CIP
INTERNATIONAL POTATO CENTER
A CGIAR RESEARCH CENTER

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

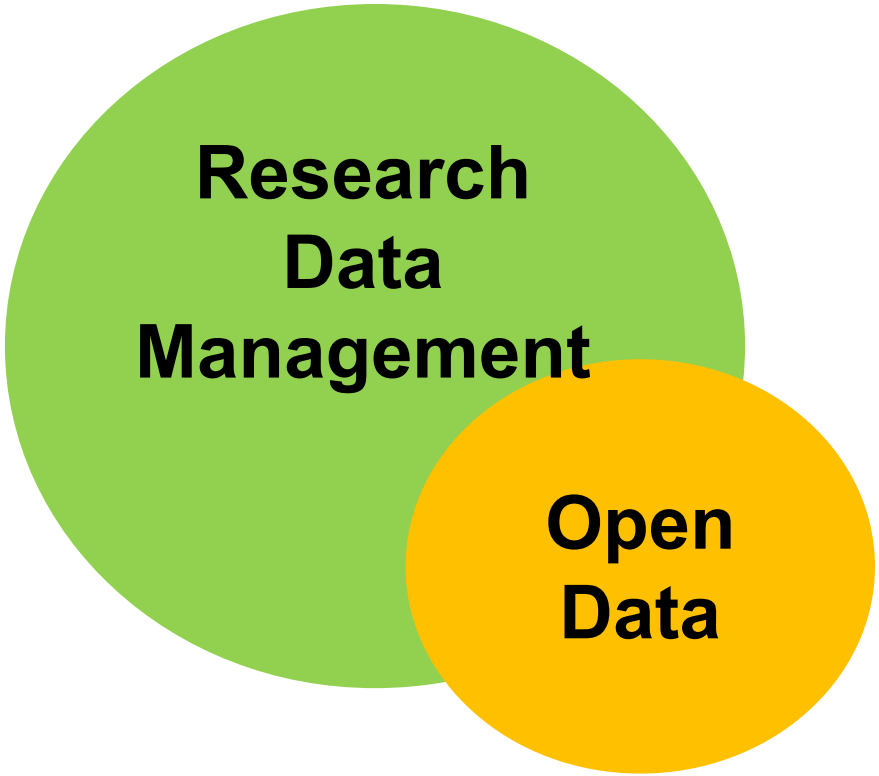
Re-usable

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)
Peer-reviewed versions of articles	As per the Policy Deposit Schedule unless OA is prohibited or subject to a longer embargo period by publisher	Ideally, at time of publication Latest: 6 months from publication
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
Data and data sets	As per Policy Deposit Schedule	As soon as possible Latest: within 12 months of completion of data collection or appropriate project milestone, or within 6 months of the publication of the information products underpinned by that data

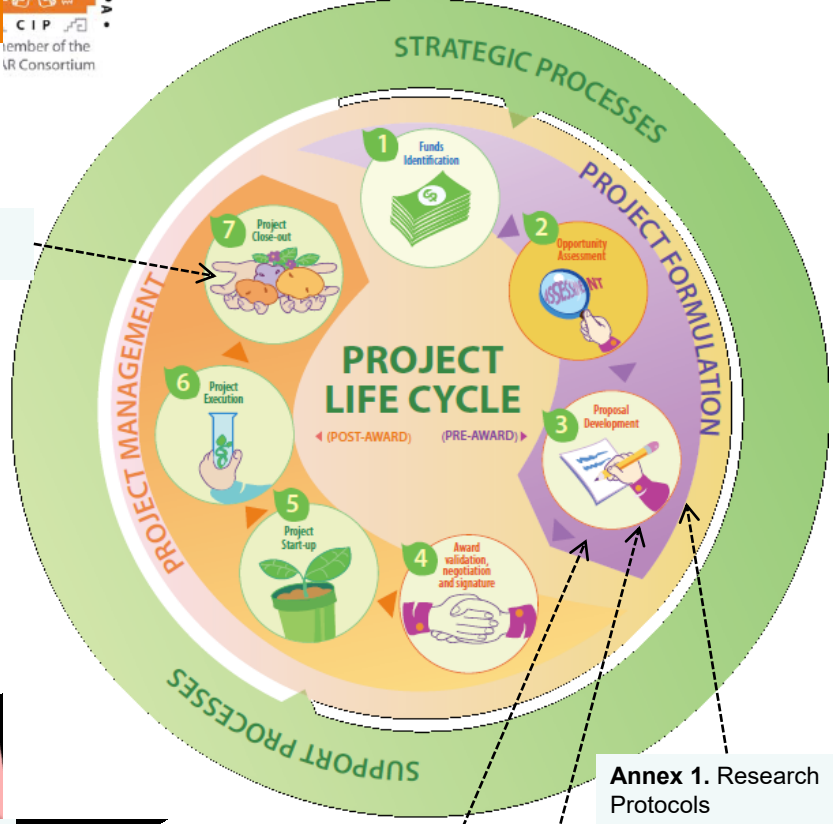
Research Data Management and Open Access



