

Microsoft Research
FacultySummit



Rensselaer

Ontologies for the Real World

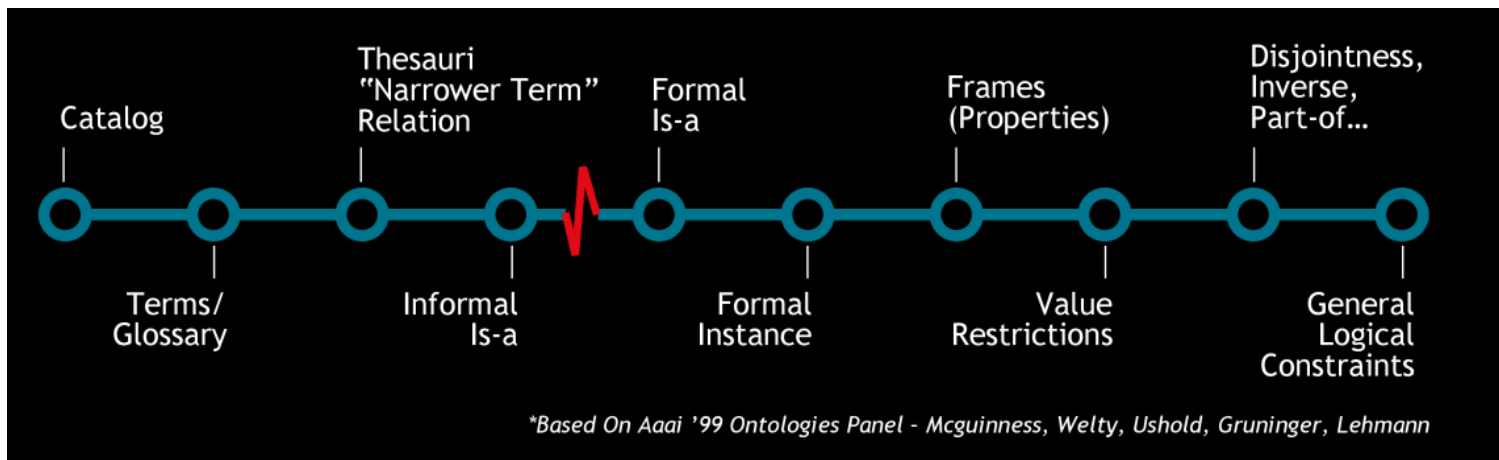
Deborah L. McGuinness

Tetherless World Senior Constellation Chair
Professor of Computer and Cognitive Science
Rensselaer Polytechnic Institute

FUTURE WORLD
2011 | 2031

Ontology Spectrum

- Ontologies continue to take off in many directions and with various styles
Ontology Spectrum from '99



Ontologies Come of Age paper from Dagstuhl Semantics for the Web meeting:

www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-%28with-citation%29.htm

New workshop – Ontologies come of age – this year – co-located with the 10th ISWC <http://ocas.mywikipaper.org/>

Ontology Generation Considerations

- Design Approach
 - Top Down - General notions encoded and specialized (e.g., SUMO, DOLCE, ...)
 - Bottom up – identify noun phrases (concepts) and relations (properties) from (volumes) of data (numerous programs... e.g., DARPA's HPKB, RKF)
 - Application / Use Case driven – identify questions / answers / required reasoning support (more specialized programs ...e.g., many virtual observatories, HALO, etc.)
- Human training
 - Knowledge Representation literate
 - Domain literate
 - Technology approach literate (e.g., machine learning, text analytics, ...)

Introduction through Examples

- Examples of ontologies in practice (2 long lived, one open data-driven)
- Identification of value
- Identification of some aspects that differ today from 10 or more years ago
- Identification of issues worthy of discussion

- Semantic Adviser (Semantic Sommelier)
- Interdisciplinary Virtual Observatory
- Environmental Informatics

