

Linked Data

What it is and the potential use in Pharmaceutical Programming

Marc Andersen, `mja@statgroup.dk`

PhUSE Copenhagen, Denmark 2014 SDE

Topics

- ▶ What it is:
 - ▶ Linked data: DbPedia, DrugBank, LinkedCT
 - ▶ Resource Descriptor Format (RDF,
<http://www.w3.org/RDF/>)
 - ▶ SPARQL Query Language for RDF
(<http://www.w3.org/TR/rdf-sparql-query/>)
- ▶ Potential Use:
 - ▶ Metadata (and data)
 - ▶ Analysis Results Metadata - PhUSE working group

Linked data concepts

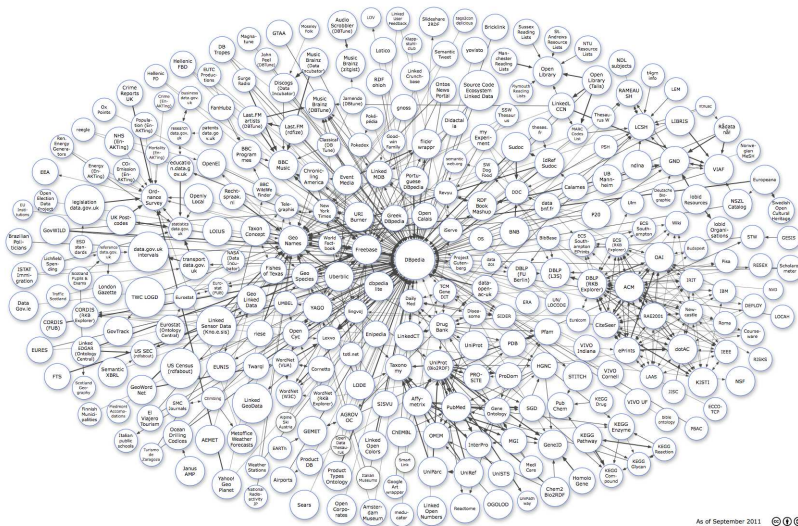
- ▶ AAA Principle “Anyone can say Anything about Any topic.”

<https://www.elsevier.com/books/semantic-web-for-the-working-ontologist/allemang/978-0-12-385965-5>

- ▶ Knowledge management
 - ▶ Open World Assumption
 - ▶ Closed World Assumption
- ▶ Ontologies
 - ▶ top-down
 - ▶ bottom up
- ▶ Standards
- ▶ URI minting

http://www.w3.org/2011/gld/wiki/223_Best_Practices_URI_Construction

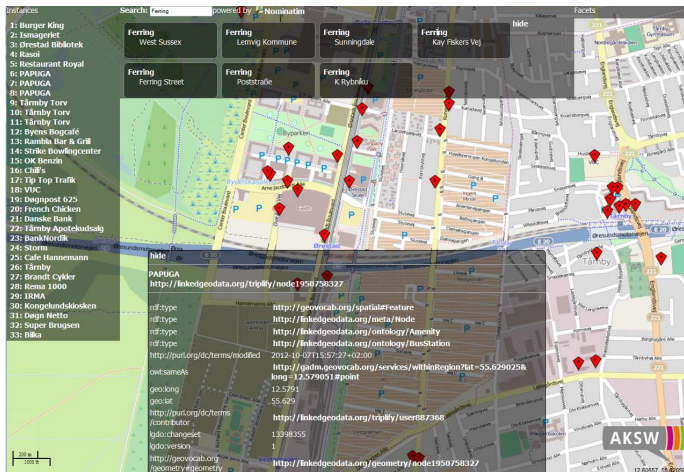
Linking Open Data cloud diagram, 2011



As of September 2011

Linking Open Data cloud diagram, 2011, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

Linked Geographical data



<http://browser.linkedgedata.org/> - enter Ferring in search box, select Ferring Kay fiskers vej

