GUIDELINES FOR CREATING CROP-SPECIFIC ONTOLOGIES TO ANNOTATE PHENOTYPIC DATA

Crop Ontology
for agricultural data

VERSION 2
2020

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

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

CGIAR
Platform for Big Data in Agriculture

CIMMYT
International Maize and Wheat Improvement Center

TO BE USED WITH THE TRAIT DICTIONARY
TEMPLATE VERSION 5.1

AUTHORS

Julian Pietragalla, Integrated Breeding Platform
Léo Valette, Bioversity International until 2015
Rosemary Shrestha, CIMMYT
Marie-Angélique Laporte, Alliance Bioversity-CIAT
Tom Hazekamp, Consultant for Alliance Bioversity-CIAT
Elizabeth Arnaud, Alliance Bioversity-CIAT

COPYRIGHT

CITATION

Pietragalla J., Valette L., Shrestha R., Laporte M.-A., Hazekamp T., Arnaud E. Guidelines for creating crop-specific ontologies to annotate phenotypic data, version 2.0, Alliance Bioversity International-CIAT, December 2020

CONTENT

Acknowledgement

The Crop Ontology has existed for 12 years thanks to the dedication of the CGIAR curators for maintaining the quality of the ontologies specific to their Center's mandate crops. We also acknowledge the crucial contribution of research partners from public and private sectors that comply with the guidelines and submit traits and variables used in their multi-partner projects to expand the ontology content.

The Crop Ontology was created with the financial support of Bioversity International and the Integrated Breeding Platform, and is maintained thanks to the CGIAR Research Programmes, the CGIAR Platform for Big Data in Agriculture, all funded by the CGIAR Fund Council and, from 2014 to 2019, with the award of the US National Science Foundation (NSF) to the cROP-Planteome Project (IOS:1340112 award).

We thank our colleagues for their valuable feedback on the guidelines and the template, namely:

Célia Michotey, INRAe, URGI, Forest Tree Data Manager, France

Cyril Pommier, INRAe, URGI, Information System and Data Integration Deputy Leader, France

Françoise Potier, CIRAD, Genetic Improvement and Adaptation of Mediterranean and Tropical Plants, Genetics and Variety Innovation Team, France

David Waring, Research Support Specialist, Cornell University, Department of Plant Breeding and Genetics, USA

Principal curators per species in 2020

Bambara Groundnut	Liliana Andres, South King Cross University
Banana	Marie-Angélique Laporte, Alliance Bioversity-CIAT
Barley	Ramesh Verna and Fawzy Nawar, ICARDA
Beet	Daphne Verdelet and Cyril Pommier, INRAE
Bracharia	Valheria Castiblanco and Luis Miguel Hernandez, Alliance Bioversity-CIAT
Brassica	Wiktor Jurkowski, Earlham Institute
Cassava	Afolabi Agbona, IITA and Naama Menda, Boyce Thompson Institute
Castor Bean	Salihu, B. Z., NCRI, Nigeria
Chickpea	Roma Das and Abhishek Rathore, ICRISAT
Common Bean	Guerrero Alberto Fabio, Alliance Bioversity-CIAT
Cotton	Jing Yu, Washington State University
Cowpea	Sam Ofodile and Tunde Agbaje, IITA
Faba Bean	Fouad Maalouf, ICARDA
Groundnut	Abhishek Rathore, ICRISAT
Lentil	Julian Pietragalla , IBP
Maize	Rosemary Shrestha, Kate Dreher, CIMMYT and Julian Pietragalla, IBP
Mungbean	Julian Pietragalla, IBP
Oats	David Waring, Cornell University
Pearl Millet	Roma Das and Abhishek Rathore, ICRISAT
Pigeon Pea	Roma Das and Abhishek Rathore, ICRISAT
Potato	Vilma Hualla, Elisa Salas, Thiago Mendes, CIP
Rice	Jeffrey Detrás and Julia Mae Pasuquin, IRRI
Sorghum	Abhishek Rathore, ICRISAT
Soybean	Rex T. Nelson, USDA
Sugar Kelp	David Waring, Cornell University
Sunflower	Evan Staton, University of British Columbia
Sweet Potato	Jolien Swanckaert, CIP
Vitis	Eric Duchêne, INRA
Wheat	Rosemary Shrestha, CIMMYT and Julian Pietragalla, IBP
Woody Species	Celia Michotey, INRAE and Ines Chaves, IBET

Introduction

The quality and consistency of agricultural data has greatly improved since the use of digital technologies applies to all steps of the research data lifecycle, particularly at the collection and management stages that integrate standardized ontology terms, such as lists of defined crop traits and variables. Created in 2008 by the CGIAR, the Crop Ontology (CO; http://www.cropontology.org) is an essential source of traits and variables to support the standardization of the breeding databases such as the Integrated Breeding Platform's BMS (IBP; https://www.integratedbreeding.net/) and the Boyce Thompson Institute's Breedbase (https://breedbase.org/) (Arnaud et al., 2020). By providing descriptions of agronomic, morphological, physiological, quality, and stress traits along with their definitions and relationships, also including a standard nomenclature for composing the variables, the CO enables digital capture and aggregation of crop trait data, as well as comparison across projects and locations (Shrestha et al, 2012). The crop ontologies follow a **conceptual model** that defines a phenotypic variable as a combination of a trait, a method and a scale. This model aims to support the creation and management of breeders' field books and the generation of annotated trial data. Annotated data are interpretable, interoperable and reusable.

The development of a crop-specific ontology is a community-driven effort which is usually coordinated by a curator (or curators) nominated from within the community itself. The curator(s) is responsible for coordinating discussions with domain experts and developing a quality **Trait Dictionary** (TD) using the template. The TD is a structured format which can be used to compile, curate and harmonize the phenotypic variables for the crop. Once the TD is finalised and is considered to be stable, it can be uploaded and published on cropontology.org (http://www.cropontology.org/add-ontology). **We strongly recommend reading and applying the Guidelines to develop a high-quality Trait Dictionary containing all necessary information for a variable.** This is the condition to enable the reuse of Crop ontology by a wide community, including industries and robust mapping with other ontologies.

The CO was included in the Planteome's ontology project funded by the National Science Foundation, US (IOS:1340112 award; http://planteome.org). CO's Traits that are properly described following the guidelines are progressively mapped to the Planteome species-neutral Trait Ontology (TO) maintained by Oregon State University, thus enabling users to search for a trait without consideration of the species (Arnaud *et al.*, 2012; Laporte et *al.*, 2016). This is useful for studies in comparative genomics or for grouping traits for a family or a clade (*e.g.* legumes) (Cooper *et al.*, 2018). The CO is listed among the most popular ontologies used in agriculture (Leonelli *et al.*, 2017; Harper *et al.*, 2018). To further support the standardization of the breeding data sets, the CO format was adopted by the metadata schema called the Minimum Information About a Plant Phenotype Experiment (MIAPPE https://www.miappe.org/; Ćwiek-Kupczyńska *et al.*, 2016; Papoutsoglou, et al., 2020) and also by the Breeding Application Programming Interface (BrAPI; https://brapi.org/; Selby *et al.*, 2019) that enables the extraction of genotype and phenotype data across databases (Arnaud *et al.*, 2020).

The CO provides the crop ontologies under a CC BY 4.0 licence and is regularly synchronised with the Ontology Lookup Service of the European Bioinformatics Institute

(EBI, https://www.ebi.ac.uk/ols/search?q=Crop+ontology) and Agroportal, the registry of ontologies in agriculture and related domains (http://agroportal.lirmm.fr/).

What is new in Version 2.0 ?

The Crop Ontology is evolving to address the feedback of curators as well as the new types of trait measurement methods and modelling. Therefore, the Guidelines need regular revision. The version 2 also aligns its content to few modifications brought to the Trait Dictionary Template version 5.1.

In version 2, some sections and tables have been revised to improve their clarity. New sections were added, such as 'The boundaries of the CO Model' about the metadata that describes the trait but is not part of the CO model, such as time series, time stamp, subsamples and experiment factors, Frequently Asked Questions and a bibliography for further reading were added as well as guidance in defining the abiotic and biotic stress variables in Annex 1.

The list of method types has been augmented with the *prediction* type to include trait predictive tools such as Near-Infrared Spectroscopy (NIRS) which are not calculations with a formula. Two new method classes were added to include (a) trait *classification* methods using Machine Learning algorithms and (b) *description* class for trait description in free text for example, it is collected in Participatory Varietal Selection (PVS) or surveys.

We acknowledge the valuable contribution of the curators to the maintenance of a quality content addressing the guidelines recommendations and partners' active collaboration.

The CO phenotype annotation model

This section presents the full model used to create a crop ontology and will guide you to properly fill in the Trait Dictionary Template explained in page 19.

