2017

A Life Cycle Approach to the Development and Validation of an Ontology of the U.S. Common Rule (45 C.F.R. § 46)

Francis Manion
University of Texas Health Science Center at Houston, Francis.J.Manion@uth.tmc.edu

A Life Cycle Approach to the Development and Validation of an Ontology of the U.S. Common Rule (45 C.F.R. § 46)

A
Dissertation

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
School of Biomedical Informatics
in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

By

Frank J. Manion, M.S.E.

University of Texas Health Science Center at Houston

2017

Dissertation Committee:

Cui Tao, PhD[1], Advisor
Hua Xu, PhD[1]
William A. Weems, PhD[2]
Marcelline R. Harris, PhD[3]
Charles P. Friedman, PhD[4]

[1]The School of Biomedical Informatics
[2]University of Texas McGovern Medical School
[3]University of Michigan School of Nursing, Department of Systems, Populations, and Leadership
[4]University of Michigan Medical School, Department of Learning Health Sciences

Dedication

I dedicate this dissertation to my wife Carol R. Manion whose constant encouragement has been essential to both this work, and to my life. May you continue to be the music in my life!

I also dedicate this work to all those brave individuals who advance the human condition by taking part in human subjects research.

Acknowledgements

In the eloquent acknowledgment section of his recent Ph.D. dissertation from the School, Jose Franck DiazVasquez, to paraphrase, said ‘it all begins with people, it grows from interactions with people, and it is people who provide the opportunities.’ He is so right. There are many people who I must acknowledge for standing alongside me in this walk and providing constant support, encouragement, scholarship, and perhaps most importantly, their precious time.

First and foremost is my wife, Carol, who has continually encouraged me for many years while I worked on first my Master’s degree and now this dissertation. She has been rewarded by many lonely days and nights while I studied or worked on projects. I could not have accomplished this without her love and support and I can never repay her.

Next, I want to thank my advisor Dr. Cui Tao and all the members of my committee. Dr. Tao has taken me from knowing little about the field to this point and has been a constant source of wisdom, encouragement and a true help in this work. My colleague and friend Dr. Marcy Harris from the University of Michigan has been enormously generous with her time without which this work would have never been finished.

Dr. Bill Weems, who was a friend and colleague long before he was a member of my dissertation committee, has spent countless hours over the years discussing and tutoring me in certain technical aspects of privacy, authentication and authorization. Dr. Hua Xu

and his lab have assisted me in the field of NLP and been a constant positive influence. Dr. Charles Friedman at the University of Michigan provided a bit of an 'academic home' for me in Ann Arbor, and constantly has looked out for my best interests. I thank you all for the important roles you have played.

Many others from the Tao and Xu labs have assisted with this work. In particular Muhammad 'Tuan' Amith, Hsing-Yi 'Judy' Song, Madhuri Sankaranarayanapillai, and Yaoyun Zhang. Given my role as a distance student, I'm sure there are others who have given time and effort that I'm either unaware of or simply have forgotten. I hope they will forgive me and know that I am grateful to them all.

Many people from the University of Michigan and other academic medical centers have played a supporting role. Those who contributed their knowledge of the domain as subject area experts were Drs. Thomas Carey, Blake Roessler, Richard Redman, Victoria Blanc, Raymond Hutchinson, Michael Geisser, Jodyn Platt, Raymond De Vries, Terry VandenBosch, and Allan Loup, J.D. Members of the OBO consortium also provided advice with biomedical ontologies and the OBO Foundry. These include Drs. Yongqun 'Oliver' He from the University of Michigan, Drs. Christian Stockert and Jie Zheng from the University of Pennsylvania, and Dr. Mathias Brochhausen from the University of Arkansas. Dr. Jihad Obeid from the Medical University of South Carolina provided help with informed consent, and Ms. Helena Ellis, managing director at Biobanking Without Borders, LLC provided additional advice on biobanking.

Marisa Conte, M.L.I.S. from the University of Michigan Taubman Health Sciences Library assisted with the literature searches.

Finally, to all those who provided encouragement and support along the way. These include Dr. Jack London from Thomas Jefferson University, and innumerable colleagues at the University of Michigan, most particularly Ms. Jin Park. M.P.H. and Mr. Wei Shi M.S., M.B.A. I am grateful to you all.

This work was supported in part by grant HG009454 from the National Human Genome Research Institute. This work used the Protégé editor, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Abstract

Requirements for the protection of human research subjects stem from directly from federal regulation by the Department of Health and Human Services in Title 45 of the Code of Federal Regulations (C.F.R.) part 46. 15 other federal agencies include subpart A of part 46 verbatim in their own body of regulation. Hence 45 C.F.R. part 46 subpart A has come to be called colloquially the 'Common Rule.'

Overall motivation for this study began as a desire to facilitate the ethical sharing of biospecimen samples from large biospecimen collections by using ontologies. Previous work demonstrated that in general the informed consent process and subsequent decision making about data and specimen release still relies heavily on paper-based informed consent forms and processes. Consequently, well-validated computable models are needed to provide an enhanced foundation for data sharing.

This dissertation describes the development and validation of a Common Rule Ontology (CRO), expressed in the OWL-2 Web Ontology Language, and is intended to provide a computable semantic knowledge model for assessing and representing components of the information artifacts of required as part of regulated research under 45 C.F.R. § 46. I examine if the alignment of this ontology with the Basic Formal Ontology and other ontologies from the Open Biomedical Ontology (OBO) Foundry provide a good fit for the regulatory aspects of the Common Rule Ontology.

The dissertation also examines and proposes a new method for ongoing evaluation of ontology such as CRO across the ontology development lifecycle and suggest methods to achieve high quality, validated ontologies.

While the CRO is not in itself intended to be a complete solution to the data and specimen sharing problems outlined above, it is intended to produce a well-validated computationally grounded framework upon which others can build. This model can be used in future work to build decision support systems to assist Institutional Review Boards (IRBs), regulatory personnel, honest brokers, tissue bank managers, and other individuals in the decision-making process involving biorepository specimen and data sharing.

Vita

1976.....B.S., Music Education, Baldwin-Wallace University

1985.....M.S.E., Computer and Information Science, University of Pennsylvania

1976 – 1983.....Project Leader, CHI Research Incorporated, Audubon, NJ

1983 – 1984.....Member, Engineering Staff, RCA Adv. Technology Labs, Camden, NJ

1984 – 1986.....Systems Programmer, Fox Chase Cancer Center, Philadelphia, PA

1986 – 1990.....Senior Systems Programmer, Fox Chase Cancer Center, Philadelphia, PA

1990 – 1994.....Manager, Systems and Applications, Fox Chase Cancer Center, Philadelphia, PA

1990 – 1994.....Project Leader, Cooperative Human Linkage Center Fox Chase Cancer Center and University of Iowa

1994 – 2002.....Director, Research Computing Services & Bioinformatics Facility, Fox Chase Cancer Center, Philadelphia, PA

2002 – 2009.....Director, High-Performance Computing Facility, Fox Chase Cancer Center, Philadelphia, PA

2002 – 2009.....Member, Bioinformatics Program, Fox Chase Cancer Center, Philadelphia, PA

2002 – 2009.....Sr. Director & Chief Technology Officer/Deputy CIO, Information Science & Technology, Fox Chase Cancer Center, Philadelphia, PA

2009 – 2017.....Chief Informatics Officer & Director, Cancer Center Informatics Facility, University of Michigan Comprehensive Cancer Center, Ann Arbor, MI

Publications

Fenton, S.H., **Manion, F.**, Hsieh, K.L., and Harris, M., Informed Consent: Does Anyone Really Understand What is Contained in the Medical Record?*, Applied Clinical Informatics*, 6(3):466-477, 2015. PMID:26448792. PMCID: PMC4586336. doi: 10.4338/ACI-2014-09-SOA-0081.

Schumacher, F.R., Schmit, S.L., Jiao, S., Edlund, C.K., Wang, H., Zhang, B., Hsu, L., Huang, S.C., Fischer, C.P., Harju, J.F., Idos, G.E., Lejbkowicz, F., **Manion, F.J.** et. al. (2015). Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nature Communications*, 6:7138. PMID: 26151821. http://doi.org/10.1038/ncomms8138.

Manion, F.J., He, Y., Eisenhauer, E., Lin, Y., Karnovsky, A., Harris, M.R., Towards a Common Semantic Representation of Informed Consent for Biobank Specimens., Workshop on Starting an OBI-Based Biobank Ontology in *Proceeding of the 5th International Conference on Biomedical Ontologies (ICBO 2014)*. R.D. Boyce et. al. editors. Pp 61-63. Houston, TX. (2014).

Lin, Y., Harris, M.R., Manion, F.J., Eisenhauer, E., Zhao, B., Shi, W., Karnovsky, A., He., Y., Development of a BFO-Based Informed Consent Ontology (ICO), *Proceeding of the 5th International Conference on Biomedical Ontologies (ICBO 2014)*. R.D. Boyce et. al. editors. Houston, TX. (2014).

Hanauer, D.A., Liu, Y., Mei, Q., **Manion, F.J.**, Balis, U.J., Zheng, K., Hedging their mets: the use of uncertainty terms in clinical documents and its potential implications when sharing the documents with patients. AMIA Annu Symp Proc. 2012;2012:321-30. PMID: 23304302. PMCID: PMC3540426.

Manion, F.J., Harris, M.R., Buyuktur, A.G., Clark, P.M., An, L.C., Hanauer, D.A. Leveraging EHR Data for Outcomes and Comparative Effectiveness Research in Oncology. Curr Oncol Rep. 2012:Dec;14(6):494–501. PMID: 22948276. doi: 10.1007/s11912-012-0272-6.

Zheng, K., Mei, Q., Yang, L., **Manion, F.J.**, Balis, U.J., Hanauer D.A. (2011) Voice-Dictated versus Typed-in Clinician Notes: Linguistic Properties and the Potential Implications on Natural Language Processing. AMIA *Annu Symp Proc.* 2011:1630–1638. PMID:22195229.

Meropol, N.J., Daly, M.B., Vig, H.S., **Manion, F.J.**, Manne, S.L., Mazar, C., Solarino, N., Zubarev, V. (2011) Delivery of Internet-based cancer genetic counseling services to patients' homes: a feasibility study. *Journal of Telemedicine and Telecare*. 17:36-40, 2011. PMID: 21097566. doi 10.1258/jtt.2010.100116.

Manion, F.J., Weems, W.A., McNamee, J., and Robbins, R.J.: Federated Authentication and Authorization. In M. Ochs. J. Casavant, and R.V. Davuluri editors, *Biomedical Informatics for Cancer Research*, Springer, New York, 2010. doi: 10.1007/978-1-4419-5714-6.

Myers, R.B., Lomax, J.W. III, **Manion, F.J.**, Wood, N.M., Johnson, T.R., (2010) Data visualization of teen birth rate data using freely available rapid prototyping tools. In *IHI '10 Proceedings of the 1st ACM International Health Informatics Symposium*. T. Veinot editor. pp 381-385. New York, NY, USA. doi:10.1145/1882992.1883047.

Min, H., **Manion, F.J.**, Goralczyk, E., Wong, Y-N., Ross, E., Beck, J.R. (2009) Integration of Prostate Cancer Clinical Data Using an Ontology. *Journal of Biomedical Informatics*. 42(6):1034, 2009. PMID: 19497389. doi 10.1016/j.jbi.2009.05.007.

Manion, F. J., Robbins, R. J., Weems, W. A., Crowley, R. S. (2008) Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study. *BMC Medical Informatics and Decision Making*. 9:31, 2009. PMID: 19527521. doi:10.1186/1472-6947-9-31.

Qian, Y., Tchuvatkina, O., Spidlen, J., Wilkinson, P., Gasparetto, M., Jones, A. R., **Manion, F.J.**, Scheuermann, R.H., Sekaly R-P., Brinkman, R.R. (2009)

FuGEFlow: data model and markup language for flow cytometry. *BMC Bioinformatics*. 10:184, 2009. PMID: 19531228. doi:10.1186/1471-2105-10-184.

Bidaut, G., **Manion, F.J.**, Garcia, C., and Ochs, M.F. WaveRead: Automatic measurement of relative gene expression levels from microarrays using wavelet analysis. *Journal of Biomedical Informatics*, 39(4), 379, 2006, PMID: 1629855. doi:10.1016/j.jbi.2005.10.001.

Ochs MF, Goralczyk EM, Grant JD, **Manion FJ**, Yeung A, Seeholzer S, Mathew G, Hardy RR and Beck JR. A unified laboratory information management system for research data. Medinfo 2004: 1784.

Grant, J.D., Somers, L.A., Zhang, Y., **Manion, F.J.**, Bidaut, G., and Ochs, M.F. FGDP: functional genomics data pipeline for automated, multiple microarray data analyses. *Bioinformatics*, 20, 282, 2004. PMID: 14734324. doi: 10.1093/bioinformatics/btg407.

Kossenkov, V., **Manion, F.J.**, Korotkov, E., Moloshok, T.D., and Ochs, M.F. ASAP: Automated Sequence Annotation Pipeline for web-based updating of sequence information with a local dynamic database. *Bioinformatics*, 19, 675, 2003. PMID: 12651736. doi: 10.1093/bioinformatics/btg056.

Bidaut, G., Grant, J.D., Moloshok, T.D., **Manion, F.J.**, and Ochs, M.F.: Bayesian Decomposition analysis of gene expression in yeast deletion mutants. In K.

Johnson and S. Lin, Editors, *Methods of Microarray Data Analysis II*, 105 - 122, Kluwer Academic, Boston, 2002.

Moloshok, T.D., Klevecz, R.R., Grant, J.D., **Manion, F.J.**, Speier, W.F. IV, and Ochs, M.F. Application of Bayesian Decomposition for analyzing microarray data. *Bioinformatics*, 18, 566, 2002. PMID: 12016054.

Grant, J.D., Dunbrack, R.L., **Manion, F.J.**, and Ochs, M.F. BeoBLAST: A Package for Distributed BLAST and PSI-BLAST on a Beowulf Cluster. *Bioinformatics*, 18, 566, 2002. PMID: 12050075.

Grant, J.D., Broccoli, D., Muquit, M., **Manion, F.J.**, Tisdall, J., and Ochs, M.F. Telometric: a tool providing simplified, reproducible measurements of telomeric DNA from constant field agarose gels. *BioTechniques*, 31, 1314, 2001. PMID: 11768660.

Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., **Manion, F.**, Quillen, J., Sheffield, V.C., Sunden, S., Duyk, G.M., et al. A comprehensive human linkage map with centimorgan density. *Science*, 265, 2049-2054, 1994. PMID: 8091227. doi: 10.1126/science.8091227.

Field of Study

Biomedical Informatics

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Requirements for the protection of human research subjects stem from directly from federal regulation by the Department of Health and Human Services in Title 45 of the Code of Federal Regulations (C.F.R.) part 46. Fifteen other federal agencies include subpart A of part 46 verbatim in their own body of regulations. Hence 45 C.F.R. part 46, subpart A, has become to be called colloquially the 'Common Rule (CR).' Based on a desire to reduce risks to human subjects and improve our ability to share data and biospecimens, this dissertation describes the structure, development, and evaluation of a Common Rule Ontology (CRO). The ontology is intended to provide a formal, machine-computable model of the Common Rule's artifacts and processes.

If one considers the contents of informed consent documents as a legal contract between the research community and research participants, then ethics demands that there should be adequate processes and methods to represent the choices these participants have made and to adequately enforce their decisions. Unfortunately, mechanisms to ascribe consent and enforce these decisions appear inadequate in a research environment that desires increasingly large amounts of data. Methods for handling informed consent decisions at best remain non-interoperable, and at worst are wholly paper-based. It is my (untested) hypothesis that part of the reason this situation exists is a lack of technical capability in current informatics systems stemming directly from a lack of community developed theoretical frameworks, including appropriate information models and taxonomies. This

situation can compromise the trust that well-intentioned investigators seek to establish with the volunteers who contribute their data, specimens, and even at times, their lives to furthering research to improve the human condition.

1.1 Research Questions and Specific Aims

This dissertation focuses on (a) establishing the scope of prior efforts; (b) examining how major ontology projects in the field of life sciences knowledge management address concepts derived from regulatory foundations such as the Common Rule; and (c) develops and tests a series of evaluation method coupled with a new hybrid lifecycle ontology development framework suitable for creation of such an ontology. Development of a Common Rule Ontology is performed with this framework and the robustness of the evaluation framework is explored.

1.1.1 Research question one. *Is there evidence that the Common Rule is represented in any computational format?* My hypothesis is that such a base of legislative and regulatory knowledge does not currently exist in a formal ontology.

- S***pecific aim one*** is to conduct a literature review, and couple the results with a survey of ontologies that contain terms and classes of relevance to regulation of human subjects research and the Common Rule.

1.1.2 Research question two. *Does the Basic Formal Ontology (BFO) (Grenon & Smith, 2004) together with the ontologies and principles from the Open Biomedical Ontologies (OBO) Foundry (Smith et al., 2007) provide a suitable foundation for ontologies describing regulation such as the Common Rule?* The BFO and OBO Foundry

will be introduced in detail later in this chapter. To address this question specific aims two through five will provide the background work necessary construct an ontology of the Common Rule and ultimately align it with the BFO and OBO.

- ***Specific aim two*** is to use document analysis and knowledge elicitation techniques to develop dictionaries of terms and relationships between processes.
- ***Specific aim three*** is to validate this material with experts in regulated human subjects research and/or the Common Rule.
- ***Specific aim four*** is to develop an ontology of the Common Rule that integrates the upper-level Basic Formal Ontology and relevant OBO Foundry ontologies while adhering to the necessary OBO Foundry principles.
- ***Specific aim five*** is to use mixed evaluation methods to examine the ontology quality across a number of axes of quality described by the literature.

1.1.3 Research question three. *Do approaches exist that provide integrated and comprehensive evaluation across the ontology lifecycle?* My contention is that continuous, iterative, evaluation integrated with the development of an ontology is necessary to assure high-quality ontologies, particularly in a field dealing with risk mitigation and management to the extent the Common Rule does.

- ***Specific aim six*** is to develop an integrated evaluation and ontology development framework using state of the art approaches, and use it to both construct and validate the quality of the ontology across the ontology lifecycle.

1.2 Motivation and Background

Although focused on the U.S. Common Rule, the motivating factor for this work started as a desire to facilitate data sharing about biospecimens in moderate- to large-scale biorepositories. Data sharing and interoperability of multiple biorepositories in a federated manner has been an ongoing research area for some years. Projects such as the eMERGE network (Kho et al., 2011; McCarty et al., 2011; Pathak et al., 2011), and the NIH-funded Shared Pathology Informatics Network (SPIN) (Becich, 2007) aimed to develop mechanisms to search multiple collections of biospecimens for a variety of purposes, including molecular medicine, public health research and bio-surveillance. Important results from this prior work include models for honest broker systems and the development of suitable trust frameworks.

Biorepositories, however, are only a microcosm of the larger picture of both human subjects research and data sharing. Large-scale data sharing networks involving both clinical and research data are now being constructed. Examples of such networks include the American Society of Clinical Oncology's CancerLinq network (Schilsky, Michels, Kearbey, Yu, & Hudis, 2014; Yu, 2017) and PCORnet (Fleurence et al., 2014), along with many other initiatives sharing everything from whole genome sequencing data to biospecimens to clinical trials data.

The need for computable models to support privacy and security for data sharing in such research networks was a finding that emerged from a study my colleagues and I conducted in 2007 as part of the cancer Bioinformatics Grid (caBIG) project, a National Cancer Institute (NCI) grid computing project meant to support cancer research through

both tissue and data sharing (Beck & caBIG Strategic Planning Workspace, 2007; F. Manion, Robbins, Weems, & Crowley, 2009). While some of these recommendations ultimately influenced the design of the caBIG security infrastructure (Langella et al., 2007) and resulted in contributions to the Globus Toolkit (Foster, 2005), there has been no further work based on these recommendations.

In practice, many research networks attempt to address data and biospecimen exchange by manually pre-coordinating consent forms and consenting practices. Additionally, prior work in the field has not attempted to categorize aspects of physical and other forms of potential harm, or ethical constraints on research, in a computable manner. Instead, research networks such as SPIN relied on existing, precoordination methods such as deidentification of clinical notes and pathology reports, and maintenance of separate codebooks to protect the confidentiality and privacy of research subjects. While these methods are useful and important, they only operate at a syntactic level. Since they do not operate at the semantic level they cannot answer questions such as 'are the researchers disclosing exactly the information and possibly specimens for which I (the donor) consented.' Additionally, these approaches cannot address the challenges that one encounters when trying to use or combine large collections of either retrospective data or biospecimens. Such collections can have hundreds of thousands of specimens or data records, with the specimens consented over time with different consent models, consent forms, and restrictions on use. The only approach that is feasible is constructing a database of metadata regarding what the consent document allows. One can easily see

that for large amounts of data this is a difficult, open-ended problem that would require substantial effort to harmonize the metadata elements.

By developing a series of formal ontology at the domain and application level it will be possible to represent restrictions on data and biospecimen use at the semantic level. It should also be possible to use theorem-proving reasoners to enforce these restrictions. Figure 1 is a conceptual diagram of the sources of influence on regulatory policy and shows some of the ontology that will ultimately be needed. Domain-level ontology will allow for the leveraging of semantic web technologies such as the Web Ontology Language (OWL) (Boris Motik, Peter F. Patel-Schneider, & Bijan Parsia, 2012) and Linked Open Data. This, in turn, can be used to drive existing policy enforcement mechanisms (termed Policy Enforcement Points in the security literature) and web service-based security and authorization languages such as SAML and XACML (OASIS eXtensible Access Control Markup Language (XACML) Technical Committee, 2013; OASIS Security Services Technical Committee, 2005).

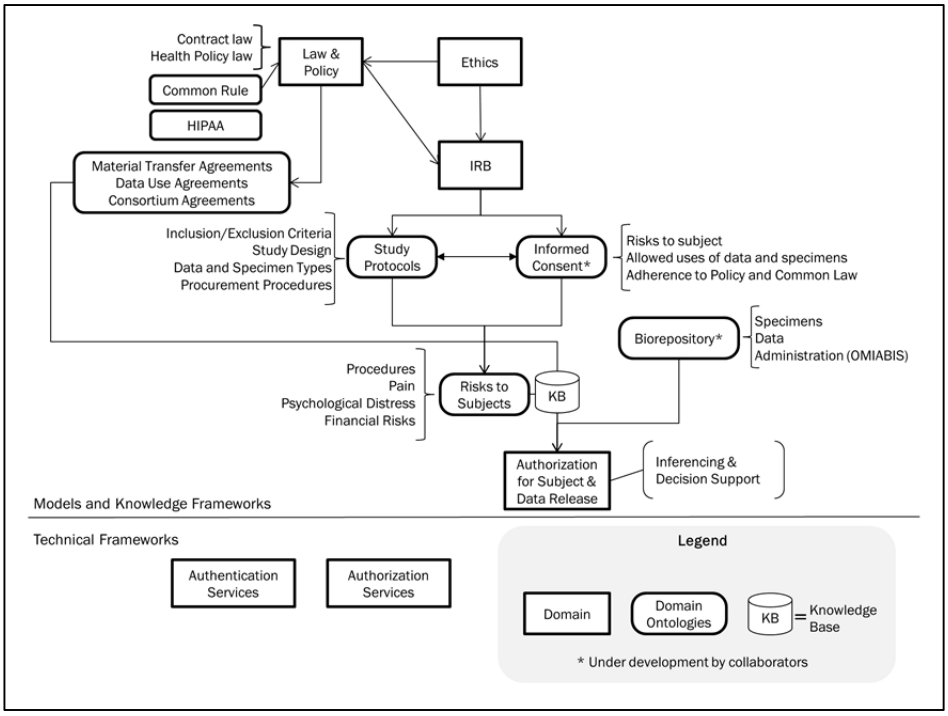


Figure 1. Conceptual Model of the Regulatory Space in Human Subjects Research.

1.3 Significance.

The management of informed consent for both research participation and data sharing still largely revolves around paper-based consent instruments. Systematic coding systems for these documents, whether electronic or paper, and the concepts contained therein do not exist. Consequently, it is difficult to leverage the substantial intellectual and financial resources devoted to developing institutional and other large-scale data warehouses and shared resources such as biorepositories. As an example, the head and neck oncology program at the University of Michigan has a collection of over 100,000 specimens, collected over two decades, with many different consent forms tied to specific subprojects. The gastrointestinal program has a similar situation containing over 400,000

specimens. These programs are not atypical — there are many similar specimen collections both at Michigan and elsewhere. Similar situations exist with clinical patient registries even if not linked with biospecimens. With paper-based systems it is difficult if not impossible to scan thousands of consent documents when building large cohorts or searching for biospecimens that exhibit not just the correct biomedical selection criteria, but also appropriate consent criteria. The problem is especially acute when data and queries must be exchanged between multiple research partners and at substantial scale, exactly the case required to achieve the vision of the Learning Health System (Grossman, Powers, & McGinnis, 2011; Institute of Medicine, 2007). Developing machine-readable formal models and systems for representing consent parameters will provide an opportunity to move beyond the problems described above.

This dissertation examines whether a high-quality ontology of the Common Rule can be constructed and aligned with an important integrated source of biomedical ontologies covering all aspects of the field of biomedical research, including biobanking. The Common Rule Ontology (CRO) is intended to provide a computable semantic knowledge model of the requirements necessary to adhere to one part of the required regulatory frameworks for human subjects research. The results of this work can be used for assessing informed consent, protocol documents, data use agreement, and for describing the output of IRB processes that are a required part of regulated research. While it is not in itself intended to be a complete solution to the problems outlined above, it is intended to produce a computational framework grounded in the actual federal regulations upon which others can build. Developing this capability will facilitate data and specimen

sharing while hopefully allowing for the reduction of risk to donors, improved regulatory review efficiency, human subject risk reduction, and enhanced compliance with the expressed wishes of the specimen donors. This ontology can be used in future work to build decision support systems to assist Institutional Review Boards (IRBs), regulatory personnel, honest brokers, tissue bank managers, and other individuals in the decision-making process involving biorepository specimen and data sharing. The work should be of interest to anyone involved in decision making for human subjects research including IRBs, honest brokers, biorepository managers, administrators, research networks, and standards groups.

Another important part of this work involves the development of a new, integrated framework incorporating continuous evaluation and lifecycle methods that will be of general utility as a method for high-quality ontology construction.

1.4 Background on the Common Rule

This section briefly describes the evolution of the Common Rule along with aspects of the regulations involving informed consent, including the basic models and requirements for consenting human subjects.

1.4.1 Research ethics and the evolution of the Common Rule.

Research involving human subjects requires that the research participants or their legal guardians be adequately informed about potential risks that might befall them should they choose to participate. The ethical principles underlying this requirement date back to the Nuremberg Code (Nuremberg Military Tribunal, 1949). Later well-known work in ethics

further refined these principles. The Declaration of Helsinki originated in 1964 with the World Medical Association and incorporated many of the principles from the Nuremberg Code. It is still considered a major ethical framework, and even now, in its seventh revision, it continues to influence research policy in many countries, including the United States (World Medical Association, 2013). The Belmont Report, commissioned under auspices of the National Research Act of 1974, laid out a series of ethical principles including respect for persons, beneficence, and justice. Additionally, it outlined requirements for appropriate separation between clinical practice and research. The report proposed that informed consent be required for all research participants, that a careful and systematic assessment of all risks and benefits be undertaken, and that selection of research subjects be just and equitable (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). It was at about this time that the Department of Health and Human Services (DHHS) began development of the regulations now known as the Common Rule (2017 Common Rule, 2017). The Common Rule was revised slightly in 2005 and revised again in 2017. An important change in the 2017 regulations is the addition of the concept of broad consent for secondary research with biospecimens and data (Menikoff, Kaneshiro, & Pritchard, 2017). The term 'secondary research' is somewhat vague and undefined in the final regulation. Prior to this revision, investigators wishing to study data or biospecimens derived from a prior research project containing identifiable data were required to seek additional, project-specific consent, or to obtain a waiver of consent from the IRB. The new regulations now allow an investigator to obtain prospective consent for unspecified future research using

identifiable private information and biospecimens at the time this material is collected. The revised Common Rule was slated to take effect on January 19, 2018, however, as of October 7, 2017, DHHS has suspended implementation of the 2017 revised Common Rule pending review by the Office of Management and Budget.

During this project, there has been substantial uncertainty about whether the 2017 revision of the Common Rule would become law, and if it would be implemented. It was only signed into law on the last day, January 19, 2017, of the Obama administration. Consequently, work done on this project mostly used the 2005 version of the regulation, which is still in effect at the moment. The work was extended slightly to include the major changes surrounding the concept of broad consent. Throughout this document, the term 'Common Rule' or the abbreviation 'CR' will mean the 2005 version, unless I specifically note otherwise.

1.4.2 Informed consent. Under the Common Rule, informed consent is required to ensure that participants in research (a) understand and agree to the potential risks and benefits of that research, and (b) understand their rights. The general principles on which this legislation is founded stem directly from principles discussed in the previous section. They consist of the precepts that (a) the participation in research is voluntary; (b) participation in the research is free from coercion; (c) participants can withdraw from research at any time without penalty; (d) subjects must be of sound mind and be allowed to have a reasonable time to decide to participate in research; (e) subjects should be reasonably informed about processes and implications involved in the research; and (f) certain groups of individuals are vulnerable and require special treatment and may not

even be fully capable of giving informed consent as a consequence of their circumstances. The Common Rule requires that research subjects receive information about eight so-called basic elements in the informed consent process, and possibly about six additional elements. For example, pregnant women must receive information that unforeseeable risks to a fetus may exist. While not addressed in this study, other countries and jurisdictions use different standards. For example, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), a body which seeks to harmonize regulatory requirements for interventional research between the European Union, Japan, and the United States, lays out 20 required elements for informed consent documents.

1.4.3 Common Rule protections for vulnerable populations. Vulnerable populations that the Common Rule requires special rules for in the administration of informed consent are defined in 45 C.F.R. § 46 subpart B as pregnant women, fetuses, and neonates; in 45 C.F.R. § 46 subpart C as prisoners and other detained individuals; and in 45 C.F.R. § 46 subpart D as children. The Food and Drug Administration regulations governing informed consent for clinical trials contains protections for members of the armed forces, as well as exceptions from general requirements for military personnel in certain military situations. See 21 C.F.R. § 50 for details. In addition to *permission* (essentially informed consent) from a parent or legal guardian, research on pediatric patients requires assent from the child, defined under 45 C.F.R. §46.402(d) as ‘a child's affirmative agreement to participate in research’. In addition to assent, children who are wards of any institution, including the state, require

the approval from an independent third-party advocate unaffiliated with the organization conducting the research as per 45 C.F.R. § 46.409. Subparts B, C, and D of 45 C.F.R. § 46 are included in subpart A by reference, and consequently, an ontology of the Common Rule needs to include these subparts.

1.4.4 Models of informed consent. In general, two basic models of informed consent are used for subjects, especially those contributing specimens or data intended for future research. These consist of *general informed consent*, and *broad, universal or global consent*. In both forms of consent, a consent instrument is presented to the subject by a researcher or research coordinator via a document, or more recently a mobile device or web form. As per the preceding paragraph, this document contains content stipulated by the U.S. Common Rule, 45 C.F.R. § 46, as well as additional constraints on the collection and use of specimens that may have been imposed during regulatory review by the IRB or an ethics board. If the consent form gives a research participant additional choices regarding the use of specimens or data derived from their person during research, the consent is said to be *tiered*. Since large biobanks and research registries often attempt to accrue specimens for future unspecified research uses, use of tiered consent documents with future-looking questions has been a fairly common practice. Examples of questions from a tiered consent form might include:

- Can we do genome sequencing on DNA derived from your blood?
- If we release [some or part of] your specimen to a researcher, can they also have access to data from your medical record relating to your condition?
- Will you allow your data to be combined with data obtained from aborted fetuses?

Changes proposed to the Common Rule in the 2017 version were meant to specifically allow an investigator options for collection and use of specimens for secondary research, assuming that certain information required by the Common Rule revision was presented to the study subject in the consent document. In the common vernacular this is termed *broad* consent.

1.4.5 Informed consent, genetic testing and other considerations. Due to the various legal, financial, and other risks associated with genetic testing, some organizations use a separate consent template when constructing informed consent documents for genetic testing. These risks include disclosure of unknown familial relationships; discrimination in the ability to purchase and increased rates for various insurance products such as disability and long-term care; required release of results to insurers; the potential need for additional testing; and the detection of untreatable conditions (Columbia University, 2014). Other organizations use special consent forms for functional Magnetic Resonance Imaging testing linked to behaviors (Stanford IRB, 2015), and for clinical trials and research linked to stem cells (Johns Hopkins Medicine IRB, 2013; Stanford IRB, 2015). Often these documents include sections that address the collection of various tissue-based specimens, the creation of cell lines, and collection and use of data from the research subject.

1.4.6 Broad, blanket, or universal consent. In the case of a *blanket* or *universal consent*, a particular organization attempts to use a single consent form to collect biospecimens of a particular type from all research participants. These forms are typically used in the creation of shared resources such as an institutional biorepository where large

numbers of individuals are asked to contribute specimens to the biorepository, typically to construct a strategic resource capable of being used in unspecified future research. These contributions may be as simple as a buccal swab, or as complicated as excess tissue surgically removed during a biopsy or therapeutic resection.

1.4.7 Opt-in and opt-out models of consent. A consideration in the construction of informed consent materials and research protocols is whether the protocol employs an *opt-in* or an *opt-out* research participation strategy. An opt-in strategy requires informed consent at enrollment by each participant. While this is feasible and logistically possible for small studies, it can pose a substantial financial and logistical burden for the creation of large-scale resources. When information from medical records, biospecimens, or other sources such as public archives can be linked, it may be possible to construct large cohorts without additional effort. In an opt-out strategy, participants are identified and given the opportunity to actively withdraw from the study if they do not wish to participate. Opt-out methods are typically used where there is minimal risk to participants and where a universal consent approach has been used. Some medical ethicists argue that the use of opt-out methods is problematic, mostly due to a violation of confidentiality, particularly if data is linked to medical records data collected over long periods of time or with the use discarded tissue specimens (Regidor, 2004). Opt-in methods have been associated with a poor response rate for some studies, and are known to introduce the possibility of selection bias into studies involving linkage with medical records data (Hewison & Haines, 2006; Junghans, Feder, Hemingway, Timmis, & Jones, 2005). Research on individual preferences for opt-in versus opt-out consent show that while

participation in research remains high using both approaches, individuals prefer opt-in methods (Kaufman, Bollinger, Dvoskin, & Scott, 2012).

1.5 Ontologies and Knowledge Modeling

Modern frameworks for developing, organizing, and distributing machine-readable computable models consist of information models, typically expressed in Unified Modeling Language (UML), and ontologies (Booch, Rumbaugh, & Jacobson, 2005). UML models represent three major aspects of systems (structure, behavior, and interaction) as types of formal diagrams. An important characteristic of these models is that they can represent state and process of a computerized software system. Gruber (1993), defined ontology as 'an explicit specification of a conceptualization' (Gruber, 1993). To the uninitiated, this definition is unsatisfying and vague. A better definition perhaps is in Noy (Noy & McGuinness, 2001) who defines them as 'explicit formal specifications of the terms in a domain and the relations between them.' In the realm of informatics and knowledge management, ontologies strive to represent knowledge in the form of objects (termed classes), and relationships or functions between these classes. Axioms are used to permit machine reasoning on instances of data expressed in the ontology. Consequently, ontologies can be linked directly to known axioms, proven theorems, and other theory. With some restrictions, classes, relations, axioms, etc. can represent anything about a field that can be stated or inferred. Generally, these are represented as a triple of the form:

Subject, predicate, object.

For example, our ontology might make the assertion:

<an adult> <can be a> <research subject>.

Ontologies are typically represented in a graph language such as the Web Ontology Language (OWL) (Boris Motik et al., 2012).

An ontology can be classified by how expressive it is, or put another way, what sort of semantic constructs it can represent. A well-known 'spectrum' of ontology that indicates the level of expressivity of model systems from weak to strong semantics classifies *taxonomies* as having the potential for syntactic interoperability, *thesauri* as having structural interoperability potential, *conceptual models* as having weak semantic interoperability, and logical theory-based models or *ontologies* as having strong semantics (Albert & Steiner, 2005; Carnot, Feltovich, Hoffman, Feltovich, & Novak, 2003; Eppler, 2006; Obrst, 2010). This is shown pictorially in Figure 2.

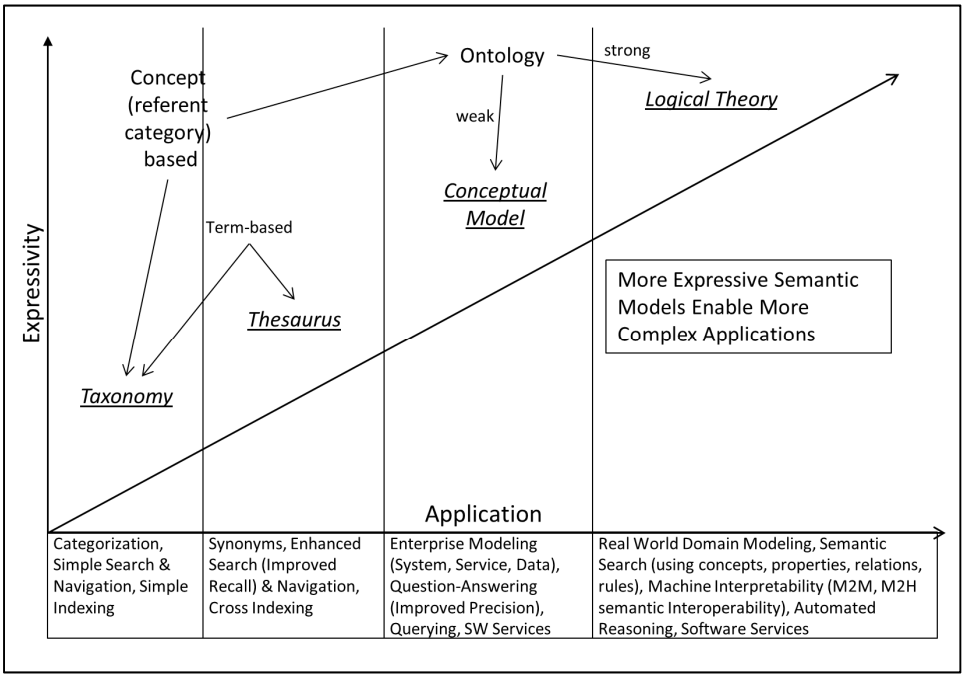


Figure 2. The semantic spectrum adapted from Obrst (Obrst, 2010).

It should be noted that ontologies gain their semantic expressivity not only from the formal definitions of terms, classes, and relations of the ontology given by the user, but in the manner that the resulting graph of these items is bound together or connected. OWL allows for the expression of concepts in description logics that represent a portion of first-order logic. Consequently, semantics inherent in fundamental theorems from logic and mathematics such as *completeness*, *validity*, *satisfiability*, and *decidability* can be applied using logical inference via algorithms called *reasoners*. Various versions of the OWL language allow for different degrees of expression of logical constructs. The current version used is OWL-2. There are two major subsets of OWL-2. The first is OWL-2 DL which is a somewhat restricted form of OWL-2. A major aspect of OWL-2 DL is that ontologies expressed in it are mathematically decidable. OWL-2 Full does not guarantee decidability and for this reason is considered more difficult to work with. In addition to these two variants of OWL-2, the language contains three so-called profiles intended to make certain tasks easier, but which will not be discussed further here. The interested reader is referred to the OWL 2 Web Ontology Language Primer (OWL 2 Primer, 2012).

1.6 Innovation and Contribution to the Field

The innovation in this proposal is the creation of a formal, computable model in the form of an ontology represented in OWL-2 DL to describe the domain of human research described in the Common Rule. As will be discussed in Chapter 2, literature and ontological term searches performed in PubMed, the National Center for Biomedical Ontologies (NCBO), and the OBO Foundry reveal that no meaningful computational

models of the Common Rule exist. Without models of the Common Rule and other regulation to provide grounding, the modeling of consent, authorization, rights, and obligations across the translational research spectrum, or from multiple data sources and institutions, will remain a difficult, manually intensive problem.

The proposal is also innovative in the creation and use of a hybrid development and evaluation framework that incorporates (a) an ontology lifecycle model; (b) a detailed ontology construction framework; (c) the use of concept maps for knowledge elicitation; and (d) multiple ongoing quality evaluation techniques for knowledge verification and validation across the ontology lifecycle. This approach is intended to elicit knowledge from domain experts in a fashion easily understood by them. It is also intended to preserve the details of the structure and syntax of the relationships and classes while the data is transformed into an OWL-2-based ontology.

1.7 Overview of the document

The outline of the remainder of the dissertation is shown below.

Chapter 2 — presents a literature review and gap analysis. The chapter contains a review of the literature and a survey of existing ontologies that have some representation of regulation of human subject research. It is intended to provide a contrast of the different approaches used to represent regulatory knowledge deriving from the Common Rule.

Chapter 3 — Common Rule Ontology construction, methods, and frameworks. Introduces the formal upper- and mid-level ontologies used in this project, describes the knowledge elicitation techniques used with domain experts, reviews the current state-of-

the-art in ontology construction and lifecycle methods, and presents the technical methods used for the construction of the ontology.

Chapter 4 — describes the various artifacts and technical results of the ontology construction process, the design patterns used, and the integration with the BFO, OBO Foundry, and other ontologies.

Chapter 5 — Ontology evaluation methods, describes the current state of ontology quality evaluation and introduces the methods used to validate the CRO.

Chapter 6 — presents the results of the evaluation of the CRO along multiple quality axes.

Chapter 7 — summarizes the results, limitations, and future directions of the work.

1.8 Relating Chapters to Specific Aims

Work described in Chapter 2 addresses specific aim one. Methods and results described in Chapter 3 and Chapter 4 specific aims two through four. Chapter 5 and Chapter 6 address specific aim five. Specific aim six is addressed in Chapter 3 through Chapter 6.

Chapter 2. Literature Review and Gap Analysis

This chapter conducts a systematized literature review to address specific aim one and explore what studies have attempted to develop ontology involving regulatory processes associated with informed consent and derived from the Common Rule. The chapter has an additional focus on the motivating biobanking use cases. The goal is to explore gaps in the literature as well as the scope and coverage of any regulatory ontology that exist. Defined inclusion and exclusion criteria were used to examine the biomedical research literature, as well as to search a number of repositories of biomedical ontologies, including the Open Biomedical Ontologies (OBO) Foundry, and the National Center for Biomedical Ontologies (Musen et al., 2012; Smith et al., 2007). The analysis reveals little work in this area. Of 944 publications initially reviewed only three address aspects of data from biorepositories, and only one of those concerns itself with informed consent. Similarly, outside of our own preliminary ontology of informed consent, only three other ontologies exist, and these mostly concern themselves with aspects of biospecimen data, not regulatory processes.

2.1 Search Methodology for PubMed

A protocol for searching PubMed was created with the assistance of a research librarian (Marisa Conte) at the University of Michigan Taubman Health Sciences Library. The protocol was designed to discover formal models of informed consent that are used in

conjunction with biorepositories or collections of biospecimens from clinical trials. The searches were initially performed in 2015 and again on May 30, 2017, to refresh the results in the intervening period. Searches were conducted using keywords and phrases in the title and abstract, as well as Medical Subject Headings (MeSH) as appropriate. Separate searches were conducted for the following topic areas:

- Ontologies, or the Unified Medical Language System (Bodenreider, 2004),
- Biobanks, biorepositories, or other biorepository-based research,
- Clinical trials or research protocols, due to the widespread use of biospecimens and subsequent biobanking of these specimens in clinical trials,
- Informed consent.

The actual PubMed search terms used for each of these searches are shown in Table 1.

Table 1

PubMed Topics and Accompanying Search Expressions

Search #	Topic	PubMed search expression
1.	Ontologies, or the Unified Medical Language System	(ontology[tiab] OR ontologies[tiab] OR vocabularies, controlled[mh] OR unified medical language system[mh] OR umls[tiab])
2.	Biobanks, biorepositories, or other biorepository-based research	('biological specimen banks'[MeSH Terms] OR biobank[tiab] OR biobanks[tiab] OR 'tissue bank'[tiab] OR 'biorepository'[tiab] OR 'biorepositories'[tiab])
3.	Clinical trials, or research protocols	(clinical trials as topic[mh] OR 'clinical trial'[tiab] OR 'clinical trials'[tiab] OR clinical protocols[mh] OR 'research protocol'[tiab])
4.	Informed consent	('informed consent'[MeSH Terms] OR 'informed consent'[tiab])

Key: [mh] or [MeSH Terms] are search qualifiers for MeSH terms
[tiab] is a search for the keyword in the article title or abstract

For each search number one through four, counts of the number of articles retrieved were recorded, and the actual lists of the articles were separately recorded using the 'Send to' option of the PubMed website (https://www.ncbi.nlm.nih.gov/pubmed/). The articles retrieved were recorded both as text files in MEDLINE format and as Microsoft Excel CSV files. The advantage of this latter format is that the data can be manipulated to examine article counts, sort by a variety of fields, and add annotation columns and other information for later analysis. Results of the PubMed searches were combined in the following way (as shown in PubMed search language): (#1 AND #2) OR (#1 AND #3) OR (#1 AND #4). The results were again separately recorded as both text files in

MEDLINE format, and as Excel CSV files. The purpose of recording the individual searches separately, and well as in a final set, was to provide a technical control for accuracy of the results.

2.1.1 Manual screening of candidate papers. Papers were manually reviewed for relevance. They were excluded if their principal focus was not in the topical area of the search, or if no abstract was available and relevance couldn't be determined from the title alone. For example, an abstract might have mentioned that samples were taken from a biorepository, but the title or abstract made clear that the article was focused on an analysis of proteins and RNA derived from the specimens rather than the consent processes or artifacts. Such an article would not be relevant to this review and consequently was excluded.

Results of the review were recorded for each paper. A categorical variable recording relevance to the study coded as 'Yes' or 'No' was used. Principal topic area of the paper, based on MeSH terms if available, or by my judgment, was recorded into one of 13 categories: basic research; biobanks; clinical research; ethics; health policy/services research; informatics research; infrastructure; mental disorders; ontology, semantics and information models; public health research; quality improvement; substance abuse; or translational research. Included papers were reviewed manually looking for evidence of the development of a theory or formal model of (a) informed consent processes; (b) the U.S. Common Rule; or (c) other regulatory processes involved with the collection and distribution of biospecimens. The full texts of papers passing this initial screening procedure were reviewed and a final decision made regarding the relevance of the work.

2.1.2 Methodology for searching ontology repositories. Searches were performed on two major repositories of biomedical ontologies. The first repository was the OBO Foundry, a collection of orthogonally developed ontologies from the field of biomedicine that are based on and aligned with the Basic Formal Ontology (BFO) (Arp, Smith, & Spear, 2015). The other repository was the NCBO Ontology Library (Musen et al., 2012). Searches proceeded as follows:

1. Searches in the OBO Foundry were performed using the Ontobee search engine (Ontobee, 2017; Xiang, Mungall, Ruttenberg, & He, 2011).
2. Searches in the NCBO Ontology Library were performed using the NCBO BioPortal (NCBO BioPortal, 2017; Whetzel et al., 2011).
3. Since these search engines are generally restricted to searching for classes or property names in the ontologies, searches were only performed looking for the term 'consent.'
4. The documentation or 'annotation properties' of the resulting ontologies were reviewed to determine the exact nature of their primary topic.
5. A record was made of the number of ontologies or vocabularies found, the type (structured vocabulary or ontology), the topical focus, and a judgment of whether the ontology was relevant.

Ontology initially deemed relevant were retrieved and examined with the Protégé ontology editor (Gennari et al., 2003) for inclusion in the final results. To be included in the results, the ontology needed to demonstrate that it contained concepts related to biorepositories, biobanks, or regulation. Simple classes or properties representing consent

were not deemed sufficient. An ontology had to demonstrate that it contained some type of semantic model of one or more components of the informed consent process, such as the processes, participants, roles, artifacts, obligations, or rights of the participants involved in consent. Information on the existence of a model, and the completeness of the model with regard to which of the aforementioned elements it contains were recorded.

2.2 Results

2.2.1 Results from the PubMed search. The results from the PubMed literature searchers are shown in Figure 3.

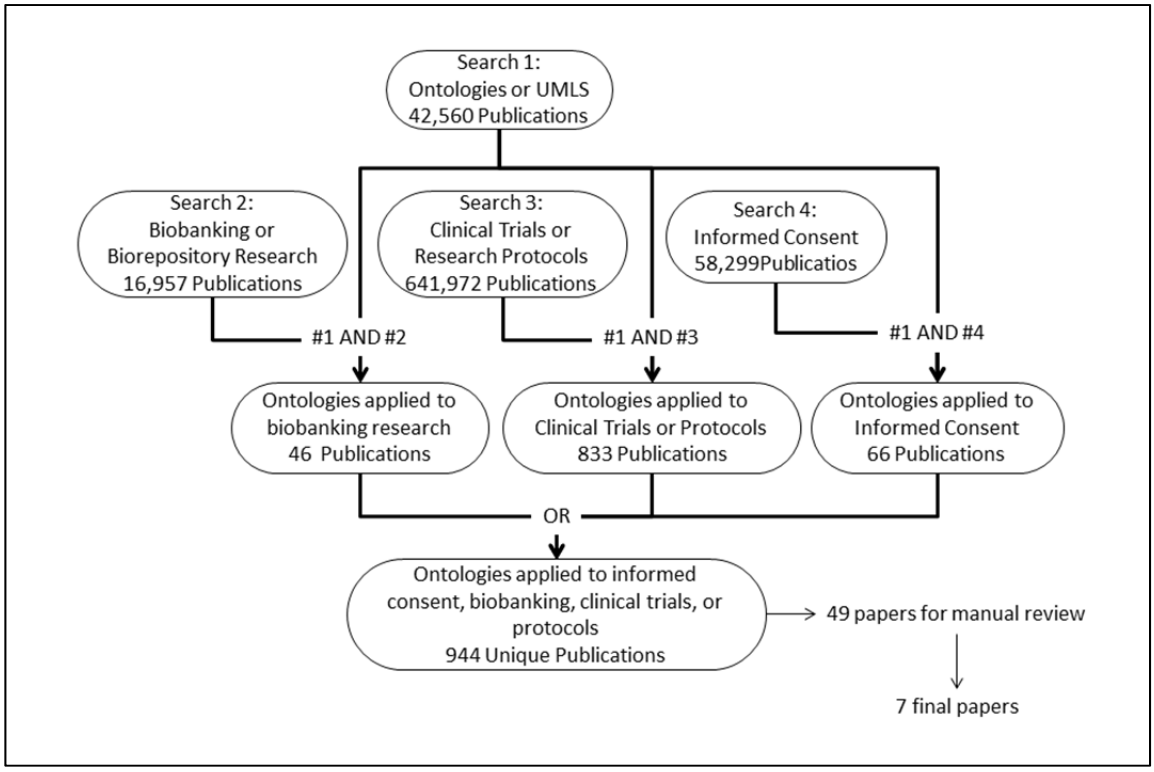


Figure 3. Results of Literature Search in PubMed.

Table 2 presents the result of the literature retrieval summarizing the thirteen principal categories of analysis after review by title and abstract. Based on the manual review of titles and abstracts, only 49 papers were judged to be of sufficient interest to review further.

Table 2

Principal article topics and relevant articles resulting from PubMed search

Subject	Relevant	Not Relevant	Totals
Ontology, semantics, and information models	27	55	82
Biobanks	12	3	15
Informatics research	5	103	108
Infrastructure	4	18	22
Ethics	1	2	3
Basic research		13	13
Clinical research		495	495
Health policy/services research		19	19
Mental disorders		115	115
Public health research		1	1
Quality improvement		4	4
Substance abuse		29	29
Translational research		38	38
Totals	49	895	944

Manual review of the 49 papers revealed that two of the papers were actually duplicates. Of the remaining 47 papers, only seven addressed aspects of regulatory science or data

representation in biobanks. These papers underwent full-text review for final determination of relevance.

2.2.2 Results from the ontology searches. Error! Reference source not found. shows the results of the search for the term 'consent' as a component of a class or property name in the OBO Foundry and NCBO Ontology libraries. Only two ontologies, the Informed Consent Ontology (ICO), which represents this author's own work, and the Vaccine Informed Consent Ontology (VICO), an extension of ICO, contain enough classes to have a sophisticated model of informed consent (Lin et al., 2014; Lin, Zheng, & He, 2016). Fifteen ontologies contained three or fewer class names with 'consent' in them, indicating they lacked sufficient coverage to be relevant. Analysis of these fifteen ontologies showed that they principally were inheriting classes from the Ontology for Biomedical Investigation (OBI), which has a small number of classes covering informed consent, and two classes specifically mentioning consent (Brinkman et al., 2010).

