

# The PopSciGrid Portal: A Tool for Visualization and Exploration of Data

Abdul R Shaikh, PhD, MHSc

Program Director

Health Communication and Informatics Research Branch

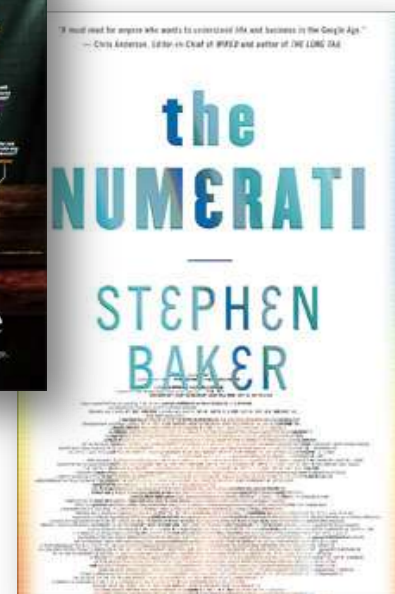
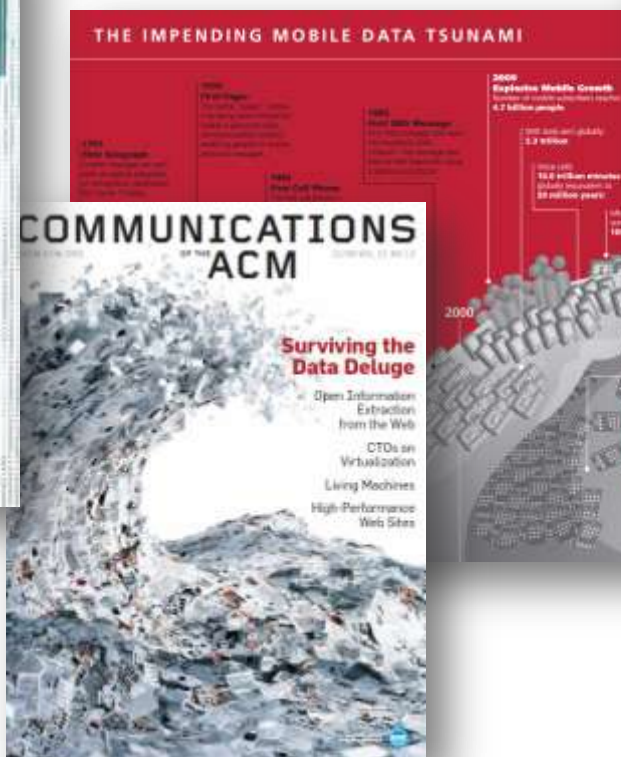
Division of Cancer Control and Population Sciences

National Cancer Institute

[shaikhab@mail.nih.gov](mailto:shaikhab@mail.nih.gov)

# In this age of 'Big Data'...

*...a grand experiment is unfolding on the utilization of data for public health impact.*



