After the successful D2SE publishment, additional workshops/seminars were given to the agricultural research communities, such as AG2PI Field Day (Apr. 24, 2024) and D2SE Workshops for Purdue Graduate Students (Mar. 2024), as of the final report submission.

		 *Figure 11. D2SE Workshop*
Educa- tional Material	For anyone who are interested in using D2SE but cannot attend workshops, a user manual (document and video) with sample data have been published.	**Online User Manual:** https://docs.gdsl.org/data-to-science-engine **Video Manual:** https://www.youtube.com/watch?v=Kzc8ZHK2NCI&t=8428s  *Figure 12. Educational Materials:* *(left) online user manual (right) video manual on YouTube* **Sample Data:** Multitemporal agricultural UAS data are available through these links: https://nappn.d2s.org/sample_data/project_boundary.zip https://nappn.d2s.org/sample_data/sample_data_for_uploading.zi p https://nappn.d2s.org/sample_data/additional_sample_data.zip **Website for experience:** https://workshop.d2s.org/

Audience

With whom has this work been targeted to and shared? Please describe how this project and its products have been disseminated to a community of interest. Include any outreach activity or information sharing as well as training or professional development opportunities provided in this project.

An online platform, Data to Science Engine (D2SE), is a straightforward and comprehensive UAS data management system and has been developed for any phenotyping communities that are using UAS HTP data but are having practical difficulties handling them in this project. D2SE is an open-source project; therefore, its entire code has been publicly available through GitHub Repo. (https://github.com/gdslab/data-to-science), and the platforms are deployed through two federally funded cloud platforms, such as Cyverse (https://cyverse.d2s.org/), and Anvil (https://d2s.anvilclou d.rcac.purdue.edu/). The outcomes of this project have been introduced in different academic events, and its training workshops and seminars have been held several times. Furthermore, for those who cannot attend the D2SE workshops/seminars, the online user manual as well as video instruction are open to the public (https://docs.gdsl.org/data-to-science-engine, https://www.youtube.com/watch?v=Kzc8ZHK2NCI&t=8428s) with sample data and a demonstration website (http://workshop.d2s.org). Through such outreach activities, we have been consistently broadening D2SE's user community and thus fostering D2SE's sustainable open-source ecosystem.