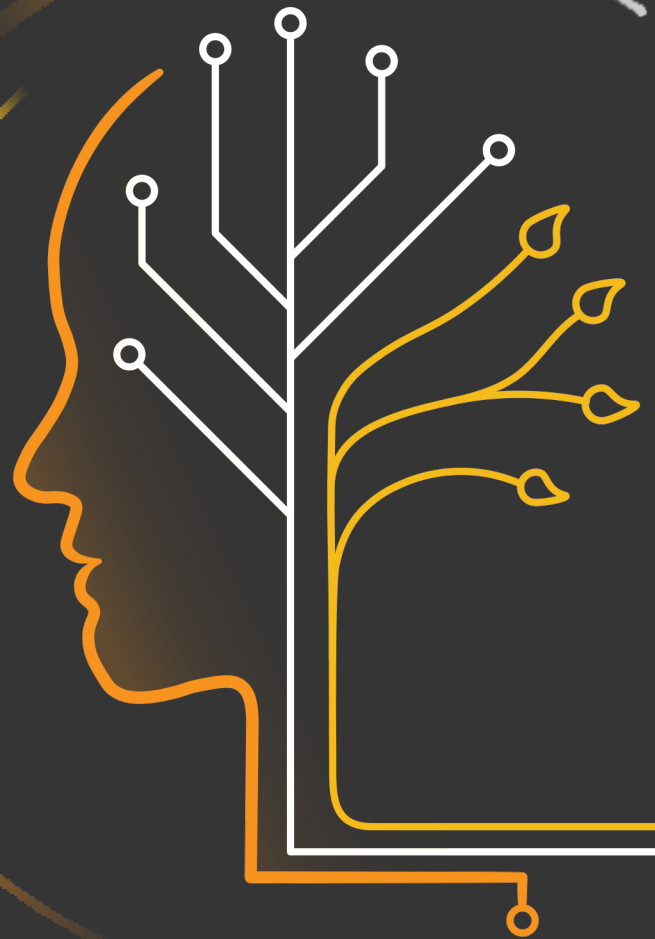




Platform for
Big Data
in Agriculture

CGIAR PLATFORM FOR BIG DATA IN AGRICULTURE ANNUAL REPORT 2019



Alliance



INTERNATIONAL
FOOD POLICY
RESEARCH
INSTITUTE



Platform for
Big Data
in Agriculture

Name of the Platform:

BigData

Name of Lead Center:

CIAT

Module lead institutions (CGIAR Centers or lead partners)

Module 1: Organize

Module 2: Convene

Module 3: Inspire

Other participating Centers:

AfricaRice, BIOVERSITY International, CIAT, Center for International Forestry Research (CIFOR), International Maize and Wheat Improvement Center (CIMMYT), International Potato Center (CIP), International Center for Agricultural Research in Dry Areas (ICARDA), World Agroforestry Center (ICRAF), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), International Food Policy Research Institute (IFPRI), International Institute of Tropical Agriculture (IITA), International Livestock Research Institute (ILRI), International Rice Research Institute (IRRI), International Water Management Institute (IWMI), and WorldFish.

TABLE OF CONTENTS

1

KEY RESULTS	5
1.1 Highlight Platform Achievements	5
1.2 Platform Progress towards Outputs and Outcomes	7
1.2.1 Overall Platform progress	7
1.2.2 Progress by modules	7
1.2.3 Variance from Planned Program for this year	16
1.2.4 Altmetric and Publication highlights	16
1.3 Cross-cutting dimensions	17
1.3.1 Gender	17
1.3.2 Youth and other aspects of Social inclusion/ "Leaving No-one Behind"	18
1.3.3 Capacity Development	19
1.3.4 Climate Change	20

2

EFFECTIVENESS AND EFFICIENCY	21
2.1 Management and governance	21
2.2 Partnerships	21
2.2.1. Highlights of External Partnerships	21
2.2.2. Cross-CGIAR Partnerships	22
2.3. Intellectual Assets	22
2.4 Monitoring, Evaluation, Impact Assessment and Learning (MELIA)	23
2.5 Efficiency	23
2.6 Management of Risks to Your Platform	23
2.7 Use of W1-2 Funding	23

3

FINANCIAL SUMMARY	24
TABLES	25

EXECUTIVE SUMMARY

Digital disruption is changing all industries, and agriculture is no exception. This is due in part to a common business model: the digital platform. A platform is a multi-sided, technology-enabled network that facilitates the interaction of stakeholders. The BIG DATA Platform is both a cross-cutting program to better integrate CGIAR and a digital platform strategy for transforming our global organization through its three modules: Organize, Convene, and Inspire.



ORGANIZE

The Organize module builds fundamental technology and data standards to support CGIAR's digital strategy. Our data discovery and analysis flagship data harvesting tool -- the Global Agricultural Research Data and Innovation Network (**GARDIAN**) -- now points to more than 155,000 publications and 23,000 datasets from several partners alongside all CGIAR Centers. In 2019, Organize **contributed** to leveraging semantic standards for describing agronomic, socioeconomic, and survey data; updating the CGIAR metadata standard; and **enabling digital collection of standards-compliant agronomy data**. A secure analytic environment for researchers to find data and collaborate on analyses was created in the GARDIAN ecosystem, integrating single sign-on (SSO) technology with **Globus**, a secure data sharing service. The new **Expert Finder** showcases CGIAR research, fosters new collaborations, and creates visualizations of institutional partnerships and expertise by location. Building community and tools around data discovery and standards realizes a key component of a digital platform strategy for CGIAR.



CONVENE

Digital platforms are multi-sided and driven by communities. In 2019, the Convene module made significant strides in developing CGIAR's digital partner networks. Six **Communities of Practice** (CoPs) dramatically increased their communications output, growing to 3,500 CGIAR and non-CGIAR members. The CoPs released important community-driven outputs including an early warning system for wheat blast; draft ground data standards to facilitate machine learning analysis of satellite data; a survey of CGIAR digital extension efforts. Convene culminated in the annual **BIG DATA Convention**, where 700 public, private, and non-profit actors assembled to examine digital models relevant for global collective action, and the CoPs organized technical meetings on livestock data, geospatial analysis, data-driven agronomy, socioeconomic data, crop modeling, and ontologies. During the event, the Platform showcased a **reference** Internet of Things architecture for agronomy and breeding, ground-breaking gender research methods that leverage telecom data, and technology standards for managing sensitive data.



INSPIRE

The Platform made four start-up grants to data-driven partnerships under its Inspire Challenge innovation initiative, and four scale-up grants to winning projects from 2018 and 2017, awarding 1.025 million USD. Its growing portfolio, totaling 14 **projects** in 2019, demonstrates impressive early-stage results. For example, one advisory **service** uses crowd-sourced smartphone images, reached 33,000 Indian wheat farmers, increasing crop insurance efficacy and knowledge on agricultural practices by 78%. A **project** tracing informal food flows by leveraging free Wi-Fi has collected data from more than four million smartphones and has been adopted at a national scale in Vietnam as a tool to assess and predict COVID-19-related food security shocks. A near real-time small-scale fisheries monitoring **system** is being scaled to seven countries in Africa and Asia. The Inspire Challenge, attracting interest from funders and partners, generated 200,000 USD in external commitments in 2019.

01

KEY RESULTS

1.1 HIGHLIGHT PLATFORM ACHIEVEMENTS



ORGANIZE

The Organize module builds fundamental technology and data standards to support CGIAR's digital strategy. GARDIAN, our data discovery and analysis flagship, now points to over 155,000 publications and 23,000 datasets from several partners alongside all CGIAR Centers. In 2019, Organize significantly **contributed** to leveraging semantic **standards** for describing **agronomic, socioeconomic, and survey data, updating the CGIAR metadata standard, and enabling digital collection of standards-compliant agronomy data**. A secure analytic environment for researchers to find data and collaborate on analyses was created in the GARDIAN ecosystem, integrating single sign-on (SSO) with **Globus**. The new **Expert Finder** showcases CGIAR research, enables new collaborations, and visualization of institutional partnerships and expertise by location. Building community and tools around data discovery and standards realizes a key component of a digital platform strategy for CGIAR.



CONVENE

In 2019, the Convene module made significant strides in developing CGIAR's digital partner networks. Six Communities of Practice (**CoPs**) dramatically increased their communication output, growing to 3,500 CGIAR and non-CGIAR members. The CoPs released important community-driven outputs including an early warning system for wheat blast; draft ground data standards to facilitate the machine learning analysis of satellite data; and a survey of CGIAR digital extension efforts. Convene culminates in the annual BIG DATA Convention, where 700 public, private, and non-profit actors assemble to examine digital models relevant for global collective action, and the CoPs organize technical meetings on livestock data, geospatial analysis, data-driven agronomy, socioeconomic data, crop modeling, and ontologies. The Platform showcased a **reference** Internet of Things architecture for agronomy and breeding; ground-breaking gender research methods that leverage telecom data; and technology standards for managing sensitive data.



INSPIRE

The Platform made four start-up grants to data-driven partnerships under its Inspire Challenge innovation process and four scale-up grants to winning projects from 2018 and 2017, awarding 1.025 million USD. Its growing portfolio, totaling 14 **projects** in 2019, demonstrates impressive early-stage results. For example, one advisory **service** that uses crowd-sourced smartphone images reached more than 33,000 Indian wheat farmers, increasing crop insurance efficacy and agricultural practices knowledge by 78%. A **project** tracing informal food flows by leveraging free Wi-Fi has collected data from over four million smartphones and has been adopted on a national scale in Vietnam as a tool to assess and predict COVID-19-related food security shocks. A near real-time small-scale fisheries monitoring **system** is being scaled to seven countries in Africa and Asia.

BIG DATA AT A GLANCE

INSPIRING INNOVATIONS



PARTNERSHIPS



BIG DATA
WORKS WITH
AN STRATEGIC
NETWORK OF
PARTNERS

DATA MANAGEMENT



155,000
PUBLICATIONS
23,000
DATASETS
DISCOVERABLE VIA GARDIAN



10
DATA MANAGEMENT
TOOLS / APPROACHES

COMMUNICATIONS



1.2 PLATFORM PROGRESS TOWARDS OUTPUTS AND OUTCOMES

1.2.1 OVERALL PLATFORM PROGRESS

Module One built fundamental technology and data standards of a CGIAR platform strategy. Module efforts in 2019 demonstrated CGIAR research outputs, created a compelling data and knowledge base through linking CGIAR and its partners' data repositories, and developed the standards and services for CGIAR and our development partners to manage data responsibly and make data output Findable, Accessible, Interoperable, and Reusable (FAIR). The team built on this data pool through efforts to **improve scientist capacity** to apply big data analysis in research and the development of local, timely solutions.

Module Two is expanding CGIAR's digital partner networks. CoPs increased their engagement and grew to more than 3,000 members. They released community-driven outputs including an early warning system for wheat blast; draft data standards for ground-truth data; a survey of CGIAR digital crop advisory efforts; and a standardized household survey methodology.

The BIG DATA **convention** attracted 700 attendees from 236 public, private, and non-profit organizations to discuss solutions for digitally-enabling global food security. CoPs curated tracks in livestock data, geospatial analysis, data-driven agronomy, socioeconomic data, crop modeling, and ontologies. The Platform's investments were featured. These included a reference IoT architecture for agricultural research; big data-enabled gender research; and technology standards for information systems that can be used to manage sensitive data. The Convention raised 70,000 USD in funding, a solid step towards establishing CGIAR as a voice of authority in digital agriculture.

Module Three increased its portfolio of Inspire Challenge projects to 14, generating 10 innovations and collecting more evidence of data-driven impact linked to CGIAR data and its partners. The Platform improved its Inspire Challenge submission and evaluation process to better target innovations. The 2019 process demonstrated a significant increases in proposals outlining incremental and disruptive innovations, use of unstructured data sources, and targeting under-represented groups.

1.2.2 PROGRESS BY MODULES

1

MODULE

ORGANIZE PROGRESS

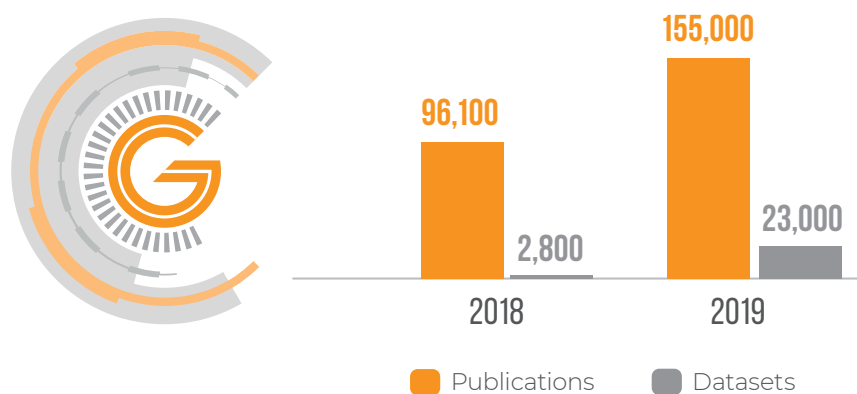
'Organize' builds the technology and data standards of a CGIAR platform strategy. Efforts in 2019 showcased CGIAR research outputs, created a compelling data and knowledge base through linking CGIAR and partner data repositories, and developed the standards and services to CGIAR and our development partners to manage its data responsibly and make outputs Findable, Accessible, Interoperable, and Reusable (FAIR). New institutional partners including development funding and finance organizations (USAID, DFID, World Bank), national agricultural research organizations (US Department of Agriculture's Ag Data Commons, the Indian Council for Agricultural Research) and partner governments (the Open Government Portal of India) linked their repositories to **GARDIAN**, building an important network of institutions joining

CGIAR in their commitment to digitize agricultural development through FAIR data. The team built on this data pool and improved scientist **capacity** to apply big data analysis in research and the development of local, timely solutions. The team launched the first version of “Collaborative GARDIAN (CG) Labs ” a secure analytic environment with tools and services to enable researchers to collaborate in finding and sharing GARDIAN (or other) data securely and to use R and Python-based scripts and other analytic approaches.

