



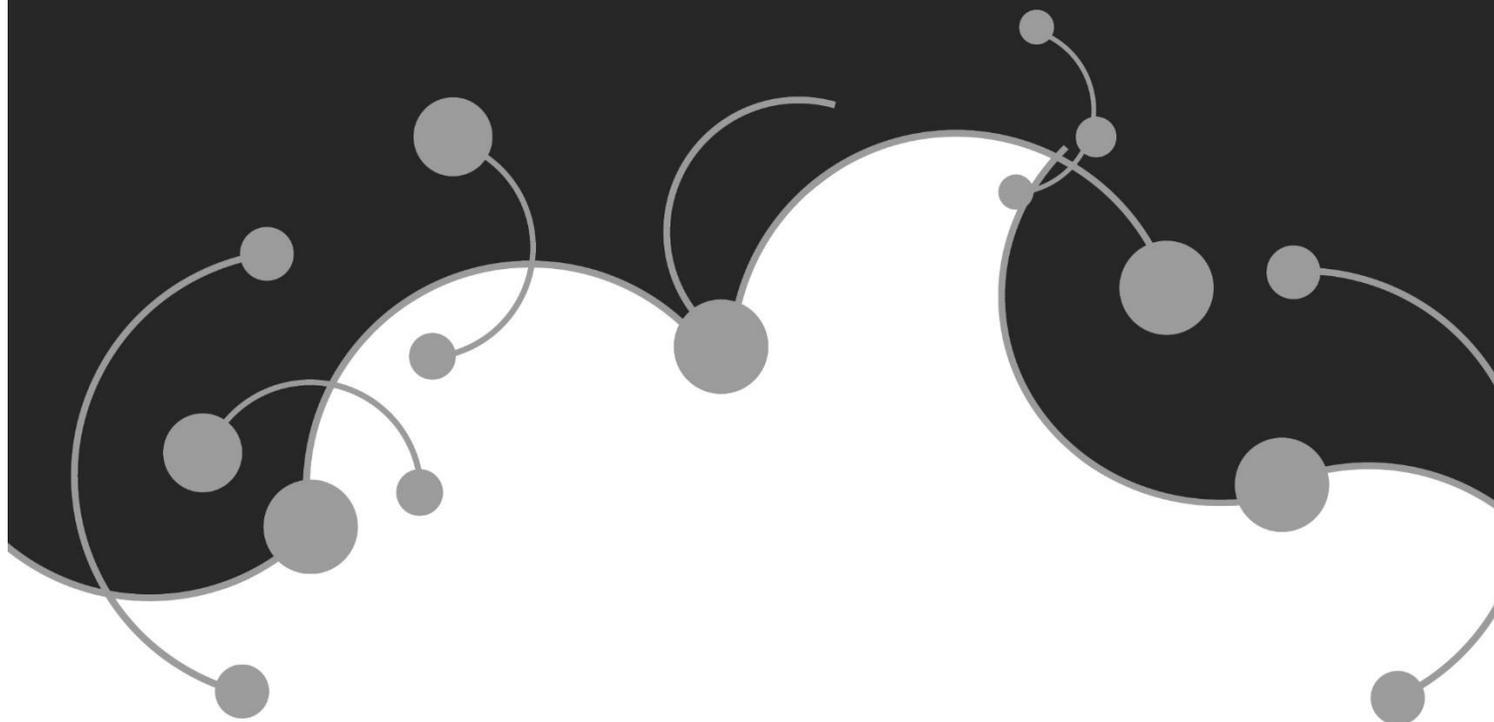
CGIAR

Big Data
in Agriculture
CONVENTION

19-22 SEPTEMBER. CALI, COLOMBIA

2017

EVENT SUMMARY
& AGENDA FOR ACTION



FIRST ANNUAL CGIAR CONVENTION ON BIG DATA IN AGRICULTURE: "Alliance For A Data Revolution"



CGIAR

Led by



INTERNATIONAL
FOOD POLICY
RESEARCH
INSTITUTE

IFPRI

SUMMARY

On September 19-22, 2017 the Consultative Group for International Agricultural Research¹ (CGIAR) gathered over 300 local and international researchers, non-profits, public and private sector actors for the first CGIAR Platform for Big Data in Agriculture Convention, hosted by the International Center for Tropical Agriculture (CIAT) in Palmira, Colombia. The Convention marked the programmatic launch of the Platform, which aims to enable the development sector to embrace data and other digital technology approaches to *solve agricultural development problems faster, better and at greater scale*.

