Towards a Methodology of Data Management Metrics

Mario Cervi
Flávio de Almeida Pires
Assesso, Sao Paulo, Brazil

Li-hsin Chang
Lisa Novier
Richard Wang
Institute of Chief Data Officers
University of Arkansas at Little Rock

***Abstract*—Gaining competitive advantage is dependent upon an organization's ability to effectively leverage high quality data in business processes. Key Performance Indicators (KPIs) are frequently used to measure progress toward business objectives and are prevalent in project management methodologies. In addition, Return on Investment (ROI) is used to determine the net worth of capital investments. However, current methodologies lack the ability to link improvements in data quality resulting from investments in data management projects to these organizational metrics. This research extends the existing concepts of ROI and KPIs to propose a methodology for deriving data management metrics to support data management decision making in alignment with an organization's vision and strategic objectives.**

***Index Terms*—data quality, information quality, data governance, key performance indicators, return of investment, cost benefit analysis, business process**

I. Introduction

Gaining competitive advantage is increasingly dependent upon an organization's ability to effectively leverage data in business processes, resulting in a business need for higher quality data. To fulfill requirements for improved data quality, organizations are making large capital investments in data management projects and technology.

To meet the increasing demand for data management solutions, technology companies are offering data management products that will enable the automated enforcement of data governance policies and improve overall data quality. The viability of continued investment in data management projects requires data management professionals to demonstrate direct business value and impact to business leaders using data management metrics.

Key Performance Indicators (KPIs) are frequently used to measure progress toward business objectives and are prevalent in project management and Six Sigma methodologies. In addition, Return on Investment (ROI) is widely used to determine the net worth of capital investments. However, current methodologies lack the ability to link improvements in data quality to organizational metrics such as KPIs and ROI.

This research extends the existing concepts of KPIs and ROI to propose a methodology for deriving data management metrics to support data management investment decision making in alignment with an organization's vision and strategic objectives.

II. Literature Review

A summary of the literature related to data management, KPIs, and ROI is presented to provide operational definitions, describe the process for calculating metrics, and understand how to measure business process effectiveness and business benefit or value.

A. Data Management Definition

According to The DAMA Guide to The Data Management Body of Knowledge (DMBOK), [1] data management "consists of the planning and execution of policies, practices, and projects that acquire, control, protect, deliver, and enhance the value of data and information assets."

The American Health Information Management Association (AHIMA) [2] defines data management as "the business processes that ensure the integrity of an organization's data during collection, application, warehousing, and analysis."

Data management is the high-level planning and management practices focused on ensuring an organization properly manages and increases the usefulness of data and information assets to attain the most value. Data management is the execution of the daily processes within the business that assess and improve data quality in alignment with the organization's data management priorities and plan.

The DMBOK [1] identifies the following ten functions within data management:

1) *Data Governance*
2) *Data Architecture Management*
3) *Data Development*

4) *Data Operations Management*
5) *Data Security Management*
6) *Reference and Master Data Management*
7) *Data Warehousing and Business Intelligence Management*
8) *Document and Content Management*
9) *Meta-data Management*
10) *Data Quality Management*

B. Identifying Data Quality Perceptions and Issues

To determine the appropriate data management actions for an organization, the current state of data quality must be assessed and used to inform decisions. The Information Quality Assessment (IQA) [3] was devised to ascertain employee perceptions regarding the current state of an organization's data quality and level of knowledge of programs and processes within the organizations to manage data quality. The IQA is a survey tool for individuals who either, collect, care for, or use data in their job role. The results of the IQA are then compared to objective measures of data quality. The IQA or a similar instrument can be used to determine the gaps between perceptions and data quality metrics as well as a comparison between the perceptions of each job role.

C. Key Performance Indicator Definition and Purpose

KPIs measure improvements or decline in an organization's performance toward achieving critical business objectives [4], [5]. The purpose of KPIs is to monitor efficiency and cost effectiveness of an organization [6]. Haustein and Bauer [6], [7], [8] both agree KPIs are intended to measure the overall health of an organization.

