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!!***

Rensselaer

Tetherless World

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!

Rensselaer

Tetherless World

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

Tetherless World

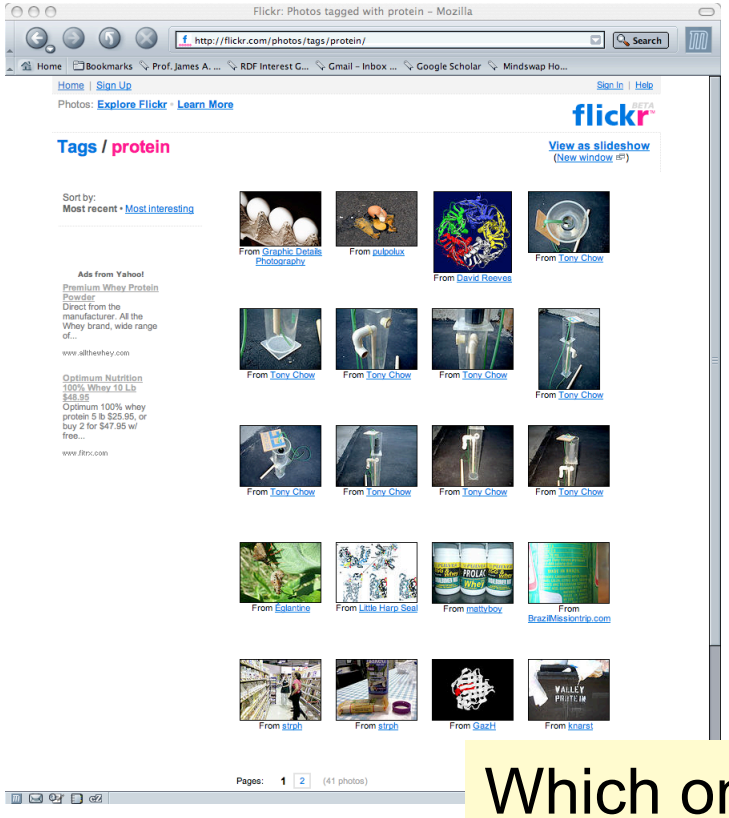
O*ntology*: the traditional view

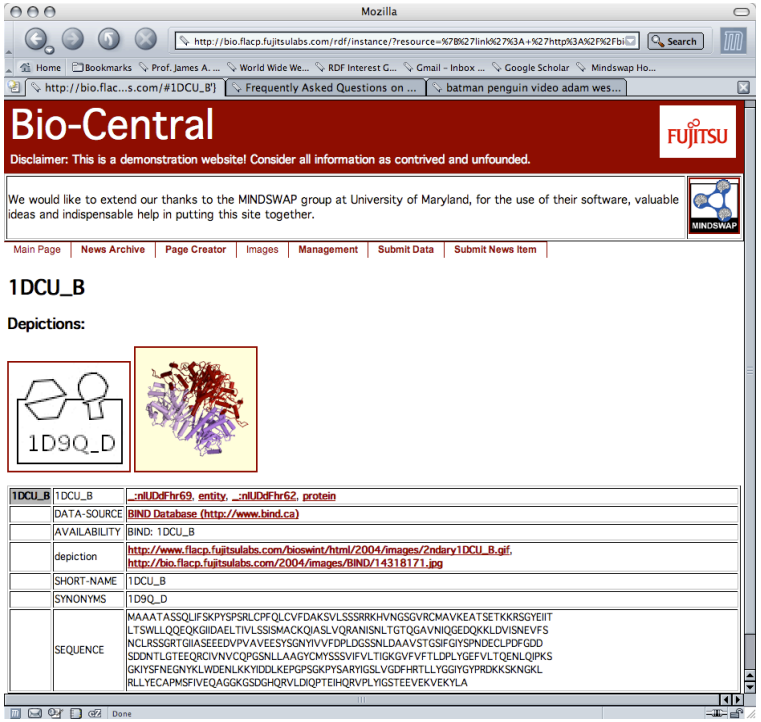
- **O**ntology as Barad-Dur (Sauron's tower):
  - Extremely powerful!
  - ~~Patrolled by Orcs~~ Decidable Logic basis
    - Let one little ~~hobbit~~ inconsistency in, and the whole thing could come crashing down

