Towards a Common Semantic Representation of Informed Consent for Biobank Specimens

Frank J. Manion[1*], Yongqun He[2], Elizabeth Eisenhauer[3], Yu Lin[2], Alla Karnovsky[4], Marcelline R. Harris[3]

[1] Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109
[2] Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, and Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA
[3] Division of Systems Leadership and Effectiveness Science, University of Michigan School of Nursing, Ann Arbor, MI 48109
[4] Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109

***Abstract* —** Biospecimen-based research is rapidly growing in the post genomic era, and includes the need to retrieve specimens from distributed biobanks of various size and complexity in a fashion that ethically preserves the expressed wishes of specimen donors as represented by the informed consent process and its artifacts. This paper briefly describes existing work along these lines, presents some challenges unique to biobanks, and presents our own work on an ontology of informed consent.

Keywords— BioBank, Informed consent; ontology; ICO; OBO Foundry, Basic Formal Ontology (BFO), OBI ontology

I. INTRODUCTION

Research in the post-genomic era requires access to high quality biospeciments, often annotated with or linked to clinical data. Many groups at varying levels of institutional complexity, ranging from small scale individual laboratories to distributed international collaboratives, have established and operate biorepositories (also refered to by various names such as *biobanks*, *biolibraries*, and even *collections*). Often, there are needs to share data and specimens among multiple biobanks [1-3]. The act of requesting specimens from a biorespository may demand a complex series of transactions, each of which in turn may convey a series of rights, obligations, and permissions for access to specimens and data. Despite over a decade of experience incorporating biospecimens in the research process, formal models that describe the use of biorepositories in human research are a relatively recent development. Without a common formal model of consent and the associated permissions on collection and distribution of specimens and data, integration of data across the translational spectrum, or from multiple banks and institutions will remain a difficult, manually intensive problem.

In this paper we briefly review current efforts toward such models, describe our own work toward a formal model for informed consent, and describe what we consider challenges and opportunities for supporting biorespository-based research with ontologies. A simple example that provides a motivation for this effort follows.

II. EXAMPLE OF THE CHALLENGE

Clinical or translational research often involves the extraction and usage of biospecimen from humans. Different biospecimens may be stored and processed differently, and may be collected under different models of consent. A typical scenario might read something like this:

> "For my study, I want to use samples from my organizations's biobank, collected under a blanket biobank informed consent form. I discover that I will need more samples, so I contact another organization's biobank to determine if they hold relevant and available specimens. That organization's samples were collected under a tiered biobank informed consent form. While some samples are shared with me, I still need more samples to address the requirements of my study. I then collect additional samples using a consent form specific to my study."

In this example there are three informed consents forms to account for – a blanket consent, a tiered consent, and the investigator's single study consent. In an effort to support the expressed wishes of the donors, informed consent documents impose a series of legal and ethical restrictions, obligations, and permissions to biobank operators and research teams using the specimens and data collected in these banks. Often these rights, obligations, and permissions accrue from multiple sources of authority and are represented in multiple legal documents. Consequently, the biobanking domain presents a series of modeling challenges, including:

The operational model of the biobank. A biobank can be a single, dedicated resource that provides samples to single or closely allied groups of studies using a common consent model. It might be a virtual or distributed biorepository using precoordinated consent models. Another organization structure might be that of a shared biobank facility containing multiple sets of tissues from multiple projects and attempting to maximize use of these tissue resources by making them available to requestors.

The consent model used for the biobank. This can be opt-in or opt-out. In the case of an opt-in consent model, a tiered consent may be used to present the participant or volunteer with choices of the type of data the participant may want shared, and for what types of research or other constraints.

The protocol model the bank operates under. Typically a biobank serving more than one project would operate under one or more Institutional Review Board (IRB)-approved collection protocols and Health Insurance Portability and Accountability Act (HIPAA) authorizations. Researchers

* Corresponding author.

subsequently requesting specimens and data would operate under separate IRB-approved protocols, and depending on this protocol, separate consent and HIPAA authorization may be required for use of a previously banked specimen. Such a model is sometimes called a *two-protocol* model [4].

Rights, obligations, and permissions accrue from multiple sources and must be consistent across time. Properly modeling the decisions typically made by human review boards and regulatory personnel considering sample and data distribution for research requires modeling not just the consent documents, but the protocols, data use agreements, and possibly other information artifacts used in both depositing samples into a biobank, and withdrawing them for subsequent research.

In a research oriented university such as the University of Michigan, thousands of informed consent forms have been generated, and there are over 100 biobanks in the Medical School alone. Queries supporting appropriate use of banked biospecimens and data must be linked to the signed informed consent agremments with the biospecimen donor.