The CO concepts

Variable

Breeders' observations or measurement are Variables

The CO model is grounded in breeder's datasets where the column headers generally integrate a lot of codified information. Table 1 shows an example of a breeder's dataset subset.

Table 1: example of a subset taken from a breeder's dataset

Germplasm ID	PH	GC OL	GY
24530	80	2	35
85432	120	4	24
8452	90	4	30
…			

The columns' headers, PH, GCOL and GY, are breeders' abbreviations that respectively stand for in the example above:

- PH = *Plant height*, defined as the distance from the ground to the top of the plant, that is measured with a ruler and expressed in cm
- GCOL = *Grain colour*, visually assessed at maturity and expressed on a codified categorical scale where 2=White/red and 4=Red
- GY = *Grain yield* which is derived by dividing the weight of dehulled grains harvested from the plot by the surface of the plot and which is reported in g per m^2

The CO model was developed considering that the value of an observation or a measurement made by a breeder on a trait is associated with a **variable**. Therefore, the CO model provides a standard framework for the definition of breeders' **variable** that measures a **trait** such as Plant height (*PH)*, Grain colour (*GCOL)* and Grain yield (*GY)*. The essence of the CO model is to decompose a variable recorded by the breeders into:

- A **trait**: "what is observed"
- A **method**: "how the observation is made"
- A **scale**: "how the observation is expressed"

In other words, **a variable is the combination of 1 trait, 1 method and 1 scale**.

1 variable = {1 trait, 1 method, 1 scale}

The purpose of this is two-fold:

1. The model provides breeders with clearly defined **variables** that can be readily incorporated into **breeding field books**
2. The model makes more granular **data integration** possible by formalising the different pieces of information that compose a variable

To illustrate this, Table 2 shows the breakdown of the PH, GCOL and GY examples.

Table 2: Breakdown of PH, GCOL and GY into trait, method and scale

Variable_name	Trait_name	Trait_abbreviation	Method_name	Method_ description	Scale_name
PH_M_cm	Plant height	PH	PH measurement	The distance from the ground to the top of the plant measured with a ruler	cm
GCol_E_1to3	Grain colour	GCOL	GCOL estimation	Visually assessed at maturity using a standard colour chart	1=White 2=White/red 3=Red/white 4=Red 5=Purple
GY_Calc_gm2	Dehulled grain yield	GY	GY computation	Measured on harvested and dehulled grain. Divide weight by plot area.	g/m^2

Breaking down the annotation of variables into trait, method and scale allows querying heterogeneous databases such as datasets of breeding trials, gene markers, agronomy trials, and retrieving the data points that reference the same trait and/or method and/or scale. CO annotated data can be queried across multiple data sources to retrieve all the related data points, e.g. retrieve all annotations of the trait plant height, regardless of the observation method (direct measurement, visual assessment, mean computation of direct measurements, etc.) and regardless of the reporting unit or the scoring system (cm, in, m).

Note: The Trait Dictionary Template version 5.1 does not currently hold a '*Variable Description'* column that would enable adding a human-readable description that can be used to ease the variable selection in field book builder interfaces instead a concatenation of abbreviations. Additionally, the 'lower_limit' and 'upper_limit' columns that are currently indicated at the scale-level should be associated with the variable. This will be corrected in the next version of the Template.

How to create the variable name?

The variable name must be composed as follows:

Trait abbreviation_Method abbreviation_Unit/scale abbreviation

This combination will precisely document the value of a measurement or an observation for its storage in databases and reuse in analytical pipelines. The use of the underscore is imposed by the format of the computerised field books. Therefore, when creating a variable name, it is advised to fill in first the Trait Dictionary Template's columns related to the trait, then the method and the scale/unit. Combining their abbreviations will ease the creation of the variable name and will secure that its format corresponds to the related information. Some collaborators have developed scripts that automatically create the variable name out of the 3 elements. Additionally, it is recommended to check how traits, methods and scales were previously abbreviated for your crop, and also for other crops, in order to reuse the existing abbreviations and secure consistency.

As shown hereunder, diverse abbreviations were already used to compose a variable name for the trait '100 Grain weight' with a similar measurement method:

GW_M_g100grn
HGW_Meas_g100gr
GW_100grnM_g

This situation complicates the integration of the data and should ideally be corrected.

The ordinal scale is always integrated in the variable name by indicating the categories used:

StayGreen_E_1to5

Trait

Combination of an entity and an attribute

The CO model defines a trait as **a character or property of an individual plant or of a group of plants that can be observed and that results from the expression of its genes and their interaction with the environment (phenotype)**. This trait definition can be summarised by "what is observed". Examples of traits are *plant height, pod colour, grain yield, seed germination rate, plant time to flowering, panicle shape, plant resistance to blight*, etc.

Traits can be formalized by a **meaningful combination** of **entities** and **attributes**. The entity is the **observed part of the plant**. The entity can be defined at the level of the cell, the tissue, the organ, the whole organism or the sub-product of the crop. The **attribute** is the **feature of an entity**. Examples are weight, length, area, colour, chemical content, grain-filling duration, photosynthesis rate, senescence rate.

A direct consequence of the trait decomposition into entity and attribute is that two traits that share the same combination of entities and attributes are identical even though they are named differently. Thus, decomposing a trait into an entity/attribute is a good practice to prevent trait duplication. Table 3 shows examples of trait breakdown into entity/attribute.

A trait can be described by a general entity or as a specimen, something that obviously belongs to a particular entity but is noticed by reason of an individual distinguishing characteristic. Examples of specimens are primary branch, first flower, first grade commercial fruit, flag leaf, main stem, fertile stem.

The traits related to biotic stress are usually observed as an interaction of two organisms, a plant (host) and a stressor (pathogen or pest). Therefore, the plant or plant part is considered as the entity and the stressor as part of the attribute. A further discussion of stress related traits is included as an annex 1 of this document.

Table 3: Examples of trait breakdown into entity and attribute

Trait	Entity	Attribute
Plant height	Plant	Height
Number of secondary branches	Secondary branches	Number
Plant flowering time	Plant	Flowering time
Plant phenotypic acceptability	Plant	Phenotypic acceptability
Leaf chlorophyll content	Leaf	Chlorophyll content
Leaf area index	Leaf	Area index
Barley leaf rust severity	Plant	Barley leaf rust severity
Flour gluten content	Flour (milled grain)	Gluten content
Dough elasticity	Dough	Elasticity
Plant rust severity	Plant	Rust severity
Leaf miner damage	Leaf	Miner damage

Trait classes

To provide the crop ontologies with an overall structure, the traits are grouped into a number of trait classes. The participants of the 2014 Crop Ontology workshop agreed on a set of 9 trait classes. **Table 4** lists these trait classes and provides a definition.

Table 4: Definition of the trait classes

Trait class	Definition	Examples of traits
Abiotic stress	All traits related to stress caused by non-living stressors. Abiotic stress is defined as the negative impact of non-living factors on the plants. Most common abiotic stressors are drought, waterlogging, high/low temperatures, mineral toxicities/deficiencies, hail and wind.	○ Plant aluminium tolerance ○ Plant drought susceptibility ○ Plant frost damage
Agronomic	All main traits contributing to yield and related to the agronomic performance of plants.	○ Seed/fruit/tuber yield ○ Biomass yield ○ Plant lodging incidence
Biochemical	All traits related to chemical components of a plant entity.	○ Leaf ABA content ○ Seed Proline content ○ Tuber carotenoid content
Biotic stress	All traits related to stress caused by living stressors. Biotic stress is defined as the negative impact on the crop/plants of living organisms such as bacteria, viruses, fungi, parasites, nematodes, weeds, invertebrate and vertebrate pests.	○ Plant disease severity ○ Plant disease incidence ○ Plant pest damage ○ Disease plant response
Fertility	Traits specifically related to fertility aspects of importance to breeding	○ Outcrossing extent ○ Fertility restoration ○ Male sterility group ○ Pollen sterility
Morphological	All traits related to anatomical (internal) and morphological (external) structure of the plant, its organs and tissues.	○ Fruit shape ○ Seed colour ○ Stem diameter ○ Seed length
Phenological	All traits related to growth/developmental stages and periods of crop/plants.	○ Plant flowering time ○ Plant maturity time ○ Plant vegetative period length
Physiological	All traits related to functioning of the crop/plant and its response/adaptation to the environment.	○ Leaf senescence rate ○ Canopy temperature ○ Canopy NDVI ○ Leaf stomatal conductance
Quality	All traits related to key characteristics that influence end-use quality of crop/plant products (seed, fruit, leaf, root/tuber, etc.) and sub-products (flour, dough, pulp, etc.).	○ Seed protein content ○ Fruit sugar content ○ Dough colour ○ Pasta consumer acceptability

Note: An ontology developed with an ontology development tool, such as Protégé, usually allows a concept to be classified in multiple classes. However, due to the Excel based format of CO, a trait can only be assigned to 1 trait class. If a trait can be classified in multiple classes, select the most representative class.

