An Ontology Blueprint for Constructing Qualitative and Quantitative Scientific Variables

Maria Stoica$^{1[0000-0002-6612-3439]}$ and Scott D. Peckham$^{1[0000-0002-1373-2396]}$

Institute of Arctic and Alpine Research
University of Colorado, Boulder 80309, USA
`maria.stoica@colorado.edu`

Abstract. This work presents an ongoing effort to develop simple ontological design patterns for describing scientific variables with a high level of specificity in resource description format (RDF). The application of the ontology design patterns discussed here were used to create a variables ontology for the geosciences. The long-term aim of this work is to develop an ontological blueprint for automated ontology generation from a corpus. Such ontologies can be used for semantic mediation in automated scientific workflows and semantic alignment of content in heterogeneous resources.

Keywords: ontology design pattern · scientific variables · semantic mediation.

1 Introduction

The Ontology for Constructing Scientific Variables (OSV) is a mechanism for storing conceptual information necessary for identifying, disambiguating, and assembling scientific variables. OSV is a successor to the Geoscience Ontology (GSN)[4] and extends the principles introduced in the CF standard names [2] and the CSDMS standard names (CSN) [3]; whereas the aforementioned naming efforts relied on encoding scientific variables using controlled vocabularies and one-dimensional strings, the OSV is terminology-agnostic and encodes relational and contextual information via the Resource Description Format (RDF), resulting in a richer representation with more degrees of freedom. OSV is a critical tool for semantic mediation, providing the language to link unstructured information contained in large corpora to structured information captured in data sets and used by computational models. Along with other interpretative tools, OSV is designed to enable automated alignment and integration of distributed scientific information.

There are a wide range of scientific ontologies available, see e.g., [5,6,7]. However, although these ontologies are useful for specific applications, there is, to the authors' knowledge, no available ontology that (a) provides the desired specificity for distinguishing variables at a highly granular level within a domain, (b) comprises patterns that are readily extensible to other domains, and (c) defines mandatory components of a variable. The ontology we present in this work aims to decompose and modularize the construction of scientific variables, explicitly labeling *required* elements that *must* be provided in order to completely and unambiguously identify the concepts represented by a scientific variable—namely an object of observation, a corresponding property, and a quantity with units. We start by identifying the core ontology building blocks in Section 2 and then describing how the building blocks are combined to build complex concept representations.

2 Concept Class Definitions

2.1 Physical Concepts

A **Phenomenon** is a fact or situation that is observed or could be observed to *exist* or *happen* in the physical world. A phenomenon that is observed to *exist* is at equilibrium, whether dynamic, chemical or static, and one that is observed to *happen* is removed from equilibrium, experiencing a change of state as a result of certain processes. A phenomenon consists of the substance of which it is made (**Matter**), a **Form** that defines its occupation of space, and possibly, one or more **Processes**. Phenomena are defined recursively[1], where any given phenomenon can be decomposed into smaller phenomena and can be combined with other phenomena to build larger, more complex phenomena. A **Body** is a phenomenon at equilibrium that is identified by its Matter and Form. A **Process** is a set of actions that may occur in parallel or sequentially.

2.2 Abstract Concepts

A **System** is the abstracted, diagrammatic representation of a phenomenon, and includes any applied, human-contrived physical or mathematical abstractions or models. In OSV, a system that has a relatively unchanging state is *static*, while a changing system is *dynamic*. A static system may comprise multiple dynamic systems which together are at equilibrium. Like Phenomena, Systems are defined recursively.

A **Property** is a characteristic or feature of a system. A **Value** may be numerical or categorical and represents a system state, evaluated either objectively or subjectively; it is associated with a property. A **Quantity** is a numerical value with associated units. An **Attribute** is a property-value pair. It is important to note that some properties may be observable but may not be able to be measured directly and may be assessed through manipulation of other attributes; examples include severity and resilience.

A **Variable** is a phenomenon-property pair. It *must* comprise an object of measurement—one or more Phenomena—as well as a Property. As an example, ‘precipitation’ is not a complete variable, as it only identifies a process, and neither is ‘rainfall’, as it only identifies a phenomenon—the precipitation of water from clouds. In order to properly identify a variable, a property (such as ‘volume flux’ or ‘duration’ in the case of rainfall) must also be identified.

3 Building a Variable

The steps for identifying the components of a scientific variable are:

1. Select a phenomenon of interest for study–this is called the *object* of observation and will be the object of the variable.
2. Select one or more properties of that phenomenon to evaluate.
3. Diagram that phenomenon for the desired analysis, and if necessary, identify any applied abstractions, such as approximate mathematical or physical models (e.g., surface, ellipsoid, etc.).

