Evaluation of CGIAR Platform for Big Data in Agriculture

I. Jouini, D. Leibovici, M. Kurian, E. Bongcam-Rudloff, J. Kieti, and S. Sellitti

Evaluation of CGIAR Platform for Big Data in Agriculture

I. Jouini, D. Leibovici, M. Kurian, E. Bongcam-Rudloff, J. Kieti, and S. Sellitti

December 2021

Acknowledgments

Table of Contents

Annexes are provided at this link.

List of Tables

List of Figures

List of Boxes

Table of Acronyms & Abbreviations

AEA	American Evaluation Association
AI	Artificial Intelligence
API	Application Programming Interface
AR4D	Agricultural Research-for-Development
CapDev	Capacity Development
CAS	CGIAR Advisory Services
CGIAR	A global research partnership for a food secure future
CGLabs	Collaborative GARDIAN Labs
CGSpace	CGIAR Digital Repository
CIAT	International Center for Tropical Agriculture
CoP	Community of Practice
CRPs	CGIAR Research Programs
CS/s	Component Study/Studies
EiA	Excellence in Agronomy Platform
ET	Evaluation Team
FAIR	Findability, accessibility, interoperability, and reusability
FISH	CGIAR Research Program on Fish Agri-Food Systems
GARDIAN	Global Agriculture Research Data Innovation Acceleration Network
GDI	Gender, Diversity and Inclusion
GDPR	General Data Protection Regulation
GIZ	Gesellschaft für Internationale Zusammenarbeit
IAB	International Advisory Board
ICT4D	Information and Communications Technology for Development
IDO	Intermediate Development Outcome
IFPRI	International Food Policy Research Institute
ISDC	Independent Science for Development Council
M&E	Monitoring and Evaluation
MARLO	Managing Agricultural Research for Learning and Outcomes
MEL	Monitoring, Evaluation, and Learning
MELIA	Monitoring, Evaluation, Learning, and Impact Assessment
NGO	Nongovernmental Organization
NARES	National Research and Agriculture and Extension Systems
OA	Open Access
OD	Open Data
OECD DAC	Organization for Economic Co-operation and Development/Development Assistance Committee

PII	Personally Identifiable Information
Platform	CGIAR Platform for Big Data in Agriculture
QA	Quality Assurance
SC	Steering Committee
SED-CoP	Socio-Economic Data Community of Practice
SIMEC	Strategic Impact, Monitoring and Evaluation Committee
SLO	System-Level Outcome
SME	Subject Matter Expert
SPIA	Standing Panel on Impact Assessment
ToC	Theory of Change
ToR	Terms of Reference
WGs	Working Groups

Executive Summary

Background and Context

In 2017, CGIAR launched a five-year Platform for Big Data in Agriculture (hereinafter the Platform) to enhance the development, curation, and maintenance of its data and knowledge assets and to stay at the cutting-edge of the rapidly evolving digital world. The Platform is a coordinating mechanism to deliver a coherent data-driven and data-intensive CGIAR strategy through collaboration among CGIAR Research Programs (CRPs) and CGIAR Centers (hereinafter Centers), and by leveraging external expertise to enable unrestricted discoverability of linked open datasets.

While the Platform is expected to run until the end of 2021, the digital revolution remains a priority to CGIAR as it transitions to One CGIAR, making the findings of this evaluation especially pertinent. The digital revolution is specifically singled out as one of seven implementation approaches prioritized in the new CGIAR 2030 Research and Innovation Strategy (hereinafter the CGIAR 2030 Strategy). The designation of the digital revolution as the 'seventh way of working' seeks to use the opportunities it presents to expertly manage data and use it effectively to accelerate progress towards the achievement of the Sustainable Development Goals (SDGs).

Purpose and Scope

The CGIAR Advisory Services Shared Secretariat (CAS Secretariat) was mandated to conduct an external evaluation of the Platform, serving the dual purposes of accountability and learning, as part of its 2021 approved workplan and budget. It assessed the Platform in terms of design, scope, implementation status, and capacity to achieve its objectives, covering all its activities from its initiation in 2017 through mid-2021, considering the available evidence, the transition of CGIAR to One CGIAR[1], and the COVID-19 pandemic.

Methods for the Evaluation

The evaluation approach followed a mixed-methods design (qualitative and quantitative) to collect data and assess the Platform's achievements and outcomes. Qualitative methods consisted of 51 semi-structured interviews, document analysis, and the preparation of two short case studies. Quantitative methods included an online survey (110 Platform stakeholders responded) and usage analytics (GARDIAN[2]). Taking a two-stage analysis approach, the evaluation team first conducted three Component Studies (CSs), each of which covered a cluster of activities corresponding to the three Platform modules: ORGANIZE (Module 1), CONVENE (Module 2), and INSPIRE (Module 3). The results of the CSs were first peer-reviewed and validated which then served as the main input for this final evaluation report. The evaluation implementation faced some limitations that were addressed to the extent possible including the limited availability of stakeholders, and the underrepresentation of three (both internal and external) stakeholder groups: women, youth, and Inspire Challenge[3] (grant scheme) applicants. These limitations impeded the collection of enough evidence to make conclusive claims specific to these groups.

---

[1] At the time of publishing this report, CGIAR is transitioning to One CGIAR.
[2] The Global Agriculture Research Data Innovation Acceleration Network (GARDIAN) is an online portal that provides access to agricultural research data produced by Centers and their implementing partners.
[3] The Inspire Challenge is CGIAR's signature digital innovation process. It leverages the global footprint and deep food security subject matter expertise of CGIAR with expert industry partners to link digital technologies to impact in developing economies.

Key Findings

Key findings presented here are structured around the four evaluation criteria: Relevance, Efficiency, Effectiveness, and Sustainability:

Relevance

1. Consulted stakeholders considered the Platform outputs highly relevant and aligned with One CGIAR and Centers' priorities but unable to sufficiently streamline Centers' varying data management approaches or leverage existing capabilities. The lack of explicit articulation of the Platform's theory of change (ToC) as a 'transformer' for agricultural research-for-development (AR4D) was identified as a design weakness.
2. Gender was not strongly embedded in the Platform's design. Consequently, adequate resources (budget and expertise) were not provided systematically to guide efforts to mainstream gender. responsiveness.
3. Capacity development activities have been steadily integrated into the Platform modules, guided by the CGIAR Capacity Development (CapDev) Framework.
4. Youth engagement was planned in the design of the Platform during implementation, but not enough attention was given to building specific and continuous mechanisms to reach foreseen outcomes and to include youth in decision-making (i.e., the Platform steering committee, Inspire Challenge selection committee).
5. Climate change was considered in the Platform design and tools created were appropriate with expected outcomes.
6. The Platform demonstrated the capacity to adapt to evolving environments and constraints (GARDIAN syntactic interoperability, deployment of COVID 19 Rapid Response grants).
7. Much work remains to enhance the semantic, syntactic, and structural interoperability, which brings great value for predictive analytics and data reuse.

Efficiency

8. The Platform sought to leverage CGIAR's 'convening power,' the management team was the chief decision-making body complemented by mechanisms for technical collaboration such as open technical Communities of Practice (CoPs) and an innovation process to move the agenda of (big) data management and analytics in agriculture research forward, yet this was not enough to foster Centers' engagement with the Platform's outputs.
9. Lack of knowledge about the Platform's mandate among internal CGIAR stakeholders affected the quality of engagement (of stakeholder groups) with its modules' activities and outputs.
10. Stronger linkages between the three modules would have increased the efficiency of the Platform to advance the use of big data analytics.
11. Related activities on cross-cutting themes (i.e., gender, youth, climate change, capacity development) have been implemented but results are still preliminary and need to be properly monitored and evaluated, particularly for CONVENE and ORGANIZE module outputs.

Effectiveness

12. In Centers, the evaluation indicated an increased awareness of Open Access/Open Data (OAOD) including Findable, Accessible, Interoperable, Reusable (FAIR) principles. The Platform contributed to this outcome by constant advocacy during its annual conventions and capacity development activities. However, evidence shows that relatively more effort is still needed to progress data Interoperability and Reusability.
13. Weak evidence was found about GARDIAN usage from end-users. Nonetheless, users are spending more time browsing the results of their queries. Not enough evidence was found about the usage of the Platform's data analytic facility (CG Labs) to make an assessment.
14. The Platform contributed to increased engagement (frequency and depth of interactions) between CGIAR researchers and stakeholders from the digital ecosystem. It resulted in standardization efforts: development and use of terms for data comparisons and reuse

(e.g., Ontologies CoP, and Information and Data Managers CoP), digital extension (e.g., Data-Driven Agronomy CoP), and modeling (Crop Modeling CoP). The consulted stakeholders considered the Ontologies CoP as instrumental in persuading CGIAR colleagues to pay attention to ontologies, not as mere academic standards but as mechanisms for unlocking the potential of big data through semantic interoperability).

15. Platform outputs added value to CGIAR efforts to map data, methods, and tools, to support the delivery of research.

16. Proposals selected under the Inspire Challenge helped to catalyze the development of new digital methods for research or delivery of research at CGIAR, meeting one of the Platform's objectives. Yet, it could have implemented a more robust Monitoring and Evaluation (M&E) system to harvest the outcomes of these projects.

17. The Platform prepared CGIAR for compliance with the General Data Protection Regulation (GDPR), it developed a tool to help data managers detect the appearance of any personally identifiable information (PII) that might appear in their repositories and contributed to raising awareness about issues of data sharing in addition to enabling discoverability and findability of CGIAR data. However, data were not checked for reliability and were devoid of markers that can provide a context for research and analysis, which the Platform team considered beyond its mandate. The absence of reliability checks and markers, in addition to other factors, severely limits the ability of the Platform to advance data interoperability and reusability.

18. More work is required to advance the data interoperability principle to safeguard the quality of CGIAR proposals/recommendations and strengthen its claim to the attribution of its value in AR4D from its existing data. Limited evidence was found on whether the Platform's quality control efforts improved (or not) the discoverability of CGIAR data.

19. The Platform's Monitoring, Evaluation, and Learning (MEL) activities did not adequately support programmatic learning and reporting (successes, failures, and lessons learned). More could be done so that the Platform can serve as a mechanism that promotes CGIAR's function as a "learning organization".

Sustainability

20. The Platform was instrumental in the revision of the CGIAR OAOD policy and large challenges await in the transition to One CGIAR. The policy implies harmonizing and standardizing publications and data libraries – work that the Platform has already started mainly through GARDIAN and CoPs.

21. CGIAR is well-positioned but not sufficiently prepared to have a leadership voice in international digital agriculture according to views of its internal stakeholders. Overall, stakeholders valued its outputs and seek its continuity. External partners' engagement was undermined by the funding uncertainty of its e-infrastructure (GARDIAN).

Lessons Learned

Key lessons learned during the evaluation include:

1. Data curation needs to be standardized, fit a well-defined set of requirements, and be made available to end-users with proper incentives and training in quality assurance and documentation. Without interoperability, big data, and big answers will never be achieved.

2. End-users will not be able to easily reproduce the prototypes (fully described from these cross-module activities as an open science contribution) without cross-module activities (i.e., interoperability, analytics, innovation) that lead to demonstrable proofs of concept and useful prototype capabilities (along with thorough documentation and transparent description).

3. "If you build it, they will come" does not apply to initiatives and digital artifacts for big data in AR4D. Additional efforts are required to increase awareness and uptake and early-adopter feedback is critical to iteratively refine products. This feedback increases usefulness and ease of use which are important antecedents of large-scale adoption Davis (1989).

4. Building trust with and engaging all Centers in decision-making are important to ensure wide acceptance and adoption of any new, centralized technology and solutions. Existing digital solutions (such as CGSpace, digital workflows to collect standardized data) and their value should

be acknowledged and built upon as One CGIAR develops a more unified digital strategy, data and digital technology governance, and an approach for rationalizing information technology investments.

5. The integration of gender as a transversal theme remains patchy with no gender expertise in-house. Piecemeal gender is less effective and less cost-effective in the long run.

Recommendations

Recommendations[4] to improve the design and implementation of the Platform for Big Data in Agriculture that are also relevant to other CGIAR projects with a similar goal include:

1. Prioritize specific digital solutions for specific data (domains) aligned with agricultural research needs to demonstrate the value of the answer (big) data can provide to support CGIAR's key priorities.
2. Prioritize and advance the interoperability agenda, building on CGIAR's wide variety of datasets.
3. Strengthen the conceptualization (theory of change) of how the impact of agricultural development can be increased by embracing big data and ICT approaches to solve development problems faster, better, and at a greater scale.
4. Raise Centers[5]' (CGIAR Entities') engagement to ensure technology solutions uptake: this can be achieved by an inclusive governance system and leveraging existing tools and incentives.
5. Build a new harmonized and interoperable analytical environment in CGIAR based on accumulated knowledge from the experience of the Platform's implementation.
6. Improve grant scheme management, monitoring, and governance to foster the Platform's (or successors') relevance to contribute to solving agriculture development challenges.

In light of the incoming implementation of One CGIAR new Initiatives, CGIAR's 7th way of working, and the CGIAR 2030 Strategy, this evaluation of the Platform for Big Data in Agriculture recommends the CGIAR System to:

7. Develop a One CGIAR (research) digital capability model and ensure the funding for a long-term digital plan with successive phases and a clear mandate that builds on the Strategic Research on Digital Transformation assessment.
8. Lead the way in hosting open data and providing analytic tools for CGIAR and its partners as well as increasing the data and funding (by showing its value).
9. Develop data synthesis tools that are amenable for use by decision-makers to support data co-curation.
10. Develop a data curation and transformation dashboard to enable CGIAR and partners to access tools and technical support to undertake data harvesting, data harmonization, and visualization.

---

[4] Operational recommendations can be consulted in section five of the report 'Conclusions and Recommendations.'
[5] 'CGIAR Entities' going forward in One CGIAR

1 Introduction

1.1 Structure of the Report

This evaluation report presents the results of the Platform evaluation in five sections. Following this introductory section, the report sets out an overview of the rationale and background of the evaluation of the CGIAR Platform for Big Data in Agriculture (hereinafter the Platform) as well as an overview of the Platform structure and objectives. It then presents the evaluation objectives and questions as laid out in the evaluation Terms of Reference (ToR), followed by a description of the evaluation approach, methodology, phases, data collection methods, and limitations. This is followed by the presentation of the key findings structured around the evaluation criteria: Relevance, Efficiency, Effectiveness, and Sustainability. Finally, it presents the conclusions and recommendations. Annexes are an important part of the report, they are gathered in a separate document (titled: CGIAR Big Data in Agriculture Platform Evaluation Report: Annexes). The Annexes include among others, executive summaries of the three Component Studies, a Community of Practice (CoP) case study, a detailed methodology, a revised evaluation matrix, the results of consultations, and the profile of people consulted.

1.2 Background of the Evaluation[6]

As part of its 2021 approved workplan and budget, the evaluation function of the Advisory Services Shared Secretariat (CAS) was mandated to conduct an external evaluation of CGIAR's Big Data in Agriculture Platform. In addition to its independent evaluation function, the CAS Secretariat also supports and facilitates the CGIAR's Independent Advisory Services (CAS), comprising the Independent Science for Development Council (ISDC) and the Standing Panel on Impact Assessment (SPIA). The evaluation function supports the implementation of the CGIAR System Multi-Year Evaluation Plan to meet CGIAR System needs for rigorous high-quality independent evaluations that inform decision-making across the System.