DETAILED ANNEX

GARDIAN added new functions in 2019. Highlights include the ability to map and spatially query production estimates for more than 30 crops; the visualization of a 7 Terabyte (TB) climate dataset; and the creation of an analytic workbench that enables CGIAR researchers to apply machine learning analytics over the growing GARDIAN data pool. GARDIAN now includes a service that helps scan research datasets and flags personally identifiable information before the data is made open. This minimizes the risks to both vulnerable individuals and to CGIAR's reputation. The team launched the first version of Collaborative GARDIAN (CG) Labs, a secure analytic environment with tools and services to enable researchers to collaborate in finding and sharing GARDIAN (or other) data securely and to use R and Python-based scripts and other analytic approaches. Researchers can now search for data via GARDIAN, import it into a shared analytic workspace in CG Labs, and find colleagues via the **Find a CGIAR Expert** feature to spark new collaborations.

GARDIAN NOW POINTS TO OVER 155,000 PUBLICATIONS AND 23,000 DATASETS FROM SEVERAL PARTNERS ALONGSIDE ALL CGIAR CENTERS



Organize released v. 1.0 of the Agronomy Field Information Management System (AgroFIMS) that employs these data quality standards to generate FAIR data at collection. Life sciences research is moving inexorably towards data annotation that leverages standard semantics and logic and, in 2019, BIG DATA made important contributions to these community **standards** for describing agronomic, socioeconomic, and survey data, and updating the CGIAR metadata standard. Module One is beginning to mutually reinforce the efforts of other Platform modules. GARDIAN **CG Labs**, for example, has been showcased in data science workshops and webinars developed with technical CoPs under Module Two.

Organize investments are linking the array of CGIAR data assets in new ways to facilitate new partnerships and innovation. In 2019, for example, the International Rice Research Institute (IRRI) updated its Rice Functional Genomics and Breeding (RFGB) database, which contains 3,000 rice genomes (3K-RG), to include new features, annotations,

DETAILED ANNEX

and data related to these genetic resources (see Wang *et al.*, 2020 and <http://www.rmbreeding.cn/snp3k>). In the new RFGB v. 2.0, new phenotypes and haplotypes allow associations to be inferred, enabling breeders and geneticists to narrow the number of candidate gene targets for validation. The new version of the database complements others that use 3K-RG data as a foundation, leverages partner contributions, and builds stronger partnerships with an array of organizations including the Shenzhen Institute of Breeding and Innovation for the Chinese Academy of Agricultural Sciences; Institute of Genetics and Developmental Biology Chinese Academy of Sciences; Nanjing AU; China AU; Beijing Genomics Institute Shenzhen; and Shanghai Jiao Tong University.

Organize engaged with CGIAR Centers and several external partners to achieve key milestones. With the Platform's support, a Bioversity International (BIOVERSITY)-based team worked to enhance agrisemantics standards and ontologies. The Environment Ontology, Food Ontology, SDGs Interface Ontology, and Planteome project teams, and WorldFish, for example, began development of a fish ontology. In 2019, almost all Centers implemented the CG Core Metadata Schema v. 2.0 and applied the AGROVOC controlled vocabulary in annotation. Several others also reported that they are using ontologies. Among them the International Center for Tropical Agriculture (CIAT)-BIOVERSITY International Alliance (the Alliance), International Institute of Tropical Agriculture (IITA), IRRI, International Water Management Institute (IWMI), Maize and Wheat Improvement Center (CIMMYT), and International Potato Center (CIP). IRRI enhanced the Rice Ontology, adapted its Farm Household Survey Database for machine-readability, developed a pilot database for unmanned aerial vehicle (UAV) images and metadata, and shared UAV-controlled vocabulary terms. IWMI improved metadata workflows by incorporating existing ontologies.

In 2019, the Organize Module tested data mining techniques to apply machine learning to GARDIAN's large data pool, worked with the University of Florida to auto-generate harmonized datasets for input to crop models and worked, also, with the University of California, Davis, to identify an approach for harmonizing key data variables and for applying machine learning and spatial approaches to GARDIAN data. To harmonize data at the collection point, the Module engaged with CIP, the Alliance, and International Food Policy Research Institute (IFPRI), soliciting researcher input from CIMMYT, IITA, Rothamsted Research, University of California, Davis (UC Davis), and the University of Florida to release AgroFIMS.

The Platform backstopped the adoption of data-related outputs, standards, and tools via guidance and webinars, facilitating Centers by offering a multi-module online course on best practices in open, FAIR, and ethical data assets as well as guidance, training sessions, and webinars. Several Centers -- including the Alliance, CIP, International Livestock Research Institute (ILRI), and IITA -- developed domain-specific materials on data management best practices, often involving researcher data champions. Through UC Davis regional workshops were offered at CIMMYT, CIP, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), and IFPRI on data science-related topics including advanced R, spatial predictive modeling, and machine learning. A data wrangling/processing guide was also developed. Facilitated by the Organize Module, Centers engaged researchers on data management through data sprints, data clinics, hackathons, "cure-a-thons," and other modalities. The Alliance held three training sessions on data management planning across CIAT regions, while BIOVERSITY-based staff participated in data sprints. ILRI organized helpdesks and data clinics. IRRI hosted a data sprint and offered four training sessions on basic research management. IITA launched a campaign on best practices for data quality. Also, the Center for International Forestry Research (CIFOR) embedded data teams into its projects; CIMMYT worked with stakeholders to promote good data management; and ICRISAT organized two hackathons and four data management training sessions.

2

MODULE

CONVENE PROGRESS

Platforms are multi-sided by definition; they are driven by communities. BIG DATA developed CGIAR digital partner networks through Module Two: Convene. Six BIG DATA CoPs increased their engagement and grew to 3,500 members. They released important community-driven output including an early warning system for wheat blast; draft ground data standards to facilitate the machine learning analysis of satellite data; a survey of CGIAR digitally-enabled crop advisory efforts; and a harmonized household survey methodology to bring more standardization to socioeconomic research across CGIAR. These outputs were shared and discussed widely through CoP-led [webinars](#) and workshops.

Convene activities apexed with the annual Big Data in Agriculture [Convention](#) at which 700 attendees from 236 public, private, and non-profit organizations examined digitally-enabled collective action in global food systems. CoPs curated technical tracks in livestock data, geospatial analysis, data-driven agronomy, socioeconomic data, crop modeling, and data ontologies. The Platform showcased investments overcoming barriers to the digitization of agricultural data including an Internet of Things architecture for agronomy and breeding; groundbreaking big-data enabled gender research; and information systems technology standards for managing sensitive data.

DETAILED ANNEX

The 2019 [Convention](#), hosted by ICRISAT, brought together 700 delegates, 65% of whom were external to CGIAR. These individuals were from 270 different organizations, institutions, and government bodies, as well as CGIAR centers and programs. The combined media and social media coverage reached an estimated 20 million people globally. The Convention raised nearly 70,000 USD in cash and in-kind support, a solid step towards the sustainability of an event that is establishing CGIAR as a voice of authority in digital agriculture worldwide.

The Convention provides a key means for the Platform to engage with youth. The [Youth In Data](#) workshop initiative entered its second year in 2019. Thirty young digital innovators from engineering and journalism schools local to Hyderabad and ICRISAT's doctoral program enrolled to be trained on data science and social media reporting. This was essential in enabling the Platform to reach audiences of several million through reporting on the Convention. Youth reporting also contributed to engagement with young, enthusiastic agricultural data scientists throughout the region and from around the world as well as with many young programmers and application developers through ICRISAT's Innovation Hub and through other regional and global entities.



2019 Inspire Challenge winners

The BIG DATA and GENDER Platforms spearheaded a novel approach to studying women’s economic empowerment in 2019. The partners conducted a phone-based survey of 10,000 respondents and used it to analyze billions of data points generated by mobile phone networks to predict gender and decision-making power among female farmers on a national scale in Uganda. The approach demonstrated the potential for observing changes in female farmers’ economic empowerment with greater speed and scale compared with solely survey-based methods.

BIG DATA launched two key new strategic partnerships designed to reveal emerging risks at the intersection of digital technologies and agro-ecologies, and to build CGIAR capacity to claim the benefits of these technologies in service of its mission. BIG DATA launched collaborative research with the Partnership on AI (PAI) and the University of Cambridge Center for the Study of Existential Risk (CSER) centered on identifying emerging priority topics for machine learning in agriculture --both the systemic risks and opportunities. The (French) National Institute for Research in Digital Science and Technology (INRIA) and BIG DATA began to design a program for integrating graduate students in computer science into CGIAR agricultural research for development. The program focused on diverse themes including machine-learning enhanced crop modeling, semantic data, digital architectures for agronomic research, and farmer decision making under uncertainty. BIG DATA and The (French) National Institute for Research in Digital Science and Technology (INRIA) will jointly build new and much needed links between computer science and the agricultural research domains.

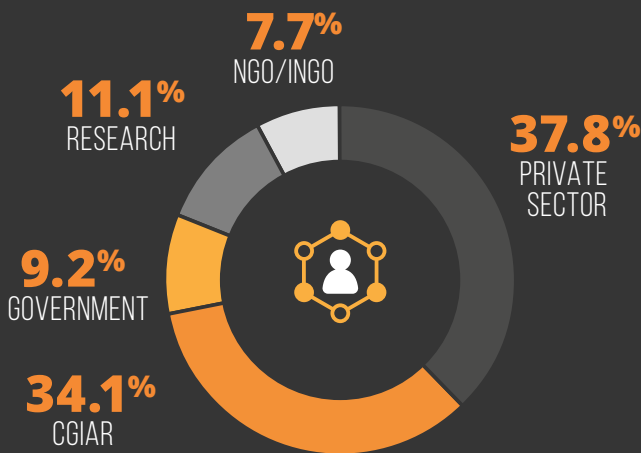
The Platform’s CoPs consolidated a strong communication strategy and significantly increased its member engagement in 2019. They published a combined total of 45 blog posts (up from five in 2018), 16 newsletters, eight webinars with more than 2,000 combined viewers, and grew the number of their combined membership to more than 3,000 individuals, with approximately 80% identified as external to CGIAR.

The Ontologies CoP made significant advancements in the development and adoption of crop ontologies in 2019. The progress was propelled by an [online curation tool](#) that makes the development and sharing of reference ontologies to describe agronomic phenotypes and the accurate mapping of these into databases easier. These critical steps enable comparative genotypic and phenotypic studies as well as gene-discovery experiments. The tool advanced with significant input from CGIAR Research Programs, the US National Science Foundation, allied research organizations, and private partners, such as PepsiCo.

2019 CONVENTION



ATTENDING SECTOR



2

MODULE

INSPIRE PROGRESS

In 2019, the Inspire Challenge grew its portfolio to 14 **projects**, generating 10 innovations this year and building the evidence base for digital agriculture. Solutions span the CGIAR portfolio, including the creation of a large-scale monitoring system for small scale fisheries; a tool using big data to reveal informal food flows; a computer vision-enabled crop disease diagnostic tool; the development of new methods for taking local rainfall measurements that leverages mobile phone networks; and machine-learning-driven seed selection.

The Platform **awarded** a combined total of 1.025 million USD to four start-up projects and to four previous Inspire winners to scale up their projects. The Inspire Challenge attracted 200,000 USD in external funding commitments in 2019, signaling it could well become a signature digital innovation process for CGIAR.

The Platform is improving its ability to source and foster innovation. In 2018 and 2019, more than 65% of applicants proposed to use machine learning on large unstructured data sources, up from 10% in 2017. Proposals targeting small producers have nearly doubled since 2017. In 2019, 69% of proposals were categorized as incremental and disruptive innovation, compared to 21% in 2017, promising more immediate and widespread impact in comparison to basic research. In 2019, 80% of proponents included a gender component in their proposals, an increase of 10% from 2018.

DETAILED ANNEX

In 2019, the Platform **awarded** four start-up grants of 100,000 USD each. Four winning projects from 2018 and 2017 that demonstrated exceptional results, proven viability, and the potential for impact were awarded a total of 525,000 USD in scale-up funds.

The eight winning teams included collaborations between a total of nine CGIAR Centers and Research Programs and 18 diverse external partners, including start-ups, governmental bodies, universities, and private sector businesses. The Inspire Challenge has begun to field test innovations that demonstrate new digital pathways to impact for CGIAR research, foster whole system agility and adaptation through digital tool use. These include:



PlantVillage Nuru is transforming farm-level pest and disease monitoring by using AI. The free-for-download phone application works offline and has proved twice as accurate in its diagnosis as extension workers. The app has been downloaded by users on all developing continents, generating more than 18,000 reports – which are generated each time the app is used correctly -- from users in more than 40 countries. The project is expanding from cassava diagnostics to the ability to diagnose additional crop diseases. An accurate model for potato diseases, for example, is expected to reach 200,000 farmers in India in 2020.



One project has put near-real-time data in the hands of fisheries officers, researchers, and local stakeholders by creating an **automated data pipeline** to highlight temporal and spatial changes in fish production. The dashboard has one of the most sophisticated data collection systems for small-scale fisheries in the world. In 2019, roughly 300 boat tracking units were deployed on fishing boats around Timor-Leste and 11 new local data collectors were trained in aforementioned coastal communities. The project is currently expanding to monitor small-scale fisheries in Egypt, Nigeria, and Zambia, with plans to scale up in Bangladesh, Malaysia, and Malawi. The existing data systems underpinning this project follow CGIAR gender-disaggregated data standards and have already highlighted the critical importance of women in fishery activities as dependable sources of household food and nutrition security in Timor-Leste.



Another project is monitoring and **informal food flows through free Wi-Fi** in and among five traditional markets in Hanoi, Vietnam. In 2019, the team collected data from more than five million smartphone devices to analyze and leverage it for improving market policies and food safety. Their efforts to gender disaggregate the data has revealed 56% of their users are women. The project has been adopted by the General Statistics Office of Vietnam as a key method for assessing and predicting unfolding COVID-19-related food security shocks on a national level.



Another project is estimating rainfall in developing nations through the unprecedented **use of commercial microwave links (CMLs)**. The team's peer-reviewed research has demonstrated the potential spatial advantage of using CMLs for monitoring rainfall, which is based on measurements in an agricultural area near Kericho, Kenya. CMLs are proving to be a sustainable, long-term solution for generating rainfall estimates in developing countries -- a critical component for improving crop yield monitoring and designing better rainfall-based index insurance.

