

The role of big data for next generation climate change scenario analysis

Gideon Kruseman, CIMMYT

Coordinator community of practice on socio-economic data @BDP | Foresight, ex-ante impact assessment and targeting research leader @CYMMYT

AGMIP GLOBAL & REGIONAL ECONOMICS GROUPS JOINT WORKSHOP, Joint Research Centre, Seville, March 8-9, 2018



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Overview

- What is the *CGIAR Platform for Big Data in Agriculture* [#BigDataInAG](#)
- Data interoperability
- What is the role of the *Community of Practice on Socio-Economic Data* [#CoP_SED](#)
- Flavor of what the platform holds in store
 - Copernicus climate and weather services
 - Data integration
- Getting involved

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



CGIAR Platform for Big data in Agriculture



Platform for
Big Data
in Agriculture

bigdata.cgiar.org

[#BigDataInAg](https://twitter.com/BigDataInAg)

CGIAR Platform for Big Data in Agriculture

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



Bringing Big Data to Agriculture,
and Agriculture to Big Data



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



A Place, A Movement, A Mission...



Platform for
Big Data
in Agriculture



Platform for
Big Data
in Agriculture

to solve agricultural
development problems faster,
better and at greater scale

ORGANIZE



#fact

**EVERY YEAR,
CGIAR SURVEYS**

1 8 0, 0 0 0

**SMALLHOLDER
FARMERS**

Kickstart a data ecosystem in agriculture

- Ensure all CGIAR data is FAIR
- Generate data sharing platforms and protocols
- Create a harvester to facilitate discoverability of data



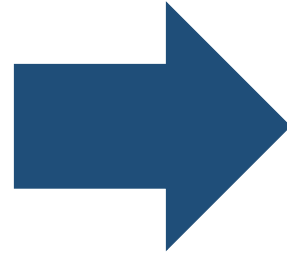
CONVENE



Supporting data generation, access, and use

So much more than just a webpage...

- Building Capacity throughout CGIAR to generate and manage big data
- Advancing open access / open data compliance to unlock our data
- Developing and sharing tools to analyze and visualize our data



Requires **open access to datasets** and harmonization on **interoperability** (standards and ontologies), **tools/platforms**, and **incentives/culture**



Platform for
Big Data
in Agriculture



CONVENE



Collaborating to put big data to use beyond the lab

Acting on data to deliver results

- **Convening the first NGO-led international** big data in agriculture annual conference
- Facilitating collaboration among partners and stakeholders to generate ideas and innovations
- Delivering data-based actionables to smallholder agriculture

15 CGIAR CENTERS

12 CGIAR RESEARCH PROGRAMS

> 48 EXTERNAL PARTNERS



CGIAR STUDIES COVER 78 COUNTRIES WORLDWIDE



Published Year
All

CGIAR Center

- | | |
|--------------------------|-----------|
| AfricaRice | ICRISAT |
| Bioversity International | IFPRI |
| CIAT | IITA |
| CIFOR | ILRI |
| CIMMYT | IRRI |
| CIP | IWMI |
| ICARDA | WorldFish |
| ICRAF | |

INSPIRE



Leading by example to deliver development results

Reaching for tomorrow's solutions, today

- Challenging partners, universities and others to use our data to create pilot opportunities that scale
- Funding novel approaches through venture capital
- Inspiring our network to use big data to create impact

#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

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 >

POWERED BY THE **Dataverse Network** PROJECT v. 3.0

CIMMYT Research Data Dataverse

Search [Create Account](#) [Log In](#)

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

The screenshot shows the CIMMYT Research Data Repository interface. On the left is a navigation menu with 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 shows a list of studies. The first study is 'Zero-Tillage Service Provision as a Business Opportunity' by Keil, Alwin; Michalski, Joel, with hdl:11529/10841, 98 downloads, and a release date of 16/01/2017. The second is 'Zero-Tillage Adoption and its Welfare Impacts at the Farm Household Level' by Keil, Alwin; Michalski, Joel, with hdl:11529/10840, 48 downloads, and a release date of 23/12/2016. The third is 'Pathways to Sustainable Intensification in Eastern and Southern Africa - Ethiopia 2010' by Marenya, Paswet; Kassie, Menale; Yirga, Chilot; Muricho, Geoffrey; Alemu, Solomon, with hdl:11529/10746, 922 downloads + analyses, and a release date of 22/11/2016. The fourth is 'Pathways to Sustainable Intensification in Eastern and Southern Africa - Tanzania 2010' by Marenya, Paswet; Kassie, Menale; Mishili, Fulgence; Muricho, Geoffrey; Alemu, Solomon, with hdl:11529/10754, 254 downloads + analyses.



