Applications of the Semantic Web for Healthcare

Poughkeepsie ACM Chapter
November 16 2015

John Sheehan
Rensselaer Polytechnic Institute
Understanding...  
Consuming...  
Building...  
**Improving Healthcare** with...  

...the Web of Data
Objectives

• The **Semantic Web** "stack“ explained
• *Review* **Linked Data** principles & practices
• **Applications to Healthcare**
• **Research Projects at RPI**
Web Architecture

First Principles of The Web:

- A standard system for **identifying** resources
- Standard formats for **representing** resources
- A standard protocol for **exchanging** resources
Web Architecture

First Principles of The Web:
- A standard system for identifying resources
- Standard formats for representing resources
- A standard protocol for exchanging resources

Relevant core standards:
- URLs: Uniform Resource Locators
- HTML: Hypertext Markup Language
- HTTP: Hypertext Transfer Protocol
What's this "Semantic Web?"

...and where can I get one???
The Semantic Web [Preview]
A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities.

By Tim Berners-Lee, James Hendler and Ora Lassila | May 2001
Semantic Web?

Web of "meaning"

Web of Data

Linked Data

Linking ideas...

meaning:
ideas:
data
How does it work???
Semantic Web Building Blocks

**Applications**

- **Interaction**
  - XHTML
  - CSS
  - SVG
  - SMIL
  - CDF
  - XForms
  - MathML
  - InkML

- **Mobile Web**
  - XHTML Basic
  - Mobile SVG
  - SMIL SVG
  - XForms Mobile
  - CC/PP
  - DD

- **Voice**
  - VoiceXML
  - SRGS
  - SSML
  - CCXML
  - EMMA

- **Web Services**
  - SOAP
  - MTOM
  - WSDL
  - WS-CDL
  - Addressing

- **Semantic Web**
  - OWL
  - SKOS

- **Privacy, Security**
  - P3P
  - APPEL
  - XML Sig
  - XML Enc
  - XKMS

**Web Accessibility / Internationalization / Device Independence**

- XML, Namespaces, Schemas, XQuery/XPath, XSLT, DOM, XML Base, XPointer

**Quality Assurance**

- RDF/XML, SPARQL

**Web Architectural Principles**

- XML InfoSet, RDF Graph

**The Web Advancing to its Full Potential**

Internet
The Semantic Web Technology Stack (not a piece of cake...)

Most apps use only a subset of the stack
Querying allows fine-grained data access
Standardized information exchange is key
Formats are necessary, but not too important
The Semantic Web is based on the Web

Linked Data uses a small selection of technologies

Identifying Web Resources

- A **global identification system** is essential...
  - ...to share information about resources
  - ...to reason about resources
  - ...to modify or exchange resources
- **Resources** are anything that can be linked to or spoken of
  - Documents, cat videos, people, ideas...
- Not all resources are *on* the Web
  - They might be referenced from the Web...
  - ...while not being retrievable from it
  - These are (so called) **information resources**

Les Carr, et.al. [http://slidesha.re/142MFrV](http://slidesha.re/142MFrV)
A Uniform Resource Identifier (URI) is used to identify the name of a resource. Enables interaction with representations of the resource over the Web.
Identifying Web Resources

● A global standard is required; the URI is it
● Others systems are possible...
  ○ ...but added value of a **single global system** of identifiers is high
  ○ Enables linking, bookmarking and other functions across **heterogeneous applications**
● How are URIs used?
  ○ All resources have URIs associated with them
  ○ Each URI identifies a single resource in a context-independent manner
  ○ URIs act as names and (usually) addresses
  ○ In general URIs are "opaque"

Identifying Web Resources

- "URIs identify and URLs locate..."
  - ...and identify
- URLs are URIs aligned with protocols
  - URLs include the "access mechanism" or "network location"
    - e.g. http:// or ftp://
  - How to "dereference" the URI and retrieve the thing
- URL examples
  - ftp://ftp.is.co.za/rfc/rfc1808.txt
  - mailto:John.Doe@example.com
  - telnet://192.0.2.16:80/

Representing Resources

- Resources are manifest as digital files
  - More precisely: *serializations* that look like files...
- The Web recognizes a (growing) set of \{file | serialization\} formats
  - The original and workhorse is HTML...
  - ...but there are many others
- Retrievable resources on the web serve multiple purposes
  - Resources *encode* information and data
  - Resources *aggregate links* to other resources
- This is what makes The Web(tm) a "web..."
Resources (nodes) aggregate links to other resources to create a Web
Retrieving Resources

- **Review:** URLs refer to retrievable resources
  - ie URIs that specify some *protocol* for retrieval
- The original and most common Web protocol is **HTTP**
- Specialized protocols are possible but resources may appear "off the grid..."
- More common case is HTTP w different formats...
RDF: Resource Description Framework
RDF: Resource Description Framework

subject

predicate

object

“article”

“has creator”

“James Hendler”
RDF: Resource Description Framework

http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/
We're missing something...

*Check:* **URIs for names:** S, P, O can be URIs  
*Check:* **HTTP URIs:** all of our examples are resolvable

*Now:* "Return something useful" when we resolve URIs  
*Format:* How do we *serialize* RDF?  
*Protocol:* How do we *retrieve* RDF?

