



Overview of Progress in OA at CIP

Research Informatics Unit - RIU



Outline

1

**CIP's Approach
in Open Access
& Research Data
Management**

2

CIP's role in
the Big Data
Platform

3

Open Access
implementation
at CIP

4

Challenges
and
Opportunities

Open Access at CIP

CGIAR Open Access and Data Management Policy

“CGIAR regards the results of its research and development activities as international public goods and is committed to their widespread dissemination and use to achieve the maximum impact to advantage the poor, especially smallholder farmers in developing countries. **CGIAR considers Open Access to be an important practical application of this commitment** as it enhances the visibility, accessibility and impact of its research and development activities. Open Access improves the speed, efficiency and efficacy of research; it enables interdisciplinary research; assists novel computation of the research literature...”

Open Access at CIP

Immediate, irrevocable, unrestricted and
free online access

Information Products

Articles, books, book chapters, Data
Training materials, extension
materials, Databases, photos, videos,
reports



Findable

Accessible

Interoperable

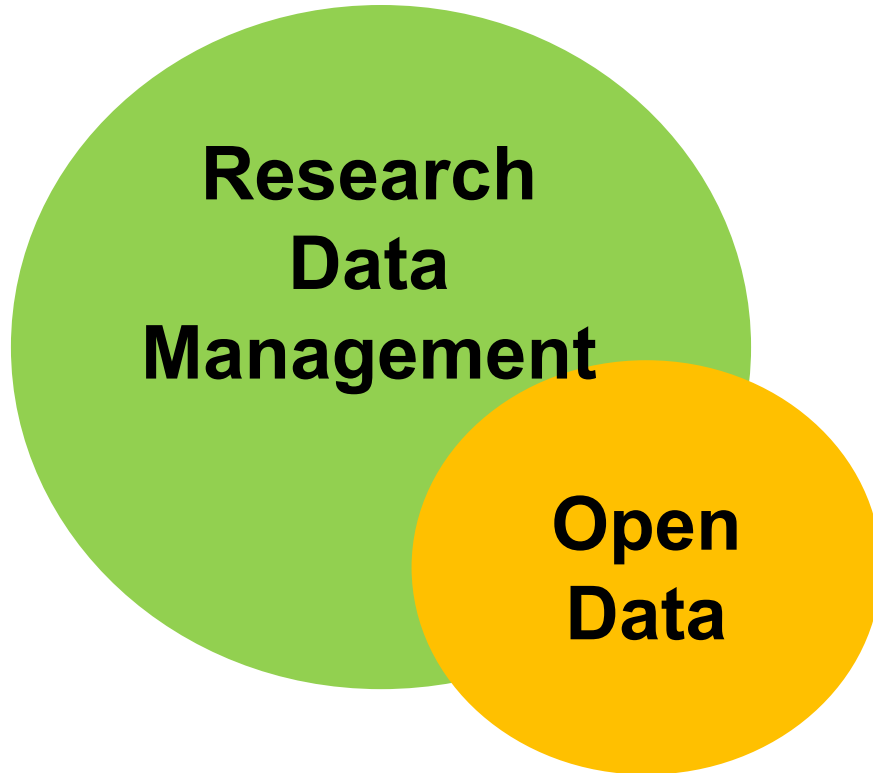
Re-usable

Deposit Timelines

CIP's towards full compliance of Open Access (OA) in 2018

Type of Information Product	Transition Deposit Schedule (until Oct 1, 2018)	Policy Deposit Schedule (Oct 2, 2018 onwards)
<u>Peer-reviewed versions of articles</u>	As per the Policy Deposit Schedule unless OA is prohibited or subject to a longer embargo period by publisher	Ideally, <u>at time of publication</u> Latest: <u>6 months from publication</u>
CIP-published journals, books, reports, etc.	Immediately	Immediately
Reports and other papers	As soon as possible Latest: within 6 months of completion	As soon as possible Latest: within 3 months of completion
Books/book chapters	As per Policy Deposit Schedule	As soon as possible Latest: within 6 months of completion
<u>Data and data sets</u>	As per Policy Deposit Schedule	As soon as possible Latest: <u>within 12 months of completion of data collection</u> or appropriate project milestone, or <u>within 6 months of the publication of the information products underpinned by that data</u>

