

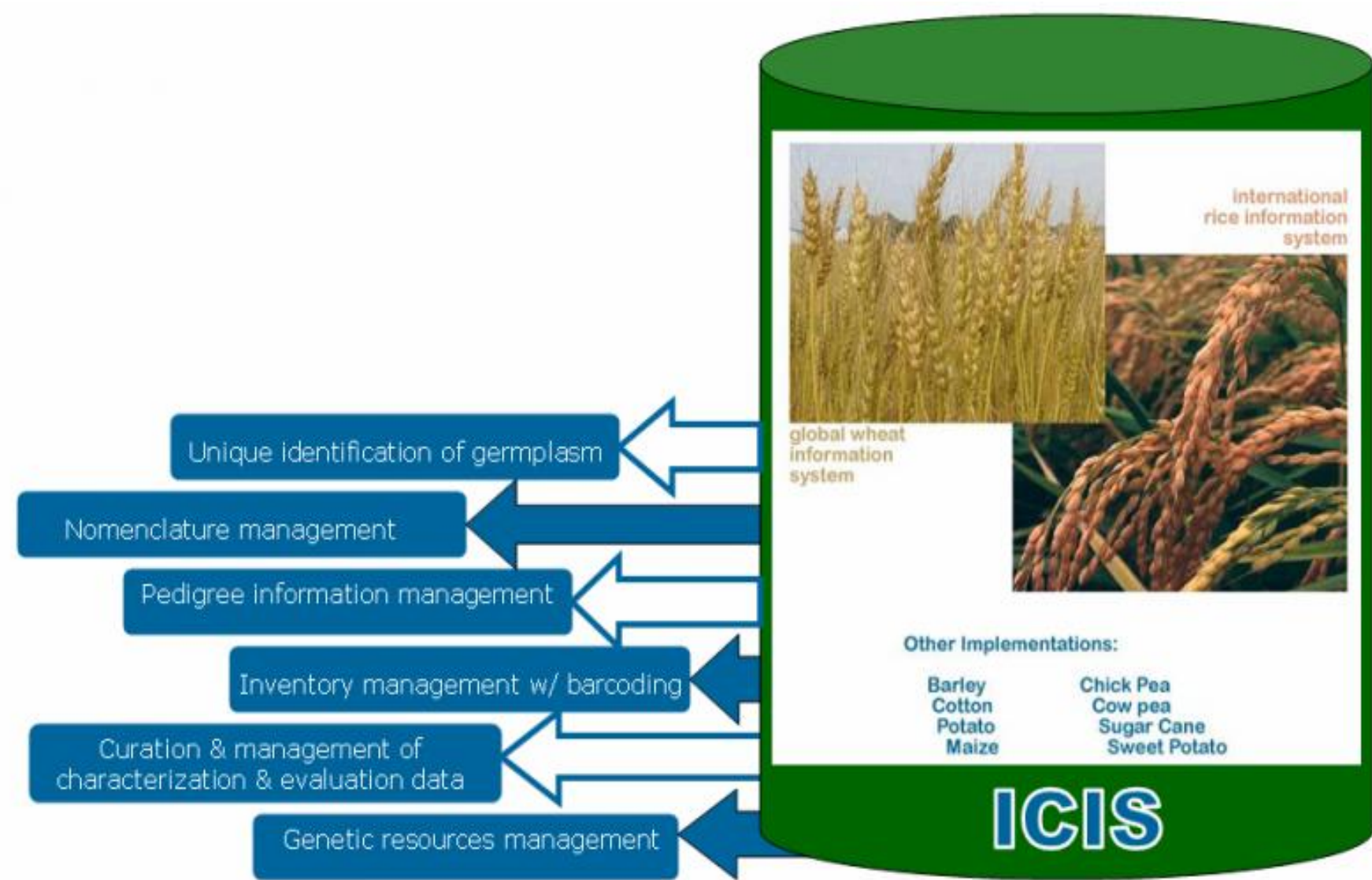
# Ontology-driven International Maize Information System (IMIS) for Phenotypic and Genotypic Data Exchange

Rosemary Shrestha<sup>1</sup>, Hector Sanchez<sup>1</sup>, Claudio Ayala<sup>1</sup>, Peter Wenzl<sup>1</sup>, and Elizabeth Arnaud<sup>2</sup>

