School of Economics and Management
TECHNICAL UNIVERSITY OF LISBON

TOWARDS CORPORATE DATA QUALITY MANAGEMENT

Ana Lucas
National Laboratory for Civil Engeneering and ISEG School of Economics and Management

Abstract

Today, it is a well known fact that poor quality data is costing large amounts of money to corporations all over the world. Despite the increasing research on methods, concepts, and tools for data quality (DQ) assessment and improvement, little has been done about corporate DQ management. The purpose of this research is to understand the nature and complexity of corporate DQ management, through various perspectives. These include the various kind of sponsorship, type and level of collaboration between business and IS/IT, organizational position of the DQ management team, scope of the DQ initiatives, roles, services provided, DQ methodologies, techniques and tools in use, etc. This paper presents, analyzes and discusses a single pilot exploratory case study, undertaken in a fixed and mobile telecommunications company in a European Union Country. The purpose of this case study is to check the validity of some initial propositions, and eventually find new ones, to be used in a subsequent multiple-case study, in order to provide an in-depth understanding of the corporate DQ management phenomenon.

Keywords: data quality management, case study

1. 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 leads to wastage of resources and gross damage to the organization externally, particularly in the area of customer relationships. Just to get an idea of the cost of poor data quality, we give a figure. The Data Warehousing Institute [TDWI] (2002) has estimated that current data quality problems cost U.S. businesses more than USD 600 billion a year.

Correspondence Address: Ana Lucas – National Laboratory for Civil Engineering – Information Technology Center – Av. do Brasil, 101 – 1700-066 Lisboa, Portugal. E-mail: ana.lucas@lnec.pt

According to Otto, Wende, Schmidt &Osl (2007), *business networking, customer management, decision-making/business intelligence and regulatory compliance* are some of the main areas inducing corporate data quality management.

One of the most commonly quoted definitions for *data quality* (DQ) comes from Juran (1989), for whom data are of high quality if they are fit for their intended usage 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. (2007, 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 (1989, 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*. The term *knowledge worker* was first coined by Peter Drucker in 1959, and it must be pointed out that current literature presents multiple definitions, although they usually only differ in small details. Sveiby (1997) considers knowledge workers as those who are highly qualified and highly educated professionals, and their work consists largely in converting information into knowledge, using their own competencies for the most part, sometimes with the assistance of suppliers of information or specialized knowledge. According to Drucker (2003, 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, it 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 data and information interchangeably in the context of this paper.

Generally speaking, DQM can be defined as the "quality-oriented management of data as an asset" (Weber *et al.*, 2009, p. 4:4). This essentially alludes to "the application of total quality management (TQM) concepts and practices in order to improve data and information quality, which includes 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 [DAMA], 2008a, p. 43). In order to be effective, data quality management must go beyond the activities of *'fixing non-quality data'*, to *'preventing data quality problems'* by effectively managing data over its entire 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. This requires a profound *cultural change* across the spectrum of *leadership*, authority, control and allocation of resources, which means governance, specifically *data governance*. With data governance, companies are able to implement corporate-wide accountabilities for DQM, encompassing professionals from both business and IT units (Wende, 2007).

According to English (1999), the *costs of poor data quality* can be classified into three categories:

- *Process failure costs*, which occur when processes do not perform properly due to poor quality data, such as costs associated with wrongly delivered or undeliverable mail due to inaccurate mailing addresses;
- *Information scrap and rework*, such as costs associated with resending mail (rework) 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 cannot be reached in advertising campaigns, resulting in lower revenues. Another example can be a customer loss due to incorrect billing.

Bitterer & Newman (2007) present, by business function, some impacts of poor quality data, which are in line with the one's presented by English (1999). The former add two new cost categories, which are budget overruns, fines and possible jail time that can occur due to the lack of financial data quality. That work also presents savings obtained through data quality management, by corporations operating in different business industries.

Considering the type of our research questions- which are 'why do corporations engage themselves in a data quality management initiative', or put another way, 'which are the drivers to such initiative', and 'how do they implement DQM (sponsorship, scope, roles, services, methodologies, techniques and tools,...)', and the little research literature concerning these subjects, we have decided to use the case study as our research strategy. This is because: a) we are going to investigate a "contemporary phenomenon within its real-life context" (Yin, 2003, p. 13); b) our questions are of the type "how" and "why", so we need to understand the nature and complexity of DQM implementation; c) little research has been carried out about this subject (Benbasat, Goldstein and Mead, 1987, p. 370). We will begin with a pilot exploratory case study.

This is a 'pilot', because we are at "the outset of theory generation" about the DQM implementation phenomenon (Benbasat *et al.*, 1987, p. 373) and 'exploratory' because there is very little in the present literature about the subject. We will proceed with a multiple-case study across different businesses. The initial case study is for evaluating and possibly modifying the research propositions and testing the overall research design, and will be followed by multiple case studies to provide an in-depth understanding of the "ways" corporations implement DQM-like sponsorship, the organizational position of the DQM team, the scope of the DQ initiative, roles, services provided, DQ methodologies, techniques and tools, etc.

