Operationalizing the Semantic

Fifth Semantic Interoperability and eGovernment conference, McLean, VA Oct 10, 2006

Eric Miller
President and Founder
The Kadomo Group
em@kadomo.com

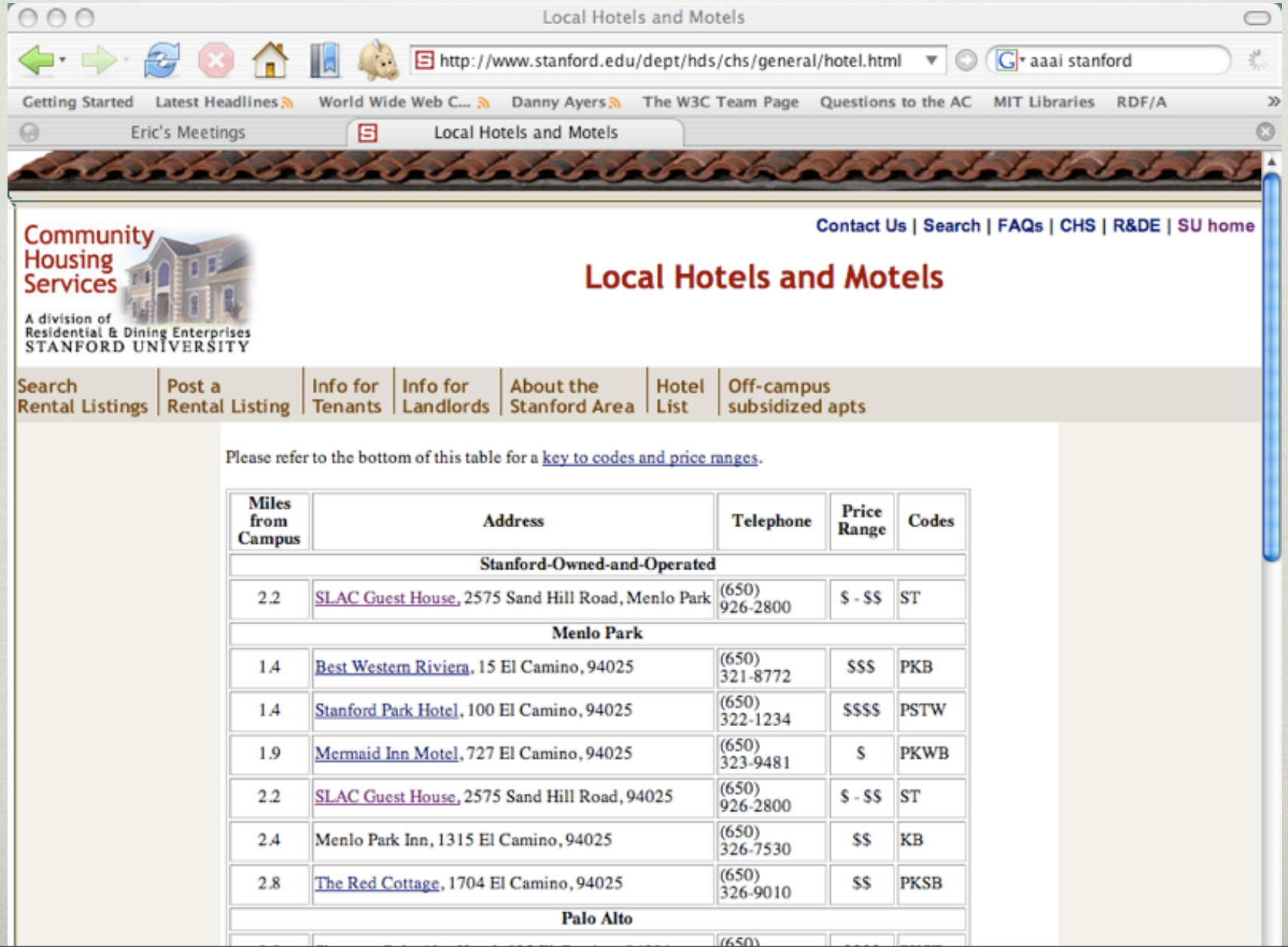
Semantic Web

- A Web of Data
- Freeing the data from the application that created it
- Enabling recombinant data - data integration and reuse across application, organizational and community boundaries.

