

A Data Quality Model for Master Data Repositories

Fernando GUALO^{1,2}, Ismael CABALLERO^{2,*}, Moisés RODRÍGUEZ³,
Mario PIATTINI²

¹ *DQTeam, Camino de Moledores s/n, 13071, Ciudad Real Spain*

² *University of Castilla-La Mancha, Paseo de la Univerisdad 4, 13071, Ciudad Real Spain*

³ *AQCLab, Camino de Moledores s/n, 13071, Ciudad Real Spain*

*e-mail: fgualo@dqteam.es, Ismael.Caballero@uclm.es, mrodriguez@aqclab.es,
Mario.Piattini@uclm.es*

Received: May 2023; accepted: October 2023

Abstract. Master data has been revealed as one of the most potent instruments to guarantee adequate levels of data quality. The main contribution of this paper is a data quality model to guide repeatable and homogeneous evaluations of the level of data quality of master data repositories. This data quality model follows several international open standards: ISO/IEC 25012, ISO/IEC 25024, and ISO 8000-1000, enabling compliance certification. A case study of applying the data quality model to an organizational master data repository has been carried out to demonstrate the applicability of the data quality model.

Key words: master data management, data quality certification, master data quality, ISO 8000-100 series, ISO/IEC 25012, ISO/IEC 25024.

1. Introduction

With the growing need to share data to empower companies, organizations, and societies to be more competitive and sustainable (i.e. European Data Strategy¹), data must be propagated within and outside the organizations. To avoid the possible propagation of errors and to assure the achievement of the expected benefits, it is more necessary than ever to monitor the levels of quality of the data repositories by strategically enabling data quality evaluation cycles regularly; and, if feasible, granting data quality certifications to maximize the trustability and usability of organizational data (Gualo *et al.*, 2021).

Implementation of master data initiatives creates a unified and internally and externally shareable vision of the critical business data elements and associated standard business rules (Fan *et al.*, 2013; Silvola *et al.*, 2011). The resulting master data repository is an integrated, consistent, trusted, and commonly shared representation of the “*single version of the truth*” populated with the golden records. One advantage is that this repository can be used as an authoritative source of data to reliably and consistently feed business

*Corresponding author.

¹European data strategy: <https://ec.europa.eu/digital-single-market/en/european-strategy-data>

processes in different scenarios by different actors in other contexts (Allen and Cervo, 2015; Haneem *et al.*, 2017, 2019). Investments in Master Data Management (MDM) allow the minimization of redundant efforts in repeatedly fixing the same data quality concerns through different departments (Allen and Cervo, 2015; Gualo *et al.*, 2020). Consequently, the better the levels of data quality of the master data repositories, the better the intra- and inter-organizational performance.

However, dealing with poor data quality in the master data repositories is still one of the biggest challenges in master data management (Benkherourou and Bourouis, 2022; Ibrahim *et al.*, 2021; Silvola *et al.*, 2011). Several works stated that master data quality issues go beyond consistency and deduplication (Cleven and Wortmann, 2010; Fan *et al.*, 2013). Furthermore, it is interesting to define master data quality measurements to explore and monitor the execution of specific managerial concerns (i.e. redefinition of master data models, Gartner, 2021) or technological projects (i.e. entity resolution processes, Talburt, 2011).

To systematically and rigorously conduct the corresponding data quality evaluation and certification projects, it is possible to use the international standards ISO/IEC 25012 (ISO/IEC, 2008), ISO/IEC 25024 (ISO/IEC, 2015), and ISO/IEC 25040 (ISO/IEC, 2011). In this sense, we pose that organizations can be benefited from using the framework we developed based on these international standards tailoring a data quality model for the master data repositories conveniently. This repository type has specific features that differentiate them from transactional data repositories making it unique and justifying a particular investigation on master data quality (see Section 2.1). In this paper, considering these differences and based on learned lessons in our previous experience, we propose a **quality model tailored for master data repositories based on international standards**. ISO/IEC 25012's data quality characteristics are revisited for this aim. The related properties and measurement methods defined in ISO/IEC 25024 are also reviewed, considering the unique features that make master data quality different from regular data quality. To this latter aim, specific business rules for master data quality are required (Fan *et al.*, 2013; Valencia-Parra *et al.*, 2021). To our knowledge, only the ISO 8000-100 series (more specifically, ISO 8000-110, ISO, 2009) could be a source of potential business rules for evaluation and certification. As the business rules are not explicitly stated in these standards, we inferred them from various parts of the ISO 8000-100 series. Finally, to grant a certificate, some characteristics of master data quality susceptible to being certifiable were identified and selected by stating and applying some criteria.

The remainder of the paper is structured as follows: Section 2 briefly reviews concepts about master data management, data quality evaluation, and ISO 8000-100 series. Section 3 describes the data quality model for master data quality evaluation and certification. Section 4 introduces a case study, also showing some learned lessons. Section 5 identifies some threats to the validity of our proposal, and finally, Section 6 presents some conclusions.

2. Literature Review

2.1. Master Data and Master Data Management Foundations

According to ISO/IEC 25024 (ISO/IEC, 2015), master data is “*the data held by an organization that describes the key entities that are independent and fundamental for an enterprise that needs to reference to perform its transactions and that are essential for the core business of an enterprise*”. Data and metadata should be shared either internally within the organization (i.e. different applications supporting various business processes) and externally with some other third partners (Allen and Cervo, 2015; Dreibelbis *et al.*, 2008; Otto, 2015; Rivas *et al.*, 2017; Silvola *et al.*, 2011). The sharing typically includes the critical entities with the essential properties or attributes and associated metadata, attributes, definitions, roles, connections, and taxonomies (Silvola *et al.*, 2011). Key entities can be **partners** (e.g. organizations, customers), **places** (e.g. locations, offices), or **things** (e.g. products, services) (Silvola *et al.*, 2011). Master data have the following features: independent existence, in contrast to other types of data (i.e. transactional), low change frequency (due to its stability across time), and constant data volume (the domain of the organization limits the amount and type of knowledge on the business) (Cleven and Wortmann, 2010).

Master data differs from any other types of data in these four aspects (Hüner *et al.*, 2011): (1) Master data refers to key business entities, data references have a lower granularity level for master data attributes, and it is not limited to a given domain (Cleven and Wortmann, 2010); (2) Master data always describes essential characteristics (i.e. product name) that can be part of the transactional data (i.e. orders) or the inventory data (i.e. product data); (3) Master data is kept unaltered longer than other types of data, and, typically, for some business, the provided initial description is the basis for the progressive evolution of the essential characteristics observed by the organization over time; and (4) Master data records (systems of records) are usually more constant in volume than transactional data.

Transactional data typically use master data as a reference. The implementation of master data management systems typically includes four main components (Haneem *et al.*, 2019): data integration module, master data repository – whose quality is the focus of this paper-metadata repository (i.e. data dictionary), and data quality module. These components use a wide range of the following artifacts (Allen and Cervo, 2015): data models, data dictionaries with reference data (the commonest metadata repository), functional architecture, source to target mapping, data life cycle, and CRUD analysis indicating the assignation of permissions to user groups to operate with master data.

Data dictionaries contain the semantic definition of terms for master data that would provide the business and technical description of the master data through metadata and valid values (reference data) for master data attributes. Reference data can be defined as a list of acceptable values (i.e. a list of currency or nation codes as those gathered in ISO 4217 or ISO 3166) or using procedures to build these good values (i.e. regular expressions to validate an email address). To ensure that master data attributes (especially identifiers)

take the correct values from the set of valid values, syntactic and semantic codification should be enabled, as explained in ISO 800-110 (ISO, 2009). This codification is required to identify a specific master entity within a dataset unambiguously (Berson and Dubov, 2011), when it is necessary to anonymize the information for personal protection purposes, or even when it is required to cipher information for security purposes (Piedrabuena et al., 2015). To better support codification, the creation of a data dictionary is convenient. These semantic definitions are later used to build master data repositories that contain what has also been called the golden record. A “golden record” is a superset of all master data attributes whose values have been integrated from all available data sources for the same real-world entity (Loshin, 2010). These aspects become essential for the quality of the proper master data repository and make the management of master data quality different than other types of data (e.g. transactional data).

However, master data is much more than technology (Haneem et al., 2017). Dahlberg et al. (2011) pointed out that poor master data management is the cause for which business initiatives can suffer from loss of performance or even lack of data quality. Master data management requires a combination of processes, data governance, and technology implementation to enable organizations to achieve the maximum benefits from the data (Allen and Cervo, 2015; Berson and Dubov, 2011; Haneem et al., 2017).

2.2. Management of Quality of Master Data Repositories

Organizations must face many data quality concerns, like inconsistencies in data definitions, data formats, and values or lack of understanding of the semantics of the data definitions (Cleven and Wortmann, 2010). Master data management is instrumental in addressing these data quality problems and concerns (Dahlberg et al., 2011; Haneem et al., 2017; Silvola et al., 2011; Vilminko-Heikkinen and Pekkola, 2017).

As master data are still data (Allen and Cervo, 2015), it could be said that the body of knowledge on data quality management for regular (i.e. transactional) data is a good starting point for studying the quality of master data, but taking into consideration the unique features of master data shown in Section 2.1. However, even when several works have dealt explicitly with how master data management has helped organizations to solve data quality issues (Silvola et al., 2011), only a few works address the evaluation of the quality of the master data repositories (Ibrahim et al., 2021; Vilminko-Heikkinen and Pekkola, 2017).

Master Data quality evaluation is also closely associated with data quality dimension or characteristics (Gualo et al., 2021; Wand and Wang, 1996). These data quality characteristics can be understood as criteria used to evaluate the quality and can be used as a basis for improving quality (Valencia-Parra et al., 2021). A data quality model is the set of data quality characteristics that best represent user requirements in a specific context. Consequently, different data quality evaluations require particular data quality models. This fact is also true for the quality management of master data. More specifically, several authors have proposed other data quality models for different contexts of using master data. Table 1 gathers and compares the various data quality characteristics that the most

Table 1
Data quality dimensions/characteristics for master data in the literature.