PNMsoft Ltd. [9] offers additional insight by summarizing the types of KPIs into four high-level categories based on what is being measured:

1) *Revenue improvement*
2) *Cost reduction*
3) *Process cycle-time improvement*
4) *Increased customer satisfaction*

D. Creating effective Key Performance Indicators

Resources for selecting KPIs are plentiful. The internet contains KPI libraries and enterprise business systems such as enterprise resource planning, supply chain management, and customer relationship management systems come with sets of predefined KPIs [8]. Determining the optimal set of KPIs for an individual organization can be overwhelming.

To be effective, KPIs must be based on clear business objectives [4]. Business objectives are outcome-based high-level business goal that can be expressed in quantifiable and measurable terms. Clearly established business objectives can be translated into effective KPIs using a SMART goal framework which means the KPIs must be specific, measureable, attainable, relevant, and time-based [10], [11]. To be specific, the KPI description must provide enough information for the objective to be well understood by stakeholders. Measureable means success can be expressed in quantifiable terms and assessed against a specific target measurement or within a target range. Attainable means the goal is achievable with the resources and time allocated. Relevant means the KPI measures criteria that is meaningful to business objective. Finally, time-based means the KPI has a clearly defined time period or point in time when it will be measured.

Eckerson [12] adds to the criteria for effective KPIs by defining the following twelve characteristics, some of which overlap with the SMART framework:

1) *Aligned - KPIs must be aligned with corporate strategy and objectives*
2) *Owned – There must be an individual or group who "owns" each KPI on the business side and is accountable for its outcome*
3) *Predictive – KPIs are leading indicators of performance desired by the organization*
4) *Actionable - KPIs data is provided in a timely and actionable method so owners can intervene to improve performance*
5) *Few in number - KPIs should focus use efforts on a few high-value targeted tasks*
6) *Easy to understand - KPIs should be straightforward and not based on complex indexes that users do not understand how to influence directly*
7) *Balanced and linked - KPIs should balance and reinforce each other, not undermine each other and sub-optimize processes*
8) *Trigger changes - KPIs should trigger positive behaviors to create a chain reaction of changes in the organization to improve performance*
9) *Standardized - KPIs are based on standard definitions, rules and calculations so they can be understood throughout the organization*
10) *Context driven - KPIs contextualize performance by applying targets and thresholds to performance so users can gauge progress over time*
11) *Reinforced with incentives - Organizations can amplify the impact by attaching compensation or incentives to well understood and stable KPIs*
12) *Relevant - KPIs gradually lose their impact over time, so they must be periodically reviewed and refreshed (p. 16)*

Bauer [7], [8] agrees with Eckerson [12] that KPIs should be aligned to strategic business goals and objective rather than individual business processes and activities to optimize results and drive productive behaviors toward a common goal. A strategic alignment pyramid can be a useful tool to delineate the steps for developing KPIs to ensure strategic alignment [7].

E. Linking Data Management and Business Performance with KPIs

The primary objective of data management projects is to improve systemic data quality issues that are negatively impacting business processes and thereby reducing revenue. Redman [13] noted three studies that estimated the cost of poor data quality to be in the range of 8-12% of an organization's revenue. However, the studies [13] were proprietary and the

approach for calculating was not revealed. Gartner, Inc. [14] conducted a research study highlighting the following three findings, demonstrating additional impacts of poor data quality on businesses:

1) *40% of all business initiatives fail to achieve their targeted benefits primarily due to poor data quality*
2) *Data quality effects overall labor productivity by as much as 20%*
3) *As more business processes become automated, data quality becomes the primary limiting factor for overall process quality*

Both Redman [13] and the Gartner, Inc. study [14] demonstrate the criticality of establishing high data quality within an organization to fully realize business goals and objectives. The challenge is to develop a methodology organizations can consistently use to measure the business impact of poor data quality and the benefits gained through proper data management. The key to meeting this challenge is link the results of data management improvements to improved overall organizational performance and key success measures.