# DCCPS Cancer Control Framework

## Cancer Control Research Activities



***Reducing the cancer burden***

# Population Science

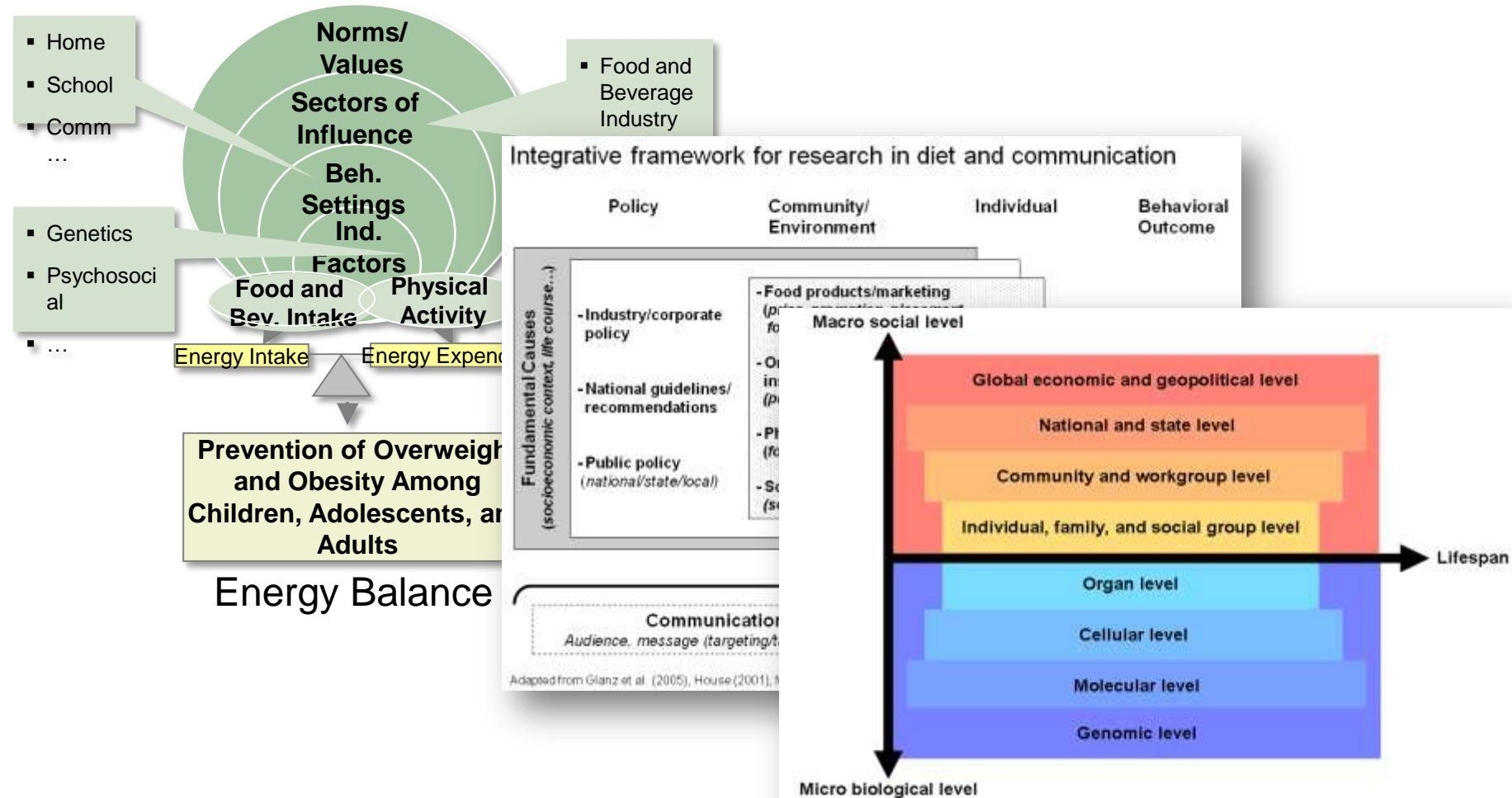


Figure 1. Health as a continuum between biological and social factors across the lifespan. (Adapted from Glass & McAtee, 2006).

# Recent Developments in the U.S. Federal HIT Landscape

**ARRA & Health Information Technology for Economic and Clinical Health (HITECH) Act** - *“ties payments specifically to the achievement of advances in health care processes and outcomes.”*  
(Blumenthal, 2010)

**Meaningful Use of EHRs:** patient-focused care, quality, efficiency, research, public health

**Healthcare Reform**

**CER and learning healthcare systems**

**Open Government Directive and CHDI:** Linking and harmonizing data, sharing and analyzing information = ***Democratizing data for impact***

# Phase I Proof of Concept: PopSciGrid (2008)

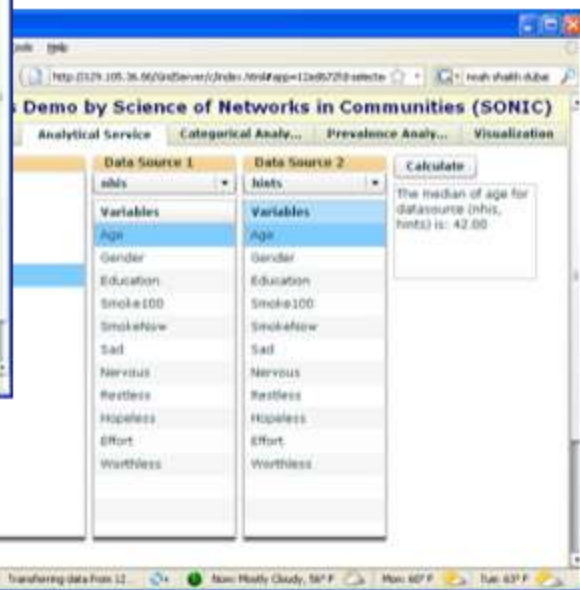
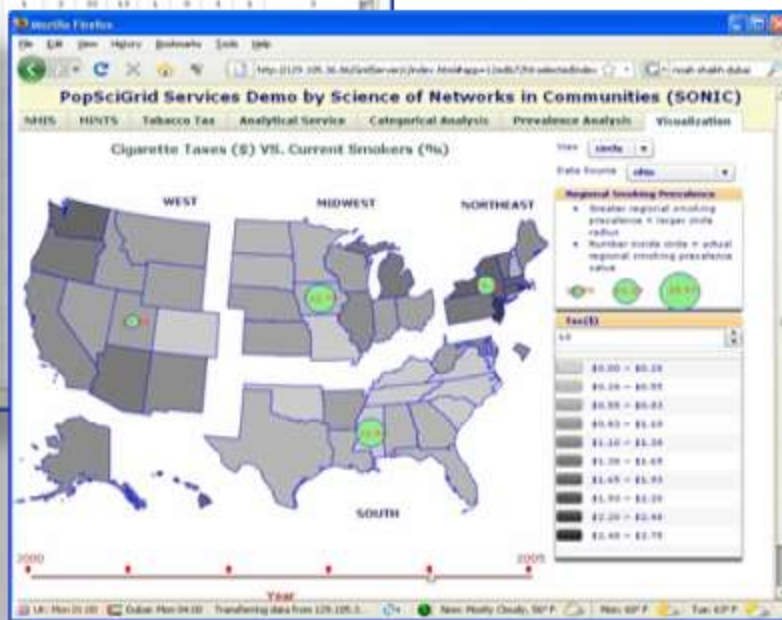
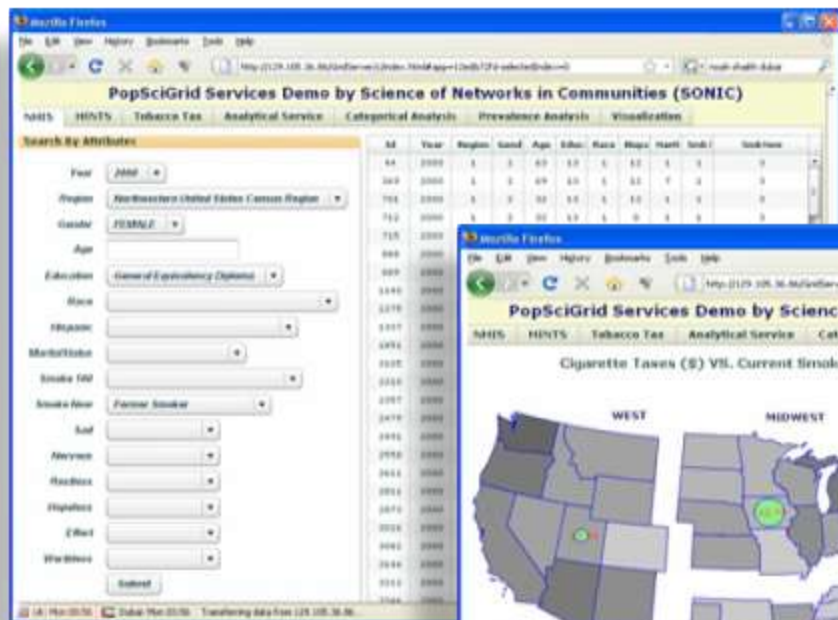
- Access & utilization of population science data
- Real-time integration and analysis of multiple types of data (e.g., population health, economic, geo-spatial)
- Decrease the time it takes to translate research into practice and policy at local and state levels

