



Data Analytics and HIT

Brian Worley

**Presentation to Heart of America
HIMMS Chapter**

June 12, 2013

Outline of Presentation

- Big Data Science
- Health Data
- Healthcare Analytics

First, a trivia quiz

Topic: Computer Scientists

The Internet? We are not interested in it.



Bill Gates
Co-Founder of Microsoft

I think there is a world market for maybe five computers.



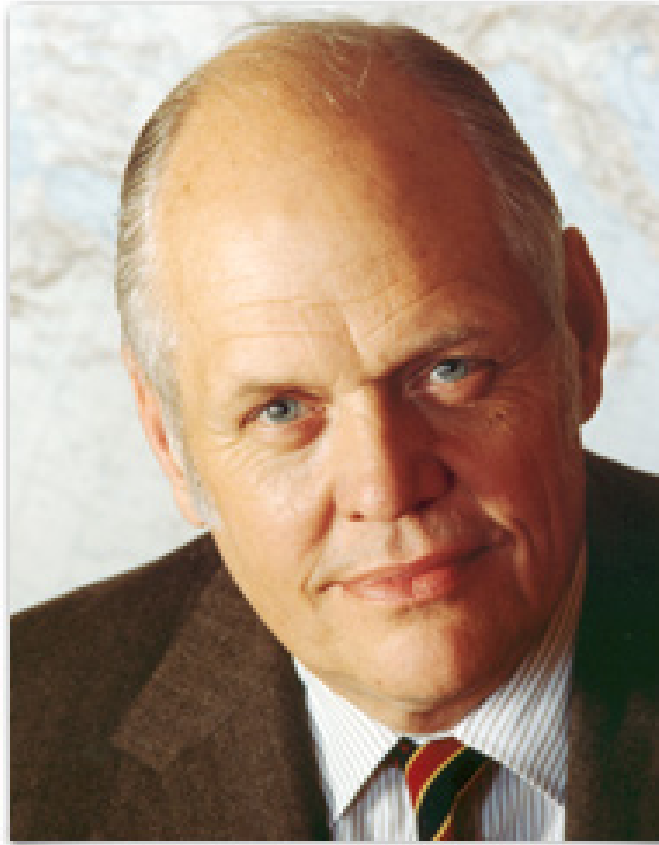
Thomas J. Watson, Sr.
Founder of IBM

If Al Gore invented the Internet, I invented
spell check.



Dan Quayle
former Vice President of the United States

I see no reason why anyone would want a computer in their home.



Kenneth Olsen , Founder of Digital Equipment

Man is still the most extraordinary
computer of all.



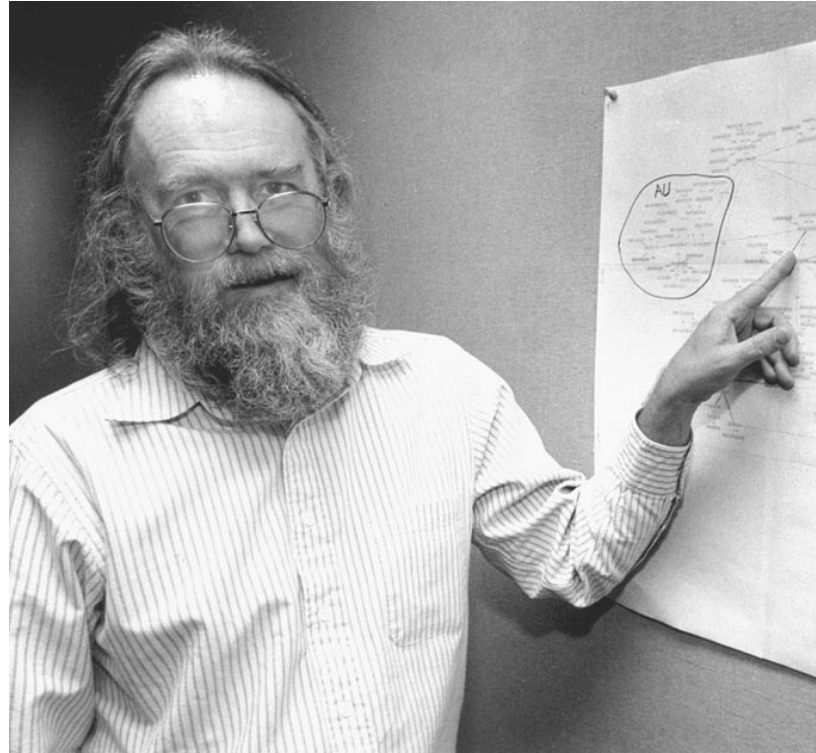
John F. Kennedy
former President of the United States

To create a new standard it takes something that's not just a little bit different. It takes something that's really new and really captures people's imagination. And the Macintosh, of all the machines I've ever seen, is the only one that meets that standard.



Bill Gates

Be conservative in what you do, be liberal in what you accept from others.