Method

Definition

The method describes how the trait is observed which covers two aspects: the sampling and the protocol.

Sampling

The **sampling** specifies if the observation or measurement is done either on one individual plant entity or on a collection of plants.

Observation procedures

The method must also detail the **procedure** to follow in order to observe the sampled entity/entities and include within its definition the growth stage at which the observation should be done. A dedicated column named ‘Growth stage’ exists in the Trait Dictionary Template to capture this specific information. If the procedure is described in the literature, the reference must be added in the dedicated “Reference” column.

Example of protocols are:

- Measurement of the plant height using a ruler
- Leaf area derived from image analysis. Lay the leaf flat and take a picture with the lens set parallel to the leaf. Single out the pixels of the leaf by filtering the image for contrast using <software>. Count the number of leaf pixels and multiply the count by the area represented by each pixel.
- Visual estimation of the peduncle length.
- Grain weighing with a scale.
- Visual assessment of the leaf colour based on a standard colour chart.

It is important to clarify here that **the observation procedure is different from the experimental protocol**. The experimental protocol combines the trial design, treatment factors and the experimental conditions. The observation procedure focuses on strictly defining how the observation is made, not how the trial is led. See also section ‘*Boundaries of the CO model for Phenotypic Data Annotation*’.

The method classes

Methods are categorized in 7 classes: “measurement”, “counting”, “estimation”, “computation”, “prediction”, “description” or “classification” as listed in Table 5.

Table 5: Classification system of the Methods

		Method class	Examples
The trait is observed directly	By using a measuring device, a sensor, a trained sensory panel using a lexicon of properties with a categorical scale	Measurement	○ Plant height measuring with a ruler ○ Fruit weighing on a scale
	By counting entities	Counting	○ Leaf counting
	By an assessment that only relies on the experience and subjectivity of the observer. The assessment is not supported by	Estimation	○ Grain colour estimation ○ Damage on leaves visual estimation

	a measuring device. It includes the hedonic sensory evaluation		o Plant height visual estimation
The trait observation derives from an aggregation of observations (regardless of how they were observed i.e., measured, counted, estimated, computed or predictive)		Computation	o 1000 grain weight calculation (1000 * measured grain weight / grain count) o Harvest index calculation (Grain yield / Aboveground biomass) o Grain protein content calculation (Grain Nitrogen content * 6.25)
The trait value is predicted as a result of a model or a spectrum before it is directly observed*		Prediction	o Lignin content of the dry matter predicted by NIRS
The trait is described by the observer in free text		Description	o Coffee flavour notes o Farmer's variety acceptability notes
The trait is obtained by processing one or more variables using algorithms, machine learning techniques and others for identifying to which category, out of a set of categories, a new observation belongs. The result is usually reported as an ordinal or nominal category		Classification	o Crop maturity by images classification (Early, Medium, Late) o n-Disease severity by Image analysis (None, Low, Medium, High)

Note: The Trait Dictionary Template version 5.1 does not have a 'Method Abbreviation' column and this will be corrected in the next version of the template. Also, note that including the trait abbreviation in the method name allows the identification of trait-specific methods.

Scale

The scale describes how the trait observation is **expressed**. When the observation is expressed by a quantitative value, the CO scale describes the unit. Alternatively, when the observation is expressed by predefined categories, the CO scale describes the possible values and their meaning. When creating a new variable, if the scale or unit already exists in the crop ontology, it is mandatory to reuse the existing identifier (CO ID).

Example for Wheat:

Trait	Method	Unit	Unit ID
Above ground biomass at maturity	BM computation	kg/ha	CO_321:0000804
Grain Yield	GY computation	kg/ha	CO_321:0000804

Note: For some crops, different scales IDs were attributed to the same scale. This situation must be regularised with the help of the ontology team.

Table 6 details the CO classification system for scales.

Table 6: Scale classes

Scale name	Description
Date	The date class is for events expressed in a time format, e.g. "yyyymmdd hh:mm:ss – UTC" or "dd-mm-yy". A good practice recommended by the Breeding API (BrAPI) is to use the Date and timestamp fields coded in the ISO 8601 standard, extended format. Check https://github.com/plantbreeding/API/blob/master/Specification/GeneralInfo/Date_Time_Encoding.md)
Duration	The duration class is for time elapsed between two events expressed in a time format, e.g. "days", "hours", "months".
Nominal	Categorical scale that can take one of a limited number of categories. There is no intrinsic ordering to the categories e.g. r="red", g="green", p="purple".
Numerical	Numerical scales express the trait with real numbers. The numerical scale defines the unit e.g. centimetre, ton per hectare, number of branches.
Ordinal	Ordinal scales are composed of ordered and fixed number of categories e.g. 1=low, 2=moderate, 3=high
Text	A free text is used to express the scale value. Also known as Character variable (varchar) e.g. "Preferred when slightly undercooked".
Code	This scale class is exceptionally used to express complex traits. Code is a nominal scale that combines the expressions of the different traits composing the complex trait. For example, a disease related code might be expressed by a 2-digit code for intensity and 2-character code for severity. The first 2 digits are the proportion of plants affected by a fungus and the 2 characters refer to the severity, e.g. "75HD" means "75% of the plants are infected and plants are highly damaged". It is recommended to create variables for every component of the code.

Note: The Trait Dictionary Template version 5.1 does not currently propose a 'Scale Abbreviation' column, and this will be corrected in the next version. The columns indicating the minimum and maximum values of a variable are currently in the Scale section of the Template v. 5.1 and will be moved in the variable's section in a future version.

Boundaries of the CO model for Phenotypic Data Annotation

Experimental conditions or factors

The CO model provides a standardised framework to develop crop ontologies. These ontologies facilitate the description of crop phenotypes and the annotation of the associated phenotypic data. It defines a variable as a combination of a trait (what is observed), a method (how the observation is done) and a scale (how the observation is expressed). However, a scientific experiment usually generates much more data to describe the context and circumstances under which the variable was assessed. These would typically be covering aspects such as (see figure 1)

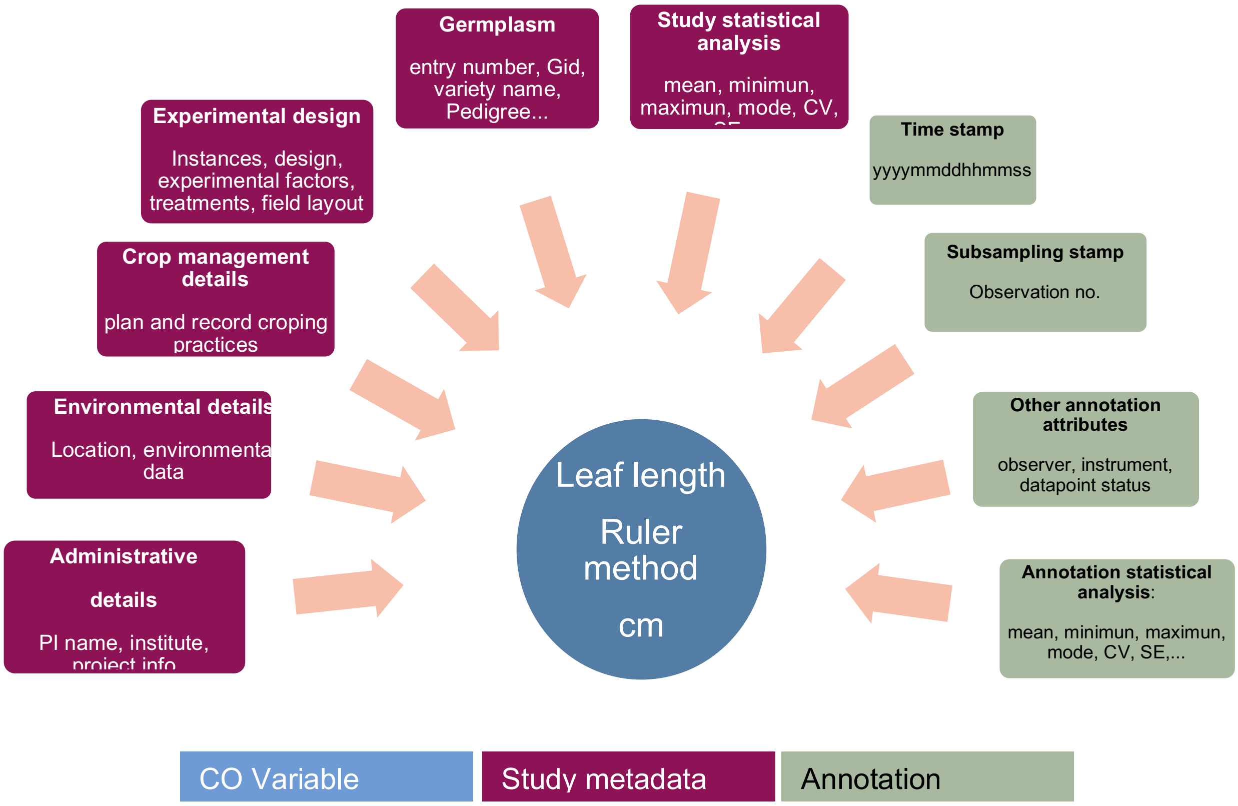
- Where and when the observation was made
- Which specific plants/entities were observed
- Which treatments were applied
- In which environmental conditions the plant/plants grew
- In which experimental conditions
- How the trial was designed and managed
- Who did the observation
- Statistical analysis, information resulting from an analysis of the variable

These experimental metadata can be divided in two main groups, study metadata that is information about a trial which describes its contents and context, and annotation metadata that is information describing the single data-point annotation. Metadata can be very useful and crucial information for storing, analysing and sharing files as well as helping collaborators understand the context of its data.