An evaluative study related to the Platform was commissioned by CAS Secretariat's predecessor, CGIAR's Independent Evaluation Arrangement in 2018; a review of CGIAR's open access/open data policy and implementation support. In addition, since its inception in 2017, the Platform has been the subject of several other reviews and studies, worthy of mention:

- In response to CGIAR System Management Board's request for a digital strategy that identified CGIAR's comparative advantage relating to big data,[7] a strategic research study was conducted on digital transformation in food, land, and water systems in a climate crisis in support of the CGIAR 2030 Strategy. The Strategic Research on Digital Transformation assessment was published in 2021 by the Platform.

- In 2021, an internal review assessed the Platform's Inspire Challenge Program (2017-2020) and its broader contributions to catalyze partnerships and digital agricultural innovations.

Against this backdrop, the recent 2021 Synthesis of Learning from a Decade of CGIAR Research Programs brought to the fore thematic evidence gaps related to digital innovations that revealed limitations on the evaluative evidence available on digital innovations. Hitherto, there has been no comprehensive independent evaluation of the Platform in its entirety.

1.3 Overview of the Platform for Big Data in Agriculture

1.3.1 Purpose and Objectives

The Platform is a five-year initiative established to develop, curate, and maintain CGIAR data and knowledge products, leveraging data capabilities and infrastructure through opportunities presented by the rapidly developing digital world. The Platform operates as a coordinating mechanism to deliver a

---

[6] This section was excerpted from the Platform ToR
[7] Chair's Summary, 13th CGIAR System Management Board (SMB) Meeting. Approved May 3, 2019

coherent data-driven and data-intensive strategy on behalf of CGIAR through collaboration among CGIAR Research Programs (CRPs) and Centers and by leveraging external expertise to enable unrestricted discoverability of linked open datasets. The Platform, approved by System Council, was launched in 2017 with the ultimate goal to "… *harness the capabilities of Big Data to accelerate and enhance the impact of international agricultural research. It will support CGIAR's mission by creating an enabling environment where data are expertly managed and used effectively to strengthen delivery on CGIAR SRF's System Level Outcome (SLO) targets".*[8]

The Platform aims to increase the impact of agricultural development by embracing big data and ICT approaches to **solve development problems faster, better, and at a greater scale.** As outlined in the CGIAR Strategy and Results Framework (SRF), this will initially be across CGIAR but will be extensible to agriculture at large.

According to the final July 2016 Proposal (hereinafter the 2016 Proposal), the Platform focuses *on "enhancing CGIAR and partner capacity to deliver big data management, analytics and ICT-focused solutions to CGIAR target geographies and communities through its ambitious partnerships with both upstream and downstream partners."* CGIAR defines big data in the proposal as *"harmonized, interoperable, and contextually integrated datasets and publications from multiple disciplines relevant for CGIAR's research and development goals[9]."*

In addition to developing new partnership models with big data leaders at the global level, the Platform seeks to promote CGIAR-wide collaboration across CRPs and Centers.

The Platform's theory of change focuses on increasing the capacity of CGIAR and partners to embrace big data and information and communication technology (ICT) approaches as shown in Figure 1.

---

[8] Big Data Coordination Platform: Full Proposal 2017-2020

[9] CGIAR Consortium 2015a

Figure 1. Initial impact pathways for the CGIAR Platform for Big Data in Agriculture

In the 2016 Proposal, the authors acknowledge that the primary **assumption** of this ToC is that data are a valued commodity that can be harnessed to deliver growth in agriculture in developing countries. It also assumes that CGIAR and its partners can identify business opportunities where rural institutions are weak to deliver benefits to marginalized smallholder farmers. The Platform operationalizes its tripartite objectives via three modules – (1) ORGANIZE, (2) CONVENE, (3) INSPIRE – as illustrated in Figure 2.

Figure 2. Platform modules and their objectives

Module 1:

ORGANIZE

Organizes extant data and draws them together for unified and interlinked discoverability. It assesses their status and fitness for use, identifies what and where gaps exist, and strengthens its analytical capacity for data-driven impact.

Objective 1: Support and improve data generation, access, and management in CGIAR.

Module 2:

CONVENE

Convenes the scientific resources across CGIAR with a range of partners to generate new collaborative opportunities and bring big data to agriculture, and likewise, agriculture to big data.

Objective 2: Collaborate and convene around big data and agricultural development.

Module 3:

INSPIRE

Inspire works on big data by funding research by CGIAR scientists with partners to innovate new ways to handle big data relevant to agriculture for the benefit of poor smallholder farmers.

Objective 3: Lead by example and inspire how big data can deliver development outcomes.

Source: Adapted from CIAT and IFPRI (2016). Big Data Coordination Platform. Full Proposal (final version). July 2016.

1.3.2 Management, Governance, and Funding

The Platform is co-led by the International Center of Tropical Agriculture (CIAT)[10] and the International Food Policy Research Institute (IFPRI), CGIAR centers. CIAT's leadership takes fiduciary responsibility, signing the performance contract with the CGIAR system office, and consults IFPRI's leadership as needed. Implementation is through a secretariat. The Platform has an eight-member steering committee reporting formally to the Alliance board, whose Chair and Director-General, in turn, report to the System Office on the Platform as a whole. A five-member International Advisory Board (IAB) was set up. Its role was to explicitly examine how the Platform connects effectively with other global and regional efforts for continued relevance and novelty.

According to the 2016 Proposal, the Platform had a 6-year budget of US$30.2m primarily from Windows 1 & 2, representing an annual budget ranging from US$3.9m to US$6.7m. In terms of the budget allocation per module, ORGANIZE (Module 1) received the largest budget share in 2017 (68% total) and 2018 (58%) with the main cost driver being funding to Centers aimed at improving the effective management of CGIAR data and compliance with the Open Access, Open Data (M) Policy. CONVENE (Module 2)'s budget in 2017 was US$1.46 with a progressive growth by a standard 5% annually to maintain the fixed costs associated with creating an enabling environment. Similarly, INSPIRE (Module 3)'s budget was projected to double by its fourth year from Year 1 (US$0.6m) to year 4 (US$1.31m). The budgeted cost for the Platform Secretariat was pegged at US$300k in the 2016 Proposal and was covered under CONVENE. The Platform Secretariat budget made allocations to cross-cutting themes such as capacity building (40%), gender- and youth-related activities (17%).

1.3.3 Looking Ahead to Big Data in One CGIAR

While the Platform is expected to run until the end of 2021, the digital revolution is one of seven new implementation approaches prioritized in the new CGIAR 2030 Research and Innovation Strategy (hereinafter the CGIAR 2030 Strategy). The 'seventh way of working' seeks to use the unprecedented opportunity provided by today's digital revolution to accelerate progress towards the achievement of the Sustainable Development Goals (SDGs). Key elements of the digital revolution approach include engagement with partners to develop cutting-edge, context-appropriate digital solutions; improving access to and the use of data and digital innovations that target small-scale farmers; and pursuing new

---

[10] The International Center for Tropical Agriculture (CIAT) became part of the Alliance of Bioversity International and CIAT in 2020.

digital applications to accelerate learning and knowledge sharing among partners underpinned by principles of findability, accessibility, interoperability, and reusability (FAIR) for all CGIAR data.

2 Evaluation Scope, Objectives, and Questions

This evaluation serves the dual purposes of accountability and learning. It is both summative and formative and it assesses the design, scope, implementation status of the Platform, and its capacity to achieve its objectives. The evaluation has collated and analyzed lessons learned, challenges faced, and best practices as a guide for future planning. It provides essential evaluative evidence for decision-making by the CGIAR System Council, Big Data Platform management, and its partners.

The evaluation **scope** covers all the activities of the Platform from its initiation in 2017 through mid-2021 considering the need for timely evidence, the transition to One CGIAR, and the COVID-19 pandemic. The evaluation integrated cross-cutting themes of Gender, Youth, Climate Change, and Capacity Development as well as the key issue of Open Data and Open Access (ODOA).

The main **objectives** of the evaluation of the Big Data in Agriculture Platform (2016) are to:

A. Assess the relevance of the Platform design, theory of change (ToC), and its role in positioning CGIAR as a learning organization, its ability to cultivate new digital alliances and, pursue data innovation in support of its mission.

B. Identify the supporting factors and constraints behind the achievements of the Platform and each of its modules and the validity of the ToC assumptions in light of the results achieved, including its response to COVID-19

C. Assess the Platform's governance, management, and implementation processes.

D. Provide recommendations relevant to the future development and implementation aligned with One CGIAR Way of Working 7 – Making the Digital Revolution Central to Our Way of Working and One CGIAR initiatives related to digital technologies. This is to include inter alia "Harnessing Digital Technologies for Timely Decision-Making across Food, Land, and Water Systems" (Systems Transformation Action Area) and other system-wide recommendations where applicable.

The evaluation examined the Platform implementation against the hereunder Development Assistance Committee (DAC) criteria - defined by the OECD DAC Network on Development Evaluation- by addressing the following (broad but not exhaustive) questions.

Table 1. Evaluation criteria and questions

DAC Criteria	Key Evaluation Questions[11]
Relevance	1. To what extent are the Platform's objectives relevant to the needs of its internal and external partners, including end-users in target groups?
Efficiency	2. Have resources (funds, human resources, time, expertise, etc.) been allocated strategically and timely to achieve Platform outcomes?
Effectiveness	3. To what extent did the Platform achieve progress towards outcomes? 4. How effective has the Platform been in building digital capabilities and partnerships supporting CGIAR research?
Sustainability	5. To what extent are the Platform products and communities positioned to be effective in the future, seen from the perspectives of scientists and of the end-users of digital agriculture products and innovations? 6. To what extent would the Platform outputs outlive the existence of the Platform in its current form?

---

[11] Evaluation questions were revised by the evaluation team and CAS Secretariat.

3 Evaluation Methodology

The evaluation approach followed a mixed-methods design (qualitative and quantitative) to collect data and assess the Platform's achievements and outcomes. Among quantitative methods the evaluation team designed and administered an online survey to reach a wide range of stakeholder groups —110 data users and partners responded to the survey. Initially, the survey was sent to 2,803 subscribers of the Platform's newsletter. The qualitative methods consisted of 51 semi-structured interviews (53 people participated), document analysis, and two short case studies (on a community of practice and a winner of the Inspire Challenge[12]). Both qualitative and quantitative methods complemented each other in ways that brought robust evidence to answer the evaluation questions. The evaluation methods tended to embrace a systems-thinking approach to capture interlinked issues in this innovative program. Mixed methods were chosen to deal with the complexity in the evaluation based on the five dimensions of complexity presented by (Bamberger *et al* 2016). Qualitative inquiry tended to be exploratory using open questions and snowball sampling. The assessment of data management and stewardship followed the 'FAIR Guiding Principles' (Wilkinson et al., 2016).

A two-stage analysis was conducted: Three Component Studies (CSs) results were first peer-reviewed and validated which then served as the main input for this final evaluation report. The evaluation matrix forms the main analytical framework for the Platform evaluation (see Annex 2). Several data collection methods were used under the framework of the three CSs; each of which covered a cluster of activities corresponding to the three Platform modules: ORGANIZE (Module 1), CONVENE (Module 2) and INSPIRE (Module 3). The studies followed the same analytical framework centered on the evaluation criteria and questions outlined in the evaluation matrix (see Annex 2). However, not all the CSs benefited from the same team level of effort (there was more focus on ORGANIZE and CONVENE.) The Inspire study was designed to build on the findings of the recent internal review conducted by the Platform team (Mockshell et al. 2021). Some data collection methods were used exclusively for a single CS (i.e., testing the GARDIAN platform) while others have been commonly used across the three studies (Online survey, interviews of the Platform management team, center focal points' interviews...).

Table 2. Key features of the Three Component Studies (CSs)

	CS1- ORGANIZE	CS2- CONVENE	CS3- INSPIRE
Focus	Data generation, access, and management.	Collaboration and conventions around big data and agricultural development.	Inspire challenge competition
Main Target Group	Data users Data managers	Members of Communities of Practice (CoPs), convention participants, and capacity development beneficiaries	Competition candidates
Cross-cutting themes	Gender, Youth, Climate Change, Capacity Development (CapDev)		
Study design	Mixed quantitative and qualitative	Mainly qualitative	Mainly qualitative
Data collection methods*	Platform statistical analytics Interviews stakeholders' survey Desk review	Interviews stakeholders' survey Desk review CoP case study	Interviews Descriptive case study complementing the 2021 Inspire Challenge review Desk review
Sampling technique	Representative and purposeful	Purposeful	Purposeful

* Here the main method for each CS is specified but all CSs used and triangulated with other data collection methods results.

---

[12] The second short case study was not included among the annexes because the subject (individual participant to an Inspire Challenge Competition) preferred to keep his/her identity and story confidential. Nonetheless, his/her experiences and testimony have been included in the evaluation report but not openly assigned to him/her.

The validation of results and quality assurance (QA) relied on triangulating data and findings from different sources and using different methods. This approach also helped to ensure to the extent possible transparency, independence of judgment, and minimizing of bias. In addition, internal and external peer reviews helped strengthen the soundness of the findings and recommendations.

In line with CGIAR evaluation policy and international evaluation quality standards,[13] the evaluation approach ensured the integration of the following principles: participatory, learning-oriented, utilization-focused, and gender-responsive. To the extent that time allowed, the Platform management team was invited to participate actively in the evaluation through the review of the inception report, data collection instruments (online survey protocol and its administration), and in the interpretation of the results. Likewise, data collection was conducted in a way that ensured full understanding, respect, and complete confidentiality of stakeholders' views and perceptions (see interview guide in Annex 3).

Finally, the Platform evaluation was designed as an entirely desk-based exercise; it followed a pre-determined process, guided by the validated terms of reference. The evaluation study considered findings and information elaborated under the framework of internal monitoring and evaluation (M&E) efforts across the Platform implementation and previous efforts of 2021 Synthesis of Learning from a Decade of CGIAR Research Programs. CAS Secretariat's processes guided, and quality assured the evaluation process and deliverables, such as the inception report, data collection instruments, and final report.

For more details about the evaluation design and methodology, see Annex 1.

3.1 Limitations and Mitigation Actions

Remote consultations with stakeholders: Both the nature of the evaluand – a Platform whose stakeholders and end-users are spread all over the world– and the current travel restrictions caused by the COVID-19 pandemic prohibited travel for the evaluation. As a result, the evaluation design was built around data collection methods that could be carried out remotely. To mitigate the shortcomings of conducting online consultations and the effects this model can have on the quality of interactions, evaluators and interviewees had their cameras turned on during meetings. In addition, interviews were designed to be semi-structured and sometimes totally open. Stakeholders were also strongly encouraged to share their ideas freely in the interviews, about what they considered important.

Limited availability of stakeholders and multiplicity of consultations: Some key individuals involved in the Platform had limited availability. In addition, Platform key stakeholders had recently been consulted by the Platform itself around themes and questions similar to those of this evaluation, i.e., in 2020-21, the Platform conducted a review of the Inspire Challenge and the Strategic Research on Digital Transformation assessment. For a few respondents, it gave the impression of repetition to internally driven assessments, asking about the usefulness of the evaluation results. The team presented the evaluation objectives and framework in the invitation email and at the beginning of the interviews. The interviewees were also invited to ask any questions they had before starting, and many did.

