



Data Management Plan (DMP)

Data Repositories & Data Citation

Research Informatics Unit - RIU



Outline

1

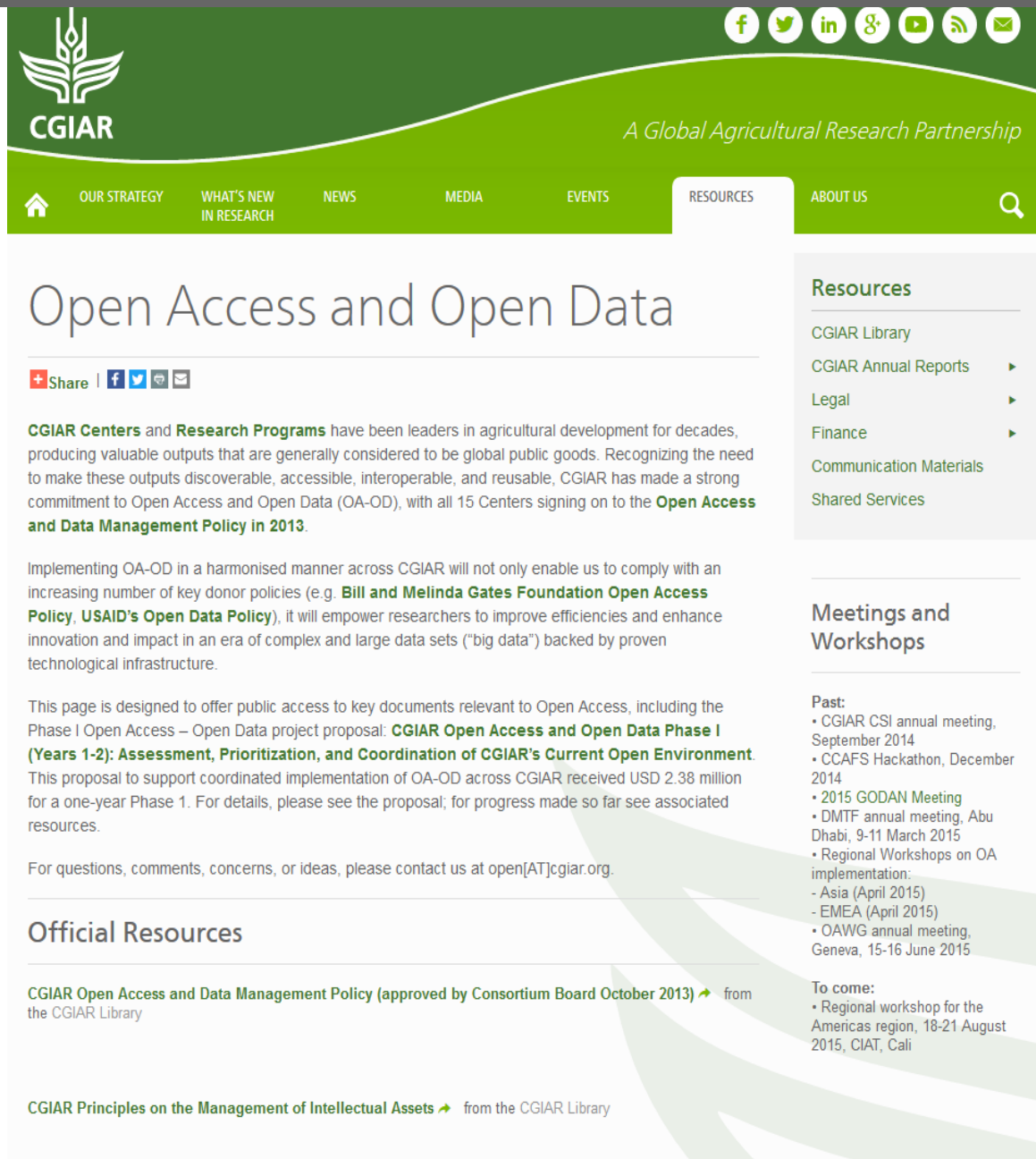
**Data Management
Plan (DMP)**

2

**Data Repositories &
Data Citation**



- According to Article 4.1.9 of the Open Access and Data Management Policy: *“Open Access and Data Management Plans should be prepared in order to ensure implementation of this Policy. Such Plans shall, in particular, outline a strategy for maximizing opportunities to make in formation products Open Access.”*



CGIAR
A Global Agricultural Research Partnership

Home OUR STRATEGY WHAT'S NEW IN RESEARCH NEWS MEDIA EVENTS RESOURCES ABOUT US

Open Access and Open Data

Share | Facebook | Twitter | LinkedIn | Email

CGIAR Centers and **Research Programs** have been leaders in agricultural development for decades, producing valuable outputs that are generally considered to be global public goods. Recognizing the need to make these outputs discoverable, accessible, interoperable, and reusable, CGIAR has made a strong commitment to Open Access and Open Data (OA-OD), with all 15 Centers signing on to the **Open Access and Data Management Policy in 2013**.

Implementing OA-OD in a harmonised manner across CGIAR will not only enable us to comply with an increasing number of key donor policies (e.g. **Bill and Melinda Gates Foundation Open Access Policy**, **USAID's Open Data Policy**), it will empower researchers to improve efficiencies and enhance innovation and impact in an era of complex and large data sets ("big data") backed by proven technological infrastructure.

This page is designed to offer public access to key documents relevant to Open Access, including the Phase I Open Access – Open Data project proposal: **CGIAR Open Access and Open Data Phase I (Years 1-2): Assessment, Prioritization, and Coordination of CGIAR's Current Open Environment**. This proposal to support coordinated implementation of OA-OD across CGIAR received USD 2.38 million for a one-year Phase 1. For details, please see the proposal; for progress made so far see associated resources.

