



# Crop Ontology: Integration of Standard Variables



grains

Léo Valette\*, Julian Pietragalla, Marie-Angélique Laporte, Agbona Afolabi, Ousmane Boukar, Steven B. Cannon, Brian W. Diers, Kate A. Dreher, Pooran M. Gaur, Alberto Fabio Guerrero, C. Tom Hash, Vilma Hualla, Drabo Inoussa, Scott R. Kalberer, Clarisse Kondombo-Barro, Shiv Kumar, Antonio Lopez-Montes, Naama Menda, Randall Nelson, Sam Ofodile, Suyash Patil, Peteti Prasad, Karthika Rajendran, Jean-Francois Rami, Abhishek Rathore, Ruaraidh Sackville Hamilton, Rosemary Shrestha, Reinhard Simon, Niaba Teme, Eva Weltzien-Rattunde and Elizabeth Arnaud\*.

\* Bioversity International, Montpellier, France. Email: e.arnaud@cgiar.org

## The Crop Ontology development

## **Trait Dictionary template**

- Developed by the Crop Ontology and the Integrated Breeding Platform to support the creation of ontologies
- Populated by breeders with their traits, observation methods and reporting scales

The **Breeding Management System** uses the Trait Dictionaries to:

- Create breeders' fieldbook
- Annotate and store breeders' data

http://integratedbreeding.net/

### The Crop Ontology online tool features:

- The **publication** of ontologies from the Trait Dictionary template or OBO files
- The **browsing** of term definitions and relations
- An **Application Programming Interface** to provide databases and web applications with ontologies in Excel, OBO, RDF, JSON formats

http://www.cropontology.org (https://github.com/bioversity/Crop-Ontology)

dehulled grains

## Improvement in phenotype annotation

To annotate phenotypes, Crop Ontology supplies breeders with:

- **Traits** i.e. the observed plant entities (e.g. leaf, grain) and attributes (e.g. color, weight)
- **Methods** i.e. the protocols to observe the trait
- **Scales** i.e. the units or categories that can express the trait observation

Thus, a plant phenotype had to be annotated with 3 identifiers for the trait, the method and the scale, respectively. Yet, breeders' fieldbooks and phenotype databases are often designed to annotate a datapoint with only one identifier. In May 2015, Crop Ontology has consequently been revised to integrate the variables. |GW100\_Meas\_g

1 Variable = {1 Trait, 1 Method, 1 Scale}

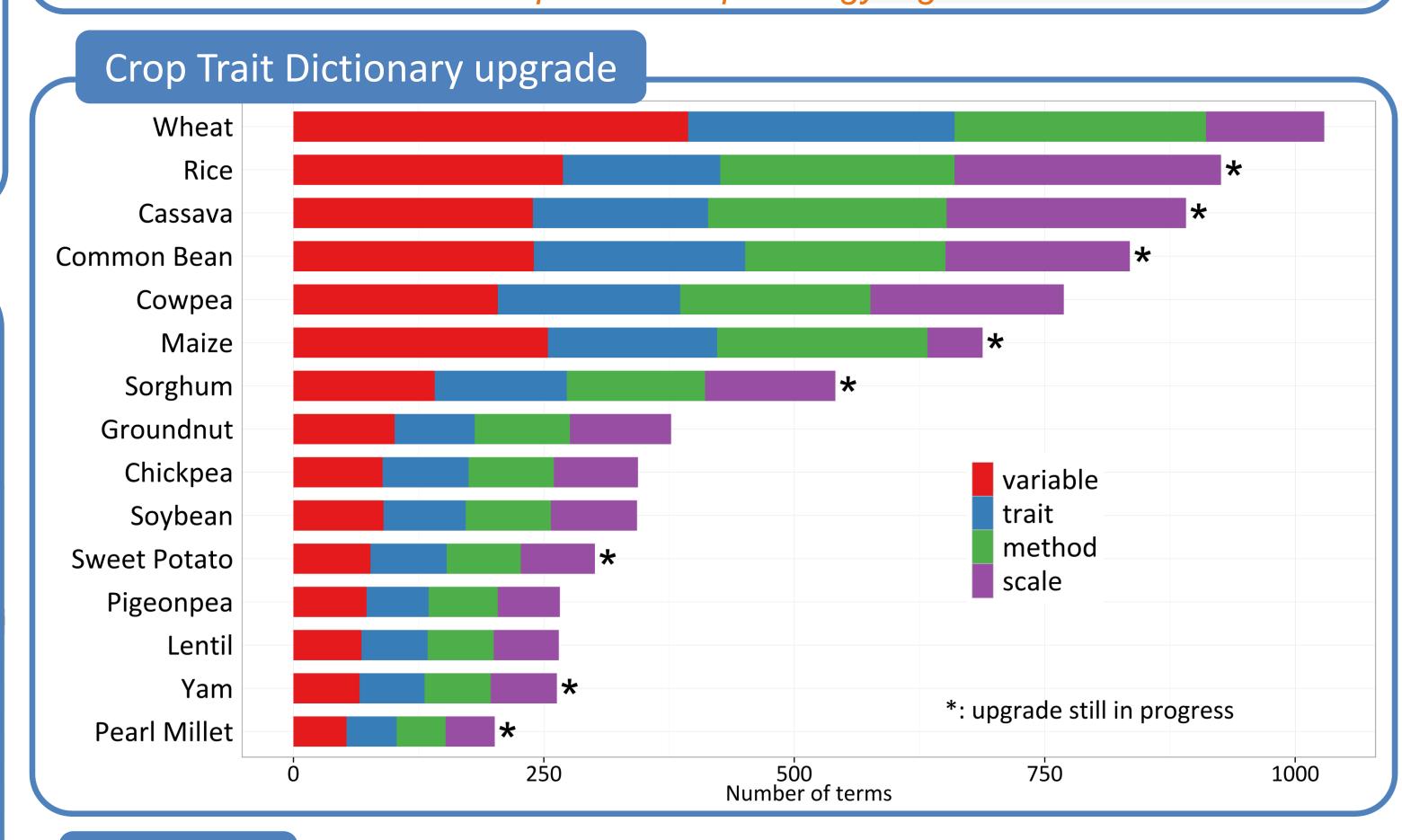
#### **Examples of variables:** Variable **Trait** Method Scale Measure the height of a plant with a ruler PltHt\_Meas\_cm cm Plant height i.e., the distance from ground Compute the average of 5 to 10 individual PltHt\_Av\_cm cm level to the tip of the plant height measurements spike PltHt\_Est\_Oto5 Visually estimate the average plant height Score Weigh a defined number of grains. Then, 100 grain weight i.e., g per 100 divide the weight by the number of grains the weight of 100