As telecommunications and banking organizations usually collect and manage huge amounts of data, we have decided that our case site should be an organization in these industries having an ongoing Data Quality Project. Accordingly, we contacted some of the main telecommunications and banking organizations in a specific EU Country. Very few firms accepted our request, and we have chosen a telecommunications company for our pilot case study, that will be addressed by the fictitious name of My Telecom

This paper is organized as follows: In the background section we present a framework for data management related concepts and roles/responsibilities, as well as the data quality concept and it's most important dimensions, together with a brief presentation of some approaches and methodologies used by organizations to assess and improve the quality of their data. The case study is presented, analyzed and discussed in the third section, which is followed by the conclusions, the limitations of the work and some guidelines for future research.

2. BACKGROUND

This section involves an introduction to the fundamentals underlying the work, namely the presentation of a framework for the concepts related to data management, the data quality concept and its most important dimensions, as well as some approaches and methodologies used by organizations to assess and improve the quality of their data.

2.1 Data Management Approach

The *data management approach* comprises all the disciplines and functions related to managing data as a valuable resource. According to DAMA (2008a, p. 39), "*data management* is:

1. The business function that develops and executes plans, policies, practices and projects that acquire, control, protect, deliver and enhance the value of data and information;

2. A program for implementation and performance evaluation of the data management function;
3. The field of disciplines required to perform the data management function;
4. The profession of individuals who perform data management disciplines;
5. In some cases, a synonym for a 'Data Management Services Organization' that actually performs data management activities."

Literature, whether from academic or professional sources, presents a set of concepts and roles, as well as their definitions, related to the management of data which is considered as a corporate asset. In order to clarify and organize these concepts, we define a philosophy underlying all of them and, in Table I, present a (non exhaustive) list of concepts and roles/responsibilities related to the data management approach and in Figure 1 some of their interrelationships.

TABLE I

Concepts and Roles/Responsibilities for the Data Management Approach

Concept or Role/ Responsibilities	Definition
Corporate Data (CDPh) Philosophy	*Corporate Business Philosophy* is a long term corporate *vision* and consists of a *set of values* that have to be considered above all kinds of policies, strategies, roles or technologies. A *Corporate Data Philosophy*, in line with *Corporate Business Philosophy*, considers data as a corporate asset across the organization, which means that "it turns its focus away from the expense associated with acquiring, managing and storing data, and focuses on the business value that can be obtained from using the data and its full strategic lifecycle" (Hewlett Packard [HP], 2007, p. 4).
Corporate Data Policy (CDP)	CDP recognizes CD philosophy and defines the broad guidelines governing data, such as: a) data must be shared and reused in order to support cross-process integration or, in other words, transformation of data property from being departmental or personal to being corporate; b) prescribes the maximization of the value created by data assets. According to Redman (1996), CDP must cover the following interrelated categories: – DQ dimensions that apply to each particular dataset; – Data assets catalog; – Data sharing, availability and accessibility; – Data architecture; – Data security and appropriate use; – Data planning.
Corporate Data Strategy (CDS)	A CDS is a long term plan of action designed to achieve the directions prescribed by CDP in line with the *Corporate Business Strategy.*

Continue

Concept or Role/ Responsibilities	Definition
Data Governance (DG)	– DG is the "formal orchestration of *people*, *processes*, and *technology* to enable an organization to *leverage data as an enterprise asset*" (The Customer Data Integration Institute [TCDII], 2006, p. 1). – 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" (DAMA, 2008b, p. 38). – In our opinion, DG mainly refers to: a) *strong leadership* over the management of data assets; b) defining *corporate data strategy* (CDS), in line with CDP; c) providing resources and organizational structures to facilitate the implementation of the strategy and goals d) cascading CDS and goals down into the organization.
Data Quality Management (DQM)	– 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" (DAMA, 2008a, p. 43). – Moreover, adapting the vocabulary presented in ISO (2005) to *data quality*, we can define DQM as a set of coordinated activities to direct and control data quality in an organization.
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 right from input information describing a given application context, defines a rational process to assess and improve the quality of data" (Batini, Cappiello, Francalanci & Maurino, 2009, 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*. 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, ie, different records that represent the same real world entity) or standardization techniques (comparing data with lookup tables, and updating it accordingly). Process driven techniques are used to describe, analyze and reengineer the information production processes, and they are mainly of two types: process control and process redesign (Batini *et al.*, 2009).
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 Friedman and Bitterer (2009).
Data Quality Assurance (DQA)	Data Quality Assurance (DQA) is that part of data quality management which is focused on providing the confidence that quality requirements will be fulfilled (adapted from ISO (2005) 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 (2005) to DQM).

