**Modeling biomedical experimental processes with OBI**

The OBI consortium (in alphabetical order): Ryan Brinkman1, Bill Bug2, Helen Causton3, Kevin Clancy, Christian Cocos, Mélanie Courtot1, Dirk Derom4, Eric Deutsch5, Liju Fan6, Dawn Field7, Jennifer Fostel8, Gilberto Fragoso, Frank Gibson9, Tanya Gray, Jason Greenbaum10, Pierre Grenon, Jeff Grethe7, Yongqun He11, Mervi Heiskanen, Tina Hernandez-Boussard7, Philip Lord, Allyson Lister12, James Malone, Elisabetta Manduchi, Luisa Montecchi, Norman Morrison, Chris Mungall, Helen Parkinson3, Bjoern Peters10, Matthew Pocock, Philippe Rocca-Serra3, Daniel Rubin, Alan Ruttenberg13, Susanna-Assunta Sansone3, Richard Scheuermann14, Daniel Schober3, 18, Barry Smith15, Larisa N. Soldatova\*16, Holger Stenzhorn, Chris Stoeckert17, Chris Taylor, Randi Vita10, John Westbrook, Joe White, Trish Whetzel7, Stefan Wiemann, Jie Zheng17.

1British Columbia Cancer Research Center, Vancouver, Canada; 2National Center for Microscopy Imaging Research, UCSD, CA, USA; 3The European Bioinformatics Institute, Cambridge, UK; 4Victoria University of Wellington, New Zealand; 5ISB, Seatle, WA, USA; 6Ontology Workshop, LLC, Columbia, MD, USA; 7Stanford University Medical School, Stanford, Ca, USA; 8SRA Global Health Sector, Durham, NC, USA; 9Abcam plc, Cambridge, UK; 10La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA; 11University of Michigan, Ann Arbor, USA; 12Newcastle University, Newcastle upon Tyne, UK; 13Science Commons, Cambridge, MA, USA; UT, 14Southwestern Medical Center, USA; 15University at Buffalo, Buffalo, NY, USA; 16Aberystwyth University, Wales, UK; 17University of Pennsylvania School of Medicine, Philadelphia, PA, USA, 18Institute of Medical Biometry and Medical Informatics (IMBI), University Medical Center, 70104 Freiburg, Germany.

§Corresponding author

Email addresses:

CAD: [charles@darwin.co.uk](mailto:charles@darwin.co.uk)

JED: [jane@darwin.co.uk](mailto:jane@darwin.co.uk)

JRSS: [johnsmith@darwin.co.uk](mailto:johnsmith@darwin.co.uk)

# NOTEs:

# <http://www.jbiomedsem.com/info/instructions/>

# <http://obi-ontology.org/page/Sig_paper_vote>

# Abstract

## Background

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.

## Results

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.

## Conclusions

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

# Background

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].

This work by the transcriptomics, proteomics and metabolomics communities were originally developed 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 [Smith *et.al*., 2007] which requires all member ontologies to define a single class only once (orthogonality) . Besides common naming conventions [Schober et al. 2009], 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, Brinkman 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.

# Results

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 [ref?] is inadequate for biology. The ‘material entity’ class will be submitted to BFO for inclusion. 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>).

In order to a) test whether OBI in its present state is already sufficiently complete for a practical annotation effort and b) to show how such an annotation will look like, we selected two simple but typical modeling use cases. These 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.

## We have to stress that, as such, we model parts of reality itself and not some database entries or pieces of text. In this respect the two examples given are self-sufficient models and do not contain any bindings to database entries or text. Hence no instances can presently be found, which usual annotation efforts would require.

## Use case 1: Dirk’s neuroscience use case

The experiment tries studies the role of the caudate nucleus in the control of action and more specifically targets the reward factor. It is found that the caudate nucleus is active before the actual presentation of the stimulus and fires more for the reward related position.

It contains two processes:

1. ‘stimulating monkey with flash light’ is modeled as an ‘stimulating host with stimulus’ that realizes some ‘stimulus role’ borne by some flash light and some ‘host role’ borne by some ‘monkey’. The ‘flash light’ is an independent\_continuant.
2. ‘measuring neural activity in monkey caudate nucleus’ is modeled as a ‘measuring in vivo activity’ process that realizes some ‘host role’ borne by some ‘monkey’. It has specified output of spike train measurement. The measurement is measured by device ‘SU device + electrodes’. The location is caudate nucleus.

