Functional Trait Ontology
Workshop Report

Montpellier, France

November 5-7, 2009

Farshid Ahrestani
TraitNet project,
Columbia University
Department of Ecology, Evolution and Environmental Biology
New York, NY USA

Contents

1. Summary

Nineteen scientists met in Montpellier, France, November 5-7 to discuss, design and develop a 'Functional Trait Ontology' for ecologists. As the majority of ecologists present studied plants, it was decided that the ontology would initially focus on plant traits. The ontology would be developed by merging plant trait concepts listed in two categories, plant entities and plant characteristics. Plant entities are the physical parts of the plant, and plant characteristics are either direct, or derived, measurements made by ecologists/biologists of plant entities. A semantic web-based application tool developed by the group would be used to coordinate the clarification of the definitions, and define a hierarchal structure and relatedness between the plant trait concepts. It was decided to build the ontology using the OWL language rather than the OBO standard. Besides the ontology, it was decided that a semantic based registry of databases relevant to ecologists would be developed to showcase the usefulness of working with data in a semantic framework. The group has also decided to publish 2-3 papers related to the scientific output that emerges in 2010.

2. Background

Trait-based research spans a wide array of ecological and evolutionary disciplines. Answering relevant ecological questions using a trait-based approach requires the integration and synthesis of trait data across different disciplines. Several European and US initiatives are currently focusing on this issue. As one of these, TraitNet, a Research Coordination Network established in 2007, identified as one of its goals to develop cross-disciplinary computational tools for merging, disseminating, and sharing trait data. The traditional approach to achieve this goal would involve collection and assimilation of various data resources, a central relational database, and web-based user access interfaces. Recognizing the growing trend of the semantic approach to data management—supported by the recent establishment of standards for multiple Semantic Web technologies by the World Wide Web Consortium (W3C)—it was decided that TraitNet would approach the task of discovering and accessing, both distributed and centralized, trait data within a semantic framework. Ontologies—key components that facilitate discovering data in a semantic environment—define the relational and hierarchical nature of the concepts in a knowledge domain.

Organized by Eric Garnier, Isabelle Mougenot and Dan Bunker—and financially supported by the USNSF-funded TraitNet project, CNRS, and Université Montpellier 2, France —19 scientists from Europe and USA (Appendix 1) met for a workshop in Montpellier, France at the SupAgro-INRA Campus de La Gaillarde, from 5 to 7 November 2010, to discuss the development of a Functional Trait Ontology.

3. The Functional Trait Ontology meeting

3.1 Objectives and themes

The objectives of the workshop were:

- To develop a trait-based ontology, and a protocol for future additions to the ontology. A first version of the ontology will focus on plants.
    - Discuss the writing of a manuscript to describe the ontology.
- To develop a mechanism for continued development of the ontology
    - A communication mechanism
    - An oversight framework so that many can contribute and others can validate/approve contributions

Five major themes were to be discussed during the workshop:

1. What should be the scope of the ontology? *e.g.* Do we include community and/or ecosystem level?). Also, relatedness to other fields (taxonomy, vegetation relevés, soil and climate environments…)?
2. How does a trait-based ontology relate to existing ontologies in ecology?
3. Identify the relevant concepts in the domain knowledge of traits.
4. Discuss the best available ontological framework.
5. Technical choices to implement the ontology.

3.2 Talks/Presentations

November 5

The meeting began with plenary introductory talks by Eric Garnier (Senior Scientist, CNRS (CEFE), Montpellier, France) and Daniel Bunker (Assistant Professor, NJIT, Newark, USA) welcoming delegates to the workshop and outlining the goals and objectives of the workshop. The organizers stressed that while the workshop was to provide a platform for learning and discussing ontologies, the Semantic Web, ontological modeling approaches, and data resources available for integration with the ontology, one major goal of the workshop was to get participants to work on developing the structure of the ontology by defining the concepts within the functional trait knowledge domain.

- **What are ontologies? Trait ontologies in relation to other ontologies in ecology** : Isabelle Mougenot (Lecturer in Bioinformatics, Université Montpellier 2, Montpellier, France) defined in her talk what is an ontology; how ontologies are important for the distribution and reuse of data; the design and classification of ontologies; the software engineering and the use of Unified Modeling Language to develop ontologies; and efforts made by Marie-Angélique and herself in integrating the OBOE Ontology (developed by NCEAS) within the framework of Functional Trait Ontology.

