Big Data Symposia

Story: Federating Big Data for Big Innovation

Government BIG DATA Symposium

Day 1: Tuesday, November 19, 2013
Day 2: Wednesday, November 20, 2013

Brochure

I. The Latest Federal Government Strategies, Plans, Needs and Initiatives
   - DOC Perspectives and Initiatives
   - Department of Homeland Security – Big Data Analytics Needs Big Data Governance
   - HHS Perspectives and Initiatives
   - GSA Perspectives and Initiatives
   - Big Data – A Threat Perspective
   - DOE Perspectives and Initiatives
   - The NIST Big Data Public Working Group
   - The Big Data Senior Steering Group

II. Big Data Analytics and Applications for the Defense and Intelligence Missions – Opportunities, Challenges, Needs and Initiatives
   - Transforming DoD Decision Making Through a Focus on Information and Data Management
   - Big Data R&D at DARPA
   - Big Data and the Defense ISR Mission
   - Weapons Data Optimization
   - Civil Information Integration in Support of the US Government
   - The Challenges of Big Data for Intelligence Missions
   - Predictive Analytics and Intelligence Operations

III. Government Large-Scale Analytic Programs and Initiatives – Status and Forecast
   - Exploration of Structured and Unstructured Clinical Data in the Veterans Administration
   - Federating Big Data for Big Innovation
   - The Health Data Initiative
   - Smart Grid Data Initiatives – Green Button

IV. Selecting and Developing Missions and Mission Apps
   - Putting Data Analytics to Work for Government
Applying Big Data Analytics to Government Missions
Big Data for the Mission Side of the House
Unlocking Value from Your Big Data

V. The Latest Tools, Techniques and Emerging Lessons Learned
Operational Challenges and Considerations in Large-Scale Data Analytics
Developing/Deploying the Big Data Analytics Architecture
Big Data Security – Really? Think Again!
Big Data, Fast Data and Cyber Security
Data Tactics: A Blended Approach to Big Data Analytics

Spotfire Dashboard
Research Notes

W3C eGov eParticipation and Open Data
New Data.gov Catalog
New Data.gov Testing

Introduction
Preparation
Question 1
Question 2
Question 3
Question 4
Question 5
Question 6
Question 7
Question 8
Question 9
Question 10
Question 11
Task 1
Task 2
Task 3
Task 4
Task 5
Task 6
Task 7
Task 8
Task 9
Task 10
Task 11
Access to testing for the new Data.gov site

EU Data Portal Catalogue

Story: Mining the Big Data Symposium for Big Data Sets and Ideas

Big Data Symposium: Analytics and Applications for Defense, Intelligence and Homeland Security

Day 1: Tuesday, September 24, 2013

Day 2: Wednesday, September 25, 2013

Research Notes

Story: Big Data Conferences

Slides

Slide 1 Big Data Conference: Analytics and Applications for Federal Big Data

Slide 2 Preface

Slide 3 Strata 2013 Conference: Making Data Work

Slide 4 Big Data Symposia

Slide 5 The Connection

Slide 6 Strata 2013 Conference: Broad Data

Slide 7 Big Data is Going Broad According to Government Internet Guru Jim Hendler

Slide 8 World Wide Web Expert Jim Hendler Receives Inaugural Strata "Big Data" Award

Slide 9 IOGDS: International Open Government Dataset Search

Slide 10 Critique

Slide 11 A Solution: Data Science

Slide 12 A Solution: Spotfire

Slide 13 Gartner Magic Quadrant: Business Intelligence and Analytics Platforms
Slide 14 My Process
Slide 15 Results
Slide 16 Big Data Symposia: Knowledge Base in MindTouch
Slide 17 IOGDS: Excel Spreadsheet
Slide 18 DataCatalogs.org: Excel Spreadsheet
Slide 19 DataCatalogs.org: Spotfire
Slide 20 IOGDS Countries and Catalogs: Spotfire
Slide 21 IOGDS France: Spotfire
Slide 22 US Data.gov Catalog: Spotfire
Slide 23 U.S. Census Bureau/Small Area Health Insurance (SAHIE) Program: Spotfire
Slide 24 Conclusions and Recommendations

Story: Matrix of First Symposium Presentations and Pilots
Pilot: Before Second Symposium
Pilot: Based on Second Symposium Presentations
Spotfire Dashboard
Research Notes
DataCatalogs.org
Data Hub
Open Government Data Catalogues
Data Catalog Vocabulary (DCAT)
Government Linked Data (GLD) Working Group

Emails
Peter Krantz
Christos Koumenides
Charles Ruelle
Bastiaan Debileck
John Erickson
Mohamed
Pierre Andrews
Vassilios Peristeras
Asuncion Gomez Perez
Carlos Iglesias
Carlos Iglesias 1
Carlos Iglesias 2
Martin Alvarez-Espinart
David Milton
David Milton 1
Ed Summers
ACT-IAC Big Data Symposium

Questions

Objectives

Symposium Overview
- Making the Business Case for Big Data Analytics / Skill Sets / Training
- Managing and Securing Large and Complex Data
- Developing a Data Analytics Capability (Tools, Pilot Projects, Resources)

Agenda
- Big Data Committee Leadership
- For More Information

Strata Conference 2013: Making Data Work

Sessions of Interest
- Data Visualization Design Using Shneiderman’s Mantra: Overview First, Zoom and Filter, Then Details-on-Demand
- Broad Data: What Happens When the Web of Data Becomes Real?
- Crowdfunded Open Doctor Data
- Maps Not Lists: Network Graphs for Data Exploration
- A Model Strategy for Data Journalism in a Country Without Open Data
- Data Science vs. Analytics -- Approaches to Problem Solving
- The BigData Top100 List
- The Web As The Greatest Dataset Of All Time
- Great Debate: Design Matters More Than Math
- Big Data vs The Beltway: The Regulatory Risks to Data-Driven Businesses
- The Victory Lab

Terracotta

Tableau

Gartner Says Big Data Makes Organizations Smarter, But Open Data Makes Them Richer

Magic Quadrant for Business Intelligence and Analytics Platforms

VIEW SUMMARY

Market Definition/Description
Integration
Information Delivery
Analysis

Magic Quadrant
Quadrant Descriptions
   Leaders
   Challengers
   Visionaries
   Niche Players

Context
Market Overview
Tibco Spotfire
   Strengths
   Cautions

Big Data Analytics and Applications for Defense, Intelligence and Homeland Security Symposium

Day 1: Wednesday, April 24, 2013
Day 2: Thursday, April 25, 2013
Brochure

I. Government Needs, Initiatives, Opportunities and Challenges
   Emerging Technical Challenges and Capabilities
   NGA Big Data R&D — Perspectives & Initiatives
   Big Data and the Intelligence Mission
   INSCOM Perspectives and Initiatives
   Air Force Perspectives and Initiatives
   Big Data and Army Intelligence
   Joint Staff Perspectives and Initiatives
   SPAWAR Perspectives and Initiatives

II. Emerging Applications for Defense and Intelligence
   Data Analytics and DCGS
   Large Scale Analytics and Network Management
   ISR in a Tactical Environment
   The DARPA UPSIDE Program
   Big Data Analytics and National Security
   GeoINT and Large Scale Analytics
   Open Source, GOTS, and GOSS - Empowering Big Data and Enabling the Intelligence Mission

III. The Latest Tools, Techniques and Technologies – Data Collection/Discovery, Deep/Predictive Analytics, Cloud, Scalability, Security, etc.
   ISR as a Candidate Solution Framework for ISR Big Data Volume, Velocity and Variety
Big Data – Not Just About Processing Data with Analytics; It’s Also About How to Collect, Manage, and Discover Large Volumes of Real Time Data Across Globally Distributed Enterprises

Challenges in Big Data Analytics – Applications and Capabilities
Large-Scale Analytics
Big Data Trends – Challenges, Solutions, Use Cases
Social Data - What is It and Why Should You Care?
Social Media Analytics
Human-Driven Data Analytics

Government Big Data Symposium
Day 1: Tuesday, March 5, 2013
Day 2: Wednesday, March 6, 2013
Brochure

I. The Latest Federal Government Strategies, Plans, Needs and Initiatives
   NITRD Perspectives and Initiatives
   NSF Perspectives and Initiatives Related to Big Data
   NASA Perspectives and Initiatives
   FCC Perspectives and Initiatives and the Role of Data Officers
   DHS Perspectives and Initiatives
   Big Data for Defense
   FEMA Perspectives and Initiatives
   CMS Perspectives on Data and Analytics to Drive Health System Transformation
   IRS Perspectives and Initiatives
   OMB Perspectives and Initiatives

II. Technical Challenges and Mission Strategies
   Big Data Challenges, Opportunities: Role of Measurement, Standards and Interoperability
   Open Gov 2.0 and Safety.Data.Gov
   Smart Disclosure and Data Analytics
   Technical Challenges for Defense
   DOI’s eERDMS Program – Update and Forecast
   The Education Data Initiative
   Operational Challenges and Considerations in Large-Scale Data Analytics

III. Advanced Tools and Techniques
   Large-Scale Text Analytics and Mining
   Dealing with Structured and Unstructured Data
   The Art of Predicting with Big Data
   The Value and Challenges of Large Scale Entity Analysis for National Security

IV. Implementation Strategies and Lessons Learned
   Big Data Implementation Strategies
I got the idea for this story from three recent activities of Dr. Jeanne Holm, Data.gov Evangelist:

- W3C eGov eParticipation and Open Data (Online Meeting)
- New Data.gov Catalog (Email Request)
- Federating Big Data for Big Innovation (Conference Presentation)

This story is both an Interface to this big page about big data conferences and to several big data applications.

I attended 4 big data conferences in a series in 2013:

- Government Big Data Symposium: March 5-6, 2013, Rosslyn, Virginia (Agenda and Brochure)
- Big Data Analytics and Applications for Defense, Intelligence and Homeland Security Symposium, April 24-25, 2013, Rosslyn, Virginia (Agenda and Brochure)
- Big Data Symposium: Analytics and Applications for Defense, Intelligence and Homeland Security, September 24-25, 2013, Rosslyn, Virginia (Agenda and Brochure)
- Government BIG DATA Symposium, November 19-20, 2013, Rosslyn, Virginia (Agenda and Brochure)

I participated remotely in two other big data conferences:

- Strata Conference 2013: Making Data Work, February 26-28, 2013, Santa Clara, California
  - Data Visualization Design Using Shneiderman’s Mantra: Overview First, Zoom and Filter, Then Details-on-Demand (See for example Tibco Spotfire)
  - Broad Data: What Happens When the Web of Data Becomes Real?
  - World Wide Web Expert Jim Hendler Receives Inaugural Strata "Big Data" Award
- ACT-IAC Big Data Symposium, March 15, 2013, Washington, DC

I read two Gartner reports about big open data and platforms:

- Gartner Says Big Data Makes Organizations Smarter, But Open Data Makes Them Richer
- Magic Quadrant for Business Intelligence and Analytics Platforms

I created three sets of Research Notes:

- Research Notes (Semantic Medline and IBM Watson) FINISH
- DataCatalogs.org
- Data Hub
• Open Government Data Catalogues
• Data Catalog Vocabulary (DCAT)
• Government Linked Data (GLD) Working Group
• Email

and

• W3C eGov eParticipation and Open Data
• New Data.gov Catalog
• New Data.gov Testing
• Access to testing for the new Data.gov site
• EU Data Portal Catalogue

I created three spreadsheets:

• BDS2013.xlsx (Knowledge Base)
• Data.gov_Catalog.csv (Catalog in CSV)
• EUDataPortalDump.xls (Catalogue in CSV that I converted to XLS)

for two Spotfire dashboards:

• Spotfire Dashboard (Harmonization of Next Data.gov and EU Open Data Portal Catalogs)
• Spotfire Dashboard (Earlier Data Catalog Analytics)
Finally, I created a matrix table to compare the interoperability of the two main catalogs:

<table>
<thead>
<tr>
<th>Item</th>
<th>Data.gov_Catalog.csv</th>
<th>EUDataPortalDump.xls</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Site</td>
<td>Data.gov</td>
<td>EU Data Portal</td>
<td></td>
</tr>
<tr>
<td>Row</td>
<td>8088</td>
<td>6269</td>
<td>About the same</td>
</tr>
<tr>
<td>Columns</td>
<td>27</td>
<td>256</td>
<td>Big difference</td>
</tr>
<tr>
<td>Web Address</td>
<td>External Links</td>
<td>url</td>
<td>Different name</td>
</tr>
<tr>
<td>Data Set Name</td>
<td>Name</td>
<td>title</td>
<td>Different name</td>
</tr>
<tr>
<td>Organization</td>
<td>Agency</td>
<td>groups</td>
<td>Different name</td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
<td>notes</td>
<td>Different name</td>
</tr>
<tr>
<td>MORE?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

using the DCAT vocabulary below.

MORE IN PROCESS

DCAT is an RDF vocabulary well-suited to representing government data catalogs such as [Data.gov](http://data.gov) and [data.gov.uk](http://data.gov.uk). DCAT defines three main classes:

- `dcat:Catalog` represents the catalog
- dcat:Dataset represents a dataset in a catalog
- dcat:Distribution represents an accessible form of a dataset as for example a downloadable file, an RSS feed or a web service that provides the data.

Another important class in DCAT is dcat:CatalogRecord which describes a dataset entry in the catalog. Notice that while dcat:Dataset represents the dataset itself, dcat:CatalogRecord represents the record that describes a dataset in the catalog. The use of the CatalogRecord is considered optional. It is used to capture provenance information about dataset entries in a catalog. If this distinction is not necessary then CatalogRecord can be safely ignored.

Government BIG DATA Symposium

Source: http://www.bigdataconference.net/ PDF

Big Data and Government R&D – Turning Overload into Exploitable Information Assets

In March 2012, the White House announced funding for a large new research and development initiative aimed at extracting and exploiting the “knowledge and insights from large and complex collections of digital data to help solve
some of the Nation’s most pressing challenges.” This is in addition to numerous ongoing “Big Data” programs that have been initiated across the Federal Government, including Homeland Security, Defense, Intelligence, Education, Energy, Health and Human Services, NASA, and NIST aimed at spurring scientific discovery and innovation.

This outstanding conference brings together the key government and industry experts who are shaping the direction of big data research and development across the Federal Government. They will provide you with an in-depth understanding of Federal agency strategy and plans, the status and forecast for key big data initiatives, and the latest tools and technologies being developed to exploit the massive amounts of information being collected at the Federal level.

- What are the most recent lessons learned from the commercial world?
- What applications are being developed for homeland security and intelligence analysis?
- What is the promise of health data analytics for developing new approaches to population health management, generating informatics-based treatments to major diseases, and coordinating response to public health crises?
- How is large-scale analytics being applied to ease overload for our warfighters?
- What is the analytic basis for the next-generation of energy capabilities?
- What role are new tools, techniques, and technologies – predictive analytics, cloud computing, metadata, etc. – playing in making big data analytics at the Federal level a reality? What role can industry play?

These and many other critical questions will be examined during this outstanding two-day event.

---

**Day 1: Tuesday, November 19, 2013**

Holiday Inn Rosslyn at Key Bridge, Arlington VA

9:00-9:05 Administrative Announcements

MR. TED MALONE, Big Data Architecture Lead, Microsoft Federal, Moderator

9:05-9:35 MR. WILCO van GINKEL, Strategist, Verizon Enterprise Solutions; Co-Chair, Cloud Security Alliance (CSA) Big Data Working Group

“Big Data Security—Really? Think Again” Slides

DNP: DR. SASI K PILLAY, Chief Technology Officer, Office of the Chief Information Officer, National Aeronautics and Space Agency (NASA)

“NASA Perspectives and Initiatives”


10:05-10:35 DR. CEREN SUSUT, Physical Scientist, Office of Science, Department of Energy (DOE) DOE Perspectives and Initiatives” Slides

DNP: MR. DAVID MUNTZ, Principal Deputy National Coordinator, Office of the National Coordinator for Health Information Technology, Health Human Services (HHS)

“HHS Perspectives and Initiatives”

10:35-10:55 Coffee and Networking
10:55-11:25 MR. MARK KRZYSKO, Deputy Director, Acquisition Resources and Analysis, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics (OUSD/AT&L)
“Transforming DoD Decision Making Through a Focus on Information and Data Management” Slides

11:25-11:55 DR. RANDY GARRETT, Program Manager, Defense Advanced Research Projects Agency (DARPA)
“Big Data R&D at DARPA” Slides
DNP: DR. KIRIT AMIN, Deputy Chief Information Officer, Department of Commerce (DOC) “DOC Perspectives and Initiatives”

11:55-12:25 MR. BILL FRANKS, Chief Analytics Officer, Teradata
“Putting Data Analytics to Work for Government” Slides

12:25-1:45 Lunch Break

1:45-2:15 MR. KEITH BRYARS, Client Executive, Federal Law Enforcement and National Security, Harris Corp; former Senior Executive Special Agent, Federal Bureau of Investigation (FBI)
“Big Data at the FBI” Slides

2:15-2:45 DR. NANCY GRADY, Technical Fellow, Data Scientist, Homeland and Civilian Solutions, SAIC
“Big Data for the Mission Side of the House” Slides
DNP: MR. JOHAN BOS-BEIJER, Director of Strategic Solutions and Senior Advisor, General Services Administration (GSA)
“GSA Perspectives and Initiatives”

2:45-3:15 MR. TED MALONE, Big Data Architecture Lead, Microsoft Federal
“Developing/Deploying the Big Data Analytics Architecture” Slides
SEE ABOVE: DR. CEREN SUSUT, Physical Scientist, Office of Science, Department of Energy (DOE)
DOE Perspectives and Initiatives”

3:15-3:45 Refreshments and Networking

3:45-4:15 MR. MARK JOHNSON, Director, Engineered Systems for MR. TOM PLUNKETT, Big Data Evangelist, Oracle Public Sector
“Big Data, Fast Data and Cyber Security” Slides
SEE ABOVE: DR. RANDY GARRETT, Program Manager, Defense Advanced Research Projects Agency (DARPA)
“Big Data R&D at DARPA”

4:15-4:45 DR. ERIC LITTLE, Vice President/Chief Scientist, Modus Operandi
“The Challenges of Big Data for Intelligence Missions” Slides
SEE ABOVE: MR. WILCO van GINKEL, Strategist, Verizon Enterprise Solutions; Co-Chair, Cloud Security Alliance (CSA) Big Data Working Group
“Big Data Security—Really? Think Again”

4:45-5:15 NO PRESENTATION
SEE ABOVE: DR. NANCY GRADY, Technical Fellow, Data Scientist, Homeland and Civilian Solutions, SAIC
“Big Data for the Mission Side of the House”

5:15-5:45 NO PRESENTATION
DNP: MR. AARON BURCIAGO, Senior Research Scientist, Elder Research; former Head Analyst/Director, Logistics Operations Analysis Office, Headquarters, US Marine Corps (USMC)
“Weapons Data Optimization”

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by mindtouch
Day 2: Wednesday, November 20, 2013

Holiday Inn Rosslyn at Key Bridge, Arlington VA

8:30-8:35 Opening Announcements

MR. TED MALONE, Big Data Architecture Lead, Microsoft Federal, Moderator

8:35-9:05 MR. PRESTON SMITH, Federal Government Manager, Tableau Software

Slides

9:05-9:35 MS. CARON KOGAN, Strategic Planning Director-Big Data, Lockheed Martin
“Unlocking Value from Your Big Data” Slides

9:35-10:05 MR. Hermanth SETTY, Chief Technology Officer, for MR. SHAWN KINGSBERRY, Chief Information Officer, Recovery Accountability and Transparency Board (RATB)
“Operational Challenges and Considerations in Large-Scale Data Analytics” Slides

10:05-10:35 MR. WO CHANG, Digital Data Advisor, Information Technology Laboratory and Co-Chair, Big Data Working Group, National Institute of Standards and Technology
“The NIST Big Data Public Working Group” Slides
DNP: DR. AUGIE TURANO, Information Technology Director, Veterans Informatics and Computer Infrastructure (VINCI), Department of Veterans Affairs (DVA)
“Exploration of Structured and Unstructured Clinical Data in the Veterans Administration”

10:35-10:55 Coffee and Networking

10:55-11:25 DR. JEANNE HOLM, Evangelist, data.gov, General Services Administration (GSA)
“Federating Big Data for Big Innovation” Slides

11:25-11:55 MR. DAMON DAVIS, Deputy Director, Health Data Initiative, Department of Health and Human Services (HHS)
“The Health Data Initiative” Slides

11:55-12:25 DR. DAVID WOLLMAN, Deputy Director, Smart Grid and Cyber-Physical Systems Program Office, NIST
“Smart Grid Data Initiatives—Green Button” Slides

12:25-1:30 Lunch Break

1:30-2:00 MR. RICHARD HEIMAN, Lead Data Scientist, Data Tactics Corporation
“Data Tactics: A Blended Approach to Big Data Analytics” Slides
DNP MR. TOM CONWAY, Senior Engineer, Office of the Project Manager, Night Vision/Reconnaissance, Surveillance and Target Acquisition (NV/RSTA), Program Executive Office—Intelligence, Electronic Warfare & Sensors (PEO-IEW&S), US Army
“Big Data and the Defense ISR Mission”

2:00-2:30 MR. GERARD CHRISTMAN, Senior Systems Engineer, Femme Corp Inc.; Contractor, Office of the Chief Information Officer, Department of Defense (DoD/CIO)
“Civil Information Integration in Support of the US Government” Slides

2:30-3:00 MR. FRANK STEIN, Director, Analytics Solution Center, IBM
“Applying Big Data Analytics to Government Missions” Slides
SEE ABOVE: MR. TOM PLUNKETT, Big Data Evangelist, Oracle Public Sector
“Big Data, Fast Data and Cyber Security”

3:00-3:20 Refreshments and Networking

3:20-3:50 JULIA SKAPIK, MD, MPH, Medical Officer for DR. KEVIN LARSEN, Medical Director, Meaningful Use, Office of the National Coordinator for Health Information Technology, U.S. Department of Health & Human Services
“HHS Perspectives on Health IT and Big Data” Slides
SEE ABOVE: MR. TED MALONE, Big Data Architecture Lead, Microsoft Federal
“Developing/Deploying the Big Data Analytics Architecture”

3:50-4:20 MR. JOSEPH FARGNOLI, Senior Fellow, The RITRE Corporation
“Predictive Analytics and Intelligence Operations” Slides
SEE ABOVE: MR. FRANK STEIN, Director, Analytics Solution Center, IBM
“Applying Big Data Analytics to Government Missions”

4:20-4:50 NO PRESENTATION
SEE ABOVE: DR. ERIC LITTLE, Vice President/Chief Scientist, Modus Operandi
“The Challenges of Big Data for Intelligence Missions”

4:50-5:20 NO PRESENTATION
SEE ABOVE: MR. JOSEPH FARGNOLI, Senior Fellow, The RITRE Corporation
“Predictive Analytics and Intelligence Operations”

Brochure

Source: http://www.bigdataconference.net/agenda-big-data-conference/ PDF

I. The Latest Federal Government Strategies, Plans, Needs and Initiatives

DOC Perspectives and Initiatives

DR. KIRIT AMIN
Deputy Chief Information Officer, Department of Commerce (DOC)

