



# **International Open Government Data Conference**



# International Open Government Data Conference

## Data Sharing to Solve Mission Problems— Are We Doing Enough?

Alan Vander Mallie, *Federal Program Manager for DataGov PMO, Office of Citizen Services & Innovative Technologies, U.S. General Services Administration (GSA, Moderator)*

Robert Schaefer, *Ph.D., Deputy Project Scientist, Johns Hopkins University Applied Physics Lab*

Steve Young, *Associate Director, Office of Information Access and Analysis, Office of Environmental Information, U.S. Environmental Protection Agency*



**DATA.GOV**  
EMPOWERING PEOPLE

*Data Sharing  
to Solve Mission Problems –  
Are We Doing Enough?*

November 2010

*Changing the Landscape of Data Sharing*

Official Site of the United States Government Sunday, August 08, 2010 Text: A+ A- A Show

**DATA.GOV**  
EMPOWERING PEOPLE

HOME DATA TOOLS COMMUNITY METRICS DIALOGUE GALLERY WHAT'S NEW

### SHERMAN ACT VIOLATIONS YIELDING FINES OF \$10 MILLION OR MORE

The Antitrust Division's chart of criminal fines depicts corporate fines of \$10 million or more that have been imposed for violations of the Sherman Antitrust Act. Sherman Act violations typically involve price fixing, bid rigging, or allocations of markets, customers, or sales or production volumes of products and services purchased by consumers and manufacturers. Such crimes rob purchasers, inflate prices, restrict supply, inhibit efficiency, reduce innovation, and undermine our system of free enterprise.

[VIEW MORE](#)

### Most Popular Datasets

1. Food and Drug Administration--Recalls
2. Worldwide M4 Earthquakes, Past 7 Days
3. U.S. Army Human Interest
4. RSS Feed for Graphical Tropical Weather ...
5. Uniform Tire Quality Grading System (UTQGS)

SEARCH OUR CATALOGS

Search our catalogs... [SEARCH](#)

#### APPS

With so much government data to work with, developers are creating a wide variety of applications, mashups, and visualizations. From crime statistics by neighborhood to the best towns to find a job to seeing the environmental health of your community—these applications arm citizens with the information they need to make decisions every day. Enjoy these highlights of the hundreds of applications available.

[READ MORE](#)

#### COMMUNITY

Data.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful applications that allow for new views and public analysis. This is a work in progress, but this movement is spreading to cities, states, and other countries. After just one year a community is born around open government data.

**Just look at the numbers:**

- 6 Other nations establishing open data
- 8 States now offering data sites
- 8 Cities in America with open data
- 236 New applications from Data.gov datasets
- 253 Data contacts in Federal Agencies
- 272,677 Datasets available on Data.gov

[VIEW MORE](#)

#### SEMANTIC WEB

As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to realize the promise of Linked Open Government Data.

Thanks to our collaboration with the **Tetherless World Constellation** at the **Rensselaer Polytechnic Institute**, Data.gov is now hosting one of the largest open collections of RDF datasets in the world! Check out some of their semantic mashups **we're featuring** and read our blog entry to learn more about where we are, where we're going, and why we think this platform will add tremendous value to democratized data.

[READ MORE](#)

DATA.gov | About | FAQ | Contact Info | Data Policy | Accessibility | Privacy Policy | Sitemap



### A government that is:

- **Transparent**
  - Promotes accountability
  - Provides information for citizens on what their government is doing
- **Participatory**
  - Provides citizens opportunities to participate in policymaking
  - Solicits ideas from citizens on how to improve those opportunities
- **Collaborative**
  - Uses innovative tools to enable collaboration across and at all levels of government

## Who is Presenting

- Alan Vander Mallie, GSA and Data.Gov PMO, Moderator
- Steve Young, EPA
- Robert Schaefer, JHU/APL



the WHITE HOUSE PRESIDENT BARACK OBAMA

★★★★ THE WHITE HOUSE WASHINGTON ★★★★★

Get Email Updates | Contact Us

BLOG PHOTOS & VIDEO BRIEFING ROOM ISSUES the ADMINISTRATION the WHITE HOUSE our GOVERNMENT

Home • The Administration • Open Government Initiative

Search WhiteHouse.gov Search

## Open Government Initiative

TRANSPARENCY + PARTICIPATION + COLLABORATION

OPEN GOV

About Open Government | Open Gov Blog | Around the Government | Innovations Gallery

My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.

- PRESIDENT OBAMA, 01/21/09

Promotes accountability

Allows people to contribute ideas/expertise; government benefits from broad knowledge sharing

Encourages cooperation within government and with industry

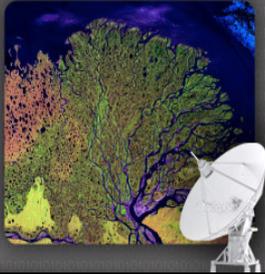
# Goals of Data Sharing

FEATURED TOOL:

**U.S. GEOLOGICAL SURVEY (USGS)**

USGS Global Visualization Viewer for Aerial and Satellite Data

Ten million archive images of the Earth's surface are available for immediate selection and free download via the USGS Earth Resources Observation and Science (EROS) Center's Global Visualization Viewer at <http://glovis.usgs.gov/>. Users can preview thumbnails, browse images and download full-image selections from 1.5 million aerial photos of U.S. sites and 8.5 million images captured worldwide by U.S. Earth-observing satellites.



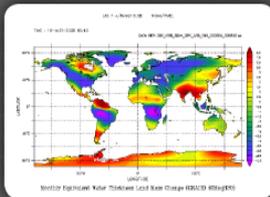
VIEW *THIS TOOL* ▶

FEATURED TOOL:

**NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA)**

MyNASA

Mentoring and inquiry using NASA data on Atmospheric and earth science for Teachers and Amateurs. The MY NASA DATA Live Access Server (LAS) is now available to create your own microsets for your class or your interests. The LAS contains over 149 parameters in atmospheric and earth science from five NASA scientific projects.



VIEW *THIS TOOL* ▶

- **Break down stovepipes** to increase access to data
- **Instant access:** downloadable data only one click away
- Data that can be **manipulated** and **mashed up**
- Encourage development of **innovative applications**
- **Tools that provide access** to complex data stores
- **Widgets**

## What this Session is About

- Highlighting specific examples of Solving Mission Problems from GSA, EPA and Johns Hopkins University APL

- **Creating a Transparent Government**
  - Data.gov provides a way for government agencies to share data on performance and their mission with the public.

**DEEPWATER HORIZON RESPONSE**

Due to the high level of interest in the on-going oil spill in the Gulf of Mexico, Data.gov is featuring data from the Department of Energy (DOE), the Environmental Protection Agency (EPA), the National Oceanic and Atmospheric Administration (NOAA), and the states of Florida and Louisiana related to the spill, its effects, and the cleanup effort. Data include oil and gas flow and recovery measurements, air and water sample data, oil spill-related exposure information, and other data of interest to scientists, recovery workers, and citizens. Data.gov will continue to provide access to as much oil spill-related information as possible.

**BP OIL SPILL DATA AND OTHER INFORMATION FROM FEDERAL AGENCIES AND STATE GOVERNMENT**

- Department of Energy**
  - Data Summary from Deepwater Horizon
    - Well Configuration
    - Timeline of Key Events
    - Containment System
    - BOP Stack Details
    - Top Kill Operations
- Department of Homeland Security**
  - GeoPlatform
    - Mapping Deepwater Horizon MC252
- Environmental Protection Agency**
  - EPA Response to BP Spill
    - Air Data
    - Water Data
    - Sediment Data
    - Waste Information
- National Aeronautics and Space Administration**
  - NASA Imagery of Oil Spill
    - MODIS Rapid Response System
    - AVIRIS Aids in Gulf Oil Spill Response
- National Oceanic and Atmospheric Administration**
  - Software and Data Sets
    - Surface Oil Maps
    - Digital Coast
- State of Florida**
  - Deep Water Horizon Oil
    - Latest News
    - Authorization
    - Response Numbers
    - Natural Resource Damage Assessment
- State of Louisiana**
  - Oil Spill Resources
    - Health Surveillance
    - Video Coverage
    - News from Louisiana

## Check It and See

- Using Environmental Protection Agency data, this app lets you search by zip code or Google map interface to find environmental issues in your neighborhood.
- Links directly to detailed EPA pages for more information and how to report an issue.
- <http://checkitnsee.com/>



# Goals of Data Sharing



- **Creating collaborative government**

Agencies and Data.gov working together create a knowledge management capability across the government at the data layer.

## Data.gov

## Data Sharing to Solve Mission Problems

### -Are We Doing Enough?

- Hear from agencies/organizations that are actually sharing data to improve government efficiency and drive innovation.
- Helps us to better understand the changing landscape of Data Sharing, and
- How openness and transparency is improving Government.

## Current & Evolving Data Environments

- Current data activities focus/emphasis on collecting, cleansing, transforming, integrating, storing, and reporting
- Evolving data activities focus/emphasis:
  - Exploring, visualizing, analyzing, communicating, monitoring
  - Used for forecasting/predicting future – explaining past events
  - Used for decision making & understanding data

### GSA was first to Post Govt-wide Datasets

- Federal Advisory Committee Act Datasets for last 10 years on DataGov in tools and raw csv datasets
  - Federal agency activity for over 1,000 advisory committees government-wide
  - Congress, the Public, the Media, and others use datasets to stay abreast of important activities
  - True shared-system providing access to data

GSA shares other Government-wide data:

- FAQs for USA.gov—Federal Customer Information Center provides real-time feeds on most popular government Q&A's
- USA Spending Grants & Loans
- USA Spending Contracts & Purchases
- USA Spending Federal Procurements

GSA engaged the Public using IdeaScale

-Public asked for:

- Location-enabled Vacancy Rates for Federal Buildings (utilization, condition, op costs)
- Performance data on obtaining a GSA Schedule (how long, how much)
- Sharing of contracting information across regions (works in Region 7...)
- More transparency for Primes/Large Business partnering with Small (Who to partner with...)

