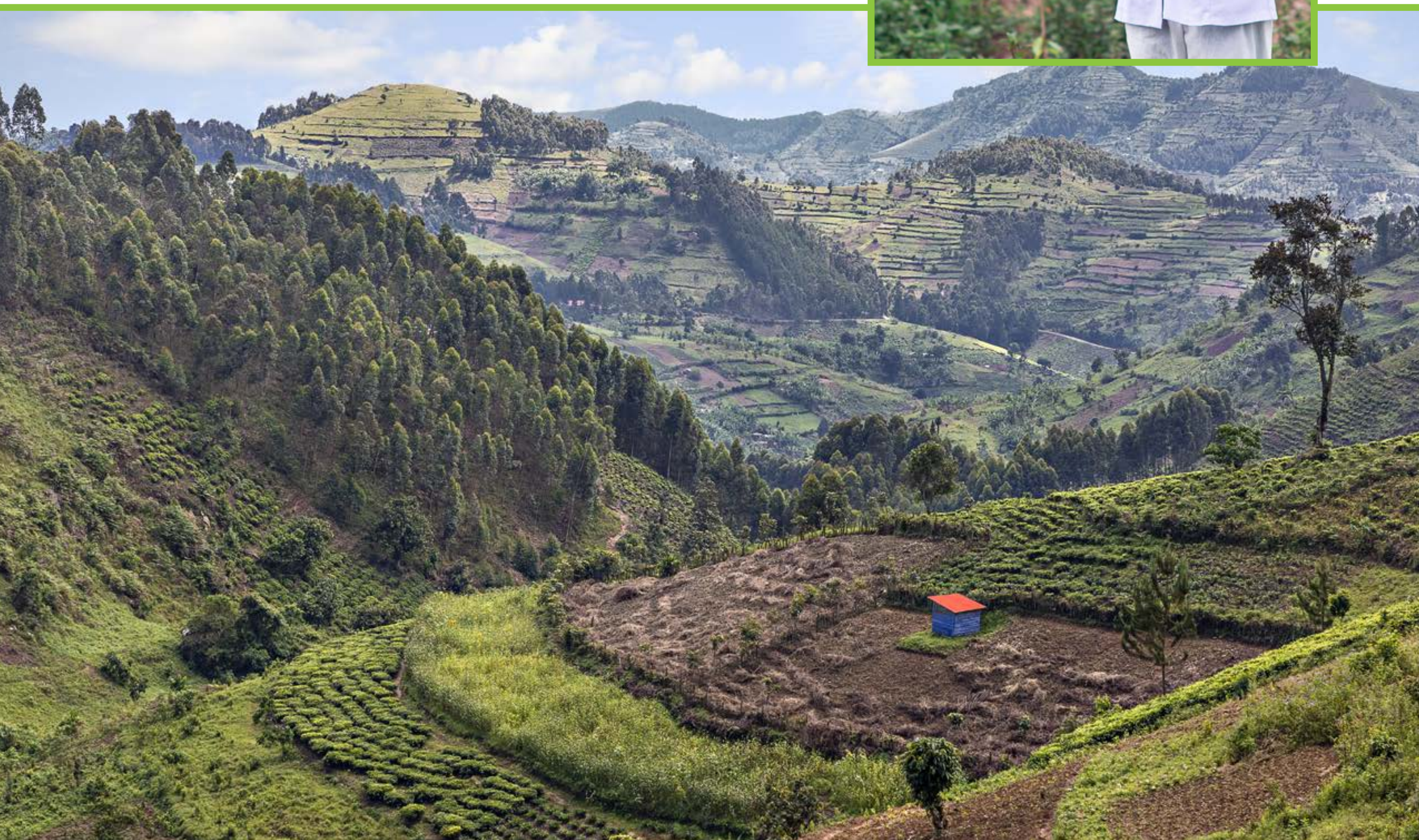
TETRA TECH

Enabling Crop Analytics At Scale

ENABLING CROP ANALYTICS AT SCALE (ECAAS)

Stakeholder Landscape Assessment



Contents

Acronyms

ACCESS	›	Advancing Collaborative Connections for Earth System Science
AGRISurvey	›	FAO Agricultural Integrated Survey
AI	›	Artificial Intelligence
AIMS	›	Agricultural Information Management Standards
API	›	Application Programming Interfaces
ARDC	›	Africa Regional Data Cube
BLDS	›	Building and Land Development Specification
CEMA	›	European Agricultural Machinery Association
CEOS	›	Committee on Earth Observations
CGIAR	›	Consultative Group on International Agricultural Research
COG	›	Cloud Optimized GeoTIFF
COPA-COGECA	›	Committee of Professional Agricultural Organizations-General Confederation of Agricultural Cooperatives
ECAAS	›	Enabling Crop Analytics at Scale
EO	›	Earth Observation
EOSDIS	›	Earth Observing System Data and Information System
ESA	›	European Space Agency
FAO	›	The Food and Agriculture Organization of the United Nations
FEWS NET	›	The Famine Early Warning Systems Network
G.E.M.S.	›	Genetic, Environmental, Management, and Socioeconomic Data
GDBX	›	Geospatial Big Data Platform
GDPR	›	General Data Protection Regulation
GEOGLAM	›	Group on Earth Observations Global Agricultural Monitoring
GeoHIVE	›	Geospatial Human Imagery Verification Effort
GPS	›	Global Positioning System
GTT	›	Ground Truth and Training Data
HR	›	High Resolution
INGO	›	International Non-Governmental Organization
JAXA	›	Japan Aerospace Exploration Agency
JECAM	›	Joint Experiment for Crop Assessment and Monitoring
KARLO	›	Kenya Agricultural Livestock and Research Organization
KIAAR	›	K.J. Somaiya Institute Applied Agricultural Research
LSMS-ISA	›	Living Standards Measurement Study-Integrated Surveys on Agriculture
ML	›	Machine Learning
MLHub	›	Machine Learning Hub
NASA	›	National Aeronautics and Space Administration
NCAR	›	National Center for Atmospheric Research
NGO	›	Non-Governmental Organization
OECD	›	Organization for Economic Co-operation and Development
PPP	›	Public-Private Partnership
SLA	›	Stakeholder Landscape Assessment

SMS	›	Short Message Service
SNA	›	Social Network Analysis
STAC	›	The SpatioTemporal Asset Catalog
TIFF	›	Tagged Image File Format
UAV	›	Unmanned Aerial Vehicles
USAID	›	United States Agency for International Development
USGS	›	United States Geological Survey
VHR	›	Very High Resolution

Executive Summary

The proliferation of mobile, remote sensing, and computer technologies is leading to the rapid digitalization of agricultural data. In turn, data collected on and off the farm is increasingly used to generate analytics and insights that farmers and agribusinesses can use to improve productivity, resilience, and livelihoods. Machine learning (ML) and artificial intelligence (AI) tools are advancing to enable analysis at the field level but across broad geographies. However, information asymmetries and inconsistent coordination of efforts lead to disparate data that lack interoperability and shareability. This hinders the ability of organizations to generate advanced crop analytics like crop maps, crop yield estimates, and crop production predictions which in turn affect the cost, accuracy, and effectiveness of extension services and other benefits for smallholder farmers. Most acutely, a lack of adequate ground truth training (GTT) data on parameters such as field boundaries, crop type, and yield is limiting the pursuit of these opportunities, especially in smallholder agricultural systems.

This Stakeholder Landscape Assessment (SLA) seeks to identify, categorize, and map the ecosystem of public, private, and civil society organizations operating in the crop analytics space for smallholder agricultural systems. The objective of the SLA is to provide key crop analytics stakeholders with a basis for developing partnerships and data-sharing arrangements that accelerate the application of advanced crop analytics to support smallholder farmers at scale. The SLA builds upon desk review, community consultations (50), and survey responses (32) from some of the 192 organizations identified through the implementation of the Enabling Crop Analytics at Scale (ECAAS) initiative thus far.

Key Findings

› Geographic Distribution

Geographically, the stakeholder landscape is characterized by two main clusters of mature organizations in North America and Europe, owing in part to the long-running NASA and ESA Earth observation programs. India, China, and Kenya host the largest emerging clusters of advanced analytics organizations working with smallholder populations (**Figure 1**). Across the landscape, existing clusters are largely siloed from one another, with few lasting relationships connecting each cluster to the others. Organizations are collecting data and deploying tools and applications most significantly in North America, Europe, and Brazil, with large amounts of emerging activity in India, Kenya, Nigeria, and Côte d'Ivoire.

Country Count

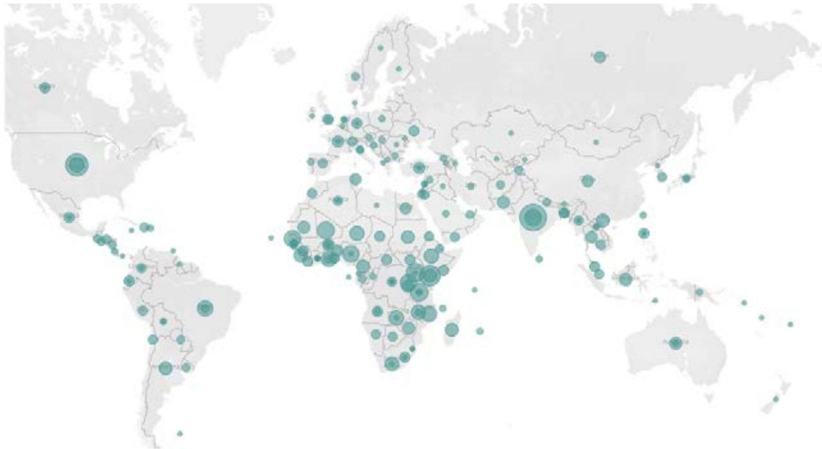


Figure 1:
Circles indicate focal locations for actors within the ECAAS stakeholder database. The size of the circle positively correlates with the quantity of actors per location.

Data Chain Gaps

Most private and public sector actors within the network focus on data capture and analytic services that prioritize two use cases: on-farm extension management and food security monitoring. These findings highlight the need to consider models and incentive structures for data collection and sharing, which comprehensively feed different use cases (**See Figure 9**).

Network Influencers

Across the data chain, the network comprises key influencers that are at most one step removed from any organization in the network and whose actions can have an outsized impact on the rest of the network. While there are many important organizations in this ecosystem, our survey, which received responses from mostly North American actors, found the top influencers included Planet Labs, Amazon Web Services, Google, ESRI, and Airbus. However, no single organizational leader emerged that specifically convenes actors and encourages sharing of data and adopting collective standards for agricultural ground truth data. While consortia (NASA Harvest) and small communities of practice have formed around specific issues, broader coordination mechanisms have failed to develop in the absence of any central convening entity or donor funding. Other organizations such as FAO Hand-in-Hand and ESA are also actively building ecosystems for agricultural data, but did not appear as key influencers in the survey data.

Challenges and Opportunities

The landscape of actors working to generate, disseminate, analyze, and improve the use of remote sensing technologies for smallholder crop analytics includes a diverse range of organizations in the public sector, private sector, and civil society. These organizations are driven by diverse mandates and business models and have widely divergent incentives to standardize data collection and formatting or reasons to share data with other organizations in the ecosystem. As a result, the community remains fragmented and relies upon donor initiatives to drive collaboration. Organizations in the sector identify eight key challenges inhibiting coordination across the landscape:

1. Lack of a coherent community of practice or a sustained coordinating mechanism;
2. Disagreements about how much and what type of data should be collected to develop valuable models while still protecting privacy;
3. Lack of interoperability among datasets and technologies driven by a lack of standardization and diverse data collection requirements;
4. Lack of independent validation and benchmarking for both datasets and analytic models;
5. Differing data privacy and ownership requirements across jurisdictions and cultures;
6. Differing perceptions of value for the ground-truth data itself, especially if collected for a narrow use case or purpose;
7. Inconsistent incentive structures to coordinate and collaborate, especially between public, private, and civil society organizations; and
8. Allowing stakeholders and end-users to drive technology and datasets development

The ECAAS Innovation Agenda seeks to advance solutions to help address several of these key bottlenecks and technical challenges, yet additional work remains. There are still many opportunities for ECAAS or similar programs to support the community to collectively create a set of solutions (as with formalizing a community of practice) or promote existing solutions in the landscape (as with interoperability and standards). These opportunities and approaches are further detailed throughout this report and summarized in **Section 4**.

Background on ECAAS

The Enabling Crop Analytics as Scale (ECAAS) initiative, funded by the Bill & Melinda Gates Foundation and implemented by Tetra Tech, seeks to unlock the tremendous potential of remote sensing and Earth observation in ways that could transform smallholder agriculture. ECAAS aims to do this by investing in long-term platforms for collecting and sharing ground-truth data that can be used in advanced crop analytics to support smallholder farmers. ECAAS is establishing a network of public and private sector actors who can work together to realize the potential benefits of the agricultural remote sensing and ground data ecosystem applied at scale.

ECAAS operates through three principles to catalyze an improved data sharing ecosystem:

1. Maximize impact for smallholder farmers by increasing productivity, incomes, market linkages, nutritional outcomes, and expanded sources of relevant extension information.
2. Target scalable solutions that drive availability and uptake of new technologies and other improvements at scale across geographies and are grounded in the financial sustainability of approaches.
3. Help bridge gaps in the stakeholder landscape and capitalize on opportunities in existing digital systems and networks, rather than creating duplicative networks or channels.

Introduction

Advances in data analytics continue to accrue at a rapid pace, allowing organizations to address previously unsolvable problems across multiple industries [1]. In agriculture, farmers and food providers must feed a rapidly growing population while protecting and improving their livelihoods. This is especially true in smallholder farming systems, which typically operate on less than 2 ha of land and account for more than 80% of all farms worldwide [2]. These individuals produce up to 70% of food supplies in certain regions but face significant gaps against potential yields in most crops [3][4]. Smallholder farmers are often most impacted by the adverse impacts of a rapidly changing climate and require more precise and real-time data to understand spatially distributed weather variability and the impact on production decision-making. Through the applications of advanced agricultural analytics, smallholder farmers could benefit from significant gains in productivity, efficiency, and environmental protection. However, accumulating the size, relevancy, interoperability, and diversity of datasets needed to reap such benefits for smallholder farmers is currently a challenge.

One of the most significant barriers preventing the widespread use of crop analytics is the lack of ground truth data required to train machine learning (ML) models which leverage remote sensing technologies. Ground truth training data (GTT) refers to the real features and conditions on the ground at a given point and time. The collection of properly formatted GTT enables calibration of remote-sensing data, and aids in the interpretation and analysis of what is being remotely detected [5]. Organizations collect GTT in various ways, driven by the perceived value of this data to their derived products or analytics, the cost required to collect and store the data, funder or client requirements, and other reasons. Accordingly, the type and quality of this data vary significantly across the landscape. As an example, **Figure 2** on the next page depicts six different survey methods employed to collect field boundary GTT data.

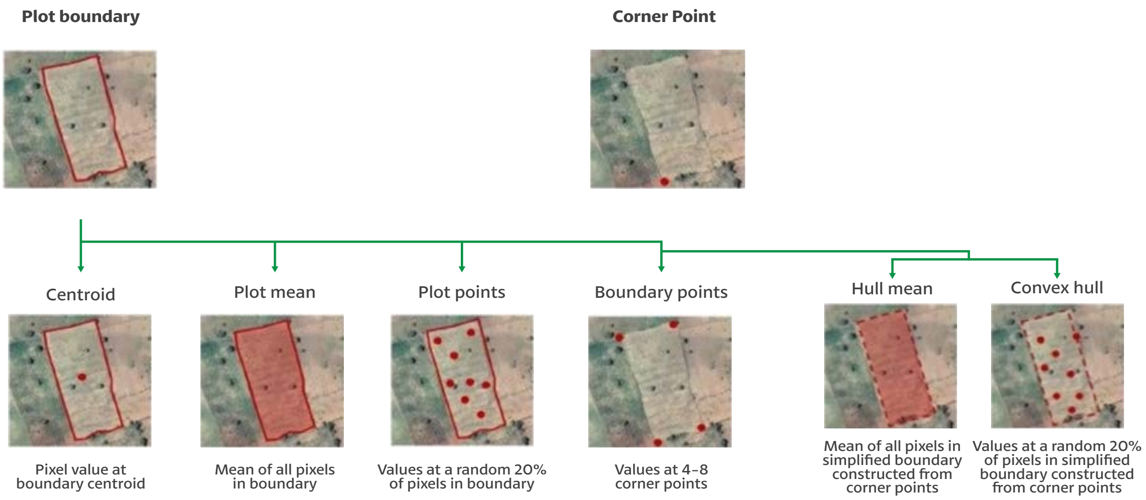


Figure 2:
Examples of Field Boundary data points [6]

Accurate and timely ground data are essential to improve smallholder farming systems typified by small field size, irregular field boundaries, and in-field heterogeneity. Yet, the cost and complexity of collection must not outweigh the value of the derived products and analytics [7]. This is especially true as analytics providers move past field boundary detection into other, more difficult core parameters such as crop type identification, crop health monitoring, and yield predication.

Beyond being time-intensive and costly to collect, only a fraction of smallholder datasets are available in the public domain because many are siloed within institutional and individual databases, are proprietary intellectual property, or are subject to legal barriers which prevent information sharing. Industry standards for defining quality GTT are also lacking as community guidelines on how to collect, format, and exchange such data are not widely accepted. Publicly available agricultural data like census surveys are generally poor sources of data to train ML models and are often not timely or responsive enough to solve emerging agricultural problems throughout the growing season. Likewise, most publicly funded satellites produce large amounts of remotely sensed open-access data but have resolutions that are generally too low to capture the intricacies of small, intercropped fields for smallholder farmers[a].

[a] Planet, a private company, provides high resolution and high frequency earth observation imagery. While Planet provides many commercial products through paid licenses, it also provides openly accessible data for the pan-tropics free of charge through the Norway International Climate and Forest Initiative (NICFI), assuming the data is used for non-commercial purposes.

Ultimately, developing cost-effective models for GTT data capture and finding strategies for disseminating information and training data in an accessible manner will require a network approach. A network based approach means looking at the relationships between organizations as well as the characteristics of each organization. To begin, we must develop a basic understanding of the crop analytics landscape through the use of this Stakeholder Landscape Assessment (SLA).

