

Towards a Context-Dependent Approach for Evaluating Data Quality Cost

Meryam Belhiah¹, Bouchaïb Bounabat²
AL QualSADI Team-ENSIAS, Rabat IT Center
Mohammed V University in Rabat, Rabat, Morocco

Abstract—Data-related expertise is a central and determining factor in the success of many organizations. Big Tech companies have developed an operational environment that extracts benefit from collected data to increase the efficiency and effectiveness of daily operations and services offered. However, in a complex economic environment, with transparent accounting and financial management, it is not possible to solve data quality issues with “dollars” without justifications and measurable indicators beforehand. The overall goal is not to improve data quality by any means, but to plan cost-effective data quality projects that benefit the organization. This knowledge is particularly relevant for organizations with little or no experience in the field of data quality assessment and improvement. Indeed, it is important that the costs and benefits associated with data quality are explicit and above all, quantifiable for both business managers and IT analysts. Organizations must also evaluate the different scenarios related to the implementation of data quality projects. The optimal scenario must provide the best financial and business value and meet the specifications in terms of time, resources and cost. The approach presented in this paper is an evaluation-oriented approach. For data quality projects, it evaluates the positive impact on the organization’s financial and business objectives, which could be linked to the positive value of quality improvement and the implementation complexity, which could be coupled with the costs of quality improvement. This paper tries also to translate empirically the implementation complexity to costs expressed in monetary terms.

Keywords—Data quality improvement project; cost of data quality; data quality assessment and improvement; cost/benefit analysis

I. INTRODUCTION

Repositioning into a data-driven organization, or at least turning the data available into a real asset [1], inevitably imposes the need to improve their quality. Indeed, organizations tend to depend on their data to make informed decisions. In addition, quality customer data is the most essential component of a CRM. It is also worth mentioning the various innovative uses of information for increasing operational efficiency, offering better products and services, reducing costs and controlling risks. Research in the area of data quality has shown that non-quality absorbs a considerable margin of an organization’s revenue. In the United States and at the end of FY 2012, The Postal Service estimated the cost of processing addressed and undelivered mail at \$ 1.5 billion [2]. A report published in 2011 by Gartner reveals that approximately 40% of the value anticipated by business initiatives is not achieved due to poor data quality. Indeed, the

latter affects operational efficiency, productivity, decision-making and downstream analysis [3].

A plethora of research in the academic and industrial spheres provides approaches to measuring the costs of poor data quality as well as the financial value of the improvement initiatives. However, generic and tangible metrics, based on a cost-benefit analysis, which can be adopted by organizations operating in diverse contexts, are lacking. The work presented in this paper tries to evaluate the business value of data quality projects, using a cost-benefit analysis. This approach can assist beneficiary organizations in determining the optimal investment to be allocated to data quality improvement projects. This paper tries also to translate empirically the implementation complexity to costs, expressed in monetary terms.

The organization of this paper is addressed as follows: Section II presents data quality definitions and dimensions. Section III summarizes the literature in both academic and industrial area that is focused on measuring the business and financial value of data quality. Sections IV and V describe the main steps of our approach. In Section VI, the conclusions and future work are summarized.

II. DATA QUALITY: DEFINITIONS AND DIMENSIONS

A. Data Quality Definitions

Data Quality is largely conceived as a multidimensional concept. It is commonly defined as “the degree to which information meets the requirements and expectations of all stakeholders who need it to execute their process” [4]. This concept is echoed by the expression “Fitness for use” [5] [6].

Particular attention is given to the context in which the data quality is considered, since it cannot be evaluated and analyzed independently of the environment of the organization in question. The environment refers to the direct environment of the organization, namely: its customers, competitors, suppliers, etc. as well as its macro environment: technological, geopolitical, economic, social, legal, etc.

The data is created or collected, stored and manipulated by the Information System through the various business processes deployed. Given the variety of fields of application, the heterogeneity of information systems and the increasing volume of available data (social networks, open data, data retrieved from connected objects, etc.), various data quality issues have emerged.

B. Data Quality Dimensions

Data quality dimensions describe an aspect of the data that can be measured and evaluated against a reference quality level in order to characterize the current level of quality.

Initially, researchers identified 179 attributes of data quality [3]. As it is a high number of dimensions to work with, advanced statistical methods have been applied to reduce the number of dimensions, in a consequent way, to 15, broken down into four categories [4]; which are: (i) intrinsic, (ii) contextual, (iii) representation, and (iv) accessibility.

