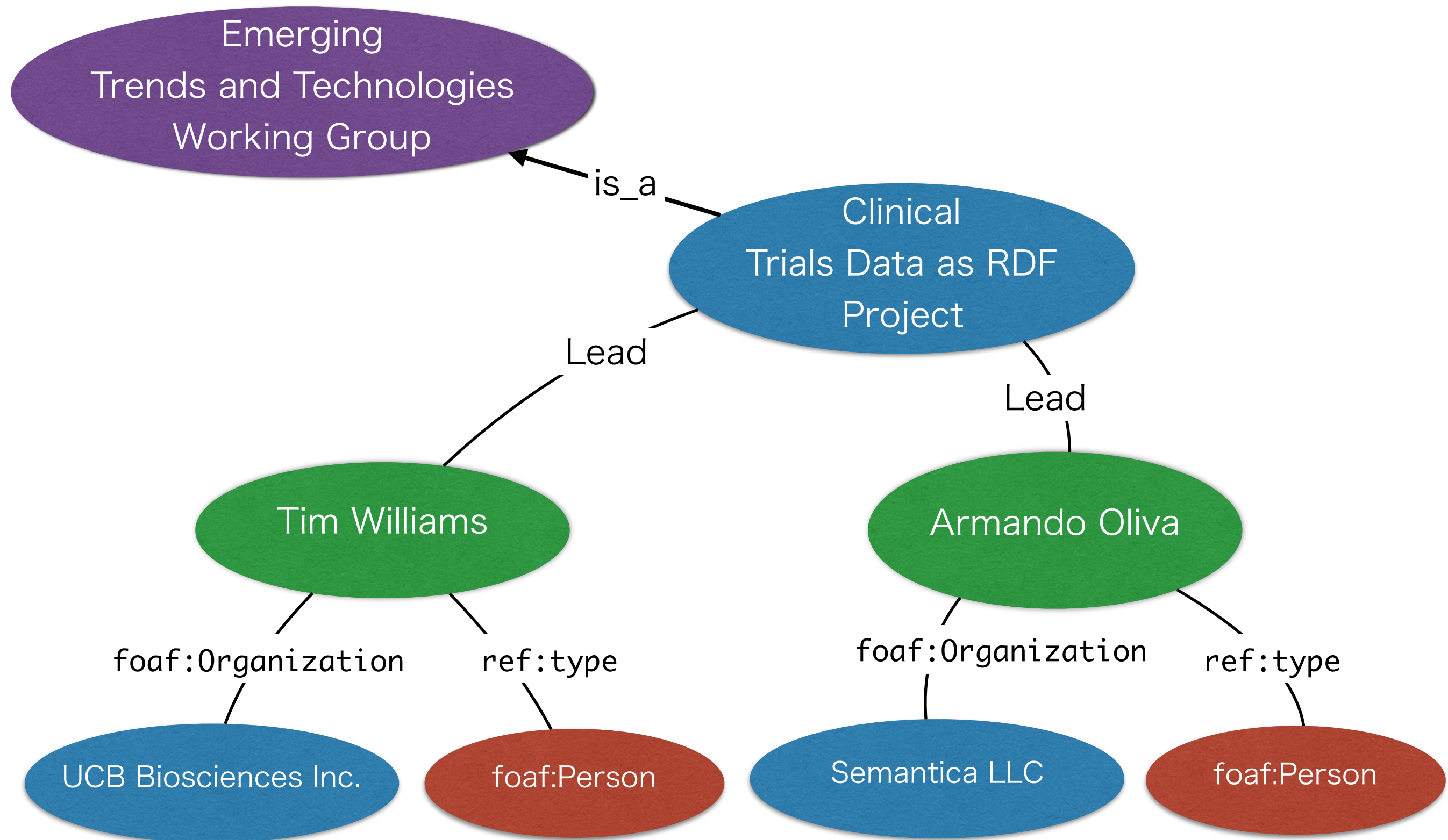


Automatically Generate Define-XML from Clinical Trials Data as RDF with SPARQL Query

Ippei Akiya, DataDriven, Inc.

Clinical Trials Data as RDF Project



What we do

The CTDasRDF project was started from the PhUSE CSS conference on March, 2017

Motivation

FDA CDER's survey* indicated 26% of submissions had a least one error. We expects our new approach with RDF and other related technologies will lead to the highly compliant submission data.

Approach

To generate the highly compliant submission data, including SDTM data, metadata(define.xml), protocol information into the RDF datasets, and extract as SDTM and define.xml by utilizing SPARQL Query and SPIN rule.

* Allard, Crystal , Technical Rejection Criteria for Study Data - Preliminary Findings. , PhUSE Single Day Event (SDE) in Boston, 2017.