	(Allen and Cervo, 2015)	(Dahlberg <i>et al.</i> , 2011)	(Haneem <i>et al.</i> , 2017)	(Hüner <i>et al.</i> , 2011)	(Loshin, 2010)	(Fan <i>et al.</i> , 2013)	(Otto <i>et al.</i> , 2010)
Accuracy	✓		✓	✓	✓		
Availability		✓					
Change frequency				✓			
Completeness	✓			✓	✓	✓	
Conformity	✓						
Consistency	✓		✓		✓	✓	
Correctness							✓
Currency					✓	✓	
Deduplication			✓			✓	
Integrity	✓	✓					
Plausibility							✓
Timeliness	✓	✓		✓	✓		
Uniqueness	✓				✓		
Validity	✓						

relevant master data quality works have included. These data quality characteristics enable the measurement of specific aspects.

In addition to these works, given its benefits, we pose that the most convenient way to deal with the selection of the data quality characteristics is using ISO/IEC 25012 (ISO/IEC, 2008) (see Table 2 – Type “I” stands for “*Inherent*”, and “SD” stands for “*System Dependent*”). For evaluating data quality characteristics, the underlying data quality properties and corresponding measurements must be defined to measure the quality of the master data records. For a generic description of data quality properties and measurements, a view on ISO/IEC 25024 (ISO/IEC, 2015) is encouraged in complementing ISO/IEC 25012. Additionally, ISO/IEC 25040 (ISO/IEC, 2011) provides a generic software quality evaluation framework that can be also used for evaluating data quality.

Based on these standards, we developed a complete framework – initially proposed by Merino (2017) and later revised and improved by Gualo *et al.* (2021) – that can be used to evaluate and certify the quality of various data types. This framework can be used with tailored data quality models for several data types and domains. The tailoring of suitable data quality models can be defined by conducting the appropriate research like the present work for Master Data. The tailoring involves the interpretation of the characteristics included in the data quality models and the consequent definition of the data quality measurements.

Part of the definition of data quality measurements involves the identification of some business rules describing the fitness for using a data record (that is, defining a list of valid values for attributes). Previously in Section 2.1, the importance of data dictionaries as typical containers of the set of valid values for the golden records in the master data repository was raised. For instance, data dictionaries can help the organization in the

Table 2
Data Quality characteristics in ISO/IEC 25012.

Characteristic	Type	Definition
Accessibility	I/SD	The degree to which data can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.
Accuracy	I	The degree to which the data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.
Availability	SD	The degree to which data has attributes that enable it to be retrieved by authorized users and/or applications in a specific context of use.
Completeness	I	The degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use.
Compliance	I/SD	The degree to which data has attributes that adhere to standards, conventions, or regulations in force and similar rules relating to data quality in a specific context of use.
Confidentiality	I/SD	The degree to which data has attributes that ensure that it is only accessible and interpretable by authorized users in a specific context of use.
Consistency	I	The degree to which data has attributes free from contradiction and coherent with other data in a specific context of use.
Credibility	I	The degree to which data has attributes regarded as true and believable by users in a specific context of use.
Currentness	I	The degree to which data has attributes of the right age in a specific context of use.
Efficiency	I/SD	The degree to which data has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.
Portability	SD	The degree to which data has attributes that enable it to be installed, replaced, or moved from one system to another, preserving the existing quality in a specific context of use.
Precision	I/SD	The degree to which data has attributes that are exact or that provide discrimination in a specific context of use.
Recoverability	SD	The degree to which data has attributes that enable it to maintain and preserve a specified level of operations and quality, even in the event of failure, in a specific context of use.
Traceability	I/SD	The degree to which data has attributes that provide an audit trail of access to the data and any changes made to the data in a specific context of use.
Understandability	I/SD	The degree to which data has attributes that enable it to be read and interpreted by users and are expressed in appropriate languages, symbols, and units in a specific context of use.

banking or insurance domain to adopt the Global Legal Entity for Identifiers (GLEI²) to build and use specific identifiers for their partners, or ISO 21586 for their related products, or when a manufacturing company has decided to adopt Global Trade Item Number (GTIN) to codify the identifiers for its products (Hüner *et al.*, 2011). Consequently, it can be said that, if correctly implemented, data dictionaries are one of the most valuable resources to guarantee adequate levels of quality for data. To some extent, it is possible to state that the quality of the master data repositories largely depends on the quality of the data dictionary. Open international standards like the ISO 22745 series can be consulted

²Global Legal Entity for Identifiers website: <https://www.gleif.org/en>

Table 3
Overview of the ISO 8000-100 series.

	ISO 8000-100 – Overview (ISO, 2016a)	ISO/TS 8000-150 – Quality management framework (ISO, 2011)	
ISO Master	Exchange of characteristic data	ISO 8000-110 – Syntax, semantic encoding, and conformance to data specification (ISO, 2009)	
Data Standard		ISO 8000-120 – Provenance (ISO, 2016b)	ISO 8000-130 – Accuracy (ISO, 2016c)
			ISO 8000-140 – Completeness (ISO, 2016d)
	ISO 8000-115 – Exchange of quality identifiers: Syntactic, semantic, and resolution requirements (ISO, 2018)		
	ISO 8000-116 – Exchange of quality identifiers: Application of ISO 8000-115 to authoritative legal entity identifiers (ISO, 2019)		

and followed to improve the definition of data dictionaries. However, studying the quality of data dictionaries is outside the scope of this investigation.

2.3. General Overview of ISO 8000-100 Series

Evaluating data quality properties requires identifying, validating, and grouping the business rules (Caballero *et al.*, 2022). For the interest of this investigation, we considered parts 100 to 150 of the ISO 8000-100 series as the primary source of business rules. These standards describe some aspects of the exchange of master data between applications (or organizations) using the exchange of master data messages. Each part of the family is summarized in Table 3.

3. Tailoring a Quality Model for Master Data Quality Evaluation

This investigation is motivated by some organizations' need to evaluate the quality of the master data repositories. Based on our experience in previous data quality evaluations and certification projects (Gualo *et al.*, 2020), and on the inspection of the scarce related literature we found, in this section, we present the master data quality model we tailored to be used with the framework introduced in Section 2.2. The adaptation involves revising the ISO/IEC 25012 data quality model and related ISO/IEC 25024 data quality properties from the point of view of master data and considering its unique features.

This revision generates a master data quality model agnostic to (1) the domain of application of the master data (i.e. Manufacturing, Healthcare) (Allen and Cervo, 2015); (2) the technology used to implement the master data repository, although typically, relational technology is preferred (Berson and Dubov, 2011; Otto, 2015), and to the technology used to build and deploy services to access to the master data repository (Dreibelbis *et al.*, 2008); (3) the number of data sources to feed the master data repository, as well as the number of destinations that consume data from the master data repository (i.e. business processes) (Gualo *et al.*, 2020; Ofner *et al.*, 2012); (4) the specific implementation of the integration processes feeding the master data repository (Dreibelbis *et al.*, 2008); and (5) the existing data quality processes already implemented to control and monitor the level of quality of master data repositories.

3.1. Revisiting Data Quality Characteristics for Master Data

Since master data repositories are a particular type of data repository with specific business rules and a unique environment, we grounded our investigation on the works on the quality of regular data based on ISO/IEC 25012 and ISO/IEC 25024. In our investigation, we comprehensively analyse master data quality and consider their specific unique features that differentiate master data from other types of data (see Section 2.2) to provide interpretations of the effect of master data quality for business and IT. In the following paragraphs, we introduce some results of this analysis. The analysis is made by first recalling the definition of the data quality characteristics according to ISO/IEC 25012 presented in Table 2.

Accessibility. Typically, this concern of accessibility is mainly referred to the user interfaces and users' disabilities. Master data is generally accessed using web services. Furthermore, this type of service would not essentially differ from others for regular data. So, this characteristic does not necessarily need to be tailored for master data.

Accuracy. Master data records must contain valid values (i.e. semantically or syntactically encoded reference data) that represent the entities in real life (Allen and Cervo, 2015). In Master Data Systems, as mentioned in Section 2.1, these reference values are typically stored and accessible in data dictionaries (i.e. eOTD by ECCMA³ in the “*product data*” domain or ICD9/ICD10 in the “*healthcare*” domain), some of which may be publicly available or through a pre-set fee (ISO, 2009). Thus, inadequate levels of accuracy may be a symptom that the coding processes need to be fixed, regardless of whether the process is manual or has been automated to the maximum extent possible. The more relevant concerns can be measured on the master data values (through “*Semantic Accuracy*” or “*Syntactic Accuracy*” properties) or the data model (using the “*master data model accuracy*” property). Low levels of this data quality characteristic are symptomatic of an inefficient standardization process or preliminary design and implementation of the ETL process. This characteristic is strongly related to others, such as Understandability and Credibility; in fact, they maintain a direct relationship: the higher the Understandability, the higher the Accuracy and the higher the Credibility.

Availability refers to the fact that master data, as a reference for different business processes, should be available and accessible at all times for stakeholders to use whenever they need it (Li *et al.*, 2013). It should also be mentioned that this data quality characteristic is related to the ability of systems to provide master data records in an acceptable response time (Allen and Cervo, 2015; Loshin, 2010). This data quality characteristic can be measured through properties like the “*Data Availability Ratio*” or “*Architectural Data Element Availability*”, which can help to detect when the master data records or some of the elements of the architecture are not available (i.e. discontinuities of the web services providing master data). In the context of master data, low levels for measuring this data quality characteristic can be symptomatic of inadequate implementation of the ETL process.

³eOTD | ECCMA OPEN TECHNICAL DICTIONARY website: <https://www.eotd.org/>

Completeness. Completeness can be measured from both points of view of the master data model and the master data values. To meet the requirement, it would be desirable that the master data design performed on a conceptual master data model be able to solve all the data demands of the organization. Therefore, this master data model should be agreed upon and complete as possible, considering as many contexts of using master data within the organization as possible (Allen and Cervo, 2015). As stated in Section 2.1, master data records are built from the integration and consolidation of master data attributes from the different data sources integrated as part of the design and implementation of the master data model (Allen and Cervo, 2015). During master data integration processes (performed through ETL processes), it should be possible to generate values for all attributes defined for master data entities (i.e. entity resolution, Talburt, 2011). Completeness measurements can be used to diagnose some relevant concerns, such as (1) missing values for an entire attribute of a data item in the master data repository, (2) missing values for some values of master data records, or (3) missing certain records in master data of the data file. Considering that the conceptual model is adequate, having null values in those data elements may be a symptom of (1) the ETLs processes are not adequately extracting values from all data sources for all master attributes specified in the conceptual model; and (2) that the data sources are not sufficiently complete. In the first case, an exhaustive testing process of the ETLs processes is to be used; additionally, it may happen that, in the process for master data record consolidation, the entity resolution algorithms do not generate any values for specific attributes. On the other hand, in the second case, the idea is to resort to a complete analysis of the completeness of each source involved in generating the master data values.