Department of Homeland Security – Big Data Analytics Needs Big Data Governance

MR. MICHAEL SIMCOCK
Chief Data Architect, and

MR. KELLY FAHEY
Senior Data Architect, Information Sharing Environment, Department of Homeland Security (DHS)

• DHS Data Management for Big Data Efforts
• New Information Sharing Challenges
• Meeting Mission Requirements for Actionable Information to Secure Borders, Facilitate Immigration and Respond to Situations Using Big Data Analytics
• DHS Data Management Best Practices
• DHS Efforts to Develop Data Tagging Strategies for Big Data Feeds
• DHS Efforts to Manage and Safeguard Information Sharing Access

HHS Perspectives and Initiatives

MR. DAVID MUNTZ
Principal Deputy National Coordinator, Office of the National Coordinator for Health Information Technology, Health and Human Services (HHS)

GSA Perspectives and Initiatives

MR. JOHAN BOS-BEIJER
Director of Strategic Solutions and Senior Advisor, General Services Administration (GSA)

Big Data – A Threat Perspective

MR. KEITH BRYARS
Client Executive, Federal Law Enforcement and National Security, Harris Corp; former Senior Executive Special Agent, Federal Bureau of Investigation (FBI)

• Context of Big Data from an FBI Perspective
• Overview of FBI Systems/Challenges Approach Post 9/11
• Transforming the FBI into an Intelligence-Led, Threat-Driven Organization FedLE vs. NSB (Unclassified vs. Classified)
• Future Challenges

DOE Perspectives and Initiatives

DR. CEREN SUSUT
Physical Scientist, Office of Science, Department of Energy (DOE)

The NIST Big Data Public Working Group

MR. WO CHANG
Digital Data Advisor, Information Technology Laboratory and Co-Chair, Big Data Working Group, National Institute of Standards and Technology (NIST)

The Big Data Senior Steering Group

MS. WENDY WIGEN
Co-Chair, Big Data Senior Steering Group, Networking and Information Technology Research and Development (NITRD) (invited)
II. Big Data Analytics and Applications for the Defense and Intelligence Missions – Opportunities, Challenges, Needs and Initiatives

Transforming DoD Decision Making Through a Focus on Information and Data Management

MR. MARK KRZYSKO
Deputy Director, Acquisition Resources and Analysis, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics (OUSD/AT&L)

• The Need for Acquisition Decision-making Data
• What Acquisition Visibility Offers
• The Criticality of Data Governance
• Managing Opportunities When Everyone Needs Data

Big Data R&D at DARPA

DR. RANDY GARRETT
Program Manager, Defense Advanced Research Projects Agency (DARPA)

Big Data and the Defense ISR Mission

MR. TOM CONWAY
Senior Engineer, Office of the Project Manager, Night Vision/Reconnaissance, Surveillance and Target Acquisition (NV/RSTA), Program Executive Office – Intelligence, Electronic Warfare & Sensors (PEO-IEW&S), US Army

Weapons Data Optimization

Mr. AARON BURCIAGO
Senior Research Scientist, Elder Research; former Head Analyst/Director, Logistics Operations Analysis Office, Headquarters, US Marine Corps (USMC)

• First Proofs in Big Data Analytics (Marine Corps)
• An Expeditionary Construct for Big Data
• Focus on the Warfighter
• Governance and Programme

Civil Information Integration in Support of the US Government

MR. GERARD CHRISTMAN
Senior Systems Engineer, Femme Comp Inc.; Contractor, Office of the Chief Information Officer, Department of Defense (DoD/CIO)

• NIEM (National Information Exchange Model)
• Shared Enterprise Services
The Challenges of Big Data for Intelligence Missions

DR. ERIC LITTLE
Vice President/Chief Scientist, Modus Operandi

• Handling the Problems of Scalability for Large Amounts of Intelligence Data
• Treating Information Needs for Different Types of Intelligence Analysts – Analysts at Government Agencies vs. Analysts in the Armed Services Operating at the Tactical Edge
• Issues of Scale Utilizing Semantic Technologies in Cloud Applications – Hybrid Approaches are Needed
• User-Friendly Applications for Driving Complex Intelligence Information to Analysts

Predictive Analytics and Intelligence Operations

MR. JOSEPH FARGNOLI
Senior Fellow, The RITRE Corporation

III. Government Large-Scale Analytic Programs and Initiatives – Status and Forecast

Exploration of Structured and Unstructured Clinical Data in the Veterans Administration

DR. AUGIE TURANO
Information Technology Director, Veterans Informatics and Computer Infrastructure (VINCI), Department of Veterans Affairs (DVA)

• Use of Structured Queries as Preparation to Research in Clinical Analysis, How are They Formulated and Refined?
• Use of Novel Unstructured Data (NLP) Analysis for Concept Extraction and Relationship Derivation
• Use of Hadoop and Open Source Tools in a NoSQL Environment
• Practical Application of Research Software and Findings as an Aid to the Clinician

Federating Big Data for Big Innovation

DR. JEANNE HOLM
Evangelist, data.gov, General Services Administration (GSA)

• Data.gov Federates Vast Data Resources from Across the Nation and the World
• Building Citizen-Driven Open Source Capabilities and Policies with the White House
• Federating Big Data and Providing Analytic Tools Allows Game-Changing Innovation
• Building Big Data and Open Data into the President’s Management Agenda and Open Data Policy

The Health Data Initiative

MR. DAMON DAVIS
Deputy Director, Health Data Initiative, Department of Health and Human Services (HHS)
Smart Grid Data Initiatives – Green Button

DR. DAVID WOLLMAN
Deputy Director, Smart Grid and Cyber-Physical Systems Program Office, NIST

• Energy Usage Information Standardization
• Green Button Utility Implementations
• Testing and Certification
• Data Privacy
• Customer Engagement and Applications

IV. Selecting and Developing Missions and Mission Apps

Putting Data Analytics to Work for Government

MR. BILL FRANKS
Chief Analytics Officer, Teradata

• Learn How to Cut Through the Hype Surrounding Big Data and Focus on What is Important
• Understand Common Pitfalls and How to Avoid Them
• Gain Perspective on the Cultural Challenges You’ll Face Along with Your Technological Challenges
• Leave Better Prepared to Tackle Big Data Analytics in Your Organization

Applying Big Data Analytics to Government Missions

MR. FRANK STEIN
Director, Analytics Solution Center, IBM

Big Data for the Mission Side of the House

DR. NANCY GRADY
Technical Fellow, Data Science, SAIC

• Organizing Big Data Concepts
• Big Data Engineering
• The Value to the Mission is in Diversity
• What Do I Do Next?

Unlocking Value from Your Big Data

MS. CARON KOGAN
Strategic Planning Director-Big Data, Lockheed Martin
• The Potential with Big Data • New Analytics Dynamics
• Techniques and Processes for Executing Big Data Analytics
• Examples of Ways the Big Data Analytics are Enriching Organization’s Revenue and Operational Imperatives

V. The Latest Tools, Techniques and Emerging Lessons Learned

Operational Challenges and Considerations in Large-Scale Data Analytics

MR. SHAWN KINGSBERRY
Chief Information Officer, Recovery Accountability and Transparency Board (RATB)

Developing/Deploying the Big Data Analytics Architecture

MR. TED MALONE
Big Data Architecture Lead, Microsoft Federal

Special Focus: Big Data and Cyber Security

Big Data Security – Really? Think Again!

MR. WILCO van GINKEL
 Strategist, Verizon Enterprise Solutions; Co-Chair, Cloud Security Alliance (CSA) Big Data Working Group

• Concerns About Big Data Security – Current Approaches: Why We Need a New Approach
• Why We Need “Data Self” to Touch on the Core of the (Big) Data Privacy Concerns
• Trust is Good, But Why a “Trust Spectrum of Verifiable Trust” is Better
• Why “Security ANTalytics” Might Be the Better Answer to a Better Big Data Security Eco-System”

Big Data, Fast Data and Cyber Security

MR. TOM PLUNKETT
Big Data Evangelist, Oracle Public Sector

• Using Big Data for Fast, Real Time Cyber Security Solutions that Analyze All of Your Data
• Integrating Big Data with Relational Databases
• Integrating Big Data with Event Processing to Handle High Velocity Streams of Data

Data Tactics: A Blended Approach to Big Data Analytics

MR. RICHARD HEIMANN
Lead Analytics Engineer, Data Tactics Corporation
Spotfire Dashboard

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

Research Notes

Also see: Research Notes

Irving’s site is worth checking out: http://blog.irvingwb.com/blog/2013/1...ions.html#more

This is where IBM is headed with cognitive computing: http://www-03.ibm.com/innovation/us/...cosystem.shtml

Here is an additional link based on IBM's announcement about Watson in the Cloud. IBM Releases a Legion of Watsons: http://www.popsci.com/blog-network/z...legion-watsons

Also, I am enclosing some information about Semantic Verses, as well as some contact info for Dr. Walid Saba. In addition to the approach and app concepts discussed in the presentation, Walid and I were discussing using Magnet technology to construct natural language front-ends for enterprise data (and big data) integration across multiple sources. I think there might be a good fit for Semantic Verses capabilities with the semantic medline demo that you’ve been working on with YARC Data and others. Also, it would be interesting to explore the potential fit and advantages of combining Be Informed, YARC, Semantic Verses and big data analytics for other public and private sector demos.

W3C eGov eParticipation and Open Data

Source: http://www.linkedin.com/
groupItem?view=&gid=1800648&type=member&item=5809394179612106752&commentID=5810771264133963776&goback=%2BBBjkmxwkEOwBjLE28YyDlxycyEO7_TA_giuRN#commentID_5810771264133963776

November 22, 2013, 11-12:30 pm Eastern US time

Jeanne Holm, Data.gov, Evangelist at GSA

The upcoming meetings of the W3C eGovernment Community should be exciting! We have a full agenda with presentations from a variety of countries. A focus for this meeting is on eParticipation, smart governance, and open data ecosystems. The best paper and key findings from the recent ICEGOV (http://icegov.org) conference on eGovernment will be highlighted.

The next meeting is tomorrow, November 22, at 11:00 am-12:30 pm (US EST). Time zone converter: http://www.timeanddate.com/worldclock/
Agenda

8:00-8:10: Welcome and introductions (All)
8:10-8:45: "Harnessing the duality of e-Participation--Social Software Infrastructure Design"
by Lukasz Porwol, Adegboyega Ojo, and John Breslin from Insight and National University of Ireland, Galway
8:45-9:00: Smart Governance for Smart Industries, Antonio Cordella, London School of Economics
9:00-9:10: Participatory Governance and Social Media, Jeanne Holm
9:10-9:20: Open Data Ecosystem, Jeanne Holm
9:20-9:30: Feedback, announcements, suggestions (All)

Scribe: TBD

--Telecon line: Dial +1-617-761-6200 or sip: zakim@voip.w3.org then conference code 3468# ("EGOV#")
--W3C IRC channel #egov, see http://www.w3.org/Project/IRC/ or use http://irc.w3.org/?channels=egov
--Group access via the W3C at http://www.w3.org/community/egovernance/ (archive at http://www.w3.org/egov/) and also via LinkedIn at the W3C eGovernment Interest Group: http://www.linkedin.com/groups?gid=1800648&trk=hb_side_g

New Data.gov Catalog

From: Holm, Jeanne M (1760) [mailto:jeanne.m.holm@jpl.nasa.gov]
Sent: Tuesday, November 19, 2013 5:29 PM
To: Brand Niemann
Cc: Strawn, George
Subject: Re: Access to testing for the new Data.gov site

Brand--

Just making sure you see the Data.gov catalog file continues to be available at http://catalog.data.gov/dataset/datalog-catalog

--Jeanne

Jeanne Holm
Evangelist, Data.gov
U.S. General Services Administration
Cell: (818) 434-5037
Twitter/Facebook/LinkedIn: JeanneHolm

From: Brand Niemann <bniemann@cox.net>
Date: Monday, November 18, 2013 4:03 AM
To: Jeanne Holm <Jeanne.M.Holm@jpl.nasa.gov>
Cc: "Strawn, George" <Strawn@nitrd.gov>
Subject: RE: Access to testing for the new Data.gov site

Jeanne, You are welcome. Please see: http://semanticommunity.info/Census_Semantic_Knowledge_Base

George, FYI, Brand

From: Holm, Jeanne M (1760) [mailto:jeanne.m.holm@jpl.nasa.gov]
Sent: Friday, November 15, 2013 11:49 AM
To: Brand Niemann
Subject: Re: Access to testing for the new Data.gov site

Thanks so much Brand! Looking forward to seeing your site.

--Jeanne

******************************************************************************
Jeanne Holm
Evangelist, Data.gov
U.S. General Services Administration
Cell: (818) 434-5037
Twitter/Facebook/LinkedIn: JeanneHolm
******************************************************************************

New Data.gov Testing

From: Brand Niemann <bniemann@cox.net>
Date: Friday, November 15, 2013 9:38 AM
To: Jeanne Holm <Jeanne.M.Holm@jpl.nasa.gov>
Subject: RE: Access to testing for the new Data.gov site

Jeanne, I did this for you. I am working on a SemanticData.gov for Census.

Best regards, Brand

Introduction

http://www.loop11.com/usability-test...6/introduction

Usability Test for Data Site
Thank you for participating in this usability test for a new design for a government web site.

During this test, you will be asked to complete tasks using a new version of a site that is currently in development. This site is a work in progress, and so there may be times when the task is difficult to complete or the site doesn't work the
way you expect it to. Your feedback will help to improve the site and make it better for future users. Some questions will ask you to type in your response, and others will open up a web page and have a question at the very top. Whenever you have particular difficulty with a task, please make a note explaining what you expected to be able to find or have happen and, if possible, what would have made the task easier. At the end of the test, there will be a question asking you to put in any comments.

The tasks should be done starting at the test site at http://Next.Data.gov, although in some cases this will link you out to other sites or even to the current version of Data.gov. For this test, see if you can find it through the content in the site itself first before typing in a URL directly of another site.

The final outputs of this test will help to inform the redesign of the site and make it easier for people to find government data. Thank you so much for your time and attention. If you have any questions or problems during the test, please contact Jeanne Holm: 818-434-5037, jholm@jpl.nasa.gov, or @JeanneHolm.

**Preparation**

http://www.loop11.com/usability-test...8/preparation/

In this exercise, you will be asked to carry out a number of tasks on a website. At the top of the page, you will see the current task you are to complete.

**Task**

To complete each task, you will need to navigate through the website to the page that contains the information for the task or the page which you think best answers or completes the task.

When you have navigated to the page which contains the information for the task, select 'Task Complete'.

**Task complete**

If you can't find the page or are having difficulty you can select 'Abandon Task'.

**Abandon task**

Please remember – we are not evaluating you, we are evaluating the website.

When you have read these instructions click NEXT to begin the first task.

**Question 1**

http://www.loop11.com/usability-test...uestion/75735/

What is your name?
Brand Niemann

**Question 2**

http://www.loop11.com/usability-test...uestion/75822/

What do you do at work?
Semantic Data Science
Question 3

http://www.loop11.com/usability-test...uestion/75823/
What is the name of your organization or agency?
Semantic Community

http://www.loop11.com/usability-test...uestion/75824/
Have you used Data.gov before?
Yes

Question 4

http://www.loop11.com/usability-test...uestion/75825/
If you have used Data.gov before, how often do you use it?
As little as possible

Question 5

http://www.loop11.com/usability-test...uestion/75826/
Would you identify yourself as a (choose one or two):
Developer Researcher Entrepreneur/innovator Data journalist Data scientist Data owner/manager Government employee Citizen Other, please specify
Data journalist Data scientist

Question 6

http://www.loop11.com/usability-test...uestion/75827/
What platform and device are you using for testing (e.g., computer, mobile, or tablet)? Please be specific as to the operating system, hardware, and browser.
Desktop computer, Microsoft Windows XP, Media Center Version, Intell Pentium 2.8 GHz and 3.00 GB RAM Google Chrome

Question 7

http://www.loop11.com/usability-test...uestion/76419/
Great, let's get started! Spend a few minutes exploring the site at http://next.data.gov. What website is this and how can you tell?
Data.gov with the confusing interface and no obvious link to download the entire catalog in a spreadsheet

Question 8

http://www.loop11.com/usability-test...uestion/76431/
What do you think is represented by the blue graphic at the top of the page? What do you think of this graphic?
No opinion
Question 9

http://www.loop11.com/usability-test...uestion/76432/
What do you think is the purpose of the white box on the left side of the blue graphic?
Start Searching

Question 10

http://www.loop11.com/usability-test...uestion/76433/
What do you think you'll find when you go to the listing of words on the left side of the page under the white box?
Another web page without actual data

Question 11

http://www.loop11.com/usability-test...uestion/76434/
What is the purpose of the posts (boxes) in the body of the page below the graphic?
More non-data information

Task 1

http://www.loop11.com/usability-test...98/task/65848/
You are writing a blog post about this site, where would you go to find the purpose of the site

Task 2

http://www.loop11.com/usability-test...98/task/65849/
You are looking for US government data, where on the site would you go to find the data

Task 3

http://www.loop11.com/usability-test...op11/new/task/
Find the most important or highlighted data on the site

Task 4

http://www.loop11.com/usability-test...op11/new/task/
You are a developer looking to create a new web site. Where would you find web services that you could use?

Task 5

http://www.loop11.com/usability-test...op11/new/task/
You're cooking a dinner for friends and interested in finding nutrition information about various foods, a friend told you about an app that provides nutrition information. Where would you find that?
Task 6

http://www.loop11.com/usability-test...op11/new/task/
Find a listing of other open data sites in the United States

Task 7

http://www.loop11.com/usability-test...op11/new/task/
You have a question for the Data.gov team, where would you go to ask that question?

Task 8

http://www.loop11.com/usability-test...op11/new/task/
Where would you go to find information about safety-related topics and data?

Task 9

http://www.loop11.com/usability-test...op11/new/task/
Find information and data about oceans

Task 10

http://www.loop11.com/usability-test...op11/new/task/
For this task, open another window from your browser and start at http://Next.Data.gov.
You are interested in buying a new, energy-efficient computer, find a dataset about the energy rating of notebook computers. Enter the URL of the dataset as your answer.
I would use a Google Search for this to see what my data options are

Task 11

http://www.loop11.com/usability-test...op11/new/task/
Look at the full dataset, are you able to see the data?
I am not sure what you mean by "see the data"?

Task 12

http://www.loop11.com/usability-test...op11/new/task/
You have a question about this dataset, who would you contact to ask that question?
I would use the Google Search to find the original source of the data for that and also for who else may provide the data and their interpretation of the data

Task 13

http://www.loop11.com/usability-test...op11/new/task/
Do you have any comments about the tasks that you just completed? Did any cause you particular difficulty or frustration?
They are just not the data science process I follow to discover, cleanup, make something that works, and publish the results

**Task 14**

http://www.loop11.com/usability-test...op11/new/task/
What is your first impression of Next.Data.gov?
Still not something a data scientist/data journalist would really use in their work

**Task 15**

http://www.loop11.com/usability-test...op11/new/task/
What do you think about the look and feel of this layout and design?
Just put the catalog in my face with three clicks: See, Sort, and Download

**Task 16**

http://www.loop11.com/usability-test...op11/new/task/
Was it easy to find what you were looking for?
No

**Task 17**

http://www.loop11.com/usability-test...op11/new/task/
Are the navigation names clear?
No, I prefer faceted browsing

**Task 18**

http://www.loop11.com/usability-test...op11/new/task/
What features did you like the most or would find the most useful?
A link at the very top to download the entire catalog in a well-formatted spreadsheet

**Task 19**

http://www.loop11.com/usability-test...op11/new/task/
What features did you like the least or would find the least useful?
All the eye candy on the site

**Task 20**

http://www.loop11.com/usability-test...op11/new/task/
How would you rate the quality of the content and data provided on Data.gov?
I would not find that a useful exercise since I would rely on the originators of the data sets for the quality of the data metadata
Task 21

http://www.loop11.com/usability-test...op11/new/task/

Did you find the content you expected to when you went to each page or searched for data or content?
No, I want the spreadsheet I keep mentioning. I want the data about the data! That is the Digital Government Strategy - all your content as data

Task 22

http://www.loop11.com/usability-test...op11/new/task/

If you could change something about the site, what would you change?
Nothing

Task 23

http://www.loop11.com/usability-test...op11/new/task/

What type of content would you like to see on this site?
I would like to see a new site with Data Science Training and Products built by data scientists and not technologists

Task 24

http://www.loop11.com/usability-test...op11/new/task/

What are three government data resources that you use in your daily work?
Census, Japan, and European that I have work with at http://semanticommunity.info/

Task 25

http://www.loop11.com/usability-test...op11/new/task/

What were the last 2-3 useful sites you went to for help in creating something with data or analyzing data?
See: http://semanticommunity.info/#Events_Calendar

Task 26

http://www.loop11.com/usability-test...op11/new/task/

Is there anything else you would like to share about the site, your reaction to it, or the test?
Focus on helping with real data problems like: http://semanticommunity.info/Healthcare.gov

Task 27

http://www.loop11.com/usability-test...op11/new/task/

Usability Test for Data Site

Thank you for completing this test and survey. If you have any questions, please contact Jeanne Holm (Jeanne.Holm@jpl.nasa.gov or 818-434-5037). Please complete this test by November 19.
Hi all--

Thanks so much for volunteering (or being volunteered) to help out with testing the new design concept for Data.gov. The test is starting a little later than planned, but you are able to jump on and test at any time. The test should take about 45 minutes and is done online by starting at http://www.loop11.com/usability-test.../introduction/ No plug-ins or special browser are needed and the test site is unrestricted. Remember, we are testing the site, we are not testing you. You can't fail! Just be sure to make a note if you find anything frustrating or particularly difficult to do. If you have general suggestions for improving the new design, those are appreciated too and you can add those in at the end of the test.

If you would like an orientation or a chance to discuss it in advance, just drop me a line and I'm happy to walk you through it.

If you complete the test before end of day, Tuesday, November 19, your input will be in the first round of recommendations. However, we will also need some people to look at the site late this month after the team has made some of the recommended improvements. Let me know if it's more convenient for you to be a tester in the second round.

Thanks again for your time and attention. I really appreciate it!

--Jeanne

**********************************************************************
Jeanne Holm
Evangelist, Data.gov
U.S. General Services Administration
Cell: (818) 434-5037
Twitter/Facebook/LinkedIn: JeanneHolm
**********************************************************************

Hi--

Our team is in the midst of redesigning Data.gov and would like you to help us by testing the new site in development. You were selected because you represent one or more of Data.gov's primary audiences (developers, researchers, businesses, entrepreneurs, data scientists, data journalists, publishers, and/or citizens).
The testing will be conducted virtually this week (November 12-15) through a simple online form that will guide you through the site. If you are in the Washington DC area, I invite you to consider in-person testing as an option. The testing will prompt you to look for data or complete tasks on the site, and then note any comments you have about the process. The final outputs of this test will help to inform the redesign of the site and make it easier for people to find U.S. government data. The test will take about 45 minutes.

The test site is at http://Next.Data.gov. If you are interested, just send me a return email and I'll send you the instructions. If you have any questions, don't hesitate to call (818-434-5037) or email me directly.

Thank you so much for your time and attention.

EU Data Portal Catalogue

From: ODP HELPDESK [mailto:ODP-HELPDESK@publications.europa.eu]
Sent: Thursday, November 21, 2013 9:08 AM
To: 'bniemann@cox.net'
Subject: FW: ID103378 - FW: [Contact Us] Request: Please send the catalogue of the 6135 data sets in a spreadsheet

Dear Sir,

Further to your request here below, please find attached the dump of the data base with the metadata of all the records.

