RDF for Semantic Interoperability

A subgroup of the HL7 ITS work group, working jointly with the W3C Healthcare and Life Sciences group

David Booth, Hawaii Resource Group
Claude Nanjo, Cognitive Medical Systems

November 2014

Download the latest version of these slides from
http://dbooth.org/2014/hl7/

Outline

- Executive Overview
  - Mission and goals of this work group
  - How this work fits into the larger goal of achieving interoperability – The Yosemite Project
  - Initial proposed work items
- Background
  - Semantic interoperability
  - What is RDF?
- Details
  - Goals
  - Purpose
  - People

EXECUTIVE OVERVIEW

Timeline

- May 2014:
  - Idea for this effort first suggested by Stan Huff, though RDF work in general has gone on for years
- Jun-Sep 2014:
  - Draft charter, circulate and revise
  - Initiate approval process
- Sep 2014 Chicago HL7 meetings:
  - Numerous hallway discussions about starting this work
  - Pros/cons of starting as a separate HL7 work group versus as a subgroup of an existing group
- Oct 2014 Formal creation of HL7 ITS subgroup on RDF for Semantic Interoperability – Joint meetings with W3C Healthcare and Lifesciences group
- (PLANNED) Oct-Dec 2014:
  - Begin teleconferences, develop work plan, prioritize deliverables, start work
- (PLANNED) Jan 2015 San Antonio HL7 meetings:
  - First face-to-face meeting

Mission

- Establish semantic interoperability of structured healthcare information
- Using RDF and related standards to express machine-processable meaning

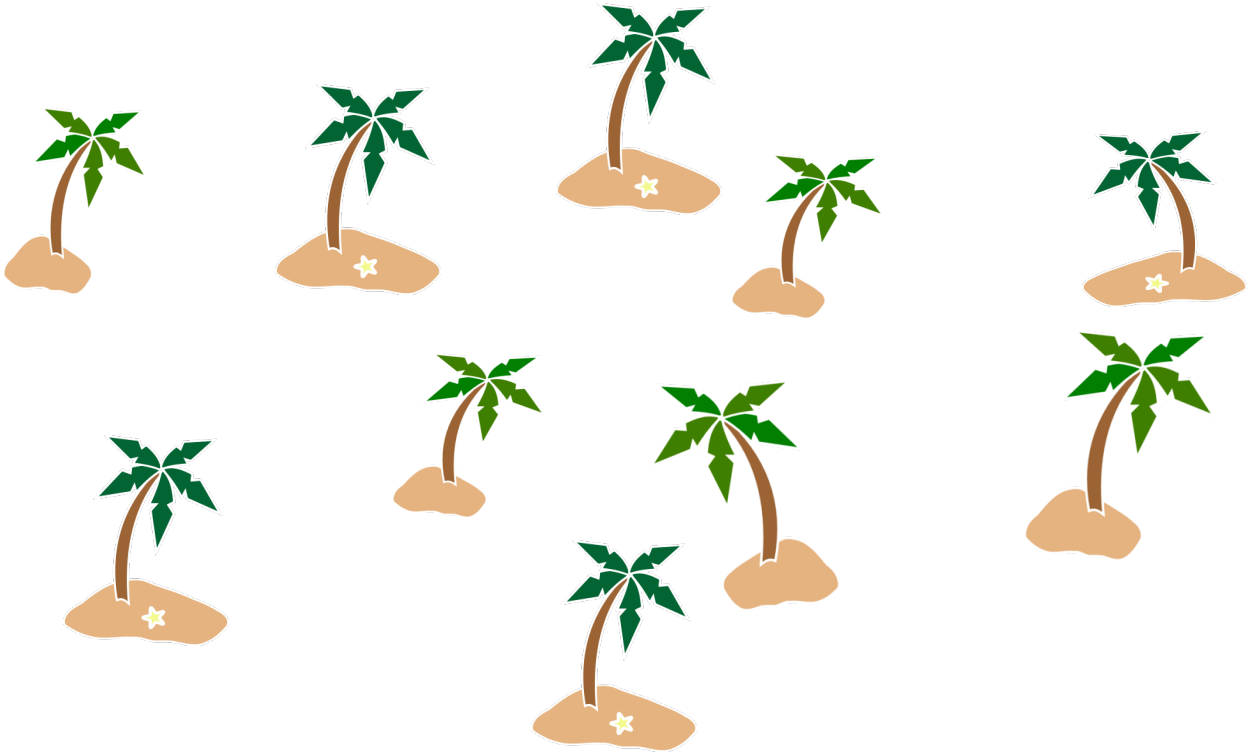
=> *Increase the effectiveness and efficiency of healthcare delivery*

Semantic interoperability:

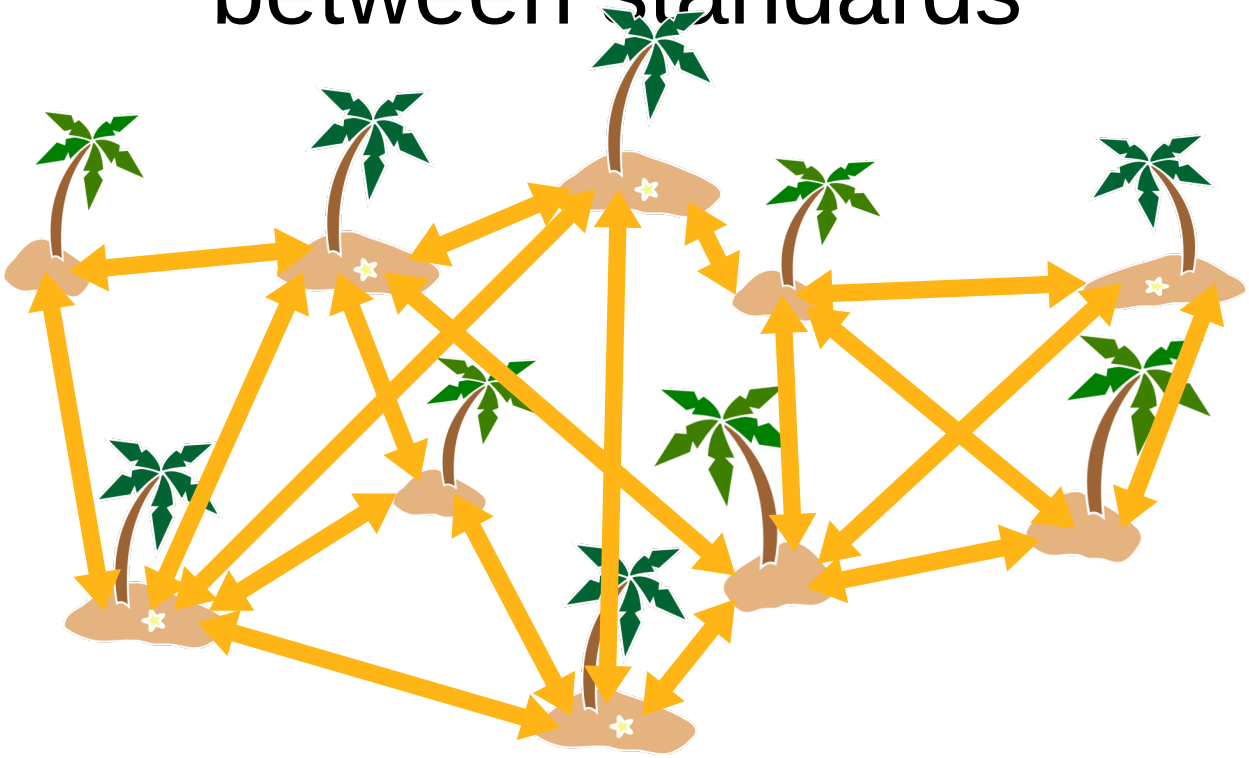
The ability of computer systems
to exchange data
with unambiguous, shared meaning.