Systems managing data present diverse database structures; there is not a single approach on how to record study and annotation metadata to data points. If a particular database does not support metadata annotation, following procedures and conventions will aid to associate metadata to CO variables.

Subsampling tags and time stamps are the most common information associated with a variable. Other key aspects usually attached to a variable as metadata are: study administrative details, crop management details, treatment factors and environmental details. All of this valuable information is not part of the variable and must be kept outside the variable construction. Therefore, we recommend the use of the Minimum Information About a Plant Phenotyping Experiment (MIAPPE) and of the Agronomy Ontology (AgrO) for the field management practices.

Figure 1: CO variable composition, annotation metadata and study metadata



The Minimum Information About a Plant Phenotyping Experiment (MIAPPE) v. 1.1

MIAPPE (https://www.miappe.org/) is a metadata standard recommended to fully document phenotyping experiments. MIAPPE is an open, community-driven project to harmonise data from plant phenotyping experiments. MIAPPE comprises both a conceptual checklist of metadata required to adequately describe a plant phenotyping experiment, and software to validate, store and disseminate MIAPPE-compliant data. ***This metadata standard is compatible with the CO format.***

The Agronomy Ontology

The Agronomy Ontology (AgrO; https://bigdata.cgiar.org/resources/agronomy-ontology/) provides terms from the agronomy domain that are semantically organised and can facilitate collection, storage and use of agronomic data, enabling easy interpretation and reuse of the data by humans and machines alike. **The entry point of the ontology is the agricultural experimental plot**. This plot can be an entire field or part of it. It is basically where the experiment takes place. The class ‘’plot’’ is linked to different concepts like:

- the variables that can be measured on that plot: the plot length, width and area.
- the activities happening on that plot, e.g. planting a crop or applying fertiliser is grouped in a class called ‘agricultural process’. A process is everything that is

time-bounded so this class enables describing what is happening in the plot or field over time. For each process, a list of 'participants' is provided.

- A 'participant' can be tools used during the process or material added to the field (e.g. crop, fertiliser), and can be linked to different techniques to perform this process.

To learn more about AgrO and how to use it, contact M.A. Laporte, m.a.laporte@cgiar.org

Trait grouping in CO

Observation protocols, in some cases, include and describe processing operations such as grinding, milling, polishing, drying. These operations give detailed information regarding the entity that is observed, e.g. fresh or dry plant, hulled or dehulled grain. Based on the definition of the CO "trait" concept, this information about the entity is to be defined in the trait description.

Yet, two variables that differ by the entity that is observed are sometimes differentiated at the level of the method. To illustrate this, Table 7 shows two alternative trait decompositions into trait, method and scale for "plant biomass in g" and "plant dry biomass in g".

Table 7: Alternative trait decompositions

	Trait-Method-scale decomposition	Variable 1: "plant biomass in gram"	Variable 2: "dry plant biomass in gram"
Option A: 2 traits, 1 method	Trait	Fresh plant biomass	Dry plant biomass
	(2 different trait entities)	(Fresh plant)	(Dry plant)
	Method	Weight measurement	Weight measurement
	Scale	Gram	Gram
Option B: 1 trait, 2 methods	Trait	Plant biomass	Plant biomass
	(1 trait entity)	(Plant)	(Plant)
	Method	Fresh weight measurement	Dry weight measurement
	Scale	Gram	Gram

Both options (A and B) can be used, although option A is usually the option preferred by biologists/breeders as they consider that traits having a different biological meaning should be distinguished at the trait level. Option B is sometimes preferred by data managers to compensate for the fact that traits only have a one-tier hierarchy. Option B offers the advantage to ease the ontology browsing online and secondly provides in the method details such as the percentage of humidity.

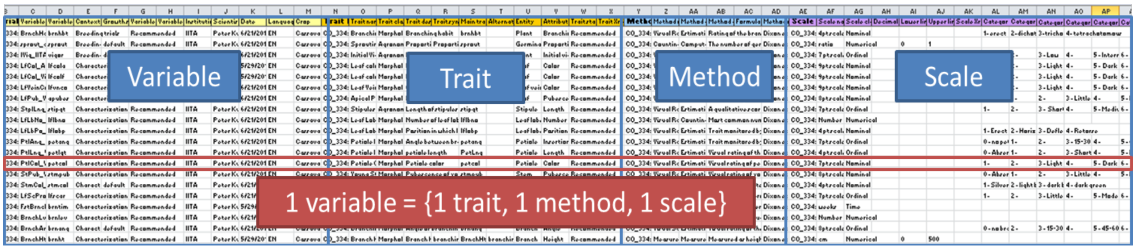
The Trait Dictionary Template v. 5.1

The crop ontologies and a number of associated ontologies are available for browsing and download at the Crop Ontology website (http://www.cropontology.org/). New crop ontologies and/or additional terms can be submitted using a Trait Dictionary (TD) Template. The Trait Dictionary is an Excel file which provides a versioned template to facilitate the definition of new terms. It can be downloaded from the Crop Ontology website.

Structure of the TD Template Version 5.1

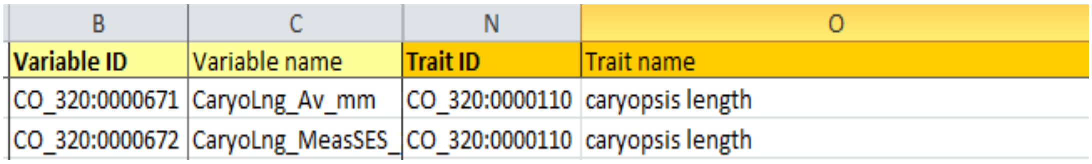
The TD template is a two-dimension table. The columns represent the properties of the variable, trait, method and scale concepts. The rows represent the instances of variables and their corresponding trait, method and scale (see Figure 2).

Figure 2: Structure of the TD template



Within a TD, each variable is unique and can only appear once. However, a given trait, method or scale might appear in more than one variable. If they do, all instances of that term need to be identical copies. See figure 3 for an example of a trait being used in 2 different variables whereby all trait elements (“Trait ID”, “Trait name” etc.) are exact copies.

Figure 3: Example of the trait “caryopsis length”. It is duplicated in the TD because it is part of the “CaryoLng_Av_mm” and “CaryoLng_MeasSES_1to7” variables



B	C	N	O
Variable ID	Variable name	**Trait ID**	Trait name
CO_320:0000671	CaryoLng_Av_mm	CO_320:0000110	caryopsis length
CO_320:0000672	CaryoLng_MeasSES_	CO_320:0000110	caryopsis length

Note: To guide the creation and the update of variables, the Triticeae toolbox team, at Cornell University, has developed an intermediate Excel workbook format that holds a worksheet for each ontology data type (one for variable, trait, method, scale). This way repeating terms such as traits, methods or scales have to be entered only once. If you wish to test their format along with the scripts, particularly if you are a user of Breedbase, you can access it on the Github: https://github.com/Triticeaetoolbox/ontology-scripts

The CO identifiers

The CO model associates a unique identifier (ID) with each individual ontology term that follows the global recommended format. These IDs are constructed by concatenating a crop code, a separator (“:”) and a 7-digit number (see Figure 4). The list of the already assigned crop codes evolves over time with the submission of new ontologies, therefore the code list is accessible from the CO website. To get a new crop code, contact the Crop Ontology help desk.