Objectives

The Stakeholder Landscape Assessment's (SLA) objective is to identify, categorize, and map the ecosystem of public, private, and civil society organizations operating in the crop analytics space and are at least partially focused on smallholder agricultural systems. This document will help the community to prioritize areas for potential investment and partnership to ultimately formalize a network that can accelerate the application of advanced crop analytics in the long term. Other networks have successfully developed coalitions that focus specifically on collaboration in their respective data chains. Some of these networks include but are not limited to the WorldWide Antimalarial Resistance Network, Parkinson's Progression Markers Initiative, Alzheimer's Disease Data Initiative, Building and Land Development Specification, and the OASIS project. These networks have developed and promoted collaborative efforts to create shared standards for data in their respective sectors, even among competitors. We anticipate that a similar collaborative community of practice could catalyze comparable achievements in the crop analytics sector.

This SLA has five sections in addition to this introduction. Each section focuses on the key characteristics of stakeholders working in the space. The first section identifies stakeholders and the focal use cases for advanced analytics to improve smallholder farmer productivity, incomes, and food security. The following section highlights stakeholders and their services based on their role in the data chain and the network, from data capture to processing, storage, sharing, and finally looping back to new data generation for end-users. It further highlights the data regulations and policy environment governing their work. The next section discusses the complex relationships and dynamics of the stakeholder within the existing crop analytic network. The fourth section focuses on the geographic distribution of the landscape and any gaps in our understanding of the network. The final section concludes with critical challenges and opportunities for the future network to consider. For additional information on the organizations referenced in this document, please see **Annex 2**.

Methods

To better understand the crop analytics landscape, we utilized a mixed-methods approach that yielded qualitative and quantitative data. Our process consisted of three distinct but related phases: 1) desk review, 2) key informant interviews and consultations, and 3) online survey. First, our team conducted an extensive review of the literature while examining publicly available information on organizational websites to formulate our general understanding of the global crop analytics landscape. Next, this information was compiled into an initial stakeholder database of 191 organizations, grouped according to organizational type and role in the crop analytics data chain (**Figure 3**).

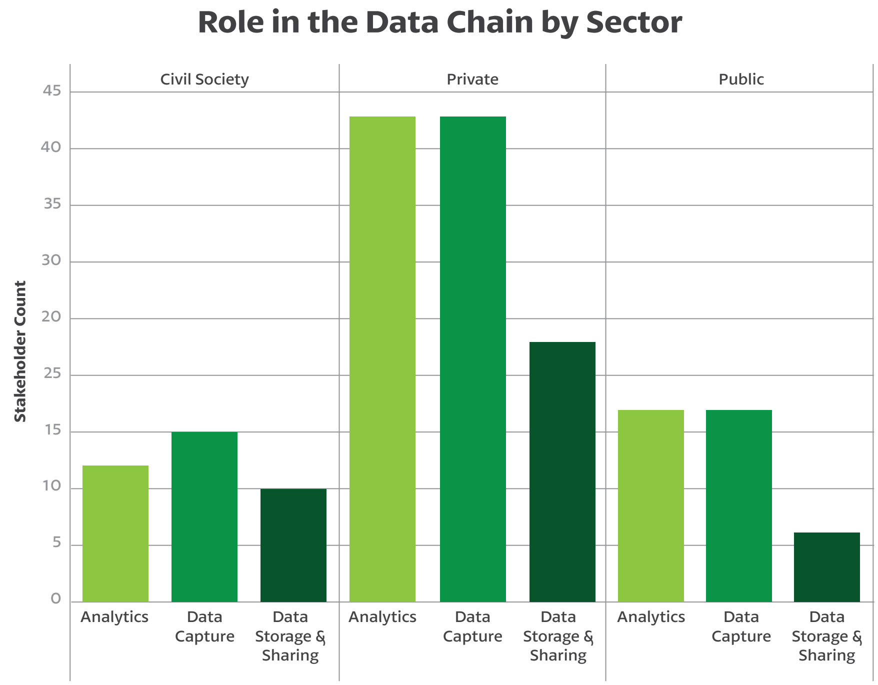


Figure 3:
Role in the Crop Analytics Data Chain by Sector.
Note: Many organizations play multiple roles in the data chain.

The desk review process generated specific insights incorporated throughout this assessment and informed the prioritization of a sub-set of the most prominent organizations for continued dialogue. Over the next six months, we conducted 50 consultations and Key Informant Interviews using semi-structured questions to understand where organizations were positioned within the data chain, emerging challenges, and opportunities they faced within their current and anticipated roles, opportunities for innovative or non-traditional collaborations, and other qualitative concepts.

To help quantify our findings and especially the relational aspects of the landscape, we followed up interviews with an online survey that helped identify specific relationships and network nodes in the space. Organizations self-reported the strength of their relationships with the organizations previously identified and wrote in organizations that had not been previously identified. This information was used as the basis of a baseline Social Network Analysis. Thirty-eight organizations responded to the survey, representing all stages of the data chain. Additional information about the survey questionnaire and methods can be found in **Annex 1**.

The findings from the survey, desk review, and key informant interviews serve as the basis of this document. However, with full transparency, we note that the organizations that completed the survey are not a complete representative sample of actors working on crop analytics at a global scale. Most notably, the responses are highly concentrated among organizations based in North America. With this said, our network analysis did not reveal any distinct sub-networks. Survey responses helped triangulate overall findings and direction, but the limits of the survey sample should be kept in mind. The ECAAS project team will continue to expand both the network of known actors and the number of survey responders to develop a more complete picture of the evolving crop analytics network.

1

Use cases for AI/ML to support smallholder farmers

Improving availability and access to actionable data and crop analytics can empower smallholder farmers prone to the risks resulting from a changing climate. Timely advisory and information can help drastically improve productivity and facilitate more efficient production, harvesting, processing, and marketing of crops, especially for rainfed agricultural systems, which many smallholders depend on[8]. Likewise, the ability to more urgently identify when regional or national agricultural productivity is low can help policymakers effectively allocate resources and mitigate potential food insecurity[9]. Below we categorize potential use cases for crop analytics and highlight where many analytics providers and their partners are focused within these categories.

1.1 Uses for Crop Analytic Products

Advanced crop analytic products support a broad array of end-use case applications which span the entire agricultural value chain from farm to fork. Most established analytical products target medium and large-scale farming systems with significant in-field homogeneity. Data commercialization occurs through service provision to large-scale farmers, input suppliers, commodity traders, governments, or others. However, despite the inherent difficulties involved, including commercialization, the number of advanced analytic services focused on smallholder farming systems is rapidly increasing.

Today, a smallholder maize farmer in Kenya can access improved seed bundled with insurance (Pula) to invest in other farm management practices more confidently as informed by tailored advisory mobile applications (6th Grain, CropIn). Smallholders also use aggregation platforms (DeHaat, eProd) to link with buyers based on production estimates before harvest. By combining these transaction data with others and layering in remote monitoring systems (FEWS NET, NASA Harvest), the Kenyan government can, in a bad year, be warned of a disruption in the regional food system and act before food scarcity reaches critical levels. While promising, this example also highlights the difficulty of directly supporting end uses in the crop analytics landscape given the number of actors, analytic products, and derivative products involved.

To analyze and categorize end-use stakeholders in this crowded environment, we first clustered applications based on the proximity to smallholder farmers within the agricultural value chain:

1. On-farm applications (extension and production management, service provision);
2. Applications upstream of the farmer (access to finance, input supply optimization);
3. Applications downstream of the farmer (output marketing and trading);
4. System-level applications (food security and early warning, public sector land management).

Crop analytics underpin a variety of applications and services for actors across the value chain, from digital extension and advisory services to tailored lending products to improved subsidy allocation. The ECAAS initiative is driven explicitly by three use cases selected chosen for their potential impact upon the smallholder farmer, scalability, and potential to fill gaps in existing digital systems and networks. Our priority use cases are **1)** integrated farm extension and management, **2)** improved financial services for smallholders, and **3)** enhanced food security monitoring and response. Below we depict the range of advanced crop analytics use cases, highlighting the ECAAS initiative's focal areas in dark blue along with the GTT data concepts required to support them and other use cases across the value chain (**Figure 4)**.

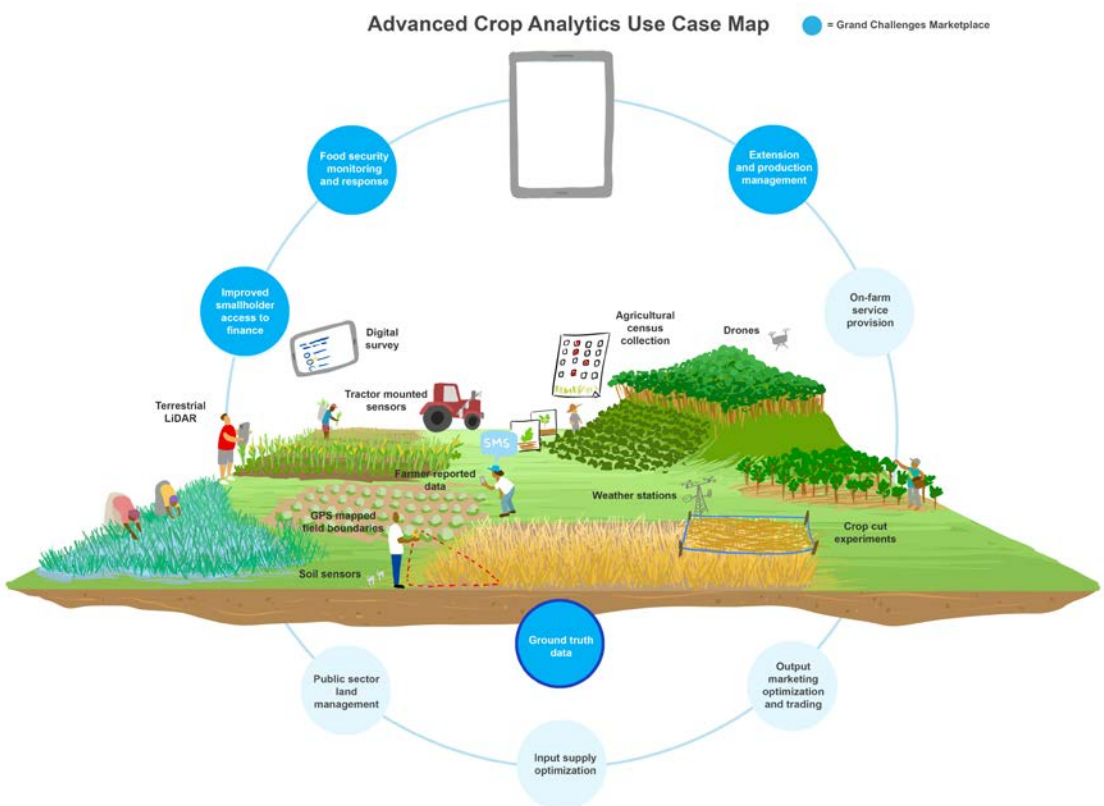


Figure 4:
Use Case Map Highlighting Ground-Truth Data Concepts

Among the total landscape of organizations in our database, the largest number by far focused upon extension production and management applications or service delivery at the farm level (38% of respondents). This further illustrates the growing market facing individual farmers using such products (**Figure 5**). Private actors dominate this use case category, as well as organizations focused on improving access to finance. Conversely, the relatively lower number of public and civil society actors are overwhelmingly represented in food security and monitoring.

These findings highlight the need to think through models and incentive structures for data collection and sharing, which feed different use cases. Data structures that can accommodate the unique balance of organizational needs will be more sustainable and will better support a data sharing network long term. It also highlights the need to balance data asymmetries to provide equitable access and use across all stakeholders, especially smallholder farmers.

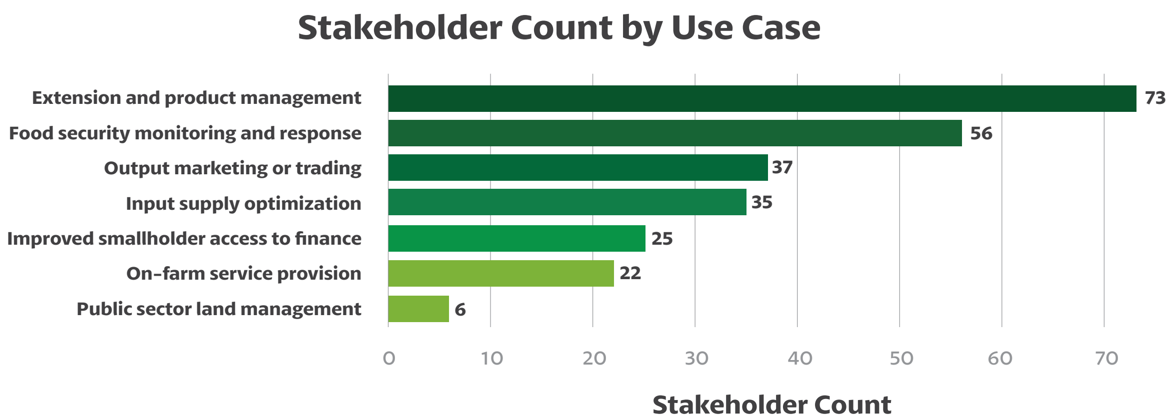


Figure 5:
Stakeholder Focus by Use Case based on the Stakeholder Database

On-farm applications

Integrated farm extension and management applications provide accurate and timely crop advice to farmers and are frequently delivered via mobile phone. In smallholder settings, analytics firms frequently partner with extension networks, input and seed suppliers, or other existing channels for last-mile delivery to leverage existing communication and trust mechanisms. This space contains active private players (6th Grain, CropIn) deploying user-facing mobile dashboards, scheduling services (for optimal spraying, weeding, etc.), or simple SMS reminders driven by ML models. While research shows that user-centric design and actionable advice are required for practical use, many stakeholders continue to "overengineer" applications and fail to add value for the final smallholder farmer user [10] [11]. **Figure 6** provides an illustrative overview of specific target functions within this space.

In the public sector, extension mandates generally fall within Ministries of Agriculture which often face financial and human resource constraints in smallholder environments. Deploying tailored farm-level and near real-time advice solely through these channels is seldom feasible, absent significant external funding. Some public entities develop static and crop-specific applications to optimize production as a middle ground. For example, the Kenya Agricultural Livestock and Research Organization (KALRO) links to 32 mobile applications from its website developed with USAID support. While helpful, smallholder farmers generally trust other local farmers over information technologies for advice [12] [13]. Equally as challenging to overcome is the daunting task of convincing smallholder farmers to trust remote sensing imagery to develop practical farm management advice. The mistrust of remotely sensed information is not unfounded. For example, some farmers have been deprived of payouts for significant losses from index insurance products relying only on satellite data [14] [15].

In addition to pushing out information, on-farm applications can serve as essential data ingestion points to collect and feed updated GTT back into analytics firms and refine models and products, creating dynamic feedback loops. Farmer-centric information loops are required for analytics providers to be able to adjust product offerings based on shifting demand, and can also help capture new data or information required by improving or changing models as they are developed and deployed. Interoperability constraints will be discussed in **Section 2.3**, but apply here as both private and public bodies collect agricultural data for various purposes, some of which do not lend themselves to ML-ready datasets. Some private sector or research organizations are willing to share collected data and compete entirely on model performance and user experience, but the sentiment is not widely shared. Meanwhile, public sector actors in smallholder settings are often underfunded or cannot ensure that data are collected to the quality and standard required by analytics firms[16]. Simultaneously, relying on private-sector data collectors does not solve the problem of geographic gaps in GTT coverage. Commercial organizations need an incentive to enter and operate within a given geographic market or agro-ecological zone that is often lacking in vulnerable smallholder farming communities.

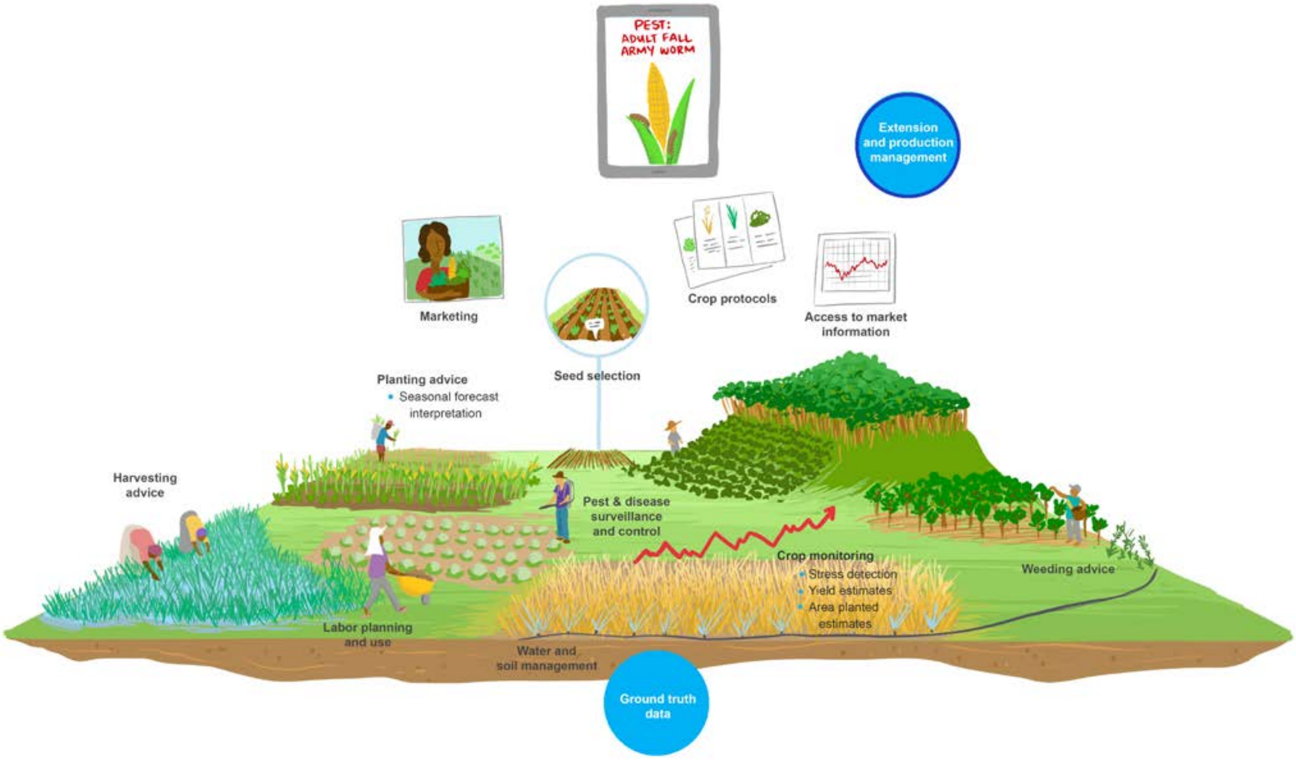


Figure 6:
Use Case Map Highlighting On-farm Extension and Production Management

Applications upstream of the farmer

Many analytics providers working in smallholder settings framed offerings around financial products, improving risk management for financial institutions and insurance providers (**Figure 7**). MercyCorp's AgriFin program, for example, connects analytic providers, small-holder farmers, and financial institutions in their market facilitation approach to digital development in LMICs. In markets such as India, pioneer organizations (Skymet, Satsure) recognized that banks and government-mandated insurance schemes have created a market of well-funded bodies willing to pay for near real-time analytics. By providing timely and geo-referenced information about crop health and growth combined with historical producer and weather data, analytics providers improve risk profiling and increase the accuracy of loss prediction in season. Insurers use analytic services for volatility measurement and develop indexes that trigger payouts, facilitating more rapid and accurate claims processing.

