

When the data *is* a graph:

An Introduction to Directed Graphs for Modelling and Visualizing Clinical Trials Data & Metadata.

Tim Williams
Principal Statistical Solutions Analyst
UCB Biosciences, Inc.
Raleigh, NC

PhUSE SDE
04 June 2015

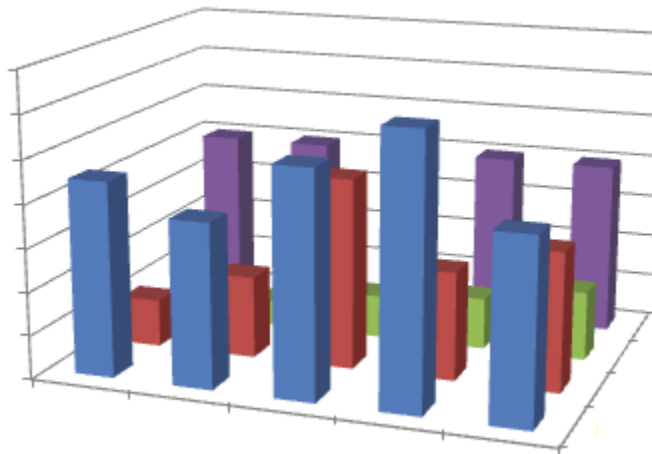


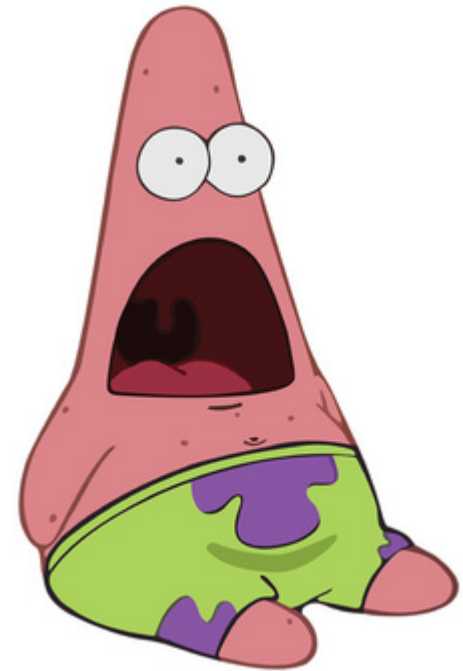
"I want to take the **clinical trials results...**"

	Placebo	LowDose	HighDose
Baseline	N=28	N=30	N=29

Sex			
F	12 (42.9)	14 (46.7)	16 (55.2)
M	16 (57.1)	16 (53.3)	13 (44.8)

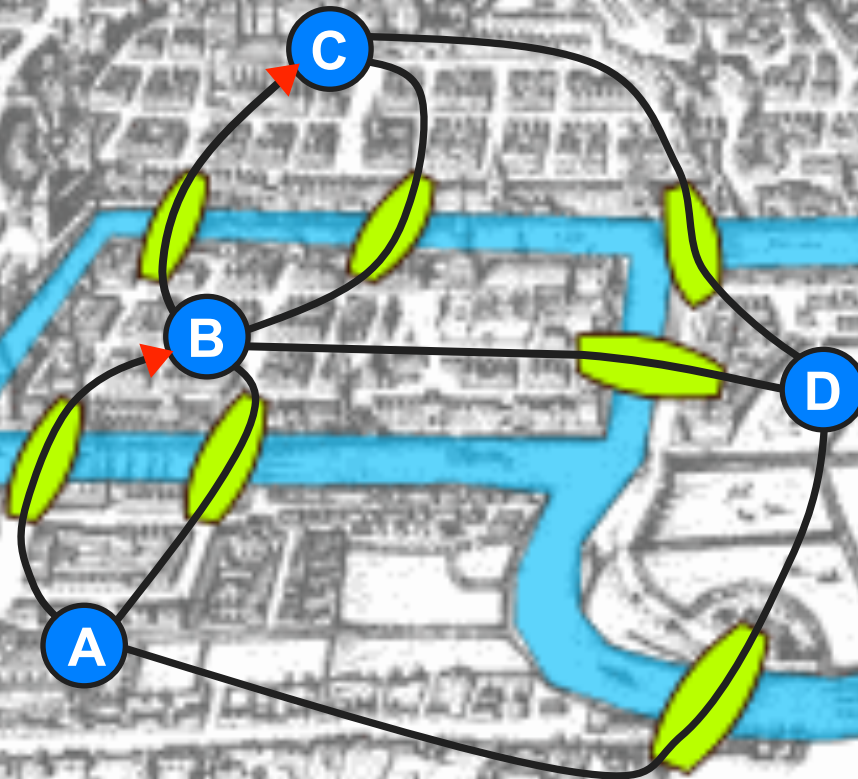
"..and put them in a pretty graph!"





Introduction to Directed Graphs

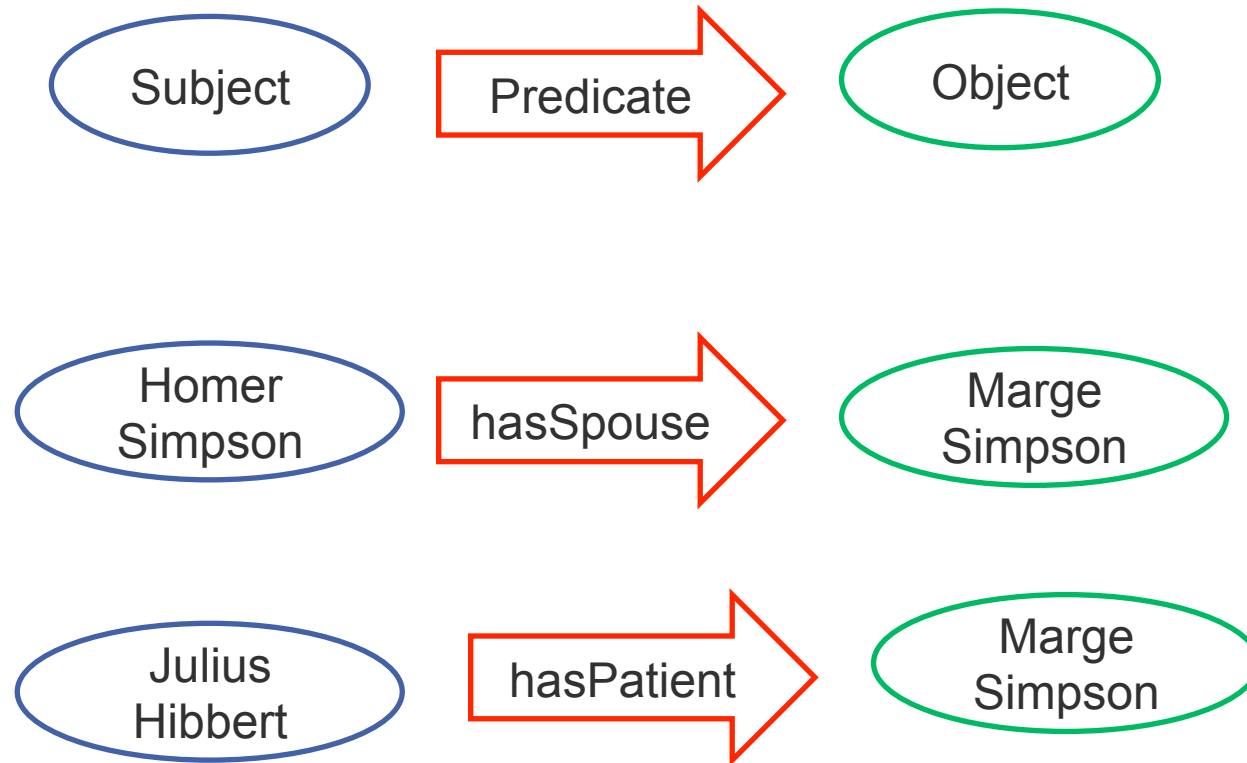
KONINGSBERGA

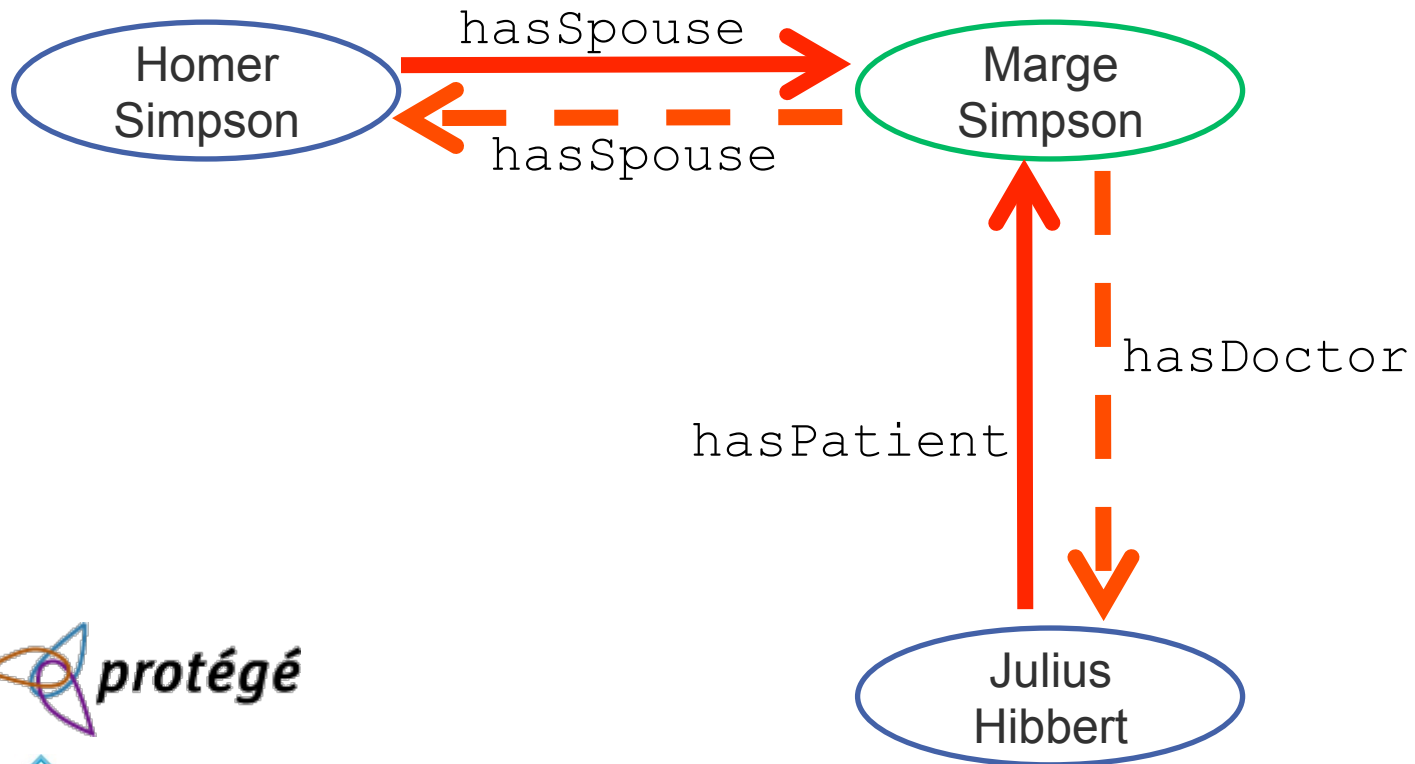


1735



Resource Description Framework (RDF)





TopBraid Composer™

What does it look like in RDF (TURTLE) syntax ?

