CORPORATE DATA QUALITY MANAGEMENT IN CONTEXT
(Research-in-Progress)

Ana Lucas

Laboratório Nacional de Engenharia Civil and Instituto Superior de Economia e Gestão, Portugal
ana.lucas@lnec.pt

Abstract: Presently, we are well aware that poor quality data is costing large amounts of money to corporations all over the world. Nevertheless, little research has been done about the way Organizations are dealing with data quality management and the strategies they are using.

This work aims to find some answers to the following questions: *which business drivers motivate the organizations to engage in a data quality management initiative?, how do they implement data quality management*? and *which objectives have been achieved, so far?*

Due to the kind of research questions involved, a decision was made to adopt the use of multiple exploratory case studies as research strategy [32]. The case studies were developed in a telecommunications company (MyTelecom), a public bank (PublicBank) and in the central bank (CentralBank) of one European Union Country.

The results show that the main drivers to data quality (DQ) initiatives were the reduction in non quality costs, risk management, mergers, and the improvement of the company's image among its customers, those aspects being in line with literature [7, 8, 20]. The commercial corporations (MyTelecom and PublicBank) began their DQ projects with customer data, this being in accordance with literature [18], while CentralBank, which mainly works with analytical systems, began with data source metadata characterization and reuse. None of the organizations uses a formal DQ methodology, but they are using tools for data profiling, standardization and cleaning. PublicBank and CentralBank are working towards a Corporate Data Policy, aligned with their Business Policy, which is not the case of MyTelecom. The findings enabled us to prepare a first draft of a "Data Governance strategic impact grid", adapted from Nolan& MacFarlan IT Governance strategic impact grid [17], this framework needing further empirical support.

Key Words: Corporate Data Quality Management, Case Study, Master Data Management, Metadata Management

INTRODUCTION

Data quality management (DQM) is an issue of growing importance for the academic and professional communities. Today, there is a great concern for the quality of corporate data, as data of poor quality means inaccurate information, which in turn means wasting resources and harming the organization, in particular the regulatory compliance and the relationships with its customers. Just to get an idea of poor data quality costs, The Data Warehousing Institute [27] estimated that current data quality problems cost U.S. businesses more than USD 600 billion a year.

English [7] classified the costs of poor data quality into three categories:

- *Process failure costs*, which occur when processes do not perform properly due to poor quality data, such as costs associated with misdelivered or undeliverable mail due to inaccurate mailing addresses;
- *Information scrap and rework*, such as costs associated with resending mail or with scrapping of defective data and their rework to achieve the desired quality levels;
- *Opportunity costs*, due to the lost and missed revenues. For example, due to the low accuracy in customers' addresses associated with "loyalty cards", a percentage of those card owners are not reached in advertising campaigns, resulting in lower revenues. Another example can be a customer loss due to incorrect billing.

According to [18] *business networking, customer management, decision-making/business intelligence*

and regulatory compliance are some of the main areas in which corporate data quality management plays an important role.

One of the most widespread definitions for *data quality* (DQ) comes from Juran [13], for whom *data are of high quality if they are fit for their intended uses in operations, decision making and planning.* Since this definition is very high level, and requires further operationalization, we will return to this subject later on.

Hoffer et al. [10, p. 601, 604] define *data* as "stored representations of objects and events that have meaning and importance in the user's environment" and *information* as "data that have been processed in such a way as to increase the knowledge of the person who uses the data". According to Drucker [6, p. 251], *knowledge* is "information that changes something or somebody, either by becoming grounds for actions, or by making an individual (or an institution) capable of different or more effective action", and, hence, data quality is ultimately intended *to increase the productivity of the knowledge worker* so as to create *value for business* as well as to assure *data risk management and compliance* [11, p. 8]. The term knowledge worker was first coined by Peter Drucker in 1959, and it must be pointed out that literature presents currently multiple definitions, although they usually only differ in small details. Sveiby [22] considers knowledge workers as those who are highly qualified and highly educated professionals, and their work consists largely in converting information to knowledge, using their own competencies for the most part, sometimes with the assistance of suppliers of information or specialized knowledge. According to Drucker [5, p. 169], "the most valuable asset of a twenty-first-century institution, whether business or non-business, will be its knowledge workers and their productivity", and although the knowledge worker's productivity depends on multiple factors, one is certainly related to the quality of data and information available.

Although data and information mean slightly different things, for reasons of simplicity, and in line with other research approaches to data quality, we will use, in the context of this paper, data and information interchangeably.

Generally speaking, DQM can be defined as the "*quality-oriented management of data as an asset*" [30, p. 4:4], that is, the "the application of total quality management (TQM) concepts and practices to improve data and information quality, including setting data quality policies and guidelines, data quality measurement (including data quality auditing and certification), data quality analysis, data cleansing and correction, data quality process improvement, and data quality education" (The Data Management Association [26, p. 43]). To be effective, data quality management must go beyond the activities of *fixing non-quality* data, to *preventing data quality problems* by managing data over its lifecycle to meet the information needs of their stakeholders. Moreover, DQM requires *breaking down the stovepipes* separating data across business units and *creating collaboration between business and IT functions*, in order to address both organizational and technical perspectives, requiring a profound *cultural change* demanding *leadership*, *authority*, *control and allocation of resources*, which means governance, specifically *data governance* (DG). Although DG does not equal DQM, either with regard to who makes the decisions, or to their scope, *DG being the responsibility of the board of directors and executive management* and more focused on corporate environment and strategic directions, including other areas beyond DQM, like Data Security and Privacy and Information Life-Cycle Management [11], with data governance, organizations are able to implement corporate-wide accountabilities for DQM, encompassing professionals from both business and IT units [31].

Due to the type of our research questions, which are: *which business drivers motivate the organizations to engage in a data quality management initiative?* and *how do they implement DQM?* we have adopted the use of multiple exploratory case studies as our initial research strategy [32], the goal being to understand, in its context, why some organizations have decided to embark in a data quality management initiative, how they did that and *which objectives have been achieved*, so far. Empirical evidence collected is expected to help to identifying the topics to be covered in a subsequent research project.

The paper is organized as follows: in this section we introduce the research problem; in the background section we present the main data qualty management related concepts and roles/responsibilities, as well as the data quality concept, together with a brief overview of the methodologies used by organizations to assess and improve the quality of their data, as well as the data quality and data governance maturity models. In the data collection, results and discussion section we present the research process, as well as the results and we discuss the findings. In the last section we present the conclusions, the limitations of the work and some guidelines for future research.

BACKGROUND

This section contains an introduction to the fundamentals underlying the work, in particular the main concepts related to data qualty management, the data quality concept and its most important dimensions, as well as some notions concerning data quality management maturity models and data quality methodologies.