Jon Postel

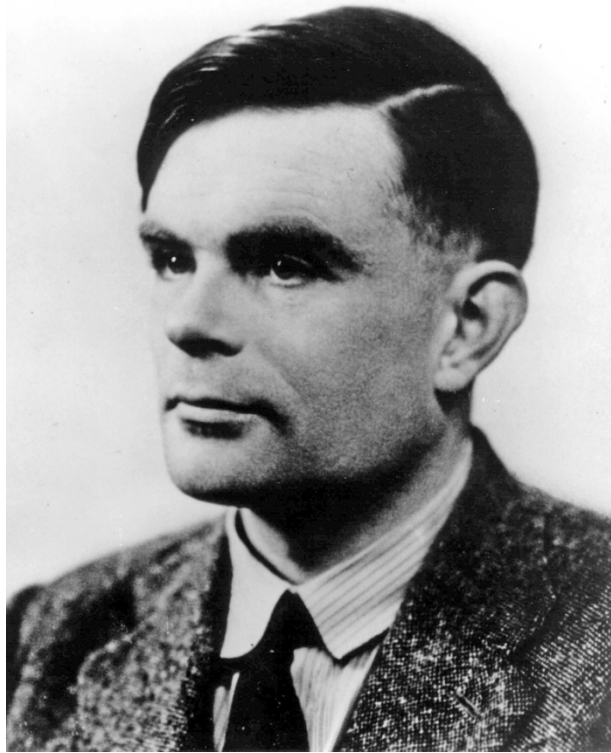
Internet pioneer and original Editor
of Internet Requests for Comments (RFCs)

It would appear that we have reached the limits of what it is possible to achieve with computer technology, although one should be careful with such statements, as they tend to sound pretty silly in 5 years. (~1949)



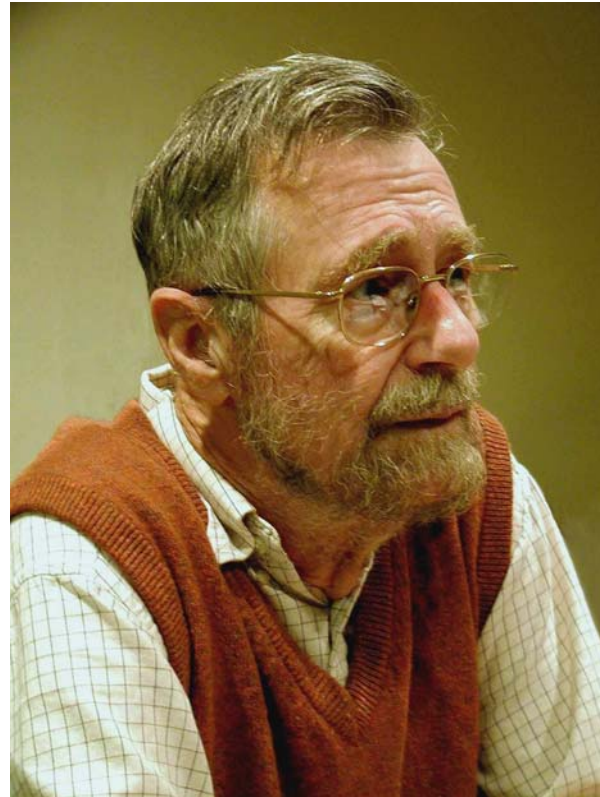
John Von Neumann
20th century mathematician

No, I'm not interested in developing a powerful brain. All I'm after is just a mediocre brain, something like the President of the American Telephone and Telegraph Company.



Alan Turing
English mathematician, cryptanalyst, and
computer scientist.

The question of whether computers can think is just like the question of whether submarines can swim.

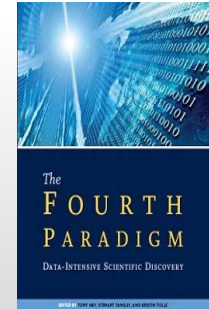
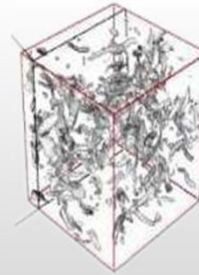


Edsger W. Dijkstra
Dutch computer scientist

The changing nature of research



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



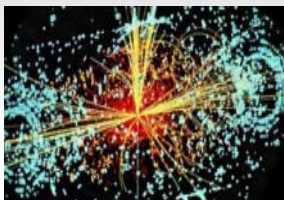
Experiment	Theory	Computation	Data
<p>Thousand years ago</p> <p><i>Description of natural phenomena</i></p>	<p>Last few hundred years</p> <p><i>Newton's laws, Maxwell's equations...</i></p>	<p>Last few decades</p> <p><i>Simulation of complex phenomena</i></p>	<p>Today and the Future</p> <p><i>Knowledge Discovery from Disparate and Dynamic Data</i></p>

The Data Explosion

Experiments



Simulations



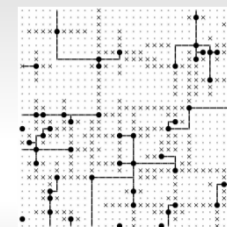
Archives



Social Media



Sensors



Information Technology

The Challenge
Enable Discovery

Deliver the capability to
mine, search, and analyze
this data in near real time

Petabytes
Exabytes
Zettabytes

The Response

Science itself is evolving

Big Data = Volume, Variety, Velocity



The business of big data

- **\$300 billion** annual value of big data for the U.S. health care system, two-thirds of which would come in reduced expenditures (McKinsey).
- **\$165 billion** worth of value for big clinical data (McKinsey).
- **966 petabytes** data stored by discrete manufacturing companies in the U.S. during 2009; 848 petabytes of data stored by government in the same year (McKinsey).
- By 2020, IT departments will have 10 times more servers and 50 times more data to look after than they do now.

