



NTIS Federal **Data** Meeting

An aerial photograph of a large industrial facility, possibly a power plant or refinery, with a complex network of pipes and structures. The image is overlaid with semi-transparent colored shapes: a large green triangle in the upper right, a blue triangle in the lower left, and a grey triangle in the lower right. The text is centered over the image.

Welcome

Greg Capella, Deputy Director NTIS

Introduction and Meeting Objectives

Avi Bender, Director NTIS

FEDERAL DATA MEETING

National Technical Information Service (NTIS)

U.S. Department of Commerce

April 4, 2017

Federal-Private Sector Joint Venture Partnership for Data Innovation

Avi Bender, Director, NTIS

avi.bender@ntis.gov

703-605-6401

Why Are we Meeting Today?

Data is a catalyst for:

- *Supporting government operational excellence and innovation*
- *Improving mission effectiveness and efficiency*
- *Driving IT Modernization at the intersection of Data Science and Technology*

How Does all of this (Innovation) Happen?

It starts with an early Federal-Private Sector conversation about mission and business outcome:

- 1. Critical success factors (CSF) for your agency*
- 2. Data driven business requirements to achieve the*
- 3. CSF Well informed functional requirements*
- 4. Technical requirements for agile development*

Realization Through an Operating Model Based on Federal-Private Sector Joint Venture Partnership

What Common Challenges Do Federal Agencies Face?

CAPACITY TO SCALE

TIMELY ACCESS TO THE SKILLS, RESOURCES, TOOLS, INFRASTRUCTURE

Consume

Process

Publish

- DATA ANALYTICS
- DATA COLLECTION
- MACHINE LEARNING DATA
- INTEROPERABILITY
- CLOUD INFRASTRUCTURE
- SECURITY/PRIVACY
- DATA MINING

- VISUALIZATION
- DATA DISSEMINATION
- STATISTICS
- GEOSPATIAL ANALYTICS
- REMOTE SENSING
- OPEN SOURCE TOOLS
- HIGH PERFORMANCE COMPUTING

INFRASTRUCTURE —APPLICATIONS—SECURITY--SKILLS

NTIS-JVP Addressing Data Driven Decisions Facing Federal Agencies

↑
INSIGHT &
VALUE



FUELING THE BUILDING BLOCKS OF INNOVATION

Sample “Pain Points” Where NTIS and JVP’s Are Focusing

Federal agencies and the private sector are challenged in their capacity to work together (time, skills, infrastructure) using data and technology to benefit the public and to improve government operations. A JVP Agreement can take as little as 90 days.

Agency Data Challenges (Sample)*

- Citizen services (health, safety,...)
- Fraud Detection (HHS/OIG, USDA)
- Customs Immigration Services
- Improper Payments
- Pricing equity (purchasing)
- Food and drug safety (FDA)
- Emergency Preparedness

* *Current clients DHS, HHS, NTIA*

Private Sector Innovation

- ❖ DATA ANALYTICS
- ❖ DATA COLLECTION
- ❖ MACHINE LEARNING/AI
- ❖ SMART CITIES
- ❖ DATA INTEROPERABILITY
- ❖ CLOUD INFRASTRUCTURE
- ❖ SECURITY/PRIVACY
- ❖ DATA MINING
- ❖ SOCIOECONOMIC RESEARCH
- ❖ DATA INTEGRATION
- ❖ VISUALIZATION
- ❖ DATA DISSEMINATION
- ❖ STATISTICAL RESEARCH
- ❖ GEOSPATIAL ANALYTICS
- ❖ REMOTE SENSING
- ❖ OPEN SOURCE TOOLS
- ❖ FRAUD DETECTION
- ❖ HIGH PERFORMANCE COMPUTING
- ❖ PREDICTIVE ANALYTICS
- ❖ PATTERN RECOGNITION

NTIS
JVP

A Structured and Merit Based Process to Support Data Innovation Through Federal Private Sector Collaboration

60-90 Days target

Ideate

Conceptualize

Propose

Implement



Federal agencies with NTIS formulate the Problem Statement. All JVP's invited (together) to discuss the opportunity with NTIS. The discussion enhances the problem definition process.



Draft, review, and refine the Problem Statements in collaboration with all JVP's, Federal agencies, and NTIS.



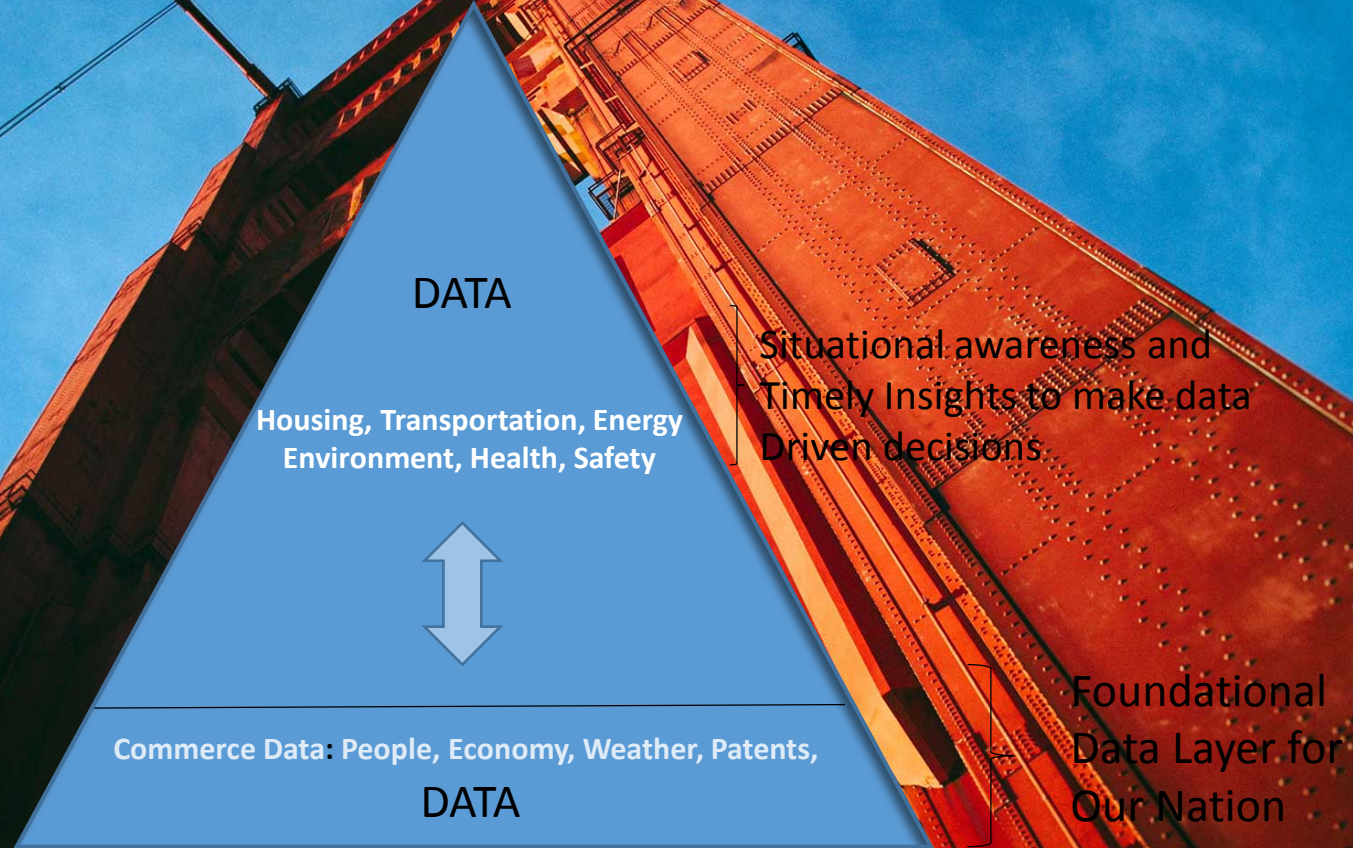
JVP's propose solutions and are selected based on a merit process managed by NTIS.



Solutions are varied but usually apply data science tools, and emerging technologies to discover and connect data sets, and develop and implement analytic tools.

Commerce Data is a Renewable National Resource

Delivering Insights When Combined with Federal and Commercial Data



New Data Services Joint Venture Partners*

JV Partners include 18 small businesses, 4 Universities consortiums, 2 non-profits, and 11 large businesses...the “who is who” in Data Science. All selected based on A rigorous merit based selection process completed October 2016

- 1901 Group, LLC
- Amazon Web Services, Inc.
- Arcadia Data, Inc.
- ASET Partners Corporation
- Bayes Impact
- Berico Technologies, LLC
- Booz Allen Hamilton, Inc.
- Cherokee Nation Technology Solutions, LLC
- Civis Analytics
- Columbia University on behalf of Northeast Big Data Innovation Hub (lead organization), in partnership with the West Big Data Innovation Hub, the Midwest Big Data Innovation Hub, and the South Big Data Innovation Hub

- data.world
- Deloitte Consulting LLP
- D&B Government Solutions
- EBO, Inc.
- Esri
- Excella Consulting, Inc.
- Govini
- HP Enterprise Services, LLC
- Hypatia Project, LLC /dba/Dendrite Technologies
- i3solutions, Inc.
- IBM Corporation
- ICF Incorporated, LLC
- National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

- Onu Technology, Inc.
- Palantir Technologies Inc.
- People Power Company, Maalka, Inc., and Urban Systems
- Qubole
- Socrata
- Splice Machine, Inc.
- SRI International
- Stanford University
- Uncharted Software Inc.
- IPUMS at the University of Minnesota
- Intel

Agreements in process

NTIS Data Joint Venture: Who We Are and What We Do

NTIS is a fee-for-service trusted advisor that provides innovative data services to All Federal agencies, through a joint venture partnership authority with the private sector.

NTIS delivers benefit to the public by advancing Federal data priorities, promoting economic growth, and enabling operational excellence.


Advancing Federal Data Priorities: Addressing Pain Points that Hinder Innovation

Economic Growth (outward public focused working with the private sector):

- Unleash Federal data and make it easier for businesses, communities, citizens to discover, access, analyze, and use.
- Enable the private sector to develop and/or use new and improved data products and services that serve businesses, communities, and citizens.
- Improve quality, scope and timeliness of economic statistics.

Operational Excellence (internally focused- making government more efficient and effective working with the private sector):

- Transform the delivery of critical government data services for the American people.
- Leverage advances in data science, software development, and cybersecurity to increase efficiency/effectiveness of government modernization programs.

An aerial, low-angle view of a large steel truss bridge, likely the Chesapeake Bay Bridge-Tunnel. The bridge's structure is composed of numerous steel beams and girders, creating a complex geometric pattern. The image is overlaid with several semi-transparent, colorful geometric shapes: a large green triangle in the upper right, a blue triangle in the lower left, and a grey triangle in the lower right. The background is a clear, light blue sky.

Keynote

The Honorable Wilbur L. Ross, Jr.
Secretary of Commerce

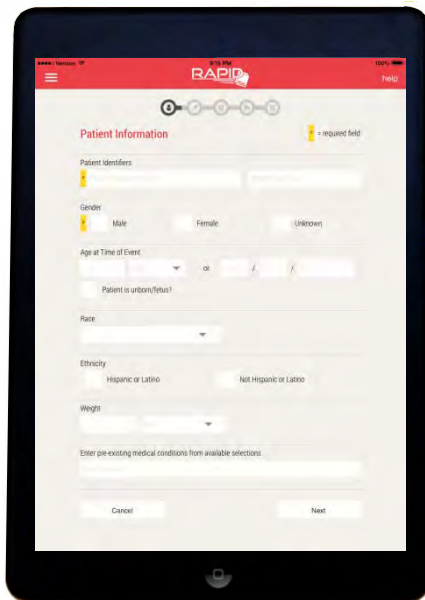


Dr. Henry Francis

FDA Case Study

Addressing Federal Priorities Through
Data Analytics

Real Time Application for Portable Interactive Devices



Henry Francis, M.D.
FDA/CDER/OTS

FDA launched the Medical Counter Measures (MCM) initiative to improve our nation's capacity to respond to CBRN and pandemic threats

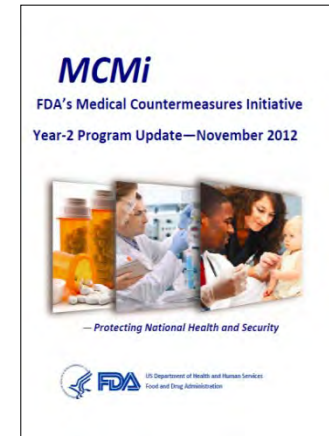
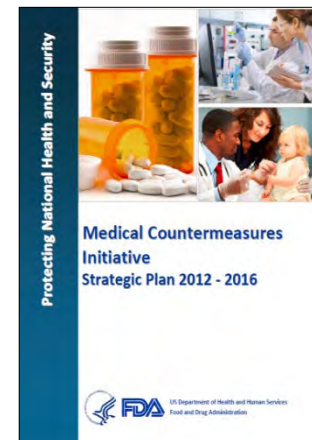
MCMi Overview

- FDA launched its medical countermeasure initiative (MCMi) in response to a comprehensive, year-long review conducted by the HHS assessing the nation's ability to respond to public health emergencies
- Mission: Strengthen and transform the MCM enterprise by augmenting FDA's activities in review, regulatory science, and policy to expedite its ability to respond to CBRN and EID threat



Three Pillar Strategy

- Enhance FDA's product review and approval processes for the highest priority MCMs and related technologies
- **Build the necessary science base for MCM regulatory review and identify clear, efficient regulatory pathways for developing critical MCMs**
- Modernize the legal, regulatory, and policy framework to facilitate MCM development and ensure an effective public health response



What are the Public Health controls in and FDA MCM response?



