The Yosemite Project
A Roadmap for Healthcare Information Interoperability

David Booth, Hawaii Resource Group
Conor Dowling, Caregraf
Michel Dumontier, Stanford University
Josh Mandel, Harvard University
Claude Nanjo, Cognitive Medical Systems
Rafael Richards, Veterans Affairs

Semantic Technology and Business Conference
21-Aug-2014

SEE LATEST VERSION:
http://tinyurl.com/YosemiteRoadmap20150709slides
Outline

• Mission and strategy
• Semantic interoperability
  – Standards
  – Translations
• Roadmap for interoperability
• Cost
MISSION:

Semantic interoperability of all structured healthcare information
MISSION:

Semantic interoperability of all structured healthcare information
STRATEGY:

*RDF as a universal information representation*
Universal information representation

• Q: What does instance data X mean?
• A: Determine its RDF information content

<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>
Why RDF?

- "Captures information content, not syntax"
- "Multi-schema friendly"
- " Allows data models and vocabularies to evolve"
- "Good for model transformation"
- "Allows diverse data to be connected and harmonized"
- "Supports inference"

- Endorsed by over 100 thought leaders in healthcare and technology as the best available candidate for a universal healthcare exchange language
  - See http://YosemiteManifesto.org/

http://dbooth.org/2014/why-rdf/
Semantic interoperability:

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

– Wikipedia
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
Obviously we prefer standards.

But . . . .
Standardization takes time

COMING SOON!
COMPREHENSIVE STANDARD

DUE
2016
2036
2096
Standards trilemma: Pick any two

- **Timely**: Completed quickly
- **Good**: High quality
- **Comprehensive**: Handles all use cases
Modernization takes time

- Existing systems cannot be updated all at once
Diverse use cases

• Different use cases need different data, granularity and representations

One standard does not fit all!
Standards evolve

• Version $n+1$ improves on version $n$
Healthcare terminologies rate of change

Slide credit: Rafael Richards (VA)
Translation is unavoidable!

• Standardization takes time
• Modernization takes time
• Diverse use cases
• Standards evolve
A realistic strategy for semantic interoperability must address both standards and translations.
Interoperability achieved by standards vs. translations
How RDF Helps Standards
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 ICPCITAE ICPCITAPI ICPCNOR ICPCPOL ICPCSPA ICPCW53AE ICD9 LCH LNC AD LNC MDS30 MCM MEDLINEPLUS MSHCZE MSHDUT MSHFIN MSHFRE MSHGER MSHITA MSHJPN MSHLAV MSHNOR MSHPOL MSHPOR MSHRUS MSHSCR MSHPSPA MSHPSEWE 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 WHOPOPOR WHOSPA
How Standards Proliferate:
(See: A/C chargers, character encodings, instant messaging, etc)

SITUATION:
There are 14 competing standards.

14?! Ridiculous!
We need to develop one universal standard that covers everyone's use cases. Yeah!

SOON:

SITUATION:
There are 15 competing standards.

http://xkcd.com/927/
Used by permission
Each standard is an island

- Each has its "sweet spot" of use
- Lots of duplication
RDF and OWL enable semantic bridges between standards

• Goal: a cohesive mesh of standards that act as a single comprehensive standard
• RDF also helps avoid the bike shed effect . . .
Bike shed effect
a/k/a Parkinson's Law of Triviality

Organizations spend disproportionate time on trivial issues. -- C.N. Parkinson, 1957

1. Nuclear Plant
   Cost: $28,000,000
   Discussion: 2.5 minutes

2. Bike Shed
   Cost: $1,000
   Discussion: 45 minutes
Standards committees and the bike shed effect

- Committees spend hours deciding on data formats, syntax and naming
  - Irrelevant to the computable information content
RDF helps avoid the bike shed effect

- Each group can use its favorite data format, syntax and names
- RDF can uniformly capture the information content
Needed: Collaborative Standards Hub

- A cross between BioPortal, GitHub, WikiData, Web Protege, CIMI repository, HL7 model forge, UMLS Semantic Network and Metathesaurus
  - Next generation BioPortal?
Collaborative Standards Hub

- Repository of healthcare information standards
- Supports standards groups and implementers
- Holds RDF/OWL definitions of data models, vocabularies and terms
- Encourages:
  - Semantic linkage
  - Standards convergence
Collaborative Standards Hub

- Suggests related concepts
- Checks and notifies of inconsistencies – within and across standards
- Can be accessed by browser or RESTful API
Collaborative Standards Hub

• Can scrape or reference definitions held elsewhere

• Provides metrics:
  – Objective (e.g., size, number of views, linkage degree)
  – Subjective (ratings)

• Uses RDF and OWL under the hood
  – Users should not need to know RDF or OWL
iCat: Web Protege tool for ICD-11
iCat development of ICD-11

In three years:

• 270 domain experts around the world
• 45,000+ classes
• 260,000+ changes
• 17,000 links to external terminologies
FIBO development process

- Financial standards are developed in RDF/OWL
How RDF Helps Translation
How RDF helps translation

• RDF supports **inference**
  – Can be used for translation
• RDF acts as a **universal information representation**
• Enables data model and vocabulary translations to be **shared**
Translating patient data

- Steps 1 & 3 map between source/target syntax and RDF
- Step 2 translates instance data between data models and vocabularies (RDF-to-RDF)
  - A/k/a semantic alignment, model alignment

Crowd-Sourced Translation Rules Hub

Source

1. Lift to RDF

2. Translate

3. Drop from RDF

Target

v2.5

v2.5
How should this translation be done?