For questions, comments, concerns, or ideas, please contact us at [open\[AT\]cgiar.org](mailto:open[AT]cgiar.org).

Official Resources

CGIAR Open Access and Data Management Policy (approved by Consortium Board October 2013) ➔ from the CGIAR Library

CGIAR Principles on the Management of Intellectual Assets ➔ from the CGIAR Library

Resources

- CGIAR Library
- CGIAR Annual Reports ➔
- Legal ➔
- Finance ➔
- Communication Materials
- Shared Services

Meetings and Workshops

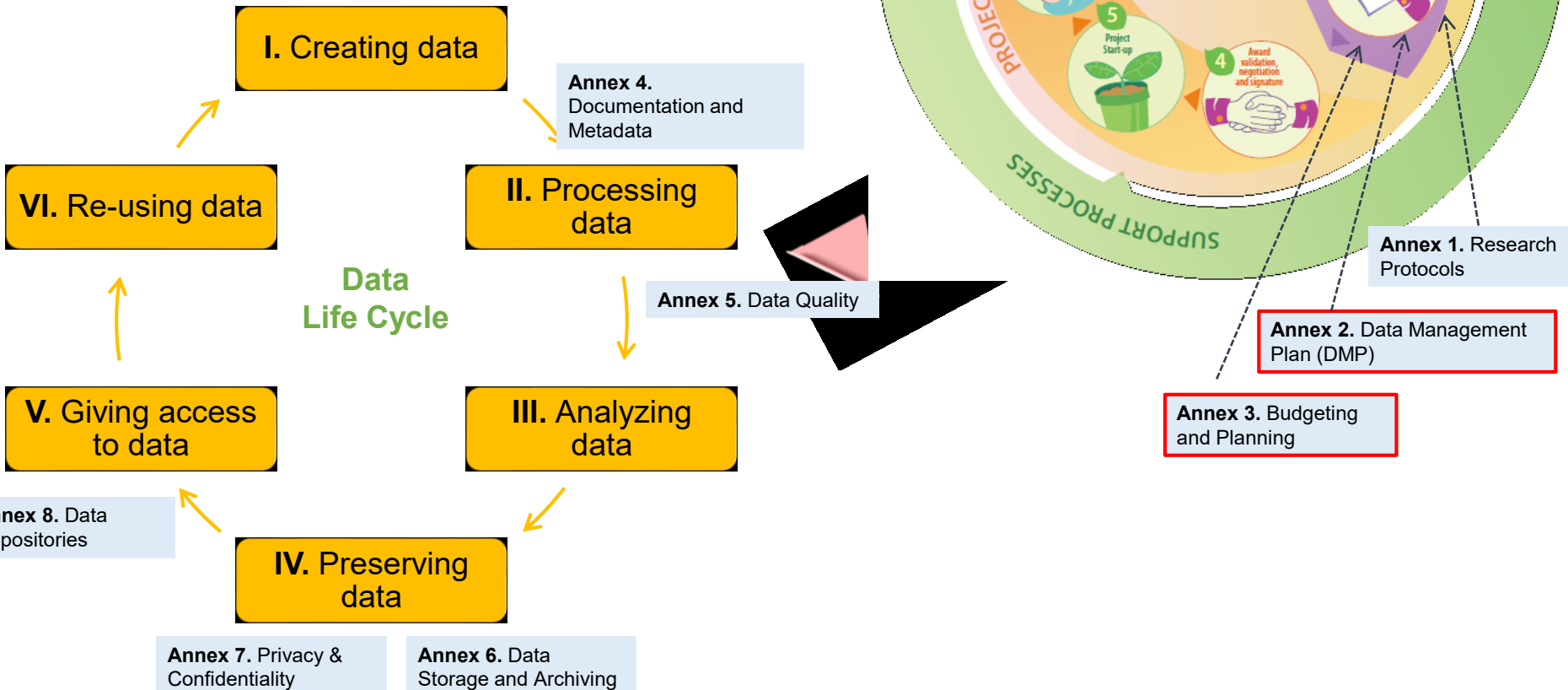
Past:

- CGIAR CSI annual meeting, September 2014
- CCAFS Hackathon, December 2014
- 2015 GODAN Meeting
- DMTF annual meeting, Abu Dhabi, 9-11 March 2015
- Regional Workshops on OA implementation:
 - Asia (April 2015)
 - EMEA (April 2015)
 - OAWG annual meeting, Geneva, 15-16 June 2015

To come:

- Regional workshop for the Americas region, 18-21 August 2015, CIAT, Cali

The Data Management Plan is the starting point in the Data Life Cycle. However the plan should be revisited often throughout the project to ensure proper data documentation and management.



What is a Data Management Plan?

- Formal document
- At project start-up meeting all new projects are required to have a DMP
- Outlines what you will do with your data during and after you complete your research
- Ensures your data is safe and open for the present and the future



- This DMP outlines what you will do with your data during and after a research project.



Data Management Plan Template

Project Title:	Sweetpotato based Job Opportunity for Youth
Sub program (if relevant):	RTB Productive varieties and quality seed
Project Lead Center:	International Potato Center
Project Investigator:	Temesgen Bocher
Individual responsible for Data Management:	Temesgen Bocher
Donor	(GIZ) GmbH
Agreement Id or cost Center	1351-GIZ0
BUS	1351-GIZ: 1391-1000:1351-1000-10



I. SUMMARY:

Describe your approach to Open Access and Data Management and any arrangements already in place to assist implementing the CIP Open Data and Data Management Policy

II. DATA AND DATABASES:

1. Describe the nature and scope of the data that will be generated under the project.
2. Describe your anticipated way of data storage and archiving that will be generated under the project.
3. Set out your plan for regular metadata capture using CG core metadata and the standards you will use and the steps you will take to ensure data quality control.

