CGIAR

Platform for Big Data in Agriculture

# ANNUAL REPORT 2017

Led by

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

IFPRI
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

# CGIAR Platform for Big Data in Agriculture Annual Report 2017

Led by:

# Table of Contents

# 1. Key Results

## 1.1 Highlight Platform Outputs

The Platform for Big Data in Agriculture is CGIAR's signature initiative, designed to leverage e-research and build new data-driven impact on the path to reducing poverty, improving food and nutrition security, and improving natural resources and ecosystems worldwide. To contribute to these outcomes the Platform aims to affect digital transformations inside and outside of CGIAR, including: mobilizing CGIAR data to accelerate research and spur new data-driven innovations, building data collaboration across the organization and with the wider agricultural sector, and leveraging CGIAR expertise while claiming a unique leadership voice in digital agriculture.

The Platform launched in May 2017, the culmination of nearly two years of consultation with more than 40 private, non-profit, and public stakeholders in digital agriculture worldwide. The venue for the launch was the 9th edition of the Information and Communications Technologies for Development (ICT4D) Conference, a global gathering of reference for the digital development community. The 800 attendees were primarily from non-profits from developing economies who specialize in using digital technologies to engage underserved communities.

A main focus of the Platform in its first year was establishing all critical components for effective governance and operations: constituting the management team and steering committee, hiring a Platform leader and a Communications Coordinator, establishing and supporting six technical Communities of Practice (CoPs), and establishing legal agreements and financial management processes with all CGIAR Centers, and promoting improved data management and use across the CGIAR System. The Platform finished the year with a solid foundation for execution.

### Module 1

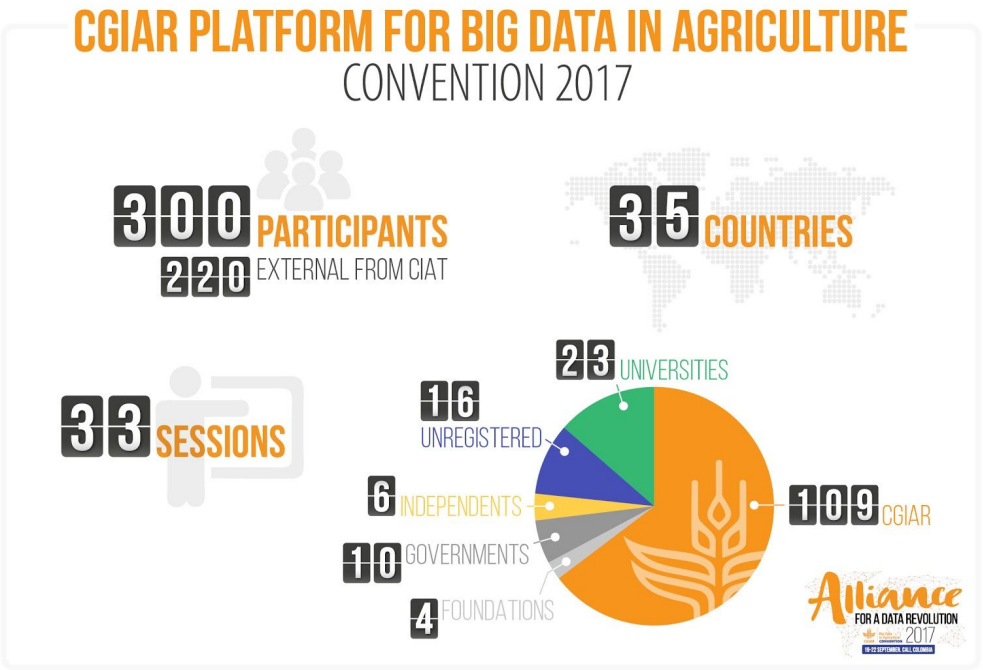
The Platform generated important outputs in its first year. Module 1 ("Organize") saw increased support and momentum for Centers to comply with CGIAR's Open Access and Data Management (OADM) Policy. Technical guidance and seed funding issued to all 15 Centers in support of implementing this policy contributed to a significant increase -- of 10 percent or more at most Centers -- in the number of public datasets and publications made available via their data repositories. To monitor and accelerate Centers' progress towards making their data Findable, Accessible, Interoperable and Reusable (FAIR), the Platform developed and launched a prototype of the first pan-CGIAR data search tool, enabling any user to execute keyword searches and discover available CGIAR publications and datasets across more than 30 open databases system-wide. The tool was provisionally named CGIAR e-Research (CeRes).

## Module 2

Under Module 2 ("Convene") the Platform established Communities of Practice (CoPs) in six key areas for data-enabling food security research: socioeconomic data, geospatial data, data-driven agronomy, crop modelling, livestock data for decision-making, and ontologies. For each, key actors across the CGIAR as well as external partners collectively identified priorities for data-enabling food security research and developed multi-stakeholder work plans.