….

## Use case 2: Vaccine protection study

The vaccine protection study (or called 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).

‘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).

‘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).

## Use case 3: Larisa’s use case

Robot Scientist “Adam” is one of the most advanced laboratory automation systems in existence [King et al 2009]. Adam is designed to measure, in high-throughput, growth curves (phenotypes) of selected microbial strains (genotypes) in a defined media (environment). Adam is able to run “lights out” for days at a time, and is capable of designing and initiating >1,000 new strain/defined growth-medium experiments each day (from a selection of several thousand yeast strains).

The Robots require a complete, precise, and fully formalised description of all experiment actions (for example, if you forgot to include an action “remove a plate lid”, it will try to pierce pipettes through the lid). Robot Scientists provide unsurpassed test beds for the development of technologies for the curation and annotation of experiment processes: for the first time it is possible to completely capture and digitally curate all aspects of the scientific process to record and fully understand how and why a particular experiment was conceived and executed, and to remove all subjectivity in experiment actions. This enables all aspects of investigation to be fully repeatable. The Robot Scientist project led to the first fully automatic discovery of new scientific knowledge, and has the potential to revolutionise the way science is done [Science, reply].

We can consider three typical processes: planning of an experiment, executing an assay, and analysing the results (fig 1,2,3). We can discuss here robot has role or function "investigator".

References: King et al. (2009) The Automation of Science. Science 324: 85-89. King, et al. (2009) Make Way for Robot Scientists. Science 325: 945. King, et al.(2009) The Robot Scientist Adam. Computer 42/8: 46-54. King et al. (2004) Functional Genomics Hypothesis Generation by a Robot Scientist, Nature 427(6971): 247-252.

Oliver: Here I provided some draft idea here for further discussion:

It contains two processes:

1. ‘performing plate layout’ is modeled as an ‘material combination’ (OBI\_0000652) that has specific input of metabolites and yeast strain delta\_YER152C. The Robot Scientist “Adam” will have ‘investigator role’.
2. ‘analyzing growth curve’ is modeled as an ‘assay’ (OBI\_0000070) that has specific input of yeast strain, and specific output of growth curve. The yeast has role of ‘evaluent role’….

Note: see what you think …

# Discussion

OBI aims for a robust and versatile symbolic representation of biomedical investigations. Two distinct biological use cases are modeled by representing statements using representational units defined in 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. 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’.

We found that all these required processe, as well as all other entities described in the use cases could be represented using OBI idioms. . 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 standard 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.

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, 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 repre-sentatives of the communities, members of the OBI consor-tium, 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.

# Conclusions

Text for this section.

# Methods

## Sub- heading for this section

Text for this sub-section.

## Sub- heading for this section

Text for this sub-section.

## Sub- heading for this section

Text for this sub-section.

# Competing interests

Text for this section.

# Authors' contributions

Please see sample text in the instructions for authors.

# 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

1. Shuman S: **Structure, mechanism, and evolution of the mRNA capping apparatus.** *Prog. Nucleic Acid Res. Mol. Biol* 2000, **66**:1-40

(NOTE: The above is the reference template)

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., 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) Biodynamic Ontology: Applying BFO in the Biomedical Domain](http://ontology.buffalo.edu/medo/biodynamic.pdf), *Ontologies in Medicine*, IOS Press: 20-32.

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

Schober D. Smith B, Lewis S, Kusnierczyk W, Lomax J, Mungall C, Taylor C, Rocca-Serra P, Sansone AS: Naming Conventions for OBO Foundry Ontology engineering. BMC Bioinformatics 2009, 10:125, <http://www.biomedcentral.com/1471-2105/10/125>

Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, Mungall CJ, Neuhaus F, Rector A, Rosse C (2005) [Relations in Biomedical Ontologies](http://genomebiology.com/2005/6/5/R46). *Genome Biology*, **6**: R46

Smith 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., Parkinson, H. E, Causton, H. C., Fan, L., Fostel, J., et al. (2006) The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics*, **7**: 866-873.

Whetzel, P. L., R. R. Brinkman, et al. (2006). "Development of FuGO: an ontology for functional genomics investigations." Omics **10**(2): 199-204.

Xiang Z, Todd T, Ku KP, Kovacic BL, Larson CB, 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>.

