PhUSE SDE Meeting, NJ 2015

Delivering Statistical Results as an RDF Data Cube: a Proof Of Concept

Presented by Marcelina Hungria / DIcore Group, LLC

Acknowledgement

AR&M Team: sub-team of the PhUSE CS Semantic Technology Working Group

Outline

- PhUSE CS Analysis Results & Metadata (AR&M)
  - Project background and definition
    - Foundational Standards
  - Benefits
- Proof of Concept (POC)
  - Scope
  - Technology Tools
  - Building and querying an RDF Cube
- Roadmap
- Questions

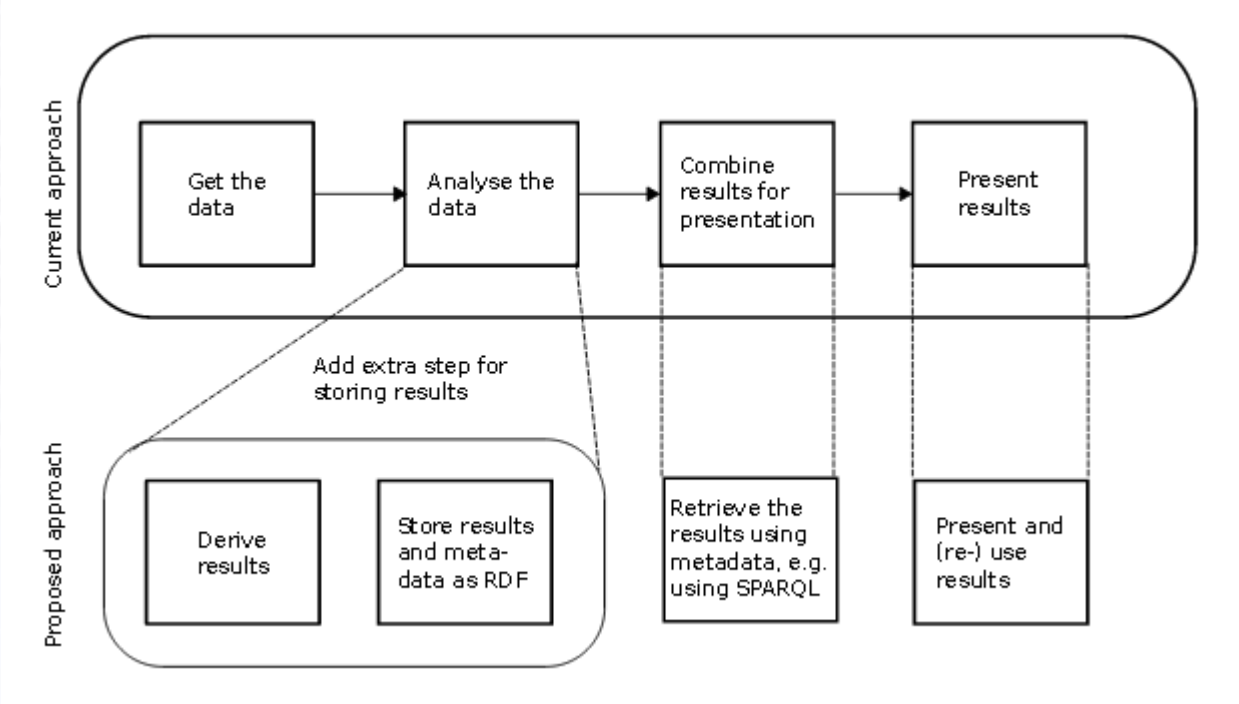
A&RM Project

- One of the current projects in the PhUSE CS Semantic Technology WG
  - Overarching goal: Investigate how formal semantic standards can support the clinical and non-clinical trial data life cycle from protocol to submission

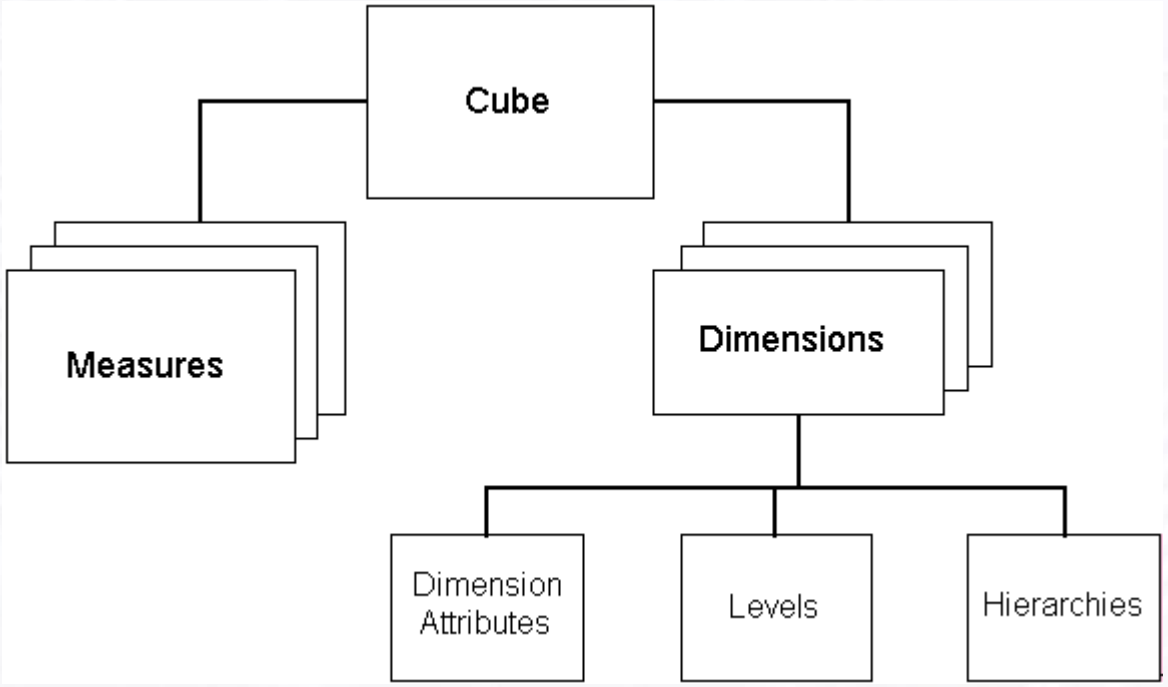
- Project goal: Logical model for the representation of analysis results and their associated metadata for clinical and non-clinical applications
  - Leverage *linked data* technology where data and metadata coexist → RDF Data Cube Vocabulary

A&RM Project

– Generation of Statistical Results Process



AR&M Project – Data Cube*



Source: Wikipedia - *Diagram of the Logical Multidimensional Model*

- Data cube = multidimensional dataset (>3 dimensions known as hypercube)

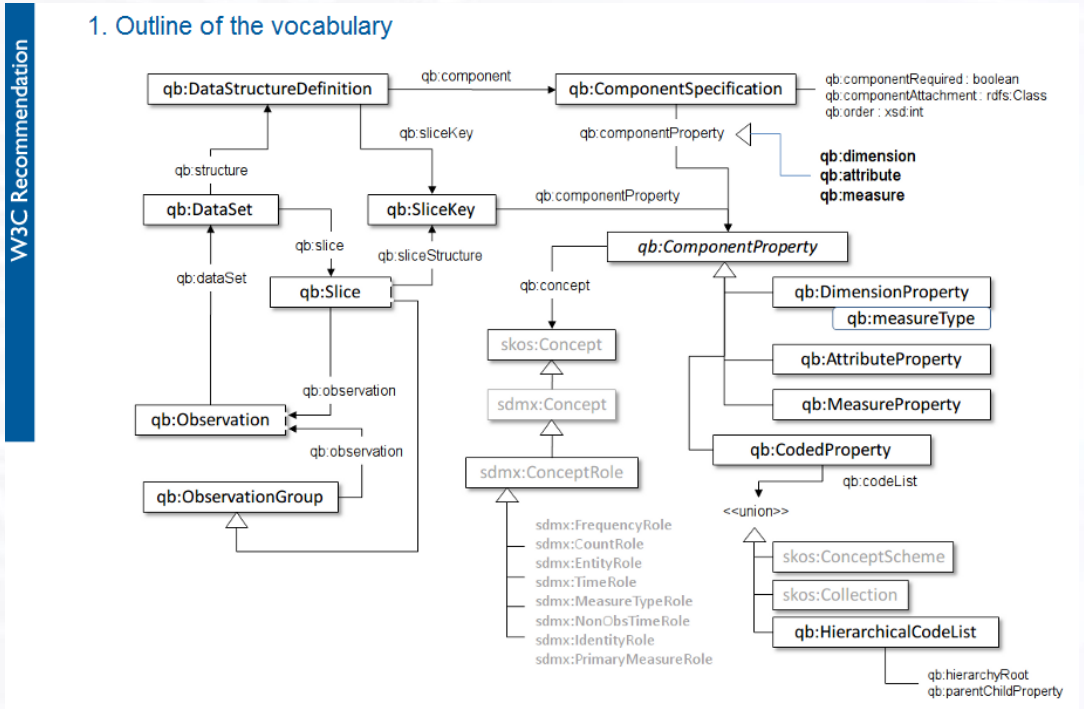
AR&M Project – DC Example

- Basic Demographics Characteristics
  - Continuous variable *Age* analyzed under five Dimensions, for levels indicated
- Challenge: What to store in the Data Cube?