The primary method for understanding successes in business performance is through the monitoring of KPIs. Hence, there is a need for establishing a framework and methodology for linking data management actions to business KPIs

F. Calculating Return on Investment

Return on investment (ROI) "measures the monetary value of the results and costs for the program, usually expressed as a percentage" [15]. ROI is used as a decision tool to evaluate the net benefit of an individual investment or for a comparative evaluation of different investments by measuring the monetary benefit or value relative to the cost. Phillips [15] uses the following formula for calculating ROI:

$$ROI = \frac{benefit\ from\ investment - cost\ of\ investment}{cost\ of\ investment} \times 100 \qquad (1)$$

Since not all project benefits are realized immediately, a complete quantification of return on investment should include an analysis of costs and benefits over time. The calculation of ROI over time can be achieved by estimating how long it will take for the benefits to pay for the project cost.

G. Models for Assessing Costs and Benefits of Improving Data Quality

Measuring the benefits and costs for improving data quality should consider the trade-off between the cost of assuring data quality and the cost of implementing a data management project [18]. A certain level of data quality issues will always exist. Correcting all data defects is not cost-effective and therefore undesirable. An optimal level of effort can be obtained by comparing the cost of poor data quality to the cost of assuring data quality [18]. Attaining a higher level of data quality requires higher costs to assure the data quality. Additionally, the costs inflicted by poor quality data will gradually decrease, and approach zero as data quality improves.

Multiple models for assessing the costs and benefits of improving data quality exist in the literature. The effects of poor data quality can be categorized into four categories [18] by identifying the hidden versus direct costs of poor data quality with whether the data is used for strategic decision making or to perform operational tasks. The resulting four categories were:

1) *Hidden costs of poor quality data used for strategic decision making*
2) *Direct costs of poor quality data used for strategic decision making*
3) *Hidden costs of poor quality data used to perform operational tasks*
4) *Direct costs of poor quality data used to perform operational tasks*

Differentiating between these four types of incurred costs for poor data quality is important to ensure hidden costs are not overlooked, thereby underestimating the true impact of poor data quality. Additionally, organizations can use [18] these classifications to prioritize data quality issues for improvement based on whether strategic decision making or operational performance is the more critical success factor.

Eppler and Helfer [19] expand the classification of data quality costs to include both the costs incurred from poor data quality and the costs associated with improving or assuring data quality. This classification [19] incudes a useful taxonomy to identify how an organization can reduce the cost incurred by poor data quality by making specific investments to assure data quality in three ways, prevention, detection, and repair.

To help business users understand the value of a data management program, Villar [20] uses three categories, Process Benefits, Costs Benefits, and Quality Benefits. Process Benefits arise from increasing process efficiency, including both reduced process time and increased worker productivity from reducing or eliminating manual processes. Costs Benefits include, but are not limited to cost avoidance actions to prevent future problems and avoiding penalties for non-compliance with regulatory requirements. Finally, Quality Benefits refers to the benefits achieve by improving data quality, including reduce effort for data cleansing, and increased revenue from high quality data used in business processes or decisions. Capturing intangible costs and benefits are equally valuable and important as tangible benefits when assessing the overall costs and benefits of improving data quality.

Models for classifying data quality costs and benefits can be used in conjunction with the "Benefit versus Cost Matrix" [21] which divides data quality projects into four categories:

1) *High Benefit/Low Cost*
2) *High Benefit/High Cost*
3) *Low Benefit/Low Cost*
4) *Low Benefit/High Cost*

The "Benefit versus Cost Matrix" [21] is a prioritization tool for data quality projects with High Benefit/Low Cost being the most desirable project for initial priority.

III. Discussion and Analysis

KPIs are widely used to evaluate organizational performance. To be effective, KPIs must be aligned to an organization's strategic objectives and based on standardized rules and calculations that are understandable across the organization.