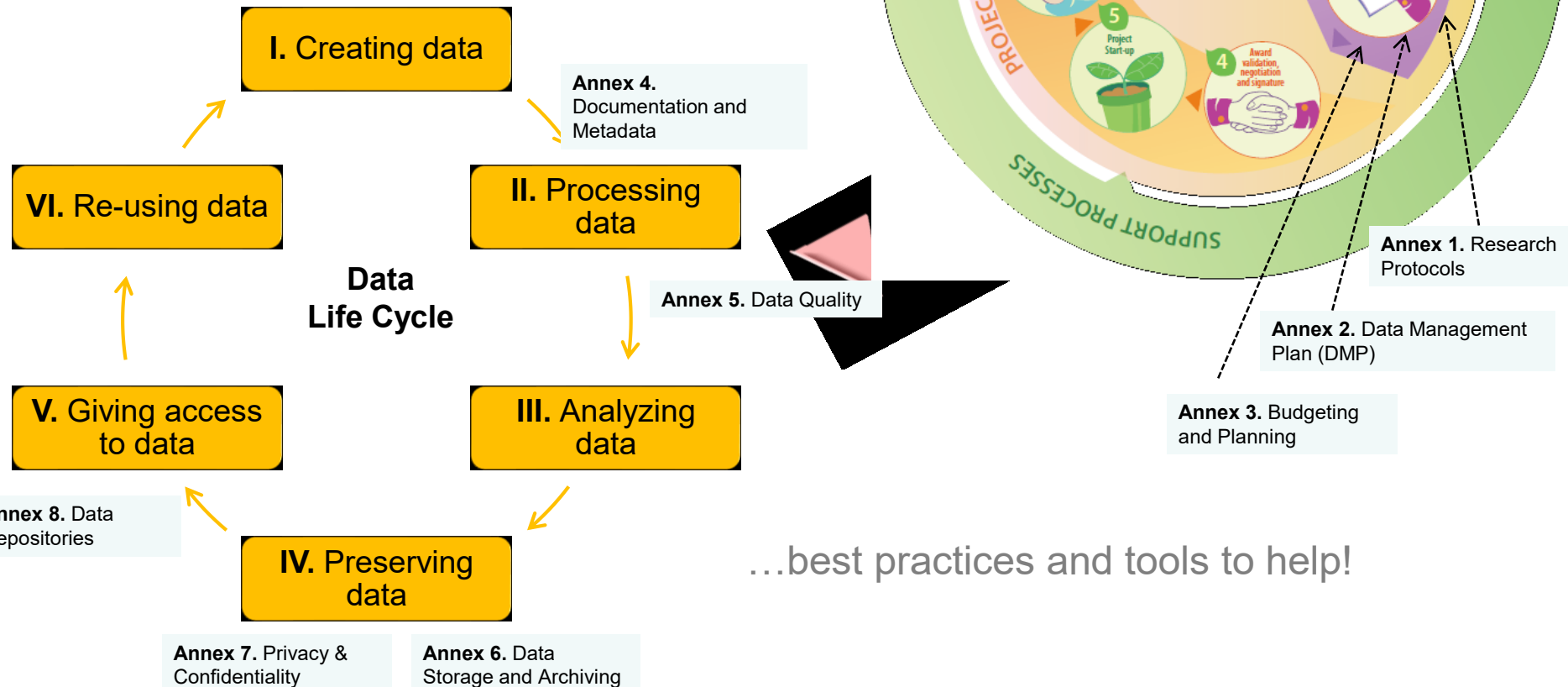
Research Data Management and Open Access



- ✓ **Research Data Management** all aspects of creating and supporting research data (cleaning, processing, analyzing, archiving, preserving datasets)
- ✓ **Open Data**: The process of making research data openly-accessible for others to find, access, and use

