Modeling Analysis Results and Metadata as Linked Data

PhUSE Single Day Event
Frenchtown, NJ, USA
4th August 2016

Presented by
Marc Andersen
StatGroup ApS, Denmark
mja@statgroup.dk

In place of
Marcelina Hungria,
Dlcore Group

The Analysis Results and Metadata (AR&M) Team is a sub-team of the PhUSE/FDA-CSC Semantic Technologies Working Group
http://www.phusewiki.org/wiki/index.php?title=Analysis_Results_Model

Outline

- Motivation
- AR&M sub-team
- RDF Data cube
- Status of deliverables
- Sample application

Disclaimer: Mistakes etc – my responsability / Marc

Acknowledgements

- Members of AR&M sub-team, discussion at every 2 week TC since 2014
- Tim Williams for discussions and ”Semantics 101 for Pharma” slides and authoring Technical Specification
- Marcelina Hungaria for discussions and authoring final version of the white paper

Motivation: Why Semantic Web / Linked Data?

CDISC SDTM/ADaM Pilot Project CDISCPILOT01

10. STUDY SUBJECTS

10.1. Disposition of Subjects

Table 14-1.01 provides a summary of the subjects in each analysis population. A total of 306 subjects have demographic information in the study tabulation database (Figure 10-1). Fifty-two (52) subjects were not randomized, and thus were screen failures. The remaining 254 subjects were randomized and entered the double-blind treatment phase. These subjects comprise the Intent-to-Treat population. The number of subjects randomized to each treatment arm was: 86 to placebo, 84 to the xanomeline low dose treatment group and 84 to the xanomeline high dose treatment group. Of the 254 subjects randomized to treatment, 118 completed the treatment phase (Week 24), and 110 completed the study through Week 26.

	Placebo (N=86)	Xanomeline Low Dose (N=84)	Xanomeline High Dose (N=84)	Total (N=254)
Intent-To-Treat (ITT)	86 (100%)	84 (100%)	84 (100%)	254 (100%)
Safety	86 (100%)	84 (100%)	84 (100%)	254 (100%)
Efficacy	79 (92%)	81 (96%)	74 (88%)	234 (92%)
Complete Week 24	60 (70%)	28 (33%)	30 (36%)	118 (46%)
Complete Study	58 (67%)	25 (30%)	27 (32%)	110 (43%)

http://www.cdisc.org/sdtmadam-pilot-project

Motivation: Why Semantic Web / Linked Data?

	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Total	
Intent-To-Treat (ITT)	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Safety	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Efficacy	79	(92%)	81	(96%)	74	(88%)	234	(92%)
Complete Week 24	60	(70%)	28	(33%)	30	(36%)	118	(46%)
Complete Study	58	(67%)	25	(30%)	27	(32%)	110	(43%)

10.1. Disposition of Subjects

Table 14-1.01 provides a summary of the subjects in each analysis population. A total of 306 subjects have demographic information in the study tabulation database (Figure 10-1). Fifty-two (52) subjects were not randomized, and thus were screen failures. The remaining 254 subjects were randomized and entered the double-blind treatment phase. These subjects comprise the Intent-to-Treat population. The number of subjects randomized to each treatment arm was: 86 to placebo, 84 to the xanomeline low dose treatment group and 84 to the xanomeline high dose treatment group. Of the 254 subjects randomized to treatment, 118 completed the treatment phase (Week 24), and 110 completed the study through Week 26.

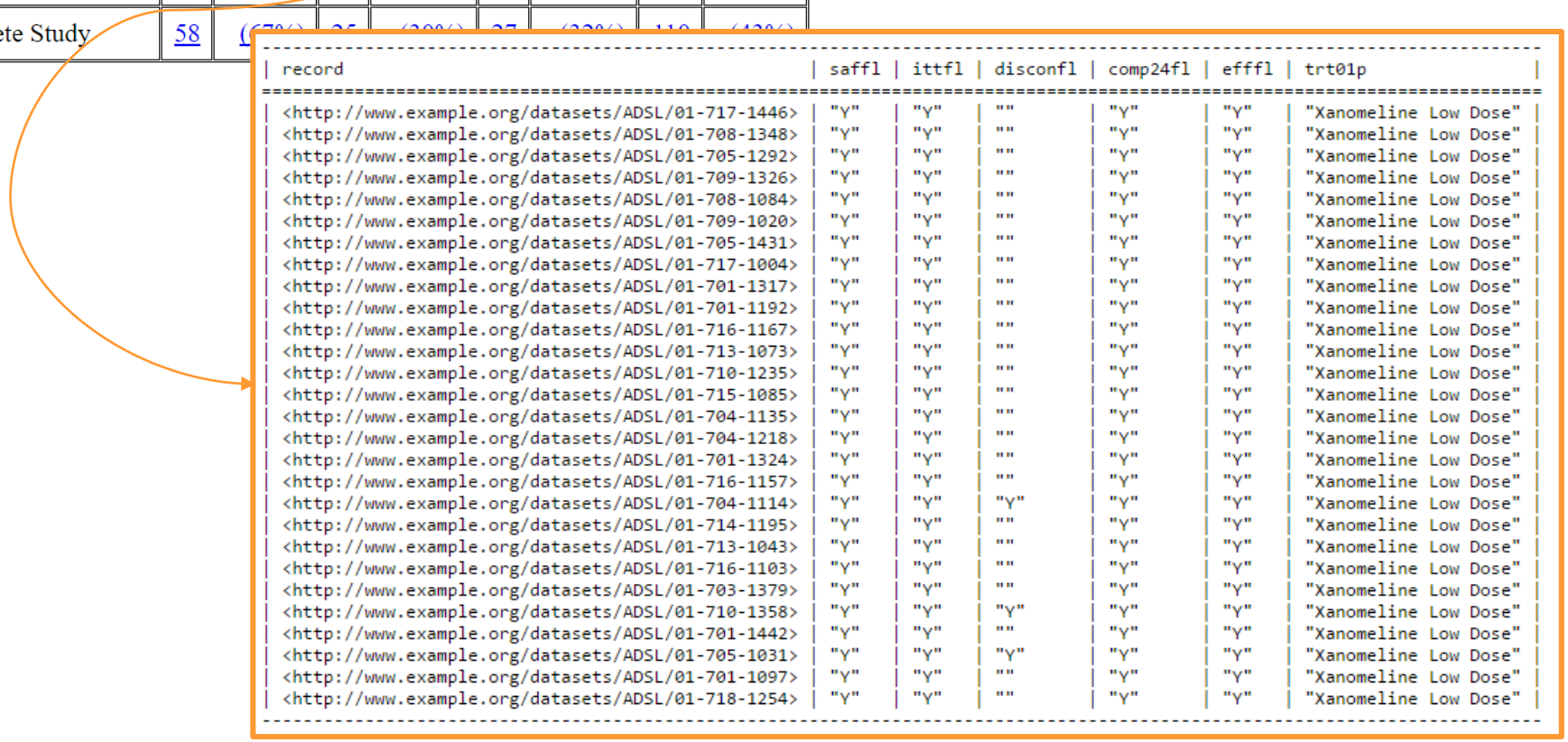
Motivation: Why Semantic Web / Linked Data?