Data: *National Health Interview Survey (NHIS):  
Health Information National Trends Survey (HINTS)  
State-level tobacco tax data 2000-2007  
(Orzechowski & Walker, 2007)*

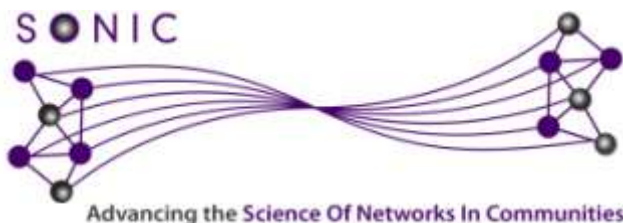


# PopSciGrid Proof of Concept

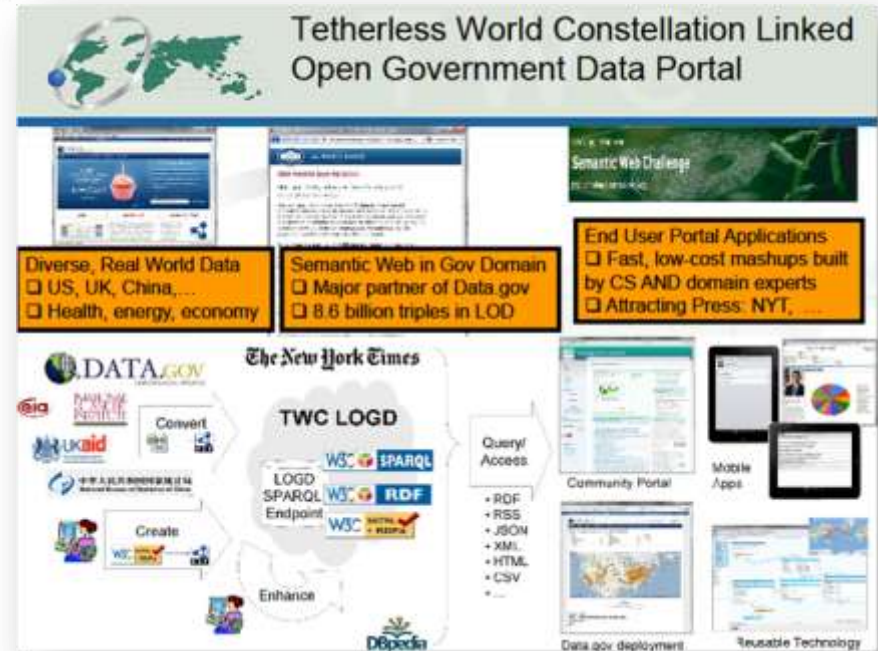
- 14 datasets spanning 6 years
- Real-time access/analysis of public health and economic data



- Prospective geo-spatial analytics
- Dr. Noshir Contractor



# Phase II: PopSciGrid Community Health Data Portal



<http://logd.tw.rpi.edu/project/popscigrd>

<http://cancercontrol.cancer.gov/hcirb/cyberinfrastructure/popsci.html>



# PopSciGrid

## Community Health Data Portal

- Builds on the original PopSciGrid proof of concept
  - Data sharing, collaboration, communication
  - Convey complex information to consumers and decision-makers
  - Data for community impact
  - Inform development of future research opportunities: communicating health info online, utility of conceptual frameworks, 'apomediaries' for information/services, semantic interoperability, data visualization, integration/analysis of multi-level data...

# PopSciGrid

## Community Health Data Portal

### Data

**Smoking prevalence:** Percentage US adults who currently smoke (BRFSS 1991-2007) \*\*

**Smoke free bars (%):** Percentage of population affected by the smoke-free policy for bars \*

**Smoke free workplace (%):** Percentage of population affected by the smoke-free policy for workplaces \*

**Smoke free restaurant (%):** Percent of population affected by the smoke-free policy for restaurants \*

**Price per pack in dollars (\$):** Average price per pack of cigarettes adjusted for inflation (1991 dollars) \*\*

**Tax per pack in dollars (\$):** State excise tax per pack of cigarettes adjusted for inflation (1991 dollars) \*\*

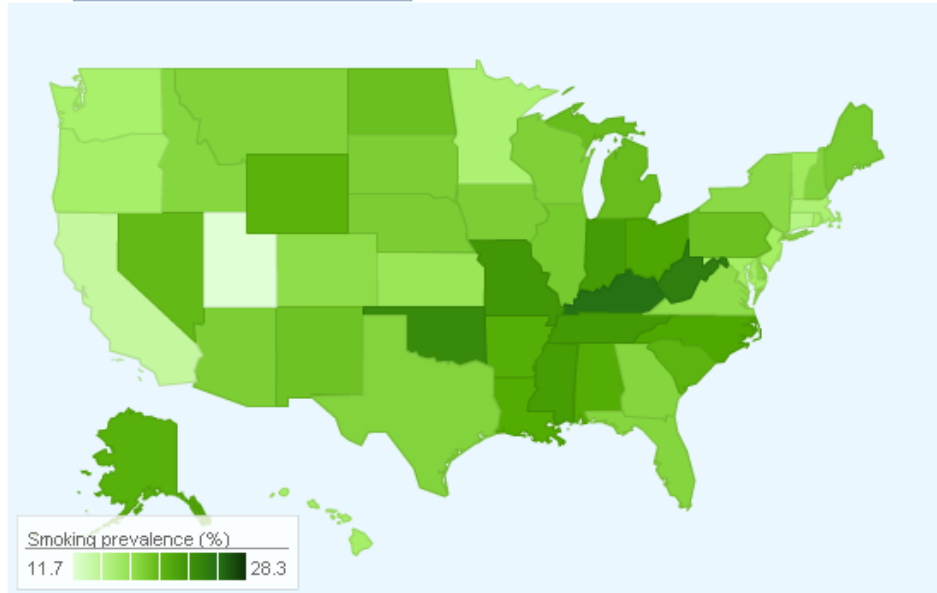
\* Tatalovich, Hunt, Howlander, Marcus & Stinchcomb. "Reconstructing the spatio-temporal patterns of tobacco policy coverage in the United States" (forthcoming)

