FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

PUTTING DATA AT THE SERVICE OF AGRICULTURE

A Case Study of CIAT

USAID
FROM THE AMERICAN PEOPLE

ABBREVIATIONS

Asohofrucol	Asociación Hortofrutícola de Colombia [Horticulture Association of Colombia]
CCAFS	CGIAR's Research Program on Climate Change, Agriculture and Food Security
CGAIR	Consultative Group on International Agricultural Research.
CIAT	International Center for Tropical Agriculture
CIMMYT	The International Maize and Wheat Improvement Center
FAIR Data principles	Findable, Accessible, Interoperable, Reusable
FLAR	Latin American Fund for Irrigated Rice
ICA	Colombian Agricultural Institute
IDEAM	National Institute of Hydrology, Meteorology and Environment Studies
INTA	National Agricultural Technology Institute
LTACs	Local Technical Agro-climatic Committees
MADR	the Ministry of Agriculture and Rural Development of Colombia
MasAgro	The Sustainable Modernization of Traditional Agriculture program
SAGARPA	Secretary of Agriculture, Livestock, Rural Development, Fisheries and Food
SIRIA	Sistema de Recolección de Información Agrícola [Collection system for Agricultural Information]
UNFCCC	United Nations Framework Convention on Climate Change
3G	a mobile communications standard that allows mobile phones, computers, and other portable electronic devices to access the Internet wirelessly
4G	a mobile communications standard intended to replace 3G, allowing wireless internet access at a much higher speed

Author: Cristina Manfre and Wesley Laytham; July, 2018

This case study describes the data-driven agronomy work led by the International Center for Tropical Agriculture (CIAT), a CGIAR center headquartered in Cali, Colombia. It is part of a series highlighting the integration of digital technologies into agricultural programs. Each case study examines different approaches to adoption and how digital tools impact organizational culture, operations, and programs.

EXECUTIVE SUMMARY

In 2008, CIAT began exploring a new approach to use big data tools to analyze information and help farmers make better on-farm decisions that lead to improved agricultural outcomes. This approach, referred to as data-driven agronomy, enhances agronomy through increased use of observational information, data mining, and contextual information to provide farmers with tailored recommendations about what to plant, when, and how to manage crops. CIAT has been using the approach over the last ten years with a range of partners across Latin America. This case study focused on CIAT's experiences in Colombia working with the rice, fruit, and maize growers' associations – Fedearroz, Asohofrucol, and Fenalce, respectively – and the Ministry of Agriculture and Rural Development in the region of Pereira.

CIAT's data-driven agronomy work has three distinct phases: data collection, data analysis, and knowledge management.

DATA COLLECTION:

Data-driven agronomy requires multiple types of data which are not always publicly available or in a format useful for analysis. Datasets are aggregated from existing sources including open datasets, experimental data, and partners' historical data, and are complemented by new data. This section describes CIAT's experiences with gaining access to data from its government and private partners. It also highlights the challenges associated with aggregating datasets in terms of data management, entry, and cleaning. Capacity building in terms of data collection was important during this phase to ensure data quality and sustain the analytical tools in Colombia after the projects with CIAT end.

DATA ANALYSIS:

Data-driven agronomy complements traditional controlled experiments by analyzing observational data that captures the soil and climate diversity among farmers' fields. Through various data mining and machine learning techniques, CIAT can examine multiple types of data to reveal patterns across crop varieties, climatic conditions, and location to identify optimal site-specific management practices and cultivars. These recommendations are reviewed and validated by an interdisciplinary team of data scientists, agronomists, and anthropologists.

KNOWLEDGE MANAGEMENT:

Data-driven agronomy leads to improved outcomes when model outputs are converted into relevant, understandable, and actionable information that is disseminated through accessible and trustworthy channels. While CIAT is well-positioned to assume the lead research role, the growers' associations and government partners are better positioned to engage farmers. CIAT coached partner staff on interpreting information to ensure that recommendations were

made with an understanding of the accuracy and variation of the models. In some projects, they built online tools to provide farmers and technicians with data about their specific farms. More recently, under CGIAR, CIAT established local committees through which farmers, alongside representatives from the private sector, research organizations, and technicians interpret climate, weather, and observational data to identify specific adaptation measures for adoption.

Digital agriculture advances agricultural outcomes by driving greater financial inclusion, more precise agriculture, better data collection and analytics, and more effective information dissemination. Data-driven agronomy leads to impacts that contribute to these outcomes in three ways:

1. **Using information to improve crop management decisions.** Farmers receive better information for evidence-based decisions, leading to more precise and more productive agriculture. In one case, CIAT and Fedearroz predicted a dry spell and recommended farmers plant later than usual. This led to farmers avoiding crop losses valued at over **USD 3.6 million**.

2. **Improving the services and credibility of farmer-serving organizations.** Through the collaboration with CIAT, these organizations improve their data collection and analysis skills. This enables them to improve the quality and accuracy of agronomic recommendations provided to their farmer members.

3. **Expanding the analytical avenues for agricultural research.** Data-driven agronomy leverages observational data from farmers and data mining techniques to complement conventional agronomic research and achieve improved yields. CIAT leads this work within the CGIAR and has successfully scaled the approach through partnerships with the Latin American Fund for Irrigated Rice, the Government of Honduras, and the National Agricultural Technology Institute in Argentina.

The case study concludes by highlighting valuable lessons for organizations wishing to pursue similar digital integration.

PUTTING DATA AT THE SERVICE OF AGRICULTURE

Series Overview: This case study is part of a series highlighting the integration of digital technologies into agricultural programs. Over the past ten years, and particularly over the past five, the use of mobile phones and Internet-based, digital tools in farming activities has skyrocketed. This is largely due to the widespread adoption of mobile phones in developing and emerging markets, coupled with the increased spread of 3G and 4G connectivity. What has emerged is a broad set of digitally-based applications that have driven greater financial inclusion, more precision agriculture, better data collection and analytics and more effective information dissemination. Agricultural organizations and programs are increasingly embracing these tools to advance their goals. Each case study in this series looks at different approaches to adoption and how the tools are impacting organizational culture, operations, and programming.