Intrinsic dimensions describe how the data has a quality in itself; in other words, the way in which data must be accurate, credible, objective and reputable. The contextual dimensions underline the requirement that the quality of the data must be taken into account in the context of the task at hand. The final categories, the representation and accessibility dimensions, emphasize the role of systems and tools to facilitate interactions between users and data.

In general, data quality dimensions are categories used to characterize the data and their fitness for use. These make it possible to characterize the current state of quality and to communicate around the desired state.

In concrete terms, this qualification makes it possible to:

- Act as a reference framework and guide to quality standards;
- Act as an instrument for segmenting efforts to improve DQ;
- Match the dimensions of the DQ with the needs of the organization;
- Prioritize improvement scenarios of the DQ.

In an economic context where financial resources are scarce, the goal is not to achieve superior data quality regardless of cost. Indeed, in addition to the characterization of quality levels, it is important to integrate the cost dimension into any strategy that aims at the improvement of data quality. This approach intends to give value to data quality initiative in monetary terms. As such, the problem of quality improvement is approached both efficiently and effectively.

III. MEASURING THE BUSINESS VALUE OF DATA QUALITY

One of the most important topics of data quality management is how to define and measure the value of the data. This section presents an overview of the work of leading experts, in both industry and academia, on how data closes or produces value.

A. Research in the Field of Industry

In *“Data Quality for the Information Age”*, Thomas Redman identifies several ways in which poor data quality affects an organization's bottom line. These include: customer attrition, incidental cost induction, decreased employee satisfaction, negative impact on the organization's reputation, negative impact on decision-making, induction of costs related to process reengineering and the negative impact on the organization's long-term strategy, to name a few. He also

emphasizes that the production and maintenance of quality data can be a unique source of competitive advantage [7].

In *“Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits”*, Larry English focuses on the high cost of poor quality data. He cites examples of direct and indirect costs caused by inaccurate and incomplete data and false and misleading information. English also provides recommendations on how to measure the costs associated with poor data quality [8].

In *“Enterprise Knowledge Management: The Data Quality Approach”*, David Loshin describes the essential and incremental costs associated with poor data quality at the operational, tactical, and strategic levels. The categories identified form a framework that can be used to identify and evaluate the costs imputed to poor quality and to the same extent, the relative benefits of a high quality level within an organization. It defines incidental costs as those that are clearly identified, but that remain difficult to measure, such as the difficulty of making decisions as well as organizational conflicts. In contrast, essential costs such as customer attrition, scrap and rework and operational delays are costs that can be estimated and measured. Loshin also presents a process for using this framework to create an aggregated dashboard that “synthesizes the cost associated with poor data quality” [9].

In *“Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM)”*, Dannette McGilvray introduces several techniques for measuring the impact of data quality issues in both quantitative and qualitative terms. For this, she presents a framework to facilitate this analysis. These techniques include:

- Collection and analysis of background and examples of the impact of poor data quality on the organization;
- Creating a repository of current and future uses of data;
- Analysis of data quality issues;
- Creating a benefit versus cost matrix to understand the effects of poor data quality;
- Classification of problems in order of importance, as well as plausible solutions.

The aim of this evaluation is to build an optimal business case and to guarantee the necessary support from the management to carry out a data quality improvement project. As a result, the decision-making process for investments related to this activity is improved [10].

In *“Information quality applied: Best practices for improving business information, processes and systems”*, Larry English provides a summary of the costs associated with poor information quality [11]. He presents well known cases that have been widely publicized:

- In 1999, NASA lost the \$ 125 million *Mars Climate Orbiter* spacecraft and all the knowledge this spacecraft had to collect;
- In 2000, the United States Supreme Court discredited the vote of 4.6 million voters because of data quality issues.

English consolidated a list of poor software and data quality costs including references from more than 120 organizations, for a total of \$ 1.25 trillion. English concludes that in many industries, these costs account for 20-25% of the company's operating revenues. These costs are broken down into the costs of recovery after a process failure and corrective actions.

“DAMA Book of Knowledge” uses a similar approach to describe the value of the data in terms of the benefits derived from the use of quality data and the costs associated with the deterioration of data. DAMA recommends assessing the effects of potential changes in revenue, cost and exposure to various risks [12].

According to the statistics collected by English, as well as the categories and techniques presented by Redman, Loshin and McGilvray, the first step in improving the quality of data within an organization is to understand its value. The business value of data can be either negative, through the costs generated by the poor quality, or positive through the benefits of a high quality level. The quality of the data therefore has a direct impact on the value of the data.