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ts: <http://www.example.com/ns/theSimpsons#> .
```

```
ts:i7832
```

```
  ts:firstName "Homer" ;
  ts:lastName "Simpson" ;
  ts:hasSpouse ts:i5628 .
```

```
ts:i2314
```

```
  ts:firstName "Julius" ;
  ts:lastName "Hibbert";
  ts:hasPatient ts:i5628 .
```

```
ts:i5628
```

```
  ts:firstName "Marge";
  ts:lastName "Simpson";
```

```
ts:hasSpouse
```

```
  rdf:type      owl:Symmetric Property ;
  rdfs:comment  "Identifies someone's spouse" .
```

```
ts:hasPatient
```

```
  rdf:type rdf:Property
  rdfs:comment "Identifies a doctor's patient" .
```

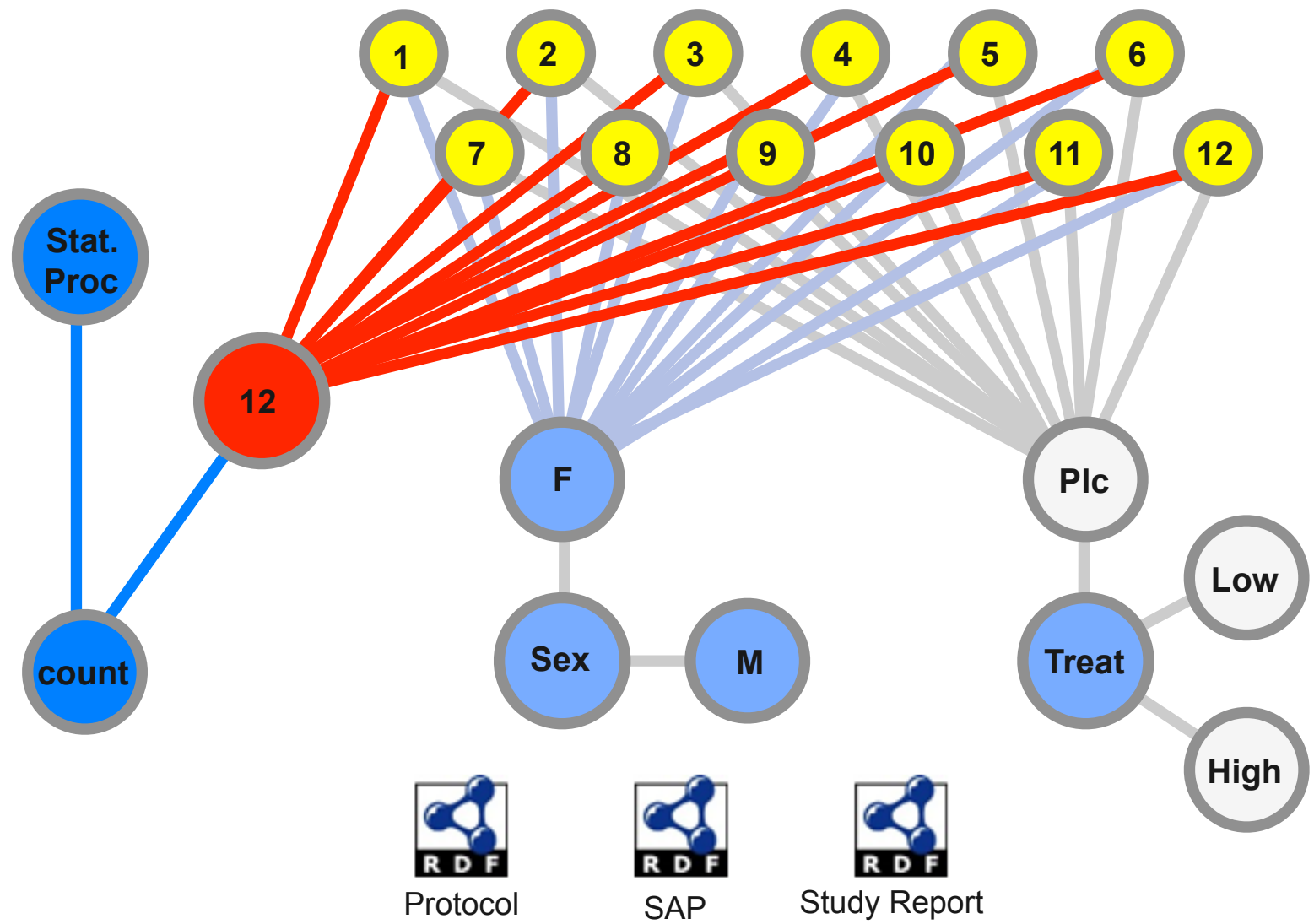
```
ts:hasDoctor
```

