Vol 28 (2)

Ingeniería y Competitividad
Vol 28 (2) 2026 doi: 10.25100/iyc.v28i2.15823

ISSN
0123-3033
e- 2027-8284

Original Article

Measuring Data Comprehensibility in Business Analytics: The B-DataCom Framework

Medición de la comprensibilidad de datos en analítica de negocio: el framework B-DataCom

Silvia Alejandra Molina España[1] Sandra Lorena Buitrón Ruiz[1] Francisco José Pino Correa[1]

[1]Facultad de Ingeniería Electrónica y Telecomunicaciones (FIET), Universidad del Cauca, Popayán, Colombia

How to cite?

Molina SA, Buitrón SL, Pino FJ. Measuring Data Comprehensibility in Business Analytics: The B-DataCom Framework. Ingeniería y Competitividad, 2026, 28(2) e-21115823

https://doi.org/10.25100/iyc.v28i2.15823

Received: 29/04/26
Reviewed: 6/07/26
Accepted: 28/07/26
Online: 26/08/26

Correspondence

silviamolina@unicauca.edu.co

Spanish version

Abstract

Objectives: This article presents a systematic mapping of the measurement of data comprehensibility in data analytics environments. The purpose of this research is to analyze how comprehensibility has been addressed in previous studies and to examine the impact this characteristic has on organizational decision-making.
Methodology: A systematic mapping of the literature was conducted; based on the results obtained, both descriptive aspects as part of its characterization and elements for measuring data quality were identified, revealing a lack of both for this characteristic.
Results: The findings of this study reinforce the existence of a research gap regarding the measurement of data comprehensibility in analytical processes. Additionally, the importance of data quality for strengthening data analytics and business decision-making is evident.
Conclusions: Primary studies in literature do not present a defined measurement for comprehensibility. Nor is there a formal conceptualization or characterization of it, which has made it difficult to establish it as a measurable attribute. Therefore, we propose the conceptual model of the B-DataCom Framework, which aims to serve as a mechanism for organizations to measure the comprehensibility of business data, to support data analysis for organizational decision-making.

Keywords: Understandability; Data Analytics; Software Development; Data Quality; Measurement.

Resumen

Objetivos: Este artículo presenta el mapeo sistemático sobre la medición de la comprensibilidad de los datos en entornos de la analítica de datos. El propósito de esta investigación es analizar cómo ha sido abordada la comprensibilidad en investigaciones previas y cuál es el impacto que genera esta característica en la toma de decisiones organizacionales.
Metodología: Se realizó un mapeo sistemático de la literatura; a partir de los resultados obtenidos, se identificaron tanto aspectos descriptivos como parte de su caracterización, como elementos de medición de calidad del dato, evidenciando una ausencia de ambas para esta característica.
Resultados: Los hallazgos de este trabajo refuerzan que existe una brecha investigativa sobre la medición de la comprensibilidad de los datos en procesos analíticos. También, se evidencia la importancia de la calidad de datos para fortalecer la analítica de datos y la toma de decisiones empresariales.
Conclusiones: Los estudios primarios de la literatura no presentan una medición definida para la comprensibilidad. Tampoco existe una conceptualización y caracterización formal de esta, lo que ha dificultado establecerla como un atributo medible. Por ende, se propone el modelo conceptual del Framework B-DataCom, que busca servir como mecanismo para que las organizaciones puedan medir la comprensibilidad de los datos de negocio, con el fin de respaldar el análisis de datos para la toma de decisiones organizacionales.

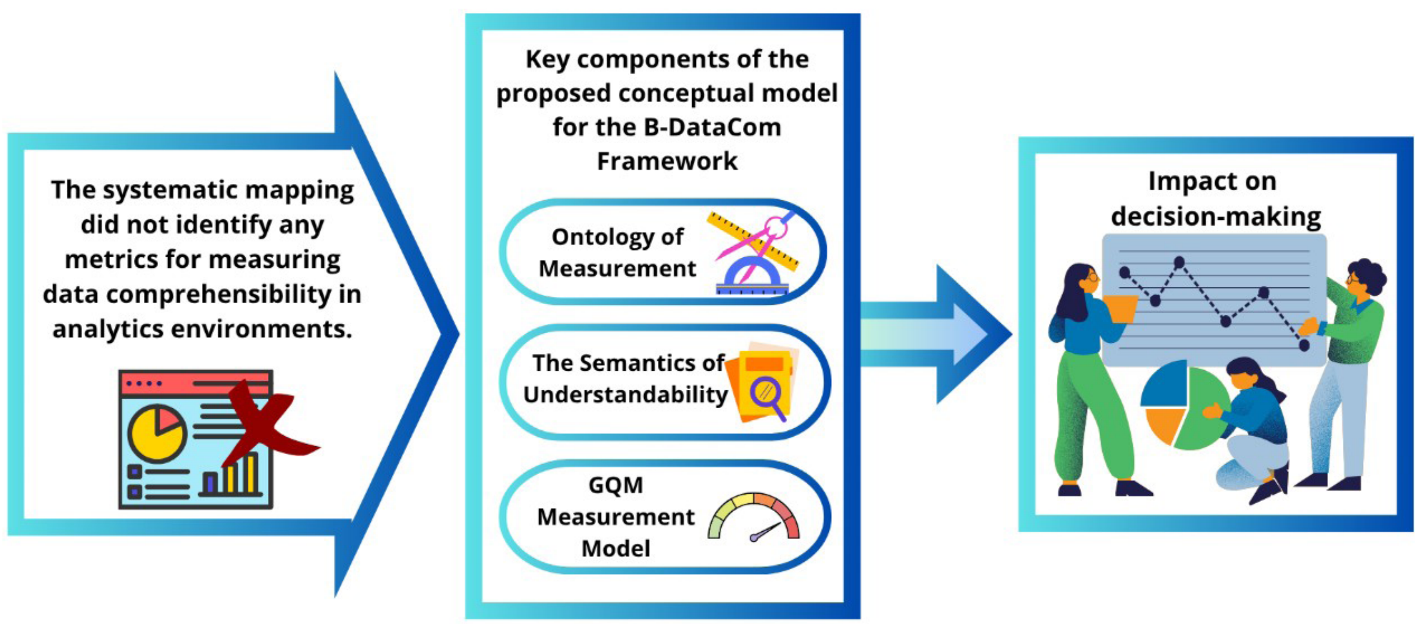
Palabras clave: Comprensibilidad; Analítica de Datos; Desarrollo de software; Calidad de los datos; Medición.