Data Analytics Expertise is in High Demand

- The U.S. will face shortages of:
 - between **140,000** and **190,000** individuals with “deep analytical skills” capable of working with very large data sets;
 - between **300,000** and **400,000** skilled technicians and support staff;
 - about 1.5 million “data-savvy” managers and analysts. (McKinsey)

Knowledge Discovery Science

Actionable insights from massive, dynamic, disparate data

Systems

- Sharing and trust
- Social media
- Streaming
- Architecture
- Sensors
- Mobile
- Workflow

Data Analytics

- Text Analysis
- Multi-modality fusion
- Clock-constrained
- Large-scale
- Geo-temporal
- Social networks

Modeling & Simulation

- Discrete-event
- Agent-based
- Predictive
- Real Time
- Physics-based

Cyber Security

- Storage
- Transmission
- Legal Governance
- Insider Threats
- Trust

Mobile platforms (Cell phones, iPad, PDA, UAV)

Desktops, Clusters, and Cloud Computing (Homogeneous and Heterogeneous)

Distributed (Sensor networks, computational platform mixes)

Extreme-scale computing (HPC)

Information platforms based on social media features

Reference by URL

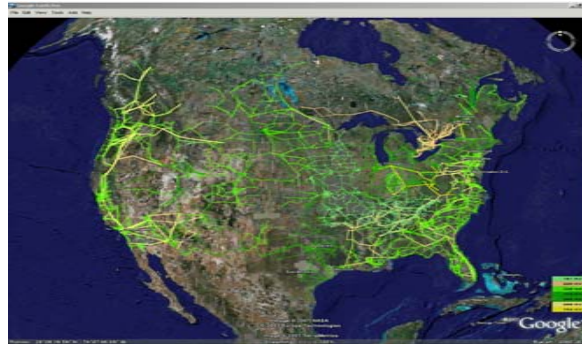
Contributed content

Reputation system

Tagging + search

User-defined mashups

Publish-subscribe



VERDE Electric Grid Status
(Real-Time Grid Awareness)



Tracking 2.0
(Cradle-to-grave tracking)



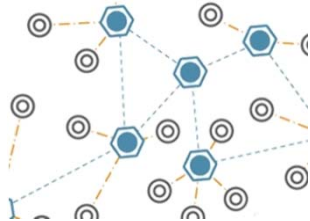
Sensorpedia
(The "Wikipedia of Sensors")



Knowledge Discovery Framework
(National Biomass Distributions)

Architecture concept for big data

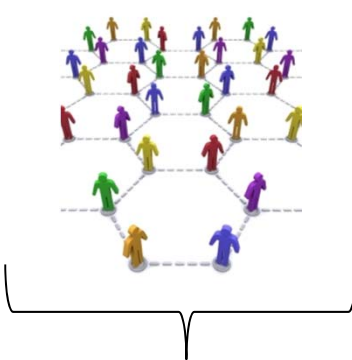
Sensor Networks



Distributed Global Databases



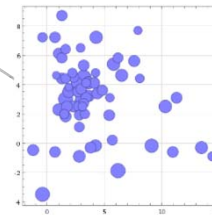
Social Media



Tagging



Analytics



End User



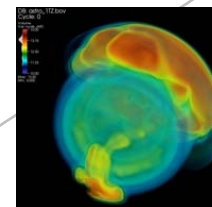
Faceted Classification Systems

- Classify by TYPE
 - Type A
 - Type B
 - Type C
- Classify by ZONE
 - Zone A
 - Zone B
 - Zone C
- Classify by TIME
 - Time A
 - Time B
 - Time C

Facet

Value

Visualization



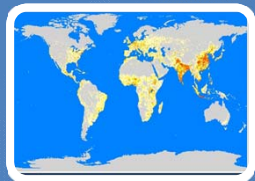
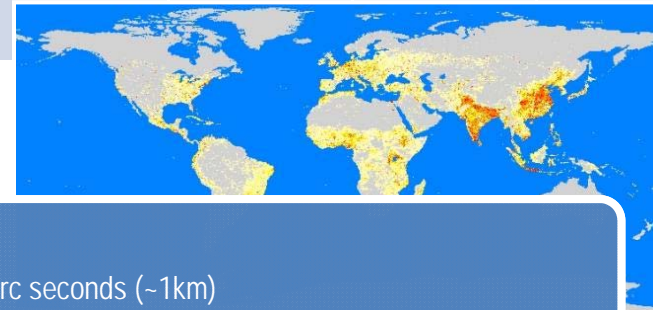
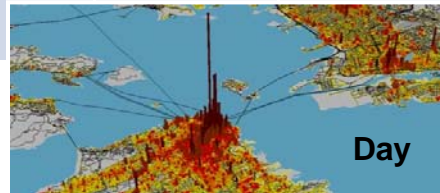
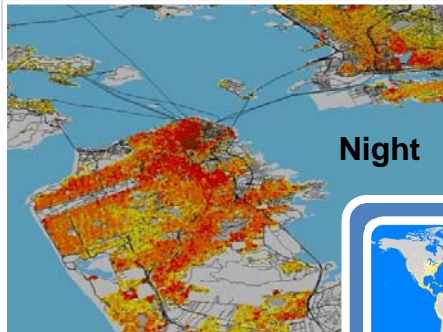
Sources

Architectures

Tools

Big data fusion for prediction of population distributions

Population	Road	Railroads	Land cover/land use	Slope	Academic institutions	Prisons	Hospitals	Business employment	Imagery
<ul style="list-style-type: none"> Census Polygons Tract-to-tract worker flow BLS quarterly updates 	<ul style="list-style-type: none"> VMAP TeleAtlas Multinet TIGER; 	<ul style="list-style-type: none"> 1:100K national railway network NTAD 	<ul style="list-style-type: none"> Geocover MODIS National Land Cover Data (NLCD) State GIS 	<ul style="list-style-type: none"> DTED LiDAR National Elevation Data (NED) 	<ul style="list-style-type: none"> Department of Education HSIP Schools ESRI GDT 	<ul style="list-style-type: none"> National Jail Census HSIP Prisons 	<ul style="list-style-type: none"> American Hospital Association (AHA) 	<ul style="list-style-type: none"> InfoUSA Pitney Bowes Dunn and Bradstreet 	<ul style="list-style-type: none"> Earth-Viewer Terra Server Google



LandScan Global

- Spatial resolution of 30 arc seconds (~1km)
- Ambient population (average of 24 hours)
- Remote sensing based global data modeling and mapping



LandScan USA

- Spatial resolution of 3 arc seconds (~90m)
- Nighttime and daytime population
- Integration of infrastructure and activity databases

Health Data

Open Government Initiative



Transparency promotes accountability

Participation allows people to contribute ideas

Collaboration encourages cooperation within government and with industry

Data.gov

DATA.GOV
TRANSPARENT. OPEN. ACCESSIBLE.

