3 Principles for Maximum Participation in Open Data on the Web

Bob Schloss, IBM Research ***April 19, 2013***

with John M Cohn, Kelvin Lawrence, Susan Malaika, John Reinhardt, David Rook, Biplav Srivastava

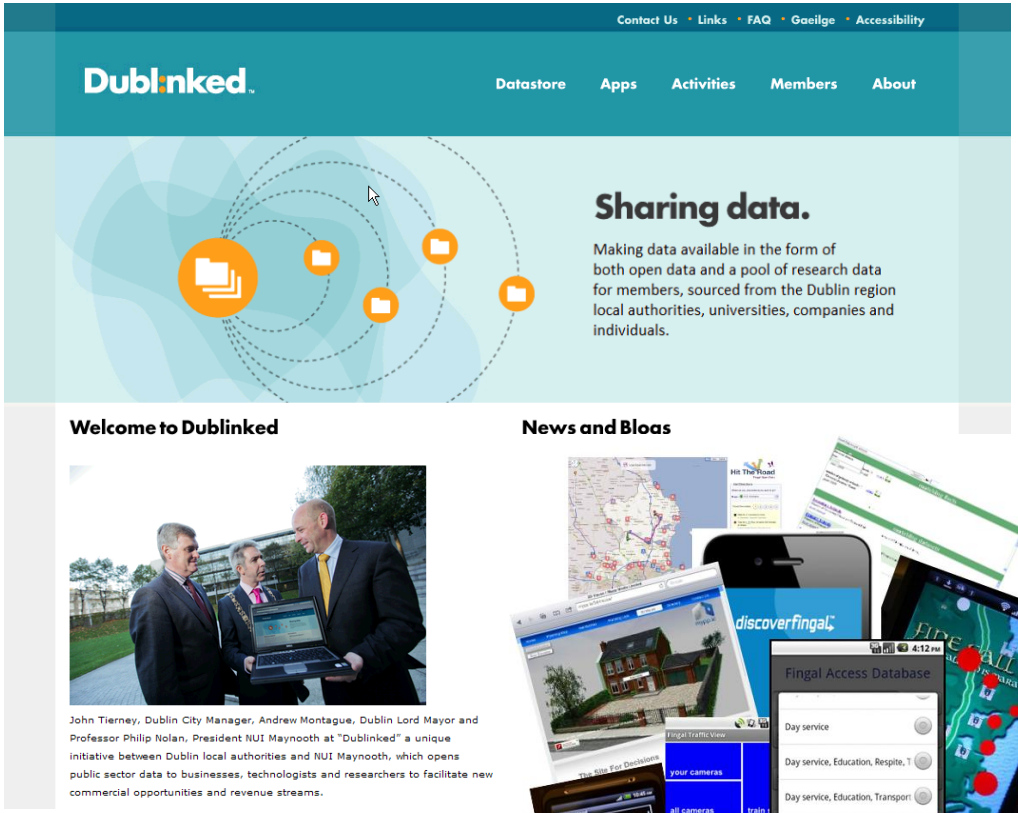
rschloss@us.ibm.com

Today's Discussion In a Nutshell

0. ***Background on our experience (2 slides, more online)***
1. ***Incentives must be in place to have "economically significant" amounts of Open Data***
2. ***Trust (Multiple Levels) with appropriate Security (where needed)***
3. ***Data Ownership and Provenance***

Background on our experience (slide 1 of 2)

Open data for innovative communities: Dublinked
(powered by Open Innovation Platform)



- "A city intelligence platform sharing information between Dublin's local authorities, Universities and major businesses.
- Launched in October 2011 with over 3,000 individual data sets across environment, planning, water and transport.
- The purpose of the partnership is to exploit the intelligence platform to transform city systems and stimulate the creation and growth of high-value businesses with international markets.
- 10's of applications have been implemented from the Open Data platform resulting in the first new businesses backed by VC with global markets
- The partnership has generated specific business cases for Dublin to invest in transformations of city systems with proven social, economic and environmental benefits.

https://www-304.ibm.com/connections/blogs/ibmireland/entry/innovative_new_data_sharing_network_to_create_commercial_opportunities1

Background on our experience (slide 2 of 2)

What We Have Learned

The Key Lies in Delivering Trusted Information -- On Multiple Levels

Useful

Combinable yielding meaning for your stakeholders

Can receiver merge & deliver to make it all valuable?