Continue

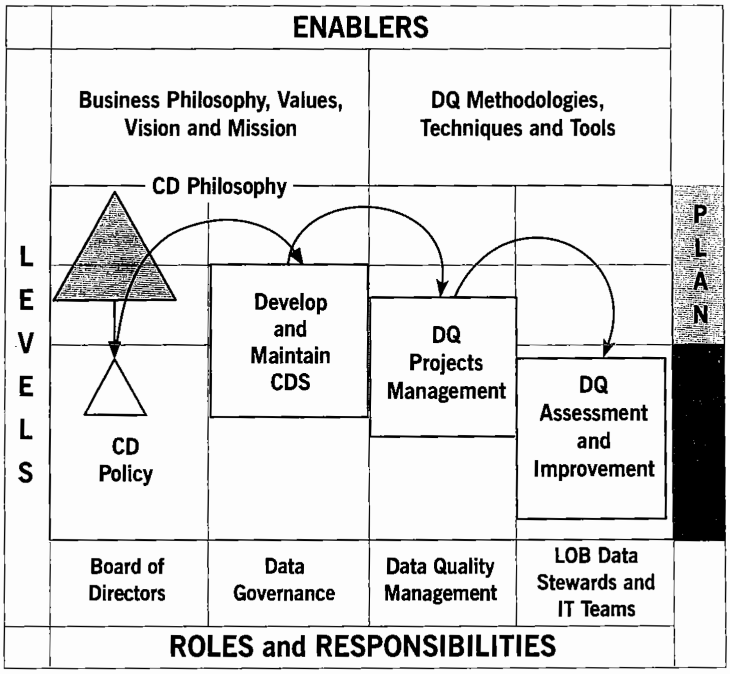
Concept or Role/ Responsibilities	Definition
Master Data (MD)	– Master Data is "the consistent and uniform set of identifiers and extended attributes that describe the core entities of the Organization and are used across multiple business processes". Examples are Customers and Employees (Radcliffe, 2009, 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 (Berson & Dubov, 2007).
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" (Radcliffe, 2009, p.3). – MDM is the framework of processes and technologies aimed at creating and maintaining an authoritative, reliable, sustainable, accurate, and secure data environment that represents a "single version of truth", constituting itself 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 (Berson & Dubov, 2007). – 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 (2008a, 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 Tee et al. (2007, p. 338), DQC "are managers who actively and vigorously promote their personal vision for using data quality related technology innovations". They push projects through approvals, provide political support, keep participants informed, and allocate resources to data quality projects.

Fig.1 presents some of the main data management related activities at the strategic, tactical and operational levels, as well as the corresponding enablers and roles and responsibilities involved at each level.

Organizations that consider data as a corporate asset, have a quality culture, and are at a higher DQ maturity level, tend to run activities in all levels, while those who are just solving data quality problems in specific datasets, being at a low DQ maturity level, tend to run only activities at the operational and perhaps tactical level, which is what is happening, as we shall see, in our case study.

FIGURE 1

Data Management Approach Framework



2.2 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 (Juran, 1989). Although widely accepted, this broad definition needs in depth specification, as data quality is presented in the literature as a multidimensional concept (Wand & Wang, 1996; Wang & Strong, 1996; Redman, 1996), and operationalized through its dimensions, which are data characteristics that are valued by data consumers, like accuracy, timeliness, understandability, completeness, relevancy, etc.

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

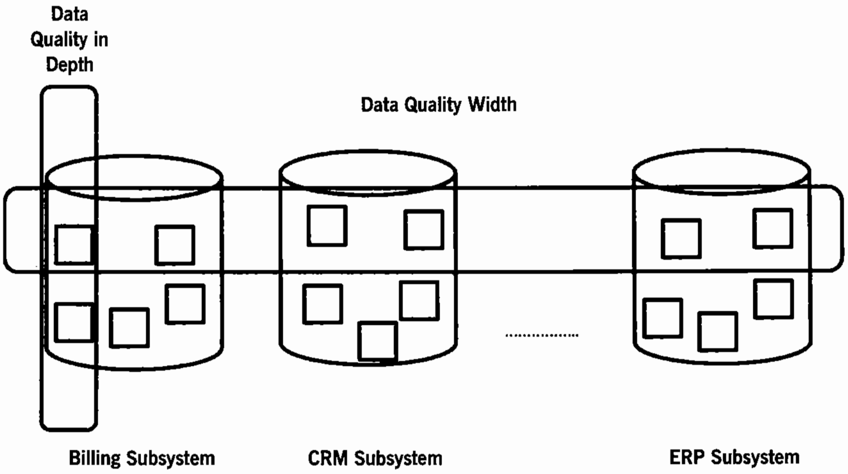
a) Theoretical (Wand & Wang, 1996);
b) Empirical (Wang & Strong, 1996) and
c) Intuitive (Redman, 1996).

These approaches refer to both the data in extension, i.e., *their values* (Wand & Wang, 1996; Wang & Strong, 1996; Redman, 1996; English 2009) and in intention, i.e., their models or *database schemas* (Redman, 1996; English, 2009). Moreover, some authors (Wang & Strong, 1996; Redman, 1996; English, 2009) still consider data *presentation* dimensions, like understandability as important. With regard to dimensions of data models, either the professional or the research literature only considers *intra-data models dimensions* (those related to the schema of a database supporting some specific applications).