<sup>1</sup>IRRI-CIMMYT Crop Research Informatics Laboratory (CRIL) - Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico; <sup>2</sup>Bioversity International, via dei Tre Denari, 472/a, 00057 Maccaresse, Rome, Italy

## The International Crop Information System (ICIS) Model

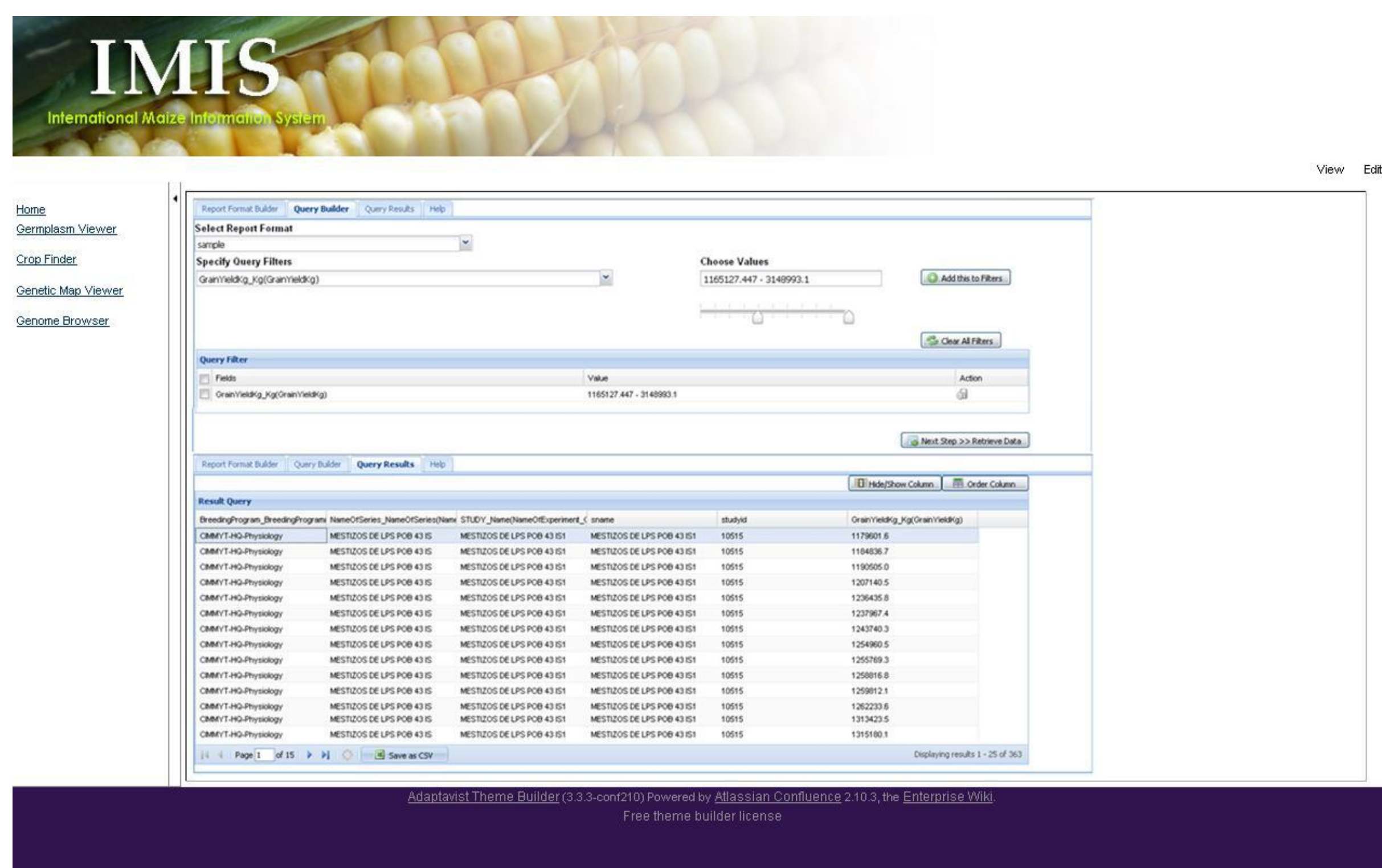
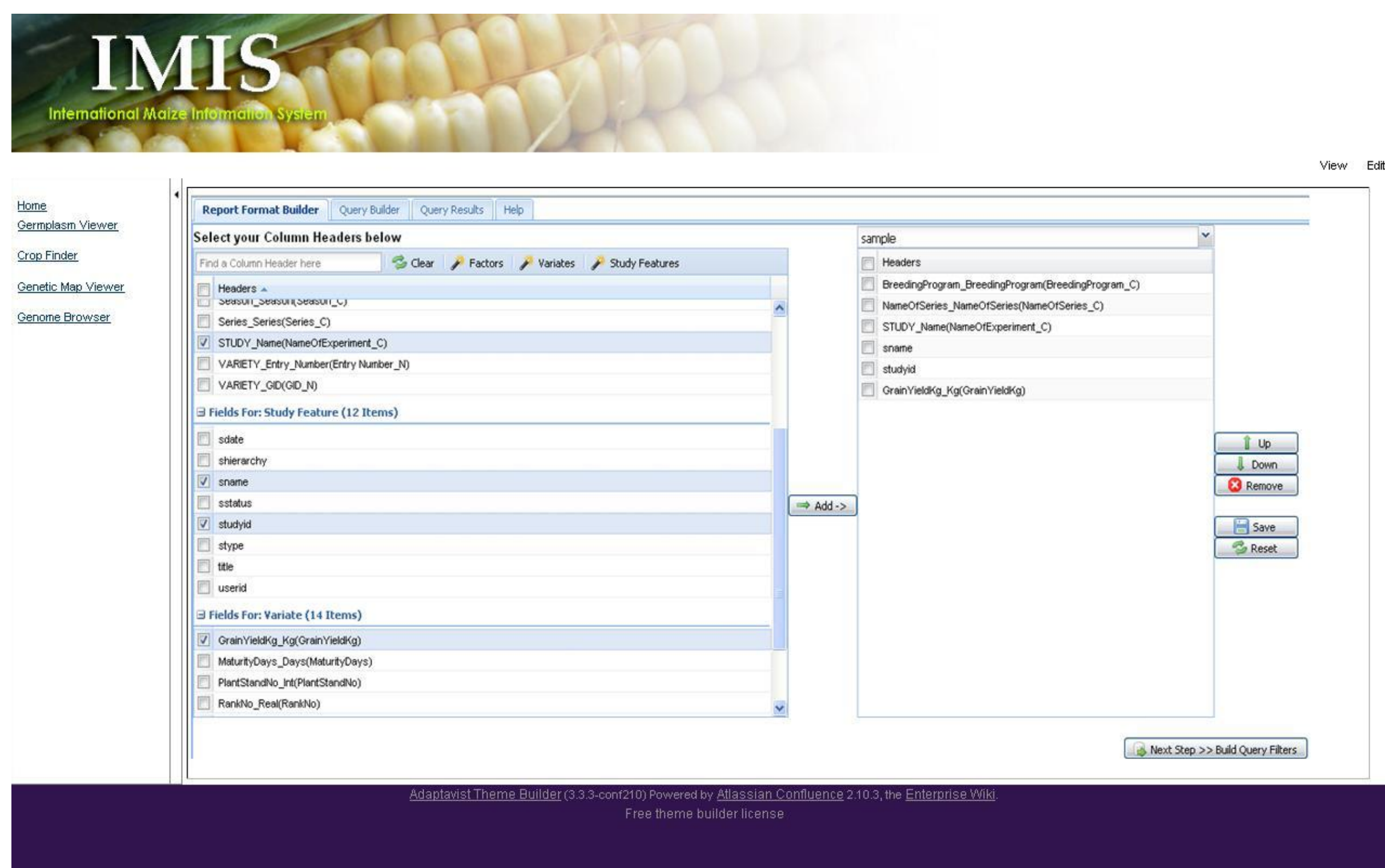
- Open-source and open-licensed generic crop information system.
- Developed by the Consultative Group on International Agricultural Research (CGIAR), national agricultural research and extension systems, agricultural research institutes, and private-sector partners.
- Designed to fully document germplasm genealogies with associated meta-data such as passport data and to accurately cross-link germplasm entries with associated experimental observations from evaluations undertaken in the field, greenhouse, or laboratory.



**Figure 1.** International Crop Information System (ICIS) model for addressing the problem of ambiguous germplasm identification, difficulty in tracing pedigree and lack of integration between genetic resources, breeding, evaluation, utilization and management trait data.

