CGIAR

**Platform for
Big Data
in Agriculture**

CGIAR PLATFORM FOR
BIG DATA IN AGRICULTURE
ANNUAL REPORT
2018

CIAT
International Center for Tropical Agriculture
Since 1967 Science to cultivate change

INTERNATIONAL
FOOD POLICY
RESEARCH
INSTITUTE
IFPRI

CONTENTS

EXECUTIVE SUMMARY

The Platform is driving the digital transformation of CGIAR and agricultural development around the world through data sharing, development of new multi-stakeholder collaborations, and supporting innovations that change the game for food security. The Platform calls these goals "Organize", "Convene", and "Inspire" and each broke new ground in 2018 to build a more integrated, cohesive digital vision for CGIAR.

ORGANIZE

BIG DATA launched the Global Agricultural Research Data Innovation & Acceleration Network (GARDIAN), a pan-CGIAR data search and discoverability portal. For the first time, datasets, publications, and crop varieties across all 15 centers and 11 genebanks of CGIAR can be easily found. The Platform helped centers enhance data management practices, standards, and tools, and capacity to deliver Findable, Accessible, Interoperable, and Reusable (FAIR) research data, resulting in nearly 100,000 publications and 3,000 datasets in GARDIAN. The Platform developed pipelines to streamline entry of well-described data into GARDIAN and leverage it into crop models and spatial analysis.

CONVENE

The Platform Communities of Practice (CoP) reached several hundred members in 2018. When the European Union's General Data Protection Regulation went into effect, the Platform and the CGIAR System Management Office (SMO) developed an online course to help agriculture development researchers understand its wide-ranging implications for legal and ethical use of research data. The "Convene" module culminated in the Annual Big Data in Agriculture Convention, co-hosted by the International Livestock Research Institute (ILRI) and World Agroforestry (ICRAF), with 400 attendees, 2500 remote participants, and millions of social media views. The Convention hit its target of 60% non-CGIAR participants, with global IT firms, startups, governments, and other research institutions forging new partnerships and seeing the breadth and depth of CGIAR research.

INSPIRE

In 2018, the Platform made five startup grants to data-driven projects under our innovation process, the Inspire Challenge, and scale up awards were granted to three projects from the 2017 cohort. *Seeing is Believing* integrates machine-learning analysis of cellphone camera images of crops into an insurance product and farm advisory service. A Randomized ControlLED Trial in 20 villages demonstrated significant improvements in farmers' practices, willingness to pay premiums, and the overall cost-effectiveness of the insurance. Another project, *Real-Time Diagnostics for Devastating Wheat Rust* demonstrates that a hand-held gene sequencer used in Ebola diagnosis can be used in the field to identify a particularly devastating strain of wheat rust—dramatically enhancing crop disease surveillance and reducing the cycle time of response. The *Analysing Livestock Social Media Data for Farmer Chatbot* project grew its user base from around 30,000 at launch to over 120,000 farmers in 2018, of which 40% are youth. This project enables animal health researchers to reach 24,000 dairy farmers, 92% of whom report improving their farming practices as a result of what they learned through the service.

The Platform studied the innovation strategy of the Inspire Challenge and improved the application process to source fewer "basic research" innovations. As a result, over 90% of submissions in 2018 targeted small producers as users of data innovation, an achievement in line with our aspiration to solve agricultural problems better, faster, and at greater scale.

1 KEY RESULTS

1.1 HIGHLIGHT PLATFORM ACHIEVEMENTS

This was the second year of implementation for the CGIAR Platform for Big Data in Agriculture, and the first full year of implementation with the leadership team in place and governance system fully functioning. As per the programmatic Theory of Change (ToC), the Platform focused 2018 on building new data-driven capabilities across CGIAR, with special emphasis on unlocking CGIAR data for impact.

The Global Agricultural Research Data Innovation Acceleration Network (GARDIAN) enabled discovery and access to nearly 100,000 publications and 3,000 datasets and crop traits in 2018, bridging data repositories across the 15 centers and 11 genebanks of CGIAR. Platform support helped centers to significantly enhance data management and upload to institutional repositories. This was achieved through improved capacity to deliver open and FAIR data and co-development of data management standards and tools.

BIG DATA developed CGIAR as a learning organization through support for data sprints at centers, shared services, CoPs, and new high-level data products. It positioned CGIAR as a go-to institution for digital agriculture through new partnerships and an annual Convention. Some key collaborative products developed include a design of 100 questions (a harmonized set of common questions in socio economics research), an inventory of standard vocabularies for livestock research, and a CGIAR-wide analysis of gaps in modeling capabilities. Additionally, trainings were held on best practices in data management, data mining, agrisemantics, image analysis, and compliance with privacy and ethics guidelines. The first prototype of a shared image processing service for drone imagery was developed for rice and cassava agronomy and breeding use-cases. The Platform's activities culminated at the Big Data in Agriculture Convention in Nairobi with 400 attendees, 2,500 remote participants, and millions of social media views. The Platform led an organization-wide assessment of digital strategy and developed an action plan for CGIAR's digital transformation.

CGIAR's signature digital innovation process, the Inspire Challenge, continued to de-risk

innovative projects aligned with CGIAR Research Programs. The Platform awarded five pilot grants to data-driven startup projects in 2018, and three scale-out projects from the 2017 cohort that demonstrated early-impact change in food and farming systems during the year. The Inspire process has started to deliver data-driven solutions to development. For example, an innovative AI-based chatbot advisory service for livestock farmers in Kenya has over 200,000 beneficiaries, and field-based sequencing of wheat rust across Ethiopia informed extension systems to respond faster and more effectively to potentially devastating biotic threats.

1.2 PLATFORM PROGRESS TOWARDS OUTPUTS AND OUTCOMES

1.2.1 OVERALL PLATFORM PROGRESS

The CGIAR Platform for Big Data in Agriculture seeks to drive digital transformations in the agriculture development sector by leveraging CGIAR's global footprint, expansive partnership networks, and deep food security expertise. In 2018 the Platform broke new ground, creating new data-driven capabilities for both our organization and our sector. We built new digital partnerships and alliances, and developed digital innovations that could change the game for global food security. We call these goals "Organize", "Convene", and "Inspire."

Under "Organize", the Platform launched the Global Agricultural Research Data Innovation Acceleration Network (GARDIAN), a pan-CGIAR data search and discoverability portal. For the first time, datasets, publications, and crop traits became discoverable and easily accessible in one portal, regardless of where they were archived across CGIAR's 15 centers and 11 genebanks. Centers nearly doubled the open datasets and publications indexed by the portal (from 50,000 publications and 1,800 datasets in December 2017 to over 96,000 publications and 2,800 datasets in December 2018). CG Core Metadata Schema V.2 was released through the collaboration of all centers, providing a standard for data annotation across repositories and software tools, improving their discoverability and applicability for analysis. A stable Agronomy Ontology spearheaded by Bioversity International and the International Food Policy Research Institute (IFPRI) was released and incorporated into the Agronomy Field Information Management System (AgroFIMS). AgroFIMS is a system for enabling electronic data collection tools to use standard ontologies and vocabularies—improving data quality and interoperability at the source. This effort involved the International Potato Center (CIP), Bioversity International, IFPRI, the International Maize and Wheat Improvement Center (CIMMYT), the International Institute of Tropical Agriculture (IITA), the International Center for Tropical Agriculture (CIAT), the University of Florida, University of California Davis, Rothamsted Research, Ontocale Inc., and Diversity Arrays Technology (DArT). A mature Crop Ontology with new species and trait classes was released, as well as an early draft Socioeconomic Ontology—the former through collaboration led by Bioversity International, and the latter led by CIMMYT and IFPRI. The

development and use of these ontologies enables the cross-domain data querying and exploration necessary for forming and addressing complex research questions across the organization.