The cloud-based RAPID Bio-surveillance System will support collaboration between FDA and other Federal agencies to enhance monitoring emerging health threats

RAPID Biosurveillance System

Tier 1: Regulatory Action/Guidance

- FDA and external partners issue guidance to ensure patient safety

Tier 2: Data Visualization

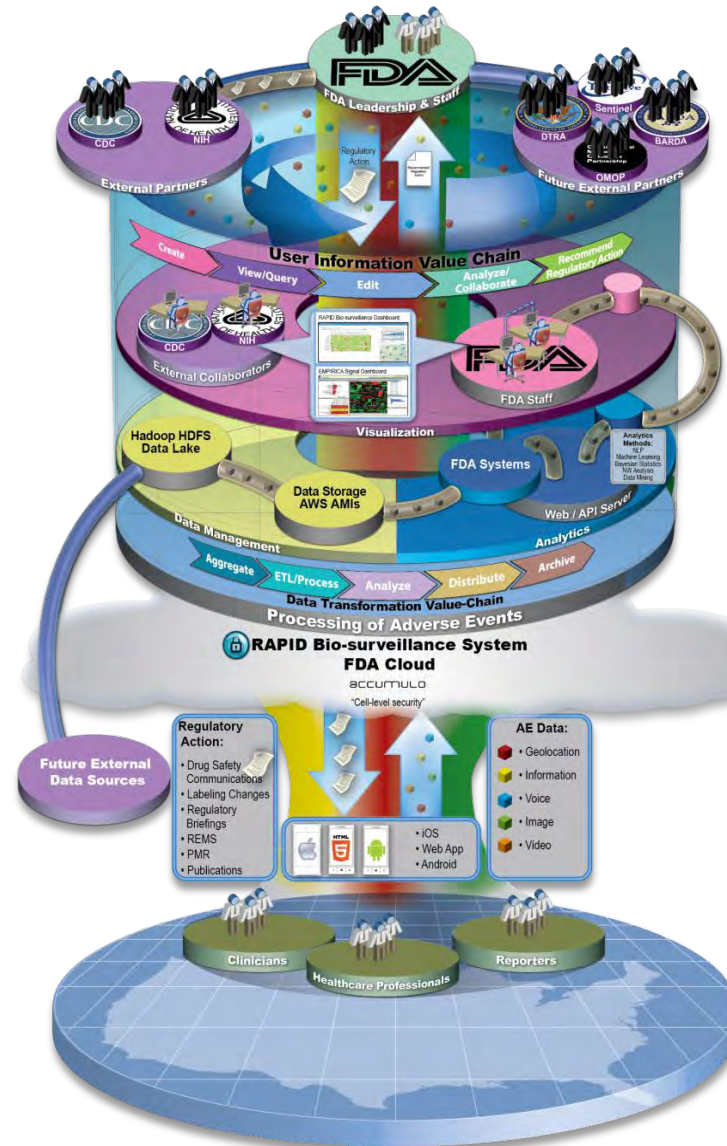
- Dashboards summarizing key information provide safety alerts
- Disproportionality metrics and detailed analyses allow FDA and collaborators to understand emerging issues

Tier 3: Data Management & Analytics

- RAPID data is combined with existing MedWatch and **Medwatcher** reports and data from external collaborators
- Advanced analytics support AE signal detection

Tier 4: Processing of Adverse Events

- Healthcare professionals submit AE data via the RAPID mobile app
- AE data is stored in a “data lake” to support real-time access



m
e
d
w
a
t
h
e
r

D
r
u
g
s
h
o
r
t
a
g
e

FDA requires a real-time active surveillance application to support pharmacovigilance and adverse event reporting in MCM situation

The Solution

- ▶ Real-time Application for Portable Interactive Device (RAPID) will facilitate the real-time collection, analysis, and communication of MCM product and health information during national public health emergencies
- ▶ Flexible Mobile platform to use during MCM events
- ▶ Flexible FDA cloud design complementary to FAERS
- ▶ Adaptable data management and analysis system
- ▶ Bidirectional CDER multimedia communications
- ▶ Decision maker data work bench

The screenshot shows the RAPID mobile application interface on a tablet. The top status bar indicates Verizon service, 2:15 PM, and 100% battery. The app header is red with the 'RAPID' logo and a 'help' link. Below the header is a navigation bar with icons for Home, Add, Edit, and Done. The main content area is titled 'Patient Information' and includes a legend for required fields. The form contains the following sections:

- Patient Identifiers:** A text input field with a yellow asterisk indicating it is a required field.
- Gender:** Radio buttons for Male, Female, and Unknown, with a yellow asterisk next to the Male option.
- Age at Time of Event:** A dropdown menu followed by 'or' and two date input fields (MM/DD/YYYY).
- Patient is unborn/fetus?:** A checkbox.
- Race:** A dropdown menu.
- Ethnicity:** Radio buttons for Hispanic or Latino and Not Hispanic or Latino.
- Weight:** A dropdown menu.
- Enter pre-existing medical conditions from available selections:** A text input field.

At the bottom of the form are two buttons: 'Cancel' and 'Next'.

Selected Drug: Peramivir

To: Case #1114 Reporter All Peramivir Reporters

Enter Message:

Dear Dr. Teller,

Thank you for your recent adverse event submission. We would like to provide you with additional information about the adverse event report you submitted. Please use the link or QR code below to access and share this information.

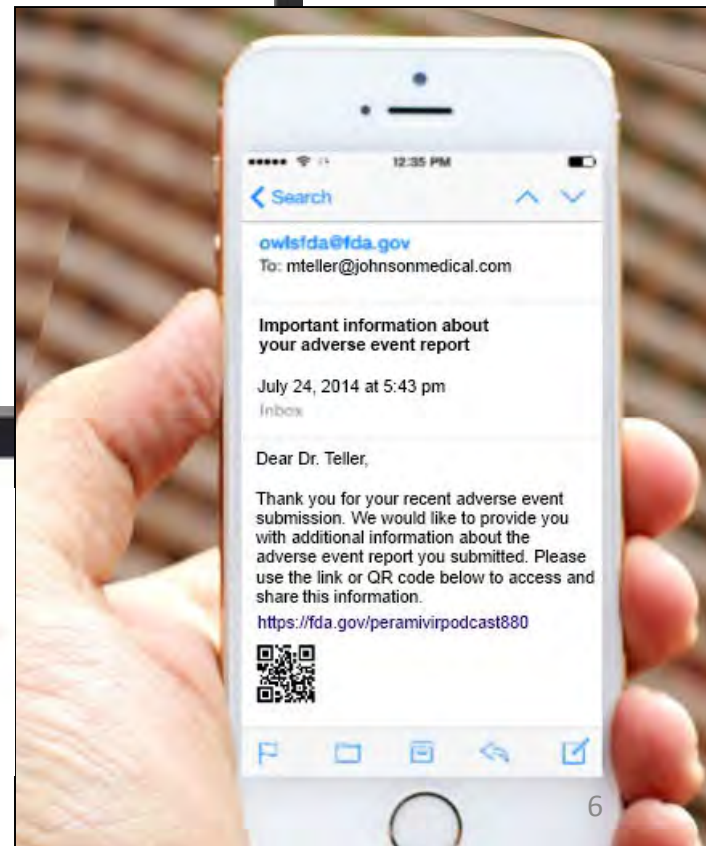
<https://fda.gov/peramivirpodcase880>



Thank you,
FDA

Cancel

Send Message



DPV - Adverse Event Report Use Case



DPV Stakeholders

- Scott Proestel
- Robert Levin
- Samantha Cotter

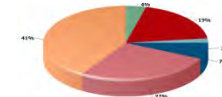
RAPID External Data (Future)



Adverse Event Analytical Tools

RAPID AE Dashboard & Data

Results by Adverse Event Outcomes (All Time)



Geovisualization



RAPID Heatmap



Empirica



1 Mobile Data Collection:
Clinicians, reporters, and patients create general surveillance and emergency surveillance reports

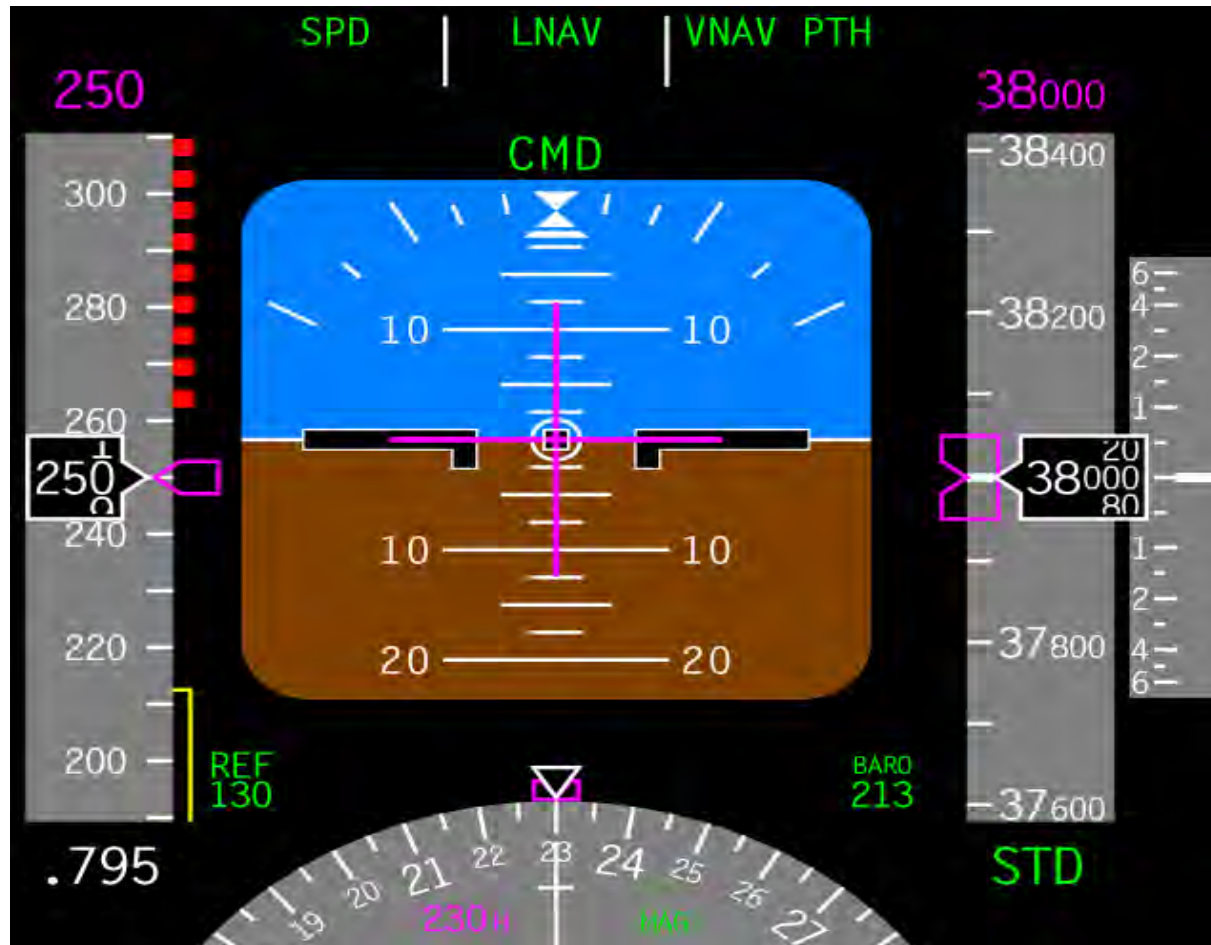
2 Data Transferred to FDA:
Data is submitted from mobile device and sent over Cellular network or WiFi to FDA GovCloud via web services

3 Perform Analytics :
Utilize RAPID Dashboard, location based data, and Empirica (signal detection) for analysis

4 Response Sent to Reporter:
Within 24 hours a targeted response is sent via email containing links to multimedia files like images and podcasts

5 Obtain External Data:
Data will be shared from external sources via web services (ex. HL7 ICSR) and increase the effectiveness of analysis performed using RAPID analytical tools

RAPID : an FDA Public health display for Emergencies



RAPID Phase II

The proposed RAPID Biosurveillance Platform includes a cloud-based open source big data analytic tool to facilitate detection of adverse event signals in near real-time

[Alerts] Display Drugs with > 50% increase in the number of reports in the last 2 weeks

My Projects

- Project 1
- Project 2
- Project 3

Create Project

Drug-Event Tracker

Drug	% Change in Number of Reports (daily)	Algorithm & Ranking Statistic (all data)
<u>Paramivir-H1N1</u>	+ 3	2.0
Avandia-Diabetes	0.17	1.05
MPA-Steroid Injections	1.0	1.3
Zanamivir-H1N1	0.17	1
ALL	+ 10	2.5

Drug-AE Filter:

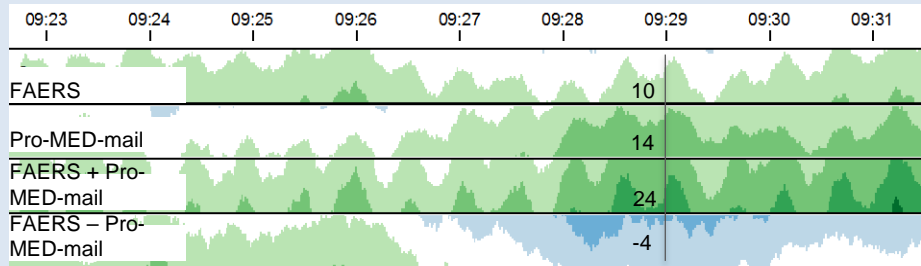
- All drugs in which 1 or more cases are fatal
- All drugs in which 50% of the adverse event cases are reported in pediatrics

Disproportionality Metric Filter:

- ROR
- PRR
- GPS
- IC
- Corresponding Ranking statistic: e.g., 5 per cent quantile $Q_{0.05}$ of the posterior distribution

Clicking on a drug in the list above activates the time series and geographic visualization

Time Series Visualization for Number of Reports for Drug of Interest: Adverse Event Data and Social Media Data



*Mouseover or use the arrow keys to inspect values



*Mouseover bubbles to view information on adverse events reported at different geographic locations

FDA collaborators in MCM events

- Federal agencies
- Healthcare providers
- Healthcare facilities
- Individuals
- Sponsors

RAPID Network of Networks in an Public Health event



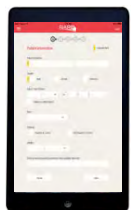
The Real-time Application for Portable Interactive Devices (RAPID) System can inform decision-making at the physician, hospital and Federal level

Data from RAPID combined with other systems

Key questions and capabilities that can be addressed with RAPID data



RAPID Secure Data Broker (SDB) – Provides access to adverse event reports and electronic health records to inform decisions made by physicians, hospitals and the Assistant Secretary.



RAPID Mobile Application



ASPR – RAPID enables improved management of MCM stockpiles

- Number and location of stockpiles (drugs, vaccines, diagnostics)
- Type, severity and location of potential CBRN threats
- Syndromic surveillance data from state/local public health agencies

Monthly

- How to manage stockpile inventory and location?
- How to get MCM resources to patients and at-risk populations?
- What types of threats are most likely to occur in the short- and long-term?



Geographic visualizations



Hospital Administrators – RAPID supports allocation of healthcare resources

- Number of occupied beds
- Patient characteristics (diagnoses, level of care, etc.)
- Patient status (waiting for treatment, ready for hospital discharge, etc.)

Daily

- How to manage current patient flow?
- How many patients are projected to require care in the short- & long-term?
- What number and type of medical equipment are required to diagnose and treat patients in short- & long-term?



Hospital Alerts



Physicians – RAPID informs diagnosis and treatment of patients

- How to diagnose the patient based on symptoms, history, and data from others?
- How can data from other patients influence the current patient's treatment plan?
- What drugs, vaccines and/or diagnostics should be ordered to treat the patient?

