Geospatial Semantic Web

Harry Chen
harryc@imagemattersllc.com
Image Matters LLC

An invited talk at UMBC (CMSC 491/691M)
March 27, 2007

IMAGE MATTERS
Locate. Innovate.

Outline

- What's Geospatial Semantic Web
- Real world problems
- State-of-the-art solutions
- Demo
- Q&A

??

+

IMAGE MATTERS
Locate. Innovate.

What's Geospatial Semantic Web
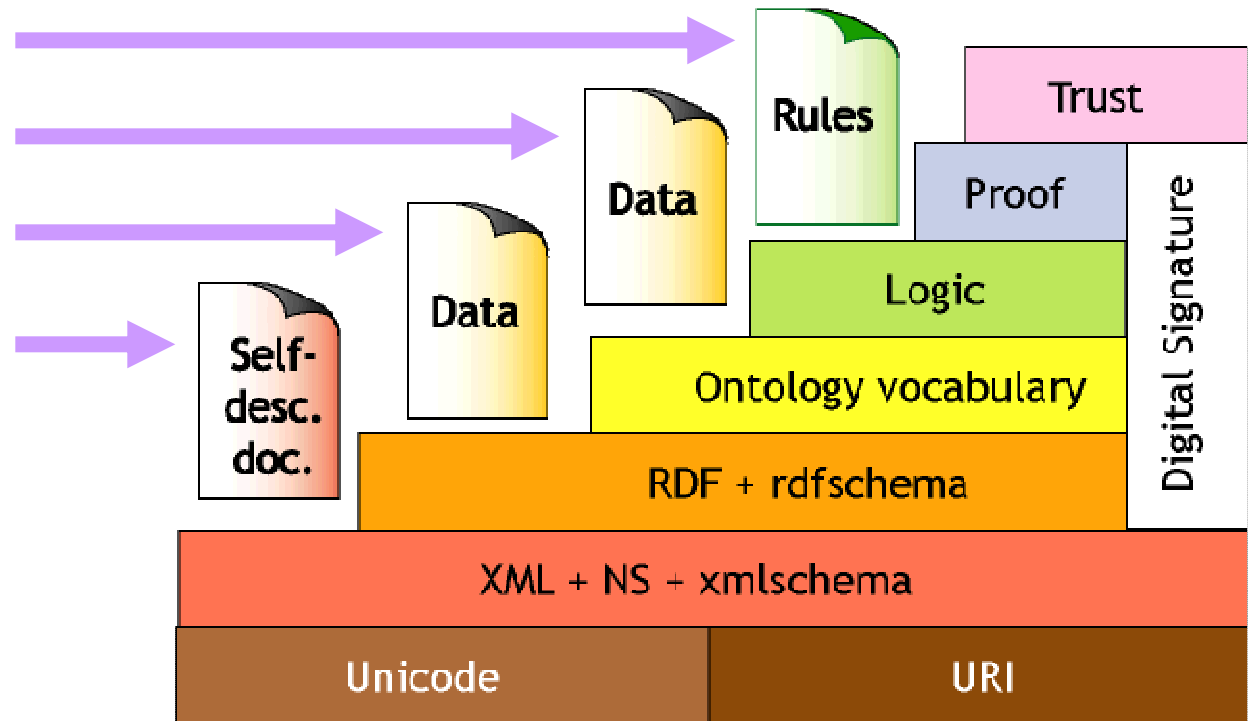
Geospatial + Semantic Web

Geospatial Data
- Geographic info
- Photo imagery
- GIS data
- Maps

- Spatial relations
- Places, landmarks etc.
- Locations, lat/long etc.

- Time
- Temporal relations

- People
- Organizations
- Other things …

Rules

Trust

Data

Proof

Data

Logic

Self-desc. doc.

Data

Ontology vocabulary

RDF + rdfschema

Digital Signature

XML + NS + xmlschema

Unicode

URI

IMAGE MATTERS
Locate. Innovate.

Why is this interesting?

- **"Location" is ubiquitous on the Web**
  - Where do you go to school or work?
  - Where did you take your flickr photos?
  - Where is the nearest gas-station from "here"?
  - Where are my friends now?
  - What's the avg. housing price in my neighborhood?
  - What's ski condition in MD and PA?
  - …

IMAGE MATTERS
Locate. Innovate.

The present Web is for human

http://hchen1.com/contact-me/biosketch/

Your browser doesn't know that my wife and I live in Columbia Maryland.
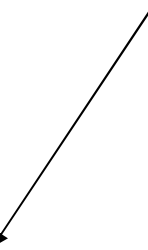
What's "london"?

It's "London UK" stupid!

IMAGE MATTERS
Locate. Innovate.

What's "london" to a machine?

	Name	Country	Feature class	Latitude	Longitude
					2544 records found for "london"
1	London Mountain	United States, Colorado	mountain elevation 4018m	N 39° 17' 3"	W 106° 9' 30"
2	New London County	United States, Connecticut	second-order administrative division population 254,957, elevation 22m	N 41° 29' 0"	W 72° 3' 58"
3	New London Nova Londres	United States, Connecticut	populated place population 26,294, elevation 17m	N 41° 21' 20"	W 72° 5' 58"
4	London	United States, Kentucky	populated place population 7,844, elevation 378m	N 37° 7' 44"	W 84° 4' 59"
5	London	United States, Ohio	populated place population 8,771, elevation 321m	N 39° 53' 11"	W 83° 26' 53"
6	New London	United States, Iowa	populated place population 1,857, elevation 232m	N 40° 55' 37"	W 91° 23' 58"
7	New London	United States, Minnesota	populated place population 1,123, elevation 368m	N 45° 18' 3"	W 94° 56' 39"
8	New London	United States, Wisconsin	populated place population 7,043, elevation 234m	N 44° 23' 33"	W 88° 44' 23"
9	London Londres	United States, California	populated place population 1,985, elevation 91m	N 36° 28' 33"	W 119° 26' 35"

london   United States   search   show on map   [advanced search]

