

FDA/PhUSE Semantic Technology Project

Represent Existing CDISC Standards in RDF

Introduction

The FDA/PhUSE Semantic Technology project investigates how formal semantic standards can support the clinical and non-clinical trial data life cycle from protocol to submission. Within this project, several teams work on dedicated tasks to execute this investigation. The first outcome and deliverable of this project is the representation of existing CDISC standards in RDF, including representations of the following CDISC standards:

- CDASH 1.1
- SDTM 1.2 and SDTM Implementation Guide 3.1.2
- SDTM 1.3 and SDTM Implementation Guide 3.1.3
- SEND Implementation Guide 3.0
- ADaM 2.1 and ADaM Implementation Guide 1.0
- Controlled Terminology

Today, CDISC publishes these standards in a paper based format and partly in Excel, which makes it difficult to consistently represent and process this information. The RDF representation addresses both issues by providing at the same time a formal model, a machine readable representation, and an exchange format. Accordingly, we have decided not to reproduce a lot of paper based documentation. This document provides only a few pointers, other than that the RDF models are self-describing and we therefore encourage you to use an ontology editor and explore the RDF datasets this way.

The main goal of this project is to represent the existing CDISC standards in RDF, not to remodel the standards, even when sometimes certain deficits became obvious. This is a conscious decision since the ownership of the standards lies within CDISC and any changes to the standards should be addressed directly by CDISC.

Distribution

The **rdm.cdisc.org** distribution contains the following directories.

import-files	Excel import files used to import the content into RDF
resources	Imported ontologies such as SKOS and Dublin Core
schemas	The schemas from which the ontologies are built
std	The RDF representation of the CDISC standards
terminology-2013-06-28	CDISC terminology published by the NCI

import-files

There is an import file for each CDISC standard in RDF file. Each sheet contains a description of the instances of a single OWL class. The header of the first column is the name of OWL class. The other headers are the names of the predicates that can be applied to instances of this OWL class. Each row describes an instance. The first column contains the URI of the instance, the other columns describe

the properties of the instance, which is either a literal for an owl:DatatypeProperty or a resource URI for an owl:ObjectProperty. Every resource URI is explicitly represented by its qname, we don't assume any default namespace.

If you are new to RDF but familiar with CDISC standards, the import files should give you a very good idea of what is being represented in RDF. In that case, we would recommend to first explore the import files before turning to the RDF datasets.

resources

The resources directory contains the RDF files for SKOS and Dublin Core. We use these vocabularies to annotate and document the CDISC ontologies.

schemas

The schemas define the classes and predicates that constitute the ontology to represent the CDISC standards. There are three schemas. The **Meta Model Schema** and the **CT Schema** are already in use today by the NCI and are used to publish the CDISC controlled terminology in RDF. It gives a fairly compact vocabulary to represent metadata in the spirit of the ISO 11179 standard for metadata registries (MDR). The **CDISC Schema** defines some additional vocabulary to deal with more CDISC specific properties such as assumptions specified in the Implementation Guide, CRF completion instructions in CDASH etc.

The core element of the Metadata Model is that of an **Administered Item**, which may represent any item that is managed in an MDR. An administered item must be named and defined within a **Context** that defines the scope within which the administered item has meaning. A key administered item is that of **Data Element** that represents a generic unit of data. A **Classifier** or a defined subclass thereof offers a generic way to classify administered items. This applies in particular to data elements, e.g. an SDTM data element may be classified as *Character* or *Numeric*, or also according to compliance classifiers such as *Required*, *Expected*, or *Permissible*. A term of the controlled terminology of a data element is defined by a **Permissible Value** that is part of an **Enumerated Value Domain**.

Every data element belongs to a context, e.g. the data element **VSTESTCD** belongs to the context of the CDISC domain **VS**. A context can recursively be contained in a higher level context, e.g. **VS** belongs to the context **Findings Observation Class**, which in turn belongs to the context **Model SDTMIG 3-1-2**. One way to explore the RDF datasets is to start from a high-level context, e.g. instances of the **Model** class and then to follow the links to the lower level contexts.

The following table gives the file names, preferred namespace prefixes, namespaces, and OWL import dependencies of the schema files.

File	Prefix	Namespace	OWL Import
meta-model-schema	mms	http://rdf.cdisc.org/mms#	skos
ct-schema	cts	http://rdf.cdisc.org/ct/schema#	mms
cdisc-schema	cdiscs	http://rdf.cdisc.org/std/schema#	mms, cts

terminology-2013-06-28

The NCI already publishes CDISC controlled terminology in RDF based on the Meta Model Schema and the CT Schema. The files included here are the ones published by the NCI in June 2013. To obtain the latest files, please follow the appropriate link from the **cdisc.org** web site. The files provided here are for demonstration purposes only. The following table gives the file names, preferred namespace prefixes, namespaces, and OWL import dependencies of the terminology files.

File	Prefix	Namespace	OWL Import
adam-terminology	adamct	http://rdf.cdisc.org/adam-terminology#	mms, skos
cdash-terminology	cdashct	http://rdf.cdisc.org/cdash-terminology #	mms, skos
glossary-terminology		http://rdf.cdisc.org/glossary-terminology #	mms, skos
qs-terminology		http://rdf.cdisc.org/qs-terminology #	mms, skos
sdtm-terminology	sdtmct	http://rdf.cdisc.org/sdtm-terminology #	mms, skos
send-terminology	sendct	http://rdf.cdisc.org/send-terminology #	mms, skos

std

There is an RDF dataset in Turtle format (file extension ttl) for each RDF representation of a CDISC standard. The following table gives the file names, preferred namespace prefixes, namespaces, and OWL import dependencies of the data standard files.