\*\* <http://www.impactteen.org/tobaccodata.htm>

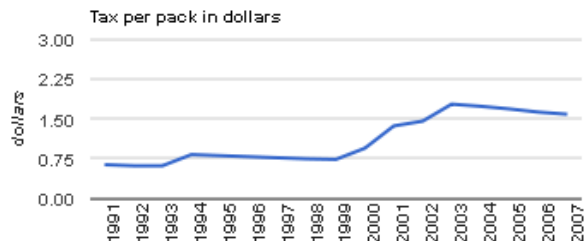
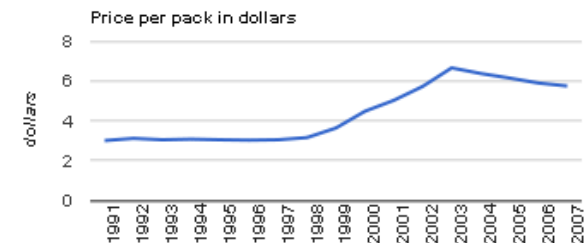
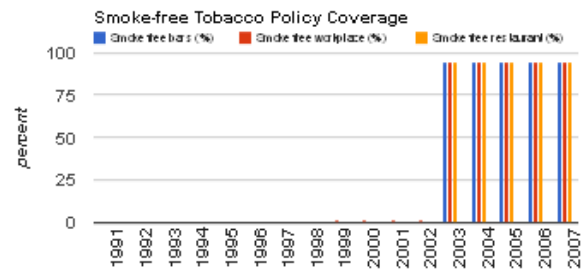
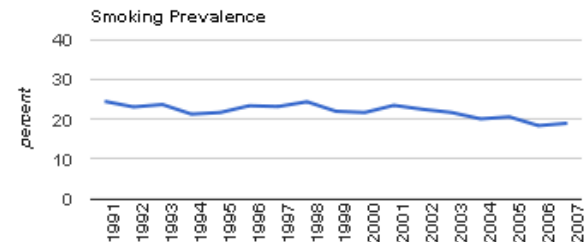
# Trends in Smoking Prevalence, Tobacco Policy Coverage and Tobacco Prices (1991-2007)

year: 2007

data: Smoking prevalence (%)