– Wikipedia

Each standard is an island



RDF and OWL are semantic bridges between standards



RDF and OWL

- W3C standards
- RDF captures the semantics or information content of structured data
  - Any data formats, models and vocabularies
- OWL builds on RDF to define classes, properties and relationships between them

Major goals

- Create **RDF and OWL definitions** of healthcare information standards
  - Enable semantics to be uniformly interpreted in RDF
- Act as a **knowledge resource** for other work groups

FAQ: Yet another domain model?

- No! Domain models and vocabularies will still be developed by other groups
- Charter: *This work group shall remain neutral about specific healthcare domain models and vocabularies.*
- Purpose:
  - **Bridge** between standards
  - **Facilitate convergence** toward common models

FAQ: Why an HL7 work group?

- Raise the visibility of this work
- Support multiple projects in multiple HL7 groups
- Coordinate efforts across groups
  - RDF and OWL ontologies
- Establish a focal point to rally the community
  - Spanning HL7 and W3C communities

FAQ: Will this compete with other HL7 activities or work group membership?

- A: No! Purpose is to support other groups.
- Membership overlap is intentional
- Deliverables will largely be for other groups
  - E.g., developing OWL ontology for FHIR

FAQ: Why work with W3C?

- Better link HL7 and W3C communities
- Leverage W3C expertise
  - W3C is the epicenter of RDF technology
- Expose HL7 standards to a wider community

FAQ: Why start as a sub-group of an existing HL7 work group?

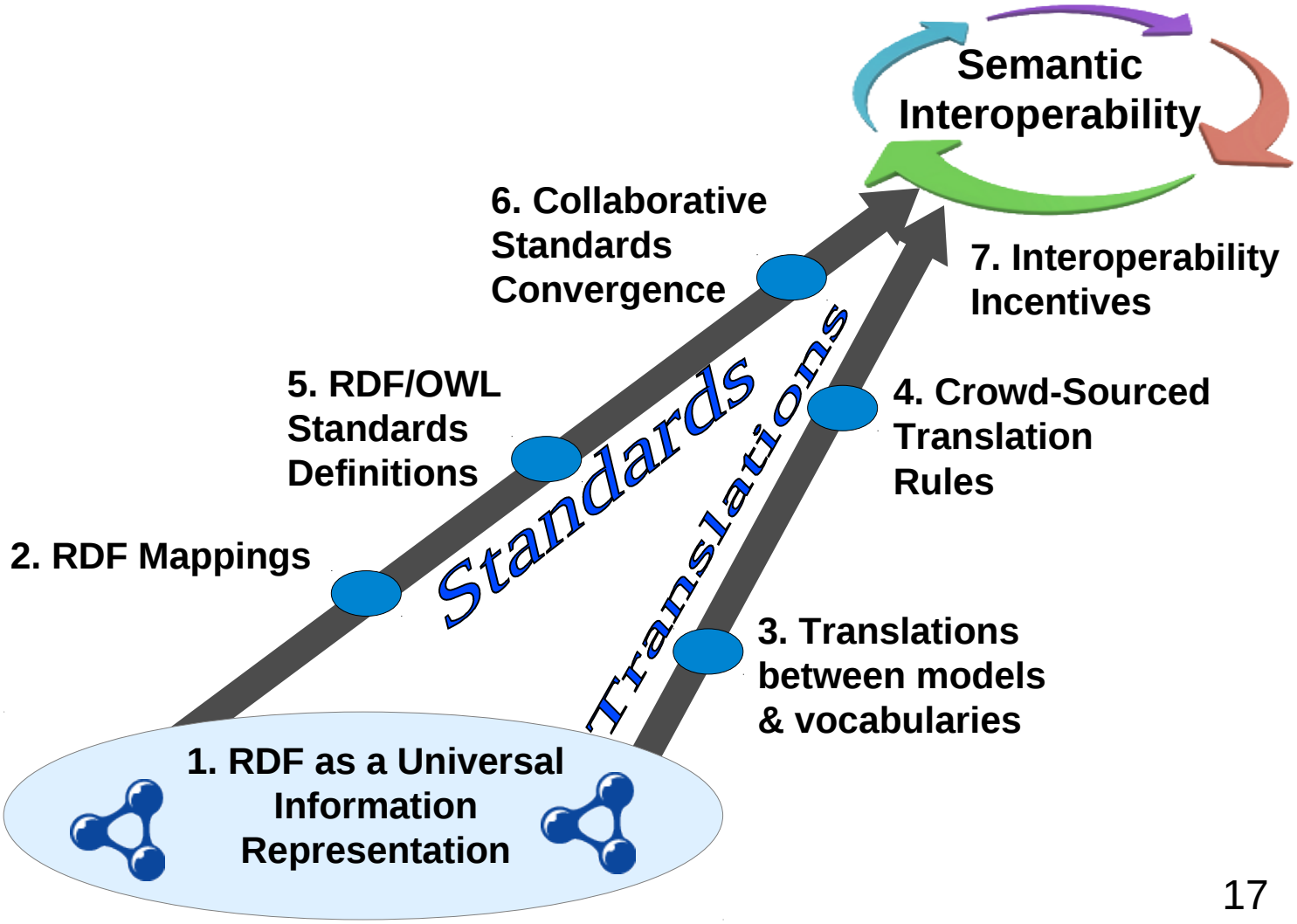
- Easier assimilation into HL7 processes
  - Less admin burden
- Better coordination and networking within HL7
- Avoid (incorrect) perception of this work competing with existing HL7 efforts

