

## A dictionary of variables to harmonize data from agro-ecological experiments on sugarcane

AUZOUX S<sup>1</sup>, CHRISTINA M<sup>1</sup>, GOEBEL R<sup>1</sup>, MANSUY A<sup>2</sup>, MARION D<sup>2</sup>

<sup>1</sup>CIRAD, Montpellier, France; <sup>2</sup>eRcane, La Bretagne, La Réunion

E-mail: [sandrine.auzoux@cirad.fr](mailto:sandrine.auzoux@cirad.fr)

### Abstract

With changing circumstances, sugarcane industries are adopting agro-ecological approaches to reduce environmental impacts of sugarcane production. These include fallow and/or companion crops cultivated in association with cane as alternatives to chemical weed control, organic material as substitutes for mineral fertilizer, and new cropping practices such as reduced tillage, crop rotation and crop diversification. Agro-ecological experiments generate many heterogeneous datasets that integrate functional biodiversity and multiple levels of spatial and temporal scales of observations. Each has its own structure. Observed variables include homonyms and synonyms. Units of measurement and methods of measurement are different or not specified. Plant and pest are identified and named in different ways.

To solve the problem of organization and standardization of these datasets, a controlled vocabulary was used for annotating data, increasing the accuracy of terms used by agronomists and harmonizing data capture. A dictionary of variables was developed using terms from reference ontologies such as plant physiology, anatomy, morphology, environmental science, ecosystems, habitats, agronomy, soil science, etc. Experts were involved throughout the process of creating the dictionary of variables. This included awareness of their data, details of the way they usually managed their data, and feedback from groups and individuals. The aim was to develop the dictionary of variables into a tool that could be integrated into daily practices. The main challenge was to reconcile agronomic and plant knowledge with data engineering, and to define a common language for close collaboration between agronomists and data scientists.

The structure of the dictionary of variables is based on the trait dictionary template used to develop crop ontology. A variable is defined by the combination of an entity, a trait, a method and a unit of measurement. The variables are classified by type (experimental design, plant, biological pest, soil, weather, crop management) and sub-type to describe the type more precisely (sub-type "physiology traits" for type "plant").

The dictionary of variables is continually being enriched by experts. It is used to support agronomists' field-books and to define the structure of agro-ecological experiment datasets. The resulting standardized datasets are made accessible to interested users and are easily reusable for modelling or statistical analyses. The dictionary of variables is useful for designing the structure of agro-ecological databases and especially to ensure interoperability with other agronomic databases. It can also be used as a tool to enhance and share expert knowledge.

**Keywords:** metadata, ontology, interoperability, heterogeneity