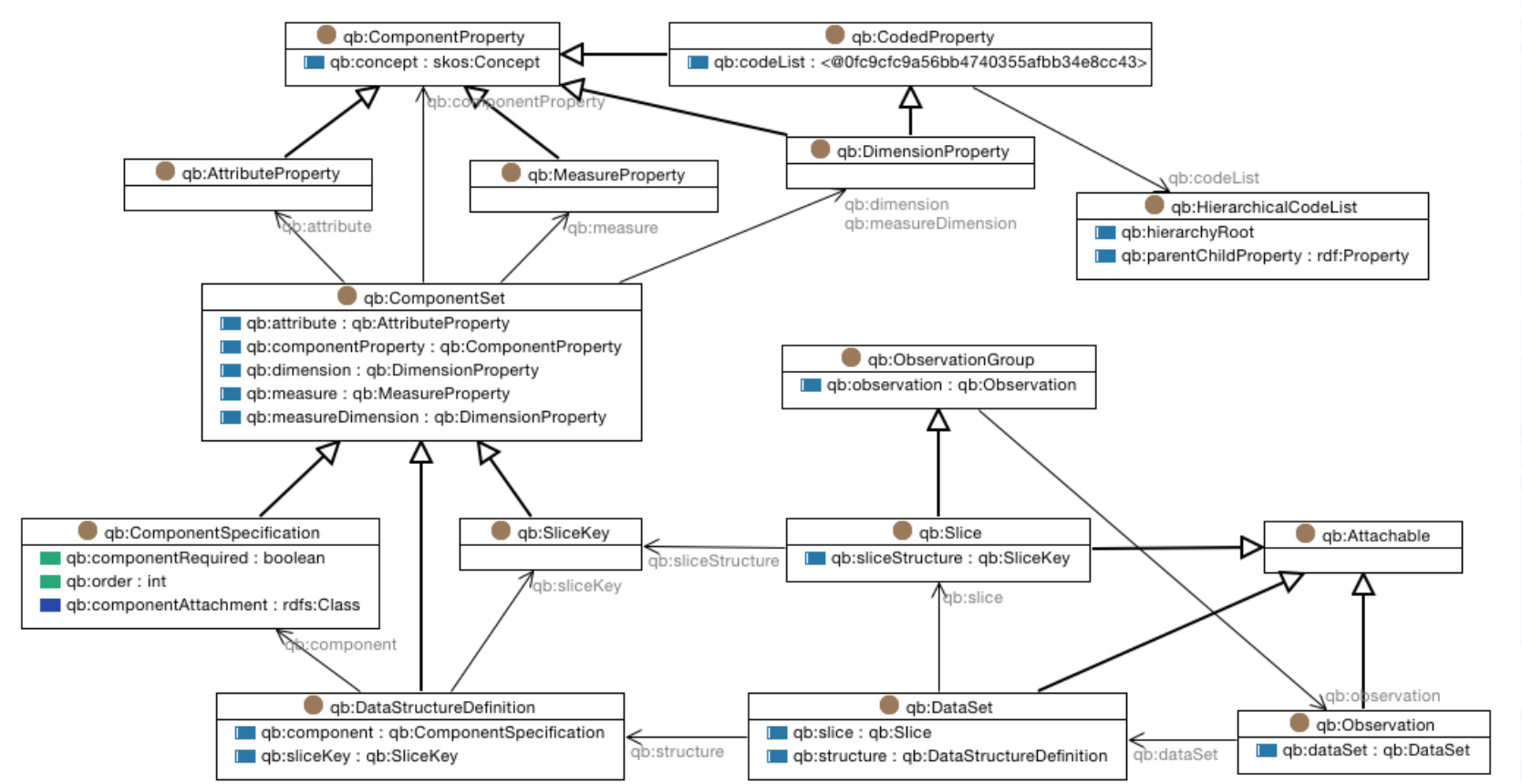
	A	B	C	D	E	F	G	H	I
1	Dimension Level					Statistic (For Variable Age, in Unit=*Years*)	Result of Statistic (hypothetical example)	Includes	Potential # cells in a Data Cube (For Variable Age, in Unit=Years)
2	PopulationFlag (Dimension1)	Treatment (Dimension2)	AgeGroup (Dimension3)	Sex (Dimension4)	Race (Dimension5)				
3						Count	600	Subjects in All PopulationFlag, Treatment, AgeGroup , Sex, and Race groups	(# of Statistics needed) *(Dimension1.Levels) *(Dimension2.Levels) *(Dimension3.Levels) *(Dimension4.Levels) *(Dimension5.Levels)
4						Mean	54.0		
5						SD	6.1		
6						Min	34		
7						Max	70		
8						...			
9			<65			Count	585	Subjects in All PopulationFlag, Treatment, Sex and Race groups for AgeGroup= "<65"	
10						...			
11				F		...	325	Subjects in All PopulationFlag, Treatment, Age and Race groups for Sex=*F*	
12					WHITE	...	167	Subjects in All PopulationFlag, Treatment, Age and Sex groups for Race=*WHITE*	
13			...					Subjects in remaining AgeGroup groups under All PopulationFlag, Age, Sex and Race groups	
14		Active1				Count	285	Subjects in All population flag, age, sex and race groups for Treatment=*Active1*	
15						...			
16			...						
17		...							
18	Safety					Count	582	Subjects under All Treatment, AgeGroup, Sex, and Race groups for PopulationFlag="Safety"	
19						...			
20		...							
21	...							Subjects in remaining PopulationFlag groups under All Treatment, Age, Sex, and Race groups	

A&RM Project – Base Vocabulary

- Evaluation of the RDF Data Cube Vocabulary for storing and documenting statistical results
  - As all W3C recommended standards, RDF Data Cube Vocabulary is an open, non-proprietary standard



A&RM Project – CRND Data Cube Vocabulary



Source: CRND RDF Data Cube Structure - Draft Technical Guidance

Foundational Standards

- *Linked Data* Standards – W3C Semantic Standards
  - RDF (Resource Description Framework)
  - SPARQL[*1] Query Language for RDF
- CDISC Standards, in RDF representation

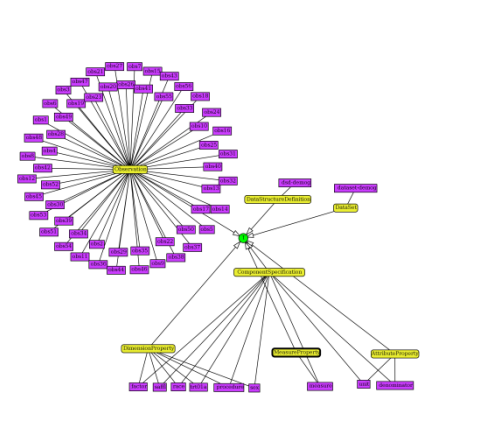
[*1] Pronounced “sparkle”

Linked Data: RDF

- RDF: directed graph
  - set of triples: *Subject Predicate Object*
    - Each component, a Uniform Resource Identifier (URI)

      e.g., http://www.example.org/dc/dm/ds/obs15

1	Subject	Predicate	Object
367	<http://www.example.org/dc/dm/ds/obs15>	<http://www.example.org/dc/dm/prop/measure>	33^^<http://www.w3.org/2001/XMLSchema#int>
368	<http://www.example.org/dc/dm/ds/obs15>	<http://www.example.org/dc/dm/prop/saffl>	<http://www.example.org/dc/code/saffl-Y>



```
ds:obs15  a                qb:Observation ;
          rdfs:label       "15" ;
          qb:dataSet       ds:dataset-DM ;
          prop:denominator "_ALL_" ;
          prop:factor      code:factor-quantity ;
          prop:measure     "33"^^xsd:int ;
          prop:procedure   code:procedure-count ;
          prop:race        code:race-_ALL_ ;
          prop:saffl       code:saffl-Y ;
          prop:sex         code:sex-M ;
          prop:trt01a      code:trt01a-Placebo ;
          prop:unit        "_NULL_"^^xsd:string .
```

A&RM Data Cube - Benefits

- Potentially provides better Traceability mechanisms
- Facilitates Analysis Results Validation
- Enables Reusability of Results → Data Integration
- Facilitates exploratory analysis
- Potentially facilitates the Regulatory Review Process
  - By minimizing or complementing the use of PDF files as a delivery mechanism for statistical results