A problem

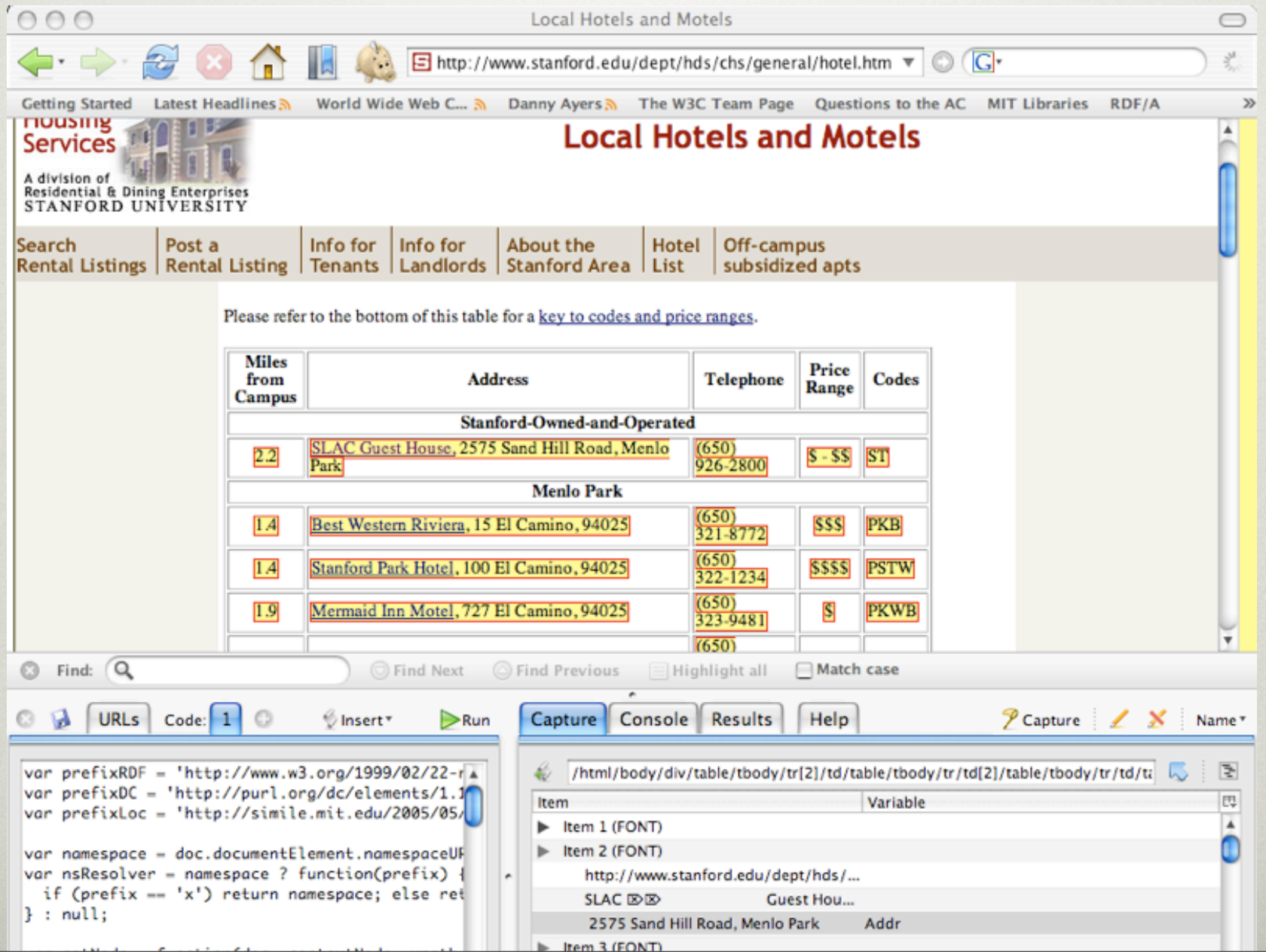
- Where to stay?
  - Hotel information on the web
  - Meeting location information on my computer