Technology Background

Clinical Trials Data as RDF is built based on graph data technology

Typical data structures

Tabular data

SAS Dataset

CSV

Excel

Tree data

XML

JSON

Graph data

RDF

Neo4j

GraphX

Example

Untitled1* x DM x				
Filter				
	STUDYID Study Identifier	DOMAIN Domain Abbreviation	USUBJID Unique Subject Identifier	SUBJID Subject Identifier for the Study
1	CDISCPILOT01	DM	01-701-1015	1015
2	CDISCPILOT01	DM	01-701-1023	1023
3	CDISCPILOT01	DM	01-701-1028	1028
4	CDISCPILOT01	DM	01-701-1033	1033
5	CDISCPILOT01	DM	01-701-1034	1034
6	CDISCPILOT01	DM	01-701-1047	1047
7	CDISCPILOT01	DM	01-701-1057	1057
8	CDISCPILOT01	DM	01-701-1097	1097
9	CDISCPILOT01	DM	01-701-1111	1111
10	CDISCPILOT01	DM	01-701-1115	1115
11	CDISCPILOT01	DM	01-701-1118	1118
12	CDISCPILOT01	DM	01-701-1130	1130
13	CDISCPILOT01	DM	01-701-1133	1133
14	CDISCPILOT01	DM	01-701-1145	1145
15	CDISCPILOT01	DM	01-701-1146	1146
16	CDISCPILOT01	DM	01-701-1148	1148
17	CDISCPILOT01	DM	01-701-1153	1153
18	CDISCPILOT01	DM	01-701-1162	1162
19	CDISCPILOT01	DM	01-701-1176	1176

```
<ItemGroupDef OID="IG.DM"
  Domain="DM"
  Name="DM"
  Repeating="No"
  IsReferenceData="No"
  SASDatasetName="DM"
  Purpose="Tabulation"
  def:Structure="One record per subject"
  def:Class="SPECIAL PURPOSE"
  def:CommentOID="COM.DOMAIN.DM"
  def:ArchiveLocationID="LF.DM">
  <Description>
    <TranslatedText xml:lang="en">Demographics</TranslatedText>
  </Description>
  <ItemRef ItemOID="IT.STUDYID" OrderNumber="1" Mandatory="Yes" KeySequence="1"/>
  <ItemRef ItemOID="IT.DM.DOMAIN" OrderNumber="2" Mandatory="Yes"/>
  <ItemRef ItemOID="IT.USUBJID" OrderNumber="3" Mandatory="Yes" KeySequence="2" Met
  <ItemRef ItemOID="IT.DM.SUBJID" OrderNumber="4" Mandatory="Yes"/>
  <ItemRef ItemOID="IT.DM.RFSTDTC" OrderNumber="5" Mandatory="No" MethodOID="MT.RFS
  <ItemRef ItemOID="IT.DM.RFENDTC" OrderNumber="6" Mandatory="No" MethodOID="MT.RFE
  <ItemRef ItemOID="IT.DM.SITEID" OrderNumber="7" Mandatory="Yes"/>
  <ItemRef ItemOID="IT.DM.BRTHDTC" OrderNumber="8" Mandatory="No"/>
  <ItemRef ItemOID="IT.DM.AGE" OrderNumber="9" Mandatory="Yes" MethodOID="MT.AGE"/>
  <ItemRef ItemOID="IT.DM.AGEU" OrderNumber="10" Mandatory="No"/>
  <ItemRef ItemOID="IT.DM.SEX" OrderNumber="11" Mandatory="Yes"/>
  <ItemRef ItemOID="IT.DM.RACE" OrderNumber="12" Mandatory="No"/>
  <ItemRef ItemOID="IT.DM.ETHNIC" OrderNumber="13" Mandatory="Yes"/>
  <ItemRef ItemOID="IT.DM.ARMCD" OrderNumber="14" Mandatory="Yes"/>
  <ItemRef ItemOID="IT.DM.ARM" OrderNumber="15" Mandatory="Yes"/>
  <ItemRef ItemOID="IT.DM.COUNTRY" OrderNumber="16" Mandatory="Yes"/>
  <def:leaf ID="LF.DM" xlink:href="dm.xdt">
```

Why we should use RDF

RDF is Simple Structure, Simple Expression, and Well Defined Vocabulary.

Subject	Predicate	Object
sdtmig-3-1-3:Column.DM.ACTARM	sdtm:additionalMetadata	study:Arm
sdtmig-3-1-3:Column.DM.ACTARM	sdtm:variableComment	"According to randomization list"^^xsd:string
sdtmig-3-1-3:Column.DM.ACTARM	code:hasMethod	code:ActivityMethod_2
sdtmig-3-1-3:Column.DM.ACTARM	sdtm:additionalMetadata	study:Arm
code:ActivityMethod_2	rdf:type	code:Method
code:ActivityMethod_2	rdfs:comment	"Y for death subject"^^xsd:string
code:ActivityMethod_2	rdfs:label	"DERIVED"^^xsd:string
sdtm:variableComment	skos:definition	"Comments for SDTM variable"^^xsd:string

RDF Tripples

Why we should use RDF

Sparql Query Language

Sparql is Query Language that specification is released by World Wide Web Consortium(W3C).

Similar to SQL, But absolutely different.

```
SELECT ?SDTM_Variable ?Ordinal
WHERE{
    ?s          rdf:type          sdm:SDTMDefineRecord_DM;
                sdm:hasDFVariable ?SDTM_Variable.
    ?SDTM_Variable rdf:type      mms:Column;
                mms:ordinal      ?Ordinal.
}
ORDER BY ?Ordinal
```

Why we should use RDF

RDF related Technology includes

- **Standardized** Data Specification
- **Standardized** Query Language
- Ontology - **Standardized** Vocabulary



Useful Resources

- W3C RDF 1.1 Primer
<https://www.w3.org/TR/rdf11-primer/>
- W3C RDF Schema 1.1
<https://www.w3.org/TR/rdf-schema/>
- SPIN
<http://spinrdf.org>

Project Wiki

The screenshot shows a web browser window with the URL [phusewiki.org](http://www.phusewiki.org). The page title is "Clinical Trials Data as RDF - PhUSE Wiki". The browser's address bar shows the URL, and the page content is displayed below. The page has a navigation trail: "PhUSE Archived Projects » Linked Data & Graph Database » Evaluation Of Alternative Transport Formats » CSS Working Groups » Clinical Trials Data as RDF". The main content area is titled "Clinical Trials Data as RDF" and contains a "Project Overview" section. The "Project Overview" section includes a "Why are we doing this project?" paragraph, a "High level project goals" paragraph, and a "Philosophy" paragraph. To the right of the "Project Overview" section is a "Contents [hide]" list with 8 items: 1 Project Overview, 2 Project Leads, 3 Project Participants, 4 Project Updates, 5 Objectives and Timelines, 6 Project Activities, 7 Meeting Minutes, and 8 Archived Content. Below the "Project Overview" section is a "Project Leads" section with a table listing two project leads: Tim Williams (Project Lead, UCB, Tim.Williams@ucb.com) and Armando Oliva (Project Lead, Industry, AOlivamd@gmail.com). Below the "Project Leads" section is a "Project Participants" section with a table listing three participants: Daniel Choi (Participant, FDA, Daniel.Choi@fda.hhs.gov), Amit Jain (Participant, ICON, Amit.Jain@iconplc.com), and Andrew Miskell (Participant, Lilly, Andrew.Miskell@lilly.com).