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Initial FAIR working groups

- 100Q
Define common questions and indicators
- SociO!
Concepts and classifications (ontology) needed to understand the data content
- Ontology Independent 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



New working groups

- Blockchain Coalition
Solving the insuperable problems in value chains with data – bridging the trust gap and democratizing VCs
- Ethics, privacy, data confidentiality and cybersecurity WG
Moving goal posts regarding human subjects data – how to not do harm with really sensitive data
- CAPI WG
Making the most of Computer assisted personal interviews
- Gender WG
Making gender data FAIR (Findable, Accessible, Interoperable, Reusable)



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED



Flavor of what the platform holds in store



Platform for
Big Data
in Agriculture





Climate Change

Climate Change Service Indicators for Agriculture

climate.copernicus.eu





Climate
Change

C o p e r n i c u s C l i m a t e D a t a S t o r e

Under construction:



you can help shape it to your needs!

Fully *open access* to global climate products (graphics/maps + full data)

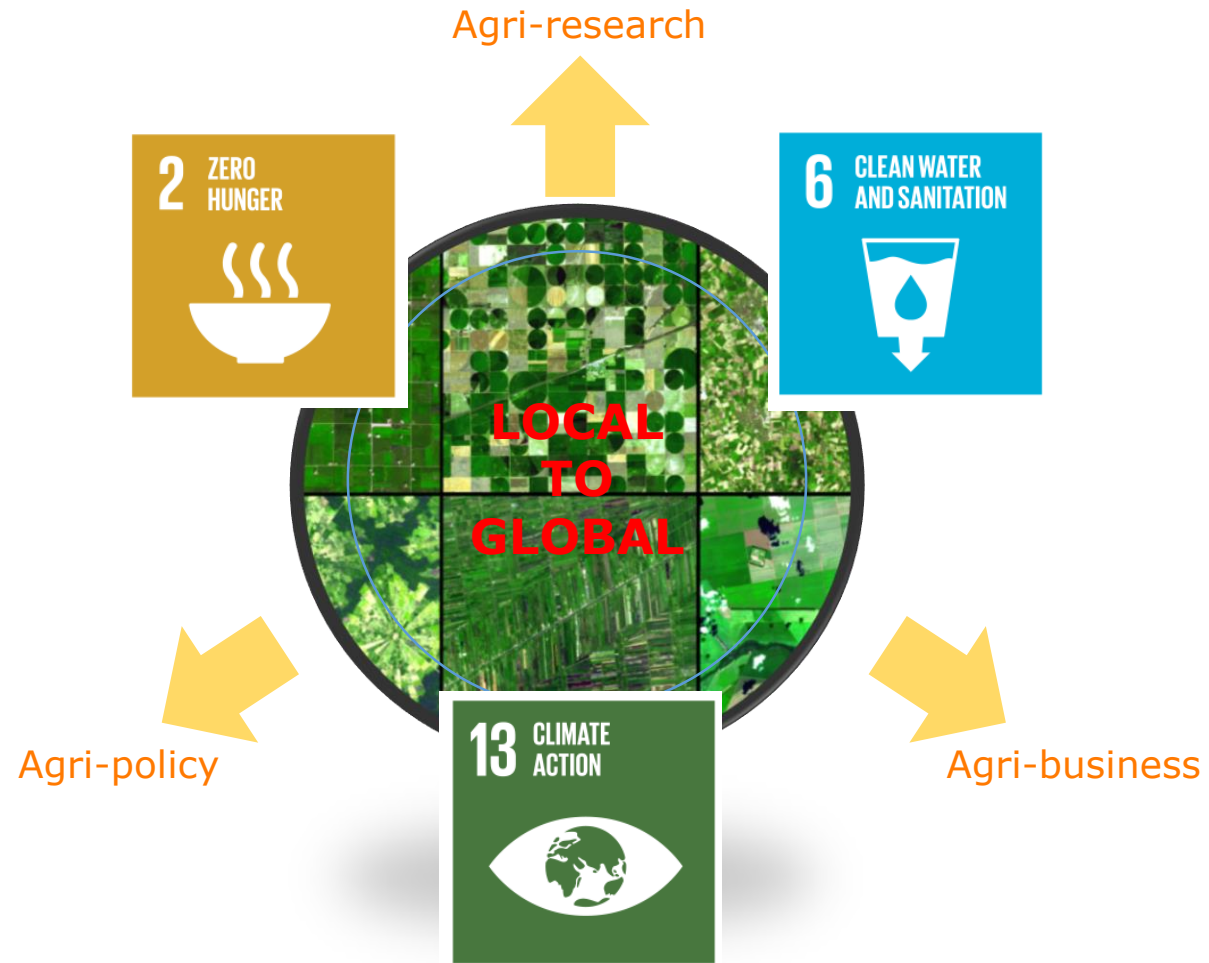
- Historic-Near Real Time: ERA5
- Seasonal Forecasts: multi model ensemble 3 SFS's
- Climate Change: multi model ensemble - CMIP5 (6)

