

INTRODUCTION TO LINKED DATA

BUILDING RELATIONSHIPS IN CLINICAL TRIALS DATA WHICH CURRENT STANDARDS DON'T
PROVIDE



WHO AM I?



Tim Williams

Principal Statistical Systems Analyst

UCB BioSciences

Raleigh, NC USA

PHUSE "STREET CRED"

- PhUSE CSS Steering Committee
- Semantic Technology Working Group - Co-lead
- Analysis Results & Metadata Project - Co-lead
- CSS Linked Data Hands-on Workshop - Instructor



LINKED DATA

- Semantic Web
- Resource Description Framework (RDF)
- Graph Data and Graph Analytics
- Property Graphs (Neo4j)



AWESOME THINGS ABOUT LINKED DATA

THAT I WILL NOT TALK ABOUT TODAY

- Integrated metadata & context
- Traceability across the data lifecycle
- Granular audit trail
- Visualization and drill down to data
- Common terminology
- Inferencing, Reasoning

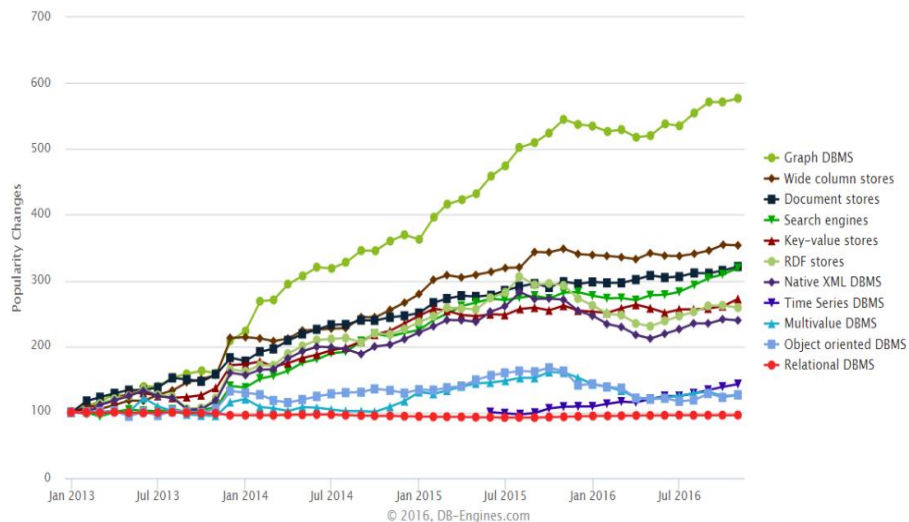


A screenshot of a Google search results page for the query "keppra". The top navigation bar includes the Google logo, the search term "keppra", and icons for web, shopping, news, images, books, more, and search tools. Below the navigation bar, it shows "About 520,000 results (0.20 seconds)". The first search result is from "Keppra USA, Dosage & Side Effects - Drugs.com" with the URL "www.drugs.com/keppra.html". The snippet states: "Keppra is a salt and extended-release drug used to treat partial-onset seizures in people with epilepsy. Learn about side effects, interactions and indications. Read Efficacy, Keppra Drug Information, Levitinacetam Dosage..." To the right of this result is a red-bordered box containing information about "Levetiracetam (Keppra)" from "Prescription Drug". It lists brand names (Keppra, Levetiracetam), frequency (twice a day), side effects (dizziness, headache, nausea, etc.), pregnancy risk (category C), possible side effects (somnolence, headache, dizziness, asthenia), drug class (antiepileptic agent), other drugs in same class (Lamotrigine, Phenytoin, Valproic acid), may treat (Epilepsy, Partial seizure), and a link to "People also search for" which lists Lamotrigine (Lamictal), Valproic Acid (Depakote), Phenytoin (Dilantin), Gabapentin (Neurontin), and Topiramate (Topamax). At the bottom of the screenshot, there are two additional search results: "PMN Keppra Medication Guide - Food and Drug Administration www.fda.gov/.../CDER/cder/drug/infoset/pmn_..." and "Levetiracetam (Keppra) | US FDA Approved Antiepileptic | Epilepsy.com".



