

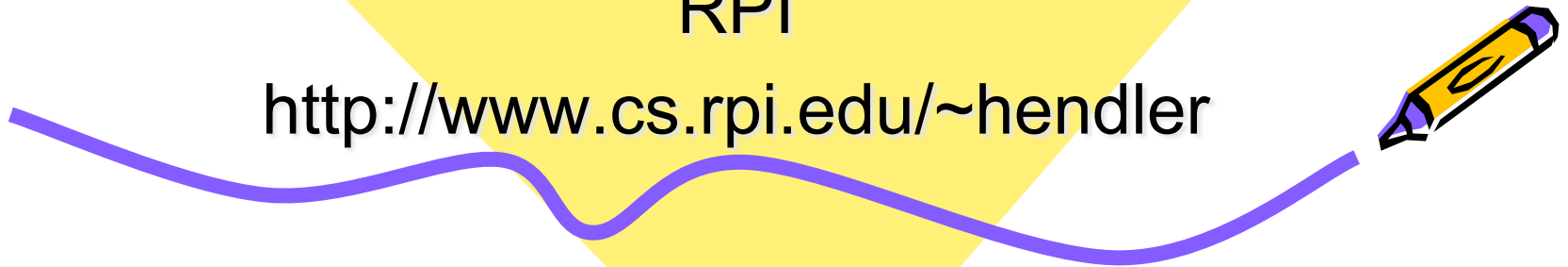


# The Semantic Web: Lighter, Faster, Easier

Jim Hendler

RPI

<http://www.cs.rpi.edu/~hendler>



# A myth that needs debunking

- ***The Semantic Web needs Ontologies (true)***
- ***But Ontologies are***
  - Inefficient (slow)
  - Complicated to express (Heavy)
  - Difficult to Build (Hard)
- ***(false)***
- We can build them:
  - ***Faster, Lighter, Easier!!***



# Traditional AI ontology

- *cf.* US National Center for Biotechnology Information, "Oncology Metathesaurus"
  - 50,000+ classes, ~8 people supporting full time, monthly updates, mandated for use by NIH-funded cancer researchers
    - OWL DL rigorously followed
    - Provably consistent



Rensselaer



# Sem Web use case

- *cf.* Friend of a Friend (Foaf)
  - 30+ classes, Dan Brickley and Libby Miller made it, maintained by consensus in a small community of developers
    - Violates DL rules (undecidable)
    - Used in many unexpected contexts
- FOAF
  - 10s of millions of Foaf people
    - (not necessarily distinct individuals)
  - Exported by a growing number of providers
    - If you use LiveJournal, you have a FOAF file
      - Also flickr, ecademy, tribe, joost, ...
      - Apps to export Foaf from Facebook and other soc netw sites
  - Becoming de facto standard for open social networking

A lot more users than the NCI ontology!



Rensselaer



# Why?

- NCBI view: Formal properties
  - Based on a decidable subset of KR
    - Description logics
  - For which much scaling research has been happening
    - *Ca. 2000 - 10,000 axioms, no facts, 1 day*
    - *Ca. 2008 - 50,000 axioms, million facts, 10 min.*
      - Not just faster computers (but Moore's Law helps), significant research into optimization, "average case"
      - Moving to parallel (Web server)
  - But still not "Web Scale"

In this view OWL is a formal *knowledge representation* standard



Rensselaer



# Ontology: the traditional view



- **Ontology as Barad-Dur (Sauron's tower):**

- Extremely powerful!

