The Public Health Grid (PHGrid): Overview and Value Proposition

Tom Savel, MD
Lead – Grid Computing
Director – NCPHI Research & Development Lab
National Center for Public Health Informatics, CDC

Statewide Epidemiology and Informatics Conference
Columbus, Ohio
June 17, 2009

Overview

- Public Health Challenges
- What is Grid
- Value of Grid to Public Health
- Our Activities
- Looking Ahead
- Q&A

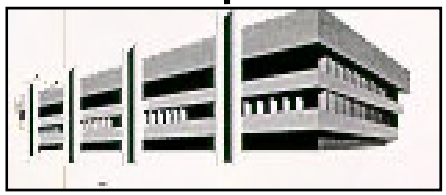
Current Challenges in Public Health

- Public health data widely distributed
- Volume of public health data growing rapidly
- Many cultural, social and political impediments to data sharing
- Requires a stronger economic model for long-term financial sustainability
- Uniquely dynamic, complex and global in scale
  - Situational awareness, population health, event detection, inventory tracking, countermeasures administration, alerting, etc
- Many redundant systems, application silos and data silos.

Current State of Public Health Surveillance (Data Flow)

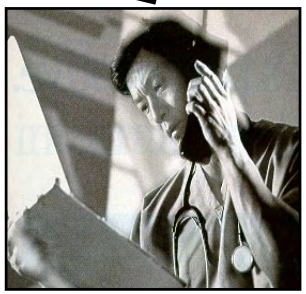
- ❑Intensive data gathering from medical facilities, state & locals into a giant CDC owned data warehouse
- ❑Heavy use of statistical algorithms to detect anomalies in the data and trigger investigations
- ❑CDC Centric Approach to developing and deploying software