- ✓ **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



Annex 9. Data Management Checklist

I. Creating data

Annex 4. Documentation and Metadata

II. Processing data

Annex 5. Data Quality

III. Analyzing data

IV. Preserving data

V. Giving access to data

VI. Re-using data

Data Life Cycle

Annex 8. Data Repositories

Annex 7. Privacy & Confidentiality

Annex 6. Data Storage and Archiving

Annex 1. Research Protocols

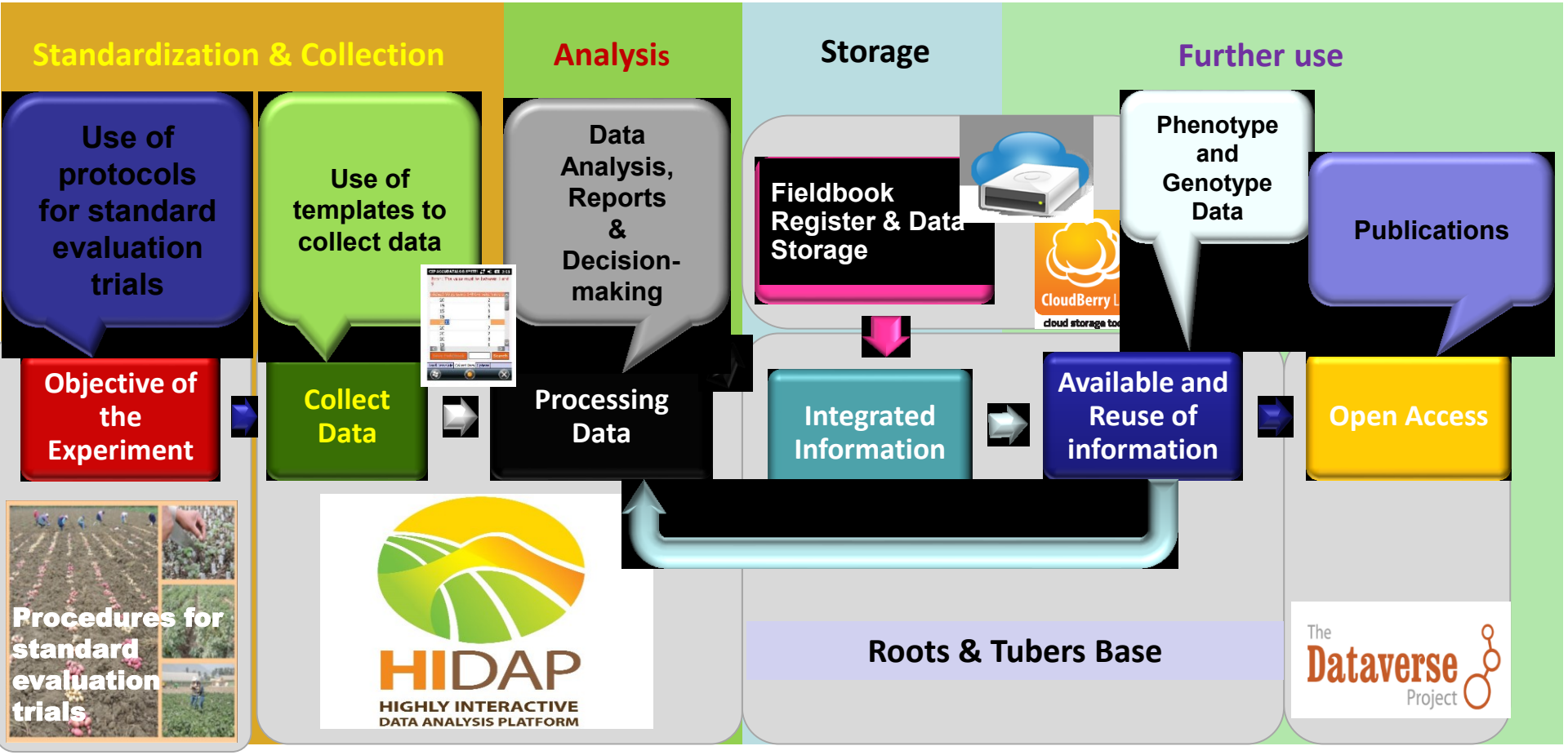
Annex 2. Data Management Plan (DMP)