OVERVIEW

In 2008, the International Center for Tropical Agriculture (CIAT) began applying innovative data mining[1] techniques to analyze existing data on crop yields, farming practices, climate, and soil and provided farmers with tailored recommendations about what to plant, when, and how to manage crops. Using sophisticated data analysis techniques, CIAT is forging new avenues for agricultural research and development. While conventional crop breeding and agricultural research efforts often draw on research conducted in labs and on research stations, this complementary approach uses data about and from farmers and their fields to refine the understanding of productivity constraints associated with specific soil and climate conditions, as well as management practices. Site-specific agriculture[2] (Box 1) harnesses the power of big data[3] to identify timely, relevant, and locally-specific recommendations, and in this way, is "putting data at the service of agriculture."[4]

This is rapidly changing the culture of farming in communities where CIAT and its partners are working; one in which the value of data and information has increased exponentially. Farmers once relied on wisdom from family members and neighbors and well-researched but generalized recommendations from extension officers. Now, they are regularly analyzing yield and climate data to make evidence-based management decisions. They are using more accurate, complete, and reliable information than before, even though this information is delivered through many of the same channels. Extension officers are providing tailored recommendations about what crops to grow and how, based on the specific conditions of a farmer's fields. National growers' associations are no longer filing away seasonal crop data they collect about their farmers, but instead they are opening it up for analysis by researchers and data scientists. Although there is still significant room to apply even more advanced data mining to predict agricultural outcomes, CIAT already

1 Data mining is the practice of automatically searching large stores of data to discover patterns and trends. Data mining uses sophisticated mathematical algorithms to examine data and evaluate the probability of future events. Data mining can answer questions that cannot be addressed through simple analysis techniques (Surampudi 2017).

2 CIAT's site-specific agriculture work has evolved over the years and the team now uses the term "data-driven agronomy" to describe the current work which includes "a set of complementary approaches that enhance traditional agronomy…[that includes] three main principles: (i) increased use of observational information (ii) data mining and (iii) contextualized information." (http://bigdata.cgiar.org/communities-of-practice/data-driven-agronomy/)

3 While the definition of Big Data varies between disciplines, it is generally characterized as datasets having high volume, velocity, variety, and variability. The CGIAR Platform for Big Data in Agriculture defines it as "harmonized, interoperable, and contextually integrated datasets and publications from multiple disciplines relevant for CGIAR's research and development goals," (CIAT and IFPRI 2016).

4 CIAT 2015.

Farmers and Fedearroz technician from a local technical agroclimatic committee in San Marcos discuss climate forecasts to identify production recommendations.

received awards from the UN Global Pulse Big Data Climate Challenge (2014), the World Bank Innovation Challenge (2015)[5] , and the UN Climate Change Momentum for Change (2017), in all cases for its use of Big Data techniques.

This case study provides an overview of how CIAT is using data analytics to help farmers make better on-farm decisions. It begins with an overview of the digital landscape in Colombia. Then it describes the CIAT data-driven agronomy team's overall approach to digital integration and specifically its experience using data-driven agronomy with partners. It reviews the impact of the digital techniques on farmers, CIAT's partners and stakeholders in targeted value chains, and offers lessons learned about its digital integration experience drawn from the reflections of staff, partners, and other stakeholders.

BOX 1

SITE-SPECIFIC AGRICULTURE

CIAT defines **site-specific agriculture** as examining "variation between sites or management units. A management unit is a set of plots of fields with relatively homogenous environmental conditions and reasonably uniform agricultural management practices." This is different from *precision agriculture*, which examines within plot or field variation.

Source: Jiménez et al. 2016

5 This award is for collaboration with Erick Fernandes of the World Bank titled Big Data for Climate Smart Agriculture - Enhancing & Sustaining Rice Systems for Latin America and the World.

THE DIGITAL LANDSCAPE

As a CGIAR center, CIAT works in 53 countries in Latin America, Africa, and Asia. Its research focuses on agriculture in the tropics, with a strong emphasis on climate and environmental degradation. Although its work is global, this case study focuses on the data-driven agronomy work being implemented in Colombia, where CIAT is headquartered. The work was first piloted in Colombia and over the last ten years. CIAT has iterated on its original data mining techniques and developed strong relationships with various public and private sector Colombian stakeholders. The work benefits from a favorable local digital landscape, which is marked by the country's commitment to open data.

The 2015 Global Open Data Index ranked Colombia fourth among 122 countries.[6] Major open datasets focus on national statistics, elections, urban life, and anti-corruption efforts like those found in the Datos Abiertos online portal for the Colombian Government.[7] In recent years the government of Colombia has begun providing open climate data through the National Institute of Hydrology, Meteorology and Environment Studies (IDEAM). However, datasets about economic development, agriculture, and climate are not as open. Many of these are held by private entities like national crop growers' associations. These associations are legal entities that represent farmers in the private and semi-public sector and possess important (and proprietary) historical data about their farmers and crop events. Some also have their own weather stations and climate datasets. Despite these obstacles, the culture of open data, led by the Government of Colombia, provides a foundation for working with various stakeholders to open the data repositories.

BOX 2
THE CGIAR

CIAT defines site-specific agriculture as examining "variation between sites or management units. A management unit is a set of plots of fields with relatively homogenous environmental conditions and reasonably uniform agricultural management practices." This is different from *precision agriculture*, which examines within plot or field variation.

Source: www.cgiar.org

APPROACH TO DIGITAL INTEGRATION

Precision agriculture is increasingly used in countries with a high degree of on-farm mechanization. In Europe and North America, a combination of sensors, drones, and satellites are used to monitor soil, climate, and productivity conditions on fields to improve yields and mitigate diseases and pests. This kind of digital integration in the agriculture sector is still nascent or does not exist in many developing countries, particularly those in the tropics. The need, however, to harness these digital tools to increase productivity and mitigate losses is great and CIAT has seen significant potential for exploring how the principles of precision agriculture could be adapted to different contexts using data mining techniques. As will become clear, the digital integration approach pursued by CIAT is about more than just data analysis: the CIAT data-driven agronomy team is interested in ushering in a data ecosystem in which there is two-way movement of information between farmers and researchers.

