TETRA TECH

Enabling Crop Analytics At Scale

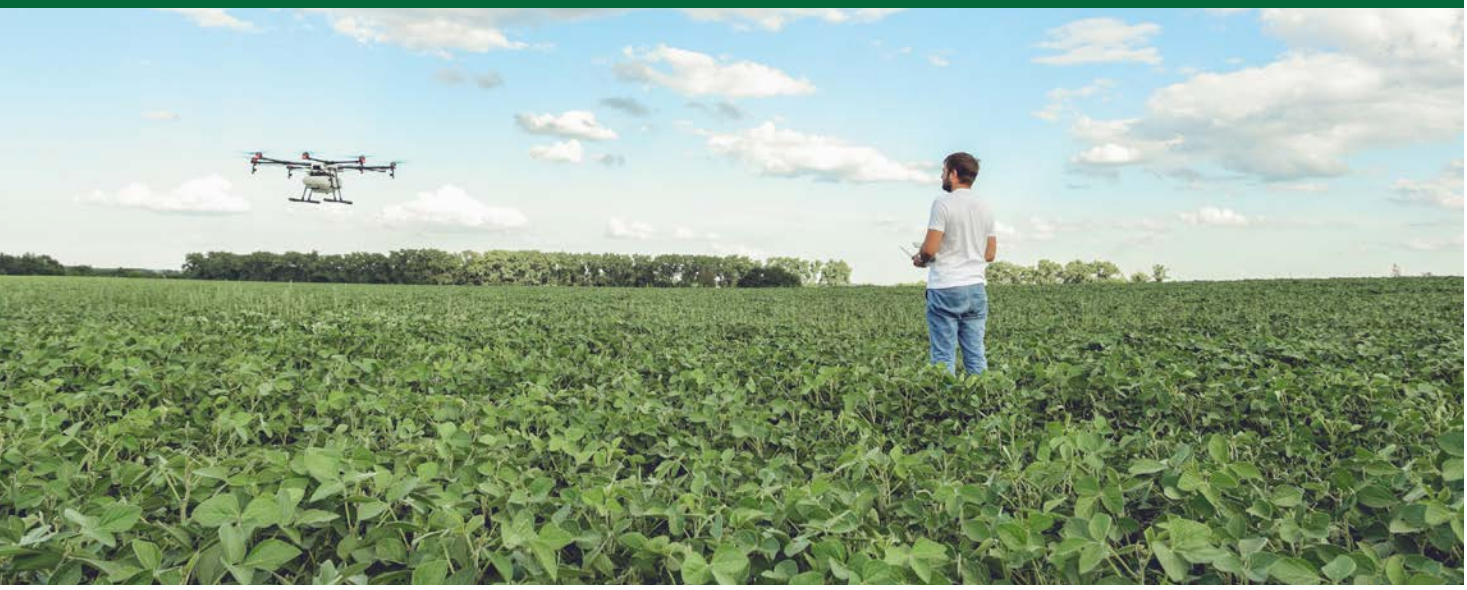
ENABLING CROP ANALYTICS AT SCALE (ECAAS)

6th Grain's Data Sharing Platform Enhances the Value of Farm Data Across Sectors

Contents

Executive Summary

One of the major challenges for remote sensing applications in smallholder settings is the lack of private sector engagement in sharing datasets that could help improve model performance. This project focused on engaging with private sector companies to pilot a scalable platform for sharing and utilizing high quality, repeatable ground data on African agriculture practices and digital support tools to support local agronomists, agricultural retailers, and farmers. We developed a Data Sharing Platform (DSP): an online tool that brings together a data production approach via our crop monitoring and survey tools in a systematic data environment. We found significant interest in using the DSP for commercial use within financial companies providing agronomic advice and loans to farmers across Africa. The DSP presents a significant opportunity for small scale businesses, microenterprises, farmers, and banks to accelerate their ability to meet the needs of their clients.



Private Sector, Sustainable Development and Data Sharing

The Sustainable Development Goals (SDGs) were adopted by all United Nations Member countries in 2015 as a universal call to end poverty and protect the planet by 2030. SDG goal 2 aims to end hunger, achieve food security, improve nutrition, and promote sustainable agriculture. The private sector is central in pursuing these goals through investment in agriculture, technology, and datasets. Large and small-scale businesses, microenterprises, wholesalers and retail stores as well as large multi-national corporations all have a role to play in lifting people out of poverty and hunger through responsible and productive investment, innovation, enhanced efficiency and employment creation (Gerpacio, 2003; Mason-D'Croz et al., 2019).

To harness private sector investment, high quality, spatially-explicit, transparent datasets are needed to ensure that every dollar spent contributes to increased agricultural productivity, improves agricultural efficiency and includes the most marginalized segments of agricultural communities (Rashed and Shah, 2020). Sustainable development goals are focused on measuring change over time and include metrics of performance, including yield, nutrition of the population, and economic equality (Brown, 2021). In this project, we focus on providing a platform where datasets and data products can be presented to encourage further collaboration and enable monitoring of key metrics that are relevant for smallholder settings.

This project coupled a data production approach via our crop monitoring and survey tools with a data sharing platform (DSP) that seeks to provide a data environment. The DSP allows individuals and institutions to 1) upload and share their agriculture data on field boundaries, crop type, planting date and yield in an easy-to-use structured database, 2) access datasets collected by other private companies via a discoverable geographic catalog of all geo-referenced datasets, and 3) provide visualization of proprietary models based on data shared within the platform to enable exploration of how data can be used. By combining digital applications that allow individuals and growers working on small plots to generate new data with a platform that displays the information, we will encourage private sector companies to provide their datasets for public use.

Clean Boundary Data for Machine Learning-Based Data Analysis

Private companies operating in African countries do not know which smallholders are using their products, since seeds and fertilizer travel through multiple distributors, retailers, and agents before being used on small (sub-hectare) plots. Lack of information and field records also keeps farmers from knowing if the inputs they used improved their yields or if a specific crop was profitable. Analysts can use machine learning-based algorithms to create new datasets that can support decision making.

The ability to visualize crop type, variety, yield and individual field boundary, visualize crop mapping/classification, and linking these locations to simple indices that provide evidence of vegetation on each field is the main selling point for commercial users. Being able to find, download and share field data, validate its accuracy and completeness in an easy, online platform brings great value to users.

Machine learning (ML) is a form of artificial intelligence that imitates how humans learn to make predictions from data. 6th Grain (6G) and partners work in data scarce environments with very few reliable datasets that enable and accelerate business decisions; as such, ML algorithms are central to producing timely and affordable datasets.

All MI Systems Have Three Basic Steps:

1. A problem statement: ML algorithms are used to make a prediction or classification which results in improved insights or better decisions. The ML algorithm uses input data, which can be labeled or unlabeled, to produce an estimate about a pattern in the data.
2. An error estimation: This function serves to evaluate the accuracy of the model and to improve it as more data becomes available.
3. A model optimization process: By understanding the accuracy requirements of the business decision, the model can be modified to better fit the data points in the training data set via weighting functions to reduce the discrepancy between the known example and the model estimate. By using accuracy thresholds, the model can be optimized to reduce cost while increasing the usability of the output (UC Berkeley, 2020).

All ML processes require high quality training data through which algorithms can be built. The problems that can be solved by ML models will increase as more data become available. A key objective of the DSP project is to expose more data for ML experts across public, provide and academic sectors. In this section we describe use cases for the data provided in the DSP and how it can be used to improve decision making in the agriculture sector.

Our engagement with commercial companies has resulted in significant interest in the platform and its capabilities, but commercial entities want to see practical uses of data sharing before they commit their own data to the platform. Discussion on the availability of boundary and crop type data with commercial partners has revealed that much more effort will need to be put into creating high quality datasets, since many agriculture companies collect only general address information on their clients and not specific field boundaries. Translating high quality geospatial data into new insights with commercial value is the key to creating large new datasets.

We therefore set up a series of demonstrations where we use data provided by commercial entities to transform the field data into information. The quality and amount of training data available is directly related to the commercial value of the ML output. As we implement new case studies, the value of participating in the DSP will become clear. Here we provide three use cases which demonstrate the utility of these data.

