VA HEALTH CARE Defining EXCELLENCE in the 21st Century

Linked Vitals

Linked Vitals

A Linked Data Approach to Semantic Interoperability

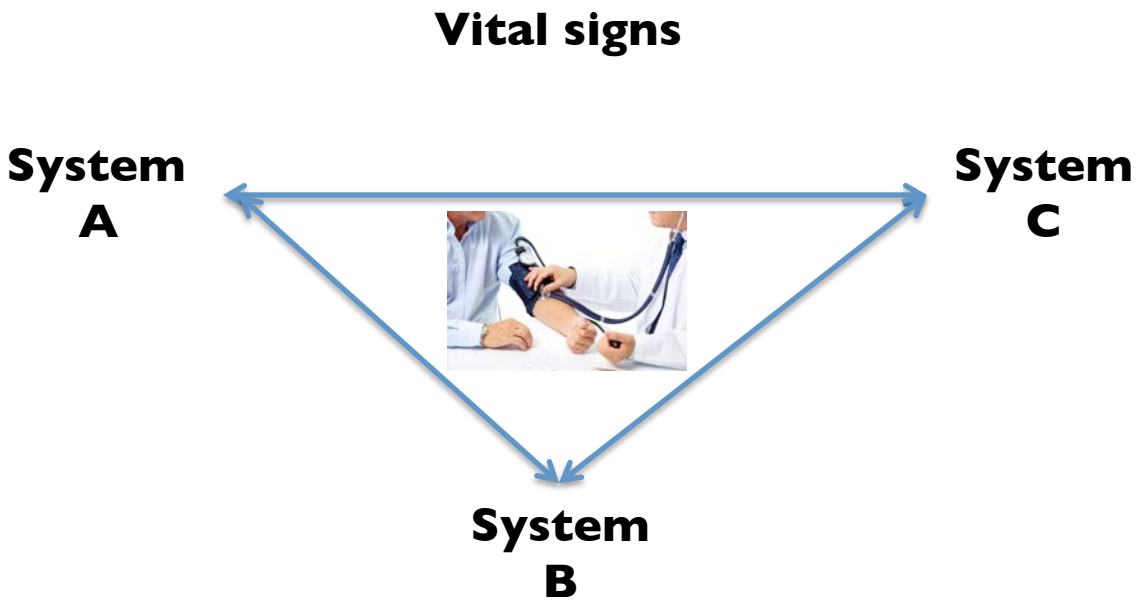
SemTech Conference
San Jose, CA
August 20, 2014

Rafael M Richards MD MS
Physician Informaticist
Office of Informatics and Analytics
Veterans Health Administration
U.S. Department of Veterans Affairs

Problem Statement: General

Linked Vitals

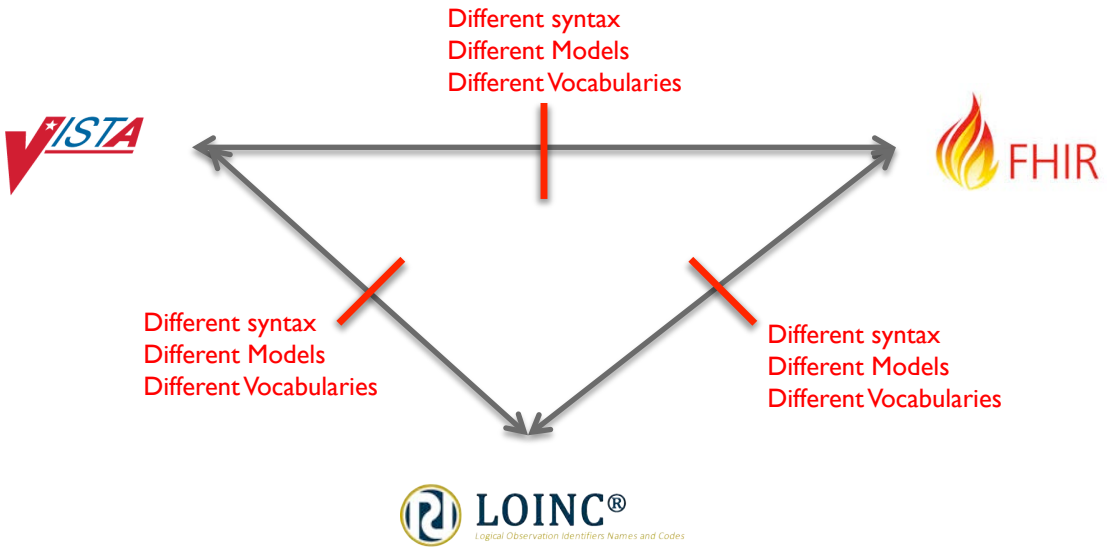
How do we integrate data such as vital signs *meaningfully* from different information systems?



Problem Statement: Specific

Linked Vitals

How do we integrate vital signs between VistA and FHIR each with different syntax, different models, and different vocabularies ?

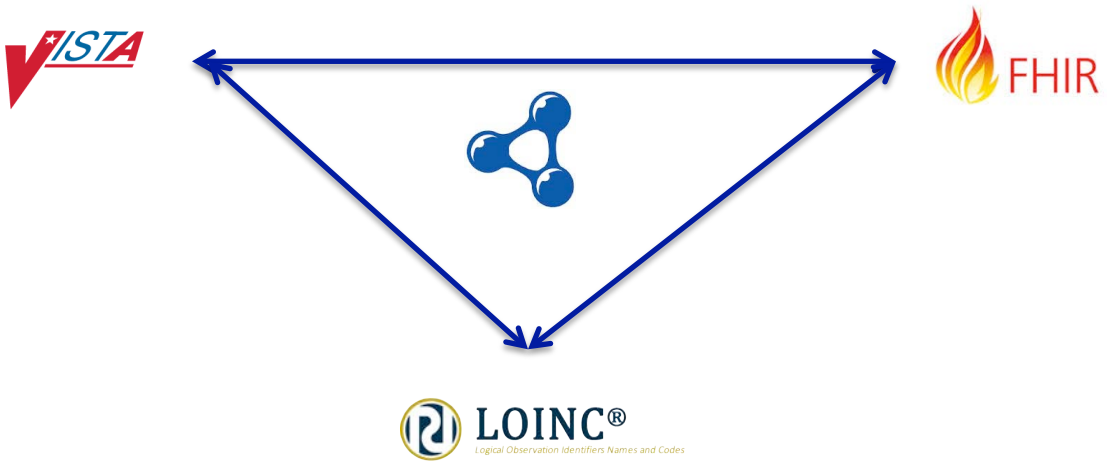


Language barriers to exchange

Solution Statement

Linked Vitals

How do we integrate vital signs between VistA and FHIR each with different syntax, different models, and different vocabularies ?



A common exchange language

Overview of Translation Process

Linked Vitals

Source Data	Syntatic alignment	SMART	Semantic alignment	Integrated Data

Syntax A
Model A
Vocabulary A

VISTA

FM Schema

Vocabulary B

LOINC®
Logical Observation Identifiers Names and

CSV

Syntax C
Model C
Vocabulary C

FHIR

XML

SPIN
SPARQL Inferencing Notation

Shared Syntax
Shared Model
Shared Vocabulary
Shared Meaning

Different Syntax
Different Models
Different Vocabularies

Common syntax within model-flexible medium (Linked Data)

Rule-based mapping
Model alignment
Vocabulary alignment

Linked Data (RDF): What is it?

Linked Vitals

The World Wide Web Consortium (W3C) standard for semantic information integration for the Internet of Data.

HTML (hypertext markup language)
For **humans** to exchange information

enables → **Linked Documents** (Document Web)

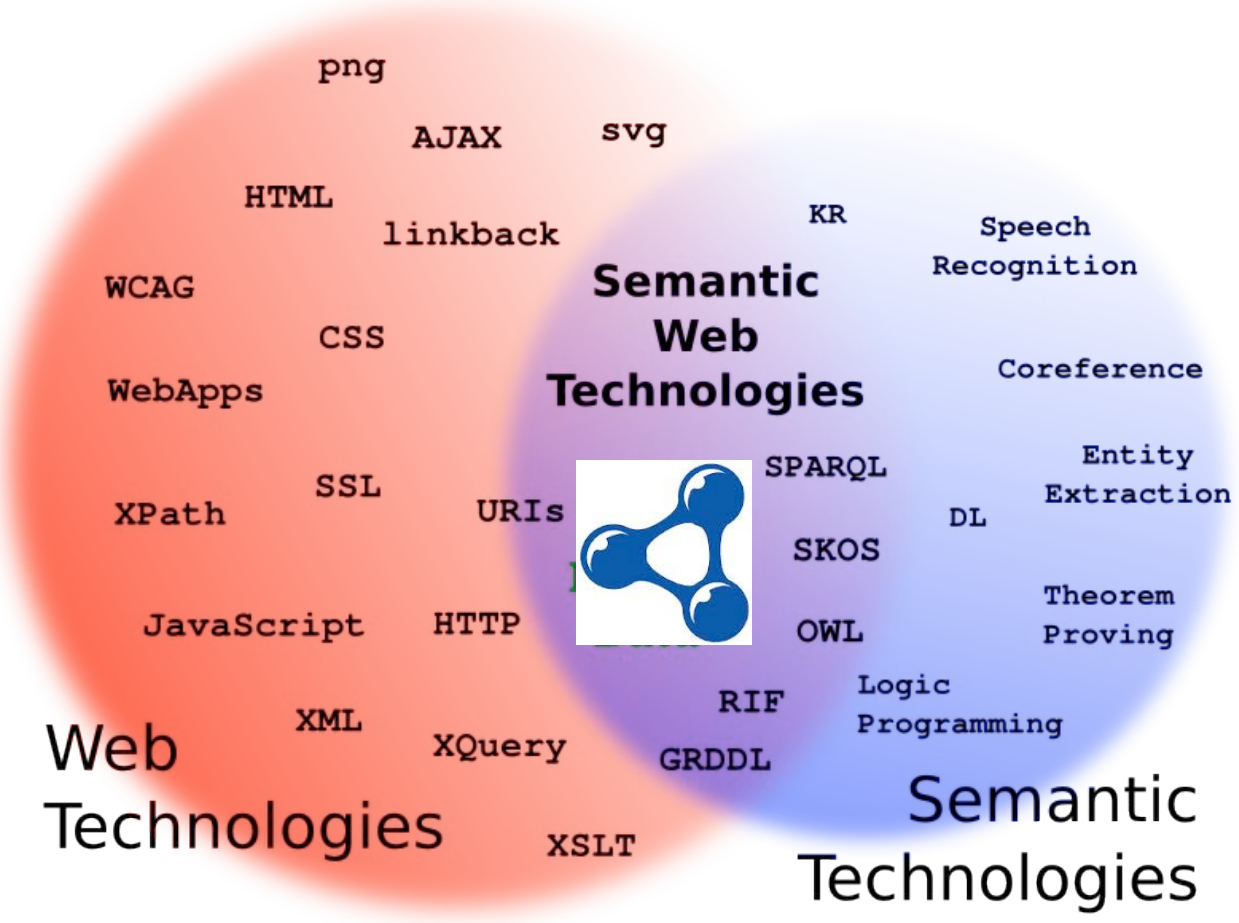
RDF (resource description framework)
For **computers** to exchange information

enables → **Linked Data** (Semantic Web)

