AG2PI SEED GRANT PROPOSAL

Title of Proposal:
Open-Source Online Platform for UAS High Throughput Phenotyping Data Management

Lead PI (Name, Title, Affiliation(s), email)

- Jinha Jung, Assistant Professor, Lyles School of Civil Engineering, Purdue University, jinha@purdue.edu

Co-PI (Name, Title, Affiliation(s), email)

- Mitch Tuinstra, Professor, Department of Agronomy, Purdue University, drmitch@purdue.edu
- Diane Wang, Assistant Professor, Department of Agronomy, Purdue University, drwang@purdue.edu
- Carol Song, Chief Scientist, Rosen Center for Advanced Computing, Purdue University, cxsong@purdue.edu
- Jeffrey Gillan, Data Scientist, BIO5 Institute, University of Arizona, jgillan@arizona.edu
- Mahendra Bhandari, Assistant Professor, Texas A&M AgriLife Research, mahendra.bhandari@ag.tamu.edu
- Amir Ibrahim, Professor, Department of Soil and Crop Sciences, Texas A&M University, aibrahim@tamu.edu

Collaborator (Name, Title, Affiliation, email):

- Lan Zhao, Research Scientist, Rosen Center for Advanced Computing, Purdue University, zhao4@purdue.edu
- Tyson Swetnam, Research Assistant Professor, Geoinformatics/BIO5, University of Arizona, tswetnam@arizona.edu
- Bryan Barker, Senior Research Advocate, Oracle for Research, bryan.barker@oracle.com

Grant Administrator:

- Jennifer Campbell, Senior Sponsored Funding Administrator, Sponsored Program Service, Purdue University, (765) 494-6090, jstyer@purdue.edu

Keywords:
Unoccupied Aircraft System (UAS), High Throughput Phenotyping (HTP), Big Data, FAIR Principle, Open Data Science

1. Objectives/aims

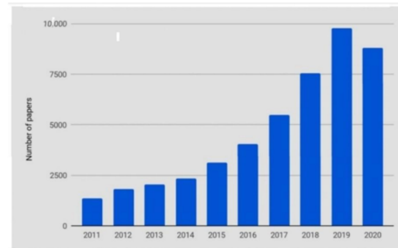

Figure 1. A number of articles published on UAS applications for agriculture (Rahman et al., 2021)

Recent advances in Unoccupied Aerial Systems (UAS) and sensor technology are now making it possible to accurately assess overall crop health status with fine spatial and high temporal resolutions at a relatively low cost. When UAS is appropriately equipped with sensors, these platforms enable fast and accurate data collection throughout the growing season. For these reasons, the UAS-based High Throughput Phenotyping (HTP) system is becoming a standard tool in plant science research as it can provide more consistent phenotypic measures and seamless coverage of the whole experimental field. UAS offers an innovative opportunity to develop the HTP system for precision agriculture, including crop yield prediction (Ashapure et al., 2020), plant-level phenotyping (Oh et al., 2020), and crop precision management (Bhandari et al., 2021) for field-level production, to name a few. Accordingly, the number of research articles on the use of UAS for agricultural applications has grown exponentially, as shown in Figure 1. As plant scientists increasingly gain access to tools for collecting big UAS HTP data, there is a growing need to generate biologically-informative, quantitative phenotypic information from the collected geospatial data (Jung et al., 2021). The massive volume of geospatial data generated by the research scientists and lack of software packages customized for processing these data make it challenging to develop transdisciplinary research collaboration around these data. This project aims to address the disengagement between big data and agricultural research scientists by developing an open-source online platform for big UAS HTP data management. The online platform will serve as a "virtual center" for 1) managing and visualizing massive volumes of UAS HTP data, 2) promoting active discussion between scientists and engineers without geographical limitations, and 3) communicating research findings to the public. The proposed online platform is expected to serve as a one-stop shop for accessing and analyzing UAS HTP data. We will implement the FAIR (Findable, Accessible, Interoperable, and Reproducible) principle (Stall et al., 2019) as a fundamental architecture so that the online platform will cultivate a synergetic collaboration ecosystem between scientists and engineers by providing an open online platform for transparent communication. The open-source online platform will adopt a bottom-up approach to build a sustainable ecosystem in the long run. The initial phase of the project will work with crop research scientists of four major row crops: (1) corn (Dr. Tuinstra), (2) soybean (Dr. Wang), (3) wheat (Dr. Ibrahim), and (4) cotton (Dr. Bhandari). The smaller initial group will help us make progress quicker and demonstrate a successful project, which will provide a firm foundation to pursue a bigger grant from federal agencies such as NSF and USDA for a larger national center. The proposed project consists of four major tasks: 1) Design a geospatial database for the UAS HTP data, 2) Develop core tools for big data management, visualization, and analysis, 3) Develop an online platform to integrate the geospatial database and the tools, and 4) Implement the developed online platform in cloud platforms.

