

WHITE PAPER:

# THE SECRETS TO SUCCESSFUL DATA GOVERNANCE AND DATA QUALITY



## GETTING TO NOW

It seems like just yesterday that we would regularly encounter passionate data professionals who were frustrated trying to win support for data quality initiatives. They would quote industry figures about the business cost of poor data quality and show their own data profiling results to stakeholders, only to get blank stares or the “So what?” question. It seemed the business was more interested in speed than accuracy – there wasn’t a need to be right, just fast. Even if they did manage to win support for a data cleanup, stakeholders were frustrated months later when the same problems would crop up, requiring ongoing data cleansing.

Then came Data Governance (DG). In the early days, DG was often a thinly disguised rebranding of Data Quality (DQ). Voilà, a struggling DQ program was suddenly a DG program with a new lease on life. Vendors also had a new reason to sell their wares, having done what they could, given the challenges around selling DQ. Fortunately, real business needs have moved us past that rebranding.

Over time, people have begun to distinguish between DQ and DG, but not necessarily in the same way. Perspectives include the following:

**The Natural Evolution of DQ** – The broader scope of DG is the way they had wanted to address DQ all along, if they had the support. “Rebrand it as Data Governance, if that helps”, say the DQ veterans.

**The DG Umbrella** – Data Governance is a broader capability which includes DQ management, among other things.

**Partners** – DG and DQ are separate, but complementary disciplines –either can be done on its own, but collaborate if doing both.

Garbage-In, Garbage-Out (GIGO) is still a truism, but the stakes are now much higher in terms of potential impact to the organization. After all, when people were managing more by the seat of their pants, the real impact of misunderstood, misused, poor quality data was minimal.

Today, changes across industries have put an increased focus on DG and DQ due to the explosion of big data (i.e. the data swamp), the promise of valuable insights from advanced analytics, organizations going through digital transformations, data privacy concerns and increased regulation/penalties. Fortunately, it is also now cool to actually manage with data, versus just pretending to do so.

Unfortunately, we still regularly encounter passionate data professionals who are frustrated trying to win support for, and deliver value from, DG and/or DQ initiatives. This white paper illustrates the power of a business value-driven approach for Data Governance and Data Quality, especially when they are joined in a happy union.



## PARTNERS TOGETHER IN DATA VALUE OPTIMIZATION

### **The Common Goal of Data Value Optimization**

Regardless of how we distinguish between DG and DQ, there is a unifying goal – to enhance the business value of the data. Although more granular DQ and DG metrics abound, at the end of the day it's about using better data for increased business value. To optimize data value, we can lower data costs and/or increase data worth.

We want to minimize costs wherever we can, including those associated with:

- Poor data quality
- Misunderstood data
- Regulatory and security risks/violations
- Lack of data lineage
- Redundant data-related work, including “fixing” DQ issues
- Inappropriate (and sometimes expensive) data retention and storage

We want to optimize data worth by:

- Ensuring data quality
- Creating shared understanding of data
- Ensuring accessibility and timeliness of data
- Establishing trust in data
- Making it easier to locate the right data

A key stumbling block for many is that traditional costs in dealing with data seem so much more quantifiable (2 TB of tier one storage cost \$2,000/year) than quantifying the worth of it. People who treat data in the abstract fail to recognize that data has no value on its own - only from what they do with it. We'll talk more about how to address this, later.

### **What the Partners Bring to the Union**

When we recognize that DQ and DG are coming together for the common purpose of Data Value Optimization, we don't need long arguments over which responsibilities lie where. The truth is that each organization handles this differently.

We also know that some may disagree with the characterizations of each. The important thing is to cover the superset of capabilities between DG and DQ. For purposes of our discussion, we'll characterize them as follows.

<b>DATA QUALITY</b>	<b>DATA GOVERNANCE</b>
Data Discovery & Profiling	Data Accountability & Stewardship
Data Cleansing	Data Definition & Understanding
Data Quality Monitoring	Orchestration of Data Stakeholders
Reactive	Proactive
IT Focus	Business Focus

These are generalizations, but the characteristics are not value statements, rather a reflection of what we have seen in this space across many organizations.