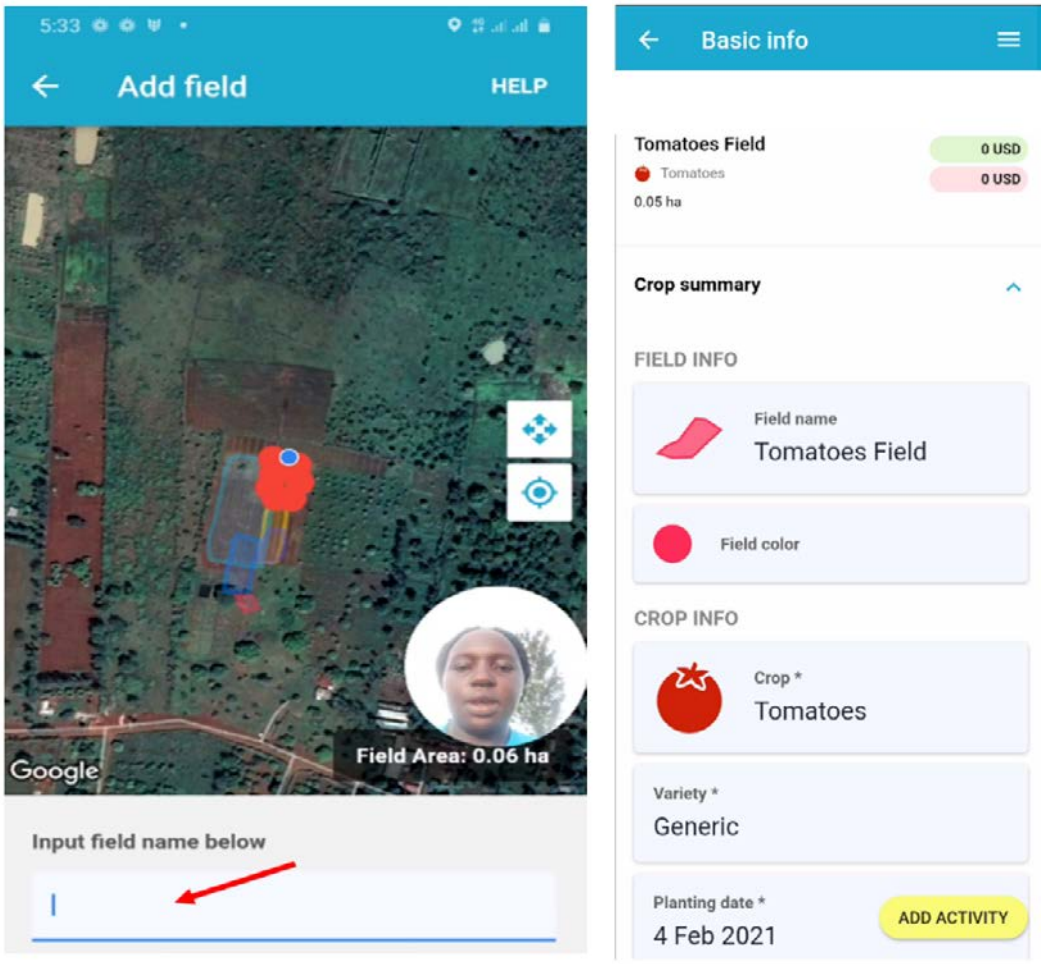


Figure 1:
FieldFocus Light Android app used to digitize field boundaries

1. Data Production in Ghana with AFAP Agents

Working with agents engaged with the African Fertilizer Agribusiness Partnership (AFAP) in Ghana, we mapped over six thousand fields using the FieldFocus Light app. The application is an Android phone system that allows the user to capture field boundary, crop type, planting date and yield information via a friendly and easy-to-use application. AFAP is a provider of technical business development services for small and medium enterprises (SMEs) and agribusiness firms engaged in agriculture in African countries. AFAP focuses on ensuring that farmers have access to agrotechnology and inputs through competitive market-driven approach. They engage with agrodealers to get agricultural inputs and output markets closer to farmers, cut transaction costs for businesses and farmers and increase impacts at the farmer level.

The development problem that AFAP addresses is that although African countries are well endowed with land, water, soil, climate, labor, forest and mineral resources, smallholder agricultural productivity is low. Agricultural productivity and income growth are not increasing fast enough to keep pace with the population growth rate to drive agricultural transformation. This is resulting in vulnerabilities among smallholder farming households and the national states. AFAP works to reduce barriers to access by building private businesses, networks, and relationships that must be put in place for the modernization of agriculture.

There are over 20,000 extension agents in Ghana who are supported by a wide range of public, private and non-profit organizations. AFAP trained 200 agents working with their agrodealers to digitize their farmers' fields, keep records of activities, manage costs, and understand trends in total production to estimate profitability and farm income. Sharing these records in networks with agronomists can help agrodealers, retailers and other participants in the agriculture value chain improve their visibility into farmer and agent activities. Field boundaries help agrodealers understand their customers through improved insights into the biotic and abiotic stresses being experienced by the crop, understanding their relative location across the agents' territory and provide advice and recommendations that are field-specific. We also focused on improving efficiency and increasing cost effectiveness for agents and agrodealers through better information on farm location and the individual need for support.

We modified the DSP to allow for the display of Points of Interest (POI) so that AFAP agents and agrodealers could view their fields and the locations of the agrodealers participating in AFAP activities together on the map. This has encouraged the agents to further engage with farmers outside their normal rounds. AFAP is also recruiting additional agrodealers on the border of Burkina Faso to the north, where there are many farmers but few dealers in the network.

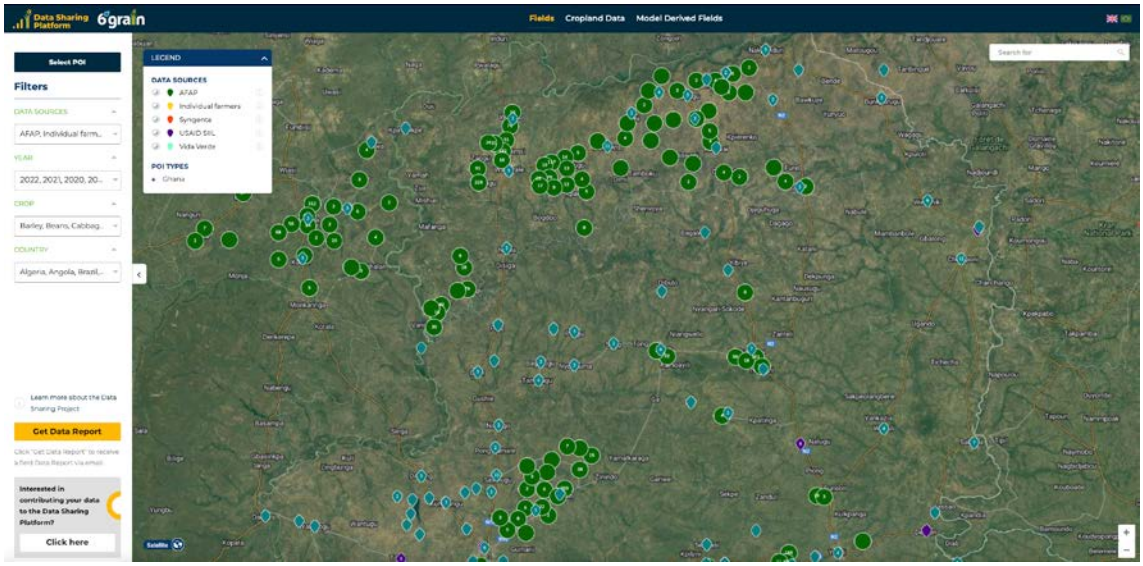
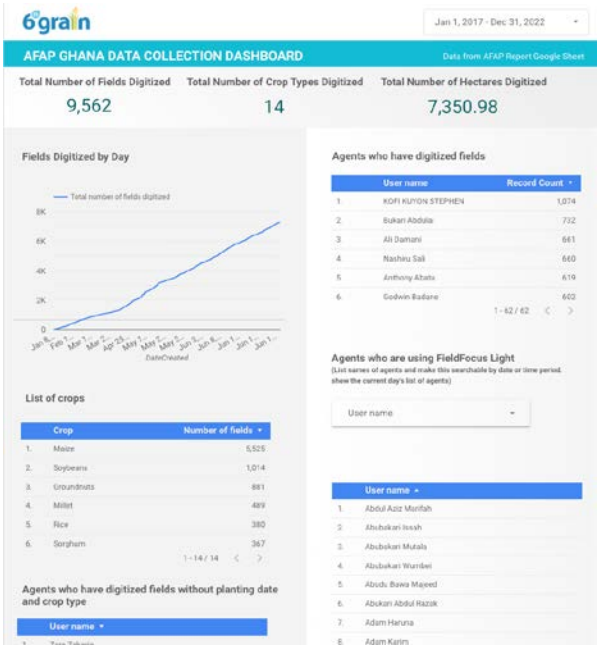


Figure 2:
DSP AFAP fields with points of interest overlaid (teal diamonds)

The agents working with AFAP continue to use FieldFocus Light, digitize the fields of the farmers they are working with and view the information in DSP. The visualization of AFAP activities is valuable to AFAP agents as it allows them to identify new markets and regions where retail and wholesale companies should be located and can be recruited into the AFAP network. The agents working with AFAP use the information to monitor crop health on their farmers' fields, to prepare advice and recommendations on actions that can improve field outcomes, determine when to visit each farmer, and continues to be useful to agents as they progress through the growing season.

Figure 3:
AFAP data collection progress since January 2022



2. Cropped Area Estimation Using MI Tools

Field boundaries can be used as training data to estimate cropped area via machine learning algorithms that integrate 12 months of satellite remote sensing information from Sentinel 2a/b with ground data to determine if an area has been cultivated in a particular year. Cropped area estimates are an essential part of understanding total production, since the area under cultivation changes year to year in response to economic, social, and biophysical drivers. Here we demonstrate cropped area estimation for Senegal in 2019 and 2021 as a foundation for total production estimation.

6G has developed a remote sensing-based operational crop mapping and production system that provides quantitative analysis of cropped area in regions dominated by small fields. There are few high quality, operational agricultural statistics available in low and middle-income countries that can be used to develop effective agricultural development and commercial programs. We seek to provide information on cropped area extent, production, in-season planting and food security assessment. Field-level data is part of that estimation to provide training and ground truth for the model.

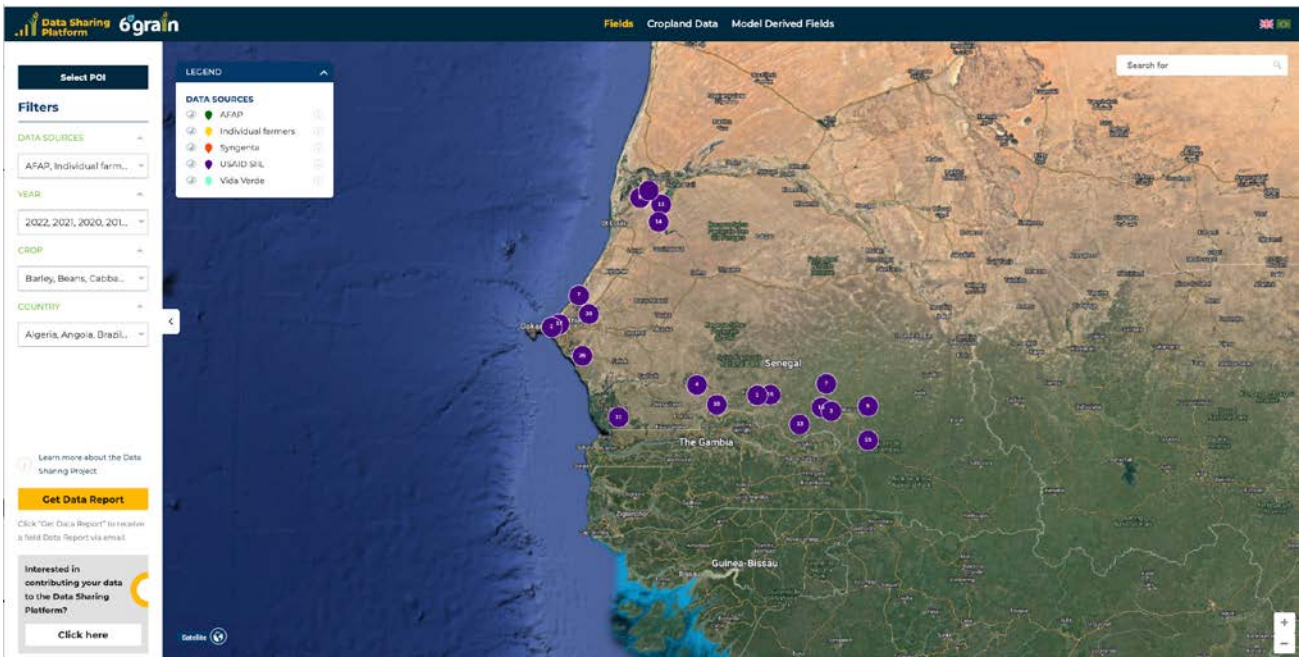


Figure 4:
Locations with boundary data in Senegal

Through a contract with USAID Sustainable Intensification Innovation Lab (https://www.k-state.edu/siil/) we gathered field data and used it to train a machine learning algorithm. The resulting data was put in a 'Cropland Area' section of the DSP, which allows the user to explore, but not download, the data.

