# Position paper for W3C SW-LS Workshop

Hui Dong, Robert Morris
Department of Computer Science
UMASS Boston

## Position

Most life science ontology development has focused on molecular data, but many aspects of biology are described at higher levels of organization, including anatomical, physiological, and ecological levels. Furthermore, at all these levels substantial quantities of data are rapidly being digitized and applications against them are in need of semantically based discovery and integration with related data. For example, the Global Biodiversity Information Facility (GBIF)[1] now serves over 41 million digital museum specimen records from over 300 collections in 77 organizations. Available both as service oriented XML and with human-centric web interfaces, this data and related resources such as Integrated Taxonomic Information System (ITIS)[2] and the GBIF Electronic Catalog of Names[3] form important sources of data for applications about higher (and occasionally molecular) organization of life. Semantic approaches at this level are lacking, with the main contributions being in the SEEK[4] and SPIRE[5] projects. Although the work shares technical difficulties with molecular ontologies, it is too early to say that all solutions are independent of the organizational level of the biological models. However, it seems likely that the solutions to problems of organizational complexities at one level can inform ontology and system design at another.

## Perspective:

We have been collaborating with a botanist who is an expert in invasive species to build an ontology for applications in the identification and control of biological alien invasive species. The invasive species ontology was written in OWL and extends from SEEK ecological ontology[6]. It defines the concepts and relationships including taxon concept, habitat, growth form, life cycle, reproductive mode, and control methods for invasive species. Using the ontological based approach, it is efficient to discover services and datasets semantically compliant to user requirements, and facilitates data integration by global terminology and local assertions. However, as with many others, its problem for inference is that presently uses OWL Full. As a consequence, its semantics is beyond the standard description logic framework, and reasoning over assertions and expressions in OWL FULL happens to be intractable. To avoid this problem, algorithms for categorizing an instance OWL document to respective categories and converting an OWL Full ontology to an OWL DL ontology or an OWL Lite ontology without significantly losing the expressive power is needed. In addition, reasoning performance issues must also be considered to provide satisfactory utilities in biology knowledge systems.

## Reference

1) http://www.gbif.org

2) http://www.itis.usda.gov/
3) Electronic Catalogue of Names of Known Organisms
4) Science Environment for Ecological Knowledge(SEEK)
5) Semantic Prototype In Research Ecoinformatics
6) SEEK ecological concept ontology