A Biomedical Research Permissions Ontology: Cognitive and Knowledge Representation Considerations

Jihad Obeid, MD
Medical University of South Carolina
125 Doughty St.
Charleston, SC, 29425
jobeid@musc.edu

Davera Gabriel, RN
University of California, Davis
2921 Stockton Blvd
Sacramento, CA , 95817
davera@ucdavis.edu

Iain Sanderson, M.Sc, FRCA
Medical University of South Carolina
125 Doughty St.
Charleston, SC, 29425
sandersi@musc.edu

ABSTRACT
In designing a comprehensive mechanism for managing informed consents and permissions for biomedical research involving human participants, a significant effort is dedicated to the development of standardized classification of these consents and permissions. In this paper, we describe the considerations and implications of this effort that should be addressed during the development of a Biomedical Research Permissions Ontology (RPO). It is hoped that this standardization will allow disparate research institutions to pool research data and associated consents and permissions in order to facilitate collaborative translational research projects across multiple institutions and subsequent new breakthroughs in medicine while providing: 1) essential built in protections for privacy and confidentiality of research participants and 2) a mechanism for insuring that researchers adhere to patient's intent whether to participate in research or not.

Categories and Subject Descriptors
K.4.1 [Public Policy Issue]: Privacy.

General Terms
Human Factors, Standardization.

Keywords
ACM proceedings, Biomedical Research, Ontology, Research Permissions, Patient, Consent, Privacy.

1. INTRODUCTION
Obtaining an informed consent for participation in biomedical research is a legal requirement and an ethical obligation. It is an indispensible component of the research process involving human participants [3]. As stated in the US Code of Federal Regulations, an informed consent must address at a minimum several basic elements, including, but not limited to, a description of the proposed research, risks and benefits, appropriate alternatives, and contact information [9]. Although there have been significant strides in computerizing medical records and clinical trial systems in recent years, the informed consent process in most research institutions remains predominantly paper-based. As such informed consents and other permissions related to research are non-standardized, giving rise to ambiguity and the potential for misinterpretation. Inconsistency in comprehension and interpretation exposes participants on one hand, and clinical investigators and research institutions on the other, to health and medicolegal risks respectively [14].

Moving consent management to an electronic platform has the potential to standardize collection, sharing and retrieval of research permissions across institutions, remove ambiguity and provide better education during the consent process using multimedia which is more readily available in electronic format thus rendering consents more informed.

GTIP 2010 Dec. 7, 2010, Austin, Texas USA

2. BACKGROUND
Health Sciences South Carolina (HSSC), a collaborative of three principal research universities and four major health systems across the state of South Carolina, is developing a Research Permissions Management System (RPMS) that will provide a comprehensive mechanism for managing informed consents and other research permissions. An essential component in this effort is the development of a Research Permissions Ontology. This will enable the research permissions information from multiple institutions to be combined into a single computable data representation. It will provide the semantic foundation for representing and validating permissions data in a variety of data capture forms, relational databases and portlets that could be expressed via a web-based system or embedded in point-of-care clinical and research applications..

3. THE ONTOLOGY
The development of the RPO began with the analysis of the permission processes at four HSSC member medical institutions. The terminology and language in the various hospital forms, Institutional Review Board (IRB) templates, and hospital privacy practice notices were reviewed along with government websites, United States Department of Health and Human Services (HHS), the Food and Drug Administration (FDA) and the Office for Civil Rights in order to ensure a more comprehensive analysis of Health Insurance Portability and Accountability Act (HIPAA) and other federal regulations. Moreover other bodies of standards work were examined with particular attention to permission and consent data standards where applicable including Healthcare Information Technology Standards Panel (HITSP) [15], HL7 Composite Privacy Consent Directive Domain Analysis Model [12] and the Integrating the Healthcare Enterprise (IHE) Basic Patient Privacy

Consents (BPPC). Other known ontologies such as SNOMED and National Cancer Institute (NCI) thesaurus [20] were investigated to look for existing concepts related to permissions and consents that could be leveraged for this effort. For example many of the concepts in RPO are rooted in NCI thesaurus concepts such as document types and HIPAA authorization.

3.1 Methods

After an extensive analysis of workflow in various research institutions as described above and examination of previous work, a list of potential terms and concepts that related to the research permission process was created. Protégé was used to layout the hierarchal order of the ontology [11]. Many key concepts, synonyms and hierarchal "is-a" relationships that were already defined in the NCI thesaurus, were imported directly into the Protégé ontology and added to the new concepts and classes that specifically address research permissions. The classification of permissions was then reviewed by domain experts in the healthcare research regulatory domain. This work is also being validated by the newly formed subgroup of the Data Standards and Interoperability Affinity Group (DIAG), the Permissions Ontology DIAG Subgroup (PODS) in the Clinical and Translational Science Award (CTSA) consortium [8].

