12 Vs of Big Data Governance

Richard J Self
Impact of the “12 Vs” of Big Data on Questions of Ethics, Trust, Stewardship and Governance of Analytics

SAS Analytics Europe 2014

Richard J Self, University of Derby

r.j.self@derby.ac.uk
http://computing.derby.ac.uk/wordpress/people-2/richard-j-self/

Acknowledging research by students from the School of Computing and Maths at the University of Derby in Spring 2014
Context (1)

• Many IT Governance Issues (COBIT and ISO 27k, etc.)
  – Challenges of Ethics, Trust and Compliance
• 3 Vs define Big Data
• Other Vs can
  – Improve understanding of Governance
  – Guide implementation
  – Manage expectations
Context (2)

• Governance Framework Style
  – COBiT – strictly defined approach
    • Tick the boxes
  – ISO27k – defined questions (ISO 27002)
    • Ask context based questions
Context (3)

• Analytics Skills Shortage
  – Approx 28K new analytics experts needed by 2017 (SAS report, 2013)
  – Implies 300 recruits to analytics programmes at every UK University in Sept 2014

• Unlikely to happen

• Plan B? – significant Corporate governance issue
ISO 27002

Context Sensitive Governance Questions

- Risk Evaluation and Management
- Compliance with External Obligations
- Compliance with Internal Obligations

ISO 27002:2013, Section 0.2
Answers or Questions?

If we are taught answers we stop learning!
To ask is to know.
The Power to Know (SAS®)
The Power to Say No!
Agenda

• The Vs
• The Questions
• Some Consequences
• A (new?) Governance Framework
• Meeting the Skills Shortage
12 Vs of Big Data

• Traditional 3 Vs
  – Volume (size)
  – Velocity (speed)
  – Variety (sources/format)

• New Vs
  – Variability (temporal?)
  – Value (to whom?)
  – Veracity (truth)
  – Validity (applicable)
  – Volatility (temporal)
  – Verbosity (text)
  – Vulnerability (security)
  – Verification (trust?)
  – Visualisation (presentation)
Key Themes

• Ethics
• Trust
• Stewardship
• Governance

Definitions?
Ethics

That branch of philosophy dealing with values relating to human conduct, with respect to the rightness and wrongness of certain actions and to the goodness and badness of the motives and ends of such actions

Dictionary.com
Trust

Reliance on the integrity, strength, ability, surety, etc., of a person or thing; confidence.

Dictionary.com
Stewardship

The responsible overseeing and protection of something considered worth caring for and preserving

A data steward is a job role that involves planning, implementing and managing the sourcing, use and maintenance of data assets in an organization. Data stewards enable an organization to take control and govern all the types and forms of data and their associated libraries or repositories

• Problems
  – A common result is that data availability, accuracy, timeliness, protection and accessibility issues abound
Governance

Corporate governance consists of the set of processes, customs, policies, laws and institutions affecting the way people direct, administer or control a corporation. Corporate governance also includes the relationships among the many players involved (the stakeholders) and the corporate goals.

Wikipedia

Doing the right thing in the right way at the right time with the right resources to the right quality in the right place for the right reasons.

To Ask is to Know!
Volume \((S, G)\)

- How much?
- Who?
- How?

Mountains

Brian Bielmann

Wikimedia

Oceans (Teahupo’o)

- Storage
- Dangers
- Technology?
Velocity \((S, G)\)

- Infrastructure
- Need for immediacy?
- Time to analyse
- How fast?

[Images: Wikimedia, Twitter]
Variety (T, S, G)

• What?
• Connections?
• Similarity?
• Original purpose?

Fruit salad?

Choices
• Data
• Structures
• Ownership

http://applesororanges.wordpress.com/cell-phones/
Variability \((T, S, G)\)

- Consistency
- In-consistency
- Periodicity

Frankfurt weather forecast

- Data
- Analyses
- Predictions

seasons

[Image of weather forecast and earth四季图]
Value ($G$)

- Define Value!
- Monetary?
- Understanding?
- Whose value?

Gold crystals

 Wikimedia

- Functionality?
- Convenience?
- Questions
- Answers?

Applications
Veracity \((T, S, G)\)

- Truthful when?
- Identifying the dross?

YouGov

We know we err

Wikimedia

Finding the Truth

- Data
- Algorithms
- Data cleansing
Validity (E, T, S, G)

- Context?
- Where?
- When?
- Correlation or Causation?

- Data
- Algorithms / Models
- Prediction

UK Penny Black

Johnson, 2008

Logical Deduction / Induction
Volatility \( (T, E, S, G) \)

- How quick?
- Value left?
- Data
- Predictions
- Sources of expertise

Snapchat

Wikimedia

Salt Pans in France
Verbosity \( (T, G) \)

- Context
- Meaning / semantics
- Grammar

- Syntax
- Redundant
- Knowledge
Vulnerability \((T, E, S, G)\)

- Legal (e.g. DPA and Right to be Forgotten)
- Reputation
- Financial
- Data
- Policies
- Processes
- Access
- Who and what skills?

OWASP: relationship between threat agent and business impact
Verification (T, E, S, G)

• How?
• Reliability?
• Context?

• Data
• Algorithms and models

1st Japanese Passport 1866

Digital Signature process
Visualisation \( (T, E, G) \)

- Manipulation
- 2D / 3D
- Video

As presented by the UK Treasury

Log scale
Breached
Official Guidelines

As corrected by Sir Andrew Dilnot

- Who chooses?
- Guidelines and standards?
Consequences

The Power to Know
And
The Power to Say No!
Clear Critical Analysis

• Implicit / intuitive choices are like habits or assumptions
  – Thinking Fast (Daniel Kahneman, 2013)
  – Often wrong, dangerous

• Explicit choices reflect Knowledge
  – Thinking Slow (Daniel Kahneman, 2013)
  – Are clearly justified
  – More accurately reflect current, specific situation
Protect the Organisation

• Use Governance Frameworks
  – 12 Vs
  – ISO 27002-2013, Section 0.2
    • Risk Evaluation and Management
    • Compliance with External Obligations
    • Compliance with Internal Obligations
  – Any other sets of useful questions

• Active, deliberate Analysis, not ticking Boxes
Prevent Danger

- To organisation
- To data
- To staff
- To customers
- To shareholders
Skills Shortage

• Requirement
  – Working together
    • Analytics providers
    • Businesses
    • Education
  – More effective Teaching by Research (doing)
Skills Shortage

• Requirement
  – Working together
    • Analytics providers
    • Businesses
    • Education
  – More effective Teaching by Research (doing)
Learning by Research (1)

• Very broad assignment topic (Vs research)
  – Big Data for SMEs:- Questions of Opportunities, Challenges, Benefits and Operations] and [Ethics, Trust, Governance, Security, Audit and Provenance]

• E-book publications
  – 82 Articles from 110 students (>= 70% mark)
    – http://commerce3.derby.ac.uk/ojs/index.php/itpsme
Learning by Research (2)

• Base SAS 9.3 by research (not “20 Chapters”)
  – UCLA Tutorial
  – SAS Manuals

• Focussed Project - Target
  – Menu driven system
  – 6th Normal data

• 3 students used SAS/AF to prototype SAS VA
References


• Kahneman, D, 2013, *Thinking, Fast and Slow*, Farrar, Straus and Giroux


• Teradata, 2008, Establishing data stewards, [http://apps.teradata.com//tdmo/v08n03/Features/EstablishingDataStewards.aspx](http://apps.teradata.com//tdmo/v08n03/Features/EstablishingDataStewards.aspx)