

Meet-up #3

CoP-SED

September 12, 2017

WebEx / CIMMYT HQ, El Batan, Mexico

Gideon Kruseman



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Agenda

1. Opening and quick introductions
2. Announcements
3. Brief introduction to the platform and the CoP
4. Big data Convention
5. Latest from the working groups
6. Closing



CGIAR Platform for Big Data in Agriculture



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



The problem

CGIAR currently lacks the tools necessary to get the most out of our data,

Insights from our research take (too) long to reach those that need them the most,

Our data is not always easy for those inside CGIAR to use,

Let alone those outside CGIAR



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED

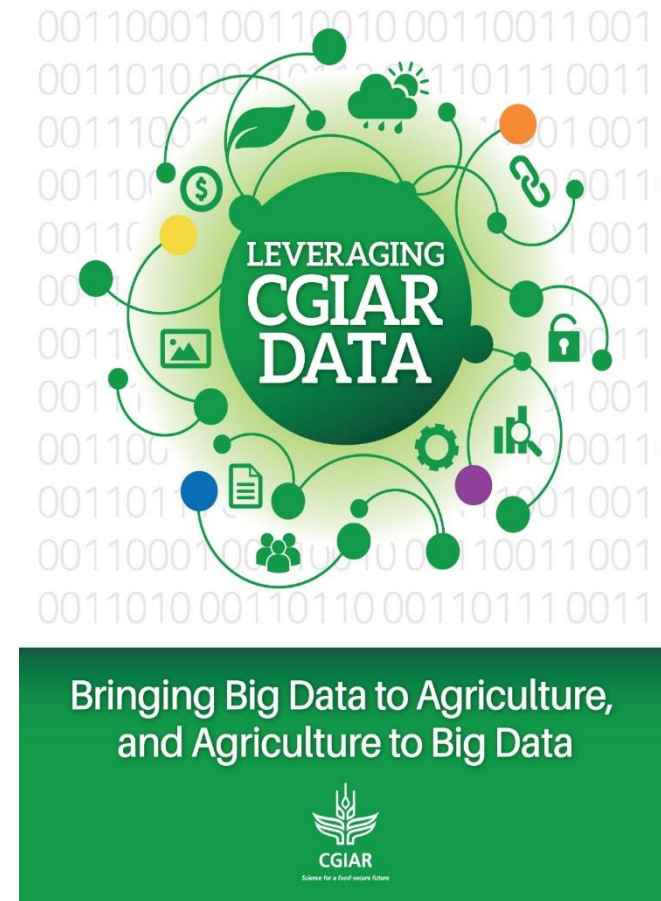


The objective

Harness the capabilities of big data to accelerate and enhance the impact of international agricultural research



Platform for
Big Data
in Agriculture



Bringing Big Data to Agriculture,
and Agriculture to Big Data



Community of practice
on socio economic data
#CoP_SED



The vision

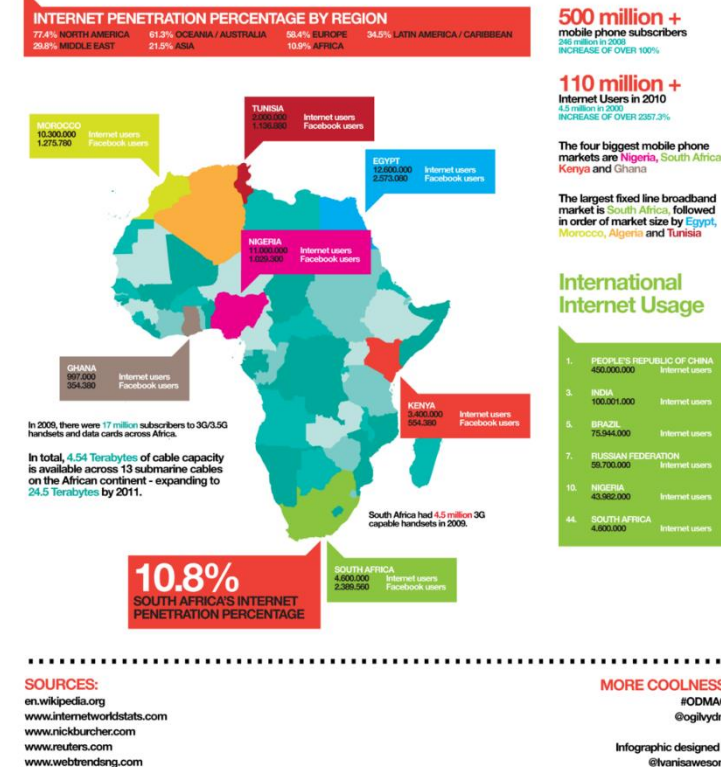
- The data revolution is changing the role, reach and modus operandi of research and development organizations. It represents an unprecedented opportunity to find new ways of reducing hunger and poverty, but also has its risks: unequal access to and use of information
- **How do we close equity gaps instead of widening them?** This is at the core of the CGIAR mission.