	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Total	
Intent-To-Treat (ITT)	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Safety	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Efficacy	79	(92%)	81	(96%)	74	(88%)	234	(92%)
Complete Week 24	60	(70%)	28	(33%)	30	(36%)	118	(46%)
Complete Study	58	(67%)	25	(30%)	27	(32%)	110	(43%)

```
ds:obs41  a                              qb:Observation ;
          rdfs:comment                   "Statistic for number of records/S
          rdfs:label                     "41"^^xsd:string ;
          qb:dataSet                     ds:dataset-TAB1X01 ;
          crnd-attribute:denominator     "" ;
          crnd-attribute:unit            "NA"^^xsd:string ;
          crnd-dimension:comp24fl        code:comp24fl-Y ;
          crnd-dimension:disconfl        code:disconfl-_ALL_ ;
          crnd-dimension:efffl           code:efffl-_ALL_ ;
          crnd-dimension:factor          code:factor-quantity ;
          crnd-dimension:ittfl           code:ittfl-_ALL_ ;
          crnd-dimension:procedure       code:procedure-count ;
          crnd-dimension:saffl           code:saffl-_ALL_ ;
          crnd-dimension:trt01p          code:trt01p-Xanomeline_Low_Dose ;
          crnd-measure:measure           "28"^^xsd:double .
```

Motivation: Why Semantic Web / Linked Data?

	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Total	
Intent-To-Treat (ITT)	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Safety	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Efficacy	79	(92%)	81	(96%)	74	(88%)	234	(92%)
Complete Week 24	60	(70%)	28	(33%)	30	(36%)	118	(46%)
Complete Study	58	(						



record	saffl	ittfl	disconfl	comp24fl	efffl	trt01p
<http://www.example.org/datasets/ADSL/01-717-1446>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-708-1348>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-705-1292>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-709-1326>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-708-1084>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-709-1020>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-705-1431>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-717-1004>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-701-1317>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-701-1192>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-716-1167>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-713-1073>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-710-1235>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-715-1085>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-704-1135>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-704-1218>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-701-1324>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-716-1157>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-704-1114>	"Y"	"Y"	"Y"	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-714-1195>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-713-1043>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-716-1103>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-703-1379>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-710-1358>	"Y"	"Y"	"Y"	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-701-1442>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-705-1031>	"Y"	"Y"	"Y"	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-701-1097>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"
<http://www.example.org/datasets/ADSL/01-718-1254>	"Y"	"Y"	""	"Y"	"Y"	"Xanomeline Low Dose"

Sub-team and status of deliverables

The Analysis Results and Metadata (AR&M) Team is a sub-team of the PhUSE/FDA-CSC Semantic Technologies Working Group

http://www.phusewiki.org/wiki/index.php?title=Analysis_Results_Model

Tim Williams Fri, 01 Jul 2016:

> Today marks a significant milestone for the AR&M Project.
>
> The White Paper, Technical Specification, and R Package were delivered to PhUSE for review and comment by the PhUSE CSS Steering Committee.