4. Set out any repositories you will use for making the data and metadata open / use CIP's



5. Description of the data to be generated and anticipated timing for making the project data and metadata available?

Output or Outcome milestone	Description of data to be generated	File types	Primary investigator for datasets	Expected research outputs type [type of datasets]	Deadline for metadata creation and data public access
Milestone #1: Project Planning Report:	o Baseline Data Collection	Dataset	Sindi Kirimi	March, 2018	December, 2018
Milestone #2: First Performance Report.	o GPS and baseline data collected from cooperatives participating in the seed trial/bulking activity.	Dataset	Sindi Kirimi	March, 2018	December, 2018

6. Explain whether any of the project data will not be made publicly available and why (e.g., ownership or access to pre-existing data; license rights to project data; personal privacy concerns; competitive advantages; data sovereignty). Describe the anticipated benefits and uses that could be achieved by making the project data available in the manner you describe above:

Not applicable for this project. All the datasets will be open.

III. VIDEO, AUDIO

Set out any repositories (finalized and in working progress) you will use for storing and sharing these information products.

Not applicable for this project.

IV. IMAGES

Set out any repositories (finalized and in working progress) you will use for storing and sharing these information products.

Not applicable for this project.

V. COMPUTER SOFTWARE:

Set out any repositories you will use for storing and sharing your software codes.

Not applicable

VI. RESOURCES AND BUDGETING:

1. Describe the anticipated total costs involved with making data widely available (if any):

Roughly 3,000 USD per manuscript.

Roughly 620 USD per dataset.



Outline

1

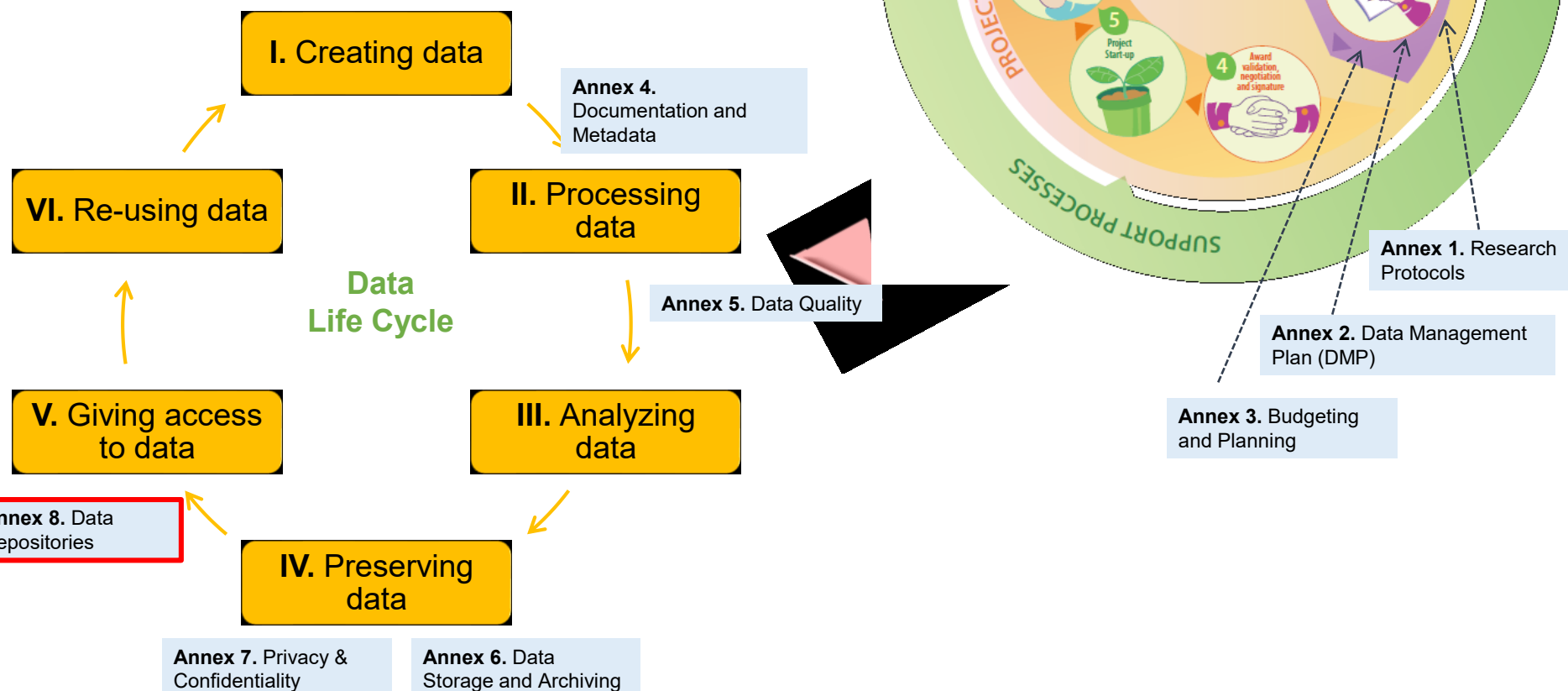
Data Management
Plan (DMP)

