

Leveraging Cloud-Based Solutions to Address High Demand, Large Scale Surveys

R. Mark Adams, Ph.D.
Booz Allen Hamilton

FedCASIC
Washington, DC
March 2011



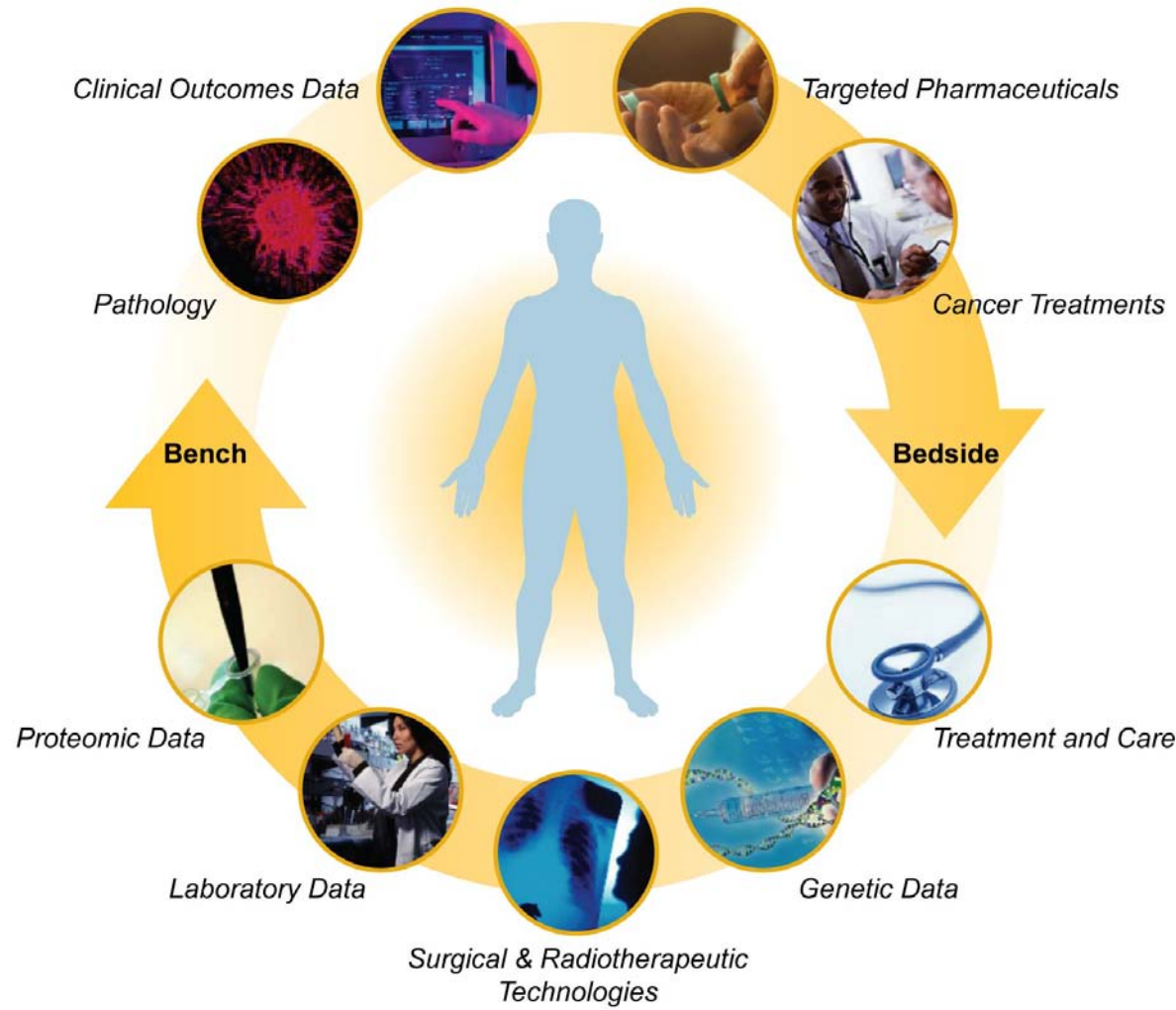
Acknowledgements in advance

- **In keeping with the caBIG philosophy of community engagement, many, many academic departments, companies, NIH/NCI divisions and individuals have contributed to the work being shown. The specific acknowledgements are in the back of the deck, and apologies to any of the literally thousands of active contributors and participants that I missed.**

From Bench to Bedside

Personalized Medicine

- Applies genomic, molecular, and clinical data to target the delivery of health care.
- Allow treatments to be tailored based on the genetic make up of the tumor.
- Focuses on variations, to better match patients with the therapy most effective for them.



caBIG[®] Vision and Goals

caBIG[®] Vision

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[®] Goals

- Connect the cancer research community through a shareable, interoperable infrastructure
- Deploy and extend standard rules and a common language to more easily share information
- Build or adapt tools for collecting, analyzing, integrating, and disseminating information associated with cancer research and care