http://YosemiteProject.org/

MISSION:
Semantic interoperability
of
all structured healthcare information

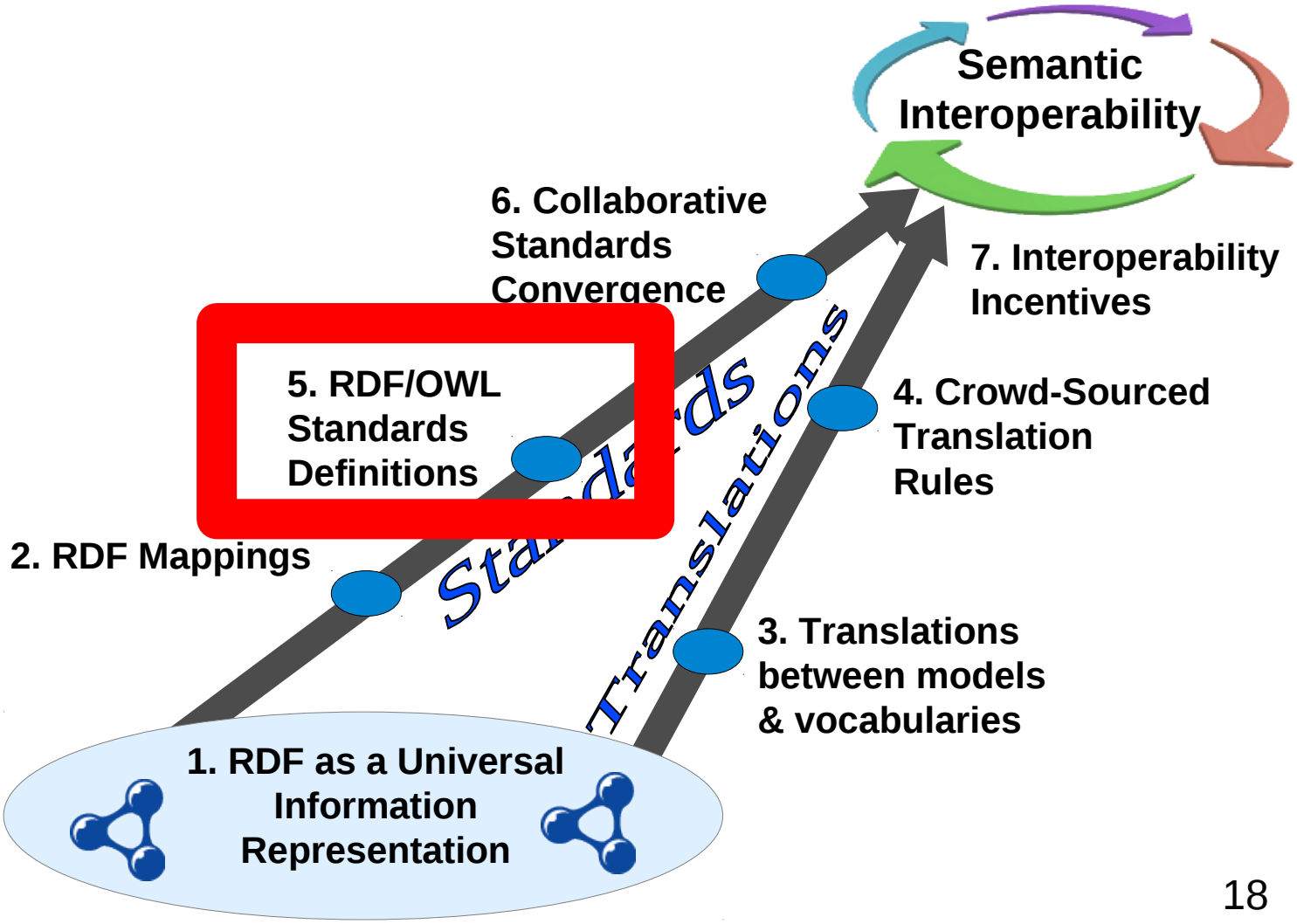
http://YosemiteProject.org/

A Roadmap for Healthcare Information Interoperability



http://YosemiteProject.org/

A Roadmap for Healthcare Information Interoperability



Proposed work items

- FHIR ontology
- ICD-11 and SNOMED
- Bridging VA, Intermountain and other models
- Use Cases
- PhUSE-FDA project (formerly CDISC2RDF)
- C-CDA RDF representations
- High-level concept mapping to RDF (AR typeCodes, etc.)

Weekly Agenda:
http://wiki.hl7.org/index.php?title=ITS_RDF_ConCall_Agenda

BACKGROUND

Two ways to achieve interoperability

- Standards:
  - Make everyone speak the same language
  - I.e., same data models and vocabularies
- Translations:
  - Translate between languages
  - I.e., translate between data models and vocabularies

Standards are preferred, but we cannot eliminate the need for translations. RDF helps with both.

What is RDF?

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

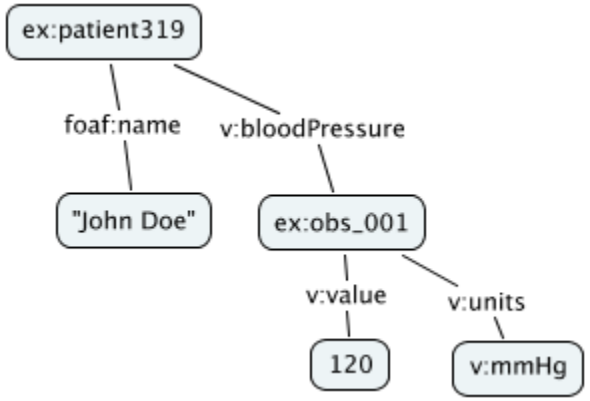
English assertions:

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"):