"*The Semantic Web [Linked Data] provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.*"

Tim Berners-Lee, MIT Professor and Inventor of the World Wide Web (HTML and RDF protocols)

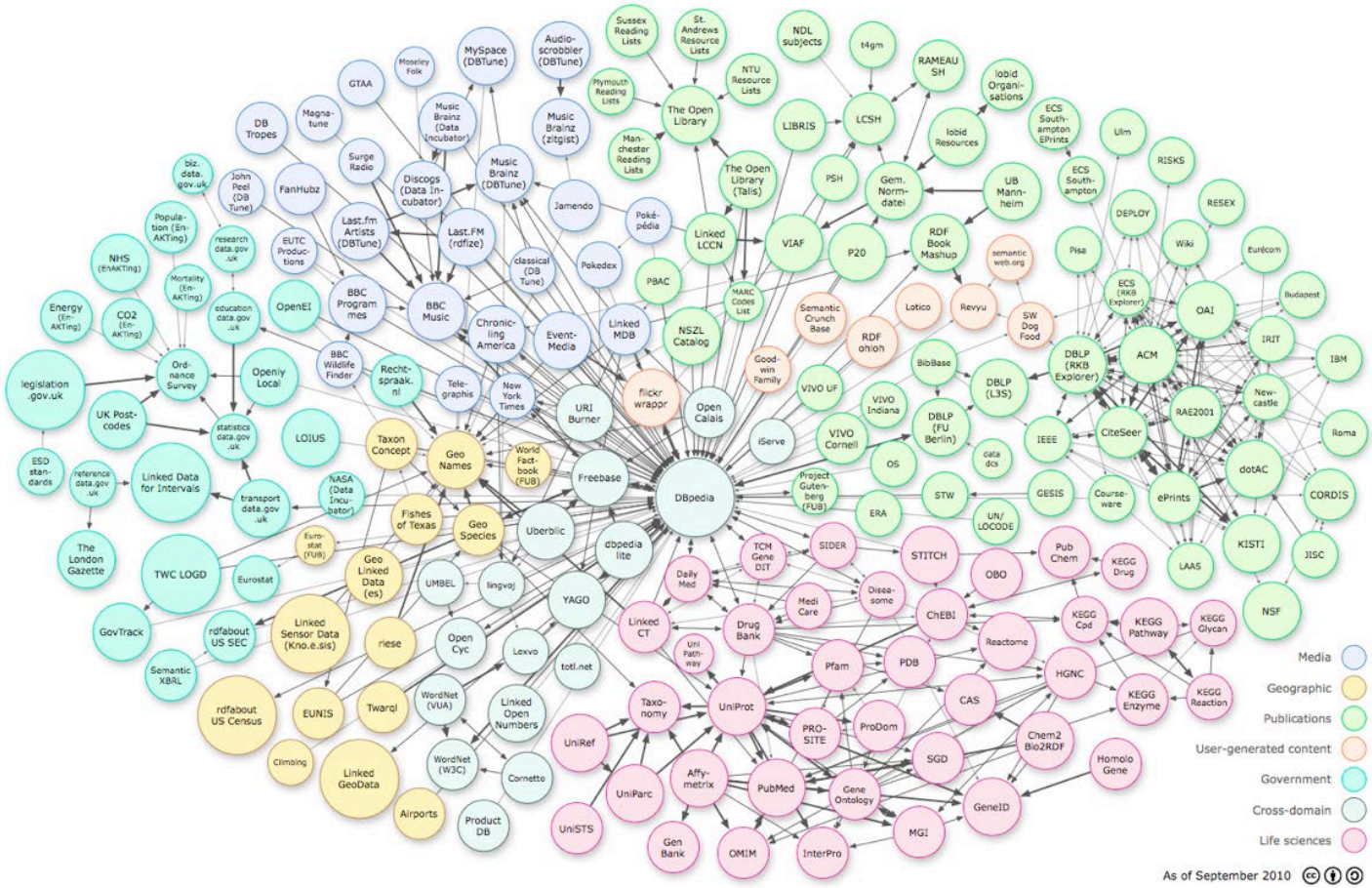
Linked Data:
Intersection of Semantic Technologies and Web Technologies



Linked Data and the Internet of Data

Linked Vitals

Linking media, geographic, publications, government, and life sciences …



Linked Data

This represents over 300 linked data sources and databases, comprising billions of data elements and millions of semantic links.

Each on of these circles represents a data source, which is semantically linked to other data sources, creating one virtual federated queryable web of data.

Wikipedia is one of the resources converted to Linked Data, and is called DBpedia (center circle).

Why not linked healthcare?

What is VISTA?

- Veterans Information Systems and Technology Architecture
- Information system of all VA care sites
- Foundation of other public health information systems
  - VA (VISTA): 1200+ care sites
  - DoD(CHCS): 900+ care sites
  - IHS(RPMS): 500+ care sites
  - NY State: 24 hospitals
- Most physicians in U.S. have used VISTA
- Open source
  - Deployed in many other settings in U.S. and internationally
  - Many developments by open source community

VISTA in the U.S.

VistA Architecture: Overview

Linked Vitals

All 150+ VistA applications are integrated within one shared multidimensional data engine.

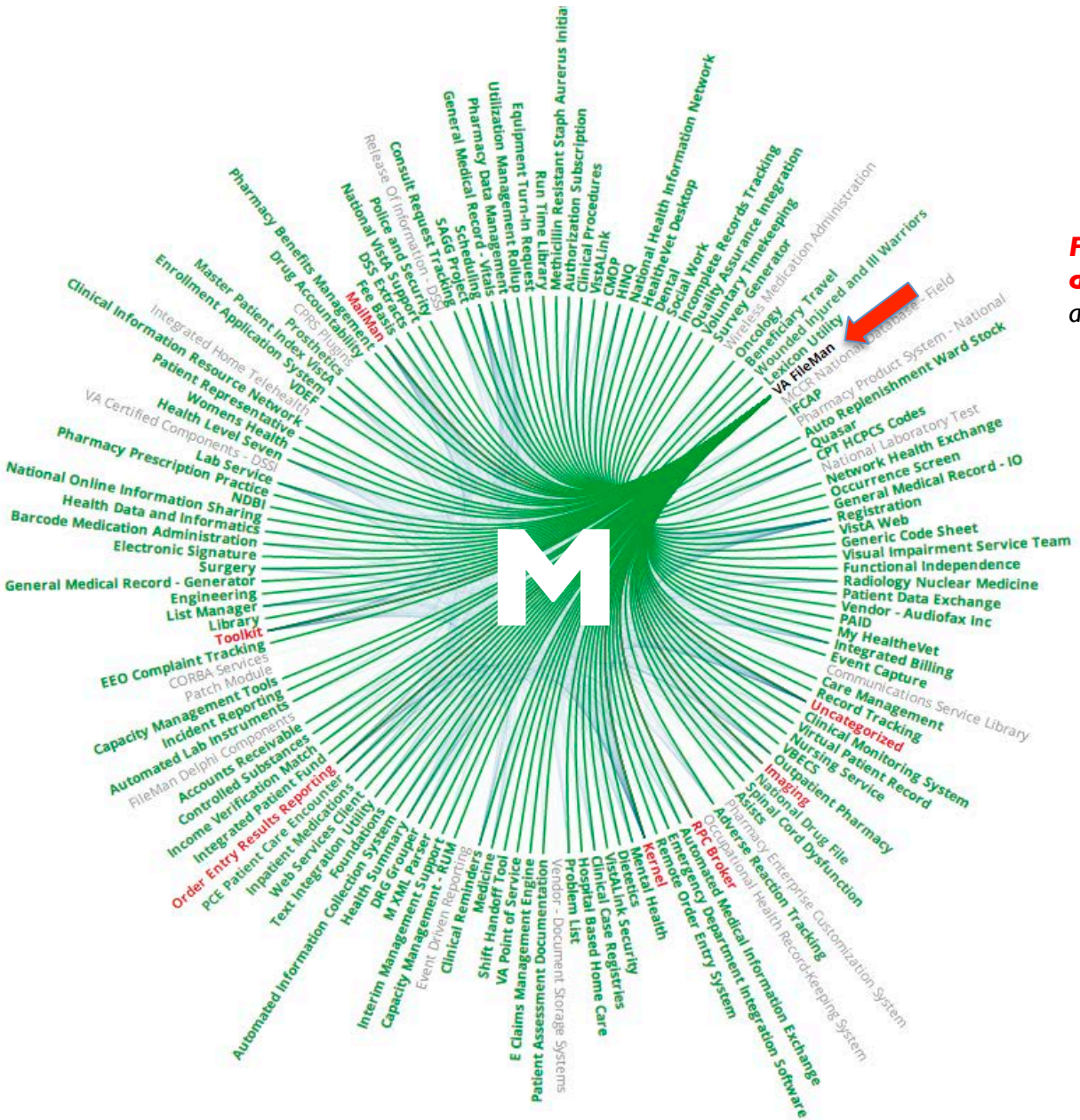
FMQL is the Fileman Query Language

Fileman is VistA's database management system

M is VistA's application-data integration engine.

VistA's Data Manager: Fileman

Linked Vitals



Fileman is the gateway to all VistA data. *All VistA applications use Fileman to access, structure, and query all data in VistA.*

VistA's integrated data engine. All 150+ applications are integrated to their data inside the multidimensional data engine, represented by the inner circle.

Reference:
code.osehra.org/vivian

Exposing VistA's native data model

Linked Vitals

Objective: Make machine-processable schemas, vocabularies, and datasets from VISTA

http://vista.caregraf.info/analytics

Challenge: Poor RDF/OWL model

Solution: Refine RDF/OWL model (example here)

Exposing VistA's native data model

Linked Vitals

FileMan Schema Browser

Populated Files

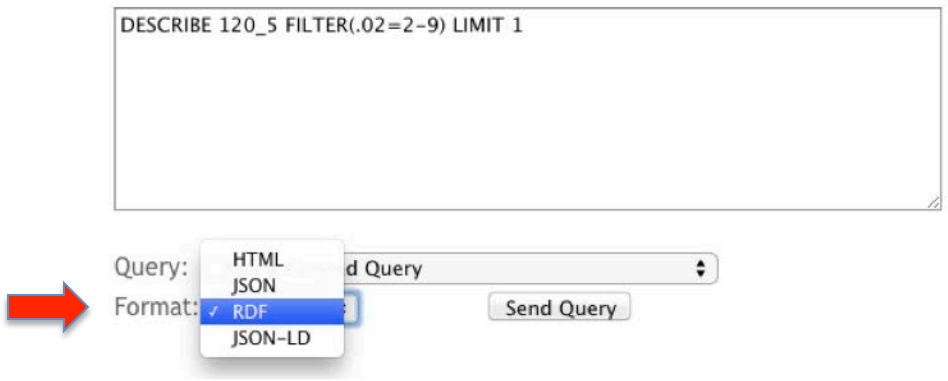
In this system 1245 out of 2356 have entries.