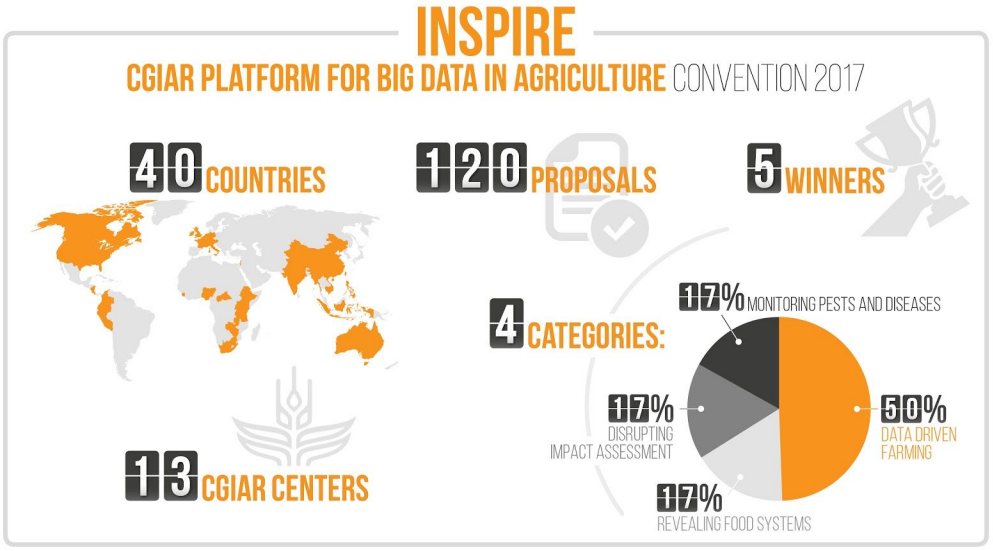

In September 2017 the Platform convened some 300 global innovators, researchers, and thought leaders from public, private, and non-profit partners in Palmira, Colombia, for the first annual CGIAR Convention on Big Data in Agriculture -- an event with content and attendees carefully curated to foster new alliances and advance the discipline of digital agriculture for both CGIAR and the sector.

The Convention opened the way to build or deepen novel partnerships that leverage CGIAR expertise to shape the future of digital agriculture in development economies. Attendees identified critical policy, infrastructure, and investment priorities for the agriculture development sector and contributed to a genuinely multi-stakeholder plan of action for the Platform for the coming year. Following the Convention, the Platform began discussions with providers of cloud computing and storage solutions, high resolution satellite imagery, data analytical infrastructures, machine learning services, and partners providing new pathways to data analysis for effecting positive change in food and farming systems worldwide.



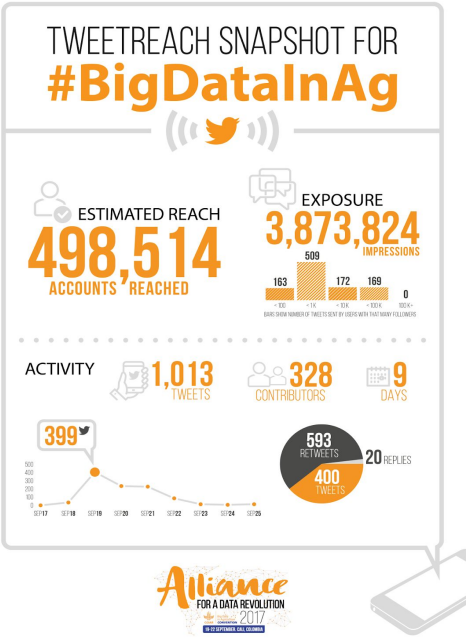
## Module 3

Module 3 ("Inspire") was designed to help CGIAR and its partners apply data to help solve development problems faster, cheaper, and more efficiently. In 2017, the Platform created an open innovation process called the Inspire Challenge, which attracted submissions from 120 teams from 37 countries. Challenge categories were defined in consultation with the CGIAR Research Programs, the comprehensive CGIAR research portfolio for building resilient food systems worldwide. External expert judges chose five teams to receive startup grants of $100,000 each in support of leading-edge applications of data including in-field disease gene sequencing and identification, social media-enabled animal health, and intermediating a whole market system over mobile phones.




## Communications

The Platform built its communications presence in 2017, with the objective of helping CGIAR claim a leadership voice in digital agriculture. All major social media platforms, the Platform website, and webpages for Communities of Practice became regular publication channels. Engagement with the Platform during, and as a result of, the Big Data Convention was significant, with Twitter reports showing approximately 500,000 unique accounts were reached during the event.




High profile external media outlets also engaged with CGIAR on event and Inspire Challenge projects. Media coverage momentum built throughout year with the Platform appearing in or on, among other outlets, globally reputable media organizations, such as Devex, the Australian Broadcasting Corporation (ABC) and Reuters.

### Contribution to Development Outcomes

The Platform serves a supporting role in helping CGIAR develop as a learning organization, cultivate new digital alliances, and pursue data innovation in support of its mission. As a result, in 2017 the Platform cannot claim direct links to the external Development Outcomes and Intermediate Development Outcomes enshrined in the CGIAR strategy.

## 1.2 Platform-specific Quality Control Activities

### Data Ontologies

The Platform, working through the Ontologies Community of Practice (Ontologies CoP), achieved several key results related to quality control in 2017. The Ontologies CoP developed new trait ontologies in Crop Ontology and updated the Agronomy Ontology ("AgrO"), integrating concepts from other ontologies including: the Environmental Ontology ("ENVO"); the Chemical Ontology ("CHEBI"); the Unit Ontology ("UO"); and the Plant Trait Ontology ("PATO") -- among others. These updated ontologies will improve data integration with digital field books and enable seamless communication across databases -- a critical step for standardizing and improving the quality, interoperability, and reusability of agronomic research data.

### Prototypes for Data Quality and Synthesis

#### Data Search and Discovery

The Platform completed a prototype of a pan-CGIAR global data search and discovery tool, provisionally branded as CGIAR e-Research ("CeRes") and created dedicated time to capture stakeholder input and user experiences from some 300 agriculture development experts at the Big Data Convention. CeRes was created specifically to enable access and download of datasets across all CGIAR open data repositories, and it includes features to ensure that data that are described and harmonized using standard, uniform metadata and ontological keywords.

#### Metadata assessment

To support the use of the CG Core metadata standard and ease data curation, the Bioversity Center developed a prototype tool to enable visualization of the metadata status of datasets uploaded to CGIAR open data repositories. The tool helps assess the completeness of metadata per data set, and provides a link to directly access datasets. The prototype of the tool was deployed and tested in the Dataverse repositories of Bioversity and CIAT prior to being demonstrated to the CGIAR Data Management Task Force. The code is available on the Bioversity Github and the Bioversity version is available here. Four CGIAR Centers developed plans to deploy the tool in 2018.