CDC



State, County, and Local Health Departments

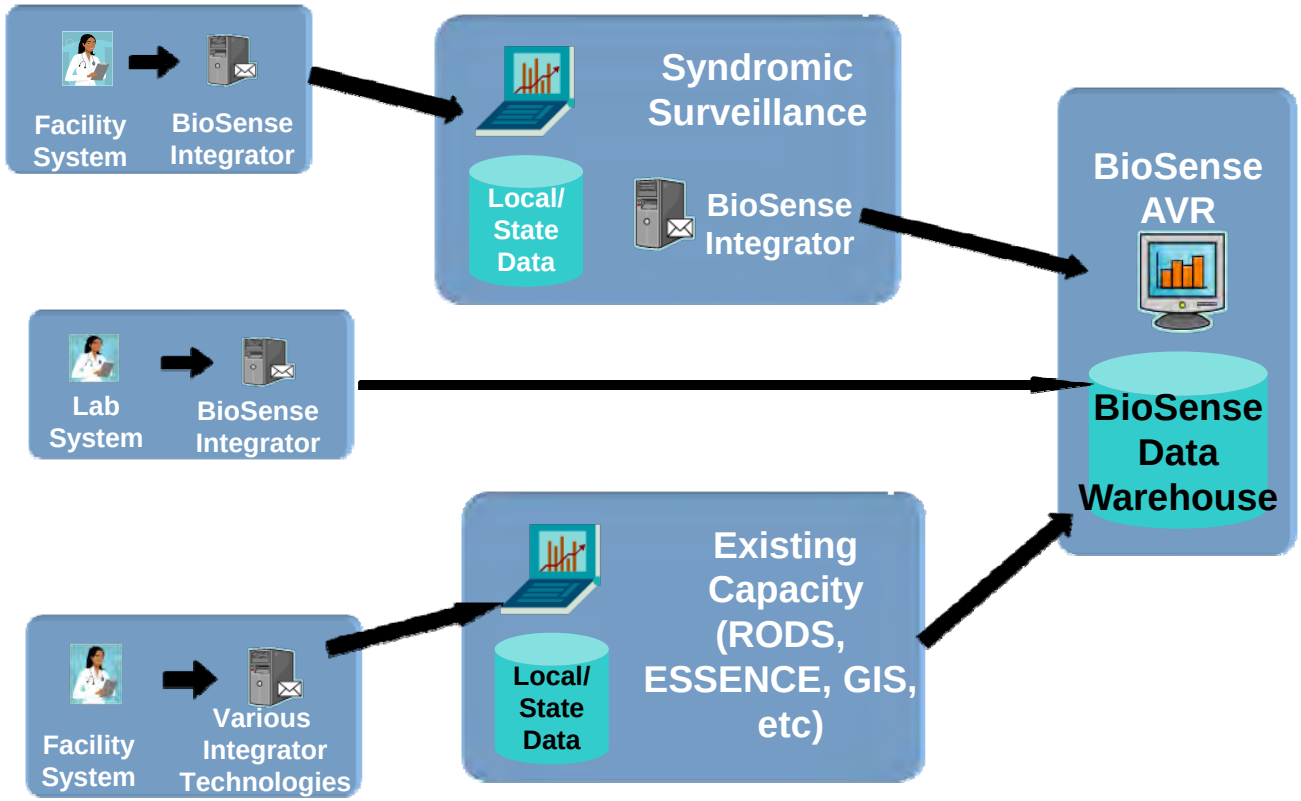
Laboratories



Practitioners

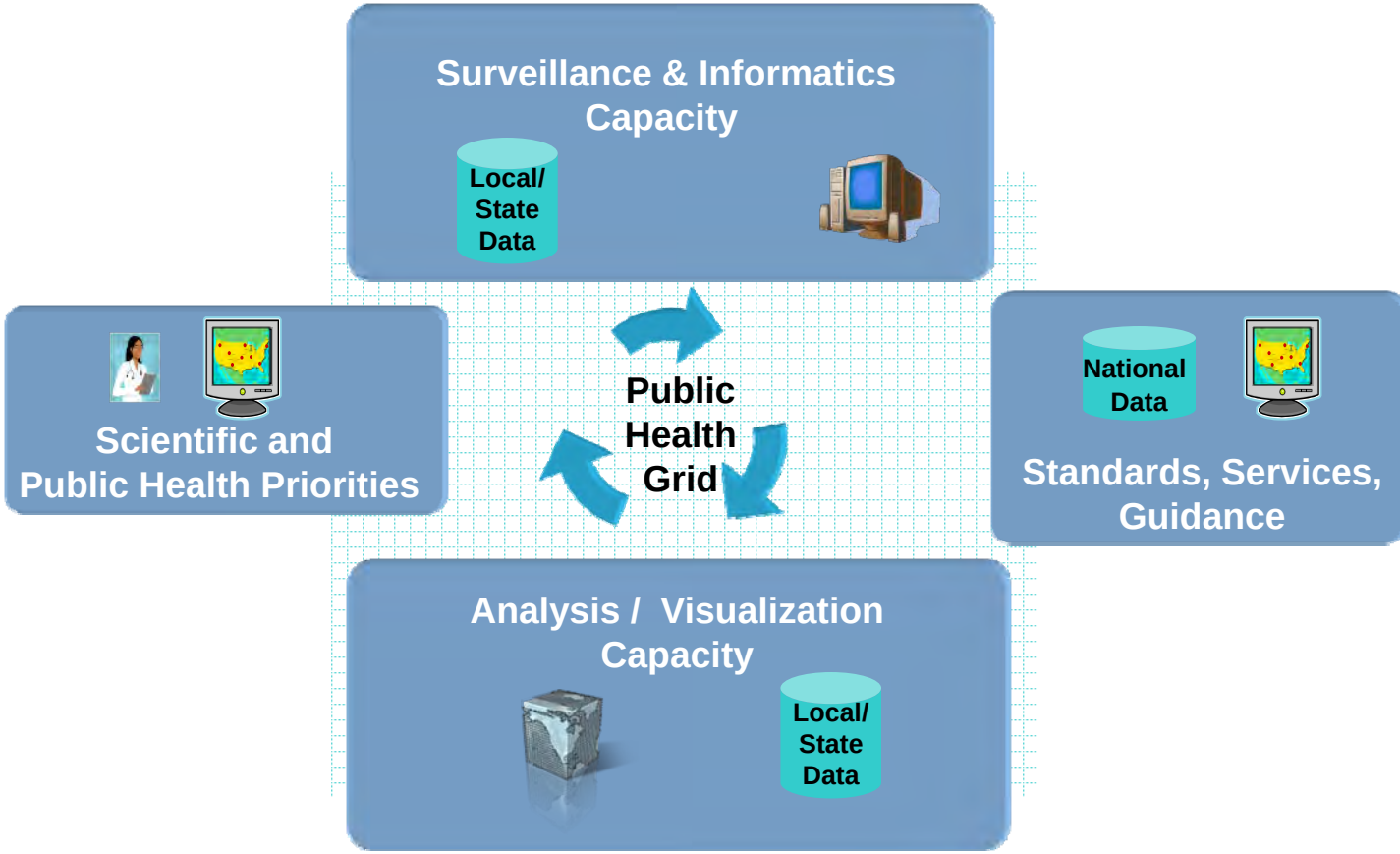
Current Situation and Challenges

- Politics of ***control*** of data has been the primary obstacle to formation of a national system
- Much existing data remains ***siloed at the Local/ State level*** – accessibility and visualization limited
- Building systems non collaboratively leads to ***low adoption*** rates



Future Goal: Federated Architecture (Grid)

- Leverage *Existing Capacity*
  - **Distribute resources and infrastructure**
  - **Increase flexibility and scalability**
- Provide *Local Control* of data and services
  - **Reduces political barriers**
  - **Address many privacy concerns**
- Foster *Collaboration* to define requirements, priorities, develop, and deploy technology



What is a Grid?

