



Semantic Technologies and CDISC Standards

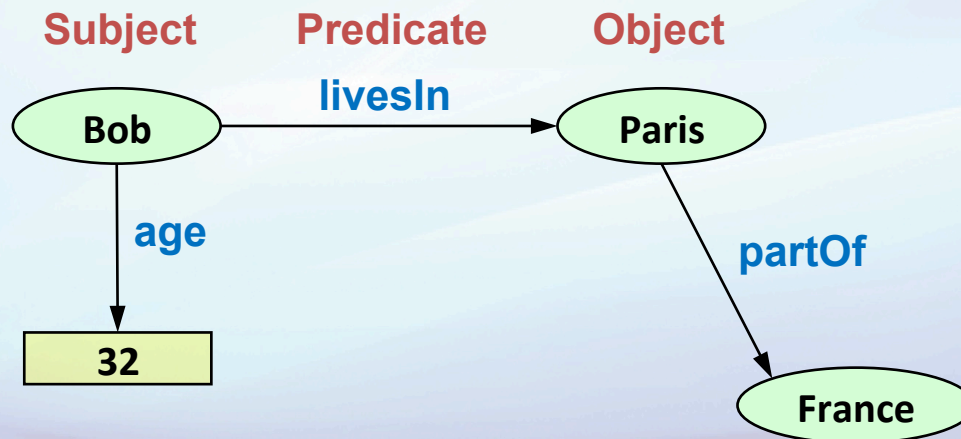
Frederik Malfait, Information Architect, IMOS Consulting
Scott Bahlavooni, Independent