2

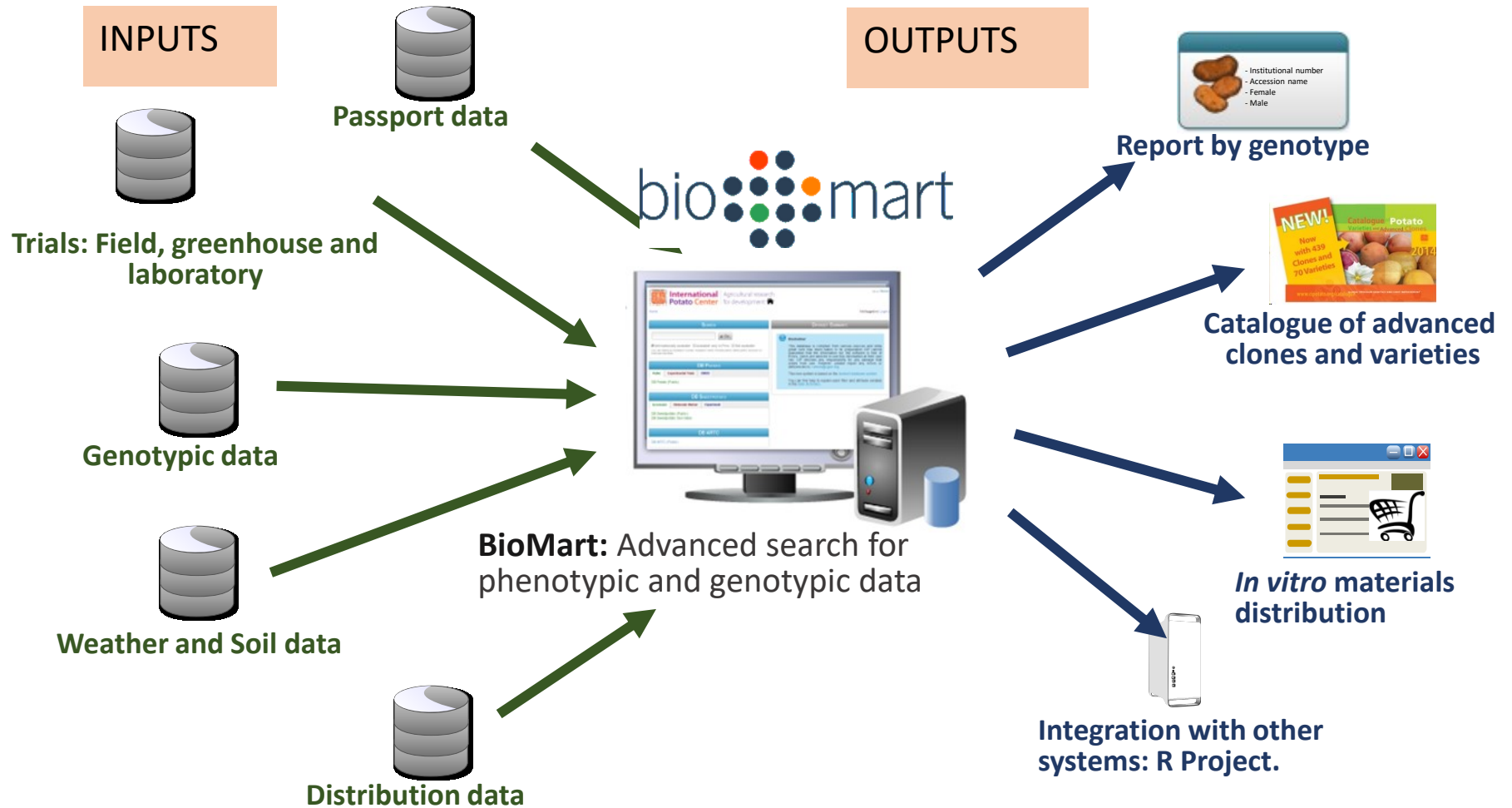
Data Repositories &
Data Citation



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



BioMart: Institutional Warehouse for Research Data



SEARCH

Go

can search by Accession number, Accession name, Female parent, or Subtaxon)

DB POTATO

Public

Experimental Trials

Mapping Population

Public Potato (Public)

DB SWEETPOTATO

Public

Experiment Trials

Molecular Marker

Experiment

Sweetpotato (Public)

Sweetpotato Gen Index

DATASET SUMMARY:

base is compiled from various sources and while has been taken in its preparation CIP cannot that the information nor the software is free of ers are advised to use this information at their own declines any responsibility for any damage that m use. However, please report any errors or s to r.simon@cgiar.org.

system is based on the [biomart database system](#).

nd help to explain each filter and attribute variable dictionary.

Username:

Password:

Login

Cancel



Findable (partially)
Accessible (partially)
Interoperable (yes)
Re-usable (partially)

The decision about CIP's Dataverse:

- ✓ Information products must be all “FAIR”
- ✓ Data must be deposited and shared via suitable repositories for all types of datasets.
- ✓ The repositories have to provide future preservation, permanent identifier (DOI / Handle) and author recognition.
- ✓ CIP has implemented a locally-hosted Dataverse installation to be compliant with CG Core and to allow for more customization.



What is Dataverse?

- Dataverse is an open source web application to share, preserve, cite, explore and analyze research data.
- Data authors, and affiliated institutions receive academic credit and web visibility.

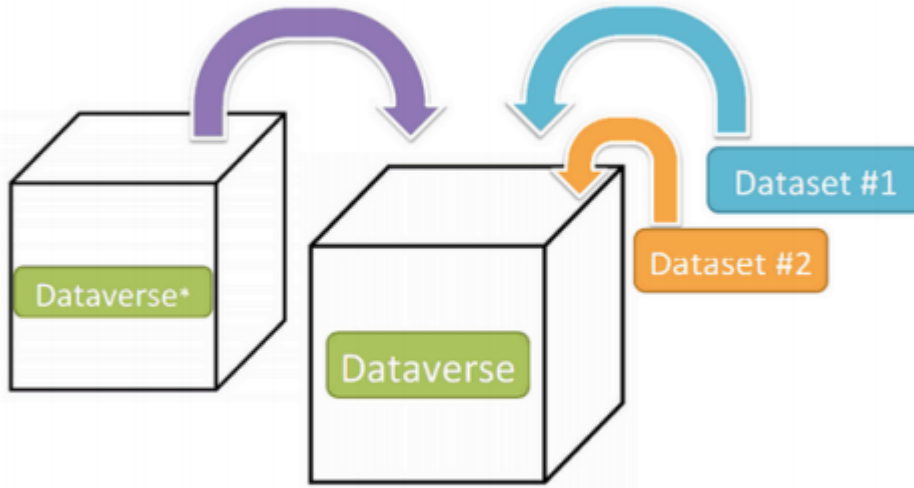
Provides incentives for researchers to share:

- Recognition & credit via data citations
- Control over data & branding
- Fulfill Data Management Plan requirements



What is a Dataverse and Dataset?