phuse

Home
Categories
Recent changes
New Pages
Popular Pages

Projects

- CSS Working Groups
- Annual Conference
- Therapeutic Areas
- Good Programming Practice
- Data Transparency
- QSPI

PhUSE Wiki Help

- Help
- FAQ
- Training
- Wiki Feedback
- Wiki Admin Working Group

Tools

- What links here
- Related changes

Page Discussion Read View source View history Search PhUSE Wiki

Navigation trail: PhUSE Archived Projects » Linked Data & Graph Database » Evaluation Of Alternative Transport Formats » CSS Working Groups » Clinical Trials Data as RDF

Clinical Trials Data as RDF

Project Overview

Why are we doing this project?

A recent survey showed 26% of CDER SDTM applications had at least one error. This doesn't even cover data quality or correctness. The current data models are clearly not sufficient. We will look at the data models for these standards so we can move into the future.

High level project goals

Future proof data models that will lead to highly compliant SDTM domains and Integrated DEFINE - having the data and meta data available directly within the data to create the DEFINE based on a query. So create validated SDTM domains by query, and rule sets.

Philosophy

We will model to the data and not to the standards. We are not taking SDTM and modelling to SDTM, we are taking the data that makes up the SDTM domains and modelling those appropriately. We will then use some of the existing rule sets and terminology files like SDTM terminologies, RDF plus others to bring that validated SDTM domain data out as a deliverable from the query, using things like SPARQL/SPIN rules. We'll be using data that is within existing SDTM domains, convert it to linked data then roundtrip it back out to an SDTM domain and do a comparison, in the hope it will be as good, if not better.

Contents [hide]

- 1 Project Overview
- 2 Project Leads
- 3 Project Participants
- 4 Project Updates
- 5 Objectives and Timelines
- 6 Project Activities
- 7 Meeting Minutes
- 8 Archived Content

Project Leads

Name	Role	Organization	E-mail
Tim Williams	Project Lead	UCB	Tim.Williams@ucb.com
Armando Oliva	Project Lead	Industry	AOlivamd@gmail.com

Project Participants

Name	Role	Organisation	E-Mail
Daniel Choi	Participant	FDA	Daniel.Choi@fda.hhs.gov
Amit Jain	Participant	ICON	Amit.Jain@iconplc.com
Andrew Miskell	Participant	Lilly	Andrew.Miskell@lilly.com

http://www.phusewiki.org/wiki/index.php?title=Clinical_Trials_Data_as_RDF

Github Site

The screenshot displays the GitHub interface for the repository `phuse-org/CTDasRDF`. The browser's address bar shows the URL `https://github.com/phuse-org/CTDasRDF`. The repository page includes a header with the repository name, a search bar, and navigation links for Pull requests, Issues, Marketplace, and Explore. Below the header, the repository statistics are shown: 447 commits, 3 branches, 0 releases, and 2 contributors. A horizontal bar indicates the progress of the repository's development. The file list section shows the following files and folders:

File/Folder	Description	Commit Date
<code>SPARQL</code>	RDF-Driver, metadata triples converted to use redland pkg	5 months ago
<code>Wiki/Images</code>	new program flow image	8 months ago
<code>data</code>	minor correction	2 months ago
<code>doc</code>	latest changes	2 months ago
<code>r</code>	Example code for SPARQL package query to Stardog triplestore. Preferr...	3 months ago
<code>vis</code>	working example of the ontologies chart	4 months ago
<code>CTDasRDF.html</code>	commit all WIP prior to merge to Master	3 months ago
<code>CTDasRDF.rmd</code>	commit all WIP prior to merge to Master	3 months ago
<code>README.md</code>	Patient 1 data created without error and Apache RIOT validation passe...	3 months ago

The README section is titled **CTDasRDF**.

<https://github.com/phuse-org/CTDasRDF>

RDF Datasets of Clinical Trials Data as RDF

1. `code.ttl` – this file contains or links to resources representing defined concepts such as controlled terminologies. It includes Defined Activities. It currently provides links to SDTM terminology and the time ontology. In the future, it can be expanded to link to other terminologies in RDF such as MedDRA, LOINC, and the WHO Drug Dictionary. It is anticipated that this file will reside and be maintained on a public site for all implementers to reference, although various links to proprietary terminologies may be restricted based on licensing agreements.
2. `study.ttl` – contains the study metamodel in OWL. It contains the core classes and relationships previously discussed that are common to all studies. This ontology imports `code.ttl`. It is anticipated that this file will also be publicly available on the web.
3. `cdiscpilot01-protocol.ttl` – contains the concepts and relationships specific to the pilot study protocol, including the protocol-specified activities, rule sets, and controlled terms/value sets. It imports the `study.ttl` ontology. It is expected that this file will be the primary source to generate the blank case report form and the `define.xml` contents. Since study protocols are considered proprietary, the file will likely reside behind a firewall with restricted access. It also defines a separate namespace called `custom:` to store protocol-specific concepts and custom terms that are not present in `code.ttl`.
4. `cdiscpilot01.ttl` – contains the instance data for the study. It imports the `cdiscpilot01-protocol.ttl` file. This file also resides behind a firewall.
5. `sdtm.ttl` – contains or links to the SDTM ontologies that are useful in creating valid SDTM datasets from the knowledgebase. This file is publicly available.
6. `sdtm-cdiscpilot01.ttl` – links the instance data in `cdiscpilot01.ttl` with the SDTM ontology in `sdtm.ttl` from which the SDTM datasets are derived. Any protocol-specific SDTM implementation information is contained herein.

code.ttl

code.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/code.ttl) : [/Users/ippei/develop/CTDasRDF/data/rdf/code.ttl]

code.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/code.ttl) Search...