Other approaches in the industrial field have also addressed this issue [3] [13].

B. Research in the Academic Field

Industrial researchers have placed particular emphasis on assessing the positive impact of improving data quality. The different approaches mentioned above propose formulas for calculating the financial benefits that would potentially result from this improvement.

In addition to the references cited above, other research work in the academic field, focused on assessing the value of the data through the analysis of costs associated with poor quality and with improving data quality [14] [15] [16]. It would also be fair to say that research in the academic sphere has paid particular attention to the definition and evaluation of costs attributable to poor data quality.

However, these approaches offer few quantifiable measures or valuations to express these costs in monetary terms, thus qualifying the importance and priority of data quality improvement initiatives.

IV. PORTFOLIO DATA QUALITY ASSESSMENT FRAMEWORK

Among the 15 dimensions of data quality, this paper focuses on the dimension of accuracy. However, it should be recalled that this methodology is reusable and transferable to other data quality dimensions.

Portfolio Data Quality Assessment Framework (PortfolioDQAF) enables the identification of the most efficient data quality improvement projects, through the suggestion of two aggregate indicators of positive impact and implementation complexity. The goal of PortfolioDQAF is to:

- Evaluate the positive impact of data quality improvement on the overall organization's business objectives;
- Evaluate the complexity of data quality improvement actions;

- Recommend, through the analysis of a proposed cost of quality model, the optimal business case for improvement.

These results will make it possible to select data quality projects, based on the benefits provided to the organization, compared to the complexity of implementation.

A. Identifying the Business Objectives

In order to understand how execution performance and quality of business processes affect an organization's success, business objectives, and results, the following key factors are considered: operational efficiency, increased revenue, improved productivity, reduced costs, improved customer satisfaction, compliance with regulatory authorities, and improved decision-making.

B. Identifying Evaluation Factors

1) *Positive impact criteria:* The positive impact factor is dividable into several criteria. These are the success factors identified in the beginning of this section, augmented by other criteria such as:

- **The transversal nature of the process** - improvement of the quality of critical data used by a transversal process has more impact compared to a vertical process;
- **The nature of the data** - the data is classified into: (i) master data; (ii) transactional data; and (iii) historical data. It can be assumed that improving the quality of master data has more impact compared to improving transactional or historical data;
- **The frequency of access to the data** - if the critical data is used several times by the process, the improvement of its quality will yield a more positive impact;
- **The completion deadline** - like other technological projects, a short completion time for the project, allows for quick results. Indeed, the extended delays may induce the demotivation of the project team, the change in the scope of the project, the evolution of the regulations, among others.

2) *Implementation complexity criteria:* Similarly, the implementation complexity factor is split into several criteria. The following criteria concern the evaluation of the complexity of improving accuracy. These criteria originate from the literature review, but also from interviews with IT managers.

The criteria considered for the improvement complexity are:

- **Risk level** - a high level of risk is proportional to a considerable level of implementation complexity. Risks can include data loss, shutdown, systemic risks, chain reactions, and cost, delays, allocated resources overruns, etc.;

- **Existence of standards to validate data** - the existence of standards for verifying and validating data reduces the complexity of detecting erroneous values;
- **Existence of a data repository** - the existence of a reference data source, even outside the organization's IS, reduces the complexity of correcting inaccuracies at the data level, through a data-match;
- **Key identification potential** - the existence of a primary key / global identifier, that is consistent across different data sources, reduces the complexity of data cleansing, making it easy to confront and cross-check data;
- **Nature of data processing** - from a technical point of view, the accuracy improvement project may consist of automatic, semi-automatic or manual processes. Manual processing, depending on the volume of the data, corresponds to a high load and complexity;
- **Volume of data to be processed** - high data volume leads to high load and complexity.

C. Measuring Evaluation Factors

Due to the fact that each organization has specific aspects and sets of success factors, and in order to provide a generic approach that can be implemented without any adjustment, the third step of our approach introduces the context-aware and configurable weighting coefficients.

Following are few examples where using different weighting coefficients is relevant:

- Public organizations may have more concerns about increasing end-users satisfaction (citizens in this particular case), than increasing revenues;
- Healthcare actors may devote more attention to meeting regulatory driven compliance than to the other factors, while still important, owing to the fact that norms and standards are mandatory in the field of healthcare;
- Industrial companies may give the same importance to all the factors above.

