Data Science for Big Data Analytics

Story
Data Science Data Publication for Federal Data Center Consolidation 2014
Data Science Data Publication for Performance.gov
Data Science Data Publication for FDA Data
Data Science Data Publication for FDA Data Visualization Gallery 1
Data Science Data Publication for FDA Data Visualization Gallery 2
Data Science Data Publication for National Science Board

Cover Page
Data Ecosystem
Global R&D: Measuring Commitment to Innovation
U.S. R&D: Funding and Performance
U.S. R&D: Federal Portfolio
U.S. and Global STEM Education
U.S. S&E Workforce: Trends and Composition
Research Outputs: Publications and Patents
Public Research Universities

Slides
Slide 1 Federal Big Data Initiative: Analytics for Data Science, Data Infrastructure, and Data Publications
Slide 2 Symposium on Predictive Analytics For Defense and Government, November 18-19, Washington, DC
Slide 3 Challenges and Solutions for Big Data in the Public Sector
Slide 4 Performance.gov
Slide 5 Performance.gov2014-Spotfire
Slide 6 Federal Data Center Consolidation Initiative: AOL Government Data Journalism
Slide 7 Federal Data Center Consolidation Initiative: Semantic Community Data Science Data Publication 2014
Slide 8 Strategic Use of Analytics in Government: Report
Slide 9 Strategic Use of Analytics in Government: An Analytic Data Center
Slide 10 OMB Data Visualization Tool Requirements Analysis: Gartner Magic Quadrant
Slide 11 OMB Data Visualization Tool Requirements Analysis: Results
Slide 12 Data Science Central: Big Data – From Descriptive to Prescriptive

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics
Updated: Wed, 23 Sep 2015 08:41:59 GMT
Powered by mindtouch
Slide 13 Data Science Central: Meteors Descriptive
Slide 14 Data Science Central: Meteors Predictive
Slide 15 Predictive Analytics World Government: 2013
Slide 17 Fourth Paradigm and Fourth Question
Slide 18 Federal Big Data Working Group Meetup: Mission Statement
Slide 19 Federal Big Data Working Group Meetup: What Are We Doing?
Slide 20 Federal Big Data Working Group Meetup: How Are we Doing it?
Slide 21 Federal Big Data Working Group Meetup: NIH Data Commons
Slide 22 Federal Big Data Working Group Meetup: NITRD FASTER
Slide 23 Federal Big Data Working Group Meetup: NSF Strategic Plan Knowledge Base
Slide 24 Data Science for NITRD-NSF-JHU DIBBs Project: Knowledge Bases
Slide 25 Data Science for NITRD-NSF-JHU DIBBs Project: Analytics & Visualizations
Slide 26 Data Science for NITRD-NSF-JHU DIBBs Project: Conclusions
Slide 27 Data Science Data Publication: FDA Data Mining Process
Slide 28 FDA-TRACK: Description
Slide 29 FDA-TRACK: Data
Slide 30 Data Science Data Publication: FDA Data in Spotfire
Slide 31 Cover Page-Performance Analytics: FDA TRACK
Slide 32 Content Analytics: Summary Statistics
Slide 33 Content Analytics: HealthData.gov Statistics 09172014
Slide 34 Content Analytics: FDA @ HealthData.gov
Slide 35 Network Analytics: FDA Glossary & Site Map
Slide 36 Data Analytics: FDA Drug Approvals & Databases
Slide 37 Data Analytics: FDA Drug Approvals & Databases Spreadsheet
Slide 38 Data Analytics: FDA Drug Approvals & Databases File Folder
Slide 39 Data Science Data Publication: National Science Board Process
Slide 40 Data Science Data Publication: National Science Board Spotfire

Research Notes
National Science Board
HealthData.gov
HealthCostInstitute Webinar
Data Sets for Office of the Assistant Secretary for Preparedness and Response
Followup from PAWG Today
PAWG2014 Slides
Symposium on Predictive Analytics For Defense and Government, November 18-19, Washington, DC

Predictive Analytics World Government 2014
Day 1: Monday, September 15, 2014
NSB Indicators Digest 2014

Cover Page

National Science Board

Chairs
Members
Member Ex Officio

Cover Page

Preface

Table of Contents

Introduction

Major S&E Indicators

What These Indicators Tell the Nation

GLOBAL R&D: MEASURING COMMITMENT TO INNOVATION

WHY IS THIS IMPORTANT?

A. HOW MUCH?

Figure A Estimated R&D expenditures worldwide: 1996–2011

B. WHERE?

Figure B1 Regional share of worldwide R&D expenditures: 1996 and 2011
Figure B2 Domestic R&D expenditures for selected countries: 1996–2011

C. GROWTH

Figure C1 Contributions of selected countries/regions/economies to growth of worldwide R&D expenditures: 2001–11
Figure C2 Average annual growth in domestic R&D expenditures of selected countries/ economies: 2001–11

D. INTENSITY

Figure D R&D expenditures as a share of economic output for selected countries/economies: 1996–2011

U.S. R&D: FUNDING AND PERFORMANCE

WHY IS THIS IMPORTANT?
A. TRENDS

Figure A U.S. R&D expenditures, by source of funds: 1990–2011

B. DEVELOPMENT AND APPLIED RESEARCH

Figure B1 Funding sources for U.S. development and applied research: 1990–2011
Figure B2 Performers of U.S. development and applied research: 1990–2011

C. BASIC RESEARCH

Figure C1 Funding sources for U.S. basic research: 1990–2011
Figure C2 Performers of U.S. basic research: 1990–2011

D. ACADEMIC R&D

Figure D Funding sources for U.S. academic R&D: 1990–2012

U.S. R&D: FEDERAL PORTFOLIO

WHY IS THIS IMPORTANT?

A. TRENDS

Figure A Federal spending on R&D: 1990–2011

B. FOCUS

Figure B Federal R&D budget, by national objectives: FYs 1990–2011

C. PERFORMERS

Figure C Federal spending on R&D, by performer: 1990–2011

D. TYPE OF WORK

Figure D Federal R&D funds, by type of work: 1990–2011

E. S&E FIELDS

Figure E1 Federal basic and applied research funds, by S&E field: 1990–2011
Figure E2 Inflation-adjusted change in federal basic and applied research funds, by S&E field: 2001–11

U.S. AND GLOBAL STEM EDUCATION

WHY IS THIS IMPORTANT?

A. K-12 MATHEMATICS AND SCIENCE

Figure A1 Average NAEP mathematics scores of U.S. students in grades 4 and 8: 1990–2011
Figure A2 Average mathematics and science PISA test scores of U.S. and OECD 15-year olds: 2003–09

B. U.S. BACCALAUREATES

Figure B U.S. bachelor's degrees in selected S&E fields per 1,000 20–24-year olds: 1991–2011

C. INTERNATIONAL BACCALAUREATES

Figure C1 Bachelor's degrees in natural sciences, by selected country/economy: 2000–10
Figure C2 Bachelor's degrees in engineering, by selected country/economy: 2000–10

D. INTERNATIONAL DOCTORATES

Figure D Doctoral degrees in natural sciences and engineering, by selected country/economy: 2000–11

U.S. S&E WORKFORCE: TRENDS AND COMPOSITION
WHY IS THIS IMPORTANT?

A. WORKFORCE GROWTH

Figure A Individuals in S&E occupations in the United States: 1960–2011

B. UNEMPLOYMENT

Figure B Unemployment rates for selected groups of workers: 1990–2012

C. WOMEN AND UNDERREPRESENTED MINORITIES

Figure C1 Men and women in S&E occupations: 1993, 2003, and 2010
Figure C2 Underrepresented minorities in S&E occupations: 1993, 2003, and 2010
Figure C3 Women and underrepresented minorities in S&E occupations: 2010

D. IMMIGRANTS

Figure D Foreign-born graduates whose highest degree is in S&E, by level and field of degree: 2010

RESEARCH OUTPUTS: PUBLICATIONS AND PATENTS

WHY IS THIS IMPORTANT?

A. PUBLICATIONS

Figure A S&E articles, by selected country/region/economy: 1997–2011

B. ENGINEERING ARTICLES

Figure B Engineering articles, by selected country/region/economy: 1997–2011

C. PATENTS

Figure C1 U.S. patents granted to non-U.S. inventors, by country/region/economy: 1995–2012
Figure C2 U.S. patents granted in ICT and health-related technologies: 2003 and 2012

D. SCIENCE-PATENT LINKAGE

Figure D U.S. patents that cite S&E literature, by ownership: 2003–12

PUBLIC RESEARCH UNIVERSITIES

WHY IS THIS IMPORTANT?

A. ENROLLMENT AND DEGREES

Figure A1 S&E degrees produced by research and other universities, by level of degree: 2011
Figure A2 Enrollment in research universities: 1990–2011

B. R&D

Figure B S&E R&D expenditures of research universities: 1990–2012

C. AFFORDABILITY

Figure C1 Tuition and state and local appropriations in U.S. public research universities: 1987–2010
Figure C2 Tuition and fees in U.S. research universities: 1987 and 2010

D. INVESTMENT IN HIGHER EDUCATION

Figure D Spending on higher education as a share of GDP for selected countries or country group: 2009

Glossary and Key to Acronyms

Applied research
Data Science Data Publications for Big Data Analytics

The National Science Board Science & Engineering Indicators Report says:

Nevertheless, the ongoing economic recovery has brought with it indications of emerging changes in S&E education and R&D. Potentially disruptive developments include the emergence of massive open online courses as an avenue for trying to reduce the cost of higher education and the continuing R&D budget uncertainty that accompanies a difficult fiscal environment. Source

Predictive Analytic World Government 2013 was very interesting and useful to me as a data scientist/data journalist. However, Predictive Analytics World Government 2014 did not seem to be as interesting and useful, but I am including it in my review of recent events on government big data analytics for an upcoming presentation at the Symposium on Predictive Analytics For Defense and Government. See my Slides below that contain examples (in process) as follows:

- Data Science Data Publication for Federal Data Center Consolidation 2014
- Data Science Data Publication for Performance.gov
- Data Science Data Publication for FDA Data
- Data Science Data Publication for FDA Data Visualization Gallery 1
- Data Science Data Publication for FDA Data Visualization Gallery 2
- Data Science Data Publication for National Science Board

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics

Updated: Wed, 23 Sep 2015 08:41:59 GMT

Powered by mindtouch
Our Federal Big Data Working Group Meetup is a massive open online course (MOOC) and this is an example of its tutorials and data science data publications in data browsers.

I am waiting to see the slides for the 6th Symposium on Big Data which contained some very good and relevant presentations for me to follow up on with the presentors.

MORE IN PROCESS

Data Science Data Publication for Federal Data Center Consolidation 2014

In 2011 and 2012, I published three stories on the Federal Data Consolidation Initiative because of the poor quality and incompleteness of the data. It was one of the first non-federal applications of analytics I did after leaving government service. I decided to revisit the data for this and was please to find that the quality and completeness had improved considerably and so I decided to import the new spreadsheet into Spotfire and explore the results in multiple dynamically linked adjacent visualizations.

Of the 3,665 data centers in the data set now, only 976 have been closed since the beginning of the program and 2,686 are yet to be closed in 2014-2015! The vast majority of these (2,254) belong to the Department of Agriculture.

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Data Science Data Publication for Performance.gov

The Knowledge Base (API), Spreadsheet (Excel), and Dashboard (Spotfire-in process below).