3.2 Content

The Web Ontology Language, or OWL is a knowledge representation language for authoring ontologies and was used in developing PRO. Every node in an OWL representation is a member of the class OWL: thing, and the remainder of ontology is represented as a subclass of this root node. The RPO begins with some top level concepts under research permissions such as: consent, permission to contact, assent, and parental permission. Other research permissions concepts were added under other top level branches. For example under document additions included waiver of consent, authorization, and notice of privacy practices (figure 1).

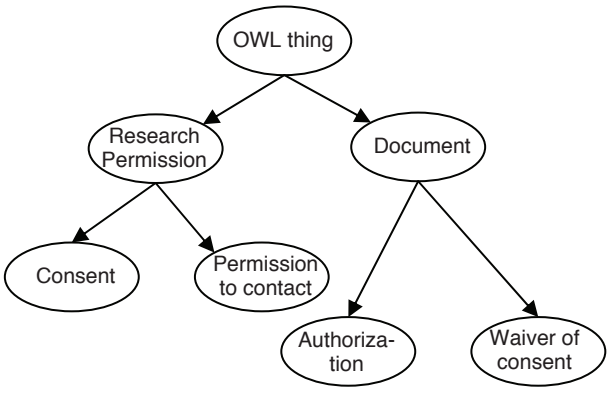

Figure 1: A portion of the ontology showing some top level concepts.

The representation of a multitude of permutations of diverse research activities was a challenging task and is an ongoing activity of PODS. For example the concept of permissions to donate tissue for research, has to be considered in a variety of potential tissue collection protocols in addition to the level of personal identifiers retained with specimens (fully identified vs. no identifiers with potential links to patients vs. no identifiers without possibility of linking back to patients).

The relationships between concepts goes beyond "is-a" parent-child relations. There are several cases of interrelationships and dependencies of the classes or concepts. For example, a consent is given by a patient, documented on a specific consent form and observed by a witness. Using Protégé, we were able to start defining and tying some of these interrelationships together as a framework to help support future computable reasoning.

4. DISCUSSION

4.1 Human Interface

One of the explicit requirements for gaining consent for treatment and the goals of an IRB is to protect the participant and thus the requirement to engage in an informed consent process. This is intended to support the patient or participant decision-making with the knowledge of the risks and benefits of participating in the treatment and / or research process. Much of the evidence base in human decision-making has focused on cognitive processes as observed in the behaviors of subjects taking part in the risk / reward outcomes of gambling..

In a series of gambling experiments, subjects were observed to favor decisions based on covert biases over those involving overt reasoning based on available facts. The ability to overcome innate biases was only achieved by repetitive experiences, thus allowing participants to develop a situational knowledge base as a result of a repetitive series of experiences [2, 4]. In another test, participants were observed to use emotionally-derived information to enhance their decision-making. Both of these experiments used repetitive experiences as conditions relating to the decision-making processes. Other experiments involve an evaluation of success as expressed as a percentage of success of winning [5]. These studies showed that there are parallel cognitive and emotional paths to decision-making, and that executive functions are only employed when the participants have some knowledge of the otherwise ambiguous process. Conceivably, these laboratory scenarios have corollaries to informed consent decision-making and the expression of risk involved in receiving a treatment or undergoing a procedure. Many of these lab cognitive scenarios can be correlated to patient medical decision-making, even as the presence of a laboratory "knowledge bias" may manifest in the form of the IRB-required education preceding research and medical consent. As such, we proposed that informed consent decision-making by participants and patients falls into similar cognitive processing patterns as observed in the gambling experiments...

Unlike the gambling experiments, patients may or may not have the opportunity to achieve repeated experiences of procedures, nor have any knowledge of what types of test may be performed on their tissue in the future. Therefore there is evidence to support that cognitively, a participant providing permission may be overtly affected in their permission decisions more by their innate biases and the lack of prior experience. Thus it is of particular concern to the development of permission ontology to accurately reflect and provide as much detail concerning the actual context of the consent or permission to effectively represent the intentions of the participants for consumption by a machine processor.