Governed & Secure

Rules are in place and tools are deployed to limit visibility, secure sensitive information, and protect privacy

Accurate & Complete

Complex and disparate data transformed, cleansed, reconcile and delivered

Can I Trust My Partner and My Own Agency? ***Can I Trust The Information?***

Background on our experience (slide 2 of 2)

What We Have Learned

The Key Lies in Delivering Trusted Information -- On Multiple Levels

Useful

Combinable yielding meaning for your stakeholders

Can receiver merge & deliver to make it all valuable?

Governed & Secure

Rules are in place and tools are deployed to limit visibility, secure sensitive information, and protect privacy

Can I Trust My Partner and My Own Agency?

Accurate & Complete

Complex and disparate data transformed, cleansed, reconcile and delivered

Can I Trust The Information?

Prior Path to Success

- Sense and Respond
- Instinct and Intuition
- Skilled Analytics Experts
- Back Office Decision Support
- Automated Processes

→

- Predict and Act
- Real-time, Fact-driven
- Everyone
- Point of Impact
- Optimized

1. ***Incentives must be in place to have "economically significant" amounts of Open Data***

1. Recognition: At this level, the producer wants simply to be recognized for the data produced and found valuable.
2. Indirect Benefit: At this level, the data producer wants the consumer to benefit using open data and credit for participating.
3. Direct Benefit: At this level, the data producer wants value from usage by consumer.

1 ***Incentives must be in place to have "economically significant" amounts of Open Data***

1. Recognition: At this level, the producer wants simply to be recognized for the data produced and found valuable.
2. Indirect Benefit: At this level, the data producer wants the consumer to benefit using open data and credit for participating.
3. Direct Benefit: At this level, the data producer wants value from usage by consumer.

For Data Providers:

-a way (if they choose to use it) to guarantee that their (multimedia) Logo appears in front of the relevant consuming enterprise participant as a function of time, volume, and location of the consuming programmatic logic.

-a way (if they choose to use it) to receive identities (minimally, e-mail addresses or URIs) when their data is fetched. Current web technologies, such as Browser-held Cookies, can simplify the annoyance factor to web users of providing this identity information repeatedly.

- a way (if they choose to use it) to charge for the data delivered, not only as a function of the data itself, but as a function of the requester organization and the technical protocols used to transfer the data (which could vary considerably in their resources used, time to execute, etc.)

2 *Trust and Security Mechanisms for Ensuring Data Quality*

1. who collected this data (including what versions of what systems did any preprocessing of it), was it opt-in
2. who authorized distribution of this data
3. which version of this data was actually transferred to satisfy a particular request
4. has any assertion of "exclusive access" or "exclusive until after date-time access" be made by the data supplier
5. has any assertion of "only complete redistribution permitted" (no partial redistribution) been made
6. has any assertion been made as to the authenticity of the data and are there guarantees in place that it has not been tampered with.

2 *Trust and Security Mechanisms for Ensuring Data Quality*

1. who collected this data (including what versions of what systems did any preprocessing of it), was it opt-in
2. who authorized distribution of this data
3. which version of this data was actually transferred to satisfy a particular request
4. has any assertion of "exclusive access" or "exclusive until after date-time access" be made by the data supplier
5. has any assertion of "only complete redistribution permitted" (no partial redistribution) been made
6. has any assertion been made as to the authenticity of the data and are there guarantees in place that it has not been tampered with.

2 *Trust and Security Mechanisms for Ensuring Data Quality*

1. who collected this data (including what versions of what systems did any preprocessing of it), was it opt-in
2. who authorized distribution of this data
3. which version of this data was actually transferred to satisfy a particular request
4. has any assertion of "exclusive access" or "exclusive until after date-time access" be made by the data supplier
5. has any assertion of "only complete redistribution permitted" (no partial redistribution) been made
6. has any assertion been made as to the authenticity of the data and are there guarantees in place that it has not been tampered with.