# *Questions*

Alan Vander Mallie, PMP  
Data.Gov Program Management Office  
Office of Citizen Services and Innovative Technologies  
U.S. General Services Administration  
Phone 202-501-6901  
Email: [alan.vandermallie@gsa.gov](mailto:alan.vandermallie@gsa.gov)

# Panel Presentation: “Data Sharing to Solve Mission Problems – Are We Doing Enough?”

International Open Government Data Conference

November 15, 2010

U.S. Department of Commerce, Washington, DC USA

Panelist: Steve Young,

U.S. Environmental Protection Agency

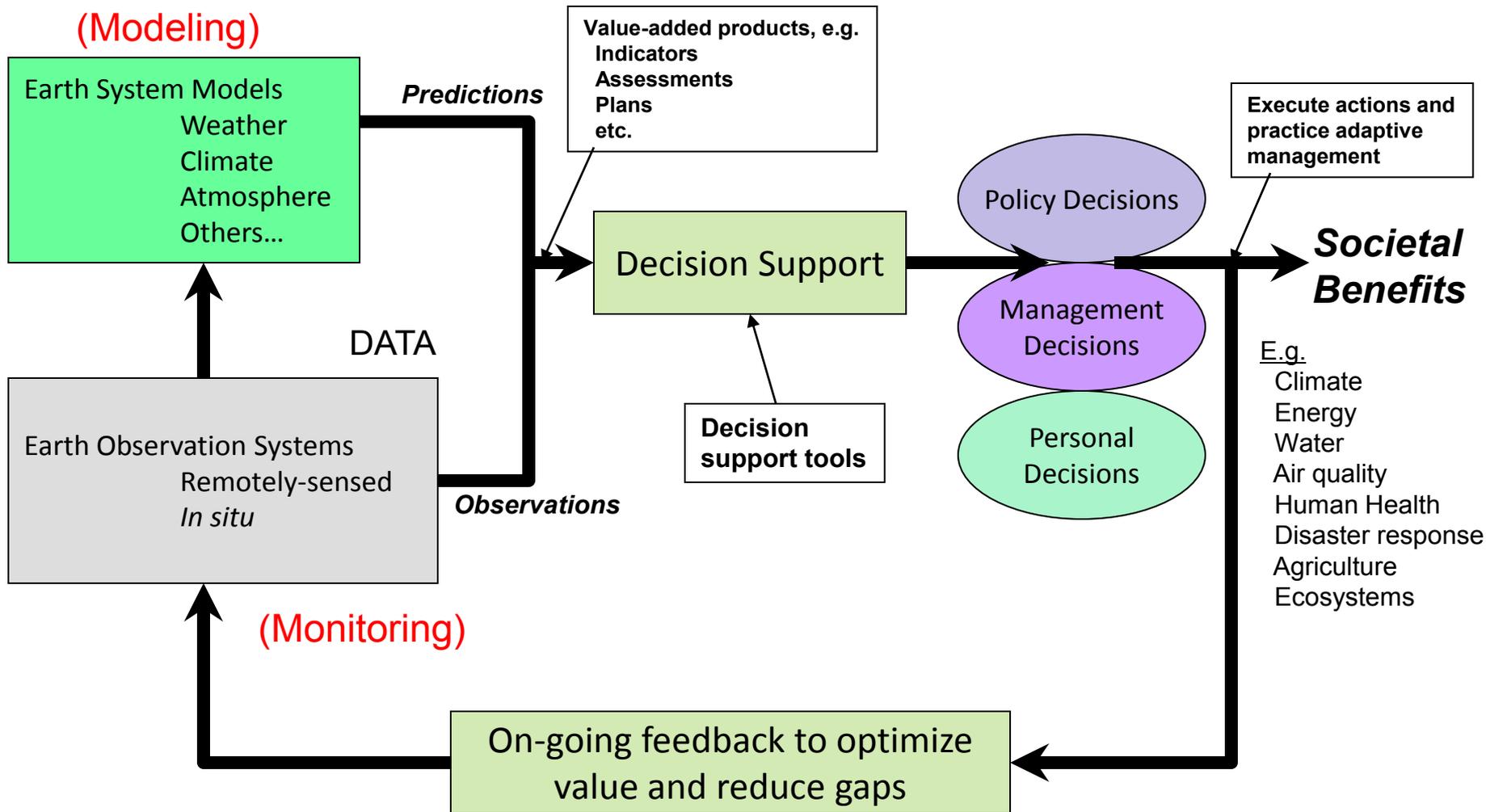
# At the Community Level

- “Community Engagement” – EPA’s Open Government Flagship Initiative
- Building on almost 25 years of “Community Right to Know”
- New users for the Toxics Release Inventory (TRI)
- Driver to think in new ways about how our data can support community efforts

# At the Global Level

- Exciting potential synergies with the international Global Earth Observation System of Systems (GEOSS) program
- Earth observations data – a strategic, high-value data domain with vast additional potential to bring societal benefits
- Open Government and GEOSS – 2 natural partners; enormous synergies

# The GEOSS/IEOS Architecture



Adapted from: *Linking Earth Observations to Societal Benefits*  
(IEOS Strategic Plan, p. 17 Figure 1) (IEOS = Integrated Earth Observation System)

<http://usgeo.gov/images/USGEOMain/EOCStrategicPlan.pdf>

# From Data to Knowledge

*R. Schaefer*

Collaborators: L. Paxton, N. Bos, S. Babin, B. Fortner, J. Holm, D. Morrison, M. Osorno, C. Parker, C. Pikas, D. Simmons, S. Simpkins, S. Strong, B. Swartz, M Weiss,

The logo for Applied Physics Laboratory (APL) consists of the letters 'APL' in a large, bold, blue, sans-serif font.

*The Johns Hopkins University*  
APPLIED PHYSICS LABORATORY

•NASA/JPL, Pasadena, CA

•NC State University, Raleigh NC

# Outline

- **Open and easily browse-able data, mashups, etc. – very good things.**
- **One more step is necessary for many missions – converting data to knowledge to support decision and policy makers.**
- **What does this mean?**
- **Example problems: Climate Change & Space Weather – critical needs**
- **How do we get there? Knowledge management and transfer through Virtual Organizations**

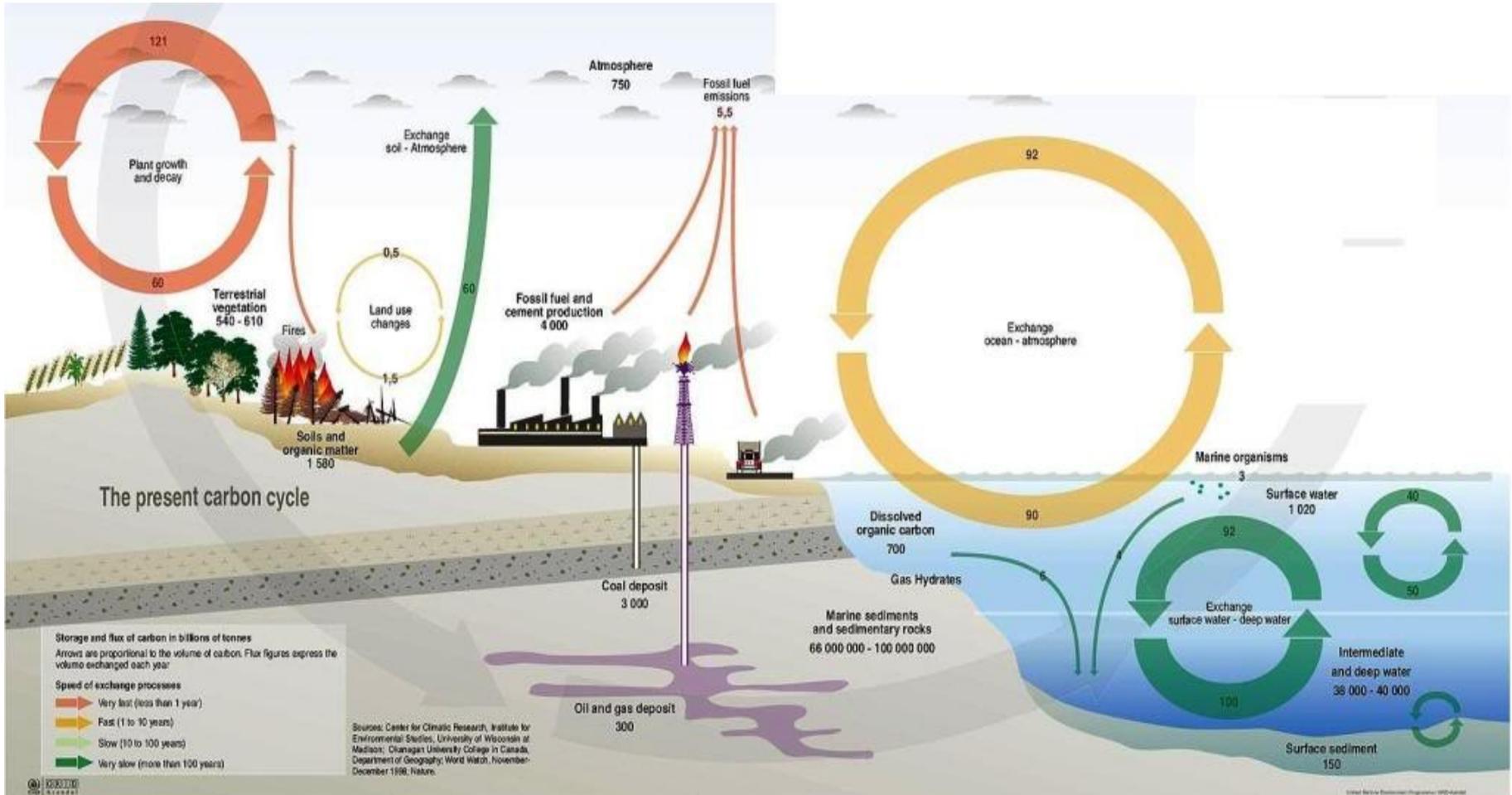
# Data/Model Results Aggregation

- Many examples of good use of data - portals that provide an interface to many datasets that are related by a specific mission.
- This enables discovery of new information through combining multiple perspectives.
- Typically these toolsets are used by analysts and scientists to extract information.
- Knowledge gained from investigations using these mashups needs to get to policy and decision makers.
- This last step can be sped up by building a virtual organization (VO)
  - to tailor and capture the knowledge gained from data analysis
  - To make available to decision and policy makers.