Traditionally, business leaders have focused on improving business processes to improve KPIs. As organizations become more reliant on data to drive business decisions, KPI calculations and the underlying business processes can both be negatively impacted by poor quality data resulting from inadequate data management functions within an organization.

Additionally, data management improvement decisions are often made from a bottom-up approach and not linked to organizational performance outcomes. Data management programs that are not linked to strategic organizational objectives become unsustainable over time because business executives are unable to discern the long-term business value of the investment.

This paper proposes a methodology for deriving data management metrics based on the existing concepts of KPIs and ROI to support data management investment decision making in alignment with an organization's vision and strategic objectives.

A. Framework for Linking KPIs to Data Management Projects

Data management decisions and actions must be aligned to an organization's strategic vision and strategy to gain executive support and funding. This can be achieved by demonstrating how the organization vision, strategy, and business objectives are positively impacted by the data management actions through the use of KPIs.

A Strategic Alignment Pyramid [7] can be expanded to include data quality critical success factors, data quality metrics, and data management actions. This expanded framework, the Data Management Project Pyramid (DMPP), ensures alignment from the organizational vision down to the data management actions. Figure 1. illustrates the DMPP.

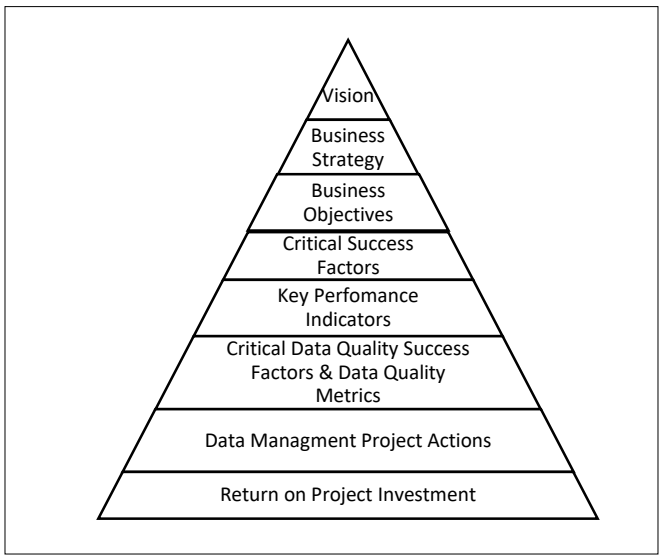

Fig. 1. Data Management Project Pyramid

The DMPP framework is implemented using a top-down approach. Critical Success Factors (CSFs) must be identified before KPIs are determined. CSFs are areas whose high performance or success is critical to the overall success of an organization [22]. CSFs are qualitative in nature and identify what should be done to achieve success compared to KPIs which are quantitative and measure the level of success gained from the CSFs [6].

The purpose of KPIs is to measure progress toward achieving strategic business goals. The DMPP places the development of KPIs in the middle of the process aligning the KPIs to both the business objectives above and the key data management action below. The selected KPIs, aligned to business objectives, would be used to monitor the business impact of data management actions. The Information Technology Infrastructure Library (ITIL) [23] recommends no more than two to five CFSs per process and two to five KPIs per CSF.

The selected KPIs should be high value indicators that measure progress toward achieving strategic business objectives. The previously discussed four categories of KPIs [9] are a useful guide for determining the type of KPI that has the greatest impact on overall organizational performance.

During the process of selecting KPIs, the business should determine the impact poor data quality has on the targeted KPIs. This process requires business process managers to determine the profit gained by increasing a KPI by a specific amount.

Critical to linking KPIs to data quality metrics is data management professionals collaborating with business process managers to understand the impact data quality metrics have on the selected KPIs [20]. Understanding the association of data quality to each KPI will link the data management actions to specific data quality issues that are negatively impacting organizational performance.

As a component of determining the data management project actions in the DMPP, the Information Quality Assessment Survey [3] or similar instrument can be utilized to assess the current state of data quality and identify the data management actions needed to improve data quality, as evidenced by the data quality metrics, and which support the Critical Success Factors.