<u>Task 1: Establish a geospatial database for UAS HTP data</u>

UAS HTP data collected are unstructured in nature. Managing and exploring unstructured data, especially when the volume of data is increasing exponentially, can be a daunting and time-consuming task for users with no advanced computing background. PIs will collaborate to develop a geospatial database schema so that the geospatial database can 1) enhance data accessibility and reusability, 2) reduce redundant data storage, and 3) support additional data-driven tool development for other tasks. PostgreSQL, an open-source RDBMS (Relational Database Management System), will be used as a database engine as it can be extended using PostGIS to handle geospatial data natively. Once the database schema is defined and implemented in a PostgreSQL server, additional tools will be developed to query data from the database. Application Programming Interface (API) that follows the OpenAPI (https://www.openapis.org/) standard will be developed using the FastAPI framework (https://fastapi.tiangolo.com/) so that UAS HTP data are FAIR. PIs of the project will collaborate closely together to accomplish this task since input from scientists is crucial for the database to be interoperable.

Task 2: Develop online tools for data management, visualization, and analysis

Using the geospatial database schema developed in the first task, we will develop core tools using Free and Open Source Software (FOSS). Three core tools will be developed to demonstrate potential software development scenarios in the online platform. UAS HTP data acquired from the PIs will be used for the demonstration. **Core tool 1 – User/Group and Data Management:** User and group management will be implemented as a first step so that we can control access to data and tools efficiently. We will also create a data upload tool so users can upload their own data into the database. The data upload tool will let users upload both raw and processed data to the online platform. We will also develop a tool to allow users to query the data to increase accessibility to the data. **Core tool 2 – Visualization:** Visualization tools will be developed so that users can quickly inspect the collected data in the online platform without downloading large files. The visualization tools will be able to handle raster (RGB, Multispectral), 3D point cloud (LiDAR and Structure from Motion), and vector (point, line, and polygon) data. The visualization tool will be developed using FOSS (Free and Open Source Software) libraries such as OpenCV, GDAL, Numpy, Scipy, and PIL and implemented in Python so that they can be easily integrated into commonly used web environments.

Task 3: Develop an online platform to integrate the geospatial database and tools

The geospatial database structure developed in Task 1 and core tools developed in Task 2 will be integrated into a web environment so that end users can easily access all the data and advanced analysis tools without downloading raw data or installing any sophisticated software. Figure 2 shows the overall design of the proposed online platform. HTML/CSS and JavaScript will be used to develop the front-end UIs, Nginx will be used as a reverse proxy to load balance web services, PostgreSQL will be used as a database engine, and the FastAPI framework will be used to communicate interactive user inputs between the front-end and back-end.

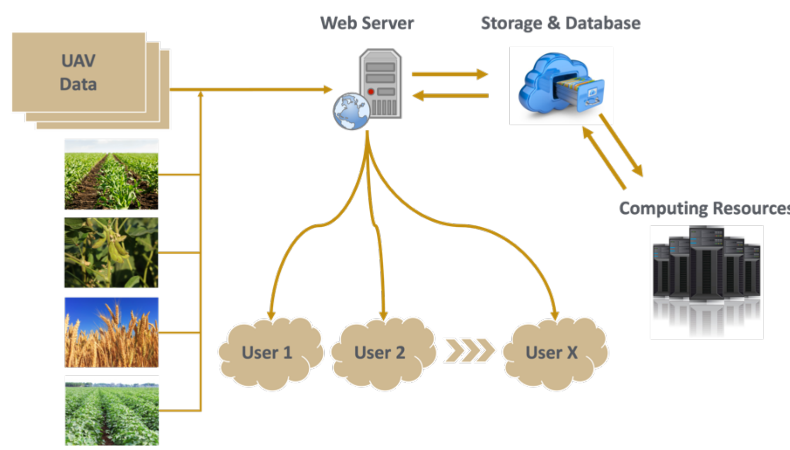

Figure 2. System design of the proposed open-source online platform

Task 4: Deploy the UAS online platform in cloud platforms

The online platform developed in Task 3 will be implemented in various computing platforms. NSF-funded open computing platforms such as Cyverse/Anvil and a commercial cloud service provider such as Oracle Cloud Infrastructure (OCI) will be used as testbeds to implement the developed online platform.