# Figures

## Figure 1 - OBI modeling of … study.

Figure legend text

## Figure 2 - OBI modeling of vaccine protection study.

Figure legend text

## Figure 3 - Hierarchical structure of processes used in the two use cases.

Figure legend text.

…

# Tables

## Table 1 - Sample table title

Table legend text.

## Table 2 - Another sample table title

Table legend text.

…

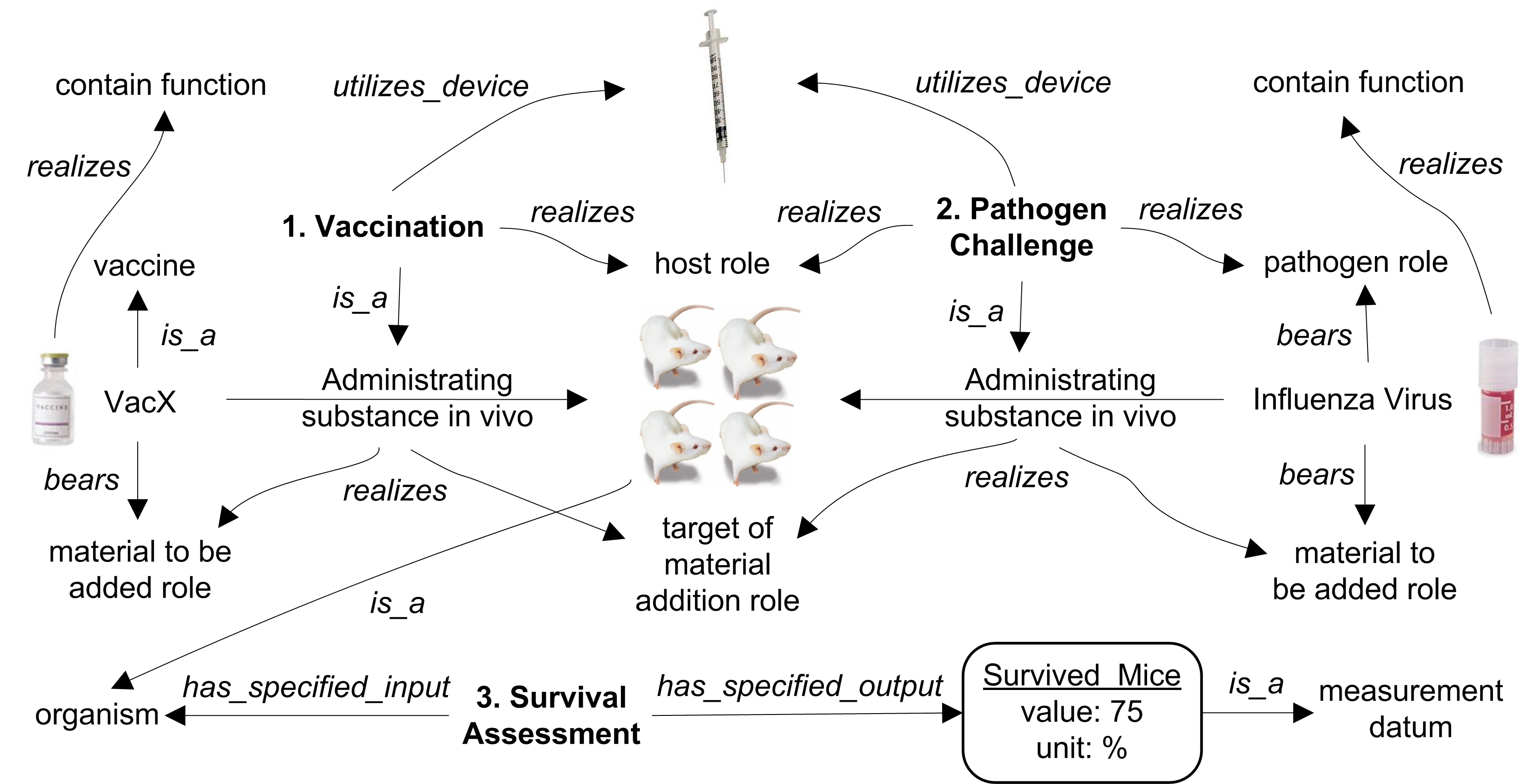
# Additional files

### Additional file 1 – Sample additional file title

Additional file descriptions text (including details of how to view the file, if it is in a non-standard format).