Useful data on the Web

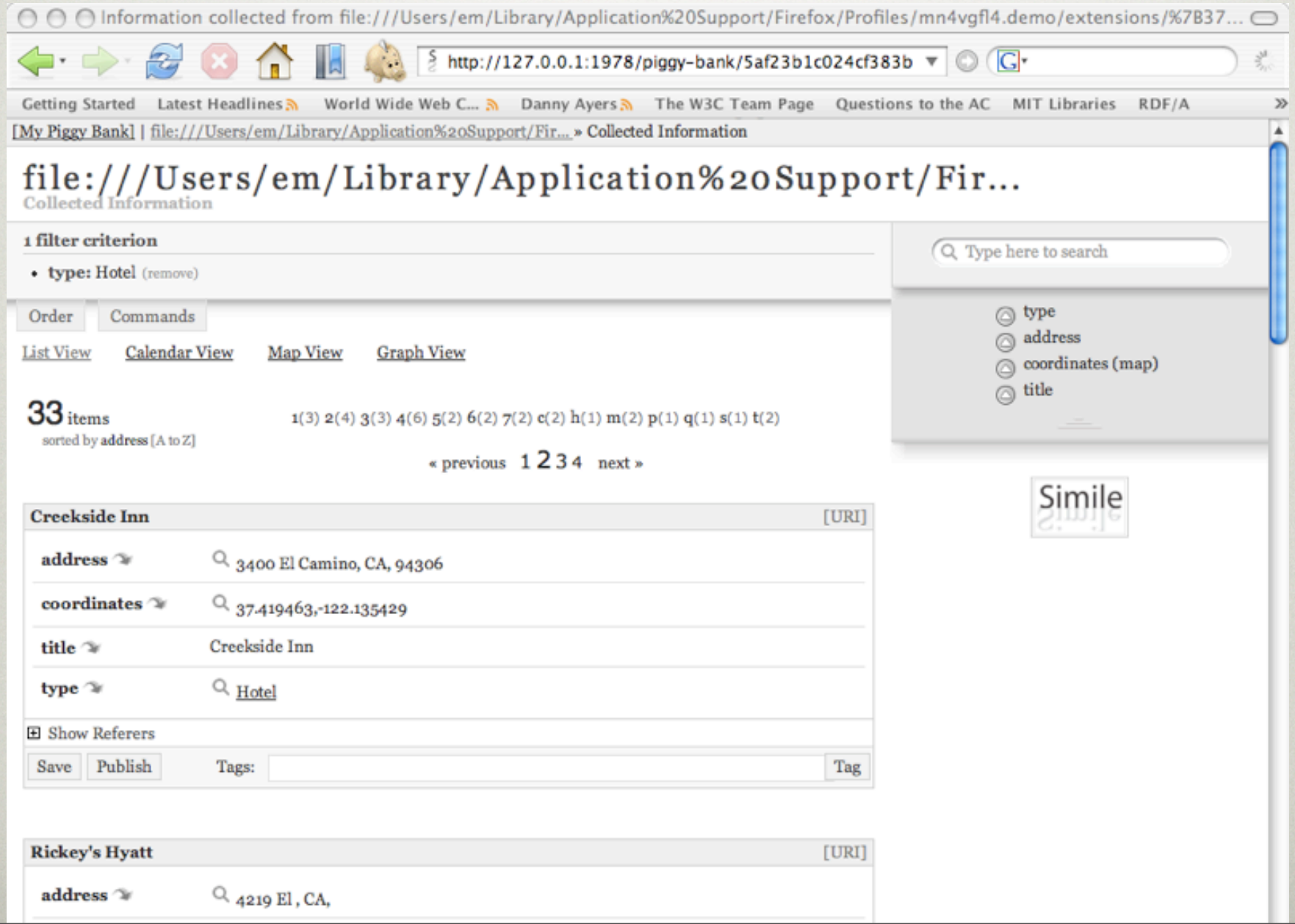


Miles from Campus	Address	Telephone	Price Range	Codes
Stanford-Owned-and-Operated				
2.2	SLAC Guest House, 2575 Sand Hill Road, Menlo Park	(650) 926-2800	$ - $$	ST
Menlo Park				
1.4	Best Western Riviera, 15 El Camino, 94025	(650) 321-8772	$$$	PKB
1.4	Stanford Park Hotel, 100 El Camino, 94025	(650) 322-1234	$$$$	PSTW
1.9	Mermaid Inn Motel, 727 El Camino, 94025	(650) 323-9481	$	PKWB
2.2	SLAC Guest House, 2575 Sand Hill Road, 94025	(650) 926-2800	$ - $$	ST
2.4	Menlo Park Inn, 1315 El Camino, 94025	(650) 326-7530	$$	KB
2.8	The Red Cottage, 1704 El Camino, 94025	(650) 326-9010	$$	PKSB
Palo Alto				