2. Furthering the aims of the AG2PI

The Agricultural Genome to Phenome Initiative (AG2PI) was created to address the challenges of agriculture, from genome to phenome, by building a cross-kingdom and multi-disciplinary research community. Among various program scope areas, this proposal addresses the priority area "Identify best practices, tools, and techniques for AG2P data sharing and storage" by creating an open-source FAIR-compliant online platform that can be used by any research group contributing to a synergistic collaboration ecosystem. Although this proposal focuses on four major row crops (corn, soybean, wheat, and cotton) in this round, the online platform developed in this project could be easily extended to manage big UAS data beyond plants, e.g., for livestock sciences. In addition, the online platform will be designed with the FAIR principle in mind and provide an OpenAPI-compliant API, which will help research scientists develop collaborative research to understand the interaction between genotype and environment by utilizing already the existing Breedbase databases (Morales et al., 2022).

The success of this project will be evaluated by a number of funding proposals submitted to external funding agencies and the awarded funding amount leveraged by this seed grant. The PIs of the proposal are confident that this project will provide great visibility in the national and global HTP community, as an open-source platform to manage big geospatial data is often listed as a top priority. Research findings and the proof-of-concept online platform will be presented in professional meetings, including the IEEE Geoscience and Remote Sensing Society Symposium, American Society of Agronomy, Super Computing, and published in peer-reviewed journals such as IEEE Transactions on Geoscience and Remote Sensing, Computers and Electronics in Agriculture. In addition to the publications and professional meeting presentations, online platform users' activities will be logged and analyzed to measure the impact on the research community and the general public.

3. Expected outcomes & deliverables

This project is expected to generate multiple outcomes that will impact the AG2PI community. All outcomes and deliverables will be shared with the public via a public GitHub repository.

3.1. Geospatial database schema: The geospatial database schema developed in Task 1 will serve as a potential data standard to store big UAS HTP data. The database schema will be critical to make the UAS HTP data to be FAIR. The developed geospatial data schema will be shared with the public in the form of a database diagram.

3.2. Programming source code of core tools: Python will be used as the main programming language to develop core tools described in Task 2. All core tools will be developed using FOSS libraries. The source code of the core tools will be shared with the public via GitHub so that not only anyone can use the code to process UAS HTP data on their own machine but also anyone with programming skills can contribute to the code base.

3.3. Source code of the online platforms: Online platform development involves multiple components such as server configuration, front-end UI, middleware, backend services, and scripts for deployment and testing. A tutorial document will be developed to demonstrate how each component is configured and implemented. The tutorial will be shared with the public via GitHub.

In addition to the deliverables listed above, PIs will provide an online workshop in the AG2PI workshop series. The workshop will provide an opportunity to disseminate the findings of the project and serve as training material for potential users of the online platform.

4. Qualifications of the project team

Dr. Jinha Jung (PI), Assistant Professor of Civil Engineering, is the Geospatial Data Science Lab Director at Purdue University. Dr. Jung is noted for his work in high-performance computing for advanced remote sensing data processing and WebGIS software development using Free and Open Source Software (FOSS). He has developed various cloud-based data portals for managing big geospatial data, including UAV and satellite images. Dr. Jung will take the lead in developing an open-source online platform for UAS HTP data management.

Dr. Mitch Tuinstra (Co-PI), Professor of Plant Breeding and Genetics and Wickersham Chair of Excellence in Agricultural Research at Purdue University, studies how crop plants grow in stressful environments. His research focuses on identifying genes and genetic resources that contribute to improved crop performance in stressful environments.

Dr. Diane Wang (Co-PI), Assistant Professor of Agronomy, is an ecophysiological modeler with formal training in plant breeding and genetics. Her research focuses on understanding the processes that under genotype by environment interactions in various crop species, including rice, cotton, and soybean. Dr. Wang has been recently collaborating with PI Jung on evaluating diverse soybean using UAS HTP technology.