#	Name	Global	Count
.11	INDEX	^DD("IX",	644
.31	KEY	^DD("KEY",	64
.4	PRINT TEMPLATE	^DIPT(	1234
.401	SORT TEMPLATE	^DIBT(	773
.402	INPUT TEMPLATE	^DIE(	1519
.403	FORM	^DIST(.403,	144
.404	BLOCK	^DIST(.404,	519
.44	FOREIGN FORMAT	^DIST(.44,	11
.5	FUNCTION	^DD("FUNC",	169
.7	MUMPS OPERATING SYSTEM	^DD("OS",	7
.81	DATA TYPE	^DI(.81,	11
.84	DIALOG	^DI(.84,	2574
.85	LANGUAGE	^DI(.85,	11
1	FILE	^DIC(	2356
1.2	ALTERNATE EDITOR	^DIST(1.2,	4
2	PATIENT	^DPT(	40
3.07	PROGRAMMER MODE LOG	^%ZUA(3.07,	930

VistA's native data model is comprised of hierarchical files and subfiles, each which addresses a specific M Global storage.

Fileman Query Language (FMQL)

Linked Vitals



FMQL is the Fileman Query Language. This provides real-time web-based query access to the entirety of VistA's data.

It exposes the native hierarchical data model of Fileman in web standard forms including HTML, JSON, and RDF.

HTML:	hypertext markup language (visual document markup)
JSON:	javascript object notation (data serialization / packaging)
RDF:	resource description framework (linked data / semantics)
JSON-LD:	JSON with linked data capability

VistA Vitals: HTML output

Linked Vitals

FMQL query of VistA for vital signs with output in HTML.

```
DESCRIBE 120_5 FILTER(.02=2-9) LIMIT 2
```

FMQL Query Maker

Query a live VistA!

Help? Caregraf Support

Query: Pick a Canned Query

Format: HTML Send Query

1. GMRV VITAL MEASUREMENT > 2005-09-01T13:00:00Z (1)

date time vitals taken
2005-09-01T13:00:00Z
patient
PATIENT/JONES,CHRISTOPHER
vital type
GMRV VITAL TYPE/BLOOD PRESSURE (VA:4500634)
date time vitals entered
2005-12-28T13:48:44Z
hospital location
HOSPITAL LOCATION/4 SOUTH - MED
entered by
NEW PERSON/NOTHER,NADA (LOCAL)
rate
150/10
entered in error
true
error entered by
NEW PERSON/MANAGER,SYSTEM (LOCAL)
reason entered in error

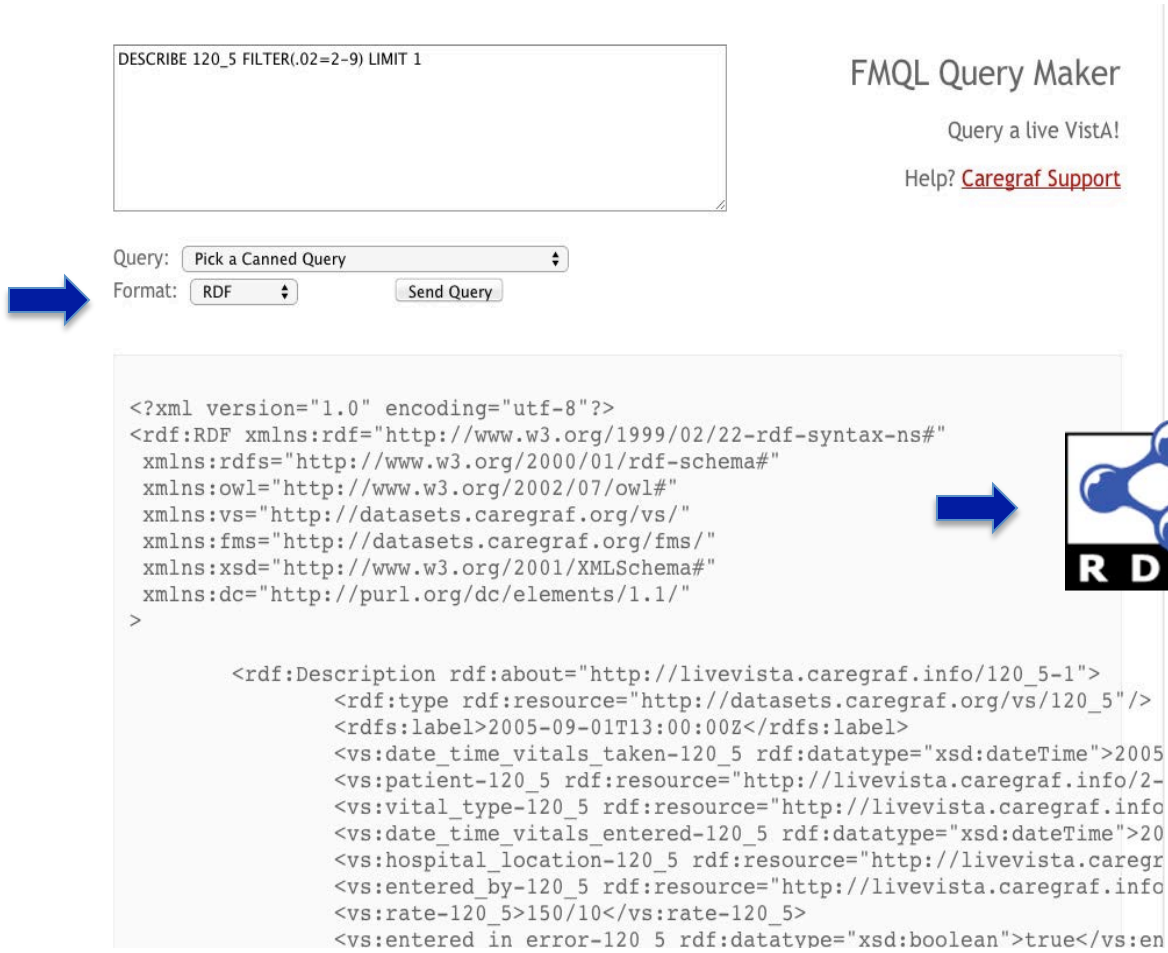
HTML

HTML output: Human-readable

VistA Vitals: RDF output

Linked Vitals

FMQL query of VistA for vital signs with output in RDF.

