G00259288

Twelve Ways to Improve Your Data Quality

Published: 17 February 2014

Analyst(s): Saul Judah, Ted Friedman

Achieving business outcomes based on trusted, high-quality data is a core enterprise requirement. Unless CIOs, chief data officers and information leaders get this right by pragmatically improving their data quality, they will be unable to take full advantage of new information-driven opportunities.

Key Challenges

- There is a widespread failure in IT departments to focus on the fidelity of information, and this impacts business value. Where data quality improvements occur, they tend to be disconnected with how they impact the business's P&L.
- CIOs, chief data officers and information leaders continue to struggle with getting data quality onto their business agendas. This is due to an overemphasis on technology rather than a focus on organizational culture, people and processes.
- Few organizations attempt to use a consistent, common language for understanding business data quality. Instead, they maintain divergent and often-conflicting definitions of the same data.
- IT leaders struggle to make data quality improvements beyond the level of a project and do not embed them as part of their business information culture.

Recommendations

- Evaluate the 12 improvement actions in this document, and assess the extent to which you have addressed them and the value they could deliver.
- Identify what actions you should take and construct a plan for moving forward with the areas that will bring the quickest and greatest benefits. Do this after your evaluation and assessment.
- Assess how this plan can be used to supplement ongoing data quality initiatives where they exist and as the basis for a data quality improvement proposal where they do not.

Strategic Planning Assumption

By 2017, 33% of the largest global companies will experience an information crisis due to their inability to adequately value, govern and trust their enterprise information.

Introduction

Based on research carried out by Gartner, we anticipate that organizations without the right information governance, processes and controls for information management (IM) will experience an information crisis.[1] Data quality plays a key role in this. Therefore, it is essential that your enterprise information management strategy is:

- Fit for purpose.
- Forward-looking in its ability to support future business requirements and risks.
- Established on solid information foundations of known data quality metrics.

Organizations of any significant size and complexity — with multiple business units (BUs), customers, employees, suppliers and products — will inevitably face data quality issues. The more mature organizations accept that this is something which will always be the case as long as they are in business. Not only do BUs reorganize, but processes, products and services evolve, mergers and acquisitions occur, and business rules change within organizations. This also applies to the customers, suppliers and regulators that organizations engage with. Furthermore, the daily transactions that are executed based on enterprise information add to the volume and complexity of data quality issues within the organization. Unless tangible and pragmatic steps are taken by the organization to understand, address and control data quality, the situation will worsen.

Analysis

Twelve actionable opportunities for improving data quality in a practical and pragmatic manner are discussed here. Some organizations may already have well-established practices, in which case this research could be used to benchmark current practices and address existing gaps. Other organizations may want to improve their data quality either through a program of coherent activity or with more immediate, tactical steps. In this situation this research should be used as an input into the data quality strategy where one is being developed, or as tactical improvement steps that can be taken until funding can support a strategic data quality program. These steps should be used as the basis for action when business outcomes are impacted as a result of poor data quality. However, regardless of their information maturity level, organizations can:

- Improve the current level of their data quality by focusing on the things that bring business benefits.
- Ensure they understand what fit-for-purpose data quality means to them.
- Assign business accountability for data quality.

- Establish a continuous improvement mindset.

Focus on the Right Things

Senior business leaders — and their direct reports — deeply understand their business objectives, business area profit and loss (P&L), key business processes, and the key performance indicators (KPIs) and key risk indicators (KRIs) that underpin them (see Table 1). Data issues that negatively impact their key priorities are therefore of most interest. As a result, it is crucial that you understand which data quality improvements will have the greatest impact on business objectives, and use this as the basis for dialogue.

Table 1. Actions 1 to 3 — Focus on the Right Things

Action	Inputs	Processes	Outputs
1. Establish a clear line of sight between the KPI/KRI impact of data and data quality improvement.	▪ Identification of the KPIs and KRIs used in each BU, and identification of their composition and value. ▪ Assessment of the business requirements and associated data used for meeting these requirements that contribute to the BU P&L, and their KPIs and KRIs. ▪ Identification of the regulatory and compliance requirements that require data quality improvement.	▪ Identifying the application systems that produce data to support business activities which impact KPIs/KRIs. ▪ Running DQ metrics against data-impacting KPIs/KRIs for each BU. ▪ Assessing the current business value in terms of the existing DQ level, and assessment of the business value for improving specific DQ items.	▪ Investment case for targeted DQ improvement, managed through IG.
2. Use data profiling early and often.	▪ The business requirement for the leveraging of data for specific results. ▪ Data integration/ migration efforts. ▪ Data scope and business rules.	▪ Using built/bought technologies to define data queries. ▪ Applying queries to key data sources and repositories. ▪ Verifying and validating results.	▪ Profiling results presented in a way that humans can understand and act upon. ▪ Analysis of profiling results against applied business rules. ▪ Balanced business and IT decisions.
3. Design and implement DQ dashboards for critical information such as master data.	▪ Requirements for the DQ dashboard to be aligned with the IG scope. ▪ Assessment of technologies that are suitable for presentation.	▪ Putting into production the DQ dashboard and embed as a BAU IT process. ▪ Identifying the quantitative/qualitative mix and responsibilities.	▪ Production packs delivered ahead of IG steering groups. ▪ Leveraging of the same modular DQ pack for different audiences. ▪ Initial foray into information