Table 3

Number of Ontology Classes Mentioning 'Consent' Resulting from Searches in the OntoBee and NCBO Ontology Repositories Against Class and Relation Names

Ontology Name	BFO-Based	OntoBee	BioPortal
ICO – Informed Consent Ontology	Yes	68	60
VICO – Vaccine Informed Consent Ontology	Yes	34	9
NCIT – National Cancer Institute Thesaurus	No*	14	13
HL7 – Health Level 7 Version 3 Reference Information Model	No		14
MESH – Medical Subject Headings	No		7
OMIT – Ontology for MIRNA Target	Yes	6	
IAO – Information Artifact Ontology	Yes	2	2

Ontology Name	BFO-Based	OntoBee	BioPortal
OBI – Ontology for Biomedical Investigations	Yes	2	2
DUO – Data Use Ontology	Yes	3	
GENEPIO – Genomic Epidemiology Ontology	Yes	3	
NCIBTaxon – NCBI organismal classification	Yes	3	
ERO – Eagle-I Research Resource Ontology	Yes	1	1
PR – Protein Ontology	Yes	1	1
RADLEX – Radiology Lexicon	No		2
APAONTO – Psychology Ontology	No		1
CRISP – Computer Retrieval of Information on Scientific Projects Thesaurus	No		1
DCM – DICOM Controlled Terminology	No		1
GAZ – Gazetteer	Yes	1	
ONTOAD -- Bilingual Ontology of Alzheimer's Disease and Related Diseases	No		1
PCORI – Patient-Centered Outcome Research Institute Ontology	No		1
PMA – Portfolio Management Application	No		1
RNPRIO – Research Network and Patient Registry Inventory Ontology	No		1
Totals		138	118

Note. Counts shown represent Universal Research Identifiers (URI's) that are unique for a given class. Thus, if a class appears with two different labels, it is counted only once.

*NCIT is not modeled as a BFO-based ontology, however a searchable copy of it does exist in OntoBee

2.3 Discussion of Literature Search Results

The literature search identified two major projects and two separate but interrelated projects that focused at least part of their efforts on developing ontology for representing

the informed consent properties necessary for collection and reuse of biospecimens and data. The major ontologies discovered by the ontology search are part of these projects.

2.3.1 The MIABIS project. Three papers discuss the development of a data model termed Minimal Information About BIobank data Sharing (MIABIS), the creation of a related BFO-based ontology in the OBO Foundry (OMIABIS), and the subsequent use of the data model (Brochhausen et al., 2013a; Fransson, Rial-Sebbag, Brochhausen, & Litton, 2014; Merino-Martinez et al., 2016). MIABIS, developed in 2012, was intended to support the goals of the Biobanking and BioMolecular Resources Research Infrastructure project in Europe (Norlin et al., 2012). The scope of this model, and consequently that of the resulting OBIABIS ontology, is the biobank administration domain. It supports federated queries about the types of specimen collections available in a biorepository and associated metadata about the collection. The level of detail in the ontology at the level of the biobank, not the biospecimen, and consequently does not cover the individual specimen's consent for use.

2.3.2 The iDASH project. Two papers describe projects that were part of the Integrating Data for Analysis, Anonymization, and Sharing (iDASH) project (Grando & Schwab, 2013; Sim et al., 2012). The paper by Grando and colleagues reports on work based on the HL7 Security and Privacy Ontology (HL7 Security Work Group, 2014) to build a system based on the eXtensible Access Control Markup Language (OASIS eXtensible Access Control Markup Language (XACML) Technical Committee, 2013) version 2.0 standard to allow access to data according to an ontological model. The second paper by Sim reports on the development of an ontology of clinical research

named OCRe. Both of these works appeared promising, but the iDASH program has now ended, and it is not clear that this work has been carried forward.

2.3.3 The ICO and OBIB projects. Two other works identified from the literature search include the Ontology for Biobanking (OBIB) (Brochhausen et al., 2016), and the Vaccine Informed Consent Ontology (VICO) (Lin et al., 2016). Both of these are OBO Foundry ontologies and use the informed consent model from the Informed Consent Ontology (ICO) (Lin et al., 2014), of which I am a co-developer. OBIB is currently under development by a team led by Dr. Christian Stockert at the University of Pennsylvania. I am collaborating with this group on enhancements to ICO. VICO is an ontology that builds off of the classes in ICO to represent specific informed consent forms used in human vaccination. It does not attempt to expand the aspects of ICO that are involved with regulations, IRB review, or other processes. Both OBIB and VICO are under active development and build off other major bodies of work in the OBO Foundry.

2.4 Review of Ontology Repository Search Results.

The small number of classes in most ontology mentioning 'consent' in the OBO Foundry and NCBO ontology repositories suggests that very little work has been done in developing formal models of the regulatory processes involved with informed consent. A complete ontology of consent would need to model aspects of the U.S. Common Rule, and required processes such as determination of research eligibility, type of IRB review required, type of informed consent required, and representations of the informed consent process itself. It is readily apparent from Table 3 that only the ICO, VICO, the National

Cancer Institute Thesaurus (NCIT), and Health Level 7 Reference Information Model (HL7-RIM) have a sufficient number of classes to model at least some of these aspects.

2.4.1 Ontology with more than 15 classes — the ICO group of results. ICO is a preliminary ontology still under development that mostly focuses on the information content of informed consent forms. While informed by the requirements of the Common Rule in 45 C.F.R. §46.116 it is not grounded in other relevant parts of the Common Rule. It currently does not address vulnerable populations, issues of rights and obligations of the research participants and study teams, or restrictions on data and specimen distribution and use. VICO is an ontology that builds off of the classes in ICO to represent specific informed consent forms used in human vaccination. It does not attempt to expand the aspects of ICO that are involved with regulations, IRB review, or other processes.

2.4.2 Ontology with between six and 15 classes — NCIT, HL7, MeSH, and OMIT. The NCIT has a number of terms related to informed consent, such as 'Consent Form', 'Date and Time of Informed Consent', but there is no evidence of an overarching semantic model tied to underlying regulatory requirements. The HL7-RIM does have classes and value sets attempting to model aspects of consent, however, these appear to be under defined, with just high-level qualifiers, such as referring to the U.S. Common Rule as a value for the governing regulation for an action, without specifying more detail. The HL7-RIM model also appears incomplete, mostly focused on what HL7 terms 'consent directives' that are more general than informed consent, and it does not model other parts of the informed consent process. It is also not clear that the HL7 model is

applicable or well suited to the research environment or to secondary research. MeSH terms are not applicable to an information model. The OMIT ontology is an ontology about micro-RNA targets and as such is not relevant to a model of consent for biospecimens; the six classes mentioning consent are inherited from the complete MeSH hierarchy that is incorporated into the ontology. These classes lack formal definitions.

2.4.3 Ontology with fewer than six classes. Other ontologies found from the ontology search had between one and three classes, not enough to model the complex regulatory requirements of informed consent or the Common Rule. Review of these ontologies using the Protégé editor indicated they were inheriting and using classes from core mid-level ontologies in the OBO Foundry, principally the Ontology for Biomedical Investigations (OBI) (Brinkman et al., 2010). OBI is an ontology focused on all manner of processes and entities used in biomedical research. Overall it has about seven concepts relating to informed consent. Review of the annotation properties of these other ontologies also indicated they were off-topic and not modeling informed consent or regulatory processes.

2.5 Critical Analysis of the Combined Results

In general, work to develop semantically interoperable representations of research protocols and consent involving biorepositories is in an early stage of development. Efforts to date include work focused on the development of stand-alone ontologies for various types of clinical research, and larger efforts focused on developing integrated

ontology knowledge bases that include interlinked constituent domain ontologies or information models.

Efforts in the former category include the Ontology of Clinical Research (OCRe), an ontology focused on providing a framework for describing objects, workflows, and events involving human research. The authors specifically state that this ontology could be extended in the future to provide for studies involving specimen collection. Other important work in this category includes the Ontology of Clinical Investigations (OCI) and the Ontology of Biomedical Investigations (OBI). OCI, a part of OBI, focuses on the curation and organization of terms used in clinical investigations.

Efforts in the latter category, that of integrated projects, include the iDASH consortium (http://www.idash.org) which developed an informed consent ontology focused on obligations and access to clinical data. The iDASH work is based on the HL7 Security and Permissions Ontology (HL7 Security Work Group, 2014). Brochhausen and colleagues have perhaps the most compelling implementation to date in the OMIABIS ontology. This ontology covers the biobank-administration domain which builds on the MIABIS (Norlin et al., 2012) project from Europe.

2.6 Conclusions from the combined literature and ontology review

While projects and ontologies exist that have attempted to model aspects of the informed consent or research regulatory processes exist, in general work to develop semantically interoperable representations of requirements for human subjects research is in an early stage of development. With the possible exception of the HL7-RIM, none of the existing models and ontology appear to have systematically derived models from a detailed

examination of law and regulatory statutes. Related efforts to date include work focused on the development of stand-alone ontologies for various types of clinical research, such as OCRe, and larger efforts focused on developing integrated ontology-based knowledge bases that include interlinked constituent domain ontologies or information models, such as the HL7-RIM, and OBIB. As noted before, a complete ontology of consent would need to model substantial parts of the U.S. Common Rule including processes and constraints defined by that regulation. The conclusion is that from an ontological standpoint, even for small subsets of the Common Rule, no cohesive body of work exists. Further, what little there is appears to be scattered among multiple ontologies developed with different focal topic areas.

2.7 Limitations of the Literature Review

The review in this chapter was specifically focused on exploring what work has been done to represent United States federal regulations in informed consent, with a focus on biobanking. As such, I did not examine foreign, state, local, regional, or tribal regulations. Note that the OMIABIS initiative is based in Europe, but its focus is not really informed consent or the representation of regulatory concepts except in the broadest, most general fashion. The ontology and terminology searches did not look at details of the HL7-RIM, or other clinical vocabularies. A search in the Unified Medical Language System (UMLS) was not conducted. Multiple raters were not used for the literature review.

Chapter 3. Common Rule Ontology Construction: Methods and Frameworks

This chapter lays out the development processes associated construction of the Common Rule Ontology. I first present the formal ontological underpinnings of the upper- and mid-level ontologies that undergird the structure and semantics of the CRO. Next I present an overview and background of the formal methods involved with the ontology development lifecycle. Subsequently I discuss the alignment of the high-level NIST lifecycle model with the more detailed METHONTOLOGY ontology construction method used to guide specific tasks. Then I show the alignment of formal evaluation methods with the ontology lifecycle and construction processes and demonstrate how they were iteratively used to provide formative evaluation. The final part of the chapter discusses details of the steps used to construct the ontology.

3.1 Ontology Frameworks: Upper- and Mid-Level Ontologies

3.1.1 The Basic Formal Ontology. The Basic Formal Ontology (BFO) is based on a theory of basic structures of reality (Grenon & Smith, 2004). The ontology consists of a hierarchy of two top-level classes, *continuant*, and *occurrent*, described below, and a variety of subclasses existing underneath them. Core sets of relations are also defined by the BFO. A full description of the BFO and the theory of ontological realism on which it is based are beyond the scope of this dissertation, however key concepts are summarized

below. The interested reader is referred to the excellent book by Arp, Smith, and Spear that discusses this topic in more detail (Arp et al., 2015)

3.1.1.1 Continuants. A continuant is an entity that exists through time. Continuants can exist at different levels of granularity, depending on the scope of what is being modeled in the ontology. Examples of continuants are people, the role a person plays in an organization, a document, a desk, a chair, a cell, an atom, a quark, and so forth. Continuants are further subdivided into three major subclasses; independent continuants, generically dependent continuants, and specifically dependent continuants. As previously mentioned there are additional subclasses below these major subdivisions. This is shown in Figure 4, along with commonly used classes from the Information Artifact Ontology (IAO).

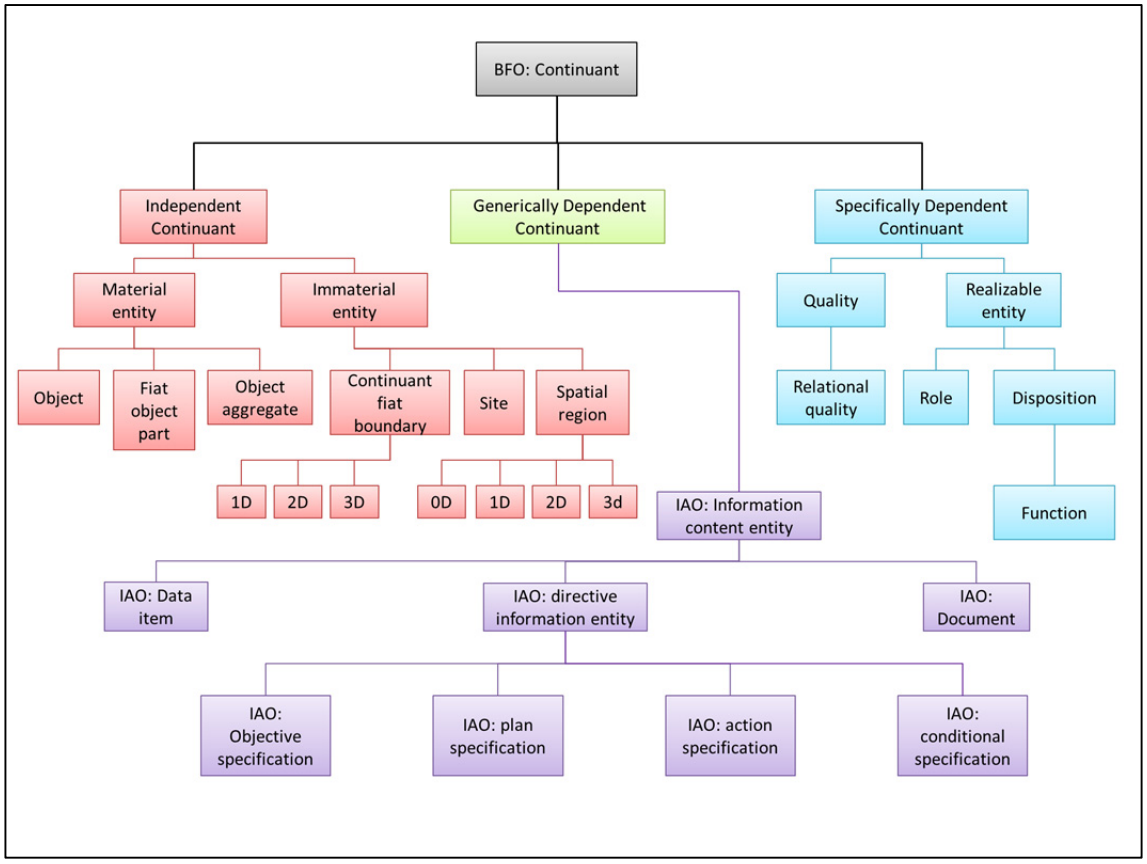


Figure 4. BFO continuant structure. Labels in grey, red, light green, and blue are formally part of the BFO. Labels starting with 'IAO:' in purple indicate commonly used classes from the Information Artifact Ontology.

Independent continuants are entities that do not depend on anything else for their existence. They may gain or lose parts but maintain their identity. A human being or an automobile are examples of independent continuants. Specifically dependent continuants exhibit existential dependence on an independent continuant (referred to as *a bearer*), that is, they don't exist unless the associated independent continuant exists. Examples of a specifically dependent continuant are the color of an object or the role of a person in a corporation. Generically dependent continuants are continuants that can migrate from one

bearer to another; they are often used to represent information in some form or another. Arp, in the aforementioned text (Arp et al., 2015), uses the example of the PDF file moving from computer to computer as an example of a generically dependent continuant, with the file representing information, and the computer the independent continuant.

3.1.1.2 Occurrents. Occurrents are entities that happen or occur in time. For example, the process of an institutional review board (IRB) assessing a research study for risk to a human subject would be a *BFO:Process*, the process would likely be informed by an *IAO: plan specification* for the process, and a standard operating document represented as an *IAO: action specification*. Figure 5 shows the BFO occurrent hierarchy. Also indicated in this diagram are classes from the Ontology for Biomedical Investigations (OBI) and the IAO that were deemed important for an ontology representing the Common Rule, or were frequently used in the CRO.

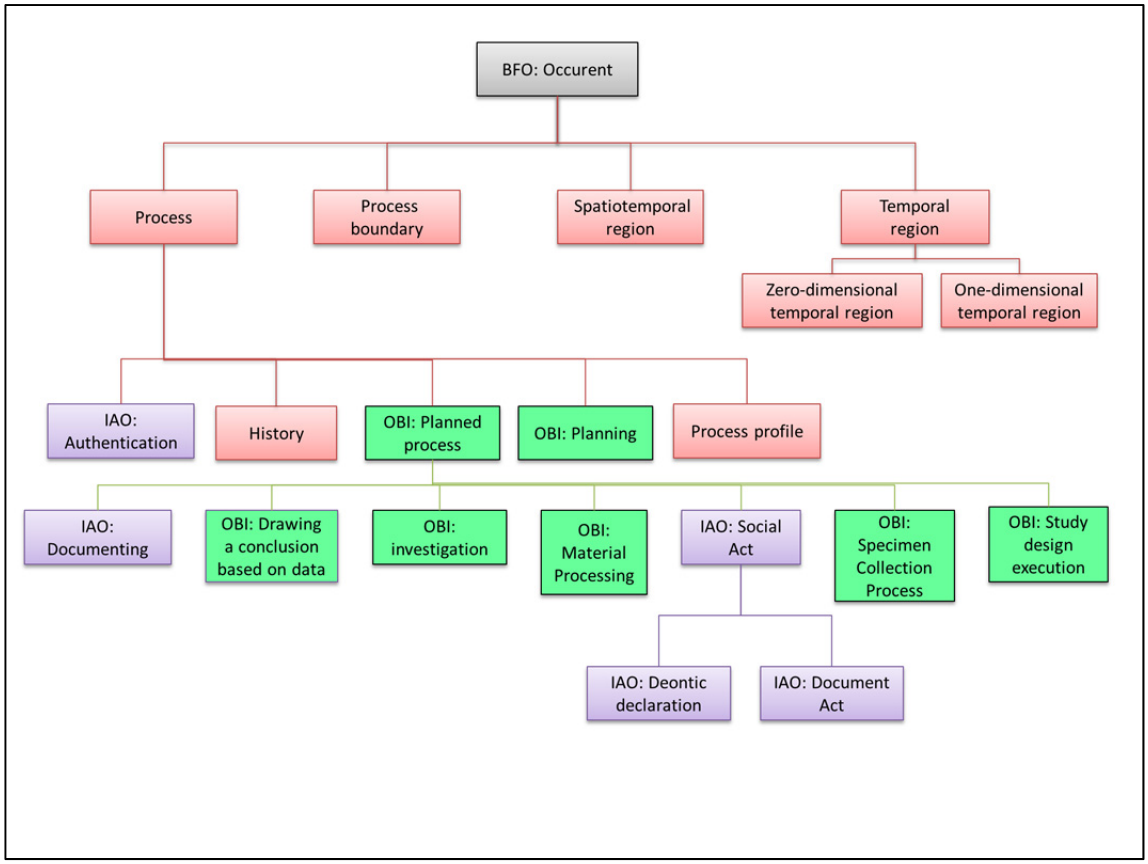


Figure 5. BFO occurrent structure. Boxes that are grey or red are parts of the BFO. Labels starting with 'IAO:' in purple indicate commonly used classes from the Information Artifact Ontology. Labels starting with 'OBI:' in green are classes from the Ontology for Biomedical Investigation that were frequently used in the CRO.

3.1.2 The Open Biomedical Ontologies Foundry. The Open Biomedical Ontologies (OBO) Foundry is an open community of developers of life-sciences ontologies (Smith et al., 2007). OBO uses a set of shared principles, coordinated terms and identifiers, well-defined syntax, and upper-level ontologies to ensure OBO Foundry projects can interoperate. The structure of classes and axioms that OBO Foundry ontologies are based on are aligned with the BFO model. This helps to ensure that any

two ontologies that follow recommended OBO Foundry construction practices and principles will generally be interoperable; although it should be noted that context between any pair of ontologies in the OBO Foundry can be different. Consequently care still needs to be taken when incorporating terms from one OBO Foundry ontology with another. Relations in OBO Foundry ontologies are typically expressed in the Relation Ontology (RO), providing a common semantic glue for linking assertions about relationships between objects (Smith et al., 2005). Consequently, OBO Foundry-based ontologies are generally interoperable with one another.

The OBO Foundry encourages ontology development based on a small set of top- and mid-level ontologies and based on the set of principles shown below in Table 4. Adherence to these principles is part of what ensures interoperability of OBO Foundry-based ontology.

Table 4

OBO Foundry Principles

Principle	Description
Openness	The ontology must be openly available.
Common Format	Defines the specific technical representation language to be used, such as RDF/XML.
URI/Identifier Space	It is required that each class and relation in the ontology have a unique Universal Research Identifier (URI); that this URI has a unique prefix; and be represented in a numeric form with no semantics interpretable by humans.
Versioning	Each change to the ontology must be versioned.
Scope	The ontology has a clearly specified scope.
Textual Definitions	A substantial portion of the ontology must have definitions on the terms, and each term must be unique with regard to definitions.
Relations	Relations should be clearly defined and congruent with the Relations Ontology (RO).
Documentation	The ontology should have as much documentation as possible. These are often provided as metadata annotation properties in the ontology itself.
Documented plurality of users	The ontology should provide evidence that multiple groups are interested in using it.
Commitment to collaboration	There must be evidence of collaboration with other OBO Foundry developers.
Locus of authority	There should be a single person in control of the ontology.
Naming conventions	Defines specifications for the manner in which terms in the ontology are represented, and what annotation properties to use.
Maintenance	There should be evidence that the ontology is actively maintained.

Note: Descriptions in this table are only a short summary of the requirements for ontology in the OBO Foundry. Those seeking to develop an ontology for inclusion in the Foundry should carefully examine the requirements found at http:obofoundry.org.

In addition to BFO, the core of the OBO Foundry consists of the Relation Ontology (RO), and the Information Artifact Ontology (IAO). The RO is an ontology that contains core OWL relationships with a heavy emphasis on biological relationships (OBO Foundry, July 3, 2015/2017). The IAO is an ontology that describes various features of

how basic information is represented and transmitted (OBO Foundry, July 31, 2015/2017). For example, it contains classes that represent documents, parts of documents, and types of information.

Ontologies in the OBO Foundry are generally considered to be *orthogonal* to one another. The OBO Foundry encourages reuse of ontology terms from other OBO Foundry ontologies to the extent possible.

3.1.3 The Ontology for Biomedical Investigations. The Ontology for Biomedical Investigations (OBI) (Brinkman et al., 2010) is an important mid-level ontology in the OBO Foundry. OBI is a project to develop a set of biomedical ontologies all aligned with the Basic Formal Ontology upper-level ontology. OBI was developed to provide an integrative framework that defines terms, relationships, and objects commonly used in many biomedical and clinical investigations. Current components of OBI were developed from 19 international biomedical communities. OBI also represents phases of experimental processes, including entities such as study designs, protocols, instrumentation, biological material, data collection, and analysis. Because of this expressivity in the domain of biomedical research, it is a good fit for work on aspects of the Common Rule. My preliminary work on the Ontology of Informed Consent (ICO) already uses this framework (Lin et al., 2014; F. J. Manion et al., 2014).

3.2 Brief Survey of Ontology Development Lifecycle Approaches

Ontology construction is generally seen as an iterative development process, much in the way software systems are developed (Noy & McGuinness, 2001; Sanya & Shehab, 2015). Methods by separate authors have classified the ontology development lifecycle in

different manners, but all suggest similar steps that can often be readily mapped one to another. Many of these different classifications often appear to be a result of the level of granularity with which a particular aspect of the construction process was perceived. In the next few sections, I briefly review the lifecycle steps in suggested by different authors. Note that most of these guidelines do not provide specific task-level steps, but rather provide only a general framework to work from in constructing an ontology. This material is presented here because it was important to apply best practices to the extent possible in knowledge acquisition and evaluation during the construction of the Common Rule Ontology. The topic of the ontology development lifecycle will be returned to later in this chapter, and again in the chapters discussing evaluation methods.

3.2.1 The Uschold model. Uschold (Uschold & King, 1995) proposed a four-stage methodology that includes (a) identifying the purpose of the ontology; (b) building the ontology; (c) evaluating the ontology; (d) documenting it. They defined important sub-steps needed for the effective building of the ontology as: capture of the ontology by identifying key concepts and relationships, producing definitions, naming these terms, coding these terms in some formal language, and integrating other existing ontologies. This framework has been criticized for its lack of evaluation criteria throughout the lifecycle (Sanya & Shehab, 2015).

3.2.2 The Noy lifecycle model. The lifecycle stages proposed by Noy (Noy & McGuinness, 2001) are completely iterative and include determining the domain and scope of the ontology, reusing other ontologies, enumeration of important terms in the

ontology, defining the class hierarchy, defining class properties, determining facets of the class properties, and creating instance data.

3.2.3 The Pinto lifecycle model. Pinto (Pinto & Martins, 2004) described five lifecycle stages consisting of specification, conceptualization, formalization, implementation, and maintenance. They noted that knowledge acquisition, evaluation, and documentation should be incorporated throughout the entire process of constructing an ontology.

3.2.4 The NIST lifecycle model. During the 2013 Ontology Summit organized by the National Institute of Standards and Technology (NIST), which is an important annual workshop in the field, participants developed an unnamed ontology lifecycle model (Neuhaus et al., 2013). This model proposes eight stages. The authors are careful to note that the activities in their proposed model can occur in sequence or in parallel. They also propose general evaluation criteria to be used in both formative and summative fashion throughout each phase. Due to the complexity of this method, lifecycle stages are shown in Table 5.

The first phase, *requirements development,* defines the need for the ontology, its scope and use, competency questions the ontology needs to support, and an analysis of the groups the ontology must support, along with their intended use of the ontology.

The *ontological analysis phase* is where important elements such as classes, individuals, and the relationships among them are captured. This phase is also where the terminology used to express the names or labels within the ontology are developed and linked to common vocabulary in the relevant community of practice. The separation of names from

the underlying terminology is an important concept and proved to be important in the evaluation of the CRO.

In the *ontology design phase,* top-level classes are developed, and the choice of technical representation language (such as XML, RDF, or OWL-2) is decided on. During this phase, the design team will choose whether or not to align with popular upper-level ontology such as the Basic Formal Ontology (BFO), the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), or the Suggested Upper Merged Ontology (SUMO).

In the *system design phase,* technical issues involving the implementation within an information system such as an individual application running on a single computer or an enterprise or cloud-based application environment are decided on.

Conference participants defined the *ontology development phase* as having an iterative structure consisting of four activities: *informal modeling*, *formalization of competency questions*, *formal modeling*, and *operational adaption*. Informal modeling is typically a refinement and extension of work done in the ontology analysis phase. Formalization of competency questions involves assessing them for completeness, translating them to the necessary query language, and evaluating the translation for correctness. Formal modeling is simply the transformation of the informal model into an ontology language such as OWL-2 DL.

In *operational adaption*, the results of formal modeling are put in practice and assessed for performance characteristics.

The *system development and integration phase,* the *deployment phase,* and the *operation and maintenance phase* are all concerned with the manner in which components of the ontology developed in earlier phases are integrated into a functioning information system, deployed in practice, and supported and maintained. They are similar to phases involved in the operational deployment of any information system, except the major focus is on the ontological components of the system.

Five high-level evaluation criteria were proposed at the NIST workshop, although no specific methods were proposed to implement these. The five criteria are *intelligibility*, *fidelity*, *craftsmanship*, *fitness*, and *deployability*. Intelligibility refers to the ability of humans to correctly interpret the ontology. The authors (Neuhaus et al., 2013) indirectly suggest that having experts review sentences generated by the ontology is a potential method to assess *intelligibility*. *Fidelity* is intended to measure if the ontology is an accurate model, within its scope, of the domain being modeled. *Craftsmanship* refers to whether good ontology design and technique have been followed. *Fitness* refers to 'fitness for use' of the ontology, i.e., does it fulfill the stated requirements for use. *Deployability* refers to the ability of the ontology to be successfully implemented in the information system that it was designed to be part of. The authors note that evaluation should take place in varying degrees throughout the entire ontology lifecycle.

Table 5

Summary of the Ontology Lifecycle as Defined at the 2013 NIST Ontology Summit

Phase	Major Tasks	Evaluation
Requirements development	Define need for ontology Define scope of domain Define resources Define use of existing standards or ontology Define competency questions	None suggested.
Ontological analysis	Identify key entities and name them using terminology that is understood by domain experts	*Intelligibility*. Sentence-based evaluation measures done by domain experts and ontology experts was suggested. Completeness measures of the ontological analysis.
Ontology design	Determine top-level classes Decide on use of upper-level ontology Decide on implementation language	Structural evaluation that all classes and instances are expressed and belong to high-level concepts.
System design	Decide on technology issues	Derived from systems engineering and considered out of scope.
Ontology development phase (four sub-phases)		
Informal modeling	Refinement and extension of work done in the ontology analysis phase	*Intelligibility*. Completeness measures of the ontological analysis. Conciseness and clarity-based measures.
Formalization of competency questions	Translating competency questions into the necessary query language	Assess the set of questions for completeness. Evaluate the translation for correctness
Formal modeling	Transforms the informal model into a formal model language such as OWL-2	Evaluate for *fidelity* Evaluate for *craftsmanship* Evaluate for *fitness* for purpose
Operational adaption	Modify the ontology as required to put into operation	Assess for performance, precision, recall, etc.
System development and integration	Build overall information system including ontology-based components	Assess for successful integration of ontology. Assess that system achieves all requirements related to ontology.

Phase	Major Tasks	Evaluation
Deployment	Deploy completed system into production	*Deployability*, regression testing, risk/benefit analysis.
Operation and Maintenance	Sustain operations	Operational monitoring and reporting.

Comparison of the lifecycle models. Table 6.

3.2.6 *Comparison of Ontology Lifecycle Models against the NIST Model from the 2013 Ontology Summit* compares how the four ontology lifecycle models surveyed relate to each other. Because the unnamed model from the 2013 NIST Ontology Summit was the most granular, was constructed by noted experts in the field working in concert, and was the newest lifecycle model reported, it is used as the standard to which others are compared to.

Table 6

Comparison of Ontology Lifecycle Models against the NIST Model from the 2013 Ontology Summit

NIST Phases	Uschold Phases	Pinto Phases*	Noy Phases
Requirements development	Identify purpose	Specification	Domain and scope Competency question development Reuse of other ontology
Ontological analysis	Ontology construction: ontology capture	Conceptualization	Enumerate terms Define classes Define class hierarchy.
Ontology design	Ontology construction: ontology capture	Conceptualization	Define classes Define class hierarchy.
System design			
Ontology development	Ontology construction:	Conceptualization	Define classes Define class hierarchy.

NIST Phases	Uschold Phases	Pinto Phases*	Noy Phases
	capture & coding		
Informal modeling	Ontology construction: ontology capture	Conceptualization, Formalization	Define classes Define class hierarchy.
Formalization of competency questions	Evaluation		Competency question development
Formal modeling	Ontology construction: coding Evaluation	Formalization, Implementation	Defining class properties. Determine class properties facets. Create instance data.
Operational adaption			
System development and integration			
Deployment			
Operation and Maintenance		Maintenance	
(happens throughout)	Documentation		

*Note: All phases of the Pinto model call for knowledge acquisition, evaluation, and documentation. The NIST model calls for evaluation and documentation in almost every phase, while the knowledge acquisition steps are most granular and specific. Thus these two models are heavily aligned in their general design philosophy. They are also the two most recent models surveyed.

As can be seen from Table 6.

Comparison of Ontology Lifecycle Models against the NIST Model from the 2013 Ontology Summit, the lifecycle stages in each of these models have rough correspondence to each other. Unsurprisingly, much of this overlap relates to requirement capture, knowledge acquisition, conceptualization, the incorporation of other formally modeled knowledge sources (e.g., ontologies), the technical considerations of construction of the

ontology into a language such as OWL-2, and the evaluation and documentation of the ontology.

3.3 Ontology Construction Methods

Ontology lifecycle frameworks generally provide high-level guidance to the ontology engineer. Ontology construction methods, on the other hand, define detailed activities and tasks that can be used by the practicing ontologist in constructing an ontology. There is considerable overlap between an ontology construction lifecycle model and ontology construction methods themselves. Some authors (Corcho, Fernández-López, & Gómez-Pérez, 2003; Sanya & Shehab, 2015) for example, consider the work of Uschold (Uschold & King, 1995) an ontology development method even though it only addresses high-level guidelines and lacks detail about knowledge acquisition and other tasks. After a short literature review (Ahmed, 2011; Corcho et al., 2003; Öhgren & Sandkuhl, 2005; Sanya & Shehab, 2015), I settled on the use of METHONTOLOGY (Fernández, Gómez-Pérez, & Juristo, 1997) to guide the technical construction of CRO ontology.

3.3.1 The METHONTOLOGY construction method. This method was developed at the Technical University of Madrid in their Artificial Intelligence Lab and is one of the more mature ontology development frameworks. It is detailed in both the tasks to be accomplished and the procedures to use and is suited to 'construction of ontologies at the knowledge level' (Corcho et al., 2003; Sanya & Shehab, 2015). To guide the construction of an ontology METHONTOLOGY leads the developer through a series of seven phases with associated tasks and strategies, including construction of a glossary of

terms, creating concept taxonomies and converting them to classes and subclasses, constructing relationships and axioms, describing rules, and finally defining value sets for the ontology. The details of these steps are the most comprehensive in the areas of *specification*, *knowledge acquisition*, and *conceptualization*, with other steps such as *evaluation* only offering general guidance. Table 7 introduces the phases and tasks for METHONTOLOGY.

Table 7

METHONTOLOGY Ontology Development Phases and Tasks

Phase	Suggested Tasks	Suggested Outputs
Planify	No detailed guidance was given	
Specification	Define: a) The purpose of the ontology. b) The level of formality of the ontology. c) The scope of the ontology. d) Sources of knowledge to be used.	A formal specification document
Knowledge Acquisition	Non-structured interviews with experts. Informal text analysis of various corpora. Structured interviews with experts. Formal text analysis. Preliminary glossaries of terms. Reviews of other ontologies.	
Conceptualization	Construct glossaries of terms. Classify concepts and identify verbs. Classify relationships between concepts and verbs. Data dictionaries. Construct tables of class attributes. Construct tables of instances. Construct tables of constants. Classify attributes using classification trees. Construct verb diagrams. Construct a dictionary of verbs. Construct tables of conditions. Merge the results of the conceptualization into: • a table of formulas; and • a table of rules.	Glossary of terms Classification hierarchy Data dictionary Table of class attributes Table of instances Tables of constants Attribute hierarchy Verb diagrams Verb dictionary Table of conditions Table of formula Table of rules
Integration	Inspect upper level or meta-ontology for terms. Inspect other ontology for terms. Develop alignment with upper- and mid-level	Table of included terms

	ontologies and incorporate terms and classes from other ontologies as appropriate.	
Implementation	Construct the ontology in a formal language.	Ontology representation in OWL-2 or another format.
Evaluation (Formal)	Use verification techniques to verify the technical correctness of the ontology. Use validation techniques to assess the fitness for use of the ontology.	Evaluation document
Documentation	Document each phase of the METHONTOLOGY process.	Documentation set
Maintain	No specific guidance was given.	

While overall it appears that these sequences of tasks are linear, the authors note that their method is not intended to impose a strict order on the development process, but rather merely the steps to be carried out. In fact, the authors propose a lifecycle model for development that they term 'evolving prototype', which allows the development team to move between steps as required. They also note that knowledge acquisition, evaluation, and documentation are steps that should be carried out throughout the construction process.

3.3.2 Correspondence of the NIST lifecycle model with METHONTOLOGY.

It should be readily apparent from the description of the NIST lifecycle model and the METHONTOLOGY construction method that they are well-aligned. For the convenience of the reader, Table 8 shows the correspondence of the two.

Table 8

Correspondence of the NIST Lifecycle Model and METHONTOLOGY Phases

NIST Phase	Corresponding METHONTOLOGY Phases
Requirements development	Planify Specification
Ontological analysis	Knowledge Acquisition Conceptualization
Ontology design	Knowledge Acquisition Conceptualization Integration
System design	
Ontology development (four sub-phases)	
Informal modeling	Conceptualization
Formalization of competency questions	Specification
Formal modeling	Implementation Integration Evaluation
Operational adaption	
System development and integration	Maintenance
Deployment	Maintenance
Operation and Maintenance	Maintenance
(Documentation happens throughout)	Documentation

3.3.3 The merged lifecycle, construction, and evaluation process for the CRO.

Development of the CRO took place along the spectrum of the NIST ontology lifecycle, with specific tasks from METHONTOLOGY being used to create specific artifacts. Throughout the process, which was highly iterative, formative evaluation was used to refine knowledge collected from domain experts, and to assess the ontology as it was being constructed. Table 9 shows the NIST lifecycle stage used, the METHONTOLOGY tasks used, and the evaluation methods used at each step.

Table 9.

NIST/METHONTOLOGY (MO) Stages and Evaluation Methods

NIST and MO Stage	Artifacts	Evaluation Methods
Requirements Development		Establish face validity of concept maps (*fidelity, fitness*).
Planify	Structured and unstructured interviews with domain experts. Concept Map One of the domain of regulated human subjects research.	Critique and iterative refinement of Concept Map One with domain experts
Specification	Requirement document describing the level of formality, scope, main documents and ontologies to be included. Initial competency questions.	n/a
Ontological Analysis		n/a
Knowledge Acquisition	Initial text review of the 1991 and 2017 Common Rule texts. Deconstruction of the Common Rule and Notice of Proposed Rule Making for the 2017 revision of the Common Rule. Review of other ontologies containing terms of relevance.	n/a
Conceptualization	Concept Map Two: preliminary glossary of top-level terms, relations, axioms and other characteristics.	Review and documentation of deconstructed elements against specific sections of the complete Common Rule text (*fidelity, fitness*)
Ontology Design		Iterative reviews with both domain and ontology experts (*intelligibility*, *fidelity, fitness*).
Knowledge Acquisition	Refined glossaries of terms.	Per above.
Conceptualization	Concept Map Three: Classification according to knowledge constructs	Per above.
Integration	Initial alignment with BFO and OBI upper- and mid-level ontology.	Per above.
Ontology Development		
1. Informal modeling		

NIST and MO Stage	Artifacts	Evaluation Methods
Conceptualization	Refining of Concept Map Three and classification hierarchy Determination of top-level formal classes Extension and refinement of alignment of classes with BFO hierarchy	Further iterative reviews with both domain and ontology experts (*intelligibility, fidelity, craftsmanship, fitness*).
2. Formalization of competency questions		Derived from University IRB screening questions for consent forms (*fidelity, fitness*).
Specification	Determine final sources of competency questions and representation mechanisms within the CRO ontology	Per above.
3. Formal modeling		
Implementation	Creation of OWL2 RDF/XML representation developed with Ontorat and the Protégé editor.	FOCA analysis (*craftsmanship, intelligibility*) Natural Language sentence analysis (*fidelity, fitness*).
Integration	Alignment of all class and property structures with BFO and major OBO Foundry structures. Incorporation of terms from other OBO-Foundry Ontologies. Structural error checking.	Iterative checks for logical *satisfiability* with HermiT 1.3.8.413 reasoner Inherent *fidelity* from the use of BFO framework.
Evaluation	Summative evaluation procedures.	Evaluation by Competency Questions (*fidelity*, *fitness*) Burton-Jones Analysis (*intelligibility*, *fidelity, craftsmanship, fitness*). Corpus assessment with domain experts (*fidelity*, *fitness*).

Note: Stages not used are not shown. For example, this dissertation does not cover Operation and Maintenance lifecycle stage of the CRO in a production capacity.
In practice, all phases and sub-phases are iterative, as is the lifecycle as a whole.
The italicized words in the evaluation criteria are the aspects of the ontology being measured as taken from the NIST methodology (Neuhaus et al., 2013).

3.4 Concept maps

Concept maps were used as a major vehicle for capturing terms, concepts, and relationships during knowledge acquisition and synthesis. Concept maps generally consist of a set of major concepts from the subject being represented as nodes in a graph, along with arcs (directed or not) representing the relationships between the major concepts.

3.5 Details of the Construction of the Common Rule Ontology

The overall approach to developing the CRO was to use a series of concept maps based on and constructed from the text of the Common Rule, coupled with successive interviews with domain experts. Three concept maps were used as part of this work. Table 10 shows the description of these maps, their major organizational unit, and the phase of the NIST model informed by the map.

Table 10

Concept Maps Used in the Development of the CRO

Concept map	Description	Organization unit	Used in NIST phases
Concept Map One	General representation of concepts and processes involved in human subjects research	Nodes are research concepts, arcs are generally processes	Requirements development
Concept Map Two	Decomposition of the Common Rule for terms, concepts, potential axioms, relationships	Organized by section of the Common Rule, *e.g.* §46.116(a)	Ontological analysis
Concept Map Three	Terms and relationship typology	Organized in a typology of knowledge constructs	Ontology design, ontology development

Artifacts developed using this approach formed parts of the overall merged NIST / METHONTOLOGY construction method, with ongoing evaluation based on formal validation methods where feasible.

3.5.1 Requirements development: 'planification.' The goal of this phase was to develop an understanding of how the Common Rule and informed consent apply in the area of regulated human subjects research involving collection and distribution of specimens and data.

Two of us (Harris, Manion) constructed a preliminary concept map, which is termed 'Concept Map One,' of the regulated research domain, including biorepository research. Using interviews with domain experts the concept map was refined and face validity was

established. This map was used as the initial basis for guiding discussions with domain experts, and for developing preliminary classes for the CRO. The map examined how various processes and information artifacts (e.g., specimens, data, documentation) related to one another. XMind Version 7.5 (XMind Ltd, 2016) mind mapping software was used to capture the results of the interviews and represent the resulting model. A portion of this concept map is shown in Figure 6; the full concept map can be found in Appendix B.

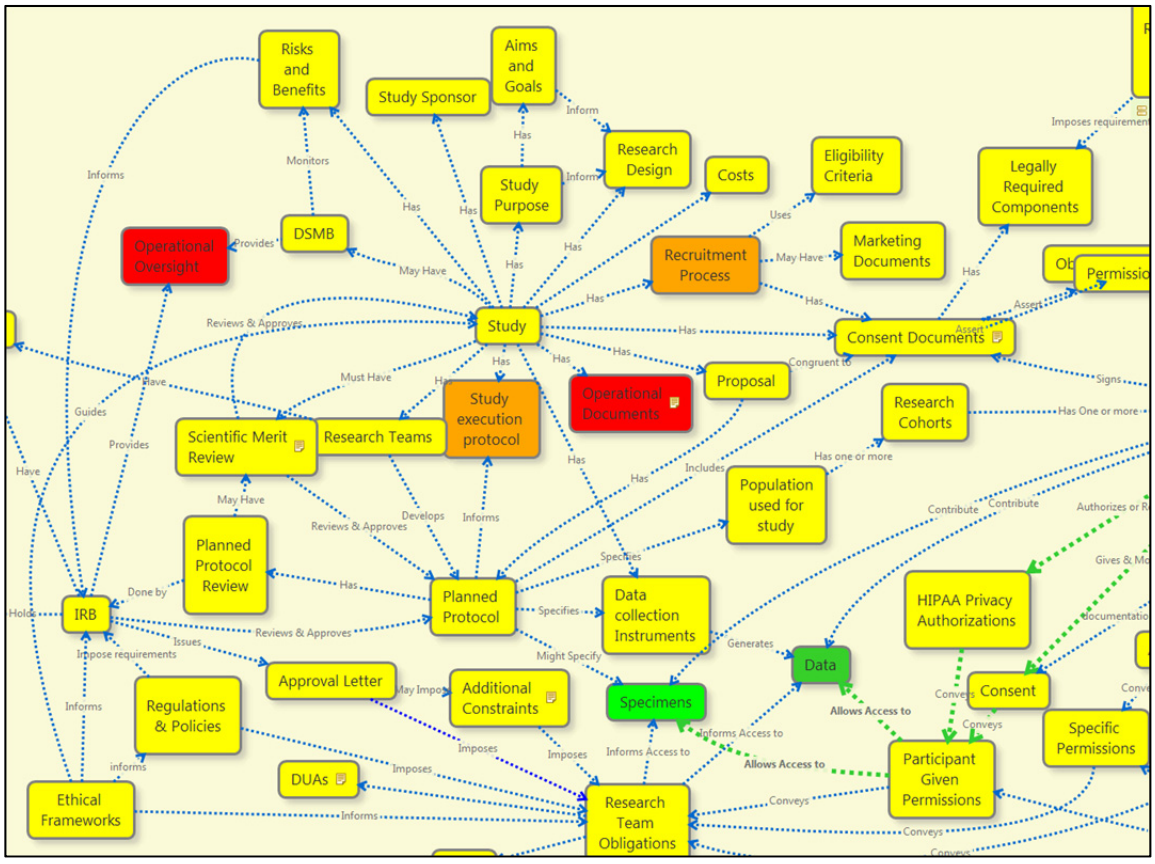


Figure 6. A portion of Concept Map One examining the domain of regulated research.

The concept map was then shown to four domain experts, as shown in Table 11. These individuals were asked if the model appeared accurate, and they were asked if they could 'break' the model by examining the major concepts and relationships between them. A common technique used in the informal validation of concept maps for face-level validity is to see if sentences that a domain expert considers accurate can be constructed by selecting a path through the map. For example, as indicated in Figure 6 a 'planned protocol' 'specifies' the 'population used for (the) study' which 'has one or more' 'research cohorts,' and so on. The model was subsequently refined based on comments received from these individuals to incorporate the changes suggested.

Table 11

Roles of Domain Experts Interviewed in the METHONTOLOGY 'Planification' Phase

The chairman of a health and behavioral science IRB that oversees large-scale biorepositories.
The director of a large-scale institutional biorepository.
A clinical researcher involved with procurement and use of biospecimens.
A university-level senior administrator involved in research compliance activities.

3.5.2 Requirements development: specification. The CRO ontology is intended to be a faithful representation of the U.S. Common Rule, 45 CFR § 56 subparts A through D. To develop the specification for the ontology the results of the literature review and ontology library analysis reported in Chapter 2 were used to develop the preliminary purpose, scope, and level of formality of the ontology. As described there, this consisted of a manual review of a number of ontologies purporting to have coverage of the research lifecycle (Bandrowski et al., 2016; Brinkman et al., 2010; Sim et al.,

2014), informed consent for clinical research (Lin et al., 2014, 2016), or clinical care (Health Level Seven International, 2014, p. 7), data use (Courtot, 2017), biospecimen sharing (Brochhausen et al., 2013b, 2016), and research permissions (Grando & Schwab, 2013). I also drew on direct personal experience working with groups developing ontologies for the OBO Foundry. It was a specific aim (specific aim four) of this proposal that CRO provide a knowledge base to integrate with OBO Foundry ontologies, and provide a basis for a meta-data ontology for informed consent in biobanking in support of a U01 award at the School of Biomedical Informatics. Consequently, it was decided that the level of formality for CRO should be 'formal', that it would use the Basic Formal Ontology (BFO) as an upper-level ontology, and would incorporate classes, terms, and relations from other ontologies in the OBO Foundry.

3.5.3 Ontological analysis: knowledge acquisition. The 1991 version of the Common Rule was reviewed to develop an initial understanding of the various subparts and sections of the regulations. During the time this work was being done it was not clear if the draft of the 2017 version of the Common Rule would be approved. Consequently work was initially restricted to the 2007 version. After the 2017 revision was approved and published in the Federal Register in January of 2017, the revised text was reviewed for substantive changes to the Common Rule, with a focus on those impacting the area of broad consent for secondary research involving specimens and data. It should be noted that the revision of the Common Rule only impacts subpart A, which is the 'basic HHS policy for protection of human research subjects'. The new regulations do not impact

subparts B, C, or D, which deal with research on fetuses, pregnant women, and neonates; research on prisoners; and research on children; respectively.

Each major section and high-level concept derived from the preliminary review of the Common Rule was entered into a node on a concept map, which I term 'Concept Map Two.' This concept map was organized by the legal section of the Common Rule and represented the initial coding of the Common Rule. Terms from other ontologies describing consent or regulatory processes, as described and reported on in Chapter 2, were included and those that were considered important were recorded. During the ontology development phase, these were incorporated into the CRO.

3.5.4 Ontological analysis: conceptualization. A preliminary glossary of top-level terms, properties, axioms, instances, and other characteristics was created. This was done by manually decomposing the complete text of Common Rule Subparts A through D into key concepts. The results were iteratively reviewed by three of us (Manion, Tao, Harris). The resulting phrases were added to Concept Map Two. Potential formal properties, axioms, equivalence classes, and instances were also captured in the concept map, resulting in a preliminary glossary of terms. An example of the decomposition at this stage of development is shown in Figure 7. Items in the concept map were also recorded into a database implemented in Microsoft Excel.

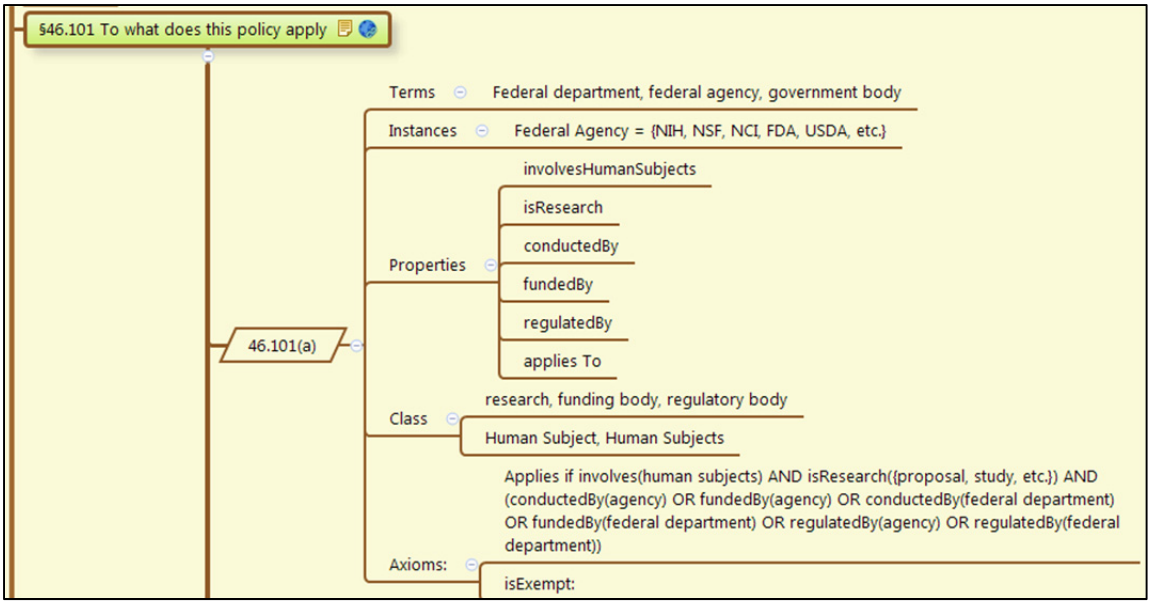


Figure 7. Formal decomposition of Common Rule text using concept maps.

During the decomposition, the legal section of the Common Rule, and the definition that was the source of the term was also recorded in the Excel database. For example, the definition of an institutional review board as defined in §46.102(g) was recorded in the database as:

Legal section: §46.102(g)**,**

Definition: 'IRB means an institutional review board established in accord with and for the purposes expressed in this policy.'

This database was used later on in the construction process to programmatically generate the terms and class hierarchy for the preliminary version of the ontology. It was also used to attach OWL annotation properties to each class and relation in the ontology.

‘Definition’ was mapped to the annotation property *‘definition’ (IAO:0000115)*, and ‘Legal section’ was mapped to annotation property *‘definition source’ (IAO:0000119)*.

3.5.5 Ontology design: conceptualization. At this point, a third concept map, Concept Map Three, was constructed. This third concept map was no longer organized by Common Rule section or text but rather was organized as a typology according to the underlying concepts and associated terms. Figure 8 shows a small example of the concept map. Note that while this hierarchy was a typological representation of the concepts from the Common Rule, it was not yet aligned with any ontology. At this point, the Excel database was updated to reflect the new classification structure, including for each term the parent term in the hierarchy.