```
  rdf:type      rdf:Property ;
  rdfs:comment  "Identifies a patient's doctor" ;
  owl:inverseOf ts:hasPatient
```

Semantic Web & Linked Data: Key Concepts

- Addressable observations, groups, concepts
 - **Unique, linkable** identifiers
 - Uniform (Universal) Resource Identifier (URI)
 - HTTP URI for lookup and linkage
 - Provides a standard for way to represent values and concepts
 - *Links* to common vocabularies, code lists and concepts allow data reconciliation and standardization
 - *Enables* reuse, standard tools and components
 - *Facilitates* data integration: Combine similar and diverse data sources.
 - Integrate data across studies, products.
 - Machine readable [*interpretable*]
 - Inferencing, logic and decision-making
- ***NOT limited to the web!***

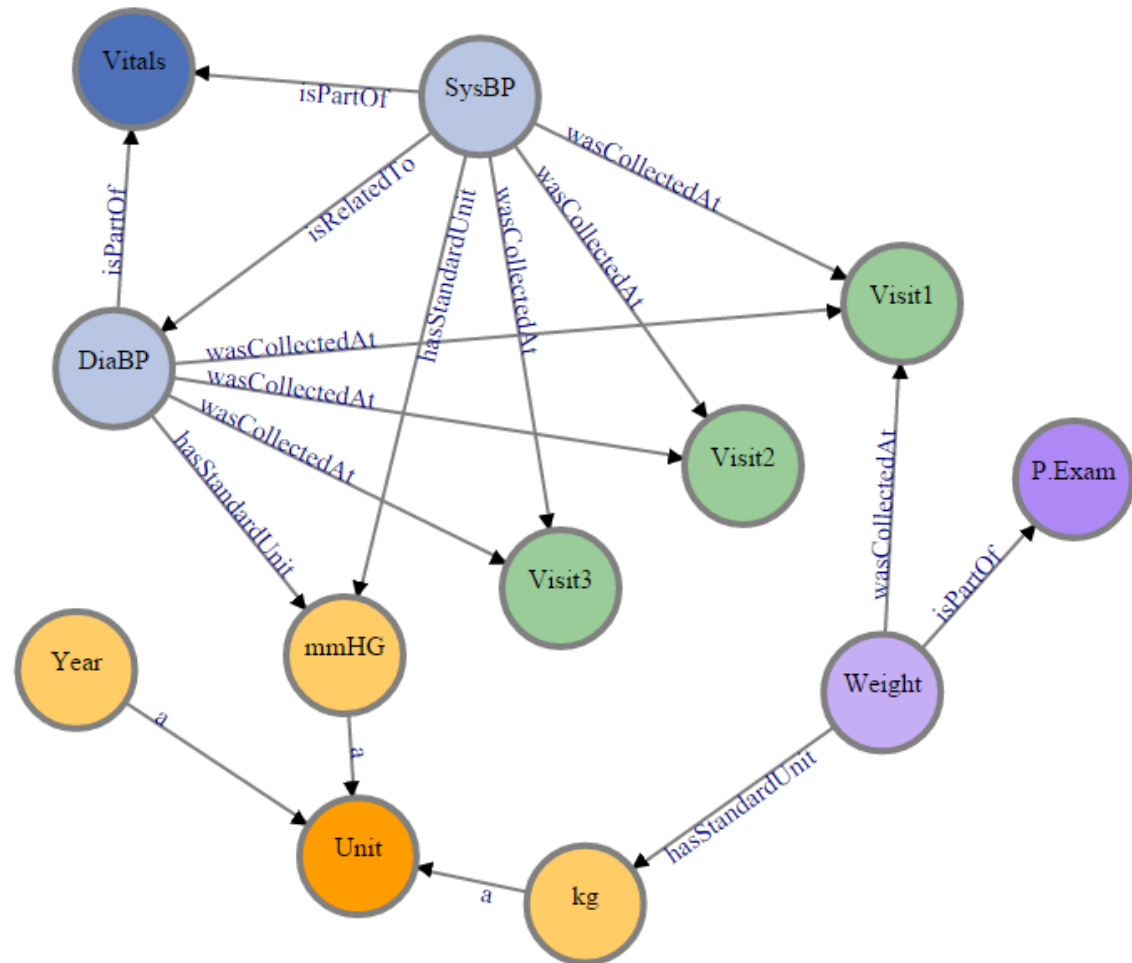
Results Traceability



LIVE DEMO: Blood Pressure Model as Linked Data

Blood Pressure (partial graph)

Adapted from Ian Fleming, PhUSE 2014



LIVE DEMO : DBPedia mmHG triples at DBPedia

About: Millimeter of mercury

An Entity of Type : [Thing](#), from Named Graph : <http://dbpedia.org>, within Data Space : dbpedia.org



A millimeter of mercury is a manometric unit of pressure, formerly defined as the extra pressure generated by a column of mercury one millimetre high. It is now defined precisely as $13.5951 \times 9.80665 = 133.322387415$ pascals. It is denoted by the symbol "mm Hg". Although not an SI unit, the millimeter of mercury is still routinely used in many fields like medicine etc]. One millimeter of mercury is approximately 1 Torr, or 1/760 of standard atmospheric pressure.

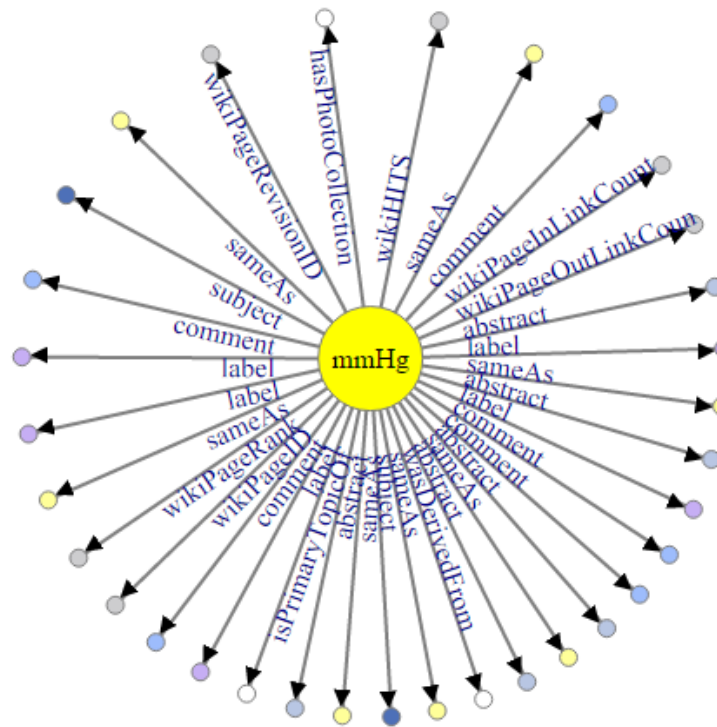
Property	Value
dbpedia-owl:abstract	A millimeter of mercury is a manometric unit of pressure, formerly defined as the extra pressure generated by a column of mercury one millimetre high. It is now defined precisely as $13.5951 \times 9.80665 = 133.322387415$ pascals. It is denoted by the symbol "mm Hg". Although not an SI unit, the millimeter of mercury is still routinely used in many fields like medicine etc]. One millimeter of mercury is approximately 1 Torr, or 1/760 of standard atmospheric pressure. The two units are not exactly equal; however, the difference (less than 0.000015%) is negligible for most practical uses.
dbpedia-owl:wikiPageID	1294851 (xsd:integer)
dbpedia-owl:wikiPageRevisionID	605881240 (xsd:integer)
dbpprop:hasPhotoCollection	http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/photos/Millimeter_of_mercury
dcterms:subject	- category:Units_of_pressure - category:Mercury_(element)
rdfs:comment	A millimeter of mercury is a manometric unit of pressure, formerly defined as the extra pressure generated by a column of mercury one millimetre high. It is now defined precisely as $13.5951 \times 9.80665 = 133.322387415$ pascals. It is denoted by the symbol "mm Hg". Although not an SI unit, the millimeter of mercury is still routinely used in many fields like medicine etc]. One millimeter of mercury is approximately 1 Torr, or 1/760 of standard atmospheric pressure.
rdfs:label	Millimeter of mercury
owl:sameAs	- fbase:Millimeter of mercury http://wikidata.dbpedia.org/resource/Q6859652 http://ko.dbpedia.org/resource/MmHg http://nl.dbpedia.org/resource/Mm_Hg http://pt.dbpedia.org/resource/Milímetro_de_mercúrio http://wikidata.org/entity/Q6859652 http://th.dbpedia.org/resource/มิลลิเมตรปรอท
prov:wasDerivedFrom	http://en.wikipedia.org/wiki/Millimeter_of_mercury?oldid=605881240
foaf:isPrimaryTopicOf	http://en.wikipedia.org/wiki/Millimeter_of_mercury
is dbpedia-owl:wikiPageRedirects of	- dbpedia:Millimetre_of_mercury - dbpedia:Millimeters_of_mercury - dbpedia:MmHg
is owl:sameAs of	http://wikidata.dbpedia.org/resource/Q6859652 http://th.dbpedia.org/resource/มิลลิเมตรปรอท
is foaf:primaryTopic of	http://en.wikipedia.org/wiki/Millimeter_of_mercury

LIVE DEMO : DBpedia mmHg triples as a Force Network Graph



mmHg

Triples from DBpedia SPARQL Endpoint



Analysis Results as RDF Data Cubes

"I want to take the **clinical trials results...**"

	Placebo	LowDose	HighDose
Baseline	N=28	N=30	N=29

Sex			
F	12 (42.9)	14 (46.7)	16 (55.2)
M	16 (57.1)	16 (53.3)	13 (44.8)

"..and put them in an **RDF Data Cube!**"

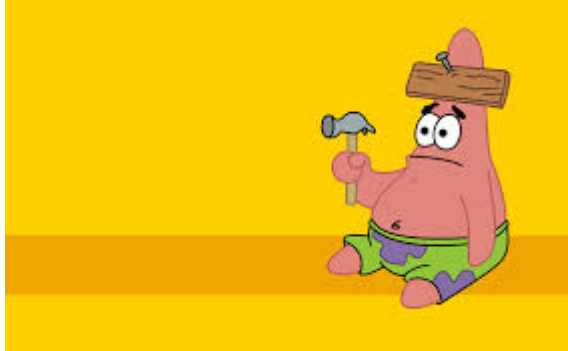


```
ds:obs1 a qb:Observation ;
    dim:treat      code:trt-Placebo ;
    dim:sex        code:sex-F ;
    dim:procedure  code:procedure-count ;
    meas:measure   "12"^^xsd:int ;
    qb:dataSet ds:dataset-DM .
```

```
ds:obs2 a qb:Observation ;
```

```
...
```

How?



<https://github.com/MarcJAndersen/rrdfqbc rnd0>

GitHub This repository Search Explore Features Enterprise Blog Sign up Sign in

MarcJAndersen / rrdfqbc rnd0 Watch 6 Star 0 Fork 0

Generating R-RDF Data Cube for Clinical Research & Development, work from a subgroup of PhUSE Semantic Technology Project

520 commits 6 branches 0 releases 3 contributors

branch: master rrdfqbc rnd0 / +

New version

File	Change	Time
ReleasePackages	New version	5 days ago
rrdfqbc rnd0	Updated version	5 days ago
.gitignore	Updated wording	22 days ago
README.md	Updated dependencies	13 days ago

latest commit 226ee3e518

README.md

Repository for R-RDF Data Cube for Clinical Research & Development

Code Issues 7 Pull requests 0 Pulse Graphs

HTTPS clone URL <https://github.com/>

You can clone with [HTTPS](#) or [Subversion](#)

Clone in Desktop Download ZIP



Analysis Results & Metadata Working Group





The RDF Data Cube Vocabulary

W3C Recommendation 16 January 2014

This version:

<http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/>

Latest published version:

<http://www.w3.org/TR/vocab-data-cube/>

Implementation report:

http://www.w3.org/2011/gld/wiki/Data_Cube_Implementations

Previous version:

<http://www.w3.org/TR/2013/PR-vocab-data-cube-20131217/>

Editors:

[Richard Cyganiak](#), [DERI, NUI Galway](#)

[Dave Reynolds](#), [Epimorphics Ltd](#)

Contributors:

[Jeni Tennison](#)

Please refer to the [errata](#), a list of issues with this document discovered after publication.

This document is also available in this non-normative format: [diff to previous version](#)

The English version of this specification is the only normative version. Non-normative [translations](#) may also be available.

[Copyright](#) © 2012-2014 [W3C](#)® ([MIT](#), [ERCIM](#), [Keio](#), [Beihang](#)), All Rights Reserved. W3C [liability](#), [trademark](#) and [document use](#) rules apply.

Abstract

There are many situations where it would be useful to be able to publish multi-dimensional data, such as statistics, on the

3 Main Components in the Cube Model

- **Attributes**

- metadata
- `status=final, issued="2014-08-06T00:00:00"^^xsd:dateTime ;`

- **Measure (or Primary Measure)**

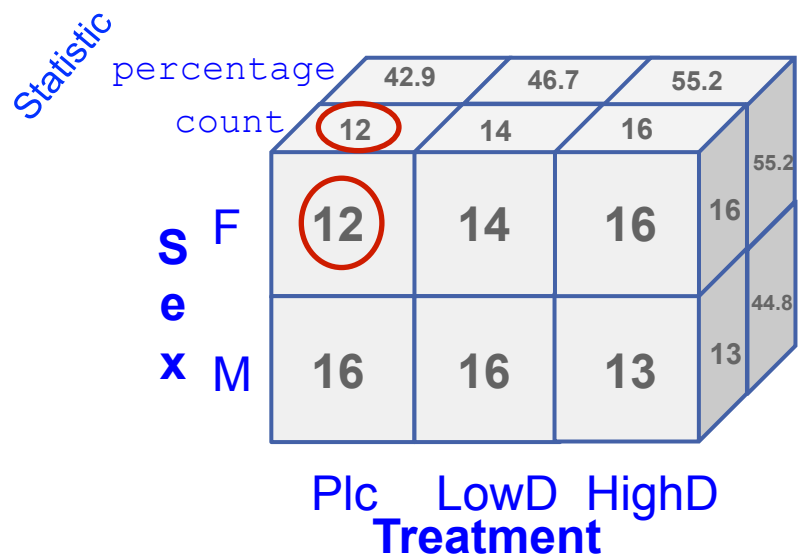
- the observed value of primary interest
- `count=12`

- **Dimensions**

- value keys or indices that identify the measure
- `treatment="Plc" , sex="F", statistic="count"`

		Treatment			Statistic • count • percentage
Baseline Characteristic		Placebo	LowDose	HighDose	
		N=28	N=30	N=29	

Sex					
	F	12 (42.9)	14 (46.7)	16 (55.2)	
	M	16 (57.1)	16 (53.3)	13 (44.8)	



Clinical Trials Data as RDF

Baseline Characteristic	Placebo N=28	LowDose N=30	HighDose N=29
Sex			
F	12 (42.9)	14 (46.7)	16 (55.2)
M	16 (57.1)	16 (53.3)	13 (44.8)