Annex 3. Budgeting and Planning

…best practices and tools to help!

Workflow 1: data management of breeding information

Division: Genetics and Crop Improvement

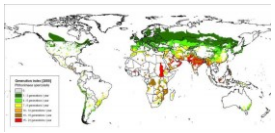


- The final dataset is published in Dataverse to make it OA as the final step in the Data Life Cycle.

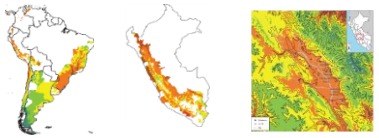
Workflow 2: Development of pest risk maps using ILCYM

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/)

Con y sin filtros de cultivos

Clima futuro: 2050

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

ERI

GI

AI

$$\mathbf{ERI} = \frac{(\sum_{i=1}^{365} \widehat{Ro}_i)/365}{max(\widehat{Ro}_i)}$$

$$\mathbf{GI} = \frac{\left(\sum_{i=1}^{365} 365 / \widehat{GL}_i\right)}{365}$$

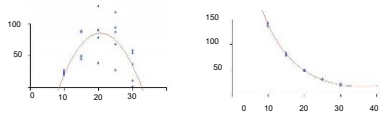
$$\mathbf{AI} = \log_{10}\left(\prod_{i=1}^{365} \hat{\lambda}_i\right)$$

Índices de riesgo

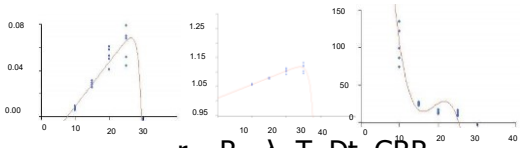
Validación del modelo

"Model builder"

Simulación de parámetros de tabla de vida

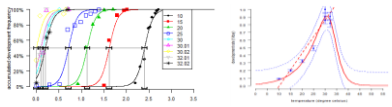


Parámetros poblacionales:

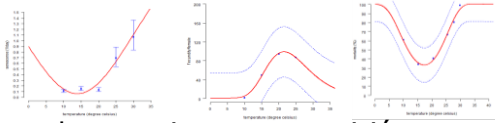


r_m, R_0, λ, T, Dt, GRR

Modelos fenológicos basados en la temperatura



Tasa y tiempo de desarrollo



Tasa de supervivencia, oviposición, mortalidad

Estudios de tabla de vida

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

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

CGIAR 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

CIP's very active!

