Modeling biomedical experimental processes with OBI

The OBI consortium (in alphabetical order): Ryan Brinkman[1], Bill Bug[2], Helen Causton[3], Kevin Clancy, Christian Cocos, Mélanie Courtot[1], Dirk Derom[4], Eric Deutsch[5], Liju Fan[6], Dawn Field[7], Jennifer Fostel[8], Gilberto Fragoso[3], Frank Gibson[9], Tanya Gray, Jason Greenbaum[10], Pierre Grenon, Jeff Grethe[7], Yongqun He[11], Mervi Heiskanen, Tina Hernandez-Boussard[7], Philip Lord, Allyson Lister[12], James Malone, Elisabetta Manduchi, Luisa Montecchi, Norman Morrison[3], Chris Mungall[13], Helen Parkinson[3], Bjoern Peters[10], Matthew Pocock, Philippe Rocca-Serra[3], Daniel Rubin, Alan Ruttenberg[14], Susanna-Assunta Sansone[3], Richard Scheuermann[15], Daniel Schober[16], Barry Smith[17], Larisa N. Soldatova[18], Holger Stenzhorn, Chris Stoeckert[19], Chris Taylor[3], Randi Vita[10], John Westbrook, Joe White, Patricia L. Whetzel[7], Stefan Wiemann, Jie Zheng[19]*

[1]BC Cancer Agency, Vancouver, Canada; [2]National Center for Microscopy Imaging Research, UCSD, CA, USA; [3]The European Bioinformatics Institute, Cambridge, UK; [4]Victoria University of Wellington, New Zealand; [5]ISB, Seattle, WA, USA; [6]Ontology Workshop, LLC, Columbia, MD, USA; [7]Stanford University Medical School, Stanford, CA, USA; [8]SRA Global Health Sector, Durham, NC, USA; [9]Abcam plc, Cambridge, UK; [10]La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA; [11]University of Michigan, Ann Arbor, USA; [12]Newcastle University, Newcastle upon Tyne, UK; [13]Lawrence Berkeley National Lab, Berkeley, CA, USA;[14]Science Commons, Cambridge, MA, USA; [15]Southwestern Medical Center, USA; [16]IMBI, University Medical Center, Freiburg, Germany; [17]University at Buffalo, Buffalo, NY, USA; [18]Aberystwyth University, Wales, UK; [19]University of Pennsylvania School of Medicine, Philadelphia, PA, USA

ABSTRACT

Motivation: Experimental metadata are stored in many different formats and styles, creating challenges in comparison, reproduction and analysis. These difficulties impose severe limitations on the usability of such metadata in a wider context. The Ontology for Biomedical Investigations (OBI), developed as part of a global, cross-community effort, provides an approach to represent biological and clinical investigations in an explicit and integrative framework which facilitates computational processing and semantic web compatibility. Here we detail two real-world applications of OBI and show how OBI satisfies such use cases.

Availability: OBI is available at http://purl.obofoundry.org/obo/obi.owl

1 INTRODUCTION

Biomedical investigations use empirical approaches to investigate causal relationships among a large range of variables. The breadth of investigations presents a number of challenges when attempting to describe experimental processes; there are varying levels of complexity, quantities and types of data, material and equipment. Furthermore, the use of terminology across the numerous communities can in itself present issues of ambiguity. This presents a further challenge when attempting to consistently represent biomedical investigations, regardless of the specific field of study.

An ontological approach for the representation of a consensus-based controlled vocabulary of terms and relations has previously shown to be successful in achieving some of these goals [the GO consortium, 2000; Matos *et.al*., 2006]. The requirement for an efficient representation of investigations is recognized as a pressing problem, and there are multiple efforts developing ontologies to increase semantic content and standardization (GO, MGED, LABORS, MSI ontology) [Whetzel *et al*., 2006; Brazma *et al*., 2001; King *et al*., 2009; Sansone *et al*., 2007].

These developments by the transcriptomics, proteomics and metabolomics communities were originally in parallel, producing similar ontologies with overlapping scope. Though each focuses on particular types of experimental processes, many generic terms such as 'investigation', 'assay' or 'hypothesis' will be common to them all. Merging common aspects of these formalisms is useful as it provides a common mechanism by which a class can be defined and understood by all, removing ambiguity and potential conflict. This practice is endorsed by policy providers such as the OBO Foundry which requires all member ontologies to define a single class only once (orthogonality) [Smith *et al*., 2007]. OBO Foundry members are required to use the same defined set of relations and upper-level ontology in order to facilitate integration and to promote automated reasoning. The Relation Ontology (RO) [Smith *et al*., 2005] contains the relations that may be used within an OBO Foundry-compliant ontology, while the

Basic Formal Ontology (BFO) [Grenon *et al*, 2004] contains all of the top-level classes under which member ontologies should build.