Underrepresentation of three stakeholders' groups: **Women and youth** were underrepresented in interviews. Indeed, this gap reflects the underrepresentation of these groups in the Platform's internal stakeholders. To address this imbalance, the evaluation team reached out to more women toward the end of the data collection phase to raise the proportion from 20% to 30% in interviews; women represent 31% among the survey's respondents. Among the survey respondents, 32% were aged below 35. The response rate in the online survey from **Inspire Challenge applicants** was low (only nine [9] respondents out of 110 participated in the Inspire Challenge and only one [1] was selected as a finalist). This made the analysis of the survey results inconclusive for the INSPIRE module (C3). To comply with General Data Protection Regulation (GDPR) requirements, the Platform assisted the external evaluation team to disseminate the survey to a large swathe of stakeholder groups through their curated lists e.g., Communities of Practice.

---

[13] Including that of American Evaluation Association's [AEA] Guiding Principles for Evaluators.

4 Key Findings

Key findings are structured around the four evaluation criteria: Relevance, Efficiency, Effectiveness, and Sustainability.

4.1 Relevance

1: Consulted stakeholders consider the Platform outputs as highly relevant and aligned with One CGIAR and CGIAR Centers' priorities but unable to sufficiently streamline Centers' varying data management approaches and leverage existing capabilities. The lack of explicit articulation of The Platform's Theory of Change (ToC) as a 'transformer' for agricultural research-for-development (AR4D) was identified as a design weakness.
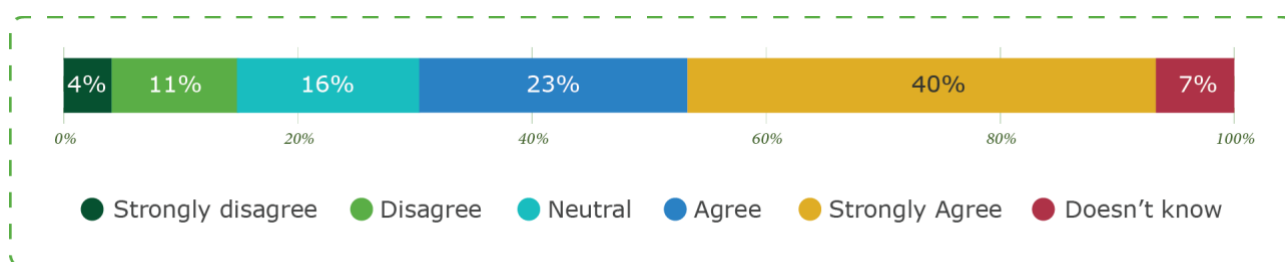
The Platform's ToC illustrates the impact pathways in terms of applications of artificial intelligence, digital services, digital trust, and collective action (See Annex 9). The premise behind the ToC as stated in the 2016 Proposal was "*There is unprecedented innovation at the intersection of digital technologies and life sciences that—if harnessed and applied—can provide the tools humanity needs to adapt to or mitigate some of its most pressing food security challenges*". Consultations showed that there is broad enthusiasm for the goal of maintaining a system for open science and open data. The 2016 Proposal's alignment to open data initiatives and interoperability across many scientific domains existing within CGIAR was deemed appropriate by many stakeholders consulted, e.g., scientists and data managers. Indeed, the movement for open science started in the early 2000's but only became widely adopted as the way forward around 2010 (Kelty 2001, Gewin 2016, and Vicente-Saez and Martinez-Fuentes 2018).

With regards to the completeness of the Platform ToC, the articulation of how its impact pathways could be shaped by incentives, needs, and skillsets of CGIAR researchers was not indicated. Crucially, the Platform's efforts did not explicitly approach big data analytics as a 'transformer' of AR4D. To overcome this gap, the Platform team conducted a pan-CGIAR assessment, which resulted in its Strategic Research on Digital Transformation assessment being released in March 2021. This strategy was developed at the request of the Board, and it is currently pending review of the System Board under One CGIAR.

With regards to the relevance of Platform modules, interviews revealed mixed opinions about the pertinence of the Inspire Challenge. Critics principally noted mainly the lack of feedback to applicants and that the resulting data and software products, did not adequately link back to CGIAR's research or GARDIAN. GARDIAN, which ranks publications and datasets equally, was considered relevant according to stakeholders' testimonies which noted that this allowed better integration of knowledge and understanding of manipulated datasets thereby contributing towards data-driven innovation. The majority of interview participants showed an understanding of the potential benefits of data sharing and OAOD for both datasets and publications. The objectives and *a posteriori* the activities performed in relation to the ORGANIZE module were well-aligned with the perspectives of the most consulted end-users towards increasing the visibility or adding value to their work together with building up collaborations.

Centers have, however, varying capabilities and resources in terms of data management and stewardship, as well as leadership engagement with OAOD. Some stakeholders pointed out that the Platform missed the opportunity to build on an existing bottom-up created tool (cgspace.cgiar.org) and that Centers with fewer resources could have received more support to overcome the capacity gap.

Figure 3. Online survey responses – Big Data Platform's products, analytical tools, and/or activities I engaged in add value to my work (online survey, n=75)

2: Gender was not strongly embedded in the Platform's design. Consequently, adequate resources (budget and expertise) were not provided systematically to guide efforts to mainstream gender responsiveness.

Despite notable efforts to mainstream gender from 2017 to 2020, the Platform was unable to substantially address gender as a cross-cutting dimension through all modules because of a lack of in-house expertise to steadily tackle the digital gender gap. The Platform deployed efforts to integrate gender including involving members of the CGIAR Gender Platform in the review of Inspire Challenge applicants to oversee the gender aspects. For example, each of the Inspire winners' information pages includes a Gender and Youth Inclusion section (e.g., The Croppie – the PhotoCropping app project and the Rapid, Low-Cost Aflatoxin Detection using AI project).

In the 2016 Proposal, it was envisioned that big data and ICT solutions would contribute to CGIAR's gender Intermediate Development Outcomes (IDOs). The Platform was expected to take leadership in bridging the gap of the technology world that is male-dominated. Gender-sensitive approaches were identified as an emphasized element of the Platform's capacity development strategy. Datasets that were made open access through ORGANIZE activities and their metadata were expected to represent the gender dimension with gender-disaggregated data wherever applicable. This was to be achieved in collaboration with the CGIAR Community of Practice on Gender and Working Groups (WGs).

3: Capacity development activities have been steadily integrated into the Platform modules, guided by the CGIAR Capacity Development (CapDev) Framework.

CapDev activities have been a strong part of the effort towards FAIR and OA improvement for the Centers via training opportunities and data sprints. The 2016 Proposal stated that all nine elements of the CGIAR CapDev Framework[14] will be addressed. The CONVENE module was to emphasize capacity development by emphasizing gender-sensitive approaches, institutional strengthening, and organizational development. Interviews revealed that external partners benefitted from discussions on open data and open access via training programs organized by the Platform. For example, in 2019, Inspire Challenge applicants received feedback on links to digital research and impact with 10 finalists receiving coaching on capacity building from an external expert panel.

4: Youth engagement was planned in the design of the Platform but during implementation, not enough attention was given to building specific and continuous mechanisms to reach the foreseen outcomes and to include youth in decision-making (i.e., Platform steering committee, Inspire Challenge selection committee).

CGIAR does not have a common framework for youth engagement. The Platform – as CRPs – had to create its own path to engage youth. Several youth engagement activities have been realized mainly to increase participation in annual conventions and capacity building on social media. The evaluation team's access to this group of stakeholders was very limited (The Platform provided a list of potential informants to assist the evaluation but only one potential informant was under 24). Youth-directed activities did not benefit from a clear vision nor were there relevant activities systematically designed or implemented under the three modules.

At the start, the 2016 proposal envisaged big data and ICT solutions to contribute to CGIAR's youth IDOs. It was anticipated that increased access to agricultural data and ICT-based applications could help retain and attract more rural youth in agriculture. Young men and women were to be considered a major target user group for the annual data/knowledge consultations and design of communication materials. To bring new ideas to the table from young people, significant youth involvement from within CGIAR and partner organizations was envisaged both at the annual conventions and in CoPs. Some Inspire challenge project topics were expected to specifically address youth-related impacts. Impact assessment efforts were also expected to look at the youth-related impacts of different innovations. Addressing youth as a cross-cutting theme of the Platform was allocated an annual budget of USD 350,000.

---

[14] The nine elements of the CGIAR CapDev Framework are: 1. Capacity needs assessment and intervention strategy design 2. Design and delivery of innovative learning materials and approaches 3. Develop CRPs and Centers' partnering capacities 4. Developing future research leaders through fellowships 5. Gender-sensitive approaches throughout capacity development 6. Institutional strengthening 7. Monitoring and evaluation (M&E) of capacity development 8. Organizational development 9. Research on capacity development

5: Climate change was considered in the Platform design and tools created were appropriate with the expected outcomes.

Interviews confirmed coverage of climate change via the Community of Practice. Tools created under ORGANIZE (CGlabs) to include climate data as part of bundled services were shown to facilitate the development of climate scenarios in sub-Saharan Africa and Vietnam. The tools were also used to provide a critical context for informing agroecological interventions based on assessments of returns to fertilizer usage. In the 2016 Proposal, climate change/variability was enumerated among the pressing challenges of our time for which data were helping to accelerate the development of robust responses. The Platform was envisaged to contribute to improved climate change forecasting. Ensuring standards of compliance on climate change was envisaged as a linkage needed under CONVENE. This was done particularly to enhance the discoverability and interoperability of the data in existing CGIAR efforts.

6: The Platform demonstrated the capacity to adapt to evolving environments and constraints (GARDIAN syntactic interoperability, deployment of COVID 19 Rapid Response grants).

At the design level and focusing on syntactic interoperability, ORGANIZE allows a high level of flexibility (i.e., protocols and standards to query and format the answers known in advance). Usually, specifications using current technologies can easily be transferred to another technology; if well specified, data and metadata models can be upgraded with minimal effort. The back-end technology is less flexible but has no direct interference with this interoperability contract. Efforts linked to model specification, ontology development, algorithms for metadata enrichment can endure, being adaptable to new technologies and being typically transferrable to new system architectures. Interoperability enables seamless integration of old and new standards. So altogether, the design proposed by ORGANIZE was flexible across disciplines and scientific questions. Nonetheless, each Center has its own local scientific specificity, which was not considered in the GARDIAN interface, altogether resulting in available data not being relevant as reported in one interview.

In response to the issues brought about by the COVID-19 pandemic, the Platform awarded US$100,000 in Rapid Response grants to big-data-enabled projects working to tackle food system challenges. The Platform's Rapid Response Grant application process was made available to current or previous Inspire Challenge winners to build on their learnings as pilot or scale-up projects and leverage their innovative, data-driven designs to respond to the situation with agility. The selection allowed previously funded projects to pitch ideas on how they would use their digital innovations for COVID-19 response, recovery, and resilience. Additional funding will help three innovative projects[15] roll out digitally-enabled, data-driven responses to food security challenges presented by the COVID-19 crisis.

7: Much work remains to enhance the semantic, syntactic, and structural interoperability, which brings great value for predictive analytics and data reuse.[16]

The 2021 CGIAR Synthesis of Learning from a Decade of CGIAR Research Programs emphasized the need to enhance the MELIA pathways of CGIAR research. Notably, boundary partners need to be engaged, and Quality of Science reviews are required for specific CRPs to enhance policy attribution (see Runzel et.al, 2021). The synthesis also recommended that AR4D adopt a landscape focus by engaging with innovative financing mechanisms. Interviews with CGIAR staff pointed to the potential of big data to improve predictive analytics capacity within CGIAR and to give new insights to data that may not be possible using current analytic technology. Whether it was an application to inform decisions on when and to what

---

[15] The three Rapid Response winning projects were: (1) Eyes on the ground for agricultural microcredit. IFPRI, A 2017 winner & 2018 scale up winner project (2) Gamifying weather forecasting: "Let it rain" campaign. CIAT. A 2019 winner project and, (3) Herd opportunity. ILRI. A 2017 winner & scale up runner 2018 project.
[16] The different levels of interoperability enabling the dialogue machine to machine (or system to system) are syntactic interoperability (e.g. format, encoding, protocol and standard) ensuring communication and information exchange, structural interoperability (e.g. resolution, spatial accuracy, temporal accuracy, other structural quality, orthorectification) ensuring expected intrinsic characteristic of the data and semantic interoperability (e.g. classification, measurement attributes, quality of the data, ontology of the data, reference of the information model) ensuring understanding of the meaning of the data and of its properties, i.e., enabling to grasp the knowledge behind the information shared. To this end and as part of the semantic interoperability the quality of the metadata is also paramount, e.g., FAIR principles. Leibovici *et al* (2021)

extent to apply crop nutrients or matching crop selection to food and nutritional needs of a particular geography, big data has an important role to play in enhancing the relevance of CGIAR research. By focusing on technology, the Platform was able to make progress on the Findability and Access dimensions of FAIR, but much effort is still needed for data Interoperability and improving the capacity for predictive analytics and data Reuse. This limits the flexibility of the Platform to respond to local and evolving conditions such as changes in crop prices or consumption patterns and dietary patterns.

The Platform team noted the absence of enforceable organizational governance at a CGIAR level related to data and building unified analytic pipelines. As a result, the only tools at the Platform's disposal to drive the 'I' and 'R' in FAIR were technical collaboration and internal coalition-building and helping develop the relevant ontologies where there were gaps (e.g., working with WorldFish scientists on Small-scale Fisheries and Aquaculture Ontology). Developing prototypes and pipelines also helped to demonstrate the power of this interoperability including AgroFIMS[17].

4.2 Efficiency

8: The Platform sought to leverage CGIAR "convening power," The management team was the chief decision-making body, complemented by mechanisms for technical collaboration such as open technical Communities of Practice (COPs) and an innovation process to move the agenda of (big) data management and analytics in agriculture research forward, yet this was not enough to foster Centers' engagement with the Platform's outputs.

Interviewees' opinions about the Platform's governance were mostly negative with mixed opinions about the management performance. Consulted stakeholders emphasized the ambiguous role of the steering committees (SC), lack of knowledge of its composition and its responsibilities (e.g., if the SC members were entitled to take strategic decisions or only keep informed and endorse management team decisions). Likewise, critics pointed out the lack of transparency on how priorities were managed and the rationale for budget allocation. Nevertheless, budget allocation and spending hewed fairly closely to the 2016 Proposal, deviations were approved by the SC as documented in SC minutes.

Despite their positive opinion about the pertinence of GARDIAN, Center data managers felt excluded from the process. Data managers were uncomfortable with what they considered a top-down approach to shaping the development of GARDIAN vis-à-vis the envisaged agile[18] approach meant to benefit from iterations between delivery and feedback. Several interviews reported a lack of continuing funding under ORGANIZE. The budget and reporting show a large allocation of spending in the first two years, The Platform team reported that in 2017 and 2018 about 50% of the ORGANIZE module budget was disbursed to Centers to facilitate the incorporation of FAIR data management into center data management plans and building further support and sustainability through project formulation and management going forward. With regards to data curation, it commonly does not attract funders as per subject matter experts' opinions, so the time spent on data curation (underfunded in projects) had to be considered during the life of the Big Data Platform and for future data curation.