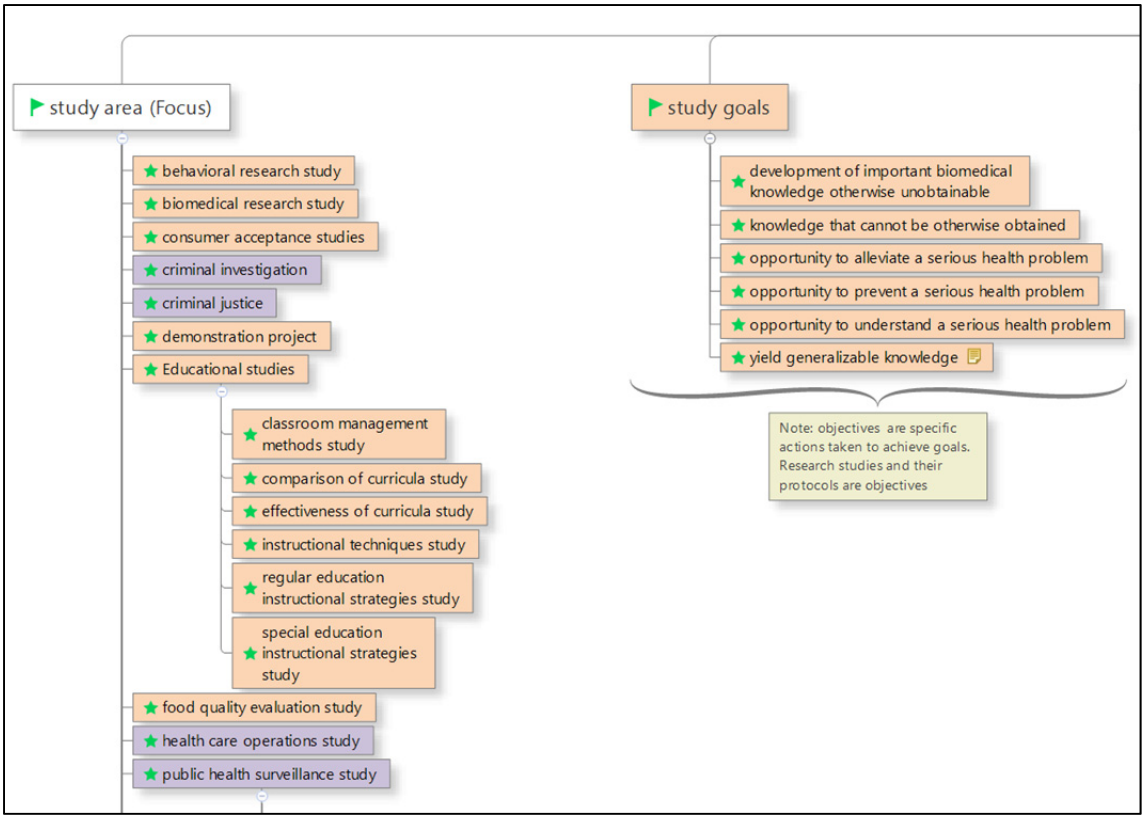


Figure 8. Example of Common Rule concept map showing typological classification.

The term glossary was reviewed and used to develop a series of dependencies associated with each term and concept in the hierarchy. For example, for each individual or organizational group that was mentioned as having specific roles in the performance of duties in the Common Rule, the corresponding role was recorded.

For each term, the following information was recorded:

1) the term name,
2) a unique identifier,
3) the major category in the hierarchy,
4) the definition,

5) the definition source,
6) the parent term in the hierarchy,
7) roles associated with the term,
8) where risks or benefits related to the term accrue to, and
9) qualities of the term.

3.5.6 Ontology design: integration. A specific aim (specific aim four) and design goal of this work was that the CRO ultimately be integrated with the BFO, the OBO Foundry, and the OBI; and to reuse terms, classes, relations, axioms, and other material from these ontologies to the extent possible. This choice was made to leverage the existing work done within the OBO Foundry on representing biobanking processes, as well as to leverage our own initial work in the form of the Informed Consent Ontology (ICO). Consequently, it was necessary for the CRO to be considered a candidate for inclusion into the OBO Foundry, which is highly desired for dissemination to the community.

During this phase, the12 high-level concepts and some of the underlying terms in the initial taxonomy were assigned to classes in the BFO, IAO, and OBI hierarchies to create an initial alignment. For example, the major classification of 'people' was assigned to *BFO:object*, which is a subclass of *BFO:material entity*, which in turn is a *BFO:independent continuant*.

3.5.7 Ontology development/informal modeling: conceptualization. Two of us (Harris, Manion) conducted a secondary review of the Common Rule to develop top-level 'design patterns' for the ontology. This review was intended to refine top-level

classes against which all concepts from the Common Rule could be placed as ontology classes.

3.5.7.1 Concept map refinement. The initial class hierarchy was refined to improve the alignment with the major elements of the Common Rule. Two of us (Manion, Harris) performed successive rounds of refinement against Concept Map Three until the hierarchy as expressed in the concept map was stable and an accurate representation of the concepts in the Common Rule. This resulted in a preliminary class hierarchy for the initial ontology.

3.5.7.2* I*terative evaluation and refinement of Concept Map Three by domain experts. Five domain experts in regulatory affairs, IRB management, and compliance were recruited to do a semi-structured qualitative review of Concept Map Three. The roles of the persons interviewed are listed in Table 12. Because of the size of Concept Map Three, the map was broken into its 12 major constituent parts and these were viewed individually, along with a top-level diagram describing these major parts. The concept maps used during these interviews are shown in Appendix D. Changes to the concept maps suggested by each interviewee were incorporated into the next round of interviews until a saturation of term placement was achieved.

Table 12

Roles of Domain Experts Interviewed for Review of Concept Map Three

An Associate Director of a Center for Bioethics and Social Sciences in Medicine
The Director of Regulatory Affairs for a large biorepository facility in a medical school
A faculty member who studies informed consenting processes
The chairman of a medical school institutional review board
An Associate Dean for Regulatory Affairs at a medical school

3.5.8 Ontology development/formalization of competency questions: specification. Formal competency questions were developed and refined by including a checklist obtained from the University of Michigan Medical School IRB at 'https://research.medicine.umich.edu/sites/default/files/res_irbmed_Informed Consent Checklist.doc.' These questions were initially derived from the Department of Health and Human Services Office of Human Research Protections as taken from 45 C.F.R. §46.116 and from the Food and Drug Administration regulations 21 C.F.R. §50.25. The resulting competency questions are shown in Table 31, below. The original source material from the Michigan IRB is found in Appendix A.

3.5.9 Ontology development/formal modeling: implementation. At this point, ontology development continued using platforms and tools for creating and working with ontologies containing semantics. In general, the interactive Protégé ontology editor developed at the National Center for Biomedical Ontologies (NCBO) at Stanford University (Gennari et al., 2003; Musen et al., 2012) and shown in Figure 4 was used.

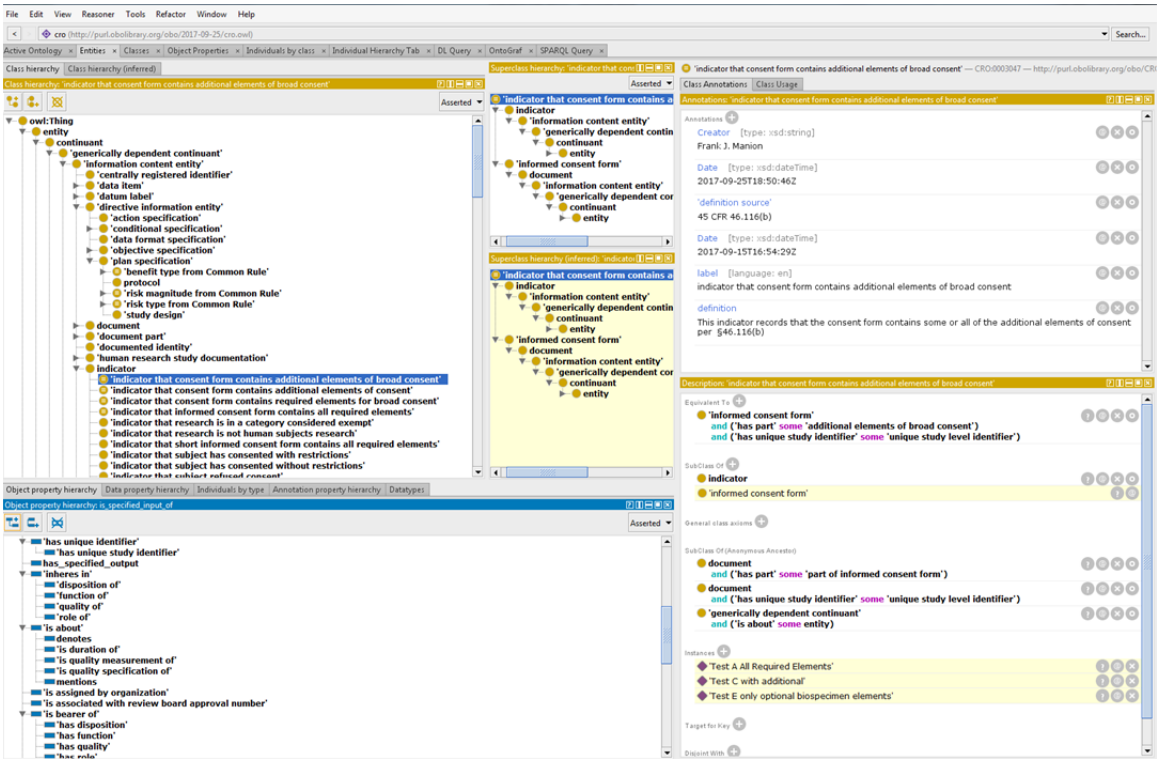


Figure 9. The interactive Protégé ontology workbench showing parts of the CRO.

Several of the ‘Onto-animal’ tools developed in the He group at the University of Michigan were also employed. These were Ontobee, the official search engine of the OBO Foundry ontology repository (Xiang et al., 2011); Ontorat, a program which generates ontology terms using design patterns and user provided templates (Xiang, Zheng, Lin, & He, 2015); and Ontofox, which is a software implementation of the MIREOT methodology of Courtot and colleagues (Courtot et al., 2009; Xiang, Courtot, Brinkman, Ruttenberg, & He, 2010). MIREOT stands for the ‘Minimum Information to Reference an External Ontology Term.’ The method is used to import only the transitive closure of classes, properties, and relationships actually required when including terms

from other ontologies. It was developed to prevent the need to import very large external ontologies when referencing only a few items from them.

3.5.9.1 Creation of the draft CRO ontology. To create the initial version of CRO, the core BFO, IAO, and RO ontologies were downloaded and used to form an initial BFO-based foundation for CRO. Individual Excel spreadsheets were created to act as input templates to Ontorat for each of the 12 major subcategories in Concept Map Three. An additional, separate, small database of definitions for each subsection of the Common Rule was created. These definitions were then linked to each term as appropriate and added to the Excel spreadsheet, based on the legal section from which each term originated.

Each term was mapped to classes and annotation properties as shown in Table 13.

Table 13

Initial Mapping of Term-based Data to the Draft Common Rule Ontology

Field in excel template	Mapping to Ontology	Type of ontology element
Name of term	Class *label*	Annotation property
Assigned URI	Class	URI
Term category	Initial superclass	Class relationship
Superclass of term	*rdfs:subClassOf* the superclass	Class relationship
Major legal section of CR	n/a	n/a
Legal source of the definition (e.g. 45 CFR 46 § 116)	*definition source*	Annotation property
Legal definition of the term	*definition*	Annotation property
General notes	*editor note*	Annotation property
'FJM' (initials of this author)	*term editor, definition editor*	Annotation properties

Note: Words shown in *italics* are formal classes or properties in the CRO ontology.

Ontorat was used to generate the initial OWL-2 files from the 12 input templates described above. The Protégé editor was then used to import these OWL-2 files into the initial CRO foundation.

3.5.10 Ontology development/formal modeling: integration. At this point, it was necessary to fully align classes and properties with the BFO, IAO, RO, and OBI ontologies. This was done by utilizing the Protégé editor, with successive rounds of review involving three of us (Harris, Manion, Tao) and iterative refinement. The HermiT 1.3.8.413 reasoner (Glimm, Horrocks, Motik, Stoilos, & Wang, 2014) was used to verify that the resulting ontology was mathematically *consistent* and *satisfiable*.

3.5.10.1 Use of the BFO top-level ontology and other OBO Foundry ontologies. To align with the BFO and OBO Foundry, Ontobee (Xiang et al., 2011) was used to search for the prior definition of CRO terms in existing OBO Foundry ontologies. Ontobee is the official ontology search engine for the OBO Foundry repository. Existing terms found from the OBO Foundry were substituted into the mind maps and the ontology, assuming these terms were well-defined and appropriate to the context of the Common Rule. Additional terms representing concepts that were deemed important but not already present in the Common Rule Ontology were added as well.

To preserve the semantics of imported terms, all new terms added to CRO were imported using the MIREOT methodology (Courtot et al., 2009) as implemented by the Ontofox program of Xiang et. al. (Xiang et al., 2010) so that any appropriate axioms would be included as well.

3.5.10.2 Structural error checking. The ontology and corresponding excel files were reviewed to ensure that (a) all terms in the initial concept maps and excel files were represented in the ontology, (b) no inadvertent duplicate terms were introduced, and (c) that the concept maps were an accurate reflection of the actual ontology. The latter step was done as the concept maps, particularly Concept Map Three, remain a useful tool for working with domain experts, users of the ontology, and for documentation purposes.

3.5.11 Ontology development/formal modeling: Evaluation. At this point, the CRO was ready for summative evaluation. Five types of evaluation were conducted: competency question evaluation, FOCA analysis, Burton-Jones semiotic analysis, sentence analysis, and an assessment by domain experts. Methods for these evaluations are discussed in detail in later chapters.

Chapter 4. Results of Ontology Construction

This chapter describes the results of the CRO construction process, including the concept maps, the design patterns used, and the ultimate hierarchical class structure of the ontology. It also discusses the major organization of the ontology in terms of its alignment with the major BFO and OBI frameworks, its incorporation of external ontologies, and coverage of the Common Rule. The artifacts discussed are presented in the order in which they are generated as the result of the NIST lifecycle and METHONTOGY (MO) development phases. Where artifacts were iteratively refined and/or span multiple steps of the NIST/MO process they are presented only once, at the final point in the process.

4.1 NIST Requirement Analysis Phase

4.1.1 Planification stage. In this phase, Concept Map One was generated using informal interviews and review of regulatory materials. Because of its size, the map is presented in Appendix B. Concept Map One was shown to four domain experts. Responses from subject matter experts are shown in Table 14.

Table 14

Comments from Subject Matter Experts Validating the Preliminary Concept Maps

Role	Paraphrased comments
Director, University of Michigan Medical School Biorepository	'I never thought of the domain in this way, and this is very intriguing. We generally keep these models in our mind, and this really demonstrates the complexity of the domain. The model appears accurate to me.'
Director, CTSA DNA Biolibrary, clinical researcher	'The model appears accurate.'
Senior compliance officer	'I never thought of the domain in this way… The model appears accurate.'
Vice-Chairman, IRB for Health Sciences and Behavioral Sciences	'I never thought of the domain in this way… The model appears accurate.'

4.1.2 Specifications. Based on the specific aims of this dissertation, and on the requirements of NHGRI-U01-HG009454 for development of 'metadata applications on informed content to facilitate biorepository data regulation and sharing,' the following specifications were developed, as shown in Table 15. The table is adapted from the content and format suggested by Fernandez, et al., in his paper on METHONTOLOGY (Fernández et al., 1997).

Table 15

CRO Requirement Specification Document

Domain: Regulation of Human Subjects Research Date: October 7, 2017 Conceptualized-by: Frank J. Manion, M.S.; Marcelline Harris, Ph.D., RN; Cui Tao, Ph.D. Implemented-by: Frank J. Manion
Purpose: This ontology is about the content of the United States Common Rule, as published in the Code of Federal Regulations (C.F.R.), Title 45 – Public Welfare, Chapter A – Department of Health and Human Services, Subchapter A – General Administration, Part 46, Subparts A through D, version 1991, with extension by new material from the January 2017 revision. The ontology is intended as a knowledge model for a metadata model incorporating required elements of consent, and optional data elements about specimen and data use in subsequent research, sometimes termed secondary research. It is also intended to be contributed to the Open Biomedical Ontologies (OBO) Foundry (Smith et al., 2007) to supplement the Informed Consent Ontology (ICO) with additional information about IRB review and approval of required processes, and to contribute to the Ontology for Biobanking (OBIB).
Level of Formality: Formal — Aligned with the Basic Formal Ontology upper-level ontology.
Scope: List of major terms from determined through text analysis of 45 C.F.R. § 46, subparts A through D dealing with what is covered, general requirements for informed consent, IRB processes, and special protections for pregnant women, fetuses and neonates, prisoners, and wards of the state.
Sources of (Domain) Knowledge: 45 C.F.R. § 46, regulatory affairs experts, legal experts, IRB members, and biobanking professionals. Legal source material from various online legal dictionaries.

4.2 NIST Ontological Analysis Phase

4.2.1 Knowledge acquisition and conceptualization. The initial series of concepts and relationships were represented in Concept Map Two and organized by Common Rule section, as discussed in methods Section 3.5. Analysis of Concept Map Two resulted in the preliminary glossary of lexical and syntactic elements such as top-level classes, terms, relations, and other data as shown in Table 16. The metadata column in the table shows the initial classification of text from the Common Rule. Note that these tags are somewhat arbitrary and simply served as an initial organizing paradigm; for example, in a formal sense many of the so-called 'axioms' and 'restrictions' are 'properties'. The actual concept map is too large to be included in this document, however, a full listing of the results of the analysis, which follows the same structure as Concept Map Two, can be found in Appendix C.

Table 16

Summary of Data from Concept Map Two

Metadata Tag	Count
Authority records	4
Axioms	65
Classes	377
Equivalence classes	27
Properties	70
Restriction	5
Individuals	2
Total	550

Authority records were statements that are intermixed in the content of the Common Rule itself describing where the regulation draws its authority from. For example, the literal text:

> Authority: 5 U.S.C. 301; 42 U.S.C. 289(a); 42 U.S.C. 300v-1(b).

appears in the body of the text directly after the Subpart A heading. This was a convenient way of capturing material for possible use in a table of authorities. The tag 'axioms' was used to represent complex expressions for possible restriction classes in the CRO. The tags for 'classes', 'equivalence classes', 'properties', and 'individuals' all roughly adhere to the normal use of these terms in an ontology. The tag 'restriction' generally was used to represent disjoint classes or properties.

Once the glossary of terms and relations was complete, Ontobee was used to search for occurrences of them in existing OBO Foundry ontologies. These were noted and if relevant ultimately imported into the CRO by using the MIREOT method of Courtot as implemented by Xiang (Courtot et al., 2009; Xiang et al., 2010). Table 17 lists the other OBO Foundry ontologies that contributed material to the CRO.

Table 17

Count of Terms Included in the CRO by Source Ontology

Ontology	Ontology Name	Unique URIs
IAO	Information Artifact Ontology	129
OMRSE	Ontology of Medically Relevant Social Entities	47
BFO	Basic Formal Ontology	43
OBI	Ontology of Biomedical Investigations	36
PATO	Phenotypic Quality Ontology	35
RO	Relation Ontology	29
ICO	Informed Consent Ontology	16
NCBITaxon	NCBI organismal classification	13
geneontology	Gene Ontology	11
CARO	Common Anatomy Reference Ontology	9
dc	Dublin Core	8
OAE	Ontology of Adverse Events	4
UO	Units of measurement Ontology	3
OGMS	Ontology for General Medical Science	2
ONTONEO	Obstetric and Neonatal Ontology	1
GAZ	Gazetteer	1
CL	Cell Ontology	1
Grand Total		388

Note: Counts shown are those in the final CRO and represent the results of a transitive closure of the classes, relations, and axioms due to the application of the MIREOT methodology.

BFO, IAO, OBI, ICO, and RO were all expected to contribute substantially when this work began. Terms included from OMRSE were used to represent people and roles involved in research and health care, such as nurse, ‘party to a legal agreement’, and so

forth. Terms from PATO were used to represent qualities of a person such as 'alive', 'dead', 'viable', 'nonviable', etc. The other ontologies were mostly used for inclusion of a small number of very specific terms.

4.3 NIST Ontology Design Phase.

4.3.1 Knowledge acquisition and conceptualization. At this point in the design Concept Map Three was started. In this concept map terms from the glossary were classified into a taxonomic relationship through successive rounds of manual classification by two of us (Manion, Harris). This resulted in 12 high-level conceptual areas, as shown in

4.3.2 Table 18, which also reports the number of terms in each of these major areas.

4.3.3 Integration. Resulting top-level classes were assigned an initial classification at a high level in the BFO or OBI taxonomic hierarchies, e.g., as an independent continuant, a generically dependent continuant, etc. Initially all terms in the taxonomic hierarchy were made subclasses of the BFO or OBI class chosen, but eventually, this was greatly refined. The initial alignment is also shown in

4.3.4 Table 18.

Table 18

Initial Top-level Taxonomic Classification

Top-level classification	# of terms	Description	Initial BFO/OBI alignment
Study area	58	Major classification of research, such as public health surveillance, educational studies, studies of existing data, etc.	*BFO:'specifically dependent continuant'*
Study goals	7	One of six categories enumerated in the Common Rule. One of these is the important category 'yield generalizable knowledge.'	*BFO:'generically dependent continuant'*
Materials	123	Real-world objects, e.g., cells, tissues, documents, food, etc.	*BFO:'material entity'*

Top-level classification	# of terms	Description	Initial BFO/OBI alignment
Procedures	91	All processes and procedures, such as the 'process for ensuring risks to subjects are reasonable.'	*OBI: 'planned process'*
Governance	21	Organizations or people that have authority over decisions.	*BFO: 'independent continuant'*
People & organizations	65	People or groups involved in the research process. IRB's and the parent of a child being studied are examples.	*BFO: 'independent continuant'*
Roles	60	A role played by an individual or thing, e.g., the role of a neonate.	*BFO: 'specifically dependent continuant'*
Qualities	84	Generally an attribute of something. The legal age of a person, for instance.	*BFO: 'specifically dependent continuant'*
Protections	45	Represents protections given to a research subject. Protections can be part of a process, or stand alone. The informed consent process is an example of the former, whereas protections afforded a research subject due to federal law are an example of the latter.	Initially assigned as a *BFO: 'generically dependent continuant'*
Events	18	Events are defined as significant occurrences that change the state of a research study. An adverse event, or a withdrawal of consent, for example.	BFO:'process boundary'
Decisions	37	The output of a process, e.g., the decision of an IRB to approve a research study.	*BFO: 'realizable entity'*
Spatial location	12	Generally, a physical location which is part of a legal jurisdiction or a location at which research takes place, e.g., an academic medical center.	*BFO: 'independent continuant'*
Total:	621		

4.4 NIST Ontology Development Phase.

The ontology development phase of the construction process was highly iterative and took multiple rounds of iteration between the implementation, integration, and to some extent the evaluation phase. This involved much back and forth between me, Drs. Harris

and Tao, and a number of ad-hoc members of the OBO community (Drs. Stockert, Brochhausen, and Obeid).

4.4.1 Informal modeling — conceptualization.

4.4.1.1 Concept map validation with domain experts. As described in Section 3.5.7.2, five subject matter experts in regulatory affairs, IRB management, and compliance were recruited to do one-and-a-half-hour semi-structured qualitative reviews of Concept Map Three. The roles of the persons interviewed are listed in Table 12. The concept maps used during these interviews are shown in Appendix D. Changes to the concept maps suggested by each interviewee were incorporated into the next round of interviews until a saturation of term placement was achieved. Nineteen total comments were made by the five interviewees; of these five were already represented in the concept map. One comment specifically verified the structure of the governance sub-hierarchy. Thirteen comments were ultimately addressed through these interviews. All of the 13 comments dealt with concepts that the domain experts felt were either missing or unclear. There was little if any disagreement with the typographical structure of Concept Map Three, including the structure of the top-level classes.

4.4.2 Formal modeling — initial implementation. At this point in the process the initial version of the CRO was constructed using methods discussed in Section 3.5.9.1, 'Creation of the draft CRO ontology.'

4.4.3 Formal modeling — integration. Following a final review of Concept Map Three, terms were aligned with the BFO and OBI hierarchy. This work was completed

utilizing the Protégé editor with iterative review by Drs. Harris and Tao. Additional or missing terms suggested by the domain experts were added to the ontology using the Ontorat program as described earlier.

4.4.3.1 BFO and OBI design patterns. In order to integrate the terms from the final iteration of Concept Map Three into the CRO, a number of BFO and OBI design patterns were used. One of the most important of these was the OBI-based design pattern that extends the notion of *BFO:process* as shown in Figure 10.

Figures in this section will show in a general fashion how all the top level concepts and associated taxonomies were integrated into the BFO. The reader should be mindful, however, that details of semantic relations for each individual terms vary significantly. The details of the semantic dimensions are difficult to show in these figures.

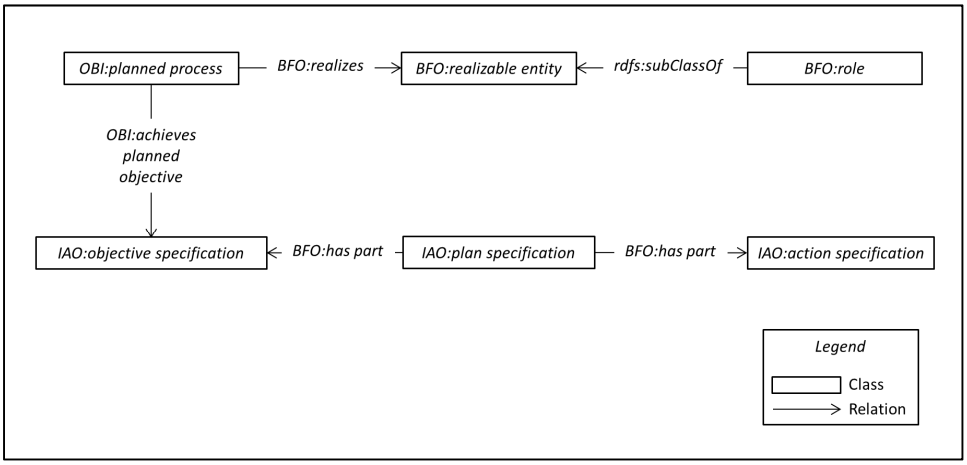


Figure 10. OBI 'planned process' design pattern.

4.4.3.2 Mapping processes and study goals into the CRO. Terms derived from Concept Map Three were aligned with the design pattern from Figure 10 as shown in Figure 11. Processes were represented as an *OBI:planned process*. This class contains semantic relationships to classes in the Information Artifact Ontology (IAO) that represent objective specifications (*IAO:objective specification*), plan specifications (*IAO:plan specifications*), and the specifications for actions to be taken (*IAO:action specification*). Study goals from Concept Map Three were mapped to *IAO:objective specification.*

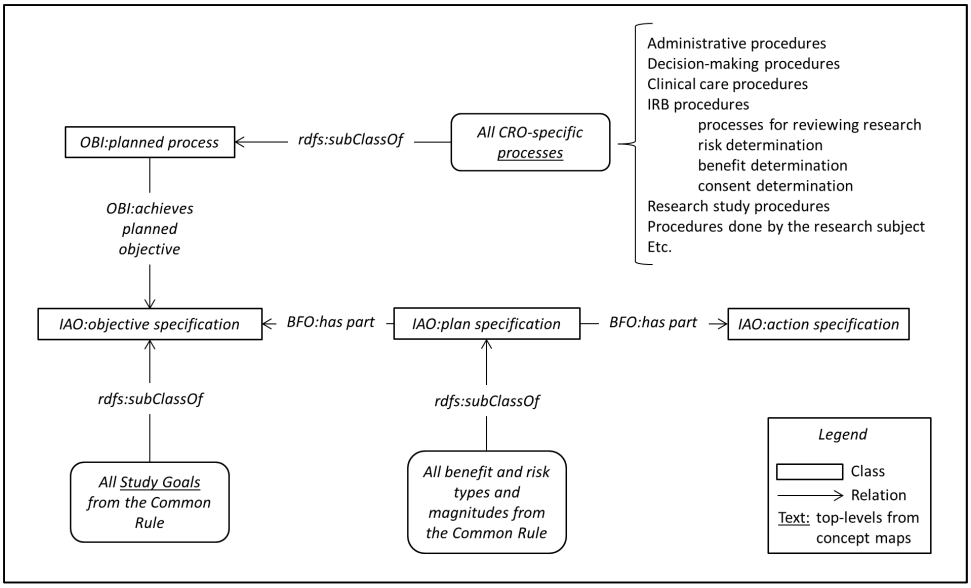


Figure 11. Result of alignment of Common Rule terms involved in processes from Concept Map Three into CRO using OBO Foundry design patterns.

4.4.3.3 Mapping 'materials' terms into the CRO. Figure 12 shows how the top-level concept 'materials' and the subparts of that section were aligned with the BFO hierarchy. Much of the content of §46.116 dealing with required and optional elements of informed consent was mapped onto an *IAO:document* part, and semantically connected with other properties, not shown in the figure, such as short, long, and verbally administered informed consent 'forms.' Cells and tissues specified in the Common Rule were aligned with classes and relationships from CARO, the Common Anatomy Reference Ontology.

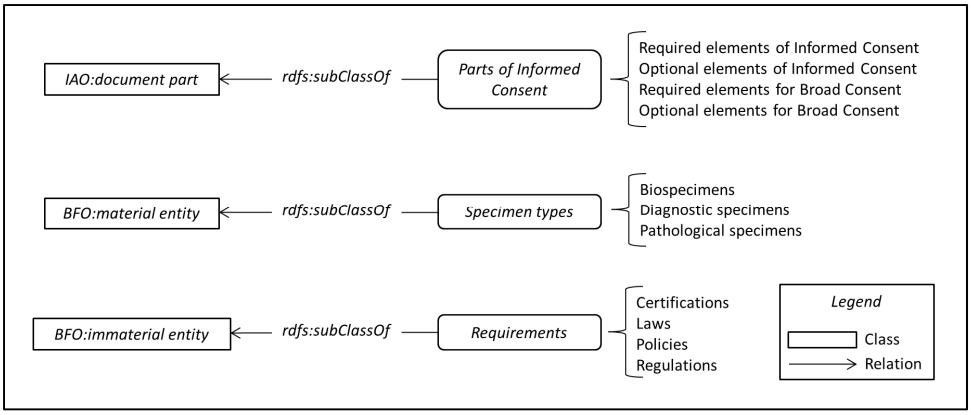


Figure 12. Integration of components of the top-level 'material' section of Concept Map Three and their mapping onto the BFO.

4.4.3.4 Mapping governance terms into the CRO. Governance was modeled as a series of roles, which were implemented as subclasses of *BFO:role*, which is a specifically dependent continuant. Roles 'inhere in' independent continuants. Mapping of the governance top-level concepts to the BFO is shown in Figure 13.

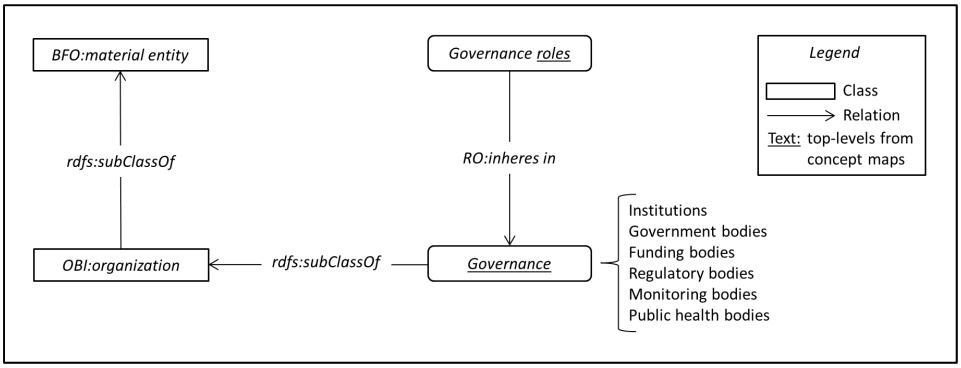


Figure 13. Mapping of concept map 'governance' to the BFO.

4.4.3.5 Mapping people and organizations into the CRO. Mapping of the concept 'people and organizations' was straightforward. People were mapped to the class *NCBITaxon:Homo sapiens*, which is typically used throughout the OBO Foundry for representing people. This class is a *BFO:material entity*, which is an independent continuant in the BFO. Organizations were mapped to the OBI class *OBI:organization*, which again is a *BFO:material entity*. These mappings are shown in Figure 14.

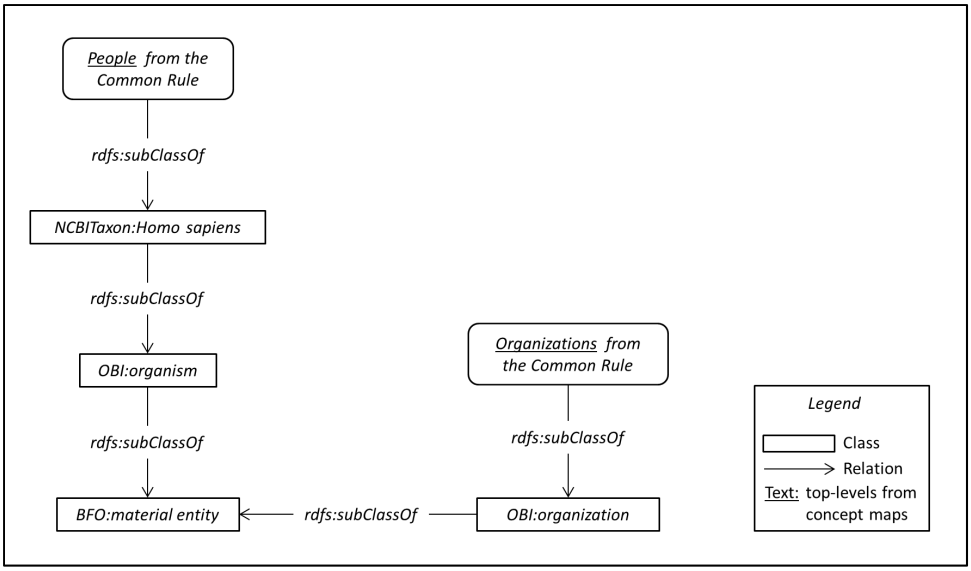


Figure 14. Mapping of concept map 'people and organizations' category to the BFO.

4.4.3.6 Mapping roles and qualities into the CRO. Roles were mapped onto the BFO class *BFO:role*, while qualities were mapped onto *BFO:quality*. Both of these BFO classes are specifically dependent continuants. *BFO:role*, however, is a *BFO:realizable entity*, which means classes and instances in this taxon are 'realized', or spring into being, through an associated *BFO:process*. The 'study goals' category was implemented as an *IAO:objective specification,* which participates in processes through the relation *OBI:achieves planned objective*. This is shown pictorially in Figure 15, along with the manner in which people and organizations interact with this structure. People and organizations that take some part in the research process were assigned specific roles or qualities pertaining to the research process.

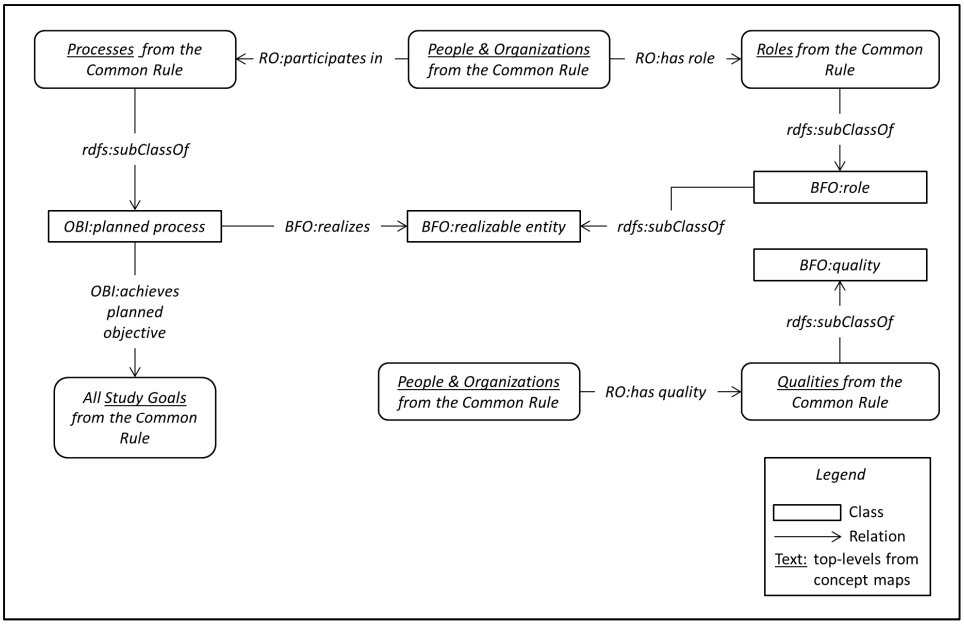


Figure 15. Mapping of top-level concepts 'roles', 'study goals', and their interaction with 'processes', 'people', and 'organizations.'

4.4.3.7 Mapping protections into the CRO.

Protections from Concept Map Three were mapped to the IAO class *IAO:rule*. This class is a subclass of *IAO:directive information entity* and is linked semantically to a realizable entity through an *IAO:is about* relation. Figure 16 extends the previous figure to illustrate this concept.

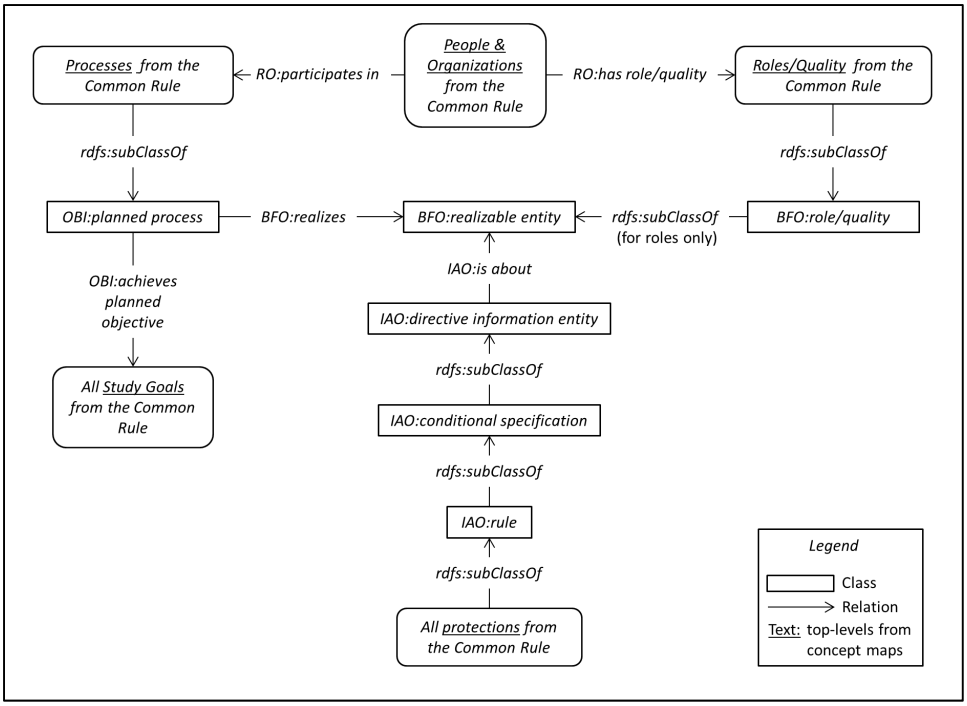


Figure 16. Mapping of protections to the BFO framework of the CRO.

4.4.3.8 Mapping decisions into the CRO. Decisions from Concept Map Three were mapped as subclasses of a new class, *CRO:decision*, which was implemented as a direct descendant of *BFO:realizable entity*. Decisions are the result of some decision-making process, consequently, they are modeled with the semantic relationship *OBI:is_specified_output_of* relating them to the appropriate decision process. Once again, the previous diagram is extended to Figure 17 to demonstrate how decisions were modeled in the CRO and how they fit into the overall design.

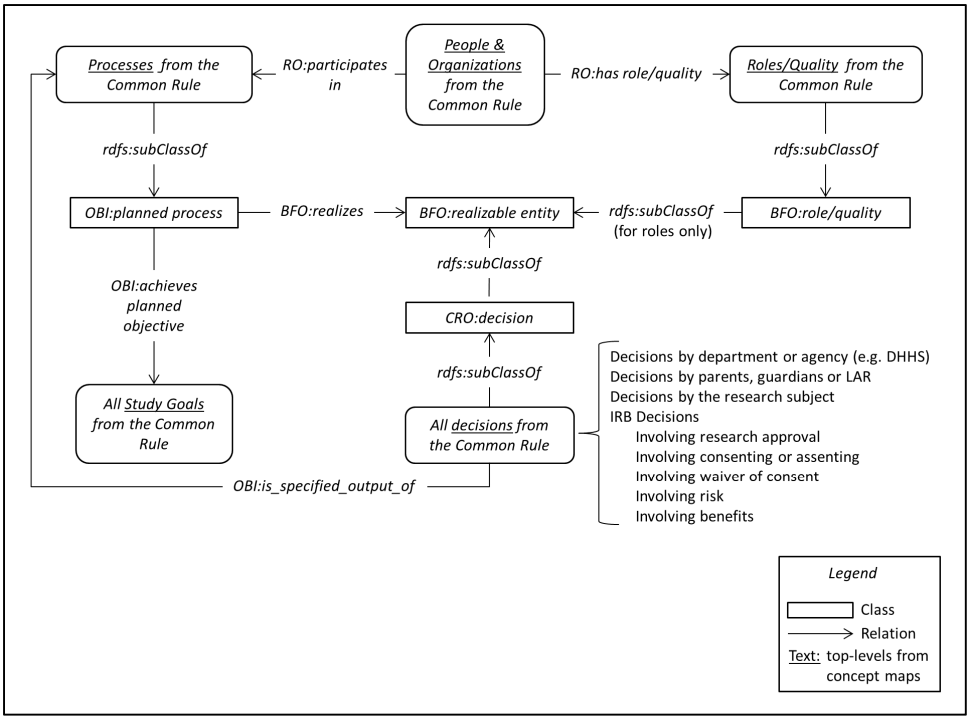


Figure 17. Representation of decisions in the CRO.

4.4.3.9 Mapping study areas into the CRO. Study areas identified from Concept Map Three were mapped as direct descendants of *BFO:realizable entity*. Figure 18 demonstrates how study areas were modeled in the CRO.

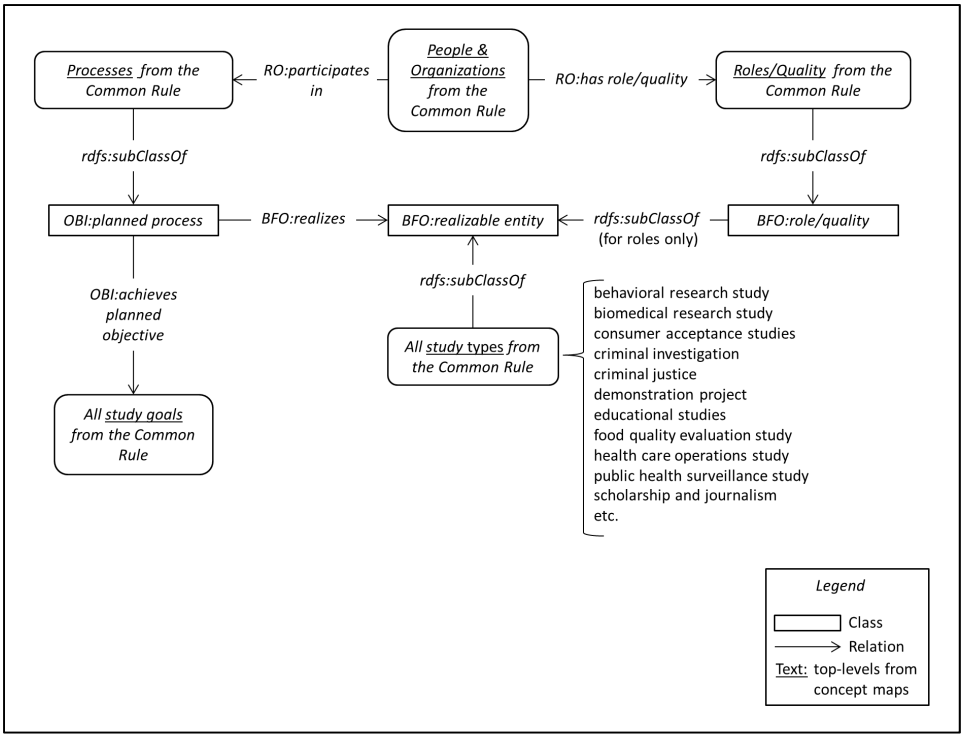


Figure 18. Representation of study areas in the CRO.

4.4.3.10 Mapping events into the CRO. Events were mapped as process boundaries, an occurrent within the BFO hierarchy. A new, CRO-specific semantic relation '*CRO:results in event*' with a domain of *BFO:process* and a range of '*BFO:process boundary*,' and its inverse relationship '*CRO:event results from'* with a domain of '*BFO:process boundary'* and a range of *BFO:process,* were introduced in mapping events. This is shown in Figure 19.

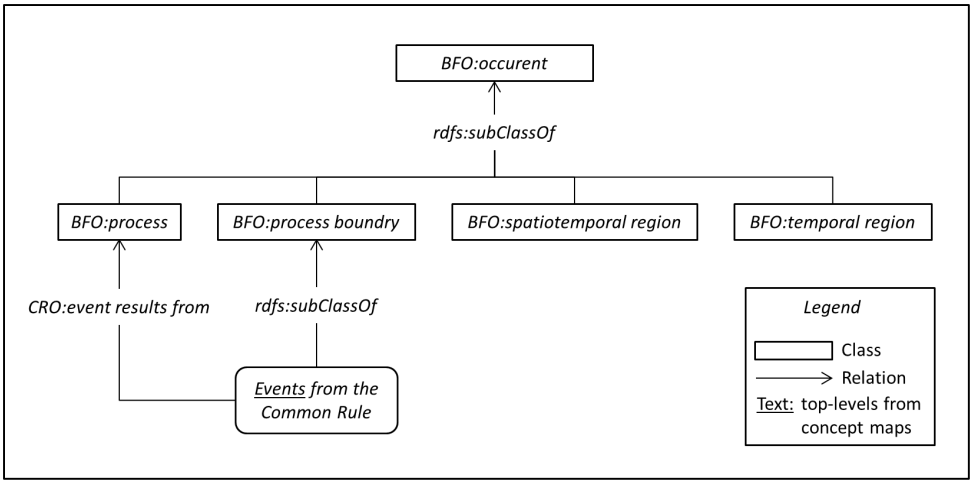


Figure 19. Representation of events within the CRO.

4.4.3.11 Mapping sites into the CRO. Sites from Concept Map Three were mapped to *BFO:site*, a *BFO:immaterial entity*, shown in Figure 20. This is an independent continuant within the BFO hierarchy.

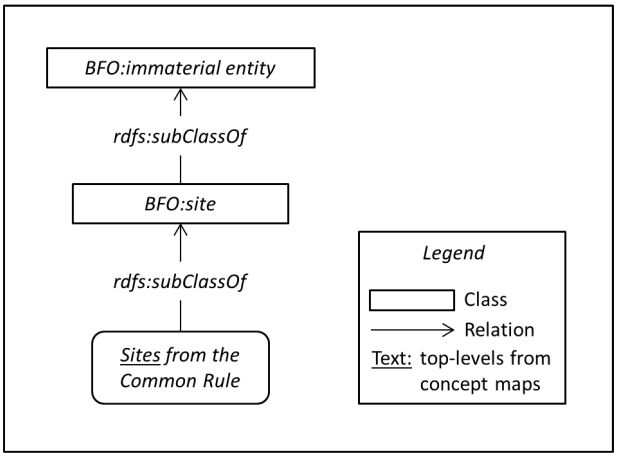


Figure 20. Representation of sites named in the Common Rule to the CRO.

4.4.3.12 Evaluation of the final ontology with the HermiT reasoner.

Once the ontology was constructed, a final check was made with the ontology reasoner for coherence, satisfiability, and consistency. The output of the Protégé ontology debugger plugin is shown below in. In reality, the reasoner is essential to the overall ontology development process and was constantly used to assure coherence and consistency while the ontology was being constructed.

Figure 21. Output from HermiT reasoner and Protégé debugger plug-in module.

4.5 Descriptive Metrics of the CRO.

In this section, various descriptive characteristics of the CRO ontology are presented. Basic metrics include the number of classes, relations, axioms, and other characteristics of the CRO. These values are reported directly from the Protégé editor. Protégé also calculates the description logic (DL) expressivity of the ontology. Understanding the DL expressivity is important for, among other reasons, choosing the correct inferential reasoner.

4.5.1 Basic counts of CRO components. Shown in Table 19 are the component counts of the CRO ontology, as reported by Protégé.

Table 19

CRO Metrics from the Protégé Ontology Editor

Metrics		Class axioms	
Axioms:	8571	SubClassOf:	1040
Logical axiom count:	1536	EquivalentClasses:	56
Declaration axioms count:	1142	DisjointClasses:	35
Class count:	864	GCI count:	0
Object property count:	70	Hidden GCI Count:	50
Data property count:	6		
Individual count:	95		
Annotation Property count:	102		
DL expressivity:	SROIQ(D)		
Object property axioms		**Data property axioms**	
SubObjectPropertyOf:	43	SubDataPropertyOf:	4
EquivalentObjectProperties:	36	EquivalentDataProperties:	0
InverseObjectProperties:	21	DisjointDataProperties:	0
DisjointObjectProperties:	0	FunctionalDataProperty:	4
FunctionalObjectProperty:	6	DataPropertyDomain:	1
InverseFunctionalObjectProperty:	0	DataPropertyRange:	4
TransitiveObjectProperty:	6		
SymmetricObjectProperty:	0		
AsymmetricObjectProperty:	2		
ReflexiveObjectProperty:	0		
IrrefexiveObjectProperty:	4		
ObjectPropertyDomain:	34		
ObjectPropertyRange:	36		
SubPropertyChainOf:	5		
Individual axioms		**Annotation axioms**	
ClassAssertion:	122	AnnotationAssertion:	5887
ObjectPropertyAssertion:	73	AnnotationPropertyDomain:	0
DataPropertyAssertion:	1	AnnotationPropertyRangeOf:	0
NegativeObjectPropertyAssertion:	0		
NegativeDataPropertyAssertion:	0		
SameIndividual:	0		
DifferentIndividuals:	2		

4.5.2 DL expressivity of the CRO. The CRO has a DL expressivity designation of $\mathcal{SROIQ}(\mathcal{D})$. The key to interpreting this designation is as follows.

$\mathcal{S}$: A shorthand for $\mathcal{ALC}$ with transitive properties.

$\mathcal{AL}$: Attributive language. This is the base language which allows (a) atomic negation, (b) concept intersection, (c) universal restrictions, and (d) limited existential quantification.

$\mathcal{C}$: Complex concept negation.

$\mathcal{R}$: Role hierarchy, i.e., sub-properties such as *rdfs:subPropertyOf.*

$\mathcal{O}$: Nominals (enumerated classes or object value restrictions such as *owl:oneOf* or *owl:hasValue*).

$\mathcal{I}$: Inverse properties.

$\mathcal{Q}$: Qualified cardinality restrictions.

$(\mathcal{D})$: Use of datatype properties, data values, or data types.

Chapter 5. Ontology Evaluation Methods and Evaluation of the CRO

This chapter describes the ontology evaluation methods used for both formative and summative evaluation of the CRO. The first section presents a short background on the theoretical basis of ontology evaluation, followed by a description of models proposed by several different authors. Criteria from the NIST Ontology Lifecycle Model as introduced in Chapter 3 is then described. The second section of the chapter details the methods actually used in evaluating the CRO, and closes by relating them to the NIST evaluation model.

5.1 A Short Review of Ontology Evaluation Frameworks

Even after over two decades of work on the subject, the evaluation of ontologies is still considered an emerging field (Gómez-Pérez, 2004; Vrandečić, 2009). Ontologies can be evaluated along a number of axes, and this section attempts to give the reader a sense of the different approaches proposed since the mid-1990's. The often cited ontology spectrum described by Obrst (Obrst, 2010) suggests that ontology evaluation methods should at a minimum be grounded in their ability to make valid and reliable measurements along that spectrum. Evaluations for assessing an ontology should, therefore, be able to make assertions about the organization of terms or concepts, the conceptual models used, and the logical theories and expressivity of the ontology. In an important paper, Obrst himself proposed evaluation criteria in the three general

categories: *quality criteria*, *philosophical foundations*, and *verification and validation* (Obrst, Ceusters, Mani, Ray, & Smith, 2007).

5.1.1 Ontology verification and validation. The terms *ontology verification* and *ontology validation* in the field were first proposed by Gómez-Pérez, an important researcher in the field, in 2004 (Gómez-Pérez, 2004). Ontology verification was defined to mean the assessment of whether or not the ontology correctly meets the stated requirements, whereas ontology validation means that the ontology is a faithful representation of the domain.

5.1.2 Brank's evaluation criteria. Brank, in a paper in 2005 surveyed and summarized a variety of approaches to evaluation (Brank, Grobelnik, & Mladenić, 2005). He determined at that time that most teams evaluated ontologies by (a) comparing them to a 'gold standard,' (b) assessing the results of the use of the ontology, or (c) by human evaluation against a set of requirements. He also proposed a classification system based on the 'levels' or structural and functional sub-parts of the ontology, noting that it is difficult to attempt the evaluation of an ontology as a whole. The categorization model he proposed consists of the following six levels:

1. The *lexical, vocabulary, and data layer* level attempts to measure what has been included in the ontology.
2. The *hierarchy* level attempts to assess the taxonomic structure of the ontology by examining the *is-a* structure of the terms.
3. The *other semantic relations* level examines the relations in the ontology.

4. The *context or application* level attempts to assess the use of the ontology within a broader context, such as membership in a larger collection of ontologies like the OBO Foundry.
5. The *syntactic* level evaluation examines the use constructs in the formal language in which it is represented.
6. The *structure, architecture, and design* level examines the organization of the ontology, and assess factors such as the potential for reuse, or alignment with upper- and mid-level ontologies.

