

# The Semantic Web: Lighter, Faster, Easier

Jim Hendler

Tetherless World Prof. of Information  
Technology and Web Science

RPI

@jahendler

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



# The Semantic Web (ca. 2001)



## Scientific American Article notes

### [Joint starting place:]

- I. Semantic Web Vision (TBL)
  - II. What are the enablers? (in sequence)
    - Screen Scraping (Ora and TBL)
    - Data on Web (Ora and TBL)
    - Zip code link between Data Bases (TBL)
    - Ontology Independence (JAH)
- Effect of Scale (TBL)

## “Then, a miracle occurs”

- III. What can you do with it? (not necessarily in sequence)
  - Self-describing documents (JAH)
  - Logic to encode... (TBL)
  - Services and Advertising (Ora)
  - Devices (Ora)
  - Digital Signatures, Authentication, and Trust (TBL)



Rensselaer

(Berners-Lee, Hendler, Lassila; 2001)



# Semantic Web ca. 2009

- Semantic Web finding success even in tough market
  - Lots of small companies in the market: Altova... Zepheira (eg. C&P, Franz, Intellidimension, Intellisophic, Ontology Works, Siderean, SandPiper, SiberLogic, TopQuadrant ...)
  - Web 3.0 new buzzword: Garlik, Twine, Freebase, Bintro, Siri, Talis, ...
  - Semantic Search taking off - Powerset bought by Microsoft for over \$100,000,000, hakia, bing, ...
- Bigger players buying in
  - 2009 announcements at SemTech (June): Google, New York Times, Oracle, IBM, Yahoo, MS Live Labs, Siri, ...
  - 2008: Gartner identifies Corporate Semantic Web as one of three "High impact" Web technologies
  - Tool market forming: AllegroGraph, TopBraid, Pellet2, ...
- Government projects in and across agencies
  - Recent open data announcements by UK and US
  - Projects/demos in EU, Japan, Korea, China, India...
  - SKOS update in govt (and private) libraries
- Several "verticals" heavily using Semantic Web technologies
  - Health Care and Life Sciences
    - Interest Group at W3C
  - Financial services
  - Human Resources
  - Sciences other than Life Science
    - Virtual observatory, Geo ontology, ...
- Many open source tools available
  - Kowari, RDFLib, Jena, Sesame, Protégé, SWOOP, Pellet, ...



SEMANTIC WEB & LINKED DATA BUSINESS STRATEGIES

October 16-17, 2008 • Santa Clara, CA

JupitermediaEvents

## The Next Generation Web



**Web 3.0**

Web 3.0 Conference & Expo is focused on bringing together the key proponents and components delivering the promise of next generation *web applications, technologies and business utilization*.

Web 3.0 showcases in case-study format, the explosive, game-changing promise and disruptive opportunities as they develop in order to help the entire community realize the promise of Web 3.0.

**Semantic Web**

Semantic Web takes that hyper-data and enriches it with meaning through semantic web standards, technologies and strategies, and ultimately leads to the next generation Web.

**Linked Data**

Linked Data brings traditional hyperlinks and today's "web of documents" into the era of an interconnected, standards-based "web of linked data".

### Applications, Technologies and Business Utilization

Different terms being used in different ways



Rensselaer





# Then build your app on top



Dbpedia mobile



HealthFinland



Semaplorer



Rensselaer



# Builds over RDF DBs

- The Profile Manager enables you to store information about users and services. It is a Resource Description Framework (RDF) data store and is general nature, so you can store any information that is required by your system. ... There are two main benefits offered by a profile store that has been created by using RDF. The first is that RDF enables you to store data in a flexible schema so you can store additional types of information that you might have been unaware of when you originally designed the schema. The second is that **it helps you to create Web-like relationships between data, which is not easily done in a typical relational database.**