DBMS POPULARITY CHANGES PER CATEGORY, NOV. 2016

Complete trend, starting with January 2013



TOP RDF AND GRAPH DATABASES, 2015



Source: "Graph and RDF databases 2015" - Bloor Market Update www.BloorResearch.com

DATA AS A GRAPH?

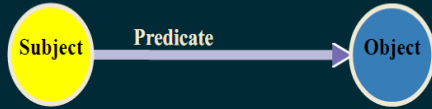


A Comparison

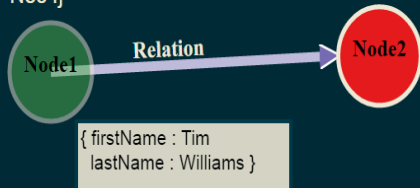


RDF, NEO4J COMPARISON

RDF Triple



Neo4j



CORE RDF CONCEPTS



- All **things** have a **name**
- All names are **unique** and **addressable**
 - HTTP URI
- Things are **linked**
 - Directed graphs
- Links have **meaning**
 - Semantics

RDF IN HEALTHCARE & PHARMA

- HL7 and FHIR
 - [RDF for Semantic Interoperability](#) group
- Foundational CDISC Standards and Terminology (SDTM, CDASH, ADaM, SEND)
 - [CDISC.org](#)
 - [cancer.gov](#)
- National Cancer Institute
 - [NCI Thesaurus](#)
- FDA
 - [Janus Clinical Trials Repository](#) can support* RDF



SDTM TERMINOLOGY AS RDF

HOORAY!....?

"Female" SDTM Terminology as RDF: [Local Graph](#)

```
SELECT * FROM <http://localhost:8890/CDISC>  
WHERE {  
  <http://rdf.cdisc.org/sdtm-terminology#C66731.C16576> ?p ?o  
}
```



SPARQL Execution


[Query](#) [Saved Queries](#)

 Default Graph IRI

Query

```
SELECT *
FROM <http://localhost:8890/CDISC>
WHERE { <http://rdf.cdisc.org/sdtm-terminology#C66731.C16576> ?p ?o
}
```

[Execute](#) [Save](#) [Load](#) [Clear](#)

p	o
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://rdf.cdisc.org/mms#PermissibleValue
http://rdf.cdisc.org/ct/schema#cdiscSubmissionValue	"F" http://www.w3.org/2001/XMLSchema#string
http://rdf.cdisc.org/ct/schema#cdiscDefinition	"A person who belongs to a sex that normally produces ova. The term is used to indicate biological sex distinctions, or cultural gender role distinctions, or both. (NCI)" http://www.w3.org/2001/XMLSchema#string
http://rdf.cdisc.org/ct/schema#cdiscSynonyms	"Female" http://www.w3.org/2001/XMLSchema#string
http://rdf.cdisc.org/mms#inValueDomain	http://rdf.cdisc.org/sdtm-terminology#C66731
http://rdf.cdisc.org/ct/schema#nciCode	"C16576" http://www.w3.org/2001/XMLSchema#string
http://rdf.cdisc.org/ct/schema#nciPreferredTerm	"Female" http://www.w3.org/2001/XMLSchema#string

TWITTER RANT



Tim Williams @NovasTaylor · 4h

If you are just straight-up converting a row x column data source to [#LinkedData](#), you are seriously missing the point.



2



2



CDISC

**YOU JUST
GOT
SERVED...**



BY LINKED DATA

- Think about data in a new way.
- Adopt a graph approach to both data and standards.

Take this row-by-column structure...