CoPs did not have a formal structure or membership charter, only Terms of Reference and an annual process of developing workplans for validation at annual conventions; yet CoP coordinators had full control of decision-making around the CoP workplan and budget. This governance model (or lack thereof) presents the risk that successes and failures are in the hands of one person. This raises questions about accountability and sustainability. In the 2016 Proposal, the needs assessment conducted through a consultation reported that "*When asked about governance, stakeholder opinions were varied, with a majority preferring either a self-governed, democratic approach or a rotating governance body.*

---

[17] AgroFIMS is an in-field data collection tools that farmers can use and is based on AgroO (ontology) enrichment of the data records.
[18] Agile practices include requirements discovery and solutions improvement through the collaborative effort of self-organizing and cross-functional teams with their customer(s)/end user(s), adaptive planning, evolutionary development, early delivery, continual improvement, and flexible responses to changes in requirements, capacity, and understanding of the problems to be solved.

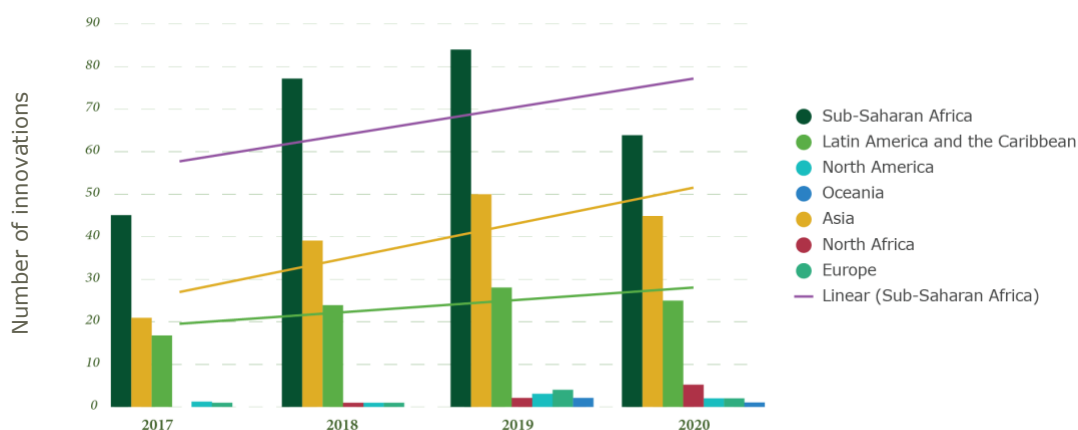*Stakeholder opinions were split whether leadership should be organized by modules or thematic task forces."*

Thirteen million USD was budgeted for ORGANIZE (Module 1). IFPRI required 8 million USD[19] as the module lead and the co-leading center of the Platform, and each of the other Centers received between 250k to 400k. The Platform team mentioned that 3.5 to 4 million USD was disbursed to Centers, nevertheless, IFPRI had one of the lowest OA% for publications as of August 2021, 6.2%. One interviewee explained this low level by the fact that IFPRI's work is about socio-economic domains involving human science fields and issues such as data privacy were slowing the process of opening the data, but the OA rate for datasets was 89%.

> *"(...) it is also important to ensure that money meant for initiatives such as Big Data Platform are not monopolized by a handful of CGIAR Centers. In some cases when individual centers do not allocate their resources on such initiatives, they tend to lose qualified staff". Excerpt from an interviewee's testimony*

There were promising activities under the framework of CoPs, *e.g.,* geospatial and ontologies. However, there was insufficient evidence that the CoP's work was incorporated in the capabilities of GARDIAN at the metadata level in relation to specificities of CGIAR data. For example, the metadata instances queried via GARDIAN did not show geospatial and temporal coverage as present in Dublin Core[20] and the CG Core[21] metadata schema despite such information existing in the Dataverse version of the metadata. The metadata cataloging was useful in federating the Centers catalogs, including extracting geographical names and license information from the original metadata. It was however reported in several interviews to result in the lost richness of metadata information. Some interviewees also questioned the sustainability of the CG Core beyond the life of the Platform.

Regarding the Inspire Challenge competition, there was an unequal distribution of project grants between winning Centers - 50% of the grants went to projects managed by Centers with Headquarters located in Latin America (CIAT (10) and CIMMYT (3). The fact that CIAT has almost 50% of the Inspire Challenge grants increased the feeling of unfairness among other Centers consulted. In addition, the March 2021 INSPIRE module review conducted by Mockshell et al. (2021) identified regional trends in submissions to the Inspire Challenge (see Figure 4) that highlighted areas that are digital innovation hotspots and others that are digital innovation deserts, with unequal coverage across regions.

Figure 4. The number of applications from different continents to the 2017, 2018, 2019, and 2020 Inspire Challenge (Mockshell et al., 2021).

---

[19] Table page 36 of the Big Data Coordination Platform: Full Proposal 2017-2020
[20] The Dublin Core™ Metadata Initiative supports innovation in metadata design and best practices. DCMI is supported by its members and is a project of ASIS&T.
21 A set of metadata elements used by CGIAR Research Center and CRP repositories, in order to facilitate cross-repository searching and enhance discovery of CGIAR information products through Open Access.

A few interviewed stakeholders considered that the selection process favored already existing partnerships and that it was more difficult for new projects starting new collaborations. In one case, someone reported that they did not partner with an existing (internal to CGIAR) similar initiative because they feared their project will not be considered innovative if there is already a project doing similar things. According to a few people, this created the feeling of *"why waste time writing a proposal when there is little chance to win"*. In contrast, the Platform team explained that they were looking more at what they were doing and how, rather than at how long, the partnership was *"We weren't seeing how much the length of the partnerships was, although probably the length was reflecting into the quality of what and how since a relationship was better established. Maybe there was a correlation there."*
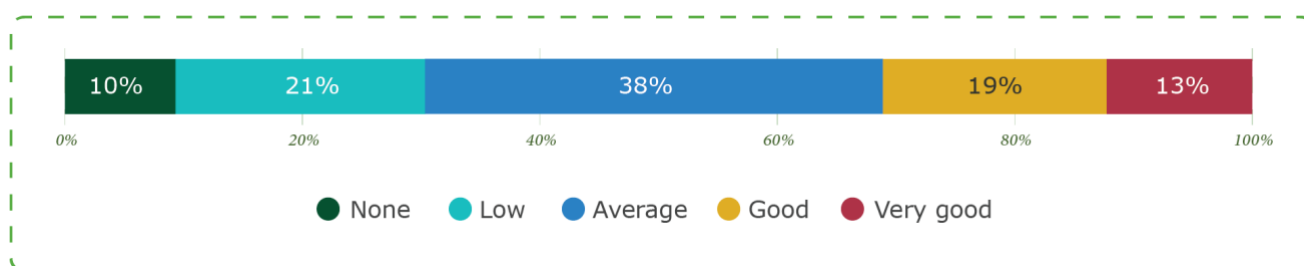
Moreover, for some stakeholders consulted (participants and non-participants in the Inspire Challenge) it was not clear how the governance of the Inspire Challenge operated or if the winning projects had continuous follow-up including contact/coordination between the Inspire winners and the Platform. A few projects mentioned also that there was a delay in the disbursement of funding which resulted in an adjustment of the implementation calendar.

9: Lack of knowledge about the Platform's mandate among internal CGIAR stakeholders affected the quality of engagement of stakeholder groups with its modules' activities and outputs.

Most of the Platform's stakeholders interviewed revealed being unfamiliar with its mandate, and so did not have specific expectations from their engagement. This finding was confirmed by the online survey where 60% of respondents qualified their knowledge level of its mandate as either low or average (Figure 5). Some believed that using the "Big Data" buzzword in the title was misleading and added extra confusion that also affected engagement. Notably, however, big data, in the sense of volume and velocity, is relevant to CGIAR's research and some Centers may use or even produce big data that meets all four criteria[22].

Platform collaboration and linkages with the private sector raised criticisms among some interviewees (mostly scientists). On their part, the critics held the view that public tax-payer money should not be spent on supporting private sector initiatives but on building the capacity of Centers to engage more effectively with opportunities offered by big data analytics. On its part, the Platform maintained that it positioned CGIAR as an enabler by demonstrating that public interest actors like CGIAR have a neutral position in the overall digital ecosystem and that developing partnerships with the private sector is an explicit feature of the 2016 Proposal. To this extent, the Platform was deemed to have convening power crossing the public, private, and non-profit divide in the view that *"we need the whole system in the room if we wish to change it,"* as per the testimony of a member of the Platform management team.

Figure 5. Self-assessment of the level of knowledge of the Platform's mandate (online Survey; n=102)

The Platform contributed to increased awareness of the importance of data sharing, open access, and FAIR principles as per the testimonies of most end-users and data managers. However, the institutional engagement of Centers was variable. The differential engagement was explained by the lack of Center staff time due to funding scarcity. Indeed, the Platform did not have a sufficiently developed and implemented communication and dissemination strategy at the start. The 2016 Proposal mentioned

---

[22] Volume, variety, velocity and veracity

Hackathons and Killer app competitions, but they do not seem to have been fully implemented. Dissemination via training to end-users and stakeholders, webinars, and the conventions were nonetheless successful and well appraised. Although communication and dissemination activities were reported in the annual reports there was insufficient evidence that such activities effectively promoted the uptake of outputs such as GARDIAN, CG Labs, AgroFIMS, and the ontologies developed by the CoPs.

10: Stronger linkages among the three modules would have increased the efficiency of the Platform to advance the use of big data analytics.

In line with the review of CGIAR's OAOD Policy and the 2021 Synthesis of Learning, we can say that research on "scaling and implementation science" can vastly improve the efficiency and timeliness of resources devoted to enhancing big data analytics within the CGIAR. Interviews with CGIAR staff highlighted the fact that there was enormous scope to explore the use of methodological innovations such as the use of typologies and composite indices on a large scale within the CGIAR by leveraging big data resources. The emergence of this thinking can be attributed to the discussions as part of the CoPs and WGs of the CONVENE module.

Over half of CGIAR interviewed scientists agreed that beyond an emphasis of *"moving or uploading data"*, the Platform should focus on establishing the computing environment (via widgets/interfaces) that would make the creation of tailor-made databases possible. The Platform acknowledges that they did not purposely require data by research domain or theme but welcomed anything Centers wanted to upload so long as it was well-described with a focus on semantics. The rationale was that diversity of data was suitable for developing the ability to run complex queries across diverse data in the future.

The approach of combining and devoting resources towards the development of a CGIAR specific analytic workbench with some common tools (e.g., R, Jupyter notebooks) was in line with the requirements for pursuing the construction of a CGIAR-wide database. However, the value of a platform for big data is the establishment of standard operating procedures that can support the co-curation of tailor-made databases to support combining databases via linked Application Programming Interface (API) protocols to support research on agri-food systems based on interoperable datasets. The examination of a host of technical and management options emerging from CGIAR research with the potential to break down siloes between CRPs, Centers, and disciplines of individual researchers can be pursued by future initiatives.

11: Related activities on cross-cutting themes (i.e., gender, youth, climate change, capacity development) have been implemented but results are still preliminary and need to be properly monitored and evaluated particularly for CONVENE and ORGANIZE outputs.

The budget was allocated to Gender, Youth, and Capacity Development (to implement OAOD) themes. However, a deliberate plan of how such investments (e.g., at the Convention or in the selection of Inspire Challenge awardees) may advance FAIR principles in AR4D was missing. There is a need to properly monitor and evaluate particularly the research on developing big data-enabled methods and the examination of inclusivity in the design of digital initiatives with the GENDER platform under CONVENE, and the work on gender data annotation and disaggregation under ORGANIZE.

Interviews revealed that many Centers have data from previous studies on gender and yet they are not being utilized in the design of new studies. As a result, "interview fatigue" occurs when respondents are visited repeatedly to ask similar questions that were covered by previous CGIAR gender studies. It is encouraging to note, however, that under CONVENE, the Platform decided that co-designed research with the Gender Platform on 'big data' methodologies would likely yield results that leveraged the rigor of gender specialists with new forms of analytics. This yielded promising results in a study aimed to predict certain aspects of small-scale women farmers' economic empowerment in Uganda. The Gender Platform and the Big Data Platform have a follow-up project underway presently to further develop these methods. In addition, the CoP on socio-economic data (SED-CoP), in collaboration with the ontology CoP, developed a standardized set of 100 farm household survey questions to improve dataset comparability over space and time.

Regarding climate change, compliance with standards in relation to data integration was planned. Effectively, a range of climate data[23] is part of external data accessed via the Platform (Data Exploration tool). However, the interface is not sufficiently describing how the data can be made available outside the Data Exploration tool.
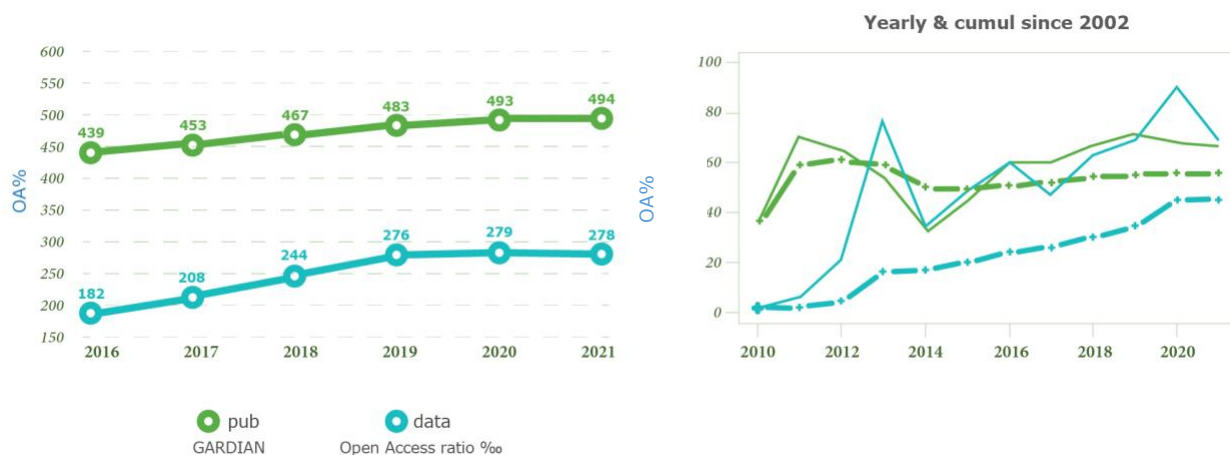
The Inspire winning projects had an elaborated declaration on how to integrate cross-cutting themes evidenced in the related website information. The resources allocated by the Platform to the work on each project were a fixed sum. However, cross-cutting themes were not particularly highlighted in the design and reporting of activities and products under ORGANIZE and CONVENE (e.g., gender dimensions). As it stands, the lack of advanced semantics in GARDIAN precludes such focus.

4.3 Effectiveness

12: In CGIAR Centers, the evaluation indicated an increased awareness of Open Access/Open Data (OAOD) including Findable, Accessible, Interoperable, Reusable (FAIR) principles. The Platform contributed to this outcome by constant advocacy during its annual conventions and capacity development activities. However, evidence shows that relatively more effort is still needed to progress data Interoperability and Reusability.

The GARDIAN Annual Reports for 2018 to 2020 showed constant and impressive progression of the number of available publications and datasets *via* the Platform. However, as shown in Figure 6 (left), the % OA did not increase substantially since the launch of the Platform, +5.5%, +9.6% respectively for publications and datasets between 2016 and 2021, but this has a lot to do with older datasets not being OA. A cumulative plot for published materials since 2002 showed a higher increase (Figure 6 right): +8% and +19% respectively between 2016 and 2021. The yearly rate of OA showed a more substantial trend on Figure 6 (right – single lines), than on the cumulative data: around 45% of datasets (45% for publications) before the launch of the platform and above 75% (70% for publications) after 2019 are OA. So, overall, for both datasets and publications, there was around a +30% increase in making research outputs open access since the Big Data initiative was launched.