III. Existing Efforts

Several current efforts are evident, focused on modeling aspects of the biobanking domain. At least two BFO-aligned ontologies relate to biobanking. The Ontologized Minimum Information About BIobank data Sharing (OMIABIS) expresses data concepts in an ontology of biobank administration [5]. OMIABIS is based on work by Norlin and colleagues [6] to develop a minimum data set for eight countries participating in the EU Biobanking and Biomolecular Resources Research Infrastructure project. Limitations of this effort are that it is intended to serve only as a description of a biobank contents, and does not describe collection critera, consenting, and protocol provenance of individual specimens. A group at the University of Pennsylvania is developing an ontology for the representation of biobanks, although the work is in early stages [7]. Similarly, we are aware that a group at Duke University is working on a collaborative effort to develop a normative set of data elements and terms to recommend as best practice to the International Society for Biological and Environmental Repositories (ISBER), although this work is not yet published [8, 9].

There are also non-BFO aligned ontologies in related areas, including a Permission Ontology used for development and evaluation of software tools for reasoning about consent permission, published by a group at the University of California San Diego (UCSD) [10]. Related work to build a Research Permission Management System was done at the Medical University of South Carolina (MUSC) to support a statewide research network [11]. A search of the term "consent" in the NCBO biportal identified the notion of informed consent at the class level in 19 different systems (http://bioportal.bioontology.org/search).

Our efforts to develop a BFO-aligned informed consent ontology (ICO) emphasizes the broad domain of informed consent. Although motivated by a biobanking use case, initial development reported here is not restricted to that domain.

IV. The Informed Consent Ontology (ICO)

Development of ICO, a BFO-based ontology represented in the Web Ontology Language (OWL2) [12], follows OBO Foundry principles of openness and collaboration. ICO is aligned with the BFO [13], making it possible to align and integrate with other BFO-based ontologies. The initial release of the ontology focuses on modeling informed consent documents. As for Aug. 14, 2014, ICO contains 471 terms including 137 ICO-specific terms and other terms imported from other BFO-aligned ontologies. Detailed ICO statistics can be found on the Ontobee ICO web page: http://www.ontobee.org/ontostat.php?ontology=ICO. ICO is released under an open Creative Commons 3.0 License.

The ontology was developed using a combination of top-down and bottom-up approaches. Protégé-OWL 4.2 was used for the ontology authoring and editing. To build the OBI-based framework of ICO we manually identified informed consent concepts from existing OBO Foundry library ontologies. These were imported to ICO using Ontodog [14] and OntoFox [15] which allowed for recursive inclusion of all defined axioms and related terms. The results were then manually reviewed for final approval before inclusion in the ICO framework.

Bottom up construction proceeded by manually identifying and extracting a list of candidate terms from two informed consent templates used at the University of Michigan (one from the Medical School Institutional Review Board, another from the Health Sciences and Behavioral Sciences Institutional Review Board). We also identified terms from a consent form used for the University of Michigan Medical School biorepository, and from World Health Organization (WHO) informed consent templates. The candidate terms identified from these templates were then enriched with metadata including definitions, concept identifiers, preferred terms, synonyms, and URIs extracted from three ontology repositories: the National Library of Medicine's Unified Medical Language System (UMLS®) Metathesaurus [16]; the National Center for Biomedical Ontology (NCBO) BioPortal [17]; and Ontobee [18]. When textual definitions were not provided, other sources such as clinical research glossaries or the current literature were used. These enriched candidate terms were manually mapped to several pre-identified resources containing terms and definitions developed and vetted by the United States regulatory community. This process yielded candidate preferred terms contining definitions accepted as robust and well defined by that community. Resources used in this step included the National Cancer Institute Thesaurus (NCIt), the Biomedical Research Integrated Domain Group (BRIDG) [19], the Ontology of Clinical Research (OCRe) [20], the Consumer Health Vocabulary (CHV) and the University of California San Diego permission ontology [10].

The pool of enriched candidate terms was organized into categories of like terms according to their definitions. For example, the category 'authorization' included the terms '*authorization for medical records release*', '*authorization documentation*' or '*authorization*'. Enriched candidate terms grouped by categories formed to-be-included terms in ICO. The final set of categories (or modeling units) was then

mapped to branches of BFO. For example, terms categorized under 'authorization' were considered to be subclasses of BFO:process. Informed consent workflows in a typical clinical research study were modeled as three processes: (i) pre-informed consent processes, (ii) obtaining informed consent processes, and (iii) processes after signing informed consent documents. Relations between entities involved in the above processes were defined. Finally, all terms and relations were aligned with BFO.

V. Discussion and Conclusions

Modeling informed consent is a necessary but not sufficient part of the modeling needed to support responsible use of biospecimens and data in research. Biospecimen and data release is complex, and informed consent plays a major role in the regulatory and scientific governance used by biorepositories to release specimens and data. In follow on work we plan to examine the specific area of specimen and data release involving the longitudinal agreements of rights, permissions, and obligations. Other work is needed in the complex areas of protocol representation, data use agreements and material transfer agreements.

Limitations of our preliminary work will inform further development efforts toward a robust Informed Consent Ontology. First, the ICO is admittedly preliminary work and is currently focused on informed consent documents and processes. More work is needed to validate the coverage and completeness in the domain. Concepts from the US Common Rule and the EU Prior Informed Consent legislation need to be included. Our current models of informed consent processes likely lack the richness and complexity of real-life informed consent processes, and they need validation with research study teams from a variety of domain areas. Aspects of rights, obligations, permissions, and ethics must be modeled and used to extend the ontology. Finally, axioms must be developed and competency validation of the ICO must be conducted using a series of still to be defined use case derived competency questions.