Dr. Carol Song (Co-PI), Chief Scientist of the Rosen Center for Advanced Computing at Purdue University, is an expert in advanced computing cyberinfrastructure (CI), data frameworks, and science gateways. She has extensive experience leading large CI projects, serving as PI for the NSF Anvil system, NSF GABBs, and GeoEDF projects, and collaborating with domain science researchers to develop science online platforms. Dr. Song will connect the team to national and campus CI resources, and lead the research software engineering team at RCAC to integrate the proposed open-source UAS platform with a high-performance and cloud computing ecosystem.

Dr. Jeffrey Gillan (Co-PI) is a Data Scientist with the BIO5 Institute at the University of Arizona and focuses on the NSF-funded Cyverse project. He has 11 years of experience developing aerial imagery processing and analysis pipelines, with most of the work centered around livestock agriculture in pasture and rangelands. Dr. Gillan will lead the implementation of the online platform on the Cyverse Kubernetes cluster.

Dr. Mahendra Bhandari (Co-PI), Assistant Professor at Texas A&M AgriLife Research, has significant experience in UAS-based applications in crop physiology and precision agriculture research. Dr. Bhandari will provide the UAS data collected over the cotton field, test the developed tools and provide feedback regarding their applicability in HTP, crop physiology, and precision agriculture research.

Dr. Ibrahim (Co-PI) leads the Small Grains Breeding and Genetics Program at College Station, Texas. The primary areas of research of his program include biotic and abiotic stress tolerance and end-use quality characteristics in small grain crops, hybrid wheat, and HTP using UAS.

5. Proposal timeline

Task	Q2 2023	Q3 2023	Q4 2023	Q1 2024
Task 1: Establish geospatial database schema				
Task 2: Develop core tools				
Task 3: Develop an online platform				
Task 4: Deployment on cloud platform				

6. Engaging AG2PI scientific communities & underrepresented groups

Although this project focuses on four major row crops (corn, soybean, wheat, and cotton) to demonstrate a proof-of-concept online platform to manage big UAS HTP data for breeding research, the developed online platform will be flexible enough to accommodate other crops with minimal modification. The online platform will also be implemented in (1) open computing platforms such as Cyverse and Anvil and (2) commercial cloud platforms such as OCI so that any research group can adapt this technology flexibly depending on the situation of each group. We believe this will provide a level playground for everyone so that the high cost of big data management won't be a bottleneck to making progress on their research. Having an open-source online platform will also cultivate an environment where collaboration between academia and industries is encouraged. This project will also provide an opportunity to support Hispanic students as the Texas A&M University – College Station and the University of Arizona are Hispanic Serving Institutions because students from these institutions will be encouraged to get engaged in this project. We will also leverage Dr. Bhandari's involvement with another AG2PI-funded project (Facilitating community unoccupied aerial systems knowledge, communication, and data processing) to engage a wider audience from AG2PI scientific communities.

Bibliography/References cited

1. Morales, N., Ogbonna, A. C., Ellerbrock, B. J., Bauchet, G. J., Tantikanjana, T., Tecle, I. Y., ... & Mueller, L. A. (2022). Breedbase: a digital ecosystem for modern plant breeding. G3-Genes Genomes Genetics.
2. Jung, J., Maeda, M., Chang, A., Bhandari, M., Ashapure, A., & Landivar-Bowles, J. (2021). The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. Current Opinion in Biotechnology, 70, 15-22.
3. Rahman, M. F. F., Fan, S., Zhang, Y., & Chen, L. (2021). A comparative study on application of unmanned aerial vehicle systems in agriculture. Agriculture, 11(1), 22.
4. Bhandari, M., Baker, S., Rudd, J. C., Ibrahim, A. M., Chang, A., Xue, Q., ... & Auvermann, B. (2021). Assessing the effect of drought on winter wheat growth using unmanned aerial system (UAS)-based phenotyping. Remote Sensing, 13(6), 1144.
5. Ashapure, A., Jung, J., Chang, A., Oh, S., Yeom, J., Maeda, M., ... & Smith, W. (2020). Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. ISPRS Journal of Photogrammetry and Remote Sensing, 169, 180-194.
6. Oh, S., Chang, A., Ashapure, A., Jung, J., Dube, N., Maeda, M., ... & Landivar, J. (2020). Plant counting of cotton from UAS imagery using deep learning-based object detection framework. Remote Sensing, 12(18), 2981.
7. Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., ... & Wyborn, L. (2019). Make scientific data FAIR.