*Let's go to the graph...*
Source: Programming the Web
http://bit.ly/1aZwr40
"Raw" Triples

### Triples of the Data Model

<table>
<thead>
<tr>
<th>Number</th>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><a href="http://kiwitobes.com/toby.rdf#ts">http://kiwitobes.com/toby.rdf#ts</a></td>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a></td>
<td><a href="http://xmlns.com/foaf/0.1/Person">http://xmlns.com/foaf/0.1/Person</a></td>
</tr>
<tr>
<td>2</td>
<td><a href="http://kiwitobes.com/toby.rdf#ts">http://kiwitobes.com/toby.rdf#ts</a></td>
<td><a href="http://xmlns.com/foaf/0.1/name">http://xmlns.com/foaf/0.1/name</a></td>
<td>&quot;Toby Segaran&quot;</td>
</tr>
<tr>
<td>4</td>
<td><a href="http://kiwitobes.com/toby.rdf#ts">http://kiwitobes.com/toby.rdf#ts</a></td>
<td><a href="http://xmlns.com/foaf/0.1/nick">http://xmlns.com/foaf/0.1/nick</a></td>
<td>&quot;kiwitobes&quot;</td>
</tr>
<tr>
<td>5</td>
<td><a href="http://kiwitobes.com/toby.rdf#ts">http://kiwitobes.com/toby.rdf#ts</a></td>
<td><a href="http://xmlns.com/foaf/0.1/mbox">http://xmlns.com/foaf/0.1/mbox</a></td>
<td><a href="mailto:toby@segaran.com">mailto:toby@segaran.com</a></td>
</tr>
<tr>
<td>9</td>
<td><a href="http://semprog.com/people/colin">http://semprog.com/people/colin</a></td>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a></td>
<td><a href="http://xmlns.com/foaf/0.1/Person">http://xmlns.com/foaf/0.1/Person</a></td>
</tr>
<tr>
<td>11</td>
<td><a href="http://semprog.com/people/colin">http://semprog.com/people/colin</a></td>
<td><a href="http://xmlns.com/foaf/0.1/name">http://xmlns.com/foaf/0.1/name</a></td>
<td>&quot;Colin Evans&quot;</td>
</tr>
<tr>
<td>12</td>
<td><a href="http://semprog.com/people/colin">http://semprog.com/people/colin</a></td>
<td><a href="http://xmlns.com/foaf/0.1/mbox">http://xmlns.com/foaf/0.1/mbox</a></td>
<td><a href="mailto:colin@semprog.com">mailto:colin@semprog.com</a></td>
</tr>
<tr>
<td>13</td>
<td>genid:A708592</td>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a></td>
<td><a href="http://xmlns.com/foaf/0.1/Person">http://xmlns.com/foaf/0.1/Person</a></td>
</tr>
<tr>
<td>15</td>
<td>genid:A708592</td>
<td><a href="http://xmlns.com/foaf/0.1/name">http://xmlns.com/foaf/0.1/name</a></td>
<td>&quot;Jamie Taylor&quot;</td>
</tr>
<tr>
<td>16</td>
<td>genid:A708592</td>
<td><a href="http://xmlns.com/foaf/0.1/mbox">http://xmlns.com/foaf/0.1/mbox</a></td>
<td><a href="mailto:jamie@semprog.com">mailto:jamie@semprog.com</a></td>
</tr>
</tbody>
</table>

Via the W3C RDF VValidator Service:
http://www.w3.org/RDF/Validator/
N-Triples...

<http://kiwitobes.com/toby.rdf#ts> <http://xmlns.com/foaf/0.1/name> "Toby Segaran".


_:jamie <http://xmlns.com/foaf/0.1/name> "Jamie Taylor".
_:jamie <http://xmlns.com/foaf/0.1/mbox> <mailto:jamie@semprog.com>.
_:jamie <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person>.


@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix semperp: <http://semprog.com/people/>.
@prefix tobes: <http://kiwitobes.com/toby.rdf#>.


tobes:ts a foaf:Person;
  foaf:homepage <http://kiwitobes.com/>;
  foaf:interest <http://semprog.com>;
  foaf:knows semperp:colin,
    [ a foaf:Person;
      foaf:mbox <mailto:jamie@semprog.com>;
      foaf:name "Jamie Taylor" ];
  foaf:mbox <mailto:toby@segaran.com>;
  foaf:name "Toby Segaran";
  foaf:nick "kiwitobes".

<http://semprog.com> a foaf:Document;
  rdfs:label "Semantic Programming".

semperp:colin a foaf:Person;
  foaf:mbox <mailto:colin@semprog.com>;
  foaf:name "Colin Evans".

N3...
RDFa...


Name: <span property="foaf:name">Toby Segaran</span><br/>
Nickname: <span property="foaf:nick">kiwitobes</span><br/>
Interests: <a rel="foaf:interest" href="http://semprog.org">
  <span property="rdfs:label">Semantic Programming</span></a>
Homepage: <a rel="foaf:homepage" href="http://kiwkitobes.com/">KiwiTobes</a><p/>

Friends:<br/>
<ul rel="foaf:knows">
  <li about="http://semprog.com/people/colin" property="foaf:name">Colin Evans</li>
  <li name="Jamie Taylor" typeof="foaf:Person">
    Email: <a rel="foaf:mbox" href="mailto:jamie@semprog.com">jamie@semprog.com</a><br/>
  </li>
</ul>
</div>:
<rdf:RDF
xmlns:foaf='http://xmlns.com/foaf/0.1/
xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#
xmlns:rdfs='http://www.w3.org/2000/01/rdf-schema#'>

<foaf:Person rdf:about="http://kiwitobes.com/toby.rdf#ts">
  <foaf:name>Toby Segaran</foaf:name>
  <foaf:nick>kiwitobes</foaf:nick>
  <foaf:mbox rdf:resource="mailto:toby@segaran.com"/>
  
  <foaf:interest>
    <foaf:Document rdf:about="http://semprog.com">
      <rdfs:label>Semantic Programming</rdfs:label>
    </foaf:Document>
  </foaf:interest>

  <foaf:knows>
    <foaf:Person rdf:about="http://semprog.com/people/colin">
      <foaf:name>Colin Evans</foaf:name>
      <foaf:mbox rdf:resource="mailto:colin@semprog.com"/>
    </foaf:Person>
  </foaf:knows>

  <foaf:knows>
    <foaf:Person>
      <foaf:name>Jamie Taylor</foaf:name>
      <foaf:mbox rdf:resource="mailto:jamie@semprog.com"/>
    </foaf:Person>
  </foaf:knows>

</foaf:Person>
</rdf:RDF>
RDF Serialization: Summary

**N-Triples**: Verbose, "pedagogical"
**N3**: Concise, in common use;
**RDFa**: Commonly used for embedded RDF
**RDF/XML**: Some use in government & "enterprise"
**JSON-LD**: Fast-rising LD standard
**RDF/JSON**: Older convention for LD applications
Things we still haven't discussed...