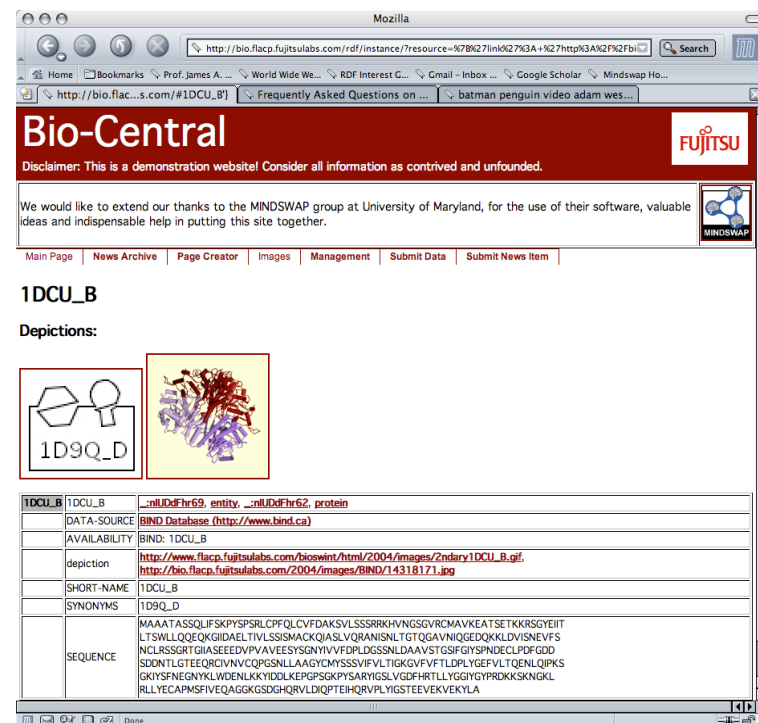
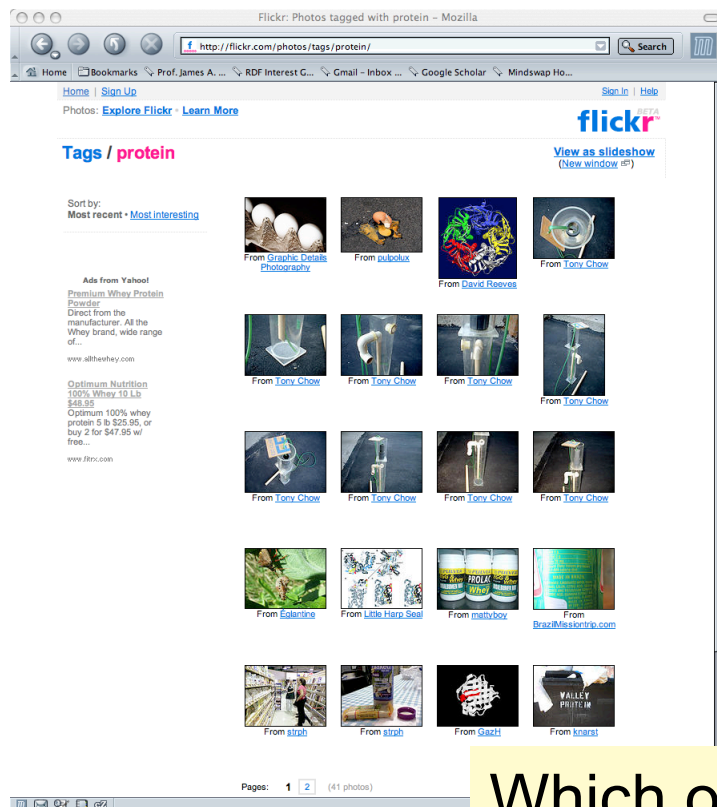
- Decidable Logic basis
    - ~~Patrolled by Orcs~~
      - inconsistency
      - Let one little ~~hobbit~~ in, and the whole thing could come crashing down



Rensselaer



# The argument for this seems compelling



Which one do you want *your* doctor to use?



Rensselaer



# But the cost is high

- Formal modeling finds its use cases in verticals and enterprises
  - Where the vocabulary can be controlled
  - Where finding things in the data is important
- Example
  - Drug discovery from data
    - **Model** the molecule (site, chemical properties, etc) as **faithfully** and expressively as possible
    - Use "Realization" to categorize data assets against the ontology
      - Bad or missed answers are money down the drain
- **The modeling is very expensive and the return on investment must be very high!**

Analogy: the pre-Web hypertext book



Rensselaer



# A better alternative for Web Development

- RDFS and OWL are based on RDF, a language designed for the (Semantic) Web
  - Built with Web architecture in mind
    - Exploits Web infrastructure, respects W3C TAG recommendations
      - Internationalization, accessibility, extensibility
  - Fits the **Web culture**
    - Open and extensible, supports communities of interest
      - *If you don't like my ontology, extend it, change it, or build your own*
    - Fits the Web application development paradigm
      - Scales like "databases"