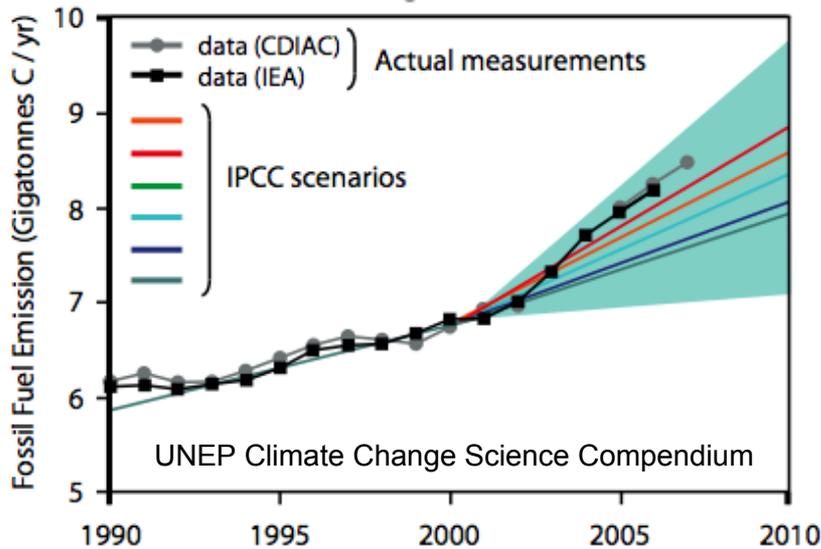
# Two Example Problem Cases:

- The world is facing problems of a global scale that are very challenging:
  - **Dealing with the Impacts of Climate Change** - effects threaten lives, societies, and political stability
  - **Technological Susceptibility to Space Weather** - events can disrupt our increasingly high tech dependent society
- These problems require coordinated action from a variety of agencies and institutions by people who are not experts in space weather or climate change.
- A wealth of data and models exist that can be analyzed by experts to translate the data and model results into knowledge.
- Bringing together data and models through a unified interface is not enough; we must bring together the community: data providers, subject matter experts, scientists, policy analysts, etc. into a virtual organization to get the appropriate knowledge to the people who need it.

# Understanding **Climate Change** Mixes Traditional Earth Science Disciplines



# Climate Change Consequences Are Wide Ranging, Serious, and Urgent



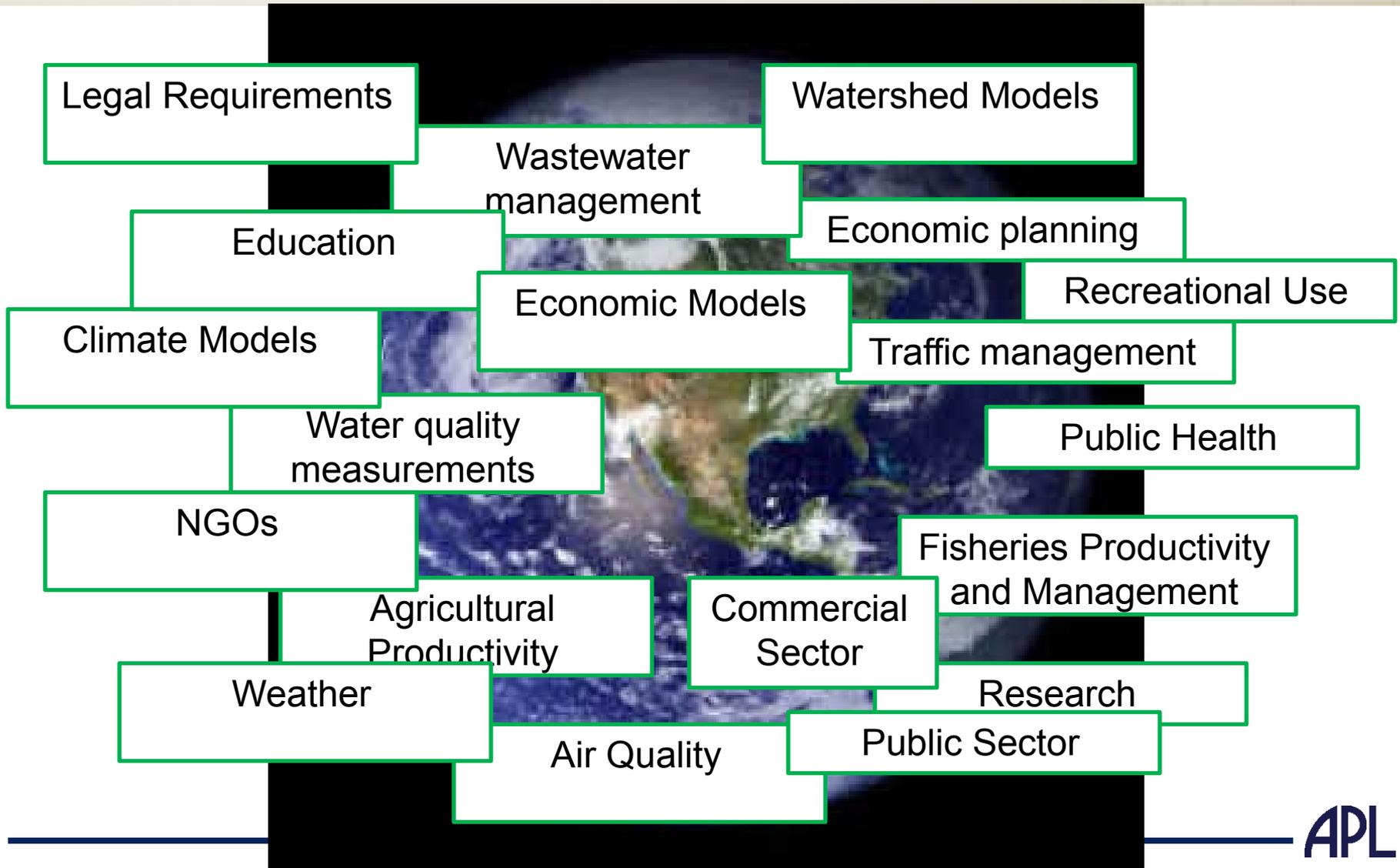
- Current CO<sub>2</sub> emission rate higher than IPCC “worst case” scenario
- Billions of People would experience serious consequences from climate change

*Relative vulnerability of coastal deltas as shown by the indicative population potentially displaced by current sea-level trends to 2050 (Extreme = >1 million; High = 1 million to 50,000; Medium = 50,000 to 5,000; following Ericson et al., 2006). Source: IPCC*

**Governments and Organizations need to re-orient policies and procedures to prepare for this eventuality**



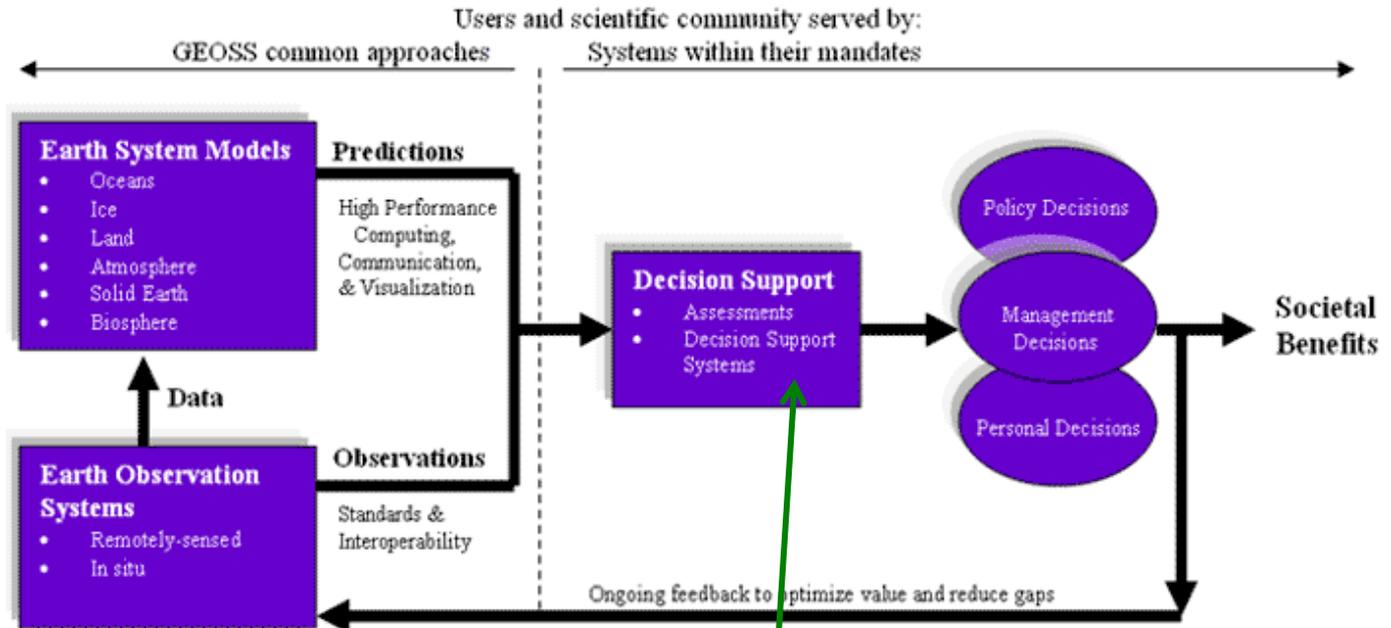
# But Tools for decision makers are fractionated and deciders are isolated within their communities



# Next step after making data accessible is to provide decision support

- Aggregation and dissemination of data and models is needed
- Also need to find a way to translate data into information (knowledge) that can be acted on

Figure from EPA's GEOSS web page



Decision support is a key element that needs attention

*NAS Report on America's Climate Choices: “..information should be tailored to user needs, provided at space and timescales to support decision-making, communicated clearly, and accompanied by decision support tools that allow exploration of alternative pathways.”*