\_\_\_ <http://msdn2.microsoft.com/en-us/library/aa303446.aspx> - 12/06

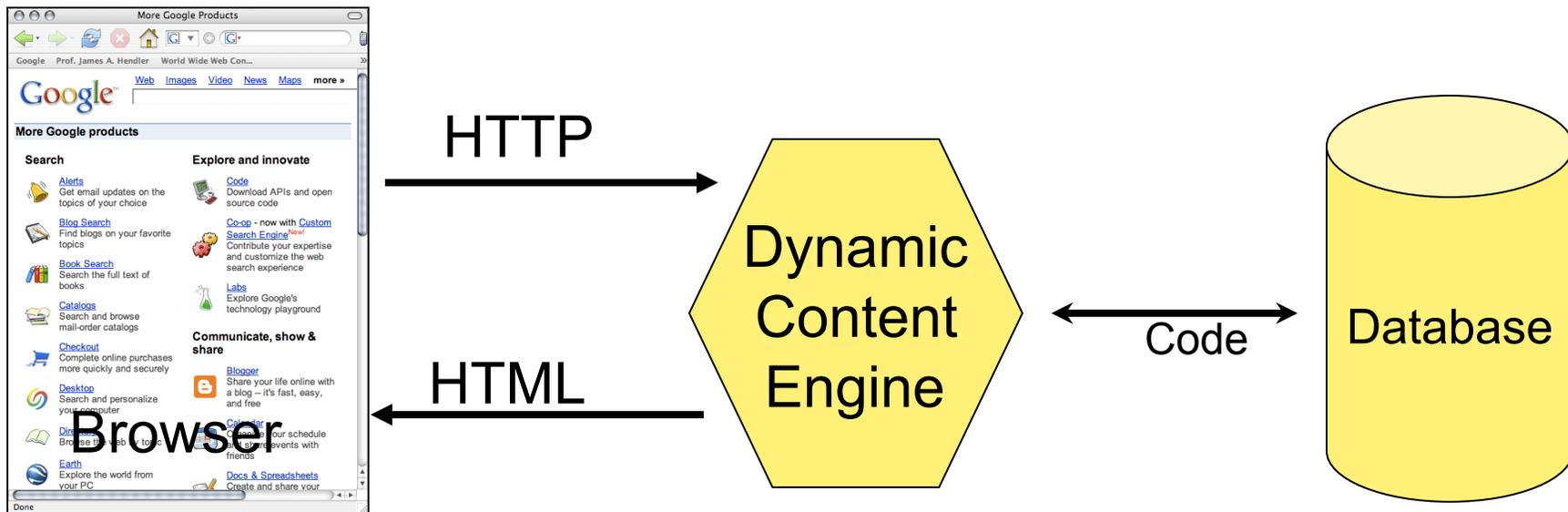


Rensselaer

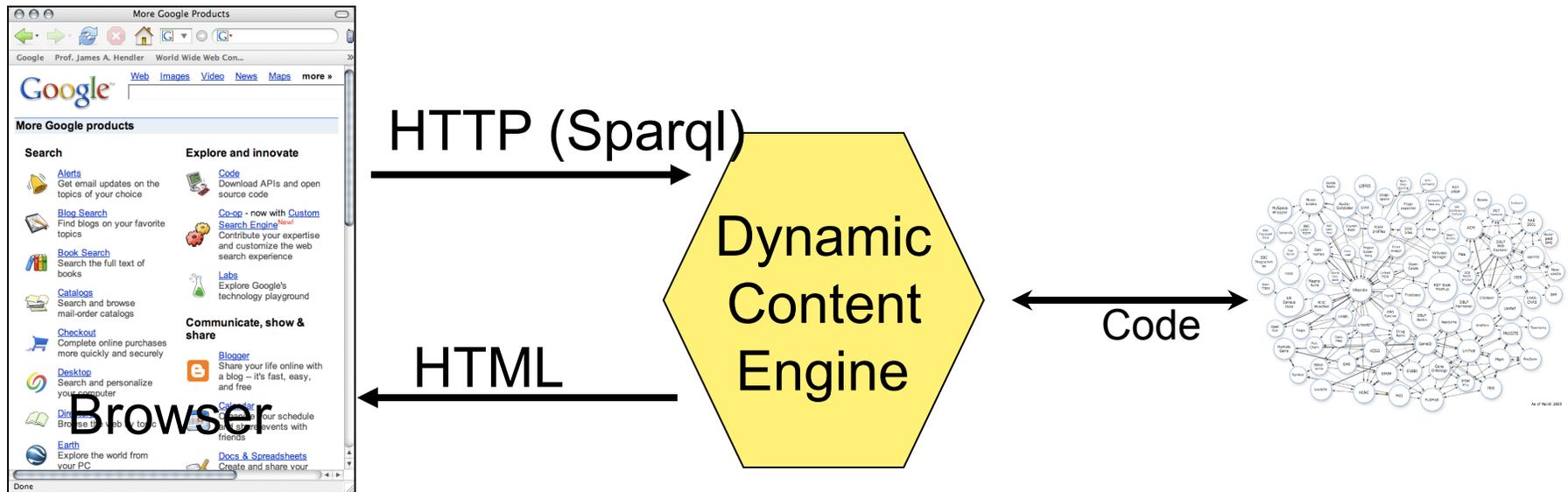




# Traditional Web applications



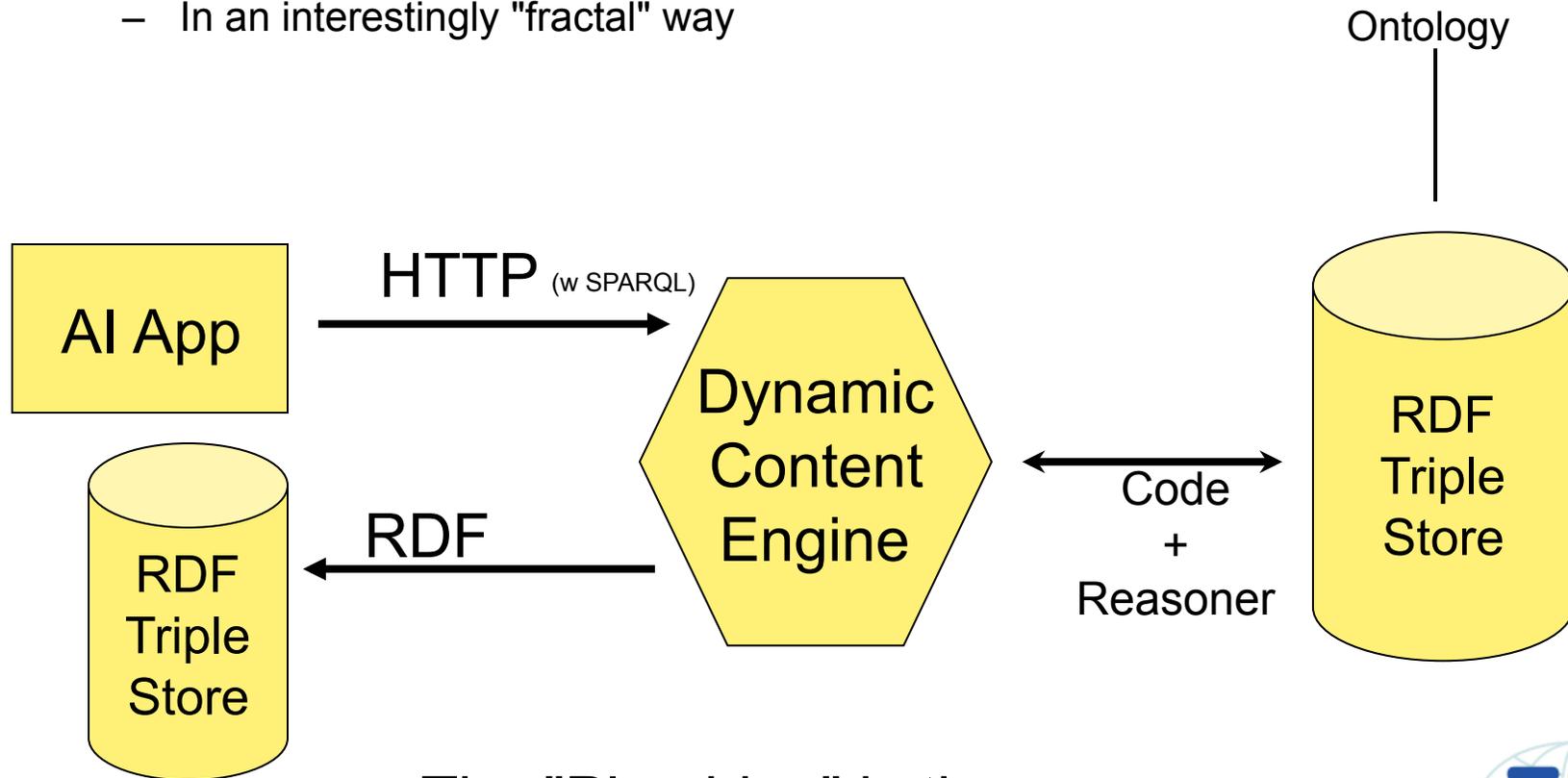
# Semantic Web applications



Do your mashup on the underlying data instead of presentations thereof