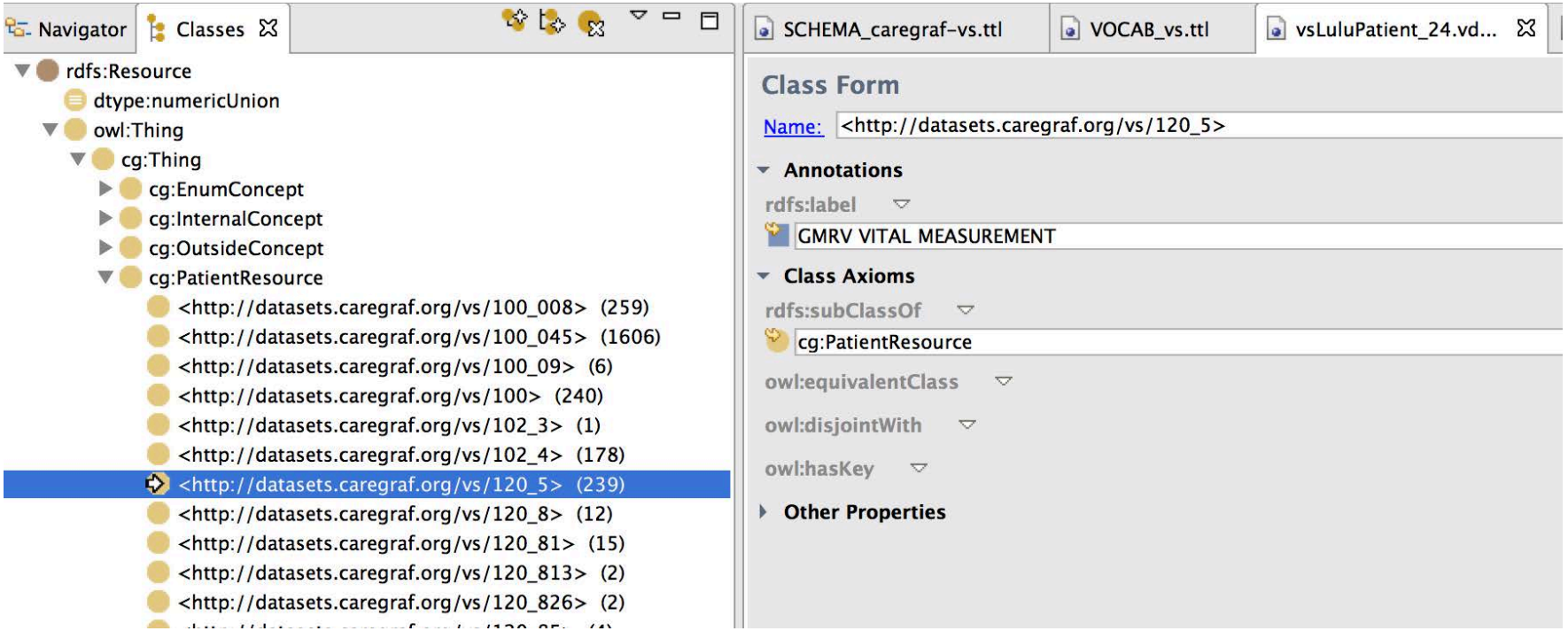


RDF

RDF output:
Machine readable

VistA Vitals in RDF

Linked Vitals

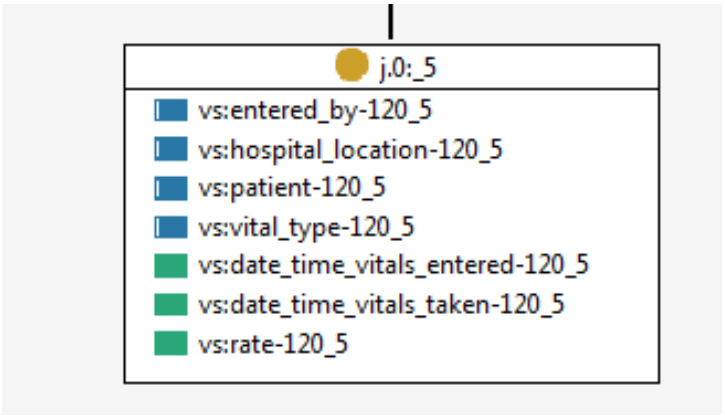


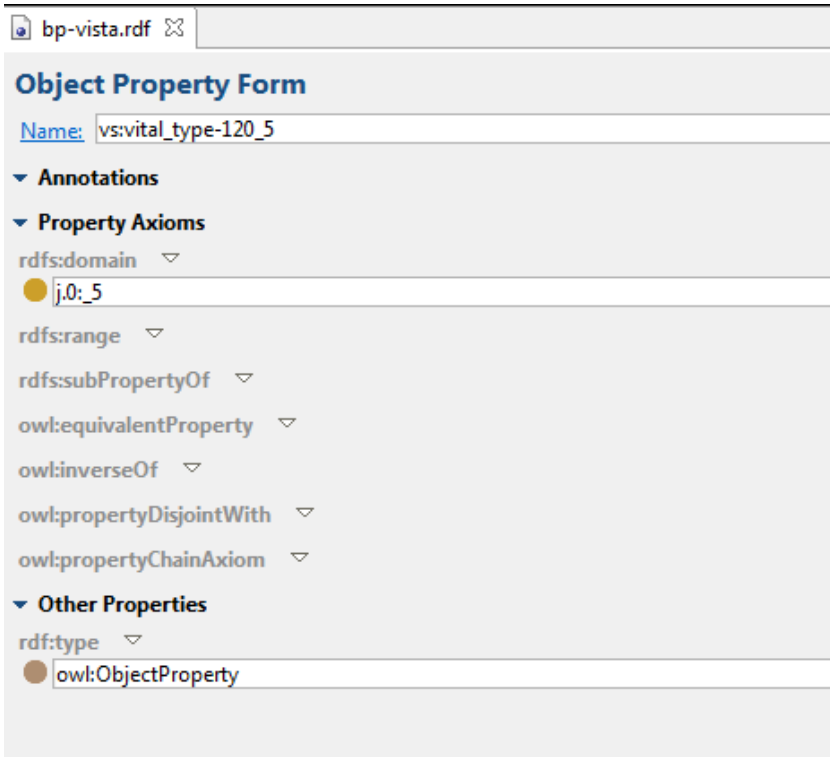
239 instances in the sample dataset

VistA Vitals in RDF

Linked Vitals

Exposing VistA's intrinsic data model for vitals

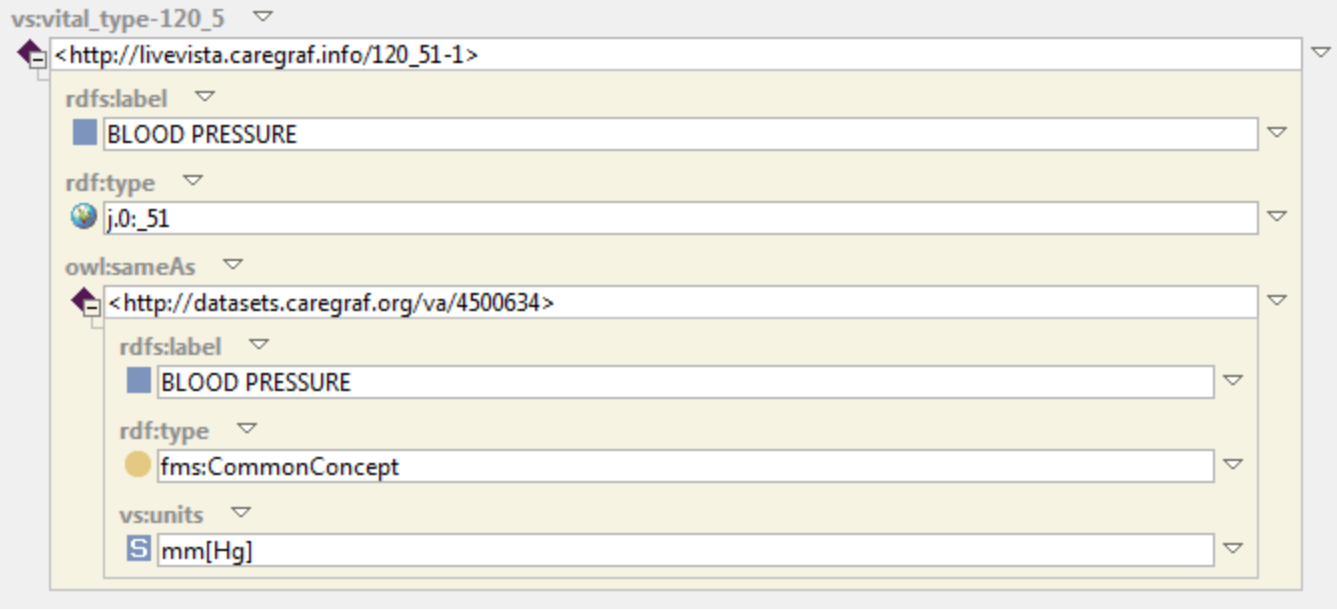




VistA Vitals in RDF

Linked Vitals

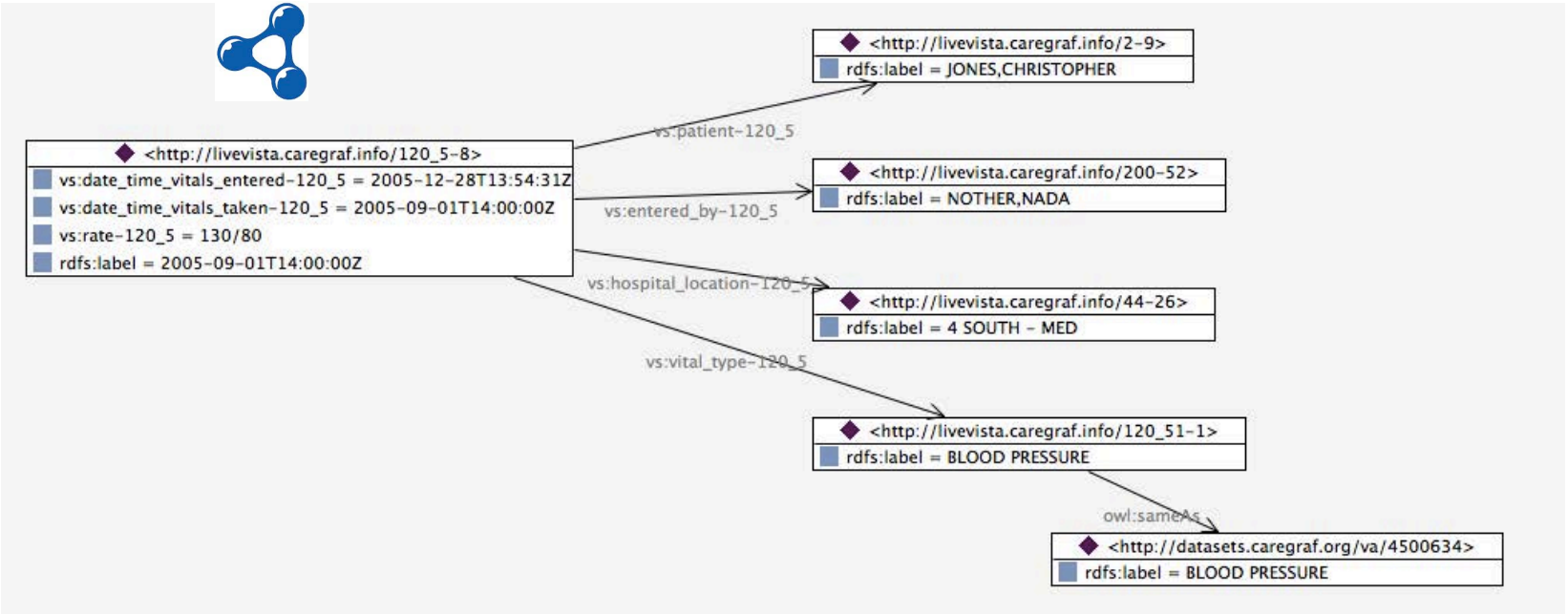
- Native VistA in RDF did not contain "common concepts" (so we created some).
- URIs are "messed up" (so we fixed them)



VistA Vitals in RDF

Linked Vitals

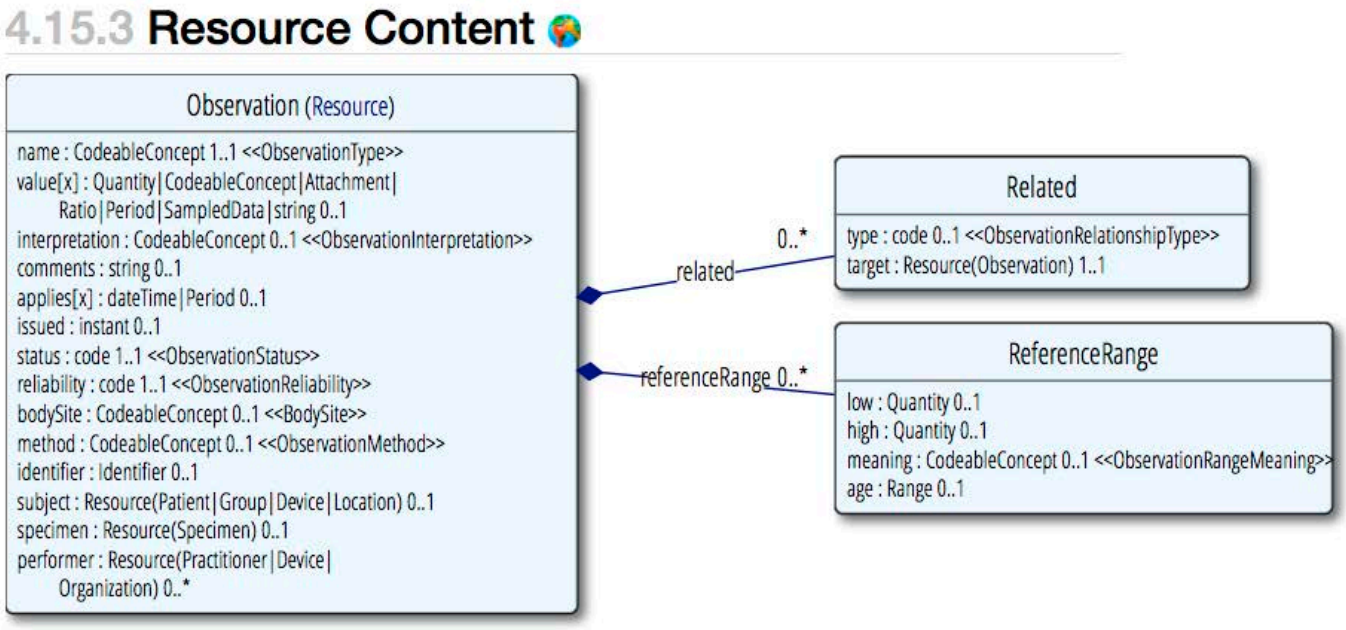
Exposing VistA's native data model for vitals