## The International Maize Information System (IMIS)

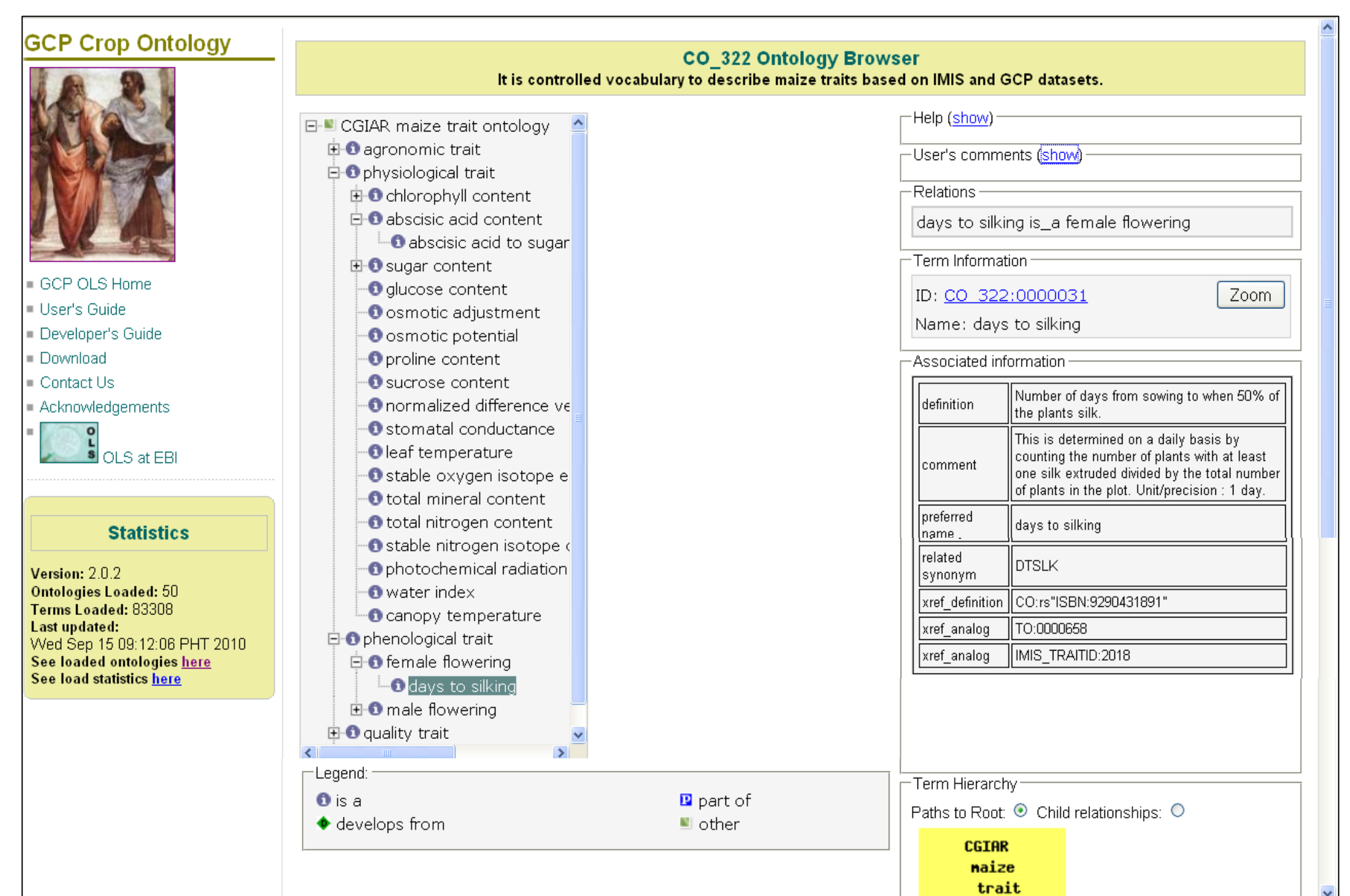
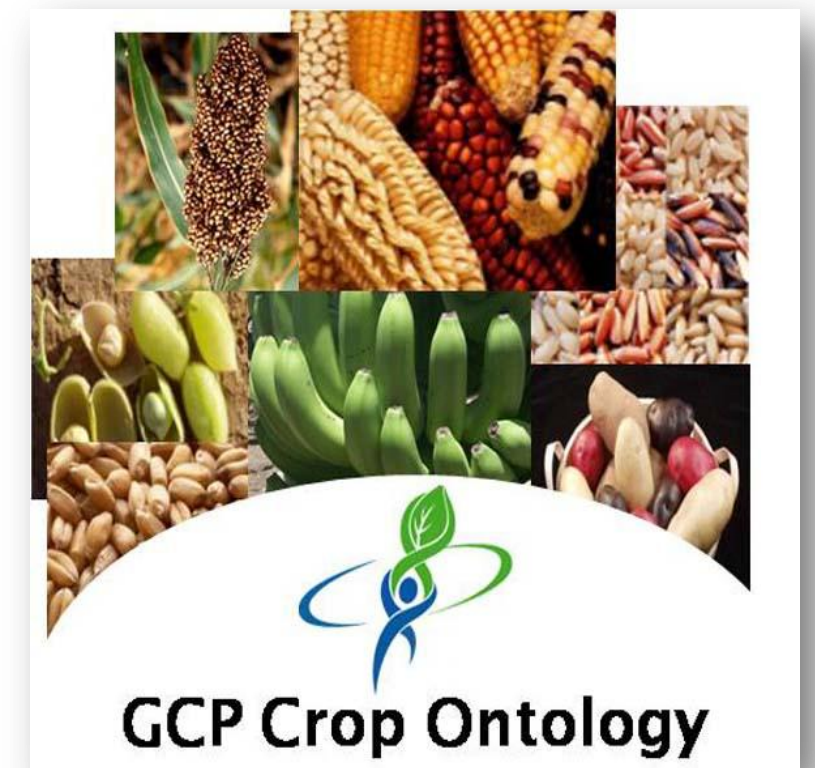
- The IMIS is the public maize implementation of ICIS and CIMMYT's flagship germplasm database.
- Contains about 33,000 germplasm ID (GID) entries with thousands of associated data points in hundreds of experimental studies, including many phenotypic observations and growing numbers of genotypic measurements.
- Query interface is available for researchers to search germplasm and associated maize breeding data.
- Many of these data sets are published in IMIS and initiated sharing data with collaborating databases such as MaizeGDB.