The Platform helped CGIAR develop as a learning organization and progressively claim a voice of authority in digital agriculture development under "Convene". When new privacy regulations went into effect in the European Union, the Platform launched online training to help agriculture development researchers understand the wide-ranging implications for their work, and developed high-level guidelines on responsible data management. Six Platform CoPs continued to help define the technical agenda for digital agriculture, and reached several hundred members during the year. Some key collaborative products developed in 2018 include: design of 100 questions (a harmonized set of common questions in socio economics research); an inventory of standard vocabularies for livestock research (necessary for development of a livestock ontology); and a CGIAR-wide gap analysis of modeling capabilities. Additionally, trainings were held for best practices in data management, data mining, agrisemantics, image analysis, and compliance with privacy and ethics guidelines. The first prototype of a shared image processing service for drone imagery was developed for rice and cassava agronomy, and breeding use-cases.

"Convene" culminated in the Annual Big Data in Agriculture Convention, 'Decoding the Data Ecosystem', with 400 attendees and 2500 remote participants, and several million social media views. The Convention hit its target of 60% non-CGIAR participants and demonstrated new ways for private sector, governments, and other research institutions to engage while showcasing the breadth and depth of CGIAR research.

CGIAR's signature digital innovation process, the Inspire Challenge, continued to de-risk innovative projects aligned with CGIAR Research Programs. The Platform awarded five US$100,000 grants to data-driven startup projects in 2018, and a total of US$500,000 in scaling funds to three projects from the 2017 cohort that demonstrated they were creating positive change in food and farming systems. The Platform evaluated and improved the submission process to better target innovations. "Basic research" submissions dropped significantly from 2017, and those targeting smallholder farmers went from 18% to over 70%, consistent with the Platform's aspiration to help CGIAR and the sector solve agricultural problems faster and at a greater scale.

The Platform aims to effect several digital transformations both inside and outside CGIAR by mobilizing CGIAR data to accelerate research, spur new data-driven innovations, build collaboration across our organization and with the wider sector, and leverage CGIAR expertise while claiming a leadership voice in digital innovation for agriculture. To begin to achieve this, the Platform—with help from Accenture Development Partnerships—led an expansive assessment of the state of digital strategy in the organization. Together, a high-level action plan was developed to build a unified digital vision, address leadership, data access and use,

the role of partnerships, and the right information infrastructure to help the organization deliver on reducing malnutrition, poverty, and building sustainable ecosystems worldwide.

1.2.2 PROGRESS BY MODULES

Module 1 - ORGANIZE progress

The Platform continued to help CGIAR mobilize its data to accelerate research and build more data-driven capabilities across the organization. Collaborative projects were undertaken to select, download, clean, annotate, and process datasets in GARDIAN toward the outcome of a common analytics environment that envisions seamless analysis, decision support via model-ready data, and aggregated CGIAR data products of value.

The Platform supported several key multi-center working groups within the CGIAR Data and Information Management CoP to facilitate progress and co-learning towards making data sets FAIR and publishable to open repositories. These include the Metadata and Ontology Working Groups led by Bioversity International, the Dataverse Working Group led by CIMMYT, CIAT, and IFPRI. Their activities are essential to making data FAIR, for example, by developing the CGIAR-wide CG Core metadata standard through collaborative ontology and tool development to apply these standards and enable CGIAR-responsive data repositories.

With involvement from several centers, the Platform developed guidance for the organization on complying with privacy and ethics considerations in research data management. Additionally, the Platform developed algorithms to help detect personally identifiable
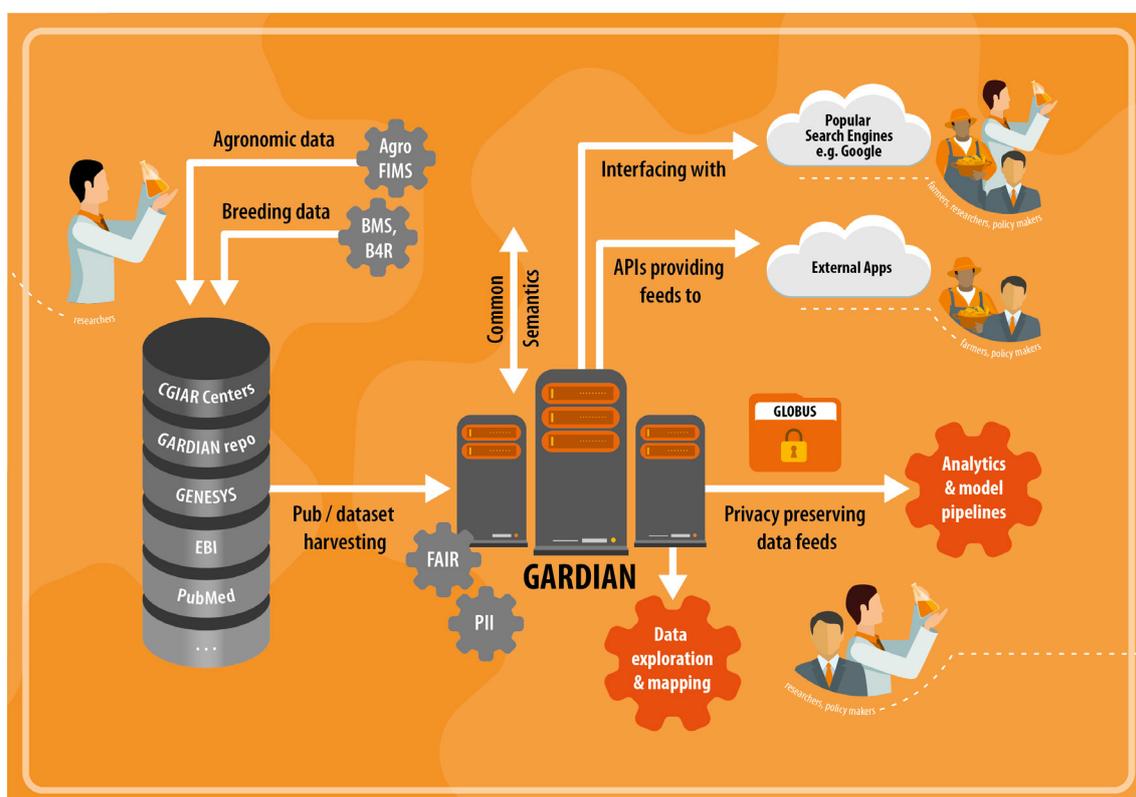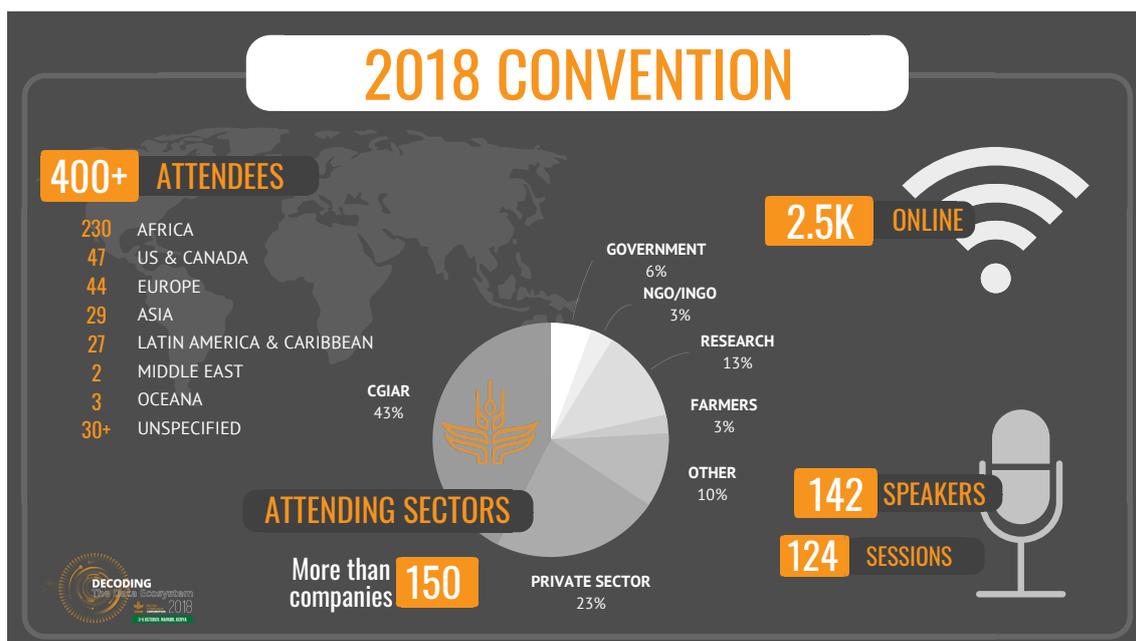
information in CGIAR repositories, facilitating regulatory compliance to reduce privacy-related organizational risk. CGIAR's FAIR assets are showcased through GARDIAN, and datasets and publications discoverable through GARDIAN almost doubled in 2018, with several new geospatial, filtering, and querying/data mining features implemented.