Corporate data quality brings new challenges that have not hitherto been provided by the various approaches. In line with Ryu, Park & Park (2006), we stress that corporate data quality also depends on *inter-data models issues* or, to put it differently, on *corporate data architecture*, which requires consideration of the inter-data models' quality dimensions, as they will ensure the *corporate data model integration and consistency*, as presented in Fig. 2. For example, although the Billing Subsystem has adequate customer DQ, CRM and ERP, It will not do, if only in depth data quality is being considered.

FIGURE 2

Corporate Data Quality. Adapted from Ryu, Park & Park (2006)



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

Table II presents the most significant "in extension" Data Quality Dimensions.

Beyond what was said, based on our own research (Lucas et al., 2009), as well as on other researchers' work, we are convinced that data quality dimensions

TABLE II

"In extension" Data Quality Dimensions

Group of Dimensions	Category	Dimension	Definition: The extent to which data or business	Authors
Data Values	Intrinsic	Accuracy	are correct, reliable end certified free of error	Wang & Strong (1996)
		Coherence	are compatible with each other and satisfies the rules applicable to it	Lucas, Palma-dos-Reis & Caldeira (2009)
	Contextual	Relevancy	are aplicable and useful for the task at hand	Wang & Strong (1996)
		Timeliness	has appropriate age for the task at hand	Wang & Strong (1996)
		Completeness	are of sufficient depth, breath and scope for the task at hand	Wang & Strong (1996)
		Preservation	can be accessed and interpreted by any user or system in the context in which they were created	Lucas, Palma-dos-Reis & Caldeira (2009)
Data Presentation		Uniformity	are always presented in the same "look and feel" format	
		Understandability	are presented in an ease, intuitive and un-biased format	English (2009)
Data Access		Accessibility	are easily and quickly retrieved	Wang & Strong (1996)
		Security	has its access resctrited when necessary	Wang & Strong (1996)

and their relative importance highly depends on the specific field of application (*domain specific dimensions*). Therefore, each organization should choose the dimensions that fit each *specific data domain and user role*, adapt the definitions and define their metrics and measurement methods (Redman, 1996).

Batini *et al.* (2009) and Batini & Scannapieco (2006) provide comparable definitions for data quality dimensions and present and discuss other dimensions that are available in the literature.

Strong *et al.* (1997) identified three main roles within the information manufacturing system and associated a process with each role: a) *information producers* generate and provide data; b) *information custodians* are IT staff who provide and manage computing resources for storing, maintaining, and securing data; and c) *information consumers* access and use information for their tasks. Weber *et al.* (2009) deepens those roles and defines four types of interaction: responsible, accountable, consulted and informed.

2.3 Data Quality Management and Data Governance Maturity Models

As previously pointed out, Data Governance is a broader concept than Corporate Data Quality Management, and includes many other subjects, such as Data Protection and Data Security, although the two concepts have much in common.

There has been limited research about the instruments to assess the progress and performance of DQM initiates, usually named data quality management maturity models (DQMM). To our best knowledge, the exception are the models developed by Caballero, Caro, Calero & Piattini (2008) and Ryu *et al.* (2006).

From the consultants' viewpoint, there is the DQMM from English (1999) and several Data Governance Maturity Models (DGMM), like IBM Data Governance Council Maturity Model (IBM, 2007) and Gartner Enterprise Information Management Maturity Model (Newman & Logan, 2008)

The DQ maturity level of the case study will be, although superficially, assessed using English (1999), IBM (2007) and Newman & Logan (2008) levels characterization.

2.4 Methodologies for Data Quality Assessment and Improvement

According to Batini *et al.* (2009), data quality methodologies (DQm) apply two types of strategies in their improvement-related activities: *data-driven* and *process-driven*, although some of them might also adopt mixed ones. Roughly speaking, data-driven strategies improve the quality of data by directly modifying their values, 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, with regard to 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. (2009) present and compare some of the most widespread methodologies and Table III contextualizes some of the concepts mentioned above.

3. DATA COLLECTION, RESULTS AND DISCUSSION

3.1 The Company

The company studied here provides fixed and mobile telecommunications services and, for confidentiality reasons, will be designated by the fictitious name of MyTelecom. It operates in one of the European Union Countries, has about 1200 employees, and in 2009, had a turnover of around Eur. 870 million and a consolidated result of about Eur. 5,250 million. Its 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 offered

TABLE III

DQ Management in Context

HOW? / WHAT?	Data Quality Control	Data Quality Assurance
DQM Activities	Fixing non-quality data	Preventing data quality problems from arising
Maturity Stages	Reactive	Proactive
Data Quality Methodologies Strategies	Data-driven	Process-driven

over 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."