Rensselaer

Tetherless World

The argument for this seems compelling





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

Rensselaer

Tetherless World

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

Tetherless World

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

Rensselaer

Tetherless World

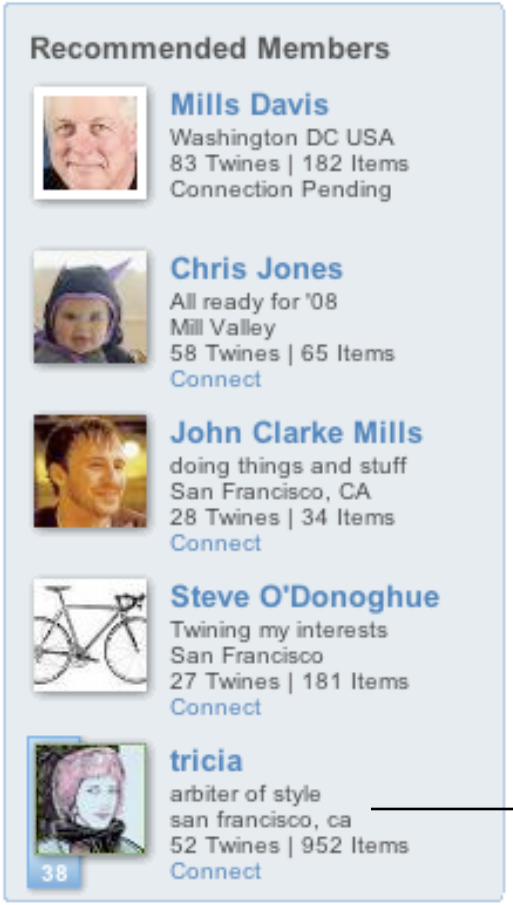
Linked Data Web

- "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*!

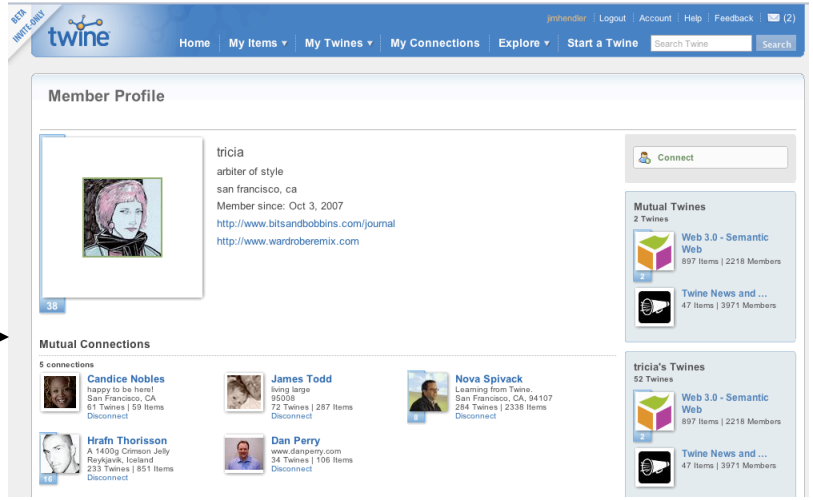
Rensselaer

Tetherless World

Very simple "reasoning"