Figure 4: Structure of a CO identifier

The properties of the CO concepts

The variable properties

Column name	Required /optional	Format	Specifications
Variable ID	Required	<Crop code>:<7-digit identifier>	For existing terms, give the unique identifier using the specification <crop code>:<7 digit>. For new terms leave the field blank. The Ontology help desk will generate a term ID for new terms.
Variable name	Required	formatted text: <trait abbreviation>_<method abbreviation>_<scale abbreviation>	Formatted name of the variable. The variable name has to be unique in a given Trait Dictionary. Its main function is to facilitate the identification of the variable and be used as a column header in data matrices. CO and IBP recommend a standard way to name variables. Names must have no space

			and no special characters except underscores, must not start with a numerical character and, by convention, be the concatenation of the trait abbreviation, the method abbreviation and the scale abbreviation. The 3 abbreviations must be separated by the underscore sign: <TraitAbbreviation>_<MethodAbbreviation >_< ScaleAbbreviation >.
Variable description	Optional	Text	The variable description will provide clarification about the variable name that is composed of abbreviations. The description will ease the selection of the variable in the field book builder interfaces of the databases. E.g. ‘100 grain weight in grams’
Variable synonyms	Optional	Free comma-separated text	List any alternative names of this variable. These synonyms do not have to follow the specification of the CO variable name
Context of use	Optional	Free text	If applicable, indicate the type of trials/research projects/models in which this variable is generally observed. If there is more than one context of use, separate the entries by a coma ",". This column is optional, but it is particularly recommended indicating the usage of the variable when different communities (e.g. breeders, physiologists, genebank managers, etc.) have contributed to the TD. There is no standard list of context options available.
Growth stage	Optional	Free text	If applicable, indicate the growth stage when the observation is generally made. If more than one growth stage, separate the entries by a comma. This column intends to give some information to a trial manager prior to the data capture. It indicates which variables could be observed at a given growth stage. However, do not use this growth stage column of timestamps.
Variable status	Required	‘recommended', 'standard <institution/community>', 'obsolete', 'legacy'	Indicate the status of the variable. If the variable is presently commonly observed by a community or an institution, indicate "standard <institution/community>". In addition, if the community requests the variable to be integrated in the BMS, indicate "recommended". If the variable is no longer observed by a community, indicate "legacy". If the variable description or the formalisation is incorrect, indicate "Obsolete" and record the reason for making the term obsolete.

Variable Xref	Optional	Free comma-separated text	Indicate the cross-reference of the variable term to an external ontology term or a database term of a major system.
Institution	Required	Free text	Indicate the name/acronym of the institution affiliated to the scientist. It is advised to use codes from the Research Organisation Registry (ROR) (https://ror.org/) already recommended by the Breeding API.
Scientist	Required	Preferably the name and ORCID	Indicate the name of the scientist who can be contacted for further information regarding the variable. It is recommended to add the ORCID (Open Researcher and Contributor ID, https://orcid.org/) which is non-proprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors. NB All the contributors to a crop TD are acknowledged on the header of the http://www.cropontology.org/ontology/<crop>/ page. Contact the crop ontology help desk to report updates on the list of contributors.
Date	Required	dd/mm/yyyy	Indicate the date of submission.
Language	Required	2 letter ISO code	Indicate the language of variable submission using the 2-letter ISO 639-1 code. On Trait Dictionary cannot contain multiple languages. A file per language must be generated for the template submission.
Crop	Required	Free text	Indicate the name of the crop/species that the variable applies to.

The trait properties

Column name	Required/ optional	Format	Specifications
Trait ID	Required	<Crop code>:<7-digit identifier>	Give a unique identifier for the term using the specification <crop code>:<7 digit> or leave blank for new terms. If left blank, the upload system will automatically generate a term ID.
Trait name	Required	Text	Name the trait (i.e. feature) that is observed.
Trait class	Required	'Morphological', 'Phenological', 'Agronomic', 'Physiological', 'Abiotic stress', 'Biotic stress', 'Biochemical', 'Quality ' or 'Fertility'	Categorise the trait in one of the consensus trait classes: either 'Morphological, 'Phenological’, 'Agronomic’, 'Physiological’, 'Abiotic stress’, 'Biotic stress’, 'Biochemical’, 'Quality' or 'Fertility’.
Trait description	Required	Text	Define the trait, explain what the trait entity and attribute are.
Trait synonyms	Optional	Comma-separated text	Indicate any alternative trait names. Separate synonyms with commas.
Main trait abbreviation	Required	Spaceless text with no special characters except ‘_’	Main abbreviation of the trait name. It has to be unique within a crop TD. By convention, this abbreviation must not start with a digit, must have no space and no special characters.
Alternative trait abbreviations	Optional	Text	Other frequent abbreviations of the trait, if any. These abbreviations do not have to follow a convention. Separate abbreviations with commas.
Entity	Required	Text	Indicate the plant part that the trait refers to e.g. “plant”, “stem”, "grain" etc.
Attribute	Required	Text	Indicate the characteristic of the plant part entity: “height”, “diameter”, "colour"
Trait status	Required	‘recommended', 'standard <institution/community>', 'obsolete' or 'legacy'	Indicate the status of the trait. If the trait is presently commonly observed by a community or an institution, indicate "standard <institution/community>". In addition, if the community requests the trait to be integrated in the BMS, indicate "recommended". If the trait is no longer observed by a community, indicate "legacy". If the description or the formalisation of the trait is incorrect, indicate "Obsolete" and record the reason for making the term obsolete.
Trait Xref	Optional	Comma-separated text	Indicate the cross-reference of the trait to an external ontology term (e.g. TO) or a database term of a major system.

The method properties

Column name	Required/ optional	Format	Specifications
Method ID	Required	<Crop code>:<7-digit identifier>	Give a unique identifier for the term using the specification <crop code>:<7 digit> or leave blank for new terms. If left blank, the upload system will automatically generate a term ID.
Method name	Required	Text	Name the method used to score the trait. A common practice is to construct the method name using the method class e.g. "Grain number per spikelet counting"
Method class	Required	'Measurement', 'Counting', 'Estimation', 'Computation', 'Description', 'Prediction', 'Classification'	Class of the method. Entries can be "Measurement", "Counting", "Estimation", "Computation", "Description", "Classification" or "Prediction" (see Table 5)
Method description	Required	Text	Describe the method
Formula	Optional	Text	Only if the method class is 'computation', the mathematical formula used to derive the trait may be indicated. Indicate only the quantities involved in the formula, not the prescribed units that the quantity is expressed in. For information, at present, no formalisation system has been agreed on.
Method reference	Optional	Text	Reference the literature that describes the method

The scale properties

Column name	Required/optional	Format	Specifications
Scale ID	Required	<Crop code>:<7-digit identifier>	Give a unique identifier for the term using the specification <crop code>:<7 digit> or leave blank for new terms. If left blank, the upload system will automatically generate a term ID.
Scale name	Required	Text	Name the scale. If the scale is of class 'Numerical' 'Date', Duration', or 'Count', indicate the unit. If the scale is 'Ordinal', it is recommended to indicate the number of categories e.g. '1 to 5 damage score'
Scale class	Required	'Numerical', 'Nominal', 'Ordinal', 'Text', 'Code', 'Time' or 'Duration'	Classify the scale in Class of the scale, entries can be 'Numerical', 'Nominal', 'Ordinal', 'Text', 'Code', 'Time', 'Duration' (see Table 6)
Decimal places	Optional	Real number	For 'Numerical' scales, specify the number of reporting decimal numbers

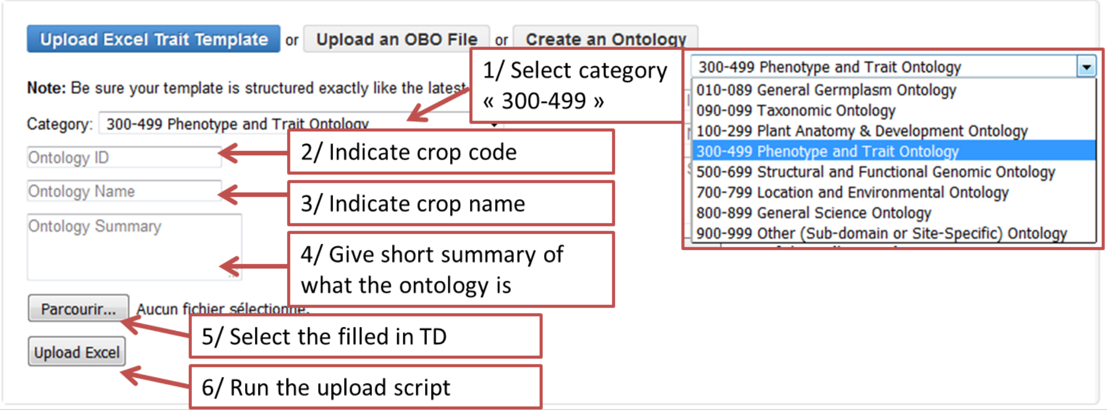
			to inform about the expected accuracy of the measurement.
Lower limit	Optional	Real number	Indicate the minimum possible (biological limit) value for data quality control.
Upper limit	Optional	Real number	Indicate the maximum possible (biological limit) value for data quality control.
Scale Xref	Optional	Comma-separated text	Indicate the cross-reference of the scale to an external ontology (e.g. UO) or to a database term of a major database
Category i	Required for 'Nominal' and 'Ordinal' scales	'i= category description' or 'i= category abbreviation = category description'	Enter the category of 'nominal' or 'ordinal scale' e.g. "1=Low"; P=Purple; MR=Moderately resistant

TD Upload and Update

The TD upload procedure

The ontology is generated by submitting an Excel TD using the cropontology.org upload procedure: http://www.cropontology.org/add-ontology#!/add/upload_excel. Figure 5 illustrates the upload procedure. Options to "Upload an OBO-file" or "Create an Ontology" should not be used any more.

