Why RDF as a Universal Healthcare Exchange Language?

David Booth, Ph.D.
Hawaii Resource Group
david@dbooth.org

Semantic Technology and Business Conference
21-Aug-2014

See latest version:

http://yosemiteproject.org/2015/webinars/why-rdf/

Outline

- Why RDF (in general)?
- Why RDF as a universal healthcare exchange language?

What is RDF?

- "Resource Description Framework"
  - *But think "Reusable Data Framework"*
- Language for representing information
- International standard by W3C
- Mature – 10+ years
- Used in many domains, including biomedical and pharma

RDF graph

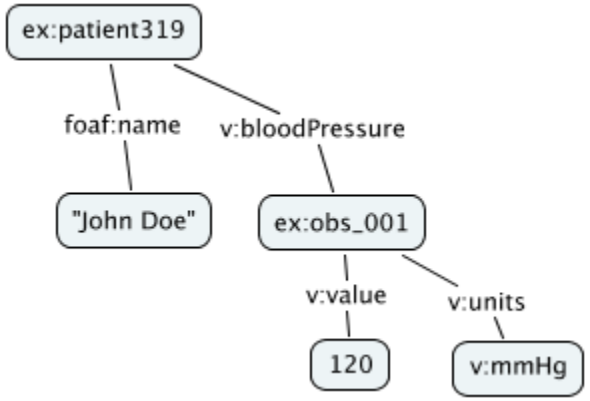
English assertions:

Patient319 has name "John Doe".
Patient319 has systolic blood pressure observation Obs_001.
Obs_001 value was 120.
Obs_001 units was mmHg.

RDF* assertions ("triples"):

ex:patient319 foaf:name "John Doe" .
ex:patient319 v:systolicBP ex:obs_001 .
ex:obs_001 v:value 120 .
ex:obs_001 v:units v:mmHg .

RDF graph:



*Namespace definitions omitted

Why RDF (in general)?

#5: RDF is self describing

- RDF uses URIs as identifiers

#4: RDF is easy to map from other data representations

- RDF data is made of assertions

#3: RDF captures information – not syntax

- RDF is format independent

#2: Multiple data models and vocabularies can be easily combined and interrelated

- RDF is multi-schema friendly

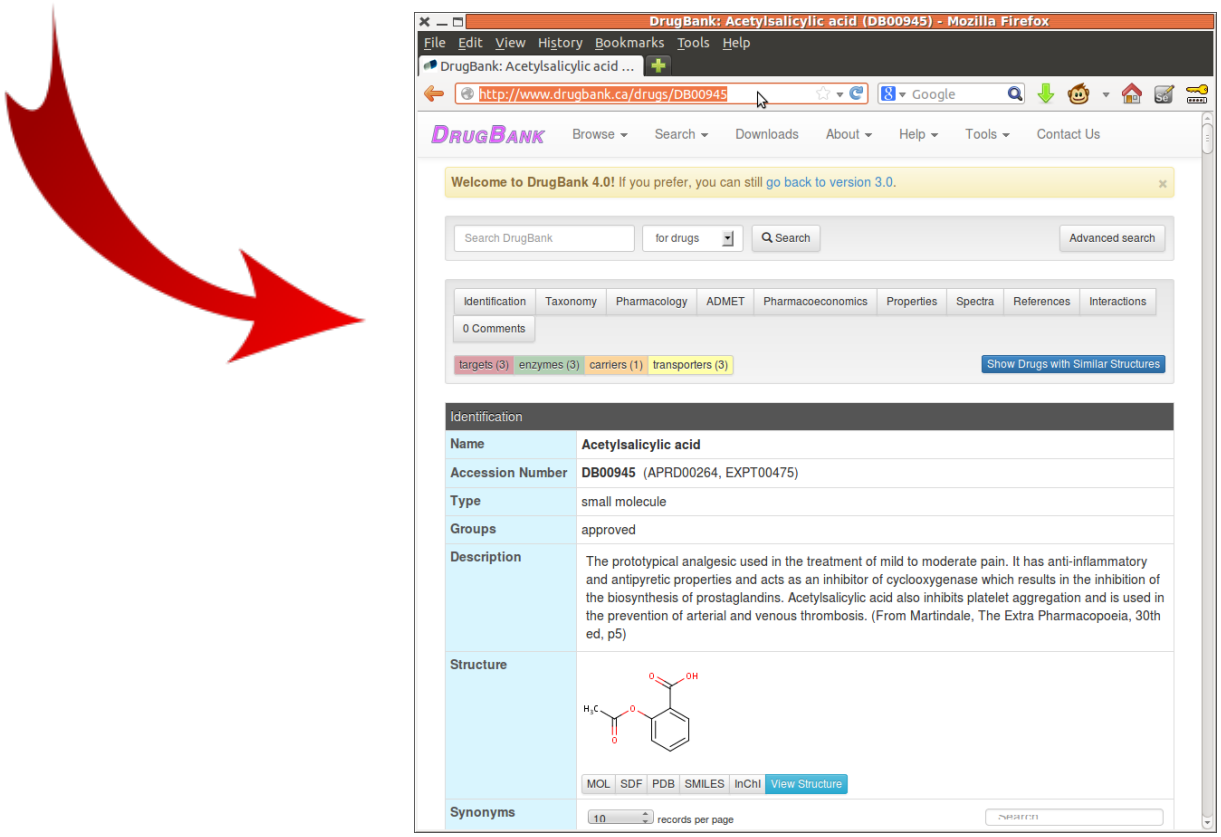
#1: RDF enables smarter data use and automated data translation

- RDF enables inference

#5: RDF is self describing

- Uses URIs as identifiers

http://www.drugbank.ca/drugs/DB00945



Why is this important?