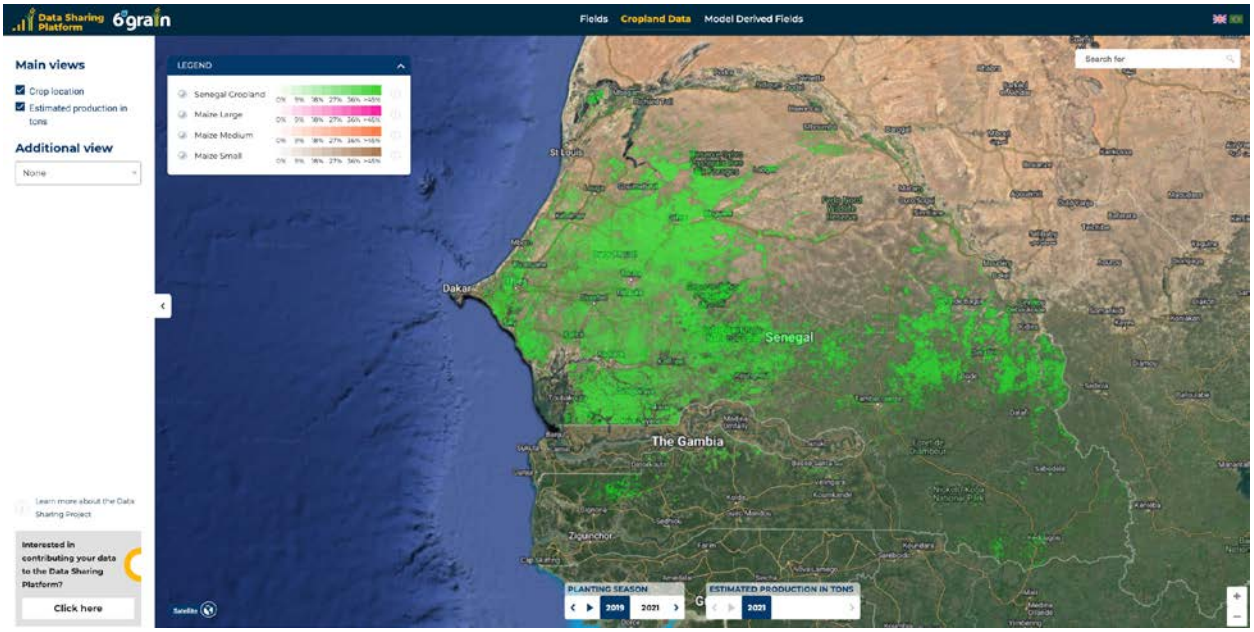


Figure 5:
Cropped area map in Senegal derived from an ML model.

Although this map was produced by ML algorithms, we do not currently have a client in Senegal who can help determine the map's accuracy. Validation of cropped area maps requires a clear understanding of the use of the maps for decision making. Since there is very little data available in the region of existing agricultural use, and the semi-arid landscape and agricultural system of Senegal makes it very difficult to identify fields.

For example, **Figure 4** shows a field in Foundiougne, Senegal, which is north of the Gambia near Sokone (13.746498, -16.450149). This millet field has been harvested, but the delineations between one field and another, and the type of agriculture conducted in the region, makes identifying and using field boundaries through an ML algorithm very challenging. Planting density is very low and although there is plenty of moisture, transforming field data into useful information for decision making, such as response to drought or investment in new agricultural input provision, requires an active partner on the ground.

This example shows how critical ground data is—specifically field outlines, crop type and planting date. Without this information obtained directly from farmers satellite remote sensing information is not as useful and cannot be transformed into actionable guidance. Estimating yield using field-level production data can be done through time series remote sensing (Maguranyanga and Murwira, 2014), which can lead to total production estimates that are useful for planning labor, products and sales activities. However, to scale these yield estimates up to a whole region the area in cultivation, we need field boundary data which excludes non-cultivated pixels like trees, bushes, bare ground, or standing water. The example above shows how important farmer-provided boundary data is to understand where the field is located and its size. Boundary data can then be used to connect total production, crop type, planting date and boundary data to provide actionable information across large areas to companies and service providers.

Figure 5:
Harvested Millet Field in Foundiougne, Senegal

3. Crop Health Monitoring Leading to Total Annual Maize Production

As mentioned above, when connected to total production information in a specific growing season, boundary data can be used to transform time series for each pixel in the cropped area map into productivity. Remote sensing information allows for the quantification of the impact of biotic and abiotic stresses, but without field-level production data, it remains as a simple anomaly and cannot be transformed into total amount of grain produced. Production information by field can be used to gain in-season intelligence on total area planted by crop, crop health conditions during the season, and yield estimates by field which would translate to total production for a country. That data can be used by the government to ensure favorable contracts if there is likely to be a grain surplus that can be sold abroad, or if a deficit will occur, to secure affordable grain from neighboring countries. Early information can be extremely helpful at the national level, and can ensure that wholesalers and retailers can anticipate constant access to affordable food for their commercial processes.

Figure 6 shows an example of crop health monitoring in Zimbabwe. The data is based on total cropped area for 2019 and 2021, which was then used as the basis for a yield model. The yield model is based on 245 fields in Zimbabwe with measured yields provided for the 2019-2020 cropping season. The crop health monitoring system uses the previous year's yield information from the best fields in the region to create a visual representation of the '6G Optimal Crop Band', which allows the user to relate the evolution of their crop to known high performing fields each day of the season. The ML algorithm identifies problems sufficiently early to allow for changes in national level purchase or sale decisions for commodities. If problems persist, yield deficits can be planned for, and crop protocols adjusted in the latter part of the season to ensure overall agricultural profitability. Observations of yields from the previous year are used to estimate the simulated field performance index for every field within the country. Because most variability in production from year to year is due to variations in growing conditions and cropped area and not changes in management (e.g., changes in variety, fertilizer use or other inputs), we use this benchmark as an input to our machine learning yield estimate.

Figure 7:
Optimal growth curve for field-level yield analysis.

Using the cropped health estimate, we combine cropped area and yield to estimate total production. We used 301 fields with yield in Mashonaland West and Mashonaland Central. The model is run three times – once over large fields that are typically large-scale commercial agriculture using hybrid seeds and irrigation, once over medium, and once over small fields. Each type of field has different productivity levels. We then use this information to estimate total maize production in the region.

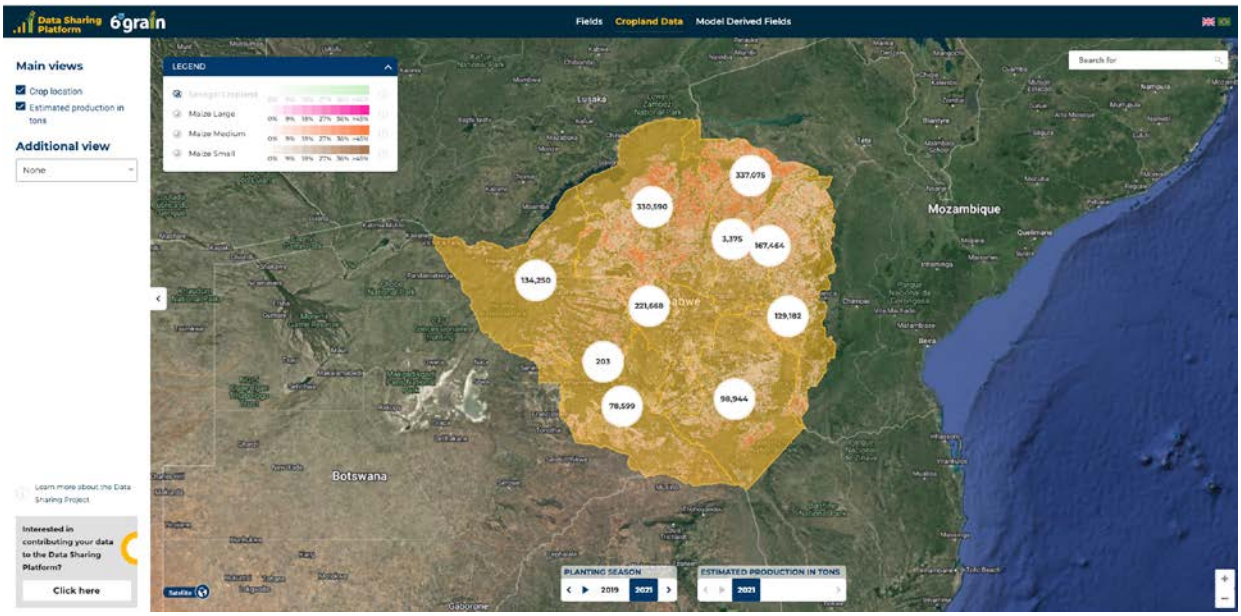
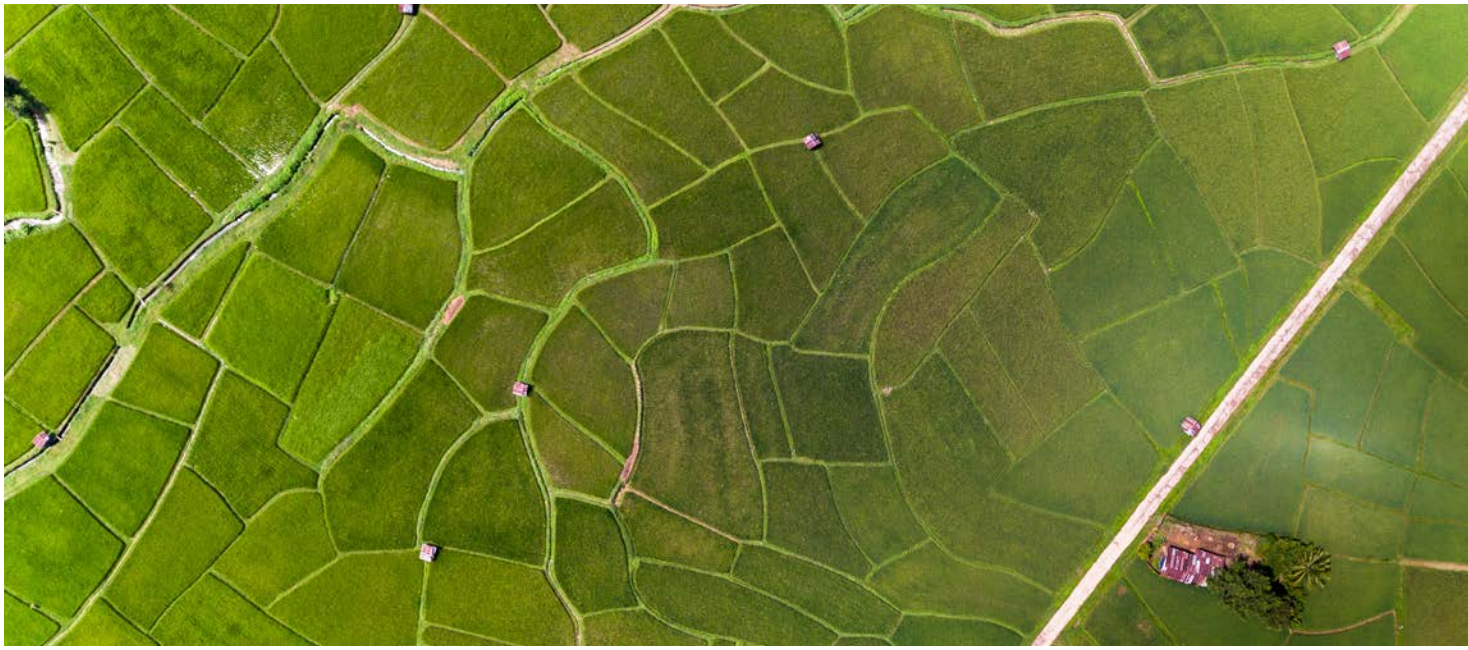