Figure 5: TD upload procedure on http://www.cropontology.org/add-ontology#!/add/upload_excel



1/ Select category « 300-499 »

2/ Indicate crop code

3/ Indicate crop name

4/ Give short summary of what the ontology is

5/ Select the filled in TD

6/ Run the upload script

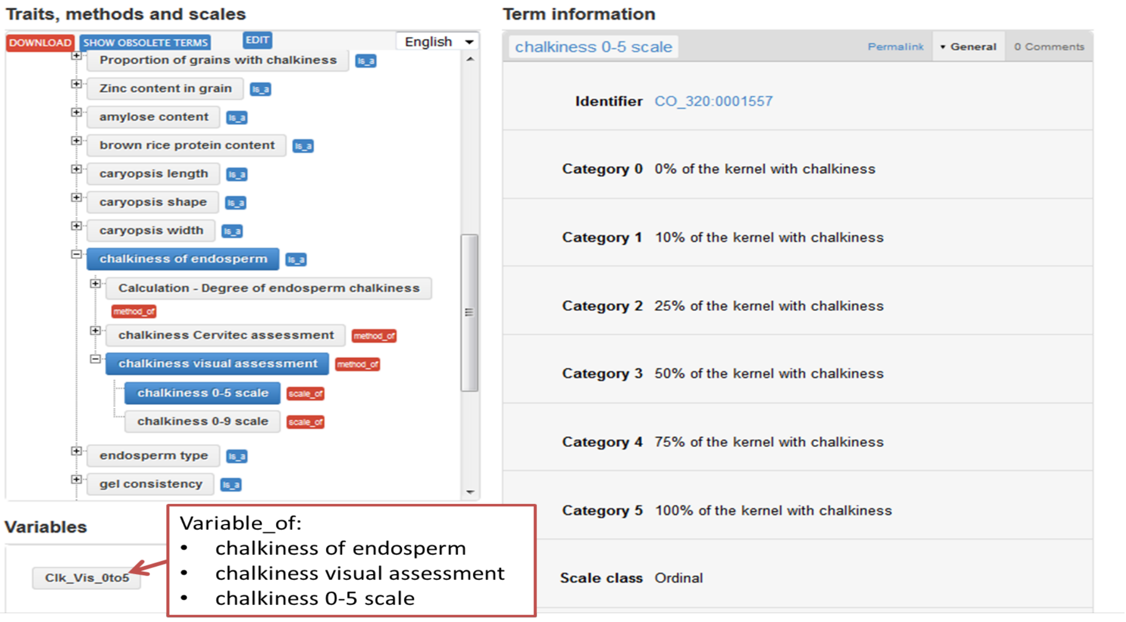
Note: Be sure your template is structured exactly like the latest

Category: 300-499 Phenotype and Trait Ontology

The upload script reads the TD row by row from left to right. It decomposes each row into a variable, a trait, a method and a scale. Note that in the TD, the information regarding the variables, traits, methods and scales are in the second sheet as the first sheet contains the TD element description. This is important for the upload procedure. Each time the upload script finds a TD term without an ID (i.e. <null> value), it automatically assigns the next available term ID. This ID management technique works for TD terms that are defined only once.

For replicated TD terms (see example in Figure 3), it is compulsory to manually determine the next available ID and assign it to all replicated terms BEFORE submitting the TD to the upload script! The script writes each term in the cropontology.org database and creates the links between the terms: a variable is a combination of a trait, a method and a scale. The terms are then browsable on the ontology display (see Figure 6) and can be served by the API ((Application Programming Interface).

Figure 6: Ontology terms display on http://www.cropontology.org/ontology/<crop code>/<crop name>



Variable_of:

- chalkiness of endosperm
- chalkiness visual assessment
- chalkiness 0-5 scale

chalkiness 0-5 scale

Identifier CO_320:0001557

Category 0 0% of the kernel with chalkiness

Category 1 10% of the kernel with chalkiness

Category 2 25% of the kernel with chalkiness

Category 3 50% of the kernel with chalkiness

Category 4 75% of the kernel with chalkiness

Category 5 100% of the kernel with chalkiness

Scale class Ordinal

Once the TD is uploaded, the ontology will appear in the list of ontologies on the home page (http://www.cropontology.org/). The creator of the ontology is listed as the owner.

Note**: Only the owner and the CO system administrators can make changes to the ontology.**

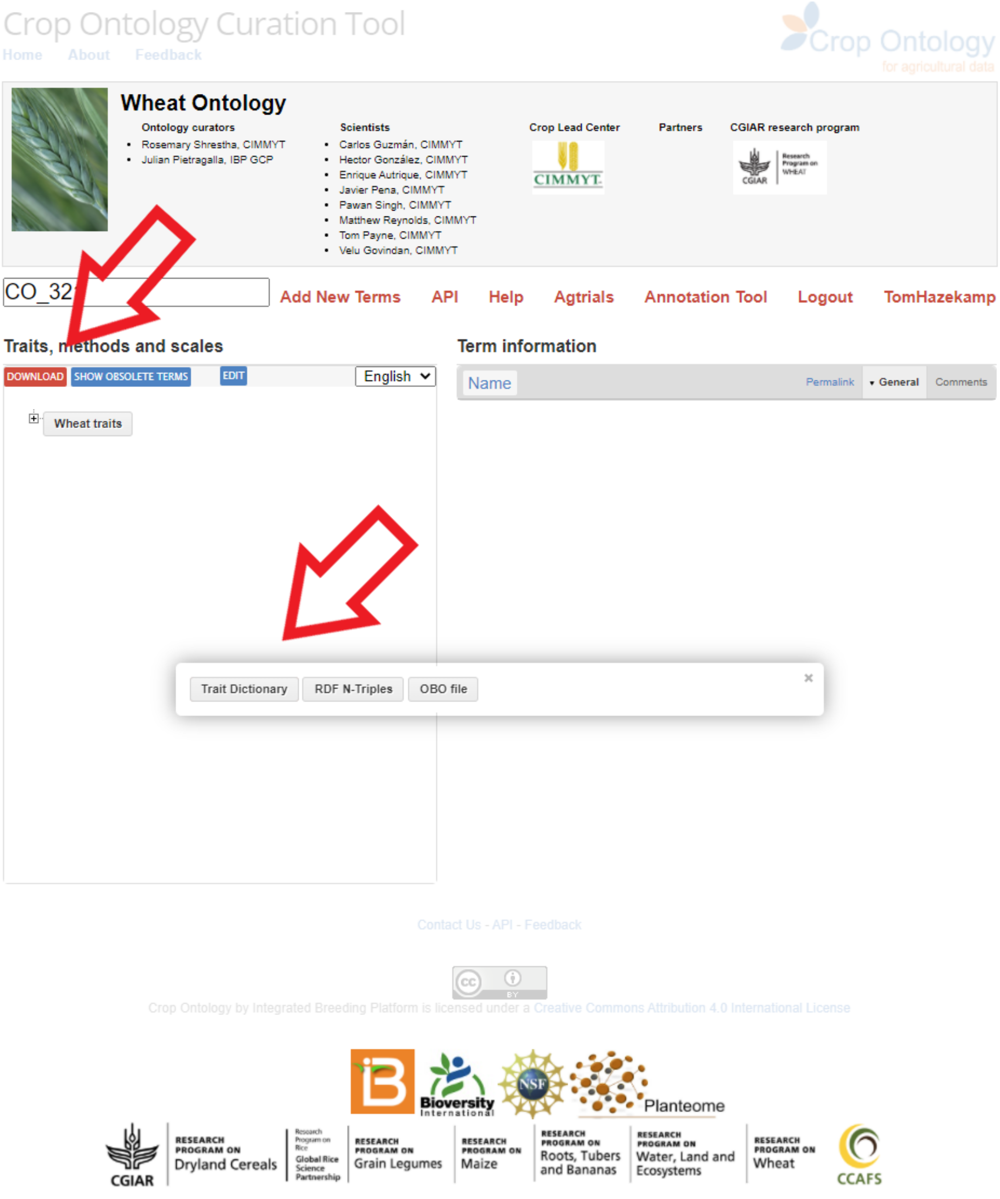
When a crop ontology is published on the CO site, it is available in the public domain under the **Creative Common license CC-By 4.0 (details provided at https://creativecommons.org/licenses/by/4.0/deed.en)**. This means that anyone can start using it with proper attribution with a link to the licence.

