

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

CDISCPIL01

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?

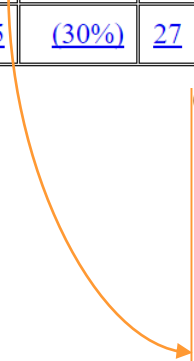
	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Total	
Intent-To-Treat (ITT)	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Safety	86	(100%)	84	(100%)	84	(100%)	254	(100%)
Efficacy	79	(92%)	81	(96%)	74	(88%)	234	(92%)
Complete Week 24	60	(70%)	28	(33%)	30	(36%)	118	(46%)
Complete Study	58	(67%)	25	(30%)	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
  rdfs:comment
  rdfs:label
  qb:dataSet
  crnd-attribute:denominator
  crnd-attribute:unit
  crnd-dimension:comp24f1
  crnd-dimension:disconf1
  crnd-dimension:efff1
  crnd-dimension:factor
  crnd-dimension:ittf1
  crnd-dimension:procedure
  crnd-dimension:saff1
  crnd-dimension:trt01p
  crnd-measure:measure
```

```
qb:Observation ;
  "Statistic for number of records/S
  "41"^^xsd:string ;
  ds:dataset-TAB1X01 ;
  "" ;
  "NA"^^xsd:string ;
  code:comp24f1-Y ;
  code:disconf1-_ALL_ ;
  code:efff1-_ALL_ ;
  code:factor-quantity ;
  code:ittf1-_ALL_ ;
  code:procedure-count ;
  code:saff1-_ALL_ ;
  code:trt01p-Xanomeline_Low_Dose ;
  "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	(67%)	27	(33%)	27	(33%)	110	(43%)

record	saffl	ittfl	disconf1	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](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