HOME DATA APPS COMMUNITY METRICS OPEN DATA SITES GALLERY WHAT'S NEW

Earthquake and Tsunami Datasets and Information

- Worldwide Major Earthquakes, Part 2 Days
- RadNet Map Interface for Near-Real-Time Radiation Monitoring Data
- Search other related datasets
- World Earthquake Interactive Map Demo

SEARCH OUR CATALOGS
Search our catalogs... SEARCH

DATA AND APPS

- 379,931 raw and geospatial datasets
- 364 government apps
- 236 citizen-developed apps
- 44 mobile apps
- 172 agencies and subagencies
- Suggest a dataset or app!
- 2011 Next Generation Data.gov is interactive, explorative, and social.

COMMUNITIES

Come explore, discuss, meet others in the same field, and develop the data and apps in the community that you care about. Join in the discussions by going to communities below that interest you.

OPEN DATA SEMANTIC WEB
HEALTH LAW

OPEN GOVERNMENT

Latest News: Japanese Earthquake and Radiation Data

Read live alerts and data related to the earthquakes off the coast of Japan and radiation levels in US.

What's coming up on Data.gov? Check out Barry's blog.

LEARN

Are you a teacher, professor, or student? Young people across America are learning about Data.gov and we want you to be part of getting data into the classroom.

See what teachers are doing, share a lesson plan, showcase an app. Learn what's going on in classrooms today!

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. We and our collaborators at the Tetherless World Constellation at the Rochester Polytechnic Institute are helping to lead the way in this exciting area!

DEVELOPER'S CORNER

Are you interested in sharing your mashups, apps, and ideas? Do you want to learn how to create app and mashups with some of the data hosted here on Data.gov? Whether you are here to share, learn, collaborate, or innovate—you've come to the right place.

Look at these great government mobile apps and help us to develop more!

About | FAQ | Contact Us | Data Policy | Accessibility | Privacy Policy | Sitemap

DATA.GOV

- Break down stovepipes
- Provide instant access to raw data in easy to use formats
- Encourage development of innovative applications
- Drive innovation and knowledge use across the globe

Data.gov Communities



Community

Restore the Gulf ✓

Open Data ✓

Semantic Web ✓

Health ✓

Law ✓

Energy ✓

Education

Ocean

Research and Development

Public Safety

Human rights

+ many more...

Health Data is Being Digitized

- More than a third of physician practices plan to purchase, replace or upgrade ambulatory EHR systems, according to HIMSS Analytics' newest Ambulatory Electronic Health Record & Practice Management Study. Meanwhile, nearly half of physician groups say they'll join an HIE.

Meaningful Use Means Dollars

- **Set of national priorities that would help healthcare performance and improvement efforts.**
- **Initial focus: improved population health, coordination of care, improved safety, increased efficiency, reduction of racial disparities, and patient engagement. Privacy and security were later added to the list. The ideals were used to create a framework to determine the meaningful use of EHRs.**
- **Meaningful use is a qualification to receive federal funding for health information technology, specifically, the use of electronic health records. According to the provisions of the Healthcare Information Technology for Economic and Clinical Health Act (HITECH), healthcare organizations that have achieved meaningful use by 2011 will be eligible for incentive payments, and those who have failed to achieve that standard by 2015 may be penalized.**

Criteria for Meaningful Use Have Been Established

- **Stage 1 meaningful use criteria set the baseline for electronic data capture and information sharing. Stage 2 and Stage 3 (expected to be implemented in 2015) will continue to expand on that baseline.**
- **The federal government released the final rules for meaningful use Stage 2. The rule is set to become effective for program participants no earlier than 2014.**
- **For eligible professionals, there are 25 meaningful use objectives for Stage 1. To qualify for payment, 20 of those 25 objectives must be met.**
- **For eligible hospitals and CAHs, there are a total of 24 meaningful use objectives. To qualify for an incentive payment, 19 of these 24 objectives must be met.**

The Health Information Exchange (HIE) Means Dollars

- **Health information exchange is the transmission of healthcare-related data among facilities, health information organizations and government agencies, according to national standards for interoperability, security and confidentiality.**
- **The HIE implementation challenge will be to create a standardized interoperable model that is patient-centric, trusted, longitudinal, scalable, reliable and financially sustainable.**

ICD-10

- The ICD-10-CM revision has more than 68,000 diagnostic codes, compared to the 13,000 found in ICD-9-CM.
- The revision also includes twice as many categories, and it introduces alphanumeric category classifications for the first time.
- The United States was set to begin using ICD-10 on Oct. 1, 2013. The new deadline for ICD-10 is October 14, 2012.

Significant HIT Challenges

- **understanding of who, exactly, owns the data**
- **data capture**
- **storage capacity**
- **search/retrieval**
- **meaningful analytics**
- **information visualization**
- **distribution and sharing of information**
- **talent**

Cloud Services

- Hosted services over the Internet.
- Infrastructure-as-a-service (IaaS)
- Platform-as-a-service (PaaS)
- Software-as-a-service (SaaS).
- Sold on-demand, elastic (the user can have as much or as little of the service they want), and fully managed by the provider.
- A cloud can be either public or private.
- Risk assumed by the provider.
- Amazon cloud is the most popular. Recently endorsed by HHS for hosting HIPPA data. Approved to sign BAAs.

