Big Data, Bad Data, Good Data
The link between information governance and big data outcomes

Danijela Strmac, ECM Sales Leader Gulf&Levant
Big Data is being used to solve problems across every industry
Use Cases in Every Industry….

- Understand customer preferences
- Target buyers with tailored offers
- Optimize supply chain
- Anticipate product problems/warranty issues
- Improve performance of enterprise assets
- Improve demand forecasts
- Build smarter grids
- Reduce outages
- Optimize production
- Optimize care
- Improve patient outcomes
- Accelerate medical/scientific discovery
- Intelligence
- National security
- Mission support/planning
- Detect/prevent fraud

© 2015 IBM Corporation
Q10. Which of the following areas have been identified as significant drivers of your organization's big data and analytics initiatives over the next 12 to 24 months?

IDC Big Data and Analytics Maturity Survey, July 2013

And in Every Major Operational Area

- Product or Service Improvement and Innovation
- Customer Service and Support
- Pricing Strategies and Programs
- Market and Competitive Intelligence
- Process, Operations Optimization and Control
- Customer Acquisition and/or Retention
- IT Optimization and/or Modernization
- Regulatory Compliance, Financial Controls
- Supply Chain Management and Logistics
- Human Capital Management
- Operational/Fraud/Malicious Risk Management
- Plant/Facilities/Equipment Design, Maint & Use

% of respondents
However, most implementations are missing a big piece of the puzzle…
Unstructured Data

➔ 90% of enterprise information is unstructured

* Source: IDC, 2014 Study
The potential for insight is HUGE

But there are particular challenges to tackle ...
Unstructured data takes different shapes and lives in many different places:

- Hundreds of formats
  - Documents, emails, wikis, etc
- Dozens of locations
  - Content management system(s)
  - Team collaborative sites
  - Email
  - Network fileshares
  - User laptops
- Both on-premise and in the cloud
And often isn’t well governed…

- What isn’t worth keeping?
  – Redundant data
  – Obsolete data
  – Trivial data

- What is risky to keep around?
  – Legal risk
  – Personal information

- What data can you trust?
- Who are the experts?

Active Data Intelligence

Understand | Analyze | Manage
The answer is Data Curation

Delivers relevant, qualified and governed content collections

Increases user confidence by improving information relevance and quality
What does Data Curation involve?
Striking the Big Data vs. Good Data Balance:

Aligning the "keep everything" needs of Big Data with the "defensible disposal" mandates of Information Governance

Keep data

Delete data
• Current value to the business
• Potential for future insights

• Is the data from a trusted source?
• Data Breach
  • Personal information
  • Corporate IP
• Handling per corporate regulatory requirements

• Storage – Present & Future
• Potential for inclusion in litigation proceedings
Why can’t I just search for it?

- Too many results
- Too time consuming
- Not enough man power to sift through everything
- Doesn't scale
- No access to the right sources (repositories)
Without Data Curation...
Important insights can be buried...
Data from untrusted sources can lead to incorrect conclusions
Privacy may be violated

A single breach of sensitive personal data cost $3.5 million in 2014

Ponemon Institute
2014 Global Cost of Data Breach Study
Sponsored by IBM

ibm.com/services/costofbreach

© 2015 IBM Corporation
What makes a good Data Curation solution?
Relevant: Provides verified content from trusted sources

Dynamic: Remains current as information evolves

Defensible: Is auditable to confirm responses

Transparent: Provides provenance and lineage
Connects to data in its native environment

IBM Information Lifecycle Governance Platform
Helps you understand your data
Part of a comprehensive Information Governance strategy

**IBM Information Lifecycle Governance**

**DISCOVER | RECOGNIZE | ACT**

**INDEX DATA IN PLACE**

- Archive Platform
- ECM
- Forensic Images/Tapes
- File Servers
- Email Servers
- Desktops
- SharePoint & Enterprise Collaboration
- Cloud
- Media
How do I get started?
1 Understand your information and its lifecycle
Define and enforce an overall retention schedule
3 Determine departmental legal requirements
4 Throw out the trash
5 Extract analyzable data
6 Sanitize before you analyze
Recommendations

- Be methodical in your approach
- Create a solution that will scale for the enterprise
- Foster strong communication among stakeholders - include the big data analytics team
- Make governance a best practice
Thank You!