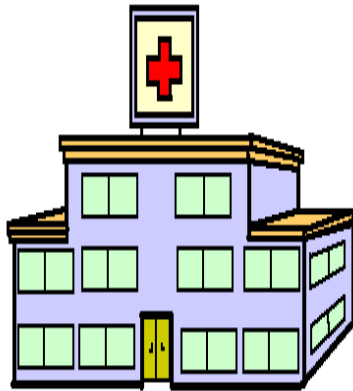
Streaming



Clinical Decision Support

RAPID Secure data sharing Scheme during Public Health Emergencies

Healthcare Facility



Secure data
Broker

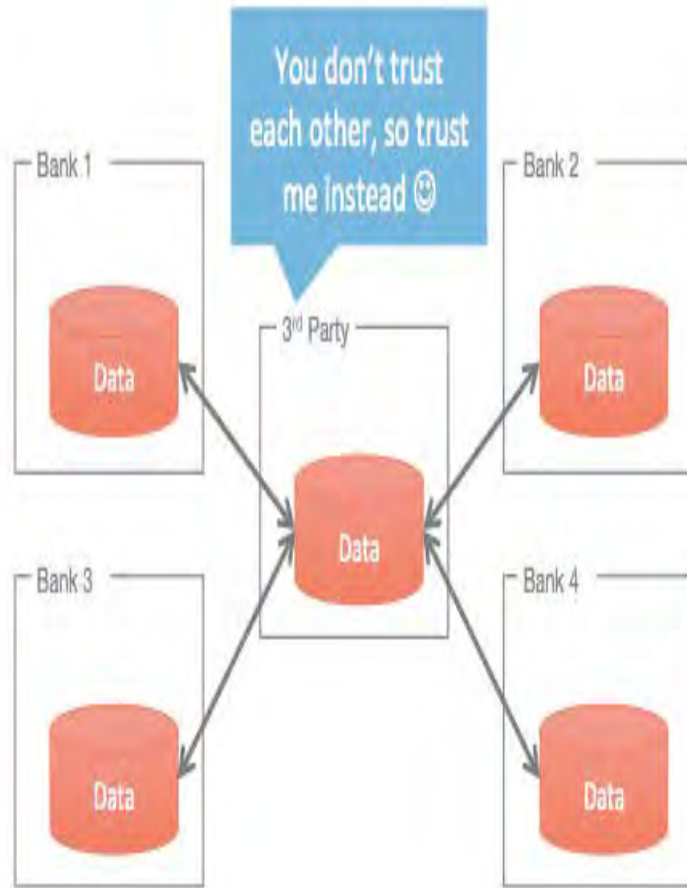


FDA, federal and non
Fed

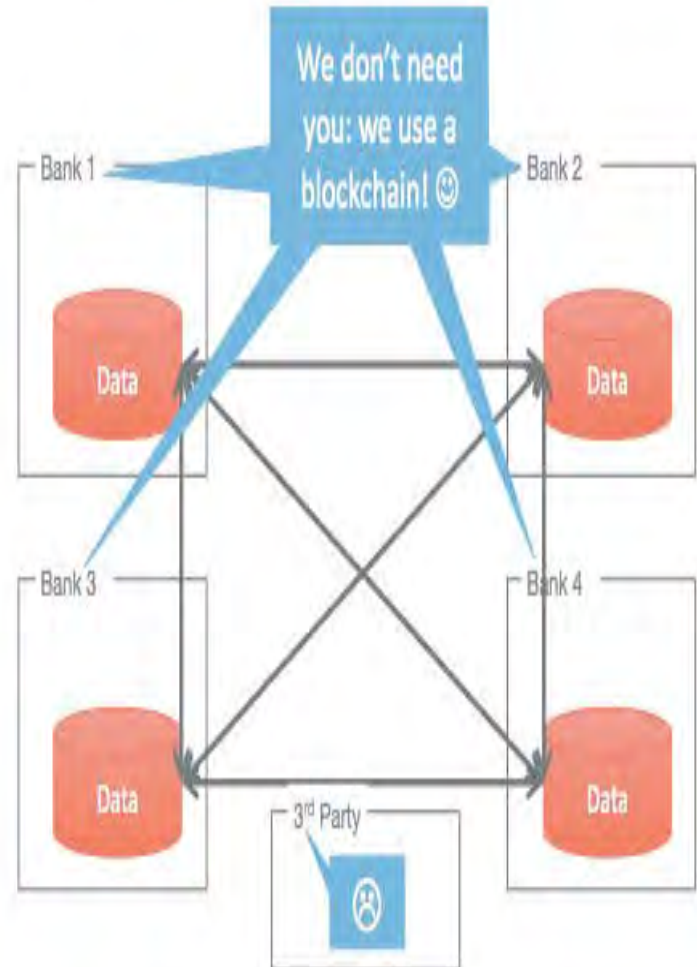


What does a Blockchain connection offer?

Before blockchains...



With blockchains...



Duke University Medical Center

Brigham and Women's Hospital

University of Southern California Hospital

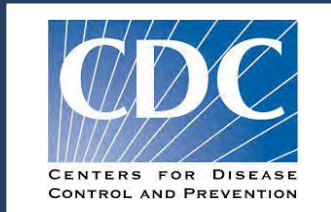
University of Maryland Medical Center



Trigger examples: influenza order, blood culture order, sputum culture order, CXR order
(ESP) Safety or Clinical data



Smart on FHIR API
Clinical Data Element Profile for Severe Acute Respiratory Infection data elements



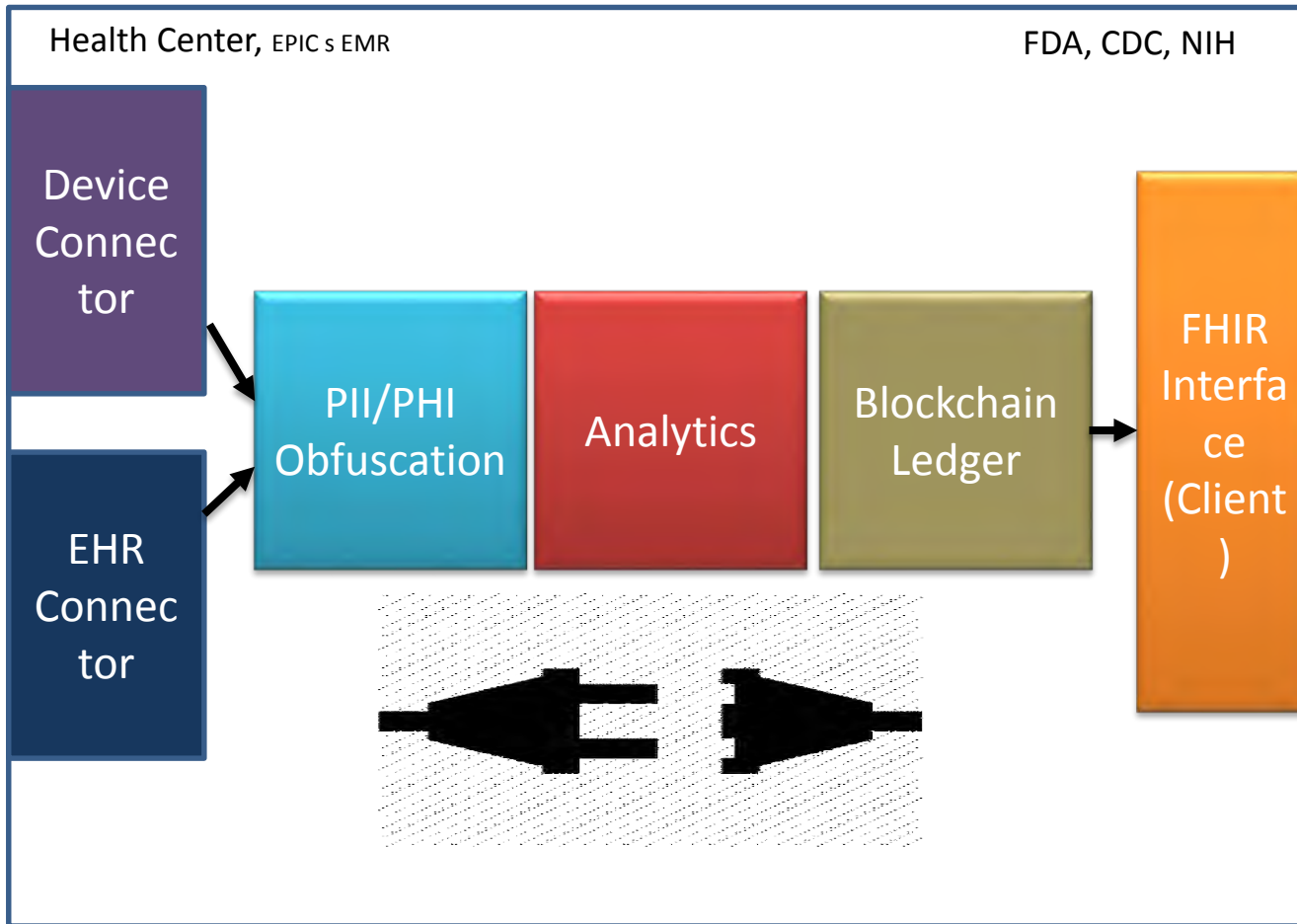
Secure Data Broker
HHS & Booz Allen Hamilton
Data Aggregated & De-identified



Visualization layer with streaming data for research case and patient level



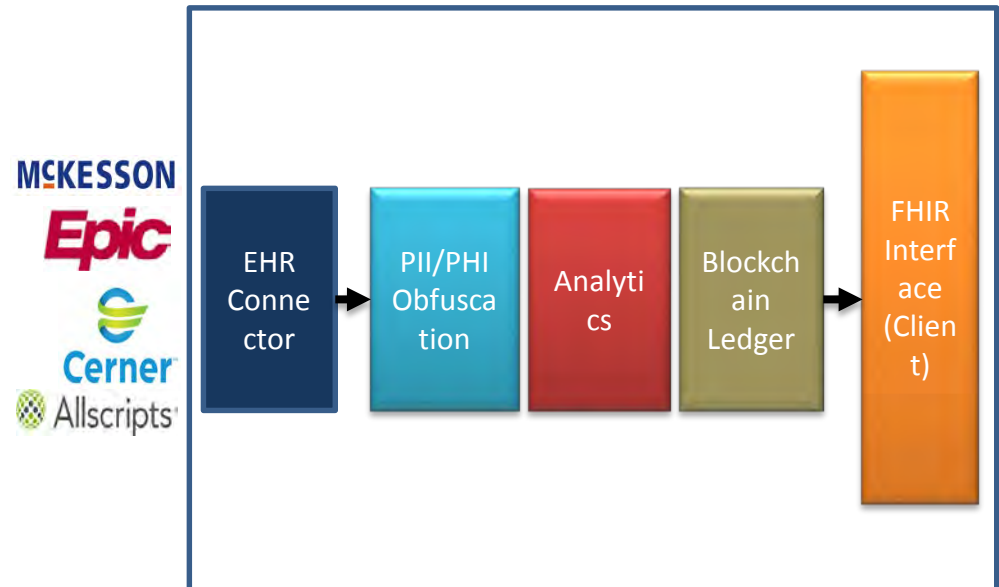
RAPID HealthChain Plugin (RHP v2)



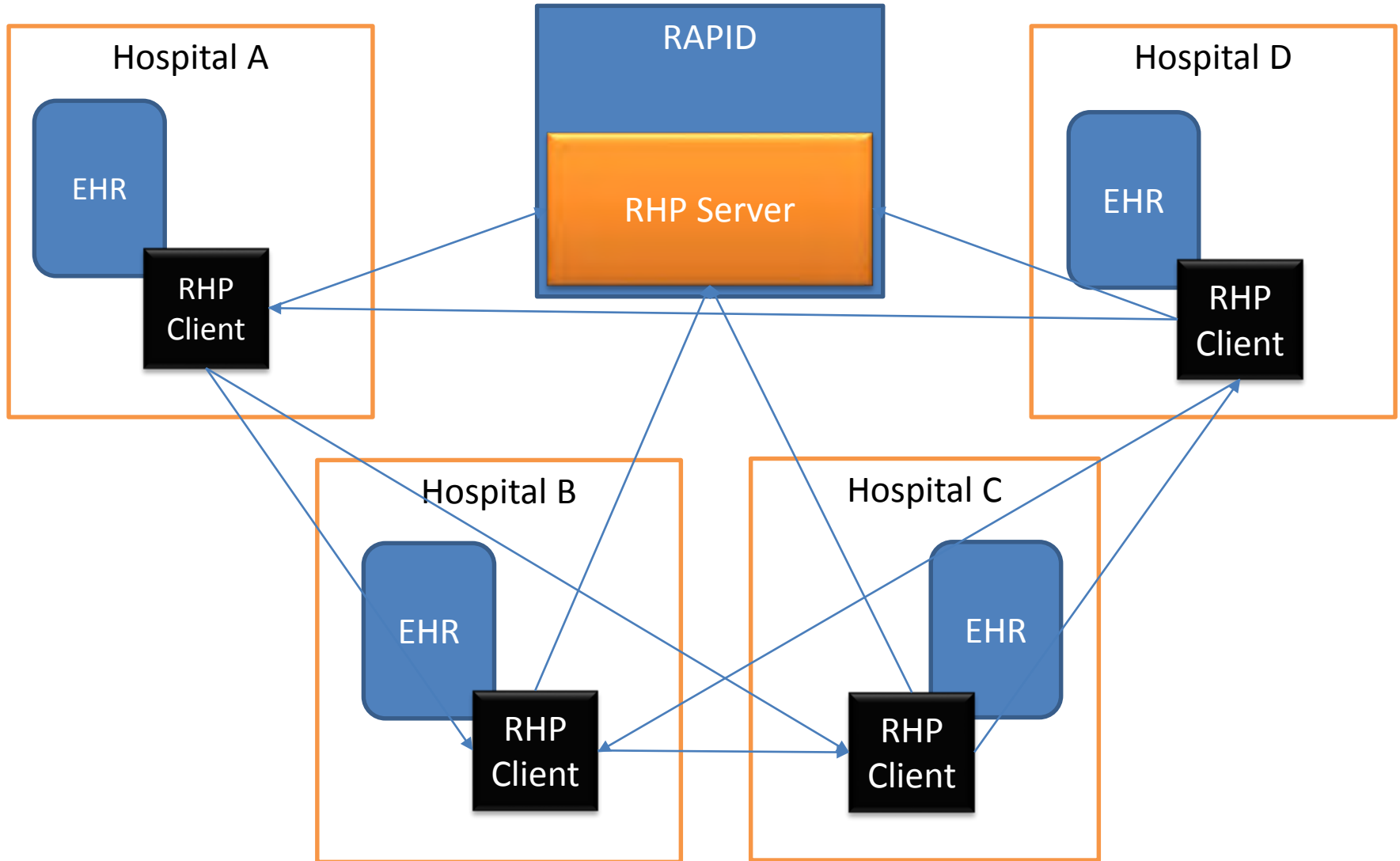
HealthChain Ledger

- Implements Blockchain technology
 - Shadow chain – i.e., a chain that at first copies the performance of more established methods of information sharing to test out the concept
- A ledger & block is associated with a hospital/provider
- **The plugin will contain a ledger so that all transfers of EHR records & streaming data are known and communicated to all participating hospital/partner organizations**
- Transmissions of records are equivalent to transactions