1) *Setting the relative importance of the criteria:* For each criterion, a weighting coefficient is defined by the business managers, thus making it possible to express the importance of the contribution of each criterion to construct the aggregate impact factor. As cited before, the weights are specific to each organization and describe its context and strategy. Table I depicts the configuration canvas for positive impact calculation.

A similar canvas, represented by Table II, is adopted for evaluating implementation complexity. Unlike the analysis of the impact factor, the definition of the weighting factors of the implementation complexity factor is the responsibility of the IT managers.

2) *Measurement of the positive impact:* Business and IT managers, who are in charge of data quality projects, must:

- List all the key business processes;
- Configure the importance of each factor by acting on the associated weighting coefficient. The sum of all weighing coefficient must be equal to 100;
- For each factor in column 1, select the corresponding value in column 2 (each value is associated with a notation in column 3).

In the case of an organization with several key business processes, the positive impact of each process is calculated using the weighted sum formula below:

$$\sum_{i=1}^m (R_i * I_i) / \sum_{i=1}^m (I_i) \tag{1}$$

Where R_i is the rating for the factor “i” and I_i is the weighing coefficient that is associated with the factor “i”, that was previously defined by both business and IT leaders. The obtained score ranges between 0 and 5, where “0” refers to “unnoticed impact” and “5” refers to “high positive impact”.

3) *Measurement of the implementation complexity:* The implementation complexity will be calculated as follows:

$$\sum_{i=1}^m (R_i * C_i) / \sum_{i=1}^m (C_i) \tag{2}$$

TABLE I. CONFIGURATION CANVAS FOR POSITIVE IMPACT CALCULATION

Factor	Values	Rating (R)	Weighting coefficient (I)
Impact on daily operations	- True - False	1 0	
Impact on short-term business/financial objectives	- Increasing revenues	0.15	
	- Increasing productivity	0.15	
	- Reducing costs	0.15	
	- Increasing end-user satisfaction	0.15	
	- Meeting regulatory driven compliance	0.15	
- Other	0.15		
Impact on decision making	- True - False	1 0	
Impact on downstream analysis	- True - False	1 0	
Is the process cross-functional?	- True - False	1 0	
Reasonable deadline	- True - False	1 0	
Nature of Data	- Master data	4	
	- Transactional data	2	
	- Historical data	0	

TABLE II. CONFIGURATION CANVAS FOR COMPLEXITY CALCULATION

Factor	Values	Rating (R)	Weighting coefficient (IC)
Risk Level	- Severe/inacceptable - Major - Medium - Minor - Imperceptible	1 0,75 0,5 0,25 0	
Existence of standards to validate data	- False - True	1 0	
Existence of a data repository	- False - True	1 0	
Key identification potential	- False - True	0 1	
Nature of Data Processing	- Manual - Semi-automatic - Automatic	1 0.5 0.25	
Volume of Data	- Very large - Large - Medium - Minor	1 0.75 0.5 0.25	

Where R_i is the rating for the factor “i” and C_i is the weighing coefficient that is associated with the factor “i”, that was defined previously by both business and IT leaders. The obtained score ranges between 0 and 5, here where “0” refers to “minimal complexity” and “5” refers to “severe complexity”.

4) *Analysis and recommendation*: Fig. 1 depicts the main steps of the PortfolioDQAF approach:

- Identification of the key processes that contribute the most to the organization’s objectives and results;
- Measurement of the complexity of the improvement of the given objects;
- Recommendation of the optimal scenario for improving data quality.

Defining the key processes is therefore about defining who is involved in achieving the organization’s goals. These processes depend on the data needed to achieve these goals.

Fig. 1 depicts the phases of

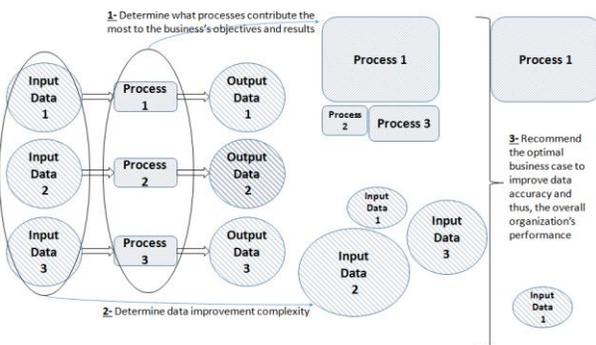


Fig. 1. Main Process [17].

The quantitative metrics of PortfolioDQAF approach correspond to the factors of positive impact and complexity of implementation. Each factor is expressed with a score, ranging from 0 to 5, corresponding to a level of impact or complexity.