Part I

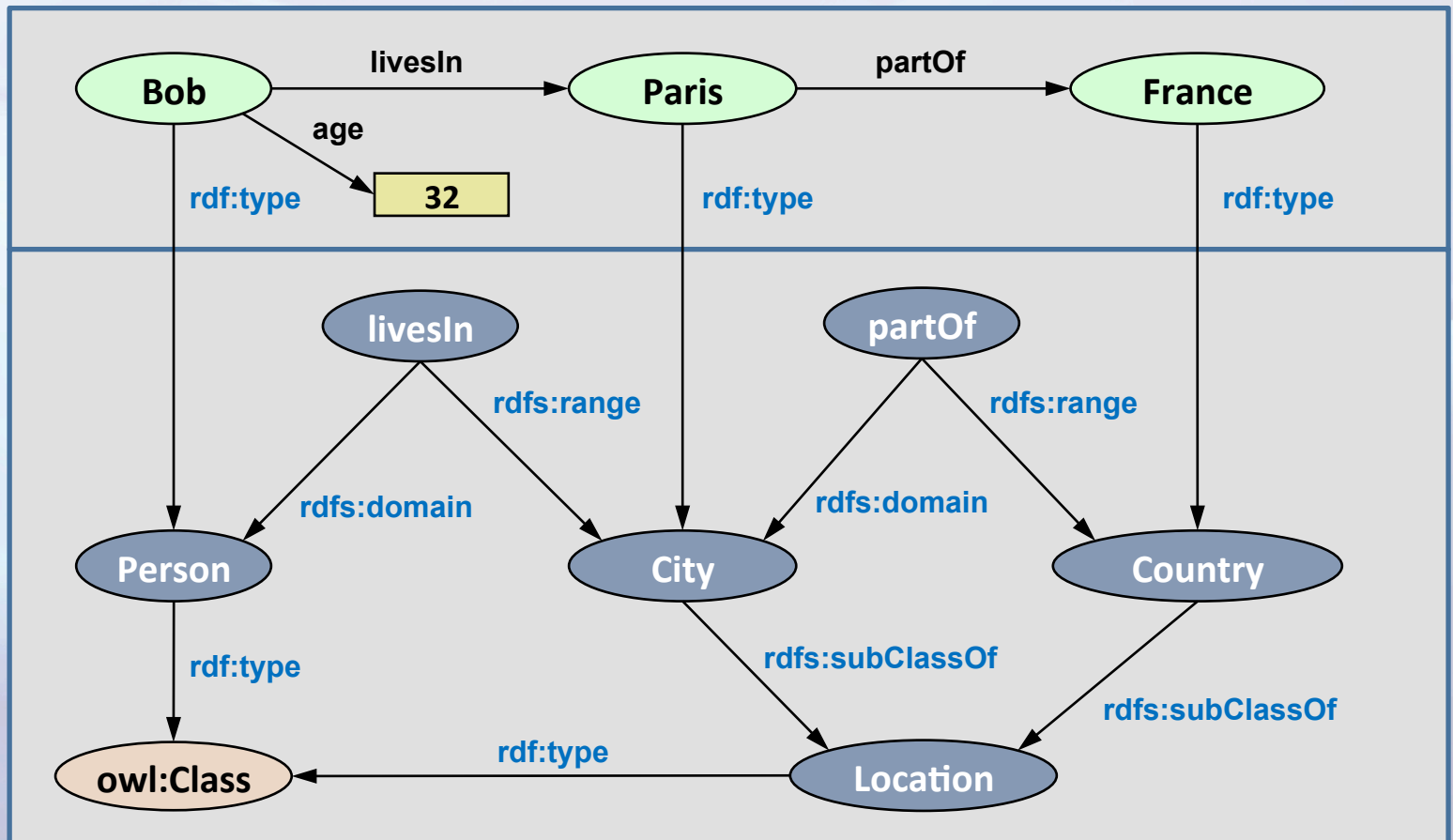
Introduction to Semantic Technology

Directed graph of <Subject Predicate Object> triples



Fact Base

Ontology





Uniform Resource Identifier - URI

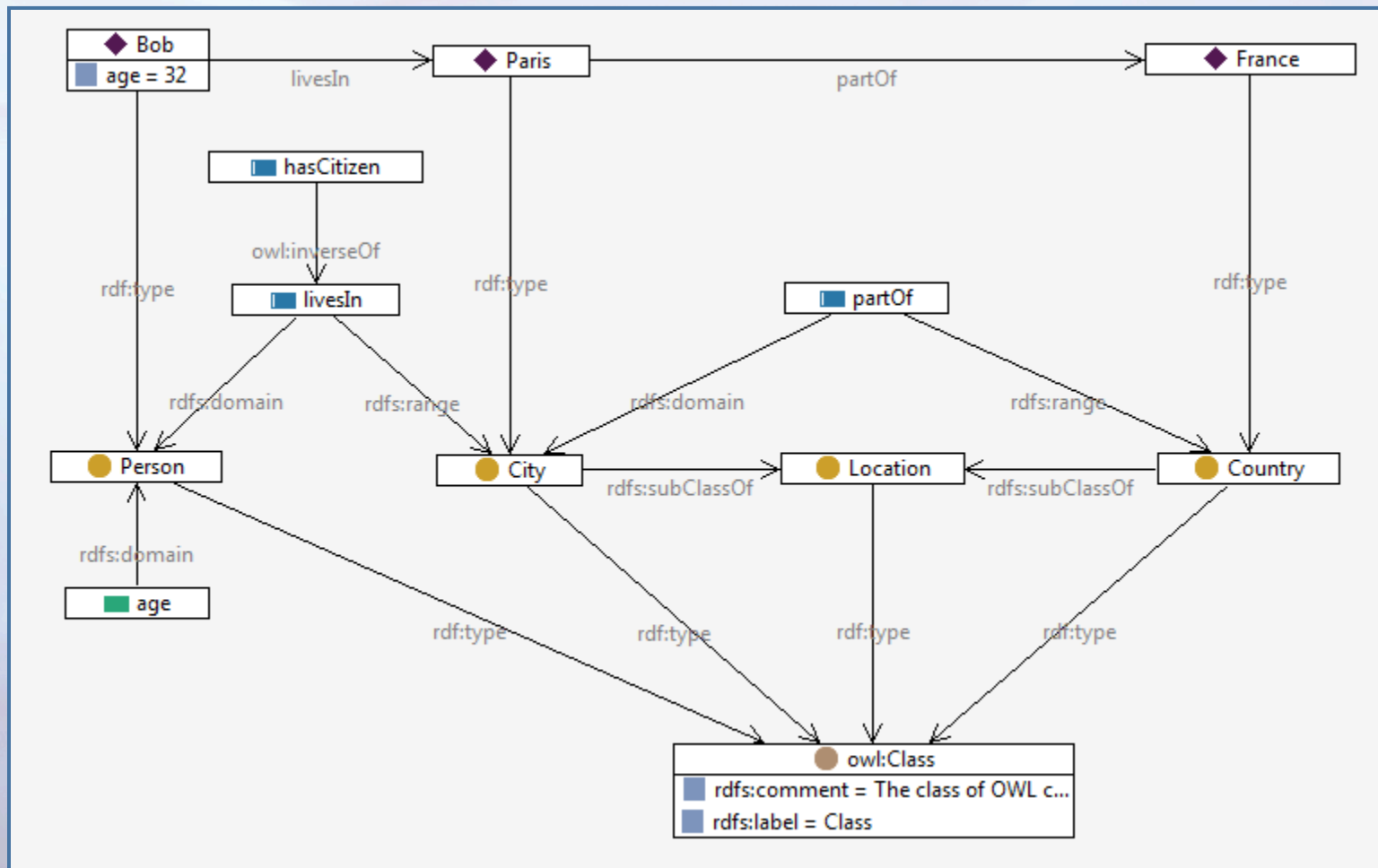
- In RDF everything is a triple built from resources and literals
 - <Subject Predicate Object>
 - <Subject Predicate Literal>
- Every RDF resource has a Uniform Resource Identifier (URI)
 - NCI Thesaurus, term Mild Adverse Event
<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C41338>
- Namespaces provide a convenient way to group related resources together
 - Define the namespace prefix nci
[xmlns:nci="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#"](http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#)
 - To obtain the Qualified Name (qname)
[nci:C41338](#)



Inference

- RDFS and OWL define an extensive vocabulary for building ontologies
 - Example: `<hasCitizen owl:inverseOf livesIn>`
- Meaning is defined by the way new triples are inferred from existing triples
 - Stated Triples
 - `<Bob livesIn Paris>`
 - `<hasCitizen owl:inverseOf livesIn>`
 - Inferred Triple
 - `<Paris hasCitizen Bob>`
- Inference engines provide reasoning capabilities for formal ontologies
- Formal ontologies have a precise mathematical foundation based on first-order logic, set theory, and model theory

RDF Representation - Graphs





RDF Representation - Ontology Editors

Navigator Classes Properties Basket

- owl:Thing
 - Location
 - City (1)
 - Country (1)
 - Person (1)

Navigator Classes Properties Basket

Grouped by Namespace

- <Default Namespace>
 - age
 - hasCitizen
 - livesIn
 - partOf
- owl
- rdf
- rdfs
- xsd

people.owl

Object Property Form

URI:

Annotations

Other Properties

rdf:type

Property Axioms

rdfs:domain

rdfs:range

Form Browser Graph Source Code



RDF Representation - Serializations

RDF/XML

```
<rdf:Description rdf:about="#Bob">
  <rdf:type rdf:resource="#Person"/>
  <age rdf:datatype="http://www.w3.org/2001/XMLSchema#positiveInteger">32</age>
  <livesIn rdf:resource="#Paris"/>
</rdf:Description>
```

Turtle

```
:Bob a      :Person ;
      :age   "32"^^xsd:positiveInteger ;
      :livesIn :Paris .
```

N-Triple

```
<http://gdsr.roche.com/people#Bob> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://gdsr.roche.com/people#Person> .
<http://gdsr.roche.com/people#Bob> <http://gdsr.roche.com/people#age> "32"^^<http://www.w3.org/2001/XMLSchema#positiveInteger> .
<http://gdsr.roche.com/people#Bob> <http://gdsr.roche.com/people#livesIn> <http://gdsr.roche.com/people#Paris> .
```




RDF Query Language - SPARQL

```
Query Editor Query Library
SELECT ?raveForm ?formOrdinal ?formLabel
WHERE {
  ?raveForm rdf:type mms:DataCollectionForm .
  ?raveForm mdr:context mms:Model.PD-Biometrics.Rave .
  ?raveForm mms:extendsModelComponent/mms:subDataElementDomainOf daco:DED.Diabetes .
  ?raveForm mms:extendsModelComponent/mms:domainOrdinal ?formOrdinal .
  ?raveForm mms:extendsModelComponent/mms:domainLabel ?formLabel .
}
ORDER BY ?formOrdinal
```

raveForm	formOrdinal	formLabel
◆ Ext.Form.Diabetes.History	50201	S Diabetes History
◆ Ext.Form.Diabetes.HypoglycemicEvents	50202	S Hypoglycemic Events
◆ Ext.Form.Diabetes.HomeGlucometerReadings	50203	S Home Glucometer Readings (for hypoglycemic events)
◆ Ext.Form.Diabetes.AllergicReactionEvents	50204	S Allergic Reaction Events
◆ Ext.Form.Diabetes.PancreatitisEvents	50205	S Pancreatitis Events



