Mapping Data Annotated with the Minimum Information About BIobank data Sharing (MIABIS) 2.0 to the Ontology for Biobanking (OBIB)

Sarah Bost[a], Niina Eklund[b], Kaisa Silander[b], Jie Zheng[c], Christian J. Stoeckert[c], Mathias Brochhausen[a]

[a] *Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States,*
[b] *National Institute for Health and Welfare (THL), Helsinki, Finland,*
[c] *Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States*

Abstract

MIABIS 2.0, an updated, modular version of MIABIS, is a set of minimum data elements for biobanks and studies using human biospecimens. Though MIABIS was previously ontologized and became part of the Ontology for Biobanking (OBIB), MIABIS 2.0 necessitated an updated mapping and extension of OBIB. A working group of MIABIS 2.0 and OBIB developers communicated via teleconference and email to ensure that the new concepts were accurately represented in the ontology either by existing OBIB classes, classes that could be imported from other OBO Foundry ontologies, or in newly-requested terms. Mapping for 84.44% of the elements has been completed and we expect to conclude the process by July 2019. Our efforts show that OBO Foundry ontologies support the mapping of developing sets of standard data elements but that one-to-one mappings are not always been feasible.

Keywords:

Biobanking, Biomedical ontologies, Mappings

Introduction

Biobanks are a key resource for novel developments in biomedical research and healthcare, but in order to be useful, information therein must be easily accessible to researchers. To meet this need, the Minimum Information About BIobank data Sharing (MIABIS) was developed by the Biobanking and BioMolecular Resources Research Infrastructure of Sweden (BBMRI.se) (1). Brochhausen *et al.* converted MIABIS into an ontology of biobank administration called Ontologized MIABIS (OMIABIS) (2). In 2014, OMIABIS was merged with the Biobank Ontology (BO), a more specimen-focused ontology created at the University of Pennsylvania, to create the Ontology for Biobanking (OBIB) (3).

MIABIS was widely adopted and in 2013, a working group began the continued development of the standard as a modular structure, leading to MIABIS 2.0 (4). MIABIS 2.0 presents a major extension of the original MIABIS. Hence, to ensure a complete mapping of data annotated with MIABIS 2.0 to OBIB, an updated mapping is necessary. Providing a mapping file also necessitated an extension of OBIB.

Methods

Best practice in ontology development requires a clear understanding of the entities to be represented and following established principles when writing definitions. Since the MIABIS 2.0 working group did not include any OBIB developers, a small working group of OBIB developers and MIABIS 2.0 developers was established to provide the mapping and ensure continued consistency between data captured with MIABIS 2.0 and OBIB-annotated data. In telephone conference and by asynchronous communication, MIABIS data elements and their descriptions were explained and potential OBIB mappings and alternative representations were suggested. Finally, each mapping was discussed and agreed upon by the working group.

In accordance with best practices (5), we also reviewed OBO Foundry ontologies for existing classes that could be mapped to MIABIS 2.0 terms. For instance, 'specimen from organism,' already present in the Ontology for Biomedical Investigations (OBI) (6), mapped to 'biological samples.'

If a MIABIS term was not mappable to any term in an OBO Foundry ontology, we requested it be added. For example, while a number of FFPE (formalin-fixed paraffin-embedded) protocols were represented in the OBO Foundry, no ontologies included an 'FFPE specimen' class. We rectified this by submitting a request to the OBI issue tracker [https://github.com/obi-ontology/obi/issues/1019]. Once all of our required terms are created, they will be imported to OBIB and the mapping table will be finalized.

Results

Currently, 84.44% of MIABIS 2.0 elements are mapped to classes in OBIB or RDF statements using OBIB classes or relationships. We expect the totality of all MIABIS 2.0 elements to be mapped by July 2019. The mappings are available here: https://docs.google.com/spreadsheets/d/1CnGP7O3_qzXhzd4lKZI7o3Ig4cxksWDdANw-1qG3R-8.