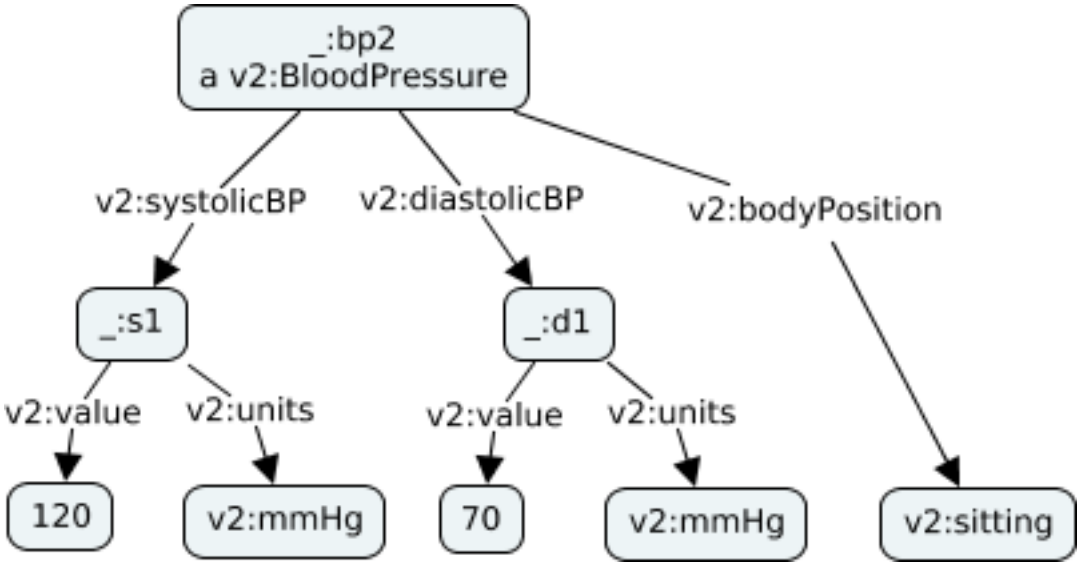
ex:patient319 foaf:name "John Doe" .
ex:patient319 v:systolicBP ex:obs_001 .
ex:obs_001 v:value 120 .
ex:obs_001 v:units v:mmHg .

RDF graph:



*Namespace definitions omitted

Hierarchical data model in RDF



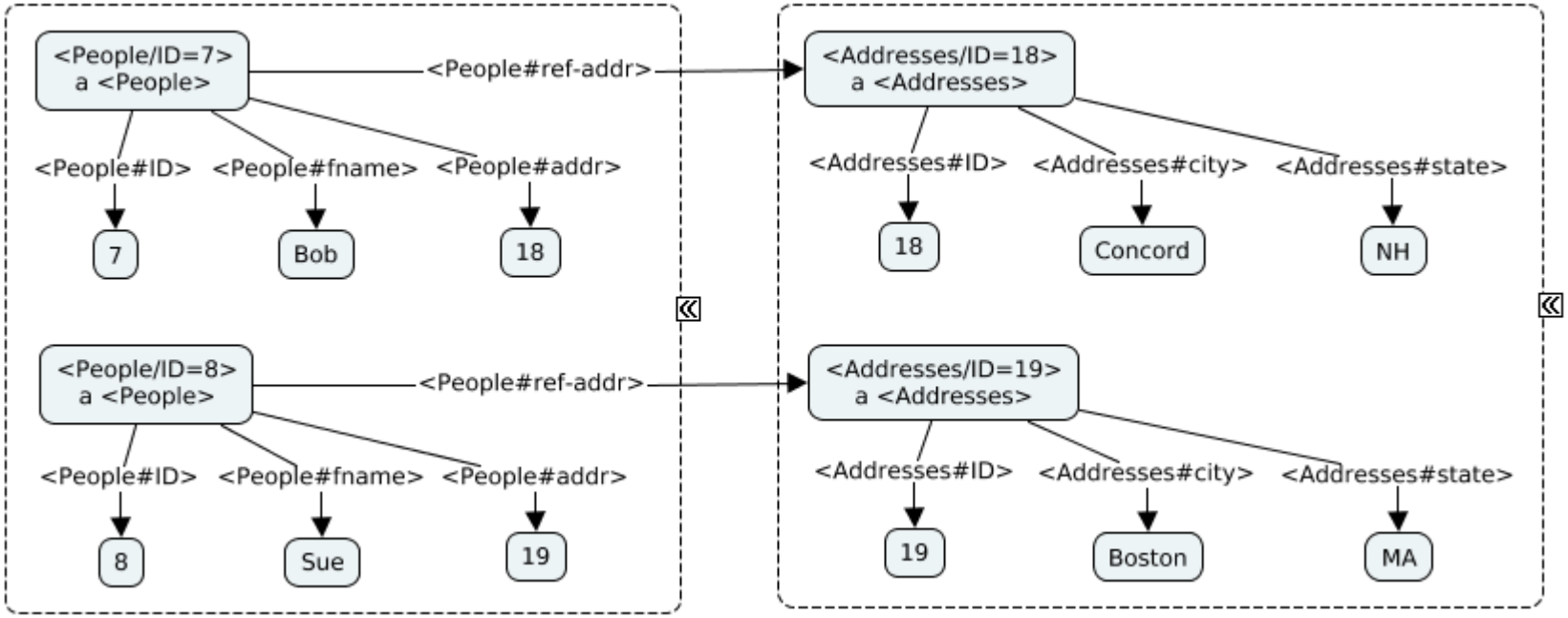
Relational data model in RDF

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 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