Recent studies have focused on not making a decision and consequences of voluntary omission of an action. These processes are difficult to study behaviorally due to the lack of a dependent measure. Nonetheless, neurobiological scans reveal that omissions to act are stored in the brain exactly the same way that an actual negation of an act is [16]. However, when an act is

voluntarily omitted, this information is stored differently than when subjects are instructed to not act as a part of the study methodology [17]. Thus it is of particular ethical concern to the development of the RPO to accurately reflect the explicit and intentional refusal to give permission and permissions gathering applications a user-interface design which captures that information accordingly.

.

4.2 Ontological Issues

4.2.1 Ambiguity and Context

Ambiguity is a given feature of most spoken and written language. Disambiguation is the pursuit of computational linguists and computer scientists. The goal of disambiguation in ontology development is to transfer knowledge or the intent of a thought or spoken concept to that of a reliably replicated and computable knowledge unit. The aforementioned goal of the RPO is to provide an encoded, computable unit representing the intention of a research participant after receiving an appropriate level of information regarding the research process, methodology, risks and potential benefits of participating in a research treatment or protocol. Nonetheless there remain linguistic and cultural factors inherent in the current, paper-based informed consent process that may lead to some language-based ambiguity. In clinical care, for example, if someone assisting with the informed consent process does not speak the native language of the person receiving treatment, it is less likely that the patient will receive documentation about procedures to be performed [21], they are more likely to receive less overall health education, worse interpersonal care and experience lower overall satisfaction with the care process [19]. This phenomenon, which should be compensated for under the scrutiny and per the mandate of a research Institutional Review Board (IRB) could, be compounded through the employ of data encoded in a RPO. This may arise simply by the fact that the permissions ontology unit of knowledge is gathered under an IRB-sanctioned set of human-aware circumstances then stored in an information system with the intention that it becomes executable knowledge later without the support of similar human processing. The act of accurately capturing and encoding the research participant's intention should take cultural and contextual queues into account as a part of some future interpretation, but may, in fact, fall short of that goal. As such, the issue of ambiguity in the development of RPO concepts is one of great concern, so as to minimize unintended consequences of a lack of spoken or written language context.

A recent survey of spoken languages has yielded the Linguistic Niche Hypothesis, which posits that languages developed in geographic or cultural isolation are morphologically more complex than those languages that service a wider geography or ethnic group, even though a high level of specification is unnecessary for communication [18]. It is not uncommon for medical colleagues or lay-persons to complain of "geek speak" when communicating with their more technical companions. It is of special concern to the development of the permissions ontology for the group to not succumb to this phenomenon and overly specialize or contextualize the concepts developed to present day circumstances without considering future states and data reuse implications, thus narrowing the utility of the RPO permission codes stored and accessed in the future.

4.2.2 Negation

One of the issues encountered when developing systems for knowledge representation, is the issue of negation, or negative findings and how that kind of information is modeled into the overall knowledge schema. Studies in the development of coding systems and ontologies for the purpose of representing in an electronic format the contents of a medical record yield numerous examples of negation, negative findings, or expressions of the absence of all or part of a health status or phenomenon [10]. In fact it is a requirement for reimbursement by the US Medicare and Medicaid regulations that patient records contain "abnormal and relevant negative findings" [6]. Ceusters, Elkin & Smith suggest that to adequately accommodate the variety of negation and negative findings required in health documentation that a health ontology provide a "lacks or lacks part of" relationship instantiated in an appropriately expressive medical ontology [7]. As a permissions ontology for use in research will need to address both ambiguously named and / or partially defined permission for use of data or tissue use in the future, it is a consideration of the project as to how to adequately express permissions that are explicitly not given as differentiated from those that are not expressly asked.

The Health Level Seven Version 3 Reference Information Model standard (HL7 V3 RIM) has as a feature of its data types a property called "nullFlavor" which aims to address shades of ambiguity of unknown data when available for use in a health data transaction. This feature within HL7 V3 accommodates the declaration of data in the same way that known data is represented, and further clarifies the data collection state beyond the simple declaration of "absent" versus "present." Although the use of the nullFlavor code set has its critics among health ontologists [22] and other health data standards developers, [1] conceivably this code system or a solution that is substantially similar may have application to address some of the negation ambiguity inherent in the collection of patient research permissions. In table 1 below, is a sample of the HL7 NullFlavor code set.