In addition to enhancing data annotation and upload, centers used Platform support to develop policies, trainings, and tools on topics including: semantics for data annotation; metadata entry and compliance tools; licensing, and best practices for privacy/ethics compliance in research data management. Centers organized 34 trainings and 18 data sprints through Platform support in 2018, with almost 40% of the 1,000+ uploaded datasets a direct result of these data sprints.

Module 2 - CONVENE progress
The Platform helped CGIAR claim a voice of authority in digital agriculture. Together with SMO, the Platform launched an online training platform and course to help international agriculture development researchers understand the implications of the General Data Protection Regulation for their work.

Six Platform CoPs continued to help define the technical agenda for digital agriculture and grew to several hundred members. The Ontologies CoP organized PhenoHarmonIS, the reference event for harmonizing data standards and reference ontologies for food security. 200 attendees from public, private, and non-profit partners were trained in the use of Crop Ontology, Agronomy Ontology, Plant Stress Ontology, co-design of the AgroFIMS tool, the Breeding API, Minimum Plant Phenotypic Information for Evaluation (MIAPPE) and other phenotypic standards.

2018 CONVENTION

400+ ATTENDEES

230	AFRICA
47	US & CANADA
44	EUROPE
29	ASIA
27	LATIN AMERICA & CARIBBEAN
2	MIDDLE EAST
3	OCEANA
30+	UNSPECIFIED

2.5K ONLINE

ATTENDING SECTORS

More than **150** companies

CGIAR 43%
PRIVATE SECTOR 23%
GOVERNMENT 6%
NGO/INGO 3%
RESEARCH 13%
FARMERS 3%
OTHER 10%

142 SPEAKERS
124 SESSIONS

DECODING the Data Ecosystem 2018

The Platform developed shared services in support of CGIAR research. The subscription to high resolution satellite imagery continued through Digital Globe. The Platform sourced datasets—including gridded weather data from IBM and high-resolution population data from Oak Ridge National Lab—to further support CGIAR's geospatial analysis capacity. The Platform activated a subscription to Globus, a grid computing alliance providing tools for secure management of data access rights, sharing, and transfer of large datasets. The Platform is testing Globus to determine if it can become a pan-CGIAR data infrastructure.

Convene culminated in the Annual Big Data in Agriculture Convention, co-hosted in Nairobi by ICRAF/ILRI. It was an opportunity to highlight CGIAR data-driven research and the role of animal science in food security. The event had 400 attendees, 2,500 remote participants, and several million social media views, and hit its target of 60% non-CGIAR participants. It demonstrated new ways for private sector, governments, and research institutions to engage while showcasing the breadth and depth of CGIAR research.

Module 3 - INSPIRE progress

In 2018 the Platform made five pilot grants to projects under our innovation process, the Inspire Challenge, and a total of three scale up awards to three projects from the 2017 cohort which generated solid evidence while piloting innovative digital business models. The five 2017 awardees of pilot grants demonstrated how data-driven innovations create new pathways to impact. Three pilot projects from this first cohort were able to demonstrate quantifiable early results and promising partnerships for achieving greater scale:

Seeing is Believing: a project that integrates machine-learning analysis of cellphone camera images of crops into the functioning of an insurance and farm advisory service.

Early results from a randomized trial demonstrated advisory messages, which reached 32,237 wheat producers, increased knowledge on best agricultural practices by 78%. The trial showed respondents were willing to pay an extra 8.7% of their insurance premium when picture-based assessment was embedded in the insurance application.

MARPLE: *Real-Time Diagnostics for Devastating Wheat Rust*, demonstrated that a hand-held gene sequencer could be used in the field to identify a devastating strain of wheat rust within 48 hours—dramatically enhancing crop disease surveillance and reducing the cycle time of response. The pilot demonstrated the first accurate identification of individual yellow rust strains as the disease was starting to emerge. The data has been incorporated into Ethiopia's early wheat rust warning systems.

Farm.Ink: *Analyzing Livestock Social Media Data for Farmer Chatbot*, grew its user base from 30,000 to over 120,000 farmers (40% of them youth), enabling animal health researchers to reach 24,000 dairy farmers. A survey of 400 dairy chatbot users showed 92% improved their farming practices as a result of the service. The process has proven valuable for sourcing high-quality digital innovations and generating new evidence on their effectiveness.

1.2.3 VARIANCE FROM PLANNED PROGRAM FOR THIS YEAR

A) Have any promising research areas been significantly expanded?

Complying with privacy recommendations in research data is a relatively new area of focus and challenging as the environment and regulations are still evolving depending on location. It became clear through center consultations in 2018 that the way centers deal with personally identifiable information and human subjects (e.g. for surveys) varies substantially. The Platform invested Module 1 and 2 resources in addressing these issues in a consultative fashion with center staff from funds saved when other activities (e.g. collaborations with Accenture Development Partnership and the Minnesota Supercomputing Institute) proved not as expensive as anticipated.

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

Although an agreement was developed between the Platform and the University of Minnesota's International Agroinformatics Alliance to enable collaboration and leverage their GEMS Platform on effective handling of sensitive data to enable analysis, this work was deferred to 2019. Several conversations did occur on how the two Platforms might move forward. A joint concept note was developed around semantic standards to harmonize data for easier analysis, but these activities will be taken further in 2019. The funds earmarked for this activity were reallocated towards a user workshop for the Agronomy Field Information Management System (AgroFIMS) and staffing a consultant to address privacy and ethics concerns.

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

No modules or research areas have significantly changed.

1.2.4 ALTMETRIC AND PUBLICATION HIGHLIGHTS

The CGIAR Platform for Big Data in Agriculture 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 Platform's CoPs published several publications for the Platform in 2018, recording a combined altmetric of 148. The Data-driven Agronomy CoP published two position papers with non-CGIAR international partners, one with IBM ('*How CGIAR and CIAT are helping to bridge the human capital divide and grow the digital revolution in the developing world's agriculture industry*') and one with University of British Columbia (positional paper published in Nature stating that '*Smallholders need access to big-data agronomy too*'), achieving the highest altmetric of 75. The Crop Modeling CoP produced a partial report that contains the gaps and expertise of CG centers in terms of modeling capacity identified during 2018. The Ontology CoP contributed to the peer-reviewed article AgBioData consortium recommendations for sustainable genomics and genetics databases for agriculture. Finally, the Geospatial CoP has been working on the magazine article '*Precision agriculture for smallholder farmers*'.

Building the brand of the Platform and establishing CGIAR as an authority in the digital agriculture space was a key priority for 2018. Analysis of social media platform statistics and Google Analytics showed significant audience growth across all publishing platforms. In 2018, we published 126 blogs and videos on the website. This attracted a 300% (20,000) increase of new users and over 200% increase (70,000) in page views.

Our social media channels (Twitter, Instagram, LinkedIn, YouTube and Facebook) recorded an average increase of 500% across all platforms, translating to a 200% increase of socially-driven website traffic, when compared to 2017. The most notable social media result of 2018 was during the annual Convention when we surpassed the audience reach of the entire 2017 convention period during the first four hours of the 2018 event. Over the course of the convention we achieved the top trending twitter topic in Kenya, serving content to a potential 25.3 million Twitter users globally.

In 2018 the Platform was featured in a variety of global news outlets, most notably Nature, Reuters (Foundation), Food Tank and over 30 prominent media outlets in Serbia (state television, agricultural TV and breakfast shows, the most-read daily newspaper and weekly magazines). The combined monthly audience of these outlets, using online tool SimilarWeb, is conservatively estimated at over 17 million.

