

# Applying Mechanisms of Data Profiling for Assuring Data Quality in the software: a first approach

**Abstract**--For any organization is necessary to satisfy their business objectives besides using data to implement organizational processes, for that reason, it is indispensable to have knowledge of how these data satisfy the preset quality requirements. Thus, these requirements could be expressed by means of some data quality dimensions. In some scenarios, models and methodologies of data quality assessment require of mechanisms to control and monitor the level of quality of data. Thus, proposing a methodology with a qualitative diagnosis of the data quality dimensions and using data profiling techniques to measure some of these dimensions, will have a significant impact on the processes of appropriate use of the data. The main contribution of this paper is a methodology that assesses the data quality, by diagnosing its dimensions through a survey and data profiling techniques.

**Keywords**-- *process improvement; data quality dimensions*

## I. INTRODUCTION

Nowadays, both public and private organizations understand and assess the value of data. Well managed information has an incalculable value for the organizations and enterprises [1]. During years, the data management has acquired a growing importance in companies, because data constitute one of the main assets of the organization, and without them, is almost impossible that corporations can align with their organizational strategy [2].

The data are becoming a key asset to improve the efficiency in today's dynamic and competitive business environment [3]. The economic and social impact of poor data quality (hereafter DQ) has a significant economic cost for organizations [4]. Although significant works of research have been done on the notion of data and information quality, their main focus is on the objective quality attributes of data [5]. The quality of data products and data warehouse directly determines the quality of the business operations and decision-making through data [6]. Because electronic

data are so pervasive, data quality plays a critical role in all business and governmental applications [7], and it is recognized as a relevant performance issue of operating processes [8].

According to standard ISO/IEC 25012 [9], the concept of DQ is defined as "the degree to which the characteristics of the data are suggested conditions and needs when used under specific conditions".

Thus, before any operation or process, it is very important to assess the suitability degree of the use of data involved in the task, according to the context in which they are. In this sense, the technique of data profiling is one of that could help to diagnose the DQ in specific contexts, which is the "data analysis systems to understand its content, structure, quality and dependencies" [10]. Indeed, doing data profiling and monitoring the defects of data, are useful activities for assessing DQ in specific contexts.

Although, nowadays there exists some models, methodologies and tools to carry out the data profiling processes, in our context, it is possible to find some needs such as: assessment of some characteristics (dimensions) of DQ using techniques and tools of data profiling, definition of roles and responsibilities for the DQ control, organization of the process through the use of artifacts and documents and the frequent reporting to the organization about the DQ diagnosis, depending on user types and level securities, in order to involve and engage members and roles that interacts with these data.

With the aim of establishing all the previous points in our proposal methodological, the specific process of *Project Planning* established in the standard ISO 12207:2008 was used as a reference, this process determines "the scope of project and technical activities, identifies process outputs, project activities and deliverables, establishes schedules for project task conduct, including achievement criteria, and required methods and techniques to accomplish project activities".

This work is organized as follow, section II describes existing methodologies for DQ control and assessment, and data profiling models that currently exists. Section III presents the proposed methodology and a description of

their characteristics, principles, scope, processes, activities and people in charge. Finally, Section IV presents the conclusions and intentions for future work.

## II. AREAS RELATED

### A. Methodologies for data quality assessment

Different data quality dimensions (characteristics) have been defined by several authors (from different point of view and use) and even they have been defined in the international standard ISO 25012 [9]. This makes see the importance that the data quality topic has been gaining in the last years, and to perform their definition influences the context in which it is to be used. This proposal was based on the data quality dimensions introduced in ISO/IEC 25012, which are: accuracy, completeness, consistency, credibility, timeliness, accessibility, compliance, confidentiality, efficiency, traceability, portability, understandability, availability and recoverability.