How to **retrieve** this "linked data" of which I speak
How (and where?) to query RDF "graphs"
How to use LD in applications
How to create visualizations & "mashups"

Also:
How to create and publish linked data...
Consuming Linked Data

Querying RDF using SPARQL
Endpoints and triple stores
Mashing data in the query
Mashing data in the application
URIs, HTTP, many formats...

http://www.w3.org/2006/Talks/0521-sb-AC-management/ReCTechStack-bg.png
Principles for creating a healthy Web

Use **URIs** as *names* for things

Use **HTTP URIs** so people can *look up* those names

When someone *looks up* a URI, return useful information

use standard representation formats to express it

Include links to other URIs, so consumers can discover more things

By "consumers" we mean *people* or *applications*

**Why is linking important??**

Tim Berners-Lee [http://www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)
Implications of a well-connected Web

Links to other nodes are a "vote" of quality and/or relevance

Google PageRank

PageRank https://en.wikipedia.org/wiki/PageRank
Perspective: *Consuming the Web of Data*

1. From the *Web of Documents* to the *Web of Data*

2. *Linked Data*: Building Blocks of the Web (of Data)

3. *Mashups*: Consuming Linked Data
Open Linked Data Gains Momentum

Tim Berners Lee's TED talk on Linked Data
SPARQL: pattern matching over RDF graphs

SPARQL Query Language for RDF
W3C Recommendation 15 January 2008

This version: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/
Latest version: http://www.w3.org/TR/rdf-sparql-query/
Previous version: http://www.w3.org/TR/2007/PR-rdf-sparql-query-20071112/
Editors:
Eric Prud'hommeaux, W3C <eric@w3.org>
Andy Seaborne, Hewlett-Packard Laboratories, Bristol <andy.seaborne@hp.com>
Improving the Bart blackboard query.

Since I first wrote on sending DBpedia SPARQL queries about Bart's blackboard messages at the start of Simpsons episodes, I've learned a lot more about SPARQL (reading the spec helped) and I wanted to walk through some of the things I've learned by expanding on and refining my original query.

I had finished that entry by wondering how to list Bart's blackboard entries for all episodes instead of for just one season. Vaclav Synacek showed me one way and I recently realized that there's a much simpler way—maybe too simple.

(all queries shown assume the namespace declarations shown on the SNORQL interface form for sending SPARQL queries to DBpedia):

```
SELECT ?blackboard WHERE {
}
```
Improving the Bart blackboard query.

Since I first wrote on sending DBpedia SPARQL queries about Bart's blackboard messages at the start of Simpsons episodes, I’ve learned a lot more about SPARQL (reading the spec helped) and I wanted to walk through some of the things I’ve learned by expanding on and refining my original query.

I had finished that entry by wondering how to list Bart’s blackboard entries for all episodes instead of just one season. Vaclav Synacek showed me one way, and I recently realized that there’s much simpler way—maybe too simple (all queries shown assume the namespace declarations shown on the SNORQL site).

```
SELECT ?blackboard WHERE {
}
```

Results: Browse Go! Reset
Improving the Bart blackboard query.

Since I first wrote on sending DBpedia SPARQL queries about Bart's blackboard messages at the start of Simpsons episodes, I’ve learned a lot more about SPARQL (reading the spec helped) and I wanted to walk through some of the things I’ve learned by expanding on and refining my original query.

I had finished that entry by wondering how to list Bart's blackboard entries for all episodes instead of just for just one season. I changed the namespace declaration (all queries shown assume the namespace declaration shown on the SNORQL interface)

```sparql
SELECT ?blackboard WHERE {
}
```

Results: Browse ▼ Go! Reset
"There is no such month as "Rocktober.""
"Double, Double, Boy in Trouble" is the third episode of The Simpsons' nineteenth season, and first aired on October 19, 2008. In this episode, the two decide to switch homes; Simon enjoys his time with the Simpsons while Bart discovers his new neighbor, football great Joe Montana, guest stars as himself.

Property | Value
--- | ---
Abstract | "Double, Double, Boy in Trouble" is the third episode of The Simpsons' nineteenth season, and first aired on October 19, 2008. In this episode, the two decide to switch homes; Simon enjoys his time with the Simpsons while Bart discovers his new neighbor, football great Joe Montana, guest stars as himself.

Thumbail | ![Double, Double, Boy in Trouble](http://upload.wikimedia.org/wikipedia/commons/thumb/d/0/Double_double_boy_in_trouble.png)

Properties:
- `episodeName`: `Double, Double, Boy in Trouble`
- `director`: `Michael Poonsino`
- `guestStar`: `Joe Montana`
- `scriptBy`: `Duff McGage`
- `aired`: `2008-10-19` (xsd:date)
- `date`: `2008-10-19` (xsd:date)
- `blackboard`: `There is no such month as "Rocktober".`
- `couchGag`: "In a parody of The Wizard of Oz, the family gets in a tornado accident.
- `episodeNumber`: 423 (xsd:integer)
- `inStory`: `Bart meets his double.`
"Double, Double, Boy in Trouble" is the third episode of the The Simpsons' twentieth season, and first aired on January 19, 2008. In this episode, the two decide to switch homes; Simon enjoys his time with the Simpsons while Bart discovers his new roommate is a football great Joe Montana guest stars as himself.
LOGD Tutorial SPARQL Query

# this query returns the agi and population data from two data.gov datasets

SELECT distinct ?state_abbv ?agi ?population
WHERE {
  }
  }
} order by ?state_fipscode
### SPARQL Query Results