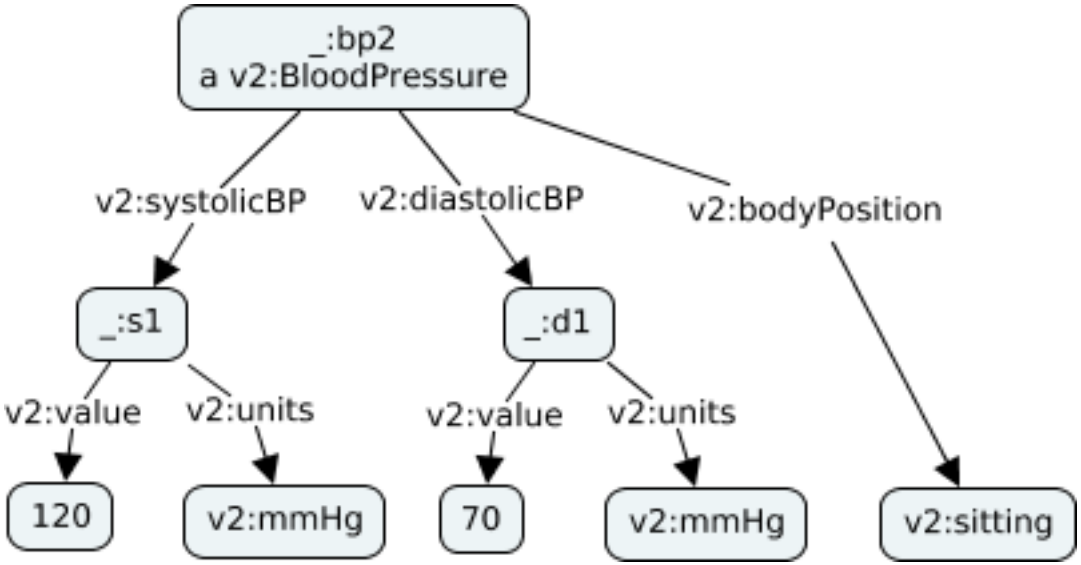
- Terms, data models, vocabularies, etc., can be linked to definitions
- Definition can be found by any party
  - Reduces ambiguity
- Aids in bootstrapping new terms toward standardization

Supports standards and diversity

#4: RDF is easy to map from other data representations

- RDF is made up of lots of small, atomic statements, called *assertions* or *triples*
- Easy to represent any data
- Easy to incorporate any data model
  - Hierarchical, relational, graph, etc.