# Semantic Web applications

- And a similar model can power the "high end" Semantic Web applications
  - In an interestingly "fractal" way



The "Plumbing" is the same



Rensselaer



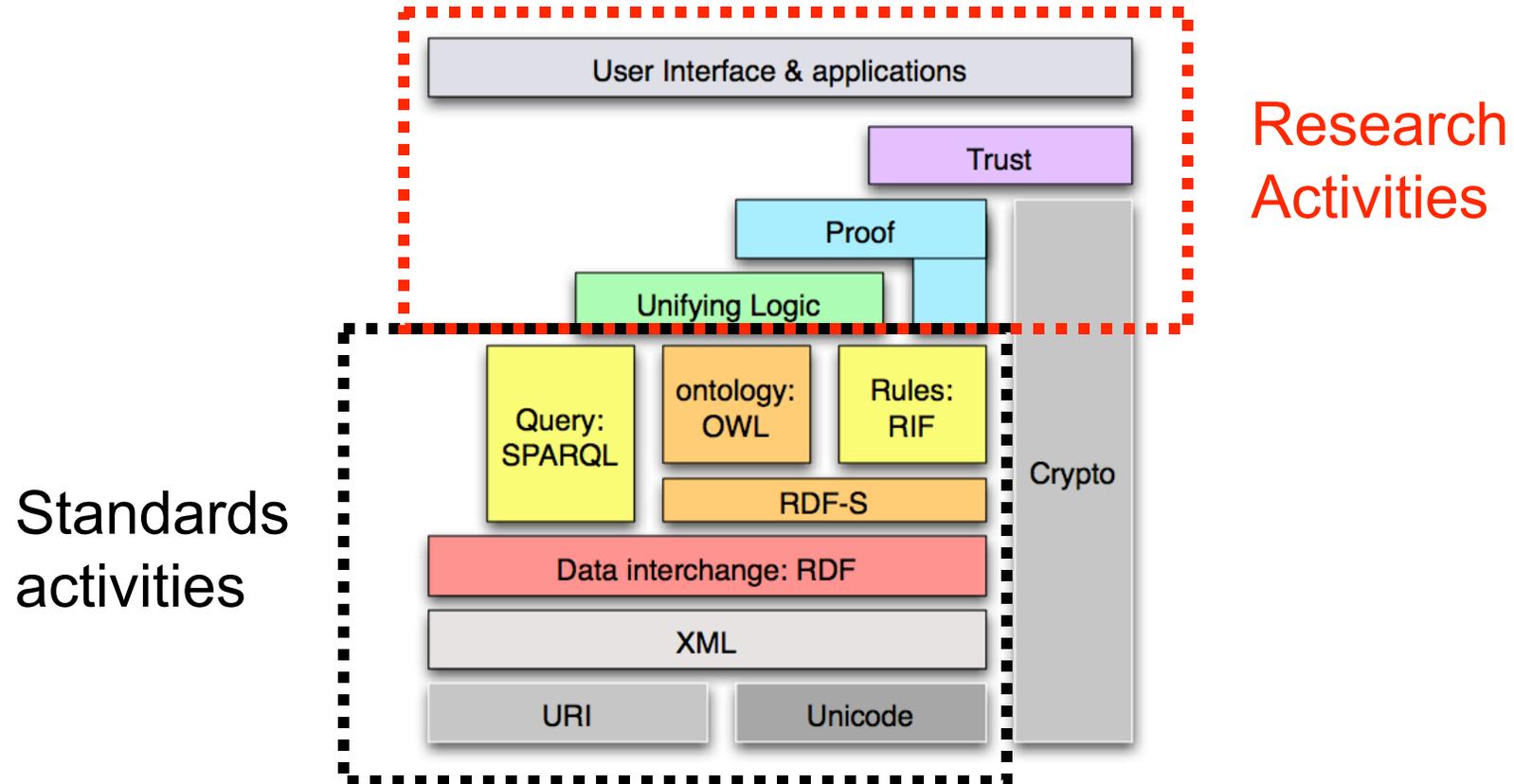
# Linked Data + Semantics

- "Linked Data" approach finds its use cases in Web Applications (at Web scales)
  - **A lot of data, a little semantics**
  - **Finding anything in the mess can be a win!**