6 In the most recent list for 2016, the Global Open Data Index ranked Colombia 14 among 94 countries. The ranking however is not comparable over time as the methodology used changed significantly between 2015 and 2016. See https://index.okfn.org/ for more information.
7 Datos Abiertos is the online portal that houses all the open data of the Government of Colombia.

BOX 3
NATIONAL GROWERS' ASSOCIATIONS AND OTHER CIAT PARTNERS IN COLOMBIA

Asociación Hortifructícola de Colombia (Asohofrucol), the Fruit and Horticulture Growers Association of Colombia

Federación Nacional de Arroceros (Fedearroz), National Federation of Rice Growers

Federación Nacional de Cultivadores de Cereales y Leguminosas (Fenalce), National Federation of Cereal and Legume Growers

CIAT defines data-driven agronomy as a set of complementary approaches that enhance traditional agronomy through the use of increased observational information, data mining, and contextualized information.[8] These approaches allow research teams to explain historical events or predict relationships to provide partners, extension service providers, researchers, and farmers with new information and best practices about the relationships between different crop, environmental, and socioeconomic conditions. Simply put, these approaches allow researchers and agronomists to analyze different types of data from different sources to make tailored, site-specific recommendations to help farmers know what to plant, if they should plant, and when to plant.

Over the course of the ten years that CIAT has been implementing data-driven agronomy, the work has evolved and engaged different partners (Box 3). In its initial stages, the work focused heavily on defining and understanding the suitability of management units for different crops. This process, known as site-specific agriculture, was used by partners like the Asociación Hortifructícola de Colombia (Asohofrucol), and the Ministry of Agriculture and Rural Development (MADR) for the Municipality of Pereira. Currently the work is being used to identify productivity limiting factors to inform adaptation measures for different climate conditions, such as El Nino. Growers' associations, like Fedearroz and Fenalce, are partnering with CIAT for this work.

While the data mining techniques form a critical part of CIAT's data-driven agronomy approach, the work encompasses far more. On their own, the digital techniques do little to change agricultural outcomes. For farmers to benefit, data must be converted into information that is understandable, relevant, and actionable. And it needs to be disseminated through accessible and trustworthy channels. As described by one team member, "blunt instruments are useless on their own without having someone who can interpret and understand." Complementing the data collection and analysis elements of data-driven agronomy is a knowledge management and communications effort to deliver the information to farmers and the technicians who serve them.

For Colombian farmers, data-driven agronomy comes at a critical moment. Farmers are experiencing declining yields of rice, cereals, and other important crops. In the last five years, the national average yield

8 http://bigdata.cgiar.org/communities-of-practice/data-driven-agronomy/

for irrigated rice dropped from 6,000 kilograms per hectare to 5,000 kilograms per hectare; a change largely attributed to climate variability.[9] Farmers are also noticing a rise in the incidence of disease and pests. The national growers' associations, who deliver inputs and technical assistance to farmers, are challenged to keep up with the changing farming conditions, despite their agricultural research experience. And although they have access to an immense amount of current and historical data about their farmers, technicians and researchers within these associations lack the knowledge and capacity to put that data to use in new ways to serve their clients. This is now changing.

The discussion that follows describes three important elements of the data-driven agronomy work: 1. Data Collection and Aggregation; 2. Data Analysis; and 3. Knowledge Management and Communications.

Data Collection and Aggregation

Access to data is necessary for this approach to work, and while the data exists, it is often not publicly available or in a format that makes it immediately useful for analysis. Sufficiently large and detailed datasets are required to produce reliable information about the interaction of different variables. CIAT's data-driven agronomy work uses a broad range of data – crop yields, farming practices, climate, weather, and soils. Datasets are aggregated from existing sources including open data sets, experimental data, and partners' historical data and is complemented by new data, for example, from weather stations. While access to open datasets is easy, the institutions that hold farmer-level data are often reluctant to share it. At least this was CIAT's experience, which meant investing time and effort to build trust and understanding with a range of partners.

When the data-driven agronomy team sought partners for their work, national growers' associations were ideal: they have data about farmers' crop management and yields and they have access to farmers. For example, Fedearroz had two decades worth of data about its farmers and their management practices. Yet, like many private entities, the growers' associations place a high value on their data and consider it proprietary, even if it is not being used. To entice these partners into collaborating, CIAT requested small samples of data

9 Delerce et al. 2016

A Fedearroz technician describes how data-driven agronomy has changed their rice trials and research

A weather station at Fedearroz used to capture local climate information for data-driven agronomy.

which it used to demonstrate what the data mining techniques could deliver. This demonstration, which resembles how agricultural technicians work with farmers, helped to open the growers' associations data archives. And as more data was shared with CIAT, the data-driven agronomy team was able to produce more useful findings.

Additionally, CIAT helped to design data collection systems and provided extensive capacity building for partner staff on data entry, running models, and interpreting outputs. This was in part because data quality was a major challenge. This meant that initially CIAT had to identify proxy data or methods that would accommodate missing data. CIAT worked closely with partners to improve data collection and data management systems, an investment for which they had not originally planned. Although significant time and expense was spent on cleaning the data, the partner organizations are now better positioned to ensure data quality and sustain the analytical tools after the work with CIAT has ended.

Data Analysis

Traditional controlled experiments take time, are expensive, and are conducted on a limited set of fields that do not represent the soil or climate diversity among farmers' fields. Further, the modeling techniques used in conventional experiments require a preconceived understanding of relationships between variables and are affected by outliers and noisy data.[10] In contrast, the data mining techniques used in data-driven agronomy offer the flexibility to explore observational data from farmers without needing to specify complex relationship between input and output variables ahead of time. These techniques are useful for datasets that are variable, noisy, incomplete, imprecise, and qualitative. They allow for the exploration of crops and productivity limiting factors that have not been sufficiently studied and for specific weather and soil conditions.