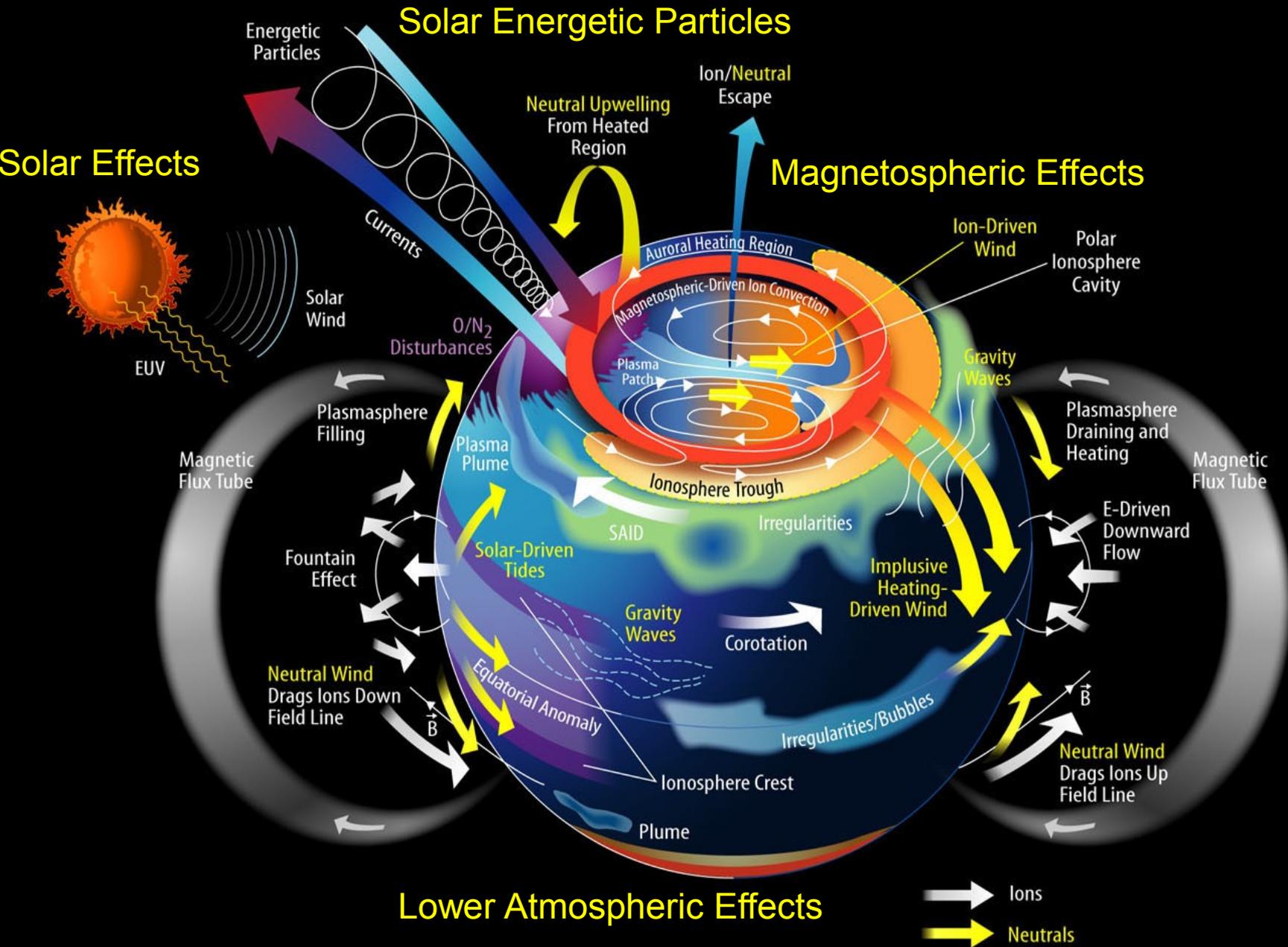
# Towards Supporting Decisions

- **Enabling decision support is a non-trivial task:**
  - **How do we know what information decision makers need?**
  - **How can we determine if we have the proper data to support decision and policy makers?**
  - **How do we create new tools tailored to the needs of policy and decision makers?**
- **Data to may already be available, but the knowledge to interpret and understand the data must be transferred from the subject matter experts to decision makers in order for the data collection effort to be effective.**
- **Need a knowledge management system maintained by an organization – a virtual organization**

# Another Example: Space Weather

- A similar problem exists in the Space Weather impact domain

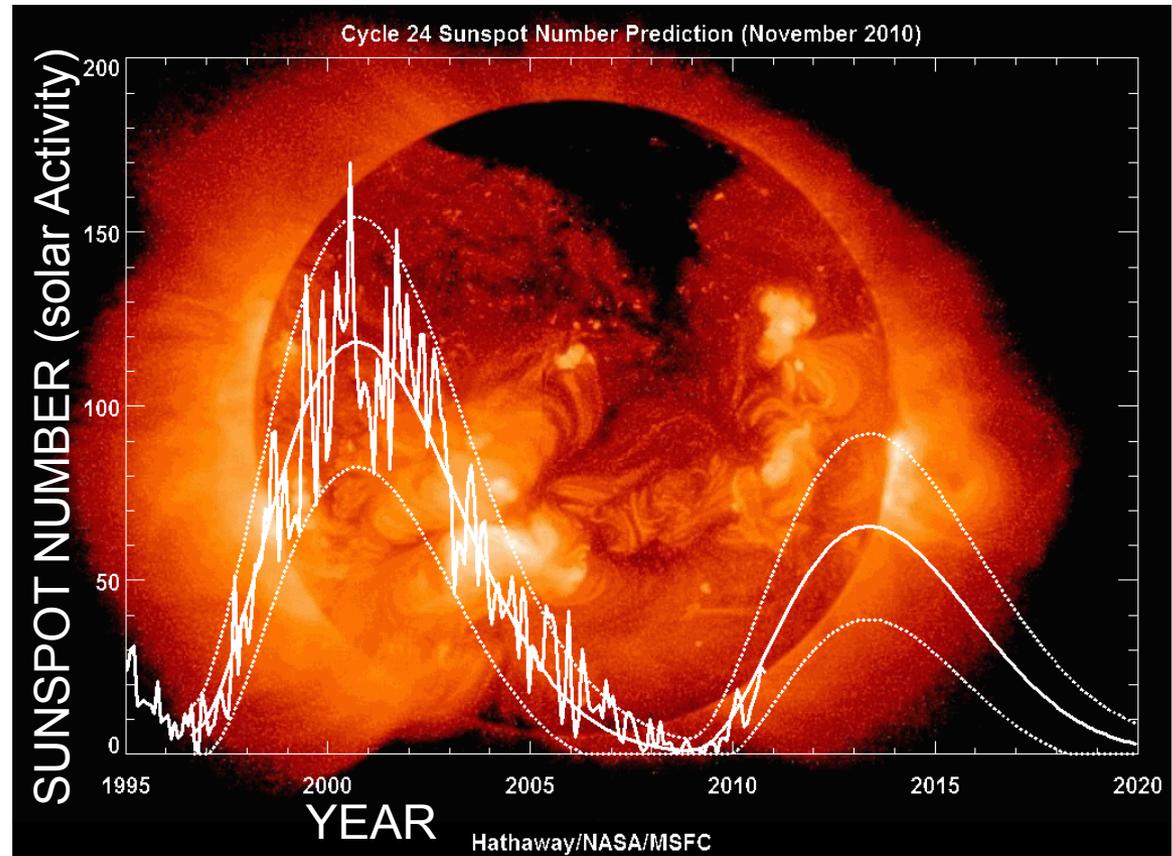
# Space Weather Also Crosses Scientific Disciplines



# Space Weather prediction and monitoring is also an urgent need

- **Reliance on space weather vulnerable technology (particularly GPS) has increased greatly in the last few years during a very quiet period of solar activity (The sun is the driver of space weather).**
- **In a couple of years the sun will be a LOT more active.**

Predicted solar activity level



# Disparate Communities – need custom information

Agencies:  
NASA, FEMA  
NOAA, FAA

Commercial  
Air, Ground

Military

## Space Weather: Common Needs

Utilities  
Internet,  
Power

Homeland  
Security

**MARCH 1989 SOLAR FLARE AND GEOMAGNETIC STORM**

- INTERFERENCE AT SIX MISSILE WARNING RADAR SITES
- OVER 1300 SPACE OBJECTS TEMPORARILY MISPLACED
- WORLDWIDE HF RADIO BLACKOUTS; SATCOM AND AF SATELLITE CONTROL NETWORK INTERFERENCE
- LOST IMAGERY AND COMMUNICATIONS ON GOES-7
- PERMANENT LOSS OF HALF THE COMMAND CIRCUITRY ON A JAPANESE COMMUNICATIONS SATELLITE
- 7 COMMERCIAL SATELLITES REQUIRED 177 MANUAL OPERATOR INTERVENTIONS (MORE THAN USUALLY REQUIRED IN A YEAR)
- MEMORY UPSETS ON TDRSS AND INTELSAT
- 9-HOUR CANADIAN POWER OUTAGE AFFECTED 6 MILLION PEOPLE; POWER GRID COLLAPSE ALMOST SPREAD INTO NORTHEAST US, WHERE IT WOULD HAVE HAD AN ESTIMATED \$3-6 BILLION IMPACT ON THE US ECONOMY
- LORAN NAVIGATION PROBLEMS AND COMPASS ALIGNMENT ERRORS

Scientific  
research

Education

Public

# Need for a VO

- **Decision and policy makers in the various agencies:**
  - do not always know how to interpret data on space weather web sites.
  - Need to transform the data into knowledge relevant to needs of users.
- **Do this by bringing in all parties (data providers, application developers, subject matter experts, end users, decision makers) to:**
  - understand requirements of decision makers and end users
  - create tools tailored to those needs of users.
- **Do this through a virtual organization (VO)**
  - **Use Web technologies to facilitate this**
    - On-line collaboration (wikis, forums, workflow sharing, on-line conference tools)
    - Web enabled data portals with easy data visualization and searching
  - *A viable collaborating VO is non-trivial to establish*

# Elements of a Successful VO

- Work has been done on understanding VOs, see e.g., NSF report: [“Beyond Being There”](#) – report from workshop on building Effective VOs (and references therein)
- Elements that need to be considered
  - Get the community together to define needs and uses of the VO
  - Have the community define most pressing needs and lay out a path to meet those needs
  - Provide easy access to data and model results to enable collaboration and discovery.
  - Bring in social scientists to aid and improve collaboration
  - Bring knowledge management experts to define and improve design of knowledge management system.
  - Create a (web based) collaborative, knowledge management system that will enable users to communicate freely
  - Establish a management plan

# How to enable a VO

- **Identify relevant subject matter experts and end users of knowledge.**
- **Bring them together to identify the most pressing problems – you cannot build it and expect them to come**
- **Identify the datasets that are most relevant to addressing the problems**
- **Enable collaboration to create tools to tailor the information**
- **Capture the knowledge from the subject matter experts into an accessible knowledge management**
- **Provide management and oversight to ensure smooth operation and to do self assessment to continually improve the VO**
- **Bring in social scientists to help enable successful collaboration**

# Two VO concepts under development at JHU/APL

- **SWIFTER-ACTION - Space Weather Informatics, Forecasting, and Technology through Enabling Research - Actionable Content & Timely Information On the Network**
- **GAIA – Global Assimilation of Information for Action this project is now under way:**
  - **Starting with 3 sets of focused workshops to bring communities together to define most pressing needs:**
  - **March 2011 – Climate Change and Energy**
  - **April 2011 – Climate Change and Public Health**
  - **August 2011 – Climate Change and National Security**