state: New York

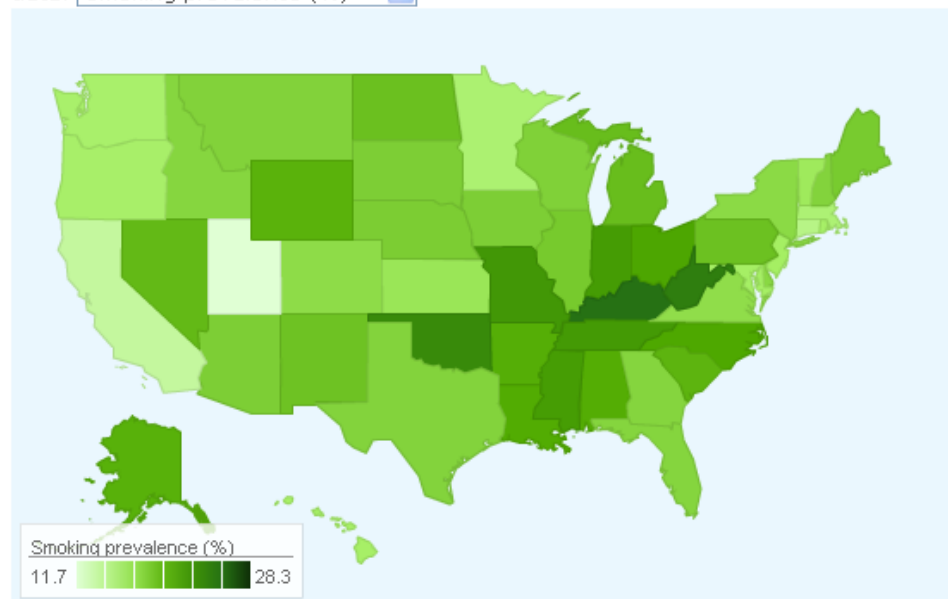


<http://cancercontrol.cancer.gov/hcirb/cyberinfrastructure/popsoci.html>

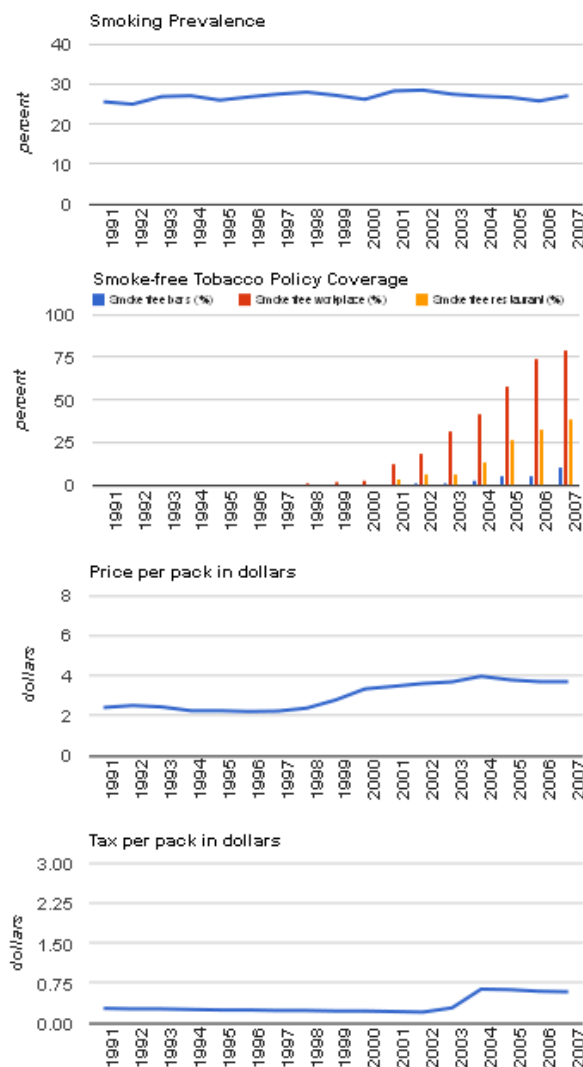
# Trends in Smoking Prevalence, Tobacco Policy Coverage and Tobacco Prices (1991-2007)

year: 2007

data: Smoking prevalence (%)



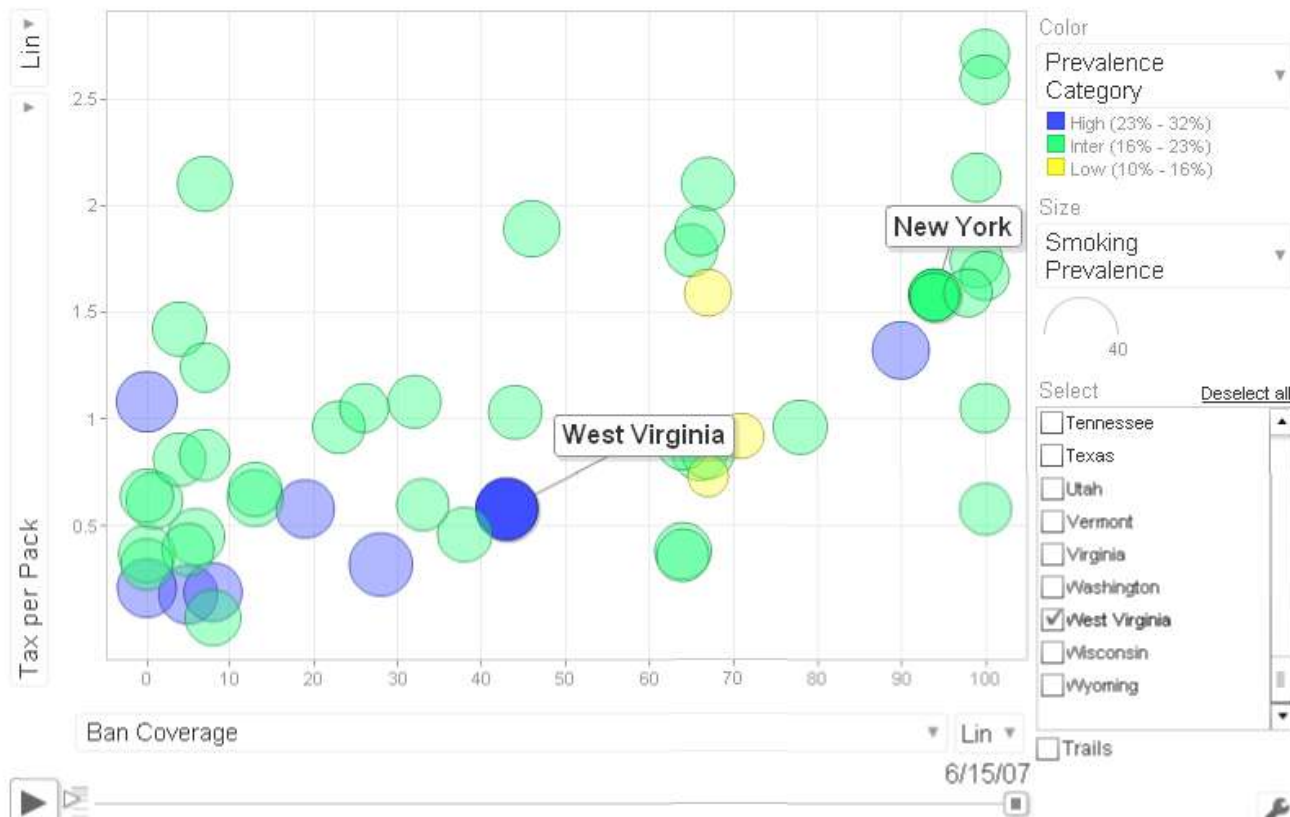
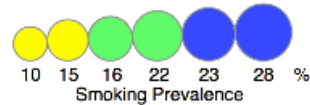
state: West Virginia



<http://cancercontrol.cancer.gov/hcirb/cyberinfrastructure/popsoci.html>

# Inter-relationships between cigarette taxes, smoke-free policy, and smoking prevalence

**Using the demo:** Press play (▶) in the chart below to show how each State's Cigarette Tax (\$) and Smoking Ban Coverage (%) changes from 1994 to 2007. The size of the circles corresponds to the percentage of smokers in the state, while the color of the circles corresponds to a "High", "Intermediate", or "Low" degree of the same quantity. The time point shown can be adjusted using the slider at the bottom of this chart.