Active Ontology x Entities x Individuals by class x DL Query x

Class hierarchy: 'Unit term' [?][][][][]

Asserted []

'Unit term' — https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/code.ttl#Unit

Class Annotations Class Usage

Annotations: 'Unit term' [?][][][][]

Annotations +

rdfs:label [type: xsd:string] @ x o
Unit term

definition [type: xsd:string] @ x o

rdfs:comment [type: xsd:string] @ x o

Description: 'Unit term' [?][][][][]

Equivalent To +

SubClass Of +

'Controlled Terminology' ? @ x o

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

BEATS/MIN ? @ x

C ? @ x

cm ? @ x

F ? @ x

IN ? @ x

kg ? @ x

LB ? @ x

mg ? @ x

mmHg ? @ x

Target for Key +

Annotation property hierarchy Datatypes

Object property hierarchy Data property hierarchy Individuals by type

Object property hierarchy: [?][][][][]

Asserted []

owl:topObjectProperty

No Reasoner set. Select a reasoner from the Reasoner menu ☒ Show Inferences

study.ttl

study.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/study.ttl) : [/Users/ippei/develop/CTDasRDF/data/rdf/study.ttl]

study.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/study.ttl) Search...

Active Ontology x Entities x Individuals by class x DL Query x

Class hierarchy: 'DEMOG DATA COLLECTION' [Icons]

Annotations: [Icons]

Asserted [Dropdown]

Class Annotations Class Usage

Annotations: 'DEMOG DATA COLLECTION' [Icons]

Annotations +

- rdfs:label [type: xsd:string] DEMOG DATA COLLECTION [Icons]
- 'preferred label' [type: xsd:string] DEMOG DATA COLLECTION [Icons]
- definition [type: xsd:string] [Icons]
- rdfs:comment [type: xsd:string] [Icons]

Description: 'DEMOG DATA COLLECTION' [Icons]

Equivalent To +

SubClass Of +

- 'has age' only Age [Icons]
- 'Observation term' [Icons]
- ethnicity only 'Ethnic outcome' [Icons]
- Observation [Icons]
- race only 'Race outcome' [Icons]
- sex only 'Sex outcome' [Icons]

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

Target for Key +

Annotation property hierarchy Datatypes

Object property hierarchy Data property hierarchy Individuals by type

Object property hierarchy: [Icons]

Asserted [Dropdown]

owl:topObjectProperty

No Reasoner set. Select a reasoner from the Reasoner menu ☒ Show Inferences

cdiscpilot01-protocol.ttl

cdiscpilot01-protocol.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/cdiscpilot01-protocol.ttl) : [/Users/ippei/develop/CTDasRDF/data/rdf/cdiscpilot01-protocol.ttl]

Active Ontology x Entities x Individuals by class x DL Query x

Class hierarchy: 'Visit Screening 1 Blood Pressure' [?][][][]

Class Annotations Class Usage

Annotations: 'Visit Screening 1 Blood Pressure' [?][][][]

Annotations +

'preferred label' [type: xsd:string]

Visit Screening 1 Blood Pressure

Description: 'Visit Screening 1 Blood Pressure' [?][][][]

Equivalent To +

SubClass Of +

'SCREENING 1 ACTIVITY' [?][][][]

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

Target for Key +

Disjoint With +

Disjoint Union Of +

Annotation property hierarchy Datatypes

Object property hierarchy Data property hierarchy Individuals by type

Object property hierarchy: [?][][][]

owl:topObjectProperty

No Reasoner set. Select a reasoner from the Reasoner menu ☒ Show Inferences

cdiscpilot01.ttl

cdiscpilot01.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/cdiscpilot01.ttl) : [/Users/ippei/develop/CTDasRDF/data/rdf/cdiscpilot01.ttl]

Active Ontology x Entities x Individuals by class x DL Query x

Class hierarchy: 'Unique subject identifier' [?][][][]

Asserted []

'Study component'

- Activity
- Entity
 - 'Activity outcome'
 - 'Administrative activity outcome'
 - 'Analysis outcome'
 - 'Drug administration outcome'
 - 'Intervention outcome'
 - 'Activity status'
 - 'Anatomic location'
 - Category
 - Country
 - CRFComponent
 - Document
 - Event
 - Identifier
 - 'Investigator identifier'
 - 'Site identifier'
 - 'Study identifier'
 - 'Subject identifier'
 - 'Unique subject identifier'
 - Party
 - Assessor
 - Organization
 - Person
 - QEVAL
 - Product
 - 'Active product'
 - 'Investigational product' $\equiv$ StudyAgent
 - 'Placebo product'

Annotation property hierarchy | Datatypes

Object property hierarchy | Data property hierarchy | Individuals by type

Object property hierarchy: [?][][][]

Asserted []

owl:topObjectProperty

'Unique subject identifier' — https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/study.ttl

Class Annotations | Class Usage

Annotations: 'Unique subject identifier' [?][][][]

Annotations +

rdfs:label [type: xsd:string]

Unique subject identifier

Description: 'Unique subject identifier' [?][][][]

Equivalent To +

SubClass Of +

Identifier

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

- 01-701-1015
- 01-701-1023
- 01-701-1028

Target for Key +

Disjoint With +

Disjoint Union Of +

No Reasoner set. Select a reasoner from the Reasoner menu [] Show Inferences

sdtm.ttl

sdtm.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/sdtm.ttl) : [/Users/ippei/develop/CTDasRDF/data/rdf/sdtm.ttl]

< > sdtm.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/sdtm.ttl) Search...