# Summary

- **Difficult problems facing society – examples:**
  - **Climate Change will impact the world in ways that will change societies in potentially catastrophic ways – we need to start making policies that mitigate those effects now.**
  - **Heavy dependence on Space Weather vulnerable technology as we approach the maximum solar activity period after a long quiet period**
- **Making data and model results available is not enough for pressing global problems**
- **Need to enable decision and policy makers with actionable information – knowledge**
- **Pathway to this is through a collaborative virtual organization uniting data providers, subject matter experts, end users, and decision makers in a collaborative, discovery enabled knowledge management environment.**



# **International Open Government Data Conference**

0110  
011  
01  
0



# International Open Government Data Conference

## *Break*

## *2:30pm – 2:50pm*

*Sponsored by*







# **International Open Government Data Conference**

0110  
0111  
0111  
0111  
0111



# International Open Government Data Conference

## Putting Data to Work—Creating Value-Added Functionality and Applications Through Mashups

**Dominic James DiFranzo**, *Graduate Student, Computer Science,  
Rensselaer Polytechnic Institute*

**George Komatsoulis**, *Deputy Director of Center for Biomedical  
Informatics and Information Technology, NCI, NIH*

**Ryan McKeel**, *Digital Assets Applications Engineer, U.S. National  
Renewable Energy Laboratory*

**Susan Turnbull**, *Senior Program Advisor, Office of Citizen Services  
and Innovative Technologies (Moderator)*

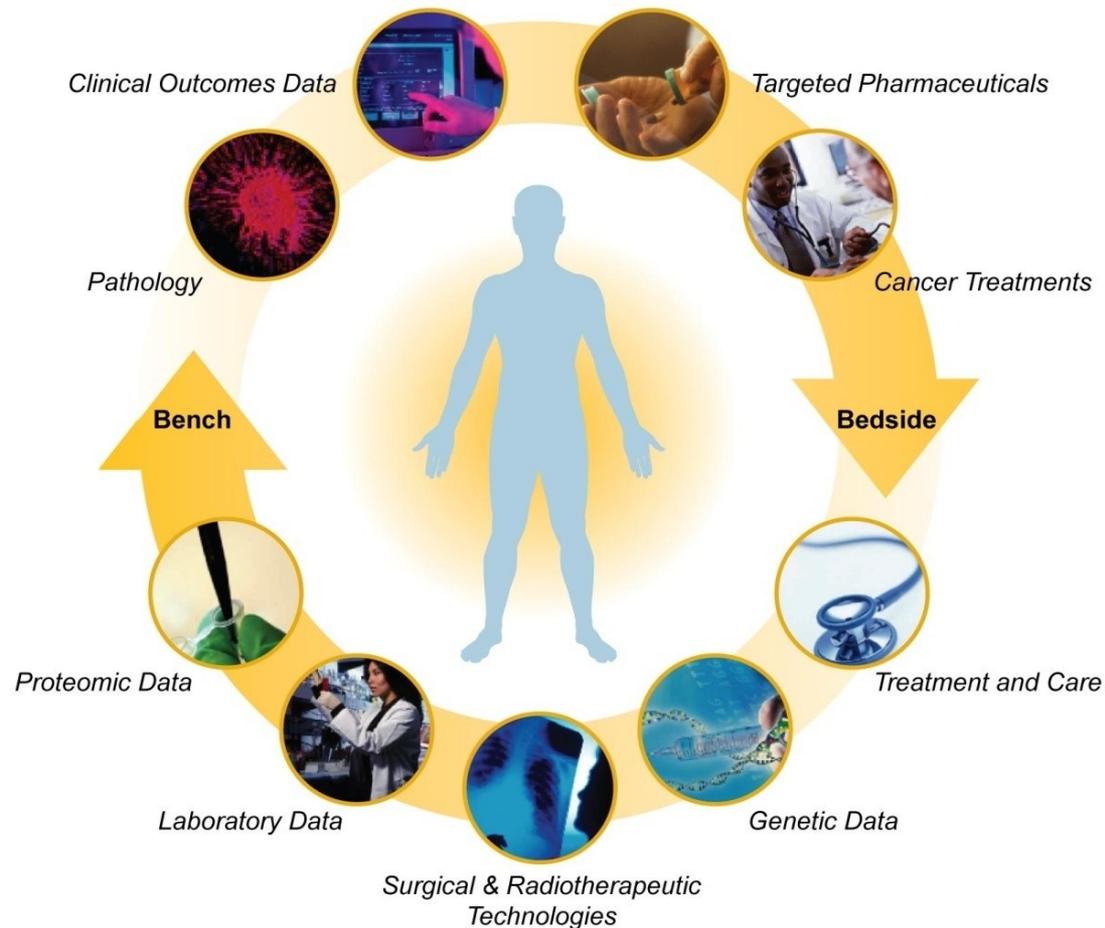
# The cancer Biomedical Informatics Grid

International Open Government Data Conference  
November 15, 2010

*George A. Komatsoulis, Ph.D.*  
*Deputy Director, Center for Bioinformatics and Information Technology, NCI*

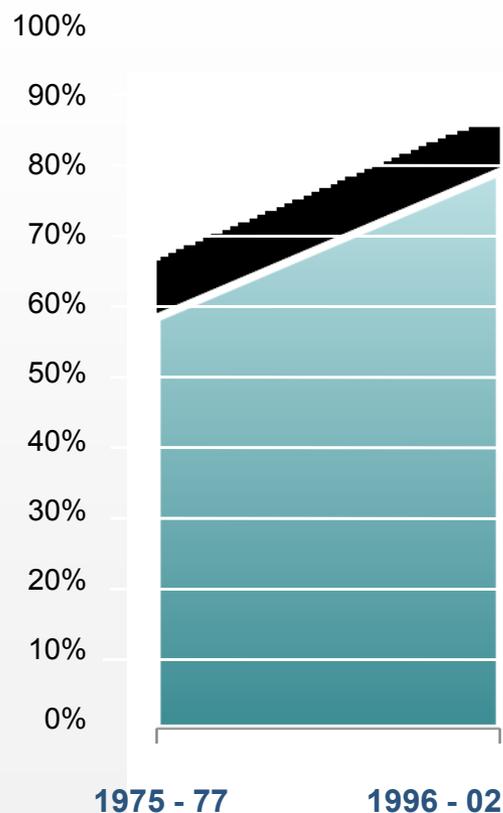
# 21<sup>st</sup> Century Biomedical Research and Care

- Personalized, Predictive, Preemptive, Participatory.....
- Unifies discovery, clinical research, and clinical care, (bench-bedside-bed) into a seamless continuum
- Results in improved clinical outcomes
- Accelerates the time from discovery to patient benefit
- Enables a health care system, not a disparate “sector”
- Empowers consumers in managing health over a lifetime



# Molecular Medicine and Childhood ALL

**Increased 5-year relative survival rate among children 10-14 years old diagnosed with ALL**



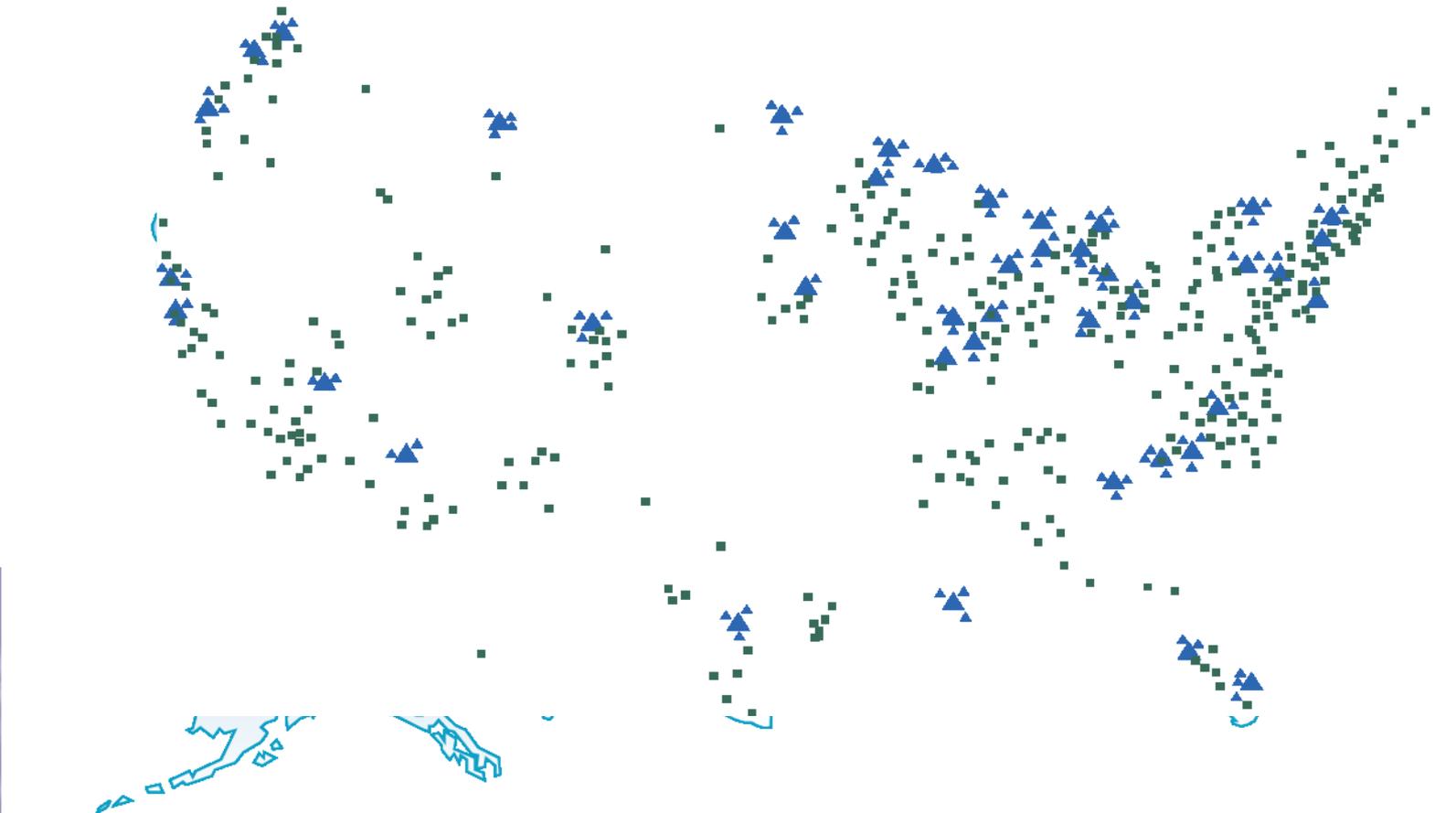
**Childhood cancer treatment now blends care delivery and clinical research. Most children with cancer are enrolled in clinical trials. Using this approach:**

- Researchers and practitioners are able to correlate experimental laboratory data with clinical data such as treatment, history, pathology and patient outcomes
- Clinical data are used to continuously evaluate outcomes
- Researchers can develop and refine evidence-based strategies at an individualized level
- Care providers improve quality by adherence to care standards

# Linking the NCI Supported Cancer Community

■ - NCI Desi

▲ - Community Oncology Practices



# caBIG®: Creating a Worldwide Web of Cancer Research

**caBIG®** is a virtual network of interconnected data, individuals, and organizations that redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise.