# PopSciGrid – an Exploratory Tool

## Communication & Behavior

- Which policies (taxation, smoking bans, etc) impact health and health care?
- What data should be displayed to help consumers/policy makers evaluate related issues?
- How can data visualization help facilitate behavior change?

## Computer Science

- How can semantic technologies be used to integrate, present, and analyze data for a wide range of users?\*
- Can tools allow lay people to build their own demos and support public usage and accurate interpretation?\*
- Facilitate collaboration and make applications 'go viral.'

# Science (Feb. 11, 2011)



**Gary King (Harvard)**

*Social scientists need to continue to build a **common, open-source, collaborative infrastructure** that makes data sharing & analysis easy.*

**Peter Fox & James Hendler (RPI)**

*Visualization is not the end product of scientific analysis, but an **exploration tool** for hypothesis development in the continually more data-intensive scientific process*

*Challenge of modeling integrated data is exponentially harder = **systems-level perspectives & transdisciplinary science***

# Data for Community Impact: HHS Community Health Data Initiative



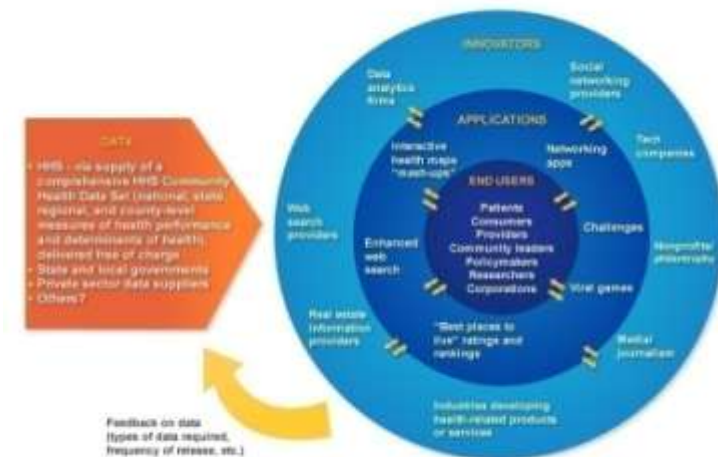
## Dec 8, 2009: White House Open Government Directive

- Transparency and data-sharing
- Public and community health
- Public-private innovation



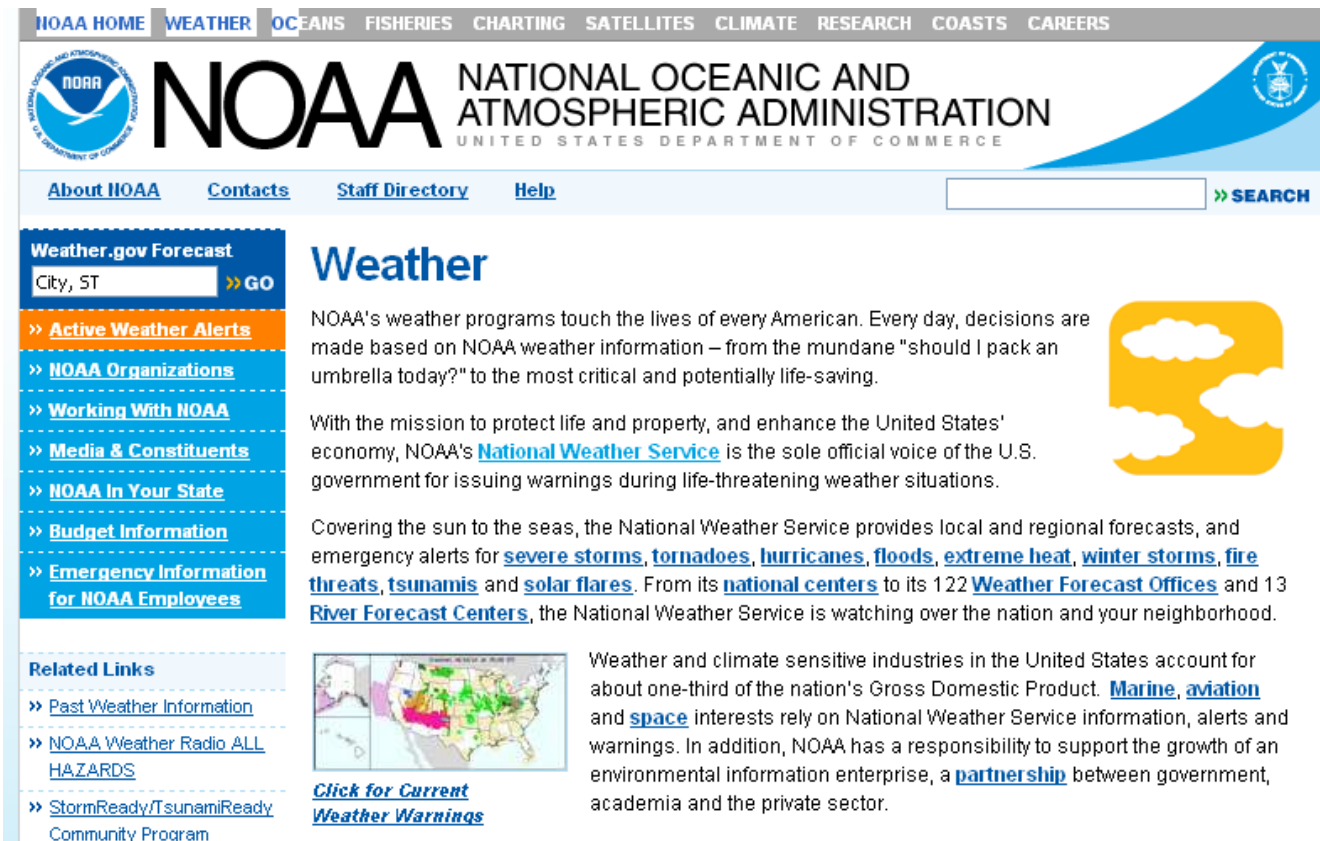
## June 2, 2010: HHS, IOM & NCHS – Community Health Data Initiative (CHDI)

- Enhance understanding of health and healthcare system performance in communities
- Data-sharing, collaboration, communication
- Who: government agencies, community organizations, foundations, healthcare & technology companies, software developers, etc.



