CROP ONTOLOGY: A reference controlled vocabulary on crop trait information for maize, wheat, chickpea, sorghum, *Musa*, potato, and rice

Rosemary Shrestha[1], Ramil Mauleon[2], Reinhard Simon[3], Jayashree Balaji[4], Mike Butterfield[4], Stephanie Channelière[5], Adriana Alercia[6], Martin Senger[2], Kevin Manansala[2], Thomas Metz[2], Guy Davenport[1], Richard Bruskiewich[2], Graham McLaren[7], Elizabeth Arnaud[5]*

[1]Crop Research Informatics Laboratory - Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Mexico; [2]Crop Research Informatics Laboratory - International Rice Research Institute (IRRI), Philippines; [3]Centro Internacional de la Papa (CIP), Peru; [4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India; [5]Bioversity International, Montpellier, France; [6]Bioversity International, Rome, Italy; [7]Generation Challenge Programme (GCP), Mexico. *Principal Investigator

The scope

- Focuses on crop-specific anatomy, development, and trait ontology for the GCP mandate crops (Figure 1).
- Provides the exact meaning of terms related to phenotypes described by crop physiologists, plant breeders, and other crop scientists.
- Consolidates across crops and adds to existing public ontology.

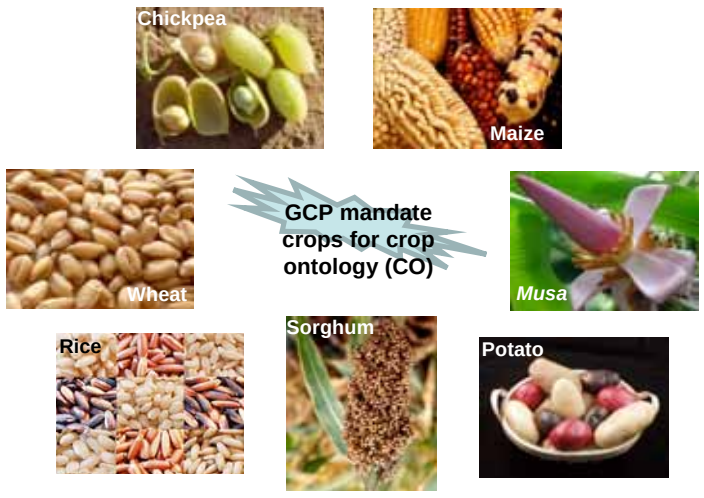


Figure 1. GCP mandate crops for developing crop-specific ontology.

Importance of deploying the crop ontology

- To facilitate the use of comparative biology infrastructure for crops by integrating diverse crop data and providing end-users and researchers with a user-friendly tool (Figure 2).
- Data related to phenotype and genotype available in CGIAR databases cannot be annotated by current ontologies because necessary terms are missing.
- Free text is used for complex descriptions of phenotypes and trait names which is almost impossible to index and search in a meaningful way.