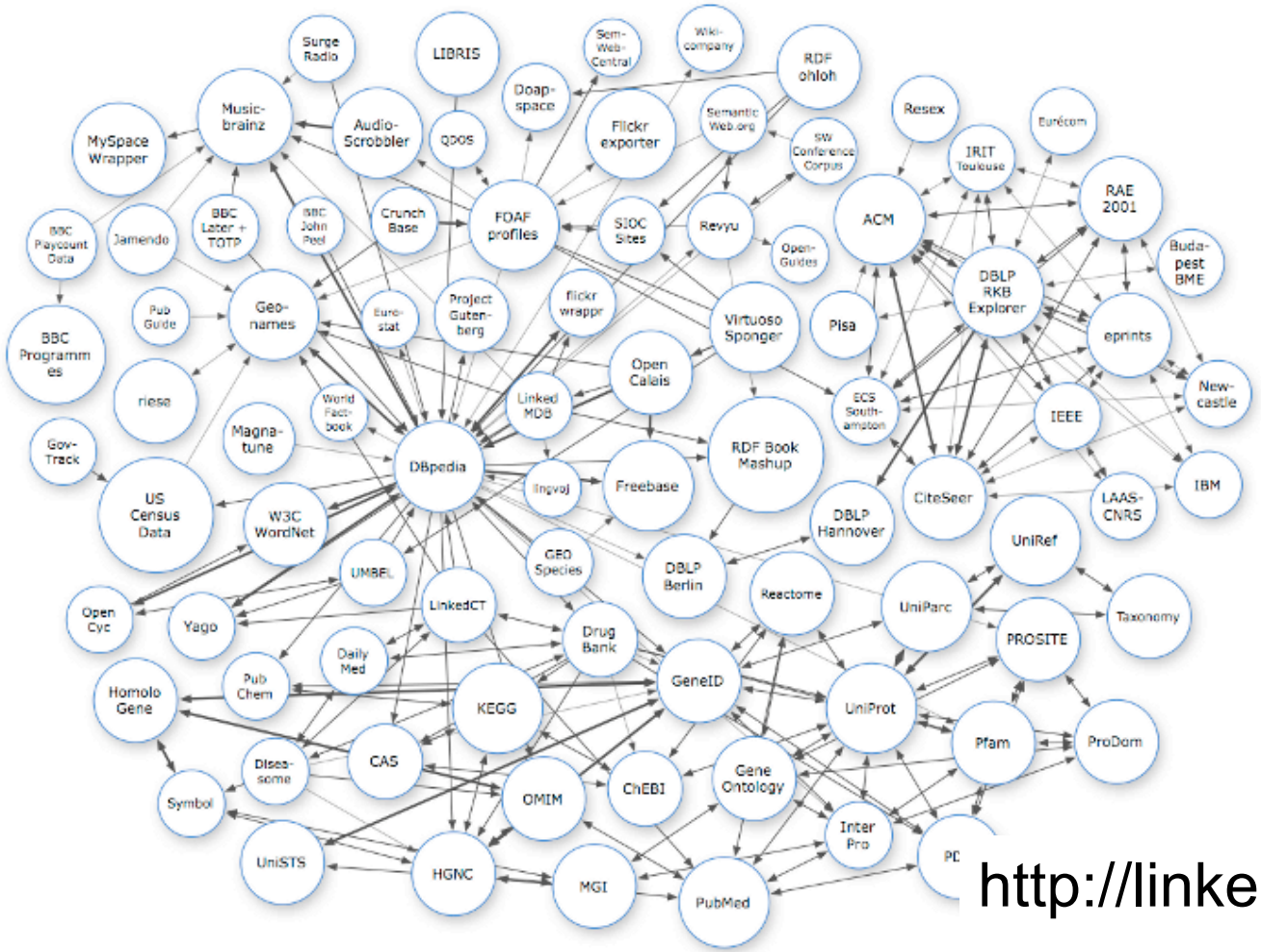


- Twine recommends some people I may want to connect to
  - What is correctness in this case?
    - If I find some folks I like this way, I use twine more. Surprises can be fun.
  - I'm only seeing a few of a very large set so "first" is more important than "there somewhere"