2 *Trust and Security Mechanisms for Ensuring Data Quality*

1. who collected this data (including what versions of what systems did any preprocessing of it), was it opt-in
2. who authorized distribution of this data
3. which version of this data was actually transferred to satisfy a particular request
4. has any assertion of "exclusive access" or "exclusive until after date-time access" be made by the data supplier
5. has any assertion of "only complete redistribution permitted" (no partial redistribution) been made
6. has any assertion been made as to the authenticity of the data and are there guarantees in place that it has not been tampered with.

3 *Data Ownership and Provenance*

1. Before open data transfer time, data consumers understand their obligations with respect to later 3rd party allegations of improper distribution or use.

2. Before open data transfer time, data providers understand the obligations with respect to later 3rd party allegations of improper distribution they will need to manage because data consumers are not taking full responsibility to handle them.

Today's Discussion In a Nutshell

1. ***Incentives must be in place to have "economically significant" amounts of Open Data***
2. ***Trust (Multiple Levels) with appropriate Security (where needed) – avoids "maldata"***
3. ***Data Ownership and Provenance***

We look forward to working with you to turn open data into Plentiful Big Open Data!

rschloss@us.ibm.com

Backup

Big Data Commission
Getting Started – Recommendations For Government Agencies

Big Data is here -- and it is accelerating.

For government agency leaders focused on taking the first steps, the Commission makes the following recommendations:

1. Understand the “Art of the Possible”.
2. Start with a clear mission or business requirement, and fully define a discrete set of use cases.
3. Take inventory and understand your data assets.
4. Assess your current set of capabilities and technical architecture against what is required to support your initial use cases.
5. Explore which data assets can be exposed for public consumption, to drive innovation and the development of Big Data solutions.