Compliance. Compliance can be measured from an inherent and a system-dependent point of view. For example, the master data values for the master data of the disease registry in the healthcare domain must comply with ICD10 codes (Alonso *et al.*, 2020), or in the case of product data management, the GTIN to describe how to construct such codes. The level of compliance with a normative value or format can be measured by the “*Regulatory compliance of value and/or format*” property. It is common to use specific technologies (i.e. JSON, XML) to support master data exchange. “*Regulatory compliance due to technology*” property can determine the level of compliance of a given technology. Low levels of compliance can be symptomatic of an inefficient observation of required standards or even regulations.

Confidentiality. It is essential to distinguish between availability (being able to access the master data record) and confidentiality (being able to read, interpret and use the master data record if the user permissions are adequate to the sensitiveness of the data). This characteristic can be measured using “*Encryption usage*” and “*non-vulnerability*” properties. The foundations of confidentiality measurements for master data do not differ much from those established for regular data. It is a matter of the user access setup of the master data repository.

Consistency can be measured between attributes of the same target entity or comparable target master data entities. Consistency can be measured for a single system or more than one system in the same environment. Considering the master data repository as an

authoritative source containing a single source of truth, inconsistencies should be avoided using reference data to prevent the master data repository from becoming cluttered with inconsistent values. This characteristic's quality level can be determined by using several properties: "Master data format consistency", "Master data semantic consistency", "Referential Integrity", or "Risk of master data inconsistency". Standardization and propagation of data reference values across data sources are good practices. In this sense, the apparition of values in the master data records for attributes that are not calculated and do not exist in the source tables can be a symptom of two possible problems. These problems are (1) referential integrity mechanisms have not been conveniently enabled in the design of the master tables, and (2) the ETLs processes are not well designed and do not consider all sources that can minimally provide values to create a minimum master data record. It is interesting to note that reference data can collaterally increase the accuracy and consistency levels. On the other hand, it is possible to include additional data quality characteristics under the ISO 25012's consistency umbrella, such as completeness, uniqueness, conformity, and validity, as Allen and Cervo (2015) did.

Credibility. Organizations that consider their master data repositories as the sole source of truth establish internally that the master data must be inherently credible. This statement can be made implicitly through agreements or by creating specific data policies that must be implemented and enforced. In this way, top management is responsible for any shortcomings that any stakeholder may cause when using the data in its specific context. To some extent, it could be said that the credibility of the master data repository, if any policy is in place, is as high as possible. The "Master data repository credibility" property can measure the level of credibility. On the other hand, most master data records contain reference data. They are generated through the integration and entity resolution of data values from various sources using specific algorithms (Talbert, 2011). These algorithms are supposed to observe rules of thumb, as well as possible exceptions, to create a master data record that is credible and believable by users in a specific context (Otto et al., 2010). This characteristic can be measured by the "Credibility of the master data value" property. Low levels of Credibility, combined with low levels of Consistency, are typically symptomatic that the credibility of the data sources is inadequate, and some inspection works must be done to isolate and fix the corresponding ones. The most crucial challenge in measuring this characteristic of data quality in master data regarding regular data is the need for solid integration and cohesion of the master data records.

Currentness. By their nature, master data repositories, once master data records are consolidated, tend to have a minimal rate of change, so they are typically expected to change little over time (Dahlberg et al., 2011; Hüner et al., 2011). However, they are not exempt from changes, being some of the reasons for such changes are some planned update operations, the need to execute some standardization process of the master data, or some update of the reference data (Loshin, 2010) as in the case of the updates from ICD9 to ICD10 in clinical data coding (Santos et al., 2021). That implies that to maintain adequate levels of the Currentness of master data records, these changes to the values of the attributes of master data records should be made and propagated as soon as possible (Fan et al., 2013). Currentness can be measured using two properties: "Timeliness of update"

and “*Update frequency*”. Low levels of Currentness are symptomatic of inefficient update processes. For this reason, Currentness is typically included as a relevant characteristic in master data quality management (Allen and Cervo, 2015).

Efficiency. One of the most common concerns in the literature on master data quality is the duplicity of master data records (Berson and Dubov, 2011; Fan *et al.*, 2013; Haneem *et al.*, 2017; Loshin, 2010) due to the inability or inefficiencies in entity resolution (Talburt, 2011), or to the election of a format type for master attribute values that do not optimize the space occupied or the performance of operations (Allen and Cervo, 2015). Optimizing Efficiency requires avoiding duplicate master data entries to solve redundancy issues, avoiding wasted storage space, and ensuring each master attribute has a usable format. The Efficiency can be measured through three properties: “*Efficient master data item format*” to ensure that the data format of master data records is correct, “*Risk of wasted space*” to ensure that the size of master data records is correct, and no unnecessary memory space is wasted, and finally, the “*Space occupied by master data records duplication*” is used to ensure that there are no duplicate master data records that may affect the efficiency of the master data repository. It is important to note that one of the worst consequences of low levels of Efficiency in the case of potentially duplicated records is the probability of propagating inaccurate values to other transactional systems.

Portability. Organizations need to exchange data with other partners during the execution of their business process (Hüner *et al.*, 2011; ISO, 2009; Rivas *et al.*, 2017). Different systems can have other implementations of the same master data, being less portable. Therefore, the more portable the data, the lower the costs of exchange and integration (Silvola *et al.*, 2011). For example, when backing up data for recovery purposes, it would be necessary to allow maximum portability of the master data to lose information. Portability can be measured with the properties “*Master Data Portability ratio*”, “*Prospective data portability*”, or “*Architecture element portability*”. Low levels of portability will thus anticipate low chances of success when installing or moving master data from one system to another, with the significant risk of not being able to preserve existing levels of quality of master data.

Precision. This data quality characteristic should be differentiated from Accuracy. Precision is more related to concerns related to the ability to understand that 5.0001 and 5.0002 are different values. It can be measured by “*Precision of data values*” or “*Precision of data format*” properties. Low levels of Precision are symptomatic of not adequately propagating the corresponding precision policies to data source repositories. The study of this characteristic does not differ much from that of regular data, although it is still vital to study the quality of the master data.

Traceability measures the extent to which it is possible to create and follow evidence for audits about the access and any other changes made to the master data record values or master data models. In the specific case of master data, it is desirable to have high levels of traceability to validate the master data life cycle. This characteristic can be measured through the properties “*Traceability of data values*”, “*Users access traceability*”, and “*Data Value traceability*”. Low levels of traceability are symptomatic of a not well-documented master data life cycle.

Table 4
Example of clause extracted from ISO 8000-110.

“The data dictionary shall be accessible in electronic form under the terms of its license in one of the following ways:

- a) The data dictionary entry that contains the supporting terminology is downloadable from the Internet.*
 - b) The data dictionary supports an application programming interface (API), available over the Internet, for the resolution of an entity reference to the data dictionary entry that contains the supporting terminology. The reference shall be in the form of an unambiguous identifier.*
 - c) The data dictionary entry that contains the supporting terminology is contained physically in the same data set as the property value.”*
-

Understandability. To conveniently exploit master data, users must understand the meaning of the metadata in dictionaries and other metadata sources that characterize the data (ISO, 2009). Understandability covers several concerns that can be measured through several properties like “*Symbol understandability*”, “*Semantic understandability*”, “*Master data understandability*”, and “*Data model understandability*”, to cite a few. Low levels of Understandability are symptomatic of an inefficient design of the master data system. The most challenging consequence of low levels of understandability is the risk of producing inadequate master data values that can hurt the design.

3.2. Business Rules for Master Data Inferred from ISO 8000-100 Series

This section presents the minimum set of business rules inferred from the ISO 8000-100 series study to be used as a reference for master data quality evaluation and certification. To our knowledge, no other open standard (i.e. developed by ISO or IEEE) explicitly addresses the business rules to be met by master data. We found only a family of standards containing requirements for the exchange of characteristic data in master data (see Section 2.3) that can be considered as the source of mandatory business rules that master data repositories must comply with. Consequently, we consider only the normative clauses of the ISO 8000-100 (mainly ISO 8000-110) series to identify the mandatory business rules. For instance, Table 4. gathers one of the clauses extracted from ISO 8000-110, which we inspected to infer and state from sub-clauses a) and b) the following business rule: “*BR.01. Master data dictionary terminology must be available online, downloadable or using an API and it should be portable*”. Following this process, business rules were inferred (see Table 5). These business rules have been grouped for the various characteristics and properties (see Table 6) following the BR4DQ methodology (Caballero et al., 2022).

It is essential to state that in data quality evaluation projects, these business rules are typically outside the development scope. So, as part of the evaluation, the discovered business rules must be mapped against the ones presented here by comparing the corresponding statements.

3.3. Tailoring the Evaluation of the Data Quality Characteristics

The customization of the measurement of the properties of the master data quality model involves the development of the following components:

Table 5
Business rules inferred from ISO 8000-110.