Sparql Query

SPARQL query using dbpedia

<http://dbpedia-live.openlinksw.com/sparql>

```
select * where {  
<http://dbpedia.org/resource/Ferring_Pharmaceuticals> ?p ?o.  
}
```

p	o
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://schema.org/Organization
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://dbpedia.org/class/yago/BiotechnologyCompanies
http://dbpedia.org/ontology/foundingYear	*1950+02:00 http://www.w3.org/2001/XMLSchema#gYear
http://dbpedia.org/ontology/numberOfEmployees	4500
http://www.w3.org/2002/07/owl#sameAs	http://rdf.freebase.com/ns/m.09mb0r
http://dbpedia.org/ontology/abstract	"Ferring Pharmaceuticals is a multinational pharmaceutical company ... In 2012, Ferring Pharmaceuticals funded the Ferring Research Infertility and Gynaecology Grant (FRIGGA)." en

SPARQL result (abbreviated)

LinkedCT - linked data version of ClinicalTrials.gov

<http://data.linkedct.org/about/>

Query <http://static.linkedct.org/snorql/>

```
PREFIX linkedct: <http://static.linkedct.org/resource/linkedct/>
SELECT ?id ?btitle ?acronym WHERE {
?s linkedct:id ?id.
?s linkedct:lead_sponsor_agency "Ferring Pharmaceuticals".
?s linkedct:brief_title ?btitle .
?s linkedct:acronym ?acronym. }
```

id	btitle	acronym
NCT00209261	A 6-Week Open Label Cross-Over Study With 2 Different Daily Doses of Minirin [®] Oral Lyophilisate in Children and Adolescents With Primary Nocturnal Enuresis (PNE)	PALAT
NCT00230594	Desmopressin Response in the Young	DRY
NCT00245479	A Study of Oral Desmopressin in Previously Untreated Children Aged 5 to 15 Years With Primary Nocturnal Enuresis	DRIP
NCT00451958	A Study Evaluating a One-Month Dosing Regimen of Degarelix in Prostate Cancer Requiring Androgen Ablation Therapy	ICHGCP
NCT00587327	Effect of Oxytocin and Vasopressin Antagonists on Uterine Contractions	OVANCON
NCT00603733	Canadian Active & Maintenance Modified Pentasa Study	CAMP
NCT00862121	A Study With Pentasa in Patients With Active Crohn's Disease	PEACE
NCT00884221	MENOPUR in GnRH Antagonist Cycles With Single Embryo Transfer	MEGASET
NCT00930319	Effectiveness and Safety of Firmagon	FAST

SPARQL result

SPARQL query <http://drugbank.bio2rdf.org/sparql>

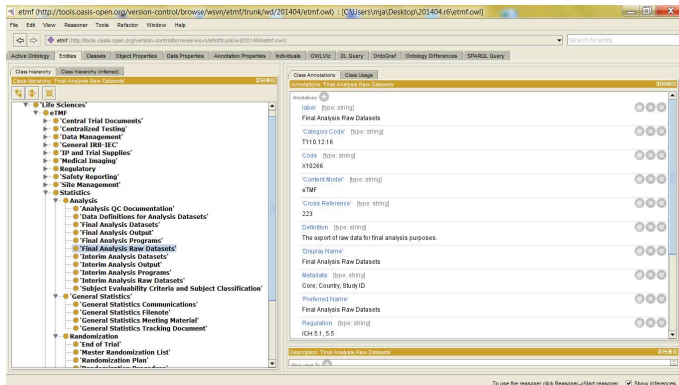
```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX db: <http://bio2rdf.org/drugbank_vocabulary:>
SELECT ?drug_name ?packager_name ?dosage_name
WHERE {
  ?drug a db:Drug .
  ?drug rdfs:label ?drug_name .
  ?drug db:packager ?packager .
  ?packager rdfs:label ?packager_name .
  OPTIONAL { ?drug db:dosage ?dosage . }
  OPTIONAL {?dosage dcterms:description ?dosage_name .}
  FILTER(regex(str(?packager_name), "ferring", "i"))
} LIMIT 5
```