Several major re-insurers (SwissRe, ContinentalRe) have been at the leading edge of deploying these technologies in smallholder environments where existing insurance penetration rates are very low. Other sovereign level index-based insurance schemes such as the Caribbean Catastrophe Risk Insurance Facility (CCRIF), the African Risk Capacity (ARC), and the Pacific Catastrophe Risk Insurance Company (PCRIC), are developing new products that will enable countries to strengthen their disaster risk management systems for agriculture. ML analytics would improve the performance of the insurance models and reduce biases in payouts that benefit smallholders. The World Bank, International Finance Corporation, WFP, and others continue to support and drive efforts to improve governments' capacities to better plan, prepare, and respond to extreme weather events and natural disasters through public-private partnerships.

Another common access point to apply analytics for smallholder benefit is input suppliers, which rely upon advanced analytics for supply chain and footprint optimization, and marketing efforts. Major multi-national organizations with significant exposure in smallholder markets (Syngenta, Yara) leverage both in-house analytics capacities and partnerships with dedicated analytics firms to inform business planning, product distribution, and other core functions. This planning results in better agent coverage among farmer areas and tailored product application advice for plant protection and similar inputs. To date, major competitors in this space have not been willing to share GTT directly with one another, but several are now exploring data-sharing models.

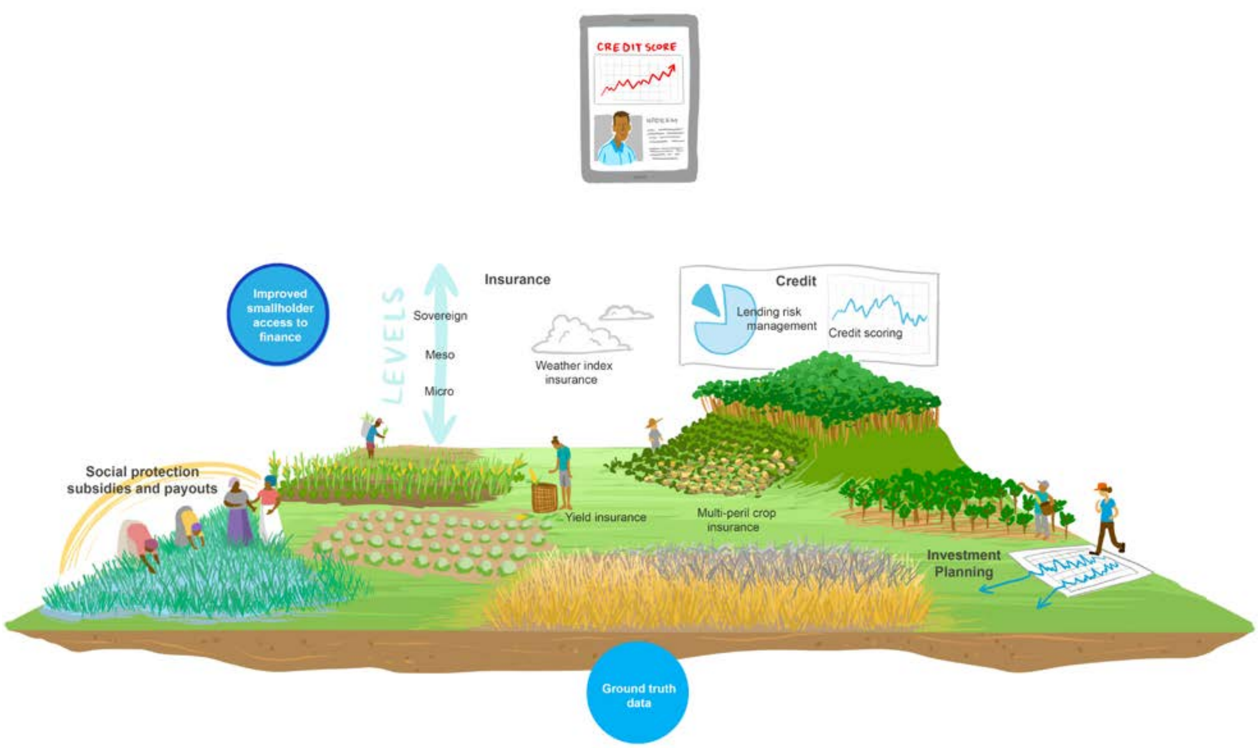


Figure 7:

Use Case Map Highlighting Financial Applications Upstream of the farmer

Applications downstream of the farmer

Applications downstream from the farmer can predict regional yield aggregates, optimize off-take and origination strategies for buyers, and provide local and macro supply forecasts and pricing estimates (**Figure 8)**. Traditionally these applications have been more heavily concentrated in large commercial agriculture, but increasingly players such as CropIn and eProd are working in high-value crops within smallholder systems. Private analytic firms find willing buyers of this information in futures markets and supply chain managers, and these analytics feed into the systemic level applications discussed next.

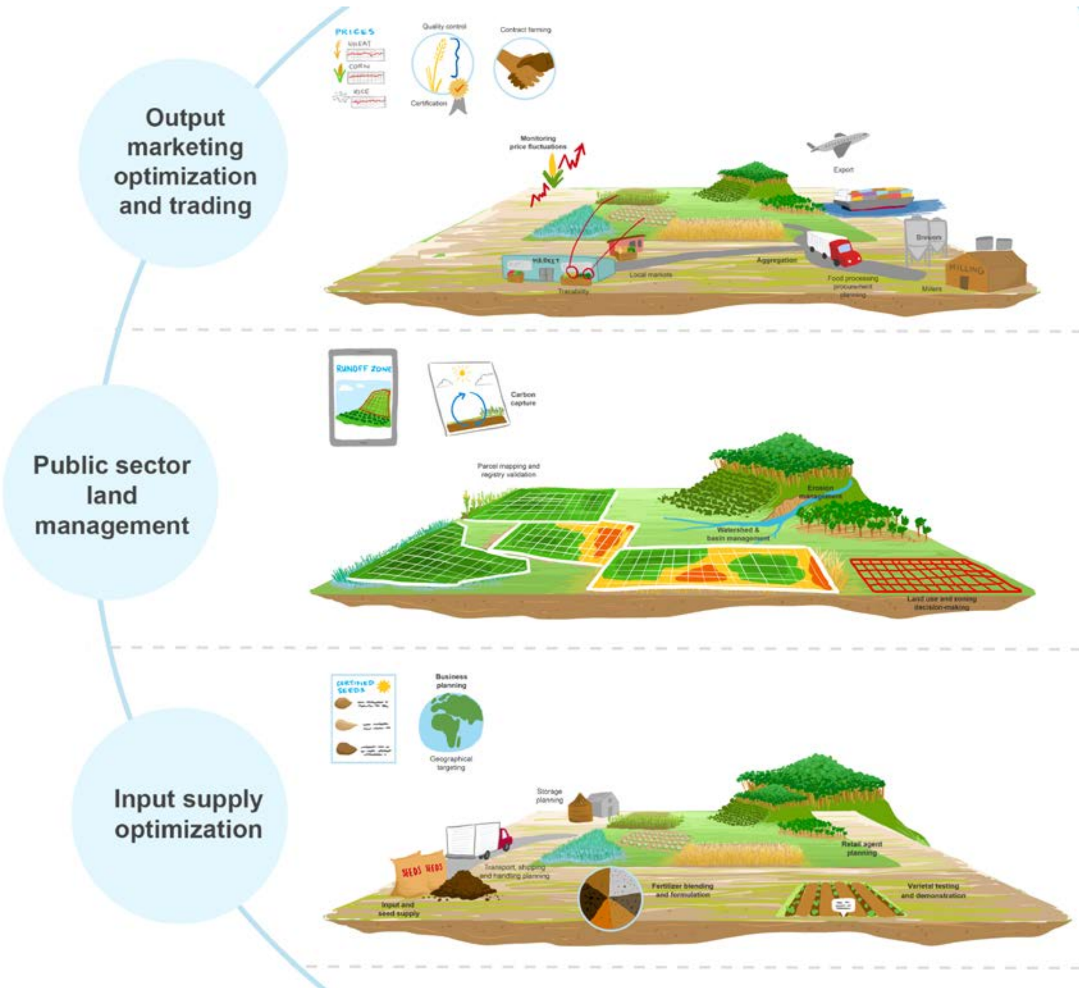


Figure 8:
Use Case Map Highlighting Downstream Application

Systemic level applications

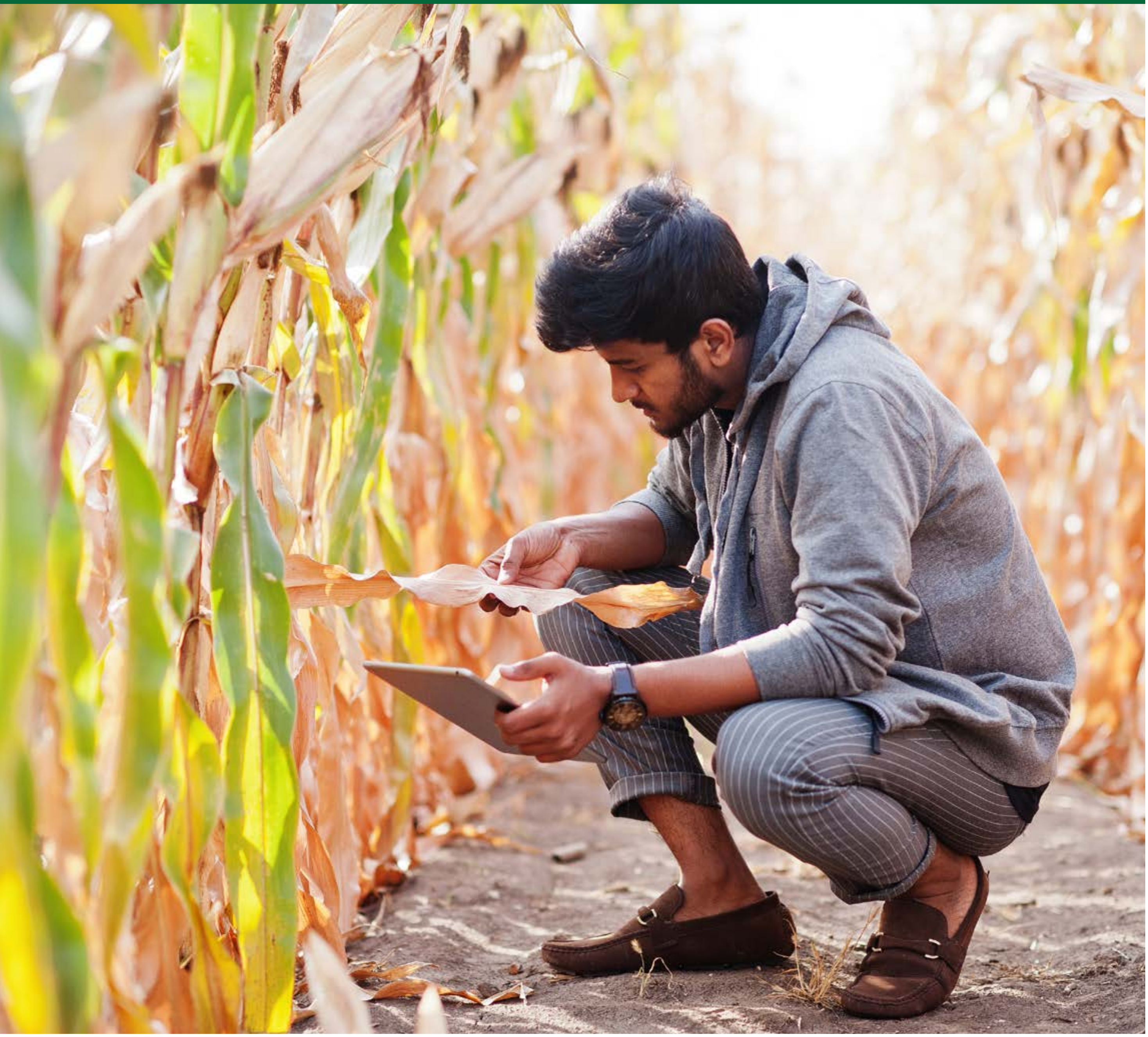
Advanced analytic products can support accurate and timely productivity mapping that international organizations, governments, and NGOs can use to monitor and address food insecurity in rapidly changing environments. Many large donors or publicly funded organizations (FEWS NET, NASA Harvest) assist national partner governments and civil society actors to develop and provide decision-support and related insights and will likely continue to dedicate resources to this area in the future.

Increasingly, donors realize the potential for data-derived AI and ML approaches to improve accuracy, forecasting capacity, and decision-support tools. As a result, they are standardizing their funded data collection campaigns to support interoperability and deepened data lakes within smallholder ecosystems. USAID, for example, has specific policies in place requiring that data collected by awardees be submitted to a central database, but assuring the quality of that data and the accessibility of it for later use has remained a significant challenge in implementation. More emphasis on standards and quality data collection through such mechanisms will improve overall dataset interoperability and value.

Stakeholders in this space work closely with and build the capacity of public sector actors to anticipate, plan for, and respond to disruptions to agriculture and food production. This often includes providing a clear perspective of food production trends in the long term and acute shortages in the short term. The number of organizations working in this space has increased during the evolving COVID 19 pandemic, primarily due to local and global food supply chain disruptions. Analytics can also help demarcate and digitize smallholder plots, map parcels, validate land registries, and resolve land use disputes. Some stakeholders use predictive erosion and runoff monitoring to inform watershed management decisions for national and regional public bodies. These issues are often highly politicized, with complex political economies involved. Advanced analytics, while not immune to political interference, can help provide an additional and neutral data point to inform and guide decision-making and discussion in this space, but the data from which those analytics are derived should be closely scrutinized for any distortion, intentional or otherwise [17].

2

The data chain and stakeholders



The data chain required to develop and deploy crop analytic products flows from data capture to processing, storage, AI/ML model development, applications in end-use cases, and finally loops back to new data generation for end-users (**Figure 9**). This section explores these components of the chain and the organizations working within and between each.

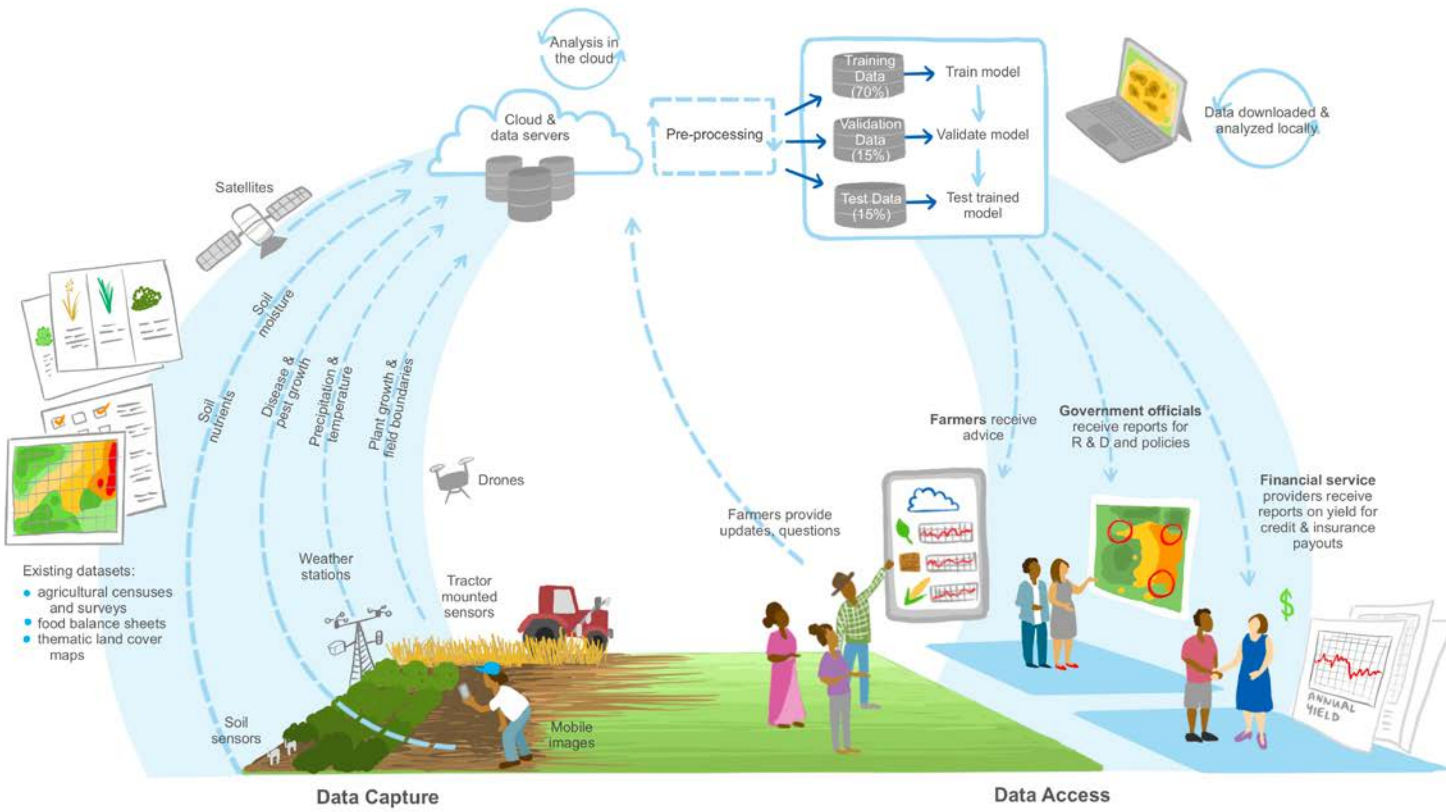


Figure 9:
Crop Analytics Data Cycle

2.1 Data Capture

Data capture is the process of collecting either ground-truth or remotely sensed information and converting it into data readable by AI/ML models. To understand the landscape, it is helpful to know what data organizations collect and how they are collecting them. Our stakeholder survey respondents named dozens of collected data types, with crop type, field boundary, and yield data cited most often (**Figure 10**).