Data Quality Management Approach

Data quality management (DQM) is "the application of total quality management (TQM) concepts and practices to improve data and information quality, including setting data quality policies and guidelines, data quality measurement (including data quality auditing and certification), data quality analysis, data cleansing and correction, data quality process improvement, and data quality education" [26, p. 43].

Moreover, by adapting to data quality the vocabulary presented in ISO [12], we can define DQM as as a set of coordinated activities aimed to direct and control an organization with regard to data quality.

Table 1 presents a set of concepts and roles/responsibilities related to the management of data, considered as a corporate asset, collected from academic or professional sources.

Concept or Role/ Responsibility	Definition
Data Policy (DP)	Redman [20, p. 290], defines Data Policy as "a statement that delineates management responsibility for data and activities that touch and/or impact data and information". According to the same author, Data Policy can cover the following interrelated categories [20, p. 40-41]: – Data Quality in its broadest sense; – Data assets inventory; – Data sharing and availalibity; – Data architecture; – Data security, privacy and appropriate use; – Data planning.
Data Governance (DG)	– DG "specifies the framework for decision rights and accountabilities to encourage desirable behaviour in the use of data. To promote desirable behaviour, data governance develops and implements corporate-wide data policies, guidelines, and standards that are consistent with the organization's mission, strategy, values, norms, and culture" [30, p. 4:6]. – DG is "the exercise of authority, control and shared decision-making (planning, monitoring and enforcement) over the management of data assets. Data Governance is high-level planning and control over data management and coordinates the collaboration between IT and the enterprise" [25, p. 38].

Concept or Role/ Responsibility	Definition
	− DG is the "formal orchestration of *people*, *processes*, and *technology* to enable an organization to *leverage data as an enterprise asset* [24, p. 1].
Data Owner (DO)	DO is the entity (usually a business unit) having responsibility and authority for a specific dataset.
Data Quality Methodology (DQm)	− A DQm is "a set of guidelines and techniques that, starting from input information describing a given application context, defines a rational process to assess and improve the quality of data" [1 , p. 16:2]. − A DQm is made of phases and activities and uses techniques (DQT) and tools (DQt) to accomplish its work.
Data Quality Techniques (DQT)	DQTs can be *data* and *process driven* [1]: − The *data driven* DQTs correspond to algorithms, heuristics, knowledge-based and learning processes that provide a solution for specific DQ problems, like record linkage (eg finding and merging duplicates, i.e. different records that represent the same real world entity), standardization techniques (comparing data with lookup tables, and updating it accordingly) or data and schema integration; − *Process driven* techniques are used to describe, analyze and reengineer the information production processes [1].
Data Quality Tools (DQt)	DQt are software products that implement specific DQTs to address the core functional requirements of the data quality discipline, in particular profiling, parsing and standardization, generalized "cleansing", matching, monitoring and enrichment. Adapted from [9].
Data Quality Assurance (DQA)	Data Quality Assurance (DQA) is the part of data quality management focused on providing confidence that quality requirements will be fulfilled (adapted from ISO [12] to DQM).
Data Quality Control (DQC)	Data Quality Control (DQC) is the part of data quality management focused on fulfilling quality requirements (adapted from ISO [12] to DQM).
Master Data (MD)	− Master Data is "the consistent and uniform set of identifiers and extended attributes that describe the core entities of the organization and are used across multiple business processes". Examples are Customers and Employees [19, p. 3]. − Master Data can be defined as the data that has been cleansed, rationalized, and integrated into an enterprise-wide "system of record" for core business activities [2].
Master Data Management (MDM)	− MDM is "a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, semantic consistency and accountability of the organization's official, shared master data assets" [19, p.3]. − MDM is the framework of processes and technologies aimed at creating and mainting an authoritative, reliable, sustainable, accurate, and secure data environment that represents a "single version of truth", constituting itelf as an accepted system of record used both intra and inter-enterprise across a diverse set of application systems, lines of business, and user communities [2]. − MDM is designed around the concept of a (virtual or physical) central repository to store and manage master data and can be implemented according to various architectural styles.

Data Steward (DS)	According to DAMA [26, p. 45], DS is a business leader and/or subject matter expert designated as accountable for: − The identification of operational and business intelligence data requirements within an assigned subject area; − The quality of data names, business definitions and domain values within an assigned subject area; − Compliance with regulatory requirements and conformance to internal data policies and data standards; − Application of appropriate security controls; − Analysis and improving of data quality; − Identification and solution of data related issues.
Data Quality Champion (DQC)	According to [23, p. 338], DQC "are managers who actively and vigorously promote their personal vision for using data quality related technology innovations". They push projects over approval, provide political support, keep participants informed, and allocate resources to data quality projects.

Table 1 – Concepts and Roles/ Responsibilities for the Data Governance and DQM Approach

The Data Quality Concept

As stated before, *data are of high quality if they are fit for their intended uses in operations, decision making and planning* [13]. Although widely accepted, this broad definition needs in depth specification, with data quality being presented in the literature as a multidimensional concept [28, 29, 20] and operationalized through its dimensions, which are data characteristics that are valued by data consumers, like accuracy, timeliness, interpretability, completeness, relevancy, etc.

Three main approaches to the identification and definition of universal (domain independent) dimensions available in literature are as follows:

a) Theoretical [28];
b) Empirical [29] ;
c) Intuitive [20].

These approaches refer to both the data in extension, i.e., their values [28, 29, 20, 7] and in intention, i.e., their models or database schemas [7, 20]. Moreover, some authors [29, 20, 7] also consider data presentation dimensions and data security and accessibility dimensions [29]. With regard to dimensions of data models, both the professional and the research literature only consider intra-data model dimensions (those related to the schema of a specific database).

Corporate data quality demands an holistic view over the data asset and brings new challenges that have not hitherto been provided by the various approaches to data quality. In line with [21], we stress that corporate data quality also depends on inter-data models issues or, put otherwise, on corporate data architecture, which requires considering the inter-data model quality dimensions, as they will ensure the corporate data model integration and consistency, as presented in Fig. 1. For example, although the Billing Subsystem may have adequate customer data quality, CRM and ERP may not, if only in depth data quality is being considered.
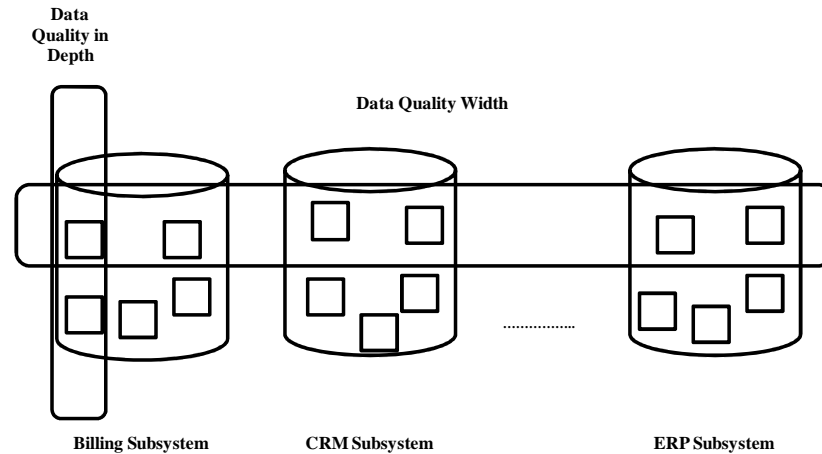
Fig. 1 - Corporate Data Quality, adapted from [21]

The corporate data model integration and consistency, represented by inter-data model dimensions is largely valued by data consumers, and should theoretically support the much publicized practitioners' concept of Master Data Management (MDM).