10 Delerce et al. 2016

The data-driven agronomy approach has been used by CIAT with crops including Andean blackberry, sugarcane, limes, plantain, avocado, lulo, beans, mango, maize, rice, and banana. A small project on Andean blackberries first demonstrated the promise of big data methodologies for site-specific agriculture. In this case, it saved years of experimental cropping systems research to identify ideal management practices for an understudied and relatively poorly understood crop.

Data driven agronomy utilizes multiple approaches to data analysis. Sometimes traditional statistical models are sufficient; in other cases, newer techniques including supervised and unsupervised machine learning[11] are better suited for the types of data available and specific research questions. The CIAT teams select analytical approaches according to specific needs of a project. Different modeling approaches involve different trade-offs between the accuracy of model predictions, the interoperability of model inputs, and ability to handle "messy" data.[12] Through various data mining and machine learning techniques, they can layer multiple types of data to look at patterns that emerge across specific crop varieties, climatic conditions, and location to identify optimal site-specific management practices and cultivars. The CIAT teams have used multiple approaches in their data driven agronomy work, many of which are detailed in the papers listed in Box 4.

Analytical methods continue to improve. In the beginning, CIAT's partners wanted to understand the decline in farmers' yields. CIAT developed explanatory models to identify the most limiting factors causing variance in expected yield where there were knowledge gaps.[13] By identifying the

BOX 4
CIAT SCIENTIFIC PAPERS

For more information on the specific data mining techniques and approaches used by CIAT and partners, please review the following papers:

Delerce, S., H. Dorado, A. Grillon, M. C. Rebolledo, S. D. Prager, V. H. Patiño, G. G. Varón, and D. Jiménez. 2016. Assessing Weather-Yield Relationships in Rice at Local Scale Using Data Mining Approaches. PLoS ONE 11(8): e0161620 https://doi.org/10.1371/journal.pone.0161620

Jiménez, D., H. Dorado, J. Cock, S. D. Prager, S. Delerce, A. Grillon, M. A. Bejarano, H. Benavides, and A. Jarvis. 2016. From Observation to Information: Data-Driven Understanding of on Farm Yield Variation. PLoS ONE 11(3): e0150015 https://doi.org/10.1371/journal.pone.0150015

Jiménez, D., J. Cock, H. F. Satizábal, M. A. Barreto S, A. Pérez-Uribe, A. Jarvis, and P. Van Damme. 2009. Analysis of Andean blackberry (Rubus glaucus) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data. Computers and Electronics in Agriculture 69 (2009) 198-208. Author's personal copy accessed through CGIAR http://hdl.handle.net/10568/43181

11 Machine learning is a method of data analysis that automates analytical model buildin based on the idea that machines should be able to learn and adapt through experience. Machine learning techniques represent an approach to allow the discovery of embedded knowledge that may be present in data. (SAS https://www.sas.com/en_us/insights/analytics/machine-learning.html)

12 Delerce et al. 2016

13 Explanatory models analyze observational data for associations to test causal theories. Predictive models are designed to use models and data mining to accurately predict new or future out-of-sample observations. Predictions may be used for developing new theories but require a high degree of accuracy (Shmueli 2010)

most limiting productivity factors for a single crop, in a particular location, with specific weather events, CIAT and partners can provide recommendations to farmers about varieties and site-specific management practices, based on soil type and altitude/temperature, and seasonal forecasts. Despite the initial focus on explanatory models, the team at CIAT saw great potential to use the same approaches for developing predictive models. Jiménez notes that when CIAT tested the performance of its models. Daniel Jiménez, Coordinator of the Community of Practice for Data-Driven Agronomy at CIAT, notes that when CIAT tested the performance of its models on more complete datasets like CIMMYT's[14] MasAgro database, it saw very good performance with a higher degree of confidence, which will allow the models to predict and not just explain variation.[15] This implies that as CIAT and partners improve data accuracy, quality, and management in the future with more observations, they can use the current explanatory models to make crop behavior predictions. Both explanatory and predictive models require quality data, though predictive models may require even better data to ensure a higher degree of accuracy.

Throughout the analytic process, the data-driven agronomy work has been informed, validated, and improved by an interdisciplinary team made up of data scientists, agronomists, and anthropologists, from CIAT and partner organizations. Each expert is a filter to verify, based on their knowledge and position in the ecosystem whether the analytical outputs are logical. The team uses other models and approaches to validate outputs from the data-driven agronomy models and, when necessary, adjusts the models by discarding poor data or adding more data to ensure a high degree of accuracy in the results.

KNOWLEDGE MANAGEMENT AND COMMUNICATIONS

To affect agricultural development outcomes, CIAT had to ensure that the analytical results produced by the data mining techniques reached farmers in actionable formats they would understand. The information needs to be analyzed and interpreted, packaged appropriately for the target audience, and then disseminated through accessible and trustworthy channels, like growers' associations. While CIAT is well-positioned to assume the lead research role, growers' associations and government partners are better positioned to effectively engage farmers. For many of the farmers, the last-mile interaction with advisory service providers has not changed with data-driven agronomy. Agricultural extension officers and technicians remain the key information providers for cultivar varieties, planting dates, and management practices. Data-driven agronomy strengthens the role of farmer-facing organizations and underscores how even with modern high-tech approaches, person-to-person interaction is still necessary.

A significant investment was made to build capacity among farmers and technicians to understand the research outputs. This began with an assessment of what farmers and technicians understood. Results of an assessment of Colombian fruit farmers showed that farmers, overall, were receptive to sharing information within their own immediate circles. The assessment

14 The International Maize and Wheat Improvement Center (CIMMYT) is a CGIAR research center. The Sustainable Modernization of Traditional Agriculture program (MasAgro) is a research for rural development project of Mexico's Secretary of Agriculture, Livestock, Rural Development, Fisheries and Food (SAGARPA) and CIMMYT that promotes a sustainable intensification of maize and wheat production in Mexico.