- Translation is hard!
- Many different models and vocabularies
- Currently done in proprietary, black-box integration engines
Translation strategies

- Point-to-point is easier/faster for each translation
- Hub-and-spoke requires fewer translations: $O(n)$ instead of $O(n^2)$
- Hub-and-spoke requires a common data model
- Both strategies can be used!
Which common data model?

Hub-and-Spoke

- Standardization may choose a common data model:
  - Moving target
  - Must be able to represent (but not require) the finest granularity needed by any use case
- Different use cases may use other data models, mapped to/from the common data model
  - Speeds standardization of common data model – Avoids bike shed effect
Where are these translation rules?

- By manipulating RDF data, rules can be mixed, matched and shared.
Needed:
Crowd-Sourced Translation Rules Hub

- Based on GitHub, WikiData, BioPortal, Web Protege or other
- Hosts translation rules
- Agnostic about "rules" language:
  - Any executable language that translates RDF-to-RDF (or between RDF and source/target syntax)
Translation rules metadata

- Source and target language / class
- Rules language
  - E.g. SPARQL/SPIN, N3, JenaRules, Java, Shell, etc.
- Dependencies
- Test data / validation
- License (free and open source)
- Maintainer
- Usage metrics/ratings
  - Objective: Number of downloads, Author, Date, etc.
  - Subjective: Who/how many like it, reviews, etc.
  - Digital signatures of endorsers?
Patient data privacy

- Download translation rules as needed – plug-and-play
- Run rules locally
  - Patient data is not sent to the rules hub
Roadmap for Interoperability
Roadmap

1. RDF as a Universal Information Representation
2. RDF Mappings
3. Translations between models & vocabularies
4. Crowd-Sourced Translation Rules
5. RDF/OWL Standards Definitions
6. Collaborative Standards Convergence
7. Interoperability Policies

Semantic Interoperability
Roadmap - 1

Use RDF as a common semantic foundation

1. RDF as a Universal Information Representation
2. RDF Mappings
3. Translations between models & vocabularies
4. Crowd-Sourced Translation Rules
5. RDF/OWL Standards Definitions
6. Collaborative Standards Convergence
7. Interoperability Policies

Semantic Interoperability
Roadmap - 2

For common healthcare information representations*, define an **RDF mapping** to/from each format, data model and vocabulary – "lift" and "drop".

*Both standard and proprietary
Define **translation rules** for instance data that is expressed in RDF representations.
Create a hub for crowd-sourcing translation rules

1. RDF as a Universal Information Representation
2. RDF Mappings
3. Translations between models & vocabularies
4. Crowd-Sourced Translation Rules
5. RDF/OWL Standards Definitions
6. Collaborative Standards Convergence
7. Interoperability Policies

Semantic Interoperability
Roadmap - 5

Create **RDF / OWL definitions** of the data models and vocabularies defined by healthcare standards

1. RDF as a Universal Information Representation
2. RDF Mappings
3. Translations between models & vocabularies
4. Crowd-Sourced Translation Rules
5. RDF/OWL Standards Definitions
Create a collaborative standards hub for RDF/OWL standards definitions, to facilitate standards convergence.
Adopt **policy incentives** for healthcare providers to achieve semantic interoperability.
Roadmap - 7

(a) Adopt **free and open interoperability standards**.

**Why?**
Eliminate **IP barriers to interoperability**.
(b) Adopt **policy incentives** for healthcare providers to achieve semantic interoperability.

Why?

A healthcare provider has no **natural business incentive** to make its data interoperable with competitors.
1. RDF as a Universal Information Representation

2. RDF Mappings

3. Translations between models & vocabularies

4. Crowd-Sourced Translation Rules

5. RDF/OWL Standards Definitions

6. Collaborative Standards Convergence

7. Interoperability Policies

Semantic Interoperability
What will semantic interoperability cost?

My guesses . . .

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Ongoing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standards</td>
<td>$40-500M</td>
<td>+ $30-400M / year</td>
</tr>
<tr>
<td>Translations</td>
<td>$30-400M</td>
<td>+ $20-300M / year</td>
</tr>
<tr>
<td>Total</td>
<td>$60-900M</td>
<td>+ $50-700M / year</td>
</tr>
</tbody>
</table>

What are yours?
Opportunity cost

Non-interoperability

$30,000 Million per year*


Interoperability

$700 Million per year?
Biggest payoff opportunities

1. RDF as a Universal Information Representation
2. RDF Mappings
3. Translations between models & vocabularies
4. Crowd-Sourced Translation Rules
5. RDF/OWL Standards Definitions
6. Collaborative Standards Convergence
7. Interoperability Policies

Semantic Interoperability
Questions?
BACKUP SLIDES
Related Activities

• New HL7 group on "RDF for Semantic Interoperability":
  http://wiki.hl7.org/index.php?title=ITS_RDF_ConCall_Agenda

• ONC's "Interoperability Roadmap" (draft):
  http://tinyurl.com/mgtwwr8
A Roadmap for Healthcare Information Interoperability

1. RDF as a Universal Information Representation

2. Standardize the Standards in RDF

3. Lift to RDF

4. Collaborate standards convergence

5. Translations between models & vocabularies

6. Crowd-sourced translation rules

7. Interoperability incentives

Semantic Interoperability

http://YosemiteProject.org/
Steps 2 and 5

Existing Standard Definition

XML Schema + Text

Describes

XML Instance Data

RDF / OWL Standard Definition

RDF / OWL Ontology + Text

Describes

RDF Instance Data

Lift/Drop Mapping

<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>
Roadmap

1. RDF as a Universal Information Representation

Semantic Interoperability