Please let us know if this corresponds to what you expected.

Thank you.

Regards,

Tommaso Materossi
Open Data Portal Helpdesk
Publications Office of the European Union
E-mail: odp-helpdesk@publications.europa.eu
Website: http://open-data.europa.eu/
Privacy Statement: http://publications.europa.eu/others...y/index_en.htm

-----Original Message-----
From: ODP HELPDESK On Behalf Of bniemann@cox.net
Sent: mardi 15 octobre 2013 17:56
To: ODP HELPDESK
Subject: [Contact Us] Request: Please send the catalogue of the 6135 data sets in a spreadsheet

Brand Niemann (bniemann@cox.net) sent a message using the contact form at http://open-data.europa.eu/en/contact.

I am not sure my request was received so I am sending it again.
Story: Mining the Big Data Symposium for Big Data Sets and Ideas

So I keep going to these meetings week-after-week looking for big data sets and ideas and amazingly I keep finding both!

I was disappointed not to hear from Gus Hunt, especially since he said of the Intelligence Community: *We 'Try To Collect Everything And Hang On To It Forever'*, long before the NSA denied it was doing that!

But I was not disappointed when I heard the two part keynote presentations by John Marshall, Deputy Director, National System for Geospatial Intelligence (NSG) Program Management Office, National Geospatial-Intelligence Agency (NGA), on "Big Data: Benefits of Universal Data Access", and Todd Myers, Lead Global Computer Architect, National Geospatial-Intelligence Agency, (NGA), on "NSG/NGA Architecture and Big Data". John Marshall stressed the need for "data in context" and Todd Myers said they need three things: "metadata, entity extraction, and contextual resolution." These are what I and others call fundamental Semantic Web concepts.

I asked Todd Myers my usual tough question at these conferences: How much of what you described is building big data bureaucracy and infrastructure and how much is actual new data science products and services? His answer was the usual: More the former than the latter. But I really liked that he said the new ICITE architecture is about giving every data set a well-defined URL in the cloud. This is like Sir Tim Berners-Lee *Getting to Five Stars of Linked Open Data* for the Semantic Web.

This was something I also asked Donna Roy, Director, Information Sharing Environment (ISE), Department of Homeland Security, who presented "Using Large-Scale Analytics to Leverage the Information Sharing Challenge" and also presented for Carrie Boyle, Division Lead, Standards and Architecture, Program Manager– Information Sharing Environment (PM-ISE), Office of the Director of National Intelligence (DNI), on "Architectures and Standards for Big Data". Her reply was: That NIEM and the ISE is moving toward the ICITE and the NGA architecture for the IC involving the use of well-defined URLs for every data set. I was intrigued by Donna's mention of a new Presidential Order #10 calling for a Data Aggregation Architecture.

The speakers with big data sets besides those mentioned above were:


The industry vendor presentation that impressed me the most was by Dr. Rod Fontecilla, Vice President, Application Modernization, Unisys Federal Systems, on “Critical Steps in Building Your Data Analytics Environment” which included:

- Hadoop Implementations
- Our Experience Building Large and Complex Data Analytics – “Hadoop-la” is Misguided and Can Potentially Drive Precious Resources in the Wrong Direction
- Clear and Actionable Roadmap to Build a Flexible, Scalable and Reliable Data Analytics Environment, Tightly Integrated to Your Existing Data Architecture
• Reference Architecture for Data Analytics, and an Analytics Framework for the Creation of Data Products that Help Drive Business Decisions

My question to him afterwards in an email was: Does it always have to be this complicated?

Finally I liked the comment of Jill Singer, Partner, Deep Water Point; and former Chief Information Officer, National Reconnaissance Office (NRO), whose keynote: “Accelerating Mission Advantage through Cloud and Big Data Convergence”, included the statement that ‘there are very few data scientists in the Washington, DC, area that have both the theoretical and practical qualifications need.’ I agree completely and that is why I am helping the Federal Big Data Senior Steering Committee with Workforce Training.

I also like the suggestion by Alan Briggs, Data Scientist Elder Research, Inc., speaking on "Higher Order Leadership for Data Science in the IC: Herding Cats and Leading Elephants", who suggested fostering Learn-Network-Do environments. This is what the Meetups like Data Community and Data Science DC do with thousands of members and hundreds of meetup attendees.

Finally, I plan to download and try MapR which describes itself as "One Platform for Enterprise Hadoop", Platfora which says it does not require Java programmers, participate in the October 1st DHS Industry Day, and work with the Hurricane Sandy data sets for the NGA Pilot. I have worked previously with the DHS-Funded Global Terrorism Database.

Big Data Symposium: Analytics and Applications for Defense, Intelligence and Homeland Security

Source: http://www.bigdataevent.net/, http://www.bigdataevent.net/download-brochure/ (PDF), http://www.bigdataevent.net/download-agenda/ (PDF)

Holiday Inn Rosslyn at Key Bridge, Arlington VA
Updated as of Monday, September 09, 2013
Subject to Change

Today’s Warfighters and intelligence analysts have access to an ever-increasing number of sensors, imagers, internet artifacts, open source and other sophisticated collection devices and mechanisms, to the point that a major challenge has become how to sift through this massive amount of information to find the most critical and actionable items of intelligence. Increasingly, this must be accomplished in near-real time and the information must be packaged in a format capable of being shared with all other pertinent parties. The result is that sensor, computer and communication technologies are being strained beyond capacity to keep pace with current and future information management and analysis needs. ‘Big Data’ tools, techniques, and technologies seek to provide the means to analyze, exploit and share conclusions drawn from this seemingly overwhelming information load.

This outstanding symposium brings together the key government and industry experts who are shaping the direction of big data research and development for defense, intelligence and homeland security. What are the latest DoD, Intelligence Community, and Homeland Security needs and initiatives in big data? How is big data analytics being applied to ISR, intelligence sharing, GEOINT fusion, video analytics, atmospherics, identity, biometrics, and a whole range of other critical mission applications? What role are new tools, techniques, and technologies – predictive analytics, cloud computing, metadata, etc. – playing in making big data analytics a reality? What are the future challenges and opportunities? What role(s) can industry play? These and many other critical questions will be examined during this outstanding two-day event.

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by mindtouch
Day 1 Tuesday, September 24, 2013

9:00-9:05 Opening Announcements

9:05-9:35 MR. GUS HUNT, Chief Technology Officer, Central Intelligence Agency (CIA)
Keynote: “Emerging Technical Challenges and Capabilities”
My Note: Unable to attend. MS. JILL SINGER spoke in this place. See below.

9:35-10:05 MR. JOHN MARSHALL, Deputy Director, National System for Geospatial Intelligence (NSG) Program Management Office, National Geospatial-Intelligence Agency (NGA)
Keynote: “Big Data: Benefits of Universal Data Access” Slides

“NSG/NGA Architecture and Big Data” Slides

10:35-10:50 Networking and Coffee Break

10:50-11:20 MS. JILL SINGER, Partner, Deep Water Point; former Chief Information Officer, National Reconnaissance Office (NRO)
Keynote: “Accelerating Mission Advantage through Cloud and Big Data Convergence” Slides
• Cloud and Big Data Convergence • Cloud Computing Revolutionized IT Infrastructure Services
• Computing Power to the Masses with the Swipe of a Credit Card • Reduced Barriers to Market Entry for Products and Services • Is Big Data the Next “Battering Ram” to Knock Down More Doors Currently in the Way of New Competition? • Convergence of Cloud Computing and Big Data as a Potentially Powerful New Platform for Mission and Business Advantage

11:20-11:50 MS. LISA SHALER-CLARK, Deputy Director/Program Manager, Futures, Intelligence and Security Command (INSCOM), US Army
“INSCOM Perspectives and Initiatives” Slides

“Big Data Analytics for Homeland Security” Slides
• DHS S&T HSARPA Big Data Initiatives, Including Needs Assessments for DHS Components
• Description of the Engagements Methodology and S&T Big Data Enclave
• Partnerships in Technology Development and Delivery
• Transition Strategy and Initial Lessons Learned

12:20-12:35 Lunch Sponsor: MR. BOBBY CAUDILL, Teradata Slides

12:35-1:35 Lunch Break

1:35- 2:05 DR. CAREY SCHWARTZ, Program Officer and Data to Decision Lead, Data-to-Decision Science & Technology Initiative, Office of Naval Research (ONR)
“Big Data and Data-to-Decision” Slides
• Why is DoD Interested in Data to Decisions?
• What is DoD Trying to Accomplish in Data to Decisions?
• Identification of S&T Gaps for Data to Decisions

2:05-2:35 MR. BOBBY CAUDILL, Program Director Government Industry Marketing and Solutions, Teradata
“Future Data Science— Challenges and Capabilities” Slides
10:55-11:25 MS. CARRIE BOYLE, Division Lead, Standards and Architecture, Program Manager– Information Sharing Environment (PM-ISE), Office of the Director of National Intelligence (DNI) “Architectures and Standards for Big Data” Slides

11:25-11:55 DR. MARK A. LIVINGSTON, Head, Visual Analytics Laboratory, Naval Research Laboratory “Human Factors for Big Data Visualization” Slides
  • Overview of Relevant Perceptual and Cognitive Theories
  • Discussion of Multivariate Visualization Techniques
  • Summaries of User Studies of Multivariate Visualization
  • Discussion of Results and Implications

11:55-12:25 DR. ROD FONTECILLA, Vice President, Application Modernization, Unisys Federal Systems “Critical Steps in Building Your Data Analytics Environment” Slides
  • Hadoop Implementations • Our Experience Building Large and Complex Data Analytics – “Hadoop-la” is Misguided and Can Potentially Drive Precious Resources in the Wrong Direction
  • Clear and Actionable Roadmap to Build a Flexible, Scalable and Reliable Data Analytics Environment, Tightly Integrated to Your Existing Data Architecture • Reference Architecture for Data Analytics, and an Analytics Framework for the Creation of Data Products that Help Drive Business Decisions

12:25-1:30 Lunch Break

1:30-2:00 MR. DERRICK DICKEY, Identity and Access Management Specialist, Public Sector/Intelligence Communities, Dell Software “Leveraging Big Data to Defeat the Insider Threat — Combining Behavioral & Conversational Data into One Holistic Audit” Slides

2:00-2:30 MR. TASSO ARGYROS, Co-President, Teradata-Aster “The latest R&D in Big Data Analytics” Slides

2:30-3:00 MR. TIM PAYDOS, Director, Worldwide Business Analytics and Optimization, IBM
  • Intensifying Mission and Threat Challenges vs. Explosion in Volume, Variety and Velocity of Potential Data Sources
  • Transforming the Way We Harness the Value of Big Data and Analytics
  • Best Practices - Common Challenges Mission Leaders Face; Efforts Undertaken; What We have Learned are the Keys to Success

3:00-3:15 Networking and Refreshment Break

3:15-3:45 MR. MARK JOHNSON, Director, Engineered Systems, and MR. TOM PLUNKETT, (Moderator) Big Data Evangelist, Oracle Public Sector “Big Data, Cyber Security, and the Information Architecture” Slides

3:45-4:15 MR. JEREMY WALSH, Lead Associate, Booz Allen Hamilton “Data Science for Mission Effects: Case Studies on Transforming the Role and Effectiveness of the Analyst” Slides

4:15-4:45 MR. JIM FIORI, Senior Principal Solutions Architect, and MR. TOM WHITE, Senior Director, Business Development, MapR Technologies “Architecture Matters: One Platform for Enterprise Hadoop” Slides

My Note: Did not presented
MR. TOAN DO, Senior Director, Federal Programs, MapR Technologies
“Avoiding the ‘Gotchas’ of Hadoop Deployments” Slides

DR. SUZANNE YOAKUM-STOVER, Director, Institute for Modern Intelligence (IMI)
“Large-Scale Analytics and the Intelligence Mission” Slides

MR. DAN GREEN, Director, Data Sciences Team, Space and Naval Warfare (SPAWAR) Systems Center, US Navy
“SPAWAR Perspectives and Initiatives” Slides

Research Notes

I suggested conference attendees become familiar with the Semantic Medline - YARCDATA Graph Appliance work our Data Science Team is doing. One question was how does that compare to the IBM Watson work described below:

Saturday, August 17, 2013 1:49 PM

Here is a presentation [1] [2] by Jim Hendler that sheds light on the many aspects of IBM Watson. In particular, it sheds light on the collective roles of Linked Data, LOD Cloud (e.g., DBpedia [3]), Relation Semantics, NLP, and Machine Learning.

Links:

1. http://www.slideshare.net/jahendler/watson-summer-review82013final
   -- basic URL that denotes the Presentation.

   -- Linked Data URI that denotes the Presentation.

3. http://www.slideshare.net/jahendler/watson-summer-review82013final/22
   -- basic URL that denotes a specific Slide that covers use of DBpedia.

Kingsley Idehen
Founder & CEO
OpenLink Software

Company Web: http://www.openlinksw.com

Personal Weblog: http://www.openlinksw.com/blog/~kidehen

Twitter/Identi.ca handle: @kidehen

Google+ Profile: https://plus.google.com/112399767740508618350/about

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by mindtouch
Big Data Conferences

I have attended and written about many big data conferences. One of the biggest in terms of number of conferences (3 per year for 3 years), attendees (1000s), and Tweets (1000s) is the recent Strata 2013: Making Data Work. I thought it would be valuable to distill that conference in preparation for next week's Government Big Data Symposium.

The Strata 2013 Conference topics were:
- Hadoop in Practice
- Beyond Hadoop
- Connected World
- Data Science
- Design
- Law, Ethics, and Open Data
- Internet of Things
- Data Driven Business Data Design
- Enterprise IT

Someone new to big data would find some unfamiliar terms in the above list like Hadoop which Wikipedia defines as: Apache Hadoop is an open-source software framework that supports data-intensive distributed applications.

The Government Big Data Symposium topics are:
- The Latest Federal Government Strategies, Plans, Needs and Initiatives
- Technical Challenges and Mission Strategies
- Advanced Tools and Techniques
- Implementation Strategies and Lessons Learned

The Big Data Analytics and Applications for Defense, Intelligence and Homeland Security Symposium topics are:
- Government Needs, Initiatives, Opportunities and Challenges
- Emerging Applications for Defense and Intelligence
- The Latest Tools, Techniques and Technologies – Data Collection/Discovery, Deep/Predictive Analytics, Cloud, Scalability, Security, etc.

The connection between these three conferences is that Gartner Says Big Data Makes Organizations Smarter, But Open Data Makes Them Richer and the Gartner Magic Quadrant for Business Intelligence and Analytics Platforms shows one how to do that with tools that work well with open data.

Strata 2013: Making Data Work emphasized the need for smart and adult data, fast and agile methods and tools that could solve difficult problems and provide real business value - just data is not a business model.
There were 11 sessions of interest, of particular interest and relevance to government big data, especially one that I had written about previously entitled Broad Data: What Happens When the Web of Data Becomes Real?, by James Hendler (RPI) which said:

Recently we have begun to see the emergence of a new online data challenge—that of the “Broad data” that emerges from millions and millions of raw datasets available on the World Wide Web. For broad data, the new challenges that emerge include Web-scale data search and discovery, rapid and potentially ad hoc integration of datasets, visualization and analysis of only-partially modeled datasets, and issues relating to the policies for data use, reuse and combination. In this talk, we present the broad data challenge and discuss potential starting points for solutions. We illustrate these approaches using data from a “meta-catalog” of over 1,000,000 open datasets that have been collected from about two hundred governments from around the world.

I found the RPI International Data Set Catalog difficult to use and the research paper that explains why linking open data is important difficult to re-create because the raw data was not provided.

The first problem with data catalogs is that their format is not standardized - they do not have a standard set of descriptive items like a library card catalog for example. Second, they do all contain sufficient metadata (data about the data) to allow one to work with the data without a subject matter expert. Third, they do not always contain links to the actual data and in a format that can be readily used. And fourth, they may not contain a data dictionary.

I have found that the best source of metadata and data for data sets that can be integrated comes from government statistical agencies like the United States (Census Bureau Annual Statistical Abstract), Europe (Eurostat Annual Statistical Yearbook) and Japan (Statistical Agency Annual Statistical Yearbook).

I have suggested a data science and system of systems approach to this problem and need for data integration as follows: World Catalog - that helps identify the best Individual Catalogs - that help identify the best Value-added Data Sets. This makes it more than just "an IT project that turns the crank on lots of data".

This is illustrated in a Spotfire dashboard I created using data sets and visualization of each these three parts of the system. Spotfire is a leader in the Gartner Magic Quadrant for Business Intelligence and Analytics Platforms for the reasons given in their report. It also shows Data Visualization Design Using Shneiderman’s Mantra: Overview First, Zoom and Filter, Then Details-on-Demand presented at the Strata 2013 Conference.

So here is what I did:

- Started with:
  - Big Data Symposia Content
  - Gartner Article and Magic Quadrant Reports
  - Strata 2013 Conference Highlights
  - Data Catalogs and Research Notes
- Copied it to MindTouch to make it Digital Government Strategy compliant
- Made it Big Data and Machine Readable

The results are presented in:

- A MindTouch Knowledge Base
My conclusions and recommendations are:

• Jim Hendler’s “meta-catalog” of over 1,000,000 open datasets that have been collected from about two hundred governments from around the world cannot be verified.
  ◦ Specifically he says US Data.gov has 441,339 data sets, but the catalog has only 5,999 and France has 353,406, while there are only 205, 595 CSV and XLS data sets!

• This big data analytics and application show four problems with data catalogs: format is not standardized; insufficient metadata (data about the data) to allow one to work with the data without a subject matter expert; they do not always contain links to the actual data and in a format that can be readily used; and they may not contain a data dictionary.
  ◦ The example of the SAHIE Program data set does!

• Bottom Line: All the work with Data Catalogs does not really help with data integration as I have been able to show!
  ◦ Recall Slide 11!

Slides

Slide 1 Big Data Conference: Analytics and Applications for Federal Big Data

http://semanticommunity.info/

http://gov.aol.com/bloggers/brand-niemann/

http://semanticommunity.info/Big_Data_Symposia
Preface

- I have attended and written about many big data conferences. One of the biggest in terms of number of conferences (3 per year for 3 years), attendees (1000s), and Tweets (1000s) is the recent Strata 2013: Making Data Work. I thought it would be valuable to distill that conference in preparation for next week’s Government Big Data Symposium and next month’s Big Data Analytics and Applications for Defense, Intelligence and Homeland Security Symposium.

Strata 2013 Conference: Making Data Work

- The Strata 2013 Conference topics were:
  - Hadoop in Practice
  - Beyond Hadoop
  - Connected World
  - Data Science
  - Design
  - Law, Ethics, and Open Data
  - Internet of Things
  - Data Driven Business Data Design
  - Enterprise IT
- Someone new to big data would find some unfamiliar terms in the above list like Hadoop which Wikipedia defines as:
  - Apache Hadoop is an open-source software framework that supports data-intensive distributed applications.

Big Data Symposia
Big Data Symposia

- The Government Big Data Symposium topics are:
  - The Latest Federal Government Strategies, Plans, Needs and Initiatives
  - Technical Challenges and Mission Strategies
  - Advanced Tools and Techniques
  - Implementation Strategies and Lessons Learned
- The Big Data Analytics and Applications for Defense, Intelligence and Homeland Security Symposium topics are:
  - Government Needs, Initiatives, Opportunities and Challenges
  - Emerging Applications for Defense and Intelligence
  - The Latest Tools, Techniques and Technologies – Data Collection/Discovery, Deep/Predictive Analytics, Cloud, Scalability, Security, etc.

Slide 5 The Connection

http://semanticommunity.info/
Big_Data_Symposia#Gartner_Says_Big_Data_Makes_Organizations_Smarter.2C_But_Open_Data_Makes_Them_Richer

http://semanticommunity.info/
Big_Data_Symposia#Magic_Quadrant_for_Business_Intelligence_and_Analytics_Platforms

The Connection

- The connection between these three conferences is that:
  - Gartner Says Big Data Makes Organizations Smarter, But Open Data Makes Them Richer and
  - The Gartner Magic Quadrant for Business Intelligence and Analytics Platforms shows one how to do that with tools that work well with open data.

Slide 6 Strata 2013 Conference: Broad Data

http://strataconf.com/strata2013/

http://semanticommunity.info/Big_Data_Symposia#Sessions_of_Interest

http://semanticommunity.info/AOL_Government/
Big_Data_is_going_Broad_According_to_Government_Internet_Guru_Jim_Hendler
Strata 2013 Conference: Broad Data

- **Strata 2013: Making Data Work** emphasized the need for smart and adult data, fast and agile methods and tools that could solve difficult problems and provide real business value - just data is not a business model.
- There were **11 sessions of interest**, of particular interest and relevance to government big data, especially one that I had written about previously entitled **Broad Data: What Happens When the Web of Data Becomes Real**, by James Hendler (RPI) which said:
  - Recently we have begun to see the emergence of a new online data challenge—that of the “Broad data” that emerges from millions and millions of raw datasets available on the World Wide Web. For broad data, the new challenges that emerge include Web-scale data search and discovery, rapid and potentially ad hoc integration of datasets, visualization and analysis of only-partially-modeled datasets, and issues relating to the policies for data use, reuse and combination. In this talk, we present the broad data challenge and discuss potential starting points for solutions. We illustrate these approaches using data from a “meta-catalog” of over 1,000,000 open datasets that have been collected from about two hundred governments from around the world.

Slide 7 Big Data is Going Broad According to Government Internet Guru Jim Hendler

Slide 8 World Wide Web Expert Jim Hendler Receives Inaugural Strata "Big Data" Award

http://news.rpi.edu/update.do?artcenterkey=3101
Slide 9 IOGDS: International Open Government Dataset Search

http://logd.tw.rpi.edu/demo/international_dataset_catalog_search

Slide 10 Critique

http://logd.tw.rpi.edu/demo/international_dataset_catalog_search

http://tw.rpi.edu/wiki/Study_Supreme_Court_Decision_Making_with_Linked_Data

http://semanticommunity.info/Data.gov/Study_Supreme_Court_Decision_Making_with_Linked_Data
Critique

- I found the RPI International Data Set Catalog difficult to use and the research paper that explains why linking open data is important difficult to re-create because the raw data was not provided.
- The first problem with data catalogs is that their format is not standardized - they do not have a standard set of descriptive items like a library card catalog for example. Second, they do all contain sufficient metadata (data about the data) to allow one to work with the data without a subject matter expert. Third, they do not always contain links to the actual data and in a format that can be readily used. And fourth, they may not contain a data dictionary.

Slide 11 A Solution: Data Science

A Solution: Data Science

- I have found that the best source of metadata and data for data sets that can be integrated comes from government statistical agencies like the United States (Census Bureau Annual Statistical Abstract), Europe (Eurostat Annual Statistical Yearbook) and Japan (Statistical Agency Annual Statistical Yearbook).
- I have suggested a data science and system of systems approach to this problem and need for data integration as follows:
  - World Catalog - that helps identify the best Individual Catalogs - that help identify the best Value-added Data Sets.
  - This makes it more than just "an IT project that turns the crank on lots of data".