# 1.3 Progress by Platform Modules

## Module 1 "Organize"

In 2017, the Big Data Platform Module 1 provided funding and guidance to all 15 Centers to address critical gaps and build capacity in, and further operationalize the CGIAR Open Access and Data Management (OADM) Policy. The data managers and information specialists at all Centers convened under the aegis of Module 1 to share progress and learning from their efforts to implement Findable, Accessible, Interoperable, and Reusable (FAIR) publications and data, and to explore collaborations on shared issues and concerns. Throughout the year, Centers were able to update workflows and software for data sharing, and to contribute more staff time toward open and FAIR data and publications. As a result, there was a significant increase in the number of discoverable publications and datasets in Center repositories in 2017, with an increase of more than 10% compared to the previous years for most Centers. Center efforts increasingly focused not only on the number of open datasets, but also the quality of data and data annotations, with the concept of FAIR data resources gaining wider understanding and acceptance by staff.

The capability to measure progress and monitor adherence to the FAIR standard was made possible due to the launch of a prototype of a system enabling discovery of research outputs across CGIAR Centers, showing the range, depth, and importance of CGIAR research and enabling others to build on it to innovate further -- an exciting first. The tool was provisionally called CGIAR e-Research (CeRes). By the end of 2017, CeRes enabled discoverability of about 50,000 publications and 1,800 datasets; with several new features planned for the coming year.

## Module 2 "Convene"

The Big Data Platform formed six Communities of Practice (CoPs) open to all Centers and partners to help better define and shape global good practice, address bottleneck issues, and engage on standards for technical disciplines related to digital agriculture. In 2017, the Platform and CoPs actively recruited members and the Big Data in Agriculture Convention served as an opportunity to define priorities and agendas for action for the coming year.

**The Crop Modelling CoP** funded three mini proposals designed to support activities related to modelling that would achieve a demonstrable boost or impact through a small amount of catalytic funding. In addition, the Crop Modelling CoP initiated the development of four draft review articles that seek to document and level-set crop modeling activities and potential impacts from CGIAR Centers and partners, with a specific focus on identify the main Crop Modelling gaps within the CGIAR and how best to leverage integrated genotypic, environmental, crop management, and socioeconomic data. During 2017, the CoP grew to 204 members, including CGIAR scientists (40%), research institutes and academia (30%), private sector organizations (20%), and NGOs, government and multilateral organizations (10%).

**The Socioeconomic CoP** formed working groups on gender, ontology-independent structural metadata schema, ethics, privacy and cyber-security, and blockchain. It also identified the need for a socioeconomic data ontology and roadmap for its development. In 2017, the CoP had about 170 members, of which around 25% were CGIAR scientists, 25% were from academia, 20% were from the private sector, and the rest represented NGOs, government, multilateral organizations.

**The Data-Driven Agronomy CoP** identified key actors and initiatives working on data-driven agronomy and gauged the strengths and critical gaps across CGIAR and its strategic partners; and, in the process, further defined the vision of the CoP. The CoP developed an exhaustive list of actors working in weather and climate, soils, crop management, and digital support for extension services. The Data-Driven Agronomy CoP grew to 280 members in 2017. It is a diverse group with participants from the CGIAR Centers, research institutes, academia, NGOs, and the public and private sector.

**The Geospatial CoP** developed the geospatial data cataloging and analysis features of CeRes that were soft-launched at the Convention, available at http://geo.cgiar.io. This tool was developed to catalogue, visualize, and perform spatial analysis of large geospatial datasets published by CGIAR. The Geospatial CoP developed a community website at http://cgiarcsi.community and used it as a communication platform to showcase CGIAR's geospatial research. The Geospatial CoP includes about 100 members across all 15 Centers and academic partner institutions, including the University of California, Davis, and the University of Twente in the Netherlands.

**The Ontologies CoP** developed new trait ontologies in Crop Ontology and updated the Agronomy Ontology (AgrO) to better integrate with existing digital fieldbooks. This constituted a critical step toward improving the quality and organization of agronomic research data from the point of collection. A collaboration with the Socioeconomic Data CoP was launched during the Convention to start a working group that will address the ontology gaps for socioeconomic data. Interoperability of biophysical and socioeconomic data is a key inhibitor to the effective study of critical issues for the CGIAR mission as: the potential adoption of breeding products, and gender- and youth-sensitive data, and meta-analyses bridging discrete research domains.

One priority activity in 2017 for the Ontology CoP was improving and expanding the Crop Ontology species. This increased the number of species to 27, with the latest addition being an algae (sugar Kelp) by Cornell University. More physiological and quality traits were added that will be used by data platforms including the Roots, Tubers, and Bananas Research Program's NExtGeneration breeding databases. The Ontologies CoP grew to 140 members in 2017.

**Livestock Data for Decisions (LD4D).** In late 2017, the Platform Steering Committee approved the addition of an external Community of Practice, Livestock Data for Decisions (LD4D), which is led by the University of Edinburgh. Key collaboration areas include livestock ontologies and promoting the role of animal science in building resilient food systems.

### The CGIAR Big Data in Agriculture Convention

The Platform produced the first annual CGIAR Convention for Big Data in Agriculture in September 2017, convening some 300 global innovators, researchers, and thought leaders from public, private and non-profit partners in Palmira, Colombia. The content and attendee list was carefully curated to foster new alliances and advance the discipline of digital agriculture for CGIAR and the sector. The Convention was an important vehicle for recruiting new CoP members, and attendees identified critical policy, infrastructure, and investment priorities for the agriculture development sector and contributed to a genuinely multi-stakeholder plan of action for the Platform.