<table>
<thead>
<tr>
<th>state_abbv</th>
<th>agi</th>
<th>population</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;AL&quot;</td>
<td>&quot;92162773&quot;</td>
<td>&quot;4599030&quot;</td>
</tr>
<tr>
<td>&quot;AK&quot;</td>
<td>&quot;17312636&quot;</td>
<td>&quot;670053&quot;</td>
</tr>
<tr>
<td>&quot;AZ&quot;</td>
<td>&quot;134442007&quot;</td>
<td>&quot;6166318&quot;</td>
</tr>
<tr>
<td>&quot;AR&quot;</td>
<td>&quot;49783294&quot;</td>
<td>&quot;2810872&quot;</td>
</tr>
<tr>
<td>&quot;CA&quot;</td>
<td>&quot;913619942&quot;</td>
<td>&quot;36457549&quot;</td>
</tr>
<tr>
<td>&quot;CO&quot;</td>
<td>&quot;128175529&quot;</td>
<td>&quot;4753377&quot;</td>
</tr>
<tr>
<td>&quot;CT&quot;</td>
<td>&quot;122697142&quot;</td>
<td>&quot;3504809&quot;</td>
</tr>
<tr>
<td>&quot;DE&quot;</td>
<td>&quot;22983204&quot;</td>
<td>&quot;853476&quot;</td>
</tr>
<tr>
<td>&quot;DC&quot;</td>
<td>&quot;18177370&quot;</td>
<td>&quot;581530&quot;</td>
</tr>
<tr>
<td>&quot;FL&quot;</td>
<td>&quot;429785960&quot;</td>
<td>&quot;18089888&quot;</td>
</tr>
<tr>
<td>&quot;GA&quot;</td>
<td>&quot;199864840&quot;</td>
<td>&quot;9363941&quot;</td>
</tr>
<tr>
<td>&quot;HI&quot;</td>
<td>&quot;30592983&quot;</td>
<td>&quot;1285498&quot;</td>
</tr>
<tr>
<td>&quot;ID&quot;</td>
<td>&quot;30292717&quot;</td>
<td>&quot;1406465&quot;</td>
</tr>
<tr>
<td>&quot;IL&quot;</td>
<td>&quot;339217881&quot;</td>
<td>&quot;12831970&quot;</td>
</tr>
<tr>
<td>&quot;IN&quot;</td>
<td>&quot;140610570&quot;</td>
<td>&quot;6313520&quot;</td>
</tr>
<tr>
<td>&quot;IA&quot;</td>
<td>&quot;68946837&quot;</td>
<td>&quot;2982085&quot;</td>
</tr>
<tr>
<td>&quot;KS&quot;</td>
<td>&quot;65216515&quot;</td>
<td>&quot;2764075&quot;</td>
</tr>
<tr>
<td>&quot;KY&quot;</td>
<td>&quot;81721206&quot;</td>
<td>&quot;4206074&quot;</td>
</tr>
<tr>
<td>&quot;LA&quot;</td>
<td>&quot;84029967&quot;</td>
<td>&quot;4287768&quot;</td>
</tr>
<tr>
<td>&quot;ME&quot;</td>
<td>&quot;28954363&quot;</td>
<td>&quot;1321574&quot;</td>
</tr>
<tr>
<td>&quot;MD&quot;</td>
<td>&quot;168047138&quot;</td>
<td>&quot;5015727&quot;</td>
</tr>
<tr>
<td>&quot;MA&quot;</td>
<td>&quot;202226349&quot;</td>
<td>&quot;6437191&quot;</td>
</tr>
<tr>
<td>&quot;MI&quot;</td>
<td>&quot;227233854&quot;</td>
<td>&quot;10095643&quot;</td>
</tr>
<tr>
<td>&quot;MN&quot;</td>
<td>&quot;143482070&quot;</td>
<td>&quot;5167110&quot;</td>
</tr>
<tr>
<td>&quot;MS&quot;</td>
<td>&quot;47387966&quot;</td>
<td>&quot;2910540&quot;</td>
</tr>
<tr>
<td>&quot;MO&quot;</td>
<td>&quot;131166510&quot;</td>
<td>&quot;5842712&quot;</td>
</tr>
<tr>
<td>&quot;MT&quot;</td>
<td>&quot;20045504&quot;</td>
<td>&quot;944632&quot;</td>
</tr>
<tr>
<td>&quot;NE&quot;</td>
<td>&quot;41569440&quot;</td>
<td>&quot;1768331&quot;</td>
</tr>
<tr>
<td>&quot;NV&quot;</td>
<td>&quot;65272642&quot;</td>
<td>&quot;2495525&quot;</td>
</tr>
<tr>
<td>&quot;NH&quot;</td>
<td>&quot;38175000&quot;</td>
<td>&quot;1314895&quot;</td>
</tr>
<tr>
<td>&quot;NJ&quot;</td>
<td>&quot;283024874&quot;</td>
<td>&quot;8724506&quot;</td>
</tr>
<tr>
<td>&quot;NM&quot;</td>
<td>&quot;38144029&quot;</td>
<td>&quot;1954590&quot;</td>
</tr>
<tr>
<td>&quot;NY&quot;</td>
<td>&quot;513598458&quot;</td>
<td>&quot;19306133&quot;</td>
</tr>
<tr>
<td>&quot;NC&quot;</td>
<td>&quot;195374554&quot;</td>
<td>&quot;8856502&quot;</td>
</tr>
<tr>
<td>&quot;ND&quot;</td>
<td>&quot;14923738&quot;</td>
<td>&quot;635867&quot;</td>
</tr>
<tr>
<td>&quot;OH&quot;</td>
<td>&quot;259099675&quot;</td>
<td>&quot;11478066&quot;</td>
</tr>
<tr>
<td>&quot;OK&quot;</td>
<td>&quot;70394493&quot;</td>
<td>&quot;3579212&quot;</td>
</tr>
<tr>
<td>&quot;OR&quot;</td>
<td>&quot;85591882&quot;</td>
<td>&quot;3700758&quot;</td>
</tr>
<tr>
<td>&quot;PA&quot;</td>
<td>&quot;313289892&quot;</td>
<td>&quot;12440611&quot;</td>
</tr>
<tr>
<td>&quot;RI&quot;</td>
<td>&quot;205323233&quot;</td>
<td>&quot;1067610&quot;</td>
</tr>
<tr>
<td>&quot;SC&quot;</td>
<td>&quot;88615194&quot;</td>
<td>&quot;4321249&quot;</td>
</tr>
<tr>
<td>&quot;SD&quot;</td>
<td>&quot;17825580&quot;</td>
<td>&quot;781919&quot;</td>
</tr>
<tr>
<td>&quot;TN&quot;</td>
<td>&quot;126270760&quot;</td>
<td>&quot;6038802&quot;</td>
</tr>
<tr>
<td>&quot;TX&quot;</td>
<td>&quot;504386602&quot;</td>
<td>&quot;23507783&quot;</td>
</tr>
<tr>
<td>&quot;UT&quot;</td>
<td>&quot;55926179&quot;</td>
<td>&quot;2550063&quot;</td>
</tr>
<tr>
<td>&quot;VT&quot;</td>
<td>&quot;15246152&quot;</td>
<td>&quot;623908&quot;</td>
</tr>
<tr>
<td>&quot;WA&quot;</td>
<td>&quot;217677476&quot;</td>
<td>&quot;7642884&quot;</td>
</tr>
<tr>
<td>&quot;WV&quot;</td>
<td>&quot;175730868&quot;</td>
<td>&quot;6395798&quot;</td>
</tr>
<tr>
<td>&quot;WI&quot;</td>
<td>&quot;140513694&quot;</td>
<td>&quot;5556506&quot;</td>
</tr>
<tr>
<td>&quot;WY&quot;</td>
<td>&quot;15216840&quot;</td>
<td>&quot;515004&quot;</td>
</tr>
</tbody>
</table>
SPARQL Query results (JSON)