```

ds:obs1 a qb:Observation ;
    dim:treat      code:trt-Placebo ;
    dim:sex        code:sex-F ;
    dim:procedure  code:procedure-count ;
    meas:measure   '12'^^xsd:int ;
    qb:dataSet     ds:dataset-DM . .
  
```

```
ds:obs2 a qb:Observation ;
```

```
...
```

Statistic

count

Sex

Plc

Treatment

	42.9	46.7	55.2	
F	12	14	16	55.2
	12	14	16	16
	16	16	13	44.8
				13

Measure

Dimensions

`treatment="Plc",`
`sex="F",`
`statistic="count"`

It's a hit!!
 count=12



Query the data with SPARQL



```
SELECT ?value
WHERE
```

```
{
  ?obs dim:treat      code:trt-Placebo;
       dim:sex        code:sex-F;
       dim:procedure  code:procedure-count;
       meas:measure   ?value.
}
```

Where did I go
wrong with this
child?

I blame *The
Internets*, honey.



Components for Analysis and Visualization



rrdf, rrdflibs



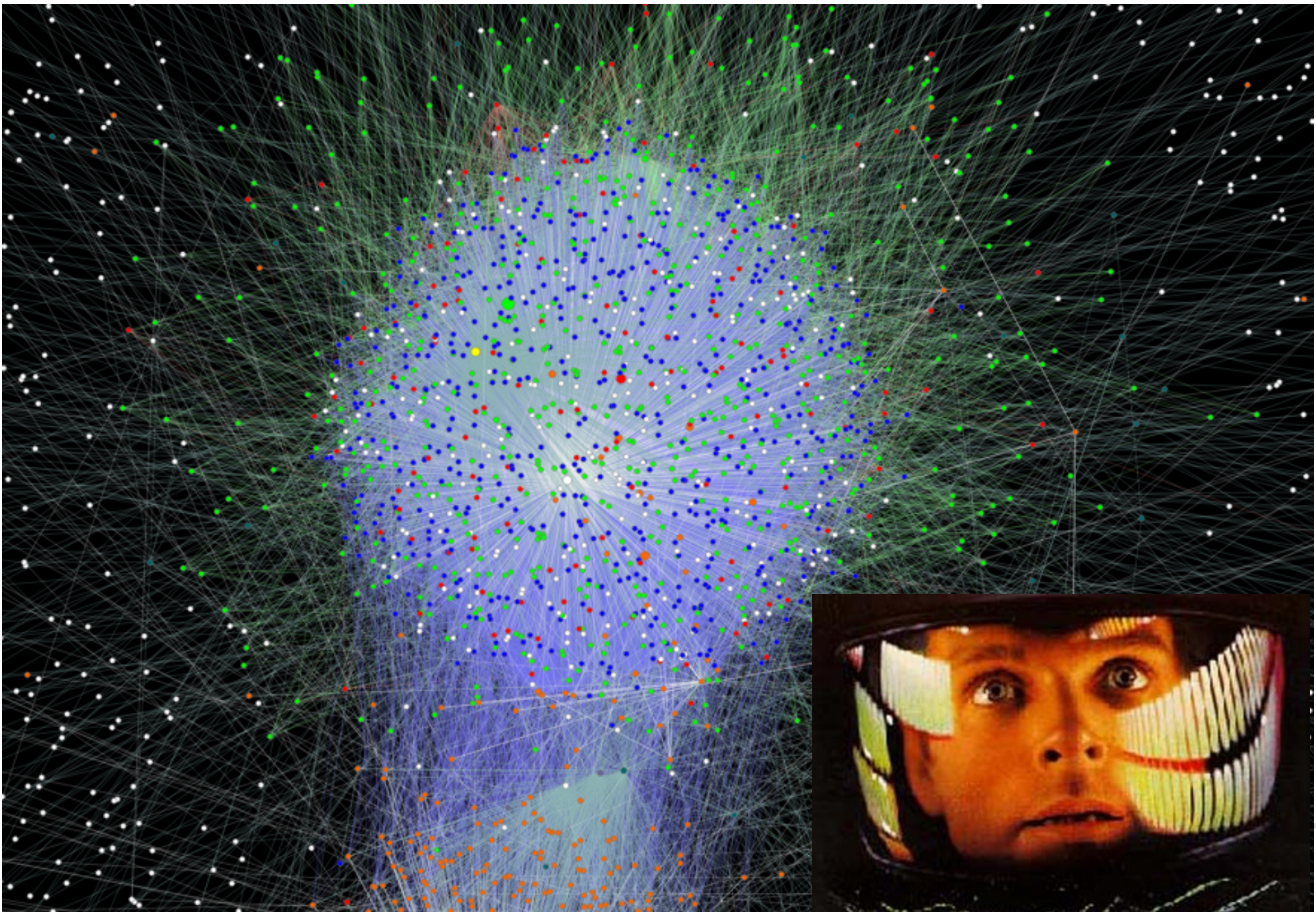
jsonLite



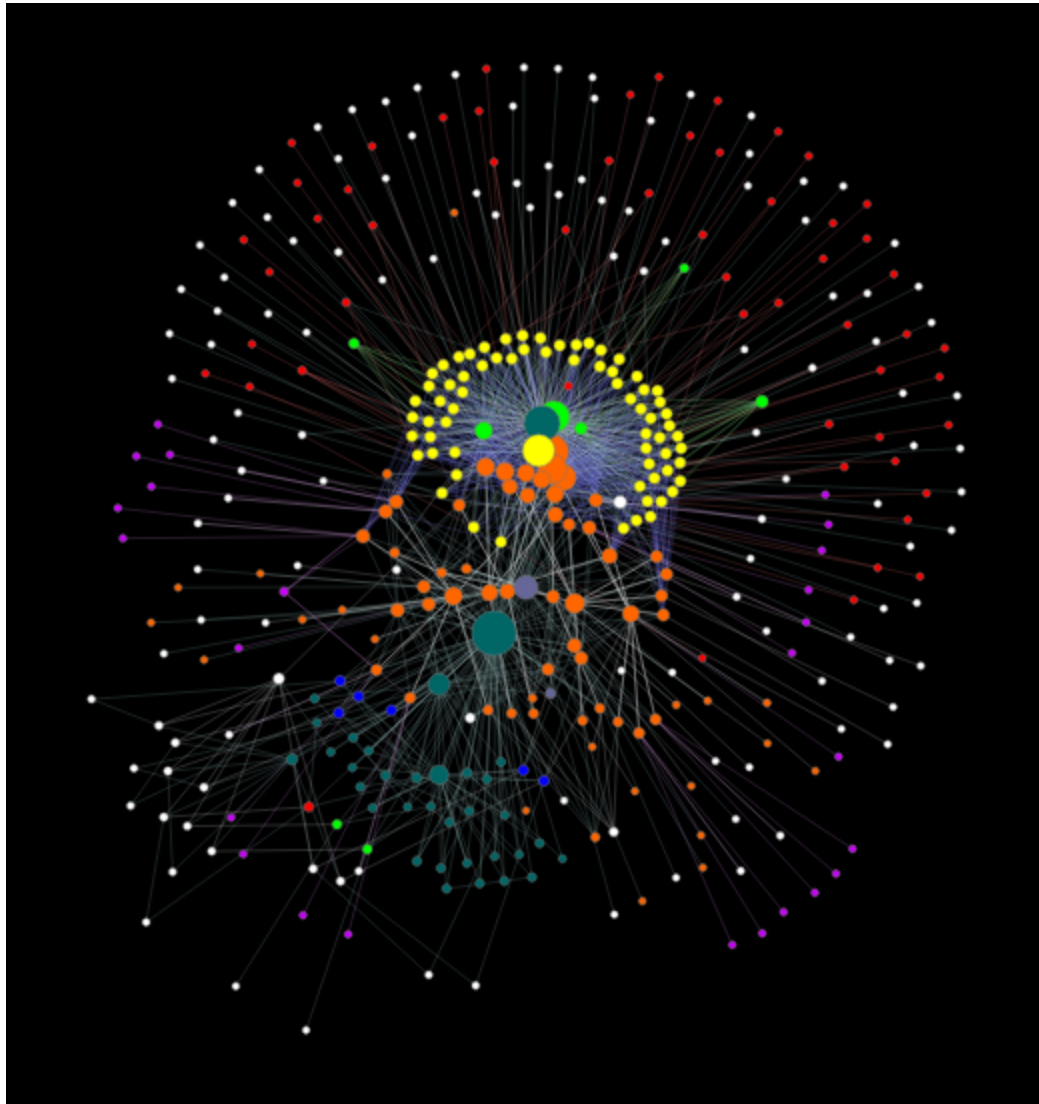
Result



Full AE RDF Data Cube as Force Network Graph



Live DEMO: DM Cube Subset as FN Graph



LIVE DEMO: Interactive DM Table

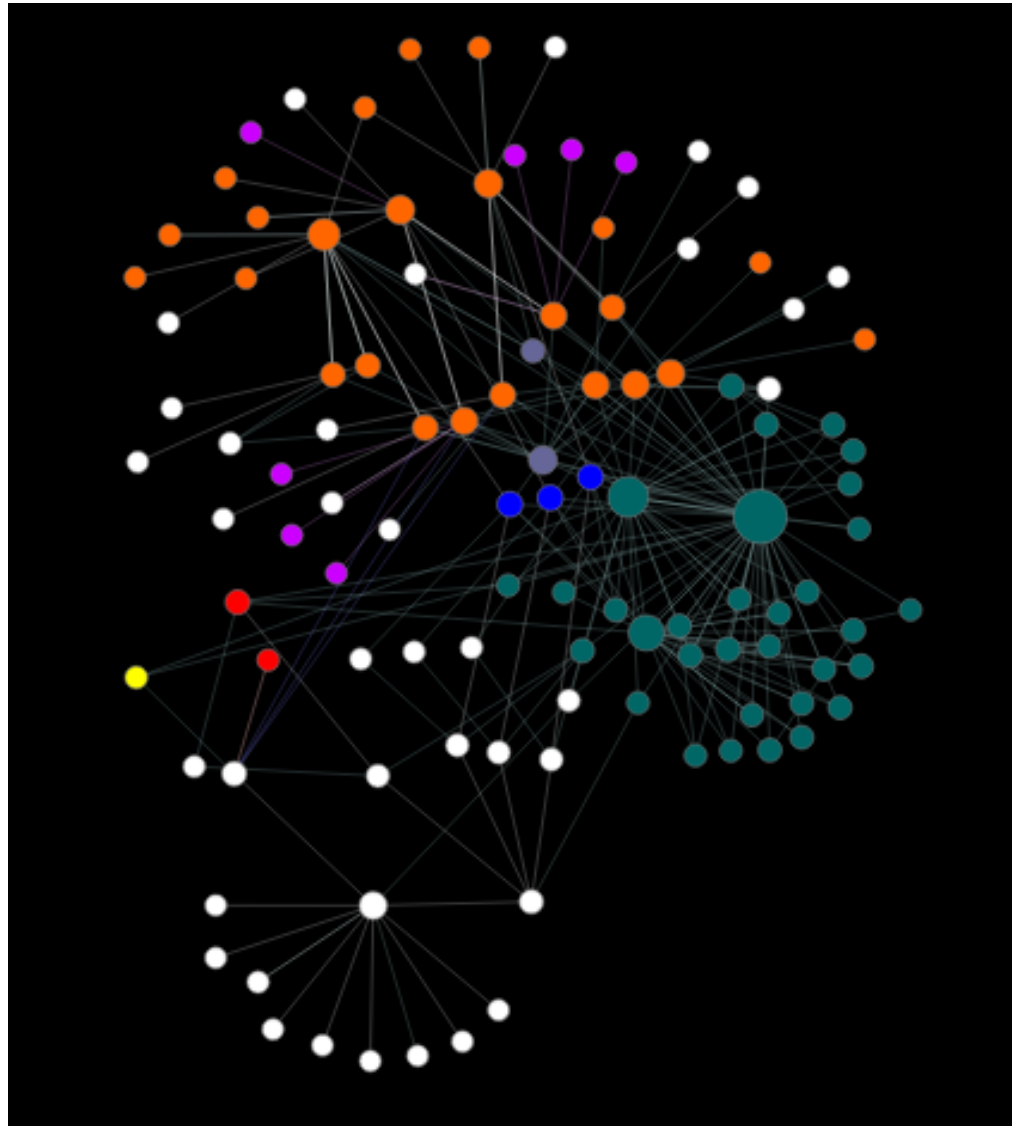
Demographics Summary Table

	Placebo	LowDose	HighDose
Baseline	N=28	N=30	N=29

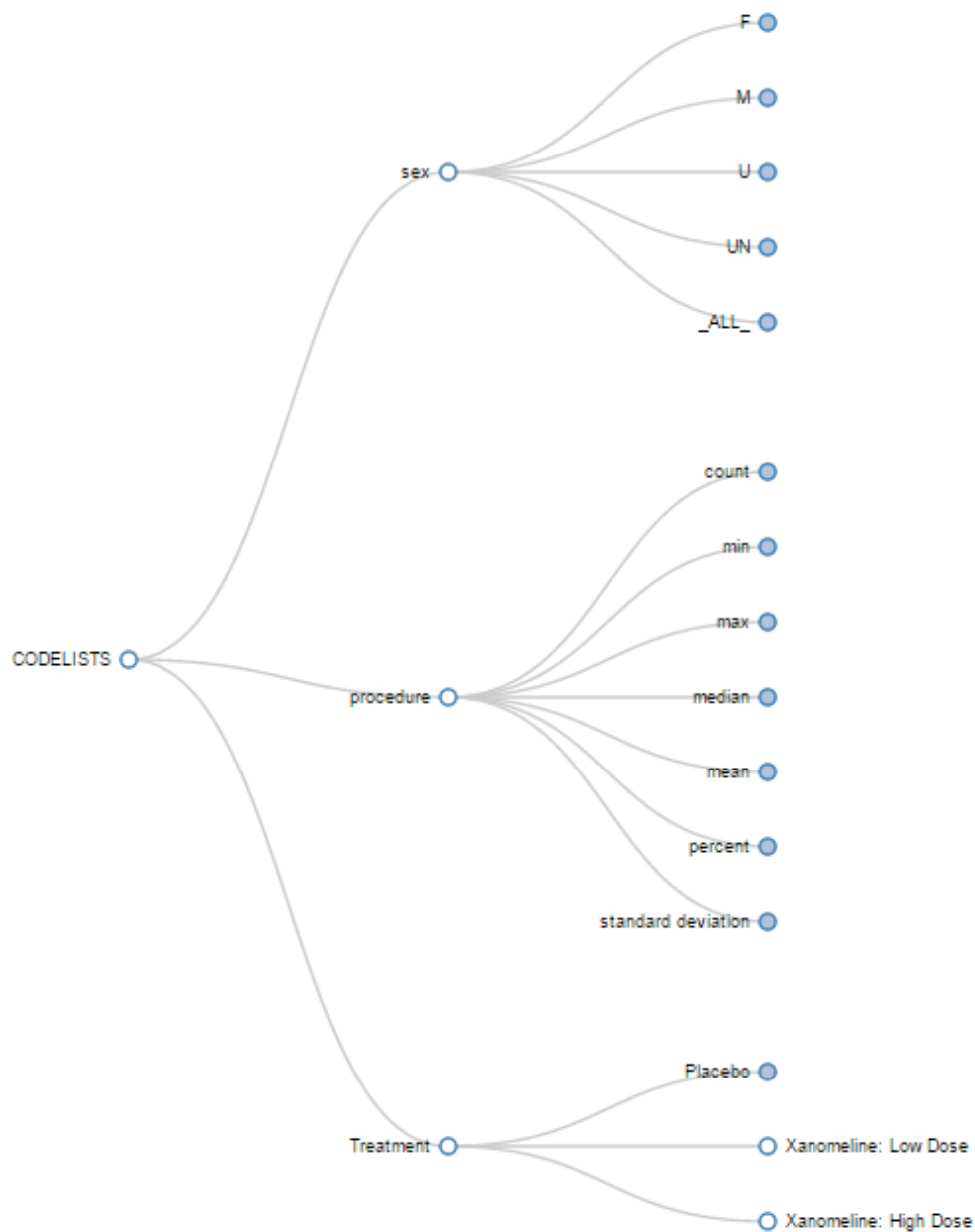