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

# Semantic Web "Stack"



ca. 2006

(Tim Berners-Lee)



Rensselaer



# 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

# 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!



# 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

# Ontology: the traditional view



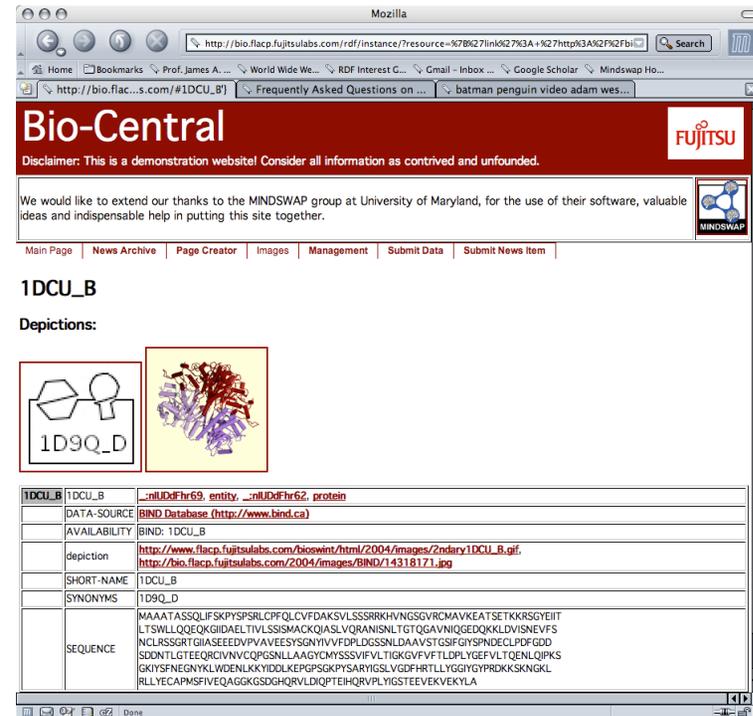
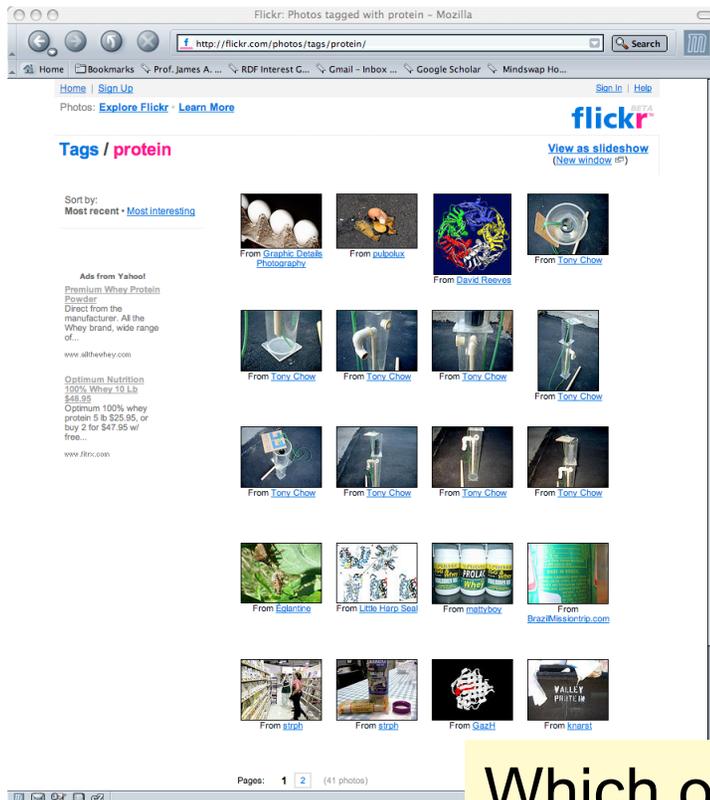
- **Ontology as Barad-Dur (Sauron's tower):**

- Extremely powerful!

- ~~Patrolled by Orcs~~ <sup>Decidable Logic basis</sup> <sub>inconsistency</sub>

- Let one little ~~hobbit~~ in, and the whole thing could come crashing down

# The argument for this seems compelling



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

# 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

# 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"

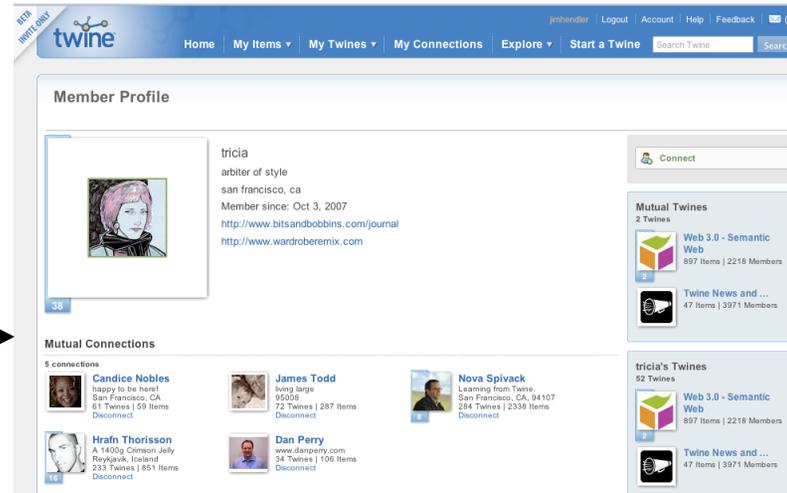
Analogy: HTML

# Very simple "reasoning"

**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 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"



The screenshot shows the Twine website interface. At the top, there's a navigation bar with 'twine' logo and links for Home, My Items, My Twines, My Connections, Explore, Start a Twine, and a search bar. Below the navigation bar is the 'Member Profile' section for 'tricia'. It includes a profile picture, a 'Connect' button, and a list of 'Mutual Twines' such as 'Web 3.0 - Semantic Web' and 'Twine News and ...'. Below the profile, there's a 'Mutual Connections' section showing five other members: Candice Nobles, James Todd, Nova Spivack, Hrafn Thorisson, and Dan Perry, each with a small profile picture and basic stats.