1.3 CROSS-CUTTING DIMENSIONS (AT PLATFORM LEVEL)

1.3.1 GENDER

A) List any important PTF research findings

Explicit attention was given to encouraging women to attend all capacity development events and Annual Big Data in Agriculture Convention sessions, where 30% of speakers were women. At the Annual Convention, total female participation was 34.9%, reflective of the often male-dominated subject matter. Efforts will be made to improve on this in 2019.

Proposal entries for the 2018 Inspire Challenge were explicitly assessed on the gender dimension, with a rubric and scoring matrix that evaluated if and how proposals dealt with gender issues. Thus, effective and persuasive attempts to include the gender dimension in Inspire grant proposals was actively recognized and rewarded. The fruits of this effort will be awarded during implementation in 2019.

Attempts were also made to explore collaboration with representatives of the CGIAR Collaborative Platform for Gender Research to assess if and how GARDIAN might assist with the Platform's efforts to identify datasets with disaggregated gender outputs. Currently, 7,060 publications related to gender are discoverable through GARDIAN, and 124 datasets have an explicit gender-related dimension. As centers manage the datasets made discoverable through GARDIAN, this dimension likely needs more attention at the center level. However, the Platform will play an important role in facilitating the standards of development and data management practices to enable this.

Towards the goal of reducing the gender data gap, the BIG DATA and Gender Platforms conducted a design session at the Big Data in Agriculture Convention to assess how new data types might be leveraged beyond surveys. For example, the session discussed the possibility of working with satellite imagery or mobile network metadata to reveal gender in food systems in new ways, and plan to launch this research in 2019.

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

Since various centers manage the datasets made discoverable through GARDIAN, annotation of these resources to clearly indicate a gender component is not always consistent or adequate. We're attempting to make this easier by encouraging tagging of research resources with appropriate keywords indicating these cross-cutting dimensions. An in-development data management course slated for 2019 will also include guidance on ensuring good practices on annotation of research resources using gender and other cross-cutting dimensions as examples.

The joint BIG DATA - Gender Platform design process during the 2018 Convention unearthed the presence of multiple country data sets on women's economic

empowerment in agriculture from the USAID Feed the Future program. This will provide a good indication of where to look for proxies in the non-traditional data, and could be an opportunity for the Platform to contribute more meaningfully to the gender dimension in 2019.

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

The Platform is not technically a research delivery program, though we do seek ways to build new knowledge and experience within CGIAR about new ways of doing business. This is not a problem *per se*, but an evolving understanding of our own role vis a vis the Gender Platform.

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

GARDIAN provides access to a total of 1,399 publications which address youth issues, but only three CGIAR datasets. This demonstrates the need to integrate youth more effectively into CGIAR research, and the Platform will promote inclusion of youth-related standards in data uploaded to CGIAR repositories. A concrete step towards this in 2018 was to integrate youth related indicators into the 100Q initiative, ensuring that future CGIAR surveys adequately address youth and social inclusion in the data that is collected.

The Platform created the Youth In Data Initiative, and engaged a group of young digital innovators from Africa and provided them with training on social media and in journalistic data reporting. A group of 12 of these digital innovators (half from CGIAR institutions and half from other youth groups in Kenya and Nigeria) then participated actively in the Annual Big Data in Agriculture Convention, building a youth outreach platform connecting with other innovators and enabling them to interact with some of the key institutions and experts in the digital agriculture space. The Convention also attracted many young, avid agricultural data scientists in the region and also from around the world. The youngest speaker was Lillian Petersen, a 15-year-old high school student from New Mexico, USA, presenting her Africa-wide food security monitoring analysis using remote sensing data. Lillian's work was first developed

as a school project, further developed with new research directions and ideas from others in the field including geospatial scientists in CGIAR, and later published as an academic journal article.

Innovation projects supported by the Platform are yielding new data and (anonymized) intelligence and insight into youth user behavior of these digital tools. For instance, the Inspire awardee *Farm.Ink: Analyzing Livestock Social Media Data for Farmer Chatbot*, reached over 120,000 farmers in 2018, 44,000 (almost 40%) of whom are estimated to be between 15-24 years old. Likewise, 28% of beneficiaries of the Seeing is Believing innovation in the insurance industry were youth. The Platform is actively examining how to systematically learn how to target youth in innovation processes, and leverage this opportunity to engage youth more effectively in agriculture.

A) List any important PTF research findings

Not applicable; the Platform is not a research delivery program; rather it supports delivery by other programs.

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

Innovation projects supported by the Platform are yielding new data and (anonymized) intelligence and insight into youth user behavior of digital tools and farmer-facing apps, and suggest that 40% or more are young users. The Platform has begun to explore opportunities to learn more on how youth may be more effectively engaged in the agricultural sector.

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

No problems have arisen in relation to integrating youth. The user-base of digital innovations tends to be youthful, and we have thus far leveraged this to develop active youth collaborations on communications and promotion in relation to the Convention. The Platform will continue to examine ways to build on the opportunity to more actively engage youth.

1.3.3 CAPACITY DEVELOPMENT

The Platform supported capacity enhancement across CGIAR and its partners in a variety of ways, including best practices in data management throughout the data life cycle, launching an online learning platform and development of an online course on the General Data Protection Regulation (GDPR), and a series of webinars on digital extension, interactive voice response, privacy guidelines and more. Fund disbursement to centers through Module 1 included guidelines with capacity enhancement as an explicit outcome, and suggested activities including trainings, hackathons, workshops, webinars, and data sprints to improve

data annotations and uploads.

Almost every center achieved the milestone of at least two training events in 2018, with a total of 34 trainings and 18 data sprints organized. Centers reported that almost 40% of the 1,000+ datasets uploaded to institutional repositories were a direct result of these data sprints. Working Groups and CoPs under Modules 1 and 2 involving all centers also coordinated several capacity development efforts across a wide range of topics towards building readiness to leverage big data capabilities. Workshops helped to strengthen capabilities and collaborations for the use of semantics in harmonizing breeding, agronomic, and agro-ecological data. Other workshops focused on data science, licensing, ontology and metadata how-to's, and identifying a set of core questions common to most CGIAR surveys.

The Big Data in Agriculture Convention included several training/capacity building sessions: R for image analysis, good practice in electronic data collection, introduction to new technologies, overview of innovation strategy, introduction to human-centered design, and more. Shared services were established which enhanced the research capacity of centers (e.g. access to Digital Globe satellite imagery, VIAMO voice response system, and IBM/TWC and aWhere global historic weather data).

1.3.4 CLIMATE CHANGE

The platform did not develop or support any research specifically addressing climate change in 2018, although its leadership in and enabling of digital agriculture is expected to accelerate the ability of CGIAR and its partners across the agricultural research sector to effectively tackle this challenge.

2 EFFECTIVENESS AND EFFICIENCY

2.1 MANAGEMENT AND GOVERNANCE

Throughout 2018, the Management Team and Steering Committee met monthly for two-hour teleconferences to ensure a smooth first full year of implementation. Meetings are fully documented with annotated agendas, proposed decisions and subsequently the minutes are finalized. It was considered important to have monthly meetings during the startup phase for the Platform, as many management and governance-relevant decisions were being made on a regular basis. The Independent Steering Committee, as per the proposal, comprises three independent members (from IBM, University of Florida and Bill & Melinda Gates Foundation), plus a representative from the lead centers and a representative each from the CRPs and from the centers. The arrangement is working well.

The Platform, as per the proposal, constituted in 2018 an entirely external International Advisory Board comprised of members of the Food and Agriculture Organization, the Global Open Data for Agriculture and Nutrition Alliance (GODAN), Google, Mars Inc., GIZ, and Ag Gateway. The function of this Board is to keep the Platform apprised of efforts in the larger digital agriculture landscape to either synergize with or differentiate the Platform as we strive to drive digital transformation and build new relevance for CGIAR. This Advisory Board met face-to-face once, and in 2019 will hold two meetings.