In order to make a comparative study of existing methodologies for DQ assessment, we consider several aspects, including: dimensions used, cost and types of data and information systems involved. The methodologies for DQ assessment and improvement have been classified in four categories [11]: a) *complete methodologies*, which provide support to both the assessment and improvement phases, and address both technical and economic issues; b) *audit methodologies*, which focus on the assessment phase and provide limited support to the improvement phase; c) *operational methodologies*, which focus on the technical issues of both the assessment and improvement phases, but do not address economic issues; d) *economic related methodologies*: which focus on the evaluation of costs.

This research is based on audit methodologies. Some of these methodologies are AIMQ [12], CIHI [13], AMEQ [14] and IQM [15].

The AIMQ methodology (A Methodology for Information Quality Assessment): It is the only methodology of information quality based on benchmarking. It draws heavily on the PSP/IQ model (Table 1), which classifies the DQ dimensions according to the interest and priority of users and administrators. AIMQ contains four classifications for DQ: comprehensive, reliable, useful and usable, into which DQ dimensions fall. It uses questionnaires for the identification and diagnosis of both DQ dimensions and measures of information quality.

**Table 1.** The PSP/IQ model.

	Conforms to specifications	Meets or exceeds the customer expectations
--	----------------------------	--

Product quality	Sound information	Useful information
Service quality	Dependable information	Usable information

CIHI methodology (*Canadian Institute for Health Information*): CIHI focus on the control of DQ of data stored in the Canadian Institute of Health Information, specifically in the monitoring of the size, heterogeneity and quality of the stored data. The data quality evaluation is based on a four-level hierarchical model. At the first level, 86 basic quality criteria are defined. These criteria are aggregated by means of algorithms of composition into 24 quality characteristics at the second hierarchical level, and finally, these are aggregated into five DQ dimensions at the third level. Finally, the five dimensions are aggregated into one overall database evaluation at the fourth level.

AMEQ methodology (*Activity-based Measuring and Evaluating of Product information Quality [14]*): AMEQ provides a rigorous basis for Product Information Quality assessment and improvement in compliance with organizational goals. The methodology is specific for the evaluation of DQ in manufacturing companies, where product information represents the main component of operational databases. In manufacturing companies, the association between product information and production processes is straightforward and relatively standard across companies [7]. AMEQ has five phases. The first one assesses the cultural preparation of the organization. The second one focuses on all information related to the product by process modeling and identification of critical areas. One of the outputs of this phase is a model of measurement techniques. The third phase focuses on the implementation of all activities and techniques for the measurement and evaluation. During the fourth phase the causes of DQ problems that have been detected after diagnosis of the dimensions will be investigated. The last one is responsible for monitoring and improving the quality of product information, through mechanisms of accountability of the processes and data.

IQM methodology (Information Quality Measurement [15]): IQM conceived the provision of a quality framework adapted to the Web data. Among its entries, besides of the quality criteria, it has the tools and techniques used to measure the DQ. The result of evaluation is the most important outputs, which is a valuable guide for selecting and customization of the tools used by web administrators for creating and managing of websites. IQM describes the following main phases: assessment planning, assessment configuration, Measurement and follow-up activities, where the most important processes are: the diagnosis of the data, the requirements analysis and evaluation of the DQ.

After studying the characteristics of these audit methodologies, we consider that they are very useful, depending on its features and goals. However, according with some aspects like: the focus on the business processes of the organization, the definition of roles and responsibilities, the use of artifacts for document the process and the inclusion of data profiling techniques for the DQ evaluation; we conclude that, except AMEQ that utilizes the organizational processes in its process modeling, the rest of the methodologies are not based on business processes. They do not use roles and responsibilities in its phases and activities, they do not include data profiling techniques for the DQ evaluation, and only CIHI has a well-defined documentation process.

### B. Data profiling models, techniques and tools

At present, some data profiling methods and techniques also contribute to the necessary assessment for the DQ control in any kind of information systems, where the fundamental approach is performed on the data collections. The DQ dimensions more widely used to assess DQ are: correctness, completeness and accuracy. One of the models available today is [10], which consists of one or more inputs of data and metadata, the application of research techniques, and as outputs, corrected metadata and information related with data, as shown in Figure 1.