Convene

Bring together big data practitioners, private sectos, 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

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

- ✓ 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

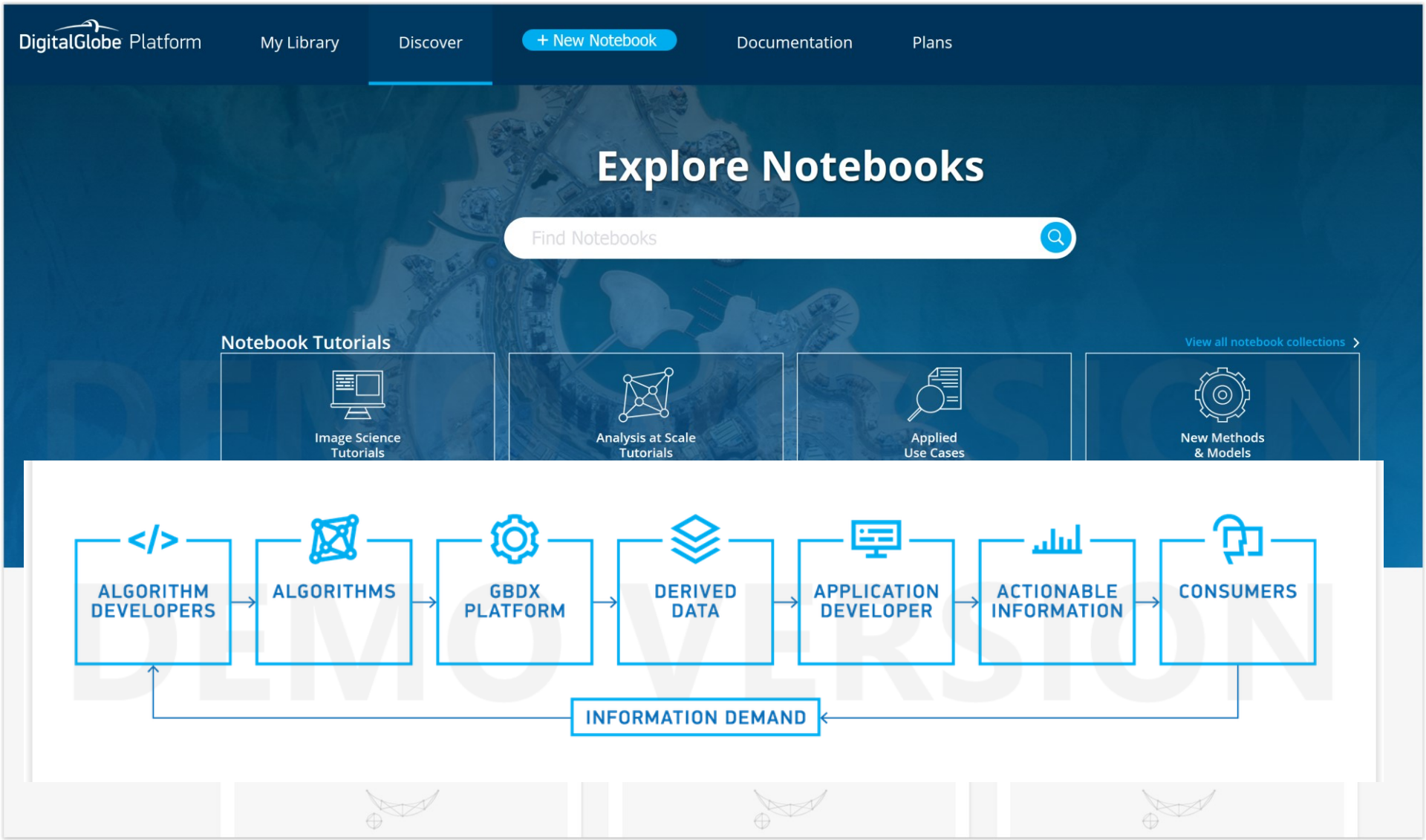
Active participation on:

- ✓ 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



Inspire

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

- ✓ The 2017 CGIAR Inspire Challenge

2017 WINNERS

Real Time Diagnostics for Wheat Rust | WHEAT | CIMMYT, EIAR, John In

IVR Marketing Service | MAIZE | CIMMYT & VOTO Mobile

Livestock Disease Detection using Social Media | LIVESTOCK | ILRI & Farm.ink

Pest and disease monitoring by using AI | RTB | CIAT, CIP, Bioversity International, Google, PSU

Picture-based Insurance and Extension | PIM | IFPRI & CABI

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

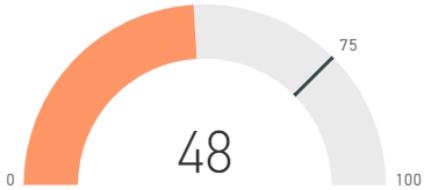
- Reached the goal for the first Data Sprint in 2017 with 125 datasets with good metadata description and vocabularies.
- OA/OD much more accepted now at CIP.
- Data Sprint considered as a good practice to incentive OA and it's being replicated in other centers.

OPEN ACCESS

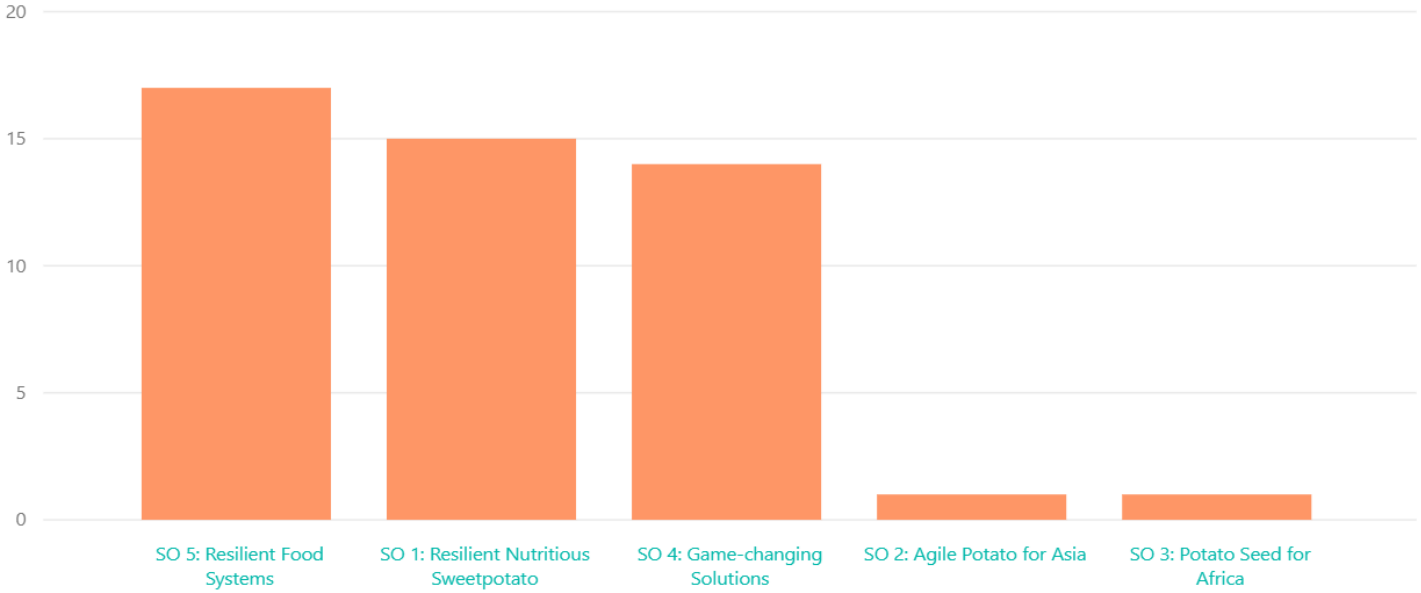
Competition 2018

Summary of metrics from datasets published in Dataverse in 2018.