# caBIG<sup>®</sup> Core Principles

- **Open Access** – caBIG<sup>®</sup> is open to all, enabling wide-spread access to tools, data, and infrastructure
- **Open Development** – Planning, testing, validation, and deployment of caBIG<sup>®</sup> tools and infrastructure are open to the entire research community
- **Open Source** – The underlying software code of caBIG<sup>®</sup> tools is available for use and modification
- **Federation** – Resources can be controlled locally, or integrated across multiple sites



# caBIG<sup>®</sup> Strategy

- **Community**
  - Establish an **open community** of participants from the spectrum of disciplines, geographies, types of institutions, etc.
  - Facilitate the work of others who are building capabilities
  - Adopt a “**federated**” model to allow local control of sharing and partnerships and to support individual labs and institutions
- **Content**
  - Facilitate access to rich primary data
  - Leverage existing academic and commercial software, wherever possible, to avoid unnecessary time and expense
  - Invest primarily in **open source** tools that the community does not have
- **Connectivity**
  - Recognize legacy IT systems to avoid “rip and replace” costs
  - Wherever feasible, make disparate applications compatible for “plug-and-play” compatibility and data-sharing through **standards-based interoperable infrastructure**

# caBIG<sup>®</sup> Operational Approach

## Domain-level

Clinical  
Trials  
Management  
Systems  
Workspace  
(CTMS)

Integrative  
Cancer  
Research  
Workspace  
(ICR)

*In Vivo*  
Imaging  
Workspace  
(IMAG)

Tissue  
Banks &  
Pathology  
Tools  
Workspace  
(TBPT)

## Strategic-level

Strategic  
Planning  
Workspace  
(SP)

Training  
Workspace  
(D&T)

Data Sharing  
& Intellectual  
Capital  
Workspace  
(DSIC)

## Cross-cutting

caBIG<sup>®</sup> Vocabularies and Common Data Elements Workspace (VCDE)

caBIG<sup>®</sup> Architecture Workspace (ARCH)

# Interoperable Applications to Support Biomedical Research

- Track clinical trial registrations
- Facilitate automatic capture of clinical laboratory data
- Manage reports describing adverse events during clinical trials



**Clinical Research**

- Combine proteomics, gene expression, and other basic research data
- Submit and annotate microarray data
- Integrate microarray data from multiple manufacturers and permit analysis and visualization of data



**Molecular Biology**



**Imaging**

- Utilize the National Biomedical Imaging Archive repository for medical images including CAT scans and MRIs
- Visualize images using DICOM-compliant tools
- Annotated Images with distributed tools

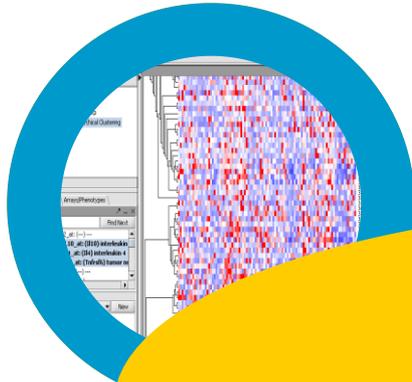


**Pathology**

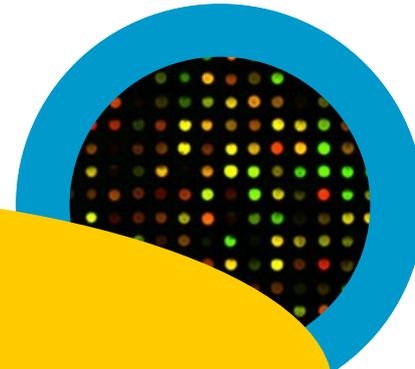
- Access a library of well characterized, clinically annotated biospecimens
- Use tools to keep an inventory of a user's own samples
- Track the storage, distribution, and quality assurance of specimens

# Supporting Individual and Institutional Needs

## Analytical Tools



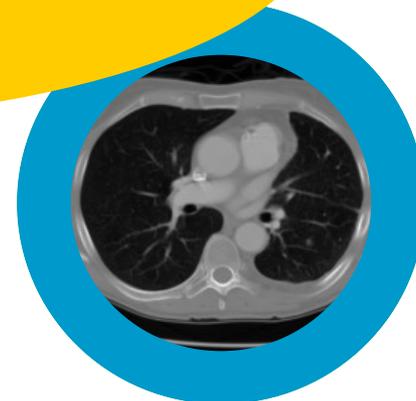
## Molecular Analysis



Bios



als



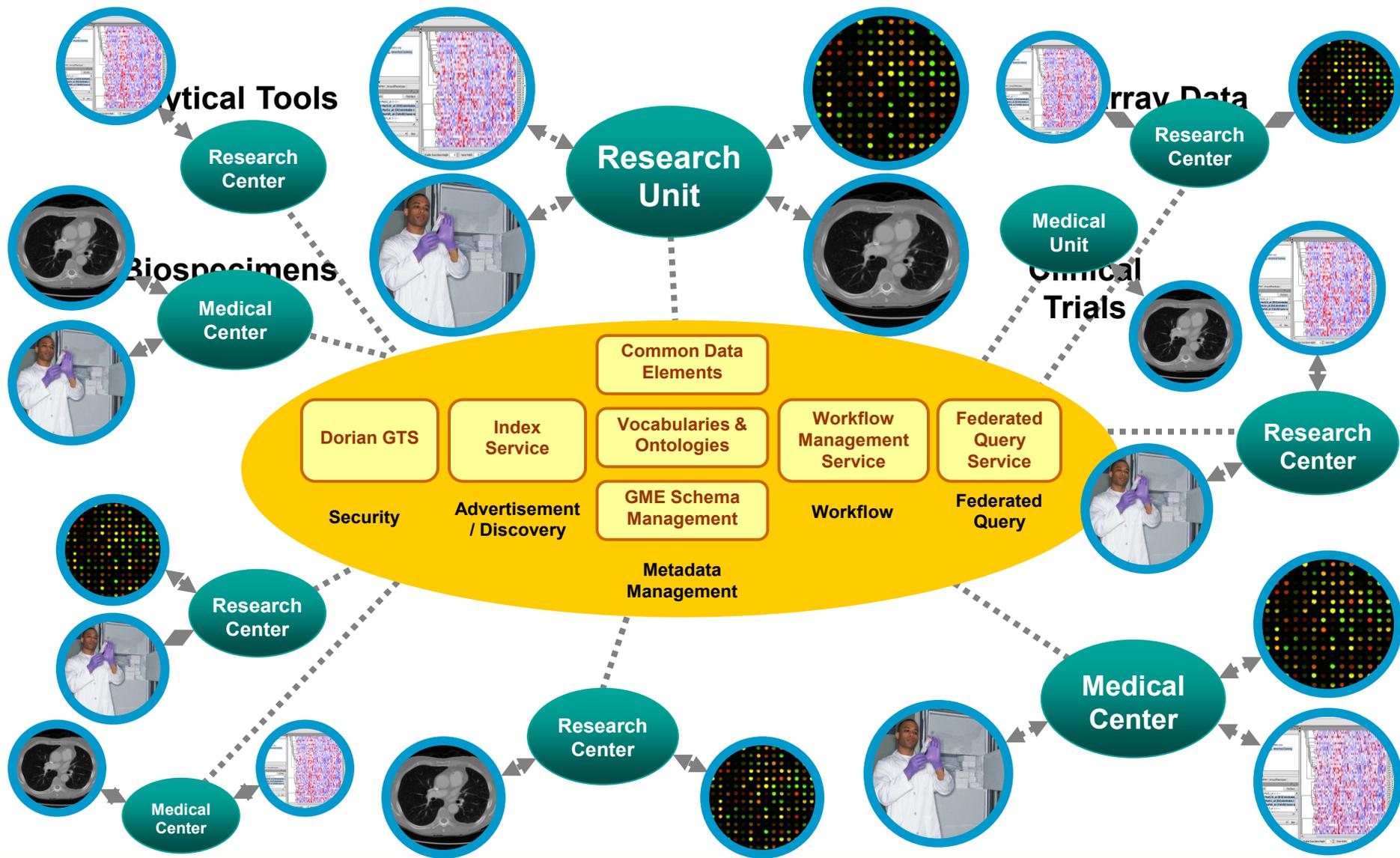
Common Data  
Elements

Vocabularies &  
Ontologies

Information  
Models

# caBIG Integrates Organizational IT Infrastructure

- **caGrid Technology can be used locally to integrate Biomedical IT elements within an institution, examples include:**
  - **Ohio State University:** Using caGrid to create the TRIAD system to support clinical research integration
  - **Washington University:** Integrating clinical trials, EHR and biospecimen resources
  - **University of Alabama, Birmingham:** Using caGrid as their internal integration technology
  - **University of Arkansas Medical School:** Integrated clinical trials information system leveraging caGrid
  - **Aga Khan University, Karachi, Pakistan**



# NCI Hosted caGrid Deployment



## caGrid in use outside of caBIG®

- **A variety of organizations are adopting or adapting caGrid to build interoperability Infrastructures. Examples include:**
  - **National Cancer Research Institute (NCRI) UK:** Using caGrid as the basis for their ONIX infrastructure
  - **NHLBI CardioVascular Research Grid (CVRG):** NCI caGrid and CVRG have begun cross-indexing services to allow access to capabilities available on both Grids
  - **Clinical Translational Science Awards (CTSA):** Ohio State University implementing caGrid for CTSA awardees
  - **Centre for the Development of Advanced Computing (CDAC), Ministry of Information Technology, Government of India:** Implementing caGrid to bring Indian research organizations into the community of cancer research.
- **caGrid Modular Infrastructure allows the creation of a “Grid of Grids”**