Why was this study conducted?
To address the lack of a structured approach for measuring business data understandability in software development organizations, with the aim of supporting data analytics and organizational decision-making.

What were the most significant findings?
The systematic mapping allowed for the identification of the main descriptive aspects and associated factors that impact data understandability, highlighted the absence of specific proposals for its measurement, and, based on these results, proposed the B-DataCom Framework as a first conceptual approach for its evaluation.

What do these results contribute?
The results provide a conceptual basis for measuring data understandability by integrating measurement ontology, research findings (descriptive aspects and factors impacting understandability), and the GQM model, with the purpose of strengthening future measurement strategies and supporting decision-making in software development organizations.

Graphical Abstract



doi: 10.25100/iyc.v28i2.15823

Introduction

In the software industry, organizations must ensure that the quality of business data is maintained throughout all their processes, so that data remain reliable and accurate for their intended use, thereby enabling this information to support effective business decision-making (1). The integration of approaches such as data science, technologies like Big Data, and practices such as data analytics has significantly transformed the information quality industry (2). Big Data addresses the complex challenges posed by the wide variety and rapid growth of large volumes of information, as well as the processing and analysis of multiple sources and formats, with the aim of extracting true value at different organizational levels. The exploratory data analysis (EDA) process also contributes valuable insights within the same organizational context (3). This compels organizations to advance their data analytics tools to address increasingly complex challenges arising from the abundant and dynamic information present in business processes (4).

From the perspective of data quality, various measurable characteristics can be considered, including those defined by the ISO/IEC 25012 standard (5). However, although information exists regarding the aforementioned data quality model, there remains a gap in this context concerning how understandable the data are in supporting and strengthening competitive strategies for organizational decision-making (6,7).

One of the challenges faced by software development companies regarding data understandability is the rapid growth in data volume, along with the multiplicity of formats and sources such as Excel, TXT, PDF, or CSV files. This is compounded by the high likelihood of containing inconsistent, incomplete, or erroneous data (8). These issues further highlight the need for comprehensible, reliable, and useful data that can be integrated into tools supporting the assessment of data quality, thus benefiting business decision-making (9).

This article presents progress on the research project entitled “Framework for Measuring the Understandability of Business Data in Software Development Organizations,” within which a systematic mapping study on data understandability has been conducted. The aim is to examine the role of this characteristic within the business context of software development and to analyze how data understandability has been addressed in previous research related to data analytics processes. Additionally, this research progress includes the definition of an initial conceptual model for the B-DataCom Framework, whose purpose is to measure the degree of business data understandability in the specified context, thus serving as a mechanism for the evaluation and improvement of business data quality in support of enhanced organizational decision-making. Accordingly, the document is organized as follows: Section Two presents the methodology used for the systematic mapping study; Section Three describes the results obtained from the mapping. Finally, the initial version of the B-DataCom Framework is introduced, along with its limitations and future work. Section Four presents the preliminary conclusions of this research.

Methodology

To develop the systematic mapping study on data understandability in the context of software development companies, and considering its impact on organizational decision-making, the protocol proposed by (10) was adopted. This protocol is structured into the following four stages:

doi: 10.25100/iyc.v28i2.15823

research question, keywords, information search strategy, and quality assessment of primary studies. Each of these stages is detailed below.

Research Question

¿ What proposals exist regarding data understandability that support business decision-making in software development companies?

Clave Keywords

To construct the search string, the following keywords were used: Data Analytics, Software Development, Understandability, Data Quality, Decision-Making. Based on these, the search string was defined as follows:

((Data Analytics) OR (Understandability)) AND (Software Development)

Information Search Strategy

For the information search process, the Scopus database was used, as it is a platform with broad coverage of peer-reviewed scientific literature and indexes the leading publication sources in computer science and engineering. This enabled the retrieval of a representative overview of the state of the art in the scientific literature relevant to the subject under study. Only research articles in the areas of computer science and engineering, open access, and published between 2016 and 2025 were considered. Upon executing the information search strategy, a total of 263 articles were found. The screening process for these studies was conducted by the two principal investigators of the project, who applied the established inclusion and exclusion criteria and discussed any discrepancies in article selection until consensus was reached. The selection process for the studies was carried out in accordance with the systematic mapping methodology adopted for this research. Subsequently, the results were filtered according to the following inclusion and exclusion criteria:

Inclusion: Articles whose title, abstract, and conclusions reference business data analytics or data quality in terms related to the keywords, defined manually based on the focus of the study: data analysis, Big Data, software, predictive analytics, data mining, decision-making, visual analytics, metadata, data science, data analytics, data quality, data management, data acquisition, data-based. After applying these criteria, a total of 20 relevant documents were identified.

