Bottom-Up Modelling of Permissions to Reuse Residual Clinical Biospecimens and Health Data

by

Elizabeth Umberfield

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nursing)
in the University of Michigan
2020

Doctoral Committee:

Associate Professor Marcelline Harris, Chair
Assistant Professor Yun Jiang
Professor Sharon Kardia
Assistant Professor Andrea Thomer

Elizabeth Umberfield

eliewolf@umich.edu

ORCID iD: 0000-0001-5029-1158

Dedication

I dedicate this dissertation to Scott Umberfield, my husband, encourager, editor extraordinaire, cheerleader, and BEST friend. I couldn't have done this without you. I cannot wait to be your partner through your own grad school journey. I am so glad that we get to do life together. You make the journey worthwhile.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Consent forms serve as evidence of permissions granted by patients for clinical procedures. As the recognized value of biospecimens and health data increases, many clinical consent forms also seek permission from patients or their legally authorized representative to reuse residual clinical biospecimens and health data for secondary purposes, such as research. Such permissions are also granted by the government, which regulates how residual clinical biospecimens may be reused with or without consent.

There is a need for increasingly capable information systems to facilitate discovery, access, and responsible reuse of residual clinical biospecimens and health data in accordance with these permissions. Semantic web technologies, especially ontologies, hold great promise as infrastructure for scalable, semantically interoperable approaches in healthcare and research. While there are many published ontologies for the biomedical domain, there is not yet ontological representation of the permissions relevant for reuse of residual clinical biospecimens and health data.

The Informed Consent Ontology (ICO), originally designed for representing consent in research procedures, may already contain core classes necessary for representing clinical consent processes. However, formal evaluation is needed to make this determination and to extend the ontology to cover the new domain. This dissertation focuses on identifying the necessary information required for facilitating responsible reuse of residual clinical biospecimens and health data, and evaluating its representation within ICO. The questions guiding these studies include:

- ***RQ1)*** *What is the necessary information regarding permissions for facilitating responsible reuse of residual clinical biospecimens and health data?*
- ***RQ2)*** *How well does the Informed Consent Ontology represent the identified information regarding permissions and obligations for reuse of residual clinical biospecimens and health data?*

We performed three sequential studies to answer these questions. First, we conducted a scoping review to identify regulations and norms that bear authority or give guidance over reuse of residual clinical biospecimens and health data in the US, the permissions by which reuse of residual clinical biospecimens and health data may occur, and key issues that must be considered when interpreting these regulations and norms. Second, we developed and tested an annotation scheme to identify permissions within clinical consent forms. Lastly, we used these findings as source data for bottom-up modelling and evaluation of ICO for representation of this new domain. We found considerable overlap in classes already in ICO and those necessary for representing permissions to reuse residual clinical biospecimens and health data. However, we also identified more than fifty classes that should be added to or imported into ICO.

These efforts provide a foundation for comprehensively representing permissions to reuse residual clinical biospecimens and health data. Such representation fills a critical gap for developing applications which safeguard biospecimen resources and enable querying based on their permissions for use. By modeling information about permissions in an ontology, the heterogeneity of these permissions at a range of levels (e.g., federal regulations, consent forms) can be richly represented using entity-relationship links and embedded rules of inference and inheritance. Furthermore, by developing this content in ICO, missing content will be added to the Open Biological and Biomedical Ontology (OBO) Foundry, enabling use alongside other widely

adopted ontologies and providing a valuable resource for biospecimen and information management. These methods may also serve as a model for domain experts to interact with ontology development communities to improve ontologies and address gaps which hinder successful uptake.

Chapter 1 Introduction

1. Background

Patients commonly sign consent forms for a range of diagnostic and therapeutic clinical procedures. Upon signing, these forms become artifacts of administrative and legal significance. Signatures serve as indicators that an informed consent process has occurred between the patients (or their legally authorized representatives) and clinicians, and that patients are willing to participate in and grant permission for the procedures included on the form and discussed with them by their provider.[1,2] The forms become valid as a legal document of permission upon signature by all involved parties.[3]

Clinical procedures often include collection of blood, body fluids, or tissues (i.e., biospecimens) from the patients' bodies. Once the clinical indication for collection is complete, remaining portions of these biospecimens may be discarded. However, the recognized value of residual clinical biospecimens and health data is increasing, largely related to expanding methods for analyzing and aggregating genomic data and associated clinical data. As a result, clinical consent forms may also include statements seeking permission to save residual clinical biospecimens and health data for secondary purposes.[4] These material and information resources have a range of applications for reuse, including genetic and molecular research,[5] commercial development for products such as of pharmaceuticals and medical devices,[6] precision medicine,[7] and population-health research.[8] Progress in these areas depends on accessing massive collections of biospecimens and health data from as many people as possible,

with as much heterogeneity as possible.[9] This suggests the need to maximize accessibility and sharing of biospecimens and their associated data for multiple uses.[10]

However, the need for biospecimens alone does not warrant their sharing or release. Authority over residual clinical biospecimens and permissions granted for them originate from the donors of the specimens, the institutions involved in collecting them, and relevant regulatory bodies, including but not limited to the federal government and overseeing organizations. As an added layer of complexity, regulation of biospecimen reuse varies based on the context of the biospecimens' collection, such as during clinical care or for primary research use.[4,11]

Regardless of whether consent documents were for health care or research, we lack methods to support automated or semi-automated querying of permissions for reuse of biospecimens and data, based on a foundation of contemporary semantic technologies. It is not uncommon for biorepositories to archive hundreds of thousands of specimens,[12] which vary in terms of the consented (i.e., permitted) modes of reuse. Retrieving and reviewing this information about permissions using the actual or scanned image of original consent form is time-intensive and requires case-by-case review by some human actor who must understand the complex interactions of the multiple sources of authority over reuse of these biospecimens. As such, manual review lacks scalability and may be prone to inconsistencies or error. Presently, when this information is represented in biorepositories' information systems, it is "locked in" proprietary and/or non-standard formats.[13] These burdens are further captured in the suggestion that, "methods for handling informed consent decisions at best remain non-interoperable, and at worst are wholly paper-based."[14] eConsent forms do not yet address the need for interoperable queries of permissions within the consent document. And while the Office of the National Coordination for Health Information Technology has called for standards-based

approaches to expressing consent,[15] the *focus* is on documents, not the permissions within the consent documents.

There is a need for increasingly capable information systems to facilitate the capture, discovery, access, and responsible reuse of stored biospecimens and data, whether from research or clinical processes, retaining permission information expressed in signed consent documents. Semantic web technologies, and especially ontologies, hold great promise as infrastructure for scalable, semantically interoperable approaches in healthcare and research.[16] Ontologies are classically defined as common vocabularies to share information in a domain which "includes machine-interpretable definitions of basic concepts in the domain and relations among them."[17] Ontologies thereby enable integration of multiple bodies of data through 'annotating' or mapping their individual fields to the structural elements of a common , shared ontology, as well as interoperability at the level of software interfaces, without manual mapping.

Many ontologies for the biomedical domain are published and in-use.[18] A study to identify the minimum metadata necessary for biobank information systems to share biospecimens data called out the need for ontologies which represent the "ethical standards under which the samples are collected, any restrictions on research use, and access requirements to the samples to expressing the permissions for reuse."[19] Such work has been started through development of the Informed Consent Ontology, which was originally designed for representing consent in research procedures.[20] However, there is not yet ontological representation of the permissions relevant for reuse residual clinical biospecimens and health data, i.e., those which were collected during clinical care processes.

2. Research Questions and Specific Aims

This dissertation focuses on (a) exploring the necessary information regarding permissions to reuse of residual clinical biospecimens and health data, and (b) evaluating the representation of this information within an ontology of informed consent (ICO).

2.1. Research Question 1. *What is the necessary information regarding permissions for facilitating responsible reuse of residual clinical biospecimens and health data?*

Specific Aim 1:

- ***1a:*** *To identify relevant federal regulations and norms (e.g., best practices, guidance documents) that bear authority or give guidance over reuse of residual clinical biospecimens and health data in the US.*
- ***1b:*** *To gather domain experts' interpretations of the permissions by which reuse of residual clinical biospecimens and health data may occur.*
- ***1c:*** *To gather domain experts' interpretations of the key issues and concepts – including the overlap, congruency, and gaps – that must be considered when interpreting regulations and norms for reuse of residual clinical biospecimens and health data.*

We conducted a scoping review to identify sources of regulation or guidance that bear authority on reuse of residual clinical biospecimens and health data. To obtain a broad array of expert perspectives on permissions for residual clinical biospecimen and health data reuse, we searched a range of biomedical, legal, and health ethics and policy databases.

Specific Aim 2: *To develop and test an annotation scheme for identifying permissions within clinical consent forms.* We followed Pustejovsky and colleagues' MAMA (Model-Annotate-Model-Annotate) cycle as the method for developing and refining our annotation

scheme, and evaluated our annotation scheme instructions using a subset of the data and revised the annotation scheme based on our findings.

2.2. Research Question 2. *How well does the Informed Consent Ontology (which was designed for research procedures) represent the identified information from aims 1 and 2 regarding permissions and obligations for sharing and use of specimens and data collected in the context of clinical procedures?*

Specific Aim 3: *To evaluate ICO for representing clinical permissions, focusing on reuse of residual clinical biospecimens and health data.* Evaluation of the ontology was performed using methods for evaluating ontologies from the literature during extension to new domains.[21] This evaluation focuses on domain-specific knowledge representation rather than a computational structure or operational requirements.

3. Significance

In completing these aims, this dissertation identifies necessary content for permissions to reuse residual clinical biospecimens and health data, and systematically evaluates the coverage of this content within an existing formal and semantically interoperable resource. Such representation fills a critical gap for developing applications which safeguard biospecimen resources and enable querying based on their permissions for access and use. By modeling information regarding permissions in an ontology, the tremendous heterogeneity of these permissions at a range of levels (e.g., federal regulations, norms, patients' consent documents) can be richly represented using entity-relationship links and embedded rules of inference and inheritance. Furthermore, by developing this content in the Informed Consent Ontology, missing content will be added to the Open Biological and Biomedical Ontology (OBO) Foundry, a Basic Formal Ontology-based biobanking suite of ontologies, enabling use alongside other widely-

adopted biomedical ontologies (e.g., Data Use Ontology, Information Artifacts Ontology) and serves as a valuable contribution for biobank specimen and information management.

Chapter 2 Exploring Regulations and Norms That Bear Authority Over or Guide Reuse of Residual Clinical Biospecimens and Health Data in the United States

1. Introduction

When patients seek health care services, portions of their blood, body fluid, or tissues may be removed for diagnostic or therapeutic purposes. Portions of these specimens often remain after their clinical indication is fulfilled. These portions, hereafter referred to as *residual clinical biospecimens,* are either discarded or sent to a biorepository.[22] Besides their abundance and relative representativeness of the healthcare-accessing and -seeking population, residual clinical biospecimens are desirable for their quality; they are processed and stored according to stringent standards and analyzed within CLIA-certified labs.[23] Residual clinical biospecimens are also linked or linkable to patients' electronic health records, which are cornucopias of information. These features make residual clinical biospecimens particularly valuable to researchers. Furthermore, developing biospecimen collections through direct-to-biobank donation or research participation is resource intensive and faces challenges for participant recruitment. In response to such challenges, biobanks and researchers often turn to "real world" collections of residual clinical biospecimens.[22]

The United States of America (US) has clearly defined requirements for human subjects' research, which includes reuse of biospecimens collected during research procedures. In addition to analysis in the primary study, investigators may request consent for reuse of biospecimens for broad categories of future research studies, unforeseen at the time of recruitment.[24,25]

However, reuse of residual clinical biospecimens is governed by multiple regulations. Because of the multiple sources of authority over reuse of residual clinical biospecimens, navigating the complex ecosystem of permissions to reuse clinical biospecimens is challenging for biorepositories and those who wish to identify biospecimens that are stored across biorepositories. Our long-term goal is to develop a computable semantic model, or ontology, which can be referenced for a variety of applications, including for example software applications that assist in identifying and making decisions about the permitted reuse of residual specimens and text-based identification of permissions within consent documents.

The purposes of this scoping review were threefold. First, we aimed to identify relevant federal regulations and norms (e.g., best practices, guidance documents) that bear authority or give guidance over reuse of residual clinical biospecimens and health data in the US. Second, we aimed to gather domain experts' interpretations of the permissions by which such reuse may occur. Third, we aimed to gather domain experts' interpretations of the key issues and concepts – including the overlap, congruency, and gaps – that must be considered when interpreting these regulations and norms. The assumption is that the literature serves as a vetted, peer-reviewed source of domain experts' interpretations. We fully recognize and appreciate the importance of the ethical considerations around the issues of reusing residual clinical biospecimens and health data, but this paper does not directly address ethics. Rather, we solely focus on the authority and guidance that the identified regulations and norms bear on such reuse.

2. Methods

2.1. Design

We conducted a scoping review to identify sources of regulation or guidance that bear authority on reuse of residual clinical biospecimens and health data. Scoping reviews enable

exploration and synthesis of a breadth of knowledge through mapping key information or evidence. Their design is well-suited to examining fields that are complex, are emerging or developing, or have not yet been comprehensively reviewed.[26,27]

2.2. Search Strategy

To obtain a broad perspective of permissions for residual clinical biospecimen and health data reuse, a variety of biomedical, legal, health ethics and policy databases were searched, including: PubMed, HeinOnline, NexisUni, Health Policy Reference Center, Philosopher's Index, and Public Affairs Index. Selection of these databases and queries for each database were advised and reviewed by a law librarian and a health informationist. The base set of search terms included synonyms for specimens, sharing, and consent or permission. However, use of these terms in HeinOnline yielded surfeit and extraneous results (n= 3378) that were unrelated to our topic of interest. Under the guidance of the law librarian, we added names of known federal regulations that have influence on biospecimen use to the search terms. Likewise, due to the scarcity of results in Health Policy Reference Center (n= 31), Philosopher's Index (n= 4), and Public Affairs Index (n= 4), the search was broadened by only using the specimen and consent/permission terms (see table A1 for a complete list of search terms for each database). Filters were set to include only reports written in English and published after 1996 (i.e., the year Health Insurance Portability and Accountability Act (HIPAA) was enacted). Filters were set in HeinOnline and NexisUni to only include law reviews and journal articles written on policies within USA jurisdiction. All searches were finalized on May 14, 2019.

2.3. Screening

All retrieved records were imported to a citation management software, and duplicates were removed. Two reviewers, both trained as clinical research scientists, independently

screened titles and abstracts followed by full-text manuscripts. They met three times for each of the screening stages to discuss challenges and uncertainties in study selection and to refine the inclusion and exclusion criteria. In scoping reviews, methods allow for inclusion and exclusion criteria to evolve as retrieved manuscripts are explored.[27] The final inclusion criteria incorporated manuscripts that discussed the conditions by which residual clinical biospecimens and their data could be shared or reused. Figure 1 depicts the entirety of the screening pipeline, from manuscript retrieval to final inclusion, as well as exclusion criteria.

2.4. Management and Extraction

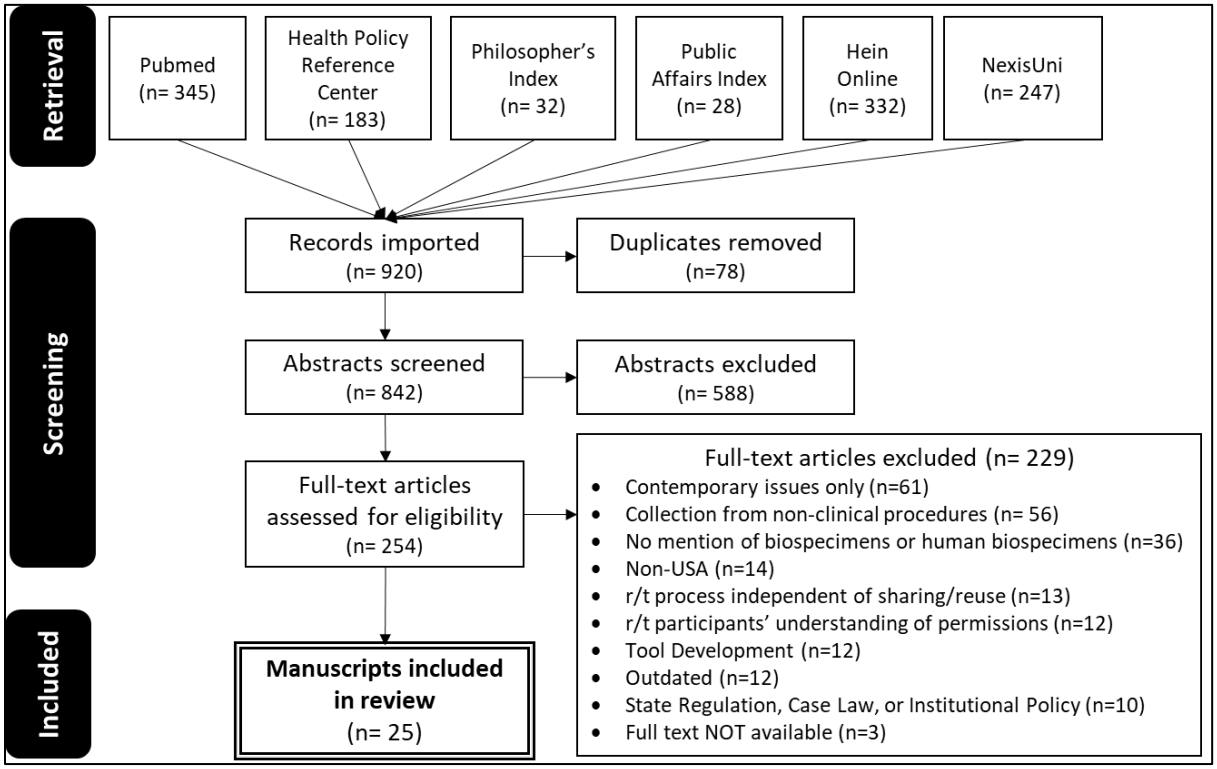
Information was extracted from the manuscripts and organized in spreadsheets. Each row included citation information, the authors' stated purpose of the manuscript, key concepts used to frame discussions of residual clinical biospecimen and health data reuse, and the key issues or assertions made by the authors. For each regulation or norm relevant to residual clinical biospecimen and health data reuse in the US discussed in these manuscripts, a sub record was created that included its name, date passed (and dates of significant revision), purpose or aim, topics used by the authors to frame discussions of that document, all conditions by which residual clinical (or all) biospecimens can be shared or reused, and the authors' evaluation of the regulation or norm.

3. Results

The final database search yielded 842 unique records after duplicates were removed. Titles and abstracts were screened and included if they discussed specimens collected during clinical procedures or biospecimens generally, regulations or norms applicable in the US, and any term related to sharing or reuse. Of note, much of the legal literature did not include abstracts, which explains the high number of manuscripts that qualified for full-text review but

were subsequently excluded. Two hundred fifty-four full-text manuscripts were further screened for inclusion. Twenty-five manuscripts were included in the final review.

Figure 1. PRISMA Flow Diagram for Identifying Regulations or Norms that Bear Authority on Reuse of Residual Clinical Biospecimens and Health Data in the US.



3.1. Identified Regulations and Norms

From the included manuscripts, we identified eleven regulations and fifteen norms (n=26) that bear authority over or guide reuse of residual clinical biospecimens and health data in the United States (Table 1 presents all identified regulations and norms identified in this review). Only one manuscript specifically examined conditions for reuse of residual clinical biospecimens, and its analysis only included HIPAA and reuse in research.[28] While there was little explicit focus on residual clinical biospecimens across the literature, regulations, and norms, these documents did often target concepts (e.g., genetic information, personally identifiable health information, all human biospecimens) that overlapped with and motivated our interests with reuse of residual clinical biospecimens. Identified norms came from both national and international sources, largely professional and nonprofit organizations and multinational consortiums with US participation.

3.2. Permissions for Reuse of Residual Clinical Biospecimens

Identified permissions for reusing residual clinical biospecimens and health data varied widely across the regulations and norms. With an eye towards thematic analysis methods, we identified several themes of permissions across the regulations and norms (see Table 1):

3.2.1. Permissions Conditional on Consent or Authorization.

Half of the identified regulations and norms (n= 13) emphasized the importance of consent or authorization when reusing residual clinical biospecimens and health data for non-clinical purposes. HIPAA and the 21st Century Cures Act permits these biospecimens and health data to be used for future research so long as patient authorization is obtained.[28,29] Likewise, the American Medical Association (AMA) and International Society for Biological and

Environmental Repositories (ISBER) echo these sentiments; they specify that clinicians should obtain consent from patients

Table 1. Federal Regulations and Norms that Bear Authority on Reuse of Residual Clinical Biospecimens and Heath Data

Regulation/Norm (Key Dates)	**Purpose of the Regulation/ Norm**	**Interpretations of Permissions for Reusing Residual Clinical Biospecimens and Health Data from the Literature**
Federal Regulations		
21st Century Cures Act (2016)	To modernize and expediate the research and approval process of the FDA for drugs and medical devices.[29]	The following enact changes to the HIPAA Privacy Rule: • PHI disclosure between covered entities is exempt if for healthcare operations or data aggregation.[29] • Sharing PHI for quality, safety, or effectiveness research of FDA-regulated products is exempt as a "public health activity," even if shared with drug companies and commercial entities.[29] • Authorizations for future research are allowable if the purpose of the research, dates for authorization, and statement that authorization will remain in effect unless revoked are clearly stated.[29]
Affordable Care Act (ACA) (2010; 2014 revised)	To extend health coverage to uninsured Americans.[30]	• Genetic information cannot be used by health insurers to determine pre-existing conditions, eligibility, or rates on premiums.[30,31]
Americans with Disabilities Act (ADA) (1990; 2008 revised)	To protect the civil rights of persons with disabilities.[32]	• Employers, employment agencies, labor organizations, joint labor management committees, and public services cannot use genetic information to discriminate based on disability.[30,32]
Clinical Laboratory Improvement Act (CLIA) (1988)	To regulate labs that test for diagnosis, prevention, treatment, or assessment of human beings.[33]	• Labs must ensure patients' confidentiality when specimens are under the labs' control.[33] • Test results may only be released to authorized persons or the initiator of the request.[33]
Common Rule (1991; 2018 revised)	To protect participants from the risks involved in research.[34]	• Researchers can use residual clinical biospecimens and health data without patients' consent: o If an IRB determines that the data are deidentified.[33,35–39] o If an IRB grants a consent waiver.[33,35,36] o If the data is publicly available.[33] • Investigators must get IRB approval and consent when prospectively intending to use residual clinical biospecimens (or collect extra material) and health data for a specific research study.[35–37]
Food and Drug Administration Amendments Act (FDAAA) (2007)	To regulate research for drugs or medical devices with human subjects, including use of biospecimens and health data.[33,40]	• Biobanks and investigators that are engaged in research with specimens used to investigate drugs or medical devices must be subject to FDA Human Subjects protections, unless: o An exception is met.[33] o An IRB deems that a biospecimen is not individually identifiable.[33] • The FDA may retrospectively retrieve entire medical records or previously stored biospecimens during surveillance studies without individuals' consent or authorization.[40] o However, if new specimens or data are needed prospectively, the FDA must obtain consent.[40]
Genetic Information Nondiscrimination Act (GINA) (2008)	To protect from genetic discrimination by insurers & employers.[41]	*NOTE: Enactment of GINA modified HIPAA, explicitly designating genetic information as PHI.*[30] • Health insurers cannot request, require, or use genetic information to make any underwriting decisions, including determining coverage, eligibility, or premiums.[31–33,41,42] • Employers must not make employment-related decisions based on genetic information.[31,33,41,42]

Regulation/Norm (Key Dates)	**Purpose of the Regulation/ Norm**	**Interpretations of Permissions for Reusing Residual Clinical Biospecimens and Health Data from the Literature**
Health Insurance and Portability and Accountability Act (HIPAA) (1996; 2003 revised; 2013 revised)	To adopt national standards for electronic health information exchange.[28] *The Privacy Rule*: To protect uses and disclosures of individually identifiable health information.[38]	• Covered entities must obtain patients' authorization to disclose PHI, including biospecimens, unless otherwise permitted.[28,30,33,42–44] Authorization must include: 1. Description of PHI to be used or disclosed 2. Who is authorized to disclose the PHI 3. Who will receive or use the PHI 4. Description of purpose of use or disclosure 5. Expiration date or event 6. Statement of right to revoke authorization 7. Statement of exceptions for right to revoke 8. Statement that the covered entity may not "condition treatment, payment, enrollment, or eligibility on the authorization" 9. Statement that further disclosure of the PHI by the recipients may no longer be protected 10. Patients' signatures and dates of signature [28] • When authorization cannot be obtained, o Covered entities may share patient information, including biospecimens, when identifiers are removed or an expert determines the risk of reidentification is very small.[39,41,45] o An IRB issues a waiver of authorization [28,43] for accessing PHI to perform eligibility screening for recruitment, or when PHI is about deceased individuals.[28,33] • PHI, including biospecimens, can be disclosed for treatment, payment, health care operations; about victims of abuse, neglect; for judicial, administrative, law enforcement; disclosures about decedents or for cadaveric tissue donation; averting serious threats to self or others; if required by law; for specialized government functions; for public health surveillance; and workers' compensation.[29,30,33,42,43] • Limited data sets may be released when recipients sign data use agreements.[28,33,39,41,43] • Recipients of limited data sets must require "any agents to whom it provides the limited data set agree to the same restrictions and conditions that apply to the limited data set recipient."[33] • Covered entities may sell PHI for research and when remuneration is limited to the cost of preparation and transmission.[33] • Health insurers cannot use genetic information to determine preexisting conditions, eligibility, or premiums.[30,44]
Newborn Screen Saves Lives Reauthorization Act (2014)	To provide federal funding for education programs and recommendations for newborn screening.[46]	• Research on newborn dried blood spots is human subjects research which means consent and IRB oversight must be obtained.[33]
Rehabilitation Act (1973)	To prevent discrimination based on disability.[32]	• Federal agencies, employers with >$2500 federal contracts, or any federally-funded program or activity cannot discriminate based on disability, including inferences made based on genetic information.[32]
42 U.S. Code § 241. Research and Investigations Generally (2011)	To protect the privacy of research participants.	• Researchers collecting sensitive and identifying data can request a certificate of confidentiality (which authorizes researchers to avoid compelled disclosure) from DHHS.[31]