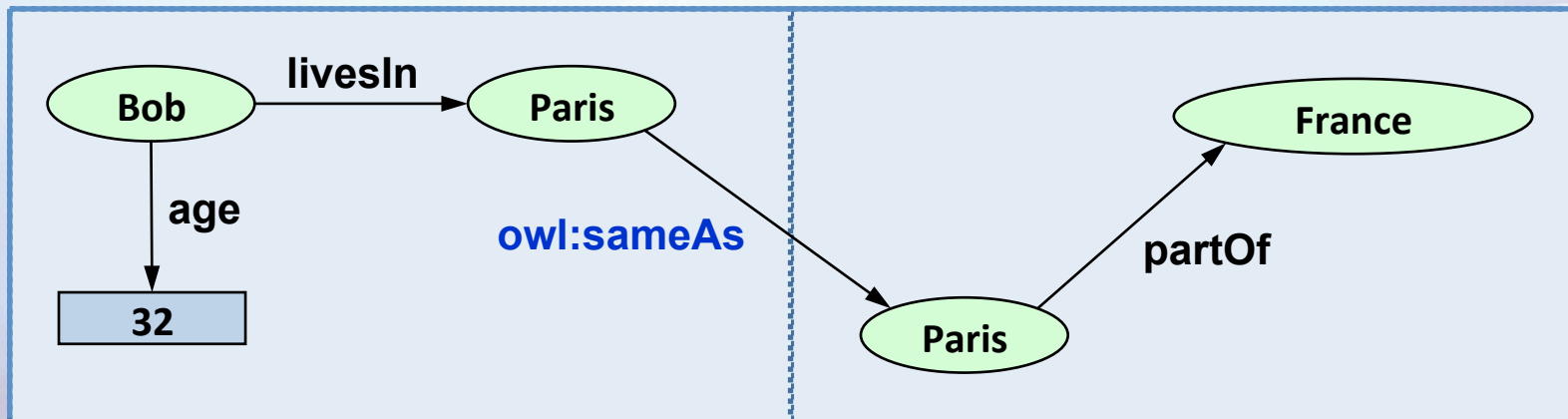
W3C Semantic Standards

- Resource Description Framework (RDF)
 - RDF defines how to express information as a directed graph of resources
 - Every resource has a URI and is part of a namespace
- RDF Schema (RDFS) and Web Ontology Language (OWL)
 - A set of standard predicates to build ontologies (vocabularies, schemas)
 - Inference capabilities
- SPARQL Protocol and RDF Query Language (SPARQL)
 - Language to query a RDF data sets
- Simple Knowledge Organization System (SKOS)
 - Small RDF vocabulary for building concept models (thesauri, etc.)

- Semantic models in RDF format are easy to federate
- Federation of Data = union of triples from both graphs
- Linked Data = referencing between graphs

<http://example.org/people#>

<http://example.org/geo/#>

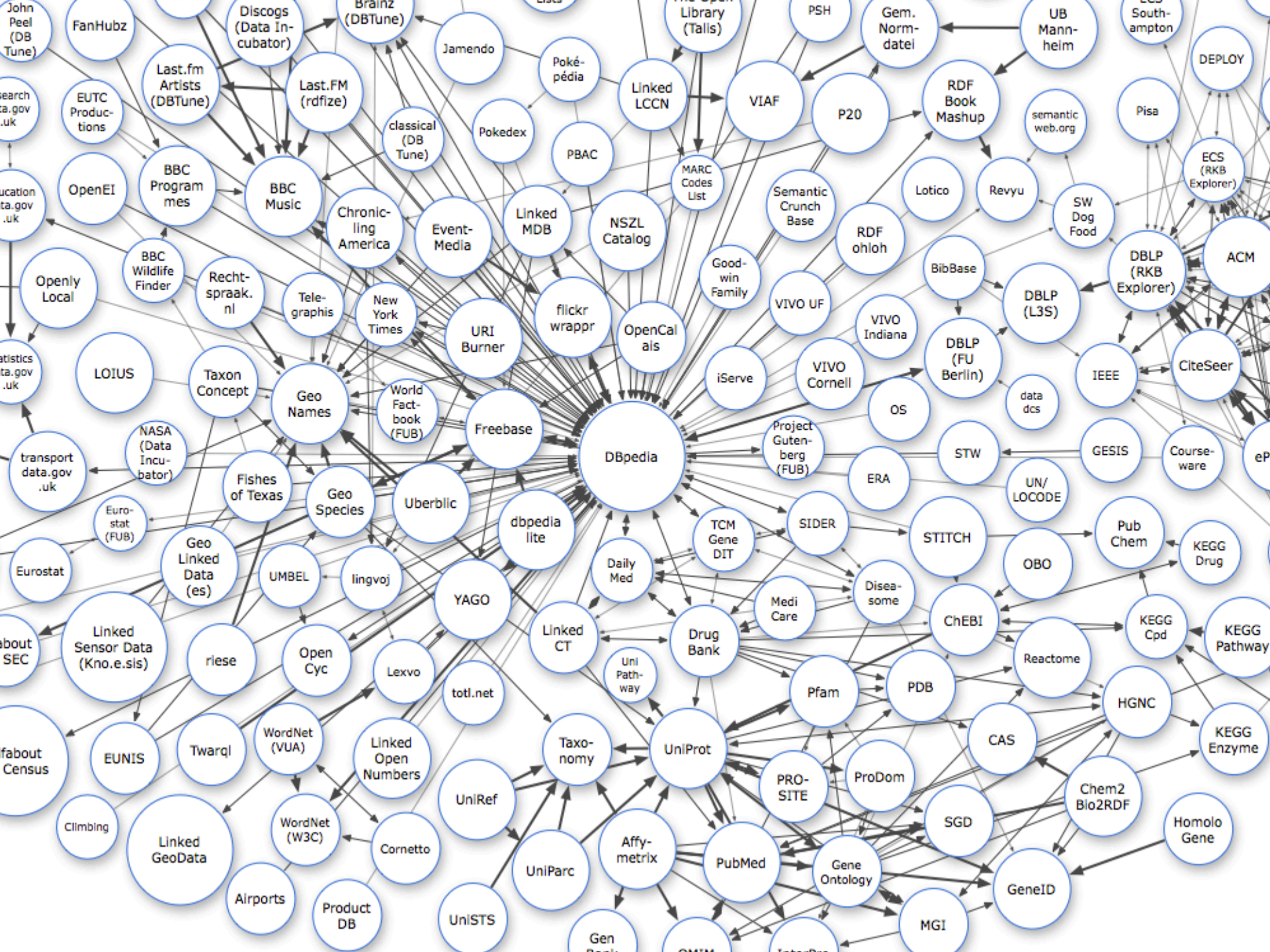


Federated Graph



dbpedia.org

In addition, we provide localized versions of DBpedia in 111 languages. All these versions together describe 20.8 million things, out of which 10.5 million overlap (are interlinked) with concepts from the English DBpedia. The full DBpedia data set features labels and abstracts for 10.3 million unique things in up to 111 different languages; 8.0 million links to images and 24.4 million HTML links to external web pages; 27.2 million data links into external RDF data sets, 55.8 million links to Wikipedia categories, and 8.2 million YAGO categories. The dataset consists of 1.89 billion pieces of information (RDF triples) out of which 400 million were extracted from the English edition of Wikipedia, 1.46 billion were extracted from other language editions, and about 27 million are data links to external RDF data sets. The [Datasets](#) page provides more information about the overall structure of the dataset. [Dataset Statistics](#) provides detailed statistics about 22 of the 111 localized versions.