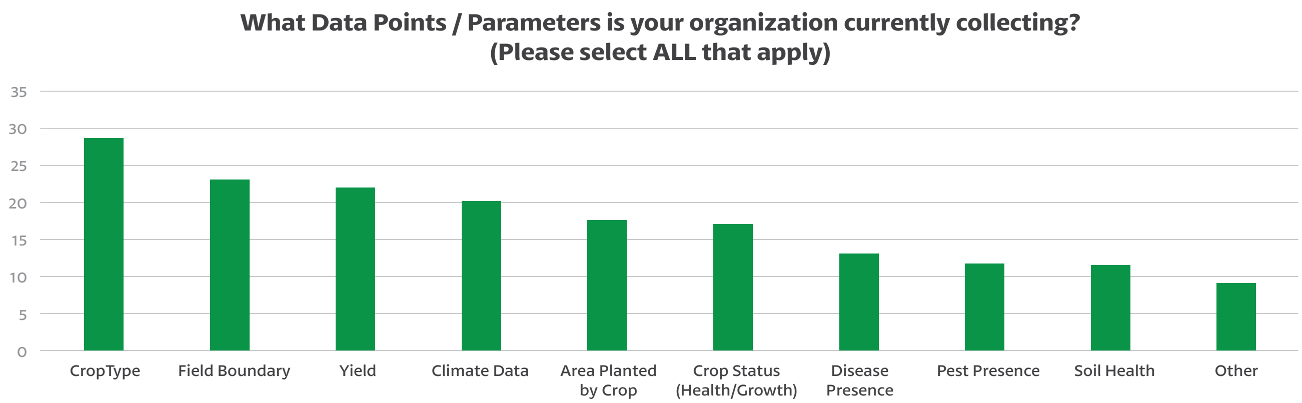
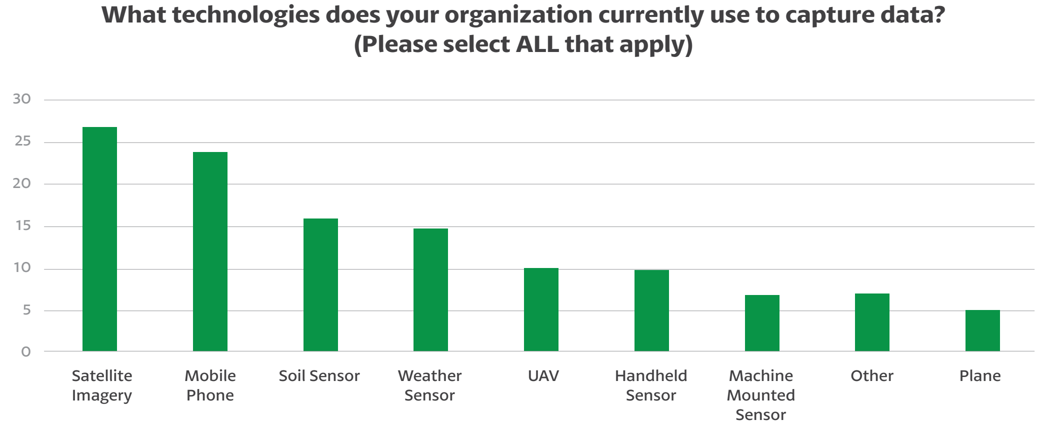


Figure 10:
Type of data collected by organizations

Data capture techniques have evolved, beginning with traditional surveys and crop cuts to now include a variety of emerging technologies and practices. Among survey respondents, the majority report using satellite imagery (27) and mobile phones (24) for data capture (**Figure 11**). The "other" types of agricultural data capture technologies reported by survey respondents include various handheld GPS units, tablets, and other portable positioning technology. Some organizations are also beginning to combine multiple data capture mechanisms using third-party sensors linked via API or using a single sensor station for multiple data streams (e.g., soil, weather, crop imagery). Below we focus on the landscape of actors working at the intersection of GTT and remote sensing data capture.

Figure 11:
Data Capture technologies



Ground-truth Data Capture: Overview and Partnerships

The landscape of organizations capturing agricultural data is vast. In certain geographies and for specific crops, researchers, analytics providers, and field-based project implementers have continued to better coordinate efforts to align data collection and sharing efforts. Yet often, the agricultural datasets generated by those operating in the field lack the geospatial or temporal characteristics that make the data usable in machine learning applications and derived crop analytics. To realize the potential of AI and ML technologies, crop analytic providers require training and validation or "ground truth" data to match the conditions on the ground at a given time and place with corresponding earth observation imagery. Ground truth data are used to train algorithms and improve accuracy or other performance metrics.

Despite recent gains in developed markets, there is not enough high-quality, timely, and interoperable GTT in smallholder agriculture systems to realize the potential benefits of advanced crop analytics. Smallholder farms are often dominated by heterogeneous cropping patterns and characterized by limited connectivity infrastructure, increasing the cost and difficulty of gathering GTT. The return on investment in data collection is difficult to achieve, especially for those organizations lacking extensive networks on the ground. In such settings, private sector actors often seek out unique partnership arrangements for ML-ready ground data capture. For example, analytic providers such as 6th Grain partner with BASF and other input suppliers to gather point of sale data, farmer profiles and locations, crop types, and sowing dates. This collection of proxy and ground data, combined with remote sensing information, can help drive yield models and ultimately tailor extension advice to farmers. Other actors such as the Lacuna Fund try to level the playing field by pulling together a collaborative of public funders and private foundations to provide data scientists, researchers, and social entrepreneurs with the resources they need to produce labeled datasets in low- and middle-income countries.

When considering the return on investment for GTT collection, various commercial and non-commercial actors value the collection and use of this data differently. For instance, some organizations prefer to compete on the quantity and quality of GTT they directly collect. For other organizations, GTT is a small step in the process of producing marketable products—these groups therefore welcome interventions that reduce the acquisition cost and effort of GTT, even if done by others. The World Bank and others are researching how to create a minimum viable standard for GTT collection, so stakeholders can save time and money when organizing collection campaigns [6].

It is important to define the real value of ground truth training data to coalesce a group of stakeholders around sharing data. For instance, the Marine Environmental Data Information and Network (MEDIN) quantified the sharing of marine data at a value of 8:1. For every British Pound spent on the marine data-sharing platform, users receive 8 Pounds worth of value [17]. With support from ECAAS, 6th Grain is testing a private sector GTT data-sharing platform with input and extension service providers willing to cooperate and see common value in a pooled GTT dataset. ECAAS will continue to quantify the value in future developments and business plans for various stakeholders and map where stakeholders perceive the value in ground-truth training data.

Table 1.

Ground Truth Capture Organizations

	Commercial	Non-Commercial
Data Capture Services	Vassar Labs Mesur.io Cropin Pula Premise 6th Grain Trimble CropIn Understory Intello Labs Skymet One Acre Fund Mesur.io Arable Hello Tractor Hummingbird Technologies aWhere	CGIAR The World Bank ICRISAT United Nations NASA Harvest (EO-Farm) Plant Village United Nations USAID Precision Agriculture for Development

The smallholder agricultural space has also seen an emergence of innovative Public-Private Partnerships (PPPs) between ag-tech companies, intergovernmental organizations, and public sector actors (**Box 1**). These partnerships are often formed to collect data for or help evaluate a specific program, but many have evolved to expand types of data collected or to identify new markets or funding opportunities. This is especially the case in smallholder agricultural settings where purely commercial models fail to return sufficient investment and require support from public entities or datasets.

BOX 1:
Innovative PPPs for Ground Truth Data Collection and Exchange

- The U.S. National Weather Service maintains a data exchange with multiple private weather companies.
- Ag-tech company Impact Terra runs a data exchange with Sathapana Bank for credit risk scoring, in partnership with Myanmar Ministry of Agriculture, Wageningen University, and others.
- The Maharashtra state government operates a data exchange with drone startup Pigeon Innovation, the World Economic Forum, and India Flying Labs.

Finally, organizations such as Farmer's Edge and Mesur.io, which leverage proprietary hardware and software for tailored extension in medium and large farm systems in high-income countries, increasingly seek to test models in smallholder settings. Recognizing the difficulties involved in such settings, many are willing to collaborate on data collection or sharing, as discussed in length in **Section 2.3**.

Ground-truth Data Capture – Processes

Traditionally, field and household surveys conducted by professional enumerators captured a significant portion of GTT used in training ML models. Survey data are well-documented, and enumerators attempt to follow accepted standards for capture, albeit with inconsistent application. Surveys can create a high volume of data that is helpful in identifying sampling points and developing sampling schemes for ML models but are typically expensive to deploy at scale. Periodicity is also an issue with AI/ML model calibration, as surveys are often not conducted frequently enough to build robust plant monitoring schemes or similar models. In some cases, survey data are inaccessible due to data security or privacy concerns, or ineffective data management practices. Even if accessible, datasets produced from surveys generally require significant amounts of data cleaning and processing to become machine-readable. Despite the challenges, several organizations are working to improve traditional survey methods and optimize sampling frames. These include the FAO Agricultural Integrated Survey Programme (AGRISurvey) and the World Bank's Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA).

Virtually all crop analytics actors also leverage a combination of non-survey data collection tools and methods, as summarized below (**Table 2**).

Table 2:

Non-survey Ground-Truth Capture Methods by Type of Data Collected

Data Type	Data Capture Methods
Climate data	Meteorological records, On-farm weather stations
Soil Data	Soil sensors, Laboratory testing
Field Boundary Data	Machine-mounted sensor, GPS-enabled mobile device, Drone, Cadastral records
Crop Type	Drone, Mobile photos, Input supply sales data
Sowing Date	Mobile phones, Input supply sales data
Crop health, including pest/disease prevalence	Drone, Mobile photos, Laboratory testing, Machine mounted sensor, Handheld sensor
Yield measurement	Drone, Crop cuts, Machine mounted sensor, Field trials

Many of these datasets often create other issues which limit use in AI and ML applications. Meteorological departments capture climate and weather data, but there is a problematic lack of weather stations in dozens of countries which can lead to inaccurate forecasts in those areas [18]. Ministries of Agriculture develop and often publicize soil health cards, yet sampling resolutions are often too coarse (e.g., 5x5 km) to be helpful in customized analytic products, and characteristics usually do not extend into micronutrient compositions. This significantly impacts the accuracy of predictive yield models, especially. Cadastral maps depicting boundary information in most smallholder systems are not digitized and infrequently updated, failing to capture changes in field shapes or land use across seasons. Stated security and privacy concerns and unstated norms resisting downstream monetization of public data result in public datasets becoming difficult to access for private analytic firms to access and use.

Academia and research institutions such as the Group on Earth Observations Global Agricultural Monitoring (GEOGLAM) 's Joint Experiment for Crop Assessment and Monitoring (JECAM), Consultative Group on International Agricultural Research (CGIAR), and institutions such as the K.J. Somaiya Institute Applied Agricultural Research (KIAAR) collect crop health and yield data through field trials. However, the scale of these campaigns is typically too small to be useful for many ML applications, is expensive, and is often very specific to a type of research finding or test. The competitive nature of research funding also inhibits sharing of pre-published research or data in the absence of specific coordination or agreement structures.

Promisingly, the recent proliferation of ag-tech hardware opens a new pathway for less-expensive GTT capture at scale. Soil sensors (Arable), private weather stations (Skymet), and machine-mounted (John Deere) or handheld sensors offer the ability to capture highly accurate localized data quickly. Organizations such as the National Center for Atmospheric Research (NCAR) and Vinduino are developing low-cost and offline-capable data capture technologies such as 3D printed weather stations and gypsum soil sensors. The widespread penetration of mobile phones and specifically smartphones present a pivotal opportunity to scale data collection. Organizations such as Penn State's PlantVillage team use a combination of web and app-based surveys, camera imagery, GPS capacities, and other tools to conduct crowdsourced data collection. Government extension networks rely on calling and SMS messaging, which are being deployed at scale and complimented in the private sector by Interactive Voice Response (IVR) by Farm.Ink and others.

Unmanned aerial vehicles (UAVs) such as drones have provided another pathway for collecting high-resolution imagery at the field level, using multisensory technologies to capture "wall to wall" data for validation. These are used most widely within the private sector but are increasingly supported by public authorities through, at a minimum, a reduction in airspace restrictions. Depending upon their loadout and operating parameters, drones also occupy an interesting position at the intersection of ground truth and remote sensing data capture, which we delve into in greater detail in the following section.

Remote Sensing Data Capture

Remote sensing agricultural data capture began primarily with a combination of aircraft and low-resolution satellite-based optical imagery platforms. As additional public and later private space programs came online, technologies and datasets expanded to include higher resolution imagery, radar, and multi-spectral bands. As noted above, drones have emerged as cost-efficient, albeit scale-dependent, means to capture very high-resolution (VHR) imagery at small extents, allowing for further expansion beyond the established aerospace industries in developed nations. As a result, today, the marketplace of agricultural remote sensing data providers includes a wide variety of organizations across the world (**Table 3**).

Table 3:
Organizations in Remote Sensing for Data Capture

	Commercial	Non-commercial
Satellite Providers	Earthi Maxar Technologies Airbus Planet Spire Capella Space Satellogic Iceye	NASA ESA JAXA CNSA CSA ASC POCKOCMOC KARI
Drone Providers	FluroSat Trimble IdeaForge Microdrones PrecisionHawk American Robotics AUS DJI Parrot	Deseret WeRobotics
Satellite Standards Developers		Committee on Earth Observation Satellites Global Agricultural Monitoring Open Geospatial Consortium (OGC) SpatioTemporal Asset Catalogs (STAC)*

*Supported by an active developer community from multiple organizations

The public sector has traditionally led satellite data capture and provides much of the imagery used in crop analytics today. Several national governments operate satellite systems that capture a broad range of earth observation data, with traditional leaders spread across North America (NASA), Western Europe (ESA), and East Asia (JAXA). Almost all crop analytics organizations use public satellite data for at least a portion of their base imagery because it is freely available, has a constant revisit rate, and has a deep archive of data from which to draw. Actors from the public and civil society sectors such as Technische Universitat (TU) Berlin's BigEarthNet may solely use public imagery, as commercial imagery can be cost-prohibitive in the absence of a data-sharing agreement and may not have the required coverage. Often organizations will favor imagery sets from their public funding counterparts even if not required to do so.

Despite the benefits of public satellite data, the spatial resolution is generally insufficient for smallholder mapping and monitoring [19] [20]. The fragmented and unclear nature of field boundaries and complexity of intercropping practices require the use of high spatial resolution data (less than 5m), and the highest resolution entirely public satellite is 10m (ESA's Sentinel-2). Periodicity is another critical issue of public satellite data. The Sentinel-2 mission has a 5-day global revisit periodicity while NASA's Landsat satellite, on the other hand, has a weekly revisit rate. There are also a few other public satellites that orbit faster, like the Soil Moisture Active Passive (SMAP) satellite, with a 3.5-day revisit rate, or Geostationary satellites which collect a continuous stream of data for a fixed point on earth. Public satellites tend to be large and expensive to build and launch. The delay between initial design and operation can be as long as a decade.

Commercial investment in satellite imagery infrastructure has rapidly created new sets of data sources in the landscape. Commercial satellites are generally smaller, less expensive, and more conducive for rapid iteration and innovation. Recent satellite iterations can capture imagery at much higher spatial resolution (as high as 0.25-meters) with faster revisit rates. In addition, the CubeSat open standards enable start-ups to build and test satellites quickly at low costs (**Box 2**). Leading actors in the commercial satellite imagery space, such as Planet Labs and Maxar, are concentrated in North America, though the field is rapidly expanding. For example, 18 new satellite start-ups from Western Asia, Australia, and Western Europe were launched in 2018.

BOX 2: CubeSat

A CubeSat is a type of small satellite for space research that is 10 cubic centimeters in size and weighs no more than 1.33kg. CubeSats often use commercial off-the-shelf components for their electronics and structure, and are deployed as parts of larger missions. Their open-sourced specifications for launching and space exploration make them an accessible satellite for public and private sector players. Cubesats are also a source of high-resolution imagery that could be used in crop analytics. As of August 2021, more than 1,500 CubeSats had been deployed in space.

Commercial satellite imagery has challenges that limit application in crop analytics. Specifically, in smallholder farming systems, the willingness or ability of farmers to pay for analytic products is often limited, making it more difficult for analytic providers to achieve a return on investment for satellite imagery access licenses. In certain situations, licensing structures prevented partnerships from forming as satellite companies were bound only to provide imagery to specific organizations, but these arrangements are decreasing in prevalence. Despite the capital-intensive nature of the business, some commercial imagery vendors provide imagery at a lower cost or even free of charge to organizations with a social mission, such as intergovernmental organizations, national government agencies, academia, and non-profit organizations. For example, researchers from academia can apply for Planet's Education and Research Program for limited access to new and archived high-resolution (HR) imagery.