More about DQ and DG are described below, but it is worth making a few quick notes about the last two items in each column in the table.

**Reactive and Proactive** – Being reactive is not a bad thing. If there is a problematic shop floor process churning out flawed data and IT can't get support for the business process improvements, isn't a DQ job to cleanse the data to maintain optimum inventory reorder points worthwhile? Oftentimes, DQ working out of IT meant the best they could do was react to, and treat the symptoms. Data Governance increases awareness within the business about their role and accountability in generating bad data and promotes more proactive practices, aimed at preventing DQ issues from arising in the first place. Clearly uniting both approaches benefits organizations in the very messy real world where not all issues can be addressed proactively.

**IT and Business Focus** – DG helps the business understand that it is primarily the business processes, not IT, that generate bad data. IT tends to have a better detailed understanding of the technical aspects of data and what happens to it as it is stored in, and transferred between, applications and data stores. The business has a better understanding of the process that generates it and how it is used. Unfortunately, communication issues can arise between the two perspectives. DG and DQ can focus both on developing a common context and clear definitions of business terms to establish a common language around data. Collaboration improves as IT and the business understand their respective roles in optimizing the business value of data.



## FINDING A STARTING POINT

We are often asked, “Where should an organization start – with Data Quality or Data Governance?” Frequently, our answer is to start where the organization is today. For example, if there is an effective and respected DQ capability in place, start to leverage the DQ insights to craft messages about the need for complementary DG to scale efforts while keeping costs contained. If, on the other hand, DG has been initiated, but there are missing, or inadequate DQ management capabilities in place, leverage the DG strength to show the need for a more formal DQ management capability.

If neither DG or DQ exist, consider the nature of the current pain. Can management attention be leveraged around that pain to win support for DG or DQ? For instance, if there is a need for a new ERP and related data migration, an initial DQ focus may make more sense. However, if regulations (e.g. SOX compliance, CCAR, etc.) are mandated or a cause for concern, starting with a DG focus may be the right choice.

If executive leadership is trying to be proactive, or if the general need is recognized, we recommend leading with Data Governance, with a plan to launch DQ management early in the formation of this initiative. There are several reasons for this, including the following:

- It's difficult to talk about the quality of specific data when there's no agreement on common definitions. DG will address business definitions for the data. In one instance, we were brought in to deal with three different business functions arguing about the quality of a new customer metric – each department reported a different number, almost every month. It didn't take long to realize that there was no common definition of what constitutes a “new customer”.
- DQ is more easily addressed when Subject Matter Experts (SMEs) are identified to help understand the data and define business rules. DG establishes data stewardship and stakeholders. Too often we see SMEs engaged to shed light on DQ issues based on who is willing and available, rather than who the true expert(s) is/are.
- Data Governance typically provides context, perspective and priority to Data Quality issues that would otherwise languish, unresolved.

Be careful not to overplay one's hand, it's better to get a smaller win than no win at all. Think of it as a wedge to open the door to develop the capabilities ultimately needed.

Of course, there is always the question of where these DQ and DG capabilities should be in the organization structure.

- If neither are established, an ideal world would establish DQ as part of DG or as sister organizations reporting to the same leader.
- Best practice is going to locate DG in the business, as opposed to within IT, to help overcome legacy perceptions that data is an IT issue.
- DQ, if already established in IT, can remain there given the technical skills typically required for data quality management.

Ultimately, organizational location is less important than effective collaboration between DG and DQ, which can sometimes blaze the trail for more effective collaboration between the business and IT overall. Always bear in mind political realities and culture in the organization when considering where in the structure they should go.



