tcs | **TATA CONSULTANCY SERVICES**

Why Data Quality Matters More in a Business 4.0™ World

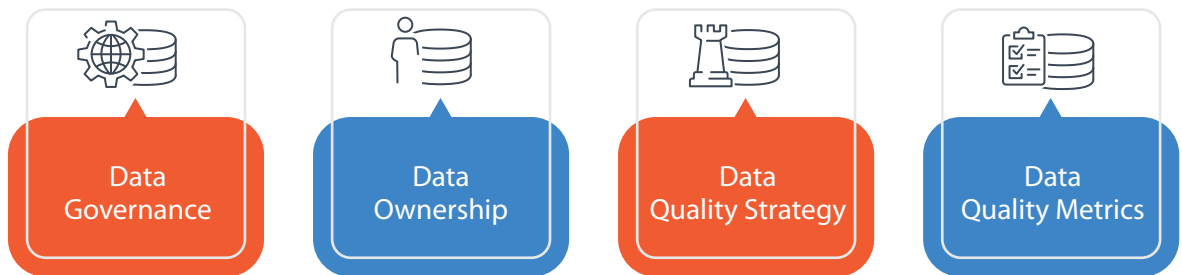| Data Governance | Data Ownership | Data Quality Strategy | Data Quality Metrics |

Executive Summary

In today's data-centric era, 'the more, the merrier' adage clearly doesn't apply to data – unless organizations ensure a high level of data quality. It is no secret that even as enterprises are drowning in a data deluge, they often struggle to find quality data at the right time to empower them with actionable and accurate business insights. The impact of poor data quality on business performance is much higher than what most organizations account for. In Experian's recent global survey, 95% of respondents indicated that poor data quality undermines business performance. A recent Gartner survey pegs the average cost of poor data quality at $12.8 million per year per organization. Clearly, there is a strong correlation between data and analytics maturity and business effectiveness, implying that data quality directly impacts a range of business outcomes.

Ensuring data quality needs intervention at multiple levels. In this paper, we identify the five key steps organizations should take to protect good data and remediate poor quality data.

Decoding the data quality challenge

Data quality is becoming increasingly important as organizations rely more heavily on data to drive business strategy. Poor quality data can result in a plethora of economic consequences ranging from poor customer service, inefficient operations, increased risk of fraud and regulatory fines or other sanctions for inaccurate financial or regulatory reporting. Data scientists report that they spend 50-80% of their time on mundane data gathering and cleansing activities rather than the real analytics activities. For the majority of organizations there is a considerable untapped business opportunity from improving data quality.

Data quality refers to fitness for purpose - does the data serve an intended purpose?

Six data quality dimensions are commonly identified to which we add availability. As illustrated in Figure 1, these are:

Figure 1: Data quality dimensions

Completeness and timeliness are usually relatively straightforward to detect and measure, for example null values and data available at the required time. Data accuracy and validity are typically harder to determine. Reconciliation processes are often deployed to determine accuracy and validity, although where used internally, this typically is an indication of data duplication. It is better to have clear, authoritative sources of data, eliminate duplication and thus, through quality data at source, remove the need for internal data reconciliations. Consistency and uniqueness can be particularly challenging for mature organizations with legacy and over-lapping systems.

Traditionally, data quality efforts have focused on structured data, but increasingly business analytics is additionally consuming unstructured or semi-structured data. The scope of data quality must now consider the full gamut of data types from structured through to unstructured. Collectively, the four Vs of data – variety, volume, velocity and veracity contribute to the data quality challenge.

In search of high-quality data: Five key steps

Good quality data not only increases customer awareness, but also helps organizations build an agile and iterative business strategy – key asks in today's Business 4.0™ world. Here are five proven steps that can help organizations improve data quality.

#1 Detect and log data quality issues

A well-defined set of metrics on business outcomes helps detect when results fall short of the desired level. Systems such as HPQC and ServiceNow can be used for monitoring and managing data and logging any issues where poor business results are a consequence of data quality.
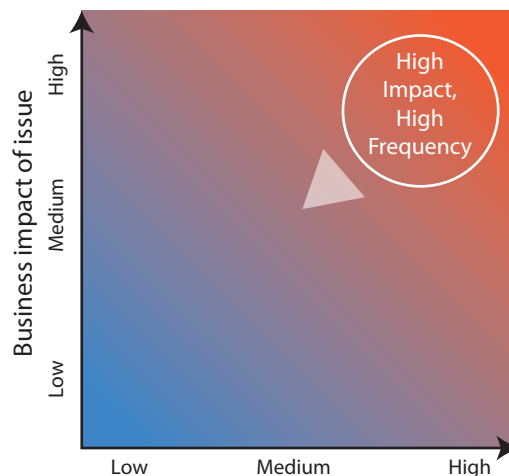

Figure 2: Resolving data quality issues

Many organizations have a mature system incident or outage logging and resolution system but do not treat data issues similarly. This despite the fact that the business impact of bad data can be just as damaging. As illustrated in Figure 2, organizations should start with high business impact, high frequency data issues. After these are resolved, they should shift the focus to issues with high/medium, medium/high and medium/medium impact and frequency respectively.

#2 Enable data quality issue resolution and ownership

An issue resolution process should be clearly defined and owned. It should encompass issue capture and logging, disposition, development of the technical fix, testing, production implementation and sign-off.