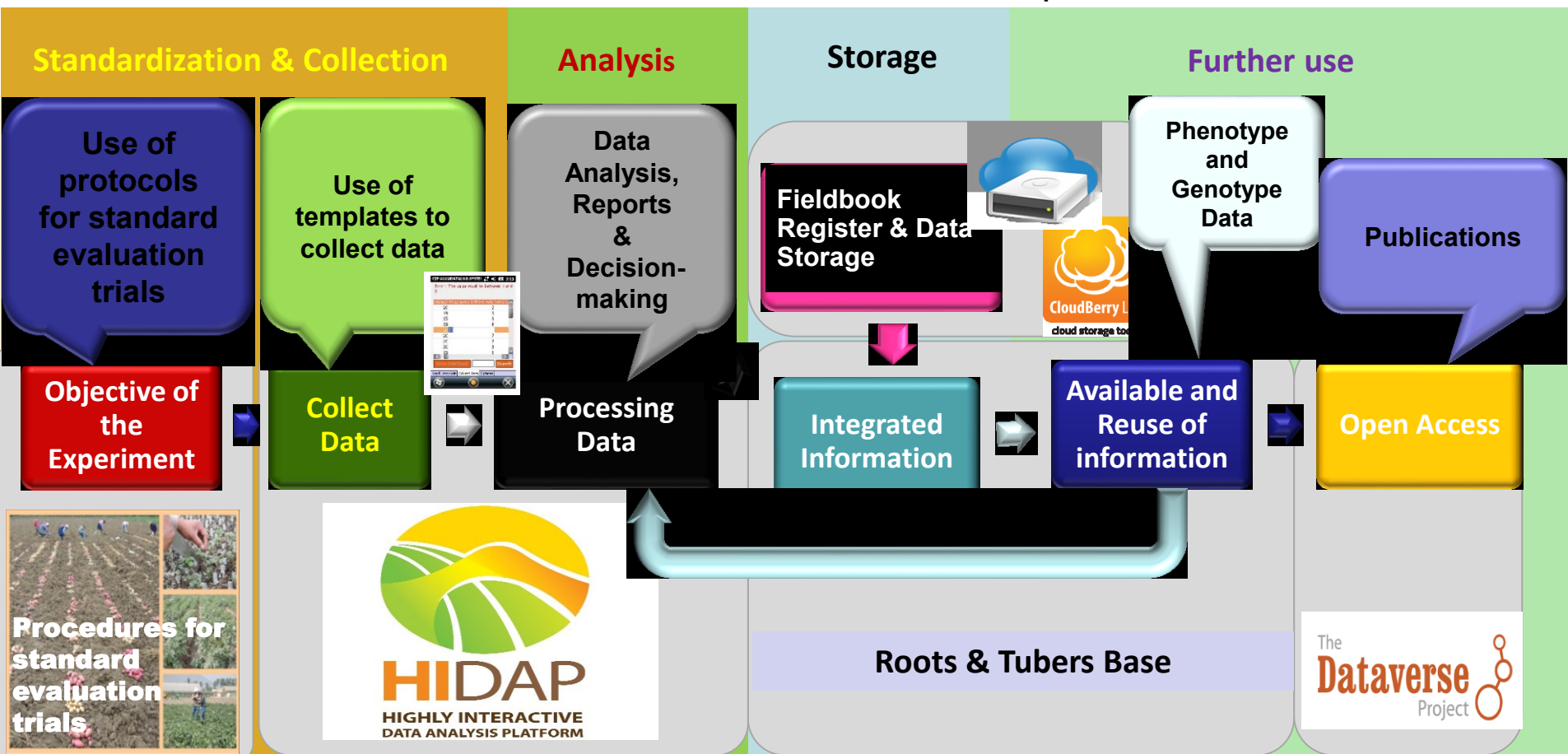
Integrated approach..

- The Research Data Management Policy
- Research Data Management Guidelines and Procedures



Workflow 1: data management of breeding information

Division: Genetics and Crop Improvement

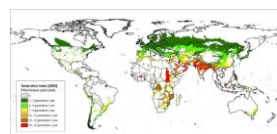


- The final dataset is published in Dataverse to make it OA as the final step in the Data Life Cycle.

Workflow 2: Development of pest risk maps using ILCYM

Division: Crop Systems
Intensification and Climate
Change

"Risk mapping"



Distribución potencial a nivel mundial, regional
y local



Clima actual: 2000

(1950-2000: www.worldclim.org/)

$$ERI = \frac{(\sum_{i=1}^{365} R\hat{o}_i) / 365}{\max(R\hat{o}_i)}$$

$$GI = \frac{(\sum_{i=1}^{365} 365 / \hat{G}L_i)}{365}$$

Con y sin filtros de cultivos

ERI

GI

AI

Clima futuro: 2050

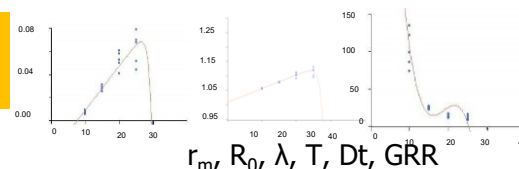
Down-scaled data SRES-A1B, IPCC (2007):
<http://gisweb.ciat.cgiar.org/GCMPPage>

$$AI = \log_{10} \left(\prod_{i=1}^{365} \hat{\lambda}_i \right)$$

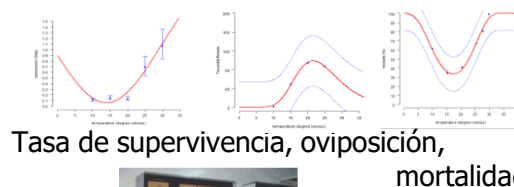
Índices de riesgo

Validación del modelo

Simulación de parámetros de tabla
de vida

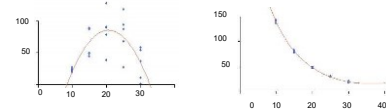


Modelos fenológicos basados en la
temperatura

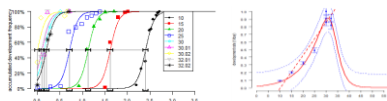


Estudios de tabla de vida

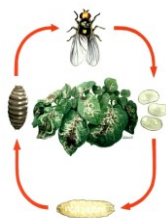
A temperaturas constante y
fluctuante: Oviposición, tiempo de
supervivencia, tiempo de desarrollo,
mortalidad