B. Return on Investment

This article proposes that traditional ROI methodologies can be adapted to generate data management metrics to measure the value of data management projects.

Demonstrating business value by calculating the ROI for a data management project is essential to the long-term success of the project. Maintaining high data quality is an ongoing effort. In order to continue to fund the maintenance cost, business leaders must understand the net benefit or value derived from the data management investment.

The existing definition of ROI can be adapted to propose a new metric termed Return on Project Investment (ROPI) which

can measure the specific return of investment of data management projects.

ROPI = (value of project benefits – cost of project investment) / (cost of project investment) x 100

The project benefits should consider tangible benefits of the data management project, including the monetary value of the increases achieved in KPI metrics that is attributable to improvements in data quality. Although difficult to be evaluated, intangible benefits should also be indicated. ROPI is intended to evaluate the benefits and costs of data management project investments, analyzing opportunities for reducing data quality costs and increasing business profitability.

C. Implementing the Data Management Project Pyramid

The DMPP provides eight key steps an organization should undertake to make key data management project action decisions. Table 1. outlines the activities within each of these steps.

TABLE I. EIGHT STEPS OF THE DATA MANAGEMENT DECISION MAKING PROCESS

Vison	Meet with business executives to identify and discuss the organization vison long-term organizational goals.
Business Strategy	Meet with business leaders to understand the overall organization business strategy for achieving the organization vison and high-level goals.
Business Objectives	Engage the business management teams to identify the key business specific objectives that have the most impact on achieving the vison and strategy.
Critical Success Factors	Survey both business team members and management to determine the critical success factors that must occur to achieve the business objectives.
Key Performance Indicators	Collaborate with business management teams to identify existing KPIs or select KPIs that effectively measure progress toward the identified business objectives. Once KPIs are identified, baseline the KPIs prior to determining or implementing data management actions.
Critical Data Quality Success Factors & Data Quality Metrics	Meet with "data collectors, custodians, and consumers" [3] within the organization who have data-related responsibilities or use data in the business processes having the greatest impact on the business objectives in Step 3 to determine the critical data quality success factors. Survey "data collectors, custodians, and consumers" using the IQA Survey [3] or a similar instrument to determine the current state of data quality and inform data management action decisions. Collaborate with "data collectors, custodians, and consumers" to determine which data quality metrics will be used to assess the conformance of the data to the business rules used within the process.
Data Management Project Actions	Based on the results of the IQA survey [3] or similar instrument determine the appropriate Data Management Project Actions.
Return of Project Investment (ROPI)	Once the proposed Data Management Project Actions are known, ROPI is calculated to demonstrate the overall business value of the project. An estimated ROPI can also be calculated to assist business leaders with selecting from multiple possible Data Management Project Actions.

The activities in the Data Management Decision Making Process outlined in Table 1. can be used to build a business case for investments in data management projects. The KPIs and data quality metrics should be utilized over time to demonstrate the business impact data management actions have on improving organization performance and outcomes.

IV. CONCLUSION AND FURTHER RESEARCH CONSIDERATIONS

Data management metrics are vital to the success and long-term viability of data management projects. Existing business methodologies for aligning KPIs to an organization's vison and strategy can be leveraged to measure the business impact of data management actions with the DMPP. The Data Management Decision Making Process uses the eight elements of the DMPP to align data management decision making to key business processes to generate improved business performance. The business impact and value of data management project actions can be measured using both KPIs and by calculating ROPI. The DMPP and Data Management Decision Making Process provide a framework for measuring the business impact and value of the investment as well as progress toward overall organizational goals.

To confirm this theory, future research is needed in the form of a case study to test and fully develop the proposed framework. The goal is to provide a reusable methodology organizations can follow to align data management actions with strategic objectives and measure improvements in business performance due to improvements in data quality.

REFERENCES

[1] *DAMA Guide to the Data Management Body of Knowledge (DMBOK)*, 1st ed., Data Administration Management Association (DAMA), 2010, pp. 6-7.