Sex			
F	12 (42.9)	14 (46.7)	16 (55.2)
M	16 (57.1)	16 (53.3)	13 (44.8)

Baseline	Placebo	LowDose	HighDose
	N=28	N=30	N=29
Sex			
F	12 (42.9)	14 (46.7)	16 (55.2)
M	16 (57.1)	16 (53.3)	13 (44.8)

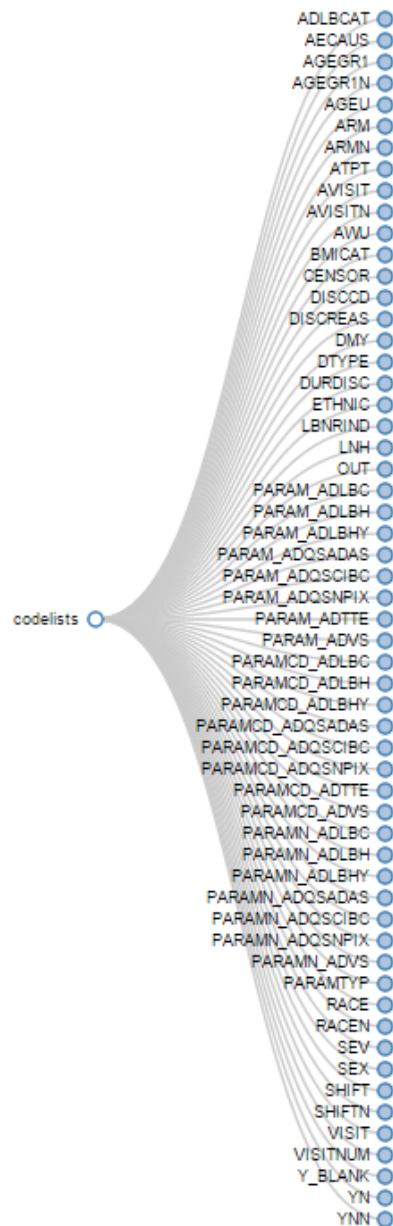
LIVE DEMO: Count Females in Placebo as FN graph

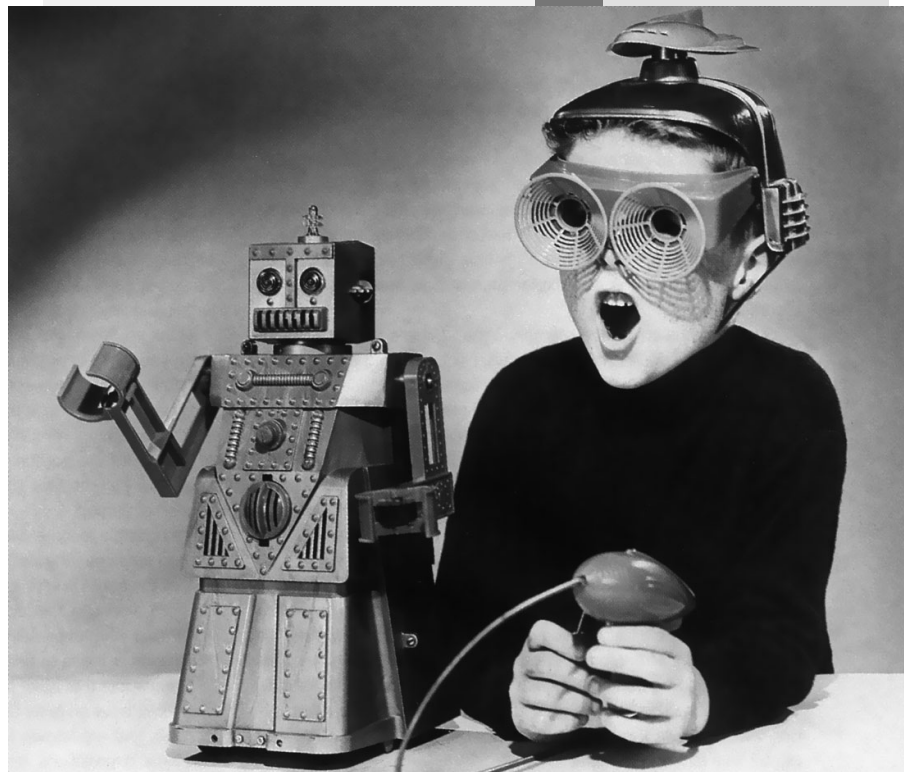


LIVE DEMO: Code List Subset



LIVE DEMO: Code List





FDA, PhUSE, and Industry

The Future is Now?

Janus Clinical Trials Repository (CTR)

“..can support other emerging study data exchange standards, such as HL7 v3, FHIR, and RDF.”

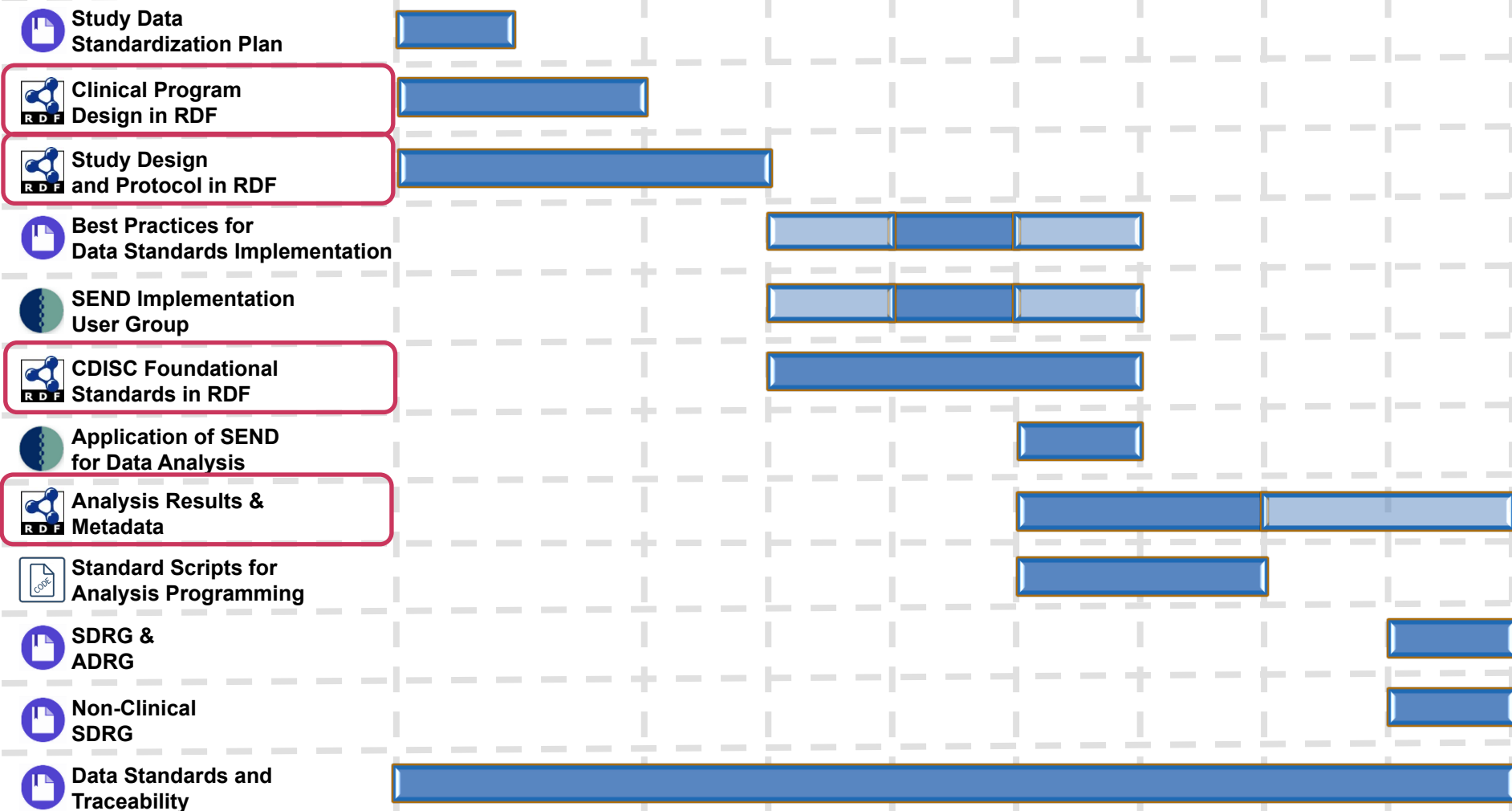
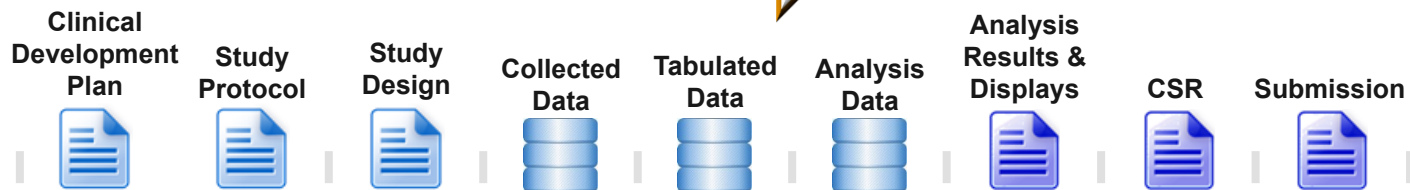
“FDA is exploring semantic web technologies for study information exchange. Further details to follow soon.” – April 2015



- **CDASH 1.1**
- **SDTM 1.2 + Impl. Guide 3.1.2**
- **SDTM 1.3 + Impl. Guide 3.1.3**
- **SEND Impl. Guide 3.0**
- **ADaM 2.1 + Impl. Guide 1.0**
- **Controlled Terminology**
- **CDISC e-share! (October 2014)**

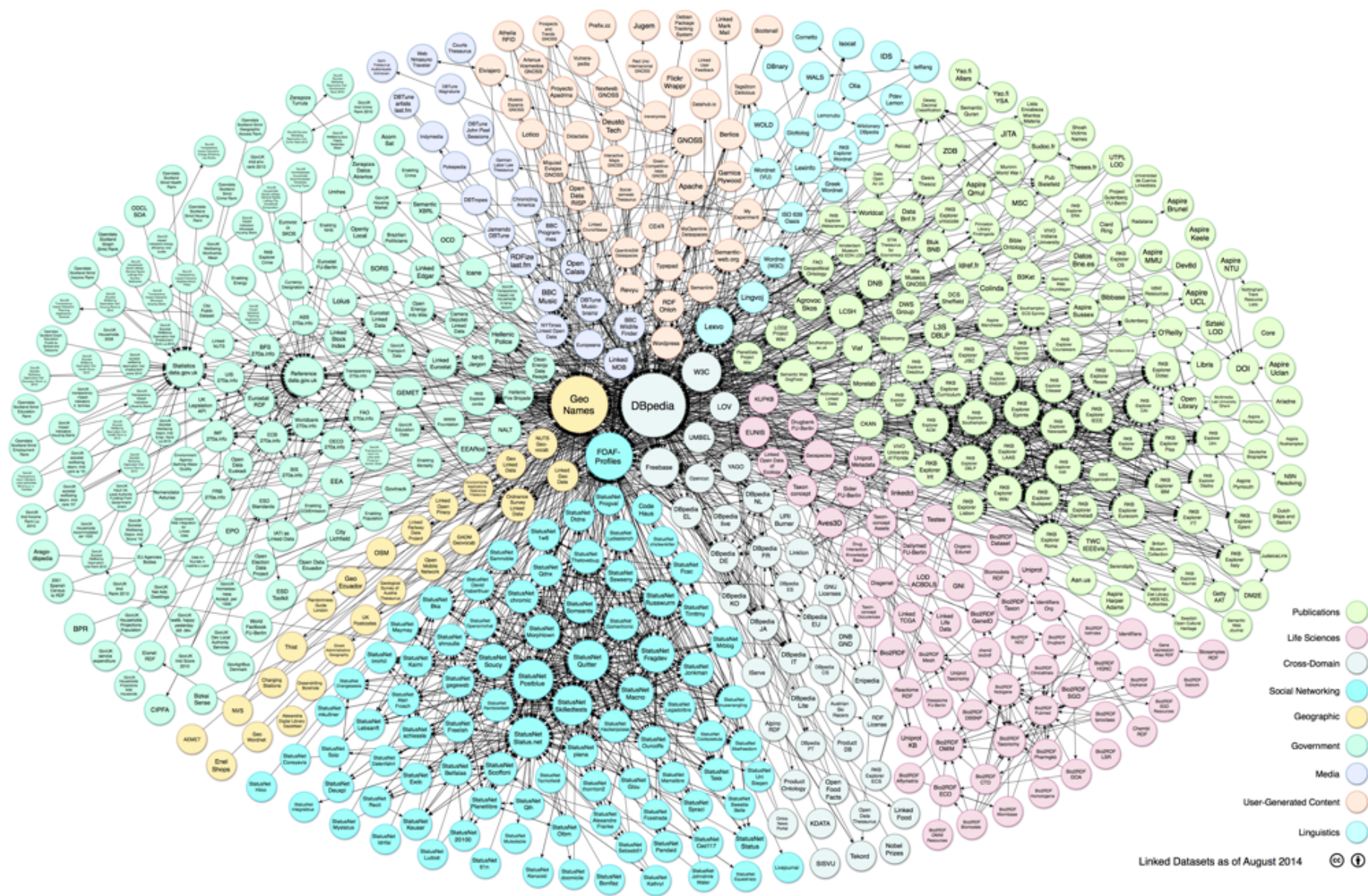


CS WG Projects & Deliverables



Linked Open Data Cloud, 2014

33



DBpedia

BIO2RDF PubMed

Medline, Life Sciences Citations

5,005,343,905 triples

BIO2RDF ClinicalTrials

ClinicalTrials.gov
