International Potato Center

Documentation and Metadata

Research Informatics Unit - RIU

OPEN ACCESS

1 2 3 4

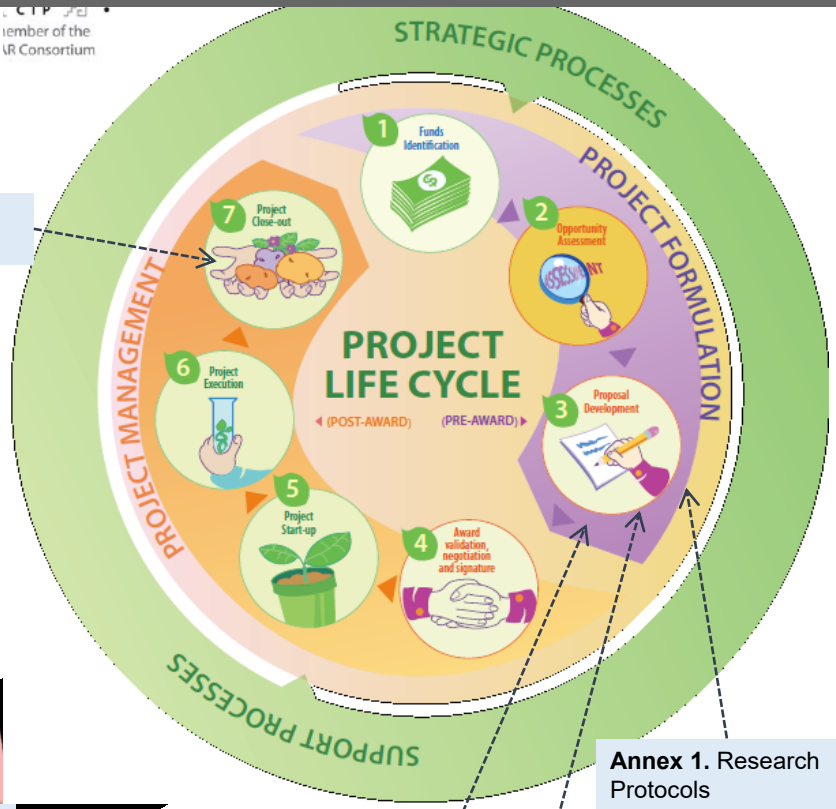
Introduction to Documentation

Metadata for Datasets

The Data Dictionary

Crop Ontology

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



Annex 9. Data Management Checklist

Annex 1. Research Protocols

Annex 2. Data Management Plan (DMP)

Annex 3. Budgeting and Planning

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

…there are best practices and tools to help!

Definition of Datasets

A collection of data and associated files.

What is Metadata?

- ✓ Metadata is data about data. It describes the content, quality, condition, and other characteristics of a dataset.

What is a data dictionary?

- ✓ Data dictionary is the definition of all elements or variables of a dataset(s)