	A	B	C	D	E	O	P	Q	R	S	T	U
1		studyid	domain	usubjid	subjid	age	ageu	sex	race	ethnic	armcd	arm
2	1	CDISCPIL0T01	DM	01-701-10	1015	63 YEARS	F	WHITE	HISPANIC OR LATINO		Pbo	Placebo
3	2	CDISCPIL0T01	DM	01-701-10	1023	64 YEARS	M	WHITE	HISPANIC OR LATINO		Pbo	Placebo
4	3	CDISCPIL0T01	DM	01-701-10	1028	71 YEARS	M	WHITE	NOT HISPANIC OR LATINO		Xan_Hi	Xanomelir
5	4	CDISCPIL0T01	DM	01-701-10	1033	74 YEARS	M	WHITE	NOT HISPANIC OR LATINO		Xan_Lo	Xanomelir
6	5	CDISCPIL0T01	DM	01-701-10	1034	77 YEARS	F	WHITE	NOT HISPANIC OR LATINO		Xan_Hi	Xanomelir
7	6	CDISCPIL0T01	DM	01-701-10	1047	85 YEARS	F	WHITE	NOT HISPANIC OR LATINO		Pbo	Placebo

... and transform it into multidimensional graph data.

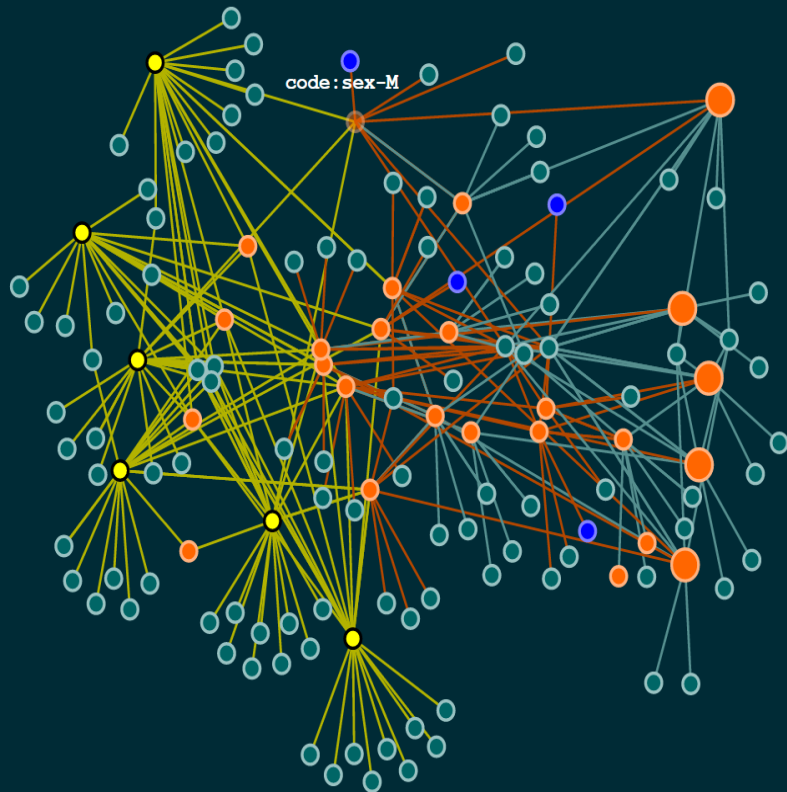


SDTM DM FOR CDISCPIL0T01

SUBSET OF SIX SUBJECTS

Legend

- Human Study Subject
- CONSC Code
- Literal Value
- Coded Value



UNDERLYING RDF

```
code:site-701 a          study:StudySite , code:Site , skos:Concept ;
    study:hasSiteId      "701" ;
    study:locatedIn      code:country-USA ;
    skos:inScheme         code:site ;
    skos:topConceptOf    code:site .

pers:pers2 a             sdtm:HumanStudySubject ;
    sdtm:hasSUBJID        pers:1023 ;
    sdtm:hasUSUBJID       pers:01-701-1023 ;
    sdtm:attendsSite      code:site-701 ;
    sdtm:hasAGE           "64"^^xsd:int ;
    sdtm:hasAGEU          "YEARS"^^xsd:string ;
```



ROUND-TRIPPING BACK TO YOUR ROW-BY-COLUMN "COMFORT ZONE"

```
PREFIX sdm: <http://www.example.org/CompoundX/Study1/SDTM/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core/>
SELECT ?USUBJID ?SUBJID ?AGE ?AGEU ?SEX
FROM <http://localhost:8890/SDTMTORDF>
WHERE {
  ?person a          sdm:HumanStudySubject ;
          sdm:hasUSUBJID ?USUBJID ;
          sdm:hasSUBJID ?SUBJID ;
          sdm:hasAGE ?AGE ;
          sdm:hasAGEU ?ageCode;
          sdm:hasSEX ?sexCode.
  ?ageCode skos:prefLabel ?AGEU .
  ?sexCode skos:prefLabel ?SEX .
}
```