Rensselaer

Tetherless World



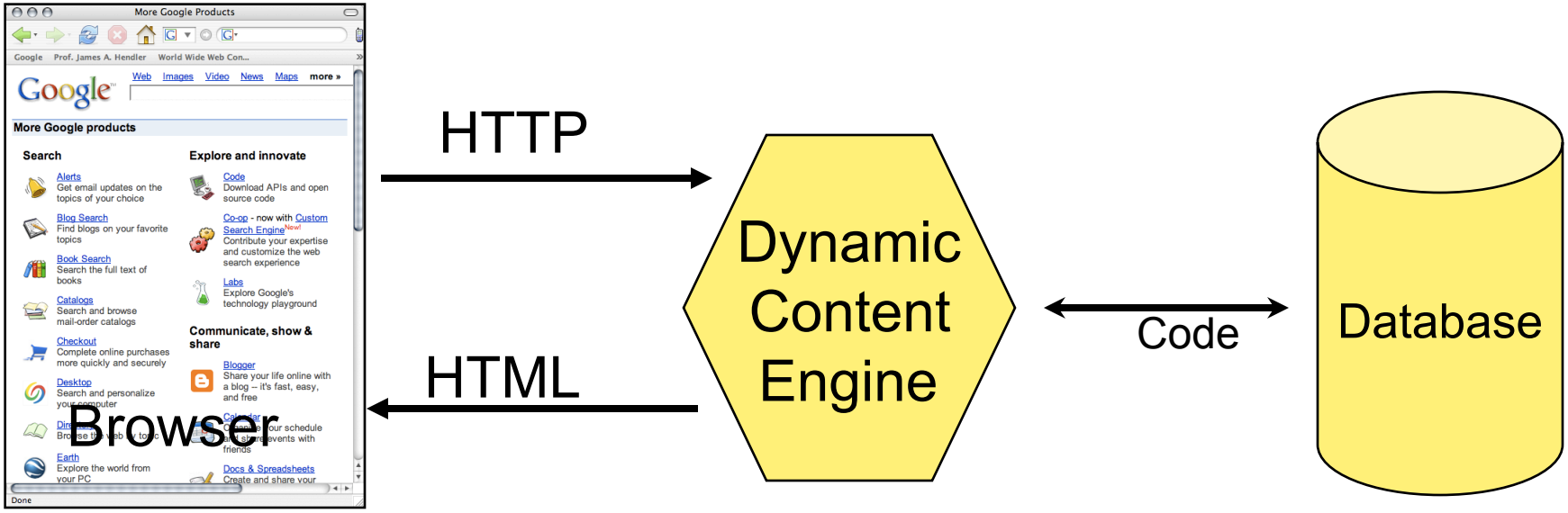
http://linkeddata.org/

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

Rensselaer

Tetherless World

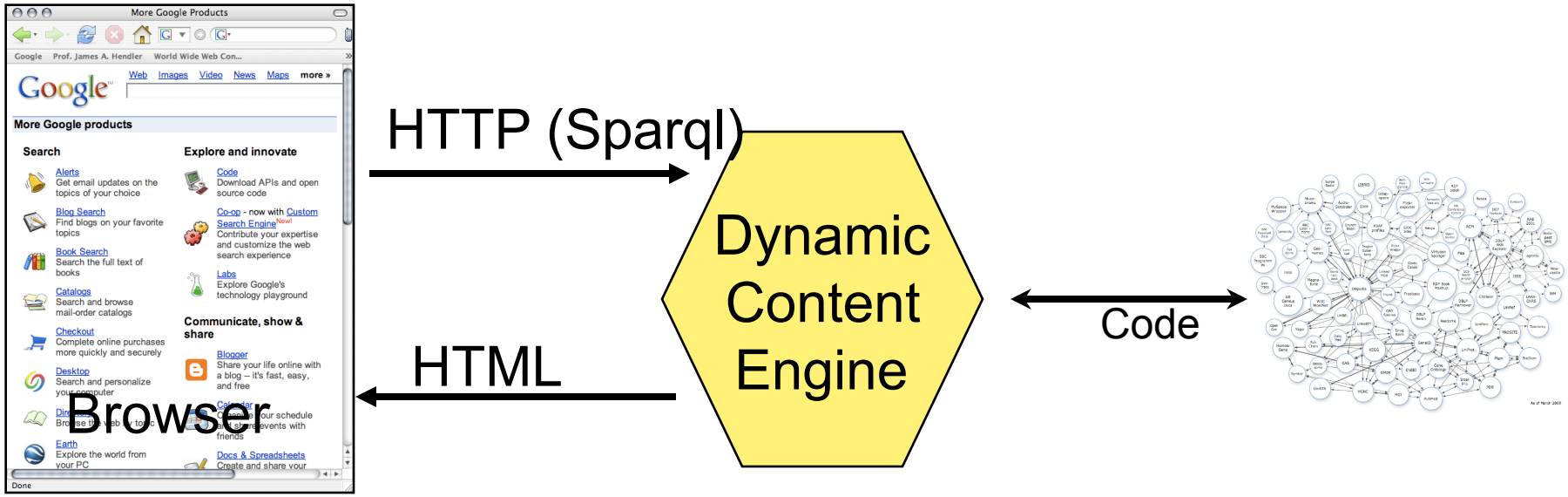
Traditional Web applications



Rensselaer

Tetherless World

Semantic Web applications



Do your mashup on the underlying data instead of presentations thereof

Rensselaer

Tetherless World

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)