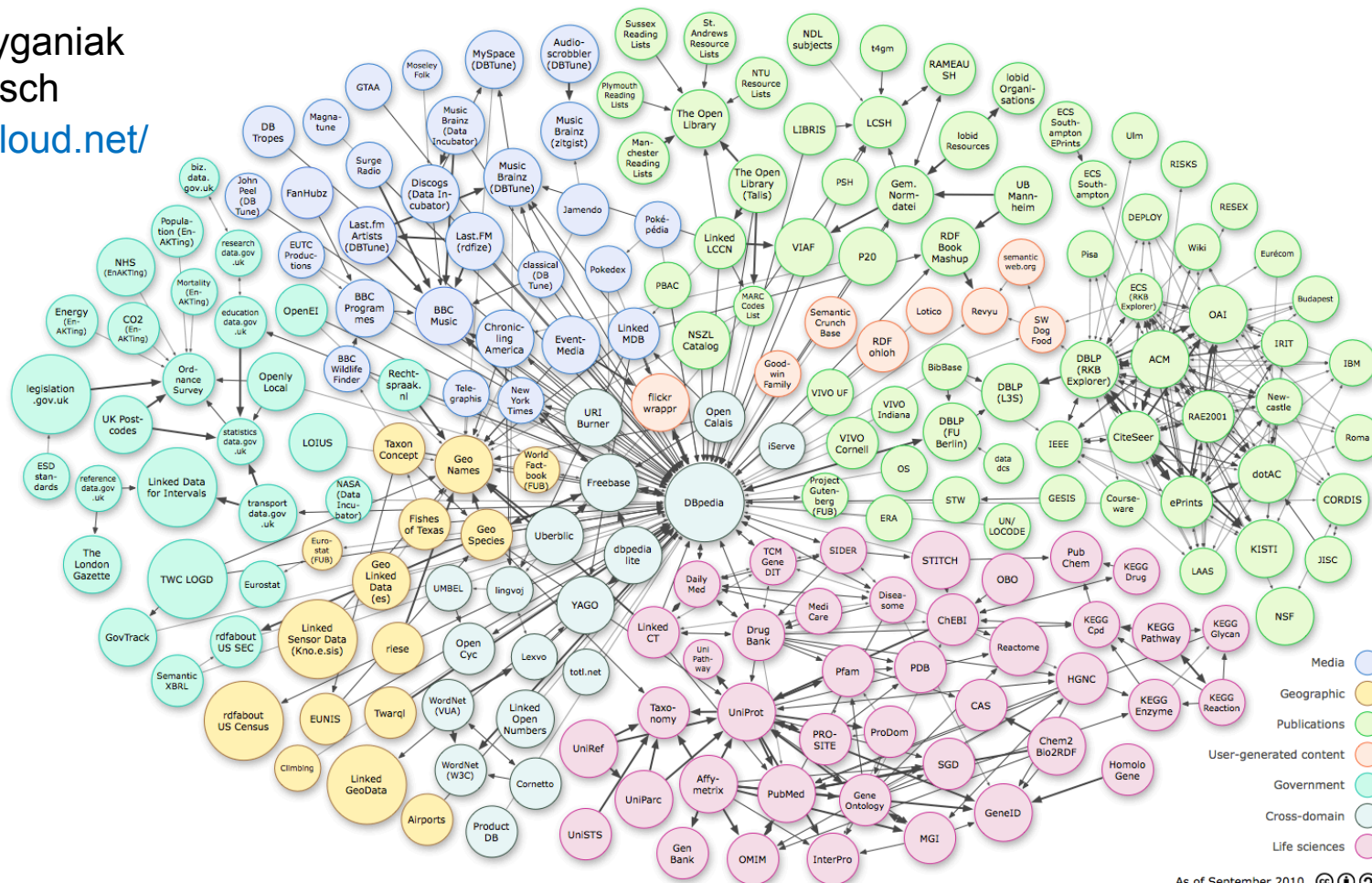


Linked Open Data (LOD) Cloud

Richard Cyganiak

Anja Jentzsch

<http://lod-cloud.net/>





Semantic Technology Application Areas

- Semantic Web
 - Move the web beyond a web of documents, add semantics
 - Large scale web data integration
- Knowledge Management
 - Vocabularies, taxonomies, thesauri, nomenclature, dictionaries e.g. SNOMED CT, NCI Thesaurus, HGNC, etc.
 - Domain ontologies, i.e. formal knowledge representation and reasoning capabilities for a certain domain of interest
- Roche Data Standards Office
 - Data standards and metadata driven workflow
 - Metadata registries, e.g. based on ISO 11179
 - Data integration and master data management



Part II

Metadata Management at Hoffmann-La Roche



Roche Data Standards Office

- Design and manage end-to-end clinical trial data standards
 - from protocol to submission
 - with supporting business processes
- Drive adoption of current and future industry standards
- Provide lifecycle management and governance for all data standards
- Contribute to external standards bodies
 - CDISC
 - TransCelerate
 - Phuse/FDA



Roche Global Data Standards Repository

- Single point of access for end-to-end data standards
 - Protocol and Study Design
 - Data Collection, Data Tabulation, Data Analysis
 - Regulatory Submission
- Enable consistent metadata management
 - Operational and conceptual metadata
 - Versioning and life cycle management
- Follow industry standards when available and fit for purpose
 - CDISC standards
 - ISO 11179 Metadata Registry (MDR) standards
 - W3C semantic standards (RDF, OWL, SKOS)



GDSR Data Standards Content

- External content
 - SDTM 1.2, SDTMIG 3.1.2, and amendments
 - CDISC Controlled Terminology
 - NCI Thesaurus
- Integrated data standards, Roche and Genentech
 - Data collection: Safety and every Roche TA (2000 data elements)
 - Data tabulation: SDTM Sponsor Extensions
- Value level metadata
 - Lab measurements (2000)
 - Preferred units and unit conversions
 - Questionnaires (150)



GDSR Semantic Information Model

Production

Partial / Future

Sponsor Extensions

PRM

CDASH

SDTM

ADaM

define

Biomedical Concept Model [partial]

ISO 11179 MDR Schema [subset]