OPEN ACCESS

1 Introduction to Documentation

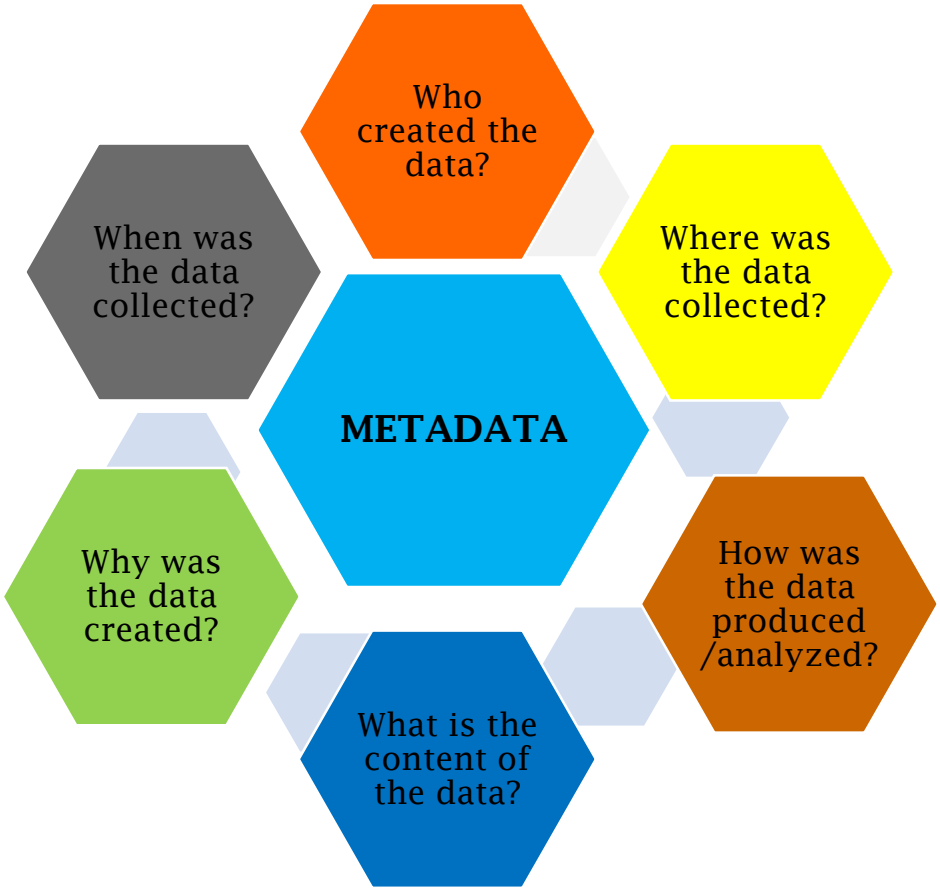
2 **Metadata for Datasets**

3 The Data Dictionary

4 Crop Ontology

Metadata

Metadata is data about data. It describes the content, quality, condition, and other characteristics of a dataset.



Metadata…

- ✓ CIP has adopted the CG Core Metadata. See the Annex 4 from Guidelines and procedures.
- ✓ The CIP's Template has the CG Core & Dataverse Metadata.

1 Metadata for common Datasets

Fieldbook1.xls

Fieldbook2.xls

Fieldbook3.xls

PTLB200109_OXAPMP_B3C102-17
PTLB200209_OXAPMP_B3C103-05
PTLB200409_OXAPMP_B3C105-03
PTLB200509_OXAPMP_B3C106-02
PTLB200112_VIENA_B3C1COM02-17
PTLB200212_VIENA_B3C1COM03-05
PTLB200412_VIENA_B3C1COM05-03
PTLB200512_VIENA_B3C1COM06-02

Dataset for: Stability of resistance and yield of 15 clones advanced clones across environments in Peru

1 Metadata for each Dataset

Fieldbook3.xls

What should be deposited

1. All research data belonging to publications
2. Datasets from projects
3. Legacy data? Depends in their value and resources to make it open.

Data underpinning Journal Articles

- ✓ Submit journal article for review.

- ✓ Article approved and published.
- ✓ Submit data to CIP Dataverse.

- ✓ Data published on CIP Dataverse and includes a reference to the article.

Data underpinning Journal Articles (Prior Publication)

- ✓ Submit data to CIP's Dataverse.
- ✓ Data is not publicly published.