IT-Enabled Research

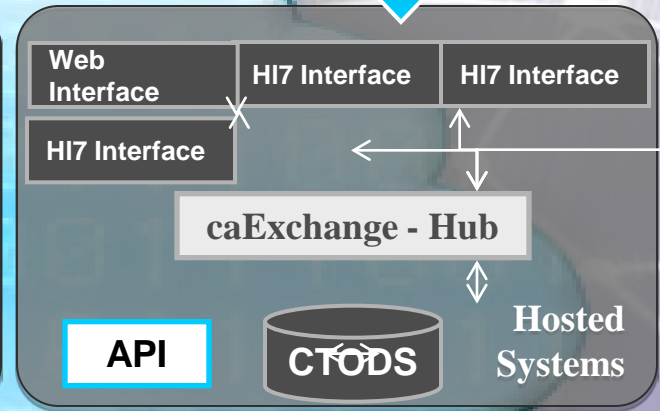
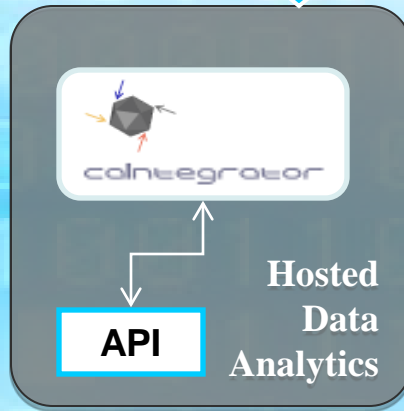
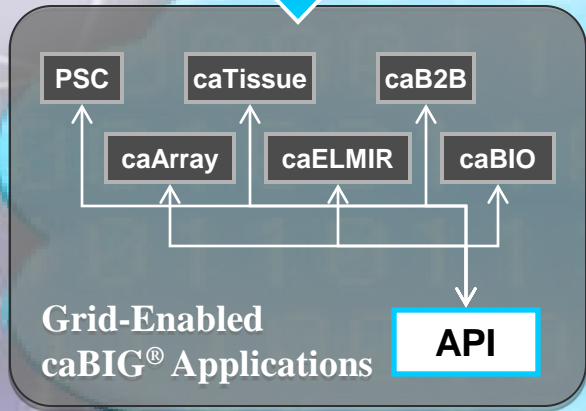
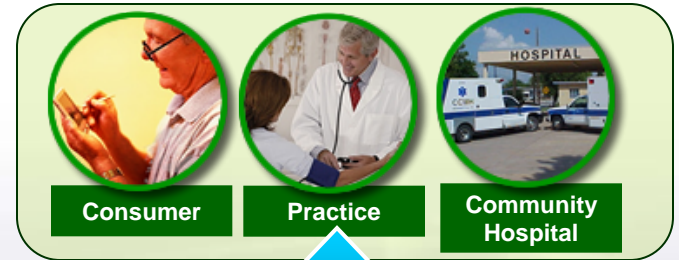
Molecular and Clinical Research Information



Clinical and Genomic Research



Producers of Clinical Information



caGRID

caBIG[®] Capabilities: Population Science

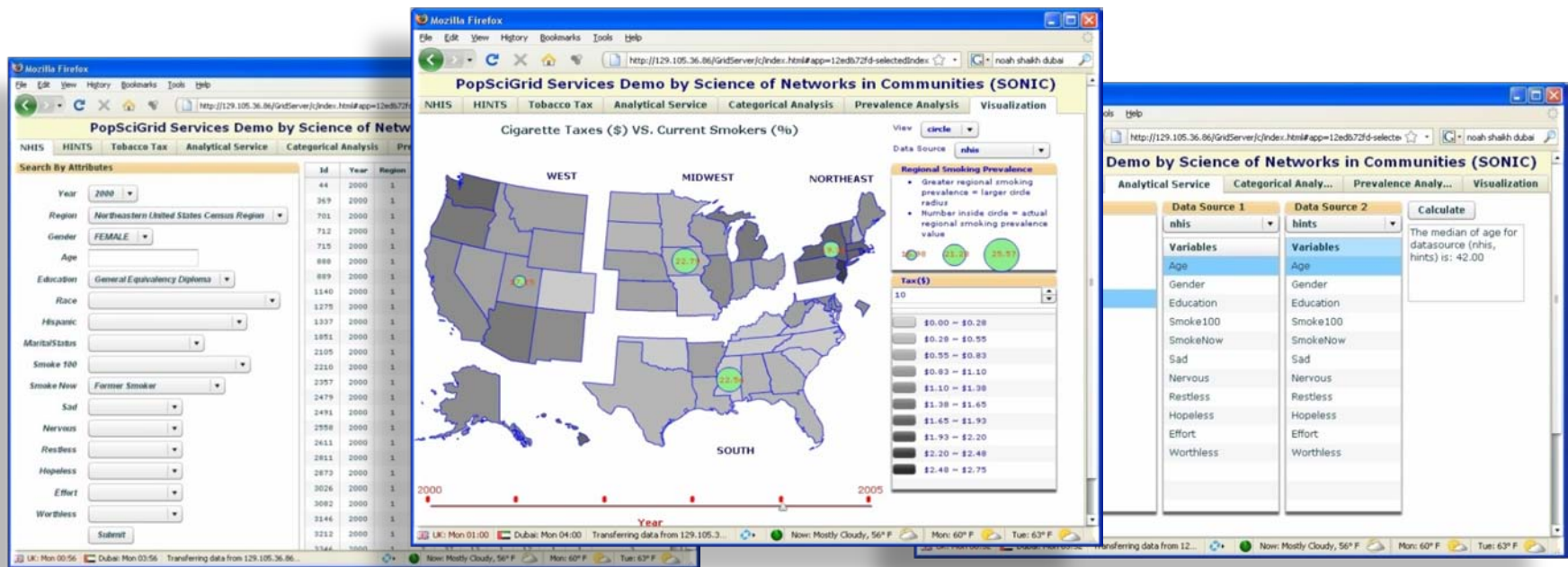
Population Sciences SIG Vision

- **Vision**

- The Population Sciences SIG in caBIG[®] is the vehicle for bringing together population scientists and informaticists in order to discover together how the caBIG[®] infrastructure and tools can advance population health as well as bringing population scientists into the dialogue on the subject of clinical and translational cancer research.

PopSciGrid 1.0: <http://ciknow2.northwestern.edu/GridServer/c/index.html>

- Aggregate analysis of 14 datasets spanning 6 years (HINTS / NHIS tobacco tax data)
- Datasets located at multiple geographically dispersed sites
- Real-time access/analysis of public health and economic data
- Prospective geospatial analytics



Shaikh AR, Contractor N, Moser R, et al.. PopSciGrid: Using cyberinfrastructure to enable data harmonization, collaboration, and advanced computation of nationally representative data. American Public Health Association Annual Meeting; November 11, 2009; Philadelphia, PA.

PopSciGrid 2.0 (Coming online 2011)



Widgets



Application Layer (e.g., Enhanced disease modeling, dashboards, data widgets...)