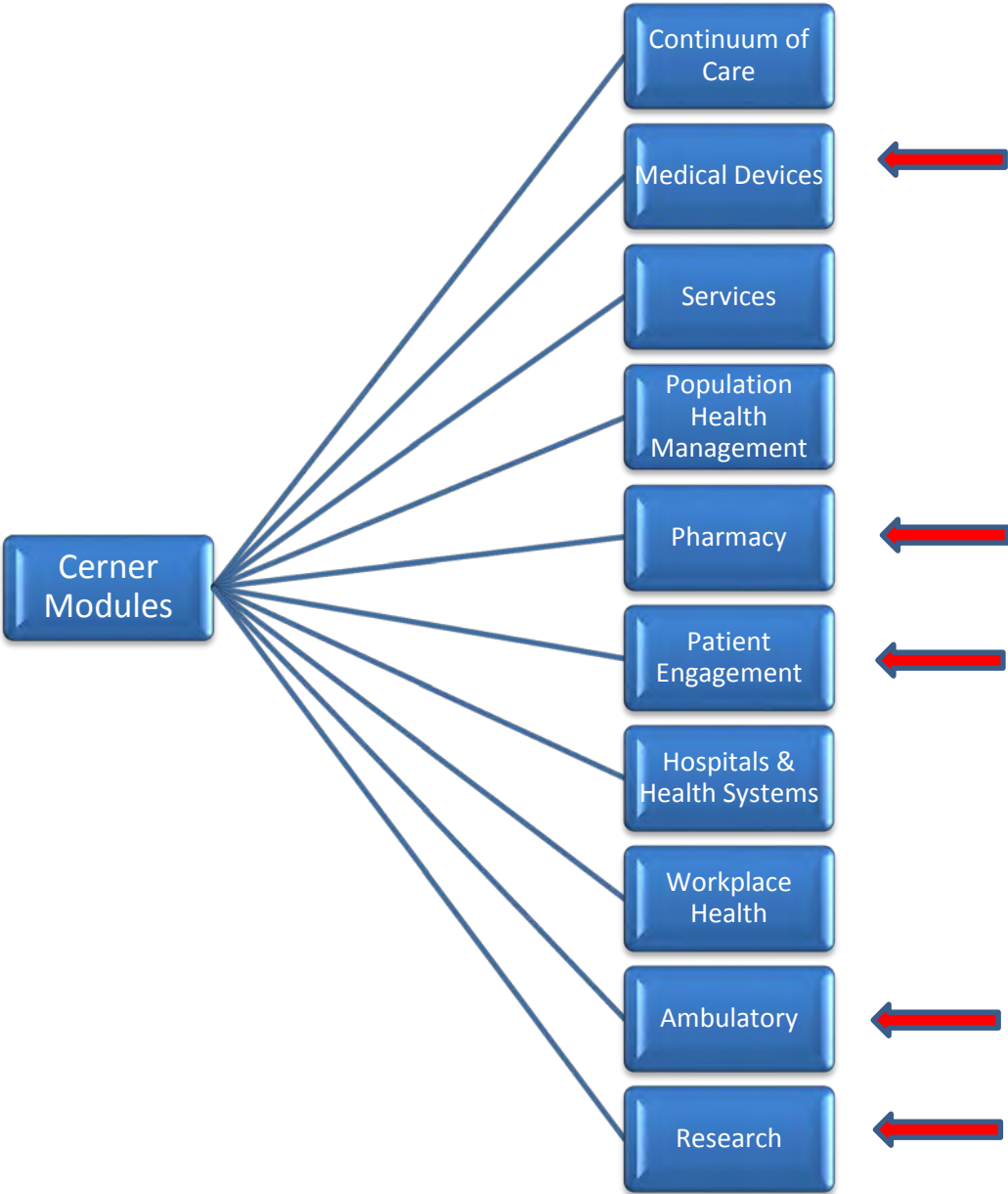
RAPID HealthChain Plugin (RHP v1)



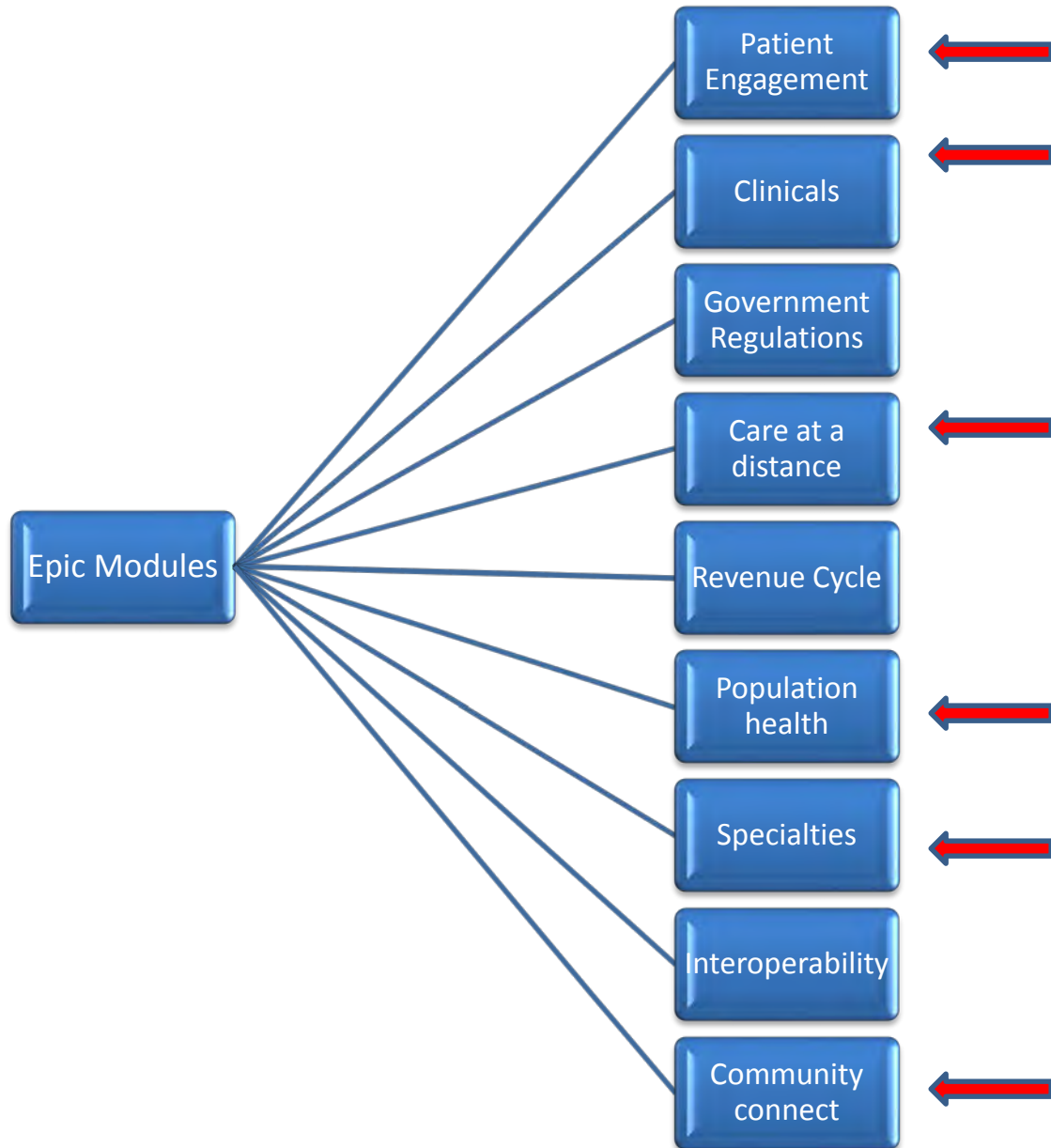
Target Environment (Health Chain)



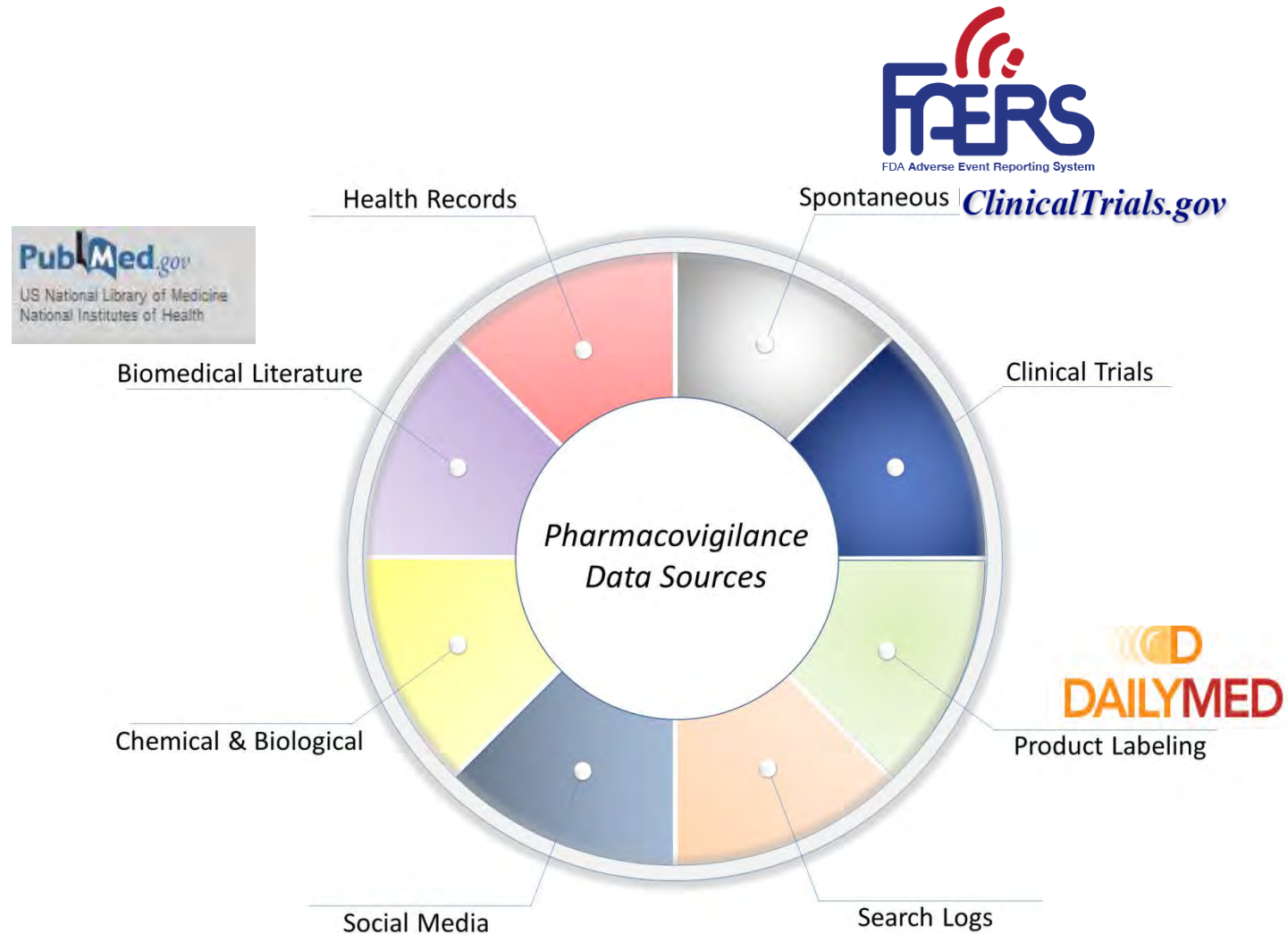
Cerner Modules



Epic Modules



Consortium of Drug Safety Information



RAPID II Summary

- Bidirectional cloud based system
- Capable of real time information and analytics
- Flexible data and analytic platform
- Beginning multimedia information analysis
- **Key issues**
 - How would the transdisciplinary group work?
 - What data, not quantity of data is important?



Dr. Caryl Brzymialkiewicz

HHS Case Study

Addressing Federal Priorities Through
Data Analytics



Predictive Analytics at HHS-OIG

Caryl Brzymialkiewicz, Ph.D.

Assistant Inspector General / Chief Data Officer
Department of Health and Human Services (HHS)
Office of Inspector General (OIG)





Who We Are

Our Mission: Protect the integrity of HHS programs

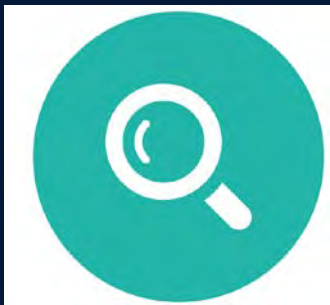
Our Vision: Drive Positive Change





What We Do

We fight **fraud**, **waste**, and **abuse** in Medicare, Medicaid, and more than 100 other HHS programs



Audit



Evaluate



Investigate



Counsel





Medicare and Medicaid

**1 in 3
Americans**

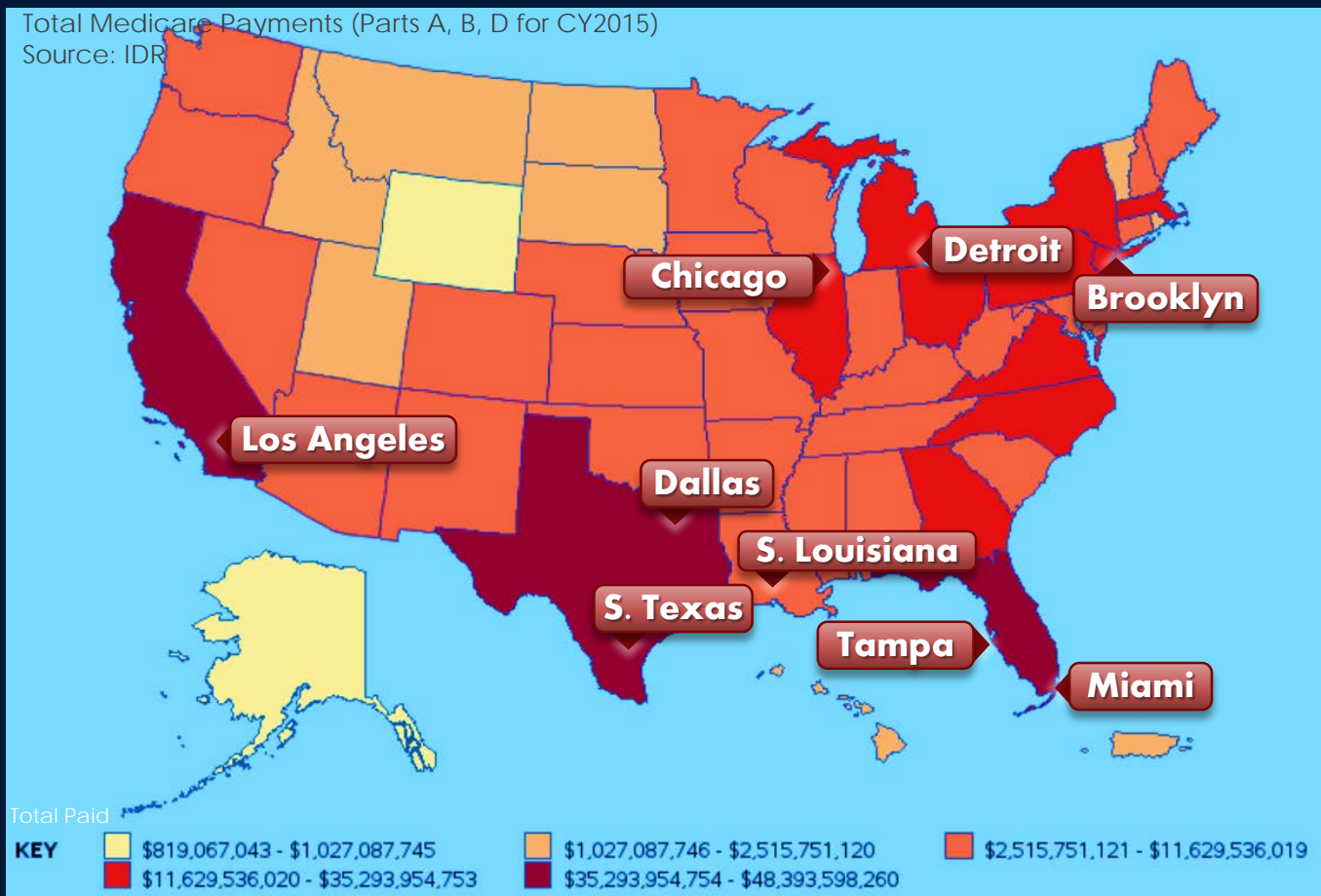




Medicare Payments

Total Medicare Payments (Parts A, B, D for CY2015)

Source: IDR



California	\$48.3B
Florida	\$36.1B
Texas	\$35.3B
New York	\$32.7B
Pennsylvania	\$21.2B
Illinois	\$20.7B
New Jersey	\$18.1B
Michigan	\$17.9B
Ohio	\$17.6B
North Carolina	\$15.6B
United States	\$494.4B

Indicates Strike Force Location





National Takedown: 300 defendants, \$900 million in billings



The Department of Health and Human Services Office of Inspector General, along with our state and federal law enforcement partners, participated in the **largest health care fraud takedown in history** in June 2016. Approximately **300 defendants in 36 judicial districts** were charged with participating in fraud schemes involving about **\$900 million** in false billings to Medicare and Medicaid. Takedowns protect Medicare and Medicaid and deter fraud -- sending a strong signal that theft from these taxpayer-funded programs will not be tolerated. The money taxpayers spend fighting fraud is an excellent investment: For every \$1.00 spent on health care-related fraud and abuse investigations in the last three years, more than \$6.10 has been recovered.