	▪ PoC dashboards and metrics that have been agreed with key stakeholders.	▪ Delivering the DQ dashboard as a modular pack.	stewardship applications.

BAU = business as usual; BU = business unit; DQ = data quality; IG = information governance; KPI = key performance indicator; KRI = key risk indicator; P&L = profit and loss; PoC = proof of concept

Source: Gartner (February 2014)

As a first step, the relationship between the P&L, business requirements, business processes, KPIs/KRIs and key data types — for example, master data and transactional data — must be understood and modeled (see "The Gartner Business Value Model: A Framework for Measuring Business Performance"). For example, an organization seeking to grow revenue by 5% — that can be attributable to customer spend — could have a key business requirement for reducing high-value customer attrition by 15% over 18 months. The relationships between customer attrition, KPIs and leading KRIs may point to high volumes of customer complaints and poor customer service due to incomplete and inaccurate data being a key cause.

The next steps are to profile the current level of data quality within and across the applications that support the key business processes. This requires the application of data profiling tools — if you have them — or the creation of scripts that run against the data repositories if you don't. Good data profiling tools anticipate the type of data profiling that is needed and offer basic out-of-the-box functionality. This functionality offers insight into the current levels of data quality within a very short period of time. It also includes graphical or programmatic tools that configure more complex profiling requirements (see "Magic Quadrant for Data Quality Tools"). If vendor profiling tools are unavailable — or if there is no funding yet in place — it is still relatively easy to profile your data by establishing some of the simpler data quality rules and running scripts (see "Toolkit: Assessing Key Data Quality Dimensions"). However, the high level of complexity in cross-application data quality analysis will challenge even the more experienced IT practitioners if the right tools are not available to them.

Profiling is not, however, something that is done once and then forgotten about — it is an activity that should be done often. For example, profiling could reveal that some critical customer contact information is missing. This may have directly contributed to a high volume of customer complaints, and would make good customer service difficult. Data quality improvement in this context now becomes a high-priority activity and, as a result, gaining support for an improvement program should become much easier.

Eventually, the business will need data quality dashboards to demonstrate the performance over time of data that is critical for key business processes. This is so that the business can make the right business decisions to achieve the desired business objectives based on trusted quality data. Where data quality improvement activity is undertaken and operational business processes take account of new data practices, the impact on key data is reflected in data quality dashboards. These dashboards can then be used by information governance bodies to understand the impact of their decisions where further action is required.

Define "Fit-for-Purpose" Data Quality and How to Achieve It

To improve data quality it is imperative to understand, agree, document and communicate what is meant by "fit-for-purpose" (see Table 2). Fit-for-purpose data quality will ensure that the aspects of data quality improvement which yield the desired business results are the focus of investment.

Table 2. Actions 4 to 6 — Defining Fit-for-Purpose Data Quality and How to Achieve It

Action	Inputs	Processes	Outputs
4. Clearly define what is meant by "good enough" data quality.	▪ DQ best practice guidance and DQ framework for analysis. ▪ A list of existing DQ issues and constraints that have been encountered. ▪ Business process KPIs/KRIs.	▪ Defining common data and variances. ▪ Applying common data quality rules, including thresholds. ▪ Applying a common DQ framework to ensure a consistent application of DQ dimensions across BUs.	▪ Improved DQ standard and an improved ability to consistently and appropriately apply it across the enterprise. ▪ Common DQ language, better communication of issues, and less confusion and better positioning of governance.
5. Establish a DQ standard across the organization.	▪ A list of existing DQ standards (where they exist), including information about how they are used and by whom. ▪ A reporting structure for existing DQ standards, including operational risk and business performance information.	▪ Analyzing scope, overlaps and gaps in DQ standards. ▪ Defining a common DQ standard.	▪ Consistency and appropriateness in the way key enterprise data is applied and reported across the organization. ▪ A DQ standard that is an active instrument of IG.
6. Move from a truth-based semantic model to a trust-based semantic model.*	▪ An understanding of what information is created inside the organization and what is created outside the organization. ▪ An understanding of the key enterprise data using the current data model, and of how and where key data is used.	▪ Extending the data quality strategy by adding an information trust model. ▪ Extending the enterprise data model with trust "characteristics."	▪ Investment in DQ improvements targeted at critical data areas. ▪ A review of information governance policies and their implementation across key business processes.