# The Community Health Data Initiative (CHDI)

- ***Vision laid out on March 11, 2010: turn HHS into the “NOAA of health data”***





# What happened next: publication of a CHDI data website + an explosion of innovation



Source: Greg Downing, Office of the Secretary, HHS, 1/2011





[About](#) [Challenges](#) [Teams](#) [People](#) [Resources](#) [LIVE CODE!](#) [Partners](#) [Sponsor](#) [Gallery](#) [In the News](#) [Blog](#)

## Enabling Community Use of Data for Cancer Prevention and Control

Using data available through the National Cancer Institute and Population Sciences (DCCPS) and their partners, communication technology applications to enable cancer prevention and control.

DCCPS supports research in behavioral science, cancer epidemiology, and survivorship. Data from this research includes the burden of cancer in the population, cancer risk behaviors, cancer services, and cancer-related knowledge, attitudes and beliefs. The DCCPS also supports research on the use of consumer health information portals, which supports the use of

The screenshot shows the Challenge.gov website. At the top, there's a navigation bar with 'HOME' and 'FIND CHALLENGES' buttons. A search bar contains the text 'e.g. "Apps" or "Health"'. Below the navigation bar, the challenge title 'Enabling Community Use of Data for Cancer Prevention and Control' is prominently displayed. Underneath the title, it says 'From: U.S. Department of Health and Human Services' and 'Category: Science &amp; Technology, Health, Personal and Public Safety'. To the right, there's a 'Post a solution' button. Below the challenge title, there's a 'Submissions' section indicating 'End in about 1 month' and '13 supporters'. A red button says 'I support this challenge!'. At the bottom, there's a 'Share' section. The National Cancer Institute logo is visible in the bottom left corner of the screenshot.

- ***13 challenges issued to date by different organizations (govt, private, foundation)***
- ***Hundreds of developers signing up to compete***

# NCI Developer Challenge (Sept. 2010)

## **Co-developer:**

Lila J. Rutten, PhD, MPH (Behavioral Scientist,  
SAIC-Frederick)



**Goal:** Develop web/mobile apps to mash-up  
population data for cancer prevention & control

**Launched:** Sept 26, 2010      **Deadline:** Nov 22, 2010

**Awards:** Two travel awards to present at the 2011 Hawaii International Conference on Systems Sciences (Jan 4-7, 2011); winning apps promoted by Health 2.0 and NCI.

# NCI Developer Challenge Winners

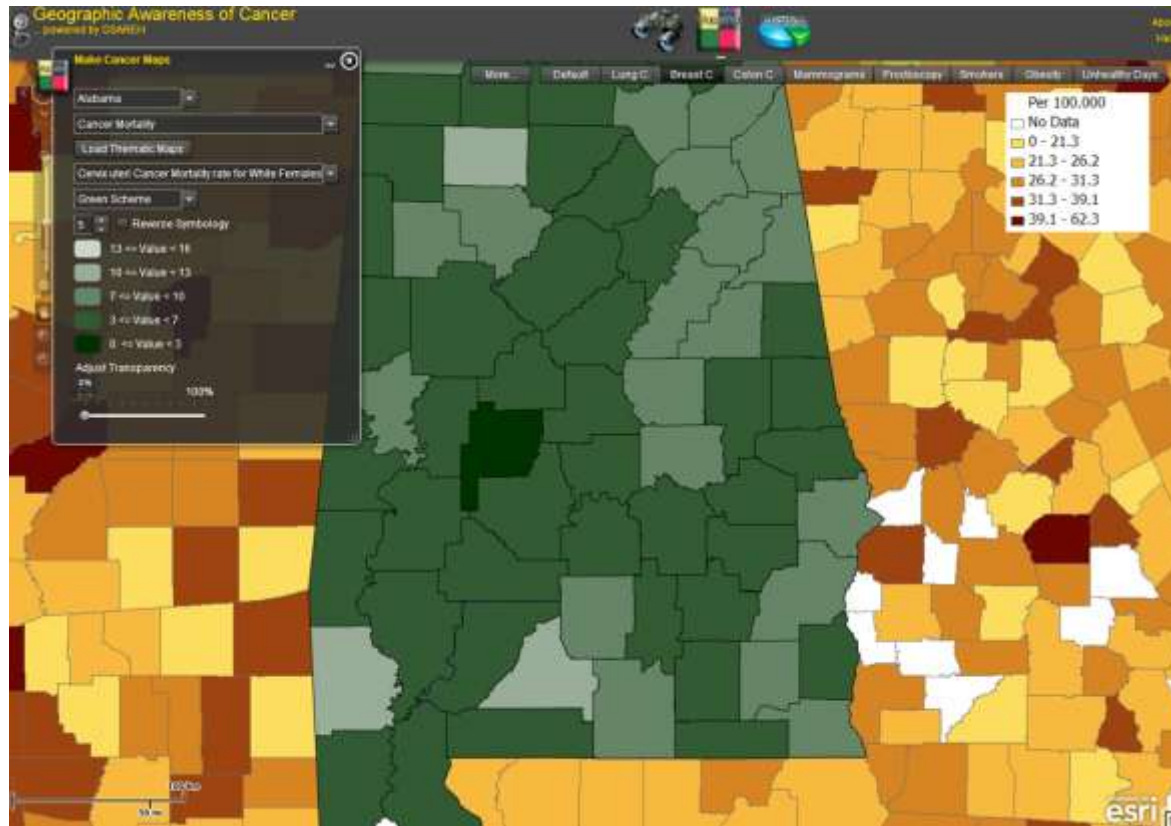
**ozioma**  
GOOD NEWS FOR YOUR COMMUNITY

Data gathering

Localization

[Washington University Health Communication Research Lab](#) – an online tool that provides locally-relevant cancer information in communities, Ozioma aggregates multiple data sources, including the NCI HINTS and SEER datasets, to enhance the relevancy of health related news stories for media sources.