My Note: I was unable to retrieve the Indicators at: http://www.performance.gov/api/perfo...-indicator/csv. The message said "The Performance.gov web site is unavailable for viewing at this time due to a planned maintenance. We apologize for any inconvenience this may have caused", but it worked for the other 4. A principal result is the amount of missing data in the three data sets so there is a lack of preformance in submitting performance data!
Data Science Data Publication for FDA Data

Mined HealthData.gov for FDA data and built linked data *spreadsheets* (17) for Spotfire and the principal results were:

- Most programs do not have a Strategic Plan!
- Of the 5 HHS agencies that come under the Holdren Memo, CDC and FDA have by far the most and almost equal number of data sets!
- See how few of these data sets are in readily useable media!
- A Dashboard to the FDA Dashboards!
- The FDA Site Map and Glossary as a Linked Data Network!
- Inventory to prioritize further data science data publication work!
Data Science Data Publication for FDA Data Visualization Gallery 1

Mined FDA Site Map for data:

- Found Two: Data Standards and FDA Drug Approvals & Databases.
- Downloaded and inventoried files (41) (ZIP, CSV & XLS) for Spotfire.
- Used for FDA Data Innovation Lab Visualization Gallery.

See FDA Data Innovation Lab Visualization Gallery Slides

Conclusions

- We have participated in Meetups & Demos to understand the OpenFDA Data & the HHS Ignite Application & Evaluation Criteria.
- We have created an FDA Data Innovation Lab Visualization Gallery. There are some problems with the FDA Data sets.
- We are creating Data Science Data Publications for FDA using the Data Science Data Mining Process.
- Semantic Community has a platform for the community to interact with each other and FDA domain experts with the goal of spurring innovation around FDA data and creating new partnerships and opportunities between the public and private sector.

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

Data Science Data Publication for FDA Data Visualization Gallery 2

Mined FDA Site Map for data:

- Found Two: Data Standards and FDA Drug Approvals & Databases.
- Downloaded and inventoried files (41) (ZIP, CSV & XLS) for Spotfire.
- Used for FDA Data Innovation Lab Visualization Gallery.
The ndc package.txt and product.txt data sets were visualized in a separate Spotfire file and showed that HUMAN PRESCRIPTION DRUG (43,458 out of 83,167 rows, and most of the STARTMARKETINGDATE were after the Year 2000.

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Data Science Data Publication for National Science Board

See Knowledge Base (MindTouch: NSB Indicators Digest 2014), Spreadsheet (Excel), and Dashboard (Spotfire screen captures and Web Player below).

Following Tufte's recent suggestions, small adjacent dynamically linked visualizations were created in Spotfire and were captured to PowerPoint slides to make the slides much more than "chartjunk or chartoons". All of this is packaged as a data science data publication in a data browser for the NSB 2014 Science and Engineering Indicators Digest and the complete report.

Cover Page

Data Ecosystem
Global R&D: Measuring Commitment to Innovation

U.S. R&D: Funding and Performance

U.S. R&D: Federal Portfolio
U.S. and Global STEM Education

U.S. S&E Workforce: Trends and Composition

Research Outputs: Publications and Patents

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics
Updated: Wed, 23 Sep 2015 08:41:59 GMT
Powered by mindtouch
Public Research Universities

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

Slides

Slide 1 Federal Big Data Initiative: Analytics for Data Science, Data Infrastructure, and Data Publications

http://semanticommunity.info/

http://www.meetup.com/Federal-Big-Data-Working-Group/

http://semanticommunity.info/Data_Science/Federal_Big_Data_Working_Group_Meetup
# Slide 2 Symposium on Predictive Analytics For Defense and Government, November 18-19, Washington, DC

## Big Data, Analytics, and Data Science

<table>
<thead>
<tr>
<th>Big Data</th>
<th>Analytics</th>
<th>Data Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Content - Structured and Unstructured</td>
<td>Results and Decisions</td>
<td>Mining and Discovery</td>
</tr>
<tr>
<td>Data Ecosystem</td>
<td>Performance Content Network Data Descriptive Prescriptive</td>
<td>Microscope and Telescope: Szaley (JHU) Data FAIRport: Strawin (NSF) Data Commons: Bourne (NIH) Data Publications in a Data Browser: Semantic Community</td>
</tr>
</tbody>
</table>
Slide 3 Challenges and Solutions for Big Data in the Public Sector

Challenges and Solutions for Big Data in the Public Sector

- Performance.gov:
  - Federal Data Center Consolidation Initiative: AOL Government Data Journalism
- Strategic Use of Analytics in Government:
  - An Analytic Data Center: Not where but how you put your data
- OMB Data Visualization Tool Requirements Analysis:
  - Gartner Magic Quadrant for Business Intelligence and Analytics Platforms: Common data sets dynamically linked small adjacent multiples
- Data Science Central:
  - Big Data – From Descriptive to Prescriptive: Meteors
- Predictive Analytics World Government 2014:
  - Health Data, Regulatory, Law Enforcement, Medicare Savings, Financial Risk, etc.: HealthData.gov, OpenFDA and Whitehouse Holdren Memo
- Federal Big Data Working Group Meetup:
  - Data Science Data Publications: NITRD-NFS-IHUI DIBBs Project, FDA Data, and NSB Science & Engineering Indicators

Slide 4 Performance.gov

http://www.performance.gov/api

http://www.performance.gov/about

Performance.gov

- Success should be judged by results, and data is a powerful tool to determine results. We can’t ignore facts. We can’t ignore data.
  - President Barack Obama, July 24, 2009
- Performance.gov is a window to the Administration’s efforts to deliver a more effective, smarter, and leaner government. The site gives the public, agencies, members of Congress, and the media a view of progress underway in cutting waste, streamlining government, and improving performance.
  - Launched in 2009, public in 2010, with changes in leadership. Also PART (every 5 years), USA Spending.gov, and Data.gov.
- To learn about agency data-driven reviews click here. http://www.performance.gov/about

Slide 5 Performance.gov2014-Spotfire

Web Player
Slide 6 Federal Data Center Consolidation Initiative: AOL Government Data Journalism


Slide 7 Federal Data Center Consolidation Initiative: Semantic Community Data Science Data Publication 2014

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics

Updated: Wed, 23 Sep 2015 08:41:59 GMT
Powered by mindtouch
Slide 8 Strategic Use of Analytics in Government: Report

http://semanticommunity.info/Analytics/Strategic_Use_of_Analytics_in_Government

Strategic Use of Analytics in Government: Report

  - Analytics is the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.
  - Examples of Analytics in Government:
    - Health Care: Using Analytics to Prevent Fraud
    - Supply Chain: Using Analytics to Plan Capacity
    - Tax Collection: Using Analytics to Forecast Revenue
    - Intelligence: Using Analytics in Human Intelligence

http://semanticommunity.info/Analytics/Strategic_Use_of_Analytics_in_Government

Slide 9 Strategic Use of Analytics in Government: An Analytic Data Center

Jeffrey Hammerbacher

Is There a Business Case for Open Government Data?

http://semanticommunity.info/Analytics

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics
Updated: Wed, 23 Sep 2015 08:41:59 GMT
Powered by mindtouch
Slide 10 OMB Data Visualization Tool Requirements Analysis: Gartner Magic Quadrant

Magic Quadrant for Business Intelligence and Analytics Platforms

OMB Data Visualization Tool Requirements Analysis: Gartner Magic Quadrant

- Gartner Magic Quadrant (2013)
- Leaders in Business Intelligence and Analytics Platforms:
  - Birst
  - IBM
  - Information Builders
  - Logi Analytics
  - Microsoft
  - Microstrategy
  - Oracle
  - QlikView
  - SAP
  - SAS
  - Spotfire
  - Tableau
  - Alteryx
  - Power BI
  - Pentaho

Slide 11 OMB Data Visualization Tool Requirements Analysis: Results

Lessons for Managing Dashboards

Use of Dashboards in Government - Been There and Done That

Free Data Visualization and Analysis Tools
OMB Data Visualization Tool Requirements Analysis: Results

- **Background:**
  - Computerworld: 22 free tools for data visualization and analysis, Sharon Machlis, April 20, 2011.
  - Federal News Radio Interview: Lessons for Managing Dashboards, November 10, 2011. Also see: Use of Dashboards in Government - Been There and Done That
  - Computerworld: Chart and image gallery: 30+ free tools for data visualization and analysis, Sharon Machlis, March 15, 2013.

- **OMB Request:**
  - Expert familiar with the strengths and weaknesses of all these tools.

- **My Conclusion:**
  - Spotfire 5.5 is a comprehensive Tool that supports the 15 Categories, provides Multi-purpose Visualizations that are dynamically linked to one another, a Desktop and Browser-based Wapping Platform for all 4 Skill Levels, with Data Stored or Processed In-memory and In-Database, that is Designed for Web Publishing in the Cloud.

- **My OMB Data Visualization Tool Requirements Analysis:**
  - Used common data sets to see if 11 Gartner Magic Quadrant Leaders could produce dynamically-linked, small adjacent multiples (Ed Tufte). They could not!

---

Slide 12 Data Science Central: Big Data – From Descriptive to Prescriptive

Data Science Central

Slide 13 Data Science Central: Meteors Descriptive

Web Player
Predictive Analytics World
Government: 2014

Data Science for Big Data Analytics

- Health Data:
  - The Health Data Initiative: Past, Present, and Future
    - Daren Davis, Director for the Health Data Initiative, Chief Technology Office, HHS
- Regulatory:
  - Anticipatory Analytics: The Path Towards Future-Truth
    - Paul Saleen, Chief Data, Insight & Analytics Officer, Dun & Bradstreet
- Law Enforcement:
  - From Analytics to Action and Results – A Law Enforcement Perspective
    - Gary Canell, Deputy Inspector General for Investigations, Department of Health and Human Services, Office of Investigations (OIG)
  - Predictive Policing: Preventing Crime with Data and Analytics
    - Dr. Jennifer Bachner, Director, M.S. in Government Analytics, Johns Hopkins University
- Medicare Savings:
  - Predictive Analytics Savings and The Future
    - Elizabeth Handley, Deputy Operations Director, CMS Center for Program Integrity
- Financial Risk:
  - Proactive Risk Monitoring Using Advanced Analytics at Genevieve Mae
    - Gregory A. Kin, Senior Vice President and Chief Risk Officer, Ginnie Mae
- Intelligence:
  - Delivering Anticipatory Analytics at NGA
    - Kevin Aynes, Senior Analyst, National Geospatial Intelligence Agency (NGA)

Data Science for Big Data Analytics
Slide 17 Fourth Paradigm and Fourth Question

Fourth Paradigm and Fourth Question

- The Fourth Paradigm of Science (1):
  - First Paradigm. Observation, descriptions of natural phenomena, and experimentation.
  - Second Paradigm. Theoretical science such as Newton’s laws of motion and Maxwell’s equations.
  - Third Paradigm. Simulation and modelling, such as in astronomy.
  - Fourth Paradigm. Data-intensive science that exploits the large volumes of data in new ways for scientific exploration, such as the International Virtual Observatory Alliance in astronomy.
- The Fourth Question of Big Data for Science (2):
  - How was the data collected?
  - Where is the data stored?
  - What are the data results?
  - Does the data story persuade?