File	Prefix	Namespace	OWL Import
cdash-1-1	cdash-1-1	http://rdf.cdisc.org/std/cdash-1-1#	cdiscs, cdashct, sdtmct
sdtm-1-2	sdtm-1-2	http://rdf.cdisc.org/std/sdtm-1-2#	cdiscs
sdtmig-3-1-2	sdtmig-3-1-2	http://rdf.cdisc.org/std/sdtmig-3-1-2#	sdtm-1-2, sdtmct
sendig-3-0	sendig-3-0	http://rdf.cdisc.org/std/sendig-3-0#	sdtm-1-2, sdtmct
sdtm-1-3	sdtm-1-3	http://rdf.cdisc.org/std/sdtm-1-3#	cdiscs
sdtmig-3-1-3	sdtmig-3-1-3	http://rdf.cdisc.org/std/sdtmig-3-1-3#	sdtm-1-3, sdtmct
adam-2-1	adam-2-1	http://rdf.cdisc.org/std/adam-2-1#	cdiscs
adamig-1-0	adamig-1-0	http://rdf.cdisc.org/std/adamig-1-0#	adam-2-1, adamct
all-standards		http://rdf.cdisc.org/std/all-standards#	[all of the above]

There is a file **all-standards** that imports all the other standards, but does nothing else. This is a good starting point if you would like to explore all standards from a single entry point.

Known Issues

File Formats

The terminology and schema files are written in RDF/XML since this is the format the NCI currently uses. The CDISC standards in RDF are written in Turtle, which is a more popular format. In the future we would like to provide a distribution for both formats.

Meta Model Schema

The following changes to the Meta Model Schema will need to be coordinated with the NCI who currently uses this schema to publish the CDISC controlled terminology.

- The use of **mms:broader** has been replaced by the new **mms:dataElement** object property, which relates a Data Element Context to its Data Element. The Meta Model Schema still contains the **mms:broader** object property, but it is expected to be removed at a later date.
- The range of **mms:dataElementType** has been changed from `xsd:QName` to `xsd:anySimpleType` to avoid warnings from ontology editors. This is currently the only difference with the NCI version, and it is expected that the NCI version will be updated at a later time.
- The following classes have been added to the CDISC Schema and are expected to move into the Meta Model Schema at a later date. For any standard importing the CDISC Schema, nothing will change once these classes are moved since they already use the correct namespace prefix.

- § `mms:DataCollectionForm`
 - § `mms:Domain`
 - § `mms:DomainContext`
 - § `mms:DataElementContext`
 - § `mms:DataCollectionField`
 - § `mms:Column`

- The ontology `rdfs:label` of the CT schema currently reads “SDTM Terminology Schema”. This needs to be changed to “CDISC Terminology Schema” and will be fed back to the NCI.

SEND Implementation Guide

- The SEND Implementation Guide 3.0 is based on SDTM 1.2, but includes some variables that relate to the following SDTM 1.3 data elements: --ANTREG, --CSTATE, --DETECT, --DIR, --EXCLFL, --LAT, --LEAD, --PORTOT, --REASEX, --SPCUFL.
- No SDTM data elements were available for --ENINT, --STINT, and --METHOD (in the interventions domain).

Analysis Data Model

The analysis data model has generally given more difficulties to map within the existing schemas.

- Time to Event (TTE) has not been included since the documentation seems to provide only an example and is probably not to be considered a standard. Also, there are new variables in TTE that are not part of the ADaM Implementation Guide, e.g. the ASEQ variable on page 6.
- The ADAE dataset is currently part of the class ADAE, but the ADAE documentation states on page 7 that the class of this dataset may change in a future version when a more general occurrence model becomes available.

Future Topics

- Model which SDTM variables are possibly permitted in SDTMIG
- Model which SDTM variables are disallowed in SDTMIG.
- Model data element traceability from ADaM to SDTM to CDASH.
- Model conformance checks
- Generate RDF data models from the corresponding metadata models.

Online References

1. Go to w3.org/2001/sw/wiki/Main_Page for an entry page to the W3C semantic web standards.
2. Go to metadata-standards.org/11179 for the six parts reference documents on ISO 11179.
3. Go to cdisc.org for comprehensive documentation on all CDISC standards.
4. Go to omg.org/spec for a directory of all OMG specifications.
5. Go to cabig.nci.nih.gov/concepts/EVS/ to see what the National Cancer Institute (NCI) is doing in the area of controlled terminologies and ontology modeling.
6. Go to bioontology.org to visit the National Center for Biomedical Ontology (NCBO), a great resource for biomedical ontologies and related technologies. It provides a repository and federated search across a large number of biomedical ontologies.

Recommended Reading

Many people interested in CDISC standards may be less familiar with semantic web technologies. The following is a shortlist for recommended reading.

1. Dean Allemang and Jim Hendler. *Semantic Web for the Working Ontologist*. Second Edition. Morgan Kaufmann, 2011. This is an excellent book, well-written, specifically on the modeling aspects of RDF and OWL. If you are going to read one book on the subject, we suggest to make it this one.
2. Christopher Walton. *Agency and the Semantic Web*. Oxford University Press, 2007. This book gives a broad outlook on knowledge systems and the semantic web, including more academic background on the computational aspects of the subject.
3. Dragan Gasevic, Dragan Djuric, and Vladan Devedzic. *Model Driven Engineering and Ontology Management*. Second Edition. Springer, 2009. This book provides valuable insight on knowledge engineering and the relationship between the different modeling spaces.

Contact

For more information about the FDA/PhUSE Semantic Technology project, please visit the Wiki at phusewiki.org or go straight to phusewiki.org/wiki/index.php?title=Semantic_Technology.

FDA/PhUSE Semantic Technology Project Team
28.10.2013