Moreover, the interpretability dimension [20, 28] requires the availability of metadata related to the most critical data elements.

In short, we postulate a broad concept of data quality, in line with its definition [13], which goes beyond the strict concept of data values quality (according to relevant dimensions), also including data architecture integration, metadata management and the organization environment, like data quality roles and responsibilities. In accordance with [11], we exclude from this broad concept, due to their specificity, the *Information Life-Cycle Management* and the *Information Security and Privacy* issues.

Methodologies for Data Quality Assessment and Improvement

According to [1], data quality methodologies (DQm) apply two types of strategies in their improvement activities: data-driven and process-driven, although some of them adopt mixed ones. Roughly speaking, data-driven strategies improve the quality of data by directly modifying their value, whereas process-driven strategies improve quality by redesigning the processes that create or modify data.

Although the various methodologies use different strategies, phases, activities and data quality dimensions, they ordinarily have two main common phases: assessment and improvement. In the assessment phase, a diagnosis of data quality, along the relevant quality dimensions, is performed using adequate data quality tools (DQt). Improvement mainly involves: a) identifying the causes for errors; b) correcting errors using appropriate DQt and c) redesigning the processes that create or modify data in order to improve their quality. Batini et al. [1] present and compare some of the most widespread methodologies.

Data Quality and Data Governance Maturity Models

Although Data Governance does not equal Corporate Data Quality Management, it has been pointed out in the introduction section, that the two concepts are closely related.

There has been limited research about the instruments to assess the progress and performance of DQM initiates, usually named data quality management maturity models[1] (DQMM), the exception being, to our best knowledge, the models developed in [4] and [21].

---

[1] In line with the Capability Maturity Model Integration (CMMI) from Software Engineering Institute

From the consultants' viewpoint, there is the DQMM from [7], and several Data Governance Maturity Models (DGMM), like IBM Data Governance Council Maturity Model [11] and Gartner Enterprise Information Management Maturity Model [16].

The DQ maturity level (DQML) of the case studies will be, although superficially, assessed using [7], [11] and [16] levels characterization.

DATA COLLECTION, RESULTS AND DISCUSSION

Considering the exploratory nature of this research study, as well the very nature of the research questions and the holistic nature of data quality management, the case study research method has been chosen to carry out the study.

Three cases have been chosen, in order to replicate and cross-case the findings. Each individual case is a holistic one [32], its unit of analysis being one corporation.

The organizations have been selected under the following conditions:
- Having an ongoing Data Quality Initiative;
- Operating in markets with different needs for data risk management and compliance, and different levels of value creation through data assets.

All the three organizations have headquarters in the same European Union Country (EUC), and will be henceforth designated by fictitious names. MyTelecom operates in the fixed and mobile telecommunications market, PublicBank operates mostly in the financial market and CentralBank is the Central Bank of that country.

Data Collection

The research questions being *which business drivers motivate the organizations to engage in a data quality management initiative?*, *how do they implement DQM?(sponsorship, scope, roles, projects, operations, methodologies, techniques and tools,…)* and sidewards *which objectives have been achieved*, so far?, the followed propositions have been stated [32, p. 26]:

1 *The main business drivers to DQM are cost control, risk management and revenue optimization [8];*
2 *The main candidate areas for DQM are business networking, customer management, decision-making/business intelligence and regulatory compliance [18];*
3 *DQM activities are mainly sponsored by IS/IT executives;*
4 *The organizational positioning of DQ Teams is directly related to the Organizations' DQ maturity level (DQML).Adapted from[7]:*
   4.1 *In Organizations with low to medium DQML, DQ Teams report to the Applications Development Unit of the IS/IT;*
   4.2 *In high DQML Organizations, DQ Teams report to the CIO or even to the Board of Directors*
5 *The level of Organizations' commitment to DQM depends on the Industry they operate within, represented by the mode they are in"DG strategic impact grid" presented in Fig. 3 and adapted from [17].*

In order to organize the data collection, we prepared a draft of the relationships between main data quality concepts and data quality roles and responsibilities, using a class diagram [3].
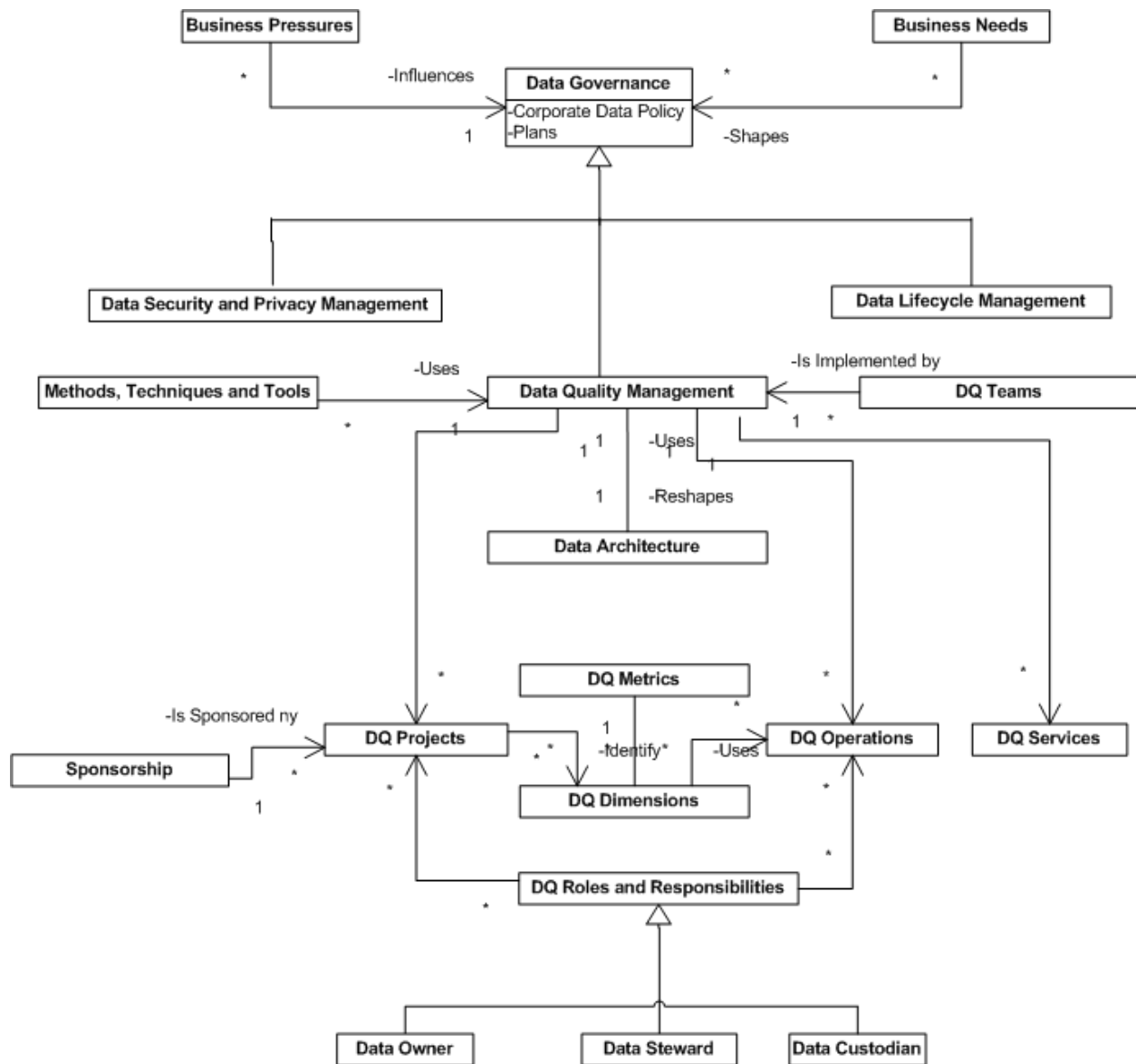
Fig. 2 - Draft of Relationships between main Concepts and Roles/ Responsibilities in Data Quality Management