15 Interview with D. Jimenez, October 30, 2017.

also revealed a need to build capacity around understanding probability for seasonal forecasts and interpreting graphs. According to CIAT interviews, while farmers and agricultural technicians have a basic grasp of concepts like weather, they often do not understand the nuances of weather forecasts, agro-climatic forecasts, and climatology.[16] When the weather is not what was forecast, farmers assume that the information was poor and therefore, untrustworthy. Assessments showed that farmers and technicians do not understand how to interpret forecasts, and specifically they do not understand how probability and uncertainty relates to forecasts. While workshops to understand how to present research results revealed a low level of understanding from technicians and farmers alike, CIAT found that farmers and technicians can rapidly grasp and understand technical information when presented in a simplified format by facilitators.[17]

Another challenge that CIAT faced was a desire by partners to provide technical recommendations to farmers before understanding the degree of accuracy of the models or what variation means for the recommendations they make. CIAT worked with farmers, technicians, and other staff from partner organizations to build their knowledge of basic concepts, like probability. This provided a foundation upon which the research results could be disseminated and discussed. As a result, CIAT explored different visual representations of data with participants to determine the most effective ways to present information. They also coached technicians and researchers from the partner organizations to strengthen their ability to use the digital tools and interpret the information. Empowering farmers and technicians with information is a key objective of CIAT's work. While identifying site-specific productivity limiting factors is important, the data-driven agronomy team is equally interested in providing information to farmers and technicians that they can use to make decisions. This is illustrated by the integration of the data-driven agronomy work in the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) in Cordoba, Cauca, and other departments in Colombia. Drawing on a model used in Senegal, in 2016 Fedearroz and the MADR, along with CIAT and CCAFS, established local technical agro-climatic committees (LTACs). The LTACs are designed to increase farmers' understanding of climate variability and uncertainty, and become familiar with interpreting data from weather stations, satellites, and other observations from the community. Each LTAC is made up of farmers, as well as representatives from the private sector, agricultural research organizations, and national and local government. At monthly meetings, the group validates the previous month's forecast, and then looks at the current forecast and model information to determine specific adaptation measures. The forecast data and adaptation measures are collaboratively developed into a bulletin that is circulated to farmers with the recommendations also made available through radio and WhatsApp.

Working with CIAT, partner organizations have built various online tools for storing, analyzing, and visualizing data. Several have online information platforms, like Aclímate Colombia, a partnership between CIAT, the MADR, and crop growers' associations like Fedearroz.

16 CIAT n.d.
17 Howland et al. 2015; Staiger-Rivas et al. 2014.

With these platforms, farmers can access current and historical data about their specific farming businesses, and technicians can consult information about their clients' farms.[18] While it is possible for farmers to access this information directly, the current emphasis is on providing access to data-driven agronomy information through groups. The number of individual farmers using the online platforms for decision-making is still unknown.

UNDERSTANDING THE IMPACT OF DIGITAL TOOLS

Despite being relatively young, data-driven agronomy has already led to notable impacts on farmers, partner organizations, and CIAT. As the discussion below reveals, the approach has drastically increased the number of farmers who have access to reliable, tailored information with which they can make decisions. It is modernizing and improving the extension services delivered by the government and growers' associations and providing new ways for agricultural research centers to use data to close yield gaps and adapt to climate variability.

USING INFORMATION TO MAKE CROP MANAGEMENT DECISIONS

The biggest change for farmers has been increased access to better quality and relevant information. Farmers in Montería (Córdoba) described a culture change with respect to information and data. They have a better appreciation for the value of data and have learned to use climate, soil, and observational information in their practices to validate whether their historical management practices will be effective. If not, they assess what kinds of alternative practices may be more suited to the current conditions and decide whether to adopt them. Many farmers are seeing the benefit of using and analyzing information, for example, from the LTACs to sustain productivity or mitigate losses.

In 2014, during the first year of Aclímate Colombia, CIAT and Fedearroz predicted a dry spell in Córdoba and recommended farmers delay planting rice. About 170 farmers followed this advice, planting during the window recommended by CIAT and Fedearroz. Evaluations estimate that the farmers who delayed planting avoided crop losses valued at over USD 3.6 million and fared significantly better than neighboring communities that planted according to the usual schedule.[19]

CIAT and its partners estimate that over 150,000 farmers received climate and/or crop management information through the LTACs in Colombia.[20] An estimated 6,000 farmers adopted climate-smart practices after receiving advisory information. In Córdoba, Fenalce and ICA (Colombian Agricultural Institute) endorsed recommendations about when to plant maize. Farmers that followed this recommendation experienced better results than those who did not comply.[21] Yet, not all farmers make different decisions with the new information they are receiving. Farmers are less likely to adopt a new strategy unless they see something proven to be effective. During interviews for this case study, one farmer indicated that he will always take a recommendation if it matches his own knowledge and the way he has farmed in the

18 Additional online platforms include Agricultura Específica Por Sitio (AEPS), Sistema de Recolección de Información Agrícola (SIRIA), and Frutisitio.
19 Young and Verhulst 2017.
20 CCAFS 2015a.
21 CCAFS 2015b.

past. Another said that most farmers will not alter their practices the first time a recommendation is provided, even if they have been losing money with their own practices. It is when they see others experience success that they are willing to change their behavior. Understanding what drives farmers to make different decisions is an area that CIAT is in the process of exploring. The interdisciplinary research team would like to conduct an impact assessment on data-driven agronomy and changes in farmer decision-making, but at this time, only has anecdotal information about why farmers are making decisions based on the analytical information CIAT provides them.