# HOW TO WORK TOGETHER FOR DATA VALUE OPTIMIZATION

## Guiding Principles

Our experience in working with data in different organizations and industries has shown wide variances in culture, business drivers, data maturity, data architectures, and organizational structures and capabilities. To address those differences, it is necessary to tailor the approach to DG and DQ to one's organization. There are enduring guiding principles that will direct the approach for successful DG and DQ, including:

**Base on Business Value** – Focus on optimizing business value through coordinated pragmatic action. Historically, DQ efforts often struggled to connect with the business on the impact of the DQ issues they discovered – data profiling results were clear, but advocates struggled to answer the “So what?” question in business terms. Talk to the business in their language and about things they care about thus ensuring data plays a role. Even simple things like referring to business rules can make a difference.

**Leverage Data as Lifeblood** – Don't try to separate data from the business – treat the whole patient. On its own, data has no value – the value comes from what we do with it. Because data flows through every process in the organization and involves people and (usually) technology, it is unlikely that lasting business value can be delivered just from a single focus. For instance, on their own, updating values in a database or defining a business term won't deliver business value. Focus DG and DQ efforts as enabling capabilities in more comprehensive solutions (people, process, technology) for business problems.

**Orchestrate the Organization** – Coordinate the work that isn't necessarily called DG or DQ in the organization to optimize value. Most of the orchestration role usually falls to DG, but it is not unusual to find many time-consuming and redundant processes that exist to address DQ issues. Different parts of the organization often have their own understandings of, and perspectives on, the same data – no one person can be as expert as the larger community when they start communicating with each other. Furthermore, what may be an acceptable standard of DQ for one group may fall far short of what is required by another.

**Build for (Better) BAU** – Plan from the beginning to make DG and DQ part of Business-as-Usual (BAU), not an oppressive bureaucratic overlay but rather a partner and advocate. Plan DG changes that will reduce the effort required for DQ cleanup. Focus on tweaking existing business processes for better data outcomes rather than creating a new process or policy.

## Practical Approaches for Success

The following approaches to DG and DQ have been proven effective and are presented at a high level below. It is beyond the scope of this white paper to detail both approaches, but the examples of the interaction between DQ and DG are designed to be instructive.

### DATA GOVERNANCE APPROACH OVERVIEW

Data Governance goes beyond determining who gets to make what decisions about what data. DG is responsible for bringing a data perspective to the table where often only people, process, and technology are discussed. DG establishes the process and structure to support those who make the decisions about data. DG also facilitates and coordinates the people, processes, and technologies. This is critical because the same data often cuts across departments, functions and applications. With that perspective represented, and leveraging the guiding principles, DG is foundational in delivering business value. Data Governance is not a project nor a program, but rather an ongoing function within the organization. Certainly, there may be a project to initiate formal DG, but once underway it is no more a program than the accounting function is a program. The approach to establishing successful Data Governance follows this path.



**1. Assess & Plan** – We consider over two dozen factors to determine how to best implement DG in the given environment and identify the best opportunities to deliver business value through business projects. We tailor the approach and roadmap to leverage organizational strengths.

**2. Orchestrate Implementation** – We focus on the execution of the roadmap, following the guiding principles. The core DG function is organized to orchestrate better data outcomes, through the targeted business projects, across the enterprise with a heavy focus on building partnerships, education, and communication.

**3. Business as Usual (BAU)** – The focus shifts from building to maintaining, by reinforcing monitoring, controls, and incentives to accommodate an evolving business and preserve positive changes, producing better data outcomes.

Of course, different parts of the organization are not likely to be at the same places in their Data Governance journey. For instance, it is not unusual for the Finance function to be more advanced in this area than the Sales function.

## DATA QUALITY APPROACH OVERVIEW

A mature DQ approach evaluates data in operational and decision-making contexts to identify the value of the data and focus on what matters most. It should set the stage for IT and the business to continuously collaborate on data improvement. Like DG, it is an ongoing function and should be able to quickly adapt to the changing needs of the business, such as the inclusion of external datasets or big data based on this approach.



**1. Discovery** - We analyze both the systems and how the data is being used in business operations and decision-making. This is where data value and the impact of Data Quality issues are realized.

**2. Profile** - With the proper context, we conduct focused, advanced analysis of known and unknown data issues. Working with the business, we discover and prioritize issues based on both data metrics and business impact.