Exclusión: Articles that do not explicitly address, within their domain of study, factors and impacts related to data understandability in the context of software development. Based on these exclusions, a final set of 12 primary articles was obtained.

The process of identification, screening, eligibility assessment, and inclusion of studies is presented in Figure 1.

doi: 10.25100/iyc.v28i2.15823

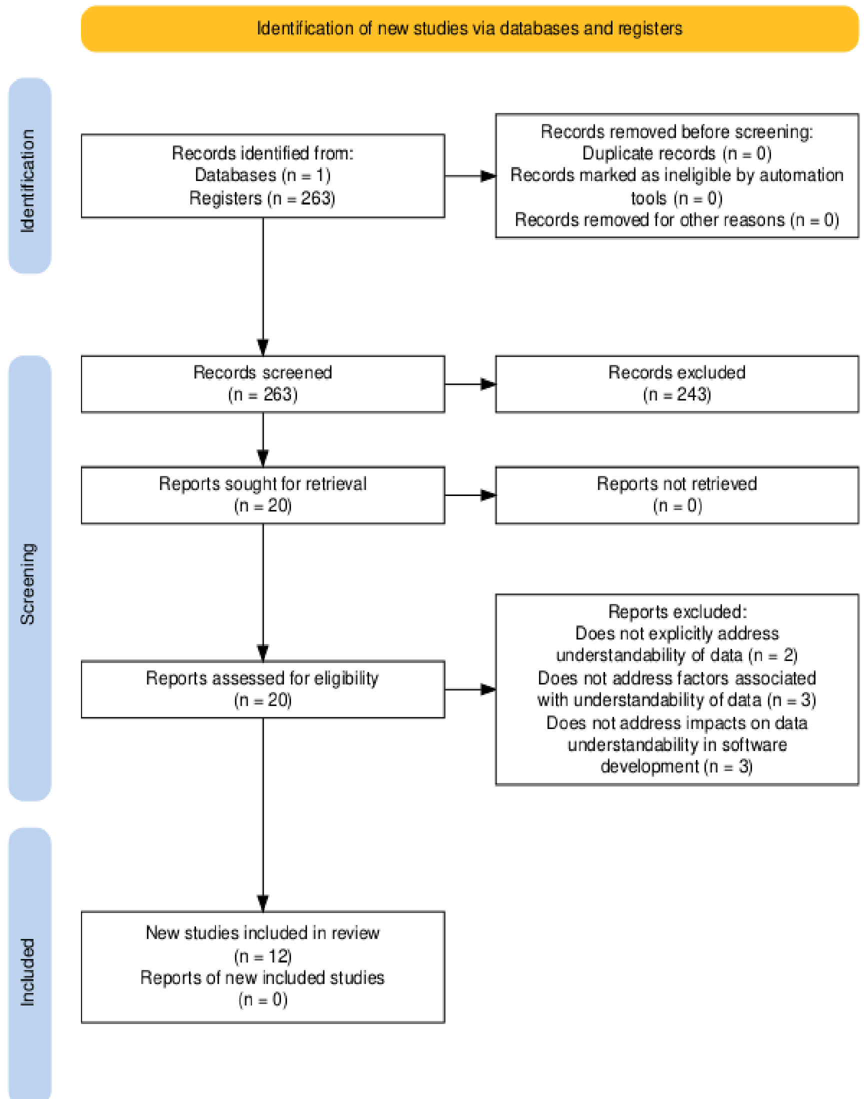


Figure 1. Flow diagram of the study selection process (adapted from PRISMA 2020 (11)). Source: Own elaboration using the PRISMA2020 Shiny App tool (12)

Quality Assessment of Primary Studies

To determine the rigor and relevance of the primary studies (10), a quality assessment was conducted on the 12 proposals. For this purpose, the following three probing questions were applied: **(i)** Do they include data quality measurement techniques? **(ii)** What is the importance of data quality for business analytics according to the authors? **(iii)** Do they contain aspects related to understandability that allow its description?

For each of the questions posed, the degree of quality contribution (GAC) to the research work was assessed. The evaluation was conducted independently by the two principal investigators of the project, who assigned the corresponding ratings to each study using a Likert scale (13), which classifies the level of contribution as follows: 1 for Low contribution, 2 for Medium contribution, and 3 for High contribution. The result of this quality evaluation process is presented in Table 1, in terms of the mode of the ratings.

doi: 10.25100/iyc.v28i2.15823

Table 1. Mode per proposal in the GCA evaluation

Ref.	Likert scale rating per question			Mode
	(i)	**(ii)**	**(iii)**	
(4)	2	2	1	2
(9)	2	3	2	2 y 3
(14)	2	3	1	3
(15)	2	3	1	3
(16)	2	3	2	2 y 3
(17)	2	3	1	3
(18)	2	2	1	2
(6)	2	2	2	2
(19)	2	2	2	2
(20)	2	3	2	2 y 3
(21)	2	3	2	2 y 3
(22)	2	3	2	2 y 3

As shown in Table 1, the most frequent ratings are found on the high and medium scales. Bimodal distributions were observed, indicating a balanced perception with these same scales. This lends greater reliability to the research by demonstrating consistency in the assessment of the contribution of the studies. In cases where the ratings exhibited a bimodal distribution, both values were retained in the table, respecting the nature of the mode obtained and ensuring transparency in the presentation of the results.