#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|
```

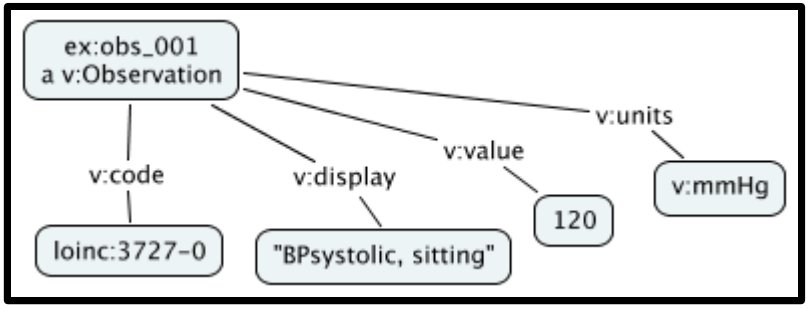
Maps to

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?

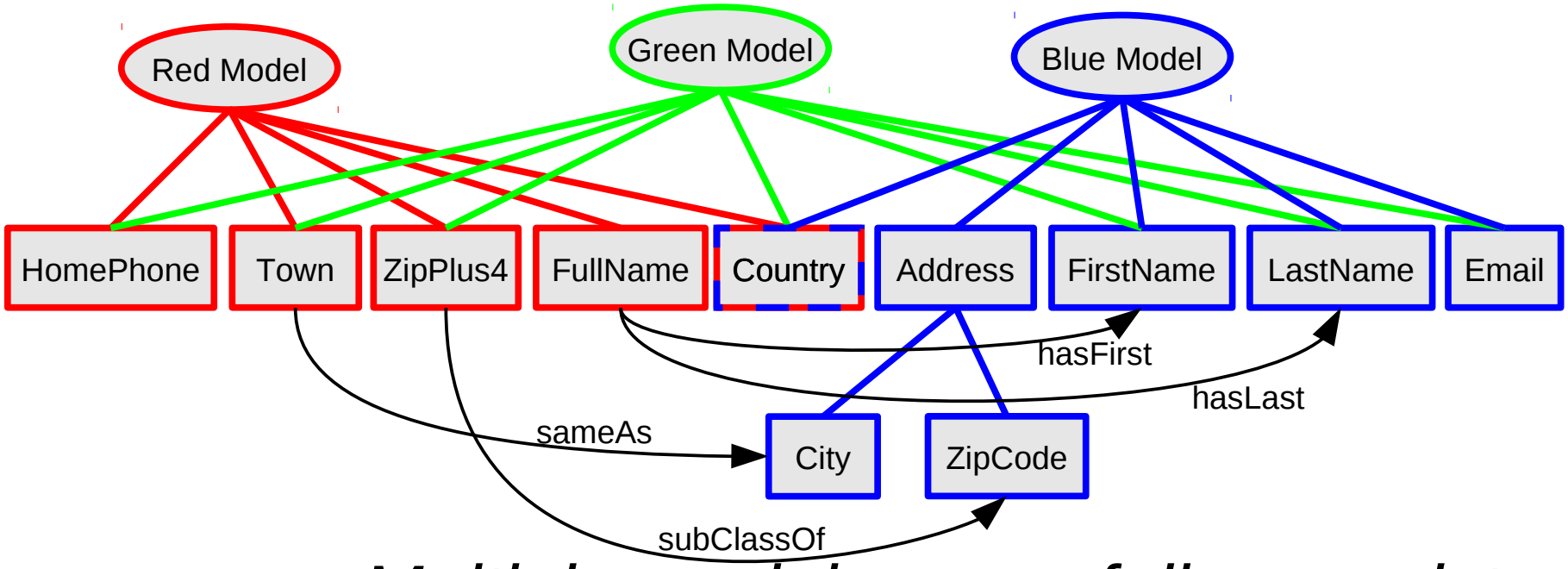
- Emphasis is on the meaning (where it should be)
- RDF acts as a common information representation

#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.

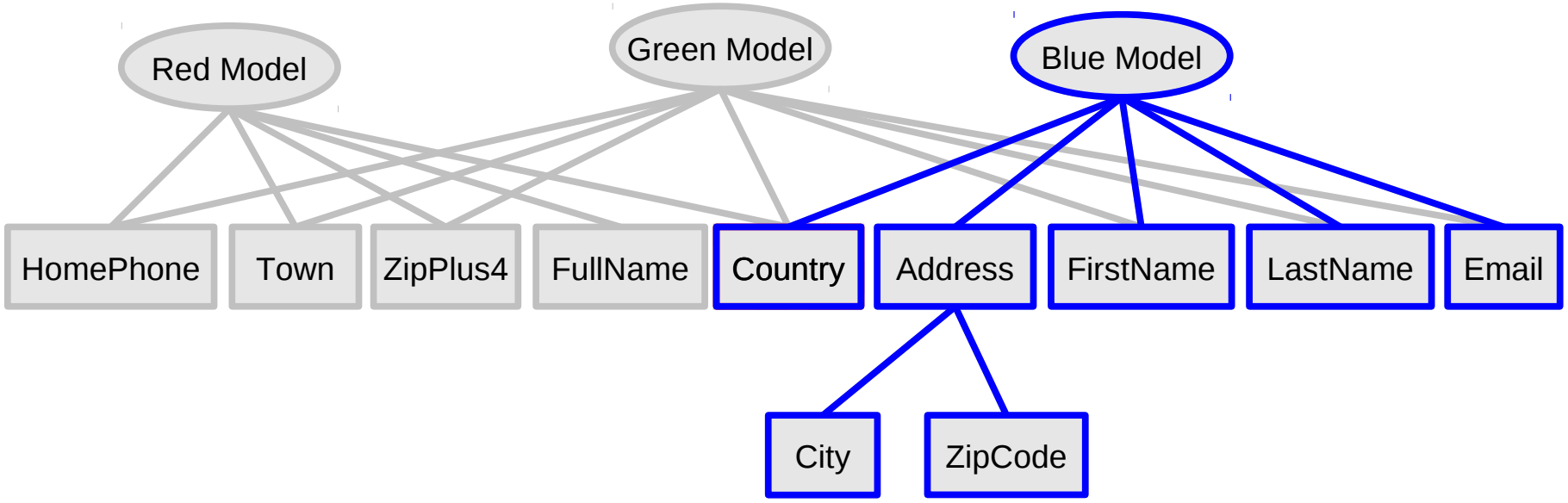
Multi-schema friendly



Multiple models peacefully co-exist

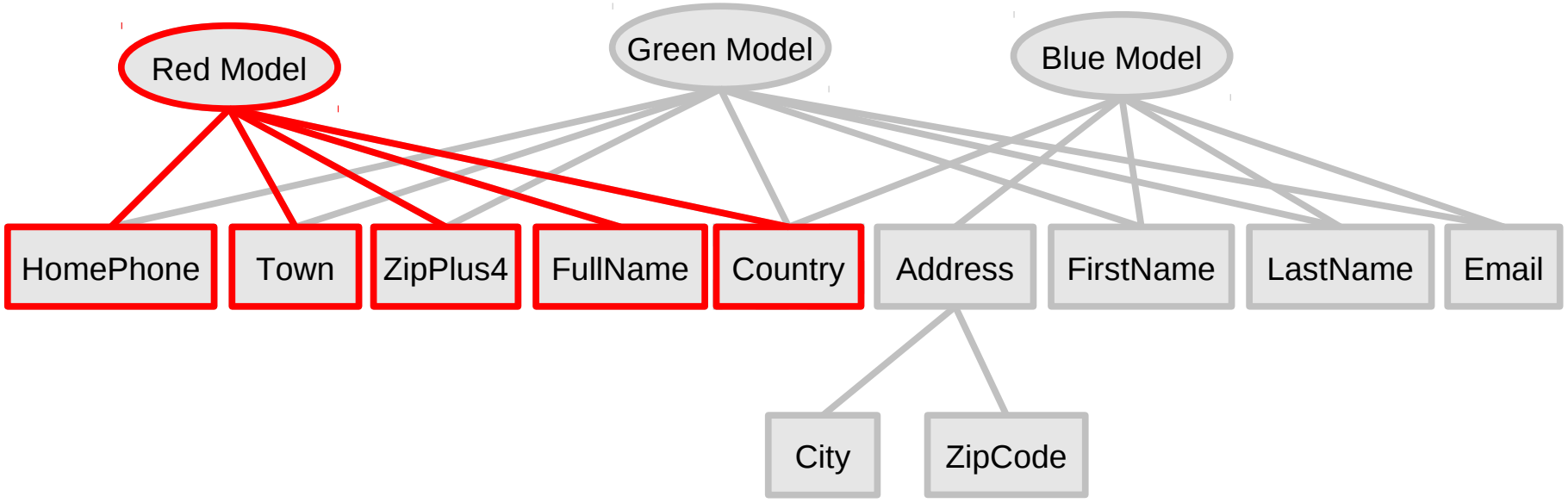
Multi-schema friendly

- Blue app sees Blue model



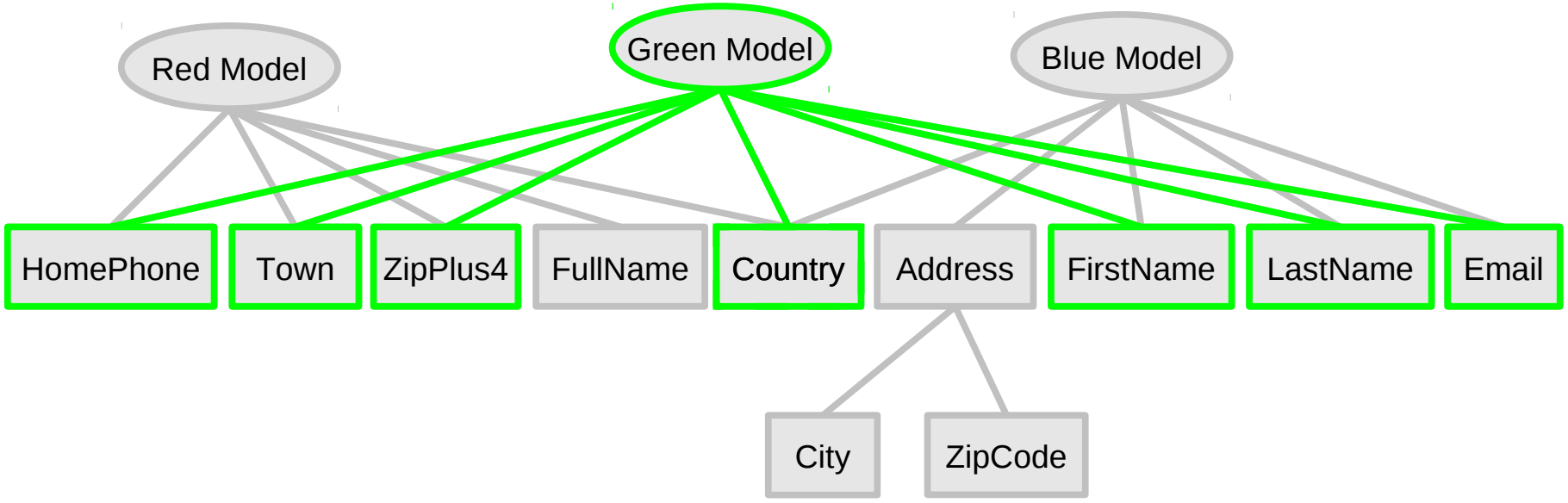
Multi-schema friendly

- Red app sees Red model



Multi-schema friendly

- Green app sees Green model



Why is this important?

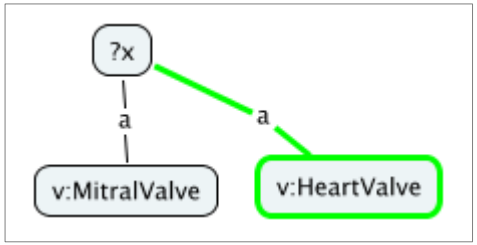
- Different formats, data models and vocabularies can be:
  - used together harmoniously
  - semantically linked
- New ones (or new versions) can be gracefully incorporated
  - Healthcare vocabularies are revised ~3-8% per year

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 .

New Work Group:
RDF for Semantic Interoperability

Charter

- DRAFT FOR DISCUSSION - NO OFFICIAL STATUS -
RDF for Semantic Interoperability Work Group

Health Level Seven

Semantic interoperability is the ability of computer systems to exchange data with unambiguous, shared meaning. (Source: Wikipedia)

Mission

The *RDF for Semantic Interoperability* work group seeks to establish semantic interoperability of structured healthcare information by using W3C Resource Description Framework (RDF) and related standards to express machine-processable meaning. It supports the HL7 mission to increase the effectiveness and efficiency of healthcare delivery.

Charter

The goal of this work group is to facilitate the use of RDF as a common semantic model for interpreting instance data that may originate in any format, data model or vocabulary. The purpose is to enable data to be computable and semantically interoperable. Losslessly representing heterogeneous data in a semantically consistent way is critical to delivering quality healthcare. Semantic clarity is essential for knowledge representation that is sharable and actionable.

This work complements other standardization work, both inside and outside HL7, by offering a common semantic foundation for all healthcare-related information representations. Although this work group will encourage and indirectly facilitate convergence toward a set of common, semantically interoperable healthcare data models and vocabularies – potentially spanning many standards – choices about specific healthcare domain models and vocabularies are the responsibility of other groups whose missions are more specifically devoted to making those decisions. This work group shall remain neutral about specific healthcare domain models and vocabularies.