**3. Design** - We work with business and IT to design solutions and define business rules for data cleansing and monitoring. We establish clear, ongoing responsibilities for business and IT stakeholders.

**4. Implement** - We iteratively implement the Data Quality solution by focusing on creating reusable objects and automating processes to validate, cleanse, standardize, and de-duplicate data.

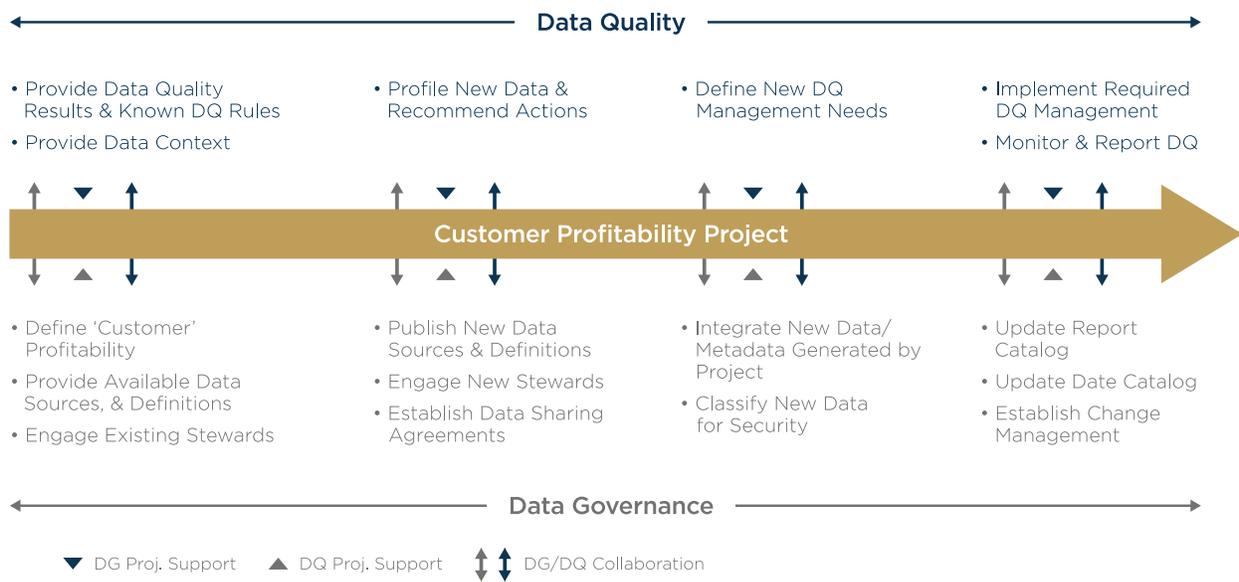
**5. Monitor** - To accommodate a dynamic business environment, we implement monitoring to ensure appropriate Data Quality is maintained for business operations and decision-making, enabling organizations to take corrective action when needed.

## A Happy Union Requires Give and Take

As noted earlier, it's important not to distinguish between DG and DQ in an organization. However, if done so, and the characterizations provided in this white paper are accepted, this example should provide further insight on the collaboration between DG and DQ. We fully acknowledge that other parties would be involved in the scenario, but we assume organizationally distinct DQ and DG groups and focus on them and their interaction. The scenario is kept relatively simple for the purposes of this white paper.

In this example, the organization conducts a Customer Profitability project. The objective is to increase average profit margin per customer by two percent by the end of the following fiscal year. Leadership would also like the ability to forecast overall customer profitability. Hence the scope of the project covers not only the analytics team and the ability to measure customer margin, but the people, process, and technology to make necessary changes to realize the objective. Initial stakeholders include Pricing, Marketing, Customer Support, Accounts Receivable, and Product Management. The list of stakeholders may expand, depending on the data and the solutions designed.

Much of the data is available within the organization, but there is often the need to acquire external data as well. Some, but not all, of the needed data is already under DQ or DG management. The DQ and DG functions have been formally established in the organization for about 18 months when the project starts.