Define-XML Analysis Results Metadata vs. AR&M DC

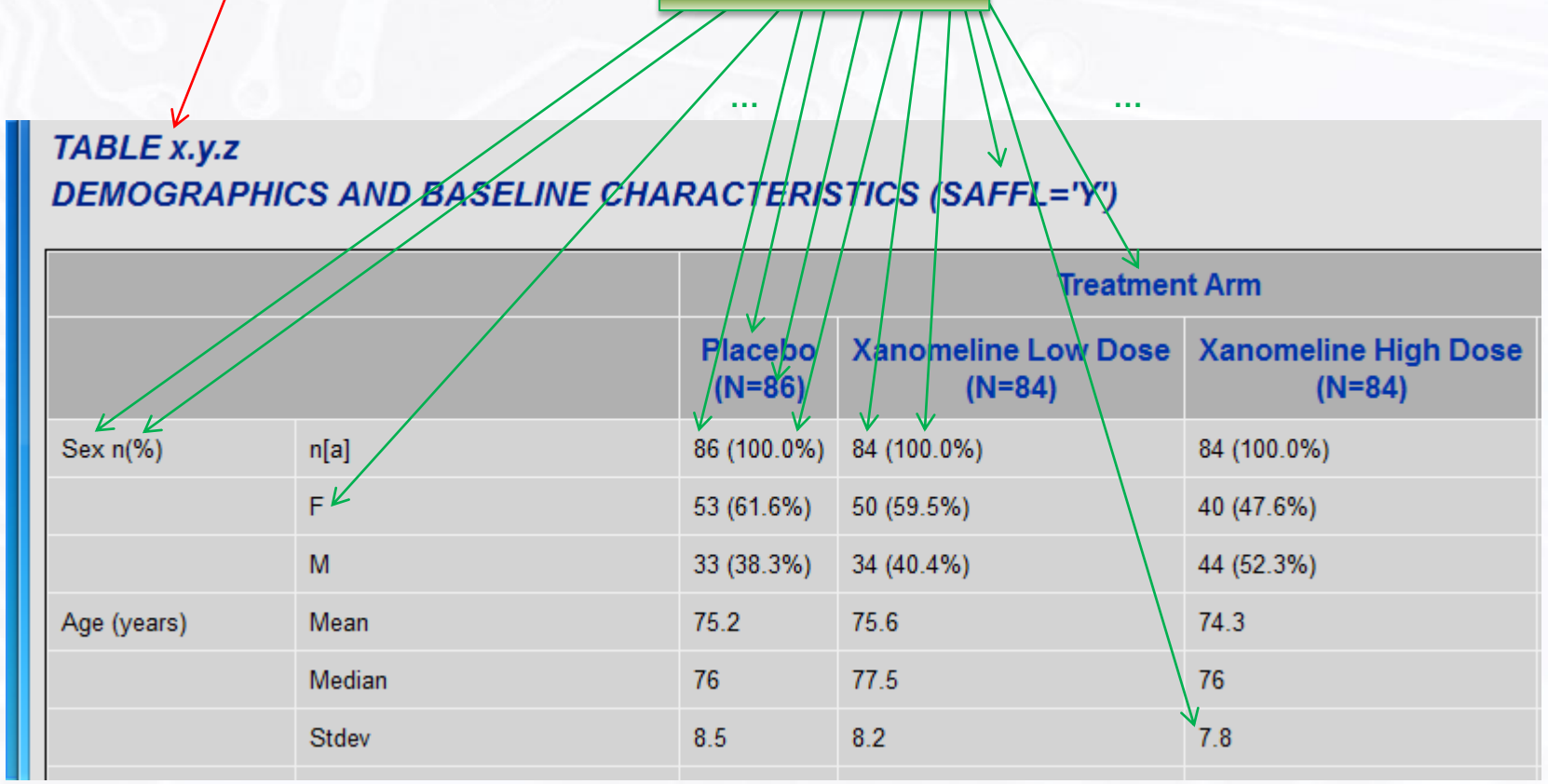


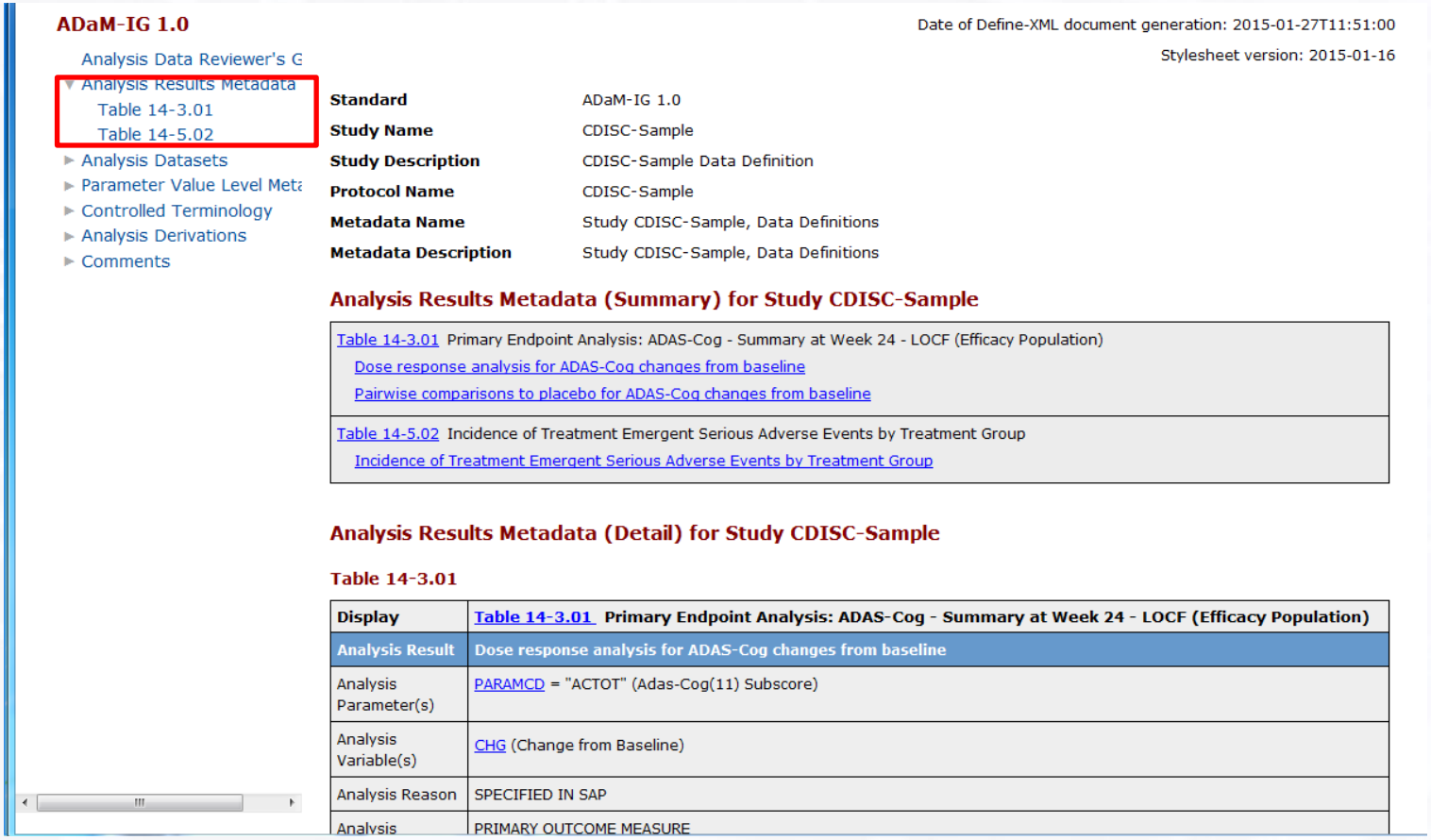
TABLE x.y.z
DEMOGRAPHICS AND BASELINE CHARACTERISTICS (SAFFL='Y')

		Treatment Arm		
		Placebo (N=86)	Xanomeline Low Dose (N=84)	Xanomeline High Dose (N=84)
Sex n(%)	n[a]	86 (100.0%)	84 (100.0%)	84 (100.0%)
	F	53 (61.6%)	50 (59.5%)	40 (47.6%)
	M	33 (38.3%)	34 (40.4%)	44 (52.3%)
Age (years)	Mean	75.2	75.6	74.3
	Median	76	77.5	76
	Stdev	8.5	8.2	7.8

Viewer Software Tools could interactively extract and display associated metadata (upon hover over, drill down options)

Define-XML 2.0 Analysis Results Metadata

(displayed w/ default stylesheet)



ADaM-IG 1.0

Date of Define-XML document generation: 2015-01-27T11:51:00

Stylesheet version: 2015-01-16

- Analysis Data Reviewer's G
- Analysis Results Metadata
  - Table 14-3.01
  - Table 14-5.02
- Analysis Datasets
- Parameter Value Level Meta
- Controlled Terminology
- Analysis Derivations
- Comments

Standard	ADaM-IG 1.0
Study Name	CDISC-Sample
Study Description	CDISC-Sample Data Definition
Protocol Name	CDISC-Sample
Metadata Name	Study CDISC-Sample, Data Definitions
Metadata Description	Study CDISC-Sample, Data Definitions

Analysis Results Metadata (Summary) for Study CDISC-Sample

Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population) Dose response analysis for ADAS-Cog changes from baseline Pairwise comparisons to placebo for ADAS-Cog changes from baseline
Table 14-5.02 Incidence of Treatment Emergent Serious Adverse Events by Treatment Group Incidence of Treatment Emergent Serious Adverse Events by Treatment Group