Platform for
Big Data
in Agriculture



The African Digital Frontier



Community of practice
on socio economic data
#CoP_SED



Platform by name,
Innovation Hub by nature

Organize



Support data generation and management, practices, enable FAIR data

Convene



Bring together big data practitioners, private sector, academia, (I)NGOs++ to produce innovative solutions

Inspire



Employ analytics and ICTs to deliver info to farmers, monitor ag and food security, and inform policy

#CoP_SED



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Our challenge

CGIAR collects lots of socio-economic data.

This data is not always made available for further use, other than the primary purpose of collection.

If it is archived for further use, it is not interoperable with other data sets.

Besides CGIAR data there are other data sets that are relevant: *e.g.* LSMS-ISA

Our challenge is to start making the data available and interoperable.



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Our challenge

Socio-economic data is by far the most complicated type of data:

- Structured, semi-structured, unstructured
- No standards for kind of data collected
- No standards for values collected

We do not seek to enforce standards that will be obsolete before they are developed!

Where useful, we will use standards

But, we seek interoperability through different means



Platform for
Big Data
in Agriculture

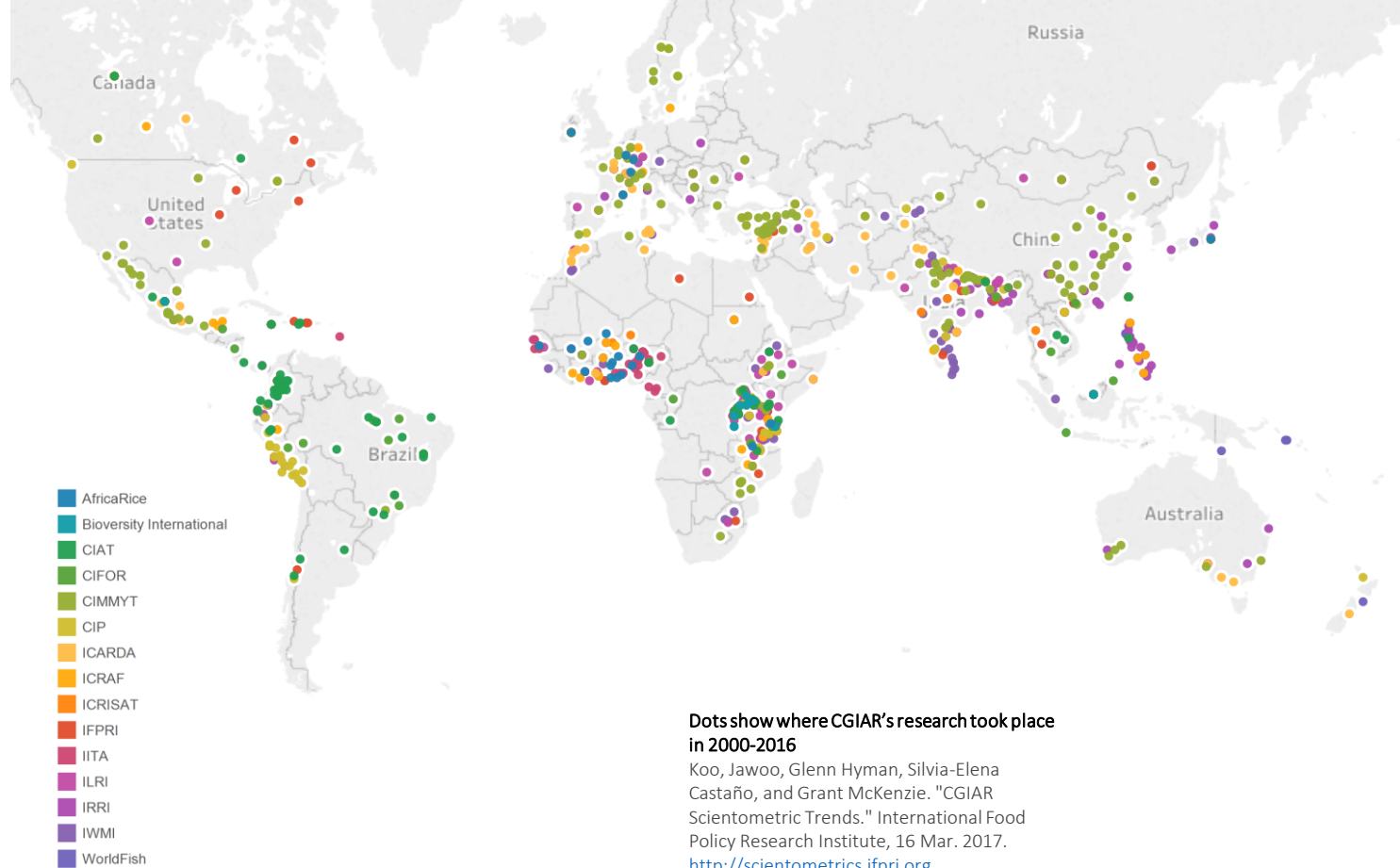


Community of practice
on socio economic data
#CoP_SED



Challenge 1

A minimum set of questions to be included in ongoing and future surveys and extractable from historic data that will allow rapid use of data for research, and policy decisions.



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Challenge 2

Ensure that data sets have relevant metadata attached.

1. Technical metadata: format, location
2. Descriptive metadata: CGCore, DublinCore
3. Structural metadata: machine readable information that allows harvesting data from wherever it is deployed into workflows and analytical tools



Research Data
Repository

CIMMYT Dataverse Network >

CIMMYT Research Data Dataverse

POWERED BY THE **Dataverse Network** PROJECT v. 3.0
Create Account Log In