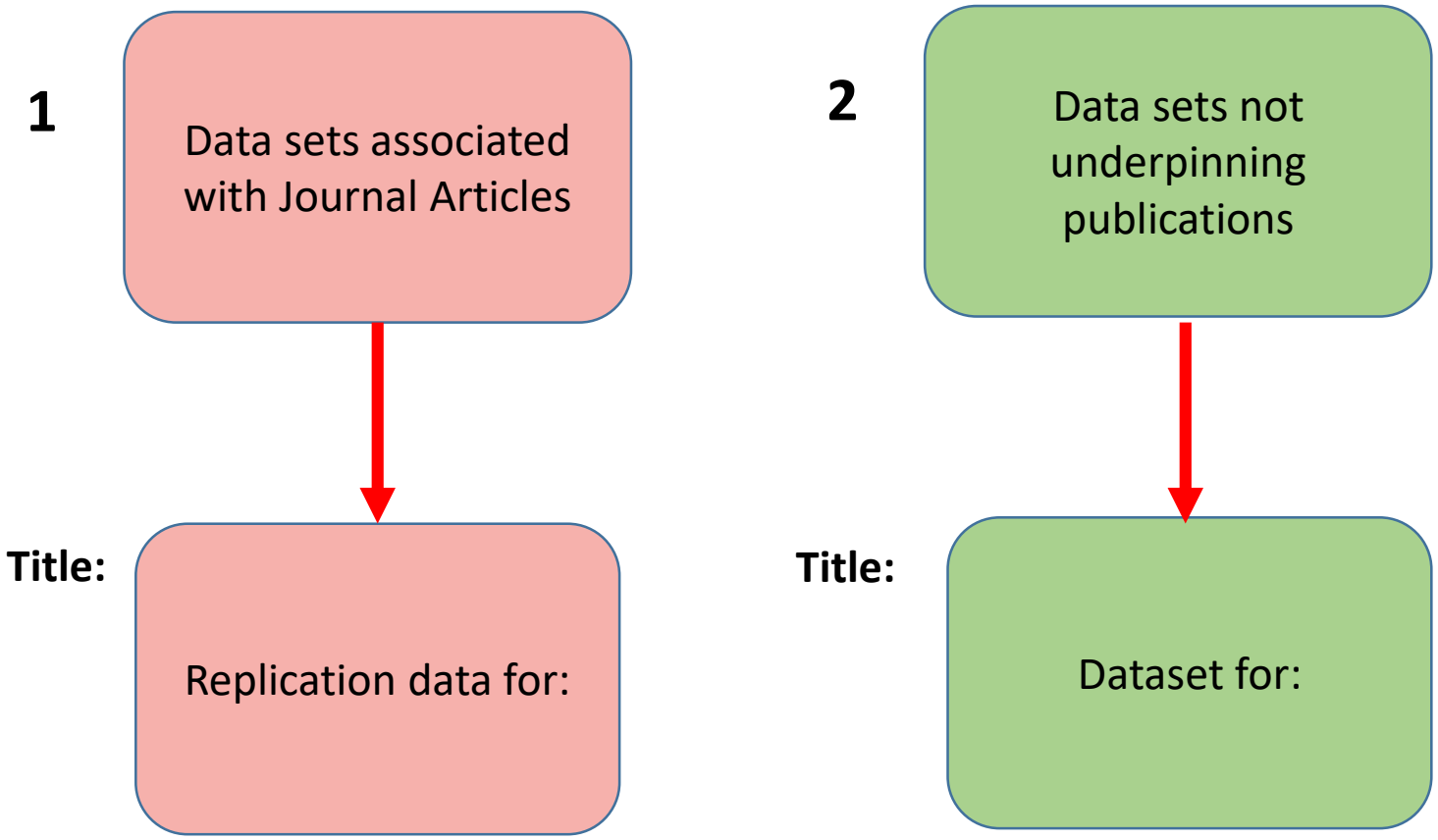
- ✓ Submit article for review.
- ✓ Submit data for review via a private link.

- ✓ Article published by journal.
- ✓ Data published by CIP.
- ✓ Article references the dataset.

Data under a research and/or development project

- ✓ Submit data to CIP Dataverse.

- ✓ Data published on CIP Dataverse and includes a permanent link for the dataset.