**Figure 2.** A screen captures of the International Maize Information System (IMIS) query interface for searching genotypic or phenotypic data. In the figures, query steps and result for the trait data "grain yield" are shown in detail.

## The Crop Ontology (CO)

- The CO provides controlled vocabulary sets for several economically important plant species and facilitates biocurators working in genebanks of plant genetic resources (PGR) and crop breeding data curation and annotation.
- It also facilitates data exchange within and between globally distributed databases, developing user-friendly query interface for searching crop data and allowing researchers to perform comparative studies across crops.
- The maize-trait ontology is developed based on the traits that are in the Data management system (DMS) trait table of IMIS database along with the traits that are described in maize crop descriptor and GCP datasets.



**Figure 3.** A screen capture of the GCP Crop Ontology Lookup Service (OLS) browsing maize trait ontology. The information for the highlighted trait "Days to silking" is shown such as trait ID, trait description, synonym.

## Implementation of the CO in the IMIS database

- The IMIS traits are mapped to those of maize-trait ontology and linked them to ontology databases by term (trait) identifier.
- Similarly, the scales for trait measurement, locations for germplasm collection [which are listed in the table of Genealogy Management System (GMS) of IMIS] and experimental sites are mapped to the generic ontologies such as scale ontology, gazetteer ontology.