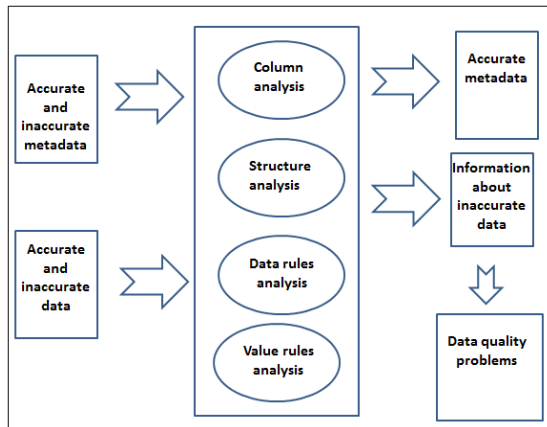


Figure 1. Data profiling model of [10].

Oracle Corporation, a company that has developed a system for profile data, is oriented to the thorough investigation and close monitoring of its quality [16]. With a tool named Oracle Data Profiling, the user has the possibility to discover and infer rules based on data, and monitor their quality over time. As shown in Figure 2, the inputs and outputs are well defined, where data and metadata that were profiled, can be profiled again.

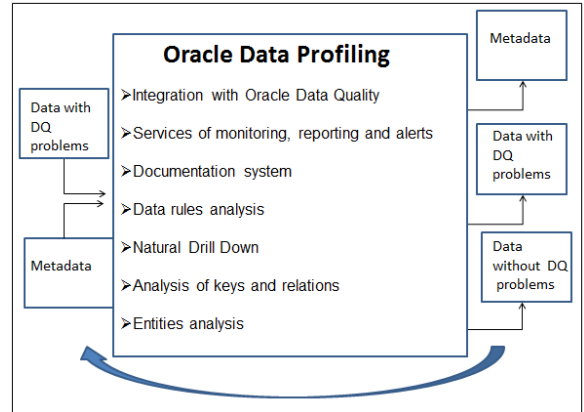


Figure 2. Data profiling techniques and process.

Microsoft offers a tool named Data Quality Services 2008 [17], with techniques and mechanism of data profiling, such as: candidates keys profiling, column profiling, data profiling using patterns, functional dependences profiling, and inclusion values profiling.

The Embarcadero Company is noted for its software design: ER/Studio. Through its CA ERwin Data Profiler tool, the user can combine the analysis and the data modeling in a practical way. In its own model highlights four key activities: analysis column, integration with data models, the discovery of keys and Extended Analysis of attributes [18].

Informatica Corporation [19], with its tool named Informatica PowerCenter, an enterprise platform that offers access, research, data profiling and data integration from any data source, and any format. It is a very important tool for data profiling and diagnoses the DQ. Informatica PowerCenter has five subsystems: Access, Discovery, Cleaning, Integration and Delivery.

### III. METHODOLOGY FOR ANALYSIS, EVALUATION AND CONTROL OF DQ BASED ON DATA PROFILING TECHNIQUES

The contribution of this paper is a methodology to control, analyze and evaluating of DQ through the use of data profiling techniques and diagnosis of the DQ attributes. It consists of three phases, each of which contains processes, activities, artifacts, people in charge and tools.

#### A. Scope and fundamentals of the methodology

The methodology guides to the establishment of the control of data quality, based on the diagnosis of DQ dimensions and processes related with data profiling of relational databases, where activities, techniques and mechanisms are involved, as a guide for its implementation. This methodology to unlike the others

audit methodologies is based on business processes of the organization, besides it defines roles and responsibilities for a better organization of the execution of its phases and activities. It also proposes well-defined artifacts that help for documenting the implementation of the methodology, and it includes an added value: the using of the results of data profiling techniques for the data quality evaluation.