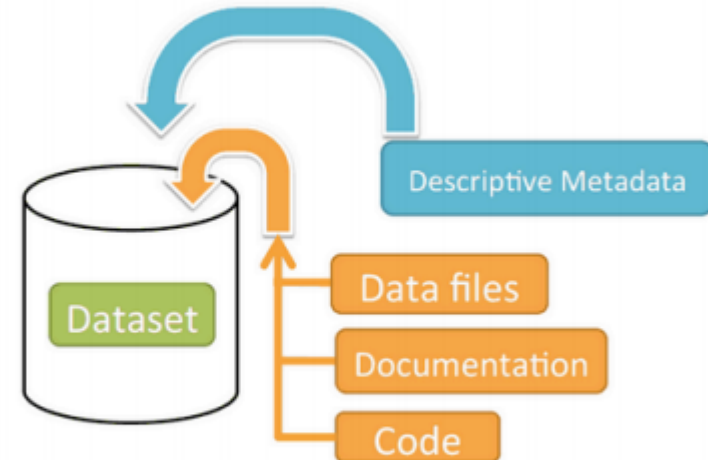
Schematic Diagram of a **Dataverse** in Dataverse 4.0



Container for your **Datasets** and/or **Dataverses***

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

Schematic Diagram of a **Dataset** in Dataverse 4.0




Container for your data, documentation, and code.

- ✓ A Dataverse is a container for Datasets, but also can contain more Dataverses.
- ✓ A Dataset is a container of Data + Metadata + Supplementary Files, etc



How Dataverse is organized at CIP?

 **Dataverse**

Q About Guides ▾ Support Sign Up Log In

CIP Research Data Dataverse (CIP)


International Potato Center Dataverse > CIP Research Data Dataverse


✉


🔗

Search this dataverse...

Q Find Advanced Search

☒  **Dataverses (3)**

☐  **Datasets (1)**

☐  **Files (5)**

Dataverse Category

Research Project (2)

Researcher (1)

Publication Date

2015 (2)

2016 (1)


Subject

Agricultural Sciences (1)


Earth and Environmental Sciences (1)


1 to 3 of 3 Results

⬆️ Sort ▾





Social and Health Science Dataverse (CIP)
Jul 6, 2016






Sweetpotato Biological Science Dataverse (CIP)
Nov 13, 2015



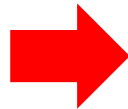
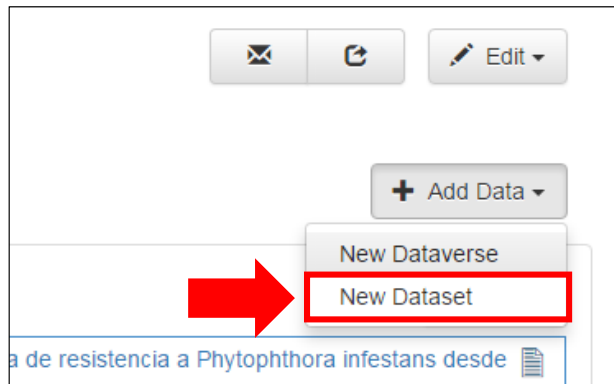


Potato Biological Science Dataverse (CIP)
Nov 13, 2015





Create a dataset



A screenshot of the 'New Dataset' form in the International Potato Center Dataverse. The form is titled 'International Potato Center Dataverse > New Dataset'. It includes a 'Host Dataverse' field set to 'International Potato Center Dataverse' and a 'Dataset Template' dropdown set to 'None'. A note states: 'Changing the template will clear any fields you may have entered data into.' Below this, a note says: '*Asterisks indicate required fields'. The form is divided into sections: 'Citation Metadata' and 'Description'. The 'Citation Metadata' section includes fields for 'Title' (with a placeholder 'Enter title...' and a button 'Add "Replication Data for" to Title'), 'Author' (with fields for 'Name' and 'Affiliation'), 'Identifier Scheme' (a dropdown), and 'Identifier' (a text field). The 'Contact' section includes fields for 'Name', 'Affiliation', and 'E-mail'. The 'Description' section includes a 'Text' field (with a note: 'This field supports only certain HTML tags.') and a 'Date' field (with a placeholder 'YYYY-MM-DD'). The 'Subject' section includes a list of categories: 'Agricultural Sciences', 'Arts and Humanities', 'Astronomy and Astrophysics', 'Business and Management', 'Chemistry', and 'Earth and Environmental Sciences'. The 'Agricultural Sciences' category is selected.

After creating a Dataset container we use the information from our CG Core Metadata template...