Review ongoing

Deliverables

White Paper

- Describe approach and discusses benefit

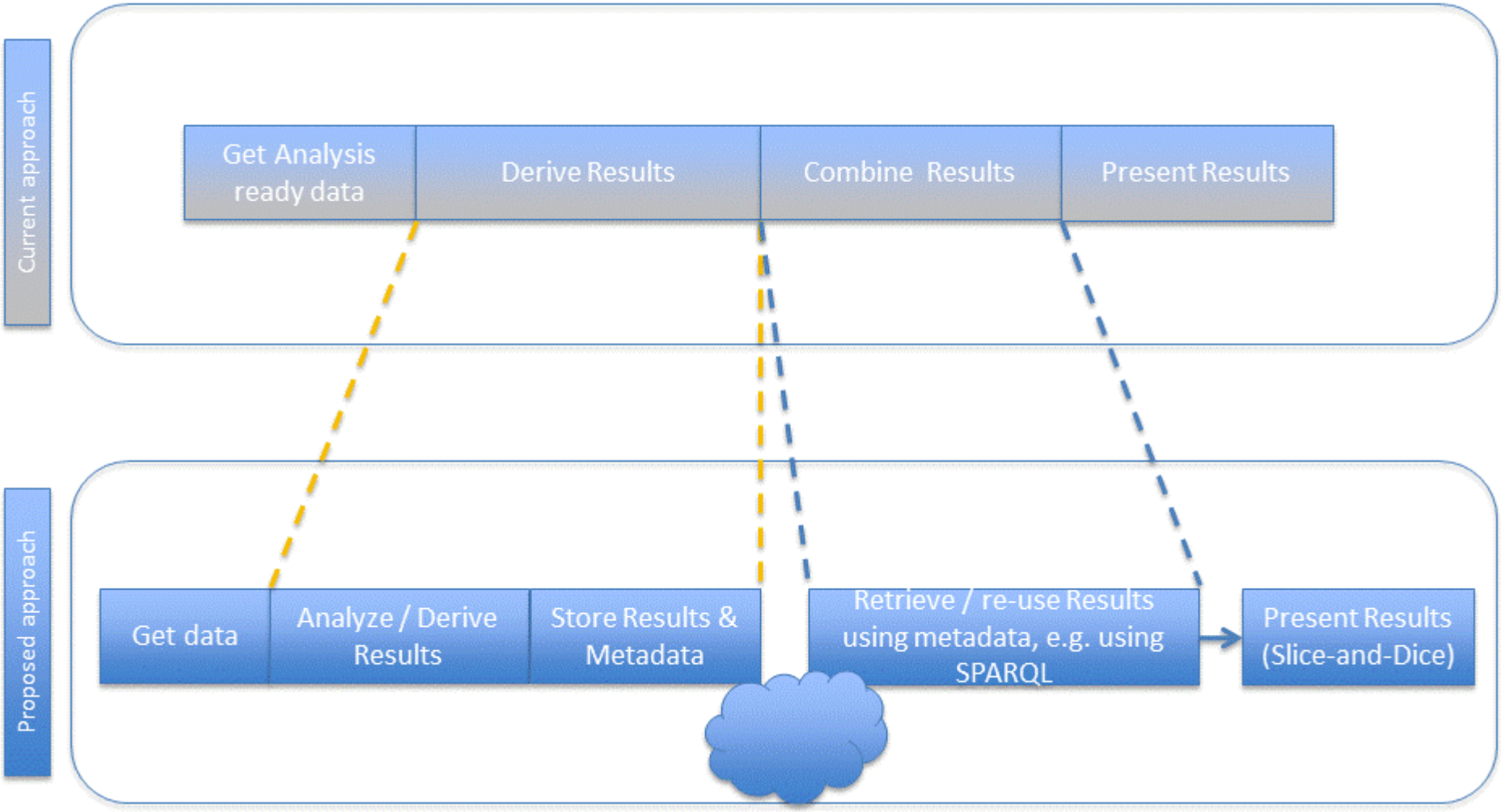
Technical Specification

- Specification of how to implement using RDF Data Cube vocabularey

R Package

- R code for creating RDF Data Cube from .csv files
  (https://github.com/phuse-org/rrdfqbcrnd)