Ontologies by Example

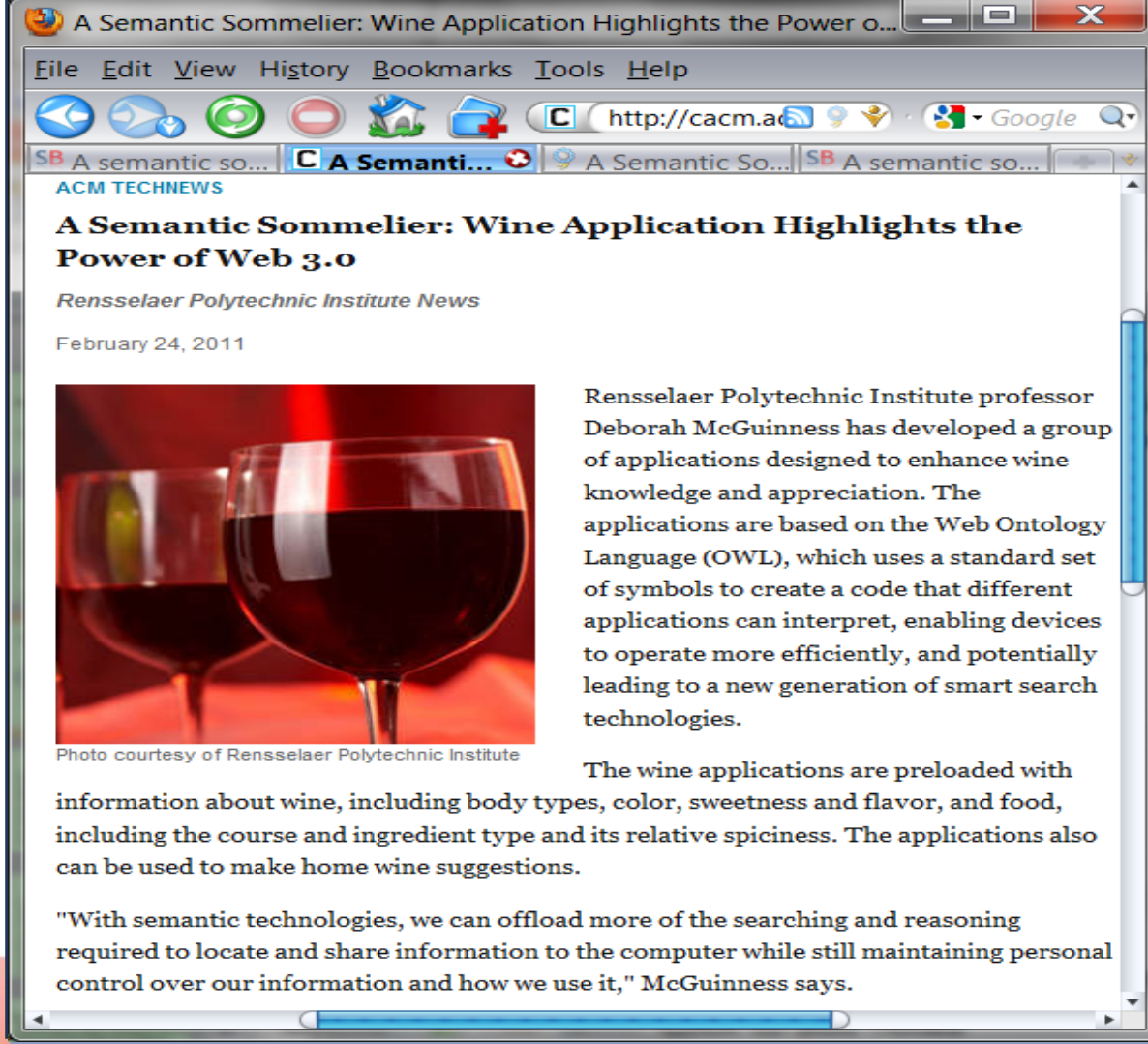
Semantic Agent

Semantically-enabled advisors utilize:

- Ontologies
- Reasoning
- Social
- Mobile
- Provenance
- Context

Patton & McGuinness.

tw.rpi.edu/web/project/Wineagent



A Semantic Sommelier: Wine Application Highlights the Power of Web 3.0

ACM TECHNEWS

A Semantic Sommelier: Wine Application Highlights the Power of Web 3.0

Rensselaer Polytechnic Institute News

February 24, 2011




Photo courtesy of Rensselaer Polytechnic Institute

Rensselaer Polytechnic Institute professor Deborah McGuinness has developed a group of applications designed to enhance wine knowledge and appreciation. The applications are based on the Web Ontology Language (OWL), which uses a standard set of symbols to create a code that different applications can interpret, enabling devices to operate more efficiently, and potentially leading to a new generation of smart search technologies.

The wine applications are preloaded with information about wine, including body types, color, sweetness and flavor, and food, including the course and ingredient type and its relative spiciness. The applications also can be used to make home wine suggestions.

"With semantic technologies, we can offload more of the searching and reasoning required to locate and share information to the computer while still maintaining personal control over our information and how we use it," McGuinness says.

Mobile Wine Agent

- Previous versions used ontologies to infer descriptions of wines for meals and query for wines
- New version uses context, Twitter, Facebook, Wiki, mobile, ...
- Source variability in quality, contradictions exist,
- Maintenance is an issue... however new models emerging