The Ontology for Biomedical Investigations (OBI) is being developed to address the need for a common, integrated ontology for the description of biological and clinical investigations [Whetzel *et al.*, 2006]. OBI is developed through collaborations among 19 biomedical communities and is part of the OBO Foundry. The representatives from the ontology projects mentioned above (MGED, LABORS, MSI ontology) are now actively involved into the OBI project. OBI includes a set of 'universal' semantic identifiers applicable across various biomedical and technological domains, and domain-specific terms relevant only to a given domain. OBI aims at representing various experimental processes e.g., investigation, study, assay, and the entities involved into those processes e.g., the study design, the protocols and instrumentation used, the material used, the data generated and the type of analysis performed on the data. OBI intends to support the logically consistent annotation of biomedical experimental processes regardless of the particular field of study. In this paper, OBI will be applied to two use cases exemplifying such annotation: 1) a blood glucose assay, and 2) a vaccine protection study.

2 USE CASES

Biomedical experimental processes can involve numerous sub-processes, where each step can involve various material entities e.g., whole organisms, organ sections, cell culture, cell pellets, devices. OBI defines the class 'investigation' as a process with the objective to generate an information entity by planning an overall study design, executing it, and documenting the results. An investigation can include a sub process of interpreting the data to draw conclusions. There can be multiple study designs in one investigation. There can be investigations that are part of another investigation. The 'material entity' class is defined in OBI as the subclass of the class 'independent continuant' that is spatially extended whose identity is independent of that of other entities and can be maintained through time. Note: material entity subsumes object, fiat object part, and object aggregate, because the three level theory of granularity is inadequate for biology. Material entities realize distinct roles given the context of the process they are used in e.g. study subject role, host role, specimen role, patient role; and distinct functions e.g. measure, separation, environment control (see the definitions of other entities at: http://purl.obofoundry.org/obo/obi.owl).

The following two use cases demonstrate how to model the entities and relations between those entities involved in the experimental processes using OBI. The first use case is applicable to several biological domains and shows how OBI can be used to model biological measurements, for example in the context of an assay. The second use case focuses on a specific implementation from one of the OBI communities, and details how OBI is used to model Vaccine studies.

2.1 Use case 1: Blood glucose assay

The blood glucose analyte assay measures the glucose concentration from blood extracted from an organism such as mouse or human. This experimental process contains two sub-processes (or steps):

(1) 'taking sample from organism' is done to obtain a blood specimen from the mouse. In this process, a device syringe is used to draw the blood from the mouse, contain it and transfer it to the glucometer. At the end of the process, a test tube contains the blood specimen ready for assaying via some device.

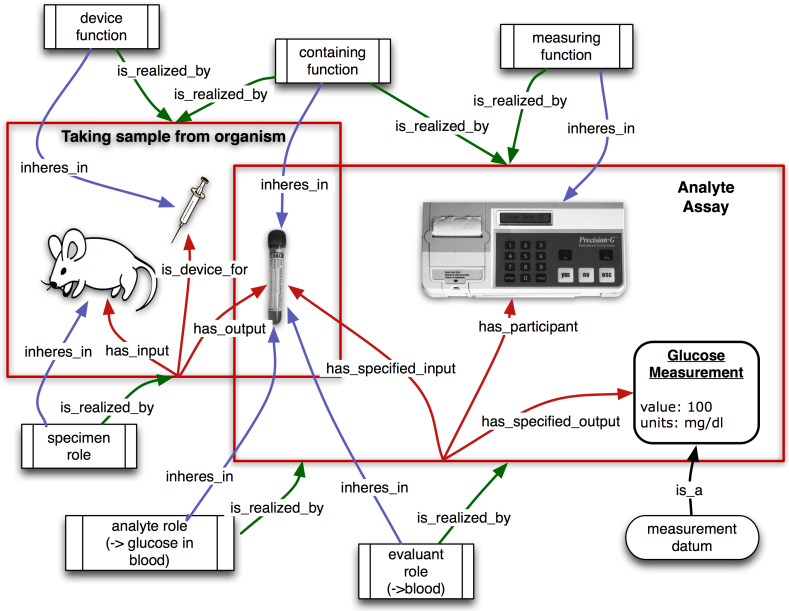

Fig.1. OBI modeling of a blood glucose assay. Analyte role inheres in scattered aggregate of glucose molecules in blood. In general, the thing that bears analyte role is a part of the thing that bears evaluant role.