Analogy: HTML


# Linked Data Web

- "Data Web" approach finds its use cases in Web Applications (at Web scales)
  - Finding anything in the mess can be a win!
- Which is different because
  - A lot of data, very little semantics
  - Used mainly for query (think Google, not Cyc)
    - not every answer must be right
    - *And time = money!*


# Very simple "reasoning"

- Twine recommends some people I may want to connect to
  - What is correctness in this case?
    - If I find some folks I like this way, I use twine more. Surprises can be fun.
  - I'm only seeing a few of a very large set so "first" is more important than "there somewhere"


**Recommended Members**




**Mills Davis**  
Washington DC USA  
83 Twines | 182 Items  
Connection Pending




**Chris Jones**  
All ready for '08  
Mill Valley  
58 Twines | 65 Items  
[Connect](#)



**John Clarke Mills**  
doing things and stuff  
San Francisco, CA  
28 Twines | 34 Items  
[Connect](#)



**Steve O'Donoghue**  
Twining my interests  
San Francisco  
27 Twines | 181 Items  
[Connect](#)




**tricia**  
arbiter of style  
san francisco, ca  
52 Twines | 952 Items  
[Connect](#)

twine



Home My Items My Twines My Connections Explore Start a Twine Search Twine Search

**Member Profile**





tricia  
arbiter of style  
san francisco, ca  
Member since: Oct 3, 2007  
<http://www.bitsandbobbins.com/journal>  
<http://www.wardroberemix.com>


**Mutual Twines**  
2 Twines

-  Web 3.0 - Semantic Web  
897 Items | 2218 Members
-  Twine News and ...  
47 Items | 3971 Members


**tricia's Twines**  
52 Twines

-  Web 3.0 - Semantic Web  
897 Items | 2218 Members
-  Twine News and ...  
47 Items | 3971 Members