TRAITID	TRITNAME	Ontology	SCTYPE	SCNAME	Scales OLS
2022	CommonRust1_5	CO_322000085	C	(1-10)	CO_042200048
2023	IronDeficiency1_5	CO_322000086	C	(1-5)	CO_042200049
2024	GrayLeafSpot1_5	CO_322000088	C	(1-5)	CO_042200057
2027	CurvulariaLeafSpot1_5	CO_322000089	C	(1-5)	CO_042200057
2048	Phaeosphaeria1_5	CO_322000090	C	(1-5)	CO_042200057
2046	TarsoComplex1_5	CO_322000091	C	(1-5)	CO_042200057
2044	LeafBlightTurfum1_5	CO_322000092	C	(1-5)	CO_042200057
2124	SeedlingDiseaser1_5	CO_322000093	C	(1-5)	CO_042200057
2129	VirusMaizeDwarfing1_5	CO_322000094	C	(1-5)	CO_042200057
2125	VirusMaizeStunt1_5	CO_322000095	C	(1-5)	CO_042200057
2126	VirusMaizeStunt2_5	CO_322000096	C	(1-5)	CO_042200057
2128	VirusMaizeStunt3_5	CO_322000097	C	(1-5)	CO_042200057
2129	VirusMaizeStunt4_5	CO_322000098	C	(1-5)	CO_042200057
2069	VirusMaizeStunt5_5	CO_322000099	C	(1-5)	CO_042200057
2070	VirusMaizeStunt6_5	CO_322000100	C	(1-5)	CO_042200057
2127	VirusMaizeStunt7_5	CO_322000101	C	(1-5)	CO_042200057
2128	VirusMaizeStunt8_5	CO_322000102	C	(1-5)	CO_042200057
2104	TryptophanPer	CO_322000103	C	Percentage	CO_042200057
2099	LysinePer	CO_322000104	C	Percentage	CO_042200057
2123	TryptophanPer_Endospem	CO_322000105	D	Real	CO_042200046
2068	TryptophanPer_WholeGrain	CO_322000107	C	Percentage	CO_042200057
2044	LysinePer_WholeGrain	CO_322000108	C	Percentage	CO_042200057
2099	LysinePer_Endospem	CO_322000109	D	PPM	CO_042200057
2093	Iron_ppm	CO_322000110	D	PPM	CO_042200057
2141	Iron_ppm	CO_322000111	D	Integer	CO_042200047
2096	Selectioendex	CO_322000112	D	Percentage	CO_042200057
2100	ProteinPer	CO_322000121	C	Percentage	CO_042200057
2050	ProteinPer_WholeGrain	CO_322000122	C	Percentage	CO_042200057
2106	ProteinPer_Endospem	CO_322000123	C	Percentage	CO_042200057
2107	ProteinPer_WholeGrain	CO_322000125	D	Percentage	CO_042200057
2048	MaizeStunt1_5	CO_322000126	C	(1-5)	CO_042200057
2046	MaizeStunt2_5	CO_322000127	C	(1-5)	CO_042200057
2002	ApHn1_5	CO_322000128	C	(1-5)	CO_042200057
2046	MaizeStunt3_5	CO_322000129	C	(1-5)	CO_042200057
2071	WheatYielding	CO_322000130	D	Real	CO_042200046
2072	WheatYieldingPer	CO_322000131	C	Percentage	CO_042200057
2057	StomatalConduc	CO_322000132	C	Integer	CO_042200047
2044	StomatalConduc	CO_322000134	D	Real	CO_042200046
2039	InfraredThermo	CO_322000142	D	Real	CO_042200046
2087	GrainC1_5	CO_322000144	C	(1-5)	CO_042200057
2101	RelativeGrainYieldPer	CO_322000147	C	Percentage	CO_042200057

**Figure 4.** The Data Management System (DMS) table of the IMIS showing mapping traits to the maize\_trait ontology and also IMIS trait scales to the ICIS scale ontology. The curation is expended for linking phenotypic data to genotypic data by mapping the IMIS traits with molecular data such as QTL to the maize\_trait ontology

## REFERENCES

- Bruskiewich R, Senger M, Davenport G, et al. 2008. The Generation Challenge Programme Platform: Semantic Standards and Workbench for Crop Science. International Journal of Plant Genomics Article ID 369601, 6 pp.
- McLaren CG, Bruskiewich RM, Portugal AM et al. 2005. The International Rice Information System. A Platform for Meta-Analysis of Rice Crop Data. Plant Physiol. 139:637-642.
- Shrestha R, Arnaud E, Mauleon R, et al. 2010. Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature. AoB Plants. plq008.

## AVAILABILITY

- The ICIS information is available at <http://www.icis.cgiar.org>.
- The IMIS query interface is available at <http://imis.cimmyt.org/> for searching maize data.
- The GCP crop ontology and platform are described on the pantheon website at <http://pantheon.generationcp.org>.
- The GCP Lookup Service is available at <http://cropontology.org> to browse ontologies.
- The CropForge software project management site is available at <http://cropforge.org/projects/gcpontology> for the collaborators and users.