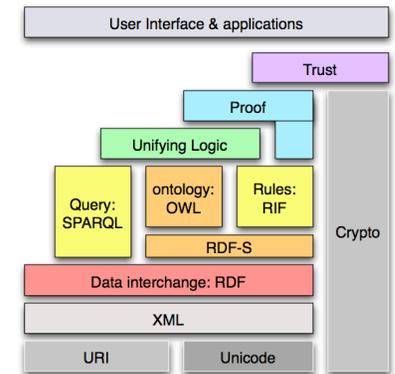
# Ontologies?

- 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)

# Evolving standards



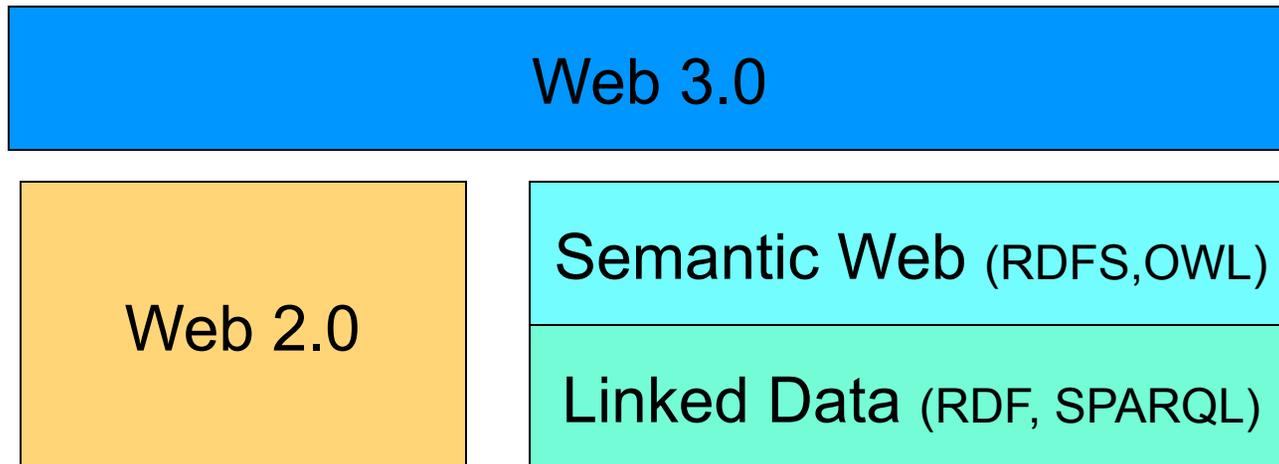
- **SPARQL: Query language for (distributed) triple stores**
  - the “SQL of the Semantic Web”
- **GRDDL/RDFa: Integration of HTML and Semantic Web**
  - “embedding” RDF-based annotation on traditional Web pages
  - Both Yahoo! and Google now supporting RDFa
- **OWL 2.0: New features, specialized subsets**
  - **OWL RL – simplification, identity, scaling to large datasets**
- **RIF: Rules Interchange Format**
  - representing rules on the Web
  - linking rule-based systems together
- **And more...**
  - **SKOS** thesaurus standard
  - Multimedia annotation, Web-page metadata annotation, Health Care and Life Sciences (LSID), privacy, Sem Web Service, etc.



Rensselaer



# The new meme: Web 3.0



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

# The making of a 3.0 app...

- The Wine Agent ca '85 : Reasoner with knowledge of wine and food pairings

Swordfish subclass BlandFish subclass Fish  
⇒ Dry + medium-body + White  
⇒ Drink EdenValleyChardonnay

- Used the wines in a particular wine cellar
  - Hard wired in
- Eventually completed with "correct" wine recommendations
  - You disagree, tough! You're wrong.

# Wine Agent 2007

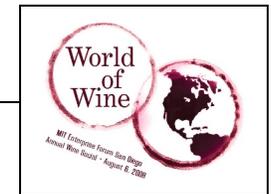


## TW Wine Agent

[Overview](#)  
[Acknowledgements](#)

To view recommendations for a given type of food, click the desired food in the menu below.

- + \* [Meat \(1 / 13 below\)](#)
- + \* [Fowl \(0 / 6 below\)](#)
- + \* [OtherTomatoBasedFood \(1 / 1 below\)](#)
- \* [Seafood \(2 / 24 below\)](#)
  - \* [Shrimp \(1 / 0 below\)](#)
  - \* [Shellfish \(1 / 8 below\)](#)
    - \* [Non Oyster Shellfish \(1 / 5 below\)](#)
      - \* [Crab \(2 / 0 below\)](#)
      - \* [Mussels \(1 / 0 below\)](#)
      - \* [Lobster \(1 / 0 below\)](#)
      - \* [Clams \(1 / 0 below\)](#)
    - \* [Oyster Shellfish \(1 / 1 below\)](#)
      - \* [Oysters \(1 / 0 below\)](#)
  - \* [Fish \(3 / 11 below\)](#)
    - \* [Non Bland Fish \(2 / 4 below\)](#)
      - \* [Tuna \(1 / 0 below\)](#)
      - \* [Swordfish \(3 / 0 below\)](#)
    - \* [Bland Fish \(2 / 3 below\)](#)
      - \* [Flounder \(1 / 0 below\)](#)
      - \* [Scrod \(1 / 0 below\)](#)
      - \* [Halibut \(1 / 0 below\)](#)
- + \* [Fruit \(1 / 5 below\)](#)
- + \* [Dessert \(1 / 8 below\)](#)
- + \* [Pasta \(3 / 9 below\)](#)