-
- BR.01. The possible values of an attribute of the master data record must be stored in a metadata repository (i.e. master data dictionary, reference data), which must be available online, downloadable, or accessible by using an API and should be portable.
 - BR.02 Master data dictionary and master data repository must be accessible anytime, anywhere, from the same system or externally deployed.
 - BR.03. All master data records must be uniquely identified, and this identifier must follow a standardized syntax broadly recognized by the organization.
 - BR.04. All master data records must have a timestamp attribute in a specific format as UTC and the user who modifies each attribute.
 - BR.05. The relationship between master data attributes must be defined using their unique identifier.
 - BR.06. The information about data provenance of master data records must be reachable.
 - BR.07. Master data records must be coded using concepts and references to a data dictionary.
 - BR.08. Master data must conform to a data specification partially or totally expressed from their metadata.
 - BR.09. Master data must conform to a formal syntax following the metadata.
 - BR.10. Master data records must indicate their unit of measure if it applies, and this unit of measure must be in defined ranges.
 - BR.11. Master data record semantics must be defined through their metadata.
 - BR.12. Understandability of Master data must be raised and defined through their metadata.
 - BR.13. The accuracy of the master data records must be described and achievable through their metadata or data specification.
 - BR.14. Master data records can follow more than a data specification; in this case, it must comply with more than one formal syntax.
 - BR.15. The completeness of master data records must be described using the data specification or the metadata needs.
 - BR.16. Using unique identifiers focuses on reducing duplicated records.
 - BR.17. Master data records must occupy the space indicated by the metadata to increase system efficiency.
-

- Define the measurement methods (labelled as ❶ in Fig. 1) for the tailored data quality properties introduced in Section 3.1 (see Table 7 for the original description provided in ISO/IEC 25024). The tailoring requires specific business rules (labelled as ❷ in Fig. 1), which have been previously presented. For example, in Fig. 1, with the corresponding measurement methods, we would obtain 60 for “*syntactic precision*”, 84 for “*semantic precision*”, and 45 for “*master data accuracy assurance*”.
- Define the Quality Value Range Matrix (labelled as ❸ in Fig. 1) to transform the DQ properties’ measurements into quality levels. In the framework, the levels of quality range between the Likert labels of 1 (the worst) and 5 (the best). Since the original ranges cannot be shown because they belong to the Intellectual Property of the owners of the framework, we can only present an example. This matrix would transform the previously measured value into 3, 4, and 3. A 5-dimensional vector called quality profile (labelled as ❹ in Fig. 1) must be mounted with these values. The *i*-eth position of the vector represents the number of properties with an ‘i’ level of quality according to the quality. For example, it is possible to mount the quality profile, vector [0, 0, 2, 1, 0]: the third position has a value of 2 because two properties have a level “3” of quality. The fourth position has a value of 1 because there is only a property with a level of quality of “4”.

Table 6
Business rules inferred from ISO 8000-110 grouped for data quality characteristics.

Data quality characteristic	Data quality property	Business rules
Accuracy	Range of Accuracy	BR.08, BR.10, BR.13
	Semantic Accuracy	BR.01, BR.07, BR.08, BR.11, BR.13
	Syntactic Accuracy	BR.03, BR.04, BR.08, BR.09, BR.13, BR.14
Availability	Architecture Elements Availability	BR.01, BR.02
	Data Availability Ratio	BR.01, BR.02
	Data Portability Ratio	BR.01
Completeness	Data Value Completeness	BR.05, BR.08, BR.10, BR.15
	False Completeness of File	BR.08, BR.15
	File Completeness	BR.05, BR.08, BR.15, BR.16
	Record Completeness	BR.03, BR.08, BR.10, BR.14, BR.15
Compliance	Regulatory compliance due to technology	BR.01, BR.02
	Regulatory compliance of value and/or format	BR.03, BR.04, BR.07, BR.08, BR.09, BR.14
Consistency	Format Consistency	BR.04, BR.05, BR.07, BR.08, BR.09, BR.11
	Referential Integrity	BR.05
	Risk of Inconsistency	BR.05, BR.07, BR.08, BR.16
	Semantic Consistency	BR.05, BR.07, BR.08,
Credibility	Data Values Credibility	BR.03, BR.06, BR.07, BR.08, BR.09, BR.11, BR.14
	Source Credibility	BR.06, BR.08
Currentness	Timeliness of Update	BR.06
	Update Frequency	BR.06
Efficiency	Data Format Efficiency	BR.03, BR.04, BR.07, BR.08, BR.09, BR.11, BR.14
	Risk of Wasted Efficiency	BR.02, BR.05, BR.07, BR.08, BR.09, BR.16, BR.17
	Space occupied by records duplication	BR.05, BR.08, BR.16
Portability	Architecture Elements Portability	BR.01
Understandability	Linked Master Data Understandability	BR.03, BR.05, BR.07, BR.08, BR.09, BR.11, BR.12, BR.14
	Master Data Understandability	BR.03, BR.05, BR.06, BR.08, BR.09, BR.11, BR.12, BR.14

Table 7
Measurement of 'Master data understandability' property adapted from ISO/IEC 25024.

Data quality characteristic	Understandability
Data quality property	Master data understandability (MAD_UND)
Measurement description	Master data understandability focuses on checking if master data is due to the metadata definition.
Calculation formula	$X = A/B$ A = number of data items of master data files with existing metadata. B = number of data items of master data files.
Scale	Ratio
Value range	[0.0, 1.0]

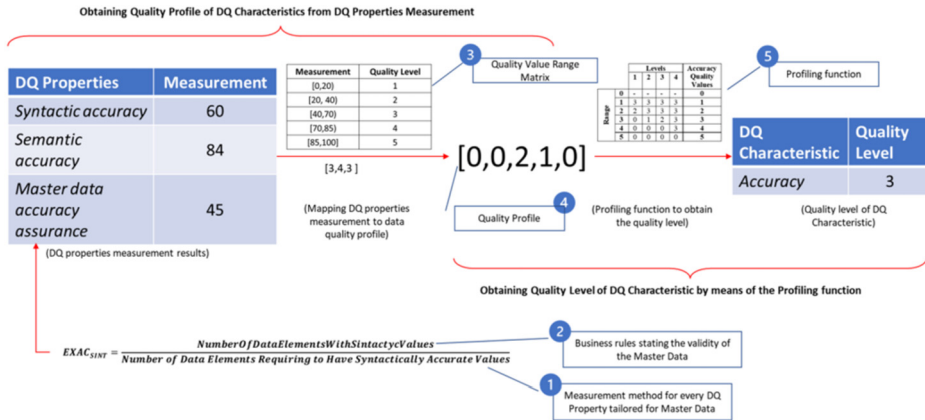


Fig. 1. Obtaining the quality level of a DQ Characteristic from DQ properties measurements.

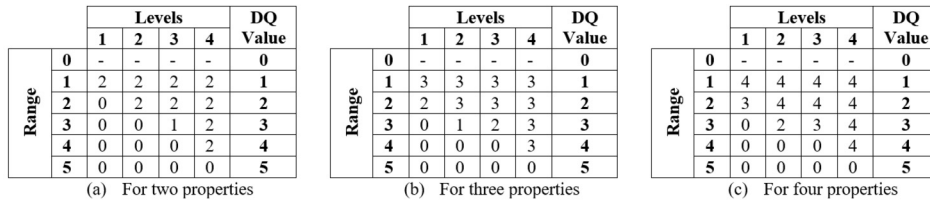


Fig. 2. Profile functions proposed to evaluate data quality properties characteristics.

- Define the profiling functions (labelled as 5 in Fig. 1) for every data quality characteristic to produce a quality level. These profiling functions transform the quality profile into a quality value. Once again, the profiling functions defined in the framework for master data cannot be shown due to Intellectual Property restrictions. However, we introduce in Fig. 2 several generic profiling functions as an example to evaluate data quality characteristics with (a) two, (b) three, and (c) four measurable properties. Every row (range) shows the maximum number of properties at a specific level. The columns on the right include the value of quality for the data quality characteristics. The way to produce the characteristics of the quality value for the DQ is to find a range containing the maximum number of properties at the specific level. Regarding the example of ‘accuracy’ in Fig. 1, with the evaluation profile [0, 0, 2, 1, 0], since accuracy has three properties in our master data quality model, the profiling function (b) must be chosen. As there are two properties with level 3 and 1 with level 4, it is easy to see that the value of quality for accuracy is 3.

3.4. Certifiable Data Quality Characteristics

Granting a certification is a form of precise pointing by independent third parties that a set of characteristics, merits, or conditions of a fact or good are acceptably adequate. Certifica-

tion requires an evaluation with results that are necessarily objective, absolute, repeatable, unambiguous, unbiased, and comparable without the need to explain the specific context in which it has been measured. The data quality evaluation framework can also be used to certify the level of quality of the master data repository. Following AENOR's guidelines (as stated in Gualo *et al.*, 2021) for software and data quality certification, the certification of an adequate level of quality for a master data quality characteristic is granted when, at least, a value of quality of "3" over "5" is obtained.

The certification process typically involves costs that include the evaluation and the management of the granting of the certification. Not all the data quality characteristics and properties introduced in Section 3.1 are susceptible to be certifiable (i.e. they do not provide relevant information) or worthy (i.e. they change very quickly over time or because the cost/benefit analysis reveals that they are not good benefits are obtained). To optimize the costs, we decided to establish a set of selection criteria to discriminate those being susceptible or worthy to be certified. The criteria are:

- C1. Eligible master data quality characteristics should provide relevant information on the data quality status of the master data repository so that they show **non-transitory states** of upgrading, installing, deploying, or maintaining master data management systems. This criterion facilitates achieving **absolute results over time**, i.e. not relative to projects in progress. However, it is important to note that not including such characteristics in the certification does not mean they are no longer useful. Rather, they would not give a potentially actionable idea of an absolute status over time.
- C2. Selected master data quality characteristics have the property(ies) that must rely on well-proven measurement methods to produce results **independent of any non-standard aspects that condition the measurement results**, such as the purely technical aspects of the systems they manage. These aspects facilitate **comparability**; that is, two different organizations measuring the same properties on their respective master data repositories can make a clear and independent interpretation of the results regarding the other organization.
- C3. Measurement methods for the chosen properties should rely on formal audit techniques to systematically limit or eliminate **ambiguities in capturing and evaluating the evidence** needed to measure the chosen properties. In addition, such techniques should promote **repeatability of measurements**, i.e. different and independent groups of auditors must obtain the same results on the same master data repositories with the same evaluation scope.
- C4. Since the intention is to develop a model that certifies the quality of master data concerning the requirements introduced in the ISO 8000-100 series of standards, following the philosophy introduced in this work, the established criterion is that **there must be specific business rules inferred from ISO 8000-100 that can be associated to the certifiable master data quality characteristic**.
- C5. We intend to develop a consistent, simple model with a minimum but enough number of data quality characteristics and properties. During the selection process, some eligible characteristics may be prioritized to the detriment of others. This prioritization is based on the **relevance and positive impact of possible results for optimizing the organization's business processes**.