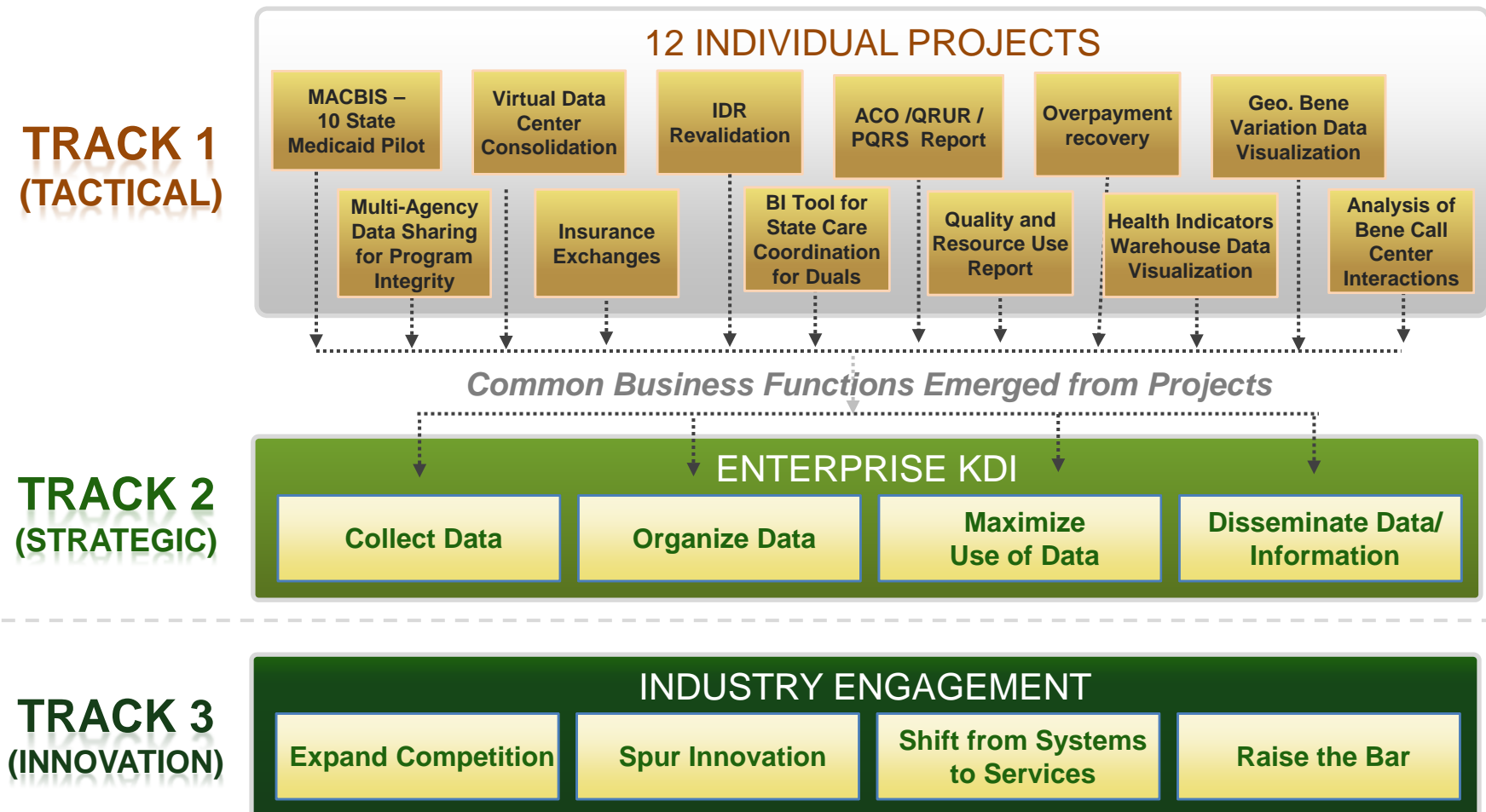
Big Data Architectures

TERADATA.ASTER

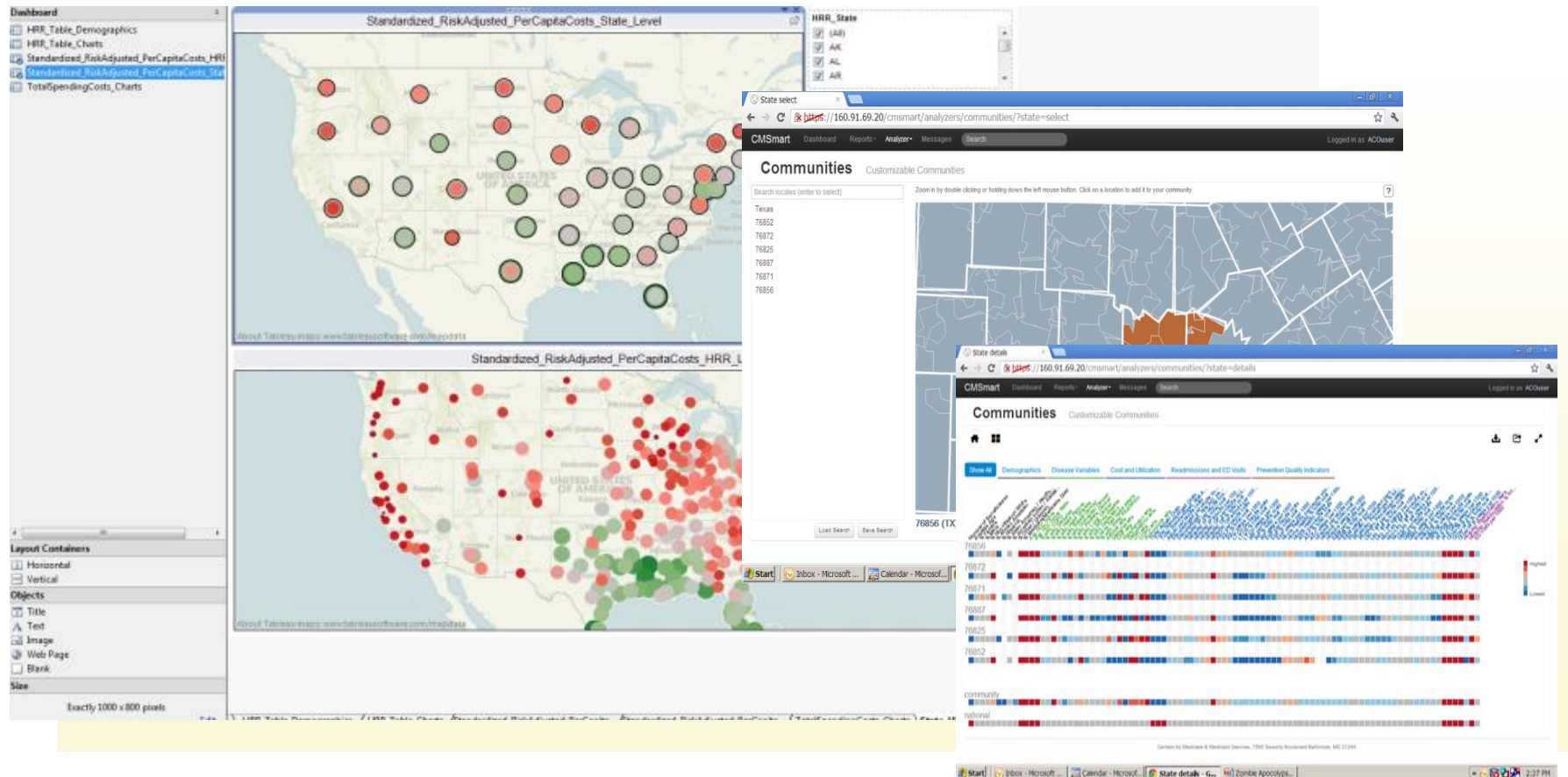


Healthcare Analytics

ORNL Work with CMS

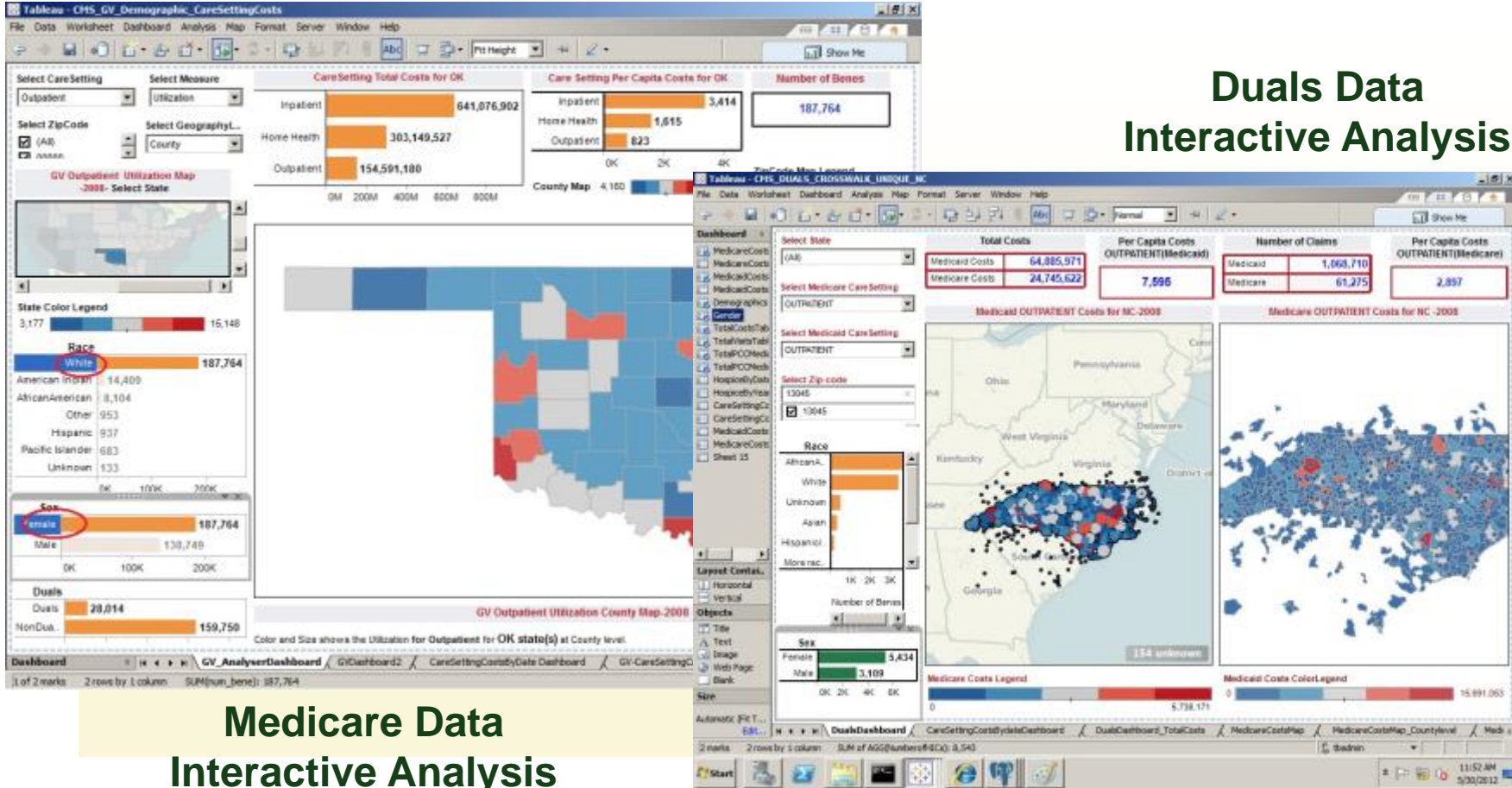


Healthcare Regional Status – Anticipating ACA - Pay by Region for Insurers



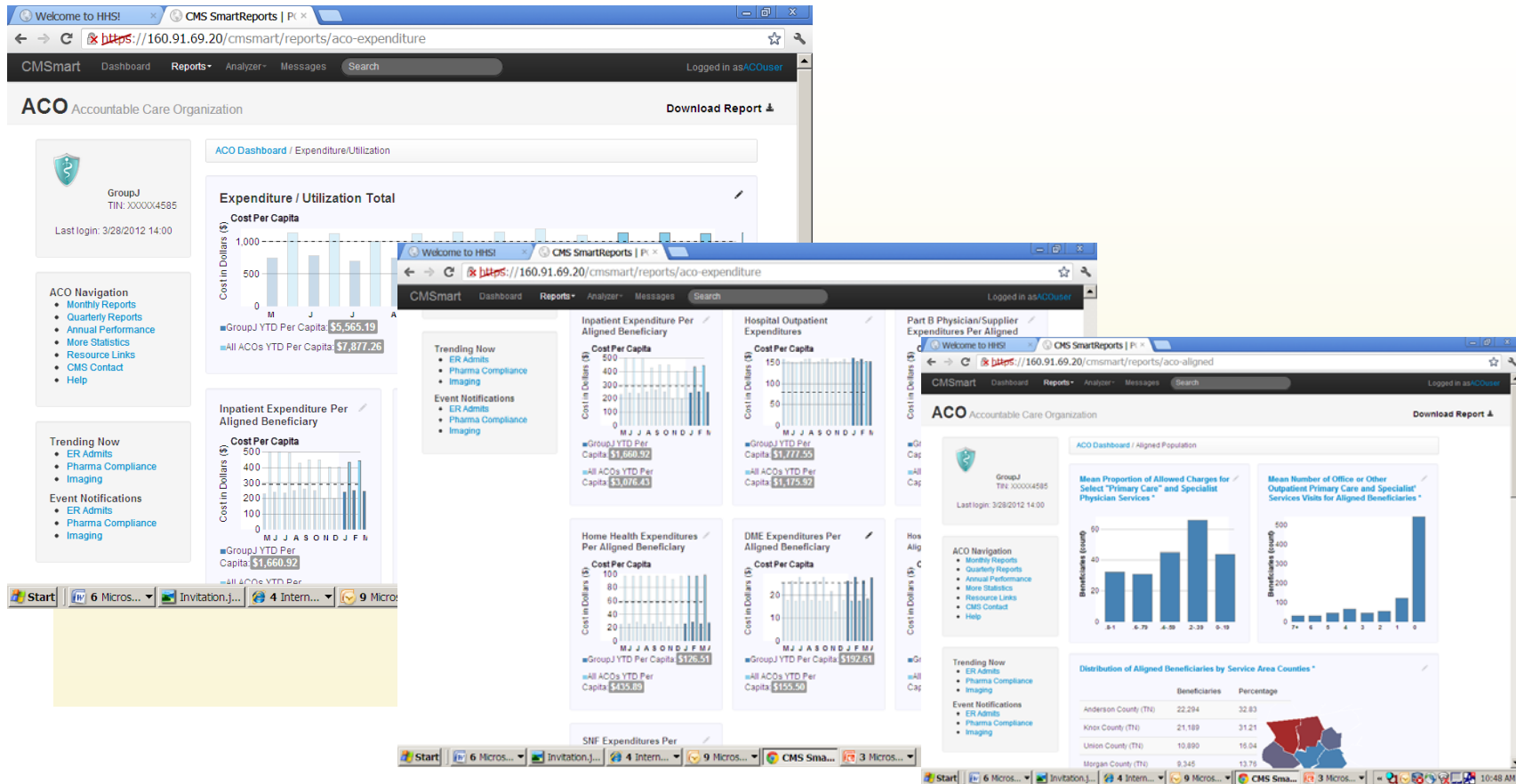
