DEVELOPMENT OF GCP ONTOLOGY FOR SHARING CROP INFORMATION

Rosemary Shrestha[1], Mauleon Ramil[2], Reinhard Simon[3], Jayashree Balaji[4], Stephanie Channelière[5], Martin Senger[2], Kevin Manansala[2], Thomas Metz[2], Guy Davenport[1], Richard Bruskiewich[2], Graham McLaren,[6] and 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]Generation Challenge Programme (GCP), Mexico. *Principal Investigator

INTRODUCTION

Within the Consultative Group on International Agricultural Research (CGIAR), there is an enormous volume of agriculture-related information that is generated. In the past six or seven decades the CGIAR has accumulated historic crop datasets that are related to phenotype, breeding, germplasm, pedigree, traits, etc. However, available data is not systematically formatted and annotated which complicates data management, retrieval, and accessibility. Therefore, the Generation Challenge Programme (GCP) is deploying crop ontology to semantically characterize these datasets. The GCP crop ontology and associated open source software framework is described on the GCP "Pantheon" project web site (http://www.pantheon.generationcp.org).

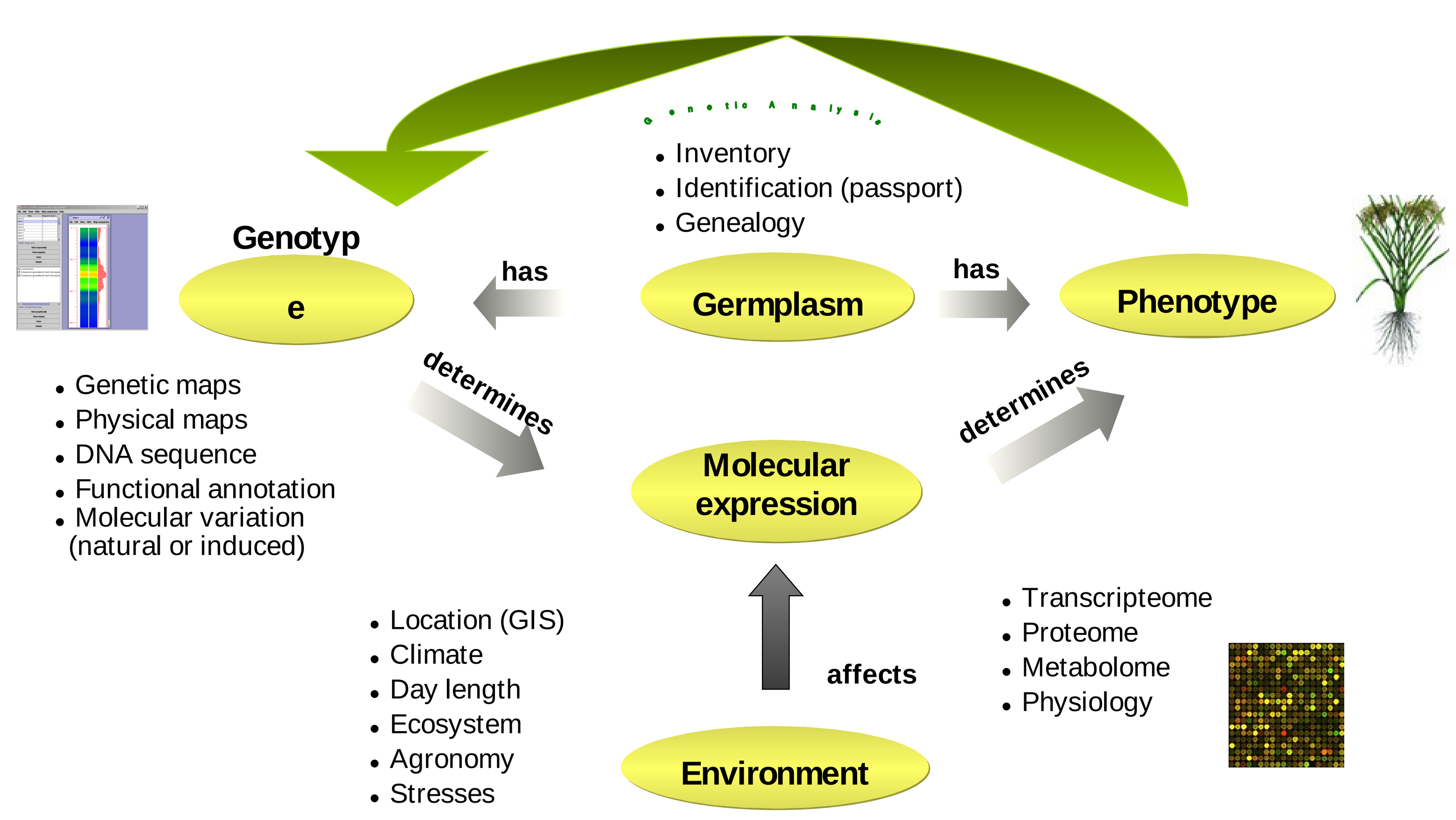

Figure 1. Integration needs across diverse crop data

I. GCP crop ontology

The GCP ontology consolidates and adds to existing public ontology by indexing each ontology within the model in a consistent fashion.

