Ontology mapping for wheat trait information management

Claire Nédellec[1], Sophie Aubin[2], Clara Sauvion[1,5], Liliana Ibanescu[3], Sonia Bravo[2], Jacques Le Gouis[4], Thierry Marcel[5], Cyril Pommier[6], Robert Bossy[1], Michaël Alaux[6]

1. MaIAGE, INRAE, Université Paris-Saclay, Jouy-en-Josas, France
2. DiPSO, INRAE, Angers, France
3. MIA-Paris-Saclay, INRAE, AgroParisTech, Université Paris-Saclay, Palaiseau, France
4. BIOGER, INRAE, Université Paris-Saclay, Thiverval-Grignon, France
5. GDEC, INRAE, Clermont-Ferrand, France
6. URGI, INRAE, Université Paris-Saclay, Versailles, France

Our interoperability ambition for the WheatIS knowledge base system focuses on the phenotypic values observed in-field or in controlled conditions and described in the scientific literature. They differ by their scope and types. The document traits mostly qualify the general properties of wheat varieties or cultivars. On the other hand, experimental data qualify given measurable properties of the plant within a limited spatial and temporal scope that need to be aggregated and experimentally confirmed to derive the general properties of the observed wheat variety.

In the WheatIS framework, the two types of data are indexed by two different ontologies, respectively, the *Wheat Trait and Phenotype Ontology* (WTO; http://agroportal.lirmm.fr/ontologies/WHEATPHENOTYPE) (Nédellec et al., 2019) and the *Wheat Crop Ontology* (CO_321; https://cropontology.org/ontology/CO_321).

Irrespective of the differences in ontology structures and the resulting inference choices, we encountered differences in the trait classes stemming from the two ontology purposes and expert disagreements that prevent the use of automatic mapping tools.

The highly technical nature of the phenotyping field calls for the participation of two cutting-edge experts and five knowledge engineers who elicit the knowledge of the domain experts and represent it according to the formal framework. Characterizing the relationship between the two ontology sets of classes involves broad expertise not only in phenotyping measurement but also in plant biology, physiology, pathology, agronomy, and food processing.

We manually defined a set of mappings that meet the requirements of the target knowledge base. Approximated and multiple mappings were needed to deal with the ontology differences in class granularity and disagreements on pathogen agents causing diseases. A set of rules and a mapping typology formalize the mapping principles for the sake of consistency and traceability. Table 1 shows an example of mappings. We represent the mapping dataset in the *Simple Standard for Sharing* Ontological *Mappings* (SSOM) TSV-based and made it available on the National Research Data portal. 435 mappings involve 262 WTO classes and 366 CO_321 classes, among which 226 WTO classes and 308 CO_321 classes are mapped through formal equivalence or subsumption inference. The remaining mappings represent complex relationships.

Table 1. Example of mapping between WTO and CO_321 classes

WTO_ID	WTO class name(PrefLabel)	WTO synonyms	Authors	Mapping type	CO_321 ID	CO_321 trait name	CO321 definition	Reason for alignement	Rules
0000484	resistance to Helminthosporium Leaf Blight	resistance to HLB	Clara et al.	1.4	0000686	Helminthosporium species severity	The disease severity in the plants caused by the agent Helminthosporium species.	Helminthosporium leaf blight is a general term for several diseases caused by several fungi formerly known as Helminthosporium spp.	Bio_Severity
0000488	resistance to Sclerotium Wilt	resistance to Southern Blight, resistance to Sclerotium rolfsii, resistance to Corticum rolfsii	Clara et al.	1.4, 1.5	0000115	Southern blight plant response	Southern blight response in the plants caused by the agent Corticum rolfsii (Sclerotium rolfsii).	Sclerotium wilt is a synonym of Southern blight according to WTO. Same fungal pathogen name.	Bio_Plant_resp

0000489	resistance to Sharp Eyespot	resistance to Rhizoctonia cerealis, resistance to Ceratobasidium cereale	Clara et al.	1.4	0000683	Sharp eyespot incidence	Main shoots are assessed for sharp eyespot disease.		Bio_Incidence