Grid Cyberinfrastructure



Common Vocabularies (Shared ontologies, common data elements)



Public Surveillance

- BRFSS
- HINTS
- NHIS
- Tax
- US Census,...

Grantees

- CECCRS
- CPHHD
- GEI
- TREC
- TTURCS

Clinical/Health System

- CRN
- QCCC projects
- Registries (SEER)

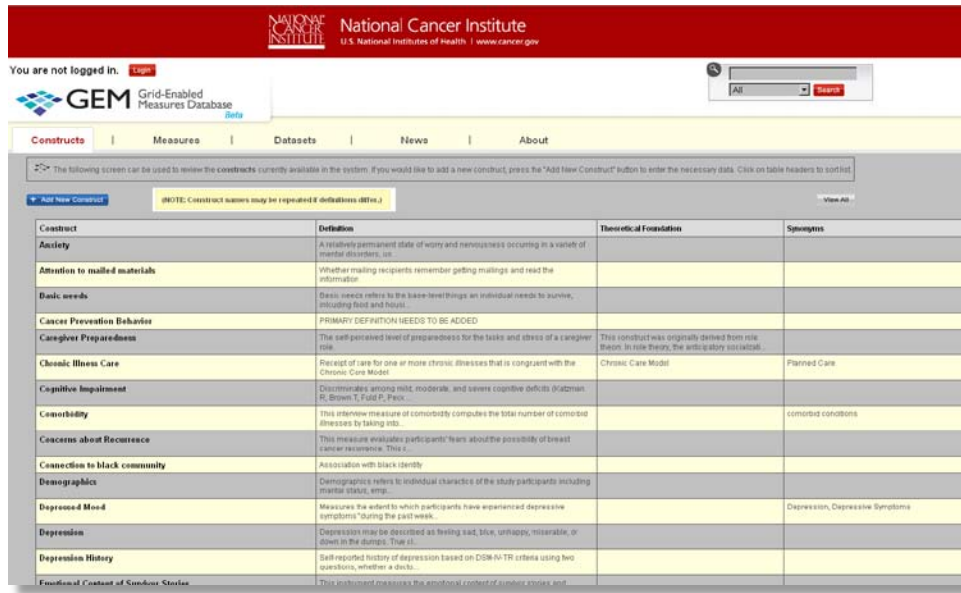
Biomedical

- Biological
- Genomic/proteomic

Community/Contextual

- 'Community health labs'
- GIS (geo-spatial data)
- Physical/Built environment
- Real-time data capture

Grid Enabled Measures (GEM): Science 2.0



Construct	Definition	Theoretical Foundation	Synonyms
Anxiety	A relatively permanent state of worry and nervousness occurring in a variety of mental disorders, etc.		
Attention to mailed materials	Whether mailing recipients remember getting mailings and read the information.		
Basic needs	Basic needs refers to the basic requirements an individual needs to survive, including food and housing.		
Cancer Prevention Behavior	PRIMARY DEFINITION: NEEDS TO BE ADOPTED		
Caregiver Preparedness	The self-perceived level of preparedness for the tasks and stress of a caregiver role.	This construct was originally derived from role theory. In role theory, the anticipated social role.	
Chronic Illness Care	Receipt of care for one or more chronic illnesses that is congruent with the Chronic Care Model.	Chronic Care Model	Planned Care
Cognitive Impairment	Discriminates among mild, moderate, and severe cognitive deficits (Fuldman, R, Brown, T, Fuld, P, Peck).		
Comorbidity	This interview measure of comorbidity computes the total number of comorbid illnesses by taking into.		comorbid conditions
Concerns about Recurrence	This measure evaluates participants' fears about the possibility of breast cancer recurrence. This is.		
Connection to black community	Association with black identity		
Demographics	Demographics refers to individual characteristics of the study participants including marital status, etc.		
Depressed Mood	Measures the extent to which participants have experienced depressive symptoms during the past week.		Depression, Depressive Symptoms
Depression	Depression may be described as feeling sad, blue, unhappy, miserable, or down in the dumps. That is.		
Depression History	Self-reported history of depression based on DSM-IV-TR criteria using two questions, whether a doctor.		
Functional Context of Construct Studies	This construct is used to track the functional context of construct studies.		