drug_name	packager_name	dosage_name
"Desmopressin"@en	"Ferring Pharmaceuticals Inc."@en	"Spray by Nasal"@en
"Menotropins"@en	"Ferring Pharmaceuticals Inc."@en	"Powder, for solution by Subcutaneous"@en
"Chorionadotropin alfa"@en	"Ferring Pharmaceuticals Inc."@en	"Injection, solution by Subcutaneous"@en
"Desmopressin"@en	"Ferring Pharmaceuticals Inc."@en	"Spray, metered by Nasal"@en
"Desmopressin"@en	"Ferring Pharmaceuticals Inc."@en	"Liquid by Parenteral"@en

SPARQL result (edited). For more <http://www.cambridgesemantics.com/semantic-university/sparql-by-example>

OASIS Electronic Trial Master File (eTMF)

Ontology viewed in Protege



<https://tools.oasis-open.org/version-control/browse/wsvn/etmf/trunk/wd/201404/etmf.owl>

eTMF <https://www.oasis-open.org/committees/etmf/>

Linked Data

Marc Andersen, mja@statgroup.dk

PhUSE Copenhagen, Denmark 2014 SDE

CDISC as RDF

<https://github.com/phuse-org/rdf.cdisc.org>

“The FDA/PhUSE Semantic Technology project investigates how formal semantic standards can support the clinical and non-clinical trial data life cycle from protocol to submission.”

...

“Today, CDISC publishes these standards in a paper based format and partly in Excel, which makes it difficult to consistently represent and process this information. The RDF representation addresses both issues by providing at the same time a formal model, a machine readable representation, and an exchange format.”

Is your Linked Open Data 5 Star?

Tim Berners-Lee, 2010

*	Available on the web (whatever format) but with an open licence, to be Open Data
**	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
***	as (2) plus non-proprietary format (e.g. CSV instead of excel)
****	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
*****	All the above, plus: Link your data to other people's data to provide context

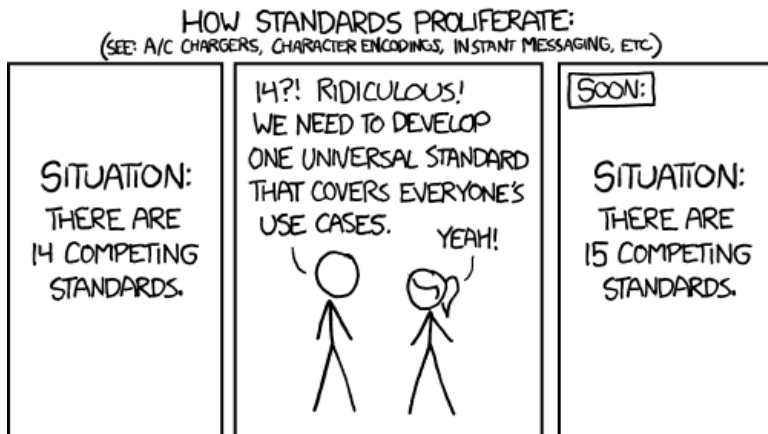
<http://www.w3.org/DesignIssues/LinkedData.html>

PhUSE Analysis Results Metadata work group

<http://www.phusewiki.org/wiki/index.php?t>

- ▶ Scope
 - ▶ Focus on Analysis Results definition
 - ▶ Aware of: Getting the analysis data
 - ▶ Aware of: Presenting the analysis data
- ▶ Approach
 - ▶ Minimalistic
 - ▶ Understand
 - ▶ Identify existing standards, ontologies etc
 - ▶ Proof of concept, simple tools
 - ▶ Discuss
 - ▶ Present
 - ▶ Identify key concepts: describe / prescribe
- ▶ Practically
 - ▶ TC every two weeks (Post presentation: Corrected from bi-weekly)

Standards



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

Analysis Results Metadata:

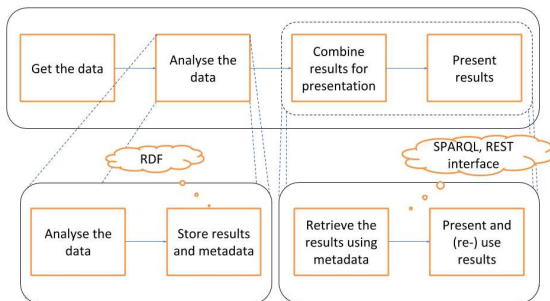
A meaningful 7-# classification similar to TBL five-star?