Regulation/Norm (Key Dates)	**Purpose of the Regulation/ Norm**	**Interpretations of Permissions for Reusing Residual Clinical Biospecimens and Health Data from the Literature**
Norms *(e.g., Best Practices, Guidance Documents)*		
AMA: Commercial Use of Human Tissues (2008)	To provide guidance on physicians using patients' biospecimens for commercial purposes.[47]	• Physicians should obtain consent from patients in order to use their tissues in research.[47] • Biospecimens should not be used for commercial purposes without patients' informed consent.[47] • If patients' biospecimens are used commercially, profit-sharing between patient and doctor should be facilitated under lawful contract.[47]
CIOMS: International Ethical Guidelines for Epidemiological Studies (1982; 1991 revised; 2009 revised)	To provide guidelines to indicate ethical principles when conducting epidemiological research.[48]	Secondary use of identifiable medical records and biospecimens is allowable when all the following requirements are achieved: • Ethics committee concludes the research is "minimal risk," • The rights of patients will not be violated, • Privacy and anonymity are assured, • The research is designed to answer an important question, or • Consent is impracticable.[49]
GA4GH: Framework for Responsible Sharing of Genomic and Health-Related Data (2014)	To provide a framework for harmonized and interoperable sharing of genomic and health related data.[49,50]	• Secondary use of biospecimens and data may occur when the uses are consented for or approved.[49] • When consent to data sharing has been obtained, the rights of citizens to benefit from the advances in medicine and of scientists to be recognized for their contribution should be recognized.[51] • Privacy and security safeguards should be proportionate to the real risks of data breaches.[51]
HUGO: Bioethics Committee 1995	To advocate for privacy protections in the context of genomics.[49]	• Secondary use of biospecimens may occur if general notification is provided, the participant does not object, and the sample is coded or anonymized.[49]
International Declaration on Human Genetic Data (IDHGD) (2003)	To protect confidentiality of medical data, including genetic and proteomic.[49]	• Secondary use of biospecimens and data are allowable when the analysis is in the public interest or when "irretrievably linked."[49]
ISBER: Best Practices for Repositories 2012 (revised)	To guide standardizing methods for collection, long-term storage, retrieval, and distribution of biospecimens to enable their future use.[52]	• If biospecimens are still necessary for clinical care, they must not be used for research. [52] • "Specimens and/or data should only be made available for ethical and scientifically appropriate research that is expected to contribute to scientific discovery."[52] • Donor consent (specific or broad) should be obtained unless waived by an ethics committee.[52] • Biorepositories should protect donor confidentiality and ensure that specimens and associated data are only shared according to the authorization/consent and privacy standards they were obtained under.[52] • Biospecimens and associated data should be deidentified prior to release, unless identifiers are absolutely necessary and an IRB approves transmission of identifiers.[52]

Regulation/Norm (Key Dates)	**Purpose of the Regulation/ Norm**	**Interpretations of Permissions for Reusing Residual Clinical Biospecimens and Health Data from the Literature**
OECD: Guidelines on Human Biobanks and Genetic Research Databases (2009)	To guide management of Human Biobanks and Genetic Research Databases (HBGRD).[53]	• The operators of the HBGRD should strive to make data and materials rapidly and widely available to researchers to advance knowledge and understanding.[50] • Secondary use of biospecimens and genetic data can only occur following ethics board approval.[49]
EEOC: Policy Guidelines (1995)	To protect applicants and employees from discrimination.[54]	• Genetic information cannot be used to discriminate for employment opportunities.[32]
UNESCO: International Declaration on Human Genetic Data (2003)	To address ethical issues related to medicine, life sciences and associated technologies.[55]	• "Researchers should… encourage the free circulation of human genetic data and human proteomic data in order to foster the sharing of scientific knowledge."[50]
UNESCO: Universal Declaration on Bioethics and Human Rights (2005)	To address ethical issues related to medicine, life sciences and associated technologies.[55]	• "Benefits resulting from any scientific research and its applications should be shared with society as a whole and within the international community, in particular with developing countries."[50] • "States should foster international dissemination of scientific information and encourage the free flow and sharing of scientific and technological knowledge."[50]
UNESCO: Universal Declaration on the Human Genome and Human Rights (1997)	To provide ethical standards for human genetic research.[56]	• "States should make every effort… to continue fostering the international dissemination of scientific knowledge concerning the human genome, human diversity, and genetic research and, in that regard, to foster scientific and cultural co-operation, particularly between industrialized and developing countries."[50]
UNESCO: Report of the International Bioethics Committee on Consent (2008)	To guide obtaining informed consent for clinical and research purposes.[57]	• Secondary use of clinical biospecimens and health data is only allowed when an expert deems that re-consent is impracticable.[49] • Participants must have a right to withdraw their specimens.[49]
WMA: Declaration of Helsinki (1964; 2008 revised)	To outline physicians' duties in research.[51]	• Physicians may pursue secondary use on "identifiable human material or data" for research without consent only when re-consent is impossible or impracticable, and only with ethics review committee approval.[49]
WMA: Declaration of Taipei (2002; 2016 revised)	To guide ethical storage and use of identifiable data and biospecimens.[58]	• Biobanks' and health databases' requirements for informed consent for using identifiable biospecimens may be waived, only when there is a "clearly identified, serious, and immediate threat where anonymous data will not suffice."[38]
Key: *AMA: American Medical Association; CIOMS: Council for International Organizations of Medical Sciences; EEOC: Equal Employment Opportunities Commission; GA4GH: Global Alliance for Genomics and Health; HUGO: Human Genome Organization; IDHGD: International Declaration on Human Genetic Data; ISBER: International Society for Biological and Environmental Repositories; NCI: National Cancer Institute; OECD: Organization for Economic Cooperation and Development; UNESCO: United Nations Educational, Scientific, and Cultural Organization; WMA: World Medical Association*		

in order to use their biospecimens in research, and that all future uses must correspond to the conditions specified in the consent.[47,52]

However, there are multiple permissible options for those who wish to reuse residual clinical biospecimens without obtaining authorization or consent. As an example, the Common Rule (CR), Food and Drug Administration Amendments Act (FDAAA), HIPAA, ISBER, and United Nations Educational Scientific and Cultural Organization (UNESCO) permit reuse of residual clinical biospecimens and health data when (a) an Ethics Committee approves the intended uses, (b) an ethics committee waives the authority of another regulation, or (c) the biospecimens and data are appropriately deidentified.[28,33,49,52] Of note, the Council for International Organizations of Medical Sciences (CIOMS), UNESCO, and World Medical Association (WMA) state that such options should only be used when consent is not possible or is "impracticable."[38,48,49] It is also interesting to note that research on newborn dried blood spots is not eligible for such exceptions, and consent and IRB approval must be obtained for this type of residual clinical biospecimens under the Newborn Screen Saves Lives Reauthorization Act.[33]

3.2.2. Permissions Conditional on Identifiability. Identifiability plays a pivotal role in reuse of residual clinical biospecimens and health data. The nature of if and how identifiable these resources determines which regulations or norms apply to their reuse. As an example, permission to use deidentified residual clinical biospecimens and health data for research is exempt from regulation under the US Common Rule and FDAAA, as research using deidentified resources is not considered human subjects research.[33,34] HIPAA similarly applies only to protected health information, i.e., that which is individually identifiable.[41] Rothstein points out that US federal regulations demonstrate a bimodal perspective of

identifiability; residual clinical biospecimens and health data are either identifiable and all of the aforementioned regulations apply, or are not identifiable and therefore are completely unprotected.[39] Multiple norms emphasize the importance of deidentification, including the International Declaration on Human Genetic Data (IDHGD), which permits biospecimen reuse when they are "irretrievably linked" to identifying information.[49]

3.2.3. Permissions Conditional based on Specified Uses. While some permissions within regulations and norms are specific to the context of specimens' and data's *collection* (i.e., whether they were collected during clinical or research procedures), many others instead focused on the downstream *uses* of all biospecimens or health data. Such downstream uses include use for discrimination, conditions for commercialization, and promoting public or individual benefit for specimen contributors. There was consensus among authors of the included manuscripts that it was never permissible, citing multiple regulations and norms, to use genetic information – regardless of context of collection – to make employment- or insurance-related decisions. These regulations and norms include the Affordable Care Act (ACA), Americans with Disabilities Act (ADA), Genetic Information Nondiscrimination Act (GINA), HIPAA, Rehabilitation Act, and the Equal Employment Opportunities Commission's (EEOC) Policy Guidelines.[30–33,41,42,44]

Permissions for commercialization of biospecimen and health data are considerably more varied across regulations and norms. The AMA specifies that physicians are never permitted to release or use their patients' biospecimens for commercial purposes without their consent.[47] However, the 21st Century Cures Act takes a much different stance, that protected health information (PHI; including biospecimens) may be shared with commercial entities for examining the quality, safety, or effectiveness of FDA-regulated products without adhering to

HIPAA's authorization requirements.[29] HIPAA further permits covered entities to sell PHI for research and when there is remuneration, renumeration must be limited to the cost of preparation and transmission.[33]

Lastly, whether using patients' biospecimens and health data for commercial transactions or not-for-profit research, there was consensus across norms that benefit should extend beyond the investigators. UNESCO's International Declaration on Human Genetic Data, Universal Declaration on Bioethics and Human Rights, and Universal Declaration on the Human Genome and Human Rights specify that investigators should freely and widely share biospecimen resources, scientific knowledge, and benefits from discovery, particularly between industrialized and developing countries.[50] The Organisation for Economic Co-operation and Development (OECD) likewise prescribes that biospecimens and health data be made widely available to advance scientific inquiry and discovery; however, IDHGD and ISBER only permit such use when the research is expected to contribute to ethical and scientifically appropriate research or is in the public's interest.[49,52]

3.3. Key Issues.

The authors of the identified manuscripts widely agreed that the most prevalent issue related to reuse of residual clinical biospecimens and health data is that such reuse remains largely unregulated, or the regulations are disjointed and difficult to interpret, across national and international contexts.[30,32,33,41,43,50,52] Harrell and Rothstein argue that, "when viewed in the context of international biobanking it frequently becomes difficult to conclude with any certainty whether international sharing of specimens and health data is lawful."[33] HIPAA and the Common Rule are designed to protect individuals privacy and risk from harm. However, the alternatives to consent provided in these regulations may be more expedient and less

administratively burdensome than tracing consent permissions. As such, “the benefits of [HIPAA’s] Privacy Rule [are] no longer be commensurate with its costs in terms of aggravation, expense, and widespread misunderstandings and misapplications"[33] and “researchers have little reason to avail themselves of the new broad consent option.”[35]

Other major issues when sharing or reusing residual clinical biospecimens and health data include lack of consensus surrounding prioritization of individual’s rights to privacy, confidentiality, and autonomy over materials that were once part of them or information about them, or the public benefit that may be derived from having the most resources possible to advance scientific discovery. These tensions are exemplified by Dove, who stated, “Unduly restricting data sharing across national borders or the matching of biospecimens with health registries and patient records limits the possibility of validating biological findings in larger cohorts… But sharing personally identifiable individual and familial data requires [consideration of] ethical and legal interests.”[50] Major issues and key concepts identified in the literature are presented in Table 2.

4. Discussion

We identified a set of twenty-six regulations and norms that apply to reuse of residual clinical biospecimens and health data in the US through this scoping review. We further identified and described the permissions in these regulations and norms that condition how and when residual clinical biospecimens and health data can be reused, and the major issues when examining these conditions. Inconsistencies in permissions coalesce in a patchwork of rules that make it difficult to tell which regulations or norms apply to a given use, or if the applicable regulations or norms even agree with each other.[59] We are not aware of other reviews that

Table 2. Major Issues and Key Concepts in the Literature for Reuse of Residual Clinical Biospecimens and Heath Data

Citation	Purpose of Manuscript	Major Issues	Key Concepts
Ahn, 2015 [41]	To advocate for a balance between privacy rights and availability of genomic information for research	Genomic data may be linked to other publicly available information. The implications of genomic info are unknown, and balancing data availability and privacy has yet to be regulated.	• Identifiability • Discrimination • Privacy
Beskow, 2016 [37]	To reflect how public opinion should affect biospecimen research policy.	Public input on approaches for shared interests should inform biospecimen research policies. Consent approaches should reflect realistic aspirations and highlight specimen contributors' autonomy.	• Consent • Confidentiality • Trust • Autonomy
Bierer, Barnes, & Lynch, 2017 [34]	To highlight revisions to the Common Rule and their implications for research participants.	The new provisions will likely be "interpreted and implemented in ways that will insulate the regulated community from compliance enforcement… the crucial challenge will be to continue upholding their commitment to improving human health and benefiting the public, while appropriately protecting research participants."	• Human Subjects Protections • Benefit Sharing
Bledsoe, 2017 [38]	To review policy developments around evolving ethical issues for biobanking and provide perspectives on future issues.	Biobanking has experienced dramatic changes in the research environment, regulations and policies, and evolving ethical issues. As technologies evolve, these issues should continue to be debated.	• Identifiability • Transparency • Accountability
Bregman-Eschet, 2012 [60]	To examine how regulations of genetic information should consider privacy and autonomy of individuals.	Genetic databases, despite their highly sensitive nature, are largely unregulated in the United States.	• Privacy • Autonomy • Property rights
Campbell et al., 2012 [52]	To provide guidance on the standardization of methods for collection, long-term storage, retrieval, and distribution of specimens that will enable their future use.	Biorepositories are the intermediary between biospecimen contributors and those who wish to use biospecimens. Biospecimen resources are "currently regulated by an amalgam of differing and occasionally conflicting laws and policies." Repositories must consider discrepant sources of authority when sharing samples and data across geographic borders and jurisdictions.	• Human Subjects Protections • Privacy & Confidentiality • Identifiability • Consent • Benefit Sharing
Dove, 2015 [50]	To describe international regulation of biobanks' privacy protections, areas of consideration for such regulations, and the need for harmonizing regulations within a global framework.	"Unduly restricting data sharing across national borders or the matching of biospecimens with medical registries and patient records limits the possibility of validating biological findings in larger cohorts… But sharing personally identifiable individual and familial data requires [consideration of] ethical and legal interests."	• Promote Sharing • Privacy & Confidentiality • Identifiability • Consent
Edwards, 2008 [47]	To argue that United States' regulations do not adequately balance promotion of future research and patients' autonomy over their biospecimens.	Many patients are unaware that blood or tissue collected from their bodies during clinical care are not thrown out but rather saved. Although hospitals require patients to sign consent forms, these forms are heterogeneous and largely unregulated. "Thus, patients are generally left with little, if any, information about how their samples will be used or who might be using them."	• Scientific Benefit • Autonomy and Rights • Ownership • Informed consent • The right to withdraw

Citation	Purpose of Manuscript	Major Issues	Key Concepts
Evans, 2009 [40]	To review regulations for drug development, including analysis of medical records and "previously stored tissue specimens."	FDAAA allows analysis of health records and residual clinical biospecimens for post-market research. However, it defers to voluntary arrangements to collect such information, which may create barriers that impede discovery.	• Scientific Benefit vs. Individual Harms
Harrell & Rothstein, 2016 [33]	To review regulations of biobanking research and privacy laws in the US and global biobank research.	"American biobanking laws and practices are diffuse, disjointed, and largely indecipherable… When viewed in the context of international biobanking it frequently becomes difficult to conclude with any certainty whether international sharing of specimens and health information is lawful."	• Privacy & Confidentiality • Informed Consent • Security
Hsu, 2010 [30]	To examine US regulations that address privacy and discrimination concerns related to genetic information.	"Current legislation [does not balance] competing interests of privacy and discrimination… The legal framework of genetic privacy and nondiscrimination remains muddled."	• Privacy • Discrimination
Hudson, 2011 [31]	To describe evolving policies pertinent to genetic and genomic research, the integration of genetics into clinical care, and issues raised by genetic technologies.	"First, research using collections of biologic specimens, genomic data, and information from medical records has amplified the long-standing yet unresolved issue about consent for future research that is unanticipated at the time of specimen collection. Second, the push for broad access to research data sets has raised privacy concerns. Third, as researchers seek to share data with colleagues, the issue of whether and how to share research results with study participants remains vexing, particularly in the absence of explicit prior consent from participants."	• Scientific Benefit • Patients and Human Subjects Protections • Consent • Confidentiality • Return of Research Results
Ireni-Saban, 2014 [32]	To compare genetic information policies in the US and UK for privacy, non-discrimination, and benefit sharing.	"Governing genetic data across countries reveals an ongoing absence and inconsistency at the level of regulatory systems."	• Privacy • Discrimination • Benefit Sharing
Lynch, 2017 [36]	To evaluate the Common Rule's options for obtaining consent.	The Common Rule's deidentification or broad consent requirements lack transparency, but societal benefits outweigh personal autonomy risks.	• Consent Models
Lynch & Meyer, 2017 [35]	To describe the shortcomings of the revised Common Rule, particularly those related to broad consent.	Considering the burden of tracking conditions specified when using broad consent, "researchers have little reason to avail themselves of the new broad consent option." Additionally, it remains unclear if biospecimens ought to be treated as inherently identifiable.	• Burden of tracking consent specifications • Autonomy • Identifiability
Maschke, 2005 [43]	To describe the myriad of biobanking regulations internationally and present relevant ethical issues.	"No binding international regulatory framework addresses [use of biospecimens]. Instead, a patchwork of national laws, regulations and ethics advisory body guidelines govern the collection, storage and research use of biological samples."	• Consent • Commercialization • Public Benefit • Property Rights
Rosati, 2005 [28]	To discuss how the HIPAA Privacy Rule allows release of biospecimens and protected health information for research.	The Privacy Rule's requirements for the release of biospecimens and protected health information is complicated and a source of confusion for covered entities and researchers alike.	• Identifiability • Consent vs. Authorization

Citation	Purpose of Manuscript	Major Issues	Key Concepts
Rothstein, 2010 [39]	To consider the effect of using deidentified biospecimens and health information in research without the knowledge, consent, or authorization of the individual on privacy.	In addition to current strategies for deidentifying health records and biological specimens being insufficient to protect privacy and respect autonomy in research, "the deidentification regulations of the Common Rule and the Privacy Rule are inexplicably and unjustifiably inconsistent."	• Privacy • Identifiability • Autonomy • Trust • Commercial Exploitation
Rothstein, 2016 [29]	To describe how the 21st Century Cures Act and Common Rule revisions weaken HIPAA's privacy protection.	"The benefits of the Privacy Rule will no longer be commensurate with its costs in terms of aggravation, expense, and widespread misunderstandings and misapplications."	• Identifiability
Rothstein, Knoppers, & Harrell, 2016 [51]	To highlight ethical issues, regulation, and guidance of international biobanking, and focus on the range of laws, policies, and practices for biobanking and privacy protections in the US.	Agreements, guidelines, contracts, memoranda of understanding, and polices create up a complex context for international biobanking. The US lacks regulations that specifically address biobank research. Development and harmonization of biobanking regulations is needed.	• Privacy & Security • Consent • Identifiability
Shayeb, 2017 [61]	To argue that individuals have property rights over their genetic information.	US regulations do not protect genetic information as the property of individuals. "The question of whether there should also be a universally recognized property interest in genes remains wide open for further discussion."	• Property Rights • Consent • Autonomy • Privacy
Shea, 2003 [44]	To review and evaluate the development of HIPAA, including the intersection of the Privacy Rule and genetic information.	While deidentification must occur for health information to be used for secondary purposes, it is unclear if genetic information is always protected under the Privacy Rule.	• Privacy • Security • Authorization vs. consent
Spector-Bagdady et al., 2019 [62]	To describe considerations for cell line research following revisions to the Common Rule and offer recommendations for regulation of biospecimen research.	Considering risk of reidentification is significantly lower using cell lines compared to other types of biospecimens, "lack of nuance in defining… specific classes of biospecimens [in regulations] may unintentionally impede the rate of research progress and drive up costs associated with specific biospecimens."	• Public benefit • Identifiability • Property Rights • Autonomy • Privacy
Thorogood & Zawati, 2015 [49]	To review international privacy norms governing human biobanks and genomic databases.	Biobanking is guided by a range of overlapping national and International norms, sometimes generated without consideration of biobanks directly. Approaches to privacy differ across cultures and legal traditions.	• Privacy • Consent • Identifiability • Access • Security
Wolf, 2017 [42]	To explain risks and protections for participants in genomic research.	US regulations largely require authorization to disclose genomic and health information, but there are multiple exceptions. Patients may not realize and may object to how their information is being used without their consent.	• Consent • Confidentiality

specifically examine the multiple sources of authority and guidance around permissions to reuse patients' biospecimens and health data.

The requirement to obtain consent or authorization from specimen contributors, both patients and research subjects alike, is widely inconsistent across regulations and norms. Whereas the Common Rule provides rigorous consent and ethics review board requirements for prospective human subjects' research, HIPAA authorization forms may contain much less information and do not require oversight from an ethics review board. Likewise, recent discussions surrounding the Common Rule revisions revealed consensus that blanket consent (i.e., permits any and all future uses) is unethical as it cannot yield truly *informed* consent.[63] However, HIPAA authorization requirements allow for almost universal reuse of protected health information, including biospecimens, so long as a statement is "briefly disclosed in a general consent-to-treat form."[37] In other words, authorization for reuse of residual clinical biospecimens and health data may be lumped into consent forms for health care services, and may not necessarily allow for opting out of some permissions named in the form. There is an argument that it is unreasonable to assume that patients automatically agree to contribute their personal bodily materials and health records for secondary purposes, when they seek health care services.[29] Even if such blanket consent requirements and lack of ethics review board oversight is deemed acceptable, regulations for reuse of residual clinical biospecimens in research differ based on the *type* of biospecimen. For example, the Newborn Screens Saves Lives Act explicitly classifies research using newborn dried blood spots, collected from babies' heels shortly after birth, as human subjects research.[33] Therefore any research using these newborn blood spots requires Common Rule-compliant consent processes at the time of collection and ethics review board oversight. There is little, if any, meaningful difference between dried blood

spots and other types of blood or tissue samples collected during clinical care; these differences between consent requirements are piecemeal and unaligned.

However, it should be noted that, in the case of research using cell lines (i.e., a collection of "potentially immortal cells… that can continue to divide indefinitely"), some argue that different biospecimen types *should* be regulated differently, as some biospecimen types are more identifiable and pose more privacy risks to individuals than others.[62] Authors disagreed about when genetic information, which is almost certainly derivable from human biospecimens, was regulated as PHI. Some authors only considered genetic information as protected when it was linked to any of the eighteen HIPAA-specified identifiers.[33,44] Others considered genetic information as always potentially individually identifiable, and thus should always protected under DHHS requirements.[30] A recent study simulating attacks on databases containing patients' deidentified health records and biospecimens found that as many as 26% of biospecimens could be reidentified.[64] Patients' privacy may continue to be at risk so long as such disagreement and discrepancies in how residual clinical biospecimens and health data are shared remain.

It is a legal and ethical ideal that residual clinical biospecimens and health data are always shared in accordance with applicable regulations and norms, and that patients will have autonomy and agency over if and how their biospecimens and health data are reused. However, there is considerable administrative and logistical burden for tracing and interpreting such decisions over time, physical location, and across jurisdiction.[35,52,65] When it comes to regulation and consent tracing, even opt-out mechanisms place sometimes insurmountable burdens on IRBs and biospecimen brokers in deciding which biospecimen resources cannot be

shared.[66] The responsible conduct of research is challenged given the complexity of navigating these regulations and norms.

While we agree that human interpretation of these permissions is not reasonable on a case-by-case basis, there is growing need for mechanisms to "flag" residual clinical biospecimens and health data as usable or unusable for a range of given activities, including research purposes.[65] To be scalable, knowledge and interpretations of regulations and norms concerning permissions to reuse specimens need to be represented in standardized, computable, and interoperable formats. No such resource presently exists, which is regarded as a major challenge to collaborations in biospecimen research, within the US and particularly internationally.[67,68] Future work should focus on development of such an interoperable resource (e.g., an ontology) to support responsible reuse of residual clinical biospecimens and health data.

We recognize several limitations of this scoping review. We reviewed the biomedical, legal, and health ethics and policy literature to identify and summarize experts' interpretations on how residual clinical biospecimens may be reused in the US. In doing so – rather than identifying and analyzing the regulations and norms directly – it is possible that we may have missed regulations or norms relevant to reusing residual clinical biospecimens and health data in the United States. We likewise may have missed relevant literature by searching within domain-specific databases rather than multidisciplinary databases such as Scopus or JSTOR. Lastly, this review did not include a legal or ethics expert, and it is possible that our interpretation of the authors' discussions may be incomplete or have unrecognized bias.

5. Conclusion

The results of this review demonstrate the patchwork of ill-aligned and occasionally conflicting sources of authority and guidance. Considering such complexity, it is expectedly difficult for those who safeguard residual clinical biospecimens and health data to manage, interpret, and apply information from all applicable regulations and norms. While further clarification, harmonization, and alignment is necessary at both national and international scale, development of a computable resource which references permissions from these regulations and norms could support consistent, expert-informed, and responsible reuse of residual clinical biospecimens and health data.

Chapter 3 Development and Testing of an Annotation Scheme for Identifying Permissions within Clinical Consent Forms

1. Introduction

1.1. Background and Significance.

Informed consent is woven into the fabric of contemporary healthcare. In the context of clinical care, informed consent is the process by which clinicians disclose information to patients or their proxies so they may voluntarily choose whether to grant permission for or refuse a given procedure or treatment, and activities performed either directly or incidentally related to the procedure or treatment.[69] Clinicians are ethically and professionally obligated to obtain consent in advance of clinical procedures; not obtaining consent is legally considered an act of negligence or battery.[70,71] Documentation of the informed consent process must be included in patients' health records, including copies of signed consent forms, whenever possible, as evidence of patients' express permission.[71]

Currently, permissions are not easily discoverable in automated modes, due in part to their many formats. Identification of these permissions is typically only available through processes such as manually retrieving and reviewing informed consent forms.[72] Consent forms remain "locked away" in non-machine readable files (e.g., scanned image files of paper-based consent forms), non-standardized eConsent forms, or as paper copies in physical storage. This means that for each downstream decision which references patient-granted permissions there is typically a manual process to review and interpret patient-granted permissions. While there are information systems that may record patient's permissions, it is still a manual process to read and

interpret permissions and enter a value for the permission status in the information system. Even when the value is entered in the information system, the absence of standard representations of permissions presents challenges to information sharing.