The methodology is based on the following pillars:

- *It is focused on the data quality control of each organizational business process:* its main objective is to implement a data quality monitoring system of the organizational process analyzed, and if a data quality problem raises, enable the possibility of detecting when occurred, the area, database or information system where the problems happened, and who are the people in charge.

- *Implication of roles that manage data:* It is based on committing all the roles involved in data management, which are in charge for monitoring or controlling the data quality.

- *Iterative and incremental:* Once a development cycle of the methodology is completed, it should be executed again so that each iteration will cover each of the organizational processes involved in access, control and management of data in the organization.

## B. Representation and description of the methodology

The methodology is represented in three key phases: *Analysis*, *Evaluation* and *Transition*.

The complete description of each phase of methodology is shown as follow.

**1. ANALYSIS.** At this phase, the current status of a particular organizational process is studied, this implies to take into account the types of existing users, the data types, database administrators, etc., for preparing the infrastructure for the application of data profiling techniques and for the survey of diagnostic of DQ dimensions. In next iterations new organizational processes should be diagnosed. Table 2 shows their characteristics.

**Table 2. Phase of Analysis.**

Input products	Description of the organizational process
Output products	Identified information, selected dimensions for the DQ evaluation
Activities	1.1. Diagnosis. 1.2. Election of Requirements.
Methods, techniques and tools	Expert judgment, brainstorming, artifact for the Diagnosis of organizational process (see artifacts presented in Appendix A).
Roles	Business analyst, Data Quality analyst.

This phase encompasses the next two activities.

**1.1 Diagnosis,** It must be executed to get a first assessment of the current status of the selected organizational processes. For doing so, it is necessary to take into account the databases used by the selected organizational processes, the user types and the existing roles, the types and formats of data handled by the organization and database administrators. As input of this activity, aspects related with the diagnostic process should be provided, and as output, the identified information related with the organization is to be generated.

**1.2 Election of Requirements,** The DQ dimensions selected will be involved in the entire cycle of execution of the methodology for each one of the selected organizational processes. As input of this activity, some aspects of the diagnostic process must be provided, and as output, the list with the selected DQ dimensions should be generated.

**2. EVALUATION.** In this phase, an evaluation of the level of DQ of a relational database should be performed. This implies the use of some techniques like structure profiling, relational profiling, data rules profiling and the implementation of surveys for the diagnosis of DQ dimensions. Table 3 shows their characteristics.

**Table 3. Phase of Evaluation.**

Input products	Result of the diagnosis of organizational process, data source, metadata source.
Output products	Data profiled, metadata profiled, result of the survey for the diagnostic of DQ dimensions
Activities	2.1. Structure profiling. 2.2. Relational profiling. 2.3. Data rule profiling. 2.4. Conductions of a Survey for the diagnostic of DQ dimensions.
Methods, techniques and tools	Profiling of table structures, and its functional dependences, data rules profiling, questionnaire of the survey for the diagnostic of DQ dimensions (see Appendix B), data profiling tools.
Roles	Business analyst, DQ analyst, database administrator, database designer.

In order to achieve the goals of this phase, the team should execute the following activities:

**2.1 Structure profiling.** It consists of thoroughly investigate each one of the columns and rows of tables in the source systems, applying a set of techniques to calculate statistical information and their metadata. The most significant DQ dimensions are completeness, accuracy and precision. As input of this activity, services of data access, profiled and not-profiled data and metadata should be provided, and as output, artifacts, data profiled and metadata profiled should be generated.

*Property profiling*: It refers to applying profiling techniques to determine table properties, such as number and percent of null values, unique, duplicates, blanks, data types, minimum and maximum size of characters, maximum and minimum values and domains, among others. Techniques as *Regular Expressions Profiling* and *Language Profiling* could be applied.