Rensselaer

Tetherless World

Example LD applications

Dbpedia mobile

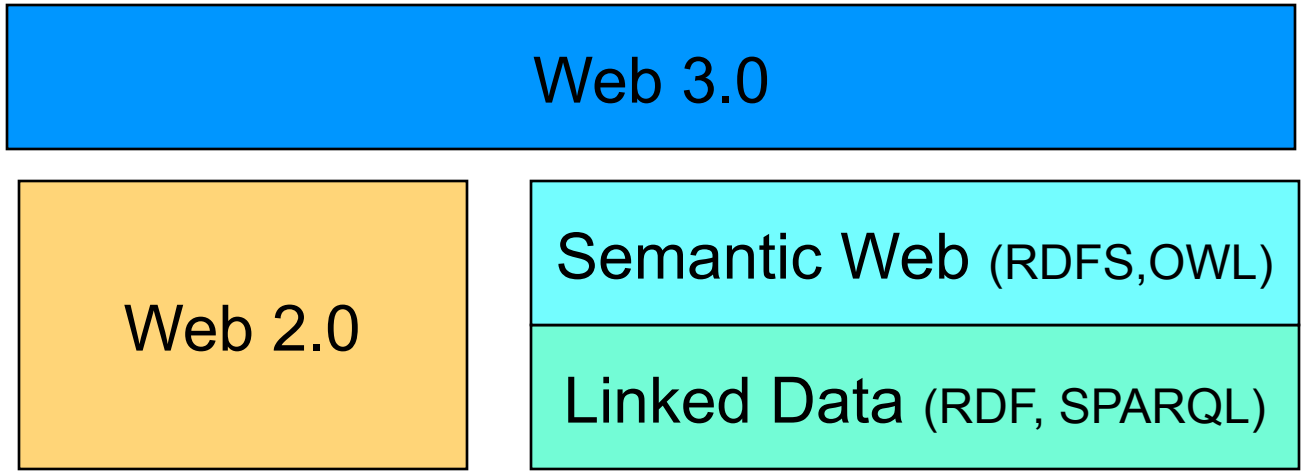
HealthFinland

Semaplorer

Rensselaer

Tetherless World

The industrial "meme"