FHIR: Native model

Linked Vitals

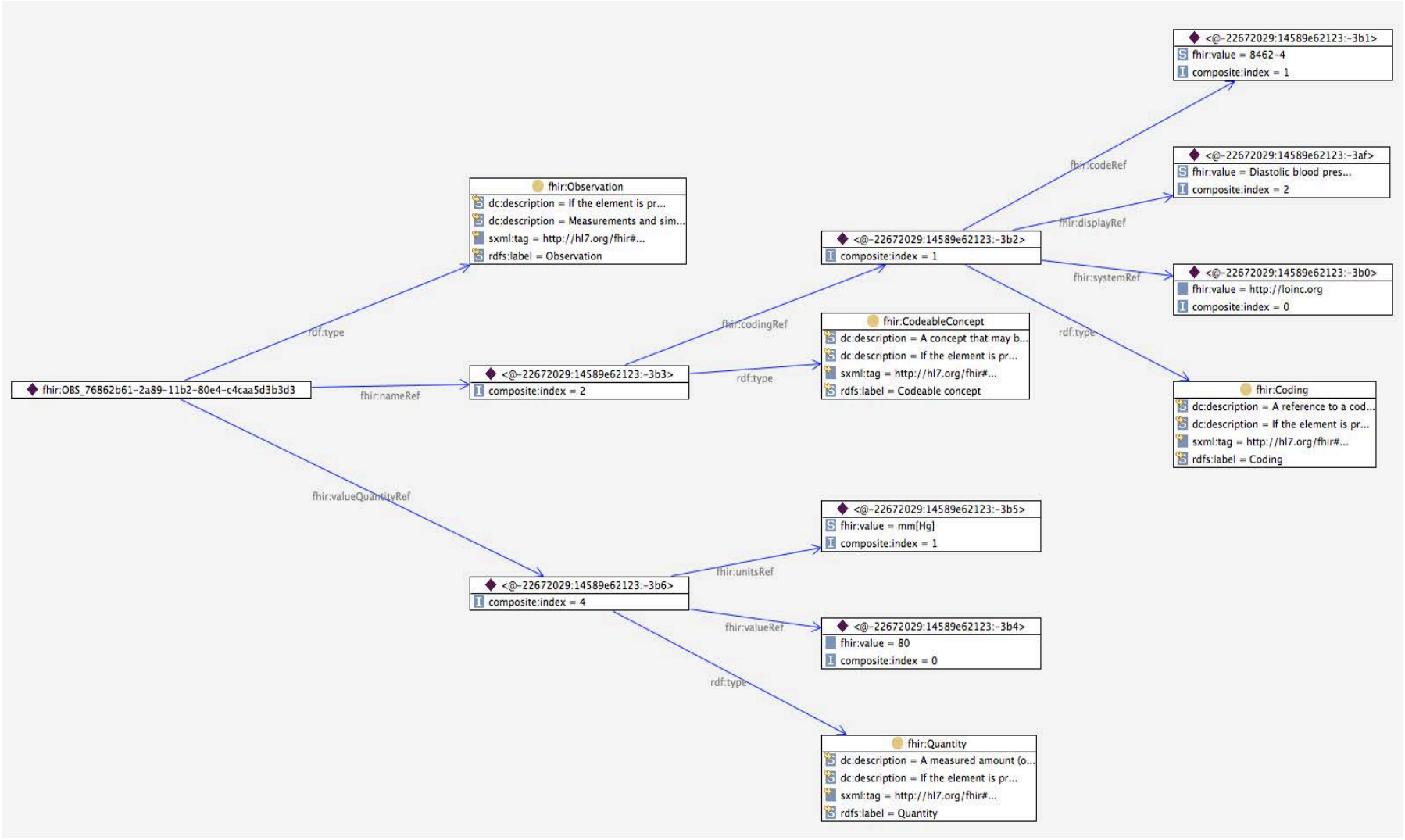
- FHIR - Observation
- XML model in XML Schema



FHIR Vitals in RDF

Linked Vitals

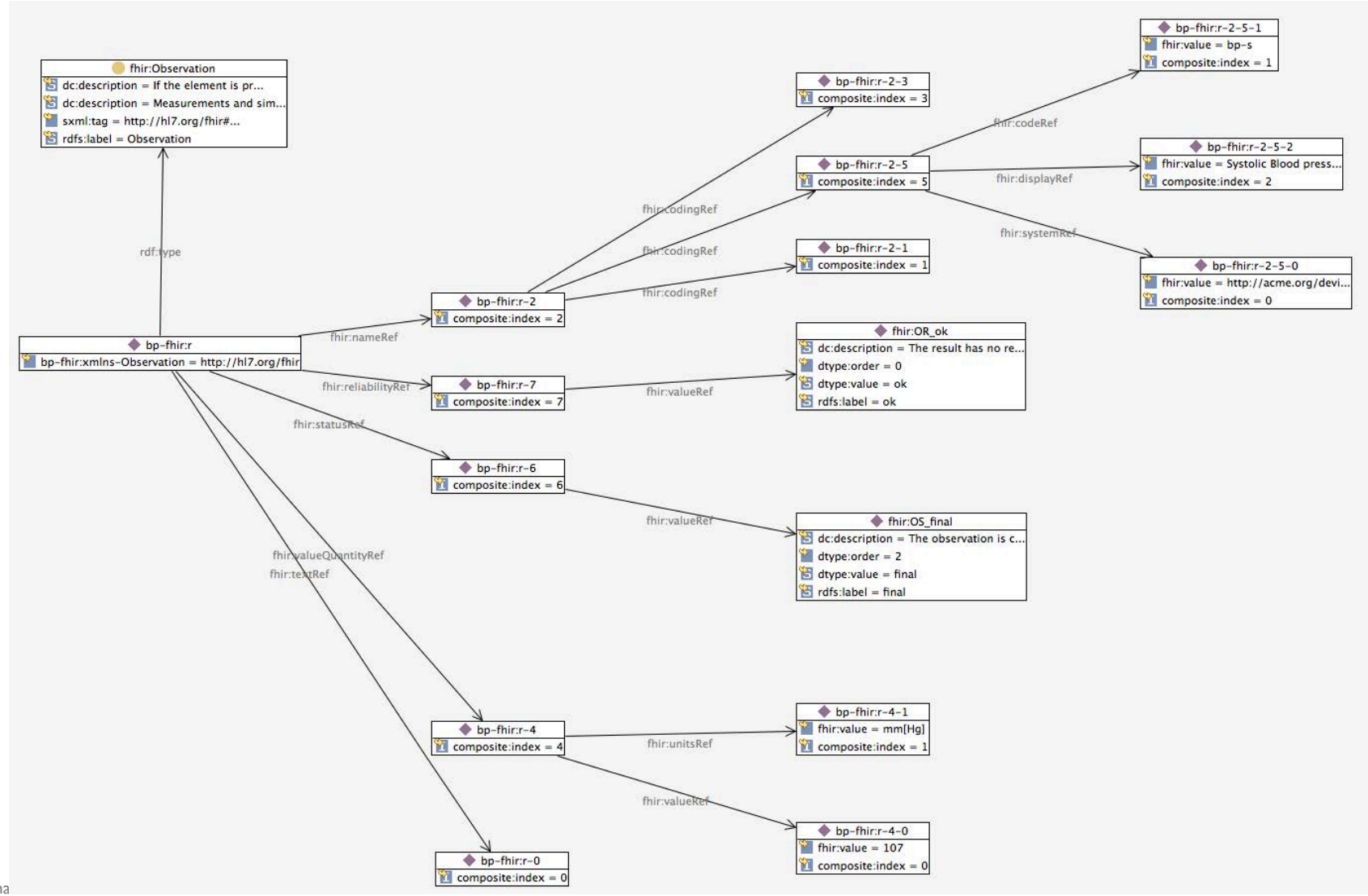
Automated transformation from FHIR XML Schema -> RDF



