Create ADSL as TTL file

mja@statgroup.dk2016-02-02

Contents

Create turtle version of ADSL dataset	1
Setup	1

Create turtle version of ADSL dataset

Setup

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

Loading rrdfqbcrndex

This script creates a turtle version of the ADSL dataset.

The script uses SQLite to export the ADSL 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)
fnadsl<- system.file("extdata/sample-xpt", "adsl.xpt", package="rrdfqbcrndex")
adsl<-read.xport(fnadsl)</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, "adsl", adsl)</pre>
```

```
dbDisconnect(con)
```

[1] TRUE

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

SQLite database stored as /tmp/RtmpDGdzpS/file27213e85fedd

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

SQLite database dump in /tmp/RtmpDGdzpS/file27217334e2e2

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 "adsl" with adsl * after "MMSETOT" REAL add a comma (",") and a new line with PRIMARY KEY (USUBJID)

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/RtmpDGdzpS/file272168506a56

```
" -o ", adslttlFn,
" -l ", dumpAfterSedFn,
" ", adslmapttlFn
))

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

File /tmp/RtmpDGdzpS/adsl-map.ttl copied to directory /home/ma/projects/R-projects/rrdfqbcrnd0/rr

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

Read in the file

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

First the data are loaded into a RDF model.

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

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

[1] "/home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/extdata/sample-rdf/adsl.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: 13160"

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': 48 obs. of 3 variables:
## $ mapColumn : chr "file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#A
## $ d2rqcolumn : chr "ADSL.USUBJID" "ADSL.TRT01A" "ADSL.TRT01P" "ADSL.STUDYID" ...
## $ d2rqdatatype: chr NA NA NA NA ...
```

knitr::kable(resmapping)

m mapColumn d2rqcolumn

file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_USUBJID file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_TRT01A

ADSL.USUBJID ADSL.TRT01A mapColumn d2rqcolumn file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL TRT01P ADSL.TRT01P file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL STUDYID ADSL.STUDYID file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL DISCONFL ADSL.DISCONFI file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL AGEU ADSL.AGEU ADSL.ITTFL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL ITTFL $file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_DCDECOD$ ADSL.DCDECOL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_BMIBL ADSL.BMIBL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_DURDIS ADSL.DURDIS file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_SUBJID ADSL.SUBJID file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL ETHNIC ADSL.ETHNIC file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL EFFFL ADSL.EFFFL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL SITEGR1 ADSL.SITEGR1 file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_COMP8FL ADSL.COMP8FL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL TRTSDT ADSL.TRTSDT file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL SEX ADSL.SEX file: ///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL-WEIGHTBL-raw/filesADSL.WEIGHTE file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL AGE ADSL.AGE file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL CUMDOSE ADSL.CUMDOSE file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL COMP16FL ADSL.COMP16F3 file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL AGEGR1 ADSL.AGEGR1 file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL BMIBLGR1 ADSL.BMIBLGR file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL COMP24FL ADSL.COMP24F file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_TRT01PN ADSL.TRT01PN ADSL.RFENDT file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_RFENDT file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_DURDSGR1 ADSL.DURDSGR file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_RFSTDTC ADSL.RFSTDTC file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL AVGDD ADSL.AVGDD file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL TRTDUR ADSL.TRTDUR file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL HEIGHTBL ADSL.HEIGHTB file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_AGEGR1N ADSL.AGEGR1N file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL RACEN ADSL.RACEN file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL RFENDTC ADSL.RFENDTC file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL DISONSDT ADSL.DISONSDT file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL SITEID ADSL.SITEID $file: ///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL \ EDUCLVL$ ADSL.EDUCLVL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL DSRAEFL ADSL.DSRAEFL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL ARM ADSL.ARM file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL SAFFL ADSL.SAFFL file:///home/ma/projects/R-projects/rrdfgbcrnd0/rrdfgbcrndex/inst/data-raw/#ADSL_RACE ADSL.RACE file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_TRT01AN ADSL.TRT01AN file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL MMSETOT ADSL.MMSETO file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_VISNUMEN ADSL.VISNUME file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_VISIT1DT ADSL.VISIT1DT file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL DTHFL ADSL.DTHFL file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL DCREASCD ADSL.DCREASC file:///home/ma/projects/R-projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL TRTEDT ADSL.TRTEDT

The code below gets the all values in one record

```
s<- paste0("<", d2rqbaseURL,c("ADSL/01-718-1254"), ">")
records.rq<-paste("select * ",</pre>
```

```
paste("(",s,")", collapse="\n"),
           "}", sep="\n", collapse="\n")
cat(records.rq,"\n")
## select *
## where { ?s ?p ?o.
   values(?s) {
## ( <http://www.example.org/datasets/ADSL/01-718-1254> )
## }
## }
records.res<- data.frame(sparql.rdf(store, records.rq),stringsAsFactors=FALSE)
str(records.res)
   'data.frame':
                    51 obs. of 3 variables:
              "http://www.example.org/datasets/ADSL/01-718-1254" "http://www.example.org/datasets/ADSL/
    $ s: chr
              "http://www.example.org/datasets/vocab/ADSL_EFFFL" "http://www.example.org/datasets/vocab
    $ p: chr
    $ o: chr
              "Y" "Completed" "54.0E0" "Y" ...
records.res$pclean<- gsub(paste0(d2rqbaseURL, "vocab/ADSL_"), "", records.res$p)
knitr::kable(records.res[,c("s", "p","o")])
```

```
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL EFFFL
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_DCREASCD
                                                 http://www.example.org/datasets/vocab/ADSL_TRT01PN
http://www.example.org/datasets/ADSL/01-718-1254
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL\_COMP8FL
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL CUMDOSE
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_TRTSDT
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL SITEGR1
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_TRT01AN
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL AVGDD
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL DURDIS
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL TRTDUR
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL HEIGHTBL
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL COMP16FL
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_DCDECOD
                                                 http://www.example.org/datasets/vocab/ADSL\_WEIGHTBL
http://www.example.org/datasets/ADSL/01-718-1254
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL BMIBLGR1
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_AGE
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL SUBJID
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_TRT01A
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_DURDSGR1
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_RFSTDTC
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL SEX
                                                 http://www.example.org/datasets/vocab/ADSL\_DISCONFL
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL DSRAEFL
http://www.example.org/datasets/ADSL/01-718-1254
http://www.example.org/datasets/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_COMP24FL
```

1

1

S	p
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_RFENDT
http://www.example.org/datasets/ADSL/01-718-1254	http://www.w3.org/2000/01/rdf-schema#label
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_SAFFL
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AGEU
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_ETHNIC
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AGEGR1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_DISONSDT
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRTEDT
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_DTHFL
http://www.example.org/datasets/ADSL/01-718-1254	$http://www.example.org/datasets/vocab/ADSL_MMSETOT$
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_VISNUMEN
http://www.example.org/datasets/ADSL/01-718-1254	$http://www.example.org/datasets/vocab/ADSL_USUBJID$
http://www.example.org/datasets/ADSL/01-718-1254	$http://www.example.org/datasets/vocab/ADSL_ITTFL$
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_BMIBL
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_RFENDTC
http://www.example.org/datasets/ADSL/01-718-1254	http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_STUDYID
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_RACEN
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_VISIT1DT
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AGEGR1N
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRT01P
http://www.example.org/datasets/ADSL/01-718-1254	$http://www.example.org/datasets/vocab/ADSL_RACE$
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_ARM
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_SITEID
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_EDUCLVL
http://www.example.org/datasets/ADSL/01-718-1254	http://www.w3.org/1999/02/22-rdf-syntax-ns#type