A system is defined recursively in the ontology and comprises one or more participants, the role of each participant, and accompanying processes. Participants are recursively defined as distinct subsystems of the larger whole to provide the desired level of granularity. The granularity of any system may be further refined by identifying system attributes (system state) that are constant for the scope of measurement.

Figure 1 provides an overview of the different systems that can be modeled. Static systems involve processes that are at equilibrium while dynamic systems are removed from equilibrium. The single-body, static system is equivalent to the Body class. Matter is a type of multiple-body, static system. When enclosed with a boundary, a multiple-body, static system may be turned into a static, single-body system. When a Form is applied to Matter, a Body system results.

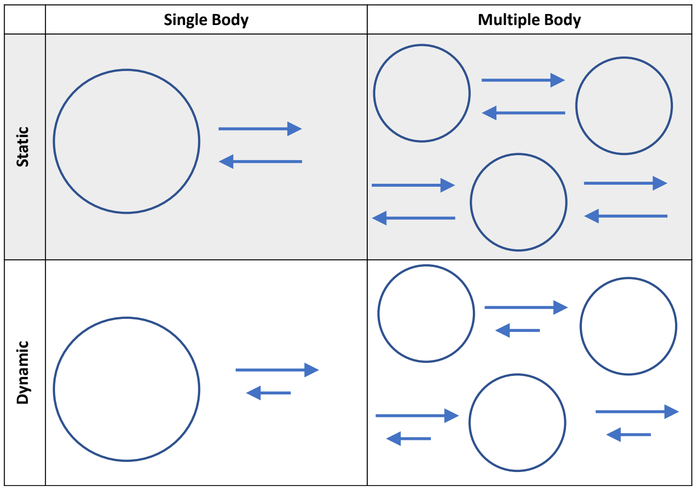

Fig. 1. The four types of systems. Circles represent Body systems and may be described by their Matter and Form, equal length arrow pairs represent processes at equilibrium, and unequal arrow pairs represent processes removed from equilibrium.

A variable is assembled by linking the system of interest to the desired property. If applicable, a variable may also include a reference frame for the evaluation of the property, as well as context phenomena. Figure 2 shows an example of how the building blocks are used to build a variable.

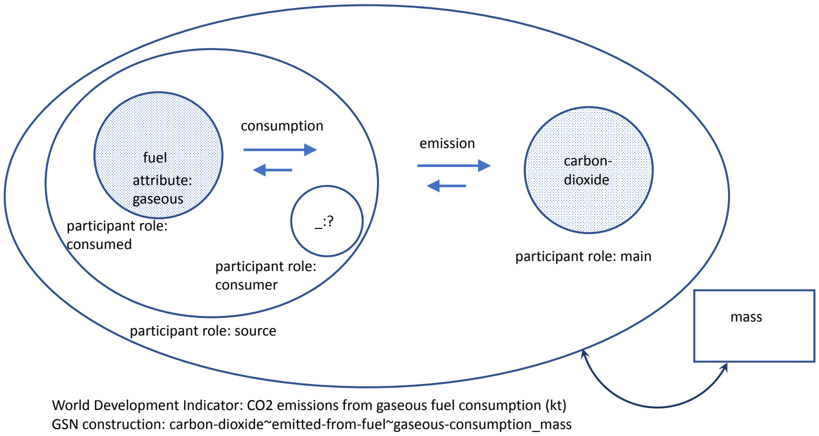

Fig. 2. Depiction of how a variable from the World Development Indicators list is represented as a dynamic, two-body system in OSV. The patterned circles indicate instances of Matter. A blank node is a stand-in for a participant that is not explicitly identified.

4 Implementation

The Geoscience Ontology[4] is an example of a domain-specific OSV application which expresses a wide range of scientific variables. The linked website provides a web interface to a SPARQL endpoint to query a beta version of the ontology.

References

1. Hooft, G.: In Search of the Ultimate Building Blocks. Cambridge University Press, Cambridge, UK (1997)
2. Guidelines for Construction of CF Standard Names, `http://cfconventions.org/Data/cf-standard-names/docs/guidelines.html`. Last accessed 1 June 2018
3. CSDMS Standard Names, `https://csdms.colorado.edu/wiki/CSDMS_Standard_Names`. Last accessed 1 June 2018
4. Geoscience Ontology, `http://www.geoscienceontology.org`. Last accessed 1 June 2018
5. SWEET Ontology, `https://sweet.jpl.nasa.gov/`. Last accessed 24 July 2018
6. QUDT Ontology, `http://www.qudt.org/`. Last accessed 24 July 2018
7. GCOS Ontology, `http://vocab-test.ceda.ac.uk/ontology/gcos/gcos-content/`. Last accessed 24 July 2018