Free, open access repository of research studies developed by CIMMYT scientists.

The screenshot shows the CIMMYT Research Data Dataverse interface. On the left is a sidebar with a tree view of categories: Agricultural Systems, Genetic Resources, Maize, Socio-Economics (selected), Adoption Pathways Project, Wheat, and Seeds of Discovery. The main content area is titled 'Socio-Economics' and includes a search bar, a 'Go' button, and a link to 'Advanced Search Tips'. Below the title, there's a 'Sort By' dropdown and a pagination bar showing 'Studies: 23' with navigation arrows. A list of studies is displayed, each with a title, author, abstract, HDI number, download count, and release date.

Study Title	Author	Abstract	HDI	Downloads	Last Released
Zero-Tillage Service Provision as a Business Opportunity	by Keil, Alwin; Michalski, Joel	Abstract: The purpose of the study was (1) to assess the economic viability of custom-hire service for ZT in CSISA's target districts in Bihar, the development of ZT service provision businesses over time, and constraints to further business expansi... Continue [+]	hdl:11529/10841	98 downloads	16/01/2017
Zero-Tillage Adoption and its Welfare Impacts at the Farm Household Level	by Keil, Alwin; Michalski, Joel	Abstract: The purpose of the study was (1) to assess the performance of ZT wheat as compared to conventional-tillage wheat in farmers' fields in six CSISA target districts in Bihar; (2) to assess farmers' resource endowment, risk exposure, risk pref... Continue [+]	hdl:11529/10840	48 downloads	23/12/2016
Pathways to Sustainable Intensification in Eastern and Southern Africa - Ethiopia 2010	by Marenja, Paswel; Kassie, Menale; Yirga, Chilot; Muricho, Geoffrey; Alemu, Solomon	Abstract: A multi-stage sampling was employed to identify households. In the first stage 9 districts (five from Oromiya region, three from SNNP region and one from Benshangul region) were selected purposely. Accordingly, Bako Tibe, Gubuesyo, Shalla,... Continue [+]	hdl:11529/10746	922 downloads + analyses	22/11/2016
Pathways to Sustainable Intensification in Eastern and Southern Africa - Tanzania 2010	by Marenja, Paswel; Kassie, Menale; Mishili, Fulgence; Muricho, Geoffrey; Alemu, Solomon		hdl:11529/10754	254 downloads + analyses	