2.2 Data Processing

Newly captured agricultural data (also called "raw data") often cannot be immediately used for analytics. Whether raw data are captured through field collection methods or remote sensing, it must be processed (cleaned and labeled) and made machine-readable before it can be used as training or validation data. Data cleaning is required to edit noisy data-sets, remove or change erroneous values, or otherwise manipulate data to overcome issues caused by inconsistent collection or data formatting. Data labeling involves taking remotely sensed imagery data (images, text files, videos, etc.) and adding one or more meaningful and informative tags to it for context so that an ML model can learn from the data. To avoid the concept of "garbage in, garbage out," as frequently articulated by the ML community, models need large volumes of high quality and labeled data during their training phases to produce properly calibrated results. While a necessary step for crop analytics, the process of cleaning and labeling datasets can be expensive and time-consuming – the Committee on Earth Observations (CEOS) estimates that users of Earth Observation data spend around 80% of their effort on this step. Organizations in the data processing space include those that provide platforms and services for data labeling and cleaning (**Table 4**).

Table 4:
Data Processing Organizations

	Commercial	Non-commercial
Labeling Platform	Azavea Google Earth Engine* Mapbox Maxar	OpenStreetMap GEO.Wiki Collect Earth Online Humanitarian Open Street Map Team ESA Sentinel Hub Training Dataset Platform (TDS-Platform**
Labeling Services	Amazon Mechanical Turk Hive Cloud Factory GEOhive MAXAR	BigEarth Digital Earth Africa Humanitarian Open Street Map Team
Data Cleaning Platform	Microsoft Azure ESRI Google Earth Engine GeoTrellis Amazon Web Service	Digital Earth Africa G.E.M.S
Data cleaning Services	Descartes Labs Azavea Atlas AI	Radiant Earth Foundation CGIAR NASA Harvest Wageningen University

*Some GEE users maintain non-commercial access to the commercial platform
**Project is in current development

To reduce this effort, many groups outsource data labeling to organizations such as Azavea or other service providers that employ trained labelers to produce high-quality labeled datasets. Crowdsourced labeling services (Amazon Web Services' Mechanical Turk, Premise, Maxar's GeoHIVE) offer an inexpensive alternative to trained labelers. Yet, these labelers may lack the subject matter expertise required to support a robust quality control system to ensure accuracy, especially in agriculture. Different approaches have been tested and can be contextually appropriate to engage crowd labelers. These approaches include pay-for-work models, gamifying the labeling process, and engaging volunteers based on humanitarian needs.

For data cleaning, if data are stored in or linked to an external repository, the organization that operates the repository may assist in the cleaning process. For example, the Radiant Earth Foundation supports the data cleaning and processing of GTT in exchange for making datasets publicly shareable on the MLHub. In contrast, the Genetic, Environmental, Management, and Socioeconomic (G.E.M.S.) data platform offers tools for the owner to clean GTT before uploading it to the platform and later charges a fee for access to meet costs involved with the repository (**Box 3 next page**). Automatic cleaning of remotely sensed imagery is further ahead than automatic GTT cleaning, with NASA's ACCESS and EOSDIS projects using multi-temporal anomaly detection as part of the pipeline for ingesting Synthetic Aperture Radar data. NASA's Land Processes Distributed Active Archive Center (LP DAAC) also developed a fusion processing method that provides unprecedented 30-meter spatial resolution imagery every two to three days for the Harmonized Landsat Sentinel-2 (HLS) dataset. In addition, Descartes Labs has built custom pipelines for several publicly available data sources to automatically ingest and process data as soon as they are published. As computing power improves, the volume of data used in training has increased, making manual cleaning increasingly unfeasible and advances in automated data cleaning even more critical. The remote sensing community at large has invested lots of effort in promoting analysis-ready data (ARD), which is a great example for the GTT community as it drastically increased the feasibility of automated data cleaning.

BOX 3: G.E.M.S

Led by a partnership with University of Minnesota and Minnesota Supercomputing Institute (MSI), the G.E.M.S platform advances machine learning techniques to better monitor global agricultural and environmental change. GEMS is a secure web-based platform for exploring, sharing, and analyzing data, workflows, and results. A subscription is required to access all GEMS features, functionality, and tools, including both the GEMShare and GEMSTools features. The GEMShare function of the platform enables data providers to control who sees what, and when, thus using the intrinsic, multi-faceted value in the data to incentivize innovation partnerships for the mutual benefit of the collaborating partners. GEMSTools allows analysts to select from the suite of already available analytical tools or load their own analytical tools on the platform.

2.3 Data Storage

The volume of data generated by stakeholders operating in smallholder environments and the potential of these data for improving smallholder analytics products is massive. However, much of this value is locked away in siloed datasets, separated by disparate storage and sharing approaches. In some cases, stakeholders are unwilling to share data at all. That said, among survey respondents, 63% reported that they share some data either publicly on their websites or via external websites or platforms (**Figure 12**).

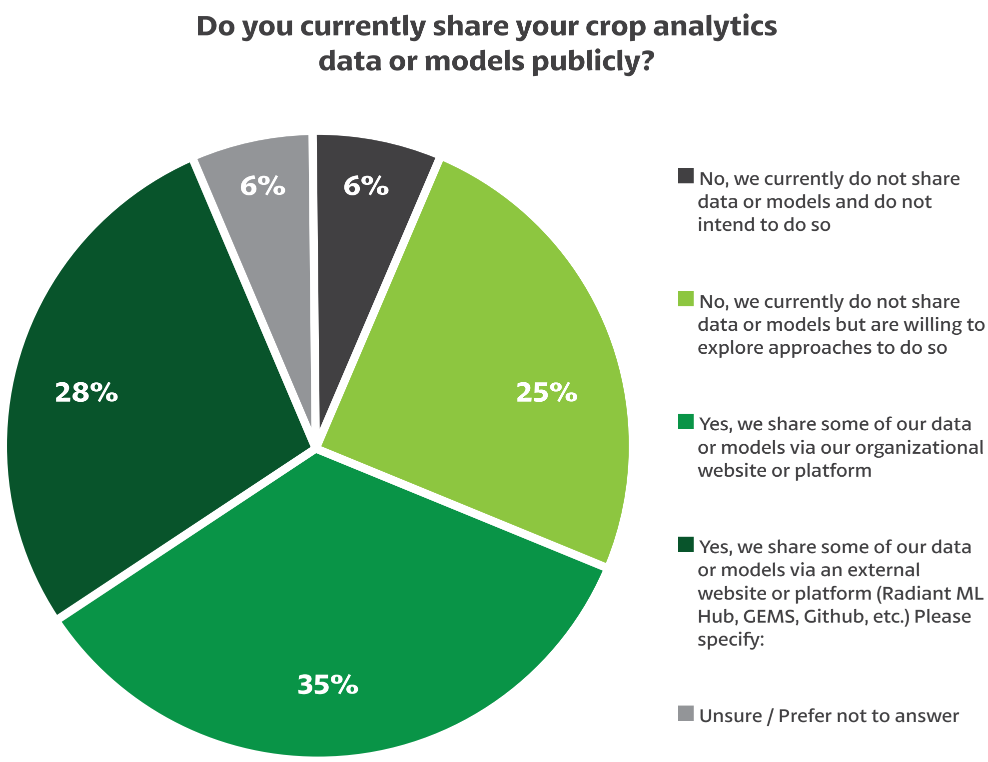


Figure 12:
Organization's participation in data sharing

Even when actors are willing to share data, other issues like storage methods and inoperable data formats interfere with the integration of data into those systems and subsequent re-use by others. Therefore, to democratize access to data, storage methods and data formatting become critical issues to address. Technological advances such as cloud storage and The SpatioTemporal Asset Catalog (STAC) specifications are helping to overcome these barriers, yet adoption remains limited and inconsistent.

The barriers to improved data sharing are not only technical, as decisions around data sharing are influenced by a range of factors, including data privacy, regulatory constraints, and limited incentive structures for sharing between actors. Several organizations, including Radiant Earth and the Open Data Institute, are working to improve incentives for public data sharing and crowd in actors that traditionally do not share data, especially in the private sector. However, these efforts remain fragmented and require coordination and community buy-in to scale. No commercial actors were identified that are working specifically on agricultural data repositories for smallholder production systems, despite all the major infrastructure providers being commercial entities (**Table 5**).

Table 5:

Organizations working in data storage and sharing

	Commercial	Non-commercial
Storage Infrastructure Provider	Amazon Web Services Google Cloud Microsoft Azure	N/A
Data Repositories		Big Earth SpaceNet G.E.M.S Digital Earth Africa Radiant Earth Foundation Wageningen University Laco-wiki ESA TDS Platform One Map*

*Project currently in development. It is created in partnership with the FAO, World Bank, GEOGLAM, CGIAR, Digital Green, HPE, World Economic Forum, and Mineral (Google).

Data Storage Systems

Historically agricultural data have been stored via physical records or locally on private servers. Many organizations still follow these practices to maintain proprietary data in-house or to abide by strict security protocols for potentially sensitive datasets. Academia, research organizations, and NGOs collect data for specific projects and store it locally for ease of access and use, but often do not share data even after the project has closed or publication has occurred. There are limited incentives to share near real-time data within the ecosystem actively. However, the AgroSTAC initiative and others seek better ways to connect research teams and their datasets to find opportunistic synergies (**Box 4**).

BOX 4:
AGROSTAC

AgroSTAC is a global crop trial repository developed by VITO and Alterra in the FP-7 SIGMA project to visualize and explore time series of in-situ data for improved agronomy. Field data are collected from researchers at sites around the world. The catalogue accepts agronomic data relevant for calibration/validation purposes in the domain of crop production monitoring by satellite data. Examples of data catalogued by STAC include crop type, phenology, biomass, and yield.

With the massive growth of cloud computing, all major providers are beginning to offer and develop agriculture-specific services that reduce the cost of storing and managing large datasets. In descending order, the most used services are Amazon (Amazon Web Services), Alphabet (Google Cloud), and Microsoft (Azure). While many private sector actors use cloud-based storage, the cost of this service is often prohibitive for NGOs or other organizations which manage small amounts of data and do not frequently share it. In these cases, open data initiatives offer storage at no cost in exchange for publicly sharing datasets. Amazon's Open Data Sponsorship Program covers the cost of storage for publicly available high-value cloud-optimized datasets in specific domains. Google's Cloud Platform provides free storage for public datasets, aggregating them in a single location and integrating with Google's analytics platforms for efficient in-place analysis. Both Amazon and Google offer these services to achieve corporate social impact goals and attract users to other revenue-generation services on their platforms.

Data Formatting

Crop analytic stakeholders use a wide array of multi-dimensional datasets from a disparate set of sources. Stakeholders most commonly use scene-based file download, but this requires significant storage and computing power. Earth Observation (EO) data cubes gained considerable attention within the community for their potential to improve storing, processing, and accessing of EO data like the now deprecated Africa Regional Data Cub (ARDC) (Box 5). Data cubes and similar structures house and process huge amounts of EO data from multiple sensors on cloud servers, freeing up space on devices and improving accessibility to the massive amount of remote-sensing data for AI/ML training. For instance, data cubes allow analysts to select one dimension (e.g., a specific field) and view all data related to that dimension (i.e., the temperature in a given location from 1990-2020) through an efficient and flexible programming interface that significantly simplifies access to freely available satellite data about the world.

BOX 5: Africa Regional Data Cube

The Africa Regional Data Cube (ARDC), decommissioned in 2020, was an open-source multidimensional infrastructure which combined datasets from multiple sources, geographies, timelines, and types in an analysis-ready format. This enabled the data user to invest less time in discovery and processing and enabled higher quality analysis. ARDC covered five countries and provided a prototype for Digital Earth Africa to scale an accessible platform for accessing high quality, analysis ready, Sentinel and Landsat data across Africa.

For data formats, the community has long used tagged Image File Format (TIFF) graphic files. The Radiant Earth Foundation, Maxar, and others are now investing in Cloud Optimized GeoTIFFs (COGs) to further optimize satellite imagery access, and many other community actors are beginning to follow suit. COGs trace back to the 2016 Open Source Geospatial Foundation. They allow users to request portions of a dataset rather than the entire file, significantly reducing storage and processing loads and associated costs. Within GTT formats, there are nascent standards for data formats and storage, yet none have been widely adopted among the entire community.

2.4 Data Sharing

The data-sharing environment describes the extent to which the current systems exchange data, and the ability to interpret that shared data. For two systems to be interoperable, they must be able to exchange data and subsequently present that data such that it can be understood. Frictionless access to information and tools to use data would improve the data sharing environment and generate new policy solutions, lower costs of data collection, and accelerate innovation.

Accessibility and Sharing

High-quality ML-ready data currently reside within a variety of disparate locations and storage platforms. Public organizations such as NASA, ESA, and the USGS generally publish remote sensing data openly on their respective websites. While technically accessible, users must search and download large datasets from a wide range of locations, requiring significant computer and storage capacity. The EO data cubes provide a solution to this issue as they pull large amounts of EO data from multiple sensors onto cloud servers, freeing up space on devices and improving accessibility.

To further reduce storage costs, data owners may also elect to store or share data via an external repository or platform such as Radiant Earth Foundation's Radiant MLHub, University of Minnesota's G.E.M.S. platform, or TU Berlin's BigEarthNet. Some existing repositories contain datasets that are geographically concentrated (G.E.M.S primarily in North America), while others include homogeneous data source types (MLHub and AGROSTAC contain open data primarily from donor and publicly funded initiatives and universities). Some repositories do not explicitly focus on agricultural ground data used for advanced crop analytics but provide expansive earth observation datasets that analytic organizations can use for new ML models (BigEarthNet, ARDC in Africa).

Organizations may prefer to store and publish data internally for various reasons, including concerns over security, fear of losing control of the dataset, or interest in a user accessing the dataset via an organizational website or platform. Several organizations have presented innovative structures to address these concerns, including API-enabled repositories like Radiant Earth's MLHub, data marketplaces such as Google Cloud's Public Datasets Marketplace, and robust security controls such as those built into the G.E.M.S Platform. Yet there is not a leading repository structure that is widely used among many types of stakeholders working across a wide variety of smallholder production regions.

Interoperability and Standards

The advent of accessible technologies has improved the availability of data. However, to be maximally useful, such data should be compatible. To date, data collection efforts do not follow specific standards, thereby limiting the potential to use them in combination with other datasets. There is also less focus around higher-level standards that could improve data collection at scale.

Several organizations have attempted to create a set of standards (e.g., STAC, AIMS), but additional efforts are needed to collaborate towards and endorse common standards and data collection procedures to promote maximum program effectiveness. The SpatioTemporal Asset Catalog (STAC), as championed by the Radiant Earth Foundation (Box 6), is one attempt to normalize and gain broad acceptance around a specific set of standards. While many organizations are moving toward adopting these standards and mechanisms, others continue to follow either their own custom data standards or those developed by the FAO's Agricultural Information Management Standards (AIMS) standards and guidelines for community data documentation.

BOX 6:
SpatioTemporal Asset Catalog (STAC)

STAC is an open specification which provides common metadata and API mechanics to access geospatial data. The aim of the STAC specification is to provide a standard structure for dataset storage and tagging to enable any API to discover the data. The end objective is to develop a global index of all imagery and derived data products.

Beyond data standards, tangible planning around collection procedures and collection guidelines could improve the quality, interoperability, and scalability of data. The World Bank's 50x2030 Initiative recently produced best practices for GTT collection surveys report that supports satellite-based crop type mapping in Sub-Saharan Africa. The report demonstrates how better sampling strategies lead to better insights, and how sampling the right amount of data needed for models, not more and not less would reduce data costs associated with data collection.

Examples from other domains have demonstrated that such data standards and collection guidelines are not only possible but desirable. For instance, a centralized catalog accessed via Application Programming Interfaces (API) could connect various existing repositories and storage platforms. Whether data are stored locally or in an external repository, data can be made discoverable and potentially accessible in a central location via such a mechanism. The Alzheimer's Disease Data Initiative and Building and Land Development Specification (BLDS) saw success utilizing a similar mechanism, albeit in different technical domains.

2.5 Data Analysis

Properly tested ML applications benefit actors throughout the agricultural ecosystem by aggregating disparate data and information and providing helpful insight and analysis. Typical applications of ML for crop analytics include classification (field boundary identification, land cover), crop condition assessment (growth stages, health), and yield (estimation and prediction). Leading organizations in this space include Analytic services providers, geospatial data platform providers, and challenge platform coordinators (**Table 6**).

Academic and research institutions are essential early-stage innovators in ML-based crop analytics. Leaders are geographically clustered in North America (Stanford University, University of Maryland) and Western Europe (Wageningen University, TU Berlin). Universities conduct frontier research in machine learning approaches, including new models and techniques for applications in crop analytics. Occasionally, these move to full application development and field deployment, as with the "Nuru" application from Penn State University. However, their innovations often remain siloed within academia, diminishing their scalability in the field. The NASA Harvest program highlighted below provides an example of a consortium designed to scale initial research through partnerships with service providers from the private, public, and voluntary sectors (**Box 7**).

BOX 7:
NASA Harvest Consortium

Led by the University of Maryland, NASA Harvest is a multidisciplinary consortium of partners from more than 50 institutions with the goal of enabling and advancing adoption of satellite Earth observations by public and private organizations to benefit food security, agriculture, and human and environmental resiliency. This mission is achieved through a series of partnerships and funding for projects which are co-developed with end users to ensure long-term sustainability. In these exchanges, the NASA Harvest team provides analysis to the end users for land use and management planning, policy development, production forecasting, and project design, monitoring, and evaluation.

Outside of academia, those working in publicly or philanthropically funded analytics include CGIAR, NASA, the European Space Agency, the World Bank, FAO, and various for-profit and non-profit companies. This work is generally project or geography-specific but often focuses on improving productivity or reducing risk in smallholder farming systems. The result of this work is generally publicly available analytical products, even if these are difficult to access due to previously mentioned data silos.