Hierarchical data model in RDF



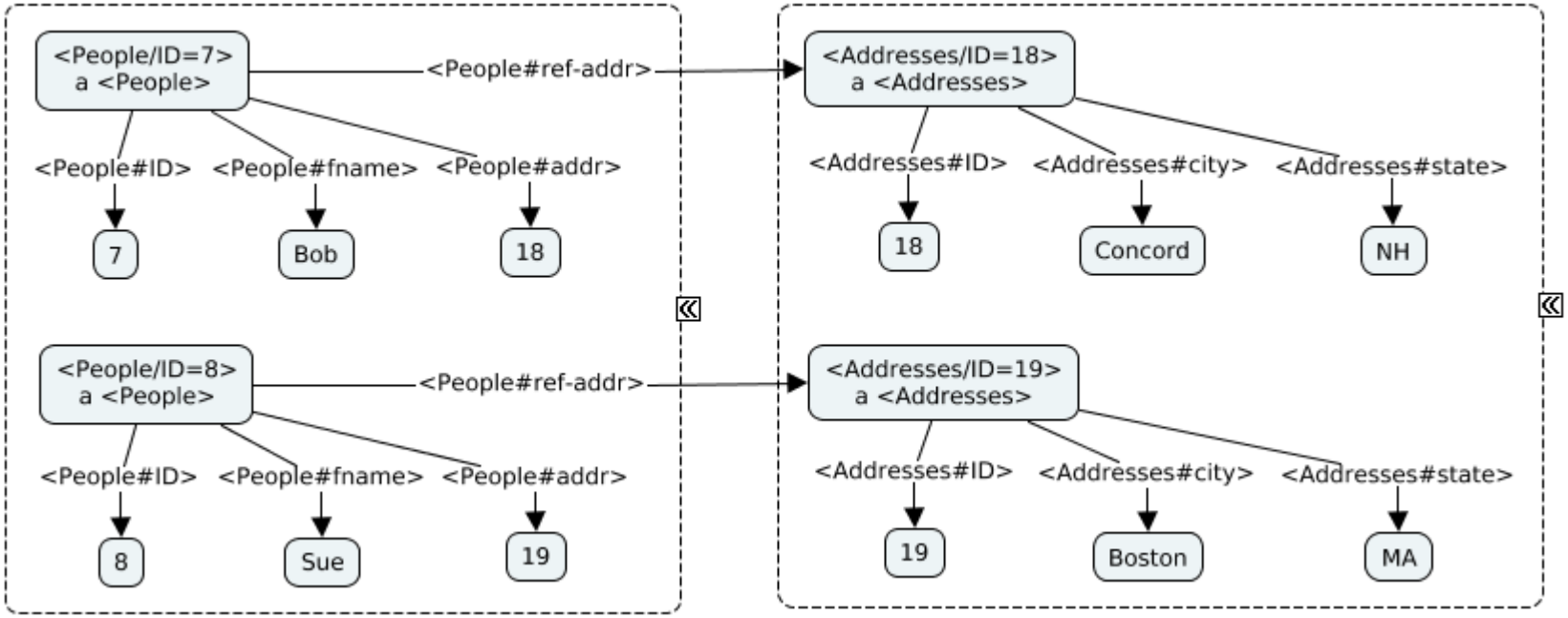
Relational data model in RDF

People

ID	fname	addr
7	Bob	18
8	Sue	19

Addresses

ID	City	State
18	Concord	NH
19	Boston	MA



See W3C Direct Mapping of Relational Data to RDF:
http://www.w3.org/TR/rdb-direct-mapping/

Why does this matter?

- Easy to map any data format to RDF
  - E.g., XML, JSON, CSV, SQL tables, etc.

#3: RDF captures information – not syntax

- RDF is format independent
- There are multiple RDF syntaxes: Turtle, N-Triples, JSON-LD, RDF/XML, etc.
- The same information can be written in different formats
- Any data format can be mapped to RDF

Different source formats, same RDF

HL7 v2.x

```
OBX|1|CE|3727-0^BPsystolic,
sitting||120||mmHg|
```

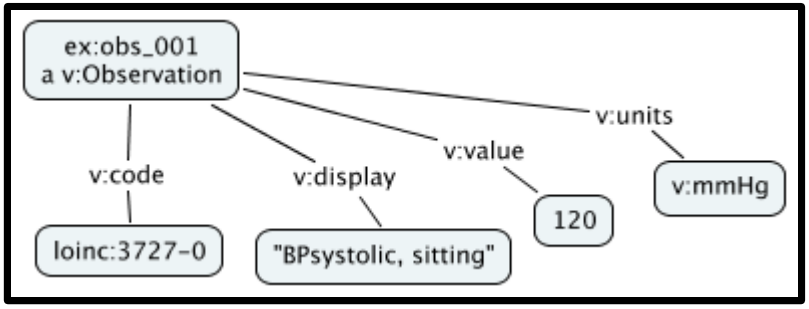
Maps to

FHIR

```
<Observation
      xmlns="http://hl7.org/fhir">
   <system value="http://loinc.org"/>
   <code value="3727-0"/>
   <display value="BPsystolic, sitting"/>
   <value value="120"/>
   <units value="mmHg"/>
</Observation>
```

Maps to

RDF graph



Why does this matter?

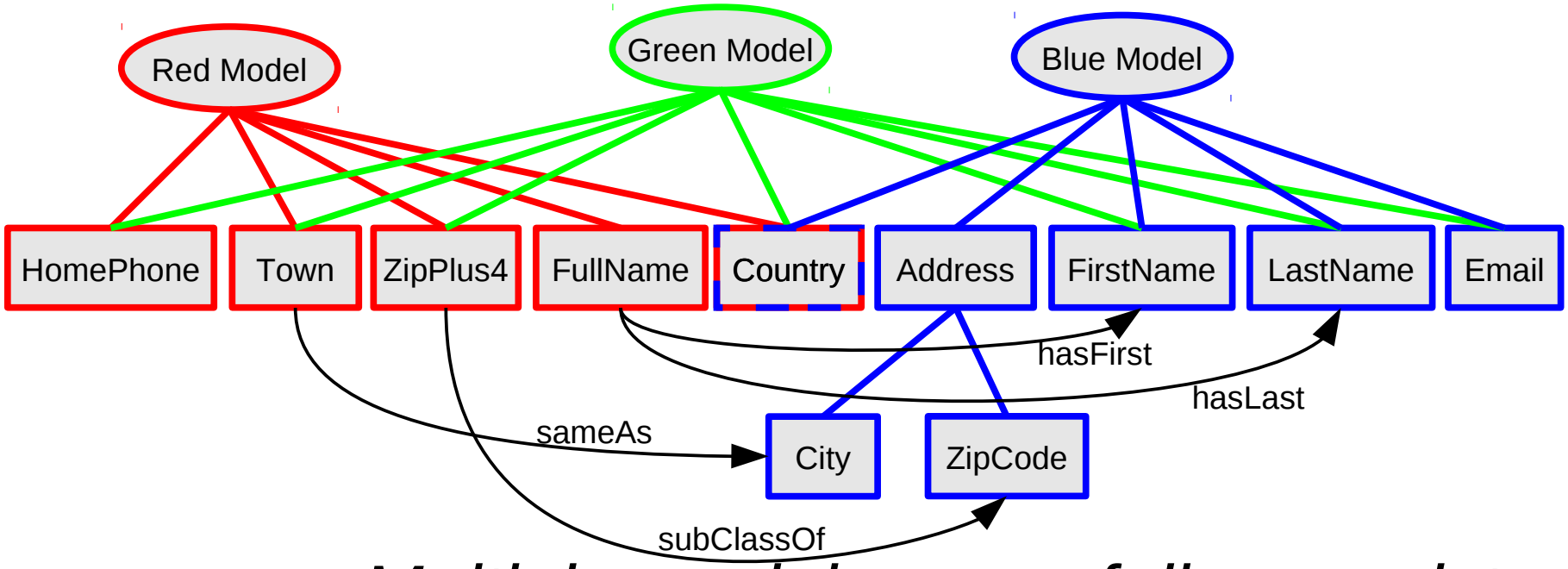
- Emphasis is on the meaning (where it should be)
- RDF acts as a common information representation
- Helps avoid the bike shed effect, a/k/a Parkinson's Law of Triviality
  - Syntax is irrelevant