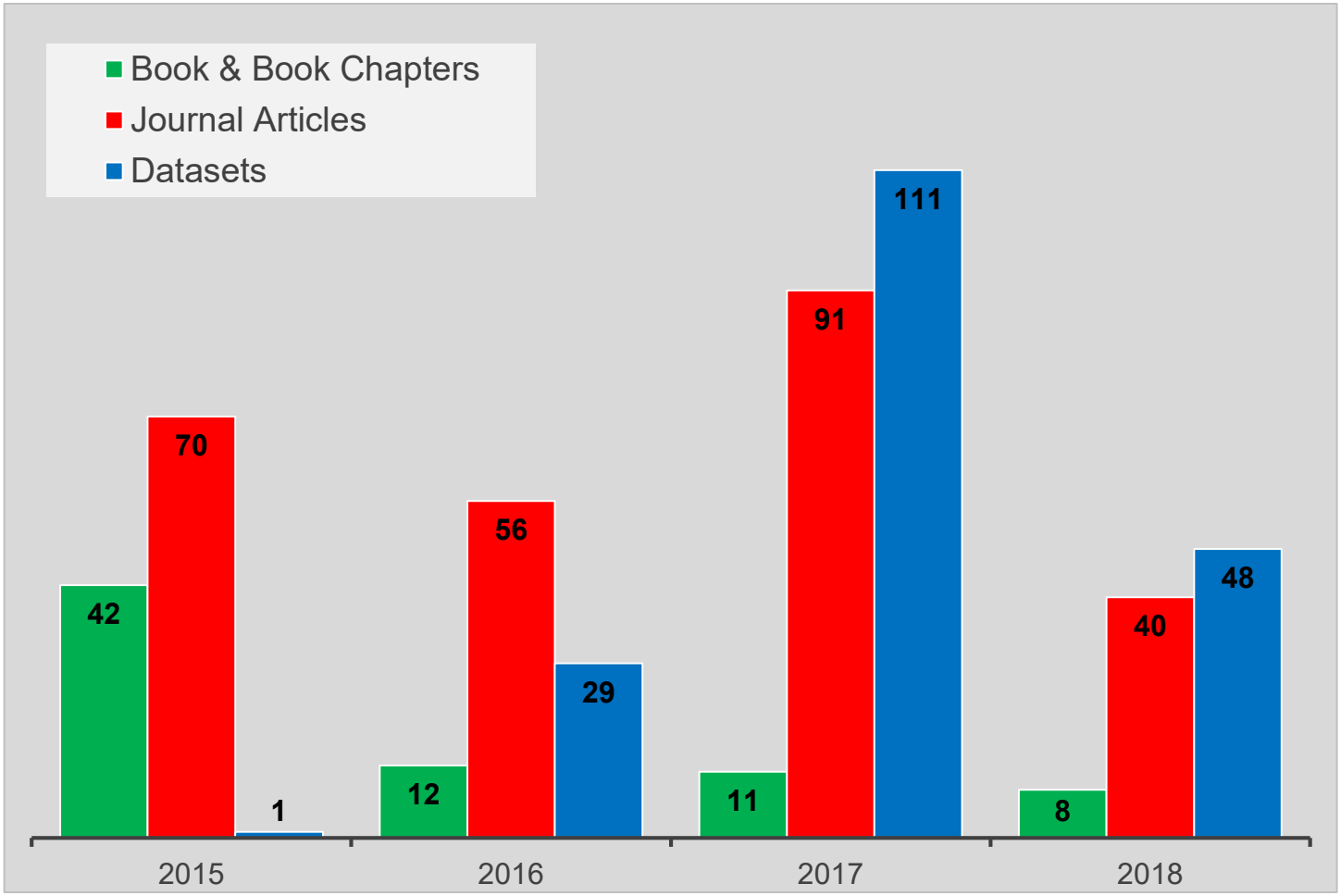
Number of Datasets



Number of Datasets by SO

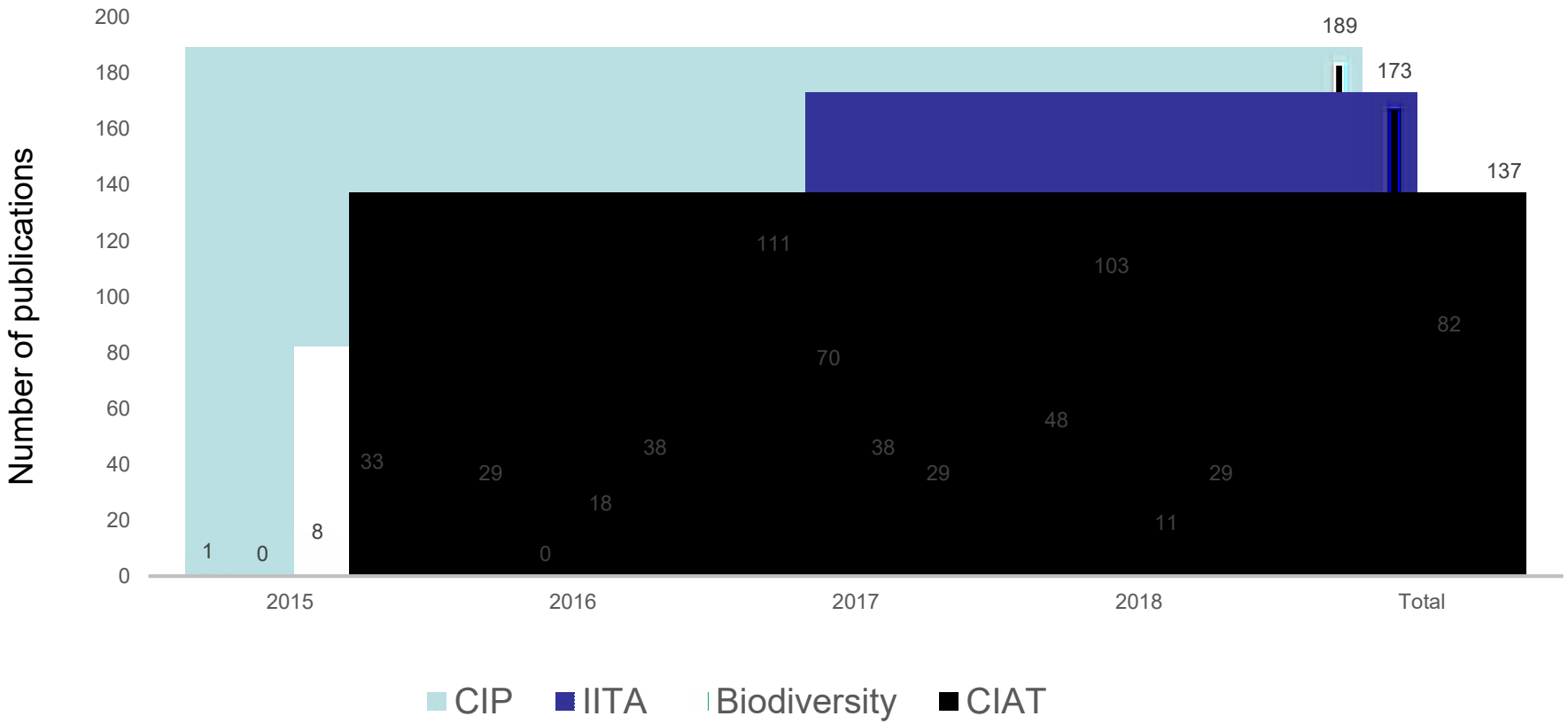


OA Implementation



✓ 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

