Linked Data: the view from 30,000 feet

Jodi Schneider
CARLI webinar
2019-02-07

jodi@illinois.edu
http://jodischneider.com/jodi.html
@jschneider
Hyperlinks
Paul Otlet’s Mundaneum

School of Information Sciences
The iSchool at Illinois
Van nevar Bush’s Memex
Douglas Engelbart & the Augmentation Research Center
The World Wide Web
Tim Berners Lee at CERN
Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control
The Web today

• 5 billion pages indexed
• 200 million domain names
• Largely commercial: 75% .com
Data
Linked Data
Linked Data: like a bag of chips
Built from many vocabularies

"If it was English English, it would say potato crisps"

American English
Built from many vocabularies

"Nutrition facts... are ... a little sublanguage understood by... the US food industry."

Data like:
- Serving size
- Servings per container
- Number of calories
Built from many vocabularies

"UPC is a language that is understood by all retailers, not just food industry, globally."
Built from many vocabularies

Information about the bag

“I ignore the stuff I don’t understand”
Linked Data: like a bag of chips

- Multiple different vocabularies co-exist
- Shared meaning arises from shared vocabularies
- “Cherry pick” the relevant vocabularies
- Structured data for different audiences, each can use what they recognize
Linked Data: like a bag of chips

Tim Berners-Lee
"Open, Linked Data for a Global Community"
Gov 2.0 Expo 2010
(minutes 4:58-5:51)

https://www.youtube.com/watch?v=ga1aSJXCFe0&feature=youtu.be&t=298
What’s different about Linked Data?
Linked Data:

- Can integrate information from multiple sources
- Focuses on bottom up agreement on terms and vocabularies
- Can be used to enrich data
- Can remove silos
- Focuses on linking DATA as opposed to DOCUMENTS
What is Linked Data?
Linked Data principles*

• Use web links to identify things
• Store useful info at the web links
  – info about the thing it identifies
  – relationships to other things

*Slightly simplifying
https://www.w3.org/DesignIssues/LinkedData.html
Kiwifruit (often shortened to kiwi) or Chinese gooseberry is the name given to the edible berries of several species of woody vines in the genus Actinidia. The most common cultivar group of kiwifruit (‘Hayward’) is oval, about the size of a large hen’s egg (5–8 cm (2.0–3.1 in) in length and 4.5–5.5 cm (1.8–2.2 in) in diameter). It has a fibrous, dull greenish-brown skin and bright green or golden flesh with rows of tiny, black, edible seeds. The fruit has a soft texture and a sweet but unique flavor. It is a commercial crop in several countries, such as Italy, New Zealand, Chile, Greece, and France.
kiwifruit (Q13194)
edible berry of several species of woody vines in the genus Actinidia, native to China
kiwi fruit | kiwi | Chinese gooseberry

### In more languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>kiwifruit</td>
<td>edible berry of several species of woody vines in the genus Actinidia, native to China</td>
<td>kiwi fruit, kiwi, Chinese gooseberry</td>
</tr>
<tr>
<td>Spanish</td>
<td>kiwi</td>
<td>fruta comestible</td>
<td></td>
</tr>
<tr>
<td>Traditional Chinese</td>
<td>奇異果</td>
<td>No description defined</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>奇異果</td>
<td>No description defined</td>
<td>猕猴桃</td>
</tr>
</tbody>
</table>

### All entered languages

#### Statements

- **instance of**
  - berry
    - 0 references
    - + add reference
Who is using Linked Data?
Linked Data Cloud
Linked Data Cloud
Examples of Linked Data
Mapping and Integrating Data
UBC Institute of Fisheries Field Records

Linked data web application providing a georeferenced visualization of UBC Open Collection Institute of Fisheries Field Records metadata, linked to Encyclopedia of Life.

Link for the application on CARTO.
Using the Web as a Database
Easy Apple Lattice Pie | Marie Callender's® Desserts

http://www.mariecallendersmeals.com

Savor The Made-From-Scratch Touches You'll Love To Share With Your Family Today!
Slow Down & Savor The Taste Of Home With Made-From-Scratch Touch Of Marie's!

Recipes

Total time

- <30 mins
- 30-60 mins
- >60 mins

Apple pie

All recipes

4.5/5 ★★★★★ (787) · 5 pins

Perfect Apple Pie Recipe - Pillsbury.com

https://www.pillsbury.com/recipes/perfect-apple-pie

Filling: In large bowl, gently mix filling ingredients; spoon into crust-lined pie plate. Top with second crust. Wrap excess top crust under bottom crust edge, pressing edges ... 4.5/5 ★★★★★ (787) · Servings: 8 · Total Time: 3 hrs

Category: Dessert

Place of origin: England

Nutrition facts

Apple pie

Amount Per 1 piece (1/6 of 8" pie) (117 g)

Calories 277

Total Fat 12.9 g · 20%

Saturated fat 4.4 g · 22%

Polyunsaturated fat 2.6 g

% Daily Value*
Recipe
Canonical URL: http://schema.org/Recipe

A recipe. For dietary restrictions covered by the recipe, a few common restrictions are enumerated via suitableForDiet. The keywords property can also be used to add more detail.