- **The Crop Trait Ontology**: Elizabeth Arnaud (Principal Investigator, Crop Ontology project, CGIAR (Consultative Group on International Agricultural Research) Generation Challenge Programme) shared with delegates the experience and the challenges faced by her group in developing the Crop Trait Ontology. The talk not only exposed delegates to the nuances of developing an ontology, but also brought into focus into existing developed ontologies that the Functional Trait Ontology can build upon and link to.

- **SONet and SemTools**: **two ontology-based efforts supporting data integration in the earth sciences:** Mark Schildhauer (Director of Computing, National Center for Ecological Analysis and Synthesis, Santa Barbara, California, USA) highlighted two new efforts by his Center being developed to better semantic interoperability by providing cross-disciplinary data interoperability solutions, while also improving networking between semantic practitioners using recent software development tools like Protégé, Pellet, and open-source triple stores by the semantic community using W3C standards. He also highlighted how best to link the new Functional Trait Ontology to NCEAS's open observations data model OBOE semantic ontology.

- **Owl Ontologies and Protégé :** Farshid Ahrestani (Associate Director, TraitNet Project, Columbia University, New York, USA) introduced delegates to OWL (W3C's Web Ontology Language); the important building blocks (Classes, Properties, and Individuals) of an OWL ontology; the types, classifications, characteristics, nuances and restrictions of the different OWL ontology building blocks; and fundamental concepts and good development practices to use while building an OWL ontology.

November 6

- **Entity-quality formalism for computable phenotypes**: data integration across disparate fields within the Phenoscape project : Hilmar Lapp (Assistant Director of Informatics, NESCent, Washington DC, USA) shared with delegates NESCent's experience with the Phenoscape project. Phenoscape is completely semantic project that includes an ontology-driven triple store database; semantic data integration with other relevant databases; and a semantic-based data discovery user interface.

- **ThesauForm**: Marie-Angélique Laporte (PhD student, CEFE and LIRMM, Montpellier, France) presented the collaborative tool ThesauForm she developed to be used for information exchange on trait characteristics and entities. All participants gave feedback on the first version, which was put online before the workshop.

November 7: Database presentations

- **LEDA** : Michael Kleyer explained why, what and how he and his collaborators established the LEDA database, a database effort aimed at integrating plant trait data from several European based projects. Michael also explained what he considers are LEDA's biggest successes as well as shortcomings, and what lies ahead for its future.

- **TRY** : Jens Kattge (Coordinator, TRY project) explained the structure of the TRY database which is a collaborative effort of different plant ecologists to provide data for DGVM modelers and share amongst themselves. Jens explained the foundations for TRY's relational database model and discussed what lies ahead for the project in the future. A high number of scientists working with both TraitNet and TRY makes this database the primary choice as a database integrator for the Trait Ontology.

- **SALVIAS/BIEN** : Brian Enquist (Co-Leader, SALVIAS/BIEN Project) explained the vision, scope and projected outcome of the SALVIAS/BIEN project: a data integration, and discovery, project of plant ecology projects in the Americas across different spatial and temporal scales.

- **e-FLORA-SYS** : Bernard Amiaud talked about e-FLORA-SYS, a database of grassland projects in France. The database was established to serve as a repository of grassland projects in France, to facilitate sharing and increased collaboration between different French grassland scientists, and help predict changes to plant diversity in grasslands that are subjected to different impacts.

- **BASECO, GALIUM & ALTA** : Sophie Gachet shared her experiences in establishing three different plant trait databases: BASECO, a database of traits of Mediterranean flora; GALIUM a database of traits of Ile-de-France flora; and ALTA a database of traits of French alpine flora.

- **VISTA/PLANTRAITS** : Eric Garnier explained the rationale for the EU-funded VISTA (Vulnerability of Ecosystem Services to Land Use Change in Traditional Agricultural Landscapes) project, and its related database of plant traits, community structure and ecosystem properties.

3.3 Theme-based discussions

At the end of day 1 (Nov 5) 18 participating scientists expressed their opinions within groups of 4-5 on what they thought should be **the scope of the Trait Ontology** (*Theme 1*). The group concurred that to begin with "the Trait Ontology should be limited to plant functional traits, while being able to accommodate both expansion to a broader ecological knowledge domain and integration with other related ontologies". Also found to be important was the need to identify and formalize ontological interfaces with other relevant ecology disciplines like taxonomy, climate, soil etc.