2.2 PARTNERSHIPS

2.2.1. HIGHLIGHTS OF EXTERNAL PARTNERSHIPS

Broadening CGIAR's partnership network around digital approaches is at the core of the Platform's efforts, and the primary focus of Module 2: Convene. During 2018, a number of high-profile partnerships were established. The Platform engaged Microsoft Research on the collaborative design of the Internet of Things network for agriculture. Accenture

Development Partnerships provided key information on global trends in digital agriculture, as well as evolving best practice on data science and how organizations are leveraging it for organizational change. IBM Weather Company provided a research subscription to validated global weather data and invited CGIAR researchers to be early user-validators of a new global information service centered on short term (six month) seasonal probabilistic forecasts. Google, GODAN, and Mars Inc. joined the BIG DATA Platform Advisory Board to help inform and shape overall digital transformation in CGIAR and help keep CGIAR apprised of trends in the sector. A strategic partnership with Digital Globe enabled centers to access high resolution satellite imagery, AWhere provided access to site-specific management information, and ESRI provided geospatial solutions for all 15 centers.

Several other partnerships across the public and private sectors were built or enhanced in 2018 to augment CGIAR's capacity in a variety of areas, including agrisemantics (Earlham Institute), data science and analytics (University of Minnesota), data discovery (University of Florida), mining (UCDavis), and visualization (AgroKnow), digital data collection, and improving CGIAR's ability to deal with data privacy and ethics.

The Annual Big Data in Agriculture Convention in Nairobi had over 150 different partner participants at the event, with 23% participation coming from the private sector, 13% from university systems, 6% from government, and 16% civil society. The broad external participation is consistent with the Platform goal to effectively place CGIAR in the center of the dialogue around digital innovation in agriculture.

2.2.2. CROSS-CGIAR PARTNERSHIPS

The Platform co-designed 'big data-enabled' gender research with the Gender Platform, hosted six CoPs on technical big data themes with hundreds of participants from centers and CRPs. Additionally, the Platform offered cost-share on six CRP technical proposals (of which one was selected for funding), supported development of a drone image processing pipeline for the CGIAR Research Program on Rice (RICE), and the CGIAR Research Program on Roots, Tubers and Bananas (RTB) (centered on rice and cassava), and facilitated seasonal forecast data from IBM Weather Company to the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). The Platform identified priority topics for digital innovation with CRPs and issued calls for new INSPIRE data innovations aligned to CRPs' needs. Ultimately, eight of these were funded in 2018 that align to the CGIAR Research Program on Livestock (LIVESTOCK), CGIAR Research Program on Fish (FISH), CGIAR Research Program on Wheat (WHEAT), CGIAR Research Program on Policies, Institutions and Markets (PIM), CGIAR Research Program on Agriculture for Nutrition and Health (A4NH), CCAFS, RTB, and MAIZE. The Platform also collaborated with GENEBANKS around benefit sharing for the ITPGRFA in the digital era.

The Data and Information Management Community of Practice, while not a formal CoP, operates under the aegis of and with support from the Platform, and consists of representatives from every center working collectively on a number of groups to consistently address common needs (e.g. metadata schema, repository platforms, data annotation and more). The Agronomy Field Information Management System (AgroFIMS) is a cross-center effort led by CIP, Bioversity International, and IFPRI, with strong cross-sectoral collaboration from CIMMYT, IITA, CIAT, the University of California Davis, the University of Florida, Rothamsted Research, Ontocale, Inc. and Diversity Arrays Technology (DArT).

2.3. INTELLECTUAL ASSETS

Have any intellectual assets been strategically managed by the PTF (together with the relevant Center) this year?
The Platform's intellectual assets are generally Open Access and any code for tools/software is Open Source (available via GitHub). The suite of Creative Commons licenses are used for Platform assets and centers are consistently encouraged and supported in adopting these licenses as well. For instance, ontologies developed as Platform deliverables are collective knowledge, created and shared by members of CGIAR CoPs, and are published as global public goods under CC-BY licensing. Software, databases, and other relevant code is managed for clear usage rights through the industry standard (MIT or GNU licensing).

Key assets managed by the Platform in 2018 include all ontologies developed, intellectual property as it intersected with the Inspire Challenge, code developed for key Platform tools (e.g. AgroFIMS and GARDIAN), and training content. The Inspire Challenge specifies that any proprietary IP of the partners that precedes their partnership under Inspire may remain proprietary, and that any new IP developed under the Inspire Challenge should be made FAIR.

Indicate any published patents and/or plant variety right applications (or equivalent)
No patents published; the plant variety rights area is not directly relevant in the work of the Platform.

List any critical issues or challenges encountered in the management of intellectual assets
No critical issues or challenges were encountered by the Platform in the management of intellectual assets.

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

To catalyze the digital transformation of the agriculture for development domain, the Platform conducted an assessment of the state of digital strategy and readiness across CGIAR. A key learning outcome was the development of a high-level action plan addressing how CGIAR can: claim a lead role in articulating and implementing a digital vision; improve data access and usage gaps, needs, and solutions; strengthen the role of partnerships; and appropriate infrastructure needs to help the System deliver on its strategic research goals.

The Platform studied the Inspire Challenge process of 2017 in light of its innovation strategy, examining how the process could better target smallholder farmers with game-changing innovations. The resulting synthesis paper informed some changes to the 2018 application process (namely requiring more explicit problem definition, identification of domain knowledge to solve the problem, and intended target beneficiaries). "Basic research" proposals declined significantly and over 90% of proposals targeted small producers (up from just 18% in 2017).

2.5 EFFICIENCY

The Platform better targeted the Inspire Challenge innovation process to source more innovations for reaching smallholder farmers, in accordance with the overall goals of the Platform. Data indexing has been automated across all CGIAR centers. A standards-based agronomy field book template and digital data collection system was user-tested. This was done in preparation for large-scale field testing to improve data collection and interoperability from the moment of collection. The Platform harmonized data curation methods and standards across all centers, as embodied in the new CG Core Metadata Schema v.2.0. Additionally, the Platform harmonized collaborative improvement and adoption of semantics and repository choices through the working groups of the Data and Information Management CoP.

2.6 MANAGEMENT OF RISKS TO YOUR PLATFORM

At the design phase, the Platform anticipated that privacy and responsible management of farmer data would become a charged issue. In 2018 the General Data Protection Regulation came into effect even as several high-profile breaches of data privacy appeared in the news. This posed a legal and reputational risk to both the Platform and the System. The Platform developed and launched an online course and webinar to help CGIAR centers and researchers understand the implications of this, and with the help of three centers developed a machine-aided method to flag the presence of personally identifiable

information in CGIAR data repositories, helping the Platform and the System to avert these risks.

Platforms support the delivery of the CGIAR research portfolio and, as a result, fulfilling this mandate requires gaining a complete understanding of CGIAR's needs with respect to digital agriculture in terms of vision and strategy, data access, partnerships, and information infrastructure—or risk irrelevancy. To manage this risk the Platform led, with help from Accenture Development Partnerships, an assessment of digital strategy and readiness across CGIAR and centers and arrived at a high-level action plan proposing ways to address gaps. This ensures that the Platform is driven by center needs and can help the CGIAR System deliver on its strategic research goals.

2.7 USE OF W1+2 FUNDING

Over 95% of the Platform budget was W1+2 as centers are not being requested nor incentivized to align bilateral activities to the Platform. Hence, practically all results reported in this Annual Report can be attributed to W1+2 funding.

Through W1+2 grants to all Centers the Platform supported updating of all data repositories and mobilizing new open data across the whole organization. For the first time CGIAR open data and publications are keyword searchable and supporting ever more sophisticated searches and visualizations. The Platform continued to support six cross-cutting CoPs on deep technical themes, to develop new shared services for secure data sharing and access to data services, and awarded US$1 million in innovation funding to CRP-aligned innovation projects.

3 FINANCIAL SUMMARY

2018 was the first full programmatic year of the BIG DATA Platform and, as a result, began the year with significant carryover from 2017 (US$4.4 million). The Platform transitioned into full implementation in 2018, executing US$7.6 million, and ending the year with a much-reduced carryover (compared to 2017) of 20.1%. It is important to note that around half of this carryover is due to US$1 million in grants awarded under the innovation process under Module 3. Selection for which only happens during the fourth quarter of the fiscal year for projects that will be implemented in 2019 and 2020 – contributing to significant carryover under Module 3. Setting aside these grants under Module 3, the Platform carryover is around 10%, which is fully committed for 2019. It is expected that the Platform will fully execute the budget for 2019.

