

Increasing Patient Findability of Medical Research: Annotating Clinical Trials Using Standard Vocabularies

by Michael Panzer

Information Standards

EDITOR'S SUMMARY

Multiple groups at Mayo Clinic organize knowledge with the aid of metadata for a variety of purposes. The ontology group focuses on consumer-oriented health information using several controlled vocabularies to support and coordinate care providers, consumers, clinical knowledge and, as part of its research management, information on clinical trials. Poor findability, inconsistent indexing and specialized language undermined the goal of increasing trial participation. The ontology group designed a metadata framework addressing disorders and procedures, investigational drugs and clinical departments, adopted and translated the clinical terminology of SNOMED CT and RxNorm vocabularies to consumer language and coordinated terminology with Mayo's Consumer Health Vocabulary. The result enables retrieval of clinical trial information from multiple access points including conditions, procedures, drug names, organizations involved and trial phase. The jump in inquiries since the search site was revised and vocabularies were modified show evidence of success.

KEYWORDS

biomedical information
ontologies
metadata
subject indexing
information retrieval
findability
access points

Michael Panzer is the former editor-in-chief for the Dewey Decimal System and chief ontologist for the Mayo Clinic. He can be reached at michael.preuss@atgmail.com.

Among industries that rely heavily on the use of terminology management in a broad sense, health care, along with libraries and financial institutions, might be facing the toughest challenges in developing appropriate strategies for the use of vocabularies or ontologies to organize its knowledge.

One reason is the variety of knowledge assets encountered in a clinical environment, which goes far beyond the classical paradigm of the text-based document or bibliographic resource. Rather, medical institutions like Mayo Clinic are confronted with the need to transform these assets (from symptom lists over care process models to clinical decision rules) into actionable artifacts that transport a specific standard of patient care in a way that supports physicians and patients alike in their shared decision making processes.

Which role can metadata play in operationalizing this knowledge? At Mayo, several groups are involved at the same time in different forms of knowledge representation, for example, clinical knowledge management with its focus on physician support or medical informatics with its focus on automated data and natural language processing (in addition to data governance and other data standardization efforts).

In contrast, the ontology group (by taking a more LIS-centric approach) collaborates closely with content generators inside Mayo, such as editors of consumer-oriented health information for the mayoclinic.org website. A specific strength of the group is the application and curation of a variety of controlled vocabularies (from value lists to ontologies) in an advanced environment using semantic technology. Through directly working with editorial content, the annotation workflow informs the ways the underlying knowledge standards evolve. While other groups are working with unmediated clinical data, the expertise in metadata design and implementation of standard vocabularies unique to the ontology group oftentimes serves as

connective tissue between research, physicians, consumers and clinical knowledge assets.

One specific example for the role of standards and metadata framework design in clinical information management is the reworking of the way clinical trials are published on the Mayo Clinic websites. The overhaul of clinical trial management was part of a larger initiative to improve comprehensive research management, but in this article I focus specifically on issues of interoperability and findability addressed by implementing standard value vocabularies, leveraging a metadata element set derived from domain modeling.

Publication of Clinical Trials – Status Quo Ante

What is a clinical trial (or, more precisely, a clinical study; I am using the two terms interchangeably from this point on)? According to the definition of the National Institutes of Health, “clinical trials are research studies that test how well new medical approaches work in people. Each study answers scientific questions and tries to find better ways to prevent, screen for, diagnose or treat a disease.” Recruitment of participants becomes crucial for the success of clinical studies. A smaller sample size than optimal based on the study design usually results in less reliable scientific outcomes. Recruiting participants, of course, requires that people are able to find appropriate studies in the first place, preferably based on their conditions or health interests.

Indeed, one of the key challenges for clinical research institutions is recruiting and retaining participants in clinical trials and other research studies. One of the main goals of redesigning the publication process and website for clinical trials at Mayo Clinic by leveraging standard vocabularies is to increase trial participation.

