Data Governance and Big Data - A Necessary Convergence

Richard Goldberg
Chief Data Governance Officer
Citibank Global Consumer Bank
Data Governance and Big Data – A Necessary Convergence

Discussion Focus Area

- As our businesses continue to expand its use and dependency on big data, it is also evident that we need to extend the reach of data governance to ensure the data is properly managed and leveraged within the confines of regulations and associated business controls.

- The challenge is when and how to establish a data governance foundation; and what are the main priorities on managing data in this new horizon.

- Key areas we will discuss include:
  - When should data governance become involved in “big data” initiatives?
  - What are the main issues associated with data governance and “big data”?
  - What are the main priorities associated with data governance and “big data”?
  - What is needed to address a data governance foundation around “big data”?
As The World is Changing….

- Data is exploding
- Compliance pressure continues to increase
- Business demands better information for faster decisions
- More data available to manage
.....Today’s Data Challenges Continue to Proliferate

Big Data content doubles every 12 to 18 months!

Data Governance is needed to manage this environment

Disconnected systems

Data Privacy

Conflicting insights

Manual reconciliations

Incomplete data

Proliferation

Scattered information

Department-level reporting

Isolated applications

Tool proliferation

Spreadsheets, homegrown databases

Master data inconsistency
**Definition of Data Governance**

**Data Governance:** A formal system of accountability designed to enforce proper management of data assets and the performance of data functions. Data governance encompasses the actions (includes people, processes, and technology) used to ensure that **key information** delivered throughout the organization is appropriately defined, used and maintained.
Why do we need to Govern Data?

- Information is a vital enterprise resource and is critical to business success
- Consistent and accurate information is fundamental for financial reporting, business intelligence and measuring the execution of corporate strategies
- Regulatory compliance has come to the forefront of information initiatives
- Historically, data has not been governed with the same rigor given to other vital assets, as evidenced by the lack of quality, timeliness, transparency and reuse
- The introduction of leveraging unstructured data ("big data") has increased risk and exposure if not managed properly
End-to-End Operational Model for Big Data

Management & Governance
- Steering Committees
- Program Management
- Strategy & Planning
- Community Management
- Controls & Compliance
- Project Management
- Data & Analytics Roadmap
- Service Management
- Operational Support

Analytics
- Idea Management
  - Idea Inventory
  - Portfolio Management
  - Value Management
  - Prioritization
- Execution
  - Model Creation
  - Model Management
  - Analytics Execution
  - Analytics Support
  - Analytics Peer Review

Data Governance
- Data De-Identification
- Data Provisioning
- Data Integration
- Data Access
- Data Privacy
- Data Transforming
- Data Cleansing
- Data Security
- Data Ingestion

Technology
- Tools
  - Application Development
  - Hadoop Engineering
  - Knowledge Management
  - Analytics Tools
  - Analytics Training
- Infrastructure
  - Hadoop Environment
  - Big Data Sandbox
Access to data must be approved by local business unit and Data Governance Officer.

Access will be on a limited and timed basis.

Data analytics will start out as R&D only and may shift to production after approvals.

Data Privacy controls will be enforced.

There will be no access to PII data.

All data needs to be de-identified to avoid ability to link back to account owner.

Access will be limited to a security zone (i.e. product/region) and users can see all of the data sets associated only to that zone.

If an analyst needs access to data from another line of Business, Data Governance Office (DGO) pre-approval is required.
Discussion/Questions