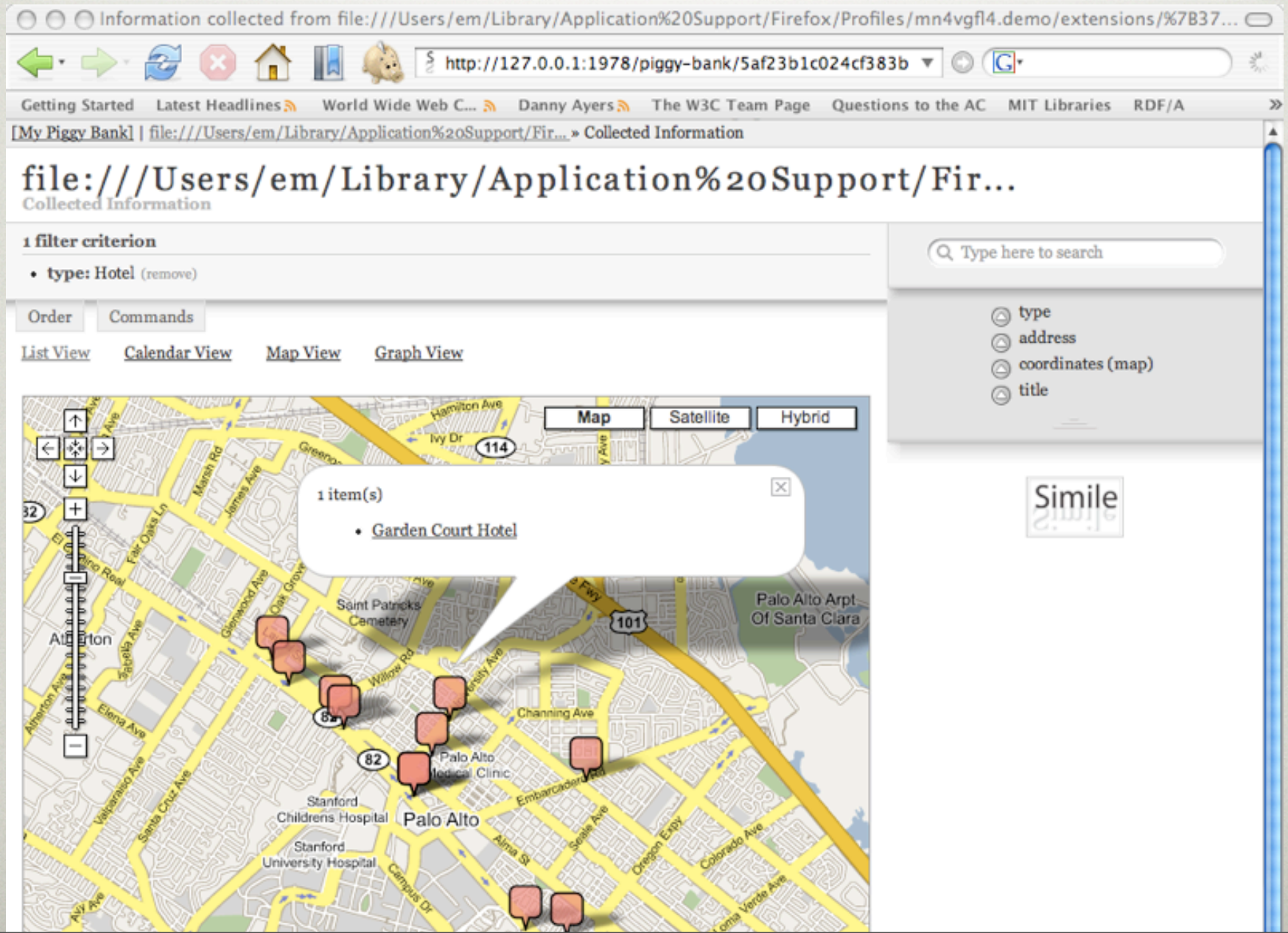
Representation Problem



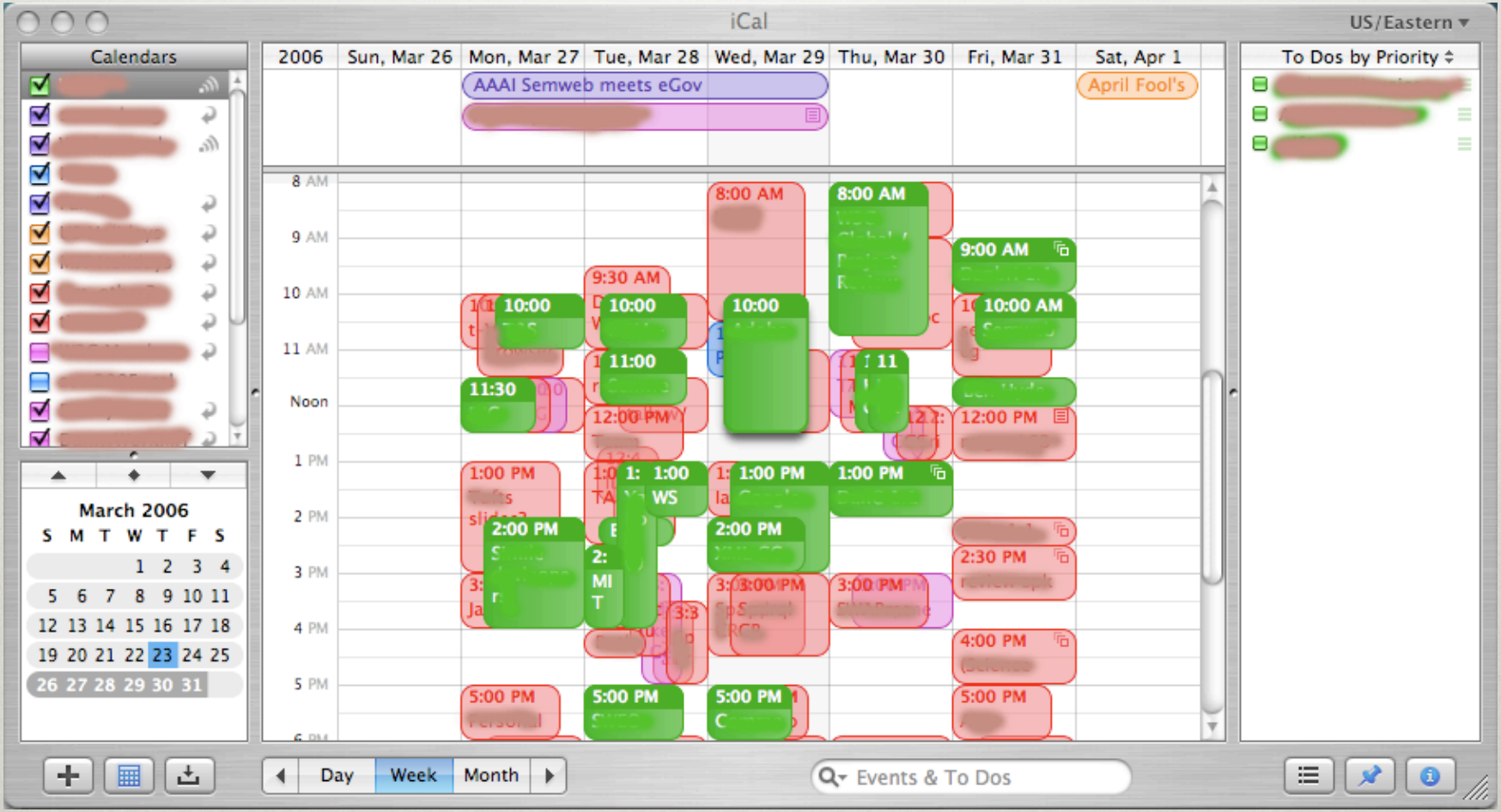
Problem Solved



Benefits of RDF



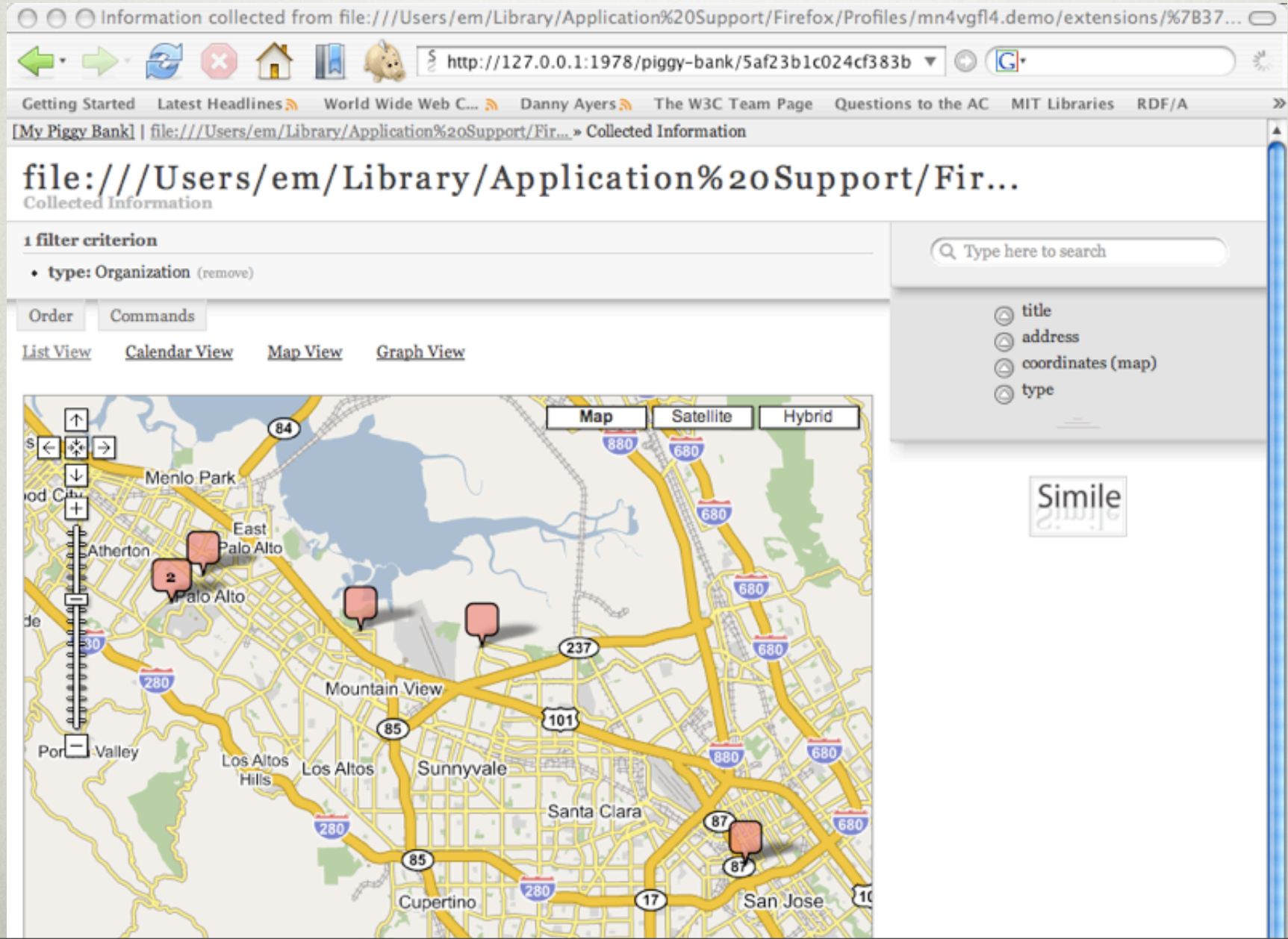
My Web


iCal
US/Eastern
Calendars
2006
Sun, Mar 26
Mon, Mar 27
Tue, Mar 28
Wed, Mar 29
Thu, Mar 30
Fri, Mar 31
Sat, Apr 1
AAAI Semweb meets eGov
April Fool's
To Dos by Priority
March 2006
S M T W T F S
1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28 29 30 31
8 AM
9 AM
10 AM
11 AM
Noon
1 PM
2 PM
3 PM
4 PM
5 PM
8:00 AM
9:00 AM
9:30 AM
10:00
10:00 AM
11:00
11:30
12:00 PM
1:00 PM
2:00 PM
2:30 PM
3:00 PM
4:00 PM
5:00 PM
Day
Week
Month
Events & To Dos