5.1.3 Extrinsic and intrinsic evaluation criteria. Zhu and colleagues reviewed ontology auditing practices and characterized aspects of various quality factors (Zhu, Fan, Baorto, Weng, & Cimino, 2009). They present five quality criteria, shown in Table 20. Zhu also introduces the notion of *intrinsic* and *extrinsic* quality factors. Intrinsic factors can be thought of as factors that inherently derive from the domain and the ontology or terminology model. Examples of intrinsic factors could be questions such as 'are the hierarchical relationships correct', 'is a concept linked to more than one parent where the parents are hierarchically related', 'does a BFO-aligned ontology properly represent material objects in the real world as independent continuants?', etc. Extrinsic factors are those that depend on sources of knowledge that are not part of the ontology. Examples of extrinsic factors might include the Unified Medical Language System (UMLS) (Bodenreider, 2004) or WordNet (Fellbaum, 2010), for example, or to an external ontology such as the Gene Ontology (Ashburner et al., 2000). In a sense (although with some overlap), intrinsic factors are meant to answer the question 'is the

ontology a quality ontology,' whereas extrinsic factors are meant to answer the question 'was the right ontology built'?

Table 20

Ontology Evaluation Criteria Suggested by Zhu et al.

Criteria	Definition
Intrinsic criteria	
Concept orientation	Terms in the ontology should not be vague, have no more than one meaning, and should be represented by unique identifiers aligned with concepts rather than human-readable labels.
Consistency	Classification of concepts is done in a consistent fashion across the ontology.
Non-redundancy	Information should not be repeated in the ontology as it can introduce ambiguity and taxonomic problems.
Soundness	Is the knowledge represented in the ontology accurate?
Extrinsic criteria	
Comprehensive coverage	Does the ontology contain the necessary and sufficient information to make it fit for a particular purpose?

5.1.4 Vrandečić's criteria and aspects of evaluation. In 2009, Vrandečić presented a framework as part of a 'Handbook on Ontologies' (Vrandečić, 2009). His framework describes eight evaluation criteria, and six 'aspects' of evaluation. The criteria are presented in Table 21, and the evaluation aspects are described in Table 22.

Table 21

Eight Criteria for Evaluation in Vrandečić Framework for Ontology Evaluation

Criteria	Definition
Accuracy	Refers to if the ontology correctly represents the domain of interest in the real world.
Adaptability	Refers to the ease of adapting and extending the ontology.
Clarity	Are definitions and other parts of the ontology readily usable by the intended community?
Completeness	Does the ontology have sufficient coverage of the domain of interest?
Computational efficiency	Can computational reasoners work with the ontology to classify instances, check satisfiability, and process queries in a suitable time period?
Conciseness	Does the ontology include relations or classes that are irrelevant to the domain?
Consistency	Refers to items such as 'do axioms or relations lead to logical contradictions?', and 'does the documentation match the actual implementation'?
Organizational fitness	Refers to the ability to actually deploy the ontology within the intended socio-technical setting.

Table 22

Six Aspects of Evaluation in Vrandečić Framework for Ontology Evaluation

Criteria	Definition
Vocabulary	Deals with the labels and their bindings to URIs.
Syntax	In Vrandečić's framework, this refers only to the different technical serialization aspects, such as choice of RDF/XML (Fabien Gandon, Guus Schreiber, & Dave Beckett, 2014), or OWL Abstract Syntax (Boris Motik et al., 2012), etc. Note some authors interpret syntax in a more traditional fashion.
Structure	Refers to the evaluation of aspects of the RDF graph that underlies the ontology. An example of such checks is that the class hierarchy is non-circular.
Semantics	Refers to the semantic models represented by the ontology, and their degree of expressivity. It also refers to the completeness of the ontology with respect to what can be expressed to what is actually present in the ontology.
Representation	Tries to measure the relationships between the structural and semantic aspects.
Context	Refers to how well the ontology works with other aspects of the environment in which it is used. An example of context assessment is competency question-based evaluation.

5.1.5 Evaluation by semiotic approaches. Burton-Jones has developed a suite of metrics for ontology evaluation that rely on the underlying semiotics, based on a framework developed by Stamper (Burton-Jones, Storey, Sugumaran, & Ahluwalia, 2005; Stamper, Liu, Hafkamp, & Ades, 2000). Stamper's framework consists of six layers, as follows:

1) *Physical* — does it have a physical form?

2) *Empiric* — can it be seen?

3) *Syntactic* — can it be read?

4) *Semantic* — can it be understood?

5) *Pragmatic* — is it useful?

6) *Social* — can it be trusted?

Evaluation by Stamper's framework proceeds through each layer, in order. The Burton-Jones suite implements metrics for the syntactic, semantic, pragmatic, and social components. It should be noted that this method attempts to measure both the intrinsic and extrinsic aspects of the ontology, as suggested in Zhu, above (Zhu et al., 2009). The method, as implemented by Amith (M. Amith & Tao, 2015), is used for evaluating CRO and is described in more detail later in this chapter.

5.1.6 Evaluation criteria suggested by the NIST lifecycle model. The NIST lifecycle model that was introduced earlier (Neuhaus et al., 2013) notes that there are three types of evaluations needed for an ontology, namely how well they can be used by people, machines, and as part of an integrated system. The NIST workshop participants suggested five criteria as shown in Table 23.

Table 23

Evaluation Criteria Suggested by NIST Lifecycle Model

Criteria	Description
Intelligibility	Describes how well humans can understand and work correctly with the ontology. Important for end users and maintainers of both the ontology and underlying information systems.
Fidelity	Refers to how well the ontology accurately represents the domain that is modeled.
Craftsmanship	Does the ontology use good design decisions and are these decisions used in a consistent fashion through the ontology?
Fitness	Does the ontology adhere to the requirements needed to make it fit for its intended use?
Deployability	Can the ontology be deployed within the information system context of its intended use and does it fulfill all the requirements imposed on it by that context?

5.1.7 Summary of evaluation approaches. As can be seen from the preceding sections, there are many different approaches to developing an evaluation strategy for an ontology. Zhu's classification of quality evaluation methods along intrinsic and extrinsic factors is attractive as it neatly partitions those evaluations that are possible based on descriptive measures and graph-theoretic components of the ontology from those that rely on adherence to fitness for purpose, domain completeness, and the like. The four Burton-Jones semiotic criteria are attractive as they are often those discussed by practitioners of ontology development. However, for the purposes of this work, I will relate the results of ontology quality evaluation to the NIST lifecycle model, as it was the overall framework used for the development of the CRO.

Readers interested in a more thorough review of ontology quality evaluation methods are referred to a forthcoming paper by Amith and colleagues that relates various methods to each other (M. F. Amith, He, Bian, Lossio-Ventura, & Tao, 2017). Bandeira also provides a nice, brief introduction to the topic in his paper on the FOCA methodology covering some additional framework proposed by others (Bandeira, Bittencourt, Espinheira, & Isotani, 2016).

5.2 Evaluation in the Context of the Ontology Lifecycle

The following sections discuss the evaluation methods actually used for evaluation at various parts of the lifecycle. As mentioned the Zhu intrinsic and extrinsic classification of evaluation methods forms a nice organizing principle, therefore this section is organized according to that framework. Because overall the NIST lifecycle was used, the section also highlights where the evaluation method occurred in the development lifecycle.

5.3 Intrinsic Evaluation Methods

As described earlier, intrinsic evaluation methods attempt to characterize components of the ontology based on its asserted classes, relations, instances, and axioms; and provide a basis for contrasting against other ontologies, or known standards derived from graph theory and best practices. They consist of measures such as classification, and attempt to demonstrate how complex in terms of description logics such ontologies are.

5.3.1 Descriptive Characterization of the Common Rule Ontology (CRO).

Metrics describing the number of axioms, classes, objects, data properties, and individuals were obtained from the Protégé ontology editor using the 'Ontology metrics' tab.

The description logics expressivity of the ontology was also computed by and obtained from the Protégé editor using the 'DL metrics' tab in the editor.

Since ontologies can be directly or indirectly imported into other ontology, simply tracking the number of ontologies imported manually from the OBO foundry is not sufficient to adequately understand where underlying classes and properties come from. Consequently, it was necessary to compute the transitive closure of the ontologies used. The number of ontologies imported into the CRO was determined by parsing the OWL file of the CRO ontology and developing a break down based on the Universal Resource Identifier (URI) of all the terms in the ontology. This approach allowed the unique source ontology of any duplicate terms imported to be definitively associated with the defining ontology within the OBO library.

5.3.2 Does the ontology demonstrate logical errors? The ontology was assessed for logical errors using the HermiT 1.3.8.413 reasoner. HermiT is a fully compliant OWL-2 reasoner based on hypertableau calculus (Glimm et al., 2014). It has been shown to outperform other commonly used reasoners and, relevant to this work, in particular on OBO Foundry-based ontologies. In general, description logic reasoners such as HermiT examine an OWL-2 DL knowledge base (i.e. the ontology) and check if it is

mathematically satisfiable. This implies that the model is both mathematically sound and complete.

5.3.3 Evaluation by ontology experts of ontology structure and integration with BFO and OBO-foundry ontologies. Formative feedback was obtained during the ontology construction from OBO community members and other domain area experts.

5.3.4 Details of the FOCA methodology. The CRO was assessed using the FOCA method (Bandeira et al., 2016). FOCA is based on a Goal, Question, Metric approach to ontology evaluation and is unique in that it tries to address (a) type of ontology (top level, domain, task, or application ontologies), and (b) variance between the level of experience of evaluators. The method is derived from ontology criteria proposed in evaluation models by a number of authors (Gangemi, Catenacci, Ciaramita, & Lehmann, 2006; Gómez-Pérez, 2001; Gruber, 1995; Hlomani & Stacey, 2014; Obrst et al., 2007; Vrandečić, 2009). Some of these characteristics were discussed in Section 5.1. Bandeira proposes a set of criteria, as described in Table 24, that he terms 'roles of knowledge representation'. These roles are considered the goals used to define a set of evaluation questions and criteria that are finally mapped to evaluation metrics in the FOCA methodology.

Table 24

FOCA Roles of Knowledge Representation

Role	Definition
Substitute	Is the knowledge representation an accurate representation of the real world?
Ontological Commitments	How close is the knowledge representation to the aspect of the world being modeled?
Intelligent Reasoning	Can the knowledge representation correctly infer components of the real world?
Efficient Computation	Can the knowledge representation be used by a computer in a reasonable time period?
Human Expression	How easy is it for a human to understand the knowledge representation?

To use this methodology, evaluators are asked to examine the ontology being evaluated and score the ontology based on 13 questions. As shown in Table 25, questions one through three measure aspects of the Substitute goal.

Table 25

Questions for FOCA Goal One: 'Substitute'

Question	Instructions
Q1	Did the ontology developer defined 'competencies' in the form of 'competency questions' or through other means? If not, the grade is 0. Otherwise, answer the following sub-questions Q1a-Q1c. For each sub-question, give one of these grades: 25,50,75, or 100.
Q1a	Is the ontology objective defined? (e.g. 'This ontology models the domain of...');
Q1b	Are the ontology stakeholders defined? (e.g. 'This ontology should be used by...');
Q1c	Are scenarios of use defined? (i.e., the situations in which the ontology must be used).
Q2	If competencies are not defined, the grade is 0. If competencies exist, see if the ontology tests for them. Use a grading scale of 25,50,75,or 100.
Q3	Does the ontology reuse other ontologies? If it does not, the grade is 0. If it does, the grade is 100.

Note: These questions are adapted from the tables in Bandeira (Bandeira et al., 2016)

The score for question Q1 is computed as either 0 or the mean of questions Q1a–Q1c. The overall score for the *substitute* goal is then computed as the mean of Q1 through Q3. This score is shown as parameter Cov_S in Equation (1) below.

Questions four through six, as shown in Table 26, relate to the *ontological commitments* demonstrated by the ontology. That is to say, how relevant are the terms in the ontology to the part of the real world being represented. For the Common Rule Ontology, this means that one expects to see aspects of the law and legal statements being modeled

rather than aspects of cells and organ systems that make up the humans who are research subjects. Note that since the Common Rule Ontology is a domain ontology, only question five is used.

Table 26

Questions for FOCA Goal Two: 'Ontological Commitments'

Question	Instructions
Q4	This question is only used if the ontology is an 'application' ontology. Does the ontology use too much abstraction to define the concepts? If the ontology is full of abstraction the grade is 0. If there are only some abstractions, give a grade between: 25 (very specific), 50 (moderate abstraction), 75 (many abstractions), 100 (full of abstractions).
Q5	This question is only used if the ontology is a 'domain' or 'task' ontology. Does the ontology use primitive concepts to define the evaluated domain (for example, an ontology which models a person, uses the concepts *thing* → *living being* → *human being* → *person* to define the concept of a person)? If the ontology does not use abstractions, the grade is 0. If there are only some abstractions, give a grade between these: 25 (very specific), 50 (moderate abstraction), 75 (much abstractions), 100 (full of abstractions).
Q6	Are the classes and properties coherent with the modeled domain? If the ontology is full of incoherences (for example, an ontology which models the concept car has a class lion and the property quantityOfPaws, that do not exist in the domain), the grade is 0. If there are some incoherences, give a grade between these: 25,50,75. If there is no incoherence, the grade is 100.

Note: These questions are adapted from the tables in Bandeira (Bandeira et al., 2016)

The overall score for the *ontological commitments* goal is computed as the mean of Q4 and Q5 or as Q4 and Q6, as appropriate depending on the type of the ontology. This score is the parameter Cov_{OC} in Equation (1) below.

As shown in Table 27, Questions seven and eight relate to how well the ontology represents *intelligent reasoning*. Essentially these questions attempt to measure how accurately statements derived from the ontology, and any inferences they produce, represent the domain being modeled.

Table 27

Questions for FOCA Goal Three: 'Intelligent Reasoning'

Question	Instructions
Q7	Check if the classes and properties (functional, transitive, reflexive and others) characteristics contradict the domain (for example *LivingBeing* is a subclass of *Person* in an ontology which models the person concept or *socialSecurityNumber* is not a functional property because a person cannot have more than one Social Security Number). If the ontology is full of contradictions, the grade is 0. If there are some contradictions, give a grade between these: 25,50,75. If there are no contradictions, the grade is 100.
Q8	Check if there are classes or properties which model the same thing with the same meaning (for example, using 'mouse' for both hardware and animals). If the ontology is full of redundancies, the grade is 0. If there are some redundancies, give a grade between these: 25,50,75. If there are no contradictions, the grade is 100.

Note: These questions are adapted from the tables in Bandeira (Bandeira et al., 2016)

The score for the *intelligent reasoning* goal is simply the mean of Q7 and Q8 and is shown as parameter Cov_{IR} in Equation (1).

The questions shown in Table 28 attempt to measure whether the ontology can be practically be used by seeing if the machine reasoner used operates quickly enough.

Table 28

Questions for FOCA Goal Four: 'Efficient Computation'

Question	Instructions
Q9	Check if the ontology reasoner returns some kind of error. If the ontology is full of errors (or the software stops responding), the grade is 0. If there are some errors, give a grade between these: 25,50,75. If there are no errors, the grade is 100.
Q10	Check if the reasoner is running quickly. If the reasoner stops, the grade is 0. If there is any delay, give a grade of 25,50, or 75. If it runs quickly, the grade is 100.

Note: These questions are adapted from the tables in Bandeira (Bandeira et al., 2016)

The score for the goal of *efficient computation* is the mean of Q9 and Q10 and is shown as parameter Cov_{EC} in Equation (1).

Finally, the questions for goal 5, *human expression*, as shown in Table 29, attempt to measure how easily human beings can use the ontology. These questions measure characteristics of the annotations used in the ontology, and the definitions in any accompanying documentation.

Table 29

Questions for FOCA Goal Five: 'Human Expression'

Question	Instructions
Q11	Check if the documentation of ontology exists. If it does not exist, the grade is 0. If documentation exists, answer two sub-questions Q11a and Q11b.
Q11a	Are the written terms in the documentation the same as the modeling?; Give a grade of: 25,50,75, or 100.
Q11b	Does the documentation explain what each term is and does it justify each detail of modeling? Give a grade of: 25,50,75, or 100.
Q12	Check if the classes or properties of ontology are written in an understandable and correct form (according to English or another language). If the ontology is difficult to understand or full of poorly written terms, the grade is 0. If there are some errors or a mix of languages, give a grade of 25,50, or 75. If the ontology is well written and one language was used, 100.
Q13	Check if the existing annotations represent definitions of the modeled concepts. If there are no annotations, the grade is 0. If there are some annotations, give a grade of 25,50, or 75. If all the concepts have annotations, the grade is 100.

Note: These questions are adapted from the tables in Bandeira (Bandeira et al., 2016)

The score for question Q11 is either 0 or the mean of questions Q11a and Q11b. The overall score for *human expression* goal is computed as the mean of Q11 through Q13. The overall FOCA quality score is calculated for each reviewer by a beta regression model (Ferrari & Cribari-Neto, 2004) as shown in Equation (1). The human expression

score *HE* is not currently used as a covariate in FOCA and simply reported separately as a score focused on human factors.

$$\widehat{\mu_i} = \frac{\exp\{-0.44+(0.33\times Cov_S)_i+(0.02\times Cov_{OC})_i+(0.01\times Cov_{IR})_i+(0.02\times Cov_{EC})_i-0.66\times LExp_i-25(0.1\times Nl)_i\}}{1+\exp\{-0.44+(0.33\times Cov_S)_i+(0.02\times Cov_{OC})_i+(0.01\times Cov_{IR})_i+(0.02\times Cov_{EC})_i-0.66\times LExp_i-25(0.1\times Nl)_i\}} \quad (1)$$

Where for each reviewer *i*:

Cov_{Si} is the calculated grade from the substitute goal;

Cov_{OCi} is the calculated grade from the ontological commitments goal;

Cov_{IRi} is the grade from the intelligent reasoning goal;

Cov_{ECi} is the grade from the efficient computation goal;

$LExp_i$ is the level of experience of reviewer *i*; a reviewer who considers themselves experienced gives themselves a 1, otherwise, they give themselves a 0;

Nl_i is set to 1 if a reviewer left a question unanswered, otherwise 0.

5.3.5 Application of the FOCA methodology for Summative Evaluation. Two experts in data modeling and ontology construction (Harris, Tao) manually reviewed the constructed CRO using the FOCA method as described in the preceding section.

5.4 Extrinsic Evaluation Methods

Extrinsic evaluation methods focus on the aspects of the ontology that make an ontology suitable for a particular purpose or use. This means the ontology must be tested against concepts like completeness and pragmatic factors (Gruber, 1995). The methods used for assessing the CRO against known parts of the Common Rule and assessing the adherence

of the representation to the original source material, i.e., 45 CFR § 46 subparts A through D, as well as the parts of the revised Common Rule dealing with biospecimen management, are described below.

5.4.1 Domain coverage analysis. Coverage analysis was done by examining which sections of the Common Rule the classes and relationships in the CRO are derived from. For example, 45 C.F.R. §46.116, which defines the basic elements required for informed consent has 113 terms defined in the CRO; 82 derived from the 1991 version of the Common Rule and 31 from the 2017 revision that includes broad consent for secondary research. Coverage analysis is carried out specific to the domain knowledge represented in the CRO and does not include the classes and other terms imported from the foundational upper level and mid-level ontologies, such as BFO, RO, and OBI. For example, the relation '*BFO:inheres in*' was not included in the coverage analysis since it is not relevant to the domain knowledge itself, even though it is a foundational relation upon which many restriction classes are defined in the CRO. Similarly, classes such as '*BFO:independent continuant*' were not counted.

5.4.2 Corpus-based assessment. I and other members of the Tao and Xu labs at the School of Biomedical Informatics collected 178 informed consent form templates from CTSA and IRB websites at academic medical centers. From this corpus two reviewers randomly sampled ten templates. This sample corpus was independently annotated by two annotators familiar with the Common Rule (Sankaranarayanapillai, Zhang) using the CLAMP program (Soysal et al., 2017) to indicate the subsection of the Common Rule each annotation represented and the term in the CRO ontology. Due to

resource constraints, one annotator only finished four templates, while the other completed the whole set. Consequently, analysis was only done on the four templates completed by both reviewers. The annotators were also asked to record any required concepts from the Common Rule which were not present in the informed consent templates. These were then scored against the Common Rule. Data was recorded for each term and summarized into a two by four table as shown in Table 30, below. Precision, recall, accuracy, and F_1-score were calculated.

Table 30

Sample Table of Results for Corpus-based Assessment of Completeness and Accuracy

Common Rule Annotation	Annotated term is required by Common Rule	Annotated term is not required by Common Rule
Correctly annotated	True positives	False positives
Incorrectly annotated	False negatives	True negatives

5.4.3 Competency question-based evaluation. Competency questions are a straightforward way of assuring that an ontology meets the stated requirements of its intended use. The method relies on using a set of questions regarding a scenario that users would like to know answers about. These are then translated to description logics-based queries, run against the ontology, and the results are checked to see if they are correct. Ideally, competency questions are used iteratively throughout the development process (Bezerra, Freitas, & Santana, 2013; Ren et al., 2014).

Competency questions for the CRO were derived from a checklist developed by the University of Michigan IRB from DHHS Office of Human Research Protections in 45 CFR §46.116 and 21 CFR §50.25 Food and Drug Administration. The competency

questions are shown in Table 31, below. The original source material from the Michigan IRB is found in Appendix A.

Table 31

Competency Questions for Evaluating the CRO

Information content questions:
Does the informed consent form contain a statement that the study involves research?
Does the informed consent form contain a statement explaining the purposes of the research?
Does the informed consent form contain the expected duration of the subject's participation?
Does the informed consent form contain a description of the procedures to be followed?
Does the informed consent form identify procedures which are experimental?
Does the informed consent form identify foreseeable risks or discomforts to the subject?
Does the informed consent form describe benefits to the subject or others?
Does the informed consent form disclose alternative procedures or treatments that might be advantageous to the subject?
Does the informed consent form describe how confidentiality of records identifying the subject will be maintained?
Does the informed consent form contain, for research with greater than minimal risk, an explanation if compensation for injury is available?
Does the informed consent form contain, for research with greater than minimal risk, an explanation of what medical treatments are available if an injury occurs?
Does the informed consent form contain, for research with greater than minimal risk, an explanation of what medical treatments for injury consist of?
Does the informed consent form contain, for research with greater than minimal risk, an explanation of where further information may be obtained?
Does the informed consent form contain an explanation of who to contact with questions about the research?
Does the informed consent form contain an explanation of who to contact with questions about research subjects rights?
Does the informed consent form contain an explanation of who to contact in the event of a research-related injury?
Does the informed consent form contain a statement that participation is voluntary?
Does the informed consent form contain a statement that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled?
Does the informed consent form contain a statement that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled?
What type of people are considered public officials in the Common Rule?
What type of people are protected individuals under 45 CFR 46 Subpart B?
What type of people are protected individuals under 45 CFR 46 Subpart C?
What type of people are protected individuals under 45 CFR 46 Subpart D?
What type of people are considered vulnerable individuals under the Common Rule?
Who has authority to give consent or permission on behalf of another individual?

Completeness questions:

Is the informed consent form valid (i.e. does it contain all the necessary required elements of informed consent per §46.116)?

Does the informed consent form contain any optional items of informed consent per §46.116?

If so, what are the optional items present on the consent form?

Does the consent contain the required statements for use of biospecimens in secondary research?

So that checking of the competency questions could be done iteratively during the later phases of ontology development, a class *CRO:indicator* was introduced into the ontology, as an '*IAO:information content entity*,' that contained as subclasses all the competency questions from Table 31. Test data was implemented as instance data (sometimes termed *individuals*) within the appropriate classes in the ontology. Competency question testing was then achieved simply by running the HermiT reasoner and querying the associated CRO subclass to see if the appropriate inferred instance data appeared. This proved an effective way to iteratively perform competency question testing.

5.4.4 Natural language sentence evaluation. Natural language sentences were generated from the classes, properties, and individuals in the ontology. For example, Abacha (Ben Abacha, Dos Reis, Mrabet, Pruski, & Da Silveira, 2016) suggests using patterns such as the following for generating sentences:

- *A rdfs:subClassOf B*
- *P rdfs:subPropertyOf Q*
- *P rdfs:domain D*
- *P rdfs:range R*
- *I rdf:type A*

- *I P J* (individuals I and J are linked by the property P)

Where *A* and *B* represent class labels, *D* is a domain name, *I* and *J* are individuals, *P* and *Q* are property names, and *R* is a range. For example, within the ontology the fact that all benefit types defined by the Common Rule should be considered when assigned a benefit rating by an Institutional Review Board (IRB) is expressed as:

$$CRO\_0001606 \equiv \exists\, OBI\_0000295.CRO\_0000088 \quad (2)$$

which the sentence generator (correctly) converts to: *'every benefit type from the Common Rule'* (CRO_0001606) is something that '*is the specified input of*' (OBI_0000295) a *'process for assessing the type of benefit*' (CRO_0000088).

In practice, the generation of natural language sentences from the complex relationships in an ontology is a hard problem, but a number of tools do exist. In this case, natural language sentences were generated from the CRO using the Hootination tool of Amith (Muhammad Amith et al., 2017). These sentences were then assessed by Drs. Harris and Tao to answer the question 'is the sentence correct relative to the domain' and scored on a categorical scale of 'correct', 'incorrect', or 'uncertain'. Results were recorded for each sentence in an excel spreadsheet and analyzed with Cohen's Kappa.

5.4.5 Burton-Jones quality evaluation. Quality of the CRO was computed using the Burton-Jones methodology (Burton-Jones et al., 2005) as implemented via the OntoKeeper tool of Muhammad Amith (M. Amith & Tao, 2015). In earlier work, Amith and his colleagues have shown this approach to be a valid quality metric using a group of

ontologies sampled from the National Center for Biomedical Ontologies ontology repository.

The statistic relies on the underlying semiotics of the ontology under evaluation, and consequently, the method allows for direct comparison of two or more ontologies. The method can be customized to account for parameters that cannot reasonably be assessed. Burton-Jones calculates an overall quality score, and subscores from the following four areas:

- *Semantic quality* – are the terms used in the ontology meaningful and interpretable, do they have clarity, and are terms used in a consistent way? Variables measured in this area are *interpretability*, *clarity*, and *consistency*.
- *Syntactic quality* — is the syntax correct and has appropriate breadth? Variables measured in this area are *lawfulness* and *richness*.
- *Pragmatic quality* — does the number of classes and properties provide comprehensive coverage of what is being modeled, are they relevant to the tasks and entities being modeled, and is the information contained relevant? These variables are termed *comprehensiveness*, *accuracy*, and *relevance*.
- *Social quality* — what is the perceived authority and history of the ontology, that is, do other ontologies rely on it and how many times has it been used. The variables measured in this category are termed *authority* and *history*.

Note that the Burton-Jones framework contains both intrinsic and extrinsic evaluation methods.

Overall quality Q of an ontology under evaluation is deemed to be a weighted sum of all the variables:

$$Q = w_1 \cdot \sum Q_{semantics} + w_2 \cdot \sum Q_{syntactics} + w_3 \cdot \sum Q_{pragmatics} + w_4 \cdot \sum Q_{social} \quad (3)$$

Weights $\{w_1, w_2, w_3, w_4\}$ must sum to one and are usually assumed to be equal. The resulting score Q is in the range of $(0,1)$. Since the CRO ontology is a new ontology, it is presumed not to possess a social quality yet; hence this term was not included the assessment. Consequently, w_1, w_2, and w_3 were all set to 1/3. Details of the calculation of the actual variables are generally straightforward and are described in the following paragraphs.

Semantic quality is computed using the variables *interpretability*, *clarity*, and *consistency*. Computation of *clarity* and *interpretability* rely on a suitable well-characterized corpus of terms. Consequently, the WordNet corpus (Fellbaum, 2010) was used. WordNet is a general corpus of words and words senses typically used for this purpose. The variable $Q_{semantics}$ (representing semantic quality) is computed as follows:

$$Q_{semantics} = \left(\frac{1}{3} \cdot interpretability\right) + \left(\frac{1}{3} \cdot consistency\right) + \left(\frac{1}{3} \cdot clarity\right) \quad (4)$$

Where:

- *Interpretability* is defined as the total number of terms (defined here as classes, properties, and instances) with a word sense as listed in WordNet, divided by the total number of terms used in the ontology.

- *Consistency* is computed as the number of misused terms divided by the total number of terms in the ontology or number of duplicate terms over the total terms.
- *Clarity* is computed as the average number of word senses from WordNet for all the terms divided by the total number of terms in the ontology.

The Burton-Jones syntactic quality metric measures the degree of use of the expressive power of the OWL-2 language. Syntactic quality is computed as shown in equation (5) below:

$$Q_{syntactics} = \left(\frac{1}{2} \cdot lawfulness\right) + \left(\frac{1}{2} \cdot richness\right) \quad (5)$$

Where:

- *Lawfulness* is the total number of syntactic violations over the total number of statements in the ontology. In practice since modern ontology editors such as Protégé enforce semantic correctness in the underlying syntax this number is very low if not zero.
- *Richness* is the number of syntactic elements utilized over the total syntactic elements available in the underlying syntax.

The computation of pragmatic quality is shown in equation (6):

$$Q_{pragmatics} = \left(\frac{1}{3} \cdot comprehensiveness\right) + \left(\frac{1}{3} \cdot accuracy\right) + \left(\frac{1}{3} \cdot relevance\right) \quad (6)$$

Where:

- *Comprehensiveness* is the total number of classes and properties in the ontology being measured, divided by the average number of classes and properties in a (similar) collection of ontologies. To calculate the denominator the average of the number of classes and properties in the ICO, DUO, CRO, OMIABIS, OBIB, and d-acts ontologies were used, as recommended by the author.
- *Accuracy* is the number of 'false' or inaccurate statements over the total number of statements scored from the ontology. Scoring requires assessment by the domain experts to ascertain the percentage of correct statements in the ontology. This was accomplished by using the results of the natural language sentence evaluation described earlier in Section 5.4.4.
- *Relevance* is similar to *accuracy*, and similarly requires the input of the domain experts, but it addresses the question *'how many of the scored statements are actually relevant to the decisions I care about.'* It is computed as the number of classes and properties deemed relevant by the domain experts over the total number of statements being scored.

The OntoKeeper tool does not implement relevance at this time so this parameter was not used. Consequently, the weights for the pragmatics section were adjusted to be (1/2, 1/2, 0) for the measures comprehensiveness, accuracy, and relevance, respectively.

The overall quality Q of CRO is consequently given as:

$$Q = \left(\frac{1}{3} \cdot Q_{semantics}\right) + \left(\frac{1}{3} \cdot Q_{syntactics}\right) + \left(\frac{1}{3} \cdot Q_{pragmatics}\right) \quad (7)$$

Results of the quality assessment allow us to derive a direct and comparable measure of CRO against the similar statistics derived from ontologies identified in Chapter 2. In addition to the overall score, the individual component scores of the Burton-Jones model are comparable and can be used to gain insight into the various strengths and weaknesses a single ontology such as CRO, or a set of ontologies representing a given domain.

5.4.6 Quality comparison of CRO to OBO Foundry ontologies. To assess the quality of the CRO against other OBO Foundry ontologies containing regulatory terms, the Burton-Jones framework was applied to the principal ontologies described in Chapter 2 that are relevant to human subjects regulation. The five ontologies selected were:

- ICO – Informed Consent Ontology;
- DUO – Data Use Ontology;
- OMIABIS – Ontologized Minimal Information About BIobank data Sharing;
- OBIB – Ontology for Biobanking; and
- d-acts – Document acts.

Since neither accuracy nor relevance could be calculated for these five ontologies, the weights for the pragmatics section were adjusted to be (1, 0, 0) for the measures comprehensiveness, accuracy, and relevance, respectively.

5.5 Relating Methods to Other Evaluation Frameworks

This section relates the evaluation methods utilized during the construction and evaluation of CRO to measures suggested by other authors. These mappings are my own and are based solely on my understanding of the methods and quality evaluation

frameworks involved. **Error! Reference source not found.** relates methods to Vrandečić's framework, and Table 33 relates the methods to the NIST lifecycle model criteria. The intent of these tables is to demonstrate that the quality evaluation undertaken for the CRO has at least one method in each category, addressing a broad spectrum of the types of criteria of evaluation proposed by the ontology evaluation community.

Table 32

Evaluation Methods Applied to Components of Vrandečić's Model

	Vrandečić's Components							
Methods	Acc	Adapt	Clarity	Complete	Comp	Concise	Consist	Org
Intrinsic measures								
Descriptive characterization				X	X			
HermiT reasoner					X			X
Informal feedback	X		X	X		X	X	
FOCA analysis	X	X	X		X		X	
Burton-Jones analysis	X		X	X		X	X	
Extrinsic measures								
Qualitative review of concept maps with subject matter experts	X		X	X				
Domain coverage analysis				X				
Corpus-based assessment	X			X				
Competency questions	X		X					
Natural language sentences							X	
Burton-Jones assessment of CRO	X		X	X		X	X	

	Vrandečić's Components							
Methods	Acc	Adapt	Clarity	Complete	Comp	Concise	Consist	Org
Burton-Jones assessment against OBO Foundry ontologies	X		X	X		X	X	

Abbreviations used: Acc = accuracy; adapt = adaptability; complete = completeness; comp = computational efficiency; concise = conciseness; consist = consistency; org = organizational fitness.

Table 33

Evaluation Methods Used and Components Assessed in the NIST Model

Intrinsic measures	Intelligibility	Fidelity	Craftsmanship	Fitness
Descriptive characterization			X	
HermiT reasoner			X	X
Informal feedback	X	X	X	X
FOCA analysis	X		X	
Burton-Jones analysis	X	X	X	X
Extrinsic measures				
Qualitative review of concept maps with subject matter experts		X		X
Domain coverage analysis		X		X
Corpus-based assessment		X		X
Competency questions		X		X
Natural language sentences	X	X		X
Burton-Jones assessment of CRO	X	X	X	X
Burton-Jones assessment against OBO Foundry ontologies	X	X	X	X

Note: Refer to Table 23 for the definitions of the column labels. Deployability is omitted since it is not relevant to this work at this time.

Note that the Burton-Jones methods span the characterization of intrinsic and extrinsic evaluation. The syntactic measures of Burton-Jones are meant to measure intrinsic

characteristics, while the pragmatic measures address extrinsic factors, and the semantic measures address both intrinsic and extrinsic factors.

5.6 Relating Evaluation Methods to the NIST Lifecycle Model

This chapter ends by demonstrating what elements were measured at each phase of the ontology development process. Refer to Table 9 to review which evaluation task was done in each phase.

Table 34

Evaluation Methods Used and Quality Measures Assessed in the NIST Model by Phase

NIST / METHONTOLOGY Phase	Intelligibility	Fidelity	Craftsmanship	Fitness
Requirements Development				
Planify		X		X
Specification		X		X
Ontological Analysis				
Knowledge Acquisition		X		X
Conceptualization		X		X
Ontology Design				
Knowledge Acquisition	X	X		X
Conceptualization	X	X		X
Integration	X	X	X	X
Ontology Development				
1. Informal modeling				
Conceptualization	X	X	X	X
2. Formalization of competency questions				
Specification		X		X
3. Formal modeling				
Implementation	X	X	X	X
Integration		X	X	
Evaluation	X	X	X	X

Note: Refer to Table 23 for the definitions of the column labels. Deployability is omitted since it is not relevant to this work at this time.

As can be seen, the intrinsic measure of craftsmanship is not used in the earlier phases of the lifecycle, and the extrinsic measures of intelligibility depend to some extent on the details of the ontology construction. This is as one would expect.

Chapter 6. Results of Ontology Evaluation

This chapter presents the results of the evaluation of the CRO and describes triangulation of these measures relative to the various axes of evaluation described in the previous chapter.

6.1 Competency Question Evaluation

Test data for seven competency questions were entered into the CRO. These seven questions were directly derived from the competency questions described in Table 31 and are as shown in Table 35.

Table 35

Competency Question Test Scenarios

Test	Description
Test A	Consent form contains all required elements of informed consent.
Test B	Consent form contains none of the required elements of informed consent, but the study has been assessed a risk magnitude of minimum risk.
Test C	The consent form contains one or more additional elements of informed consent.
Test D	Contains no required elements of consent, one optional element of consent, and a risk magnitude.
Test E	Contains only optional elements of broad consent for biospecimen use.
Test F	Contains all required elements for broad consent, but no risk magnitude.
Test G	Contains only an indicator that research is greater than minimal risk

As noted before, tests were run by using the *CRO:indicator* classes asserted in the ontology to query the results by using the HermiT reasoner. This is shown for test A in Figure 22.

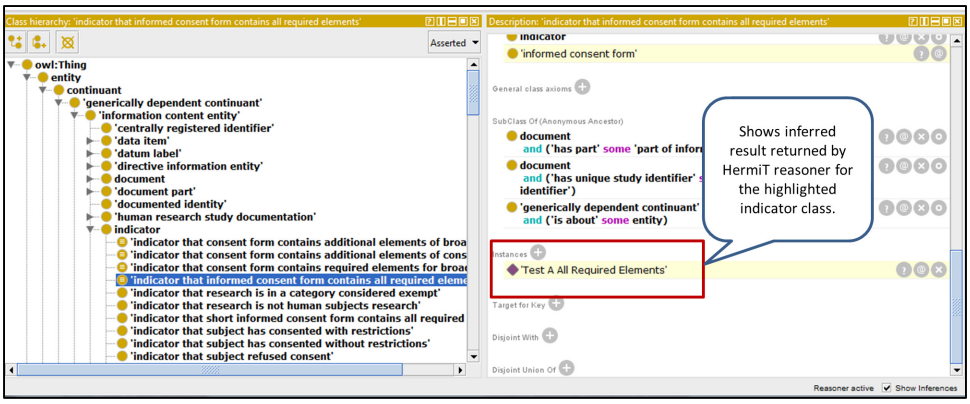


Figure 22. Use of indicator classes to query competency question test data.

The expected and obtained results and the time in milliseconds to compute the instance-level results for each test are shown in Table 36.

Table 36

Results of Competency Question Evaluation by Test Scenario

	Risk magnitude		Regular IC		Broad IC	
Description	Min risk	>Min Risk	Required	Optional	Required	Optional
Expected result	B, D	A, G	A	A, C, E	F	A, C, E
Observed result	B, D	A, G	A	A, C, E	F	A, C, E
Time (m.s.)	405	383	973	488	805	487

Note: IC – Informed Consent. A – G refer to the test scenarios in Table 35.

6.2 Coverage of the Common Rule.

To perform coverage analysis, the following SPARQL code was executed to extract the URI's of all terms in the ontology along with their definition and definition source.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX IAO: <http://purl.obolibrary.org/obo/>


SELECT ?uri ?label ?definition_source

        WHERE {

        ?uri IAO:IAO_0000119 ?definition_source .

