Repercussions of Poor Data Quality (DQ) on Business Process Improvement (BPI) Projects

Krishna Valluru

Independent Researcher
kdvalluru@gmail.com
ORCID: 0000-0002-7438-3361

Abstract:
Business environment of the modern day is at the mercy of accurate and reliable data for informed decision making and continuous business process improvement (BPI). Business Process Improvement (BPI) projects rely significantly on high quality data to identify, assess, and improve processes that add value to the organization as a whole. Redman (2016) states that IBM's estimate of the yearly cost of poor data quality is $3.1 trillion in 2016 in the US alone [1]. Poor data quality poses significant challenges and severely undermines these projects, leading to inappropriate and inaccurate data-driven decisions, establishing futile strategies that barely come to fruition, unexploited time, money, and resources, and attracting anti-trust from key stakeholders. When it comes to BPI projects, if organizations are unable to put their money where their mouth is, their credibility and reputation is at stake which eventually leads to negative impact, plummeted monetary gains and loss of sustenance for organization's goals. This paper explores the multifaceted ramifications of poor data quality on BPI projects and highlights the importance of data quality management (DQM) for the success of these initiatives.

Keywords: Data Quality (DQ), Business Process Improvement (BPI), Operational Excellence (OpEx), Lean Six Sigma, DMAIC.

1. INTRODUCTION

Data is often pictured as the substantial factor in decision making in information driven organizations nowadays. However, if data quality is not up to the set threshold can cause the data to be inaccurate, inconsistent or untimely. The resultant effect undermines Business Process Improvement (BPI) initiatives and risk management efforts. The core objective of BPI projects is to augment efficiency in the current state processes of the organization, shrink the cost of operational processes, enhance customer satisfaction, and improve overall process performance. Business process improvement is strategic approach and a fundamental component of any successful process improvement project is highly reliable and high-quality data and the success of such projects is heavily contingent upon the quality of underlying data. However, operational excellence teams implementing process improvement projects across organizations are repeatedly exposed to fiascos due to issues arising from bad data which is inaccurate, inconsistent, and incomplete, leading to undesirable repercussions for BPI initiatives.

Organizations across myriad industries incessantly strive for operational excellence (OpEx), seeking to demonstrate how a lean based BPI eliminates non-value-added activities, enhance process efficiency, minimize undesirable costs, and ultimately improve customer satisfaction through Business Process Improvement (BPI) initiatives [2]. BPI methodologies drive continuous improvement which include process mapping, structured approach to analyze, redesign, and optimize workflows [3]. However, the success capability of these initiatives pivots critically on data quality (DQ) fueling them which is often a misjudged factor. Poor data quality can sternly undermine BPI projects at every stage leading to suboptimal outcomes, unexploited resources, and substantial undesirable business consequences [4].

Business Process Improvement projects embrace the saying "garbage in, garbage out" which is predominantly pertinent in the context of BPI and garnering the recognition of poor data quality input leading to unreliable data output [5]. The quality of the output is only as good as the quality of the input. Redman (1998) nonchalantly claims that 40-60% of the expense of a service organization may be consumed as an effect of poor data and these ranges are “good working estimates” of the cost of poor data quality [6].