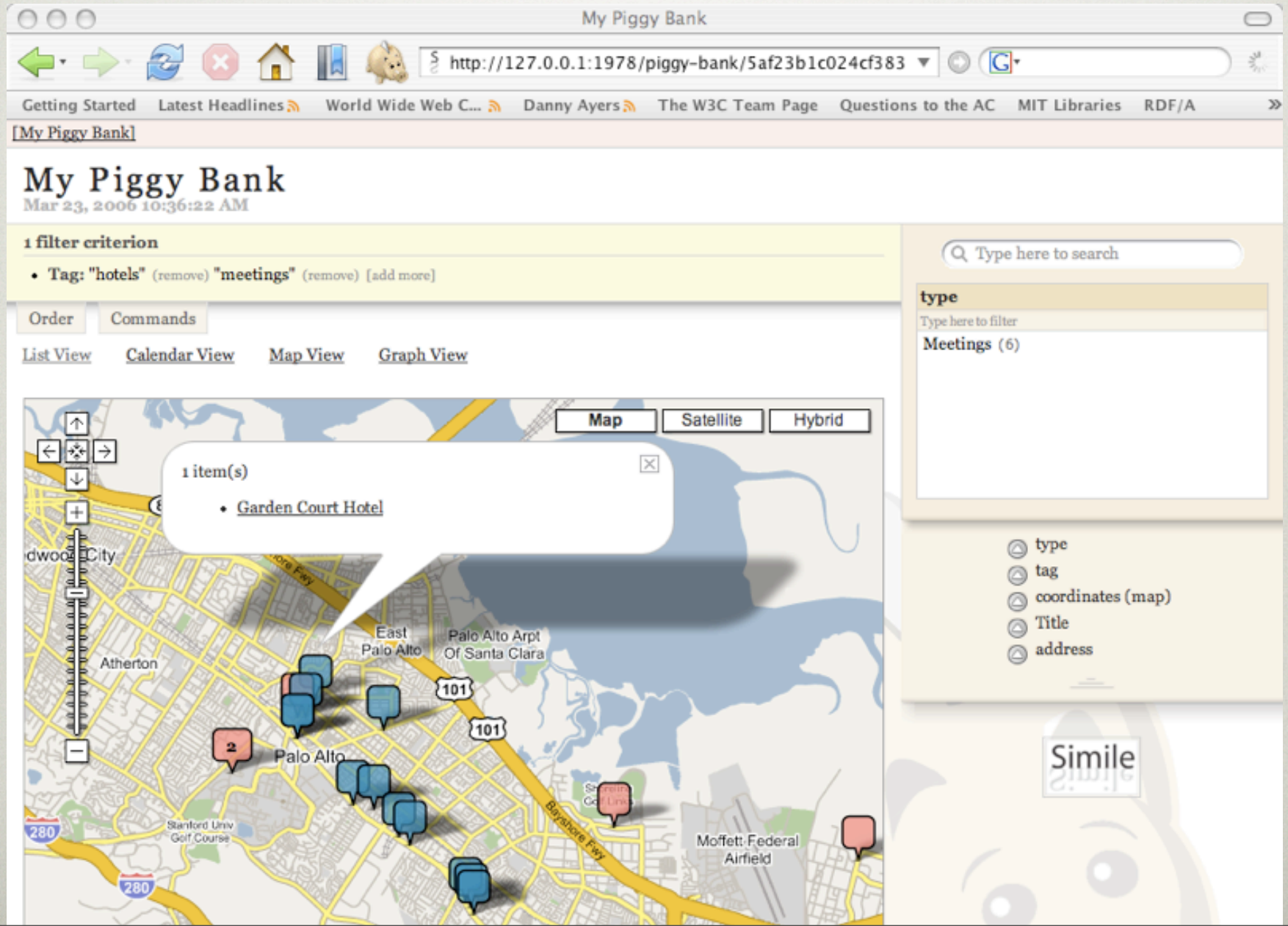
Another Problem

- Data is in iCal format
  - useful beyond just Calendar application, but not accessible
- Solution
  - iCal to RDF ... 3 seconds later