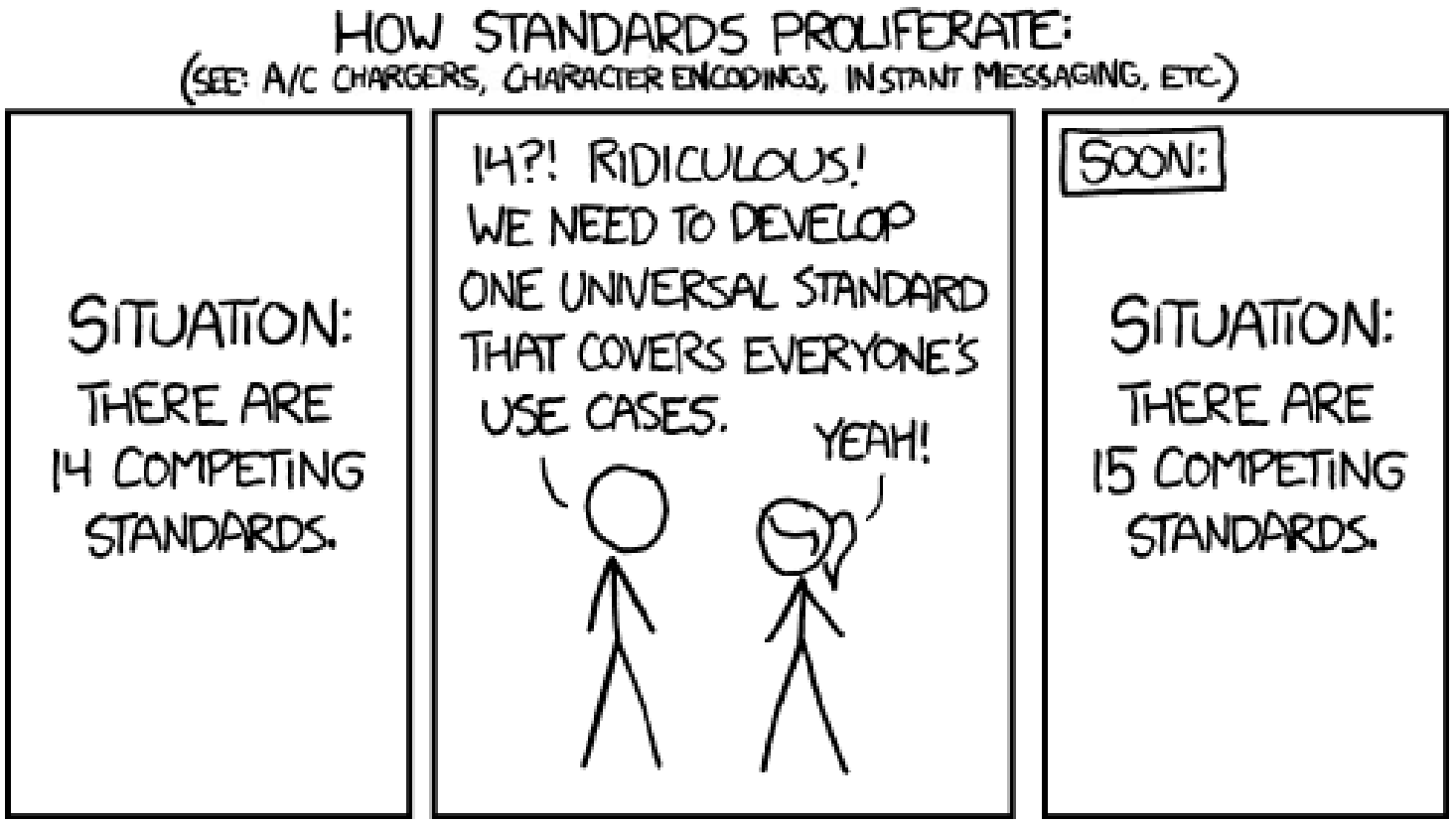
The idea: storing the results as RDF



Storing results using RDF Data Cube

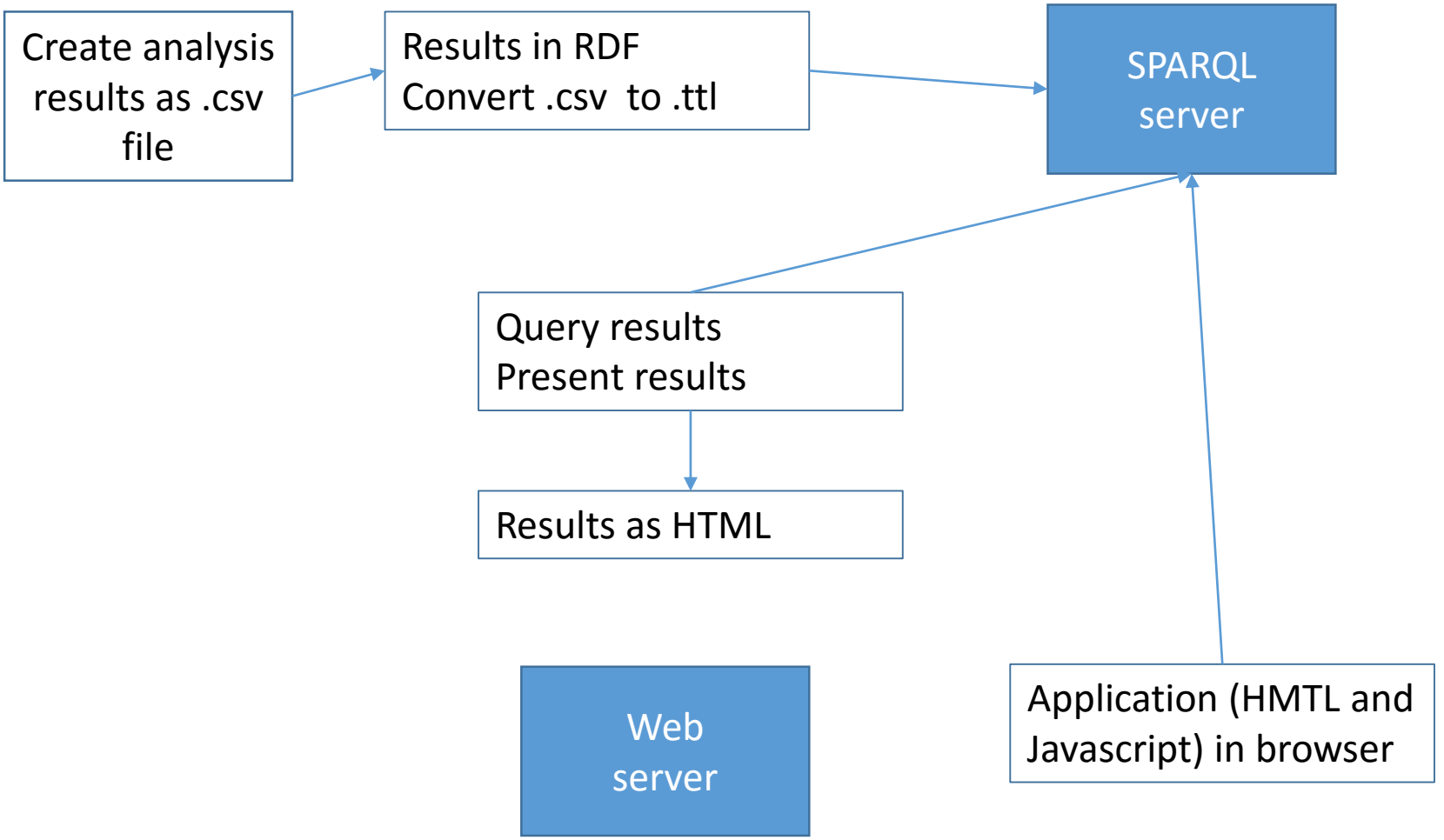
https://www.w3.org/TR/vocab-data-cube/

"Best Practice: Publish Statistical Data In Linked Data Format"
https://www.w3.org/2013/share-psi/bp/stats/ - 25 July 2016