Table 1: A sample of HL7 V3 nullFlavor code set [13]

Code	Name	Description
NI	NoInformation	The value is exceptional (missing, omitted, incomplete, improper). No information as to the reason for being an exceptional value is provided. This is the most general exceptional value.
INV	invalid	The value as represented in the instance is not a member of the set of permitted data values in the constrained value domain of a variable.
DER	derived	An actual value may exist, but it must be derived from the provided information (usually an EXPR generic data type extension will be used to convey the derivation expression.
OTH	other	The actual value is not a member of the set of permitted data values in the constrained value domain of a variable. (e.g., concept not provided by

UNC	un-encoded	The actual value has not yet been encoded within the approved value set for the domain. Example: Original text or a local code has been specified but translation or encoding to the approved value set has not yet occurred.
MSK	masked	There is information on this item available but it has not been provided by the sender due to security, privacy or other reasons. There may be an alternate mechanism for gaining access to this information. Note: using this null flavor does provide information that may be a breach of confidentiality, even though no detail data is provided. Its primary purpose is for those circumstances where it is necessary to inform the receiver that the information does exist without providing any detail.
NA	not applicable	Known to have no proper value (e.g., last menstrual period for a male).
UNK	unknown	A proper value is applicable, but not known.
ASKU	asked but unknown	Information was sought but not found (e.g., patient was asked but didn't know)
NAV	temporarily unavailable	Information is not available at this time but it is expected that it will be available later
NASK	not asked	This information has not been sought (e.g., patient was not asked)

The utility of applying an approach such as is employ by the HL7 nullFlavor code set would allow for future interpretation of permissions gathered and encoded in permissions ontology under new circumstances with contextual information concerning the data gathering event when complete information is not available. Applying this method could allow for a greater number of stored RPO data points, tissue samples or other information to be legally and ethically used in the future under novel or revised circumstances otherwise that were not anticipated by the researcher or the approving IRB at the time of data creation or tissue storage.

5. FUTURE CONSIDERATIONS

There are numerous ethical legal and social issues which could arise in the future that may have bearing on the intent or execution of permissions granted in a current-day context. With great attention paid to the specific working of research documents, it is possible to be quite explicit with regard to the actual nature of the permission being granted and thus stored in a database as information token. However, should the working of the permission statement be ambiguous, either purposefully or as an oversight, it is difficult to predict what a future human being or a computer program may interpret. In the case of tissue banking, it is impossible to know what laboratory or scientific discoveries may arise in the future which may affect the use or re-use of a tissue sample. Further it is conceivable that a genetic array or information contained therein may someday be proxy for a person identifier as a medical record or social security number is today, thus changing the nature of the use of the data created from said tissue sample, and not expressly addressed as a part of the RPO development scope of work.

Current thinking holds to the notion that there might be a permission to use tissue in "genetic testing." However genetic testing is, even in present day terms, is a rather broad and ambiguous concept. Should a research coordinator review all of the various kinds of genetic testing that are presently possible, if this is not one of the objectives of a study? Is the research coordinator qualified or capable of educating a participant in the various kinds of genetic testing that are currently available, sufficient to answer any scientific, legal or ethical questions? What should a program / IRB / study coordinator ethically say about potential tests or procedures that could be performed in the future? How can we reconcile the goal of a comprehensive ontology that can accurately reflect the complexity and ambiguity of these kinds of questions relevant in today's clinical environment, with the seemingly divergent trends of ensuring research participants remain fully informed while clinical research becomes more and more complex with the advancement of medical science.

6. CONCLUSION

In the short term, the RPO, as a crucial component of RPMS, will standardize collection, sharing and retrieval of research permissions across institutions, make permissions and consent assumptions more explicit, and open the door for potential semantic reasoning. Standardizing collection of research permissions with careful consideration for novel and future circumstances will facilitate research by ensuring that patients' intentions for inclusion or exclusion in research projects are being met while ensuring privacy and confidentiality.

7. ACKNOWLEDGMENTS

This publication was made possible by the National Institutes of Health (NIH) Grant Numbers UL1RR029882 and UL1RR024146 from the National Center for Research for Medical Research and RC2LM010796 from the National Library of Medicine. This work is a collaborative effort between several academic institutions: HSSC, Medical University of South Carolina, University of California, Davis, Clemson University and Duke University with industry consultants. Special thanks go to Lisa J Macdonald BS MT (ASCP), Douglas Owan BSCS, Steven E. Reynolds MSCS, from Science Applications International Corporation (SAIC) and Matvey B Palchuk MD MS from Recombinant Data Corp for their valued research and contribution to the ontology. We would also like to acknowledge the contributions of: Laura Beskow, Ph.D., Lawrence (Doc) Muhlbaier, Ph.D. and Kevin Weinfurt, Ph.D. from Duke University, and the support of Nick Anderson, Ph.D. of the University of Washington. The contents in this publication are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. Information on Re-engineering the Clinical Research Enterprise can be obtained from http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp.