Discussion

Some of the mappings link to generic classes from OBO Foundry ontologies and do not by themselves represent the entirety of the meaning intended by MIABIS. For example, MIABIS captures information about the juristic person (the institution or organization) twice: once in the context of the biobank component and once in the context of the sample collection and study components. It is obvious that this difference in contexts must be represented in the mappings. The case of the juristic person in the biobank context is fairly straightforward. It is defined in MIABIS 2.0 as a "Textual string of letters denoting the juristic person e.g. a university, concern, county council etc. for the biobank." OBIB contains a one-to-one mapping for that specific meaning: "biobank organization" or alternatively, "biobank juristic person" [http://www.ontobee.org/ontology/OBIB?iri=http://purl.obolibrary.org/obo/OMIABIS_0000010]. The textual definition for this class is "An organization bearing legal personality that owns or administrates a biobank." Axiomatically, the class is defined as being equivalent to:

> organization AND ((owns SOME biobank) OR (administrates SOME biobank)) AND (bearer of SOME legal person role).

A one-to-one mapping for this data element exists because this element was included in MIABIS 1.0 and a class representing the exact meaning in OWL was created during the OMIABIS curation.

The second use of juristic person is completely different. The data element defines juristic person as "Textual string of letters denoting the juristic person e.g. a university, concern, county council etc." The fact that this element is part of the "Researcher information" subset suggests that this field records the juridic entity with which a researcher is affiliated. We propose to map this data element to the OMRSE class "legal person role" [http://purl.obolibrary.org/obo/OMRSE_00000038]. Obviously, that role alone does not convey the entire meaning of the data element in question. This means that we have two approaches: (1) create an OWL class that can be mapped one-to-one to the data element, or (2) map this data element to an RDF pattern without enforcing a one-to-one mapping.

In the first approach, we would create a class with a necessary and sufficient condition like:

> (Homo sapiens AND bearer of SOME investigation agent role) AND is member of organization SOME (organization AND bearer of SOME legal person role)

Such a class would enable a one-to-one mapping for the data element in question. However, providing an axiomatically-rich and very specific class for every data element or data item that may occur in managing medical data would come at the burden of creating an ontology that takes very long to reason over, even if just for the sake of checking its consistency.

With the second approach, we only need to ensure that our data about each researcher is represented in a way that we have information about the organization that has legal personhood. For each researcher we then need to represent of which legal personality they are a member. The legal person role itself can remain anonymous. Once the information is represented as such in RDF, we can run a SPARQL query to retrieve the same class of objects that we defined above. The advantage of this approach is that the computational cost of the query is only paid when the query to retrieve the data is run and not every time the ontology or a triple store containing the ontology is checked for consistency.

Conclusions

The results demonstrate the ability of OBO Foundry ontologies and their maintenance procedures to foster mappings for evolving sets of standard data elements, such as MIABIS. It also shows that one-to-one mappings from OWL classes to data elements can be problematic. There is a natural mismatch between the fundamental information representation paradigms of tabular data representation and semantic web methodologies.

Address for correspondence

Sarah Bost, sbost@uams.edu

References

1. Norlin L, Fransson MN, Eriksson M, Merino-Martinez R, Anderberg M, Kurtovic S, et al. A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS. Biopreservation Biobanking. 2012 Aug;10(4):343–8.

2. Brochhausen M, Fransson MN, Kanaskar NV, Eriksson M, Merino-Martinez R, Hall RA, et al. Developing a semantically rich ontology for the biobank-administration domain. J Biomed Semant. 2013;4(1):23–23.

3. Brochhausen M, Zheng J, Birtwell D, Williams H, Masci AM, Ellis HJ, et al. OBIB-a novel ontology for biobanking. J Biomed Semant. 2016 May 2;7(1):23.

4. Merino-Martinez R, Norlin L, van Enckevort D, Anton G, Schuffenhauer S, Silander K, et al. Toward Global Biobank Integration by Implementation of the Minimum Information About BIobank Data Sharing (MIABIS 2.0 Core). Biopreservation Biobanking. 2016 Aug;14(4):298–306.

5. Principles: Overview [Internet]. The OBO Foundry. [cited 2019 May 23]. Available from: http://www.obofoundry.org/principles/fp-000-summary.html

6. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. PLOS ONE. 2016 Apr 29;11(4):e0154556.