http://logd.tw.rpi.edu/demo/building-logd-visualizations/mashup-353-population-1356-agi.js
<table>
<thead>
<tr>
<th>state_abbv</th>
<th>agi</th>
<th>population</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>92162773</td>
<td>4599030</td>
</tr>
<tr>
<td>AK</td>
<td>134442007</td>
<td>6166318</td>
</tr>
<tr>
<td>AZ</td>
<td>199864840</td>
<td>9363941</td>
</tr>
<tr>
<td>AR</td>
<td>18177370</td>
<td>581530</td>
</tr>
<tr>
<td>CA</td>
<td>198536905</td>
<td>18089888</td>
</tr>
<tr>
<td>CO</td>
<td>128175529</td>
<td>4753377</td>
</tr>
<tr>
<td>CT</td>
<td>18177370</td>
<td>581530</td>
</tr>
<tr>
<td>DE</td>
<td>22983204</td>
<td>853476</td>
</tr>
<tr>
<td>FL</td>
<td>199864840</td>
<td>9363941</td>
</tr>
<tr>
<td>GA</td>
<td>30592983</td>
<td>1285498</td>
</tr>
<tr>
<td>ID</td>
<td>32092717</td>
<td>14664651</td>
</tr>
<tr>
<td>IL</td>
<td>339217881</td>
<td>12831970</td>
</tr>
<tr>
<td>IN</td>
<td>140616570</td>
<td>6313520</td>
</tr>
<tr>
<td>IA</td>
<td>68946837</td>
<td>2982085</td>
</tr>
<tr>
<td>KS</td>
<td>65216515</td>
<td>2764075</td>
</tr>
<tr>
<td>KY</td>
<td>81721206</td>
<td>4206074</td>
</tr>
<tr>
<td>LA</td>
<td>84029967</td>
<td>4287768</td>
</tr>
<tr>
<td>ME</td>
<td>28954363</td>
<td>1321574</td>
</tr>
<tr>
<td>MD</td>
<td>168647138</td>
<td>5615727</td>
</tr>
<tr>
<td>MA</td>
<td>202226349</td>
<td>6437193</td>
</tr>
<tr>
<td>MI</td>
<td>227233854</td>
<td>10095643</td>
</tr>
<tr>
<td>MN</td>
<td>143482070</td>
<td>5167101</td>
</tr>
<tr>
<td>MS</td>
<td>47387966</td>
<td>2910540</td>
</tr>
<tr>
<td>MO</td>
<td>131166510</td>
<td>5842713</td>
</tr>
<tr>
<td>MT</td>
<td>20045504</td>
<td>944632</td>
</tr>
<tr>
<td>NE</td>
<td>41569440</td>
<td>1768331</td>
</tr>
<tr>
<td>NV</td>
<td>65272642</td>
<td>2495529</td>
</tr>
<tr>
<td>NH</td>
<td>38175000</td>
<td>1314895</td>
</tr>
<tr>
<td>NC</td>
<td>195374554</td>
<td>8856505</td>
</tr>
<tr>
<td>ND</td>
<td>19423738</td>
<td>635867</td>
</tr>
<tr>
<td>OK</td>
<td>70394493</td>
<td>3579212</td>
</tr>
<tr>
<td>OR</td>
<td>85591882</td>
<td>3700758</td>
</tr>
<tr>
<td>PA</td>
<td>31328992</td>
<td>12440621</td>
</tr>
<tr>
<td>RI</td>
<td>26532233</td>
<td>1067610</td>
</tr>
<tr>
<td>SC</td>
<td>88615194</td>
<td>4321249</td>
</tr>
<tr>
<td>SD</td>
<td>17825580</td>
<td>781919</td>
</tr>
<tr>
<td>TN</td>
<td>126270760</td>
<td>6038803</td>
</tr>
<tr>
<td>TX</td>
<td>50438660</td>
<td>23507783</td>
</tr>
<tr>
<td>UT</td>
<td>55426179</td>
<td>2550063</td>
</tr>
<tr>
<td>WA</td>
<td>15246152</td>
<td>623908</td>
</tr>
<tr>
<td>WV</td>
<td>32423697</td>
<td>1818470</td>
</tr>
<tr>
<td>WI</td>
<td>140516394</td>
<td>5556506</td>
</tr>
</tbody>
</table>

http://logd.tw.rpi.edu/demo/building-logd-visualizations/mashup-353-population-1356-agi.js
Defining HTML Layout

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title>AGI per Capita Map</title>
</head>
<body>
  <div>AGI per Capita Map: average adjusted gross income per person in dollar amount in US states.</div>
  <div id='map_canvas'>Loading Map ...</div>
</body>
</html>

http://logd.tw.rpi.edu/tutorial/building_logd_visualizations
1. Load the appropriate Google Visualization API packages (in this case, the GeoMap package).