Several key weaknesses were identified in the legacy process for recruitment, which lead to a situation where not only prospective participants had trouble finding trials, but also study coordinators. The system was mainly browse-focused, based on tags assigned in an ad hoc manner by IT teams. The lack of consistency in breadth and depth of indexing was aggravated by the trial summaries being written for specialists, not including participants as one of the audiences. At some point, MeSH (Medical Subject Headings, edited by the National Library of Medicine) was

adopted as a standard source for medical terms, yet it was only applied in a fairly limited, lexical way (using the preferred headings only).

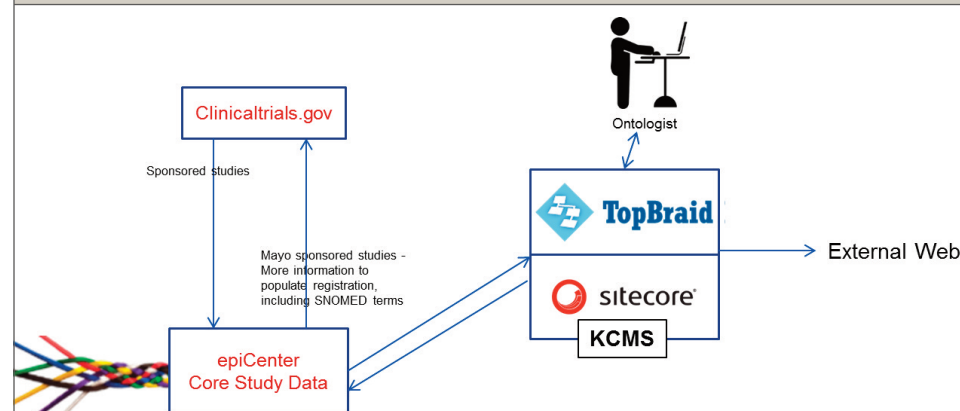
Excursion: Standard Metadata Needs Robust and Reliable Data

A prerequisite to annotating clinical trials with standard metadata was shifting the workflow of capturing core study data to include the ontology group’s TopBraid environment (called semantic services environment, or SSE). Together with Sitecore as the content management system, these two components form an integrated KCMS (knowledge content management system). Without going into more detail here, core study data from various sources such as clinicaltrials.gov is integrated by epiCenter (a study protocol information system), which sends an appropriate subset of such studies to KCMS, that is, Sitecore and SSE. An ontologist in the workflow annotates the clinical trial before the annotated catalog item gets published to the web in various ways (see below). Also, the annotations are then shared back to epiCenter and also, for Mayo-sponsored studies, all the way back to the original registration at clinicaltrials.gov (see Figure 1).

Designing a Metadata Framework for Clinical Trials

The metadata framework relies on the development of a clinical study domain model/ontology, as well as on the selection of candidate clinical

FIGURE 1. Mayo Clinic work flow for annotating clinical trials



PANZER, continued

vocabularies. Clinical trials are conceptualized as a subset and extension of a larger domain model encapsulating a broad view of the entities and relationships involved in Mayo research as a whole, as represented by the research web. The research web domain model tries to account for all entities that play a role in the domain such as Person, Organization (with a subclass for Department, etc.), Location and Information Resource (with a subclass for Publication).

The annotation design, on the other hand, attempts to be much more specific, as you can see in Figure 2. The class of clinical trials is at the center of the design. The relationships can be grouped into three broad categories: Which disorder(s) and medical procedure(s) is the study investigating? Which drug will be investigated as an intervention? Which clinical departments are conducting the research study?

In order to convey the first two categories in a standardized way, SNOMED CT and RxNorm, respectively, were selected as value vocabularies. SNOMED CT is a comprehensive clinical health terminology with more than 300,000

concepts, governed by an international body from 28 member countries. SNOMED CT allows us to capture the investigated condition and intervention at a very granular level and also provides for post-coordination of concepts and inference of properties. For drug intervention, RxNorm as a vocabulary provides normalized names for clinical drugs, but also links to many other drug vocabularies.

