OBCS: The Ontology of Biological and Clinical Statistics

Jie Zheng[1]*, Marcelline R. Harris[2], Anna Maria Masci[3], Yu Lin[4], Alfred Hero[5,6], Barry Smith[7], Yongqun He[4]*

[1] Department of Genetics, University of Pennsylvania Perelman School of Medicine, PA 19104, USA
[2] Division of Systems Leadership and Effectiveness Science, University of Michigan School of Nursing, Ann Arbor, MI 48109, USA
[3] Department of Immunology, Duke University Medical Center, Durham, NC 27710, USA
[4] Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, Center for Computational Medicine and Bioinformatics, Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109, USA
[5] Department of Electrical Engineering and Computer Science and Department of Biomedical Engineering, University of Michigan College of Engineering, Ann Arbor, MI 48109, USA
[6] Department of Statistics, University of Michigan College of Literature, Sciences and the Arts, Ann Arbor, MI 48109, USA
[7] Department of Philosophy, University at Buffalo, Buffalo, NY 14203, USA

* Co-corresponding authors

Abstract— Statistics play a critical role in biological and clinical research. In the era of Big (clinical) Data, this role becomes even more prominent, since statistics will serve as the central tool in the virtual clinical trials and meta-trials of the future. To promote logically consistent representation and classification of statistical entities, we have developed the Ontology of Biological and Clinical Statistics (OBCS). OBCS extends the Ontology of Biomedical Investigations (OBI) that is an ontology of the Open Biological/Biomedical Ontologies (OBO) Foundry and is supported by some 20 communities. OBCS imports all statistics-related terms from OBI. In addition, many other statistics related terms were added to OBCS to fill up the gaps in statistics representation. A combination of top-down and bottom-up methods is used in the OBCS development. The top-down approach works by surveying statistics workflows from the perspective of high-level structuring and generating new terms in OBCS when they are missing. The bottom-up method is applied by studying specific biological and clinical statistical analysis use cases, creating corresponding terms under existing high-level ontology classes. Currently, OBCS contains 686 terms, including 381 classes imported from OBI and 147 classes specific to OBCS. In this paper, we will introduce the rationale, history, and current status of the OBCS development. Furthermore, one biological and one clinical use cases are provided to illustrate potential applications of OBCS. The biological use case involves an OBCS representation of a statistical data analysis of a microarray experiment conducted using blood samples from human subjects vaccinated with a trivalent inactivated influenza vaccine. The clinical use case analyzes clinical outcomes of nursing services using data obtained from electronic hospital discharge abstracts. The OBCS will be further developed. More statistics terms will be included based on community needs and biological/clinical use cases. The OBCS project and source code are available at http://obcs.googlecode.com.

Keywords— ontology, statistics, data analysis, biological and clinical research