Rensselaer

Recommendation Wiki

(Michaelis, 07)





# TW Wine Agent

[Overview](#)  
[Acknowledgements](#)

Why MountEdenVineyardEdnaValleyChardonnay was selected for Fish

### Wine Properties

NAME: MountEdenVineyardEdnaValleyChardonnay  
COLOR: White  
BODY: Medium  
FLAVOR: Moderate  
SUGAR: Dry

### List of recs being considered

### Supporting Recs

TOTAL IN SUPPORT: 9

ID	COLOR	BODY	FLAVOR	SUGAR
<b>MountEdenVineyardEdnaValleyChardonnay</b>	White	Medium	Moderate	Dry
Bland-2Dfish	White	Medium U Full	Moderate U Strong	
RecDLM Swordfish	White	Medium	Moderate	Dry
RecDLM Tuna	White	Medium	Moderate	Dry
RecSwordfish	White	Medium		
RecNonBlandFish	White		Moderate	
RecDLM NonBlandFish	White	Medium	Moderate	Dry
RecFish	White		Moderate	Dry
RecDLM Fish		Medium		Dry
RecSeafood	White			

### Opposing Recs

TOTAL IN CONFLICT: 6

ID	COLOR	BODY	FLAVOR	SUGAR
<b>MountEdenVineyardEdnaValleyChardonnay</b>	White	Medium	Moderate	Dry
RecDLM_Scrod	White	Medium	Delicate ☒	Dry
Melville_Estate_Chardonnay_2006	White	Light ☒	Strong ☒	Sweet ☒
RecDLM_Halibut	White	Medium	Delicate ☒	Dry
Rec-2Dhendler	Red ☒	Light ☒		Dry
RecDLM_Flounder	White	Medium	Delicate ☒	Dry
RecDLM_BlandFish	White	Medium	Delicate ☒	Dry



# Rensselaer





# TW Wine Agent

[Overview](#)  
[Acknowledgements](#)

Why LongridgeMerlot was selected for Swordfish

## Wine Properties

NAME: LongridgeMerlot  
COLOR: Red  
BODY: Light  
FLAVOR: Moderate  
SUGAR: Dry

## List of recs being considered

### Supporting Recs

TOTAL IN SUPPORT: 1

ID	COLOR	BODY	FLAVOR	SUGAR
<b>LongridgeMerlot</b>	Red	Light	Moderate	Dry
Rec-2Dhendler	Red	Light		Dry

### Opposing Recs

TOTAL IN CONFLICT: 6

ID	COLOR	BODY	FLAVOR	SUGAR
<b>LongridgeMerlot</b>	Red	Light	Moderate	Dry
RecSwordfish	White ☒	Medium ☒		
RecNonBlandFish	White ☒		Moderate	
Melville_Estate_Chardonnay_2006	White ☒	Light	Strong ☒	Sweet ☒
RecFish	White ☒		Moderate	Dry
RecDLM_Swordfish	White ☒	Medium ☒	Moderate	Dry
RecSeafood	White ☒			

[back](#)



# Rensselaer

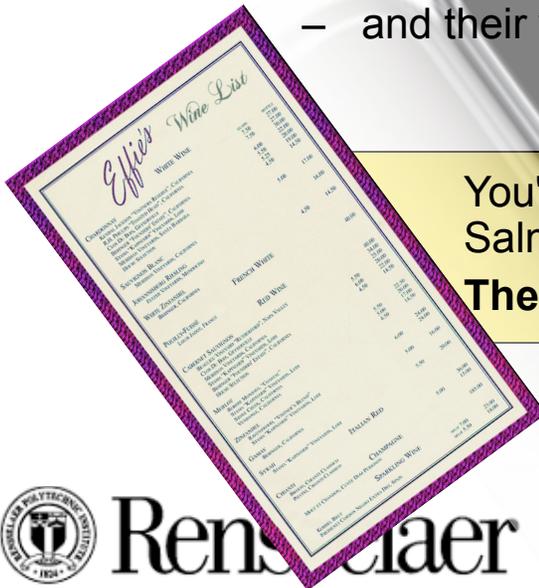


# Wine Agent 3.0

- Coming soon: Add the Data!
  - Phone knows your location (and thus what restaurant you are in)
    - And the menu
    - And the wine list
  - Phone knows who else is there
    - Your Facebook network is there with you
    - So are other people with the application
  - and their wine preferences

You're having sole, Jane beef, and Fred the Salmon Special

**The Flowers Pinot would be a great choice!**



Patton 08



Rensselaer



# Web 3.0 examples

The screenshot shows a web browser window with the URL <http://www.powerset.com/explore/go/james-hendler>. The page is titled "James Hendler - Powerset" and features a search bar with the text "James Hendler". Below the search bar, there is a profile for James Hendler, including a photo and a brief biography. The profile also lists his date of birth (1957), place of birth (Queens), profession (Computer scientist, Computer Science), and religion (Jewish). Below the profile, there is a section titled "Facts from Wikipedia: we found the following about James Hendler" which lists several facts, such as "James Hendler developed SHOE." and "James Hendler saw parallell.". At the bottom of the page, there is a list of Wikipedia articles related to James Hendler, including "James Hendler", "AI winter", "Hendler James Hendler", "Simple HTML Ontology Extensions", "List of Rensselaer Polytechnic Institute people", "Ontology (information science)", "Rensselaer Polytechnic Institute", "Semantic Web Rule Language", "HTTP body data", and "Variant object".