The Platform works across the CGIAR network and CGIAR Research Programs (CRPs) and with the gamut of stakeholders in the agriculture sector as they grapple with creation, curation, and sharing data to enable new approaches to complex development challenges.

The Platform is designed around three strategic pillars: *Organize, Convene, and Inspire*. The first aims to organize data so datasets are findable, accessible, and interoperable so they can be used increasingly in big data analytics. In addition, this pillar will develop open digital infrastructures for the sector that support the CGIAR's work and enable new partnerships and innovations. The aim to convene analysts, researchers and public, private and non-profit actors in the agriculture sector will build new partnerships that both shape and fully leverage digital technologies in support of global agricultural development. The final pillar is to inspire these actors to push the limits of research and innovation to generate new data-driven approaches that solve real world development problems faster, cheaper, and more efficiently.

WHAT DO WE MEAN BY "BIG DATA"?

Open, harmonized, interoperable, and integrated datasets from multiple domains aimed to accelerate agricultural research and data use in service of development goals.

¹ For those who may wish to cite this report, may we suggest: *Event Summary and Agenda for Action, First Annual CGIAR Convention on Big Data in Agriculture: Alliance for a Data Revolution, September 19-22, 2017*. November 02, 2017.

The CGIAR Platform for Big Data in Agriculture Convention 2017: Alliance for a Data Revolution served as a consultative process with a rich cross-section of the agriculture development sector during which the following priorities were identified and affirmed:

Drive Data Discoverability and Use

For the full potential of data to be unlocked it must not only be discoverable in digital format but also be accessible enough that others are able to re-use it to reproduce research or build responsive and available new digital services. Currently, interoperability and flow of data remains constrained within major classes of sector data infrastructure, including farmer advisory services, breeding programs, and predictive or geospatial analytics.

The Platform for Big Data in Agriculture is well-positioned to help the CGIAR network and its partners leverage open data assets to develop new digital capabilities for building resilient food systems. During the convention, the Platform unveiled *CGIAR e-Research (CeRES)*, a robust prototype of a searchable CGIAR-wide data harvester. The Platform will continue to support Centers' execution of open data policies, and will evaluate new data architectures to enable meaningful applications of these data.

Define responsible data

Data sharing is critical for digitally-enabling the agriculture development sector. While the enthusiasm for sharing grows, we also recognize the need to continue examining the ethical dimensions of data collection, sharing and use. The convention participants agreed that, rather than continuing in a circular discussion – the potential of big data on the one hand, privacy and ethics concerns on the other – we need to find ways to balance these responsibly so that data can contribute to improved agriculture outcomes without creating unacceptable risks of harm to those who provide it or are affected by the resulting decisions. The Platform for Big Data in Agriculture is actively addressing this. Leveraging external and internal expertise, and drawing from similar efforts in other sectors, we are preparing to draft responsible data guidelines for the agriculture development sector that aim to balance ethical concerns with unlocking with the power of data sharing.

Spur Sector-Specific Innovations

Big data analytics is an emerging discipline for the agriculture development sector. Digital innovation “processes and places” could have a key role to play in spurring adoption and cutting-edge applications of data. The Convention was the final stage of the Platform’s 2017 Inspire Challenge, an innovation grant program that drew submissions from 120 teams across 37 countries seeking to advance leading-edge uses of data for enabling market systems, disrupting impact assessment, diagnosing and managing pests and disease, and advancing data-driven farming in developing economies. The process laid bare the creative power of CGIAR Centers and their partners to push the limits of the possible in research and application of data for agriculture development. The Platform for Big Data in Agriculture will support the [five winning proposals](#) and continue to advance the learning agenda about how these innovations can drive impact in agriculture development.