Note: Some crop ontologies are versioned in the Planteome GitHub. In that case, the reference file is the GitHub master branch that must be synchronised with www.cropontology.org. The update workflow is under construction, so please contact the help desk if you wish to contribute to these crop TDs.

Crop Ontology synchronisation with 3rd-party web sites

The Crop Ontology Application Programming Interface (API) is used by third-party websites and database like the EMBL-EBI Ontology Lookup Service (https://www.ebi.ac.uk/ols/index) that replicates CO and provides term search access through its own portal. Agroportal (http://agroportal.lirmm.fr/), a repository of ontologies in agriculture, also regularly synchronises their files. Crop ontology curators are therefore urged to limit publication on the CO site to ontology versions which are **considered fairly stable** and to refrain from publishing draft versions on the CO site. Draft versions should be posted on GitHub (https://github.com/Planteome) instead which provides a much better platform to solicit and manage feedback.

TD update



Ontology curators
- Rosemary Shrestha, CIMMYT
- Julian Pietragalla, IBP GCP

Scientists
- Carlos Guzmán, CIMMYT
- Hector González, CIMMYT
- Enrique Autrique, CIMMYT
- Javier Pena, CIMMYT
- Pawan Singh, CIMMYT
- Matthew Reynolds, CIMMYT
- Tom Payne, CIMMYT
- Velu Govindan, CIMMYT

Crop Lead Center

Partners

CGIAR research program

Crop Ontology by Integrated Breeding Platform is licensed under a Creative Commons Attribution 4.0 International License

If updates are necessary (e.g. the addition of a new set of variables), the trait dictionary should be downloaded by clicking the “DOWNLOAD” button and choosing the “Trait

Dictionary" option (see arrows in figure above). The necessary changes should be made in the downloaded Excel format. When all changes have been made, the ontology owner can login again, delete the existing on-line version of the ontology and upload the modified Excel Trait Template (see "The TD upload procedure" section).

Modification of the TD structure

To ensure that crop ontologies are uploaded AND downloaded correctly the TD template should not be modified.

Though uploading a Trait Dictionary with a modified template on http://www.cropontology.org/add-ontology#!/add/upload_excel works to a certain extent, downloading the Trait Dictionary from cropontology.org can only be done in a template v5.1.

The download feature of cropontology.org discards any added column to the template. NB Changing a column name is equivalent to deleting the column and adding an extra-column.

Despite the aforementioned disclaimer, if modifying the template for a specific crop/community is necessary, it is important to take into account that the cropontology.org upload feature will only handle:

- columns added to the variable concept by inserting columns in between the "Variable ID" and "Trait ID" columns
- columns added to the trait concept by inserting columns in between the "Trait ID" and "Trait Xref" columns
- columns added to the method concept by inserting columns in between the "Method ID" and "Method reference" columns
- columns added to the scale concept by inserting columns in between the "Scale ID" and "Category 1" columns

Frequently asked questions

This section covers some of the most frequently asked questions raised by the Crop Ontology community.

How to annotate repeated observations in space and in time?

Observations with timestamp and time series

Use of a time stamp is a standard practice when observing variables in a time series or to keep record of the time when the observation was made. Repeated observations of a variable on an experimental unit generate time series data that usually are processed to derive variables as senescence rate, canopy temperature averaged during grain-filling, etc. It also supports calculation of the area under the **disease** progress curve (**AUDPC**) which is a useful quantitative summary of **disease** intensity over time, for comparison across years, locations, or management tactics (definition by the American Phytopathological Society (APS).

The variable name can be extended by adding a component after an underscore using a "T" and a time code as follows:

(Trait)_(method)_(scale)**_T**(**time code)**

It would be understood as "Trait" by "Method" in "Scale" **at "time"**. See example hereunder and Table 8. Note that these variables should not be part of the Trait Dictionary when submitting new variables but should be stored in the database where the collected data is stored.

Example: Adding time codes to a variable name

Variable + time code: LfLng_rlrMtd_cm _Tday1

"leaf length" by "ruler measurement method" in "cm" at "day 1"

It is highly recommended using a widely adopted general or crop specific scale for the Developmental/Growth stages to codify phenological stages or periods, see Table 4 for time code examples.

Table 8. Examples of Time Codes

Stage or event date	Time Code	Example
Cereals Zadoks growth/development stages	GS(00-99) or DS(00-99)	DS65=heading DS75=anthesis DS87=physiological maturity
Maize growth/development stages	VE, V1-Vn, VT V= vegetative R1-R6 R=reproductive	VE=emergence V3=third leaf VT=tasselling R1=silking R6=physiological maturity
Soybean growth/development stages	VE, VC, V1-Vn, R1-R8	VE=emergence V4=second trifoliate R2=full flowering R8=full maturity
Generic code	Heading, flowering, vegetative period, grain filling period, maturity	hd, flw, vg, gf, mat
Days after emergence	dae	45dae, 65dae…ndae
Days after sowing	das	45das, 65das…ndas
Days after planting	dap	45dap, 65dap…ndap
Weeks after planting	wap	1wap, 2wap….
Months after planting	map	1map,2map…
Date	yyyymmdd	20150315
Date+hour+min	yyyymmddhhmm	201503151135
Time	t	t1, t2, t3…tn

How to manage observations on subsampled entities?

In the context of high-throughput phenotyping or advanced physiology trials particularly, scientists observe the same variable on several subsamples. An example is the repetition of the trait “leaf colour” by a visual estimation method and expressed in colour chart categories on the leaves of a number of the plants in the plot.

Example of a solution proposed by the BMS

Regarding the management of repeated observations in the BMS, multiple datasets may be created to manage the same CO variable at different levels like observation (plot) and sub-observations (plant, sample, time series). The principle is that for each CO variable with multiple observations, a level is defined and given a unique identifier. Repeated observations of the same CO variables in time and space are managed at the level of the sub-observations datasets where the user defines levels/labels, i.e., plant, quadrat, leaf, time, custom. For further information, please contact Julian Pietragalla, IBP regional deployment manager, at j.pietragalla@integratedbreeding.net.

How to deal with study details, environmental details and treatment factors without an adapted database?

If a particular database does not accommodate the recording of study metadata, a convention must be established for attaching metadata to a variable. As already mentioned, (in the section on Experimental conditions or factors), all experimental information (metadata) associated with studies must be separately recorded from the variables defined in the Trait Dictionary – we recommend using the metadata schema called Minimum Information About a Plant Phenotyping Experiment (MIAPPE; www.miappe.org).

Experimental information can be classified as:

- Study administrative details: administrative details to be tracked per study as nursery/trial name, study description, PI name, Collaborator name, etc.
- Crop management details: irrigation date, irrigation amount, fertiliser use, fertilisation date, pesticide use, planting date, harvest date, previous use of the field.
- Environmental details: all information associated with environmental characterisation, including:
    - Site condition: location name, location code, location coordinates, environment, etc.
    - Experimental condition: field, green-house, screen-house, lab, contained or confined, pot, hydroponic.
    - Soil condition: soil moisture, top-soil texture, soil pH, Phosphorus content, etc.
    - Weather information: precipitation, frost event, etc.
    - Abiotic condition: lodging, waterlogging, frost damage, etc.
    - Biotic condition: disease pressure, weeds pressure, insect damage, etc.
- Treatment factors: treatments to be applied as part of a factorial or multifactorial experimental design, like: low/mid/high nitrogen fertilisation or with/without fungicide.

The most common specific study information attached to variables are environmental details, crop management details and treatment factors. Users can utilise a convention to extend the variable's name by adding a suffix representing a specific condition, E" for environmental detail, "M" for crop management details and a "F" for treatment factors and a code assigned to each specific condition:

(Trait)_(method)_(scale)**_E, M or F (study metadata)**

It would be understood as "Trait" by "Method" in "Scale" **from/under "specific" condition**. See Table 9 for examples.

Table 9. Examples of a Variable + study metadata code

Variable + environmental detail:
LfLng_rlrMtd_cm_EFld "leaf length" by "ruler measurement method" in "cm" from/under "field" condition
Variable + crop management detail:
LfLng_rlrMtd_cm_MIrri "leaf length" by "ruler measurement method" in "cm" from/under "irrigated" condition
Variable + treatment factor:
LfLng_rlrMtd_cm_FhNF "leaf length" by "ruler measurement method" in "cm" from/under "High-Nitrogen fertilization" condition

It is highly recommended adopting a convention on the coding used for describing study metadata; see Table 10 for examples of subsampling codes.