3.2 Data Collection

The research questions being *'why do corporations engage themselves in a data quality management initiative?'*, *'how do they implement DQM'* and *'which objectives have been achieved, so far?'*, the following propositions have been stated (Yin, 2003, p. 22):

1. The main drivers of DQM are cost control, risk management and revenue optimization (Fisher, 2009);
2. The main candidate areas for DQM are business networking, customer management, decision-making/business intelligence and regulatory compliance (Otto *et al.*, 2007);
3. The level of Corporations commitment to DQM is directly related to the value their Executive Leadership assign to data and to the degree of market competitiveness in which they operate;
4. DQM activities are mainly sponsored by IS/IT executives;
5. DQ Teams belong to the Application Development Unit of the IS/IT Department.

Data collection was done through two sources (Benbasat et al, 1987, p. 374): two semi-structured interviews (which were recorded and transcribed) to the Quality Management Systems (QMS) Coordinator, and by the observation of data governance reports (of which the content is analyzed below), followed by email exchanges to clarify some aspects. The QMS belongs to the Quality Technical and Business Information Systems Unit (QTEC&BIS) of the IT Department and consists of about ten people, three of whom are assigned to data quality tasks.

Our respondent is clearly a data quality champion (Tee *et al.*, 2007), i.e., someone who provides political support, keeps participants informed and allocates resources to data quality projects.

In November 2007, driven by the business process of billing, MyTelecom launched the "Enterprise Data Governance Initiative" within the IT, and began to focus on the name, address and zip code customers' attributes. Although there are many data quality problems at the Data Warehouse (DW), the IT Unit believes that improving data at the operational level will partially resolve the DW data quality problems.

The company has incurred extra post office costs due to poor quality of its customer data because, in spite of having a contract with the post office company stating that they must use all means to deliver their bills, including manual correction, this contract has non-negligible costs. They are hoping to reduce these costs by improving the quality of customer data as a result of this initiative.

IT Unit is responsible for the data quality management initiative, the main project sponsor being the CIO, although they are "winning" business sponsors at the directors' and managers' levels.

They do not have, till now, a Corporate Data Policy (CDP) or a Master Data Management (MDM) in place. Hence, a number of anomalies still exist in the data. They can still have, for instance, the same attribute with different codifications, such as the agent identifier; the "range" of codes that each department of marketing & sales (Corporate Marketing and SMEs) should use to identify "their" agents is still lacking a clear definition. This aspect is a risk and requires manual control that is dispensable. The company finds it very difficult to implement policies and architectural options, such as CDPs and MDM, in a very competitive industry, where projects have very short time to market.

Asked about the processes used under this initiative, our respondent identified them in the following order: Awareness, Exploration, Reporting, Fixing and Preventing.

He believes that the initiative was reactive at first, aimed at correcting customers' attributes. Nevertheless, they then began analyzing the root causes of the errors and so they are now, in his words, betting on proactive and reactive processes.

IT people believe that DQ problems are more of an IT problem, than what is typically considered in the literature. Indeed, in the life cycle of the development projects, strong requirements should be sought and consolidated to validate the input. At worst, they may "even find a stakeholder who is aggrieved by another stakeholder decision".

At the level of 'tools' and after assessing various proposals and their costs, they have chosen Trillium, although it seems obvious that this choice was mainly dictated by costs. Our interviewee informed us that MyTelecom will be looking very carefully at other available open source solutions.

No formal data quality methodology has been adopted, so the method they use is entirely empirical and based on intuition and "common sense".

The DQM team started the cleaning process with all clients, but then felt the need to target, "because customers do not all have the same value for the company": "fixed, identified and active customers"; "fixed, identified and inactive"; "fixed, unidentified and active (prepaid)", and so on. They have also assigned priorities to customers' attributes, so as to give priority to the cleaning process: 1) customer code; 2) tax identification number; 3) name, address, postal code; 4) email address etc and they are analyzing impacts of poor data quality on some attributes, like the one of having a client's incorrect civil state. Customers' addresses and postal codes are validated against Official Postal Office reference and, in the impossibility of doing so because the address does not exist in that reference (it can be a new address), and in case it is related to a fiber client, they confirm it at the ground.

Generally speaking, the most privileged dimensions are accuracy, completeness and relevancy, since these data quality dimensions are essential for operations and decision making.

They identify DQ problems through various means, such as, directly by consumers, remedy tickets, IT projects in the testing phase, and even through social networks, in particular twitter. Currently, data consumers can inform the application support person of all the non-conformities, the latter situation leading to strengthening input data validation.

Several sources of DQ problems have been identified, including data entry errors by producers, lack of data entry validations and data integration between systems. As validations are discouraged strongly by data producers, they decided to implement strong data validations, but with great usability, for e.g., they started giving suggestions (best matching). A centralized rules management has been implemented, which allows reusing standard validation rules by the various applications, and is based on regular expressions. This system is being implemented in several phases to validate various attributes, such as postal codes, dates, phone numbers, etc. The integrity rules are defined, whenever possible, at the data base management system (DBMS) level, releasing the applications from these tasks, in this case, Due to the DQ initiative, some business processes have been changed, in particular data entry validations and data transfer between the Billing system and the Customer Relation Management (CRM).

The organization has standards, rules and classifications for all development projects, the standard being to use, whenever possible, classifications that cut across all the development areas. On the other hand, there are standards and classifications, general and specific, for the development of specific data models associated with the projects.