        ?uri rdfs:label ?label

}
```

Results, organized by the source of the term definition are shown in Table 37.

Table 37

Sources of Definitions of Terms Used in the CRO

Definition Source	Total terms
1991 Common Rule	409
2017 Common Rule	61
Defined by this author	110
Various online dictionaries	11
Various online legal dictionaries	10
Federal government websites (HHS, etc.)	4
NLM – UMLS	3
www.research.olemiss.edu/irb	1
Totals:	609

Error! Reference source not found. shows the number of terms with definitions taken directly from the Common Rule. Note that while the CRO ontology has a total of 1,076 terms, not all of these terms were directly asserted in CRO; some were imported from other ontologies such as BFO, IAO, and OBI. As shown in **Error! Reference source not found.**, 470 terms were directly derived from the Common Rule. Other terms were introduced to represent processes and concepts inferred by the text of the Common Rule. An example of this is the concept of age. Legal decisions regarding what consenting practices are necessary involve considerations of emancipation status and biological age, and these notions had to be introduced in the CRO. They are not technically part of the Common Rule itself and are not represented in these counts. In total 27 of the possible 44 sections of 45 C.F.R. §46 subparts A – D, or 61%, are covered directly by terms in the CRO.

Table 38

CRO Coverage of the 1991 and 2017 Versions of the Common Rule by Section

Section	1991	2017	Total
Subpart A. Basic HHS Policy for Protection of Human Research Subjects			
§45 CFR 46 Preamble	5		5
§46.101 To what does this policy apply?	76		76
§46.102 Definitions.	38	20	58
§46.103 Assuring compliance with this policy--research conducted or supported by any Federal Department or Agency.	3		3
§46.104 Exempt research (2017 version only, unused in 1991 version)		8	8
§46.105 - §46.106 [Reserved]			
§46.107 IRB membership.	1		1
§46.108 IRB functions and operations.			0
§46.109 IRB review of research.	1		1
§46.110 Expedited review procedures for certain kinds of research involving no more than minimal risk, and for minor changes in approved research.	1		1
§46.111 Criteria for IRB approval of research.	21	2	23
§46.112 Review by institution.			0
§46.113 Suspension or termination of IRB approval of research.			0
§46.114 Cooperative research.			0
§46.115 IRB records.	2		2
§46.116 General requirements for informed consent.	87	31	118
§46.117 Documentation of informed consent.	16		16
§46.118 Applications and proposals lacking definite plans for involvement of human subjects.			0
§46.119 Research undertaken without the intention of involving human subjects.			0
§46.120 Evaluation and disposition of applications and proposals for research to be conducted or supported by a Federal Department or Agency.			0

Section	1991	2017	Total
§46.121 [Reserved]			0
§46.122 Use of Federal funds			0
§46.123 Early termination of research support: Evaluation of applications and proposals.			0
§46.124 Conditions.			0
Totals:	251	61	312
Subpart B. Additional Protections for Pregnant Women, Human Fetuses and Neonates Involved in Research			
§46.201 To what do these regulations apply?	3		3
§46.202 Definitions.	11		11
§46.203 Duties of IRBs in connection with research involving pregnant women, fetuses, and neonates.			0
§46.204 Research involving pregnant women or fetuses.	18		18
§46.205 Research involving neonates.	11		11
§46.206 Research involving, after delivery, the placenta, the dead fetus or fetal material.	13		13
§46.207 Research not otherwise approvable which presents an opportunity to understand, prevent, or alleviate a serious problem affecting the health or welfare of pregnant women, fetuses, or neonates.	8		8
Totals:	64	0	64
Subpart C. Additional Protections Pertaining to Biomedical and Behavioral Research Involving Prisoners as Subjects			
§46.301 Applicability			0
§46.302 Purpose.			0
§46.303 Definitions.	11		11
§46.304 Composition of Institutional Review Boards where prisoners are involved.			0
§46.305 Additional duties of the Institutional Review Boards where prisoners are involved.			0

Section	1991	2017	Total
§46.306 Permitted research involving prisoners.	13		13
Totals:	24	0	24
Subpart D. Additional Protections for Children Involved as Subjects in Research			
§46.401 To what do these regulations apply?	3		3
§46.402 Definitions.	9		9
§46.403 IRB duties.			0
§46.404 Research not involving greater than minimal risk.			0
§46.405 Research involving greater than minimal risk but presenting the prospect of direct benefit to the individual subjects.	11		11
§46.406 Research involving greater than minimal risk and no prospect of direct benefit to individual subjects, but likely to yield generalizable knowledge about the subject's disorder or condition.	10		10
§46.407 Research not otherwise approvable which presents an opportunity to understand, prevent, or alleviate a serious problem affecting the health or welfare of children.	8		8
§46.408 Requirements for permission by parents or guardians and for assent by children.	19		19
§46.409 Wards.	10		10
Totals:	70	0	70
Overall Totals:	409	61	470

6.3 Corpus-based Assessment

Due to resource constraints, the two independent annotators were only able to complete manual annotation of four randomly sampled informed consent forms. This resulted in the identification of 1,125 unique annotations. 102 concepts that should have been annotated from the Common Rule were missed by the annotators. In 22 cases the wrong

CRO concept was used for the annotation. The annotators introduced four new terms into the annotations, resulting in 35 annotations. The new terms, however, were found not to be directly derived from the Common Rule, and consequently were treated as false negatives in the analysis. Results of the assessment of the performance of the CRO using these manual annotations is shown in Table 39. H_0 is 'CRO can't be used to find a relevant set of terms for informed consent documents.'

Table 39

Precision/Recall Results for Corpus-based Assessment

Common Rule annotation	Found by annotator	Missed by annotator	Totals	Precision
Proper use of CRO	1068	102	1170	0.9128
Improper use of CRO	22	35	57	
Totals	1090	137		
Recall	0.9798			$F1 = 0.9451$
Accuracy	0.8989			

Recall is a good estimate of coverage of the ontology, as viewed by two annotators attempting to use it. Recall was 0.9798, $p<.001$, CI [0.9687, 0.9909]. F_1 is 0.9451. In this case, accuracy used to test how well the CRO performs on both true positive and false negative results. Accuracy was 0.8989, $p<.001$, CI [0.8219,0.9759]. Because in this case we are not attempting to evaluate the quality of the annotation but rather the quality of coverage, the precision statistic is less relevant. The eventual use of the CRO with NLP and machine learning methods will ultimately give a more precise measure of precision.

6.4 FOCA Evaluation

Results of the FOCA evaluation are shown in Table 40. As a reminder to the reader, FOCA is an intrinsic evaluation measure performed by humans that tries to assess ontology quality. The scores range on the interval (0-1).

Table 40

Results of FOCA Evaluation

Rater	FOCA	Human expression score
Harris	0.999	.9166
Tao	0.998	.9583
Mean	0.9985	.9375

6.5 Sentence Construction Evaluation

Table 41 shows the distribution of sentences generated for evaluation by human reviewers (Drs. Tao, Harris). 1,494 sentences were generated. Sentences only containing terms from the BFO, RO, and OBO Foundry ontologies were eliminated, leaving 999 sentences. Of these, some terms of the ontology are only intended as internal metadata, such as '*obsolete classes'*, and some test data used during the development, were eliminated, resulting in 872 sentences for evaluation, as shown in below.

Table 41

OWL Properties Used in Sentence Level Analysis

OWL Property Type	Eliminated from Sentence Analysis	Used in Analysis	Total
AsymmetricObjectProperty	0		0
ClassAssertion	35	56	91
DataPropertyAssertion	0		0
DataPropertyDomain	0		0
DataPropertyRange	0		0
DifferentIndividuals	0		0
DisjointClasses	0	13	13
EquivalentClasses	0	16	16
FunctionalDataProperty	0		0
FunctionalObjectProperty	0		0
InverseObjectProperties	0	2	2
ObjectPropertyAssertion	73		73
ObjectPropertyDomain	2	2	4
ObjectPropertyRange	2	2	4
SubClassOf	15	779	794
SubDataPropertyOf	0		0
SubObjectPropertyOf	0	2	2
SubPropertyChainOf	0		0
TransitiveObjectProperty	0		0
Totals	127	872	999

Agreement statistics for the two raters are shown in Table 42.

Table 42

Results of Sentence Evaluation for Accuracy by Multiple Raters (N=861)

Rater	Agreed with Sentence	Disagreed with Sentence	Uncertain	Unanswered	Total
Harris	822 (95.47%)	20 (2.32%)	19 (2.21%)	0 (0.0%)	861
Tao	816 (94.77%)	0 (0.0%)	27 (3.14%)	18 (2.09%)	861
Total	1638 (95.12%)	20 (1.16%)	46 (2.67%)	18 (1.05%)	1722

One rater did not score 18 sentences, consequently, these were eliminated from further analysis. The variable 'uncertain' was also recoded as disagreement with the sentence being evaluated, leaving the final results as shown in Table 43.

Table 43

Final Results of Sentence Evaluation for Accuracy by Multiple Raters (N=843)

Rater	Agreed with Sentence	Disagreed with Sentence	Total
Harris	804 (95.37%)	39 (4.63%)	843
Tao	816 (96.80%)	27 (3.20%)	843
Total	1620 (96.09%)	66 (3.91%)	1686

Tests of Kendall's coefficient of concordance (*W)* and Cohen's kappa (κ) with Fleiss' adjustment for nominal ratings for two raters were calculated. Kendall's coefficient is a non-parametric test of agreement between the two raters when both either agree or disagree. The kappa statistic includes the level of disagreement between the raters as part of the calculation of kappa. The test of concordance between the two raters was found to be statistically significant (W = 0.57502 $F(841,841)=1.353$, $p<.0001$). Despite statistically significant overall agreement (κ=0.14848 SE=0.034442, z=4.31118, $p<.0001$)

kappa did not reach the 0.61 value generally used for 'substantial' agreement. Further analysis of the results revealed that while there was significant agreement on sentences judged correct, there was disagreement between the raters on sentences deemed incorrect. Both raters agreed that 783 of the 843 sentences were correct, and both judged only six (6) sentences incorrect, leaving 54 sentences on which the raters disagreed. This analysis clearly indicates opportunities to improve the coverage, accuracy, and clarity of the ontology.

6.6 Burton-Jones Analysis

To calculate the Burton-Jones *comprehensiveness* parameter, an estimate of the so-called 'library mean' of the total number of classes and object properties was required. This was calculated from the corresponding values in the ontologies being tested, as shown in Table 44.

Table 44

Comparison of Classes and Object Properties in CRO and other OBO Foundry Ontologies with Regulatory Components

Ontology	Classes	Object Properties	Totals
ICO	409	48	457
DUO	249	53	302
CRO	864	70	934
OMIABIS	427	75	502
OBIB	1021	72	1093
d-acts	232	65	297
Total	3202	383	3585
Average	534	64	598*

*This value is used as the library mean for use in calculating the Burton-Jones *comprehensiveness* results.

6.6.1 Burton-Jones Analysis of CRO. Table 45 shows the results of the individual calculation of the Burton-Jones score, both by subsection and for the overall quality measures as determined by OntoKeeper (M. Amith & Tao, 2015). Since the *accuracy* variable of Burton-Jones depends on the sentence evaluation described above, the mean value for agreement (96.09%) was used and the result for the *pragmatic* section was calculated manually.

Table 45

Results of Full Burton-Jones analysis of the Common Rule Ontology

Burton-Jones Metric	Feature Weight	CRO	Burton-Jones score
Syntactic Features			
Lawfulness	0.50	1	0.5
breached rules		0	
total axioms		8,571	
Richness	0.50	0.69	0.345
features used		27	
Total syntactic score			0.845
Semantic Features			
Interpretability	0.33	0.97	0.3201
number of terms		1,025	
terms with senses		999	
Consistency	0.33	1	0.33
inconsistent terms		0	
Clarity	0.33	0.98	0.3234
word senses		17,004	
Total semantic score			0.9735
Pragmatic Features			
Comprehensiveness	0.50	1.71	0.855
Accuracy	0.50	0.9609	0.4804
Relevancy	0	0	0
Total pragmatic score			1.3354
Aggregate Scores			
Syntactic Features	0.33	84.5	.2789
Semantic Features	0.33	97.35	.3213
Pragmatic Features	0.33	1.335	.4451
Burton-Jones Score			1.0452

6.6.2 Comparison of CRO to regulatory ontologies from the OBO Foundry.

Comparison of the CRO to other ontologies from the OBO Foundry that had regulatory components dealing with consent, data sharing, or biospecimen sharing resulted in the following Burton-Jones scores, as shown in Table 46. The OBO Foundry ontologies compared to CRO in the table were those discovered as a consequence of the joint literature review and ontology repository search described in Chapter 2.

Table 46

Comparison of Burton-Jones Results by Ontology

Burton-Jones Metric	CRO	ICO	DUO	OMIABIS	OBIB	d-acts	Mean	*sd*
Syntactic Features								
Lawfulness	1	1	1	1	1	1	1	0
breached rules	0	1	1	3	43	2	8.33	15.53
total axioms	8,571	4,687	2,974	5,093	9,618	3,191	5,689	2,540.13
Richness	0.69	0.59	0.62	0.56	0.59	0.64	0.62	0.04
features used	27	23	24	22	23	25	24.0	1.63
Semantic Features								
Interpretability	0.97	0.95	0.83	0.95	0.94	0.95	0.93	0.05
number of terms	1,025	468	335	516	1,151	330	637.5	327.46
terms with senses	999	445	277	492	1,085	314	602.0	320.47
Consistency	1	1	1	1	1	1	1	0
inconsistent terms	0	0	1	0	0	1	0.33	0.47
Clarity	0.98	0.97	0.97	0.97	0.99	0.97	0.98	0.01
word senses	17,004	7,311	3,286	7,610	15,875	3,578	9,110.66	5,448.07
Pragmatic Features								
Comprehensiveness	1.71	0.78	0.56	0.86	1.92	0.55	1.06	0.55
Accuracy	0	0	0	0	0	0	0	0
Relevancy	0	0	0	0	0	0	0	0
Aggregate Scores								
Syntactic Features	84	79	80	78	79	82	80.33	2.05
Semantic Features	97	96	92	96	96	96	95.50	1.61
Pragmatic Features	171	78	56	86	192	55	106.33	54.63
Burton-Jones Score	1.17	0.84	0.76	0.86	1.22	0.77	0.94	0.19

Significance results are not calculated for this table, as the use of common upper and mid-level ontologies introduces covariates that are not possible to control for. The overall

quality score of the CRO, however, compares quite favorably with other members of the cohort.

6.7 Relating Evaluation Results to Goals of Ontology Quality

This chapter summarized the results of the various intrinsic and extrinsic evaluation criteria for the CRO across the NIST suggested quality dimensions of intelligibility, fidelity, craftsmanship, and fitness. Table 47 attempts to summarize these results and allows one to see the triangulation of these results. Qualitative or descriptive results are only noted, while quantitative results are given.

Table 47

Evaluation Methods Used and Components Assessed in the NIST Model

Intrinsic measures	Intelligibility	Fidelity	Craftsmanship	Fitness
Descriptive characterization			X	
HermiT reasoner test			Successful	Successful
Informal feedback	X	X	X	X
FOCA analysis	0.9375 (human expression)		0.9985 (other components)	
Burton-Jones analysis	0.9735 (semantics)	1.335 (pragmatics)	.84 (syntactic)	1.0452 (overall)
Extrinsic measures				
Qualitative review of concept maps with subject matter experts		X		X
Domain coverage analysis		61% of CR sections		61% of CR sections
Corpus-based assessment		0.8989 (accuracy)		0.9798 (recall)
Competency questions		100%		100%
Natural language sentences	96.09% agreement	96.09% agreement		96.09% agreement
Burton-Jones assessment of CRO	0.9735 (semantics)	1.335 (pragmatics)	.84 (syntactic)	1.0452 (overall)
Burton-Jones assessment against OBO Foundry ontologies	0.9735 (semantics)	1.71 (pragmatics)	.84 (syntactic)	1.17 (overall)

Note: Deployability is omitted since it is not relevant to this work at this time.

Chapter 7. Conclusion and Future Directions

An ontology is a physical and scholarly artifact that represents a 'conceptualization of a domain' in a computational manner. The work presented in this dissertation describes the first steps toward such a computable model of the domain of regulated human research and is grounded by connections to the relevant regulatory frameworks. It represents an important step in the field as the need to share data across multiple research teams and institutions is becoming a major emphasis for both research teams and funding agencies such as the NIH. Currently, participant consent, for research involving collection and sharing of specimens and data, is collected using consent forms with considerable variation between groups or projects. This makes it difficult to understand the precise desires of the research subject, especially when aggregating specimens or data from multiple studies, research teams, or institutions. The use of structured taxonomy and extensive metadata frameworks are an approach to this problem.

As with any ontology, it is expected that this work will evolve over time. The ontology presented in this dissertation was constructed from knowledge derived from the Common Rule and human subject matter experts. It was encoded into an ontology using first principles that align it with a realism philosophy, and with an important, community-driven body of work in the field. As noted earlier, there is substantial interest in this work

as there is a lack of work in the field focused on facilitating sharing of data and other associated physical artifacts such as biospecimens.

This chapter summarizes the results from the literature searches, the knowledge elicitation process, the ontology construction process itself and the hybrid methodology used, and the results of the evaluation process. It concludes with a discussion of the importance of the work and how it should evolve in the future.

7.1 Summary of accomplishments and contributions

Materials presented in this dissertation describe work done to represent the U.S. Common Rule, 45 C.F.R. §46 subparts A – D as a formal OWL-2 ontology aligned with the BFO and the principles of the OBO Foundry. Best practices in knowledge elicitation, representation, ontology construction, and lifecycle methods were employed in its development. Extensive, iterative approaches were used for both qualitative and quantitative evaluation. The resulting ontology can answer competency questions regarding aspects of the Common Rule itself and can provide value sets for database systems attempting to store elements of informed consent and the risk factors involved in research. It also provides models of IRB processes and artifacts, such as the documentation of IRB decisions as required by the Common Rule. It is hoped that this work, as a starting point, will ultimately make a significant contribution in the field, and be taken up by the OBO Foundry as described under future plans.

Another accomplishment of this work is the creation of a hybrid method for ontology development, construction, and evaluation that uses a lifecycle based approach. It incorporates research from the area of ontology lifecycle design, detailed tasks for

ontology construction, makes suggestions for knowledge elicitation, and incorporates ongoing evaluation across the knowledge elicitation and construction lifecycles.

In Chapter 2 gaps were described in the field of informatics regarding the state of ontological frameworks for representing the regulation of human subjects research in the United States. This was accomplished through a review of the literature and a survey of two major ontology repositories, the Ontology for Biomedical Ontology (OBO) Foundry, and the National Center for Biomedical Ontology. It was also indicated informally by conversations with colleagues, including members of the OBO Foundry and OBI consortium who are now attempting to extend existing OBO Foundry ontologies to fill these gaps. The gaps appeared especially acute in the rapidly emerging areas of large-scale data interchange and biobanking.

Chapter 3 provided a short review of contemporary lifecycle frameworks, construction methods, and best practices. A hybrid model for ontology construction was described that integrated a NIST-developed lifecycle model (Neuhaus et al., 2013) and the METHONTOLOGY (Fernández et al., 1997) ontology construction framework. The remainder of the chapter discussed in detail how the hybrid model would be used in the ontology construction.

Chapter 4 described the knowledge elicitation and technical construction processes of the CRO in considerable detail and related the technical artifacts directly back to the hybrid lifecycle and construction model that was developed in Chapter 3.

A short review of ontology evaluation techniques was presented in Chapter 5. The chapter also described evaluation criteria that were used to measure different

characteristics of the CRO across both intrinsic and extrinsic dimensions and along different axes of evaluation. It demonstrated how the evaluation methods align with the hybrid ontology construction and evaluation method. The rest of Chapter 5 presented details for the application of the evaluation criteria to the CRO.

Chapter 6 presented results associated with the various quality evaluations performed on the CRO, and via triangulation of results, showed that major ontological quality dimensions were achieved, leading to the conclusion that the CRO is a high-quality ontology.

7.2 Generalizability and Range of Applications

The hybrid lifecycle construction and evaluation method developed as part of specific aim six is an extensible framework that can support additional construction or evaluation tasks. It provides both a high-level lifecycle framework coupled with practical details for guiding an ontology developer. This is especially useful for first-time ontology developers or people with expertise in other fields but lacking substantial ontology development backgrounds.

There are a broad range of applications for the CRO itself. The first, and simplest, is as a foundational knowledge base of aspects of the Common Rule. Terms in the ontology are linked directly back to their source paragraphs in the Common Rule, and when warranted, other federal regulations via the use of OWL annotation properties. The work should provide a basis for extension to specific state and local laws and regulations regarding human subjects research. Second, the work provides a basis for developing natural language processing applications to support data release and regulatory

workflows based on understanding consent and protocol documents. Large retrospective collections of biospecimens and linked data often have only written (i.e., paper-based) informed consent forms associated with them, limiting the ability of researchers to share this data at scale. Using the CRO as the basis for an annotation system should provide a useful framework for named entity recognition and other machine-classification algorithms to categorize the metadata relating to data and specimen type, participant sharing preferences, and other similar data encoded in consent forms, research study protocol documents, and existing databases containing subject preferences.

7.3 Discussion

This study has demonstrated the feasibility and value of ontology evaluation across the development lifecycle. This is one of the few projects to use, evaluate, and report on a lifecycle grounded approach to continuous and comprehensive evaluation. Recent work by Amith, Tao, and colleagues demonstrated that of 200 randomly sampled ontologies from the National Center for Biomedical Ontologies BioPortal, only 15 had evidence of any formal evaluation (M. F. Amith et al., 2017).

Research question one asked if there were adequate representations of the Common Rule in existing ontology libraries. The findings from Chapter 2 indicated that this was not the case. Not only was little evidence found of systematic work on the problem, save for my own work on the Informed Consent Ontology, what work there was appears scattered in many ontologies and showed little evidence of being grounded in the actual legal regulations.

Research question two sought to discover if the BFO and OBO Foundry provide a suitable foundation for ontologies grounded in regulation such as the Common Rule. The Common Rule is primarily about the consenting processes, prevention of risk to subjects, providing assurances for protections to human subjects, and IRB processes indirectly concerned with the same. The realism perspective of the BFO implies that regulatory concepts can only be realized in processes that create continuant artifacts, such as documents. Because documents are represented as generically dependent continuants, this felt especially awkward to both this modeler and to those attempting to evaluate the ontology. It is also worth noting that the major document class used in the BFO is *IAO:document*, which was originally created to represent scholarly journal publications. This is not necessarily well aligned with representing scientific protocol and informed consent documents, although other OBO Library ontologies have used *IAO:document* in that fashion. This representation presented modeling challenges that made it difficult for domain experts to confirm the correctness of the resulting model. An example of this is the representation of one-to-many relationships that involve one or more processes which create one or more representations of an input object. A demonstration of this difficulty is the need to represent informed consent forms, which are real entities, as a general information entity given to a research subject. It becomes difficult to differentiate that first entity from the later representation of signed, executed informed consent forms. The consent forms are temporally separate entities from the original templates. One of the independent reviewers of the sentence constructs found that use of the BFO constructs

such as above introduced a certain degree lack of clarity and consequential uncertainty in the reviewing the sentences generated by the ontology.

OBO Foundry principles requiring alignment with other Foundry-based ontologies further constrained the development, sometimes in a semantically non-congruent fashion that limited precise representation. Again, this was especially noticeable in the results of the sentence evaluation. As noted earlier, many of the terms dealing with data sharing and regulation are scattered in the supposedly non-overlapping, orthogonal OBO Foundry ontologies. Each of these ontologies had its own focal area that was not research regulation. When the transitive closure of these terms and their related axioms were included in the CRO via the MIREOT methodology the resulting implied semantics were difficult to integrate. Interestingly, this appeared to have more impact on the subclass, or so-called *is-a*, structure of the ontology then on the axioms and relations.

The OBO Foundry ontologies were also found to lack representation of fundamental legal and regulatory concepts. For example, they do not contain the concept of jurisdiction, which can vary depending on the type of law, the geographic location, or the governance hierarchy of federal, state, and local laws and regulations. Existing representations of such concepts would have been very helpful in this project. While the d-acts ontology of Brochhausen provides some work in this area, it is mostly based on the notion of describing contracts as ‘document acts’. Simply stated, ‘acts’ are a legal concept of recording what has been said or done. The d-acts ontology is still in an early stage of development, and at this point in time appears to lack the expressive power needed to

fully represent legal concepts found in the Common Rule; however, the CRO is aligned with it to the extent possible.

Research question three asked if approaches exist that provide an integrated and comprehensive evaluation framework across the ontology lifecycle. While there have been many suggestions for high-level evaluation frameworks, and a few lifecycle models have been developed, to date there appear to be few models that have proposed extensive, iterative evaluation across the whole lifecycle. The frameworks that do exist appear to focus on only a single aspect of development. There are a number of high-level lifecycle approaches, several ontology construction frameworks, and a large number of suggested evaluation methods. However, descriptions of methods that attempt to integrate these aspects in a pragmatic fashion appear to be lacking. Consequently, it was necessary to evolve my own framework as specific aim six, incorporating work from various aspects of the upper-level lifecycle, ontology construction, and ongoing intrinsic and extrinsic evaluation.

The FOCA evaluation method proved more suited to a cursory surface level structural evaluation than to a deep evaluation of the strengths and weaknesses of an ontology. The method was found to have inconsistencies in its measurements of the *ontological commitments* dimension. It is unclear why the *human expression dimension* was not included into FOCA's beta regression formula as a covariate. Finally, some of the questions concerning consistency and satisfiability don't appear particularly relevant when one considers that most ontologies are built using ontology workbenches like Protégé and would typically be undergoing continuous reasoning checks for consistency.

The method could possibly be improved with more granular questions. The goal, question, metric (GQM) method used in the evaluation did provide a nice framework for the lightweight evaluation of the ontology and may prove useful for ontology developers and domain reviewers wanting a quick quality assessment. It should be noted, though, that tools such as OntoKeeper (M. Amith & Tao, 2015) are probably more robust and require much less work for both domain experts and ontology engineers. These tools provide for rapid and high-quality evaluation whether used in a formative or summative fashion.

The ten Burton-Jones criteria based on semiotics proved to be very effective, even though only seven of them were used in this work due to resource constraints. It had the distinct advantage of evaluating the ontology simultaneously across multiple aspects of the intrinsic and extrinsic spectrum and provided a link between syntactic and semantic factors. The ability to directly compare to other ontologies in a library allowed for direct assessment to achieve at least comparable, if not superior, quality as the arithmetic mean of the library. Weaknesses in other methods became immediately apparent when they are compared to the Burton-Jones method. As implemented in OntoKeeper, this method appears superior in the level of granularity of its assessment, the time required for the analysis, the modular nature of the analysis design, and its ability to directly compare multiple ontologies. Unfortunately, it is not yet clear how to properly control for the effects of the upper- and mid-level ontologies when computing Burton-Jones on a library of BFO-based ontologies, preventing statements about the significance of findings using the method.

The suggested NIST evaluation criteria proved insufficiently granular to provide meaningful diagnostic quality measures for either formative or summative evaluation of the ontology. Again, the Burton-Jones method proved far superior, both as a rubric of evaluation, and pragmatically as implemented in OntoKeeper.

Finally, demonstrating the difficulty with issues of trust and information sharing that transcend technology, early in the project I hoped to collect a substantial number of informed consent and research protocol documents from both the University of Michigan and the University of Texas. This was to explore if information extraction methods and named entity recognition could be applied to the corpus to enrich the CRO. Unfortunately, obtaining access to this material ultimately proved infeasible, mostly due to internal policies of the groups involved. This has been recognized by federal agencies such as DHHS, and the 2017 revision of the Common Rule requires, in section § __.116(h) that:

> for each clinical trial conducted or supported by a federal department or agency, one IRB-approved informed consent form used to enroll subjects must be posted by the awardee or federal department or agency component conducting the trial on a publicly available federal Web site that is established as a repository for such informed consent forms. The informed consent form must be published on the federal Web site after the trial is closed to recruitment, and no later than 60 days after the last study visit by any subject, as required by the protocol (Federal Policy for the Protection of Human Subjects, 2017).