Proposition 5 has been drawn according to the following criteria:
1. Some recent works [30, 14] have drawn parallelisms between IT Governance and Data Governance, although the reasons for that decision were not explained;
2. Although IT Governance (ITG) is different from Data Governance (DG), they have in common the most important component of ITG, which is *data (and information)*. In fact, hardware, software and communications are increasingly commoditized, and can only be considered as vehicles to store and manage what is really important: the data assets[2]. Therefore, the most important assets that ITG is supposed to govern is data (and information), as well as people and processes.

---

[2] See "Letters to The Editor" concerning Nolan & McFarlan paper in Harvard Business Review, 84 (2), pp. 155-157.
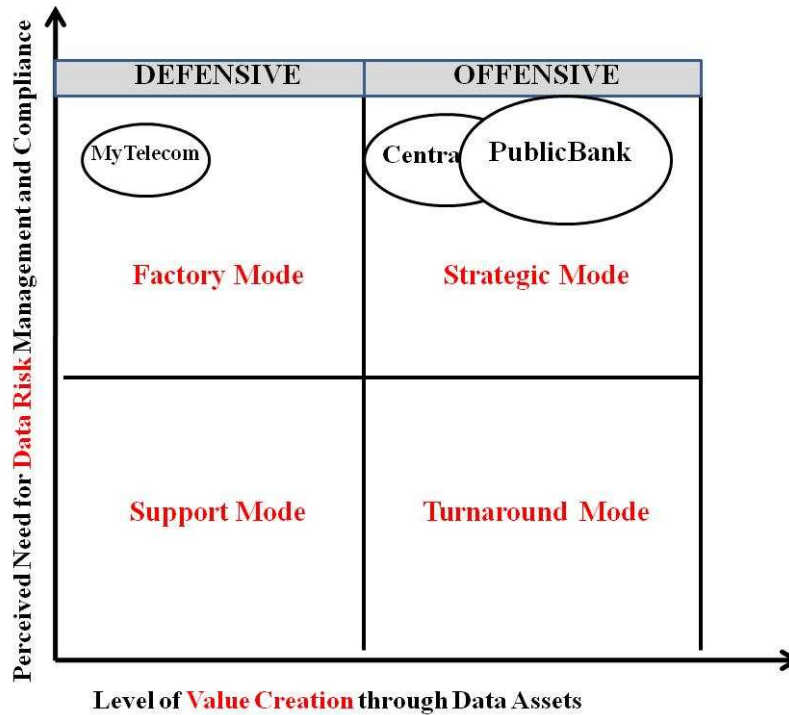
Fig. 3 – The DG Strategic Impact Grid. Adapted from [17]

This grid is a contingency approach to data governance and has two dimensions, which we borrowed from the DG outcomes presented in [11, p. 8-10]. Reference [11, p. 10] defines **value creation** as "the process by which data assets are qualified and quantified to enable the business to maximize the value created by data assets", that depending on how "data intensive" is the Industry the business operates within. **Data risk management and compliance** is defined as "the methodology by which risks are identified, qualified, quantified, avoided, accepted, mitigated or transferred out" [11, p. 10], which requires data with superior quality. The horizontal dimension represents the strategic impact level of data assets on business performance and the vertical one the need for "reliable" data. Organizations in the **support mode** "have both a relatively low need for quality data and a low value creation level by data assets. In **factory mode**, organizations need highly quality data for risk mitigation and compliance, but that data do not really create much value for business performance, whereas **turnaround mode** usually involves the strategic transformation of a business model supported by new data and does not need really highly reliable data. Once the change is made, organizations "move to either factory mode or strategic mode" [17, p. 101]. In **strategic mode,** organizations need high quality data and can create high value from that data.

MyTelecom
The Enterprise product and services catalog consists of mobile communications services (mobile and Internet), pre-paid, post-paid and fixed (telephone, digital television and Internet), which are provided by optical fiber structure or ADSL (Asymmetric Digital Subscriber Line). The company belongs to a national business group, with interests in multiple sectors all around the world.
Its mission stands out:
– "... Whose ambition is to be the best communications services provider in this country ..."
– "... Striving to consistently create products, services and innovative solutions that fully meet the needs of its markets and generate superior economic value."

MyTelecom has about 1200 employees and had, in 2009, a turnover of around Eur. 870 million and a consolidated result of about Eur. 5,250 million.

Data collection was done through two semi-structured interviews (which were recorded and transcribed) to the Quality Management Systems (QMS) Coordinator, by the observation of data governance reports (their content being analyzed below), followed by email exchanges to clarify some aspects and by analysis of its web site.
The evidences observed in MyTelecom are summarized in Table 2.