#	Analysis Results available in electronic format (scan, PDF, word)
##	Analysis Results available as datasets (SAS, R, relational database, excel, etc)
###	Analysis Results available in non-proprietary format (e.g. CSV instead of excel)
####	Uniform structure for the analysis results within trial
#####	Common uniform structure for the analysis results across trials
#####	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff - from 5 star
#####	All the above, plus: Link your data to other people's data to provide context - from 5 star

Confidentiality, privacy?

Analysis Results Generation Process

Steps in using Analysis Results



Phuse subgroup analysis results meta data Status. 25 apr 2014 for Emerging Technologies Semantic Technologies TC

Possible uses / how it may change TFL production

Two approaches - same results

ADAM to analysis results to RDF

ADAM to RDF to analysis results

Usages - with RDF - every results has a URI

Provide results with links back to definitions

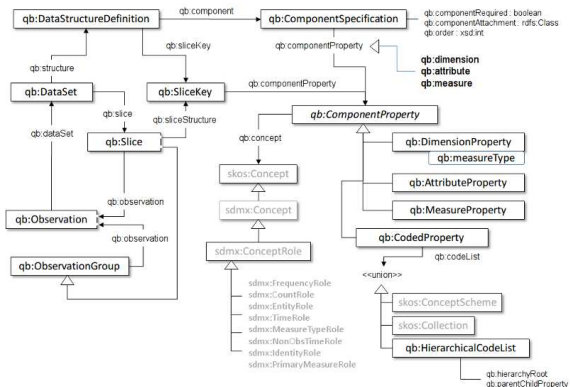
Combine results by cut-paste value and link (URI)

Publish trial results for submission (how to get them in correct format)

Summary

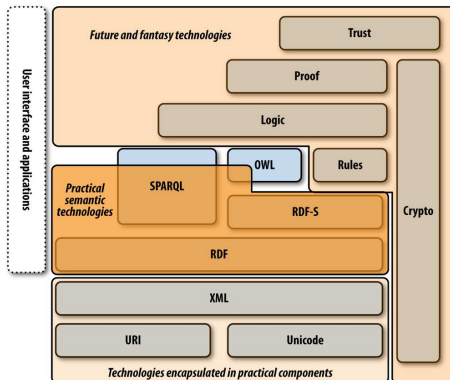
- ▶ What it is:
 - ▶ Linked data: DbPedia, DrugBank, LinkedCT
 - ▶ Resource Descriptor Format (RDF,
<http://www.w3.org/RDF/>)
 - ▶ SPARQL Query Language for RDF
(<http://www.w3.org/TR/rdf-sparql-query/>)
- ▶ Potential Use:
 - ▶ Metadata (and data)
 - ▶ Analysis Results Metadata - PhUSE working group

RDF data cube vocabulary, W3 recommendation



Source Figure 1 in RDF Data cube: <http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/>

A practical view of the semantic stack



Source: Figure 11-2. A practical view of the semantic stack; from Programming the Semantic Web, 2009. By: Toby Segaran; Colin Evans; Jamie Taylor; <http://shop.oreilly.com/product/9780596153823.do>

Tools

Triple Stores and SPARQL endpoints

See <http://en.wikipedia.org/wiki/Triplestore>

RDF Tools

Protege <http://protege.stanford.edu/>

TopBraid Composer

<http://www.topquadrant.com/downloads/topbraid-composer-install/>

Statistical Software

R: rrdf <http://cran.r-project.org/package=rrdf>

SAS: SAS macro for accessing SPARQL endpoint

<http://github.com/MarcJAndersen/SAS-SPARQLwrapper>