- Architecture for participation
- Data driven decisions
- Wisdom of the masses

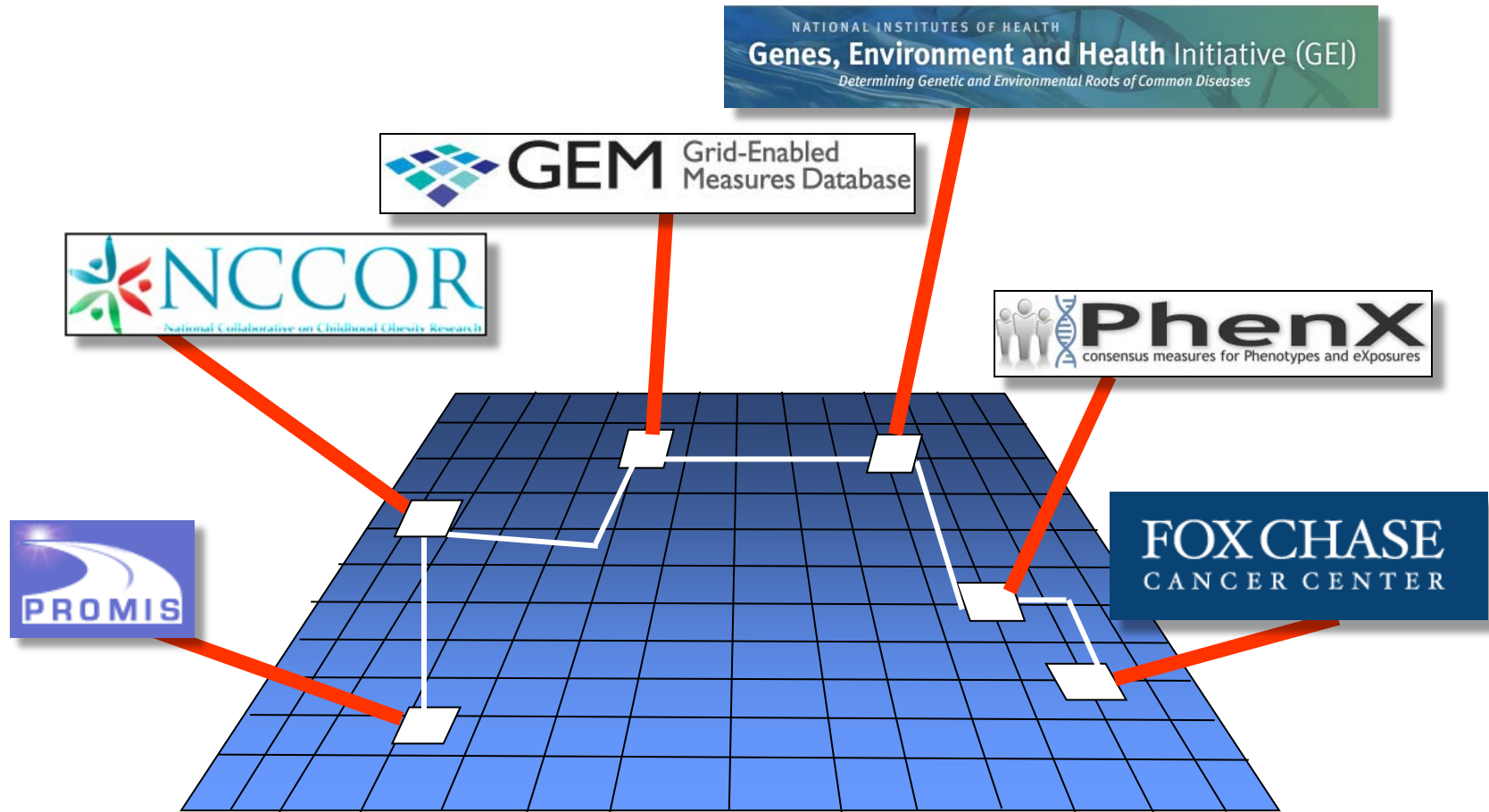


Overall Goals:

To facilitate a virtual community of scientists using collaborative web technology to:

- vet and promote the use of standardized measures— based on theoretically-meaningful constructs;
- share the resulting harmonized data.

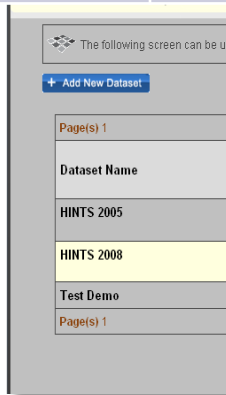
The Power of the Grid: Connecting Data Systems



Dataset Tab: Use Case Scenario

ID	VarX	VarY
101	2.3	1
102	1.2	1
103	4.3	

ID	VarX	VarY
201	3.5	0
202	2.5	0
203	1.8	0



101
102
103
201
202
203



Current State of GEM

Constructs: 73

Measures: 110

Registered Users: 204



How Can You Be Involved?

- Identify, provide meta-data for constructs (domains) for use by others
- Identify, upload, rate measures for use in studies
 - Templates available
- Data sharing capabilities
- GEM wants your input!
 - Email: moserr@mail.nih.gov



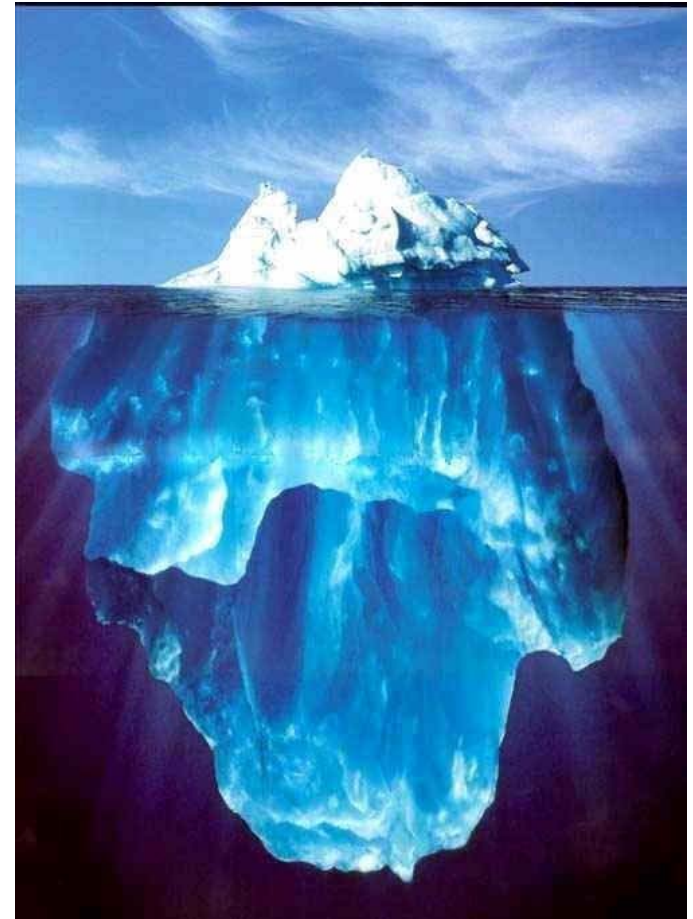
Cloud Computing Overview

“Cloud-computing will help to optimize the Federal data facility environment and create a platform to provide services to a broader audience of customers.”

President's Budget for FY 2010
Section 9, Cross Cutting Programs

Nearly every federal agency is either moving toward adopting cloud computing or thinking about how to do it

- **GSA**
- **Department of the Interior**
 - National Business Center
 - Bureau of Land Management
- **NASA**
- **NOAA**
- **DISA**
- **NSA**
- **FBI**
- **And on, and on...and on**



The issue...