# caBIG<sup>®</sup> at a glance

## Community

- 2,300+ participants from more than 700 institutions
  - 56 NCI-designated Centers
  - **30** Community Centers
- **1000+ registrants** for the **2010** caBIG<sup>®</sup> Annual Meeting
- **19** licensed Support Service Providers to sustain the biomedical community as they deploy caBIG<sup>®</sup> tools and technology
- 15 countries using caBIG<sup>®</sup> tools and technology to facilitate

## Connectivity

- **78** applications supporting full continuum of biomedical research
- **149** “nodes” connected to National Grid via caGrid

## Content

- **2.17** million biospecimens available through caGrid
- **4.76** million images stored in the National Biomedical Imaging Archive
- **39,952** microarray experiments available for research use on caGrid

# Expanding Across Sectors and Internationally

## 15 countries engaged with and/or using caBIG® tools and technologies, including:

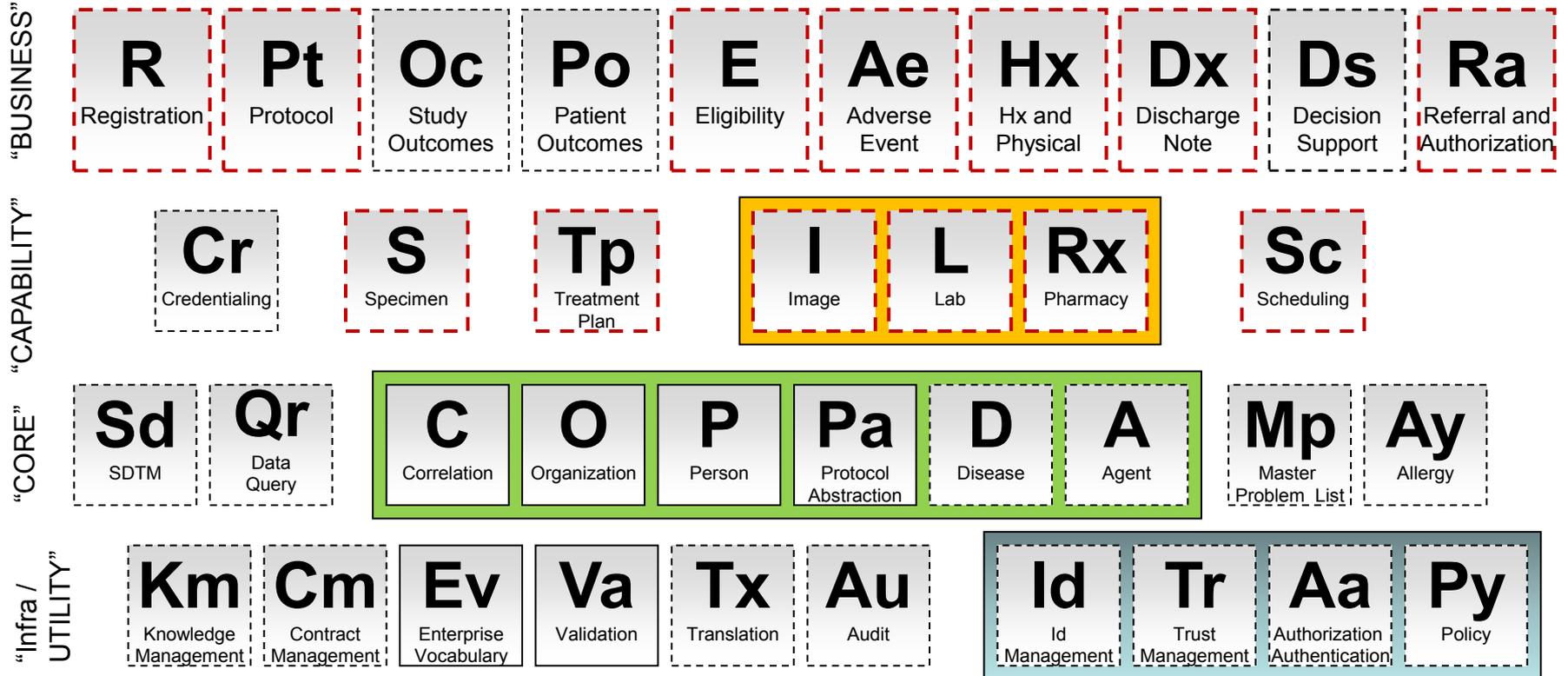
- United Kingdom
- Latin America
- India
- China
- Mexico, Brazil, Uruguay, Argentina, Chile
- Czech Republic
- The Netherlands
- Germany
- Finland
- Jordan
- Pakistan
- Australia
- New Zealand



- **United Kingdom:** NCI collaboration with National Cancer Research Initiative (NCRI) focused on the use of caGrid technology to connect researchers and enable exchange of research data.
- **Latin America:** Latin American Cancer Pilot Program began patient enrollment for two clinical trials on molecularly characterized stage II and III breast cancer patients, using a broad suite of caBIG® tools to ensure that researchers can compare data across partner sites.
- **China:** Duke University Comprehensive Cancer Center and Beijing University Cancer Hospital launched a collaboration using caBIG® capabilities to conduct first clinical trials in China where all patients are registered electronically.

**India:** NCI engaged with the All India Institute of Medical Sciences, the Center for Development of Advanced Computing, and the Tata Memorial Hospital of Mumbai over the use of grid computing for managing clinical trials data.

# Enterprise Services for Research and Care



## Next Generation: caBIG<sup>®</sup> 2.0

**caBIG<sup>®</sup> capabilities are evolving to support the rapidly changing needs of Cancer Centers as they move to EHRs and genomically-guided medicine:**

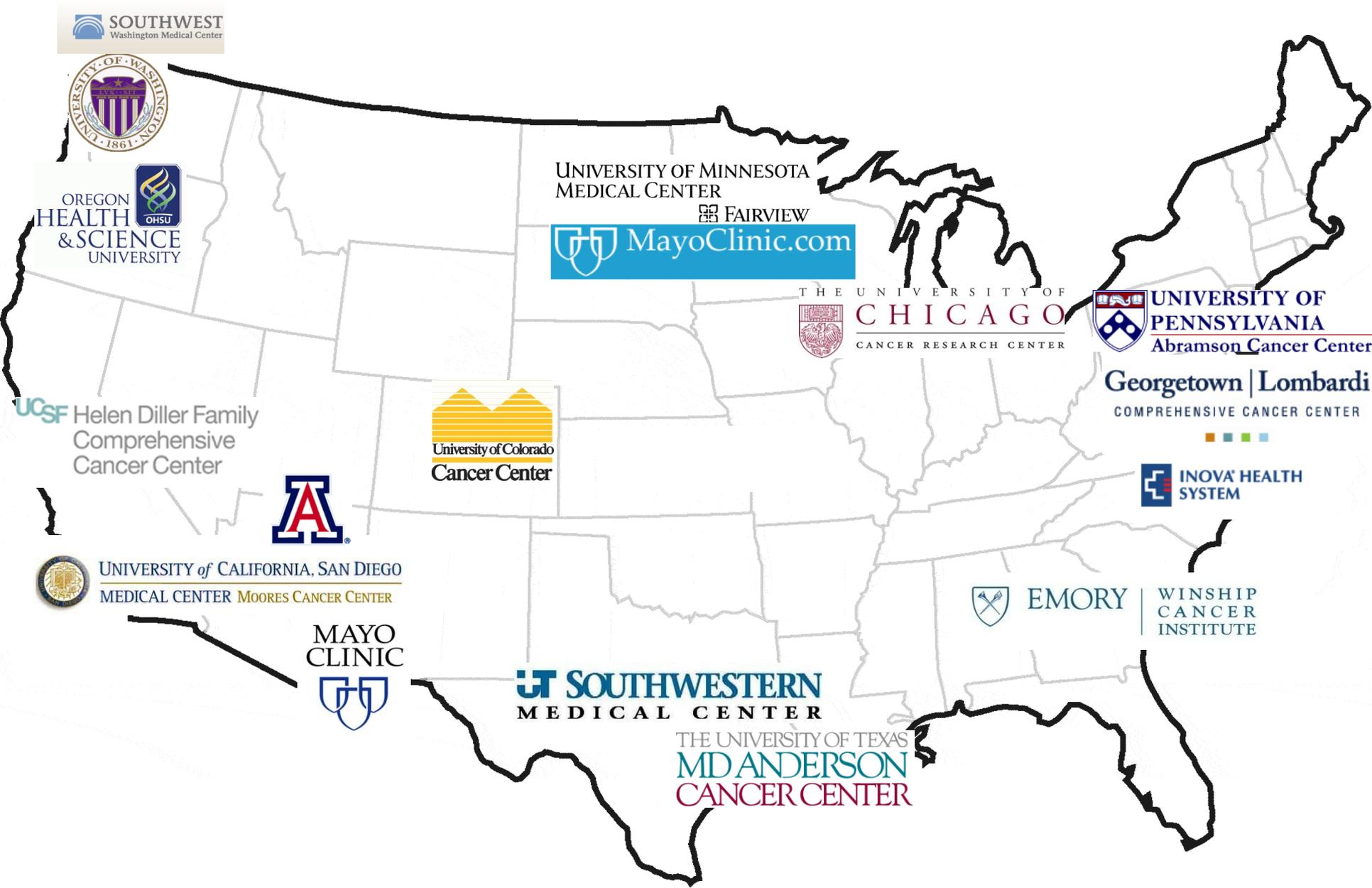
- Collaborate with ASCO on caEHRs
- Capitalize on tech advances: cloud computing, Service-Oriented Architecture, mobile devices
- Lower barriers: make it easier and easier for all stakeholders to use caBIG<sup>®</sup> research capabilities
- Leverage increasingly mature collection of publicly-available open source infrastructure

**The I-SPY TRIAL (Investigation of  
Serial studies to Predict Your  
Therapeutic Response with  
Imaging And molecular analysis):**

A national study to leverage biomarkers  
in predicting response to combinatorial therapy for  
women with Stage 3 breast cancer.