http://www.geonames.org/search.html?q=london&country=US
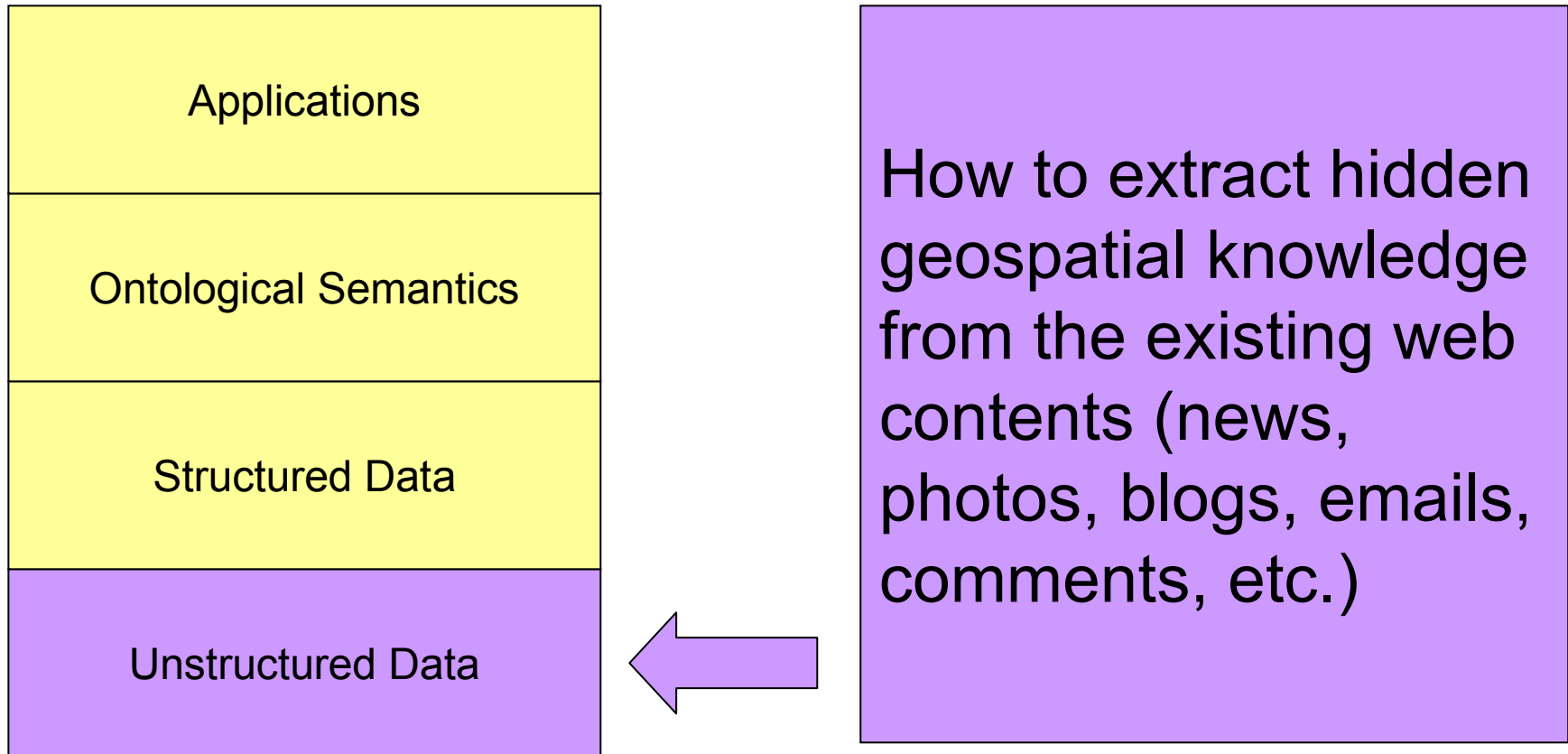
IMAGE MATTERS
Locate. Innovate.

What did we learn?

- Most of the information on the Web today is meant for human consumption.
- Without an explicit semantic description, it's difficult for machines to consume Web information.
- The study of geospatial semantic web is to exploit Semantic Web and geospatial technology to improve human productivity
  - i.e., get machines to do more work for us.

Real World Problems

Hidden Knowledge on the Web

Applications
Ontological Semantics
Structured Data
Unstructured Data

How to extract hidden geospatial knowledge from the existing web contents (news, photos, blogs, emails, comments, etc.)

IMAGE MATTERS
Locate. Innovate.

Scraping semantics from free text

Location information hidden inside unstructured free text

Scraping semantics from tags

Knowledge integration

Applications
Ontological Semantics
Structured Data
Unstructured Data

How to enable heterogeneous data stores to integrate and interoperate (RDBMS, shapefiles, XML, gazetteer, photo annotations, etc.)

Fuse knowledge from multiple datastores
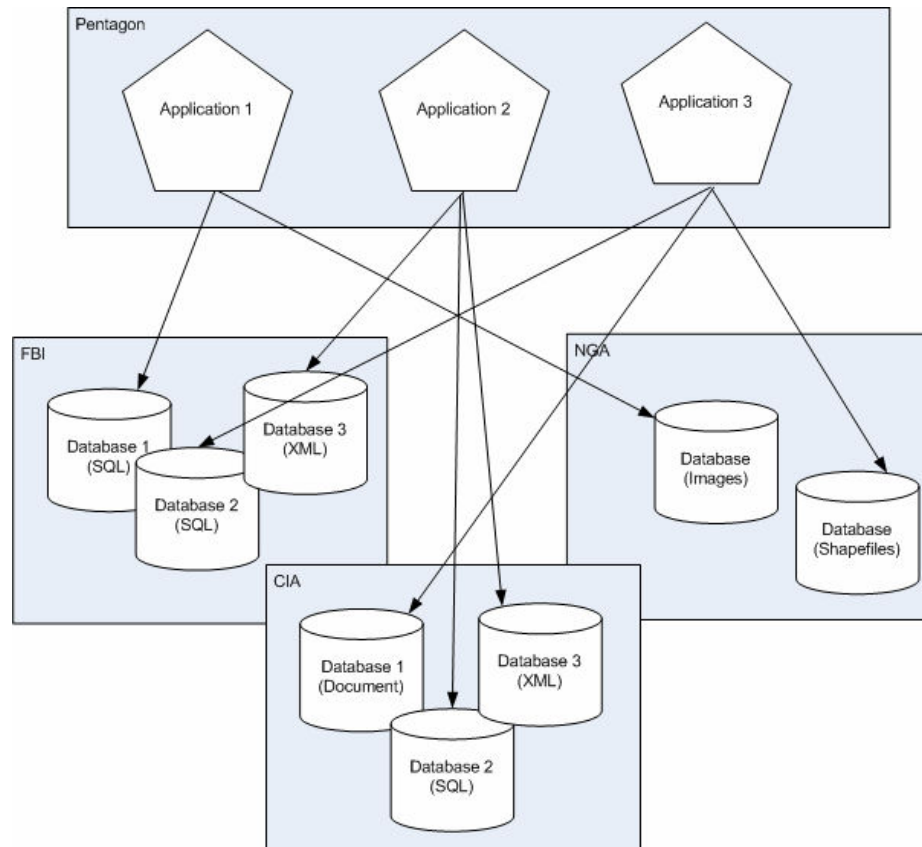
Ontology sharing & reuse
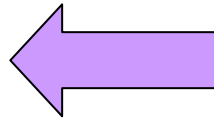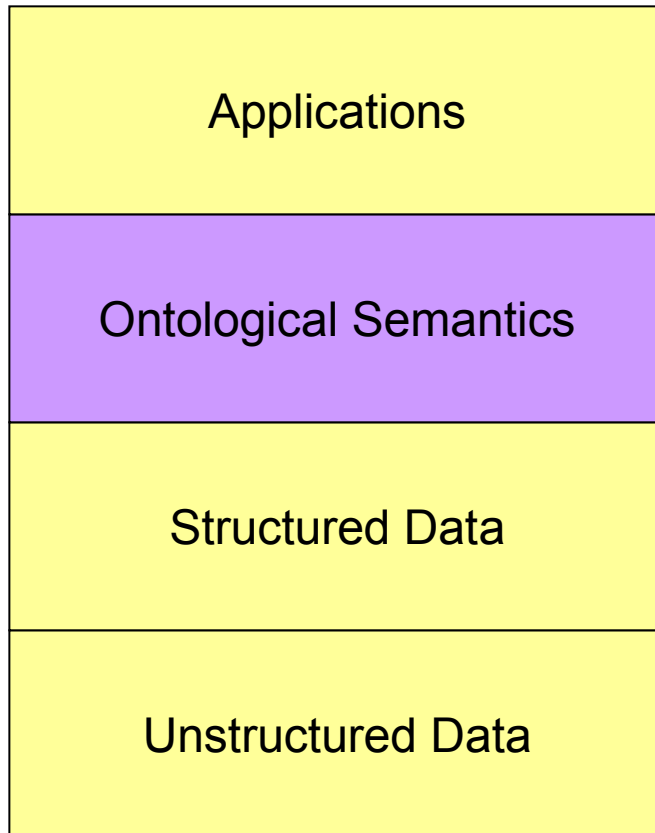
Many different ways to describe Geo