While manual review may be reasonable when searching for permissions from a specific patient, the lack of machine-interpretable representations of consent precludes the development of tools which search and identify such information at scale, i.e. across many consent forms. As an example, when patients sign consent forms during clinical care, these forms may contain additional permissions to reuse their residual clinical biospecimens and health data for secondary purposes. Should an investigator or other entity wish to access these materials, some human broker must access and interpret the consent form, also accounting for the complex environment of regulations and norms at multiple jurisdictions (i.e., institutional, state, federal, international norms), prior to acting on the request to access materials. If machine-interpretable and interoperable representations of permissions existed, tools which reference this information may increase the discovery of and access to biospecimen and health data resources for scientific discovery, while also respecting specimen contributors by observing and heeding how they choose and permit their biological material and to be used.[73]

Before such tools can be developed, a requisite first step is to ensure that humans can reliably identify permissions as they appear in real-world consent forms. Creating a human-readable blueprint (i.e., annotation scheme) for manual discovery of the permissions in consent forms also enables quantitative comparison of the reliability of annotations between annotators.[74] Because of the complexity of the phenomenon and the heterogeneity of permissions as they appear in consent forms, it is important to start with manual annotations to

ensure that the full scope of permissions are discovered, and to create a preliminary dataset for training automated annotation or text processing tools.[75]

1.2. Objective.

The purpose of this study was to develop and test an annotation scheme for identifying permission-sentences within clinical consent forms. This study was a part of a larger investigation focused on modelling permissions to reuse residual clinical biospecimens and health data for secondary, non-clinical purposes. Permissions identified through this study will inform the extension of The Informed Consent Ontology (ICO), a machine-readable, semantic resource for representing consent processes, including permissions granted during such processes. ICO was originally developed to reflect research consent processes, and we propose that ICO may can be extended to provide a common resource for the development of multiple resources that require interoperable representations of permissions (e.g., text processing systems, at-scale information systems that that are aware of consent and permitted actions) across both research and clinical contexts.

2. Methods

2.1. Design.

This was a retrospective analysis of existing clinical consent forms. Three members of the research team were involved throughout the entirety of the data analysis and annotation process: the principle investigator (PI; a nurse scientist), a trained research assistant (RA; provided a healthcare consumer perspective), and a practicing registered nurse (RN; provided a healthcare clinician perspective). In addition, a data scientist with experience in text processing was engaged to support the technical requirements of the annotation effort. Institutional Review

Board review was not required because human subjects and their data were not involved in this analysis; only blank, unsigned consent forms were collected and analyzed.

2.2. Sampling Strategies.

Consent forms were collected using two methods: direct contribution by healthcare facilities and systematic web searching. Direct contribution involved convenience sampling [76] by soliciting voluntary contribution of consent forms from healthcare facilities. With support from leadership within the Michigan Health Information Management Association (MHIMA), a recruitment email was sent to twenty-nine directors of health information management departments who were members of MHIMA. Because we could not assume that this limited sample was representative of health care facilities broadly, we chose to supplement our sample with the following purposive sampling approaches.

First, case sampling – cases that are representative of the *typical* or *average* [76] – was conducted by randomly selecting healthcare facilities from two U.S. government published, open-access lists. Two hundred facilities were randomly selected from a list of all hospitals registered with Medicare [77], and 50 facilities were randomly selected from a list of all ambulatory surgical centers participating in the Ambulatory Surgical Center Quality Reporting Program.[78] These lists included facilities' names, unique identifiers, location information, hospital type (critical access, acute care, children's, or psychiatric), and hospital ownership (e.g., proprietary, government, non-profit, etc.).

Second, criterion sampling [76] was used to discover clinical consent forms from academic medical centers associated with Clinical and Translational Science Award (CTSA) hubs funded for the years 2014-2018.[79] CTSA hubs are research institutions "that work together to improve the translational research process to get more treatments to more patients

more quickly."[79] We selected the criterion of healthcare facilities affiliated with CTSA hubs based on the assumption that, because their mission includes biomedical and health research, their clinical consent forms may be more specific in requesting permission to reuse residual clinical biospecimens and health data for secondary purposes (such as research) than forms from other facilities. Although some CTSA hubs were associated with multiple healthcare facilities, only one facility was selected for each CTSA hub (i.e., a one-to-one match; n= 63), typically the affiliated academic medical center or first facility listed by the CTSA on its website.

2.3. Consent Form Management

2.3.1. Data Collection.

Per directions sent by email to the MHIMA members, consent forms were emailed to the PI directly or through our MHIMA partner as an intermediary. All consent forms were saved as PDFs, some as native files and others as scanned images of hard copies of forms. One facility gave our team permission to download consent forms directly from their internal health information management website, which the PI had access to as an employee of that facility.

For each of the web-identified healthcare facilities (i.e., randomly selected hospitals, randomly selected ambulatory surgical centers, and facilities affiliated with CTSA hubs), a two-fold internet search strategy was used. First, facility names and locations were used to query Google and identify each site's website, if available. Queries for "consent form" were submitted in the websites' internal search tools, and webpages with patient resources were reviewed for consent forms. Second, a combination of facility names, locations, and "consent form" were used to query Google for any publicly available consent forms not yet retrieved through the website searches. Forms were retrieved if they included a request to conduct any clinical care process or procedure (i.e., therapeutic or diagnostic). If the relevance of a form was not immediately clear,

the form was retrieved and later screened. Again, we did not search for consent forms that were developed for specific research projects. This study focuses on consent forms for clinical procedures, and our interest is identification of permissions for research use of residual biospecimens.

2.3.2. Data Management, Transformation, and Curation

We established separate folders for archiving downloaded forms according to each sampling approach. Digital files of all consent forms and corresponding documentation were stored using a Business Associate Agreement (BAA) compliant cloud-based storage service (i.e., Box). Files were only accessible to members of the research team and, for all files that were not publicly available (i.e., facility-contributed), facilities' identifiers (e.g., logos, names) were obscured.

Records containing descriptive information about selected facilities and identified forms were entered into Excel spreadsheets. For all internet searches, the following metadata were generated: if the facility's website was found (yes/no), number of forms discovered, and the curator's field notes. For each discovered consent form, file name, header or title of the form, search terms used, URL, date uploaded (if available), date last modified (if available), date retrieved, and if the document was a consent form template (yes/no) was documented. Facility metadata assigned by Centers for Medicare and Medicaid Services, including unique identifiers, name, location, facility type, and facility ownership, were mapped to and recorded for each form.

2.3.3. Screening for Inclusion

All retrieved consent forms were screened for inclusion based on relevance to this study. Forms were included in the analysis if they documented permission for clinical procedures (e.g., diagnostic, therapeutic, general/admission). Forms were excluded if they were duplicates of other

retrieved forms, were for hospital operations or other non-clinical purposes (e.g., health information exchange, payment, research), were written in languages other than English, or were not human-readable after optical character recognition (OCR) and conversion to .txt file. For large, many-page files that contained multiple forms with varied purposes (e.g., a patient registration or preoperative packet), the screeners selected which pages should be included in the final sample and omitted pages according to the eligibility criteria. Screening occurred independently by the PI and RA, i.e., decisions regarding inclusion vs. exclusion were blind to the other reviewer. Eligibility decisions were discussed by the PI and RA, and disagreements were settled by deliberation the PI's graduate supervisor. See Figure 2 for a summary of data collection and management procedures.

2.4. Annotation Scheme Development and Testing

2.4.1. Annotation Scheme Development.

We followed Pustejovsky and colleagues' MAMA (Model-Annotate-Model-Annotate) cycle as the method for developing and refining our annotation scheme (see Figure 3).[80] That is, after the study team proposed an initial set of instructions for annotations, the annotators applied those instructions using a small subset of the data, the PI and data scientist tested and evaluated the reliability of those annotations across annotators, and the study team revised the annotation scheme based on these findings. In iteratively completing these steps, we refined the specifications within the annotation scheme, increasing the clarity for human-interpretation of the annotation scheme and, ultimately, the quality of their downstream annotations. Our annotation scheme was informed by deontic logic as it informed ICO.[81] Deontic logic is a branch of logic rooted in representations of normative statements (i.e., what *ought* to be) and includes the notions of permission, obligation, and prohibition.[82]

Figure 2. Data Collection and Management Procedures

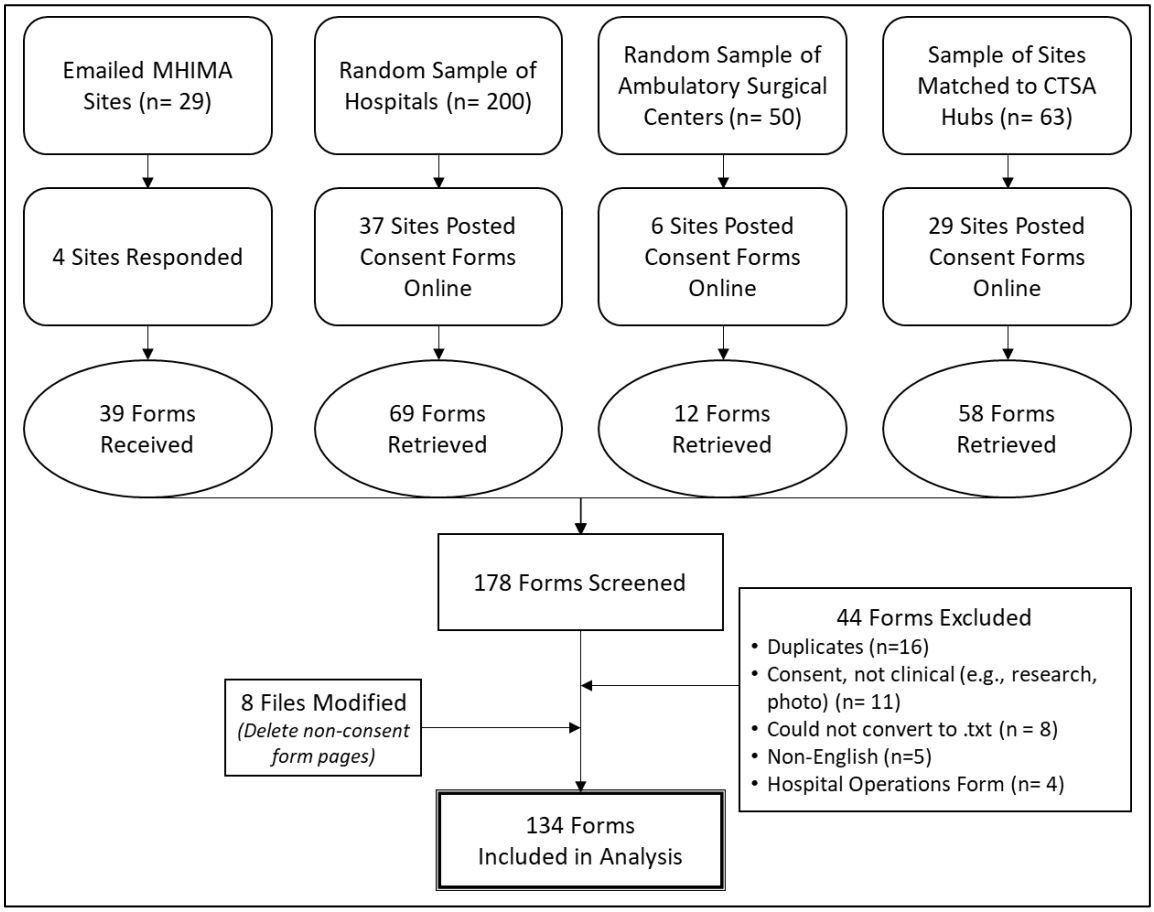
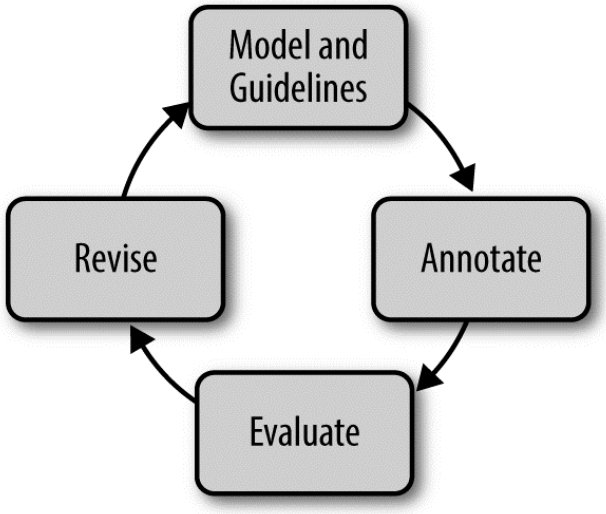


Figure 3. MAMA cycle for annotation scheme development.

Reprinted from "Designing Annotation Schemes: From Theory to Model." Pustejovsky J, Bunt H, Zaenen A. In: Ide N, Pustejovsky J, editors. Handbook of Linguistic Annotation. Dordrecht: Springer Netherlands; 2017. p. 21–72.



To begin, two team members independently reviewed five randomly selected consent forms using the prompt, "Identify any content within the forms that would be essential to understand what was consented to." Whereas the PI marked up large blocks of text including descriptive content, signature lines, and facility information, the RA only marked up what she called "consent language." Consent language was verb phrases indicating the act of consent; such phrases included: agree that, consent to, hereby authorize, request that, etc. When examining the overlap of the PI's and RA's markup, we noticed that the RA's consent language almost always was included in the larger blocks of text marked by the PI. We subsequently decided to limit the annotation task to only those overlapping sentences. We identified these as "permission-sentences" which were direct statements within the consent form that, when the form was signed by the patient or their legally authorized representative, permitted the healthcare facility or its clinicians to do some action or activity.

We iteratively annotated unique sets of five randomly selected consent forms at a time, manually comparing our output after each iteration and adapting the annotation scheme to clarify its specifications. We found that the number of permission-sentences and their structure and content varied by consent documents. After five iterations, in which we met to review annotations and revise the annotation scheme to clarify procedures, the research team agreed that reasonable agreement was achieved based on qualitative evaluation.

The final definition for permission-sentences was, "statement(s) that, upon signature of the consent form, authorizes any new action or activity that may, must, or must not be done." While broad, this definition allowed us to discriminate permission-sentences from those that do not allow or require some new action or activity, such as descriptions of care, statements that reinforce patients' rights, etc. We did not annotate sentences that included agreements or

permissions related to payment or other hospital operations (e.g., “In consideration of the services provided at the Hospital, I agree to pay the Hospital, my physician(s) and other professionals involved in my care for all services, facilities and supplies provided to me.”) The tag *A:Is* was used to annotate sentences that annotators were certain were permission-sentences (i.e., This is a permission-sentence). Due to complex and inconsistent language across consent forms, we also chose to allow for annotator uncertainty in identifying permissions, denoted by the tag *B:MightBe* (i.e., This might be a permission-sentence). Appendix C includes the complete final annotation scheme. For all sentences not tagged by the annotator, we forced automatic encoding of a third tag, *C:Not* (i.e., This is not a permission-sentence). Because the language of permission-sentences is often ambiguous and dependent on context, we allowed annotators to identify a level of uncertainty but did not impose criteria regarding what made them uncertain. This decision was made with the goal of being as inclusive as possible.

2.4.2. Data Preprocessing.

All consent forms were converted from their original file format (most often .pdf files) to text files (.txt) using optical character recognition (OCR) and conversion tools built into Adobe Acrobat DC or MS Word (for .doc). Text files were uploaded to DataTurks, an open-source web-service annotation platform. Three annotators used DataTurks and the developed annotation scheme to markup sentences in consent forms, and their level of certainty that a given sentence was a permission-sentence. (See annotation scheme in Appendix C.)

While the DataTurks software was easy for annotators to learn and use, considerable pre-processing of files exported from DataTurks was required prior to calculating agreement statistics at the sentence level for multiple annotators and labels. We first constructed a pipeline written in Python 3.7 for processing DataTurks’ export files (lists of *.json* objects). We forced

standard character encoding ('ASCII') for all documents and replaced all non- ASCII characters (such as Chinese characters) with spaces (whitespace characters). We removed signature line characters ('_') and redundant white space characters. The result of these preprocessing steps was a data structure that contained (a) the full raw text of the consent forms (as a string), (b) a list of annotations, and (c) metadata about each annotation task. Each annotation in (b) was further broken down into the annotated text string, starting and ending character positions relative to the character counts in the consent form, and the assigned label for each annotator. Parts (a-c) of the data structure were separated by annotator.

We then used spaCy, an open-source software library for natural language processing, to parse the raw text files in order to generate a list of all sentences for each informed consent form. We employed the pre-trained word vectors from the En_core_web_lg model. This semantic model was used to detect sentence boundaries and was trained on OntoNotes Common Crawl.[83] Sentence parsing using the spaCy default parsing logic resulted in 9951 sentences across all forms (n=134). Due to the noise introduced by OCR and heterogenous sentence structures across consent forms (e.g., header phrases with bullets and other symbols rather than full sentences) we excluded parsed text identified by spaCY as a "sentence" if the text lacked *any* English alphabet characters, was less than nine characters long, or was less than three words long. These data cleaning steps resulted in a final set of 6399 sentences, from across the 134 consent forms. These sentences served as reference for the sentences that were marked up by each annotator.

We next developed an approach to align individual annotations by multiple annotators to the spaCy-parsed sentences. First, we aligned all annotations that matched a sentence in the informed consent form exactly. For those annotations that did not have exact matches, we

resorted to fuzzy matching using an implementation of Gestalt pattern matching algorithm.[84] This algorithm finds the longest contiguous matching subsequence, then recursively matches the remaining unmatched elements, in this case words. Using this approach, we were able to align annotations from multiple annotators to a given spaCY-identified sentence in the informed consent document. These steps yielded two data structures: (a) a document-map containing the raw consent forms and their sentence-parsed representations, and (b) a tabular structure where each record (row) was a spaCy-parsed sentence text span and metadata relating to that text span, and three columns reflecting the three annotators, so that the value of each cell in the table corresponded to the assigned label its corresponding annotator.

The ease of use of DataTurks by the annotators was somewhat offset by the considerable amount of data preprocessing necessary to conduct the analysis. While we evaluated other freely available annotation tools, there were limitations to each that would have introduced other types of data cleaning procedures prior to analyzing inter-rater agreement.

2.4.3. Annotation Evaluation.

We evaluated inter-annotator agreement of permission-sentence annotations using multiple methods. The first approach was to calculate raw or observed agreement (denoted as A_o) by summing the count of agreed upon annotations for all tags and dividing by the count of all sentences. Although easy to calculate and easy to understand, output of this measure does not take accidental agreement into consideration. This bias was particularly evident in cases when annotated categories are sparse.[85] We therefore also calculated coefficients which demonstrate the level of inter-annotator agreement beyond what was attributable to chance or arbitrary annotation. Table 3 provides benchmarks for interpreting agreement coefficients.

We first calculated Cohen's κ to provide the proportion of agreement between annotators when treating categories as categorical.[85] However, *A:Is, B:MightBe,* and *C:Not* can also be conceptualized as rank-ordered, ranging from strongly agreeing that a given sentence is a permission sentence (*A:Is*) to strongly disagreeing that a given sentence is a permission-sentence (*C:Not*) with a neutral option allowing for uncertainty in the middle (*B:MightBe)*. As such, we calculat weighted kappas (κ_w) to account for greater disagreement between *A:Is* and *C:Not* than between *A:Is* and *B:MightBe* or *B:MightBe* and *C:Not*. While Cohen's κ treats all disagreements equally; κ_w quantifies the *amount* of agreement between categories.[86] We likewise calculated inter-annotator agreement across all three annotators, using Krippendorff's α for agreement across categories of tags (categorically coded variables) and the amount of difference between categories (ordinally coded variables).[85] All statistical analyses were performed using R for Statistical Computing,[87] particularly the *psych* and *irr* packages.

Lastly, because of the conceptual overlap of the tags (i.e., *A:Is* indicates annotators' certainty for identifying a permission-sentence while *B:MightBe* encodes uncertainty around the same phenomenon of interest), we also chose to evaluate agreement by combining *A:Is* and *B:MightBe* into a single class (tag *A=B)*. Cohen's κ and Krippendorff's α were recalculated using this approach. While acknowledging this approach may introduce false positives, this approach allowed us to be highly inclusive of all potential permission-sentences regardless of certainty of the annotators. This information will be useful in future refinement for automatic annotation.

***Landis & Koch, 1977* [88]**		***Koo & Li, 2017* [89]**	
<0.00	Poor	--	--
0.00-0.20	Slight	--	--
0.21-0.40	Fair	<0.50	Poor
0.41-0.60	Moderate	0.50-0.75	Moderate
0.61-0.80	Substantial	0.75-0.90	Good
0.81-1.00	Almost Perfect	>0.90	Excellent

Table 3 Benchmarks for Interpreting Inter-annotator Agreement Coefficients.

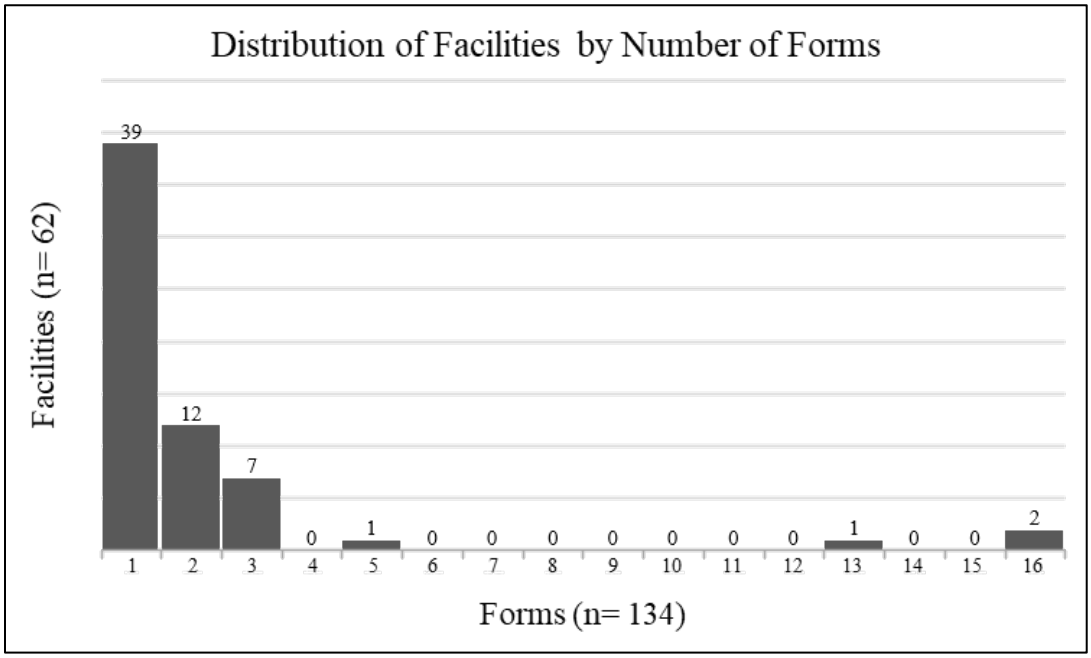
3. Results

The final dataset included 134 clinical consent forms. These forms were retrieved from 62 unique healthcare facilities, which predominately included acute care hospitals (n= 48). Table 4 describes the forms discovered by facility type. Although we only identified one form for most healthcare facilities (n= 39), the average number of consent forms by facility was considerably higher (μ= 4.25). This is largely explained by three outlier facilities which yielded a higher number of forms: Saint Marcy Mercy (n=13), University of Michigan (n=16), and Strong Memorial Hospital (n=16). Forms from Saint Mary Mercy and University of Michigan were collected through contribution from MHIMA facilities (i.e., they were not publicly available), while Strong Memorial Hospital's consent forms were discovered online. Figure 4 depicts the

Table 4. Forms per Facility Types

Facility Type	*Unique Facilities*	*Consent Forms per Facility*					*Total Forms*
		Min	*Max*	*Median*	*Mode*	*Mean*	
Acute Care Hospitals	48	1	16	1	1	4.61	**113**
Children's Hospitals	1	1	1	1	1	1	**1**
Critical Access Hospitals	4	1	2	1.5	2	1.5	**6**
Psychiatric Hospitals	3	1	1	1	1	1	**3**
Ambulatory Surgical Centers	6	1	3	2	2	1.83	**11**
Total Sample	**62**	**1**	**134**	**1**	**1**	**4.25**	**134**

Figure 4. Distribution of facilities by number of forms included in the analysis.



right-skewed distribution of number of identified forms by facility. A complete list of facilities, their location, ownership, sampling strategy used to identify forms, and number of forms included in the final sample is provided in Appendix B.

After preparing the data, the final corpus included 6399 sentences from across the 134 forms. Consent forms varied widely in length, ranging from 4-432 sentences with an average of 47.75 sentences per form. Of these 6399 sentences, 739 (11.5%) were annotated as possible permission-sentences (i.e., *A:Is* or *B:MightBe*) by at least one annotator; 5660 (88.5%) were not annotated by any annotator and therefore was assigned the label of *C:Not* by default. All three annotators identified 211 permission-sentences (28.6% of possible permission-sentences, 3.3% of total corpus) using *A:Is,* and at least two of the annotators identified 351 sentences (47.5% of possible permission-sentences, 5.5% of total corpus) using this label. Table 5 depicts the number and proportion of permission-sentences as the threshold for agreement was relaxed. As expected, the count of permission-sentences diminishes as the threshold for agreement increases. Note that the number of sentences annotated as *A:Is* (n= 635) and *B:MightBe* (n=139) do not add to 739. This was explained by one annotator marking a given sentence with *A:*Is and another annotator marking the same sentence as *B:MightBe.* Sentences could be counted multiple times on this table.

Table 6 compares inter-annotator agreement measures, including percent observed agreement (A_o), Cohen's κ and κ_w, Krippendorff's α, and their associated confidence intervals. Observed agreement (A_o) was greater than 93% for all combinations of annotators (pairs and between all three). Agreement coefficients are presented for all combinations of annotators. The highest pairwise agreement was achieved between PI and RN (κ = 0.6165; κ_w = 0.6546), followed by PI and RA (κ= 0.5606; κ_w= 0.6040, then RA and RN (κ= 0.5466; κ_w= 0.5796).

Krippendorff's α was 0.5737 when treating labels as categorical, and 0.5988 when treating labels as ordinal. These values reflect moderate agreement to substantial agreement.[88]

As previously discussed, we also evaluated agreement while treating *A:Is* and *B:MightBe* as equivalent (henceforth *A=B*). Although annotators varied in their certainty when assigning labels, we observed an overall increase in agreement when combing *A:Is* and *B:MightBe* labels (α= 0.5927) compared to analysis of *A:Is*, *B:MightBe,* and *C:Not* categorically (α= 0.5737), but not higher than analysis of the three label ordinally (α= 0.5988). Tables 7 and 8 describe the annotations and inter-annotator agreement when using *A=B.*

Table 5. Identification of permission-sentences based on number of annotators. Percentages indicate the proportion of identified sentences to possible permission-sentences (n= 739) and the entire corpus (n=6399).

	Agreement by 3 annotators	Agreement by ≥2 annotators	Annotation by ≥1 annotator
A:Is: This is a permission-sentence	211 *211/739 (28.6%)* *211/6399 (3.3%)*	351 *351/739 (47.5%)* *351/6399 (5.5%)*	635 *635/739 (85.9%)* *635/6399 (9.9%)*
B:MightBe: This might be a permission-sentence	0 *0/739 (0%)* *0/6399 (0%)*	5 *5/739 (0.7%)* *5/6399 (0.0%)*	139 *139/739 (18.8%)* *139/6399 (2.1%)*
C:Not: This is not a permission-sentence	5660 -- *5660/6399 (88.5%)*	6028 -- *6028/6399 (94.8%)*	6168 -- *6168/6399 (96.4%)*
			Total Sentences 6399

Table 6. Agreement measures among subgroups of and all annotators.