Table 8
Proposed characteristics for master data quality certification.

DQ Charac.	C1	C2	C3	C4	C5	Selected?	Rationale under the selection as certifiable
Accessibility	Y	N	N	N	N	N	The measurement of Accessibility properties heavily depends on various types of user disabilities and devices with which users can access master data records. For this reason and simplicity, accessibility is not initially selected as certifiable.
Accuracy	Y	Y	Y	Y	Y	Y	This characteristic meets all criteria, and it is usually very relevant. In addition, it is the basis for measuring other characteristics, so it is selected.
Availability	Y	Y	Y	Y	N	N	This data quality characteristic fulfills the first three principles and is relevant to business processes, but the measurement depends on many technological concerns; consequently, it is not included in this first version.
Completeness	Y	Y	Y	Y	Y	Y	Consideration of this data quality characteristic is always desirable and relevant for master data records used by different business processors.
Compliance	Y	Y	Y	Y	Y	Y	This data quality characteristic meets the first three principles and is relevant to business processes, so it is included in this first version of the model.
Confidentiality	Y	Y	Y	N	N	N	Despite complying with the first three principles, confidentiality can be on a second level when it comes to prioritizing the data quality characteristics so that in this first version, it would not be included in the model.
Consistency	Y	Y	Y	Y	Y	Y	This characteristic complies with the first three principles, and, in addition, it is considered that its measurement and evaluation would add value to the organization's business processes, so it is included in this first version.
Credibility	Y	Y	Y	Y	Y	Y	Credibility is relevant to the data used by the organization's business processes and satisfies the first three requirements.
Currentness	Y	Y	Y	Y	Y	Y	This data quality characteristic meets all the above criteria and is also a very relevant data quality characteristic for the data in the organization's business processes.
Efficiency	Y	Y	Y	Y	Y	Y	Efficiency fulfills all criteria and is also very relevant for master data management.
Portability	N	Y	Y	Y	N	N	Although this characteristic could meet criteria 2 and 3, it is not initially selected because it would not be an initially relevant characteristic, and its study would be included in very restricted and specific areas in managing master data.
Precision	Y	Y	Y	Y	Y	Y	This characteristic is directly related to accuracy, and it is also crucial to study efficiency. Therefore, it is very relevant for certification. In addition, it would meet the first three criteria.
Recoverability	Y	Y	N	N	N	N	It is considered that this characteristic would have a second-order level of relevance as it deals with specific aspects of master data management, so it would not be included in the first version of the model.
Traceability	Y	Y	Y	N	N	N	This data quality characteristic is very relevant in the case of master data since the values of master data attributes come from different data repositories. However, it is not included in this first version for two reasons: (1) it does not have a significant contribution directly to the exploitation of master data, and (2) its implementation would require the inclusion of much metadata that may be difficult to include for most organizations, at least for the less mature ones.
Understandability	N	Y	N	Y	N	N	This characteristic would be considered the basis for credibility to establish a good relationship between understandability and credibility. In this sense, to simplify the model, we decided not to include it in this first version.

Table 8 shows a complete overview of the proposed certifiable subset of data quality characteristics for the evaluation of master data repositories once the criteria are applied.

Once the certifiable data quality characteristics were selected, the next step was to identify the underlying data quality properties that can contribute notably to evaluating

Table 9
Proposed data quality characteristics and properties for master data certification.

Certifiable DQ Characteristic	# DQ properties considered	DQ Property Acronym	Selected DQ Properties
Accuracy	3	EXAC_SEMAN	Semantic Accuracy
		EXAC_SINT	Syntatic Accuracy
		ASEG_EXAC_MD	Master Data Accuracy Assurance
Completeness	4	COMP_REG_MD	Master data record completeness
		COMP_FICH_MDM	Master data file completeness
		FAL_COMP_FICH_MD	Empty master data record in a master data file
		COMP_VAL_DM	Completeness of Master Data Values
Compliance	2	COMPL_REG_TEC	Regulatory compliance due to technology
		COMPL_REG_VAL	Regulatory compliance of value and/or format
Consistency	4	CONS_FORM_DM	Master data format consistency
		CONS_SEMAN_DM	Master data semantic consistency
		INT_REF	Referential Integrity
		RIES_INCO_DM	Risk of master data inconsistency
Credibility	2	CRED_FUEN	Master data repository credibility
		CRED_VAL_DM	Credibility of master data value
Currentness	2	CONV_ACT	Timeliness of update
		FREC_ACT	Update frequency
Efficiency	3	EFIC_FORM_DM	Efficient master data item format
		EFIC_DUP_REGDM	Space occupied by master data records duplication
		RIES_ESPAC_DESP	Risk of wasted space
Precision	1	PRE_VAL_MD	Precision of Master Data Values

the data quality characteristics. For this second level of selection, we listed three criteria: (1) the measurement of the property must be done in an objective and repeatable way; (2) there must be specific business rules inferred from ISO 8000-100 that can be used as the basis for the measurement; and (3) the measurement of the property must be relevant and consistent to measure the data quality characteristics. Table 9 shows the data quality model with the list of certifiable data quality characteristics and the selected specific data quality properties.

4. Case Study

The primary purpose of the case study is to validate that the model is complete, minimal, and contains all the data quality characteristics relevant to master data. It is entirely usable in the real world. Thus, this section describes the application of the framework for master data evaluation to a real case study following the methodology based on ISO/IEC 25040, which is further described in Gualo *et al.* (2021). This methodology consists of five steps. To make the paper brief, we only present some specific details for only two data quality characteristics (Accuracy and Efficiency), although all results are shown.

Table 10
List attributes, descriptions, and example values from the master data repository.

Attribute	Example valid values	Description
User_Id	"12345678A"	The Spanish identifier for each person.
Email	"nsurname@company.es"	The email account for each person
Categories	["Software Engineer I", "Software Engineer I", "Hardware Engineer", "Analyst-Developer"]	A list of the different categories in the company.
Department	["Human Resources", "Accountability", "IT area", "Management"]	A list of the several departments in the company.
DateAdded	"2017-01-08T19:12:13Z"	The date the persona was first added to the master data repository.

Table 11
Example of metadata in data specification.

Attribute	"User_Id"		
Type	String	Nullable	False
Default value	00000000A	Unique	Yes
Min Length	9	Max length	10
Specific range	No	Encoded	Yes
Specific syntax	Yes	Mandatory	Yes
REGEXP	[0-9]{8} [A-Z]{1}		

4.1. Description of the Data from the Evaluated Master Data Repository

The evaluated master data repository integrates several data sources of a software company (called the 'Client') to provide a 360° view of employees. In this sense, the master data repository is part of the organizational data lake, and the whole organization uses it through different departments with different goals. Unfortunately, we were not allowed to provide further information about the organization that owns the master data repository.

This master data repository and data dictionary was implemented in a relational database management system (MySQL 5.7.6) to enable the required SQL queries necessary to compute the data quality measures. The master data repository comprises 48 master data attributes with 59,387 records. Several triggers were implemented in the master data repository to track the changes in the master data records values. Table 10 shows an excerpt of the most representative attributes, their description, and an example of the valid value they can take. An example of metadata for the attribute "User_Id" is shown in Table 11. Finally, there is a data dictionary containing the reference data that should be used in assigning a value to an attribute when necessary. In this case, the data dictionary is organized into levels. Table 12 shows an excerpt from this data dictionary.

4.2. Evaluation of the Quality of the Master Data Repository

As part of **Activity 1. Establishment of the evaluation requirement**, the Client's team presented the master data repository along with design and implementation details, including the business rules governing the validity of the data in the master data repository.

Table 12
Excerpt of Data dictionary for the case study.

Level 1	Level 2	
	Reference	Value
Department	DPT0001	Human Resources
Department	DPT0002	Accountability
Department	DPT0003	IT area
Category	CAT0001	Software Engineer I
Category	CAT0002	Hardware Engineer I

Table 13
Matching between sets of mandatory and obtained business rules for the master data repository.

Mandatory business rules	Business rules obtained from the master data repository of the case study.
BR.01	Master data dictionary and master data repository are portable in CSV and JSON format.
BR.02	Master data dictionary and master data repository are deployed in MySQL 5.7.6.
BR.03	The “ <i>Id_User</i> ” or “ <i>Email</i> ” attribute uniquely identifies each master data record.
BR.04	The “ <i>DateLastUpdated</i> ” and “ <i>UserLastUpdated</i> ” attribute identifies the last update date.
BR.05	The relation between data values with the data dictionary entries is using this unique identifier (i.e. the “ <i>Department</i> ” attribute takes the value “ <i>DPT0001</i> ”, which is the entry in the data dictionary which represents the Human Resources department.)
BR.06	Every master data record has the attributes “ <i>DateLastUpdated</i> ” and “ <i>UserLastUpdated</i> ” to reach the changes in the master data repository.
BR.07	Some attributes like “ <i>User_Id</i> ” must be encoded.
BR.08	The “ <i>User_Id</i> ” attribute must not be null or unique and follow the specified regular expression.
BR.09	“ <i>User_Id</i> ” must follow this syntax (i.e. 00000000A)
BR.10	The “ <i>EnglishLanguageLevel</i> ” attribute indicates the unit of measure for their language aptitudes (i.e. Cambridge B1, Trinity B1).
BR.11	Data dictionary entries specify the semantics for data values.
BR.12	Data dictionary entries specify the semantics for the understandability of data values.
BR.13	“ <i>User_Id</i> ” accuracy is defined with min and max lengths in their metadata or data specification.
BR.14	In this case, it is not applicable.
BR.15	“ <i>Id</i> ” must not be null because it is part of the unique identifier for each master data record.
BR.16	The “ <i>Id_User</i> ” and “ <i>Email</i> ” attributes help to reduce the duplicity.
BR.17	The data specification defines the maximum length of space occupied for each data type.