- **GML** (Geographical Markup Language) backed by OGC (Open Geospatial Consortium)
- **W3C geo ontology** backed by W3C
- **GeoRSS** – a language for embedding location information in RSS
- **Geonames ontology** – an open and free geographical data service
- **Google KML** – used by Google Maps and Google Earth

Application research issues

Applications
Ontological Semantics
Structured Data
Unstructured Data
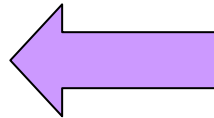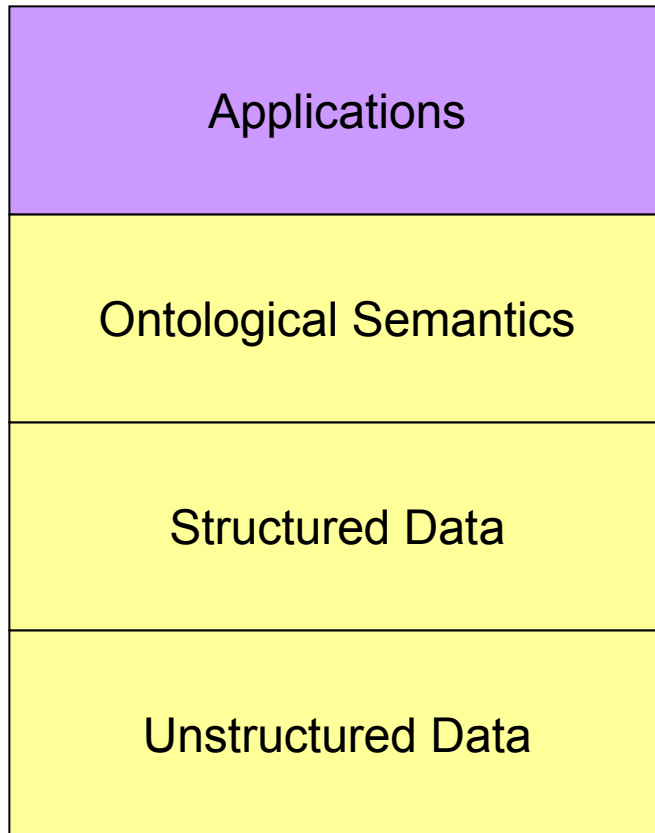
How to make easy for people to publish, annotate and search geospatial semantic information on the Web

IMAGE MATTERS
Locate. Innovate.

A mystery to many people

```
<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="http://hchen1.com/people/harrychen"/>
  <foaf:primaryTopic rdf:resource="http://hchen1.com/people/harrychen"/>
  <admin:generatorAgent rdf:resource="http://www.ldodds.com/foaf/
foaf-a-matic"/>
  <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>

<foaf:Person rdf:about="http://hchen1.com/people/harrychen">
  <foaf:name>Harry Chen</foaf:name>
  <foaf:title>Dr</foaf:title>
  <foaf:givenname>Harry</foaf:givenname>
  <foaf:family_name>Chen</foaf:family_name>
  <foaf:nick>hchen1</foaf:nick>
  <foaf:mbox_sha1sum>80368445d942c6e207c6694353355c3bd37f3e7a</foaf:mbox_sha1sum>
  <foaf:homepage rdf:resource="http://harry.hchen1.com"/>
  <foaf:depiction rdf:resource="http://harry.hchen1.com/images/hchen-small.jpg"/>
  <foaf:workplaceHomepage rdf:resource="http://www.imagemattersllc.com"/>
  <foaf:schoolHomepage rdf:resource="http://www.umbc.edu"/>

  <foaf:knows>
    <foaf:Person>
      <foaf:name>Tim Finin</foaf:name>
      <foaf:mbox_sha1sum>49953f47b9c33484a753eaf14102af56c0148d37</foaf:mbox_sha1sum>
      <rdfs:seeAlso rdf:resource="http://www.cs.umbc.edu/~finin/foaf.rdf"/>
    </foaf:Person>
  </foaf:knows>
```

- Do you really believe companies can make money from products that require people to edit and query data like this…

IMAGE MATTERS
Locate. Innovate.

State-of-the-art Solutions

Technology overview

- **S**emantic **W**eb vs. **s**emantic **w**eb
- Publishing W3C geo data on the Web
  - RDF, Microformats & RDFa
  - GeoRSS & flickr machine tags
- Knowledge integration technology
- Mashups and browser add-ons

IMAGE MATTERS
Locate. Innovate.