Note: Agreement is reported using all labels (A:Is, B:MightBe, and C:Not). Ao is observed agreement, Ae is expected agreement or agreement by chance, and κ is the Kappa coefficient used to measure inter-annotator agreement, taking both Ao and Ae into account. Cohen's κ is used to measure inter-annotator agreement between pairs of annotators, and Krippendorff's α is used to measure inter-annotator agreement across all three annotators.

	A:Is, B:MightBe, C:Not (n= 6399)				
		Categorical		*Ordinal*	
	A_o	Cohen's κ	95% CI	κ_w	95% CI
PI-RA	0.9444	0.5606	0.5212-0.6001	0.6040	0.4152-0.7928
RA-RN	0.9373	0.5466	0.5078-0.5854	0.5796	0.3722-0.7869
RN-PI	0.9509	0.6165	0.5787-0.6544	0.6546	0.4586-0.8506
	$\overline{A_o}$	Krippendorff's α	95% CI	Krippendorff's α	95% CI
PI-RA-RN	0.9442	0.5737	0.5425-0.6028	0.5988	0.5656-0.6309

Table 7. Identification of permission-sentences based on number of annotators. Percentages indicate the proportion of identified sentences to possible permission-sentences (n= 739) and the entire corpus (n=6399).

	Agreement by 3 annotators	Agreement by ≥2 annotators	Annotation by ≥1 annotator
	231	371	739
A=B: This [is/might be] a permission-sentence	*231/739* (31.3%) *231/6399 (3.6%)*	*371/739 (50.2%)* *371/6399 (5.8%)*	*739/739 (100%)* *739/6399 (11.5%)*
	5660	6028	6168
C:Not: This is not a permission-sentence	-- *5660/6399 (88.5%)*	-- *6028/6399 (94.8%)*	-- *6168/6399 (96.4%)*
			Total Sentences *6399*

Table 8. Agreement measures among subgroups of and all annotators.

Note: Agreement is reported based on treating A:Is and B:MightBe as equivalent (A=B, C:Not). Ao is observed agreement, Ae is expected agreement or agreement by chance, and κ is the Kappa coefficient used to measure inter-annotator agreement, taking both Ao and Ae into account. Cohen's κ is used to measure inter-annotator agreement between pairs of annotators, and Krippendorff's α is used to measure inter-annotator agreement across all three annotators.

	A=B, C:Not (n=6399)		
	A_o	Cohen's κ	95% CI
PI-RA	0.9483	0.5881	0.5479-0.6282
RA-RN	0.9389	0.5550	0.5157-0.5943
RN-PI	0.9541	0.6385	0.6003-0.6767
	$\overline{A_o}$	Krippendorff's α	95% CI
PI-RA-RN	0.9471	0.5927	0.5589-0.6243

4. Discussion

The purpose of this study was to develop and test an annotation scheme to identify permission-sentences within clinical consent forms. Consent forms include essential information about what activities were permitted by the signers. However, this information is often inaccessible to computable information systems. One way to approach the extraction of this information from consent forms is through text processing. While the presented results may not yet be sufficient for information retrieval tasks, this work provides a foundation for discovering potential permission-sentences in new samples of consent forms at a larger scale. Automated classification models for discovery of potential permission sentences may yield data necessary for further developing and iteratively revising the annotation scheme downstream. This is the first study we are aware of that has attempted to identify permission-sentences within clinical consent forms.

The developed annotation scheme and process yielded moderate to substantial reliability in providing guidance on identifying permission-sentences. Overall, these results should be interpreted cautiously and acknowledgement given that disagreement or uncertainty around identifying permission-sentences may be the result ambiguously written consent forms. Agreement was highest when evaluating labels, *A:Is, B:MightBe,* and *C:Not* using a weighted Kappa statistic. This is explained by the label, *B:MightBe* – which encodes the annotators' uncertainty about assigning either of the two labels – is more similar to the other two classes than *A:Is* and *C:Not* are to each other. We also tested the combination of classes (i.e., *A=B* vs. *C:Not*) with improved results compared to the unweighted kappa, but these results were not superior to the weighted kappa. By using the weighted kappa, we reduce the quantitative penalty for annotators disagreeing about which particular instances they were uncertain about classifying

using an *A:Is* or *C:Not* label. Despite use of *B:MightBe* by all annotators, very few sentences were classified with this label by any two annotators (n= 5) and none by all three annotators. Artstein reminds us that when labels share similarity or overlap, "it may make sense to test reliability at multiple levels at once, since working at multiple levels reflects the process the annotators go through when making their choices."[85] By providing multiple analyses, we demonstrate the complexity of identifying permission-sentences within consent forms.

When examining pairwise agreement between annotators, agreement was highest between the PI and RN, and lowest between the RA and RN. However, the difference in agreement between pairs of annotators was minimal. Slightly higher agreement between the PI and RN may be explained by both having health care clinician backgrounds, i.e. they may have a different view of health care consent processes than the RA, who is representative of the health care consumer perspective. It is arguable that "since these [documents] are written by [those with health care perspective], it is reasonable to assume that in this case, the [health care provider's] interpretation is a better reflection of the writer's intention."[85] Whereas reliability of annotations may benefit from only using clinicians in the future, this does not bode well for the nature of the consent forms themselves – which should be ethically understandable by lay readers and health care consumers.[90]

We found that it was difficult to obtain clinical consent forms. Few facilities responded to a direct request for consent forms (13.8%, n= 4), and only 18.5% (n= 58) of randomly and purposively sampled facilities made their consent forms publicly available on the web. There are known reasons for this; Retrievability of consent forms may often be blocked for legal, risk management, or proprietary protections.[91] As such, those consent forms that are published may reflect biased selectivity or projection of the values and attributes of the healthcare facilities

that publish their forms. It warrants questioning that so few clinical consent forms are made publicly available when the basis of consent in healthcare is rooted in beneficence, transparency, and trust.[69,92] Furthermore, in research contexts, open access to consent forms is a clearly articulated goal.[93,94]

As an incidental finding, we observed that permission-sentences which specified reuse of residual clinical biospecimens and health data predominately permitted these activities without restriction or description of the future reuses, which merits further investigation. During recent debates about informed consent within human subjects research, consent to all future research without limitations or conditions was deemed unethical as it could not possibly be considered "informed."[37,95,96] As such, the 2018 revisions to the Common Rule instead prescribe inclusion of specifications, categories, and purposes of future research, now referred to as "broad consent."[11] Although standards for clinical consent specify that complete information about diagnosis, plan of care, alternatives, and risks be enumerated, there is no broadly enforceable consent requirements to protect patients whose residual clinical biospecimens or health data may be used for downstream purposes.[92] In our sample of consent forms, when permission to reuse residual clinical biospecimens or health data were specified, this permission was linked to all other permissions in the form, without a traceable method that allowed patients to consent to some permissions but not all. As currently practiced, informed consent "does not offer the granularity that is required for data donors to exert meaningful control."[10] In addition, some authors have identified that patients prefer that they be able to permit or forbid individual actions or activities, and not have multiple permissions be aggregated in an "all-or-nothing" consent form.[7]

This project lays an important foundation for automated discovery of potential permission-sentences within clinical consent forms. The next step, machine-interpretation of identified permissions, further requires a semantically rich knowledge base (i.e., ontology) and mapping of consent information to that knowledge base. Identified permission-sentences from this study will serve as input to testing and extending the Informed Consent Ontology for representation of clinical consent processes. The internal structure of permission-sentences must also be further deconstructed and modeled so that information about permissions can be mapped to ICO classes and enable downstream queries and inferencing based on permissions. While there is important work underway to standardize consent statements going forward (e.g., GA4GH and HL7) those efforts are grounded in research consent forms. They may however be useful to clinical consent forms. Even should such clauses become standardized, there will remain a need to support more automated processes to manage permissions in existing consent forms.

Other future work should include refining a final list of permission-sentences which will serve as training data for automated text processing of clinical consent forms to identify permission-sentences, and later classify these sentences according to their content.[97] Automatic annotation, which largely employs rule-based methods or training a model to identify a phenomenon of interest, requires extensive programing or training of the tagging system.[75] Model training requires machine learning from set of reliably-detected instances of a phenomenon so that future instances can be detected from new data sources.[98] While our results demonstrate moderate to substantial inter-annotator agreement, enhancement of the set of identified permission-sentences is yet needed before automated text processing can be performed reliably.[97]

Our findings point to the complexity of identifying permission-sentences within the forms. While it is known that incomprehensibility of clinical consent forms is ubiquitous,[99] our findings may also point to such complexity and obfuscation within the forms that even two clinicians (PI and RN) had difficulty identifying permission-sentences within the sample of forms. Future research should examine the understandability permissions in consent forms to determine if signers, clinicians, and form authors can reliably agree regarding which sentences permit some new action or activity, and later which actions or activities are being permitted. Further work might also include standardization of clinical consent forms and their content, with emphasis on increasing their understandability by patients and clinicians alike.

This study has several limitations. First, it is unknown whether the sample is broadly representative of clinical consent forms, or if bias was unintentionally introduced to the sample through collection of publicly available consent forms. Access to certain types of documents are often blocked for legal or proprietary protections, and those that are published may suggest "biased selectivity" or projection of the values or attributes of the units that publish the documents.[91] It was also not possible to determine if the consent forms retrieved from the web were the most current version in use by the institution since most were not dated. Another limitation was that additional levels of error variance were not accounted for when calculating inter-annotator agreement because we did not nest permission-sentences by form or form by facility. It is possible that particular forms or facilities used language that was either highly agreed or disagreed upon by the annotators, at levels different from the annotations overall. Lastly, annotations by the developer of the annotation scheme and trainer of the RA (i.e., PI) may have introduced bias to the findings. Inter-annotator agreement is calculated under the assumption that annotators and their classifications are independent of one another.[88] This

limitation was addressed in-part by the PI not instructing the RA on how to interpret the annotation scheme, but rather iteratively revising it and pointing the RA back to the scheme during each annotation round. We also addressed this limitation by adding a third, untrained annotator (i.e., RN) whose classification of permission-sentences were, other than sharing an annotation guide, independent of the other annotators.

5. Conclusion

We developed and tested an annotation scheme for identifying permission-sentences within clinical consent forms that performed with moderate to substantial reliability among three annotators. The permission-sentences identified through this work will serve as the input for future modeling the structure of the internal structure and content of permission-sentences, and be iteratively enhanced to serve as a gold standard training set for natural language processing and automated-extraction of permission-sentences. Extending and leveraging the Informed Consent Ontology will contribute to laying a foundation for machine-interpretation of permission information in clinical and clinical research information systems. It will also contribute to scientific discovery by increasing the accessibility of materials and information resources that are permitted to be shared and protect patients by honoring their agency and decision-making around complex issues.

Chapter 4 Evaluating the Informed Consent Ontology for Coverage of Permissions to Reuse Residual Clinical Biospecimens and Health Data

1. Introduction

1.1. Background and Significance

Informed consent is a foundational requirement in both clinical care and research studies. The documentation of informed consent on forms serves as one source of evidence that an informed consent process occurred, and that the consenter has received enough information to make an informed decision regarding the permissions they are being asked to either grant or deny.[100] As the demand for vast and heterogeneous collections of human biospecimens increases, so does the frequency with which individuals are being asked if their biospecimens can be used for secondary purposes.[101] While considerable efforts are underway for tracing such permissions for biospecimens collected during research studies, significantly less attention has been given been given to clinical consent forms and permissions to reuse residual clinical biospecimens. These residual clinical biospecimens are those portions of biospecimens that remain after their clinical indications are fulfilled, and are recognized as an increasingly valuable resource.

There is a need for increasingly capable information systems to facilitate the discovery, access, and responsible sharing and use of stored biospecimens and data, and to facilitate data integration and knowledge discovery within a contemporary connected research environment. This vision requires development of technology which supports expectations around FAIR

(Findable, Accessible, Interoperable, Reusable) Principles, including metadata that supports discovery of biospecimens and data according to their permitted and restricted uses.[102] Semantic web technologies such as ontologies hold great promise as an infrastructure solution for scalable, interoperable approaches in healthcare and research.[16] Ontologies enable integration of multiple bodies of data through 'annotating' their individual fields to the structural elements of a common ontology, which map knowledge in terms of entities and relationships in a domain. Many ontologies for the biomedical domain are published and in-use.[18,103]

The Basic Formal Ontology (BFO) is realism-based, upper level ontology informed by the perspective that reality is independent of human experience, and the real world is comprised of objective entities.[104] BFO's structure is composed of objects (i.e., continuants), processes (i.e., occurrents), and their relationships with time.[105] This common structure is the basis for enforcing logical rules across all ontologies which import or refer to BFO. As an upper level or reference ontology, BFO does not contain any physical, chemical, biological or other specialist terms.[106] Rather, the Open Biomedical Ontologies (OBO) Foundry is a family of ontologies aligned with a set of shared principles for ontology development. One core principle is use of a shared logical structure (i.e., BFO) of the ontologies as the source for common classes (i.e., the maximal collection of all particulars or instances of that thing) and object properties (i.e., the relationships which link classes). Other principles include open access and use, community-based collaborative development, and non-overlapping and strictly-scoped content. The OBO Foundry enforces these design principles to achieve semantic interoperability, a strong and accurate logical basis, and scientific accuracy.[18,103]

While multiple OBO Foundry ontologies are useful for representing information relevant to biomedical science and research, there is one ontology that is particularly relevant and may be

fit for use when representing relevant permissions in the context of clinical care – the Informed Consent Ontology (ICO). ICO was designed to represent documents and processes relevant to consent, specifically for biomedical and health research.[18,107] As currently published, the ontology contains class representations of processes such as signing an informed consent form and IRB (or other regulating body) approval of the consent form, as well as identifiers for the investigator and participant roles.[20] Development of ICO followed OBO Foundry principles and extends from the top-level Basic Formal Ontology (BFO), enabling its use a reference ontology (i.e., represents universals and relations) in this domain.[18,107,108] ICO, however, was not developed to represent the nuanced information relevant to consent in health care, including permissions to reuse residual clinical biospecimens and health data for secondary purposes. While some or all of the existing classes may be transferable to this new domain, ICO must be formally evaluated for reuse in this domain and extended or revised as necessary.

This paper addresses the need to evaluate ICO for use in clinical consents. However, there are well-recognized challenges in ontology evaluation: namely that, despite more than a decade of publication on the issue, there continues to be no accepted standard methodology for ontology evaluation.[109,110] The scope of potential criteria by which ontologies are evaluated include "its coverage of a particular domain and the richness, complexity, and granularity of that coverage; the specific use cases, scenarios, requirements, applications, and data sources it was developed to address; and formal properties such as the consistency and completeness of the ontology and the representation language in which it is modeled."[111] Overall, formal evaluation is underutilized in ontology development, leading to the release of poor ontologies and ultimately hindering "the successful deployment of ontologies as a technology."[21]

Currently, systematic evaluation of ontologies require that evaluators assemble methods from across various evaluation schema.

1.2. Purpose

The purpose of this study is to apply a hybrid model of ontology evaluation to ICO, with a specific focus of evaluating ICO for its completeness expressing permissions from regulations and clinical consent forms with a focus on reuse of residual clinical biospecimens and health data, and developing recommendations for extension and revision.

2. Methods

This study follows a formative evaluation design, in which we examine an information resource (ICO) under development.[112] We adapted Friedman and Wyatt's multistep evaluation process to iteratively examine and check our findings, integrating key steps from the ontology evaluation literature. Figure 5 depicts our multistep evaluation process. Our analysis was also guided by evaluation methods and questions abstracted from the literature, including the National Institute of Standards and Technology's (NIST) Ontology Summit's guidance for evaluating ontologies across ontology life cycles.[21] Figure 6 presents the phases of the Ontology Life Cycle Model. Importantly, we solicited stakeholder involvement throughout the evaluation process, primarily ICO developers and the OBO foundry community.

Figure 5. Multistep, Iterative Ontology Evaluation Process.

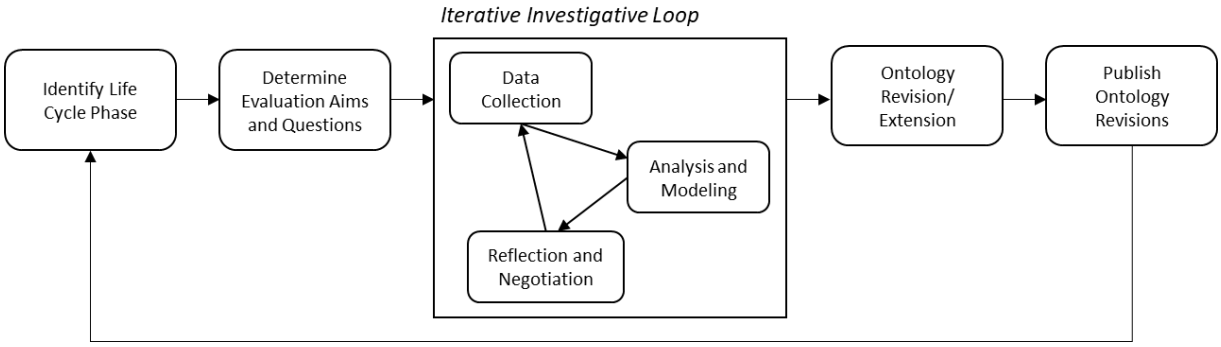
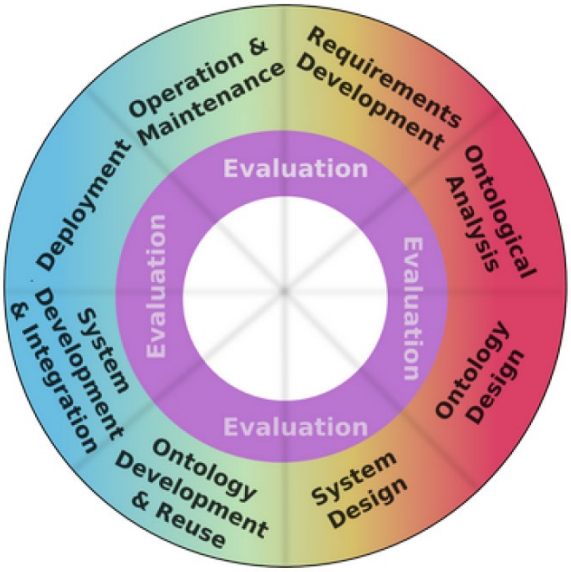


Figure 6. Ontology Life Cycle Model.

Retrieved from Neuhaus, F., Ray, S., & Sriram, R. D. (2014). Toward ontology evaluation across the life cycle (NIST IR 8008). National Institute of Standards and Technology.



2.1. Identify Life Cycle Phase

Because our purpose was to evaluate any modifications to ICO necessary for representing permissions in clinical consent regarding residual clinical biospecimen and health data reuse, we entered ICO's life cycle and this evaluation at the Ontology Development & Reuse Phase. The Ontology Development & Reuse Phase includes a focus on ontology adaptation to fit a new domain. Four primary evaluation tasks are recommended at this phase: informal modeling, formalization of competency questions, formal modeling, and operational adaptation.[21] Because this evaluation focuses on completeness (i.e., if the new domain is appropriately covered in the ontology),[113] we limit our analysis and reporting to informal modeling of content identified from relevant regulations and clinical consent forms.

2.2. Determine Evaluation Aims and Questions

Consistent with both evaluation literature and the available ontology evaluation literature, evaluation often largely depend on the contribution and feedback of domain experts.[21,110] Accordingly, we developed evaluation aims and questions together with a community of ICO developers and OBO Foundry stakeholders. The questions guiding this evaluation were:

A) Does ICO contain the necessary classes and relationships to represent permissions granted through US federal regulations to reuse residual clinical biospecimens and health data?

B) Does ICO contain the necessary classes and relationships to represent permissions named in clinical informed consent forms, both for reuse of residual clinical biospecimens and health data and other clinical procedures?

2.3. Iterative Investigative Loop

The Iterative Investigative Loop is comprised of three steps: data collection, analysis and modeling, and reflection and negotiation. These steps are done both iteratively and in tandem with one another, enabling continuous revisiting of the data, revision of the models, and collaboration with stakeholders.

2.3.1. Data Collection

Data collected and analyzed for this evaluation study included permissions to reuse residual clinical biospecimens and health data, abstracted from two sources. First, permissions to reuse residual clinical biospecimens provided through US Federal regulations were identified through a review of biomedical, legal, and health policy literature. By identifying these permissions through the literature, we attempt to reduce introducing our own biases into the information resource by instead extracting the interpretations of what is legally permissible by experts in their respective fields. The included regulations were: the 21st Century Cures Act, Affordable Care Act (ACA), Americans with Disabilities Act, Clinical Laboratory Improvement Act (CLIA), Common Rule, Food and Drug Administration Amendments Act (FDAAA), Genetic Information Nondiscrimination Act (GINA), Health Insurance and Portability and Accountability Act (HIPAA), Newborn Screen Saves Lives Reauthorization Act, Rehabilitation Act, and 42 U.S. Code § 241 Research and Investigations Generally. While the literature review additionally collected information about norms (e.g. best practices and guidance documents), we chose to limit the present evaluation to US Federal regulations because of the legal authority and weight they bear. Norms are important for guiding practice, but do not bear legal authority.

Second, permission-sentences identified through an annotation study of clinical consent forms. Permission-sentences were eligible for inclusion if they were positively annotated as a

permission-sentences by three annotators of the consent forms. Of this collection of 211 permission-sentences, a sample of 15 were included in this evaluation. Six permission-sentences were purposively selected because they expressly permitted sharing or reuse of residual clinical biospecimens or health data, i.e., the primary phenomenon of interest in this study. Four sentences were selected to for the heterogeneity of the activities they permitted (e.g. videotaping, surgery, anesthesia). Five additional sentences were randomly selected to account for and reduce potential bias introduced through purposive sampling. All permission-sentences included in this analysis are listed in Table 9.

2.3.2 Analysis and Modeling

The goal of informal modeling is to identify all relevant ontological entities (including classes and relationships), the entities' important attributes, and appropriate terminology for the new domain. These results are then modeled in an informal way, such as through concept maps.[21]

We started by looking at each of the identified permissions from the source data, both from US federal regulations and consent forms, and identifying all entities and relationships between those entities within each permission. We then referenced key parent classes modeled in ICO – including information content entity, material entity, process, and role – and began sorting the identified entities into these categories without defining them or their hierarchies. At the same time, we began graphically modeling the loosely defined classes from the consent forms and their relationships using a mind-mapping software. For regulations, we focused just on identifying classes rather than graphically modeling their relationships, as there is not yet a legal ontology in the OBO foundry, and we recognize that the complexity of regulations warrants intensive involvement by a legal domain expert. We carefully examined ICO and OBO foundry

Table 9. List of Permission-Sentences from Clinical Consent Forms which were included in this Evaluation.

Note: By agreement with recruited facilities, we obscured facility names whose consent forms were not publicly available using XXX. For all facilities whose consent forms were publicly available, we did not remove identifiers.

Purposively Sampled Permission-Sentences
Reuse of Residual Clinical Biospecimens and/or Health Data
1. I hereby authorize XXX to retain, preserve and use for scientific or teaching purposes, or to dispose at its discretion or convenience, any specimen or tissues taken from my body during my visit.
2. I am a New York state resident and I give permission for GeneDx to retain any remaining sample longer than 60 days after completion of testing and use my de-identified data for scientific and medical research purposes.
3. I DONATE and authorize XXX to own, use, retain, preserve, manipulate, analyze, or dispose of any excess tissues, specimens, or parts of organs that are removed from my body during the procedures described above and are not necessary for my diagnosis or treatment.
4. I agree that any excess tissue, fluids or specimens removed from my body during my outpatient visit or hospital stay (my specimens) that would otherwise be disposed of by the Hospital may be used for such educational purposes and research, including research on the genetic materials (DNA).
5. I authorize the pathologist, at his or her discretion, to retain, preserve, use, or dispose of any tissues, organs, bones, bodily fluid or medical devices that may be removed during the operation(s) or procedure(s).
6. I hereby consent to the use and disclosure of my protected health information as described in the Notice of Privacy Practices.
Other Clinical Procedures and Activities
7. By signing this form, I am requesting and giving my consent for MHSM and the doctors and/or nurses to give me blood and/o blood products during this admission or series of treatments.
8. Your signature below indicates that you understand to your satisfaction the information about the genetic testing ordered by your health care provider and that you consent to having this testing performed.
9. I consent to the Facility videotaping, photographing, video monitoring, or taking other recordings of me or parts of my body for diagnosis, treatment, research, or patient safety purposes.
10. I also consent to diagnostic studies, tests, anesthesia, x-ray examinations and any other treatment or courses of treatment relating to the diagnosis or procedure described herein.
Randomly Sampled Permission-Sentences
11. I, , request and consent to the start or induction of my labor by my provider: and other assistants as may be selected by him/her.
12. I voluntarily consent to receive medical and health care services that may include diagnostic procedures, examination, and treatment.
13. I consent to the use of closed-circuit television, taking of photographs (including videos), and the preparation of drawings and similar illustrative graphic material for scientific purposes providing my identity is not revealed.
14. I hereby consent to engaging in virtual health/telemedicine services, where available, as part of my treatment.
15. In the event a healthcare worker is exposed to my blood or body fluids in connection with my procedure, or during my hospital stay, I agree to the collection and testing of my blood for HIV.

ontology classes in terms of their labels, textual definitions, and formal definitions for fidelity to their use in the context of US Federal regulations and clinical consent forms. These steps were performed both in tandem and iteratively until a final model for each individual permission from the source data was developed and vetted by multiple team members and collaborators. The resulting graphic models serve as representations of proposed design patterns for ICO.

2.3.3. Reflection and Negotiation

As previously mentioned, methods for evaluation of ontologies is not mature or standardized. As such, we abstracted evaluation questions from the literature to guide reflection on the verification (i.e., did we build it right?) and validation (i.e., did we build the right thing?) of the proposed ICO revisions. These reflection questions were used to guide negotiation with regular meetings with ICO team members and biweekly calls with members of the Ontology for Biobanking (OBIB) community. The list of reflective evaluation questions includes:

Verification

1. Are the proposed revisions adherent to OBO Foundry principles?[114]
2. Are the proposed and revised classes consistent across the ontology's hierarchy?[115]
3. Are proposed classes non-redundant?[115]
4. Does the model capture only entities within the specified scope of the ontology? [21]
5. Is the documentation sufficiently unambiguous to enable a consistent use of the terminology? [21]

Validation

6. Does the ontology contain the necessary and sufficient info (identified through the source data) to make it fit for our particular purpose? Are all entities within the scope of the ontology captured?[21,115]
7. Do the domain experts agree with the ontological analysis? [21]

3. Results

3.1. Analysis and Modeling

Prior to this evaluation and its proposed revisions, ICO was comprised of 893 classes and 95 object properties. Table 10 provides summary counts of all classes used in the informal modeling process, including those ICO classes that were transferable to this new domain or required revision, and those that should either be imported or added to extend ICO or the OBO suite of ontologies for representing permissions from clinical care, including those to reuse residual clinical biospecimens and health data. Informal modeling of permission-sentences abstracted from US federal regulations and clinical consent forms used fifty-two classes and eleven object properties already within ICO, demonstrating the appropriateness of extending ICO for the clinical domain rather than developing a new ontology. Six of these classes from ICO were used in the informal models but were flagged for recommended revision. These proposed revisions were always regarding either the classes' formal definitions (i.e., position within the hierarchy) or human-readable definitions, which may have been ambiguous or overly specific with regard to use in models of permissions. Table 11 presents all terms used to express permissions from the source data that were already present in ICO.