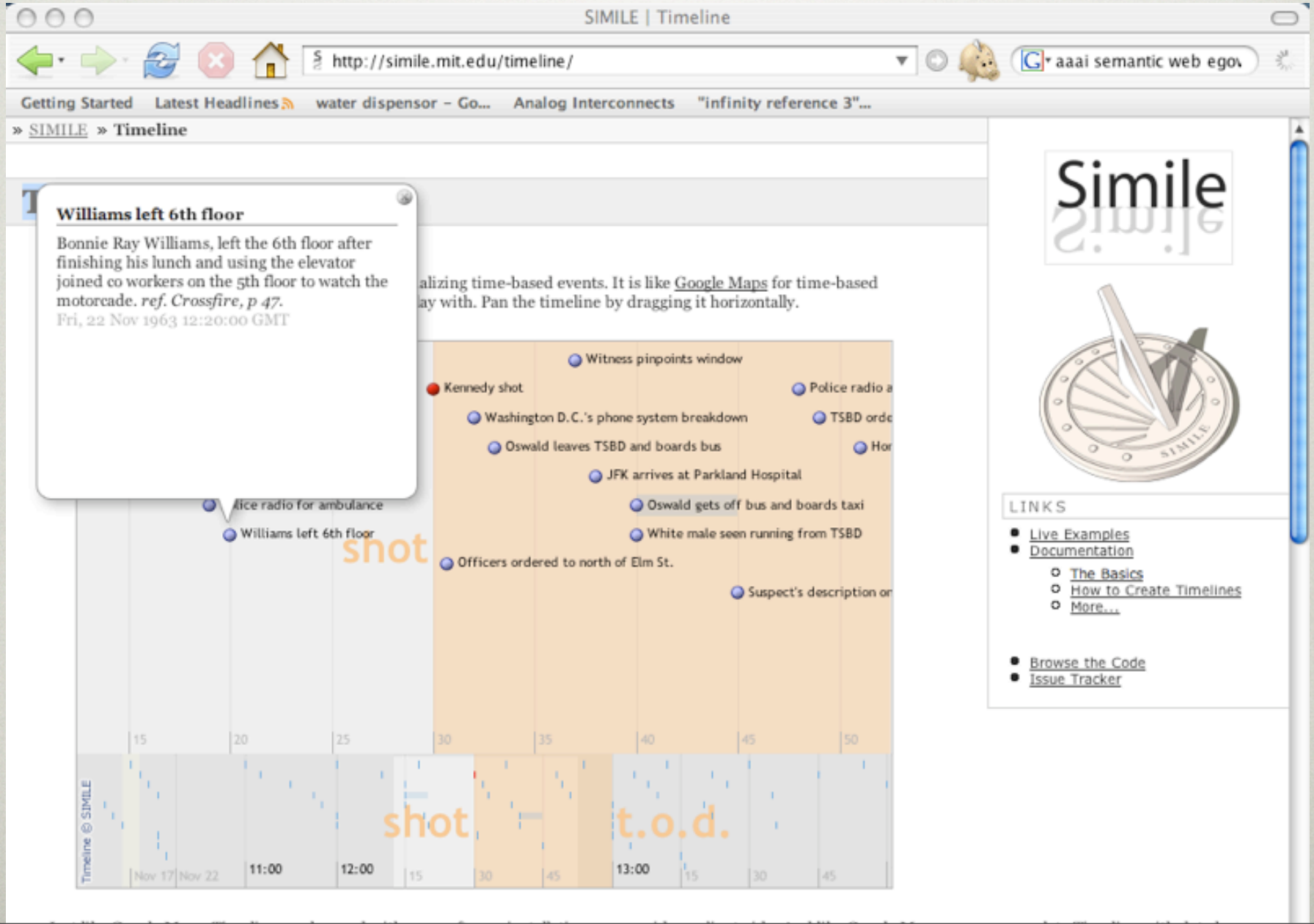
Problem solved



Recombinant Data



But wait! There's more - Timeline

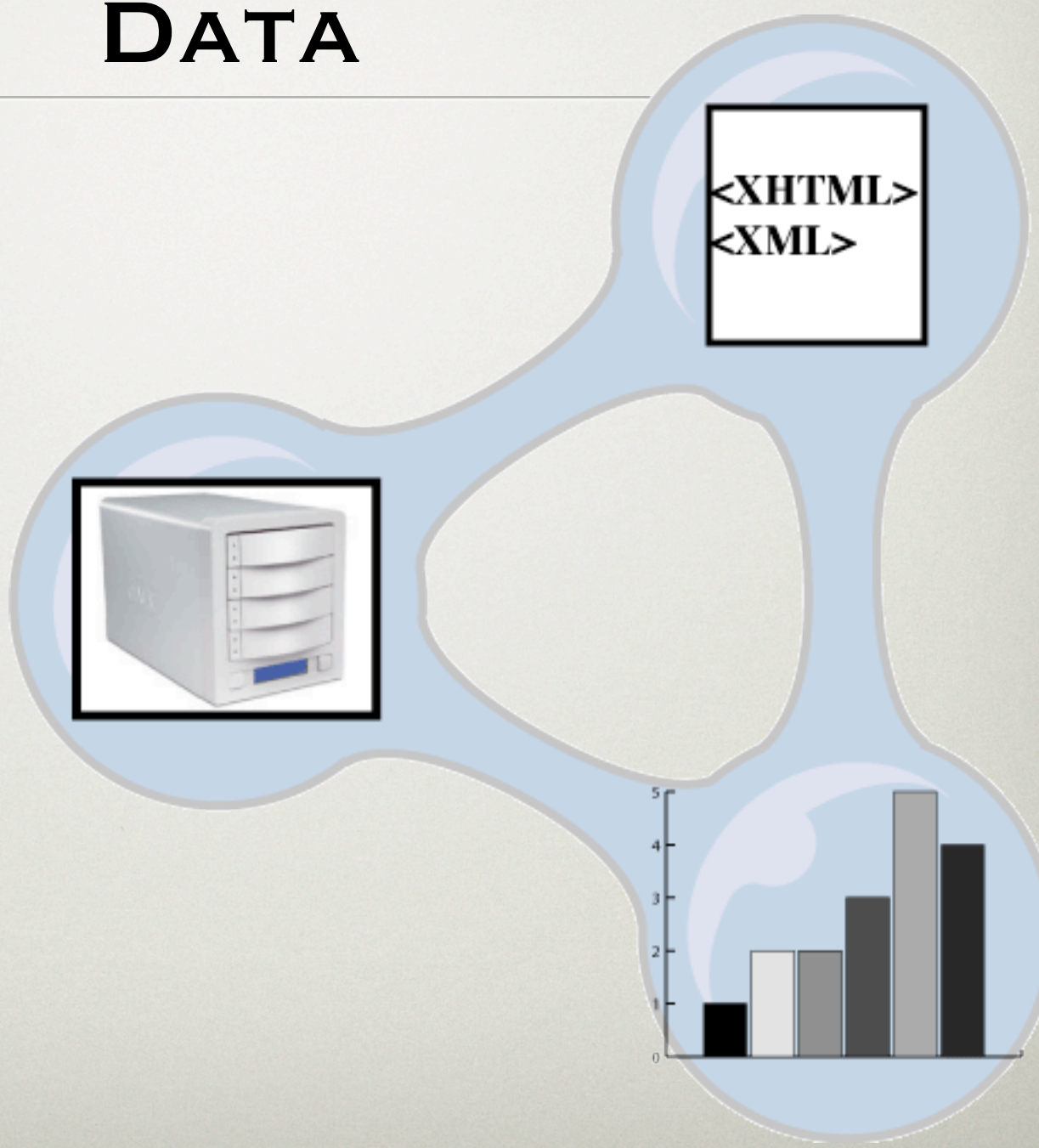


How?

- Web Standards for supporting data integration
- Utilizing tools / services that are based on these standards
  - In this case - MIT Simile Project
- Model real entities, not documents or database tables

Leveraging "Reuseful" Data

- Wrapping, enhancing the existing Web
- Exposing data hidden in documents, servers, applications
- Web evolution not revolution



Enabling Technologies in Place

- RDF, RDF Schema, OWL
- SPARQL
- GRDDL

- General Best Practices / Guidelines

Market Indicators - Products

- Oracle, Adobe, IBM, ...
- (see next door)

Market Indicators - Acquisitions

- Northrup Grumman - Tucana
- WebMethods - Cerebra
- Fortent (Warburg Pincus) - Semagix
- IBM - Unicorn

Market Indicators - VC

- ...
- Vulcan / Leapfrog - Radarnetworks
- ...

Market Indicators - Communities

- Health care and Life Sciences
- eGovernment
- Oil and Gas / Energy
- Geospatial
- Financial Services

The Good

- Generalized solutions
  - common data architecture
  - address common problems of data integration and reuse across many domains
  - serendipitous interoperability

The Bad

- Generality cuts both ways
- Solution for all, solution for none
- Difficulty in understanding in how technology relates to “my needs” now

The Ugly

- Miscommunication
- Misunderstanding
- Fear

Lots of Applications - Open vs Closed

- Enterprise Data Integration
- Community
  - e.g. Healthcare / Life Sciences, eGov, etc.
- (stay tuned to next session!)

Filling the Gap

- Between industry and enabling technologies
- Reduce costs for deployment
  - Social, technical, financial

See a need, Fill a need

- Education, awareness and training
- Mentoring
- “Best of breed” open source tools and applications
  - reduce implementation costs
  - facilitate deployment

Transferring Lessons Learned

- Vodaphone
  - data integration from diverse content providers