2. Define a callback function for loading visualization code, which is called upon the loading of the HTML page.

3. Obtain data from a given source to pass to our GeoMap instance. The Google Visualization API is designed to accept data in the form of specially-formatted JSON (represented by a URI) which can then be fed to a JSON processing function.

4. Following a call to the JSON processor, verify that it successfully processed the passed file.

5. Get back a response from the query processor, containing the data from the JSON file.

6. Define a data table to store the response data in. This process starts by defining header entries of the form `TABLE.addColumn(DATATYPE, NAME).

[http://logd.tw.rpi.edu/tutorial/building_logd_visualizations](http://logd.tw.rpi.edu/tutorial/building_logd_visualizations)
1. Load the appropriate Google Visualization API packages (in this case, the GeoMap package).

```javascript
// Load Google Visualization API
google.load('visualization', '1', { packages: ['geomap'] });
```

2. Define a callback function for loading visualization code, which is called upon the loading of the HTML page.

```javascript
// Set callback function for drawing visualizations
google.setOnLoadCallback(drawMap);
```

```javascript
function drawMap() {
  // Load static data
  var queryUrl = 'http://logd.tw.rpi.edu/tutorial/building_logd_visualizations/mashup.3S3.population.1356agination.js';
  var query = new google.visualization.Query(queryUrl);
  query.send(handleQueryResponse);
}
```

3. Obtain data from a given source in the form of specially-formatted JSON (represented by a URI) which can then be fed to a JSON processing function.

```javascript
function handleQueryResponse(response) {
  // Check for query response errors
  if (response.isError()) {
    alert('Error in query: ' + response.getMessage() + ' ' + response.getDetailedMessage());
    return;
  }
}
```

4. Following a call to the JSON processor, verify that it successfully processed the passed file.

```javascript
// Read data
var data = response.getDataTable();
```

5. Get back a response from the query processor, containing the data from the JSON file.

```javascript
// Create new data
var newData = new google.visualization.DataTable();
newData.addColumn('string', 'State');
newData.addColumn('number', 'AGI per Capita');
```

```javascript
// Populate each row
var rows = data.getNumberOfRows();
for (var i = 0; i < rows; i++) {
  var state = 'US.' + data.getValue(i, 0);
  var value = Math.round(data.getValue(i, 1) * 1000); // AGI figure uses thousand-dollar unit
  newData.addRow([state, value]);
}
```

6. Define a data table to store the response data in. This process starts by defining header entries of the form `TABLE.addColumn(DATATYPE, NAME)`.

```javascript
// Configure map options
var options = {};
options['region'] = 'US';  // Show US map
options['dataMode'] = 'regions';
options['width'] = 900;
options['height'] = 500;
```

```javascript
// Define GeoMap instance
var viz = document.getElementById('map_canvas');
new google.visualization.GeoMap(viz).draw(newData, options);
```

[http://logd.tw.rpi.edu/tutorial/building_logd_visualizations](http://logd.tw.rpi.edu/tutorial/building_logd_visualizations)
Dynamic Visualizations

Loading data using SPARQL queries

```javascript
//load data using SPARQL query
var sparqlproxy = "http://logd.tw.rpi.edu/ws/sparqlproxy.php";
var queryloc = "http://logd.tw.rpi.edu/demo/building-logd-visualizations/mashup-353-
population-1356-agi.sparql";
var service = "http://logd.tw.rpi.edu/sparql";
var queryurl = sparqlproxy
  + "?" + "output=gvds"
  + "&service-uri=" + encodeURICOMPONENT(service)
  + "&query-uri=" + encodeURICOMPONENT(queryloc) ;
```
Advanced Concepts

Vocabulary design/RDFS
Knowledge Organization
Ontology design
Provenance
Inference

Inference: Discovering New Relationships

On the Semantic Web, data is modeled as a set of (named) relationships between resources. **Inference** means using automatic procedures to generate **new relationships** based on the data...

...and some **additional information** in the form of a **vocabulary** or a **set of rules**

The new relationships may **explicitly added** to the set of data, or may be **returned at query time** (implementation issue)

The source of **additional information** is defined through vocabularies or rule sets
Both approaches draw upon **knowledge representation** techniques

**Ontologies** provide classification methods, putting an emphasis on defining 'classes', 'subclasses', on how individual resources can be associated to such classes, and characterizing the relationships among classes and their instances

**Rules** define mechanisms for discovering and generating new relationships based on existing ones, much like logic programs (Prolog)

In the Semantic Web toolkit, **RDFS**, **OWL**, or **SKOS** are used for defining ontologies

**RIF** covers rule based approaches
Vocabulary Design: W3C RDFS

RDF Vocabulary Description Language

RDF has no mechanism for:
- describing properties
- describing the relationships between properties and other resources

**RDF Schema** defines classes and properties for describing classes, properties and other resources

RDF Schema vocabulary descriptions are written in RDF

http://www.w3.org/TR/2004/REC-rdf-schema-20040210/
RDF Schema: Classes

rdfs:Resource ...is the class of everything
rdfs:Class ...declares a resource as a class for other resources
rdfs:Literal ...literal values such as strings and integers
rdfs:Datatype ...the class of datatypes
rdf:XMLLiteral ...the class of XML literal values
rdf:Property ...the class of properties

http://www.w3.org/TR/2004/REC-rdf-schema-20040210/
RDF Schema: Properties

- **rdfs:domain** ...declares the class of the subject in a triple whose second component is the predicate.
- **rdfs:range** ...declares the class or datatype of the object in a triple whose second part is the predicate.
  
  ex:employer rdfs:domain foaf:Person
  ex:employer rdfs:range foaf:Organization

- **rdf:type** ...state that resource is an instance of a class.