Picture from Chris Dagdigan, BioTeam

A working definition of Cloud Computing for us:

Cloud Computing is...

- **A system that can distributed across multiple locations, or run on a single platform**
- **Tools/infrastructure that can provide access at multiple logical levels to the software and underlying services**
- **A stack that allows access, management, integration and provisioning of services transparently between systems local, remote and distributed**
- **Software that can be flexibly accessed from multiple remote locations using installation-free thin clients**
- **Infrastructure that promotes access via software-as-service to tools and to underlying data via data-as-service to a wide range of users both human and machine**

Cloud Computing is not a panacea for IT – some capabilities are not right for the cloud

- **Cloud Computing Readiness Assessment & Transition Models – not everything belongs in the cloud:**
 - Real-time applications, Command & Control systems are probably not cloud candidates
- **Legal & Regulatory Challenges:**
 - HIPPA, SOX, privacy issues with physical data located outside US
- **Multi-Tenant Environment:**
 - All resources are shared (disk storage, network) causes difficulty cleaning ‘spills’
- **Security**
 - Can security actually be *increased* by moving to a cloud?
 - What new security vulnerabilities are introduced?
- **Maturity**
 - Cloud computing is proven at Google, Microsoft, Amazon, using proprietary technology –the open-source cloud technologies continue to mature

Love Avon Health of Women Study

caBIG[®] and the Love Army of Women

caBIG[®] is partnering with the Love/Avon Army of Women to build a **consumer-controlled** online cohort of **one million women**, called the Health of Women (HOW) Study



AOW is Speeding National Research Investigations...

- **Sister Study: NIH/NIEH study to determine how environment and genes affect breast cancer risk**
 - 5,000 women needed... **54,411 total responses from AOW**
 - 1600 recruited in 48 hours
 - 2300 recruited in 2 weeks
- **Variations in Health Needs of Breast Cancer Survivors: NIH-funded study to reduce disparity in cancer burden due to sexual orientation**
 - 100 lesbian, bisexual women required... **15,412 responses from AOW**
 - 158 recruited

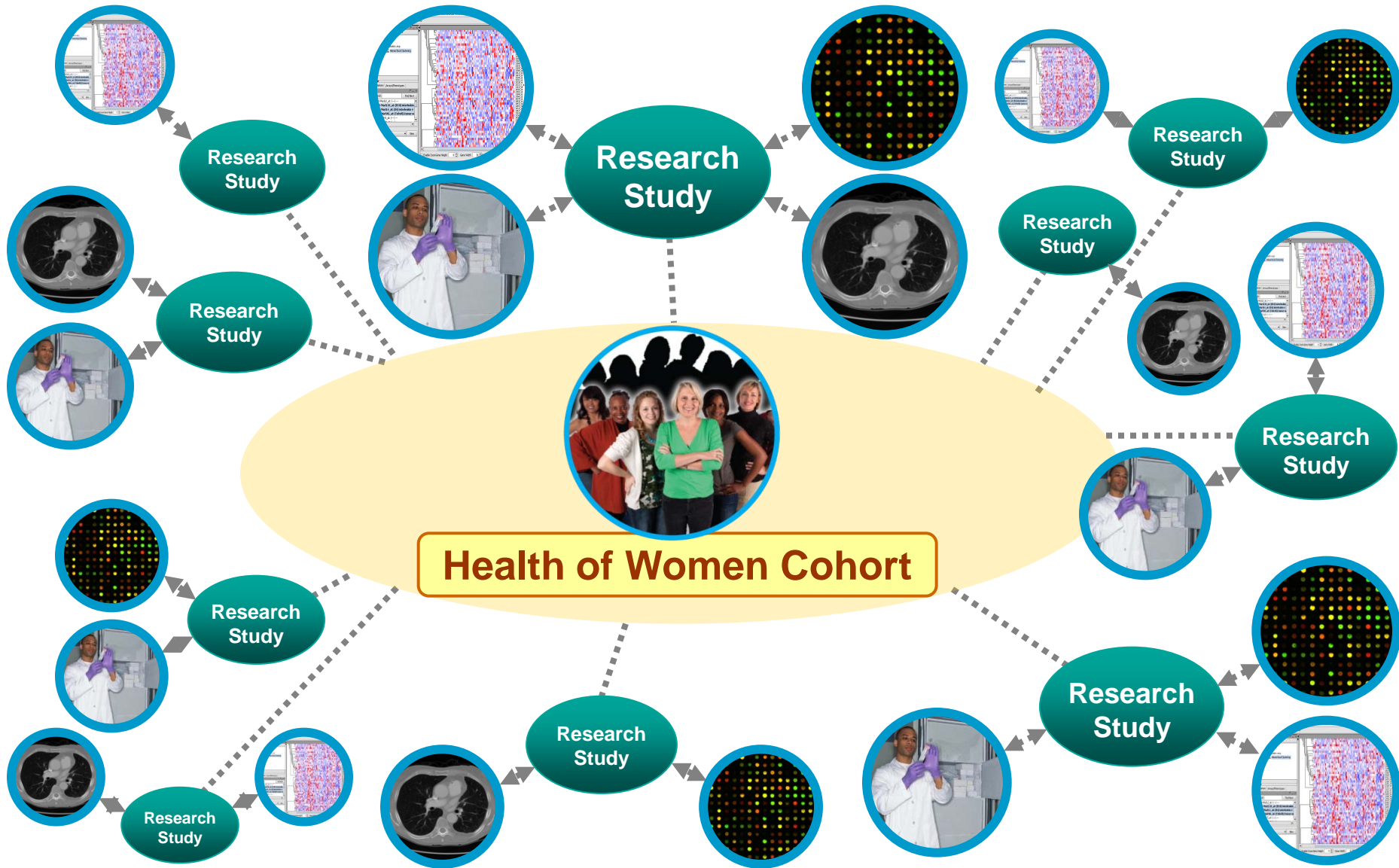
And Local Studies...