*(PI Laura Esserman, UCSF )*

# Projected I-SPY 2 study sites



# I-SPY Adaptive Trial Outline



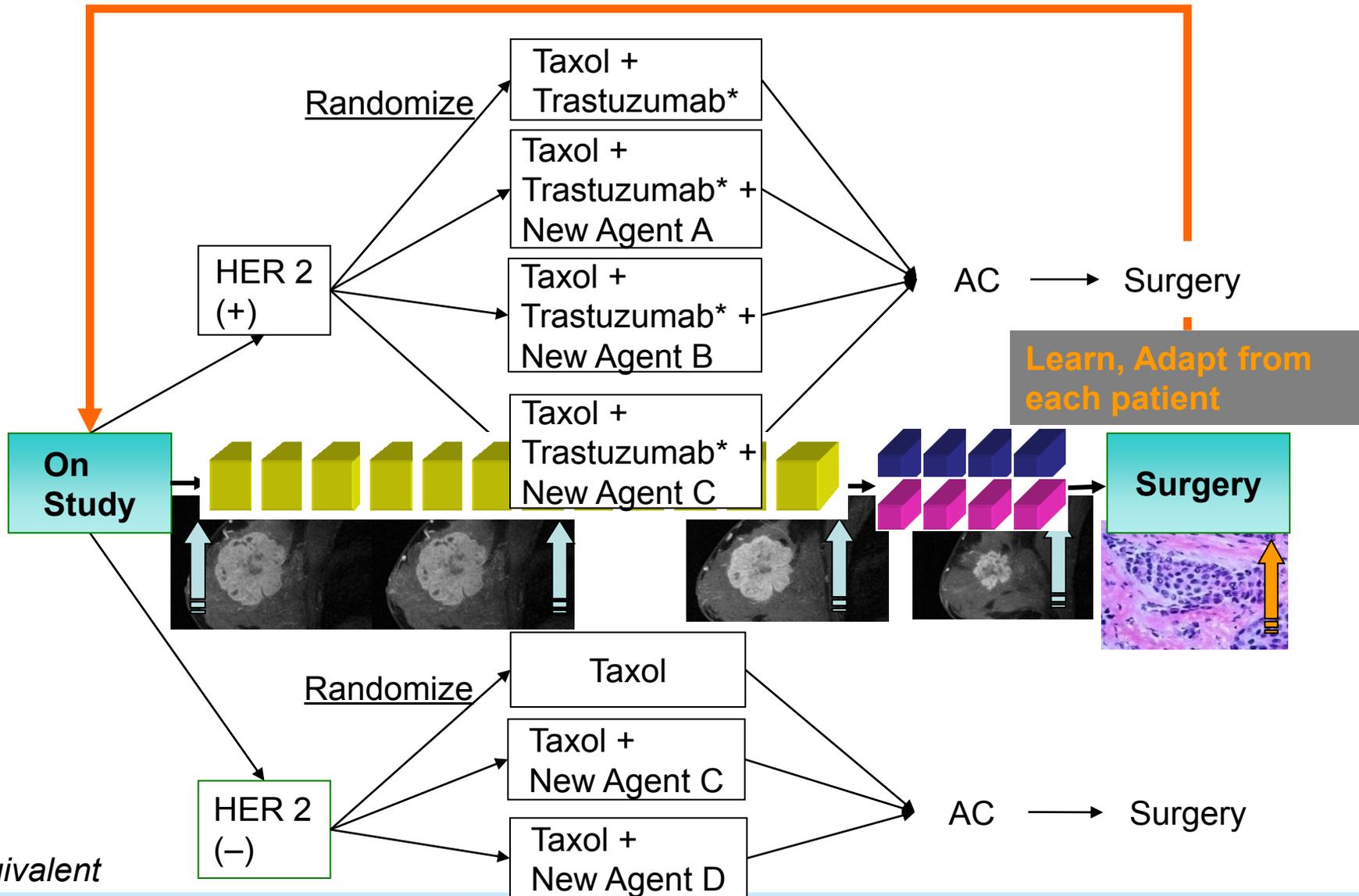
Accrual: Anticipate 800 patients over 3–4 years

Enroll ~20 patients per month

Participating Sites: 15–20 across US and Canada

# I-SPY Adaptive Trial:

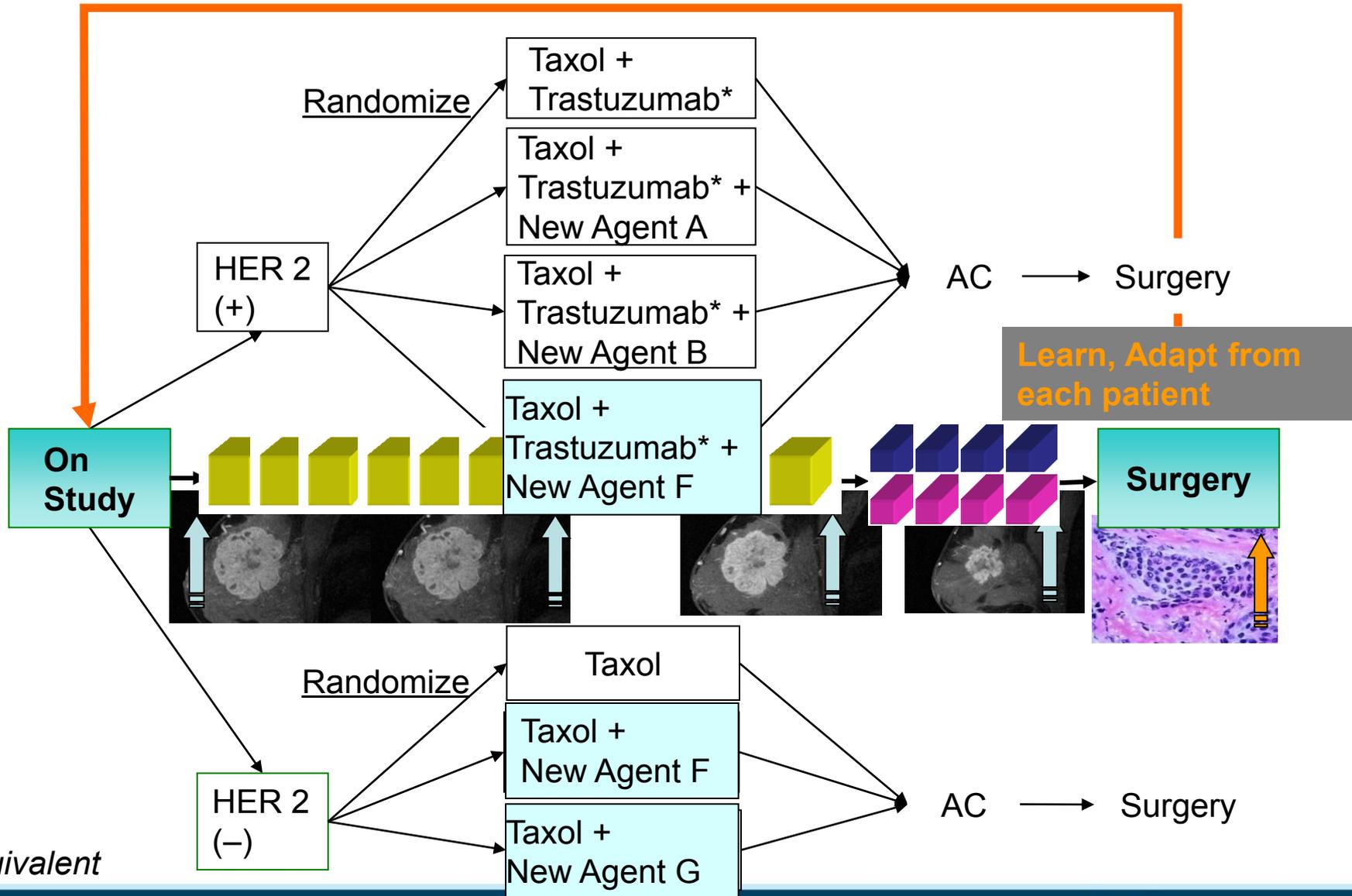
Introduce several new agents for a given profile



\*Or Equivalent

# I-SPY Adaptive Trial:

Introduce several new agents for a given profile



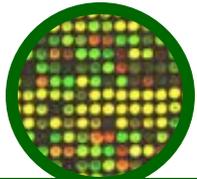
\*Or Equivalent

# I-SPY TRIAL IT Infrastructure

## Research Environment



SNP Array Data



Expression Array Data



Patient Samples

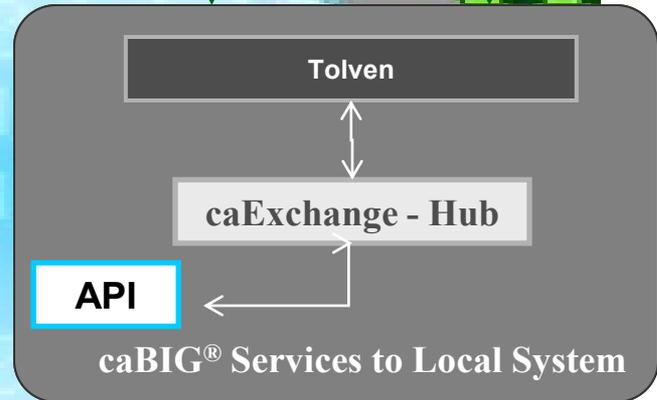
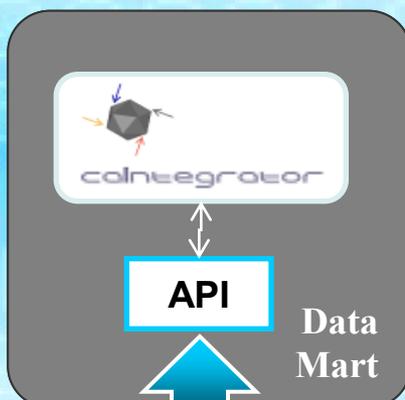
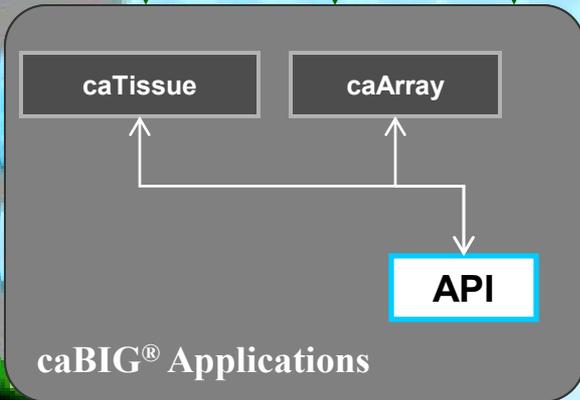
## Clinical Care Environment



Clinical Data



Radiological Data



**For more information, please visit:**

**<http://caBIG.cancer.gov>**

**<http://www.bighealthconsortium.org>**

# Thank You



# **International Open Government Data Conference**