Figure 8:
Total cropped area and productivity of maize from 2021 Zimbabwe.



4. Use of Very High Resolution Imagery for Boundary Detection

In collaboration with colleague Lyndon Estes of Clark University, we have implemented an example of how very high-resolution imagery from Unmanned Aerial Vehicles (UAVs) or drones can be assist in creating training data for ML models that can automatically identify fields. Dr. Estes generated training data with hand-digitized fields using the UAV imagery, along with field digitized by their partner Farmerline in Ghana.

UAV data provides extremely detailed imagery, with the ability to see every tree and bush canopy and areas of cultivation and disturbance. With this detail comes problems of access – the size of the imagery make it very difficult to load, access and share, even with high speed internet, it can take several hours to download and display data-dense UAV files. Problems of format and interpretation are also very common, since UAV data are typically not stored in easy-to-access formats (Tamiminia et al., 2020). Similar to field boundary and metadata, private sector companies spend hundreds of thousands of dollars every year on collecting and processing data, but fail to share this data across different divisions, collaborators and partners due to the technical difficulties of moving and sharing the file. The DSP display allows for sharing and development of uses of UAV data within a modeling and analysis context. In the image below, the blue fields are places where the FarmerLine agents digitized fields by walking around them in the presence of the farmer. Additional confirmation of farming can be made with these images, and support for the display, analysis and sharing of UAV data within the DSP.

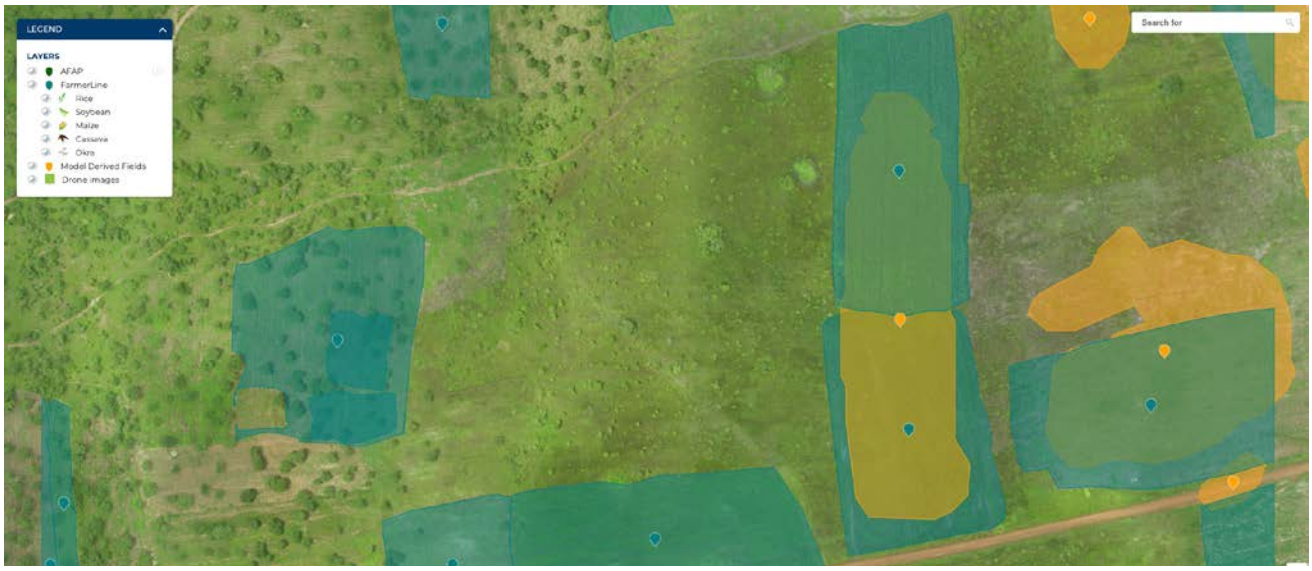


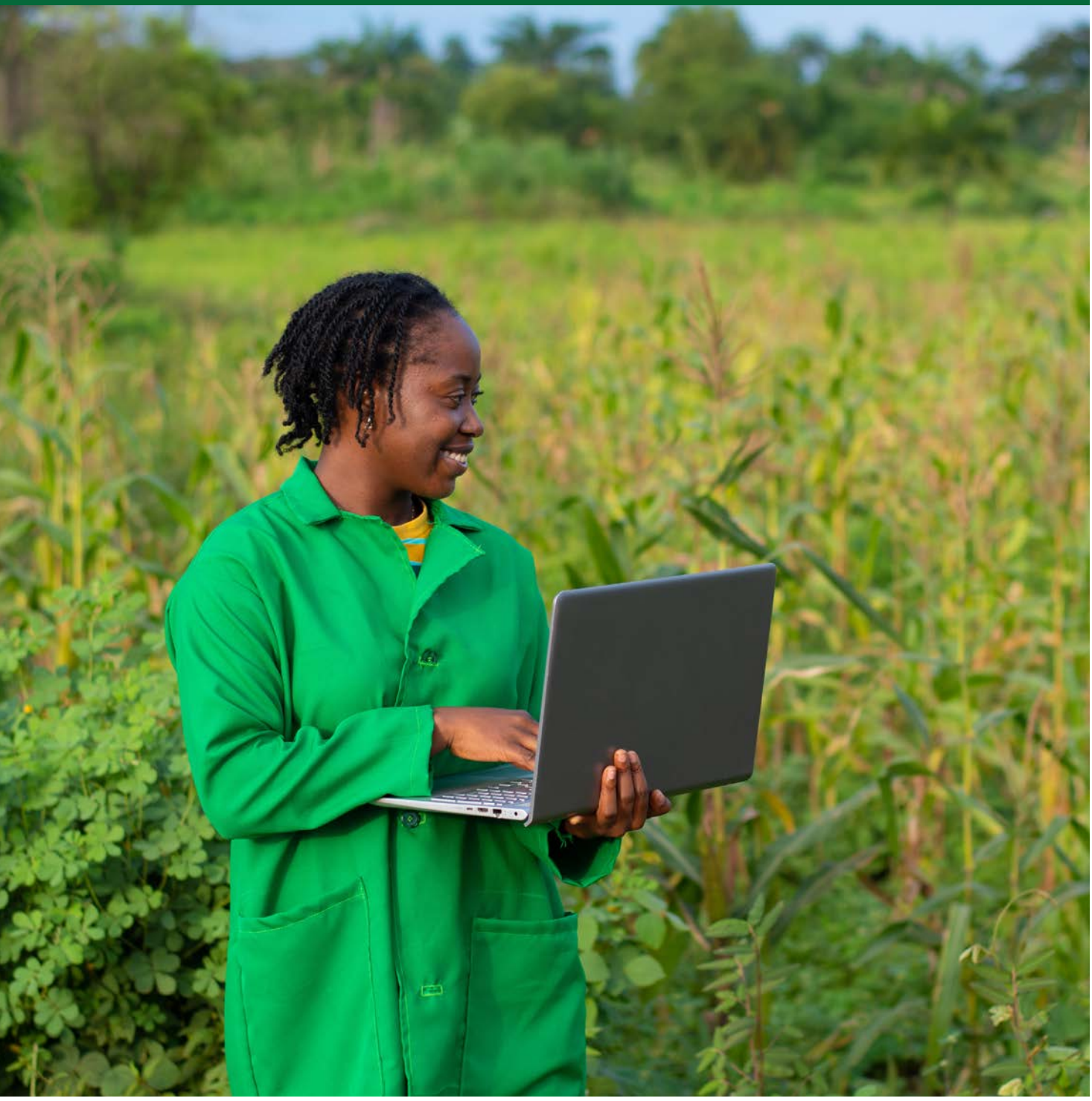
Figure 9:
Model derived fields (yellow), training data (blue) and UAV images (green image) in Ghana.

This example shows the use of a ML model to estimate field location and how challenging it is in Northern Ghana to use a model to identify field boundaries. Having sufficient training data is the primary limiting factor in this project. The data derived from the model is not sufficiently accurate for use in commercial decision making, as it seems to be unable to capture observed variability. This model is described in Estes et al. (2021).