- **rdfs:subClassOf** ...allows to declare hierarchies of classes.
  
  e.g. "Every Person is an Agent": foaf:Person rdfs:subClassOf foaf:Agent

- **rdfs:subPropertyOf** ...states that all resources related by one property are also related by another

- **rdfs:label** ...used to provide a human-readable version of a resource's name

- **rdfs:comment** ...provides a human-readable description of a resource

- **rdfs:seeAlso** ...indicates a resource that might provide additional information about the subject resource.

- **rdfs:isDefinedBy** ...indicates a resource defining the subject resource

http://www.w3.org/TR/2004/REC-rdf-schema-20040210/
Knowledge Organization: W3C OWL

Web Ontology Language
RDFS too weak to describe resources in sufficient detail
   No localised range and domain constraints
      Can’t say that the range of hasChild is person when applied to persons and elephant
           when applied to elephants
   No existence/cardinality constraints
       Can’t say that all instances of person have a mother that is also a person, or that
           persons have exactly 2 parents
   No transitive, inverse or symmetrical properties
      Can’t say that isPartOf is a transitive property, that hasPart is the inverse of isPartOf or
           that touches is symmetrical

Difficult to provide reasoning support
   No “native” reasoners for non-standard semantics

Desirable features identified for Web Ontology Language:

- Extends existing Web standards
  - Such as XML, RDF, RDFS
- Easy to understand and use
  - Should be based on familiar KR* idioms
- Formally specified
- Of “adequate” expressive power
- Possible to provide automated reasoning support

KR* = knowledge representation

OWL Tools: Protege-OWL Editor

http://protege.stanford.edu/overview/protege-owl.html
Knowledge Organization: W3C SKOS

Simple Knowledge Organization System
An application of RDFS and OWL
Provides a way to represent controlled vocabularies, taxonomies and thesauri

**controlled vocabulary**: a list of terms which a community or organization has agreed upon

**taxonomy**: a controlled vocabulary organized in a hierarchy

**thesaurus**: a taxonomy with more information about each concept including preferred and alternative terms.

A thesaurus may also contain relationships to related concepts

SKOS is an OWL ontology; it can be written out in any RDF syntax

http://www.w3.org/2004/02/skos/ or http://slidesha.re/1etWDue or http://bit.ly/1etYLIE
Provenance: The W3C PROV Model

A set of W3C recommendations and notes on modelling provenance
PROV-O is the "core..."

http://www.w3.org/TR/prov-primer/
Provenance in a Nutshell

prov:Entity is a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary.

prov:Activity is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities.

prov:Agent is something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent’s activity.

These three classes provide a basis for the rest of PROV.

http://www.w3.org/TR/prov-primer/ or http://www.provbook.org/
Inference and W3C RIF

**Production Rules**
Analogous to instruction in a program: If a certain condition holds, then some action is carried out
Example: "If a customer has flown more than 100,000 miles, then upgrade him to Gold Member status."

**Declarative Rules**
Stating a fact about the world
Understood as sentences of the form "If P, then Q"
Example: "If a person is currently president of the United States of America, then his or her current residence is the White House."

There are many rule systems, esp. in the expert systems domain
The W3C **Rule Interchange Format** is an interchange format between existing rule systems

http://www.w3.org/TR/2013/NOTE-rif-primer-20130205/
The Future....
The Future (from the past)

Web 4.0 (2020 - 2030)
- Intelligent personal agents

Web 3.0 (2010 - 2020)
- Distributed Search
- Semantic Databases
- Widgets

Web 2.0 (2000 - 2010)
- Office 2.0
- Mashups
- Social Media Sharing
- Wikis
- SaaS
- Social Networking

Web 1.0 (1990 - 2000)
- Online Community
- Keyword Search
- Lightweight Collaboration

PC Era (1980 - 1990)
- Email
- FTP
- IRC
- USENET

Semantics of Social Connections

Semantics of Information Connections

How can the Semantic Web be used to improve Healthcare?

- Interoperability of electronic record systems
- Knowledge representation of medical experts
- Data analytics
Semantic interoperability is the ability to exchange data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems.

Semantic interoperability is therefore concerned not just with the packaging of data (syntax), but the simultaneous transmission of the meaning with the data (semantics).

This is accomplished by adding data about the data (metadata), linking each data element to a controlled, shared vocabulary. The meaning of the data is transmitted with the data itself, in one self-describing "information package" that is independent of any information system.

It is this shared vocabulary, and its associated links to an ontology, which provides the foundation and capability of machine interpretation, inferencing, and logic.

Syntactic interoperability is a prerequisite for semantic interoperability. Syntactic interoperability refers to the packaging and transmission mechanisms for data. In healthcare, HL7 has been in use for over thirty years to provide syntactic interoperability.

Interoperability of Electronic Health Records (EHRs)
Cleveland Clinic Use Case

Improve the Clinic’s ability to use patient data for generating new knowledge to improve future patient care through outcomes-based and longitudinal clinical research. **Leverage expressiveness and versatility of formats** to provide individual patients an appropriate terminology and accessible view of summary data.