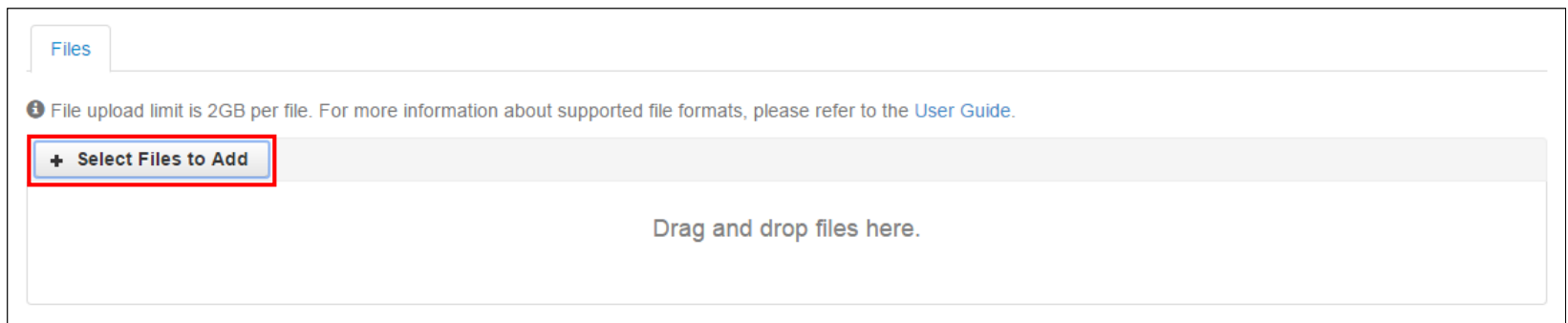


Upload files to the dataset container

Once done you need to upload the files to the dataset container. Data, data dictionaries, and supplementary files...



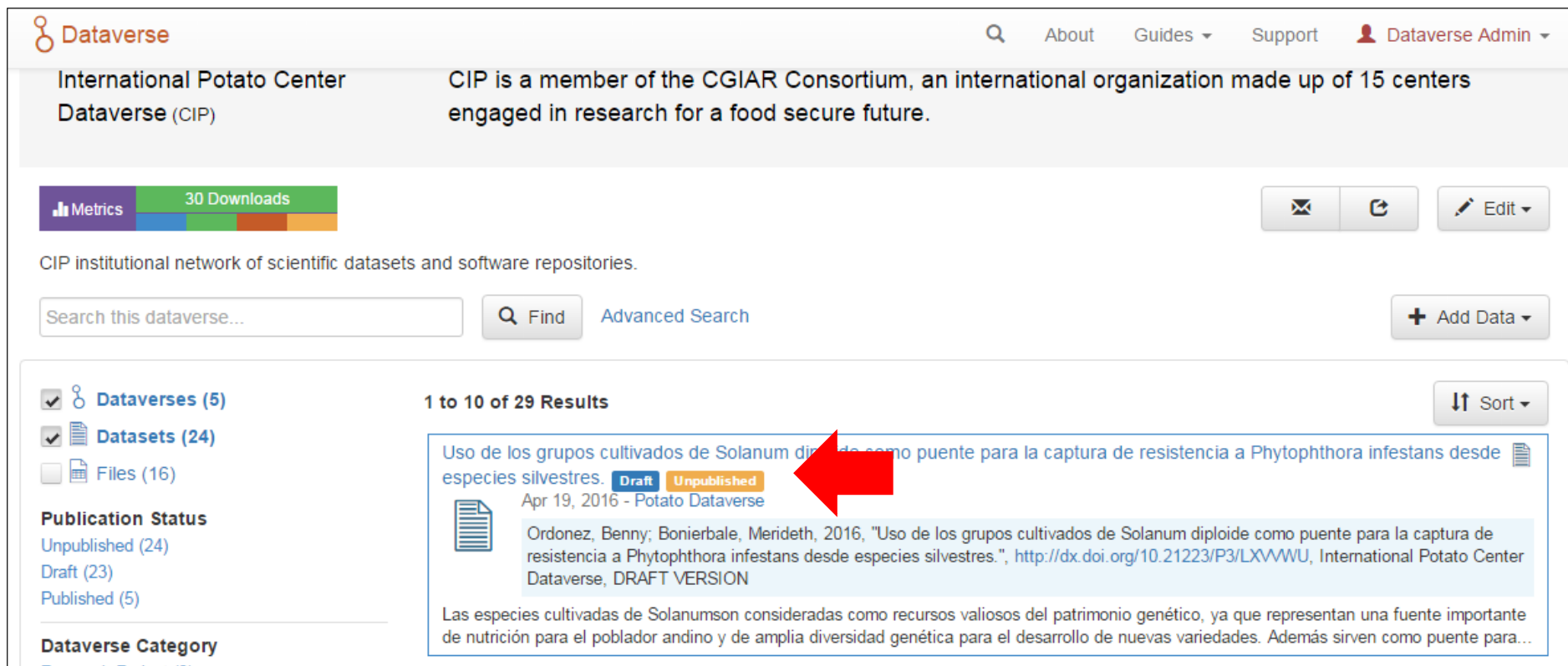
The screenshot shows the 'Files' tab of a dataset container. At the top, there are tabs for 'Files', 'Metadata', 'Terms', and 'Versions'. Below these is a search bar with the placeholder text 'Search this dataset...' and a 'Find' button. On the right side, there are two buttons: '+ Upload Files' (highlighted with a red box) and 'Edit Files'. Below the buttons, it says '0 Files' and 'There are no files in this dataset.'



The screenshot shows the 'Files' tab of a dataset container. At the top, there is a 'Files' tab. Below it, there is a message: 'File upload limit is 2GB per file. For more information about supported file formats, please refer to the [User Guide](#).' Below this message, there is a button labeled '+ Select Files to Add' (highlighted with a red box). Below the button, there is a large area with the text 'Drag and drop files here.'



Click “Save Dataset” button when you are done. Your unpublished dataset is now created.



Dataverse Search About Guides Support Dataverse Admin

International Potato Center Dataverse (CIP) CIP is a member of the CGIAR Consortium, an international organization made up of 15 centers engaged in research for a food secure future.

Metrics 30 Downloads

CIP institutional network of scientific datasets and software repositories.

