

Using Semantic Web Technologies to Manage FDA Electronic Submissions Policy Information

Mitra Rocca, Ph.D. – CDER Office of Translational Sciences, Armando Oliva, M.D. – CDER Office of Computational Science, Syed Haider, DDS, MPH – CDER Office of Translational Sciences

Abstract

FDA requirements, policies, and preferences for eSubmissions are documented in various sources in formats that are not favorable to support automated eSubmission-creation processes. Semantic Web Technologies (SWT) provide a promising approach to represent eSubmission policy information in a machine-computable format, which we believe will facilitate the creation of high quality eSubmissions that meet FDA requirements and expectations. We are forming a team under the PhUSE Computational Science Symposium Emerging Technologies: Semantic Technologies working group to explore this approach. Interested participants should contact cdercsc@fda.hhs.gov.

Background

Current FDA requirements, policies, and preferences regarding the format and content of electronic regulatory submissions (eSubmissions) are documented in various sources: legislation (e.g. the Food and Drug Administration Safety and Innovation Act), federal regulations (e.g. 21 CFR 312 and 314), guidances (e.g. the electronic Common Technical Document (eCTD) guidance, ICH E3 Structure and Content of Clinical Study Reports, ICH E6 Good Clinical Practice), technical specifications (e.g. the Study Data Specifications document), and web-site content (e.g. the Study Data Standards web page on fda.gov). A submitter wishing to prepare and provide an electronic submission has to navigate, understand, and process a myriad of documents from many sources; some of which are quite technical and challenging to understand. This is a manual, labor intensive, complex process that, understandably, can lead to errors and omissions, decreased eSubmission quality, and inefficient regulatory review. If this information were available in a machine-readable and machine-computable format, then it would be possible for computers to help guide the creation of high quality electronic submissions that meet FDA requirements and expectations. This would benefit both submitters and the Agency.

Semantic Web Technologies (SWT)

Semantic Web Technologies (SWT) enables the linking of data and their meaning across multiple sources in a machine-readable format so computers can perform increasingly sophisticated tasks and queries on our behalf. The word *semantic* itself implies *meaning or understanding*. As such, the fundamental difference between Semantic Web technologies and other technologies related to data (such as relational databases or the World Wide Web itself) is that the Semantic Web is concerned with the *meaning* and not the structure of data. SWT consist of a family of technology standards that “play nicely together” and provide: a flexible data model, an expressive ontology language, and a distributed query language, which together drive data integration, business intelligence, and the creation and maintenance of large knowledge bases. When data are modeled and represented correctly using semantic web technology standards, computers can analyze data in ways not previously possible or practical.

The existing SWT standards include, but are not limited to:

- **RDF** (the Resource Description Framework language)
RDF is a formal language for describing structured information and is a foundation for processing metadata. The goal of RDF is to enable applications to exchange data on the Web while still preserving their original meaning.
- **RDFS** (the Resource Description Framework Schema language)
Is a set of classes with certain properties using the RDF extensible knowledge representation language, providing basic elements for the description of ontologies, otherwise called RDF vocabularies, intended to structure RDF resources.
- **OWL** (the Web Ontology Language)
The schema language, or knowledge representation (KR) language, of the Semantic Web. OWL enables you to define concepts composably so that concepts can be reused as much and as often as possible
- **SPARQL** (the SPARQL Protocol and RDF Query Language)
The query language of the Semantic Web. It is specifically designed to query data across various systems.

Using SWT to Manage FDA eSubmissions Policy Information

If we can represent FDA electronic submissions requirements, policies, and preferences using SWT technologies, we hope to enable end users to perform useful queries such as:

- *What is CDER’s currently supported format(s) to submit Case Report Forms?*
- *Which study data standards does CBER currently support?*
- *Which study data standards does CDER require?*

If future electronic submissions were also to be expressed using SWT, then it may be possible one day to enable the “holy grail” of useful eSubmission queries, namely:

- *Does my electronic submission conform to existing FDA eSubmission policies?*

Initial Exploration

To explore the use of SWT for this use case, and to familiarize ourselves with this technology, we attempted to represent the FDA Data Standards Catalog information in RDF/OWL. The Catalog contains a description of the study data standards that CBER, CDER, and CDRH supports for study data eSubmissions, along with additional useful information on their use¹. Our first step was to model the information using a Mind Map. This allowed us to express, in a graphical form, the key concepts and relationships that are explicitly and implicitly provided in the catalog. We found this facilitated the creation of OWL classes and properties. The following two figures illustrate the possible use of mind maps for this activity. To create the ontology, we used Protégé, an open source ontology editor freely distributed by Stanford University (see Reference #5). Once the ontology was complete, we added individuals (e.g. data standards, implementation guides) and made known assertions about the individuals. This preliminary work demonstrated the ability to represent FDA data standards catalog information using SWT. The utility of this representation remains to be demonstrated.