# NCI Developer Challenge Winners



Geographic Awareness of Cancer Tool (GSAREH) – an interactive geospatial visualization tool that integrates complex national health-related data, including the NCI Atlas of Cancer Mortality, at the state and county levels with multiple options for comparing data and creating customizable maps, tables, and charts.



Fueling the emerging “health data ecosystem” – with ever more data, made ever more accessible

## **Additional HHS Data Commitments**

- 1 Community Data: new HHS Health Indicators Warehouse, launching 1/2011**
- 2 Provider Quality Data: COMPARE APIs launched 9/2010, COMPARE enhancements on the way**
- 3 Medical/Scientific Data: National Library of Medicine API portal, launched 9/2010**
- 4 Insurance Product Data from HealthCare.gov: coming soon**
- 5 New 2x/year data publication review process by each HHS agency, beginning 1Q 2011**
- 6 *HEALTHDATA.GOV*, new “one-stop resource” – debuting 2/2011**

Source: Greg Downing, Office of the Secretary, HHS, 1/2011<sup>23</sup>



# The goal: a “self-propelled” ecosystem of health data supply and use that creates escalating benefit

The collage features several digital health resources:

- Sonoma County Network of Care**: A website with a green and yellow header, navigation links (Library, Link), and a 'Healthy People 2010 : Progress Tracker' section.
- doximity**: A logo with the tagline 'Connected at Point-of-Care' and a smartphone displaying the app interface.
- Gallup · Healthways Well-Being Index™**: A logo and a snippet of text describing the index as a first-ever daily assessment of U.S. residents' health and well-being.
- U.S. Department of Health and Human Services**: A webpage titled 'Find a Health Center' with a search bar and a red 'Find Centers' button.
- healthTap**: A logo with a red 'WE'RE HIRING' sign above it.
- HRSA**: A logo for the U.S. Department of Health and Human Services.
- Microsoft Amalgam**: A logo with the tagline 'Unified Intelligence'.
- Triage**: A mobile app interface showing a red cross icon and buttons for 'Call 911', 'Find Medical Treatment', 'Find A Doctor', 'Look up your Symptoms', 'Research Diseases', 'Learn about Procedures', 'Your Health Record', and 'About and Help'.

- **Expanding use of apps developed thus far**
- **New app development work from startups, established companies, and agencies**
- **Private sector organizations beginning to release free and open data into the ecosystem**

Source: Greg Downing, Office of the Secretary, HHS, 1/2011

# Big Data

## Established Data Sources

- Surveillance
- Research
- Administrative
- Health systems
- Biological
- Genomic/proteomic
- Socio-demographic
- Policy...

## Emerging Data Platforms

- EHRs/PHRs
  - Mobile devices
  - Sensors & GIS
  - Data.gov
  - Web 2.0
- (blogs/microblogs/  
wikis, video/photo  
sharing, social  
networks, PHRs,...)

# Cyberinfrastructure

(Grid and Cloud Computing)



## Common Vocabularies & Middleware



# Applications for Consumer Health



- Apps
- Widgets
- EHRs
- PHRs
- Dashboards
- Medical devices  
& technologies



Re-Entry into the Information Ecology

# PCAST - “Designing a Digital Future: Federally Funded R&D in NIT” (Dec. 2010)

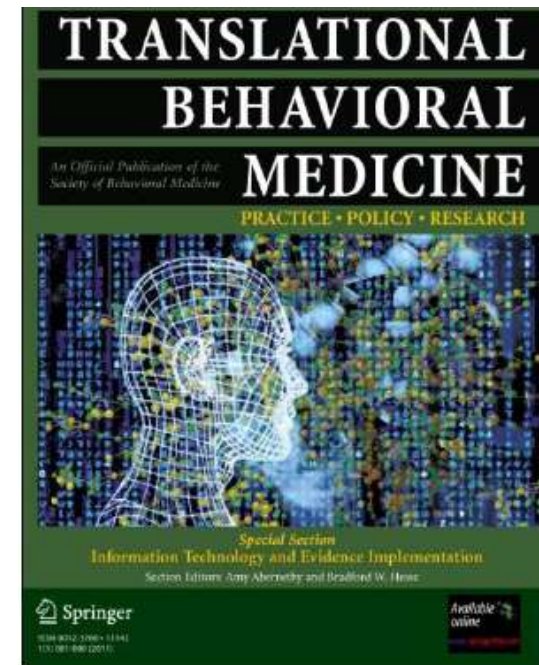
- I. “Every Federal agency needs to have a **BIG DATA** strategy”
- II. “Improvements in health & QOL...depend upon knowledge and information that can drive appropriate actions”
  - Comprehensive ‘**individual health record**’
  - From data to knowledge for empowering health (**scalable architecture**)
  - Tools & apps for **personalized healthcare**
- III. “Democratization of research & education”
- IV. **NIH & NSF collaboration:**
  - Data sharing & computation
  - Environment and sensor data (contextualization)
  - HCI and ‘human-NIT’ interactions



# Upcoming NCI Knowledge Products: Cyberinfrastructure & Data Liquidity



Shaikh, A.R., Prabhu Das, I., Vinson, C., Spring, B. (eds.). **Cyberinfrastructure for Consumer Health.** *American Journal of Preventive Medicine.* April 2011 (in press)



Abernethy, A., Hesse, B., Spring, B (eds.). **Information Technology and Evidence Implementation.** *Translational Behavioral Medicine.* April 2011 (in press)