Modeling a microbial community and biodiversity assay with OBO Foundry ontologies: the interoperability gains of a modular approach.

Philippe Rocca-Serra[1], Ramona Walls[2], Jacob Parnell[3], Rachel Gallery[4], Jie Zheng[5], Susanna Assunta Sansone[1] and Alejandra Gonzalez-Beltran[1]

1. Oxford e-Research Centre, University of Oxford, OX1 3QG, Oxford, United Kingdom
2. The iPlant Collaborative, University of Arizona, Thomas J. Keating Bioresearch Building, 1657 East Helen St., Tucson, AZ 85721, USA.
3. NEON, Inc., 1685 38th St., Ste. 100, Boulder, CO 80301, USA.
4. University of Arizona, 1311 E. 4th St. Tucson, AZ 85721, USA.
5. Department of Genetics and Institute of Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract— The advent of affordable sequencing technology provides for a new generation of explorers who probe the world's microbial diversity. Projects such as Tara Oceans, Moorea Biocode Project and Gut Microbiome rely on sequencing technologies to probe community diversity. Either targeted gene surveys (also known as community surveys) or complete metagenomes are evaluated. The former, being the less costly of the two methods, relies on the identification of specific genomic regions, which can be used as a proxy to estimate genetic distance between related species in a Phylum. For instance, 16S ribosomal RNA gene surveys are used to probe bacterial communities while Internal Transcribed Spacer (ITS) surveys, for example, can be used for probing fungal communities. With the explosion of projects and frenzy to explore new domains of life, scientists in the field have issued guidelines to report minimal information (following a checklist), ensuring that information is contextualized in a meaningful way. Yet the semantic of a checklist is not explicit. We demonstrate here how a tabular template can be used to collect information on microbial diversity using an explicit representation in the Resource Description Framework (RDF) that is consistent with community agreed-upon knowledge representation patterns found in the Ontology for Biomedical Investigations (OBI).

Keywords— microbial diversity, targeted gene survey, ontology, assay, OBO Foundry, ISA framework