* For more information on "trust models" see "Certify Data to Foster Trust and Consistent Use."
BU = business unit; DQ = data quality; IG = information governance

Source: Gartner (February 2014)

When data quality rules have been agreed upon and set up, a consistent, minimum data quality standard must be then be established. This will enable BUs across the enterprise to understand and execute their business operations in accordance with the defined and agreed standard. Because it is likely that different BUs in an enterprise will have different levels of business sensitivity, culture and maturity, the manner and speed in which requirements of data quality standards are met may differ. For example, an organization may include, in its definition of its data quality standard requirements, the need for:

- Completing data quality training.
- Establishing a governance body for data quality.
- Creating specific data quality control metrics that feed into a dashboard.

However, the thresholds that apply to each BU may vary, depending on the levels of commercial sensitivity and organizational maturity. This is so that the more sensitive and mature parts of the organization can work toward — and achieve compliance with — the data quality standard earlier than the less sensitive or mature areas.

Furthermore, since information is also acquired from sources outside the organization — where the data quality rules, authorship and levels of governance are often unknown — organizations must move toward a "trust model" (see "Certify Data to Foster Trust and Consistent Use"). This means that rather than thinking about key enterprise data as being absolute, its origin, jurisdiction and governance — and therefore the degree to which it can be used in decision making — must also be considered. Business operations and IT departments using information based on such a trust model are better able to use data lineage and context as the basis for their actions.

Assign Business Accountability for Data Quality

Although IT has a very large part to play in data quality improvement, it is usually business users within a business processes context who are the authors and primary consumers of data. It is therefore essential that business stakeholders recognize and accept that they are accountable for the quality of data they create (see Table 3). However, it is the responsibility of the IT department to clarify which data provided to the business stakeholders was created inside the organization, and which was created outside. This is so that the appropriate judgment can be applied by business consumers about its use.

Table 3. Actions 7 to 9 — Moving Accountability for Data Quality Into the Business

Action	Inputs	Processes	Outputs
7. Make data quality an agenda item at the IG steering group meetings.	▪ Definitions of the business value of data in terms of business objectives. ▪ Backing agreements from the IG program sponsor. ▪ Presentation of the PoC to the IG steering group, to establish a case for DQ.	▪ Modifying the IG steering group's agenda. ▪ Assigning responsibilities and expectations for DQ. ▪ Identifying the right decisions based on the presented data.	▪ DQ dashboard requirements. ▪ Correct identification of the stakeholders and objectives/key indicators that are impacted. ▪ DQ improvement road map that is aligned with the IG vision and strategy.
8. Establish data quality responsibilities as part of the data steward role.	▪ IG scope and stakeholder map. ▪ Business area process/data targets and investment plan. ▪ Existing data steward role profiles and best practices.	▪ Reviewing how data quality issues are managed. ▪ Analyzing any issues with the current approach, and what the impact is on the current investment plan and performance targets. ▪ Identifying what action is not being taken where it is required.	▪ Correct identification of the issues caused by gaps in DQ management that impact business area objectives. ▪ Review of recommended changes to the data steward role and inclusion in appropriate performance objectives. ▪ RASCI matrix.
9. Establish a cross-BU and IT special interest group for data quality.	▪ Identification of common data quality issues across business areas. ▪ Identification of individuals within the business and IT departments who can benefit from becoming more	▪ Proposing an initiative and gaining support from the IG steering group or senior sponsor, and communicating your intention. ▪ Setting up a regular forum with those impacted using the	▪ Understanding of common DQ issues, and shared techniques/models/ dashboards that have been successful. ▪ Forum for identifying new requirements and opportunities to improve business data quality.

	engaged in addressing data quality issues.	appropriate external input.	

BU = business unit; DQ = data quality; IG = information governance; PoC = proof of concept; RASCI = responsible, accountable, supporting, consulted and informed

Source: Gartner (February 2014)

Earlier in this document, we discussed how creating a causal link between business outcomes and the quality of data used to achieve those outcomes was of primary interest to senior business stakeholders. Having understood this connection, the same stakeholders still need to track their investments in data quality improvement alongside the achievement of business objectives. Making data quality an important agenda item with the information governance steering group will provide a level of visibility that the data quality improvement program needs. For example, if an information governance group meets every quarter, it will benefit from reviewing a data quality dashboard that charts the performance of data quality against KPIs, KRIs and key business operational processes. Where issues are identified, senior business stakeholders can direct the corrective action taken within their BUs. Increasingly, the business role of data steward is being established in more mature organizations to champion good IM practice and actively monitor, control and escalate data quality issues where they occur. In some cases organizations are taking this further by enabling data stewards to formally collaborate across BUs through special interest groups for data quality and master data management (MDM). This enables better organizational management of information risk. It also creates more opportunities for reducing operational cost, and encourages growth through shared and consistent best practices.

