

A Simple Tool to Enrich Clinical Trial Data with Multiontology-Based Conceptual Tags

Holger Stenzhorn^{1,2}(✉)

¹ Chair of Methods in Medical Informatics,
University of Tübingen, Tübingen, Germany
`holger.stenzhorn@uni-tuebingen.de`

² Faculty of Medicine, Saarland University, Homburg, Saar, Germany

1 Introduction

The use of ontologies to ease planning and execution of clinical trials and the handling of the resulting data has been proposed in various forms over the past years ranging from dedicated ontologies to ontology-driven software. ObTiMA [11] is of the latter type and provides a complete web-based clinical trial management system which allows to define all data collection forms and items visually and then automatically generates both user interface and the database to enter patient data. Using the system in several projects proofed its ontological base as useful and powerful but also revealed four major issues:

- The initial focus for using ontological concepts was on semantically defining forms for data collection and the contained questions. Only little attention was paid on enabling the automatic or manual enrichment of other trial items.
- Despite the visual design of those forms, its user interface to semantically define questions (based on an ontology tree) was judged as difficult to use by clinical experts without ontology experts (and thus time-consuming and error-prone).
- The concepts could only be selected from a single ontology which was hard-coded into the system (first the ACGT Master Ontology [2], later (after some re-coding) the Health Data Ontology Trunk [9]). Thus established ontologies, like NCI Thesaurus [8] or SNOMED CT [10] could not be employed.
- Albeit trial data could be exported in standard CDISC ODM [3] extended with ontological concepts, no standard tool was capable to interpret this additional information and an export in RDF format for further processing was missing.

To target those issues, an extensive reengineering took place and is described in detail below: Now, any external ontology can be imported and its concepts used to specify additional data for all logical parts of a trial. A simpler user interface hides the inherent complexity of the ontologies and finally, all trial data can now be exported via RDF.

2 Ontology Management

The restriction to a single, hard-coded ontology was lifted and the possibility to import and use ontologies dynamically at runtime was introduced. Figure 1

Acronym ^	Name ^	Namespace		Enabled ^	
		Prefix ^	IRI ^		
ChEBI	Chemical Entities of Biological Interest	obo	http://purl.obolibrary.org/obo/	✓	
CTCAE	Common Terminology Criteria for Adverse Events	ctcae	http://ncicb.nci.nih.gov/xml/owl/EVS/ctcae.owl#	✓	
DO	Disease Ontology	obo	http://purl.obolibrary.org/obo/	✓	
FMA	Foundational Model of Anatomy	fma	http://purl.org/sig/ont/fma/	✓	
GO	Gene Ontology	obo	http://purl.obolibrary.org/obo/	✓	
LOINC	Logical Observation Identifiers Names and Codes	loinc	http://loinc.org/owl#	✓	
MedDRA	Medical Dictionary for Regulatory Activities	meddra	http://purl.bioontology.org/ontology/MEDDRA/	✓	
NCIt	NCI Thesaurus	ncit	http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#	✓	
OBI	Ontology for Biomedical Investigations	obo	http://purl.obolibrary.org/obo/	✓	
SNOMED-CT	Systematized Nomenclature of Medicine Clinical Terms	snomed-ct	http://snomed.info/	✓	

Fig. 1. List of the ontologies currently loaded in the system.

presents an (example) excerpt of all currently loaded ontologies. To add an ontology, the user simply clicks on the respective button and provides in a pop-up dialog (optional) acronym, name, namespace, (optional) description and source location of the ontology where the latter can either be any remote Web location or a local file reference (e.g. if the ontology artifacts are too large or not openly available). During the loading process, the ontologies are scanned for concepts and their respective identifiers and (possibly multilingual) labels are stored in a Lucene-based index [1] for performant subsequent retrieval. The ontologies can be either in OWL format or a text file with each line containing the concept identifier and label. (To extract this data, a regular expression has to be provided).

After the import, the ontologies and their concepts are immediately available system-wide and can be added to existing trials or used when creating new ones. Within a trial it is then possible to select the relevant ontologies, as shown in Fig. 2, so that users in that trial are only presented with the concepts from those ontologies (and irrelevant concepts from others are hidden).

Available Ontologies	
<input type="checkbox"/>	Acronym ^ Name ^
<input type="checkbox"/>	ChEBI Chemical Entities of Biological Interest
<input checked="" type="checkbox"/>	CTCAE Common Terminology Criteria for Adverse Events
<input checked="" type="checkbox"/>	DO Disease Ontology
<input type="checkbox"/>	GO Gene Ontology
<input checked="" type="checkbox"/>	LOINC Logical Observation Identifiers Names and Codes

Fig. 2. Ontologies selected for being used within a trial.

3 Conceptual Tagging

The tagging of the different trial elements with concepts is realized as auto-complete field widgets with drop-down lists of all matching concepts. This means that in order to find a suitable concept, the user does not have to navigate through complex ontology trees but simply types one or more terms (or sub-terms) in the text field and all ontologies selected for the trial are searched for the fitting concepts where their labels contains all elements of the entered query. Note that the drop-down lists are updated dynamically when typing or erasing

characters in the field. (To keep the user interface highly responsive even with long queries and several large ontologies at once, this is realized by employing the Lucene index – cf. above.) The concepts displayed are grouped according to their source ontology and for each concept its identifier and all attached (possibly multilingual) labels are shown too, cf. Figs. 3, 4 and 5.