Other notable emergent innovations from the 2019 Inspire portfolio include:

- ☑ **CubicA**, which piloted a digital-advisory system that reached 3,000 smallholder banana farmers in Uganda.
- ☑ **Smart Seed Selection** successfully developed a machine-learning-driven process for the selection of the best maize seeds for expected productivity in specific environmental conditions.
- ☑ **Seeing is Believing** improved a method for remote monitoring crop phenology that leverages smartphone images and links these to yield predictions.

The Challenge attracted some 200,000 USD in external funding commitments in 2019, signalling it could well become a signature digital innovation process for CGIAR.

INSPIRE 2019

8
WINNERS



4
SCALE-UP
AWARDS



4
INSPIRE
CHALLENGE
AWARDS

1.25
MILLION
TOTAL
AWARDED



150
APPLICATIONS



STRATEGIC WINNING
PARTNERSHIPS:
9 CGIAR CENTERS
& RESEARCH PROGRAMS
PARTNERED WITH

18 DIVERSE
EXTERNAL
PARTNERS



1.2.3 VARIANCE FROM PLANNED PROGRAM FOR THIS YEAR

A. Have any promising research areas been significantly expanded?

The Platform continues to follow the original proposal closely, although with a more explicit focus on the development of both Center and pan-CGIAR digital strategies than originally conceived. This has led to an expanded role for more explicit organizational and strategy development research that looks at strategic trends, organizational capabilities, and building a more unified pan-CGIAR information vision.

B. Have any research lines been dropped or significantly cut back?

The Platform continues to follow the original proposal closely, although with a more explicit focus on the development of both Center and pan-CGIAR digital strategies than originally conceived.

C. Have any modules or specific research areas changed direction?

No modules have changed direction. The Platform for Big Data agreed to serve as a ready Window Two contracting mechanism for some donor funds earmarked specifically for impact assessment.

1.2.4 ALTMETRIC AND PUBLICATION HIGHLIGHTS

BIG DATA does not prioritize producing research publications, but rather serves to support the centers and programs in their research, which is reported through CRPs.

Nonetheless, the CoPs and Inspire Challenge teams published several publications for the Platform in 2019, recording a combined altmetric of 513. Most notably, the Data-driven Agronomy CoP published a peer-reviewed [paper](#), in collaboration with the Colombian Ministry of Agriculture and Rural Development – “A scalable scheme to implement data-driven agriculture for small-scale farmers” -- that achieved an altmetric score of 106. One Inspire Challenge [team](#) published a [dataset](#) and [paper](#) detailing how remote monitoring of crop phenology that leverages smartphone images can improve crop modeling and insurance for small-holder farmers. Another [project’s](#) paper – “MARPLE, a point-of-care, strain-level disease diagnostics and surveillance tool for complex fungal pathogens” -- recorded an altmetric score of 88, was mentioned by four news outlets, and tweeted about 90 times. The Ontology CoP collaborated with the INRAe Unit (France) on a [paper](#) about “Applying FAIR Principles to plant phenotypic data management in GnpIS” published in the journal *Plant Phenomics*. Through their mini-grant programs, the Geospatial CoP contributed -- in collaboration with the University of Twente -- on “A suite of global accessibility indicators,” which was [published](#) in the *Nature journal Scientific Data*, and the Crop Modeling CoP participated in a paper on “Different uncertainty distribution between high and low latitudes in modeling warming impacts on wheat,” which was [published](#) in *Nature Food*.

The Platform showed success in building its brand and establishing CGIAR as an authority in the digital agriculture space. A Google Analytics analysis of the Platform’s social media platforms and website showed significant audience growth. In 2019, the Platform published approximately 150 pages, blog posts, and videos, 13 [webinars](#), and 54 newsletters that recorded an open-rate of 5% greater than the industry average. The website showed a 68% increase in user traffic (44.7K) and a 48% increase in unique page views (116.8K). Our social media audience (Facebook, LinkedIn, Instagram, YouTube, and Twitter) grew 78% (10.5K), which, when combined with our website audience, shows a total audience growth of 68% (55.3K).

In 2019, the Platform was featured in numerous global news outlets, such as The Economist, Reuters (Foundation), Telangana Today, Hans India, and Food Tank, recording a total potential reach of 139 million (Meltwater). The most notable media result of 2019 occurred during the annual Convention, when the combined media and social media coverage reached an estimated 20 million people globally.

1.3 CROSS-CUTTING DIMENSIONS (AT PLATFORM LEVEL)

1.3.1 GENDER

A. List any important PTF research findings

Collaborative research across BIG DATA Modules revealed that the data annotation standards for gender research data need attention if CGIAR gender-relevant data is to be discovered and used in new research. Some preliminary findings have informed the recommendations that have been made to the CGIAR Metadata Working Group to update metadata annotations. These would enhance discovery as well as provide users with the ability to reuse gender-disaggregated data. This will be essential for advancing all forms of gender research across the System and for unlocking new big data-enabled methods that can be used for researching and advancing gender equality. As a result, the annotation, discovery, and re-use of gender disaggregated data has become a pillar of the Platform's data strategy and will inform future collaboration with the Gender Platform.

Gender research is nuanced and multidisciplinary. For this reason, BIG DATA decided that an open innovation process might not yield the quality of research innovation needed, so we decided to co-design research with the Gender Platform. This approach was validated in 2019. During the 2019 management team retreat, the team recognized the need to formalize an official gender strategy for the Platform. The team nominated a gender focal point and decided upon specific action items to explicitly target in 2020, agreeing that the first be the creation of an official Gender Strategy document.

Since 2018, the Platform has made effective progress in including a gender dimension in our Inspire innovation challenge using a rubric and scoring matrix to explicitly assess whether and how proposals dealt with gender issues. These efforts manifested in 2019 with 80% of Inspire Challenge participants including a gender component, up from 70% in 2018.

While gender review has been effective in detecting and averting potential biases in digital innovation projects supported by this module, through collaborative research with the Gender Platform we have identified further opportunities to build upon and use the Challenge process as a source of insight and positive action for mainstreaming gender equality. In 2019, the team decided to modify the digital innovation challenge process to include key points related to gender mainstreaming. Specifically, these included requesting a gender balance of proposal teams and detailing a gender equality mainstreaming hypothesis for proposed projects. We anticipate that this will result in positive action for mainstreaming gender equality and will highlight the role of digital innovation in achieving gender parity in the agricultural research space.

B. What have you learned? What are you doing differently?

Through collaborative research with the Gender Platform, we learned that some effort is required to help translate between the gender and big data research disciplines. Such translation promotes rich discussion, presents new opportunities for collaboration, and reinforces the need for research co-design. We found that gender review, while helpful in averting biases, was not

sufficient to mainstream gender equality throughout our programs. Collaborative research across the Organize and Convene modules revealed that the data annotation standards for gender research data need attention if CGIAR gender-relevant data is to be discovered and used in new research.

We decided that we could more intentionally use our innovation process to test hypotheses about mainstreaming gender equality through requesting that teams have at least one participating female and that they state their hypotheses about gender mainstreaming. We hope that this will yield insight into how to best leverage big data research methods to ensure gender equality.

C. Have any problems arisen in relation to gender issues or integrating gender into the PTF's research?

We found that gender review, while helpful in averting biases, was not sufficient to mainstream the gender equality concept through our programs. We decided that we could more intentionally use our innovation process to test hypotheses about mainstreaming gender equality through requesting that teams have at least one participating female and that they state their hypotheses about gender mainstreaming. We hope that this will yield insight into how best to leverage big data research methods to ensure gender equality.

1.3.2 YOUTH AND OTHER ASPECTS OF SOCIAL INCLUSION / "LEAVING NO-ONE BEHIND"

GARDIAN currently provides access to about 2,500 publications and 380 datasets addressing youth. Of these 380 datasets, most are from the World Bank's Living Standards Measurements Survey (LSMS), as might be expected; however, there are 40 datasets from CGIAR Centers with a youth component, up from just three in 2018. While the youth dimension of CGIAR's research could be improved, along with implementing tagging protocols to enhance data discoverability, this still represents an increase of roughly 10-fold from the previous year. The Platform will continue to assess ways to better promote the inclusion of youth-related data standards. The Platform released a new tool in 2019 for harmonizing well-annotated socioeconomic research surveys, called 100Q, which includes key youth-related indicators.



The Platform built upon its Youth In Data **initiative**, which was created in 2018 to engage with young digital innovators in a digital agriculture workshop during that year's Convention in Nairobi. This model proved very successful and was expanded in 2019 with the launch of the Youth in Data Connect **platform**. We received more than 80 applications for the 2019 workshop, in which 30 youth were enrolled to learn about the basics of big data-applied digital agriculture and how to report on it using digital media tools. The 2019 Youth in Data group consisted of students from engineering and journalism universities local to Hyderabad, as well as doctoral scholars from the ICRISAT campus. These young digital innovators interacted with and **interviewed** experts and participants at the Convention and were key in the Platform's success in reaching a combined potential audience of 20 million across its social media platforms and through global media coverage during the event.

As in previous years, the Convention attracted young, enthusiastic agricultural data scientists in the region and from around the world. This also included many young programmers and application developers from ICRISAT's Innovation Hub as well as from other regional and global entities.

The launch of the new **Youth in Data Connect** platform, a database that compiles global youth-focused or youth-driven digital agriculture initiatives, is the BIG DATA Platform's first step towards mapping the global landscape of youth engaged in this innovation space. The objective is to connect these young innovators with industry leaders and experts and to build informed infrastructure that will support and engage young people in digital agriculture.

We have found that the user base of digital innovations supported by the Platform tend to skew young, but that the most notable contributions and engagement have flowed from engaging youth in data science and digital agriculture reporting in collaboration with our communications team. We aim to make this a standard feature of our communications protocol going forward. We think it highlights the need for more targeted youth engagement in other aspects of the program.

A. List any important PTF research findings

Platforms are generally not involved in research; they have an organizational development role. That said, we found that involving youth in reporting on data science and digital agriculture strengthened our online engagement and built new digital connections across digital agriculture startups in East Africa and India (the two places where the Youth In Data initiative was leveraged in support of the BIG DATA Convention). More detail is provided in the Module Two section.

B. What have you learned? What are you doing differently?

Building on our experience in involving youth in reporting on data science and digital agriculture, we will develop this as a standard approach throughout the BIG DATA program.

C. Have any problems arisen in relation to youth issues or integrating youth into the PTF's research?

Youth inclusion was conceived from the outset, rightly pointing out that the user base of digital innovation tends to skew young. That said, we found that the more explicit targeting of youth is an effective approach.

1.3.3 CAPACITY DEVELOPMENT

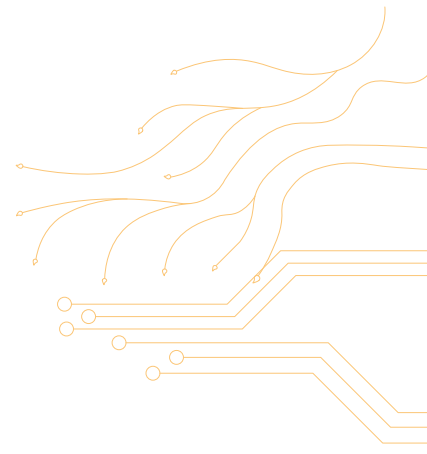
The Platform's engagement in capacity building grew significantly in 2019 through six active CoPs, a combination of online and in-person events, development of new self-paced online training content, and by co-funding a computer science PhD candidate working on digitally-enabled decision support systems for farmers. In total, the Platform reached more than 4,000 researchers with these capacity building efforts in 2019.

Most Centers organized at least one hands-on “data sprint” workshop to promote best practices in managing and uploading data assets towards open, FAIR, and ethical outcomes. The 2019 sprints resulted in the creation of more than 750 annotated datasets that were uploaded to repositories across the Centers. A webinar series provided further support for these efforts and harmonized key messages, culminating in the development of a new five-module, self-paced online course outlining good practices in creating FAIR data and of best practices for ethical data management throughout the research lifecycle. In addition, the Platform conducted user workshops on key tools being developed for field trial information management and data annotation. The Platform also supported CGIAR’s efforts in the digital agriculture space in other ways as well, through year-round, multi-day regional workshops at IFPRI, CIP, ICRISAT, and CIMMYT, to synergize cross-Center collaboration and address topics spanning advanced R, spatial predictive modeling, and machine learning.

The Inspire Challenge, CGIAR’s signature digital innovation process, received 150 applications from CGIAR/partner teams. All of these received feedback on how to more effectively link their ideas for digital research from innovation to impact. Ten Inspire Challenge finalist teams were brought to the Big Data in Agriculture Convention, where they interacted with each other as well as with subject matter experts and potential investors to get a better idea on how to match their products with desired impact. All 10 received coaching on their presentations and feedback from an external expert panel.

1.3.4 CLIMATE CHANGE

Under Module One, the Platform equipped CGIAR with new standards-driven tools for climate change research, leveraging a common pool of interpretable and interoperable data assets in a common analytic environment -- which features, among other key data assets, a 7 Terabyte global climate dataset --, and held multiple Center workshops on key related data science skills including advanced R, spatial predictive modeling, and machine learning. 2019 saw the development of several key tools and datasets for examining climate impact on agriculture under Module Two, including an updated weather dataset for modeling, updates to the Spatial Production Allocation Model (a key, widely used dataset), and several research publications centered on managing climate impacts on specific crops and regions. The Inspire Challenge sourced and developed a new method for more precise and localized rainfall measurement than either stations or the Climate Hazards Group InfraRed Precipitation with Station data dataset offer. It also uses mobile telephone network data to make forecasts.



EFFECTIVENESS AND EFFICIENCY

2.1 MANAGEMENT AND GOVERNANCE

The Platform has an internal/external Steering Committee that consists of four CGIAR representatives -- permanent seats are allocated to IFPRI and CIAT as lead Centers, the remaining seats are given to another Center representative and a CRP representative -- and three external representatives (research, funder, private sector). By the end of 2019, two external members rotated out, one extended their term for one year, and a new CRP representative was instated in late 2019.

2.2 PARTNERSHIPS

2.2.1. HIGHLIGHTS OF EXTERNAL PARTNERSHIPS

The Big Data Platform CoPs are open partnership networks that grew to more than 3,500 CGIAR and non-CGIAR members in 2019.