http://www-05.ibm.com/fi/ibm/wdc2012/pdf/IBM_SCC_Helsinki__English.pdf

Open data creating change: Helsinki

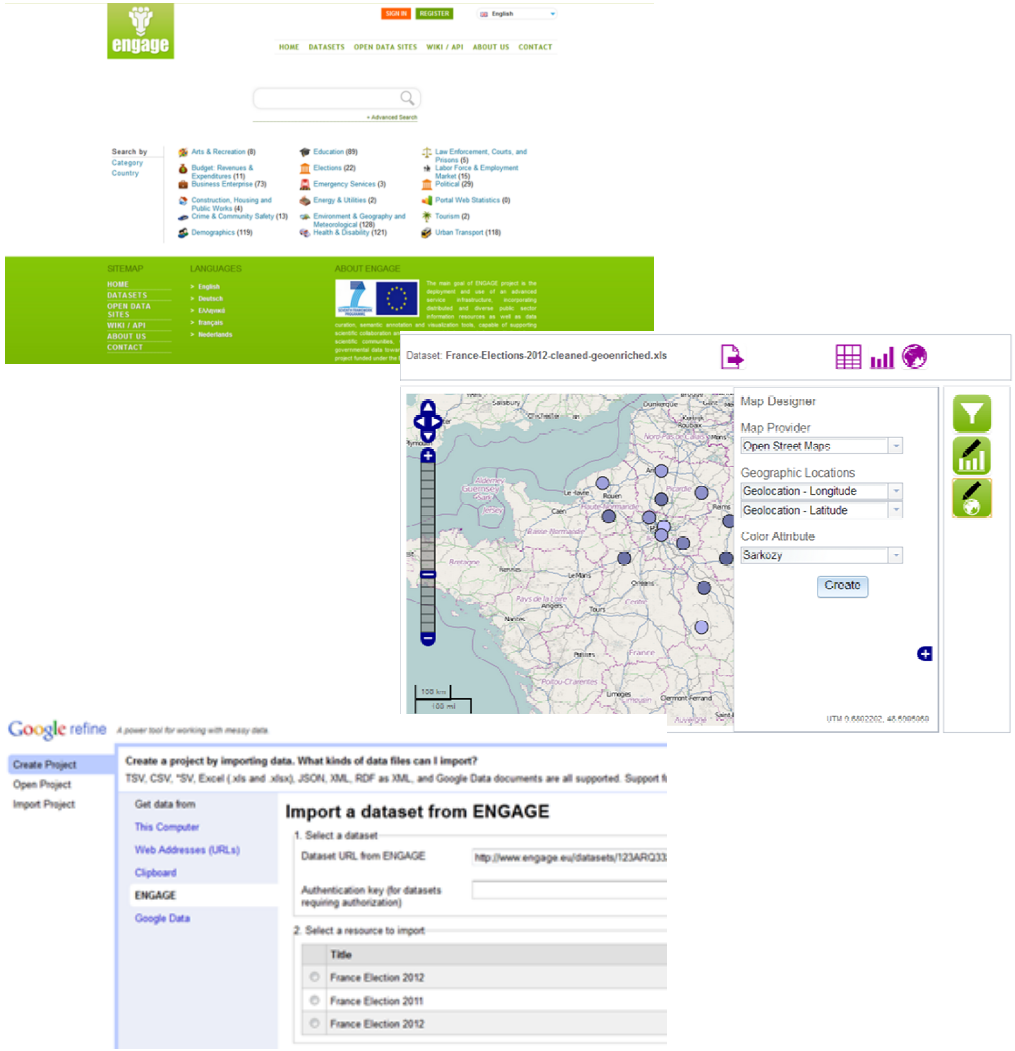
The IBM Smarter City Challenge

Examples of design sketches which were produced as part of the vision for visualization based on open data

- A team of six IBM experts worked for 3 weeks to deliver recommendations on visualizing open data to Mayor Jussi Pajunen and city stakeholders.
- The challenge addressed the challenge of making digital data from the city's open data initiative accessible and useful to citizens through engaging visualizations to support the World Design Capital event in Helsinki in 2012.
- The IBM team developed recommendations for:
  - Defining a vision for the future based on citizen engagement through visualization of open data
  - Defining the components necessary to grow a sustainable, repeatable platform, process and ecosystem to leverage the principles of open data, and turn data into information, into action and into change.
- The City have announced that it plans to adopt the recommendations made in the Challenge report.