Some specific goals:

Create and adopt standard mappings for instance data from existing and future healthcare information formats, data models and vocabularies to RDF and vice versa.
Create and adopt W3C Web Ontology Language (OWL) semantic definitions for existing and future healthcare information specifications.
Encourage the adoption of RDF as the authoritative semantic interpretation of structured healthcare information.
Encourage the use of RDF as a lossless information exchange language, particularly in situations where the use of other representations could involve information loss.
Encourage the use of W3C Linked Data principles.
Facilitate convergence on common URIs and URI patterns.
Act as a knowledge resource for other work groups, on the most effective use of RDF and related standards.

These goals may be achieved partially through this group's own efforts and partially through the work of others. Much of the group's purpose will be to assist and coordinate with work done by others, both within HL7 and externally.

The purpose of this work is to:

enable the meaning of any structured instance data to be accurately determined, regardless of the source format, data model or vocabulary in which that data is represented;
enable domain models and vocabularies to be described and related, using OWL as a common description language;
encourage community convergence around common RDF domain models and vocabularies, in order to minimize model and vocabulary transformations or mappings during data exchange;
facilitate accurate semantic transformation and integration of instance data that originates in different formats, data models or vocabularies, using RDF as a common semantic representation;
facilitate the sharing of model and vocabulary transformation rules and procedures, based on RDF as a common semantic representation; and
facilitate semantic alignment and consistency within and between healthcare information specifications, using RDF, OWL and related standards.

Formal Relationships with Other HL7 Groups

This work group will collaborate with other HL7 work groups involved in domain models or vocabularies as appropriate.

Formal Relationships with Groups Outside of HL7

This work group will collaborate with the W3C Semantic Web for Healthcare and Life Sciences Interest Group. This work group will also collaborate with other relevant external efforts as appropriate, such as IHTSDO (SNOMED CT) and WHO (ICD-11).

Date of Last Revision