Slide 12 A Solution: Spotfire

http://semanticommunity.info/Big_Data_Symposia#Magic_Quadrant

http://semanticommunity.info/Big_Data_Symposia#Tibco_Spotfire

http://semanticommunity.info/Big_Data_Symposia#Data_Visualization_Design_Using_Shneiderman.E2.80.99s_Mantra_Overview_First.2C_Zoom_and_Filter.2C_Then_Details-on-Demand
A Solution: Spotfire

- This is illustrated in a Spotfire dashboard I created using data sets and visualization of each these three parts of the system.
- Spotfire is a leader in the Gartner Magic Quadrant for Business Intelligence and Analytics Platforms for the reasons given in their report.
- It also shows Data Visualization Design Using Shneiderman’s Mantra: Overview First, Zoom and Filter, Then Details-on-Demand presented at the Strata 2013 Conference.

Slide 13 Gartner Magic Quadrant: Business Intelligence and Analytics Platforms

http://semanticommunity.info/Big_Data_Symposia#Magic_Quadrant

Slide 14 My Process

http://semanticommunity.info/AOL_Government/Building_a_Digital_Government
My Process

• So here is what I did:
  – Started with:
    • Big Data Symposia Content
    • Gartner Article and Magic Quadrant Reports
    • Strata 2013 Conference Highlights
    • Data Catalogs and Research Notes
  – Copied it to MindTouch to make it Digital Government Strategy compliant
  – Made it Big Data and Machine Readable

Slide 15 Results

http://semanticommunity.info/Big_Data_Symposia

http://semanticommunity.info/@api/deki/files/23179/BDS2013.xlsx


http://semanticommunity.info/@api/deki/files/23185/BrandNiemann03052013.pptx

Results

• The results are presented in:
  – A MindTouch Knowledge Base
  – An Excel Spreadsheet
  – A Spotfire Dashboard
  – Tutorial slides in PowerPoint

Slide 16 Big Data Symposia: Knowledge Base in MindTouch

http://semanticommunity.info/Big_Data_Symposia#Government_Big_Data_Symposium
Slide 17 IOGDS: Excel Spreadsheet

http://semanticommunity.info/@api/deki/files/23179/BDS2013.xlsx

IOGDS: Excel Spreadsheet

http://semanticommunity.info/@api/deki/files/23179/BDS2013.xlsx

Slide 18 DataCatalogs.org: Excel Spreadsheet

http://semanticommunity.info/@api/deki/files/23179/BDS2013.xlsx
Slide 19 DataCatalogs.org: Spotfire

https://silverspotfire.tibco.com/us/...taSymposia2013

Slide 20 IOGDS Countries and Catalogs: Spotfire

https://silverspotfire.tibco.com/us/...taSymposia2013
IOGDS Countries and Catalogs: Spotfire

Note: This does not really tell about the data!

https://silverspotfire.tibco.com/us/library#users/bniemann/Public/7BigDataSymposia2013

Slide 21 IOGDS France: Spotfire

https://silverspotfire.tibco.com/us/...taSymposia2013

IOGDS France: Spotfire

Note: Mostly CSV, HTML, and XLS!

Note: 552,885 rows by 20 columns!

https://silverspotfire.tibco.com/us/library#users/bniemann/Public/7BigDataSymposia2013

Slide 22 US Data.gov Catalog: Spotfire

https://silverspotfire.tibco.com/us/...taSymposia2013
US Data.gov Catalog: Spotfire


Slide 23 U.S. Census Bureau/Small Area Health Insurance (SAHIE) Program: Spotfire

https://silverspotfire.tibco.com/us/...taSymposia2013

U.S. Census Bureau/Small Area Health Insurance (SAHIE) Program-Spotfire

Note: Mostly US EPA where I worked!


Note: Percent Uninsured:

Note: 204,295 rows by 19 columns!
Conclusions and Recommendations

- Jim Hendler’s “meta-catalog” of over 1,000,000 open datasets that have been collected from about two hundred governments from around the world cannot be verified.
  - Specifically he says US Data.gov has 441,339 data sets, but the catalog has only 5,999!
- This big data analytics and application show four problems with data catalogs: format is not standardized; insufficient metadata (data about the data) to allow one to work with the data without a subject matter expert; they do not always contain links to the actual data and in a format that can be readily used; and they may not contain a data dictionary.
  - The previous example of the SAHIE Program data set does.
- Bottom Line: All the work with Data Catalogs does not really help with data integration as I have been able to show.
  - Recall Slide 11!

Story: Matrix of First Symposium Presentations and Pilots

My Note: The afternoon speakers were cancelled due to the snow storm.

<table>
<thead>
<tr>
<th>Name</th>
<th>Title and Organization</th>
<th>Presentation Title and Slides</th>
<th>Comments</th>
<th>Data Sets and Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR. MARK LUKER</td>
<td>Associate Director, National Coordination Office of Networking and Information Technology Research and Development</td>
<td>NITRD Perspectives and Initiatives Slides</td>
<td>Excellent presentation that mentioned the 20th Anniversary Workshop, Blue Book with annual budget, and February 22nd Memo by Dr. John Holdren to provide public access to scientific publications and data with an implementation plan and timelines.</td>
<td></td>
</tr>
<tr>
<td>DR. SASTRY PANTULA</td>
<td>Director, Division of Mathematical Sciences, National Science Foundation</td>
<td>NSF Perspectives and Initiatives Related to Big Data Slides</td>
<td>Think what you might want to do with data, consider Theories, Methods, Analyses - not just Tools, and</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Title and Organization</td>
<td>Presentation Notes</td>
<td>Email Address</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-------------------------------------</td>
<td></td>
</tr>
<tr>
<td>DR. SASI K. PILLAY</td>
<td>Chief Technology Officer, Office of the Chief Information Officer, National Aeronautics and Space Agency (NASA)</td>
<td>NASA Perspectives and Initiatives <a href="http://semanticommunity.info/Big_Data_Symposia">Slides</a></td>
<td><a href="mailto:ralph.kahn@nasa.gov">ralph.kahn@nasa.gov</a></td>
<td></td>
</tr>
<tr>
<td>MS. JO STRANG</td>
<td>Associate Administrator, Safety, Federal Railroad Administration, Department of Transportation</td>
<td>Open Gov 2.0 and Safety.Data.Gov <a href="http://semanticommunity.info/Big_Data_Symposia">Slides</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR. MICHAEL SIMCOCK and MR. PAUL REYNOLDS</td>
<td>Chief Data Architect and Senior Information Architect, respectively, Department of Homeland Security (DHS)</td>
<td>DHS Perspectives and Initiatives <a href="http://semanticommunity.info/Big_Data_Symposia">Slides</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR. NANCY GRADY</td>
<td>Technical Fellow, Data Scientist, Homeland and Civilian Solutions, SAIC</td>
<td>Big Data Across the Clouds <a href="http://semanticommunity.info/Big_Data_Symposia">Slides</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR. ASHIT TALUKDER</td>
<td>Chief, Information Access Division, National Institute of Standards and Big Data Challenges, Opportunities: Role, NIST Cloud Computing and Big</td>
<td>I need to see the slides for this <a href="http://semanticommunity.info/Big_Data_Symposia">Slides</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) Dr. Ralph Kahn  
Senior Research Scientist  
NASA Goddard  
Space Flight Center  
Code 613  
Greenbelt, MD  
20771  
Phone: 301-614-6193  
FAX: 301-614-6307  
e-mail: ralph.kahn@nasa.gov
<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Slides</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR. BRUCE WEED and MR. BILL HARTMAN</td>
<td>Program Director, Worldwide Big Data Business Development, and President, respectively</td>
<td>IBM Software Group and TerraEchos Slides</td>
<td>I need to see the slides for this</td>
</tr>
<tr>
<td>MR. SHAWN KINGSBERRY</td>
<td>Chief Information Officer</td>
<td>Recovery Accountability and Transparency Board Slides</td>
<td>Recovery.gov website going away, but infrastructure to work on waste, fraud, and abuse will be used in support of the new legislation - stay tuned!</td>
</tr>
<tr>
<td>MR. TED MALONE</td>
<td>Big Data Architecture Lead</td>
<td>Microsoft Federal Slides</td>
<td>Tim O'Reilly: Data is becoming more important than software. Excel is the most important database and analytics tool. My Note: Spotfire works well on Excel!</td>
</tr>
<tr>
<td>DR. FREDERICA DAREMA</td>
<td>Director, (Member, Senior Executive Service) Mathematics, Information and Life Sciences, Former Senior Science and Technology Advisor, National Science Foundation</td>
<td>Air Force Office of Scientific Research (AFOSR) Slides</td>
<td>I need to see the slides for this</td>
</tr>
<tr>
<td>MR. KEVIN JACKSON</td>
<td>Vice President and General Manager, Cloud Services</td>
<td>NJVC Slides</td>
<td>I am familiar with this from the NCOIC Cloud Computing Pilot for the National Geospatial Intelligence Agency</td>
</tr>
<tr>
<td>MR. TOM PLUNKETT and</td>
<td>Senior Consultant and Director,</td>
<td>Oracle Public Sector Slides</td>
<td>I need to see the slides for this</td>
</tr>
</tbody>
</table>

Technology (NIST) Slides of Measurement, Standards and Interoperability Data Workshop and produced a knowledge base.
<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Role</th>
<th>Organization/Department</th>
<th>Presentation Title</th>
<th>Additional Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR. MARK JOHNSON</td>
<td>Engineered Systems Program, respectively</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPTAIN AARON D. BURCIAGA</td>
<td>Director, Operations Research, Logistics</td>
<td>Headquarters US Marine Corps</td>
<td>Military Operational Perspectives</td>
<td>Impressive presentation on informing senior military leadership about data science</td>
</tr>
<tr>
<td>CARON KOGAN</td>
<td>Strategic Planning Director–Big Data</td>
<td>Lockheed Martin</td>
<td>The Art of Predicting with Big Data</td>
<td>This seemed routine to me</td>
</tr>
<tr>
<td>MR. SCOTT GNAU</td>
<td>President</td>
<td>Teradata Labs</td>
<td>Unify Your (Big) Data Analytic Strategy</td>
<td></td>
</tr>
<tr>
<td>MS. SOPHIE RASEMAN</td>
<td>Director for Smart Disclosure</td>
<td>Department of the Treasury</td>
<td>Smart Disclosure and Data Analytics</td>
<td>She spoke without slides and her bio said that the Department of Treasury recently launched the Finance Data Directory <a href="http://www.treasury.gov/financedata">http://www.treasury.gov/financedata</a> My Note: I have worked with some of these data</td>
</tr>
<tr>
<td>MR. MICHAEL SCHULMAN</td>
<td>Director of Business Development and Marketing</td>
<td>ScaleMP</td>
<td>Virtual SMPs</td>
<td>I heard this at a previous big data conference.</td>
</tr>
<tr>
<td>JEFF BUTLER</td>
<td>Director, Research Databases</td>
<td>IRS, Research, Analysis, and Statistics</td>
<td>Big Data and Analytics at the IRS: Perspectives and Initiatives</td>
<td>Amazing!</td>
</tr>
<tr>
<td>CARON B. KOGAN</td>
<td>Strategic Planning Director</td>
<td>Lockheed Martin</td>
<td>The Process Of Predicting with Big Data: Art and Science Applied</td>
<td>Activity Based Intelligence</td>
</tr>
</tbody>
</table>

Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by [mindtouch](http://www.mindtouch.us)
Title: Why we need huge datasets of space-based Earth observations, examples of what we do with them for studying airborne dust, smoke, and pollution, and how an involved statistician might help out

Abstract: From a human perspective, Earth is a huge planet, and environmental conditions are enormously diverse. Yet we care very much about even small-scale and short-lived phenomena, as they affect climate and determine habitability. As such, satellite-borne instruments that can make frequent, global observations are central to our study of current conditions, and are indispensable for efforts to predict future change. As a window into the nature of massive Earth science data sets, I will use space-based measurements of aerosols: desert dust, volcanic ash, wildfire smoke, and pollution particle context for these measurements, general data set attributes, key questions these data are intended to address, and the need for coupling such observations with climate and air quality numerical models, will be covered. The final aspect of the seminar, how statisticians might help out, will be explored during discussion at the end of the presentation.

My Notes:

PILOTS IN PROCESS
WILL SEE SLIDES IN ABOUT A WEEK
ESPECIALLY INTERESTED IN THOSE FROM THE SPEAKERS THAT WERE NOT PRESENT

Pilot: Before Second Symposium

BJS

Pilot: Based on Second Symposium Presentations

FEMA.gov

Spotfire Dashboard

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

Previous work: Big Data Innovation and Social Media & Web Analytics Innovation, September 13-14, Boston, MA. Wiki Blogs Slides

Include brochures: Yes, February 20th

Provide Tutorial: Federal Big Data Senior Steering Group, Wiki and Wiki. January 24, Ballston, VA Slides, Slides

Gartner Magic Quadrant for Business Intelligence and Analytics Platforms defines Big Data: The ability to find patterns, correlations and insights across multistructured data will become a mainstream requirement as companies try to better innovate and find operational efficiencies across business processes that leverage data. These include capabilities that enable the collection, storage, management, correlation, organization, exploration and analysis of multistructured data.

Thought: I might use NIEM as a tutorial example - https://www.niem.gov/Pages/sitemap.aspx

DataCatalogs.org

Source: http://datacatalogs.org/

DataCatalogs.org aims to be the most comprehensive list of open data catalogs in the world. It is curated by a group of leading open data experts from around the world - including representatives from local, regional and national governments, international organisations such as the World Bank, and numerous NGOs.

The alpha version of DataCatalogs.org was launched at OKCon 2011 in Berlin. We have plenty of useful improvements and features in the pipeline, which will be launched over the coming months.

If you have an idea for something you think we should add, please let us know on the data-catalogs list. If you're interested talking to other people interested in open government data, you can join the open-government list or follow #opendata on Twitter.

Previous:

<table>
<thead>
<tr>
<th>Date (s)</th>
<th>Title</th>
<th>Metadata</th>
<th>Data</th>
<th>Analytics</th>
<th>Slides</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>July 8, 2011</td>
<td>DataCatalogs.org</td>
<td>MindTouch</td>
<td>Excel</td>
<td>Spotfire</td>
<td>PowerPoint (NA)</td>
<td>NA</td>
</tr>
</tbody>
</table>
Data Hub

Source: http://datahub.io/dataset/datacatalogs-org

Scrape all datasets from datacatalogs.org, geocode them, tidy them up a little, and output as a JSON list of dictionaries in a single file. The output file has been uploaded to thedatahub here: http://thedatahub.org/dataset/datacatalogs-org

My Note: No CSV with data

Open Government Data Catalogues

Source: http://wiki.okfn.org/Wg/government#Open_Government_Data_Catalogues

DataCatalogs.org aims to be the most comprehensive list of open data catalogs in the world. It is curated by a group of leading open data experts from around the world – including representatives from local, regional and national governments, international organisations such as the World Bank, and numerous NGOs.

The aim of the Open Government Data Working Group is to support development of open government data catalogues around the world, and ensure different platforms are technically interoperable. Work will include:

- Technical support for setting up instances of CKAN in countries around the world
- Introductory guide to open government data catalogues

Data Catalog Vocabulary (DCAT)

Source: https://dvcs.w3.org/hg/gld/raw-file/...cat/index.html

Excerpt

DCAT is an RDF vocabulary well-suited to representing government data catalogs such as Data.gov and data.gov.uk. DCAT defines three main classes:

- dcat:Catalog represents the catalog
- dcat:Dataset represents a dataset in a catalog
- dcat:Distribution represents an accessible form of a dataset as for example a downloadable file, an RSS feed or a web service that provides the data.

Another important class in DCAT is dcat:CatalogRecord which describes a dataset entry in the catalog. Notice that while dcat:Dataset represents the dataset itself, dcat:CatalogRecord represents the record that describes a dataset in the catalog. The use of the CatalogRecord is considered optional. It is used to capture provenance information about dataset entries in a catalog. If this distinction is not necessary then CatalogRecord can be safely ignored.
Government Linked Data (GLD) Working Group

Source: [http://www.w3.org/2011/gld/wiki/Main_Page](http://www.w3.org/2011/gld/wiki/Main_Page)

Excerpt

*Developing standards which help governments publish their data as effective and usable Linked Data, using Semantic Web technologies.* (See [Charter](http://www.w3.org/2011/gld/wiki/Main_Page))

<table>
<thead>
<tr>
<th>Editor's Draft</th>
<th>Latest Public</th>
<th>Next Public</th>
<th>Tracker</th>
<th>Contact</th>
<th>Status / Next Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>best practices</td>
<td></td>
<td></td>
<td>tracker</td>
<td>Bernadette, Boris</td>
<td></td>
</tr>
<tr>
<td>dcat</td>
<td>2012-04 WD1</td>
<td></td>
<td>tracker</td>
<td>Fadi, John</td>
<td></td>
</tr>
<tr>
<td>dcat UCR</td>
<td></td>
<td></td>
<td></td>
<td>Fadi</td>
<td></td>
</tr>
<tr>
<td>adms</td>
<td></td>
<td></td>
<td></td>
<td>Phil</td>
<td></td>
</tr>
<tr>
<td>data cube</td>
<td>2012-04 WD1</td>
<td></td>
<td>tracker</td>
<td>Richard</td>
<td></td>
</tr>
<tr>
<td>data cube UCR</td>
<td></td>
<td></td>
<td></td>
<td>Benedikt</td>
<td></td>
</tr>
<tr>
<td>people</td>
<td>2012-04 WD1</td>
<td></td>
<td>tracker</td>
<td>MichaelH</td>
<td></td>
</tr>
</tbody>
</table>

wiki version recently updated from feedback since [presentation](http://www.w3.org/2011/gld/wiki/Main_Page) in call from Feb 23 2012.

not expected to progress in this form
**Emails**

My Comment: Statistical Data Books have this and can be used for this

My Summary of the emails below

PK: classification of datasets followed a shared taxonomy.

CK: W3C's DCAT would be a candidate in this case

CR: data.gouv.fr (the french open data portal) we currently use Eurovoc to describe our datasets (I USED THIS)

BD: The Open Data Portal of the EU uses DCAT/DCT and is aligned in general terms to be compatible with ADMS


JE: Lexvo not only defines global IDs (URIs) for language-related objects, but also ensures that these identifiers are dereferenceable and highly interconnected as well as externally linked to a variety of resources on the Web...."

M: what about using Wikidata ?

PA: to link vocabularies or KOS such as lexvo or eurovoc to your datasets, you might want to check one of the document annotation voc

VP: In my opinion, the most appropriate vehicle to "harmonize classification of datasets" would be DCAT.

AGP: I would suggest to follow the W3C Ontology-Lexica Community Group

CI: the Spanish National Catalogue [1] one [2]. It is based on EUROSTAT [3], the World Bank Data Classification [4], the OECD Stats Classification [5]
CI 1: Do we need to agree on common vocabularies for the DCAT property values: definitely yes

CI 2: IMHO EUROVOC is a really good option for the EC Open Data portal, but I am not sure whether it could be the same for the pan-European one.

MA-E: Today, the Spanish Government released a Technical Interoperability Standard on PSI reuse [1]. This document is a collection of guidelines addressed to all the Spanish public bodies involved in Open Data. These guidelines are about publishing using standards, identification through URIs, exposing catalog metadata based on DCAT, etc.

DM: These links should give you plenty to review including the final link which is a mapping between classifications

DM 1: Harmonizing codes - happy to discuss our experiences of getting communities to adopt a single standard.

ES: It would be great to see some demonstrations of using a thesaurus like Eurovoc to provide cross-repository topical views of datasets. If anyone has seen anything like that (using Eurovoc or another thesaurus) I would greatly appreciate a pointer. However it does seem entirely reasonable to me for a data consumer or service operator to say "this is the data I understand, please use controlled vocab lists A, B or C if you want me to understand you.

PA: I don't think there is a single answer. Creating a global "everyone should use this central list of enumerated terms" list is the way forward.

PK 1: If I have understood correctly, each language version of a concept gets its own URI. That should make it harder to do comparisons without loading it all into a triple store.

BH: We'll do our bit to fold in the salient points from this thread to the forthcoming Best Practices for Linked Data document and DCAT Vocab emerging from the Gov't Linked Data WG (antic. May 2013).

PK 2: Yes, thank you, a lot of good feedback and a list of potential candidates for dataset topic annotations. It seems that a mix of a big broad taxonomy in combination with smaller national one has been used in some cases.

BN: Bottom Line: All the work with Data Catalogs does not really help with data integration as I have been able to show!

PA 1: I think you might be hearing from Martin Kaltenböck rather soon, Peter. And someone from Top Quadrant and ...

MK: for sure - this is an easy task - Peter please see: http://poolparty.biz/ Go for a free demo account here: http://poolparty.biz/try-it/demoaccount/And - yes this is not open source but an enterprise terminology tool that is widely used and thereby an easy to use SKOS editor with linked (open) data capability...

Peter Krantz

From: Peter Krantz [mailto:peter@peterkrantz.se]
Sent: Friday, March 01, 2013 4:32 AM
To: euopendata@lists.okfn.org; public-egov-ig
Subject: Classification of open datasets...

Hi!

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by mindtouch
Many countries are developing national portals with metadata about open datasets from the public sector. To make datasets easier to find and to lower the threshold for pan-european (or global) re-use it would be great if classification of datasets followed a shared taxonomy.

There are many candidates that could be used, e.g. Eurovoc [1], NACE [2]. I would be grateful for any pointers if there is work going on to harmonize classification of datasets on a global or European level.

Regards,

Peter Krantz
http://www.peterkrantz.com
@peterkz_swe

[1]: http://eurovoc.europa.eu/ - available as LOD
[2]: http://ec.europa.eu/competition/merger.../nace_all.html

Christos Koumenides

From: koumenides c.l. (clk1v07) [mailto:clk1v07@ecs.soton.ac.uk]
Sent: Friday, March 01, 2013 5:25 AM
To: Peter Krantz; euopendata@lists.okfn.org; public-egov-ig
Subject: RE: Classification of open datasets...

Hi

I suppose W3C’s DCAT would be a candidate in this case. http://www.w3.org/TR/vocab-dcat/

Regards,

Christos

Charles Ruelle

From: Charles RUELLE [mailto:charles.ruelle@gmail.com]
Sent: Friday, March 01, 2013 7:17 AM
To: Peter Krantz
Cc: euopendata@lists.okfn.org; public-egov-ig
Subject: Re: [euopendata] Classification of open datasets...

Hi,

This topic is very interesting.
In France, for data.gouv.fr (the french open data portal) we currently use Eurovoc to describe our datasets.
Do you know who is using Eurovoc ? What are others classifications that are used ?

Best regards,
Hello,

Please allow me to give the TenForce view on this situation. At TenForce we are collaborating very closely with the EC and with other open data initiatives throughout Europe. As a commercial company we follow research activities and apply their results in our projects. With regards to this discussion we are strong supporters of:

- DCAT: [http://www.w3.org/2011/gld/wiki/Data...log_Vocabulary](http://www.w3.org/2011/gld/wiki/Data...log_Vocabulary)

We are convinced that these are excellent vocabularies to facilitate multilingual data exchange and linking. We have been using and will be using these "tools" in our projects for government and industry. From this experience we know that EUROVOC is/will be key in anything the EC does in the area of open data. The Open Data Portal of the EU uses DCAT/DCT and is aligned in general terms to be compatible with ADMS, cf. [http://open-data.europa.eu/open-data/linked-data](http://open-data.europa.eu/open-data/linked-data)

Future activities in the area of linked open data on a European scale will almost certainly involve these vocabularies. Contracts like this [http://epsiplatform.eu/content/ec-pu...en-data-tender](http://epsiplatform.eu/content/ec-pu...en-data-tender) have been attributed and are moving forward.