+ Sector specific derived products: **Agriculture**, Hydrology, Energy,...



Climate
Change

Climate Indicators for Agriculture





Climate
Change

Climate Indicators for Agriculture

4 data product groups:

1. Climate forcing data for crop models
 - all time scales, downscaled, bias corrected
2. Water based indicators
 - Soil moisture, groundwater recharge, reservoir inflow, river flow, ET, SPI, ...
3. Agroclimatic indicators
 - Growing degree days, huglin index, cold/heat stress days, insect flight index, ...
4. EarthObservation based indicators
 - Dry Matter Productivity, ET_{actual} , ...

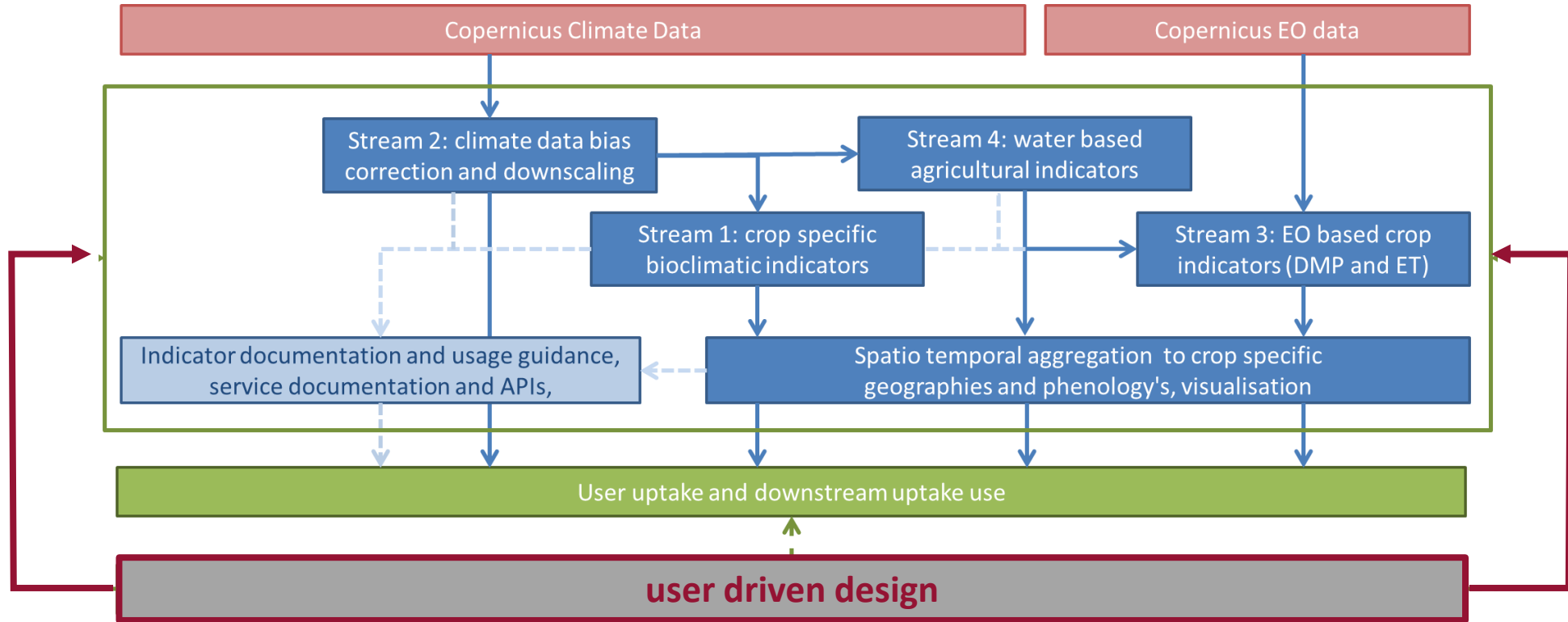
All aggregated to
crop specific

- phenological calendars
- growing areas



Climate
Change

Climate Indicators for Agriculture



Getting engaged

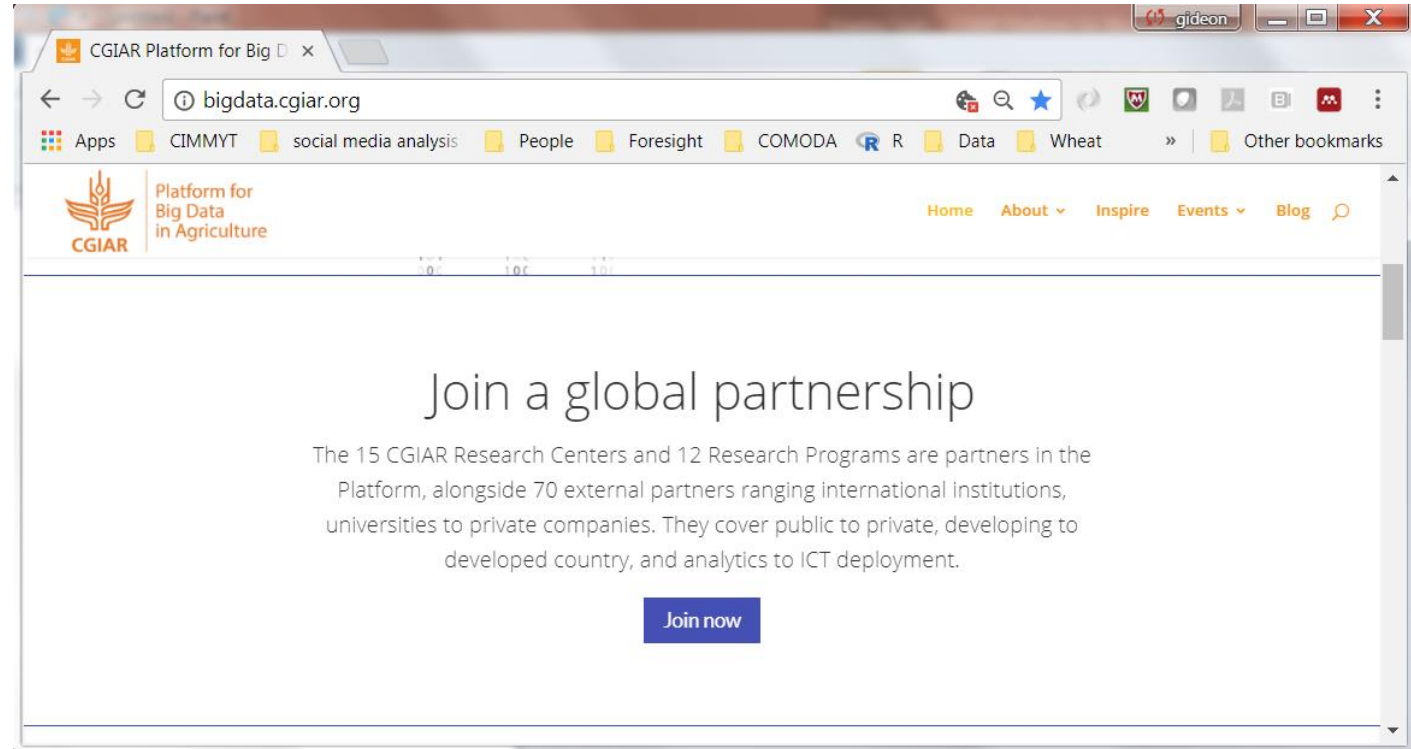


Platform for
Big Data
in Agriculture



What can you do

1. Sign-up on bigdata.cgiar.org



Platform for
Big Data
in Agriculture



Community of practice
on socio economic data
#CoP_SED





Platform for
Big Data
in Agriculture

bigdata.cgiar.org

#BigDataInAg