- Nokia Forums
  - ontology driven portal environments
- GoPubMed
  - value added, ontology organization

The Venice Project

- Distributing TV shows / video over the web
  - from founders of Skype, Kazaa
- PC as a TV++
  - streaming video
- Peer to Peer
- Using RDF to model TV shows / video

Reducing costs

- Persistent URLs
- Decentralized Data Dictionary
- “Slinked-in” (expertise location)

Common Themes

- Things change, plan for it
- No “right way” / “one way” for describing digital assets
- Importance of “partial understanding”
- Recognized need to free data from application that created it

Common Themes - Continued

- The value of “as needed” data integration
- The power of links, network effect for data
- Big wins come from integrating small ones
- Open world, open solutions are cost effective

eGovernment

Law	Security
Policy	Processes
Regulations	Best Practices
Services	Internationalization

Challenges - Social

- Security
- High granularity access control
- Privacy
- Intellectual property management
- Translating lessons / best practices among organizations

Challenge - Technical

- URIs for entities
- Simple things simple, complex things possible
- Scalable storage
  - “unference”
- Scalable inference

Challenge - Content

- Creation
- Annotation
- Contextualization
- Exposing existing collections of data
  - Relational, LDAP, spreadsheets, scrapers, etc.

Challenge - Marketplace

- Build the marketplace for investing in each others URIs - “coopetition”
- Persistence and usage policies
- 3D - “Decentralized Data Dictionary”

Challenge - Bringing it together

- (insert fancy name here)
  - DOAO - Description of a Organization
  - DOAP - Description of a Project
  - FOAF - Description of a Person
  - SKOS - Taxonomy of subjects

Conclusions

- Semantic Web enables a Web of Data
- Impressive vendor and open source tools
- Market indicators are positive
- Gap between industry and technology is being addressed
- Specific eGovernment challenges
- Opportunities are enormous