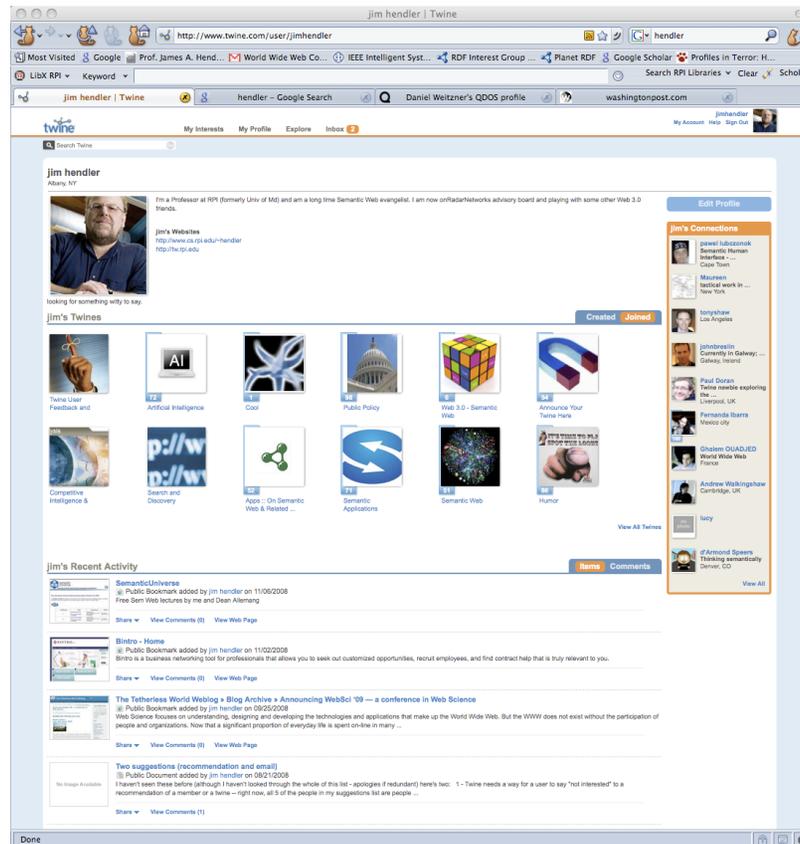
## Semantic Search (Powerset.com)



Rensselaer



# Web 3.0 examples



Enhanced Social Networks (twine.com)

# Web 3.0 examples

The screenshot shows the BINTRO BETA website interface. The user is logged in as 'James'. The main navigation includes HOME, MESSAGES, MY BROADCASTS, COMPOSE BROADCAST, CONTACTS, and PROFILE. A banner at the top encourages users to broadcast their needs or opportunities to business professionals. The main content area is titled 'Activity For Web 3.0 Expertise' and shows a list of broadcast results. The first result is a match for 'Adam Glick', who is located in New York, United States, and is a Vice President in the Commercial Mortgage Loans industry. The match score is 75%. The interface also includes a 'Send a Message To Adam' button and a footer with copyright information for Paradigm5 Inc. DBA Bintro © 2008.

Semantic Match (bintro.com)

# Web 3.0 examples

The screenshot shows a web browser window displaying the BINTRO BETA website. The page features a navigation menu with options like HOME, MESSAGES, MY BROADCASTS, COMPOSE BROADCAST, CONTACTS, and PROFILE. A main heading reads "BROADCAST YOUR NEED OR OPPORTUNITY" followed by a sub-heading "... to business professionals like yourself or manage and respond to those needs and/or opportunities you've previously Broadcast." Below this, there is a section titled "2 Active Broadcasts" with a sub-heading "Activity For: Web 3.0 Expertise". The main content area displays a profile for "ADAM" with the following details:

- ABOUT ADAM:** Currently Vice President of Mortgage Originations at Paradigm Capital Funding, with over 10 years of experience in commercial bridge lending, specializing in structuring non-conforming real estate transactions.
- BUSINESS EXPERTISE:**
  - Commercial Bridge Loans: LEVEL OF EXPERTISE 10
  - Real Estate: LEVEL OF EXPERTISE 10
  - Commercial Mortgage Loans: LEVEL OF EXPERTISE 10
- SKILLS:**
  - Mortgage Originations: SKILL LEVEL 8
  - Commercial Hard Money Loans: SKILL LEVEL 10
  - Underwriting: SKILL LEVEL 8
- EMPLOYMENT HISTORY:**
  - Employer: Paradigm Capital Funding
  - Title: Vice President
  - Period: 08/11/1997 - Present
- EDUCATION HISTORY:**
  - Institution: Syracuse University

At the bottom of the profile, there is a summary of the match: Match Score: 75%, 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 includes "Paradigm5 Inc DBA Bintro © 2008 | Privacy Policy | Terms and Conditions | Security | Contact Us".