The GCP ontology focuses on crop-specific anatomy, development, and trait ontology for the GCP mandate crops (Figure 2) and provides the exact meaning of terms related to phenotypes described by crop physiologists, plant breeders, and other crop scientists.

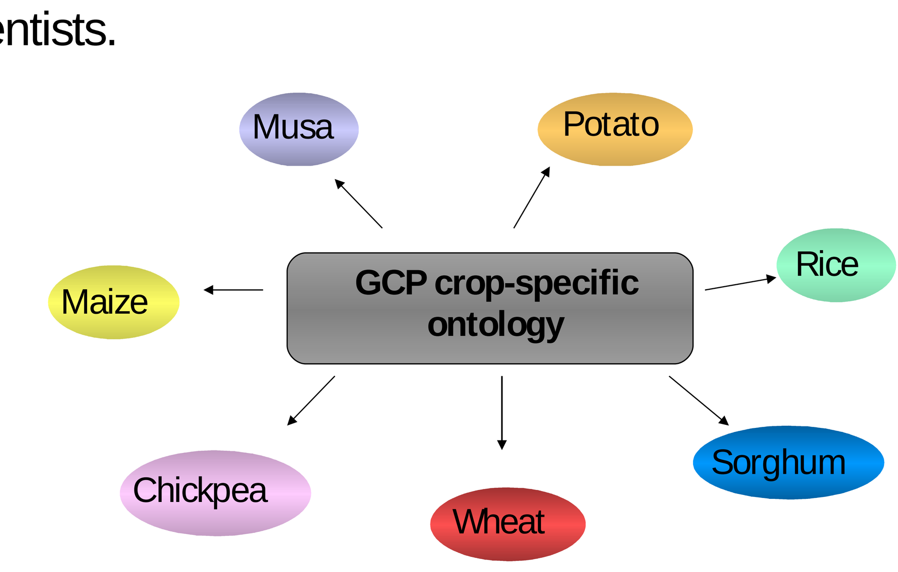

Figure 2. GCP mandate crops for developing crop-specific ontology.

II. Importance of deploying the GCP crop ontology

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 3).

a. Osmotic pressure
Osmotic pressure morning
Osmotic pressure morning well water
Osmotic pressure morning intermediate stress
Osmotic pressure morning stressed

Osmotic pressure afternoon
Osmotic pressure afternoon well water
Osmotic pressure afternoon intermediate stress
Osmotic pressure afternoon stressed

b. Grain yield
Grain yield Kg
Grain yield Kg_FieldWt
Grain yield Kg_GrainWt

Grain yield tons
Grain yield tons_FieldWt
Grain yield tons GrainWt

c. Helminthosporium sativum blotch disease resistance
Hsativum on leaf (0-5)
% Hsativum on node

% Hsativum_Flag
% Hsativum_Flag1
% Hsativum_Flag2

% Hsativum_Tiller1_Flag
% Hsativum_Tiller1_Flag1
% Hsativum_Tiller1_Flag2

% Hsativum_Tiller2_Flag
% Hsativum_Tiller2_Flag1
% Hsativum_Tiller2_Flag2

Figure 3. An example of complex trait names and variations in measurement encountered during the ontology development process.

Decomposing complex trait names into several simple terms and allocating these terms into various ontology domains (Figure 4) is necessary to properly annotate available data related to crop phenotypes.

