Create ADAE as TTL file

mja@statgroup.dk2016-03-08

Contents

Create turtle version of ADAE dataset	1
Setup	1

Create turtle version of ADAE dataset

Setup

```
library(rrdfancillary)
devtools::load_all(pkg="../..")
```

Loading rrdfqbcrndex

This script creates a turtle version of the ADAE dataset.

The script uses SQLite to export the ADAE dataset, and d2rq (http://d2rq.org/) to create a turle version of the file. In the code below it is assumed that d2rq is installed in the directory /opt/d2rq-0.8.1.

```
library(foreign)
fnadae<- system.file("extdata/sample-xpt", "adae.xpt", package="rrdfqbcrndex")
adae<-read.xport(fnadae)</pre>
```

The present code uses read.xport to input the sas dataset. TODO: Date variables are note handled correctly. Other methods are available allowing access to more of the meta data in the datasets.

Next step is to dump the data as a SQL dump, that will serve as input to D2RQ. For dumping the data sqlite is used as it is the easies to setup. The R-package RSQLite (http://cran.r-project.org/web/packages/RSQLite.pdf) provides access to sqlite.

```
## install.packages("RSQLite")
```

The SQL dump is created by transfering the R data frame to SQLite, and the invoking sqlite to create a dump.

```
library(RSQLite)
## create a SQLite instance and create one connection.
m <- dbDriver("SQLite")
## initialize a new database to a tempfile and copy some data.frame
## from the base package into it
tfile <- tempfile()
con <- dbConnect(m, dbname = tfile)
dbWriteTable(con, "adae", adae)</pre>
```

```
dbDisconnect(con)
```

[1] TRUE

```
cat("SQLite database stored as ", tfile, "\n")
```

SQLite database stored as /tmp/Rtmp9tloU2/file1e4b7472e965

```
dumpFn<- tempfile()
system(paste("sqlite3", tfile, ".dump >", dumpFn, sep=" "))
cat("SQLite database dump in ", dumpFn, "\n")
```

SQLite database dump in /tmp/Rtmp9tloU2/file1e4b262ac97a

Next step is to process the dump, so the SQL can be used as input to d2rq. The changes applied are: * change TEXT to VARCHAR(1000) * remove top 2 lines with PRAGMA * in insert statements replace "adae" with adae * after "AOCC01FL" TEXT add a comma (",") and a new line with PRIMARY KEY (USUBJID, AESEQ)

This could maybe be simplified using RSQlite on dbSendQuery-methods, see (https://stat.ethz.ch/pipermail/r-sig-db/2010q1/000813.html). For d2rq documentation see (file:///opt/d2rq-0.8.1/doc/dump-rdf.html).

SQLite database dump modified stored as /tmp/Rtmp9tloU2/file1e4b2c4af80b

```
" -b ", d2rqbaseURL,
" -o ", adaettlFn,
" -l ", dumpAfterSedFn,
" ", adaemapttlFn
))

targetDir<- system.file("extdata/sample-rdf", package="rrdfqbcrndex")
if (file.copy(adaemapttlFn, targetDir, overwrite = TRUE)) {
    cat( "File ", adaemapttlFn, " copied to directory ", targetDir, "\n")
}</pre>
```

File /tmp/Rtmp9tloU2/adae-map.ttl copied to directory /home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/

```
if (file.copy(adaettlFn, targetDir, overwrite = TRUE)) {
cat( "File ", adaettlFn, " copied to directory ", targetDir, "\n")
}
```

 $\verb|## File /tmp/Rtmp9tloU2/adae.ttl copied to directory /home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/instable for the file of t$

```
### This could maybe also work ..
## m <- dbDriver("SQLite")
## tfile <- tempfile()
## con <- dbConnect(m, dbname = tfile)
## s <- sprintf("create table %s(%s, primary key(%s))", "adae",
## paste(names(adae), collapse = ", "),
## paste(names(adae)[c(1,2)],collapse=", ")
## dbGetQuery(con, s)
## dbWriteTable(con, "adae", adae,append=TRUE)
## dbDisconnect(con)
## tfile</pre>
```

Read in the file

This section demostrates how to access the data for the ADAE dataset.