1. INTRODUCTION: WHY A BIG DATA PLATFORM FOR AGRICULTURE DEVELOPMENT?

The agriculture sector as a whole has yet to fully unlock the potential of digital technologies. For example, a recent report by McKinsey found that, even in industrialized economies like the United States, the agricultural sector ranks 23rd out of 25 industries in digitalization² and the rate of adoption is slow. When it comes to low and middle-income countries, it appears that these innovations have an even smaller foothold. Recent World Bank research in Kenya, Senegal, South Africa, and Armenia³ found that actors in the mobile application ecosystem of these countries perceived the agricultural sector to be the least impacted by digital technology and well behind other sectors (See Table 1); only 5 of 73 digital technology firms surveyed identified the agricultural sector as a source of revenue.

² *Digital America: A tale of haves and have-mores*. 2015 <https://www.mckinsey.com/industries/high-tech/our-insights/digital-america-a-tale-of-the-haves-and-have-mores>

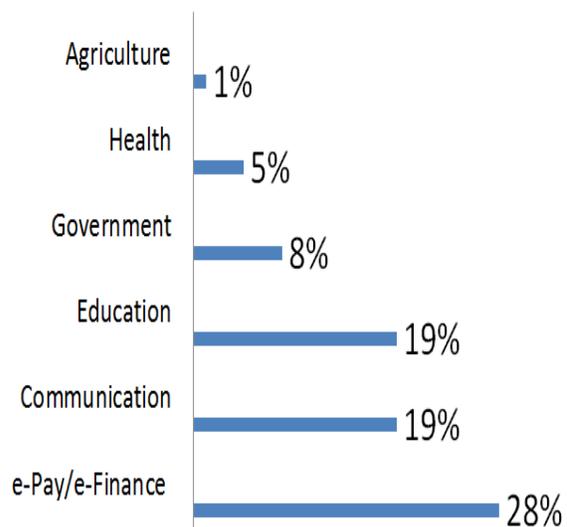
³ “Do mLabs Still Make a Difference? A Second Assessment.” 2017. Washington, DC: The World Bank Group.

This trend also appears to be reflected in projects funded by the U.S. Agency International Development (USAID) Development Innovations Ventures fund, where in August 2015 at Stage 1 (start-up phase) 25% of funded projects focused on agriculture, at Stage 2 this number fell to 14%, and by in final Stage 3 there were none.⁴

The funding and financing ecosystem for these innovations in developing economies is at a nascent stage. For example, venture capital investment in agriculture technologies (“AgTech”) in 2016 is estimated to be around USD \$3.2 billion⁵ but only a little over 3% is estimated to have flowed into developing economies. Furthermore, much of the total investment is made up of very large deals (for example, John Deere’s acquisition of the precision agriculture company Blue River Technologies for US \$305 million). Present trends in funding and financing will likely continue to favor development of digital agriculture technologies for industrialized markets, leaving small holders in developing economies underserved and in danger of missing out entirely.

In emerging economic regions, however, a number of digital trends should be harnessed for agriculture development including: the mobile revolution of widespread ownership and use of mobiles to both collect and send data; (2) rapid improvements in remote sensing data availability and resolution make it more applicable for agriculture; (3) the rapid spread of cheap sensor technologies enable “Internet of Things” business models and new approaches to on-farm measurement; and (4) distributed computing and storage are democratizing and expanding analytical power like never before; and (5) worldwide we are seeing myriad processes and places for digital innovation aiming to build startup ecosystems and spur the development of new digital business models.

Table 1: Where Has Digital Technology Impacted National Development ?



⁴ “Deconstructing USAID Innovation Strategies.” Maryland: UBDP llc, 2017.

⁵ “AgTech Investing Report Year in Review 2016, “Jan 31, 2017.

The global footprint of the CGIAR Centers and Programs, and the enormous amount of research data they generate, position the network and its partners very well to shape these trends to favor small farmers and developing economies. Collectively this community of actors can harness these digital trends and become critical infrastructure for global adaptation.

2. KEY ISSUES RAISED AT THE CONVENTION: THE DEVIL IS IN THE DETAILS

It is hard to argue with the overall proposition that the CGIAR, and the sector as a whole, needs to gather their data more systematically and make it usable by a wider range of partners. A more systematic approach to data sets offers new possibilities for innovative data analysis and development of new digital services, as working with real datasets can be more complicated in practice. The following issues were raised as key constraints or priorities for action among the gathered Centers, CRPs, university, non-profit, and private sector partners:

a. Big data challenges some incentives and approaches of traditional scientific research