Visual-Analytic Tools for Expenditures and Utilization by Race, Sex, Geography, Disease Codes, etc.

Duals Data
Interactive Analysis




Medicare Data
Interactive Analysis

Provider Feedback Vignettes (PQRS, QRUR, ACO)



Medicaid – 10 State Pilot for T-MSIS

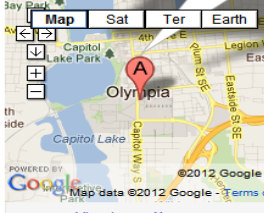
MACBIS
File Types ▾ Dashboard Reporting Messages Search



Washington
Capital: Olympia

Last login: 1/8/2012 11:00

Map Sat Ter Earth

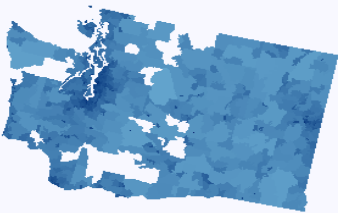


View Larger Map

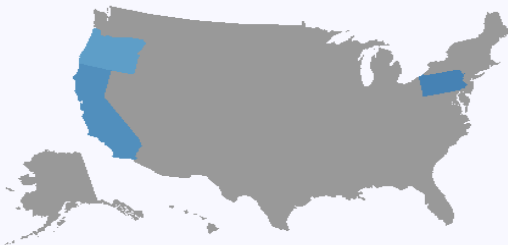
Your last submission failed.
Summary: 1/1/2012 17:49:51 Error: 5515C