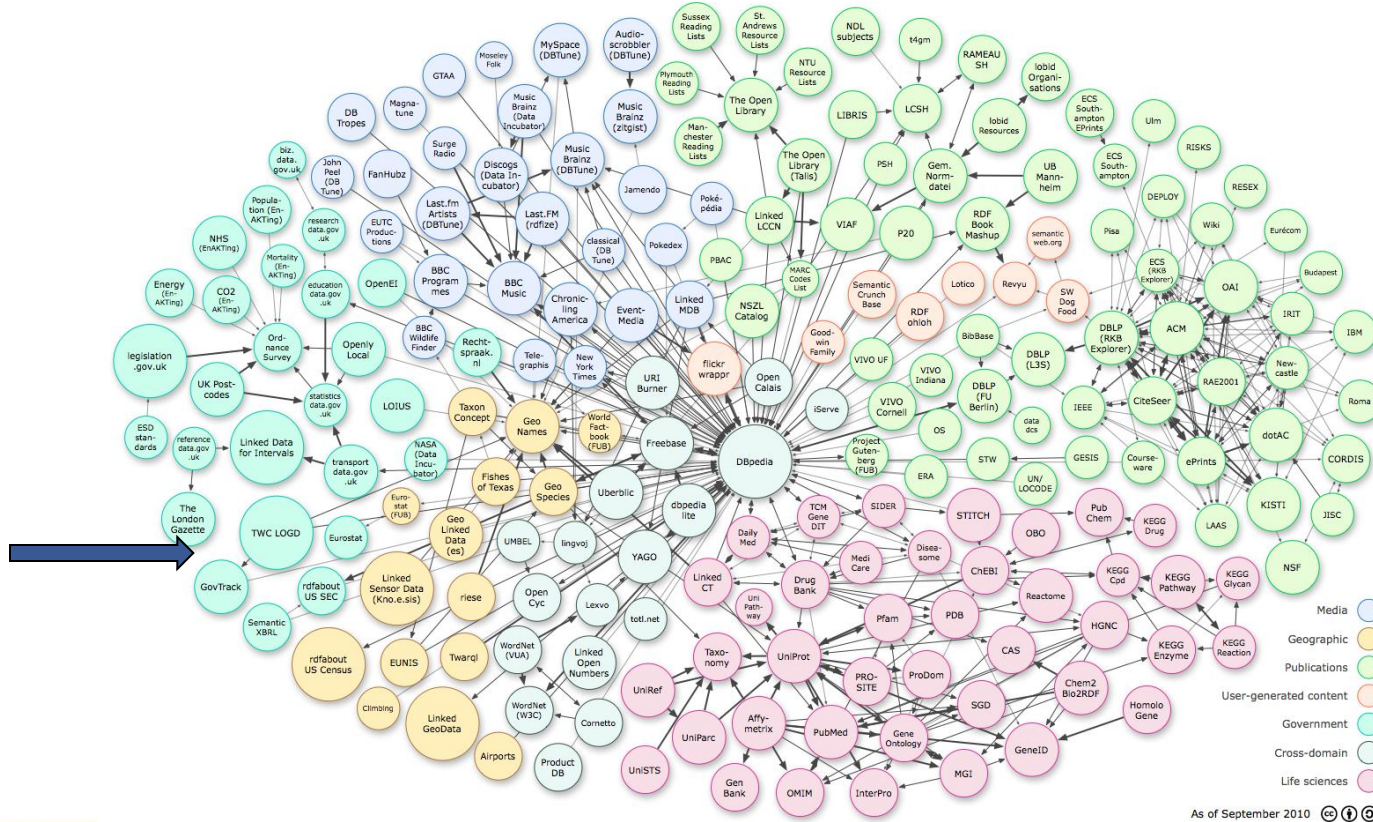





Ontology Ecosystem Discussion & Directions

- Base ontology very simple
 - Wine, Winery, Grape, Flavor, Body, Color, Sugar
 - Stood the test of time: Living with Classic (1991), CLASSIC tutorials, Ontologies 101, OWL Guide, ...
 - To scale however, need to be compatible with WIDE range of menus, wine lists, vocabularies. Not hard to obtain but significant enhancement required.
 - Needs more ecosystem support – explanation, provenance, validation, inconsistency detection, prioritization scheme, UI considerations, additional social connections, citizen-oriented maintenance and evolution schemes, ...

www.ksl.stanford.edu/people/dlm/papers/living-with-classic-abstract.html

Linked Data Cloud



As of September 2010   



Rensselaer

Microsoft Research
FacultySummit

Results 1 to 200 of 12230

next

Dataset Title	Agency	Categories	Catalog	Country
Geochemistry of Water Samples in the US from the NURE-HSSR Database	Department of the Interior	Geography and Environment	Data.gov (United States) Raw Data Catalog	United States
Sydney Futures Exchange (SFE) State Water Indexes.	State Water Corporation	n/a	Data.nsw.gov.au (New South Wales, Australia)	Australia
'WATER DEPTH and Other Data from HUDSON from 19670124 to 19670330 (Accession No. - 7100125)' from 19670201 to 19680404 (Accession No. - 7100127)	US National Oceanographic Data Center	oceans	Data.gov Geodata Catalog	United States
1992 Water-Table Contours of the Mojave River Ground-Water Basin, San Bernardino County, California.	U.S. Geological Survey	inlandWaters	Data.gov Geodata Catalog	United States
Abstractions from non-tidal surface water and groundwater by use: 1971-2007	Department for Environment, Food and Rural Affairs	Environment, Water resources, Water conservation	Data.gov.uk (United Kingdom)	United Kingdom
FOIA Request Log - Water Management	City of Chicago	Government	Data.cityofchicago.org (Chicago, USA)	United States
	Department of		Data.vic.gov.au	

Search Terms:

water

Catalogs:

Deselect All

(11571) [Data.gov --- Geodata Catalog](#)
 (251) [Data.gov.uk \(United Kingdom\)](#)
 (78) [En.openei.org \(Energy Datasets\)](#)
 (54) [Data.govt.nz \(New Zealand\)](#)
 (47) [Data.gov \(United States\) --- Raw Data Catalog](#)
 (44) [Data.gov \(United States\) --- Tool Catalog](#)
 (27) [Env.gov.bc.ca \(British Columbia, Canada\)](#)
 (22) [Data.gov.au \(Australia\)](#)

Countries:

Deselect All

(11707) United States
 (254) United Kingdom
 (81) International organization
 (64) Canada
 (54) New Zealand
 (40) Australia
 (16) Ireland
 (8) Kenya

Agencies:

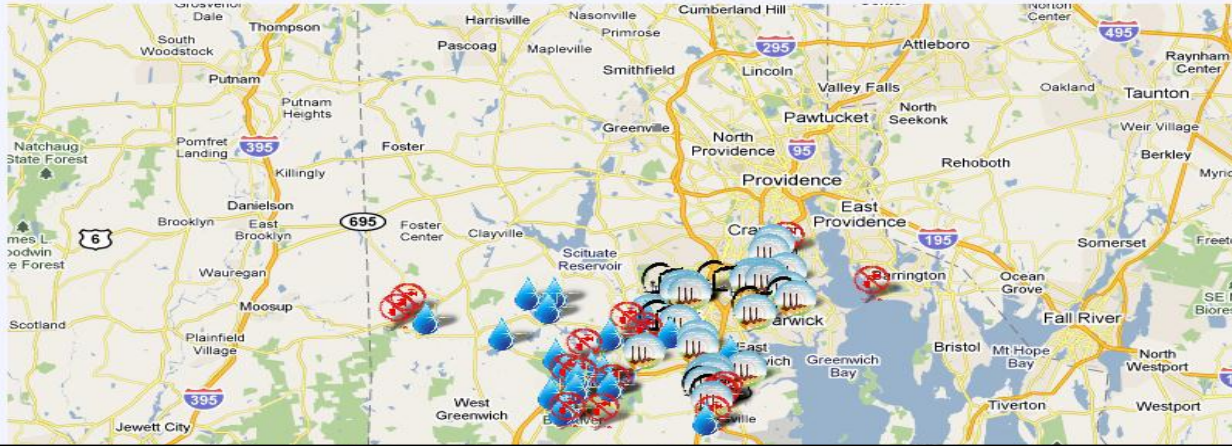
Deselect All

(11231) US National Oceanographic Data Center
 (193) U.S. Geological Survey
 (54) Department for Environment, Food and Rural Affairs
 (35) NOAA National Oceanographic Data Center
 (34) Ministry for the Environment (MfE)
 (30) NREL

TWC Semantic Water Quality Portal

Showing Data for Triples From: 1 To 5000 Clear Map

Zip Code:



Data Source

- USGS
- EPA

Regulation

- EPA Regulation
- MASS Regulation
- CA Regulation
- RI Regulation
- NY Regulation

Data Type

- Facility
- Polluting Facility
- Polluted Water
- Clean Water

Characteristic

- No Filter

Aimed at helping people investigate local water quality

- Diverse datasets, regulations, datatypes

- Uses lightweight semantic technologies to produce mashups that make data accessible that would be otherwise difficult to view in perspective

- Maintains provenance about data and manipulations

- Potential to empower citizens with contextualized data and support citizen scientist questions and reporting

Microsoft Research

FacultySummit

Provenance and Water Pollution in SWQP

Showing Data for Triples From: 1 To 5000 Clear Map

Zip Code:

pH	5.82 SU?	>=	6.5 SU?	2009-09-30
pH	6.49 SU?	>=	6.5 SU?	2009-12-31
pH	6.46 SU?	>=	6.5 SU?	2010-03-31
pH	6.11 SU?	>=	6.5 SU?	2010-06-30
pH	10.87 SU?	<=	9 SU?	2010-09-30
Oil_Grease	16 MG/L?	<=	15 MG/L?	2008-09-30
pH	10 SU?	<=	9 SU?	2007-12-31
pH	10.54 SU?	<=	9 SU?	2008-03-31
pH	9.88 SU?	>=	9 SU?	2008-06-30
pH	9.86 SU?	>=	9 SU?	2008-12-31

[Visualize contaminants](#)

next 5000 triples

About | Privacy | Contact

Data Source

USGS

EPA

Regulation

EPA Regulation

MASS Regulation

CA Regulation

RI Regulation

NY Regulation

Data Type

Facility

Polluting Facility

Polluted Water

Clean Water

Characteristic

No Filter

[select](#)

Ping Wang, Jin Zhen, Evan Patton, Tim Lebo, Joanne Luciano, Deborah L. McGuinness. tw.rpi.edu/web/project/SemantAQUA

Discussion and Directions

- Base ontology also very simple - Water, contaminant, threshold, test
- Simple use of recognition and easily extensible (e.g., recently with health impact data)
- To scale however, needs to be compatible with wide range of data source vocabularies, including a wide range of tests
- New processes create new vocabulary needs (e.g., protectingourwaters.wordpress.com/2011/06/16/black-water-and-brazenness-gas-drilling-disrupts-lives-endangers-health-in-bradford-county-pa/)
- Needs more ecosystem support – explanation, provenance, validation, inconsistency detection, prioritization scheme, UI considerations, additional social connections, citizen-oriented maintenance and evolution schemes, ...

Interdisciplinary Virtual Observatory (VSTO)

File Edit View Go Window Help

Bookmarks Location: <http://cedar-1.hao.ucar.edu:80/cgi-bin/ion-p?page=mv>

CEDARweb

UCAR CEDAR SEARC

Login

Home
Community
Workshop
Data Services
Tools
Documents
Help
Download
Contact Us

Copyright 2003, UCAR
Approved by Peter Fox

Coupling, Energetics and Dynamics of Atmospheric Regions (CEDAR) is a focused *Global Change program* sponsored by the *National Science Foundation (NSF)*. The scientific objectives of the program are described in the [CEDAR Phase III document](#) (1.8MB, acrobat reader required).

The CEDAR Data System (formerly the CEDAR Database and before that, the Incoherent Scatter Radar Database) is a cooperative project between the *High Altitude Observatory (HAO)* division of the *National Center for Atmospheric Research (NCAR)*, the *National Science Foundation (NSF)*, and numerous institutions that provide upper atmosphere data and model output for community use.