Perspectives	Empirical Evidences
Degree of Market Competitiveness	Very High.
Degree of Market Regulation	High.
Which Value does your Executive Leadership assign to Data?	An application resource that has some intrinsic value.
Data Policy (DP)	Does not exist
Drivers	The main driver was a decrease in Post Office costs related with the Business Process of Billing. A second driver was to improve the company image among its customers.
Beginning	The DQ project was launched in November 2007, and began to be focused on the name, address and zip code of the customers' attributes.
Sponsorship	The main project sponsor is CIO, although they are "winning" business sponsors at the directors and managers levels.
Scope of the DQ Initiative	Operational System − Customer Data supporting the Business Process of Billing
Data Quality Management Team	DQ Management is part of the Quality Management Systems (QMS) Team, that belongs to the Quality Technical and Business Information Systems Unit, which is one of the three Application Development Units of the IS/IT Department, and consists of about three people.
DQ Projects	DQ Assurance and Control of Customer Data supporting the Business Process of Billing − Cleaning of the attributes with the greatest impact on business; − Standardization of Addresses and Postal Codes; − Impact analysis of poor data quality attributes on business.
DQ Operations	On Customer Data supporting the Business Process of Billing. − Strong data validations have been implemented, but with great usability, like giving suggestions (best matching); − A centralized rules management has been implemented, which allows reusing standard validation rules by the various applications.
DQ Most Important Dimensions	Accuracy, Completeness, Relevancy
DQ Methodologies (DQm)	No formal DQ methodology has been adopted, so the method used is entirely empirical and based on intuition and "common sense".
DQ Techniques (DQT)	− *Data Driven*: techniques used are embedded in the tool adopted, in particular record deduplication and standardization; − *Process Driven*: enforcement of data entry validations.
DQ Tools (DQt)	Trillium – Discovery, Quality and Insight

Data Stewards (DS)	– As *Data Stewards have not been appointed yet*, the Requirements Analysis Teams have become "quality negotiators" between the various stakeholders; – Indeed, in the life cycle of the development projects, strong requirements should be demanded to validate the input and, at the limit, they may "even find a stakeholder who is aggrieved by another stakeholder 's decision".
Data Owners (DO)	Do not exist.
Master Data Management (MDM)	– Do not exist. They find it very difficult to implement policies and architectural options, such as DMS and MDM, in a very competitive industry, in which projects have very short time to market; – A reconciliation process runs daily between the Billing's Costumer Data and the Customer Relationship Management (CRM).
Communication Strategy	– Every other week, *data governance reports*, which present some information on the evolution of quality indicators (IQ), are sent by email, in newsletter format, to IS/IT professionals and business sponsors. The IQ of an entity is currently calculated as the average of the IQ of its attributes, with no weight. Tag clouds are used to show the attributes that contribute positively to the quality indicator (IQ) and the ones contributing negatively. The newsletter also provides news, such as "this issue out there": technical and management articles and they believe that this newsletter has been fairly helpful; – They organize *data governance awareness sessions*, in which they show the results of what they are doing, currently by role types - producers, custodians and customers; – Users are *trained* whenever necessary: a training session is organized for a new employees' group or when there are new versions/significant changes in applications that will impact the way data is entered. These training sessions are important to "educate" people on the data entry process.
Achievements	Until now, 51% of all customer records have been corrected (1.672.244 as in May 25th, 2010).
Benefits	They have not been calculated.
Costs	Only the costs associated with tools (acquisition and maintenance) and with external data quality consultants have been calculated, which have amounted to approximately Eur. 570 thousand, between 2008 and 2010.
Future Perspectives	– The tasks of data profiling, cleaning, standardization and enrichment are due to proceed, as well as the identification and modification of processes that induce data quality problems; – Data Stewards are expected to be appointed in the Business Units; – Their biggest challenge is to create an environment conducive to the acceptance of a Data Policy and Master Data Management.

Table 2 – MyTelecom's Data Quality Management Evidences

PublicBank

PublicBank is a market leader state-owned financial group, holding interests in the commercial banking, asset management, specialized credit, investment banking and venture capital, insurance, healthcare, real estate, etc., and having operations in European, African, American and Asian countries.

Its mission stands for:

▶ Consolidating its position as a structuring Group of the European Union Country (EUC) financial system, distinguished by strong relevance and responsibility in contributing to:
  – Economic development;
  – The strengthening of competitiveness, innovation and internationalization of the European Country companies;
  – The stability and soundness of the national financial system.

- As a market leader, finding a balanced development between profitability, growth and financial strength, always within a prudent risk management.

PublicBank has about 10600 employees and had, in 2009, a turnover of around Eur. 3.000 million and a consolidated result of about Eur. 279 million.

Data collection was done through one structured interview (which was recorded and transcribed) with the Architecture Unit Coordinator, the Coordinator and another member of the Data Management Group, which belongs to the Architecture Unit, one member of the Costumers' Development Team and the Coordinator of the Data Warehouse Team, followed by email exchanges to clarify some aspects and by analysis of the financial group web site. All the interviewees belong to the IS/IT Unit and the Data Management Group consists of five people.
Evidences collected from PublicBank are summarized in Table 3.

Perspectives	Empirical Evidences
Degree of Market Competitiveness	High.
Degree of Market Regulation	Very High.
Which Value does your Executive Leadership assign to Data?	A Critical Business Asset.
Data Policy (DP)	PublicBank is moving towards a DP.
Drivers	− The business drivers to DQ were the decrease in Post Office costs, the improvement of the company's image among its customers and the merger with another bank; − More recently data quality had an increased importance due to the need for Regulatory Compliance, in particular with Basel II and CobiT.
Beginning	− Approximately in 1993-1994 began the process of creating standards for defining data elements, glossary and centralized metadata management; − The data models analysis began later, around 1998; − PublicBank acquired the Metadata Repository in 1998, and the Data Modeling Support Tool (ERwin Data Modeler) in 2000; − The first project concerning data values quality began in 2004-2005, its target being the Customer Database; − Last year (2009) the Data Management Team (DMT), which has been in charge of corporate metadata definitions and data models quality, launched a cross-cut initiative concerning the data values quality.
Sponsorship	− The sponsor of the Customer Data Quality Project is the owner of the Customer Master Data and the sponsor of DW is its owner; − The Cross-cut Data Quality Initiatives are sponsored by one member of the Board of Directors of the Group's IS/IT Enterprise.
Scope of the DQ Initiative	Operational and Analytical Systems: − Customer Master Data; − Data Warehouse; − Corporate Metadata; − Data Models quality assurance; − Data Values quality assurance and control.
Data Quality Management Team	Although there is a Cross-cut Data Management Team (5 specialists) with DQ responsibility, DQ assurance and control is also performed by: − The Customer Team, dedicated to the Customer Data Quality (2-3 people from the Business Unit and 1-2 from the Applications Development Department of IS/IT);