Active Ontology x Entities x Individuals by class x DL Query x

Class hierarchy: Variable ? [] [] [] []

Asserted

- Aggregation
- Command
- Element
- 'Order by condition'
- Path
- Tuple
- Variable
- 'Test case'
- ResearchStaff ≡ 'Research staff'
- 'Rights Statement'
- SDTMComponent**
 - SDTMDomain**
 - SDTMRecord**
 - 'SDTMSubject level record'
 - 'SDTM DM Record'
 - 'SDTM Supplemental record'
 - 'SDTM VS Record'
- SDTMTerminology**
 - Domain**
 - Variable**
 - Standard
 - 'Study component'
 - StudyAgent ≡ 'Investigational product'
 - 'Temporal duration'
 - 'Temporal entity'
 - 'Temporal position'
 - 'Temporal Reference System'
 - 'Time Zone'
 - VocabularyEncodingScheme
 - xsd:date
 - xsd:string

Annotation property hierarchy Datatypes

Object property hierarchy Data property hierarchy Individuals by type

Individuals by type: Column.DM.RFSTDTC [] [] [] []

Variable (69)

- Column.DM.SITEID**
- Column.DM.RFSTDTC**

Column.DM.RFSTDTC — http://rdf.cdisc.org/std/sdtmig-3-1-3#Column.DM.RFSTDTC

Individual Annotations Individual Usage

Annotations: Column.DM.RFSTDTC ? [] [] [] []

Annotations +

controlledTermsOrFormat

ISO8601

Description: Column.DM.RFSTDTC ? [] [] [] []

Types +

Variable ? @ x o

Same Individual As +

Different Individuals +

Property assertions: Column.DM.RFSTDTC [] [] [] []

Object property assertions +

'has method' DERIVED ? @ x o

Data property assertions +

Negative object property assertions +

Negative data property assertions +

No Reasoner set. Select a reasoner from the Reasoner menu ☒ Show Inferences

sdtm-cdiscpilot01.ttl

sdtm-cdisc01.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/sdtm-cdisc01.ttl) : [/Users/ippei/develop/CTDasRDF/data/rdf/sdtm-cdisc01.ttl]

< > sdtm-cdisc01.ttl (https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/sdtm-cdisc01.ttl) Search...

Active Ontology x Entities x Individuals by class x DL Query x

Class hierarchy: 'SDTM DM Record' [?] [] [] [] [x]

Asserted [] [] [] [] [x]

'Person 1' — https://raw.githubusercontent.com/phuse-org/CTDasRDF/master/data/rdf/cdiscpilot01.ttl#Person_

Individual Annotations Individual Usage

Annotations: 'Person 1' [?] [] [] [] [x]

Annotations +

'preferred label' [type: xsd:string]

Person 1

Description: 'Person 1' [?] [] [] [] [x]

Property assertions: 'Person 1' [] [] [] [] [x]

Object property assertions +

'participates in' 'Randomization 1' [?] [] [] [] [x]

'has unique subject ID' 1015 [?] [] [] [] [x]

'participates in' VisitBaseline_1 [?] [] [] [] [x]

'has Site' site-701 [?] [] [] [] [x]

'participates in' PopFlagSaf_1 [?] [] [] [] [x]

'participates in' PopFlagCmpltr16Wk_1 [?] [] [] [] [x]

'participates in' 'Product administration 1' [?] [] [] [] [x]

'participates in' PopFlagEff_1 [?] [] [] [] [x]

'has study participation interval' 'Study Participation Interval 1' [?] [] [] [] [x]

'has unique subject ID' 01-701-1015 [?] [] [] [] [x]

'participates in' 'Informed consent 1' [?] [] [] [] [x]

'participates in' PopFlagCmpltr8Wk_1 [?] [] [] [] [x]

'has lifespan' 'Lifespan Interval 1' [?] [] [] [] [x]

'participates in' VisitScreening1_1 [?] [] [] [] [x]

'participates in' 'P1 Screening 1 Demographic data collection' [?] [] [] [] [x]

'participates in' PopFlagItt_1 [?] [] [] [] [x]

'participates in' VisitWk2_1 [?] [] [] [] [x]

'participates in' PopFlagCmpltr24Wk_1 [?] [] [] [] [x]

'has lifespan' 'Reference Interval 1' [?] [] [] [] [x]

'participates in' VisitWk24_1 [?] [] [] [] [x]

'actual arm' Pbo [?] [] [] [] [x]

Types +

'Enrolled subject' [?] [] [] [] [x]

Same Individual As +

Different Individuals +

Annotation property hierarchy Datatypes

Object property hierarchy Data property hierarchy Individuals by type

Individuals by type: 'Person 1' [] [] [] [] [x]

'Person 1'

'Person 3'

'Enumerated Value Domain' (141)

ETHNIC (4)

ETHNIC

ETHNIC

'Ethnic outcome' (4)

'Ethnic outcome'