To fully utilize the power of data analytics on a large scale, the sector as a whole needs to move further on the path of making data findable, accessible, interoperable and reusable; yet much of scientific research is incentivized solely by the final research product. Furthermore, big data analytics may enable new approaches to scientific research. Such as using enhanced pattern recognition to more rapidly identify meaningful relationships in data, helping to generate new hypotheses and conduct more timely observations on a greater scale.⁶

⁶ For example, strong proxies for changes in wealth have been established in mobile phone network metadata (e.g. who called whom, duration of call, amount of phone credit purchased) that can be monitored at regional or national scales.

b. Common data infrastructure can foster more meaningful data use, yet “one platform” approaches are challenging

Meaningful use of data is made easier by unified processes for its collection, organization, storage, and application, and this naturally creates incentives to build unified data platforms around specific needs. Moving towards more unified processes can also contribute to the seductive idea of building a “single definitive platform”, which may actually hinder data sharing and agility in the sector. These challenges can be compounded when trying to improve data sharing and interoperability across different (public, private, non-profit, national, global) actors. Agile, function-driven, iterative approaches to sharing and use may prove more fruitful, such as between major classes of sector data including farm trials, breeding programs, and predictive or geospatial analytics.

c. Communities of practice may be overlooking relationship synergy potential

The Platform’s communities of practice (CoPs): *Ontology, Social Science, Crop Modeling, Geospatial* and *Data Driven Agronomy*, were created to investigate and demonstrate how to best apply leading-edge data tools for agriculture development. They are open to all practitioners and are intended to anchor the programmatic decisions of the Platform in tangible benefits to specific technical disciplines in agriculture development. These communities have significant overlap and participants felt they could be more synergistic in practice. How will they interact to provide mutual support in pursuit of Platform’s goals and help build resilient agriculture and food systems?

d. Data ethics & legal frameworks are still evolving

Data sharing is critical for digitally-enabling the agriculture development sector, yet even while the enthusiasm for sharing is growing, it is important to continue to examine the ethical dimensions of new approaches to data collection, sharing and use. The ability to capture and recombine digital data, including even “live” data in

real time, can impact human subjects in a variety of new ways that need to be better understood.

Existing legal frameworks and ethics guidelines may not be evolving as quickly as these multifaceted data-driven capabilities, that may have implications for technology, confidentiality, intellectual property, consent, access and sharing of benefits.

e. There is growing consensus that research data needs to be FAIR, yet this can be challenging in practice

There is a growing consensus that research data should be findable, accessible, interoperable, and reusable (FAIR), while recognizing that the incentives of multiple actors including donors, private sector, researchers, and data subjects themselves may not always easily align. Each of these actors may have different ethical, fiduciary, or personal interests that may serve as disincentives to making data available.

f. We don't yet fully know how to link data to actual change

Local context and various other drivers and incentives independent of data determine how knowledge is consumed and acted on. In order to better understand how open data/big data can lead to benefits for smallholder farmers it is important to understand the wider ecosystem and the incentives of different stakeholders at various levels. What does the whole chain from "data" to "benefit for small shareholder farmers" look like in different contexts, and where will public-good actors like CGIAR play a role? What partners are needed and what will their roles be in deriving value out of open and/or big data? How will the benefits of these efforts actually reach farmers? How will farmers themselves shape this process?

3. PROGRAMMATIC NEXT STEPS: ACTION ITEMS RAISED DURING THE CONVENTION

Participants identified actionable next steps to further 'big-data-enable' the agriculture development enterprise that can be implemented over the coming year.

a. Invest in data discoverability across CGIAR in tandem with open data efforts at each Center

CGIAR e-Research (CeRes), presented for the first by the CGIAR Platform for Big Data in Agriculture during the convention, is a robust prototype of an open, searchable data harvester, spanning databases across all CGIAR centers. CeRes will continue to be developed and populated, and semantic web methods will be applied to CGIAR data to enable searching across domains. In parallel, the Platform will continue their support of Centers' execution of the CGIAR open data policy.