Slide 18 Federal Big Data Working Group Meetup: Mission Statement

Federal Big Data Working Group
Meetup: Mission Statement

- **Federal**: Supports the Federal Big Data Initiative, but not endorsed by the Federal Government or its Agencies;
- **Big Data**: Supports the Federal Digital Government Strategy which is “treating all content as data”, so big data = all your content;
- **Working Group**: Data Science Teams composed of Federal Government and Non-Federal Government experts producing big data products (How was the data collected; Where is it stored; What are the results; and Does the data story persuade?); and
- **Meetup**: The world’s largest network of local groups to revitalize local community and help people around the world self-organize like MOOCs (Massive Open Online Classes) being considered by the White House to reduce the cost of higher education.

Co-organizers: Brand Niemann and Katherine Goodier
Slide 19 Federal Big Data Working Group Meetup: What Are We Doing?

Federal Big Data Working Group Meetup: What Are We Doing?

- Leadership of the Semantic Data Science Team that produced Semantic Medline running on the Yarc Data Graph Appliance.
- Founding and co-organizing of the Federal Big Data Working Group Meetup.
- A graduate class prepared for GMU entitled “Practical Data Science for Data Scientists”.
- Using the Cross Industry Standard Process for Data Mining (CRISP-DM; Shearer, 2000) to build a Data Science Knowledge Base.
- Mining of the Data Science and Digital Earth scientific journals for the CODATA International Workshop on Big Data for International Scientific Programmes, June 8-9, in Beijing.
- Participation in the Data FAIRport (Findable, Accessible, Interoperable, and Reusable) with “Data Publication in Data Browsers”.
- Providing data stories that persuade and presentation materials for public education conferences like the COM Big Data Conference, August 4-6, in Washington, DC.

Slide 20 Federal Big Data Working Group Meetup: How Are we Doing it?

Federal Big Data Working Group Meetup: How Are we Doing it?

- **Federating Uses Cases**: Data Science (Brand Niemann); Environmental and Earth Science (Joan Aron); and Astronomy (Kirk Borne)
- **Federating Data Publications**: Structured Scientific Content (Papers, journals, books, reports, etc.); Data FAIRports (Findable, Accessible, Interoperable); NIH Data Commons, NITRD FASTER & NSF Strategic Plan (see next slides) and Reusable Data Stories That Persuade (Claims and Evidence)
- **Federating Solutions & Technologies**: Hand-Crafted by Individuals and Teams (Mary Galvin, STEM); Data Mining Standards and Products (Brand Niemann, Data Publications in Data Browsers); Machine Processing (Fredrik Salvesen, Semantic Data Publications on Yarc Data Graph Appliance); Reading and Reasoning (Katherine Goodier and Chuck Rehberg (Semantic Insights on Elsevier Content Text Mining); and Data Curation at Scale (Alan Wagner, Tamr on 1000s of Spreadsheets)

Slide 21 Federal Big Data Working Group Meetup: NIH Data Commons

http://semanticommunity.info/Data_Science/Data_Science_for_RDA#Slide_50_The_Power_of_the_Commons

My Note: Registries, Repositories, Clearinghouses, Portals, GitHubs, Data Commons, & Data FAIRportstoMindTouch and Spotfire
Federal Big Data Working Group Meetup: NITRD FASTER

- Faster Administration of Science and Technology Education and Research (FASTER) Community of Practice (CoP):
  - FASTER's goal is to enhance collaboration and accelerate agencies' adoption of advanced IT capabilities developed by Government-sponsored IT research. FASTER hosts Expedition and Emerging Technology workshops as well as monthly meetings with invited guest speakers to achieve this goal.
  - NITRD created FASTER for Federal agency CIOs and/or their advanced technology specialists. FASTER seeks to accelerate deployment of promising research technologies; share protocol information, standards, and best practices; and coordinate and disseminate technology assessment and testbed results. The Federal CIO Council under the leadership of the Office of Management and Budget (OMB) coordinates the use of IT systems. NITRD coordinates federally supported IT research under the leadership of OSTP (with OMB participation). FASTER, supported by the NITRD NCCO, communicates with OMB and the Federal CIO Council concerning IT R&D matters that are of general interest to Federal agencies.
  - FASTER is responding to the Open Government Directive by using the technologies of the Social Data Web (e.g., Linked Open Data and the Semantic Web).

Federal Big Data Working Group Meetup: NSF Strategic Plan Knowledge Base

http://semanticommunity.info/Data_Science/NSF_Strategic_Plan
Slide 24 Data Science for NITRD-NSF-JHU DIBBs Project: Knowledge Bases

Data Science for JHU DIBBs Project

SDSS.xlsx

Slide 25 Data Science for NITRD-NSF-JHU DIBBs Project: Analytics & Visualizations

Web Player
Data Science for NITRD-NSF-JHU DIBBs Project: Conclusions

- Science is increasingly driven by data (big and small)
- New instruments: “microscopes” & “telescopes” for data
- A major challenge on the “long tail”
- A new, Fourth Paradigm of Science is emerging...
- SDSS has been at the cusp of this transition
- Now the SciServer is continuing the legacy Gray’s Law of Data Engineering:
  - Scientific computing is revolving around data
  - Need scale-out solution for analysis
  - Take the analysis to the data!
  - Start with “20 queries”
  - Go from “working to working”

Source: Professor Szalay, August 19, 2014
Data Science Data Publication: FDA Data Mining Process

- **Steps**
  - See previous visualization and analytics for OpenFDA
  - Mined HealthData.gov for FDA Data and built linked data spreadsheets (17) for Spotfire
  - Mined fda.gov site map for data
  - Found 2: Data Standards and FDA Drug Approvals & Databases
  - Downloaded and inventoried files (13) (ZIP, CSV & XLS) for Spotfire
- Prepared Data Science Data Publication in Spotfire for FDA Briefing on September 23rd

FDA-TRACK: Description

* The Food and Drug Administration (FDA) runs FDA-TRACK, an agency-wide performance management program that monitors all 100+ FDA program offices’ key performance measures and highlighted projects. The acronym that FDA chose for this initiative succinctly captures key objectives of the Administration’s performance management approach: Transparency, Results, Accountability, Credibility, and Knowledge-Sharing. Program offices develop, monitor, and report the monthly measures, which are then analyzed by FDA’s Office of Planning and discussed during quarterly reviews with the agency senior leadership.

FDA-TRACK: Data

http://www.fda.gov/AboutFDA/Transparency/track/default.htm
Slide 30 Data Science Data Publication: FDA Data in Spotfire

Data Science Data Publication:
FDA Data in Spotfire

- Cover Page-Performance Analytics: FDA TRACK
- Content Analytics: Summary Statistics
- Content Analytics: HealthData.gov Statistics 09172014
- Content Analytics: FDA @ HealthData.gov
- Network Analytics: FDA Glossary & Site Map
- Data Analytics: FDA Drug Approvals & Databases

Slide 31 Cover Page-Performance Analytics: FDA TRACK

Web Player
Cover Page-Performance Analytics: FDA TRACK

Slide 32 Content Analytics: Summary Statistics

Content Analytics: Summary Statistics

My Note: Most programs do not have a Strategic Plan!

Slide 33 Content Analytics: HealthData.gov Statistics 09172014

Content Analytics: HealthData.gov Statistics

My Note: Of the 5 HHS agencies that come under the Holdren Memo, CDC and FDA have by far the most and almost equal number of data sets!
Content Analytics: FDA @ HealthData.gov

Network Analytics: FDA Glossary & Site Map
Network Analytics:
FDA Glossary & Site Map

My Note: The FDA Site Map and Glossary as a Linked Data Network!

Data Analytics:
FDA Drug Approvals & Databases

My Note: Inventory to prioritize further data science data publication work!

Data Analytics: FDA Drug Approvals & Databases Spreadsheet

Data Analytics:
FDA Drug Approvals & Databases Spreadsheet

My Note: This inventory is updated as one drills down into the data sets!

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics

Slide 38 Data Analytics: FDA Drug Approvals & Databases File Folder

Data Analytics:
FDA Drug Approvals & Databases File Folder

My Note: Some folders contain multiple files!

Slide 39 Data Science Data Publication: National Science Board Process

http://www.nsf.gov/statistics/seind14/
Data Science Data Publication: National Science Board Process

- Got NSB 2014 Science and Engineering Indicators Publication (an inch thick!)
  - See Conclusions (Pages O-21- O-22)
- See Web Site to download Spreadsheets
- Build data ecosystem for Spotfire Dashboard
  - Did previous dashboard

Slide 40 Data Science Data Publication: National Science Board Spotfire
Web Player

Research Notes

National Science Board

SCIENCE AND ENGINEERING INDICATORS 2014
A broad base of quantitative information on the U.S. and international science and engineering enterprise

http://www.nsf.gov/statistics/seind14/
HealthData.gov

http://www.hhs.gov/digitalstrategy/other-counting.html

Results 1 - 10 of about 10300 for data sets.

Office of the Assistant Secretary for Preparedness and Response

http://www.hhs.gov/about/orgchart/aspr.html

http://www.hhs.gov/about/foa/osleadership/aspr.html

http://www.phe.gov/about/aspr/pages/default.aspx

Search for Data Sets
1-10 of about 82 results Your search took 0.23 seconds.

HealthData.gov Spreadsheet Tabs Created for Spotfire

HealthCostInstitute Webinar

http://www.healthcostinstitute.org

Question: How is this different from HealthData.gov and the work of the HHS Idea Lab?

Comment: This seems premature until we can see some actual results.

Data Sets for Office of the Assistant Secretary for Preparedness and Response

Yesterday, Damon Davis of HHS mentioned in his talk at PAWG 2014 that 5 parts of HHS (CDC, FDA, AHRQ, NIH, and OASPR) have agreed to comply with the so-called “Whitehouse Science Advisor John Holdren Memo” to make their research data sets available to the public.

See: http://www.predictiveanalyticsworld.com/gov/2014/agenda.php#day2-815a

I have been able to find the data sets for the first four at HealthData.gov, but am unable to find your data sets.

Please send me a web link where I can find your data sets for my data science/data journalism work.

Thank you, Brand

Followup from PAWG Today

Damon, Is the healthdata.gov catalogue (says about 1600 data sets) available as a spreadsheet?
This is the first step in the data science data mining standard process to finding data sets that can be linked and integrated into a data science data publication for our future meetup.

Thank you, Brand

Damon, Enjoyed your presentation at PAWG today and wondered about the position I applied for. You said you thought someone had been selected so that probably answers my question – not me.

Linking data sets is now just a small part of a Data FAIRport (Dr. George Strawn) or a Data Commons (Dr. Phil Bourne) which we are doing for NSF (see Joint Meetup for NSF Data Scientists, Data Infrastructure, and Data Publication: http://www.meetup.com/Federal-Big-Data-Working-Group/events/201107182/ and NIH Semantic Medline (Dr. Tom Rindflesch): http://www.digitalgovernment.com/Events/Conferences/Government-Big-Data-Conference--Expo.shtml

Next week we meet on Open FDA (Dr. Taha Kass-Hout) because he liked the work we did for a recent Meetup: http://semanticommunity.info/Data_Science/Data_Science_for_OpenFDA

I am thinking a future Meetup should be a Data Science Data Publication for HealthData.gov, etc.