TABLES

TABLE 2: Condensed list of policy contributions in this reporting year (Sphere of Influence)

N/A.

TABLE 3: List of Outcome/ Impact Case Reports from this reporting year (Sphere of Influence)

N/A.

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)
Inspire Challenge Award: Farm.ink: Analyzing livestock social media data for farmer chatbot	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase). *Farm.Ink*: Analyzing Livestock Social Media Data for Farmer Chatbot, grew its user base from around 30,000 at launch to over 120,000 farmers, enabling animal health researchers to reach 24,000 dairy farmers. A recent survey of over	National, Kenya
Global Agricultural Data Innovation and Acceleration Network (GARDIAN)	Research and Communication Methodologies and Tools	Stage 3: available/ ready for uptake (AV);	Global
Agronomy Field Information Management System (AgroFIMS)	Research and Communication Methodologies and Tools	Stage 1: discovery/proof of concept (PC - end of research phase)	Global
Inspire Challenge Award: Seeing is believing – Using smartphone camera data	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase). *Seeing is Believing* integrates machine-learning analysis of cellphone camera images of crops into the functioning of an insurance and farm advisory service. Early results from a randomized controlled trial demonstrated advisory messages, which reached 32,237 wheat producers, increased knowledge on best agricultural practices by 78%. In addition, respondents were willing to pay an extra 8.7% of their insurance premium when picture-based assessment was embedded in the insurance application.	National, India

Inspire Challenge Award: Real-time diagnostics for devastating wheat rust	Production systems and Management practices	Stage 2: successful piloting (PIL - end of piloting phase). Project *MARPLE* (Mobile And Real-time PLant disEase): Real-Time Diagnostics for Devastating Wheat Rust demonstrated that a small hand-held gene sequencing device used in response to the Ebola crisis could be used in the field to identify a particularly devastating strain of wheat rust within 48 hours—dramatically enhancing crop disease surveillance and reducing the cycle time of response. The pilot demonstrated the first ever accurate identification of individual yellow rust strains just as the disease was starting to emerge in the field. The data generated by the pilot has already been incorporated into Ethiopia's national early wheat rust warning systems.	National, Ethiopia
Inspire Challenge Award: Using IVR to connect farmers to market	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase). Viamo and CIMMYT designed and implemented a *crowdsourced IVR* (Interactive Voice Response) marketing service to increase linkages between individuals in the value chain. Farmers were enabled to report their harvest/needed services through their phone and onto a web platform/app that buyers can use to locate produce. This IVR marketing service completed human-centered design studies and prepared for national launch in Nepal.	National, Nepal
Inspire Challenge Award: Pest and disease monitoring by using artificial intelligence	Production systems and Management practices	Stage 1: discovery/proof of concept (PC - end of research phase). The project aimed to radically transform *pest and disease monitoring by using artificial intelligence (AI)*, advancing sensor technology and crowdsourcing capable of connecting the global agriculture community to help smallholder farmers.	Multi-national, United Republic of Tanzania, Kenya
Metadata Quality Checking Tool for Dataverse Repositories	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase)	Global
CGIAR Core Metadata Schema version 2	Research and Communication Methodologies and Tools	Stage 2: successful piloting (PIL - end of piloting phase)	Global
Agronomy Ontology	Research and Communication Methodologies and Tools	Stage 3: available/ ready for uptake (AV);	Global

TABLE 5: Summary of status of Planned Outcomes and Milestones (Sphere of Influence-Control)

Module		Summary narrative on progress against each FP outcome this year.	Milestone	2018	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.	Module 1 engages with several partners to help CGIAR handle sensitive data, collaborate on data annotation standards, and easily aggregate and analyze large amounts of data. Modules 1 and 2 worked with Accenture to identify and overcome system-wide and center-specific gaps relating to the effective, secure, and ethical use of big data in agriculture. The value of best practices and FAIR for CGIAR and other stakeholders will be demonstrated from use cases based on GARDIAN data search, translate, and analysis, and case studies for data aggregation. Data science workshops in 2018 and 2019 backstop CGIAR researchers and enhance their capacity.	2018 - 1.1.1. Prototype data analytics environment developed and tested.	Extended	This is a relatively new and challenging area of work, requiring collaboration and new approaches. The projects related to this milestone were conceived in 2018 as long-term work with partners to fully leverage their comparative advantages, as well as the comparative advantages of the Platform and CGIAR to achieve Platform goals.
			2018 - 1.1.2. Seamless integration with 2-3 key analytical or mapping tools enabled.	Complete	Strong progress has been made on exciting analytics approaches and capacity with the University of Florida (available here) and UC Davis (available here). This work will continue through 2019.
			2018 - 1.1.3. At least two case studies using the analytical environment developed and presented.	Extended	CGIAR's data assets can provide great value through easier aggregation, querying, and analysis, e.g., across location, time, and datatype. Work towards this was conceived as longer-term, to continue beyond 2018, and includes design/implementation of GARDIAN data annotations for (1) R-based analysis pipelines, (2) decision support, and (3) aggregated high-value data products.
	M1 Outcome: 1.2. CGIAR resources are discoverable and reused.	GARDIAN improvements to better showcase CGIAR research include: improved search, filtering, and linking of related publications and datasets; a pilot to query and explore a semantically-enabled data pool; machine-readable licenses; and visualization of crop production data. The number of CGIAR publications and datasets discoverable via GARDIAN almost doubled in 2018. An overview of CGIAR licensing for GARDIAN data	2018 - 1.2.1. New interface and functionalities added to GARDIAN (Global Agricultural Research Data Innovation Acceleration Network), including tools and services to accommodate data privacy,	Complete	Querying GARDIAN and browsing through data and publications results will demonstrate achievement of this milestone.

M1					
		was undertaken and reported back to centers, and an algorithm to identify Personally-Identifiable Information (PII) was developed and tested with data from Bioversity International and ICRISAT. Centers were supported in improving resource discoverability and data annotations via funds, tools and services, and capacity development.	ethics, and licensing issues. environment developed and tested.		
			2018 - 1.2.2. At least 50 datasets in institutional repository quality-checked, and both data and metadata annotated with ontology and/or AGROVOC/GACS terms.	Complete	Improved annotations for publications and datasets archived in Center repositories, with increased use of semantic standards enhancing discovery and data exploration via GARDIAN.
			2018 - 1.2.3. Sustainable model for data annotation/ curation tested.	Complete	Best practices for data annotation/curation at centers were supported through tool and workflow development, including: a tool to assess compliance with CG Core metadata (available here); COPO (available here) for easy metadata annotation customized for CGIAR; prototype ontology annotation and PII checking tools. Implementation by centers is expected in 2019.
			2018 - 1.2.4. Data mining and machine learning methodologies tested for improving data quality or searchability.	Complete	This milestone demonstrates the value of semantically enriched data assets provided by centers through a GARDIAN pilot showing semantically-enabled querying, exploration, aggregation, and visualization of datasets (available here).
	M1 Outcome: 1.3. Standards and semantics are utilized to enable FAIR	This outcome focuses on promoting and supporting semantic and other standards across CGIAR to enhance CGIAR's FAIR data assets and effectively position CGIAR as a leader in the digital agriculture arena. Notable milestones	2018 - 1.3.1. CG Core Metadata Schema v.2.0 finalized and implemented, and/ or mapped across Center publica	Complete	The CG Core Metadata Schema v.2.0 that enables common repository-level annotation and discovery of CGIAR resources was upgraded and finalized by the cross-Center Metadata Working Group, supported by Module 1. The schema was ratified as final and acceptable in December 2018 (available here) to be implemented by centers in 2019.