- Interactive SQL (ISQL)
- WebDAV Browser
- Virtuoso Start Menu

- Documentation (web)
- Tutorials (web)

- Virtuoso Web Site
- OpenLink Software

Version: 07.20.3215
Build: Dec 10 2015

Home System Admin Database Replication Web Application Server XML Web Services Linked Data NNTF

SPARQL Sponger Statistics Graphs Schemas Namespaces Views Quad Store Upload

SPARQL Execution

Query Saved Queries

Default Graph IRI

Query

```

PREFIX sdtm: <http://www.example.org/CompoundX/Study1/SDTM/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core/>
SELECT ?USUBJID ?SUBJID ?AGE ?AGEU ?SEX
FROM <http://localhost:8890/SDTMTORDF>
WHERE {
  ?person a sdtm:HumanStudySubject ;
    sdtm:hasUSUBJID ?USUBJID ;
    sdtm:hasSUBJID ?SUBJID ;
    sdtm:hasAGE ?AGE ;
    sdtm:hasAGEU ?ageCode ;
    sdtm:hasSEX ?sexCode .
  ?ageCode skos:prefLabel ?AGEU .
  ?sexCode skos:prefLabel ?SEX .
}
```

Execute Save Load Clear

USUBJID	SUBJID	AGE	AGEU	SEX
http://www.example.org/person/01-701-1015	http://www.example.org/person/1015	"63"^^<http://www.w3.org/2001/XMLSchema#int>	"YEARS"	"F"
http://www.example.org/person/01-701-1023	http://www.example.org/person/1023	"64"^^<http://www.w3.org/2001/XMLSchema#int>	"YEARS"	"M"
http://www.example.org/person/01-701-1028	http://www.example.org/person/1028	"71"^^<http://www.w3.org/2001/XMLSchema#int>	"YEARS"	"M"
http://www.example.org/person/01-701-1033	http://www.example.org/person/1033	"74"^^<http://www.w3.org/2001/XMLSchema#int>	"YEARS"	"M"
http://www.example.org/person/01-701-1034	http://www.example.org/person/1034	"77"^^<http://www.w3.org/2001/XMLSchema#int>	"YEARS"	"F"
http://www.example.org/person/01-701-1047	http://www.example.org/person/1047	"85"^^<http://www.w3.org/2001/XMLSchema#int>	"YEARS"	"F"

MAKE IT "CDISC PRETTY"

```
PREFIX sdtm: <http://www.example.org/CompoundX/Study1/SDTM/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core/>

SELECT ?STUDY ?DOMAIN ?USUBJID ?SUBJID ?AGE ?AGEU ?SEX
FROM <http://localhost:8890/SDTMTORDF>
WHERE {
    ?person a          sdtm:HumanStudySubject ;
            sdtm:hasUSUBJID ?usubjid_ ;
            sdtm:hasSUBJID  ?subjid_ ;
            sdtm:hasAGE     ?age_ ;
            sdtm:hasAGEU    ?ageCode;
            sdtm:hasSEX     ?sexCode.
    ?ageCode skos:prefLabel ?ageu_ .
    ?sexCode skos:prefLabel ?sex_ .

    # Make so pretty-pretty
    bind( strafter(str(?usubjid_),str('person/')) as ?USUBJID )
    bind( strafter(str(?subjid_),str('person/')) as ?SUBJID )
    bind( xsd:integer(?age_) as ?AGE )
    bind( str(?sex_) as ?SEX )
    bind( str(?ageu_) as ?AGEU )

    # Cheating here. Pay no attention to that man behind the curtain
    bind("DM" as ?DOMAIN)
    bind("CDISCPIL01" as ?STUDY)
```