Clinical Trials registry

8,323,598 triples

BIO2RDF DrugBank

Bioinformatics and chemoinformatics

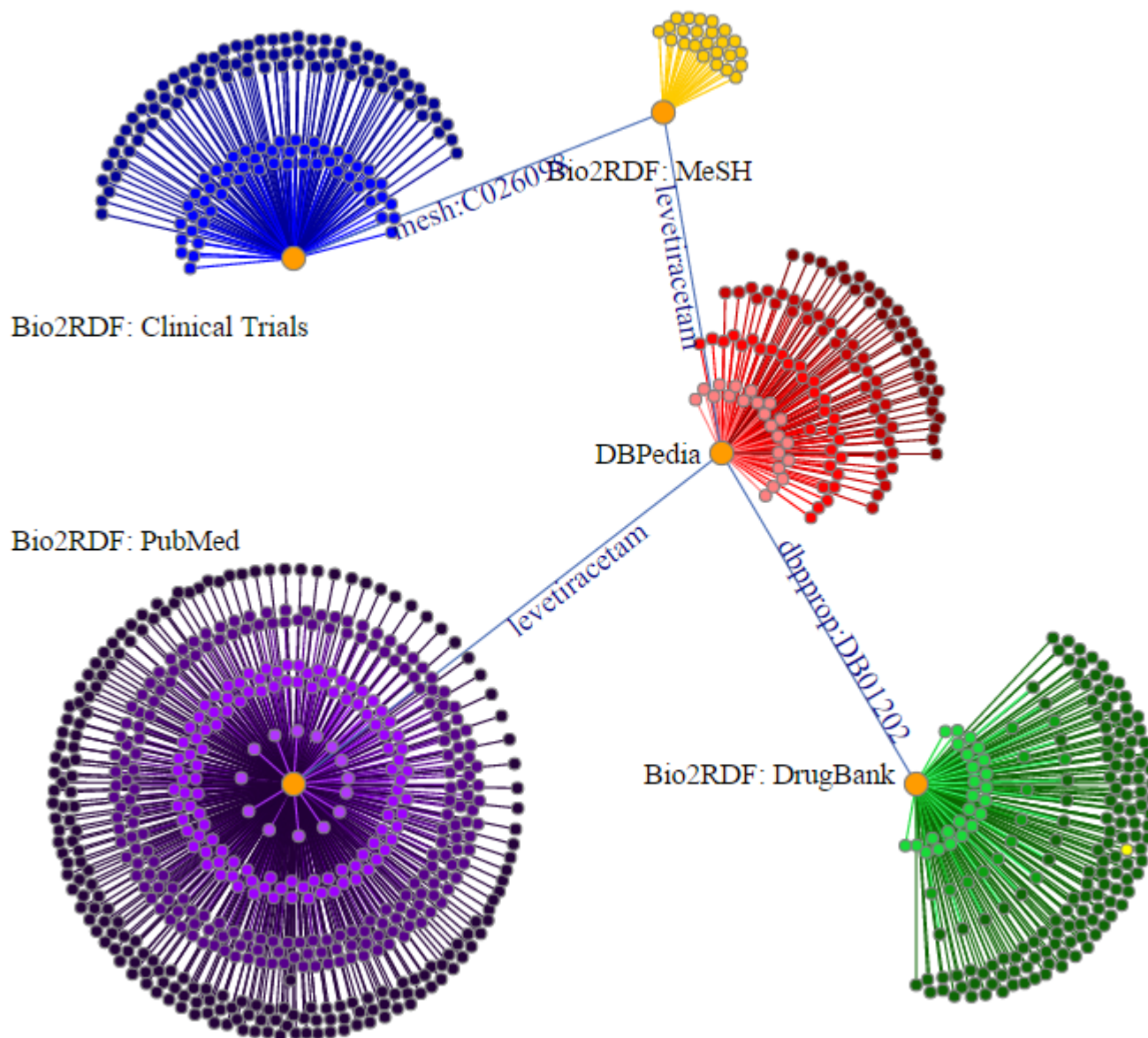
3,672,531 triples

BIO2RDF MeSH

Medical Subject Headings
7,323,864 triples

LIVE DEMO: Levetiracetam LOD Merge & Vis

Levetiracetam



LIVE DEMO: PubMed page for article



Facets (new se

About: <http://bio2rdf.org/pubmed:23416036> [Sponge](#) [Permalink](#)

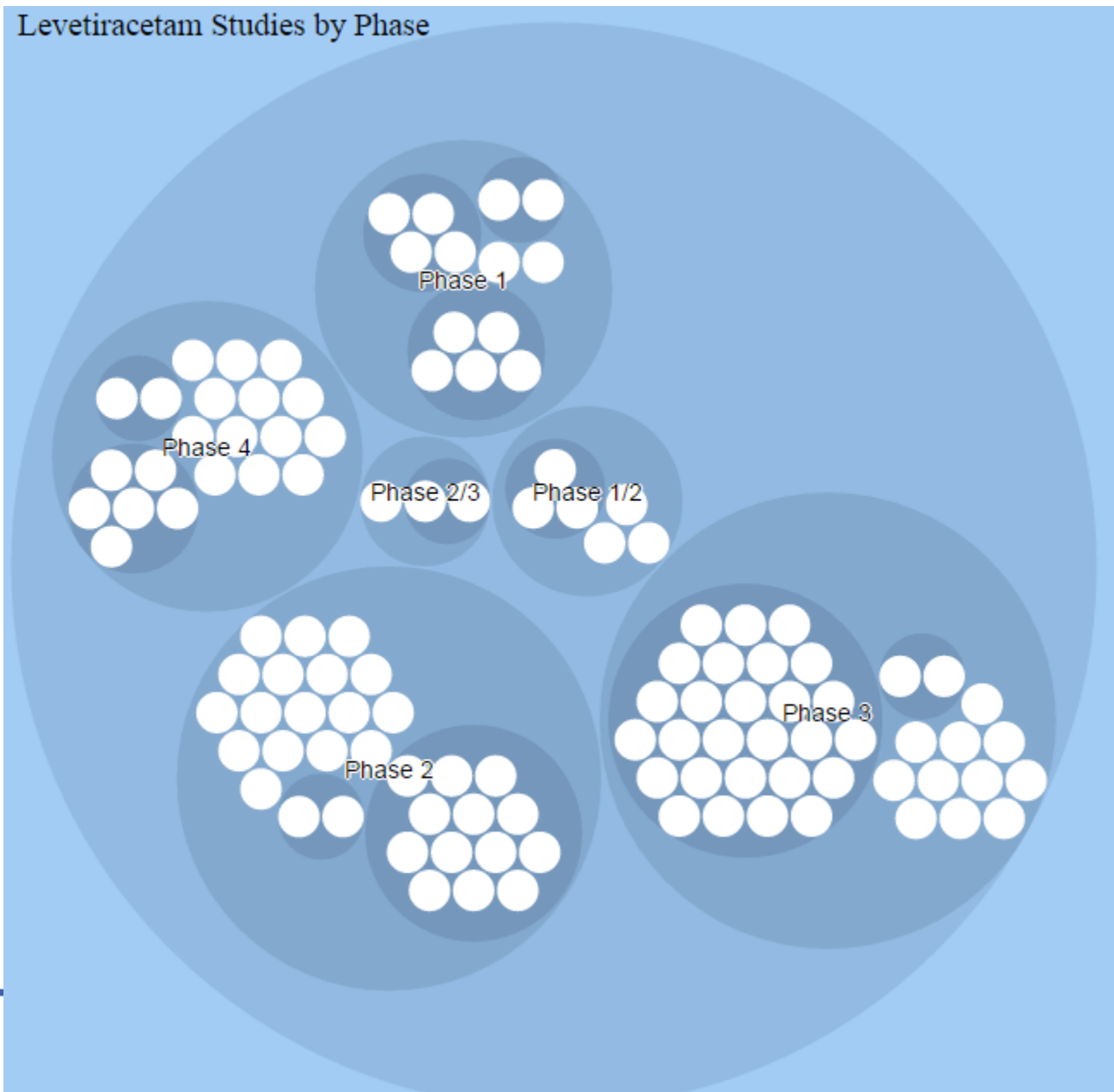
An Entity of Type : http://bio2rdf.org/pubmed_vocabulary:PubMedRecord, within Data Space : pubmed.bio2rdf.org associated with source [dataset\(s\)](#)

Type: http://bio2rdf.org/pubmed_vocabulary:PubMedRecord ▼ [New Facets Session with This Class](#)

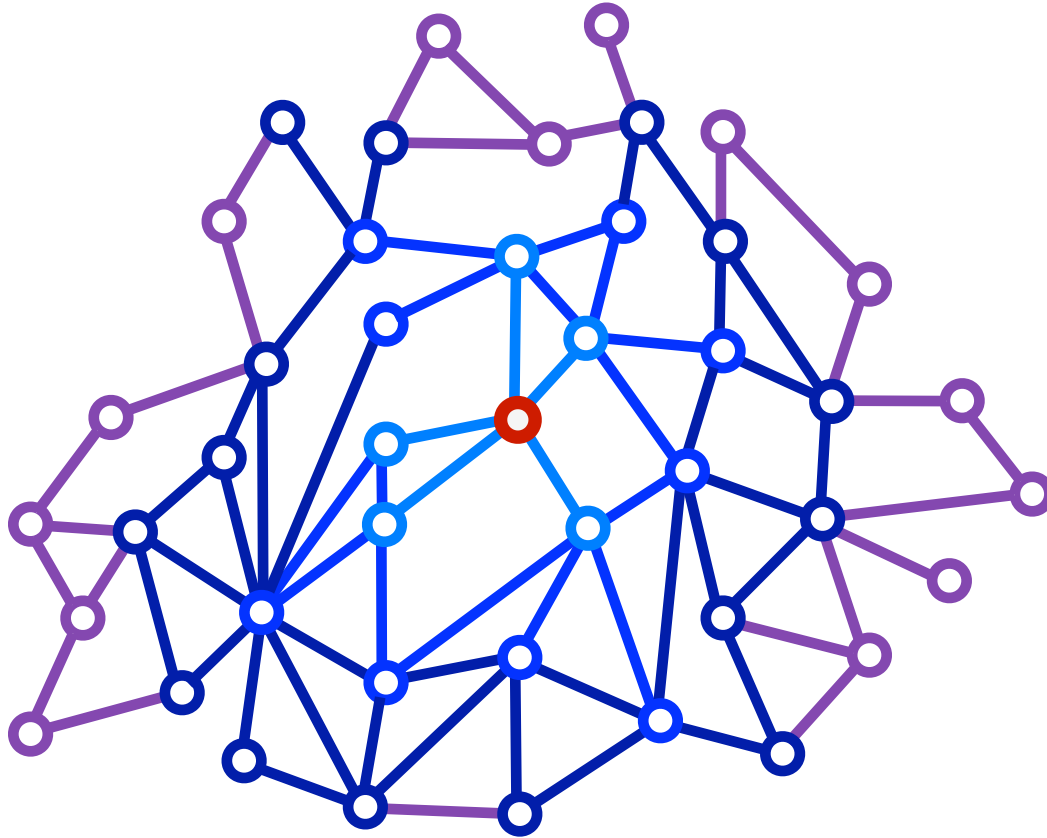
Attributes	Values
rdf:type	http://bio2rdf.org/pubmed_vocabulary:PubMedRecord
dcterms:created	2013-05-13(xsd:date)
rdfs:label	Effects of levetiracetam, an antiepileptic drug, on memory impairments associated with aging and Alzheimer's disease in mice.
http://bio2rdf.org...ry:completed_date	2013-12-23(xsd:date)
http://bio2rdf.org...vocabulary:owner	NLM
http://bio2rdf.org...vocabulary:status	MEDLINE In-Process
void:inDataset	http://bio2rdf.org/bio2rdf_dataset:bio2rdf-pubmed-20140523