The intent of this consent form repository is to improve the consenting processes for subjects, and putatively to allow information mining of the corpus for a variety of purposes, including efforts such as this work. Unfortunately, as of October 7, 2017, the Trump administration proposed suspending implementation of the 2017 Common Rule until January of 2019, pending further review by the Office of Management and Budget.

7.4 Limitations

The work to date contains a number of limitations that include both shortcomings in the model expressed by the ontology and the evaluation. First, the ontology does not contain a formal framework for legal theory. It is unclear that it is important to have this framework at this time, but as the work is extended and strengthened it will become important. Second, due to resource limitations, it was necessary to use people who had been involved in the process of developing the original concept maps during the evaluation of the ontology. Consequently, unconscious bias may have been introduced. This limitation impacts the FOCA evaluation, as described in Section 6.4, and the sentence construction evaluation as described in Section 6.5. It also impacts the Burton-Jones analysis of the CRO as the sentence generation analysis is used as input in the calculation of the Burton-Jones *Accuracy* statistic. It does not, however, impact the Burton-Jones analysis comparing CRO to other regulatory ontologies, as described in Section 0, because the *Accuracy* statistic was not used in that comparison.

Third, research study participants and legal authorities were not consulted during the development of the concept maps, and the work could be strengthened by their participation in future work. Fourth, while the CRO has been validated by a variety of

methods including the use of competency questions and sentence construct validity, both development and evaluation were limited by the lack of availability of a gold standard corpus containing consent forms, research protocols, and the like. Competency questions have only been employed across a narrow range of conditions described by the ontology, namely those concerned with judging accuracy and completeness of the construction and presentation of informed consent forms and templates. Since the CRO ontology contains representations of the IRB processes and decision artifacts involved in the process of reviewing research proposals, suitable validation and competency questions for these processes must still be addressed. This will require resources to conduct a multi-site study of IRB methods, and is beyond the scope of resources available at the present time. However, it goes to the heart of the problems attempted to be addressed by this proposal, namely decision making in the context of IRB proposal review, and the release of biospecimens and data sharing.

Currently, the CRO does not incorporate the regulatory aspects of HIPAA that pertain to regulated human subjects research. Similarly, the regulatory statutes from the FDA are not included. Ultimately it will be necessary to include these into the CRO or develop a separate but substantively aligned ontology to represent these concepts.

Finally, I did not explore in depth the Health Level Seven (HL7) transactional models involving information transfer, security, and privacy. There may be much that can extend the CRO contained in that material.

7.5 Conclusions

In this dissertation I have demonstrated that there are serious limitations in the existing corpus of biomedical research ontologies regarding their representation of concepts and processes derived from the Common Rule. None of the existing models and ontology appear to have systematically derived these models from a detailed examination of law and regulatory statutes. I have further shown that the BFO and OBO Foundry has shortcomings in representing legal constructs. Further, what semantic constructs there are tend to be scattered between different ontologies in the OBO Foundry and do not appear to be well aligned with any particular top-level legal theory or framework. The realism basis of BFO makes concepts derived from the Common Rule seem awkward to represent, and may be off-putting and a barrier to potential users of the ontology. Finally, I have demonstrated that while it appears there are no merged approaches to life-cycle-based ontology development that are coupled with strong evaluation, it is readily possible to construct such approaches. A major outcome of this work is a new, hybrid method intended to assure high-quality ontology. The method consists of strong lifecycle approaches for planning, knowledge elicitation, detailed ontology construction steps, and strong, continuous evaluation.

7.6 Future work

Four areas are identified for future work. The first is a simple dissemination of the ontology. Areas for scientific exploration include (a) incorporation of legal theory into the ontology; (b) adding additional knowledge regarding decision making, specimen and

data sharing; and (c) determination of risk. A related question is how effectively the ontology can be used to parse informed consent forms, protocol documents, and other text-based research artifacts using NLP and other information extraction and retrieval methods.

From a purely administrative and technical perspective, immediate plans include registering the CRO with the OBO Foundry so that the work can be disseminated and utilized by others. The CRO was developed along the principles required for an ontology to be included in that library. This requires external peer review by the OBO Foundry members. During the process, I hope to add additional ontological entailments derived from Concept Map Three, along with abstract components of the regulatory process derived from Concept Map One.

As noted earlier, there is a lack of legal theory expressed in the current OBO Foundry ontologies. I hope to begin to work with OBO consortium members who are interested in extending legal theory into the OBO Foundry ontologies. The CRO has the potential to grow to a much larger ontology as various components of the legal space involved with human subjects research are modeled. The HHS website contains a substantial number of guidance documents for legal and regulatory interpretation. For example, a recently published document gives recommendations on the interpretation of broad consent (hhs.gov, 2017). The document contains, among other things, who is bound by a person's refusal to give broad consent, the specificity of the description of future uses of specimens and data, and many other items of interest to this work. I hope to review this material and begin the add it to the CRO.

One of the early motivating factors for this work was the desire of individuals associated with the University of Michigan IRB to develop a system based on the Common Rule that could recognize violations of U.S. regulation and institutional policy for both new proposals and those undergoing continuing review. It was also hoped to have a knowledge base that supported determining type and magnitude of risk based on the research and clinical procedures involved. The incorporation of an elementary risk model into this ontology gives this work the opportunity to be extended in this direction. This would be a substantial effort and would involve building knowledge bases about clinical and procedural risk.

Finally, the ultimate goal of this work is to develop a framework to allow computer-based reasoning about specimen and data release. The concepts involved are closely aligned with traditional notions of authentication and authorization but transcend them due to multiple constraints imposed by underlying requirements for trust, privacy, and the need for regulatory metadata.

Ultimately, to address these final two questions it will be necessary to systematically retrieve information from both discrete electronic sources and documents. This will require exploring how the CRO can effectively support such information retrieval and will provide ultimate guidance in both future development and validation.

References

Ahmed, F. E. (2011). Biobanking perspective on challenges in sample handling, collection, processing, storage, analysis and retrieval for genomics, transcriptomics and proteomics data. *Analytical Methods*, *3*(5), 1029–1038. https://doi.org/10.1039/C0AY00544D

Albert, D., & Steiner, C. (2005). Representing Domain Knowledge by Concept Maps: How to Validate Them? In *T. Okamoto, D. Albert, T. Honda & F. W. Hesse (Eds.), The 2nd Joint Workshop of Cognition and Learning Through Media-Communication for Advanced e-Learning (JWCL), September 28-30, 2005, Tokyo, Japan.* (pp. 169–174). Retrieved from http://telearn.archives-ouvertes.fr/hal-00190401

Amith, M. F., He, Z., Bian, J., Lossio-Ventura, J. A., & Tao, C. (2017). Assessing the Practice of Ontology Evaluation for Biomedical Ontologies in BioPortal: Gaps and Opportunities. *Journal of Biomedical Informatics (Submitted)*.

Amith, M., & Tao, C. (2015). A web application towards semiotic-based evaluation of biomedical ontologies (Vol. 1428). Presented at the CEUR Workshop Proceedings.

Amith, Muhammad, Manion, F., Harris, M., Zhang, Y., Xu, H., & Tao, C. (2017). Hootation: Expressing Biomedical Ontologies in Natural Language. Presented at the Medinfo 2017, Hangzhou, China.

Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with Basic Formal Ontology*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Bandeira, J., Bittencourt, I. I., Espinheira, P., & Isotani, S. (2016). FOCA: A Methodology for Ontology Evaluation. *arXiv:1612.03353 [cs.AI]*, 1–24.

Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M. H., Bug, B., Chibucos, M. C., … Zheng, J. (2016). The Ontology for Biomedical Investigations. *PLOS ONE*, *11*(4), e0154556. https://doi.org/10.1371/journal.pone.0154556

Becich, M. J. (2007). Lessons Learned from the Shared Pathology Informatics Network (SPIN): A Scalable Network for Translational Research and Public Health. *Journal of the American Medical Informatics Association*, *14*(4), 534–535. https://doi.org/10.1197/jamia.M2477

Beck, J. R., & caBIG Strategic Planning Workspace. (2007). The Cancer Biomedical Informatics Grid(TM): Infrastructure and applications for a worldwide research community. *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems: 2007*, 330–334.

Ben Abacha, A., Dos Reis, C. J., Mrabet, Y., Pruski, C., & Da Silveira, M. (2016). Towards natural language question generation for the validation of ontologies and mappings. *Journal of Biomedical Semantics*, *7*, 48. https://doi.org/10.1186/s13326-016-0089-6

Bezerra, C., Freitas, F., & Santana, F. (2013). Evaluating Ontologies with Competency Questions (pp. 284–285). Presented at the 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), IEEE. https://doi.org/10.1109/WI-IAT.2013.199

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, *32*(suppl 1), D267–D270. https://doi.org/10.1093/nar/gkh061

Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *Unified Modeling Language User Guide, The, 2nd Edition* (2nd ed.). Addison-Wesley Professional. Part of the Addison-Wesley Object Technology Series series. Retrieved from http://www.informit.com/store/unified-modeling-language-user-guide-9780321267979

Boris Motik, Peter F. Patel-Schneider, & Bijan Parsia. (2012, December 11). OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition) [W3C Technical Report]. Retrieved October 4, 2017, from https://www.w3.org/TR/owl2-syntax/

Brank, J., Grobelnik, M., & Mladenić, D. (2005). A survey of ontology evaluation techniques. In *In In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005*.

Brinkman, R. R., Courtot, M., Derom, D., Fostel, J. M., He, Y., Lord, P., … Zheng, J. (2010). Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*, *1*(Suppl 1), S7. https://doi.org/10.1186/2041-1480-1-S1-S7

Brochhausen, M., Fransson, M. N., Kanaskar, N. V., Eriksson, M., Merino-Martinez, R., Hall, R. A., … Litton, J.-E. (2013a). Developing a semantically rich ontology for the biobank-administration domain. *Journal of Biomedical Semantics*, *4*(1), 23. https://doi.org/10.1186/2041-1480-4-23

Brochhausen, M., Fransson, M. N., Kanaskar, N. V., Eriksson, M., Merino-Martinez, R., Hall, R. A., … Litton, J.-E. (2013b). Developing a semantically rich ontology for the biobank-administration domain. *Journal of Biomedical Semantics*, *4*, 23. https://doi.org/10.1186/2041-1480-4-23

Brochhausen, M., Zheng, J., Birtwell, D., Williams, H., Masci, A. M., Ellis, H. J., & Stoeckert, C. J. (2016). OBIB-a novel ontology for biobanking. *Journal of Biomedical Semantics*, *7*(1). https://doi.org/10.1186/s13326-016-0068-y

Burton-Jones, A., Storey, V. C., Sugumaran, V., & Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, *55*(1), 84–102. https://doi.org/10.1016/j.datak.2004.11.010

Carnot, M. J., Feltovich, P., Hoffman, R. R., Feltovich, J., & Novak, J. D. (2003). A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support. *Pensacola, FL*. Retrieved from https://carmenwiki.osu.edu/download/attachments/15599855/IHMC+Literature+Review+on+Concept+Mapping.pdf

Columbia University. (2014, August 25). Columbia University Adult WES Consent. Columbia University. Retrieved from http://www.cumc.columbia.edu/dept/irb/Templates.html

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & Knowledge Engineering*, *46*(1), 41–64. https://doi.org/10.1016/S0169-023X(02)00195-7

Courtot, M. (2017). *The Data Use Ontology*. Retrieved from http://www.obofoundry.org/ontology/duo.html

Courtot, M., Gibson, F., Lister, A. L., Malone, J., Schober, D., Brinkman, R. R., & Ruttenberg, A. (2009). MIREOT: the Minimum Information to Reference an External Ontology Term. *Nature Precedings*, (713). https://doi.org/10.1038/npre.2009.3574.1

Eppler, M. J. (2006). A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge

construction and sharing. *Information Visualization*, *5*(3), 202. https://doi.org/http://dx.doi.org/10.1057/palgrave.ivs.9500131

Fabien Gandon, Guus Schreiber, & Dave Beckett. (2014, February 25). RDF 1.1 XML Syntax [W3C Technical Report]. Retrieved October 4, 2017, from https://www.w3.org/TR/rdf-syntax-grammar/

Federal Policy for the Protection of Human Subjects, 45 C.F.R. § 46 (2017). Retrieved from https://www.federalregister.gov/documents/2017/01/19/2017-01058/federal-policy-for-the-protection-of-human-subjects

Federal Policy for the Protection of Human Subjects, Pub. L. No. 82 FR 7149 (2017). Retrieved from https://www.federalregister.gov/documents/2017/01/19/2017-01058/federal-policy-for-the-protection-of-human-subjects

Fellbaum, C. (2010). WordNet. In R. Poli, M. Healy, & A. Kameas (Eds.), *Theory and Applications of Ontology: Computer Applications* (pp. 231–243). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/978-90-481-8847-5_10

Fernández, M., Gómez-Pérez, A., & Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In *AAAI Technical Report SS-97-06*. Retrieved from http://oa.upm.es/5484/1/METHONTOLOGY_.pdf

Ferrari, S., & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, *31*(7), 799–815. https://doi.org/10.1080/0266476042000214501

Fleurence, R. L., Curtis, L. H., Califf, R. M., Platt, R., Selby, J. V., & Brown, J. S. (2014). Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*, *21*(4), 578–582. https://doi.org/10.1136/amiajnl-2014-002747

Foster, I. (2005). Globus Toolkit Version 4: Software for Service-Oriented Systems. *Lecture Notes in Computer Science IFIP International Conference on Network and Parallel Computing*, *3779*, 2–13.

Fransson, M. N., Rial-Sebbag, E., Brochhausen, M., & Litton, J.-E. (2014). Toward a common language for biobanking. *European Journal of Human Genetics*. https://doi.org/10.1038/ejhg.2014.45

Gangemi, A., Catenacci, C., Ciaramita, M., & Lehmann, J. (2006). Modelling Ontology Evaluation and Validation. In Y. Sure & J. Domingue (Eds.), *The Semantic Web: Research and Applications* (pp. 140–154). Springer Berlin Heidelberg. Retrieved from http://link.springer.com.proxy.lib.umich.edu/chapter/10.1007/11762256_13

Gennari, J. H., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., … Tu, S. W. (2003). The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies*, *58*(1), 89–123. https://doi.org/10.1016/S1071-5819(02)00127-1

Glimm, B., Horrocks, I., Motik, B., Stoilos, G., & Wang, Z. (2014). HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning*, *53*(3), 245–269. https://doi.org/10.1007/s10817-014-9305-1

Gómez-Pérez, A. (2001). Evaluation of ontologies. *International Journal of Intelligent Systems*, *16*(3), 391–409. https://doi.org/10.1002/1098-111X(200103)16:3<391::AID-INT1014>3.0.CO;2-2

Gómez-Pérez, A. (2004). Ontology Evaluation. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (First edition, pp. 251–274). Springer Berlin.

Grando, A., & Schwab, R. (2013). Building and evaluating an ontology-based tool for reasoning about consent permission. *AMIA Annual Symposium Proceedings*, *2013*, 514–523.

Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial Cognition and Computation*, *4*(1), 69–104.

Grossman, C., Powers, B., & McGinnis, J. M. (2011). *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary*. Institute of Medicine. Retrieved from http://books.google.com/books?hl=en&lr=&id=cHrWk6kEh1EC&oi=fnd&pg=PR1&dq=%22Series%22+%22Grossmann,+Brian+Powers,+and+J.+Michael%22+%22and%22+&ots=jbsxerQ4dJ&sig=MQh8MpQ_zuUHc7PMHKoknU60L7Y

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220. https://doi.org/10.1006/knac.1993.1008

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, *43*(5), 907–928. https://doi.org/10.1006/ijhc.1995.1081

Health Level Seven International. (2014). *HL7 Standards Product Brief - HL7 Version 3 Standard: Security and Privacy Ontology, Release 1*. Retrieved from http://www.hl7.org/implement/standards/product_brief.cfm?product_id=348

Hewison, J., & Haines, A. (2006). Overcoming barriers to recruitment in health research. *BMJ*, *333*(7562), 300–302. https://doi.org/10.1136/bmj.333.7562.300

hhs.gov. (2017, August 23). Attachment C - Recommendations for Broad Consent Guidance [Text]. Retrieved October 11, 2017, from https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-august-2-2017/index.html

HL7 Security Work Group. (2014, May 30). HL7 Version 3 Standard: Security and Privacy Ontology, Release 1. HL7. Retrieved from http://www.hl7.org/implement/standards/product_brief.cfm?product_id=348

Hlomani, H., & Stacey, D. (2014). Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web – Interoperability, Usability, Applicability an IOS Press Journal*.

Institute of Medicine. (2007). *The Learning Healthcare System: Workshop Summary*. https://doi.org/10.17226/11903

Johns Hopkins Medicine IRB. (2013, October). Human Pluripotent Stem Cell Research consent form template. Retrieved March 25, 2015, from http://www.hopkinsmedicine.org/institutional_review_board/forms/cirb_pediatric_cf_template.docx

Junghans, C., Feder, G., Hemingway, H., Timmis, A., & Jones, M. (2005). Recruiting patients to medical research: double blind randomised trial of "opt-in" versus "opt-out" strategies. *BMJ*, *331*(7522), 940. https://doi.org/10.1136/bmj.38583.625613.AE

Kaufman, D., Bollinger, J., Dvoskin, R., & Scott, J. (2012). Preferences for opt-in and opt-out enrollment and consent models in biobank research: a national survey of Veterans Administration patients. *Genetics in Medicine*, *14*(9), 787–794. https://doi.org/10.1038/gim.2012.45

Kho, A. N., Pacheco, J. A., Peissig, P. L., Rasmussen, L., Newton, K. M., Weston, N., … Denny, J. C. (2011). Electronic Medical Records for Genetic Research: Results of the eMERGE Consortium. *Science Translational Medicine*, *3*(79), 79re1. https://doi.org/10.1126/scitranslmed.3001807

Langella, S., Oster, S., Hastings, S., Siebenlist, F., Phillips, J., Ervin, D., … Saltz, J. (2007). The Cancer Biomedical Informatics Grid (caBIG™) Security Infrastructure. *AMIA Annual Symposium Proceedings*, *2007*, 433–437.

Lin, Y., Harris, M. R., Manion, F. J., Eisenhauer, E., Zhao, B., Shi, W., … He, Y. (2014). Development of a BFO-based Informed Consent Ontology (ICO). In *Proceedings of the 5th International Conference on Biomedical Ontologies (ICBO)* (Vol. 1327, pp. 84–86). Houston, Texas, USA. Retrieved from http://ceur-ws.org/Vol-1327/icbo2014_paper_54.pdf

Lin, Y., Zheng, J., & He, Y. (2016). VICO: Ontology-based representation and integrative analysis of Vaccination Informed Consent forms. *Journal of Biomedical Semantics*, *7*, 20. https://doi.org/10.1186/s13326-016-0062-4

Manion, F. J., He, Y., Eisenhauer, E., Lin, Y., Karnovsky, A., & Harris, M. R. (2014). Towards a Common Semantic Representation of Informed Consent for Biobank Specimens. In *Starting an OBI-based Biobank Ontology Workshop (OBIB 2014)* (Vol. 1309, pp. 61–63). Houston, Texas, USA: CEUR Workshop Proceedings. Retrieved from http://ceur-ws.org/Vol-1309/paper7.pdf

Manion, F., Robbins, R., Weems, W., & Crowley, R. (2009). Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study. *BMC Medical Informatics and Decision Making*, *9*(1), 31.

McCarty, C., Chisholm, R., Chute, C., Kullo, I., Jarvik, G., Larson, E., … Team, the eMERGE. (2011). The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Medical Genomics*, *4*(1), 13. https://doi.org/doi:10.1186/1755-8794-4-13

Menikoff, J., Kaneshiro, J., & Pritchard, I. (2017). The Common Rule, Updated. *New England Journal of Medicine*, *376*(7), 613–615. https://doi.org/10.1056/NEJMp1700736

Merino-Martinez, R., Norlin, L., van Enckevort, D., Anton, G., Schuffenhauer, S., Silander, K., … Litton, J.-E. (2016). Toward Global Biobank Integration by Implementation of the Minimum Information About BIobank Data Sharing

(MIABIS 2.0 Core). *Biopreservation and Biobanking*, *14*(4), 298–306. https://doi.org/10.1089/bio.2015.0070

Musen, M. A., Noy, N. F., Shah, N. H., Whetzel, P. L., Chute, C. G., Story, M.-A., & Smith, B. (2012). The National Center for Biomedical Ontology. *Journal of the American Medical Informatics Association*, *19*(2), 190–195. https://doi.org/10.1136/amiajnl-2011-000523

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report* (Text). Department of Health, Education, and Welfare. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

NCBO BioPortal. (2017). NCBO BioPortal. Retrieved June 21, 2017, from http://bioportal.bioontology.org/

Neuhaus, F. M., Vizedom, A., Baclawski, K., Bennet, M., Dean, M., Denny, M., … Yim, P. (2013). Towards Ontology Evaluation across the Life Cycle. *Applied Ontology*. Retrieved from https://www.nist.gov/publications/towards-ontology-evaluation-across-life-cycle

Norlin, L., Fransson, M. N., Eriksson, M., Merino-Martinez, R., Anderberg, M., Kurtovic, S., & Litton, J.-E. (2012). A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS. *Biopreservation and Biobanking*, *10*(4), 343–348. https://doi.org/10.1089/bio.2012.0003

Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology* (Technical Report No. Stanford Knowledge

Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880,). Retrieved from http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html

Nuremberg Military Tribunal. (1949). Nuremberg Code. *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10*, *2*, 181–182.

OASIS eXtensible Access Control Markup Language (XACML) Technical Committee. (2013, January 22). eXtensible Access Control Markup Language (XACML) Version 3.0. Retrieved March 23, 2015, from http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html

OASIS Security Services Technical Committee. (2005, March 15). Authentication Context for the OASIS Security Assertion Markup Language (SAML) V2.0. OASIS. Retrieved from http://docs.oasis-open.org/security/saml/v2.0/

OBO Foundry. (2017a). *IAO: information artifact ontology*. Web Ontology Language, IAO. Retrieved from https://github.com/information-artifact-ontology/IAO (Original work published 2015)

OBO Foundry. (2017b). *obo-relations: RO is an ontology of relations for use with biological ontologies*. Web Ontology Language, oborel. Retrieved from https://github.com/oborel/obo-relations (Original work published 2015)

Obrst, L. (2010). Ontological Architectures. In R. Poli, M. Healy, & A. Kameas (Eds.), *Theory and Applications of Ontology: Computer Applications* (pp. 27–66).

Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/978-90-481-8847-5_2

Obrst, L., Ceusters, W., Mani, I., Ray, S., & Smith, B. (2007). The Evaluation of Ontologies. In C. J. O. Baker & K.-H. Cheung (Eds.), *Semantic Web* (pp. 139–158). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-0-387-48438-9_8

Öhgren, A., & Sandkuhl, K. (2005). Towards a Methodology for Ontology Development in Small and Medium-Sized Enterprises. In *IADIS Conference on Applied Computing, Algarve, Portugal, February 2005*. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-4936

Ontobee. (2017). Ontobee. Retrieved June 8, 2017, from http://www.ontobee.org/

OWL 2 Primer. (2012, December 11). OWL 2 Web Ontology Language Primer (Second Edition). Retrieved October 15, 2017, from https://www.w3.org/TR/2012/REC-owl2-primer-20121211/#OWL_2_Profiles

Pathak, J., Wang, J., Kashyap, S., Basford, M., Li, R., Masys, D. R., & Chute, C. G. (2011). Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. *Journal of the American Medical Informatics Association*, *18*(4), 376–386. https://doi.org/10.1136/amiajnl-2010-000061

Pinto, H. S., & Martins, J. P. (2004). Ontologies: How can They be Built? *Knowledge and Information Systems*, *6*(4), 441–464. https://doi.org/10.1007/s10115-003-0138-1

Regidor, E. (2004). The use of personal data from medical records and biological materials: ethical perspectives and the basis for legal restrictions in health research. *Social Science & Medicine*, *59*(9), 1975–1984. https://doi.org/10.1016/j.socscimed.2004.02.032

Ren, Y., Parvizi, A., Mellish, C., Pan, J. Z., Van Deemter, K., & Stevens, R. (2014). Towards competency question-driven ontology authoring. In *The Semantic Web: Trends and Challenges* (pp. 752–767). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-07443-6_50

Sanya, I. O., & Shehab, E. M. (2015). A framework for developing engineering design ontologies within the aerospace industry. *International Journal of Production Research*, *53*(8), 2383–2409. https://doi.org/10.1080/00207543.2014.965352

Schilsky, R. L., Michels, D. L., Kearbey, A. H., Yu, P. P., & Hudis, C. A. (2014). Building a Rapid Learning Health Care System for Oncology: The Regulatory Framework of CancerLinQ. *Journal of Clinical Oncology*, *32*(22), 2373–2379. https://doi.org/10.1200/JCO.2014.56.2124

Sim, I., Carini, S., Tu, S. W., Detwiler, L. T., Brinkley, J., Mollah, S. A., … Huser, V. (2012). Ontology-Based Federated Data Access to Human Studies Information. *AMIA Annual Symposium Proceedings*, *2012*, 856–865.

Sim, I., Tu, S. W., Carini, S., Lehmann, H. P., Pollock, B. H., Peleg, M., & Wittkowski, K. M. (2014). The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research. *Journal of Biomedical Informatics*, *52*(Supplement C), 78–91. https://doi.org/10.1016/j.jbi.2013.11.002

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., … Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, *25*(11), 1251–1255. https://doi.org/10.1038/nbt1346

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., … Rosse, C. (2005). Relations in biomedical ontologies. *Genome Biology*, *6*(5), R46. https://doi.org/10.1186/gb-2005-6-5-r46

Soysal, E., Wang, J., Jiang, M., Wu, Y., Pakhomov, S., Liu, H., & Xu, H. (2017). CLAMP – a toolkit for efficiently building customized clinical natural language processing pipelines. *Journal of the American Medical Informatics Association*. https://doi.org/10.1093/jamia/ocx132

Stamper, R., Liu, K., Hafkamp, M., & Ades, Y. (2000). Understanding the roles of signs and norms in organizations - a semiotic approach to information systems design. *Behaviour & Information Technology*, *19*(1), 15–27. https://doi.org/10.1080/014492900118768

Stanford IRB. (2015, March 24). Stanford IRB - Forms & Templates. Retrieved March 25, 2015, from http://humansubjects.stanford.edu/new/resources/forms_templates/

Uschold, M., & King, M. (1995). Towards a Methodology for Building Ontologies (pp. 1–13). Presented at the Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Quebec, Canada.

Vrandečić, D. (2009). Ontology Evaluation. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 293–313). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-92673-3_13

Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., & Musen, M. A. (2011). BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research*, *39*(Web Server issue), W541–W545. https://doi.org/10.1093/nar/gkr469

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Xiang, Z., Courtot, M., Brinkman, R. R., Ruttenberg, A., & He, Y. (2010). OntoFox: web-based support for ontology reuse. *BMC Research Notes*, *3*, 175. https://doi.org/10.1186/1756-0500-3-175

Xiang, Z., Mungall, C., Ruttenberg, A., & He, Y. (2011). Ontobee: A Linked Data Server and Browser for Ontology Terms. In *Proceedings of the 2nd International Conference on Biomedical Ontologies (ICBO)* (pp. 279–281). Buffalo, NY, USA. Retrieved from http://ceur-ws.org/Vol-833/paper48.pdf

Xiang, Z., Zheng, J., Lin, Y., & He, Y. (2015). Ontorat: automatic generation of new ontology terms, annotations, and axioms based on ontology design patterns. *Journal of Biomedical Semantics*, *6*(1), 4. https://doi.org/10.1186/2041-1480-6-4

XMind Ltd. (2016). XMind Version 7.5. Retrieved September 30, 2017, from http://www.xmind.net

Yu, P. P. (2017). Perspectives on the Implementation of an Ethical Rapid Learning Health System. *Journal of Oncology Practice*, *13*(3), 151–153. https://doi.org/10.1200/JOP.2016.019828

Zhu, X., Fan, J.-W., Baorto, D. M., Weng, C., & Cimino, J. J. (2009). A Review of Auditing Methods Applied to the Content of Controlled Biomedical Terminologies. *Journal of Biomedical Informatics*, *42*(3), 413–425. https://doi.org/10.1016/j.jbi.2009.03.003

Appendix A. University of Michigan Informed Consent Checklist

Below is the informed consent checklist in its original format as taken from the University of Michigan 'MEDIRB' website.

Informed Consent Checklist

DHHS Office of Human Research Protections 45 §46.116 and 21 §50.25 Food and Drug Administration (FDA)

Studies that use an informed consent template other than those provided by their IRB (or CIRB) must add the IRB required header/footer (http://www.med.umich.edu/irbmed/ict/eResearch-IC-Other-Header.doc). Also, footnote/comment where in a proposed documents each of the elements is being met. This assures all elements are included and speeds the review process.

Basic and Additional Elements (required if appropriate to the study)

	A statement that the study involves research
	An explanation of the purposes of the research
	The expected duration of the subject's participation
	A description of the procedures to be followed
	Identification of any procedures which are experimental
	A description of any reasonably foreseeable risks or discomforts to the subject
	A description of any benefits to the subject or to others which may reasonably be expected from the research
	A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject
	A statement describing the extent, if any, to which confidentiality of records

	identifying the subject will be maintained. For studies under FDA oversight it must also note the possibility that the Food and Drug Administration may inspect the records.
	For research involving more than minimal risk, an explanation as to whether any compensation, and an explanation as to whether any medical treatments are available, if injury occurs and, if so, what they consist of, or where further information may be obtained
	An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of a research-related injury to the subject
	A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits, to which the subject is otherwise entitled
Additional elements, as appropriate	
	When seeking informed consent for applicable clinical trials, as defined in 42 U.S.C. 282(j)(1)(A), the following statement shall be provided to each clinical trial subject in informed consent documents and processes. This will notify the clinical trial subject that clinical trial information has been or will be submitted for inclusion in the clinical trial registry databank under paragraph (j) of section 402 of the Public Health Service Act. The statement is: "A description of this clinical trial will be available on *http://www.ClinicalTrials.gov,* as required by U.S. Law. This Web site will not include information that can identify you. At most, the Web site will include a summary of the results. You can search this Web site at any time."