Analysis Results Metadata (Detail) for Study CDISC-Sample

Table 14-3.01

Display	**Table 14-3.01 Primary Endpoint Analysis: ADAS-Cog - Summary at Week 24 - LOCF (Efficacy Population)**
Analysis Result	**Dose response analysis for ADAS-Cog changes from baseline**
Analysis Parameter(s)	PARAMCD = "ACTOT" (Adas-Cog(11) Subscore)
Analysis Variable(s)	CHG (Change from Baseline)
Analysis Reason	SPECIFIED IN SAP
Analysis	PRIMARY OUTCOME MEASURE

Source: http://www.cdisc.org/adam

AR&M Project – POC / Software Tools

- Open Source
- Prior Technology Tools Evaluated
  - Publisci (Ruby gem), Open Refine (manual)
- Selected Technology Tools
  - R (programmatic)
    - rrdf package (RDF functionality)
    - SPARQL interface
  - Fuseki server
    - SPARQL endpoint

A&RM Project – POC Scope

- Source Data
  - ADaM datasets used in PhUSE – CSS Scriptathon 2014) *1
- Statistics (Basic Descriptive Statistics)
  - Model one data cube observation = one result
- Metadata (Codelists, from CDISC CT in RDF)
  - For Race and Sex in the examples

*1 Results could also be derived from SDTM data; however, not under scope

AR&M Project – POC Process

- The RDF Data Cube delivery process
  - Build it / Test it – Query it / Validate it
    - Develop the model / skeleton
      - Understand your data
      - Understand the RDF Data Cube Vocabulary
      - Define Components
        - » Dimension(s)
        - » Measure(s)
        - » Attribute(s)
    - Populate data and metadata
      - Develop script (programmatic approach)
      - Execute script
    - Test / Validate Results
  - Publish it
- Query it / Re-Use it → Present Results

AR&M Project – POC Process

- Building the cube, an example …
  - From Results Source Data*1 and Cube Model (skeleton)

	A	B	C	D	E	F	G	H	I
1	measure	TRT01A	sex	SAF	procedu	factor	denominat	race	unit
2	86	Placebo	_ALL_	Y	count	quantity	_ALL_	_ALL_	_NULL_
3	84	Xanomeline High Dose	_ALL_	Y	count	quantity	_ALL_	_ALL_	_NULL_
4	84	Xanomeline Low Dose	_ALL_	Y	count	quantity	_ALL_	_ALL_	_NULL_
5	143	_ALL_	F	Y	count	quantity	_ALL_	_ALL_	_NULL_
6	111	_ALL_	M	Y	count	quantity	_ALL_	_ALL_	_NULL_
7	53	Placebo	F	Y	count	quantity	_ALL_	_ALL_	_NULL_
8	33	Placebo	M	Y	count	quantity	_ALL_	_ALL_	_NULL_
9	40	Xanomeline High Dose	F	Y	count	quantity	_ALL_	_ALL_	_NULL_
10	44	Xanomeline High Dose	M	Y	count	quantity	_ALL_	_ALL_	_NULL_
11	50	Xanomeline Low Dose	F	Y	count	quantity	_ALL_	_ALL_	_NULL_
12	34	Xanomeline Low Dose	M	Y	count	quantity	_ALL_	_ALL_	_NULL_
22	230	_ALL_	_ALL_	Y	count	quantity	_ALL_	WHITE	_NULL_
25	78	Placebo	_ALL_	Y	count	quantity	_ALL_	WHITE	_NULL_
28	74	Xanomeline High Dose	_ALL_	Y	count	quantity	_ALL_	WHITE	_NULL_
31	78	Xanomeline Low Dose	_ALL_	Y	count	quantity	_ALL_	WHITE	_NULL_
43	75.20930233	Placebo	_ALL_	Y	mean	AGE	_NULL_	_ALL_	YEARS
44	74.38095238	Xanomeline High Dose	_ALL_	Y	mean	AGE	_NULL_	_ALL_	YEARS
45	75.66666667	Xanomeline Low Dose	_ALL_	Y	mean	AGE	_NULL_	_ALL_	YEARS
46	8.590167127	Placebo	_ALL_	Y	stdev	AGE	_NULL_	_ALL_	YEARS

R script showing *rrdf* statements →

*1 Data summarization process not shown here

	A	B	C	D	E
1	compType	compName	compLabel	codeType	nciDomainValue
2	dimension	trt01a	Treatment Arm	DATA	
3	dimension	sex	Sex (Gender)	SDTM	C66731
4	dimension	saffl	Safety Population Flag	DATA	
5	dimension	procedure	Statistical Procedure	DATA	" "
6	dimension	factor	Type of procedure (quantity, proportion...)	DATA	
7	dimension	race	Race	SDTM	C74457
8	measure	measure	Value of the statistical measure		

qb.buildSkeleton.R - WordPad

Home View

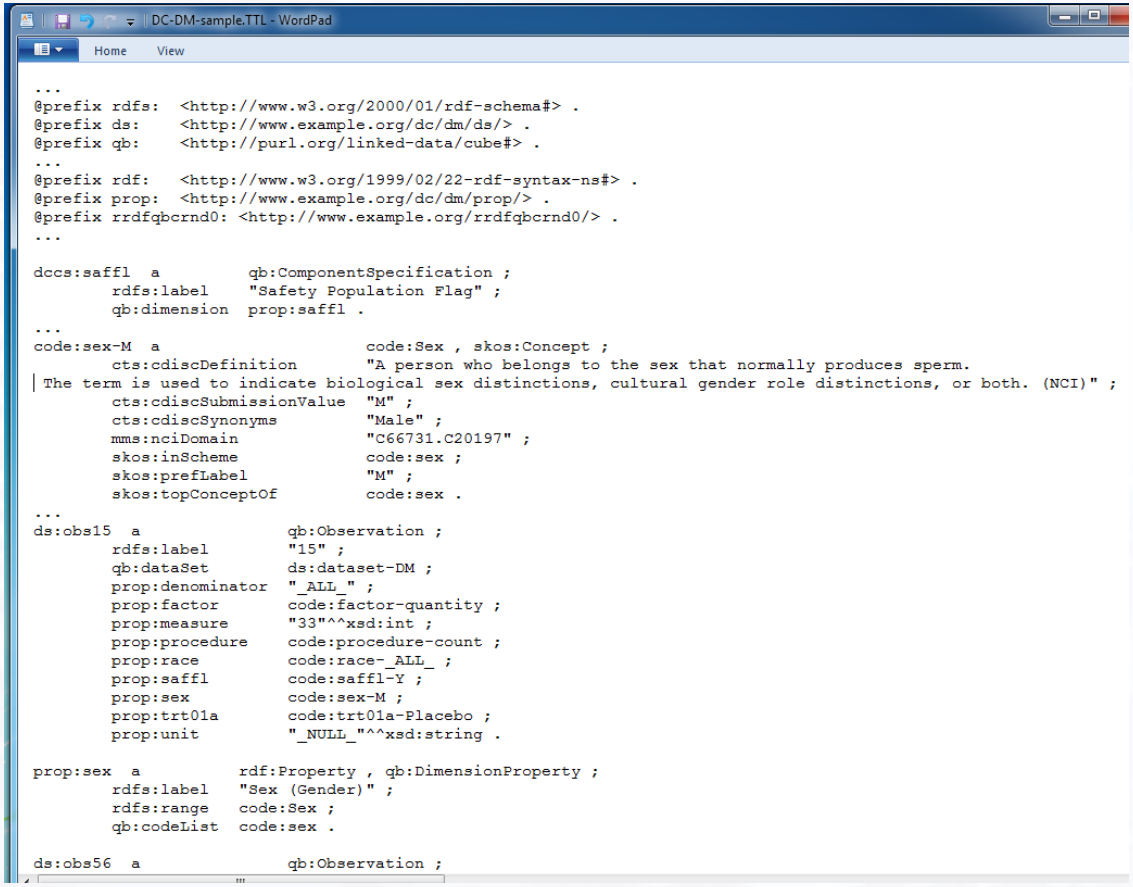
```
# Class name is the compName with an uppercase first letter.
compNameClass <- capitalize(toString(skeletonSource[i,"compName"]))
##DEBUG message ("Building: Dimension")
#  Example: prop:trt01a a qb:DimensionProperty
add.triple(store,
           paste0(prefixPROP, skeletonSource[i,"compName"]),
           "http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
           paste0(prefixQB, "DimensionProperty"))

add.triple(store,
           paste0(prefixPROP, skeletonSource[i,"compName"]),
           "http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
           paste0(prefixRDF, "Property"))

# Label for property. Example:   rdfs:label "Treatment Arm"@en .
add.data.triple(store,
                paste0(prefixPROP, skeletonSource[i,"compName"]),
                "http://www.w3.org/2000/01/rdf-schema#label",
                paste0(skeletonSource[i,"compLabel"]))
```

AR&M Project – POC Process

- Resulting RDF Data Cube extract, rendered in RDF Turtle format



AR&M Project – POC Process

- Presentation of Results – Generate TLFs from delivered RDF data cube
  - Not the current focus, just to illustrate process
  - Script (R, SAS, etc.) to present as a basic Safety Table in HTML or PDF format
    - Results and associated metadata obtained from SPARQL queries to the RDF data cube
      - » No additional data derivations in the TLF scripts
    - Displayed in any browser or saved in a repository for sharing