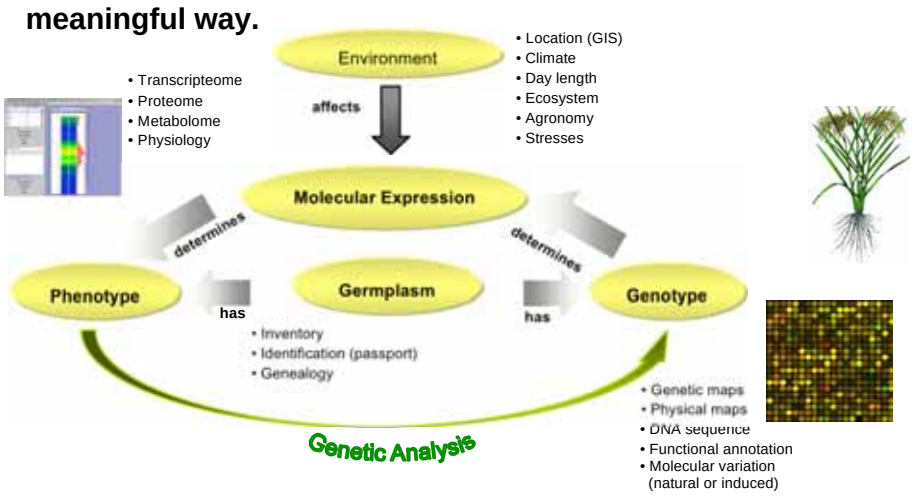


Figure 2. Integration needs across diverse crop data

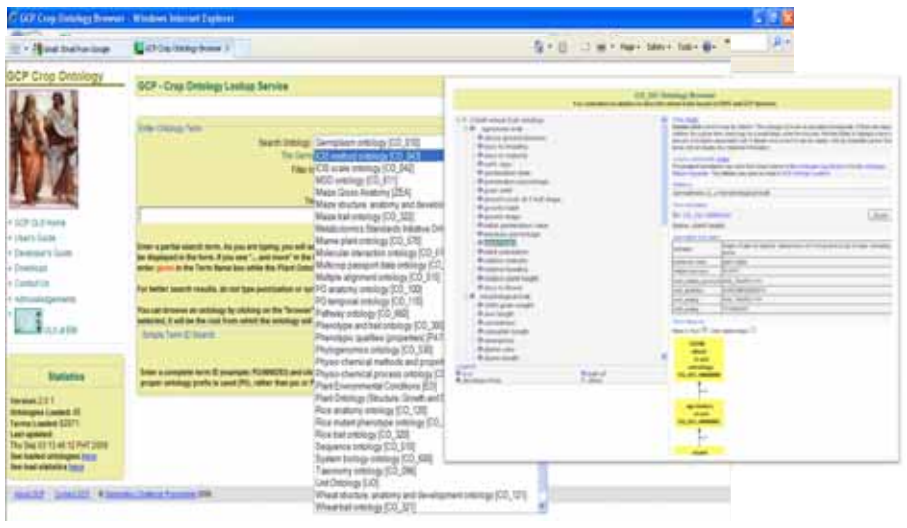
Ontology development and registration

- Decomposing complex trait names into several simple terms (Figure 3).
- Allocating simple terms into various ontology domains to properly annotate available data related to crop phenotypes (Figure 3, 4).
- Crop ontology to be registered in the OBO-Foundry.

Data annotation and curation with ontological terms

- Integrating controlled vocabulary from the crop ontology and other ontologies into the ICIS model and other Crop Information Systems (Figure 5).
- Testing Terminizer: software for assisted detection of ontological terms (http://terminizer.org/).
- Text tagging using the MS Word 2007 add-in for applying ontological terms.

Online ontology lookup service



http://koios.generationcp.org/ontology-lookup/

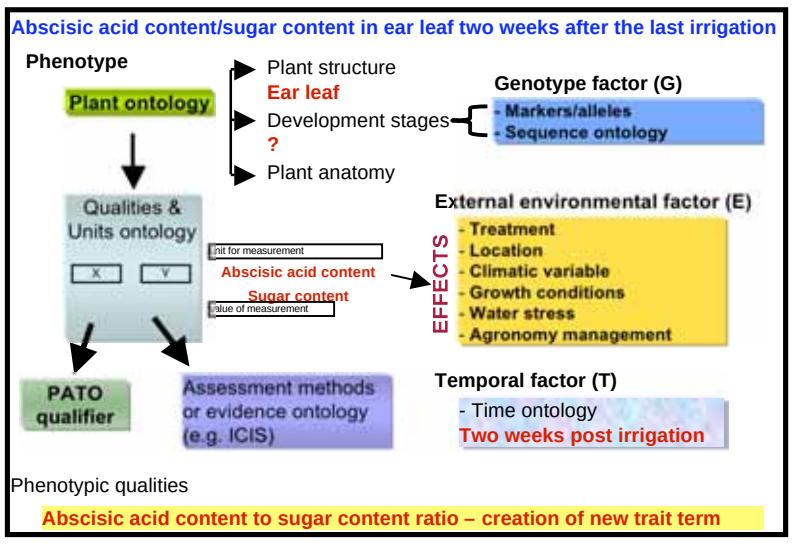


Figure 3. The concept method for dissection of complex phenotyping trait terms and several factors that affect phenotyping of crops. Red fonts represent the ontology terms decomposed from the complex term "abscisic acid content/sugar content in ear leaf two weeks after irrigation" that are embedded into respective ontology domains.

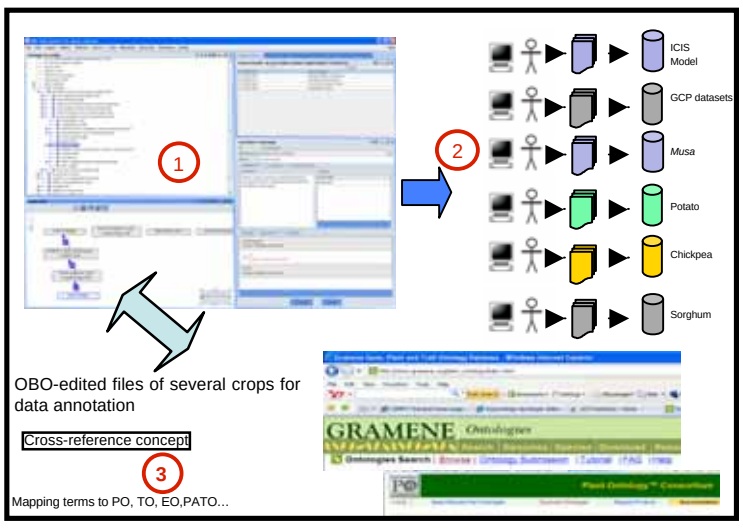


Figure 4. Ontologies development process, available data annotation and cross-referencing with related external ontologies.

Figure 5. Integration of crop ontology terms in ICIS model by adding an additional column in each database for ontology IDS.

Key actions for a sustainable crop ontology

- Develop a validation process for ontology terms involving a community of crop experts and botanists. Sustain curation by the community of practice and maintain networks with other communities developing related ontology.
- Initiate new crop ontology as necessary: e.g. cassava.
- Apply the crop ontology in the GCP phenotyping templates for annotation of representative data sets by researchers and integrate it into crop databases, e.g. the International Crop Information System (ICIS).
- Integrate ontology management into the GCP platform for distributed integration of data across the Internet.

http://pantheon.generationcp.org/ http://cropforge.org/projects/gcpontology/

http://mcclintock.generationcp.org/