- **BEAM Study: Breast Estrogen and Methylation investigation at Northwestern University and Johns Hopkins University**
 - 300 women required... **10,617 responses from the AOW**
 - 23 at NU in 16 months
 - 34 at JHU in 5 months
- **Obesity Study: Controlled diet and exercise study at UCLA monitoring markers of risk for breast cancer**
 - 20 participants needed... **16,466 total responses from the AOW**
 - 125 recruited
 - Fully enrolled in **12 hours**

With Targeted Populations

- **Milk Study: Biomarker study recruiting lactating women scheduled for a breast biopsy; conducted at UMass, funded by Avon**
 - 250 women needed... **62,826 total responses from AOW**
 - 31 recruited in 24 hours... *representing a 6 month acceleration in recruitment*
 - 324 recruited to date
 - Study enlarged to 1000 based on response

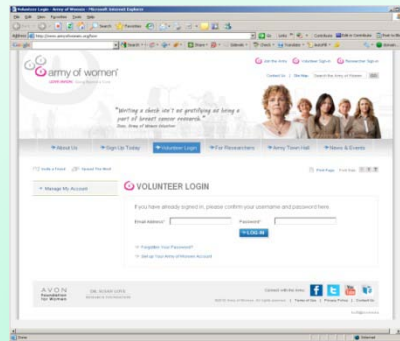
Army of Women “Ecosystem”



Registration and consent workflow

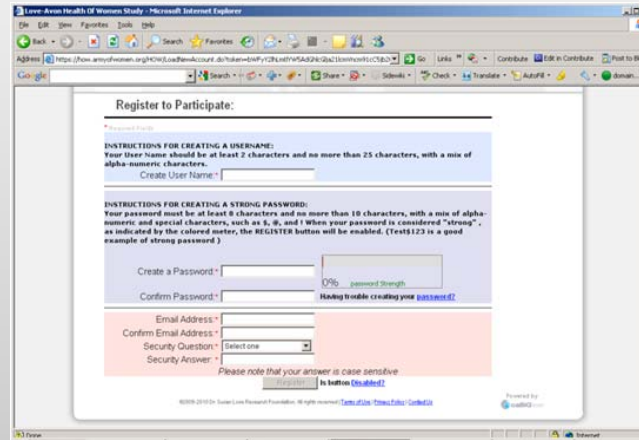
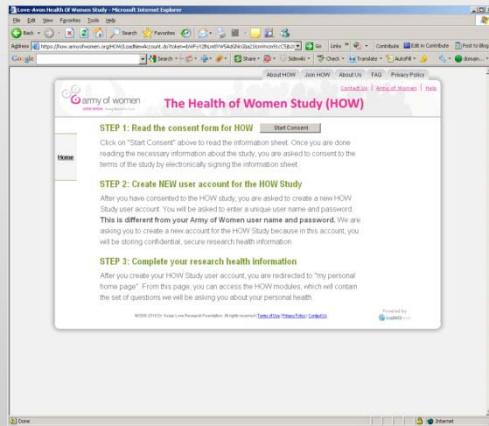


Participants first register to the Army of Women



Next, participants login to Army Of Women Site

 HealthCare IT, Inc.



On to the questionnaire

After login, users are 'passed' to the Health of Women site for consent, registration and completion of the questionnaire.

The Forms:

Love-Avon Health Of Women Study - Microsoft Internet Explorer

Address <https://how.armyofwomen.org/HOW/PatientSaveReproductiveHistory.do>

army of women
LOVE-AVON Giving Women a Choice

The Health of Women Study (HOW)

About Me My Reproductive Health **My Health as of Today**

MY HEALTH AS OF TODAY [View / Print](#)

Please note we will be asking more detailed questions related to this topic in future modules

Compared to a year ago, how would you rate your health, in general now? (Select only one.)

- Much better than a year ago
- Somewhat better now than one year ago
- About the same
- Somewhat worse now than a year ago
- Much worse than a year ago
- I'd rather not say

The following questions are about activities you might do during a typical day. Does your health limit you in these activities? If so, how much? (Select one response for each item.)

Activity	Yes, limited a lot	Yes, limited a little	No, not limited at all	I'd rather not say
Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or golf	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lifting or carrying groceries	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Climbing several flights of stairs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Climbing one flight of stairs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bending, kneeling, stooping	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Walking more than one mile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Walking several blocks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Walking one block	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bathing and dressing yourself	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Enter your height and weight

Height: Feet: Inches:

Weight: pounds

Have you ever used any tobacco products?

- Yes, previously

The Questionnaire

The questionnaire is divided into 4 sections:

About me

The 'About me' section includes a navigation menu with 'About me' and 'My Health as of Today'. The main content area contains several form fields: First Name, Middle Initial, Last Name, Street Address, City, State (dropdown), Zip, Phone Number, E-mail Address, and Date of Birth. Below these fields are three checkboxes: 'Check the box that best describes your highest level of completed schooling' (with options: Less than high school, Vocational school, High school, Some College, 4 year college, Graduate/professional school, No answer not sure), 'What is your occupation today?', and 'What is your ethnic background? Check all that apply' (with options: White/Caucasian, Black/African American, American Indian or Alaska Native, Asian/Pacific Islander).

My Reproductive Health

The 'My Reproductive Health' section asks 'Please rate me for using more detailed questions related to this topic in future modules'. It includes a question 'How old were you when you had your first menstrual period?' with radio button options from 10 to 17, plus '17 years old or older', 'I don't know', and 'No answer not sure'. It also asks 'Have you ever taken birth control pills (oral contraceptives) for one month or longer?' and 'Have you ever taken fertility drugs?'. A table of activities follows, with columns for 'Yes, I do', 'No', and 'No answer not sure'. The table lists activities like 'Vigorous activities, such as running, bicycling, sports', 'Moderate activities, such as walking & swimming', 'Light or sedentary activities', 'Drinking one 8-oz. glass of alcohol', 'Smoking, drinking, driving', 'Walking more than one mile', 'Walking one block', and 'Walking and driving around'. The section concludes with 'Enter your height and weight' and 'Have you ever used any tobacco products?'.