**Mutual Connections**  
5 connections




**Candice Nobles**  
happy to be here!  
San Francisco, CA  
61 Twines | 59 Items  
Disconnected




**James Todd**  
living large  
95008  
72 Twines | 287 Items  
Disconnected



**Nora Spivack**  
Learning from Twine  
San Francisco, CA 94107  
284 Twines | 2338 Items  
Disconnected



**Hrafn Thorisson**  
A 1400g Clemson Jelly  
Reykjavik, Iceland  
233 Twines | 851 Items  
Disconnected

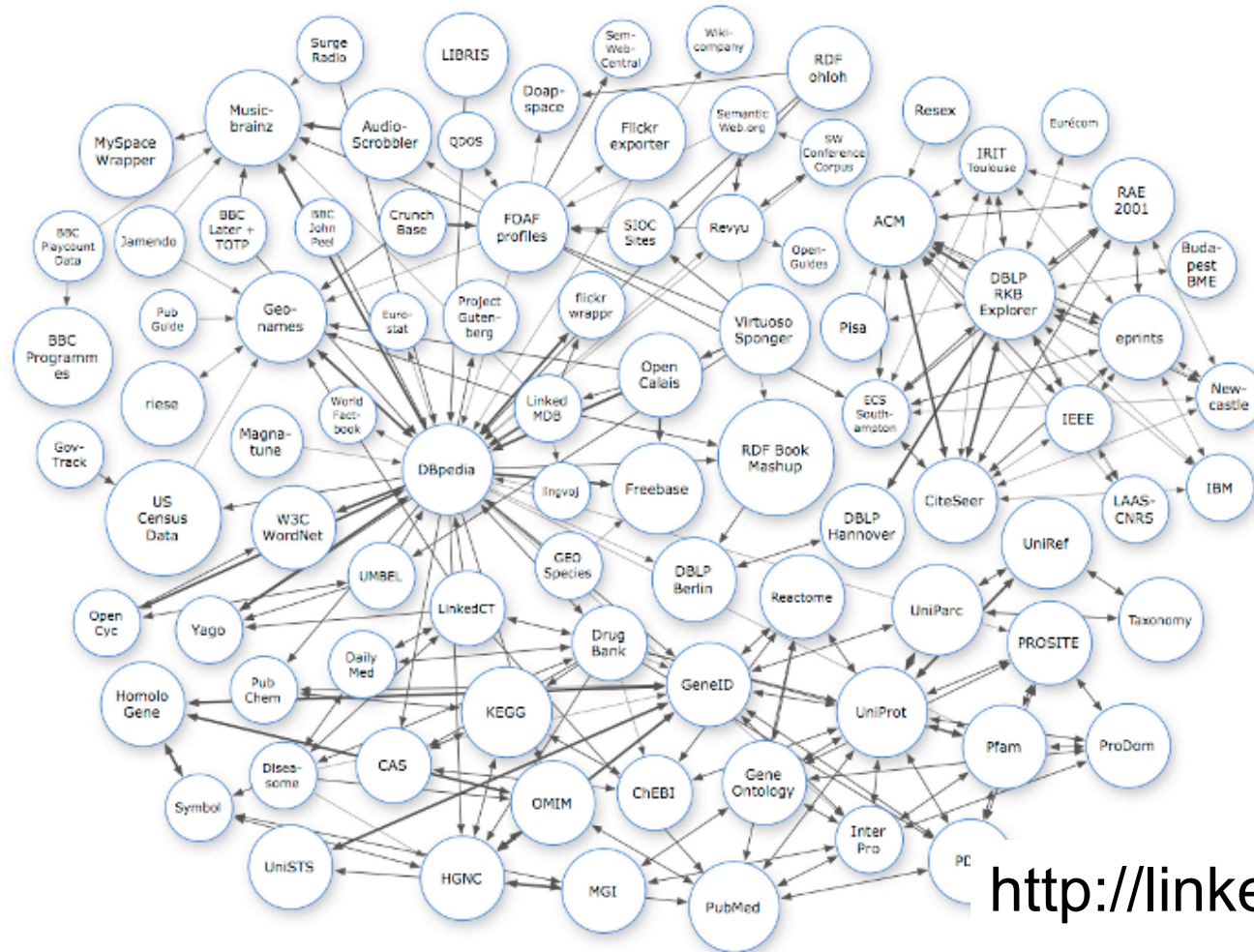


**Dan Perry**  
www.danperry.com  
34 Twines | 108 Items  
Disconnected



Rensselaer





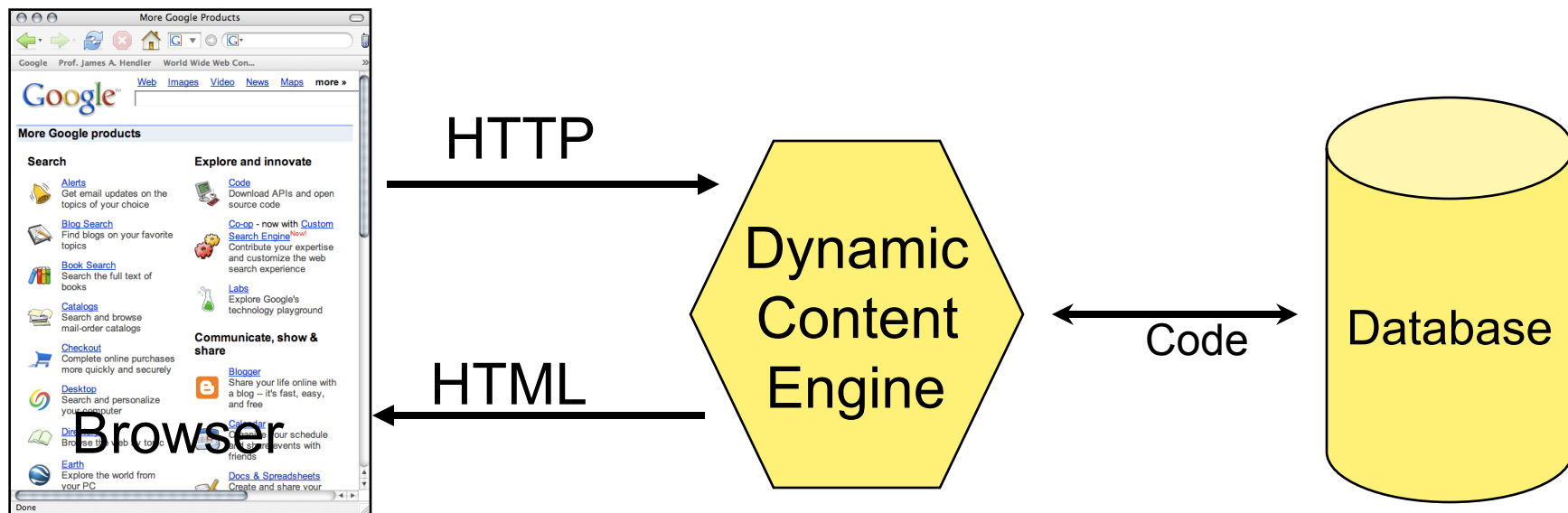
The linked open data cloud now has billions of assertions, and is growing rapidly