BRIDG and ISO 21090

NCI Thesaurus

RDF

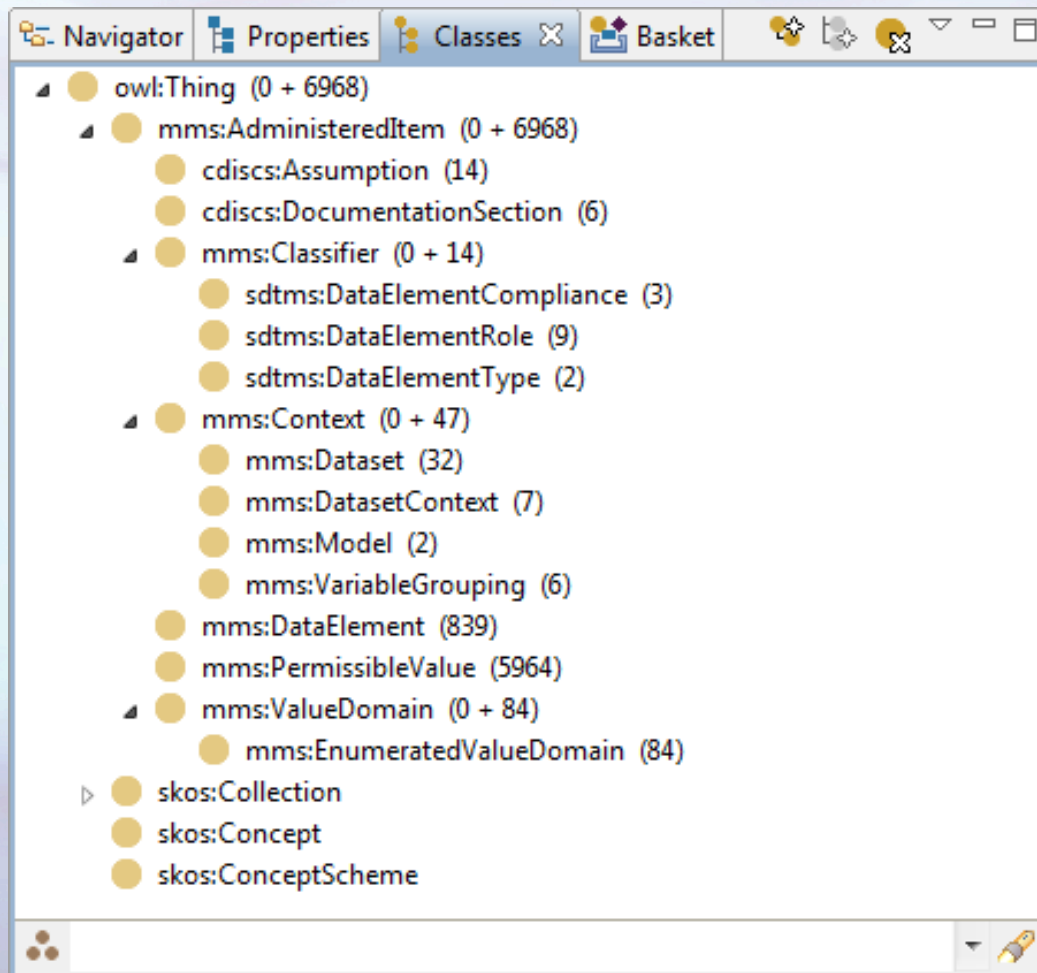
OWL

SKOS

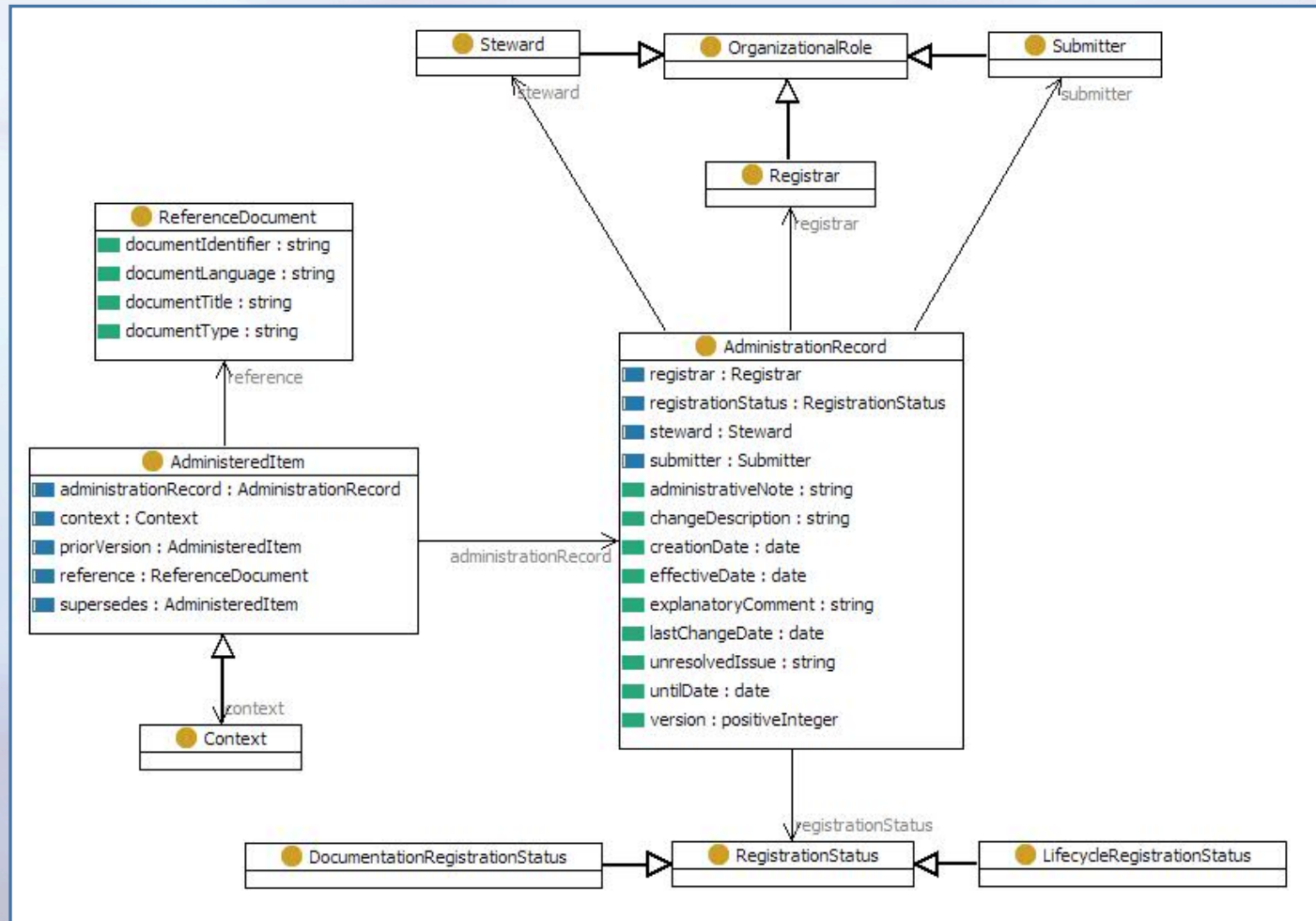
**CDISC
Standards**

**Metadata
Management**

**Knowledge
Management**

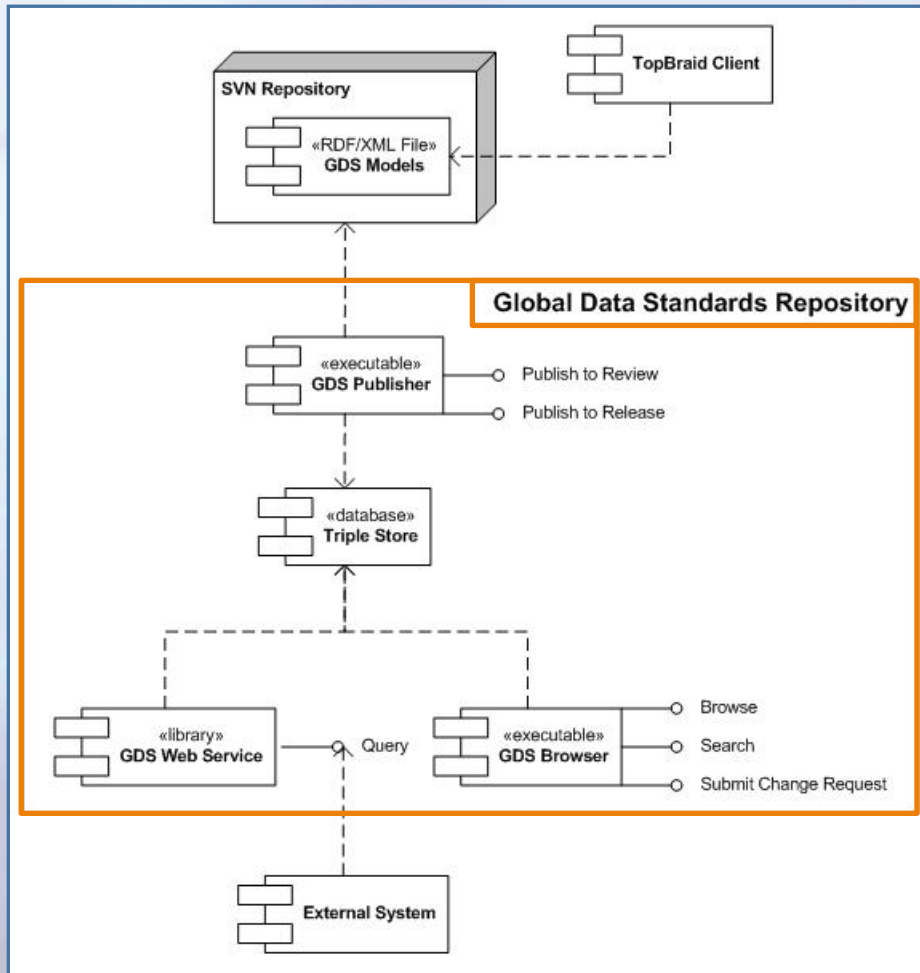


GDSR ISO 11179 Part 6





GDSR System Components



Content Management

Content Publishing

Metadata Repository

Single Point of Access

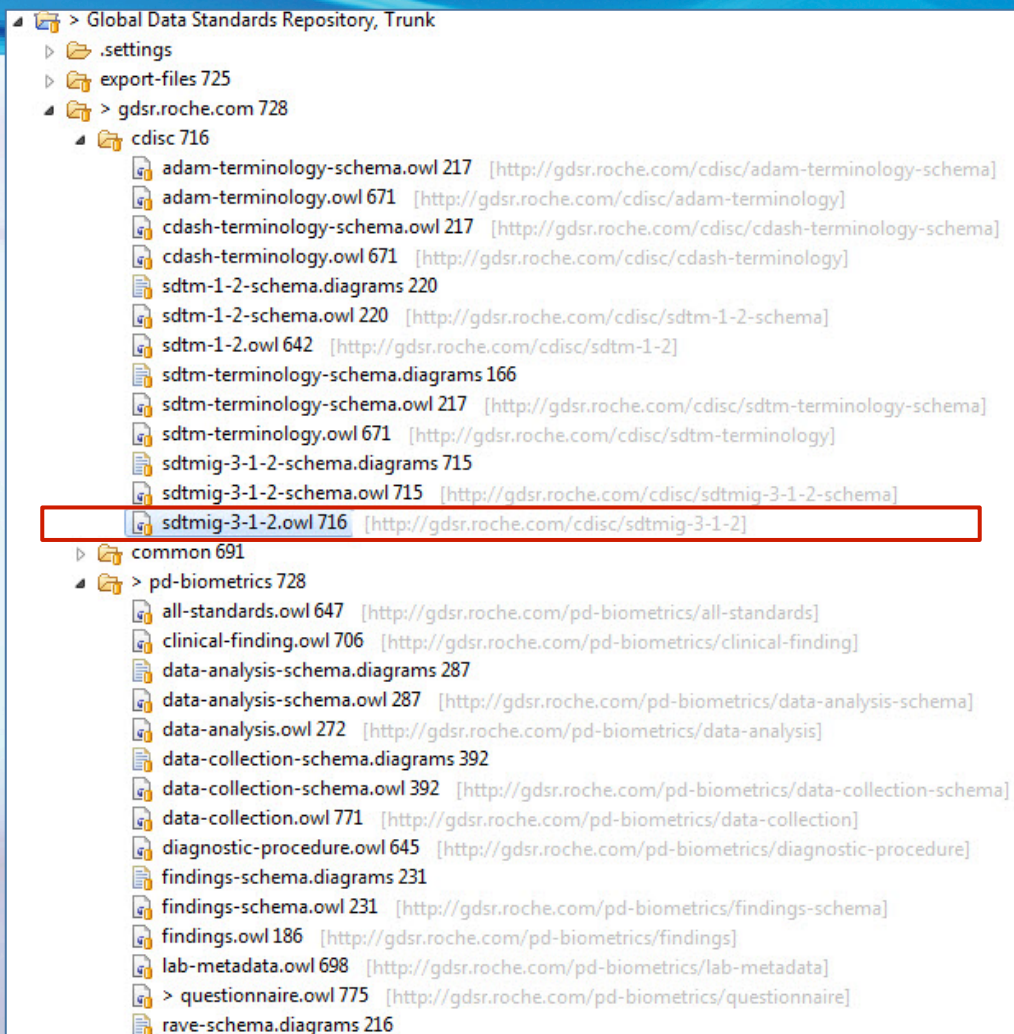


Part III

Semantic Models for CDISC Standards



SDTMIG 3.1.2 Walk-Through





SDTMIG 3.1.2 Walk-Through

Class Form

URI:

Annotations

rdfs:isDefinedBy

rdfs:label

skos:definition

Class Axioms

rdfs:subClassOf

Other Properties

rdf:type

Form Diagram Graph Form Layout Source Code

Instances