	A statement that the particular treatment or procedure may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant), which are currently unforeseeable
	Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent
	Any additional costs to the subject that may result from participation in the research
	The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject
	A statement that significant new findings developed during the course of the research, which may relate to the subject's willingness to continue participation, will be provided to the subject

	The approximate number of subjects involved in the study

If the study falls under the FDA's oversight the consent must be dated when signed.

If neonates, children, children who are wards of the state, pregnant women, or fetuses are to be enrolled in the study additional regulations apply. For studies involving children as subjects the IRB is required to determine if assent of the subjects must be obtained in addition to parent's permission before research can proceed.

Appendix B. Concept Map One

Concept Map One appears on the next page.

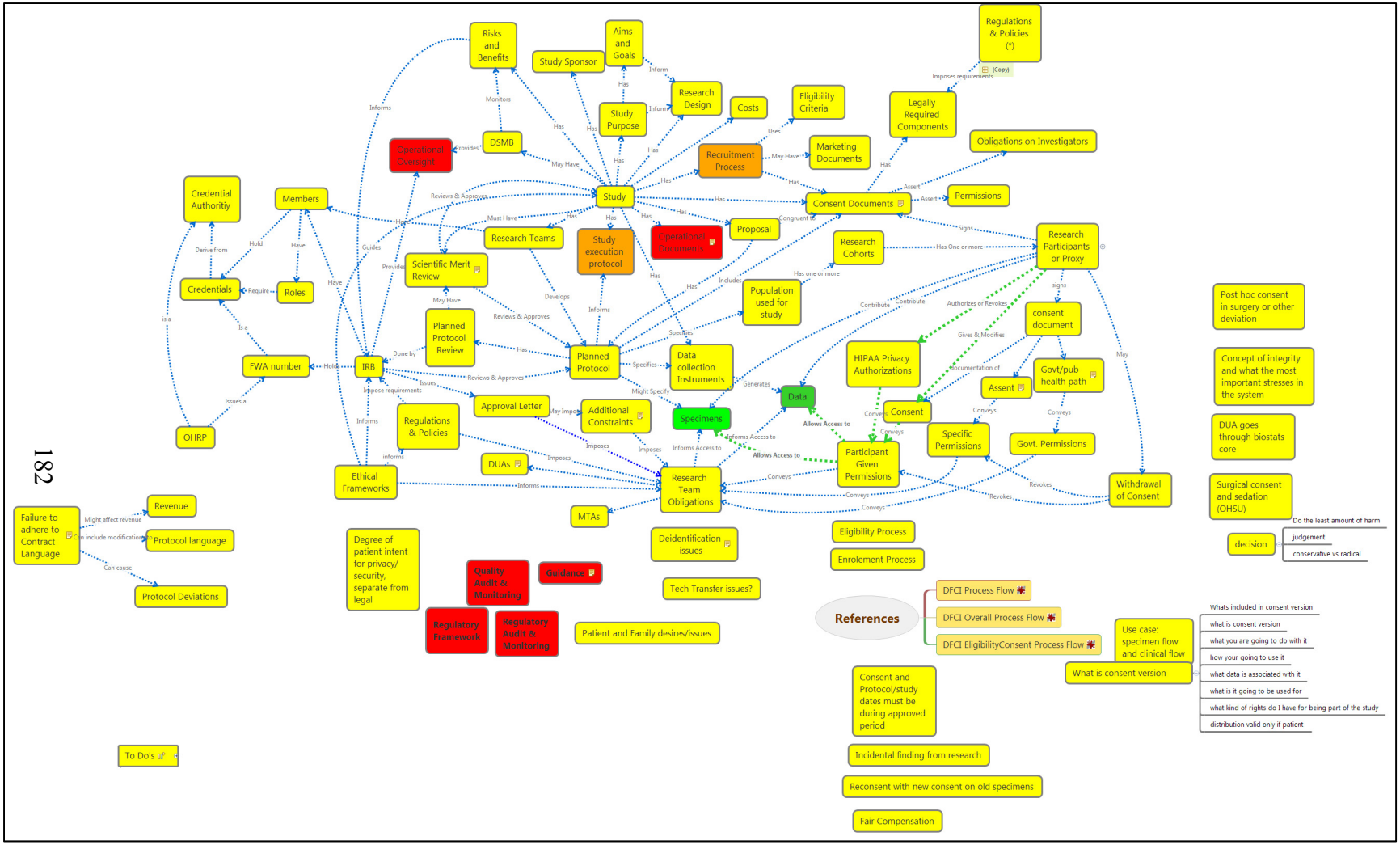


Figure 23. Concept Map One.

Appendix C. Table of Initial Classes, Terms, and Other Artifacts

Table 48

Table of Initial Designations of Text and Concept from the Common Rule

Section Text	Type	Term	Value
45 CFR 46.101	Authority:	[56 FR 28012, 28022, June 18, 1991; 56 FR 29756, June 28, 1991, as amended at70 FR 36328, June 23, 2005]	
45 CFR 46.101(a)(1-2)	Class	Funding Body	Funding Body
45 CFR 46.101(a)(1-2)	Class	Government Body	Government Body
45 CFR 46.101(a)(1-2)	Class	Regulatory Body	Regulatory Body
45 CFR 46.101(a)(1-2)	Axioms:	Applies if involves(human subjects) AND isResearch({proposal, study, etc.}) AND (conductedBy(agency) OR fundedBy(agency) OR conductedBy(federal department) OR fundedBy(federal department) OR regulatedBy(agency) OR regulatedBy(federal department))	
45 CFR 46.101(a)(1-2)	Properties	involvesHumanSubjects	
45 CFR 46.101(a)(1-2)	Properties	regulatedBy	
45 CFR 46.101(a)(1-2)	Axioms:	isExempt	
45 CFR 46.101(a)(1-2)	Properties	fundedBy	
45 CFR 46.101(a)(1-2)	Properties	applies To	
45 CFR 46.101(a)(1-2), 46.101(b)(5)	Properties	conductedBy	

Section Text	Type	Term	Value
45 CFR 46.101(a)(1-2), 46.101(e)	Class	Federal Agency = {NIH, NSF, NCI, FDA, USDA, etc.}	
45 CFR 46.101(a)(1-2), 46.102(c)	Class	Research	
45 CFR 46.101(a)(1-2), 46.102(d)	Properties	isResearch	
45 CFR 46.101(a)(1-2), 46.102(e)	Class	Federal department	
45 CFR 46.101(a)(1-2), 46.102(f)	Class	Human Subjects	
45 CFR 46.101(b)	Axioms:	exemptResearch(conductedIn(ANY(established educational setting, commonly accepted educational setting)) AND involves(normal educational practices))	
45 CFR 46.101(b)(1)	Class	commonly accepted educational settings	
45 CFR 46.101(b)(1)	Class	established educational settings	
45 CFR 46.101(b)(1)	Class	classroom management methods	
45 CFR 46.101(b)(1)	Class	Comparison of curricula	
45 CFR 46.101(b)(1)	Class	effectiveness of curricula	
45 CFR 46.101(b)(1)	Class	instructional techniques	
45 CFR 46.101(b)(1)	Class	normal educational practice	
45 CFR 46.101(b)(1)	Class	regular education instructional strategies	
45 CFR 46.101(b)(1)	Class	special education instructional strategies	
45 CFR 46.101(b)(1)	Properties	comparisonOf	

Section Text	Type	Term	Value
45 CFR 46.101(b)(1)	Axioms:	comparisonOf(ANY{instructional techniques, curricula, classroom management methods},ANY{instructional techniques, curricula, classroom management methods}) isA normal educational practice	
45 CFR 46.101(b)(1)	Properties	effectivenessResearch	
45 CFR 46.101(b)(1)	Axioms:	effectivenessResearch(ANY{instructional techniques, curricula, classroom management methods},ANY{instructional techniques, curricula, classroom management methods}) isA normal educational practice	
45 CFR 46.101(b)(1)	Axioms:	regular instructional strategy isA normal educational practice	
45 CFR 46.101(b)(1)	Axioms:	special education instructional strategy isA normal educational practice	
45 CFR 46.101(b)(2-3)	Class	cognitive educational tests	
45 CFR 46.101(b)(2-3)	Class	diagnostic educational tests	
45 CFR 46.101(b)(2-3)	Class	educational achievement tests	
45 CFR 46.101(b)(2-3)	Class	educational aptitude tests	
45 CFR 46.101(b)(2-3)	Class	interview procedures	
45 CFR 46.101(b)(2-3)	Class	observation of public behavior	
45 CFR 46.101(b)(2-3)	Class	civil liability	
45 CFR 46.101(b)(2-3)	Class	criminal liability	
45 CFR 46.101(b)(2-3)	Class	employability	
45 CFR 46.101(b)(2-3)	Class	Financial	
45 CFR 46.101(b)(2-3)	Class	Reputational Risk	

Section Text	Type	Term	Value
45 CFR 46.101(b)(2-3)	Properties	hasRisk	
45 CFR 46.101(b)(2)	Axioms:	hasRisk(ANY(criminal liability,civil liability, financial, employability, reputational))	
45 CFR 46.101(b)(2)	Axioms:	notExempt(OR(subjectIndentifiable(),hasRisk()))	
45 CFR 46.101(b)(2)	Axioms:	subjectIdentifiable(ANY(containsPHI, containsIdentifierToPHI))	
45 CFR 46.101(b)(2), 46.101(b)(3)	Class	Research Method	survey procedures
45 CFR 46.101(b)(2), 46.101(b)(3)	Properties	containsIdentifierToPHI	
45 CFR 46.101(b)(2), 46.101(b)(3)	Properties	containsPHI	
45 CFR 46.101(b)(3)	Class	Research Participant	appointed public official
45 CFR 46.101(b)(3)	Class	Research Participant	candidate for public office
45 CFR 46.101(b)(3)	Class	Research Participant	elected public official
45 CFR 46.101(b)(3)	Properties	usesInstrument	
45 CFR 46.101(b)(3)	Properties	usesMethod	
45 CFR 46.101(b)(3)	Properties	recordsInformation	
45 CFR 46.101(b)(3)	Properties	confidentialityRequiredbyLaw	
45 CFR 46.101(b)(3)	Axioms:	hasRisk(ANY(criminal liability,civil liability, financial, employability, reputational))	
45 CFR 46.101(b)(3)	Axioms:	notExempt(OR(subjectIndentifiable(),hasRisk()))	
45 CFR 46.101(b)(3)	Axioms:	subjectIdentifiable(ANY(containsPHI, containsIdentifierToPHI))	
45 CFR 46.101(b)(4)	Class	Privacy Protections	dataset Privacy Type

Section Text	Type	Term	Value
45 CFR 46.101(b)(4)	Class	Research Materials	Identifiers linked to subject
45 CFR 46.101(b)(4)	Class	Research Materials	Identifiers not linked to subject
45 CFR 46.101(b)(4)	Class	Research Materials	Study of diagnostic specimens
45 CFR 46.101(b)(4)	Class	Research Materials	Study of documents
45 CFR 46.101(b)(4)	Class	Research Materials	Study of existing Data
45 CFR 46.101(b)(4)	Class	Research Materials	Study of existing records
45 CFR 46.101(b)(4)	Class	Research Materials	Study of pathological specimens
45 CFR 46.101(b)(4)	Equivalence Class:	Biobank == Biorepository == specimen collection OR biobank collection	
45 CFR 46.101(b)(4)	Terms	DatasetType={Full, limited, anonymized}	
45 CFR 46.101(b)(4)	Axioms:	exempt(any of {existingData, documents, records, pathological specimens, diagnostic specimens) AND (hasSource(publicly available) OR subject cannot be identified)	
45 CFR 46.101(b)(4)	Properties	isDeidentified	
45 CFR 46.101(b)(4)	Properties	publicalyAvailable	
45 CFR 46.101(b)(4)	Axioms:	subject cannot be identified = (subjectIDs = anonymized OR datasetPrivacyType = anonymized)	
45 CFR 46.101(b)(4)	Terms	SubjectIDs={Identifiable, coded (or linked?), anonymized}	
45 CFR 46.101(b)(5-6)	Properties	hasStudyType	
45 CFR 46.101(b)(5)	Class	Government Body	Federal Agency Or Department Head
45 CFR 46.101(b)(5)	Class	Research Focus	demonstration Project
45 CFR 46.101(b)(5)	Class	Research Focus	public benefit or service programs

Section Text	Type	Term	Value
45 CFR 46.101(b)(5)	Class	Research Focus	public benefit or service programs
45 CFR 46.101(b)(5)	Class	Research Focus	public benefit or service programs
45 CFR 46.101(b)(5)	Class	Research Focus	public benefit or service programs
45 CFR 46.101(b)(5)	Class	Research Focus	public benefit or service programs
45 CFR 46.101(b)(5)	Class	Research Project	research Project
45 CFR 46.101(b)(5)	Properties	approvalBy	
45 CFR 46.101(b)(5)	Axioms:	exempt((researchProject OR demonstrationProject) AND (conductedBy(federalAgencyOrDepartmentHead) OR approvalBy(federalAgencyOrDepartmentHead)) AND hasStudySubject(public benefit or service program))	
45 CFR 46.101(b)(5)	Equivalence Class:	public program == public benefit or service programs	
45 CFR 46.101(b)(6)	Class	Chemicals	agricultural chemical
45 CFR 46.101(b)(6)	Class	Chemicals	environmental contaminant
45 CFR 46.101(b)(6)	Class	Food	food ingredient
45 CFR 46.101(b)(6)	Class	Food	food with additives
45 CFR 46.101(b)(6)	Class	Food	food without additives
45 CFR 46.101(b)(6)	Class	Research Focus	Consumer Acceptance Studies
45 CFR 46.101(b)(6)	Class	Research Focus	Food Quality Evaluation
45 CFR 46.101(b)(6)	Class	Research Focus	Taste Evaluation
45 CFR 46.101(b)(6)	Properties	areConsumed	

Section Text	Type	Term	Value
45 CFR 46.101(b)(6)	Axioms:	exempt = hasStudyType SOME (taste study, food quality evaluation, consumer acceptance studies) AND ((areConsumed SOME food with additives AND foundSafeBy ONLY {Environmental Protection Agency, Food Safety and Inspection Service of USDA}	
45 CFR 46.101(b)(6)	Axioms:	exempt = hasStudyType SOME {taste study, food quality evaluation, consumer acceptance studies} AND ((areConsumed SOME food without additives AND isWholesome ONLY food) OR (NOT (areConsumed SOME food with additives)))	
45 CFR 46.101(c)	Class	Administrative Authority	Agency Head
45 CFR 46.101(c)	Class	Administrative Authority	Agency Head makes Final Determination if Policy applies
45 CFR 46.101(c)	Class	Administrative Authority	Department Head
45 CFR 46.101(c)	Class	Administrative Authority	Department Head makes Final Determination if Policy applies
45 CFR 46.101(c)	Class	Scope of Policy	Activity Covered By This Policy
45 CFR 46.101(c)	Axioms:	AgencyHead Has FinalJudgment	
45 CFR 46.101(c)	Axioms:	AgencyHead Retains FinalJudgment	
45 CFR 46.101(c)	Axioms:	DeparmentHead Retains FinalJudgment	
45 CFR 46.101(c)	Axioms:	DepartementHead Has FinalJudgment	
45 CFR 46.101(c)	Properties	Retains	
45 CFR 46.101(c)	Equivalence Class:	This Policy == 45 CFR 46	
45 CFR 46.101(d)	Properties	mayRequire	

Section Text	Type	Term	Value
45 CFR 46.101(d), 46.101(e)	Properties	compliance	
45 CFR 46.101(e-g)	Class	Protections Afforded Participant	Additional Human Subject Protection
45 CFR 46.101(e)	Class	Legal Authority	Pertinent Federal Laws
45 CFR 46.101(e)	Class	Legal Authority	Pertinent Federal Regulations
45 CFR 46.101(e)	Equivalence Class:	ThisPolicy == 45 CFR 46	
45 CFR 46.101(f)	Class	Legal Authority	Local Law
45 CFR 46.101(f)	Class	Legal Authority	Local Regulation
45 CFR 46.101(f)	Class	Legal Authority	State Law
45 CFR 46.101(f)	Class	Legal Authority	State Regulation
45 CFR 46.101(f), 46.201	Class	Legal Authority	Alaska Native Tribal Government
45 CFR 46.101(f), 46.201	Class	Legal Authority	American Indian Tribal Government
45 CFR 46.101(g)	Class	Legal Authority	Foreign Laws
45 CFR 46.101(g)	Class	Legal Authority	Foreign Regulations
45 CFR 46.101(h)	Properties	DeclarationOfHelsinki	
45 CFR 46.101(h)	Properties	Foreign Countries	
45 CFR 46.101(h)	Properties	WorldMedicalAssemblyDeclaration	
45 CFR 46.101(i)	Class	Administrative Authority	Office For Human Research Protections
45 CFR 46.101(i)	Class	Administrative Notices	Advance Notice
45 CFR 46.101(i)	Class	Administrative Notices	Federal Register

Section Text	Type	Term	Value
45 CFR 46.101(i)	Class	Communication	publish
45 CFR 46.101(i)	Class	Legal Authority	Statute
45 CFR 46.101(i)	Equivalence Class:	HHS == Department of Health and Human Services	
45 CFR 46.101(i), 46.201	Class	Administrative Authority	Department of Health and Human Services
45 CFR 46.102(a)	Equivalence Class:	Departement Head == federal department AND headOF (federal departement OR headOF federal agency)	
45 CFR 46.102(a)	Equivalence Class:	Department Head == hasDelegatedAuthority(SOME(federal department, federal agency) AND SOME (department employee, agency employee)	
45 CFR 46.102(a)	Equivalence Class:	Department Head == hasDelegatedAuthority(SOME(federal department, federal agency) AND SOME (department officer, agency officer))	
45 CFR 46.102(b)	Equivalence Class:	Institution == SOME(public entity, private entity, federal agency, state agency, other agency)	
45 CFR 46.102(c)	Class	Legal Authority	Authorized Individual
45 CFR 46.102(c)	Class	Legal Authority	Authorized Judicial Body
45 CFR 46.102(c)	Class	Legal Authority	Other Authorized Legal Body
45 CFR 46.102(c)	Class	Participant Screening	prospective subject
45 CFR 46.102(c)	Class	Research Method	Research procedures
45 CFR 46.102(c)	Properties	authorizedUnderApplicableLawToConsentOnBehalfOfSubject	

Section Text	Type	Term	Value
45 CFR 46.102(c)	Equivalence Class:	legally authorized representative == authorizedUnderApplicableLawToConsentOnBehalfOfSubject SOME (individual, judical body, other legal body)	
45 CFR 46.102(c)	Axioms:	Research Procedures SubClassOf Research	
45 CFR 46.102(c)	Axioms:	research subject participatesIn Research	
45 CFR 46.102(c)	Axioms:	research subject participatiesIn research procedures	
45 CFR 46.102(d)	Class	Research Purpose	contribute to generalizable knowledge
45 CFR 46.102(d)	Class	Research Purpose	develop generalizable knowledge
45 CFR 46.102(d)	Class	Research Type	evaluation
45 CFR 46.102(d)	Class	Research Type	research development
45 CFR 46.102(d)	Class	Research Type	systematic evaluation
45 CFR 46.102(d)	Class	Research Type	testing
45 CFR 46.102(d)	Properties	hasPurpose	
45 CFR 46.102(d)	Properties	hasResearchType	
45 CFR 46.102(d)	Restriction:	RealizedResearchPurpose	
45 CFR 46.102(d)	Restriction:	RealizedResearchType	
45 CFR 46.102(e)	Class	Research Subject to regulation	
45 CFR 46.102(e)	Axioms:	Federal Department hasSpecificRegulatoryAuthorityOver Research Subject to Regulation	
45 CFR 46.102(e)	Properties	hasGeneralRegulatoryAuthorityOver	

Section Text	Type	Term	Value
45 CFR 46.102(e)	Properties	hasSpecificRegulatoryAuthorityOver	
45 CFR 46.102(e)	Properties	hasSpecificRegulatoryAuthorityOver disjointwith hasGenralRegulatoryAuthorityOver	
45 CFR 46.102(f)	Class	Information About Participant	Identifiable Private Information
45 CFR 46.102(f)	Class	Investigator	Professional Investigator
45 CFR 46.102(f)	Class	Investigator	Student Investigator
45 CFR 46.102(f)	Class	Participant Type	Living Individual
45 CFR 46.102(f)	Class	Research Method	Manipulation Of Subject
45 CFR 46.102(f)	Class	Research Method	Manipulation Of Subject Environment
45 CFR 46.102(f)	Properties	conductsInteraction	
45 CFR 46.102(f)	Properties	conductsIntervention	
45 CFR 46.102(f)	Axioms:	HumanSubject==LivingIndividual AND obtainsData some Investigator	
45 CFR 46.102(f)	Axioms:	Intervention some PhysicalProcedures OR some ManipulationOfSubject OR some ManipulationOfSubjectEnvironment	
45 CFR 46.102(f)	Properties	isLiving	
45 CFR 46.102(f)	Properties	obtainsData	
45 CFR 46.102(f)	Properties	obtainsSpecimen	
45 CFR 46.102(g)	Class	Regulatory Body	Institutional Review Board
45 CFR 46.102(g)	Class	Regulatory Body	IRB
45 CFR 46.102(g)	Class	Regulatory Body	Privacy Board
45 CFR 46.102(g)	Equivalence Class:	IRB == Institutional Review Board	

Section Text	Type	Term	Value
45 CFR 46.102(g)	Equivalence Class:	IRB == Privacy Board	
45 CFR 46.102(h)	Class	Constraints On Research	Federal Policy Constraints
45 CFR 46.102(h)	Class	Constraints On Research	Institutional Policy Constraints
45 CFR 46.102(h)	Class	Constraints On Research	IRB Imposed Constraints
45 CFR 46.102(h)	Class	Research State	Approved Research
45 CFR 46.102(h)	Class	Research State	Reviewed Research
45 CFR 46.102(h)	Class	Research State	Unapproved Research
45 CFR 46.102(h)	Class	Research State	Unreviewed Research
45 CFR 46.102(h)	Restriction:	disjoint(Approved Research, Unapproved Research)	
45 CFR 46.102(h)	Restriction:	disjoint(Reviewed Research, Unreviewed Research)	
45 CFR 46.102(h)	Properties	hasConstraints	hasIRBImposedConstraints
45 CFR 46.102(h)	Properties	hasConstraints	hasFederalPolicyConstraints
45 CFR 46.102(h)	Properties	hasConstraints	hasInstitutionalPolicyConstraints
45 CFR 46.102(i)	Class	Risk Magnitude	Minimal Risk
45 CFR 46.102(i)	Class	Risk Magnitude	Ordinary Daily Life Risk
45 CFR 46.102(i)	Class	Risk Magnitude	Routine Physical Exam Risk
45 CFR 46.102(i)	Class	Risk Magnitude	Routine Psychological Exam Risk

Section Text	Type	Term	Value
45 CFR 46.102(i)	Class	Risk Magnitude	Routine Test Risk
45 CFR 46.102(i)	Properties	hasRiskProbability	
45 CFR 46.102(i)	Restrictio n:	need to express Minimal Risk as 'Research hasRiskMagnitude (low) AND Research hasRiskProbability (low)'	
45 CFR 46.102(i), 46.116(d)(1-4)	Properties	hasRiskMagnitude	
45 CFR 46.102(j)	Class	Administrative Notices	Certification by IRB of Research
45 CFR 46.102(j)	Class	Administrative Notices	Official Notification from IRB
45 CFR 46.102(j)	Class	Research Project	Research Activity
45 CFR 46.102(j)	Equivalen ce Class:	Certification == Approved Research AND Institution officiallyNotifies some(Department,Agency)	
45 CFR 46.102(j)	Properties	officiallyNotifies	range: Department or Agency Domain: Institution
45 CFR 46.102(j)	Equivalen ce Class:	Research project == Research Activity	
45 CFR 46.116	Class	Consent Media	Oral Consent
45 CFR 46.116	Class	Consent Media	Written Consent
45 CFR 46.116	Class	Ethical Behavior	minimize possibility of coercion
45 CFR 46.116	Class	Ethical Behavior	minimize possibility of undue influence
45 CFR 46.116	Class	Ethical Behavior	No Release from Negligence
45 CFR 46.116	Class	Exculpatory Language	No release of sponsor from liability

Section Text	Type	Term	Value
45 CFR 46.116	Class	Exculpatory Language	No appearance of waiver of legal rights
45 CFR 46.116	Class	Exculpatory Language	No Exculpatory Language
45 CFR 46.116	Class	Exculpatory Language	No release of institution from liability
45 CFR 46.116	Class	Exculpatory Language	No release of investigator from liability
45 CFR 46.116	Class	Exculpatory Language	No release of legal agents from liability
45 CFR 46.116	Class	Exculpatory Language	No Waiver of legal rights
45 CFR 46.116	Class	Human Language	Language Understood by Subject
45 CFR 46.116	Class	Investigator	Investigator
45 CFR 46.116	Class	Legal Agent Of Participant	Legally authorized representative of subject
45 CFR 46.116	Class	Legal Authority	Common Rule (this policy)
45 CFR 46.116	Class	Participant Type	Human Being
45 CFR 46.116	Class	Regulatory Constraint	Sufficient time to consider consent
45 CFR 46.116	Class	Research Focus	Scope of Consent
45 CFR 46.116	Class	Research Participant	subject in research
45 CFR 46.116	Equivalence Class:	45CFR46 == Common Rule	
45 CFR 46.116	Axioms:	Exculpatory Language = (some release of investigator from liability) OR (some release of institution from liability) OR (some release of legal agents from liability)	
45 CFR 46.116	Axioms:	Legally effective informed consent = (consent_from some legally authorized representative of subject) OR (consent_from some subject in research)	

Section Text	Type	Term	Value
45 CFR 46.116	Axioms:	Research permitted = (research involves some subject in research) AND investigator has_obtained legally effective informed consent)	
45 CFR 46.116	Axioms:	Scope of Consent == research AND Covered_by 45CFR46	
45 CFR 46.116	Axioms:	Subject in research == some Human Being	
45 CFR 46.116	Axioms:	Understood language = (consent language understood_by subject in research) OR (consent language understood by Legally authorized representative of subject)	
45 CFR 46.116, 46.205(b)(1-2)	Class	Legally Effective Informed Consent	Legally Effective Informed Consent
45 CFR 46.116(a)(2)	Class	Required Statement	Description of discomforts
45 CFR 46.116(a)(2)	Class	Required Statement	Description of risks
45 CFR 46.116(a)(3)	Class	Required Statement	Description of benefits to others
45 CFR 46.116(a)(3)	Class	Required Statement	Description of benefits to subject
45 CFR 46.116(a)(4)	Class	Required Statement	advantageous alternative treatments
45 CFR 46.116(a)(4)	Class	Required Statement	Disclosure of alternative treatments
45 CFR 46.116(a)(5)	Class	Required Statement	Confidentiality of records identifying subject
45 CFR 46.116(a)(6)	Class	Required Statement	Availability of Care for Injury
45 CFR 46.116(a)(6)	Class	Required Statement	Compensation for Participation
45 CFR 46.116(a)(6)	Class	Required Statement	Type of care for injury
45 CFR 46.116(a)(6)	Class	Required Statement	Where further information can be obtained
45 CFR 46.116(a)(7)	Class	Required Statement	Who to contact

Section Text	Type	Term	Value
45 CFR 46.116(a)(7)	Class	Required Statement	Who to contact
45 CFR 46.116(a)(7)	Class	Required Statement	Who to contact
45 CFR 46.116(a)(8)	Class	Required Statement	May discontinue participation at any time
45 CFR 46.116(a)(8)	Class	Required Statement	No loss of entitled benefits for discontinuing participation
45 CFR 46.116(a)(8)	Class	Required Statement	No loss of entitled benefits for refusal to participate
45 CFR 46.116(a)(8)	Class	Required Statement	No penalty for discontinuing participation
45 CFR 46.116(a)(8)	Class	Required Statement	No penalty for refusal to participate
45 CFR 46.116(a)(8)	Class	Required Statement	Participation is Voluntary
45 CFR 46.116(b)(1)	Class	Additional Statement of Consent	Research may involve unforeseeable risks to embryo or fetus
45 CFR 46.116(b)(1)	Class	Additional Statement of Consent	Research may involve unforeseeable risks to subject
45 CFR 46.116(b)(2)	Class	Additional Statement of Consent	Circumstances of possible termination by investigator
45 CFR 46.116(b)(3)	Class	Additional Statement of Consent	Additional resulting costs to subject for participation
45 CFR 46.116(b)(4)	Class	Additional Statement of Consent	Consequences of voluntarily withdrawing
45 CFR 46.116(b)(4)	Class	Additional Statement of Consent	Procedures for termination of participation
45 CFR 46.116(b)(5)	Class	Additional Statement of Consent	Significant finding relating to subjects willingness to participate will be communicated

Section Text	Type	Term	Value
45 CFR 46.116(b)(6)	Class	Number of Participants	Approximate Number of Subjects in Study
45 CFR 46.116(c)(1-2)	Class	Altered Consent	IRB Allows Altered Consent
45 CFR 46.116(c)(1-2)	Class	Research Focus	Changes To Payment Levels
45 CFR 46.116(c)(1-2)	Class	Research Focus	Changes To Payment Methods
45 CFR 46.116(c)(1-2)	Class	Research Focus	Changes To Public Benefi tProgram
45 CFR 46.116(c)(1-2)	Class	Research Focus	Changes To Public Service Program
45 CFR 46.116(c)(1-2)	Class	Research Focus	Procedure for Obtaining Benefits
45 CFR 46.116(c)(1-2)	Class	Research Focus	Procedure for Obtaining Services
45 CFR 46.116(c)(1-2)	Class	Research Focus	Public Benefit Program Research
45 CFR 46.116(c)(1-2)	Class	Research Focus	Public Service Program Research
45 CFR 46.116(c)(1-2)	Axioms:	AllowsAlteredConsent == PublicBenefitProgramResearch	
45 CFR 46.116(c)(1-2)	Properties	hasAdministrativeProcudure	
45 CFR 46.116(c)(1-2)	Properties	HasIRBApprovedConsentProcedure	
45 CFR 46.116(c)(1-2)	Properties	HasResearchFocus	
45 CFR 46.116(c)(1-2)	Properties	IRBAlteredConsentProcedure	
45 CFR 46.116(c)(1-2)	Properties	Local Govt	
45 CFR 46.116(c)(1-2)	Axioms:	PublicBenefitProgramResearch = (PublicBenefitProgram OR PublicServiceProgram) AND (ResearchConductedBy some State Govt) OR (ResearchApprovedBy some Local Govt)	

Section Text	Type	Term	Value
45 CFR 46.116(c)(1-2)	Axioms:	PublicBenefitProgramResearch = (PublicBenefitProgram OR PublicServiceProgram) AND (ResearchConductedBy some State Govt) OR (ResearchApprovedBy some Local Govt) AND (HasResearchFocus some ChangesToPaymentMethods) OR (HasResearchFocus some ChangesToPaymentLevels) AND ResearchImpracticalWithoutConsentAlteration	
45 CFR 46.116(c)(1-2)	Axioms:	PublicBenefitProgramResearch = (PublicBenefitProgram OR PublicServiceProgram) AND (ResearchConductedBy some State Govt) OR (ResearchApprovedBy some Local Govt) AND (HasResearchFocus some ChangesToPublicServiceProgram) OR (HasResearchFocus some ChangesToPublicBenefitProgram) AND ResearchImpracticalWithoutConsentAlteration	
45 CFR 46.116(c)(1-2)	Axioms:	PublicBenefitProgramResearch = (PublicBenefitProgram OR PublicServiceProgram) AND (ResearchConductedBy some State Govt) OR (ResearchApprovedBy some Local Govt) AND (HasResearchFocus some ProcedureforObtainingServices) OR (HasResearchFocus some ProcedureforObtainingBenefits) AND ResearchImpracticalWithoutConsentAlteration	
45 CFR 46.116(c)(1-2)	Axioms:	PublicBenefitProgramResearch = (PublicBenefitProgram OR PublicServiceProgram) AND (ResearchConductedBy some State Govt) OR (ResearchApprovedBy some Local Govt) AND (HasResearchFocus some ProcedureforObtainingServices) OR (HasResearchFocus some ProcedureforObtainingBenefits) AND ResearchImpracticalWithoutConsentAlteration	

Section Text	Type	Term	Value
45 CFR 46.116(c)(1-2)	Properties	ResearchApprovedBy	
45 CFR 46.116(c)(1-2)	Properties	ResearchConductedBy	
45 CFR 46.116(c)(1-2)	Properties	State Govt	
45 CFR 46.116(c)(1-2), 46.116(d)(1-4)	Class	Altered Consent	Research Impractical Without Consent Alteration
45 CFR 46.116(d)(1-4)	Class	Additional Information Provided Participant	Additional Information Provided After Participation
45 CFR 46.116(d)(1-4)	Class	Altered Consent	Alteration Of Consent
45 CFR 46.116(d)(1-4)	Class	Altered Consent	IRB Waiver Of Consent
45 CFR 46.116(d)(1-4)	Class	Altered Consent	Subject Rights Not Impacted
45 CFR 46.116(d)(1-4)	Class	Altered Consent	Subject Welfare Not Impacted
45 CFR 46.116(d)(1-4)	Axioms:	AllowsAlteredConsent == (Research HasRiskMagnitude MinimalRisk) AND ResearchImpracticalWithoutConsentAlteration AND some AdditionalInformationProvidedAfterParticipation AND (WaiverOfConsent DoesNotImpact Subject Rights) AND (AlterationOfConsent DoesNotImpact Subject Welfare)	
45 CFR 46.116(d)(1-4)	Properties	HasAlterationofConsent	
45 CFR 46.116(d)(1-4)	Properties	HasRiskType	
45 CFR 46.116(d)(1-4)	Properties	HasWaiverofConsent	
45 CFR 46.116(d)(1-4)	Properties	Research	
45 CFR 46.116(d)(1-4), 46.303, 46.306(a)(1), 46.306(a)(2)(i-ii)	Class	Risks To Participant	Minimal Risk

Section Text	Type	Term	Value
45 CFR 46.116(d)(1-4), 46.405(a-c), 46.406(a-d)	Properties	Involves	
45 CFR 46.116(e)	Class	Limits Of Policy	Jurisdictions Of Relevance To Study
45 CFR 46.116(e)	Class	Limits Of Policy	No Preemption Of Other Governing Bodies
45 CFR 46.116(e)	Class	Limits Of Policy	Policy Does Not Preempt Federal State Local Requirements
45 CFR 46.116(f)	Class	Limits Of Policy	No Limits On Authority of Physician To Provide Emergency Care
45 CFR 46.117(a-b)	Class	Consent Media	Orally Presented Elements Of Consent
45 CFR 46.117(a-b)	Class	Consent Media	Short Written Consent
45 CFR 46.117(a-b)	Class	Consent Media	Signed Informed Consent Form
45 CFR 46.117(a-b)	Class	Consent Media	Signed Short Informed Consent Form
45 CFR 46.117(a-b)	Class	Consent Media	Written Consent Form
45 CFR 46.117(a-b)	Class	Required Statement	Summary Of Research For Waived Documentation Studies
45 CFR 46.117(a-b)	Class	Required Statement	Summary Of Short Consent
45 CFR 46.117(a-b)	Class	Required Statement	Written Summary Of Orally Presented Consent
45 CFR 46.117(a-b)	Class	Witness	Legal Witness To Oral Presentation of Consent
45 CFR 46.117(a-b)	Properties	Documented_by	
45 CFR 46.117(a-b)	Axioms:	ExecutedInformedConsentForm == ExecutedLongInformedConsentForm OR ExecutedShortInformedConsentForm	

Section Text	Type	Term	Value
45 CFR 46.117(a-b)	Axioms:	ExecutedLongInfomredConsentForm == SignedInformedConsentForm	
45 CFR 46.117(a-b)	Axioms:	ExecutedShortInformedConsentForm == ShortWrittenConsent	
45 CFR 46.117(a-b)	Axioms:	ShortWrittenConsent Requires some LegalWitnessToOralPresentation	
45 CFR 46.117(a-b)	Axioms:	ShortWrittenConsent Requires some OralyPresentedElementsOfConsent	
45 CFR 46.117(a-b)	Axioms:	ShortWrittenConsent Requires some SignatureByConsentObtainerOnSummaryOfShortConsent	
45 CFR 46.117(a-b)	Axioms:	ShortWrittenConsent Requires some SignatureByLegalWitnessOnShortConsent	
45 CFR 46.117(a-b)	Axioms:	ShortWrittenConsent Requires some SignatureByLegalWitnessOnSummaryOfShortConsent	
45 CFR 46.117(a-b)	Axioms:	SignedInformedConsentForm == WrittenConsentForm Signed_by Legally authorized representative of subject	
45 CFR 46.117(a-b)	Axioms:	SignedInformedConsentForm == WrittenConsentForm Signed_by some Research Subject	
45 CFR 46.117(c)	Class	Consent Type	Consent Normally Not Required For Procedures
45 CFR 46.117(c)	Class	Consent Type	IRB Waiver Of Signed Consent Requirement
45 CFR 46.117(c)	Class	Risk Type	Principal Risk
45 CFR 46.117(c)	Class	Risks To Participant	Breach Of Confidentiality

Section Text	Type	Term	Value
45 CFR 46.117(c)	Axioms:	IRBWaiverOfSignedConsentRequirement Requires (PrincipalRisk some BreachOfConfidentiality) AND SignedConsentDocument some UniqueIdentifier) OR ConsentNormallyNotRequiredForProcedures	
45 CFR 46.201	Class	Participant Type	Human Fetuses
45 CFR 46.201	Class	Participant Type	Pregnant women
45 CFR 46.201	Equivalence Class:	Department of Health and Human Services == DHHS	
45 CFR 46.201, 46.205(a)(1-4), 46.205(b)(1-2)	Class	Participant Type	Neonates Of Uncertain Viability
45 CFR 46.202	Class	Legal Authority	Secretary
45 CFR 46.202	Class	Medical Event	Delivery
45 CFR 46.202	Class	Medical Event	Pregnancy
45 CFR 46.202	Class	Participant Type	Fetus
45 CFR 46.202	Class	Participant Type	Neonate
45 CFR 46.202	Class	Participant Type	Viable Neonate
45 CFR 46.202	Equivalence Class:	Neonate == Newborn	
45 CFR 46.202	Equivalence Class:	Secretary == SecretaryOfDHHS	
45 CFR 46.202, 46.206(a-b)	Class	Participant Type	Dead Fetus
45 CFR 46.202,46.205(c)(1-5), 46.201, 46.205(a)(1-4)	Class	Participant Type	Nonviable Neonate
45 CFR 46.204(a)	Class	Constraints On Approval of Research	Data Exists for Assessing Potential Risks To Fetuses
45 CFR 46.204(a)	Class	Constraints On Approval of Research	Data Exists for Assessing Potential Risks To Pregnant Women

Section Text	Type	Term	Value
45 CFR 46.204(a)	Class	Constraints On Approval of Research	Prior Clinical Studies exist for assessing potential risks
45 CFR 46.204(a)	Class	Constraints On Approval of Research	Prior Pre-clinical Studies exist for assessing potential risks
45 CFR 46.204(a)	Class	Constraints On Approval of Research	Prior Studies On Nonpregnant Women exist for assessing potential risks
45 CFR 46.204(a)	Class	Constraints On Approval of Research	Prior Studies On Pregnant Animals exist for assessing potential Risks
45 CFR 46.204(b)	Class	Risk Cause	Risk caused by interventions with possible Direct Benefit To Fetus
45 CFR 46.204(b)	Class	Risk Cause	Risk caused by interventions with possible Direct Benefit To Woman
45 CFR 46.204(b)	Class	Risk Cause	Risk caused by procedures with possible Direct Benefit To Fetus
45 CFR 46.204(b)	Class	Risk Cause	Risk caused by procedures with possible Direct Benefit To Woman
45 CFR 46.204(b)	Class	Risk Magnitude	Risk to Fetus Not Greater than Minimal
45 CFR 46.204(b)	Class	Risk Purpose	Development of important biomedical knowledge otherwise unobtainable
45 CFR 46.204(c), 46.205(b)(1-2)	Class	Risk Magnitude	Least Possible Risk to achieve research objectives
45 CFR 46.204(f)	Class	Consent Administrator	Individual Providing Consent to a participant
45 CFR 46.204(f)	Class	Risk	Impact Of Research On Fetus
45 CFR 46.204(f)	Class	Risk	Impact Of Research On Neonate

Section Text	Type	Term	Value
45 CFR 46.204(g)	Class	Legal Authority	This Part
45 CFR 46.204(g)	Class	Participant Type	Children Who Are Pregnant
45 CFR 46.204(g)	Equivalence Class:	ThisPart == 45CFR46	
45 CFR 46.204(g), 46.402(b)	Class	Assent	Assent
45 CFR 46.204(g), 46.402(c)	Class	Permission Type	Permission
45 CFR 46.204(h-i)	Class	Clinical Event	Termination O fPregnancy
45 CFR 46.204(h)	Class	Disallowed Inducements	Monetary Inducements To Terminate Pregnancy
45 CFR 46.204(h)	Class	Disallowed Inducements	No Inducements To Terminate Pregnancy
45 CFR 46.204(h)	Class	Participant Clinical Status	Viability of Neonate
45 CFR 46.204(h)	Axioms:	Researchers Can'tParticipateIn some DeterminingViabilityOfNeonate	
45 CFR 46.204(i)	Class	Legal Authority	Decision Maker
45 CFR 46.204(i)	Class	Method of Pregnancy Termination	Method
45 CFR 46.204(i)	Class	Timing of Pregnancy Termination	Timing
45 CFR 46.204(i)	Properties	Decision	
45 CFR 46.204(i), 46.402(a)	Class	Medical Procedure	Medical Procedures
45 CFR 46.205(a)(1-4)	Class	Medical Status	Assessing Potential Risks To Neonates
45 CFR 46.205(a)(1-4)	Class	Research Outcomes	Reasonably Foreseeable Impact
45 CFR 46.205(a)(1-4)	Class	Scientifically Appropriate Clinical Studies	

Section Text	Type	Term	Value
45 CFR 46.205(a)(1-4)	Axioms:	Researchers Can'tParticipateIn some DeterminingViabilityOfNeonate	
45 CFR 46.205(a)(1-4), 46.205(b)(1-2)	Properties	dataExists	
45 CFR 46.205(b)(1-2)	Class	Benefit	Enhanced Probability Of Survival
45 CFR 46.205(b)(1-2)	Class	Clinical Event	Viability of Neonate Ascertained