Semantic Web vs. semantic web

- There are two different schools of thinking on how semantics should be published, exported and searched on the Web.
  - http://tantek.com/presentations/2004etech/realworldsemanticspres.html
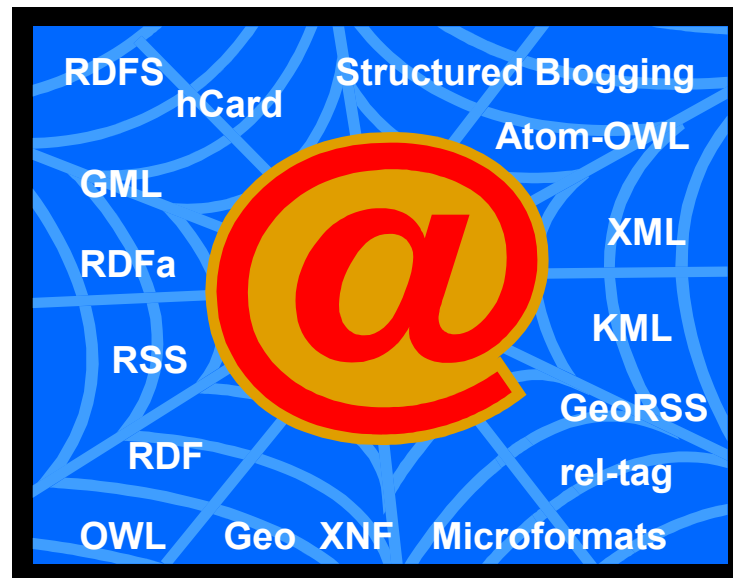
Semantic Web vs. semantic web

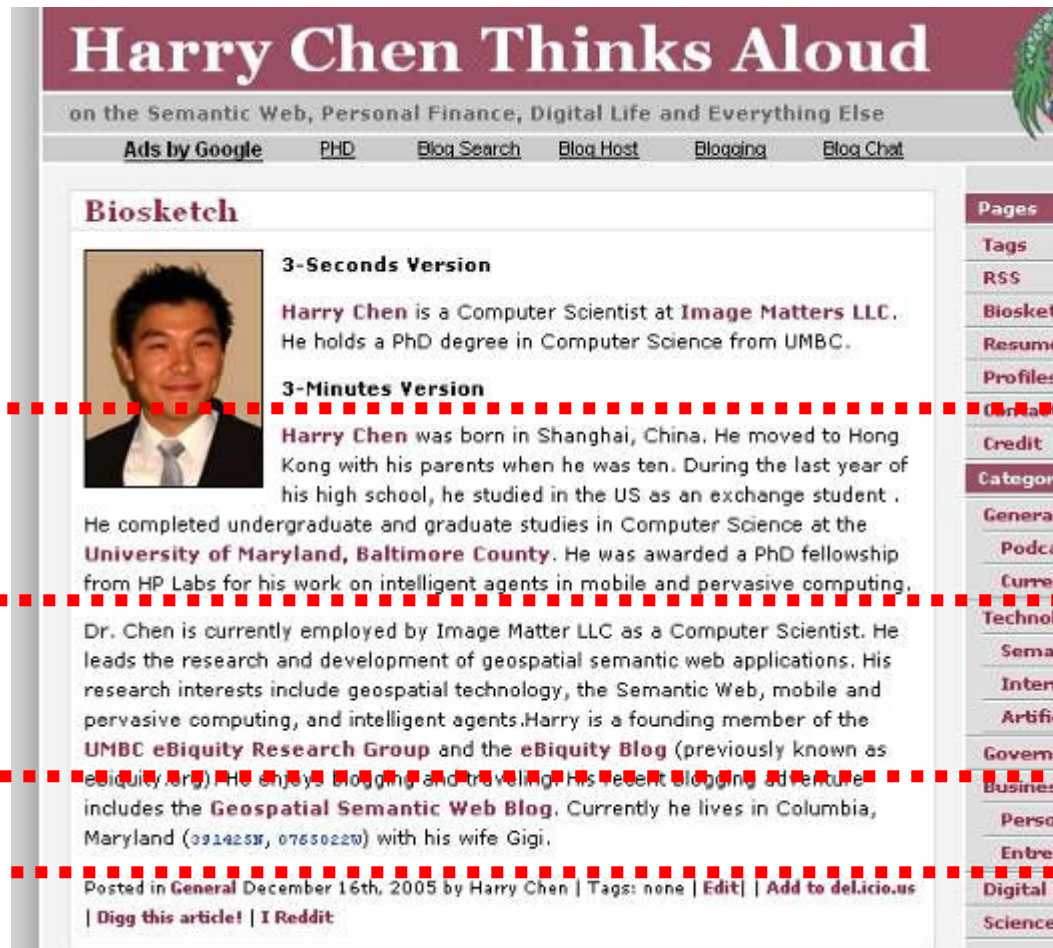|	Semantic Web	semantic web
Philosophy	Build a common data format for expressing the meaning of data. Use ontologies to help machines to understand web content.	Humans first, machines second. Encode existing Web content with special tags.
Language	**RDF, RDFS, OWL**	**Microformats (based on XHTML)**
Format	Must be well-formed RDF documents	Anything goes, as long as its XHTML
Semantic	Defined by the underlying ontology model (e.g., OWL)	Loosely defined. No formal semantic model.
Examples	FOAF, OWL-S, OWL-Time	XFN (social network), hCard (contact), hReview (opinions), rel-tag (tagging)

Using W3C Geo (RDF)

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns="http://xmlns.com/foaf/0.1/">
<Person>
  <name>Dan Brickley</name>
  <homepage dc:title="Dan's home page"
          rdf:resource="http://danbri.org/"/>
  <based_near geo:lat="51.47026" geo:long="-2.59466"/>
  <rdfs:seeAlso rdf:resource="http:/danbri.org/foaf.rdf"/>
</Person>
</rdf:RDF>
```

Subject	Predicate	Object
:_a	rdf:type	foaf:Person
:_a	foaf:name	"Dan Brickley"
:_a	foaf:base_near	:_geo
:_geo	geo:long	"-2.59466"
:_geo	geo:lat	"51.47026"

Source: http://www.w3.org/2003/01/geo/

IMAGE MATTERS
Locate. Innovate.

Using Microformats Geo (1)

Using Microformats (2)

```
<p class="vcard">
   <a class="url fn" href="http://harry.hchen1.com">Harry Chen</a> was born in Shanghai,
China. He moved to Hong Kong with his parents when he was ten. During the last year of his
high school, he studied in the US as an exchange student . He completed undergraduate and
graduate studies in Computer Science at the
<a title="UMBC" class="org" href="http://www.umbc.edu">University of Maryland, Baltimore
County</a>. He was awarded a PhD fellowship from HP Labs for his work on intelligent agents
in mobile and pervasive computing.
</p>
```

```
…
Currently he lives in Columbia, Maryland (
<span class="geo">
        <span class="latitude"> 39.207806 </span>,
        <span class="longitude">-76.825475 </span>
) with his wife Gigi.
</span>
…
```

Technorati: Microformats Search

Not yet understand "geo", but that's okay. At least it works!

http://kitchen.technorati.com

RDFa – embedding RDF in XHTML