Best Regards,
Bastiaan Deblieck

John Erickson

From: John Erickson [mailto:olyerickson@gmail.com]
Sent: Friday, March 01, 2013 2:07 PM
To: public-egov-ig
Cc: euopendata@lists.okfn.org
Subject: Re: [euopendata] Classification of open datasets...

In addition to the vocabs that have been mentioned here (esp. W3C DCAT and ADMS), a particularly interesting opportunity is Lexvo.org [http://www.lexvo.org/](http://www.lexvo.org/), which tries to solve the multi-lingual problem using canonical terms in a particularly linked data-friendly way.

From their site: "Lexvo.org brings information about languages, words, characters, and other human language-related entities to the Linked Data Web and Semantic Web. The Linked Data Web [http://linkeddata.org](http://linkeddata.org) is a worldwide initiative..."
to create a Web of Data that exposes the relationships between entities in our world. Lexvo.org adds a new perspective to this Web by exposing how everything in our world is connected in terms of language, e.g. by considering semantic relationships between multilingual labels (like book or New York).

Lexvo not only defines global IDs (URIs) for language-related objects, but also ensures that these identifiers are dereferenceable and highly interconnected as well as externally linked to a variety of resources on the Web...."

John

Mohamed

From: innovimax@gmail.com [mailto:innovimax@gmail.com] On Behalf Of Innovimax W3C
Sent: Friday, March 01, 2013 2:48 PM
To: John Erickson
Cc: public-egov-ig; euopendata@lists.okfn.org
Subject: Re: [euopendata] Classification of open datasets...

Dear all,

And what about using Wikidata? It might be a bit premature, but it's definitely something worth considering at some point

Best regards,

Mohamed

Pierre Andrews

From: Pierre Andrews [mailto:pierre.andrews@gmail.com]
Sent: Friday, March 01, 2013 2:52 PM
To: John Erickson
Cc: public-egov-ig; euopendata@lists.okfn.org
Subject: Re: [euopendata] Classification of open datasets...

Hello,

to link vocabularies or KOS such as lexvo or eurovoc to your datasets, you might want to check one of the document annotation voc:
- annotate, general annotation of resources http://www.w3.org/2000/10/annotation-ns#
- tags2con, which links "tags" to documents and to conceptual their meaning (for instance in lexvo or eurovoc) http://disi.unitn.it/~knowdive/dataset/delicious/
- the tag ontology, which is less "semantic" as it links tags to resources but doesn't provide a simple way to link the tag strings to concepts: http://www.holygoat.co.uk/projects/tags/

Pierre
Hello Peter,

> I would be grateful for any pointers if there is work going on to
> harmonize classification of datasets on a global or European level.

In my opinion, the most appropriate vehicle to "harmonize classification of datasets" would be DCAT.

This does not mean that the practical problem you pose here has been solved.

The vocs/thesaurus already mentioned by Bastian and others are all good possible candidates to be taken into account:

- DCAT is a vocabulary for describing data-catalogs AND datasets. DCAT provides a good description of a dataset and can help cross-querying (e.g. based on licensing type, spatial coverage, etc). The DCAT Theme/Category property is a placeholder where you could use controlled vocabularies to better describe your datasets content. The question remains: what to use there. An agreement is indeed needed.

- ADMS is a vocabulary to describe other vocabularies and schemata (simplified definition). As such it could also be used for describing special types of government datasets (e.g. all the codelists, XML schemata, and metadata standards a country has adopted). Very useful this may be for cross-querying catalogs for this type of datasets, but it cannot be used as a global solution to the problem you refer here.

- EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU, the European Parliament in particular. However, practically its coverage is broader and can be used for describing all governmental functions. It is really multi-lingual, quite detailed and mature work.

- NACE is the "Statistical Classification of Economic Activities in the European Community". No need for explanation.

- COFOG is a Classification of the Functions of Government. Again no need for explanation.

Some questions, with possible answers:

A) Do we want to come up with single-dimensional classification or we would like to develop a multi-facet one for describing datasets? I assume a multi-facet one would be more useful.

B) Is DCAT enough for identifying these facets? I think yes.

C) Do we need to agree on common vocabularies for the DCAT property values: definitely yes
D) Have we done anything for this? Not yet, but there is discussion already...
D) Are Eurovoc, NACE, COFOG good candidates to be used for this purpose? Most, possibly yes

Regards,
Vassilios Peristeras

Asuncion Gomez Perez

From: Asunción Gómez Pérez [mailto:asun@fi.upm.es]
Sent: Sunday, March 03, 2013 4:55 AM
To: public-egov-ig@w3.org
Subject: Re: [euopendata] Classification of open datasets...

Dear all

Expressive vocabularies/models have been introduced that allow describing morpho-syntactic properties of the linguistic realization in various languages as well as their specific relation, e.g. stating whether one is an abbreviation of another, whether two lexicalizations in different languages are literal or more lenient translations of each other etc. For this purpose, the LIR model [1] allows representing relations between lexicalizations of classes and properties in different languages, thus adding a cultural and linguistic layer to ontologies that is crucial for applications requiring ontology-based access across languages. In the MONNET project [2], LIR has been unified with LexInfo [3] to produce a more comprehensible model for the enrichment of ontologies with a linguistic layer called lemon [4].

I would suggest to follow the W3C Ontology-Lexica Community Group ...

Kind regards
Asun


Carlos Iglesias

From: carlos.iglesias.moro@gmail.com [mailto:carlos.iglesias.moro@gmail.com] On Behalf Of Carlos Iglesias
Sent: Sunday, March 03, 2013 2:15 PM
To: Peter Krantz
Cc: euopendata@lists.okfn.org; public-egov-ig; charles.ruelle@gmail.com
Subject: Re: Classification of open datasets...

Hi everybody,
Getting back to the original and interesting question from Peter, who was asking for taxonomies and not vocabularies to represent or point to them, I was involved in the development of the **Spanish National Catalogue** [1] one [2]. It is based on **EUROSTAT** [3], the **World Bank Data Classification** [4], the **OECD Stats Classification** [5] and several national references, such as the proposal at the 11/2007 Law, the 060 [6] Information Service Classification, The Spanish National Statistics Institute [7] and The EUGO Network national classification [8].

We found that EUROVOC was also a very good option, but maybe too oriented to the European Parliament activity. The hard decision was also how to balance between something more European compatible and something more national specific and complete. I Would love also to know about what is being used in other initiatives, such as the aforementioned case in France, what drove their decisions and what the challenges were.

Regards,
Carlos Iglesias, Open Data Independent Consultant.
carlosiglesias.es
@carlosiglesias

[2] - [http://datos.gob.es/datos/sites/defa...12_tax_02.pdf](http://datos.gob.es/datos/sites/defa...12_tax_02.pdf)
[5] - [http://www.oecd.org/document/31/0,37...1_1_1,00.html](http://www.oecd.org/document/31/0,37...1_1_1,00.html)
[8] - [http://www.eugo.es/POVUDS_web/appman...usqDirecto=SI](http://www.eugo.es/POVUDS_web/appman...usqDirecto=SI)

Carlos Iglesias 1

Hello Vassilios,
On 1 March 2013 21:22, Vassilios Peristeras <vassilios44@gmail.com> wrote:
[...]
C) **Do we need to agree on common vocabularies for the DCAT property values: definitely yes**
D) Have we done anything for this? Not yet, but there is discussion already...

Any pointers to the relevant discussions?

Regards,
Carlos Iglesias, Open Data Independent Consultant.
carlosiglesias.es
@carlosiglesias

Carlos Iglesias 2

From: carlos.iglesias.moro@gmail.com [mailto:carlos.iglesias.moro@gmail.com] On Behalf Of Carlos Iglesias
Sent: Sunday, March 03, 2013 2:32 PM
To: Bastiaan Deblieck
cCc: Charles RUELLE; Peter Krantz; euopendata@lists.okfn.org; public-egov-ig
Subject: Re: [euopendata] Classification of open datasets...

Hello Bastiaan and all,

I would also say that EUROVOC is the obvious option for the Commission’s own data resources and those of other European institutions, but I am not sure about its suitability for public sector bodies from the different Member States, as we may be loosing national specificity and peculiarities, not to mention if you would like also to be compatible beyond the European borders in the future.

IMHO EUROVOC is a really good option for the EC Open Data portal, but I am not sure whether it could be the same for the pan-European one. At least I think this is an issue that it is worth some analysis, and also count on the different national perspectives, opinions and expectations.

 Regards,
Carlos Iglesias, Independent Open Data Consultant.
carlosiglesias.es
@carlosiglesias

Martin Alvarez-Espinar

From: Martin Alvarez-Espinar [mailto:martin.alvarez@fundacionctic.org]
Sent: Monday, March 04, 2013 8:08 AM
To: Carlos Iglesias; Bastiaan Deblieck; Charles RUELLE; Peter Krantz
Cc: EU Open Data Working Group; public-egov-ig
Subject: Re: [euopendata] Classification of open datasets...

Hi all,

Today, the Spanish Government released a Technical Interoperability Standard on PSI reuse [1]. This document is a collection of guidelines addressed to all the Spanish public bodies involved in Open Data.

These guidelines are about publishing using standards, identification through URIs, exposing catalog metadata based on DCAT, etc. Apart from that, there is a concept scheme (which Carlos Iglesias mentioned [2]) with basic themes to be used in the classification of datasets.

Concepts are described in RDF and they need references with other resources and translations for their labels. Also, the HTML redirection doesn’t work properly yet.

See the 22-concept taxonomy translated into English at [3].

Best,

Martin
Hi, These links should give you plenty to review including the final link which is a mapping between classifications

Single classifications


COFOG: http://www.listpoint.co.uk/CodeList/details/United%20Nations%20COFOG%20Codes%20-%20Functions%20of%20Government/1.0

Classification Context

- group of UK Government and International Classifications
  http://www.listpoint.co.uk/Context/details/Classifications%20of%20Government/1.0

- group of International Codings for Industrial Classifications
  http://www.listpoint.co.uk/Context/details/International%20Coding%20Systems%20for%20Industrial%20Classification/1.1

Classification Mapping

- direct mapping between competing standards
  http://www.listpoint.co.uk/ContextMap/details/Mappings%20between%20Industrial%20Classifications/0.1

Regards David

David Milton 1

From: Milton, David [mailto:DavidMitton@liberata.com]
Sent: Monday, March 04, 2013 3:43 AM
To: 'Peter Krantz'; euopendata@lists.okfn.org; public-egov-ig
Subject: RE: [euopendata] Classification of open datasets...
...and one other:

Common Procurement Vocabulary (CPV Codes)


Harmonizing codes - happy to discuss our experiences of getting communities to adopt a single standard. E-mail me off-line or through this forum as you wish.

Regards David

Ed Summers

From: ed.summers@gmail.com [mailto:ed.summers@gmail.com] On Behalf Of Ed Summers
Sent: Monday, March 04, 2013 10:17 AM
To: public-egov-ig
Subject: Re: [euopendata] Classification of open datasets...

Maybe this was already covered in this thread, but DCAT and Eurovoc are complementary, not competing technologies. For example you could use dcat:theme [1] to reference a SKOS concept in Eurovoc. I was going to give an example, but it appears that you can't download the SKOS XML without being a registered user (which I couldn't figure out how to do). Also, it looks like the licensing around Eurovoc doesn't appear to be particularly "open" [2].

It would be great to see some demonstrations of using a thesaurus like Eurovoc to provide cross-repository topical views of datasets. If anyone has seen anything like that (using Eurovoc or another thesaurus) I would greatly appreciate a pointer.

//Ed


Phil Archer

From: Phil Archer [mailto:phila@w3.org]
Sent: Monday, March 04, 2013 11:33 AM
To: Peter Krantz
Cc: euopendata@lists.okfn.org; public-egov-ig
Subject: Re: Classification of open datasets...

Hi Peter,

You've kicked off a lot of discussion - did you get a satisfactory answer?
I've been thinking about a closely related topic recently - what controlled vocabularies do people find most useful and how can we find out? In Europe, in some circles, we think of Eurovoc, or AgroVoc - which are fine but may be seen as very Euro-centric. Owen Ambur points (not unnaturally) to what he sees in the StratML world which, so far, is largely US-centric.

The NACE codes - that describe company activity - are based on the UN's ISIC codes and it all gets turned into a country-specific set known as SIC codes here in UK. What on Earth is a data publisher to do?

I don't think there is a single answer. Creating a global "everyone should use this central list of enumerated terms" list is the way forward. "However" it does seem entirely reasonable to me for a data consumer or service operator to say "this is the data I understand, please use controlled vocab lists A, B or C if you want me to understand you." And, in similar vain maybe, something like: "you're free to use any of skos:prefLabel, rdfs:label and dcterms:title but in *my* application I treat them all the same."

WDYT?

Phil.

Peter Krantz 1

From: Peter Krantz [mailto:peter@peterkrantz.se]
Sent: Monday, March 04, 2013 12:53 PM
To: Ed Summers
Cc: public-egov-ig
Subject: Re: [euopendata] Classification of open datasets...

2013/3/4 Ed Summers <ehs@pobox.com>:
> I was going
> to give an example, but it appears that you can't download the SKOS
> XML without being a registered user (which I couldn't figure out how
> to do). Also, it looks like the licensing around Eurovoc doesn't
> appear to be particularly "open" [2].

For unknown reasons they have made it difficult to download the machine readable version. You need a "free of charge licence agreement" which is signed on paper, scanned and emailed to you if you ask for a license. Not entirely sure about what they want to prevent based on the license terms [1].

> It would be great to see some demonstrations of using a thesaurus like
> Eurovoc to provide cross-repository topical views of datasets. If
> anyone has seen anything like that (using Eurovoc or another
> thesaurus) I would greatly appreciate a pointer.
>
I have implemented DCAT over Atom on opengov.se and tried mapping some keywords to Eurovoc (see an example at [2]). This feed is syndicated to publicdata.eu (or was earlier at least).

I have a recent version of Eurovoc in SKOS [3] if you want to take a look (but please read the text in [4] before clicking the link:-). If I have understood correctly, each language version of a concept gets its own URI. That should make it harder to do comparisons without loading it all into a triple store.

Regards,

Peter

[1]: https://www.evernote.com/shard/s14/s...663e0edba458a7
[2]: http://www.opengov.se/data/115/rdf/
[3]: https://dl.dropbox.com/u/2372866/eur...rov_skos.zip
Responsibility for the reproduction and adaptation lies entirely with Peter Krantz.

Bernadette Hyland

From: Bernadette Hyland [mailto:bhyland@3roundstones.com]
Sent: Monday, March 04, 2013 1:24 PM
To: Peter Krantz; Martin Kaltenböck; Phil Archer
Cc: Fadi Maali; John Erickson; W3C public GLD WG WG; egov-ig mailing list; euopendata@lists.okfn.org
Subject: Re: Classification of open datasets...

Hi Peter,

Thank you for kicking off a thread initially on the e-gov IG and EU Open Data lists. I've broadened to include the public W3C Government Linked Data working group because we're interested stakeholders. I hope this helps ...

Today, I pinged several of the editors of the DCAT vocabulary that is on track as a W3C Recommendation document. Immediately, several responses from working group members (both in Europe) independently shared a perspective that is held by many linked data advocates:

1) There is no one ring to rule them all -- there is no one vocabulary to describe gov't data sets globally. That is a feature, not a bug.

2) Data harmonization is hard work but worth doing -- I've learned not to trivialize the effort in stitching together data sets that have been published as linked data (4 star linked data); however

3) The task of linking data sets together (from which we derive the semantic goodness linked data advocates proclaim), comes from using several core 'interlinking' vocabularies, for example SKOS, RDFS, Dublin Core, (not an exhaustive list!!) to yield 5 star linked data -- this is the high octane fuel for global innovation & discovery.

4) Whether you're publishing beautiful RDF for all your public data sets (meaning you've taken the big plunge), or getting started with your toe in the waters by publishing RDFa 1.1 Lite on your site, you've succeeded in making your data more
accessible to search engines and lowered the barrier for participation for everyone -- what is not to love?! This is what you wanted in the first place, right?

Peter -- what is very cool about your question IMO is that in one business day, open data advocates, entrepreneurs and public sector employees from all over the EU & US put forward projects we're involved in.

While you didn't get a 'multiple choice' style answer, namely use Eurovoc | DCAT | ADMS, etc., hopefully you have confidence that we're beyond the tipping point publishing open government data and in many cases as 4 and 5 star linked data. Not all data need be (or will be!!) published as linked data, but for data sets destined for access & re-use, while keeping it simple, keep in mind some basic principles (1-4 above).

We'll do our bit to fold in the salient points from this thread to the forthcoming Best Practices for Linked Data document and DCAT Vocab emerging from the Gov't Linked Data WG (antic. May 2013).

Cheers,

Bernadette Hyland, co-chair
W3C Government Linked Data Working Group
Charter: http://www.w3.org/2011/gld/

Peter Krantz 2

From: Peter Krantz [mailto:peter@peterkrantz.se]
Sent: Monday, March 04, 2013 3:14 PM
To: Phil Archer
Cc: euopendata@lists.okfn.org; public-egov-ig
Subject: Re: Classification of open datasets...

2013/3/4 Phil Archer <phila@w3.org>:

> You've kicked off a lot of discussion - did you get a satisfactory answer?

Yes, thank you, a lot of good feedback and a list of potential candidates for dataset topic annotations. It seems that a mix of a big broad taxonomy in combination with smaller national one has been used in some cases.

When looking at use cases it is also important to make it easy to classify datasets. Having a huge taxonomy to choose from may make it difficult to use. Limiting the choice to top levels may make it easier. If all we want to do is to facilitate discovery of similar datasets that may be good enough.

On a side note there seems to be a lack of simple tools to create your own material in SKOS. I have only found a handful that seems usable for non-tech people (e.g. iQvoc).

> The NACE codes - that describe company activity - are based on the
UN’s ISIC codes and it all gets turned into a country-specific set

known as SIC codes here in UK. What on Earth is a data publisher to do?

Yes, NACE looks pretty simple to use and probably covers a lot of topics. The relation could be <dataset> ---created from the activity--- <nace code>.

I don’t think there is a single answer. Creating a global "everyone should use this central list of enumerated terms" list is the way forward.

*However* it does seem entirely reasonable to me for a data consumer or service operator to say "this is the data I understand, please use controlled vocab lists A, B or C if you want me to understand you."

And, in similar vain maybe, something like: "you’re free to use any of skos:prefLabel, rdfs:label and dcterms:title but in *my* application I treat them all the same."

An idea could be to use a subset of your national wikipedia data for topics. Wikipedia is managed and wikipedia articles in one language are linked to their counterpart in other languages. They also have categories which facilitate (some) hierarchy. Using wikipedia (DBPedia) links it would be fairly easy to follow relations to other datasets.

Regards,

Peter Krantz

http://www.peterkrantz.com

@peterkz_swe

**Brand Niemann**

From: Brand Niemann [mailto:bniemann@cox.net]
Sent: Monday, March 04, 2013 3:35 PM
To: 'Bernadette Hyland'; 'Peter Krantz'; 'Martin Kaltenböck'; 'Phil Archer'
Cc: 'Fadi Maali'; 'John Erickson'; 'W3C public GLD WG WG'; 'egov-ig mailing list'; euopendata@lists.okfn.org
Subject: RE: Classification of open datasets...

I am working on this as well for an upcoming Big Data Symposium:
I also presented some work on this about two years ago in my keynote for the SEMIC.EU Conference:

Bottom Line: All the work with Data Catalogs does not really help with data integration as I have been able to show!

Dr. Brand Niemann
Director and Senior Data Scientist
Semantic Community

Phil Archer 1

From: Phil Archer [mailto:phila@w3.org]
Sent: Monday, March 04, 2013 3:49 PM
To: Peter Krantz
Cc: euopendata@lists.okfn.org; public-egov-ig; Martin Kaltenböck
Subject: Re: Classification of open datasets...

On 04/03/2013 20:14, Peter Krantz wrote:

[.]

> On a side note there seems to be a lack of simple tools to create your
> own material in SKOS. I have only found a handful that seems usable
> for non-tech people (e.g. iQvoc).

I think you might be hearing from Martin Kaltenböck rather soon, Peter.

And someone from Top Quadrant and ...

--

Phil Archer

1 From: Phil Archer [mailto:phila@w3.org] Sent: Monday, March 04, 2013 3:49 PM To: Peter Krantz Cc: euopendata@lists.okfn.org; public-egov-ig; Martin Kaltenböck Subject: Re: Classification of open datasets...
Hi Peter, all

for sure - this is an easy task - Peter please see: [http://poolparty.biz/](http://poolparty.biz/) Go for a free demo account here: [http://poolparty.biz/try-it/demoaccount/](http://poolparty.biz/try-it/demoaccount/)

And - yes this is not open source but an enterprise terminology tool that is widely used and thereby an easy to use SKOS editor with linked (open) data capability...

More infos 'off the list' if you are interested to avoid marketing spam here - best Martin

---

**ACT-IAC Big Data Symposium**

Source: [http://www.actgov.org/sigcom/SIGs/SI...default%2Easpx](http://www.actgov.org/sigcom/SIGs/SI...default%2Easpx)

ET SIG Big Data Committee Government-only Symposium

Date: Friday, March 15, 2013
Time: 8:00am - 12:30pm
Cost: Free Government-Only Event
Venue: GSA Building at 1275 1st St, NE, Washington, DC 20002 (near Union Station and New York Ave Metro)
Register: email jenny.knox@k3-solutions.com

**Questions**

Does your agency have diverse data that has been left untapped?
Do you want to enhance your mission by using this data to make better decisions or gain efficiencies in these tough budget times?

Are you making a business case for a big data pilot, but having trouble getting it off the ground or finding the right skill sets?

If the answer to any of these questions is yes, then we cordially invite you to join the ACT-IAC Emerging Technology SIG Big Data Committee at its first Big Data Symposium on Friday, March 15, 2013. My Note: We just had a Big Data Symposium this week that I did the knowledge capture for to help you.

The term Big Data has become the new buzz word in Washington, but what does it really mean for you? Now that many government leaders are aware of the potential value big data and analytics can bring, it is timely for the ACT-IAC ET SIG Big Data Committee to help move the discussion beyond the hype to discuss real-world challenges faced by practitioners from federal agencies.

How is this Big Data Event different from all the rest? My Note: It is only a half-day which is too short.

There seems to be another big data event every week, so how is this one any different? My Note: There were three this week, but they were affected by the snowstorm.

1) Government-only: This event is targeted to a government-only audience so that participants can share their challenges, ask questions of their colleagues, and collaborate in an open ‘safe harbor’ forum. My Note: It is better to have a diverse audience.

2) Interactive Discussion: This roundtable event focuses on interactive dialogue between government participants to dig into real challenges and use cases rather than an audience listening to panelists. My Note: It is good to have presentations by experts and discussions about them.

Objectives

During the event, participants will fulfill the following objectives:

Agency Cross Pollination: Data volume and sources vary across agencies which offer opportunities to develop strategies around large and small data sources. Through government-to-government discussion of solutions and key challenges, participants can take use cases, lessons learned and practical information away from the symposium to their agency.