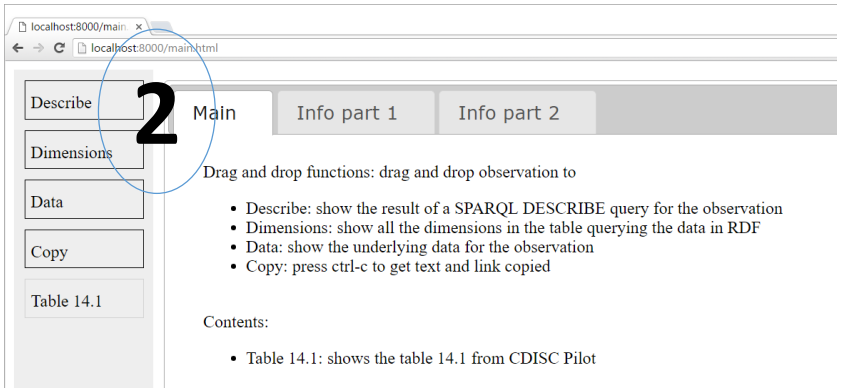
http://xkcd.com/927/

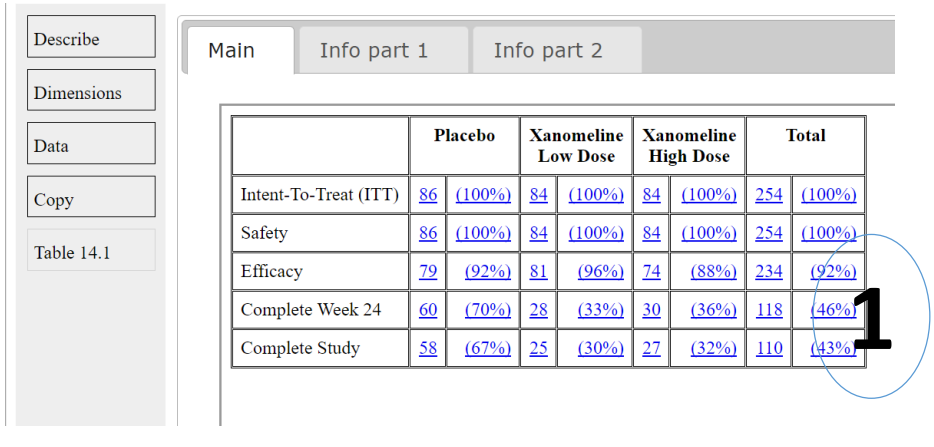
Application



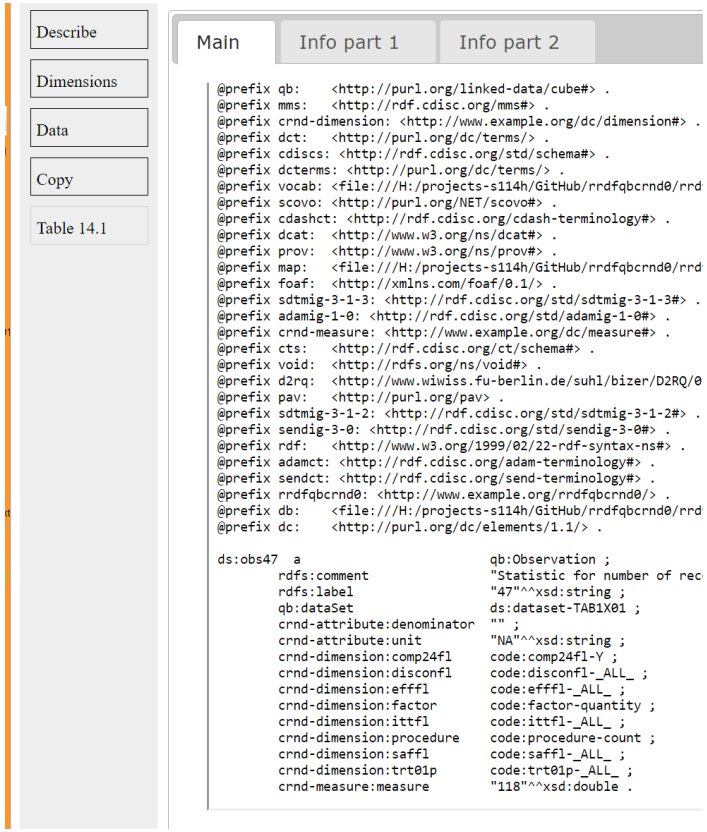
Application

1. Click and hold ”118”
2. Drag to describe
3. SPARQL describe for observation





	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Total	
Intent-To-Treat (ITT)	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Safety	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Efficacy	79	(92%)	81	(96%)	74	(88%)	234	(92%)
Complete Week 24	60	(70%)	28	(33%)	30	(36%)	118	(46%)
Complete Study	58	(67%)	25	(30%)	27	(32%)	110	(43%)



http://www.phusewiki.org/docs/Conference%202015%20TT%20Papers/TT07.pdf
http://www.phusewiki.org/docs/Conference%202015%20TT%20Presentations/TT07_Dude_Wheres_My_Graph.pptx

Application

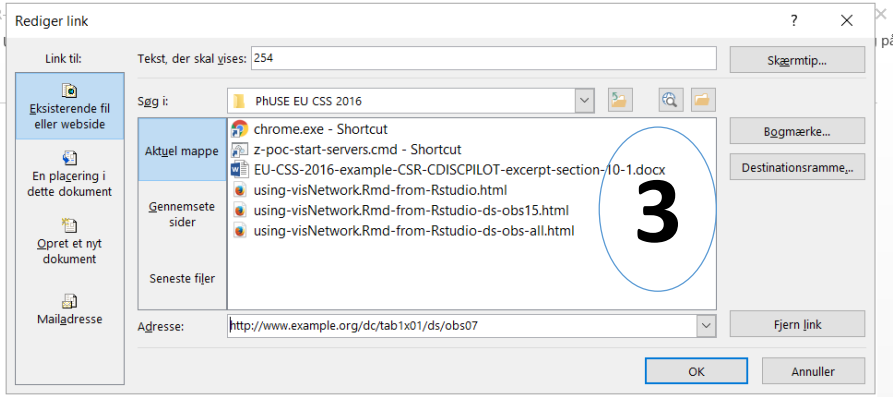
Describe
Dimensions
Data
Copy
Table 14.1

Main | Info part 1 | Info part 2

	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Total	
Intent-To-Treat (ITT)	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Safety	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Efficacy	79	(92%)	81	(96%)	74	(88%)	234	(92%)
