Create ADSL as TTL file

mja@statgroup.dk

2016-05-15

Contents

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)
```

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/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)
```

```
## [1] TRUE
```

```
dbDisconnect(con)
```

```
## [1] TRUE
```

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

```
## SQLite database stored as  /tmp/RtmpgwUgO5/file46de41d51343
```

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

```
## SQLite database dump in  /tmp/RtmpgwUgO5/file46de1f942a06
```

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).

```
dumpAfterSedFn <- tempfile()
sedCmd<- paste("sed",
               "-e 's/TEXT/VARCHAR(1000)/g; s/\"MMSETOT\" REAL/\"MMSETOT\" REAL, PRIMARY KEY (USUBJID)/
               dumpFn, ">", dumpAfterSedFn, sep=" ")
system(sedCmd)
cat("SQLite database dump modified stored as ", dumpAfterSedFn, "\n")
```

```
## SQLite database dump modified stored as  /tmp/RtmpgwUgO5/file46def4f6c57
```

```
## Check only expected changes were applied
## not run due to large output
## system(paste("diff", dumpFn, dumpAfterSedFn, sep=" "))


adslmapttlFn<- file.path(tempdir(), "adsl-map.ttl")
adslttlFn<- file.path(tempdir(), "adsl.ttl")

d2rqbaseURL<- "http://www.example.org/datasets/"

## -b option does not work with -l
## /opt/d2rq-0.8.1/generate-mapping reports Unknown argument: -b
system(paste("/opt/d2rq-0.8.1/generate-mapping",
             " -l ", dumpAfterSedFn,
             " -o ", adslmapttlFn
             ))

system(paste("/opt/d2rq-0.8.1/dump-rdf",
             " -b ", d2rqbaseURL,
```

```
            " -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")
}
```

File /tmp/RtmpgwUgO5/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/RtmpgwUgO5/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
```

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
```

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

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

```
## [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)
```

```
## [1] "Number of triples: 13160"
```

Next step is to make a query showing the mappings

```
mapping.rq<-"
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#>
prefix jdbc: <http://d2rq.org/terms/jdbc/>
select * where {
   ?mapColumn a d2rq:PropertyBridge ;
              d2rq:column ?d2rqcolumn .
   optional{ ?mapColumn d2rq:datatype ?d2rqdatatype  }
}
"

cat(mapping.rq,"\n")
```

```
##
## prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
## prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
## prefix xsd: <http://www.w3.org/2001/XMLSchema#>
## prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#>
## prefix jdbc: <http://d2rq.org/terms/jdbc/>
## select * where {
##    ?mapColumn a d2rq:PropertyBridge ;
##               d2rq:column ?d2rqcolumn .
##    optional{ ?mapColumn d2rq:datatype ?d2rqdatatype  }
## }
##
```

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

```
## 'data.frame':    48 obs. of  3 variables:
## $ mapColumn   : chr  "file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_HEIGHTBL
## $ 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:dou
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_AVGDD	ADSL.AVGDD	xsd:dou
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:dou
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_RFENDT	ADSL.RFENDT	xsd:dou
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:dou
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_RACEN	ADSL.RACEN	xsd:dou
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:dou
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:dou
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:dou
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_VISIT1DT	ADSL.VISIT1DT	xsd:dou
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:dou
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_MMSETOT	ADSL.MMSETOT	xsd:dou
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:dou
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:dou
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:dou
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:dou
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_TRTSDT	ADSL.TRTSDT	xsd:dou
file:///home/ma/projects/rrdfqbcrnd0/rrdfqbcrndex/inst/data-raw/#ADSL_CUMDOSE	ADSL.CUMDOSE	xsd:dou
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:dou
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:dou

The code below gets the all values in one record

```
s<- paste0("<", d2rqbaseURL,c("ADSL/01-718-1254"), ">")

records.rq<-paste("select * ",
          "where { ?s ?p ?o.",
          " values(?s) {",
          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:
## $ s: chr  "http://www.example.org/datasets/ADSL/01-718-1254" "http://www.example.org/datasets/ADSL/(
## $ p: chr  "http://www.example.org/datasets/vocab/ADSL_EFFFL" "http://www.example.org/datasets/vocab/
## $ 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")])
```

s	p	o
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_EFFFL	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_DCREASCD	C
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRT01PN	5
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_COMP8FL	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_CUMDOSE	9
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRTSDT	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_SITEGR1	7
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRT01AN	5
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AVGDD	5
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_DURDIS	2
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRTDUR	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_HEIGHTBL	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_COMP16FL	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_DCDECOD	C
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_WEIGHTBL	8
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_BMIBLGR1	2
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AGE	7
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_SUBJID	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRT01A	Y
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	2
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_SEX	M

s	p	o
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	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_RFENDT	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.w3.org/2000/01/rdf-schema#label	A
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_SAFFL	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AGEU	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_ETHNIC	H
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AGEGR1	6
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_DISONSDT	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRTEDT	1
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	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_VISNUMEN	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_USUBJID	0
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_ITTFL	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_BMIBL	2
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_RFENDTC	2
http://www.example.org/datasets/ADSL/01-718-1254	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	h
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_STUDYID	C
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_RACEN	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_VISIT1DT	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_AGEGR1N	2
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_TRT01P	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_RACE	V
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_ARM	Y
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_SITEID	7
http://www.example.org/datasets/ADSL/01-718-1254	http://www.example.org/datasets/vocab/ADSL_EDUCLVL	1
http://www.example.org/datasets/ADSL/01-718-1254	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	h