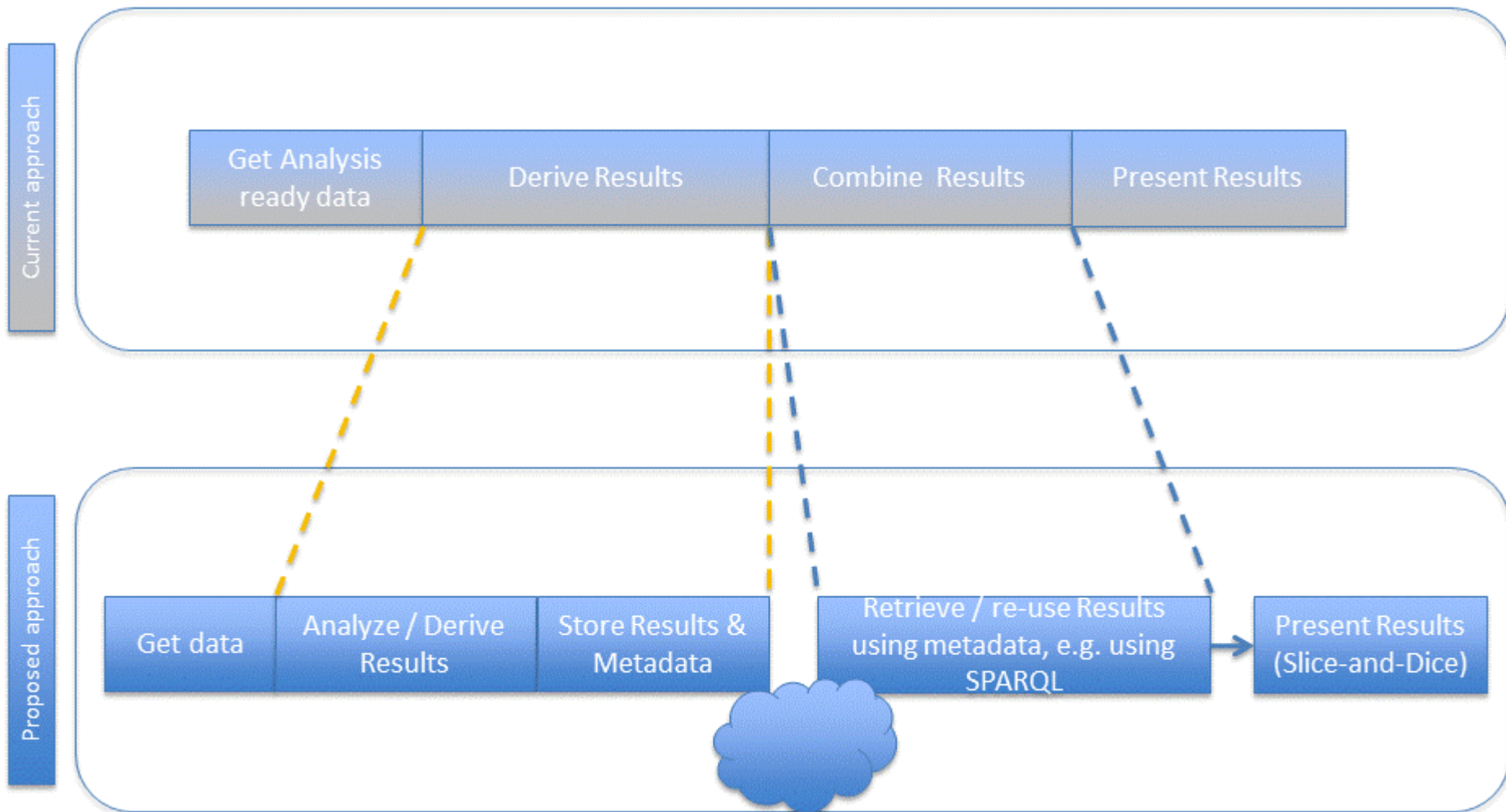
Technical Specification

- Specification of how to implement using RDF Data Cube vocabulary

R Package

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

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"

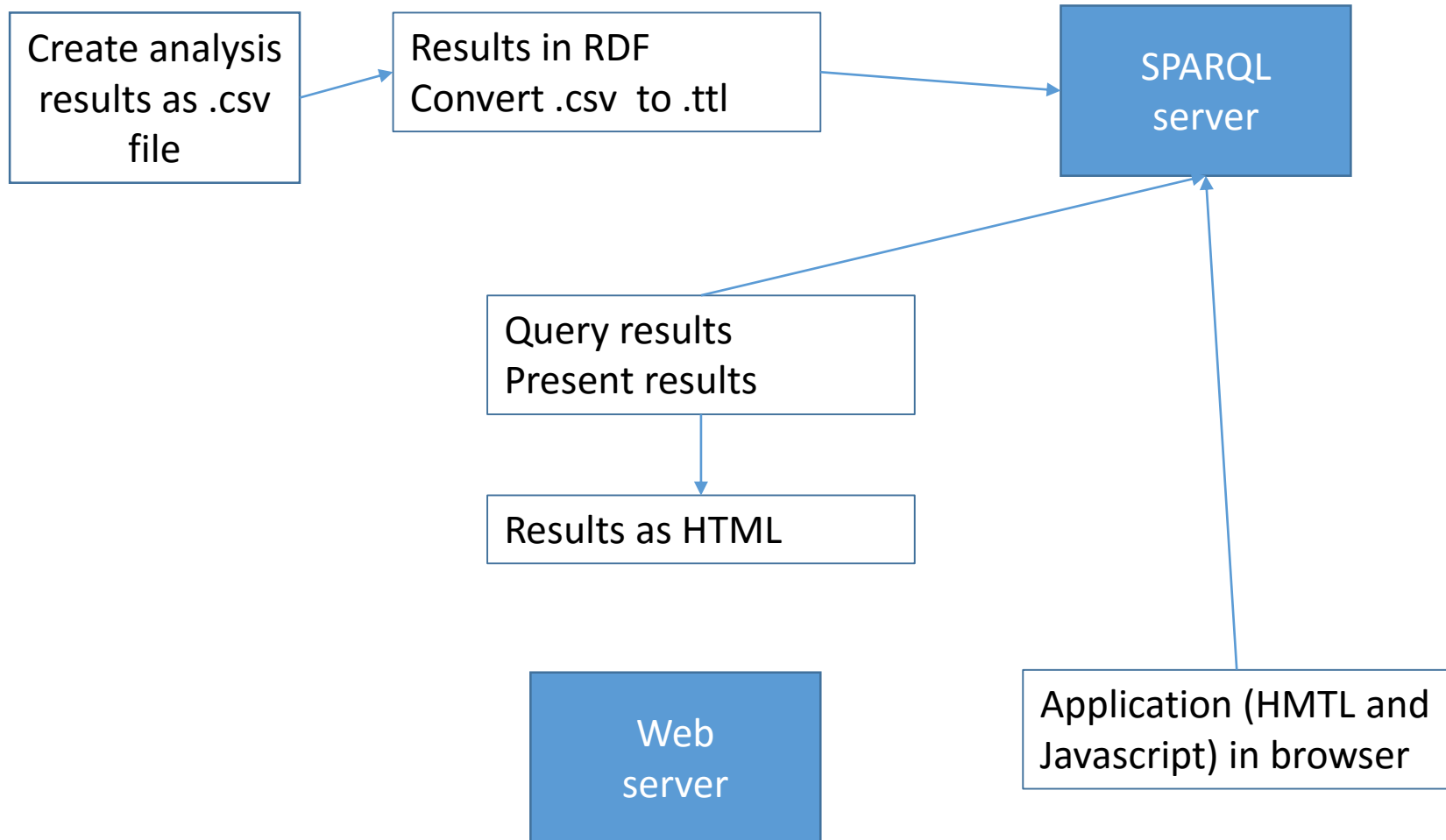
<https://www.w3.org/2013/share-psi/bp/stats/> - 25 July 2016