[Resource]	rdfs:label	rdfs:comment
Table.AE		
Table.CE		
Table.DS		
Table.DV		
Table.MH		



SDTMIG 3.1.2 Walk-Through

Resource Form

URI: <http://gdsr.roche.com/cdisc/sdtmig-3-1-2#Table.AE>

Annotations

Incoming References

← mms:dataElementDomain ▾

- ◆ Column.AE.AEACN ▾
- ◆ Column.AE.AEACNOTH ▾
- ◆ Column.AE.AEBODSYS ▾
- ◆ Column.AE.AECAT ▾
- ◆ Column.AE.AECONTRT ▾
- ◆ Column.AE.AEDECOD ▾
- ◆ Column.AE.AEDUR ▾
- ◆ Column.AE.AEENDTC ▾
- ◆ Column.AE.AEENDY ▾
- ◆ Column.AE.AEENRF ▾
- ◆ Column.AE.AEENRTPT ▾
- ◆ Column.AE.AEENTPT ▾
- ◆ Column.AE.AEGRPID ▾
- ◆ Column.AE.AELOC ▾
- ◆ Column.AE.AEMODIFY ▾
- ◆ Column.AE.AEOUT ▾
- ◆ Column.AE.AEPATT ▾
- ◆ Column.AE.AEPRESP ▾
- ◆ Column.AE.AEREFID ▾
- ◆ Column.AE.AEREL ▾