AR&M Project – POC Process

- Presentation of Results – Generate TLFs from delivered RDF data cube
  - POC target example
    - Replicate the Demographics table produced by the Scripts team, reading results and metadata from the POC *demog* RDF data cube
    - No additional data derivations in the TLF scripts!
    - One approach (Ex-1 by M. Hungria):
      - Sparql query in R to extract all DC observations and all DC properties (output as .tsv files)
      - Import .tsv files in SAS and transpose / transform structure to fit the report dataset structure designed by the Scripts team
      - Create Table using SAS Proc Report
      - All steps invoked from a SAS program executed in batch
- Further collaboration between the PhUSE CS Standard Scripts and AR&M teams planned

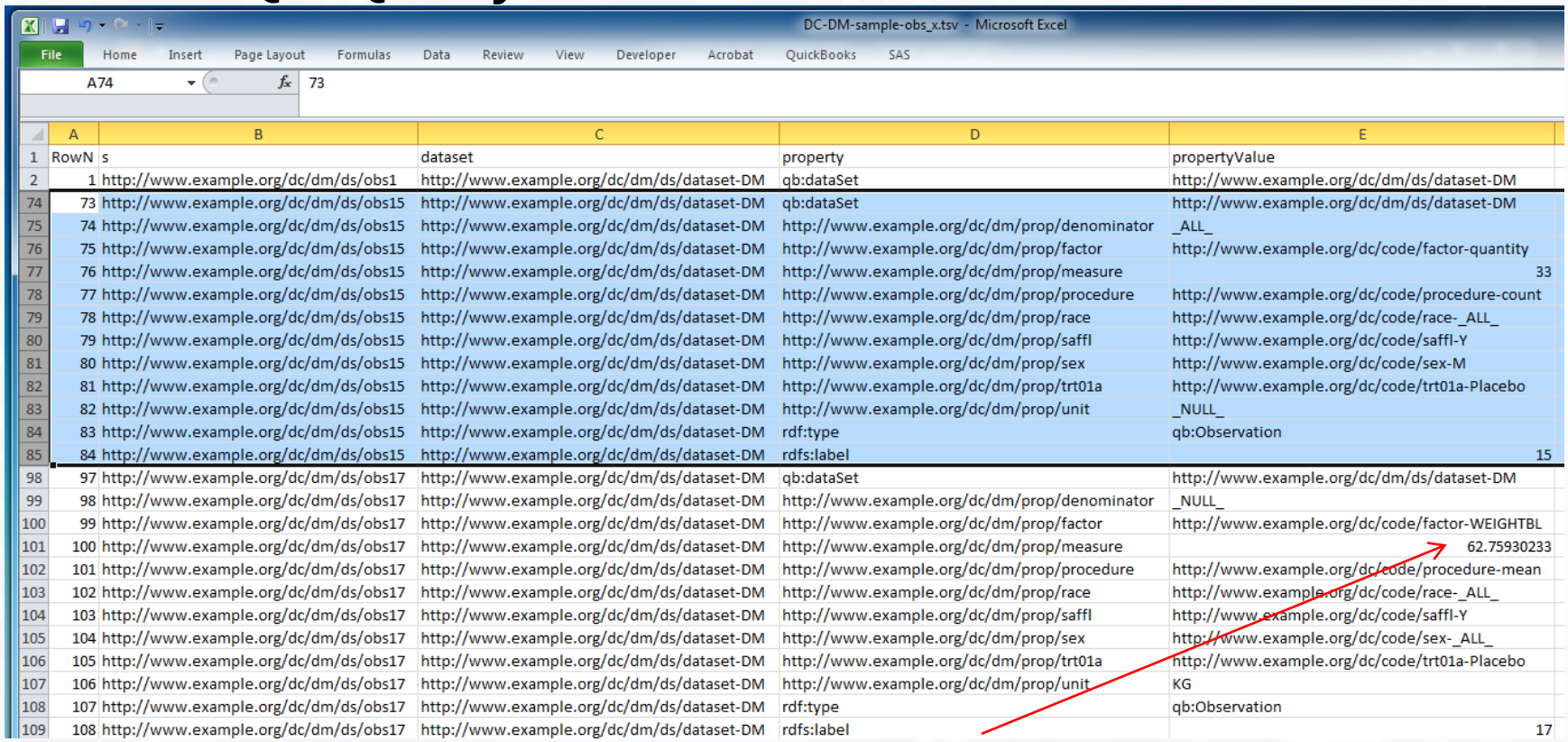
Linked Data: SPARQL

- SPARQL query to RDF Data Cube named *ds:dataset-demog*

```
> load.rdf(dataCubeFile, format="N3", appendTo=store)
> summarize.rdf(store)
[1] "Number of triples: 1192"
>
> ## ------------------------------------------------------------------------
> cube.check= sparql.rdf(store,
+                        paste(forsparqlprefix,
+                              "select * where {",
+                              "    ?s a qb:Observation  ;",
+                              "       qb:dataSet ?dataset ;",
+                              "       ?property        ?propertyValue ;        ",
+                              "} order by ?s"
+                        )
+ );
> #cube.check
>
> # write results to a cache file
>
> write.table(cube.check,file=checkCubeFileObs,sep="\t")
>
  --------------------
```

AR&M Project – POC Process

- SPARQL Query Results Ex-1, Observations



DC-DM-sample-obs_x.tsv - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Developer Acrobat QuickBooks SAS