Parámetros poblacionales:



Tasa y tiempo de desarrollo



"Model builder"

Outline

1

CIP's Approach in
Open Access &
Research Data
Management

2

**CIP's role in
the Big Data
Platform**

3

Open Access
implementation
at CIP

4

Challenges
and
Opportunities

Big Data Platform



Platform for
Big Data
in Agriculture

The CGIAR Big Data Platform aims to increase the impact of agricultural development by embracing big data approaches to solve development problems faster, better and at greater scale than before.



Organize

Support data generation
and management,
practices, enable FAIR
data



Convene

Bring together big data
practitioners, private sectors,
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

CIP's very active!

Organize

Support data generation, open access, and management

- ✓ Policies and guidelines of OA in place supporting processes and workflows linking to our framework Project Life Cycle.
- ✓ All resources for OA are available at Open Access & Open Data Toolkit (<http://cipotato.org/open-access>) along with training materials.
- ✓ CIP has set up a local installation of Dataverse <https://data.cipotato.org/> for Datasets
- ✓ ~125 researchers trained on OA/OD during 2016/2018. There's a need on more training to understand and implement OA specially in the regions.
- ✓ Open Access Sprint a success story [October 2016 – March 2017].
- ✓ DMP for all new bilateral projects at the startup meeting.
- ✓ Workflows for data management in place: protocols, ontologies, tools for data quality/ standardization / statistical analysis, databases, including OA.



Organize

Support data generation, open access, and management

Active participation on Communities of practices and Working Groups:

- ✓ CGIAR Dataverse WG.
- ✓ CGIAR Metadata WG
- ✓ CGIAR Ontology WG
- ✓ CGIAR Open Access WG



HiDAP AgroFIMS, the Agronomy Field Information Management System



The screenshot displays the HiDAP AgroFIMS web application. On the left is a dark sidebar with a logo at the top and navigation links: Fieldbook, Single Trial Analysis, Documentation, Help, and About. The main content area has a green header with the title 'HiDAP AgroFIMS' and a subtitle 'Agronomy Field Management System'. Below the header is a banner image showing a group of diverse farmers in a field. The version 'HIDAP AgroFIMS v0.0.18' is displayed. A paragraph describes the system's development on CGIAR's HiDAP platform and its use of ontologies. A bulleted list outlines its features, such as standardizing data collection and enabling data uploads. At the bottom, a line of text mentions funding from the Bill and Melinda Gates Foundation and the CGIAR Big Data Platform. A row of logos for partner organizations (CGIAR, CIAT, CIP, IFPRI, and Bioversity International) is at the very bottom.

HiDAP AgroFIMS
Agronomy Field Management System

HIDAP AgroFIMS v0.0.18

The Agronomy Field Information Management System (AgroFIMS) has been developed on CGIAR's [HiDAP](#) (Highly-interactive Data Analysis Platform created by CGIAR's International Potato Center, [CIP](#)). AgroFIMS draws fully on ontologies, particularly the Agronomy Ontology and the Crop Ontology. It consists of modules that represent the typical cycle of operations in agronomic trial management, and enables the creation of data collection sheets using the same ontology-based set of variables, terminology, units and protocols. AgroFIMS therefore:

- Standardizes data collection and description for easy aggregation and inter-linking across disparate datasets;
- Allows easy integration with HiDAP breeding data, or any other ontology-based datasets;
- Functions as a data staging repository, allowing data uploads with view/edit permissions;
- Enables data quality checks, statistical analysis of the data collected, and the generation of sophisticated statistics reports;
- Aligns a priori with CGIAR's CG Core metadata schema;
- Enables easy upload to the institutional repositories, and much more.

Funding for AgroFIMS was provided by the Bill and Melinda Gates Foundation's Open Access, Open Data Initiative, and the [CGIAR Big Data Platform](#).

 Platform for Big Data in Agriculture

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

 **CIP**
INTERNATIONAL POTATO CENTER
A CGIAR RESEARCH CENTER

 INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

 **Bioversity**
International

- ✓ AgroFIMS is being developed at CIP with ontology and agronomists experts from Bioversity, IFPRI, and Ontocale Inc.
- ✓ An AgroFIMS prototype was subject to user testing in September 2018 and garnered interest and enthusiasm.
- ✓ A first release will be available in early 2019 for field testing to capture data during the 2019 cropping season.