and multiply by 100

#### CO schema upgrade **Variable Trait class CO** identifier Name is\_a **Synonyms** variable of Status **Trait** Xref CO identifier Context of use Name Growth stage Definition Submitting scientist and institution Synonyms and abbrev. Method Date of submission **Entity** CO identifier Language of submission <u>Attribute</u> Name Status <u>Class</u> Xref Scale Description CO identifier Formula Name n Reference method of Class For ordinal and **Xref** scale\_of nominal scales Categories -: Addition of the variable Decimal places For numerical Lower/upper limit **Added term information** scales

#### Data annotation with variables In the Breeding Management System: **Integrated Breeding Platform MANAGE TRIALS** Manage Lists Today's tools for tomorrow's crops. **Make Crosses** Trial: AT-ZIM-2015B Manage Nurseries Manage Trials CO\_322:0000869: PH\_mes\_cm CO 322:0000684: "Estimation of Chilo Settings Germplasm Treatment F "Measurement MANAGEMENT partellus damage on a 1 to 5 scale" ► Define Measurement Details of plant height STATISTICAL ANALYSIS 🗓 🗄 Add Measurements MARKER-ASSISTE in cm" ChiloDmg\_est\_1to5 BREEDING 100 Showing 1 to 100 of 135 entries WORKFLOWS moderately (50%) dami<sup>calc\_day</sup> ChiloDmg\_est\_1to5 Action TRIA PLOT, BLOCK, REP\_ENTRY DESIGN GID noderately (50% ) dan emi-soft and semi-opad ADMINISTRATION ICML38 12168 highly (>80%) damaged nighly (>80%) damaged oft and opague grains Manage Program semi-soft and semi-opac 265 lightly (20%) damaged slightly (20%) damaged oft and opaque grains Manage User-Adde semi-soft and semi-opac 300 noderately (50% ) dama clean, no damage **Data Import Too** nighly (>80%) damaged oft and opaque grains **Tools and Crops** Significantly (70%) dama 61-70% de moderately (50%) dama noderately (50% ) dama 31-40% de Software License 41-50% de lightly (20%) damaged oft and opaque grains 11 41-50% de lean, no damage In Nextgen databases: □ CO:0000000 CGIAR cassava trait ontology

## Online visualization Quality traits Is a Language of EN Cooking time 133 submission Cotyledon colour [33] Grain protein content Variable ID CO\_339:0000254 Grain protein content analysis - Method Percentage Variable name GrPrtCent\_Comp\_Ptc Variables Context of use Evaluation trials CotCol\_Est\_1to3 CookT\_Meas\_min GrPrtCent\_Comp\_Ptc SdMcrNutCont\_Meas\_Ptc Variable status Recommended StrProtCont\_Meas\_Ptc http://www.cropontology.org



## Next Steps

- Complete the upgrade of the Trait Dictionaries
- Start upgrading the Trait Dictionaries of banana and potato
- Upload the Trait Dictionaries in the Breeding Management System 4.0
- Release curation guidelines
- Add crops: cacao, faba bean, sunflower, forage, grape, beetroot, woody plants
- Translate the Trait Dictionaries into relevant languages





is\_a CO:0000001 Agronomic trait

is a CO:0000361 Ease of Harvest



is\_a CO:0000281 Anthocyanin Pigmentation



VARIABLE\_OF CO:0000225 ease of harvest assessment 1-3

VARIABLE\_OF CO:0000103 anthocyanin pigmentation visual rating 0-3





NEXTGEN CASSAVA

















RESEARCH PROGRAM ON **Dryland Cereals**  Research Program on Global Rice Science Partnership

RESEARCH PROGRAM ON Grain Legumes

RESEARCH PROGRAM ON Maize

RESEARCH PROGRAM ON Roots, Tubers and Bananas

RESEARCH PROGRAM ON Water, Land and Ecosystems