Afterward, it was decided that the scope of the evaluation would include all certifiable master data quality characteristics (see Table 9).

As the master data quality granted for a data quality certification, the business rules required for evaluating the selected data quality characteristics must be assimilated to the reference set of business rules identified in Table 6.

This action was done using the metadata and the provided study of the data model. For example, through the example metadata in Table 11, some business rules were identified: “*User_Id* is mandatory, so it cannot be null and not duplicated”, “*User_Id* max length is 10, so the attribute must not support more space” and “*User_Id* must follow the specific formal syntax: [0-9]8 [A-Z]1”, . . . (see Table 13). Conducting this analysis for each attribute in the master data model and considering the data dictionary, too, 567 business

Table 14
SQL script example to evaluate *EXAC_SINT* data quality property.

```
SELECT COUNT (*)
FROM 'SwFact_MDM_Repository'
WHERE 'SwFact_MDM_Repository'.User_Id REGEXP '^([0-9]{8} [A-Za-z-]{1})$' AND 'SwFact_MDM_
Repository'.dateadded
REGEXP '^([0-9]{4}-[0-9]{2}-[0-9]{2})T([0-9]{2}):([0-9]{2}):([0-9]{2})$'
```

rules were identified for the 'Employee' repository and conveniently grouped for every data quality characteristic. The execution of this activity included, as part of the plan for master data quality evaluation, some agreements on granting of permissions to access the master data repository, the necessary security restrictions (that is, VPN connection), the working schedule restrictions for the connection in order not to interfere with the organization's operational activities, as well as the time frame in which the evaluation results should be provided. In this regard, it is worth noting that although a data access service layer has been enabled to run the evaluation more efficiently, and given the experimental nature of the case study, the Client allowed access to the master data through a repository clone.

In **Activity 2. Specification of the evaluation**, the measurement methods, and corresponding quality levels for the measures of each data quality property were introduced so that the organization can understand the evaluation results and their impact on the business process. This understanding allows refinement of the information provided for the master data repository. It also enables fine-tuning the implementation details of the business rules for the data repository. The evaluation team uses the acquired knowledge when conducting Activity 4 to develop the elements required to generate the specific measurements. In this sense, some data quality properties may be measured by inspecting the results of SQL scripts (see Table 14). In contrast, others may require other mechanisms for data extraction, such as queries to data dictionaries, system logs, or the development and execution of specific programs beyond SQL scripts.

During the execution of **Activity 3. Evaluation design**, the scope of the evaluation is refined to the core of what is to be evaluated and is fully defined in detail. Once the scope is delimited, the evaluation plan is fully established with the particularized activities and tasks to execute the evaluation.

In the execution of **Activity 4. Conducting the evaluation**, the objective is to obtain the quality level for the selected data quality characteristics following the evaluation plan defined and refined during Activities 1 to 3. The tasks required in the execution of this activity (recall Fig. 1) are the following:

T.4.1. Design the evaluation scripts and the necessary mechanisms for checking the validity of the business rules, considering the intrinsic or system-dependent nature of the characteristics to be evaluated. Taking as input (1) the business rules already assimilated to the reference set of business rules for the master data repository and grouped for each data quality characteristic and each property, (2) any other documentation provided by the organization seeking evaluation, and (3) the information obtained in the meetings during

Table 15
SQL script example to evaluate *EFIC_{FORMDM}* data quality property.

```
SELECT * FROM information_schema.TABLES
WHERE TABLE_NAME = 'SwFact_MDM_Repository'
```

Table 16
Results of the application of the measurement methods for the DQ properties.

DQ characteristics	DQ property	Measure description	Obtained value from the execution of the evaluation script
Accuracy	Syntactic accuracy	Number of data elements with syntactically correct values	740
Accuracy	Syntactic accuracy	Number of data elements required to have syntactically accurate values	870
Efficiency	Risk of wasted space	Size in bytes of space occupied due to duplicate records	250588
Efficiency	Space occupied by master data records duplication	Size in bytes occupied due to unduplicated records	225530

Activity 1, the evaluation team develops a set of evaluation scripts together with other technological mechanisms necessary to check the degree of compliance of each of the business rules defined for each entity in the scope of the evaluation. For example, Table 14 shows an example of an evaluation script that aims to check compliance with business rule BR.09 of Table 5 groups under the quality property “*Syntactic Accuracy*” for the Accuracy Data Quality characteristic, or even Table 15 shows an evaluation script for the evaluation of the property “*Efficient Master Data Item Format*” of the Efficiency characteristic.

T.4.2. Execution of the evaluation scripts and the mechanisms necessary to check the business rules. Following the evaluation plan, the evaluation scripts or the corresponding mechanism for each master data entity are executed, as well as any other instruments developed to obtain the base measures for measuring each data quality property for each chosen characteristic. From the results of this run, the weaknesses and strengths of each data quality property can be profiled. For example, Table 16 shows some of the results obtained after the execution of one of the evaluation scripts designed to measure the indicated properties on the master data repository of the case studied.

T.4.3. Production of the quality value for the quality properties. The results obtained from the evaluation scripts (see Table 16) are used to calculate the measurement values of each data quality property by applying the corresponding measurement functions conveniently tailored for every master data entity (see Section 3.3). As an example, the property “*Semantic Accuracy*” (*EXAC_{SINT}*) would be measured by the formula presented in Eq. (1), obtaining a value of 100, and the property “*Space occupied by record duplication*” (*EFIC_{DUP_REGDM}*) calculated by the formula shown in Eq. (2) obtained a value of 95. The results for the measurements of all data quality properties considered for the

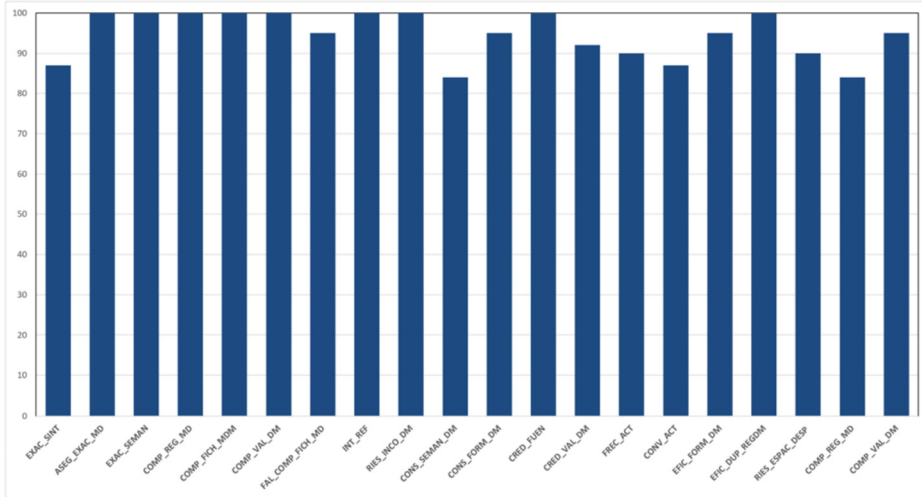


Fig. 3. The value obtained for data quality properties during the evaluation of the master data repository.

case study are shown in Fig. 3.

$$EXAC_SINT$$

$$= \frac{\text{NumberOfDataElementsWithSintactycValues}}{\text{NumberofDataElementsRequiringtoHaveSyntacticallyAccurateValues}}, \quad (1)$$

$$EFIC_DUP_REGDM$$

$$= \frac{\text{SizeInBytesOfSpaceOccupiedDueToDuplicateRecords}}{\text{SizeInBytesOccupiedDueToRecordsWithoutDuplication}}. \quad (2)$$

These measurement values should be then analysed and discussed to collate the weaknesses and strengths previously identified for each data quality property to ratify and identify critical aspects that can be improved. It is important to note that although the definition of the measurement method follows ISO/IEC 25024, its application requires customization in terms of the specific semantics of the master data entities under evaluation and the technological aspects of the system of which the evaluated data repository is part.

T.4.4. Derivation of the quality level of data quality properties from their quality value. Data quality characteristics are evaluated using a specific profiling function (see Section 3.3) that requires an intermediate step in which the quality values of each data quality property are transformed into quality levels using a quality value range matrix. As an example, using the one provided in Fig. 1, “*Syntactic Accuracy*”, which obtained a measure of 85, is mapped to 4, “*Semantic accuracy*”, which obtained a value of 100, is mapped to 5, “*Master Data Accuracy Assurance*”, obtained 100 is mapped to 5, and “*Space occupied by duplication of master records*”, which obtained a value of 90 is mapped to 5. With these results, the corresponding quality profiles for accuracy are represented by the vector [0, 0, 0, 0, 3], and for Efficiency, the corresponding vector is [0, 0, 0, 1, 2].

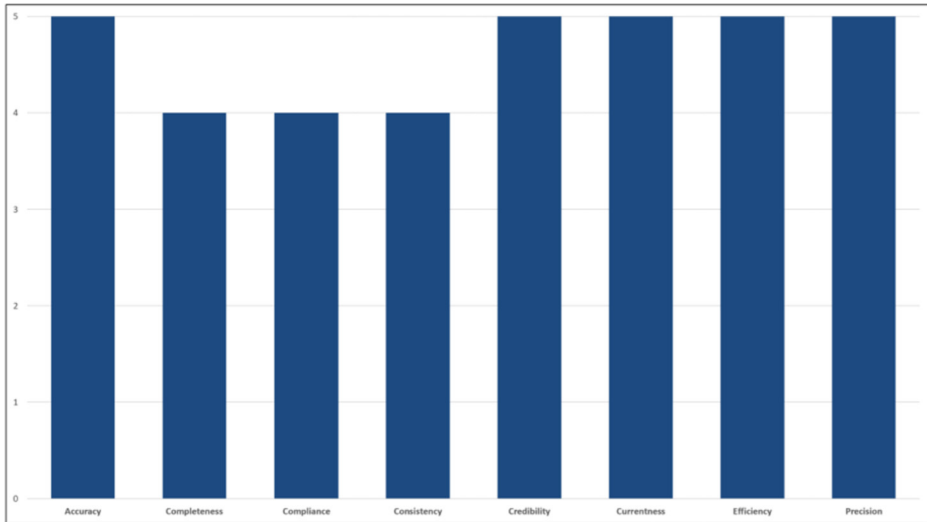


Fig. 4. Data quality level results for each data quality characteristic in the master data repository evaluation.