Private sector analytic providers often have venture capital or mixed funding streams that can help them invest in R&D and infrastructure to produce high-quality analytics. There is a nascent cluster of public-private partnerships aimed at expanding for-profit ventures into otherwise less attractive smallholder data environments, but a comprehensive study of key lessons from this work has not yet been documented in this domain. Key private sector analytics providers are clustered mainly in North America (Descartes Labs, Atlas AI) and Western Europe (VITO, Hummingbird Technologies). These organizations may provide analysis on behalf of another organization or conduct their own analysis to provide a commercial analytic service to various end-users such as farm management specialists (The Climate Corporation, Granular Inc). Analytics providers use proprietary datasets and algorithms which they can use to improve their services and gain a competitive edge in the market. Private sector organizations would benefit from open data/data sharing but could create data asymmetries when they do not share back with the community.

Traditionally, data analytics were performed locally and required significant computer processing speeds and storage. Cloud computing is changing how scientists study data by allowing for data analysis "in place" without the need to download and analyze data locally. The private sector has led cloud adoption, yet the public and voluntary sectors are increasingly following suit. The European Space Agency (ESA) and the United States Geological Survey (USGS), for example, have established strategies for moving data processing and analysis to the cloud. Cloud computing has also opened the way for developers to build and deploy cloud-based analytics applications and platforms such as Esri's ArcGIS Online, Google's Google Earth Engine, and Maxar's GBDX platform.

Table 6:

Prominent Organizations in Data Analytic Services & Platforms

	Commercial	Non-commercial
Imagery Analytics Providers	Maxar Planet Descartes Labs Orbital Insight AtlasAI Applied Geosolutions 6th Grain Indigo Ag OneSoil Doktar SatAgro	NASA The World Bank NASA Harvest CGIAR ESA USGS United Nations
GeoSpatial Data Cloud Computing Services	Orbital Insight Microsoft Azure ESRI Trimble Google Earth Engine Maxar	Digital Earth Africa Open Data Cube Open EO Pangeo
Challenge Platforms	N/A	SpaceNet Radiant Earth Foundation DrivenData

2.6 Data Regulatory Environment, Privacy, and Data Ownership

The unprecedented amount of agricultural data that has become available in the last few decades increases the risk of exacerbating digital divides, information asymmetries, and other power imbalances which could disproportionately affect smallholder farmers. ECAAS can play a role in addressing these risks through the support of responsible data regulation and policy.

The responsible collection, storage, and use of real or potentially personally identifiable or other sensitive data are essential to maintain trust in the crop analytics data ecosystem. Several ML applications such as tailored extension and advisory services require farm-level, geo-referenced data. Collecting and publishing these data presents a high risk for abuse and misuse. As a result, data at the farm or field level are often less accessible. A 2015 study by the Organization for Economic Co-operation and Development (OECD) reported that 45% of OECD government organizations surveyed do not provide farm-level data for non-government third-party actors under any circumstances. Only 41% allow access after someone requests access and specifies the intended use for the data. Partnerships such as the World Bank's 50x2030 initiative and Atlas AI are advancing methods such as geographic fuzzing to anonymize data to mitigate potential privacy concerns. Currently, more than 80 countries have implemented data protection laws underpinned by OECD data protection principles (**Box 8**).

BOX 8:
OECD Data Protection Principles

1. Purpose Specification Principle. The reason for data collection must be specified.
2. Use Limitation Principle. Personal data should not be shared, made available, or used for purposes other than those specified at the time of collection.
3. Collection Limitation Principle. Data should only be collected if they are directly relevant and necessary to accomplishing the specified purposes.
4. Transparency/Openness Principle. Policies and practices around data collection and use should be made available.

box continued on next page

5. Data Quality Principle. Records should be relevant to the stated purposes of use and should be accurate, complete, and up to date.
6. Individual Participation/Consent Principle. The knowledge and consent of the individual are required to collect, use, store, and share personally identifiable data. The individual retains the right to be informed of the existence, use, and disclosure of data relating to that individual and request access to these data. This information must be communicated to the requestee in a reasonable time, format, and cost, and legitimate reasoning must be given if the request is denied. The individual also retains the right to challenge the accuracy of data and have the data erased or amended.
7. Security Safeguards Principle. Personal data must be protected through appropriate security safeguards against risks of loss or unauthorized use, destruction, modification, or disclosure.
8. Accountability & Auditing Principle. The data controller should be held accountable for complying with the above principles, providing training to employees and contractors accessing personal data, and auditing the use of data to demonstrate compliance with these principles. An individual is able to challenge a data controllers' compliance with the above principles as it relates to that individual's data

In practice, the laws and related requirements for data capture, storage, and use vary widely from one jurisdiction to the next. By default, many crop analytics stakeholders adhere to the EU's General Data Protection Regulation (GDPR) which is, in some ways, the most restrictive framework. GDPR forms the basis for many other national regulations, such as Kenya's Data Protection act. India's Personal Data Protection Bill goes further, forcing data collected in India to be stored there. These regulations, and the application of the law at various levels of government in any given country, change somewhat often and force stakeholders working across boundaries to dedicate resources to remain in compliance.

In addition to legal frameworks, the ethics and norms of ground or agricultural data sharing and use vary significantly from one culture to another, increasing the difficulty of data exchange across cultural borders. Several non-governmental bodies have released data privacy and sharing guidelines to help provide general guidance and a baseline of data ethics and norms. They include the American Farm Bureau Federation's Privacy and Data Security Principles for Farm Data and the EU Code of Conduct on Agricultural Data Sharing by Contractual Agreement created by COPA-COGECA & CEMA. At this time, no single set of norms or guidelines have been accepted by the entire community, but most public or civil society groups start with FAIR (findability, accessibility, interoperability, and reusability) principles as a baseline position.

3

Stakeholder Network

As noted in **section 2**, a majority of survey respondents within the smallholder agricultural crop analytics ecosystem are either already sharing or are willing to share at least some amount of relevant data. To further understand the landscape and thinking within it, and to identify ongoing and potential collaborating organizations, we conducted a social network analysis (SNA). The SNA identified organizations that are key for *connecting* other organizations within the network, as measured by degree centrality. In general, organizations with a high degree of centrality are local network cluster hubs, well-connected within a specific pocket of the ecosystem, even if not within the wider network. The SNA revealed that MercyCorps, CropIn, NASA Harvest, and USAID had the highest degrees of centrality, in addition to Tetra Tech (which administered the survey). With that being said, while these results are useful to triangulate the overall findings and directions of the network, it is important to note that the findings reported in this report are dependent on our survey sample.

The SNA also identified those organizations that *bridge* different clusters of organizations within the network, as captured by the concept of betweenness. Betweenness measures how many times an organization lies on the shortest path between two other organizations. In general, organizations with high betweenness have more control over the flow of information. These groups can act as key bridges within the network for data sharing or other purposes or potential bottlenecks. The five organizations with the highest betweenness in the network are MercyCorps, CropIn, Digital Green, aWhere, and Measure.io.

The SNA revealed that the ECAAS project currently serves as one distinct hub in the network. Even though other organizations are clearly also collaborating and working together, they are not doing so as a distinct community. Organizations with the ability to influence the entire network are measured by reach in an SNA. Reach is an SNA metric that quantitatively measures the portion of the network within two steps of an organization. In general, organizations with high reach can spread information through the network through close friend-of-a-friend contacts. The network contains five organizations with a reach of 1, meaning they are at most one step removed from any organization in the network. If any of these organizations change their policies, standards, or protocols, the effects will be felt quickly throughout the entire network. These organizations include Planet Labs, Google, Amazon Web Services, ESRI, and Airbus. A concerted effort must be made to monitor the activities, and any potential policy or standards changes from Planet Labs, Google, AWS, ESRI, and Airbus as these will heavily influence the behavior of all actors in the network.

4

Geographical Distribution of the Landscape

Despite an overall fragmentation of the sector, stakeholders have formed clusters around specific geographies and technical areas in the data chain. Data from our surveys also point to clear geographic network gaps that could be addressed through intentional coordination mechanisms.

There are two major stakeholder clusters in North America and Europe, and emerging clusters in India, Kenya, and China. Geographically, the most mature cluster is concentrated in North America, in part due to the NASA/USGS Landsat program, which has provided publicly available imagery services for more than 40 years. This concentration intensified due to the emergence of Amazon, Google, and Microsoft as leaders in data hosting and cloud computing capacities. The cloud computing and hosting infrastructure offered by these companies radically decrease the cost of using big data in Artificial Intelligence (AI) and Machine Learning (ML) applications for crop analytics.

In the early 2000s, a second cluster emerged in Western Europe around the ESA's then-nascent Copernicus Program, which also produced multi-spectral imagery that was made available to the public free-of-charge. In recent years, smaller clusters of ag-tech start-ups have emerged in Bengaluru, Beijing, Nairobi, and other tech hubs with connections to and, for the most part, focuses on end-use applications for largely untapped smallholder farming markets. Regardless of office locations, the deployment of these technologies spans the globe (**Figure 13**).

Country Count

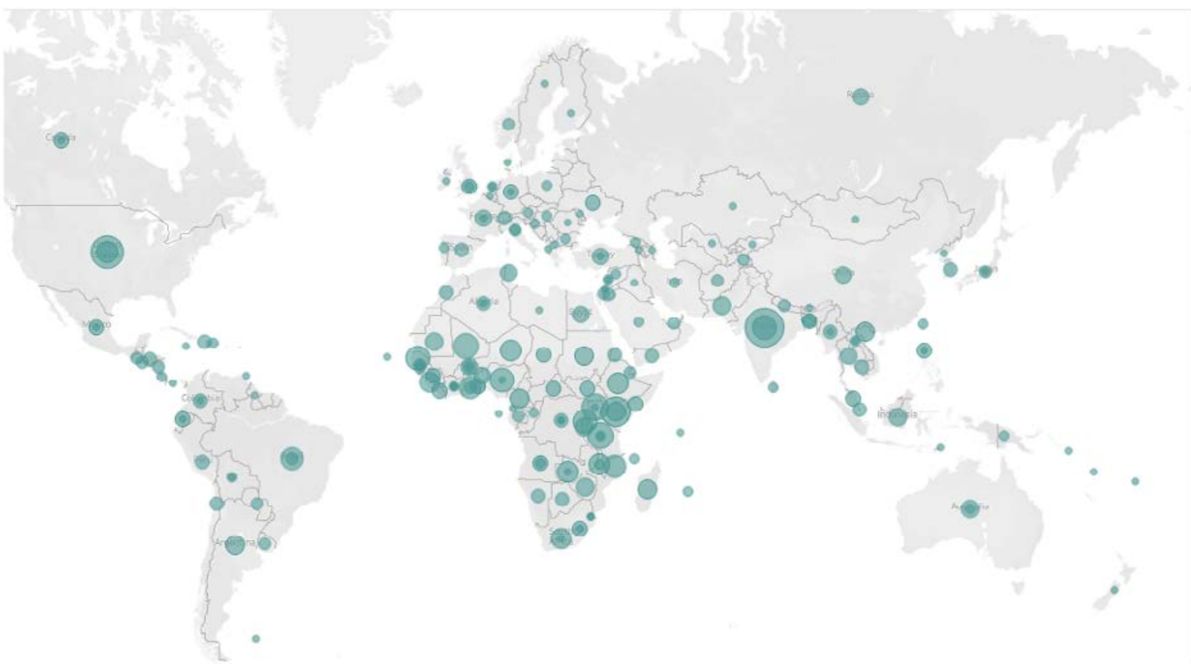


Figure 13:
Locations and focus markets of crop analytics stakeholders identified to date

While major clusters of organizations are based in North America and Europe, our survey respondents (focused on smallholder systems) indicated that they have the highest operational and data collection presence in Eastern Africa and Central & Southern Asia (focused mostly by India) (**Figure 14**). Broken down further, survey respondents indicated that their organizations have the strongest presence in the smallholder markets of Kenya, India, Nigeria, and Côte d'Ivoire (**Table 7**).

Figure 14:
Top 5 Regions of Operation

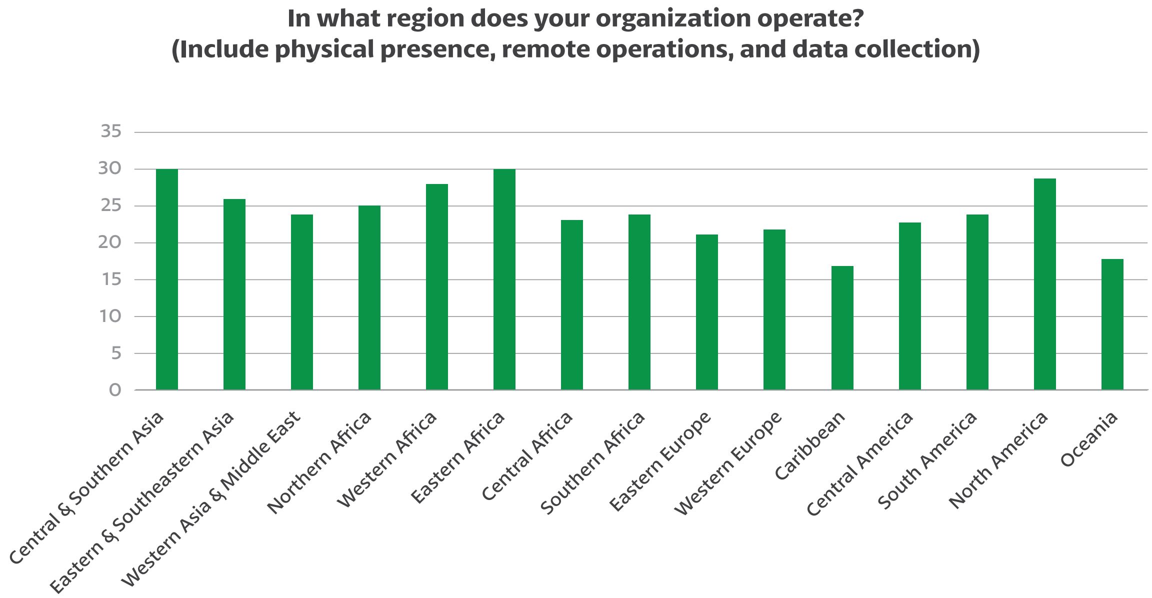


Table 7:
Countries of Operation

Country	Count	Count
Kenya	15	1
India	13	2
Nigeria	8	3
United States	7	4
Côte d'Ivoire	5	5
Brazil	4	6
Ethiopia	4	6
Indonesia	4	6
Zambia	4	6
Canada	3	10
China	3	10
South Africa	3	10
Tanzania	3	10
Uganda	3	10

Most survey respondents primarily interact with farmers through an intermediary such as a supplier, bank, or extension agent. The minority of respondents (4) reported that they primarily interact with farmers directly (either on-site or remotely) (**Figure 15**). This variety of channels further exacerbates the difficulties in standardizing GTT capture and formatting, since intermediate organizations often have requirements and processes which may not be optimized for the collection of ML-ready data.

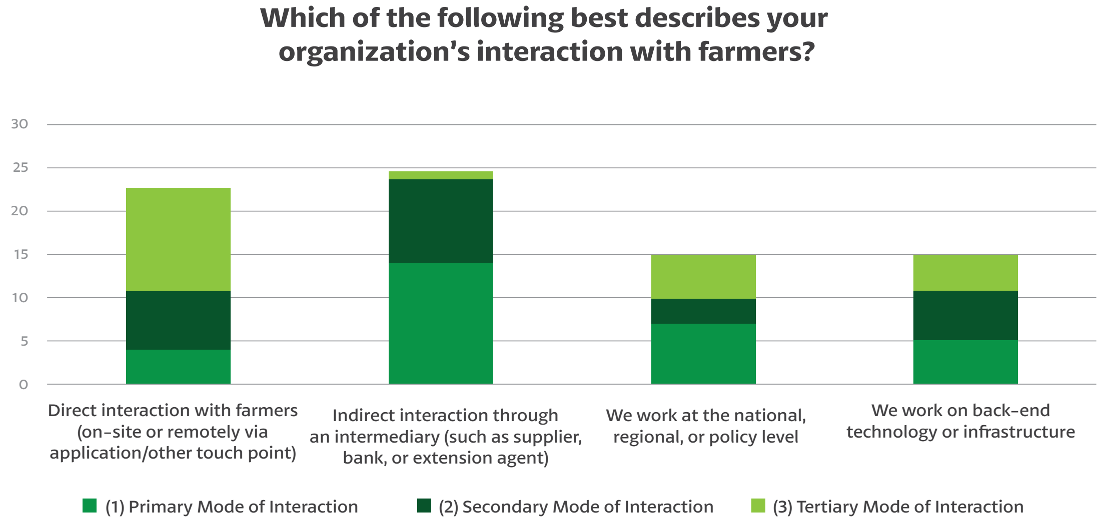


Figure 15:
Organizational Interactions with Farmers

The data show that clusters of stakeholders have formed more around specific components of the data chain than around specific geographies. However, the linkages between technical clusters are often ad hoc and reliant on individual, rather than institutional, connections. The channels and partnerships that could, for example, ensure that newly launched satellites host the sensors and technologies that analytics firms need and value, are primarily opportunistic rather than comprehensive.

This SLA serves as an initial baseline and a starting point for understanding the network. While our survey respondents were mostly members of the North American network, our broader analysis did not reveal any additional distinct regional hubs outside of North America. These findings leave open two possibilities regarding the interactions between the North American network and a European network or other regional hubs. Either 1) there is limited close collaboration between North American and European organizations, or 2) that collaboration exists but does not include organizations that constitute the core of the North American organizations that were surveyed. The best way to identify leaders and clusters within the European network and determine if or how they are collaborating with organizations in North America is to conduct outreach in the form of additional surveys and targeted key informant interviews. We will begin with a follow-up survey intentionally targeting additional actors in the European network in Q1 2022.

5

Key Challenges and Opportunities

The diverse set of actors participating in the global crop analytics landscape is robust and creates unique challenges. While the community remains fragmented, opportunities exist for synchronizing a more cohesive crop analytics landscape. Below are our key takeaways that serve as a future roadmap to guide the network through the various challenges and opportunities that emerged from the research.

