The Agronomy Ontology (AgrO): A semantic layer to standardize agronomic data in farming systems research and development

Céline AUBERT[a] *, Marie-Angélique LAPORTE[b], Elizabeth ARNAUD[b], Medha DEVARE[a]

[a]*IFPRI, Washington, DC, USA;* [b]*Alliance of Bioversity International and CIAT, Montpellier, France*

*Corresponding Author: c.aubert@cgiar.org

Keywords. Ontology, agronomy, farming system, fieldbook, standards, semantics, agriculture, AgrO, interoperability

To fully understand the implications of varying practices within cropping systems and derive insights, it is often necessary to pull together data across different disciplinary domains. For example, data on field management, soil, weather and crop phenotypes may need to be aggregated to assess performance of a particular crop under different management interventions. However, agronomic data are often collected, described, and stored in inconsistent ways, impeding data comparison, mining, interpretation, and reuse (Devare *et al.*, 2016). The use of semantic standards for the annotation of data variables plays a key role in addressing these challenges. The Agronomy Ontology (AgrO) is one such standard enabling the description of data variables in agronomic trials. Created largely through support from the CGIAR Platform for Big Data in Agriculture, AgrO forms the foundation of the Agronomy Field Information Management System (AgroFIMS). Thus, AgroFIMS allows digital collection of data which is "born" already semantically standardized.

AgrO is an open-source ontology with concepts modeled on field operations generally important in agronomic trials. As depicted in the example shown in Figure 1, variables are classified under three main classes: Experiment information, agricultural event description, and experiment measurement. Under these main classes, agronomic concepts are organized in ways that make sense to those envisioned to be the primary users of AgrO -- by field operations, or field management practices, for instance.
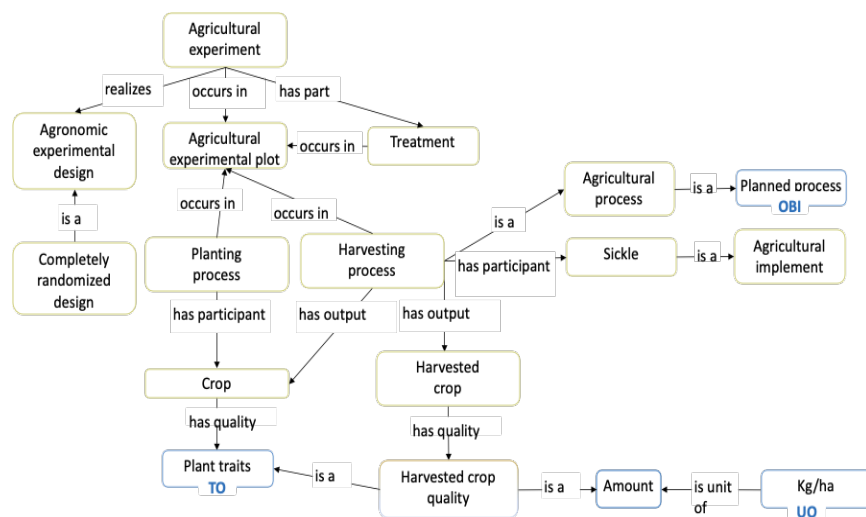

Figure 1. Structure of AgrO – an exemplar.

The ontology's 1685 concepts have been created based on their relevance to agronomic trials, and through regular consultation with a Community of Practice composed of agronomists and data stewards from across CGIAR Centers, and partner institutions like CIRAD and INRA in France; the University of Florida, and UC Davis in the US; and Rothamsted Research in the UK. AgrO is openly available on the project's GitHub repository (https://github.com/AgriculturalSemantics/agro).

The backbone of AgrO is based on formal categories from the Basic Formal Ontology (BFO) shared by the ontologies of the Open Biological and Biomedical Ontology (OBO) Foundry family (http://www.obofoundry.org/). In further efforts towards such interoperability, AgrO reuses and builds on existing terms from other ontologies of the OBO Foundry to minimize term proliferation and duplication (Figure 2). FoodOn's top-level semantics are also synchronized with AgrO (Dooley *et al.*, 2018).

Figure 2. Ontologies reused in AgrO.

Plans for the future include continuation of efforts to validate and improve AgrO through its use to describe data variables in legacy or non-digitally collected data, and via AgroFIMS. A key aspiration also centers around the recognition that agronomy research is driven not only by efforts to enhance the productivity, but also profitability of farming systems – requiring the inclusion of socioeconomic concepts as well. By providing key semantics to describe operations and data collected in field studies relevant to agricultural production, the Agronomy Ontology is one building block of the farm to fork digital highway. With the use of AgrO to annotate individual datasets or AgroFIMS to generate harmonized data, it is anticipated that the pool of semantically enabled agronomic data will grow, allowing more efficient data mining and application of machine learning tools for analysis.

References

Devare, M., C. Aubert, M.A. Laporte, L. Valette, E. Arnaud, P.L. Buttigieg. 2016. Data-driven Agricultural Research for Development: A Need for Data Harmonization Via Semantics. Proceedings of the Joint International Conference on Biological Ontology and BioCreative, Corvallis, Oregon, United States, August 1-4, 2016. http://ceur-ws.org/Vol-1747/IT205_ICBO2016.pdf

Dooley, D.M., Griffiths, E.J., Gosal, G.S. *et al.* 2018. FoodOn: A harmonized food ontology to increase global food traceability, quality control and data integration. *npj Sci Food* **2,** 23. https://doi.org/10.1038/s41538-018-0032-6