- [Fact Sheet](#)
- [Graphics Gallery](#)
- [Photo Gallery](#)
- [RAW VIDEO: 2016 Health Care Fraud Takedown \(OIG-HHS\)](#)

TAKEDOWN STATISTICS

\$900M in False Billing
300+ Defendants Charged
1,000 Law Enforcement Personnel, Including 350 OIG Agents
36 Judicial Districts

Share

[Tweet](#)

[Share via Email](#)

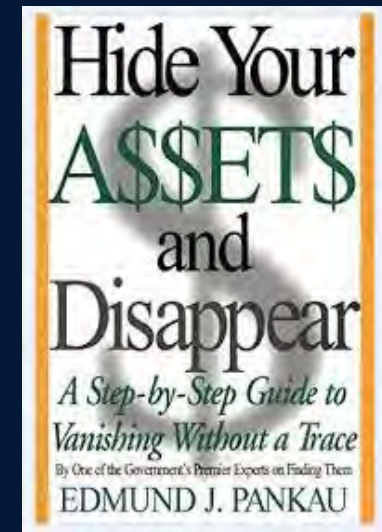
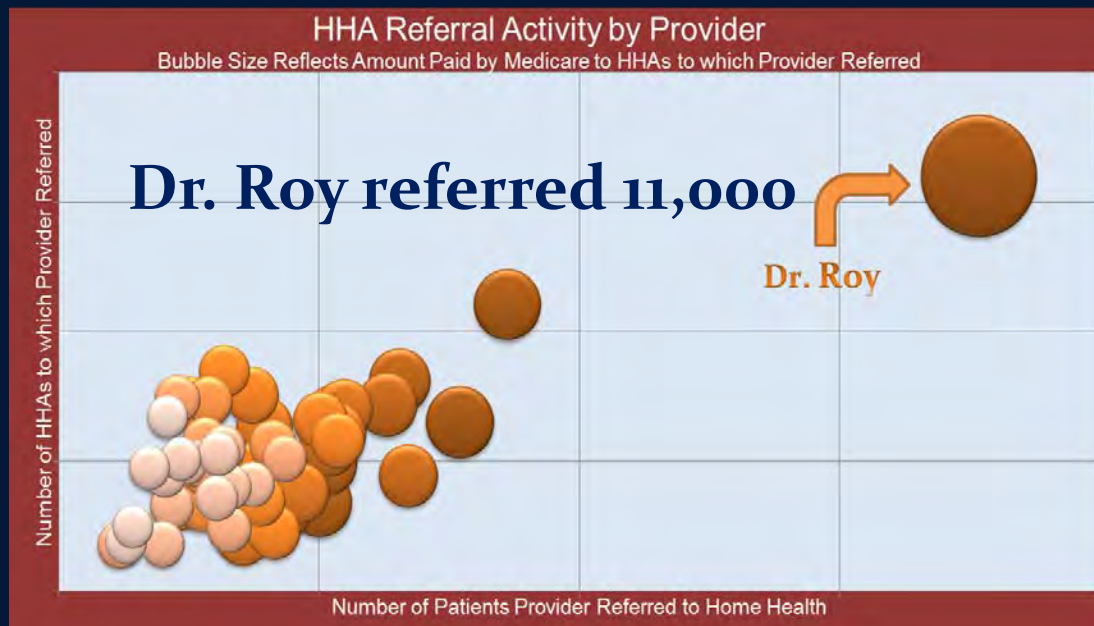
Media Contact

media@oig.hhs.gov
202-708-9855



Home Health Example

- Dr. Jacques Roy came to our attention based on analysis of suspicious billing
- Typical for a physician to refer <100 patients for Home Health



- Convicted at trial in April 2016; to be sentenced this fall





Three Charged in \$1 Billion Medicare Fraud Scheme

- Owner of more than 30 Miami-area skilled nursing homes and assisted living facilities, hospital administrator, and physician's assistant charged
 - Many of the beneficiaries did not qualify; **received medically unnecessary services** that were billed to Medicare and Medicaid
 - Co-conspirators employed sophisticated money laundering techniques to **hide the scheme and owner's identity**





Three Charged in \$1 Billion Medicare Fraud Scheme

“This is the largest single criminal health care fraud case ever ... and this is further evidence of how successful data-driven law enforcement has been as a tool in the ongoing fight against health care fraud”

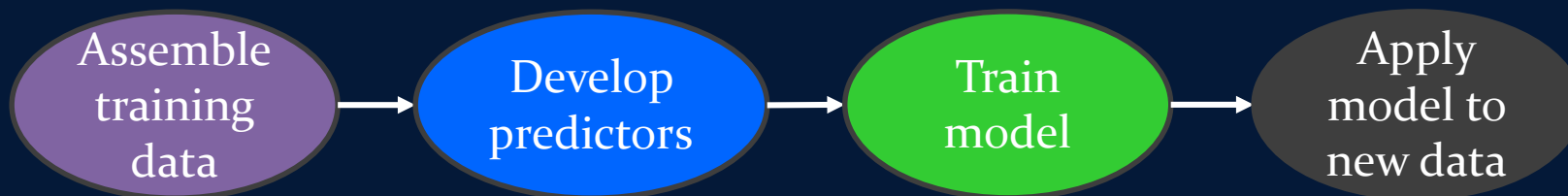
- Assistant Attorney General Leslie Caldwell



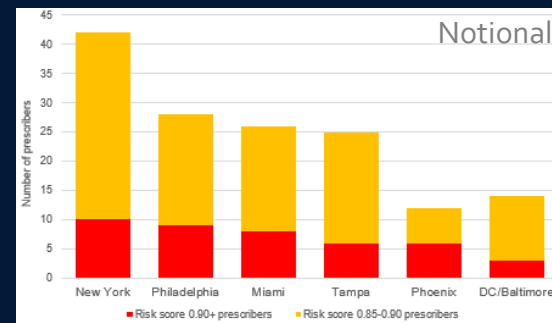
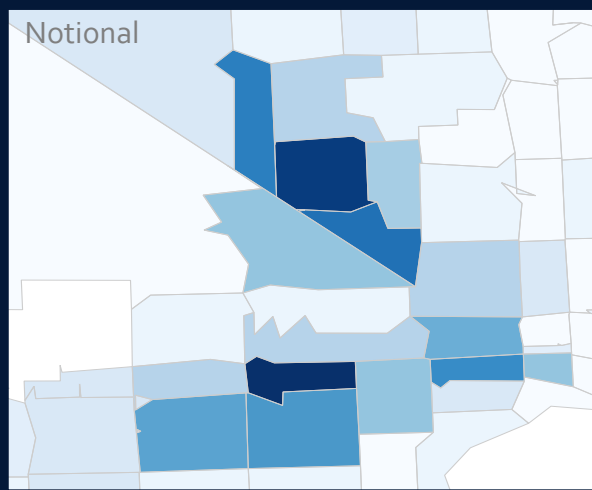


Predictive Modeling

- Developing statistical models to calculate “**risk scores**” for providers who may be defrauding Medicare



- Visualize and communicate the results





Our Challenge

How do we conduct anticipatory analysis – and be prepared for future vulnerabilities?

How do we make sure that our risk models keep up with intelligent, adaptive adversaries?

How do we better predict potential for fraud, waste, and abuse?





Stay Connected

The screenshot shows the official website of the Office of Inspector General, U.S. Department of Health & Human Services. At the top, there is a navigation bar with links for Home, FAQs, FOM, Careers, Contact, HEAT, and Download Reader. Below this is the main header with the OIG logo and a search bar. A secondary navigation menu includes links for About OIG, Reports & Publications, Fraud, Compliance, Recovery Act Oversight, Exclusions, and Newsroom. The main content area features a large article titled "OIG Examines FEDERAL MARKETPLACE CONTROLS Recommends Fixes for Enrollment Systems" with a "Learn More >>" link. To the right, there is a "I'm looking for" section with a search box and a dropdown menu. Below this are two promotional boxes: "EXCLUSIONS DATABASE" and "REPORT FRAUD". At the bottom, there are sections for "What's New" and "Latest Enforcement Actions" with several news items dated August 2015. A "Stay Connected" section offers links to Twitter, Podcasts, YouTube, and Widgets. A "Get Email Updates" button is also present.

oig.hhs.gov



@OIGatHHS



OIGatHHS





Questions



An aerial photograph of an airport terminal building, viewed from a high angle. The image is overlaid with several large, semi-transparent geometric shapes in shades of green, blue, and grey. The text is centered over the image.

Dr. Christopher Carrino

Census Case Study

Addressing Federal Priorities Through
Data Analytics

Addressing Federal Priorities Through Data Analytics

Census Case Study

Christopher Carrino, PhD

Chief Data Architect

US Census Bureau

Commerce Data Mission



- To promote the Commerce Department's and Federal data priorities, including open access and open data.



- To serve as the leading source of quality data about the nation's people and economy.



- To provide the most timely, relevant, and accurate economic accounts data in an objective and cost-effective manner.

The Past Week at Census



Census Bureau Submits Subjects for 2020 Census to Congress

March 28, 2017

Planned subjects for the 2020 Census, which include gender, age, race, ethnicity, relationship & homeownership status, were delivered to Congress.



Maricopa County Added Over 222 People Per Day in 2016

March 23, 2017

Maricopa County, Ariz., replaced Harris County, Texas, as the county with the nation's highest annual population growth.



The 2017 Census Test is Underway

March 22, 2017

This national test allows the U.S. Census Bureau to assess the integration of operations and systems needed for internet self-response.

Data Standards

WEP CSM GSM TDMA WiMAX USB IEEE 1394 CDMA JBI
CC/PP SOLARIS HPUX DGUX DYNIX IRIX OSF/1 CSS
CDM SMASH SMI-S LDAP KDM WTF GRDDL InkML
MathML OWL PICS DCML AVDL RDF SMIL SPARQL
MSXML BETAMAX HD-DVD BLUERAY BLUETOOTH VHS

**The best thing about standards is
there are so many to choose from.**

POP3 IMAP SMTP X.400 SVG URI/URL WAI WebCGM
XML AC DC NTSC PAL VISA American Express Minitel
Compuserve MOF XForms XHTML XLink CGA EGA VGA
VESA XML Processing Query XPath XPointer XSL XSLT UBL
WSDM-MUWS IMAP DCHCP WSRF WSS SAML XACML

Federal Statistics Data Standards

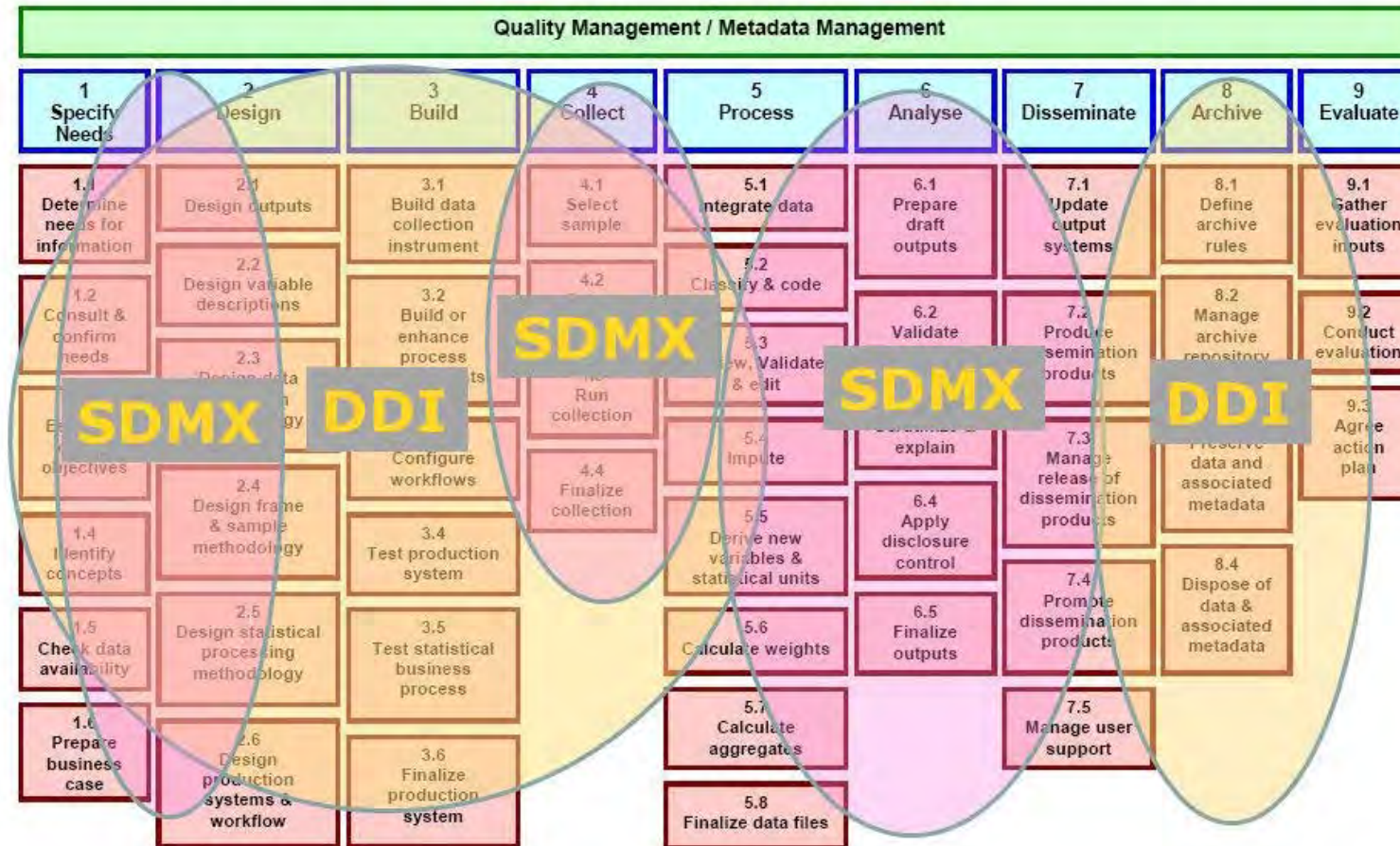
Acronym	Name	Governing Board
GSIM	Generic Statistical Information Model	UNECE
GSBPM	Generic Statistical Business Process Model	UNECE
SDMX	Statistical Data & Metadata eXchange	SDMX Board
ISO 11179	Information Technology - Metadata Registries	ISO/IEC
NIEM	National Information Exchange Model	NIEM ESC
DDI	Document, Discover & Interoperate	DDI Alliance
SPARQL	SPARQL Query Language for RDF	W3C
JSON-LD	JSON-LD	W3C