1. Establish a coherent community of practice and coordinating mechanism to encourage the adoption of shared data standards, sampling methods, and lessons learned: Several loose and overlapping communities of practice currently exist in the ecosystem. A series of alliances have formed around NASA Harvest, GEMS, Digital Earth Africa, and similar bodies, but few larger convening or coordination mechanisms exist to bridge these clusters or to support public/private partnerships at scale. This disjointed community means that for most organizations that are willing to collaborate and share data, the effort required to change disparate data collection, storage, or formatting methods does not outweigh the potential benefits of doing so.

   There is a significant interest and opportunity for the ECAAS initiative to bring key actors together to formalize the network and improve coordination through the use of a secretariat or similar function. Network stakeholders are willing to participate in such mechanisms and see the value in addressing certain issues (such as standards) through these means. Robust stakeholder buy-in and ownership must push adoption and establish protocols to keep them updated with the changing needs of the community, as well as disrupt current siloed efforts to bring the power of advanced crop analytics to smallholder farmers

   An organization should be selected, possibly on a rotating basis, to serve in a network coordination function in support of those efforts noted above. This function could provide monitoring services on behalf of network members to notify members when changes in the policies, standards, and protocols of some of the most influential service providers in the network (Planet Labs, Google, Amazon Web Services, ESRI, Airbus, etc.) occur, and what they mean for member organizations. The ECAAS team is concluding a case study and business plan analysis of similar data networks and will discuss recommendations for the selection of a coordinating body with the ECAAS Advisory Groups in the coming months. The team will bring examples of viable models with innovative public-private business models that reach smallholder farmers as options to help shift current organizational silos.

   Thus far, this research shows that most successful data networks leverage existing and well-connected organizations to serve as institutional hosts. Often an independent and network-specific governance mechanism is required, outside the existing governance of the coordinating entity. In many cases, revenue-generating activities can help cover operational costs of network coordination, but these can take time to reach maturity. A combination of philanthropic, in-kind, or other funding is often required to support such networks, especially in the early years.

2. Align data collection requirements and scale standards: There is disagreement on what types, quality, and quantity of data must be collected to begin with, and from what sources the most trusted data should be collected. Disagreements are driven in part by end users' specific needs and use cases, and in part by a lack of perceived incentive for organizations to adapt their data processes to an emerging standard or methodology. The disconnect between bottom-up stakeholders that work in farm-level data collection and top-down approaches from major donors, governments, or institutional players contributes to the divide. Standards make it easier to create, share, and integrate data across these and other groups by ensuring that data are represented and interpreted correctly. Standards also reduce the time spent cleaning and translating data, allowing more time for analysis.

   By making the value proposition of aligning on methodologies clear, ECAAS can help increase the adoption of existing scalable standards in the network. For example, the Radiant Earth Foundation has developed a "Guide for Collecting and Sharing Ground Reference Data for Machine Learning Applications," but it has not yet been adopted at scale. The World Bank's 50x2030 Initiative recently produced best practices for GTT collection surveys to support satellite-based crop type mapping in Sub-Saharan Africa. The community will need both bottom-up and systemic level coordination to scale standards. *A key opportunity for ECAAS is in communicating the gains from widespread adoption of such standards, such as by quantifying the reduction in later cleaning and processing costs required to use datasets created under this standard. This should be combined with a process to identify and promote tangible incentives so that data collectors make the extra effort to collect quality GTT.*

3. Enhance interoperability and data sharing pipelines: The inconsistency in data formats and standards noted above is a required step to improve data sharing, aggregation, and ease of use. The community also continues to request improved data pipelines to ease the automated or semi-automated data ingestion and sharing processes required to make data aggregation less manual and time-intensive. Testing and building back-end data pipelines are just as, if not more challenging to develop than front-end interfaces or products but are critical for interoperability. Innovation in defining the best data pipeline structures, reducing the burden upon individual, organizational research, and development teams

   Within the existing ECAAS network, most organizations reported being eager to share relevant data and collaborate for improved interoperability. The stakeholder network analysis (SNA) identified MercyCorps and CropIn as two key organizations with the ability to connect others and control the flow of information from one organization to another, but many others outside of our sample set are similarly positioned. *ECAAS or similar projects can continue to support these organizations and other innovations to define the best data pipeline structures.*

4. Generate more abundant validation and benchmarking models: Once collected, validation and benchmarking for GTT are inconsistent. The accuracy levels required for some use cases (like insurance) are significantly different from others, yet these needs are not always clearly articulated to develop practical algorithms and applications. One reason for this is that the proprietary nature of models and datasets serves as a disincentive for sharing. A common repository of benchmark datasets and models would allow for the comparison of models that could advance the understanding of ML families and applications and benefit the community. While direct model output comparison, if made broadly public, could disincentive participation by researchers and the private sector, identifying pre-competitive spaces and focusing on mutual lessons learned will be critical to encourage sharing.

   The ECAAS initiative is funding and can better communicate emerging research regarding the minimum viable accuracy based on the use case and can continue to promote data collection and analysis with specific end-use cases in mind. The nascent network structure can also test third-party verification and benchmarking services as a potential revenue model and value addition for participating member organizations.

5. Analyze Data Privacy and Ownership Models: Data privacy and ownership remain major hurdles, especially for multi-national initiatives. Both regulatory environments, and the ethics and norms which support them, change considerably across cultures and countries. As primary data sources, farmers have very different expectations of and legal protections for their information, whether personally identifiable or not. There is information asymmetry both within and between most governments on these issues. Some countries, such as Canada, have harmonized data sharing and use agreements at sub-national levels, but most have not yet reached that stage. While some organizations such as Digital Earth Africa are working hand in hand with government agencies, in many cases, private or NGO actors are subverting data sovereignty, intentionally or otherwise.

   These challenges are exacerbated in a smallholder farmer data environment, often typified by inconsistent or unequal access to mobile phones, unreliable or expensive data networks, and complicated political economy concerns around the use of agricultural data. In many systems, farmers work with intermediaries who collect and manage data, and farmers often lack an understanding of how their digitized data is used. The abundance of software and apps, many of which do not meet smallholder farmers' needs, further drives the divide between the farmer and their digital footprint, as their information may reside in multiple and disconnected places. Once data is collected, maintaining data privacy, aligning standards, and storing data in a common location all pose challenges to sharing GTT from the farm level.

   There remain ample opportunities to disseminate best practices and findings from data-sharing models from both the crop analytics landscape and other technical domains. In addition to building this into future network design, ECAAS will continue to amplify similar programs such as the Digital Fronters project, which seeks to test different farmer-centric data governance and ownership models and is jointly funded by USAID and the Gates Foundation.

6. Realign Inconsistent Incentive Structures: Throughout the data chain, incentives to collaborate and share data vary significantly depending on the types of organizations involved and how each perceives the value of GTT. Identifying and promoting aligned incentives across public, private, and civil society could lead to increased participation in data harmonization and a greatly expanded GTT lake. Some literature suggests that promoting pre-competitive spaces, supporting challenges to identify innovative business models, and leveraging the increasing push toward social responsibility can act as a combination of "carrots" to promote agricultural data sharing. From the opposite perspective, agricultural or public data policies, regulations, and funding requirements can strongly recommend or require that all actors operating in a given jurisdiction adopt congruent GTT standards. This type of standardization has occurred in zoning, permitting, and other spaces in many places, but has not yet permeated most agricultural data ecosystems.

Today, the crop analytics community is where the disaster risk insurance industry was before the OASIS project generated a commonly agreed upon pool of actuary data. There are opportunities for the crop analytics community to learn from OASIS and other initiatives that have faced similar data ecosystem challenges. ECAAS will continue to test the feasibility of these models in the agricultural GTT space and will discuss options with the initiative's Advisory Groups in early 2022. For example, a searchable tool that connects several existing GTT repositories could help provide value for actors across the data chain by significantly reducing the cost of GTT acquisition and processing. *ECAAS will continue to explore how to test and promote these data sharing models and data exchange marketplaces that support the development and scaling of technologies, processes, and partnerships that facilitate the deployment of advanced analytics in smallholder markets.*

References

[1] H. Alon, P. Norvig and F. Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems*, vol. 24, no. 2, 2019.

[2] S. Lowder, J. Skoet, T. Raney and S. Lowder, "The number, size, and distribution of farms, smallholder farms, and family farms worldwide.," *World Development*, vol. 87, pp. 16-29, 2016.

[3] B. Graeub, M. Jahi Chappell, H. Wittm and e. al., "The State of Family Farms in the World," *World Development*, vol. 87, pp. 1-15, 2016.

[4] L. Samberg, J. Gerber, N. Ramankutty, M. Herrero and P. West, "Subnational distribution of average farm size and smallholder contributions to global food production," *Environmental Research Letters*, vol. 11, no. 12, 2016.

[5] M. Hegarty-Crave, J. Polly, M. O'Neil, R. N. Ujeneza N and e. al., "Remote Crop Mapping at Scale: Using Satellite Imagery and UAV-Acquired Data as Ground Truth," *Remote Sensing*, vol. 12, no. 12, 2020.

[6] G. Azzari, S. Jain, G. Jefferies, T. Kilic and S. Murray, "Understanding the Requirements for Surveys to Support Satellite-Based Crop Type Mapping: Evidence from Sub-Saharan Africa," *Remote Sensing*, vol. 13, no. 23, 2021.

[7] M. Marshall, S. Crommelinck, D. Kohli, C. Perfer, M. Yang, A. Ghosh, S. Fritz, K. Bie and A. Nelson, "Crowd-Driven and Automated Mapping of Field Boundaries in Highly Fragmented Agricultural Landscapes of Ethiopia with Very High Spatial Resolution Imagery," *Remote Sensing*, vol. 11, no. 18, 2019.

[8] A. Challinor, J. Watson, D. Lobell, S. Howden, D. Smith and N. Chhetri, "A Meta-Analysis of Crop Yield under Climate Change and Adaptation," *Nature Climate Change*, vol. 4, pp. 287-291, 2014.

[9] M. Brown, "Famine early warning systems and remote sensing data," *Springer Science & Business Media*, 2008.

[10] B. Ortiz-Crespo, J. Steinke, C. Quiros, J. Van de Gevel, H. Daudi, M. Gaspar Mgimiloko and J. van Etten, "User-centred design of a digital advisory service: enhancing public agricultural extension for sustainable intensification in Tanzania," *International Journal of Agricultural Sustainability*, 2020.

[11] H. Baumuller, "The little we know: An exploratory literature review on the utility of mobile phone-enabled services for smallholder farmers," *Journal of International Development*, vol. 30, no. 1, 2018.

[12] R. Southwood and K. Wong, "Building a Data Ecosystem for Food Security and Sustainability in AgTech V3.0.," *CGIAR*, 2020.

[13] A. Kamilaris, A. Kartakoullis and F. Prenafeta-Boldu, "A review on the practice of big data analysis in agriculture," *Computers and Electronics in Agriculture*, vol. 143, 2017.

[14] N. Nagendra, G. Narayanamurthy and R. Moser, "Satellite big data analytics for ethical decision making in farmer's insurance claim settlement: minimization of type-1 and type-2 errors," *Annals of Operations Research*, 2020.

[15] D. Leeuw, A. Vrieling, A. Shee, C. Atzberger, K. Hadgu, C. Biradar, H. Keah and C. Turbey, "The Potential and Uptake of Remote Sensing in Insurance: A Review.," *Remote Sensing*, vol. 6, no. 11, 2014.

[16] C. Carletto, D. Jolliffe and R. Banerjee, From Tragedy to Renaissance: Improving Agricultural Data for Better Policies, vol. 52:2, *The Journal of Development Studies*, 2015.

[17] I. Dickle and S. Hull, "MEDIN Cost Benefit Analysis," Economics for the Environment, London, 2019.

[18] "2019 State of Climate Services: Agriculture and Food Security," World Meteorological Organization (WMO), 2019.

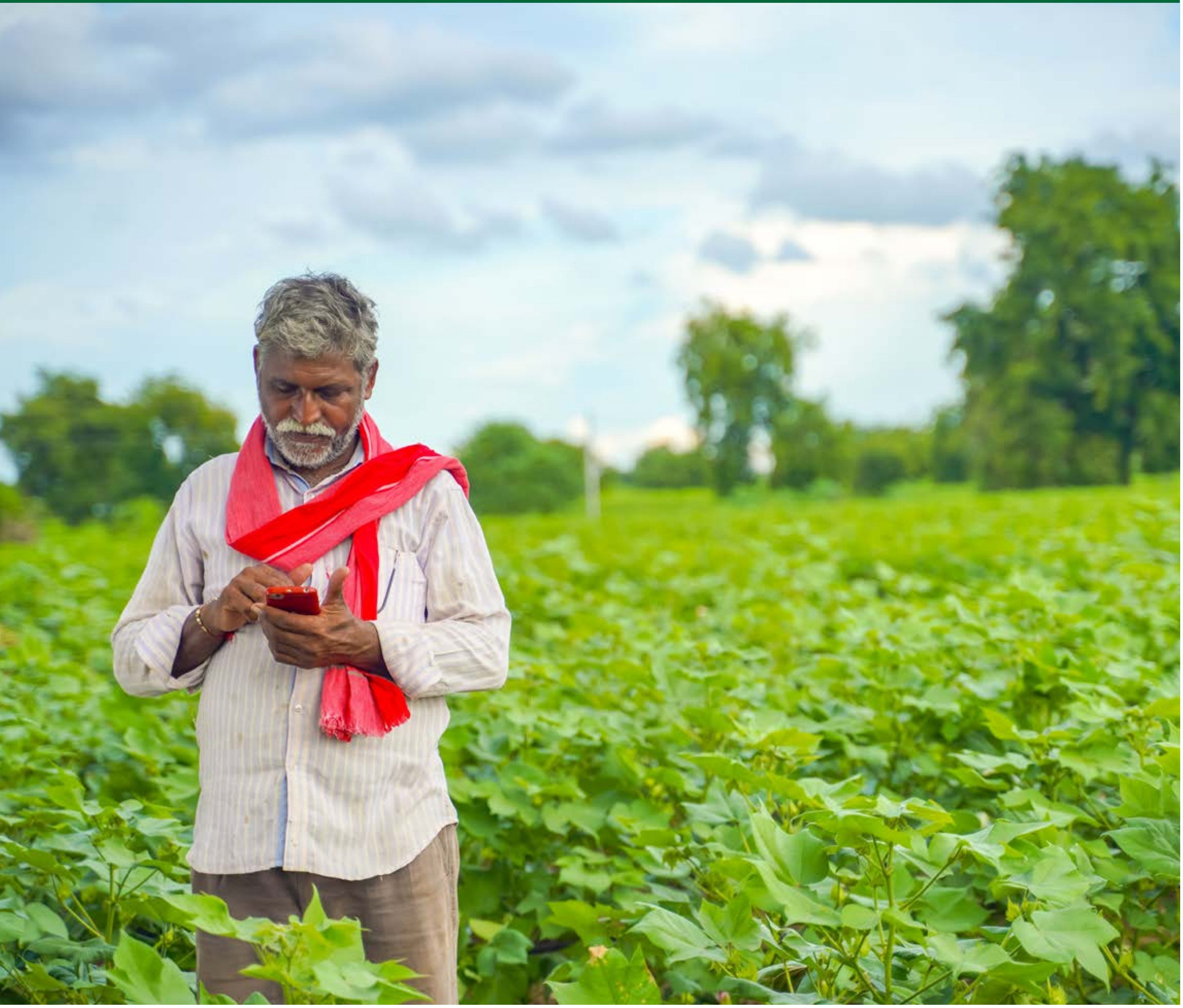
[19] A. Begue, L. Leroux, M. Soumare, J.-F. Faure, A. A. Diouf, X. Augusseau, L. Toure and J.-P. Tonneau, "Remote Sensing Products and Services in Support of Agricultural Public Policies in Africa: Overview and Challenges," *Frontiers in Sustainable Food Systems*, vol. 14, 2020.

[20] J. Duncan, J. Dash and P. Atkinson, "The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes," *Frontiers in Environmental Sciences*, vol. 11, 2015.

[21] B. Mann and R. Stanimirova, "Seeing More with Satellites: Improving Insurance for African Farmers," *International Research Institute for Climate and Society*, 2014.

1

Annex: Summary of Social Network Analysis Survey Results



Tetra Tech launched an online survey in March of 2021 targeting organizations identified during ECAAS implementation to 1) obtain more consistent information about organizations of interest and 2) systematically uncover existing relationships between these organizations. As noted in the methods section, the online survey was meant to supplement ongoing literature reviews and interviews feeding the SLA process. The survey questions focused on fundamental attributes that relate to crop analytics and the broader crop analytic network. The public-facing information elements collected from the survey are summarized in the ECAAS Stakeholder Database, available on the initiative's website.

Survey Questions

- In what regions does your organization operate? (This includes physical presence, remote operations, and data collection)
- Please indicate the top 5 countries where your organization has the strongest presence.
- Which of the following best describes your organization's interaction with farmers?
  - Direct interaction with farmers (on-site or remotely via application or other touchpoints)
  - Indirect interaction through an intermediary (such as supplier, bank, or extension agent)
  - We work at the national, regional, or policy level
  - We work on back-end technology or infrastructure
- Which of the following crop analytics end-use cases are priorities for your organization?
  - Extension and production management
  - On-farm service provision
  - Output marketing or trading Input supply optimization
  - Public sector land management
  - Improved farmer access to finance
  - Food security monitoring and response
  - Other
- What roles does your organization serve in the crop analytics data chain?
  - None of the above
  - Data Collection. Active capture of ground or remotely sensed data.
  - Data Ingestion & Processing.
  - Data aggregation, fusion, quality control, and/or labeling.
  - Data Hosting. Local or cloud-based data storage.
  - Analytics Provider. Creating analytic products or models either for their purposes or on behalf of another organization.
  - Analytics User/End User. Using analytics or derived information products (e.g., extension & advisory services, financial services) either directly or to support end-users (farmer, government agency, etc.)
  - Other (please specify)

- Do you currently share your crop analytics data or models publicly?
- What Data Points Parameters is your organization currently collecting? (Yield, Field Boundary, Crop Type, Area Planted by Crop, Crop Status (Health/Growth), Pest Presence, Disease Presence, Soil Health, Climate Data, Other)
- What technologies does your organization currently use to capture data? (Soil Sensor, Weather Sensor, UAV, Plane, Satellite Imagery, Mobile Phone, Machine Mounted Sensor, Handheld Sensor, Other)

Relationship Information

The ECAAS team tried to obtain information about existing partnerships, collaborations, and relationships between organizations via publicly available information and one-on-one discussions. Despite using a semi-structured set of interview questions, information was inconsistent and could have been biased by what was top-of-mind for a given interviewee, or who within an organization was available and willing to be interviewed. Additionally, most organizations were not willing or could not detail many of their collaborations due to competitive concerns or non-disclosure agreements. To overcome this, the ECAAS team included in the survey certain sets of relational information which would not be made public, and analyzed this more sensitive data through a Social Network Analysis (SNA).