Evaluation also revealed that considerable extension of ICO is necessary to represent all identified permissions from the source data. Thirty classes were identified from other OBO

Foundry ontologies that may/should be imported into ICO. Additionally, we recommend nine new classes to either be added to ICO or another OBO Foundry ontology (and imported into ICO) in order to express all content within the included source data. Tables 12 presents all terms used to express permissions from the source data that were not present in ICO but were identified in another OBO Foundry ontology; we recommend importing these classes. Table 13 presents proposed or new terms necessary for expressing permissions from the source data but were not identified in ICO or another OBO Foundry Ontology. Appendix E contains fifteen tables, reporting the respective classes and relationships used for each of the fifteen permission-sentences from clinical consent forms, and mapping classes to sentence content.

Early in the informal modeling process, a design pattern for the context of permission-sentences emerged. Figure 7 demonstrates this design pattern, demonstrating how the individual who is granting permission ('homo sapiens' (NCBITaxon:9606); 'consenter role' (ICO:0000086)) participates in a process of consenting ('informed consent process' (OBI:0000810)) by using a consent form which contains a permission-sentence ('permission directive' (ICO:0000244)) that prescribes some process ('planned process' (OBI:0000011)). It should be noted that this is the most simplified version of this design pattern, and there is significant heterogeneity and added complexity as each permission-sentence was modeled. As an example, the real-world person (i.e., instance) who is the consenter may also have a range of other important roles including being the patient themselves ('patient role' (OBI:0000093)) or the patient's legally authorized representative ('legal guardian role'(OMRSE:00000038)). Likewise, the processes that are prescribed by permission-sentences also varied widely, but most often included a 'health care process' (OGMS:0000096)) like surgery or blood product

administration, a ‘specimen collection process’ (OBI:0000659), or an ‘act of data sharing’ (ICO:0000228).

As the permission-sentences became more complex, so too did their informal models. Figure 8 demonstrates this complexity, and is based on the following permission-sentence: “…I give permission for GeneDx to retain any remaining sample longer than 60 days after completion of testing and use my de-identified data for scientific and medical research purposes.” In this example, not only is some process prescribed, but also specifications on how the data that emerged from that process may be used (‘data use limitation’(DUO:0000001)) and the timeline by which these processes may occur (‘temporal directive’(proposed class)). Additionally, the flow of a given specimen becoming a residual clinical biospecimen which bears an ‘excess material role’(ICO:0000313) was fleshed out to ensure that all necessary classes were modeled.

Table 10. Summary Counts and Sources for Terms Used in Informal Modelling

	Classes	*Relationships*
Present in ICO	52	11
Recommend Import from OBO Foundry Ontologies	30	--
New Classes	9	--

Table 11. Terms Present in the Informed Consent Ontology – including PURLs, Labels, Definitions, and Proposed Revisions and Comments – Used to Express Permissions Identified in the Source Data.

*Note: *** in the Ontology Label column denotes that the given term was identified through the review of US Federal Regulations via the literature. All other terms were first identified from permission-sentences from clinical consent forms.*

Ontology PURL	Ontology Label	Ontology Definition	Recommendations and Comments
Source Ontologies: Classes			
BFO:0000040	Material Entity	An independent continuant that is spatially extended whose identity is independent of that of other entities and can be maintained through time.	
DUO:0000001	Data Use Limitation	A data item that is used to indicate consent permissions for datasets and/or materials, and relates to the purposes for which datasets and/or material might be removed, stored or used.	
DUO:0000038	genetic research	Biomedical research concerning genetics (i.e., the study of genes, genetic variations and heredity).	
DUO:0000039	Drug Development Research***	Biomedical research concerning drug development.	
IAO:0000027	Data Item***	a data item is an information content entity that is intended to be a truthful statement about something (modulo, e.g., measurement precision or other systematic errors) and is constructed/acquired by a method which reliably tends to produce (approximately) truthful statements.	
IAO:0000030	Information Content Entity	A generically dependent continuant that is about some thing.	
IAO:0021006	Obligee Role***	A role that is either the specified output of an obligation generating social act or the concretization of a transferable obligation and that is realized by its bearer being the providing part of a process that fulfills the previously agreed upon requirements.	
IAO:0021300	Obligor Role***	A role that is either the specified output of an obligation generating social act or the concretization of a transferable obligation and that is realized by its bearer being the receiving part of a process that fulfills the previously agreed upon requirements.	
ICO:0000001	Informed consent form	A document that explains all relevant information to assist a human being in understanding the expectations and requirements of participation in a process, and is an instrument in obtaining consent and, after having obtained consent, is a record that such a consent has occurred.	

ICO:0000019	Signature Section	A document part that is specified as a place to receive a signature.	
ICO:0000053	Drug Testing Investigation	An investigation that targets drug testing.	
ICO:0000060	Act of storing a specimen	A planned process that involves placing a specimen in some location in order to maintain possession of it.	
ICO:0000074	authorization for release of confidential health information	An act of authorizing that is signed by a patient, or his legal representative, or clinical study participant for the use or disclosure of oral, written, or electronic form of confidential health information that identifies the individual and relates to the medical history, diagnosis, treatment, or prognosis of his condition.	This is the only child of 'act of authorizing' (ICO:0000046). Consider deleting the parent class.
ICO:0000079	Institutional Review Board***	A specially constituted organization comprised of medical, scientific and non-scientific members established and designated by an entity to ensure the protection of the rights, safety and well-being of human subjects recruited to participate in biomedical or behavioral research according to the requirements outlined in Title 38, part 16 (same as Title 45, part 46 and Title 21, part 56) of the U.S. Code of Federal Regulations. IRB responsibility include but not limited to the reviewing, approving, and providing continuing review of trial protocol and amendments and of the methods and material to be used in obtaining and documenting informed consent of the trial. Other equivalent committees with the same or similar functions are also considered to be IRBs.	
ICO:0000081	Legal Guardian Role***	A role that inheres in an individual who is authorized under applicable State or local law to consent on behalf of a child or incapable person to general medical care including participation in clinical research.	
ICO:0000083	Individually Identifiable Health Information Entity***	A protected health information entity which identifies an individual; or with respect to which there is a reasonable basis to believe the information can be used to identify an individual.	
ICO:0000086	Consenter Role	A role that inheres in a patient or the legal guardian when he/she participates the consenting process.	
ICO:0000101	Health Information Entity***	An information content entity which is created or received by a health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse;	

		and relates to the past, present, or future physical or mental health or condition of any individual, the provision of health care to an individual, or the past, present, or future payment for the provision of health care to an individual.	
ICO:0000103	Protected Health Information Entity	A health information entity which is created, received, stored, or transmitted by covered entities and their business associates in relation to the provision of healthcare, healthcare operations and payment for healthcare services.	
ICO:0000113	Anonymized information content entity	An information content entity that has been processed to prevent the identification of the person with whom the data are associated, thereby enduring in a stasis of anonymization.	
ICO:0000116	Act of Using Participant Data	A planned process that involves using data which was an output of a human being's participation in a study and which is about that human being.	Either need to make new class for patient information or abstract this concept ('Act of Using Data' or 'Act of Using Human Data').
ICO:0000181	Secondary Use***	A planned process that involves a human biological specimen originally collected for the purpose of treatment or a different investigation.	
ICO:0000186	Permission Temporal Region	A temporal region during which some act is permitted.	
ICO:0000196	Act of Informed Consenting	An act of permitting that is a process part of some informed consent process.	Reconcile hierarchies; why is “act of permitting” and “act of informed consenting” a deontic declaration, whereas “act of authorizing” is not?
ICO:0000199	Permission Role	A deontic role that inheres in an agent and which permits certain actions.	
ICO:0000218	Act of Removing Identifiable Information	An act of anonymization that alters an information content entity by removing information content which identifies that individual.	
ICO:0000222	Signature	An information content entity that is a handwritten depiction of a name.	
ICO:0000228	Act of Data Sharing	a planned process in which data possessed by one person or organization is shared with one or more other persons or organizations.	
ICO:0000233	Regulatory document	A document that either prescribes actions or prescribes the permissions, obligations, restrictions, or protections relating to public or legal policy.	

ICO:0000244	Permission Directive	A directive information entity that prescribes a deontic role that permits some action.	
ICO:0000245	Conditional Permission Directive	A conditional specification that prescribes a permission, should some trigger condition obtain.	Change text definition to “occur” (delete “obtain”).
ICO:0000248	Obligation Directive***	A directive information entity that prescribes an obligation that is realized a deontic role that is the output of a document act.	
ICO:0000268	Confidentiality Directive***	A directive information entity that prescribes the nondisclosure of information.	
ICO:0000290	Waiver Document***	A document designed for the purpose of assisting in an act of waiving, and which contains both some description of waiver and some waiver directive.	
ICO:0000307	Waiver Directive***	A directive information entity that prescribes an act of waiving.	
ICO:0000311	Act of Genetic Testing	A planned process that involves the investigation of some sequence of DNA.	
ICO:0000312	Genetic testing directive	A directive information entity that prescribes some act of genetic testing.	
ICO:0000313	Excess material role	A role inhering in some material entity that is realized in an act of artifact modification.	Unclear human readable definition. Must also include what about the act of modification results in the material entity being 'excess'
LABO:0000107	Laboratory Test	A measurement assay that has as input a specimen derived from an organism and that aims as having as output a data item that is about an entity related to the specimen.	
NCBITaxon:9606	Homo Sapiens	The bipedal primate mammal, Homo sapiens; belonging to man or mankind; pertaining to man or to the race of man; use of man as experimental subject or unit of analysis in research	
OBI:0000011	Planned process	A processual entity that realizes a plan which is the concretization of a plan specification.	
OBI:0000066	Investigation	a planned process that consists of parts: planning, study design execution, documentation and which produce conclusion(s).	
OBI:0000093	Patient role	A role which inheres in a person and is realized by the process of being under the care of a physician or health care provider	
OBI:0000112	Specimen Role	A role borne by a material entity that is gained during a specimen collection	

		process and that can be realized by use of the specimen in an investigation.	
OBI:0000202	Investigation Agent Role***	A role borne by an entity and that is realized in a process that is part of an investigation in which an objective is achieved. These processes include, among others: planning, overseeing, funding, reviewing.	
OBI:0000245	Organization	An entity that can bear roles, has members, and has a set of organization rules. Members of organizations are either organizations themselves or individual people. Members can bear specific organization member roles that are determined in the organization rules. The organization rules also determine how decisions are made on behalf of the organization by the organization members.	
OBI:0000659	Specimen Collection Process	A planned process with the objective of collecting a specimen.	
OBI:0001769	Specimen collector role	An Investigation agent role borne by a person or organization which is realized in a specimen collection process.	
OBI:0100051	Specimen	A material entity that has the specimen role.	
OGMS:0000018	Laboratory Finding	A representation of a quality of a specimen that is the output of a laboratory test and that can support an inference to an assertion about some quality of the patient.	
OGMS:0000090	Treatment	A planned process whose completion is hypothesized by a health care provider to eliminate, prevent, or alleviate the signs and symptoms of a disorder or pathological process	Incorrect hierarchy. Treatment should be child of 'Health Care Process' (OGMS:0000096).
OGMS:0000090	Documenting	A planned process in which a document is created or added to by including the specified input in it.	
Source Ontologies: Relations			
BFO:0000051	has part	*(No definition in Ontology Lookup Service or Ontobee)*	
CommonCore Ontologies/ prescribes	prescribes	*(No definition in Ontology Lookup Service or Ontobee)*	
DUO:0000010	Is restricted to	*(No definition in Ontology Lookup Service or Ontobee)*	
IAO:0000136	Is about	is:about is a (currently) primitive relation that relates an information artifact to an entity.	
OBI:0000295	is:specified:input:of	*(No definition in Ontology Lookup Service or Ontobee)*	

OBI:0000299	has:specified:output	*(No definition in Ontology Lookup Service or Ontobee)*	
OBI:0000293	has:specified:input	*(No definition in Ontology Lookup Service or Ontobee)*	
OBI:0000312	is:specified:output:of	*(No definition in Ontology Lookup Service or Ontobee)*	
RO:0000053	bearer of	*(No definition in Ontology Lookup Service or Ontobee)*	
RO:0000056	Participates in	*(No definition in Ontology Lookup Service or Ontobee)*	
RO:0002350	Member of	Inverse of has member	

Table 12. Terms from OBO Foundry Ontologies (and not in ICO) – including PURLs, Labels, Definitions, and Proposed Revisions and Comments – Used to Express Permissions Identified in the Source Data.

*Note: *** in the Ontology Label column denotes that the given term was identified through the review of US Federal Regulations via the literature. All other terms were first identified from permission-sentences from clinical consent forms.*

Ontology PURL	Ontology Label	Ontology Definition	Recommendations and Comments
Source Ontologies: Classes			
CHMO:0002832	Material Disposal	A planned process in which materials for an experiment are removed permanently from the laboratory.	Delete "for an experiment" as clinical biospecimens can similarly be disposed of.
ERO:0000333	Video	Data containing moving pictures stored in digital and analog formats.	
FBbi:00000223	Graphic Illustration	*(No definition in Ontology Lookup Service or Ontobee)*	
GEO_000000400	Government Organization***	An organization that governs the people living in a particular geographical region or aggregate of geographical regions. The geographical region it governs can change over time (such as the westward expansion of the United States and the addition of Hawaii). Note: this definition was taken over from "geopolitical organization".	
IAO:0000185	Photograph	A photograph is created by projecting an image onto a photosensitive surface such as a chemically treated plate or film, CCD receptor, etc.	
NCIT_C142640	Postmarketing Surveillance***	Programs to identify adverse events that did not appear during the drug approval process.	
NCIT:C13235	Fetus	An unborn or unhatched vertebrate in the later stages of development showing the main recognizable features of the mature being.	Consider importing all age group and at-risk population classes.
NCIT:C15380	Telemedicine	The use of telecommunications technology to provide, enhance, or expedite health care services, as by accessing off-site databases, linking clinics or physicians' offices to central hospitals, or transmitting x-rays or other diagnostic images for examination at another site.	
NCIT:C163409	Safety Monitoring	Review of safety data to ensure safety of the individuals who are participating in the study, or to identify potential safety concerns for the duration of the study lifecycle.	
NCIT:C16423	Child	An age group comprised of individuals who are not yet an adult. The specific cut-off age will vary by purpose.	Consider importing all age group and at-risk population classes.
NCIT:C43960	Health Care Organization	An organization that provides healthcare services or that is involved in the provision of health care activities. Groupings or	If fitting into OBO Foundry hierarchy, should be child of OBI's 'organization'

		subdivisions of an organization, such as departments, may also be considered as organizations where there is a need to identify them.	(OBI:0000245) and parent of 'hospital' (OBI:0000844)
NCIT:C64262	Biologic Sample Preservation Procedure	Procedures utilized to save organic substances from decay. Some preservation procedures are meant to maintain cells, tissues, or organisms in a viable state.	
NCIT:C95018	Use	To put into action or service.	
OBI_0001305	Genotype Information***	a genetic characteristics information that is about the genetic material of an organism and minimally includes information about the genetic background and can in addition contain information about specific alleles, genetic modifications, etc.	
OBI:0000145	Pathologist Role	a worker role of being responsible for making the histopathology diagnoses associated with data from a study; this activity occurs outside the study timeline	
OBI:0000207	Health Care Provider Role	a worker role of providing medical care either within or outside the study timeline	Revise to exclude requirement of within a study timeline
OBI:0600003	Performing A Clinical Assessment	A protocol application during which a series of tests are made of a patient leading to determination of disease state, or condition.	
OBIB_0000039	Blood Specimen on Blood Spot Card***	A blood specimen that is located on a blood spot card.	
OGMS:0000096	Health Care Process	A planned process with the objective to improve the health status of a patient that directly involves the treatment, diagnosis, or prevention of disease or injury of a patient	Import 'health care process' and revise hierarchy within ICO. Treatment should be a child of Health Care Process.
OGMS:0000104	Diagnostic Process	A health care process that involves the interpretation of a clinical picture from a given patient (input) and the assertion to the effect that the patient has a disease, disorder, or syndrome of a certain type, or none of these (output).	
OMIABIS_0000010	Biobank Organization***	An organization bearing legal personality that owns or administrates a biobank.	
OMRSE_00000087	Employer Role***	a role in human social processes that is realized when the bearer provides a wage or salary in exchange for some labor or services as specified by some declaration	
OMRSE_00000093	Insurance Company***	An organization that secures a group of people against pecuniary loss by payment of a sum of money if a specified event occurs.	

OMRSE_00000198	Information Content Entity-Request Process***	a communication in which some participant requests of some other participant an information content entity about some portion of reality	
OMRSE:00000054	Hospital Role	A healthcare provider role that inheres in an organization and is realized by providing inpatient and outpatient care.	
OMRSE:00000056	Hospital Organization	An organization that is the bearer of a hospital role.	
OMRSE:00002029	Education Process	A planned process with an active participant who acquires mental representations of information content entities (ICEs), which had no previous mental representation in the cognitive system, and through repeated use or application of these ICEs becomes the bearer of a new instance of some type of capability, and the participant was not previously the bearer of that instance of that type of capability.	Consider SDGIO definition. While logically accurate, the OMRSE definition is not easily understandable to non-BFO users.
SWO:3000079	Audio Format	*(No definition in Ontology Lookup Service or Ontobee)*	
VO:0000299	Company	*(No definition in Ontology Lookup Service or Ontobee)*	
XCO:0000527	Genetic Manipulation	A condition in which the genotype or the gene expression of an organism or a cell has been modified.	Unknown if there are other types of specimen manipulation. Further domain expertise should be sought to determine if this class is sufficient or if an additional class is needed.

Table 13. Proposed/New Terms Necessary to Express Permissions Identified in the Source Data.

*Note: *** in the Ontology Label column denotes that the given term was identified through the review of US Federal Regulations via the literature. All other terms were first identified from permission-sentences from clinical consent forms.*

Proposed Label	Proposed Definition	Comments and Citations
Classes		
Act of Selling***	To give up (property) to another for something of value (such as money).'	Definition of 'sell' from Merriam-Webster.[116]
Covered Entity Role***	A role based on the Health Insurance Portability and Accountability Act that inheres in individuals, organizations, and agencies that are also health care providers, health plans, or health care clearinghouse.	Definition adapted from HIPAA.[117]
Data Recipient Role***	A natural or legal person, public authority, agency or another body, to which the data are disclosed.	Definition from adapted from GDPR Article 4(9).[118]
Limited Data Set***	Protected health information that excludes HIPAA-specified identifiers of the individual or of relatives, employers, or household members of the individual.	Definition adapted from HIPAA.[117]
Owner Role	A role in a human social process that is that is based on a social act and whose bearer exercises exclusive control over a property, where this control is permitted by one or more deontic roles, which are parts of the ownership role.	After deliberation with members of the Document Acts Ontology, this term was added under PURL IAO:0020027 by Dr. Mathias Brochhausen
Owner Role Directive	A directive information entity that prescribes the owner role.	
Ownership Process	The process of exercising exclusive control over a property.	
Prospective Data Collection***	A study design in which the documentation of the presence or absence of an exposure of interest is documented at a time period preceding the onset of the condition being studied.	Definition from Sage Research Methods.[119]
Residual Clinical Biospecimen	A material entity, collected during some health care process, that bears both a specimen role and an excess material role following some act of artifact modification.	Consider also including mention that these specimens have fulfilled their clinical purpose or indication, and that they would otherwise be discarded.

Figure 7. Context for a Permission Directive (Permission-Sentence)

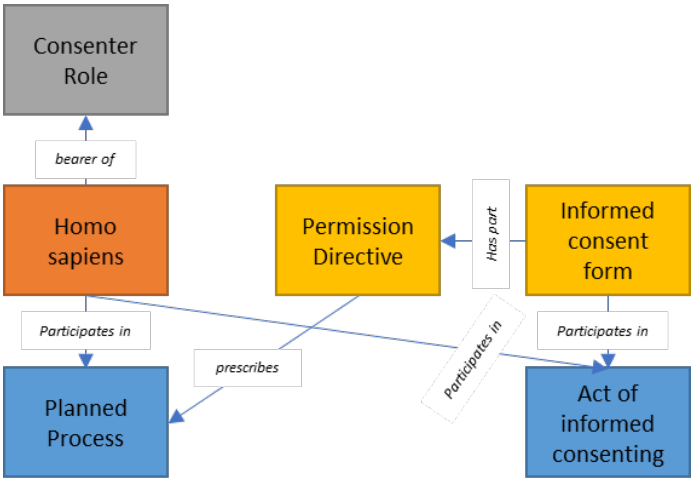
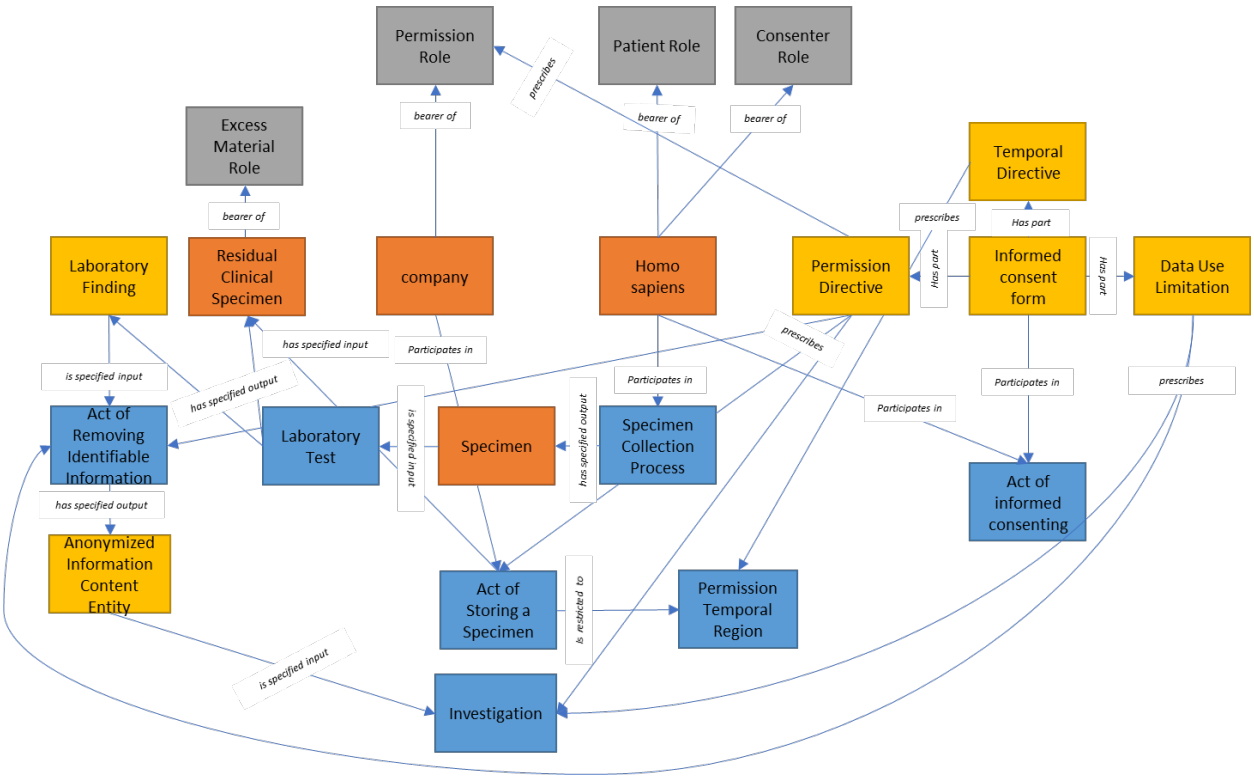


Figure 8. Informal Model for the following permission sentence: "…I give permission for GeneDx to retain any remaining sample longer than 60 days after completion of testing and use my de-identified data for scientific and medical research purposes."



3.2. Reflection and Negotiation

At this time, all of the sampled source data from US Federal Regulations and clinical consent forms have been modeled and reviewed with team members. We asked the following evaluation questions to evaluate the proposed revisions and extensions of ICO in terms of verification and validation, and to determine next steps:

3.2.1. Verification

1. **Are the proposed revisions adherent to OBO Foundry principles?**[114]

OBO Foundry principles are centered on openness and reusability to a range of users, and facilitating reuse through shared best practices such as common formats, clear definitions, community collaborations, and regular maintenance of the ontologies.[114] ICO is released under a Creative Commons 3.0 license and written in web ontology language (OWL). All additional principles have been meet under previous evaluation efforts.[21; manuscript in progress] The proposed revisions from this evaluation do not affect ICO's adherence to OBO Foundry principles, and any added classes will follow the naming, defining, and documentation requirements.

2. **Are the proposed and revised classes consistent across the ontology's hierarchy?**[115]

We have made every effort to appropriately map new classes at the same level of granularity as similar classes across the ontology. Three of the seven proposed revisions to existing ICO classes revolved around editing the location within the class hierarchy, i.e., their formal definitions.

3. **Are proposed classes non-redundant?**[115]

During the informal modeling phase, we attempted to use existing ICO classes wherever possible prior to suggesting extension of the ontology by adding classes. Likewise, we also searched Ontobee, a web-based tool for viewing the OBO Foundry suite of ontologies, to identify classes and their definitions (formal and textual) that could be imported into ICO. The shared logical structure and hierarchy of classes in the OBO Foundry facilitates interoperability and ease of importing individual classes or entire branches of the ontologies' trees.

4. **Does the model capture only entities within the specified scope of the ontology?** [21]

Permissions from the source data and their relevant entities largely included terms specific to consent processes, their forms, and the individuals involved in the consent process; however, they also included entities that are relevant outside of consent processes, including health care processes such as surgeries, organizations such hospitals or biorepositories, and processes like the act of selling some material (e.g., biospecimen). For this reason, we plan to continue to negotiate all newly proposed classes both with the ICO development team and representatives from a range of OBO Foundry ontologies to determine where they fit best. As an example, efforts have already been completed to add and define the term 'owner role' in The Document Acts Ontology (D-Acts), which shares a namespace with the Information Artifacts Ontology (IAO). Even though ownership was a necessary concept identified through our analysis of the source data, the concept of ownership extends beyond the context of informed consent, for example including deeds and land ownership so merits modeling in an ontology that provides content inherited by 'downstream' ontologies such as ICO.