Figure 6. GARDIAN Open Access changes (left- cumulated statistics up the given year, right-cumulative from 2002 onwards for the dotted lines with (+) and yearly statistics for single lines); data for 2021 is up to August GARDIAN Open Access change

Despite these trends, FAIR scores have been improving over the years of published data, showing an impact after 2017. FAIR scoring is not reflecting the OA aspect well and vice versa. The median score for Accessible was 3.5 in 2017, reaching the top score of 5 after 2017, but the range is not improving. Q1

---

[23] From coupled model intercomparison project 5[CMIP5] and coupled model intercomparison project phase 6[CMIP6] initiatives with a range of models and projection scenarios.

stays at 2.5, besides 2020 showing only a few outliers, so relative stability of this score but with the percentage of OA improving. This difference may be due to more publication with restricted access, or missing licensing information as can be inferred from the scoring description (FAIR guidelines Document). Importantly, individual researchers were not required to make their data FAIR as reported by the Platform team. As a result, there is a missed opportunity for CGIAR to take full advantage of its data. The Platform had the mandate to work on this but not to require it.

Regarding the use of the GARDIAN in finding relevant datasets and useful metadata (survey results, Figure 7), only 24,5% of the survey participants (n =106) answered the questions with 50% giving positive or very positive views (12% of the whole sample). Data-driven innovation which can be notable via CoP's activities or Inspire challenges was not demonstrated nor sufficiently enabled from the GARDIAN platform and no evidence of innovations influence to enhance and feedback the GARDIAN facilities and developments was found. Collaborative GARDIAN Labs (CG Labs) – a data analytic computing environment with a cloud-based shared data storage facility – was tailored from the existing GLOBUS[24] service and released in 2019. CG Labs had 184 members (as of Sept 2021) but the assessment of its effectiveness is impeded by the limited documentation available.

Figure 7. Respondents' perception of easiness to find, access, interoperate and reuse data through GARDIAN (Online survey; n=26)

The Annual Convention was billed as a high-profile event "designed to be the 'go-to' meeting for any individual or institution interested in engaging on big-data in agricultural development". Another objective was to establish Communities of Practice (CoP) and working groups around key topics with clear Terms of Reference (ToRs). One task identified was for CoPs to establish new collaborative research efforts and proposals that would receive bilateral funding. We can report that 45 webinars were organized in 2020 of which approximately 5,673 participated. Not enough evidence was found about follow-up activities like joint publications or proposal writing that can be attributed directly to the webinars.

13: Weak evidence was found about GARDIAN usage from end-users. Nonetheless, users are spending more time browsing the results of their queries. Not enough evidence was found about the usage of the Platform data analytic facility (CG Labs) to make an assessment.

---

[24] GLOBUS a facility to support data sharing and collaborations for the development of applicative solutions https://www.globus.org/platform/services "

Centers initially had their own metadata schemas, but the push for one schema that had begun during the OAOD initiative continued through the Platform activities, resulting in a released version of the CG Core metadata schema. Based on the Platform team testimony, most Centers have now implemented the CG Core, but political issues and lack of accountability to the Platform mean that making this a requirement was not possible - and this affected uniform uptake of other products. According to the Platform team, GARDIAN tried to plug this gap, demonstrating the value of harmonized metadata drawing data from multiple Centers (Bonilla-Cedrez *et al.*, 2021[25]).

Although the number of users increased over the short, monitored period (March 2020 to August 2021) based on available Google Analytics, a plateau was reached, followed by a regular decrease. Nonetheless, over this period a relative increase in browsing publications metadata suggests an increase of usefulness as users are spending more time browsing results of their queries. There is not enough evidence to make a conclusion about the usage of the platform data analytic facility. The GARDIAN website generated 28,993 unique pageviews over the period March 2020-Sept 2021. As a comparison, web traffic generated by GARDIAN is around one-tenth fewer (17,705 unique pageviews up to March 2021) than the Platform website (169,150 unique pageviews in 2020). The end-user survey confirms this relatively low usage of GARDIAN (from users of GARDIAN) with 60% of those who answered using GARDIAN once every six months or less (over 25 out of the 106 samples), the 40% remaining answered that they used GARDIAN at least once a month.

A comparison between 2020 and 2021 user behavior was analyzed to see the evolution of the data available. An increase of around 11% in the number of page views (21 % in unique page views) between the 2020 and 2021 periods was noted, but it was not a steady increase, as it had a very high peak on the 8th of July 2021, 10 times the daily average. All pages decreased the proportion of the number of views, search (-6%), Dataset (-1%), Exploration (-6%), and Analytics (-2%) to the benefit of publication views (+15%) and FAIR metrics (+1%). Thus, users spent more of their visits browsing the publication results including understanding the FAIR scoring than trying another search. The interest in using the Data Exploration tool diminished.

From March 2020 to September 2021, the traffic of the 10,146 users, was recorded as coming at about 50% from the Google search engine and another 35% as direct links used, *i.e.,* from a pdf file or typing/directly pasting the URL. The provenance of traffic from bigdata.cgiar.com to GARDIAN is not referred to or is unknown. Between 5 to 13 unique users visited GARDIAN daily with a period of increasing period usage to the beginning of 2021 with up to 1000 users every month and a decrease from May 2021. On average, interest appears to have grown from about 250 monthly users in April 2020 to 950 around December until May 2021 and then decreased to 350 in August 2021.

A discrepancy exists between datasets viewed from the GARDIAN and the Center's repositories. Besides the loss of richness observed (discussed earlier), the harvest process has not yet reached full coverage of the existing local metadata. The GARDIAN is harvested weekly, so, understandably, there may temporarily be some differences.

Interviews revealed that data managers at the Centers were depicted to have preferred greater investment in their existing systems than in GARDIAN, which appeared to duplicate existing repositories such as Dataverse for datasets and CGSpace for publications. Moreover, GARDIAN was deemed to inadequately represent the effort captured in center-based repositories and to suffer double reporting of resources. Its focus on data was also deemed diluted by its additional effort on publications. Effort spent on the CG Core metadata schema was depicted by some interviewees as over-emphasized vis-à-vis other improvements to data repositories.

14: The Platform contributed to increased engagement (frequency and depth of interactions) between CGIAR researchers and stakeholders from the digital ecosystem. It resulted in standardization efforts: development and use of terms for data comparisons and reuse (e.g., Ontologies CoP, and the Information and Data Managers CoP), digital extension (e.g., the Data-Driven Agronomy CoP), and modeling (Crop Modelling CoP). The consulted stakeholders considered the Ontologies CoP as instrumental in persuading CGIAR colleagues to pay attention to ontologies, not as mere academic standards but as mechanisms for unlocking the potential of big data through semantic interoperability).

---

[25] https://www.nature.com/articles/s43016-021-00370-1?proof=tv

The overall level of participation in CoP activities also shows the commitments and engagement with the digital community outside CGIAR, 82% of the 1644 members (88% as registered in the Data-Driven Agronomy CoP, 86% in Geospatial, 89% in Crop, 80% in Socio and 82% in Ontology CoPs). This is an open initiative and could potentially have an impact on CGIAR outreach, but it was not possible to assess the level of CGIAR leadership from the reporting of CoP activities.

The webinars attracted large numbers of participants (*i.e.,* 45 webinars in 2020 with 5,673 attendees). A smaller number of webinars occurred in 2021 (540 attendees for 9 webinars), so an average attendance of 60 versus 128 in 2020. Convention attendance from outside CGIAR increased every year from 57% in 2018 to 75% in 2020 with conventions characterized as attracting enthusiasm, engagement, creating common goals, and togetherness, and more so when viewed in combination with CoP activities.

Several Inspire projects enabled CGIAR's engagement with the wider agriculture data and innovation ecosystem. As an example, the 2019 Scale-up winner **PlantVillage Nuru** developed a phone App for cassava pest and disease monitoring that uses AI. The data used built on 200,000 annotated cassava plant images. This is a successful project that reached more than 200,000 farmers and continues to build significant new capabilities for transnational pest and disease surveillance for CGIAR and partners in collaboration with the MARPLE project, another Inspire Challenge awardee. The App can be a showcase for the Inspire Module as it is available both on Google Play and the Apple Store as a free download. On the other hand, this project can also be an example of the disconnection between the Inspire Module and the Platform governance and dissemination plan. The CGIAR Platform's Inspire Challenge is not mentioned on the information provided at Apple Store or Google Play from where the **PlantVillage Nuru** application by David Hughes can be downloaded.

Box 1. CoP case study: The Ontologies Community of Practice

Promoting data annotation for semantic interoperability - the Ontologies Community of Practice

Among the Platform's strategies to enable interactions and collaborations for scientists within and outside CGIAR was to establish or work with existing Communities of Practice (CoPs) and Working Groups (WGs). The Ontologies CoP adopted and enhanced the pre-existing Crop Ontology Community Project. It aimed to create best practices, recommended ontologies and guidelines in the selection, use and application of semantics for data harmonization. The CoP also sought to stimulate knowledge sharing on ontologies. The CoP is led by the Alliance of Bioversity International and CIAT and has four working groups namely: (a) *Plant Phenotypes Ontologies* (b) *Fish Ontology* (c) *Agronomy Ontology*, and (d) *Socio-Economic Ontology*. Its membership is through a LinkedIn group. The CoP's governance is not yet formalised , but governance frameworks for its individual ontology products are being explored.

The CoP delivered 17 webinars on topics related to ontologies and semantic interoperability, each of which attracted over 40 live attendees. The attendance and viewership, diversity of speakers, novelty of the content, and interactivity of sessions suggest efficient use of resources and being fruitful in fostering capacity development among stakeholders. The CoP published a descriptor article on its workings and results in the *Patterns* journal in 2020 as well as contributed to the Platform's paper on the One CGIAR digital strategy. **The CoP was deemed instrumental in persuading CGIAR colleagues to pay attention to ontologies, not as mere academic standards but as mechanisms for unlocking the potential of big data through semantic interoperability.** Thanks to the Platform's support, the Crop Ontology (CO) grew to have over 10,000 harmonized and validated variables covering 33 crops. The CoP also advanced new ontologies such as Agronomics Ontology (AgrO), Socio-economic Ontology (SEONT), and Fish Ontology (FishO). The evaluation team found no evidence to ascertain the extent of adoption of the ontologies led by the CoP. Frustration among key informants during interviews was noted about a slow uptake of ontologies in spite of the visible investments in their development. The CoP reported to having stimulated the interactions with several industries (Bayer, BASF, KWS) who shared their practices for using ontologies and knowledge graphs to manage internally their crop data.

The ontologies CoP registered steady growth of membership to its LinkedIn group and subscriptions to its newsletter. A steady growth in the CoP's organizational partnerships was observed from 10 in

2018, to 18 in 2020. The richness, diversity and value of membership and partnerships in the CoP were appreciated by the key informants. Formalizing the CoP's governance was considered an imperative for strengthening its partnership framework. The work of the CoP was appreciated by key informants to the extent of appealing to sustain it in the new One CGIAR. Proposals for new initiatives under One CGIAR commit to comply with the Open and FAIR data assets (OFDA) policy by annotating data with widely adopted ontologies. Although this suggests sustained relevance, sustainability of the Ontologies CoP is yet to be demonstrated in One CGIAR. Moreover, ontology objectives about data discovery and knowledge modelling were considered fully understood and adopted by the Platform, yet not adequately understood by the rest of CGIAR.

The Ontologies CoP has been a key mechanism for advancing semantic interoperability through data annotation thanks to the Platform's support. Although the CoP has advanced the development and maintenance of ontologies, inability to demonstrate the extent of their uptake is a concern for return on investment. The CoP efficiently harnessed and shared expertise on ontologies among stakeholders within and outside CGIAR. There is no clear indication on where the important role served by the CoP will be housed in the new One CGIAR.

More demonstrable effort is recommended to entrench within One CGIAR the capacity to develop and maintain ontologies for agricultural subdomains. This is more so for CGIAR's elevated position as a go-to institution for thought leadership on the science of data, advancing semantic interoperability across its diverse and ever-growing agricultural research datasets. It is crucial to institute mechanisms for tracking usage and aggressively increasing usage of CGIAR-led ontologies. This will allow refinement opportunities to unlock return on investment for CGIAR and its funders, and to reduce deployment delays in tandem with a rapidly evolving digital technology space.
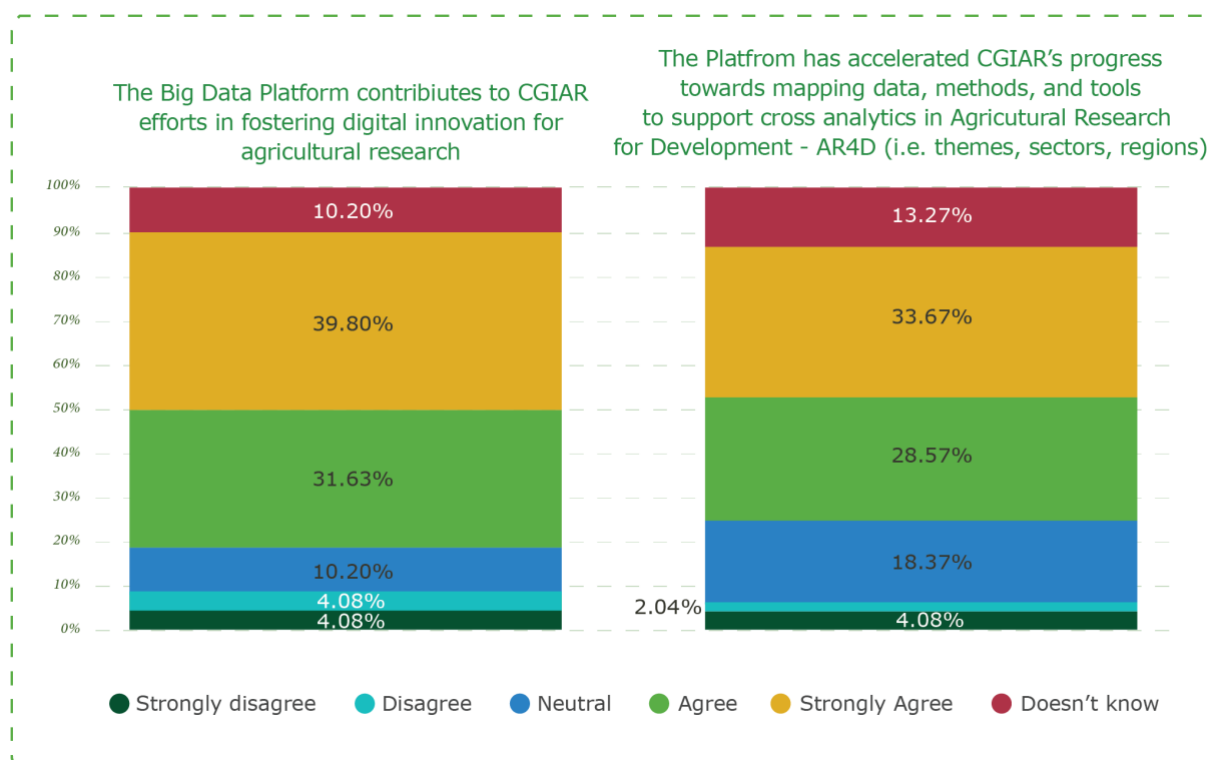
For more details see Annex 8: CoP case study

15: Platform outputs added value to CGIAR efforts to map data, methods, and tools to support the delivery of research.