During the interviews for this case study, trust in the information provider emerged as one reason farmers adopt a new practice.[22] Among the farmers that were interviewed, many cited their confidence in the information they received because it came via a trusted organization, like Federarroz or Fenalce. Such intermediaries provide cohesion and communication between farmers and experts, which also facilitates the adoption of new technologies and practices by farmers.[23] Others cited the additional benefit of knowing that a research institute like CIAT was working with their growers' association. Still, trust is difficult to win and easy to lose, which makes agricultural technicians of the growers' associations and CIAT staff worry about what will happen when models lead to a recommendation that results in poorer outcomes for farmers who adopted the recommended practices compared to those who did not. It is not possible to predict how much trust will be lost when that happens.

Farmers, however, are no longer just passive recipients of information. They are actively engaged in discussions with scientists and technicians, which creates a sense of collective ownership over the decisions farmers ultimately make, potentially mitigating any negative consequences with future decisions.

IMPROVING THE SERVICES AND CREDIBILITY OF FARMER SERVING ORGANIZATIONS

Farmer intermediary organizations are crucial actors for interpreting and delivering digestible and useful information to farmers. The organizations that partnered with CIAT learned to use their observational data to understand historical events and what happened in farmers' fields. This strengthened their ability to serve clients, improved their reputation as credible and professional organizations, and enhanced their capacity for planning and resource mobilization.

Data-driven agronomy has allowed the government and growers' associations to make targeted recommendations to maximize their resources. For example, analysis with data about the Saldaña region suggested that during a year with the El Niño phenomenon, they would face a dry spell, which led Fedearroz to advise the district to ration water. While the district planted less total crops, they experienced lower losses than a neighboring district.[24]

In Pereira, the Secretary of Agriculture and Rural Development used the principles of site-specific agriculture to identify 119 groups of agroecological zones within 47,000 hectares in the municipality as

22 Also see Blundo et al. 2016 for more information.
23 Howland et al. 2015.
24 Young and Verhulst 2017.

high potential zones for specific commercial crops.[25] This made it possible for farmers to choose crops that would increase the profitability of their farms. Using the zone information also allowed the Secretary to petition the municipal planning department to reserve specific land for agricultural use. Classifying the crop suitability of agroecological zones also improved the allocation of credit to farmers.

Additionally, the MADR uses what it knows about the suitability of crops for different plots to provide credit incentives to farmers to switch to those crops and design loan terms around the appropriate time horizon for those crops.

Using data in these ways has changed organizational attitudes about data collection and analysis. Many organizations spent time and money collecting monitoring data but were frustrated at their inability to learn something different from the data. The data-driven agronomy approach was appealing to these organizations because it increased the power and usefulness of their data. As described by a representative from Fedearroz, "data collection is fundamental to be able to continue to generate information that is useful for producers." Furthermore, when the technician is more prepared, service delivery is streamlined and time spent on-farm is significantly reduced. As representatives from the Secretary of Agriculture and Rural Development explained, a technician now has, "the entire medical history of a specific farm."

Through the collaboration, partner organizations learned the importance of relevant and quality data for data-driven methodologies. Following their work with CIAT, many organizations began collecting new types of information or altered their data management practices. For example, collecting local climate data was new for many organizations. While they understood the importance of climate information, they did not know how to effectively utilize the seasonal forecasts and weather information. This led some organizations, like Fedearroz, to invest in local weather stations. Additionally, efforts were made to improve the quality of data. Previously, as a representative from Fedearroz explained, data was not standardized and much of it reflected subjective assessments determined by technicians. Now datasets are standardized and managed in an interoperable way.[26] While the datasets themselves are not open yet, the growers' associations are more open to sharing their data.

The improvements in data analysis, bolstered by CIAT's reputation, enhanced the growers' associations credibility in front of their members, donors, and the government. The methodologies show that they are modernizing the way they manage information and can improve the extension services they offer members. For example, the free public-facing data-driven platform, SIRIA, raised Fenalce's profile as a modern organization. In the case of the Municipality of Pereira, the value of site-specific agriculture is illustrated by continued investment despite three changes in government.

EXPANDING ANALYTICAL AVENUES FOR AGRICULTURAL RESEARCH

Data-driven agronomy has expanded the range of analytical methods used in CIAT's agricultural research.

25 Alcaldia de Pereira 2015.

26 CIAT complies with the FAIR Guiding Principles for scientific data management and stewardship, which defines interoperable as (meta)data that uses "a formal, accessible, shared, and broadly applicable language for knowledge representation, …vocabularies that follow FAIR principles, …and includes qualified references to other (meta)data." (https://www.nature.com/articles/sdata201618)

The change at CIAT has been described as a slow evolution rather than a revolution. When the team first explored using big data for understanding productivity, it was seen as "a little bit of heresy," according to one team member. The approach differs significantly from conventional agronomic research: it relies heavily on observational data from farmers and data mining techniques instead of controlled experiments and is then packaged as recommendations to farmers. As a result, during the early years, the team spent time convincing peers and partners that data-driven agronomy could generate accurate research outputs and positive results for farmers.

The interest in data-driven agronomy by a range of donors strengthened the status of CIAT's research within the CGIAR, with potential partners, and with other donors. CIAT's projects won the UN Global Pulse award for the Big Data Climate Challenge in 2014, and the World Bank Innovation Challenge for Big Data for Development in 2015. In 2017, CIAT and CCAFS won an UNFCCC Momentum for Change Lighthouse Activities award in the ICT Solutions category. The awards and funding over the last decade highlighted the importance of their work, as a complement to breeding improved crop varieties, for closing farmer yield gaps and adapting to climate variability.

CIAT now leads the Data-driven Agronomy Community of Practice under CGIAR's Platform for Big Data in Agriculture. The Platform launched across the CGIAR network and research programs in 2017, with the goal of helping development practitioners use proven big data innovations to drive agricultural growth.[27] Under the Platform, "big data" is defined as, "open, harmonized, interoperable, and integrated datasets from multiple domains aimed to accelerate agricultural research and data use in service of development goals."[28] CIAT's work within the Data-Driven Agronomy Community of Practice continues their exploration of using modern big data techniques to "advance agricultural practices, in relation to particular socio-economic and environmental contexts in a way that promotes the democratization and transparency of agricultural information."