First the data are loaded into a RDF model.

```
dataFilemap<- system.file("extdata/sample-rdf", "adae-map.ttl", package="rrdfqbcrndex")
dataFilemap</pre>
```

[1] "/home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/extdata/sample-rdf/adae-map.ttl"

```
dataFile<- system.file("extdata/sample-rdf", "adae.ttl", package="rrdfqbcrndex")
dataFile</pre>
```

[1] "/home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/extdata/sample-rdf/adae.ttl"

```
store <- new.rdf() # Initialize
invisible(load.rdf(dataFilemap, format="TURTLE", appendTo= store))
invisible(load.rdf(dataFile, format="TURTLE", appendTo= store))
summarize.rdf(store)</pre>
```

```
## [1] "Number of triples: 61002"
```

Next step is to make a query showing the mappings

```
resmapping<- data.frame(sparql.rdf(store, mapping.rq),stringsAsFactors=FALSE)
str(resmapping)</pre>
```

```
## 'data.frame': 55 obs. of 3 variables:
## $ mapColumn : chr "file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AOCCPFL
## $ d2rqcolumn : chr "ADAE.AOCCPFL" "ADAE.AESCAN" "ADAE.AENDT" "ADAE.ADURU" ...
## $ d2rqdatatype: chr NA NA "xsd:double" NA ...
```

knitr::kable(resmapping)

mapColumn	d2rq c olumn	d2rqc
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AOCCPFL file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESCAN	ADAE.AOCCPFL ADAE.AESCAN	NA NA

mapColumn	d2rqcolumn	d2rqc
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AENDT	ADAE.AENDT	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_ADURU	ADAE.ADURU	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_SITEID	ADAE.SITEID	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AOCCSFL	ADAE.AOCCSFL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_RACEN	ADAE.RACEN	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESOC	ADAE.AESOC	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AOCC01FL	ADAE.AOCC01FL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESCONG	ADAE.AESCONG	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEACN	ADAE.AEACN	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEHLTCD	ADAE.AEHLTCD	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESEQ	ADAE.AESEQ	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_TRTA	ADAE.TRTA	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AELLTCD	ADAE.AELLTCD	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AOCC03FL	ADAE.AOCC03FL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_TRTEMFL	ADAE.TRTEMFL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEPTCD	ADAE.AEPTCD	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AENDY	ADAE.AENDY	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESOCCD	ADAE.AESOCCD	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AGEGR1	ADAE.AGEGR1	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESLIFE	ADAE.AESLIFE	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE RACE	ADAE.RACE	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AELLT	ADAE.AELLT	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESDTH	ADAE.AESDTH	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_SEX	ADAE.SEX	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_USUBJID	ADAE.USUBJID	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AOCC02FL	ADAE.AOCC02FL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AOCCFL	ADAE.AOCCFL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_TRTSDT	ADAE.TRTSDT	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_CQ01NAM	ADAE.CQ01NAM	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEHLGT	ADAE.AEHLGT	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESEV	ADAE.AESEV	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEOUT	ADAE.AEOUT	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESDISAB	ADAE.AESDISAB	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESOD	ADAE.AESOD	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEDECOD	ADAE.AEDECOD	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AGEGR1N	ADAE.AGEGR1N	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEREL	ADAE.AEREL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESER	ADAE.AESER	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AETERM	ADAE.AETERM	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_ASTDTF	ADAE.ASTDTF	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEBODSYS	ADAE.AEBODSYS	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AGE	ADAE.AGE	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_ASTDY	ADAE.ASTDY	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_SAFFL	ADAE.SAFFL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEHLGTCD	ADAE.AEHLGTCD	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEHEGTOD	ADAE.STUDYID	NA
file:///home/ma/projects/rrdfqbcrndo/rrdfqbcrndex/inst/data-raw/#ADAE_STODTID	ADAE.AOCC04FL	NA
file:///home/ma/projects/rrdfqbcrndo/rrdfqbcrndex/inst/data-raw/#ADAE_AOCC04rL	ADAE.AOCC04FL ADAE.TRTAN	xsd:d
file:///home/ma/projects/rrdfqbcrndo/rrdfqbcrndex/inst/data-raw/#ADAE_ADURN	ADAE.ADURN	xsd:d
	ADAE.AESHOSP	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AESHOSP file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_TRTEDT	ADAE.AESHOSP ADAE.TRTEDT	xsd:d
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_ASTDT	ADAE.ASTDT	xsd:d