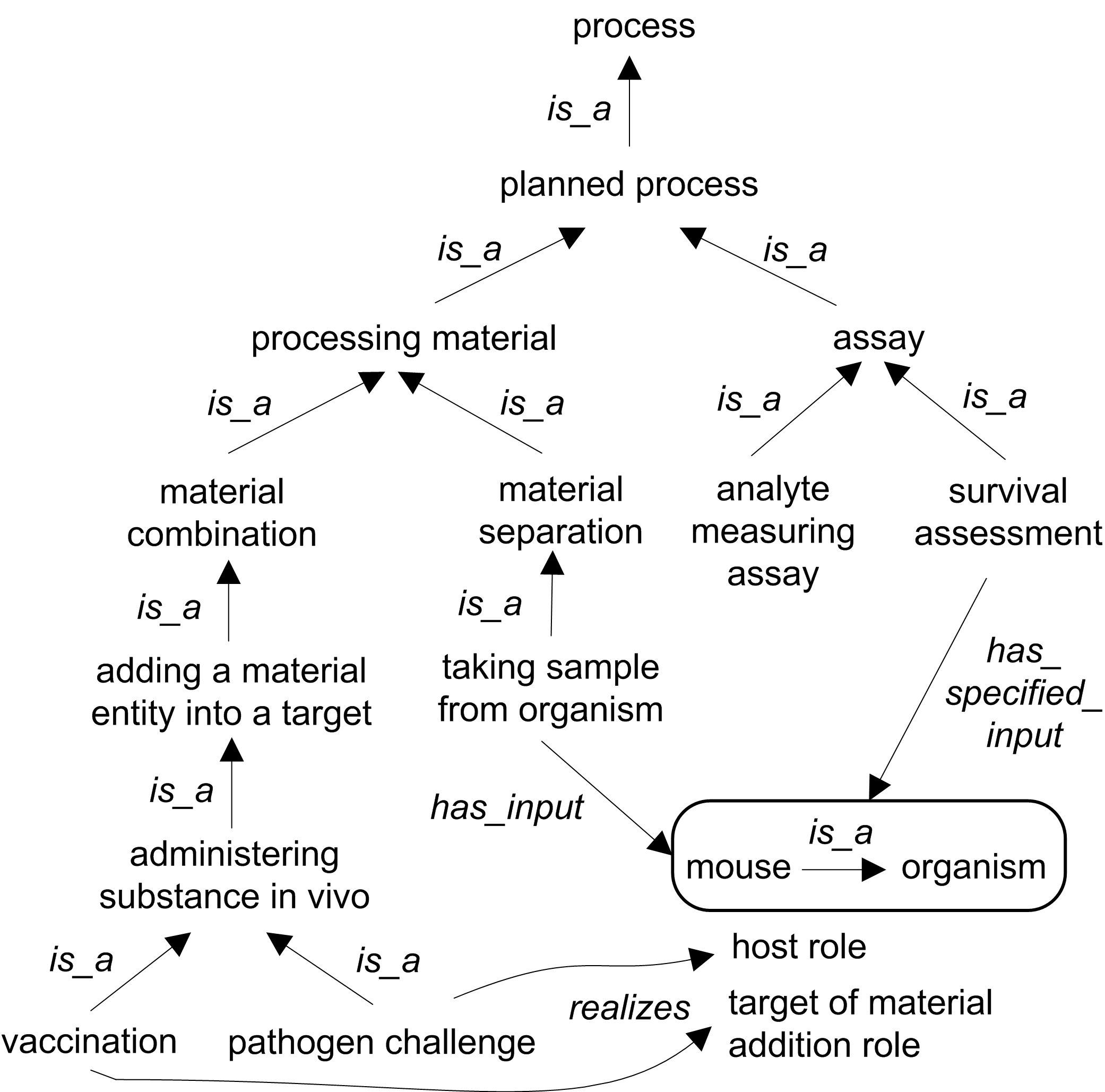
**Figure 1.** Dirk’s use case here (Draft by Oliver He based on Dirk’s wiki document and OBI devel discussion, may not accurate☺ ).



**Figure 2.** OBI modeling of vaccine protection study.

**Figure 3.** Larisa’s use case.

Oliver NOTE: The order of the three use cases can be discussed later.



**Figure 4.** Hierarchical structure of processes used in the three use cases.

Oliver NOTE: This is still the original SIG paper figure. It is possible to make a similar one with the new three use cases.