In **PLoS Track at 15th Annual International Conference on Intelligent Systems for Molecular Biology and & 6th European Conference on Computational Biology**, July 2007, pp.4.

An ontology for the semantic modelling of natural systems.

Ferdinando Villa[1,2,3], Ioannis N. Athanasiadis[4], Gary W. Johnson[1,3]

[1]Ecoinformatics Collaboratory, Gund Institute for Ecological Economics; [2]Department of Plant Biology and [3]Department of Computer Science, University of Vermont, Burlington, VT 05401, USA; [4]Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland.

In recent years, activities such as the Semantic Web have spurred a renewed emphasis on the semantic description of datasets in natural sciences. Several initiatives are trying to improve our ability to integrate and coordinate information transfer by seeking ways to semantically describe datasets. A common endeavor for these initiatives is creating a core conceptualization (ontology) that can help formalizing scientific observations, such as measurements, and the ways the observed parts of a system relate to each other. Such ontologies allow basic integration of independently collected information and provide a branching point for further specification with domain-specific ontologies, enabling semantic validation of information collected, transferred and created along data paths.

In this contribution, we take a step towards a higher-level generalization, and present a new theoretical synthesis that stems from the realization that all system models, whether static (datasets) or dynamic, incarnate the result of an observation process that can be described logically through a single, appropriately expressive ontology. Such a conceptualization, which we have distilled into a publicly available OWL ontology and supported with open source software infrastructure, not only provides a more natural semantics for data annotation, but is also key to enabling a novel integration of data, models, and applications. Our statement of the nature of scientific observations and of the ways that observations can be linked to each other incorporates:

1. formal semantics for direct and indirect observations, including measurements, classifications and counts, and for the two fundamental relationship (contingency and dependency) that allow building fully general representations of observed systems representing composite datasets as well as process models and analytical pathways;
2. a notion of an **observation context**, encompassing both physical (space, time) and conceptual (hypothesis space, uncertainty) extents, which allows a pragmatic definition of **observational scale,** and enables automated validation and negotiation of representational differences.

This novel semantic unification has important practical applications, such as:

1. Assist the description of both static (data) and dynamic (model) natural systems knowledge with a common formal semantics, linked to domain-specific ontologies, enabling integration, consistency analysis, and facilitating the automated generation of simulation models based on the semantic description of dynamic systems.
2. Allow software to automatically induce semantic compatibility by transforming heterogeneous information along datapaths. Both context-insensitive (e.g. converting units of measurement) and context-sensitive (e.g. currency conversion in different temporal contexts) conversions are made possible, with full characterization and tracking of conversion-related uncertainties.
3. Enable **model-based query**, or the automated generation of query strategies that can discover realized instances of known conceptual models (e.g. food webs or species-area relationships) in semantically annotated, distributed databases.

This previously unpublished approach is being field-tested in two large-scale integrated projects. In the EU-funded SEAMLESS, our conceptualization allows mediation and integration of agricultural, physical, and econometric data and models to predict the impact of large-scale agricultural policy changes in the EU. In the NSF-funded ARIES, it allows automated discovery of scenarios of provision of valuable ecosystem services, and helps generate spatially-explicit maps of natural assets and their values that can be used in decision making. We discuss the theoretical foundations of the logical model and illustrate case studies of its application in each context.