Following the Guiding Principles mentioned earlier, the focus for both DG and DQ is on the project. It is the project, not DQ or DG alone, that has the requisite scope and commitment to deliver business value to the organization. The DG and DQ groups also keep in regular contact and collaborates, but the project focus helps ensure they concentrate on tasks that are most useful to the larger organization.

As we follow the project, we see the interaction of each with the project, as well as the collaboration between DQ and DG.

STAGE	PROJECT INTERACTION	DQ/DG COLLABORATION
A	<p>DG and DQ are bringing available existing artifacts to the project. The project can get off to a faster start, with a better understanding of the data, where to find it, who the experts are, and its peculiarities. Information about earlier similar efforts may also be shared. A data scientist may plan to source some data from the data warehouse, but once she sees the DQ rules and filtering applied may decide sourcing from elsewhere is better. There is an opportunity for the team to manage expectations as they better understand the data – ideally, before firm commitments are made!</p>	<p>DG and DQ should already be in sync on this, but the project provides an opportunity to confirm that they each are working with the most current information. Each should assign someone to be a point of contact for the project and to coordinate directly with their counterpart.</p>
B	<p>In collaboration with the DG team, the project team identifies the full set of data and prioritizes it in terms of its importance to the project. Ideally, the DG and DQ teams can work ahead of the project team so they have a better understanding of the data when they start work on it. DG also works out any necessary data sharing agreements to ensure the project team can use the data they need.</p>	<p>With an understanding of the full scope of the data for the project, DG and DQ focus on the priority data they do not yet have under management. DG provides information on where data should be sourced from to guide the data profiling efforts. Data profiling results confirm or redirect data sourcing decisions. DG engages Data Stewards as needed who can be a resource to the DQ team. Both collaborate on recommendations back to the project team on how best to use data.</p>
C	<p>The DG team has worked with Data Stewards to create business definitions and other key metadata for the high priority data and provided it to the project team. The project team may need to make some course corrections based on DG/DQ recommendations as they better understand the realities of the data. With the design coalescing, DG facilitates the engagement of the information security, legal and regulatory teams to provide appropriate information back to the project team, usually through the appropriate Data Stewards, to avoid last minute project roadblocks.</p>	<p>The DQ team looks at possible revisions to existing DQ management processes and how best to operationalize them and any new requirements. They coordinate with the DG team to vet the changes and ensure there are no inadvertent impacts to data stakeholders beyond the project. Both collaborate on updating existing glossaries and metadata with the new understanding gained from the project and coordinate their actions to ensure accuracy and completeness.</p>
D	<p>DG and DQ help the project team focus on creating proactive processes and guidelines for the ongoing operation of customer profitability analytics to ensure better data outcomes. Effectively, they are engineering good data practice into the new process. For instance, DQ will help them establish appropriate courses of action if DQ monitoring thresholds are exceeded and DG will help establish model management processes for the data science team. If appropriate, additional Data Stewards may be engaged.</p>	<p>Both collaborate and check each other to ensure the new process data is properly documented in the business glossary, business rules, procedures, data lineage, and other metadata. They coordinate to make sure the customer profitability team has one clear, easy set of controls to make sure the quality of the results meet expectations. As operations are in production, they continue to coordinate and check in to make sure nothing is missed.</p>

Although this example is more than a systems development project, it is worth noting that the collaborative approach shown here works regardless of the systems development methodology. Agile, waterfall, and everything in between are supported. Of course, adjustments should be made to ensure a positive working relationship with the project team. In a traditional waterfall approach, much of the work through stage C may be done during requirements and preliminary design. In a more iterative, agile approach, smaller versions of the DG and DQ work would be repeated for each sprint and building a backlog of data information for development teams, and focusing on stage D for the sprints bundled into a production release.

Naturally, each project evolves the DQ and DG capabilities and adds to the data under management, creating a continuous cycle. Over time, a majority of the data needed for a project will be understood up front. Of course, this reduces project cycle times and improves the quality of the deliverables. The DG and DQ teams can focus on any changes to data and keeping all information up-to-date. An additional benefit of this maturity of the capabilities is that the organization can respond more quickly to changing needs whether they be a change in operations, deployment of new enabling technology, or changes to products and services.