Convene

Collaborate and convene around data and agricultural development

Active participation on:

- ✓ Geospatial Data
- ✓ Crop Modelling
- ✓ Ontologies Data
- ✓ Socio-economic Data

Need representatives on:

- ✓ Data-driven Agronomy
- ✓ Livestock Data



Accces to GDBX, DigitalGlobe's high-resolution Earth imagery along with analytics

DigitalGlobe Platform

My Library

Discover

+ New Notebook

Documentation

Plans

Explore Notebooks

Find Notebooks

Notebook Tutorials



Image Science Tutorials



Analysis at Scale Tutorials



Applied Use Cases



New Methods & Models

View all notebook collections >



ALGORITHM DEVELOPERS



ALGORITHMS



GDBX PLATFORM



DERIVED DATA



APPLICATION DEVELOPER



ACTIONABLE INFORMATION



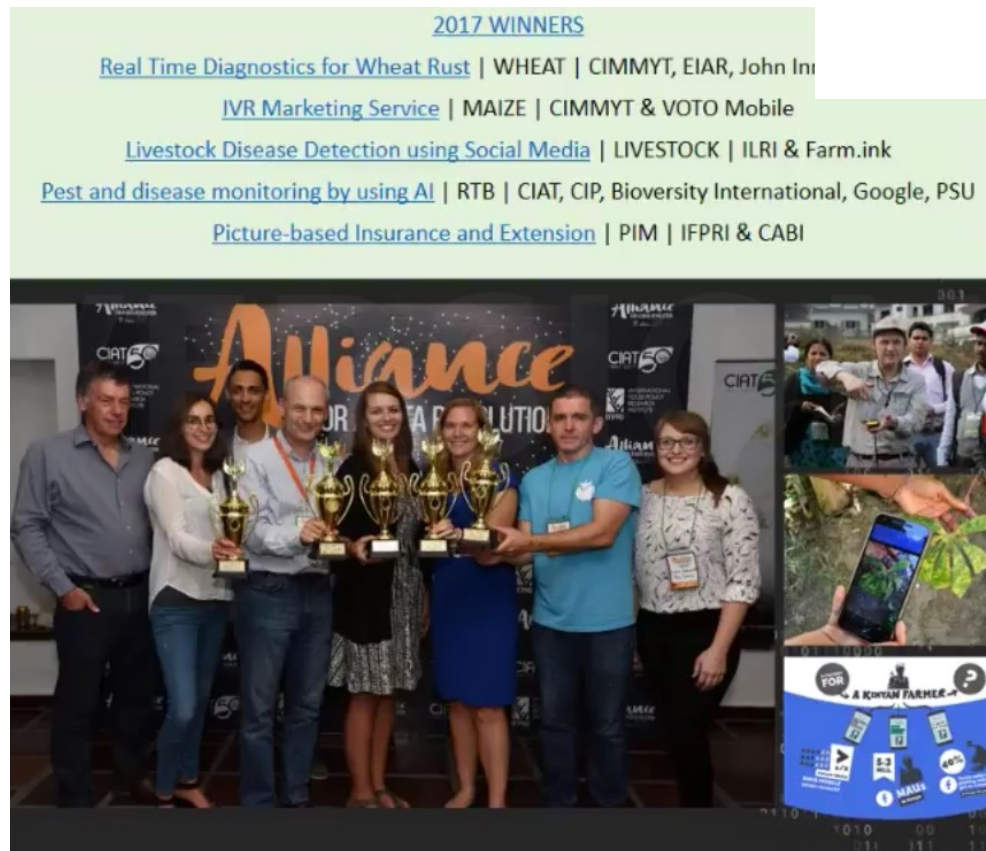
CONSUMERS

INFORMATION DEMAND

Inspire

Lead by example and inspire
how big data analytics can
monitor and deliver
development impact

- ✓ The 2017 CGIAR Inspire Challenge



Outline

1

CIP's Approach in
Open Access &
Research Data
Management

2

CIP's role in
the Big Data
Platform

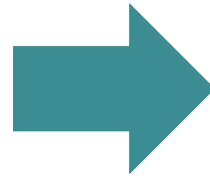
3

**Open Access
implementation
at CIP**

4

Challenges
and
Opportunities

OA Implementation



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

OA Implementation

CIP has been implementing OA for ~3 years now:

- ✓ 257 Journal Articles (67.3% OA) [2015-2018]
- ✓ 190 Datasets [2015-2018]

CIP launched the Second Open Access Competition [March 2018].

- ✓ The goal: to have ~100 datasets at the end of 2018.
- ✓ The competition is focusing mainly on data underpinning publications.

CIP generates a lot of valuable research. Our goal, in alignment with the CGIAR policies, is to be fully Open Access at the end of 2018.