## Semantic Match (bintro.com)



# Web 3.0 examples

Leicester facts - Freebase

Leicester topic

Also known as City Of Leicester, Leicester edit

Leicester (pronounced /ˈlɛstə/) 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

Contents: Location Administrative Division City/Town Dated location Statistical region English unitary authority

Location	Geolocation	latitude	longitude
more options -		52.6342	-1.1385
4 empty fields			

Contains	
edit	University of Leicester
	Filbert Street
	Walkers Stadium
	Leicester Medical School
	Leicester College
	detail view >

Contained by	
edit	United Kingdom
	England
	East Midlands
	Leicestershire
	detail view >

Area	
edit	73.32km²
	detail view >

GEOnet feature ID	
edit	6077607

Time zone(s)	
edit	Greenwich Mean Time
	detail view >

Country	
edit	United Kingdom

Administrative Division	
edit	ISO 3166-2 Code GB-LCE
more options -	
1 empty field	
edit	FIPS 10-4 Region Code UKH4

City/Town	
more options -	
0 empty fields	

Dated location	
more options -	
2 empty fields	

Statistical region	

Gallery

Map

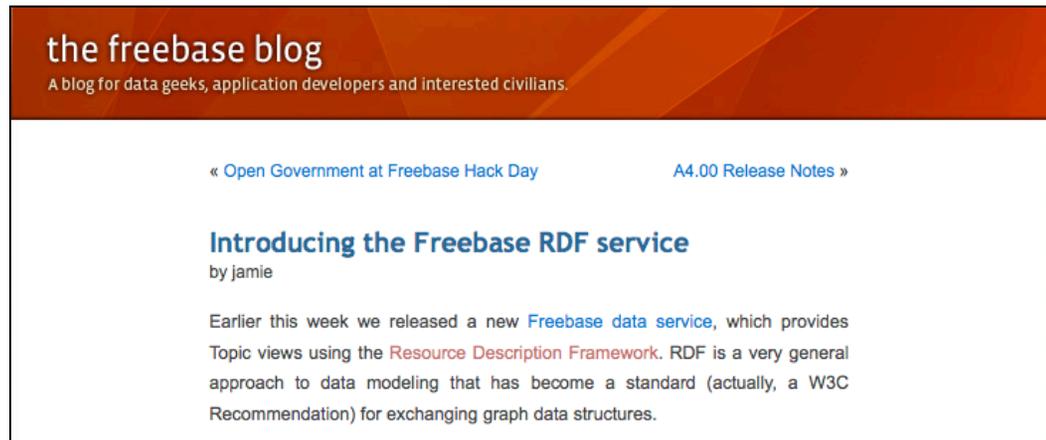
Weblinks

Wikipedia

Social database (freebase.com)

# Web 3.0 - RDF APIs

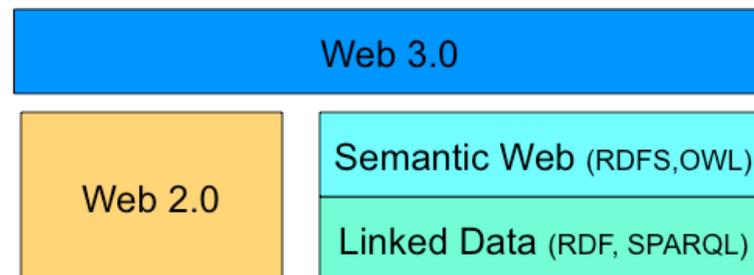
- RDF also starting to provide interoperability between Web applications in Web 2.0 and Web 3.0
  - Many Web 2.0 apps already can dump RDF
    - Flickr, mySpace, facebook, livejournal...
  - Web 3.0 apps are doing so as well



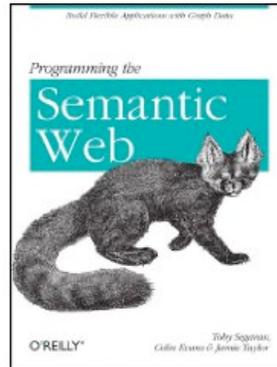
The screenshot shows a blog post from 'the freebase blog'. The header is orange with the text 'the freebase blog' and 'A blog for data geeks, application developers and interested civilians.' Below the header, there are two blue links: « Open Government at Freebase Hack Day » and « A4.00 Release Notes ». The main content of the post is titled 'Introducing the Freebase RDF service' by jamie. The text of the post reads: 'Earlier this week we released a new Freebase data service, which provides Topic views using the Resource Description Framework. RDF is a very general approach to data modeling that has become a standard (actually, a W3C Recommendation) for exchanging graph data structures.'

# Web 3.0 excitement

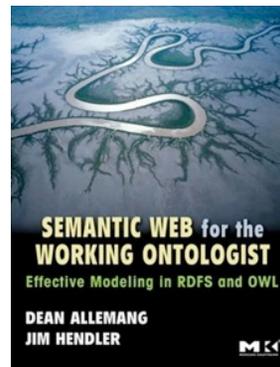
- Significant and growing commercial interest...
  - Web: Google
  - Web 2.0: Facebook, Wikipedia, YouTube, Flickr, ...
  - Web 3.0: the big one is still out there



# How can I learn more?



<http://www.amazon.com/Programming-Semantic-Web-Toby-Segaran/dp/0596153813>



<http://www.amazon.com/Semantic-Web-Working-Ontologist-Effective/dp/0123735564>

# Bottom line

- The Semantic Web, powered by technologies such as RDFS, SPARQL, and a little bit of OWL is showing tremendous promise
  - Linked Data – focus on open world and network effect
    - Mashup the data (Web like) and build you app (Web scale)
    - Traditional Web architectures work just fine
  - Web 3.0 – embed the power of the Semantic Web in large scale Web apps
    - Closer to Web 2.0 in look and feel
    - Similar implementation approach

Lighter, Faster, Easier!

# WebSci10

- <http://websci09.org> was a great conference
  - 350 people
  - #websci09 a top-10 twitter trend
- WebSci10 will be help co-located with WWW2010 in Raleigh, NC, US
  - Subscribe to [wsri-announce@webscience.org](mailto:wsri-announce@webscience.org) for continued information
- And please submit your work!