Every fortnight, data governance reports, which present some information on the evolution of quality indicators (IQ), are sent by email, in newsletter format, to IT professionals and business sponsors. The IQ of an entity is currently calculated as the sum 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 which are contributing negatively. The newsletter also provides news, such as "this issue out there": technical and management articles. They think that this newsletter has been fairly helpful.

DQM team organizes 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 about the data entry process.

Until May, 25th, 2010 1.672.244 customer records have been corrected, which means 51% of all customers. The data is updated on the billing data base and then transfered to other systems. A centralized rules management has been implemented (see above), and they are "gaining allies" among business people in various areas and hierarchical levels.

DQM team is going to proceed with the tasks of data profiling, cleaning and enrichment, as well as with the identification and modification of processes that induce data quality problems. The big challenge is to create an environment conducive to the acceptance of a Corporate Data Policy and Master Data Management.

The evidences observed in MyTelecom are summarized in Table IV.

3.3 Results and Discussion

The main driver for MyTelecom to undertake the data quality management initiative was Customer Data, which is in line with the most relevant areas presented in the academic (Otto *et al.*, 2007; Umar *et al.*, 1999) and professional

TABLE IV

MyTelecom's Data Quality Management Evidences

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.
Corporate Data Policy (CDP)	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's image among its customers.
Beginning	The DQ project was launched in November 2007, and began with a focus 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.
Level of Collaboration between Business and IS/IT (Low, Medium, High)	Low.
DQ Services	Only provided to the Customer Data supporting the Business Process of Billing. • *Data Driven* – 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. • *Process Driven* – Strong data validations have been implemented, but with great usability, like giving suggestions (best matching); – A centralized rules management has been implemented, which allows re-use of 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".

Continue

Perspectives	Empirical Evidences
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, 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 Corporate Data Policy and Master Data Management.

(Fisher, 2009; Informatica, 2005) literature, and confirms the "customer management" part of proposition number 2. However, because they have neither a Corporate Data Policy, nor a Master Data Management approach, their DQM initiative is strictly related to the customer data supporting the business process of billing, their main objectives are to reduce costs (namely post office costs), which is also in line with literature (English, 1999; Fisher, 2009), and fits with the "cost control" part of proposition number 1.

MyTelecom's DQM Team belongs to one of the Application Development Units of the IS/IT Department, thus confirming proposition number 5. Proposition number 4 is also confirmed as DQM activities are being sponsored by the CIO.

MyTelecom has a low level commitment to DQM, which is observable through the absence of:

- Metadata Management;
- Master Data Management;
- Data Owners and Data Stewards Appointment;
- Corporate Data Policy.

That low level commitment to DQM is positively associated with the low value its executive leadership assigns to data, that data is only "an application resource" (and not "a critical business asset"), confirms the first part of proposition number 3 and disconfirms its 2nd part, as MyTelecom operates in a very competitive market.

Another finding is that in the absence of use of any formal data quality methodology (DQm), their DQ activities are only supported by a data quality tool (DQt).

Given the defined objectives, the DQ initiative seems to be a success, and they are also gaining some sponsorship on the business side, at the directors' and managers' levels and have thus boosted the understanding that DQ problems are much more an "IT problem" than is typically considered in the literature. This question deserves reflection on the basis of its justification, which places the requirements analysis teams of the "IT Development" as a "quality negotiator" between stakeholders. This is being justified because of the absence of data stewardship roles. We have caught the following message, concerning DQM: "We (IT) are in charge, please work with us and keep it simple"

DQ services are provided through the *fixing of non-quality data*, like *data cleaning and standardization*, and by the prevention of DQ problems, through data validation improvement.

We highlight their decision to target customers, in order to assign different priorities for cleaning and enrichment, as well as prioritizing attributes for improvement according to their usability. This enabled them to optimize the utility/cost trade-offs associated with DQ improvement, which is in line with recent DQ literature (Even and Shankaranarayanan, 2009).

Two other interesting architectural constructs are:

- The implementation of strong validations, but with very good usability, like giving suggestions;

– The creation of a centralized rules management repository that, in addition to helping the data quality improvement through centralized validations, facilitates greater productivity of development teams.

Their biggest focus is on motivating employees through innovative forms of communication, in addition to training, which appears to be a critical success factor (CSF) (Leidecker & Bruno, 1984) for data quality management. The characterization of the quality indicators, and the dissemination of its bi-weekly trends in newsletter format using tag clouds, has had excellent results with regard to motivation and involvement of multiple users. Another positive aspect is the recent introduction, (although for IT people only), of a "data quality" key performance indicator[1] (KPI), despite not having a strong presence. Another CSF seems to be the data quality management leadership by a person with a data quality champion profile (Tee *et al.*, 2007).

This single case study shows the existent gap between DQ research efforts that develop and enforce the application of DQ methodologies for quality assessment and improvement, and what is actually done in the industry. Moreover, evidence shows that MyTelecom is working on a cost-effective business case with no use of a data quality methodology. According to literature, this takes place frequently in the first stages of the maturity models.