(2) 'analyte assay' is defined in OBI as an assay with the objective to determine the presence, concentration, or amount of one substance (bearer of analyte role) that is present in another (bearer of the evaluant role). The presence, concentration or amount is a quality of the analyte towards the evaluant.

In this use case, blood will be used as the evaluant in the analyte assay to determine the concentration of glucose in the blood (see Fig.1). The device glucometer is used for this assay. The blood is specified as input to the glucose measuring process. The blood bears the role evaluant in the analyte assay, the material entity in which the assay measures the concentration of the analyte molecule. The analyte role inheres in the glucose molecules scattered throughout the blood specimen. The objective of this planned process is to analyze the analyte (glucose) concentration. The glucometer device is used to measure the concentration of glucose in the sample transferred to the dipstick, and to calculate the concentration of glucose in the evaluant. The specified output is the glucose concentration in the evaluant at some given measurement datum. The glucose concentration in the blood specimen will then be used to describe the mouse from which the blood was taken.

2.2 Use case 2: Vaccine protection study

The vaccine protection study (or vaccine challenge experiment) measures how efficiently a vaccine or vaccine candidate induces protection against virulent pathogen infection *in vivo*. Fig. 2 demonstrates how to use OBI to represent a typical vaccine protection experiment via the following three sub-processes:

(1) 'vaccination' is modeled as an 'administering substance in vivo' that realizes some 'material to be added role' borne by some 'vaccine' (e.g., VacX) and bears some 'target of material role' borne by some 'organism' that bears some 'host role' (e.g., mouse).

(2) 'pathogen challenge' is another 'administering substance in vivo' that realizes some 'material to be added role' borne by some 'organism' and bears some 'pathogen role' (e.g., Influenza Virus) and realizes some 'target of material role' borne by some 'organism' that bears some 'host' (e.g., mouse).

(3) 'survival assessment' is a common protection efficiency assay that has specified input some organism (e.g., mouse) and has specified output some survival measurement, in this case, 75% of mice survived from pathogen challenge.

It is noted that survival assay is not the only method to assess the efficacy of a specific vaccine to protect against virulent infections *in vivo*. It is mainly due to the fact that many pathogens are not able to kill its host (e.g., mouse).

DISCUSSION

OBI aims for a robust and versatile representation of biomedical investigations. Two distinct biological use cases are modeled by OBI. Individual experimental steps are represented by the process and its subclasses in OBI. The two processes in the blood glucose assay and the three processes in the vaccine protection study all fall under 'planned process' in OBI.

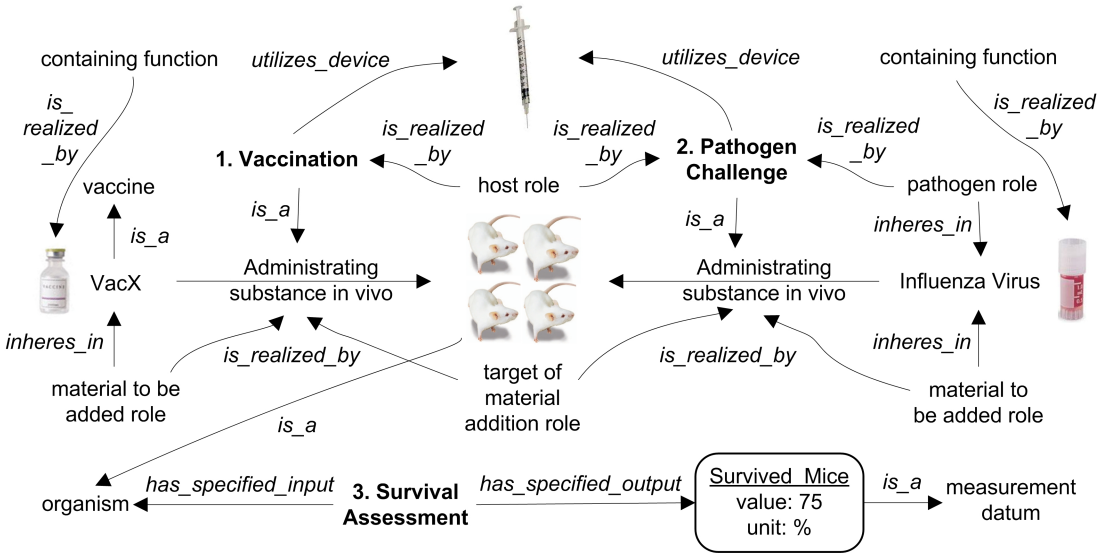

Fig. 2. OBI modeling of vaccine protection study.