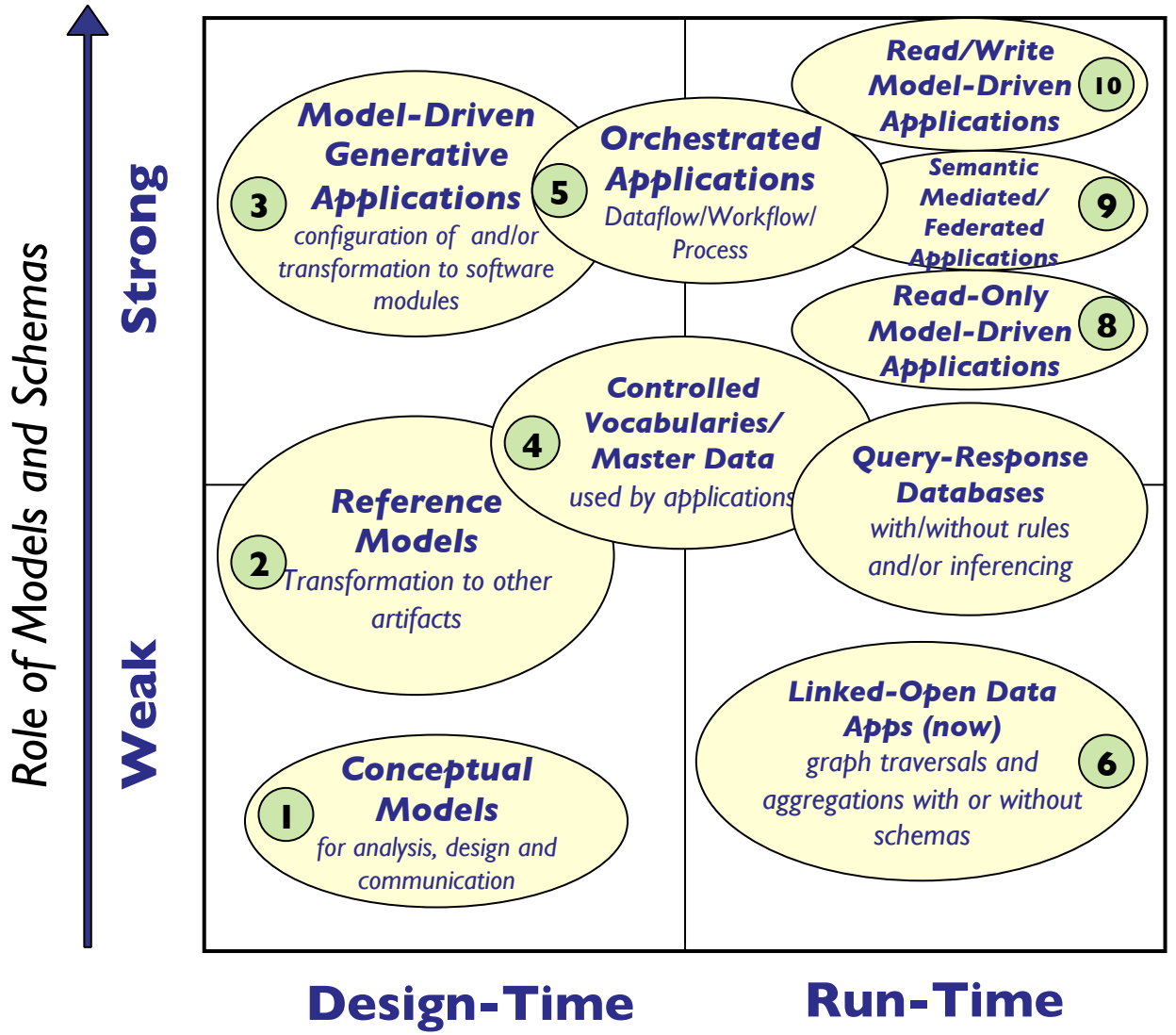
FHIR Vitals in RDF Enhanced

Linked Vitals

Enhanced bridge model for FHIR in RDF



Many ways of using "Ontologies"

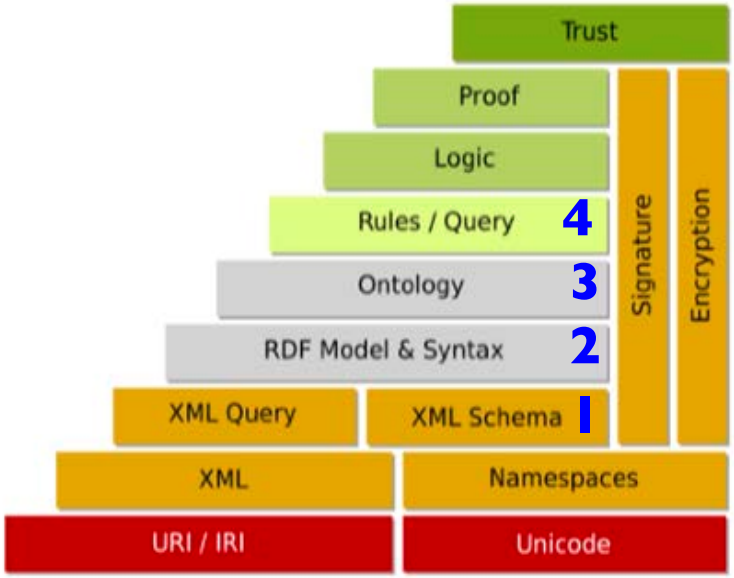


Our application

Context of Translation: RDF Technology Stack

Linked Vitals

- FHIR: XML Schema -> RDF **(1->2)**
- VISTA: Fileman model -> RDF **(2)**
- LOINC: CSV -> RDF **(2)**
- Create bridge model **(3)**
- Merge models and terminology: SPIN Map **(4)**



1. FHIR native model (XML Schema)
2. Translation (RDF Schema)
3. Bridge schema (OWL)
4. SPARQL rules (SPINMap)

SPIN: SPARQL Inferencing Notation

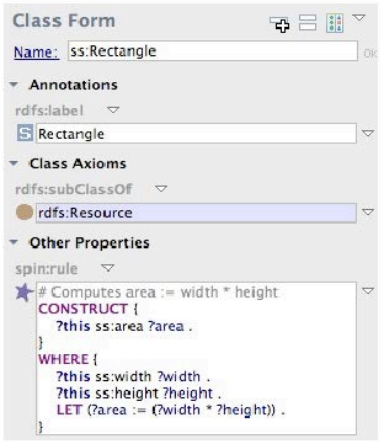
Linked Vitals

SPIN
SPARQL Inferencing Notation

- A W3C candidate standard
- A SPARQL Rules Language
- Builds on top of the SPARQL query language
- Enables exchangeable rules and transformations

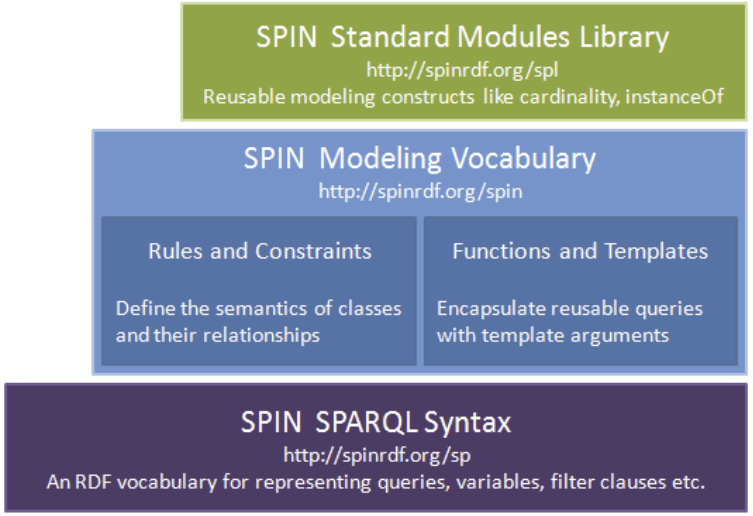
SPIN is a SPARQL Rules Language

The property spin:rule can be used to link a class with SPARQL CONSTRUCT queries that define inference rules for the members of the class



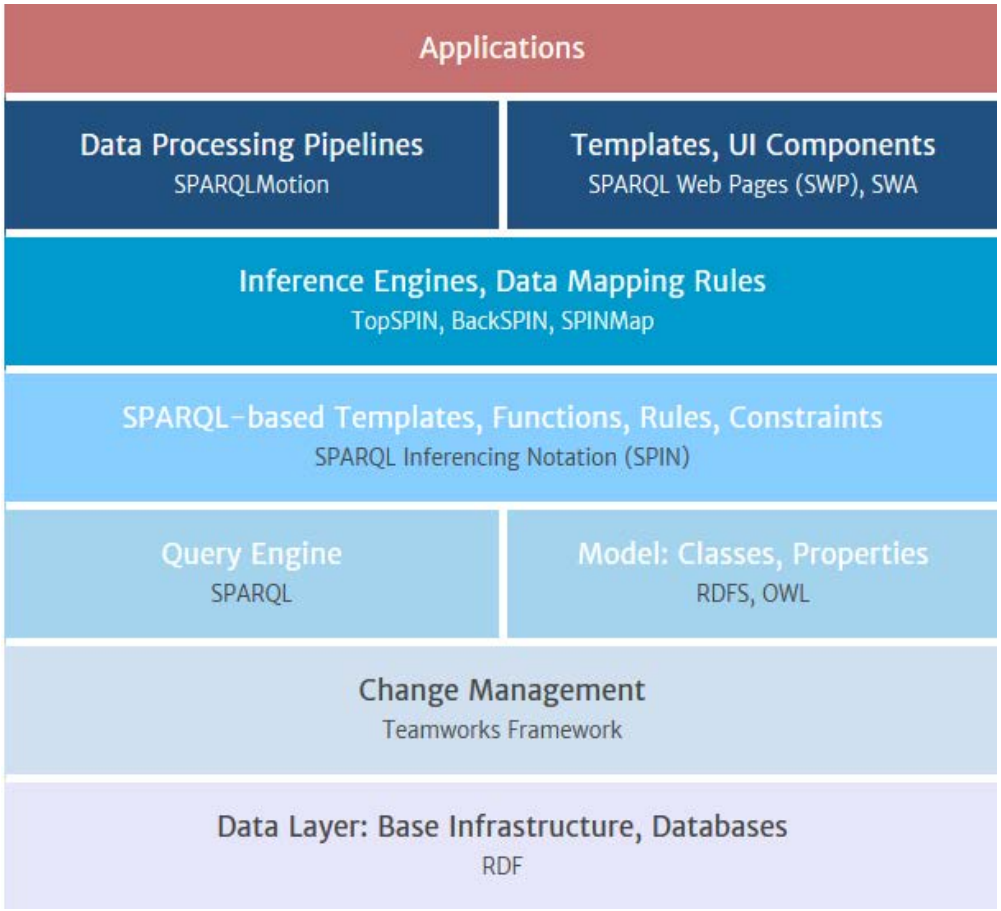
Benefits

- Natural object-oriented way of modeling
- SPARQL is very expressive
- Rules can be natively executed by SPARQL engines of the database
- Easy to combine with other SPARQL rule bases like OWL RL