5. **Is the documentation sufficiently unambiguous to enable a consistent use of the terminology?** [21]

Text definitions for all newly added terms will be entered into the ontology. Additionally, these text definitions will be double checked to ensure that they match the terms' formal definitions.

3.2.2. Validation

1. **Does the ontology contain the necessary and sufficient info (identified through the source data) to make it fit for our particular purpose? Are all entities within the scope of the ontology captured?** [21,115]

As a reminder, our purpose was to represent permissions relevant to clinical care contexts, including reuse of residual clinical biospecimens and health data for secondary purposes. By systematically extracting entities from all relevant US Federal regulations and a sample of permission-sentences from clinical consent forms, we aimed to represent all necessary and sufficient classes for our purposes in ICO. After modeling a few permission-sentences, a distinct and consistent design pattern emerged for representing permission-sentences (this design pattern is previously provided in Figure 7). It should also be noted that by following our progression of modeling permission-sentences – from those purposively sampled for our most narrow use case, then purposively selected for heterogenous clinical permissions, and lastly a random sample of permission-sentences – we achieved saturation when identifying new classes, which demonstrates that a sufficient and representative sample of permission-sentences were modeled to capture most relevant classes.

2. **Do the domain experts agree with the ontological analysis?**[21]

We had multiple discussions ICO development team and OBIB community which informed iterative improvements to our preliminary informal models. We believe that we have

achieved a degree of stability in these revisions based on expert feedback, and that proposed classes can now be pushed forward to their respective ontological communities.

4. Discussion

This evaluation revealed substantial overlap and therefore appropriateness of using ICO to represent permissions in clinical care, with a focus on reusing residual clinical biospecimens and health data. It further identified gaps and inconsistencies in ICO for representing such permissions, and proposes a series of recommendations for extension and revision of ICO and surrounding OBO Foundry ontologies. In their study to identify the minimum metadata necessary for biobank information systems to share their samples and data, Norlin et al. recognized the need for ontologies which represent the "ethical standards under which the samples are collected, any restrictions on research use, and access requirements to the samples to expressing the permissions for reuse."[19] Once revised, ICO will serve as a valuable information resource for consent information that can be referenced by systems which manage patients' residual clinical biospecimens and health data, in addition to those biospecimens and data captured from research studies.

This evaluation also demonstrates the value of developing and extending ontologies and other information resources based on a "bottom-up" modelling approach. Without use of real-world source data (regulations and consent forms), it is unlikely that the extent of the gaps in ICO and other BFO-based, OBO Foundry ontologies would have been identified.

Importantly, our evaluation and modeling process provides a roadmap for improving and expanding domain knowledge within ontologies, and addressing knowledge-representation gaps that hinder their successful uptake.[121] This role is particularly well-suited to clinical research informaticists, an emerging and recognized specialty that leverages informatics to discover and

manage new knowledge relating to health and disease and their use in research.[122] We also wish to amplify the need for collaboration and transparency in ontology development. There are hundreds to thousands of ontologies across repositories, but these ontologies are not necessarily interoperable. It is only with a shared semantic structure and collaborative negotiation of ontology structures that dynamic and changing knowledge can successfully be modeled in interoperable families of ontologies. This work must occur in open and collaborative communities to achieve transparency in knowledge representation.

Several limitations of this evaluation are recognized. First, we reviewed informal models and design patterns with ICO developers and collaborators from the OBIB community, but have not yet formally modelled the proposed classes and their relations according to the BFO hierarchy, or determined with the OBO Foundry which ontologies should be the home for each proposed term. Upon formal modeling, reasoners will be applied to the extended version of ICO and other OBO Foundry ontologies to test inferencing of the new classes. Second, while we have worked with ontologists and clinical research informaticists, we have not yet presented our models to domain experts from the legal and compliance, health information management, or biobanking domains. However, we believe that the extensive literature review and systematic collection and analysis of clinical consent forms mitigates this limitation; the literature and approved consent forms are a form of the collective voice of domain experts. Future refinements will include further checking with experts to ensure we have correctly interpreted the literature and the permissions.

5. Conclusion

Representing permissions to reuse residual clinical biospecimens and health data in an information resource that is interoperable within a family of recognized biomedical ontologies is

valuable for facilitating the responsible reuse of these resources at scale. By evaluating and proposing revisions and extension to ICO, we make a meaningful step in this direction. Our methods demonstrate the use of a bottom-up approach to modeling content from their respective domains' perspectives. We propose such methods as a valuable way for clinical research informaticists and domain experts to engage in the development and revision of semantic information resources.

Chapter 5 Conclusion

1. Introduction

When patients receive health care services, many clinical procedures involve the collection of blood or tissues (i.e., biospecimens) for some assessment, diagnostic, or therapeutic purpose. These biospecimens may be discarded once their clinical indication is complete, but the reality is that many end up in biorepositories.[123] When patients commonly sign consent documents indicating permission for clinical procedures, these clinical consent forms also often include statements which grant permission to reuse whatever leftover portions of their biospecimens remain (i.e., residual clinical biospecimens and health data) for secondary purposes, including but not limited to research, education, and commercial uses.[101]

The recognized value of these residual clinical biospecimens and associated data is rapidly increasing as methods for analyzing and aggregating these data advance. As an example, for precision health care to be actualized at the point of care, electronic health records and clinical decision support tools will require both genetic and health information to classify and identify persons according to their genotype (i.e., sequence of genetic code) and phenotype (i.e., classification into subpopulations based on variation in response or susceptibility to certain treatments or diseases).[9,124] While this type of treatment for medical conditions is yet to be performed on a broad scale, analyses of genomic information alongside health data have yielded more than 10,000 confirmed genotype-phenotype associations between 2014 and 2017.[125]

As custodians of biospecimens and heath data, biorepositories are responsible for safeguarding the interests of and the information about biospecimen donors. Biorepositories and

researchers must ensure that biospecimens are shared in a manner that is consistent with the permissions granted through regulations and by biospecimen donors during the informed consent process under which the biospecimens were obtained.[126] However, information about the permissions is often only available through the time-intensive process of manually retrieving and reviewing informed consent forms or when this information is "locked" in local information management systems.[13,72] Even when information about consent is contained within information systems, international best practices only recommend data elements such as form version identifiers, relevant dates, consent method (e.g., written or verbal), consent status (e.g., whether or not consent was provided and/or withdrawn), and identifiers for the person from whom consent was obtained.[127] Permissions for biospecimen and data use are generally not fully represented in such systems.

Although humans may read regulatory and consent documents and intuitively gain understanding of their meaning (i.e., semantics) based on contextual clues, they lack the speed and cognitive processing to retain and process large quantities of information at scale and the ability to understand information when context is lacking.[102] Computers, on the other hand, are powerful machines capable of great speed and processing, but *on their own* lack any reliable way to "understand" semantics.[102,128] How knowledge is classified (or organized) and expressed in technologies can enable or constrain discovery of that knowledge.[129,130]

The growing demand to increase the availability of biospecimens and data must be considered in light of ethical imperatives to responsibly use these highly sensitive information resources by adhering to the permissions granted by the biospecimen donors and the rules and regulations which govern those biospecimens. Identification and management of information about informed consent and relevant regulations must be 'known' to all individuals and digital

systems that interact with the repositories in which the specimen and data are stored as well as the repositories that receive shared biospecimens and data. In addition to potential misuse of specimens and data, without such information management it is possible (if not likely) that banked biospecimens may be under-utilized because the determination of shareability is not discoverable, and/or shared without traceability of these permissions.

Ontologies, or semantic web technologies, show great promise for representing such a complex field in a machine-interpretable manner. The term ontology, designated by philosophers as the study of the nature of existence,[131] has been adopted by information and web professionals to denote files which formally specify classes of entities and relationships between them.[17] Ontology files specify the rules of inference and inheritance, which enable computers to extract the contextual meaning (i.e., the *semantic* aspect of the semantic web) of terms and manipulate that information in ways similar to those of human users.[128] These technologies use machine-readable syntax to capture, process, and communicate meaning and knowledge surrounding entities and inter-entity relationships in the real world.[132] By embedding semantics in technologies, classification and organization of knowledge are enabled "in a definite and meaningful way."[133] Likewise, when semantics are encoded into web content, that information becomes actionable by computers in a way that is different from syntactically-identical sources, such as is printed in a book.[128,134]

Ontologies also address another challenge faced by computers: the need for an infrastructure which enables interoperability, or the integration and circulation of potentially diverse data across diverse contexts with minimal effort.[102,130] For example, interoperability is necessary for continuous aggregation and integration of health data through linking to databases operated independently from the biorepositories.[127] While the challenge of

interoperability can in-part be addressed using machine-readable languages and standards, "languages and standards are of no consequence without uptake, and uptake requires increasing the amount of data exposed in RDF."[134] The Resource Description Framework (RDF) is language for writing ontologies that encodes meaning about entities by encoding it as sets of three (called triples), much like the subject, predicate, and object of a sentence.[128] These subjects and objects become nodes of the semantic web, and the predicates (i.e., relationship statements) serve as the connections which link these nodes.[132] Through this common format and "understanding of terms, scientific data and findings [can] be consistently annotated thus promoting data integration and exchange across heterogeneous representations." [104]

In addition to the structure and language of ontologies, it is tremendously valuable that semantic technologies for knowledge discovery integration be developed and maintained within open communities for the vision of interoperability to be achieved. "Creating bespoke [software,] in all computer languages, for all data-types and all analytical tools that require those data-types, is not a sustainable activity."[102] To address this need, development of broadly useful and interoperable ontologies benefit when built to fit within a network of interoperable knowledge representations by an open community.

One such open community is the Open Biological and Biomedical Ontologies (OBO) Foundry, a collaborative dedicated to building and maintaining ontologies for the life sciences.[18] All OBO Foundry ontologies are realism-based and reference the shared logical structure of the Basic Formal Ontology (BFO). Realism-based ontologies often initiated through a "top-down" approach, in which "development process [start] with the definition of the most general concepts in the domain and subsequent specialization of the concepts."[17] To achieve the aim of realism-based ontologies to classify entities and knowledge about those entities in

such a way that real-world instances can be described, further classification and development is required to embed domain expertise in ontologies at the application level.[104,135] Doing so "[encapsulates] the knowledge of the world that is associated with the general terms used by scientists in the corresponding domain."[136]

However, top-down approaches to modeling may contribute to gaps in knowledge representation that do not reflect users' needs at the local level. It has previously been demonstrated that using a bottom-up approach can be used to identify these gaps and extend or expand information resources for interoperability.[20] Such a bottom-up approach leverages real-world instances and data sources and applies a data driven approach to systematically identify content that must be represented in an information resource, and evaluate the coverage of that resource for the source data's domain.

This dissertation employs a bottom-up modeling approach to first explore the necessary information regarding permissions to reuse residual clinical biospecimens and health data, informally model this information, and evaluate its representation within the Informed Consent Ontology (ICO). The following sections of this chapter summarize the results from each of the dissertation studies, discusses the significance of this work and related initiatives for computably representing consent or permissions, and concludes with envisioned next steps and future work.

2. Summary of Chapter 2

Chapter 2 of this dissertation reported on a scoping review of the literature, which served as one source of data towards our bottom-up modelling approach. The purposes of this scoping review were threefold:

1. To identify relevant federal regulations and norms (e.g., best practices, guidance documents) that bear authority or give guidance over reuse of residual clinical biospecimens and health data in the US.
2. To gather domain experts' interpretations of the permissions by which such reuse may occur.
3. To gather domain experts' interpretations of the key issues and concepts – including the overlap, congruency, and gaps – that must be considered when interpreting these regulations and norms.

Upon querying six biomedical, legal, and health policy databases; screening manuscripts; and reviewing full text articles, twenty-five manuscripts were included in the final review. These twenty-five manuscripts covered eleven US federal regulations and fifteen norms that bear authority or guide reuse of residual clinical biospecimens in the US.

We found considerable inconsistencies in permissions across a patchwork of regulations of norms that make it difficult to tell which apply to a given use, or if the applicable regulations or norms even agree with each other. Further, even once the applicability of these rules is determined, there is considerable administrative and logistical burden for tracing such decisions over time, physical location, and across jurisdiction.[35,52,65] The responsible conduct of research is challenging given the complexity of navigating these regulations and norms, and the burden of humans' interpretation of these permissions on a case-by-case basis. As such, there is growing need for tools to "flag" residual clinical biospecimens and health data as usable or unusable for a range of given activities at scale. These findings support the need for information regarding permission to reuse residual clinical biospecimens and health data to be represented in

standardized, computable, and interoperable formats. Moreover, this work provides the necessary information for such information systems.

3. Summary of Chapter 3

The second source of source data for our bottom-up modelling approach was produced through the annotation study reported in Chapter 3. The purpose of this study was to develop and test an annotation scheme for identifying permissions within clinical consent forms. This was a retrospective study of clinical consent forms, examining forms collected using two methods: direct contribution by healthcare facilities and systematic web searching. We followed Pustejovsky and colleagues' MAMA (Model-Annotate-Model-Annotate) cycle as the method for developing and refining our annotation scheme.[80] We proposed an initial set of instructions for annotations, and then iteratively tested those instructions using a small subset of the data and revised the annotation scheme accordingly. We determined inter-annotator agreement using multiple methods: raw or observed agreement, Cohen's κ, weighted κ (κ_w), and Krippendorff's α.

The final annotation scheme demonstrated moderate to substantial inter-rater reliability with three annotators identifying permission-sentences. Of note, we did not find any other approaches or projects in the literature to annotate permission-sentences. While the performance of this annotation scheme may in-part be due to its quality or sufficiency, we believe that these findings may point to the complexity of identifying permission-sentences within the forms, which aligns with a significant body of literature which speaks to clinical consent forms being convoluted and incomprehensible.[99] This is ethically problematic as the nature of *informed* consent requires that information provided in consent processes and documentation be

understandable by consenters (e.g., clinicians) and consentees (e.g., health care consumers) alike.[90]

To mitigate the amount of disagreement around identifying permission-sentences, only those sentences annotated by all three annotators (n= 211) were eligible as source data for the final ontology evaluation. Should permission-sentences ever serve as training data to automatically identify permission-sentences in new samples of consent forms, further enhancement of the annotation output must occur through careful selection of which sentences should or should not be included as training data.[97] These sentences might also be used to evaluate the coverage of ICO to verify or extend the coverage of ICO, should such analysis reveal that this set of sentences is qualitatively different from the 211 agreed-upon sentences.

Use of permission-sentences identified through real-world consent forms to extend ICO lays an important foundation for machine-interpretation of permission information in clinical and clinical research information systems. It also contributes to scientific discovery by increasing the accessibility of materials and information resources that are permitted to be shared and protecting patients by honoring their agency and decision-making around complex issues.

4. Summary of Chapter 4

Lastly, Chapter 4 reports on our evaluation of the Informed Consent Ontology. The purpose of this evaluation study was to apply a hybrid model of ontology evaluation, with focus on evaluating ICO for its completeness in expressing permissions from regulations and clinical consent forms, particularly for reuse of residual clinical biospecimens and health data. We also aimed to develop recommendations for extension and revision based on these findings. We followed a formative evaluation design, adapting Friedman and Wyatt's multistep evaluation

process to iteratively examine and check our findings, and integrated steps recommended in the ontology evaluation literature.

We found considerable overlap of classes and object properties between ICO as presently published and the necessary content identified through federal regulations and clinical informed consent forms from the previous two studies. This overlap demonstrates the appropriateness of extending ICO for the clinical domain rather than developing a new ontology outright. However, we also found that considerable extension – either in the ICO namespace or another OBO Foundry ontology – is necessary to represent permissions from clinical consent, with a focus on reuse of residual clinical biospecimens and health data. We propose two foundational domain patterns based on our informal modeling step – one of the context of a permission-sentence and one demonstrating all classes involved when a patient permits reuse of their deidentified residual clinical biospecimens and health data.

Our evaluation using questions in the literature demonstrates that, while we have already initiated collaborations to vet and extend the ICO and other OBO Foundry ontologies, these collaborations must be continued and ongoing. Once extended and revised, ICO will serve as a valuable information resource for consent information that can be referenced by systems which manage patients' residual clinical biospecimens and health data, in addition to those biospecimens and data captured from research studies.

5. Significance

This work addresses a critical gap in formally representing the permissions of how residual clinical biospecimens and health data can be responsibly reused. which is of increasing value in data-intensive health sciences (e.g., population health, translational science, precision health, etc.) which require access to such information. It also adds this missing content to the

BFO-based biobanking suite of ontologies, enabling use alongside widely-adopted biomedical ontologies and acting as a valuable contribution for biobank information management. This work produces a machine-readable semantic resource, grounded in real-world sources included regulations and clinical consent forms.

It is our hope that deployment of ICO will facilitate responsible reuse of residual clinical biospecimens and health data, both increasing their discoverability by entities that will advance scientific knowledge related to health and also protecting the agency and rights of patients, whose expressed choices regarding the disposition of their biospecimens and health data should be respected. Revising and extending ICO to this new domain enables future logical inferencing about permissions and obligations for sharing and use which can be adopted by a range of systems and applications. As examples, ICO may also be used for assessing informed consent tools such as eConsent forms and as a resource for mapping and annotating text in future natural language processing tasks. It may also be used to build query tools or decision support systems to support covered entities, biorepositories, federated research networks, institutional review boards, and other individuals in identifying eligible biospecimen and data resources that meet their needs or deciding if and when certain biospecimen and data resources should be shared.

Another strength and important feature of this work is the demonstration of a bottom-up modeling approach to evaluate and extend an existing information resource using real-world source data. We suggest that this approach may demonstrate a clear entry point for domain experts to interact with the ontology development community, to be involved in formal modeling efforts, and to develop ontologies that meet the needs of their communities both broadly and at the local levels.

6. Related Efforts

In addition to ICO, there are a number of initiatives related to computable representations of consent and permissions presented throughout the literature. Nearly a decade ago, two permissions ontologies emerged in the literature, both for representing permissions granted during research consent processes. Obeid and Sanderson introduced the Biomedical Research Permissions Ontology as a part of Health Sciences South Carolina's (HSSC) initiative to provide a "comprehensive mechanism for managing informed consents and other research permissions."[137] Grando and Schwab likewise introduced their permissions ontology, designed and tested using actual investigator queries and study participants' data to determine reasoners' conformance to the permissions granted by those study participants in consent forms.[72] Each of these efforts focused on consent permissions granted by research participants of IRB-regulated studies, not for reuse of residual clinical biospecimens. Furthermore, despite initial publication of efforts, it does not appear that either of these ontologies are actively maintained or made publicly accessible online.

Others have addressed the need for common terminology in the context of biobanks, including a common language of informed consent.[138] However, the ontology development efforts that sought to express such language as an ontology, the Ontologized Minimum Information about Biobank Data Sharing (OMIABIS), stops just short of representing informed consent. In recent years, collaborations have occurred between members of the ICO and OMIABIS development communities to sufficiently represent consent processes in the context of biobanking.[139] Lastly, the Data Use Ontology (DUO) is now adopted by the Global Alliance for Genomics & Health (GA4GH) as standard coded data elements for describing data use

categories at the data set level, but not the level of permission-sentences within a consent document.[140]

However, ontologies are not the only format for facilitating biomedical information exchange. HL7 is a family of standards for the facilitation of health information exchange, many of which automate data sharing and underlie systems for providing patient care. One standard, the Fast Healthcare Interoperability Resources (FHIR) Specification "aims to simplify implementation without sacrificing information integrity."[141,142] FHIR's most recent release includes an in-trial representation of consent, the development of which is currently in-progress by FHIR's Community Based Collaborative Care Work Group.[143] Data elements include a diverse range of patient, policy, and data information. The scope of FHIR's representation of consent targets four uses: (a) privacy, to collect, access, use or share information; (b) medical treatment, to undergo treatment; (c) research, to participate in research; and (d) advance care directives. However, only the privacy directive is modeled at this time.

This dissertation work builds on the Informed Consent Ontology. Like the aforementioned permissions ontology efforts, ICO was developed to represent permissions within research consents. In this set of studies, we used a bottom-up approach to identify detailed content and language related to permissions in clinical consent processes, with a focus on when patients are asked for permission to reuse their residual clinical biospecimens and health data for secondary purposes. We completed an in-depth review of the literature as well as clinical consent documents to identify key entities and relationships. Leveraging design patterns from ICO and other ontologies, we demonstrated a way forward for inclusion of permissions related to clinical consents within ICO. In this way, we provide a uniformly modeled framework for addressing consent permissions, whether they arise from research or clinical consents.

7. Next Steps

Going forward, we anticipate several next steps that will be essential if ICO is to be adopted. First, we envision development of a permissions ontology that directly integrates with the extended ICO. This ontology requires modeling the semantic structure of a permission-sentence, and all the variability that may occur across permission-sentences. This effort is presently underway, and is likely to be particularly useful in the identification and parsing of permission-sentences in natural language processing tasks.

Second, we envision and are making efforts towards collaborating with key standards groups, aiming for integration and interoperability among our efforts. Towards this aim, we believe ICO could be an asset to Ga4GH efforts. Ga4GH already uses DUO, which is interoperable with ICO based on the shared BFO hierarchy and compliance with OBO Foundry principles. In addition to data use codes at the data set level (a limitation of DUO), ICO can support the need for granular expressions of permissions within consent forms.

Third, we hope to participate in further iterations and extensions of FHIR/HL7's Security and Privacy Ontology and Consent Resource. This initiative is rapidly gaining increased and national attention. In 2019, "the ONC Cures Act Notice of Proposed Rulemaking (NPRM) 5 was released, which intends to advance and support seamless and secure access to, exchange of, and use of [electronic health information], particularly through apps."[144] While it is envisioned that patients and healthcare consumers will manage their health information, it is likewise expected that there will be increasing demand for computable and interoperable representations of consent and permissions within these applications. The FHIR consent resource is currently a focus of major development efforts to enable interoperability; however, as previously mentioned, only the privacy use case is modeled. Additional work is needed to test and refine this resource,

in addition to developing the Medical Treatment Consent Directive, Research Consent Directive, and Advanced Care directive resources. There is demand for these resources, and we believe our perspective and foundation in disentangling clinical consent and permissions makes us suitable candidates to contribute to such efforts.

In summary, the work presented in this dissertation provides an essential and previously missing foundation for representing informed consent and permissions related to reuse of residual biospecimens and health data. We envision that patients will not only have agency over what happens to them, their bodies, biospecimens, and data, but also that these permissions will be traceable and interpretable over time and across systems which reference these permissions. In this way, our work may contribute to protecting patients. While such applications do not yet exist, extension of ICO for clinical permissions is an initial and foundational step towards this vision.

Appendices

Appendix A: Complete Search Strategy including Search Terms and Results by Database

	PubMed	NexisUni	HeinOnline	Health Policy Reference Center	Philosopher's Index	Public Affairs Index
AND	specimen OR specimens OR biospecimen* OR "Specimen Handling"[Mesh]		specimen OR specimens OR biospecimen*	specimen OR specimens OR biospecimen*		
	sharing OR share* OR releas* OR distribut*		sharing OR share* OR releas* OR distribut*	--		
	permission OR permissions OR consent OR consents		permission OR permissions OR consent OR consents	permission OR permissions OR consent OR consents)		
	--		HIPAA OR "Health Insurance Portability and Accountability Act" OR "FDA" OR "Food and Drug Administration"	--		
Filters	1997-present; English only	1997-present; US Federal Jurisdiction, Category: Law Reviews and Journals	1997-present; Law Journal Library and (Location) United States of America, United States	1997-present; English only		
Results	**345**	**247**	**332**	**183**	**32**	**28**

Appendix B: Facilities in Sample of Clinical Consent Forms

Table 14. List and Descriptions of All Facilities in Sample of Clinical Consent Forms

Facility Name	City	State	Ownership	Search Strategy	Form Count
Acute Care Hospitals					
Bon Secours St Mary's Hospital	Richmond	VA	NFP - Church	RAND	2
Boston Medical Center	Boston	MA	NFP - Private	CTSA	1
Cape Fear Valley Health Hoke Hospital	Raeford	NC	NFP - Private	RAND	1
Columbia University Irving Medical Center	New York	NY	NFP - Private	CTSA	1
Doctors' Community Hospital	Lanham	MD	Proprietary	RAND	1
Emory Healthcare	Atlanta	GA	NFP - Private	CTSA	1
GHS Laurens County Memorial Hospital	Clinton	SC	NFP - Private	RAND	1
Hospital of The University of Pennsylvania	Philadelphia	PA	NFP - Private	CTSA	1
Indiana University Health	Indianapolis	IN	NFP - Private	CTSA, RAND	3
Indiana University Health Arnett Hospital	Lafayette	IN	NFP - Private	RAND	1
Intermountain Medical Center	Murray	UT	NFP - Private	RAND	3
Johns Hopkins Health System	Baltimore	MD	NFP - Private	CTSA	1
Karmanos Cancer Center	Detroit	MI	Proprietary	MHIMA	1
Mayo Clinic Hlth System Franciscan Med Ctr	La Crosse	WI	NFP - Private	RAND	1
Mayo Clinic Rochester	Rochester	MN	NFP - Church	CTSA, RAND	1
Mclaren Oakland	Pontiac	MI	NFP - Private	MHIMA	1
Montefiore Medical Center	Bronx	NY	NFP - Other	CTSA	1
Mount Carmel New Albany Surgical Hospital	New Albany	OH	Proprietary	RAND	1
Mount St Mary's Hospital and Health Center	Lewiston	NY	NFP - Church	RAND	1
MUSC Medical Center of Medical University of South Carolina	Charleston	SC	Govt. - State	CTSA	1
North Austin Medical Center	Austin	TX	Proprietary	RAND	2
Northwestern Memorial Hospital	Chicago	IL	NFP - Private	CTSA	2
OHSU Hospital	Portland	OR	Govt. - State	CTSA	5
Piedmont Athens Regional Medical Center	Athens	GA	NFP - Other	RAND	1
Piedmont Rockdale Hospital	Conyers	GA	Proprietary	RAND	1
St Bernard's Medical Center	Jonesboro	AR	NFP - Private	RAND	2
St Lucie Medical Center	Port Saint Lucie	FL	Proprietary	RAND	2
St Luke's Des Peres Hospital	Saint Louis	MO	Proprietary	RAND	1
St Mary Mercy Hospital	Livonia	MI	NFP - Private	MHIMA	13
Strong Memorial Hospital	Rochester	NY	NFP - Private	CTSA, RAND	16
SUNY Downstate Medical Center University Hospital	Brooklyn	NY	NFP - Private	CTSA	1

ThedaCare Regional Medical Center - Neenah Inc	Neenah	WI	NFP - Private	RAND	1
UC Health	Cincinnati	OH	NFP - Private	CTSA	1
UC San Diego Health Hillcrest - Hillcrest Med Ctr	San Diego	CA	NFP - Other	CTSA	1
UF Health Shands Hospital	Gainesville	FL	NFP - Private	CTSA	1
Facility Name	**City**	**State**	**Ownership**	**Search Strategy**	**Form Count**
UNC Health Care System	Chapel Hill	NC	NFP - Private	CTSA	3
University of California Davis Medical Center	Sacramento	CA	NFP - Other	CTSA	2
University of Iowa Hospitals and Clinics	Iowa City	IA	NFP - Private	CTSA	3
University of Kansas Hospital	Kansas City	KS	Govt. - Hospital District or Authority	CTSA	1
University of Kentucky Albert B. Chandler Hospital	Lexington	KY	Govt. - State	CTSA	1
University of Miami Health System	Miami	FL	NFP - Private	CTSA	1
University of Michigan Health System	Ann Arbor	MI	NFP - Private	CTSA, MHIMA, RAND	16
University Of New Mexico Hospitals	Albuquerque	NM	Govt. - State	CTSA	1
UPMC	Pittsburgh	PA	NFP - Private	CTSA	1
UT Southwestern University Hospital-Zale Lipshy	Dallas	TX	Govt. - State	CTSA	1
VCU Medical Center	Richmond	VA	Govt. - Hospital District or Authority	CTSA	2
Wellstar Cobb Hospital	Austell	GA	NFP - Other	RAND	3
Yale-New Haven Health System	New Haven	CT	NFP - Private	CTSA	1
Children's Hospitals					
Children's National Health System	Washington	DC	NFP - Private	CTSA	1
Critical Access Hospitals					
CHI St Alexius Health Williston	Williston	ND	NFP - Church	RAND	2
Christus Mother Frances Hospital-Winnsboro	Winnsboro	TX	NFP - Private	RAND	2
Pinckneyville Community Hospital	Pinckneyville	IL	Govt. - Hospital District or Authority	RAND	1
UHHS Memorial Hospital of Geneva	Geneva	OH	NFP - Private	RAND	1
Psychiatric Hospitals					
Pacific Grove Hospital	Riverside	CA	Proprietary	RAND	1
Springfield Hospital Center	Sykesville	MD	Govt. - State	RAND	1
Windhaven Psychiatric Hospital	Prescott Valley	AZ	NFP - Private	RAND	1
Ambulatory Surgical Centers					
Chambersburg Endoscopy Center LLC	Chambersburg	PA	*(not specified)*	RAND	3
CSA Surgical Center LLC	Columbia	MO	*(not specified)*	RAND	2
Georgia Lithotripsy and Laser Center	Athens	GA	*(not specified)*	RAND	2
Gwinnett Endoscopy Center Pc	Lawrenceville	GA	*(not specified)*	RAND	2
Hamilton Surgical Center Inc	Hamilton	NJ	*(not specified)*	RAND	1
Spivey Station Surgery Center	Jonesboro	GA	*(not specified)*	RAND	1
***Key**: Govt. = Government; NFP = Not for Profit; RAND = Randomly Selected*					

Appendix C: Final Coding Guidelines and Procedures for Annotation

Purpose

The purpose of this document is to describe annotation criteria so that human annotators can identify permissions in the provided texts (i.e., consent forms).