#2: Multiple data models and vocabularies can be easily combined and interrelated

- RDF is multi-schema friendly*
- Multiple data models/schemas and vocabularies can peacefully co-exist, semantically connected

*A/k/a schema-promiscuous, schema-flexible, schema-less, etc.

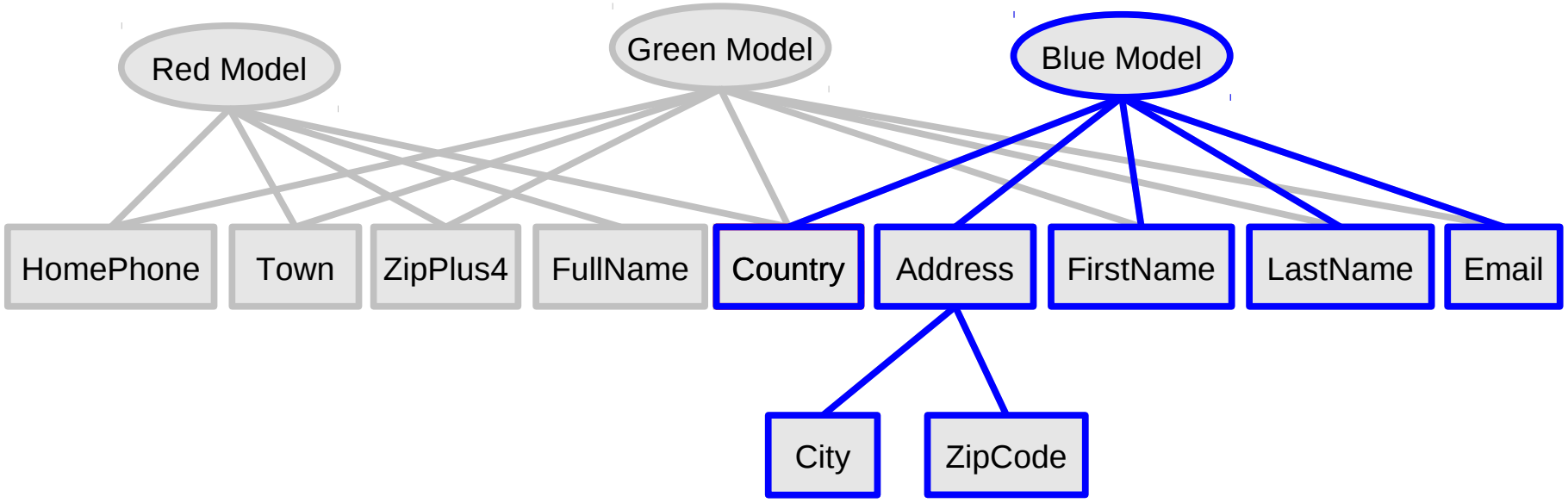
Multi-schema friendly



Multiple models peacefully co-exist

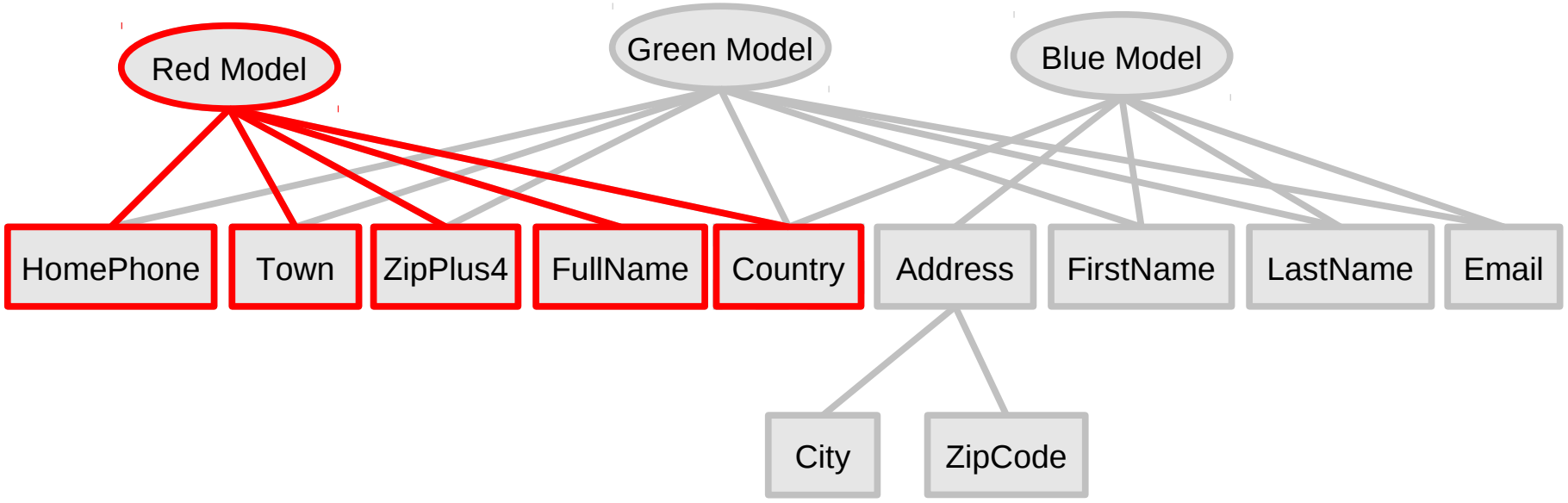
Multi-schema friendly

- Blue app sees Blue model



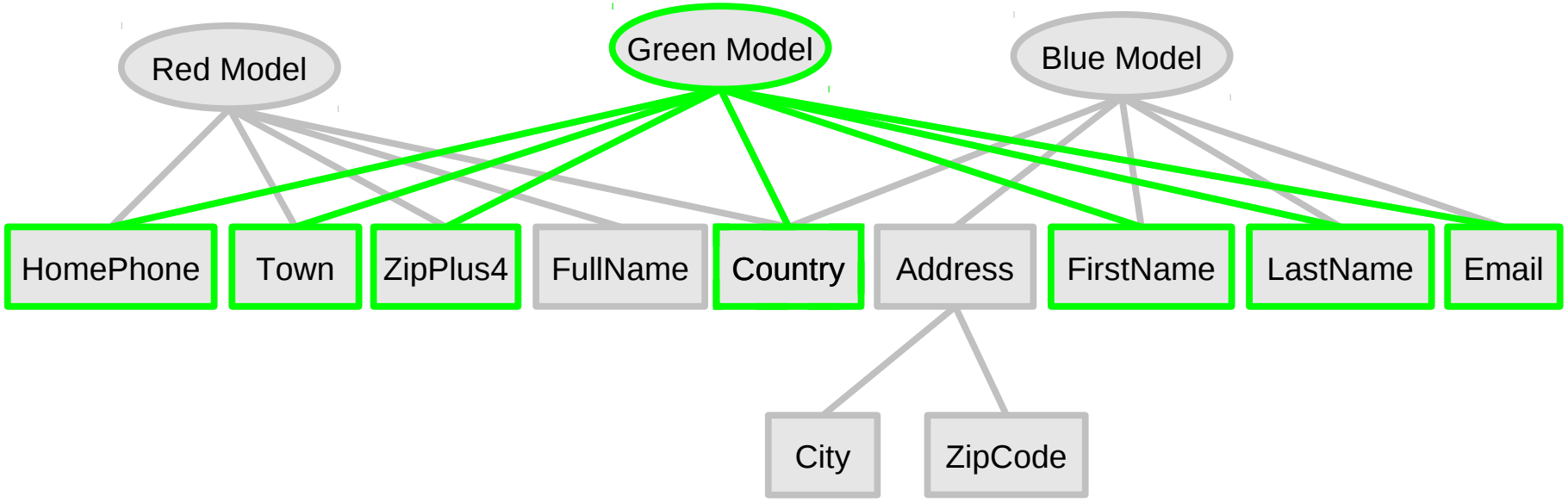
Multi-schema friendly

- Red app sees Red model



Multi-schema friendly

- Green app sees Green model



Why is this important?

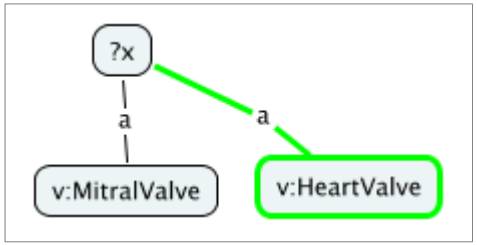
- Different formats, data models and vocabularies can be:
  - used together harmoniously
  - semantically linked
- New ones (or new versions) can be gracefully incorporated
  - Healthcare vocabularies are revised ~3-8% per year

Unified Medical Language System (UMLS) includes over 100 standard vocabularies and millions of concepts!