We need to stop thinking about data merely in terms of the questions we want to answer, and instead think in terms of the data-driven capabilities we want to build.

b. Develop draft reference architectures for key sector functions

The Platform for Big Data in Agriculture will pursue an iterative, stakeholder-driven approach to designing solutions to improve data flow, interoperability, and use in three areas over the coming year:

- Improving data flow and between gene banks, breeding programs, and field trials
- Linking analytics platforms with farmer-facing digital decision support systems
- Enabling new public-private data sharing where interests may align (e.g. pre-competitive research)

The Platform will be biased towards action, favoring agile, function-driven, iterative approaches to sharing and making meaningful use of data by actively promoting agricultural data exchange —rather than seek to be the definitive single platform.

c. Develop draft guidelines for ethics and accountability

- The Platform will take a consultative approach with data providers and users to define and promote the practices that best address the challenges of responsible data sharing. These draft guidelines will encompass technology issues (such as use of tracking systems or encryption), confidentiality, intellectual property rights, privacy protection, ethics approvals, prior informed consent, and access and benefit sharing related to traditional knowledge and genetic information, among other aspects.

d. Advance the learning agenda on digital innovation “places and processes” for agriculture

Big data analytics is an emerging discipline for the agriculture development sector. Digital innovation “processes and places” could have a key role to play in spurring adoption and cutting-edge applications of data, yet more needs to be learned about how to best match digital innovations with agriculture development challenges. The Platform for Big Data in Agriculture will seek to advance this learning agenda over the coming year by:

- Developing a synthesis document that captures lessons from the CGIAR Inspire Challenge and how to better match such processes to sector development challenges, and
- Examining in more depth what kind of digital interventions make sense at different parts of agriculture and food systems, and examining these in light of agriculture innovation strategy.

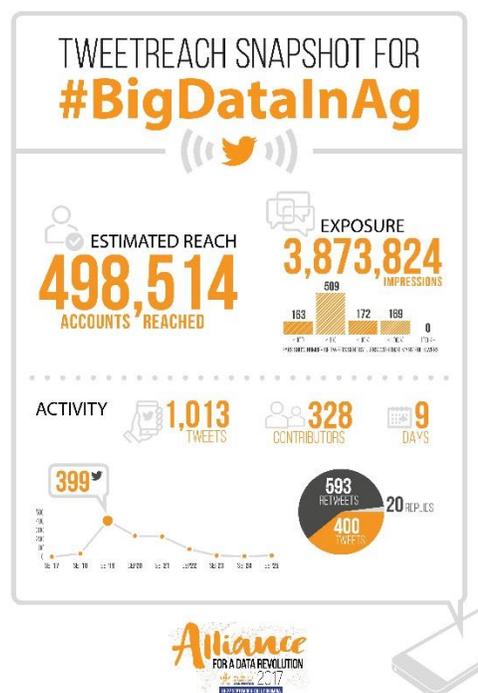
e. Learn more about data and “change”

The Platform will seek to advance the learning agenda about the impact of data-driven agriculture development overall through conducting market research, developing and testing theories of change about improved data use in food systems, and seeking to understand the impact of the concrete actions described here on outcomes such as productivity, resilience, and farmer well-being.

4. ROLE OF CGIAR IN REGARDS TO THE PLATFORM FOR BIG DATA IN AGRICULTURE

The global footprint of the CGIAR Centers and Programs and the enormous amount of research data they generate positions the network and its partners well to both leverage and shape the big data trend to favor small farmers and developing economies. Collectively this community of actors can harness these digital trends and become critical infrastructure for global adaptation. The CGIAR needs to fill several catalytic roles:

- **A Thought leader:** Generate the ideas and debates that will bring about the kinds of sector-level changes envisaged. It must be able to clarify differences so that partners can ally and move forward more quickly.
- **A Communicator:** Strive to be a key source of best information on the issues raised at the convention. The audiences are both extremely broad and very niche, including: the CGIAR network itself, the agriculture development enterprise as a whole, the nearly 500,000 Twitter accounts that engaged with the Platform during the Convention, but there are many, many different niches.



- **A Trailblazer:** Lead by example with trailblazing projects that demonstrate how the new data-driven approaches we communicate about will work.
- **A Facilitator of common standards and access:** Create or promote common frameworks (e.g. Field Books) or secure new digital services or good quality data sets to help “big data-enable” the sector.
- **A Facilitator of Innovation:** Birth new public, private and non-profit actors and approaches, with people coming out of research community and more typical start-up organizations.
- **A Facilitator of the ecosystem:** Drive down the cost of both public and private innovation. It is not about being in the business of picking winners but rather to create processes by which winners can emerge.