In 2019, BIG DATA contributed to an event produced by the Partnership on AI (PAI), a membership organization linking global IT firms, non-profits, and research organizations working on equitable use of artificial intelligence. This precipitated a partnership development visit to Silicon Valley and new collaborations with Stanford University on evidence for digital agriculture; Facebook Research on using computer vision for crop disease diagnosis and building data-driven insights into gender equality; Google Earth Engine for development of a shared service; and Google X on improving data analytics methods for linking phenotypic data and environmental analysis.



Through The Partnership on AI, BIG DATA also launched new collaborative research with the Cambridge University Center for the Study of Existential Risk, specifically on emerging topics and the potential risk factors associated with the rising use of machine learning in the agricultural sector.

The Inspire Challenge once again catalyzed new partnerships, attracting expressions of interest from 191 external partners. The eight winning teams created partnerships between nine CGIAR Centers and associated Research Programs as well as 18 diverse external partners from start-ups, governmental bodies, universities, and private sector businesses.

The Annual Big Data in Agriculture Convention once again proved to be an important partnership development mechanism, with 65% of its 700 attendees external to CGIAR. These included individuals from government, non-profit, digital startups, agribusiness, and other organizations.

2.2.2. CROSS-CGIAR PARTNERSHIPS

In 2019, the Platform initiated a cross-platform design process for building a unified pan-CGIAR information architecture vision. Excellence in Breeding, BIG DATA, and GENE BANKS with subject matter experts in breeding, genomics, scaling, agro-ecology, large-scale agronomy, animal science, and fisheries developed a unified view on key research processes and the necessary data, digital tools, and stakeholders that would be needed to support each process. This fundamental design work is being folded into BIG DATA inputs to One CGIAR.

BIG DATA and the Gender Platform conducted collaborative research on big data-enabling research into women's economic empowerment.

BIG DATA and Excellence in Breeding reinforced each organization's efforts to develop standard operating procedures and analytic pipelines for remote sensing imagery.

2.3. INTELLECTUAL ASSETS

Have any intellectual assets been strategically managed by the PTF (together with the relevant Center) this year?

The Inspire Challenge digital innovation process is designed to facilitate meaningful collaboration between CGIAR and any type of partner. These partners agree that they will retain their original intellectual property, but that the joint projects will be used in accordance with the desired agricultural development outcomes. In 2019, we put this into action in an agreement for a project in which a partner brought a proprietary algorithm to the project and we effectively managed the IP. We are finding that this approach is helpful not only for accelerating impacts, but also for transferring digital technologies from one subject area or application into agriculture. A nanopore sequencing tool used in responding to the Ebola virus outbreak, for example, is now being used at scale in wheat rust surveillance.

Indicate any published patents and/or plant variety right applications (or equivalent)

N/A

List any critical issues or challenges encountered in the management of intellectual assets

None

2.4 MONITORING, EVALUATION, IMPACT ASSESSMENT AND LEARNING (MELIA)

In 2019, BIG DATA increased its efforts to “leave no evidence behind” on the relevance of digital technologies to agricultural development. This is reflected in more granular reporting on MARLO and in the launch of the Digital Food Systems [Evidence Clearinghouse](#), where we actively seek credible, measured evidence of “value add” digital interventions somewhere in a food system. The Platform CoPs are leveraged as expert networks for sourcing and validating this evidence. Both formal studies conducted by the Platform’s awardees and wider evidence gathered by the sector inform both how we learn about our interventions and our view on how best to target our digital innovation strategy.

2.5 EFFICIENCY

No new efficiencies were noted in 2019, although there was, perhaps, an increase in efficacy. The Platform raised 70,000 USD specifically through the Big Data in Agriculture Convention in 2019 and attracted commitments of 200,000 USD in external funding for the Inspire Challenge.

2.6 MANAGEMENT OF RISKS TO YOUR PLATFORM

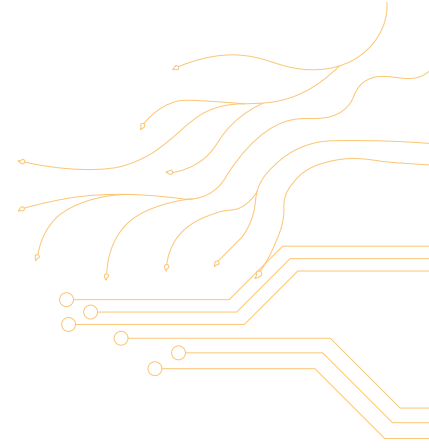
The Platform implemented a software [tool](#) for helping CGIAR data managers to detect personally identifiable information that may have inadvertently been shared on open data repositories, averting significant regulatory risk (e.g. the General Data Protection Regulation) and reputational risk (e.g. causing CGIAR to be known as an institution that violated someone’s privacy).

2.7 USE OF W1-2 FUNDING

Platforms do not typically conduct research; rather, they serve in an organizational development internal service delivery role. In 2019, the Platform offered to cover the indirect costs of those funding proposals submitted by CGIAR Research Programs in which the donor had placed a cap on admissible overhead expenses, but for which the technical substance of the project would be an opportunity to build CGIAR’s big data capabilities. Five funding proposals were submitted to donors with this promise of cost-sharing from BIG DATA; one of these proposals was selected for an award, and BIG DATA transferred this cost-share directly to the Program/Center.

Other interesting funding developments in 2019: we raised roughly 70,000 USD for production of the Big Data in Agriculture Convention, a signal that we can build this into an event of reference in the digital agriculture space; and around 200,000 USD in funding commitments for the Inspire Challenge, a signal that we can build this into a signature CGIAR digital innovation process through many more funding cycles.

03



FINANCIAL SUMMARY

BIG DATA continued to closely follow the budget as outlined in the initial proposal with a few notable differences. In 2019, BIG DATA offered cost-sharing to those CRPs applying for funding that have a significant big data component, specifically to cover any indirect costs. Module Two, through the Big Data in Agriculture Convention, raised roughly 70,000 USD in cash and in-kind sponsorship for the event, although some of this funding will only enter BIG DATA accounts in 2020. Module Three typically has the highest carryover among the modules due to the timing of a grant process the Platform runs, in which funding commitments are made at the end of the calendar year.



LIST OF TABLES

Table 2: Condensed list of policy contributions in this reporting year (Sphere of Influence)	26
Table 3: List of Outcome/ Impact Case Reports from this reporting year (Sphere of Influence)	26
Table 4: Condensed list of innovations by stage for this reporting year	27
Table 5: Summary of status of Planned Outcomes and Milestones (Sphere of Influence-Control)	29
Table 6: Numbers of peer-reviewed publications from current reporting period	42
Table 7: Participants in CapDev Activities	42
Table 8: Key external partnerships	43
Table 9: Internal Cross-CGIAR Collaborations	44
Table 10: Monitoring, Evaluation, Learning and Impact Assessment (MELIA)	47
Table 11: Update on Actions Taken in Response to Relevant Evaluations	47
Table 12: Examples of W1/2 Use in this reporting period (2019)	47
Table 13: Platform Financial Report	48
Table Annexes	49

TABLE 2: CONDENSED LIST OF POLICY CONTRIBUTIONS IN THIS REPORTING YEAR (SPHERE OF INFLUENCE)

Title of policy, legal instrument, investment or curriculum to which CGIAR contributed	Description of policy, legal instrument, investment or curriculum to which CGIAR contributed	Level of Maturity	Link to sub-IDOs	CGIAR cross-cutting marker score				Link to OICR (obligatory if Level of Maturity is 2 or 3) or link to evidence (e.g. PDF generated from MIS)
				Gender	Youth	Capdev	Climate Change	
314 - National Fisheries Strategy of Timor-Leste	Responsible, sustainable and equitable management of fisheries and living aquatic resources based on a co-management arrangement and complying with regional and international laws, voluntary instruments, agreements and covenants.	Level 1	<ul style="list-style-type: none"> Enhanced conservation of habitats and resources Enhanced individual capacity in partner research organizations through training and exchange Enhanced institutional capacity of partner research organizations 	1- Significant	1- Significant	0 - Not Targeted	1- Significant	OICR2916

TABLE 3: LIST OF OUTCOME/ IMPACT CASE REPORTS FROM THIS REPORTING YEAR (SPHERE OF INFLUENCE)

Title of Outcome/ Impact Case Report (OICR)	Link to full OICR	Maturity level	Status
OICR2916 - National adoption of digital fisheries monitoring system enables science driven development prioritisation	Link	Level 2	New Outcome/Impact Case
OICR3181 - Rice Functional and Genomic Breeding (RFGB) v2.0 database to link genotypic and phenotypic datasets, facilitates access and uploading of genomic data.	Link	Level 1	New Outcome/Impact Case

TABLE 4: CONDENSED LIST OF INNOVATIONS BY STAGE FOR THIS REPORTING YEAR

Title of innovation with link	Innovation Type	Stage of innovation	Geographic scope (with location)
741 - Infrastructure mapping and strategy engagements	Research and Communication Methodologies and Tools	Stage 1: discovery/proof of concept (PC - end of research phase)	Global
1095 - Agronomy Ontology	Research and Communication Methodologies and Tools	Stage 3: available/ ready for uptake (AV)	Global
1101 - Extension of SDG Interface Ontology with the concepts of the CGIAR Strategic Research Framework	Other	Stage 1: discovery/proof of concept (PC - end of research phase)	Global
1102 - Socio!: Agricultural Household Survey Ontology	Other	Stage 1: discovery/proof of concept (PC - end of research phase)	Global
1157 - Best practices for FAIR, ethical, and open data assets	Research and Communication Methodologies and Tools	Stage 1: discovery/proof of concept (PC - end of research phase)	Global
1158 - CGIAR Expert Finder	Research and Communication Methodologies and Tools	Stage 1: discovery/proof of concept (PC - end of research phase)	Global
1159 - Agronomy Field Information Management System (AgroFIMS)	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase)	Global
1160 - Case studies to illustrate data wrangling approaches and share R scripts	Research and Communication Methodologies and Tools	Stage 3: available/ ready for uptake (AV)	Regional, Western Africa
1161 - GARDIAN Labs, a service to enable researchers to collaborate and share data and analytic approaches and tools	Research and Communication Methodologies and Tools	Stage 3: available/ ready for uptake (AV)	Global
1162 - GARDIAN PII checker to minimize risk of personally-identifiable information in datasets	Research and Communication Methodologies and Tools	Stage 3: available/ ready for uptake (AV)	Global
1170 - Rice Functional and Genomic Breeding (RFGB) v2.0 database to link genotypic and phenotypic datasets, facilitates access and uploading of genomic data	Biophysical Research	Stage 4: uptake by next user (USE)	Global
1172 - Controlled vocabulary dictionary for UAV, phenotyping and image data annotation	Biophysical Research	Stage 2: successful piloting (PIL - end of piloting phase)	Global
1173 - Pilot database for management of drone (UAV) image data with basic functionality was developed.	Biophysical Research	Stage 2: successful piloting (PIL - end of piloting phase)	Global

Title of innovation with link	Innovation Type	Stage of innovation	Geographic scope (with location)
1259 - Gap analysis of crop modeling activities within CGIAR and collaborators	Other	Stage 1: discovery/proof of concept (PC - end of research phase)	Global
1262 - Early Warning System for Wheat Blast	Production systems and Management practices	Stage 1: discovery/proof of concept (PC - end of research phase)	National, Bangladesh
1289 - Big data technology and system to improve small-scale fisheries management	Research and Communication Methodologies and Tools	Stage 4: uptake by next user (USE)	National, Timor
1305 - Improved monitoring of crop phenology with smartphone images	Social Science	Stage 2: successful piloting (PIL - end of piloting phase)	Sub-national, India
1336 - Predicting yields using crop simulation models and gridded weather data	Production systems and Management practices	Stage 1: discovery/proof of concept (PC - end of research phase)	National, India
1357 - CubicA - The right information, at the right time, to the right farmer	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase)	National, Uganda
1358 - Farm.ink Digital Learning Platform	Research and Communication Methodologies and Tools	Stage 1: discovery/proof of concept (PC - end of research phase)	Regional, Eastern Africa
1360 - MARPLE (Mobile And Real-time PLant disEase) diagnostics. Rapid and accurate point-of-care (PoC) diagnostics for wheat yellow rust	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase)	National, Ethiopia
1362 - use of CMLs in estimating rainfall in Kenya	Research and Communication Methodologies and Tools	Stage 1: discovery/proof of concept (PC - end of research phase)	National, Kenya
1366 - Use of free Wi-Fi to better understand food flow in markets	Research and Communication Methodologies and Tools	Stage 1: discovery/proof of concept (PC - end of research phase)	National, Vietnam
1423 - Deep Learning for Image-Based Cassava Disease Detection	Production systems and Management practices	Stage 2: successful piloting (PIL - end of piloting phase)	Regional, Sub-Saharan Africa
1449 - Groundwater Game	Production systems and Management practices	Stage 1: discovery/proof of concept (PC - end of research phase)	National, India
1460 - Develop machine learning models that predict the performance of seed varieties in particular conditions	Production systems and Management practices	Stage 1: discovery/proof of concept (PC - end of research phase)	National, Mexico
1461 - Revealing Gender in Food Systems Through Big Data	Social Science	Stage 1: discovery/proof of concept (PC - end of research phase)	National, Uganda

TABLE 5: SUMMARY OF STATUS OF PLANNED OUTCOMES AND MILESTONES (SPHERE OF INFLUENCE-CONTROL)

Module	Module Outcomes 2022	Sub-IDOs	Summary narrative on progress against each Module outcome this year.	Milestone	2019 milestones status	Provide evidence for completed milestones (refer back to means of verification, and link to evidence wherever possible) or explanation for extended, cancelled or changed.
M1	M1 Outcome: 1.1. A demand-driven analytics environment is available.	Enhanced institutional capacity of partner research organizations	The four key milestones for this outcome focused on tools, approaches, and capacity enhancement leading to the development of back end tools and services to leverage data for analysis and visualization. Progress on all milestones was strong. A prototype data analytics environment, CG Labs, was developed and coupled with Globus software to allow the sharing of sensitive data. CG Labs enables researcher collaboration and access via a Single Sign-On (SSO); a search of GARDIAN data; dataset download to CG Labs; secure data upload and sharing via Globus; and collaborative development of analytics and crop models in Python and R. Jupyter notebook was implemented as a machine learning platform as part of CG Labs, and a collaboration with UC Davis resulted in the development of tools for	2019 - 1.1.1. Prototype data analytics environment developed and tested.w	Complete	A prototype data analytics environment, Collaborative GARDIAN Labs (CG Labs, was developed and made available in 2019. CG Labs includes an implementation of Globus software that allows the sharing Wand transfer of sensitive and non-sensitive data. CG Labs enables researcher collaboration via Slack-like functionality, and once a user is logged in via an SSO, the user can perform a search of GARDIAN data; download datasets to CG Labs; securely upload and share data via Globus; and collaborate on writing and running analytics and crop models in Python and R. The R-WOFOST model is currently implemented and the DSSAT model will soon be available. CG Labs is currently being tested by researchers at CGIAR and other institutions that require several TBs, large CPUs, and uninterrupted, long-timeframe data processing.
				2019 - 1.1.3. At least 2 case studies using the analytical environment developed and presented.	Complete	Two use cases were developed to design and test data interoperability and analytics pipelines. The first involves the development of model-ready data from GARDIAN for crop and new trait prediction models and required extension of data translation tools from the Agriculture Model Intercomparison and Improvement Project (AgMIP) for use in GARDIAN so that selected datasets can be translated to the AgMIP format. Development of standard unit conversion utilities was expanded to include all units found in a subset of GARDIAN datasets used for testing. Existing AgMIP data translation tools were extended to formats encountered in sample GARDIAN datasets, resulting in the release of an updated data translator to generate harmonized AgMIP-format files from GARDIAN data. A third use case focused on developing methods to enhance the ability to derive insights from large data pools, improving our ability to derive insight from data; for example, understanding the drivers of variation in crop response to fertilizer use in Sub-Saharan Africa. Library for unit alignment across ICASA and GARDIAN datasets available here .