The purpose of the next section is to translate empirically the implementation complexity to costs, expressed in monetary terms. Although this section does not claim to present a comprehensive cost theory, it attempts to introduce some elements that could be the starting point for a cost model for data quality.

V. TOWARDS A CONTEXT-DEPENDENT APPROACH FOR EVALUATING DATA QUALITY COST

As presented in Section IV, the objective of multi-criteria analysis of data quality improvement projects is to select a portfolio of projects, which produces the best cost-benefit ratio, according to several constraints. In this type of problem, there is uncertainty about the ROI of these projects. Non-linear programming helps to resolve the uncertain nature of this problem [18].

The classical Cost of Quality models are the PAF model (Prevention-Appraisal-Failure), which was first published by 1956 [19] and the Juran model [20]. In the context of data quality, reference books in terms of evaluation of data quality initiatives are [7] [8] [9].

Currently, the majority of models that measure the business value of data quality are developed in the context of the industry: [3] [13].

A. Definition of the Decision Problem

PortfolioDQAF approach qualifies the positive impact of data quality and its implementation complexity by quantitative indicators. It is now important to characterize the financial cost reflecting the implementation complexity.

Currently, there are little public information available that address the link between investment in terms of cost of data quality and expected quality levels, which makes this characterization difficult. It is however possible, from experience in the field of industry to make the following assumptions:

- The cost-quality curve of data would be convex;
- The improvement cost is equal to zero if the same level of data quality is maintained;
- The quality cost is exponentially high when approaching 100% accuracy. However the gradient becomes more important towards the maximum of the quality;
- The gradient would be a function of complexity.

This mathematical problem has a predictive model and the shape of the mathematical function is weakly defined.

To optimize this mathematical problem, the decision variables should be identified, as well as the constraints. The objective function is constructed from these same decision variables and constraints.

B. Definition of Decision Variables

To select projects, managers must make two independent but related choices:

- Which project portfolio can be selected for implementation? To model this decision, the following binary decision variables will be used: Y_i , where $i = 1, 2, 3$.
- Managers need to determine the optimal level of data quality that can be achieved. This variable will be characterized by the variables a_i , where $i = 1, 2, 3$.

C. Constraints Definition

In a real case scenario, this objective function would be subject to several constraints, among which:

- The sum of the costs of selected projects must not exceed the budget allocated to the data quality improvement program:

$$\sum_i Y_i * C_i \leq C \quad (3)$$

Where:

- C_i : corresponds to the cost associated with improving the accuracy of the business object I ,
- C : corresponds to the overall cost of improving data accuracy;
- Constraints on human resources allocated to projects.

D. Empirical Form of the Objective Function

The goal of managers is to minimize the costs associated with data quality improvement projects, while maximizing the expected impact of this improvement. This implies that the complexity factor composes the numerator and the positive impact factor is part of the denominator. The objective function is thus composed of:

- The positive impact factor;
- The implementation complexity factor;
- The initial accuracy of the business object;
- The target accuracy of the business object.

$$\min_i \sum Y_i * \frac{C_i * (a_i - a_{0i})}{I_i * (1 - a_i)} \quad (4)$$

Where:

- Y_i : is the binary variable that defines whether the business object i will be selected;
- C_i : refers to the complexity of implementing the business object i ;
- I_i : refers to the positive impact of the business object improvement i ;
- A_{0i} : refers to the initial accuracy of the business object i ;
- A_i : refers to the target accuracy of the business object i .

E. Identification and Resolution of Data Quality Issues

The final step in the business case consists of understanding the roots of data quality issues and how they can be addressed. Typically, this means reviewing the flow of information from the point of creation of the process-level data to detect when the error was introduced. Once the source of the data failure is identified, the data analyst can consider the alternatives to eliminate sources of errors, instituting preventive measures or corrective actions. Each of these alternatives will have an impact and will introduce a financial or other cost (organizational, risk, etc.), that can be measured, even conservatively, with the cost-benefit analysis proposed by PortfolioDQAF. The techniques in sections IV and V will prioritize actions that have an interesting cost-benefit ratio, subject to different financial and human constraints.

In addition, and with the objective of recommending the optimal scenario to improve the data accuracy and thus the overall performance of the organization, the model takes into consideration:

- The initial level of data quality (as-is);
- The positive impact of key processes that use the data to be improved;
- The complexity of implementing data quality improvement.

Depending on the values of these indicators and the target accuracy (to-be), one or more improvement scenarios should be considered.

