

# 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](mailto: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

Local Hotels and Motels

http://www.stanford.edu/dept/hds/chs/general/hotel.html

aaai stanford

Getting Started Latest Headlines World Wide Web C... Danny Ayers The W3C Team Page Questions to the AC MIT Libraries RDF/A

Eric's Meetings Local Hotels and Motels

Community Housing Services  
A division of Residential & Dining Enterprises  
STANFORD UNIVERSITY

Contact Us | Search | FAQs | CHS | R&DE | SU home

## Local Hotels and Motels

Search Rental Listings | Post a Rental Listing | Info for Tenants | Info for Landlords | About the Stanford Area | Hotel List | Off-campus subsidized apts

Please refer to the bottom of this table for a [key to codes and price ranges](#).

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

# REPRESENTATION PROBLEM

The screenshot shows a web browser window titled "Local Hotels and Motels" with the URL `http://www.stanford.edu/dept/hds/chs/general/hotel.htm`. The page content includes a navigation menu, a search bar, and a table of hotels. The table has columns for "Miles from Campus", "Address", "Telephone", "Price Range", and "Codes". The table is divided into sections: "Stanford-Owned-and-Operated" and "Menlo Park".

Miles from Campus	Address	Telephone	Price Range	Codes
<b>Stanford-Owned-and-Operated</b>				
2.2	SLAC Guest House, 2575 Sand Hill Road, Menlo Park	(650) 926-2800	\$ - \$\$	ST
<b>Menlo Park</b>				
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
		(650)		

The browser's developer tools are open at the bottom. The "Console" tab shows JavaScript code for namespace resolution. The "Results" tab shows the DOM tree with the selected table row highlighted.

```
var prefixRDF = 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'
var prefixDC = 'http://purl.org/dc/elements/1.1/'
var prefixLoc = 'http://simile.mit.edu/2005/05/...'

var namespace = doc.documentElement.namespaceURI
var nsResolver = namespace ? function(prefix) {
  if (prefix == 'x') return namespace; else return null;
} : null;
```

# PROBLEM SOLVED

Information collected from file:///Users/em/Library/Application%20Support/Firefox/Profiles/mn4vgfl4.demo/extensions/%7B37...

← → ↻ × 🏠 📖 🐼  🔍

Getting Started Latest Headlines World Wide Web C... Danny Ayers The W3C Team Page Questions to the AC MIT Libraries RDF/A

[My Piggy Bank] | file:///Users/em/Library/Application%20Support/Fir... » Collected Information

## file:///Users/em/Library/Application%20Support/Fir... Collected Information

**1 filter criterion**

- type: Hotel (remove)

Order Commands


List View Calendar View Map View Graph View

**33** items  
sorted by address [A to Z]

1(3) 2(4) 3(3) 4(6) 5(2) 6(2) 7(2) c(2) h(1) m(2) p(1) q(1) s(1) t(2)

« previous 1 2 3 4 next »

- ⌵ type
- ⌵ address
- ⌵ coordinates (map)
- ⌵ title



Creekside Inn	[URI]
address ↗	🔍 3400 El Camino, CA, 94306
coordinates ↗	🔍 37.419463,-122.135429
title ↗	Creekside Inn
type ↗	🔍 <u>Hotel</u>

Show Referers

Save Publish Tags:  Tag

Rickey's Hyatt	[URI]
address ↗	🔍 4219 El, CA,

# BENEFITS OF RDF

Information collected from file:///Users/em/Library/Application%20Support/Firefox/Profiles/mn4vgl4.demo/extensions/%7B37...

← → ↻ × 🏠 📖 🐼  🔍

Getting Started Latest Headlines World Wide Web C... Danny Ayers The W3C Team Page Questions to the AC MIT Libraries RDF/A

[My Piggy Bank] | file:///Users/em/Library/Application%20Support/Fir... » Collected Information

## file:///Users/em/Library/Application%20Support/Fir...

Collected Information

1 filter criterion

- type: Hotel (remove)

Order Commands

List View Calendar View Map View Graph View

Map Satellite Hybrid

1 item(s)

- [Garden Court Hotel](#)

Simile

# MY WEB

The screenshot displays the iCal application interface. The title bar indicates the application is 'iCal' and the time zone is 'US/Eastern'. The main window shows a weekly calendar view for the week of March 26 to April 1, 2006. The calendar grid is populated with various events represented by colored blocks (green, red, purple, orange). A sidebar on the left lists several calendars, each with a checkmark and a refresh icon. Below the calendar list is a monthly calendar for March 2006, with the 23rd highlighted. On the right side, there is a 'To Dos by Priority' sidebar showing a list of tasks with checkboxes and priority indicators. The bottom of the window features a navigation bar with buttons for '+', a grid icon, a download icon, and view toggles for 'Day', 'Week', and 'Month'. A search bar labeled 'Events & To Dos' and several utility icons (list, share, info) are also present at the bottom.



# ANOTHER PROBLEM

---

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

# PROBLEM SOLVED

Information collected from file:///Users/em/Library/Application%20Support/Firefox/Profiles/mn4vgf4.demo/extensions/%7B37...