The GARDIAN requirements and CoP activities, particularly on ontologies, contributed to improving quality in the delivery of research. Nevertheless, the quality assurance (Strategy #1 in the Proposal[26]) was not sufficiently addressed[27]. More than 71% of surveyed persons (with 98/106 answering) felt that CGIAR's Big Data Platform was fostering innovation in agriculture research and 62% (agree and strongly agree) thought the platform has accelerated progress in data analytics in AR4D (Figure 8). Of the two questions asked (Figure 8), 40% and 34% strongly agreed, respectively. The external evaluation interviews captured appreciation among stakeholders of the efforts in developing and maintaining products such as the Crop Ontology and the Agronomy Ontology by the Ontologies CoP for promoting semantic interoperability. Frustration was expressed in the interviews about the slow uptake of the ontologies despite the visible investments in their development.

---

[26] "*Strategy #1: Provide data quality training and develop quality control protocols*"
[27] Quality assurance of the metadata creation but indirectly reflecting on the quality assurance of the data, e.g.., protocol on assessing the quality of the data and details on the description of the protocol to acquire the data.

Figure 8. Opinion on Platform's ability to foster innovation in agriculture research and accelerate progress in data analytics in AR4D (Online survey; n=98)

In 2019, with two university partners and Rothamsted Research, the ORGANIZE module released AgroFIMS for in-field data collection using a mobile device. The generated metadata associated with the data follows the harmonized description including the AgrO Ontology, ready for the GARDIAN metadata catalog. However, the Manual v2 released in 2020 says "*By mid-2021 you will be able to easily upload this collected data through AgroFIMS to a Dublin Core or CG Core compliant Dataverse repository*" (127 downloads including v1). Thus, the contribution of ORGANIZE to this innovation is not yet achieved.

The **Crop Modelling CoP** reports on collaboration involving authors drawn from different organizations on three publications, it is difficult however to exclusively attribute this collaboration to the activities of the CONVENE module.

16: Proposals selected under the Inspire Challenge helped to catalyze the development of new digital methods for research or delivery of research at CGIAR, meeting one of the Platform's objectives. Yet, it could have implemented a more robust monitoring and evaluation (M&E) system to harvest the outcomes of these projects.

The description of granted projects indicates that the Inspire project proposals fulfilled most of the listed requirements. What is not monitored is how the results of the different projects were linked to the CRPs to enhance the technologies uptake by other researchers. There was also no clear strategy for the Inspire Challenge winners on how to showcase CGIAR-produced data or the projects and how to integrate the produced data into GARDIAN. The Platform team reported that they limited the monitoring to quarterly technical and financial reports, one synthesis learning report regarding the Challenge process and design[28], and one cross-cutting internal evaluation of its effectiveness in building innovation ecosystems[29], and had left evaluation of individual projects to the host CRPs to minimize double-reporting. In sum, the Inspire Challenge awarded 28 grants to 21 projects with a total of USD 3.225M. Additional scale-up funds were awarded to three of the five 2017 winners in 2018, and in 2019, four

---

[28] https://cgspace.cgiar.org/handle/10568/99282
[29] https://cgspace.cgiar.org/handle/10568/113597

winners from both 2017 and 2018 were also awarded. The total amount of scale-up funds granted to date is USD1.125 million.
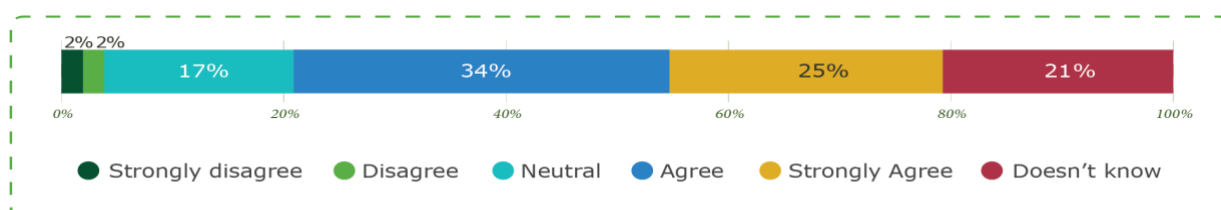
One researcher quoted:

> *"Personally, I think they triggered some really nice collaborations. The challenge system with relatively very small money started some small startup projects. Brought to us some very smart people. That collaboration continues and a number of projects, mining social media, signals of concerns, they brought to us great smart computer scientists, combination went very well."*

Another added:

> *"Additionally, the brand association with CGIAR was appreciated as valuable, providing social validation and a sense of credibility to the project's operations on the ground. The link to CGIAR was associated with an increased reach for the project and was beneficial for the project to attract new partners."*

59 % of surveyed persons (with 53/106 answering) think that the Inspire Challenge grant process made significant contributions to digital innovations for AR4D (Figure 9). The main idea of the Inspire Challenge was to demonstrate the power of using big data analytics by Centers through inspiring and innovative projects. In the 2016 Proposal, the Inspire Challenge projects were selected to (1) Bridge the deep subject matter expertise of CGIAR researchers with the capabilities of external partners; and (2) Challenge research organizations to partner with industry to leverage public good data. According to the information available on the Platform website the proposals were selected by a robust, established application and review process that evaluated the proposals on the following criteria: Scale, Innovation, Impact, Pitch Quality, Data Use, and Sustainability. In theory, this process should have encouraged the CRPs supporting the Inspire proposals to define the needs connected to the goals of CGIAR that could catalyze the usage of new digital methods.

Figure 9. Opinions on whether "The Inspire Challenge grant process made significant contributions to digital innovations for agricultural research and development" (Online survey; n=53)

17: The Platform prepared CGIAR for compliance with the General Data Protection Regulation (GDPR), it developed a tool to help data managers detect the appearance of any personally identifiable information (PII) that might appear in their repositories, and has contributed to raising awareness about issues of data sharing in addition to enabling discoverability and findability of CGIAR data. However, data were not checked for reliability and were devoid of markers that can provide a context for research and analysis, which the Platform team considered beyond its mandate. The absence of reliability checks and markers, in addition to other factors, severely limits the ability of the Platform to advance data interoperability and reuse.

Interviewees did not report strong opinions about ethical concerns, but some considered that the culture regarding responsible research was changing. The ORGANIZE module helped CGIAR get ready for compliance with the GDPR in the EU[30] by developing a training course and a tool to scan Center repositories for any personally identifiable information that could appear in their metadata.

Several Centers have different policies regarding data sharing and for making data readily accessible. While the Platform has raised awareness about issues of data sharing and enabled discoverability and

---

[30] The European Data Protection Regulation is applicable as of May 25th, 2018 in all member states to harmonize data privacy laws across Europe.

findability of CGIAR data, it has appeared to be too heavily focused on the technology side in a way that is disconnected from the AR4D/research process itself. As a result, despite establishing GARDIAN and CG Labs, data uploads have been emphasized but analysis regarding traffic on the platforms and the number of downloads of data ready for analysis is limited or non-existent. Instead, most of the download traffic appears limited to the download of CGIAR publications and data that have not been adequately checked for reliability and are devoid of relevant markers (for example, GPS coordinates, socio-economic and institutions contextualization). To that end, by limiting the focus to data uploads, the Platform has not been able to exploit the power of big data analytics and data visualization to transform data to enable AR4D to serve its key external stakeholders/NARES while at the same time missing out on the opportunity to advance the scientific understanding of the causality between technology/management models and development impact.

18: More work is required to advance the data interoperability principle to safeguard the quality of CGIAR proposals/recommendations and strengthen its claim to the attribution of its value in AR4D from its existing data. Limited evidence was found on whether the Platform's quality control efforts improved (or not) the discoverability of CGIAR data.

The public goods research function emphasizes the potential of CGIAR research to address concerns of water, land, and/or food security. The private technology function of the Platform was meant to amplify the benefits of CGIAR publicly funded research by addressing the challenge of data interoperability (a transformational role). Instead, the focus of the conventions and inspire grants was on showcasing technology innovation (blue-sky thinking) but without an emphasis on how technology innovation could potentially amplify the benefits of publicly funded CGIAR research. Indeed, this finding is corroborated by the absence of publications that cite the results of "analytics" (not raw-unstructured datasets uploads per se) performed by the Platform and the absence of funded third party proposals that refer to the reuse of anonymized and interoperable data sets that were co-curated as part of the Platform. One way this balance can be restored via future initiatives is by addressing the skills deficit within CGIAR (hiring research programmers with an understanding of both the research process and the role of data analytics in AR4D). If not properly addressed, a skills deficit within CGIAR can affect the quality of proposals/recommendations of the CGIAR and importantly weaken its claim to existing data and attribution of its value AR4D.

FAIR is promoting metadata quality which has consequences on sharing and evaluating the quality of the data itself. Quality of data curation, including assessment of the quality of the data, has been the responsibility of the Centers but support from the Platform was planned in the 2016 Proposal. Besides the metadata quality in relation to interoperability (FAIR), the proposal also mentioned as part of ORGANIZE (Module 1) "*Strategy #1: Provide data quality training and develop quality control protocols*", *i.e.*, assist to assess the quality of datasets, and "*Strategy #2: Convey the probabilistic nature of data analytics outputs with associated uncertainti*es", "*Strategy #3: Convey the dynamic nature of data input streams and analytics outputs*", *i.e.,* reporting uncertainty as metadata for data reuse. A data management pack was released to help promote this (strategy #1) along with training courses. While GARDIAN harvests native metadata from Center repositories and cleans it of frequent errors and inconsistencies. there is no strong quality control on the metadata level *per se and* including recording the actions of quality control; *i.e.,* extending the metadata model to allow this to happen. As mentioned in the proposal this is an important aspect of data analytics when using datasets from a range of providers (Strategy #2 and #3), and ORGANIZE module leader reported during the interview "more effort on data quality, making data more interoperable (Reusability)" will be needed.

19: The Platform's Monitoring, Evaluation, and Learning (MEL) activities did not adequately support programmatic learning and reporting (successes, failures, and lessons learned). More could be done so that the Platform can serve as a mechanism that promotes CGIAR's function as a 'learning organization'.

The Platform did not have its own MEL system but followed CGIAR's MEL. Annual reports do not sufficiently highlight successes, failures, lessons learned, and any corrective measures taken. Grants achievements are reported through the Managing Agricultural Research for Learning and Outcomes (MARLO) Platform, but it is perceived by some stakeholders consulted as a mechanical task that does not foster exchanges about achievements between the Platform and its partners. The information included does not reflect sufficiently the effectiveness of the projects. Besides, the evaluation noted a lack of consistency in reporting Platform results: Platform Annual Reports refer to milestones while the Platform refers to outputs. While the budget refers to annual spending, it was difficult to establish a link between

spending and produced outputs. Monitoring metrics were not rigorously defined in the original proposal due in large part to a weak theory of change (ToC).

The ORGANIZE module did not build strong monitoring and feedback mechanisms that could foster the use and uptake of GARDIAN tools. A key assumption that the Platform has worked with is that standardized data upload and ease of data asset annotation according to reference ontologies will build a corpus of interoperable data assets that will support the development of pan-CGIAR analytics. What distinguishes big data analytics is the emphasis on transformation, which if sufficiently well accomplished will differentiate it from a database. This is a weakness of the Platform approach.

The implication of this is that the Platform need not be an initiative that covers all Centers. The principles of data sharing and access may be promoted in principle without making it mandatory. A previous CGIAR assessment (Strategic Research on Digital Transformation) suggested that when carried out in mission mode, the Platform can benefit from conventions that examine CGIAR data, models, and projects to prioritize geographies and themes/CRPs. The Annual Convention in 2019 was rated excellent or above average by more than 80% of the participants regarding speaker quality, networking, and engagement opportunities, and the overall content. A report on the Platform's 2020 annual convention made recommendations to conduct an exercise of redefining goals, target audiences, and expected outcomes from the annual event. The report also recommended the development of key points of engagement with the broader digital agriculture community through virtual and in-person events throughout the calendar year. Although the convention statistics indicated a growing audience and reach, no adequate evidence was found to conclude that the improvements were facilitated by the MEL function of the Platform.

Box 2. Interviewees' suggestions for indicators to monitor and evaluate the Platform results

A discussion regarding indicators was set up during interviews. Stakeholders pointed out that it is important to have both internal and external indicators. For internal indicators, it is not sufficient to use the number of datasets made available through the Platform, since it does not cover other important aspects, such as utility and quality. It could be helpful to look at communications coming out of Inspire projects, as well as at journal articles and blog posts that use Platform's data. It would be useful to also include Altmetric scores in the evaluation of the quality of science. Furthermore, some interviewees suggest looking at individual capacity development, at how data were exchanged both internally and externally, and at the overall change in culture regarding data collection. In terms of external indicators, some interviewees said that there has not been enough effort to evaluate the impact and assess if the use of the tools changed beneficiaries' behavior. Some suggested indicators are: the number of people reached; how data are used in developing activities; how data are used to make decisions, policies, and investments.

Other suggestions are to assess users' satisfaction in a quantitative way and to use google analytics and look at changes over time. Furthermore, some found it important to have more clarity about the investments by Centers and Donors in data.

Specifically related to the Inspire Challenge, some indicators were proposed by the interviewees and according to them they better show the success of such an initiative: (a) evidence of funded proposals, (b) joint publications emerging from use of data and analysis supported by the Platform, and c) requests for technical assistance based on co-curated data and analysis via the Platform.

Last but not least, some pointed out the importance of evaluating women's empowerment.

Source: Key Informant interviews notes

4.4 Sustainability

20: The platform was instrumental in the revision of the CGIAR OAOD policy and large challenges await in the transition to One CGIAR. The policy implies harmonizing and standardizing publications and data libraries – work that the Platform has already started mainly through GARDIAN and CoPs.

In April 2021, the CGIAR Open and FAIR Data Assets Policy[31] was approved by the System Management Board. One interviewee at the System Office indicated that the Platform contribution was instrumental for the review of the previous OAOD policy. The policy is considered as important to move to the new integrated One CGIAR operational structure and compliance with the policy will be tracked. Comprehensive Center engagement with the policy is yet to be proved as per one interviewee's opinion. Initial resistance to the Platform among CRPs was also reported; linked to competition for resources. Although the approval of the Open and FAIR data assets policy in 2021 was expected to solve these alignment challenges, to some interviewees it was not the panacea as they felt that its effective implementation was not guaranteed.