This work continues to expand outside of Colombia. CIAT is working with the Latin American Fund for Irrigated Rice (FLAR), a public-private partnership, to replicate their successes with Fedearroz. Under CCAFS, the Government of Honduras established seven local technical agro-climatic committees and CIAT secured funding for capacity building with the Secretary of Agriculture and Livestock.[29] In Argentina, CIAT collaborated with the National Agricultural Technology Institute (INTA) to support a similar program that proved the data-driven agronomy concept and research to stakeholders. CIAT supported INTA to enhance their analytics much more quickly and apply the approach to more crops. The World Bank funded part of that two-year initiative. CIAT and INTA are continuing to look for additional ways to collaborate.

27 CIAT and IFPRI 2016; CIAT and IFPRI 2017.
28 CIAT and IFPRI 2017.
29 UNCC n.d.

LESSONS LEARNED

This section outlines lessons learned from CIAT's and partners' experiences with the digital tools highlighted earlier in this case study. It is organized around the Principles for Digital Development: a set of principles developed by donors and the development community to guide and inform technology-enabled development programs.[30] The discussion below includes the most significant reflections, and therefore, not all of the Principles are represented.

Principle 2: Understand the Ecosystem

The CIAT data-driven agronomy team is connected to organizations and networks within Colombia and the CGIAR that improve and facilitate its work. Within Colombia, CIAT partners with civil society institutions and the public sector who act as intermediaries to reach farmers and provide access to data, whether private or open. It was important for CIAT to work through these partners because they have strong and trusted relationships with farmers. Furthermore, these organizations had the data that drove the analytical work. Together, CIAT, the growers' associations, and the farmers have created a network with strong feedback loops that allows for continuous improvement and validation of the models.

Not all countries will have the same ecosystem, particularly with respect to the availability of local data. CIAT demonstrated that it can work with organizations whose data systems are of differing quality. However, the team initially spent more time strengthening Fenalce's data systems than with Fedearroz who benefited more quickly from the advanced analytical work.

Data access and quality are important to determine what kind of data-driven agronomy effort is possible in different contexts.

Principle 5: Be Data Driven

CIAT's data-driven agronomy program is completely data-driven. It utilizes data to make reliable information available to the right people, so they may make better decisions. An important part of CIAT's work is cleaning and aggregating diverse datasets so it's available in a form suitable for analysis. CIAT works with its partners to strengthen methodologies and infrastructure that support consistent, high-quality data collection and management. It also works to present data in formats that are easy for stakeholders to interpret, and prioritizes a data-use culture for all stakeholders.

30 http://digitalprinciples.org/

Principle 6: Use Open Standards, Open Data, Open Source, and Open Innovation

CIAT's digital intervention embodies Principle 6, and the team has been explicit about abiding by the FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles.[31] CIAT utilizes open data available in Colombia. Whenever possible they use freely available analytical and data management tools to help democratize this approach.

A big challenge facing big data in agriculture interventions is that, in most countries, agricultural data is fragmented between various actors and often goes underutilized because it is not shared or analyzed. Many groups see their data as an asset and monetize it rather than unlocking and sharing it for its potential benefits. Yet, CIAT is trying to change institutional mindsets about the value of proprietary data by demonstrating the potential generated through collectivized knowledge. Individual data sets have little value, but when linked and analyzed, it takes on value. They are working to foster open data and data sharing among and within national growers' associations. CIAT aims to kickstart a data ecosystem where there is a flow of information to and from farms and institutions so that the data each actor already possesses may be analyzed and put to additional use to serve researchers and farmers in closing the productivity gap.

Principle 7: Reuse and Improve

CIAT's intervention utilizes tools and approaches from within the CGIAR, academia, partner organizations, and large-scale mechanized agriculture in developed countries. They have adapted these lessons into transferable approaches that can be applied to different value chains in tropical and other increasingly volatile climates. CIAT has improved their approach over ten years by securing additional research funding to refine the analytics, add new data types, and explore different methodologies for disseminating information. For example, the use of LTACs was an approach seen during an exchange visit to a CCAFS site in Senegal and adapted for use in Colombia. CIAT and partners have also gained a greater understanding of the resources needed to implement data-driven agronomy. The key determining factors include:

- The area of land under cultivation and degree of variation within this area.
- The capacity of the association to collect, interpret, and disseminate data.
- The quality of data available from the government, the private sector, and other partners, as well as the time needed to develop the relationships with different partners to access data.
- The digital landscape, and specifically the connectivity level that affects the ability to upload and transfer data.

31 For more information see http://www.datafairport.org/.

Principle 9: Be Collaborative

CIAT's digital intervention is an excellent example of strong collaboration between different types of partners across disciplines. They work closely with public institutions in Colombia, Honduras, Mexico, and Argentina, with private growers' associations, and with other research entities in the CGIAR and national public spaces. The pooling of expertise and resources has allowed CIAT to advance digital interventions and expand its reach as a research institution (For example in its leadership of the CGIAR Data-Driven Agronomy Community of Practice).

The digital intervention also takes advantage of diverse experts to ensure that information is accurate and accessible to stakeholders. Each engagement includes data professionals with the analytical skills to manage data, design statistical learning models, and interpret results--agronomists, researchers, and subject matter experts that ground-truth results and improve the algorithms to reflect agricultural realities--and knowledge management and communications personnel that develop effective ways to disseminate information. CIAT's confidence in their models stems from its iterative and consultative process of developing and improving the models, validating the output with agronomists and agricultural technicians, and in building feedback loops with farmers through the LTACs. This collaboration creates a sense of ownership among CIAT's partners and end-users, which instills trust in the process, even if the recommendations do not always lead to the best outcomes for farmers.

Even with this significant collaboration, there remains a deficit of analytical experts with sufficient data science expertise within Colombia. The team acknowledges that scaling and replicating these approaches will require a sufficient pipeline of data science experts into agricultural development and research, or agricultural researchers with data science skills.