45 CFR 46.205(b)(1-2)	Class	Legally Effective Informed Consent	Legally Effective Informed Consent Of Father
45 CFR 46.205(b)(1-2)	Class	Legally Effective Informed Consent	Legally Effective Informed Consent Of Parent
45 CFR 46.205(b)(1-2)	Class	Legally Effective Informed Consent	Legally Effective Informed Consent Of Parents Authorized Representative
45 CFR 46.205(b)(1-2)	Class	Pregnancy Due to Incest	Incest
45 CFR 46.205(b)(1-2)	Class	Pregnancy Due to Rape	Rape
45 CFR 46.205(b)(1-2)	Class	Research Purpose	Knowledge That Cannot Be Otherwise Obtained
45 CFR 46.205(b)(1-2)	Properties	DeterminationByIRB	
45 CFR 46.205(b)(1-2)	Axioms:	Researchers Can'tParticipateIn some DeterminingViabilityOfNeonate	
45 CFR 46.205(b)(1-2), 46.205(c)(1-5)	Class	Risks To Participant	No Added Risk To Neonate
45 CFR 46.205(c)(1-5)	Class	Research Method	Neonate Vital Function Not Artificially Maintained
45 CFR 46.205(c)(1-5)	Class	Research Method	Research Will Not Terminate Heart Or Respiration
45 CFR 46.205(c)(1-5)	Properties	ConsentFrom	

Section Text	Type	Term	Value
45 CFR 46.205(c)(1-5)	Axioms:	DoesNotSuffice ConsentFrom some ParentsLegallyAuthorizedRepresentative	
45 CFR 46.205(c)(1-5)	Axioms:	LegallyEffectiveInformedConsent Requires some LegallyEffectiveInformedConsentOfParent	
45 CFR 46.205(c)(1-5)	Axioms:	NoWaiverAndNoAlterationOfConsent	
45 CFR 46.205(c)(1-5)	Properties	Requires	
45 CFR 46.205(c)(1-5)	Axioms:	Researchers Can'tParticipateIn some DeterminingViabilityOfNeonate	
45 CFR 46.205(d)	Class	Research Participant	Viable Neonate
45 CFR 46.206(a-b)	Class	Legal Authority	Applicable Federal Law
45 CFR 46.206(a-b)	Class	Legal Authority	Applicable Federal Regulations
45 CFR 46.206(a-b)	Class	Legal Authority	Applicable Local Law
45 CFR 46.206(a-b)	Class	Legal Authority	Applicable Local Regulation
45 CFR 46.206(a-b)	Class	Legal Authority	Applicable State Law
45 CFR 46.206(a-b)	Class	Legal Authority	Applicable State Regulations
45 CFR 46.206(a-b)	Class	Research Materials	Cells
45 CFR 46.206(a-b)	Class	Research Materials	Fetal Material
45 CFR 46.206(a-b)	Class	Research Materials	Identifiable Information
45 CFR 46.206(a-b)	Class	Research Materials	Macerated Fetal Materials
45 CFR 46.206(a-b)	Class	Research Materials	Organs
45 CFR 46.206(a-b)	Class	Research Materials	Placenta
45 CFR 46.206(a-b)	Class	Research Materials	Tissue

Section Text	Type	Term	Value
45 CFR 46.207	Class	Research Type	Research Not Otherwise Approvable
45 CFR 46.207, 46.207(a)	Class	Research Focus	Serious Problem Affecting Fetal Health
45 CFR 46.207, 46.207(a)	Class	Research Focus	Serious Problem Affecting Fetal Welfare
45 CFR 46.207, 46.207(a)	Class	Research Focus	Serious Problem Affecting Neonatal Health
45 CFR 46.207, 46.207(a)	Class	Research Focus	Serious Problem Affecting Neonatal Welfare
45 CFR 46.207, 46.207(a)	Class	Research Focus	Serious Problem Affecting Pregnant Women's Welfare
45 CFR 46.207, 46.207(a)	Class	Research Focus	Serious Problem Affecting Pregnan tWomen sHealth
45 CFR 46.207, 46.407	Class	Research Focus	Opportunity To Alleviate
45 CFR 46.207, 46.407	Class	Research Focus	Opportunity To Prevent
45 CFR 46.207, 46.407	Class	Research Focus	Opportunity To Understand
45 CFR 46.207(a-b)	Properties	ToAlleviate	
45 CFR 46.207(a-b)	Properties	ToPrevent	
45 CFR 46.207(a-b)	Properties	ToUnderstand	
45 CFR 46.207(a)	Class	Communication	Public Meeting Announcement
45 CFR 46.207(a)	Class	Legal Authority	Federal Register
45 CFR 46.207(a)	Class	Legal Authority	The Secretary
45 CFR 46.207(a)	Class	Legal Finding	IRB Finding Of Reasonable Opportunity
45 CFR 46.207(a)	Class	Legal Process	Opportunity For Public Review And Comment

Section Text	Type	Term	Value
45 CFR 46.207(a)	Class	Research Focus	Alleviation Of Serious Health Problem
45 CFR 46.207(a)	Class	Research Focus	PreventionOfSeriousHealthProblem
45 CFR 46.207(a)	Class	Research Focus	Understanding Of Serious Problem
45 CFR 46.207(a)	Class	Research Process	Conduct Research
45 CFR 46.207(a)	Class	Review Process	Consultation With Expert Panel
45 CFR 46.207(a)	Class	Review Process	Panel Of Experts In Pertinent Disciplines
45 CFR 46.207(a)	Properties	WillFundResearch	
45 CFR 46.207(b), 46.407	Class	Ethics	Sound Ethical Principles
45 CFR 46.301	Class	Research Type	Behavioral Research Involving Prisoners
45 CFR 46.301	Class	Research Type	Biomedical Research Involving Prisoners
45 CFR 46.301	Axioms:	hasAdditionalProtections some BehavioralResearchInvolvingPrisoners	
45 CFR 46.301	Axioms:	hasAdditionalProtections some BiomedicalResearchInvolvingPrisoners	
45 CFR 46.301, 46,401(a)(1-2)	Class	Protections Afforded Participant	Additional Protections Afforded Prisoners
45 CFR 46.301, 46.401(a)(1-2)	Class	Dept of Health and Human Services	Conducted By DHHS
45 CFR 46.301, 46.401(a)(1-2)	Class	Dept of Health and Human Services	Funded By DHHS
45 CFR 46.301, 46.401(a)(1-2)	Properties	hasAdditionalProtections	
45 CFR 46.302	Class	Constraints On Participants	Constraints due to incarceration
45 CFR 46.302	Class	Research Participant Ethical Attribute	Truly Uncoerced Decision
45 CFR 46.302	Class	Research Participant Ethical Attribute	Truly Voluntary Decision
45 CFR 46.303	Class	Legal Authority	Civil Statute

Section Text	Type	Term	Value
45 CFR 46.303	Class	Legal Authority	Criminal Statute
45 CFR 46.303	Class	Location Of Research	Penal Institution
45 CFR 46.303	Class	Participant Type	Individuals Detained By Law
45 CFR 46.303	Class	Participant Type	Individuals Detained In Other Settings
45 CFR 46.303	Class	Participant Type	Prisoner
45 CFR 46.303	Class	Risk Magnitude	Routine Dental Examination
45 CFR 46.303	Class	Risk Magnitude	Routine Medical Examination
45 CFR 46.303	Class	Risk Magnitude	Routine Psychological Examination
45 CFR 46.306	Class	Research Type	Permitted Research Involving Prisoners
45 CFR 46.306(a)(1)	Class	Administrative Proceedures	Institutional Certification Of IRB Approval To DHHS
45 CFR 46.306(a)(1)	Class	Dept of Health and Human Services	DHHS Approval Of Research
45 CFR 46.306(a)(1)	Class	Research Focus	Behavioral Research
45 CFR 46.306(a)(1)	Class	Research Focus	Biomedical Research
45 CFR 46.306(a)(1), 46.306(a)(2)(i-ii)	Class	Risks	Inconvenience To Incarcerated Subjects Allowed
45 CFR 46.306(a)(1), 46.306(a)(2)(i)	Class	Research Focus	Causes of Criminal Behavior
45 CFR 46.306(a)(1), 46.306(a)(2)(i)	Class	Research Focus	Causes Of Incarceration
45 CFR 46.306(a)(1), 46.306(a)(2)(ii)	Class	Research Focus	Study Of Prisoners As Incarcerated People
45 CFR 46.306(a)(1), 46.306(a)(2)(ii)	Class	Research Focus	Study Of Prisons As Institutional Structure
45 CFR 46.306(a)(2)(iii)	Class	Research Focus	Conditions Affecting Prisoners As A Class

Section Text	Type	Term	Value
45 CFR 46.306(a)(2)(iii)	Class	Research Focus	Research On Psychological Problems
45 CFR 46.306(a)(2)(iii)	Class	Research Focus	Research On Social Problems
45 CFR 46.306(a)(2)(iv)	Class	Research Focus	Improving Health of Subject
45 CFR 46.306(a)(2)(iv)	Class	Research Focus	Improving Well Being of Subject
45 CFR 46.306(a)(2)(iv)	Class	Research Focus	Research On Accepted Practices
45 CFR 46.306(a)(2)(iv)	Class	Research Focus	Research On Innovated Practices
45 CFR 46.306(a)(2)(iv)	Class	Research Method	Assignment To Control Groups
45 CFR 46.401(a)(1-2)	Class	Administrative Proceedures	Administrative Procedural Modifications
45 CFR 46.401(a)(1-2)	Class	Dept of Health and Human Services	Operating Division Of DHHS
45 CFR 46.401(a)(1-2)	Class	Location Of Research	Research Conducted Outside The US
45 CFR 46.401(a)(1-2)	Class	Research Participant	Children Involved As Research Subjects
45 CFR 46.401(a)(1-2)	Axioms:	hasAdditionalProtections some ChildrenInvolvedAsResearchSubjects	
45 CFR 46.401(a)(1-2)	Authority:	Source: 48 FR 9818, March 8, 1983, unless otherwise noted	
45 CFR 46.402(a)	Class	Legal Authority	Applicable Law
45 CFR 46.402(a)	Class	Location Of Research	Jurisdiction Where Research Is Conducted
45 CFR 46.402(a)	Class	Person	Persons
45 CFR 46.402(a)	Class	Research Method	Treatments
45 CFR 46.402(a)	Class	Research Participant	Children
45 CFR 46.402(a)	Equivalence Class:	Children == Persons AND (NOT HasAttainedLegalAgeForConsent some JurisdictionWhereResearchIsConducted)	

Section Text	Type	Term	Value
45 CFR 46.402(a)	Properties	HasAttainedLegalAgeForConsent	
45 CFR 46.402(b-c)	Class	Assent	Childs Affirmative Agreement To Participate in Research
45 CFR 46.402(b)	Class	Failure To Object To Research	Failure To Object
45 CFR 46.402(b)	Equivalence Class:	Assent == Child Gives some AffirmativeAgreementToParticipate	
45 CFR 46.402(b)	Equivalence Class:	Child == Children	
45 CFR 46.402(b)	Equivalence Class:	Child Gives some FailureToObject NOT == Assent	
45 CFR 46.402(b)	Equivalence Class:	Children == Persons AND (NOT HasAttainedLegalAgeForConsent some JurisdictionWhereResearchIsConducted)	
45 CFR 46.402(b), 46.402(c)	Properties	Gives	
45 CFR 46.402(c-d)	Class	Guardian	Parent
45 CFR 46.402(c)	Class	Participant Type	Ward
45 CFR 46.402(c)	Equivalence Class:	Permission == Parent Gives some AffirmativeAgreementToParticipate OR Guardian Gives some AffirmativeAgreementToParticipate	
45 CFR 46.402(c), 46.202(e)	Class	Guardian	Guardian
45 CFR 46.402(c), 46.402(e)	Class	Participant Type	Child
45 CFR 46.402(d)	Class	Guardian	Adoptive Parent
45 CFR 46.402(d)	Class	Guardian	Biological Parent
45 CFR 46.402(d)	Equivalence Class:	Parent == (BiologicalParent OR AdoptiveParent) AND has some Child	

Section Text	Type	Term	Value
45 CFR 46.402(e)	Class	General Medical Care	General Medical Care
45 CFR 46.402(e)	Class	Legal Authority	Authorized To Consent on behalf of child
45 CFR 46.402(e)	Equivalence Class:	Guardian == Individual AND AuthorizedToConsent OnBehalfOf some Child inContext GeneralMedicalCare	
45 CFR 46.402(e)	Properties	inContext	
45 CFR 46.404	Class	Assent Process	Provisions For Assent Of Child
45 CFR 46.404	Class	Permission Type	Permission Of Parents
45 CFR 46.404	Class	Permission Type	Permissions Of Guardians
45 CFR 46.405(a-c)	Class	Benefit	Direct Benefit To Subject
45 CFR 46.405(a-c)	Class	Risk Magnitude	Risk Benefit Ratio Favorable To Alternatives
45 CFR 46.405(a-c)	Class	Risk Magnitude	Risk Justified By Anticipated Benefit
45 CFR 46.405(a-c), 46.406(a-d)	Class	Risk Magnitude	Greater Than Minimal Risk
45 CFR 46.405(a-c), 46.406(a-d) 46.407	Class	Assent	Assent Of Child
45 CFR 46.405(a-c), 46.406(a-d), 46.407	Class	Permission Type	Permission Of Parents Or Guardians
45 CFR 46.406(a-d)	Class	Benefit	No Direct Benefit To Subject
45 CFR 46.406(a-d)	Class	Benefit	Yield Generalizable Knowledge
45 CFR 46.406(a-d)	Class	Risk Magnitude	Intervention Experience Similar To Expected Clinica lSituation
45 CFR 46.406(a-d)	Class	Risk Magnitude	Minor Increase Over Minimal Risk

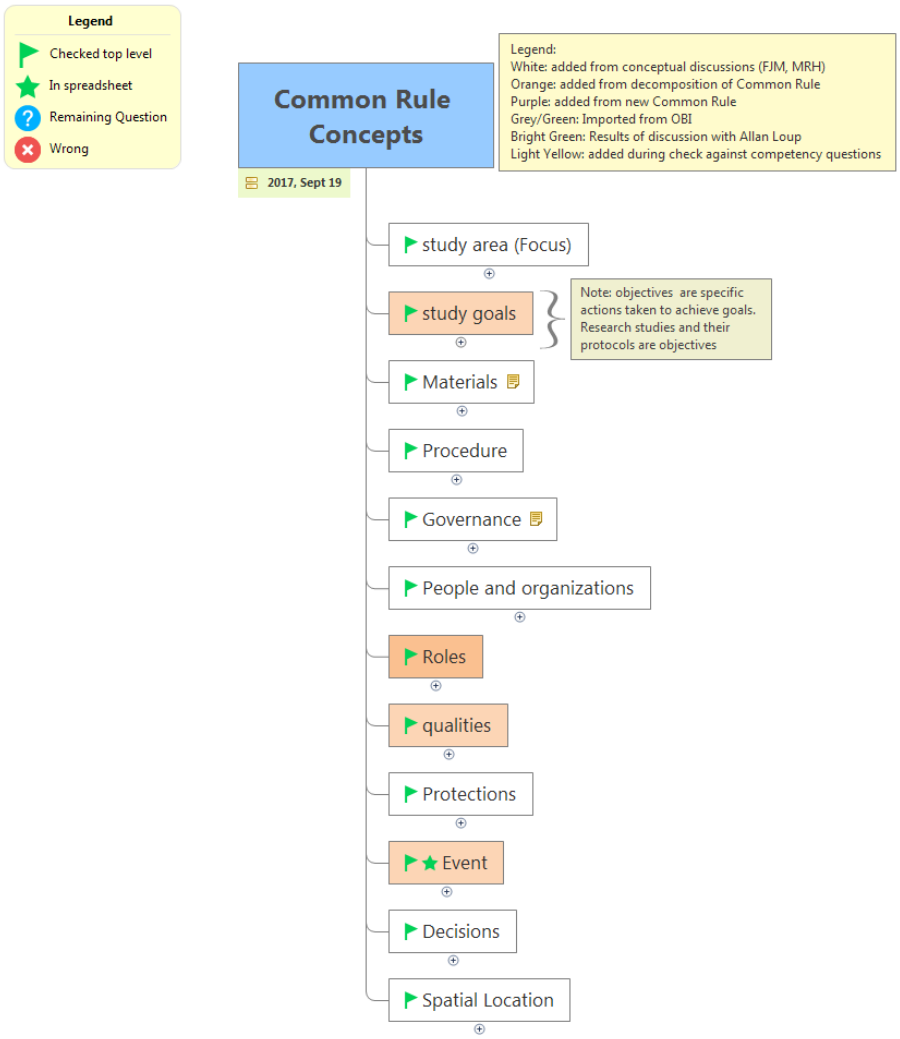
Section Text	Type	Term	Value
45 CFR 46.407	Class	Research Focus	Serious Health Problem Of Children
45 CFR 46.407	Class	Research Focus	Serious Welfare Problem Of Children
45 CFR 46.408(a)	Class	Participant Age	Age Of Child
45 CFR 46.408(a)	Class	Status of Child	Capable Of Providing Assent
45 CFR 46.408(a)	Class	Status of Child	Maturity Of Child
45 CFR 46.408(b)	Class	Parental Status	One Parent Has Legal Custody
45 CFR 46.408(b)	Class	Parental Status	Parent Deceased
45 CFR 46.408(b)	Class	Parental Status	Parent Incompetent
45 CFR 46.408(b)	Class	Parental Status	Parent Not Reasonably Available
45 CFR 46.408(b)	Class	Parental Status	Parent Unknown
45 CFR 46.408(b)	Class	Permission Type	Permission Of Both Parents
45 CFR 46.408(b)	Class	Permission Type	Permission Of One Parent
45 CFR 46.408(c)	Class	Research Method	Subject Population
45 CFR 46.408(c)	Class	Risks To Participant	Parental Consent Not Reasonable To Protect Child
45 CFR 46.408(d)	Class	Documentation Requirement	Consent By Guardians
45 CFR 46.408(d)	Class	Documentation Requirement	Consent By Parents
45 CFR 46.408(d)	Class	Documentation Requirement	IRB Determines Assent Documentation Requirements
45 CFR 46.409(a)	Class	Location Of Research	Research Conducted In Camps
45 CFR 46.409(a)	Class	Location Of Research	Research Conducted In Hospitals
45 CFR 46.409(a)	Class	Location Of Research	Research Conducted In Institutions

Section Text	Type	Term	Value
45 CFR 46.409(a)	Class	Location Of Research	Research Conducted In Schools
45 CFR 46.409(a)	Class	Participant Type	Wards Of Institution
45 CFR 46.409(a)	Class	Participant Type	Wards Of Other Agency
45 CFR 46.409(a)	Class	Participant Type	Wards Of State
45 CFR 46.409(a)	Class	Research Type	Majority Of Subjects Are Not Wards
45 CFR 46.409(b)	Class	Legal Agent Of Participant	Requires Appointment Of Advocate
45 CFR 46 Subpart C	Axioms:	hasAdditionalProtections some BehavioralResearchInvolvingPrisoners	
45 CFR 46 Subpart C	Axioms:	hasAdditionalProtections some BiomedicalResearchInvolvingPrisoners	
45 CFR 46 Subpart C	Authority:	Source: 43 FR 53655, Nov. 16, 1978, unless otherwise noted.	
45 CFR 46 Subpart D	Axioms:	hasAdditionalProtections some ChildrenInvolvedAsResearchSubjects	
45 CFR 46 Subpart D	Authority:	Source: 48 FR 9818, March 8, 1983, unless otherwise noted	

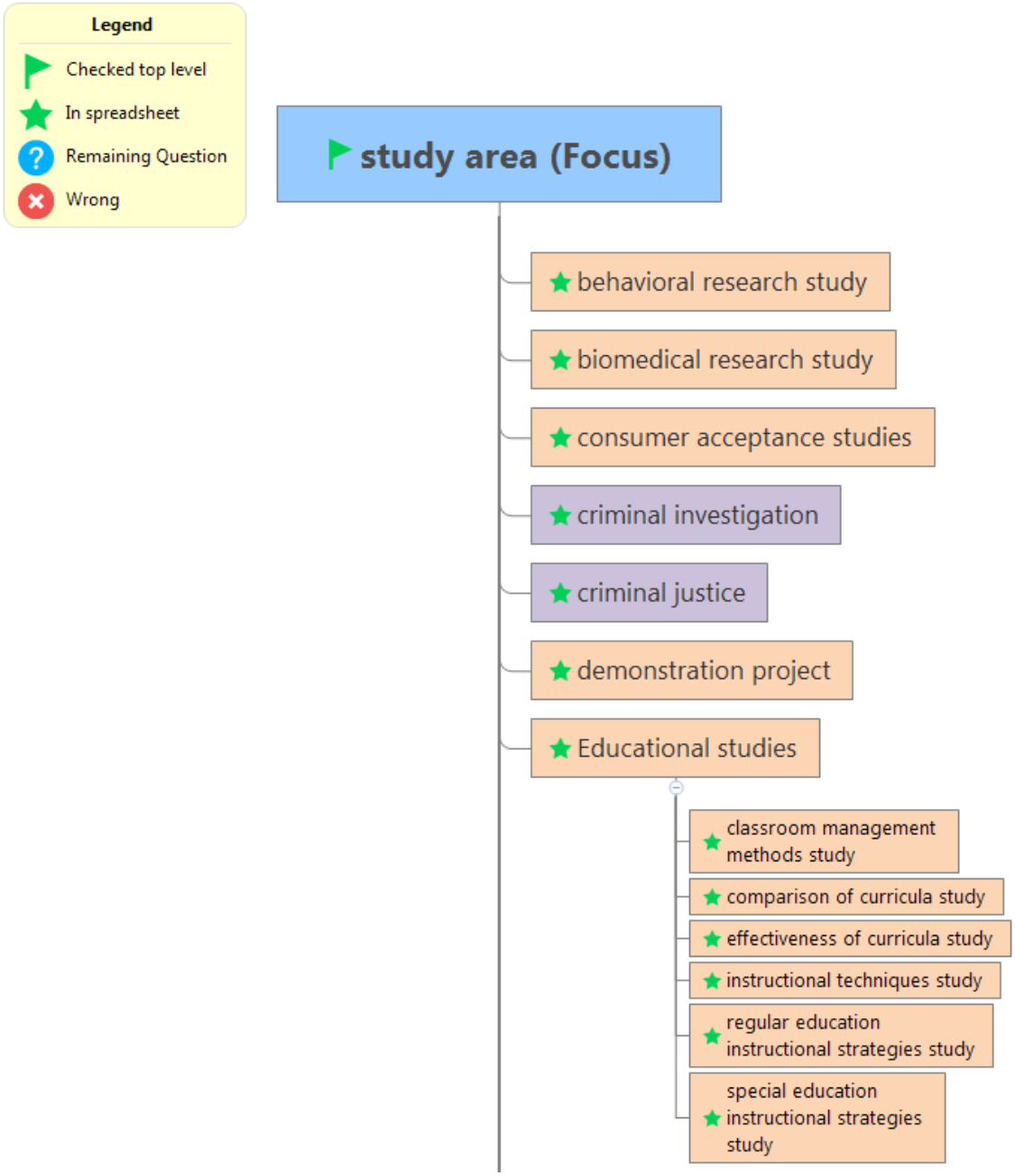
Appendix D. Concept Map Three

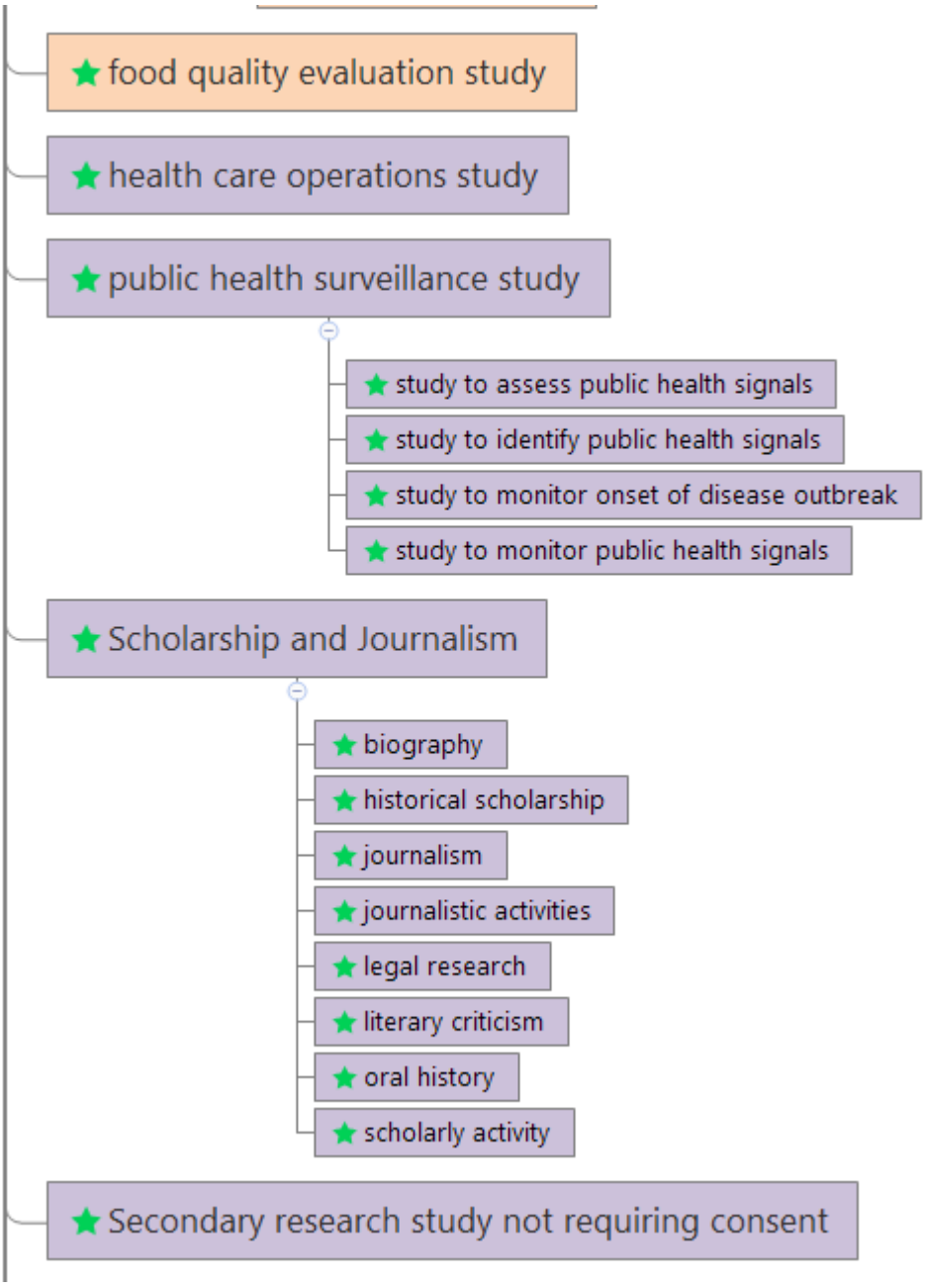
Concept Map Three is important because it was used for both refinement and iterative review with domain experts. Unfortunately, it is very large and must be broken up for presentation in this format. The overview at the top level is presented first, and then each major subsection of the map.

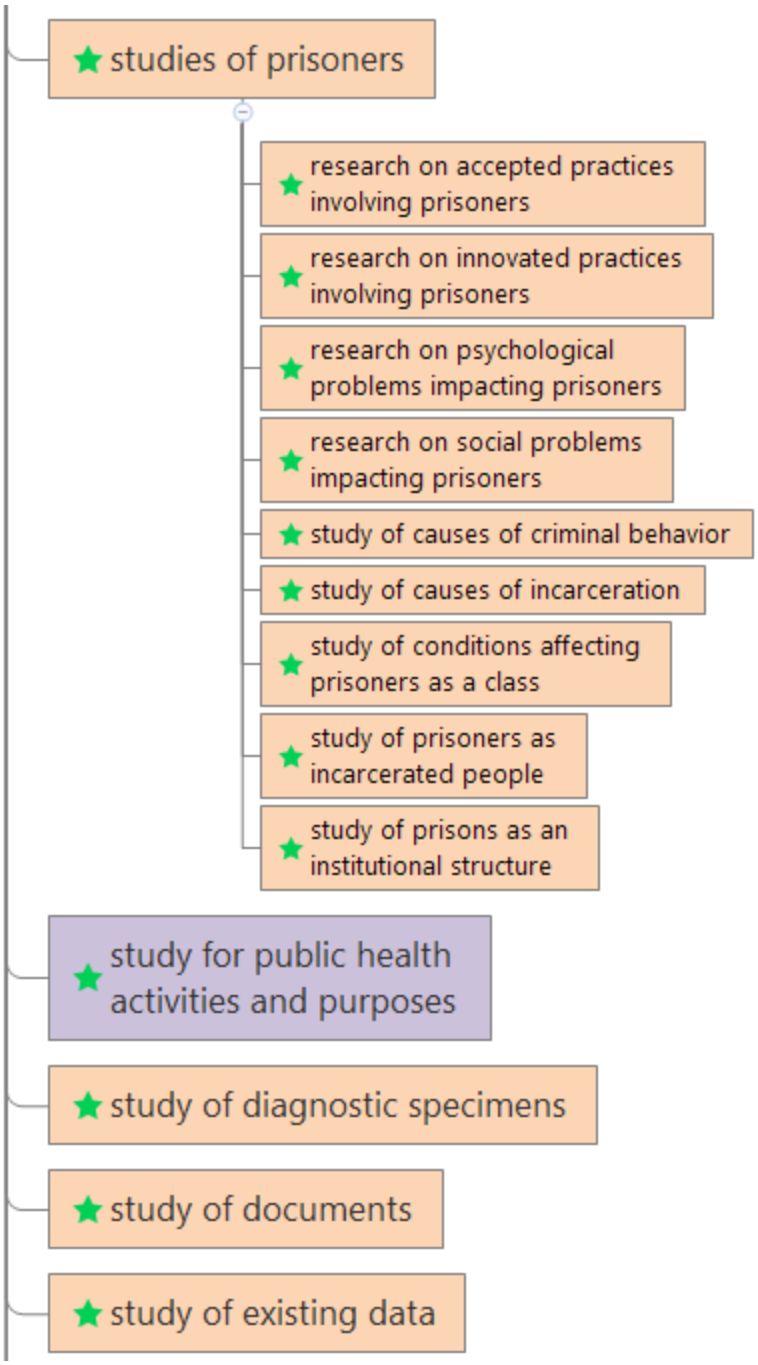
Concept Map — top level

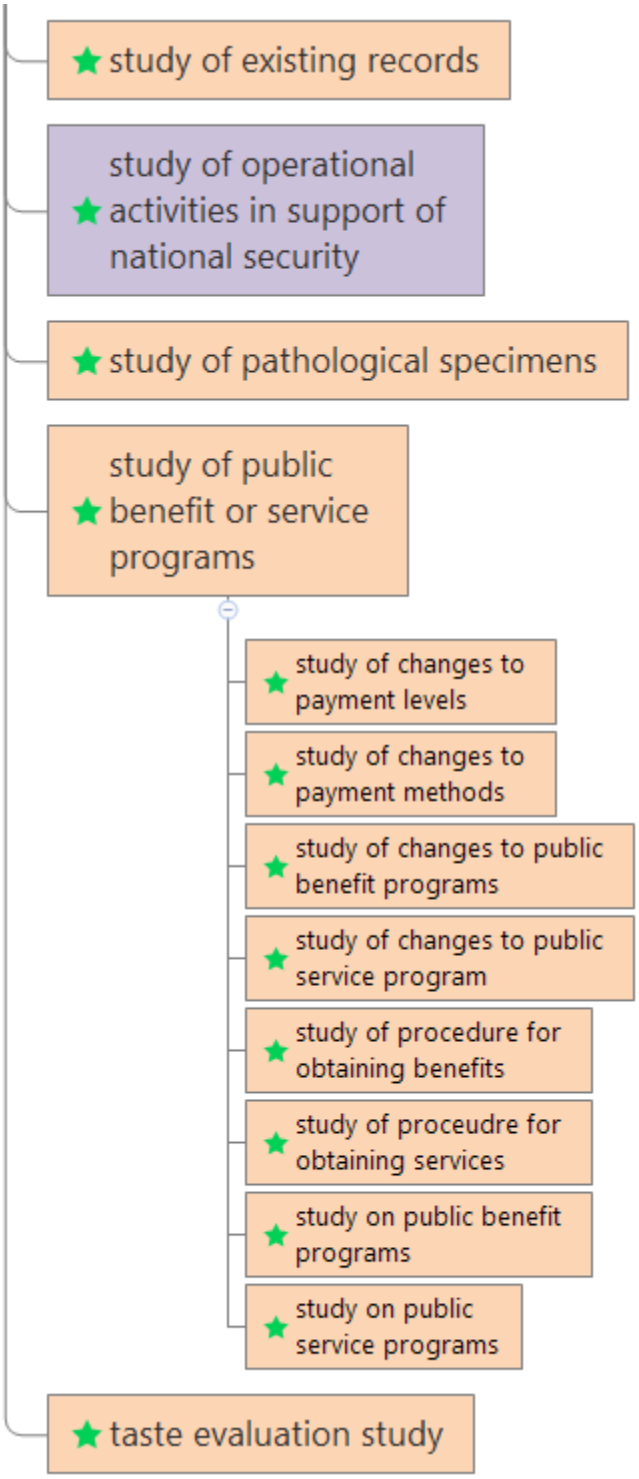


Study area or focus




food quality evaluation study
health care operations study
public health surveillance study
study to assess public health signals
study to identify public health signals
study to monitor onset of disease outbreak
study to monitor public health signals
Scholarship and Journalism
biography
historical scholarship
journalism
journalistic activities
legal research
literary criticism
oral history
scholarly activity
Secondary research study not requiring consent

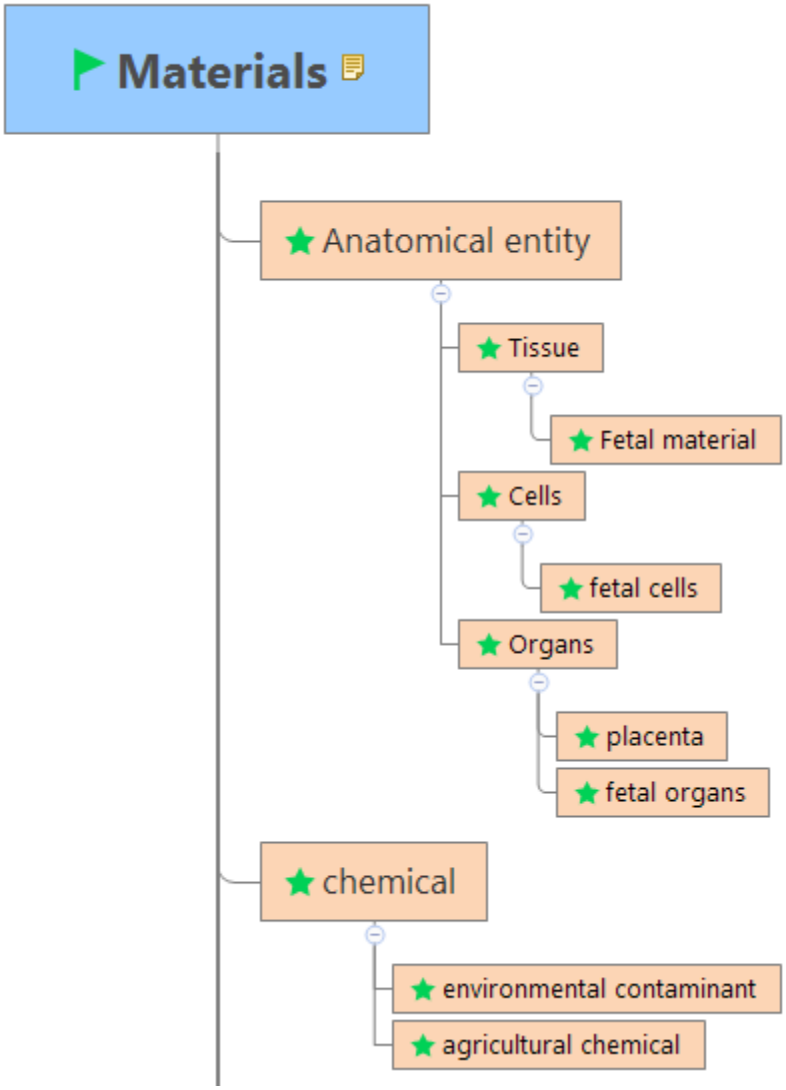

studies of prisoners
research on accepted practices involving prisoners
research on innovated practices involving prisoners
research on psychological problems impacting prisoners
research on social problems impacting prisoners
study of causes of criminal behavior
study of causes of incarceration
study of conditions affecting prisoners as a class
study of prisoners as incarcerated people
study of prisons as an institutional structure
study for public health activities and purposes
study of diagnostic specimens
study of documents
study of existing data

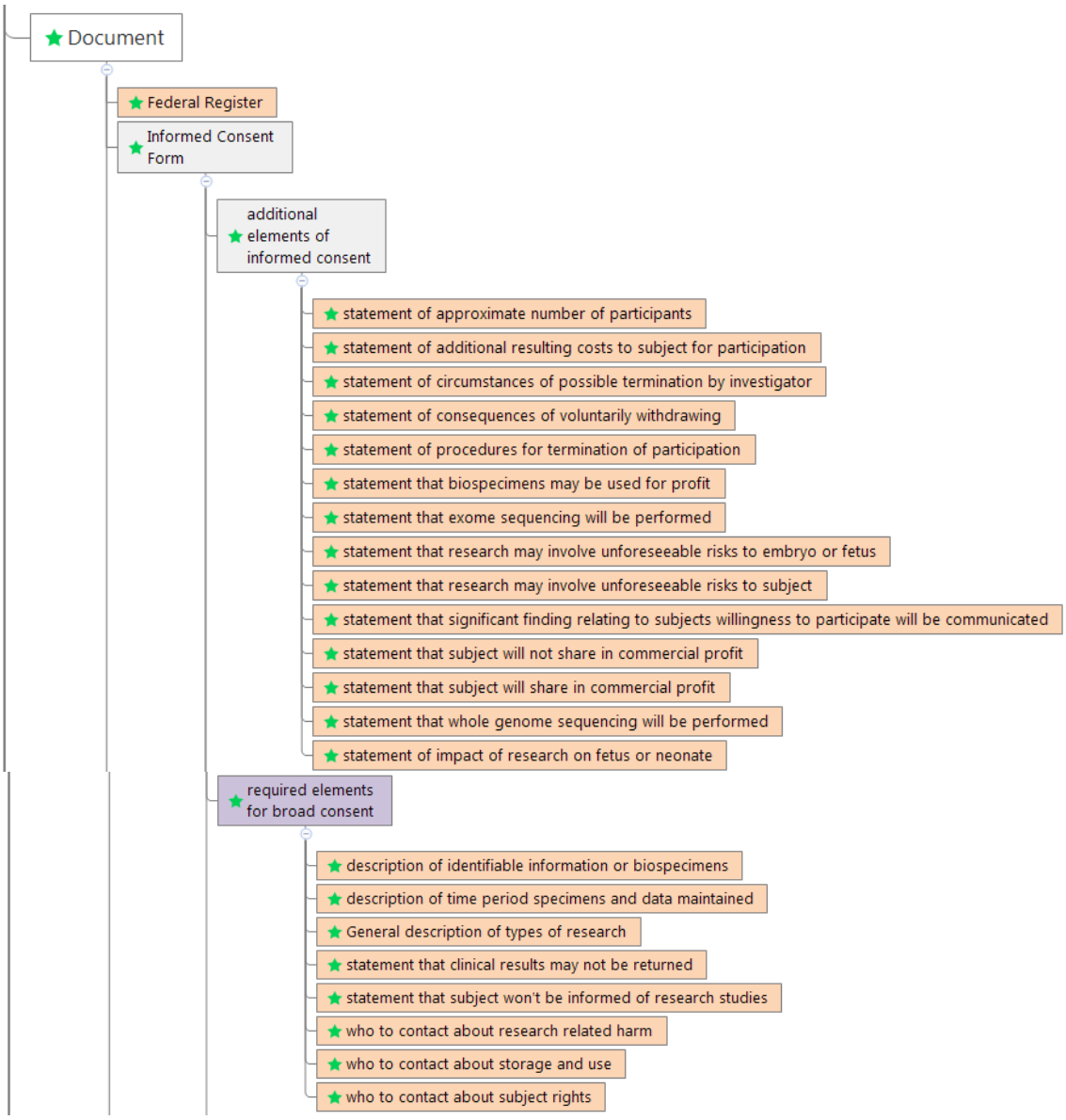

study of existing records
study of operational activities in support of national security
study of pathological specimens
study of public benefit or service programs
study of changes to payment levels
study of changes to payment methods
study of changes to public benefit programs
study of changes to public service program
study of procedure for obtaining benefits
study of proceudre for obtaining services
study on public benefit programs
study on public service programs
taste evaluation study

Study Goals

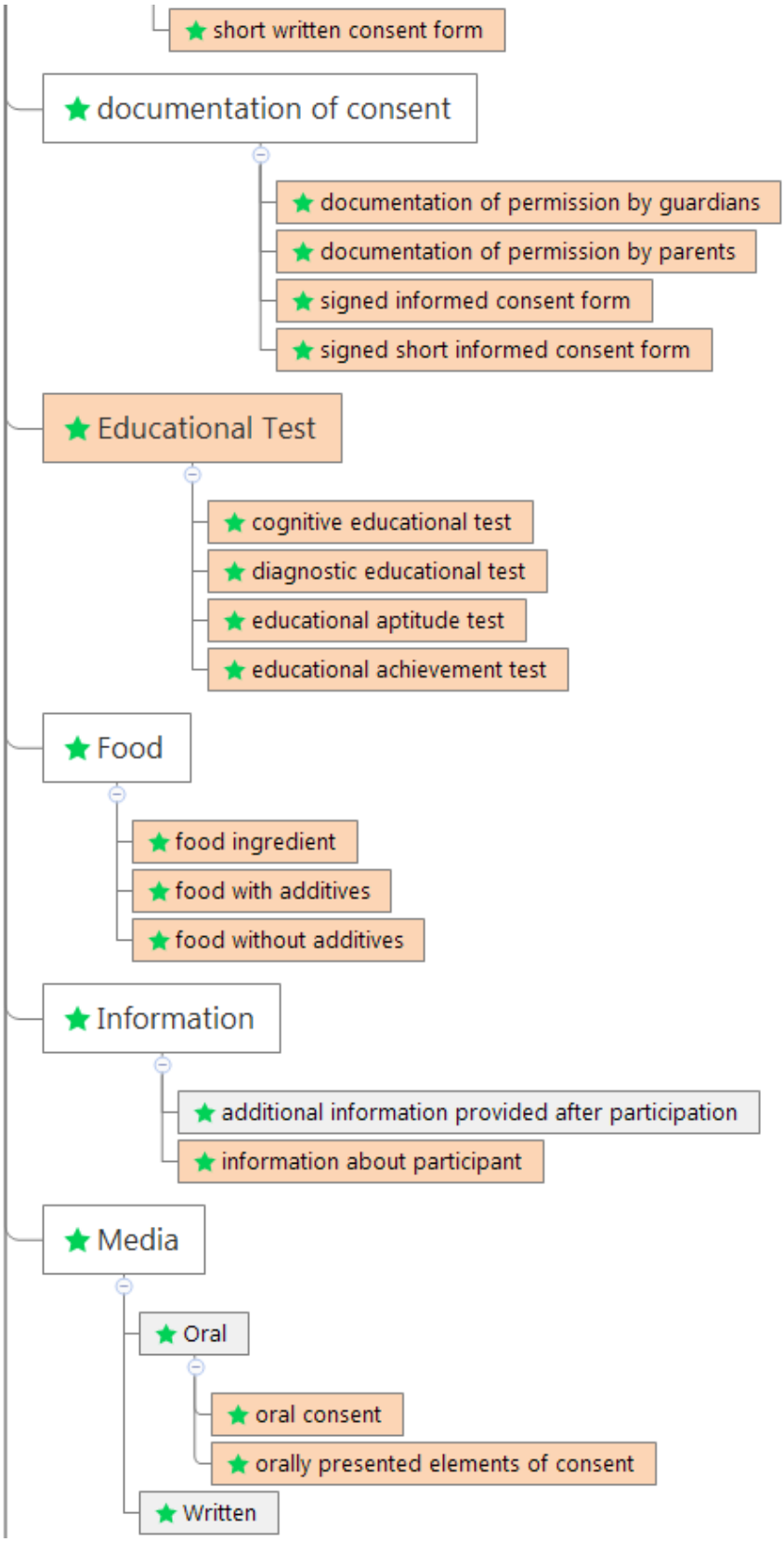

Legend
Checked top level
In spreadsheet
Remaining Question
Wrong
study goals
development of important biomedical knowledge otherwise unobtainable
knowledge that cannot be otherwise obtained
opportunity to alleviate a serious health problem
opportunity to prevent a serious health problem
opportunity to understand a serious health problem
yield generalizable knowledge

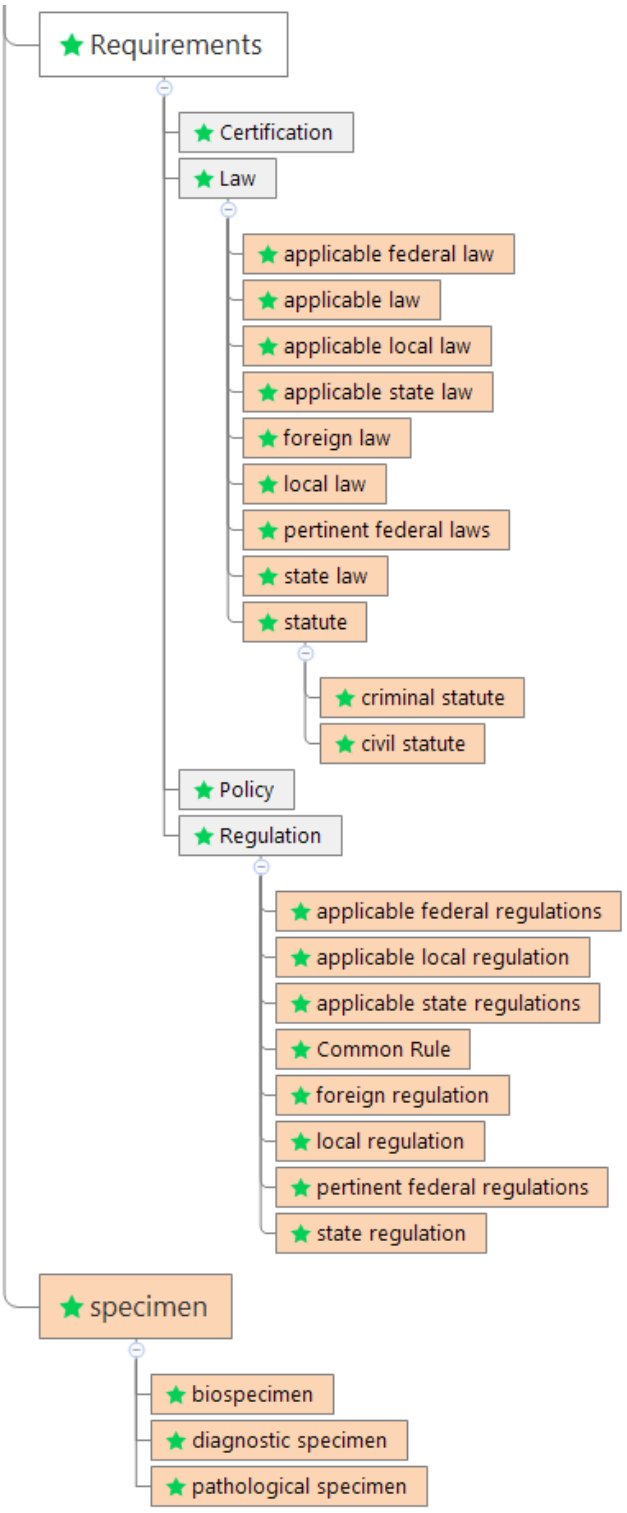
Materials


Materials
Anatomical entity
Tissue
Fetal material
Cells
fetal cells
Organs
placenta
fetal organs
chemical
environmental contaminant
agricultural chemical

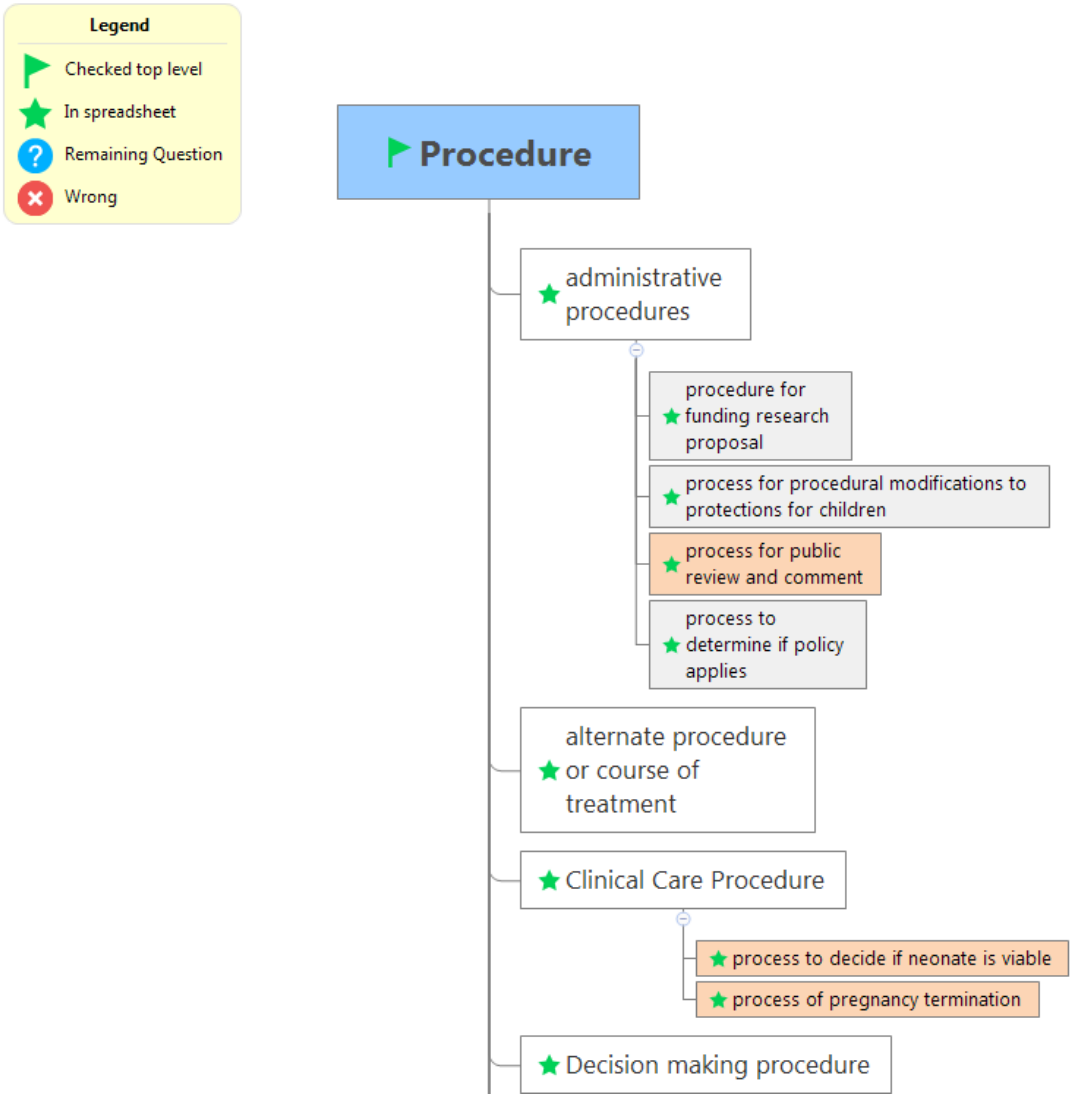

Document
Federal Register
Informed Consent Form
additional elements of informed consent
statement of approximate number of participants
statement of additional resulting costs to subject for participation
statement of circumstances of possible termination by investigator
statement of consequences of voluntarily withdrawing
statement of procedures for termination of participation
statement that biospecimens may be used for profit
statement that exome sequencing will be performed
statement that research may involve unforeseeable risks to embryo or fetus
statement that research may involve unforeseeable risks to subject
statement that significant finding relating to subjects willingness to participate will be communicated
statement that subject will not share in commercial profit
statement that subject will share in commercial profit
statement that whole genome sequencing will be performed
statement of impact of research on fetus or neonate
required elements for broad consent
description of identifiable information or biospecimens
description of time period specimens and data maintained
General description of types of research
statement that clinical results may not be returned
statement that subject won't be informed of research studies
who to contact about research related harm
who to contact about storage and use
who to contact about subject rights

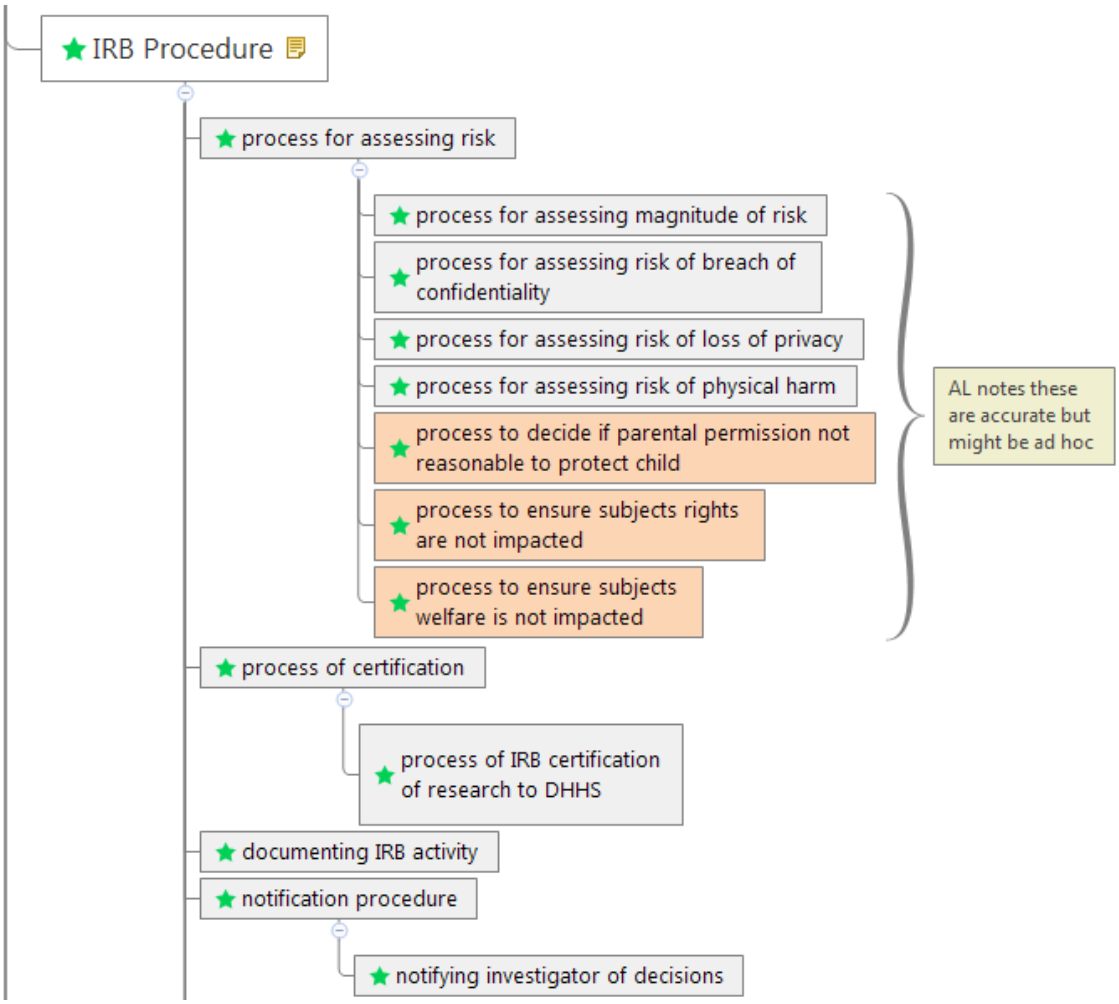

required informed consent element
statement about availability of medical care for injury
statement about compensation for participation
statement about confidentiality of records identifying subject
description of benefits to others
description of benefits to subject
description of discomforts
description of risks
disclosure of advantageous alternative procedures
disclosure of advantageous alternative treatments
statement that subject may discontinue participation at any time
statement of no loss of entitled benefits for dicontinuing participation
statement of no loss of entitled benefits for refusal to participate
statement of no penalty for discontinuing participation
statement of no penalty for refusal to participate
statement that participation is voluntary
statement of collection of identifiable biospecimens
statement of collection of identifiable private information
summary of research for waived documentation studies
summary of short consent
statement about what medical care for injury consists of
where futher information can be obtained
who to contact for questions about the research
statement that specimens and data will be used for future research
statement that specimens and data will not be used for future research
written summary of orally presented consent
statement identifying procedures which are experimental
statement describing the research procedures to be followed
statement explaining the purposes of the research
statement that the study involves research
statement about compensation for injury
who to contact for questions about the subjects rights
who to contact in the event of research related injury

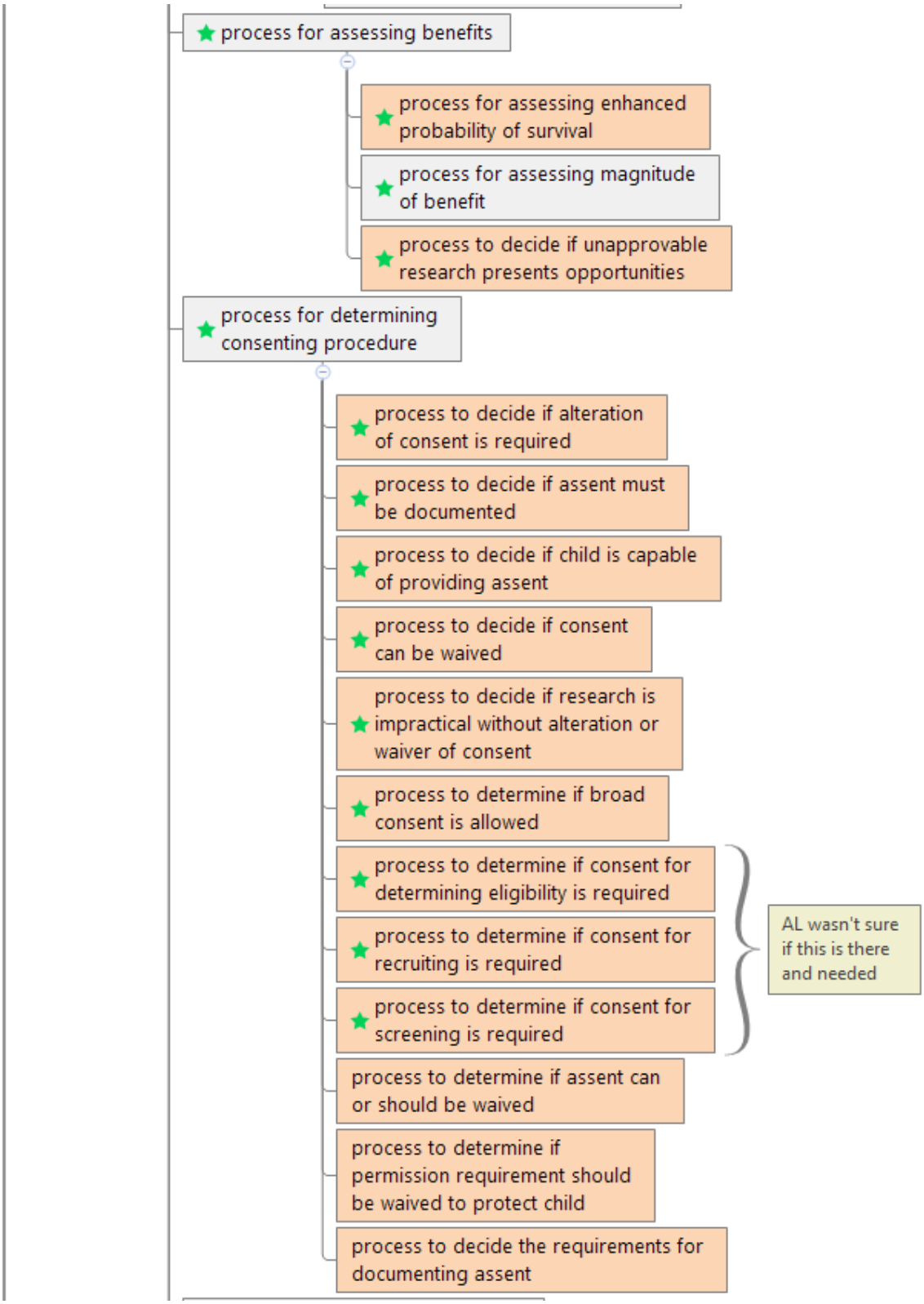

short written consent form
documentation of consent
documentation of permission by guardians
documentation of permission by parents
signed informed consent form
signed short informed consent form
Educational Test
cognitive educational test
diagnostic educational test
educational aptitude test
educational achievement test
Food
food ingredient
food with additives
food without additives
Information
additional information provided after participation
information about participant
Media
Oral
oral consent
orally presented elements of consent
Written


Requirements
Certification
Law
applicable federal law
applicable law
applicable local law
applicable state law
foreign law
local law
pertinent federal laws
state law
statute
criminal statute
civil statute
Policy
Regulation
applicable federal regulations
applicable local regulation
applicable state regulations
Common Rule
foreign regulation
local regulation
pertinent federal regulations
state regulation
specimen
biospecimen
diagnostic specimen
pathological specimen

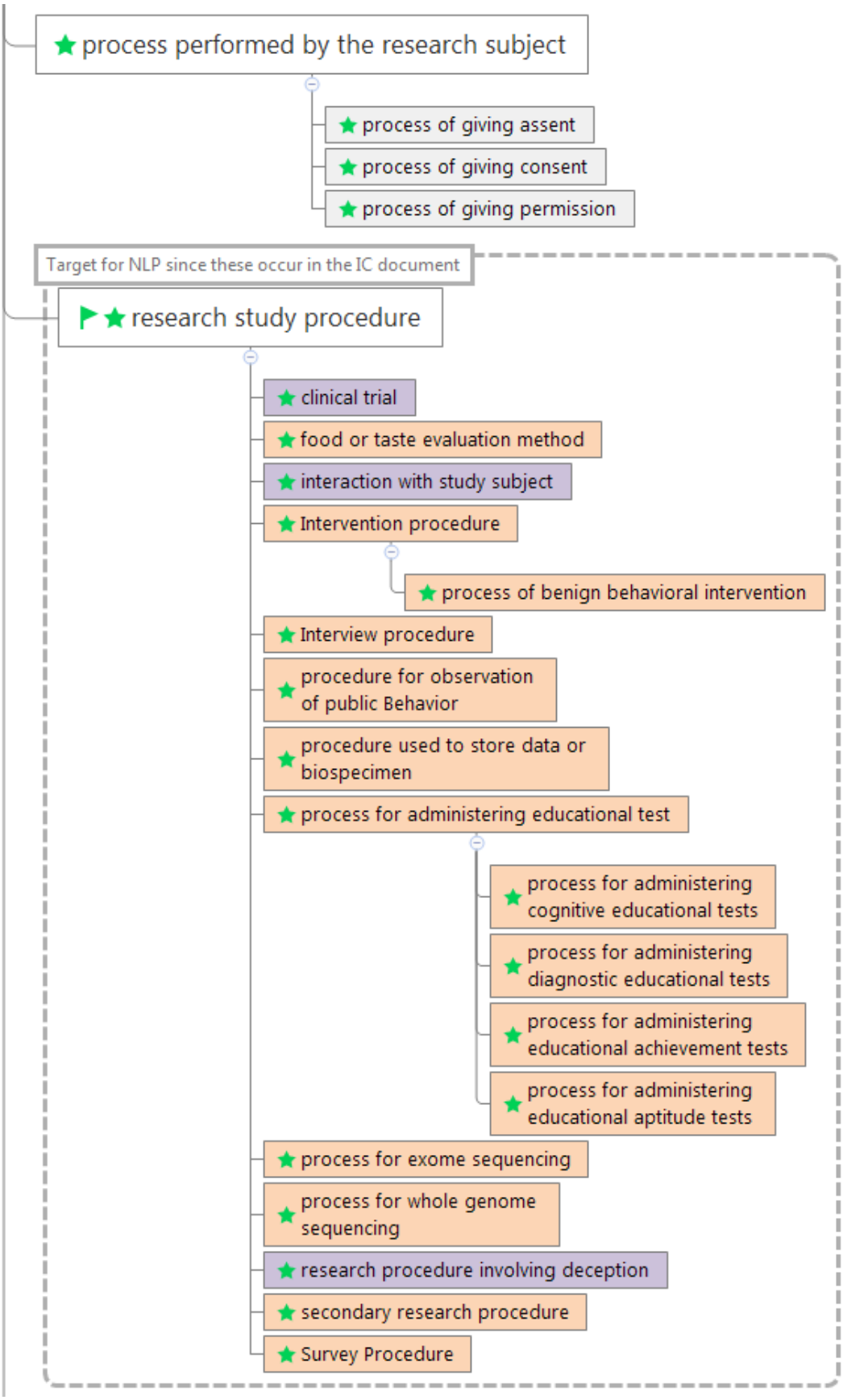
Procedures

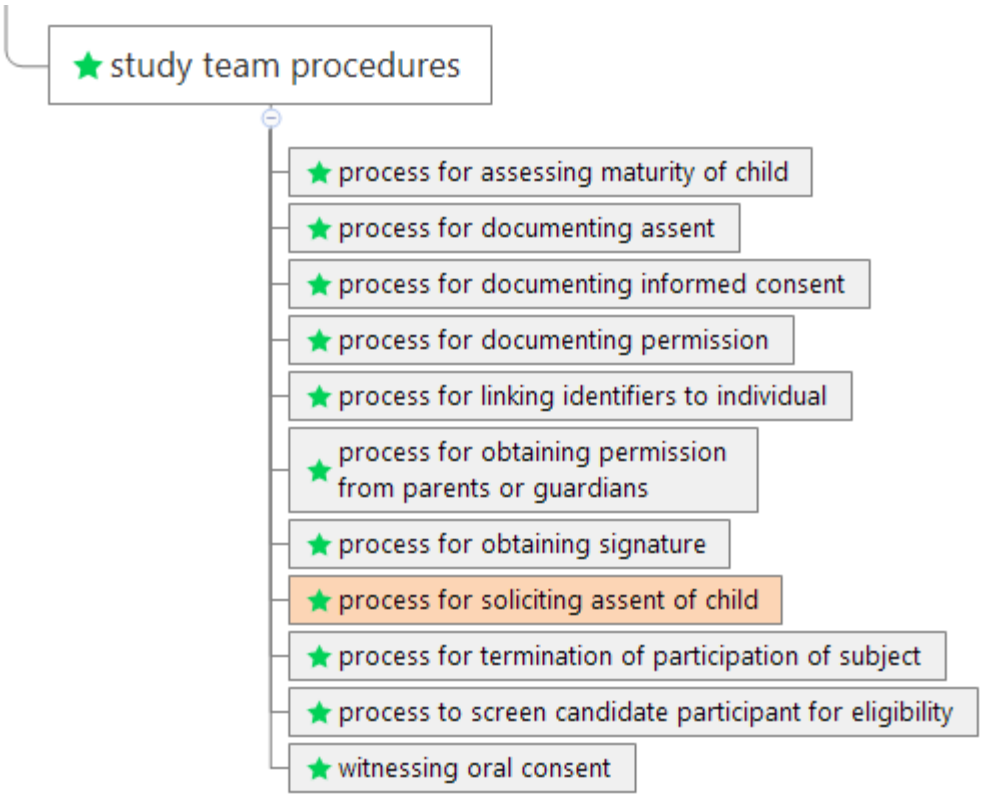

Legend
Checked top level
In spreadsheet
Remaining Question
Wrong
Procedure
administrative procedures
procedure for funding research proposal
process for procedural modifications to protections for children
process for public review and comment
process to determine if policy applies
alternate procedure or course of treatment
Clinical Care Procedure
process to decide if neonate is viable
process of pregnancy termination
Decision making procedure


IRB Procedure
process for assessing risk
process for assessing magnitude of risk
process for assessing risk of breach of confidentiality
process for assessing risk of loss of privacy
process for assessing risk of physical harm
process to decide if parental permission not reasonable to protect child
process to ensure subjects rights are not impacted
process to ensure subjects welfare is not impacted
AL notes these are accurate but might be ad hoc
process of certification
process of IRB certification of research to DHHS
documenting IRB activity
notification procedure
notifying investigator of decisions

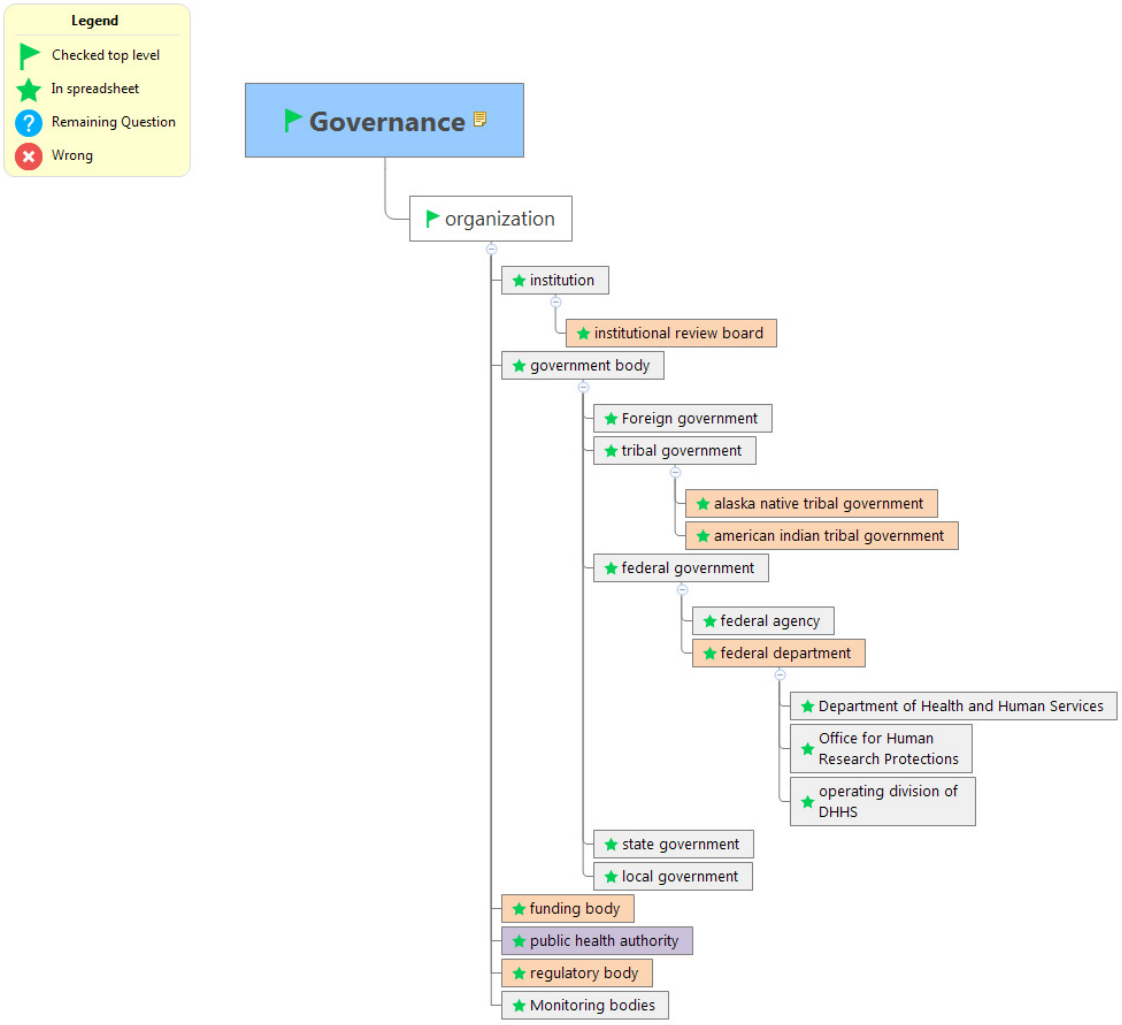

process for assessing benefits
process for assessing enhanced probability of survival
process for assessing magnitude of benefit
process to decide if unapprovable research presents opportunities
process for determining consenting procedure
process to decide if alteration of consent is required
process to decide if assent must be documented
process to decide if child is capable of providing assent
process to decide if consent can be waived
process to decide if research is impractical without alteration or waiver of consent
process to determine if broad consent is allowed
process to determine if consent for determining eligibility is required
process to determine if consent for recruiting is required
process to determine if consent for screening is required
AL wasn't sure if this is there and needed
process to determine if assent can or should be waived
process to determine if permission requirement should be waived to protect child
process to decide the requirements for documenting assent


procedure for reviewing proposal
process for continuing review of research
process for expedited review
process for initial review of research
process for limited review
process to decide if research is exempt
criteria for IRB approval of research
process for ensuring risks to subjects are minimized
process for ensuring risks to subjects are reasonable
process for ensuring that selection of subjects is equitable
process for ensuring that consent is sought
process for ensuring that consent is documented
process for reviewing provisions for data and safety monitoring
process for reviewing privacy and confidentiality provisions
process for ensuring additional safeguards are included for vulnerable populations
already exist as protections

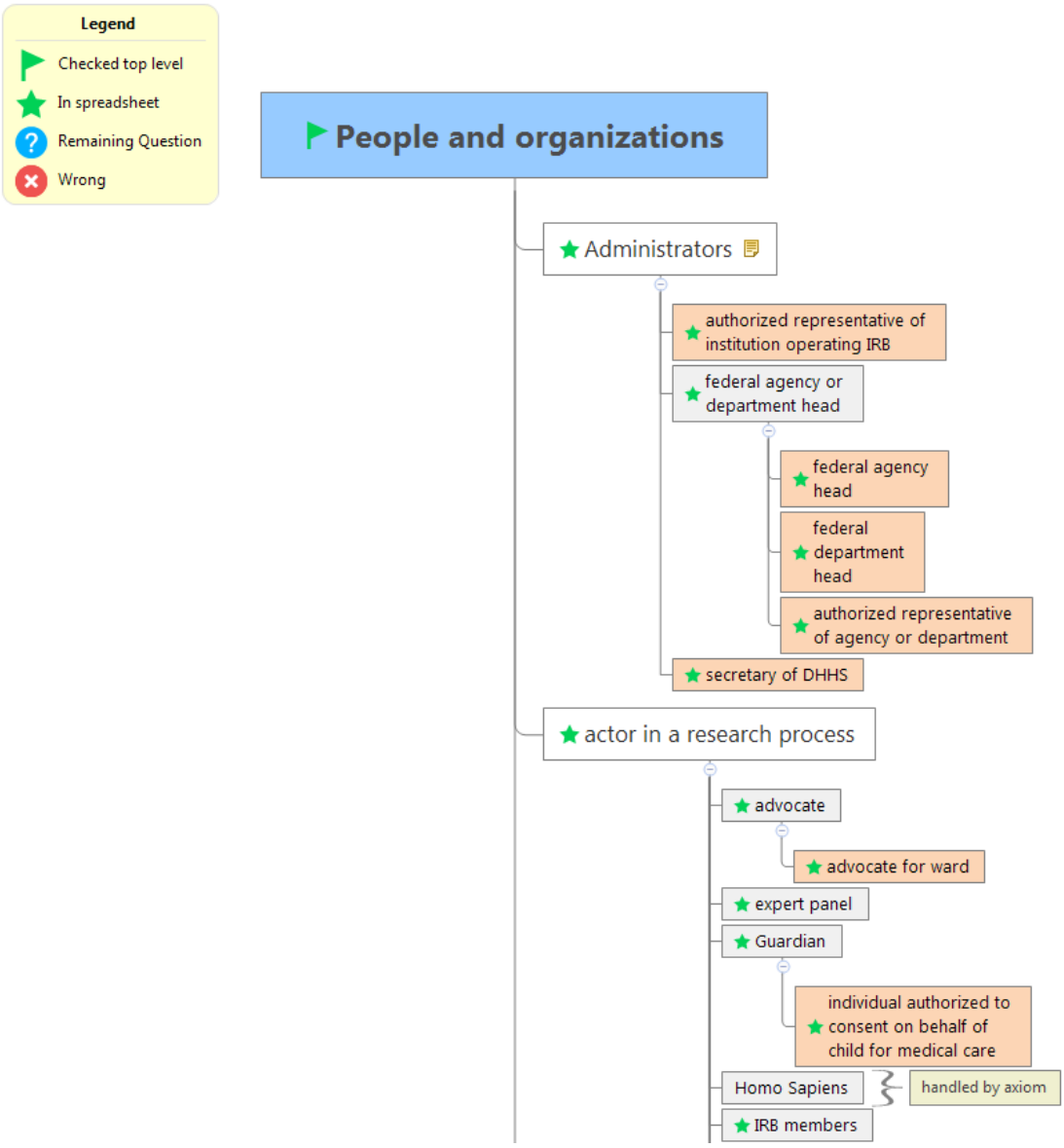

process performed by the research subject
process of giving assent
process of giving consent
process of giving permission
Target for NLP since these occur in the IC document
research study procedure
clinical trial
food or taste evaluation method
interaction with study subject
Intervention procedure
process of benign behavioral intervention
Interview procedure
procedure for observation of public Behavior
procedure used to store data or biospecimen
process for administering educational test
process for administering cognitive educational tests
process for administering diagnostic educational tests
process for administering educational achievement tests
process for administering educational aptitude tests
process for exome sequencing
process for whole genome sequencing
research procedure involving deception
secondary research procedure
Survey Procedure

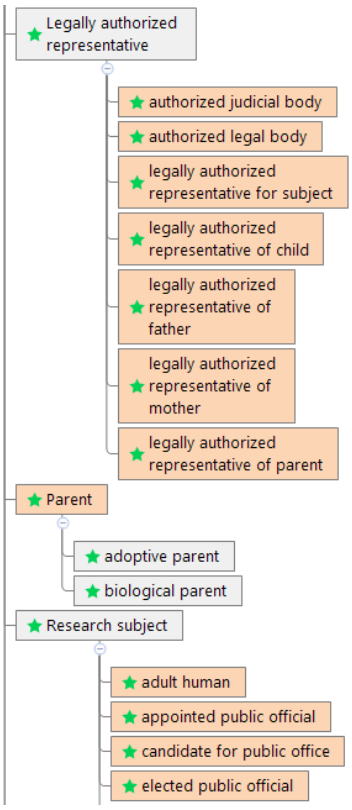

study team procedures
process for assessing maturity of child
process for documenting assent
process for documenting informed consent
process for documenting permission
process for linking identifiers to individual
process for obtaining permission from parents or guardians
process for obtaining signature
process for soliciting assent of child
process for termination of participation of subject
process to screen candidate participant for eligibility
witnessing oral consent

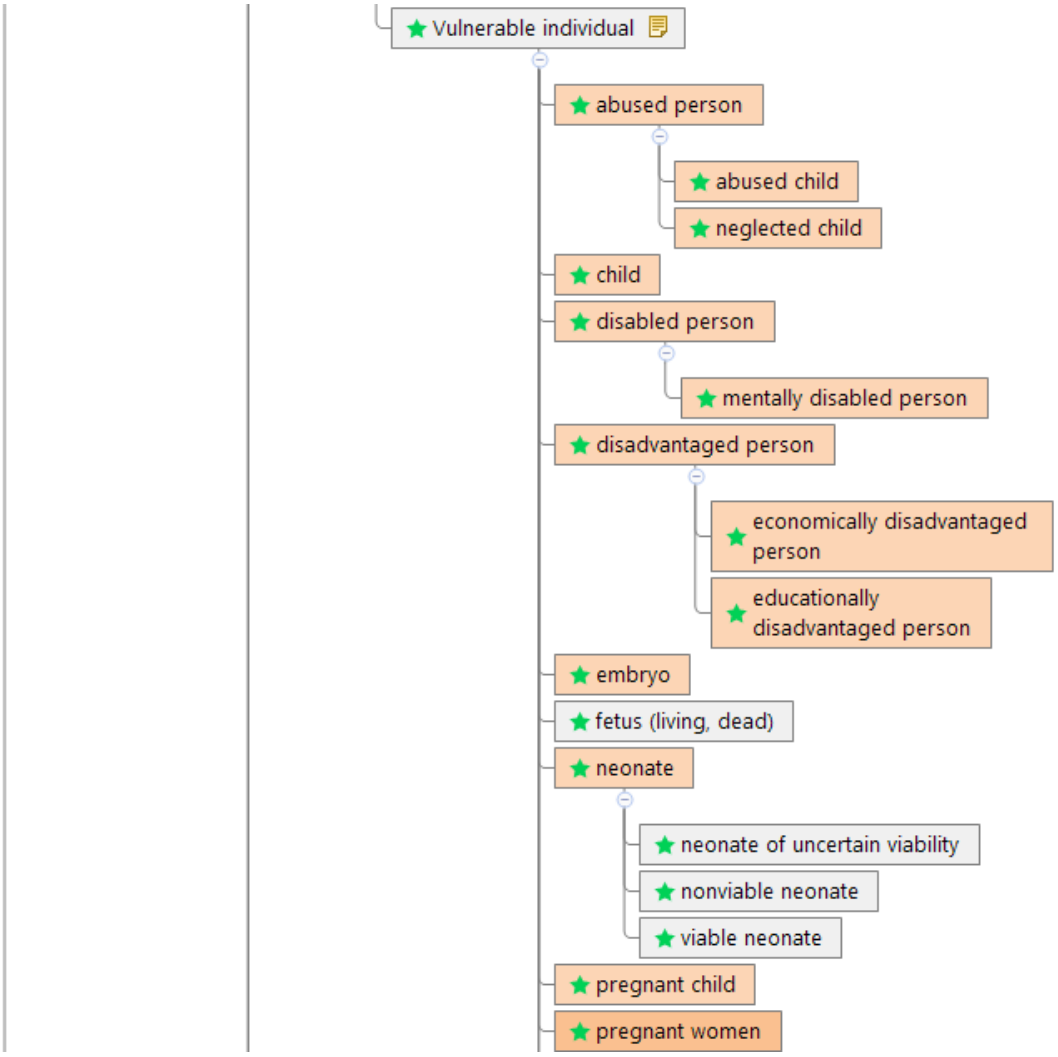
Governance

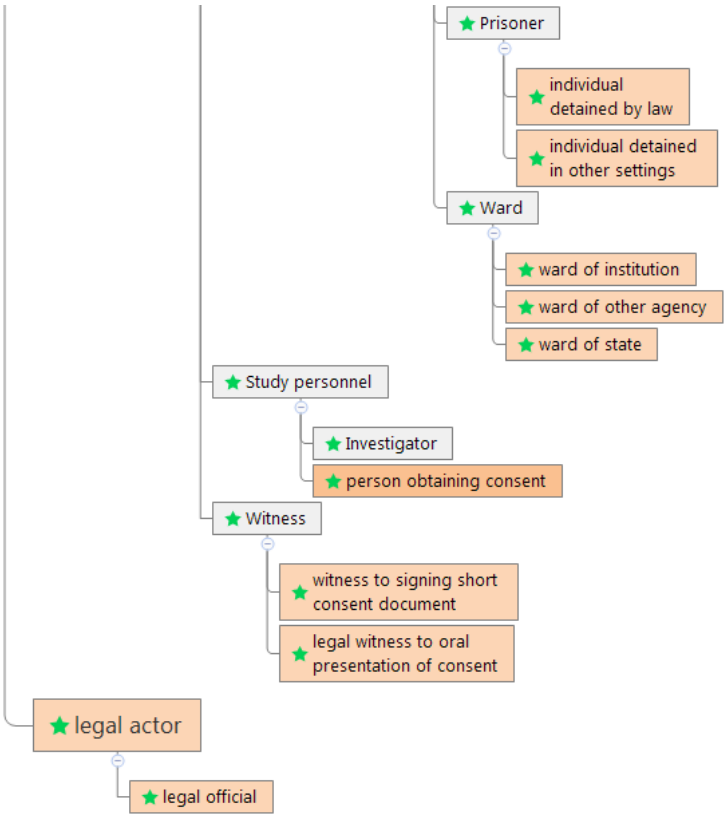

Legend
Checked top level
In spreadsheet
Remaining Question
Wrong
Governance
organization
institution
institutional review board
government body
Foreign government
tribal government
alaska native tribal government
american indian tribal government
federal government
federal agency
federal department
Department of Health and Human Services
Office for Human Research Protections
operating division of DHHS
state government
local government
funding body
public health authority
regulatory body
Monitoring bodies

People and Organizations

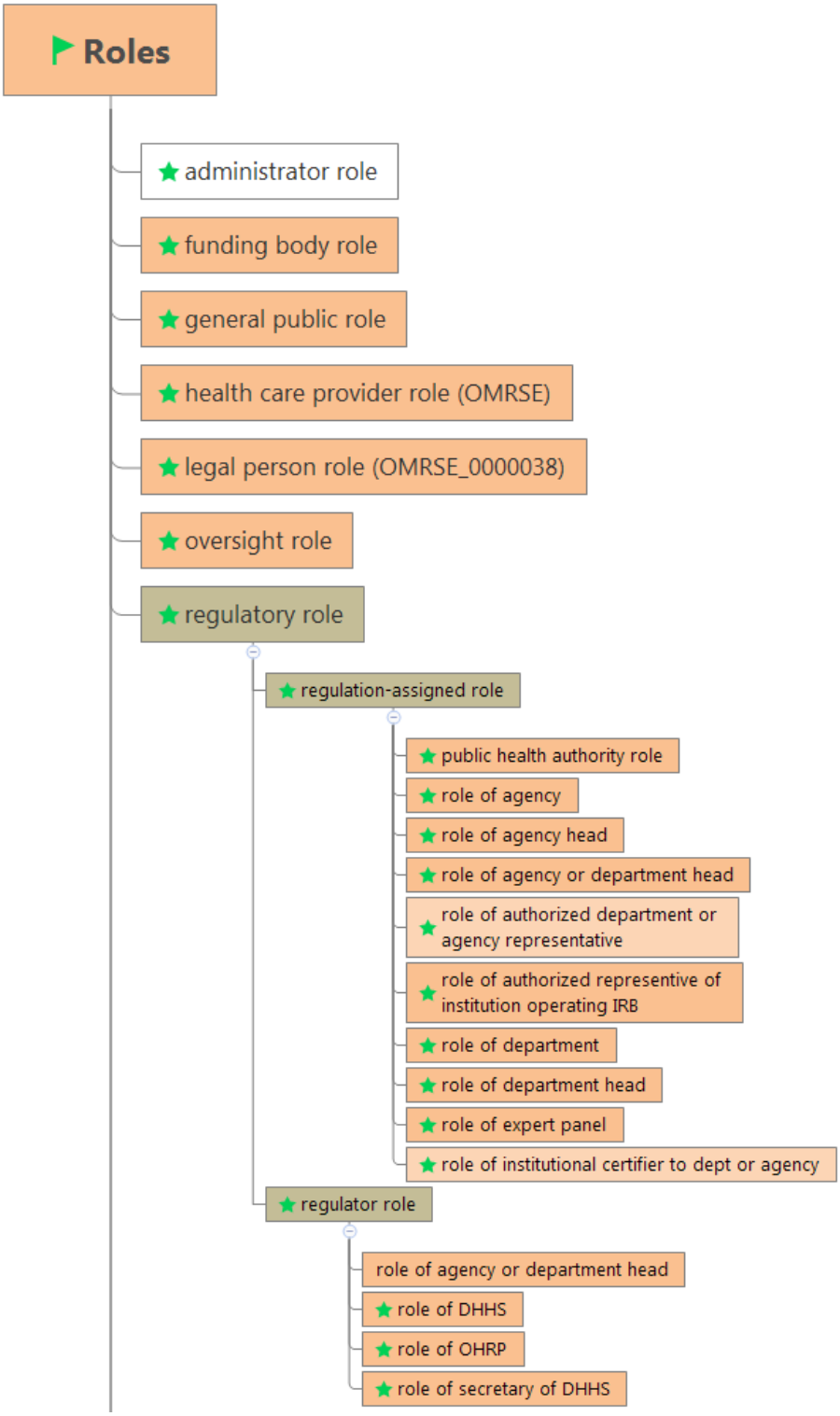


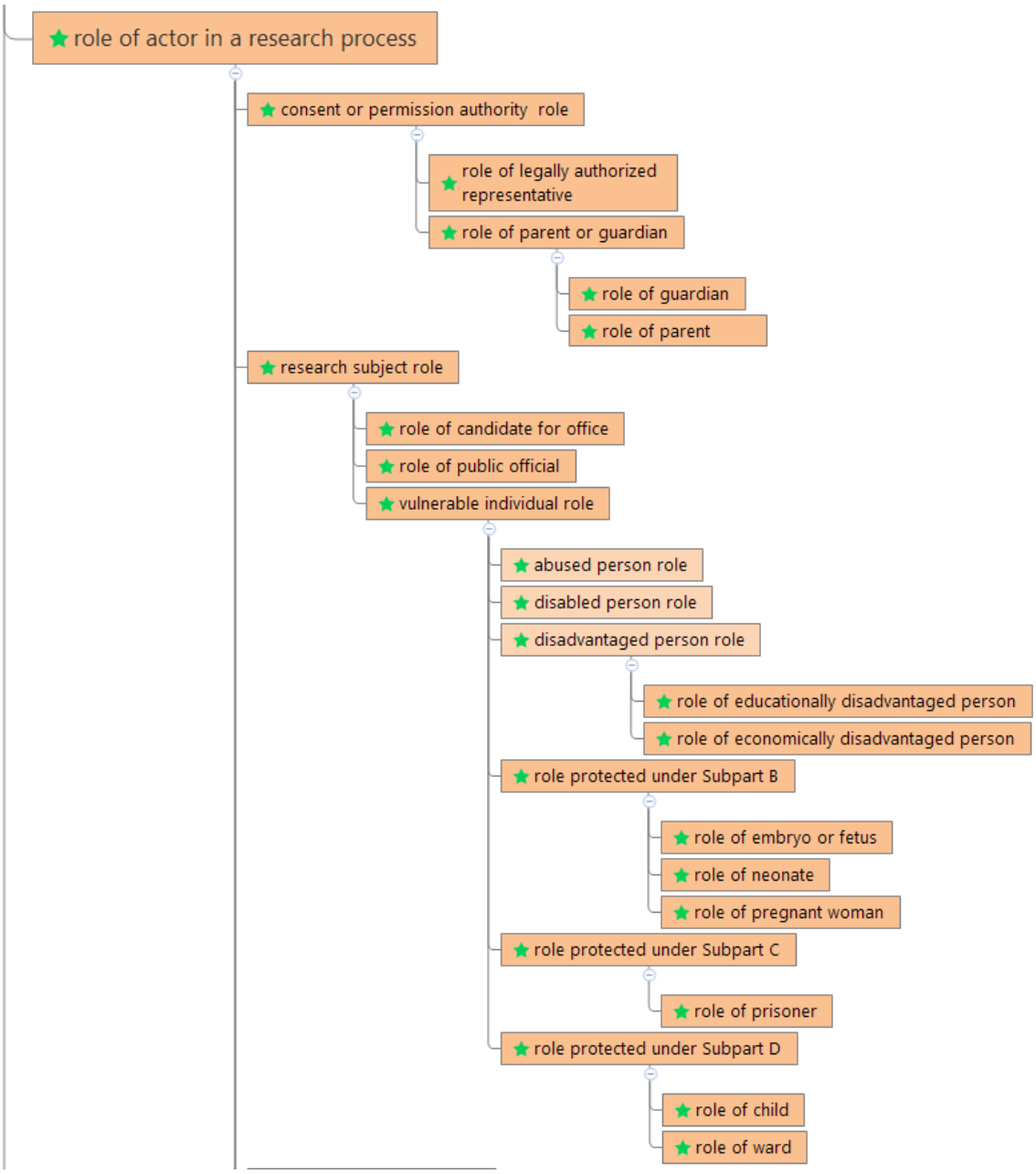

Legally authorized representative
authorized judicial body
authorized legal body
legally authorized representative for subject
legally authorized representative of child
legally authorized representative of father
legally authorized representative of mother
legally authorized representative of parent
Parent
adoptive parent
biological parent
Research subject
adult human
appointed public official
candidate for public office
elected public official

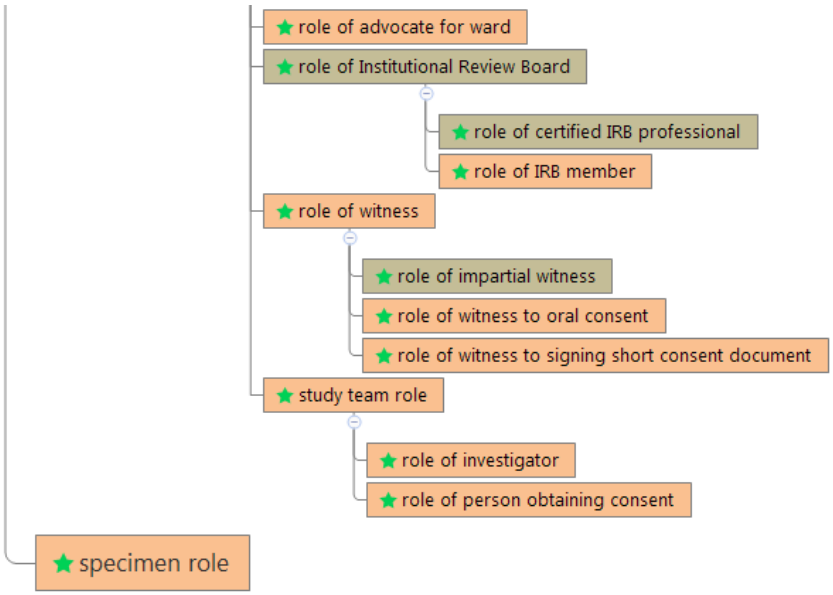

Vulnerable individual
abused person
abused child
neglected child
child
disabled person
mentally disabled person
disadvantaged person
economically disadvantaged person
educationally disadvantaged person
embryo
fetus (living, dead)
neonate
neonate of uncertain viability
nonviable neonate
viable neonate
pregnant child
pregnant women


Prisoner
individual detained by law
individual detained in other settings
Ward
ward of institution
ward of other agency
ward of state
Study personnel
Investigator
person obtaining consent
Witness
witness to signing short consent document
legal witness to oral presentation of consent
legal actor
legal official

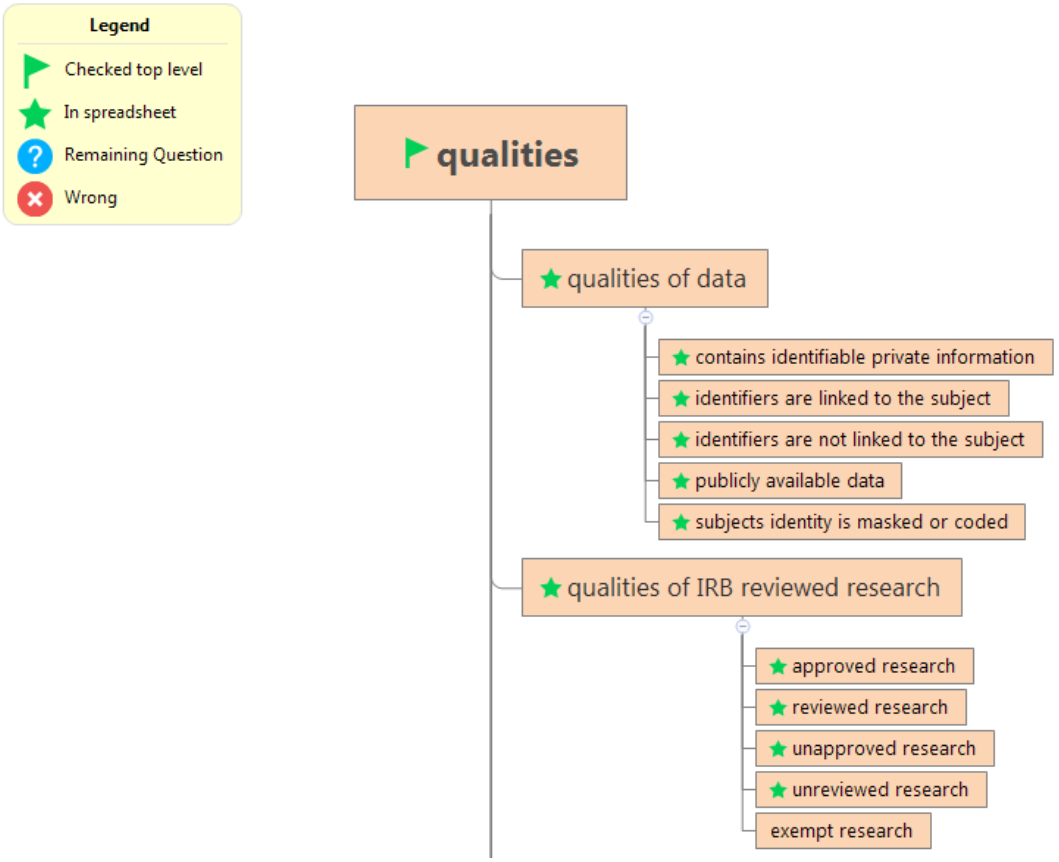
Roles

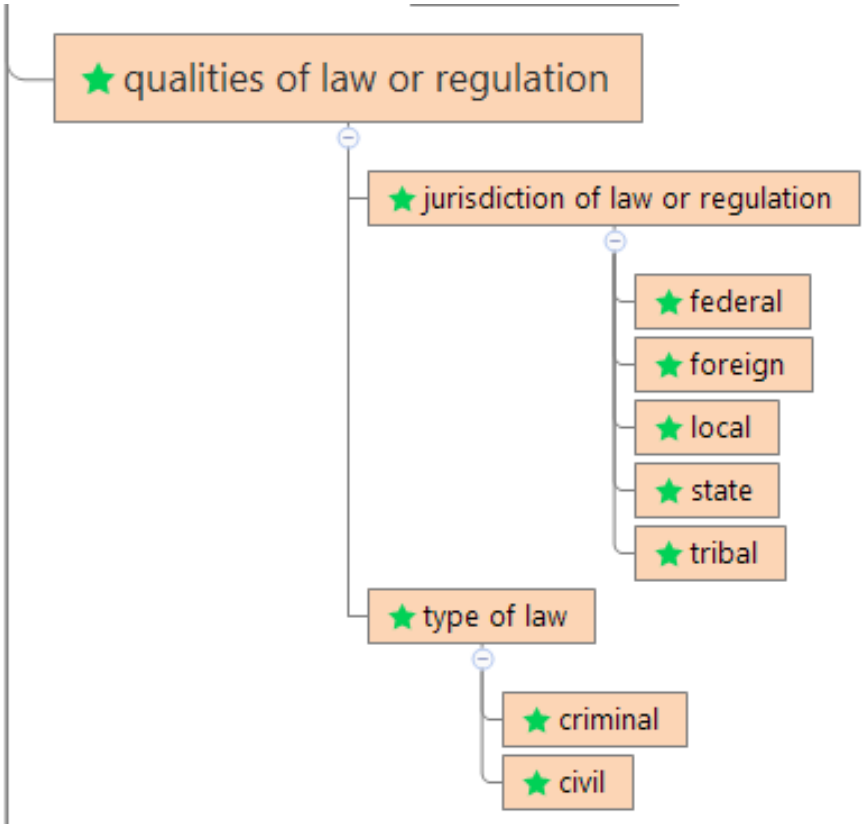

Roles
administrator role
funding body role
general public role
health care provider role (OMRSE)
legal person role (OMRSE_0000038)
oversight role
regulatory role
regulation-assigned role
public health authority role
role of agency
role of agency head
role of agency or department head
role of authorized department or agency representative
role of authorized representive of institution operating IRB
role of department
role of department head
role of expert panel
role of institutional certifier to dept or agency
regulator role
role of agency or department head
role of DHHS
role of OHRP
role of secretary of DHHS

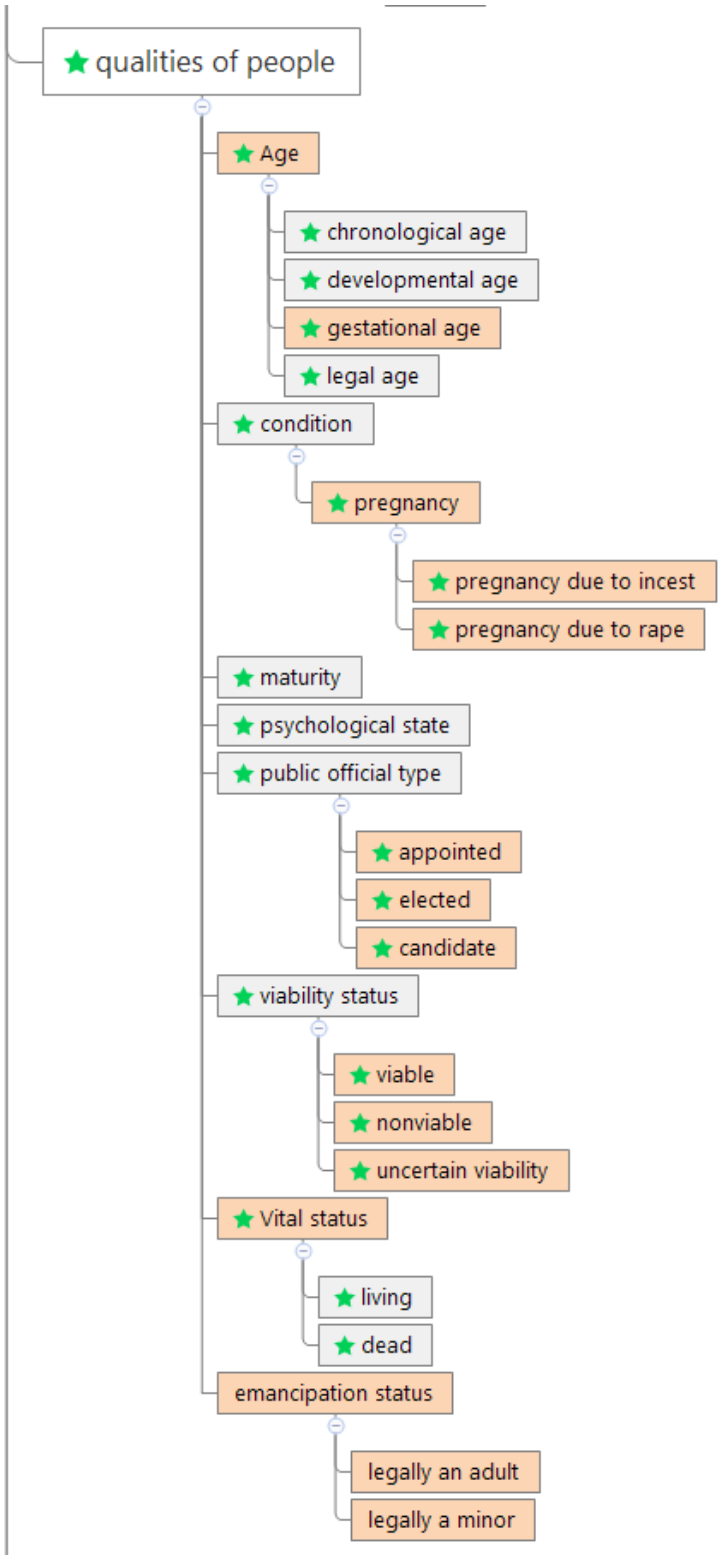

role of actor in a research process
consent or permission authority role
role of legally authorized representative
role of parent or guardian
role of guardian
role of parent
research subject role
role of candidate for office
role of public official
vulnerable individual role
abused person role
disabled person role
disadvantaged person role
role of educationally disadvantaged person
role of economically disadvantaged person
role protected under Subpart B
role of embryo or fetus
role of neonate
role of pregnant woman
role protected under Subpart C
role of prisoner
role protected under Subpart D
role of child
role of ward

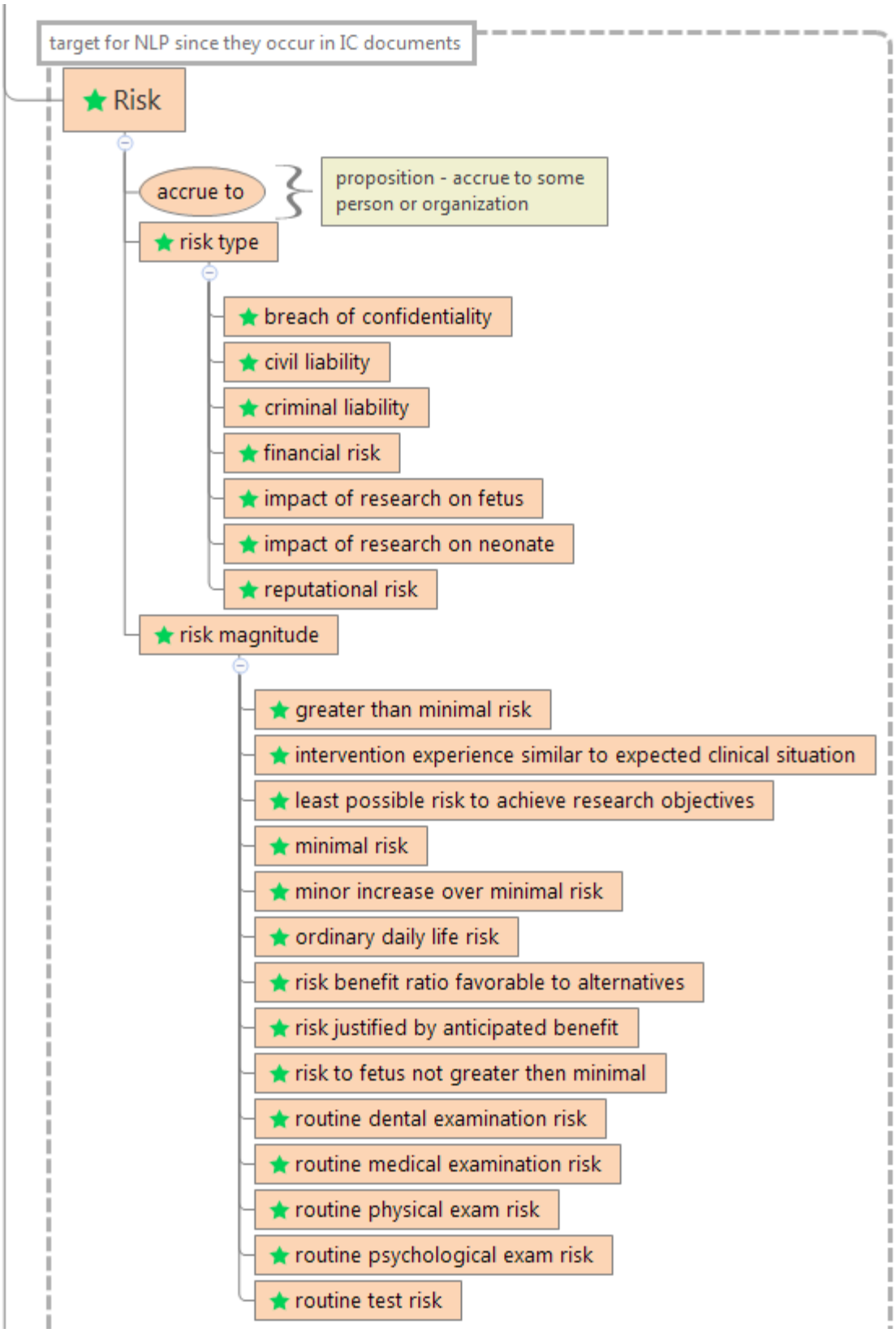

role of advocate for ward
role of Institutional Review Board
role of certified IRB professional
role of IRB member
role of witness
role of impartial witness
role of witness to oral consent
role of witness to signing short consent document
study team role
role of investigator
role of person obtaining consent
specimen role

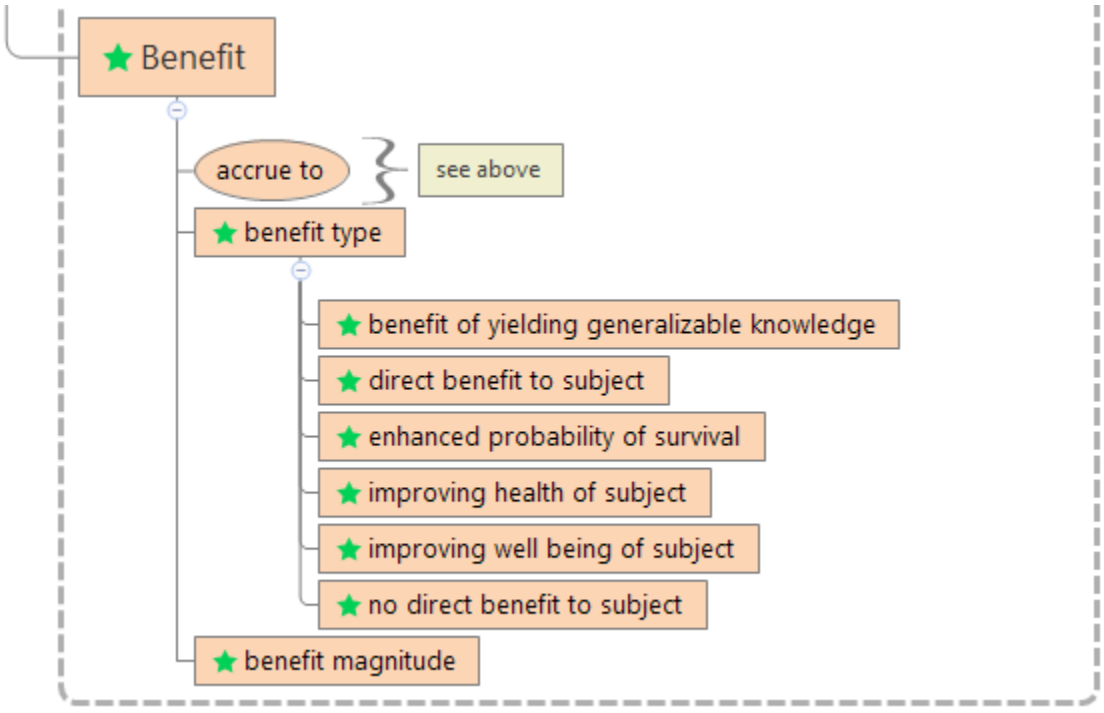
Qualities


Legend
Checked top level
In spreadsheet
Remaining Question
Wrong
qualities
qualities of data
contains identifiable private information
identifiers are linked to the subject
identifiers are not linked to the subject
publicly available data
subjects identity is masked or coded
qualities of IRB reviewed research
approved research
reviewed research
unapproved research
unreviewed research
exempt research

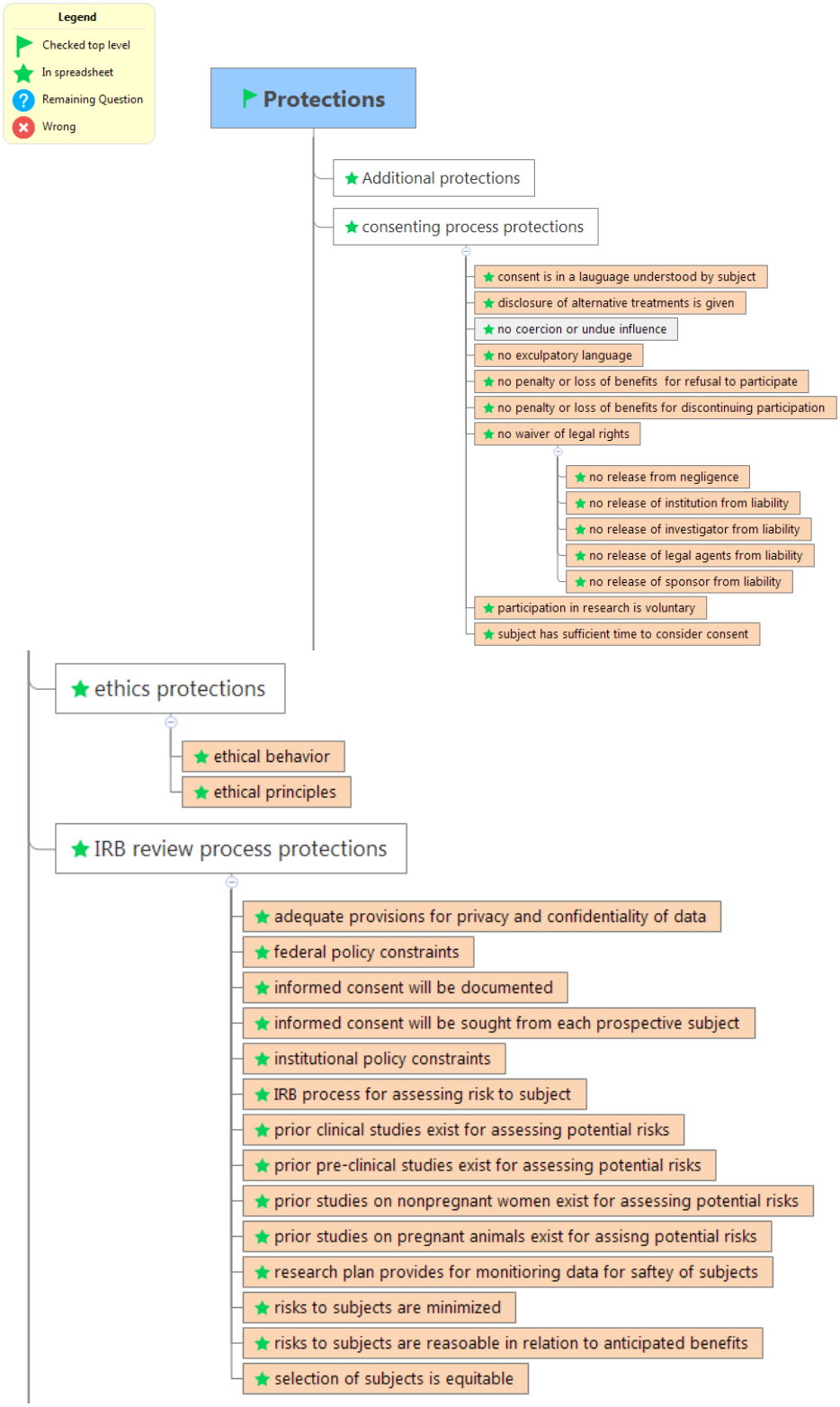

qualities of law or regulation
jurisdiction of law or regulation
federal
foreign
local
state
tribal
type of law
criminal
civil

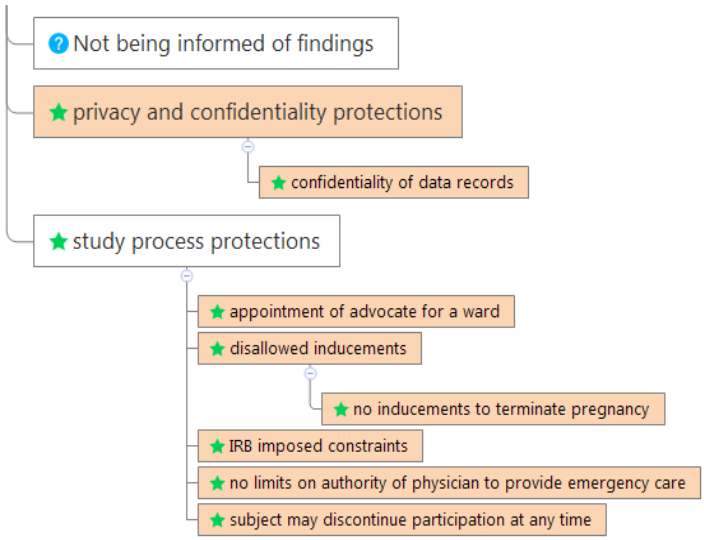

qualities of people
Age
chronological age
developmental age
gestational age
legal age
condition
pregnancy
pregnancy due to incest
pregnancy due to rape
maturity
psychological state
public official type
appointed
elected
candidate
viability status
viable
nonviable
uncertain viability
Vital status
living
dead
emancipation status
legally an adult
legally a minor


target for NLP since they occur in IC documents
Risk
accrue to
proposition - accrue to some person or organization
risk type
breach of confidentiality
civil liability
criminal liability
financial risk
impact of research on fetus
impact of research on neonate
reputational risk
risk magnitude
greater than minimal risk
intervention experience similar to expected clinical situation
least possible risk to achieve research objectives
minimal risk
minor increase over minimal risk
ordinary daily life risk
risk benefit ratio favorable to alternatives
risk justified by anticipated benefit
risk to fetus not greater then minimal
routine dental examination risk
routine medical examination risk
routine physical exam risk
routine psychological exam risk
routine test risk

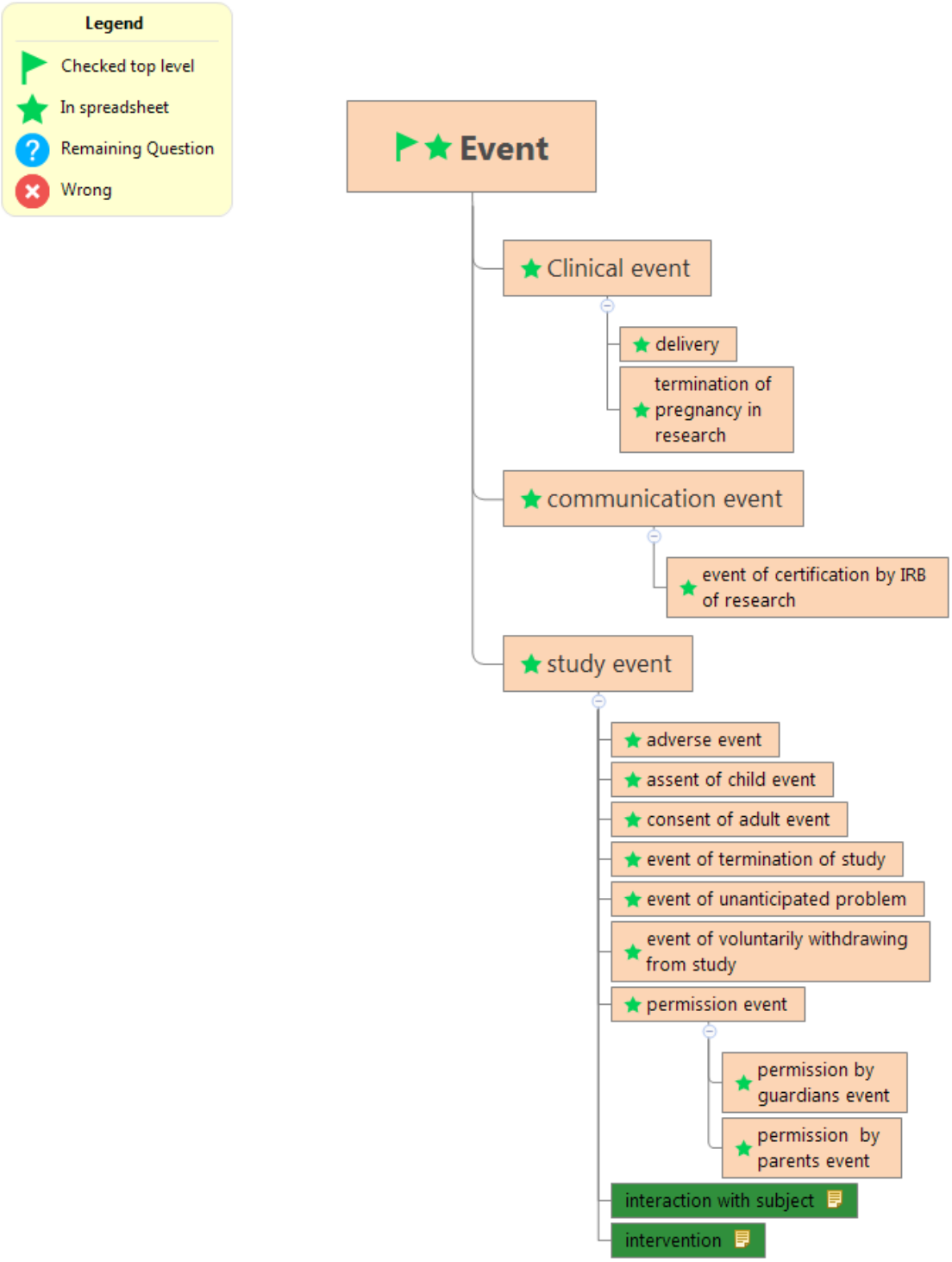

Benefit
accrue to
see above
benefit type
benefit of yielding generalizable knowledge
direct benefit to subject
enhanced probability of survival
improving health of subject
improving well being of subject
no direct benefit to subject
benefit magnitude

Protections

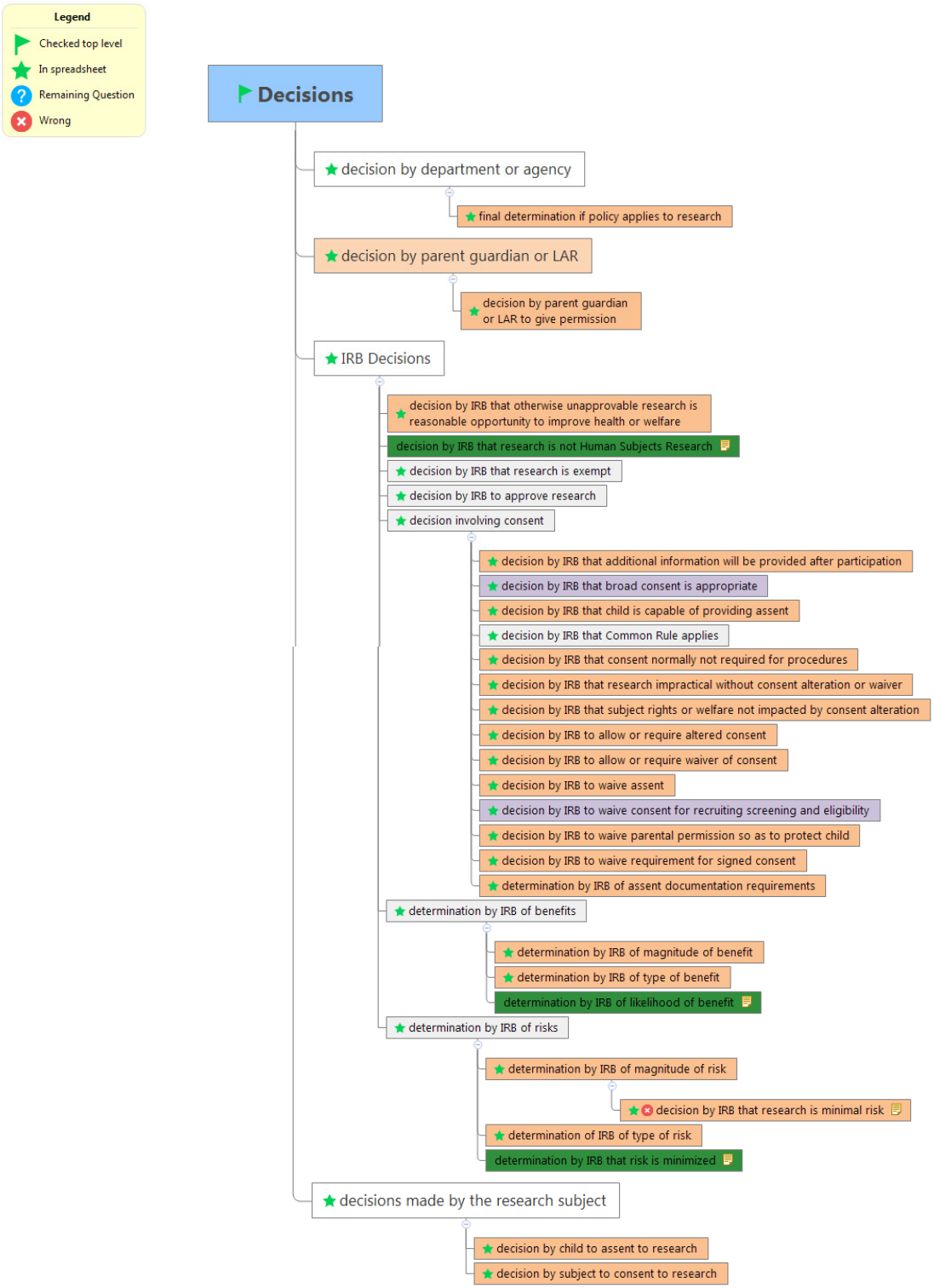



Not being informed of findings
privacy and confidentiality protections
confidentiality of data records
study process protections
appointment of advocate for a ward
disallowed inducements
no inducements to terminate pregnancy
IRB imposed constraints
no limits on authority of physician to provide emergency care
subject may discontinue participation at any time

Events


Legend
Checked top level
In spreadsheet
Remaining Question
Wrong
Event
Clinical event
delivery
termination of pregnancy in research
communication event
event of certification by IRB of research
study event
adverse event
assent of child event
consent of adult event
event of termination of study
event of unanticipated problem
event of voluntarily withdrawing from study
permission event
permission by guardians event
permission by parents event
interaction with subject
intervention

Decisions


Legend
Checked top level
In spreadsheet
Remaining Question
Wrong
Decisions
decision by department or agency
final determination if policy applies to research
decision by parent guardian or LAR
decision by parent guardian or LAR to give permission
IRB Decisions
decision by IRB that otherwise unapprovable research is reasonable opportunity to improve health or welfare
decision by IRB that research is not Human Subjects Research
decision by IRB that research is exempt
decision by IRB to approve research
decision involving consent
decision by IRB that additional information will be provided after participation
decision by IRB that broad consent is appropriate
decision by IRB that child is capable of providing assent
decision by IRB that Common Rule applies
decision by IRB that consent normally not required for procedures
decision by IRB that research impractical without consent alteration or waiver
decision by IRB that subject rights or welfare not impacted by consent alteration
decision by IRB to allow or require altered consent
decision by IRB to allow or require waiver of consent
decision by IRB to waive assent
decision by IRB to waive consent for recruiting screening and eligibility
decision by IRB to waive parental permission so as to protect child
decision by IRB to waive requirement for signed consent
determination by IRB of assent documentation requirements
determination by IRB of benefits
determination by IRB of magnitude of benefit
determination by IRB of type of benefit
determination by IRB of likelihood of benefit
determination by IRB of risks
determination by IRB of magnitude of risk
decision by IRB that research is minimal risk
determination of IRB of type of risk
determination by IRB that risk is minimized
decisions made by the research subject
decision by child to assent to research
decision by subject to consent to research

Spatial Locations