- A computing infrastructure…
  - **provides dependable, consistent, pervasive and inexpensive access to data and applications.**
  - **By pooling federated assets into a virtual system, a grid allows data owners to share data and applications while maintaining control.**

Grid Represents…

- Different way of thinking
- Different way of solving problems
- A long-term, fiercely collaborative approach
- Large-scale Computer Trends

Types of Grids

- Computational Grids
  - Focused on the application of several computers to a single problem at the same time – usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data.
  - (virtual super-computer)
- Collaboration / Access Grids
  - Focused on the collection of resources and technologies that enables large format audio and video based collaboration between groups of people in different locations.
- Data Grids
  - Focused on he controlled sharing and management of large amounts of distributed data.

- Dynamic combinations or all on same platform

Computational Grids

Most famous/infamous

A massively distributed computing environment composed of over 3 million Internet-connected computers launched in May 1999—has led to a unique public involvement in science.

Three million PCs deliver 6,000 CPU years per day—the fastest (admittedly special-purpose) computer in the world

@Home Model Extended

Folding@home distributed computing

Grid application models protein folding & misfolding (1224 teraflops, as of 23 Sept 2007)

@FRICA home

Grid application models the way malaria spreads in Africa and the potential impact that new anti-malarial drugs may have on the region

Grid application models the design of new anti-HIV drugs based on molecular structure (in silico)[1]

Collaboration Grid Example

- **Presentation, visualization and interactive environments**
- **Runs on the same grid as the computational grid**
- **These combined resources are used to support group-to-group interactions, large-scale distributed meetings, collaborative work sessions, seminars, lectures, tutorials, and training**

Access Grid aids SARS patients

During the 2003 Sudden Acute Respiratory Syndrome (SARS) virus outbreak in Taiwan, Argonne's Access Grid technology enabled radiologists from across the country to review patients' X-rays without added risk of infection. Three members of Argonne's Mathematics and Computer Science Division assisted in the fast deployment of the specialized Grid.

Access Grid technology, conceived and developed at Argonne, connects people across town or around the world to interact and exchange ideas with each other in real time through high-resolution video and voice over the Internet. Each "node" in the grid may be able to display a dozen or more real-time images of other users, documents, Web sites and other useful data.

SARS GRID — Access Grid technology enabled radiologists across Taiwan to review SARS patients' daily X-rays without risk of infection.

Normally, laboratory and academic groups use the Access Grid for training and education, conferences and workshops, site

Data Grid Examples

caBIG – Cancer Research Datagrid

GEON – Geosciences Network Datagrid

EGEE / CERN - The world's largest particle physics laboratory... where the web was born (LHC – The Large Hadron Collider, May, 2008)

DataGrid –EU funded resource of shared large-scale databases

TeraGrid –Shares resources at San Diego Supercomputer Center, Indiana University, Oak Ridge National Laboratory, National Center for Supercomputing Applications, Pittsburgh Supercomputing Center, Purdue University, Texas Advanced Computing Center, University of Chicago/Argonne National Laboratory, and the National Center for Atmospheric Research

Commercial Grid Products

Tier 2

Tier 1

Open Source Grid Software/Projects

PRAGMA
PACIFIC RIM APPLICATIONS AND GRID MIDDLEWARE ASSEMBLY

the globus alliance

egee
Enabling Grids for E-sciencE

HealthGrid

NGrid
Open source grid computing

gLite
Lightweight Middleware for Grid Computing

DutchGrid
Large-scale Distributed Computing in The Netherlands

CLEVERSAFE.ORG

OpenGridForum
OPEN FORUM | OPEN STANDARDS

Commercial Grid “Consumers”