Success stories

- Reached the goal for the first Data Sprint in 2017 with 125 datasets with good metadata description and vocabularies.
- OA/OD much more accepted now at CIP.
- Data Sprint considered as a good practice to incentive OA and it's being replicated in other centers.





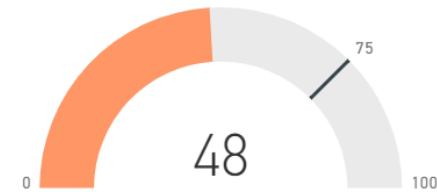
OPEN ACCESS

Competition
2018

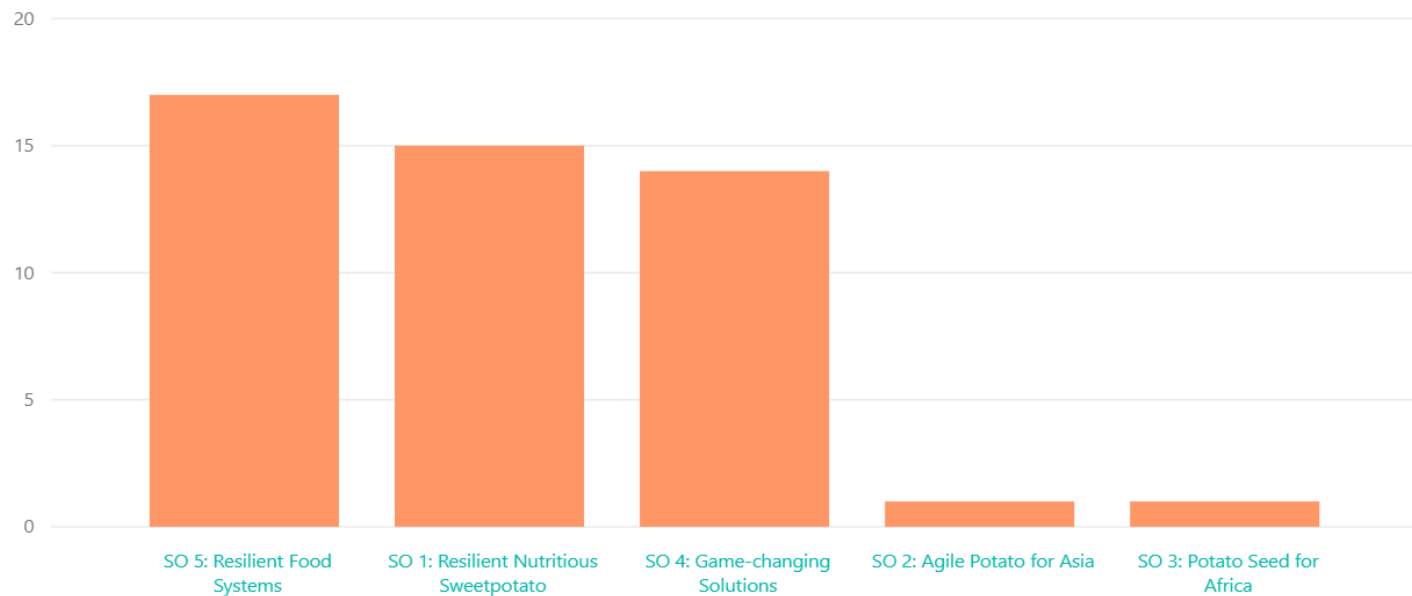


Summary of metrics from datasets published in Dataverse in 2018.