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Working groups

- Survey of surveys
Define common questions
- SociO!
Concepts and classifications (ontology) needed to understand the data content
- Structural metadata schema
Make the data interoperable in a flexible and extensible way



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



01101100	01100101	01110110	01100101	01110010	01100001
01100111	01101001	01101110	01100111	00100000	01100011
01100111	01101001	01100001	01110010	00100000	01100010
01101001	01100111	00100000	01100100	01100001	01110100
01100001	00100000	01100010	01110010	01101001	01101110
01100111	00100000	01100010	01101001	01100111	00100000
01100100	01100001	01110100	01100001	00100000	01110100
01101111	00100000	01100001	01100111	01110010	01101001
01100011	01110101	01101100	01110100	01110101	01110010
01100101	00101100	00100000	01100001	01101110	01100100
00100000	01100001	01100111	01110010	01101001	01100011
01110101	01101100	01110100	01110101	01110010	01100101
00100000	01110100	01101111	00100000	01100010	01101001
01100111	00100000	01100100	01100001	01110100	01100001

Big Data Convention

- What is the role of the CoP's in the convention?
- Where are the CoP's visible:
 - Tuesday afternoon: CoP pitches
 - Wednesday afternoon: outward session
 - Thursday morning: lightning report
 - Friday mid-morning onwards: inward session
- Opportunities for cross CoP link-ups



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Wednesday

- There will be three hours for CoP's to pursue some of their high-level goals such as:
 - Articulating what we'll be looking to produce in our work plan
 - Defining more clearly any new tools, services or product requirements
 - Our vision for what sort of data might be shared and how it will flow amongst Centers and partners
 - Clarity on what sorts of standards/protocols/tools etc. are required to enable easier data sharing, discovery, integration
 - State of the industry, any new themes or advances therein
 - Emphasizing the scope and agreeing on areas that you will not be focusing on



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Friday

- 10am until... we decide to call it quits
- This is an open-ended CoP Business day, a chance for our CoP to share amongst members, conduct internal business, cross-pollinate between the CoPs, and start hacking away at some of the products agreed to on Wednesday, a chance to have a beer and do some group bonding, team building etc. (whatever we decide we want to do).



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



01101100	01100101	01110110	01100101	01110010	01100001
01100111	01101001	01101110	01100111	00100000	01100011
01100111	01101001	01100001	01110010	00100000	01100010
01101001	01100111	00100000	01100100	01100001	01110100
01100001	00100000	01100010	01110010	01101001	01101110
01100111	00100000	01100010	01101001	01100111	00100000
01100100	01100001	01110100	01100001	00100000	01110100
01101111	00100000	01100001	01100111	01110010	01101001
01100011	01110101	01101100	01110100	01110101	01110010
01100101	00101100	00100000	01100001	01101110	01100100
00100000	01100001	01100111	01110010	01101001	01100011
01110101	01101100	01110100	01110101	01110010	01100101
00100000	01110100	01101111	00100000	01100010	01101001
01100111	00100000	01100100	01100001	01110100	01100001

Next CoP meet-ups:

- Big Data Convention (live)
- November (virtual)



Platform for
Big Data
in Agriculture

- Learn more about us at bigdata.cgiar.org
- Follow us at Twitter [@CGIAR_data](https://twitter.com/CGIAR_data)
- #BigDataInAg #CoP_SED