1
Data sets associated with Journal Articles
Title:
Replication data for:
2
Data sets not underpinning publications
Title:
Dataset for:

Metadata Schema Template

I	II	III
General Metadata (Mandatory)	Journal Metadata (Mandatory if the Dataset is "Replication Dataset for an Article")	Geospatial Metadata (Not Mandatory)

Metadata Schema Template

Title	**Title**	**dc**	Official or unofficial title of the document. If comes from publication the title must start "**Replication data for**" , if not "**Dataset for**"
			Replication Data for: Chlorophyll concentration in leaves is an indicator of potato tuber yield in water-shortage conditions
Author		**dc**	Creators of the item typically a person. Could be an organization in case of corporate authors (e.g. Center reports)
			Lindqvist-Kreuze, Hannele (International Potato Center) CIP
			Ramirez, David (International Potato Center) CIP Yactayo, Wendy (International Potato Center) CIP, Gutierrez, Raymundo (International Potato Center) CIP Mares, Victor (International Potato Center) CIP, De Mendiburu, Felipe (International Potato Center) CIP Posadas, Adolfo (International Potato Center) CIP Quiroz, Roberto (International Potato Center) CIP
Contact:		**dc**	The contact(s) for this Dataset.
	Name:		Ramirez, David (CIP)
	Affiliation:		International Potato Center
	E-mail:		d.ramirez@cgiar.org
Description	**Description**	**dc**	Abstract or longer description of the item (Description of dataset).
			The phenotyping of secondary characters is a common practice in breeding programs aiming at finding physiological mechanism related to drought tolerance. However, the dynamics of these characters depend on crop phenology, levels of water shortage, and other factors that affect their relationship with yield and limit their capacity to be used for predictive purposes……
Subject	**Subject**	**dc**	Subject matter of the research, technologies tested
			Agricultural Sciences

Keyword		**dc**	Key terms that describe important aspects of the Dataset
			Drought (AGROVOC) http://aims.fao.org/aos/agrovoc/c_926
			Solanum tuberosum (AGROVOC)
			Senescence (AGROVOC)
			Deficit irrigation
Crop	**Crop**	**cc**	Cultivated plants or agricultural produce, such as tuber, roots, grain, vegetables, or fruit, considered as a group, allows the registration of crop used in the experiment or bioassay in laboratory, greenhouse or field. (E.g. Sweetpotato, potato, ullucus, etc.)
			Potato
Citation		**dc**	Publications that use the data from this Dataset.
	Citation		Ramirez, D.A. (CIP); Yactayo, W. (CIP); Gutierrez, R. (CIP); Mares, V. (CIP); Mendiburu, F. de(CIP); Posadas, A. (CIP); Quiroz, R. (CIP). 2014. Chlorophyll concentration in leaves is an indicator of potato tuber yield in water-shortage conditions. Scientia Horticulturae. (Netherlands). ISSN 0304-4238. 168:202-209. (AN=77785) . REP.21080.
	ID Type		doi
	URL		https://cgspace.cgiar.org/handle/10568/64888

http://aims.fao.org/standards/agrovoc/functionalities/search

Contributor		**dc**	**Person, organization, or service making contributions to resource content; CGIAR affiliation**
	Center	cg	
	CRP	cg	Roots, Tubers and Bananas (RTB)
	Funder	cg	Research Program on Water, Land and Ecosystems
	Project	cg	
	Researcher	cg	
	Research Group	cg	

Time Period Covered		dc	Time period to which the data refer. This item reflects the time period covered by the data.
	Start		2008-09-01
	End		2009-11-30
Embargo end date		cg	Used when an item has an embargo by publisher (ex: 6 or 12-month embargo): YYYY-MM-DD
Format		dc	File format of item e.g.: PDF; jpg etc.
			xls
Related Material		dc	Any material related to this Dataset
Software		dc	Information about the software used to generate the Dataset.
			SAS, Version: v8.02
			R, Version: v3.0.1