Rensselaer



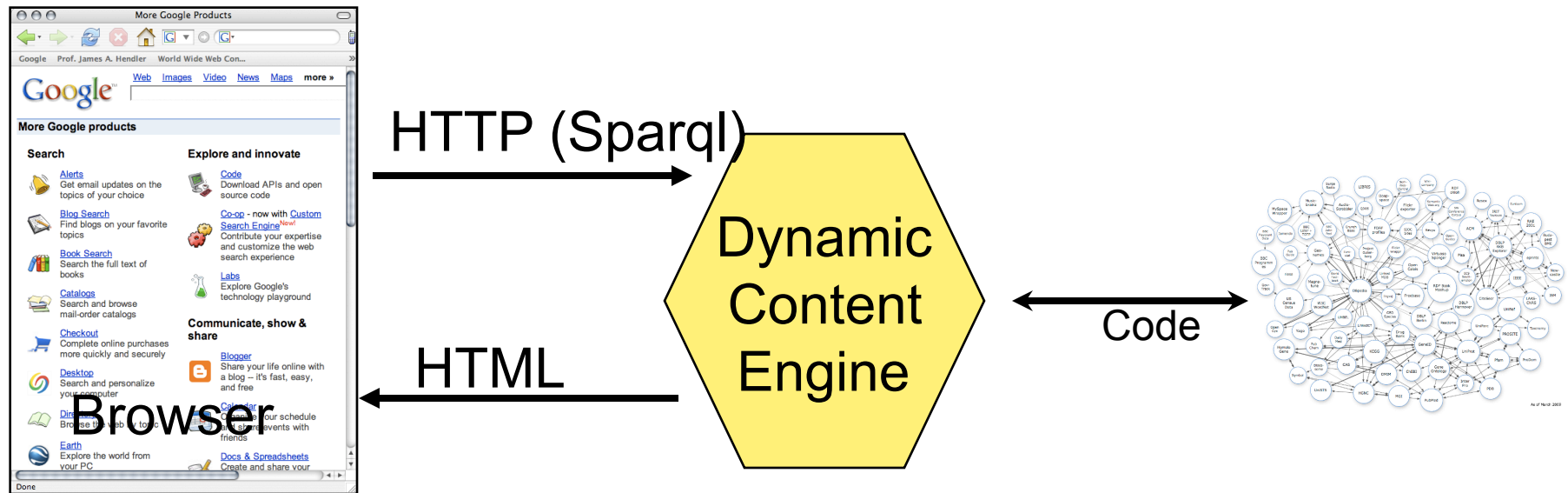
# Traditional Web applications



Rensselaer



# Semantic Web applications



Do your mashup on the underlying data  
instead of presentations thereof



Rensselaer



# Ontologies?

- Mostly reuse of a few simple ones (Dbpedia terms, foaf, doap, etc.)
  - Faster
- Uses simple parts of language (RDFS and a very small amount of OWL)
  - Lighter (sometimes called "lightweight ontologies")
- Mostly small and "local"
  - Easier



# Reasoning?

- Very little
  - Mainly just which data in one sphere is related to another
    - (easy)
  - Mainly based on small vocabularies
    - (Light)
  - Mainly procedural
    - (fast)