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)



<http://xkcd.com/927/>

Application



<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

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

Describe	Main	Info part 1	Info part 2
Dimensions			
Data			
Copy			
Table 14.1	<pre> @prefix qb: <http://purl.org/linked-data/cube#> . @prefix mms: <http://rdf.cdisc.org/mms#> . @prefix crnd-dimension: <http://www.example.org/dc/dimension#> . @prefix dct: <http://purl.org/dc/terms/> . @prefix cdiscs: <http://rdf.cdisc.org/std/schema#> . @prefix dcterms: <http://purl.org/dc/terms/> . @prefix vocab: <file:///H:/projects-s114h/GitHub/rndfqbcrrnd0/rnd @prefix scovo: <http://purl.org/NET/scovo#> . @prefix cdashct: <http://rdf.cdisc.org/cdash-terminology#> . @prefix dcat: <http://www.w3.org/ns/dcat#> . @prefix prov: <http://www.w3.org/ns/prov#> . @prefix map: <file:///H:/projects-s114h/GitHub/rndfqbcrrnd0/rnd @prefix foaf: <http://xmlns.com/foaf/0.1/> . @prefix sdtmig-3-1-3: <http://rdf.cdisc.org/std/sdtmig-3-1-3#> . @prefix adamig-1-0: <http://rdf.cdisc.org/std/adamig-1-0#> . @prefix crnd-measure: <http://www.example.org/dc/measure#> . @prefix cts: <http://rdf.cdisc.org/ct/schema#> . @prefix void: <http://rdfs.org/ns/void#> . @prefix d2rq: <http://www.wiwiiss.fu-berlin.de/suhl/bizer/D2RQ/0 @prefix pav: <http://purl.org/pav/> . @prefix sdtmig-3-1-2: <http://rdf.cdisc.org/std/sdtmig-3-1-2#> . @prefix sendig-3-0: <http://rdf.cdisc.org/std/sendig-3-0#> . @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> . @prefix adamct: <http://rdf.cdisc.org/adam-terminology#> . @prefix sendct: <http://rdf.cdisc.org/send-terminology#> . @prefix rndfqbcrrnd0: <http://www.example.org/rndfqbcrrnd0/> . @prefix db: <file:///H:/projects-s114h/GitHub/rndfqbcrrnd0/rnd @prefix dc: <http://purl.org/dc/elements/1.1/> . ds:obs47 a rdfs:comment qb:Observation ; rdfs:label "Statistic for number of rec qb:dataSet "47""^^xsd:string ; crnd-attribute:denominator ds:dataset-TAB1X01 ; crnd-attribute:unit "" ; crnd-dimension:comp24fl "NA""^^xsd:string ; crnd-dimension:disconf1 code:comp24fl-Y ; crnd-dimension:efffl code:disconf1-ALL_ ; crnd-dimension:factor code:efffl-ALL_ ; crnd-dimension:ittfl code:factor-quantity ; crnd-dimension:procedure code:ittfl-ALL_ ; crnd-dimension:saffl code:procedure-count ; crnd-dimension:trt01p code:saffl-ALL_ ; crnd-measure:measure code:trt01p-ALL_ ; "118""^^xsd:double . </pre>		