dcterms:title	Effects of levetiracetam, an antiepileptic drug, on memory impairments associated with aging and Alzheimer's disease in mice.
http://bio2rdf.org...cabulary:chemical	http://bio2rdf.org/pubmed_resource:chemical:2c53fe606117663c3d9e9b280cd58fb0 http://bio2rdf.org/pubmed_resource:chemical:c8cb73fc47c02e84975161d0dafa8855 http://bio2rdf.org/pubmed_resource:chemical:3352605dc14d7f0e09469973fc77bc61 http://bio2rdf.org/pubmed_resource:chemical:9aaff71db9108c0d256c6af146e8ded2
http://bio2rdf.org...urnal_publication	http://bio2rdf.org/pubmed_resource:journal_publication:3c26f276d3c6cc83340ea2c1ee0ce30a
dcterms:contributor	http://bio2rdf.org/pubmed_resource:23416036_AUTHOR_1 http://bio2rdf.org/pubmed_resource:23416036_AUTHOR_2

LIVE DEMO: Clinical Trials by Phase

Levetiracetam Studies by Phase



The Power is not in the Visualization



LINK

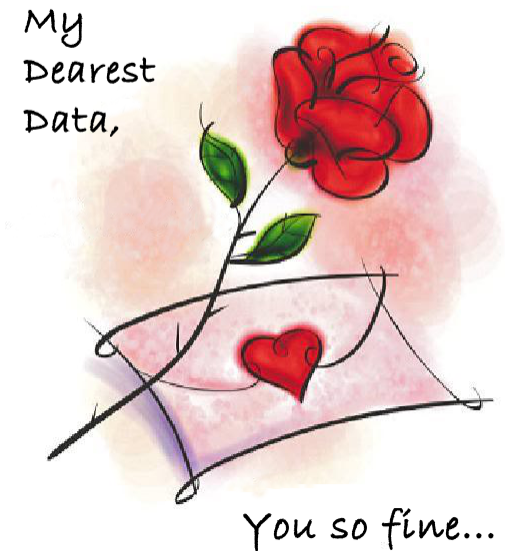


Conclusion: Why linked data?

- Individual observations and groups of observations become addressable.
- Flexibility in combining data both similar and diverse data sources (owl:sameas, etc.)
- Machine readable
- Enables reuse, standard tools and components
- Non-proprietary exchange format
- The context (metadata) comes *with* the data.

