Ontology of Experiments for Bioinformatics

Larissa N. Soldatova[1] **Ross D. King**[1]
lss@aber.ac.uk rdk@aber.ac.uk

[1] The University of Wales, Aberystwyth, Penglais, Ceredigion, SY23 3DB, Wales, UK

Keywords: experiment design, annotation, ontology, bioinformatics

1 Introduction

The efficient annotation, retrieval, and sharing of experimental results is a key problem in modern biology. As more and more biological data is stored on computer the problem of efficient retrieval and analysis of this data becomes the most important scientific bottleneck, and the problem is particularly acute in biology because biological data is notorious for its complex form and semantics. Ontologies can help in these tasks because they embody the abstract knowledge required for data integration and analysis. The utility of ontologies has been clearly demonstrated in several biological domains (e.g. Gene Ontology [4]). A formalized description of experiments is extremely important for organizing and executing experiments in biology.

2 An Ontology of Experiments

Our Ontology of EXPeriment (EXPO) aims to formalize nonspecific knowledge about experiment design, methodology, and representation of results (see Fig. 1). Many particular domains including biology, chemistry, medicine, etc. follow the same principles of organizing, caring out, and analyzing of experiments; use similar instruments, materials; and describe experiment results in the same formats and dimensional units. EXPO proposes a standardized vocabulary for the annotation of scientific experiments, allows sharing and reusing of common knowledge, and provides full interoperability with the upper ontology SUMO of the ontology society and compliant domain ontologies [5]. EXPO is based on a classic theory of experiment design [3], statistics, a theory of errors [1], philosophy of science (epistemology, logic, and methodology) [2], analysis of existing upper and domain ontologies, meta-data of experiments, and analysis of real experiments.

3 Discussions

Experimental goals, methods, requirements, experimental restrictions, rules for experiment design, etc should be kept in an ontology of experiments. The advantage of this abstraction is that generic knowledge is held in only one place - ensuring consistency, clean updating, etc. Such knowledge, as well as information on dimension units, data types, and types of bibliographic references, etc. should be represented at the level of upper ontology. Likewise, descriptions of experiment objects and subjects as a human, an animal, a plant, a robot, belong to scientific domain ontologies. The division of knowledge into corresponding levels, with the integration of the ontological knowledge for retrieval and inference will enhance the comprehensiveness, functionality of bioinformatics systems, as well as optimizing their design, and helping to avoid many mistakes made in their construction.

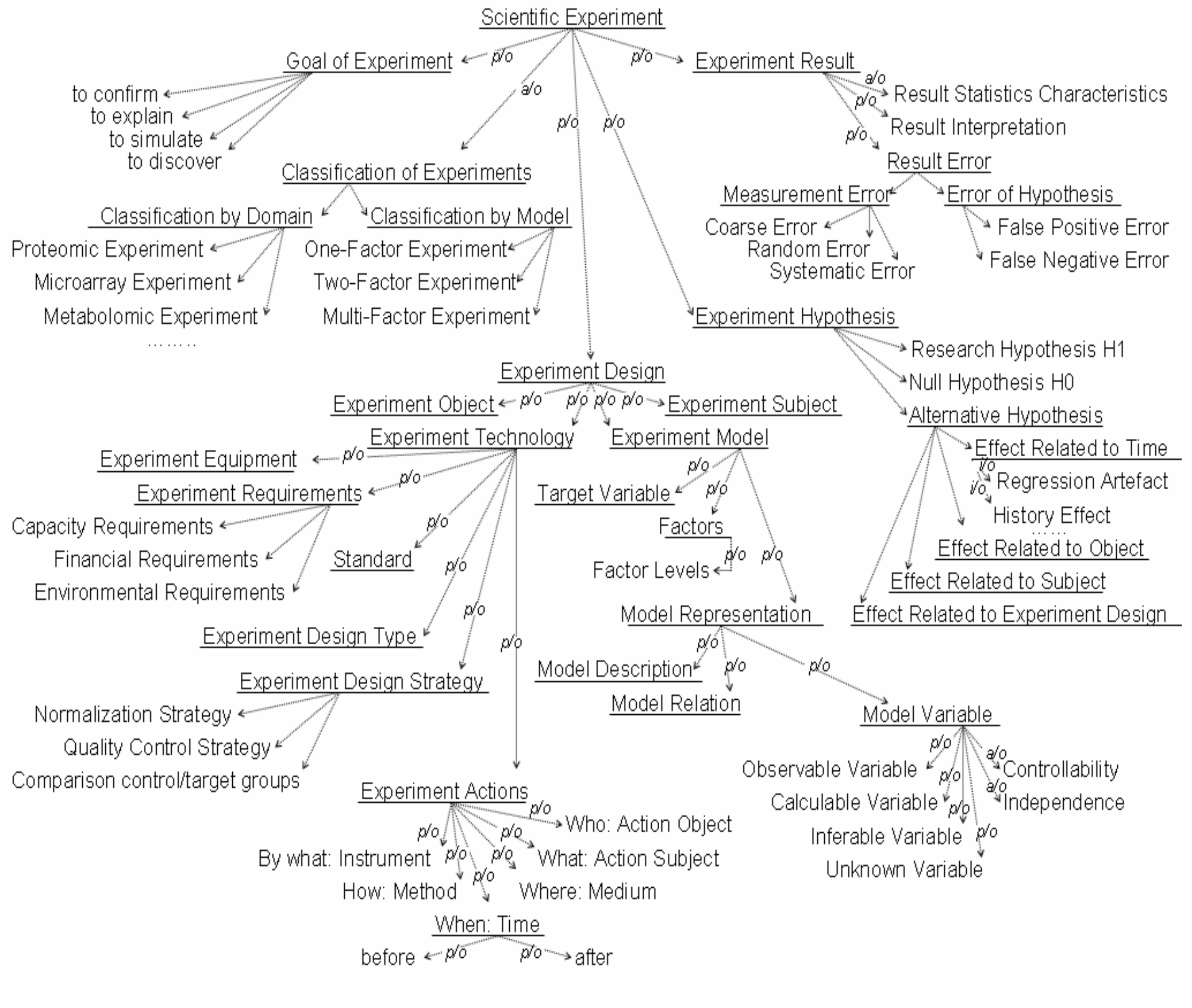

Figure 1: The Ontology of Experiment (a fragment), where p/o is a part-of relation, a/o is an attribute-of relation, i/o is an instance-of relation, and an arrow with an empty label corresponds to a class-subclass (is-a) relation.

References

[1] Arley, N. & Buch , K.R., *Introduction to the theory of probability and statistics*, New York : Wiley London. Chapman & Hall, 1950.

[2] Caws, P., *The philosophy of science: a systematic account*, Princeton, N.J. ; London : Van Nostrand, 1965.

[3] Fisher, R. A., *The design of experiments*, Oliver & Boyd, Edinburgh, 1956.

[4] http://www.geneontology.org

[5] http://suo.ieee.org/SUO/SUMO/index.html