8. REFERENCES

[1] Beale, T., Grieve, G. *OpenEHR*. Null Flavours and Boolean data in openEHR. [article on the Internet] http://www.openehr.org/wiki/display/spec/Null+Flavours+and+Boolean+data+in+openEHR.html, [accessed October 3, 2010]

[2] Bechara, A. Damasio, H, Tranel, D., & Damasio, A. R. (1997) Deciding Advantageously Before Knowing the Advantageous Strategy. *Science.* 275 1293 - 1295

[3] The Belmont report: ethical principles and guidelines for the protection of human subjects of research. Bethesda, Md.: National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978. [report on the Internet] http://ohsr.od.nih.gov/guidelines/belmont.html. [accessed October 1, 2010]

[4] Berg, E. A simple Objective Technique for Measuring Flexibility in Thinking (1948) *Journal of General Psychology*, 39 15 – 22

[5] Brand, M. et al. Neuropsycholgical correlates of decision-making in ambiguous and risky situations. *Neural Networks.* 19 (2006) 1266-1276

[6] Centers for Medicare and Medicaid Services, *Evaluation and Management Services Guide* Accessed September 30 p16. [document on the Internet]: http://www.cms.gov/MLNProducts/downloads/eval_mgmt_serv_guide.pdf. [accessed October 3, 2010]

[7] Ceusters, W., Elkin, P. Smith, B. Negative findings in electronic health records and biomedical ontologies: A realist approach. *International Journal of Medical Informatics.* 76s (2007) s326-s333

[8] Clinical and Translational Science Awards (CTSA) program. [article on the Internet] http://www.ctsaweb.org [accessed October 1, 2010]

[9] Code of Federal Regulations, Title 45, Part 46(a), 116. Washington, D. C.: Department of Health and Humans Services. [document on the Internet] http://www.hhs.gov/ohrp/humansubjects/guidance/ 45cfr46.htm [accessed October 1, 2010]

[10] Elkin, PL et al., A controlled trial of automated classification of negation from clinical notes. *BMC Medical Inform. Decision Making* (2005) 5, p13

[11] H. Knublauch, R. W. Fergerson, N. F. Noy, M. A. Musen. The Protégé OWL Plugin: An open development environment for semantic web applications. *Lecture Notes in Computer Science*, 2004, Vol 3298/2004, 229-243; doi: 10.1007/978-3-540-30475-3_17

[12] HL7 Version 3. [article on the Internet] http://www.hl7.org/v3ballot/html/welcome/environment/ [accessed October 1, 2010]

[13] HL7 Version 3 Standard: Core Principles and Properties of Version 3 Models, Release 1, Normative Ballot 4 - September 2010 [specification on the Internet, accessed October 3, 2010]

[14] Hollon, T. Researchers and regulators reflect on first gene therapy death. *Nature Medicine* 6(1), p6 (2000); doi:10.1038/71545

[15] J. D. Halamka. Making Smart Investments In Health Information Technology: Core Principles. *Health Affairs*, 28, no. 2 (2009): w385-w389; doi: 10.1377/hlthaff.28.2.w385

[16] Kuhn, S., Brass, M. The cognitive representation of intending not to act: Evidence for specific non-action-effect binding. *Cognition* (2010) 117, pg9-16

[17] Kuhn, S., Elsner, B., Prinz, W., Brass, M. P. Busy doing nothing: Evidence for nonaction-effect binding. *Psychonomic Bulletin & Review*. (2009) 16 (3) 542-549

[18] Lupyan, G., Dale, R. Language Structure is Partly Determined by Social Structure (2010) *PloSONE*, 5, Vol. 1 e8559

[19] Ngo-Metzger Q, Sorkin DH, Phillips RS, et al. Providing high-quality care for limited English proficient patients: The importance of language concordance and interpreter use. *J Gen Intern Med* 2007. 22(Suppl 2):324-30

[20] N. Sioutos, Sherri de Coronado, M. W. Haberc, F. W. Hartelb, W. Shaiud, L. W. Wright. NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. *Journal of Biomedical Informatics* 40(1), 2007, 30-43; doi:10.1016/j.jbi.2006.02.013

[21] Schenker Y, Wang F, Selig SJ et al. The importance of language barriers on documentation of informed consent with on-site interpreter services. *J Gen Intern Med* 2007. 22 (Suppl 2): 294-9

[22] Smith, B. *HL7 Watch*. Flavors of Null [article on the Internet] http://hl7-watch.blogsppot.com/20089/10/flavors-of-null.html [accessed October 3, 2010]