

Framework for Data Quality Evaluation Based on ISO/IEC 25012 and ISO/IEC 25024

Julieta Calabrese¹ , Silvia Esponda , Patricia Pesado 

Instituto de Investigación en Informática LIDI (III-LIDI)*
Facultad de Informática – Universidad Nacional de La Plata
50 y 120, La Plata, Buenos Aires, Argentina

*Partner Center of the Scientific Research Agency of the Province of Buenos Aires

¹ Fellow, UNLP

{jcalabrese, sesponda, ppesado}@lidi.info.unlp.edu.ar

Abstract. Nowadays, organizations process large volumes of data. Being able to foster and maintain data quality is one of the biggest challenges that these organizations face. To do this, there are standards, such as ISO/IEC 25012 and ISO/IEC 25024, which are intended to measure data quality based on a set of inherent and system-dependent characteristics, along with a set of associated metrics. Using standards to carry out measurements can be complex and even expensive for those with little experience in the area. In this context, we propose in this work a prototype for a tool based on ISO/IEC 25012 and ISO/IEC 25024 that, by analyzing different patterns of common errors in data, allows an organization to understand data current status.

Keywords: data quality - ISO/IEC 25012 - ISO/IEC 25024

1 Introduction

From the beginning of Software Engineering, there have always been issues related to achieving optimum quality levels in different aspects of software [1]. It is widely known that quality management is essential within any organization.

Technological advancement interferes in all sectors, from agriculture to manufacture, tourism, health care, and university; in the process, data become the most powerful organizational asset and a key aspect for decision-making. New technologies allow obtaining and storing large amounts of data (through mobile devices, sensors, and so forth) and analyzing them using different algorithms, generating an unlimited amount of processing. However, for the information obtained to be considered the most important asset for the organization, data must be of adequate quality from their

source and appropriate for the environment where they are used. The result of any decision made by the organization will be based on its information.

Currently, data have become essential in digitally transformed organizations, and data quality is essential to achieve service excellence for all stakeholders. Lack of resources to assess data quality is one of the main problems organizations currently face, as this significantly affects organizational and business effectiveness and efficiency. This context leads to focus attention on the standards defined by ISO in relation to data quality.

This work is part of a doctoral thesis for the development of a framework that simplifies data quality control for organizations.

2 ISO/IEC 25012 and ISO/IEC 25024

To organize and unify all standards related to software product quality, ISO/IEC published in 2005 the document ISO/IEC 25000:2005 - SQuaRE (System and Software Quality Requirements and Evaluation) [2], also known as the ISO 25000 family. Within this set, ISO/IEC 25012 - "Data Quality Model" [3] and ISO/IEC 25024 - "Measurement of data quality" [4] stand out for study.

ISO/IEC 25012 - "Data Quality Model" specifies a general quality model for data that are defined in a structured format within a computer system. This standard classifies quality attributes into fifteen characteristics analyzed from two points of view: inherent and system-dependent. These characteristics will be assigned different importance and priority by each evaluator based on their own specific needs. Accuracy, Completeness, Consistency, Credibility and Currentness are **inherent** characteristics; while Availability, Portability and Recoverability are **system-dependent** ones. Accessibility, Compliance, Confidentiality, Efficiency, Precision and Traceability belong to **both groups**.

On the other hand, ISO/IEC 25024 - "Measurement of data quality" provides measurements, including measurement methods and related quality measurement elements for the quality characteristics of the data quality model described above.

3 Data Quality

Due to the lack of departments specialized in data quality analysis and quality certifications, organizations face greater challenges in relation to the data they manage. As an example, a government organization published its data corresponding to Wi-Fi networks access points [5], and situations such as the following are found:

id	identificador	ubicacion	latitud	longitud
965	TUC034-02	Comuna	-26.420030	-64.775878
681	CHA002-01		0.000000	0.000000

Table 1. Wi-Fi networks access point data

The location of one of the access points is unknown and, therefore, the corresponding data about coordinates is not available. For the specific context, is it possible for this type of problem to occur? Should the missing data be mandatory? Are access points allowed that do not have location as data? Is naming fields as ID and IDENTIFIER confusing as to what they represent?

In this context, it is of interest for this work to identify a set of recurring problems with data in various organizations with different topics, generating patterns that represent common organizational failures that somehow can facilitate data analysis.

Commonly found problems are usually:

- Errors in completeness: Fields that are required but are left blank.
- Syntax errors: Some fields in upper case and others in lower case, mix of both cases in the same field, language-specific issues, such as words with tilde and others without tilde (where it should be present), problems with the "ñ", and so forth.
- Semantic errors: Errors with expected values. For example, in a field called "Country", the expected value would be the name of a country, not "María".
- Consistency errors: Contradiction errors.
- Update errors: Data are not updated with respect to the current environment.
- Traceability errors: There is no data log recording additions, modifications and/or deletions, specifying the corresponding event.
- Precision errors: Data are not accurate where they should be. This is more common in contexts where a lack of precision in a value (decimals for example) can result in a radical change.
- Understandability errors: Data are represented by symbols that cannot be understood by any type of user.
- Accessibility errors: Data cannot be accessed by users who need support due to having some type of disability.

In addition, it has been demonstrated that there are department-specific errors in data, which must be taken into account for correct evaluation.

It would be extremely useful for organizations to have a tool that can use quality standards to respond to the different patterns identified so as to understand the current status of their information through a simple process that yields an in-depth analysis of the quality of their data.

4 Proposed Development

A framework will be developed to respond to the different common failure patterns identified in current organizations from various areas. This framework will be based on the ISO standards associated with data quality, ISO/IEC 25012 - Data Quality Model and ISO/IEC 25024 - Data Quality Measurement, to define which characteristics will be evaluated, as well and the corresponding metrics. The contribution will be useful for organizations that need to understand the current state of the information they process, simplifying the application of standards.

The framework will be applied in various organizations. In particular, it will be applied in organizations that process large volumes of data and which require acceptable levels of quality.

5 Bibliography

- [1] “Guía para evaluar calidad de datos basada en ISO/IEC 25012”. Calabrese, Julieta; Esponda, Silvia; Pasini, Ariel, Boracchia, Marcos; Pesado, Patricia. Congreso Argentino de Ciencias de la Computación - CACIC 2019.
- [2] “ISO/IEC 25000:2014 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Guide to SQuaRE”.
- [3] “ISO/IEC 25012:2008. Software engineering -- Software Product Quality Requirements and Evaluation (SQuaRE) -- Data quality model”.
- [4] “ISO/IEC 25024:2015. Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Measurement of data quality”.
- [5] Data obtained from <http://datos.gob.ar>