Sharing of Best Practices: The successful cases presented will bring about lively discussion and preliminary opportunities in developing big data strategies and analytics for specific agencies. Understanding the solutions to challenges that other agencies face will enable attendees to have a better understanding of how they can improve or develop their big data analytics capability.

Governmental Awareness: Disseminating the information regarding analysis of large and complex data is valuable to the larger community regarding big data management. Through the development of public sector standards and practices, the practitioners can implement the methodologies within their agency.
We hope that you will use this forum and the ACT-IAC Big Data Committee as an ongoing resource for information sharing, concept incubation, and challenge discussion. My Note: I have built the ET SIG Big Data Committee Knowledge Base for that purpose.

Symposium Overview

This half-day symposium will feature roundtables led by government big data leaders to discuss best practices and solutions to the unique challenges of implementing data and analytics strategies within federal agencies. Government track leads from DHS, DOJ, and HHS will lead discussions around pressing challenges regarding making the case within your agency, aligning your workforce to meet your skills requirements, enabling data security and assurance as well as developing a business intelligence strategy for your agency. My Note: I have not heard any of these “government big data leaders” listed below speak at the big data meetings I have attended. Are they members of the Federal Big Data Senior Steering Work Group from the 19 member agencies?

The roundtables will focus on the following key challenges to implementing big data strategies:

My Note: See my recent presentation to the Federal Big Data Senior Steering Group on Government Challenges with Big Data (Slides) and to the ACT-IAC C&T SIG - on Activity Based Intelligence, Open Data APIs and Apps, and Web Site Modernization (Slides).

Making the Business Case for Big Data Analytics / Skill Sets / Training

Government Track Lead: Susan Stolting, Office of the CIO, DHS ICE

- How do I convince my management about big data technology and committing to a big data strategy?
- How do I engage executive endorsement and keep the championship alive as the program matures?
- How do I make this work with my constrained budget resources in the coming years? How do I get the right people to help me given the current budget environment?
- What does the future of analytics and big data promise for expansion and potential sharing of services between government agencies or partners?

My Note: See NIEM

Managing and Securing Large and Complex Data

Government Track Lead: Steven Hernandez, CISO/Director, Information Assurance, HHS OIG

- How do I start a data management group or governance mechanism within my agency to tackle this big data problem?
- How do I secure all of this data in the most effective way? And is there a way I can have a cost saving solution at the same time (i.e. cloud services)?
What are the expected cost savings with big data efforts?
What are the risk considerations for ingesting and using data?

My Note: See Health.Data.gov

Developing a Data Analytics Capability (Tools, Pilot Projects, Resources)

Government Track Lead: Jim Burch, Deputy Assistant Attorney General, DOJ

- How do I navigate the landscape of vendors providing tools and services?
- How do I determine the best mix of staffing resources (contractor and federal staff) for developing a data analytics capacity?
- How can I get started with a small pilot project in my office to gain buy-in?
- What lessons learned do I need to be aware of when selecting tools and analytics services (i.e. shared or dedicated) for my agency’s capability suite?

My Note: See FedStats.net. Also started to work on Bureau of Justice Statistics big data pilot for Ms. Jo Strang, Associate Administrator, Safety, Federal Railroad Administration, Department of Transportation for “Open Gov 2.0 and Safety.Data.Gov” from the Big Data Symposium this week.

Agenda

8:00-8:30 AM - Registration and Networking

8:30-8:45 AM - Introduction & Symposium Goals, Johan Bos-Beijer, BDC Government Co-Chair & Topics and Moderators, Mile Corrigan BDC Industry Co-Chair

8:45-9:45 AM - Discuss Topic #1 – Making the Case for Big Data Analytics & Finding the Right Skill Sets

9:45-10:45 AM - Discuss Topic #2 – Securing and Managing Big Data

10:45-11:00 AM - Break

11:00 AM-12:00 PM - Discuss Topic #3 – Developing a Data Analytics Capability (Tools, Techniques, Pilot Projects)

12:00-12:25 PM - Next steps: Define/Capture Follow-Up Action Items for white paper, future symposium topics for the Big Data Committee and SIGs

12:25-12:30 PM - Wrap-up and closing remarks, BDC Chairs

Big Data Committee Leadership

Government Co-Chair Johan Bos-Beijer, GSA
Industry Co-Chairperson Mile Corrigan, Noblis

Symposium Coordinator Chair Isaiah Goodall, Elder Research, Inc.

My Note: Knowledge Capture Chair Brand Niemann, Semantic Community

For More Information

Please contact Johan Bos-Beijer, Director and Senior Advisor at GSA and government co-chair of the ACT-IAC Big Data Committee for more information. You can reach Johan via email at johan.bos-beijer@gsa.gov.

Registration:

Registration is provided at no cost and is limited to government employees only. Please email Jenny Knox at jenny.knox@k3-solutions.com to register for the March 15 Big Data Symposium.

Strata Conference 2013: Making Data Work

Source: http://strataconf.com/strata2013/

February 26-28, 2013, Santa Clara, California

Tap into the collective intelligence of the best minds in data

If you care about big data, join its best and brightest this February at Strata 2013

• Get face-to-face with the leading minds in big data (and thousands of your peers).
• Master the skills and technologies that make data work—with three days of inspiring keynotes, practical tutorials and sessions, an expo hall with key players and products, and plenty of networking events.
• Discover how some of the world’s most successful companies (like Google and Microsoft) tackle and use big data—and how you can apply what they’ve learned.

An amazing event with the latest and greatest of the big data community dropping knowledge and taking names.
—Strata attendee

The future belongs to those who understand how to collect and use their data successfully. And that future happens at Strata.

The breadth and depth of expertise at Strata is unsurpassed—with over 120 speakers and 100 presentations and events, you’ll find solutions to your most pressing data issues. The conference program covers strategy, technology, and policy:

• Data-driven Business: Solve some of today’s thorniest business problems with big data, new interfaces, and the advent of ubiquitous computing.
• Big Data for Enterprise IT: Create big data strategy, manage your first project, demystify vendor solutions, and understand how big data differs from BI.
• Beyond Hadoop: Dive deep into Cassandra, Storm, Drill, and other emerging technologies.
• **Connected World**: Explore the implications—and opportunities—as low-cost networks and sensors create an ever-connected world.

• **Data Science**: Immerse yourself inside the world of data practitioners—from the hard science of new algorithms to cultural change and teambuilding.

• **Design**: Make data matter with highly effective user experiences, using new interfaces, interactivity, and visualization.

• **Hadoop in Practice**: Get practical lessons, integration tricks, and a glimpse of the road ahead.

• **Law, Ethics, and Open Data**: Tackle the biggest issues in compliance, governance, and ethics in the era of open data and heightened privacy concerns.


Presentation slides will be made available after the session has concluded and the speaker has given us the files. Check back if you don't see the file you're looking for—it might be available later! (However, please note some speakers choose not to share their presentations.)

---

**Sessions of Interest**

**Data Visualization Design Using Shneiderman’s Mantra: Overview First, Zoom and Filter, Then Details-on-Demand**

Source: [http://strataconf.com/strata2013/public/e/detail/27169](http://strataconf.com/strata2013/public/e/detail/27169)

Eric Legrand (Wells Fargo), Dana Zuber (Wells Fargo)

MY NOTE: Looked at slides and no graphics

This session explores applications of Shneiderman’s mantra for visual data analysis (overview first, zoom and filter, then details-on-demand) as a framework in the context of three complex analytical applications at Wells Fargo: (1) Analytics process, (2) Interactive meeting facilitation and (3) Dashboard design. Read more.

Data visualization is often an after-thought in a world of collecting and analyzing data. And yet, without clear and compelling communication, analysis will never drive insights and action.

The best analyses show us something that we aren’t expecting and when it comes to spotting surprises, “even the best statisticians often set their calculations aside for a while and let their eyes take the lead.” (Stephen Few, 2009)

This session explores applications of Shneiderman’s famous mantra for visual data analysis (overview first, zoom and filter, then details-on-demand) as a framework in the context of three complex analytical applications at Wells Fargo: (1) Analytics process, (2) Interactive meeting facilitation and (3) Dashboard design.

We will explain how an interactive process of filtering and visual representation guides not only the analytical process, but also tells a story in a way that conveys useful and actionable insights to business users. Using specific case study examples from Wells Fargo, we will describe which visualization tactics and tools have driven success and why..

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Broad Data: What Happens When the Web of Data Becomes Real?

Source: http://strataconf.com/strata2013/publication/detail/27609

This presentation will be streamed live.

James Hendler (RPI)

In this talk, we present the broad data challenge and discuss potential starting points for solutions. We illustrate these approaches using data from a "meta-catalog" of over 1,000,000 open datasets that have been collected from about two hundred governments from around the world. Read more.

Recently we have begun to see the emergence of a new online data challenge—that of the “Broad data” that emerges from millions and millions of raw datasets available on the World Wide Web. For broad data the new challenges that emerge include Web-scale data search and discovery, rapid and potentially ad hoc integration of datasets, visualization and analysis of only-partially modeled datasets, and issues relating to the policies for data use, reuse and combination. In this talk, we present the broad data challenge and discuss potential starting points for solutions. We illustrate these approaches using data from a “meta-catalog” of over 1,000,000 open datasets that have been collected from about two hundred governments from around the world.

http://tw.rpi.edu/wiki/James_A._Hendler

Crowdfunded Open Doctor Data

Source: http://strataconf.com/strata2013/publication/detail/27662

Fred Trotter (FredTrotter.com)

At Strata RX, we announced the release of DocGraph, the largest open named social graph data set that we know of. This data set included links between doctor who commonly team together in the Medicare dataset. Since then, we have added tremendous depth to the data by crowdfunding the acquisition of doctor credentialing data. Come learn how healthcare works under the cover. Read more.

The DocGraph referral graph encoded using the National Provider Identifier (NPI).

The core NPI database is already open data, including location and type of doctor. But most of the data is held at the state level by the various medical boards.

So we crowdfunded the purchase of this data, which makes the referral graph far deeper. As a result it is possible to use the graph to analyze how doctor come to trust and work with one another.

We cannot be sure, but we think this is the largest open data set, that uses real names that details social relationships. Of course Facebook et al have much larger social graphs, but they only release a silver at a time. If you are a data scientist interested in graph big data, this is a pretty deep well.

http://www.fredtrotter.com/
Maps Not Lists: Network Graphs for Data Exploration

Amy Heineike (Quid)

The majority of data we consume today are presented in lists, one-dimensional orderings that limit the users ability to understand context or perform strategic analyses. For unstructured data, we need to re-imagine what types of visualisations enable exploration in the way that geographic maps can. Read more.

The majority of data we consume today are presented in lists, for example search results, news feeds, and recommendations. The makers of these lists get to decide what’s at the top, and while that can be useful for quick filtering, it is very limiting for understanding context or for thinking strategically.

Public health officials confronting epidemics and generals planning battles interact with geographic maps to explore all the possible plans of attack. However, for decision makers facing conceptual complexity geography isn’t necessarily the right organizing principle, and so new types of maps are needed to enable them to explore multiple dimensions. This could be a business strategist trying to outmaneuver their competitor, or a technologist understanding how new scientific advances could change their field.

This talk explores the use of network graphs as maps of rich datasets. We’ll look at how unstructured data can be represented as nodes and edges using similarity metrics, how learnings from cartography inform their design and interactivity, and the toolkits available to make your own. In the process we’ll see what these maps can teach us about the evolution of big data itself.

http://quid.com/

A Model Strategy for Data Journalism in a Country Without Open Data

Sandra Crucianelli (International Center for Journalists), Angélica Peralta Ramos (La Nacion Newspaper)

A way to introduce the idea that access to Big Data in many countries – especially Argentina – is still a work in progress and somewhat politicized. Despite that, media like La Nacion Newspaper, are working with developers and experts in Data Viz to address the lack of transparency and accountability. Read more.

Join us for a model strategy for data journalism, managing big data, in a country without open data. We will present 3 Case Study for Argentina: Database transportation subsidies; Analysis of data from health insurance payments and Monitoring of efficiency fund a loan from the World Bank to the Argentine government What were the boundary conditions? No Law on Access to Information No Open Data No Open Government

We will show how we could design a news apps with data scattered in various government websites, repositories, databases, some semi-open, others closed, mostly in PDF. We will also discuss what strategies were used with simple and free tools: Create a Data Team, join to hackathons, evangelization in the newsroom from personal training and by projects and promote a the fist Latinamerican Data Fest, in spanish, to open and data mining.
Finally, we will give a preview of our work and discuss the impact of the Data Team’s work in a country with no open data.

http://www.icfj.org/about/profiles/sandra-crucianelli

Data Science vs. Analytics -- Approaches to Problem Solving

Source: http://strataconf.com/strata2013/pub...e/detail/28626

Nick Kolegraff (Accenture)

Data Science has created quite the movement in the data world, yet confusion between data science and analytics still remain across the enterprise. Rather than approach the subject talking about semantic differences between the two, we will discuss the topics as they relate to solving problems, how businesses are approaching them and what you can start doing with data science. Read more.

Data Science has created quite the movement in the data world, yet confusion between data science and analytics still remain across the enterprise. Rather than approach the subject talking about semantic differences between the two, we will discuss the topics as they relate to solving problems, how businesses are approaching them and what you can start doing with data science.

The BigData Top100 List

Source: http://strataconf.com/strata2013/pub...e/detail/28051

Milind Bhandarkar (Greenplum, A Division of EMC), Chaitan Baru (SDSC/UC San Diego)

We will describe the BigData Top100 List initiative—an new, open, community-based effort for benchmarking big data systems. Read more.

We will describe the BigData Top100 List initiative—an new, open, community-based effort for benchmarking big data systems. The BigData Top100 list will rank big data systems according to a well-defined, audited performance metric. The benchmark also provides an accompanying efficiency metric. With “big data” becoming a major force of innovation across enterprises of all sizes, new platforms for managing big data sets are being announced almost on a weekly basis with increasingly more features. Yet, there is currently a lack of any means of comparability among such platforms. While the performance of traditional database systems is well understood and measured by long-established institutions such as the Transaction Processing Performance Council, there is neither a clear definition of the performance of big data systems nor a generally agreed upon metric for comparing these systems. This session unveils a community-based effort for defining an end-to-end application-layer benchmark for big data applications, with the ability to easily adapt the benchmark specification to evolving challenges in the big data space. We actively seek community input into this process.

http://www.greenplum.com/

http://clds.sdsc.edu/
The Web As The Greatest Dataset Of All Time

Source: http://strataconf.com/strata2013/public/event/detail/27488

Lisa Green (Common Crawl), Greg Lindahl (blekko), Kevin Burton (Spinn3r)

Big data tools made it possible to gain extremely valuable insight from large scale analysis of web data, but until recently few people had access to the data. Now tools like Grep the Web and increased raw access to web data grant anyone the power to do such analysis. This presentation addresses practical applications of web data analysis that you can incorporate into your research or products. Read more.

Using the web to view pages one at a time is like using a personal computer to only store recipes. Just as most people immensely underestimated the full potential of personal computers 30 years ago, most people have yet to recognize the full potential of web data today.

The web is the largest collection of information in human history and growing at a staggering rate – the estimated number of webpages grew from 26 million in 1998 to 1 billion in 2000 and hit the 1 trillion mark back in 2008. Modern big data tools and technological advances lowering compute costs have made it possible to gain extremely valuable insight from large scale analysis of web data, but until recently few people had access to the data. Now tools like Grep the Web, indexing services that provide raw access to web data, and repositories of open data make it possible for almost anyone to extract knowledge from the web that was previously only available to large search engine companies.

This presentation will explain the various tools available and share examples of powerful insights gained from analysis of web data, including results from recent research projects. You will hear firsthand from Blekko CTO Greg Lindahl about their motivation for building the free Grep the Web service, how they built it, and what they have learned from it. Kevin Burton, Spinn3r CEO, co-inventor of RSS, and long time Apache contributor will share his experiences from a decade of analyzing web content. Lisa Green brings her insight into the relationship between data accessibility and innovation. The three speakers will discuss practical applications of web data analysis that you can incorporate into your research or products.

http://commoncrawl.org/
https://blekko.com/
http://spinn3r.com/

Great Debate: Design Matters More Than Math

Source: http://strataconf.com/strata2013/public/event/detail/27970

Alexander Gray (Skytree, Inc.), Monica Rogati (LinkedIn), Julie Steele (O'Reilly Media, Inc.), Douglas van der Molen (ClearStory Data)

The Great Debate series returns to Strata. In this Oxford-style debate, two opposing teams take opposing positions. We
poll the audience, and the teams try to sway opinions. It'll be a fast-paced, sometimes irreverent look at some of the core challenges of putting data to work. Read more.

Math is proof. Given enough data—and today, we have plenty—we can know. “The right information in the right place just changes your life,” said Stewart Brand. But your life won’t change by itself. Bruce Mau defines design as “the human capacity to plan and produce desired outcomes.” Math informs; design compels. Which matters more? A well-designed collection of flawed information—or an opaque, hard-to-parse, but unerringly accurate model? From mobile handsets to social policy, we need both good math and good design. Which is more critical? The Great Debate series returns to Strata. In this Oxford-style debate, two opposing teams take opposing positions. We poll the audience, and the teams try to sway opinions. It'll be a fast-paced, sometimes irreverent look at some of the core challenges of putting data to work.

http://www.skytree corp.com/

http://www.rogati.com/

http://radar.oreilly.com/julies/

http://www.clearstorydata.com/

**Big Data vs The Beltway: The Regulatory Risks to Data-Driven Businesses**

Source: [http://strataconf.com/strata2013/pub...e/detail/28193](http://strataconf.com/strata2013/pub...e/detail/28193)

Kenneth Cukier (The Economist)

As big data makes inroads into all aspects of society, how governments regard the technology will be critical for its success. If the past is a guide, the state will embrace big data for its own uses (both good and ill). It will recognize that its authority is threatened and lash out. Read more.

As big data makes inroads into all aspects of society, how governments regard the technology will be critical for its success. If the past is a guide, the state will embrace big data for its own uses (both good and ill). It will recognize that its authority is threatened and lash out. And government will try to place big data under the yoke of regulatory control. The talk will build on ideas outlined in the new book “Big Data: A Revolution That Will Transform How We Live, Work, and Think” (with Viktor Mayer-Schönberger) to explain where the choke points are—and how to keep big data free from governments’ grip.

http://www.cukier.com/

**The Victory Lab**

Source: [http://strataconf.com/strata2013/pub...e/detail/28555](http://strataconf.com/strata2013/pub...e/detail/28555)

Sasha Issenberg (The Victory Lab)

The Victory Lab presents a secret history of modern American politics, pulling back the curtain on the tactics and strategies used by some of the era's most important figures—including Barack Obama and Mitt Romney—with iconoclastic
insights into human decision-making, marketing and how analytics can put any business on the road to victory. Read more.

The Victory Lab presents a secret history of modern American politics, pulling back the curtain on the tactics and strategies used by some of the era's most important figures-including Barack Obama and Mitt Romney—with iconoclastic insights into human decision-making, marketing and how analytics can put any business on the road to victory. Read more.

Nerds crash the gates of a venerable American institution, shoving aside its so-called wise men and replacing them with a radical new data-driven order. We’ve seen it in sports, and now in The Victory Lab—which Politico has described as “Moneyball for politics”—Sasha Issenberg tells the hidden story of the analytical revolution upending the way political campaigns are run in the 21st century. The Victory Lab follows the renegade academics and maverick operatives rocking the war room and re-engineering a high-stakes industry previously run on little more than gut instinct and outdated assumptions. Armed with insights from behavioral psychology and randomized experiments that treat voters as unwitting guinea pigs—and reams of new individual-level data fed into microtargeting algorithms—the smartest campaigns now believe they know who you will vote for even before you do. The Victory Lab presents a secret history of modern American politics, pulling back the curtain on the tactics and strategies used by some of the era’s most important figures-including Barack Obama and Mitt Romney—with iconoclastic insights into human decision-making, marketing and how analytics can put any business on the road to victory.

http://www.thevictorylab.com/

Terracotta

Are you available for a short call to discuss the link between in-memory computing and Big Data? It would be great to understand your needs and interests relative to Big Data, to discuss BigMemory’s capabilities, and to address your questions.

Following is the link to the webcast, Enabling Real-Time Access to Big Data, we held last week with Aberdeen:
http://aberdeen.reg.meeting-stream.com/accessing_big_data/

This is the link to the corresponding report from Aberdeen:
http://info.terracotta.org/AnalystReports_Download.html

Do you have 30 minutes available on your calendar next week to schedule a brief introductory webinar?

Best regards,

Jim Leonard

Terracotta

O - 866-846-6779

C - 510-378-1995
As you know, government agencies are developing their big data strategy or deploying their big data solutions in an effort to make data based decisions. Tableau Software’s visual analytics helps organizations turn big data into pictures, enabling end users to quickly and easily find the answers to their questions and gain insights regardless or data size. You can find some examples here: [http://www.tableausoftware.com/solutions/government-reporting](http://www.tableausoftware.com/solutions/government-reporting).

I understand that you will be attending Government Big Data in D.C. next week. Sean Brophy from the Tableau team will be at our booth all day next Tuesday and from 8-12 on Wednesday. Would you have a few minutes to stop by and learn more about how Tableau is helping government organizations manage big data? Please let us know so Sean can make sure to keep a lookout for you.

Cheers,

Doreen

Doreen Jarman
PR Manager
Tableau Software
t: 206.633.3400 x5648
e: djarman@tableausoftware.com

---

**Gartner Says Big Data Makes Organizations Smarter, But Open Data Makes Them Richer**

Source: [http://www.gartner.com/newsroom/id/2131215](http://www.gartner.com/newsroom/id/2131215)

Open Data on the Agenda for Gartner Symposium/ITxpo, October 21-25, 2012, Orlando, Florida

Whereas "big data" will make organizations smarter, open data will be far more consequential for increasing revenue and business value in today's highly competitive environments, according to Gartner, Inc.

"Big data is a topic of growing interest for many business and IT leaders, and there is little doubt that it creates business value by enabling organizations to uncover previously unseen patterns and develop sharper insights about their businesses and environments," said David Newman, research vice president at Gartner. "However, for clients seeking competitive advantage through direct interactions with customers, partners and suppliers, open data is the solution. For example, more government agencies are now opening their data to the public Web to improve transparency, and more
commercial organizations are using open data to get closer to customers, share costs with partners and generate revenue by monetizing information assets."

Gartner analysts believe an open data strategy should be a top priority for any organization that uses the Web as a channel for delivering goods and services. Open data strategies support outside-in business practices that generate growth and innovation. Enterprise architects help their organization connect independent open data projects by creating actionable deliverables and information-sharing practices that generate business-focused outcomes for achieving strategic customer growth and retention objectives.

Gartner analysts said that any business that has a data warehouse should consider how it can use data as a strategic asset and revenue generator. Maturing technologies for data quality and data anonymization can help mitigate regulatory restraints and risk factors. Open data APIs provide simple, Web-oriented means for data exchange, and linked data techniques are effective for generating big datasets. When considering the long-term benefits of an open data strategy, organizations should investigate the types of data exchange now emerging where information producers and consumers share data for profit.