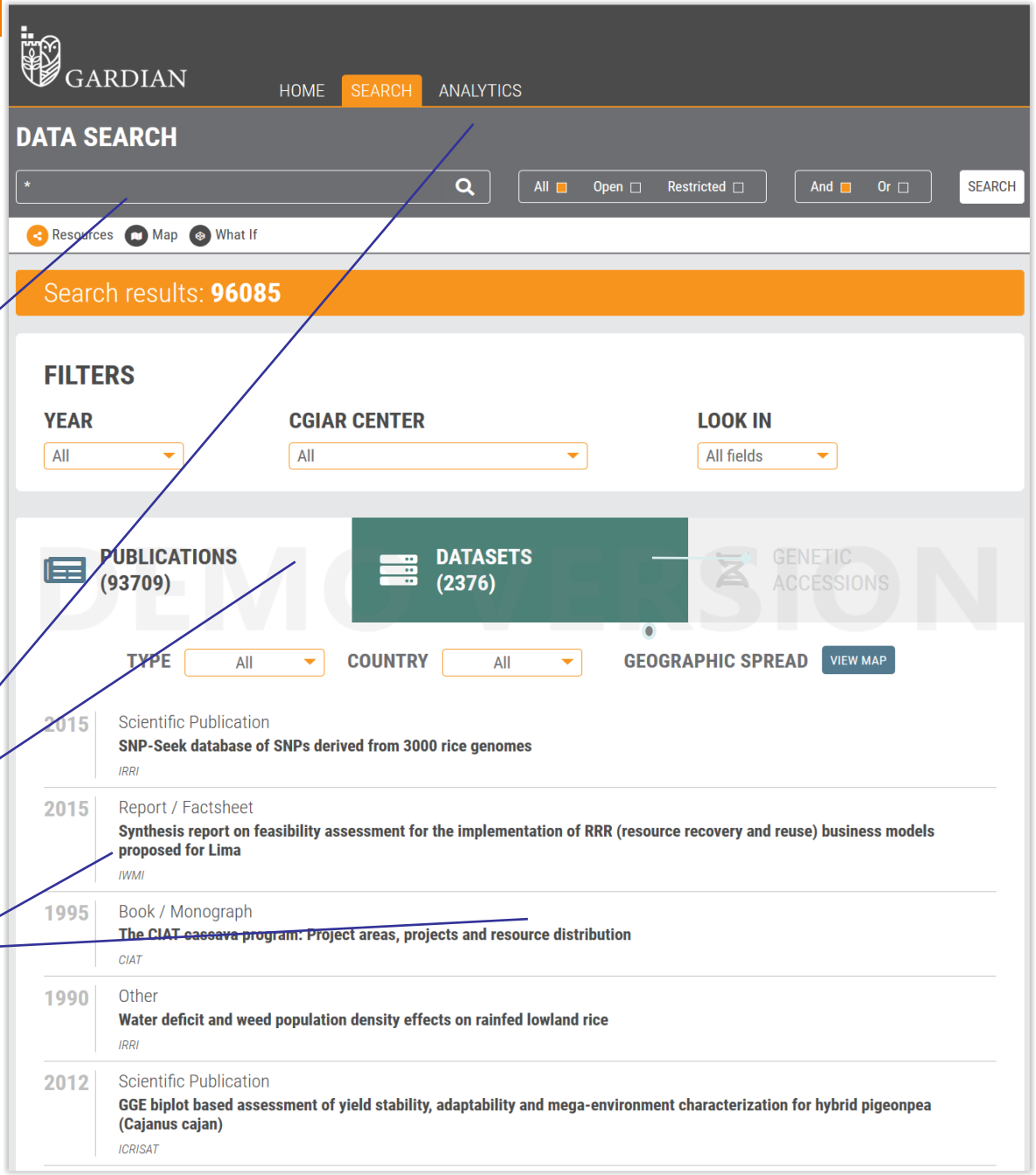
Global discovery
Federated search across centers

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



http://gardian.bigdata.cgiar.org/

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.

CIP
INTERNATIONAL POTATO CENTER
A CGIAR RESEARCH CENTER

Overview of Progress in OA at CIP

Research Informatics Unit - RIU