Confusing Data Quality with Data Governance

“Isn’t Data Quality (DQ) synonymous with Data Governance (DG)? Why does it even matter at the end of the day? “

Improved Data Quality is a key desired outcome from the implementation of Data Governance policies whereas Data Governance is the broader, strategic, enterprise vision of recognizing and managing data as a valued enterprise asset.

On the surface the two seem alike, are related and more often than not co-exist and certainly sound synonymous. One strong theme that is fairly evident is that the skill-sets required for any successful implementation of Data Quality or Data Governance initiative overlap and are very similar. Yet despite all the similarities and closely coupled existence there are some variations both subtle and distinct to classify them as their own independent and distinct domains.

Let’s start with their definitions:

**Data Quality (DQ)** is defined as the perception or assessment of fitness of data within a given context, meaning it is a measurement of data - the degree to which information consistently meets the expectations and requirements of the users/workers using it to perform their jobs. It is predominantly executed as an IT specialty and managed thru by those proficient in specialist data quality and data management tools.

**Data Governance (DG)** refers to the policies and procedures geared towards the overall management of usability, availability, integrity and security of data. In other words it is about a function owned by the business and executed by the business stewards in order to recognize and unlock the value of data to an enterprise and manage it as an enterprise asset.

If this is insufficient to make the case to separate them then let us examine their core concepts, competencies and dimensions.

Data Quality is measured against six dimensions in order to gauge its fitness:
- **Completeness:** Is there any data missing?
- **Consistency:** What data is giving conflicting information? Are the values that have the same meaning represented differently? i.e. “IL”, “Illinois”
- **Conformity:** What data is stored in non-standard formats. Do columns contain same data or different types of data
- **Accuracy:** What data does not reflect reality or is out of date. Sales shows 4 items in inventory but there is only 2 in the warehouse
- **Duplications:** What data or attributes are repeated. Two records for customer
- **Integrity:** What data is not referenced. Is all the data needed for best decision included

The key distinction here is that the quality of data exists solely in eyes the customer based on the value they perceive with respect to meeting their needs. Data not required to support any business processes, or required to make decisions, or useful in trend-analysis is irrelevant. Even if the data is accurate and has inherent quality, it has no value to the enterprise. The more data can be shared across the organization the more valuable it becomes, increasing the value of data within the enterprise, which is the primary goal for Data Governance.
Despite the distinctions, there are more synergies and relationships between the two as both work to increasing the usability and sharing of data. In order for any Governance or Quality initiative to succeed it must:

- always relate the scope to very specific and relevant business pain points or gaps in order to show the improvements or benefits to business
- start with a small independent area or project and expand from there based on progression and maturity
- have senior management or key business stakeholder championing the case and to ensure participation on front-lines of the data issues
- profile early and often: understanding the data is key to realize what data needs governance and what rules need to be applied to improve the quality
- extend metrics to measure the quality of data and how well it is being governed
- remediate compliance related issues immediately

The soft-skills really spell the difference between success and failure for both of these initiatives. These include the wider and clear understanding of the challenges faced by business and what are drivers as well as how governance and quality can improve or enhance it. Also key is having passion for it as well as leadership, diplomacy, tact, negotiation skills, communication skills, ability to simplify the problem.

**Laying the Foundation to unlock the business value**

*Everything begins and ends with Data.*

Data Quality can be approached independently on its own, in some cases all that is required is to add a data quality tool. Or alternatively it can be undertaken as the first precursor towards an MDM. Some organizations simply are not ready to implement Master Data Management or it is undertaken over long time frame like 2 to 5 years. In such situations they will unlock value by undertaking a shorter DQ effort. The other common use case is implementing DQ as part of Master Data Management project. Irrespective of the approach, improving quality in one area does not automatically make that data sharable across the enterprise. MDM is a first step towards making the data usable and sharable across the enterprise. Organizations are moving towards consolidation of environments, applications and data in order to:

- increase efficiencies of costs and operations
- reduce risks and non-compliance
- manage data as enterprise asset, enabling greater sharing of data across business operations by increasing the trustworthiness, security and auditability of the data
- improving the agility to enable timely, smarter business transformations and decisions
- improve customer satisfaction and retention

Unlike other assets such as money or equipment which are finite and are consumed as you use them, data assets become more valuable the more you use them. The moment that data is shared across business groups, it becomes an enterprise asset that must be governed in order to protect it and maximize its value to the organization as well as enable the organization to reduce costs in maintaining it. Data Quality initiatives have a greater impact and return when it is mandated through the Data Governance entity. Enterprise Risk Management and Compliance are becoming hot button items for every organization and are very data-centric and can only be managed thru an enterprise Governance program.