Google

Adobe

Second Life

AMD
Smarter Choice

amazon
web services™

BOEING

GOODYEAR

Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

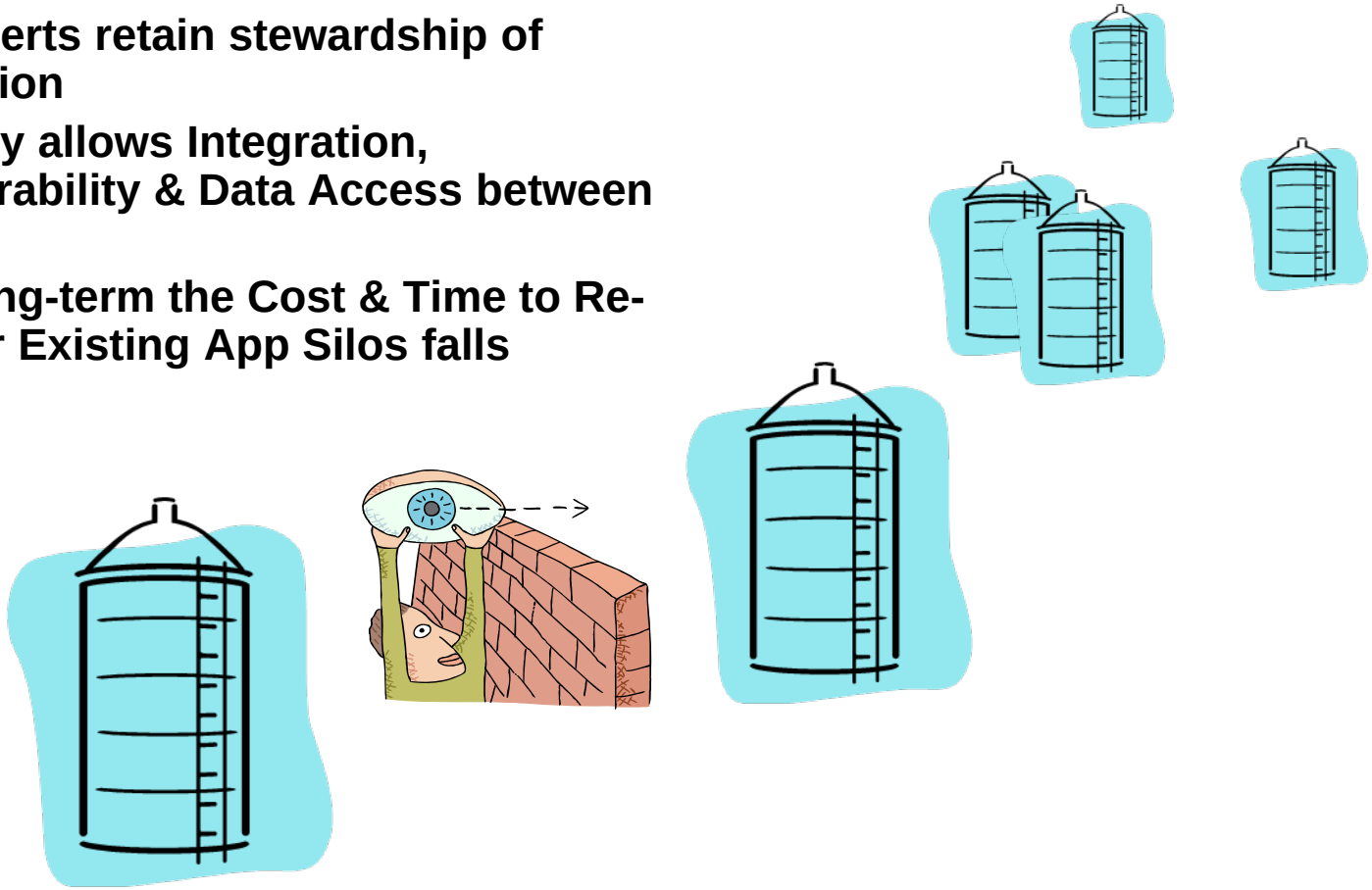
PARTNERS
HEALTHCARE

Grid and Healthcare

- International HealthGrid Association
  - 1 billion Euros
- HealthGrid.US
- US - Federal
  - National Cancer Institutes' Cancer Bioinformatics Grid (caBIG) initiative
    - Launched in 2004
    - For cancer research community
  - National Science Foundation
    - "Cyberinfrastructure"
      - » The coordinated aggregate of software and hardware and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering. The challenge of Cyberinfrastructure is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and "end-to-end" coordination^
      - » ^ SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences, Fran Berman, San Diego Supercomputer Center and UC San Diego

Grid as a supporting technical framework for public health…

- **Lets experts retain stewardship of information**
- **Flexibility allows Integration, Interoperability & Data Access between Silos**
- **In the long-term the Cost & Time to Re-engineer Existing App Silos falls**