	– Data Warehouse users and IS/IT DW Team. The Data Management Team belongs to the Architecture Unit which, in turn, depends directly on one member of the Board of Directors of the Group's IS/IT Enterprise.
DQ Services	– Corporate Business and Technical Metadata Management, and Data Models and Data Values Quality Assurance and Control.
DQ Operations	– Customer Master Data Quality Assurance and Control; – DW information testing and error detection by its users. These errors are always analyzed and their origin identified and, if coming from an Operational System, the team in charge is always informed of them.
DQ Most Important Dimensions	Accuracy, Timeliness, Completeness and Security (the latter dimension being managed by a Specific IS/IT Unit).
DQ Methodologies (DQm)	No formal DQ methodology has been adopted, although Data Management Team members have been inspired by TIQM [7].
DQ Techniques (DQT)	▶ Data Driven: – Metadata Definition; – Data modelling; – Techniques embedded in the used tool (Quality Stage), in particular record deduplication and standardization. ▶ Process Driven: As one of the main sources of DQ problems is data entry errors by producers, the data entry validations have been enforced. Now every customer 's update in the Costumer database calls for Quality Stage for name and address standardization. ▶ As the Country has recently launched a new Citizen Card with every relevant data stored in a chip, there will be fewer typos, since most relevant customer's data are directly collected from the digital support.
DQ Tools (DQt)	– *Clients' Team*: Quality Stage; – *Data Management Team*: Trillium (Discovery) for data profiling, CA Repository for Z/OS for Metadata Management, Erwin Data Modeler for Data Modeling and Erwin Model Manager for Data Models Management.
Data Stewards (DS)	Some people are carrying out a few data stewards' tasks in the bussiness unit owning the Customer Data (DSO).
Data Owners (DO)	The Operational Support Direction (DSO) is the owner of the Customers' Master Data and the Control and Planning Direction (DCP) is the owner of the Data Warehouse.
Master Data Management (MDM)	There is Centralized Customer MD, data services being made available to other areas. Some data are physically duplicated either synchronously or asynchronously. For example, CRM subscribes updates to the Customers MD.
Data Communication Strategy	The Data Management Team "sells its DQ projects" through the IS/IT Portal and meetings within IS/IT Enterprise and with the various Business Units.
Achievements	They do not know.
Benefits	They have not been calculated.
Costs	Thye have not been calculated.
Future Perspectives	Information Architecture aligned with the Business and Data Owners for each Master Data.

Table 3 – PublicBank's Data Quality Management Evidences

CentralBank

The CentralBank is part, since its inception in 1998, of the European System of Central Banks (ESCB) - which comprises the European Central Bank (ECB) and the national central banks of the European Union (EU) and is a part, since its inception in January 1, 1999, of the Eurosystem - comprising the ECB and national central banks participating in the euro.

CentralBank has about 1700 employees and had, in 2009, a consolidated result of Eur. 254 million.
Data collection was done through one structured interview (which was recorded and transcribed) with the Architecture and Innovation Group Coordinator and the Information Sharing System (SPAI) Team Leader, followed by email exchanges to clarify some aspects and by analysis of the CentralBank web site.

Evidences collected from CentralBank are summarized in Table 4.

Perspectives	Empirical Evidences
Degree of Market Competitiveness	Not applicable.
Degree of Market Regulation	Very High.
Which Value does your Executive Leadership assign to Data?	A Critical Business Asset.
Data Policy (DP)	They have a DP for the Analytical and the Operational Systems.
Drivers	The main driver was the reuse of data sources metadata by the analytical environments: statistics and supervision.
Beginning	The project began in 2002.
Sponsorship	The data quality management initiative was sponsored by the Applications Development Director and the IS/IT Assistant Director.
Scope of the DQ Initiative	Analytical Systems − Various data sources' metadata characterization, in order to induce their reuse. This scope has been chosen because it supposedly gives the best added value to the final product.
Data Quality Management Team	Information Sharing Team (SPAI) belongs to the Architecture and Organizational Innovation Group of the Applications Development Unit. SPAI is a flexible team, having a permanent staff of 1.5 specialists and using outsourcing services when necessary.
DQ Projects	− Treatment of sources: data conversion, examination of filling, etc; − Tools are provided for parameterized queries and data sources management.
DQ Services	− Guarantee of data delivery from sources provided by SPAI, according to defined SLAs; − Sources' Catalog and advice for their use, such as by identifying possible sources for a particular attribute. If a certain project does not want to refresh a specific source, than a copy of it must be done; − In some situations, SPAI Team helps customers increase their efficiency and effectiveness, such as by suggesting the use of common classification systems.
DQ Most Important Dimensions	Standard DQ dimensions are not used, the context of each project being analyzed and the DQ dimensions defined accordingly. Nevertheless, Data Reuse and Comparability (via common reference data) are the most common dimensions.
DQ Methodologies (DQm)	No formal DQ methodology has been adopted, although SPAI Team has defined some guidelines.
DQ Techniques (DQT)	Characterization of the Metadata of Sources.
DQ Tools (DQt)	No financial means have been allocated to SPAI Project, due to the lack of strong sponsorship. As a result, no commercial tool was bought and SPAI members have developed their own tools.
Data Stewards (DS)	Statistics Department defined some data steward tasks. But overall CentralBank has not implemented Data Steward roles.

Data Owners (DO)	– Whether in Analytical or in Operational Systems, every Information Unit has always a Department's Owner, who is responsible for its management. One Information Unit can either be a Table, resulting from the data processing of various sources or an Attribute of such a Table. – One Owner is assigned to a set of Information Units of the same Information Domain, eg. Currency Exchange, Securities and Territories.
Master Data Management (MDM)	There is a Centralized MDM for Information Domains managed by SPAI.
Communication Strategy	– SPAI Team began selling its project within the IS/IT Department; – Nowadays, they are selling it to the various Users' Departments, project by project; – SPAI is also planning to publicize their work in the Intranet.
Achievements	SPAI is presently in charge of managing between 20 to 30 information sources, which are reused by multiple projects.
Benefits	Benefits have not been calculated yet, but SPAI Team considers of great interest to assess to benefits generated by that system.
Costs	There have been calculated for internal staff only, but not for the consultants, who are assigned to projects. From January 2002 to May, 2010 internal staff costs have been around Eur. 905 thousand.
Future Perspectives	Continue integrating and enabling the reuse of more source' metadata .

Table 4 – CentralBank Data Quality Management Evidences

Results and Discussion

In this section, we are going to cross analise and discuss the most relevant findings.

Results

Value that Organizations' Executive Leadership assigns to data

PublicBank and CentralBank (which be henceforth named as Banks) executive leadership considers data as a Critical Business Asset, while in MyTelecom it is regarded as an Application Resource. That evidence is in line with the nonexistence of a Data Policy in MyTelecom and this first evidence points to considering MyTelecom in a lower Data Quality Maturity Level (DQML).