From the analysis, the following answers to the questions posed were obtained:

(i) What aspects are measured, with respect to data quality, in the proposals? It was identified that 100% of the articles (12) address data quality aspects. Studies (4) and (6) are based on subjective measurements that assess perception and ease of data comprehension, such as the Likert scale. Additionally, studies (9), (15), (20), (21), and (22) are based on direct quantitative measurements such as the number of errors, time, and accuracy, with the aim of evaluating data efficiency, thereby indirectly affecting its understandability. On the other hand, studies (14), (17), (16), and (18) employ theoretical models that structure the assessment through attributes such as accessibility and interoperability. Finally, study (19) involves a technical measurement focused on data processing and standardization, aiming to reduce ambiguity in understandability. This demonstrates that while the studies address certain aspects of data quality, they do not incorporate explicit techniques for its measurement.

(ii) What is the importance of data quality for business analytics according to the authors? It was found that 100% of the articles (12) highlight the role and importance of data in analytics within the business context of software development. Regarding data quality, it was identified that, of these 12 articles, 8 (66.7%) provide a high contribution as they describe contributions toward data quality improvement in a concise and direct manner, which aids data understandability and analytics processes for decision-making in the business context of software development. The remaining four articles (33.3%) provide a medium contribution, as their contributions to data quality are indirect, given that their primary focus does not specifically address this aspect.

doi: 10.25100/iyc.v28i2.15823

(iii) Do they contain aspects associated with understandability that allow its description? Among the 12 articles (100%), seven (58.3%) provide a medium contribution, as they include aspects such as ease, clarity, and intelligibility that can be associated with understandability, though not explicitly addressed as a direct attribute in their study. The remaining five articles (41.7%) provide a low contribution, as these aspects associated with understandability are addressed minimally and indirectly in the studies. This indicates the need to identify aspects that allow for the description of data understandability in the context of software development.

Based on these results, it was decided to continue with all 12 research works in the Systematic Mapping process.

Results and discussion

To organize the information found on data understandability in the context of software development organizations, a classification scheme was established, defining three perspectives of analysis: (i) relevant studies, (ii) the perspective in the domain of methodology, and (iii) the perspective in the domain of data understandability. Within each perspective, topics for analysis were identified. Each of these results is described below.

Relevant Studies

The following is an analysis of the relevant research identified during the systematic mapping process, grouped into three thematic cores: (i) data understandability and quality; (ii) use and integration of metadata; (iii) data analytics, Big Data, and decision-making.

Data Understandability and Quality

Study (9) presents the views of professionals and businesspeople regarding how understandability, usefulness, and reliability support quality models for software analysis tools. Similarly, (23) seeks to deepen data comprehension to generate greater value in services and products, contributing a framework through its research. Although both studies address understandability indirectly, it is not presented as a directly measurable attribute through specific metrics.

Use and Integration of Metadata

Study (15) demonstrates how metadata assist in the analysis and understanding of public proteomics data files through the MAGE-TAB format. Likewise, (17) discusses established metadata that support FAIR computational workflows by guiding and recording their provenance, while (24) explains how Git metadata contribute to the understanding of software projects through interactive exploration. For its part, (25) focuses on metadata as an essential element in the search for services and repositories, and (26) highlights its importance in reproducible computing (RCS). Finally, (27) uses metadata to design data analysis and requirements improvement platforms, whereby these are integrated into the DASE framework infrastructure. Although these studies recognize the role of metadata in enhancing data understandability, they propose no specific metrics to

doi: 10.25100/iyc.v28i2.15823

evaluate this characteristic in support of analytical processes within the business context of software development.

Data Analytics, Big Data, and Decision-Making

Studies (3), (4), and (6) seek to answer how Big Data analytics impacts and generates business value in SMEs, focusing on analytical tools and strategic QaSD dashboards to improve high-level decision-making. Similarly, studies (14), (16), and (18) highlight how Q-Rapids frameworks, methodologies, and digitalization processes in SMEs guide and strengthen organizations to enhance analytical processes, data quality, and business decision-making. In turn, (19), (28), and (29) analyze the best way to invest in business analytics strategies in software development projects by integrating Building Information Modeling (BIM) according to organizational capabilities, aiming to achieve competitive advantages and enable data-driven decision-making. Likewise, (30), (31), and (32) propose the integration of digital twins and frameworks to optimize data management and analysis in business contexts. Nevertheless, although these studies establish the importance of data analytics, Big Data, and decision-making in software development, they do not directly consider understandability as a potentially relevant factor within these processes.

Based on the analysis of relevant studies, there is a significant concentration on data analytics, Big Data, and decision-making, highlighting their use and importance in the business context of software development from the perspective of analytical infrastructures, predictive models, and technological frameworks, while overlooking the semantic quality of data. Additionally, metadata are shown to aid understanding, interpretation, contextualization, and information analysis. Similarly, there is a research gap regarding the integration and assessment of data understandability and quality in software development contexts, justifying the need for a framework to support and foster business decision-making.

Perspective in the Methodological Domain

From the methodological domain perspective, the primary studies identify the dimensions: types of contribution and types of validation. Each is described below:

Dimension of Type of Contribution: The types of contribution found in the primary studies are described in Table 2.

doi: 10.25100/iyc.v28i2.15823

Table 2. Types of Contribution (based on (33))

Contribution	Definition
Study (ES)	Research works that, from a specific perspective, summarize, analyze, and generate conclusions that contribute from the studied context to other research efforts. Includes: review, classification, and compendium.
Approach (EN)	Research works that conceptually describe a specific topic on a process to solve problems and questions. Includes: conceptual approach.
Methodological Proposal (PM)	Research works that, to achieve a specific function or objective, include stages, procedures, or steps. Includes: method, technique, methodology.
Tool (H)	Research works that create tools to support, understand, assist, and perform a process. Includes: model, framework, and Framework.