The processes 'vaccination' and 'pathogen challenge' are disjoint subclasses of 'administering substance in vivo', which is a subclass of 'material combination'. It is opposite to the 'material separation' used in the glucose assay. Both 'analyte assay' and 'survival assessment' are subclasses of the class 'assay'. Syringe is a 'device' that participates in different processes. Such entities as glucose and vaccine are subclasses of the class 'material entity'; analyte, evaluant, subject are subclasses of the class 'role'. All these entities are used to represent disparate experimental processes in a way suitable for computer-assisted reasoning.

We demonstrated how OBI can be applied to two different biomedical investigations and increase the computer accessible semantics of the information. The logical definitions of the entities involved allow computers to unambiguously understand and integrate different biological experimental processes and their relevant components with the help of an OWL reasoner.

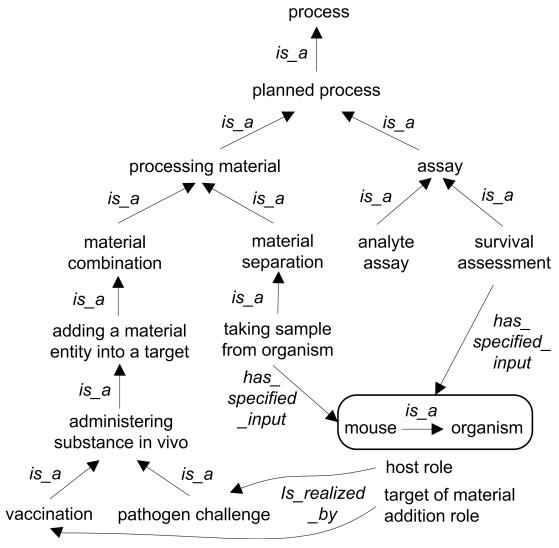

Fig. 3. Hierarchical structure of processes used in the two use cases.

OBI descriptions of experimental processes can be used for different applications. For example, the OBI modeling of a vaccine protection study provides an advanced approach to represent and mine vaccine-induced protection experimental processes. Approximately 400 vaccines have been manually curated and stored in the VIOLIN vaccine database system [Xiang *et.al.,* 2008]. The vaccine protection experimental data in VIOLIN is currently stored in plain text without ontological definitions nor term hierarchy. The lack of a common ontology support becomes an obstacle of the full use of the VIOLIN vaccine data for advanced querying and for the task of integration with data from other data sources. Once the vaccine protection data and other biomedical data are represented using OBI [VO], the information will be easily compared and analyzed among different vaccines and other biomedical domains.

We have considered only two use cases as examples of the modeling of experimental processes with OBI. The representatives of the communities, members of the OBI consortium, ensure that OBI has sufficient coverage and is suitable for the description of experimental processes from a wide range of biomedical applications. OBI will be further developed to expand the coverage and depth of biomedical investigations.

ACKNOWLEDGEMENTS

This work is partially supported by grant funding from the National Institute of Biomedical Imaging and Bioengineering, National Human Genome Institute, NIH (R01EB005034, NIH P41 HG003619), EC EMERALD project (LSHG-CT-2006-037686), EC FELICS, MUGEN, BBSRC (BB/E025080/1, BB/C008200/1), RC UK, NERC-NEBC, EU IP CarcinoGenomics (PL 037712), EU NoE NuGO (NoE 503630), CARMEN project EPSRC (EP/E002331/1), the Michael Smith Foundation for Health Research.

REFERENCES

Brazma, A., et al. (2001) Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat Genet*, **4**: 365-371.

Grenon, P., Smith, B., and Goldberg, L. (2004) Byodynamic Ontology: Applying BFO in Biomedical Domain, *Ontologies in Medicine*, IOS Press: 20-32.

King, R.D., et al. (2009) The Automation of Science. *Science*, **324/ 5923**: 85-89.

Matos, P., Ennis, M., Darsow, M., Guedj, M., Degtyarenko K. and Apweiler R. (2006) ChEBI - Chemical Entities of Biological Interest. *Nucleic Acids Research*, Database Summary paper 646.

Sansone, S., et al. (2007) Metabolomics Standards Initiative - Ontology Working Group. Work in Progress. *Metabolomics*, **3(3)**: 249-256.

Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, Mungall CJ, Neuhaus F, Rector A, and Rosse C (2005) Relations in Biomedical Ontologies. *Genome Biology*, **6**: R46

Smith, B., et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology,* **25**: 1251 - 1255.

The Gene Ontology Consortium (2000) Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, **25**: 25-29.

Whetzel, P. L., et al. (2006) The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics*, **7**: 866-873.

Xiang Z, et al. (2008) VIOLIN: vaccine investigation and online information network. *Nucleic Acids Res*. **36** (Database issue): D923-8.

The OBI Consortium. http://purl.obofoundry.org/obo/obi

VO. http://www.violinet.org/vaccineontology