II Journal Metadata

Dataset is "**Replication Dataset for** an Article")

Journal Metadata
Name:
Volume:
Issue:
Publication Date:
Type of Article:

III Geospatial Metadata

if you need to **register** 1 to 3 localities can register the data for each locality

North Latitude
South Latitude
East longitude
West Longitude

http://www.geonames.org/

Region
Country
Admin1
Admin2
Admin3
Locality

if you need to **register more the** 3 localities create a **Geographic Bounding Box** to geographical coordinates (GIS)

OPEN ACCESS

1 Introduction to Documentation

2 Metadata for Datasets

3 **The Data Dictionary**

4 Crop Ontology

What is a data dictionary?

- ✓ Data dictionary is the definition of all elements or variables of a dataset(s)
- ✓ Data dictionary indicates if the description, units, methods and the reference
- ✓ Data dictionary is used to describe and facilitate the documenting.

DATA DICTIONARY TEMPLATE

Variable name	Abbreviation	Description	Method	Formula	Unit	Reference

DATA DICTIONARY:

Variable synonyms	Trait	Trait description	Method description	Formula	Scale name/unit
RWC	Relative Water Content	Measure of plant water status evaluated on a leaflet from an expanded and sun-exposed leaf located in the upper third section of the plant canopy	The collected leaflets were weighed (fresh weight, LFW), then immersed in distilled water during 6 h at 6 ◦C and weighted again (saturated weight, SW) and finally, laflets were dried during 48 h at 80 ◦C and weighted (dry weight, LDW).	RWC=((LFW−LDW)×100/(LSW−LDW))	Percentage (%)
Chl_{SPAD}	Chlorophyll Content	Chlorophyll content estimation using a portable chlorophyll meter (SPAD-502 model, Konica Minolta)	Nine readings are averaged per leaf in one plant. The leaf must be expanded and sun-exposed, located in the upper third section of the plant canopy.		SPAD
OP	Osmotic potential	Involves the rupture of the cell membrane brought about by an abrupt leaf defrosting which causes that water potential equals OP	A leaflet circular sample (0.5 cm of diameter) is immersed in liquid nitrogen and after that conserved at −80 ◦C. The water potential is measured in defrosted samples using a dew point potentiometer (Wescor, Logan, UT, USA).		Mpa (Mega Pascales)
TY	Tuber Dry Biomass	Total dry tuber yield biomass per hectare (field experiment) and tuber dry biomass per plant (greenhouse experiment)	Tuber dry biomass per area based on the harvest of the three central rows of the plots (fiel experiment) and tuber dry biomass per plant at each pot (greenhouse experiment)		$t\ ha^{-2}$ / $g\ plant^{-1}$
CC	Canopy Cover	Percentage of the total ground area covered by the vertical projection of the canopy of the plant	Average of the proportion of vertices of a nylon grid (0.05 m × 0.05 m) intercepting leaf surface in a 0.8 m × 0.6 m frame		%
TT	Thermal Time	Summation of cumulative differences between daily mean temperature and a specified base temperature	Mean of maximum and minimum temperatures least base temperature (temperature at which development stops due to cold)	TT= ((Tmax −Tmin)/2)− Tbase	◦C day

OPEN ACCESS

1 Introduction to Documentation

2 Metadata for Datasets

3 The Data Dictionary

4 **Crop Ontology**

Ontology

- ✓ Ontology is a standard terminology used to describe crop development and agronomic traits for phenotypic information.
- ✓ The Crop Ontology (CO) http://cropontology.org/) is developed for the Integrated Breeding Platform (IBP) by several centers (bioversity, CIMMYT, CIP, ICRISAT, IITA, and IRRI).