Dimension of Type of Validation Performed: The types of validation found in the primary studies, understood as types that define the reliability and accuracy of published research, are described in Table 3.

Table 3. Types of Validation (based on (34))

Validation	Definition
Theoretical (T)	Documents that are relevant and enhance practice. They are usually methodological and may present algorithms. Includes reviews and literature studies.
Surveys (E)	Documents that present responses to questions posed to an audience knowledgeable about the field in general. They provide the results' insights to future researchers.
Case Studies (EC)	Documents that provide a detailed description of how and why the practice was applied. These are analytical and descriptive with an objective perspective.

doi: 10.25100/iyc.v28i2.15823

Experiments (EX)	Documents that present validation through an empirical study. The researcher has control over the experiment, including the target population, duration, task types, and areas.

Figure 2 presents the bubble chart illustrating the intersection between the TC and TV dimensions of the studies, thus showing which types are prioritized and guide the research.

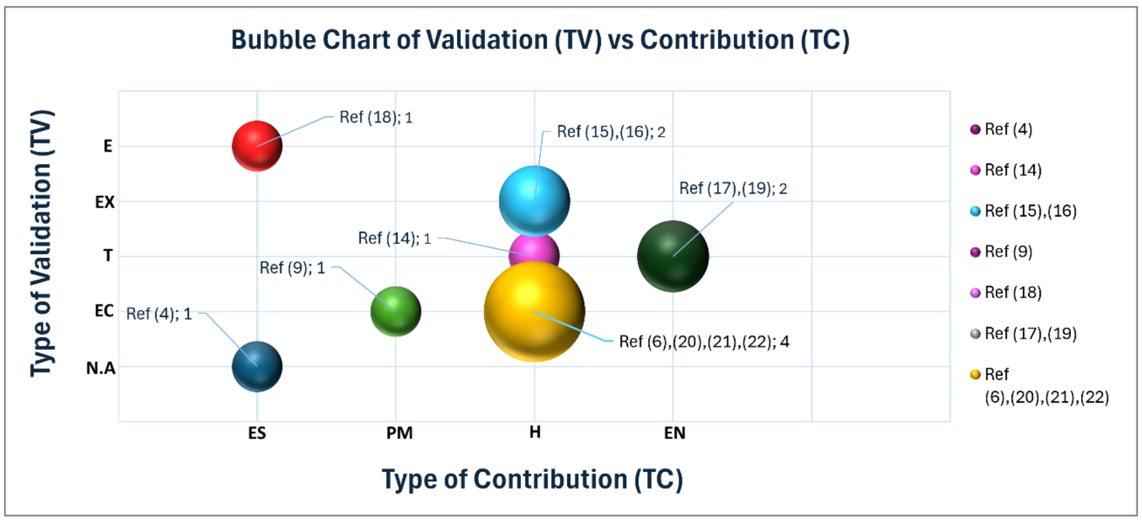


Figure 2. Bubble Chart of Validation (TV) vs Contribution (TC). Source: Own elaboration.

Based on the analysis from the methodological domain perspective, it was identified that 33.33% (4/12) of the proposals used Case Study (EC) as the validation type and Tool (H) as the contribution type; 16.67% (2/12) correspond to studies where the validation type is Experiment (EX) with the same contribution type; 16.67% (2/12) are studies where the validation type is Theoretical (T) and the contribution type is Approach (EN); 8.33% (1/12) is a study with Theoretical (T) validation and Tool (H) contribution; 8.33% (1/12) is a study with Case Study (EC) validation and Methodological Proposal (PM) contribution; 8.33% (1/12) is a study with Survey (E) validation and Study (ES) contribution; and 8.33% (1/12) presents no validation type (N.A) and Study (ES) as the contribution type.

These statistics reflect that research on data understandability in the context of software development is consistent with the nature of the topic, as researchers prioritize applicable solutions rather than only hypotheses untested in real-world contexts. This type of validation enables improvements through empirical research in different contexts and under various tool uses. Likewise, there is also evidence of a low tendency for research with contributions in Studies (1/12) and Methodological Proposals (1/12), as well as for validations oriented to Surveys (1/12). This gap can be strengthened in future research, as it demonstrates persistent deficiencies in theory regarding data quality and understandability in these same contexts.

doi: 10.25100/iyc.v28i2.15823

Perspective in the Domain of Data Understandability

From this perspective, the studies identify topics that allow for understanding the current state of data understandability in the context of software development. Each of these is listed below:

Data Quality Measurement Elements (EMC): Techniques, metrics, and methods for measuring data quality in the software development context.

Descriptive Aspects of Data Understandability (ADM): Descriptive aspects of the term "understandability" found in the literature. The identification of these aspects was carried out through manual coding of the primary studies, extracting the terms used by the authors to describe or associate data understandability, which were subsequently classified as descriptive aspects.

Table 4 below presents a summary of the 12 primary studies, together with the classification of the previously described perspectives, in terms of Data Quality Measurement Elements (EMC) (techniques, metrics, and methods) and Descriptive Aspects of Data Understandability (ADM). An analysis of this perspective is presented subsequently.

Table 4. Summary of results from the primary studies and the classification scheme of the information.