MI			<p>finding, cleaning, and processing data to enable machine learning applications. Three use cases were developed to design analytics pipelines involving development of model-ready data from GARDIAN for crop and new trait prediction models and facilitating the aggregation of datasets for analysis. Collaborators at the University of Florida's Agricultural Model Intercomparison and Improvement Project (AgMIP) successfully tested an API to develop DSSAT model-ready datasets from GARDIAN. Several data science workshops and webinars provided backstopping for these efforts.</p>	<p>2019 - 1.1.4. Cloud-based machine learning platforms configured and made accessible to CGIAR scientists.</p>	<p>Complete</p>	<p>The ability to apply machine learning techniques depends on the availability of large data pools on which these scripts can operate. Thus, one of the Organize Module's activities centered on creating workflows to download, standardize, and aggregate research data for use and the discovery of new insights through machine learning. The UC Davis collaboration resulted in R scripts that allow data compiling and processing, as well as the aggregation of large numbers of standardized datasets on which machines can operate. Further, in recognition of the need for a central environment through which to access and deploy these techniques, Jupyter Labs (available upon logging in to CG Labs, via "Analyze" feature) was customized and implemented as a machine learning platform seamlessly integrated with GARDIAN CG Labs. IRRI's Farm Household Survey Database was adapted to be fully machine readable, with more than 100 surveys annotated and further curation continuing in an Amazon Web Services S3 bucket as the Socioeconomic Ontology matures and can be used for annotation. These will be shared for a machine learning annotation project through the Convene Module's Socioeconomic Data CoP.</p>
				<p>2019 - 1.1.5. Datasets in GARDIAN organized and visualized using the Sustainable Intensification Framework.</p>	<p>Complete</p>	<p>Collaborators at the University of Florida's AgMIP project developed minimum data criteria requirements for the DSSAT model, mapped the AgMIP data dictionary (ICASA variables) to the Agronomy Ontology (AgrO), and successfully tested an API to identify GARDIAN datasets that meet minimum data criteria and with data variables that could be mapped to AgrO. AgMIP translation tools were further extended to read the datasets found in GARDIAN, and to allow users to download data in these formats for easy integration into DSSAT. A collaboration with UC Davis researchers resulted in the articulation of a viable use case that would power the development of an R library for processing and cleaning data to enable the aggregation of a large number of datasets from GARDIAN.</p>

<p>MI</p>	<p>M1 Outcome: 1.2. CGIAR resources are discoverable and reused.</p>	<p>Enhanced institutional capacity of partner research organizations</p>	<p>The Organize Module launched GARDIAN in 2019, developed on the 2018 pan-CGIAR CeRES portal. GARDIAN facilitates the discovery of CGIAR data and publications with tools for data visualization and back-end services to accommodate data privacy, ethics, and clarity on licensing. As a global knowledge base for agriculture, it enables discovery of more than 155,000 publications and 23,000 datasets, including those of all CGIAR Centers and their research partners (United States Agency for International Development, UK Aid Direct, the World Bank and others). With Organize Module support, almost 1,000 datasets were curated and uploaded to Center repositories, much more than the minimum 150 dataset target, and these are discoverable via GARDIAN. Although the number of uploads per Center varied, 13 Centers reported workshops were held to help researchers standardize data and metadata through annotation that met accepted standards prior to upload. An API-key and support for harvesting GARDIAN FAIR scores was provided to MARLO, which makes about 70 API calls per month. GARDIAN is more than a data discovery portal -- it is a data ecosystem that also enables reuse of CGIAR data via interactive visualizations (e.g. crop production estimates, 7 TB CMIP5 climate dataset), a collaborative analytics platform, and seamless model pipelines, all backed by secure data transfer.</p>	<p>2019 - 1.2.5. GARDIAN officially launched with a PII detector, a data upload functionality for partners without repositories, indexing external data from at least two partners' repositories.</p>	<p>Complete</p>	<p>GARDIAN released a containerized, easily deployable PII detection service for Centers as one part of a FAIRification workflow to help CGIAR and its partner institutions. This is a step to minimize risks for the CGIAR system related to unintentional disclosure of PII, and was tested by the Alliance Bioversity-CIAT, ICRI-SAT, and IITA. Other parts of the FAIR workflow include the ability to easily annotate data with semantic standards and the CG Core metadata, and seamless upload to institutional repositories. These pieces of the workflow have been developed and tested in 2019, but required refinement and will be released in early 2020. The GARDIAN team provided technical guidance and/or APIs to make Center/CRP data assets discoverable through GARDIAN to several Centers (e.g. IRRRI, ICRISAT, IWMI) and worked with others to test and refine the PII service. In 2019, GARDIAN indexed assets from five large repositories in 2019, including those of the World Bank, USAID, DFID, USDA, and the Government of India, and now indexes about 155,000 publications and 23,000 datasets.</p>
				<p>2019 - 1.2.6. Key quality datasets with well-annotated metadata published in institutional repository in 2019.</p>	<p>Complete</p>	<p>GARDIAN took a big step toward becoming the definitive agricultural data resource in 2019, with an almost 10-fold increase in datasets and publications made discoverable between December 2018 and December 2019. These assets are from CGIAR repositories and from other important agricultural data sources, recognizing that CGIAR scientists need relevant data wherever it resides, in the consistent and standardized format that GARDIAN provides. With support from the Organize Module, several Centers were able to implement the latest version of the CG Core metadata schema and most Centers made strong progress in 2019 on enriching metadata associated with data and publications, and applying semantic standards across all assets -- making them easy to discover and interpret, and rendering them more actionable by machine-based techniques to demonstrate value.</p>

MI				2019 - 1.2.7. An expert annotation tool is available to add quality meta-data and annotations prior to data upload in the repositories.	Complete	User testing and alignment of the COPO data annotation tool with GARDIAN functionalities was coordinated by the Organize Module with Alliance Bioversity-CIAT, which also contributed substantial staff time to this effort. COPO is now fully responsive to the newest version of the CG Core Metadata Schema, having been further tested by data and information managers during the CoP annual meeting, which was held during the Platform Convention in October 2019, and refined by December 2019. Supported by Organize, CIP, the Alliance, and IFPRI also created a release of AgroFIMS with help from CIMMYT, IITA, Rothamsted Research, UC Davis, and University of Florida researchers. AgroFIMS is another "expert tool" which automatically annotates data with semantic standards and conforms to the CG Core schema at the collection stage.
				2019 - 1.2.8. Enhance data discoverability through improved meta-data, using text mining and machine learning methodologies; also tested for ontologies development, for data quality and/or searchability.	Complete	In 2019, the Organize Module tested the data mining of GARDIAN towards applying machine learning on large pools of data. The Module worked with the University of Florida's AgMIP team, which first mapped and adapted their model data dictionary (ICASA variables) to relevant ontologies, harmonizing GARDIAN annotations to the AgMIP format. Auto-generated harmonized datasets are made available for input to the DS-SAT crop model. The UC Davis team identified and applied an approach for harmonizing key data variables and apply machine learning and spatial approaches to analyze a large number of datasets from GARDIAN, with a high-value publication envisioned. IWMI is in the process of reorganizing its data warehouse and implementing the CG Core v. 2.0 for enhanced data discovery. Similarly, IRRI's Farm Household Survey Database has been adapted to be fully machine readable, with 99 surveys annotated to render them findable via GARDIAN. Other social science survey datasets from IRRI are currently awaiting curation using the Socioeconomics Ontology in development by the Socioeconomics Development Community of Practice under the Convene Module. These will also be shared for a machine learning annotation project through the CoP. IWMI began improving its metadata workflows and WorldFish began development of a Fish Ontology for improved annotation.

MI	M1 Outcome: 1.3. Standards and semantics are utilized to enable FAIR agricultural data.	Increased capacity for innovations in partner research organizations		2019 - 1.3.6. CG Core Metadata Schema v. 2.0 implemented, and/or mapped across Center publications and data repositories.	Complete	The Organize Module continued to support Center adoption of the CG Core Metadata Schema v. 2.0 in various ways, including facilitating the Metadata Working Group's active engagement around the schema, and collaborative fine-tuning to further improve it. The schema was uploaded to GitHub for easy access with robust documentation, and an application profile to enable clarity on terms and facilitate use. The Module provided support for further development of the COPO annotation tool to fully align with the CG Core v. 2.0, allowing easier and wider use of the schema to describe publications, data, and other information resources. Almost all CGIAR Centers are now implementing the newest version of the CG Core, and several are testing COPO for suitability with their data management workflows. IWMI began a data reorganization and metadata review in 2019, developing tools to map metadata from one format to another to facilitate adherence to the CG Core Metadata Schema. CGIAR Dataverse Working Group can be accessed here .
				2019 - 1.3.7. Fieldbook based on the Agronomy Ontology tested for data collection in the field.	Complete	The Agronomy Ontology is available for wide download and deployment via a GitHub agrisemantics collection , and users can also find information on it via the CGIAR Platform for Big Data in Agriculture website . AgroFIMS v. 1.0 was released at the end of 2019. It reliably allows the digital collection of semantically standardized data through the KDSmart app via smartphone or tablet, allowing the data to be checked for consistency in the field (via heat mapping). Data values can be validated once exported back to AgroFIMS, which also allows the user to perform statistical analyses on the collected data, generating reports that can be excerpted into reports, publications, etc. A workshop was organized in 2019 with agronomists and developers from CIP, the Alliance, IFPRI, Ontocale Inc., Diversity Arrays Inc., CIMMYT, IITA, Rothamsted Research, and the University of Florida to test improved features of AgroFIMS. This led to the design of more features. Testing sessions were held during the Big Data Convention at ICRISAT and individually with interested agronomists at IIRI and AfricaRice. A user-manual was created and published. It is also available on the AgroFIMS website.

MI				2019 - 1.3.8. Ontology (Crop and Agronomy Ontology at minimum) and/or AGROVOC/ GACS terms are available as annotation tools used by Centers (e.g. COPO) for upload in repositories, in GARDIAN, GEMS.	Complete	In 2019, the Organize Module facilitated collaboration on an ontology tool developed by the GEMS team, along with support for data and information specialists via training sessions, webinars, and coordinated user testing to demonstrate the value and use of ontologies and controlled vocabularies in annotating data assets. Almost all Centers report using AGROVOC (controlled vocabulary) in 2019 in repository metadata fields; several CoPs also report using terms from the Crop, Agronomy (maintained by the Alliance Bioversity-CIAT with support from Organize), and other ontologies, among them the Alliance, IITA, IRRI, IWMI, CIMMYT, and CIP. The use of ontology concepts to describe variables in datasets for enhanced interpretability and interoperability remains rarer, with progress made by the Alliance, CIP, IITA, and IRRI. IRRI continued work to enhance the Rice Ontology in 2019, by adding/mapping terms in IRRI SNP-Seek, B4R (EBS) databases, and in Agroportal. Additionally, IRRI's UAV- controlled vocabulary terms were shared with the Ontologies CoP, and a pilot database management system for UAV images and metadata was developed for automated high-throughput phenotyping. IWMI's 2019 data reorganization and metadata review also incorporated terminology from existing ontologies, and data managers are considering how metadata forms may eventually verify matches between user input terms and ontologies. Crop Ontology: https://www.cropontology.org/ Agronomy Ontology: https://github.com/AgriculturalSemantics/agro
				2019 - 1.3.9. A first version of the socio- economic ontology developed and aligned to the 100 questions for Agricultural Household surveys and is available on Github.	Cancelled	The Socioeconomic Ontology is being developed by the Socioeconomic Development CoP under the Convene Module. Progress will be reported in that section.