The CEDAR Data System mission is to provide:

- long term archive for observations and models of the Earth's upper atmosphere and geophysical indices and parameters needed to interpret them,
- browsing capability to survey the data holdings and identify periods, instruments, models, of interest,
- reliable data access methods that are fast, stable and interactive, and
- detailed documentation on data acquisition and reduction.

Proceed to data selection now

This site also supports the CEDAR community which is represented by the [CEDAR Science Steering Committee](#) consisting of representatives from the community and NSF and meets twice a year. Read more about the community in the latest [CEDAR FOOT \(1st issue\)](#).
-- Revised 31 October 2003 by pfox@ucar.edu

● General: Find data subject to certain constraints and plot appropriately

File Edit View Go Window Help

Bookmarks Location: http://mlso.hao.ucar.edu/cgi-bin/mlso_homepage.cgi

Mauna Loa Solar Observatory HA

Welcome to the Mauna Loa Observatory (MLSO) Website. The MLSO, operated by the High Altitude Observatory in Boulder Colorado, houses several instruments designed to observe the sun at many different wavelengths.

ACOS Advanced Coronagraph Observing System. A suite of instruments designed to observe the solar atmosphere at a variety of heights. Includes Chromospheric Helium Imaging Photometer (CHIP, 1083.0nm), H-alpha prominance and solar disk monitor (PICS, 656.2nm), and the Mk4 K-coronameter, which observes the white light K-corona from 1.12-2.70 solar radii.

ECHO Euler Chromospheric Helium Imaging Photometer. A network of two instruments which observe solar oscillations as seen in the radial velocity of the solar surface.

PSPT Precision Solar Photometric Telescope. Observes the solar disk in three bandpasses: 605-610 nm (red), 408-412 nm (blue), and 393 nm (CaIIK).

ACOS | ECHO | PSPT | Hawaii W3 | Related Sites | Contact Us
Eclipses | Instruments | Publications | About MLSO

Latest MLSO Images

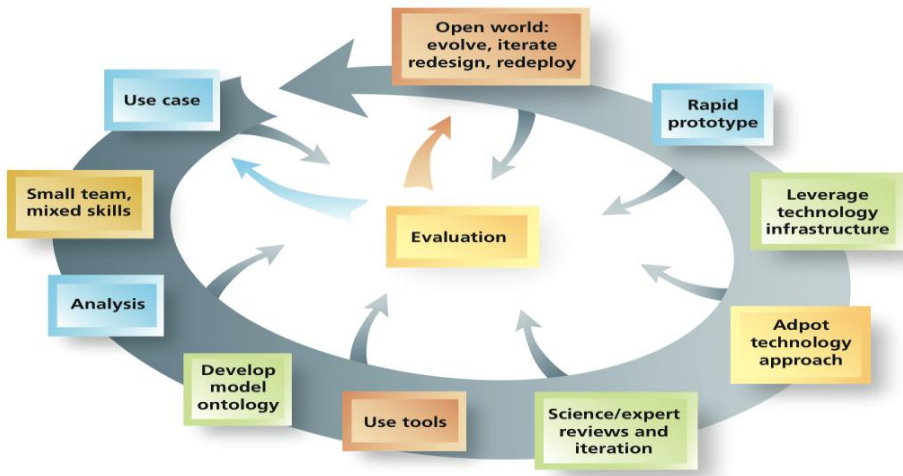
K-Corona 700-950 nm 27-Nov-2003 20:59 Movie [Interped-GIF]	H-Alpha Limb 656.3 nm 28-Nov-2003 21:11 Movie [Interped-GIF]	H-Alpha Disk 656.3 nm 28-Nov-2003 21:04 Movie [Interped-GIF]	Helium-I 1083 nm 28-Nov-2003 21:20 Movie [Interped-GIF]
PSPT CaIIK	PSPT Blue	PSPT Red	ECHO Sample Velocity Image

● Specific: Plot the observed/measured Neutral Temperature as recorded by the Millstone Hill Fabry-Perot interferometer while looking in the vertical direction at any time of high geomagnetic activity in a way that makes sense for the data.

Semantic Enhancements

Semantic Web Methodology & Technology Development Process

- ▶ Establish and improve a well-defined methodology vision for semantic technology based on application development
- ▶ Leverage controlled vocabularies, etc.



McGuinness, Fox, West, Garcia, Cinquini, Benedict, Middleton <http://www.vsto.org>

Ontology Ecosystem Discussion & Directions

- Base ontology relatively simple - Instrument, Observatory, Data Product, ...
- Initially done for solar terrestrial physics but has been used in volcanology, plate tectonics, sea ice, water, ... with relatively little rework (NSF: VSTO, SPCDIS, SESF, SSI, SONET ... NASA: SESDI, ...)
- Modularity has been key – both to reusing other ontologies (e.g., SWEET) and in expanding our reuse
- To scale and be maintainable however, need to be compatible with WIDE range of evolving vocabularies. (Unlike the wine agent and to some extent the water quality portal, this is not as uncomplicated, e.g., ESIP discussions last week)
- Needs more ecosystem support – explanation, provenance, validation, inconsistency detection, prioritization scheme, UI considerations, additional social connections, citizen-oriented maintenance and evolution schemes, ...

What is different now (10+ years later)?