Business drivers to data quality

The main business driver to DQ were, in all Organizations, the reduction in non quality costs, particularly those related to "process failure" and "information scrap and rework" [7, p. 209-212], although none of them has calculated the benefits from DQ initiatives. Both findings are in line with previous literature [20, 7, 8]. The risk management and one merger were also the drivers to PublicBank DQ, that being also in line with literature.

Sponsorship

DQ initiatives have been sponsored, in the three cases, by IS/IT executives at different levels, except for the Customer Data Quality and DW projects in the PublicBank, which were launched and sponsored by its owners, that pointing to a higher PublicBank DQML. All the Organizations are actively looking for business sponsors for the DQ projects.

Scope of the DQ initiative

– The two commercial corporations (MyTelecom and PublicBank) are working in their Customer Data quality assurance and control and have enforced data validations to prevent DQ problems. This "Customer Focus" is in line with all the quality literature, and specifically with Standards ISO [12]. Nevertheless, the way they perceive customer data is quite different, as PublicBank uses a

Customer MDM framework, while MyTelecom does not;

- PublicBank DQM activities are also focusing on decision-making/business intelligence and regulatory compliance areas. Indeed, PublicBank adopted, three years ago, the IT governance framework CobiT, due to operational risk[3] internal control required by Basel II. CobiT includes a data management process (DS11), in the "Deliver and Support Domain".
- All the three Organizations are working in data values quality, although the Banks are the only ones working on metadata characterization, documentation and disclosure;

Data Quality Management Organization

- PublicBank's DQ Team is the best hierarchically positioned, as it is located within the Architectural Unit and occupies a 2nd level position in IS/IT Enterprise, while the other two have 3rd level positions within the Application Development Units of IS/IT Departments, belonging to the Quality Team in MyTelecom and to the Architecture and Organizational Innovation Group in Central Bank;
- Banks have assigned Data Owners and some Data Steward roles, although MyTelecom has not, that also pointing to a lower DQML for MyTelecom and a higher one for PublicBank.

Data Architecture Integration

- Both Banks are working towards MDM: Public Bank has a Costumer Master Data and CentralBank has centralized some "Information Domains";
- MyTelecom finds it "very difficult to implement policies and architectural options, such MDM, in a very competitive industry, in which projects have very short time to market";
- These evidences point towards a higher DQML for Banks.

DQ Methodologies, Techniques and Tools

- None of the three organizations uses a formal DQ methodology, although Banks are using some guidelines inspired in published methodologies. MyTelecom and PublicBank are enforcing data validations, both Banks are working on metadata definitions, and PublicBank pays special attention to data models quality assurance. MyTelecom and PublicBank use commercial DQ tools, while CentralBank developed its own tool, due to the reasons explained in Table 4.

DQ Dimensions

With regard to the most important DQ dimensions, they depend on each Corporation and, in CentralBank, they depend on each project context, that being in line with previous research findings [15].

Data Communication Strategies

All the three Organizations are communicating their DQ projects, either within IS/IT Units or to Business Units, MyTelecom having the strongest and most innovative DQ communication strategy, led by its DQ Team leader, who is clearly a data quality champion.

Benefits and Costs

The benefits from DQ initiatives have not been calculated and costs have only been partially calculated.

Discussion

Chronologically, PublicBank was the first Organization to launch a DQ initiative (in 1993-1994), well before CentralBank that began its DQ efforts in 2002, MyTelecom being the last one, in 2007. Indeed, it seems that MyTelecom is just running a DQ business case, while Banks have a much broader DQM scope.

All DQ projects and operations are dealing with the *fixing of non-quality data,* like *data cleaning and standardization,* and *preventing DQ problems*, through *data validation improvement* and *corporate metadata description and standardization*, as well as *quality assurance of data models*.

None of the organizations uses any formal DQ methodology, but two of them consider that they have some methodological influences. They are using commercial tools for data profiling, standardization and

---

[3] The Basel Committee defines operational risk as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events."

cleaning, except CentralBank, of which the tools have been internally developed due to the lack of financial resources.

Banks' Executive leadership consider data as a Critical Business Asset, and they are *working towards a Data Policy*, in line with Business Policy, which is not the case of MyTelecom, that difference also applying to Master Data Management and to appointment of Data Owners. In fact, PublicBank has a Customer and Data Warehouse Master Data Management, and has appointed Data Owners for that Master Data, the same happening to CentralBank, as refers to Analytical Master Data.

We found out that the main business drivers to the DQ initiatives have been the *reduction in non quality costs* and, afterwards, the *improvement of the company's image among its customers*, a *merger* and *risk management*, in particular the compliance with regulatory issues, which is in line with literature [7, 8, 20] and confirms proposition 1, except for *revenue optimization*.

MyTelecom and PublicBank are investing in *customer management*, which is in line with literature [18], PublicBank also operating in *decision making/business intelligence and regulatory compliance*. CentralBank, which mainly works with analytical systems, is operating in *data sources metadata characterization and reuse*. This confirms proposition 2 except for *business networking*.

Sponsorship has been ensured at different levels of the IS/IT Departments, which confirms proposition 3, PublicBank also having business sponsorship for its Customer Master Data and Data Warehouse quality projects.

The three organizations are trying to raise IS/IT and business people's awareness for non-quality data problems and communicating their DQ project achievements, MyTelecom being the champion of such communication initiatives.

In MyTelecom and CentralBank, DQ Teams have 3rd level positions within the Application Development Unit of IS/IT Departments, belonging to the Quality Team in MyTelecom and to the Architecture and Organizational Innovation Group in Central Bank.PublicBank's. PublicBank DQ Team is the exception, its DQ Team being located within the Architectural Unit and occupying a 2nd level position in IS/IT Enterprise which, according to [7], denotes a higher maturity level, and seems to confirm proposition 4.

CONCLUSIONS

In this section we will discuss the findings related with the initial research questions: *which business drivers motivate the organizations to engage in a data quality management initiative? how do they implement Data Quality Management?* and sidewards *which objectives have been achieved, so far*.

Fig. 3 presents our attempt to locate the three Organizations in the "DG strategic impact grid". Although we tried to gain feedback from the DQ team leaders of the three organizations about their positioning in the grid, it has been only possible to have the opinion of MyTelecom, who agreed with its positioning. Considering that the organizations are properly positioned in the grid, it can be deduced that, although all of them need quality data for their operations, Banks' business performance is also highly dependent on their data assets, which can justify their higher commitment to data quality management activities.

The differences that were found lead us to believe that PublicBank and CentralBank have an higher DQML maturity level than MyTelecom, mainly due to the aspects as follows: data being considered as a Critical Business Asset by executive leadership; the business commitment to DQ and MDM [11, p. 12]; the appointment of Data Owners; the importance assigned to Metadata characterization, documentation and disclosure [11, p. 12] and to their strategy towards a Data Policy.