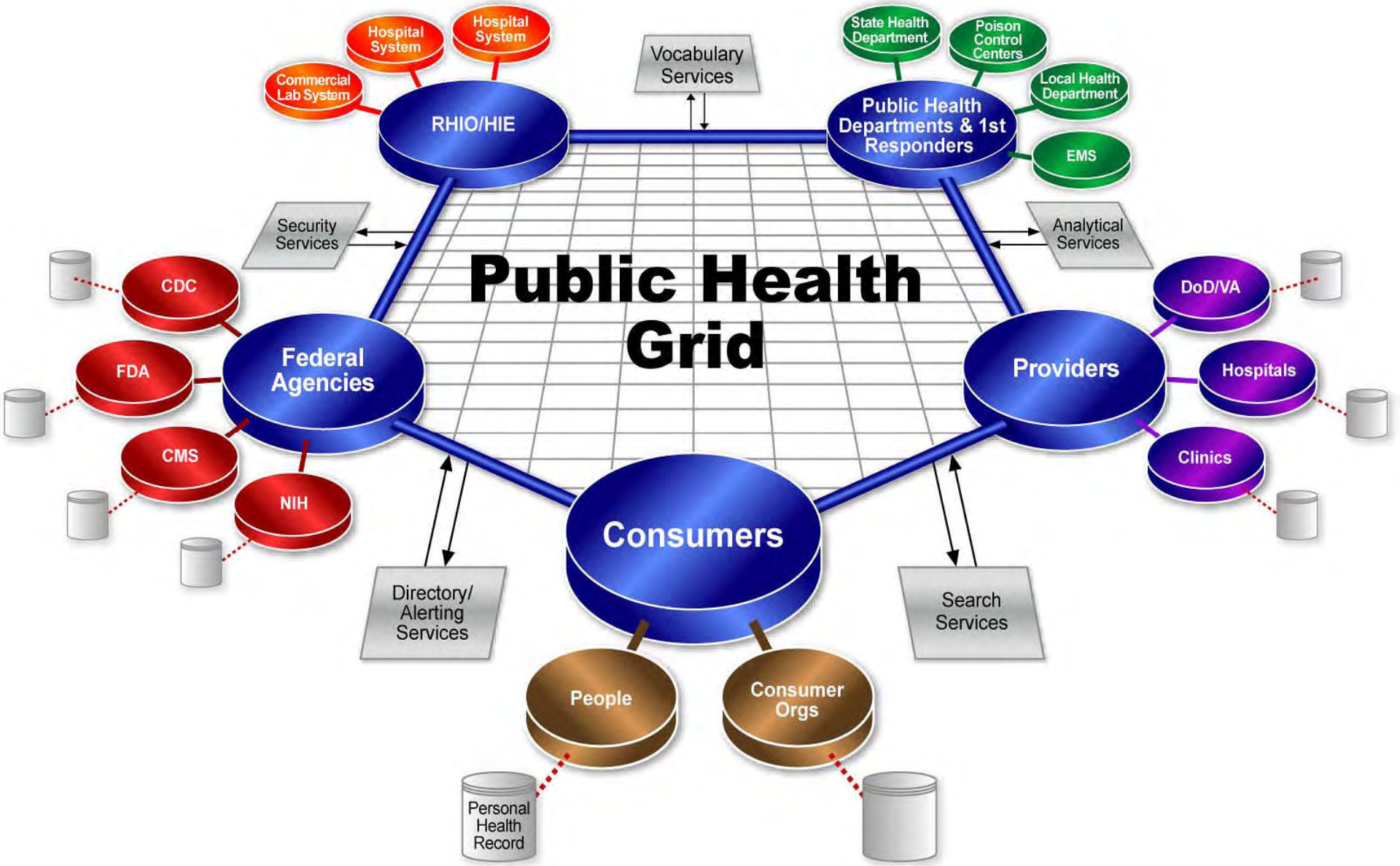
Value of Grid to Public Health – Meeting those Challenges

Ideal Attributes of a Public Health Grid:

- Open-architecture
- Federated
- Scalable
- Flexible
- Redundant
- Leveraging best practices