### Shared services

In 2017, the Platform conducted a survey among the Centers, as a first attempt at uncovering common information technology needs for scientific research across the system that the Platform can help address through shared data services. Results showed varied needs, capabilities, and budgets across Centers related to storage, computation, datasets, and building capacity for Big Data. Genomics researchers and some climate researchers, for example, reported that they need high performance storage and computing power, and found that they prefer the cost-performance of doing this on-site rather than in the cloud. Some Centers reported that some amount of cloud-based data archiving and store-and-compute services would be valuable to them. As a result, the Platform began to further segment the various Centers' needs with an eye towards developing one or more shared data services in 2018.

## Module 3 "Inspire"

In 2017, the Platform designed and implemented a CGIAR-branded open innovation process called the Inspire Challenge. More than 300 innovators from 37 countries participated in the Challenge, generating 120 submissions addressing four topic categories developed in collaboration with the 12 CGIAR Research Programs:

- Pest and Disease,
- Revealing Food Systems,
- Disrupting Impact Assessment, and
- Data-Driven Farming.

Applicants were required to have CGIAR and non-CGIAR team members as well as to develop concepts that fully leveraged the competencies of each participating organization to develop something new. Twelve finalists defended their proposals at the Big Data in Agriculture Convention in September. Five of these were selected by external judges to receive $100,000 startup grants. Among the 5 awardees, 3 were from Pest and Disease, one was from Revealing Food Systems, and 1 was from Data-Driven Farming.

# 1.4 Cross-Cutting Dimensions at Platform Level

## 1.4.1 Gender, Youth and Capacity Development

### Capacity Development

Big Data Platform Focal Points, several CGIAR Data Managers, and an array of CGIAR researchers joined the Big Data in Agriculture Convention. Attendees engaged in Community of Practice sessions to help better define and shape global good practice,s address bottleneck issues, and engage on developing standards for technical disciplines related to digital agriculture. Attendees gained new knowledge and insights on the challenges and opportunities of making CGIAR data FAIR; this contributed to a significant increase in data sharing in 2017.

### Youth and Gender

In its six months of full operation in 2017, the Platform was unable to explicitly address the youth and gender cross-cutting dimensions. New knowledge and insights gleaned from 2017 activities enabled these specific areas of focus to be further developed and integrated into the program in 2018.

## 1.4.2 Open Data

While all CGIAR Centers are committed to making final information products and research outputs Open and FAIR, to maximize users' ability to reuse and integrate datasets, Centers are at different points with respect to planning, launching, and scaling-up operations. This is particularly true in terms of staffing and infrastructure. To accelerate Centers' efforts to achieve Open and FAIR data, metadata, and repositories, a first tranche of seed funding was allocated to CGIAR Centers in 2017, along with guidelines for acceptable uses of the funds, including:

- Planning, advocacy, capacity building;
- Focal point time and travel;
- Staffing;
- Infrastructure development;
- Enhancing interoperability and the FAIRness of repository infrastructure;
- Enhancing the FAIRness of data; and
- Repository-related subscriptions.

These guidelines were appended to the Program Participant Agreement executed with all Centers.

These Platform funds for Open Data are intended to be additive, to offset initial costs including planning, initial staffing, platform development, or other work towards FAIR data, metadata, and repositories. Additional, smaller tranches are planned for 2018 and 2019, and no further grants are planned beyond 2019. Centers are requested to supplement this investment by co-funding, such as through CRP Open Access/Open Data funds or through

line items in bilateral project budgets. Several Centers were able to update workflows and software for data sharing, and to contribute more staff time to executing the OADM Policy as a result of this support. As a result, there was a significant increase in the number of discoverable publications and datasets in Center repositories in 2017, and most Centers achieved an access increase of more than 10% from the 2016 baseline for access and datasets.

The Platform debuted a new Open Data capability for the system in 2017: for the first time CGIAR can monitor the number of open publications and data as well as insure adherence to the FAIR standard at the Center or System-wide level. This is possible thanks to the launch of a prototype of CeRes, a pan-CGIAR data discovery tool that enables users to create queries to search publications and data repositories at all 15 Centers. By the end of 2017, CeRes was able to access about 50,000 publications and 1,800 datasets. The number continues to grow.

### 1.4.3 Intellectual Assets

In 2017, the Platform developed its high-level intellectual assets and data management plan, adapting several aspects of the plan put in place by the CCAFs CRP. The goals are similar: to enable multiple programs to fulfil their obligations with respect to making data products and supporting documentation from its research globally available. Similar to CCAFS, the Data Management Strategy is based on three pillars: establishing a process, supporting compliance, and enabling a data culture in alignment with the CGIAR OADM Policy. In addition, the Big Data Platform has invested in an online tool that enables users to easily search and discover open datasets and publications across databases at all CGIAR Centers, with the intention of making this a key mechanism for monitoring and measuring compliance with the CGIAR open access policy. Some key guiding principles of the plan include:

- In accordance with the CGIAR OADM Policy, the Big Data Platform is mandated to produce international public goods and ensure that these are open via FAIR principles – that is, the data are Findable, Accessible, Interoperable and Reusable. This enables the data to be used to enhance innovation, impact, and uptake.
- The Big Data Platform also provides data managers at Centers a Data Management Support Pack. This tool was designed to help the research community produce high quality, reusable, and open data from research activities. It consists of documents, templates, and videos covering a range of aspects related to data management and interoperability, ranging from overarching concepts and strategies through to day-to-day activities.
- In 2017, the Big Data Platform Communications Coordinator began developing online content to promote data sharing both across the CGIAR and with the agriculture development sector.
- The Big Data Platform coordinates and supports a monthly webinar series and a number of cross-Center groups and CoPs. These activities are designed to support the management and "FAIRification" of information resources (e.g. CGIAR Open Access Working Group, CGIAR Data Management Task Force, Dataverse Working Group, Metadata Working Group, Ontology Working Group) and has one related Community of Practice on Ontologies that helps to classify agronomic and breeding concepts and knowledge.