http://www.engage-project.eu/engage/wp/

Open Data and Visualisation for Research: Engage

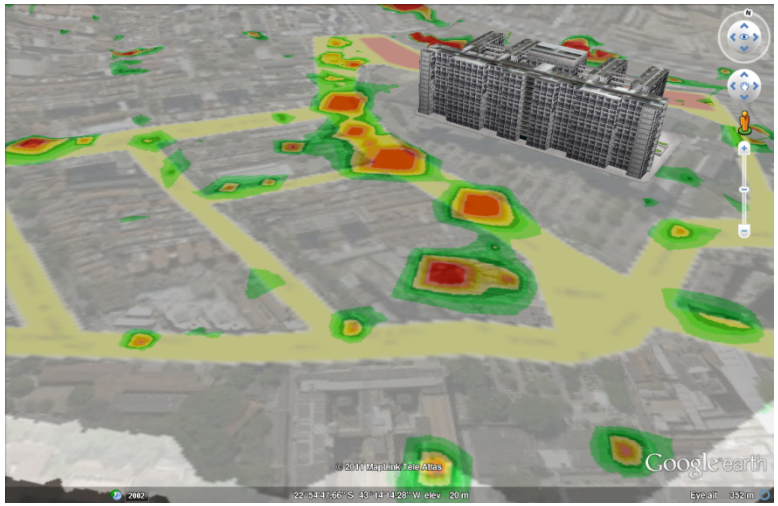


Engage data publishing and provisioning components

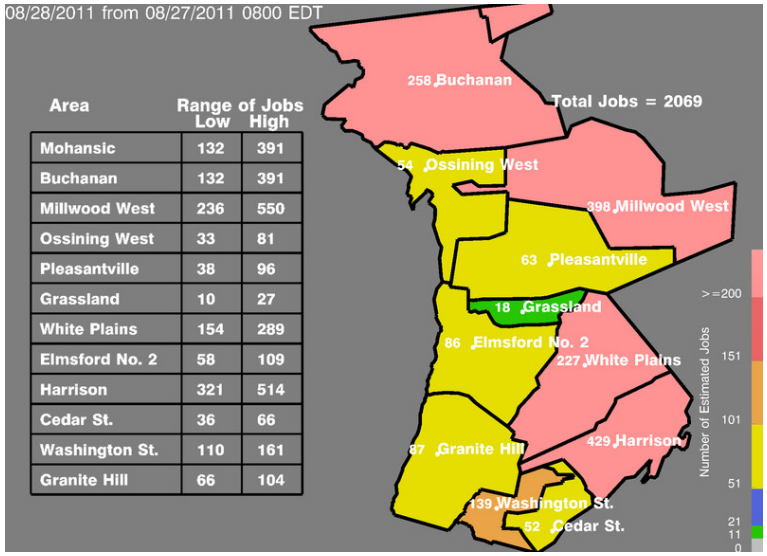
- Engage is an EU funded e-Infrastructure project providing an open data platform for the Researchers community.
- Supports dataset publishing and provisioning activities such as curation, linking, visualization, searching and more.
- IBM developed:
  - Data visualization services for tabular data (grid, charts, maps).
  - Integration with Google Refine for cleansing, curation and linkage.

http://asmarterplanet.com/blog/2012/07/18315.html

Open data and the business of weather: Deep Thunder



Rio de Janeiro - flooding model at 1m resolution



Area	Range of Jobs Low	High
Mohansic	132	391
Buchanan	132	391
Millwood West	236	550
Ossining West	33	81
Pleasantville	38	96
Grassland	10	27
White Plains	154	289
Elmsford No. 2	58	109
Harrison	321	514
Cedar St.	36	66
Washington St.	110	161
Granite Hill	66	104

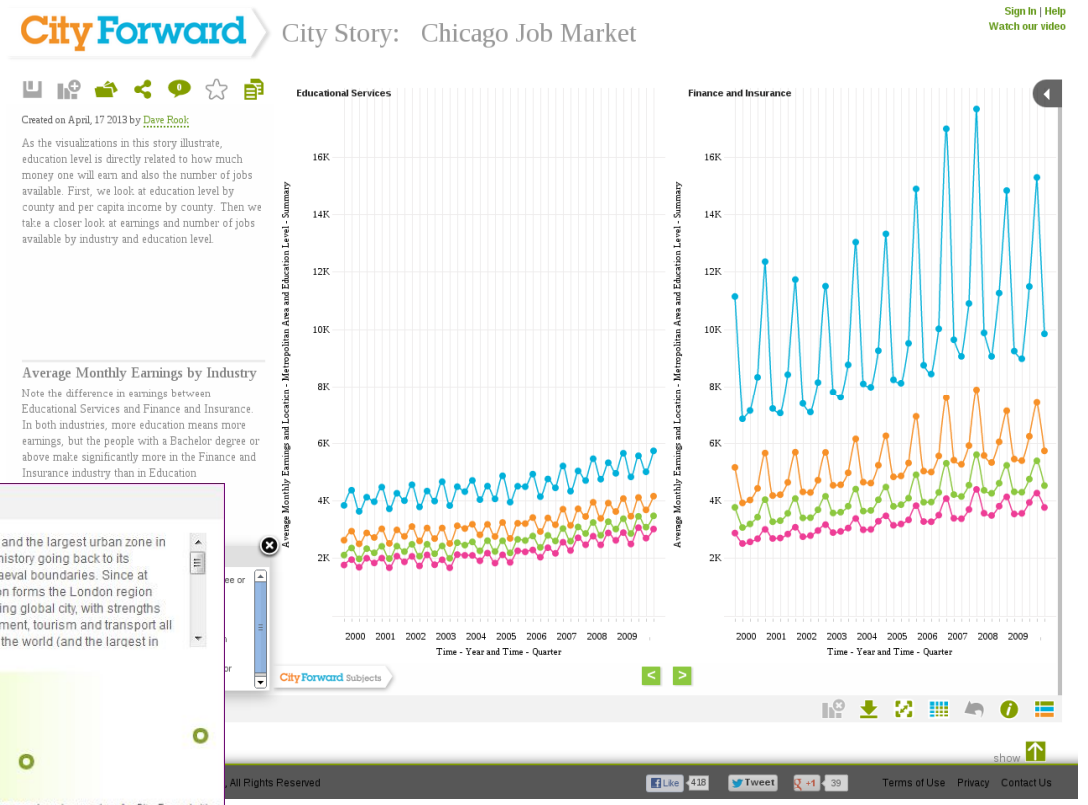
Tropical Storm Irene - Forecasted Outages per Substation