```
<html xmlns:foaf="http://xmlns.com/foaf/0.1#"
     xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
…
<p class="foaf:Person" about="#me">
  <a rel="foaf:homepage" href="http://harry.hchen1.com">Harry Chen</a> was born in
Shanghai, China. He moved to Hong Kong with his parents when he was ten. During the last
year of his high school, he studied in the US as an exchange student . He completed
undergraduate and graduate studies in Computer Science at the
<a title="UMBC" rel="foaf:schoolHomepage" href="http://www.umbc.edu">University of
Maryland, Baltimore County</a>. He was awarded a PhD fellowship from HP Labs for his work
on intelligent agents in mobile and pervasive computing.
…
Currently he lives in Columbia, Maryland (
<div rel="foaf:base_near">
<span class="geo:Point">
        <span property="geo:lat">39.207806</span>,
        <span property="geo:long">-76.825475</span>
) with his wife Gigi.
</span>
</div>
</p>
```

IMAGE MATTERS
Locate. Innovate.

Scraping Triples from RDFa

Subject	Predicate	Object
<#me>	rdf:type	foaf:Person
<#me>	foaf:homepage	http://harry.hchen1.com
<#me>	foaf:schoolHomepage	http://www.umbc.edu
<#me>	foaf:base_near	_:aa0
_:aa0	rdf:type	geo:Point
_:aa0	Geo:lat	39.207806
_:aa0	geo:long	-76.825475

GeoRSS

- GeoRSS is a language for embedding location object descriptions in RSS
- Three kinds of format representations are supported:
  - GeoRSS Simple
  - W3C Geo
  - GML Geometry
- How GeoRSS can be used:
  - Describe the reporting location of a news
  - Describe the location history of a traveler

IMAGE MATTERS
Locate. Innovate.

W3C Geo in GeoRSS

