Data Governance Demystified - Lessons From The Trenches

Jay Zaidi, PMP
December 11, 2011

Introduction

Data Governance is gaining importance lately, due to a renewed focus on regulatory compliance and risk management. Most companies already have some form of data governance in place, but may lack formality or may limit its scope to governance within business silos, rather than at the enterprise level. Governing data across an enterprise is non-trivial and companies typically attempt it a few times, before they get it right. There are many reasons for this - scope creep, a corporate culture that is resistant to change, poor change management practices, lack of sponsorship from the top, lack of education about benefits of the program, poor strategy, and budgetary challenges.

Certain fundamental concepts related to governance must be understood before embarking on the program. In addition to this, the following questions must be addressed - What is it that must be governed?, Who is supposed to do this?, How do I do this at the enterprise level?, How long will it take for me to implement a Governance program throughout my organization?, How can I do it in a cost-effective way? and How do we sustain it and continuously improve it over time?. Like any other enterprise-wide initiative there are no simple answers, but if treated as a strategic priority and built systematically, sustainable governance is achievable. The strategy and approach will vary based on the maturity of a firm's governance program and its appetite for change, but the fundamental components that must be implemented remain the same, regardless.

Challenge

Data pervades every aspect of a firm's business ecosystem and is a strategic asset, hence it must be governed and managed well. Organic growth of data stores within corporations, introduces challenges with respect to multiple definitions for a data element, lack of consistent allowable values, multiple versions of the truth, inconsistent data types and precision, and lack of formality and standards for governing data. The increasing velocity of data, coupled with larger volumes and complexity (structured, unstructured, documents, email logs, social media feeds, etc.) add more challenges.

Four Fundamental Concepts of Governance

Four fundamental concepts must be understood and bought into by key stakeholders, before embarking on a governance program.

Data Ownership

The question of data ownership is critical to governance, since it deals directly with data accountability. When asked this question across teams within an organization, I've received many answers to this. Some of them are provided below:
• Data producers own data
• Data consumers own data
• Business units that typically manage it (e.g. product line) own data
• The owner of the system where this data resides (system-of-record or trusted source) owns the data
• The first system that receives the data and processes it owns it, and
• The owner of the corporate data warehouse owns it

None of these answers is correct. Data is neither owned by a single business area nor an individual system owner, but is an enterprise asset that is owned by the corporation. However, to govern and manage it appropriately, organizations identify and assign responsibilities to Data Trustees, Data Custodians and Data Stewards.

Accountability

Having established the fact that data is a strategic asset owned by the corporation, three roles are essential - Data Trustee, Data Custodian and Data Steward. Staff in these roles will play a vital role in day-to-day governance and will be accountable for its definition and quality, across the corporation’s information supply chain. Ideally, a Trustee should be at an Officer-level and is responsible for data definition, data quality requirements, data issue management, and ongoing monitoring of data for a certain business domain. Custodians (Officer or Director-level staff) play a similar role, but for operational data residing in operational data stores, data marts or warehouses and Stewards are subject matter experts in their respective data domains and support the business unit staff and the Trustees and Custodians in their day-to-day responsibilities. Stewards also play a critical role in data certification.

Organization

Data related issues arise during the course of regular business. They could be related to data definition, data consistency, data quality, alignment with industry standards or a disagreement between business units on their use. All such issues should be captured and escalated from the local level (business unit) to one or more enterprise governance bodies for resolution. These governance councils are cross-functional and typically composed of members of the business, operations and technology teams. The specific structure of the governance councils will vary, depending on the size, culture and preference of the organizations leadership.

Transparency

Complete transparency into data-quality and data-governance related Key Value Indicators (KVI) and Key Performance Indicators (KPI) across the information supply chain is essential for the success of a governance program. This has two major benefits - first, measuring quality and process efficiency enables organizations to find and address material weaknesses and second providing every stakeholder a view into such metrics gives them the ability to not only become aware of hotspots and operational issues, but to have a fact-based discussion about their impacts and resolution with other stakeholders. It is uncomfortable for Trustees, Custodians and Stewards to have their dirty laundry aired in public, so expect some push back. But if they view data as a corporate asset, appreciate their role in proactively managing it, then this should not be an issue.
The basics of governance

Having discussed the four fundamental aspects of governance, we have laid the foundation and can now discuss the mechanics.

What should be governed?

Enterprises consume and process thousands of data elements. However, there is a sub-set that is absolutely critical to support business operations, decision-making and reporting. Engage stakeholders in each business (product lines) and corporate support function (risk, audit, procurement, customer support, finance, accounting, corporate reporting, etc.) to identify key business functions and their associated data. Ideally, this should number in the four to five hundred. Focus must be on governing this set of enterprise critical data at the enterprise level and not to boil the ocean. Business units should set up their local governance structure and processes (mimicking the enterprise one) to govern their business-critical data.