## ENABLING TECHNOLOGY FOR DQ / DG COLLABORATION

As described, DG and DQ need to be driven by clear, value-driven objectives, but there is a role for enabling technology to support the collaboration. Industry measures show about half the DQ management work is done manually, as opposed to leveraging automation. In some cases, like processing exceptions from a customer de-duplication algorithm, there is no substitute (yet) for human engagement. However, organizations can improve efficiency and effectiveness with more automation and should pursue automation where practical. DQ tools should support work of the DQ team including data profiling, definition and application of DQ rules, and monitoring of DQ metrics. In addition, some vendors are starting to embed artificial intelligence in their data quality software to improve accuracy and efficiency even further.

Likewise, DG enabling technology is useful to catalog, author, and manage the wide variety of metadata, including DQ metadata. To scale, good DG tools change from a convenience to a necessity. DG tools should support the DG mission through strong business glossary capabilities, support for data stewardship, automated integration & management of metadata, and the ability to author and customize objects and create relationships between them. Automated workflow capabilities are also desirable, including the ability to manage workflow with the DQ team.

Flexibility and ease of customization of the tool are key. Clearly, organizations are best served by tools that support an integrated perspective for DG and DQ. This doesn't mean they need to be part of the same vendor stack, but they should be integrated in a way that promotes collaboration and makes the data and metadata understandable and approachable by business and IT users. For example, a user looking to assess the suitability of an existing report should be able to easily navigate from information about the report to information about specific data elements in one place, including the following:

- Business Definition
- Security Parameters
- Data Quality (across DQ dimensions)
- Source (data lineage)
- DQ Rules / Filters applied
- Meaning of Standard Code Values (e.g. NW = Northwest Region)
- Data Steward(s)

When considering tool choices, keep in mind the guiding principle of Build for (Better) BAU. Experience has shown that people are seldom willing to add another application to the suite they already must work with to do their job. A key consideration when choosing a tool is the ability for the tools selected to be seamlessly integrated with existing processes. Of course, different departments have different systems and processes. Consider the following:

**Workflow & Communication** - If Data Stewards are required to log into the DG application daily to see DQ monitoring results, or what actions they have, then they likely won't comply and login. However, if the DG application can leverage automated workflow to send something to them by email and they can complete their action by interacting with that email, the chances of success are much better. Likewise, if the organizational culture eschews email in favor of a corporate workplace collaboration platform, look for integration opportunities there.

**Accessibility** - Consider what can be done to make it easy to leverage the knowledge and insight created by the DG / DQ efforts. For instance, having meaningful business definitions for data can be tremendously powerful, but if the only way to access them is to log in to a rarely accessed DG application people are less likely to leverage it. Instead, think about the different entry points people could have for querying the metadata and integrate with those entry points. For example, can hyperlinks in reports be created or intranet sites to the business glossary? Doing this will promote real world alignment with the standards defined through DG and DQ, because it is easier for stakeholders to be more data savvy.

## CONCLUSION

If optimizing the value of data is important to one's organization, establish a happy, collaborative union between Data Quality and Data Governance. Each brings complementary focus and skills to move the organization towards Data Value Optimization. Every organization may start in a different place, have different drivers and require tailoring of their approach to DG / DQ - but all can benefit from these guiding principles.



Base on  
Business Value



Leverage  
Data as Lifeblood



Orchestrate  
the Organization



Build for  
Business as Usual

As these Data Governance and Data Quality capabilities scale and become part of the organization's data management foundation, incorporating enabling technologies to promote and improve collaboration and stakeholder engagement will be the next step along the Data Value Optimization roadmap.

Data Governance and Data Quality are important considerations in the ongoing development and enhancement of an organization's data management strategy. The successful union of these two capabilities will result in higher quality data that's well understood across the organization, thus enabling a higher confidence level in the ability to leverage this data for key decisions. Data Governance and Data Quality: the perfect union!