Best regards, Brand

---

**PAWG2014 Slides**

Dr. John Elder YES
Paul Ballew YES
Gary Cantrell YES
Elizabeth Handley NO
Kevin Ayers YES
Dr. Jennifer Bachner YES
Damon Davis YES

---

**Symposium on Predictive Analytics For Defense and Government, November 18-19, Washington, DC**

Dr. Brand Niemann, Director and Senior Data Scientist, Semantic Community; Co-organizer of the Federal Big Data Working Group Meetup; former Senior Enterprise Architect and Data Scientist, OCIO, US EPA: "Federal Big Data Initiative: Content, Network, and Data Analytics for NITRD/NSF Data Science, Data Infrastructure, and Data Publications"

We will also be able to include 3-5 short “bullet” items outlining the presentation:

Challenges and Solutions for Big Data in the Public Sector

Fourth Paradigm and Fourth Question

---

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics

Updated: Wed, 23 Sep 2015 08:41:59 GMT

Powered by mindtouch™
In the past, the problem of poor presentation quality has been one of the primary complaints from many seminar and conference attendees. The following suggestions are offered to guide you in the preparation of your presentation to ensure maximum quality and readability:

- **Presentations should be prepared in PowerPoint only.** We are not able to support other programs, such as Macintosh files. Also, we do not support the use of external videotapes. If you wish to use video, it should be embedded in your PowerPoint presentation.

- **The background color of your slide should be dark with white text, or a light color with black text.** For example, a dark blue background with orange text may not be legible to the audience in the back of the room. Remember – your presentation will be projected on a big screen to the audience in a large room. Too many colors at once may not produce the best quality of slide.

- **The recommended point size for titles is 36 pt.; the recommended text size is 24 pt.,** with sub-bullets no smaller than 18 pt. Deviating from these sizes will cause your text to be too small to be read from the back of the room, or a title that is too large for the space allocated.

- **There should be a uniform and consistent look to your presentation.**
  - Use a similar title and format for each section and slide.
  - Do not change typefaces from slide to slide. To add variety and/or emphasis to your presentation, italics or boldface type may be used rather than switching fonts.

- **Try not to “cover the waterfront.”** Comprehension is more important than volumes of data.
  - Too much information, illustrations or graphics on one slide is confusing.
  - Each slide should have no more than 5 bullet points on it for maximum readability.
  - Frame your pages so that the audience will know where to look.

- **The presentation should be in one “voice,”** and not appear to be taken from multiple sources (unless you cite and quote a source).

- **All presentations should be paginated and in a horizontal format.**

- **Typographical errors will reduce the impact of your presentation.** Be sure to verify the accuracy of your acronyms and any jargon used. **If you are using a lot of acronyms within your presentation, please supply a list of acronyms at the end of your presentation for inclusion in the documentation.**

- **If you are planning on bringing a back up copy of your presentation to the conference (which we suggest), please bring only “USB” connection thumb drives or a cdrom so we can easily load any updates you may have.**

- **The preferred method to submit your presentation is via email. If your presentation is extremely large (greater than 5MB), please follow the uploading instructions below.**
Go to this web site to upload: [http://www.ttcusfiles.com/ttpus/](http://www.ttcusfiles.com/ttpus/)
1. Fill out the blanks noted with asterisks*
2. Note file name. (Your Name & What you call your briefing)
3. Select browse button so you can find your file
4. Send File
5. A thank you message will appear and an email notification will be sent alerting us of the file upload.

Predictive Analytics World Government 2014

Source: [http://www.predictiveanalyticsworld...a_overview.php](http://www.predictiveanalyticsworld...a_overview.php)

### Day 1: Monday, September 15, 2014

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:15-8:15am</td>
<td>REGISTRATION &amp; BREAKFAST</td>
<td>Constitution Foyer</td>
</tr>
<tr>
<td>8:15-8:30am</td>
<td>WELCOME REMARKS</td>
<td>Constitution B</td>
</tr>
<tr>
<td>8:30-9:30am</td>
<td>KEYNOTE • Room: Constitution B</td>
<td>Constitution B</td>
</tr>
<tr>
<td></td>
<td>The Peril of Vast Search (and How Target Shuffling Can Save Science)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dr. John Elder, CEO &amp; Founder, Elder Research, Inc.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slides</td>
<td></td>
</tr>
<tr>
<td>9:30-10:00am</td>
<td>Sponsored Session • Room: Constitution B</td>
<td>Constitution B</td>
</tr>
<tr>
<td></td>
<td>Anticipatory Analytics: The Path Towards Future Truth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paul Ballew, Chief Data, Insight &amp; Analytics Officer, Dun &amp; Bradstreet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slides</td>
<td></td>
</tr>
<tr>
<td>10:00-10:35am</td>
<td>BREAK • Room: Constitution E</td>
<td>Constitution E</td>
</tr>
</tbody>
</table>

**Track 1**
- Room: Constitution A
- [Finding Patterns of Emergence](#)

**Track 2**
- Room: Constitution B
- [Developing an Analytical Framework & Measuring Risk at the SBA](#)
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:20am-12:05pm</td>
<td><strong>From Analytics to Action and Results – A Law Enforcement Perspective</strong></td>
</tr>
<tr>
<td></td>
<td>Gary Cantrell, Deputy Inspector General for Investigations, Department of Health and Human Services, Office of Investigations (OI) <a href="#">Slides</a></td>
</tr>
</tbody>
</table>
| 12:05-1:00pm      | **LUNCH**  
|                   | Room: Constitution E  
|                   | **Sponsored by:**  
|                   | [Dell Software](#) |
| 1:00-1:45pm       | **KEYNOTE**  
|                   | Room: Constitution B  
|                   | [Predicting Predictive Analytics Savings and The Future](#)  
|                   | Elisabeth Handley, Deputy Operations Director, CMS Center for Program Integrity Slides Not Available |
| 1:45-2:45pm       | **PLENARY SESSION**  
|                   | Room: Constitution B  
|                   | [Abstract to Concrete: Demystifying the Model Building Process](#)  
|                   | Gerhard Pilcher, VP and Senior Scientist, Elder Research, Inc. |
| 2:45-3:00pm       | **Sponsored Session**  
|                   | Room: Constitution B  
| 3:00-3:45pm       | **BREAK**          |
| 3:45-4:30pm       | **Track 1:**  
|                   | Room: Constitution A  
|                   | [Finding a Lost Person and Evaluating State Budget - Case Studies in Government Analytics](#)  
|                   | Dr. Aric LaBarr, Assistant Professor of Analytics, Institute for Advanced Analytics, NCSU |
|                   | **Track 2:**  
|                   | Room: Constitution A  
|                   | [Geospatial Predictive Analytics Case Studies: Northern VA Sniper and LRA Leader Joseph Kony](#)  
|                   | Dr. Colleen (Kelly) McCue, Senior Director, Social Science & Quantitative Methods, DigitalGlobe |
4:30-5:15pm  | **Case Study: Chicago Department of Public Health**  
| Transforming Food Inspections Using Predictive Analytics  
| Jay Bhatt, Managing Deputy Commissioner, Chicago Department of Public Health  

| 5:15-6:15pm | **NETWORKING RECEPTION**  
| Room: Constitution E |

---

### Day 2: Tuesday, September 16, 2014

| 7:30-8:15am | **REGISTRATION & BREAKFAST**  
| Room: Constitution Foyer |

| 8:15-8:20am | **WELCOME REMARKS**  
| Room: Constitution B |

| 8:20-9:05am | **KEYNOTE**  
| Room: Constitution B  
| The Health Data Initiative: Past, Present, and Future  
| Damon Davis, Director for the Health Data Initiative, Chief Technology Office, HHS  
| [Slides](http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics) |

| 9:05-9:50am | **PRACTITIONER'S PANEL**  
| Room: Constitution B  
| Applying Analytics in the Regulatory Environment  
| Gregory Keith, Senior Vice President & Chief Risk Officer, Ginnie Mae  
| Christina Wolf, Chief Economist, The British Columbia Securities Commission  
| Elaine Ayo, Graduate Student, Georgetown University |

| 9:50-10:20am | **Sponsored Session**  
| Room: Constitution B  
| Acceptable Big Data Analytics: Required Features for Successful Deployments  
| Sanjeev Kumar, Director of Analytics and Business Intelligence Solutions Innovation, Dell Software Group |

| 10:20-10:50am | **BREAK** |

---

**Track 1**  
Room: Constitution A  
Track Sponsored by

**Track 2**  
Room: Constitution A  
Track Sponsored by

[http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics](http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics)  
Updated: Wed, 23 Sep 2015 08:41:59 GMT  
Powered by [mindtouch](https://www.mindtouch.com)
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Presenter/Details</th>
</tr>
</thead>
</table>
| 10:50-11:35am| **Case Study: Regulatory Oversight Using Predictive Risk Models at the British Columbia Securities Commission**  
Christina Wolf, Chief Economist and Director of Economic Analysis, British Columbia Securities Commission | **Building a Data-centric Strategy & Roadmap**  
Dr. Peter Aiken, Founding Director, Data Blueprint  
Lewis Broome, CEO, Data Blueprint |
| 11:35am-12:20pm| **Proactive Risk Monitoring Using Advanced Analytics at Ginnie Mae**  
Gregory A. Keith, Senior Vice President and Chief Risk Officer, Ginnie Mae | **Predictive Policing: Preventing Crime with Data and Analytics**  
Dr. Jennifer Bachner, Director, MS in Government Analytics, Johns Hopkins University [Slides](#) |
| 12:20-1:15pm | **LUNCH**  
Room: Constitution E  
*Sponsored by:* [D&B](#) | **LUNCH & LEARN**  
Room: Constitution B  
Optimizing Decisions with Information Builders and Predictive Analytics  
Eric Merkel, Senior Systems Engineer, Federal Systems Group, Information Builders  
*Sponsored by:* [Information Builders](#) |
| 1:15-2:15pm | **KEYNOTE**  
The Need for Analytic Innovation and Managing Barriers  
Dean Silverman, Senior Advisor to the Commissioner (Compliance Analytics Initiatives), IRS  
Moderator: Jason Miller, Federal News Radio | |
| 2:15-3:00pm | **Transforming Government Oversight: Collaborative Analytics**  
Bryan Jones, Deputy Assistant Inspector General, Analytics, USPS-OIG | **Enabling IT Risk Management with Analytics**  
Gleb Reznik, Director - Risk, Continuity and Security, Fannie Mae  
Moderator: Jason Miller, Federal News Radio |
| 3:00-3:30pm | **BREAK** | |

[http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics](http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics)

Updated: Wed, 23 Sep 2015 08:41:59 GMT
Powered by [mindtouch](#)
<table>
<thead>
<tr>
<th>Track 1</th>
<th>Track 2</th>
</tr>
</thead>
</table>
| **3:30-4:15pm** | **Delivering Anticipatory Analytics at NGA**  
Kevin Ayers, Senior Analyst, National Geospatial Intelligence Agency (NGA)  
Michael Falkowski, US Federal - Defense, Intelligence, & Justice, IBM Advanced Analytics [Slides](#) |
| **Big Data for Small(er) Campaigns**  
Daniel Porter, Co-Founder, BlueLabs | **The Role of the Interpreter: Standing between Vision Leaders and Action Technicians are the Interpreters - Critical for Successful Analytics Deployments**  
LTC Gary L. Good, Office of Business Transformation, Under Secretary of the Army |
| **4:15-5:00pm** | **Utah HR's Journey into Predictive Analysis**  
Lowe Rudd, Business Intelligence Consultant, Department of Human Resource Management, State of Utah  
Bart Olsen, Human Resource Director, Business Intelligence & HRIS, State of Utah  
Jeremy Marsh, Business Intelligence Consultant, Department of Human Resource Management, State of Utah |
| **6:30-9:30pm** | **Big Data Roundup**  
Location: Capitol City Brewing Company |

**Post-Conference Workshop: Wednesday, September 17, 2014**

**Full-day Workshop**
The Best and the Worst of Predictive Analytics: Predictive Modeling Methods and Common Data Mining Mistakes  
Dr. John Elder, CEO & Founder, Elder Research, Inc.