Over 4 years, Cleveland Clinic has developed a **representational methodology** for bridging data collection, document management, and knowledge representation. The result is a unified content repository called **SemanticDB**.

http://www.w3.org/2001/sw/sweo/public/UseCases/ClevelandClinic/
Cleveland Clinic Use Case

• SemanticDB internally deployed for production on top of an open source XML & RDF content repository and Firefox (with extensions).
• Methodology realized through a core set of terms that facilitate creation of a domain vocabulary (or domain model) instances of the vocabulary managed automatically by the system.
• Patient records available as both uniform, structured markup and RDF.
• Coordinated use of both representation languages enables a variety of operations on patient record:
  • form-based data entry, transformation to reporting formats, document validation, targeted inference, and querying
  • Operations can be dispatched on the patient record documents and RDF graphs over a uniform set of interfaces.

http://www.w3.org/2001/sw/swoe/public/UseCases/ClevelandClinic/
Medical Expert System

Custom applications
- Auto-generated, forms-based data entry
- User management
- Case/quality workflow management
- "Controlled" vocabulary management
- Data export for analysis

Automated generation of data management components:
- Schema(s)
- Data dictionary
- Templates
- GRDDL transformations

Cyc Knowledge Base (~70K medically-related assertions: 32K about CT domain)

Cyc Analytical Environment
(Medical Expert System)

XML/RDF Content Repository API

XML Document

RDF Graph

Patient record ontology (OWL)

Longitudinal patient record Instance

Network Protocol Servers

SOAP, HTTP, FTP, SPARQL

- Nightly cases identified from hospital system(s)

http://www.w3.org/2001/sw/sweo/public/UseCases/ClevelandClinic/
What patients were diagnosed with a Hiatal Hernia after an operation that employed a femoral percutaneous surgical incision?

Simple Questions

The patient ID for __________ is __________.

__________ occurred in operations.

__________ is an operation.

__________ is a femoral percutaneous surgical incision.

__________ comes about during a femoral percutaneous surgical incision.

The type of surgical incision employed in ________ is femoral percutaneous surgical incisions.

During the event ________ the diagnosis indicates hiatal hernia.

__________ is before ________.

__________ contains ________.

__________ is a surgery.

__________ is the diagnosis indicator hiatal hernia.

Combine selected queries

Attempt to combine the two queries

http://www.w3.org/2001/sw/sweo/public/UseCases/ClevelandClinic/
Give the values of CCF-MANAGEMENT-EVENT, ELAPSED-TIME-712, SURGICAL-PROCEDURE, CCF-EVALUATION-EVENT and ID such that all of the following are true:

- The patient ID for PATIENT is ID.
- The type of surgical incision employed in CCF-MANAGEMENT-EVENT is femoral percutaneous surgical incisions.
- CCF-MANAGEMENT-EVENT is an operation.
- PATIENT is the patient involved in CCF-MANAGEMENT-EVENT.
- SURGICAL-PROCEDURE is a surgical procedure performed during CCF-MANAGEMENT-EVENT.
- During the event CCF-EVALUATION-EVENT the diagnosis indicates bilat hernia.
- PATIENT is the patient involved in CCF-EVALUATION-EVENT.
- CCF-MANAGEMENT-EVENT started ELAPSED-TIME-712 before or after CCF-EVALUATION-EVENT.
- ELAPSED-TIME-712 is less than 2 months.

CCF-EVALUATION-EVENT < 2 months

**between 1988 and 2006**

**CCF-MANAGEMENT-EVENT**

### Answers

<table>
<thead>
<tr>
<th>Ccf management event</th>
<th>Surgical procedure</th>
<th>Id</th>
<th>Ccf evaluation event</th>
</tr>
</thead>
</table>

**http://www.w3.org/2001/sw/sweo/public/UseCases/ClevelandClinic/**
Tetherless World Constellation

What is TWC?

The Tetherless World Constellation (TWC) at Rensselaer Polytechnic Institute (RPI) is a constellation of multidisciplinary researchers who study the scientific and engineering principles that underlie the Web, to enhance the Web’s reach beyond the desktop and laptop computer, and develops new technologies and languages that expand the capabilities of the Web under three themes: Future Web, Xinformatics and Semantic Foundations.

TWC goals include making the next generation web natural to use while being responsive to the growing variety of policy, educational, societal, and scientific needs. Research areas include: web science, privacy, intellectual property, general compliance, Web-based medical and health systems, semantic esence, data-science, semantic data frameworks, next generation virtual observatories, semantic data and knowledge integration, ontologies, semantic rules and query, semantic applications, data and information visualization, and knowledge provenance, trust and explanation for science. [More...]

TWC In Action

Tim Berners-Lee (center) congratulates TWC PhD recipients Greg Williams (left) and Jesse Weaver (right) during a March 2013 visit.

http://tw.rpi.edu
Rensselaer Polytechnic Institute Launches Initiative on Healthy Birth, Growth, and Development Knowledge: Semantic and Data Analytic Support

Rensselaer receives grant to integrate data sets from diverse, focused studies into a larger body of knowledge from the Bill & Melinda Gates Foundation

http://news.rpi.edu/content/2015/08/03/rensselaer-launches-initiative-healthy-birth-growth-and-development-knowledge
“By using advanced analytical, visualization, modeling, and semantic web informatics methodologies to extract information from multiple data sources, Rensselaer and HBGD researchers will assess causal relationships and quantify risk factors in order to develop novel hypotheses and targeted approaches to impact the health of children,”

“We are extremely pleased to receive this award from the Bill & Melinda Gates Foundation to address a pressing worldwide challenge. Rensselaer is in a unique position to collaboratively design best practice ontologies and refine them as needed by virtue of the expertise of inventors, data scientists, and researchers here who are experienced at using semantic web methodologies in numerous interdisciplinary science settings.” – RPI President Shirley Ann Jackson

http://news.rpi.edu/content/2015/08/03/rensselaer-launches-initiative-healthy-birth-growth-and-development-knowledge
Recommended Reading

What does Web Science ask?
Acknowledgement

Many of the slides for this presentation provided by:

John S. Erickson, Ph.D.
Director of Operations, The Rensselaer IDEA
Deputy Director, Web Science Research Center (RPI)
<http://tw.rpi.edu> erickj4@rpi.edu

Many Thanks John!

For Originals see: https://docs.google.com/presentation/d/1PdUomp-WqCBz4E2PnLQ2suI4HUsMLVlVzEzv8WWyl8/edit#slide=id.p
Thanks for Listening!