# Example LD applications



Dbpedia mobile



HealthFinland



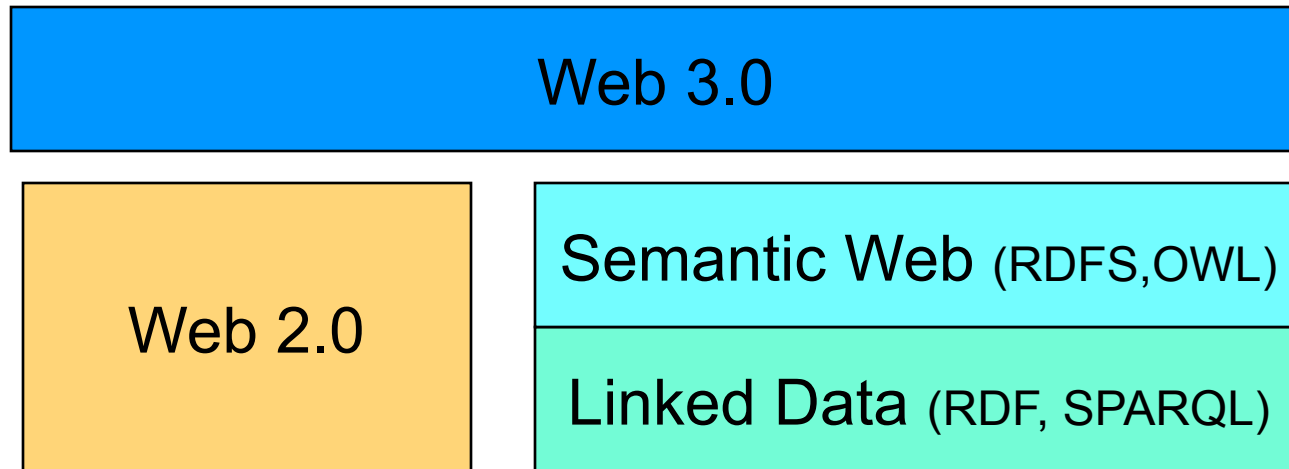
Semaplorer



Rensselaer



# The industrial "meme"



Web 3.0 extends current Web applications using Semantic Web technologies and graph-based, open data.