- Thus, meeting the financial, social, technology, and security challenges…

Conceptual Representation

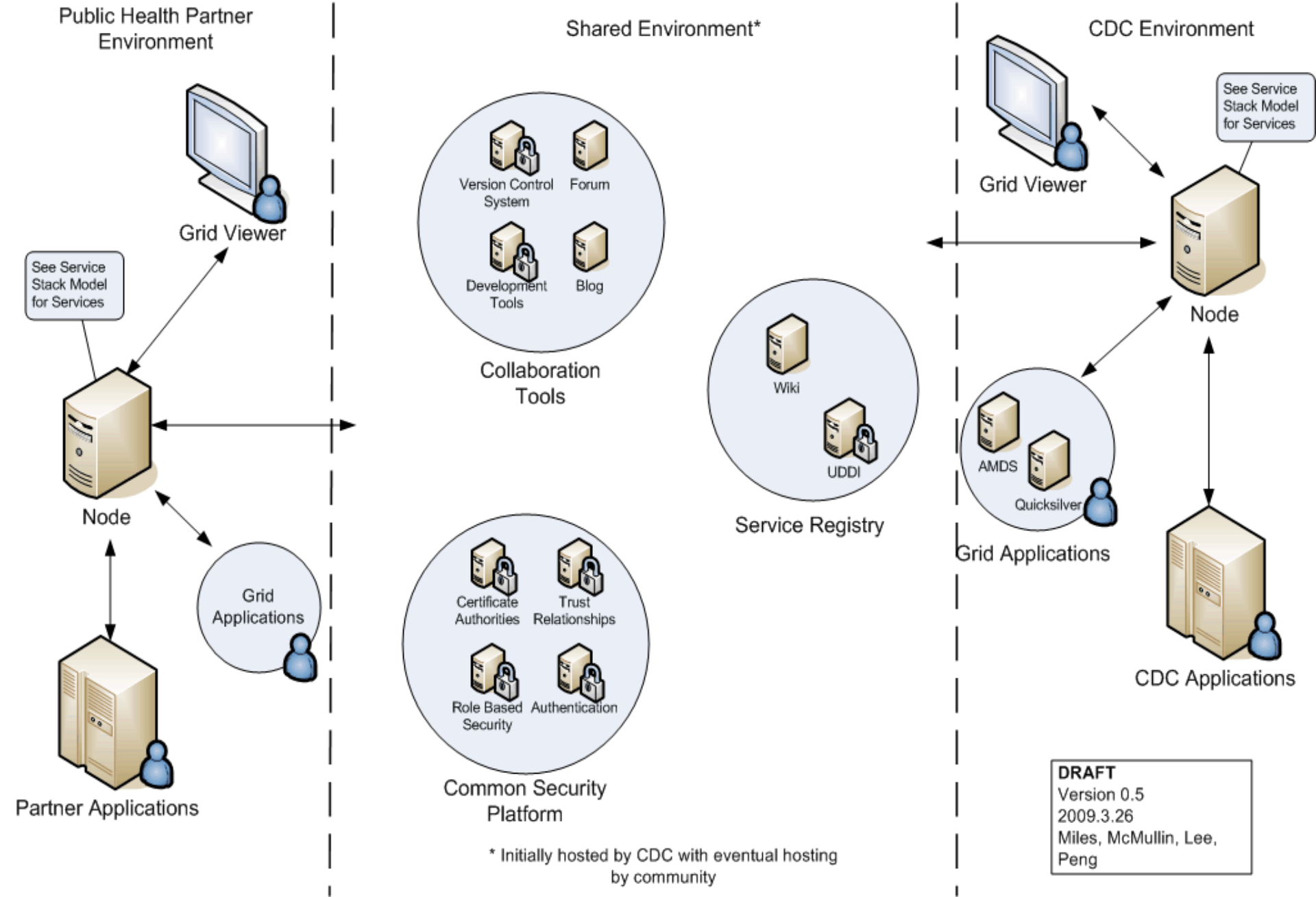


PHGrid Initiative: Purpose of Research

- Determine the viability of Federated Architecture in Public Health
- Establish relationships with key partners / collaborators
- Determine / Inform future public health informatics approaches

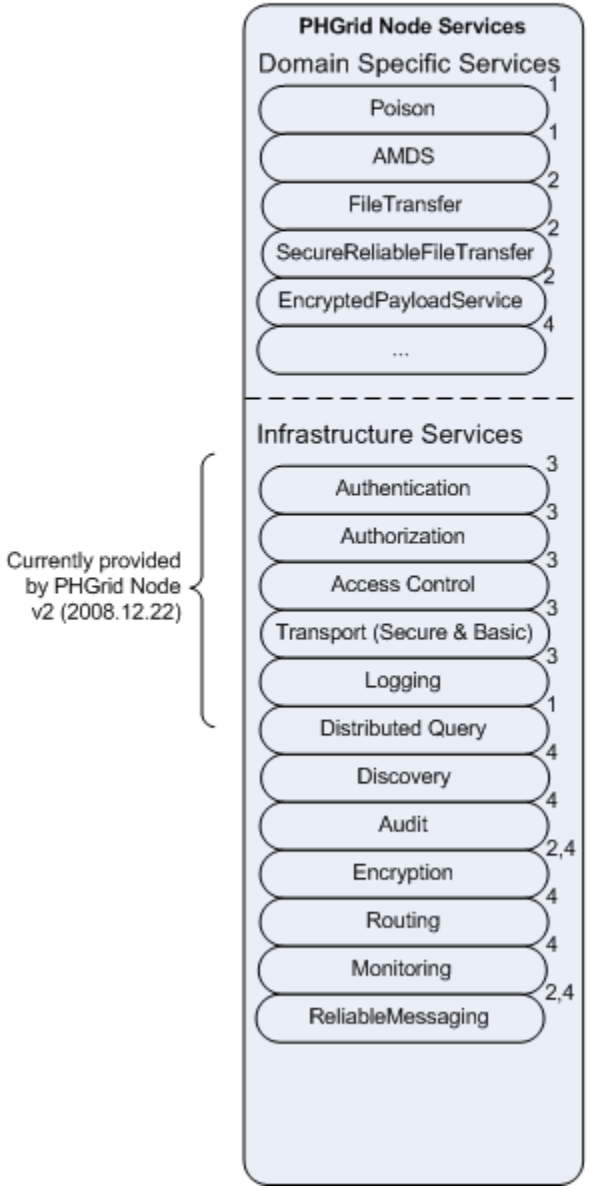
- And….to move from the conceptual to the actual

PHGrid Conceptual Architecture



PHGrid Service Stack

A configurable node installation (mix & match) allows a partner to choose which components are installed and run.