```
<item rdf:about= 'http://www.oreillynet.com/pub/wlg/8466?CMP=OTC-TY3388567169'>
    <title>Live Coverage XML 2005 (Tuesday Keynotes)</title>
    <link>http://www.oreillynet.com/pub/wlg/8466?CMP=OTC-TY3388567169</link>
    <description> <![CDATA[A live look at the XML Keynotes and seminal talks.]]>
</description>
    <dc:creator>Kurt Cagle</dc:creator>
    <dc:date>2005-11-15T07:45:58-08:00</dc:date>
    <geo:Point>
    <geo:lat>46.183</geo:lat>
    <geo:long>-123.816</geo:long>
   </geo:Point>
</item>
```

Embedding GeoRSS in a RSS 1.0 document

IMAGE MATTERS
Locate. Innovate.

GeoRSS Simple properties

- Point
- Line
- Polygon
- Box
- Elevation
- Radius
- Feature Type
- Feature Name
- Relationship Tags

For more information:
  http://www.georss.org/simple.html

IMAGE MATTERS
Locate. Innovate.

Machine tags

- Flickr machine tags are a special kind of tags for defining semantic information

> Like tags, there are no rules for machine tags beyond the syntax to specify the parts of a machine tag. For example, you could tag a photo with :
>
> * flickr:user=straup
>
> * flora:tree=coniferous
>
> * medium:paint=oil
>
> * geo:quartier="plateau mont royal"
>
> * geo:neighbourhood=geo:quartier
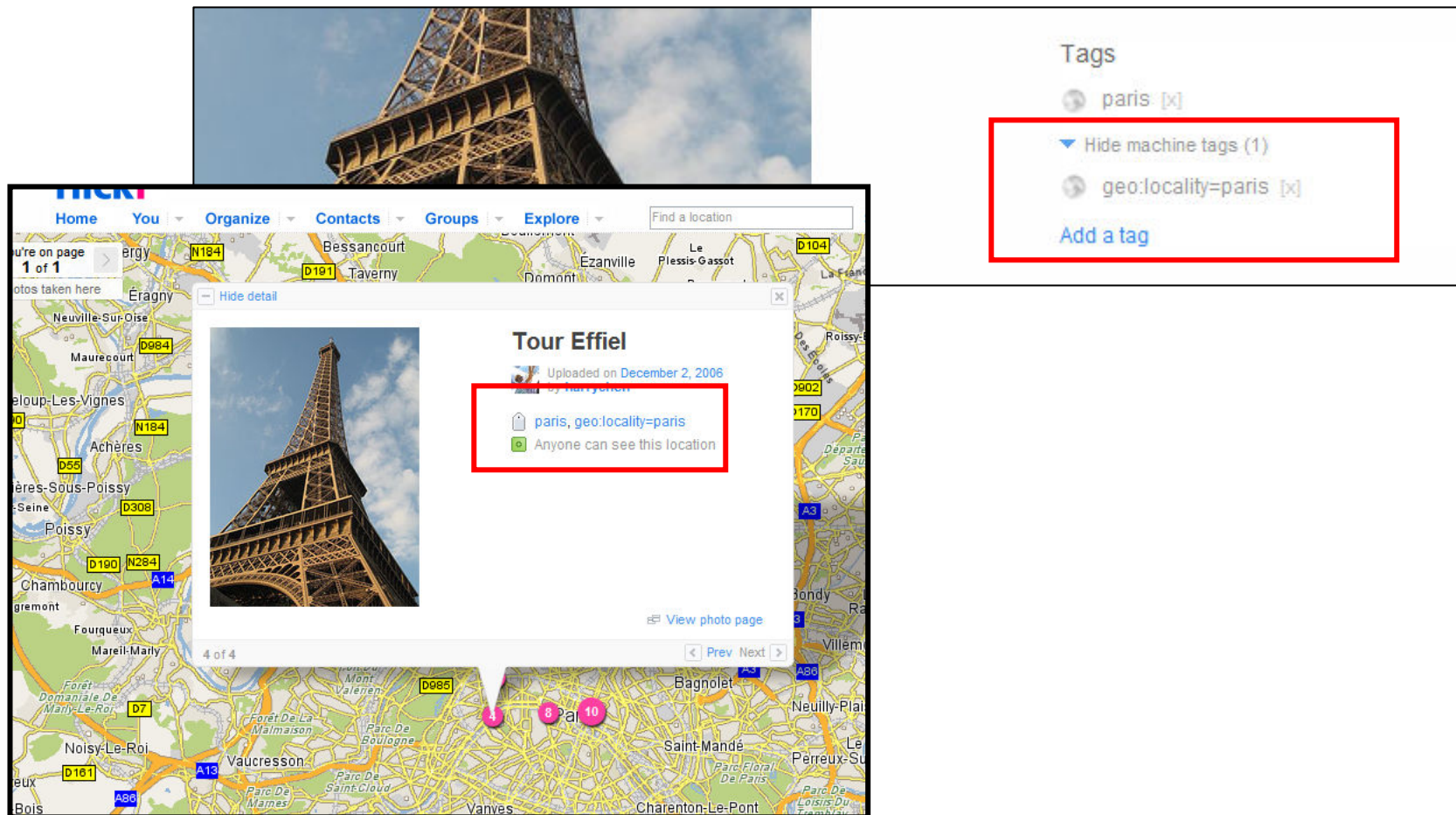
For more information:
  http://www.flickr.com/groups/api/discuss/72157594497877875/

A machine tag example

Remember our example?

How to effective integrate knowledge stored in multiple data sources with heterogeneous representations?

Why knowledge integration?

- Much data is hidden in our legacy systems. We must find ways to export this data onto the Web

Web pages are designed for people. For the Semantic Web we need to look at existing databases and the data in them.
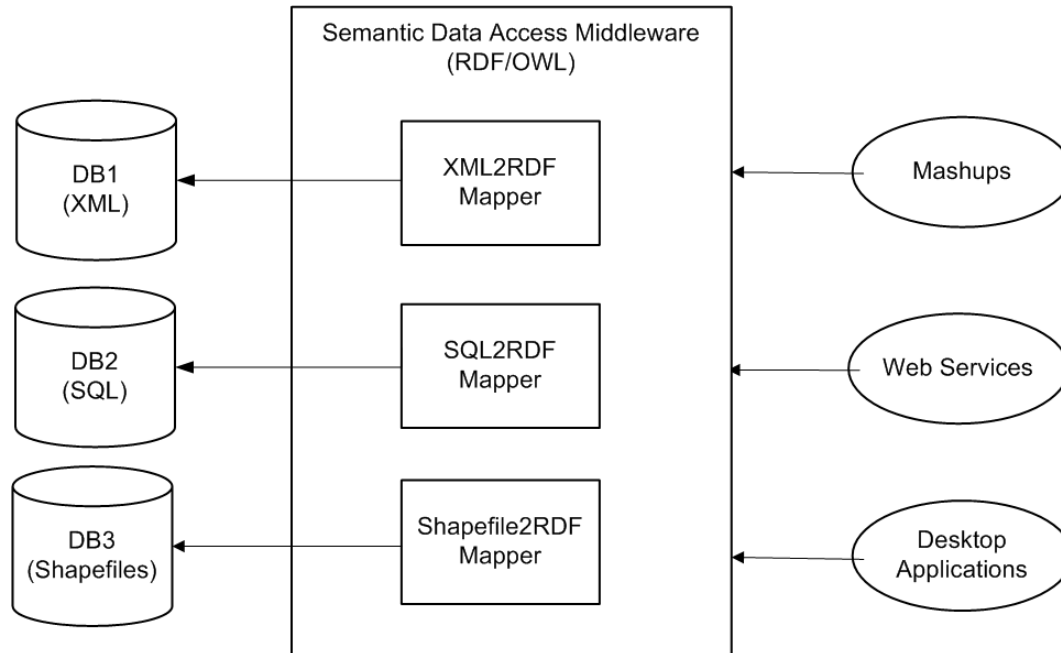
-Tim Berners-Lee, March 2006.

http://www.bcs.org/server.php?show=ConWebDoc.3337

IMAGE MATTERS
Locate. Innovate.

Getting data onto the Web

- Approach 1: Consolidate everything into a single database
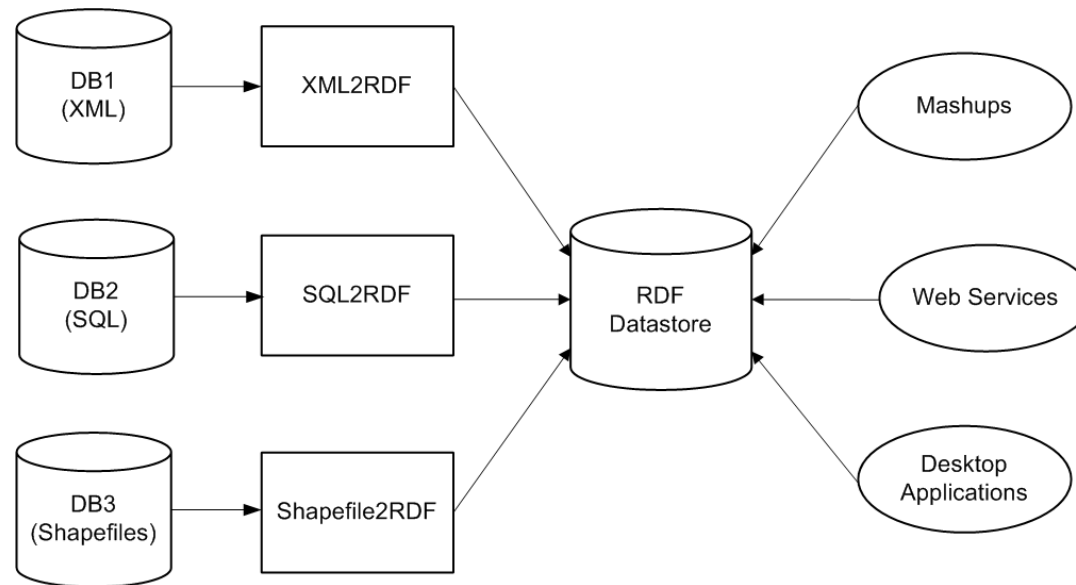
Getting data onto the Web

- Approach 2: Dynamically integrate data into a uniformed representation

Data integration systems

- **Oracle RDF database** (Oracle)
  - Supports full RDF and RDFS