Rensselaer

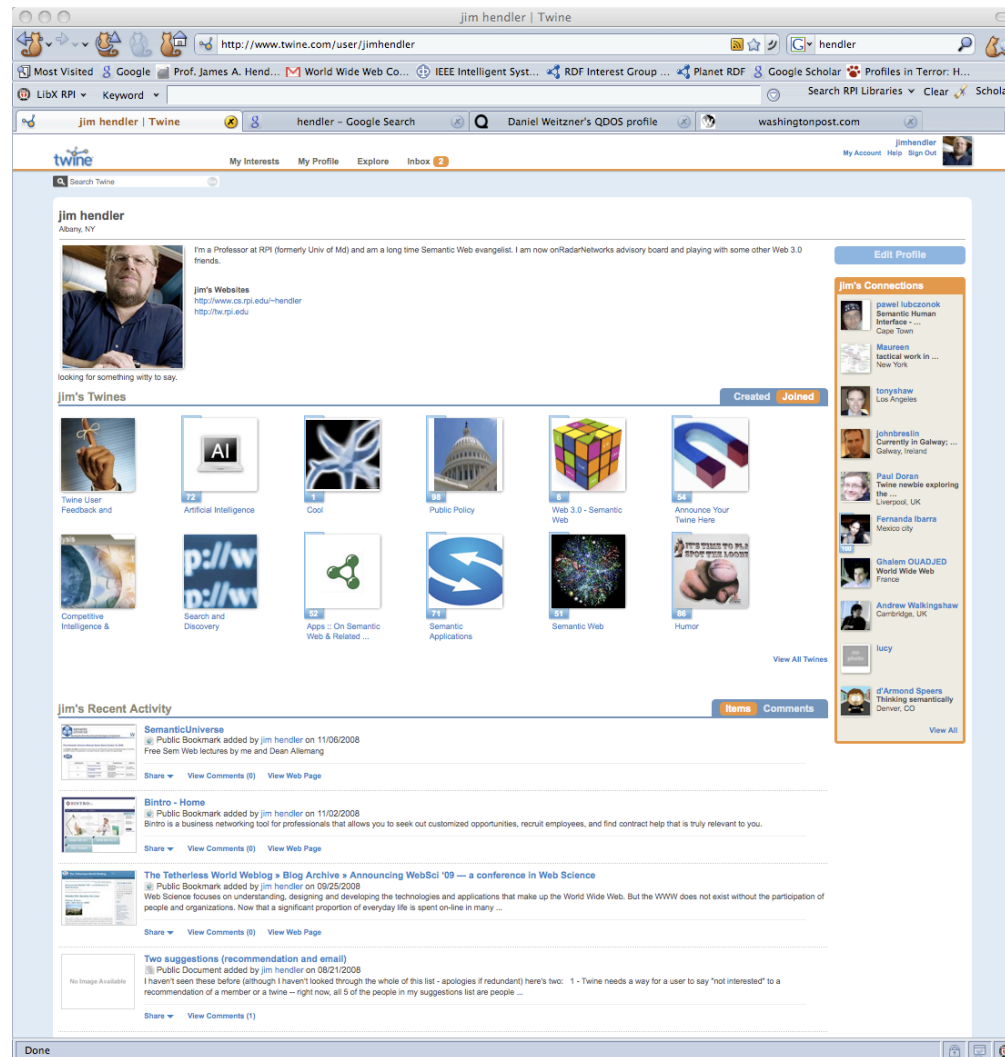


# Web 3.0 examples

The screenshot displays the Powerset.com Semantic Search interface. The browser window shows the URL <http://www.powerset.com/explore/go/james-hendler>. The search results for 'James Hendler' are presented in a structured, semantic format. A profile card for James Hendler is shown, including a photo, a brief biography, and key facts such as 'Date of Birth: 1957', 'Place of Birth: Queens', 'Profession: Computer scientist, Computer Science', and 'Religion: Jewish'. Below this, a section titled 'Facts from Wikipedia' lists various relationships, such as 'James Hendler developed SHOE' and 'James Hendler saw parallel'. The interface also includes a list of Wikipedia articles related to the search, with snippets of text from each article. The bottom of the page shows a 'Done' button and a status bar.

## Semantic Search (Powerset.com)

# Web 3.0 examples



Enhanced Social Networks (twine.com)



Rensselaer



# Web 3.0 examples

The screenshot displays the Bintro.com web application interface within a browser window. The browser's address bar shows the URL `http://bintro.com/yourbroadcasts.do?request_id=6430`. The page features a navigation bar with links for HOME, MESSAGES, MY BROADCASTS, COMPOSE BROADCAST, CONTACTS, and PROFILE. A user named James is logged in, with links for Sign Out, Account Settings, Help, and Contact Us. The main content area is titled "BROADCAST YOUR NEED OR OPPORTUNITY" and includes a description: "... to business professionals like yourself or manage and respond to those needs and/or opportunities you've previously Broadcast." Below this, there are two active broadcasts: "My Profile" (Broadcast on 28 Aug 2008) and "Web 3.0 Expertise" (Broadcast on 16 Jul 2008). The "Web 3.0 Expertise" broadcast is selected, showing a list of activity. The activity list has columns for Received, From, and Quality. One entry is visible: Received on 16 Jul 2008, from Adam Glick, with a link to "Check out my profile". Below the activity list, a detailed match for Adam Glick is shown. It includes a profile picture, name, location (New York, United States), and a match score of 75%. Other details include Position: Vice President and Industry: Commercial Mortgage Loans. There are buttons for "Add to Contacts" and "Send a Message To Adam". The footer of the page mentions "Paradigm5 Inc DBA Bintro © 2008" and provides links for Privacy Policy, Terms and Conditions, Security, and Contact Us.

Semantic Match (bintro.com)

# Web 3.0 examples

**Leicester topic** [rename](#) | [edit](#)

Also known as [City Of Leicester](#), [Leicester](#) [edit](#)

Leicester (pronounced [/ˈleɪstə/](#) [listen](#) [\(help·info\)](#)) is the largest city and unitary authority area in the East Midlands of England, and is the traditional county town of Leicestershire. Leicester lies on the River Soar and at the edge of the English National Forest. In 2004, the population of the city proper was estimated at 285,100, with 441,213 living in the urban area. It is currently, by population, the 10th largest city in England and the 13th largest in the UK. The urban area extends... [full article at wikipedia](#)

[Write new description for Freebase.com](#)