GARDIAN opted for a centralized view of all Centers with a consolidated service based on federated catalogs, updating itself from harvesting these catalogs every week. In that sense, GARDIAN contributed efficiently to the One CGIAR reform. This contribution and the service it provides is seen by several data managers and leaders as a positive outcome which many wish continuity. As laid out in the Initiative Proposal of Excellence in Agronomy for Sustainable Intensification and Climate Change Adaptation (EiA) released on September 28, 2021, *"The GARDIAN data and analytics infrastructure will be enhanced, facilitating human and machine-interpretable data, and data and analytics-dependencies in other WPs (see Outputs 1.2, 1.3, 3.2, 3.7, and 4.5; Outcome 1.2). Turnkey solutions will be developed based on validated agronomic solutions generated through the Use Cases and made accessible to additional scaling partners over time. Assumptions underlying the above processes are related to the availability and capacities of EiA scientists, NARS, and ARI colleagues to engage with standardized and actionable data and analytics assets and decision support tools for gender-, youth-, and climate-responsive recommendations."*

> *"In terms of leadership on publications and data management, Big Data/GARDIAN is the front runner for CGIAR, and big challenges await if the One CGIAR transition will imply harmonizing all publications and data libraries. Big Data/GARDIAN is, I think, uniquely well placed to help overcome those challenges." Excerpt from a Survey respondent*

Even before and in light of ongoing CGIAR reform, the Platform began new thinking on ways to organize data, and models, and modalities. This has the potential to catalyze a more centralized and holistic vision of what One CGIAR can achieve to advance food, land, and water systems in a climate crisis. This is aligned with the 2030 Research and Innovation's Strategy goal of achieving multiple SDG benefits across the 5 impact areas.[32]

21: CGIAR is well-positioned but not sufficiently prepared to have a leadership voice in international digital agriculture according to views of its internal stakeholders. Overall, stakeholders valued the outputs of the Platform and seek its continuity. External partners' engagement was undermined by the funding uncertainty of the Platform's e-infrastructure (GARDIAN).

Contrary to the negative impressions collected during interviews about CGIAR preparedness to play a leadership role in the digital agriculture landscape, the majority of survey respondents indicated that CGIAR is prepared to take on a role of leadership in the international digital agriculture landscape (67%; respondents indicating good and very good, Figure 10) and that the dream-scenario for One CGIAR is to have more data accessibility and knowledge sharing (Figure 11).

---

[31] The purpose of the CGIAR Open and Fair Data Assets Policy ("this Policy") is to clarify expectations regarding the management and dissemination of data assets (as defined in Annex 1) to ensure that they are as open as possible, always FAIR, and managed responsibly. This Policy should be read in conjunction with the CGIAR Open and FAIR Data Assets Implementation Guidelines which may be updated from time to time to reflect current recommended practices.

[32] CGIAR Five Impact areas: (1) Nutrition, Health, and Food Security (2) Poverty reduction, livelihoods, and jobs (3) Gender equality, Youth, and Social Inclusion (4) Climate Adaptation and Mitigation (5) Environmental Health and Biodiversity

"CGIAR has the experience, networks, historical and ongoing data collection efforts. Great Potential to produce standardized, open datasets representing large temporal and spatial scales." Survey respondent

Box 3. Interviewees' opinion about the challenges in using data analytics with reference to work in agriculture

> **Summary of the challenges experienced by CGIAR internal stakeholders in using data analytics with reference to work in agriculture.**
>
> **Personnel Capacity Gaps** - Key informants considered lacking or inadequate skills were impeding the use of data analytics in agriculture. Some interviewees self-assessed their expertise in data analytics as inadequate while others considered the institution-wide levels of such expertise low especially among Centers when compared to private sector entities. They called for addressing these capacity gaps through leveraging existing partnerships with universities, attracting younger talent, and encouraging the private sector to invest in the work of CGIAR. Some key informants recommended the demonstration of real-life cases and benefits in data analytics, to promote interest and individual investment as well as to address hesitance among stakeholders. Other key informants desired allocation of relevant talent to big data-related work. This was to avoid observed sub-optimal resource allocation resources cases such as Centers not sending the appropriately skilled persons to the Platform's deliberation gatherings.
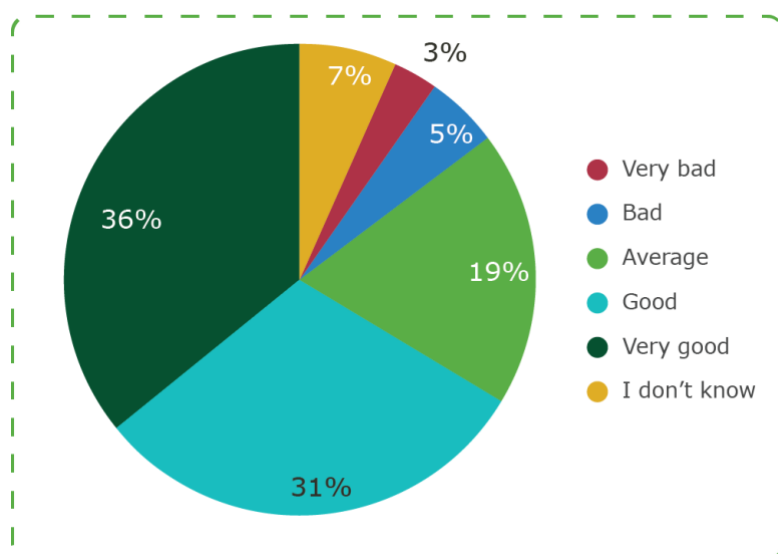>
> **Data Sharing and Interoperability** - Key informants deemed data access challenges to persist as a major hurdle for data analytics within CGIAR. Incompatible formats in which data were shared were considered to undermine its interoperability. GARDIAN was observed to have made progress addressing these challenges, but the remaining problem was linked to organizational culture where "people want to work in isolation". The key informants called for intervention on CGIAR's ways of working to reduce overlaps in engaging communities for data collection. Semantic interoperability of data generated across the organization was also considered a major challenge for data analytics. Some key informants also opined that the CGIAR lacked strong leadership on ontology as an area of improvement.
>
> **Data Quality and Relevance** - Key informants considered the quality of data and its reliability a major challenge for data analytics in agriculture. It was also opined that specificity of data and its relevance to on-the ground needs were lacking, leading to a call for emphasis on sub-national statistics. To enhance data relevance, key informants called for the adoption of Internet of Things (IoT) technologies to avail real time data for real time decision-making. They considered such real time technologies better to invest in than traditional methods such as surveys for availing real-time decision-making aids to farmers.
>
> **BIG DATA Stewardship** - Some key informants felt that inclusivity was lacking in the stewardship of the Platform. This was to the extent that they considered the Platform as not encouraging the contribution of FAIR data. Other key informants deemed feedback mechanisms for Centers on their contribution to GARDIAN unsatisfactory in ways that undermined enhancements to data analytics potential within CGIAR.
>
> Source: Key Informant interviews notes

Figure 10. Future of CGIAR – views on CGIAR's preparedness to take on a role of leadership in the international digital agriculture landscape (Online survey; n=59)

The full benefits of data findability and discoverability are yet to be achieved on a significant scale. The 2021 Strategic Research on Digital Transformation Assessment highlighted the need to pursue a skills agenda to ensure the sustainability of the Platform. Several risks that could arise in the absence of a concerted effort to improve big data skills could include the following: (a) under-developed skills can lead to erroneous conclusions, (b) the inability to capture key data because of limited digital capabilities, and (c) potential future funding opportunities being compromised if CGIAR is not able to claim full ownership and exploit the value of its data.

These are important lessons to emphasize as the sustainability of the big data platform is pursued by adopting a mission mode to prioritize engagement at selective sites worldwide to build regional capacity (by leveraging big data analytics where feasible) to respond to environmental challenges. At the same time, such a mission mode can advance understanding of local realities that can mediate the impact of CGIAR technology and management models on development outcomes.

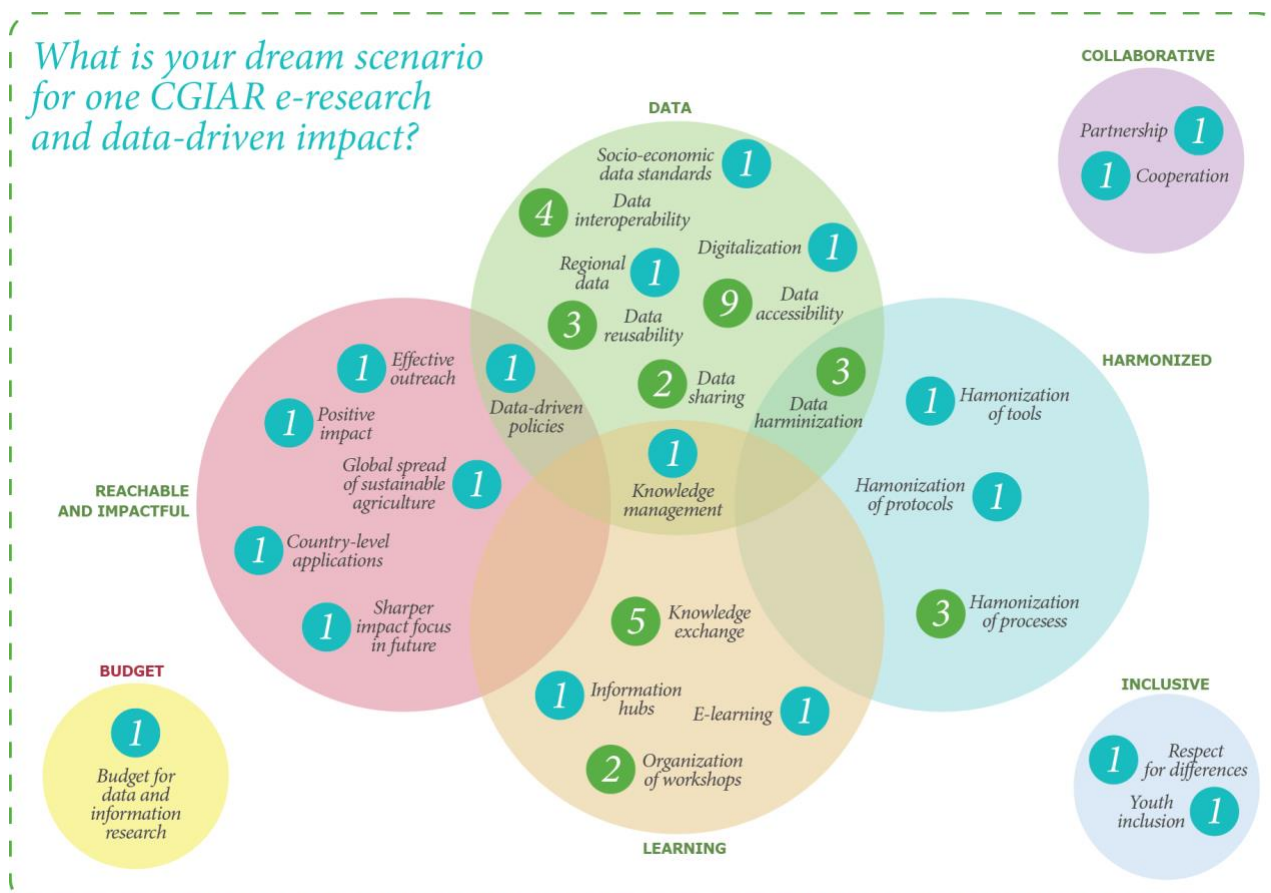***Figure 11. Dream-scenario for One CGIAR (Online survey; Open question)***

Figure 11 shows the main answers to the question "What is your dream scenario for One CGIAR e-research and data-driven impact?" Numbers represent the number of times keywords appeared. Circles represent the macro-area in which keywords can be grouped. We identified up to a maximum of three keywords used by respondents and divided them into seven macro-areas: Data, Collaboration, Budget, Reachability and Impact, Harmonization, Inclusion, and Learning. Inside each area, represented in the figure as a circular diagram, we included the keywords and the number of times they appeared.

5 Conclusions and Recommendations

5.1 Conclusions

The Platform for Big Data in Agriculture draws to a close in December 2021 as part of the CGIAR's restructuring to One CGIAR which includes a new research and innovation prospectus of Initiatives to transform food, land, and water systems in a climate crisis. Discussion with the Platform team and CGIAR staff involved in the design of these new initiatives revealed that the replication of the Platform's programmatic components will not be contained in a Platform entity as before but rather will be hosted in different initiatives (i.e., the GARDIAN e-infrastructure will be likely continued under the Excellence in Agronomy for Sustainable Intensification and Climate Change Adaptation Initiative). However, at the time of writing this report, the funding and continuity of other Platform components (CoPs, annual conventions, and the Inspire Challenge) remain unknown.

Irrefutably, the premise that drove the Platform's ToC that: *"There is unprecedented innovation at the intersection of digital technologies and life sciences that – if harnessed and applied – can provide the tools humanity needs to adapt to or mitigate some of its most pressing food security challenges"* is still valid. The Platform work added value to CGIAR's efforts to map data, methods, and tools to support the delivery of research. It did so by raising awareness around OAOD including FAIR principles, by advocating constantly during annual conventions and capacity-building activities. Yet, much effort is still needed to move forward the agenda of data interoperability and reusability.

At a higher level, the Platform's attempts and results paved the way to build more harmonized analytics towards One CGIAR. The new OAOD policy, in which the Platform team's contribution was qualified as instrumental, is a good step forward towards achieving positive results. However, it is contingent on strong governance architecture across CGIAR and the involvement of partner networks (including NARES). This vision, for which the Platform has laid fundamental blocks, needs to be driven by standards and convening efforts. CGIAR should therefore specifically adopt mission-driven digital innovation processes.

At present, some consulted stakeholders think that the "*CGIAR digital revolution claim is everywhere and nowhere"*. Referring to the 2030 Strategy in which *"Making the digital revolution central to our way of working" is* contradicted by the shortage of resources and lack of a clear vision on how to advance the research digital agenda. Communities of Practice – spaces where discussions advance thinking on how this can become a reality – are still looking for means to survive as they may disappear with the end of the Platform by the end of the year.

Having a harmonized and well-developed (research) digital data system would give CGIAR more control of the types of data that could be shared. Indeed, CGIAR researchers need a framework, a computing workflow, and tools that inform them about the *modus operandi* they can employ to make their data interoperable, which enables them to undertake a comparative analysis of biophysical processes. It is also about the institutional environment that is required to support the uptake of CGIAR technical and management innovation by end-users (Renkow, 2018). This will ensure that research is attuned to policy-relevant questions rather than being driven by the extent to which data is available for bio-physical resources and processes.

5.2 Lessons Learned

Key lessons learned during the evaluation include:

1. Data curation needs to be standardized, fit a well-defined set of requirements, and be made available to end-users with proper incentives and training in quality assurance and documentation. Without interoperability, big data, and big answers will never be achieved.

2. End-users will not be able to easily reproduce the prototypes (fully described from these cross-module activities as an open science contribution) without cross-module activities (i.e., interoperability, analytics, innovation) that lead to demonstrable proofs of concept and useful prototype capabilities (along with thorough documentation and transparent description).

3. "If you build it, they will come" does not apply to initiatives and digital artifacts for big data in AR4D. Additional efforts are required to increase awareness and uptake and early-adopter feedback is critical to iteratively refine products. This feedback increases usefulness and ease of use which are important antecedents of large-scale adoption Davis (1989).

4. Building trust with and engaging all Centers in decision-making are key success factors to ensure wide acceptance and adoption of any new, centralized technology and solutions. Existing digital solutions (such as CGSpace, digital workflows to collect standardized data) and their value should be acknowledged and built upon as One CGIAR develops a more unified digital strategy, data and digital technology governance, and an approach for rationalizing information technology investments.

5. The integration of gender as a transversal theme remains patchy with no gender expertise in-house. Piecemeal gender is less effective and less cost-effective in the long run.

5.3 Recommendations

Recommendations to improve the design and implementation of the Platform for Big Data in Agriculture and in light of upcoming new Initiatives[33] in One CGIAR with a similar goal include:

1. **Prioritize specific digital solutions for specific data (domains) aligned with agricultural research needs to demonstrate the value of the answer that (big) data can provide to support CGIAR's key priorities:**

   1.1. Develop a harmonized framework for modeling with a focus on a given geography to strengthen feedback loops between results of CGIAR field trials, and the design of policy instruments (*guidelines, standards, notifications, circulars, and directives*) through support to appropriate authorities.

