Use R RDF version of CDISC standards

PhuseSubTeamAnalysisResults@example.org

2016-02-07

Introduction

This vignette shows how to - load the CDISC standards from the packages - make a SPARQL query for extracting a codelist - make a SPARQL query for extracting the definition for the SDTM DM dataset

Setup

```
library(rrdfancillary)
```

```
## Loading required package: rJava

## Loading required package: methods

## Loading required package: rrdf

## Loading required package: rrdflibs
```

```
require("rrdfcdisc")
```

```
## Loading required package: rrdfcdisc

## Loading required package: RCurl

## Loading required package: bitops

##
## Attaching package: 'RCurl'

## The following object is masked from 'package:rJava':
##
##     clone

## Loading required package: devtools
```

Loading using Load.cdisc.standards

The code below shows the call for loading the zipped rdf file and store in a triple store. This is currently handled in the .onLoad function for the package. The rrdf model is then stored in the package environment.

```
cdisc.rdf<- Load.cdisc.standards()
```

```
## Start loading rdf.cdisc.org contents from /home/ma/R/x86_64-redhat-linux-gnu-library/3.2/rrdfqbcrnd0/
```

```
## Done loading rdf.cdisc.org contents, number of triples: 165700
```

```
message(".. total number of triples: ", summarize.rdf.noprint(cdisc.rdf) )
```

```
## .. total number of triples: 165700
```

Extract codelist

Note: consider using `qbCDISCprefixes` .

```
nciDomainValue<- "C66731"
query <- GetCDISCCodeListSparqlQuery( Get.rq.prefixlist.df(qbCDISCprefixes), nciDomainValue )
codeSource <- as.data.frame(sparql.rdf(cdisc.rdf, query),stringsAsFactors=FALSE)

knitr::kable(codeSource)
```

nciDomain	cdiscDefinition
sdtm:C66731.C20197	A person who belongs to the sex that normally produces sperm. The term is used to indicate biolog
sdtm:C66731.C17998	Not known, not observed, not recorded, or refused. (NCI)
sdtm:C66731.C45908	A person (one of unisexual specimens) who is born with genitalia and/or secondary sexual character
sdtm:C66731.C16576	A person who belongs to the sex that normally produces ova. The term is used to indicate biologica

Extract the definition for the SDTM DM dataset

```
query.rq<-'
select *
where {
  ?column a <http://rdf.cdisc.org/mms#Column> .
  ?column <http://rdf.cdisc.org/mms#context> ?dataset .
  ?dataset a <http://rdf.cdisc.org/mms#Dataset>.
  optional { ?column <http://rdf.cdisc.org/mms#dataElementLabel> ?label }
  optional { ?column <http://rdf.cdisc.org/std/schema#controlledTermsOrFormat> ?controlledTermsOrFormat
  optional { ?column <http://rdf.cdisc.org/std/schema#dataElementCompliance> ?dataElementCompliance }
  optional { ?column <http://rdf.cdisc.org/std/schema#dataElementRole> ?dataElementRole }
  optional { ?column <http://rdf.cdisc.org/std/schema#dataElementType> ?dataElementType }
  optional { ?column <http://rdf.cdisc.org/std/schema#references> ?references }
  optional { ?column <http://rdf.cdisc.org/mms#dataElement> ?dataElement }
  optional { ?column <http://rdf.cdisc.org/mms#dataElementDescription> ?dataElementDescription }
  optional { ?column <http://rdf.cdisc.org/mms#dataElementName> ?dataElementName }
  optional { ?column <http://rdf.cdisc.org/mms#dataElementValueDomain> ?dataElementValueDomain }
  optional { ?column <http://rdf.cdisc.org/mms#ordinal> ?ordinal }
  values (?dataset) {(<http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM>)}
  }
  order by ?dataset ?ordinal
```

```
'
datasetDef <- as.data.frame(sparql.rdf(cdisc.rdf, query.rq),stringsAsFactors=FALSE)
```

Here are some of the columns in the result:

```
knitr::kable(datasetDef[,c("dataset", "ordinal", "dataElementName", "label")])
```

dataset	ordinal	dataElementName	label
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	1	STUDYID	Study Identifier
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	2	DOMAIN	Domain Abbreviation
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	3	USUBJID	Unique Subject Identifier
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	4	SUBJID	Subject Identifier for the Study
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	5	RFSTDTC	Subject Reference Start Date/Time
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	6	RFENDTC	Subject Reference End Date/Time
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	7	RFXSTDTC	Date/Time of First Study Treatment
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	8	RFXENDTC	Date/Time of Last Study Treatment
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	9	RFICDTC	Date/Time of Informed Consent
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	10	RFPENDTC	Date/Time of End of Participation
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	11	DTHDTC	Date/Time of Death
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	12	DTHFL	Subject Death Flag
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	13	SITEID	Study Site Identifier
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	14	INVID	Investigator Identifier
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	15	INVNAM	Investigator Name
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	16	BRTHDTC	Date/Time of Birth
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	17	AGE	Age
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	18	AGEU	Age Units
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	19	SEX	Sex
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	20	RACE	Race
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	21	ETHNIC	Ethnicity
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	22	ARMCD	Planned Arm Code
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	23	ARM	Description of Planned Arm
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	24	ACTARMCD	Actual Arm Code
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	25	ACTARM	Description of Actual Arm
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	26	COUNTRY	Country
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	27	DMDTC	Date/Time of Collection
http://rdf.cdisc.org/std/sdtmig-3-1-3#Dataset.DM	28	DMDY	Study Day of Collection