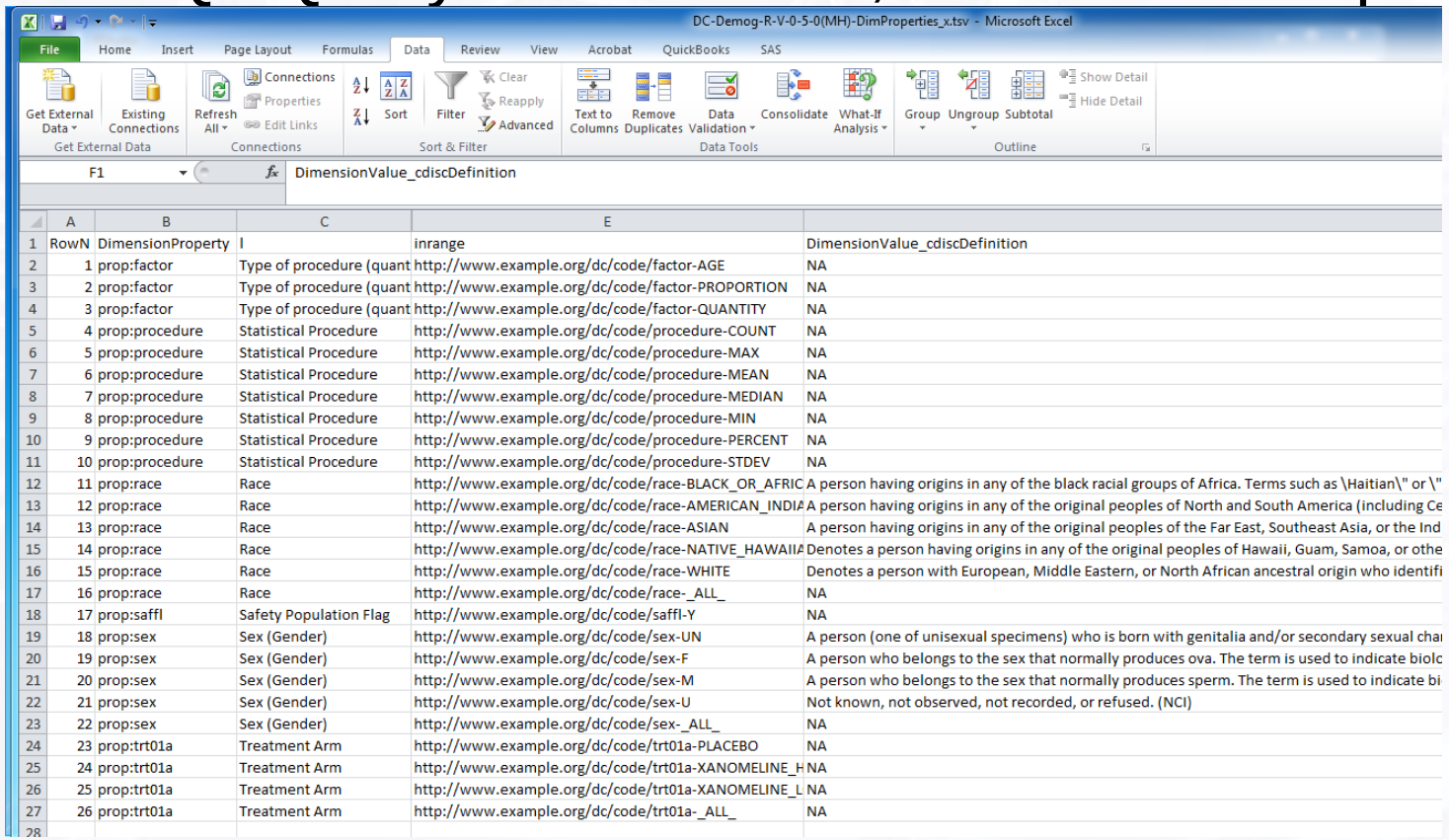
A74 fx 73

	A	B	C	D	E
1	RowN	s	dataset	property	propertyValue
2	1	http://www.example.org/dc/dm/ds/obs1	http://www.example.org/dc/dm/ds/dataset-DM	qb:dataSet	http://www.example.org/dc/dm/ds/dataset-DM
74	73	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	qb:dataSet	http://www.example.org/dc/dm/ds/dataset-DM
75	74	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/denominator	_ALL_
76	75	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/factor	http://www.example.org/dc/code/factor-quantity
77	76	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/measure	33
78	77	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/procedure	http://www.example.org/dc/code/procedure-count
79	78	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/race	http://www.example.org/dc/code/race-_ALL_
80	79	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/saffl	http://www.example.org/dc/code/saffl-Y
81	80	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/sex	http://www.example.org/dc/code/sex-M
82	81	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/trt01a	http://www.example.org/dc/code/trt01a-Placebo
83	82	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/unit	_NULL_
84	83	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	rdf:type	qb:Observation
85	84	http://www.example.org/dc/dm/ds/obs15	http://www.example.org/dc/dm/ds/dataset-DM	rdfs:label	15
98	97	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	qb:dataSet	http://www.example.org/dc/dm/ds/dataset-DM
99	98	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/denominator	_NULL_
100	99	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/factor	http://www.example.org/dc/code/factor-WEIGHTBL
101	100	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/measure	62.75930233
102	101	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/procedure	http://www.example.org/dc/code/procedure-mean
103	102	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/race	http://www.example.org/dc/code/race-_ALL_
104	103	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/saffl	http://www.example.org/dc/code/saffl-Y
105	104	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/sex	http://www.example.org/dc/code/sex-_ALL_
106	105	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/trt01a	http://www.example.org/dc/code/trt01a-Placebo
107	106	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	http://www.example.org/dc/dm/prop/unit	KG
108	107	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	rdf:type	qb:Observation
109	108	http://www.example.org/dc/dm/ds/obs17	http://www.example.org/dc/dm/ds/dataset-DM	rdfs:label	17

Displayed in Excel - tsv cell: 62.7593023255814

AR&M Project – POC Process

- SPARQL Query Results Ex-1, Dimension Properties



DC-Demog-R-V-0-5-0(MH)-DimProperties_x.tsv - Microsoft Excel

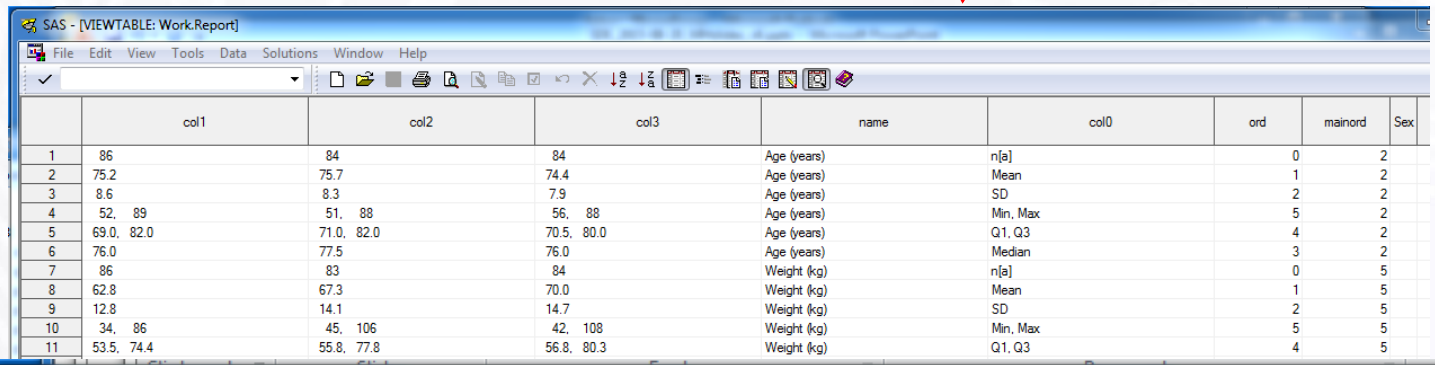
File Home Insert Page Layout Formulas Data Review View Acrobat QuickBooks SAS

Get External Data | Existing Connections | Refresh All | Connections | Properties | Edit Links | Sort | Filter | Clear | Reapply | Advanced | Text to Columns | Remove Duplicates | Data Validation | Consolidate | What-If Analysis | Group | Ungroup | Subtotal | Show Detail | Hide Detail

F1 DimensionValue_cdiscDefinition

	A	B	C	E	
1	RowN	DimensionProperty	l	inrange	DimensionValue_cdiscDefinition
2	1	prop:factor	Type of procedure (quant	http://www.example.org/dc/code/factor-AGE	NA
3	2	prop:factor	Type of procedure (quant	http://www.example.org/dc/code/factor-PROPORTION	NA
4	3	prop:factor	Type of procedure (quant	http://www.example.org/dc/code/factor-QUANTITY	NA
5	4	prop:procedure	Statistical Procedure	http://www.example.org/dc/code/procedure-COUNT	NA
6	5	prop:procedure	Statistical Procedure	http://www.example.org/dc/code/procedure-MAX	NA
7	6	prop:procedure	Statistical Procedure	http://www.example.org/dc/code/procedure-MEAN	NA
8	7	prop:procedure	Statistical Procedure	http://www.example.org/dc/code/procedure-MEDIAN	NA
9	8	prop:procedure	Statistical Procedure	http://www.example.org/dc/code/procedure-MIN	NA
10	9	prop:procedure	Statistical Procedure	http://www.example.org/dc/code/procedure-PERCENT	NA
11	10	prop:procedure	Statistical Procedure	http://www.example.org/dc/code/procedure-STDEV	NA
12	11	prop:race	Race	http://www.example.org/dc/code/race-BLACK_OR_AFRIC	A person having origins in any of the black racial groups of Africa. Terms such as \Haitian\" or \"
13	12	prop:race	Race	http://www.example.org/dc/code/race-AMERICAN_INDIA	A person having origins in any of the original peoples of North and South America (including Ce
14	13	prop:race	Race	http://www.example.org/dc/code/race-ASIAN	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Ind
15	14	prop:race	Race	http://www.example.org/dc/code/race-NATIVE_HAWAIIA	Denotes a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or othe
16	15	prop:race	Race	http://www.example.org/dc/code/race-WHITE	Denotes a person with European, Middle Eastern, or North African ancestral origin who identifi
17	16	prop:race	Race	http://www.example.org/dc/code/race-_ALL_	NA
18	17	prop:saffl	Safety Population Flag	http://www.example.org/dc/code/saffl-Y	NA
19	18	prop:sex	Sex (Gender)	http://www.example.org/dc/code/sex-UN	A person (one of unisexual specimens) who is born with genitalia and/or secondary sexual cha
20	19	prop:sex	Sex (Gender)	http://www.example.org/dc/code/sex-F	A person who belongs to the sex that normally produces ova. The term is used to indicate biolc
21	20	prop:sex	Sex (Gender)	http://www.example.org/dc/code/sex-M	A person who belongs to the sex that normally produces sperm. The term is used to indicate bi
22	21	prop:sex	Sex (Gender)	http://www.example.org/dc/code/sex-U	Not known, not observed, not recorded, or refused. (NCI)
23	22	prop:sex	Sex (Gender)	http://www.example.org/dc/code/sex-_ALL_	NA
24	23	prop:trt01a	Treatment Arm	http://www.example.org/dc/code/trt01a-PLACEBO	NA
25	24	prop:trt01a	Treatment Arm	http://www.example.org/dc/code/trt01a-XANOMELINE_H	NA
26	25	prop:trt01a	Treatment Arm	http://www.example.org/dc/code/trt01a-XANOMELINE_L	NA
27	26	prop:trt01a	Treatment Arm	http://www.example.org/dc/code/trt01a-_ALL_	NA
28					

Report data: Scripts WG vs. Ex-1


SAS - [VIEWTABLE: Work.Report]

File Edit View Tools Data Solutions Window Help

	col1	col2	col3	name	col0	ord	mainord	Sex
1	86	84	84	Age (years)	n[a]	0	2	
2	75.2	75.7	74.4	Age (years)	Mean	1	2	
3	8.6	8.3	7.9	Age (years)	SD	2	2	
4	52, 89	51, 88	56, 88	Age (years)	Min, Max	5	2	
5	69.0, 82.0	71.0, 82.0	70.5, 80.0	Age (years)	Q1, Q3	4	2	
6	76.0	77.5	76.0	Age (years)	Median	3	2	
7	86	83	84	Weight (kg)	n[a]	0	5	
8	62.8	67.3	70.0	Weight (kg)	Mean	1	5	
9	12.8	14.1	14.7	Weight (kg)	SD	2	5	
10	34, 86	45, 106	42, 108	Weight (kg)	Min, Max	5	5	
11	53.5, 74.4	55.8, 77.8	56.8, 80.3	Weight (kg)	Q1, Q3	4	5	