- Interactive SQL (ISQL)
- WebDAV Browser
- Virtuoso Start Menu

- Documentation (web)
- Tutorials (web)

- Virtuoso Web Site
- OpenLink Software

Version: 07.20.3215
Build: Dec 10 2015

- Home
- System Admin
- Database
- Replication
- Web Application Server
- XML
- Web Services
- Linked Data
- NNTP

- SPARQL
- Sponger
- Statistics
- Graphs
- Schemas
- Namespaces
- Views
- Quad Store Upload

SPARQL Execution

Query Saved Queries

Default Graph IRI

Query

```

sdm:hasAGE ?age_ ;
sdm:hasAGEU ?ageCode;
sdm:hasSEX ?sexCode.

?ageCode skos:prefLabel ?ageu .
?sexCode skos:prefLabel ?sex_ .

# Make so pretty-pretty
bind( strafter(str(?usubjid),str('person/')) as ?USUBJID )
bind( strafter(str(?subjid),str('person/')) as ?SUBJID )
bind( xsd:integer(?age_) as ?AGE )
bind( str(?sex_) as ?SEX )
bind( str(?ageu) as ?AGEU )
# Cheating here. Pay no attention to that man behind the curtain
bind("DM" as ?DOMAIN)
bind("CDISCPILLOT01" as ?STUDY)
}

```

Execute

Save

Load

Clear

STUDY	DOMAIN	USUBJID	SUBJID	AGE	AGEU	SEX
CDISCPILLOT01	DM	01-701-1015	1015	63	YEARS	F
CDISCPILLOT01	DM	01-701-1023	1023	64	YEARS	M
CDISCPILLOT01	DM	01-701-1028	1028	71	YEARS	M
CDISCPILLOT01	DM	01-701-1033	1033	74	YEARS	M
CDISCPILLOT01	DM	01-701-1034	1034	77	YEARS	F
CDISCPILLOT01	DM	01-701-1047	1047	85	YEARS	F

BUST YOUR DATA OUT OF STANDARDS PRISON

Smash the CDISC models (including existing RDF)

- Harvest the useful bits
- Make a true Graph Model
- Create SDTM as Graph Data

RESULT

- Data values become standards-independent.
- The Standard becomes just another **data layer**
- Metadata, audit data, terminology, DEFINE - are Linked or integral



SEMANTIC TECHNOLOGY WORKING GROUP

PROJECTS

- Clinical Program Design
- Regulations to RDF (Reg2RDF) [FDA in-house]
- Analysis Results and Metadata [completed]

AT THE MARCH CSS

- Hands-on Workshop (RDF, Neo4j)
- Hack the Transport format! XPT versus XML, RDF, Property Graph...



SEMANTIC TECHNOLOGY WORKING GROUP

Seeking new projects!

NEW PROJECT [*proposed*]: SDTM Data as RDF

- SDTM data as multidimensional Linked Data
- Integration with a minimal Study Ontology prototype
- Use of SDTM and BRIDG terminology
- Integrated DEFINE data and metadata
 - DEFINE becomes a SPARQL query



CONCLUSION: GRAPH IT (LIKE IT'S HOT)



- Infinitely extensible multidimensional data model.
- Future-proof data and standards.
- Facilitate data integration.
- Integrated metadata.
- End-to-End Traceability.
- Data. Once.



Thank you!



RESOURCES

Contact

tim.williams@phuse.eu

@NovasTaylor

<http://www.linkedin.com/in/timpwilliams>

PhUSE Semantic Technology Wiki

<http://www.phusewiki.org/>

Search: "Semantic Technology"

Neo4j

<http://neo4j.com>

Virtuoso

<http://http://virtuoso.openlinksw.com>

RDF Quad Store



IMAGE ATTRIBUTIONS

Images not listed below are attributed to the author.



Google, Facebook, Microsoft, Twitter



db-engines.com



"Graph and RDF databases 2015" - Bloor Market Update www.BloorResearch.com



Thing1 and Thing2 - Dr. Seuss



"The Complete Far Side" - Gary Larson



"South Park" - Comedy Central



"Drop it like It's Hot", R&G (Rhythm & Gangsta), Snoop Dogg