Considering the PublicBank compliance with an IT governance framework[4], as well as its DQ Team

---

[4] Although the relationship between IT Governance and Data Governance has not been researched yet, to our best knowledge.

hierarchical position, apart from the fact that their Customer and Data Warehouse DQ projects were sponsored by a Business Unit (and not by the IS/IT Department, as the other ones), leads us to consider that this Corporation is in a better DQML than CentralBank. MyTelecom's DQML seems to be the lower one, due the reasons presented above.

These findings are in line with the organizations' positioning in the "DG strategic impact grid" (Fig. 3), MyTelecom occupying the "Defensive Column", due to the fact that data (and information) may not be a major differentiating factor, since an important number of its mobile customers (phone and Internet) are pre-paid and therefore their identification is not business significant. Banks are placed in the "Offensive" column due to the "higher need for value creation through data assets", their business being mostly conducted by knowledge workers through information analysis, their effectiveness and efficiency depending, to a large extent, on the quality of data (and information) available to them.

All the three Organizations are highly placed in the vertical axis, that meaning they are working in very regulated markets.

LIMITATIONS

Although the draft of each case study report has been reviewed by all its informants, more sources of evidence could have been used, if more time and resources were available, which could increase the construct validity [32] of each case.

FUTURE WORK

This research work made it possible to collect rich empirical evidence, which resulted in a better understanding of the Corporate Data Quality Management phenomenon in three Organizations of one European Union Country.

We intend to deepen this research work by testing the hypotheses perceived herein through a survey, to be administered to IS/IT and business managers from Organizations involved in DQ projects, enabling the development of a "Corporate Data Quality Meta-Framework" (CDQMF), drawing on contingency factors and best practices. This CDQMF is intended to support Organizations in the development of their own "Corporate Data Quality Frameworks".

ACKNOWLEDGEMENTS

REFERENCES

[1] Batini, C. Cappiello, C. Francalanci, C and Maurino, A. "Methodologies for data quality assessment and improvement". *ACM Comput. Surv 41* (3). 2009. pp. 16:1-16:52.

[2] Berson, A., Dubov, L. *Master Data Management and Customer Data Integration for a Global Enterprise*. McGraw-Hill. 2007.

[3] Blaha, M. & Rumbaugh, J. *Object-Oriented Modeling and Design with UML*. 2nd ed. Pearson Education, NJ, 2005.

[4] Caballero, I., Caro, A. Calero, C. & Piattini, M. "IQM3: information quality management maturity model". *Journal of Universal Computer Science, 14* (22). 2008. pp. 3658-3685.

[5] Drucker, P.F. *A Functioning Society*. Transaction Publishers.Piscataway, NJ. 2003.

[6] Drucker, P.F. *The New Realities*. Harper & Row. New York, NY, 1989.

[7] English, L. *Improving Data Warehouse and Business Information Quality*. Wiley & Sons. 1999.

[8] Fisher, T. *The Data Asset – How Smart Companies Govern their Data for Business Success*. John Wiley & Sons. Hoboken, NJ, 2009.

[9] Friedman T. & Bitterer, A. *Gartner Magic Quadrant for Data Quality Tools*. 2009. Retrieved January 29, 2010 from http://www.gartner.com/technology/media-products/reprints/dataflux/167657.html

[10] Hoffer, J.A., Prescott, M.B., Mcfadden, F.R. *Modern Database Management*. 8th edition. Pearson Education Inc. Upper Saddle River, NJ, 2007.

[11] International Business Machines. *The IBM data governance council maturity model: Building a roadmap for effective data governance*. Technical report. IBM Software Group. 2007.

[12] International Standards Organization. *ISO 9000:2005 (E) - Quality Management Systems: Fundamentals and Vocabulary* . 3rd edition. 2005.

[13] Juran, J. M. *Juran on Leadership for Quality: An Executive Handbook*. The Free Press. NY, 1989.

[14] Khatri, V & Brown, C. V. "Designing Data Governance". *Communications of the ACM, 53* (1). 2010. pp. 148-152.

[15] Lucas, A., Palma-dos-Reis, A., Caldeira, M. "The Quality of Monitoring Data in Civil Engineering Works", *14th International Conference on Information Quality (IQ 2009)*. 2009.

[16] Newman, D., Logan, D. *Gartner Introduces the EIM Maturity Model (ID Number G00160425).* Gartner Research, 2008.

[17] Nolan, R. & McFarlan, F. W. "Information Technology and the Board of Directors". *Harvard Business Review, 83* (10). 2005. pp. 96-106.

[18] Otto, B., Wende, K., Schmidt, A.& Osl, P. "Towards a framework for corporate data quality management". *18th Australasian Conference on Information Systems.* 2007. pp.916-926.

[19] Radcliffe, J. *Gartner Magic Quadrant for Master Data Management of Customer Data*. Gartner. 2009.

[20] Redman, T. C. *Data Quality for the Information Age*. Artech House. 1996.

[21] Ryu, K.-S., Park, J.-S., Park, J.-H. "A data quality management maturity model". *ETRI Journal, 28* (2). 2006. pp. 191-204.

[22] Sveiby, K. *The new organisational wealth - Managing and measuring Knowledge-Based Assets.* Berret – Koehler.San Francisco, 1997.

[23] Tee, S. W., Bowen, P. L., Doyle, P., Rohde, F. H. "Factors influencing organizations to improve data quality in their information systems". *Accounting and Finance, 47*. 2007. pp. 335-355.

[24] The Customer Data Integration Institute. *Corporate Data Governance Best Practices - 2006-07 Scorecards for Data Governance in the Global 5000*. 2006. Retrieved April 20, 2010, from http://www.tcdii.com/images/Data_Governance_white_paper_-_April_2006.pdf

[25] The Data Management Association. *DAMA DMBOK Functional Framework.* Version 3.02. DAMA International. 2008b.

[26] The Data Management Association. *The DAMA Dictionary of Data Management*. Technics Publications LLC. Bradley Beach, NJ, 2008a.

[27] The Data Warehousing Institute. *Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data*. TDWI. Seattle, WA, 2002. .

[28] Wand, Y. & Wang, R. Y."Anchoring Data Quality Dimensions in Ontological Foundations". *Communications of the ACM 39* (11). 1996. pp. 86-95.

[29] Wang, R.Y.& Strong, D.M. "Beyond Accuracy: What Data Quality Means to Data Consumers". *Journal of Management Information Systems, 12* (4). 1996. pp. 5-34.

[30] Weber, K, Otto, B. & ¨Osterle, H. "One size does not fit all—a contingency approach to data governance". *ACM J. Data Inform. Quality, 1*(1). 2009. pp. 4:1-4:27.

[31] Wende, K. "A model for data governance – organizing accountabilities for data quality management", *18th Australasian Conference on Information Systems*. 2007.

[32] Yin, R. K. *Case Study Research: design and methods*. 3rd edition. Sage Publications Inc. Thousand Oaks, CA, 2003.