Despite DQ being considered a source of competitive advantage (Redman, 1996; Tee *et al.*, 2007) prerequisite to operate in highly competitive markets, our findings indicate that MyTelecom which is working in a very high competitive market is unwilling and /or lacks the resources to commit itself to Corporate DQ.

This low DQ maturity level in MyTelecom is, in the view of its DQ Team leader, precisely due to market competitiveness, which leads to projects with very short time to market and leaves no time for defining and implementing architectural options and data policies, such as CDP and MDM.

Another rival explanation for the low DQ maturity level of MyTelecom, may be due to the Industry in which it operates, in which data (and information) may be not a major differentiating factor, since it does not even know the identity of it's pre-paid customers.

4. CONCLUSIONS

From the analysis of our single-case study we can observe that our case study confirms the following (parts of the) propositions:

[1] According to Veerawat, Koronios and Gao (2009) KPIs are compilations of data measures used to assess the performance of business´s.

1. One of the drivers of DQM is cost control;
2. One of the candidate areas for DQM is customer management;
3. The level of Corporations' commitment to DQM is positively associated with the value their Executive Leadership assign to data;
4. DQM activities are mainly sponsored by IS/IT executives;
5. DQ Teams belong to the Application Development Unit of the IS/IT Department.

On the other hand, the case study does not fit the 2nd part of the proposition number 3, "The level of Corporations commitment to DQM is directly related to the degree of market competitiveness in which they operate". This requires further examination through replication in other cases.

Moreover, from the case study findings we can draw the following new prepositions:

6. The level of Corporations' commitment to DQM depends on the industry in which they operate (in which data and information are/are not a major differentiating factor);
7. The use of formal DQ Methodologies is directly related to the Corporations DQ Maturity Level.

5. LIMITATIONS AND FUTURE WORK

Our research questions being *why* a specific company decided to embark on a data quality management initiative, *how* it did that and *which objectives* have been achieved, so far, as well as a comparison of what we actually found in the case with theories from the literature have been achieved. This work being a single pilot exploratory case study, is only aimed to understand a real situation in its context, in order to check the validity of the initial propositions and find new ones to be used in a subsequent multiple-case study, to provide an in-depth understanding of the DQM implementation phenomenon, thus allowing "cross-case analysis and the extension of theory" (Benbasat *et al.*, 1987, p. 373). In fact, to the best of our knowledge (Tee *et al.*, 2007; Missier *et al.*, 2003), very few case studies concerning this broad issue are available.

Although the draft of the case study report has been reviewed by its informants, more sources of evidence could have been used, if more time and resources were available, which could increase the construct validity of the case (Yin, 2003).

This effort will continue through a multiple-case study, across different businesses, with interviews of business and IT people, if we can find available resources and organizations willing to welcome these studies.

References

Batini, C. & Scannapieco, M. (2006). Data Quality: Concepts, Methodologies and Techniques. Springer-Verlag.

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

Benbasat, I., Goldstein, D. K. & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quaterly,* 11 (3), 369-386.

Berson, A. & Dubov, L. (2007). Master Data Management and Customer Data Integration for a Global Enterprise. McGraw-Hill.

Bitterer, A. & Newman, D. (2007). Organizing for Data Quality. Standford, CT: Gartner Research.

Blaha, M. & Rumbaugh, J. (2005). Object-Oriented Modeling and Design with UML (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

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

Dreibelbis, A., Milman, I., Run, P. v., Hechler, E., Oberhofer, M. & Wolfson, D. (2008). Enterprise Master Data Management: An SOA Approach to Managing Core Information. IBM Press.

Drucker, P. F. (1989). The New Realities. New York, NY: Harper & Row.

Drucker, P. F. (1999). Management Challenges for the 21th Century. Burlington, MA: Elsevier.

Drucker, P. F. (2003). A Functioning Society. Piscataway, NJ: Transaction Publishers.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review,* 14 (4), pp. 532-550.

Elmasri, R. & Navathe, S. B. (2006). Fundamentals of Database Systems (5th ed.). Boston, MA: Pearson Addison Wesley.

English, L. (1999). Improving Data Warehouse and Business Information Quality. Wiley & Sons.

Even, A. & Shankaranarayanan, G.(2009). Dual Assessment of Data Quality in Customer Databases. *ACM J. Data Inform. Quality*, 1(3), 15:1-15:29.

Fisher, T. (2009). The Data Asset – How Smart Companies Govern their Data for Business Success. Hoboken, NJ: John Wiley & Sons.

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

Hüner, K. M., Ofner, M. & Otto, B. (2009). Towards a maturity model for corporate data quality management. In *Proceedings of the 2009 ACM symposium on Applied Computing* (pp. 231-238), Honolulu, Hawaii.