The conclusion we draw from the integration of this project data on DSP is that even with extremely high resolution, timely data from UAVs (green boxes in Model Derived Fields section), there is no substitution for data derived from walking around a field with a farmer.

The real value of the DSP field boundary data is providing geospatial data that represents farmer activity – when the farmer cultivated the field, what they grew in it, and the ultimate productivity they obtained (if available). Given the enormous diversity of the agroecological system across different countries in Africa, only with farmer and field-specific information can we extract from optical remote sensing data information that can be generalized sufficiently to create a model. To create this data, we need to leverage the agents, loan officers, agronomists and others who visit farmers and their fields in their work to gather, clean, upload and distribute this data to create high quality, transparent data for all.

Case Studies and Recommendations



The DSP should be formally launched into production across existing data partners and various government agencies in Africa. All data contributors to this project expressed commercial interest in using the platform. Our initial estimates are that for commercial use and operational stability, 6th Grain (6G) would require about 3–4 months of testing, bug fixes and AWS setup configuration. **We therefore recommend launching an operational DSP platform around December 2022.**

As a way of assessing commercial viability and practical usefulness of DSP, 6G took on a task of exposing the platform to two companies that provide financing to farmers in Nigeria who expressed interest in using the platform. These companies are: CropIT and GT Bank. 6G had no previous relationship with these two Nigerian companies. CropIT contracts with the Nigeria Central Bank to supervise credit/financing of small holders across Nigeria. The company already provides data on about 50,000 farmers, and 6G is ready to share these data on the platform upon receiving permission from CropIT. GT Bank is financing ~65,000 farmers annually across the country. The financing goes to the implementation agency, Afex, which distributes the funds to farmers through input purchases.

6G approached CropIT and GT Bank as an experiment to evaluate the DSP solution, and both expressed real commercial interest in the platform. 6G is currently in negotiations with both companies to provide access to the private version of the platform that has a username and password protected area where companies can store and visualize data without sharing with others. The primary driver of interest was the functionality of visualizing all farmers which they plan to finance on a map, and linking each farmer/field to expected agricultural productivity. **Figure 10** below shows how GT Bank in Nigeria realized that the location data they are provided by farmers has significant issues. 6G is in the process of helping GT Bank guide their agents on the ground to use FieldFocus Light to gather field boundaries, verify the location data on each farmer, validate that the farmer has a real field to cultivate, and monitor crop vegetation after planting. We are using the DSP to visualize the result in this effort.

Figure 10:
GT Bank field locations, as provided

The key outcome of this engagement in Nigeria is that local companies see value in using DSP and are ready to pay for the use of these tools. The fact that DSP combines functionality for visualizing crop mapping/classification, visualizing individual fields, and linking these locations to simple indices that provide evidence of vegetation on each field is the main selling point for these two companies in Nigeria. There are thousands of similar companies in Africa (big and small) that provide financing to farmers. Most of these companies are running at a loss because of high default rates for small farmers. Launching DSP across several African countries and providing these three modules (visualizing of crops, visualizing of fields, and linking these data to crop health indices) to all agriculture financing companies would be a major step forward in making the business of credit and financing of agriculture sustainable in Africa. 6G is working with these companies to understand how to use the tools, how they need to change their processes to gather accurate field boundary data and clean the field data that they already have, and to integrate the field crop health indices into decision making.

We are also pleased to report that several government agencies on the continent that have been exposed to the platform would like to explore its use for their own country-specific needs. For example, the Zambia Ministry of Agriculture and the Mauritania Ministry of Agriculture have expressed interest in potential adaptation of the platform. In addition, Syngenta would like to integrate DSP into their commercial launch of an Agronomist application in Kenya at the end of 2022. The agronomist-farmer connection will be implemented by Syngenta to provide timely agronomy recommendations to farmers. The non-confidential aspect of that agronomist-farmer relationship (the field boundary, crop type, and planting date) will be uploaded and available through the Data Sharing Platform.

A very similar arrangement is under discussion with the African Fertilizer and Agribusiness Partnership (AFAP). While the experiment between 6G and AFAP covered agrodealers and farmers in Ghana only, the same model could be implemented across all African countries where the AFAP agrodealer network is strong. The next set of countries where AFAP would like to launch the DSP include Kenya, Zambia, Zimbabwe, Tanzania, and Nigeria.

6G recommends selecting 2-3 countries where DSP has already generated interest from various agricultural players and formalizing these interests through a set of agreements for the official launch. In our opinion, the highest probability of success could be achieved in Nigeria and Kenya, given the initial level of interest and successful use cases we report above. With support from the Enabling Crop Analytics at Scale team, we should be able to identify several other country candidates to experiment with similar commercial arrangements.

A parallel effort should be made to expose the DSP application to the agriculture research community in Africa and around the world. The data already available on the platform, along with any additional data that potential partners will upload, should be used freely by researchers as a source for Machine Learning/Deep Learning (ML/DL) tasks. 6G could commit to providing the initial set of ML/DL algorithms within the platform and would need 2-3 months to get these algorithms fully implemented and operational within the platform.

Use of DSP and Data Collection Tools

Effective agent training requires feedback and real insights from leaders on the ground who have deep knowledge about the mindsets of local agents and farmers. To get these insights, we conducted interviews and held meetings with lead agents, including from AFAP Ghana and Koolimar Input Capital. These discussions provided 6G with an on-the-ground understanding on how local agents and farmers think and what sort of challenges they face. These insights will help us refine and create training and support for agents using FieldFocus Light and other 6G tools when they are working with farmers, and to collect clean boundary and field data.

Key Interview Learnings About Growers for Agent Training Lessons:

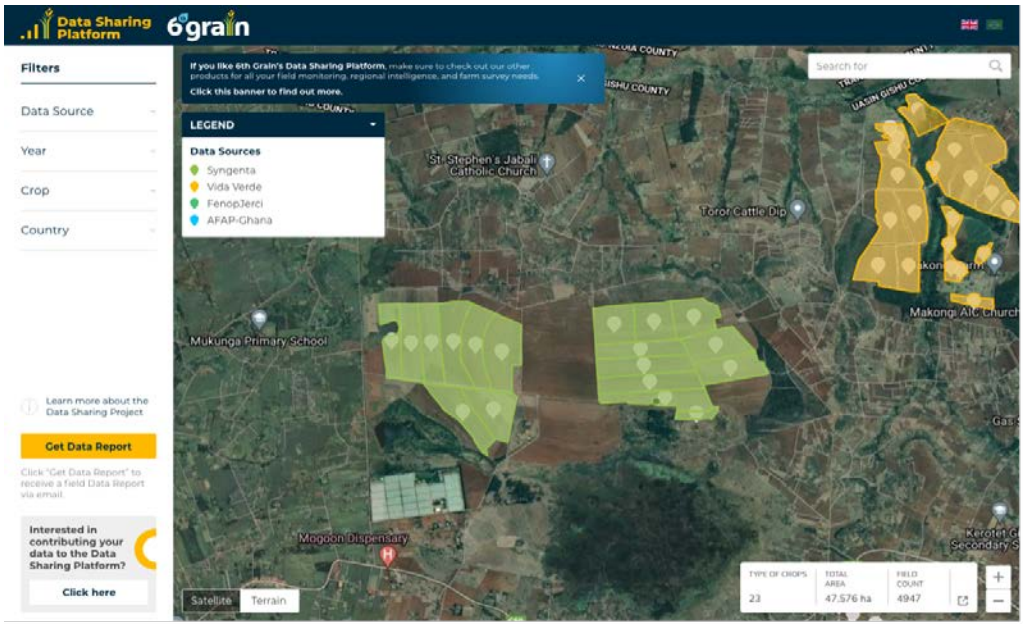
1. Smallholder growers have limited knowledge and experience in using productivity-improving digital tools and technologies. Agents emphasize the importance of clarifying and communicating FieldFocus Light benefits to growers in a way that explains the benefits and incentives.
2. Data collection needs to be done on the field with the farmer, given how complex and varied farmer activities are.
3. Unifying multiple functionalities across various 6G tools, including the survey-based tool Data Terminal and remote sensing information in FieldFocus and FieldFocus Light would make for a better user experience for agents and farmers. More investment in meeting the needs of agents is needed.
4. Agents have noted that while the most growers do not have smartphones, those growers that do have smartphones and access to data services can help drive significant adoption of FieldFocus Light through word-of-mouth.
5. Deployment of FieldFocus Light by agents for farmers represents a "paradigm shift" for farmers and agents working in the financial and agronomic sectors. Training agents to use these tools could have significant results in improved farmer performance and increased loan repayments.

DSP Features and Functionality

The Data Sharing Platform's (DSP) objective is to provide immediate visualization of all data available, provide download and data cleaning options, and enable a summary of all data available to download. The DSP prototype provides users with an online interactive visualization tool for the Platform's field data, a customizable report file with user-selected field data, and a data contribution function. The user-friendly interface places a strong emphasis on a streamlined experience for those who wish to explore the available data, receive data reports, and contribute to the Platform.

A video showing the features of this platform can be found here:
https://watch.screencastify.com/v/8f2KF81cpVyAOewXf7Ri

Interactive Map Webpage



Field Boundary Data

Interactive Map

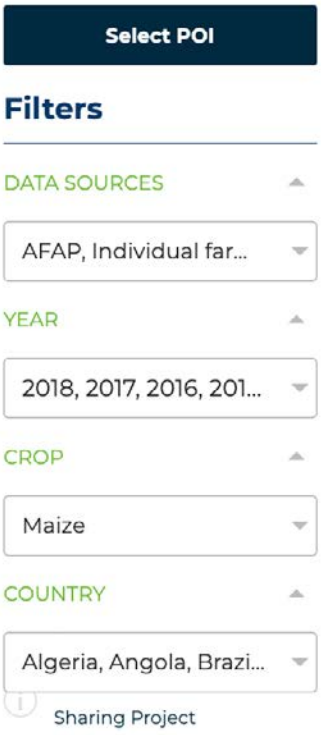
After navigating to the Platform URL (dsp.6grain.com), visitors are presented with the primary online visualization component of the Platform. This map displays all field data available on the Platform to users by way of circles when the map is zoomed out, and pins marking the location of individual fields when the map is zoomed in. The circles display the total number of fields digitized in the circle's location and are colored according to Data Source.