Sweden - September 3, 1967

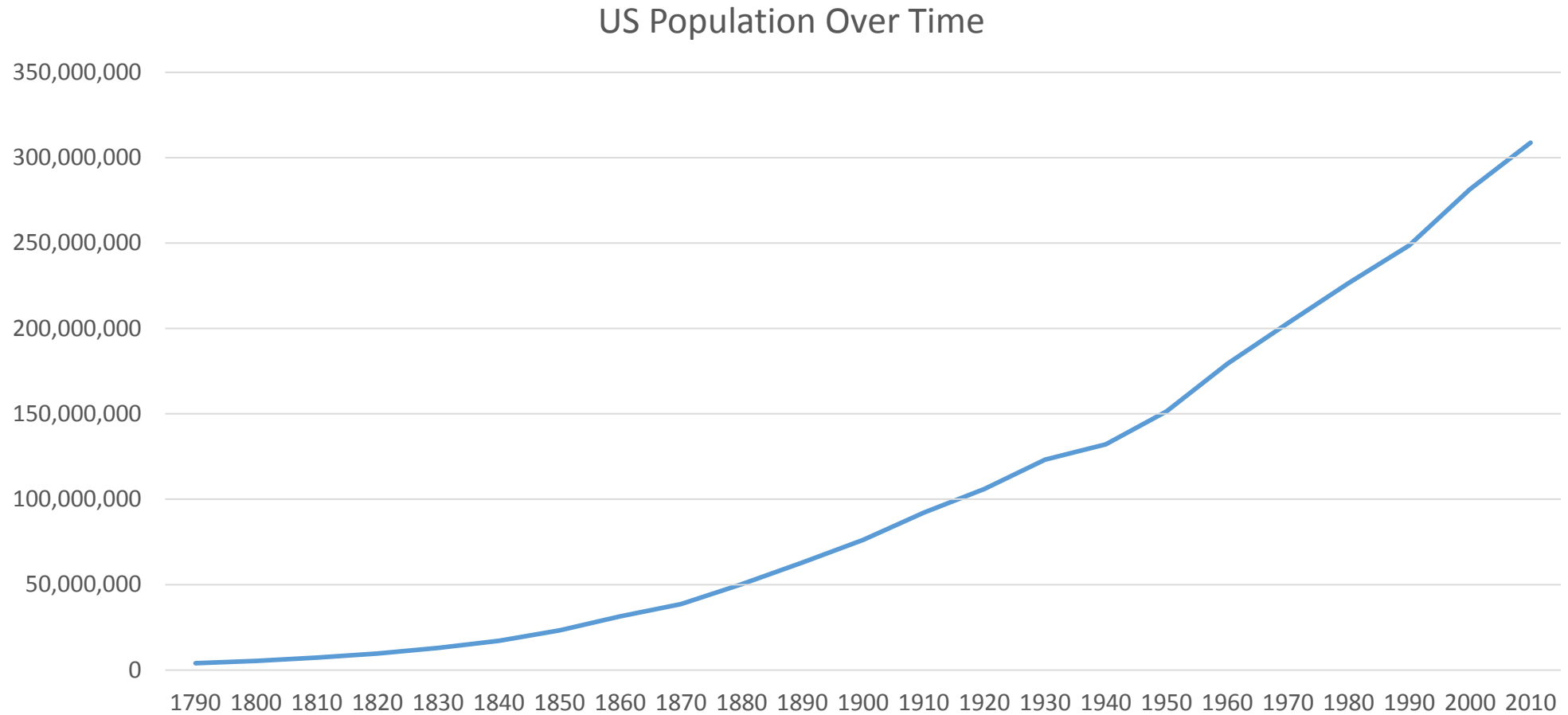


Managing Complexity

GSBPM, DDI and SDMX: towards a complete system?



Enabling Data Science through Data Standards



Thank You!



Jeff Chen

Commerce Data Service

Addressing Federal Priorities Through
Data Analytics

an ecosystem play



Jeff Chen
Acting Chief Data Officer
Commerce Data Service

Education

Privacy

Probability

Git

Open data

Experiment design

UI/UX

Sample design

Storage

Mobile Collection

Cognitive Testing

Statistical learning

Distributed Compute

Intake

Bulk

Provenance

Search

APIs

Data Management

Data processing

UI/UX

Schema development

Non Disclosure

Survey design

Prediction

Visualization





NOAA VIIRS
Satellite
Imagery



Severe
Weather Data
Inventory

NTIA
National
Broadband
Map



American
Housing Survey
+ American
Community
Survey

National
Economic
Accounts



NOAA
NEXRAD



IP +
Patents



3 25



Commodity
Flow Survey

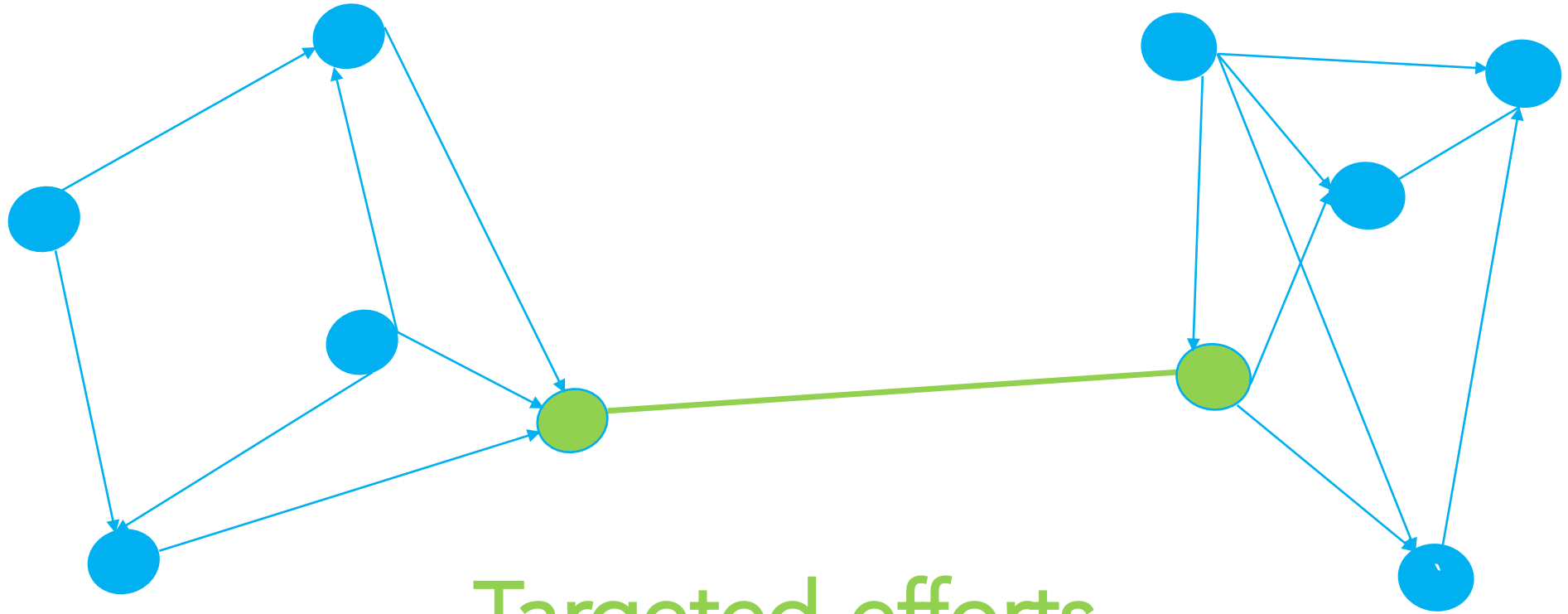


Census Foreign Trade
+ ITA Market Reports



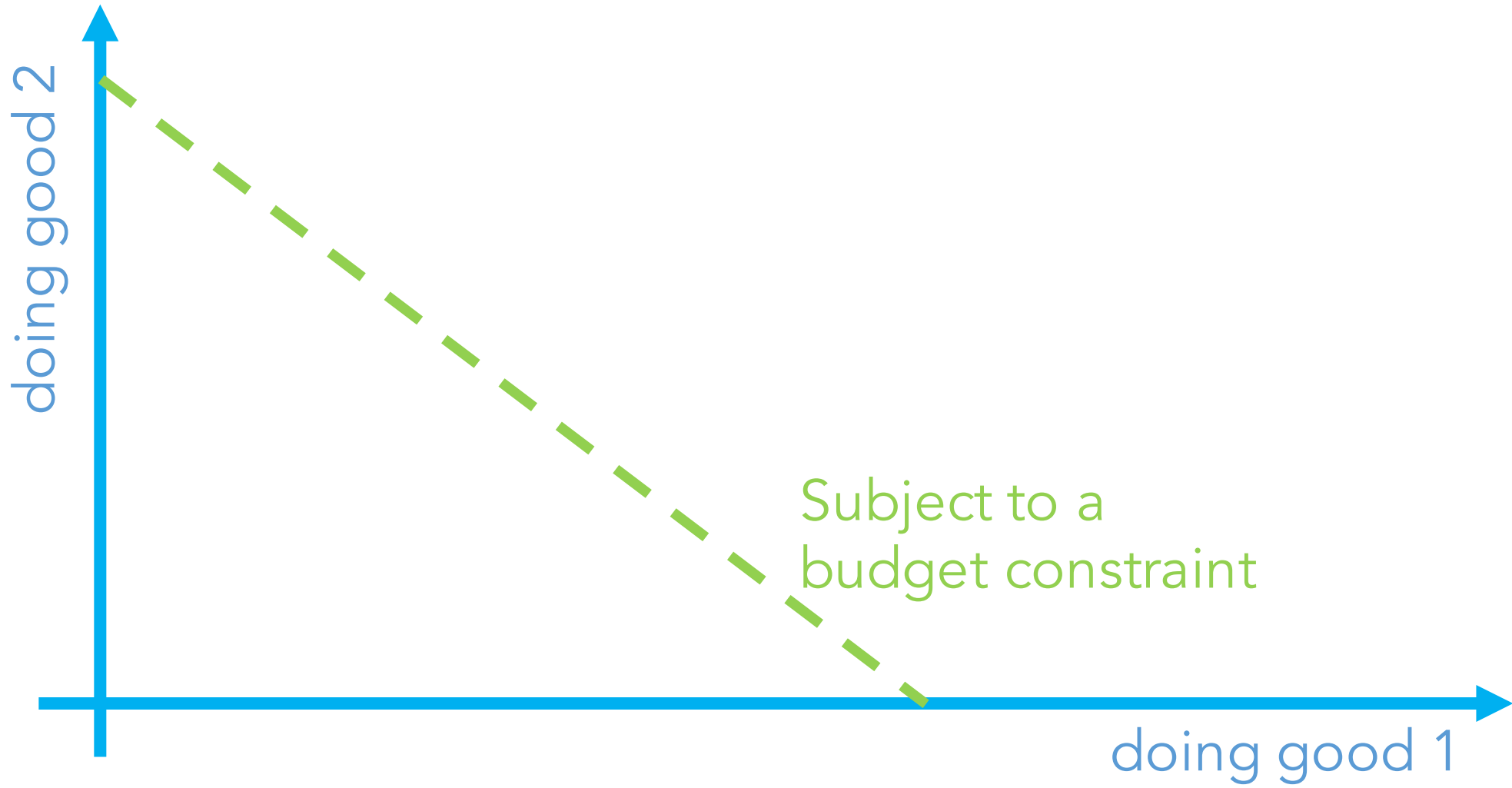
NOAA
Fishing
Landings





Targeted efforts









Exercise

on program that when
Python interpreter prints
Hello World ?

Exercise

Create a Python program that when
passed to the Python interpreter prints
Hello World ?

Exercise

Create a Python program that when
passed to the Python interpreter prints
Hello World ?

COMMERCE
RESEARCH
LIBRARY



Education First



Page Layout Formulas Data Review

Calibri (Body) 11 A A

Wrap Text Merge & Center

General \$ %

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

AutoSum Fill Clear

Sort & Filter

fx =IF(OR(ROUND(SIN(RADIANS(\$CK\$5)),0)=0),1,0)

Muahahaha!!

INSIGHT!

STATS!





COMMERCE.GOV



Search

Engage

Share



COMMERCE DATA ACADEMY

Educating and empowering Commerce employees.

[What We Offer](#)

[Current Courses](#)

[Previous Courses](#)

[FAQ](#)

[Contact Us](#)



Class Name	Date	Class Recording	Slides	Additional Materials
Intro to Agile Development	1/12/16	N/A	Presentation	
Introduction to HTML & CSS	1/14/16	N/A	Presentation	
Introduction to Storytelling with Data	1/19/16	Class Recording	Presentation	Data Set
Excel at Excel	1/28/16	N/A	Presentation	Answer Key
Data Science Basics	3/14/16	Class Recording	Presentation	
Git and GitHub	3/21/16	Class Recording	Presentation	
Intro to Python	4/6/16	Class Recording	Presentation	Answer Key
Intro to Design and Photoshop	4/19/16 and 10/13/16	Class Recording	Presentation	Class Materials
Intro to R	5/2/16	Class Recording	Presentation	
Intro to Data Analysis with R	5/17/16	Class Recording	Presentation	NBM Wireless CSV Sheet, Getting Into Data Science Resources, README.md File, Read-Me-CSV-June, Data Analysis e,
Data Storytelling with R	6/1/16	Class Recording	N/A	Step-by-Step Tutorial

dataacademy.commerce.gov

by the numbers

An award-winning, homegrown initiative to upskill DOC employees in data science, programming, and design.