#1: RDF enables smarter data use and automated data translation

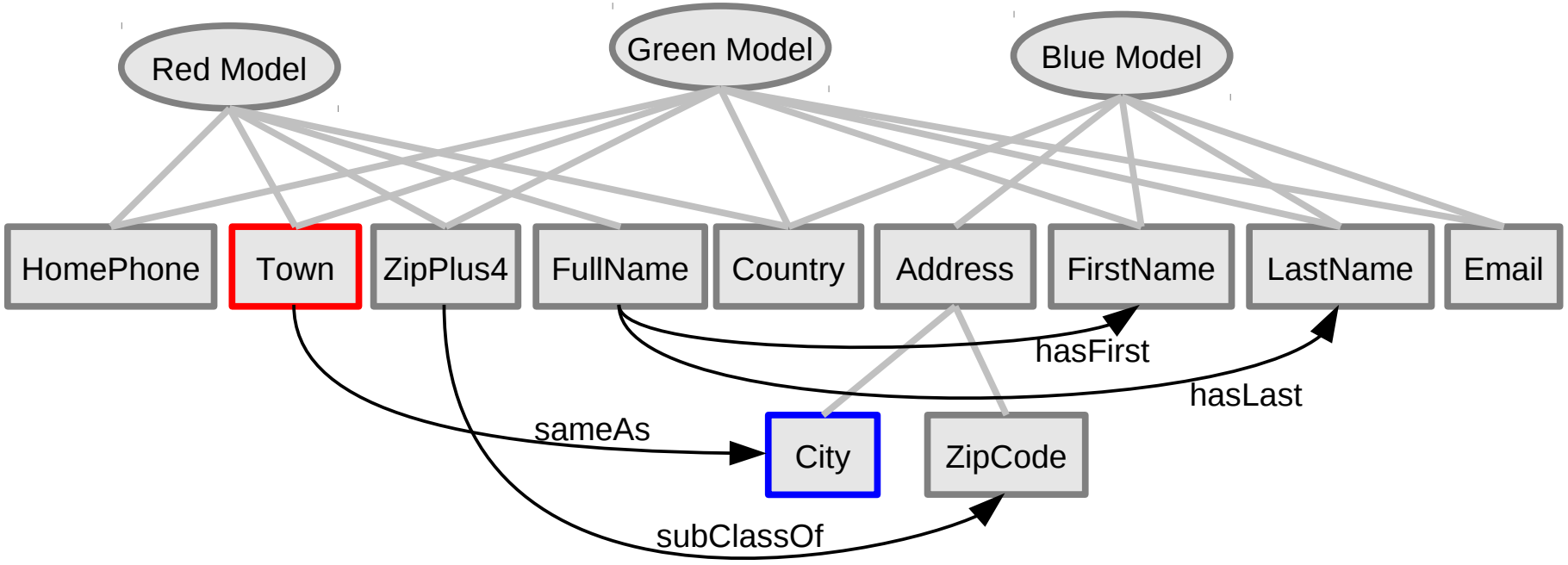
- RDF enables inference
- Inference derives new assertions from old
  - "Entailments"
- Query for v:HeartValve surgeries can find v:MitralValve surgeries

Inference example



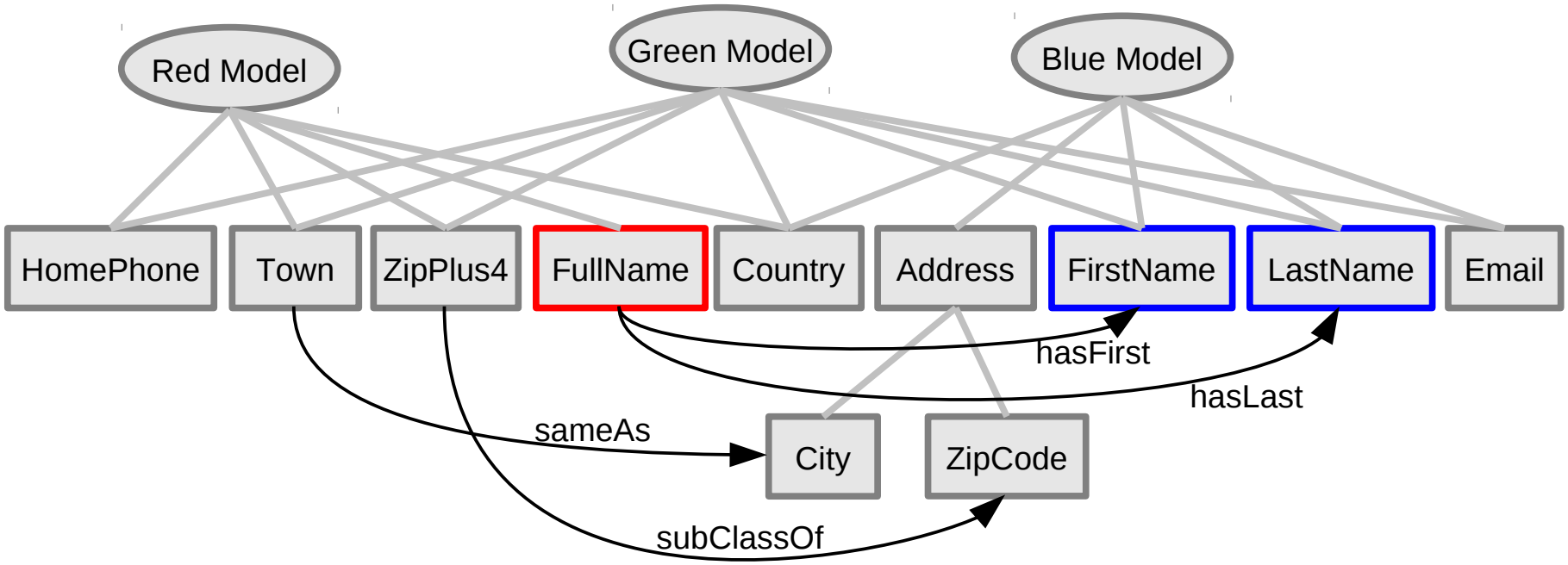
- If you know:
  ?x a v:MitralValve .
  v:MitralValve rdfs:subClassOf v:HeartValve .
- Then you can infer:
  ?x a v:HeartValve .