SAS - [VIEWTABLE: Work.Report2]

File Edit View Tools Data Solutions Window Help

	mainord	ord	cat	scat	trt_ord	treatment	value	pct	value_display	start	label
1	1	1	Sex n(%)	n[a]	0	Placebo	86	100.0	86 (100.0%)	1	86 (100.0%)
2	1	1	Sex n(%)	n[a]	54	Xanomeline Low Dose	84	100.0	84 (100.0%)	2	84 (100.0%)
3	1	1	Sex n(%)	n[a]	81	Xanomeline High Dose	84	100.0	84 (100.0%)	3	84 (100.0%)
4	1	2		F	.	All	143		143	4	143
5	1	2		F	0	Placebo	53	61.6279069767442	53 (61.6%)	5	53 (61.6%)
6	1	2		F	54	Xanomeline Low Dose	50	59.5238095238095	50 (59.5%)	6	50 (59.5%)
7	1	2		F	81	Xanomeline High Dose	40	47.6190476190476	40 (47.6%)	7	40 (47.6%)
8	1	3		M	.	All	111		111	8	111
9	1	3		M	0	Placebo	33	38.3720930232558	33 (38.3%)	9	33 (38.3%)
10	1	3		M	54	Xanomeline Low Dose	34	40.4761904761905	34 (40.4%)	10	34 (40.4%)
11	1	3		M	81	Xanomeline High Dose	44	52.3809523809524	44 (52.3%)	11	44 (52.3%)
12	2	.		Stdev	0	Placebo	8.59016712714194		8.5	12	8.5

Proc Report code: Scripts WG vs. Ex-1

```
proc report data=report ls=150 ps=45 split="@"  headline headskip nocenter
nowd missing spacing=2;
  column mainord name ord col0 col1 col2 col3;

  define mainord / order order=internal noprint;
  define ord  / order order=internal noprint;
  define name / display " " width=30 spacing=0 flow order;
  define col0 / display " " width=30 ;
  define col1 / display "&trtname1.@      (N=%qcmpres(&trt1))" width=20;
  define col2 / display "&trtname2.@      (N=%qcmpres(&trt2))" width=20;
  define col3 / display "&trtname3.@     (N=%qcmpres(&trt3))" width=20;

  break after mainord / skip;
run;
```

```
proc report data=report2 ls=150 ps=45 split="@"  headline
   headskip nocenter nowd missing spacing=2;

   column mainord ord cat scat treatment,obsn ;
   define mainord / group order=data noprint;
   define ord  / group order=data noprint;

   define cat / group " " order=data width=62 spacing=0 flow ;
   define scat / group " " order=data width=32 ;
   define treatment   / across "Treatment Arm" center order=data
     format=$trtfmt. style(column) = [asis=on cellwidth = 1.75 in];
   define obsn /analysis " " left sum format=dvalue30. width=30;

   break after cat / skip;

run;
```

The control FMTDS Dataset

start	label	fmtname	type
1	86 (100.0%)	dvalue	n
2	84 (100.0%)	dvalue	n
3	84 (100.0%)	dvalue	n
4	143	dvalue	n
5	53 (61.6%)	dvalue	n
6	50 (59.5%)	dvalue	n
7	40 (47.6%)	dvalue	n
8	111	dvalue	n
9	33 (38.3%)	dvalue	n
10	34 (40.4%)	dvalue	n
11	44 (52.3%)	dvalue	n
12	8.5	dvalue	n

Demographics Table - Scripts WG

		Placebo (N=86)	Xanomeline Low Dose (N=84)	Xanomeline High Dose (N=84)
Sex n(%)	n[a]	86 (100.0%)	84 (100.0%)	84 (100.0%)
	F	53 (61.6%)	50 (59.5%)	40 (47.6%)
	M	33 (38.4%)	34 (40.5%)	44 (52.4%)
Age (years)	n[a]	86	84	84
	Mean	75.2	75.7	74.4
	SD	8.6	8.3	7.9
	Median	76.0	77.5	76.0
	Q1, Q3	69.0, 82.0	71.0, 82.0	70.5, 80.0
	Min, Max	52, 89	51, 88	56, 88
Age categories n(%)	n[a]	86 (100.0%)	84 (100.0%)	84 (100.0%)
	<65	14 (16.3%)	8 (9.5%)	11 (13.1%)
	65-80	42 (48.8%)	47 (56.0%)	55 (65.5%)
	>80	30 (34.9%)	29 (34.5%)	18 (21.4%)
Race n(%)	n[a]	86 (100.0%)	84 (100.0%)	84 (100.0%)
	White	78 (90.7%)	78 (92.9%)	74 (88.1%)
	Black or African American	8 (9.3%)	6 (7.1%)	9 (10.7%)
	American Indian or Alaska Native	0	0	1 (1.2%)
Ethnicity n(%)	n[a]	86 (100.0%)	84 (100.0%)	84 (100.0%)
	Not Hispanic or Latino	83 (96.5%)	78 (92.9%)	81 (96.4%)
	Hispanic or Latino	3 (3.5%)	6 (7.1%)	3 (3.6%)
Weight (kg)	n[a]	86	83	84
	Mean	62.8	67.3	70.0
	SD	12.8	14.1	14.7

Demographics Table – Ex-1

TABLE x.y.z
DEMOGRAPHICS AND BASELINE CHARACTERISTICS (SAFFL='Y')

		Treatment Arm			
		Placebo (N=86)	Xanomeline Low Dose (N=84)	Xanomeline High Dose (N=84)	All
Sex n(%)	n[a]	86 (100.0%)	84 (100.0%)	84 (100.0%)	**NA**
	F	53 (61.6%)	50 (59.5%)	40 (47.6%)	143
	M	33 (38.3%)	34 (40.4%)	44 (52.3%)	111
Age (years)	Mean	75.2	75.6	74.3	**NA**
	Median	76	77.5	76	**NA**
	Stdev	8.5	8.2	7.8	**NA**
	Min	52	51	56	**NA**
	Max	89	88	88	**NA**
Age categories n(%)	**NA**				
Race n(%)	n[a]	86 (100.0%)	84 (100.0%)	84 (100.0%)	**NA**
	White	78 (90.6%)	78 (92.8%)	74 (88.0%)	230
	Black or African American	8 (9.3%)	6 (7.1%)	9 (10.7%)	23
	American Indian or Alaska Native	0 (0.%)	0 (0.%)	1 (1.1%)	1
Ethnicity n(%)	**NA**				
Weight BL (kg)	Mean	62.7	67.2	70.0	**NA**
	Median	60.5	64.9	69.2	**NA**
	Stdev	12.7	14.1	14.6	**NA**
	Min	34	45.4	41.7	**NA**
	Max	86.2	106.1	108	**NA**

A&RM Data Cube - Benefits

- Potentially provides better Traceability mechanisms
- Facilitates Analysis Results Validation
- Enables Reusability of Results → Data Integration
- Facilitates exploratory analysis
- Potentially facilitates the Regulatory Review Process
  - By minimizing or complementing the use of PDF files as a delivery mechanism for statistical results

More details in the “Representing Analysis Results & Metadata as an RDF Data Cube” White Paper – stay tuned!

Roadmap

- Planned Activities [*1]
  - POC (under way)
    - AR&M team is working on an R package with RDF Data Cube functionality[*2]
    - RDF Data Cube Structure - Draft Technical Guidance completed – R package re-design
    - AR&M White Paper
  - More source data
    - Other Data Domains (beyond Demographics and Adverse Events)
    - Other data transport formats (e.g., Dataset-XML)
  - More items to model
    - Multiple measures, complex statistics, other dimension attributes, hierarchies
  - More links
    - Link to other CDISC Standards and non-CDISC Standards in RDF
    - Link to Define-XML and other documents (e.g.., SAP, Protocol)
  - Presentation Layer

[*1] A more complete/formal roadmap and tentative timelines may be available in the future on the PhUSE wiki, PhUSE - CS Analysis Results & Metadata Team page

[*2] Draft name *rrdfqbcrnd* (R-RDF Data Cube for Clinical Research & Development) – final name TBD