- Ontologies (at many points on an expressiveness spectrum) are in use in wide variety of settings and disciplines and are built by a broad range of users
- Recommended Web Ontology Language (and business consequences), Rules recommendation, Provenance on its way
- Issues are much less about starting points for ontologies – they are now about mapping, reusability, maintenance, and sustainability
- Issues are not only technical – social issues of team acquisition and maintenance may be at least as important

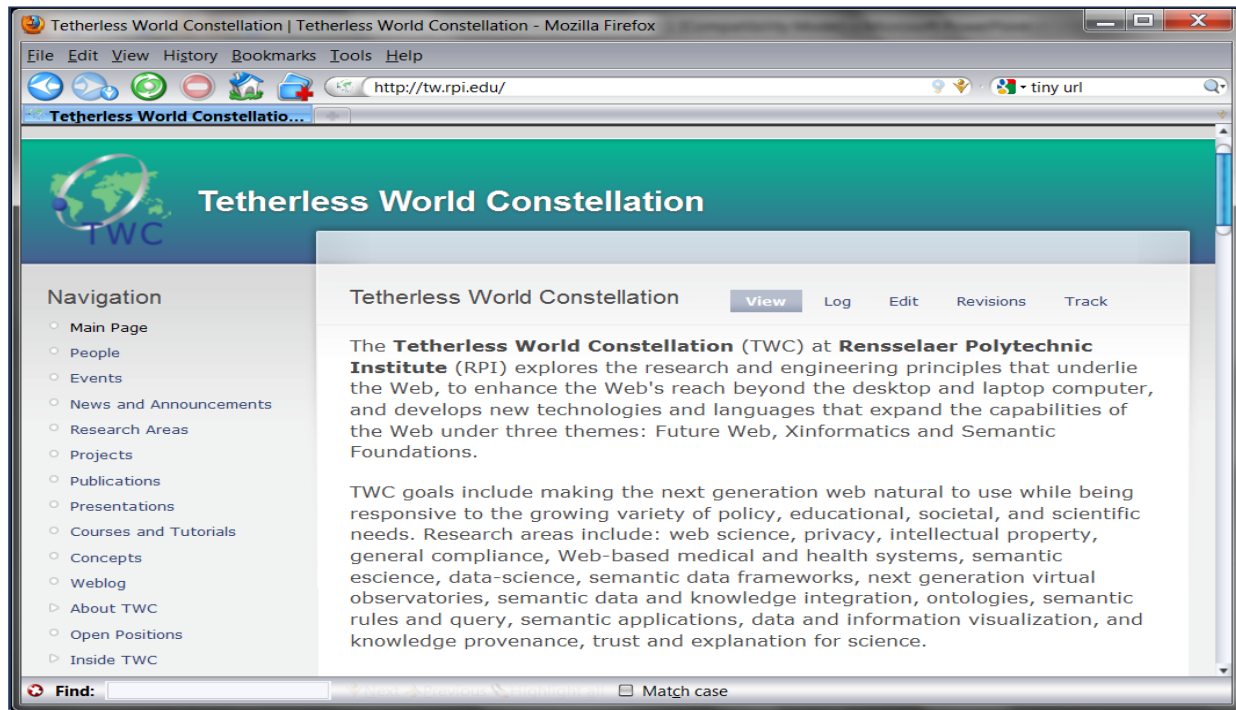
What directions might we consider?

- Guidelines for creating ontologies for reuse – modularity, limited conflict generators, ease of use considerations
- Provenance - Representation (e.g., W3C working group), Watermarking , ...
- Semi-automatic tools for ontology creation and maintenance
 - Checking
 - Expanding
 - Mapping
- Hybrid tools for working with learning / discovery tools AND humans – e.g., picking up on Mundie's example of readmission – congestive heart failure & gastric issues and/or depression
- Directions for examples such as Watson-style work
- What do you need for ontologies to be practically and sustainably used in commodity computing? - forthcoming 4th paradigm blog post