Usage: Between 10 and 100 domains

Properties from Recipe

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cookTime</td>
<td>Duration</td>
<td>The time it takes to actually cook the dish, in ISO 8601 duration format.</td>
</tr>
<tr>
<td>cookingMethod</td>
<td>Text</td>
<td>The method of cooking, such as Frying, Steaming, ...</td>
</tr>
<tr>
<td>nutrition</td>
<td>NutritionInformation</td>
<td>Nutrition information about the recipe or menu item.</td>
</tr>
<tr>
<td>recipeCategory</td>
<td>Text</td>
<td>The category of the recipe—for example, appetizer, entree, etc.</td>
</tr>
<tr>
<td>recipeCuisine</td>
<td>Text</td>
<td>The cuisine of the recipe (for example, French or Ethiopian).</td>
</tr>
</tbody>
</table>

https://schema.org/Recipe
Linked Data on Genetic Variation
The Monarch Initiative

Overview  Diseases  Phenotypes  Models  Genes

We have developed the Human Phenotype Ontology (HPO), a vocabulary to describe human disease features (phenotypes). The HPO now includes synonyms that patients, doctors, and machines can all understand.

Gene variations associated with disease

Diseases described with phenotypes

Diseases suggest models

Models suggest candidate genes

Plain language for describing human diseases

<table>
<thead>
<tr>
<th>Plain language</th>
<th>Medical term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Webbed toes</td>
<td>Syndactyly</td>
</tr>
<tr>
<td>Deformity due to premature fusing of skull bones</td>
<td>Craniosynostosis</td>
</tr>
<tr>
<td>Wide-set eyes</td>
<td>Ocular hypertelorism</td>
</tr>
</tbody>
</table>

View Announcement
Linked Data for Drug Discovery
Linked open drug data for pharmaceutical research and development
GrEDeL: A Knowledge Graph Embedding Based Method for Drug Discovery From Biomedical Literatures
Linked Data in Journalism
Effortlessly updating the Olympics and elections

Michael Phelps

Key Facts
Country
United States
Competed in
SWIMMING
Men's 100m Butterfly
Men's 200m Butterfly
Men's 200m Individual Medley
Men's 400m Individual Medley
Men's 4 x 100m Freestyle Relay
Men's 4 x 200m Freestyle Relay
Men's 4 x 100m Medley Relay
Date of birth
30 Jun 1985

Phelps through after scare
Michael Phelps, the 14-time gold medal winner, comes within 0.07 seconds of being eliminated from the men's 400m individual medley.
1 Jan 13

Best feeling in the world - Phelps
MEN'S 4 X 100M MEDLEY RELAY
5 Aug 12

Olympic great Michael Phelps describes finishing his career at the London Games in front of his family as "the best feeling in the world".

Phelps ends with 18th gold medal
MEN'S 4 X 100M MEDLEY RELAY

The BBC’s 2012 Olympics pages were generated using linked data

The BBC powered its delivery of the 2012 London Olympics with linked data: effortlessly updating hundreds of athlete, sport, and competition pages with all the recent related news.
We use linked data to curate indexes, such as constituency pages during the elections – imagine keeping 650 pages up to date at any given moment – or the massive repository of all our programmes.
We use linked data to curate indexes, such as constituency pages during the elections – imagine keeping 650 pages up to date at any given moment – or the massive repository of all our programmes.

Events and ontologies

Events can be big events or small events, and some ontologies nest these events – as if some of them could contain, or rather be composed of other events.

Take an election, for example. That’s an event itself, but there are also many smaller, other events that participate in the election:

- The campaign;
- A candidate saying something at a press conference;
- The results coming in...

A candidate touring a factory somewhere during their campaign is an event that we humans would see as linked to the bigger one: the election.

But it’s also linked to this factory’s story, and that’s a web again, because there are links everywhere.
Linked Data for Research Repositories
**Connect**
Duke University connects more than 170,000 journal articles, 2,000 books and 1,200 artistic works to REACH NC, a state-wide network of 19 organizations in North Carolina highlighting research and scholarship.

**Share**
Brown University uses VIVO to share accomplishments of faculty through integrated and automated profiles that include papers published, courses taught and grants awarded. Publication citations link to full text when available.

**Discover**
Cornell University uses VIVO to discover locations of research efforts using visualizations displaying VIVO data, helping scholars, students, administrators and the public find activity across the nation and world.
Thanks!

- Matthew Short & the CARLI Created Content Committee
- Amy Maroso at the CARLI Office
- Liliana Giusti Serra, UNSP Brazil & University of Illinois at Urbana Champaign
Upcoming Sessions

- **Linked Data Webinar 2: Linked Data in the Library**
  Tuesday Feb 19th, 10 to 11am

- **Linked Data Webinar 3: Linked Data Projects**
  Tuesday March 5th, 10 to 11:30am