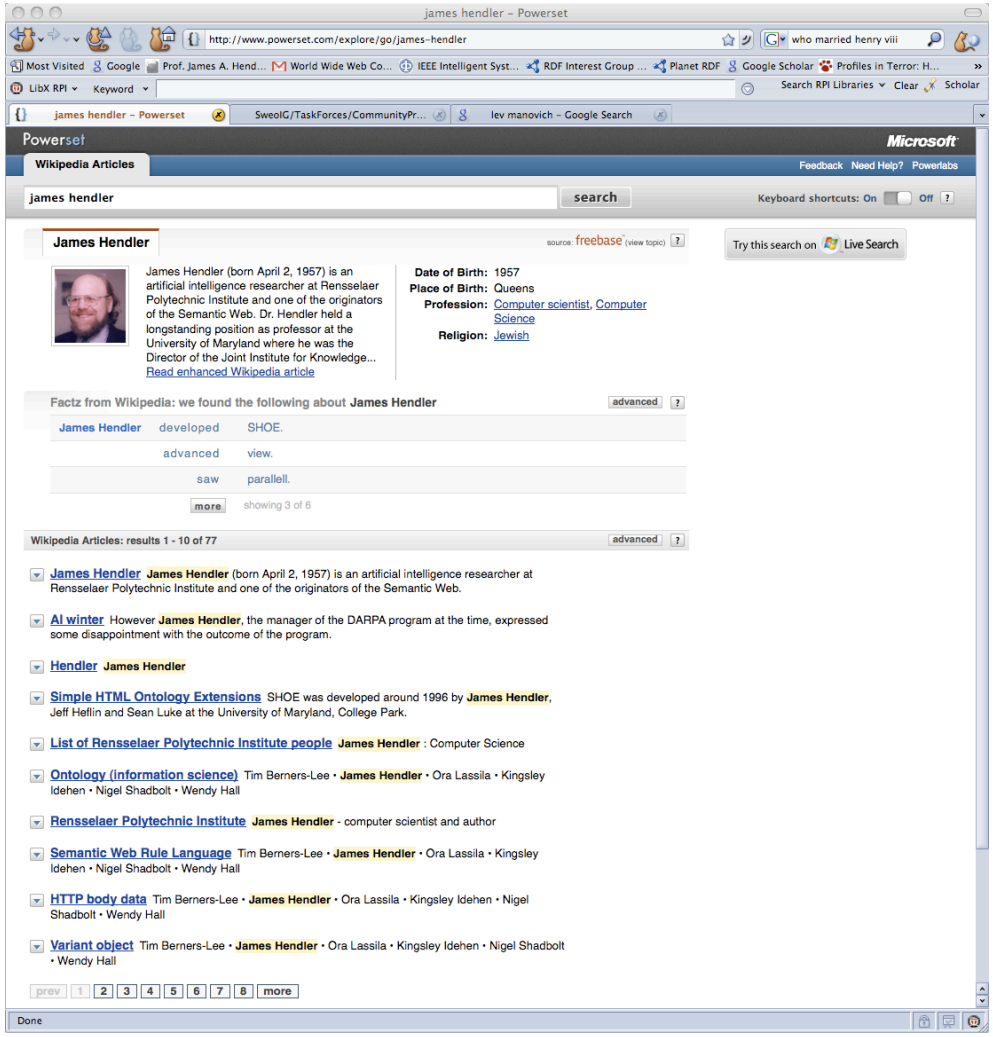


Web 3.0 extends current Web applications using Semantic Web technologies and graph-based, open data.