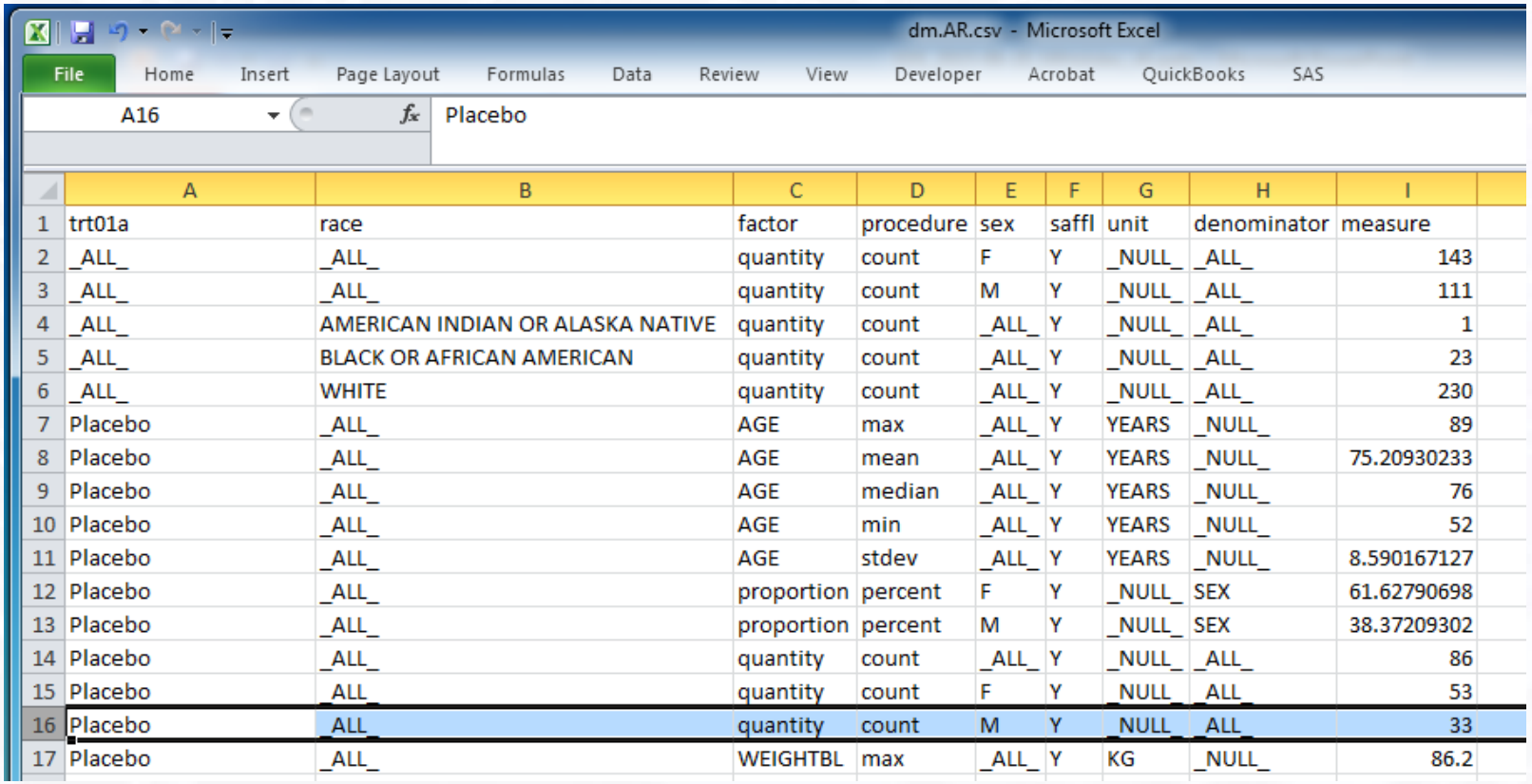
References

- W3C, http://www.w3.org
  - The RDF Data Cube Vocabulary and its Integrity Constraints, http://www.w3.org/TR/vocab-data-cube
- CDISC, http://cdisc.org
  - *CDISC Public Webinar - Standards Updates and Additions 20140724*, by Frederik Malfait and Scott Bahlavooni
- Papers and AR&M Team member's presentations
  - *Linked data to support Clinical and Non-Clinical Reporting*, Marc Andersen, SemStats 2014, Second International Workshop on Semantic Statistics, ISWC 2014, 20 October 2014 https://semstats2014.files.wordpress.com/2014/10/semstats2014_submission_5.pdf
  - *A Primer on Converting Analysis Results Data to RDF Data Cubes using Free and Open Source Tools*, Tim Williams, PhUSE 2014, Paper TT03
- AR&M Team Information and Deliverables
  - PhUSE Wiki, AR&M Team Page, http://www.phusewiki.org/wiki/index.php?title=Analysis_Results_Model
  - GitHub repository planned, https://github.com/phuse-org/rrdfqbcrnd

Supplementary Slides

AR&M Project – POC Process

- Summarized source data


dm.AR.csv - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Developer Acrobat QuickBooks SAS

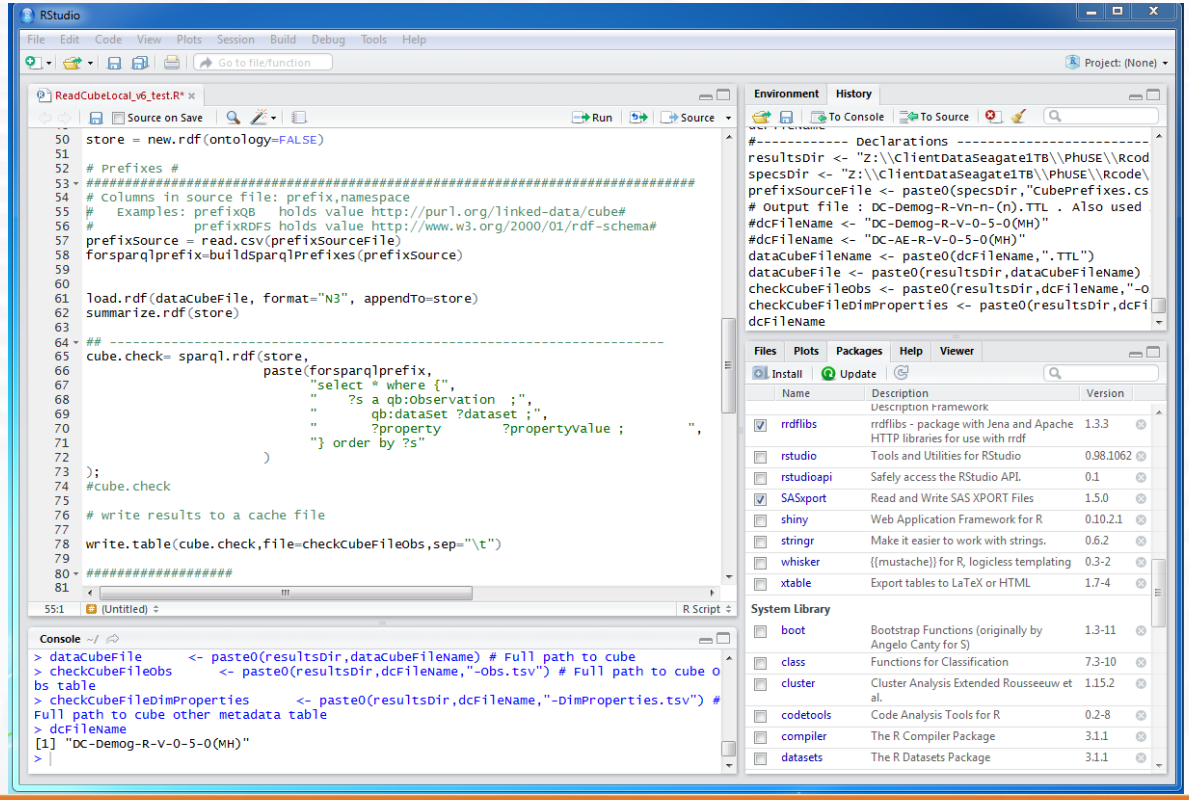
A16 Placebo

	A	B	C	D	E	F	G	H	I
1	trt01a	race	factor	procedure	sex	saffl	unit	denominator	measure
2	_ALL_	_ALL_	quantity	count	F	Y	_NULL_	_ALL_	143
3	_ALL_	_ALL_	quantity	count	M	Y	_NULL_	_ALL_	111
4	_ALL_	AMERICAN INDIAN OR ALASKA NATIVE	quantity	count	_ALL_	Y	_NULL_	_ALL_	1
5	_ALL_	BLACK OR AFRICAN AMERICAN	quantity	count	_ALL_	Y	_NULL_	_ALL_	23
6	_ALL_	WHITE	quantity	count	_ALL_	Y	_NULL_	_ALL_	230
7	Placebo	_ALL_	AGE	max	_ALL_	Y	YEARS	_NULL_	89
8	Placebo	_ALL_	AGE	mean	_ALL_	Y	YEARS	_NULL_	75.20930233
9	Placebo	_ALL_	AGE	median	_ALL_	Y	YEARS	_NULL_	76
10	Placebo	_ALL_	AGE	min	_ALL_	Y	YEARS	_NULL_	52
11	Placebo	_ALL_	AGE	stdev	_ALL_	Y	YEARS	_NULL_	8.590167127
12	Placebo	_ALL_	proportion	percent	F	Y	_NULL_	SEX	61.62790698
13	Placebo	_ALL_	proportion	percent	M	Y	_NULL_	SEX	38.37209302
14	Placebo	_ALL_	quantity	count	_ALL_	Y	_NULL_	_ALL_	86
15	Placebo	_ALL_	quantity	count	F	Y	_NULL_	_ALL_	53
16	Placebo	_ALL_	quantity	count	M	Y	_NULL_	_ALL_	33
17	Placebo	_ALL_	WEIGHTBL	max	_ALL_	Y	KG	_NULL_	86.2

AR&M Project – POC Process

– Testing the cube …

- SPARQL query directly in R – cube loaded from local repository



Questions?

PhUSE SDE Meeting, NJ 2015

Thank you!

Marcelina Hungria / DIcore Group, LLC
mhungria@dicoregroup.com