<p>MI</p>				<p>2019 - 1.3.10. Promoting CGIAR agri-se- mantics within CGIAR and to relevant exter- nal semantics groups, such as Planteome, IC- 3Food, RDA-IG- AD, and CoData in coordination with the Ontol- ogy CoP.</p>	<p>Complete</p>	<p>The Organize Module provided in-kind and funding support to create the semantic web-based CGIAR Expert Finder, enabling CGIAR to showcase its research and researchers. The Module also supported the Ontology Working Group and CoP in enhancing crop and other ontologies and in engaging with the Research Data Alliance's (RDA) agrisemantics efforts. CGIAR agrisemantics efforts were consolidated in one GitHub loca- tion for easy and organized access by users. The site includes links to the Agronomy Ontology and the in-development Socioeconomic Ontology. Module support also enabled a data cure-a-thon organized by IRRI to present development of the Rice Ontology and the CGCore Metadata Schema, with refer- ence and application materials made accessible to those IRRI staff members working on data and/or databases. With sup- port from Organize, a Bioversity-based team continued collab- orations with the Environment Ontology, Food Ontology, SDG Interface Ontology, and Planteome Project teams, and attend- ed the 2019 Plant and Animal Genomics Conference, interact- ing with the Excellence in Breeding and Integrated Breeding Platforms and the Planteome Project. CGIAR's agrisemantics efforts were represented by Bioversity-based staff at the SWAT- 4LS Conference organized by GODAN, and the Module con- tinued engaging with CABI, FAO, CTA, and GODAN on issues focused on agrisemantics and capacity development around FAIR data. MARLO link to project contribution here.</p>
<p>MI Outcome: 1.4. Enhance capacity, catalyze cultural change to further CGIAR OA/OD com- pliance and public goods mandate.</p>	<p>Enhanced institution- al capacity of partner research organiza- tions.</p>	<p>Enabling CGIAR to move further and faster towards Open, FAIR, and ethical data management necessitates enhanced capacity and cultural change that can be fueled by the provision of services, tools, and support to ease transition. Throughout 2019, the Organize Module supported the development of tools and guides covering best practices in data management to insure open, FAIR, and ethical data assets, and for data processing.</p>	<p>2019 - 1.4.4. Materials de- veloped and webinars held to support the community maximizing FAIR CGIAR resources.</p>	<p>Complete</p>	<p>The Organize and Convene Modules organized several webi- nars supporting the creation of open, FAIR and ethical data assets, and facilitated Centers in supporting their research communities in a variety of ways. A multi-module online and interactive course was developed, which will be deployed in 2020 to provide tools and knowledge to help researchers and data/information managers use best practices in ethi- cally managing data assets toward making them open and FAIR. Guidance was made available via GARDIAN to help data managers improve the FAIRness of data assets. This was pre- sented during the Platform's 2019 Convention. ILRI organized presentations (see event here) by scientists and others on "the power of open - inspiring stories and discussion" (e.g. mobile phone-based surveillance systems, making ILRI code open, open data for decision-making etc.). Several Centers including the Bioversity-CIAT Alliance, CIP, ILRI, and IITA developed do- main-specific guidance on best practices in managing data for their researchers, often involving researcher-data champi- ons. Through an Organize Module collaboration with UC</p>	

<p>MI</p>	<p>MI Outcome: 1.4. Enhance capacity, catalyze cultural change to further CGIAR OA/OD compliance and public goods mandate.</p>	<p>Enhanced institutional capacity of partner research organizations.</p>	<p>Module funds were used by Centers for data sprints to promote standards-based data annotation and upload to repositories. Module-facilitated webinars were organized, and an online interactive course on data management best practices developed for deployment in 2020. To assure that capacity development solutions meet real needs the Module organized user workshops, which included information designed to enhance the use of the COPO annotation tool with Alliance data and by information specialists, as well as COPO/Earlham Institute, GARDIAN, and University of Florida/AgMIP developers. COPO and AgroFIMS needs were also assessed through larger user workshops. Several regional workshops were held (at CIMMYT, CIP, ICRISAT, and IFPRI) to enhance CGIAR's data science and machine-learning capabilities. The GARDIAN team provided technical guidance and/or APIs to make Center/CRP data assets discoverable (e.g. to IRRI, IITA, ICRISAT, IWMI). An API key and support for harvesting GARDIAN FAIR scores was provided and is used by MARLO.</p>	<p>2019 - 1.4.4. Materials developed and webinars held to support the community maximizing FAIR CGIAR resources.</p>	<p>Complete</p>	<p>Davis, regional workshops were held at CIMMYT, CIP, ICRISAT, and IFPRI to enhance CGIAR's capabilities in data science-related topics including advanced R, spatial predictive modeling, and machine learning, and a data wrangling/processing guide produced and made available as a "living document" that continues to be enhanced. See ILRI's data management guidelines here and CIP's data management toolkit here. See the Alliance/CCAFS Support Pack here: https://ciat.cgiar.org/data-management-support-pack/ Draft course – "Best practices for FAIR, ethical, and open data assets" : https://drive.google.com/drive/u/1/folders/1qobb_8ltyroDo3l-FSivHRRqqoujSv5f</p>
------------------	--	---	--	--	-----------------	--

MI				2019 - 1.4.5. Engaging more scientists into the quality data publishing workflow, by means of at least two data sprints, or 1 relevant training and 1 data sprint, to build capacity and increase the number of well-annotated datasets in Center data repositories.	Complete	Facilitated by the Organize Module, Centers creatively and enthusiastically engaged researchers on data annotation and management efforts in 2019 through data sprints, data clinics, hackathons, “cure-a-thons,” and other training modalities. The Bioversity-CIAT Alliance held three training sessions on data management planning across all CIAT regions to ensure that researchers use good data practices from project inception, while Bioversity-based staff participated in data sprint activities. ILRI organized help-desks and data clinics, and disseminated guidance on how to “be open” across different knowledge products. IIRRI organized four training courses on RDM 101 (Basic Research Management), incorporating FAIR as well as OA/OD -- with 76 participants. Additionally, a data sprint was conducted at IIRRI for about 20 key scientists involved in data and databases. IITA launched a massive campaign across hubs in 2019 on best practices for data quality in projects. CIFOR embedded data teams into projects to promote good data management planning, development, and curation. CIMMYT data management staff worked with all stakeholders including key data stewards to promote good data management. Two hackathons were organized by ICRISAT (India) and one in Ethiopia for the ESA and WCA regions, with four digitization and data management trainings also conducted for breeding programs.
				2019 - 1.4.6. At least two workshops or training sessions for data, knowledge, and ontology managers and researchers held on ways to render datasets FAIR, in collaboration with Wageningen University, GODAN, and FAO.	Complete	The Module organized a workshop to enhance the COPO annotation tool involving Alliance data and information specialists, as well as COPO/Earlham Institute, GARDIAN, and University of Florida/AgMIP developers. Online and in-person training, the latter of which was held at the Platform’s 2019 Convention was held for data managers and ontology specialists to familiarize themselves with the COPO data annotation tool. Likewise, an AgroFIMS user workshop involving Alliance ontology experts, developers from CIP and IFPRI, and researchers from multiple Centers and partners (CIMMYT, IITA, Rothamsted Research, University of Florida, Diversity Arrays Inc., and Ontocale Inc.) allowed researchers to test AgroFIMS’ FAIR data workflows and provide input to refine them. The Organize Module also facilitated researcher-focused regional workshops to enhance CGIAR’s data science and machine-learning capabilities; these were held at CIMMYT, CIP, ICRISAT, and IFPRI and covered advanced R, spatial predictive modeling, and machine learning. Draft course – “Best practices for FAIR, ethical, and open data assets”: https://drive.google.com/drive/u/1/folders/1qobb_8ltyroDo3l-FSivHRRqqoujSv5f

M2	M2 Outcome: 2.1. CGIAR is more broadly engaged in the big data community.	Enhanced institutional capacity of partner research organizations.	The BIG DATA Platform's open, collaborative communities of practice grew to more than 3,500 members and they curated open technical tracks at the annual Big Data in Agriculture Convention (with more 700 attendees, 65% of which were external to CGIAR. The event had a media reach of several million views of print, online, and television media).The Platform launched new collaborative research	2019 - 2.1.4. Hold high-level 2019 Convention on Big Data in Agriculture, with wide participation of CGIAR and non-CGIAR actors, establishment of collaborative agreements.	Complete	High-level 2019 Convention on Big Data in Agriculture, with wide participation of CGIAR and non-CGIAR actors, was held from 16-18 October in Hyderabad, India, hosted by ICRISAT. The Convention had more than 700 attendees, most of which were external attendees, and a media reach of several million views of print, online, and television media. The Youth in Data initiative entered its second year. It was developed as an explicit strategy to engage young reporters and data scientists in digital agriculture.
			and events with Cambridge University and the Partnership on AI (a membership organization comprised of the major IT firms, researchers, startups, and civil society groups working on ethical AI) specifically on emerging topics and managing systemic risks to food security.	2019 - 2.1.3. CoPs around topics of geospatial data, socioeconomic data, ontologies, data-driven agronomy, livestock data for development, and crop modeling have produced data output that address key constraints of the CGIAR and the wider sector and strengthen CoP networks.	Complete	All news, information, and activities from CoPs in 2019, including newsletters and webinars, can be found here: https://bigdata.cgiar.org/communities-of-practice/ . 2019 Wrap up updates from the CoPs can be found here: https://us15.campaign-archive.com/?u=59a7500ef5c94dd50e2b9e2fb&id=24350e63e0
	M2 Outcome: 2.2. CGIAR increases its capacity to work on priority topics more quickly, more effectively, and at greater scale.		The Platform launched new collaborative research and events with Cambridge University and the Partnership on AI (a membership organization comprised of the major IT firms, researchers, startups, and civil society groups working on ethical AI) specifically on emerging topics and managing systemic risks to food security. Big Data Platform open, collaborative CoPs grew to more than 3,500 members and they curated open technical tracks at the annual Big Data in Agriculture Convention.	2019 - 2.2.2. Identify high priority, high impact new data products including water resources, crop mapping, field boundaries, poverty, climate change adaptation, and gender, and develop a methodical plan to produce	Complete	1) New set of MetGrid files and CLI files sets, 2) Interactive map for visualizing the simulations of different G x M options in all grids of rabi sorghum regions in India, and 3) Aggregate datasets on crop stage and growth, created from cellphone images provided by more than 30,000 wheat farmers. Dataset for Crop Monitoring Using Smartphone Based Near-Surface Remote Sensing: Ground Pictures of Wheat and Auxiliary Data from Northern India is a processed dataset of approximately 20,000 near-surface remote sensing images acquired using inexpensive smartphones within the context of a picture-based insurance (PBI) initiative of 1,685 smallholder farmers fields in northwest India. Spatial Production Allocation Model (SPAM 2015): the team processed subnational statistics data

M2				them, with an initial implementation. A special focus in 2020 will be development of training data for machine learning applications.		covering 28 priority countries in sub-Saharan Africa, Asia, LAC, and Europe, but there are still 105 more countries to process. Currently, our tentative target date for release is December 2020. Continuing working on MapSPAM 2015. Update to MapSPAM 2010, v. 1.1 was released on October 9, 2019.
	M2 Outcome: 2.3. CGIAR develops as a learning organization.		As a result of a pan-CGIAR assessment of the state of digital strategy in the organization, seven Centers worked with the BIG DATA Platform to begin or enhance their Center-level digital strategies. The Platform Coordinator presented progress on the pan-CGIAR digital strategy to the System Management Board and the Platform was tasked with developing a digital strategy in light of the 2030 plan. The Platform conducted co-design workshops across Centers and domains to develop a more unified information vision for the organization and has begun to apply this to prioritize digital investments, such as in shared services. Design outputs were presented to partners and CGIAR participants at the Big Data in Agriculture Convention. The BIG DATA and Gender Platforms spearheaded a new method for leveraging mobile phone network metadata to study the spatial extent and changes in key factors related to women's economic empowerment, with plans to mainstream this method in the organization.	2019 - 2.3.1. Map out CGIAR needs for common big data related computing and storage infrastructure.	Complete	As a result of a pan-CGIAR assessment of the state of digital strategy in the organization, seven Centers worked with the BIG DATA Platform to begin or enhance their Center-level digital strategies. The Platform Coordinator presented progress on the pan-CGIAR digital strategy to the System Management Board (SMB) and the Platform was tasked with developing a digital strategy in light of the 2030 plan (See the Chair's summary of the SMB presentation here .). The Platform conducted co-design workshops across Centers and domains to develop a more unified information vision for the organization and has begun to apply this to prioritize digital investments, such as in shared services. Design outputs were presented to partners and CGIAR participants at the Big Data in Agriculture Convention.
				2019 - 2.3.5. Map CGIAR needs for common big data-related computing and storage infrastructure, through infrastructure mapping and digital strategy services to Centers.	Complete	The Platform conducted co-design workshops across Centers and domains to develop a more unified information vision for the organization and has begun to apply this to prioritize digital investments, such as in shared services. Design outputs were presented to partners and CGIAR participants at the Big Data in Agriculture Convention.

M2				2019 - 2.3.6. Establish shared services for CGIAR by negotiating with external data utility partners for weather data, commercial satellite imagery, secure data sharing, machine learning, and other products.	Complete	The Platform has secured third-party services from aWhere and The Weather Company for weather data, access to the archive of the satellite imagery provider Digital Globe, and a trial subscription to a seasonal probabilistic forecast product from IBM.
				2019 - 2.3.7. Develop capacity-building activities (a combination of in-person, webinars, and online courses) linked to Centers' needs on themes including machine learning, data management, digital extension, and using shared services for imagery processing.	Complete	The Annual Big Data in Agriculture Convention included side workshops specifically on familiarizing CGIAR researchers with emergent data science methods and underscoring good practices: https://2019cgiarbigdatainag.sched.com/ Big Data Webinars and courses: https://bigdata.cgiar.org/webinars/ Introduction to GDPR Course: https://bigdata-cgiar.course.tc/catalog/course/gdpr-for-international-development
				2019 - 2.3.8. Monitor the capacity of CGIAR to meet the data needs of the agricultural development sector and its development as an analytics-driven organization.	Complete	The main platform mechanisms for monitoring and building capacity are technical Communities of Practice , which implement webinars throughout the year and curate technical tracks of the annual Convention . This culminates in a final work plan for the CoPs that is developed late in the calendar year, in time for inclusion in the next year's Plan of Work and Budget. Big Data Online Course Catalog for courses and webinars: https://bigdata-cgiar.course.tc/catalog

M3	M3 Outcome: 3.1 CGIAR shows how data-driven approaches yield results in poverty reduction, enhanced nutrition, or environmental benefits.	Improved forecasting of impacts of climate change and targeted technology development	M3 has built a portfolio of 14 projects to date demonstrating research innovations and new pathways to data-driven impact for CGIAR. For example: a successful pilot of machine-learning tool that enables seed selection is being leveraged for in-field seed and fertilizer recommendations in Mexico and East and Southern Africa, linking to genetic resources in a CGIAR seedbank. A system leveraging free Wi-Fi to track informal food flows has been adopted by the Government of Vietnam to diagnose and predict COVID-19-related food security shocks (58% of users are women). A method leveraging mobile telephoned network data has proven more accurate in some circumstances than weather stations and CHIRPS in measuring rainfall. A farmer-to-farmer social network linked to ILRI researchers led to significant improvement in dairy practices. 33,000 wheat farmers in India found they improved risk management and had increased willingness to pay for a mobile phone-based advisory and insurance product.	2019 - 3.1.4. Successful piloting of five Inspire grantees in Africa and South Asia, and up to four new pilot Inspire projects around Big Data related innovations assigned.	Complete	The pilot projects were implemented in 2019. See full project portfolio here: https://bigdata.cgiar.org/inspire/inspire-winners/
				2019 - 3.1.5. Scale-up of three pilot Inspire projects on pest and disease surveillance, farmer-to-farmer digital extension and ICT-enabled insurance and award of two additional scale-up grants in Q4.	Complete	Awards were made and early results are coming in. Projects included the launch of randomized control trials of picture-based insurance and advisories offered as retail services; credit linked to these; and a new advisory service pilot; scale-up of successful in-field wheat rust diagnostics; and more than 180,000 users of a farmer-to-farmer social media group. See Farm.Ink Digital Learning Platform here . Please see links to resulting dataset and papers: Dataset: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBAFZY Papers: https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-019-0684-y https://www.sciencedirect.com/science/article/pii/S2352728518300812?via%3Dihub https://www.sciencedirect.com/science/article/pii/S0168192318303484?via%3Dihub
M3				2019 - 3.1.6. Synthesis of Inspire project successes and failures in 2019; best practice guidance provided.	Extended	Preliminary findings of the analysis were presented at the Big Data Convention on what the Inspire Challenge process has revealed in terms of trends in digital innovation in the agricultural research for development space and what this may mean for CGIAR and the sector actors seeking to advance digitization.

