Alignment of Cultured Cell Modeling Across OBO Foundry Ontologies: Key Outcomes and Insights

Matthew H. Brush[1*], Jie Zheng[2], Yongqun He[3], Sirarat Sarntivijai[4], Bjoern Peters[5], Alexander D. Diehl[6], Christian J. Stoeckert[2], Melissa A. Haendel[1]

[1]Oregon Health & Science University, Portland, OR, USA
[2] University of Pennsylvania, Philadelphia, PA, USA
[3] University of Michigan, Ann Arbor, MI, USA
[4]US Food and Drug Administration, MD, USA
[5]La Jolla Institute for Allergy & Immunology, San Diego, CA, USA
[6]University at Buffalo, Buffalo, NY, USA

ABSTRACT

The *in vitro* culture of animal cells has long served as an experimentally amenable model for investigating cellular phenomena that occur in metazoan organisms. At present, several ontologies within the OBO Foundry library contain modeling that relates to the types, production, and biological attributes of cell lines and cultures. These include the Cell Line Ontology (CLO), Ontology of Biomedical Investigations (OBI), the Cell Ontology (CL), the eagle-i Resource Ontology (ERO) and the Reagent Ontology (ReO). Prompted by new informatics initiatives from the Beta Cell Biology Consortium[1], the Monarch Initiative[2], and the eagle-i network[3], each of which aim to leverage a shared model of cultured cells and lines, several critical problems were identified pertaining to interoperability of existing representations across these ontologies. To address these issues, a working group comprised of representatives from each of these efforts was formed to harmonize modeling of experimentally cultured cells in accordance with OBO Foundry principles of orthogonality and re-use. We wish to report to the ICBO community the outcomes of this work to inform users of important updates, and to share insights and best practices that emerged from our exercise in high-level ontology alignment.

As a pre-requisite for implementing an aligned model across these ontologies, it was necessary to establish consensus definitions and labels for key high-level concepts in the domain. This included careful characterization of entities such as primary cell cultures, cell lines, and immortal cell lines, as well as defining different 'scales' at which to represent cultured cell populations - including single cells, experimental 'samples' of cells, and entire cultures or lines. This work proved a surprisingly difficult but productive endeavor, and highlighted many important complexities and nuances that were accounted for in our final model. Key outcomes of the work that followed included the accommodation of the CLO move to represent specific types of lines at the level of single cells (e.g. the CLO 'HeLa' class represents a single HeLa cell rather than a population or entire line of cells). In addition, duplicate representations of 'cell line cell' classes between CL, OBI, and CLO were merged to live as a single class in CLO, which as then imported into CL and OBI. Necessary updates were made to other ontologies and data sets that use CL and CLO classes, so as to be consistent with these changes.

Another key outcome was the representation of cultured cells collections in OBI at the level of a 'cell line sample'. This new concept is defined as a population of cultured cells comprised of a defined portion of a cell line that has been successively passaged together. Here we make a distinction between a sample and the entire line of which it is a part, which includes the broader collection of all generations of cells following from the point the line was established. The definitional criteria for a 'cell line sample' are meant to clearly demarcate a collection representing what researchers actually culture, experiment on, and share in the day-to-day practice of science. This concept is important because it is at this scale where attributes such as passage number apply, and at which there is a common need to annotate biomedical data. Introduction of this newly clarified scale of representation of cultured cell collections required extensive updates to legacy modeling in OBI and ERO, to clarify the precise intention of existing representations.

In the end, alignment outcomes were implemented such that each concept was represented as a single class that resided in the appropriate ontology, and was re-used by others in a consistent manner. As a result, several classes among participating ontologies needed to be merged and obsoleted, and appropriate documentation was made to provide users with an audit trail. The importance of documentation was also one of the many instructive lessons earned in the process of our alignment work, in addition to insights regarding venues and formats for discussion, and resolution of conflicting perspectives about core concepts in the domain.

* To whom correspondence should be addressed: brushm@ohsu.edu

1 http://genomics.betacell.org/gbco/

2 http://monarchinitiative.org/

3 https://www.eagle-i.net/