11-Jul-2014

Mission

- Establish semantic interoperability of structured healthcare information
- Using RDF and related standards to express machine-processable meaning

=> *Increase the effectiveness and efficiency of healthcare delivery*

How? (1 of 2)

- RDF as a **common semantic model** for interpreting instance data
- Instance data may originate in **any format, data model or vocabulary**

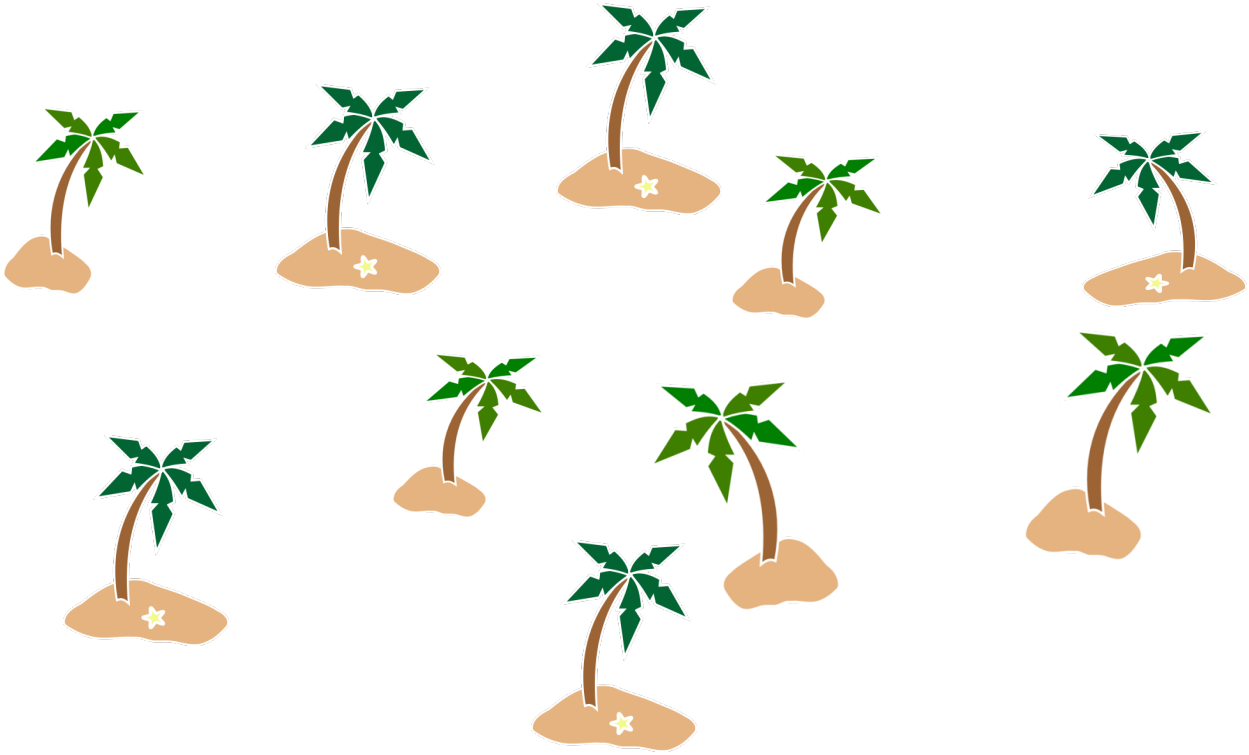
Q: Convert all data to RDF?

A: No! RDF mapping provides a consistent way to interpret the data's meaning.

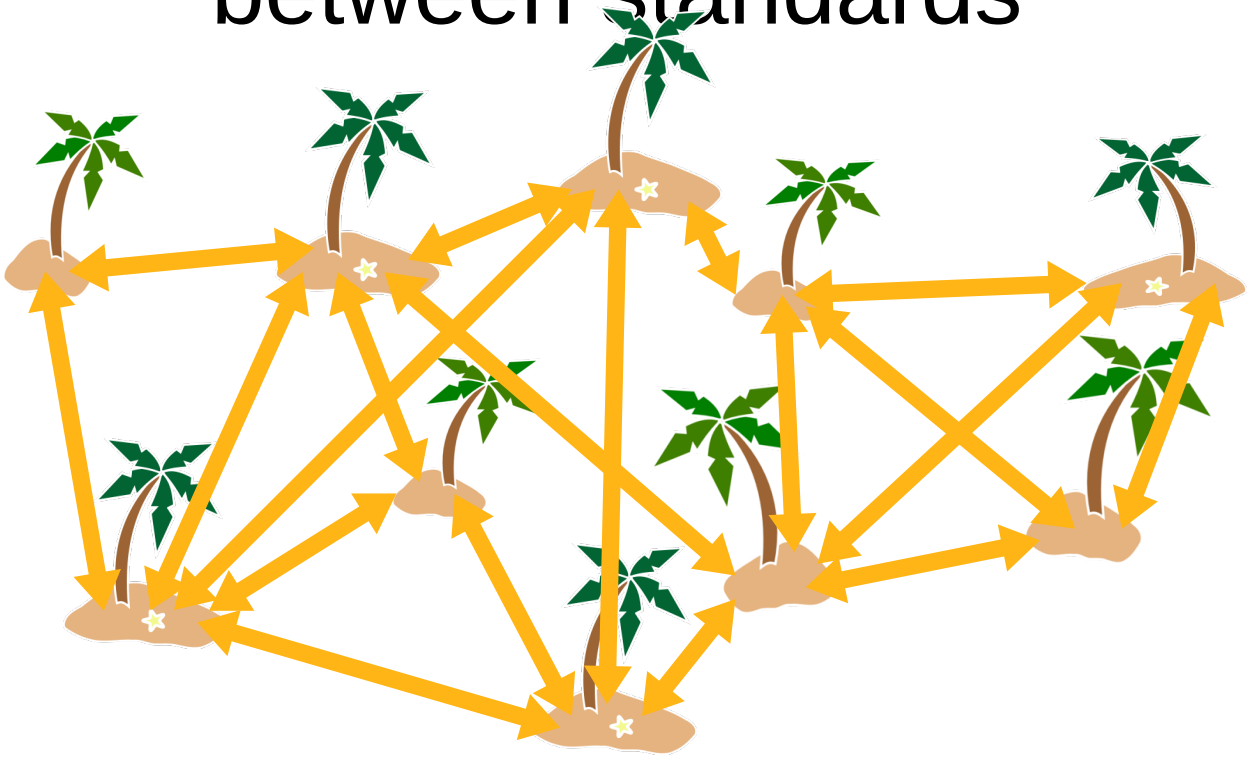
How? (2 of 2)

- Facilitate **convergence** toward common, semantically interoperable healthcare data models and vocabularies
- Spanning many standards
- Remain **neutral** about specific healthcare domain models and vocabularies

Each standard is an island



RDF and OWL are semantic bridges between standards



FAQ: Yet another model?

- No! Domain models and vocabularies will still be developed by other groups
- Purpose:
  - **Bridge** between standards
  - **Facilitate convergence** toward common models

FAQ: Why an HL7 work group?

- Raise the visibility of this work
- Support multiple projects in multiple HL7 groups
- Coordinate efforts across groups
  - RDF and OWL ontologies
- Establish a focal point to rally the community
  - Spanning HL7 and W3C communities

FAQ: Will this compete with other HL7 activities or work group membership?

- A: No! Purpose is to support other groups.
- Membership overlap is intentional
- Deliverables will largely be for other groups
  - E.g., developing OWL ontology for FHIR

Specific goals

- Create **standard mappings** from existing/future information formats, data models and vocabularies to RDF
- Create **OWL definitions** of existing/future healthcare information specifications
- Encourage the use of:
  - RDF for semantic interpretation
  - RDF as a lossless information exchange language
  - Linked Data principles (e.g., clickable URIs as identifiers)
- Facilitate convergence on common URIs and URI patterns
- Act as a **knowledge resource** for other work groups

Role

Most of the group's purpose will be to assist and coordinate with work done by others, both within HL7 and externally.

Coordination with other groups

- **W3C**
  - Semantic Web for Healthcare and Life Sciences Interest Group
- IHTSDO (SNOMED CT)
- WHO (ICD-11)

FAQ: Why work with W3C?

- Better link HL7 and W3C communities
- Tap into W3C expertise
  - W3C is the epicenter of RDF technology
- Expose HL7 standards to a wider community

Purpose

- Enable the **meaning** of any structured instance data to be accurately determined