Tech Program of the Year



Digital Government Achievement Award

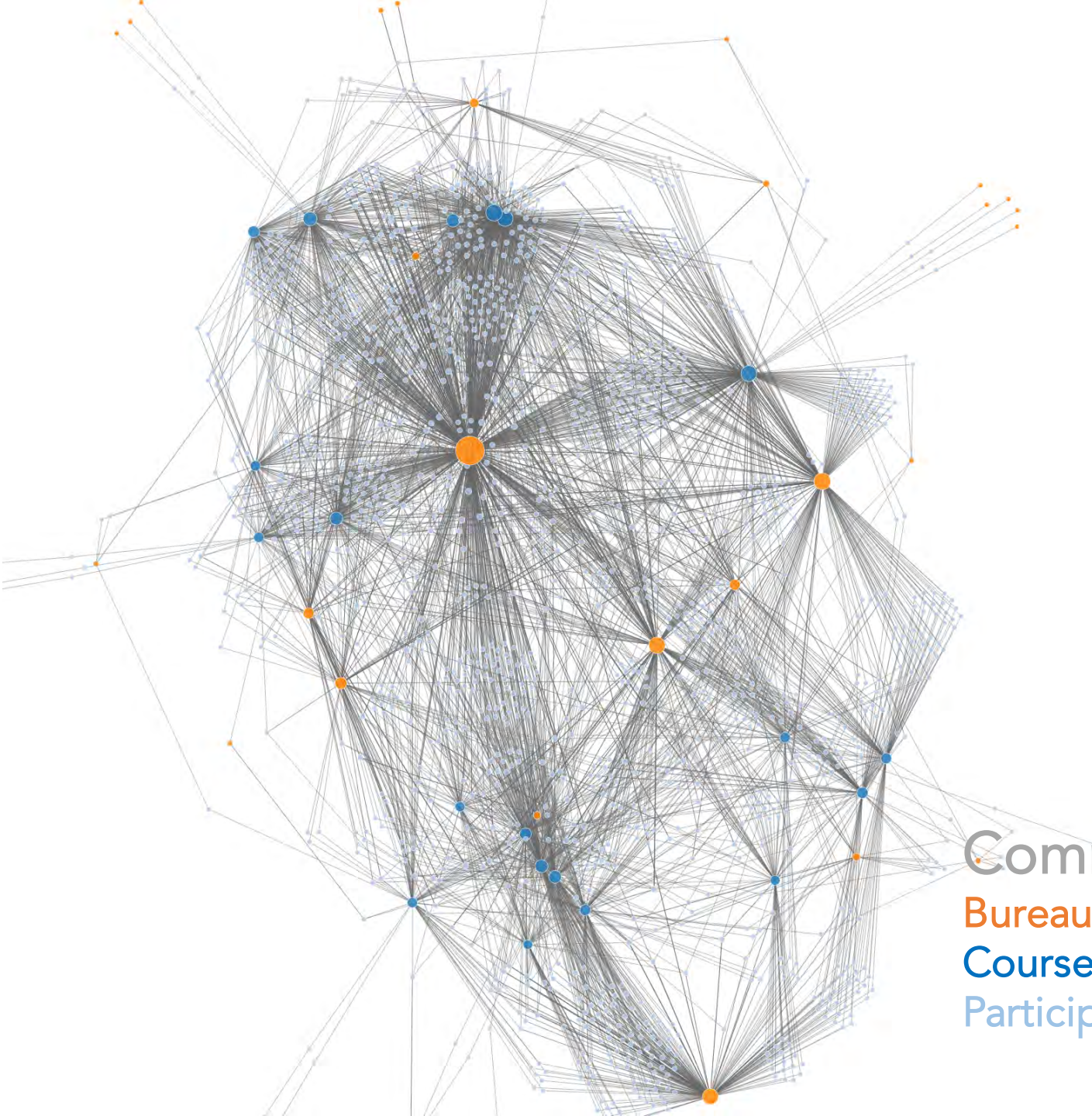


In-Residence Incubator

- 15 Participants from DOC bureaus (6 data science, 9 engineering)
- 6 Finished products in process of deployment
- 2 Products currently deployed

Short Courses

- 20 programming +tech courses
- 1400+ Commerce employees have taken at least one CDA class
- 1.9 Average number of classes taken per participant



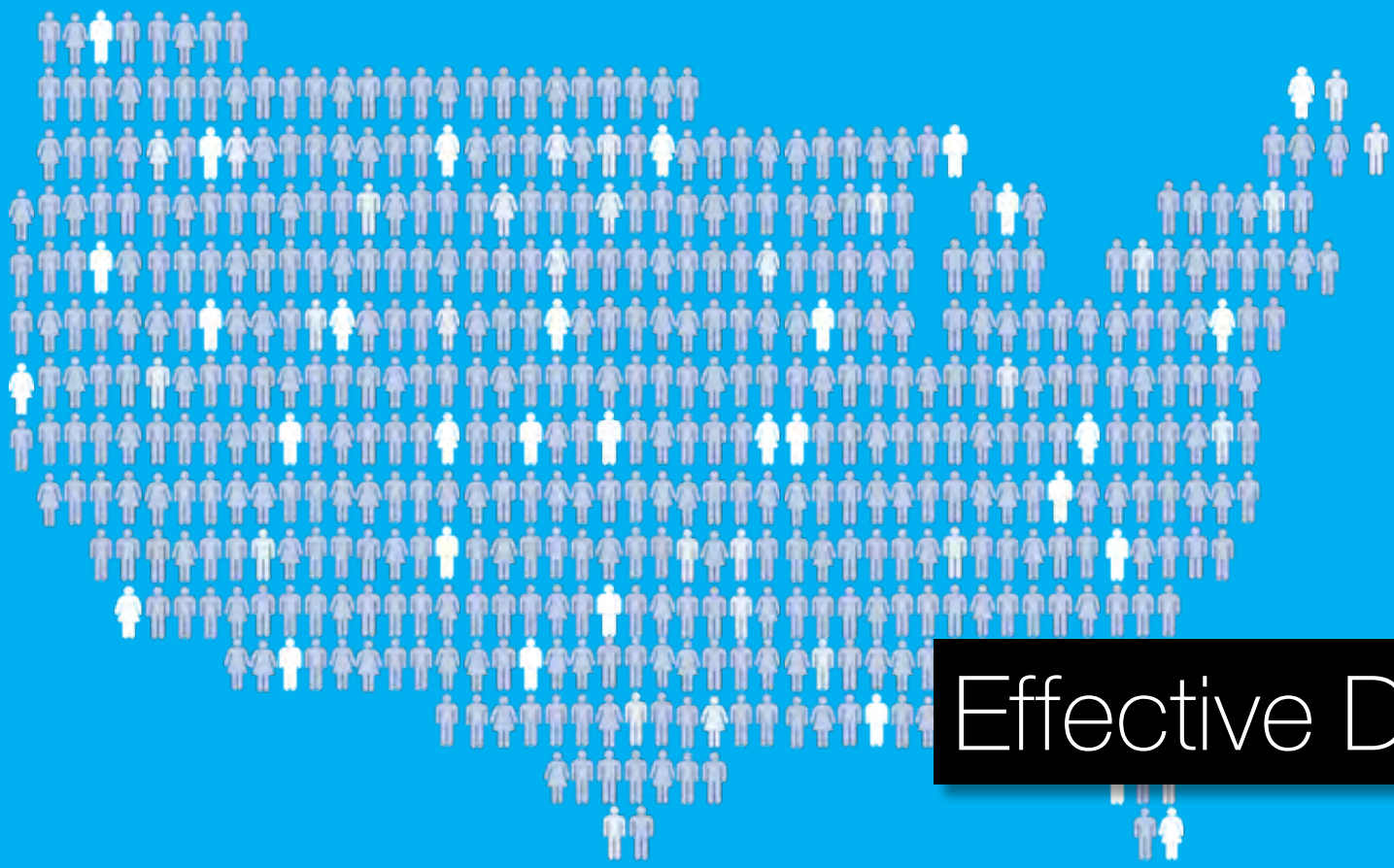
Commerce Data Academy

Bureaus (n=12)

Courses (n = 20)

Participants (n = 1460, ~2 classes/student)





Effective Data Use



Which organizations are most likely
to use government services?



MEP • MANUFACTURING
EXTENSION PARTNERSHIP

Specification

$$\Pr(Y = \textit{conversion} | \textit{features}) = f(\textit{revenue}, \textit{employment}, \textit{industry}, \dots)$$

Regularization - Lasso Estimate

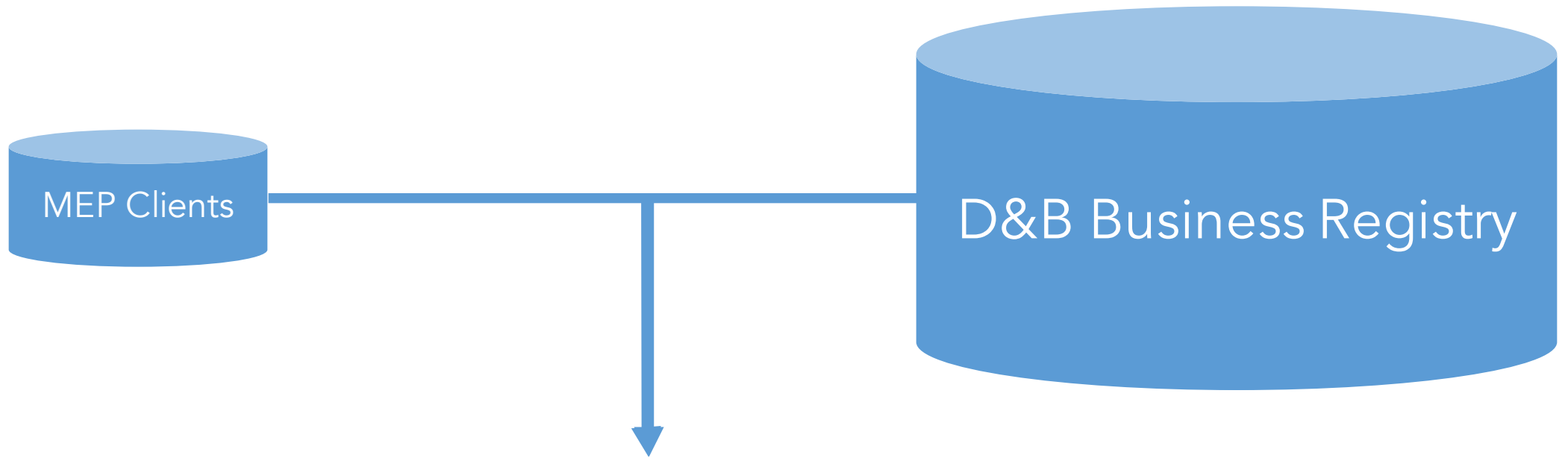
$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Loss function

l_2 penalty



MEP • MANUFACTURING
EXTENSION PARTNERSHIP



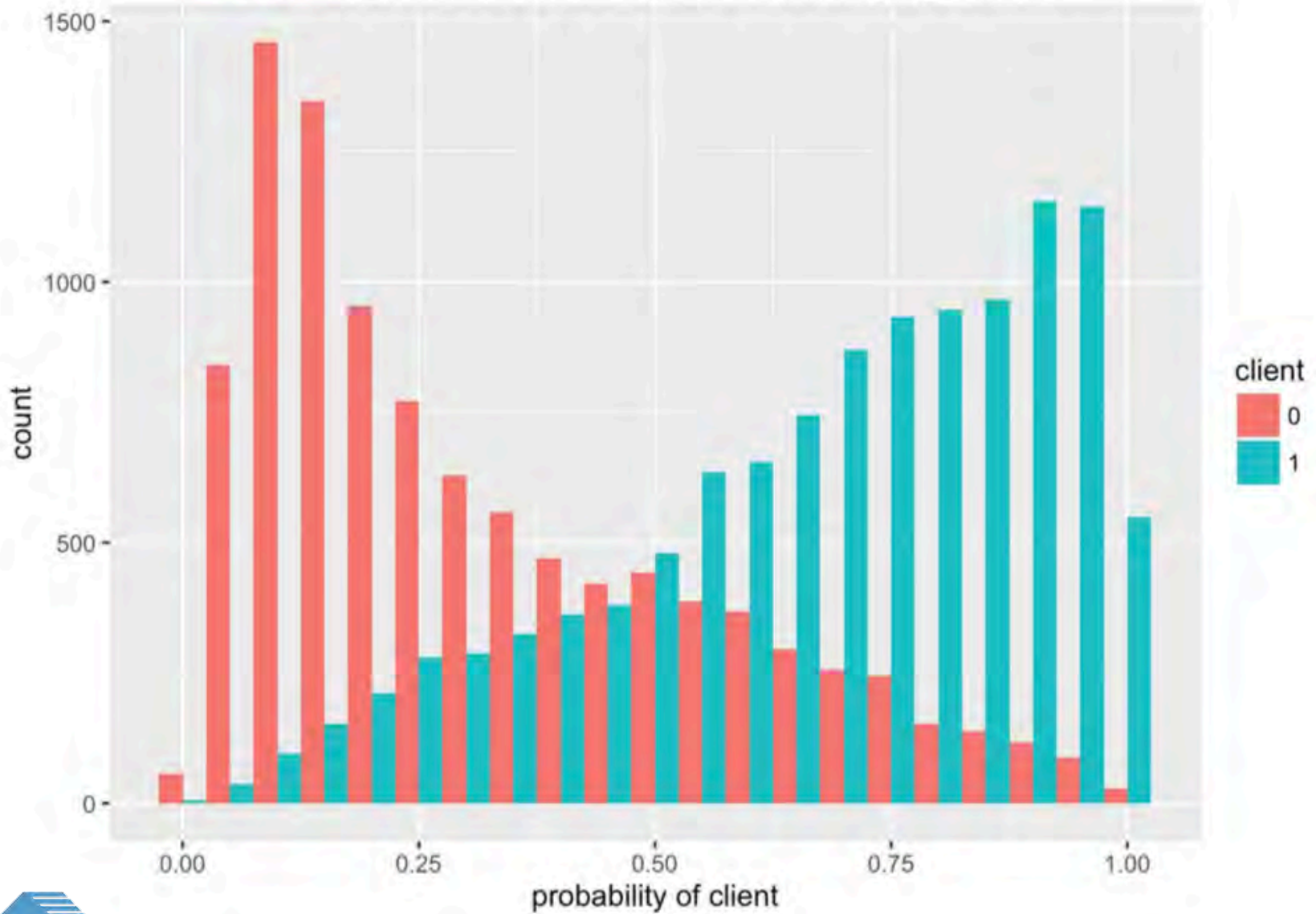
$$\hat{\beta} = \operatorname{argmin} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Pr(Conversion)

Factors that matter



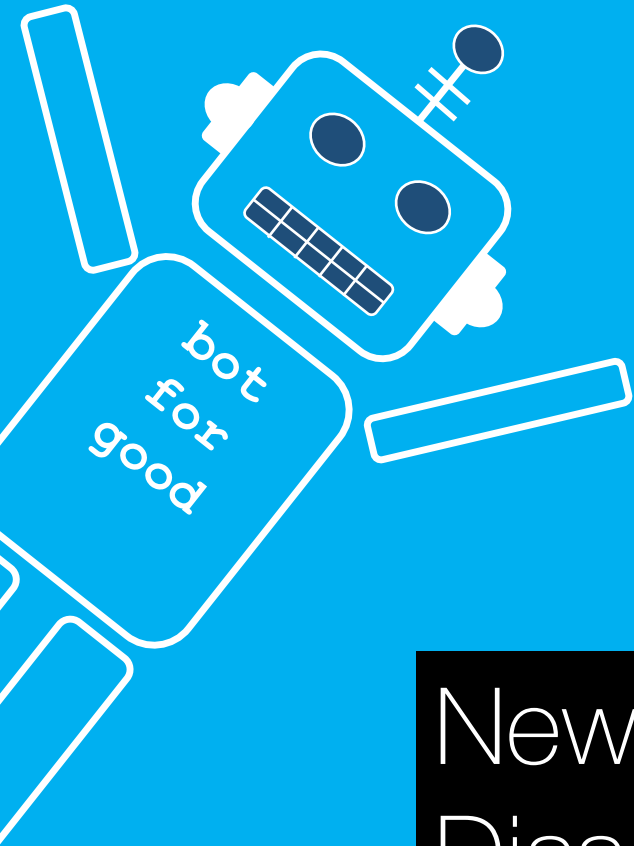
Predictive Model Performance in Identifying Clients