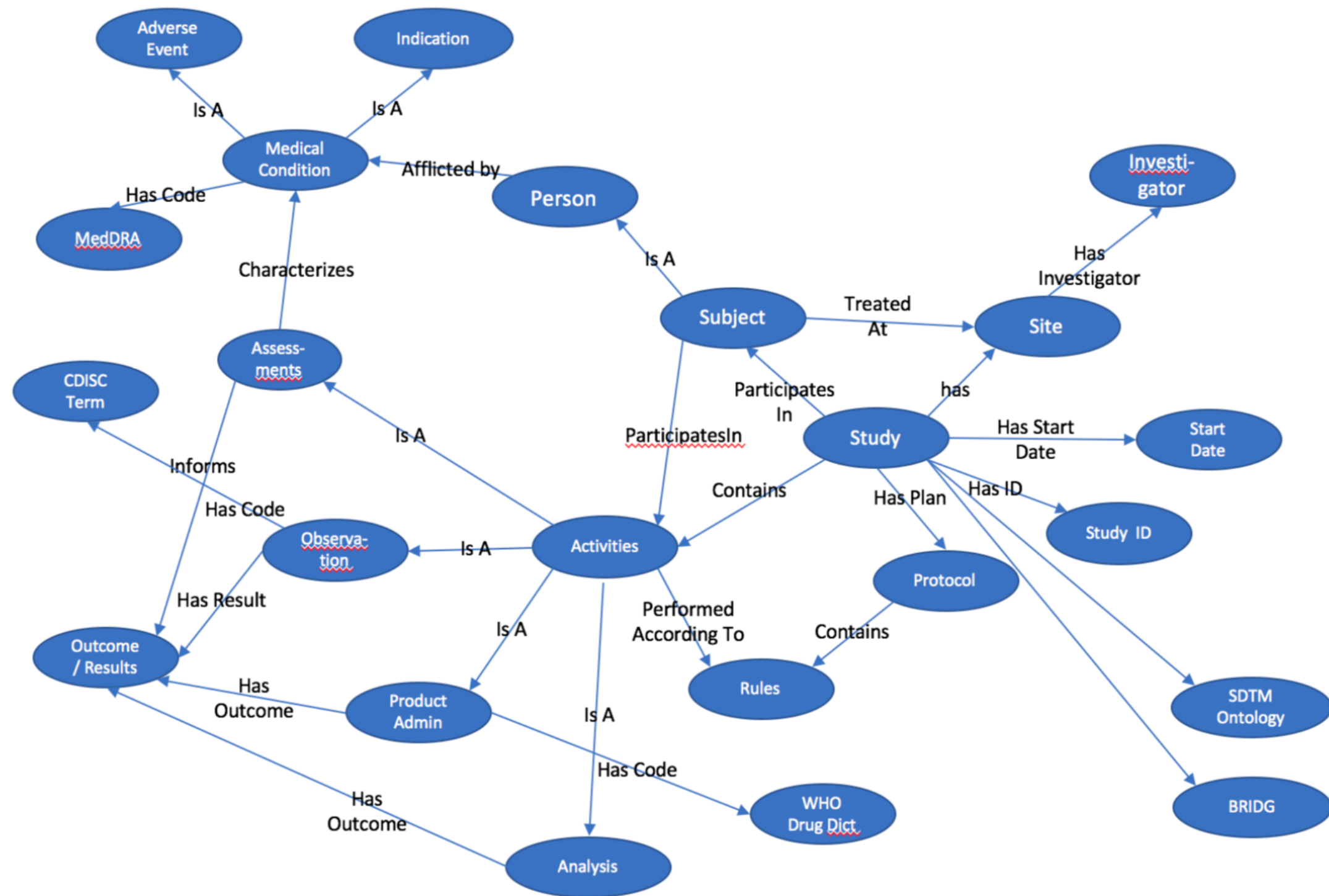
ethnicity

Extent

'File Format'

No Reasoner set. Select a reasoner from the Reasoner menu [x] Show Inferences

Graph Visualization



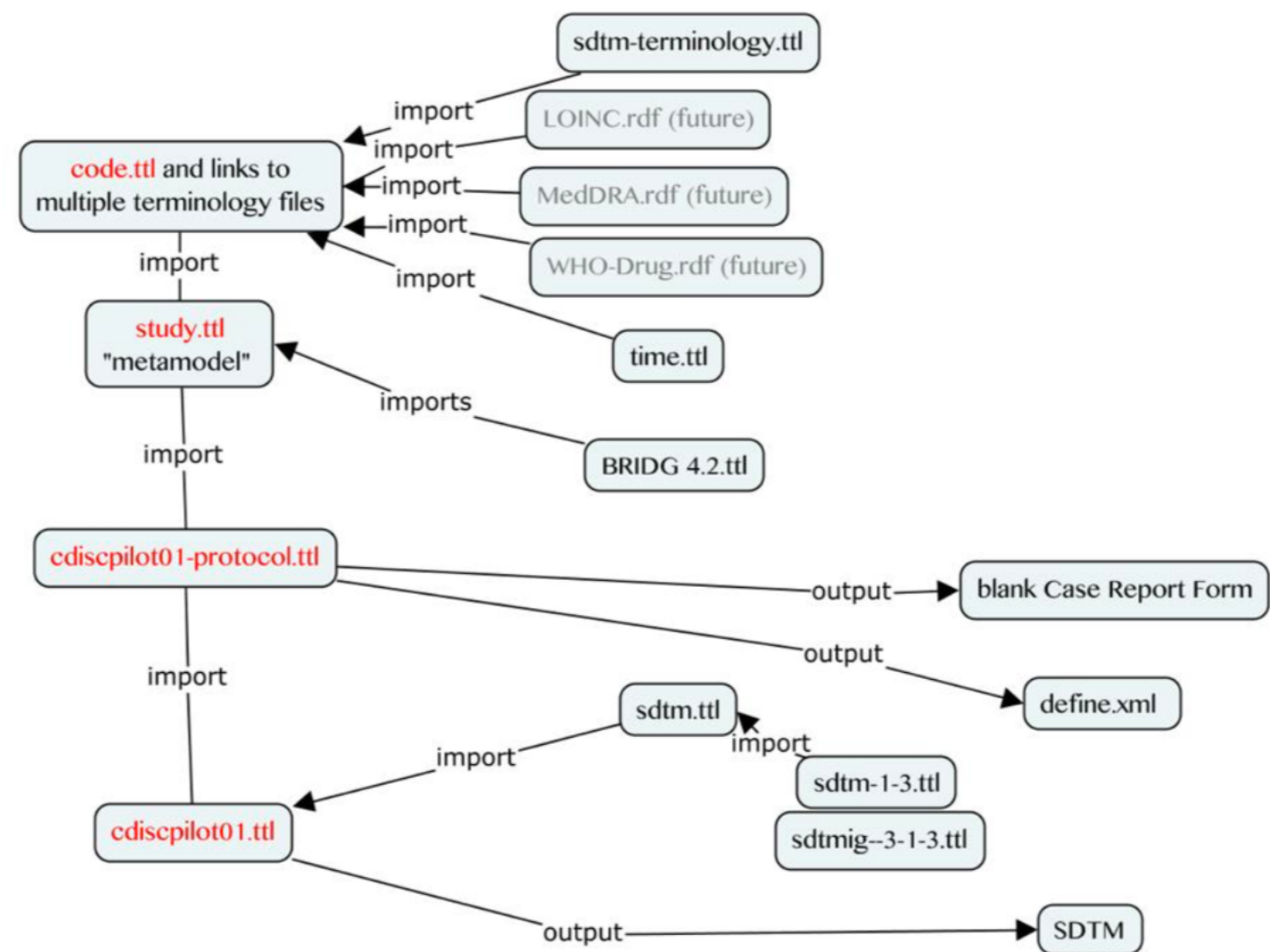
Tim Williams and Armando Oliva, Breaking the Mold: Clinical Trials Data as RDF, PhUSE 2017