Crop Ontology
for agricultural data

Crop Ontology Curation Tool

Home About Feedback

Crop Ontology for agricultural data

Integrated Breeding Platform
Today's tools for tomorrow's crops

Potato Ontology

Ontology curators

- Reinhard Simon, Integrated IT & Computational Research (IITCR) leader, CIP

Scientists

- Vilma Hualla Mamani, Research assistant (IITCR), CIP
- Elisa Salas Murrugarra, Research assistant, CIP
- Rene Gomez, potato genebank curator, IFPRI
- Raul Cordova, IT Analyst (IITCR), CIP
- Stef de Haan, Agroecologist, CIP
- Merideth Bonierbale, Genetics and Crop Improvement Global Program leader, CIP
- Vilma Hualla Mamani, Research assistant (IITCR), CIP
- Alberto Salas, Potato genebank curator, CIP
- Dieudonne Harahagazwe, Crop Ecophysiologist and Modeller, CIP-SSA

Crop Lead Center | Partners | CGIAR research program

CO_330

Add New Terms API Help Agtrials Annotation Tool Register Login

DOWNLOAD SHOW IBFIELDBOOK LIST English

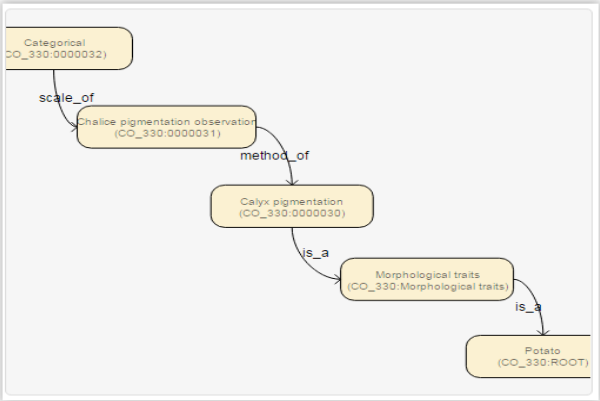
- Molecular traits
- Morphological traits
  - Anther pigmentation
    - Anther pigmentation observation (method_of)
      - SES 5 pt scale (scale_of)
  - Apex shape of the terminal leaflets
  - Apical Leaflet Shape
  - Apical Leaflet Size
  - Base shape of the terminal leaflets
  - Calyx pigmentation
    - Chalice pigmentation observation (method_of)
      - SES 7 pt scale (scale_of)
  - Color of the abaxial surface of leaves
  - Color of the adaxial surface of leaves
  - Corolla Shape

Term information

SES 7 pt scale | Permalink | General | 0 Comments

Identifier	CO_330:0000032
For Categorical: Class 1 - value = meaning	1 = Green
For Categorical: Class 2 - value = meaning	2 = Green with few pigmented spots
For Categorical: Class 3 - value = meaning	3 = Green with many pigmented spots
For Categorical: Class 4 - value = meaning	4 = Pigmented with many green spots
For Categorical: Class 5 - value = meaning	5 = Pigmented with few green spots
For Categorical: Class 6 - value = meaning	6 = Reddish
For Categorical: Class 7 - value = meaning	7 = Purple
For Categorical: Name of rating scale	SES 7 pt scale
Type of Measure (Continuous, Discrete or Categorical)	Categorical
created_at	Mon Jun 30 13:45:16 UTC 2014
name	Categorical

Add a new attribute



Ontologies and trait dictionaries are online for potato

http://www.cropontology.org/ontology/CO_330/Potato

Sweet Potato Ontology

Ontology curators

- Reinhard Simon, Integrated IT & Computational Research (IITCR) leader, CIP

Scientists

- Vilma Hualla Mamani, Research assistant (IITCR), CIP
- Raúl Eyzaguirre, CIP
- Raúl Cordova, CIP
- Robert O. M. Mwanga, NARO
- Genoveva Rossel, CIP
- Wolfgang J. Gruneberg, CIP
- Vilma Hualla Mamani
- Jorge Espinoza
- Maria Andrade
- Dieudonne Harahagazwe, Crop Ecophysiologist and Modeler, CIP-SSA
- Harrison Dapaah
- Sammy Agili
- Harrison Dapaah
- Felistus P. Ndingo-Chipungu
- Sreekanth Attaluri
- Regina Kapinga
- Tinh Nguyen
- Xie Kaiyung
- Koko Tjintokohadi
- Ted Carey
- Jan Low
- Genoveva Rossel