**Post-Conference Workshop: Thursday, September 18, 2014**

**Full-day Workshop**
Making Text Mining Work: Practical Methods and Solutions  
Dr. Andrew Fast, Chief Scientist, Elder Research Inc.  
Sponsored by:
6th Symposium on Big Data

Source: http://www.bigdataevent.net/agenda-big-data-defense/

Applications and Analytics for Defense and National Security
September 23-24, 2014
Holiday Inn Rosslyn at Key Bridge, Arlington, VA

Keynotes: 2

I. Government Needs, Initiatives, Opportunities and Challenges: 7

II. The Latest Tools, Techniques and Technologies: 5

III. High-Value Applications for Defense and Intelligence – Emerging Lessons Learned: 7

Other: I-2, III-1

Day One: September 23, 2014

9:00-9:05 Administrative Announcements; Moderator Introduction (Mr. Alan Briggs, Elder Research)

9:05-9:35 MR. JEFFREY EGGERS, SES, Chief Technology Officer, Office of the Deputy Chief of Staff for Intelligence, Surveillance and Reconnaissance, US Air Force (AF/A2D)
Keynote: “Big Data and the Air Force ISR Mission” Slides

9:35-10:05 REAR ADMIRAL (ret) ALEX MILLER, Director, Federal Consulting, Deloitte Consulting; former Chief of Staff, National Security Agency
Keynote: “Architecture to Support Advanced Analytics” Slides

10:05-10:35 MS. LAURA ODELL, Information Technology Systems, Institute for Defense Analyses; DR. ARUN MAIYA, Senior Staff Member, IDA, and MS. L. HUTCHINSON, Chief Information Officer, Office of the Secretary of Defense
II Special Presentation: “Using Natural Language Processing and Automated Machine Learning to Turn Extreme Data into Relevant Data – Quickly and Accurately” Slides

• Given the Large and Diverse Collection of Unstructured Text Documents — How to Automate Characterization of Subject Areas to
Efficiently Navigate to Critical Information for Rapid Decision Making
• Digital Investigations—Discovery of Sensitive and Classified Information Use Cases

http://semanticommunity.info/Data_Science/Data_Science_for_Big_Data_Analytics
Updated: Wed, 23 Sep 2015 08:41:59 GMT
Powered by mindtouch™
• Appraisal of Electronic Records– Demand More Automated Less Human Intensive Reviews
• Faceting Information by: Topics, Mention Search, Format, Location, Time, Author
• Automated Keyword Search Extraction (Keyword Extraction for Reports and Articles (KERA)) and Discovery Techniques
• Tag Clouds as Lenses

10:35-10:55 Coffee and Networking Break

10:55-11:25 DR. ASHIT TALUKDER, Director, Information Access, National Institute of Standards and Technology (NIST)
I “NIST Perspectives and Initiatives” Slides

11:25-11:55 MR. MARK KRZYSKO, Deputy Director, Enterprise Information, Acquisition Resources and Analysis, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics (OUSD/AT&L)
I “DoD Decision Making Through Information and Data Management” Slides

• The Case for a Data-focused Approach to Decision-making
• The Defense Acquisition Data Solution
• Acquisition Data Uses from the Components to Congress
• The Future of Defense Acquisition Data

11:55-12:25 MS. LISA SHALER-CLARK, Deputy Director, Program Manager, Futures, Intelligence and Security Command (INSCOM), US Army
I “Big Data Analytics for Army Intel” Slides

• Leveraging Big Data across our Enterprise
• Partnering on Secure, Sharable Capabilities
• Opportunities and Emerging Lessons

12:25-1:45 Luncheon

1:45-2:15 MR. MARK GEMBICKI, Chairman and Chief Architect, Resilient Corporation
III Special Presentation: “The Role of Big Data in Improving Security and Resilience to Catastrophic Events” Slides

2:15-2:45 MR. KEVIN LAUSTEN, Senior Manager, Geospatial Data, and MR. DAVID GRASON, Manager, Analytics Technologies
III? “Geospatial Big Data – Deriving Insights from Global Datasets” Slides

2:45-3:15 COLONEL (ret) LEE WIGHT, Executive Director, Fusion Center, Washington Regional Threat Analysis Center (WRTAC); former Chief, Strategy Division, HQ US Air Force
I “Fusion Centers: Current Operations, Future Challenges, and Big Data” Slides
• National Fusion Center History
• WRTAC Mission/Organization/Key Processes
• WRTAC/NCR Current/Near Term Challenges
• WRTAC “Big Data” Integration
• National Fusion Center Strategy/Long Term Outlook and Implications

3:15-3:45 Refreshments and Networking Break

3:45-4:15 MR. NEIL ZIRING, Technical Director, Information Assurance, National Security Agency (NSA)
I “NSA Perspectives and Initiatives” Slides

4:15-4:45 MR. STEPHEN DENNIS, Director, Innovation, Science and Technology Directorate, Department of Homeland Security
Big Data Analytics and Homeland Security Slides

NOT PRESENTED?

DR. ADRIAN GARDNER, Chief Information Officer, Federal Emergency Management Agency (FEMA) [tentative]
I “FEMA Perspectives and Initiatives” Slides

MR. ADAM KARCHER, Unit Chief, Headquarters, Federal Bureau of Investigation
I “FBI Perspectives and Initiatives” Slides

MR. JAMES TRAINOR
Special Agent-in-Charge and Senior Executive, Headquarters, Federal Bureau of Investigation (FBI)
I “FBI Perspectives and Initiatives” Slides

Day Two: September 24, 2014

9:00-9:05 Administrative Announcements

9:05-9:35 LT. COLONEL (P) BOBBY SAXON, Division Chief and Program Director, Enterprise Management Decision Support, G-3/5/7 Program Office, US Army
II “Building on Big Data: Predictive Analytics” Slides

• The Big Data Foundation
• Access, Visualization, Analytics
• Predictive Analytics: “Seeing” the Future
• Business before Technology

II Special Double Presentation: (1) “Decision Point Analytics” (2) “The Role of the Interpreter” Slides

1. “Decision Point Analytics”
• Decision Point: What Data Was Used? What Algorithm? Recommended Course of Action? Was Recommendation Followed?
• Tracking the Results: Comparison vs. Decision-Point Expectations; Should the Algorithm be Modified? Was there Missing Data?
• Winning Over the Decision-Maker: Building Trust

2. “The Role of the Interpreter”

• Interpreters Understand the Business of the Organization and are Technically Competent
• Facilitate Building Better Government-Contractor Partnerships
• Critical Role after the Contractor “Hands Over the Keys”

10:25-10:45 Coffee and Networking

10:45-11:15 MR. DAVID BOYD, Chief Technology Officer, Engineering, L-3 Data Tactics
II “Big Data Standards Initiatives: Building a Framework for Data Interoperability” Slides

11:15-11:45 MR. JOSEPH KENNEY, National Security Leader, Federal Big Data Industry Team, IBM Software Group, for MR. STEVE STENNETT, Chief Technology Officer, Federal Information Management, IBM
III “Big Data Analytics and National Security” Slides

11:45-12:15 DR. DARRELL YOUNG, Principal Engineering Fellow, Raytheon
III “Big Data Analytics Needed to Measure Intelligence Potential” Slides

• Data Explosion is Accelerating with No End In Sight
• Intelligence Potential is the Likelihood of Finding Something Significant
• Intelligence Potential is Needed to Navigate through the Mountains of Data
• Intelligence Potential is Data and Task Dependent
• Defining and Computing Intelligence Potential is Needed to Support Variety of Other Analytics – Many Examples
• Intelligence Potential is Needed to Help Guide Analysts in Forming and Testing Hypotheses
• Computing the Intelligence Potential of Multiple Datasets and New Phenomenologies is Emerging Opportunity Well Suited for Cloud Implementation

12:15-1:30 Luncheon

1:30-2:00 MR. AARON BURCIAGA, Senior Scientist, Accenture LLP
III “Developing Use Cases” Slides

2:00-2:30 DR. GREG GARDNER, Chief Architect, Defense and Intelligence Solutions, NetApp
II “Optimizing Data Storage for Big Data in the Federal Government” Slides

• Data Storage Best Practices
• Data Storage Essentials for Big Data
• Big Data and Cloud Storage
• Addressing the Particular Data Storage Needs of the Federal Government

2:30-3:00 Refreshments and Networking

3:00-3:30 MR. ALAN BRIGGS, Senior Research Scientist, Elder Research
III “Advanced Data Science and the Intelligence Mission” Slides

• Big Data – a Paradigm Shift from Hypothesis-Driven to Data-Driven Discovery
• NSF Big Data Investment Areas and Programs: Foundational Research, Building New Multidisciplinary Research Communities, Education and Workforce, Cyberinfrastructure
• Recent Key NSF Funded Projects, Collaborations, and Workshops in Big Data
• Supporting the Big Data Innovation Ecosystem: Multistakeholder Partnerships to Move Data to Knowledge to Action

3:30-4:00 MR. GERARD CHRISTMAN, Senior Systems Engineer, Femme Comp, Inc.; contractor support to the Office of the Chief Information Officer, DoD/CIO
III "Civil Information Integration in Support of Complex Humanitarian Operations" Slides

• Leveraging the Geosocial Space to Fill Information Gaps
• Visualization Tools to Discover Patterns and to Predict Outcomes
• Geospatial Information to Aid in Mapping for Common Situational Awareness
• Leveraging the Unclassified Information Sharing Service to Provide the Means to Share and a Place to Collaborate
• Using Unstructured and Structured Data to Complete the Picture

Other

MR. WILLIAM KEELY
Deputy Chief Technology Officer, Defense Information Systems Agency (DISA) [tentative]
I “DISA Perspectives and Initiatives” Slides

MR. DUNCAN MCCARTHY
Technical Executive, Innovation, National Geospatial-Intelligence Agency (NGA) [invited]
I “NGA Perspectives and Initiatives” Slides

MR. KEN CAMPBELL
Vice President for National Security Solutions and
DR. COLLEEN McCUE
Senior Director, Social Science and Quantitative Methods, DigitalGlobe

III "Data Mining and Predictive Analytics for Law Enforcement and Intelligence” Slides

• Geospatial Big Data for Defense and National Security Missions
• Operational Security Analytics in Support of Anticipation and Influence
• Resource Optimization Through Area Reduction
Symposium on Predictive Analytics For Defense and Government

Source: Email, 9/17/2014

November 18-19, Washington, DC

Sheraton Pentagon City
900 S. Orme Street
Arlington, VA 22204

Day One: November 18

9:00-9:05 Administrative Announcements

9:05-9:35 “FCC Perspectives and Initiatives”
DR. DAVID BRAY, Chief Information Officer, Federal Communications Commission (FCC)

9:35-10:05 “DOC Perspectives and Initiatives”
MR. STEVE COOPER, Chief Information Officer, Department of Commerce (DOC)

10:05-10:35 “Army Cost Management and Financial Transparency”
DR. MORT ANVARI, SES, Director, Strategy and Programs, Office of the Assistant Secretary of the Army for Financial Management (ASA/FM&C)

10:35-10:55 Coffee and Networking

10:55-11:25 “CMS Perspectives and Initiatives”
MR. RAYMOND WEDGEWORTH, Director, Analytics and Control Group, Centers for Medicare and Medicaid Services, Department of Health and Human Services (CMS/HHS)