**2.2. Relational profiling.** The main aim of this activity is to determine possible relationships and functional dependencies between tables or business objects, and discovering primary and foreign keys. With this activity it is possible to evaluate the degree of consistency. According to [7], the DQ dimension consistency refers to the violation of semantic rules defined on a data or a particular data set. In this case it will be profiled the violations of integrity constraints, specifically inter-relational constraints. As input, the services of data access, data and metadata profiled and unprofiled are to be provided; and as output, artifacts, primary keys, foreign keys, relationships between entities and the relational matrix should be generated. For this case, several rules, SQL statements and data mining techniques can be applied [20], for example association rules to find dependency percentages of some attributes related to others, and thus find possible foreign keys. Business analyst, DQ analyst, database designer can use the following three techniques and tools to achieve their objectives: *Analysis of primary key*, *Analysis of foreign key* and *Relational Matrix specification*.

**2.3 Data rules profiling.** Activity aimed to the researching, discovering, verification and validation of data rules. It helps to specify the degree of conformity, which determines whether the data has attributes that adhere to standards, conventions or regulations and similar rules relating to DQ in a specific context of use [21]. As input of this activity, services of data access, data and metadata profiled and unprofiled, and as output, artifacts and data rules. Business analyst and DQ analyst could use the following techniques and tools to get the specified results: *Analysis of default data rules* and *Discovery of data rules*.

**2.4 Conduction of a Survey for the diagnostic of DQ evaluation.** The survey is a system for collecting information to describe, compare and explain knowledge, attitudes and behavior [21]. In this process a qualitative diagnostic of DQ is performed from the application of the survey. As input of this activity, the result of organizational analysis is to be provided, and as output, the result of the survey of DQ dimensions should be produced. With the aim of improving the diagnosis and evaluation of DQ, there are five types of users to whom the survey is proposed. Table 4 shows some examples of types of users.

**Table 4.** Types of users and some examples

User types	Examples
Data user	Database administrators, database developers, ETL specialists, etc
Requirement user	Requirement analysts, requirement specialists, etc.
Technology user	Network administrators, server administrators, IT specialists , etc.
Bussiness user	Business analysts, executives, leaders, managers, customers, area and department directors, final user, etc.
Interface user	Web programmers, designers, ads and marketing specialists, etc.

Business analyst, DQ analyst and database administrator can use the following technique and tool to achieve their objectives.

*-Conducting of survey for the diagnostic of DQ dimensions*: This method is based on the DQ characteristics provided by the ISO/IEC 25012. It should be conducted periodically to the members of the organization that interact and manage data involved in the organizational process, with questions related to each of these DQ dimensions, so as to provide a qualitative and quantitative value of the level of quality of the data used within the organization. The Details on the survey can be seen in Appendix B. As proposed tools, the data profiling ones and the questionnaire can be suggested.

**3. TRANSITION.** In this phase the organizational process analyzed is monitored, continuing to the analysis. It should be reported the status of the DQ, to all roles and members involved in the organizational business process. It implements activities related to the process of monitoring and alerting of DQ. Table 5 shows their characteristics.

**Table 5.** Types of users and some examples

Input products	Result of the survey for the diagnostic of DQ dimensions.
Output products	Artifacts, notifications and alerts.
Activities	3.1 Monitoring and control.
Methods, techniques and tools	Notification, and alerts of Data Quality.
Roles	DQ Analyst.

**3.1 Monitoring and control.** The aim of this activity is to notify and alert events related with the detection of a poor data quality in any of the selected business processes of the organization. The people in charge should ensure the beginning for repeating the phase of analysis in a new organizational process. As input, the result of the survey for diagnostic the DQ dimensions should be entered, and as output, the specification of the artifacts, notifications and alerts should be generated. DQ Analyst can use the following technique and tool to achieve their objectives.

*-Execution of the monitoring and alert:* The methodology proposes the implementation of a reporting solution, for the notification of the DQ dimensions assessment to members and roles related with the organizational process, about the current diagnostic of DQ in that process.