Inference example: sameAs



- If you know: Town
- You can infer: City (or vice versa)

Inference example: composition



- If you know: FirstName + LastName
- You can infer: FullName
  - But not necessarily vice versa

Why is this important?

- Smarter data use
  - Query for v:HeartValve surgeries can find v:MitralValve surgeries
- Automated data transformation
  - Red Model data + Blue Model data => Green Model data

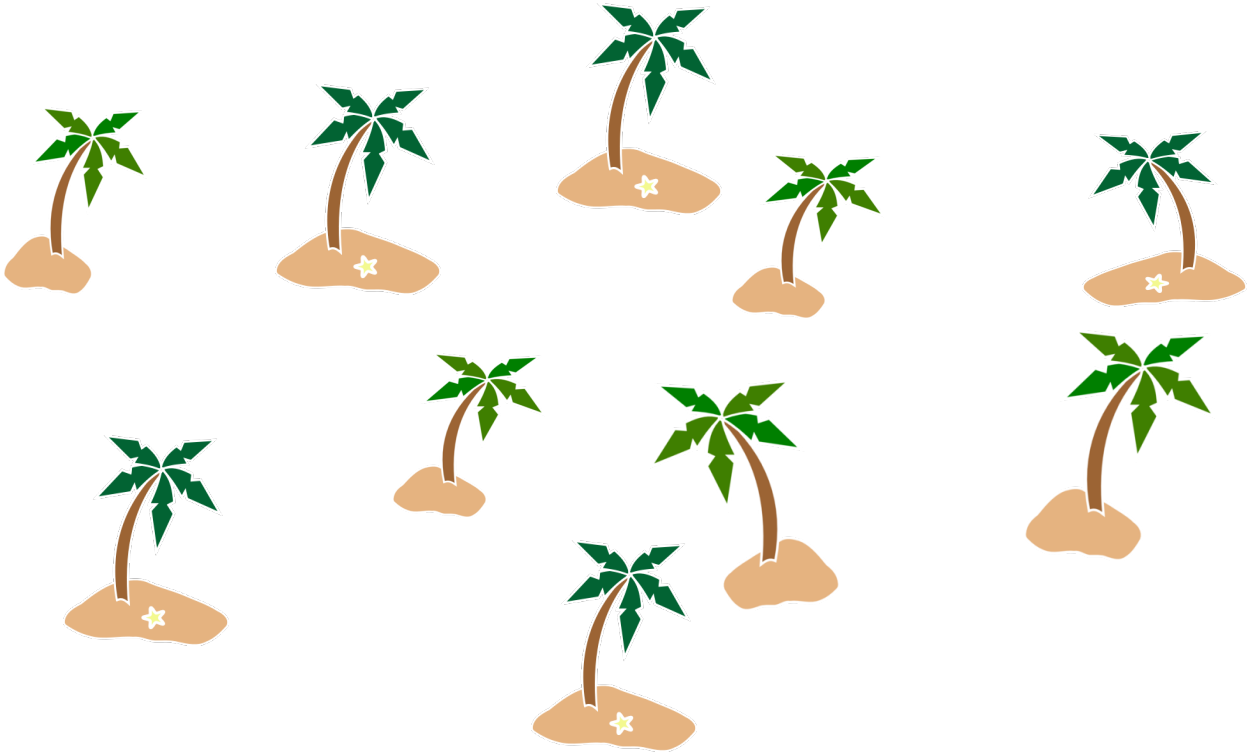
How RDF can help standards convergence

Standard Vocabularies in UMLS

AIR ALT AOD AOT BI CCC CCPSS CCS CDT CHV COSTAR CPM CPT CPTSP CSP CST DDB DMDICD10 DMDUMD DSM3R DSM4 DXP FMA HCDT HCPCS HCPT HL7V2.5 HL7V3.0 HLREL ICD10 ICD10AE ICD10AM ICD10AMAE ICD10CM ICD10DUT ICD10PCS ICD9CM ICF ICF-CY ICPC ICPC2EDUT ICPC2EENG ICPC2ICD10DUT ICPC2ICD10ENG ICPC2P ICPCBAQ ICPCDAN ICPCDUT ICPCFIN ICPCFRE ICPCGER ICPCHEB ICPCHUN ICPCITA ICPCNOR ICPCPOR ICPCSPA ICPCSWE LAB LCH LNC_AD LNC_MDS30 MCM MEDLINEPLUS MSHCZE MSHDUT MSHFIN MSHFRE MSHGER MSHITA MSHJPN MSHLAV MSHNOR MSHPOL MSHPOR MSHRUS MSHSCR MSHSPA MSHSWE MTH MTHCH