- Provides customizable, accurate, high resolution weather predictions up to three days ahead.
- Models, analyses, visualises and disseminates insight from open data, organisational data, commercial data and crowdsourcing.
- Forecasts the impact of weather on businesses and systems such as energy, cities, transportation, and agriculture.
- Cities, water, energy, and transportation systems can be Smarter, safer and more efficient if the effects of weather on underlying business processes can be predicted and optimized:
  - Energy: foresee outages; predict supply from renewable resources; proactively manage load and maintenance ...
  - Cities: foresee emergencies; proactively manage demand, water quality and maintenance ...
  - Transport: foresee disruption to road, rail, air and marine transport; proactively manage capacity and demand; proactively manage maintenance ...
  - Agriculture: foresee flooding and droughts; proactively manage planting, harvesting, fertilisation and supply chain activities ...
  - Others: insurance, mining, retail, sports, tourism ...

http://cityforward.org

City Forward – a free web platform for sharing open data

- City Forward is a free, web-based platform that enables users – city officials, researchers, academics and interested citizens world-wide – to view and interact with open city data while engaging in an ongoing public dialogue.



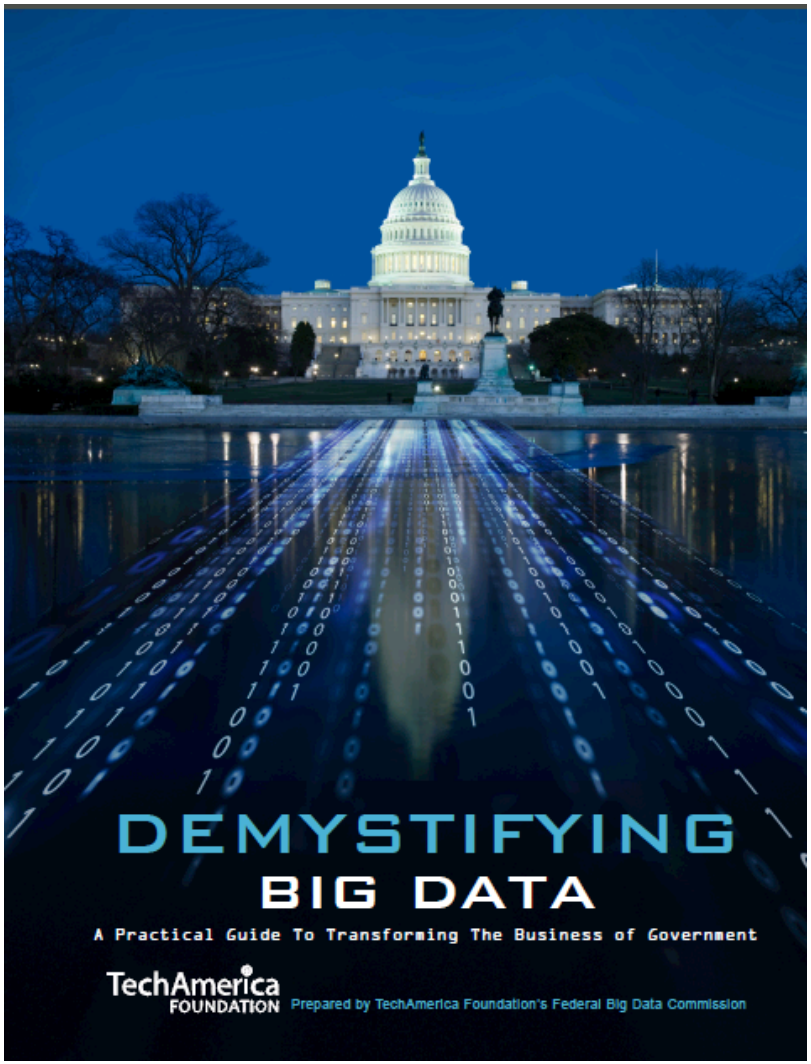
Quick Facts:

129 Cities
44 Countries
122 Datasets
147 City Stories

Quick Summary

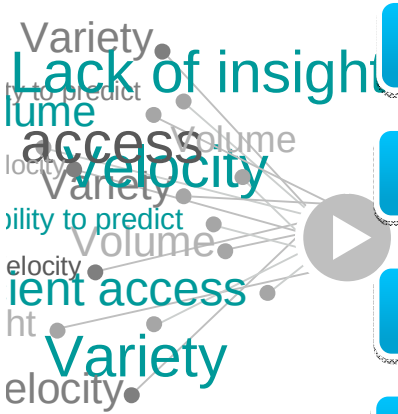
Big Data in Govt Commission Report Released 2012 October 3

- The Commission was formed in May 2012
- IBM & SAP Named as Chairs
- Brought together experts from Government, Academia and Industry
- The report seeks to “De-Mystify” Big Data, and focus on the business and mission value it will deliver
- Intent is to Provide clear recommendations and a roadmap for getting started



Gov't Agencies, Businesses, NGOs Leading The Way

Optimize decision-making with actionable insights from blended data



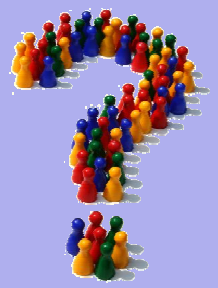
Prior Path to Success	Today's Leaders
Sense and Respond	Predict and Act
Instinct and Intuition	Real-time, Fact-driven
Skilled Analytics Experts	Everyone
Back Office Decision Support	Point of Impact
Automated Processes	Optimized

What We Have Learned

Big Data Requires A Different Approach – It Morphs The Traditional Analytics Process

Traditional Approach

Business Users
Determine what question to ask

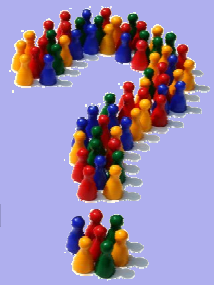


IT
Structures the data to answer that question

Big Data Approach

IT
Delivers a platform to enable creative discovery

Business
- Combines Open with Proprietary Data
- Explores what questions could be asked



Structured & Repeatable Analytics
- Query Based -- Question Drive Data
- Citizen Surveys - Push
- Monthly, Weekly, Daily
- Data At Rest

VS.

Iterative & Exploratory Analytics
- Autonomic -- Insight Drives Answers
- Citizen Sentiment -- Pull
- Persistent, Just-in-Time & Ad Hoc
- Data In Motion

Open data with impact: what we have learned so far

- "Open" is expanding how governments manage relationships with citizens and the data that both need

- The understanding of "open" is uneven *and* definitions have evolved

- Many jurisdictions are already adopting this new, evolving *style* of public administration

  1. **Providing raw data**
  2. **"Seeding" innovation**
  3. **Enabling collective problem solving**
  4. **Creating the "bazaar"**

- The potential benefits for *both* citizens and government are compelling

- A different *mental model* is needed to guide leaders in their open journey

- ***What*** – Strategically integrate, execute on *and* manage four areas

  - **Information**
  - **Engagement model**
  - **Digital Platform**
  - **Analytics competency**

- ***How*** – Take a series of three steps and repeat progressively over time

  1. **Define and measure "openness"**
  2. **"Open up":**
  3. **Capitalize on "Open"**

- To best *position* for realizing potential benefits, leadership and flexible, sustainable engagement model(s) are essential

- To *sustain* the journey, set key indicators, measure and manage the balance between different engagement models

Source: IBM Institute for Business Value

What We Have Learned

The Key Lies in Delivering Trusted Information -- On Multiple Levels

Useful

Combinable yielding meaning for your stakeholders

Can receiver merge & deliver to make it all valuable?

Governed & Secure

Rules are in place and tools are deployed to limit visibility, secure sensitive information, and protect privacy

Can I Trust My Partner and My Own Agency?

Accurate & Complete

Complex and disparate data transformed, cleansed, reconcile and delivered

Can I Trust The Information?