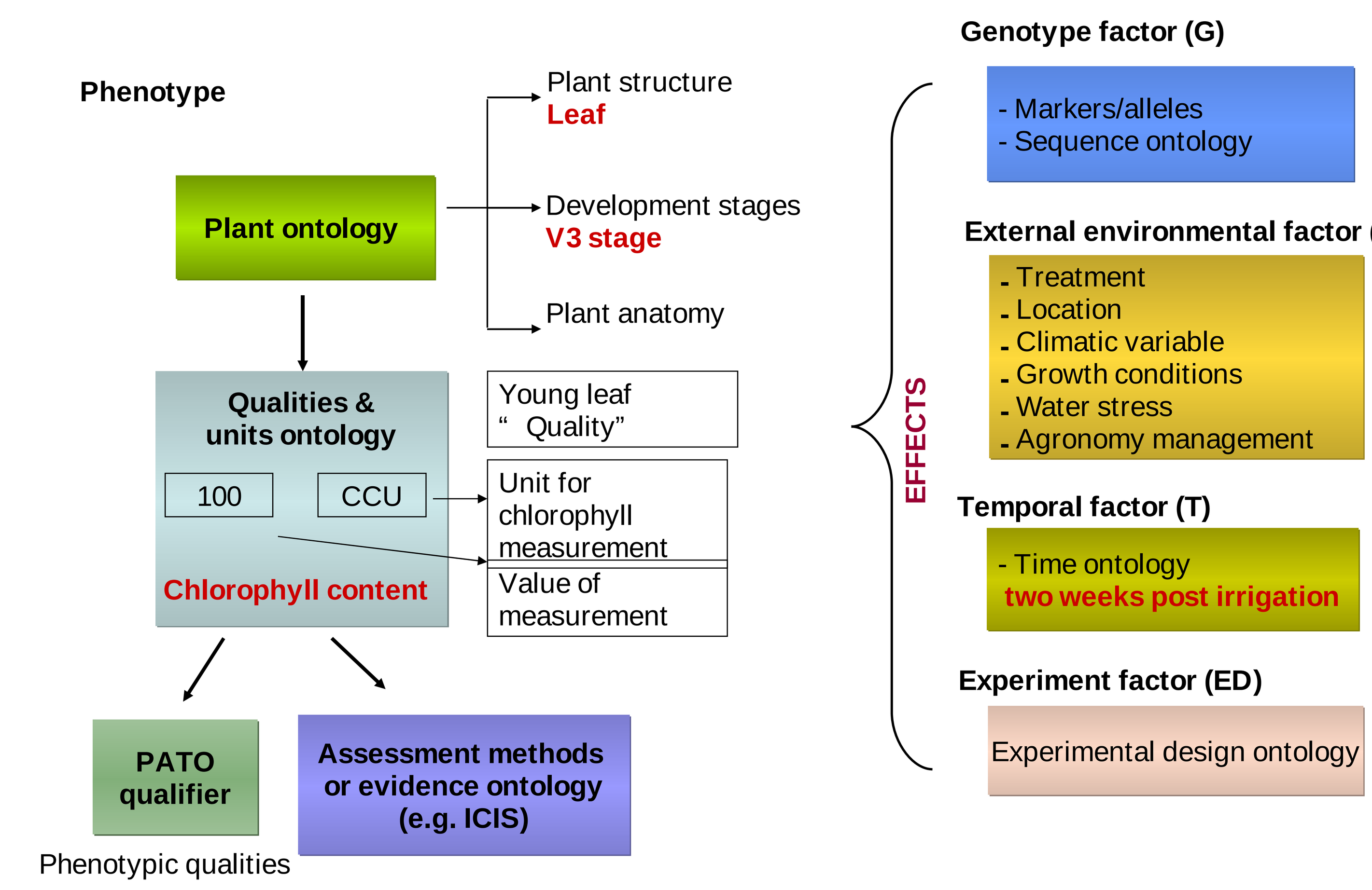

Figure 4. Schematic diagram representing the concept method for dissection of complex phenotyping trait terms and several factors that affect phenotyping of crops. Red fonts represent the ontology terms decomposed from the complex term "leaf chlorophyll content V3 stage two weeks after irrigation" that are embedded into respective ontology domains.

AVAILABILITY

The GCP ontology and platform are described on the pantheon website (http://pantheon.generationcp.org) (Figure 5).

The CropForge software project management site (http://cropforge.org/projects/gcpontology) provides both the latest releases and previous versions of ontology flat files, which describe the terms, relationships, definitions, software tools, forums, and mailing lists for communication among collaborators.

Additional documentation for scientific end users of the GCP ontology is being established (http://mcclintock.generationcp.org/).