- The Platform is examining the Intellectual Property implications of derivative products that CGIAR may generate through use of third-party data services, in light of the potential scale and impact of these products, and will update the IA plan if needed.

# 2. Platform Effectiveness & Efficiency

## 2.1 Variance from Planned Platform Activities

The Big Data Platform proposal outlines several potentially transformational actions for the CGIAR system, changing the way Centers and individual researchers interact with digital agriculture and the wider digital economy. During consultations with convention attendees, the Big Data Platform steering committee, and candidates for the International Advisory Board, it became apparent that there are two key areas that need more definition in order to target CGIAR data and digital investments for greatest effect:

- Market research into the data uses and needs of the array of public, private, and non-profit stakeholders working towards CGIAR System Level Outcomes, in order to identify and prioritize particular data users and ways that CGIAR could help to address these users' data needs.
- The key components or contours of a CGIAR digital strategy that builds on CGIAR's comparative advantages and helps guide decisions about the System's digital technology investments.

The Management Team decided to begin to address these larger strategic questions under Modules 1 and 2.

## 2.2 Use of W1-2 Funding

In 2017, nearly 100% of Platform funding was W1+2 (one Center project was partially mapped to Big Data Platform). In 2018 and 2019, we expect Centers (especially the lead Centers – IFPRI and CIAT) to align Bilateral and W3 funds to the Platform.

## 2.3 Key External Partnerships

In 2017, the Platform approached potential partners that could bring important new complementary capabilities to CGIAR as the organization seeks to shape the future of digital agriculture in developing economies. Following the Convention, the Platform began discussions with companies offering cloud computing and storage services (negotiation is

ongoing with Microsoft Azure and Amazon AWS), providers of high-resolution satellite information (DigitalGlobe), data analytical infrastructures (e.g. University of Minnesota Supercomputing Institute, IBM), machine learning service providers (Google TensorFlow, IBM Watson), and partners providing new pathways to data analysis or impact at scale (e.g. the Digital Impact Alliance to liaise with the mobile industry).

In 2017, the Platform completed negotiation of a research license to access the full imagery archive and cloud-based analysis platform of DigitalGlobe, the provider of high-resolution satellite imagery. In addition, Amazon Web Services agreed to grant credits to DigitalGlobe on CGIAR's behalf to increase the storage, downloads, and computational hours available to CGIAR researchers.

## 2.4 Cross-CGIAR Partnerships: CRPs & other Platforms

A joint infrastructure session was held with the Excellence in Breeding (EiB) and the Genebanks (GB) Platforms at the Convention in 2017 to explore information technology synergies and avoid duplication of efforts. Collaboration, beginning in 2017, centered on ensuring the outputs of the GB and EiB platforms are discoverable by the Big Data harvester prototype CeRes, in order to make them accessible and reusable within the analytics toolkit of the Platform's other services. Specialists from the EiB Platform, in particular, requested that the the crop and agronomy ontologies be maintained and improved, and that the Big Data Platform examines the infrastructure for storing and processing the vast phenotypic and genotypic datasets that are being produced. The Big Data Platform has included this demand in planning and requirements definition efforts for developing a shared storage and processing infrastructure under Module 2.

Under Module 3 ("Inspire"), the Big Data Platform made startup awards to projects aligned with the Wheat; Livestock; PIM; and the Roots, Tubers, and Bananas CRPs.

## OADM

## 2.5 Monitoring, Evaluation, Impact Assessment & Learning (MELIA)

The Platform started to adopt the Managing Agricultural Research for Learning and Outcomes (MARLO) system for program planning and reporting. MARLO has been adopted by a majority of CRPs. In addition, the MARLO and CeRes teams began to design approaches for linking the two systems to facilitate easier reporting of non-sensitive indicator data. Through MARLO, information on Centers' compliance with OADM policy and their uploading of datasets and

publications onto CeRes will become available for mapping and easier visualization of CGIAR's impact at scale. In addition, CeRes enables monitoring of Centers' progress on making their data and publications FAIR.

The Management Team of the Big Data Platform is comprised of coordinators of the Communities of Practice and, as a result, the monthly Management Team meeting serves as a key monitoring and agenda-setting opportunity across all CoPs.

## 2.6 Improving Efficiency

N/A; the Platform was only fully operational for less than six months in 2017.

# 3. Platform Management

## 3.1 Platform Management and Governance

Since the Platform was launched in 2017, a main focus for the second half of 2017 was the establishment of all critical aspects of governance and operations of the Platform: the management team, the steering committee, hiring a Platform leader and a Communications Coordinator, establishing and supporting six technical Communities of Practice (coordinators of which are members of the management team), and putting legal agreements and financial management processes in place with all Centers to promote improved data management and use across the system. The Platform finished the year with a solid foundation for execution.

## 3.2 Management of Risks to Your Platform

### Programmatic

The Platform kept a significant focus on Module 1 activities in 2017. As a result, there were potential risks associated with making data open and accessible. There was the potential risk of failure associated with the CGIAR Centers' delivery on CGIAR Open Access Data Management (OADM) Policy, for example. Late or non-delivery could have potentially eroded the Platform's reputation. Centers will continue to receive funds in 2018 to move toward implementation of the Center Open Access - Open Data (OA-OD) strategy. Distribution of those funds is subject to having an implementation plan in place to mitigate this and other potential risks. Regular

meetings of the Data Management Task Force will also continue to be held. Higher level management will continue to be engaged to ensure the correct incentives are in place in Centers.

Similarly, the Platform seeks to ensure high-level governance buy-in across the CGIAR to adopt best-in-class management of research data, and to promote Center-level investment in the correct incentives and structures to build data-related capacity; this is a key enabler of OADM compliance. To mitigate this risk, the Platform sends email updates to DDGs and DGs, and will continue to develop communications materials that can help highlight best practices across the Centers to ensure this issue has the attention of high-level management.