2. IMPORTANCE OF DATA QUALITY IN BUSINESS PROCESS IMPROVEMENT (BPI) PROJECTS

Data quality is the pillar reinforcing the accomplishment of BPI initiatives. It enables data driven decision making, apprises exact problem solving, and ensures long-term sustainability. Pipino et al., (2002) states that data quality must be measured against how well data meets process-specific needs [7]. Organizations pursuing operational excellence must treat data governance as a prerequisite and also a premeditated benefit to improving process performance. Thoughtful process performance monitoring and accurate root cause analysis is enabled if data quality is high. Poor data quality deludes process improvement efforts making the issues at hand sloppier and also blowing up the operating costs. Conversely, high-quality data enables fact-based decision making and clear-cut performance monitoring of the improved processes. Identifying pain points and gaps to get to the root cause of the issue in the process is initiated in every process improvement effort irrespective of the scale and magnitude of the BPI project. Clean and reliable data is utmost important in this drill and root cause analysis cannot be made feasible without clean and reliable data. Measure and Analyze phases of DMAIC (Define, Measure, Analyze, Improve, and Control) are directly impacted by data quality. Incorrect root cause analysis (RCA) can be caused by inconsistent and missing data values. Poor data quality leads to additional effort for data validation and clean up, which inflates the cost of process improvement projects. In majority, BPI projects span across multiple departments, so data inconsistency between teams lead to misaligned views of the same process. Hence difficulty in standardizing processes arise. Good data quality ensures that control mechanisms remain beneficial even after the initial project ends which enables early detection of process anomalies and also enable predictive insights for ongoing optimization. Data quality is the backbone of every successful business process improvement initiative which enlightens accurate problem solving, empowers objective decision making, aids cost savings, and warrants sustainability. Organizations that are thoughtful about Operational Excellence (OpEx) must treat data governance as a strategic enabler of process performance, but not as an IT function.

3. CONSEQUENCES OF POOR DATA QUALITY IN BPI PROJECTS

Poor data quality is a downfall in silence which quietly erodes the improvements organizations expect by way of Business Process Improvement (BPI) initiatives. For majority of the organizations, the cost of bad data is estimated to be an astonishing 15% to 25% of revenue [8]. Failure of data quality assessment in any of the aspects of accuracy, completeness, consistency, timeliness, and relevance can lead to erroneous conclusions and ineffective solutions which hinder business process analyses [9]. To unearth inefficiencies in the process, accurate process mapping is quintessential. Misguided verdicts are the effect of inadequate data or poor data quality which misrepresents this immense process mapping effort [10]. Inaccurate data leads to unsound process mapping which delays timelines of process improvement projects and as a result overall effectiveness is tapered. Resource misallocation is a noteworthy direct result of poor data quality which often sidetracks sincere process improvement efforts away from the intended course of action [1]. If data is inconsistent, chaotic analytical outcomes are unescapable which diminishes stakeholder faith and leading to challenge future initiatives pertaining to business process improvement [11]. Process revisions, delays due to bottlenecks, and additional efforts for data purging are caused due to poor data quality which in turn increases project costs.

4. COMMON ROOT CAUSES OF POOR DATA QUALITY (DQ) AFFECTING BPI PROJECTS

In the context of supporting BPI projects, effective deterrence and damage restore strategies are crucial to recognizing the origins of poor data quality. Data deficiencies dent process improvement efforts and quite a few common root causes are:

a. Lack of robust data validation rules during data entry or processing paves the way for invalid or inconsistent data to enter into analysis downstream.

b. Human error that leads inaccuracies and inconsistencies during manual data entry.
c. Lack of standardization in data definitions and formats which hinders data integration and causes erratic analysis.
d. Lack of unique data identifiers for duplicate records compromises uniqueness and skews data analytics and operational processes significantly.
e. Inadequate data ownership and accountability lead to ambiguity which offers little to no incentive to proactively maintain data quality.
f. Lack of regular mechanisms to update data. Data becomes stale over a period of time and there should be processes in place for validating the data degrade.

5. LEVERAGING LEAN SIX SIGMA TO IMPROVE POOR DATA QUALITY (DQ)

DMAIC (Define, Measure, Analyze, Improve, and Control) framework of Lean Six Sigma can be leveraged to address data quality issues as it is cardinal to mitigating risk and empowering process excellence across services organizations. Each phase in the DMAIC framework can be applied to improvement of data quality. An elaborated outline is below.