Other Properties

sdtmigs:domainCode ▾

S AE ▾

sdtmigs:domainStructure ▾

S One record per adverse event per subject ▾

mms:domainLabel ▾

Adverse Events ▾

mms:domainName ▾

S AE ▾

mms:domainOrdinal ▾

8 ▾

mdr:administrationRecord ▾

◆ mdr:AR.SDTMIG-3-1-2 ▾

mdr:context ▾

◆ mms:Model.SDTMIG-3-1-2 ▾

rdf:type ▾

● sdtmigs:EventsObservationClass ▾



SDTMIG 3.1.2 Walk-Through

Resource Form

URL: <http://gdsr.roche.com/cdisc/sdtmig-3-1-2#Column.AE.AEOUT> Ok

Annotations

Incoming References

Other Properties

sdtms:dataElementCompliance ▾
◆ sdtms:Classifier.PermissibleVariable ▾

sdtms:dataElementRole ▾
◆ sdtms:Classifier.RecordQualifier ▾

sdtms:dataElementType ▾
◆ sdtms:Classifier.Char ▾

sdtmigs:controlledTermsOrFormat ▾
S (OUT) ▾

sdtmigs:dataElementCodelist ▾
◆ sdtmt:C66768 ▾

sdtmigs:references ▾
S SDTM 2.2.2 ▾

mms:dataElement ▾
◆ sdtm:DE.Event.--OUT ▾

mms:dataElementDescription ▾
S Description of the outcome of an event. ▾

mms:dataElementDomain ▾
◆ Table.AE ▾

mms:dataElementLabel ▾
S Outcome of Adverse Event ▾

mms:dataElementName ▾
S AEOUT ▾



SDTMIG 3.1.2 Walk-Through

Resource Form

URI: <http://gdsr.roche.com/cdisc/sdtm-terminology#C66768>

Annotations

Incoming References

← sdtmts:parentCodelist ▾

sdmt:C66768.C17998 ▾

sdmt:C66768.C48275 ▾

sdmt:C66768.C49494 ▾

sdmt:C66768.C49495 ▾

sdmt:C66768.C49496 ▾

sdmt:C66768.C49498 ▾

← sdtmigs:dataElementCodelist ▾

Column.AE.AEOUT ▾

Other Properties

sdtmts:cdiscDefinition ▾

A condition or event that is attributed to the adverse event and is the result or conclusion of the adverse event. (NCI)

sdtmts:cdiscSubmissionValue ▾

OUT ▾

sdtmts:cdiscSynonyms ▾

Outcome of Event ▾

sdtmts:codelistName ▾

Outcome of Event ▾

sdtmts:isExtensibleCodelist ▾

false ▾

sdtmts:nciCode ▾

C66768 ▾

sdtmts:nciPreferredTerm ▾

CDISC SDTM Adverse Event Outcome Terminology ▾

rdf:type ▾

sdtmts:Codelist ▾



SDTMIG 3.1.2 Walk-Through

Resource Form

URI: <http://gdsr.roche.com/cdisc/sdtm-terminology#C66768.C49494>

Annotations

Incoming References

Other Properties

sdtmts:cdiscDefinition

One of the possible results of an adverse event outcome that indicates that the event has not improved or recuperated. (NCI)

sdtmts:cdiscSubmissionValue

NOT RECOVERED/NOT RESOLVED

sdtmts:nciCode

C49494

sdtmts:nciPreferredTerm

Not Recovered or Not Resolved

sdtmts:parentCodelist

sdtmt:C66768

rdf:type

sdtmts:CodelistElement



SDTMIG 3.1.2 Walk-Through

Resource Form

URL: <http://gdsr.roche.com/cdisc/sdtmig-3-1-2#Column.AE.AEOUT>

Annotations

Incoming References

Other Properties

sdtms:dataElementCompliance
sdtms:Classifier.PermissibleVariable

sdtms:dataElementRole
sdtms:Classifier.RecordQualifier

sdtms:dataElementType
sdtms:Classifier.Char

sdtmigs:controlledTermsOrFormat
S (OUT)

sdtmigs:dataElementCodelist
sdmt:C66768

sdtmigs:references
S SDTM 2.2.2

mms:dataElement
sdtm:DE.Event.--OUT

mms:dataElementDescription
S Description of the outcome of an event.

mms:dataElementDomain
Table.AE

mms:dataElementLabel
S Outcome of Adverse Event

mms:dataElementName
S AEOUT



Resource Form

URI:

Annotations

rdfs:isDefinedBy

rdfs:label

skos:definition

Incoming References

- ← sdtns:dataElementCompliance
- Column.AE.AEACNOTH
- Column.AE.AECAT
- Column.AE.AECONTRT
- Column.AE.AEDUR
- Column.AE.AEENDY
- Column.AE.AEENRF
- Column.AE.AEENRTPT
- Column.AE.AEENTPT

Other Properties

rdf:type



Part IV

RDF Model Driven REST Services



Model Driven REST Service Model

- True resource oriented service model using HTTP GET <URI> requests
- GET response is model driven and fully configurable in RDF describing
 - which resource properties and relationships to return
 - how and where to retrieve the response data from RDF
 - how to format the response
 - how to optionally transform an XML response using XSLT and XSL-FO
- Requests can ask for
 - data formats XML, JSON, CSV
 - schema formats XSD, Java JAXB ready classes
- Model driven REST services are specified in an ontology based on the idea of a resource facet describing a logical view of a resource
- Resources may have multiple facets and apply facet composition



GET <SDTM Variable URI>

```
1 <variable
2   xmlns="http://gdsr.roche.com/ws/data-tabulation/facet.dt.variable.default"
3   uri="http://gdsr.roche.com/pd-biometrics/sdtm-ext#ColumnExtension.VS.DOMAIN">
4   <variableOrdinal>2</variableOrdinal>
5   <name>DOMAIN</name>
6   <label>Domain Abbreviation</label>
7   <type>Char</type>
8   <length>2</length>
9   <controlledTermsOrFormat>VS</controlledTermsOrFormat>
10  <role>Identifier Variable</role>
11  <cdiscNotes>Two-character abbreviation for the domain.</cdiscNotes>
12  <sponsorNotes>Set to 'VS'.</sponsorNotes>
13  <compliance>Required Variable</compliance>
14  <reference>SDTM 2.2.4, SDTMIG 4.1.2.2, SDTMIG Appendix C2</reference>
15 </variable>
```



GET <SDTM Domain URI>

```
1 <domain xmlns="http://gdsr.roche.com/ws/data-tabulation/facet.dt.domain.default"
2   uri="http://gdsr.roche.com/cdisc/sdtmig-3-1-2#Table.VS"
3   label="Vital Signs">
4   <domainOrdinal>19</domainOrdinal>
5   <observationClass>Findings Observation Class</observationClass>
6   <datasetName>VS</datasetName>
7   <domainCode>VS</domainCode>
8   <domainLabel>Vital Signs</domainLabel>
9   <variableList>
10    <variable uri="http://gdsr.roche.com/pd-biometrics/sdtm-ext#ColumnExtension.VS.STUDYID">
11      <variableOrdinal>1</variableOrdinal>
12      <name>STUDYID</name>
13      <label>Study Identifier</label>
14      <type>Char</type>
15      <length>8</length>
16      <role>Identifier Variable</role>
17      <cdiscNotes>Unique identifier for a study.</cdiscNotes>
18      <sponsorNotes>Set to STUDYID in SDTM.DM domain.</sponsorNotes>
19      <compliance>Required Variable</compliance>
20      <reference>SDTM 2.2.4</reference>
21    </variable>
22    <variable uri="http://gdsr.roche.com/pd-biometrics/sdtm-ext#ColumnExtension.VS.DOMAIN">
23      <variableOrdinal>2</variableOrdinal>
24      <name>DOMAIN</name>
25      <label>Domain Abbreviation</label>
26      <type>Char</type>
27      <length>2</length>
28      <controlledTermsOrFormat>VS</controlledTermsOrFormat>
29      <role>Identifier Variable</role>
30      <cdiscNotes>Two-character abbreviation for the domain.</cdiscNotes>
```