Emerging data marketplaces are also places for organizations to open their data — potentially turning their "data into dollars." The challenge is to keep the barriers to entry low to enable participation by different types of business and streamlined processes for adding and vetting data sources. Monetizing data is a technological and operational challenge. If an enterprise's goal is to unlock its data's full revenue potential, it needs to be able to reach all possible data buyers efficiently.

"With tight budgets and continued economic uncertainty, organizations will need leaders who can craft breakthrough strategies that drive growth and innovation," said Mr. Newman. "As change agents, enterprise architects can help their organizations become richer through strategies such as open data."

Although openness is a pervasive and persistent issue in IT, there is very little agreement about exactly what "open" means. According to Gartner analysts, an informal definition of openness is a level playing field where everyone can play a game that can evolve. There is a positive relationship between the openness of information goods (for example, code, data, content and standards) and information services (for example, services that offer information goods, such as the Internet, Wikipedia, OpenStreetMap and GPS) and the size and diversity of the community sharing them. From the viewpoint of enterprise information architects, this is known as the information-sharing network effect: the business value of a data asset increases the more widely and easily it is shared.

Open data APIs are a lightweight approach to data exchange. Their use is now considered a best practice for opening data and functionality to developers and other businesses. Organizations use APIs to generate new sources of revenue, spur innovation, increase transparency and improve brand equity.

"The challenge for organizations is to determine how best to use APIs and how an open data strategy should align with business priorities," Mr. Newman said. "This is where enterprise architects can help. While some internal IT functions may be using APIs to fulfil local or specific application needs, the enterprise architecture process harvests and elevates good works as first-class strategic priorities that create business-focused outcomes. As a strategic enabler, APIs are a powerful means with which to build an ecosystem, and a first step toward monetizing data assets."
Additional information is available in the Gartner report "Open for Business: Learn to Profit by Open Data." The report is available on Gartner's website at http://www.gartner.com/resId=1947015.

Magic Quadrant for Business Intelligence and Analytics Platforms

Source: http://www.gartner.com/technology/re...t=130206&st=sb
5 February 2013 ID:G00239854
Analyst(s): Kurt Schlegel, Rita L. Sallam, Daniel Yuen, Joao Tapadinhas

MY NOTE: I am using TIBCO Spotifre as an example of a vendor product in my stories and tutorial. I participated in this survey and was one of the 1,702 respondents last October 2012.

VIEW SUMMARY

The dominant theme of the market in 2012 was that data discovery became a mainstream BI and analytic architecture. The market also saw increased activity in real time, content and predictive analytics.

Market Definition/Description

This document was revised on 13 February 2013. The document you are viewing is the corrected version. For more information, see the Corrections page on gartner.com.

Gartner changed the name of this Magic Quadrant from "Business Intelligence Platforms" to "Business Intelligence and Analytics Platforms" to emphasize the growing importance of analysis capabilities to the information systems that organizations are now building. Gartner defines the business intelligence (BI) and analytics platform market as a software platform that delivers 15 capabilities across three categories: integration, information delivery and analysis.

Integration

- BI infrastructure: All tools in the platform use the same security, metadata, administration, portal integration, object model and query engine, and should share the same look and feel.
- Metadata management: Tools should leverage the same metadata, and the tools should provide a robust way to search, capture, store, reuse and publish metadata objects, such as dimensions, hierarchies, measures, performance metrics and report layout objects.
- Development tools: The platform should provide a set of programmatic and visual tools, coupled with a software developer's kit for creating analytic applications, integrating them into a business process, and/or embedding them in another application.
- Collaboration: Enables users to share and discuss information and analytic content, and/or to manage hierarchies and metrics via discussion threads, chat and annotations.

Information Delivery

- Reporting: Provides the ability to create formatted and interactive reports, with or without parameters, with highly scalable distribution and scheduling capabilities.
- Dashboards: Includes the ability to publish Web-based or mobile reports with intuitive interactive displays that indicate the state of a performance metric compared with a goal or target value. Increasingly, dashboards are used
to disseminate real-time data from operational applications, or in conjunction with a complex-event processing engine.

- Ad hoc query: Enables users to ask their own questions of the data, without relying on IT to create a report. In particular, the tools must have a robust semantic layer to enable users to navigate available data sources.

- Microsoft Office integration: Sometimes, Microsoft Office (particularly Excel) acts as the reporting or analytics client. In these cases, it is vital that the tool provides integration with Microsoft Office, including support for document and presentation formats, formulas, data "refreshes" and pivot tables. Advanced integration includes cell locking and write-back.

- Search-based BI: Applies a search index to structured and unstructured data sources and maps them into a classification structure of dimensions and measures that users can easily navigate and explore using a search interface.

- Mobile BI: Enables organizations to deliver analytic content to mobile devices in a publishing and/or interactive mode, and takes advantage of the mobile client's location awareness.

**Analysis**

- Online analytical processing (OLAP): Enables users to analyze data with fast query and calculation performance, enabling a style of analysis known as "slicing and dicing." Users are able to navigate multidimensional drill paths. They also have the ability to write back values to a proprietary database for planning and "what if" modeling purposes. This capability could span a variety of data architectures (such as relational or multidimensional) and storage architectures (such as disk-based or in-memory).

- Interactive visualization: Gives users the ability to display numerous aspects of the data more efficiently by using interactive pictures and charts, instead of rows and columns.

- Predictive modeling and data mining: Enables organizations to classify categorical variables, and to estimate continuous variables using mathematical algorithms.

- Scorecards: These take the metrics displayed in a dashboard a step further by applying them to a strategy map that aligns key performance indicators (KPIs) with a strategic objective.

- Prescriptive modeling, simulation and optimization: Supports decision making by enabling organizations to select the correct value of a variable based on a set of constraints for deterministic processes, and by modeling outcomes for stochastic processes.

These capabilities enable organizations to build precise systems of classification and measurement to support decision making and improve performance. BI and analytic platforms enable companies to measure and improve the metrics that matter most to their businesses, such as sales, profits, costs, quality defects, safety incidents, customer satisfaction, on-time delivery and so on. BI and analytic platforms also enable organizations to classify the dimensions of their businesses — such as their customers, products and employees — with more granular precision. With these capabilities, marketers can better understand which customers are most likely to churn. HR managers can better understand which attributes to look for when recruiting top performers. Supply chain managers can better understand which inventory allocation levels will keep costs low without increasing out-of-stock incidents.

The BI and analytics platforms market is broad, covering many different use cases and levels of maturity that span four distinct phases: descriptive, diagnostic, predictive and prescriptive analytics.

The vast majority of applications built with BI and analytics platforms to date could be labeled "descriptive" because critical capabilities, such as reports and dashboards, are used to describe the dimensions and measures of a particular aspect of the business. So, for example, a measure such as on-time delivery could be defined in a well-governed data
model and enable users to report on the goal and actual value for that measure by various dimensions, such as customer segments or time periods.

Increasingly, Gartner sees more organizations building diagnostic analytics that leverage critical capabilities, such as interactive visualization, to enable users to drill more easily into the data to discover new insights. For example, visual patterns uncovered in the data might expose an inconsistent supply chain process that is the root cause of an organization's ability to consistently reach its goal for on-time delivery.

As organizations mature at diagnostic analysis, they become so adept at understanding the root causes in their business processes that they can identify the explanatory variables that predict what the measure will be in a future period. For example, a predictive analytic system could be built to forecast the on-time delivery measure. Solutions can be further evolved to prescriptive analytics as the insights from predictive models are integrated into business processes to take corrective or optimal actions.

Right now, most of the user activity in the BI and analytics platform market is from organizations that are trying to mature from descriptive to diagnostic analytics. The vendors in the market have overwhelmingly concentrated on meeting this user demand. If there were a single market theme in 2012, it would be that data discovery became a mainstream architecture (see "More Choices and Complexity in Selecting Data Discovery Tools" for a description of the data discovery architecture). For years, data discovery vendors — such as QlikTech, Salient Management Company, Tableau Software and Tibco Spotfire — received more positive feedback than vendors offering OLAP cube and semantic-layer-based architectures. In 2012, the market responded:

- MicroStrategy significantly improved Visual Insight.
- SAP launched Visual Intelligence.
- SAS launched Visual Analytics.
- Microsoft bolstered PowerPivot with Power View.
- IBM launched Cognos Insight.
- Oracle acquired Endeca.
- Actuate acquired Quiterian.

This emphasis on data discovery from most of the leaders in the market — which are now promoting tools with business-user-friendly data integration, coupled with embedded storage and computing layers (typically in-memory/columnar) and unfettered drilling — accelerates the trend toward decentralization and user empowerment of BI and analytics, and greatly enables organizations' ability to perform diagnostic analytics.

---

**Magic Quadrant**

**Figure 1.** Magic Quadrant for Business Intelligence and Analytics Platforms

---

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by mindtouch®
Quadrant Descriptions

Leaders

Leaders are vendors that are strong in the breadth and depth of their BI platform capabilities, and can deliver on enterprisewide implementations that support a broad BI strategy. Leaders articulate a business proposition that resonates with buyers, supported by the viability and operational capability to deliver on a global basis. Small vendors such as Tableau, QlikTech and Tibco Spotfire, which may lack strong scores for geographic or vertical strategy, or breadth of capabilities in the offering (product) criterion, are still Leaders due to the strength of their market understanding and marketing strategy. The evidence that they are market leaders comes from the fact that most of the market is trying to imitate the simplicity of their architecture and the ease of use that it provides.

Challengers

Challengers are well-positioned to succeed in the market. However, they may be limited to specific use cases, technical environments or application domains. Their vision may be hampered by a lack of coordinated strategy across the various products in their platform portfolios, or they may lack the marketing effort, sales channel, geographic presence, industry-specific content and awareness offered by the vendors in the Leaders quadrant. In this Magic Quadrant, two
vendors, Birst and LogiXML, were vaulted into the Challengers quadrant this year due to the overwhelmingly positive scores they received on the customer reference survey.

**Visionaries**

Visionaries are vendors that have a strong vision for delivering a BI platform. They are distinguished by the openness and flexibility of their application architectures, and they offer depth of functionality in the areas they address. However, they may have gaps relating to broader functionality requirements. Visionaries are market thought-leaders and innovators. However, they may still have to achieve sufficient scale — or there may be concerns about their ability to grow and provide consistent execution. There were no Visionaries in the market this year, mainly because most of the Niche Players do not provide the breadth of vision across descriptive, diagnostic, predictive and prescriptive analytics. Some vendors had moderate product breadth, but lacked vision on the other criteria, such as vertical or geographic strategy.

**Niche Players**

Niche Players do well in a specific segment of the BI platform market — such as reporting or dashboarding — or have a limited capability to innovate or outperform other vendors in the market. They may focus on a specific domain or aspect of BI, but are likely to lack depth of functionality elsewhere. They also may have gaps relating to broader platform functionality. Alternatively, Niche Players may have a reasonably broad BI platform, but have limited implementation and support capabilities or relatively limited customer bases, such as in a specific geography or industry. In addition, they may not yet have achieved the necessary scale to solidify their market positions.

**Context**

Readers should not use this Magic Quadrant in isolation as a tool for vendor selection. Gartner has defined the BI and analytics market broadly. We are including a variety of products that span a range of buyers and use cases, such as decision management suites, interactive dashboards, and tools that are better for integrated planning. Consider this Magic Quadrant to be more of an executive summary into Gartner's research into this market, and use it in combination with the critical capabilities; customer references survey; strengths, weaknesses, opportunities and threats (SWOTs); and analyst inquiries when making specific tool selection decisions.

**Market Overview**

Although this is a mature market and has been a top CIO priority for years, there is still a lot of unmet demand. Every company has numerous subject areas — such as HR, marketing, social and so on — that have yet to even start with BI and analytics. The descriptive analytics have largely been completed for most large companies in traditional subject areas, such as finance and sales, but there is still a lot of growth expected for diagnostic, predictive and prescriptive deployments. Moreover, many midsize enterprises have yet to even start their BI and analytic initiatives. Gartner’s view is that the market for BI and analytics platforms will remain one of the fastest growing software markets. The compound annual growth rate for the BI and analytics space is expected to be 7% through 2016 (see "Forecast: Enterprise Software Markets, Worldwide, 2011-2016, 4Q12 Update").
In addition, the emerging data-as-a-service trend could significantly grow the market for BI and analytics platforms. Today, the business model is largely "build" driven. Organizations license software capabilities to build analytic applications. However, organizations increasingly will subscribe to industry-specific data services that bundle a narrow set of data with BI and analytic capabilities embedded. This data-as-a-service trend will be driven from well-known, trusted data aggregators, such as Nielsen or Thomson Reuters, as well as industry-specific players, such as IMS (life sciences) or CoreLogic (financial services). In time, most companies — regardless of their business model — will need to provide a data-as-a-service offering (see "Make Customer-Facing Analytics Part of Your Business Model"). Therefore, this data-as-a-service trend has the potential to grow the market significantly as a range of vendors are looking to embed a BI and analytic platform provider's software capabilities into their data-as-a-service offerings.

Tibco Spotfire

Strengths

- Tibco Spotfire is a flexible and easy-to-use platform for business user data discovery and analysis, for authoring analytic applications, for publishing interactive and visual dashboards, and for building predictive models and applications. Tibco Spotfire's interactive visualization capabilities are now enabled by a hybrid, in-memory and newly added direct query access approach to support, and they leverage larger enterprise-managed datasets than previously possible.

- Tibco Spotfire's strong product vision has been and continues to be a key strength. Its focus on advanced and real-time analytic applications and dashboards delivered to mobile devices contributes to its vision. Unlike the other data discovery platforms (for example, QlikView and Tableau), Tibco Spotfire is leveraging the acquisition of Insightful and its newly released runtime engine for R for data mining, as well as its integration with Tibco middleware, Tibco Software's acquisition of LogLogic, and Tibco Software's social capability, tibbr, to broaden the possible spectrum of end-user-driven interactive analysis and data sources, and to incorporate business events and predictive analytics. In particular, in the latest release of Spotfire, Spotfire 5.0, the product features a commercial, integrated runtime engine for R. This engine can run any R model and will be integrated into other Tibco products, such as LogLogic, which is a competitor to Splunk. In addition to statistician-oriented tools, which allow for maximum flexibility in complex model building, new 5.0 capabilities expose a set of code-free modeling tools to analysts. These have the potential to put more advanced analytics in the hands of a broader set of users, and to enable a seamless analytic model development workflow between business analysts and data scientists. This set of capabilities makes Tibco Spotfire well-positioned to take advantage of the increase in market demand for packaged analytic applications and dashboards, which increasingly feature predictive analytics capabilities to make analytics accessible to nontraditional BI users.

- Tibco Spotfire is included in the Leaders quadrant for the first time this year because of its strong product vision combined with increased BI market momentum, which is largely driven by higher levels of investment by Tibco in Spotfire marketing, awareness, and sales and partner channels. Momentum for Spotfire has also been driven by Tibco stack positioning, where Spotfire is now prominently positioned as a critical differentiator for the Tibco stack in support of Tibco big data analytics, which is a key market initiative for Tibco. The combination of these 2012 initiatives has resulted in stronger market awareness and an increase in the shortlisting of Tibco Spotfire outside its traditional installed base of niche users.

- Much like the other data discovery vendors that are addressing increasing market requirements for intuitive, highly interactive BI platforms, Tibco Spotfire's customers are very satisfied with many aspects of the relationship. Customers have a positive view of Tibco Spotfire's future; they report success in terms of expanded usage over the past year and have an above-average view of Spotfire's product quality. The Tibco Spotfire platform also earned above-average performance scores, albeit on smaller-than-average datasets. Over the past year, Tibco Spotfire made investments, which are ongoing, in Tibco Spotfire 5.0 to improve performance on larger and larger datasets with new, direct SQL and MDX query capabilities and an updated, high-performance in-memory data engine.
• Tibco Spotfire has among the highest complexity of user analysis scores of any vendor in the Magic Quadrant, while at the same time customers rate it above average on ease of use, particularly for end users. Tibco Spotfire customers continue to choose it for its ease of use for end users and functionality more often than they do most other vendors, with above-average ratings for achievements of business benefits. Tibco Spotfire customers report the largest percentage of users of any vendor in the Magic Quadrant using the platform for interactive visualization, and among the highest percentage of users conducting ad hoc query and analysis (simple and complex), and deploying predictive analytics, with above-average functional rates in these three capabilities. Because of Tibco Spotfire’s ease of use, more users can leverage the benefits of more advanced analytics. This paradox typifies why data discovery capabilities in general, and Tibco Spotfire in particular, are so compelling for business users, and why they are proliferating.

• Tibco Spotfire’s cloud version of its software allows business users to author and share Tibco Spotfire visualizations and dashboards without having to install the software on-premises. While cloud adoption and intentions are generally low (67% say they will never put their enterprise BI in the cloud) in the survey population at large, Tibco Spotfire has among the highest percentage of customers using (14%) or planning to deploy Spotfire in the cloud in the next 12 months (19%). Spotfire’s cloud success over the past two years appears to be coming from deployments by line-of-business users and departments versus IT, where cloud investments have had lower acceptance.

Cautions

• Even though the average employee size of a company that uses Tibco Spotfire software is the among the highest in the survey, and although Tibco Spotfire has some customer references with large datasets and thousands of users, on average, its deployments tend to be focused on a department or multiple departments in global companies with below-average data volumes and numbers of users, when compared with those of other vendors. Tibco Spotfire also scored among the lowest of all vendors in the reference survey on the percentage of customers that consider it to be their BI platform standard. The combination of this result with Tibco Spotfire’s strong functionality ratings in interactive visualization, ad hoc analysis and predictive analytics suggests that, while it is not usually the enterprise standard, it has been successful in augmenting the BI standard when more flexible discovery-based and sophisticated analysis is required. This complementary position may be under increased threat over time from enterprise vendors that added data discovery features to their platforms in 2012 and are aggressively enhancing them.

• Tibco Spotfire is well-suited to building analytic content ranging from basic interactive visualizations and dashboards to advanced interactive analytic applications, but the perception of Tibco Spotfire’s license cost and packaging continues to be a factor limiting its consideration beyond high-end requirements. As a result, Tibco Spotfire is not included on shortlists as frequently as its primary competitors — QlikTech and Tableau in particular — when basic, mainstream data discovery capabilities are required, even though Spotfire’s awareness has increased substantially over the past year. While a premium for Tibco Spotfire software may be justified, given its differentiating features around collaborative, mobile, advanced and real-time analytics, Tibco Spotfire must overcome its high license cost reputation to capitalize on the buying momentum that’s driving the growth of more mainstream and competitively priced and packaged data discovery alternatives. This will become increasingly important as the stand-alone data discovery vendors, including Tibco Spotfire, face increased competitive pressures, particularly on pricing, from the enterprise BI platform vendors that have now added data discovery capabilities to their platforms, and are often bundling them without additional license costs as features of their platforms. As further evidence of its high license cost reputation, this year, like the past two years, license cost continues to be cited more frequently than most other vendors in the Magic Quadrant survey as a limitation to broader deployment, and its “total license cost per user” continues to be above the survey average. Moreover, while Tibco Spotfire customers are generally happy with most measures of the customer experience — as reflected in its strong position on the Ability to Execute axis of the Magic Quadrant — the one point of dissatisfaction they have expressed is with the sales experiences, which include presales, sales, contracting and pricing.

• While Tibco Spotfire is rated highly in the survey for ad hoc analysis, interactive visualization and predictive analytics, it is rated in the bottom third of vendors for static and parameterized reporting, and has scored below the survey average in areas that are related to enterprise readiness, such as BI infrastructure, metadata management.
and development tools, confirming that its true sweet spot is in providing a flexible, easy-to-use environment for advanced analysis. Much like for the other data discovery vendors, adding enterprise features to support larger data and user adoption requirements to compete against the traditional BI players that have now added data discovery capabilities will be an important competitive requirement in the future.

- Support scores appear to have suffered this year compared with last year, which could be a casualty of high growth. While level-of-support expertise is rated above average, response times and time to resolution scored below the survey average.

---

**Big Data Analytics and Applications for Defense, Intelligence and Homeland Security Symposium**

Source: [http://www.bigdataevent.net/](http://www.bigdataevent.net/)

Subject to Change 2/4/2013 (PDF)

Holiday Inn Rosslyn at Key Bridge, Arlington VA

**Big Data for Defense, Intelligence, and Homeland Security**

Today’s Warfighter has access to an ever-increasing number of sensors, imagers, internet artifacts, open source and other sophisticated collection devices and mechanisms, to the point that a major challenge has become how to sift through this massive amount of information to find the most critical and actionable items of intelligence. Increasingly, this must be accomplished in near-real time and the information must be packaged in a format capable of being shared with all other pertinent parties. The result is that sensor, computer and communication technologies are being strained beyond capacity to keep pace with current and future information management and analysis needs. ‘Big Data’ tools, techniques, and technologies seek to provide the means to analyze, exploit and share conclusions drawn from this seemingly overwhelming information load.

This outstanding symposium brings together the key government and industry experts who are shaping the direction of big data research and development for defense, intelligence and homeland security. What are the latest OSD and Service needs and initiatives in big data? How is big data analytics being applied to ISR, intelligence sharing, GEOINT fusion, video analytics, atmospherics, identity, biometrics, and a whole range of other critical mission applications? What role are new tools, techniques, and technologies – predictive analytics, cloud computing, metadata, etc. – playing in making big data analytics a reality? What are the future challenges and opportunities? What role(s) can industry play? These and many other critical questions will be examined during this outstanding two-day event.

**Day 1: Wednesday, April 24, 2013**

9:00-9:05 Administrative Announcements and Introduction of Moderator

9:05-9:35 MR. JOHN DELAY, Chief Strategic Officer/Architect, Geospatial Solutions, Harris Corporation “Big Data— Not Just About Processing Data with Analytics; It’s About How to Collect, Manage, and Discover Large Volumes of Real Time Data Across Globally Distributed Enterprises” Slides

My Note: Antonisse substituted for Delay who substituted for MR. GUS HUNT, Chief Technology Officer, Central Intelligence Agency Keynote: “Emerging Technical Challenges and Capabilities” Conflict

---

http://semanticommunity.info/Big_Data_Symposia

Updated: Sat, 19 Sep 2015 02:40:40 GMT

Powered by mindtouch
9:35-10:05 DR. ANN CARBONELL, Director, Innovision Information Integration Office, National Geospatial–Intelligence Agency (NGA)  
Keynote: "NGA Big Data R&D– Perspectives & Initiatives" Slides

10:05-10:35 MR. RICH CAMPBELL, Chief Technology Officer, EMC Federal "Big Data Trends– Challenges, Solutions, Use Cases " Slides

10:35-10:50 Coffee Break & Networking Period


11:20-11:50 MS. LISA SHALER-CLARK, Deputy Director/Program Manager, Futures, Intelligence and Security Command (INSCOM), US Army “INSCOM Perspectives and Initiatives” Slides?