Questions? dlm @ cs . rpi . edu

Extra

Tetherless World Constellation - RPI



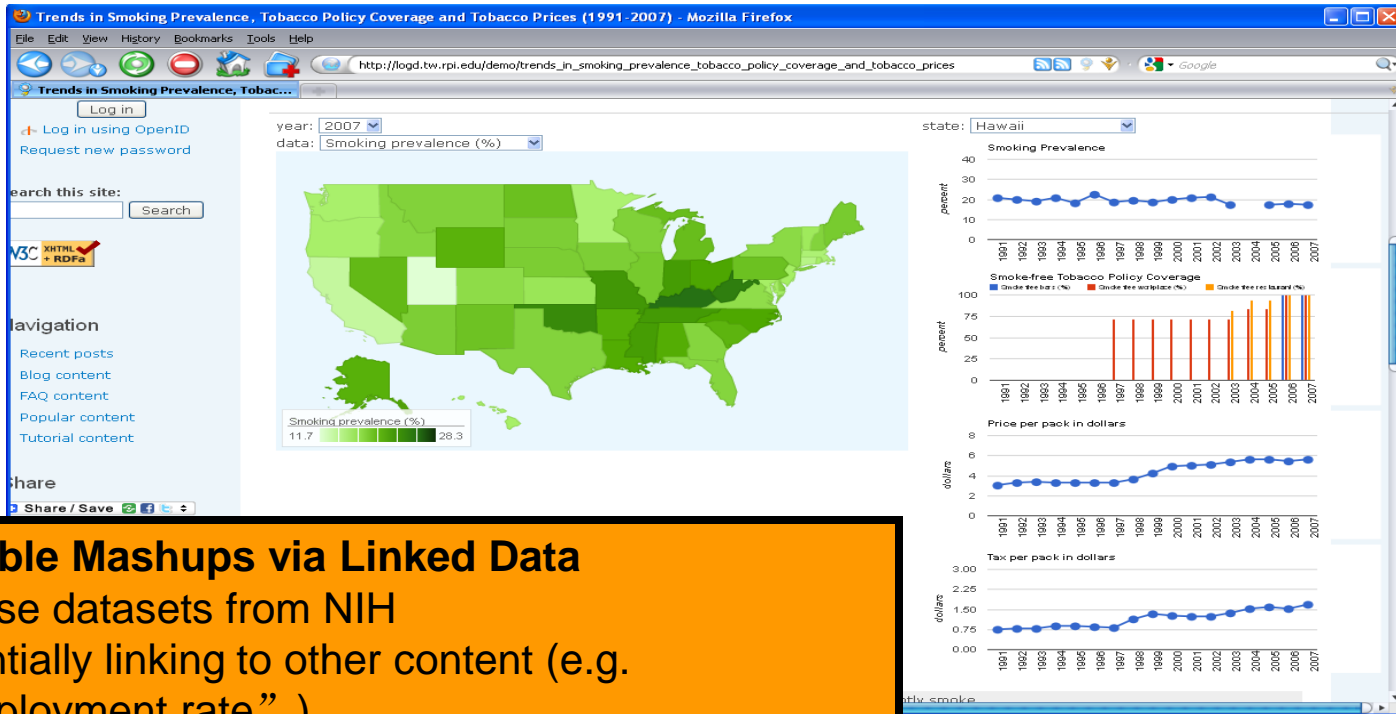
Themes:

- Semantic Foundations
 - Knowledge Provenance
 - Ontology Environments
 - Inference
 - Trust
 - Linked Data
- Xinformatics
 - Semantic eScience
 - Data Science
 - eHealth
 - eEnvironment
- Future Web
 - Web Science
 - Policy
 - Social

Chaired Professors: McGuinness, Fox, Hendler

Research Assoc. Professor: Luciano <http://tw.rpi.edu>

PopSciGrid Example State -Hawaii



Extensible Mashups via Linked Data

- ❑ Diverse datasets from NIH
- ❑ Potentially linking to other content (e.g. “unemployment rate”)