T.4.5. Determination of the quality level for the selected data quality characteristics.

Based on the data quality level for each data quality property, the next step is to generate the values of quality for the selected data quality characteristics by applying the corresponding profiling function introduced in Fig. 2 (the selection depends on the number of properties chosen for every data quality evaluation). See Fig. 4 for the results of the application to the “Employee” master data repository for all selected master data quality characteristics.

Finally, **Activity 5: Conclude data quality evaluation** produces a detailed evaluation report reflecting the quality levels achieved for the selected data quality characteristics and the values obtained for the corresponding quality properties. For example, the data problems encountered during the evaluation of the data quality characteristic Accuracy are related to the *EXAC_SINT* property. These problems happened because some attributes did not follow the specified syntax. After all, they take null values. Along the same lines, it is possible to state that some of the data problems encountered during the evaluation of the data quality characteristic Efficiency were as follows:

- *EFIC_FORM_DM* did not meet the maximum quality value for the data quality property because some master data records did not meet the specific format (i.e. “phoneExtension”).
- *RIES_ESPAC_DESP* did not reach the maximum quality value for the data quality property because the space reserved for the attributes was much larger than the space needed to store the data value (i.e. “phoneNumber”).

In addition to the evaluation report, a comprehensive improvement report can be prepared and provided to the organization, detailing the weaknesses and strengths of each measured data quality property. This improvement report focuses on the properties that did not achieve an adequate level of quality. It details the causes of those low levels so that

the organization can take steps to improve them. In addition, the organization provides a set of scripts to identify specific records that require improvement actions. Finally, to complete the evaluation, the access permissions to the master data repository and other assets needed for data quality evaluation are removed or revoked.

4.3. Discussion and Principal Findings of the Case Study

Explanations can be split into two parts to demonstrate the validity of the proposed work. On the one hand, the data quality evaluation framework has already been tested and validated (Gualo *et al.*, 2021). On the other hand, we intended to validate the proposed master data quality model. In this sense, through the case study, we could check that evaluating the data quality characteristics included and revisited in the proposed model can provide helpful information to monitor and improve the quality of the master data repository. In all contexts, master data quality analysts must select those providing this valuable information. In this sense, it is a matter for the data quality analysts to choose the minimum number of data quality characteristics that can provide the required information. On the other hand, since the case study has been conducted on a master data repository of a software industry, we can conclude that the proposed master data quality model is applicable in the real world – even when we recognize that only one case could not be enough to generalize this last statement. However, we assume some threats to the validity of our proposal (see Section 5).

In addition, after analysing the process of conducting the case study, we have reached a series of findings:

- **The need to have a complete and widely accepted shared master data definition to improve the understandability of the master data schema.** In this case study, the master data schema consists of 48 attributes. It is hard for stakeholders and data quality analysts to manage, even when it is not a large number, without an adequate description. Additionally, it is not easy to understand how to manage the level of data quality for these attributes. Summarizing: Without knowing the meaning of the master data entities and their attributes, the data quality evaluation lost its sense.
- **The need to have a complete data dictionary for usability.** In our study case, a data dictionary containing metadata about the master data repository is available. This data dictionary includes data types, business rules (i.e. describing allowed values), and other necessary metadata related to the number of elements and non-empty attributes. Something important to note is that some metadata can be used to compute the data quality measurement, but per se, they are not data quality measures. This fact has made the definition and implementation of the measurement method easier for us for the various master data quality properties selected for evaluation. In addition, to better manage the evolution of the data quality evaluations, we saw it necessary to attach the version of the data dictionary to the version of the results of measuring the data quality properties. Summarizing: the more it is known about the master data using metadata, the better definition of the master data quality measures, and the more insights it is possible to obtain from the evaluation to guide improvements to the master data repository better.

- **The need to have adequate technological and organizational support for the master data repository.** The measurements were produced outside any business context and without technological, business, or temporal restrictions for the case study. In a different business scenario, the master data quality evaluation should coexist with other data-related operations. From the technological perspective, the information system should support the evaluation processes to produce adequate evaluation results. From the business point of view, management should guide the evaluation process to not interfere with the business processes that consume data from the master data repository. However, this fact is usually mitigated as the master data follows the principle “*write once, update little, read many*”. Furthermore, from the material or timeliness point of view, even when the changing pace of the master data is plodding, the evaluation processes should be adequately synchronized with the update so that the correct values are read during the evaluation. In summary, the most mature organizations (those with enough selected and deployed technologies, well-organized master data governance processes, and correctly defined business processes) obtain better results for evaluating the quality of the master data repository.

5. Threats to the Validity

In the following subsections, the threats to the validity of this initial state of the framework are analysed based on the aspects identified (Runeson *et al.*, 2012; Runeson and Höst, 2009). **Construct validity.** In the certification model, we included eight certifiable data quality characteristics and 21 data quality properties in the proposed model (see Table 9) because, from our experience and through the study of the related literature, we considered that it could be successfully used in most scenarios. Another possible threat to construct validity could be grounded on the results’ domain dependency (i.e. CRM, ERP, PDM, ...). Still, given the genericity of the definition of the measurement procedures, we pose that it could be easily tailored to the specifics of any domain. Finally, the last threat to the construct validity is the principled threshold values we define for the quality value range matrix and the profiling functions. This decision was based on our experience conducting a data quality evaluation of relational repositories (Gualo *et al.*, 2021). However, it is essential to recall that the values we use in industrial projects cannot be shared because they belong to the owners of the data quality evaluation and certification framework.

Internal validity. A threat to the internal validity of the proposed framework could be the quality and stability of the business rules used during the definition and tailoring of the measurement of every data quality property. Especially important for the sake of a compliance certification to ISO 8000-110 is the selection of business rules that we infer from the ISO 8000-100 series. Nevertheless, to the extent of our knowledge, there are no more international open standards specifically addressing master data. However, if some new related open standards appear, the set of references to business rules can be revisited to cover the requirements introduced.

External validity. The proposed master data quality model must be applied to companies and organizations to increase external validity. However, complying with several ISO standards generally ensures a certain level of applicability to all domains.

6. Conclusions

Master data is one of the unique types of data in organizations. It holds the knowledge that the organization needs for their day-by-day operations, i.e. they are the reference for the transactional data. This capability of being referenced means that the master data values are propagated through the organizations as master data is used for other purposes, i.e. transactional purposes. Consequently, if the values stored in the master data repository do not have adequate levels of quality, then low-quality data are propagated through the entire organization and maybe externally.

In this paper, we presented a data quality model for evaluating the quality of master data repositories. This master data quality model can be used with the evaluation framework we developed. This evaluation framework has been successfully used to evaluate the quality of repositories of regular data repositories. Master data has specific features beyond regular data (i.e. transactional relational data) that make necessary the investigation of the particular concerns of master data quality. The tailoring of the data quality model involves the specific definition of the quality value range matrix and the corresponding profiling functions. As these elements are under the Intellectual Property protection of the owners of the data quality evaluation frameworks (they are exploiting the framework commercially), we have been allowed only to show examples of these elements.

Sometimes, granting a certification of the level of quality can improve the trustworthiness of a master data repository. Only some of the 15 data quality characteristics included in the data quality model are susceptible or worthy of certification. To limit the scope of certification and improve the usability of the results, we defined five criteria used to select data quality characteristics relevant to the business or the information systems if certified.

The results obtained either from evaluation or the certification process can be used to identify the most critical points or risks for a master data repository, enabling the guiding and optimization of the efforts in fixing errors and, consequently, making the master data an essential asset of the organization.

The most important conclusions we can raise are that the presented version of the data quality model is sufficiently complete and comprehensive to be used in master data quality evaluation and certification projects, easily understandable, and combined with the developed framework, it is applicable in the real world.

Acknowledgements

We would like Natalia Sánchez for her impressive support in formatting the manuscript.

Funding

The following projects and grants have partially funded this research: Industrial PhD (Ref.: DIN2018-009705), funded by the Spanish Ministry of Science, Innovation and Universities; ADAGIO (SBPLY/21/180501/000061) – Alarcos' Data Governance framework and systems generation, funded by the Department of Education, Culture and Sports

of the Junta de Comunidades de Castilla La Mancha, Spain, and the Fondo Europeo de Desarrollo Regional FEDER; and AETHER-UCLM: A smart data holistic approach for context-aware data analytics focused on Quality and Security project (Spanish Ministry of Science and Innovation, PID2020-112540RB-C42)