Table 10. Study metadata codes

Specific condition	Abbreviation	Study information	Condition code
Field	Fld	Environmental detail	EFld
Green house	Gh	Environmental detail	EGh
Late planting	LP	Environmental detail	ELP
High land	hL	Environmental detail	EhL
Low land	lL	Environmental detail	ElL
Low-Nitrogen fertilisation	lNF	Treatment factor	FlNF
High-Nitrogen fertilisation	hNF	Treatment factor	FhNF
With fungicide	wFu	Treatment factor	FwFu
Without fungicide	woFu	Treatment factor	FwoFu
Irrigated field	FldIrri	Combined environmental details	EFldIrri
Field and drought	FldDrt	Combined environmental details	EFldDrt

Bibliography

Arnaud, E., Cooper, L., Shrestha, R., Menda, N., Nelson, R.T., Matteis, L., Skofic, M., Bastow, R., Jaiswal, P., Mueller, L, et al., (2012). Towards a reference Plant Trait Ontology for modeling knowledge of plant traits and phenotypes. Proceedings of the International Conference on Knowledge Engineering and Ontology Development. SciTePress, Barcelona, Spain, pp 220–225. DOI: 10.5220/0004138302200225

Arnaud, E., Laporte, M.A., Kim, S., Aubert, C., Leonelli, S., Miro, B., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., et al. (2020). The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. Patterns *1*(7), 100105.DOI: 10.1016/j.patter.2020.100105. DOI: 10.1016/j.patter.2020.100105

Cooper, L., Meier, A., Laporte, M.-A., Elser J.L., Mungall, C.J., Sinn, B.T., Cavaliere, D., Carbon, S., Dunn, N.A., Smith, B., et al., (2018). The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. Nucleic Acids Research 46: D1168–D1180. doi.org/10.1093/nar/gkx1152

Ćwiek-Kupczyńska, H., Altmann, T., Arend, D., Arnaud, E., Chen, D., Cornut, G., Fiorani, F., Frohmberg, W., Junker, A., Klukas, C., et al., (2016). Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods. 2016 Nov 9;12:44. doi: 10.1186/s13007-016-0144-4. PMID: 27843484; PMCID: PMC5103589.

Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé Yeumo, E., Emonet, V., Graybeal, J., Laporte, M.-A., Musen, M.A., Pesce, P., Larmande, P. (2018). AgroPortal: A vocabulary and ontology repository for agronomy, Computers and Electronics in Agriculture, Volume 144, 2018, p. 126-143, doi: 10.1016/j.compag.2017.10.012.

Laporte, M.-A., Valette, L., Cooper, L., Mungall, C., Meier, A., Jaiswal, P., Arnaud, E., (2016). Comparison of ontology mapping techniques to map plant trait ontologies. Proceedings of the Joint International Conference on Biological Ontology and BioCreative, Oregon State University, Oregon, 2016

Papoutsoglou, E.A., Faria, D., Arend, D., Arnaud, E., Athanasiadis, I.N., Chaves, I., Coppens, F., Cornut, G., Costa, B.V., Ćwiek-Kupczyńska, et al., (2020). Enabling reusability of plant phenomic datasets with MIAPPE 1.1. New Phytol. doi:10.1111/nph.16544

Selby, P., Abbeloos, R., Backlund, J.E., Basterrechea Salido, M., Bauchet, G., Benites-Alfaro, O.E., Birkett, C., Calaminos, V.C., Carceller, P., Cornut, et al., (2019). BrAPI consortium. BrAPI-an application programming interface for plant breeding applications. Bioinformatics. 2019 Oct 15;35(20):4147-4155. doi: 10.1093/bioinformatics/btz190. PMID: 30903186; PMCID: PMC6792114.

Shrestha, R., Matteis, L., Skofic, M., Portugal, A., McLaren, G., Hyman, G., Arnaud, E., (2012). Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. Frontiers in Plant Physiology v. 3 Article 326: doi: 10.3389/fphys.2012.00326, ISSN: 1664-042X

Annex 1- Consistency of Trait, Method and Scale for abiotic and biotic stress traits

Abiotic and Biotic stress variables

In order to maintain harmonization on abiotic and biotic stress related traits and variables few conventions must be applied. Following described convention and procedures for traits and variables creation, curators will assure a high-quality TD.
For biotic and abiotic stresses, the consistency of the variable's components trait, method and scale used is important. For example, if the trait name is disease *severity*, the variable must be expressed on a scale defining the level of *severity* and not on a scale defining level of *incidence* or *plant response*. Also, it is better to have a standardisation of the terms used.

As far as disease assessment[1] is concerned, we suggest following the definitions below for separating *Severity*, *Incidence* and *Field response*:

Severity Trait
Severity of disease and pest is the area (relative or absolute) of the sampling unit (leaf, fruit, etc.) showing symptoms, lesions or damages of disease expressed as a percentage or proportion of the total area (Nutter *et al.*, 1991).

Severity (damage, number of lesion)

The ''Severity for lodging' for example can be defined as 'the degree of the plant inclination'.

Table 11 provides useful examples.

---

[1] The general goal of disease assessment is to provide reliable estimates of the amount of disease in an area (plot, field, farm, county, region, etc.) upon the evaluation of specific symptoms and signs of a disease, at the lowest reasonable cost with known confidence (British Mycological Society).

Table 11: Examples of disease severity recording

Trait	Definition	Method	Example
Disease severity *Disorder severity* *Pest infestation severity*	Relative or absolute area of plant tissue affected by a disease, a pest, or showing disorder.	For many plant diseases and pests, 'severity' is observed as the area of plant surface affected by lesions. Often severity is represented as a proportion or percentage of plant surface affected. For pest infestation severity, it can be measured by counting the number of larvae, pups of adults for the observed affected area.	Leaf rust severity estimated as percentage
Disease lesions count *Disease unit of infection count*	The number of lesions (or other units of infection) per plant or per area of tissue per plant or per area of plant tissue. The density of disease symptoms is often considered to be a form of severity.	Usually following a count of infection units as lesions, pustules, spots, egg, larvae, adults, etc.	Leaf rust spot infection count by method 'counting pustules in last fully expanded leaf'
Disease damage *Disorder damage* *Pest damage*	An estimate of the relative intensity of injuries caused by a disease, a disorder or a pest on plants.	The method may require the presence of a protected experimental unit to compare with.	Fruit skin thrip damage estimated and reported as 'none', 'slight', 'medium' or 'highly damaged'

Incidence Trait

Incidence of a disease or disorder or pest is the frequency of damaged individuals or their parts, commonly the portion of plant units diseased. One can, for example, determine the proportion of diseased plants, per sampling area (plot, field, etc.) or the proportion of diseased units (leaves, etc.) per plant as representations of incidence (Madden and Hughes, 1999)[2].

Response/Field response

For the *response*, consider the reaction of the host to the stressor.

Table 2: Examples of Response naming

Trait	Definition
Host response	The host reaction to a stress factor (Biotic/Abiotic) is usually rated from immune/tolerant/resistance to susceptible.
Host susceptibility	Type of reaction of a host to a stress factor.

---

[2] https://apsjournals.apsnet.org/doi/pdf/10.1094/PHYTO.1999.89.11.1088

Host tolerance	A measure of a host's ability to "deal with" a biotic/abiotic stress situation. Refers to the ability of an organism to withstand or endure stressful conditions.

Entities and attributes for abiotic and biotic stress related traits

Table 13: Examples of biotic stress trait naming

Trait name	Used for
Armyworm damage	general (crop/plant) observation of the damage by Armyworm
Stem Armyworm damage	stem specific observation of the damage by Armyworm
Leaf Armyworm damage	leaf specific observation of the damage by Armyworm
Spike Armyworm damage	spike specific observation of the damage by Armyworm
Grain Armyworm damage	grain specific observation of the damage by Armyworm
Rust severity	general (crop/plant) observation of the severity by Rust
Stem Rust severity	stem specific observation of the severity by Rust
Leaf Rust severity	leaf specific observation of the severity by Rust
Spike Rust severity	spike specific observation of the severity by Rust
Weed plant number	number of weed plants per area
Weevil egg count	number of weevil eggs per sample/plot/plant/leaf
Weevil adult count	number of weevil adults per sample/plot/plant/leaf

Also, for naming biotic stress related trait **it is recommended to use as a disease/pest name acronym whenever this is widely used in the breeding community**, for example:

- FHB incidence; where FHB stands for Fusarium head blight
- RLN damage; where RLN stands for Root lesion nematode

Abiotic stress trait naming

For abiotic stress related traits, a similar convention to biotic stress trait naming is followed by avoiding the use of the word crop or plant as entity in the trait name. Trait name will require mentioning the entity when the stress observation is made in a specific organ, tissue or cell.

Table 14: Examples on abiotic stress trait naming

Trait name	Used for
Frost damage	general (crop/plant) observation of the damage by frost
Spike frost damage	a spike specific observation of the damage by frost
Lodging incidence	general (crop/plant) observation of lodging incidence
Stem lodging incidence	stem specific observation of lodging incidence
Root lodging incidence	root specific observation of lodging incidence