TABLE 6: NUMBERS OF PEER-REVIEWED PUBLICATIONS FROM CURRENT REPORTING PERIOD

	Number	Percent
Peer-Reviewed publications	11	100.0%
Open Access	10	90.91%
ISI	9	81.82%

TABLE 7: PARTICIPANTS IN CAPDEV ACTIVITIES

Number of trainees	Female	Male
In short-term programs facilitated by CRP/PTF	106	157
In long-term programs facilitated by CRP/PTF	1	11
PhDs	0	1

TABLE 8: KEY EXTERNAL PARTNERSHIPS

Lead Module	Brief description of partnership aims (30 words)	List of key partners in partnership. Do not use acronyms.	Main area of partnership (may choose multiple)
M1	Develop the GARDIAN data ecosystem, enabling data discovery from CGIAR and R&D partners, need-driven services, and a prototype analytics environment and model pipelines, all backstopped by secure data sharing.	SCiO - Big Data in Food Systems	<ul style="list-style-type: none"> Capacity Development Delivery
M1	Develop analytic use cases to demonstrate seamless integration of GARDIAN data with agricultural R&D tools. Extend AgMIP translation tools to harmonize data discovered via GARDIAN for input to models.	UF - University of Florida	<ul style="list-style-type: none"> Delivery Capacity Development Research
M1	Develop R-based spatial data analysis environment to enable seamless analysis of data discovered via GARDIAN and promote its use through training workshops.	UC Davis - University of California, Davi	<ul style="list-style-type: none"> Capacity Development Delivery Research
M1	Support Agronomy Ontology refinement and mapping to field book; develop RDF converter for AgroFIMS; develop VIVO Expert Finder ontology and portal for CGIAR.	Ontocale	<ul style="list-style-type: none"> Capacity Development Delivery
M1	Implementation of the COPO data annotation tool to streamline metadata entry and ontology annotation of data variables prior to upload of data assets to GARDIAN or other repositories.	Earlham Institute	<ul style="list-style-type: none"> Delivery Capacity Development
M2	Co-development of events centered on the role of AI in agricultural development and collaborative research on emerging topics.	Partnership on AI	<ul style="list-style-type: none"> Capacity Development Research Policy
M2	Collaborative research on emerging priority topics in digital agriculture and AI	CAM - University of Cambridge	<ul style="list-style-type: none"> Research
M2	Co-funding of a Computer Science Phd focused on digitally enabled decision support system for farmers.	INRIA - National Institute for Research in Digital Science and Technology. CIRAD - Centre de coopération internationale en recherche agronomique pour le développement	<ul style="list-style-type: none"> Capacity Development
M3	External partner for Inspire Challenge winner: An Integrated Data Pipeline for Small Fisheries	PDS - Pelagic Data Systems	<ul style="list-style-type: none"> Research
M3	External partner for Inspire Challenge winner: CubicA - Agriculture Advisory App	alberg - The Dalberg Group Viamo	<ul style="list-style-type: none"> Research
M3	External partner for Inspire Challenge winner: Smart Seed Selection	BioSense Institute	<ul style="list-style-type: none"> Research
M3	External partner for Inspire Challenge winner: Revealing Informal Food Flows through Free Wifi	MARD - Ministry of Agriculture and Rural Development (Vietnam)	<ul style="list-style-type: none"> Research
M3	External partner for Inspire Challenge winner: Use CML to Estimate Rainfalls for Agriculture	Cornell University AtmosCell - AtmosCell	<ul style="list-style-type: none"> Research

TABLE 9: INTERNAL CROSS-CGIAR COLLABORATIONS

Brief description of the collaboration	Name(s) of collaborating CRP(s), Platform(s) or Center(s)	Optional: Value added, in a few words
AfricaRice and the Organize Module worked together towards compliance with CGIAR's Open Access and Data Management Policy and to enable FAIR outputs resulting in increasingly discoverable and interoperable metadata and data assets.	AfricaRice	Scientific and efficiency benefits
Bioversity worked with the Organize Module to publish quality, well-annotated data assets, including datasets from the Center's research portfolio in Dataverse, and associated publications in Dspace -- enabling harvesting by GARDIAN, and driving towards CGIAR's public goods goals.	BIOVERSITY	Scientific and efficiency benefits
As one of the Platform's co-leaders, CIAT continued work with the Organize Module on implementing CGIAR's strategy towards open and interoperable data and publications. This collaboration enabled the Center to manage research outputs well, support key policies by developing guidance and trainings, and making strategic investments in key human resources to support these activities.	CIAT	Scientific and efficiency benefits
CIFOR worked with the Organize Module towards implementation of the FAIR Guiding Principles for Scientific Data Management and Stewardship. The Center ensured data quality and preservation of research outputs by supporting governance and best practices for managing data assets, in line with CGIAR and funder strategies and to increase the visibility of outputs through GARDIAN.	CIFOR	Scientific and efficiency benefits
CIMMYT joined forces with the Platform and worked with the Organize Module to ensure that the Center's data followed the FAIR principles. Besides data being findable, accessible, interoperable and reusable the Center collaborated to advance the generation of data in an ethical and reproducible manner.	CIMMYT	Scientific and efficiency benefits
The purpose of the collaboration between CIP and the Organize Module of the Platform was to support implementation of the Open Access/Open Data strategy in line with CGIAR's public goods mandate and make the Center's outputs FAIR. A range of strategies was developed to further structure and institutionalize the principles of OA and DM into standard operations. The emphasis was on generating high-quality, well-annotated data using the CG Core Metadata Schema, making assets available via our repositories and providing technical expertise and operational support to enable FAIR.	CIP	Scientific and efficiency benefits
ICARDA collaborated with the Organize Module to build on implementation of the CGIAR Open Access Policy and its implementation guidelines, work initiated in 2015-2016. Collaborative activities in 2019 focused primarily on the generation of FAIR data assets to better enable machine learning and similar technologies to operate on them.	ICARDA	Scientific and efficiency benefits
ICRAF's fund allocation provided by the Organize Module was used in part to support the implementation of CGIAR's and ICRAF's Open Access/Open Data Plan and make data FAIR, through two support units at the Center responsible for operationalization of OA/OD. These are the Research Methods Group and the Knowledge Management Group, both of whose activities towards open and FAIR assets were supported through this collaboration.	ICRAF	Scientific and efficiency benefits

Brief description of the collaboration	Name(s) of collaborating CRP(s), Platform(s) or Center(s)	Optional: Value added, in a few words
ICRISAT organized trainings on good annotation and best management of data resources through the Organize Module's initiatives, with upload to various electronic platforms. The main emphasis was to generate quality research data compliant with CGIAR's mandate and the FAIR principles to more effectively leverage various advanced data analytics and decision-making tools and technologies.	ICRISAT	Scientific and efficiency benefits
Building on progress made in 2018, IFPRI continued working to operationalize Organize Module and CGIAR OA/OD goals by increasing the production and usage of open access information products including journal articles, books, discussion papers and datasets.	IFPRI	Scientific and efficiency benefits
IITA worked with the Organize Module of the Big Data platform on getting standardized data uploaded to the Center's CKAN repository. To achieve this, an effort was made to identify and work with scientists and data managers who could be champions. Three aspects were the primary focus of 2019 work: 1) A workbook that would become a standard resource for the institute and beyond; 2) Workshops and meetings for capacity building, to educate and empower scientists; 3) A major data sprint to get as much data uploaded as possible; and 4) Enhancement and streamlining of the online repository (CKAN) to make it easy and efficient to use, and assure integration with GARDIAN.	IITA	Scientific and efficiency benefits
ILRI worked with the Organize Module of the Platform towards the implementation of the Open Access and Data Management Policy and making data FAIR.	ILRI	Scientific and efficiency benefits
IRRI, through support from the Organize Module of the Big Data Platform, implemented activities to ensure compliance of research datasets, publications, and databases with the CGIAR Open Access and Data Management Policy and the FAIR Principles. These activities were (1) Development on databases (SNP-Seek, GOBII,B4R) to create APIs; datasets tagging with Rice Ontology, AGROVOC/GACS terms, and domain-specific ontologies in coordination with the Ontology WG of the Module; (2) Documentation/capture of drone imaging methodologies for HT phenotypes/GIS applications, identifying appropriate metadata/controlled vocabulary terms, and developing a pilot image management system that uses these; and (3) Continued development for interoperability (through the IRRI Interoperability WG) of various digital asset management systems (e.g. IRRI Dataverse, publications repository). These included: Updating the metadata of deposited research data in Dataverse and publications with CG Core Metadata Schema, ontology (Crop/Agronomy Ontology) and/or AGROVOC/GACS terms, and developing APIs to support FAIR compliance. Trainings and data sprint were conducted to support institute-wide awareness and implementation (by the responsible teams) of OA/OD and FAIRness in research datasets, publications, and databases/data management systems.	IRRI	Scientific and efficiency benefits
While IWMI has long had a policy of open data access, which we implement by making all of our project data available on IWMI's Water Data Portal (http://waterdata.iwmi.org/), we have struggled with capacity on keeping the portal up to date. However, the work with the Big Data Platform's Organize Module in 2019 and past years allowed the the Center to go beyond the current status quo of making data available. We have been developing a strategy for consistent review and harmonization of data and producing data products and interfaces that make our data holdings more visible, more useful, and more consistent across regions and projects.	IWMI	Scientific and efficiency benefits

Brief description of the collaboration	Name(s) of collaborating CRP(s), Platform(s) or Center(s)	Optional: Value added, in a few words
<p>WorldFish, as the lead center of the CGIAR Research Program on Fish Agri-food Systems (FISH), continued to support CGIAR efforts to comply with the CGIAR Open Access and Data Management Policy (OADM) adopted in 2013, and its 2014 Implementation Guidelines. FISH, through the Organize module of Big Data platform, ensured that all information products, deliverables and outputs related to FISH comply with the FAIR principles. The primary focus was to consolidate both governance and management structures in order to promote an open data culture within research program. In addition, it was key to prioritize high-quality datasets associated with consistent metadata for sharing via the Center's repositories based on the guidelines developed through Organize Module activities, and finally to work towards inspiring the culture change required to enable FAIR.</p>	WorldFish	Scientific and efficiency benefits
<p>In 2019 the Platform initiated a cross-platform design process for building a unified pan-CGIAR information architecture vision. Excellence in Breeding, Big Data, and Genebanks representatives and subject matter experts in breeding, genomics, scaling, agro-ecology, large-scale agronomy, animal science, and fisheries developed a unified view on key research processes and the necessary data, digital tools, and stakeholders involved in supporting each. This fundamental design work is being folded into Big Data inputs to One CGIAR.</p>	EiB	Scientific and efficiency benefits: Big Data built on EiB design achievements to extend them across domains and across Centers to define a more unified information vision for CGIAR.
<p>Big Data and the Gender Platform conducted collaborative research on big data-enabling research into women's economic empowerment.</p>	PIM	Data science collaboration bridging big data analysis and gender analysis and translation of methods and terminology were valuable aspects of this collaboration.
<p>Big Data and Excellence in Breeding reinforced each other's efforts to develop standard operating procedures and analytic pipelines for remote sensing imagery. Specifically, Big Data partially developed a pipeline targeting rice and cassava, whereas EiB focused on other crops. The two efforts were compared and contrasted, and going forward Big Data will contribute funding to a pan-CGIAR shared service led by Eib.</p>	EiB	Parallel similar efforts targeting different crops increased common understanding of best and more unified practices.