#### IV. CONCLUSIONS

This main contribution of this paper is a methodology for the analysis, control and evaluation of DQ, through data profiling techniques and the application of surveys for the diagnostic of DQ dimensions, to various types of users. As research methods during the process of developing the methodology, we used theoretical and empirical methods [22], including the method of survey. We empirically obtained the DQ dimensions used in the

methodology, the types of users to which the questionnaire should be applied, the roles and responsibilities defined, and the output products of the analysis phase. For the success of implementation of the methodology in a real scenario into an enterprise, we think that it is necessary to consider the systemic method as a combined and integrated system of all phases and activities, with an iterative and incremental approach. Besides, the survey plays a key role for the evaluation of the DQ in this methodology.

As future work, we will continue with the analysis of the rest of processes defined in the standard ISO/IEC 12207:2008, identifying their artifacts, roles and processes.

#### V. REFERENCES

1. George, E. and J. Gao. *A quality study exploring users' perception of information management challenges in the cloud*. in *International Conference on Information Quality, ICIQ 2014*. 2014. Xian, China.
2. Ferdinandi, P.L., *Data warehouse advice for managers*. 1999, AMACOM American Management Association.
3. Oliveira, P., F.t. Rodrigues, and P. Henriques. *A formal Definition of Data Quality Problems*. in *Tenth International Conference on Information Quality (ICIQ'05)*. 2005. MIT, Cambridge, MA, USA.
4. Eppler, M. and M. Helfert. *A Classification and Analysis of Data Quality Costs*. in *International Conference on Information Quality*. 2004. MIT, Cambridge, MA, USA.
5. Eshraghian, F. and S.A. Harwood. *Information product: How information consumers' perception of 'fitness for use' can be affected*. in *International Conference on Information Quality, ICIQ 2015*. 2015.
6. Liu, Q., G. Feng, and N. Wang. *Managing data quality for information systems combination from different data sources* in *International Conference on Information Quality, ICIQ 2014*. 2014. Xian, China.
7. Batini, C., et al., *Methodologies for data quality assessment and improvement*. ACM Computing Surveys, 2009. **Vol. 41, No. 3**.
8. Chengalur-Smith, I.N., D.P. Ballou, and H.L. Pazer, *The impact of data quality information on decision making: An exploratory analysis*. IEEE Trans. Knowl. Data Eng., 1999. **11(6)**: p. 853-864.
9. ISO-25012, *ISO/IEC 25012: Software Engineering-Software product Quality Requirements and Evaluation (SQuaRE)-Data Quality Model*. 2008.
10. Olson, J.E., *Data Quality: The Accuracy Dimension*. 2003, San Francisco, CA, USA: Morgan Kaufmann Publishers.
11. Jarke, M., et al., *Fundamentals of Data Warehouses*. Springer Verlag, 1995.
12. Lee, Y.W., et al., *AIMQ: a methodology for information quality assessment*. Information & Management, 2002. **40(2)**: p. 133-146.
13. Long, J. and C. Seko. *A cyclic-hierarchical method for database data-quality evaluation and improvement*. in *Advances in Management Information Systems-Information Quality Monograph (AMISIQ) Monograph*. 2005.
14. Su, Y. and Z. Jin, *A Methodology for Information Quality Assessment in the Designing and Manufacturing Process of Mechanical Products*, in *Information Quality Management: Theory and Applications*, L. Al-Hakim, Editor. 2007, Idea Group Publishing: Hershey, PA, USA. p. 190-220.
15. Eppler, M. and P. Muenzenmayer. *Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology*. in *Proceeding of the Seventh International Conference on Information Quality*. 2002.
16. Corporation, O. *ORACLE DATA QUALITY FOR DATA INTEGRATOR AND ORACLE DATA PROFILING IIG*. 2010 [cited 2018 June 2018]; Available from: <http://www.oracle.com/technetwork/middlew>