Accountable Mashups via Provenance

- ❑ Annotate datasets used in demos

PopSciGrid in Action

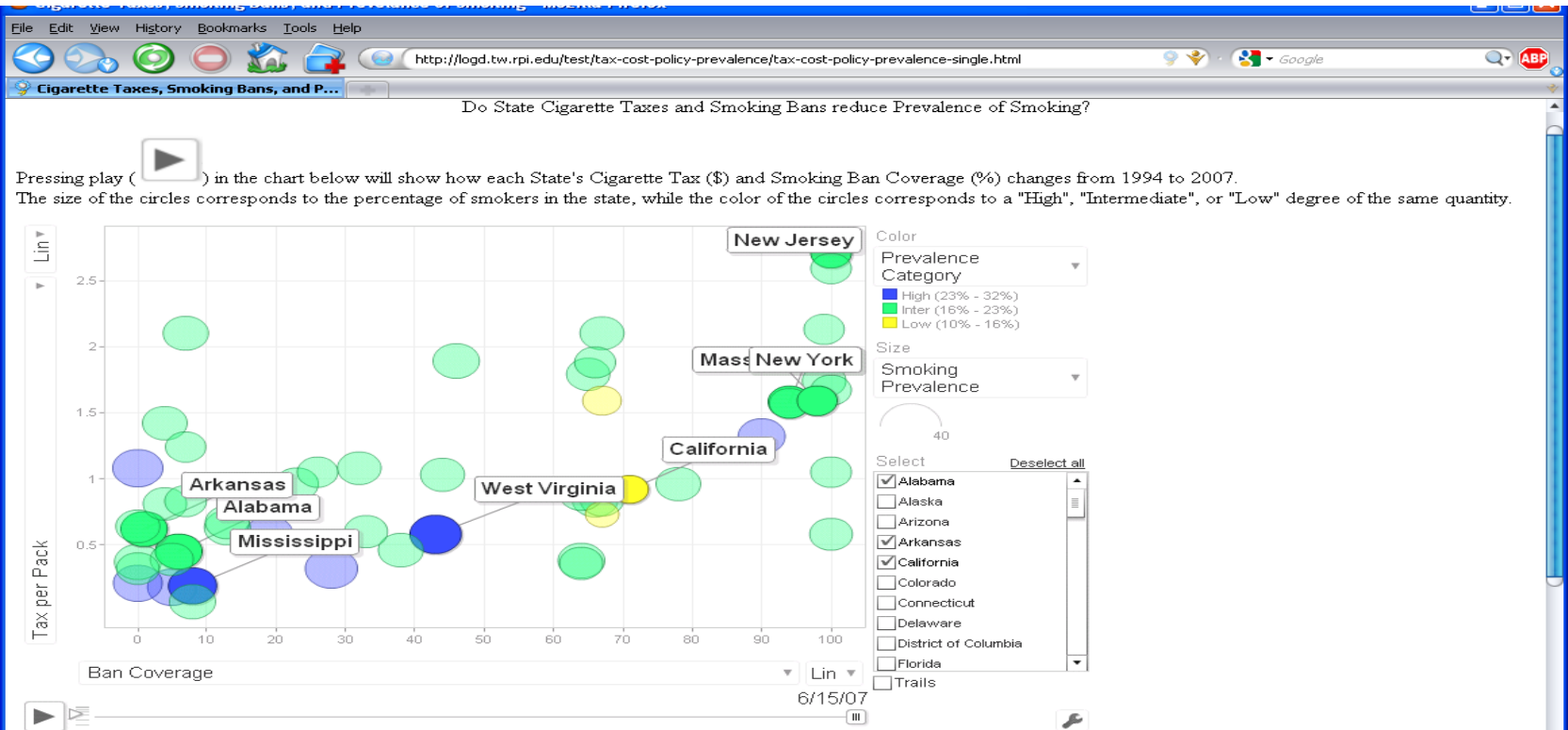


By 2007, we see that the (blue) "High" Smoking Prevalence states tend toward the low tax and low ban corner of the graph.

We also see that the (yellow) "Low" Smoking Prevalence states are in the middle of the tax range and do not have the highest ban coverage.

The following two timelines show the Tax and Cost per Pack (in US Dollars) and the Smoking Prevalence and Degree of Smoking Ban Coverage (in percentage) for the state of Alabama. To view these measures for a different state, select it here: [Alabama](#)

PopSciGrid II



By 2007, we see that the (blue) "High" Smoking Prevalence states tend toward the low tax and low ban corner of the graph. We also see that the (yellow) "Low" Smoking Prevalence states are in the middle of the tax range and do not have the highest ban coverage.

The following two timelines show the Tax and Cost per Pack (in US Dollars) and the Smoking Prevalence and Degree of Smoking Ban Coverage (in percentage) for the state of Alabama. To view these measures for a different state, select it here:

Provenance and Water Pollution in SWQP

Showing Data for Triples From: 1 To 5000 Clear Map

Zip Code:

Water Regulation Provenance: <http://tw2.tw.rpi.edu/zhengj3/owl/EPA-regulation.pml>
characteristic: Arsenic
limit: 10.0
From:
RDF Source: <http://tw2.tw.rpi.edu/zhengj3/owl/epa.owl#ExcessiveArsenicugMeasurement>
CSV Source: <http://tw2.tw.rpi.edu/zhengj3/owl/EPA-regulation.csv>
Original Source: <http://water.epa.gov/drink/contaminants/index.cfm#List>





Data Source

- USGS
- EPA

Regulation

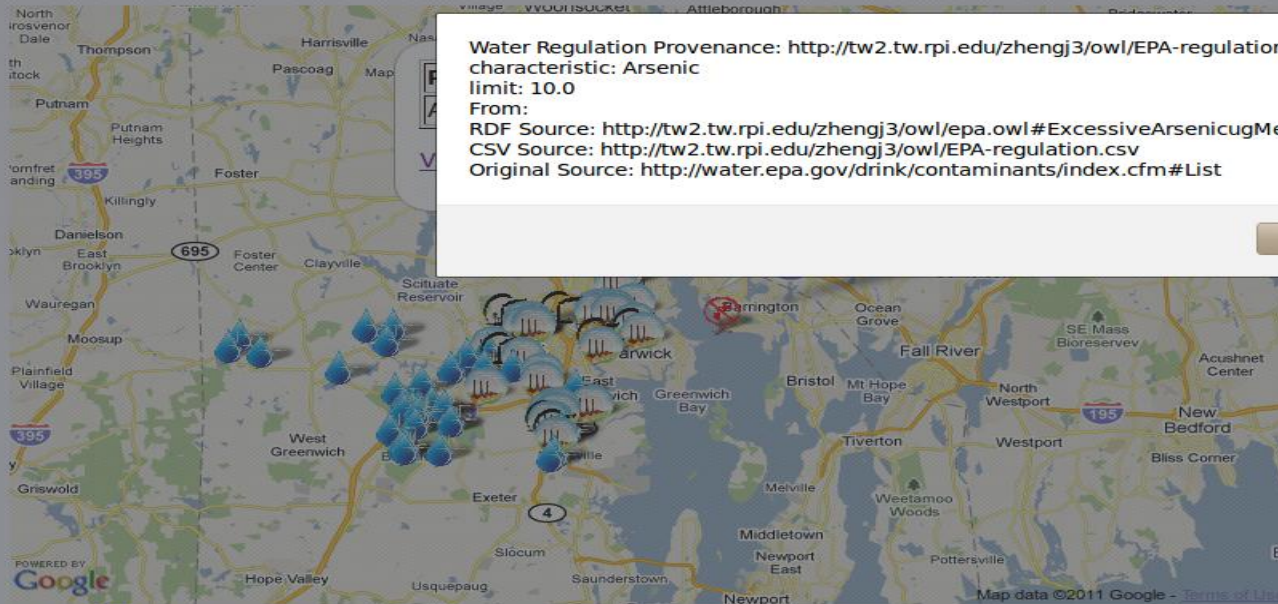
- EPA Regulation
- MASS Regulation
- CA Regulation
- RI Regulation
- NY Regulation

Data Type

-  Facility
-  Polluting Facility
-  Polluted Water
-  Clean Water

Characteristic

- No Filter



next 5000 triples