It was decided that a priority first step in developing the ontology was to discover what semantic knowledge existed of the plant traits in existing ontologies—such as the Crop Ontology (CO), Plant Ontology (PATO) and Trait Ontology (TO)—so as to optimize the reuse of semantic knowledge and structure (*Theme 2)*.

At the beginning of Day 2, it was decided to use the OBOE (Madin et al. 2007, Ecological Informatics, 2, 279-296) as a general framework to develop the plant functional trait-based ontology. Participants then split up into two groups to accomplish the task of identifying relevant concepts related to plant traits (Theme 3). Group 1 (Farshid Ahrestani, Michael Kleyer, Brian Enquist, Marie-Angélique Laporte, Hilmar Lapp, Sandra Lavorel, Isabelle Mougenot, Shahid Naeem and Evan Weiher) focused on listing physical plant entities. The group (led by Brian and Evan) listed all physical plant entities that ecologists collect data on, arranged these entities in a hierarchal manner, and began the process of identifying the best definitions for these entities. Group 2 (Bernard Amiaud, Daniel Bunker, Sophie Gachet, Eric Garnier, Jens Kattge, Mark Schildhauer,) focused on listing plant functional traits, i.e. traits that can be derived from plant trait entities. The two groups worked on these tasks during Day 2 and till lunch on Day 3. It was also decided during the meeting that choosing the most appropriate ontological framework/model to define the plant trait knowledge domain would have to wait till after a few different inference models were tested in the following months (*Theme 4)*.

With respect to choosing the best technological solutions for this project (*Theme 5)*, it was decided that Isabelle and Marie-Angélique would lead this effort. The Thesauform being developed by Marie-Angélique would be used as the main tool to aggregate information regarding choosing relevant different trait entities and characteristics, and their respective definitions, and OWL was chosen over OBO as the language that the ontology would be built in.

3.4 Proposed post-meeting goals

The following list of goals was identified, with tentative deadlines and coordinators for each goal. Participants agreed that all participants at the Trait Ontology Workshop would have the opportunity to contribute to any and all project components and resulting publications.

1) Ontology Development
   a) Stabilize controlled vocabulary of characteristics
   b) Stabilize entities ontology
   c) Unite vocabulary and entities ontologies
   d) Recruit ecologists for further development of ontology
2) Write a paper describing the usefulness of data integration within a semantic framework of ecological concepts

3) Develop a Registry of data resources
4) Write a paper describing the plant trait ontology
5) Develop an example that illustrates the usefulness of taking a semantic approach to data integration
    a) Identify an ecological question
    b) What informatics issues must this example address?
    c) Identify datasets that can be used to work out this example.
    d) Model the worked example

Appendix 1

Table 1. List of Participants

Last name	First name	Country	Domain	e-mail address
Ahrestani	Farshid	USA	IT/Ecology	fa2260@columbia.edu
Amiaud	Bernard	France	Ecology/agronomy	Bernard.Amiaud@ensaia.inpl-nancy.fr
Arnaud	Elizabeth	France	IT	e.arnaud@cgiar.org
Bunker	Dan	USA	IT/Ecology	dbunker@njit.edu
Cellinese	Nico	USA	IT	ncellinese@flmnh.ufl.edu
Enquist	Brian	USA/France	Ecology	benquist@u.arizona.edu
Gachet	Sophie	France	Ecology	sophie.gachet@univ-cezanne.fr
Garnier	Eric	France	Ecology	eric.garnier@cefe.cnrs.fr
Kattge	Jens	Germany	Ecology/IT	jkattge@bgc-jena.mpg.de
Kleyer	Michael	Germany	Ecology/IT	michael.kleyer@uni-oldenburg.de
Grigulis	Karl	France	Ecology	karl.grigulis@ujf-grenoble.fr
Laporte	Marie-Angélique	France	IT	marie-angelique.laporte@cefe.cnrs.fr
Lapp	Hilmar	USA	IT	hlapp@nescent.org
Lavorel	Sandra	France	Ecology	Sandra.Lavorel@ujf-grenoble.fr
Mougenot	Isabelle	France	IT	isabelle.mougenot@lirmm.fr
Naeem	Shahid	USA	Ecology	sn2121@columbia.edu
Navas	Marie-Laure	France	Ecology/agronomy	navas@supagro.inra.fr
Schildhauer	Mark	USA	IT	schild@nceas.ucsb.edu
Weiher	Evan	USA	Ecology	weiher@uwec.edu