Bottom-up curation of terminology for experimental variables: the Ontology of Experimental Variables and Values (OoEVV)

Gully A.P.C. Burns, Jessica A. Turner*

University of Southern California; ISI; Mind Research Network; UNM

1 INTRODUCTION

The challenges of developing an effective biomedical knowledge representation (KR) include (a) expressing the correct logic of domain-specific concepts in a reusable way, (b) excluding extraneous ontological commitments that complicate its use for a given task, (c) ensuring that the curation process scales, and (d) guaranteeing that it is understandable to domain experts. We are developing an 'ontology design pattern' (ODP, [1]) called the 'Ontology of Experimental Variables and Values' (OoEVV) as a pragmatic, reusable modular component based on earlier work [2], to capture definitions of variables and their values to be reused in other applications. We describe the design of the approach and tools we have constructed to support its use.

2 FORMULATION AND TOOLS

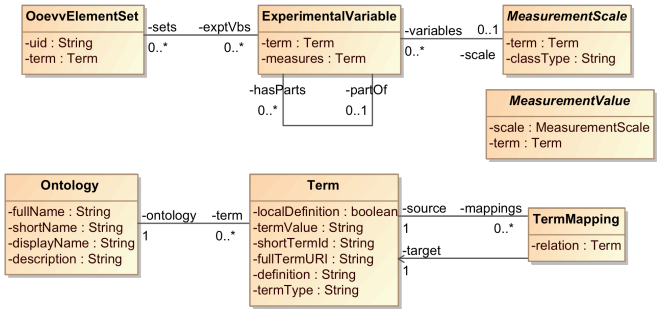

Fig. 1. The high level design of OoEVV

Our formulation uses UML (Fig. 1). This includes **`Term`**, **`TermMapping`**, and **`Ontology`** classes to both import externally-defined ontology terms and to generate our own for our internal structures. OoEVV models consist of several elements: an **`OoevvvElementSet`** is a holder for all components relevant to a specific domain. This set contains **`ExperimentalVariable`** elements that each measure a defined *quality* **`Term`** (that may be drawn from any available external ontology). The mathematical properties of the variable's values are specified by its **`MeasurementScale`** to provide a framework for managing the computations that may be performed on data measured with a particular variable. Scale subtypes include 'Binary', 'Decimal', 'Integer', 'Nominal', 'Ordinal' and 'Hierarchical' (to describe a hierarchical taxonomy). This UML-based structure is extensible to provide a practical methodology for developers to construct specialized data representations. OoEVV differentiates between 'Qualities', 'Variables' and 'Measurement-Scales' as its central core elements and provides a mechanism to capture different variables that measure the same thing. An example of this is 'Handedness' (PATO:0002201): a nominal scale might include three categories ('left', 'right' and 'ambidextrous', as with the children of the PATO term shown) whereas the 'Edinburgh Handedness Inventory' (obo:OBI_0001001) is a numerical scale providing a score derived from a questionnaire ranging from -100 to +100 (obo:OBI_0000991). We support best practices by providing good definitions and documentation, reuse of terminology wherever possible and compatibility with existing ontology formats and standards. We provide a practical curation toolset that may be used by domain experts to develop a structured lightweight terminology that may be accessed via the NCBO's Bioportal.

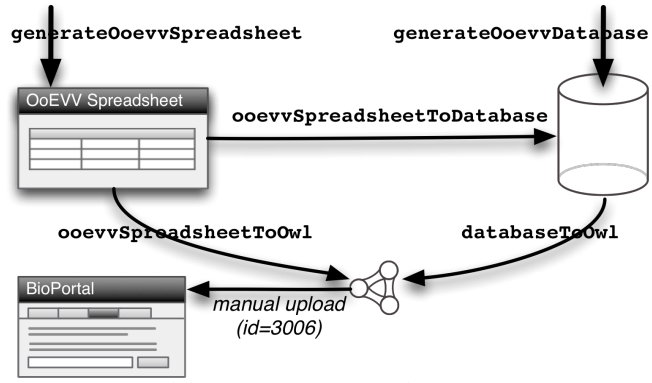

Figure 2: data flow and commands for the toolkit.

We provide a command-line application (the 'OoEVV' toolkit, see http://www.isi.edu/projects/ooevv/overview) that uses spreadsheets to curate terminology and generate an OWL file that can then be uploaded to BioPortal.

REFERENCES

[1] A. Gangemi and V. Presutti, (2009) "Ontology Design Patterns," in Handbook of Ontologies, S. Staab and R. Studer, Eds.

[2] T. Russ, C. Ramakrishnan, E. Hovy, M. Bota, and G. Burns (2011), "Knowledge Engineering Tools for Reasoning with Scientific Observations and Interpretations: a Neural Connectivity Use Case," BMC Bioinformatics, 12,:351.