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

Rediger link
?
×

Link til:

Eksisterende fil eller webside

En placering i dette dokument

Åbnet et nyt dokument

Mailadresse

Tekst, der skal yses: 254

Søg i: PhUSE EU CSS 2016

Aktuel mappe:

- chrome.exe - Shortcut
- z-poc-start-servers.cmd - Shortcut
- EU-CSS-2016-example-CSR-CDISCILOT-excerpt-section-10-1.docx
- using-visNetwork.Rmd-from-Rstudio.html
- using-visNetwork.Rmd-from-Rstudio-ds-obs15.html
- using-visNetwork.Rmd-from-Rstudio-ds-obs-all.html

Seneste filer

Adresse: http://www.example.org/dc/tab1x01/ds/obs07

OK
Annuller

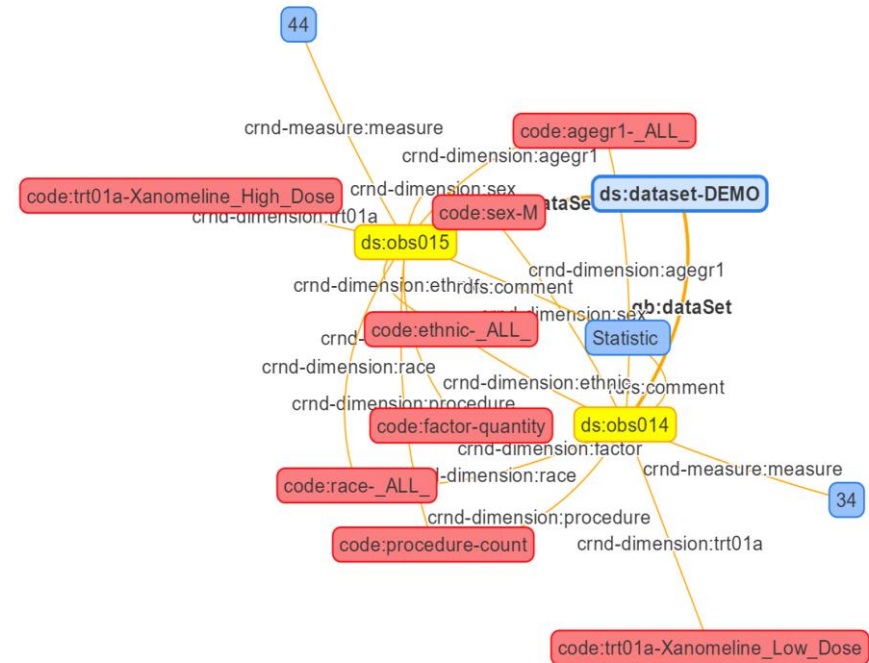
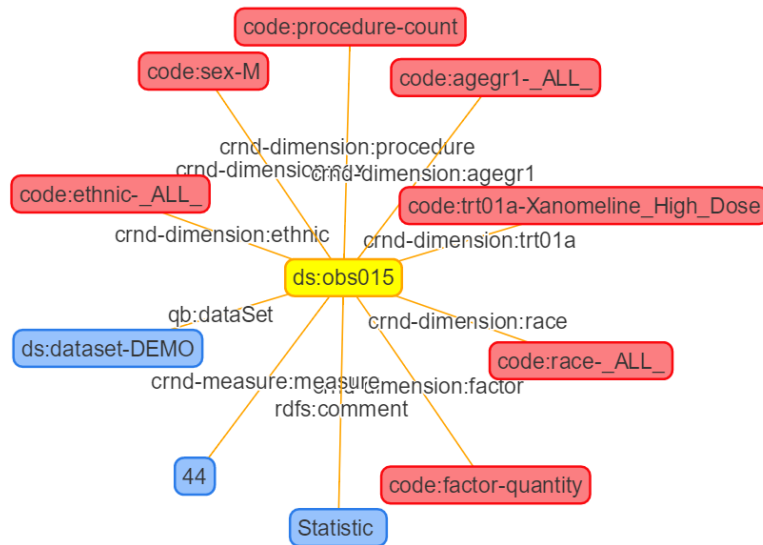
10.1. Disposition of Subjects

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.