M1	(Findable, Accessible, Interoperable and Reusable) agricultural data.	in 2018 include: Release of the CG Core Metadata Schema v. 2.0 draft; completion of the Agronomy Ontology; user-testing of the ontology-powered Agronomy Field Information Management System (AgroFIMS) prototype producing agronomy field books for digital data collection with semantic, standards-compliant metadata and data (standardized column headings within datasets etc.). This work, and development of a draft Socioeconomic Ontology (SociO), relies heavily on partnerships.	tions and data repositories.		
			2018 - 1.3.2. Agronomy Ontology completed and prototype field book tested for data collection based on semantic standards.	Complete	The Agronomy Ontology, developed under Module 1, is the semantic standard for agronomic data, harmonizing data and enabling digital data collection via AgroFIMS. AgrO is now largely stable; at the first user workshop in 2018 users reported ease of use, with no major changes to the ontology requested.
			2018 - 1.3.3. Ontology (Crop and Agronomy Ontology at minimum) and/or AGROVOC/ GACS terms adopted as metadata descriptors by Center repositories/platforms.	Complete	Adoption of semantic standards to enhance FAIRness of CGIAR's data assets was promoted through text and data mining webinars (available here and here), during sessions of the PhenoHarmonis workshop, support for CGIAR's Ontology, Metadata, and Dataverse working groups, and through sessions at the October data and information specialists meeting in Kenya (available here).
			2018 - 1.3.4. Draft for key classes and sub-classes for socioeconomic ontology developed.	Complete	Draft classes and sub-classes for SociO were developed in 2018 to capture concepts represented by 100 questions identified as common across CGIAR surveys (https://www.dropbox.com/s/rga7sghtcv67am2/100Q_v1.pdf?dl=0). This effort has been led by the Socioeconomic Development Community of Practice (SeD CoP), spearheaded by CIMMYT, IFPRI, ILRI, and others, under Modules 1 and 2.
			2018 - 1.3.5. Links to agrisemantics efforts external to CGIAR maintained or enhanced.	Complete	Module 1 supported the development and adoption of agrisemantics standards in 2018, through funding and intellectual support for the PhenoHarmonis workshop for over 100 participants from within and beyond CGIAR (https://sites.google.com/a/cgxchange.org/cropontologycommunity/2018-phenoharmonis), and wider domain engagement including 35-40 presentations, keynotes and panels discussions (e.g. RDA-IGAD, DCMI: http://www.dublincore.org/conferences/2018/keynotes/, SWAT4LS: http://www.swat4ls.org/workshops/antwerp2018/keynotes/), and two agrisemantics concept notes.

M1	M1 Outcome: 1.4. Enhance capacity, catalyze culture change to further CGIAR OA/OD compliance and public goods mandate.	2018 activities towards this outcome were focused on developing training materials and guides on achieving and managing FAIR data, including licensing, privacy/ ethics, and best practices in data management. Module 1 and 2 worked together to make available a course and online seminar on the General Data Protection Regulation (https://bigdata-cgiar.course.tc/catalog/course/gdpr-for-international-development). Support was provided to centers for data sprints to promote the upload of well-annotated datasets to repositories, and for 3 workshops on ways to improve the FAIRness of CGIAR data assets.	2018 - 1.4.1. Materials and webinars developed and/ or shared with appropriate CGIAR communities on licensing resources, data privacy and ethics, and data management and standards for maximizing FAIRness (Findability, Accessibility, Interoperability and Reuse) of CGIAR resources.	Complete	Guidance and a course on addressing privacy and ethics in CGIAR datasets were developed in consultation with Centers. Webinars and a presentations/discussions at the CGIAR data management and information specialist meeting in Kenya (referenced above) also helped build capacity on best practices in FAIR data management.
			2018 - 1.4.2. At least two data sprints or one relevant training and one data sprint held for researchers to increase the number of well-annotated datasets in center data repositories.	Complete	Module 1 funds were used by centers to organize 34 trainings and 18 data sprints to help researchers better annotate datasets using semantic standards and CG Core metadata, and to sensitize them to FAIR resources. About 40% of approximately 1,100 datasets uploaded were a direct result of these sprints.
			2018 - 1.4.3. At least two workshops/trainings for data/ information/ ontology managers and researchers held on ways to render datasets FAIR.	Complete	Several in-person and virtual presentations were organized through Platform-supported CoPs and working groups to enhance CGIAR data management and FAIRness of research outputs, including on: ontologies and vocabularies for data annotation; several tools to ease metadata entry and compliance; licensing; and best practices in addressing privacy and ethics (https://sites.google.com/a/cgxchange.org/oad-support-pack/2018-events/2018-oawg-dmtf-annual-meeting; and https://cgiar.sharepoint.com/sites/CGIARData&InformationManagementCommmunityofPractice).

M2	M2 Outcome: 2.1. CGIAR is more broadly engaged in BIG DATA community.	The Platform developed several shared services (clean historical weather data, secure transfer of large datasets, gridded global population data, and commercial satellite imagery access), started development of critical datasets (croplands, elevation) and launched a learning portal to host webinars and short online courses specifically for CGIAR. In addition, the Platform produced another edition of the Annual Big Data in Agriculture Convention.	2018 - 2.1.1. Communities of Practice around topics of geospatial data, socioeconomic data, ontologies, data-driven agronomy, livestock data for development and crop modelling, produce outputs addressing key constraints of the sector and establish CoP networks.	Complete	Six CoPs were fully operational during 2018. Each CoP has held webinars, online virtual discussions, and contributed to review and synthesis papers provided in the publications annex. They also have websites with dynamic content available here.
			2018 - 2.1.2. Hold high-level Annual Big Data in Agriculture Convention, with wide participation of CGIAR and non-CGIAR actors, establishment of collaborative agreements.	Complete	The Convention was held in October 2018 in Nairobi, co-hosted by ILRI and ICRAF. Over 400 people were in attendance and 2,500 people participated virtually (of which 35% were female). Full documentation is available here.
	M2 Outcome: 2.2. CGIAR increases its capacity to work on priority topics more quickly, more effectively and at greater scale.	The Platform developed several shared services (clean historical weather data, secure transfer of large datasets, gridded global population data, and commercial satellite imagery access), started development of critical datasets (croplands, elevation) and launched a learning portal to host webinars and short online courses specifically for CGIAR.	2018 - 2.2.1. Identify high priority, high impact new data products and develop methodological plan to produce them, with initial implementation.	Complete	All centers were given access to key enabling datasets and analytics environments as Shared Services, including Digital Globe's GBDX Platform and Satellite Imagery Archive, TWC Global Historical Weather Dataset, aWhere Weather Dataset, and LandScan Gridded Population Dataset through the Center Focal Points and CoP on Geospatial Data. Agreements with the providers are archived at the Platform SharePoint and available upon request.

M2	M2 Outcome: 2.3. CGIAR develops as a learning organization.	The Platform conducted a pan-CGIAR assessment of the state of digital strategy in the organization addressing the role of leadership, data, partnerships, infrastructure, and skills building in building a more effective, cohesive digital vision for the System. This culminated in special sessions at the Convention, where four Directors General of centers underscored the importance of building new digital strategies for the organization. This spurred requests directly from centers to look more closely at these elements of digital strategies at the center level.	2018 - 2.3.1. Map out CGIAR needs for common big data related computing and storage infrastructure.	Extended	Platform personnel and consultants conducted a pan-CGIAR assessment of these needs and gaps in developing digital strategy for the organization, and began information infrastructure mapping at centers. This is a complex issue which continues to be evaluated in 2019 with an expanding partnership with GLOBUS.
			2018 - 2.3.2 Establish shared services for CGIAR by negotiating with external data utility partners.	Complete	Shared service agreements in place with ESRI, aWhere, IBM The Weather Company, GLOBUS, LandScan, and Digital Globe. Agreements with the providers are archived at the Platform SharePoint and available upon request.
			2018 - 2.3.3 Develop capacity building activities linked to Centers' needs.	Complete	As per capacity building section, numerous training sessions held on distinct digital topics. Some of the webinars and online courses are documented here.
			2018 - 2.3.4 Build the capacity of CGIAR to meet the data needs of the agriculture development sector.	Complete	Significant capacity building efforts have been made through each of the CoPs to CGIAR institutions, and through shared service agreements centers. CRPs have access to more tools and data resources. The Platform used the food system framework as a way to develop the agenda of the Annual Big Data in Agriculture Convention, interviewing some 60 different food system participants in East Africa.
M3	M3 Outcome: 3.1 CGIAR shows how data-driven approaches yield results in poverty reduction, enhanced nutrition or environmental benefits.	Under this module the Platform has sourced 24 high-quality innovations to date (two cohorts of 12 finalists) over two years and awarded 10 startup grants and three scale-up grants. For those that got to scale-up stage, they had quantified evidence and credible plans for scale-up that were evaluated by scaling experts.	2018 - 3.1.1. New Pilot Inspire projects around big data related innovations.	Complete	Five new pilot grants were awarded in 2018, and are documented here.