Governance is inherently an integration challenge and for it to be successful it must me more than a collection of data quality initiatives.
A key component for long-term success and efficiency is the establishment of Governance framework and an optimal governance model. The framework and model should clearly establish the operating model and process for how data should be governed and establish priorities linking scope to pain points.

While there are several variations of the governance structure/office, a common theme is to base it on the segregation of responsibilities and functions it is tasked to perform:

**Strategic: Define, Agree and Approve**

- DG Strategy, Vision, Mission, Objectives
- Budget and funding for DG initiatives
- DG Guiding Principles, Rules, Procedures and Definitions
- Roles & responsibilities of the various members
- Operating Model
- Policies and Processes
- Resolution of issues needing escalation
- Priorities, objectives and goals

**Tactical:**

- Execute to the established strategy/vision and priorities
- Ensure availability of infrastructure and tools
- Coordinate and integrate efforts with on-going projects and initiatives
- Manage scope, costs, risks and issues and escalate appropriately
- Monitor and report progress against established goals
- Review and discuss DG scorecards, concerns expressed by stewards
- Revise roles & responsibilities of the stewards as required
- Update Operating Model
- Discuss program and project data needs and requirements

**Execution:**

- Review, define and revise definitions, rules for governance
- Manage exception items through resolution
- Report on metrics and issue scorecards
- Escalate data concerns to the DG Office
- Raise data quality issues and suggest remediation
- Implement projects as defined by tactical component
- Educate developers, end-users, etc., on data standards and the importance of data quality
- Profile and Audit (sampling and monitoring) of data quality to ensure compliance against standards for both internal and external data
- Participate in system-related projects to ensure standards (data model, metadata, etc.,) are incorporated in development/enhancement
A sample Governance model may look something like this:
Governance is a journey and will take time, patience and lot of fortitude. Adoption of Data Governance progresses in varying stages. You start out small with an initial individual implementation and slowly build and extend to a program level implementation till you can finally coordinate implementations that are cross-functional in nature and across the enterprise.

Adoption of Governance is linked with the maturity of Governance within an organization. The following is an illustration of the maturity levels:

Adoption will typically occur in three distinct phases:

- **Explore Phase**: The initial phase where the best way to engage with governance is linking specific governance objectives with a project and then slowly expanding it to a Program level adoption. The maturity level for the organization is either “Non Existent” or “Reactive”.

- **Enhance Phase**: Here the organization has built some traction gaining confidence, repeatability and discipline; governance maturity is slowly tracking upward to Structured. The governance coverage is now extending from local and program level implementations to a business unit level implementation getting away from a standalone adoption process. Coverage also extends from focusing on data quality to including other functionalities and capabilities. I.e. masking data per security and compliance policies.

- **Transform Phase**: This is the optimal phase where the main-theme is enterprise-wide adoption indicating the organization is now reaching optimal levels of Governance maturity of “Proactive” and “Embedded”. At this point Governance is fully embedded within all corners of an organization.

---

**Figure 2: Data Governance Adoption Phases**

**Figure 3, Increasing Value of Data as the Information Management Capability Matures**
As noted in the above figures, governance may start out initially as Data Quality specific implementations but as it matures it will extend and enhance business functions and capabilities across the enterprise. An organization that reaches its target state maturity of level 5 “Embedded” will have addressed the following objectives and capabilities:

- Optimize or introduce functions and associated capabilities around:
  - Reference Data Management – framework enabling data sharing
  - Master Data Management – enables integrity and accuracy of data
  - Business Intelligence Management – drives actionable and intelligent decision-making
  - Metadata Management – enables consistent definitions and standards
  - Information Quality Management – accurate, consistent, conformed and measurable data
  - Information Lifecycle Management – process to manage creation and retirement data
  - Content Management – administration of digital content
  - Information Security Management – policies and procedures to safeguard data

- Develop target architectures for effective implementation of IT systems required to support the Information and Data Management functions

- Include points of integrations between application systems and the effective use of integration systems

Once an organization has reached its optimal maturity level, what’s next? Business is continuously evolving and changing. With every evolution and following best practices, organizations will need to review and revise the policies and procedures. The revised policies may call for enhancements or introduction of new capabilities. A fully mature “Embedded” governance organization will find greater success in introducing and implementing new functions while minimizing the risks and costs that typically derails such projects.

So does a focus on Data Quality confuse or complicate the message related to Data Governance. While Data Quality and Data Governance are closely interlinked and very tightly coupled, they are NOT quite the same thing. Data Governance has a far reaching enterprise objective. Data Quality programs are part of the overall Governance goals and objectives. DQ is implemented through technology supporting DG objectives for the enterprise.