Photo by Megan Johnson, USAID

CONCLUSION

Data-driven agronomy is a digital integration approach that has demonstrated impact for a variety of actors and is a complementary methodology to conventional agronomic research for helping farmers close yield gaps and adapt to climate variability. It is an approach that has been applied to agricultural development only recently but contributes to narrowing the divide between agricultural research and practice and extends the boundaries of what researchers can do to support agricultural development.

Because all models are inherently uncertain, instances of "wrong" or "incorrect" recommendations are inevitable. Yet to date, CIAT and its partners have not issued recommendations that have led to significant losses or undermined the willingness of growers' associations to make data-informed decisions. When failure does happen, a scenario that most of the team acknowledges is an inevitability, the trust and ownership of the process by farmers and intermediary organizations will be tested. Yet the hope is that the work the team has done in helping organizations establish data management systems and building analytical capacity will help weather those storms.

In addition to the team's own refinements, the rapidly changing digital environment is going to alter what is possible in the future. As mobile phone access, and specifically the number of smartphones and 3G and 4G connections expand, the research results can be disseminated to many more farmers. This may enable CIAT to reach smallholder farmers, a group of farmers that is not associated with growers' associations and is more difficult to reach. The analytical possibilities of the techniques will continue to improve with higher resolution satellite imagery and on-field sensors that can provide more granular information about climate and soil quality. Finally, the expanding use of artificial intelligence and other approaches to improve processing capacity will further enhance research and development outcomes. As more of this data becomes available, CIAT is well-positioned to harness it to truly "put data at the service of agriculture."

REFERENCES

Alcaldía de Pereira. 2015. Transformación del Sector Rural en el Municipio de Pereira, Risaralda: Informe de Gestión 2008 – 2015. Pereira: Secretaria de Desarrollo Rural.

Blundo Canto, G., G. Mendez, P. Alvarez-Toro, L. Perez Marulanda, and C. Gartner Vargas. 2016. Local, reliable and timely agro-climatic information: a requirement of Colombian farmers. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Cali, Colombia. https://ccafs.cgiar.org/publications/local-reliable-and-timely-agro-climatic-information-requirement-colombian-farmers#.WodcR6gbM2y

CCAFS. 2015a. Scaling up climate advisories in Senegal and Colombia. Outcome Cases. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Cali, Colombia. https://ccafs.cgiar.org/research/results/scaling-climate-advisories-senegal-and-colombia#.WWkmVdKLQ2w

CCAFS. 2015b. Local Technical Agroclimatic Committees. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Cali, Colombia. https://ccafs.cgiar.org/local-technical-agroclimatic-committees#.WosSGq6nFhF

CIAT. 2015. Site-Specific agriculture: putting data at the service of agriculture. International Center for Tropical Agriculture (CIAT), Cali, Colombia. http://hdl.handle.net/10568/67309

CIAT. N.d. Report: Workshops for the identification of demands for the climate services web interface for maize and rice.

CIAT and IFPRI. 2016. CGIAR Big data coordination platform: Proposal to the CGIAR Fund Council, 31 March, 2016. International Center for Tropical Agriculture (CIAT), Cali, Colombia and International Food Policy Research Institute (IFPRI), Washington, DC. https://cgspace.cgiar.org/bitstream/handle/10947/4303/2-Big%20Data%20Platform%20Full%20Proposal.pdf?sequence=1

--2017. Event Summary and Agenda for Action, First Annual CGIAR Convention on Big Data in Agriculture: Alliance for a Data Revolution September 19-22, 2017. CGIAR. http://bigdata.cgiar.org/wp-content/uploads/2017/11/First-Annual-CGIAR-Platform-for-Big-Data-in-Agriculture-Convention-Summary-Report.pdf

Delerce, S., H. Dorado, A. Grillon, M. C. Rebolledo, S. D. Prager, V. H. Patiño, G. G. Varón, and D. Jiménez. 2016. Assessing Weather-Yield Relationships in Rice at Local Scale Using Data Mining Approaches. PLoS ONE 11(8): e0161620 https://doi.org/10.1371/journal.pone.0161620

Howland, F., L.A. Muñoz, S. Staiger-Rivas, J. Cock, and S. Alvarez. 2015. Data sharing and use of ICTs in agriculture: working with small farmer groups in Colombia. Knowledge Management for Development Journal 11(2): 44-63 http://journal.km4dev.org/

Jiménez, D., H. Dorado, J. Cock, S. D. Prager, S. Delerce, A. Grillon, M. A. Bejarano, H. Benavides, and A. Jarvis. 2016. From Observation to Information: Data-Driven Understanding of on Farm Yield Variation. PLoS ONE 11(3): e0150015 https://doi.org/10.1371/journal.pone.0150015

Jiménez, D., J. Cock, H. F. Satizábal, M. A. Barreto S, A. Pérez-Uribe, A. Jarvis, and P. Van Damme. 2009. Analysis of Andean blackberry (Rubus glaucus) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data. Computers and Electronics in Agriculture 69 (2009) 198-208. Author's personal copy accessed through CGIAR http://hdl.handle.net/10568/43181

Shmueli, G. 2010. To Explain or To Predict? Statistical Science, Vol. 25, No. 3, 289–310. Institute of Mathematical Studies. http://dx.doi.org/10.1214/10-STS330Surampudi, S. 2017. "Oracle Data Mining Concepts, 12c Release 1 (12.1)." ORACLE. https://docs.oracle.com/database/121/DMCON/GUID-8232ABAD-E6B9-4C70-B227-E00738040932.htm#DMCON002

United Nations Climate Change. N.d. ICTs For Small-Scale Farmers: A Game Changing Approach to Climate Smart Agriculture in Latin America. http://unfccc.int/secretariat/momentum_for_change/items/10464.php

Young, A. and S. Verhulst. 2017. Aclímate Colombia: Open Data to Improve Agricultural Resiliency. The GovLab. http://odimpact.org/case-acl%C3%ADmate-colombia.html