SPINMap: Data transformation stack

Linked Vitals

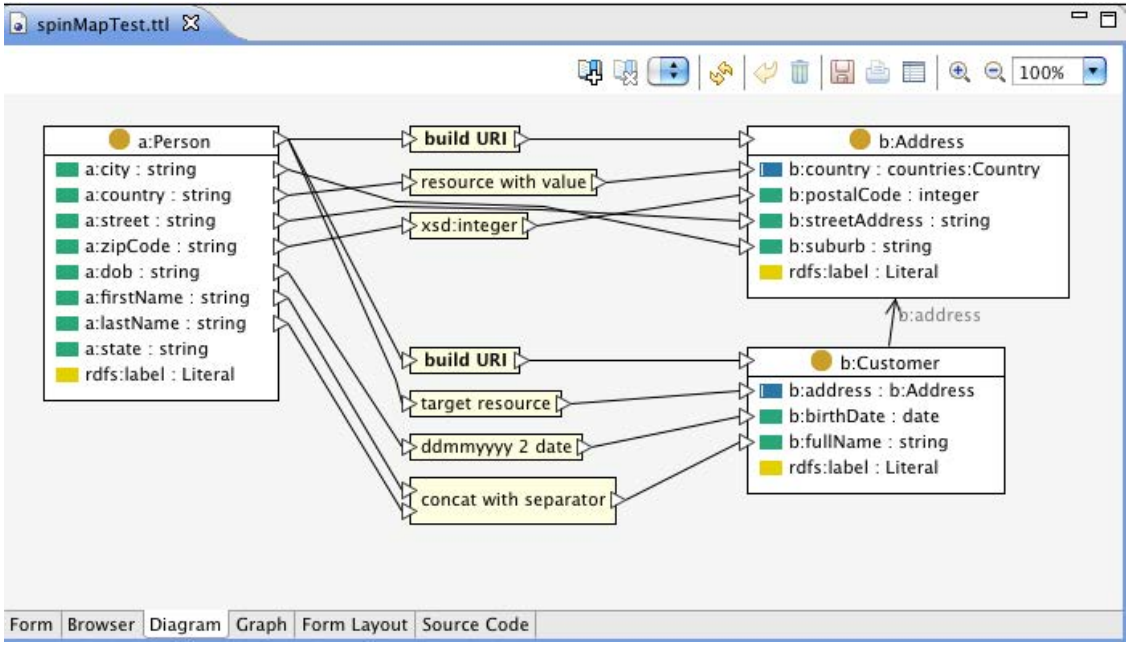


SPINMap: Data mapping rules engine

Linked Vitals

Motivation: Simplifies mappings between different models

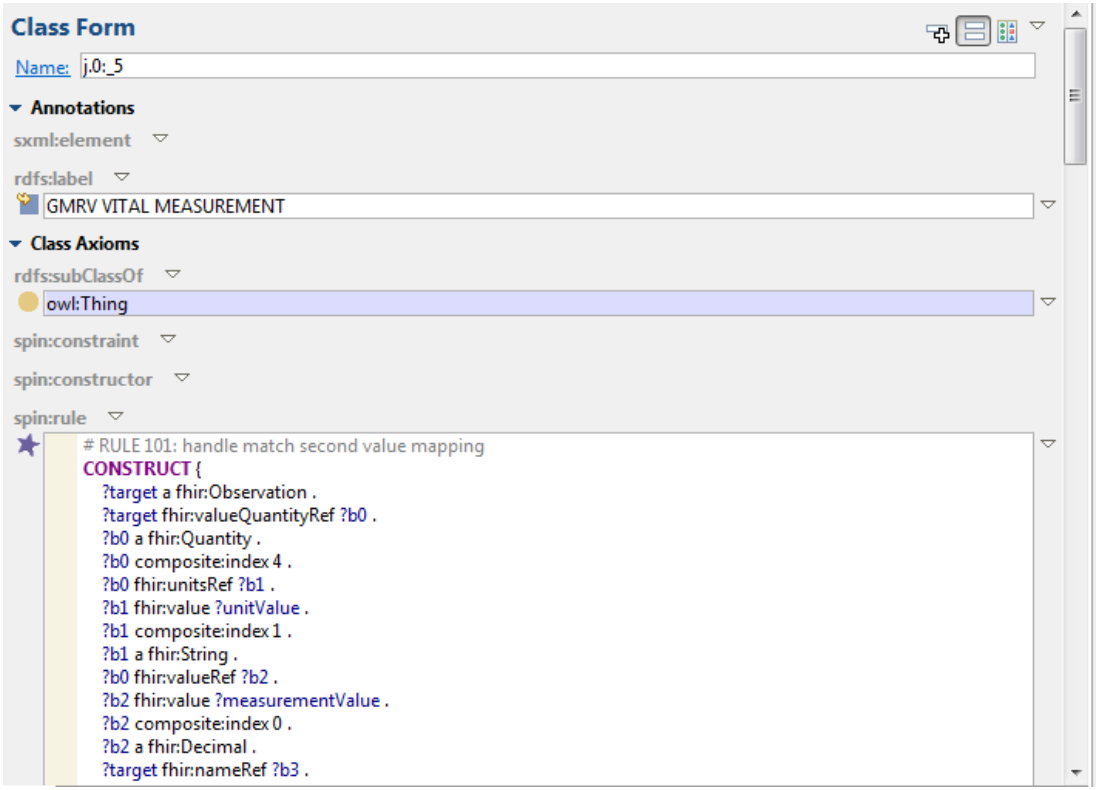
Key Features: Creates executable transformations



SPINMap: Field mapping with rules

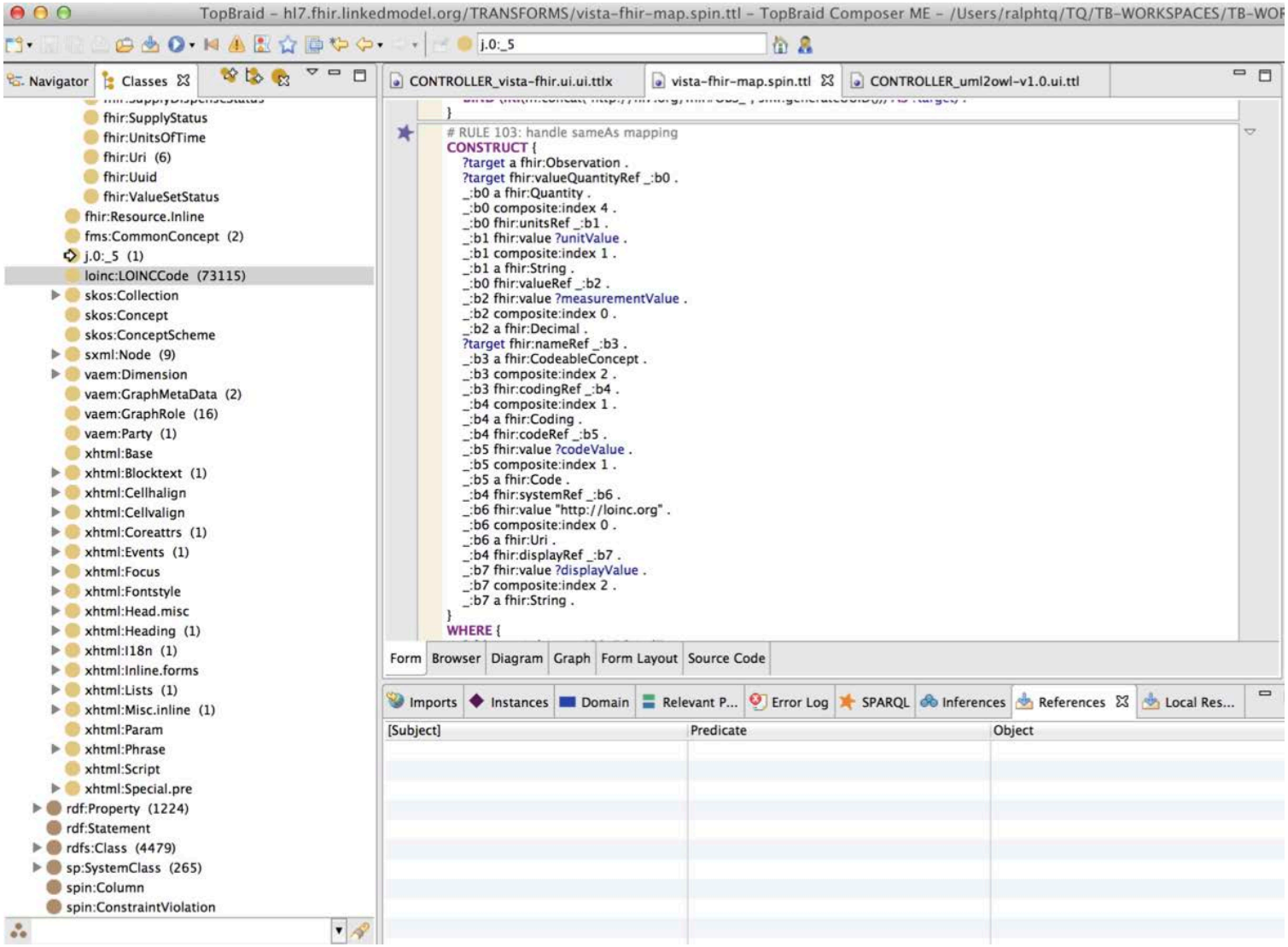
Linked Vitals

Easier to create deep nested structures in the target



SPINMap: Rules for LOINC terminology

Linked Vitals



SPINMap Output: Linked Vitals

Linked Vitals

VistA Patient Records of Interest

VistA Patient Record	Blood Pressure
http://livevista.caregraf.info/120_5-8	130/80

SPIN
SPARQL Inferencing Notation

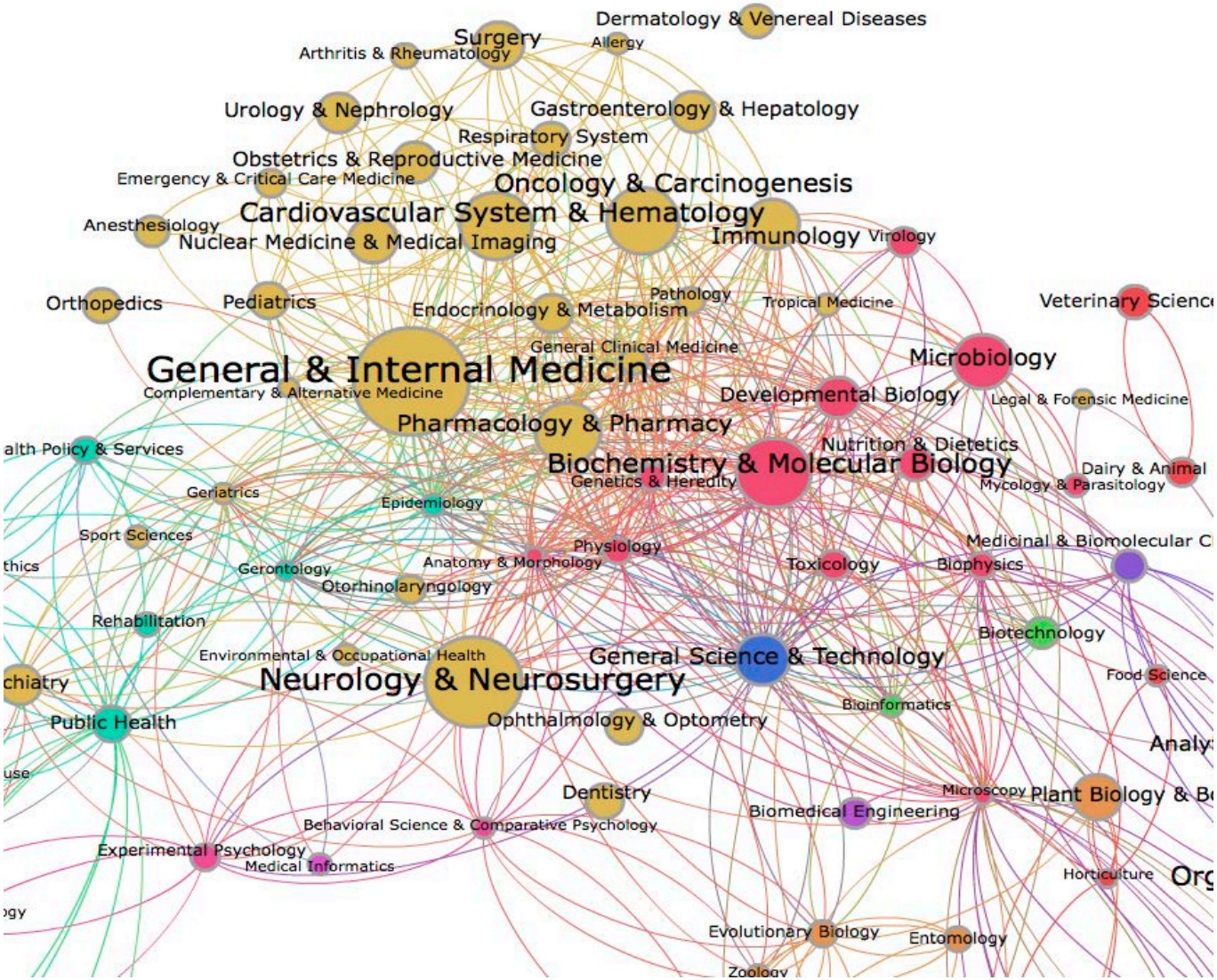
Same As

FHIR Patient Records

FHIR Patient Record	Blood Pressure	Value
http://hl7.org/fhir#OBS_cf8c9913-2aa1-11b2-80b1-f6177b67abba	Diastolic blood pressure	130/80
http://hl7.org/fhir#OBS_cf8c9912-2aa1-11b2-80b1-f6177b67abba	Systolic blood pressure	130
http://hl7.org/fhir#OBS_cf8c9909-2aa1-11b2-80b1-f6177b67abba	Diastolic blood pressure	80