Defined Concepts and Activities
and Controlled Terminologies
and Study Metamodel (how to
combine defined concepts into a protocol)

Study Metamodel
(how to combine defined concepts in
a Study)

Protocol-Specified Activities
(includes custom terms)

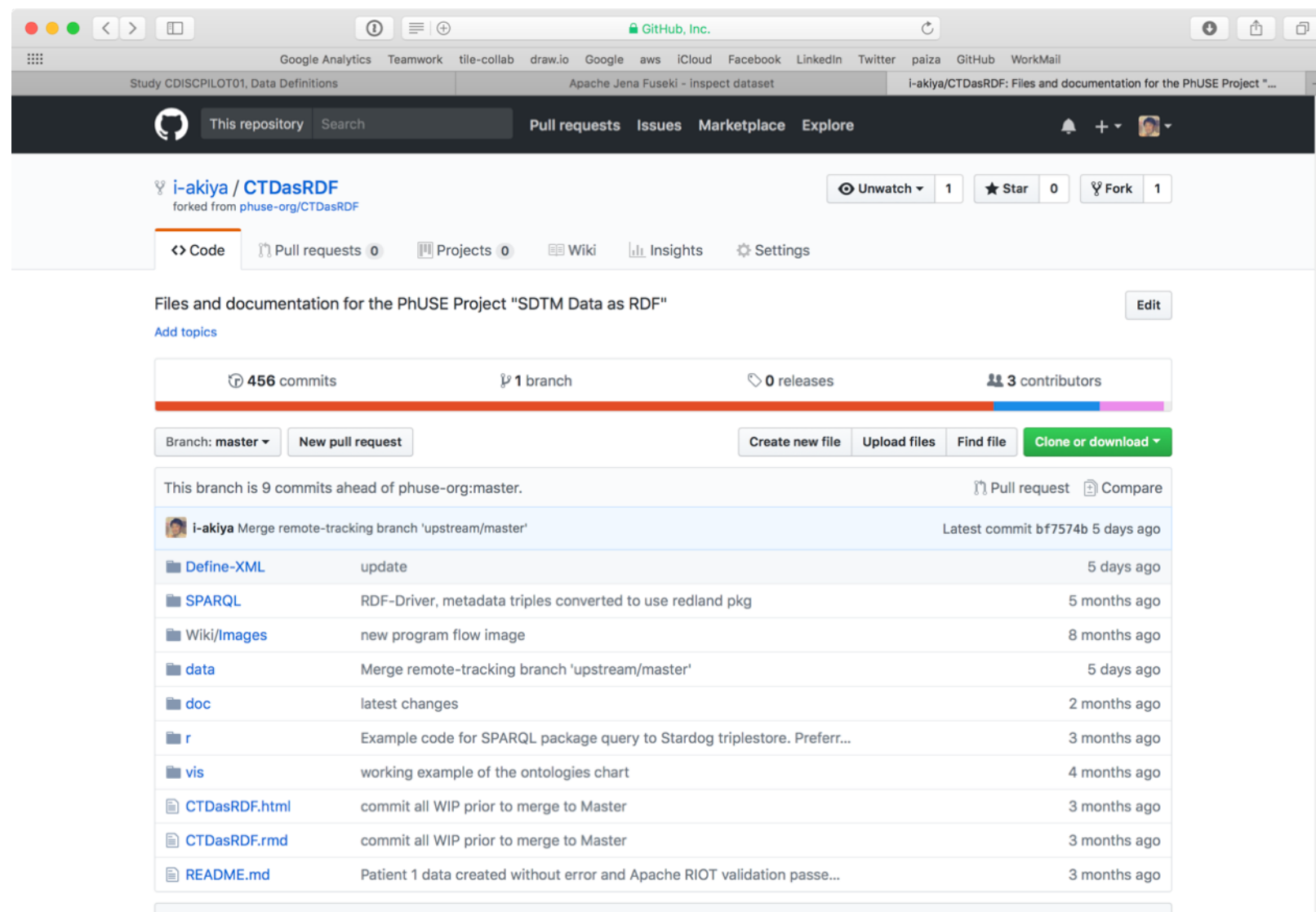
Scheduled and Performed
Activities

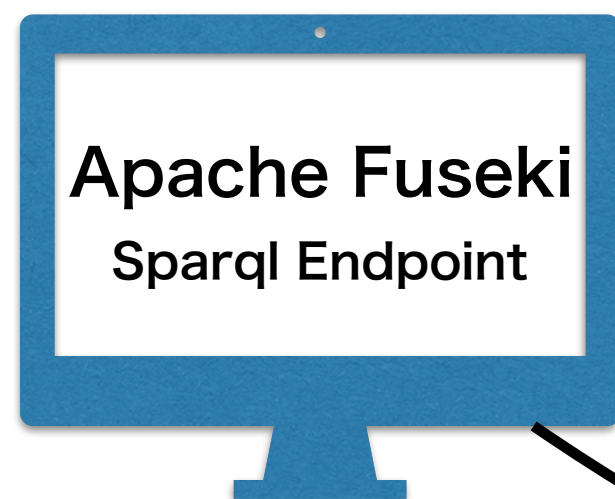


Generate define.xml

This tool is for the proof of concept that is the automatically generating of a high-quality define.xml from Clinical Trials Data and other metadata as RDF through SPARQL query.