  - Support SQL query over RDF graph model
  - Built-in subsumption support: subClassOf and subPropertyOf
- **D2RQ** (Freie Universität Berlin):
  - Declarative language for describing mappings between relational DB schemas and RDFS/OWL ontologies
  - Support SQL
  - D2RQ Server allows accesses to SQL using SPARQL queries over HTTP
- **KnowledgeSmarts** (Image Matters LLC)
  - A middle-ware system for knowledge integration over heterogeneous datastores
  - Supports SQL, Shapefiles, XML, WFS and more.
  - Optimized for applications that require spatial and temporal computation support.

IMAGE MATTERS
Locate. Innovate.

An explosion of mashups

A **mashup** is a website or web application that uses content from more than one source to create a completely new service.

Source: Wikipedia -- http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)

TOP MASHUP TAGS

- Mapping
- Photo
- Search
- Shopping
- Sports
- Transit
- News
- Travel
- Real estate
- NYC
- Other

Check Real Estate Value

Track Ski Conditions

Track Storms

IMAGE MATTERS
Locate. Innovate.

Visiting the White House
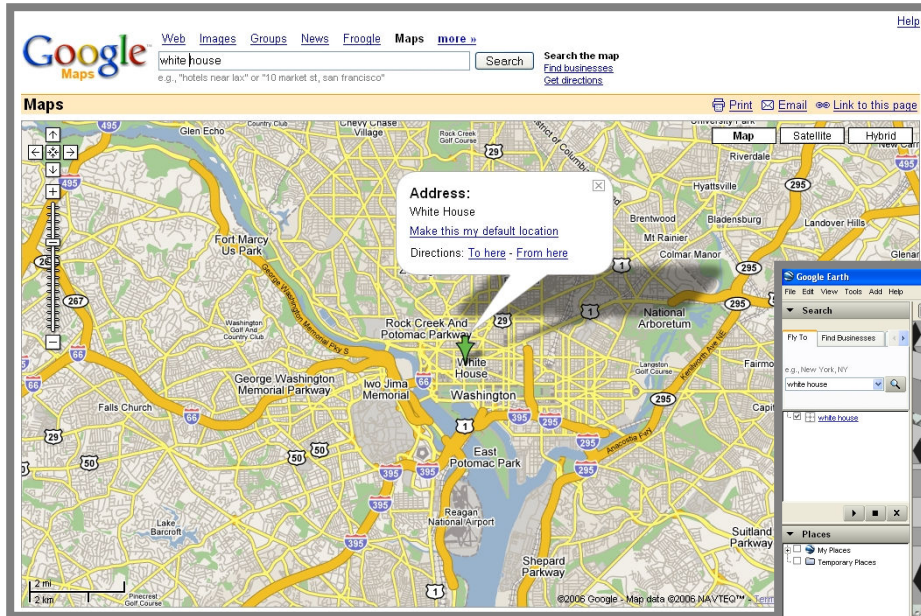
The White House in Google Maps

The White House in Google Earth

IMAGE MATTERS
Locate. Innovate.
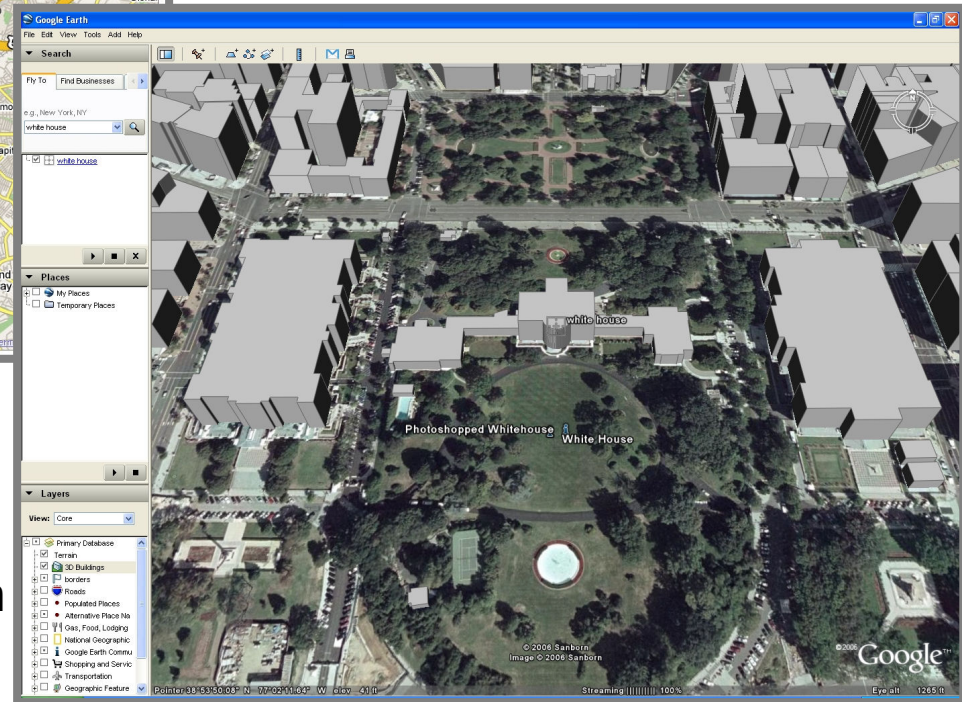
Special features in Google Earth
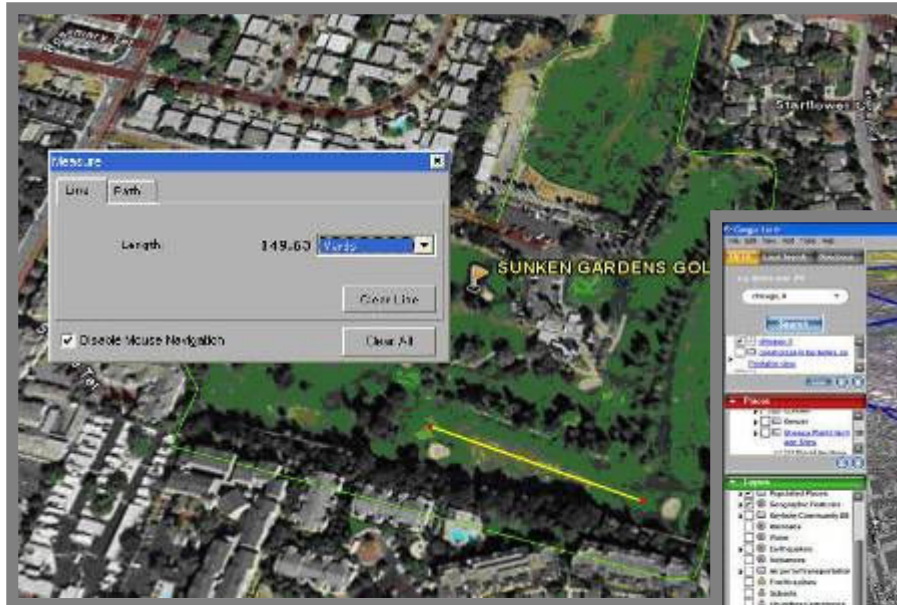
Measure Distances

3D buildings and terrian

IMAGE MATTERS
Locate. Innovate.
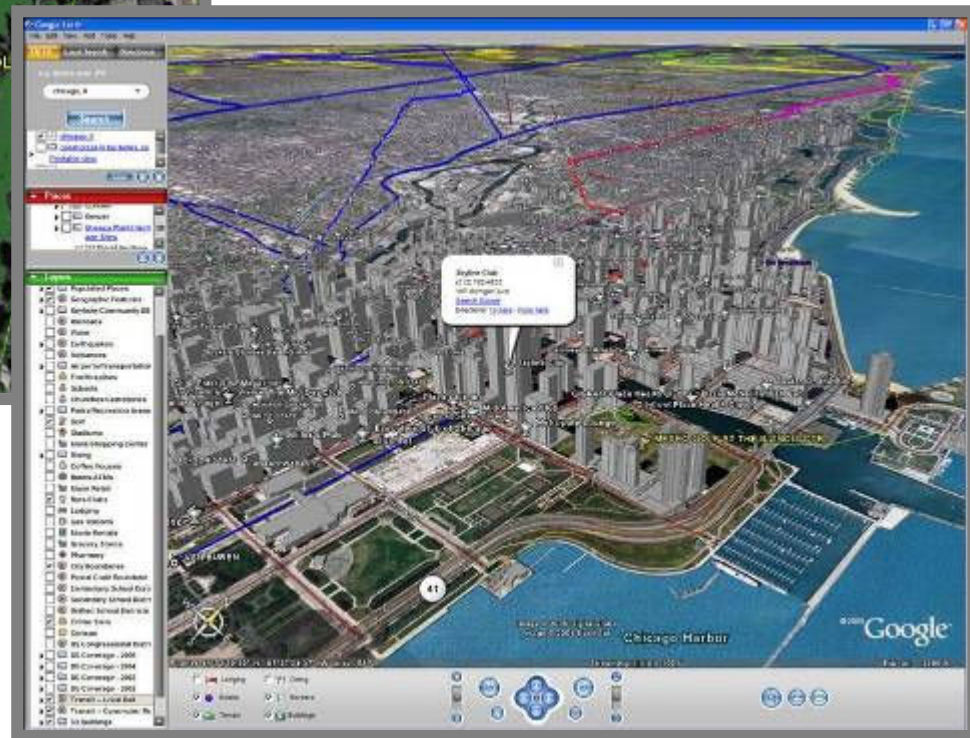
Questions

- Why is there a sudden explosion of "mashups"?

- Is it the "holy grail" in building the next generation Web?

- What's the use of semantic technology in building mashups?

Mashups are growing fast

- Ubiquitous web service API
  - Google Maps, Yahoo! Maps, Amazon, Flickr, del.icio.us, etc.
- People can create new applications by reusing the existing parts
  - The whole is more than the sum of its parts
- Maps are intuitive UI interface.

Mashup issues (1 of 3)

- The present Web is built for human users. Information is meant for humans to consume and not for computer programs.
  - A map image is a map to the humans, but is a image to the machines.

IMAGE MATTERS
Locate. Innovate.

Mashup issues (2 of 3)

- It's difficult to discover and integrate legacy data into new mashup applications.