- Enable domain models and vocabularies to be **described and related**
- Encourage **community convergence** around common RDF domain models and vocabularies
- Facilitate accurate **semantic transformation** and integration of instance data
- Facilitate the **sharing** of model and vocabulary **transformation rules**
- Facilitate **semantic alignment** and consistency within and between healthcare information specifications

Initial co-chairs

- David Booth, Ph.D., Hawaii Resource Group http://dbooth.org/
  - 10 years semantic web technology
  - 5 years applied to healthcare and life sciences
- Claude Nanjo, MS, Cognitive Medical Systems
  - 7 years semantic web technology
  - 9 years in healthcare
  - 12 years machine learning / data mining / ETL

Initial members (in addition to co-chairs)

- Rafael Richards, VA
- Anthony Mallia, Edmond Scientific
- Joshua Phillips, Semantic Bits
- Cecil Lynch, Accenture
- Harold Solbrig, Mayo Clinic
- Josh Mandel, Harvard
- Mehmet Aydar, Cleveland Clinic

Others who have expressed interest

- Michel Dumontier, Stanford
- Conor Dowling, Caregraf
- Michael Miller, Systems Biology
- Marc Twagirumukiza, Agfa
- Vipul Kashyup
- Oya Beyan
- Tim Williams, UCB
- M. Scott Marshall, HCLS
- Jeremy Carroll, Syapse
- Keith Cambell, Informatics
- Peter Hendler, Kaiser Permanente
- Jamie Ferguson, Kaiser Permanente,
- Kerstin Forsberg
- Melissa Haendel, OHSU
- Keith Boone
- Alex Garcia
- Jiang Guoqian, Mayo Clinic
- Ratnesh Sahay, DERI
- Joachim Baran
- Hans Cools
- Sofia Pinto

Timeline & Deliverables

Timeline:

- Sept 2014 (Chicago): HL7 approval of the work group
- Jan 2015 (San Antonio): First face-to-face meeting

Deliverables:

- TBD

Questions?

More Information

- **RDF as a Universal Healthcare Exchange Language** -- David Booth – http://dbooth.org/2014/rdf-as-univ/
- **Why RDF?** -- David Booth – http://dbooth.org/2014/why-rdf/
- **The Ideal Medium for Health Data? A Dive into Lab Tests** – Conor Dowling – http://schemes.caregraf.info/presentations/semtech2014/
- **Introduction and RDF Representation of Fast Healthcare Interoperability Resources (FHIR) for Clinical Data** – Josh Mandel – http://bit.ly/fhir-semtech-2014
- **Transformations for Integrating VA data with FHIR in RDF** – Rafael Richards – http://dbooth.org/2014/richards/
- **Towards a Web of Clinical Knowledge** – Claude Nanjo – http://dbooth.org/2014/nanjo/
- **Data-Driven Biomedical Research with Semantic Web Technologies** – Michel Dumontier – http://dbooth.org/2014/dumontier/
- **The Yosemite Project: A Roadmap for Healthcare Information Interoperability** -- David Booth – http://dbooth.org/2014/yosemite/

BACKUP SLIDES