11:25-11:55 “Anticipatory Intelligence Research and Development at IARPA”
DR. DEWEY MURDICK, Program Manager, Intelligence Advanced Research Projects Agency (IARPA), Office of the Director of National Intelligence (DNI)

11:55-12:25 “NSF Perspectives and Initiatives”
DR. FEN ZHAO, Staff Associate, Office of the Assistant Director for Computer and Information Science and Engineering, National Science Foundation (NSF)

12:25-1:45 Lunch

1:45-2:15 “Intelligence, Surveillance, and Reconnaissance (ISR) and Predictive Analytics”
MR. MARTIN KRUGER, Program Manager, ISR Thrust Area, Office of Naval Research (ONR)
2:15-2:45 “Anticipatory Analysis and the Intelligence Mission”
MR. ROBERT ZITZ, Executive Vice President and Chief Systems Architect, Leidos; former Deputy Under Secretary for National Security and Preparedness, Department of Homeland Security

2:45-3:15 “Air Force Mobile Enabled Predictive Analytics”
DR. ROD FONTECILLA, Vice President, Federal Systems, Big Data Center of Excellence, Unisys

3:15-3:45 Refreshments and Networking

3:45-4:15 “National Security and Open Source Information – Legal and Policy Issues”
MR. DAVID VERHEY, Managing Partner, Argent PLLC; Former Senior Advisor and Counsel, Combating Terrorism Technical Support Office, Department of Defense

4:15-4:45 “Federal Big Data Initiative – Content, Network and Data Analytics for NITRD/NSF Data Science, Data Infrastructure and Data Publications”
DR. BRAND NIEMANN, Director and Senior Data Scientist, Semantic Community; former Senior Enterprise Architect, Office of the Chief Information Officer, Environmental Protection Agency (EPA)

4:45-5:15 “Beyond Crystal Balls and Tarot Cards – Big Data and Analytics”
MR. TIM PAYDOS, Director, Public Sector Big Data Industry Team, IBM

Day Two: November 19

9:00-9:05 Administrative Announcements

9:05-9:35 “Analytics and Information Sharing”
MS. PAMELA J. WISE-MARTINEZ, Senior Strategic Enterprise Architect, Office of the Program Manager, Information Sharing Environment, Office of the Director of National Intelligence (DNI)

9:35-10:05 “Analytics and Sustainment”
COLONEL JEFFREY POWELL, Commandant, Army Financial Management School

10:05-10:35 “GAO Perspectives and Initiatives”
MR. TIM PERSONS, Chief Scientist, Government Accountability Office (GAO)

10:35-10:55 Coffee and Networking

DR. MARK LIVINGSTON, Head, Visual Analytics Laboratory, (Stephen Russell, Jonathan Decker, Eric Leadbetter, Antonio Gilliam –coauthors), Naval Research Laboratory (NRL)

11:25-11:55 “Probabilistic Programming for Advanced Machine Learning and Predictive Analytics”
DR. SURESH JAGANNATHAN, Program Manager, Defense Advanced Research Projects Agency (DARPA)
11:55-12:25 “Analytics and Architecture for the Mission”
DR. NANCY GRADY, Technical Fellow, Data Science, SAIC; Chair, Definitions and Taxonomy Subgroup, Big Data Working Group, National Institute of Standards and Technology

12:25-1:30 Lunch

1:30-2:00 “Predicting the Next Cyber Attack”
MR. MARK JOHNSON, Director, Engineered Systems and Government Cloud, North America Public Sector Strategic Programs, Oracle

2:00-2:30 “Predictive Analytics and Law Enforcement”
MR. KEITH BRYARS, Client Executive, National Security and Law Enforcement, Harris Corporation; former Executive Special Agent, Federal Bureau of Investigation

2:30-3:00 “State-of-the-Art in Predictive Analytics for Emergency Mass Egress and Crowd Movement”
MR. DOUGLAS A. SAMUELSON, President and Chief Scientist, InfoLogix, Inc.

3:00-3:15 Refreshments and Networking

3:15-3:45 "Sociocultural Behavior Sensemaking – State-of-the-Art in Understanding the Operational Environment”
DR. DYLAN SCHMORROW (CAPT, USN-RET), Chief Scientist, SoarTech; former Deputy Director, Human Performance, Training and Biosystems, Office of the Secretary of Defense

3:45-4:15 “Accelerating the OODA-Loop with Analytics”
LT. COLONEL (USAF-RET) TAMMY SCHWARTZ, Owner, Llamrai Enterprises; former Chief Technology Officer, Air Force Enterprise Networking

---

**NSB Indicators Digest 2014**


This digest draws from the National Science Board's Science and Engineering Indicators 2014 and highlights key trends and data points from across the volume.

---

**National Science Board**

**Chairs**

**Dan E. Arvizu.** Chairman, Director and Chief Executive, National Renewable Energy Laboratory
Kelvin K. Droegemeier. Vice Chairman, Vice President for Research, Regents’ Professor of Meteorology and Weathernews Chair Emeritus, University of Oklahoma

Members

Deborah L. Ball. William H. Payne Collegiate Chair, Arthur F. Thurnau Professor, Dean of the School of Education, University of Michigan

Bonnie Bassler. Howard Hughes Medical Institute Investigator, Squibb Professor of Molecular Biology, Princeton University

Arthur Bienenstock. Professor Emeritus of Photon Science, Stanford University

Ray M. Bowen. President Emeritus, Texas A&M University, Visiting Distinguished Professor, Rice University

Vinton G. Cerf. Vice President and Chief Internet Evangelist, Google

France A. Córdova. President Emeritus, Purdue University

Ruth David. President and CEO, Analytic Services, Inc.

Inez Fung. Professor of Atmospheric Science, University of California, Berkeley

Esin Gulari. Dean of Engineering and Science, Clemson University

G. Peter Lepage. Professor of Physics, College of Arts and Sciences, Cornell University

Alan I. Leshner. CEO and Executive Publisher, Science, American Association for the Advancement of Science

W. Carl Lineberger. Fellow of JILA, E. U. Condon Distinguished Professor of Chemistry, University of Colorado

Stephen Mayo. Bren Professor of Biology and Chemistry, Chair, Division of Biology, California Institute of Technology

G. P. (Bud) Peterson. President, Georgia Institute of Technology

Douglas D. Randall. Professor Emeritus and Thomas Jefferson Fellow, Director Emeritus Interdisciplinary Plant Group, University of Missouri-Columbia

Geraldine Richmond. Richard M. and Patricia H. Noyes Professor of Chemistry, University of Oregon

Annella I. Sargent. Ira S. Bowen Professor of Astronomy, Vice President for Student Affairs, California Institute of Technology

Diane L. Souvaine. Vice Provost for Research, Professor of Computer Science, Tufts University

Arnold F. Stancell. Former Vice-President Mobil Oil, Professor Emeritus of Chemical Engineering and Turner Servant Leadership Chair Emeritus, School of Chemical and Biomolecular Engineering, Georgia Institute of Technology
Claude M. Steele. Dean, School of Education, Stanford University

Robert J. Zimmer. President, University of Chicago

Maria T. Zuber. E.A. Griswold Professor of Geophysics, Vice President for Research, Massachusetts Institute of Technology

Member Ex Officio

Cora B. Marrett. Acting Director, National Science Foundation

Michael L. Van Woert. Executive Officer, National Science Board, and Director, National Science Board Office National Science Board

Cover Page

February 2014 • NSB 14-02 National Science Board Science & Engineering Indicators DIGEST National Science Foundation 2014

Preface

The National Science Board (Board) is required under the National Science Foundation (NSF) Act, 42 U.S.C. § 1863 (j) (1) to prepare and transmit the biennial Science and Engineering Indicators (SEI) report to the President and to the Congress every even-numbered year. The report is prepared by the NSF National Center for Science and Engineering Statistics (NCSES) under the guidance of the Board. It is subject to extensive review by Board members, outside experts, interested federal agencies, and NCSES internal reviewers for accuracy, coverage, and balance.

Indicators are quantitative representations relevant to the scope, quality, and vitality of the science and engineering (S&E) enterprise. SEI is a factual and policy-neutral source of high-quality U.S. and international data; it neither offers policy options nor makes policy recommendations. The indicators included in the report contribute to the understanding of the current S&E environment.

This digest of key S&E indicators draws from the Board’s Science and Engineering Indicators 2014, the 21st volume of this biennial series. The digest serves to draw attention to important trends and data points from across SEI 2014 and to introduce readers to the data resources available in the report. Readers are invited to explore each of the key indicators presented here in more detail in the full report. To that end, each indicator presented in this digest refers to the SEI 2014 chapter or chapters from which it was drawn. The complete SEI 2014 report and related resources are available on the Web at http://www.nsf.gov/statistics/indicators/.

Readers may also be interested in the online resources associated with SEI 2014. A list and description of these products appears at the end of this digest. The Board hopes that readers will take advantage of these rich sources of information.
Table of Contents

2 Introduction

4 Global R&D: Measuring Commitment to Innovation
   • How much?
   • Growth
   • Where?
   • Intensity

6 U.S. R&D: Funding and Performance
   • Trends
   • Basic research
   • Development and applied research
   • Academic R&D

8 U.S. R&D: Federal Portfolio
   • Trends
   • Type of work
   • Focus
   • S&E fields
   • Performers

10 U.S. and Global STEM Education
   • K–12 mathematics and science
   • International baccalaureates
   • U.S. baccalaureates
   • International doctorates

12 U.S. S&E Workforce: Trends and Composition
   • Workforce growth
   • Women and underrepresented minorities
   • Unemployment
   • Immigrants

14 Research Outputs: Publications and Patents
   • Publications
   • Patents
   • Engineering articles
   • Science-patent linkage
Introduction

The United States holds a preeminent position in science and engineering (S&E) in the world, derived in large part from its long history of public and private investment in S&E research and development and education. Investment in R&D, science, technology, and education correlate strongly with economic growth and with the development of a safe, healthy, and well-educated society.

Many other nations, recognizing the economic and social benefits of such investment, have increased their R&D and education spending. These trends are by now well-established and will challenge the world leadership role of the United States.

Major S&E Indicators

The National Science Board has selected 41 S&E indicators for inclusion in this digest. These indicators have been grouped into seven themes. Although each stands alone, collectively these seven themes are a snapshot of U.S. R&D capacity and outputs in the context of global trends affecting them. Exploration of areas that indicate capacity for innovation is a thread common to many of the themes presented here. As economies worldwide grow increasingly knowledge-intensive and interdependent, capacity for innovation becomes ever more critical.

Three themes provide a worldwide view, picturing R&D spending, research outputs, and STEM education. Four others share a domestic focus, providing information on U.S. R&D funding and performance, federal R&D support, the U.S. S&E workforce, and public research universities. Indicators may vary in successive volumes of the Science and Engineering Indicators series as different S&E policy issues emerge.

What These Indicators Tell the Nation

By selecting a set of indicators, the Board seeks to contribute to the assessment of the state of U.S. science and engineering and to highlight issues of current opportunity or concern. These measures address an emerging set of trends of particular interest to planners and policymakers at all levels whose decisions affect our national S&E enterprise.

GLOBAL R&D: MEASURING COMMITMENT TO INNOVATION
WHY IS THIS IMPORTANT?

Innovation in the form of new goods, services, or processes builds new knowledge and technology, contributes to national competitiveness, improves living standards, and furthers social welfare. Research and development is a major driver of innovation. R&D expenditures indicate the priority given to advancing science and technology relative to other national goals.

