Create ADSL as TTL file

mja@statgroup.dk2016-04-26

Contents

Create turtle version of ADSL dataset	1
Setup	1

Create turtle version of ADSL dataset

Setup

```
library(rrdfancillary)
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/RtmpcsNlsC/file29a465099aa9

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

SQLite database dump in /tmp/RtmpcsNlsC/file29a4779aad64

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/RtmpcsNlsC/file29a43e35de7

```
" -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/RtmpcsNlsC/adsl-map.ttl copied to directory /home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/

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

File /tmp/RtmpcsNlsC/adsl.ttl copied to directory /home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst

Note: consider using Produce W3C Direct Mapping (https://www.w3.org/TR/rdb-direct-mapping/) - this can be done using the option -w3c.

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

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>
```

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

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

[1] "/home/ma/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/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_HEIGHTB
## $ d2rqcolumn : chr "ADSL.HEIGHTBL" "ADSL.AVGDD" "ADSL.DCDECOD" "ADSL.TRT01P" ...
## $ d2rqdatatype: chr "xsd:double" "xsd:double" NA NA ...
```

knitr::kable(resmapping)

mapColumn	d2rqcolumn	d2rqda
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_HEIGHTBL$	ADSL.HEIGHTBL	xsd:doi
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_AVGDD$	ADSL.AVGDD	xsd:do
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_DCDECOD$	ADSL.DCDECOD	NA
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_TRT01P$	ADSL.TRT01P	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_SITEID$	ADSL.SITEID	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_COMP16FL	ADSL.COMP16FL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_COMP8FL	ADSL.COMP8FL	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_AGEGR1N$	ADSL.AGEGR1N	xsd:do:
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_RFENDT$	ADSL.RFENDT	xsd:doi
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_ETHNIC$	ADSL.ETHNIC	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_BMIBL$	ADSL.BMIBL	xsd:doi
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_RACEN	ADSL.RACEN	xsd:doi
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_DTHFL	ADSL.DTHFL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_DISCONFL	ADSL.DISCONFL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_TRT01A	ADSL.TRT01A	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_DURDSGR1	ADSL.DURDSGR1	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_TRT01PN	ADSL.TRT01PN	xsd:doi
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_EFFFL	ADSL.EFFFL	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_VISNUMEN	ADSL.VISNUMEN	xsd:doi
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_ARM	ADSL.ARM	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_BMIBLGR1	ADSL.BMIBLGR1	NA
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_AGE	ADSL.AGE	xsd:doi
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_VISIT1DT$	ADSL.VISIT1DT	xsd:doi
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_RFSTDTC	ADSL.RFSTDTC	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_SUBJID$	ADSL.SUBJID	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_EDUCLVL$	ADSL.EDUCLVL	xsd:doi
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_MMSETOT$	ADSL.MMSETOT	xsd:doi
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_SAFFL$	ADSL.SAFFL	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_RACE$	ADSL.RACE	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_DISONSDT$	ADSL.DISONSDT	xsd:do:
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_AGEU$	ADSL.AGEU	NA
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_TRTEDT$	ADSL.TRTEDT	xsd:doı
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_RFENDTC$	ADSL.RFENDTC	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_AGEGR1$	ADSL.AGEGR1	NA
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_COMP24FL$	ADSL.COMP24FL	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_WEIGHTBL$	ADSL.WEIGHTBL	xsd:do:
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_DSRAEFL$	ADSL.DSRAEFL	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_ITTFL$	ADSL.ITTFL	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_STUDYID$	ADSL.STUDYID	NA
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_SITEGR1$	ADSL.SITEGR1	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_TRT01AN$	ADSL.TRT01AN	xsd:doı
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_TRTSDT$	ADSL.TRTSDT	xsd:doı
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_CUMDOSE$	ADSL.CUMDOSE	xsd:do:
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_SEX$	ADSL.SEX	NA
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_DCREASCD$	ADSL.DCREASCD	NA
$file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_TRTDUR$	ADSL.TRTDUR	xsd:do:
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_USUBJID$	ADSL.USUBJID	NA
$file: ///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/\#ADSL_DURDIS$	ADSL.DURDIS	xsd:doi

The code below gets the all values in one record

```
s<- paste0("<", d2rqbaseURL,c("ADSL/01-718-1254"), ">")
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/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
    $ p: chr
              "http://www.example.org/datasets/vocab/ADSL_EFFFL" "http://www.example.org/datasets/vocab
##
              "Y" "Completed" "54.0E0" "Y" ...
    $ o: chr
records.res$pclean<- gsub(paste0(d2rqbaseURL, "vocab/ADSL_"), "", records.res$p)
knitr::kable(records.res[,c("s", "p","o")])
```

```
\mathbf{S}
                                                 p
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/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_TRT01PN
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/ADSL/01-718-1254
                                                 http://www.example.org/datasets/vocab/ADSL_WEIGHTBL
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/vocab/ADSL_RFSTDTC
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_SEX
```

Ĉ

S	p
http://www.example.org/datasets/ADSL/01-718-1254	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/vocab/ADSL_COMP24FL
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