Further Exploration

We would like to continue to explore SWT for our use case in a collaborative environment with other stakeholders. To facilitate this collaboration, we are forming a team under the PhUSE Computational Science Symposium Emerging Technologies: Semantic Technologies working group. The intent is to help maximize the utility of SWT representation of eSubmissions policy information for all interested.

Since the volume of eSubmissions regulatory policy information is large. One of the goals of this collaboration is to prioritize the available information as we move forward. The team will then work to model the information in a collaborative and iterative fashion. The team will also explore solutions to demonstrate the value of representing this information using formal semantic standards.

Getting Involved

We are actively seeking interested individuals to join this team. Please contact cdercsc@fda.hhs.gov.

References

1. <http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>
2. D. Allemang, J. Hendler, Semantic Web for the Working Ontologist, Morgan Kaufmann, 2008
3. P. Hitzler, M. Krötzsch, S. Rudolph, Foundations of Semantic Web Technologies, CRC Press, 2010
4. E. Prud’hommeaux, A. Seaborne, SPARQL Query Language for RDF W3C Recommendation. March 2013, Available at <http://www.w3.org/TR/rdf-sparql-query/>
5. Protégé, Stanford Center for Biomedical Informatics Research, <http://protege.stanford.edu>

Figure 1: Mind Map of FDA Data Standards Catalog Information

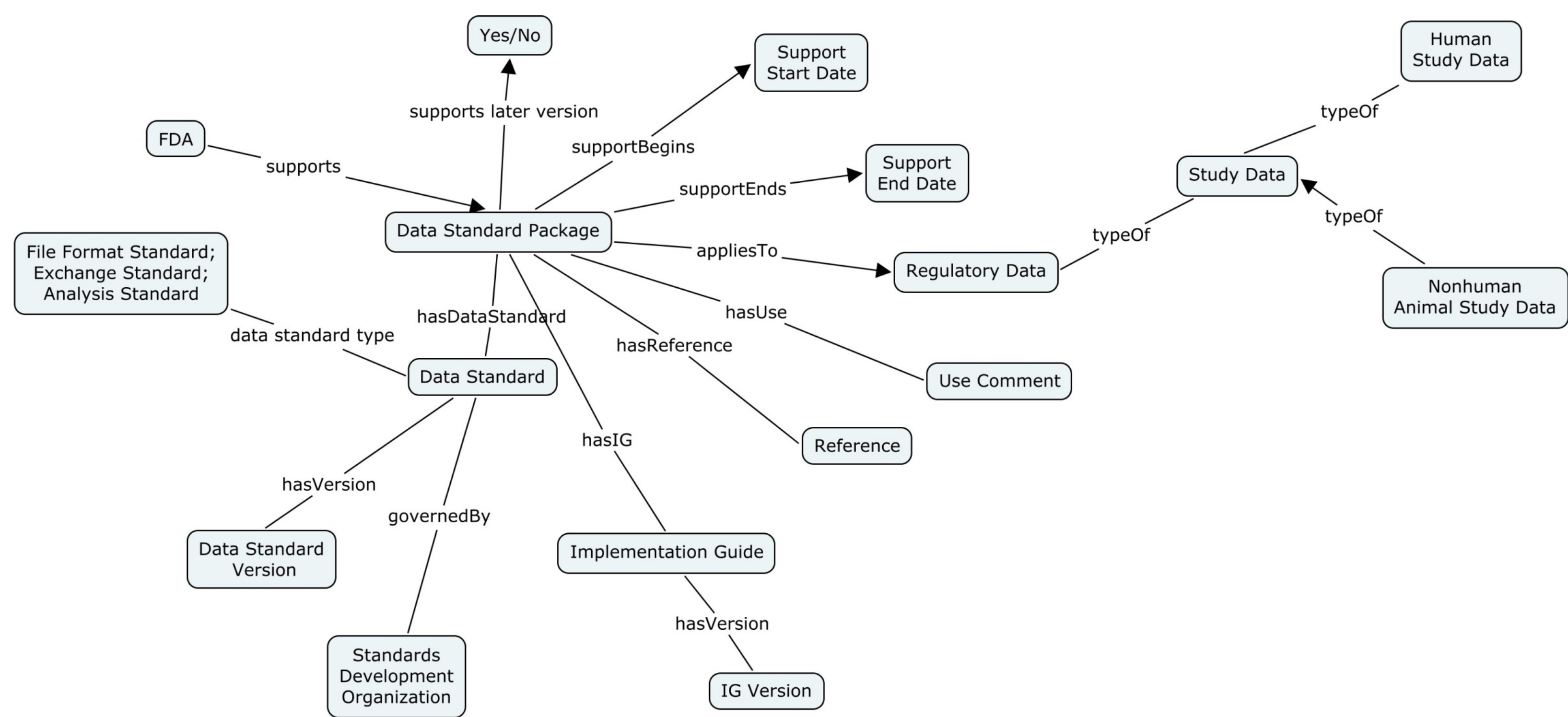


Figure 2: Mind Map of General FDA eSubmission Policy Information

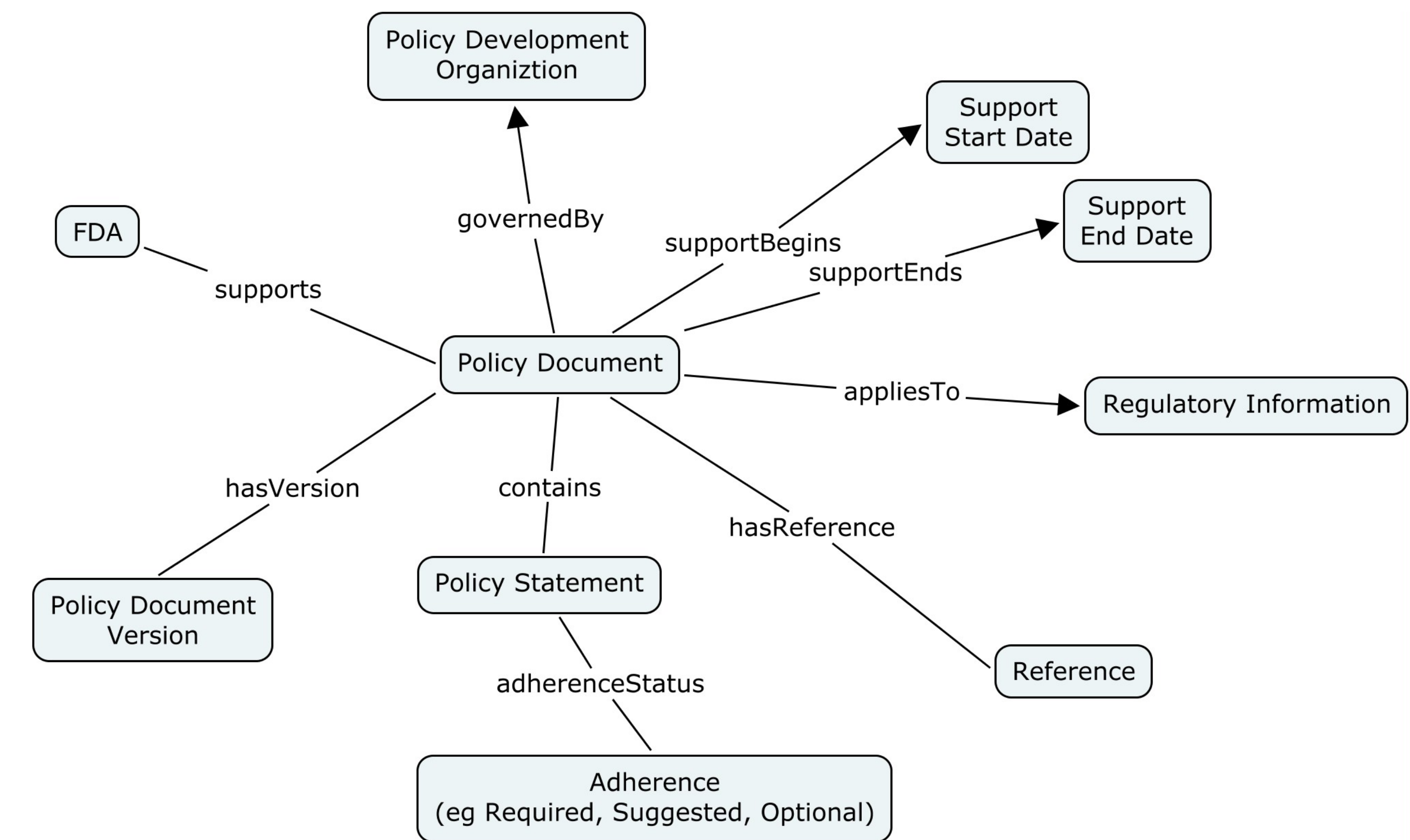


Figure 3: Mind Map of ICH and FDA Regulations and Guidances