- **DEFINE –** Identifying a specific pain point caused by poor data quality and articulating it with quantifiable impact is the starting point to develop a clear problem statement in the DEFINE phase. This phase can be utilized to set objectives to reach a certain target by providing a proper clarification of the data quality problem at hand. In the perspective of data quality, this phase evidently outlines the business process or any specific data realm suffering from data quality issues by identifying critical data elements and ensuing business impacts. A SIPOC (Suppliers, Input, Process, Outputs, Customers) diagram is supportive in understanding the flow of data at a high-level by mapping upstream and downstream data flows and is crucial for seeing where poor quality data is coming from. Identification of stakeholders and customers who use the data and how poor data quality affects them reveals where defects in data originate. Translating VOC (Voice of the Customer) into Critical-to-Quality (CTQ) requirements is also one of the most crucial steps for improving data quality. Data quality problems are often vast, and process improvement teams can get overwhelmed easily. To confront these kinds of scenarios, clear-cut boundaries and expectations should be defined to clarify the data quality problem which is critical for preventing the scope creep and further preventing poor data quality. To outline data quality issues, the DEFINE phase can be put to ritual to ensure the stage is set for measurable, cross-functional impact and strategic alignment that helps build a compelling business case to sustain financial and operational improvement.

- **MEASURE –** Data quality issues that are undermining process performance can be identified and quantified for which the MEASURE phase is key. During MEASURE phase, converting the hazy problem of "poor data quality" into quantifiable and solvable issue is critical. Baseline data quality metrics need to be established which involves creating precise and unambiguous definitions for specific and key data quality metrics such as completeness, accuracy, consistency, and timeliness. By doing so, actual understanding of data quality problem can be given a shape and form. Data quality can be improved with a systematic plan for data collection pertaining to the problem itself. Root Cause Analysis of data issues can be carried out in this phase which gets to the root of the issue to investigate why data quality is poor. Quantifying how metrics pertaining to poor data quality affect crucial business objectives is a significant contribution of MEASURE. As-Is data process map helps visualize the process, identify where data quality issues are taking cover, and quantify the current process performance. What acceptable data quality looks like can be determined using MEASURE phase.

- **ANALYZE –** The Analyze phase is the third step in the DMAIC methodology. When this phase is applied to data quality issues, root causes of data defects can be pinpointed, hypotheses can be validated using statistical analysis to know why data quality is poor, impact of each root cause on business outcomes or process performance can be quantified. The most effective tools for analyzing poor data quality are Pareto chart, Ishikawa diagram, 5 Whys, Regression Analysis, Histogram & Box Plot. These tools trace data quality issues back to their origins. Once root causes are identified, this phase enables prioritization of fixes based on

frequency and impact of data defects, enables evidence-based decision-making, and enables clear association between data issues and business pain points.

- **IMPROVE –** The IMPROVE phase in DMAIC is where root causes of poor data quality identified in Analyze phase are eliminated. This phase commences by taking advantage of the analysis work from prior Define, Measure, and Analyze stages of DMAIC. By this point, where data quality problems exist and their impact is already identified. The purpose of Improve phase in the context of poor data quality is to interpret root cause insights into actionable solutions, to enhance data accuracy, completeness, consistency, and timeliness by piloting and data validation that catch errors before they pass into data systems, to optimize processes and systems to prevent data defect recurrence. Techniques to improve data quality include, but not limited to Mistake-Proofing, Data validation rules, Pilot testing, Solution Prioritization Matrix, and Design of Experiments (DOE). The structured, data-driven approach is the chief benefit of IMPROVE phase. Solutions are not just guessed, but changes grounded in analysis and validation are implemented prior to complete deployment.

- **CONTROL –** The CONTROL phase in DMAIC is critical to reinforce the improvements with respect to data quality are sustained over the long term through data governance, data automation, and data accountability. In this phase, focus on tracking data quality metrics is continuous as the improvements tend to drift back to baseline if incessant monitoring of process performance dwindles. The purpose of CONTROL phase in the context of poor data quality is to sustain improvements, prevent recurrence of data defects, monitor performance using KPIs and dashboards, and standardize processes to ensure consistent data handling. Improvements made in the IMPROVE phase are collated into Standard Operating Procedures (SOPs) in this phase. Scheduled audits of data quality is established to grab hold of faults early before they accumulate. When data quality metrics tumble above or below the set control limits, a process to investigate root cause and take Corrective and Preventive Action (CAPA) is established in this phase. Feedback loops are established for continuous monitoring and to discover issues to steer them back into continuous improvement efforts. As vigilance diminishes, data quality improvements tend to wane over time and thus a strong CONTROL phase is essential which eventually transforms temporary improvements into persisting organizational practices.