Search this dataverse... Find Advanced Search + Add Data

☒ **Dataverses (5)**
☒ **Datasets (24)**
☐ **Files (16)**

Publication Status
Unpublished (24)
Draft (23)
Published (5)

Dataverse Category

1 to 10 of 29 Results Sort

Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres. Draft Unpublished
Apr 19, 2016 - Potato Dataverse

Ordóñez, Benny; Bonierbale, Merideth, 2016, "Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres.", <http://dx.doi.org/10.21223/P3/LXVWWU>, International Potato Center Dataverse, DRAFT VERSION

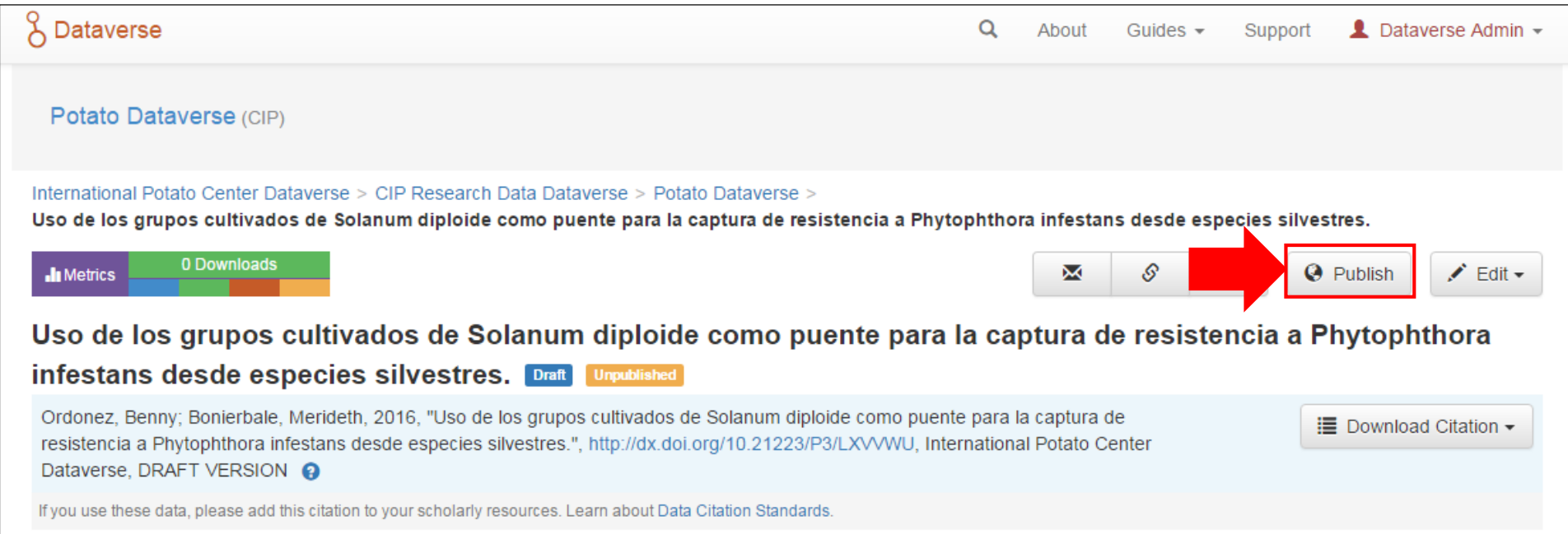
Las especies cultivadas de Solanumson consideradas como recursos valiosos del patrimonio genético, ya que representan una fuente importante de nutrición para el poblador andino y de amplia diversidad genética para el desarrollo de nuevas variedades. Además sirven como puente para...



Publish Dataset

When you publish a dataset you make it available to the public so that other users can browse or search for it.

Once a dataset is made public it can no longer be unpublished



Dataverse Search About Guides ▾ Support Dataverse Admin ▾

Potato Dataverse (CIP)

[International Potato Center Dataverse](#) > [CIP Research Data Dataverse](#) > [Potato Dataverse](#) >

Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres.

Metrics **0 Downloads**

[Email](#) [Share](#) [Publish](#) [Edit ▾](#)

Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres. **Draft** **Unpublished**

Ordonez, Benny; Bonierbale, Merideth, 2016, "Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres.", <http://dx.doi.org/10.21223/P3/LXVWU>, International Potato Center Dataverse, DRAFT VERSION [?](#)

[Download Citation ▾](#)

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)



Delete Dataset

You are able to delete your dataset as long as it is not published.

Important : A published dataset cannot be deleted.

The screenshot shows the Dataverse web interface. At the top, the Dataverse logo and navigation links (About, Guides, Support, Dataverse Admin) are visible. The main content area displays a dataset titled "Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres." The dataset is in "Draft" status and is "Unpublished". A red arrow points to the "Delete Dataset" option in the dropdown menu.

Potato Dataverse (CIP)

[International Potato Center Dataverse](#) > [CIP Research Data Dataverse](#) > [Potato Dataverse](#) >

Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres.

Metrics **0 Downloads**

[Email](#) [Link](#) [Share](#) [Publish](#) [Edit](#)

Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a P infestans desde especies silvestres. **Draft** **Unpublished**