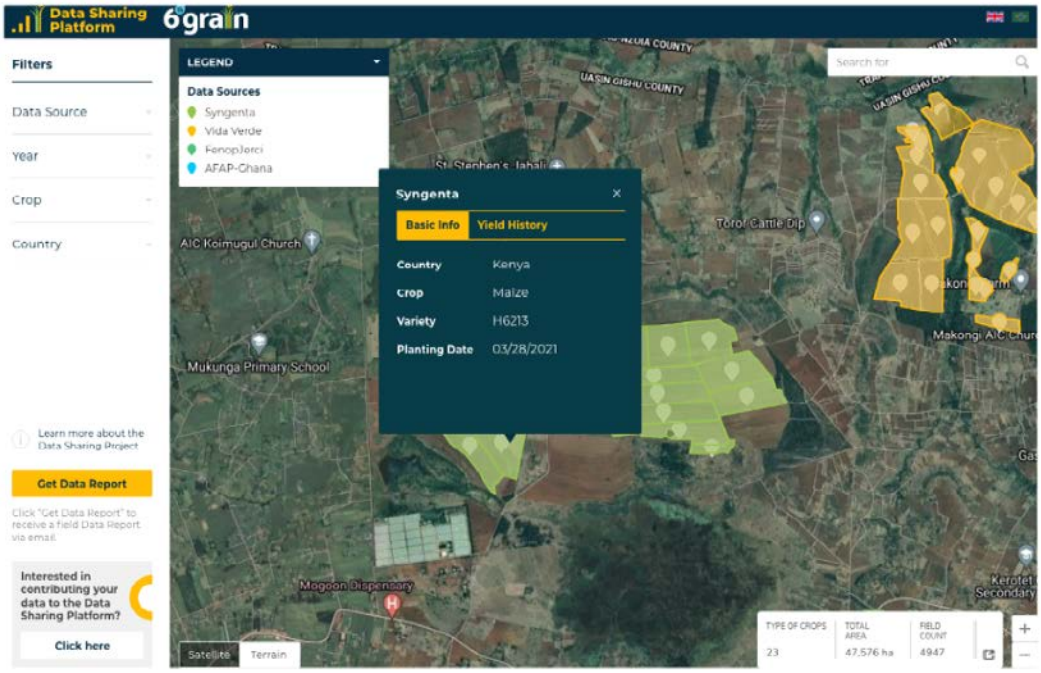
Once users zoom in the map, fields with boundary coordinates have the boundaries displayed on the map and colorized by Data Source. Fields with no boundary data are displayed as colorized pins on the map, based on field geocoordinates.

The Platform's map interface provides users with robust Filters to refine the field data results visualized on the map. Located on the left of the Platform map interface (**Figure 1**), Data Source, Year, Crop, and Country Filters allow users to select the desired field attributes users want displayed. Once selected, the map will display the fields that match the selected Filter field criteria.

The Data Source Filter lists all available companies, organizations, and sources that provided the data to 6G. The Year Filter is based on the field planting date year for a specific crop. The Crop Filter enables users to view fields according to available crop type. Lastly, the Country Filter allows users to sort the fields displayed on the map view by country where the field is located. These Filter categories provide users from a wide variety of backgrounds and objectives to quickly display fields of interest.

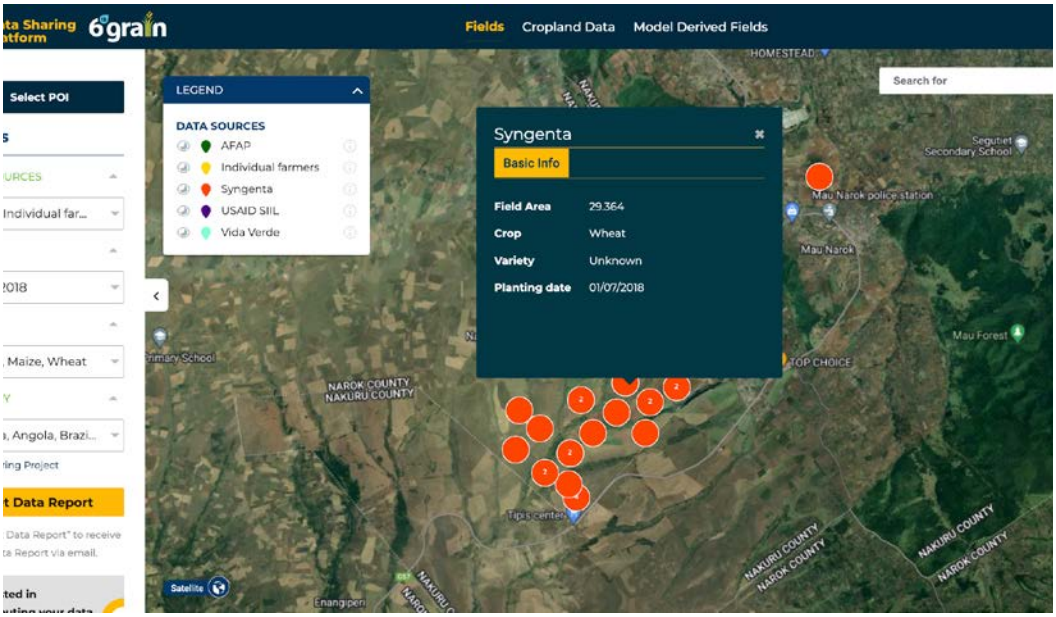
The Platform dashboard, located at the bottom right of the map interface, provides users with key aggregate field statistics, depending on any selected Filters. When all Filter attributes are elected, the dashboard shows the grand total number of Crop types, Field Area, and Field Count. As users change the Filters, the dashboard will adjust accordingly to match the selected fields on the map. The Platform Field Passport ("Passport") presents key field data in an easy-to-read, concise format. To simplify the organization of the field data, the Passport is divided into Basic Info and Yield History tabs. Users can switch between Basic Info's most recent growing season field data and Yield History's historical growing season field data. The key field attributes shown in both the Basic Info and Yield History tabs are as follows: Country, Crop, Variety, Planting Date, Data Source. Yield History displays field attribute data for all available historical growing seasons plus the total production for each historical growing season.





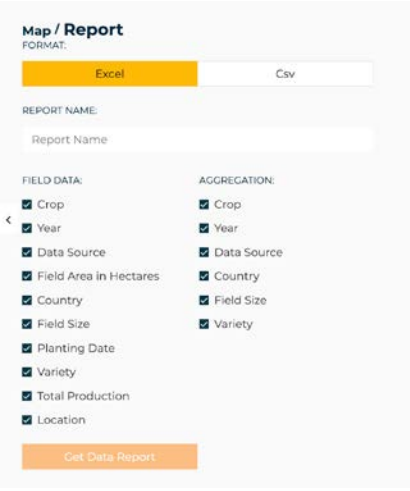
Field Passport Pop-up Displaying Basic Info on the Interactive Map.

The Data Sharing Platform's Legend provides information on the source of our boundary data, along with information on each. The Data Source is represented by a different color in the Legend and on the map. Pins and field boundaries share the same Data Source colors, enabling users to easily identify fields from specific Data Sources on the map without needing to access Field Passports. We also have a feature that allows for overlay of points of interest (POI) which allow the user to view where agrodealers are located and compare those to the fields that have been digitized.

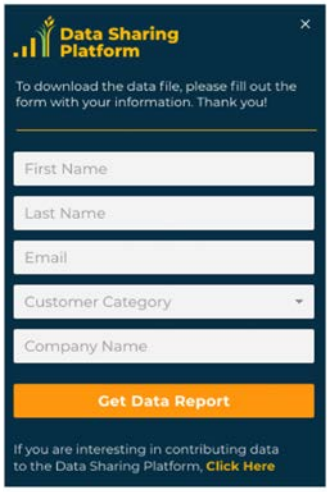


Data Report

Central to the Data Sharing Platform's features is the field Data Report. The Report is a customizable Excel or CSV file of desired field data that is emailed to users. At any point during a user's visit to the Data Sharing Platform, when a user clicks on the "Get Data Report" on the interactive map page, they are presented with a set of field attribute filters to customize the Report. Only filters that are selected by users are included in the emailed Report.



Data Report Filter Page



Data Download Form

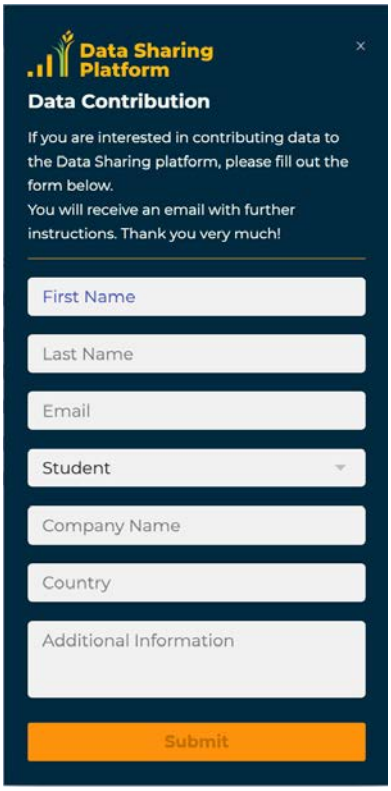
Once users select their filters and click "Next" at the bottom of the Filters page, they are presented with the User Information Form. Users fill out this form's mandatory fields: First Name, Last Name, Email Address, Company Name, and Customer Category. Users click "Get Data Report" and the customized Report file is emailed automatically to the user-provided email address. An email is automatically sent to 6G informing the team of the new user and associated information. Users can use the Data Report feature in an unlimited fashion.

6G utilizes existing partnerships and networks to invite organizations like Syngenta and BASF to contribute their field data. To expand the depth and breadth of the Platform's data beyond organizations brought on by existing partnerships, we have integrated a Data Contribution function in the Platform.

The goal is to have companies that utilize the Data Sharing Platform and want to raise company visibility to easily start the dialogue to contribute to the Platform. The Platform serves as a hub for companies to indirectly advertise by being present on the Platform, due to being visible to all Platform users. The more companies join the Platform, the more field data is available for Platform users, and the greater the incentive for companies to contribute to the Platform to gain visibility.

Due to data privacy concerns, our Data Contribution system features several steps, including a Data Sharing Agreement. 6G takes data privacy very seriously and is why users cannot upload data directly to the Platform via the Platform's user interface.