One of the challenges typically large organizations face is getting to the root cause of the problem and finding the appropriate owner to fix the issue. Data flows from system to system and the system where the data problem manifests itself may not be the cause of the problem. An organization with a collaborative, learning culture is helpful to bringing subject matter experts (SMEs) together to quickly analyze such an issue and determine where the root cause lies. A learning organization also asks "where else could the same type of issue occur?" It then takes measures to solve not only the issue that has actually occurred but other potential issues as well. Clear data ownership and governance ensures that the issue resolution process is appropriately prioritized and resources and budget are assigned as required. Data governance at an enterprise level helps create a data centric culture that accords data quality the highest priority. For example, in the financial industry, the Basel Committee on Banking Supervision (Basel Committee on Banking Supervision (2013) mandates industry players to establish unified standards on data management, adopt consistent data taxonomies, and establish uniform data architectures at the organizational level.

#3 Fix the data and the root cause

Data fixes typically take on two distinct aspects: (a) fixing the existing bad data and (b) fixing the root cause of the bad data so that the data generated in the future will be good. The first part can be a complex task where, for example, an incorrect data element on reference data causes errors on dependent transaction data. Decisions may need to be taken on the appropriateness of extrapolating values to fill in missing data or re-running part of the system processing flow. Once the data has been corrected, any reporting that uses the data - client reports, regulatory reports or internal management reports may need to be re-run. This may itself expose complexities and challenges with "as-on" date processing.

While the root cause of the bad data may be at the data source, it often presents elsewhere too in the data lineage. Multiple systems and long data lineage pathways can pose particular challenges for large enterprises in the detection and resolution of the root cause of data quality issues. An open, collaborative and learning culture fosters the right environment for teams to come together to rapidly identify and fix issues.

#4 Establish a data quality program

Due to the scale and pervasive nature of data quality, many organizations create a dedicated program to focus attention on fixing data quality issues, with both the senior management and staff across the organization being part of the program. It is important to adopt both - a top down and bottom up approach to data quality. Data programs should have senior sponsorship all the way up to the Board level. The CEO talking about the importance of data quality at a forum such as the quarterly town hall meetings can really help align the organization on the importance of data quality. Awareness and communication from the top are critical to creating a culture of data centricity that values and promotes data quality throughout the organization. Organizations should consider including data quality related goals in employee performance goals and even conduct periodic internal mock exams to test the data quality and fix issues.

#5 Define data quality outcomes and KPIs

Driving data quality outcomes requires clear KPIs that align to the overall business goals, ideally with quantitative financial impact. Examples include measuring the number of incidences of data failing quality checks, for instance, missing settlement instruction data and deriving the average cost of fixing each incident.

Data quality metrics can be categorized into two main components: (a) monitoring and (b) scorecards or dashboards. Monitors operate in real-time and provide alerts when data quality rules are broken. Scorecards are point-in-time reports of data quality usually graphing or aggregating data quality metrics over a period of time. Reporting on trends is important to determine whether data quality is improving or deteriorating. Individual scores will often be assessed against thresholds and RAG rated together with an indication of the trend to aid business interpretation.

We recommend taking a more holistic business-focused approach to metrics and spanning commercial, financial, operational and people aspects. This helps take a 360-degree view of the contribution of data to the business and becomes a powerful tool for driving data quality improvements.

Unlocking data's true potential

A recent Dutch research study, in collaboration with ICT Media, Tilberg University and TCS, reveals that there is a measurable positive correlation between high performing digital organizations and data maturity. Enterprises

that treat data as a first-class asset, have strong data governance and ownership practices, measure data quality and have robust processes to remediate data where the quality falls short. As organizations increasingly step into the digital future, it is important to take a proactive approach towards data quality and take full advantage of the value of data to drive business performance.

References:

Basel Committee on Banking Supervision (2013): "Principles for effective risk data aggregation and risk reporting", Bank for International Settlement, https://www.bis.org/publ/bcbs239.pdf

Experian (2019): "2019 Global Data Management Research", Benchmark Report.

Experian (2013): "The Data Advantage: How Accuracy Creates Opportunity", An Experian QAS 2013 Research Report.

Gartner (2020): "Cost Optimization is Crucial for Modern Data Management Programs", June 2020, https://www.gartner.com/en/documents/3986583/cost-optimization-is-crucial-for-modern-data-management-IBM (2011): "The Four V's of Big Data", https://www.ibmbigdatahub.com/infographic/four-vs-big-data

ICT Media, Tilberg University & TCS (2020): "High Performance Digital Organization", Demonstrate Data Driven Business Behavior, A Dutch Research Study on data and analytics maturity and business effectiveness.

TCS - DATOM® (2020): "Data and Analytics Target Operating Model", An Enterprise-wide Framework to enhance data and analytics program, TCS, https://www.tcs.com/data-analytics-target-operating-model

UK DAMA (2013) – Data Management Association: https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQ DimensionsWhitePaperR37.pdf

About The Authors

Paul Wood

Paul Wood has an advisory role in the Analytics and Insights team at TCS. With over 26 years of IT experience in the investment banking industry, he has led teams delivering complex transformational software development and package integration projects. Paul has worked across the trade life cycle from equity trading systems, straight-through-processing trade automation to financial management and reporting. He graduated in applied physics from Durham University in 1987.

Prithwis Kumar De

Prithwis Kumar De, Senior Consultant, has an Artificial Intelligence (AI) / Machine Learning (ML) advisory role in the Analytics and Insights team at TCS. He has managed and supported clients across industries such as BFSI, retail, CPG, life sciences and healthcare, telecom, among others, to drive accelerated business outcomes through data and analytics. In TCS, he has played multiple leadership roles across delivery, operations management, business development, pre-sales, solutions, offering development, AI/ML, and helped scale significant footprint for data and analytics business. He holds a PhD in Econometrics (applied statistics).

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at **www.tcs.com**

Experience certainty.   IT Services
                          Business Solutions
                          Consulting