Linked Vitals: A step towards Linked Health

Linked Vitals



Linked Vitals

Appendix

Review of Linked Data and its features as a sematic interoperability language

Linked Data:
Accommodates both Standardization and Innovation

Linked Vitals

Information Models:
An apparent conflict between standardization and innovation?

Standardization: need to remain static in order not to be disruptive for adopters.

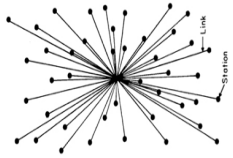
- Static, Brittle
- Centralized
- General (Lowest Common Denominator)
- Committee-driven
- Large, "all-or-nothing", disruptive updates

vs.

Innovation: requires continuous evolution of thousands of new information models.

- Adaptive / Evolutionary
- Decentralized
- Highly Specialized, "Best of breed"
- End-user / specialist – driven
- Small, continuous, low-impact updates

What are the options?



Centralized Model-rigid approach: For exchange of information to occur all models must remain fixed, and data must go through only one central 'broker' model. *Technologies that support this method are HL7 and XML.*

The current approach to healthcare data

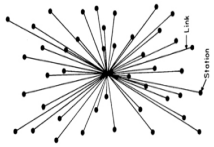
Decentralized Model-flexible approach: Multiple models peacefully co-exist and evolve, mediated by their ability to freely link to any model at all times. In this approach, all models are free to evolve AND are capable of resolving to a common standard model at all times. *The only technology that currently supports this approach is RDF (Linked Data).*

Linked Data supports both standardization and innovation

http://www.carlsterner.com/research/2009_resilience_and_decentralization.shtml

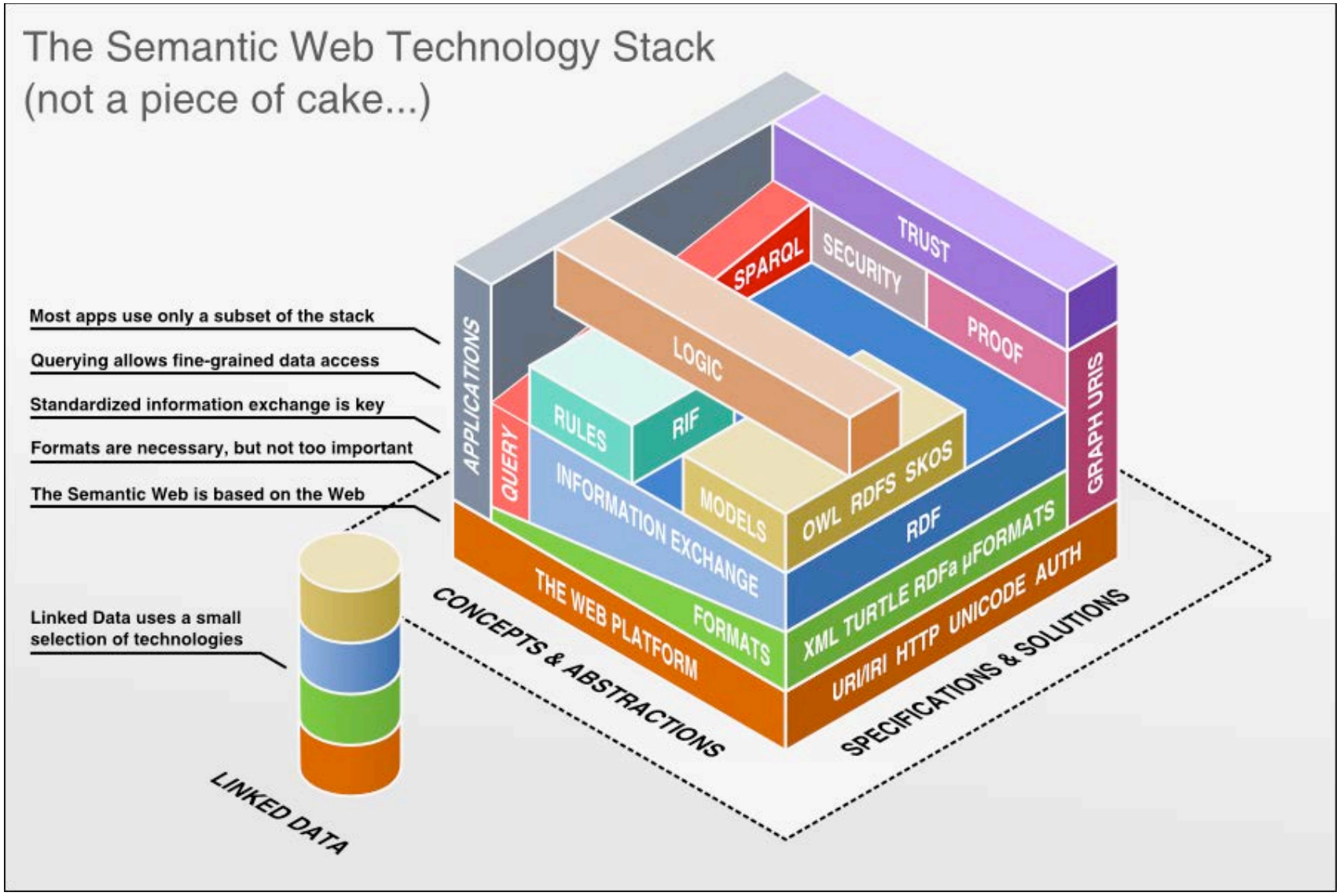
Data Integration:
Legacy vs. Linked Data Approach

Linked Vitals

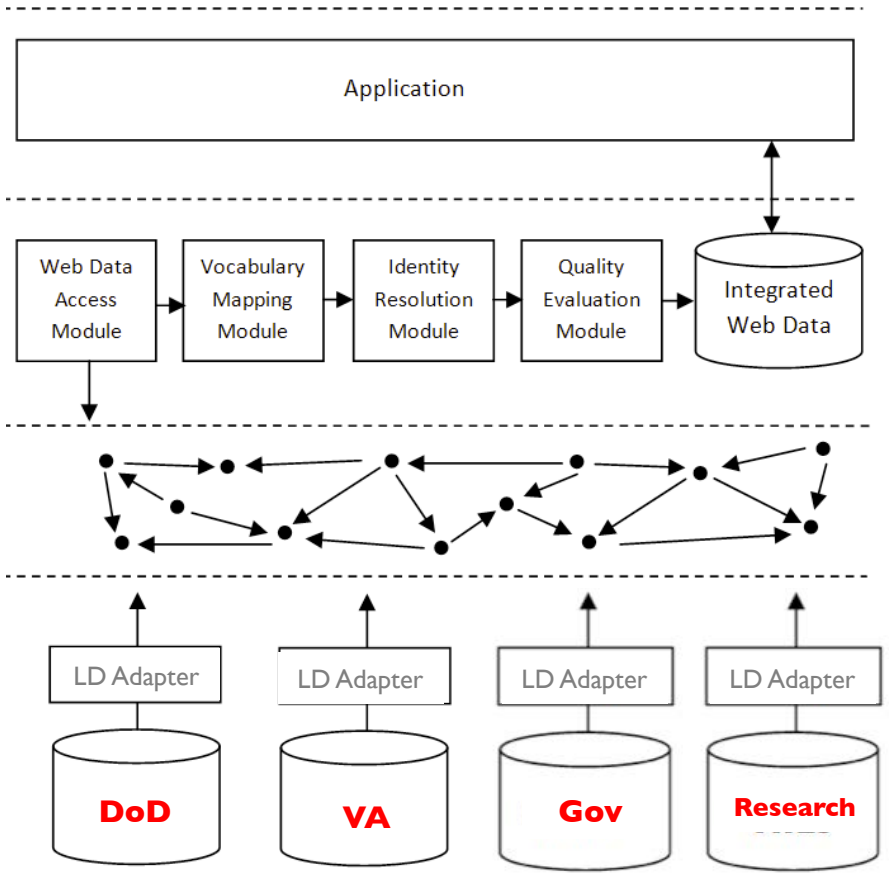


Architectural Issues	Legacy Data (HL7, XML)	Linked Data (RDF)
Function	**Serialization format**	**Data model**
Granularity	**Message-centric** (documents)	**Data-centric** (data elements)
Semantics	**Weak semantics. Extrinsic to the data.** Depends on an external data model.	**Strong semantics. Intrinsic to the data.**
Data model characteristics	**Model-rigid architecture.** Only the least common denominator of model unifies information; must remain unchanged to orchestrate. **Restrictive expression.**	**Model-flexible architecture.** All data models may independently evolve. **Maximizes expressivity.**
Data model compatibility	**No model diversity permitted** Requires one-size-fits-all mega-model	**Multiple models peacefully coexist** Data model agnostic
Data model evolution	**Costly and difficult to evolve models.** Due to model-rigid architecture.	**Cheaper and easier to evolve models.** Due to model-flexible architecture.
Data access method	Downloading + Aggregating	**Linking + Federating**
Scalability: incremental effort required to add new data sources	**Common model** must be updated	**Individual models** can be independently and incrementally semantically linked.

Linked Data: Technology Stack



Linked Data: General Approach to Semantic Integration



Integrated Systems

Transactional Systems

Application

Integrated Data

Linked Data

LD Adapter

Source Data

End