We have described our work on ICO, a preliminary ontology of informed consent that provides general classification of content contained in general informed consent documents. It requires expansion, revisions and collaboration to build a robust model, and to move toward a representation of the complex area of biobank data sharing and specimen release. We hope to collaborate with the broader community in this effort.

Acknowledgment

This research was supported by a University of Michigan interdisciplinary research award (MCubed) and by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number 2UL1TR000433-06. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] A. Cambon-Thomsen, E. Rial-Sebbag, and B. M. Knoppers, "Trends in ethical and legal frameworks for the use of human biobanks," *Eur Respir J*, vol. 30, pp. 373-82, Aug 2007.

[2] R. E. Hewitt, "Biobanking: the foundation of personalized medicine," *Curr Opin Oncol*, vol. 23, pp. 112-9, Jan 2011.

[3] G. E. Henderson, R. J. Cadigan, T. P. Edwards, I. Conlon, A. G. Nelson, J. P. Evans, A. M. Davis, C. Zimmer, and B. J. Weiner, "Characterizing biobank organizations in the U.S.: results from a national survey," *Genome Medicine*, vol. 5, no. 1, p. 3, Jan. 2013.

[4] F. J. Manion, R. J. Robbins, W. A. Weems, and R. S. Crowley, "Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study," *BMC Med Inform Decis Mak*, vol. 9, p. 31, 2009.

[5] M. Brochhausen, M. N. Fransson, N. V. Kanaskar, M. Eriksson, R. Merino-Martinez, R. A. Hall, L. Norlin, S. Kjellqvist, M. Hortlund, U. Topaloglu, W. R. Hogan, and J.-E. Litton, "Developing a semantically rich ontology for the biobank-administration domain," *J Biomed Semantics*, vol. 4, p. 23, 2013.

[6] Norlin, M. N. Fransson, M. Eriksson, R. Merino-Martinez, M. Anderberg, S. Kurtovic, and J.-E. Litton, "A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS," *Biopreservation and Biobanking*, vol. 10, no. 4, pp. 343–348, Aug. 2012.

[7] J. Zheng, Personal Communication to Yongqun He, 2014.

[8] H. Ellis, Personal Communication to Frank J. Manion, 2014.

[9] ISBER. (2014). ISBER - International Society for Biological and Environmental Repositories. Available: http://www.isber.org/

[10] A. Grando and R. Schwab, "Building and evaluating an ontology-based tool for reasoning about consent permission," *AMIA Annu Symp Proc*, vol. 2013, pp. 514-23, 2013.

[11] J. S. Obeid, K. Gerken, K. C. Madathil, D. Rugg, C. E. Alstad, K. Fryar, R. Alexander, A. K. Gramopadhye, J. Moskowitz, and I. C. Sanderson, "Development of an Electronic Research Permissions Management System to Enhance Informed Consents and Capture Research Authorizations Data," *AMIA Summits Transl Sci Proc*, vol. 2013, pp. 189–193, 18 2013.

[12] W3C, "OWL 2 Web Ontology Language document overview," pp. http://www.w3.org/TR/2009/REC-owl2-overview-20091027/. Accessed on March 1, 2014, 2009.

[13] P. Grenon and B. Smith, "SNAP and SPAN: Towards dynamic spatial ontology," *Spatial Cognition and Computation*, vol. 4, pp. 69–103, 2004.

[14] J. Zheng, Z. Xiang, C. J. Stoeckert, and Y. He, "Ontodog: a web-based ontology community view generation tool," *Bioinformatics*, vol. 30, no. 9, pp. 1340–1342, May 2014.

[15] Z. Xiang, M. Courtot, R. R. Brinkman, A. Ruttenberg, and Y. He, "OntoFox: web-based support for ontology reuse," *BMC Res Notes*, vol. 3, p. 175, Jun. 2010.

[16] NLM, "U.S. National Library of Medicine (NLM). Unified Medical Language System (UMLS®) Metathesaurus [version 2013AA][Internet]. Bethesda (MD): National Library of Medicine," 2013.

[17] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen, "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications," *Nucleic Acids Res*, vol. 39, no. Web Server issue, pp. W541–W545, Jul. 2011.

[18] Z. Xiang, C. Mungall, A. Ruttenberg, and Y. He, "Ontobee: A Linked Data Server and Browser for Ontology Terms.," in *Proceedings of the 2nd International Conference on Biomedical Ontologies (ICBO)*, Buffalo, NY, USA, 2011, pp. 279–281.

[19] D. B. Fridsma, J. Evans, and C. N. Mead, "The BRIDG Project: A Technical Report," *Journal of the American Medical Informatics Association*, vol. 15, no. 2, p. 130, 2008.

[20] I. Sim, S. W. Tu, S. Carini, H. P. Lehmann, B. H. Pollock, M. Peleg, and K. M. Wittkowski, "The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research," *Journal of Biomedical Informatics*, Nov 13 2013.