Ref.	EMC	ADM
(4)	Escala Likert Scale (1 to 5 points) (ordinal scale for quantifying subjective responses)	Easy
(9)	Number of fully specified problems (a direct data point for subsequent measurement)	Easy, Clear
(14)	Q-Rapids Quality Model (a model)	Easy, Evident
(15)	KPI (Key Performance Indicator): Workflow Efficiency, Compliance Rate, Processing Speed (TB/hour), Anomaly Detection Accuracy.	Easy
(16)	MAGE-TAB Proteomics (Standard data annotation format)	Intelligible
(17)	Q-Rapids Quality Model	Easy
(18)	KPI (Key Performance Indicator): FAIR Indicators (Findable, Accessible, Interoperable, Reusable)	Easy, Evident
(6)	Likert Scale (1 to 5 points) Attitudes, opinions, and user perceptions. Likert technique	Easy
(19)	Reliable Matching of Multi-Cultural Names (approach), Standardization (technique), Cleansing (technique), Data Profiling Tools (tool), Matching (technique), Survivorship Techniques (technique).	Easy
(20)	Quality dimension: Accuracy, Completeness, Consistency, Timeliness, and Asymmetry.	Easy

doi: 10.25100/iyc.v28i2.15823

(21)	Quality dimension: Accuracy, Completeness, Conformity, Uniqueness, Consistency, and Legibility.	Easy
(22)	Metrics: Single Distribution Metric and Two Distribution Metric.	Easy

Within the process of analyzing the information obtained through the topics, it was identified that 91.67% (11/12) associate and characterize aspects of understandability with the concept of “Easy,” 8.33% (1/12) with the concept of “Intelligible,” as well as 8.33% (1/12) with the concept of “Clear,” and 16.67% (2/12) with the concept of “Evident.” The word “Easy” is the most recurrent among the primary studies. According to the literature, the authors implicitly seek that information in the field of data quality can be understood without ambiguity or difficulty by the user. However, this superficial interpretation may hinder an adequate and objective conceptualization of data understandability according to the needs of software development companies. Furthermore, there is evidence that there is no formal definition of what data understandability is, nor are there descriptive aspects directly related to it. Therefore, the need to characterize and define understandability more rigorously in the organizational context of software development companies is demonstrated.

Regarding the data quality measurement elements shown in Table 4, it is identified that 66.67% (8/12) represent Key Performance Indicators (KPIs) in techniques such as: number of fully specified problems, Q-Rapids quality model, accuracy, completeness, consistency, timeliness, asymmetry, accuracy, completeness, conformity, uniqueness, consistency, legibility, reliable matching, data profiling, and matching techniques. This indicates that, although quantitative measurements oriented toward data performance or quality are carried out in practice, they still lack formalization, traceability, and uniformity.

Conceptual Model of the B-DataCom Framework

Regarding the definition of an initial version of the B-DataCom Framework (Business Data Comprehensibility Framework), this research adopts a perspective that enables a structured measurement of business data understandability, grounded in theoretical concepts such as Measurement Ontology (35) and centered on the semantics of understandability, supported by the GQM measurement model (36). These foundations delimit the construction scope of the framework, as illustrated in Figure 3.

doi: 10.25100/iyc.v28i2.15823

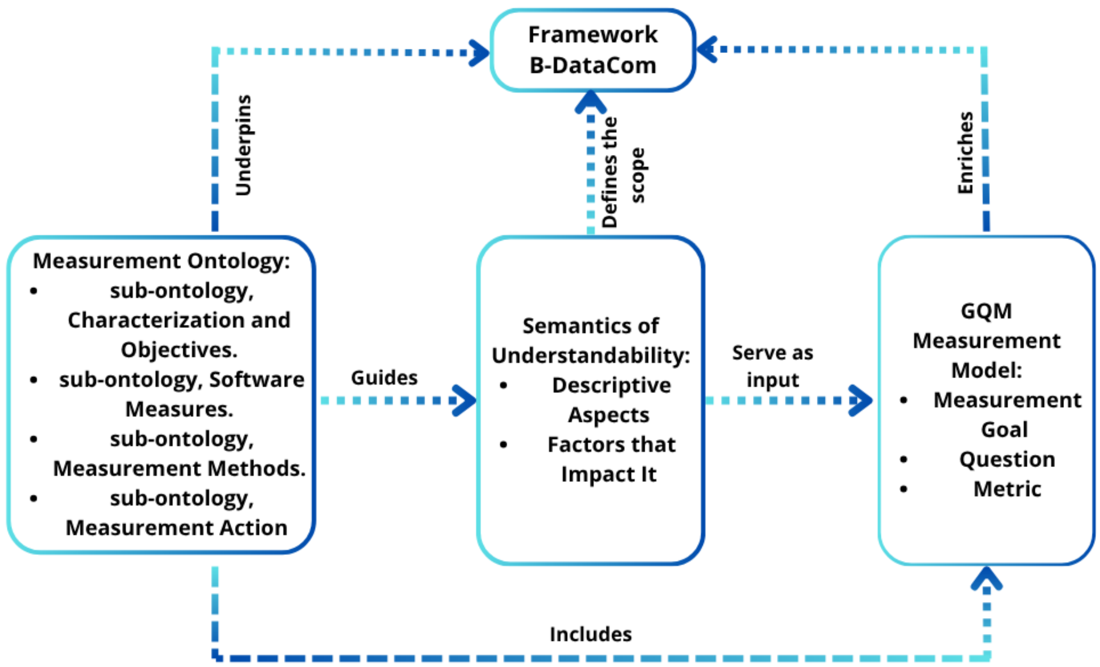


Figura 3. Conceptual Model of the B-DataCom Framework. Source: Own elaboration.

One of the pillars of the B-DataCom Framework is the Measurement Ontology, which comprises four sub-ontologies: characterization and objectives, software measures, measurement methods, and measurement actions. Together, these facilitate the interpretation and representation of the measurement context. This pillar underpins the framework by providing ontological elements that address the need to measure data understandability in the context of software development organizations. It also guides the semantic component of understandability and incorporates the GQM Measurement Model theory.