### Contextual

A constant risk for the Platform is that major privacy breach of farm and farmer data creates a controversial environment for working on Big Data related efforts. To mitigate this, the Platform monitored the external environment with the goal of using proactive communications with maximum transparency to protect the reputation of the Platform. In addition, in 2017 the Platform engaged a lawyer to survey the privacy and ethics frameworks of all Centers as well as external partners, and to develop high-level risk assessments and guidelines for the Platform and the CGIAR System as a whole. This work will be built on in 2018, to produce guidelines and actionable support for Centers and others in the sector.

### Institutional

In the startup phase of the Platform, the work demands on the Secretariat were, at times, extreme. The Platform recruited two additional staff members in 2017 to mitigate the risk of stress-related productivity losses.

## 3.3 Financial Summary

The compressed operational year of the Platform resulted in significant budget carryover from 2017 to 2018. Big Data Platform Focal Points were notified that a very aggressive -- yet responsible -- Plan of Work and Budget would be developed for 2018 in order to maintain momentum toward meeting the program goals while ensuring its continued, overall financial health.

# Tables

## Table A: Reporting against Platform Specific Indicator*

| Indicator | Description | Comments (in relation to target, if one available) |
|---|---|---|
| | | |
| | | |
| | | |

*N/A; the Big Data Platform did not have Platform Specific Indicators in 2017*

# Table B: Status of Planned Milestones

| Module | 2022 Platform outcomes (from proposal) | Milestone* | 2017 milestones status: -Complete -Extended -Cancelled | Provide evidence for completed milestones** or explanation for extended or cancelled |
|---|---|---|---|---|
| **1. Organize** | 1.1. Demand-driven analytics environment | 1.1.1. Key data and analytics needs identified through consultations via existing meetings involving CGIAR and external stakeholders | Complete | The Platform developed interactive sessions on this theme at the Big Data Convention, and conducted a survey of ICT leads and Big Data Focal points at Centers, highlighting a spectrum of needs and capabilities across Centers |
| | | 1.1.2. Stakeholder- responsive roster of 3-5 top data analytics solutions developed, to be hosted on BIG DATA's infrastructure "Toolkit" | Ongoing | Three data analytics and visualization tools developed that span CGIAR Centers, hosted on Big Data's CeRes harvester:<br>(1) CeRes "what if" feature to demo value of FAIR data; (2) CeRes Map feature, to visualize CGIAR mandate crop data; (3) CeRes "relevant resources" feature to link related resources (e.g. pubs relevant to "gender" data query) |
| | | 1.1.3. Key enabling backend services and tools (e.g., Raster Data Catalog and Analysis Tool, R Studio, Sphinx etc.) identified and implemented through the BIG DATA infrastructure | Extended to 2018 | |

| | | | | |
|---|---|---|---|---|
| | 1.2. CGIAR data made accessible | 1.2.1. Dataset prioritization framework developed consultatively and implemented to identify high-value low-risk, low-cost datasets | Partial, extended | US government's CIO data prioritization framework modified to identify high-value, low-risk, low-cost datasets: Prioritization framework, but lack of consensus on implementation across Centers |
| | | 1.2.2. Data quality assurance criteria and exemplar workflows developed | Complete | Data quality workflow exemplar for CGIAR Centers developed as an output of 2017 Data Management Task Force (DMTF) Annual Meeting breakout session |
| | | 1.2.3. 20% increase in uploads of well- annotated current and key legacy datasets to appropriate institutional repositories | Ongoing | 10% overall increase in uploads of datasets by Centers achieved: CeRes analytics |
| | 1.3. Unrestricted CGIAR resource discoverability via early prototype infrastructure | 1.3.1. Scope of work for harvesting CGIAR publications and data developed, including: user requirements, inter-linking and harvesting of institutional data and publications repositories, and planning/reporting platforms (e.g. MARLO), and acceptable budget range | Complete | CeRes user requirements were defined to determine key user requirements re: features and functionalities needed in enabling harvesting and discovery of CGIAR publications and data |
| | | 1.3.2. Criteria and process for choosing appropriate developer/s of harvester articulated and implemented as a consultative process – including key partners, SC, and IAB | Complete | Two proposals for the development of a harvester obtained, and a choice made based on cost and ease of working with developer team (the Steering Committee and International Advisory Board were not in place by then): Cascadeo Inc. bid Agroknow bid<br><br>Extra: Terms of reference for data curation model to standardize CGIAR data in harvester, without which |

| | | | | |
|---|---|---|---|---|
| | | 1.3.3. Service provider for the development of harvester identified; an early working prototype developed | Complete | interoperability and reuse are hampered: Wageningen UR agreement<br><br>CeRes - a harvester that enables discovery of data and publications across all CGIAR Centers developed |
| | 1.4. System-wide standards for information resource interoperability | 1.4.1. CGIAR Core metadata schema implemented and/or mapped across key publications and data repositories of all 15 Centers; Agronomy Ontology completed and prototype field book tested; Crop Ontology terms adopted as descriptors by at least 6 key repositories; draft for key classes and sub-classes for socioeconomic ontology developed; links to agri-semantics efforts external to CGIAR maintained | Complete | All Centers report mapping to or implementing CG Core metadata schema in their data repositories. |
| | | 1.4.2. Materials developed and shared with appropriate CGIAR communities of practice on standards and ontologies to advocate for best practices and adoption | Complete: Agronomy Ontology | The Agronomy Ontology (AgrO) is openly available on GitHub to enable interoperability of agronomic resources. |
| | | | Complete: Prototype fieldbook developed, tested | Early agronomy field book prototype employing AgrO and the Crop Ontology to standardize data at collection stage tested at Big Data Convention, with the user feedback determining further development: AgroFIMS prototype |
| | | | Complete: Four Centers | The use of ontologies and/or controlled vocabularies as descriptors for resources uploaded to Center repositories was encouraged and supported via discussion and webinars |