mapColumn	d2rq c olumn	d2rqd
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADAE_AEHLT	ADAE.AEHLT	NA

```
The code below gets the all values in one record
s<- paste0("<", d2rqbaseURL,c("ADAE/01-701-1023/2.0E0"), ">")
records.rq<-paste("select * ",
           paste("(",s,")", collapse="\n"),
           "}", sep="\n", collapse="\n")
cat(records.rq,"\n")
## select *
## where { ?s ?p ?o.
## values(?s) {
## ( <http://www.example.org/datasets/ADAE/01-701-1023/2.0E0> )
## }
## }
records.res <- data.frame(sparql.rdf(store, records.rq),stringsAsFactors=FALSE)
str(records.res)
## 'data.frame':
                    50 obs. of 3 variables:
              "http://www.example.org/datasets/ADAE/01-701-1023/2.0E0" "http://www.example.org/datasets
    $ s: chr
              "http://www.example.org/datasets/vocab/ADAE_AEHLT" "http://www.example.org/datasets/vocab
    $ p: chr
              "HLT 0284" "Y" "Placebo" "N" ...
records.res$pclean<- gsub(paste0(d2rqbaseURL, "vocab/ADAE_"), "", records.res$p)
knitr::kable(records.res[,c("s", "p","o")])
```

```
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE AEHLT
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_TRTEM
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE TRTA
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_AESCON
                                                       http://www.example.org/datasets/vocab/ADAE_ADURU
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_AOCCSI
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_SEX
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_AESOC
                                                       http://www.example.org/datasets/vocab/ADAE_AESER
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_TRTED?
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_AEBODS
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_SAFFL
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_AESDTI
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_AGE
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.example.org/datasets/vocab/ADAE_AOCC01
http://www.example.org/datasets/ADAE/01-701-1023/2.0E0
                                                       http://www.w3.org/1999/02/22-rdf-syntax-ns#type
```

http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE ASTDY http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE STUDYI http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE ASTDT http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AEREL http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AEHLG http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE SITEID http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE_TRTSD7 http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE_ASTDTI http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AESCAM http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AGEGR http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AOCC02 http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AELLT http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AESEV http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE_TRTAN http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AESOD http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE USUBJI http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AEACN http://www.example.org/datasets/vocab/ADAE AEOUT http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AEDECO http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE CQ01NA http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AESEQ http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AESHOS http://www.example.org/datasets/ADAE/01-701-1023/2.0E0http://www.example.org/datasets/vocab/ADAE_AESLIF http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AOCCPI http://www.example.org/datasets/ADAE/01-701-1023/2.0E0http://www.example.org/datasets/vocab/ADAE AESDIS. http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AOCC03 http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AOCCF1 http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AOCC04 http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AETERI http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE RACE http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.w3.org/2000/01/rdf-schema#label http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE AGEGR http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.example.org/datasets/vocab/ADAE RACEN http://www.example.org/datasets/ADAE/01-701-1023/2.0E0 http://www.w3.org/1999/02/22-rdf-syntax-ns#type