Complete Week 24	60	(70%)	28	(33%)	30	(36%)	118	(46%)
Complete Study	58	(67%)	25	(30%)	27	(32%)	110	(43%)

1

1. Drag the ”254” to copy, press CTRL-c
2. paste into word,
3. inspect result



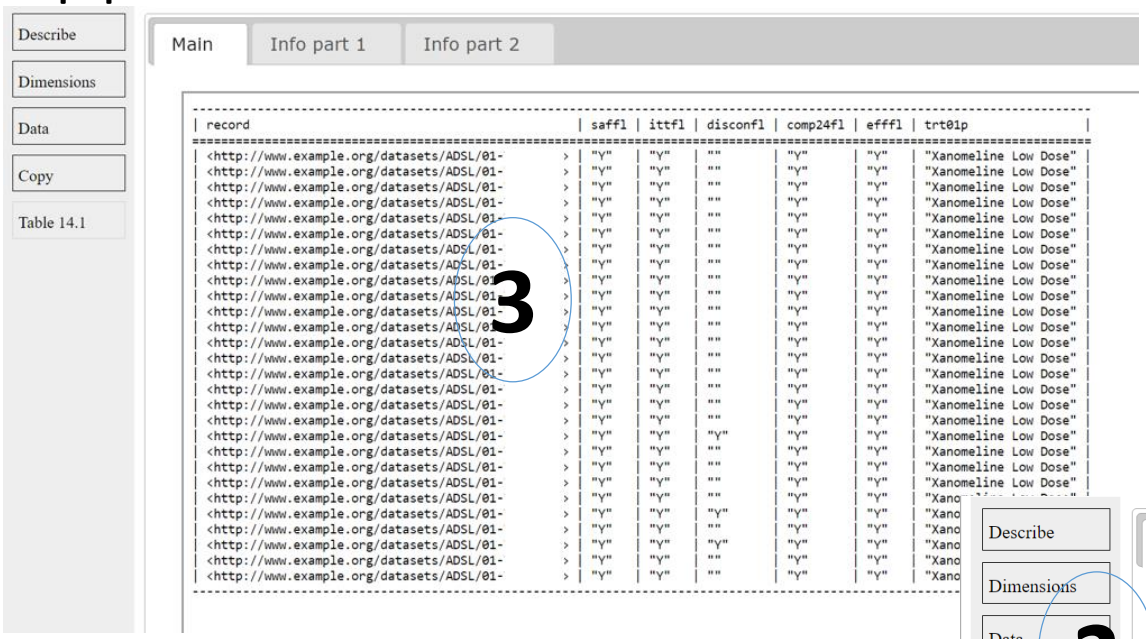
10.1. Disposition of Subjects

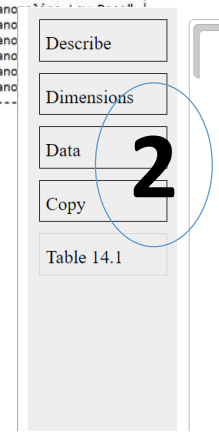
Table 14-1.01 provides a summary of the subjects in each analysis population. A total of 306 subjects have demographic information in the study tabulation database (Figure 10-1). Fifty-two (52) subjects were not randomized, and thus were screen failures. The remaining 254 subjects were randomized and entered the double-blind treatment phase. These subjects comprise the Intent-to-Treat population. The number of subjects randomized to each treatment arm was: 86 to placebo, 84 to the xanomeline low dose treatment group and 84 to the xanomeline high dose treatment group. Of the 254 subjects randomized to treatment, 118 completed the treatment phase (Week 24), and 110 completed the study through Week 26.