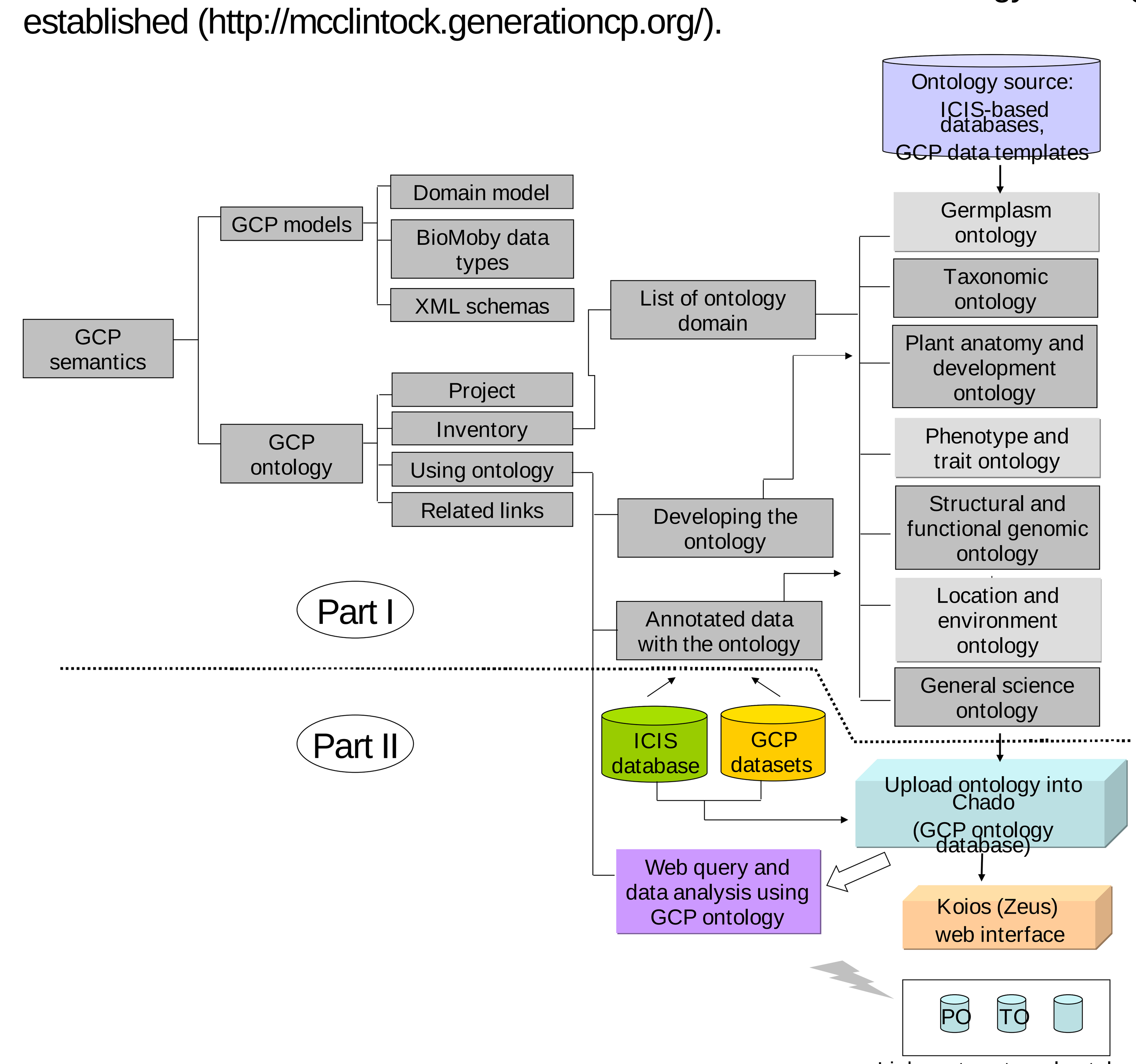

Figure 5. Overview of the GCP semantics. Part I indicates submenus under the main menu GCP semantics (http://pantheon.generationcp.org/index.php). Part II indicates an overview of the application part of crop ontology which is ongoing at present. The light rectangular box represents the major GCP target ontology domains.

FUTURE STRATEGY

Develop a validation process for ontology terms involving a community of crop experts and botanists.

Apply the GCP 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).

To provide a browser interface for searching ontology terms or specific ontology hierarchy and an annotation page for annotating GCP datasets with GCP ontology.

To integrate ontology management into the GCP platform for distributed integration of data across the Internet.

Integration with plant ontology (PO) and gramene trait ontology (GTO), to develop a common, internationally-shared crop trait and anatomy ontology.

http://pantheon.generationcp.org/index.php