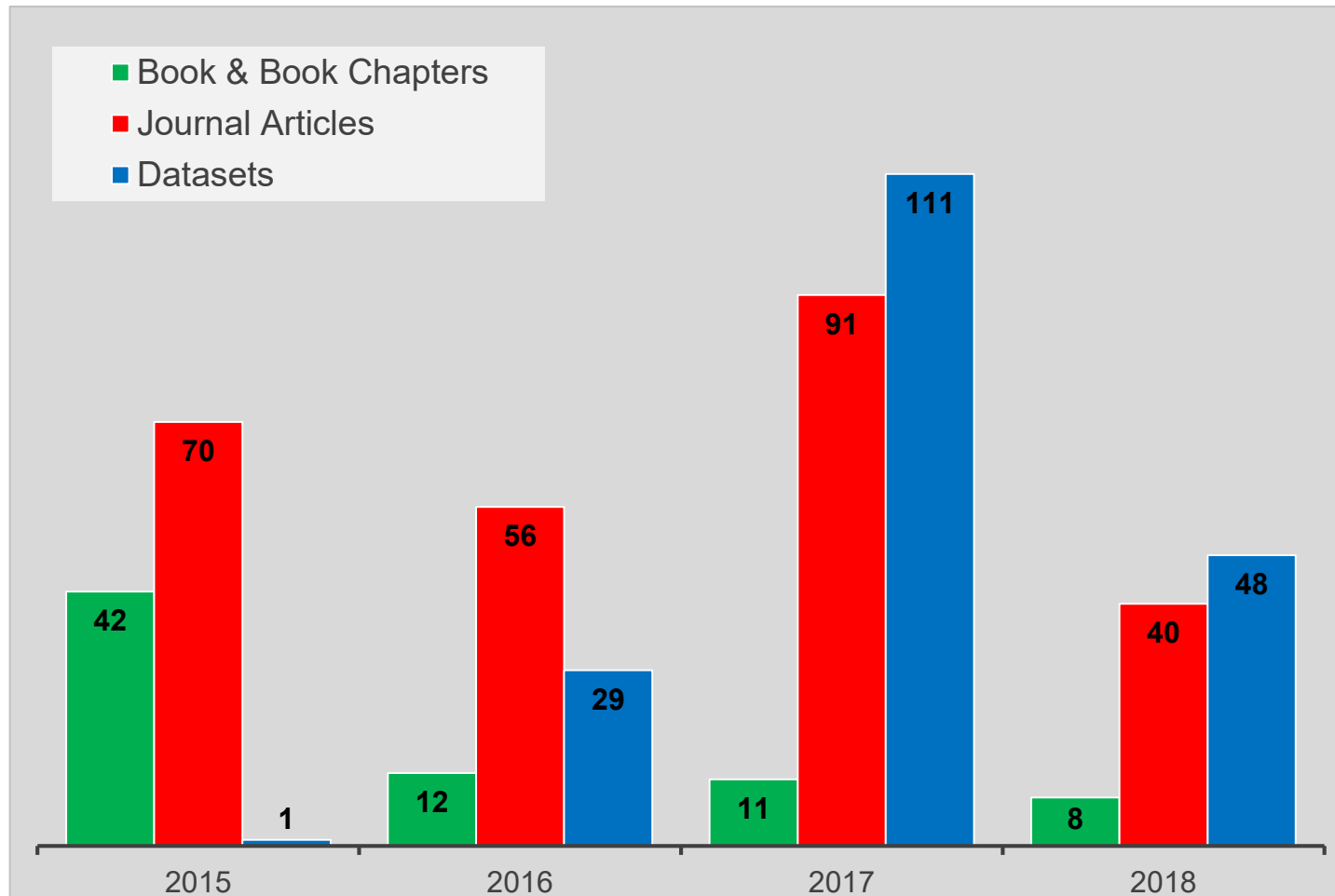
Number of Datasets



Number of Datasets by SO

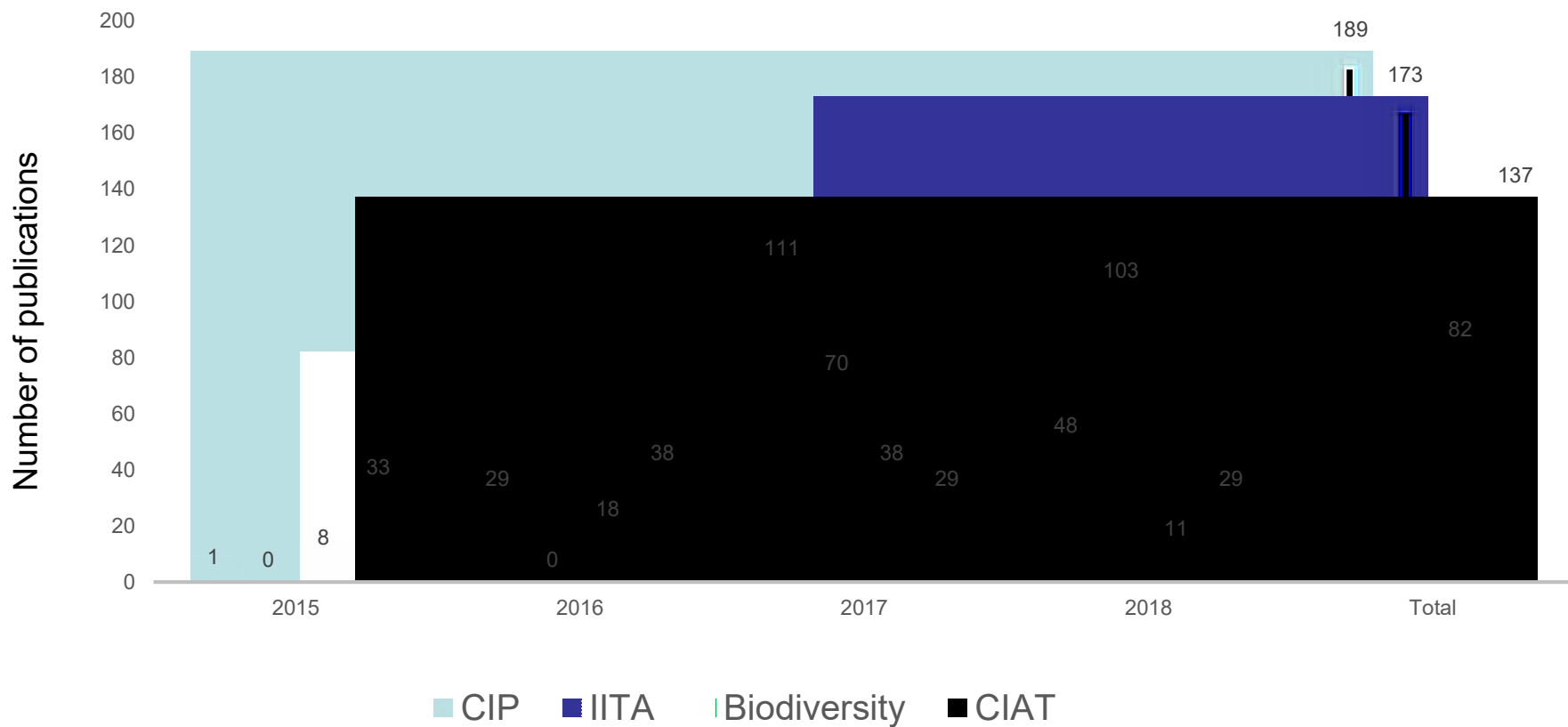


OA Implementation



- ✓ OA ratio for publications increasing at CIP (67.3% OA) and all centers.

OD implementation within our sister centers



✓ CIP is positioning in Open Data!

CIP's contribution to GARDIAN

- ✓ An integrated harmonized agricultural knowledge resources

Global discovery
Federated search
across centers



Collection of tools
Toolkit for analytics



Categorized
content type
Faceted results



Machine-readable
Human + machine
readable content



Improved access, reuse
High precision,
integration via controlled
vocabularies, ontologies

The screenshot displays the GARDIAN DATA SEARCH interface. At the top, the GARDIAN logo is visible alongside navigation links for HOME, SEARCH, and ANALYTICS. The main section is titled 'DATA SEARCH' and features a search bar with a magnifying glass icon. Below the search bar, there are filters for 'All', 'Open', and 'Restricted' content, as well as 'And' and 'Or' search options. A prominent orange banner indicates 'Search results: 96085'. The interface includes a 'FILTERS' section with dropdown menus for 'YEAR' (set to 'All'), 'CGIAR CENTER' (set to 'All'), and 'LOOK IN' (set to 'All fields'). Below the filters, there are two main categories: 'PUBLICATIONS (93709)' and 'DATASETS (2376)'. The 'PUBLICATIONS' section is further divided into 'TYPE' (set to 'All') and 'COUNTRY' (set to 'All'). A 'GEOGRAPHIC SPREAD' button with a 'VIEW MAP' link is also present. The search results list several publications, including