“Metadata is a love note to the people and machines after you.”

- Jason Scott via Twitter, 2011

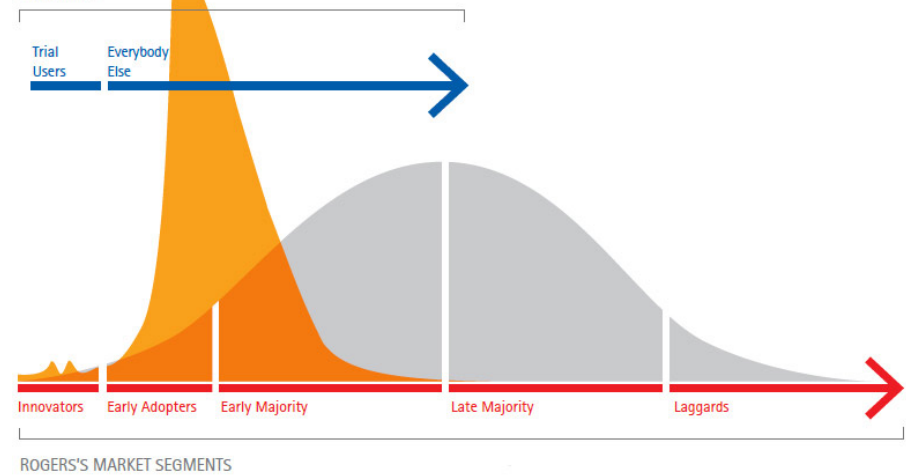


Challenges

- Resistance to change
- Vendors and consumers
- Learning Curve
- Software Tools and Applications
- Size?

Big Bang Market Adoption

BIG BANG MARKET SEGMENTS



Where to learn more?

PhUSE Wiki “Semantic Technology Curriculum”

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

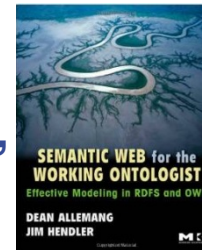
PhUSE Wiki “Semantic Technology Working Groups”

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

“Learning SPARQL” - Ducharme



“Semantic Web for the Working Ontologist” - Allemang & Hendler



Semantic University by Cambridge Semantics

<http://www.cambridgesemantics.com/semantic-university>

RDF Primer

<http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

Acknowledgements

- **PhUSE & PhUSE CSS**

- **PhUSE “Analysis Results & Metadata Working Group”**

- Leaders: Ian Fleming, Marc Andersen



- **Mike Bostock**

- Author of the D3.js library



Data-Driven Documents

- **Bio2RDF**

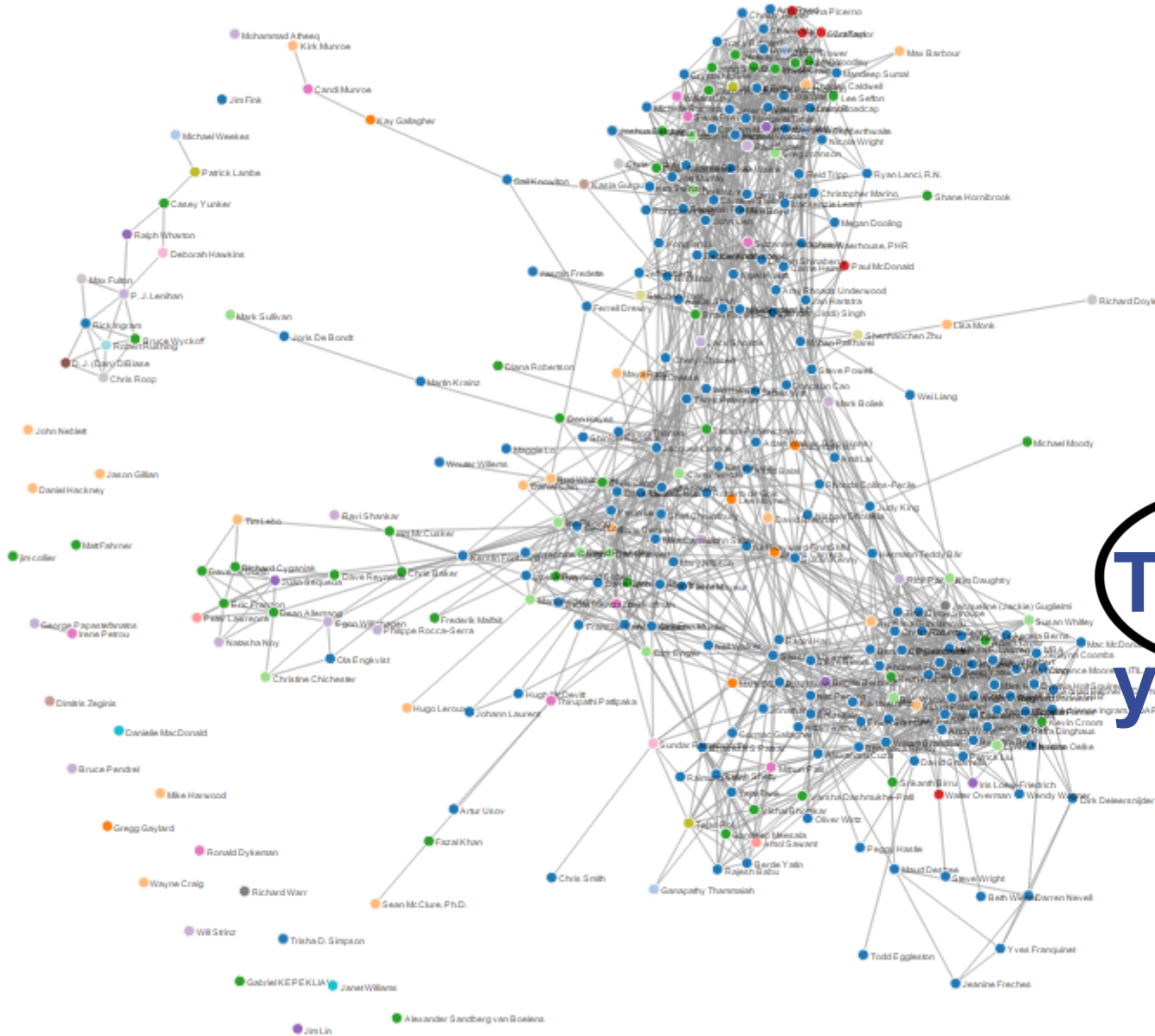


- **The R Project**



- **UCB Biosciences**





Thank
you!



@NovasTaylor

tim.williams@ucb.com