87%
CV Accuracy



MEP • MANUFACTURING
EXTENSION PARTNERSHIP



New Methods of Dissemination



WIKIPEDIA

The Free Encyclopedia

English

5 357 000+ articles

Español

1 322 000+ artículos

Deutsch

2 042 000+ Artikel

日本語

1 052 000+ 記事

Русский

1 380 000+ статей

Français

1 852 000+ articles

Italiano

1 342 000+ voci

中文

931 000+ 條目



Português

960 000+ artigos

Polski

1 212 000+ haseł

US Census Bureau|

EN ▾

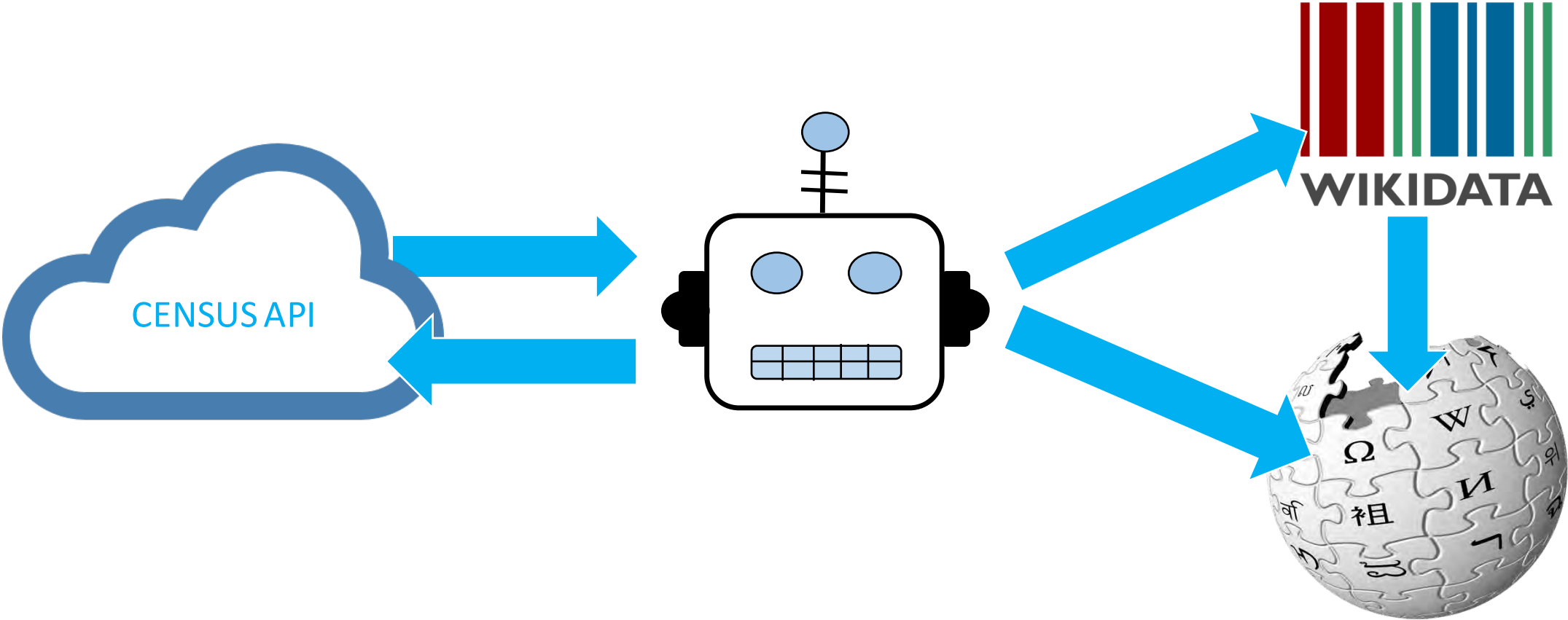


United States Census Bureau

bureau of the United States responsible for the census and related statistics



United States
Census
Bureau



United States™
Census
Bureau

WIKIPEDIA
The Free Encyclopedia

Arizona

From Wikipedia, the free encyclopedia

Coordinates: 34°N 112°W﻿ / ﻿34°N 112°W﻿ / 34; -112﻿ / 34; -112

This article is about the U.S. state of Arizona. For other uses, see Arizona (disambiguation).

Arizona (ⁱ/ˈærɪˈzoʊnə/; ⁱ/ærɪˈzoʊnə/) (Navajo: *Hoozdo Hahoodzo* [xò:ztò xàxò:tsò]; O'odham: *Alí ṣonak* [ˈali ˈʃonak]) is a state in the southwestern region of the United States. It is also part of the Western United States and the Mountain West states. It is the sixth largest and the 14th most populous of the 50 states. Its capital and largest city is Phoenix. Arizona is one of the Four Corners states. It has borders with New Mexico, Utah, Nevada, California, and Mexico, and one point in common with the southwestern corner of Colorado. Arizona's border with Mexico is 389 miles (626 km) long, on the northern border of the Mexican states of Sonora and Baja California.

Arizona is the 48th state and last of the contiguous states to be admitted to the Union, achieving statehood on February 14, 1912. Historically part of the territory of *Alta California* in New Spain, it became part of independent Mexico in 1821. After being defeated in the Mexican–American War, Mexico ceded much of this territory to the United States in 1848. The southernmost portion of the state was acquired in 1853 through the Gadsden Purchase.

Southern Arizona is known for its desert climate, with very hot summers and mild winters. Northern Arizona features forests of pine, Douglas fir, and spruce trees; the Colorado Plateau; some mountain ranges (such as the San Francisco Mountains); as well as large, deep canyons, with much more moderate summer temperatures and significant winter snowfalls. There are ski resorts in the areas of Flagstaff, Alpine, and Tucson. In addition to the Grand Canyon National Park, there are several national forests, national parks, and national monuments.

About one-quarter of the state^[8] is made up of Indian reservations that serve as the home of 27 federally recognized Native American tribes, including the Navajo Nation, the largest in the state and the United States, with more than 300,000 citizens. Although federal law gave all Native Americans the right to vote in 1924, Arizona excluded those living on reservations from voting until its state Supreme Court ruled in 1948 in favor of Native American plaintiffs.^{[9][10]}

Contents [hide]

- Etymology
- Geography and geology
 - Earthquakes
- Climate
- History
 - 20th century to present
- Demographics
 - Race and ethnicity
 - Languages
 - Important cities and towns

State of Arizona



Flag



Seal

Nickname(s): **The Grand Canyon State;**
The Copper State

Motto(s): **Ditat Deus (God enriches)**

State song(s): "The Arizona March Song" and
"Arizona"



Official language

English

Spoken languages

As of 2010

- English 74.1%
- Spanish 19.5%
- Navajo 1.9%
- Other 4.5%

Demonym

Arizonan^[1]

Population **Ranked 14th**

• **Total** **6,828,065 (2015 est)^[3]**

• **Density** **57/sq mi (22/km²)**
Ranked 33rd

• **Median household income** **\$52,248 ^[4] (33rd)**

— INFOBOX

— VALUE

Phase II

Step 1: Region vs. Industry

Step 2: Choose Multipliers

Step 3: View Shopping Cart

Checkout

Shopping
Cart

0

Step 2: Select Your Industry

You can select aggregate and/or detailed industries. ?

Select your industry type: ?

	SEARCH
1 Farms	<input type="button" value="ADD"/>
2 Forestry, fishing, and related activities	<input type="button" value="ADD"/>
3 Oil and gas extraction	<input type="button" value="ADD"/>
4 Mining, except oil and gas	<input type="button" value="ADD"/>
5 Support activities for mining	<input type="button" value="ADD"/> ●
6 Utilities*	<input type="button" value="ADD"/>
7 Construction	<input type="button" value="ADD"/>
8 Wood product manufacturing	<input type="button" value="ADD"/>
9 Nonmetallic mineral product manufacturing	<input type="button" value="ADD"/>
10 Textile mill product manufacturing	<input type="button" value="ADD"/>

ADD SELECTED
ITEMS TO CART

Industry Aggregate
4 Mining, except
oil and gas

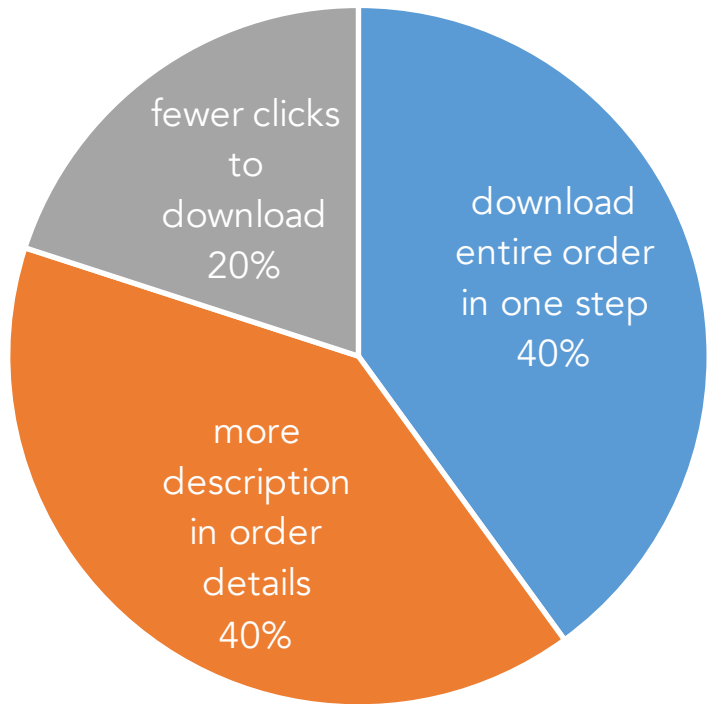
Improving
User Experience (UX)



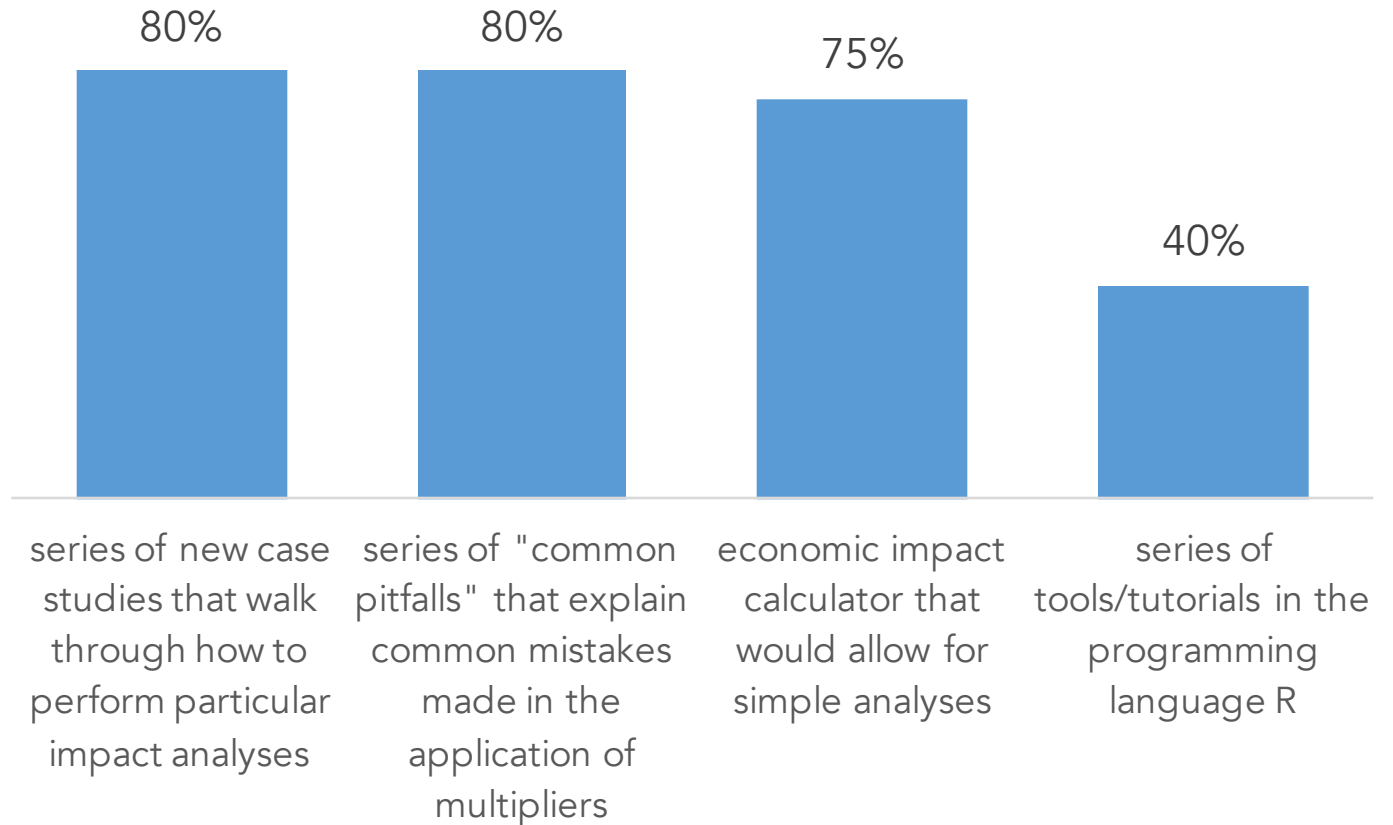
		INDUSTRIES	FINAL USE (GDP)			
			Personal consumption expenditures	Private fixed investment	Net exports	Government consumption expenditures and gross investment
COMMODITIES	Processing (See table 2.2)		Final Use			TOTAL COMMODITY OUTPUT
	Income					
VALUE ADDED	Compensation	Income				
	Taxes					
	Gross operating surplus					
		TOTAL INDUSTRY OUTPUT				









National Use Table

Top requested features
If you could make three improvements
to viewing/downloading multipliers,
what would they be?



% that responded tools would be either extremely useful or very useful



- [Sign In](#)
- [Order Multipliers](#)
-  [Shopping Cart](#)
-  [View Multipliers](#)
-  [Update Account](#)
-  [User's Guide](#)
-  [Sample Tables](#)
-  [Industry Lists](#)
-  [About RIMS](#)
-  [FAQ's](#)




[Step 1: Region vs. Industry](#) » [Step 2: Choose Multipliers](#) » [Step 3: View Shopping Cart](#) » [Step 4: Checkout](#)

Step 1: Region vs. Industry

RIMS II multipliers are based on 2007 national benchmark input-output data and 2015 regional data.

When you order by **Region** you receive multipliers for **All Industries** in that region including

- Multipliers cost \$275
- 369 detailed industries
- 64 industry aggregates



When you order by **Industry** you receive multipliers for **All 50 & the District**

- Multipliers cost \$75



[RIMS A](#)
[Things](#)



Jeff Chen
Acting Chief Data Officer
Commerce Data Service