Crop Lead Center

Partners

CGIAR research program

CGIAR Research Program on Roots, Tubers and Bananas

DOWNLOAD | SHOW OBSOLETE TERMS | English

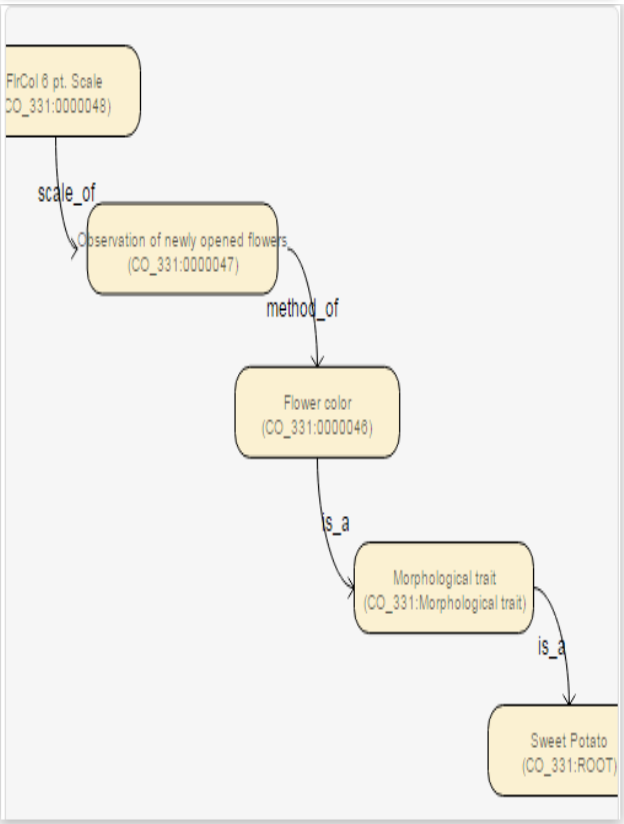
- Sweet Potato
  - Agronomic trait (is_a)
  - Biotic stress trait (is_a)
  - Morphological trait (is_a)
    - Abaxial Leaf Vein Pigmentation (is_a)
    - Distribution of Secondary Flesh color (is_a)
    - Flower color (is_a)
      - Observation of newly opened flowers (method_of)
        - FlrCol 6 pt. Scale (scale_of)
    - General Outline of the Leaf (is_a)
    - Ground Cover (is_a)
    - Immature Leaf Color (is_a)
    - Intensity of Predominant Skin color (is_a)
    - Latex Production in Storage Roots (is_a)
    - Leaf Lobe Number (is_a)

FlrCol 6 pt. Scale | Permalink | General | 0 Comments

Identifier	CO_331:0000048
Category 1	White
Category 2	White limb with purple throat
Category 3	White limb with pale purple ring and purple throat
Category 4	Pale purple limb with purple throat
Category 5	Purple
Category 6	Other
Scale class	Ordinal
Scale name	FlrCol 6 pt. Scale

created_at	Thu Jan 28 16:37:46 UTC 2016
name	FlrCol 6 pt. Scale

Add a new attribute



Ontologies and trait dictionaries are online for sweetpotato http://www.cropontology.org/ontology/CO_331/Sweetpotato

Note

- ✓ Better if you're using ontologies from the beginning using SweetpotatoBase, HiDAP, etc.
- ✓ Otherwise, you can define the variables in a Data Dictionary or reuse Data Dictionaries from other research groups.

International Potato Center

A member of the CGIAR Consortium

Documentation and Metadata

Research Informatics Unit.