   1.2. Integrate e-infrastructure design and development efforts with efforts to demonstrate e-infrastructure usability.

   1.3. Design a learning program tasked with identifying verifiable metrics to evaluate a big data pilot intervention at a dedicated site, preferably embedded within a regional network of NARES partners.

2. **Prioritize and advance the Interoperability agenda, building on CGIAR's wide variety of datasets:**

   2.1 Develop mechanisms (communication or else) to demonstrate interoperability benefits with data integration (i.e. relevant to CGIAR mandate use cases).

   2.2 Develop easy-to-use knowledge management tools from the wide variety of datasets to meet interoperability requirements.

   2.3 Allocate more resources to metadata standardization (without replicating models already available) with appropriate semantic annotations, metadata on data quality and metaquality (e.g., FAIR description and metrics), sharing metadata and data services within an interoperable manner. This should build on existing standards and ontologies developed by international bodies, e.g., the Open Geospatial Consortium (OGC) and World Wide Web Consortium (W3C).

   2.4 Add the semantic ontological knowledge base (semantic engine) in addition to the semantic enrichment of metadata from harmonized vocabulary and ontological terms, i.e., providing semantic reasoning along the keyword searching discovery capacity. Concretely, developing common definitions and standards of variables, to the extent possible, and keeping them the

---

[33] Including but not limited to Harnessing Digital Technologies for Timely Decision-Making Across Food, Land, and Water Systems. This Initiative aims to support inclusive agricultural transformation and sustainable food, land, and water management by improving information systems and strengthening digital innovation ecosystems.

same over time, where feasible, is critical. These definitions should incorporate best practices in Data Stewardship as outlined by Plotkin (2014) and be monitored and maintained over time[34].

2.5 Develop a well-thought-out and inclusive plan for designing visual analytics that is appropriate to CGIAR domains, and primarily at the basis, the geographical and temporal aspects (date and reference period) for the Platform but also in combination with semantic queries results. Engage users for feedback.

2.6 Develop the data analytics using the interoperable services provided and with an awareness of the knowledge structure.

2.7 Develop and implement a plan to empirically assess the Data and Meta-Data quality, completeness, usefulness, and shape of the data using analytical tools during upload and over time: quality of data and meta-data is critical for the adoption and use.

2.8 Conduct/commission a study on the role and integration of specific e-infrastructures, including existing CGIAR services (for example, CGSpace), becoming more known as the reference point to look for CGIAR publications metadata.

2.9 Strengthen the feedback loop: (i) develop and implement a plan to track outcomes of data and other digital artifacts developed or accessed through the Platform in terms of measurable impacts of the gathered data over time to the extent possible (ii) record end-user usage, results from data analytics methods, with feedback to the knowledge structuration, i.e., usage and results as dynamic metadata and 3rd type of the Open science aspects complementing publications and datasets, e.g., scripts, models, and software (models as statistical or machine learning but also biophysical models, crop models) and, (iii) track systematically and continuously usage analytics to evaluate the impact in terms of usability for CGIAR researchers and outside for each new launch and facilities provided, then, to be able to incorporate feedback and lessons to refine these facilities accordingly. This monitoring is also useful for the Quality of Science (e.g., views and download metrics).

2.10 Consider several 'Vs' (Volume, Value, Variety, Velocity, and Veracity) of big data in adopting and/or developing measures that go beyond FAIR to help add value to data along the continuum from storage to analysis and reporting/publishing. such an approach would facilitate the development of measures for each of the 'Vs' and improve monitoring over time.

3. **Strengthen the conceptualization (theory of change (ToC) of how the impact of agricultural development can be increased by embracing big data and ICT approaches to solve development problems faster, better, and at a greater scale:**

    3.1 Develop a ToC that articulates clearly how big data analytics can enable CGIAR research to lead to development outcomes.

    3.2 Ensure cross-cutting themes (gender and youth) are addressed more systematically and driven by a clear strategy with specific and adapted engagement mechanisms.

    3.3 Reach outside of CGIAR and/or other agri-food organizations including other sectors advanced in the digitalization process to explore what works in big data platforms/ digital transformation, etc.

4. **Raise CGIAR Entities' engagement to ensure technology solutions uptake: this can be achieved by an inclusive governance system, leveraging existing tools and incentives:**

    4.1 Give more power to CGIAR Entities at the decision-making level for example all participating Entities can be represented and have a voice in the Platform steering committee.

---

[34] It should also focus on the concepts most useful for analysis and be searchable via a variety of means, including the use of analytics tools like collaborative filters (as the data and usage grows to allow this) to suggest similar dataset that may be of interest to a user, similar to Amazon or Netflix suggesting "if you like this dataset, you might also like this one….". Granted, we are a ways from being able to do this now given the low number of users providing data, but it may be useful to at least conceptualize what could be done now so that the groundwork could be laid.

4.2 Use incentives i.e. building in a Peer-Reviewed Journal for data and meta-data[35], encouraging scientists through internal performance management to include their data in publications, etc.

4.3 Clarify through effective communication the mandate (avoid overpromising) and mission of the Platform or similar future initiatives. Use with caution the word "Big Data". The CGIAR system is characterized much more by the variety of its data rather than its size, yet its infrastructure and capability have the potential to grow into a platform that can collect and hold "big data" in the perceived "classical" sense.

5. **Build a new harmonized and interoperable analytical environment in CGIAR based on accumulated knowledge from the experience of the Platform's implementation:**

5.1. Develop a computing workflow for how data will be organized, transformed, and visualized to support the identification of a robust monitoring framework that would enable the contribution or attribution of policy changes to AR4D. Include stakeholders in the design from the beginning.

5.2. Consider the Platform's implementation experience (lessons learned, successes, and failures) and the 2021 Strategic Research on Digital Transformation assessment to inform efforts to centralize research data management and stewardship under One CGIAR. One CGIAR can build on CoP-initiated discussions on constraints and potential ways forward.

6. **Improve grant scheme management, monitoring, and governance to foster the Platform's (or successors') relevance to contribute to solving agriculture development challenges:**

6.1. Allocate more resources to deal with the high number of received proposals; enhance/create stronger feedback mechanisms throughout the process to document achievements and lessons learned.

6.2. Strengthen the link between funded projects and CGIAR digital facilities. For example; use selected projects as use cases to test new CGIAR capabilities.

6.3. Strengthen trust and ownership among CGIAR Entities by creating tighter and more transparent governance structures around grants' selection.

6.4. Rebalance distribution of grants between CGIAR Entities while ensuring the relevance of innovations selected.

6.5. In line with CGIAR's Gender and Diversity and Inclusion (GDI) strategy, ensure diversity in the decision-making body, for example through including youth in the grant selection committee.

6.6. Ensure collaboration with national innovation ecosystems to diversify applications and to harness the capacities of such innovation ecosystems for national-level advancement of big data for agriculture and AR4D.

6.7. Build a tailored monitoring and evaluation system to track results and for timely decisions.

In light of the incoming implementation of One CGIAR new Initiatives, CGIAR's 7th way of working, and the CGIAR 2030 Strategy, this evaluation of the Platform for Big Data in Agriculture recommends the CGIAR System to:

7. **Develop a One CGIAR (research) digital capability model and ensure the funding for a long-term digital plan with successive phases and a clear mandate building on the Strategic Research on Digital Transformation assessment:**

7.1 Adopt a mission-driven digital innovation process under One CGIAR.

---

[35] An additional incentive that would make it easier for funders, academics, and other researchers to have one more reason to contribute to the platform. This is building in a *Peer-Reviewed Journal* for data and meta-data. This would make it easier to include in the funding proposals, as the review team mentions, as it is a publication valuable to most researchers. It would also provide peer-review which would improve the quality of the data and meta-data. The journal *Data in Brief*, among others, is a nice example of this. Also, see (Cazier et al., 2019) for a presentation of some of the benefits tying a journal to meta-data and data standardization for pollinator data that may be relevant here.

7.2 Develop integrated cross-cutting and cross-modal analytics capabilities. For this purpose, One CGIAR can build on CoP-initiated discussions on constraints and potential ways forward.

8. **Lead** the way in hosting open data and providing analytic tools for CGIAR and its partners as well as increasing data and funding (by showing its value)**:**

   8.1 Reach out and work with international bodies and invest in the development and adoption of standards. Commission a study to map and explore open APIs required for a variety of analytical tools to interface with the data.

9. **Develop data synthesis tools[36] that are amenable for use by decision-makers to support data co-curation.**

   9.1 CGIAR should support blended learning preferably embedded with a regional network of NARES to build capacity to advance data interoperability and reuse based on use cases curated at dedicated sites. Lessons from the Agronomy and Ontologies CoP can be consolidated to support continuous learning through engagement with data and analytics.

10. **CGIAR develops a data curation and transformation dashboard to enable CGIAR and partners to access tools and technical support to undertake data harvesting, data harmonization, and visualization.**

   10.1. The impact of a data dashboard in monitoring data quality, generating anonymized datasets, and reporting on progress towards Sustainable Development Goal outcomes and the publishing of research results is likely to impact positively on the CGIAR Quality of Science (QoS).

---

[36] For example, composite indices: the principle of data co-curation is key to advancing the use of composite indices. If CGIAR can build regional capacity, then NARES would become active partners in data co-curation and by implication the goal of interoperability would be advanced.

References[37]

Arnaud, Elizabeth, Marie-Angélique Laporte, Soonho Kim, Céline Aubert, Sabina Leonelli, Berta Miro, Laurel Cooper et al. "The ontologies community of practice: A CGIAR initiative for big data in agrifood systems." Patterns 1, no. 7 (2020): 100105.

Bamberger *et all*(2015). Dealing with complexity in development evaluation: A practical approach. Sage Publications.

Bonilla-Cedrez, C., Chamberlin, J. & Hijmans, R.J. Fertilizer, and Grain Prices Constrain Food Production in sub-Saharan Africa. Nat Food 2, 766–772 (2021). https://doi.org/10.1038/s43016-021-00370-1

Bowen, G. A. (2009). Document analysis as a qualitative research method. Qualitative Research Journal. https://doi.org/10.3316/QRJ0902027

Cazier, Joseph A., Hassler, Edgar E., Wilkes, James T., Runzel, Max A., Formato, Giovanni, Brodschneider, Robert (2019) The Promise of Standardized Data and How to Achieve It. Bee Culture, November 2019 issue. Pages 43-48.

CIAT and IFPRI (2016). CGIAR Big Data Coordination Platform. Proposal to the CGIAR Fund Council, 31 March 2016. International Center for Tropical Agriculture (CIAT), Cali, Colombia, and the International Food Policy Research Institute, Washington DC, United States of America.

CGIAR (2021). One CGIAR Operational Structure. February 18, 2021

CGIAR (2020). Quality of Research for Development in the CGIAR Context, Technical Note, Independent Science for Development Council, Rome.

CGIAR System Organization (2021). CGIAR 2030 Research and Innovation Strategy: Transforming food, land, and water systems in a climate crisis. Montpellier, France: CGIAR System Organization. https://hdl.handle.net/10568/110918

Corbin, J., & Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage Publications.

Cvitanocic C. L. Norstrom and A. Reed. (2018). Building university-based boundary organizations that facilitate impacts on environmental policy and practice, PloS One 13 (9), 0203752.

Davis, F. D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", MIS Quarterly, 13 (3): 319–340, doi:10.2307/249008, JSTOR 249008

Gewin, V. (2016). Data sharing: An open mind on open data. Nature, 529(7584), 117-119.

Goldstein, A., Fink, L., & Ravid, G. (2021). A Framework for Evaluating Agricultural Ontologies. Sustainability, 13(11), 6387

Holderness, M., Howard, J., Jouini, I., Templeton, D., Iglesias, C., Molden, D., & Maxted, N. (2021). Synthesis of Learning from a Decade of CGIAR Research Programs. https://hdl.handle.net/10568/114082

Jin, X., Wah, B.W., Cheng, X. and Wang, Y. (2015). Significance and challenges of big data research. Big Data Research 2(2):59-64. https://doi.org/10.1016/j.bdr.2015.01.006

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. Journal of mixed methods research, 1(2), 112-133. https://doi.org/10.1177/1558689806298224

Kelty, C. M. (2001). Free software/free science. First Monday. doi:10.5210/fm.v6i12.902

King, B.; Devare, M.; Overduin, M.; Wong, K.; Kropff, W.; Perez, S.; Guerena, D.; McDade, M.; Kruseman, G.; Reynolds, M.; Molero, A.; Sonder, K.; Arnaud, E.; Jimenez, D.; Koo, J.; Jarvis, A.

---

[37] Including citations in the List of Annexes available as a separate document here.

(2021) Toward a digital one CGIAR: Strategic research on digital transformation in food, land, and water systems in a climate crisis. 41 p. https://hdl.handle.net/10568/113555

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data and Society 1(1): p.2053951714528481. https://doi.org/10.1177%2F2053951714528481

Kurian M. and Y. Kojima (2021). Boundary Science: Re-imagining water-energy-food interactions in the context of a data light approach to monitoring the environment- development Nexus, Elsevier, MA.

Laney, D. (2001). 3-D Data management: Controlling data volume, variety and velocity. META Group File, 949. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Leibovici DG, Santos R, Hobona G, Anand S, Kamau K, Charvat K, Schaap B and Jackson MJ (2021) Geospatial standards and their adoption through the example of the agriculture domain. In: The Routledge Handbook of Geospatial Technologies and Society, Chapter 8 (in press)

McCalla A. (2014). CGIAR Reform - Why So Difficult? Review, Reform, Renewal, Restructuring, Reform Again and then "The New CGIAR" - So Much Talk and So Little Basic Structural Change - Why?, Working Paper No. 14-001, Department of Agriculture and Resource Economics, University of California, Davis.

Madin, J.S., Bowers, S., Schildhauer, M.P. and Jones, M.B. (2008). Advancing ecological research with ontologies. Trends in Ecology and Evolution 23(3):159-168. https://doi.org/10.1016/j.tree.2007.11.007

MARLO (2018). List of Big Data Projects, 2018 Reporting Cycle for M2 (Convene) module.

Mertens, D. M. (2017). Mixed methods design in evaluation (Vol. 1). SAGE publications

Mockshell et al. (2021). CGIAR Platform for Big Data in Agriculture. Inspire Challenge Review (2017-2020). Inspiring the next generation of digital innovations in food systems through bridging institutions: Where do we go from here? https://hdl.handle.net/10568/113597

Plotkin, David (2014). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance / David Plotkin. 1st edition. Amsterdam: Elsevier, 2014. Print.

Rapley, T. (2007). Doing Conversation, Discourse and Document Analysis. Qualitative Research Kit.

Renkow, M. (2018). A Reflection on Impact and Influence of CGIAR Policy-Oriented Research. Rome, Italy: Standing Panel on Impact Assessment (SPIA), CGIAR Independent Science and Partnership Council (ISPC). 34 pp

Runzel M, P. Sarfatti and S. Negroustoueva (2021). Evaluating quality of science in CGIAR Research Programs: Use of bibliometrics, Outlook on Agriculture, Sage Publications.

Vicente-Saez, R., and Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. Journal of business research, 88, 428-436.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9. https://doi.org/10.1038/sdata.2016.18

© *CGIAR Platform for Big Data in Agriculture*