Rensselaer

Tetherless World

Web 3.0 examples

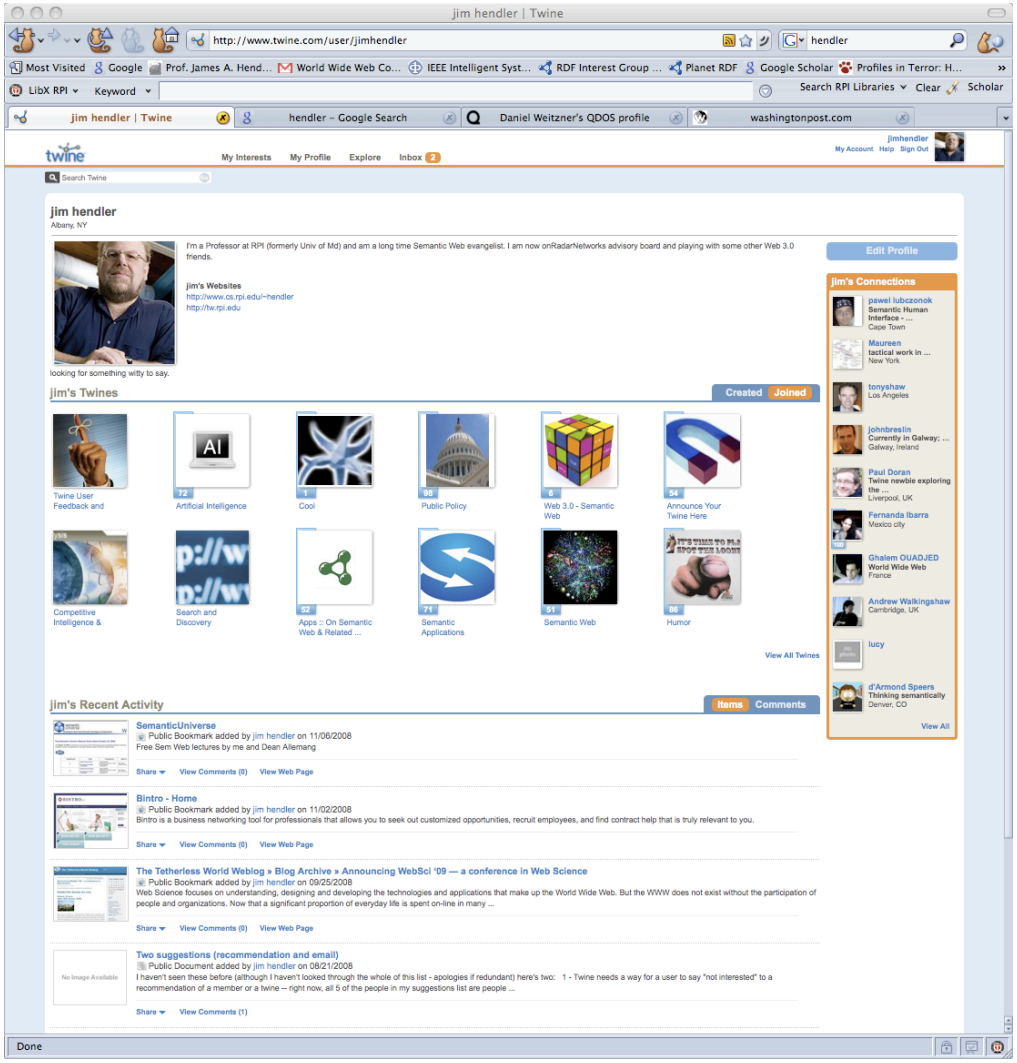


Semantic Search (Powerset.com)

Rensselaer

Tetherless World

Web 3.0 examples

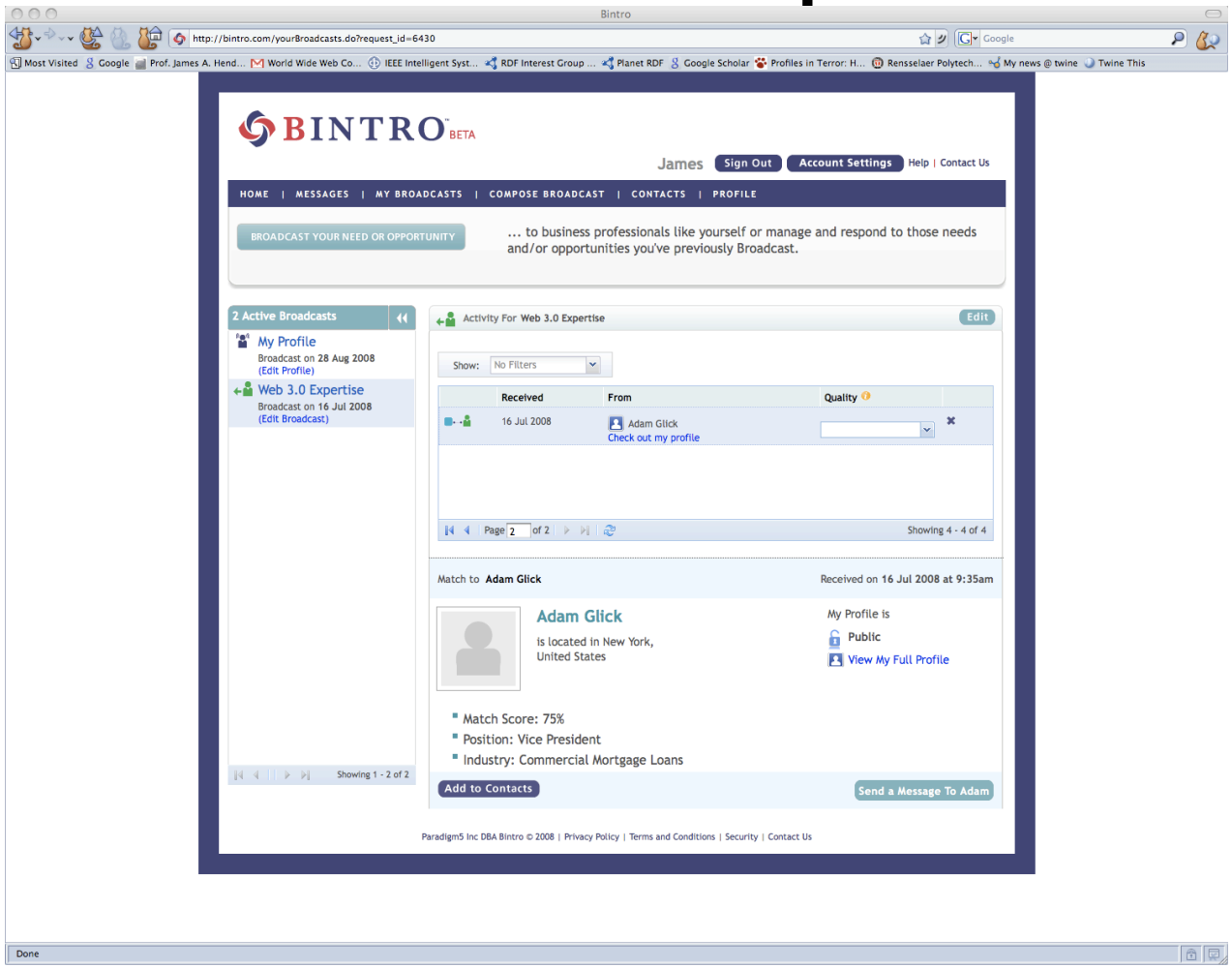


Enhanced Social Networks (twine.com)