a 2015 'Scientific Publication' titled 'SNP-Seek database of SNPs derived from 3000 rice genomes' by IRRI, a 2015 'Report / Factsheet' titled 'Synthesis report on feasibility assessment for the implementation of RRR (resource recovery and reuse) business models proposed for Lima' by IWM, a 1995 'Book / Monograph' titled 'The CIAT cassava program: Project areas, projects and resource distribution' by CIAT, a 1990 'Other' publication titled 'Water deficit and weed population density effects on rainfed lowland rice' by IRRI, and a 2012 'Scientific Publication' titled 'GGE biplot based assessment of yield stability, adaptability and mega-environment characterization for hybrid pigeonpea (Cajanus cajan)' by ICRISAT.

Outline

1

CIP's Approach in
Open Access &
Research Data
Management

2

CIP's role in
the Big Data
Platform

3

Open Access
implementation
at CIP

4

**Challenges
and
Opportunities**

Challenges and Opportunities:

- ✓ There is still a cultural change for scientists to share their data publicly. There's a need on more training to understand and implement OA specially in the regions.
- ✓ CGSpace and Dataverse repositories are constantly being updated, more publications are now linked to their datasets.
- ✓ Managing datasets and their quality require more time and resources to put datasets into CIP's standards. Scientists take longer to deliver complete datasets.
- ✓ Human Resources will include OA in Global Onboarding Project in the induction programs.
- ✓ MEL, Dataverse and CGSpace are now connected.



Overview of Progress in OA at CIP

Research Informatics Unit - RIU