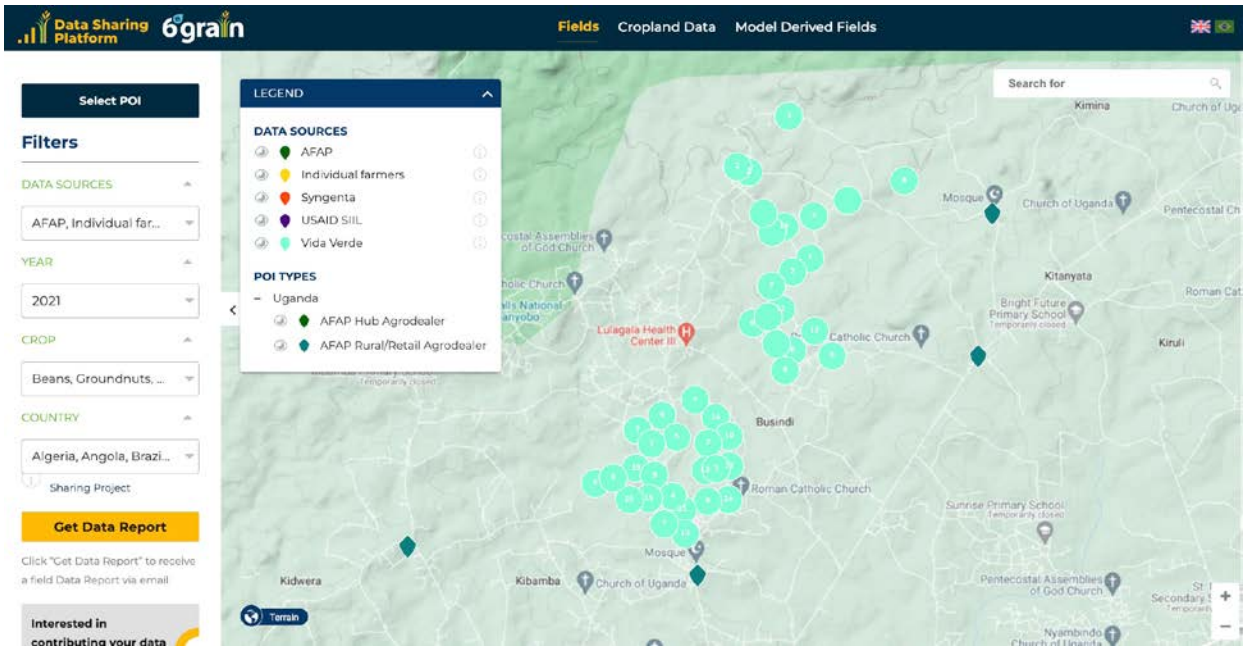
The Data Contribution process on the Platform is easy for users to initiate. Users who are interested in contributing field data to the Platform can do so by selecting the Data Contribution button found on the Interactive Map page. Users fill out the Data Contribution form, and the submitted information is sent via email to 6G. 6G then initiates a discussion by email with the interested contributor and shares a Data Sharing Agreement to the user (**see Appendix**). Once the user emails 6G the user-signed Data Sharing Agreement, 6G will email the user an Excel template that provides a structure for the data they wish to contribute to the platform. 6G accepts multiple file types from users, including shp and kml. 6G then takes these user files and bulk uploads the fields and field data into DSP. After review of the data to ensure no personal identifiable information, we will upload and colorize the data locations according to the donor institution.



Data Contribution Form

Points of Interest Overlay

To enhance the analytic capabilities of DSP, we have implemented a 'Points of Interest' (POI) overlay. For example, partners in Ghana have digitized fields (green circles), which are nearby rural retail agrodealers (diamonds). By identifying regions which are poorly serviced by retailers, or retailers with poor visibility into agricultural fields, improved access to agricultural inputs can be achieved.

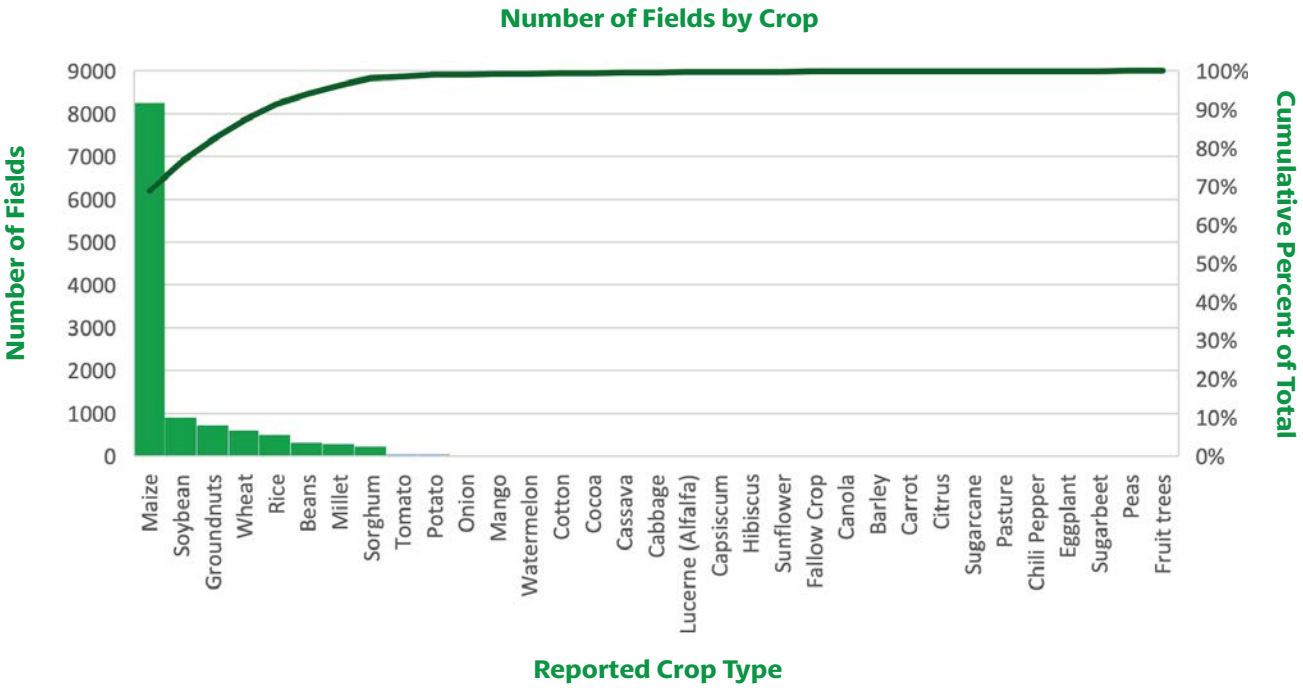


Data Available on DSP

DSP began displaying data in December 2021. Visitors who go to the URL can freely view the field data, download data reports, and initiate the data contribution process. DSP has 12021 fields, of which there are 1189 with production information.

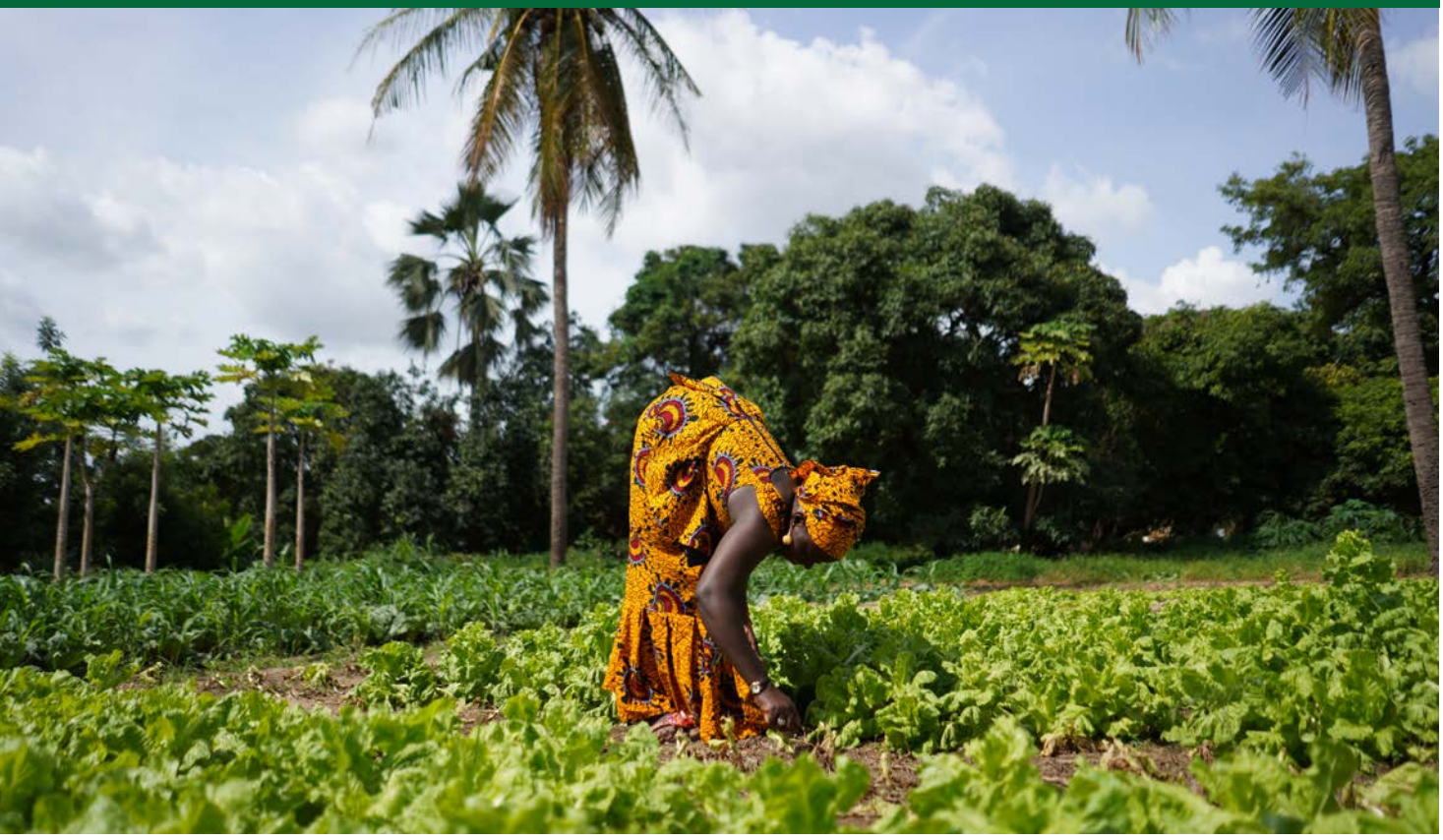
Source	Number of Fields
Individual farmers	101
USAID SIIL	233
Vida Verde	1177
Syngenta	3730
AFAP	6780
TOTAL	**12021**

As a prototype, DSP will undergo minor edits as needed with the goal to improve the user experience over time. Usage metrics will be tracked continuously through Google Analytics, enabling us to better understand how DSP is being utilized. We will also be tracking report download and contribution metrics. These metrics will enable 6G to identify and resolve inefficiencies, as well as better understand the real-world use cases of the platform.