A web platform has been developed to implement PortfolioDQAF approach and calculate the various metrics, automatically.

The main features of this application are:

- 1) Create the definition of business processes;
- 2) List all configured business processes;
- 3) Add new business objects (physically implemented by data objects), which are used by previously registered business processes;
- 4) List all registered business objects;
- 5) Evaluate data quality improvement projects.

VI. CONCLUSION AND FUTURE WORK

Data quality problems perfectly illustrate the key principle of any performance effort: "You can't control what you can't measure". This paper presents PortfolioDQAF approach, which is a metrics-based approach, to evaluate and analyze the positive impact and complexity of implementation of data quality improvement projects. This approach provides a factual basis for identifying and justifying investments in data quality. It is also a medium of communication between business managers and IT analysts. PortfolioDQAF develops a cost-benefit model, based on multi-criteria decision support analysis, to evaluate the portfolios of data quality improvement projects. It also introduces an approach to evaluate data quality cost.

It is possible to identify numerous research and future development, among which: (i) the generalization of PortfolioDQAF approach for more dimensions of data quality, in addition to accuracy. For instance, timeliness and consistency and (ii) the enrichment of the approach by an automated mechanism for reassessing cost progression as part of a continuous improvement of quality.

REFERENCES

- [1] Evans, Nina, Louis Fourie, and James Price. "Barriers to the Effective Deployment of Information Assets: The Role of the Executive Manager." *Proceedings of the European Conference on Management, Leadership & Governance*. Vol. 7. 2012.
- [2] Nixon, Ron. "Postal Service Reports Loss of \$15 Billion." *The New York Times*, *The New York Times*, 15 Nov. 2012, www.nytimes.com/2012/11/16/us/politics/postal-service-reports-a-nearly-16-billion-loss.html.
- [3] Gartner Group. *Measuring the Business Value of Data Quality*. 2011, https://www.data.com/export/sites/data/common/assets/pdf/DS_Gartner.pdf.
- [4] International Association for Information and Data Quality. 2013, <http://iaidq.org/main/glossary.shtml#I>.
- [5] Wang, Richard Y., and Diane M. Strong. "Beyond accuracy: What data quality means to data consumers." *Journal of management information systems* 12.4 (1996): 5-33.
- [6] Batini, Carlo, et al. "Methodologies for data quality assessment and improvement." *ACM computing surveys (CSUR)* 41.3 (2009): 16.
- [7] Redman, Thomas C., and A. Blanton. *Data quality for the information age*. Artech House, Inc., 1997.
- [8] English, Larry P. *Improving data warehouse and business information quality: methods for reducing costs and increasing profits*. Vol. 1. New York: Wiley, 1999.
- [9] Loshin, David. *Enterprise knowledge management: The data quality approach*. Morgan Kaufmann, 2001.
- [10] McGilvray, Danette. *Executing data quality projects: Ten steps to quality data and trusted information (TM)*. Elsevier, 2008.
- [11] English, Larry P. *Information quality applied: Best practices for improving business information, processes and systems*. Wiley Publishing, 2009.
- [12] DAMA, O. GUIA. "The DAMA guide to the data management body of knowledge." DAMA International (2012).
- [13] Laney, Douglas B. *Infonomics: how to monetize, manage, and measure information as an asset for competitive advantage*. Routledge, 2017.
- [14] Haug, Anders, Frederik Zachariassen, and Dennis Van Liempd. "The costs of poor data quality." *Journal of Industrial Engineering and Management (JIEM)* 4.2 (2011): 168-193.
- [15] Eppler, Martin, and Markus Helfert. "A classification and analysis of data quality costs." *International Conference on Information Quality*. 2004.
- [16] Kim, Won, and Byoungju Choi. "Towards Quantifying Data Quality Costs." *Journal of Object Technology* 2.4 (2003): 69-76.
- [17] Belhiah, Meryam, Mohammed Salim Benqatla, and Bouchaïb Bounabat. "Decision support system for implementing data quality projects." *International Conference on Data Management Technologies and Applications*. Springer, Cham, 2015.
- [18] Ragsdale, Cliff. *Spreadsheet modeling and decision analysis: A practical introduction to business analytics*. Nelson Education, 2014.
- [19] Feigenbaum, Armand Vallin. "Total quality-control." *Harvard business review* 34.6 (1956): 93-101.
- [20] Juran, J. M., F. M. Gryna, and R. Bingham. "Quality Control Textbook." (1975).