Embed Data Quality Improvement as Part of the Information Ecosystem

For data quality improvement to become a core part of the organization, steps must be taken that will embed data quality and its management into the organizational consciousness of IT infrastructure, enterprise architecture, business operations, risk management and internal audit stakeholders (see Table 4). It is likely that such opportunities to improve data quality maturity are achievable in organizations that have already reached a greater level of maturity than those at less-established stages of development.

Table 4. Actions 10 to 12 — Embed Data Quality Improvement as Part of the Information Ecosystem

Action	Inputs	Processes	Outputs
10. Establish a DQ review as a release management "stage-gate" review process.*	▪ Existing program governance model and new product approval frameworks. ▪ Current terms of reference and a quality review process. ▪ Approved DQ standard and thresholds for business departments.	▪ Modifying the relevant quality review process to include the DQ impact on business value and risk. ▪ Exploring the extension of existing processes; for example, including processes within data privacy checkpoints.	▪ Revised release management "stage-gate" process, with a DQ impact assessment and highlighted issues. ▪ Continuous DQ improvement as part of the BAU process.
11. Communicate the benefits of better data quality regularly to business departments.	▪ Metrics from business departments showing the business impact of poor DQ. ▪ Metrics showing related industry statistics and benchmarks.	▪ Creating a regular DQ bulletin and highlighting what could be achieved with better DQ management. ▪ Benchmarking improvements with competitors and the industry.	▪ Ongoing DQ awareness program and improved data culture. ▪ Improved maturity, leading to data opportunities being shared across business departments.
12. Leverage external/industry peer groups.	▪ Existing forums — for example, Gartner's Peer Connect — for sharing DQ ideas and concerns. ▪ Identification of feeds, blogs and podcasts that are relevant to existing DQ programs.	▪ Using the DQ SIG to discuss the best ideas. ▪ Sharing the most relevant and interesting stories where improved DQ has had a positive (or negative) impact. ▪ Inviting peers to present at SIG meetings.	▪ Improved information culture and maturity. ▪ Identified opportunities and data points that improve internal programs and ongoing projects.

* A "stage-gate" is a formal review and assurance process ensuring that any products which are released conform to acceptable standards.
BAU = business as usual; DQ = data quality; SIG = special interest group

Source: Gartner (February 2014)

The existing release management "stage gate" — or equivalent — process that organizations have in place to deliver changed and new products and/or systems, may be a very good place to start an investigation (see Note 1). Typically, such stage-gate reviews already take into consideration a number of areas, including architectural compliance, information risk and privacy requirements. Because of this it may be possible to extend the scope of the stage-gate review process to examine how data quality is being addressed, and if there is a risk of key enterprise information degrading. Because there are many diverse stakeholder groups that benefit from data quality improvements, it is important to establish, implement and monitor a communication plan alongside the training and education activities that are carried out. Connecting the enterprise with data quality peer groups encourages organizational maturity in this area and provides:

- Alternative perspectives on best practices
- Insights into the approach of others' similar challenges
- Support of a professional peer group

The Gartner Peer Connect forum is one example that can be leveraged.

Gartner Recommended Reading

Some documents may not be available as part of your current Gartner subscription.

"Measuring the Business Value of Data Quality"

"Magic Quadrant for Data Quality Tools"

"Gartner's Data Quality Maturity Model"

"Toolkit: Assessing Key Data Quality Dimensions"

"The Gartner Business Value Model: A Framework for Measuring Business Performance"

"Developing Metrics for Successful Master Data Management"

"Certify Data to Foster Trust and Consistent Use"

"Information Infrastructure Modernization Key Initiative Overview"

"Predicts 2014: Information Governance and MDM Are Critical for Digital Transformation"

Evidence

[1] The observations, findings and recommendations in this document are based primarily on interactions with CIOs, chief data officers and information leaders. The interactions — focused on data quality, information governance and MDM — came about during the course of Gartner client inquiries from the past 12 months.

Note 1 Definition of "Stage Gate"

A "stage gate" is a formal review and assurance process ensuring that any products that are released conform to acceptable standards.

GARTNER HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
USA
+1 203 964 0096

Regional Headquarters
AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

For a complete list of worldwide locations,
visit http://www.gartner.com/technology/about.jsp