| | | | | |
|---|---|---|---|---|
| | | | start using standards as metadata descriptors | in 2017 as a way to include interoperability standards to make CGIAR resources "FAIRer". CIAT, CIMMYT, CIP, and Bioversity publications and data repositories (8 repositories) employed these standards in 2017. |
| | | | Extra: Metadata curation tool | A metadata curation tool was developed toward the end of 2017 to help Centers assess completeness of CG Core metadata use; 4 Centers began using it by year-end:<br>Bioversity<br>AfricaRice<br>CIAT<br>IFPRI |
| | | 1.4.3. At least one workshop/training for data/repository managers and ontologists held on semantic standards to render datasets accessible. | Complete: Materials | A number of presentations, webinars, and standards were developed/improved/organized in 2017 for members of key CGIAR CoPs (OAWG, DMTF, Big Data Focal Points, Big Data CoPs, CGIAR leadership and others as appropriate) to advocate for and support FAIR resources. Selected examples: Ontologies overview; Module 1 2017 webinar series; FAIR ecosystem presentation; Agronomy Ontology |
| | | | Complete: Training | Two workshops were organized and held (April 2017) to bring together appropriate CoPs to share learning and tools, and consistent messaging and advocacy for best practices in data management for "FAIRer" resources:<br>DMTF 2017 agenda, notes, highlights<br>OAWG 2017 agenda, notes, highlights.<br>Support for these efforts continued throughout 2017 via the Module 1 2017 webinar series |

| | | | | |
|---|---|---|---|---|
| **2. Convene** | 2.1. CGIAR expertise leveraged in partnership and established as a thought leader | 2.1.1. Fully establish six Communities of Practice around topics of geospatial data, economic data, ontologies, data- driven agronomy, livestock data for decision-making and crop modelling | Complete | The Platform established Communities of Practice (CoPs) around six relevant topics : socioeconomic data, geospatial data, data-driven agronomy, crop modelling, livestock data for decision-making, and ontologies |
| | | 2.1.2. Identify high priority, high impact new data products and develop methodological plan to produce them, with initial implementation | Ongoing | Multiple consultative sessions at the Big Data in Agriculture Convention helped capture of key data assets to be prioritized for investment and plans for their sourcing and investment were subsequently developed |
| | | 2.1.3. Map out CGIAR needs for common Big Data-related computing and storage infrastructure. Establish shared service for CGIAR by negotiating with external data utility partners | Ongoing | A dedicated session on information architecture at the Big Data Convention started the process of identifying common infrastructure needs.<br><br>This was followed by a requirements gathering survey among the Centers, as a first attempt at surfacing some shared business needs across our network organization that the Platform can help to address through shared data services. The results showed a wide array of capabilities and needs, requiring further segmentation and analysis |
| | 2.2. Enhanced collaboration between Centers and CRPs on data analytics and ICTs | 2.2.1. Hold high-level Convention on Big Data in Agriculture, with wide participation of CGIAR and non-CGIAR actors, establishment of collaborative agreements | Complete | The Platform produced the first annual CGIAR Convention for Big Data in Agriculture in September 2017, convening some 300 global innovators, researchers, and thought leaders from public, private and non-profit partners in Palmira, Colombia |

| | | | | |
|---|---|---|---|---|
| **3. Inspire** | 3.1 Lead by example and inspire how Big Data can deliver development outcomes | 3.1.1. 5 Inspire projects selected, generating innovative, data-driven solutions to CGIAR research and development problems | Complete | Twelve finalists defended their proposals at the Big Data in Agriculture Convention, and external judges selected five of them for $100,000 innovation grants. Among the 5 awardees, 3 were from Pest and Disease, one was from Revealing Food Systems, and 1 was from Data-Driven farming |

## Table C: Cross-cutting Aspect of Outputs

| Cross-cutting | Number (%) scored 2 (Principal) | Number (%) scored 1 (significant) | Number (%) scored 0 | Total overall number of outputs |
|---|---|---|---|---|
| **Gender** | | | 100% | 24 |
| **Youth** | | | 100% | |
| **CapDev** | 25% | 25% | 50% | |

# Table D: Common Results Reporting Indicators

## Table D-1: Key Platform Results from 2017, in Numbers

| Sphere | Indicators | Data | Comments |
|---|---|---|---|
| Influence | **I1/I2*. Projected uptake (women and men) /hectares from current CRP investments (for innovations at user-ready or scaling stage only – see indicator C1)** | New indicator being introduced in 2018. | See indicator C1 |
| | **I3. Number of policies/ investments (etc) modified in 2017, informed by CGIAR research** | N/A | Not relevant |
| Control | **C1. Number of innovations by phase - new in 2017** | 4 innovations at proof of concept;<br>4 piloted successfully; and<br>4 user-ready innovations.<br>Usage unclear; analytics to be implemented in 2018 (policy change not relevant). | Four data analytics and visualization tools developed, piloted, and available for users: (1) CeRes harvester to enable discovery of CGIAR information resources; (2) "what if" feature on CeRes to demonstrate the power of the Semantic Web in data exploration and reuse; (3) Map feature on CeRes, to visualize CGIAR mandate crop data; (4) "relevant resources" algorithms powering CeRes to link related resources (e.g. pubs relevant to "gender" data query) |
| | **C2. Number of formal partnerships in 2017, by purpose (ongoing + new)** | 3 research partnerships;<br>0 policy partnerships; and<br>4 delivery partnerships.<br>See Table B for more detail. | Research partnerships include: (1) University of Minnesota's International Agroinformatics Alliance collaboration on secure data handling and cleaning and ontologies; (2) Wageningen University and Research to explore human and machine data curation; and (3) Agroknow Inc. to pilot Semantic Web application to agricultural data exploration and mining. |