			2018 - 3.1.2. Scale-up one successful pilot Inspire project (from winners 2017) around big data related innovations.	Complete	Three scale-out grants were awarded in October during the Annual Big Data in Agriculture Convention.
			2018 - 3.1.3. Synthesis of Inspire project successes and failures, policy documents, best-practice guidance.	Complete	A learning document was produced and published reflecting on the innovation process around the 2017 Inspire grants.

TABLE 6: Numbers of peer-reviewed publications from current reporting period

	Number	Percent
Peer-Reviewed publications	7	100.0%
Open Access	5	71.43%
ISI	5	71.43%

TABLE 7: Participants in CapDev Activities

Number of trainees	Female	Male
In short-term programs facilitated by PTF	169	356
In long-term programs facilitated by PTF	164	211

TABLE 8: Key external partnerships

Lead Module	Brief description of partnership aims	List of key partners in partnership. Do not use acronyms.	Main area of partnership (may choose multiple)
M1	Agroknow supported the evolution of GARDIAN by contributing with broadening the discoverability and scope of the system, enabling data integration and interoperability, and enabling a service ecosystem.	Agroknow Inc	• Delivery • Other
M1	The team is supporting the COPO tool implementation to streamline metadata entry to GARDIAN and center repositories. The tool facilitates the standardization of dataset column headings with ontology/GACS terms.	Collaborative Open Plant Omics (COPO) - Earlham Institute	• Delivery • Other
M1/M2	Mini-grant recipient that is co-investing in developing the analytic pipelines for linking GARDIAN to modeling tools and the AgMIP data translators.	University of Florida	• Delivery • Other
M1/M2	The Minnesota Supercomputing Institute is co-developing several efforts with the Platform, including: development of an ontology independent, flexible, and extensible metadata schema; testing an MSI supported software platform for secure management of sensitive data; and developing a mash-up of the existing ontologies.	University of Minnesota	• Delivery • Other
M1/M2	The team is providing guidance on managing personally identifiable information in geo-spatial data and developing a common analytics environment linked to GARDIAN technical training on the use of R/Python programming for spatial analyses.	UC Davis - University of California, Davis	• Delivery • Other
M2	The partnership will bring research collaboration around validation of data services: "Clean Historical Weather Data" and "Seasonal Probabilistic Forecast".	IBM - International Business Machines Corporation	• Delivery • Other

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
BIG DATA provided a cost-share on a proposal selected for funding by the Bill & Melinda Gates Foundation. In 2018, Inspire Challenge funding went to an AI-driven app for disease diagnosis. The Platform funded development of an image processing pipeline for drone imagery for cassava, bridging breeding and agronomy use-cases.	CIP, RTB	Fundraising benefits, scientific benefits
BIG DATA funded development of a drone image processing pipeline bridging breeding and agronomy use-cases for Rice.	RICE and RTB	Efficiency and scientific benefits
The Platform co-designed 'big data-enabled' gender research with the Gender Platform.	PIM	Scientific, in the sense of development of new methods
CGIAR partner for the Inspire Challenge winner: Seeing is believing: Using smartphone camera data.	IFPRI / PIM	Scientific, in the sense of development of new methods
CGIAR partner for the Inspire Challenge winner: Farm.ink: Analysing livestock social media data for farmer chatbot.	ILRI / Livestock	Scientific, in the sense of development of new methods
CGIAR partner for the Inspire Challenge winner: Using IVR to connect farmers to market.	CIMMYT / Wheat	Scientific, in the sense of development of new methods
CGIAR partner for the Inspire Challenge winner: Pest and disease monitoring by using artificial intelligence.	IITA, CIP, Bioversity International	Scientific, in the sense of development of new methods
CGIAR partner for the Inspire Challenge winner: Real-time diagnostics for devastating wheat rust.	CIMMYT / Wheat	Scientific, in the sense of development of new methods
PhenoHarmonIs 2018 workshop on semantic harmonization of breeding and agronomic data.	RTB, Excellence in Breeding	Fundraising benefits, scientific benefits
Support in exploring management of benefit sharing mechanisms under the ITPGRFA in a digital era.	Genebanks Platform	Strategy and policy relevant research
Shared service agreements with Digital Globe, ESRI, AWhere, IBM Weather Company, Microsoft Research.	All centers and CRPs	Access to new data tools, services, and software

TABLE 10: Monitoring, Evaluation, Learning and Impact Assessment (MELIA)

The Platform commissioned a study of the submissions to the 2017 Inspire Challenge in light of innovation strategy and with an eye towards learning more about partnerships, data types used or developed, target users, and features of enabling environments for digital agriculture innovation to take root. The resulting synthesis paper informed the 2018 process and dramatically improved targeting of these innovations towards small producers.

Studies/learning exercises planned for this year (from POWB)	Status	Type of study or activity	Please include links to MELIA publications here.
S2790 - The 2017 Inspire Challenge: Innovation Strategies for Digital Agriculture	Complete	Synthesis: reviews, systematic reviews, evidence gap maps	https://bigdata.cgiar.org/wp-content/uploads/2018/08/The-2017-CGIAR-Inspire-Challenge-3.pdf

TABLE 11: Update on Actions Taken in Response to Relevant Evaluations
N/A.

TABLE 12: Examples of W1/2 Use in this reporting period (2018)
N/A.

TABLE 13: Platform Financial Report

	Planned Budget 2018*			Actual expenditure*			Difference*			Comments
	W1/W2	W3/ Bilateral	Total	W1/W2	W3/ Bilateral	Total	W1/W2	W3/ Bilateral	Total	
M1 - Organize	US$ 5,604,297.00	US$ .00	US$ 5,604,297.00	US$ 4,804,865.00	US$ .00	US$ 4,804,865.00	US$ 799,432.00	US$ .00	US$ 799,432.00	Module 1 includes grants to all centers to reinforce their ability to curate, publish, and share data as well as several cross-cutting investments to build overall capacity of CGIAR to capture, curate, and more fully leverage its collective data assets. Carryover noted under this Module captures both cross-cutting and center-level investments.
M2 - Convene	US$ 1,854,001.00	US$ 426,489.00	US$ 2,280,490.00	US$ 1,683,191.00	US$ 313,755.00	US$ 1,996,946.00	US$ 170,810.00	US$ 112,734.00	US$ 283,544.00	The Platform saw an uptick in aligned bilateral investments from Centers in 2018, while Platform managed funds carried over 9%.
M3 - Inspire	US$ 1,360,419.00	US$ .00	US$ 1,360,419.00	US$ 387,410.00	US$ .00	US$ 387,410.00	US$ 973,009.00	US$ .00	US$ 973,009.00	The Inspire Challenge is the Platform's signature innovation process, that culminates in Q4 of the fiscal year, where previous year's awardees report on progress and compete for scaling funds, and a new cohort of finalists goes for startup funding. As a result, major program commitments happened in Q4 for projects to be implemented in 2019. This contributed significantly to overall carry over for the Platform.
Strategic Competitive Research grant	US$ .00	US$ .00	US$ .00	US$ .00	US$ .00	US$ .00	US$ .00	US$ .00	US$ .00	There were no strategic competitive research grants in 2018.
Platform Management and Support Cost	US$ 455,494.00	US$ .00	US$ 455,494.00	US$ 456,474.00	US$ .00	US$ 456,474.00	US$ -980.00	US$ .00	US$ -980.00	The budgets for Platform Secretariat and the Project Management Unit support costs were completely drawn down in 2018.
Platform Total	US$ 9,274,211.00	US$ 426,489.00	US$ 9,700,700.00	US$ 7,331,940.00	US$ 313,755.00	US$ 7,645,695.00	US$ 1,942,271.00	US$ 112,734.00	US$ 2,055,005.00	

Platform for
Big Data
in Agriculture

CGIAR