MTHHH MTHICD9 MTHICPC2EAE MTHICPC2ICD10AE MTHMST MTHMSTFRE MTHMSTITA NAN NCISEER NIC NOC OMS PCDS PDQ PNDS PPAC PSY QMR RAM RCD RCDAE RCDSA RCDSY SNM SNMI SOP SPN SRC TKMT ULT UMD USPMG UWDA WHO WHOFRE WHOGER WHOPOR WHOSPA

Over 100!

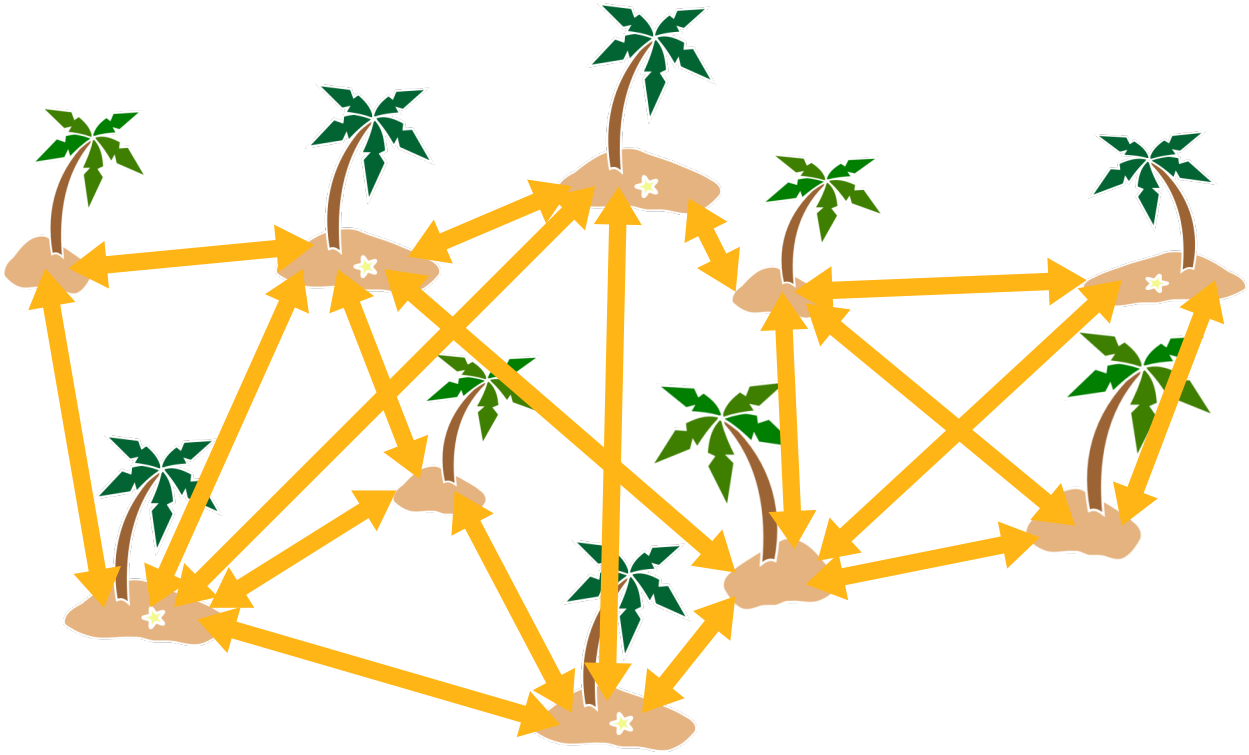
Each standard is an island



How RDF helps standards

- Enables common semantic linkage across standards
  - Use OWL to define semantics
- Encourages semantic clarity and consistency
- Distributed extensibility and late linkage

Bridging healthcare standards



Why RDF?

- Captures **information content**
- **Multi-schema friendly**
- Enables **smarter data use**
- Enables **bridging** of diverse standards
- Mature, **vendor-neutral** international standard
- *The "best available candidate" for a universal healthcare exchange language*

http://YosemiteManifesto.org/

BACKUP SLIDES

De jure versus de facto standards

- De facto standards evolve faster than de jure standards
- RDF supports both

- @@ TODO: Add slides showing how a vocab can be extended by one party, then used by other @@