A. HOW MUCH?

R&D expenditures worldwide are estimated to have exceeded $1.4 trillion in 2011, up from $753 billion a decade earlier.

Figure A Estimated R&D expenditures worldwide: 1996–2011

SEI 2014: Global Pattern of R&D Expenditures, Chapter 4.

B. WHERE?

Global R&D expenditures are highly concentrated in three regions: Asia, North America, and Europe.
NOTE: Asia/Pacific includes China, Taiwan, Japan, South Korea, Singapore, Malaysia, Thailand, Indonesia, Philippines, India, Pakistan, and Sri Lanka.


The seven countries with the largest R&D expenditures together accounted for nearly three-fourths of total global R&D in 2011. The United States remains the largest R&D performer and accounted for almost one-third of worldwide R&D total in 2011. China is now the second largest R&D performing nation, accounting for about 15% of the worldwide total.

SEI 2014: *Global Pattern of R&D Expenditures*, Chapter 4
C. GROWTH

Asian countries have led the growth in worldwide R&D expenditures over the past decade, with China accounting for about one-fourth of the total global growth.

**Figure C1 Contributions of selected countries/regions/economies to growth of worldwide R&D expenditures: 2001–11**

![Contributions of selected countries/regions/economies to growth of worldwide R&D expenditures: 2001–11](image)

**NOTE:** Other Asia includes India, Indonesia, Malaysia, Pakistan, Philippines, Singapore, Sri Lanka, Taiwan, and Thailand.

**SEI 2014: Global Pattern of R&D Expenditures, Chapter 4.**

Asian countries have led the pace of R&D expansion as well. Between 2001 and 2011, China’s R&D expanded the most rapidly followed by South Korea. By comparison, the pace of growth has been much slower in the United States and the European Union. Rapid R&D growth in Asia overall reflects private spending by domestic and foreign firms as well as increased public R&D spending.
D. INTENSITY

R&D intensity is the proportion of a country’s economic activity (gross domestic product) devoted to R&D investment. Historically, Japan has had higher R&D intensity than most other large economies, but South Korea recently surpassed it.

China’s R&D intensity has increased sharply since the mid-1990s, as growth in R&D outpaced a rapid expansion in GDP. China’s R&D intensity is approaching that of the European Union but remains well below South Korea, Japan, and the United States.

Figure D R&D expenditures as a share of economic output for selected countries/economies: 1996–2011

SEI 2014: Comparison of Country R&D Intensities, Chapter 4.
U.S. R&D: FUNDING AND PERFORMANCE

WHY IS THIS IMPORTANT?

Businesses, government, academia, and nonprofit organizations all fund and perform R&D. The outcomes and benefits depend not only on the total funds devoted to R&D but also on the types of R&D these funds support—basic research, applied research, and development.

A. TRENDS

U.S. R&D support exceeded $424 billion in 2011. R&D support rose in 2011, following a decline in 2010 in the aftermath of the economic downturn. Overall, U.S. R&D investment grew by half in the last decade (21% after adjusting for inflation). Industry, long the nation’s largest supporter of R&D, increased its 2011 funding, offsetting a decline in federal R&D funding.

Figure A U.S. R&D expenditures, by source of funds: 1990–2011

SEI 2014: Sources of R&D Funding, Chapter 4.

B. DEVELOPMENT AND APPLIED RESEARCH

Resources for development and applied research—work that aims at practical application, new products, or novel processes—rose in 2011, propelled by a nearly $21 billion increase in industry funding. Industry provides the largest share of funding for development and applied research.
Industry also performs the largest share by far of the nation’s development and applied research. The federal government and academic and other nonprofit institutions perform the remainder.

NOTE: The data for federal government are intramural only.

C. BASIC RESEARCH

Basic research is directed primarily toward increasing knowledge or understanding and has long relied on federal government support. Federal funding for basic research, however, has mostly dropped since 2004. Although funding
from academic and other nonprofit institutions combined has exceeded that from industry throughout the past decade, industry funding for basic research has risen overall since 2006. Universities and colleges are the prime performers of the nation’s basic research, a role they uniquely combine with the training of new researchers. Industry’s share of basic research performance has recently risen after years of decline; the federal government’s share has gradually diminished.

Figure C1 Funding sources for U.S. basic research: 1990–2011

![Funding sources for U.S. basic research: 1990–2011](image)

**SEI 2014: Sources of R&D Funding** and **R&D, by Character of Work**, Chapter 4.

Figure C2 Performers of U.S. basic research: 1990–2011

![Performers of U.S. basic research: 1990–2011](image)

**NOTE:** The data for federal government are intramural only.

D. ACADEMIC R&D

Most of the R&D performed by the academic sector is basic research. Funding sources for academic R&D have been relatively stable for nearly two decades, with the federal government continuing to fund about 60% of academic R&D, and institutions’ own funds constituting the next largest share.

Figure D Funding sources for U.S. academic R&D: 1990–2012

![Funding sources for U.S. academic R&D: 1990–2012](image)

**SEI 2014: Expenditures and Funding for Academic R&D, Chapter 5.**

U.S. R&D: FEDERAL PORTFOLIO

**WHY IS THIS IMPORTANT?**

The distribution of R&D funds by the U.S. federal government provides insight into the nation’s broad mission priorities for public expenditures.

**A. TRENDS**

Federal investment in R&D declined in 2010 and 2011, reaching $126 billion in 2011. However, federal R&D support increased consistently during the decade of the 2000s, rising by more than two-thirds overall between 2001 and 2011 (34% after adjusting for inflation).
Figure A Federal spending on R&D: 1990–2011

B. FOCUS

The majority of the federal R&D budget is devoted to defense. About half of the federal nondefense R&D budget is devoted to health and is funded primarily through the National Institutes of Health. In the past 20 years, as a proportion of all nondefense R&D, health and general science accounts have risen the most. The 2009 boost in nondefense, health, and general science accounts was driven by federal stimulus funds.

Figure B Federal R&D budget, by national objectives: FYs 1990–2011
C. PERFORMERS

Different institutions bring different perspectives and approaches to R&D. Academic and other nonprofit institutions, which tend to concentrate on basic research, have generally received steadily increasing federal support. Industry, heavily focused on development and applied research, has seen a doubling of federal funding (not adjusting for inflation) since 2001 after a decade of no growth.

Figure C Federal spending on R&D, by performer: 1990–2011

NOTE: The data for federal government are intramural only.

SEI 2014: R&D Funding by the Federal Government, Chapter 4.

D. TYPE OF WORK

For nearly 20 years, funding of basic and applied research has accounted for at least half of federal funding of R&D. During the 2000s, federal funding of development activities grew more than federal funding of basic and applied research, driven primarily by federal stimulus funds and reversing the trend seen in the previous decade.
Figure D Federal R&D funds, by type of work: 1990–2011

SEI 2014: R&D, by Character of Work, Chapter 4.

E. S&E FIELDS

The life sciences have accounted for about half of the federal research portfolio (basic and applied research) since 2001. Over the past decade, federal research funding has varied little in most S&E fields. The environmental sciences saw the largest decline in federal research dollars. Starting from a low base and propelled by psychology, the combined category of social sciences and psychology saw the largest increase.

Figure E1 Federal basic and applied research funds, by S&E field: 1990–2011

SEI 2014: Federal Spending on Research, by Field, Chapter 4.
U.S. AND GLOBAL STEM EDUCATION

WHY IS THIS IMPORTANT?

Education at all levels in science, technology, engineering, and mathematics—STEM—develops, preserves, and disseminates knowledge and skills that convey personal, economic, and social benefits. Higher education provides the advanced work skills needed in an increasingly knowledge-intensive, globally-integrated, and innovation-based economic landscape.

A. K-12 MATHEMATICS AND SCIENCE

Over the past two decades, U.S. students’ mathematics scores on national assessments have modestly improved.

Figure E2 Inflation-adjusted change in federal basic and applied research funds, by S&E field: 2001–11
Figure A1 Average NAEP mathematics scores of U.S. students in grades 4 and 8: 1990–2011

NAEP = National Assessment of Educational Progress.

SEI 2014: Mathematics and Science Performance in Grades 4 and 8, Chapter 1.

U.S. 15-year olds tend to score slightly lower than the international average in mathematics and have scores about equal to the international average in science.

Figure A2 Average mathematics and science PISA test scores of U.S. and OECD 15-year olds: 2003–09

PISA = Program for International Student Assessment; OECD = Organisation for Economic Co-operation and Development.

B. U.S. BACCALAUREATES

U.S. output of total bachelor’s degrees has increased by more than half over two decades, reflecting a rising trend in college attendance. Natural sciences and engineering degrees have consistently constituted about one-sixth of the total.

The increases in certain natural sciences degrees—physical sciences and mathematics—and in engineering degrees generally reflect the size of the college-age cohort rather than a rise in the number of students who major in those fields. In contrast, the number of biological and agricultural sciences degrees rose during most of the last two decades. Computer sciences degrees rose through the dot.com bubble and then declined and leveled off during the second half of the 2000s.

Figure B U.S. bachelor’s degrees in selected S&E fields per 1,000 20–24-year olds: 1991–2011

C. INTERNATIONAL BACCALAUREATES

The number of degrees in natural sciences and engineering fields has risen much faster in China than in the United States.

Natural sciences and engineering fields account for a much larger proportion of all bachelor’s degrees in China than in the United States. In 2010, these fields accounted for 44% of all bachelor’s degrees in China, compared with 16% of all bachelor’s degrees in the United States.
Figure C1 Bachelor’s degrees in natural sciences, by selected country/economy: 2000–10

![Bachelor's degrees in natural sciences graph](image1)


Figure C2 Bachelor’s degrees in engineering, by selected country/economy: 2000–10

![Bachelor's degrees in engineering graph](image2)


D. INTERNATIONAL DOCTORATES

Advanced training towards the doctorate has expanded in recent years. The numbers of doctoral degrees in natural sciences and engineering have risen dramatically in China, whereas the numbers awarded in the United States, South Korea, and many European countries have risen more modestly.

In 2010, China graduated the largest number of doctorate recipients in natural sciences and engineering of any individual country, followed by the United States. In the United States, however, only 57% of these doctorates were earned by citizens or permanent residents, and temporary visa holders earned the remainder.
U.S. S&E WORKFORCE: TRENDS AND COMPOSITION

WHY IS THIS IMPORTANT?

Workers with S&E expertise are an integral part of a nation’s innovative capacity. Their high skill level and inventiveness provide them with the ability to not only advance basic scientific knowledge, but also to transform that knowledge into useful products and services.

A. WORKFORCE GROWTH

The U.S. S&E workforce—made up of chemists, mathematicians, economists, engineers, and other such workers—has grown faster over time than the workforce overall. Over the past 50 years it has grown fivefold and now represents more than 4% of all U.S. jobs.

During and after the 2007–09 recession, employment in S&E or S&E related jobs was generally more resilient than was overall employment.
B. UNEMPLOYMENT

For decades, workers in S&E occupations have almost always had lower unemployment than workers in other types of jobs. The unemployment rate for college-graduate workers in S&E occupations is generally lower than it is for college-graduate workers in non-S&E occupations, and it is far lower than the overall unemployment rate. However, S&E workers are not immune to overall business cycles, as the spikes in S&E unemployment in the 2001 and the 2007–09 recessions illustrate.