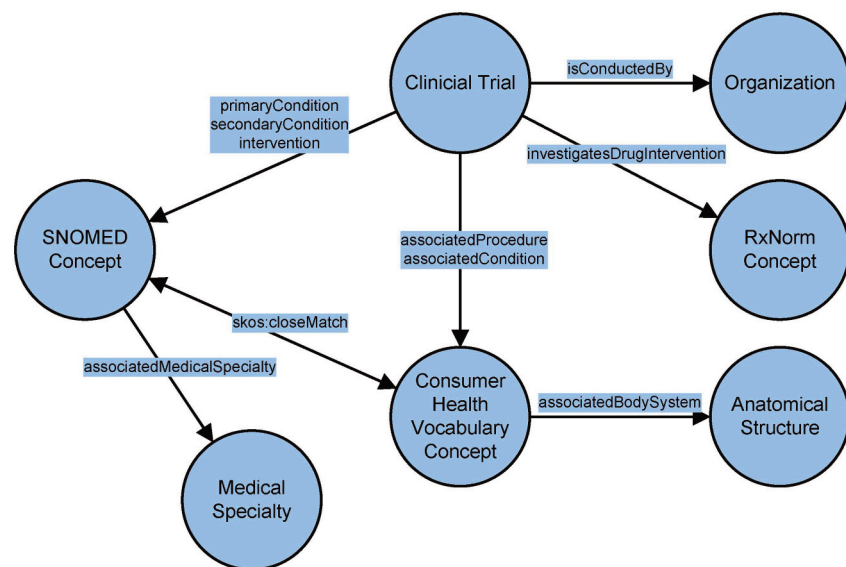
Obviously, much more needs to be specified to ensure consistent application of vocabulary terms. Such rules include that the primary condition captures the primary topic of the study, while the secondary condition may capture a condition of the population being studied if different from the primary condition, for example, diabetic neuropathy in patients with untreated diabetes. Both are constrained to concepts from the UMLS semantic group of Disorder.

Both SNOMED CT and RxNorm alone, while providing the appropriate features to code clinical concepts in an interoperable manner, do not address the problem of translating the clinical idiom into terms used by consumers of health information and, by extension, increasing findability of studies by prospective participants. RxNorm provides generic and brand names of drugs, which helps in that regard. SNOMED CT concepts contain a rich set of synonyms, which also helps closing the jargon gap.

As a third vocabulary, the Mayo-curated Consumer Health Vocabulary (CHV) was included in the design to address some of these findability issues. CHV is a fairly compact SKOS-based scheme of consumer-oriented concepts of conditions, procedures, symptoms, devices and human anatomy (~5000 concepts with a rich set of relationships connecting symptoms, diagnoses and treatments). Augmenting SNOMED CT with CHV also allows for a tighter integration with health information and patient education content on mayoclinic.org, most of which is already annotated with CHV concepts, whereas SNOMED CT allows for interoperability with clinical content.

The ontology group already curates a mapping of SNOMED CT to CHV, which is leveraged in the design to derive the relationships “associatedProcedure” and “associatedCondition.” The connection to CHV also allows acquiring related body systems to further enrich the annotation with relevant terminology.

FIGURE 2. Entity-relationship diagram of the clinical trial annotation domain



The value for the third category of properties is selected from a list of organizations curated as part of the larger research web ontology.

In summary, the design increases findability by allowing search for conditions and procedures (based on SNOMED CT and CHV, including synonyms), for drug names (brand and generic, based on RxNorm), for associated organizations (for example, clinical departments) and body systems. Figure 3 shows an example of a clinical trial instance (i.e., a catalog item) with most properties present.