- [are/data-integration/oracledq-datasheet-128404.pdf](https://docs.microsoft.com/en-us/sql/data-quality-services/introduction-to-data-quality-services?view=sql-server-2017).
17. Corporation, M. *Data Quality Services*. 2012 [cited 2018 May 2018]; Available from: <https://docs.microsoft.com/en-us/sql/data-quality-services/introduction-to-data-quality-services?view=sql-server-2017>.
  18. Studio, E. *CA Erwin Data Profiler*. 2009 [cited 2018; Available from: <http://www.dbta.com/DBTA-Downloads/Software/CA-ERwin-Data-Profiler-1581.aspx>.
  19. Corporation, I. *Informática Power Center*. 2017 [cited 2018; Available from: <https://www.informatica.com/mx/products/data-integration/powercenter.html>.
  20. Agrawal, R. and R. Srikant, *Fast algorithms for mining association rules in large databases.*, in *Research Report RJ 9839*, I.R. Center, Editor. 1994.
  21. Lawrence, S. and B. Kitchenham, *Principles of Survey Research. Part 1: Turning Lemons into Lemonade*. Software Engineering Notes, 2001. **26**(6): p. 16.
  22. Hernández, R. and S. Coello, *El paradigma cuantitativo de la investigación científica.*, ed. EDUNIV. 2002, La Habana, Cuba.: Editorial Universitaria.

## APPENDIX A

### Diagnosis of organizational process

#### Deliverable

<Organization name>

<Organizational process name>

<Version>

#### Version control

Date	Version	Description	Author
dd/mm/yy>	<x.x>	<Details>	<Name>

#### Introduction

##### Purpose

*[Define the main objective for the evaluation and diagnosis of organizational process.]*

##### Scope

*[It specifies which business processes and DQ dimensions shall apply. In this case the artifact for the diagnosis of organizational process will integrate with the survey for the Diagnosis of the DQ dimensions.]*

##### References

*[List of referenced documents]*

Code	Title
[1]	Document 1
[2]	Document 2

#### Glossary

*[In the glossary specifies a group of basic terms that are managed for the diagnosis of organizational process.]*

#### Description of the diagnosis application

*[It describes the implementation strategy for the diagnosis of the organizational process.]*

#### Summary of the diagnosis in the business process:

*[Summary of the results of the diagnosis of organizational process.]*

#### Analysis of significant results:

*[Analysis of the most relevant results obtained in the diagnosis and a summary of the main factors to consider.]*

#### Conclusions

*[Conclusions of the diagnosis of the organizational process.]*



## APPENDIX B

<b>Survey: Diagnostic of DQ dimensions</b>						
Taking into account the following initial requirements:						
The business area to diagnose: _____						
The business concept to diagnose: _____						
The source, data source or agent in charge (person or system) to enter data: _____						
The data or the data set that must be evaluated: _____						
In the range of time: From: _____ (day/month/year) To: _____ (day/month/year)						
		A	B	C	D	E
Accuracy	Does the collection have the value and the actual characteristics expected?					
Completeness	Is the collection completed and has all the expected values?					
Consistency	Is the collection free of inconsistencies, contradictions in relation to other data?					
Credibility	Does the collection have adequate credibility and reliability?					
Currentness	Do you think the collection is updated with respect to the specified time range or with respect to the current time?					
Accessibility	Can be the collection properly managed through its access?					
Compliance	Does the collection comply with business rules or restrictions?					
Confidentiality	Does the collection have the appropriate confidentiality and security?					
Efficiency	Does the collection have the expected levels of efficiency and performance?					
Precision	Does the collection have the adequate accuracy and precision?					
Traceability	Is the access to the collection being audited by traces or tracks?					
Understandability	Is the collection understandable and interpretable by users?					
Availability	Can the collection be properly retrieved by authorized users or applications?					
Portability	Will maintain the collection its quality if is moved from one system to another?					
Recoverability	Will maintain the collection its quality despite occurrences of failures?					