GET <Measurement URI>

```
1  <measurement
2      xmlns="http://gdsr.roche.com/ws/lab-metadata/facet.measurement.default"
3      uri="http://gdsr.roche.com/pd-biometrics/lab-metadata#Analyte.ALBUMIN">
4      <label>Albumin</label>
5      <code>ALBUM</code>
6      <name>ALBUM</name>
7      <lbtestcd>ALB</lbtestcd>
8      <lbtest>Albumin</lbtest>
9      <lbspec>BLOOD</lbspec>
10     <hasNoUnits>false</hasNoUnits>
11     <SI-Unit>g/L</SI-Unit>
12     <USCV-Unit>g/dL</USCV-Unit>
13 </measurement>
```



GET <Measurement Category URI>

```
1  <measurementCategory
2      xmlns="http://gdsr.roche.com/ws/lab-metadata/facet.measurementCategory.default"
3      uri="http://gdsr.roche.com/pd-biometrics/lab-metadata#MeasurementCategory.Chemistry">
4      <label>Chemistry</label>
5      <domain>LB</domain>
6      <lbcat>CHEMISTRY</lbcat>
7      <measurementList>
8          <measurement uri="http://gdsr.roche.com/pd-biometrics/lab-metadata#Analyte.A1AGLP">
9              <label>Alpha-1 Acid Glycoprotein</label>
10             <code>A1AGLP</code>
11             <name>A1AGLP</name>
12             <lbtestcd>A1AGLP</lbtestcd>
13             <lbtest>Alpha-1 Acid Glycoprotein</lbtest>
14             <lbspec>BLOOD</lbspec>
15             <hasNoUnits>false</hasNoUnits>
16             <SI-Unit>g/L</SI-Unit>
17             <USCV-Unit>mg/mL</USCV-Unit>
18             <sponsorNotes>AAG is a common synonym for this analyte.</sponsorNotes>
19         </measurement>
20         <measurement uri="http://gdsr.roche.com/pd-biometrics/lab-metadata#Analyte.A1_ANTITRYPSIN_SER">
21             <label>Alpha-1 Antitrypsin, Serum Conc</label>
22             <code>A1ANTRY</code>
23             <name>A1ANTRY</name>
24             <lbtestcd>A1ANTRY</lbtestcd>
25             <lbtest>Alpha-1 Antitrypsin</lbtest>
26             <lbspec>BLOOD</lbspec>
27             <hasNoUnits>false</hasNoUnits>
28             <SI-Unit>g/L</SI-Unit>
29             <USCV-Unit>mg/dL</USCV-Unit>
30             <sponsorNotes>LEGACY - DO NOT USE</sponsorNotes>
31         </measurement>
```



GET <Code List URI>

```
1 <sdmCodelist
2   xmlns="http://gdsr.roche.com/ws/data-tabulation/facet.dt.sdmCodelist.default"
3   uri="http://gdsr.roche.com/cdisc/sdm-terminology#C101834">
4   <code>C101834</code>
5   <label>Normal Abnormal Response</label>
6   <extensible>false</extensible>
7   <cdiscSubmissionValue>NORMABNM</cdiscSubmissionValue>
8   <cdiscSynonyms>Normal Abnormal Response</cdiscSynonyms>
9   <cdiscDefinition>
10    A terminology codelist to describe a normal or abnormal response.
11  </cdiscDefinition>
12  <nciPreferredTerm>CDISC SDTM Normal Abnormal Response Terminology</nciPreferredTerm>
13  <valueList>
14    <value uri="http://gdsr.roche.com/cdisc/sdm-terminology#C101834.C14165">
15      <code>C14165</code>
16      <submissionValue>NORMAL</submissionValue>
17    </value>
18    <value uri="http://gdsr.roche.com/cdisc/sdm-terminology#C101834.C17998">
19      <code>C17998</code>
20      <submissionValue>UNKNOWN</submissionValue>
21    </value>
22    <value uri="http://gdsr.roche.com/cdisc/sdm-terminology#C101834.C25401">
23      <code>C25401</code>
24      <submissionValue>ABNORMAL</submissionValue>
25    </value>
26    <value uri="http://gdsr.roche.com/cdisc/sdm-terminology#C101834.C48658">
27      <code>C48658</code>
28      <submissionValue>INDETERMINATE</submissionValue>
29    </value>
30  </valueList>
31 </sdmCodelist>
```



REST Based Workflow Automation

- Representation of a protocol schedule of assessments
- Creation of a CRF view of data collection standards in PDF format
 - Operational View
 - Annotated CRF Submission View
- Creation of a Medidata Rave upload file (ALS)
- Creation of a non-CRF File Format Specification
- SDTM transformation specification and code generator
- Integration with Master Data and Document Management systems



Part V

Community Work



What we would like to see...

CDISC Ownership

sdtmig-3-1-2.owl

Resource Form

URI:

Annotations

Other Properties

mms:contextDescription

The Study Data Tabulation Model Implementation Guide is a CDISC defined guide for the implementation of SDTM providing a detailed specification of the SDTM domains.

mms:contextLabel

Study Data Tabulation Model Implementation Guide (SDTMIG) Version 3.1.2

mms:contextName

sdtmig-3-1-2

rdf:type

mms:Model

Incoming References

mms:context

EventsObservationClass

FindingsAbout

FindingsObservationClass

InterventionsObservationClass

RelationshipDataset

SpecialPurposeDomain

TrialDesignModel

Type: mms:DatasetContext (Hold CTRL to navigate)

Form | Browser | Graph | Source Code



Collaborations

- cdisc2rdf
 - Precompetitive collaboration to publish CDISC Standards in RDF
 - SDTM 1.2, IG 3.1.2, CT published January 2013
 - Discontinued and folded into FDA/PhUSE CSS work
- National Cancer Institute
 - Publishing CDISC Controlled Terminology in RDF since June 2013
- FDA/Phuse CSS 2013-2014
 - Semantic Technology project within the Emerging Technologies WG
- CDISC and TransCelerate
 - Contribute to the SHARE architecture and modeling effort
 - Define requirements for RDF Export of SHARE content



FDA/PhUSE CSS 2013-2014

- Phase I - April to August 2013
 - Represent existing CDISC standards in RDF
 - ✓ CDASH 1.1
 - ✓ SDTM 1.2, SDTMIG 3.1.2, SDTM 1.3, SDTM IG 3.1.3
 - ✓ SEND IG 3.0
 - ✓ ADaM 2.1, ADaMIG 1.0
 - Including domain assumptions
 - Integrated with Controlled Terminology
- Consolidation - September 2013
- Phase II - October 2013 to February 2014
 - Protocol Representation, Schedule of Assessments, Trial Design
 - Conformance Checks
 - Represent clinical trial data in RDF