Definition

For purposes of this project, we define a permission as statement(s) that, upon signature of the consent form, authorizes any new action or activity that may, must, or must not be done.

- *We are **NOT** interested in permissions for payment or hospital operations. Do not annotate these permissions.*

Login and Technical Instructions

1. Navigate to DataTurks and Login
2. On the "My Projects" page, click on the project that you intend to work on (see project names under each annotation round below)
3. On the project's dashboard, click "Start Tagging"
4. You will be directed to eligible texts for annotation. Follow the corresponding instructions (below) for each annotation round.

Identifying Relevant Permissions (and Confidence)

1. Open your DataTurks Project
2. Read the entirety of the consent form. Do not mark up any text.
3. Read through the form again. Highlight the sentence(s) which contain permission(s), selecting a label corresponding with your confidence that it is a permission:
    - **A: I am certain that this is a permission.**
    - **B: This may be a permission.**

NOTES

For all annotations, highlight the entire sentence(s) from the first non-whitespace character to the last punctuation character.

- If there is no end punctuation, you may need to assume its presence.
- If statements indicate free text entry, highlight to the end of the text entry space (e.g., line ___) if visible in the .txt file (do not highlight trailing hard returns or white space)

There are various configurations of permissions and sentences:

- If a sentence contains a single permission, highlight that sentence once.
- If a sentence contains multiple permitted actions/activities, highlight the entire sentence once.
- If sequential sentences contain different permissions, highlight them separately.

Do not annotate permissions related to payment or hospital operations (e.g., insurance, contact)

4. Read through the form again. Verify all annotations.
5. Once you have completed all annotations on a text, click “Move to Done” at the bottom of the page (or ctrl+entr). Or, if you want to revisit a text before submitting it as done, click “Skip” at the bottom of the page (or ctrl+q).

Appendix D: Pairwise Agreement Measures and Corresponding Contingency Tables

Pairwise Inter-annotator Agreement across 3 Labels *(A:Is, B:MightBe, C:Not)*

Table 15. Inter-annotator Agreement between PI and RA across 3 Labels (A:Is, B:MightBe, C:Not)

		PI			
		A:Is	*B:MightBe*	*C:Not*	
RA	*A:Is*	237	18	158	413
	B:MightBe	7	3	46	56
	C:Not	89	38	5803	5930
		333	59	6007	**6399**

Agreement Measure		***95% CI***
A_o	0.9444	--
Cohen's k	0.5606	0.5162-0.6050
k_w	0.5997	0.5603-0.6391

Table 16. Inter-annotator Agreement between RN and RA across 3 Labels (A:Is, B:MightBe, C:Not)

		RN			
		A:Is	*B:MightBe*	*C:Not*	
RA	*A:Is*	268	3	142	413
	B:MightBe	7	1	48	56
	C:Not	176	25	5729	5930
		451	29	5919	**6399**

Agreement Measure		***95% CI***
A_o	0.9373	--
Cohen's k	0.5466	0.5036-0.5896
k_w	0.5731	0.5342-0.6120

Table 17.Inter-annotator Agreement between PI and RN across 3 Labels (A:Is, B:MightBe, C:Not)

		PI			
		A:Is	*B:MightBe*	*C:Not*	
RN	*A:Is*	268	17	166	451
	B:MightBe	3	1	25	29
	C:Not	62	41	5816	5919
		333	59	6007	**6399**

Agreement Measure		***95% CI***
A_o	0.9509	--
Cohen's k	0.6165	0.5751-0.6579
k_w	0.6504	0.6132-0.6876

Pairwise Inter-annotator Agreement across 2 Labels *(A=B, C:Not)*

Table 18. Inter-annotator Agreement between PI and RA across 2 Labels (A=B, C:Not)

		PI		
		A=B	*C:Not*	
RA	*A=B*	265	204	469
	C:Not	127	5803	5930
		392	6007	**6399**

Agreement Measure		***95% CI***
A_o	0.9483	--
Cohen's k	0.5881	0.5449-0.6313

Table 19. Inter-annotator Agreement between RN and RA across 2 Labels (A=B, C:Not)

		RN		
		A=B	*C:Not*	
RA	*A=B*	279	190	469
	C:Not	201	5729	5930
		480	5919	**6399**

Agreement Measure		***95% CI***
A_o	0.9389	--
Cohen's k	0.5550	0.5123-0.5977

Table 20. Inter-annotator Agreement between PI and RN across 2 Labels (A=B, C:Not)

		PI		
		A=B	*C:Not*	
RN	*A=B*	289	191	480
	C:Not	103	5816	5919
		392	6007	**6399**

Agreement Measure		***95% CI***
A_o	0.9541	--
Cohen's k	0.6385	0.5981-0.6789

Appendix E: Terms to Express Permission-Sentences in Clinical Consent Forms

Table 21. Terms and Sentence References to Express the Permission-Sentence: "By signing this form, I am requesting and giving my consent for MHSM and the doctors and/or nurses to give me blood and/o blood products during this admission or series of treatments."

By signing this form, I am requesting and giving my consent for MHSM and the doctors and/or nurses to give me blood and/o blood products during this admission or series of treatments.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'consent for MHSM and the doctors and/or nurses'
OBI:0000093	Patient Role	'to give me blood'
ICO:0000086	Consenter Role	'I am requesting'
OBI:0000207	health care provider role	'doctors and/or nurses'
OBI:0000245	organization	'MHSM'
OMRSE:00000054	hospital role	'MHSM'
OMRSE:00000056	Hospital organization	'MHSM'
NCBITaxon:9606	Homo Sapiens	'I', 'me', 'doctors and/or nurses'
OGMS:0000096	Health Care Process	'give me blood and/or blood products'
OGMS:0000090	Treatment	'give me blood and/or blood products'
ICO:0000019	Signature Section	'By signing this form'
ICO:0000222	Signature	'By signing this form'
ICO:0000196	act of informed consenting	'I am requesting'
ICO:0000001	Informed consent form	'this form'
ICO:0000244	Permission Directive	*entire sentence*
--	Temporal Directive	'during this admission or series of treatments'
ICO:0000186	Permission Temporal Region	'during this admission or series of treatments'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
DUO:0000010	Is restricted to	
OBI:0000293	has:specified:input	
RO:0000053	bearer of	
RO:0000056	Participates in	
RO:0002350	Member of	

Table 22. Terms and Sentence References to Express the Permission-Sentence: "Your signature below indicates that you understand to your satisfaction the information about the genetic testing ordered by your health care provider and that you consent to having this testing performed."

Your signature below indicates that you understand to your satisfaction the information about the genetic testing ordered by your health care provider and that you consent to having this testing performed.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'that you consent to having this testing performed'
OBI:0000093	Patient Role	'ordered by your health care provider' and 'to having this testing performed'
ICO:0000086	Consenter Role	'your signature below indicates... consent'
OBI:0000207	Health Care Provider Role	'health care provider'
NCBITaxon:9606	Homo Sapiens	'your', 'health care provider'
ICO:0000019	Signature Section	'signature below'
ICO:0000222	Signature	'signature below'
ICO:0000196	Act of Informed Consenting	'your signature below indicates... consent'
ICO:0000001	Informed consent form	'signature below'
ICO:0000312	Genetic testing directive	'consent to having this testing performed'
ICO:0000311	Act of Genetic Testing	'genetic testing'
ICO:0000244	Permission Directive	*entire sentence*
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 23. Terms and Sentence References to Express the Permission-Sentence: "I consent to the Facility videotaping, photographing, video monitoring, or taking other recordings of me or parts of my body for diagnosis, treatment, research, or patient safety purposes."

I consent to the Facility videotaping, photographing, video monitoring, or taking other recordings of me or parts of my body for diagnosis, treatment, research, or patient safety purposes.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'I consent to'
OBI:0000093	Patient Role	'for diagnosis, treatment, ..."
ICO:0000086	Consenter Role	'I consent to'
OBI:0000245	Organization	'the facility'
NCBITaxon:9606	Homo Sapiens	'I'
--	Health Care Organization	'the facility'
OGMS:0000096	Health Care Process	'I consent to'
ICO:0000196	act of informed consenting	'I consent to'
ICO:0000001	Informed consent form	*Implicit*
IAO:0000572	Documenting	'videotaping, photographing, video monitoring, or taking other recordings'
BFO:0000040	Material Entity	'videotaping, photographing, video monitoring, or taking other recordings'
IAO:0000030	Information Content Entity	'videotaping, photographing, video monitoring, or taking other recordings'
ERO:0000333	Video	'videotaping', 'video monitoring'
IAO:0000185	Photograph	photographing'
ICO:0000244	Permission Directive	*entire sentence*
DUO:0000001	Data Use Limitation	'for diagnosis, treatment, research, or patient safety purposes'
OGMS:0000104	Diagnostic Process	'diagnosis'
OGMS:0000090	Treatment	'treatment'
OBI:0000066	Investigation	'research'
NCIT:C163409	Safety Monitoring	'patient safety purposes'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
OBI:0000295	is:specified:input:of	
OBI:0000299	has:specified:output	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 24. Terms and Sentence References to Express the Permission-Sentence: "I hereby consent to the use and disclosure of my protected health information as described in the Notice of Privacy Practices."

I hereby consent to the use and disclosure of my protected health information as described in the Notice of Privacy Practices.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'consent to the use and disclosure of'
OBI:0000093	Patient Role	'I', 'my protected health information'
ICO:0000086	Consenter Role	'I hereby consent to'
NCBITaxon:9606	Homo Sapiens	'I', 'my'
ICO:0000196	act of informed consenting	'I hereby consent to'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000244	Permission Directive	*entire sentence*
ICO:0000233	Regulatory document	'Notice of Privacy Practices'
ICO:0000116	Act of Using Participant Data	'use'
ICO:0000228	Act of Data Sharing	'disclosure'
ICO:0000074	authorization for release of confidential health information	'I hereby consent to the use and disclosure of my protected health information'
ICO:0000103	Protected Health Information Entity	'protected health information'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
IAO:0000136	Is about	
OBI:0000295	is:specified:input:of	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 25. Terms and Sentence References to Express the Permission-Sentence: "I hereby authorize XXX to retain, preserve and use for scientific or teaching purposes, or to dispose at its discretion or convenience, any specimen or tissues taken from my body during my visit."

I hereby authorize XXX to retain, preserve and use for scientific or teaching purposes, or to dispose at its discretion or convenience, any specimen or tissues taken from my body during my visit.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'hereby authorize XXX'
OBI:0000093	Patient role	'I', 'my body during my visit'
ICO:0000086	Consenter Role	'I hereby authorize'
OBI:0000207	health care provider role	'taken from my body', *implicit for health care provider who collects the specimen or tissues*
OBI:0000245	Organization	'XXX'
OMRSE:00000054	hospital role	'XXX'
NCBITaxon:9606	Homo Sapiens	'I', 'my', *implicit for health care provider who collects the specimen or tissues*
ICO:0000196	Act of Informed Consenting	'I hereby authorize'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000244	Permission Directive	*entire sentence*
OBI:0100051	Specimen	'any specimen or tissues'
ICO:0000060	Act of storing a specimen	'to retain'
NCIT:C64262	Biologic Sample Preservation Procedure	
OBI:0000066	Investigation	'use for scientific... purposes'
OMRSE:00002029	education process	'use for... teaching purposes'
CHMO:0002832	Material disposal	'to dispose'
OBI:0000659	Specimen Collection Process	'taken from my body'
--	Temporal Directive	'during my visit'
ICO:0000186	Permission Temporal Region	'during my visit'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
OBI:0000295	is:specified:input:of	
OBI:0000299	has:specified:output	
RO:0000053	bearer of	
RO:0000056	Participates in	
RO:0002350	Member of	
Note: *XXX denotes a health care facility which voluntarily contributed consent form(s) for these studies.*		

Table 26. Terms and Sentence References to Express the Permission-Sentence: "I am a New York state resident and I give permission for GeneDx to retain any remaining sample longer than 60 days after completion of testing and use my de-identified data for scientific and medical research purposes."

I am a New York state resident and I give permission for GeneDx to retain any remaining sample longer than 60 days after completion of testing and use my de-identified data for scientific and medical research purposes.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'I give permission to GeneDx'
OBI:0000093	Patient Role	'any remaining sample', 'my deidentified data'
ICO:0000086	Consenter Role	'I give permission'
OBI:0000207	health care provider role	*Implicit health care provider who collects the sample*
VO:0000299	Company	'GeneDx"
NCBITaxon:9606	Homo Sapiens	'I', 'my', *implicit health care provider*
OBI:0000659	Specimen Collection Process	*Implicit in process by which a clinically collected specimen becomes a residual clinical biospecimen*
OBI:0100051	Specimen	'sample'
LABO:0000107	Laboratory Test	*Implicit in process by which a clinically collected specimen becomes a residual clinical biospecimen*
OGMS:0000018	Laboratory Finding	*Implicit in process by which a clinically collected specimen becomes a residual clinical biospecimen*
ICO:0000313	Excess Material Role	'remaining sample'
--	Residual Clinical Biospecimen	'remaining sample'
ICO:0000218	Act of Removing Identifiable Information	'de-identified data'
ICO:0000113	Anonymized Information Content Entity	'de-identified data'
ICO:0000060	Act of Storing a Specimen	'to retain any remaining sample'
ICO:0000196	act of informed consenting	'I give permission'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000244	Permission Directive	*Entire Sentence*
--	Temporal Directive	'give permission to... retain.. longer than 60 days'
ICO:0000186	Permission Temporal Region	'longer than 60 days'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
DUO:0000010	Is restricted to	
OBI:0000293	has:specified:input	
RO:0000053	bearer of	
RO:0000056	Participates in	
RO:0002350	Member of	

Table 27. Terms and Sentence References to Express the Permission-Sentence: "I also consent to diagnostic studies, tests, anesthesia, x-ray examinations and any other treatment or courses of treatment relating to the diagnosis or procedure described herein."

I also consent to diagnostic studies, tests, anesthesia, x-ray examinations and any other treatment or courses of treatment relating to the diagnosis or procedure described herein.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
OBI:0000093	Patient Role	'I', 'diagnostic studies, tests, anesthesia, x-ray examinations and any other treatment or courses of treatment"
ICO:0000086	Consenter Role	'I also consent to'
NCBITaxon:9606	Homo Sapiens	'I'
OGMS:0000096	Health Care Process	'diagnostic studies, tests, anesthesia, x-ray examinations and any other treatment or courses of treatment'
ICO:0000196	act of informed consenting	'I also consent to'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000245	Conditional Permission Directive	'consent to... relating to the diagnosis or procedure described herein'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 28. Terms and Sentence References to Express the Permission-Sentence: "I DONATE and authorize XXX to own, use, retain, preserve, manipulate, analyze, or dispose of any excess tissues, specimens, or parts of organs that are removed from my body during the procedures described above and are not necessary for my diagnosis or treatment."

I DONATE and authorize XXX to own, use, retain, preserve, manipulate, analyze, or dispose of any excess tissues, specimens, or parts of organs that are removed from my body during the procedures described above and are not necessary for my diagnosis or treatment.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'I... authorize XXX'
OBI:0000093	Patient role	'I', 'my diagnosis or treatment'
ICO:0000086	Consenter Role	'I DONATE and authorize'
OMRSE:00000056	Hospital organization	'XXX'
NCBITaxon:9606	Homo Sapiens	'I', 'my'
ICO:0000245	Conditional Permission Directive	'I... authorize... and are not necessary for my diagnosis or treatment.'
ICO:0000196	Act of Informed Consenting	'I DONATE and authorize'
ICO:0000001	Informed consent form	*Implicit*
--	Owner role	'to own'
--	Owner role directive	'I DONATE and authorize XXX to own'
--	Ownership process	'to own'
--	Residual clinical biospecimen	'excess tissues, specimens, …organs'
OBI:0000659	Specimen Collection Process	'removed from my body'
OBI:0100051	Specimen	'excess tissues, specimens, or parts of organs'
LABO:0000107	Laboratory Test	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
OGMS:0000018	Laboratory Finding	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
ICO:0000313	Excess Material Role	'excess tissues, specimens, …organs'
NCIT:C95018	Use	'use'
ICO:0000060	Act of storing a specimen	'retain'
NCIT:C64262	Biologic Sample Preservation Procedure	'preserve'
XCO:0000527	Genetic manipulation	'manipulate'
OBI:0000066	Investigation	'analyze'
CHMO:0002832	Material disposal	'dispose of'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
OBI:0000293	has:specified:input	
OBI:0000295	is:specified:input:of	
RO:0000053	bearer of	
RO:0000056	Participates in	
Note: *XXX denotes a health care facility which voluntarily contributed consent form(s) for these studies.*		

Table 29. Terms and Sentence References to Express the Permission-Sentence: "I agree that any excess tissue, fluids or specimens removed from my body during my outpatient visit or hospital stay (my specimens) that would otherwise be disposed of by the Hospital may be used for such educational purposes and research, including research on the genetic materials (DNA)."

I agree that any excess tissue, fluids or specimens removed from my body during my outpatient visit or hospital stay (my specimens) that would otherwise be disposed of by the Hospital may be used for such educational purposes and research, including research on the genetic materials (DNA).		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'I agree... by the hospital'
OBI:0000093	Patient Role	'during my outpatient visit or hospital stay'
ICO:0000086	Consenter Role	'I agree that'
OMRSE:00000056	Hospital organization	'the hospital'
NCBITaxon:9606	Homo Sapiens	'I', 'my'
ICO:0000196	act of informed consenting	'I agree that'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000244	Permission Directive	*Entire sentence*
OBI:0000659	Specimen Collection Process	'removed from my body'
OBI:0100051	Specimen	'tissue, fluids or specimens'
LABO:0000107	Laboratory Test	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
OGMS:0000018	Laboratory Finding	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
ICO:0000313	Excess Material Role	'excess tissue, fluids or specimens'
--	Residual clinical biospecimen	'excess tissue, fluids or specimens'
OMRSE:00002029	education process	'educational purposes'
OBI:0000066	Investigation	'research'
DUO:0000038	genetic research	'research, including research on the genetic materials (DNA).'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
OBI:0000295	is:specified:input:of	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 30. Terms and Sentence References to Express the Permission-Sentence: "I authorize the pathologist, at his or her discretion, to retain, preserve, use, or dispose of any tissues, organs, bones, bodily fluid or medical devices that may be removed during the operation(s) or procedure(s)."

I authorize the pathologist, at his or her discretion, to retain, preserve, use, or dispose of any tissues, organs, bones, bodily fluid or medical devices that may be removed during the operation(s) or procedure(s).		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'I authorize the pathologist'
OBI:0000093	Patient role	'removed during the operation(s) or procedure(s)'
ICO:0000086	Consenter Role	'I authorize'
OBI:0000145	Pathologist role	'the pathologist'
NCBITaxon:9606	Homo Sapiens	'I', 'the pathologist'
ICO:0000196	Act of Informed Consenting	'I authorize'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000244	Permission Directive	*Entire sentence*
OBI:0000659	Specimen Collection Process	'removed during the operation(s) or procedure(s)'
OBI:0100051	Specimen	'tissues, organs, bones, bodily fluid or medical devices'
LABO:0000107	Laboratory Test	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
OGMS:0000018	Laboratory Finding	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
ICO:0000313	Excess Material Role	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
--	Residual clinical biospecimen	*Implicit in process by which a specimen becomes a residual clinical biospecimen*
ICO:0000060	Act of storing a specimen	'to retain'
NCIT:C64262	Biologic Sample Preservation Procedure	'preserve'
NCIT:C95018	Use	'use'
CHMO:0002832	Material disposal	'dispose'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
DUO:0000010	Is restricted to	
OBI:0000293	has:specified:input	
OBI:0000295	is:specified:input:of	
OBI:0000299	has:specified:output	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 31. Terms and Sentence References to Express the Permission-Sentence: "I, , request and consent to the start or induction of my labor by my provider: and other assistants as may be selected by him/her."

I, , request and consent to the start or induction of my labor by my provider: and other assistants as may be selected by him/her.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'I, , request and consent... by my provider'
OBI:0000093	Patient Role	'I', 'induction of my labor'
ICO:0000086	Consenter Role	'I, , request and consent'
OBI:0000207	health care provider role	'by my provider: and other assistants...'
NCBITaxon:9606	Homo Sapiens	'I', 'my provider', 'assistants'
OGMS:0000096	Health Care Process	'start or induction of my labor'
ICO:0000196	act of informed consenting	'I, , request and consent'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000244	Permission Directive	*Entire Sentence*
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 32. Terms and Sentence References to Express the Permission-Sentence: "I voluntarily consent to receive medical and health care services that may include diagnostic procedures, examination, and treatment."

I voluntarily consent to receive medical and health care services that may include diagnostic procedures, examination, and treatment.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000086	Consenter Role	'consent to receive medical and health care services'
NCBITaxon:9606	Homo Sapiens	'I', *implicit health care provider(s) who provides the health services*
ICO:0000244	Permission Directive	*Entire Sentence*
ICO:0000196	act of informed consenting	'I voluntarily consent'
ICO:0000001	Informed consent form	*Implicit*
OGMS:0000096	Health Care Process	'health care services that may include diagnostic procedures, examination, and treatment'
OGMS:0000104	Diagnostic Process	'diagnostic procedures'
OBI:0600003	performing a clinical assessment	'examination'
OGMS:0000090	Treatment	'treatment'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 33. Terms and Sentence References to Express the Permission-Sentence: "I consent to the use of closed-circuit television, taking of photographs (including videos), and the preparation of drawings and similar illustrative graphic material for scientific purposes providing my identity is not revealed."

I consent to the use of closed-circuit television, taking of photographs (including videos), and the preparation of drawings and similar illustrative graphic material for scientific purposes providing my identity is not revealed.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'consent to the use of'
OBI:0000093	Patient Role	'I', *inferred from context of clinical consent form*
ICO:0000086	Consenter Role	'I consent to'
NCBITaxon:9606	Homo Sapiens	I', *implicit person(s) who take the photos, videos, etc.*
ICO:0000196	act of informed consenting	'I consent to'
ICO:0000001	Informed consent form	*Implicit*
ICO:0000244	Permission Directive	*Entire Sentence*
IAO:0000572	Documenting	'use of closed-circuit television, taking of photographs (including videos), and the preparation of drawings and similar illustrative graphic material'
BFO:0000040	Material Entity	'photographs', 'videos', 'illustrative graphic material'
IAO:0000030	Information Content Entity	'photographs', 'videos', 'illustrative graphic material'
IAO:0000185	Photograph	'photographs'
ERO:0000333	Video	'videos'
FBbi:00000223	Graphic Illustration	'illustrative graphic material'
DUO:0000001	Data Use Limitation	'for scientific purposes providing my identity is not revealed'
OBI:0000066	Investigation	'scientific purposes'
ICO:0000218	Act of Removing Identifiable Information	providing my identity is not revealed'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
OBI:0000295	is:specified:input:of	
OBI:0000299	has:specified:output	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 34. Terms and Sentence References to Express the Permission-Sentence: "I hereby consent to engaging in virtual health/telemedicine services, where available, as part of my treatment."

I hereby consent to engaging in virtual health/telemedicine services, where available, as part of my treatment.		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000086	Consenter Role	'hereby consent to engaging in virtual health/telemedicine services'
NCBITaxon:9606	Homo Sapiens	'I', *implicit person(s) are providing telehealth services*
ICO:0000244	Permission Directive	*Entire Sentence*
ICO:0000196	act of informed consenting	'I hereby consent to'
ICO:0000001	Informed consent form	Implicit
NCIT:C15380	Telemedicine	'virtual health/telemedicine services'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
RO:0000053	bearer of	
RO:0000056	Participates in	

Table 35. Terms and Sentence References to Express the Permission-Sentence: "I request and authorize URMC Labs to perform only the above designated test(s) on the sample from me (or my child or fetus)."