Interoperability, Findability

As we have seen, SNOMED CT as a vocabulary provides the access points from a clinical standpoint and is thus closely aligned with the language of medical research. But to be able to increase visibility and get studies in the path of potential participants visiting the websites, CHV serves as the main access point. Figure 4 gives an overview of the various locations on Mayo websites

FIGURE 4. Overview of various Mayo web locations at which clinical studies show up automatically based only on shared annotations

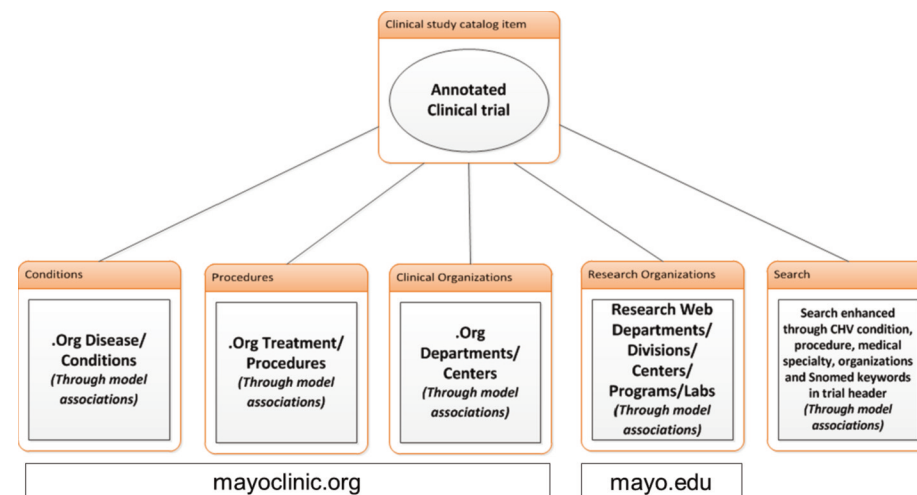


FIGURE 3. A clinical trial catalog item

Colchicine Versus Placebo in Post-Cardiac Surgery Patients to Prevent Post-Pericardiotomy Syndrome and Atrial Fibrillation (Clinical Trial)
http://KCMS.mayoclinic.org/RWV#ColchicineVersusPlaceboinPostCardiacSurgeryPatients

Labels and Definition

preferred label: Colchicine Versus Placebo in Post-Cardiac Surgery Patients to Prevent Post-Pericardiotomy Syndrome and Atrial Fibrillation
type: [ClinicalTrial](#) →

Standard Relationships

has broader: [Clinical Trial Concept](#) →

Custom Properties

IRB_Number: 13-008553
NCTID: NCT02177266
has Associated Condition: [Atrial fibrillation](#) →
has Associated Procedure: [Heart surgery](#) →
has SSEID: 24052
intervention: [Drug therapy](#) →
investigates Drug Intervention: [Colchicine](#) →
primaryCondition: [Postoperative cardiac complication](#) →
res Study Conducted By: [Cardiothoracic Surgery](#) →
[Cardiovascular Diseases \(Clinical\)](#) →
[Cardiovascular Diseases \(Research\)](#) →
[Cardiovascular Research Center](#) →
[Cardiovascular Surgery](#) →
secondaryCondition: [Atrial fibrillation](#) →
[Post-pericardiotomy syndrome](#) →

at which clinical studies show up automatically based on shared annotations alone. Clinical trials are accessible on [mayoclinic.org](#) from diseases/conditions topics (for example, [breast cancer](#)), treatment/procedure topics (for example, [liver transplant](#)) and clinical departments (for example, the [Breast Clinic](#)). On the research site [mayo.edu/research](#), they are directly integrated with research centers conducting trials (e.g., [Cardiovascular Research Center](#)).

Findability improvements are mainly reached through enhanced site search, incorporating the additional access points explicated above. A second main driver of search is a new clinical trials landing page, leveraging the core descriptive metadata from epiCenter (study phase, open/closed status, location and so forth) in the form of facets, and the model-based annotations as an autocomplete feature in the search field.

Evaluating Outcomes

No formal evaluation has been attempted since the launch of the search and (complementary) annotation enhancements went live, but some data is still available for a quick reality check. The Mayo Clinic Cancer Center saw an increase in inquiries from 3500 to 5000 a year after the improvements went live, with the volume of inquiries continuing to increase. As we have seen, successfully implementing a metadata framework that promotes interoperability and findability of assets is a multi-stage process involving strategic and operational choices, from reliably acquiring source data, designing a model and crafting an annotation workflow to enabling the use of annotation in search and information architecture. ■