Download PDF »
Download Data »

[WA] Cost by Zipcode (\$)



Other States (Cost)



MACBIS Navigation

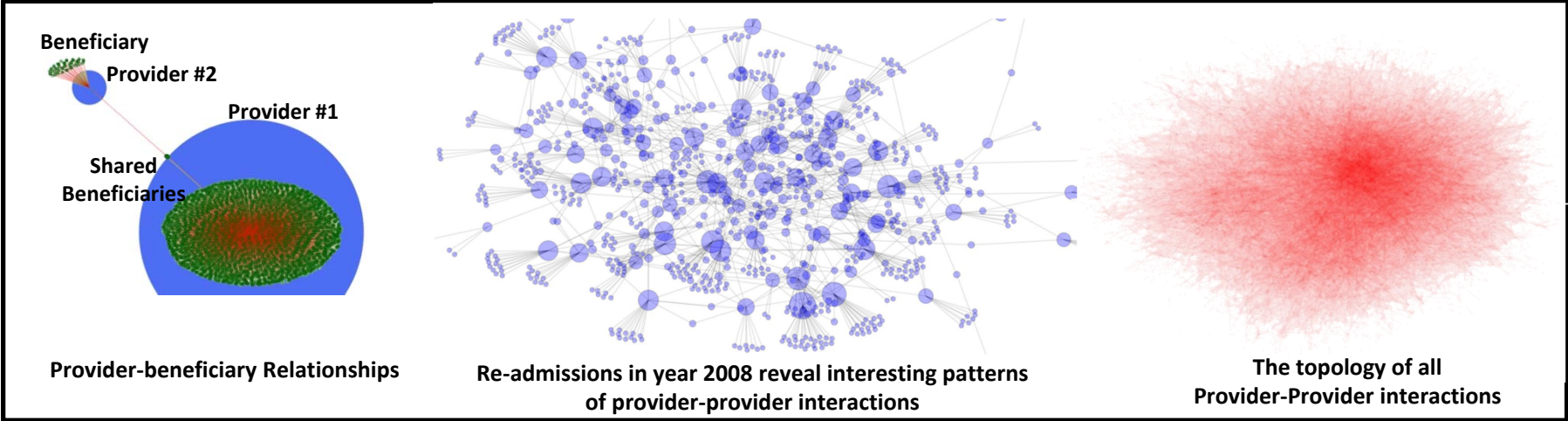
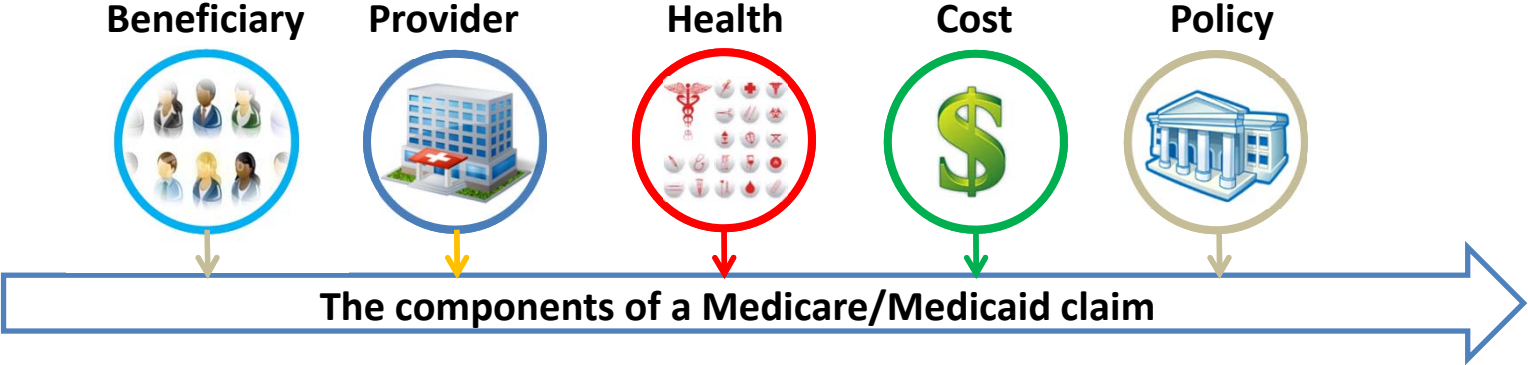
- Provider
- Costs
- Long Term
- Inpatient
- Pharmacy
- Third Party
- Managed Care
- Other
- Submission Status
- Help

[WA] Provider Distribution

Individual	30,081
Fac/Agency/Orgn/Inst	12,111
Group Practice	3,433
Billing Agent/Clearinghouse	375
Tribal Health Services	270
Managed Care Organization	53
Special Considerations	3

Centers for Medicare & Medicaid Services, 7500 Security Boulevard Baltimore, MD 21244

Healthcare Graph-based Analytics



Healthcare Analytics Will Reduce Costs by Addressing Key Issues

- Unwarranted medical procedures
- fraud, waste, and abuse
- administrative costs
- provider inefficiencies
- coordinated care
- preventable conditions

Healthcare Analytics for Administrative Cost Reduction by Improving:

- the billing process
- the coding process
- the logistics process.

Healthcare Analytics for Improving Care

- **Complementing clinical trials with social media data and HIE data**
- **Reduction of medical errors**
- **More successful drug development**
- **Data-driven preventive care**
- **Consumers of healthcare, not just patients**

Thank You

Brian A. Worley

bworley@pyaanalytics.com