<https://github.com/i-akiya/CTDasRDF/tree/master/Define-XML>





Apache Fuseki
Sparql Endpoint

Posting Sparql Query

Retrieve the metadata for
generating defile.xml

Generate define.xml

Groovy Environment

Sparql Query
Generator

XML Builder

Java Runtime Environment

Operating System

SDTM-IG 3.1.3
Study Data Reviewer's Guide
» Tabulation Datasets
» Comments

Standard: SDTM-IG 3.1.3
Study Name: CDISCPILOT01
Study Description: Study CDISCPILOT01
Protocol Name: CDISCPILOT01
Metadata Name: CDISCPILOT01, SDTM Data Definitions
Metadata Description: CDISCPILOT01, SDTM Data Definitions

Date of Define-XML document generation: 2018-02-20T12:10:31
Stylesheet version: 2015-01-16

Tabulation Datasets for Study CDISCPILOT01 (SDTM-IG 3.1.3)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
DM	Demographics	Special Purpose Domain	One record per subject	Tabulation		dm.xml	
SUPPDM	Supplemental Qualifiers for SUPPDM	Relationship Dataset	One record per IDVAR, IDVARVAL, and QNAM value per subject	Tabulation		suppdm.xml	
VS	Vital Signs	Findings Observation Class	One record per vital sign measurement per time point per visit per subject	Tabulation		vs.xml	

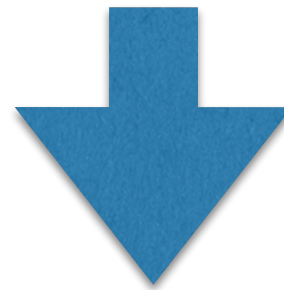
Go to the [top](#) of the define.xml

Demographics (DM) [Location: dm.xml]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier						
DOMAIN	Domain Abbreviation					Assigned	
USUBJID	Unique Subject Identifier						Concatenation of STUDYID, DM.SITEID and DM.SUBJID
SUBJID	Subject Identifier for						

Sparql Query Generator

```
SELECT ?Domain ?DatasetLabel ?Structure ?defClass
WHERE{
sdtmig-3-1-3:Dataset.${Dataset} mms:context ?Context;
                                mms:contextName ?Domain;
                                mms:contextLabel ?DatasetLabel;
                                cdiscs:datasetStructure ?Structure.
?Context mms:contextLabel ?defClass.
}
```



**Assigned \${Dataset}="DM"
and
Post the Query**

Query Result

QUERY RESULTS				
	Raw Response	Table		
	Search:		Show	50 entries
Domain	DatasetLabel	Structure	defClass	
1 "DM"	"Demographics"	"One record per subject"	"Special Purpose Domain"	
Showing 1 to 1 of 1 entries				

Groovy XML Builder

```
'OID' : "IG.${dsname}",
      'Name' : dsname,
      'Domain' : dsname,
      'SASDatasetName' : dsname,
      'Repeating' : repeating,
      'IsReferenceData' : reference,
      'Purpose' : "Tabulation",
      'def:Structure' : sol.Structure,
      'def:Class' : sol.defClass,
      'def:CommentOID' : "",
      'def:ArchiveLocationID' : "LF.${dsname}"
    ) {
      'Description' ({ 'TranslatedText' ( 'xml:lang' : "en", dslabel ) })
        for (ResultSet variableResultset = variableMetadataQE.execSelect();
variableResultset.hasNext() ; ) {
          QuerySolution sol2 = variableResultset.nextSolution()
            'ItemRef' (ItemOID: "IT.${datasetName}.${sol2.dataElementName}",
OrderNumber: sol2.ordinal_, Mandatory: sol2.Core, KeySequence: "", MethodOID: "")
          }
          'def:leaf' ( 'ID' : "LF.${dsname}", 'xlink:href' : "$
${dsname}.xpt".toLowerCase(), { 'def:title' ( "${dsname}.xpt".toLowerCase() ) }
        )
    }
```

The highlighted fragment in **red** indicate that the values from Sparql Query.

Groovy XML Builder

Generated xml code

```
<ItemGroupDef OID="IG.DM" Name="DM" Domain="DM" SASDatasetName="DM" Repeating="Yes"
IsReferenceData="No" Purpose="Tabulation" def:Structure="One record per subject"
def:Class="Special Purpose Domain" def:ArchiveLocationID="LF.DM">
  <Description>
    <TranslatedText xml:lang="en">Demographics</TranslatedText>
  </Description>
  <ItemRef ItemOID="IT.DM.STUDYID" Mandatory="Yes" />
  <ItemRef ItemOID="IT.DM.DOMAIN" Mandatory="Yes" />
  <ItemRef ItemOID="IT.DM.USUBJID" Mandatory="Yes" />
  <ItemRef ItemOID="IT.DM.SUBJID" Mandatory="Yes" />
  <ItemRef ItemOID="IT.DM.RFSTDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.RFENDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.RFXSTDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.RFXENDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.RFICDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.RFPENDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.DTHDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.DTHFL" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.SITEID" Mandatory="Yes" />
  <ItemRef ItemOID="IT.DM.INVID" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.BRTHDTC" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.AGE" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.AGEU" Mandatory="No" />
  <ItemRef ItemOID="IT.DM.SEX" Mandatory="Yes" />
```



DEMO