congenital heart disease
DO - Disease Ontology
congenital heart disease obo:DOID_1682
lethal congenital glycogen storage disease of heart
LOINC - Logical Observation Identifiers Names and Codes
Cyanotic congenital heart disease English
MedDRA - Medical Dictionary for Regulatory Activities
Congenital heart disease NOS English
Disease heart congenital (NOS) English
Heart disease congenital English
NCIt - NCI Thesaurus
Congenital Heart Disease
SNOMED-CT - Systematized Nomenclature of Medicine Clinical Terms
(Blue baby) or (cyanotic congenital heart disease NOS)

Fig. 3. Concepts from multiple ontologies matching the terms in the query.

heart surgery
MedDRA - Medical Dictionary for Regulatory Activities
Open heart surgery English meddra:10048935
开胸心脏手术 Chinese
Operace na otevřeném srdci Czech
open hartchirurgie Dutch
Chirurgie à coeur ouvert French
offene Herzoperation German
Nyitott szívműtét Hungarian
Chirurgia a cuore aperto Italian
Cirurgia de coração aberto Portuguese
Cirugía a corazón abierto Spanish

Fig. 4. Single matching concept with various labels in several languages.

cardiocranial
SNOMED-CT - Systematized Nomenclature of Medicine Clinical Terms
Cardiocranial syndrome Pfeiffer type snomed-ct:720606005
Cardiocranial syndrome Pfeiffer type (disorder)
Craniosynostosis with congenital heart disease and intellectual disability syndrome
Pfeiffer Singer Zschiesche syndrome
Sagittal craniostenosis with congenital heart disease, mental deficiency and mandibular ankylosis

Fig. 5. Single concept with alternative labels.

Note again that this interface is uniform for all “taggable” trial elements. A concrete example is the creation of a question about a patient’s gender with two answer possibilities. In here, concepts for *Gender* are chosen from both NCI Thesaurus and SNOMED CT to tag the question and its answer possibilities *Male* and *Female*, see Figs. 6 and 7. Note that this is part of the trial setup process where all necessary aspects, including the ontology-based tags are visible. But when patient data is entered during the trial execution, all tags are hidden from the users (e.g. trial nurses) as they are concerned with a quick and simple data entry but not with any ontological representations, see Fig. 8. (Yet in the background, all questions/answers are still linked to their defined tags.).

Fig. 6. Definition of a question with concepts/tags added to the question itself and each answer possibility (marked by tag icon next to them).

Fig. 7. Two concepts/tags added to one answer possibility

Fig. 8. Simple view with tags hidden when filling in a patient form.

4 Data Export

All trial-related data can still be exported in ontologically-enriched CDISC ODM as this format is currently still one of the de-facto standards for exchanging trial data. But in addition it is now further possible to export that data in the form of RDF either to a local file or by pushing it via SPARQL Update [6] to a connected triple-store. The overall structure of this RDF export is based on a newly defined vocabulary which is derived closely from the original CDISC ODM elements’ specification re-using existing vocabularies, like the Dublin Core Metadata Set [4] as much as possible. The use of this “home-grown” vocabulary has the advantage that the resulting RDF (semantically) follows the original CDISC ODM very closely and is thus easily interpretable in this context but has the disadvantage of being non-standard. Therefore, in addition, an export based on the proposed FHIR RDF representation [12] is provided that integrates the approaches of both [5, 7] for mapping CDISC ODM to FHIR.

5 Conclusions

As told above, applying ontologies in clinical trial management is nothing novel per se. The difference here lies in the strong focus on (1) ease-of-use which allows people without much (or any) “ontological background” to use ontologies and concepts in their regular clinical trial work approaches with little training through an intuitive and responsive user interface, (2) applying well established,

standard ontologies combined with current Semantic-Web technologies to foster both semantic and technical interoperability.
 (Additional information can be found at <https://purl.org/holger/monster>).

References

1. Apache Software Foundation: Lucene. <https://lucene.apache.org>
2. Brochhausen, M., Weiler, G., Cocos, C., et al.: The ACGT master ontology on cancer - a new terminol. source for oncolog. Practice. In: Proceedings of IEEE Symposium on Computer-Based Medical Systems (2008)
3. CDISC Consortium: Operational Data Model (ODM). <https://www.cdisc.org/odm>
4. DCMI: Dublin Core Meta. Elem. Set. <http://dublincore.org/documents/dces>
5. Doods, J., Neuhaus, P., Dugas, M.: Converting ODM metadata to FHIR questionnaire resources. *Stud. Health Technol. Inform.* **228**, 456–460 (2016)
6. Gearon, P., et al.: SPARQL Update. <https://www.w3.org/TR/sparql11-update>
7. Leroux, H., Metke-Jimenez, A., Lawley, M.: ODM on FHIR: towards achieving semantic interoperability of clinical study data. In: Proceedings of SWAT4LS (2015)
8. National Cancer Institute: NCI Thesaurus. <https://ncit.nci.nih.gov>
9. Sanfilippo, E., Schwarz, U., Schneider, L.: The health data ontology trunk (HDOT) - towards an ontolog. represent. of cancer-related knowledge. In: Proceedings of IARWISOCI (2012)
10. SNOMED International: SNOMED CT. <http://www.snomed.org/snomed-ct>
11. Stenzhorn, H., Weiler, G., Brochhausen, M., Schera, F., Kritsotakis, V., et al.: The ObTiMA system - ontology-based managing of clinical trials. In: Proceedings of Medinfo (2010)
12. FHIR IS WG. FHIR RDF Representation. <https://www.hl7.org/fhir/rdf.html>

Data Integration in the Life Sciences

12th International Conference, DILS 2017, Luxembourg,

Luxembourg, November 14-15, 2017, Proceedings

Da Silveira, M.; Pruski, C.; Schneider, R. (Eds.)

2017, VII, 109 p. 49 illus., Softcover

ISBN: 978-3-319-69750-5