| | | | |
|---|---|---|---|
| | | | Delivery partnerships include: (1) Agroknow Inc. to develop and deliver the CeRes harvester; (2) Spatial Dev to develop Map feature on CeRes; (3) Ontocale Inc., to support exploration of Semantic Web technologies and development of ontology-based Agronomy Fieldbook; and (4) Bioversity International for all ontology-based tools, including delivery of Agronomy Ontology (AgrO), metadata curation tool, and related training sessions. |
| | **C3. Participants in CGIAR activities 2017**<br>(new + ongoing) | XXX 'end-users' (XX% women) in on-farm trials, farmer field days and similar: N/A<br><br>360 'next users' (40% women) in innovation platforms, policy workshops and similar. | Workshops, webinar series on tools, approaches, best practices and advocacy for open and FAIR resources: 60 CGIAR personnel; Big Data Convention plenary sessions: 300 attendees (approx. 100 CGIAR); Five Big Data Platform CoPs as incubators for innovation and best practice |
| | **C4. People trained in 2017** | Long term (new + ongoing): XXX (XX% Women): N/A<br>Short term: 30 (40% women) | Approx. 30 key data management personnel across CGIAR (15 information specialists; 15 data managers). |
| | **C5. Number of peer-reviewed publications** | 1 in 2017 [please add link to full list of Platform publications]<br>of which 1 (100%) are openly published<br>XX (XX%) is with open database: N/A | **Improving global integration of crop research**<br>http://science.sciencemag.org/content/357/6349/359#BIBL |
| | **C6. Altmetrics** | New indicator being introduced in 2018 – details TBA | |

**Please note: I = Sphere of Influence and C = Sphere of Control*

## Table D-2: List of Platform Innovations in 2017 (From indicator #C1 in Table C-1)

| Title of innovation (minimum required for clarity) | Phase of research * | Novel or adaptive research | Contribution of Platform (sole, lead, contributor) | Geographic scope: innovations in phases AV* or USE* only (one country, region, multi-country, global) |
|---|---|---|---|---|
| (1) CeRes harvester to enable discovery of CGIAR information resources | AV | Novel | Lead | Global |
| (2) "what if" feature on CeRes to demonstrate the power of the Semantic Web in data exploration and reuse | AV | Novel | Lead | Global |
| (3) Map feature on CeRes, to visualize CGIAR mandate crop data | AV | Novel | Contributor | Global |
| (4) "Relevant resources" algorithms powering CeRes to link related resources (e.g. pubs relevant to "gender" data query) | AV | Novel | Lead | Global |

# Table E: Intellectual Assets*

| Year reported | Applicant(s) / owner(s) (Center or partner) | Patent or PVP Title | Additional information* | Link or PDF of published application/ registration | Public communication relevant to the application/registration |
|---|---|---|---|---|---|
| | | | | | |

**Not relevant for Platform as it has, thus far, focused on highlighting or generating resources and tools that are openly available*

# Table G: List of Key External Partnerships

| Module | Stage of research* | Name of partner | Partner type* | Main area of partnership* |
|---|---|---|---|---|
| 1. Organize | Research | 1. University of Minnesota International Agroinformatics Alliance | **Academic and Research** | Collaboration on secure data handling and cleaning |
| | Research | 2. Wageningen University and Research | **Academic and Research** | Exploration of human and machine data curation |
| | Pilot application (Semantic Web); Scaling up & scaling out | 3. Agroknow Inc | **Private sector** | Pilot of Semantic Web application to agricultural data exploration and mining, and development and delivery of the CeRes harvester |
| | Scaling up & scaling out | 4. Spatial Development International | **Private sector** | Development of Map feature on CeRes; |
| | Research | 5. Ontocale Inc. | **Private sector** | Support for exploration of Semantic Web technologies and development of ontology-based agronomy fieldbook |

# Table H: Status of Internal (CGIAR) Collaborations between the Platform and Programs and among Platforms

| CRP or Platform | Brief description of collaboration (give and take between the Platforms and CRPs) and value added* | Relevant Module |
|---|---|---|
| | Information architecture session and findings at Convention with EiB and Genebanks/Crop Trust, setting the agenda for coming year. | 2 |

# Table I: Monitoring, Evaluation, Impact Assessment and Learning

## Table I-1: Status of Evaluations, Impact Assessments and Other Learning Exercises Planned in the 2017 POWB*

| Studies/learning exercises in 2017 (from POWB) | Status | Comments |
|---|---|---|
| | | |

## Table I-2: Update on Actions Taken in Response to Relevant Evaluations (IEA, CCEEs and Others)*

| Name of the evaluation | Recommendation | Management response – Action Plan | By whom | By when | Status |
|---|---|---|---|---|---|
| | | | | | |

**N/A; the Platform was only active for a few months in 2017*

# Table J: Platform Financial Report

| | Planned budget 2017 | | | Actual expenditure 2017* | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|
| | W1/2 | W3/bilateral | Total | W1/2 | W3/bilateral | Total | W1/2 | W3/bilateral | Total |
| Module 1 | 4,285 | 8 | 4,293 | 1,365 | 8 | 1,373 | 2,920 | - | 2,920 |
| Module 2 | 1,058 | 78 | 1,136 | 326 | 63 | 389 | 733 | 15 | 748 |
| Module 3 | 835 | - | 835 | 102 | - | 102 | 733 | - | 733 |
| Strategic | - | - | - | - | - | - | - | - | - |
| Competitive Research grant | | | - | - | - | - | - | - | - |
| Platform Management & Support Cost | 527 | - | 527 | 505 | - | 505 | 22 | - | 22 |
| **Platform Total** | **6,705** | **86** | **6,791** | **2,297** | **71** | **2,368** | **4,408** | **15** | **4,423** |

**Audited lead Center financial report*