Data Contributors

During the project, 6G instituted policies regarding the data it publishes on the DSP to protect the privacy of individuals and ensure that users of the data have the legal and ethical rights to use it. 6G does not publish datasets that include Personally Identifiable Data (PID), and our system allows only information on field boundary, crop type, crop variety, planting date, and yield. The system does not provide any other fields for data sharing, and therefore protects PID through design. All organizations that have shared data have signed a Data Sharing Agreement which is on file at 6G. The agreement template is presented in the **Appendix**.



Individual Farmers

Farmers who downloaded and used the FieldFocus Light application to digitize their own fields accepted the data sharing agreement upon installation, and their data is located on the platform. These fields are in Kenya, where we implemented our Facebook campaign during the fall of 2021.

Syngenta

Syngenta Corporation is one of the leading companies offering high quality seeds, agriculture inputs and agriculture partnerships in East and Southern Africa. 6G works with Syngenta country offices in Zambia, Zimbabwe, Kenya, and Tanzania and provided data from demonstration plots, replication plots and other locations where crop, planting date, and production information.

The Syngenta Foundation for Sustainable Agriculture (SFSA) focuses on productivity and the inclusion of farmers in remunerative value chains. The focus is on 'pre-commercial' farmers, often in semi-arid areas, who display potential for agricultural growth. SFSA has worked with 6G on various projects, including designing a Fall Armyworm decision support product and implementing a farmer survey on the need for information and weather insurance.

African Fertilizer and Agribusiness Partnership (AFAP)

AFAP provides business and technical services to small and medium agribusiness across East, Southern and West Africa. AFAP takes a “hub and spoke” approach by developing wholesale and retail networks to assist inputs suppliers and off-takers in reaching the “last mile.” These linkages help local agribusiness and smallholder farmers gain greater access to global fertilizer, seed, agrochemical and food companies, technologies, financing, and equipment providers in Africa and provide trade finance for fixed assets and inventory via the Agribusiness Partnership Contract (APC) mechanism.

We provided the FieldFocus Light tool to retail locations and will ask networks of distributors, agents, and retailers participating in the AFAP network to collect boundaries of farmers. This data will be extremely useful to them because they can track where production anomalies (significantly above or below average production) in one season will affect the purchasing power of the subsequent season. We engaged specifically with the office in Ghana and have posted over 6000 field boundaries in Ghana of farmers being served by its agronomists.

Vida Verde

Vida Verde (VVL) works in Uganda to engages with a large range of associate experts across Africa in all areas of specialties in agriculture. VVL has excellent crop and animal production expertise and has market linkages for all agricultural produce locally and internationally. VVL also strives to uplift the farming standards and efficiencies of the

small holder farmers and to improve the margins of less profitable entities, and helps with risk management, loss control and turnaround strategies. They have worked with 6G for the past five years and have provided our data collection tool FieldFocus Light to their agents, who have used it to digitize fields across Uganda. This data has been shared within the DSP.

Farmerline

Farmerline launched in 2013, with operations serving 800 farmers across Ghana – and a massive ambition to change the agricultural sector as a whole. Currently, FarmerLine works with colleague Lyndon Estes at Clark University to collect field boundary data and evaluate the model-developed boundary data that Clark University has created within their system to provide inputs, financing, warehousing, and access to agricultural products across the value chain in Ghana.

References

Brown, M.E., 2021. Metrics to accelerate private sector investment in sustainable development goal 2—zero hunger. Sustainability 13, 5967.

Estes, L.D., Ye, S., Song, L., Luo, B., Eastman, J.R., Meng, Z., Zhang, Q., McRitchie, D., Debats, S.R., Muhando, J., others, 2021. High resolution, annual maps of field boundaries for smallholder-dominated croplands at national scales. Front. Artif. Intell. 4.

Gerpacio, R. V, 2003. The roles of public sector versus private sector in R&D and technology generation: the case of maize in Asia. Agric. Econ. 29, 319–330.

Maguranyanga, C., Murwira, A., 2014. Mapping maize, tobacco, and soybean fields in large-scale commercial farms of Zimbabwe based on multitemporal NDVI images in MAXENT. Can. J. Remote Sens. 40, 396–405.

Mason-D'Croz, D., Sulser, T.B., Wiebe, K., Rosegrant, M.W., Lowder, S.K., Nin-Pratt, A., Willenbockel, D., Robinson, S., Zhu, T., Cenacchi, N., others, 2019. Agricultural investments and hunger in Africa modeling potential contributions to SDG2--Zero Hunger. World Dev. 116, 38–53.

Rashed, A.H., Shah, A., 2020. The role of private sector in the implementation of sustainable development goals. Environ. Dev. Sustain. 1–18.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS J. Photogramm. Remote Sens. 164, 152–170.

UC Berkeley, 2020. What Is Machine Learning (ML)? University of California Berkeley.

Appendix: Data Sharing Agreement

Contributor Name

Data Sharing Platform Agreement

DATA SHARING AGREEMENT

BETWEEN

[Name of Contributor Here]

address

AND

6th Grain Corporation

7313 Barra Drive, Bethesda MD 20817 USA

Definitions

"6th Grain Data Sharing Platform" is a geospatial database and visualization platform where data can be accessed.

"*Anonymized Information or Data*" means geospatial data on field boundary, crop type, crop year and production, but with no personally identifiable information such as name, phone number, field "*Field*" is an area on the ground with uniform management – cultivated at the same time, planted in the same crop, application of the same crop inputs and harvested at the same time.

"*Participants*" are companies that provide information to the Data Sharing Platform.

"*Personally Identifiable Data (PID)*" for the purposes of this policy shall include Name, Phone number, Mailing or Home address, Date of birth or age, Gender or sex, Race, Income and other financial indicators, and Highly Sensitive PID as defined as National identification numbers, Bank account numbers, Driver's license numbers, Credit card numbers, and Personal health data.

"*Researchers*" are individuals or groups working for academic or public-sector organizations looking for training data to improve publicly available agricultural statistics, or for research purposes.

Purpose of Agreement

The 6th Grain Data Sharing Platform's objective is to provide an easy to use, centralized location where anonymized information on agricultural activities can be shared to benefit both participants and researchers. The long-term goal is to increase yields and income for all farmers in Africa. To achieve this, we need to greatly expand the agronomy advice and input availability information for growers in East and Southern Africa. We also need to increase these growers' ability to determine if seeds purchased and other inputs used have resulted in increased productivity. To do this, we seek to share information at the field-level, which include field boundaries, crop in cultivation, year in cultivation, and productivity in the field, that can be combined with geospatial data to improve information on effectiveness of agricultural technology.

The intent of this Data Sharing Platform is to provide new ground-truth data to participating companies and to researchers seeking to create new information on the agricultural system for the public good. Participants of the Platform will include those organizations who provide data, as well as researchers who will use the data who are from universities or other research organizations.

6th Grain Data Platform maintains the following policies regarding the data it publishes on the Data Sharing platform to protect the privacy of individuals and ensure that users of the data have the legal and ethical rights to use it.

Contributor Name

Data Sharing Platform Agreement

Duration of Agreement

This agreement will begin with the Anonymized Data being placed on the Platform and will remain in place indefinitely, or until terminated by either party in accordance with local and international data protection laws and regulations

Description of Data

The data to be shared include **Anonymized Information on field boundary, crop type, production in that field, and year during which the field was in cultivation or planting date**. A field shall be defined as a region of land that is managed uniformly throughout a growing season, with field preparation, planting, weeding, and harvesting all occurring on the same date. Additional information on use of fertilizer or variety in cultivation may also be shared, but no other information, particularly on ownership or personally identifiable information which can relate any individual or company to the field, shall be shared. In the event that personal data are nevertheless transferred or otherwise processed in any way, 6Grain will inform [Contributor] immediately and will delete the personally identifiable data. In addition, 6th Grain undertakes to comply with all applicable data protection laws applicable to such cases.

Data Security

The site will be password protected, and all users will be requested to regularly update their password. Organizations participating in the Platform will be published on the site and all will be invited to participate in quarterly seminars where the results of the system will be presented. All parties shall provide details of staff and levels of access to be assigned.

Personally Identifiable Data

6th Grain does not publish datasets that include Personally Identifiable Data (PID). Personally Identifiable Data (PID) for the purposes of this policy shall include: Name, Phone number, Mailing or Home address, Date of birth or age, Gender or sex, Race, Income and other financial indicators, and Highly Sensitive PID as defined as National identification numbers, Bank account numbers, Driver's license numbers, Credit card numbers, and Personal health data.

To ensure that no PID is published, 6th Grain follows a strict data review policy to ensure that data being provided has no PID. Data partners are instructed not to share datasets that contain PID. 6th Grain will review datasets sent and will remove any PID before uploading them to the Platform.

Consent and Rights

For data to be published by 6th Grain, the data partner must have all rights to share the data under the license they choose.

Thus, data may only be shared with 6th Grain and the Platform Partners if the data provider owns the data, has permission and rights to share from any partner organization for any component parts, has complied with any applicable local regulations on collecting and sharing data, and has consent to share from participants. By submitting the data to our platform, data providers affirm their compliance with these requirements and take all responsibility for violations of those requirements.

Signatures

[Contributor]	**6th Grain Corporation**
_____	_____
Signature	Signature **Molly E Brown, PhD**
_____	_____
Printed Name	Printed Name
	Chief Science Officer
_____	_____
Title	Title
_____	_____
Date	Date

6th Grain's Data Sharing Platform Enhances the Value of Farm Data across Sectors

info.ecaas@tetratech.com
cropanalytics.net

Final report
July 29, 2022

Prepared by:
Molly E Brown, CSO and Vladimir Eskin, CEO