← → ↻ × 🏠 📖 🐼  🔍

Getting Started Latest Headlines World Wide Web C... Danny Ayers The W3C Team Page Questions to the AC MIT Libraries RDF/A

[My Piggy Bank] | file:///Users/em/Library/Application%20Support/Fir... » Collected Information

## file:///Users/em/Library/Application%20Support/Fir...

Collected Information

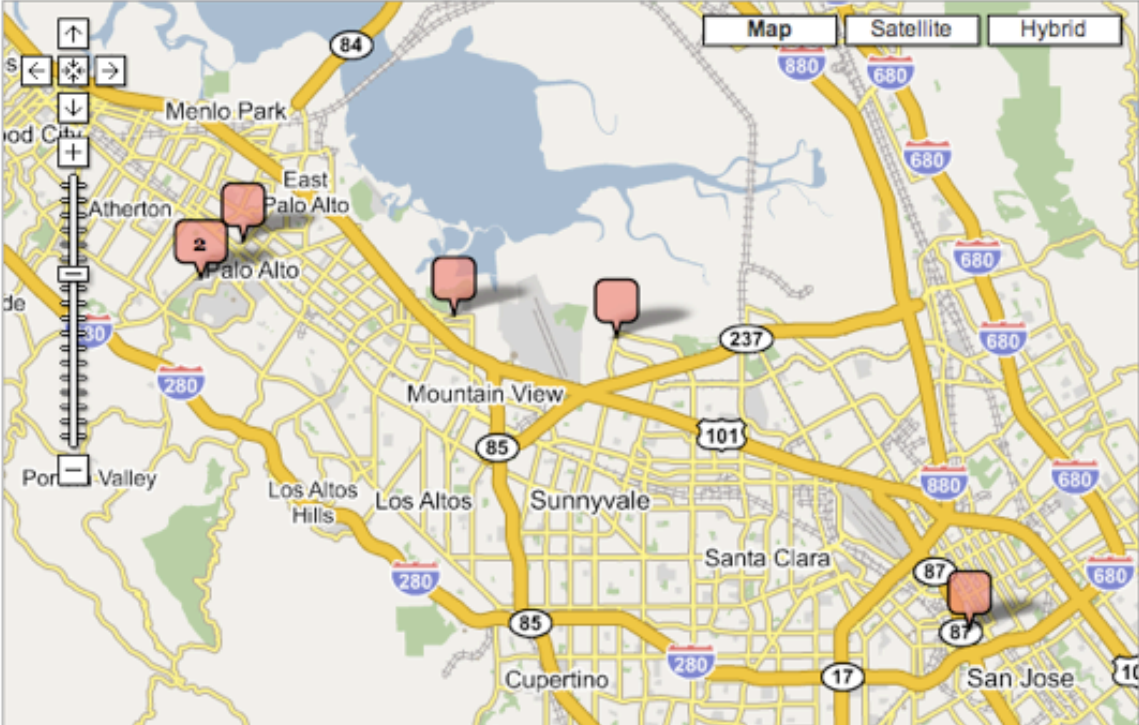
1 filter criterion

- type: Organization (remove)

Order Commands

List View Calendar View Map View Graph View

Map Satellite Hybrid



Simile

# RECOMBINANT DATA

My Piggy Bank

Mar 23, 2006 10:36:22 AM

1 filter criterion

- Tag: "hotels" (remove) "meetings" (remove) [add more]

Order Commands

List View Calendar View Map View Graph View

Map Satellite Hybrid

1 item(s)

- [Garden Court Hotel](#)

type

Type here to filter

Meetings (6)

- type
- tag
- coordinates (map)
- Title
- address

Simile

# BUT WAIT! THERE'S MORE - TIMELINE

The screenshot shows a web browser window titled "SIMILE | Timeline" with the URL "http://simile.mit.edu/timeline/". The browser's address bar and tabs are visible. The main content area displays a timeline interface with a vertical axis on the left and a horizontal axis at the bottom. A tooltip is open over the event "Williams left 6th floor".

**Williams left 6th floor**  
Bonnie Ray Williams, left the 6th floor after finishing his lunch and using the elevator joined co workers on the 5th floor to watch the motorcade. *ref. Crossfire, p 47.*  
Fri, 22 Nov 1963 12:20:00 GMT

Other events on the timeline include: Kennedy shot, Washington D.C.'s phone system breakdown, Oswald leaves TSBD and boards bus, JFK arrives at Parkland Hospital, Oswald gets off bus and boards taxi, White male seen running from TSBD, Officers ordered to north of Elm St., Witness pinpoints window, Police radio a, TSBD orde, Hor, and Suspect's description or.

Large orange text "shot" and "t.o.d." is overlaid on the timeline. The bottom of the timeline shows a date range from Nov 17 to Nov 22, with a time range from 11:00 to 13:00.

**Simile**  
Timeline

**LINKS**

- [Live Examples](#)
- [Documentation](#)
  - [The Basics](#)
  - [How to Create Timelines](#)
  - [More...](#)
- [Browse the Code](#)
- [Issue Tracker](#)

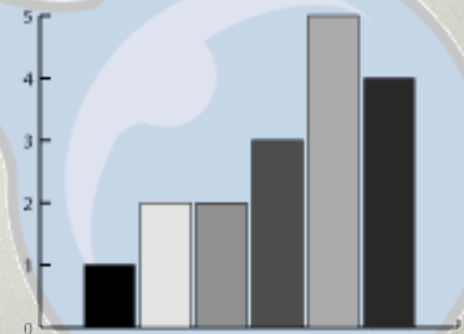
# How?

---

- 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