What is the best model for governance?

This is an extremely important topic, since the success of the program hinges on this. Governance should not be perceived as a "big brother" or a top down program - since the doers within the business units resent this. A bottom up approach, with the business units driving governance doesn't work either, since business unit staff do not have a holistic view of data across the enterprise. In my experience, a combination of top-down and bottom-up approach works best. The term I have coined for this is Federated Data Governance - where an enterprise governance team facilitates the monitoring and management of the quality of enterprise critical data with help from Custodians and Trustees (top down) and the business unit Custodians and Trustees govern the business unit critical data in collaboration with the enterprise governance team (bottom up). Business unit critical data is promoted to the enterprise critical data domain, if there is consensus among all parties that it belongs there.

What aspects of data should be governed?

Data related issues mostly arise due to multiple and inconsistent data definitions (semantics) for a particular data element within data models or data stores, poor quality interface definitions that cause data corruption during data hand offs between systems, inconsistent data types, incomplete or inconsistent set of allowable values for a data element, lack of basic data quality rules that must be applied, and the inability to trace the data element from a downstream system to its source. There are many other aspects of data that can be governed, but I've highlighted the more important ones. These are non-trivial issues and need a concerted effort to align the entire company to a centralized repository of the single version of the truth.

How do I address data issues?

Proactive identification and management of data-related issues is required to lower systemic impacts. Each enterprise critical data element should be tagged with its system of record, trusted source, custodian, trustee, steward, line of business and other pertinent metadata, to facilitate root cause analysis and remediation of issues. Issues should be logged in an enterprise issue management system and assigned to the proper Trustee or Custodian, whose role is to triage the
issues, perform root cause analysis, assign it to the appropriate owner (data, process or technology) and ensure that it is resolved on time. The enterprise governance team should mine this data to find patterns of data anomalies, run some predictive analytics on the impact of such issues to down stream systems and provide aging reports to management.

Lessons from the Trenches

A summary of the lessons from the trenches is below:

- Focus on enterprise critical first, do not attempt to boil the ocean
- Educate senior leadership on the fact that governance is an ongoing process and doesn't have an expiration date
- Roll out enterprise-wide governance, modeling and data quality related policies and standards, measure compliance and take proactive steps to address non-compliance
- Use a carrot and stick approach to ensure compliance to corporate policies and standards. Tie compliance to annual performance reviews and reward staff members that display the desired behaviors. Penalize those that don't.
- Use a Federated Data Governance model. Top-down and bottom-up approaches to data governance do not work very well, especially in medium and large enterprises
- Build a team of Trustees, Custodians and Stewards that are passionate about data and assign them clear roles, responsibilities, authority, and accountability
- Define a repeatable process for identifying and addressing data issues and an escalation process for those issues that can not be resolved within the first-tier governance council
- Emphasize continuous improvement of data definitions, data quality and governance processes, throughout the company
- Educate employees and Board Members about the importance of good governance and the role they play in it
- Highlight the importance of treating data as a strategic corporate asset by every employee in the company
- Implement automated processes for managing and governing data

A paradigm shift with respect to organizational structure, accountability management, metrics management and execution strategies is required to address the best practices listed above. Executing against such a paradigm is extremely challenging - given constraints related to legacy systems, immature data management capabilities, silo-ed business models, silo-ed data management practices, lack of work flow between components of the information supply chain, internal politics and market pressures.

Conclusion

There are proven Enterprise Data Governance (EDG) and Holistic Data Quality Management (HDQ) frameworks, processes, methodologies, design patterns and disruptive technical solutions that can be applied to address the data governance challenges discussed in this paper. It is imperative that corporations invest in them, to improve their regulatory compliance and risk management functions and the overall health of their business critical data. These are strategic programs that require sponsorship, investment and on-going political support from the Board and C-level executives. The return on investment is
significant, if the programs are implemented in a systematic manner, with the proper change management processes and incentives built-in.

About the author: Jay Zaidi is an astute, hands-on, versatile and results-oriented leader with proven success in Enterprise Data Management, Strategic Planning, and Program Management. He is passionate about solving multidimensional problems. During his professional career, Jay has conceptualized and led business transformation and change management programs in the Financial Services business vertical. He has led global data management projects to address Regulatory Compliance, Risk Management and Operational challenges. He consults with and influences all levels of management and works to bridge gaps, facilitate communication and develop integrated business solutions. Proven success in strategic guidance to leaders in Fortune 100 firms. Jay can be reached at jayzaidi@gmail.com and his LinkedIn Profile can be viewed at http://www.linkedin.com/in/javedzaidi.