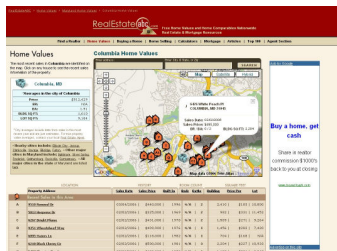
Real Estate Value Mashup

Where can I find real estate data?
Data format? Permission to use it?
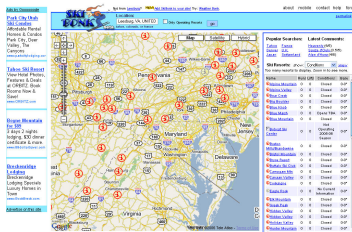
National Ski Condition Mashup

Where can I find weather data?
Data format? Permission to use it?

IMAGE MATTERS
Locate. Innovate.

Mashup issues (3 of 3)

- ## Too many wrongly think that mashups must be Google Maps on "steroid".
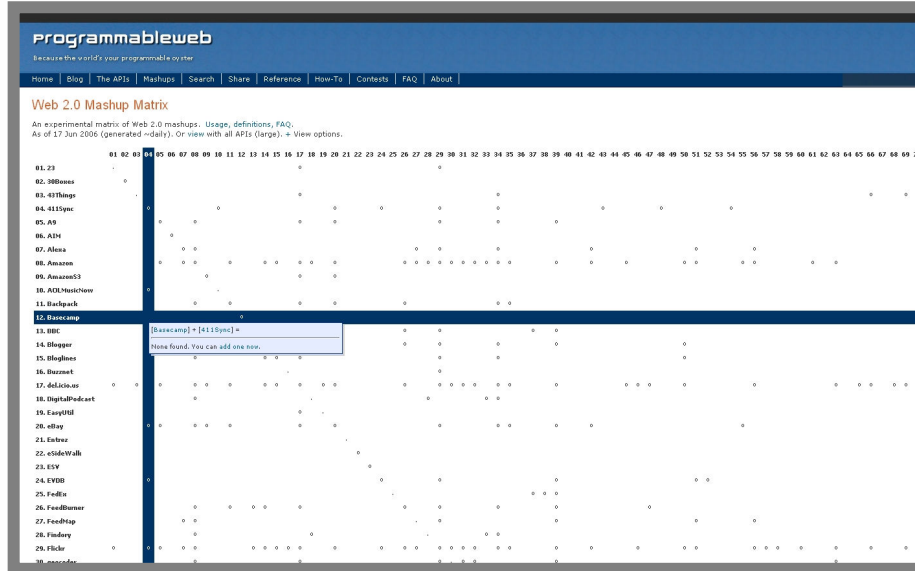
Web 2.0 Mashup Matrix

- Records 104 Web 2.0 API
- 104 x 104 possibilities
- Google Maps 1 of 104

http://www.programmableweb.com/matrix

IMAGE MATTERS
Locate. Innovate.

Semantic mashup: Piggy Bank

- **Piggy Bank** is a Firefox extension that uses JavaScript to scrape RDF triples from the Web.
  - Part of MIT's SIMILE project
  - http://simile.mit.edu/piggy-bank

Movies at Toronto.Com

Typical movies listing

Piggy Bank this information

Semantic Data in a Piggy Bank

Location information

Movies!!!!

Location! Location! Location!

Demo

Google Maps Supports GeoRSS

- Recently [Google Maps announced the support for GeoRSS](#) in addition to KML

Geonames

- ## The largest geographical database that is licensed under the Creative Common License (it's free!)

  - ### Search world locations by name

  - ### Open Source API

  - ### Query results in RDF

    - #### Geonames OWL + SKOS

IMAGE MATTERS
Locate. Innovate.
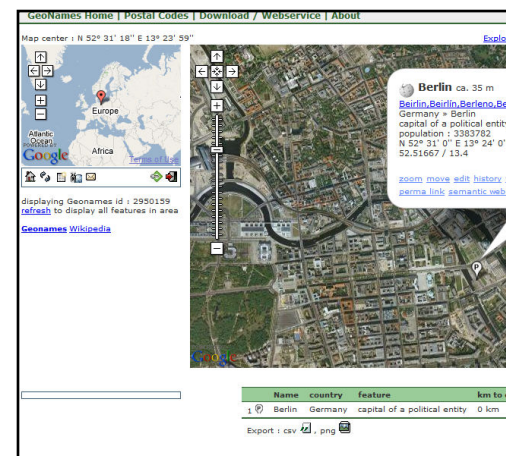
Concluding Remarks

- Geospatial data and applications are intrinsic part of the future Web

- A cross-fertilization of geospatial and Semantic Web technology will help to improve people's productivity

- We should build useful applications and not worry about the difference between Semantic Web and semantic web

IMAGE MATTERS
Locate. Innovate.

Web Resources

- Geospatial Semantic Web Blog
  - http://geospatialsemanticweb.com
- W3C Geospatial Incubator Group
  - http://www.w3.org/2005/Incubator/geo/
- W3C Geo Ontology
  - http://www.w3.org/2003/01/geo/
- GeoRSS
  - http://www.georss.org/
- Geonames
  - http://www.geonames.org
- Microformats Geo
  - http://microformats.org/wiki/geo
- RDFa (additional geo usage examples)
  - http://www.w3.org/2006/07/SWD/RDFa/syntax/