References

- Allen, M., Cervo, D. (Eds.) (2015). *Multi-Domain Master Data Management: Advanced MDM and Data Governance in Practice*. Morgan Kaufmann, Boston, USA. 978-0-12-800835-5. <https://doi.org/10.1016/B978-0-12-800835-5.09993-0>.
- Alonso, V., Santos, J.V., Pinto, M., Ferreira, J., Lema, I., Lopes, F., Freitas, A. (2020). Problems and barriers during the process of clinical coding: a focus group study of Coders' perceptions. *Journal of Medical Systems*, 44(3), 62. <https://doi.org/10.1007/s10916-020-1532-x>.
- Benkherourou, C., Bourouis, A. (2022). A framework for improving data quality throughout the MDM implementation process. In: *Proceedings of the 2nd International Conference on Industry 4.0 and Artificial Intelligence (ICIAI 2021)*. Atlantis Press International B.V., Amsterdam, Netherlands, pp. 164–169. 978-94-6239-528-2. <https://doi.org/10.2991/aisr.k.220201.029>.
- Berson, A., Dubov, L. (2011). *Master Data Management and Data Governance*, 2nd ed. McGraw-Hill Education, USA. 978-0-07-174458-4.
- Caballero, I., Gualo, F., Rodríguez, M., Piattini, M. (2022). BR4DQ: a methodology for grouping business rules for data quality evaluation. *Information Systems*, 109, 102058. <https://doi.org/10.1016/j.is.2022.102058>.
- Cleven, A., Wortmann, F. (2010). Uncovering four strategies to approach master data management. In: *Proceedings of the 43rd Hawaii International Conference on System Sciences*, pp. 1–10. <https://doi.org/10.1109/HICSS.2010.488>.
- Dahlberg, T., Heikkilä, J., Heikkilä, M. (2011). Framework and research agenda for master data management in distributed environments. In: *Proceedings of IRIS 2011, TUCS Lecture Notes*, Vol. 15, pp. 82–90. 978-952-12-2648-9.
- Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., Run, P.v., Wolfson, D. (2008). *Enterprise Master Data Management: An SOA Approach to Managing Core Information*. IBM Press/Pearson plc, London, UK. 978-0-13-236625-0.
- Fan, W., Geerts, F., Ma, S., Tang, N., Yu, W. (2013). Data Quality Problems beyond Consistency and Deduplication. In: Tannen, V., Wong, L., Libkin, L., Fan, W., Tan, W.-C., Fourman, M. (Eds.), *In Search of Elegance in the Theory and Practice of Computation: Essays Dedicated to Peter Buneman. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, pp. 237–249. 978-3-642-41660-6. https://doi.org/10.1007/978-3-642-41660-6_12.
- Gartner, I. (2021). Magic Quadrant™ for Master Data Management Solutions. <https://www.informatica.com/magic-quadrant-MDM.html>.
- Gualo, F., Caballero, I., Rodríguez, M. (2020). Towards a software quality certification of master data-based applications. *Software Quality Journal*, 28(3), 1019–1042. <https://doi.org/10.1007/s11219-019-09495-w>.
- Gualo, F., Rodríguez, M., Verdugo, J., Caballero, I., Piattini, M. (2021). Data quality certification using ISO/IEC 25012: industrial experiences. *Journal of Systems and Software*, 176, 110938. <https://doi.org/10.1016/j.jss.2021.110938>.
- Haneem, F., Ali, R., Kama, N., Basri, S. (2017). Resolving data duplication, inaccuracy and inconsistency issues using Master Data Management. In: *Proceedings of the 2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pp. 1–6. <https://doi.org/10.1109/ICRIIS.2017.8002453>.
- Haneem, F., Kama, N., Taskin, N., Pauleen, D., Abu Bakar, N.A. (2019). Determinants of master data management adoption by local government organizations: an empirical study. *International Journal of Information Management*, 45, 25–43. <https://doi.org/10.1016/j.ijinfomgt.2018.10.007>.
- Hüner, K.M., Schierning, A., Otto, B., Österle, H. (2011). Product data quality in supply chains: the case of Beiersdorf. *Electronic Markets*, 21(2), 141–154. <https://doi.org/10.1007/s12525-011-0059-x>.
- Ibrahim, A., Mohamed, I., Satar, N.S.M. (2021). Factors influencing master data quality: a systematic review. *International Journal of Advanced Computer Science and Applications*, 12(2). <https://doi.org/10.14569/IJACSA.2021.0120224>.

- ISO (2009). ISO 8000-110:2009. Data quality — Part 110: Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification. (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/51653.html>.
- ISO (2011). ISO/TS 8000-150:2011. Data quality — Part 150: Master data: Quality management framework (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/54579.html>.
- ISO (2016a). ISO 8000-100:2016. Data quality — Part 100: Master data: Exchange of characteristic data: Overview. (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/62392.html>.
- ISO (2016b). ISO 8000-120:2016. Data quality — Part 120: Master data: Exchange of characteristic data: Provenance. (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/62393.html>.
- ISO (2016c). ISO 8000-130:2016. Data quality — Part 130: Master data: Exchange of characteristic data: Accuracy. (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/62394.html>.
- ISO (2016d). ISO 8000-140:2016. Data quality — Part 140: Master data: Exchange of characteristic data: Completeness. (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/62395.html>.
- ISO (2018). ISO 8000-115:2018. Data quality — Part 115: Master data: Exchange of quality identifiers: Syntactic, semantic and resolution requirements. (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/70095.html>.
- ISO (2019). ISO 8000-116:2019. Data quality — Part 116: Master data: Exchange of quality identifiers: Application of ISO 8000-115 to authoritative legal entity identifiers. (Data Quality) [International Standard]. ISO. <https://www.iso.org/standard/75117.html>.
- ISO/IEC (2008). ISO/IEC 25012:2008. Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model. [International Standard]. ISO/IEC. <https://www.iso.org/standard/35736.html>.
- ISO/IEC (2011). ISO/IEC 25040:2011. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process. [International Standard]. ISO/IEC. <https://www.iso.org/standard/35765.html>.
- ISO/IEC (2015). ISO/IEC 25024:2015. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality. [International Standard]. ISO/IEC. <https://www.iso.org/standard/35749.html>.
- Li, S.-H., Chen, J.-L., Yen, D.C., Lin, Y.-H. (2013). Investigation on auditing principles and rules for PDM/PLM system implementation. *Computers in Industry*, 64(6), 741–753. <https://doi.org/10.1016/j.compind.2013.04.007>.
- Loshin, D. (2010). *Master Data Management*. Morgan Kaufmann, San Francisco, USA. 978-0-08-092121-1.
- Merino, J. (2017). *I25k: Environment for the Evaluation and Certification of the Quality of Data Products*. PhD thesis, University of Castilla-La Mancha.
- Ofner, M.H., Otto, B., Österle, H. (2012). Integrating a data quality perspective into business process management. *Business Process Management Journal*, 18(6), 1036–1067. <https://doi.org/10.1108/14637151211283401>.
- Otto, B. (2015). Quality and value of the data resource in large enterprises. *Information Systems Management*, 32(3), 234–251. <https://doi.org/10.1080/10580530.2015.1044344>.
- Otto, B., Ebner, V., Hüner, K. (2010). Measuring master data quality: findings from a case study. In: *AMCIS 2010 Proceedings*. <https://aisel.aisnet.org/amcis2010/454>.
- Piedrabuena, F., González, L., Ruggia, R. (2015). Enforcing Data Protection Regulations within e-Government Master Data Management Systems. In: *Proceedings of the 17th International Conference on Enterprise Information Systems (ICEIS 2015)*, ICEIS 2015, Vol. 3. SCITEPRESS – Science and Technology Publications, Lda, Setubal, PRT, pp. 316–321. 978-989-758-098-7. <https://doi.org/10.5220/0005458003160321>.
- Rivas, B., Merino, J., Caballero, I., Serrano, M., Piattini, M. (2017). Towards a service architecture for master data exchange based on ISO 8000 with support to process large datasets. *Computer Standards & Interfaces*, 54, 94–104. <https://doi.org/10.1016/j.csi.2016.10.004>.
- Runeson, P., Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2), 131–164. <https://doi.org/10.1007/s10664-008-9102-8>.
- Runeson, P., Host, M., Rainer, A., Regnell, B. (2012). *Case Study Research in Software Engineering: Guidelines and Examples*. John Wiley & Sons, New Jersey. 978-1-118-10435-4.
- Santos, J.V., Novo, R., Souza, J., Lopes, F., Freitas, A. (2021). Transition from ICD-9-CM to ICD-10-CM/PCS in Portugal: an heterogeneous implementation with potential data implications. *Health Information Management Journal*, 52(2), 128–131. <https://doi.org/10.1177/18333583211027241>.

- Silvola, R., Jaaskelainen, O., Kropsu-Vehkaperä, H., Haapasalo, H. (2011). Managing one master data – challenges and preconditions. *Industrial Management & Data Systems*, 111(1), 146–162. <https://doi.org/10.1108/02635571111099776>.
- Talbur, J.R. (Ed.) (2011). *Entity Resolution and Information Quality*. Elsevier, Boston, USA. 978-0-12-381973-4. <https://doi.org/10.1016/B978-0-12-381972-7.00016-6>.
- Valencia-Parra, Á., Parody, L., Varela-Vaca, Á.J., Caballero, I., Gómez-López, M.T. (2021). DMN4DQ: when data quality meets DMN. *Decision Support Systems*, 141, 113450. <https://doi.org/10.1016/j.dss.2020.113450>.
- Vilminko-Heikkinen, R., Pekkola, S. (2017). Master data management and its organizational implementation: an ethnographical study within the public sector. *Journal of Enterprise Information Management*, 30(3), 454–475. <https://doi.org/10.1108/JEIM-07-2015-0070>.
- Wand, Y., Wang, R.Y. (1996). Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*, 39(11), 86–95. <https://doi.org/10.1145/240455.240479>.

F. Gualo holds PhD and MSc degrees in computer science from the University of Castilla-La Mancha. He is currently the founder and CEO of the spin-off of the University of Castilla-La Mancha DQTeam. He is a Certified Information System Auditor (CISA) for the Information System Audit and Control Association (ISACA) since 2017 and a Software Engineering Auditor for AENOR since 2018. He has been a part of the ALARCOS Research Group since 2015. His research interests include master data management and data quality

I. Caballero received his MSc and PhD degrees in computer science from the University of Castilla-La Mancha (Spain) in 2004. He has worked as an associate professor in the Information Systems and Technologies Department at UCLM since 2001, and he was appointed as a training head of the spin-off DQTeam in 2017. Caballero holds CISA certification from ISACA in 2016 and CDO-1 certification by UALR-MIT since 2017. He is currently a member of ISO TC184/SC4, a project editor of several parts of the ISO 8000-60 series development project, and an ISO 8000-62 project editor. His research interests are focused on data quality management and data governance.

M. Rodríguez received the MSc and PhD degrees in computer science from the University of Castilla-La Mancha. He is currently the CEO of AQCLab, the first accredited laboratory for software quality assessment using ISO/IEC 25000. He is a Certified Information System Auditor (CISA) and a Chief Auditor for AENOR in Software Engineering and R&D&I. He is a professor at the Escuela Superior de Informática of the University of Castilla-La Mancha and a part of the ALARCOS Research Group. His research interests include software processes, products, and data quality.

M. Piattini received the MSc and PhD degrees in computer science from the Technical University of Madrid (UPM). He is a Certified Information System Auditor for the ISACA (Information System Audit and Control Association). He is a full professor at the Escuela Superior de Informática of the University of Castilla-La Mancha. He leads the ALARCOS research group, and his research interests are software engineering and information system quality.