My Health as of Today

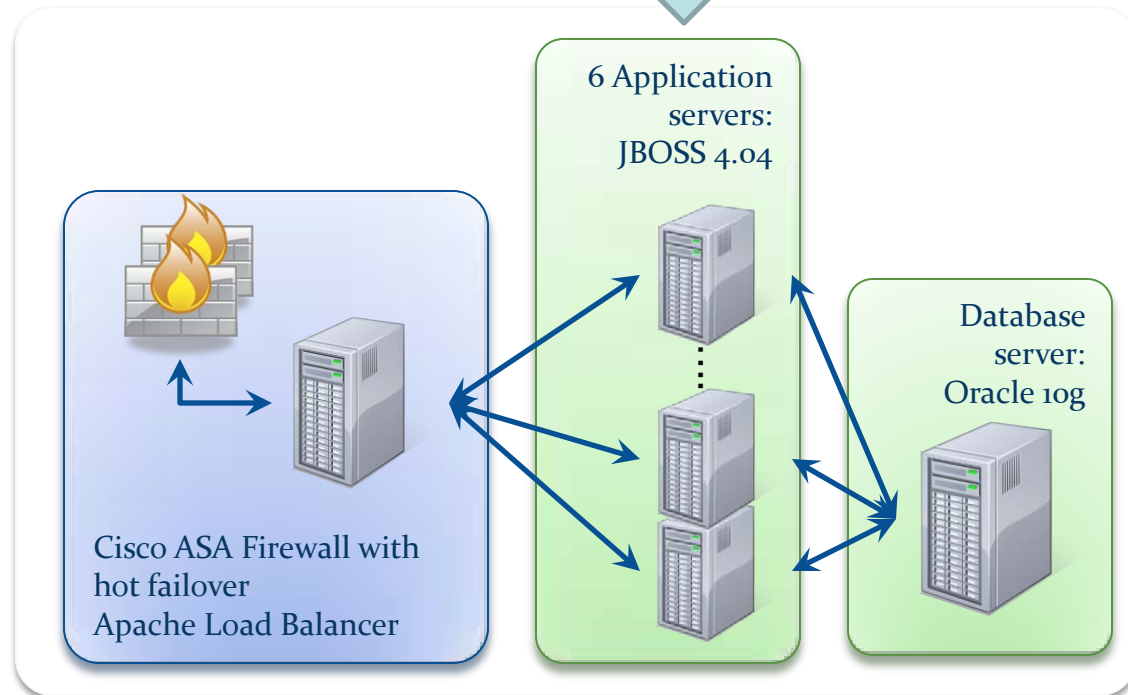
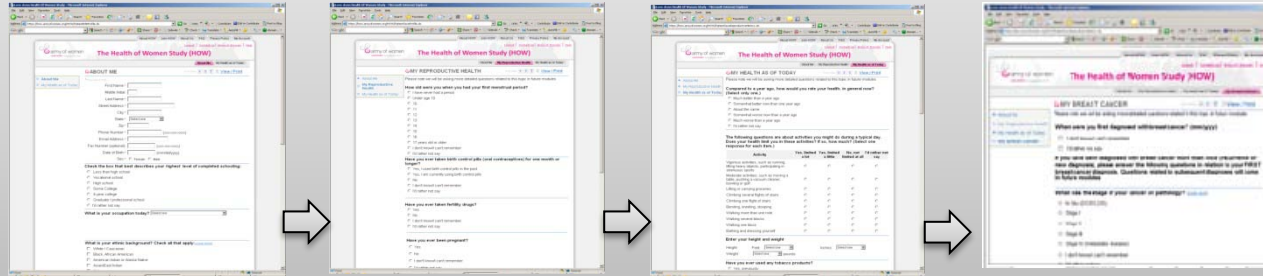
The 'My Health as of Today' section asks 'Compared to a year ago, how would you rate your health, in general now?' with radio button options: 'Much better than a year ago', 'Somewhat better than a year ago', 'About the same', 'Somewhat worse than a year ago', 'Much worse than a year ago', and 'No answer not sure'. It includes a table of activities with columns for 'Yes, I do', 'No', and 'No answer not sure'. The table lists activities like 'Vigorous activities, such as running, bicycling, sports', 'Moderate activities, such as walking & swimming', 'Light or sedentary activities', 'Drinking one 8-oz. glass of alcohol', 'Smoking, drinking, driving', 'Walking more than one mile', 'Walking one block', and 'Walking and driving around'. The section concludes with 'Enter your height and weight' and 'Have you ever used any tobacco products?'.

My Cancer

The 'My Cancer' section is titled 'MY BREAST CANCER' and asks 'Have you ever had breast cancer?' with radio button options: 'I had breast cancer', 'I didn't have it', and 'No answer not sure'. It includes a table of activities with columns for 'Yes, I do', 'No', and 'No answer not sure'. The table lists activities like 'Vigorous activities, such as running, bicycling, sports', 'Moderate activities, such as walking & swimming', 'Light or sedentary activities', 'Drinking one 8-oz. glass of alcohol', 'Smoking, drinking, driving', 'Walking more than one mile', 'Walking one block', and 'Walking and driving around'. The section concludes with 'Enter your height and weight' and 'Have you ever used any tobacco products?'.

Deployment Configuration

The questionnaire is deployed on a cluster of application servers accessed through a load balancer



Cloud Fail Over Plan



Army of Women site
www.armyofwomen.org

User registers with AoW

Peak volume



Cloud site

Email passed to cloud site.
User notified of pending communication.