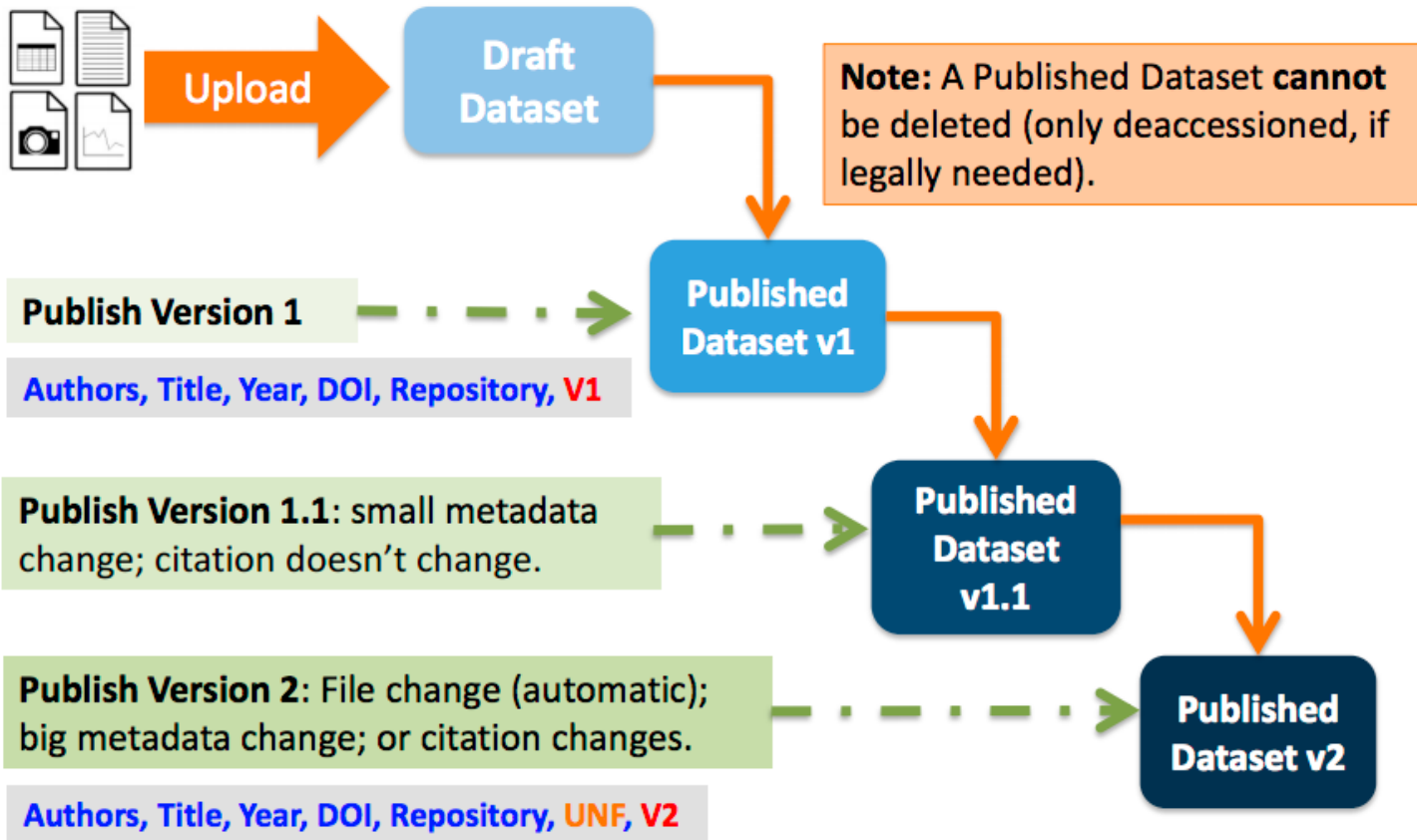
Ordenez, Benny; Bonierbale, Merideth, 2016, "Uso de los grupos cultivados de Solanum diploide como puente para la captura de resistencia a Phytophthora infestans desde especies silvestres.", <http://dx.doi.org/10.21223/P3/LXVWU>, International Potato Center Dataverse, DRAFT VERSION ?

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)

- Files (Upload)
- Metadata
- Terms
- Permissions
- Delete Dataset**



Dataset Versioning



Data citation in Dataverse

- ✓ Dataverse standardizes the citation of datasets to publish data and get credit as well as recognition for the work.
- ✓ The citation standard offers proper recognition to authors as well as permanent identification [DOI](#)



Data citation in Dataverse

1. Data belonging to publications

Andrade, Maria I.; Naico, Abdul; Ricardo, Jose; Eyzaguirre, Raul; Makunde, Godwill S.; Grüneberg, Wolfgang J.; Ortiz, Rodomiro, 2016, "Replication Data for: Genotype x environment interaction and selection for drought adaptation in sweetpotato (*Ipomoea batatas* [L.] Lam.) in Mozambique.", <http://dx.doi.org/10.21223/P3/M0HGJ4>, International Potato Center Dataverse, V1

2. Datasets from projects and legacy data

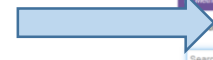
Naico, Abdul; Andrade, Maria I.; Zano, Filipe; Andre, Higino, 2016, "Dataset for: The Viable Sweetpotato Technologies in Africa (VISTA)-Mozambique: The Baseline Survey for Nampula and Zambezia Provinces", <http://dx.doi.org/10.21223/P3/NN00OC>, International Potato Center Dataverse, V1



CGSpace integration with Dataverse

The screenshot shows the CGSpace website with the International Potato Center (CIP) as the selected community. The header includes the CGSpace logo and a search bar. The main content area features a search bar for the community and its collections. A sidebar on the right lists various categories like 'Mi cuenta', 'Descubre', and 'Autor'. The bottom section displays a list of collections available in the community, such as 'CIP Book Chapters', 'CIP Books', and 'CIP Briefs'.

Citation to data



Citation to article

The screenshot shows the Dataverse website with the International Potato Center (CIP) dataset. The header includes the Dataverse logo and navigation links. The main content area displays the dataset title, description, and a search bar. A sidebar on the right lists various categories like 'Mi cuenta', 'Descubre', and 'Autor'. The bottom section displays a list of datasets available in the community, such as 'CIP Book Chapters', 'CIP Books', and 'CIP Briefs'.

Track # of article downloads and # of data sets downloads.



CIP Open Access Launch <http://cipotato.org/open-access/>



Data Management Plan (DMP)

Data Repositories & Data Citation

Research Informatics Unit - RIU