I request and authorize URMC Labs to perform only the above designated test(s) on the sample from me (or my child or fetus).		
Ontology PURL	**Ontology Label**	**Sentence Reference**
Classes		
ICO:0000199	Permission Role	'request and authorize URMC Labs'
ICO:0000086	Consenter Role	'I request and authorize'
OBI:0000245	Organization	'URMC Labs'
NCBITaxon:9606	Homo Sapiens	'I', 'sample from me (or my child or fetus)'
ICO:0000244	Permission Directive	*Entire Sentence*
LABO:0000107	Laboratory Test	'perform only the above designated test(s)'
ICO:0000196	act of informed consenting	'I request and authorize'
ICO:0000001	Informed consent form	*Implicit*
NCIT:C16423	Child	'child'
NCIT:C13235	Fetus	'fetus'
Relations		
BFO:0000051	has part	
BFO:0000054	Realized in	
CommonCore Ontologies/prescribes	prescribes	
RO:0000053	bearer of	
RO:0000056	Participates in	

Bibliography

1. Gray J, Grove SK, Sutherland S. The practice of nursing research: Appraisal, synthesis, and generation of evidence. 8th ed. St. Louis, Missouri: Elsevier; 2017.

2. Consent. In: Tabers Cycolpedic Medical Dictionary. 21st ed. Philadelphia: F. A. Davis; 2009.

3. Consent Form. In: Tabers Cycolpedic Medical Dictionary. 21st ed. Philadelphia: F. A. Davis; 2009.

4. Allen MJ, Powers MLE, Gronowski KS, Gronowski AM. Human Tissue Ownership and Use in Research: What Laboratorians and Researchers Should Know. Clin Chem. 2010 Nov 1;56(11):1675–82.

5. Poste G. Biospecimens, biomarkers, and burgeoning data: the imperative for more rigorous research standards. Trends Mol Med. 2012 Dec;18(12):717–22.

6. Clayton EW, Halverson CM, Sathe NA, Malin BA. A systematic literature review of individuals' perspectives on privacy and genetic information in the United States. PLOS ONE. 2018 Oct 31;13(10):e0204417.

7. Lee SS-J, Cho MK, Kraft SA, Varsava N, Gillespie K, Ormond KE, et al. "I Don't Want to be Henrietta Lacks": Diverse Patient Perspectives on Donating Biospecimens for Precision Medicine Research. Genet Med Off J Am Coll Med Genet. 2019 Jan;21(1):107–13.

8. Vaught J. Developments in biospecimen research. Br Med Bull. 2015;114:29–38.

9. Landry LG, Ali N, Williams DR, Rehm HL, Bonham VL. Lack Of Diversity In Genomic Databases Is A Barrier To Translating Precision Medicine Research Into Practice. Health Aff (Millwood). 2018 May;37(5):780–5.

10. Blasimme A, Fadda M, Schneider M, Vayena E. Data Sharing For Precision Medicine: Policy Lessons And Future Directions. Health Aff (Millwood). 2018 May;37(5):702–9.

11. hhs.gov. Updated FAQs Informed Consent for Use of Biospecimens [Internet]. HHS.gov. 2018 [cited 2019 Feb 14]. Available from: https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-faqs-recommendations-and-glossary-informed-consent-and-research-use-of-biospecimens-and-associated-data/index.html

12. Henderson GE, Cadigan RJ, Edwards TP, Conlon I, Nelson AG, Evans JP, et al. Characterizing biobank organizations in the U.S.: results from a national survey. Genome Med. 2013;5(1):3.

13. Blanc V. Intro to Biobanking: Laboratory Information Management System. 2018.

14. Manion FJ. A Life Cycle Approach to the Development and Validation of an Ontology of the U.S. Common Rule (45 C.F.R. § 46). [Houston, TX]: University of Texas Health Science Center at Houston; 2017.

15. The Office of the National Coordinator for Health Information Technology. Patient Consent for Electronic Health Information Exchange and Interoperability [Internet]. HealthIT.gov. 2019. Available from: https://www.healthit.gov/topic/interoperability/patient-consent-electronic-health-information-exchange-and-interoperability

16. Kock-Schoppenhauer A-K, Kamann C, Ulrich H, Duhm-Harbeck P, Ingenerf J. Linked Data Applications Through Ontology Based Data Access in Clinical Research. Stud Health Technol Inform. 2017;131–135.

17. Noy NF, Mcguinness DL. Ontology development 101: A guide to creating your first ontology. 2001.

18. The OBO Foundry [Internet]. 2018 [cited 2018 Dec 14]. Available from: http://www.obofoundry.org/

19. 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 1;10(4):343–8.

20. Lin Y, Harris MR, Manion FJ, Eisenhauer E, Zhao B, Shi W, et al. Development of a BFO-based Informed Consent Ontology (ICO). In: ICBO Conference Proceedings. Houston; 2014. p. 3.

21. Neuhaus F, Ray S, Sriram RD. Toward ontology evaluation across the life cycle. Gaithersburg, MD: National Institute of Standards and Technology; 2014. Report No.: NIST IR 8008.

22. Edwards T, Cadigan RJ, Evans JP, Henderson GE. Biobanks containing clinical specimens: Defining characteristics, policies, and practices. Clin Biochem. 2014 Mar 1;47(4):245–51.

23. Center for Devices and Radiological Health. Clinical Laboratory Improvement Amendments (CLIA). FDA [Internet]. 2019 Jul 3 [cited 2020 Feb 16]; Available from: http://www.fda.gov/medical-devices/ivd-regulatory-assistance/clinical-laboratory-improvement-amendments-clia

24. Ploug T, Holm S. The biobank consent debate: why 'meta-consent' is still the solution! J Med Ethics. 2019 Mar 14;medethics-2018-105258.

25. HHS.gov. Revised Common Rule Q&As [Internet]. Office for Human Research Protections. 2018 [cited 2019 Mar 25]. Available from: https://www.hhs.gov/ohrp/education-and-outreach/revised-common-rule/revised-common-rule-q-and-a/index.html

26. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping Reviews: Time for Clarity in Definition, Methods, and Reporting. J Clin Epidemiol. 2014 Dec;67(12):1291–4.

27. O'Brien KK, Colquhoun H, Levac D, Baxter L, Tricco AC, Straus S, et al. Advancing scoping study methodology: A web-based survey and consultation of perceptions on terminology, definition and methodological steps. BMC Health Serv Res. 2016;16(1).

28. Rosati KB. Human Genetic Sampling and the HIPAA Privacy Standards. Jurimetrics. 2005;45:251.

29. Rothstein MA. The End of the HIPAA Privacy Rule? J Law Med Ethics. 2016;44(2):352–8.

30. Hsu J. Genetic Testing: Balancing Preventative Medicine with Privacy and Nondiscrimination. J Law Policy Inf Soc. 2010 2011;6:557.

31. Hudson KL. Genomics, Health Care, and Society. Feero WG, Guttmacher AE, editors. N Engl J Med. 2011 Sep 15;365(11):1033–41.

32. Ireni-Saban L. Genomics Governance in the United States and the United Kingdom. Eur J Comp Law Gov. 2014;1:244.

33. Harrell HL, Rothstein MA. Biobanking Research and Privacy Laws in the United States. J Law Med Ethics. 2016;44:106.

34. Bierer BE, Barnes M, Lynch HF. Revised 'Common Rule' Shapes Protections for Research Participants. Health Aff (Millwood). 2017 May 1;36(5):784–8.

35. Lynch HF, Meyer MN. Regulating Research with Biospecimens under the Revised Common Rule. Hastings Cent Rep. 2017 May;47(3):3–4.

36. Lynch HF. A Functional Approach to Assessing Consent for Biospecimen Research. Am J Bioeth. 2017 Dec;17(12):20–3.

37. Beskow LM. Lessons from HeLa Cells: The Ethics and Policy of Biospecimens. Annu Rev Genomics Hum Genet. 2016;17:395–417.

38. Bledsoe MJ. Ethical Legal and Social Issues of Biobanking: Past, Present, and Future. Biopreservation Biobanking. 2017 Apr 1;15(2):142–7.

39. Rothstein MA. Is Deidentification Sufficient to Protect Health Privacy in Research? Am J Bioeth. 2010;10(9):3–11.

40. Evans BJ. Seven Pillars of a New Evidentiary Paradigm: The Food, Drug, and Cosmetic Act Enters the Genomic Era. Notre Dame Law Rev. 2009 2010;85:0.

41. Ahn S. Whose Genome Is It Anyway: Re-identification and Privacy Protection in Public and Participatory Genomics. San Diego Law Rev. 2015;52:[i]-806.

42. Wolf LE. Risks and Legal Protections in the World of Big-Data. Asia Pac J Health Law Ethics. 2017 2018;11:1.

43. Maschke KJ. Navigating an ethical patchwork—human gene banks. Nat Biotechnol. 2005 May;23(5):539.

44. Shea GR. To Ends the Most Public and Universal: An Overview of the HIPAA Privacy Rule. Labor Law J. 2003;54:22.

45. Maschke KJ. Wanted: Human Biospecimens. Hastings Cent Rep. 2010 Sep;40(5):21–3.

46. Newborn Screening Saves Lives Reauthorization Act Introduced in House of Representatives [Internet]. Parent Project Muscular Dystrophy. 2019 [cited 2020 Mar 27]. Available from: https://www.parentprojectmd.org/newborn-screening-saves-lives-reauthorization-act-introduced-in-house-of-representatives/

47. Edwards LC. Tissue Tug-of-War: A Comparison of International and U.S. Perspectives on the Regulation of Human Tissue Banks. Vanderbilt J Transnatl Law. 2008;41:639.

48. Rose S. International Ethical Guidelines for Epidemiological StudiesBy the Council for International Organizations of Medical Sciences (CIOMS). Am J Epidemiol. 2009 Dec 1;170(11):1451–2.

49. Thorogood A, Zawati MH. International Guidelines for Privacy in Genomic Biobanking (or the Unexpected Virtue of Pluralism). J Law Med Ethics. 2015 Winter;43(4):690–702.

50. Dove ES. Biobanks, Data Sharing, and the Drive for a Global Privacy Governance Framework. J Law Med Ethics. 2015;43(4):675–89.

51. Rothstein MA, Knoppers BM, Harrell HL. Comparative Approaches to Biobanks and Privacy. J Law Med Ethics. 2016;44(1):161.

52. Campbell LD, Betsou F, Giri JG, Pitt KE, Pugh RS, Sexton KC, et al. 2012 Best Practices for Repositories: Collection, Storage, Retrieval, and Distribution of Biological Materials for Research. Biopreservation Biobanking. 2012 Apr;10(2):79–161.

53. Organization for Economic Cooperation and Development. Guidelines for Human Biobanks and Genetic Research Databases (HBGRDs) - OECD [Internet]. 2009. Available from: http://www.oecd.org/sti/emerging-tech/guidelines-for-human-biobanks-and-genetic-research-databases.htm

54. U.S. Equal Employment and Opportunity Commission. Prohibited Employment Policies/Practices [Internet]. Available from: https://www.eeoc.gov/laws/practices/

55. ten Have AMJ, Jean MS. UNESCO Universal Declaration on Bioethics and Human Rights: Background, Principles and Application. UNESCO; 2009. 371 p.

56. United Nations Educational, Scientific, and Cultural Organization. Universal Declaration on the Human Genome and Human Rights [Internet]. UNESCO. 2019. Available from: https://en.unesco.org/themes/ethics-science-and-technology/human-genome-and-human-rights

57. United Nations Educational, Scientific, and Cultural Organization. Report of the International Bioethics Committee of UNESCO (IBC) on Consent [Internet]. 2008. Report No.: SHS/EST/CIB08-09/2008/1. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000178124

58. World Medical Association. Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks [Internet]. Available from: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

59. Dawson L. Common Rule Revised: Opportunities Lost. Am J Bioeth. 2017 Jul;17(7):46–8.

60. Bregman-Eschet Y. Genetic Databases and Biobanks: Who Controls Our Genetic Privacy. 2012;23:55.

61. Shayeb TY. You Are What You Own: Reopening the Discussion on Universally Recognizing a Property Right in Genetic Information and Material. Whittier Law Rev. 2017 2018;38:181.

62. Spector-Bagdady K, Fernandez Lynch H, Brenner JC, Shuman AG. Biospecimens, Research Consent, and Distinguishing Cell Line Research. JAMA Oncol. 2019 Mar 1;5(3):406.

63. Dresser R. Research Information for Reasonable People. Hastings Cent Rep. 2018 Nov 1;48(6):3–4.

64. Atreya RV, Smith JC, McCoy AB, Malin B, Miller RA. Reducing patient re-identification risk for laboratory results within research datasets. J Am Med Inform Assoc. 2013 Jan 1;20(1):95–101.

65. Rivera SM. Reasonable Research Oversight: A Work in Progress. IRB Ethics Hum Res. 2017 Dec 11;39(6):15–9.

66. Fisher CB, Layman DM. Genomics, Big Data, and Broad Consent: a New Ethics Frontier for Prevention Science. Prev Sci. 2018 Oct;19(7):871–9.

67. Kiehntopf M, Krawczak M. Biobanking and international interoperability: samples. Hum Genet. 2011 Sep;130(3):369–76.

68. Eder J, Gottweis H, Zatloukal K. IT Solutions for Privacy Protection in Biobanking. Public Health Genomics. 2012;15(5):254–62.

69. Cocanour CS. Informed consent—It's more than a signature on a piece of paper. Am J Surg. 2017 Dec;214(6):993–7.

70. Paterick TJ, Carson GV, Allen MC, Paterick TE. Medical Informed Consent: General Considerations for Physicians. Mayo Clin Proc. 2008 Mar 1;83(3):313–9.

71. American Medical Association. Informed Consent [Internet]. American Medical Association. 2020 [cited 2020 Jan 4]. Available from: https://www.ama-assn.org/delivering-care/ethics/informed-consent

72. Grando A, Schwab R. Building and evaluating an ontology-based tool for reasoning about consent permission. In: AMIA Annu Symp Proc. Washington, DC; 2013. p. 514–523.

73. Eder J, Gottweis H, Zatloukal K. IT Solutions for Privacy Protection in Biobanking. Public Health Genomics. 2012;15(5):254–62.

74. Wittenburg P. About Annotation Schemes and Terminology. In Athens, Greece; 2000.

75. Hovy E, Lavid J. Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. Int J Transl. 2010;22(1):25.

76. Polit DF, Beck CT. Essentials of Nursing Research. 9th Edition. Philadelphia: Wolters Kluwer; 2018.

77. Centers for Medicare & Medicaid Services. Hospital General Information [Internet]. Data.gov. 2019 [cited 2019 Nov 7]. Available from: https://catalog.data.gov/dataset/hospital-general-information

78. Centers for Medicare & Medicaid Services. Ambulatory Surgical Quality Measures – Facility [Internet]. Data.gov. 2019 [cited 2019 Nov 7]. Available from: https://catalog.data.gov/dataset/ambulatory-surgical-measures-facility

79. CTSA Program Hubs [Internet]. National Center for Advancing Translational Sciences. 2015 [cited 2019 Nov 26]. Available from: https://ncats.nih.gov/ctsa/about/hubs

80. Pustejovsky J, Bunt H, Zaenen A. Designing Annotation Schemes: From Theory to Model. In: Ide N, Pustejovsky J, editors. Handbook of Linguistic Annotation. Dordrecht: Springer Netherlands; 2017. p. 21–72.

81. Vajda JM, Otte JN, Stansbury C, Harris MR, Umberfield E, Manion FJ, et al. Expanding the Representation of Permissions and Deontic Roles in the Informed Consent Ontology (ICO). In: AMIA Knowledge Center. Washington D.C.; 2019.

82. McNamara P. Deontic Logic. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy [Internet]. Summer 2019. Metaphysics Research Lab, Stanford University; 2019 [cited 2020 Mar 11]. Available from: https://plato.stanford.edu/archives/sum2019/entries/logic-deontic/

83. Honnibal M, Montani I. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. 2017.

84. Ratclif JW. Pattern Matching: the Gestalt Approach [Internet]. Dr. Dobb's. [cited 2020 May 11]. Available from: http://www.drdobbs.com/database/pattern-matching-the-gestalt-approach/184407970

85. Artstein R. Inter-annotator Agreement. In: Ide N, Pustejovsky J, editors. Handbook of Linguistic Annotation [Internet]. Dordrecht: Springer Netherlands; 2017 [cited 2019 Aug 19]. p. 297–313. Available from: http://link.springer.com/10.1007/978-94-024-0881-2_11

86. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull. 1968 Oct;70(4):213–20.

87. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org/

88. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159–74.

89. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155–63.

90. Hall DE, Prochazka AV, Fink AS. Informed consent for clinical treatment. Can Med Assoc J. 2012 Mar 20;184(5):533–40.

91. Bowen GA. Document analysis as a qualitative research method. Qual Res J. 2009;9(2).

92. Rao KHS. Informed Consent: An Ethical Obligation or Legal Compulsion? J Cutan Aesthetic Surg. 2008 Jan;1(1):33.

93. 45 CFR 46 [Internet]. Available from: https://www.hhs.gov/ohrp/sites/default/files/ohrp/policy/ohrpregulations.pdf

94. Lynch HF, Largent EA, Zarin DA. Reaping the Bounty of Publicly Available Clinical Trial Consent Forms. IRB. 2017;39(6):10–5.

95. Tomlinson T. Respecting donors to biobank research. Hastings Cent Rep. 2013 Feb;43(1):41–7.

96. Garrison NA, Sathe NA, Antommaria AHM, Holm IA, Sanderson SC, Smith ME, et al. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genet Med Off J Am Coll Med Genet. 2016 Jul;18(7):663–71.

97. Dickinson M, Tufiş D. Iterative Enhancement. In: Ide N, Pustejovsky J, editors. Handbook of Linguistic Annotation. Dordrecht: Springer Netherlands; 2017. p. 257–76.

98. Pustejovsky J, Stubbs A. The Basics. In: Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications. O'Reilly Media, Inc.; 2012.

99. Eltorai AEM, Naqvi SS, Ghanian S, Eberson CP, Weiss A-PC, Born CT, et al. Readability of Invasive Procedure Consent Forms. Clin Transl Sci. 2015;8(6):830–3.

100. The Joint Commission, Division of Healthcare Improvement. Informed Consent: More than getting a Signature [Internet]. The Joint Commission; 2016 Feb [cited 2020 Jun 9]. (Quick Safety). Report No.: Issue 21. Available from: https://www.jointcommission.org/-/media/deprecated-unorganized/imported-assets/tjc/system-folders/joint-commission-online/quick_safety_issue_twenty-one_february_2016pdf.pdf?db=web&hash=5944307ED39088503A008A70D2C768AA

101. ISBER. Basics of Biobanking [Internet]. Intro to Biobanking. 2018. Available from: https://www.isber.org/page/IntrotoBiobanking

102. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018.

103. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007 Nov;25(11):1251–5.

104. Dumontier M, Hoehndorf R. Realism for scientific ontologies. In: Proceedings of the 2010 Conference on Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010) [Internet]. Amsterdam, The Netherlands, The Netherlands: IOS Press; 2010 [cited 2018 Dec 13]. p. 387–399. Available from: http://dl.acm.org/citation.cfm?id=1804715.1804755

105. Arp R, Smith B, Spear AD. Building Ontologies with Basic Formal Ontology. MIT Press; 2015. 245 p.

106. Basic Formal Ontology (BFO) [Internet]. 2017 [cited 2018 Dec 14]. Available from: http://basic-formal-ontology.org/

107. He Y. Informed Consent Ontology [Internet]. ICO-ontology; 2019 [cited 2018 Dec 14]. Available from: https://github.com/ICO-ontology/ICO

108. Informed Consent Ontology - Summary [Internet]. NCBO BioPortal. 2019 [cited 2020 Jun 7]. Available from: https://bioportal.bioontology.org/ontologies/ICO

109. Burton-Jones A, Storey VC, Sugumaran V, Ahluwalia P. A semiotic metrics suite for assessing the quality of ontologies. Data Knowl Eng. 2005 Oct;55(1):84–102.

110. Gelernter J, Jha J. Challenges in Ontology Evaluation. J Data Inf Qual. 2016 Aug 22;7(3):1–4.

111. Obrst L, Ceusters W, Mani I, Ray S, Smith B. The Evaluation of Ontologies. In: Baker CJO, Cheung K-H, editors. Semantic Web [Internet]. Boston, MA: Springer US; 2007 [cited 2019 Jan 9]. p. 139–58. Available from: http://link.springer.com/10.1007/978-0-387-48438-9_8

112. Friedman CP, Wyatt JC. Evaluation as a Field. In: Evaluation Methods in Medical Informatics [Internet]. 2nd ed. New York: Springer-Verlag; 2006 [cited 2020 Jun 6]. (Computers and Medicine). Available from: https://www.springer.com/gp/book/9781475726855

113. Raad J, Cruz C. A Survey on Ontology Evaluation Methods: In: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management [Internet]. Lisbon, Portugal: SCITEPRESS - Science and and Technology Publications; 2015 [cited 2019 Jan 17]. p. 179–86. Available from: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005591001790186

114. Principles: Overview [Internet]. The OBO Foundry. [cited 2020 Jun 1]. Available from: http://www.obofoundry.org/principles/fp-000-summary.html

115. Zhu X, Fan J-W, Baorto DM, Weng C, Cimino JJ. A review of auditing methods applied to the content of controlled biomedical terminologies. J Biomed Inform. 2009 Jun 1;42(3):413–25.

116. Sell. In: Merriam-Webster [Internet]. wowo [cited 2020 Aug 14]. Available from: https://www.merriam-webster.com/dictionary/sell

117. HIPAA Administrative Simplification Regulation Text. U.S. Department of Health and Human Services Office for Civil Rights; 2013 Mar. Report No.: 45 CFR Parts 160, 162, and 164.

118. Article 4 EU GDPR “Definitions” [Internet]. PrivazyPlan. SecureDataService; 2020. Available from: https://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm

119. Prospective Study. In: Encyclopedia of Research Design [Internet]. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.; 2010 [cited 2020 Aug 14]. Available from: http://methods.sagepub.com/reference/encyc-of-research-design/n342.xml

120. Harris MR. Methods for Ontology Evaluation. 2020.

121. Harris MR, Langford LH, Miller H, Hook M, Dykes PC, Matney SA. Harmonizing and extending standards from a domain-specific and bottom-up approach: an example from

development through use in clinical applications. J Am Med Inform Assoc JAMIA. 2015 May;22(3):545–52.

122. American Medical Informatics Association. Clinical Research Informatics [Internet]. AMIA. 2020 [cited 2020 Jun 1]. Available from: https://www.amia.org/applications-informatics/clinical-research-informatics

123. Skloot R. Opinion | Your Cells. Their Research. Your Permission? - The New York Times. The New York Times. 2015 Dec 30;A23.

124. Ginsburg GS, Phillips KA. Precision Medicine: From Science To Value. Health Aff (Millwood). 2018 May;37(5):694–701.

125. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017 Jan 4;45(D1):D896–901.

126. Campbell LD, Astrin JJ, Brody R, De Souza Y, Giri JG, Ashokkumar AP, et al., editors. Best Practices: Recommendations for Repositories [Internet]. International Society for Biological and Environmental Repositories; 2018. Available from: https://www.isber.org/page/BPR

127. ISBER. Module 9. Data Systems and Records Management [Internet]. Intro to Biobanking. 2018. Available from: https://www.isber.org/page/IntrotoBiobanking

128. Berners-Lee T, Hendler J, Lassila O. The Semantic Web. Sci Am. 2001 May;4.

129. Kwasnik BH. The Role of Classification in Knowledge Representation and Discovery. Libr Trends. 1999;48(1):22–47.

130. Millerand F, Bowker GC. Metadata standards: trajectories and enactment in the life of an ontology. In: Star SL, Lapland M, editors. Formalizing Practices: Reckoning with Standards, Numbers and Models in Science and Everyday Life. 1st Edition. Ithaca, NY: Cornell University Press; 2008. (Cornell Paperbacks).

131. Arp R, Smith B, Spear AD. Kinds of Ontologies and the Role of Taxonomies. In: Building Ontologies with Basic Formal Ontology. MIT Press; 2015.

132. Allemang D, Hendler M. Chapter 3. RDF- The basis of the Semantic Web. In: Semantic Web for the Working Ontologist. San Francisco: Morgan Kaufmann; 2008. p. 31–55.

133. Beghtol C. Semantic validity: concepts of warrant in bibliographic classification systems. Libr Resources Tech Serv. 1986;30:109–25.

134. Shadbolt N, Berners-Lee T, Hall W. The Semantic Web Revisited. IEEE Intell Syst. 2006 May;21(3):96–101.

135. Shaw M, Detwiler LT, Brinkley JF, Suciu D. Generating Application Ontologies from Reference Ontologies. AMIA Annu Symp Proc. 2008;5.

136. Arp R, Smith B, Spear AD. What Is an Ontology? In: Building Ontologies with Basic Formal Ontology. MIT Press; 2015.

137. Obeid J, Gabriel D, Sanderson I. A Biomedical Research Permissions Ontology: Cognitive and Knowledge Representation Considerations. Proc Gov Technol Inf Policies Addressing Chall Worldw Interconnectivity Workshop Gov Technol Inf Policies 2010 Austin Tex. 2010 Dec;2010:9–13.

138. Fransson MN, Rial-Sebbag E, Brochhausen M, Litton J-E. Toward a common language for biobanking. Eur J Hum Genet. 2015 Jan;23(1):22–8.

139. 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.

140. Dyke SOM, Philippakis AA, Rambla De Argila J, Paltoo DN, Luetkemeier ES, Knoppers BM, et al. Consent Codes: Upholding Standard Data Use Conditions. Barsh GS, editor. PLOS Genet. 2016 Jan 21;12(1):e1005772.

141. FHIR. 7.15.0 Appendix: The Relationship between FHIR and other HL7 Standards [Internet]. FHIR Release 3. 2017 [cited 2018 Dec 14]. Available from: https://www.hl7.org/fhir/comparison.html

142. FHIR. 2.11 FHIR overview [Internet]. FHIR Release 3. 2017 [cited 2018 Dec 14]. Available from: https://www.hl7.org/fhir/overview.html

143. FHIR. 6.2 Resource Consent - Content [Internet]. FHIR Release 3. 2017 [cited 2018 Dec 13]. Available from: https://www.hl7.org/fhir/consent.html

144. ONC Announces Interest in Applications to Address Standardization of Patient Information for Seamless Access, Exchange, and Use [Internet]. Office of the National Coordinator for Health Information Technology; 2019 Mar. Report No.: NAP-AX-18-003. Available from: https://www.healthit.gov/sites/default/files/page/2019-03/LeadingEdgeAccelerationProjectSpecialEmphasisNotice.pdf