[Add a Type](#) Contents: [Location](#) [Administrative Division](#) [City/Town](#) [Dated location](#) [Statistical region](#) [English unitary authority](#)

| Location                       | edit | Geolocation | latitude | longitude |
|--------------------------------|------|-------------|----------|-----------|
| <a href="#">more options</a>   |      |             | 52.6342  | -1.1385   |
| <a href="#">4 empty fields</a> |      |             |          |           |

| Contains                                 | edit | Contains |  |
|--|------|----------|--|
| <a href="#">more options</a>             |      |          |  |
| <a href="#">University of Leicester</a>  |      |          |  |
| <a href="#">Filbert Street</a>           |      |          |  |
| <a href="#">Walkers Stadium</a>          |      |          |  |
| <a href="#">Leicester Medical School</a> |      |          |  |
| <a href="#">Leicester College</a>        |      |          |  |

| Contained by                   | edit | Contained by |                             |
|--------------------------------|------|--------------|-----------------------------|
| <a href="#">more options</a>   |      |              |                             |
| <a href="#">United Kingdom</a> |      |              | <a href="#">detail view</a> |
| <a href="#">England</a>        |      |              |                             |
| <a href="#">East Midlands</a>  |      |              |                             |
| <a href="#">Leicestershire</a> |      |              | <a href="#">detail view</a> |

| Area                         | edit | Area |                             |
|------------------------------|------|------|-----------------------------|
| <a href="#">more options</a> |      |      |                             |
| <a href="#">73.32km²</a>     |      |      | <a href="#">detail view</a> |

| GEONet feature ID            | edit | GEONet feature ID |  |
|------------------------------|------|-------------------|--|
| <a href="#">more options</a> |      |                   |  |
| <a href="#">6077607</a>      |      |                   |  |

| Time zone(s)                        | edit | Time zone(s) |                             |
|-------------------------------------|------|--------------|-----------------------------|
| <a href="#">more options</a>        |      |              |                             |
| <a href="#">Greenwich Mean Time</a> |      |              | <a href="#">detail view</a> |

| Country                        | edit | Country |  |
|--------------------------------|------|---------|--|
| <a href="#">more options</a>   |      |         |  |
| <a href="#">United Kingdom</a> |      |         |  |

| ISO 3166-2 Code              | edit | ISO 3166-2 Code |  |
|------------------------------|------|-----------------|--|
| <a href="#">more options</a> |      |                 |  |
| <a href="#">GB-LCE</a>       |      |                 |  |

| FIPS 10-4 Region Code        | edit | FIPS 10-4 Region Code |  |
|------------------------------|------|-----------------------|--|
| <a href="#">more options</a> |      |                       |  |
| <a href="#">UKH4</a>         |      |                       |  |

| City/Town                      | edit | City/Town |  |
|--------------------------------|------|-----------|--|
| <a href="#">more options</a>   |      |           |  |
| <a href="#">0 empty fields</a> |      |           |  |

| Dated location                 | edit | Dated location |  |
|--------------------------------|------|----------------|--|
| <a href="#">more options</a>   |      |                |  |
| <a href="#">2 empty fields</a> |      |                |  |

| Statistical region             | edit | Statistical region |  |
|--------------------------------|------|--------------------|--|
| <a href="#">more options</a>   |      |                    |  |
| <a href="#">0 empty fields</a> |      |                    |  |

**Map** [edit](#)

[Wikipedia](#)

Social database (freebase.com)

# "Cutting Room Floor"

- RDF, RDFS data model/details
- Linked data Web tools
  - <http://linkeddata.org/tools>
- RDFa, GRDDL - embedding RDF in (X)HTML
- Yahoo! Search Monkey
  - <http://developer.yahoo.com/searchmonkey/>
- Advantages of RDF/Linked Data over RDBs for Ruby on Rails development
  - O'reilly: Programming the Semantic Web (coming)
    - <http://oreilly.com/catalog/9780596802066/>
- My own research work (<http://tw.rpi.edu>)
  - Scaling RDFS inference, policy/accountability



Rensselaer



# Bottom line

- The "low end Semantic Web, powered by technologies such as RDFS, SPARQL, and a little bit of OWL is showing tremendous promise
  - Can embed the power of the Semantic Web in traditional Web apps
    - Closer to Web 2.0 in look and feel
    - Similar implementation approach
- Significant and growing industrial interest
  - Web 3.0: the big one is still out there!!!!

**Lighter, Faster, Easier!**



Rensselaer