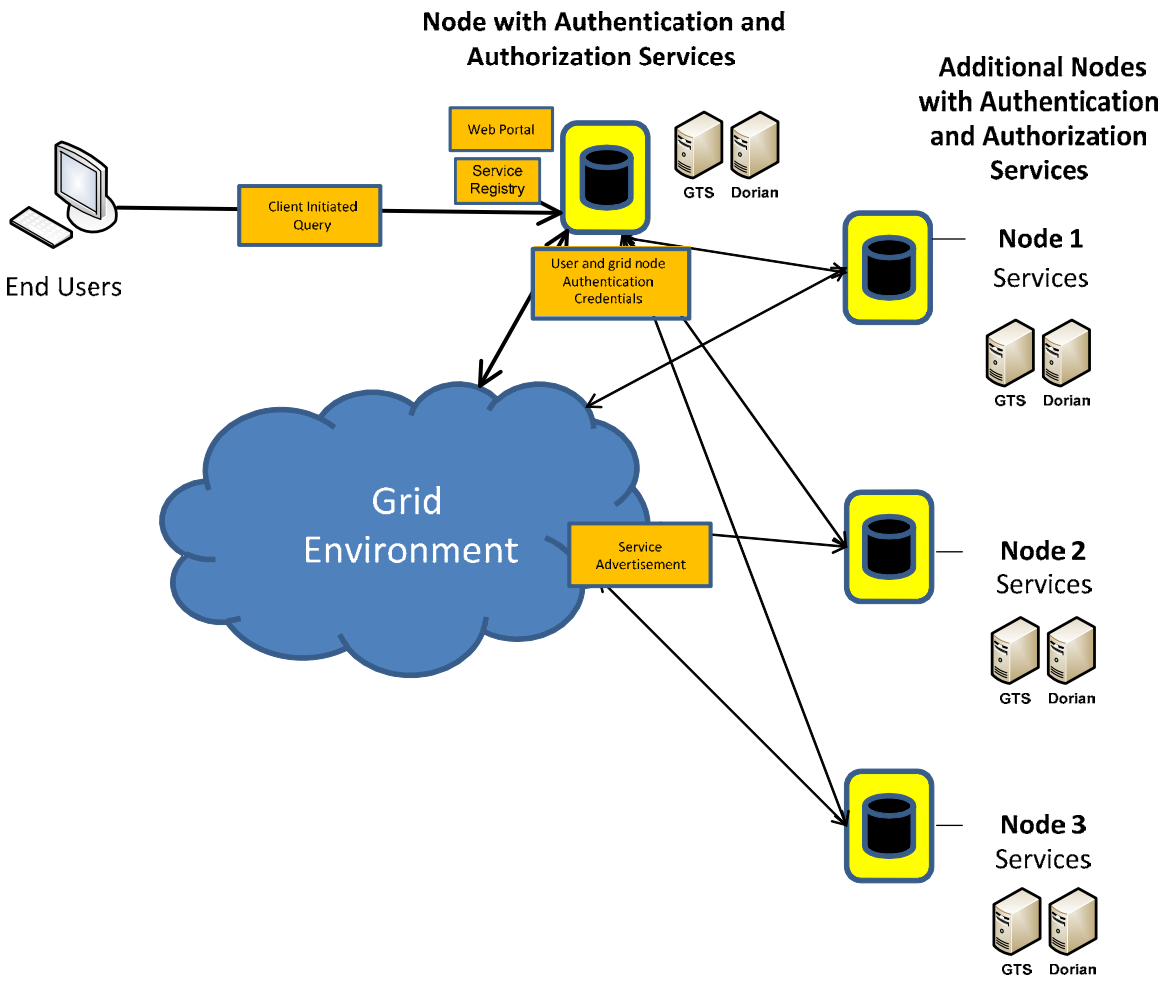
Status of Development	
1	Service Proof of Concept
2	PHIN-SRM
3	Demonstrated (Globus/GAARDS)
4	Planned

DRAFT
Version 0.5
2009.3.26
Miles, McMullin, Lee, Peng

Examining PHGrid and Security: Leveraging the Expertise of Others…

Grid Authentication and Authorization with Reliably Distributed Services (**GAARDS**) is a series of tools developed by Ohio State to enhance the open source Gobus Toolkit.

Provides enterprise level administrative tools for managing users, federated identities, trust, credential delegation, group management, access control policy, and integration between grid and non-grid-based security domains.



PHGrid Activities: Goals / Objectives

- Provide a secure, easy-to-use national technical and social infrastructure for solving public health problems
- Develop an extremely low cost grid appliance
- Simplify web services development (drag & drop)
- Simplify data access and data exchange (drag & drop)
- Connect public health grid to other grids, and to other data sources, regardless (in other words, interoperate with everything)
- Recruit local & state health departments, HIEs, RHIOs, academic institutions, national data sources, medical centers, international public health partners, and vendors

Guiding Principles - Partnerships

- Partnership Attributes
  - **Volunteerism (varying degrees)**
  - **Enthusiasm**
  - **Capability / Leadership**
    - **Public Health**
      - **Use Cases, etc.**
    - **Technical**
      - **Grid, Open Source (when possible), SOA, Distributed Systems**
    - **Ambiguity**
- What is developed needs to align with the philosophy of:
  - **Sustainability**
  - **Low barrier to entry**
    - **Technically, financially, socially**
  - **100% Standards-based**

Current State of PHGrid Activities: Research & Practice

- Our Methodology
  - **First develop hypothesis and then perform research**
  - **Develop evidence base**
  - **Make evidence-based decisions on the value of potential tools & resources**
  - **Apply selected tools to both existing and novel systems**
  - **Move systems to production**
  - **Continuous evaluation and enhancement**

Current State of PHGrid Activities: Research & Practice

- Leveraging:
  - **Existing technology applied to a novel domain (public health)**
  - **An open / collaborative development process with our partners (academic, PH, industry)**
  - **CDC as participant – not has solution owner**
  - **An evidence base (not personal preference)**
- Challenges
  - **Gain expertise in the wide variety of grid-based resources currently available**
  - **Define the gaps between the PH and Grid domains**

So…How should you think about grid?