OPTION 1

Email addresses stored for batch download

OPTION 2

Email addresses passed through for immediate notification

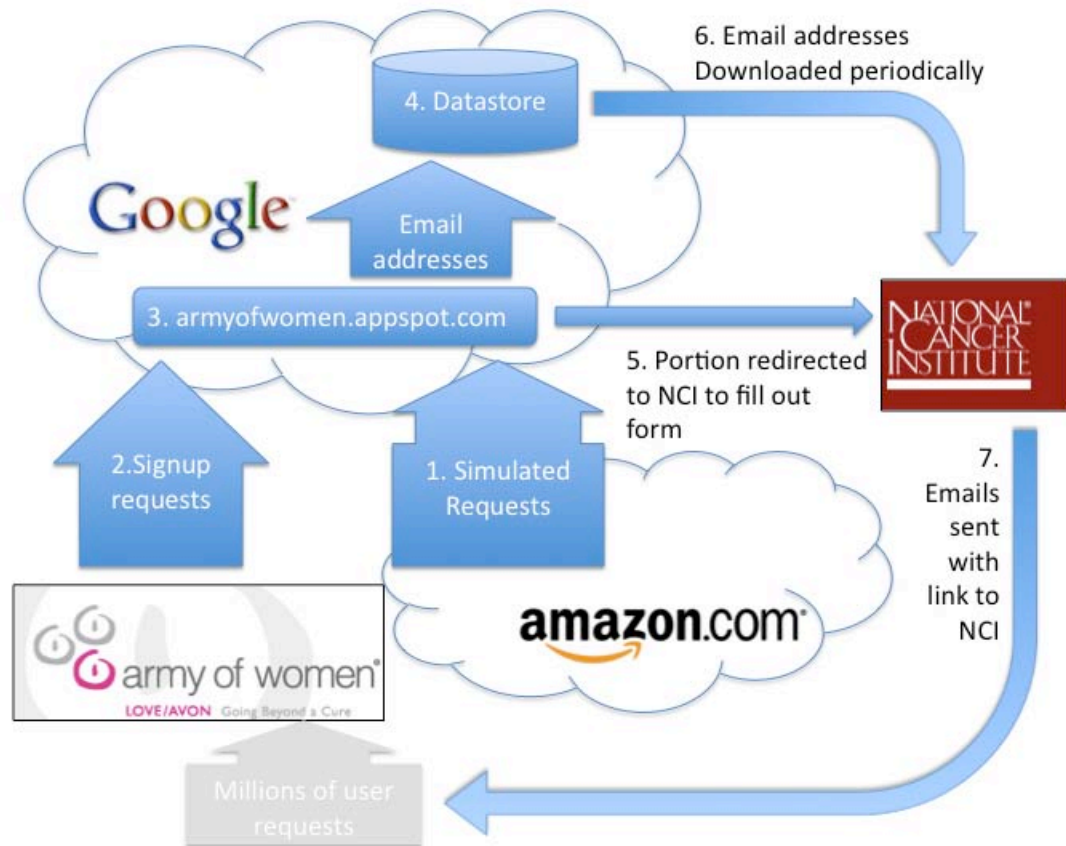
Email notification sent

User clicks on email link

NCI Health of Women site

User directed to NCI Health of Women site

Off-peak volume



AOW/caBIG[®] Collaboration: Health of Women Study (HOW)

- **Members of AOW are invited to join the HOW study and respond to periodic secure online questionnaires concerning health history**
- **Authorized researchers access data and design study protocols based on profiles and data of potential research participants**
- **Database enables “interactive” and “dynamic” process – as researchers conceive new projects and women add more health information, new theories can be generated and additional opportunities for participation arise**



Army of Women

Health of Woman Study

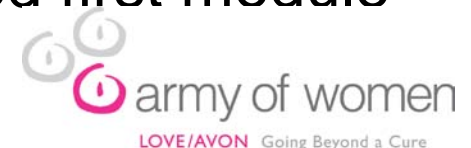
Invitation sent out in escalating batches to current AOW population (262,047) between 12/8/09 and 12/28/09

In response to this **single e-mail** invitation

- 30% viewed the invitation
- 57% who viewed the invitation clicked yes they were interested

By 2/11/10

- 28,032 users (62% of those who clicked yes)
- 25,162 have completed and submitted first module



NCI caBIG Program

NCI CBIIT

- **Ken Buetow**
 - **George Komatsoulis**
 - **John Speakman**
 - **Juli Klemm**
- ...and many, many
more**

PopSci Grid / Community Portal

Partners in development

- Tetherless World Constellation at RPI
 - Deborah McGuinness
 - Li Ding
 - Tim Lebo
 - Jim McCusker
 - Joanne Luciano
 - Maryam Fazel-Zarandi
 - Dominic DiFranzo
- SONIC Laboratories, Northwestern University
 - Nosh Contractor
 - York Yao
 - Yun Huang
 - Hugh Devlin
- NCI
 - Brad Hesse
 - Abdul Shaikh
 - Rick Moser
 - Glen Morgan
 - Erik Augustson
 - Gordon Willis
 - Yvonne Hunt
 - Dave Stinchcomb
 - Zaria Tatalovich
 - Angela Mariotto
 - Kathy Cronin
- SAIC
 - Paul Courtney
 - Lila Rutten



GEM Team

- **NCI**
 - Rick Moser
 - Brad Hesse
 - Abdul Shaikh
 - Paul Courtney
 - Erik Augustson
 - Gordon Willis
 - Glen Morgan
- **SAIC**
 - Mary Cooper
- **Westat**
 - Cindy Helba
 - Kerry Levin
 - Kisha Coa
 - Marsha Dunn
 - David Garner
 - Tom Milke

Cloud Computing

Booz Allen Hamilton

- **Paul Brown**
- **Ed Kohlwey**
- **Barb Benoit**

Love Army of Women – Health of Women Survey

Love Army

- **Susan Love**
- **Naz Sykes**

Healthcare IT

- **Bill Tulskie**

For More About caBIG®

Resource	Link
The caBIG® Web site is the best centralized resource for general caBIG® information	https://cabig.nci.nih.gov/
To send in a specific question about caBIG®, write to this e-mail	caBIGconnect@cancer.gov
Read postings to the “caBIG® Announce Listserv” where community news and notices are posted	https://list.nih.gov/archives/ca_big_announce.html
Contact NCICB Applications Support for technical help with NCICB applications and tools	ncicb@pop.nci.nih.gov