11:50-12:20 MR. KEITH JOHNSON, Director, Advanced Analytics, Lockheed Martin “Rethinking Multi-INT Analysis in a Big Data World” Slides

12:20-1:25 Lunch Break

1:25-1:55 MR. CHARLIE FLEISCHMAN, Chief Technology Officer, Boeing Intelligence Systems “Large-Scale Analytics” Slides

1:55-2:25 MR SEAN BROPHY, Senior Analyst, Tableau Software Slides  
My Note: Brophy substituted for DR. WILFRED PINFOLD, Director, Extreme Scale Programs, Intel Labs Slides

2:25-2:55 MR. CHRIS BLOW, Federal CTO, Vice President, Mark Logic Corporations Slides

2:55-3:25 MR. STEVE HAGAN, Vice President, Server Technologies, Oracle Public Sector “Challenges in Big Data Analytics– Applications and Capabilities” Slides

3:25-3:35 Afternoon Refreshments & Networking Period

3:45-4:15 LT. COL. ANDREW O. HALL, Rotational Forces Branch Chief, Joint Operations Division – Global Force Management on the Joint Staff in the Pentagon “Joint Staff Perspectives and Initiatives” Slides

4:15-4:45 MR TOAN DO, Senior Director Federal Programs, MapR Technologies Slides

4:45-5:15 MR DAVID YACHIN, Program Lead, Cyber Security Situational Awareness, HP Autonomy Slides

5:15-5:45 ???

My Note: I do now know what happended to these speakers on the initial agenda:

SENIOR REPRESENTATIVE, Space and Naval Warfare Systems Center, US Navy “SPAWAR Perspectives and Initiatives”

MR. ROBERT McCORMACK, Director of Analytics, Aptima “Social Media Analytics”

MS. PATRICIA GUITARD, SES, Deputy Director, Intelligence (G-2), US Army (tentative) “Big Data and Army Intelligence”
Day 2: Thursday, April 25, 2013

8:30-8:35 Administrative Announcements


9:05-9:35 MR. JASON DALTON, Senior Scientist, Digital Globe “GeoINT and Large Scale Analytics” Slides

My Note: Dalton substitute for MR. CHARLES J. GASSERT JR., Assistant Program Manager, Distributed Common Ground System– Navy (DCGS-N) Opening Address: “Data Analytics and DCGS”

9:35-10:05 DR. ROBERT BONNEAU, Program Director, Very Complex Networks, Air Force Office of Scientific Research (AFOSR) “Large Scale Analytics and Network Management” Slides

10:05-10:20 Coffee Break & Networking Period


11:25-11:50 ROBERT ZITZ, Senior Vice President and ISR Chief Systems Architect & NATHANIEL WIESNER, Director of Big Data, SAIC Slides

11:50-12:20 MR. TIM PAYDOS, Director, Worldwide Business Analytics and Optimization, IBM “Information and the Big Data Phenomenon– If You Only Remember Four Things” Slides

12:20-1:35 Lunch Break

1:35-2:05 MR. TONY JIMENEZ, President and Chief Executive Officer, MicroTech “Social Data – What is it and Why Should You Care?” Slides

2:05-2:35 MR. BRUCE GOLDFEDER, Division Technical Lead, Data Tactics “Open Source, GOTS, and GOSS– Empowering Big Data and Enabling the Intelligence Mission” Slides

2:35-3:05 ????

3:05-3:20 Afternoon Refreshments & Networking Period

3:20-3:50 MR. BRIAN SCHIMPF, Forward Deployed Engineer, Palantir “Human-Driven Data Analytics” Slides

My Note: Choung substituted for Schimpf
I. Government Needs, Initiatives, Opportunities and Challenges

KEYNOTES:

Emerging Technical Challenges and Capabilities
MR. GUS HUNT, Chief Technology Officer, Central Intelligence Agency
- Implications of Big Data and Intelligence
- Approaches to Analytics and Visualization
- Data Sciences and Data Democracy
- The Road Ahead

NGA Big Data R&D — Perspectives & Initiatives
DR. ANN CARBONELL, Director, Innovation Integration Office, National Geospatial-Intelligence Agency (NGA)

Big Data and the Intelligence Mission
MR. GRANT SCHNEIDER, Deputy Director/Chief Information Officer, Information Management, Defense Intelligence Agency (DIA)

INSCOM Perspectives and Initiatives
MS. LISA SHALER-CLARK, Deputy Director/Program Manager, Futures, Intelligence and Security Command (INSCOM), US Army

Air Force Perspectives and Initiatives
MR. SAMUEL OLIVER, Jr., Chief, ISR PED and Applications Division (A2CP), Office of the Deputy Chief of Staff for Intelligence, Surveillance and Reconnaissance, US Air Force

Big Data and Army Intelligence
MS. PATRICIA GUITARD, SES, Deputy Director, Intelligence (G-2), US Army (tentative)
Joint Staff Perspectives and Initiatives
DR. ANDREW HALL, Deputy Director, Intelligence, Joint Staff (tentative)

SPAWAR Perspectives and Initiatives
SENIOR REPRESENTATIVE, Space and Naval Warfare Systems Center, US Navy

II. Emerging Applications for Defense and Intelligence

Data Analytics and DCGS
MR. CHUCK GASSERT, Assistant Program Manager, Distributed Common Ground System – Navy (DCGS-N)

Large Scale Analytics and Network Management
DR. ROBERT BONNEAU, Program Director, Very Complex Networks, Air Force Office of Scientific Research (AFOSR)

ISR in a Tactical Environment
MR. TOM CONWAY, Office of the Project Manager, Night Vision/Reconnaissance, Surveillance and Target Acquisition (NV/RSTA), Program Executive Office – Intelligence, Electronic Warfare & Sensors (PEO-IEW&S), US Army

The DARPA UPSIDE Program
DR. DAN HAMMERSTROM, Program Manager, UPSIDE (Unconventional Processing of Signals for Intelligent Data Exploitation) Program, Defense Advanced Research Projects Agency (DARPA)
• Utilize the Physics of the Devices and Leverage Probabilistic Techniques
• To Achieve Orders of Magnitude Improvements in Processing Efficiency
• In Embedded Sensor Data Analysis Applications

Big Data Analytics and National Security
MR. TIM PAYDOS, Director, Worldwide Business Analytics and Optimization, IBM

GeoINT and Large Scale Analytics
MR. JASON DALTON, Senior Scientist, Digital Globe
• Using Scalable Hadoop Frameworks for Large Geospatial Computations
• Giving Users Simple Access to Powerful Geospatial Cloud Services
• Using the Open Source MrGeo Project to Power Large Scale GeoINT

Open Source, GOTS, and GOSS - Empowering Big Data and Enabling the Intelligence Mission
MR. BRUCE GOLDFEDER, Division Technical Lead, Data Tactics
• GOSS, GOTS, and Open Source Benefits
• Choosing the Bazaar over the Cathedral - Building in Modular Fashion; No OEM Costs
• Bidirectional Benefits Cloudbase - Accumulo
• Providing Government with a Marketplace of Capabilities
• Judicious Use of COTS
• Exemplars: COOP and DR for Big Data Systems; Single Global Dataspace Infinitely Linearly Scalable System
• Taking Big Data to the Battle Edge

III. The Latest Tools, Techniques and Technologies – Data Collection/Discovery, Deep/Predictive Analytics, Cloud, Scalability, Security, etc.

SPECIAL PRESENTATION:

ISR as a Candidate Solution Framework for ISR Big Data Volume, Velocity and Variety
MR. STEPHEN LONG
Director, ISR Technology Integration, Northrop Grumman; former Oversight Executive, Motion Intelligence, OSD

Big Data – Not Just About Processing Data with Analytics; It’s Also About How to Collect, Manage, and Discover Large Volumes of Real Time Data Across Globally Distributed Enterprises
MR. JOHN DELAY, Chief Strategic Officer/Architect, Geospatial Solutions, Harris Corporation

Challenges in Big Data Analytics – Applications and Capabilities
MR. STEVE HAGAN, Vice President, Server Technologies, Oracle Public Sector
• The Value of Big Data Technologies in Threat Identification, Situational Awareness, and Other Geospatially Intensive Applications
• Integrating Big Data Analytics with Spatial, Semantic and Traditional Business Intelligence Technologies
• How Engineered Systems Simplify the Creation and Configuration of Big Data Environments
• Practical Big Data, How to Go Beyond the Hype and Get Started to Achieve Data to Action

Large-Scale Analytics
MR. CHARLES FLEISCHMAN, Chief Technology Officer, Boeing Intelligence Systems
• Cutting Edge Analytics Engines; Advanced Discovery Engines; Advanced Visualization
• Workflow in Mobile Environments
• Data Security - Preventing Exfiltration in Multi-Level Secure Environments
• Next Generation Opportunities and Technologies - Information-to-Knowledge; Extraction of Deeper Meaning; Query Help; Feeding Analytics into Thought Process

Big Data Trends – Challenges, Solutions, Use Cases
MR. RICH CAMPBELL, Chief Technology Officer, EMC Federal
• What Defines Big Data Solutions
• Customer Challenges
• Use Cases

Social Data - What is It and Why Should You Care?
MR. TONY JIMENEZ, President and Chief Executive Officer, MicroTech
• The Exponential Growth of Social Media: Feeding the Big Data Machine
• Mastering the Big Data Explosion
• Integrating Social Media Analytics with Big Data
• Social Data as a Complex Nervous System
• Deriving Real World Success from Social Data
Social Media Analytics

MR. ROBERT McCORMACK, Director of Analytics, Aptima

Human-Driven Data Analytics

MR. BRIAN SCHIMPF, Forward Deployed Engineer, Palantir

• Explore Concepts and Research in Human-Driven Data Analytics
• Government and Commercial Applications
• Discussion of Large-Scale Financial Counter-Fraud Work
• Present Commercial Cyber-Security Analytics

Government Big Data Symposium

Source: http://www.bigdataconference.net/

Subject to Change 2/15/2013 (PDF)

Holiday Inn Rosslyn at Key Bridge, Arlington VA

Big Data and Government R&D – Turning Overload into Exploitable Information Assets

One year ago the White House announced funding for a large new research and development initiative aimed at extracting and exploiting the “knowledge and insights from large and complex collections of digital data … to help solve some of the Nation’s most pressing challenges.” This is in addition to numerous ongoing “Big Data” programs that have been initiated across the Federal Government, including Homeland Security, Defense, Intelligence, Education, Energy, Health and Human Services, NASA, and NSF aimed at spurring scientific discovery and innovation.

This outstanding symposium brings together the key government and industry experts who are shaping the direction of big data research and development across the Federal Government. They will provide you with an in-depth understanding of Federal agency strategy and plans, the status and forecast for key big data initiatives, and the latest tools and technologies being developed to exploit the massive amounts of information being collected at the Federal level. What are the most recent lessons learned from the commercial world? What applications are being developed for homeland security and intelligence analysis? What is the promise of health data analytics for developing new approaches to population health management, generating informatics-based treatments to major diseases, and coordinating response to public health crises? How is large-scale analytics being applied to ease overload for our warfighters? Synthesizing, sharing, and exploiting earth observation data? Providing the analytic basis for the next-generation of energy capabilities? Modernizing Education? What role are new tools, techniques, and technologies – predictive analytics, cloud computing, metadata, etc. – playing in making big data analytics at the Federal level a reality? What role can industry play? These and many other critical questions will be examined during this important two-day event.

Day 1: Tuesday, March 5, 2013

My Note: See Slides in Story: Matrix of First Symposium Presentations and Pilots

9:00-9:05 Administrative Announcements

9:05-9:35 DR. MARK LUKER, Associate Director, National Coordination Office of Networking and Information Technology Research and Development “NITRD Perspectives and Initiatives”
9:35-10:05 DR. SASTRY PANTULA, Director, Division of Mathematical Sciences, National Science Foundation “NSF Perspectives and Initiatives Related to Big Data”

10:05-10:35 DR. SASI K. PILLAY, Chief Technology Officer, Office of the Chief Information Officer, National Aeronautics and Space Agency (NASA) “NASA Perspectives and Initiatives”

10:35-10:55 Coffee Break & Networking Period

10:55-11:25 MR. GREG ELIN, Chief Data Officer, Federal Communications Commission (FCC) “FCC Perspectives and Initiatives and the Role of Data Officers”


11:55-12:25 MS. JO STRANG, Associate Administrator, Safety, Federal Railroad Administration, Department of Transportation “Open Gov 2.0 and Safety.Data.Gov”

12:25– 12:40 MR. TED MALONE (Moderator), Big Data Architecture Lead, Microsoft Federal

12:40-1:45 Lunch Sponsored by Microsoft

1:45-2:15 DR. ASHIT TALUKDER, Chief, Information Access Division, National Institute of Standards and Technology (NIST) “Big Data Challenges, Opportunities: Role of Measurement, Standards and Interoperability”


2:45-3:15 MR. SHAWN KINGSBERRY, Chief Information Officer, Recovery Accountability and Transparency Board “Operational Challenges and Considerations in Large-Scale Data”

3:15-3:45 MS. SUSIE ADAMS, Vice President, Federal Sector, Microsoft “Dealing with Structured and Unstructured Data”

3:45-4:15 Refreshments & Networking Period

4:15-4:45 DR. FREDERICA DAREMA, Director, (Member , Senior Executive Service) Mathematics, Information and Life Sciences, Air Force Office of Scientific Research (AFOSR); Former Senior Science and Technology Advisor, National Science Foundation “InfoSymbiotics/DDDAS: From Big Data to New Capabilities”

4:45-5:15 MR. KEVIN JACKSON, Vice President and General Manager, Cloud Services, NJVC “New Cloud Service Approaches for Big Data”

5:15-5:45 DR. NANCY GRADY, Technical Fellow, Data Scientist, Homeland and Civilian Solutions, SAIC “Big Data Across the Clouds”

5:45-6:15 MR. TOM PLUNKETT, Senior Consultant and MR. MARK JOHNSON, Director, Engineered Systems Program, Oracle Public Sector “Practical Big Data for Government”

Day 2: Wednesday, March 6, 2013

My Note: See Slides in Story: Matrix of First Symposium Presentations and Pilots

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by mindtouch 104
8:00-8:05 Administrative Announcements


8:35-9:05 MR. NIALL BRENNAN, Director, Office of Information Products and Data Analytics, Centers for Medicare and Medicaid Services, Department of Health Services (CMS/DHHS) “CMS Perspectives on Data and Analytics to Drive Health System”

9:05-9:35 MR. SCOTT GNAU, President, Teradata Labs “Unify Your (Big) Data Analytic Strategy”

9:35-10:05 MS. SOPHIE RASEMAN, Director for Smart Disclosure, Department of the Treasury “Smart Disclosure and Data Analytics”

10:05-10:35 MR. JEFF BUTLER, Director, Research Databases, Internal Revenue Services (IRS) “IRS Perspectives and Initiatives”

10:35-10:55 Coffee Break & Networking Period

10:55-11:25 MS. CARON KOGAN, Strategic Planning Director–Big Data, Lockheed Martin “The Art of Predicting with Big Data”

11:25-11:55 DR. MARC ABRAMS, Chief Technical Officer, Harmonia Holdings Group, LLC “Putting MapReduce on Steroids with Low Cost, Clustered GPUs”

11:55-12:25 MS. MARINA MARTIN, Entrepreneur-in-Residence and Head, Education Data Initiative, Department of Education “The Education Data Initiative”

12:25-1:30 Lunch Sponsored by AIE

1:30-2:00 DR. FLAVIO VILLANUSTRE, Vice President, Technology Architecture and Product, LexisNexis “The Value and Challenges of Large Scale Entity Analysis for National Security”

2:00-2:30 MR. DANTE RICCI, Senior Director, Public Sector, SAP Federal Innovation “Gaining Value from Big Data”

2:30-3:00 MR. JOHN MONTEL, eRecords Service Manager, Department of the Interior “Big Data, Big Records”

3:00-3:20 Afternoon Refreshments & Networking Period

3:20-3:50 MR. JUSTIN LEGARY, Director, FEMA National Exercise and Simulation Center (NESC) “FEMA Perspectives and Initiatives”

3:50-4:20 MR. DOMINIC SALE, Policy Analyst, Office of Management and Budget (OMB) “OMB Perspectives and Initiatives”

4:20-4:50 MR. SEAN BROPHY, Senior Analyst, Tableau Software “Analytics for Big Data Success”

4:50-5:20 MR. MICHAEL SCHULMAN, Director of Business Development and Marketing, ScaleMP “Virtual SMPs”

Brochure

http://semanticommunity.info/Big_Data_Symposia
Updated: Sat, 19 Sep 2015 02:40:40 GMT
Powered by mindtouch™
I. The Latest Federal Government Strategies, Plans, Needs and Initiatives

Setting the Stage – The Environment and Opportunities for Government Big Data

NITRD Perspectives and Initiatives
Dr. Mark Luker, Associate Director, National Coordination Office of Networking and Information Technology Research and Development
  • Strategy, Policy, Plans and Initiatives
  • Big Data and Related Programs Status, Updates, and Forecasts
  • Coordinating R&D Strategies and Directions Across Major Research Agencies
  • Big Data Steering Group

NSF Perspectives and Initiatives Related to Big Data
Dr. Sastry Pantula, Director, Division of Mathematical Sciences, National Science Foundation
  • Core Technologies
  • Workforce Development

NASA Perspectives and Initiatives
Dr. Sasi K. Pillay, Chief Technology Officer, Office of the Chief Information officer, National Aeronautics and Space Agency (NASA)

FCC Perspectives and Initiatives and the Role of Data Officers
Mr. Greg Elin, Chief Data Officer, Federal Communications Commission (FCC)
  • It’s About Having Great Data - Big Data Just Helps
  • Getting Great Data Without the PRA Headaches
  • What Does a Chief Data Officer Do, Anyway?
  • Success Measure: Answering Questions with URLs (with Examples from FCC)

DHS Perspectives and Initiatives
Mr. Michael Simcock, Chief Data Architect, and Mr. PAUL REYNOLDS, Senior Information Architect, Department of Homeland Security (DHS)
  • What does Big Data Mean to DHS?
  • Big Data Capabilities are being Tied to DHS Mission Focus
  • DHS is Taking an Architectural Approach to Guide Big Data Development Efforts
  • DHS is Ensuring Data Management does Play a Role with Big Data Development

Big Data for Defense
Mr. Alan Shaffer, Principal Deputy, Office of the Assistant Secretary of the Defense for Research and Engineering [invited]
FEMA Perspectives and Initiatives
Mr. Ted Okada, Senior Advisor for Technology, Office of the Administrator, Federal Emergency Management Agency (FEMA)

CMS Perspectives on Data and Analytics to Drive Health System Transformation
Mr. Niall Brennan, Director, Office of Information Products and Data Analytics, Centers for Medicare and Medicaid Services, Department of Health and Human Services (CMS/DHHS)

IRS Perspectives and Initiatives
Mr. Jeff Butler, Director, Research Databases, Internal Revenue Service (IRS)
  • IRS Enterprise Data
  • Analytic Data Environment
  • Methods, Tools, and Application Areas
  • Systems and Architecture
  • Best Practices and Lessons Learned

OMB Perspectives and Initiatives
Mr. Dominic Sale, Policy Analyst, Office of Management and Budget (OMB)

II. Technical Challenges and Mission Strategies

Big Data Challenges, Opportunities: Role of Measurement, Standards and Interoperability
Dr. Ashit Talukder, Chief, Information Access Division, National Institute of Standards and Technology (NIST)

Open Gov 2.0 and Safety.Data.Gov
MS. JO STRANG, Associate Administrator, Safety, Federal Railroad Administration, Department of Transportation (DOT)

Smart Disclosure and Data Analytics
Ms. Sophie Raseman, Director for Smart Disclosure, Department of the Treasury

Technical Challenges for Defense
Senior Representative/Program Manager (tbd), Defense Advanced Research Projects Agency (DARPA) [tentative]

DOI's eERDMS Program – Update and Forecast
Mr. John Montel, eRecords Service Manager, Department of the Interior (DOI)
  • eMail, Enterprise Records, and Document Management System (eERDMS) Program – Cloud-Based Records and Information Governance
  • Supporting Information Management for the Collaborative, Integrated Mobile Workforce
  • Capacity for 30,000 Additional Users
• FISMA Compliant
• 200 Million Annual Emails; 35 Terabytes; 100 Million Annual Digitized Records

The Education Data Initiative
Ms. Marina Martin, Entrepreneur-in-Residence and Head, Education Data Initiative, Department of Education

Operational Challenges and Considerations in Large-Scale Data Analytics
Mr. Shawn Kingsberry, Chief Information Officer, Recovery Accountability and Transparency Board (RATB)

III. Advanced Tools and Techniques

Large-Scale Text Analytics and Mining
DR. ASHOK SRIVASTAVA, Principal Scientist, Data Sciences, NASA Ames Research Center

Dealing with Structured and Unstructured Data
Ms. Susie Adams, Vice President, Federal Sector, Microsoft

The Art of Predicting with Big Data
Ms. Caron Kogan, Strategic Planning Director-Big Data, Lockheed Martin

• New Techniques for Mining Intelligences from Big Data
• Leveraging Entire Sets of Massive and Complex Data to Significantly Enhance the Accuracy of Predictive Modeling
• Predictive Analytics - An Art and a Science; but a Structured Approach is Mandatory
• Proven Guidelines for Building Predictive Analytics

The Value and Challenges of Large Scale Entity Analysis for National Security
Dr. Flavio Villanustre, Vice President, Technology Architecture and Product, LexisNexis

IV. Implementation Strategies and Lessons Learned

Big Data Implementation Strategies
Mr. Bruce Weed, Program Director, Worldwide Big Data Business Development, IBM Software Group and MR. BILL HARTMAN, President, TerraEchos

• The Importance of Implementing Big Data Now
• The Approach to Implementing Big Data
• Use Cases Around Big Data
• TerraEchos Use Case in Profile
• Why are We Implementing and Leveraging Big Data, What are the Real Benefits?

Driving Adoption and Impact with Big Data Analytics
Mr. Scott Gnau, President, Teradata Labs
Practical Big Data for Government

MR. TOM PLUNKETT, Senior Consultant, and MR. MARK JOHNSON, Director, Engineered Systems, Oracle Public Sector

• Discovering New Questions in Your Agency
• Jump Starting Big Data Projects
• Social Media for Governments
• The National Cancer Institute Genomics Project (Winner of the 2012 Government Big Data Solution Award)

Special Focus: Implementing Big Data in the Cloud

New Cloud Service Approaches for Big Data

Mr. Kevin Jackson, Vice President and General Manager, Cloud Services, NJVC

• Using Multiple Cloud Service Providers for Secure Storage
• Social Media Data Analytics
• Using Cloud Services Brokerage for Big Data Analytics
• Big Data Security in the Cloud
• Reducing the Cost of Big Data Analytics Using Spot Market Pricing

Big Data Across the Clouds

DR. NANCY GRADY, Technical Fellow, Data Scientist, Homeland and Civilian Solutions, SAIC

• Leveraging Public Clouds for Big Data
• Splitting Applications Across Hybrid Clouds
• Adding Big Data to Your Existing Infrastructure
• Architectures for Big Data Attributes

Migrating Applications to the Cloud

Mr. Mike Daconta, Vice President, Advanced Technology, InCadence Solutions; former Metadata Program Manager, Department of Homeland Security

Gaining Value from Big Data

Mr. Dante Ricci, Director, SAP Federal Innovation

• Learn How Top Performing Organizations are Investing Time Up-Front in Identifying Use Cases for Big Data, Segmenting the Targeted Users and Understanding the Value that May Result
• Understand How New Outcomes are Coming About Due to a Convergence of New Technology Enablers (Mobility, Cloud, Open, In-memory Real Time Data Platforms)
• Learn About and Discuss High Value Big Data Use Cases Scenarios

Analytics for Big Data Success

Mr. Sean Brophy, Senior Analyst, Tableau Software