On the other hand, the semantics of understandability, as the second conceptual pillar of the Framework, establishes the scope of the meaning of understandability within the Framework and serves as input for the GQM measurement model. The objective of this component is to define measurable attributes and factors that impact understandability in the organizational context.

Finally, the Framework is enriched by elements from the GQM Measurement Model (Goal-Question-Metric) (36), enabling it to obtain indicators in terms of the degree of data understandability. This includes a measurement process comprised of Measurement Goals, Measurement Questions, and Metrics.

The objective of the B-DataCom Framework is to evaluate understandability in the context of a software development organization. It establishes questions to identify direct and indirect data related to understandability within the context, and finally, metrics that answer the established measurement questions, generating indicators for data understandability. The three pillars described above are integrated and interrelated, resulting in the conceptual model of the B-DataCom Framework that will be developed toward its implementation. The aim is to support business data analysis processes through structured measurement.

doi: 10.25100/iyc.v28i2.15823

Limitations and Future Work**:** The B-DataCom Framework presented in this article corresponds to an initial conceptual version. It has not yet been implemented or empirically validated in software development organizations. Therefore, the results are limited to the theoretical foundation obtained from the systematic mapping and to the definition and construction of the conceptual components constituting the B-DataCom Framework. As a result of these limitations, the next stage of this research will advance its implementation by defining GQM metrics associated with its components as measurement mechanisms. Subsequently, it will be validated through a case study in a software development company, thus evaluating, in a real context, the applicability, usefulness, and capacity of the B-DataCom Framework to measure business data understandability and support organizational decision-making.

Conclusion

The systematic mapping identified 12 primary studies that, although they use data quality metrics for aspects slightly related to understandability, such as accuracy, completeness, and consistency, do not provide a structured measurement based on explicit measurement models for the understandability characteristic, considering its semantics. Likewise, 91.67% of the primary studies relate understandability solely to the descriptive aspect “Easy,” which evidences a limited and superficial conceptualization of this characteristic. Similarly, there is no formal conceptualization and characterization of it, which has further limited its operationalization in organizational processes, tending to be regarded as something intrinsic to quality without achieving a deep understanding of its implications and impacts on data analytics. This gap hinders the integration and establishment of understandability as a measurable attribute that supports data analytics for decision-making.

The results of the mapping highlight the relevance of data quality to strengthen analytics and business decision-making in the context of software development, as reflected in the predominant use of metrics, frameworks, models, and techniques aimed at improving data quality and management. However, it has not yet been addressed how understandability influences and impacts these processes. Based on the above, this research, through the B-DataCom Framework, seeks to contribute to narrowing this gap by conceptualizing and measuring understandability as an essential element for the effective use of business data in organizational environments, with the purpose of supporting business data analysis processes based on the understanding of various types of data present in the software development context, such as supplier, cost, product development, client, project, and training data, among others.

From a theoretical perspective, this research broadens the application of Measurement Ontology to the domain of data understandability, which includes the GQM measurement model, guides the semantics of understandability, and underpins the B-DataCom Framework. Thus, data are recognized as a fundamental and strategic asset within business data analytics, whose usefulness depends on both their quality and their degree of understandability.

At the business and societal level, the B-DataCom Framework aims to support the diagnosis of the degree of business data understandability, understood as a measurable characteristic belonging to the concept of data quality, in order to strengthen improvement actions for analytics and business decision-making. At the academic level, it provides a conceptual and methodological basis in this

doi: 10.25100/iyc.v28i2.15823

line of research, using foundations from data quality models that remain valid and essential for creating new knowledge in technologies and approaches related to Big Data and data science.

In this regard, the research demonstrates the importance of articulating the foundations established in the Framework with the dynamics of organizations, since the continuous growth of data volume and diversity requires mechanisms that enhance their understanding and use. This allows organizations to integrate data quality concepts into their business analytics practices, thereby strengthening business knowledge.

As a future research direction, it is proposed to extend the B-DataCom Framework by incorporating a new component for the automatic capture and processing of business data, supported by artificial intelligence techniques, so that the understandability indicators generated by the Framework are obtained in a timely and continuous manner for organizational management areas involved in decision-making. Additionally, it is proposed to validate the applicability and adaptability of the Framework in different organizational contexts and to expand its scope to scenarios related to Big Data and data science.

Funding

This research is funded by COLFUTURO for master's and doctoral programs through the call "Training of high-level human capital to promote social and community leadership and address the Fourth Industrial Revolution in the department of Cauca (4RI)" in 2024.

Declaration on the Use of Artificial Intelligence

Grammarly was used exclusively for the translation and correction of the English language of the manuscript. The authors reviewed and verified all modifications made and assume full responsibility for the scientific content of the article.

Statement of authorship contribution credits

Conceptualization – Ideas: Silvia Molina, Sandra Buitrón, Francisco Pino; Data curation: Silvia Molina, Sandra Buitrón, Francisco Pino; Formal analysis: Silvia Molina, Sandra Buitrón; Funding acquisition: Silvia Molina, Sandra Buitrón, Francisco Pino; Investigation: Silvia Molina, Sandra Buitrón; Methodology: Silvia Molina, Sandra Buitrón, Francisco Pino; Project administration: Silvia Molina, Sandra Buitrón, Francisco Pino; Supervision: Silvia Molina, Sandra Buitrón, Francisco Pino; Validation: Silvia Molina, Sandra Buitrón, Francisco Pino; Visualization – Preparation: Silvia Molina, Sandra Buitrón; Writing – original draft – Preparation: Silvia Molina, Sandra Buitrón; Writing – review and editing – Preparation: Silvia Molina, Sandra Buitrón, Francisco Pino; Ethical implications: none declared; Conflicts of interest: none declared.