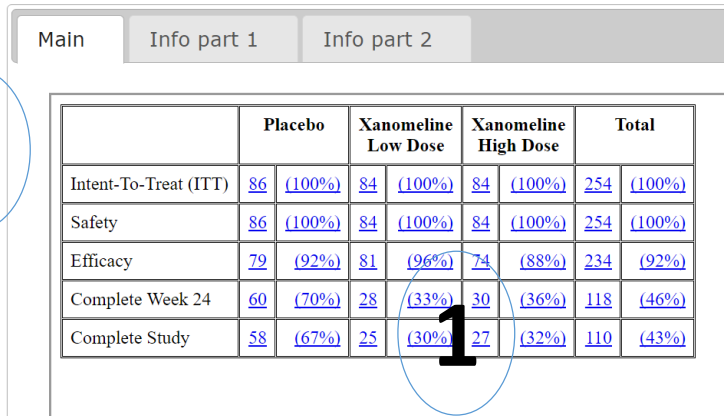
2

SIDE 1 AF 1 112 ORD ENGELSK (USA) 200 %

Application

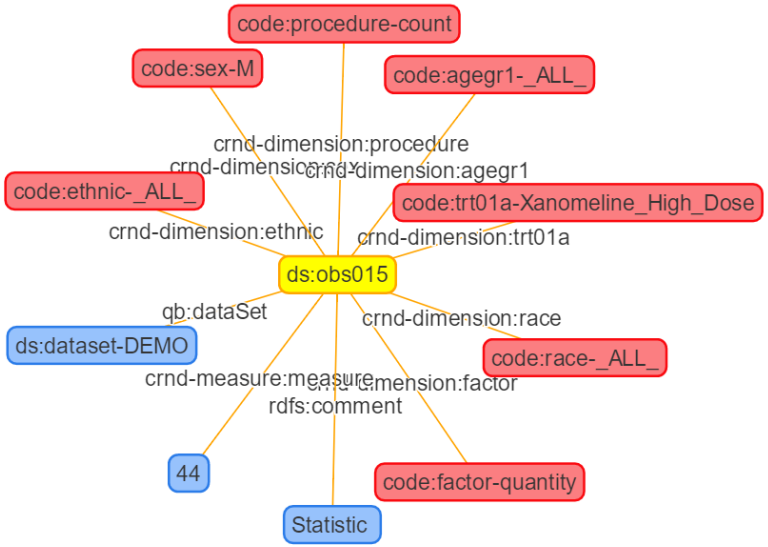


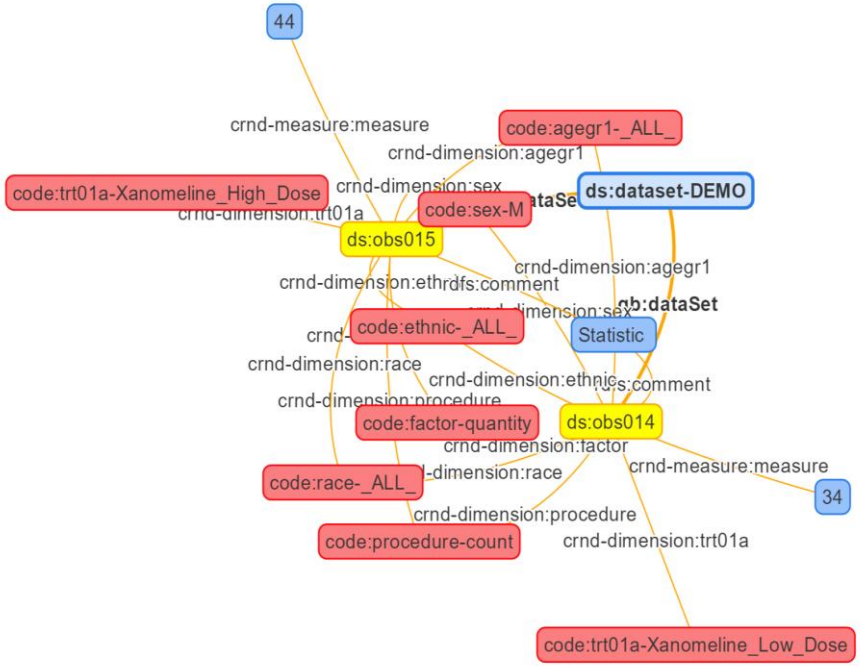




1. Click and hold ”28” in row ”Complete 24 weeks” and column ”Xanomeline Low Dose”
2. Drag to data
3. Inspect result

Visualising RDF data cube using visNetwork (R package) visjs

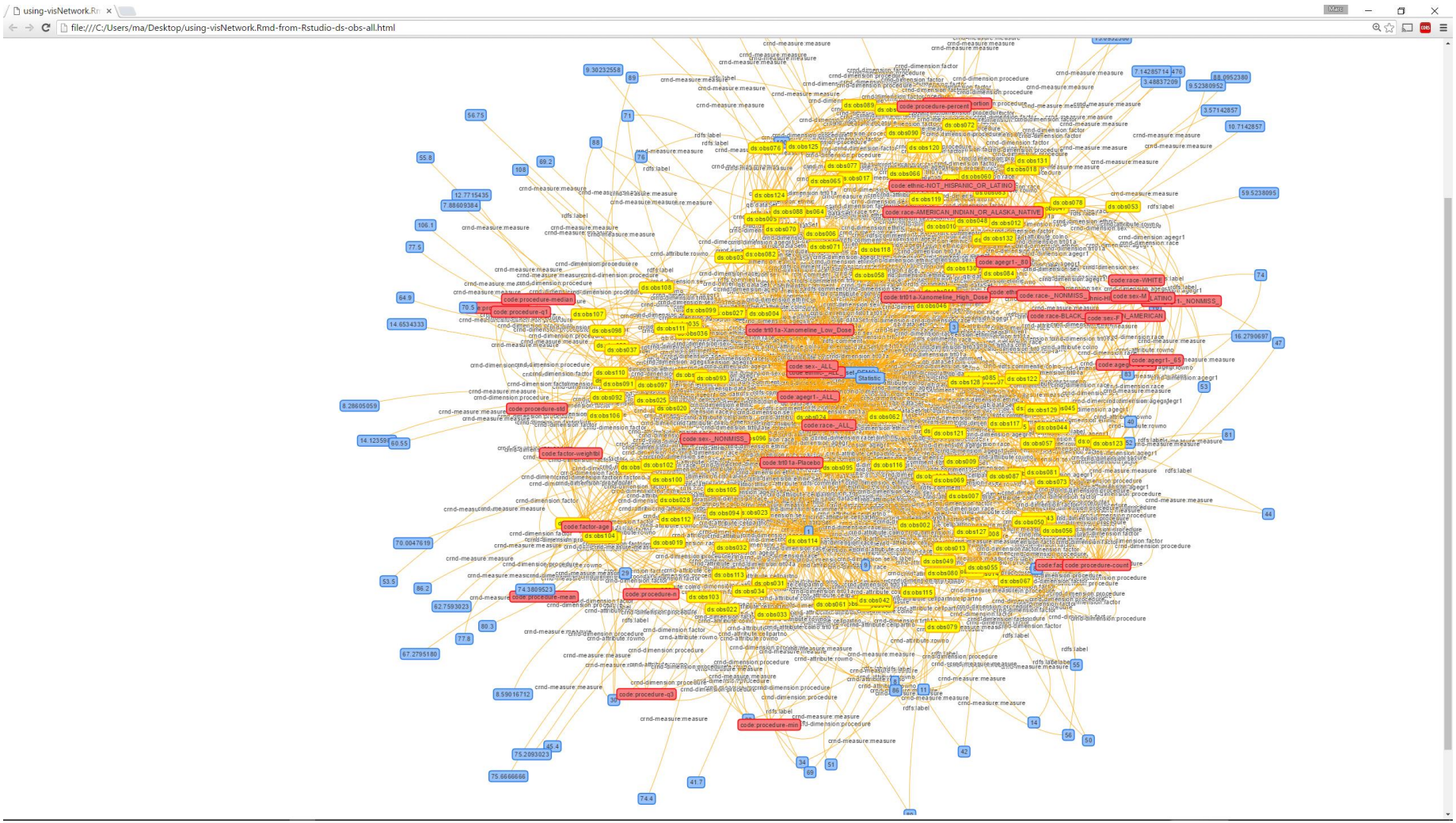




Úsing http://datastorm-open.github.io/visNetwork/, http://visjs.org/,

R-code available at

https://github.com/MarcJAndersen/rrdfqbcrnd0/blob/v.0.2.4/rrdfqbcrndex/inst/data-raw/using-visNetwork.Rmd



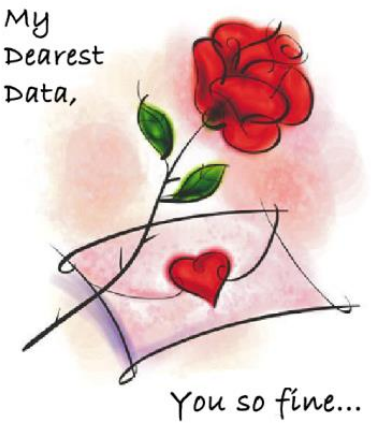
R code available at

https://github.com/MarcJAndersen/rrdfqbcrnd0/blob/v.0.2.4/rrdfqbcrndex/inst/data-raw/using-visNetwork.Rmd

Why Linked Data (from Semantics 101 for Pharma)

- Integrated metadata

  "Metadata is a love note to the people and machines after you." - *Jason Scott via Twitter, 2011*



- Data + Metadata = Context
- Traceability & Trust
- Across the data lifecycle
  - Reuse
  - Exploration
  - Flexible content presentation
- Reduce vendor Lock-in