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

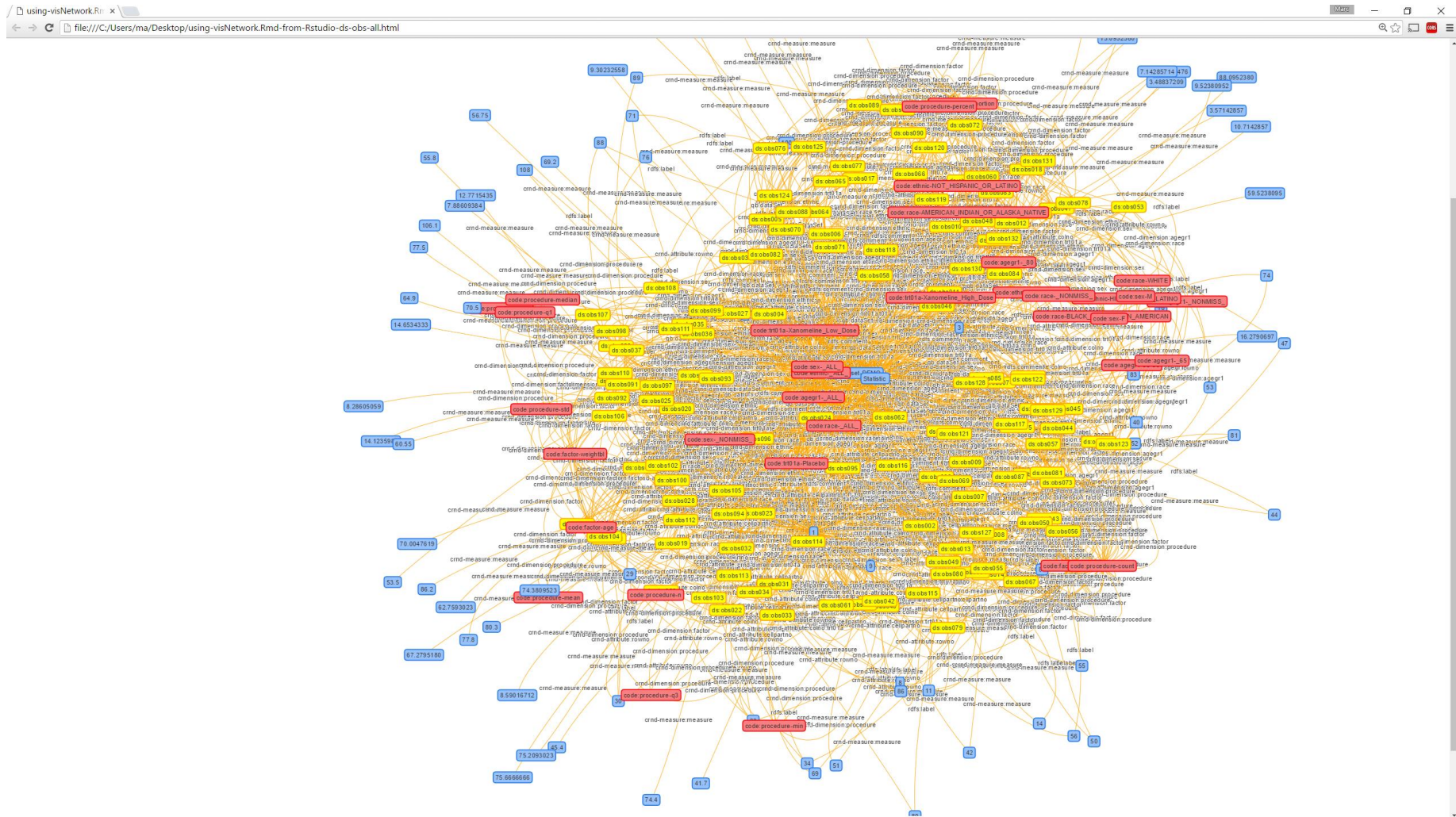
Visualising RDF data cube using visNetwork (R package) visjs



Using <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/rrdfqbcrrnd0/blob/v.0.2.4/rrdfqbcrrnd0/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