Secure….

Service Oriented Architecture (SOA)…

Ecosystem…

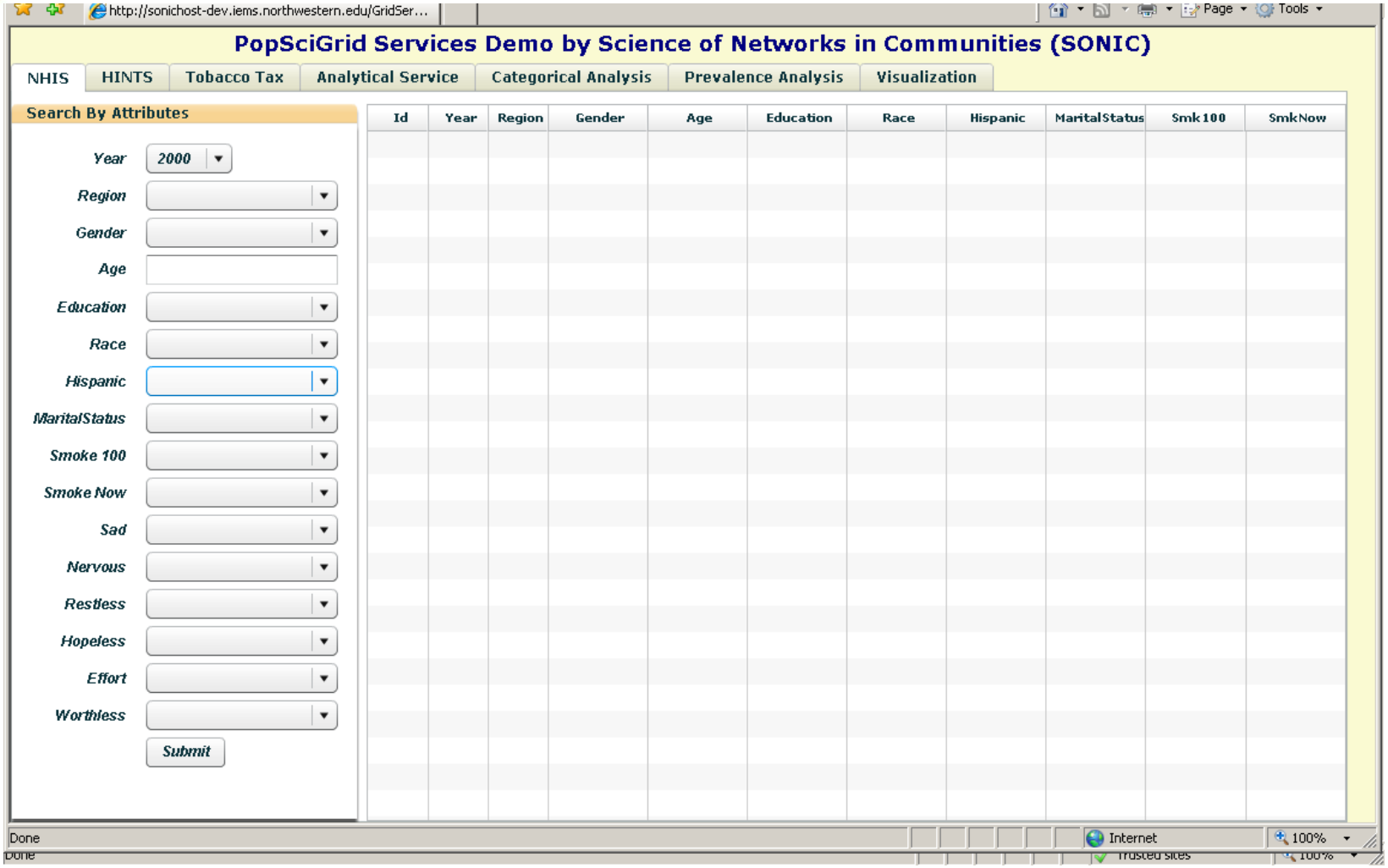
Community…

The Public Health Grid (PHGrid) Current Activities – the details…

Overview

- A Sample of others…
- Research Activities – What are we doing?
- Tools – How do we do it?
- What have we learned?
- The Future
- Questions?

Data Grids and Public Health Research



http://sonichost-dev.iems.northwestern.edu/GridServer/c/index.html

PHGrid Research Projects (examples)