6. STRATEGIES TO MITIGATE POOR DATA QUALITY IN BUSINESS PROCESSES

Customer-centric services organizations in today's world must adopt a data governance framework that is comprehensive and across-the-board to mitigate the adverse effects of poor data quality. For inclusive data quality, data cleansing and regular audits can help in significant error reduction. Data inconsistencies need to be detected and corrected before the decision making is impacted. This could be made possible by leveraging the power of lean six sigma data analytics to further foster a data-driven culture encouraging cross-functional collaboration for data management to ensure answerability and unambiguousness. Organizational awareness about the significance of data quality has to be promoted. Continuous training on data handling need to imparted to employees that can nurture a hands-on to maintaining data quality. To ensure that inputs to data sources are dependable, data quality audit needs to be performed before launching a BPI initiative. If data quality issues are discovered, they can be remedied early in the life cycle of process improvement project. Positive business process improvement outcomes can be influenced by investing in integrated data governance which can thereby facilitate improved data accuracy, consistency, and reliability.

7. CONCLUSION

Through Business Process Improvement (BPI), the pursuit of operational excellence (OpEx) is a strategic necessity for organizations of today. The repercussions of poor data quality on BPI projects are overwhelming. The success of BPI projects is unswervingly debilitated by poor data quality leading to botched initiatives and financial losses. Estimating the financial impact of poor data quality is difficult because the associated costs are voluminous and convoluted [12]. The success of BPI initiatives is inseparably associated to the quality of the data and the result ensuing from the analysis. Effective and sustainable process improvement is established via high data quality which is not only a supporting element, but also the fundamental starting point. Defective data leads to erroneous process analysis that offers a warped view of process performance. The repercussions

extend far beyond the BPI project itself and addressing this challenge requires a thoughtful and unified approach. Organizations must recognize data quality as a critical dependency for BPI. Within their chosen BPI methodologies, organizations must embed data quality improvement and monitoring activities and recognize data quality as a critical dependency for BPI. To avoid or minimize the impact of poor data quality on BPI projects, development of data quality metrics that are both subjective and objective is essential which requires an incessant awareness and effort in assessing data quality. Investing in data quality management is not only a cost of doing business but also an imperative that is strategic for organizations committed to leveraging BPI competitive advantage. Process optimization journey is arduous and simultaneously it is a precondition to commence with a commitment to trusted and reliable high-quality data that can determine the success and sustenance of process improvement efforts.

REFERENCES:

[1] Redman, T. C. (2016). Bad data costs the US $3 trillion per year. *Harvard Business Review*, *22*, 11-18.

[2] Womack, J. P., & Jones, D. T. (1997). Lean thinking—banish waste and create wealth in your corporation. *Journal of the operational research society*, *48*(11), 1148-1148.

[3] Harrington, H. J. (1991). Business process improvement: The breakthrough strategy for total quality, productivity, and competitiveness. *McGraw-Hill*.

[4] Haug, A., Zachariassen, F., & Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management, 4*(2), 168-193.

[5] Kilkenny, M. F., & Robinson, K. M. (2018). Data quality: "Garbage in–garbage out". *Health Information Management Journal*, *47*(3), 103-105.

[6] Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, *41*(2), 79-82.

[7] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211–218.

[8] Redman, T. C. (2017). "Seizing Opportunity in Data Quality." *MIT Sloan Management Review*, 27 November 2017, https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/

[9] Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33.

[10] Hammer, M. (1990). Reengineering work: Don't automate, obliterate. *Harvard Business Review*, 68(4), 104–112.

[11] Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, *29*(1), 3.

[12] Eppler, M., & Helfert, M. (2004, November). A classification and analysis of data quality costs. In *International Conference on Information Quality* (pp. 311-325). Cambridge: MIT.