SEI 2014: Growth of the S&E Workforce, Chapter 3.
C. WOMEN AND UNDERREPRESENTED MINORITIES

Despite accounting for nearly half of the college-educated workforce, women in 2010 accounted for less than one-third of S&E employment. Although the number of women in S&E jobs has risen significantly in the past two decades, the disparity has narrowed only modestly.

Figure C1 Men and women in S&E occupations: 1993, 2003, and 2010

SEI 2014: Women and Minorities in the S&EE Workforce, Chapter 3.

Similarly, although underrepresented minorities—blacks, Hispanics, and American Indians or Alaska Natives—have made substantial strides, their representation in S&E jobs remains below their proportion in the population.

Figure C2 Underrepresented minorities in S&E occupations: 1993, 2003, and 2010

For both women and underrepresented minorities, growth in participation slowed during the 2000s. Women’s presence varies widely across S&E occupations. For underrepresented minorities, variation among occupations, although present, is much less pronounced.

**Figure C3 Women and underrepresented minorities in S&E occupations: 2010**

D. IMMIGRANTS

Foreign-born scientists and engineers, whether educated in the United States or abroad, are a critical part of the U.S. S&E workforce. Among individuals with their highest degree in an S&E field, 33% of master’s degree holders and 42% of doctorate holders are foreign born. Reliance on foreign-born scientists and engineers is greatest on those with engineering and mathematics and computer sciences degrees. More than half of the doctorate holders in these fields are foreign born.
RESEARCH OUTPUTS: PUBLICATIONS AND PATENTS

WHY IS THIS IMPORTANT?

Research produces new knowledge, products, or processes. Research publications reflect contributions to knowledge, patents indicate useful inventions, and citations on patents to the scientific and technical literature indicate the linkage between research and practical application.

A. PUBLICATIONS

The United States publishes more S&E articles than any other country. The combined output of the European Union, however, is larger than that of the United States.

Asia's S&E research article output is approaching parity with the United States and the European Union. Between 1997 and 2011, Asia's output more than doubled, led primarily by China. In 2011, China produced 11% of the world's S&E articles, more than any country except the United States.
Figure A S&E articles, by selected country/region/economy: 1997–2011

NOTE: Asia total includes China, India, and Japan.

SEI 2014: S&E Article Output, Chapter 5.

B. ENGINEERING ARTICLES

Engineering is vital to knowledge-intensive and technologically advanced economies, and many Asian economies are building their engineering capabilities.

In 2011, China published 17% of the world’s engineering articles, equaling the U.S. share. U.S. publications, however, continue to receive more citations. Asia as a whole published more than twice as many engineering articles as the United States and 50% more than the European Union in 2011. The output of engineering articles is rising in the European Union and, more gradually, in the United States.

Figure B Engineering articles, by selected country/region/economy: 1997–2011
NOTE: Asia total includes China, India, and Japan.

SEI 2014: S&E Article Output, Chapter 5.

C. PATENTS

Patents protect the property rights of inventors. As knowledge-intensive economic activities expand worldwide, patent awards are rising. Inventors from around the globe seek patent protection in the United States because of its large and open market. U.S. inventors now account for just under half of all U.S. patents. Inventors in the European Union and Japan receive most of the U.S. patents awarded to non-U.S. inventors, with growing numbers of U.S. patents awarded to inventors in Taiwan and South Korea. Although growing, U.S. patents awarded to inventors in China and India remain modest.

Figure C1 U.S. patents granted to non-U.S. inventors, by country/region/economy: 1995–2012

Information and communications technologies (ICT) and health-related technologies account for nearly half of all U.S. patents granted. During the last decade, the number of ICT patents climbed much more steeply than the number of health-related patents.
D. SCIENCE-PATENT LINKAGE

Patents cite the prior scientific and technological knowledge on which they are built. A growing proportion of U.S. patents, whether awarded to U.S. or foreign inventors, cites research in published scientific articles as an influence.

SEI 2014: Citation of S&E Articles by USPTO Patents, Chapter 5.

PUBLIC RESEARCH UNIVERSITIES
WHY IS THIS IMPORTANT?

Public and private research universities contribute to innovation and economic competitiveness. Public research universities play a special role in supplying S&T expertise to state and local governments and businesses and providing a broad cross section of students with access to high-quality STEM education in a research-oriented environment. Declining funding, increasing student enrollments, and rising costs pose formidable challenges to their mission.

A. ENROLLMENT AND DEGREES

Research universities are the leading producers of S&E degrees in the United States. Public research universities grant the majority of the S&E degrees awarded by these universities.

Figure A1 S&E degrees produced by research and other universities, by level of degree: 2011

NOTE: Other universities include master’s- and bachelor’s-awarding universities and colleges, special focus institutions, associates colleges, and tribal colleges.


Public research universities also account for the majority of all research-university enrollment.
Figure A2 Enrollment in research universities: 1990–2011

![Enrollment in research universities: 1990–2011](image)


**B. R&D**

Along with their private counterparts, public research universities play an essential role in the production of academic research. Public universities account for the majority of all research-university R&D expenditures.

Figure B S&E R&D expenditures of research universities: 1990–2012

![S&E R&D expenditures of research universities: 1990–2012](image)


**C. AFFORDABILITY**

Public institutions have traditionally offered less-affluent students an avenue to a high quality, affordable education. In the last two decades, however, the cost of attending public research universities has risen steeply, coinciding with a decline in state and local appropriations, a significant source of institution revenue. Adjusted for inflation, tuition and fees
per full-time student rose sharply between 1987 and 2010—143%—in the most research-intensive public universities; state and local appropriations per full-time student fell 28% in that period.

**Figure C1 Tuition and state and local appropriations in U.S. public research universities: 1987–2010**

![Tuition and state and local appropriations in U.S. public research universities: 1987–2010](image)

NOTES: Data are per full-time equivalent student and for the most research-intensive universities. Net tuition data reflect tuition after subtracting institutional grant aid.


The sharp rise in tuition coincided with nearly stagnant household incomes. Between 1987 and 2010, median household income in the United States grew only 3% after adjusting for inflation.

Despite their steep increase in tuition and fees, public research universities remain less expensive than private research universities.
D. INVESTMENT IN HIGHER EDUCATION

The sum of U.S. economic resources devoted to higher education remains competitive in the global context. Along with Canada and South Korea, the United States devotes a larger proportion of its gross domestic product to higher education than do other developed countries.

NOTES: Data are per full-time equivalent student and for the most research-intensive universities. Net tuition data reflect tuition after subtracting institutional grant aid.


GDP = gross domestic product; OECD = Organisation for Economic Cooperation and Development.

NOTES: Data for Canada are for 2008. Data for Switzerland include public expenditures only.
Glossary and Key to Acronyms

Applied research
Systematic study to gain knowledge or understanding to meet a specific, recognized need.

Basic research
Systematic study to gain more comprehensive knowledge or understanding of the subject under study without specific applications in mind.

Development
Systematic use of the knowledge or understanding gained from research directed toward the production of useful materials, devices, systems, or methods, including the design and development of prototypes and processes.

GDP
Gross domestic product. The market value of all final goods and services produced within a country within a given period of time.

European Union (EU)
As of June 2013, the EU comprised 27 member nations: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom. Croatia joined the EU in July 2013. Unless otherwise noted, Organisation for Economic Cooperation and Development data on the EU include all 28 members; data on the EU from other sources are limited to the 27 nations that were members as of June 2013.

NCSES
National Center for Science and Engineering Statistics, National Science Foundation. NSB. National Science Board.

NSF
National Science Foundation. Natural sciences. Agricultural, biological, computer, and physical sciences (including earth, atmospheric, and ocean sciences), and mathematics.

R&D
Research and development. R&D intensity. R&D as a proportion of gross domestic product.
Research university

The Carnegie Classification of Institutions of Higher Education considers *doctorate-granting universities* that award at least 20 doctoral degrees per year to be research universities. The 2010 Carnegie Classification includes three subgroups of research universities based on the level of research activity: very high research activity (108 institutions), high research activity (99 institutions), and doctoral/ research universities (90 institutions).

**S&E**

Science and engineering.

**S&E occupations**

Biological, agricultural, and environmental life scientists; computer and mathematical scientists; physical scientists; social scientists; and engineers. S&E managers and technicians, and health-related occupations are categorized as S&E-related and are not included in S&E.

**S&T**

Science and technology.

**SEI**

Science and Engineering Indicators.

Explore Further

To read more about the themes presented in this digest, please see the Overview chapter as well as the more detailed analysis and fuller discussion of the related topics presented in *SEI 2014*. Each theme is matched with its source *SEI 2014* chapter or chapters in the list below. *SEI 2014* also provides a wealth of detailed information on public attitudes and understanding of science and engineering (Chapter 7) and state-level comparisons of selected science and engineering indicators (Chapter 8).

**Global R&D: Measuring Commitment to Innovation**

- Chapter 4. Research and Development: National Trends and International Comparisons

**U.S. R&D: Funding and Performance**

- Chapter 4. Research and Development: National Trends and International Comparisons
- Chapter 5. Academic Research and Development

**U.S. R&D: Federal Portfolio**

- Chapter 4. Research and Development: National Trends and International Comparisons

**Global STEM Education**

- Chapter 1. Elementary and Secondary Mathematics and Science Education
SEI 2014 Online Resources


**Companion piece.** The Board’s companion pieces are “companion” policy statements to *SEI 2014*. The Board focuses on trends that it believes raise important policy concerns and should be brought to the attention of the President, Congress, and the public.

**State data tool.** The state data tool allows interactive exploration of 59 indicators of state trends in science and technology education, workforce, finance, and R&D. Users have the ability to choose and explore a single indicator in depth, compare multiple indicators for pre-selected groups, customize their own graphics, or download data tables.

**Presentation graphics.** Presentation graphics, in PowerPoint slide and image (JPEG) formats and accompanied by their supporting data (Excel), are based on figures in the Overview chapter of *SEI 2014*.

**Source data.** Data supporting each figure, table, and appendix table in *SEI 2014* are available for download in Excel format. Links are provided on the *SEI 2014* main page to the lists of figures, tables, and appendix tables, each organized by chapter.

Acknowledgments

This digest was developed with guidance from the National Science Board by Beethika Khan, National Science Foundation, National Center for Science and Engineering Statistics (NCSES), under the direction of Robert Bell and assisted by the Center’s analytic staff. The volume was edited by Cheryl Roesel, NCSES. Eileen Kessler and staff at OmniStudio, Inc., designed the layout. Development of the Web version was guided by Cheryl Roesel and produced by Robin Pentola, with technical assistance from staff of Penobscot Bay Media, LLC.
The flower-like structure on the cover of *Science and Engineering Indicators 2014* is a graph that illustrates potential energy surfaces in a molecule called sym-triazine. The theoretical approach behind the graph is part of a larger effort that helped explain how sym-triazine can simultaneously break into three parts. Most molecules break apart one step at a time, so the phenomenon is rare. Researchers at the University of Southern California used computational chemistry tools to produce the graph, explaining the experimental results obtained by collaborators at the University of California, San Diego. The researchers reported their findings in the August 8, 2008, issue of the journal Science. This work was supported by the National Science Foundation under the auspices of the iOpenShell (Center for Computational Studies of Electronic Structure and Spectroscopy of Open-Shell and Electronically Excited Species). (Credit: Vadim Mozhayskiy and Anna I. Krylov, Department of Chemistry, University of Southern California.)

Scan this code for direct access to online content. (Requires mobile device with camera and QR/ barcode scanner app.)