[2] *AHIMA Body of Knowledge*, 2015, American Health Information Management Association (AHIMA), Chicago, IL, 2015, [Online]. Available: http://bok.ahima.org/doc?oid=107773#.WQfoENz_qpo

[3] Y. W. Lee, L. L. Pipino, J. D. Funk, and R. Y. Wang, "Assessing data quality, part I", *Journey to Data Quality*. Cambridge Press: MIT Press, 2006, pp. 27-51.

[4] IBM, (2017, April 17). "Key performance indicators (KPIs)", *WebSphere Business Modeler Advanced* [Online]. Available: https://www.ibm.com/support/knowledgecenter/SSBJDG_6.2.0/com.ibm.btools.modeler.advanced.help.doc/doc/concepts/measures/kpis.html

[5] Klipfolio (2017, February 22). *What is a KPI, Metric or Measure?* [Online]. Available: https://www.klipfolio.com/blog/kpi-metric-measure

[6] J. R. Haustein. (2012, June 25). *Successful Metrics* [Online]. Available: https://confluence.cornell.edu/display/metrics/Successful+Metrics

[7] K. Bauer, "KPIs-The metrics that drive performance management," *Information Management*, vol. 14, no. 9, p. 63, Sept. 2004.

[8] K. Bauer, "KPIs: Not all metrics are created equal," *Information Management*, vol. 14, no.12, p. 42, Dec. 2004.

[9] PNMsoft Ltd. (2017, April 21). *KPI examples - Types of Key Performance Indicators* [Online]. Available: http://www.pnmsoft.com/resources/bpm-tutorial/key-performance-indicators/

[10] Aveta Business Institute (2013, August 05). *Tips for Creating Key Performance Indicators*. [Online]. Available: http://www.sixsigmaonline.org/six-sigma-training-certification-information/tips-for-creating-key-performance-indicators/

[11] A. Hursman, "Measure what matters - Seven strategies for selecting relevant key performance indicators," *Information Management*, vol. 20, no. 4, p. 24, Apr. 2010.

[12] W. W. Eckerson, "Creating Effective KPIs," *Information Management*, vol. 16, no. 6, p. 15, Jun. 2006.

[13] T. C. Redman, (1998). "The impact of poor data quality on the typical enterprise," *Communications of the ACM*, vol. 41, no. 2, pp. 79-82, Feb. 1998.

[14] T. Friedman and M. Smith, "Measuring the business value of data quality," Gartner, Inc., Stanford, CT, Rep. G00218962, Oct. 2011.

[15] J. J. Phillips, "Measuring return on investment," *American Society for Training and Development*, vol. 2,p. 5, 1987.

[16] Y. W. Lee, L. L. Pipino, J. D. Funk, and R. Y. Wang, "Cost/Benefit analysis", *Journey to Data Quality*. Cambridge Press: MIT Press, 2006, pp. 13-26.

[17] J.N. Morgan, "A roadmap of financial measures for IT project ROI", IT Prof, vol. 7 no. 1, pp. 52-57, Jan/Feb 2005.

[18] A. Haug, F. Zachariassen, and D. Van Liempd, (2011). "The costs of poor data quality," *Journal of Industrial Engineering and Management*, vol. 4 no. 2, pp. 168-193, Jul. 2011.

[19] M. Eppler, and M. Helfert, "A classification and analysis of data quality costs," in *International Conference on Information Quality*, Cambridge, MA, 2004, pp. 311-325.

[20] M. Villar, "Data governance," in Institute for Chief Data Officers, Little Rock, AR, 2017, p.28.

[21] D. McGilvray 'The ten step process," *Executing data quality projects: Ten steps to quality data and trusted information™*,. Burlington, MA: Elsevier, 2008, pp. 177-178.

[22] J. F. Rockart, "Chief executives define their own data needs," *Harvard Business Review*, vol. 57, no. 2, pp. 81-93, Dec. 1978.

[23] *ITIL® Continual Service Improvement*, 2011 edition, London, UK, 2011.