TABLE 10: MONITORING, EVALUATION, LEARNING AND IMPACT ASSESSMENT (MELIA)

Studies/learning exercises planned for this year (from POWB)	Status	Type of study or activity	Description of activity / study	Links to MELIA publications
S2262 - Survey of data needs and practices among food system actors, and how CGIAR can better meet these.	On Going	Program/project evaluation/review	This data demand study is replicated from one conducted in 2018, with the aim of developing good transnational views.	
S2300 - Farm.ink Learning study on impact of digital services (bot/web) on farming practices	Extended	Other MELIA activity	Plan to conduct study through digital surveys to our user base. Focused on learning which areas of practice we are having the most influence over.	https://spark.adobe.com/page/elcejE-Jp45Jjb/
S2915 - PeskaAS: Automated analytics of relative fish abundance over time and space	On Going	Synthesis (secondary) study	N/A	https://doi.org/10.7910/DVN/TVGIJJ
S3154 - Digital technologies for financial inclusion of smallholder farmers: Needs assessment in three states of India	Completed	Ex-ante, baseline and/or foresight study	The "Seeing is Believing" project aims to deliver personalized agricultural advice and financial services to smallholder farmers, using smartphone pictures for crop monitoring. We test alternative models for scaling: a retail insurance product for commercial crops in Haryana; credit in order to unlock credit for tenant farmers in Odisha; and an integrated pest management advisory in Tamil Nadu and Pondicherry. This deliverable analyzes whether these services are perceived to provide value, with people willing to engage, and whether farmers value the reduced basis risk for insurance, the improved access to credit, and the personalization of their advisories.	http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/133534

TABLE 11: UPDATE ON ACTIONS TAKEN IN RESPONSE TO RELEVANT EVALUATIONS

N/A

TABLE 12: EXAMPLES OF W1/2 USE IN THIS REPORTING PERIOD (2019)

N/A

TABLE 13: PLATFORM FINANCIAL REPORT

	Planned Budget 2019*			Actual expenditure*			Difference*			Comments
	W1/W2	W3/ Bilateral	Total	W1/W2	W3/ Bilateral	Total	W1/W2	W3/ Bilateral	Total	
M1 - Organize	US\$ 2,407,764.98	US\$.00	US\$ 2,407,764.98	US\$ 2,264,979.71	US\$.00	US\$ 2,264,979.71	US\$ 142,785.27	US\$.00	US\$ 142,785.27	Module One continued to follow the budget as outlined in the initial proposal closely.
M2 - Convene	US\$ 3,132,442.73	US\$ 1,055,423.23	US\$ 4,187,865.96	US\$ 2,089,745.71	US\$ 525,262.46	US\$ 2,615,008.17	US\$ 1,042,697.02	US\$ 530,160.77	US\$ 1,572,857.79	Big Data continued to follow the budget as outlined in the initial proposal closely, though some fundraising in support of the annual Convention will arrive in Big Data accounts in 2020.
M3 - Inspire	US\$ 1,270,459.98	US\$.00	US\$ 1,270,459.98	US\$ 774,213.57	US\$.00	US\$ 774,213.57	US\$ 496,246.41	US\$.00	US\$ 496,246.41	Module Three typically has highest carryover among modules due to the timing of a grant process the Platform runs, in which funding commitments are made at the end of the calendar year.
Strategic Competitive Research grant	US\$.00	US\$.00	US\$.00	US\$.00	US\$.00	US\$.00	US\$.00	US\$.00	US\$.00	The Platform does not typically conduct research and does not compete with Research Programs in going after funding.
Platform Management & Support Cost	US\$ 382,602.37	US\$.00	US\$ 382,602.37	US\$ 355,966.81	US\$.00	US\$ 355,966.81	US\$ 26,635.56	US\$.00	US\$ 26,635.56	Management and support costs hewed closely to the budgeted amounts in 2019.
Platform Total	US\$ 7,193,270.06	US\$ 1,055,423.23	US\$ 8,248,693.29	US\$ 5,484,905.80	US\$ 525,262.46	US\$ 6,010,168.26	US\$ 1,708,364.26	US\$ 530,160.77	US\$ 2,238,525.03	

TABLE ANNEXES

FP	Detailed Annex
MI: Organize	<p>GARDIAN added new functions in 2019. Highlights include the ability to map and spatially query production estimates for more than 30 crops; the visualization of a 7 Terabyte (TB) climate dataset; and the creation of an analytic workbench that enables CGIAR researchers to apply machine learning analytics over the growing GARDIAN data pool. GARDIAN now includes a service that helps scan research datasets and flags personally identifiable information before the data is made open. This minimizes the risks to both vulnerable individuals and to CGIAR's reputation. The team launched the first version of Collaborative GARDIAN (CG Labs), a secure analytic environment with tools and services to enable researchers to collaborate in finding and sharing GARDIAN (or other) data securely and to use R and Python-based scripts and other analytic approaches. Researchers can now search for data via GARDIAN, import it into a shared analytic workspace in CG Labs, and find colleagues via the Find a CGIAR Expert feature to spark new collaborations. Organize released v. 1.0 of the Agronomy Field Information Management System (AgroFIMS) that employs these data quality standards to generate FAIR data at collection. Life sciences research is moving inexorably towards data annotation that leverages standard semantics and logic and, in 2019, BIG DATA made important contributions to these community standards for describing agronomic, socioeconomic, and survey data, and updating the CGIAR metadata standard. Module One is beginning to mutually reinforce the efforts of other Platform modules. GARDIAN CG Labs, for example, has been showcased in data science workshops and webinars developed with technical CoPs under Module Two. Organize investments are linking the array of CGIAR data assets in new ways to facilitate new partnerships and innovation. In 2019, for example, the International Rice Research Institute (IRRI) updated its Rice Functional Genomics and Breeding (RFGB) database, which contains 3,000 rice genomes (3K-RG), to include new features, annotations, and data related to these genetic resources (see Wang et al., 2020 and http://www.rmbreeding.cn/snp3k). In the new RFGB v. 2.0, new phenotypes and haplotypes allow associations to be inferred, enabling breeders and geneticists to narrow the number of candidate gene targets for validation. The new version of the database complements others that use 3K-RG data as a foundation, leverages partner contributions, and builds stronger partnerships with an array of organizations including the Shenzhen Institute of Breeding and Innovation for the Chinese Academy of Agricultural Sciences; Institute of Genetics and Developmental Biology Chinese Academy of Sciences; Nanjing AU; China AU; Beijing Genomics Institute Shenzhen; and Shanghai Jiao Tong University. Organize engaged with CGIAR Centers and several external partners to achieve key milestones. With the Platform's support, a Bioversity International (BIOVERSITY)-based team worked to enhance agrisemantics standards and ontologies. The Environment Ontology, Food Ontology, SDGs Interface Ontology, and Planteome project teams, and WorldFish, for example, began development of a fish ontology. In 2019, almost all Centers implemented the CG Core Metadata Schema v. 2.0 and applied the AGROVOC controlled vocabulary in annotation. Several others also reported that they are using ontologies. Among them the International Center for Tropical Agriculture (CIAT)-BIOVERSITY International Alliance (the Alliance), International Institute of Tropical Agriculture (IITA), IRRI, International Water Management Institute (IWMI), Maize and Wheat Improvement Center (CIMMYT), and International Potato Center (CIP). IRRI enhanced the Rice Ontology, adapted its Farm Household Survey Database for machine-readability, developed a pilot database for unmanned aerial vehicle (UAV) images and metadata, and shared UAV-controlled vocabulary terms. IWMI improved metadata workflows by incorporating existing ontologies. In 2019, the Organize Module tested data mining techniques to apply machine learning to GARDIAN's large data pool, worked with the University of Florida to auto-generate harmonized datasets for input to crop models and worked, also, with the University of California, Davis, to identify an approach for harmonizing key data variables and for applying machine learning and spatial approaches to GARDIAN data. To harmonize data at the collection point, the Module engaged with CIP, the Alliance, and International Food Policy Research Institute (IFPRI), soliciting researcher input from CIMMYT, IITA, Rothamsted Research, University of California, Davis (UC Davis), and the University of Florida to release AgroFIMS. The Platform backstopped the adoption of data-related outputs, standards, and tools via guidance and webinars, facilitating Centers by offering a multi-module online course on best practices in open, FAIR, and ethical data assets as well as guidance, training sessions, and webinars. Several Centers -- including the Alliance, CIP, International Livestock Research Institute (ILRI), and IITA -- developed domain-specific materials on data management best practices, often involving researcher data champions. Through UC Davis regional workshops were offered at CIMMYT, CIP, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), and IFPRI on data science-related topics including advanced R, spatial predictive modeling, and machine learning. A data wrangling/processing guide was also developed. Facilitated by the Organize Module, Centers engaged researchers on data management through data sprints, data clinics, hackathons, "cure-a-thons," and other modalities. The Alliance held three training sessions on data management planning across CIAT regions, while BIOVERSITY-based staff participated in data sprints. ILRI organized helpdesks and data clinics. IRRI hosted a data sprint and offered four training sessions on basic research management. IITA launched a campaign on best practices for data quality. Also, the Center for International Forestry Research (CIFOR) embedded data teams into its projects; CIMMYT worked with stakeholders to promote good data management; and ICRISAT organized two hackathons and four data management training sessions.</p>

FP	Detailed Annex
<p>M2: Convene</p>	<p>The 2019 Convention, hosted by ICRISAT, brought together 700 delegates, 65% of whom were external to CGIAR. These individuals were from 270 different organizations, institutions, and government bodies, as well as CGIAR centers and programs. The combined media and social media coverage reached an estimated 20 million people globally. The Convention raised nearly 70,000 USD in cash and in-kind support, a solid step towards the sustainability of an event that is establishing CGIAR as a voice of authority in digital agriculture worldwide. The Convention provides a key means for the Platform to engage with youth. The Youth In Data workshop initiative entered its second year in 2019. Thirty young digital innovators from engineering and journalism schools local to Hyderabad and ICRISAT’s doctoral program enrolled to be trained on data science and social media reporting. This was essential in enabling the Platform to reach audiences of several million through reporting on the Convention. Youth reporting also contributed to engagement with young, enthusiastic agricultural data scientists throughout the region and from around the world as well as with many young programmers and application developers through ICRISAT’s Innovation Hub and through other regional and global entities. The BIG DATA and GENDER Platforms spearheaded a novel approach to studying women’s economic empowerment in 2019. The partners conducted a phone-based survey of 10,000 respondents and used it to analyze billions of data points generated by mobile phone networks to predict gender and decision-making power among female farmers on a national scale in Uganda. The approach demonstrated the potential for observing changes in female farmers’ economic empowerment with greater speed and scale compared with solely survey-based methods. BIG DATA launched two key new strategic partnerships designed to reveal emerging risks at the intersection of digital technologies and agro-ecologies, and to build CGIAR capacity to claim the benefits of these technologies in service of its mission. BIG DATA launched collaborative research with the Partnership on AI (PAI) and the University of Cambridge Center for the Study of Existential Risk (CSER) centered on identifying emerging priority topics for machine learning in agriculture --both the systemic risks and opportunities. The (French) National Institute for Research in Digital Science and Technology (INRIA) and BIG DATA began to design a program for integrating graduate students in computer science into CGIAR agricultural research for development. The program focused on diverse themes including machine-learning enhanced crop modeling, semantic data, digital architectures for agronomic research, and farmer decision making under uncertainty. BIG DATA and The (French) National Institute for Research in Digital Science and Technology (INRIA) will jointly build new and much needed links between computer science and the agricultural research domains. The Platform’s CoPs consolidated a strong communication strategy and significantly increased its member engagement in 2019. They published a combined total of 45 blog posts (up from five in 2018), 16 newsletters, eight webinars with more than 2,000 combined viewers, and grew the number of their combined membership to more than 3,000 individuals, with approximately 80% identified as external to CGIAR. The Ontologies CoP made significant advancements in the development and adoption of crop ontologies in 2019. The progress was propelled by an online curation tool that makes the development and sharing of reference ontologies to describe agronomic phenotypes and the accurate mapping of these into databases easier. These critical steps enable comparative genotypic and phenotypic studies as well as gene-discovery experiments. The tool advanced with significant input from CGIAR Research Programs, the US National Science Foundation, allied research organizations, and private partners, such as PepsiCo.</p>
<p>M3: Inspire</p>	<p>In 2019, the Platform awarded four start-up grants of 100,000 USD each. Four winning projects from 2018 and 2017 that demonstrated exceptional results, proven viability, and the potential for impact were awarded a total of 525,000 USD in scale-up funds. The eight winning teams included collaborations between a total of nine CGIAR Centers and Research Programs and 18 diverse external partners, including start-ups, governmental bodies, universities, and private sector businesses. The Inspire Challenge has begun to field test innovations that demonstrate new digital pathways to impact for CGIAR research, foster whole system agility and adaptation through digital tool use. These include: PlantVillage Nuru is transforming farm-level pest and disease monitoring by using AI. The free-for-download phone application works offline and has proved twice as accurate in its diagnosis as extension workers. The app has been downloaded by users on all developing continents, generating more than 18,000 reports – which are generated each time the app is used correctly -- from users in more than 40 countries. The project is expanding from cassava diagnostics to the ability to diagnose additional crop diseases. An accurate model for potato diseases, for example, is expected to reach 200,000 farmers in India in 2020. One project has put near-real-time data in the hands of fisheries officers, researchers, and local stakeholders by creating an automated data pipeline to highlight temporal and spatial changes in fish production. The dashboard has one of the most sophisticated data collection systems for small-scale fisheries in the world. In 2019, roughly 300 boat tracking units were deployed on fishing boats around Timor-Leste and 11 new local data collectors were trained in aforementioned coastal communities. The project is currently expanding to monitor small-scale fisheries in Egypt, Nigeria, and Zambia, with plans to scale up in Bangladesh, Malaysia, and Malawi. The existing data systems underpinning this project follow CGIAR gender-disaggregated data standards and have already highlighted the critical importance of women in fishery activities as dependable sources of household food and nutrition security in Timor-Leste. Another project is monitoring and informal food flows through free Wi-Fi in and among five traditional markets in Hanoi, Vietnam. In 2019, the team collected data from more than five million smartphone devices to analyze and leverage it for improving market policies and food safety. Their efforts to gender disaggregate the data has revealed 56% of their users are women. The project has been adopted by the General Statistics Office of Vietnam as a key method for assessing and predicting unfolding COVID-19-related food security shocks on a national level. Another project is estimating rainfall in developing nations through the unprecedented use of commercial microwave links (CMLs).</p>

**M3:
Inspire**

The team's peer-reviewed research has demonstrated the potential spatial advantage of using CMLs for monitoring rainfall, which is based on measurements in an agricultural area near Kericho, Kenya. CMLs are proving to be a sustainable, long-term solution for generating rainfall estimates in developing countries -- a critical component for improving crop yield monitoring and designing better rainfall-based index insurance. Other notable emergent innovations from the 2019 Inspire portfolio include:

CubicA, which piloted a digital-advisory system that reached 3,000 smallholder banana farmers in Uganda.

Smart Seed Selection successfully developed a machine-learning-driven process for the selection of the best maize seeds for expected productivity in specific environmental conditions.

Seeing is Believing improved a method for remote monitoring crop phenology that leverages smartphone images and links these to yield predictions.

The Challenge attracted some 200,000 USD in external funding commitments in 2019, signalling it could well become a signature digital innovation process for CGIAR.



Platform for
Big Data
in Agriculture