Hewlett Packard (2007). Managing data as a corporate asset: three action steps toward successful data governance – White Paper. Retrieved May 11, 2009 from http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA1-2494ENA.pdf

Hoffer, J. A., Prescott, M. B. & Mcfadden, F. R. (2007). Modern Database Management (8th edition). Upper Saddle River, NJ: Pearson Education Inc.

INFORMATICA (2005). Addressing Data Quality at the Enterprise Level – White Paper.Retrieved February23, 2009 from http://www.i2.com/partners/Partner_collateral/INFA_DQ_wp_6631_lo.pdf

Inmon, W. H., O'Neil, B. & Fryman, L. (2008). Business Metadata. Burlington, MA: Morgan Kaufmann Publishers.

International Business Machines (2007). The IBM data governance council maturity model: Building a roadmap for effective data governance. Technical report. IBM Software Group.

International Standards Organization (2005). ISO 9000:2005 (E) – Quality Management Systems: Fundamentals and Vocabulary (3rd edition).

Juran, J. M. (1989). Juran on Leadership for Quality: An Executive Handbook. NY: The Free Press.

Leidecker J. K. & Bruno, A. V. (1984). Identifying and Using Critical Success Factors. *Long Range Planning*, 17(1), pp. 23-32.

Lucas, A., Palma-dos-Reis, A. & Caldeira, M. (2009). The Quality of Monitoring Data in Civil Engineering Works. In *Proc 14th International Conference on Information Quality (IQ 2009)*. Potsdam, Germany.

Missier, P., Lalk, G., Verykios, V., Grillo, F., Lorusso, T. and Angeletti, P. (2003). Improving Data Quality in Practice: A Case Study in the Italian Public Administration". *Distrib. Parallel Databases,* 13 (2).

Newman, D. & Logan, D. (2008). Gartner Introduces the EIM Maturity Model (ID Number G00160425). Gartner Research, 2008.

Otto, B., Wende, K., Schmidt, A. & Osl, P. (2007). Towards a framework for corporate data quality management. In *Proceedings of the 18th Australasian Conference on Information Systems* (pp. 916-926).

Radcliffe, J. (2009). Gartner Magic Quadrant for Master Data Management of Customer Data. Gartner.

Redman, T. C. (1996). Data Quality for the Information Age. Artech House.

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

Strong, D. M., Lee, Y. W. & Wang, R. Y. (1997). Data Quality in Context. *Communications of the ACM*, 40 (5), pp. 103-110.

Sveiby, K. (1997). The new organisational wealth – Managing and measuring Knowledge-Based Assets. San Francisco: Berret – Koehler.

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

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

The Data Management Association (2008). DAMA DMBOK Functional Framework (version 3.02). DAMA International.

The Data Management Association (2008). The DAMA Dictionary of Data Management. Bradley Beach, NJ: Technics Publications LLC.

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

Umar, A., Karabatis, G., Ness, L., Horowitz, B. & Elmagardmid, A. (1999). Enterprise Data Quality: A Pragmatic Approach. *Information Systems Frontiers*, pp. 279-301.

Veerawat, V., Koronios, A. & Gao, J. (2009). A Framework for the Development of the Business Case for the Introduction of Data Quality Program Linked to Corporate KPIs & Governance. In *Proceedings of the Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, pp. 230-235. IEEE.

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

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

Weber, K, Otto, B. & Österle, H. (2009). One size does not fit all – a contingency approach to data governance. *ACM J. Data Inform. Quality,* 1 (1), 4:1-4:27.

Wende, K. (2007). A model for data governance – organizing accountabilities for data quality management.In Proceedings of 18th Australasian Conference on Information Systems.

White, A. & Radcliffe, J. (2008). Mastering Master Data Management. Gartner.

Yin, R. K. (2003). Case Study Research: design and methods (3rd ed.). Thousand Oaks, CA: Sage Publications Inc.

Resumo

A fraca qualidade dos dados custa actualmente muito dinheiro às organizações de todo o mundo, como é sobejamente conhecido. Apesar de se assistir a um crescente interesse na investigação sobre os métodos, conceitos e ferramentas para a avaliação e melhoria da qualidade dos dados (QD), pouco tem sido feito no que respeita à gestão da QD ao nível organizacional. A presente investigação tem por objectivo compreender a natureza e a complexidade da gestão da QD ao nível organizacional, através de várias perspectivas, tais como o tipo de *sponsorship*, o tipo e o nível de colaboração entre o "negócio" e os SI/TI, a posição hierárquica da equipa de gestão da QD, o âmbito das iniciativas de QD, os papéis, os serviços prestados, os métodos, técnicas e ferramentas utilizadas, etc. Este trabalho apresenta, analisa e discute um estudo de caso piloto, de natureza exploratória, desenvolvido numa empresa de telecomunicações fixas e móveis, que opera num País da União Europeia. Este estudo de caso tem por objectivo verificar a validade de algumas proposições iniciais e, eventualmente, encontrar novas proposições, que serão utilizadas num subsequente estudo de múltiplos casos, com vista a aprofundar a compreensão do fenómeno de gestão da QD ao nível organizacional.

Palavras-chave: gestão da qualidade dos dados, estudo de caso