Rensselaer

Tetherless World

Web 3.0 examples

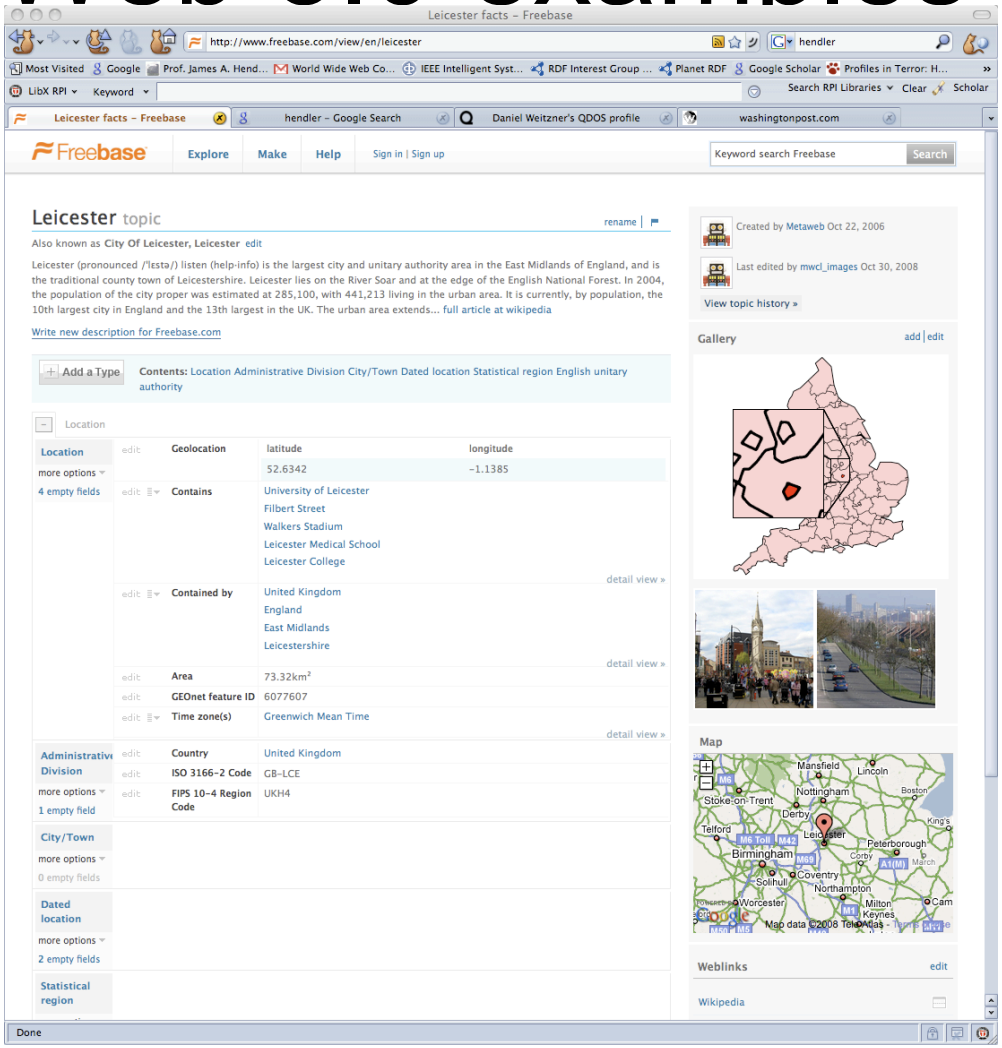


Semantic Match (bintro.com)

Rensselaer

Tetherless World

Web 3.0 examples



Social database (freebase.com)

Rensselaer

Tetherless World

"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

Tetherless World

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

Tetherless World