We asked respondents to indicate their relationships to 120 organizations included in the network roster (25 civil society, 21 public sector, and 74 private sector). Respondents were specifically asked to "Please indicate the degree to which you are collaborating with the following Private Sector organizations within the ECAAS Network":

- Unaware of this organization
- I'm aware of the organization, but we have never interacted
- Some informal contact (e.g., conferences, workshops, etc.)
- Some 1:1 discussion or interaction
- Our organizations have a formal agreement

The survey also gave respondents space to write in the names of up to 10 organizations not included in the roster to indicate any other relationships. This information was used to expand the roster and identify any recurring key organizations that we had otherwise missed.

Overview of Key SNA Analysis

We received 44 survey responses, of which 32 were unique and complete. Data were uploaded into Kumu.io software for initial analysis and visualization. The team identified the top 50 ranked organizations based on the unweighted degree (total number of connections). Of these 50 organizations, 27 had not responded to the survey directly. For those who did not respond, the team added publicly available information into the network analysis.

Key Findings of SNA

- The overall stakeholder network consists of 193 organizations (120 identified for the survey plus an additional 73 that were identified via survey responses). Screen-shots from an interactive version of this network map can be accessed via this Google Docs Link, and a visual summary overview is provided below (**Figure 16**).

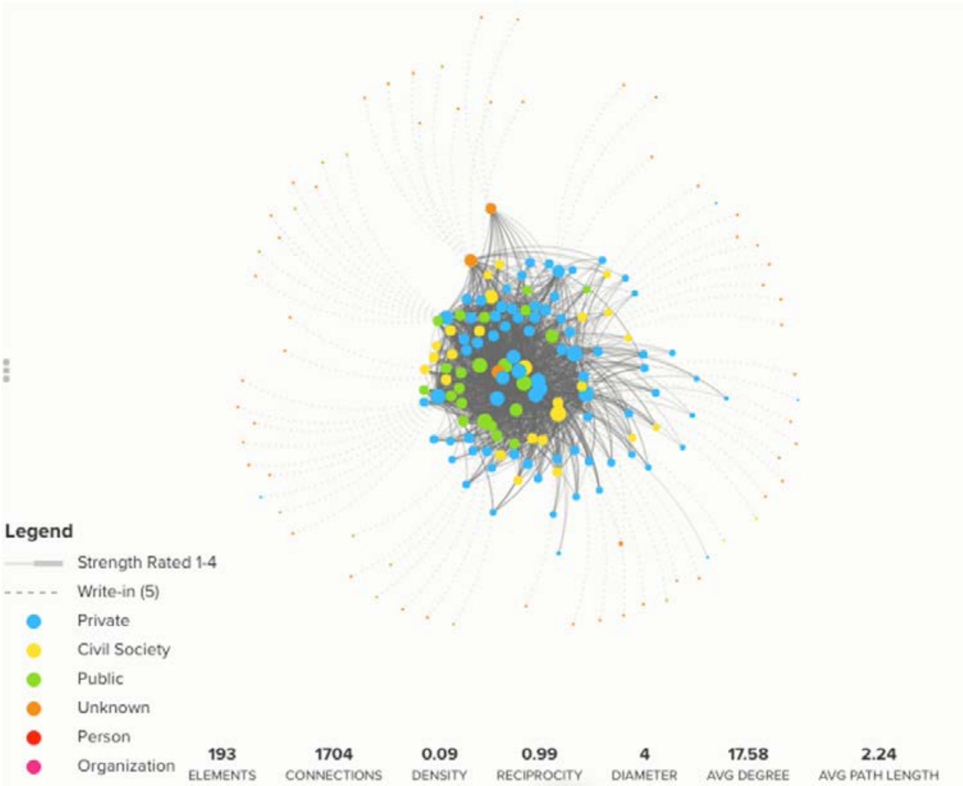


Figure 16:
Visual Representation of Network Map

- Among these organizations, there are 1,704 unique connections. The most frequent types of connections are either some level of 1:1 interaction (n=485), or simple awareness of another organization but no direct relationship (n=569) (**Figure 17**). The range of connections helps to demonstrate that the network of respondents is disjointed and lacks the cohesion of a well-organized network.
- The analysis did not detect distinct geographical hubs within the network. However, we suspect that this may change as we receive more respondents from outside of North America.

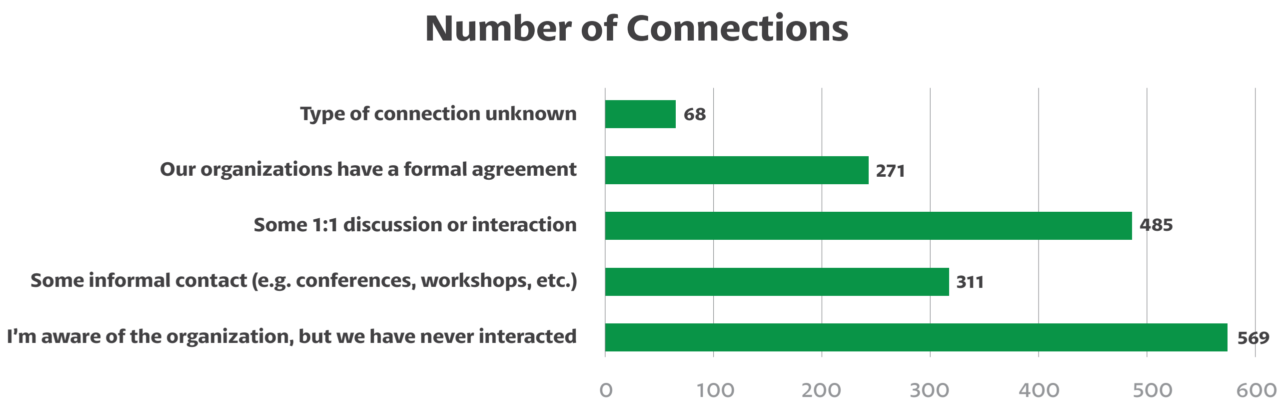


Figure 17:
Number and type of unique connections with other organizations

Indegree

- Indegree measures the number of incoming connections for an element in the network. In general, elements with high indegree are the leaders, looked to by others as a source of advice, expertise, or information (*Definition from kumu.io*). The top ten organizations that have the highest indegree based on responses received are represented in Table 8 below. Most of these organizations responded to the survey.

Table 8:
Top Ten Organizations Identified by Indegree

Rank	Organization	Value
#1	Mercy Corps	134
#2	Tetra Tech	131
#3	CropIn	128
#4	NASA Harvest	108
#5	Rabobank	106
#6	aWhere	100
#7	Digital Green	100
#8	McKinsey & Company	100
#9	CGIAR or Research Centers (IFPRI, ICRISAT, etc.)	96
#10	USAID (Feed the Future, FEWS NET, etc.)	94

Reach (two-step out)

- Reach measures the portion of the network within two steps of any other organization. In general, organizations with high reach can spread information through the network through close "friend-of-a-friend" contacts. A reach value of 1 indicates that an organization reaches 100% of the network through this level of relationship. Furthermore, a value of 1 indicates that 100% of survey respondents reported that they are at a minimum aware of the organization listed. (*Definition from kumu.io*). The top ten organizations that have the highest reach based on responses received are represented in **Table 9**.

Table 9:

Top Ten Organizations Identified by Reach.

Rank	Organization	Value
#1	Airbus	1
#2	Amazon Web Services	1
#3	ESRI	1
#4	Google	1
#5	Planet Labs	1
#6	FAO	0.995
#7	Microsoft	0.990
#8	World Bank	0.990
#9	IBM	0.990
#10	McKinsey & Company	0.990

- While most of these organizations did not complete the survey, it is worth noting that each can influence the network, primarily through changes in their platforms, standards, and policies.

Next Steps

The ECAAS team will continue to collect information about organizations through an open request and stakeholder dashboard located on the initiative's website. We recommend that in future phases of ECAAS, this survey be repeated on an annual basis to identify how the network has evolved and has grown year over year.

2

Annex: SLA Database of Top Influencers

The database below presents the information captured for the top influencers as identified through the SNA survey.

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
6th Grain	Private	Northern Africa | Southern Africa | South America | North America | Eastern & Southeastern Asia | Central & Southern Asia | Western Asia & Middle East | Eastern Europe | Western Europe	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider		Yield | Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Pest Presence | Disease Presence	Satellite Imagery | Mobile Phone	
Agriculture and Agri-Food Canada	Public	North America	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/ End User	No	Yield | Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Pest Presence | Disease Presence | Soil Health | Climate Data	Soil Sensor | Weather Sensor | UAV | Plane | Satellite Imagery | Mobile Phone | Machine Mounted Sensor | Handheld Sensor	Canada
Agritask	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Western Africa | Eastern Africa | Western Europe | Central America | South America	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider |	No	Yield | Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Pest Presence | Disease Presence | Soil Health | Climate Data | Other	Soil Sensor | Weather Sensor | Satellite Imagery | Mobile Phone | Machine Mounted Sensor | Handheld Sensor | Other	Brazil | Israel | Mexico | Peru | Chile
Airbus	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection	No	N/A	Satellite Imagery	
Amazon Web Services	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Hosting	N/A	N/A	N/A	

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
AtlasAI	Private	Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa	Data Ingestion & Processing | Analytics Provider | Analytics User/ End User	Yes	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth)	Satellite Imagery	
aWhere	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Central America | South America | North America | Oceania	Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/ End User	Yes	Climate Data	Weather Sensor | Satellite Imagery	Kenya | Uganda | Honduras | Zimbabwe | United States
CGIAR or Research Centers (IFPRI, ICRISAT, etc.)	Public	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/ End User	Yes	Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/ Growth) | Pest Presence | Disease Presence | Soil Health | Climate Data	Soil Sensor | Weather Sensor | UAV | Plane | Satellite Imagery | Mobile Phone | Machine Mounted Sensor | Handheld Sensor	Kenya | India | Rwanda | Ethiopia | France
Clark University	Public	Western Africa | Eastern Africa | Southern Africa | North America	Other	No	Field Boundary | Crop Type | Other	Weather Sensor | Satellite Imagery	United States | Zambia | Ghana | South Africa | Kenya
CropIn	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Southern Africa | Eastern Europe | Western Europe | Central America | South America | North America	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/ End User	Yes	Yield | Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Pest Presence | Disease Presence | Climate Data | Other	Soil Sensor | Weather Sensor | UAV | Satellite Imagery | Mobile Phone | Other	India | Kenya | Nigeria | Mayanmar | Phillipines

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
Digital Earth Africa	Civil Society	Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider	Yes	Other	Satellite Imagery	Kenya | Ghana | Tanzania | Senegal | Nigeria
Digital Globe	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection	Yes	N/A	Satellite Imagery	
Digital Green	Private	Central & Southern Asia | Western Africa | Eastern Africa | North America	Data Collection | Analytics User/End User | Other	No	Yield | Field Boundary | Crop Type | Area Planted by Crop | Soil Health | Other	Mobile Phone	India | Ethiopia | Nepal | Nigeria | Kenya
ESRI	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Ingestion & Processing | Analytics Provider	Yes	N/A	Satellite Imagery	
Facebook	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection | Analytics User/End User	N/A	N/A	Mobile Phone	

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
Google Cloud	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Ingestion & Processing | Data Hosting | Analytics Provider	N/A	N/A	Satellite Imagery	
Grameen Foundation	Civil Society	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection | Analytics User/End User	No	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery | Mobile Phone	
Group on Earth Observations Global Agricultural Monitoring Initiative (GEOGLAM)	Civil Society	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection | Data Ingestion	Yes	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary	Satellite Imagery | Soil Sensor | UAV | Mobile Phone	
Intel	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Ingestion & Processing | Data Hosting | Analytics Provider	N/A	N/A	N/A	

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
KJ Somaiya Institute of Applied Agricultural Research (KIAAR)	Public	Central & Southern Asia	Data Collection | Data Ingestion & Processing | Analytics User/End User	No	Yield | Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Soil Health | Climate Data	Weather Sensor | Satellite Imagery | Mobile Phone | Handheld Sensor	India | Russia | United States
McKinsey & Company	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Analytics Provider | Analytics User/End User	No	Field Boundary | Climate Data	Satellite Imagery	United States | Brazil | India | Canada | EU
Mercy Corps	Civil Society	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Caribbean | Central America | South America	Other	Yes	Yield | Field Boundary | Crop Type | Area Planted by Crop | Pest Presence | Soil Health | Climate Data	Soil Sensor | Weather Sensor | UAV | Satellite Imagery | Mobile Phone | Handheld Sensor |	Kenya | Nigeria | Ethiopia | Indonesia | Tanzania
Mesur.io	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/End User | Other	Yes	Crop Type | Crop Status (Health/Growth) | Pest Presence | Disease Presence | Soil Health | Climate Data | Other	Soil Sensor | Weather Sensor | UAV | Plane | Satellite Imagery | Mobile Phone | Machine Mounted Sensor | Handheld Sensor	United States | Canada | Japan | United Kingdom | Netherlands
Microsoft Azure	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Central America | South America | North America | Oceania	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider	N/A	N/A	N/A	

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
NASA Harvest	Civil Society	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Central America | South America | North America | Oceania	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider	Yes	Crop Type	Soil Sensor | Satellite Imagery | Mobile Phone |	
NIAS Centre for Spatial Analytics and Advanced GIS (C-SAG)	Un-known	Central & Southern Asia	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/ End User | Other	No	Yield | Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Pest Presence | Disease Presence | Soil Health | Climate Data | Other	Soil Sensor | UAV | Satellite Imagery | Mobile Phone | Other	India
Pennsylvania State University	Public	North America	Data Collection | Data Ingestion & Processing | Analytics Provider | Analytics User/End User	Yes	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery | Mobile Phone	
Rabobank	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Africa | Eastern Africa | Western Europe	Data Collection | Analytics Provider | Analytics User/End User	Yes	Yield | Field Boundary | Crop Type | Area Planted by Crop	Soil Sensor | Satellite Imagery | Mobile Phone	Kenya | India | Uganda | Peru | Indonesia
SatSure	Private	Central & Southern Asia | Eastern & Southeastern Asia | Eastern Africa	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/ End User	Yes	Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth)	Soil Sensor | Satellite Imagery | Mobile Phone	India | Phillipines | Myanmar
Tata Consultancy Services	Private	Central & Southern Asia	Data Collection | Analytics Provider | Analytics User/End User	No	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery | Mobile Phone	

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
Tetra Tech	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Eastern Europe | Central America | South America | North America	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider | Analytics User/ End User	No	Yield | Crop Type | Area Planted by Crop | Climate Data	Mobile Phone | Other	United States | Colombia | Afghanistan | Indonesia | India
Trimble	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection | Analytics Provider | Analytics User/End User	No	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery | Machinemounted Sensor | Weather Station | Handheld Sensor | Mobile Phone |	
University of California Berkeley	Public	North America	Data Collection | Analytics Provider	Yes	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery	
University of California Davis	Public	North America	Data Collection | Analytics Provider	Yes	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery	
University of California Santa Barbara	Public	Eastern Africa	Data Collection | Analytics User/End User	Yes	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery	Kenya | Zambia

Label	Type	Regions	Data Chain	Share Data Publicly?	Parameters	Technologies	Top 5 Countries
University of Colorado Boulder	Public	North America	Data Collection | Analytics Provider	Un-known	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery	
University of Minnesota	Public	North America	Data Collection | Analytics Provider	Yes	Yield | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Field Boundary | Pest Presence | Disease Presence	Satellite Imagery	
USAID (Feed the Future, FEWS NET, etc.)	Public	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Caribbean | Central America | South America | Oceania	Data Collection | Data Ingestion & Processing | Analytics User/End User | Other	Yes	Yield | Field Boundary | Crop Type | Soil Health | Climate Data	Soil Sensor | UAV | | Satellite Imagery | Mobile Phone | Handheld Sensor | Other	
Vito	Private	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Central America | South America | North America	Data Collection | Data Ingestion & Processing | Data Hosting | Analytics Provider	Yes	Yield | Field Boundary | Crop Type | Crop Status (Health/Growth) | Disease Presence	Soil Sensor | UAV | Plane | Satellite Imagery | Mobile Phone | Machine Mounted Sensor | Handheld Sensor	Belgium | China | Qatar | India | Kenya
World Bank	Public	Central & Southern Asia | Eastern & Southeastern Asia | Western Asia & Middle East | Northern Africa | Western Africa | Eastern Africa | Central Africa | Southern Africa | Eastern Europe | Western Europe | Caribbean | Central America | South America | North America | Oceania	Data Collection | Data Ingestion & Processing | Analytics Provider | Analytics User/End User		Yield | Field Boundary | Crop Type | Area Planted by Crop | Crop Status (Health/Growth) | Pest Presence | Disease Presence | Soil Health | Climate Data	Soil Sensor | Satellite Imagery | Mobile Phone | Handheld Sensor | Other	

Enabling Crop
Analytics At Scale

Stakeholder Landscape Assessment

The Enabling Satellite-based Crop Analytics at Scale (ECAAS) Initiative is a multi-phase project that aims to catalyze the development, availability, and uptake of agricultural remote-sensing data and subsequent applications in smallholder farming systems. The initiative is funded by The Bill & Melinda Gates Foundation and implemented by Tetra Tech.

info.ecaas@tetratech.com
cropanalytics.net

Final report

February 2022