doi: 10.25100/iyc.v28i2.15823

References

1. Merino J, Caballero I, Rivas B, Serrano M, Piattini M. A Data Quality in Use model for Big Data. Future Gener Comput Syst. el 1 de octubre de 2016;Modeling and Management for Big Data Analytics and Visualization63:123-30.

https://doi.org/10.1016/j.future.2015.11.024

2. Badshah A, Daud A, Alharbey R, Banjar A, Bukhari A, Alshemaimri B. Big data applications: overview, challenges and future. Artif Intell Rev. el 16 de septiembre de 2024;57(11):290.

https://doi.org/10.1007/s10462-024-10938-5

3. Yeh YT, Eden R, Fielt E, Syed R. The role of use for the business value of big data analytics. J Strateg Inf Syst. el 1 de junio de 2025;34(2):101888.

https://doi.org/10.1016/j.jsis.2025.101888

4. Willetts M, Atkins AS. Evaluation of a software positioning tool to support SMEs in adoption of big data analytics. J Electron Sci Technol. el 1 de marzo de 2024;22(1):100229.

https://doi.org/10.1016/j.jnlest.2023.100229

5. ISO/IEC 25012:2008. Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model. Geneva; 2008.

https://www.iso.org/standard/35736.html

6. López L, Manzano M, Gómez C, Oriol M, Farré C, Franch X, et al. QaSD: A Quality-aware Strategic Dashboard for supporting decision makers in Agile Software Development. Sci Comput Program. el 1 de febrero de 2021;202:102568.

https://doi.org/10.1016/j.scico.2020.102568

7. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. J Manag Inf Syst. marzo de 1996;12(4):5-33.

https://doi.org/10.1080/07421222.1996.11518099

8. Putrama IM, Martinek P. Heterogeneous data integration: Challenges and opportunities. Data Brief. el 1 de octubre de 2024;56:110853.

https://doi.org/10.1016/j.dib.2024.110853

9. Martínez-Fernández S, Vollmer AM, Jedlitschka A, Franch X, López L, Ram P, et al. Continuously Assessing and Improving Software Quality With Software Analytics Tools: A Case Study. IEEE Access. 2019;7:68219-39.

https://doi.org/10.1109/ACCESS.2019.2917403

10. Petersen K, Feldt R, Mujtaba S, Mattsson M. Systematic Mapping Studies in Software Engineering. Proc 12th Int Conf Eval Assess Softw Eng. el 26 de junio de 2008;17.

doi: 10.25100/iyc.v28i2.15823

https://doi.org/10.14236/ewic/EASE2008.8

11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. el 29 de marzo de 2021;n71

https://doi.org/10.1136/bmj.n71

12. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Syst Rev. el 1 de junio de 2022;18(2):e1230.

https://doi.org/10.1002/cl2.1230

13. Likert R. A technique for the measurement of attitudes. Arch Psychol. 1932;22:1-55.

https://legacy.voteview.com/pdf/Likert_1932.pdf

14. Fannouch A, Gharib J, Gahi Y. Enhancing DataOps practices through innovative collaborative models: A systematic review. Int J Inf Manag Data Insights. el 1 de junio de 2025;5(1):100321.

https://doi.org/10.1016/j.jjimei.2025.100321

15. Dai C, Füllgrabe A, Pfeuffer J, Solovyeva EM, Deng J, Moreno P, et al. A proteomics sample metadata representation for multiomics integration and big data analysis. Nat Commun. el 6 de octubre de 2021;12(1):5854.

https://doi.org/10.1038/s41467-021-26111-3

16. Kozik R, Chora□ M, Puchalski D, Renk R. Q-Rapids framework for advanced data analysis to improve rapid software development. J Ambient Intell Humaniz Comput. el 1 de mayo de 2019;10(5):1927-36.

https://doi.org/10.1007/s12652-018-0784-5

17. Goble C, Cohen-Boulakia S, Soiland-Reyes S, Garijo D, Gil Y, Crusoe MR, et al. FAIR Computational Workflows. Data Intell. el 13 de agosto de 2024;2(1-2):108-21.

https://doi.org/10.1162/dint_a_00033

18. Kusumawardhani FK, Ratmono D, Wibowo ST, Darsono D, Widyatmoko S, Rokhman N. The impact of digitalization in accounting systems on information quality, cost reduction and decision making: Evidence from SMEs. Int J Data Netw Sci. el 1 de marzo de 2024;8(2):1111-6.

https://doi.org/10.5267/j.ijdns.2023.11.023

19. Liberatore MJ, Pollack-Johnson B, Clain SH. Analytics Capabilities and the Decision to Invest in Analytics. J Comput Inf Syst. el 2 de octubre de 2017;57(4):364-73.

https://doi.org/10.1080/08874417.2016.1232995

20. Bayram F, Ahmed BS, Hallin E, Engman A. DQSOps: Data Quality Scoring Operations Framework for Data-Driven Applications. En: Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering [Internet]. New York, NY, USA: Association for Computing

Machinery; 2023 [citado el 2 de octubre de 2025]. p. 32-41. (EASE '23).

https://doi.org/10.1145/3593434.3593445

21. Elouataoui W, El Mendili S, Gahi Y. An Automated Big Data Quality Anomaly Correction Framework Using Predictive Analysis. Data. diciembre de 2023;8(12):182.

https://doi.org/10.3390/data8120182

22. Tu D, He Y, Cui W, Ge S, Zhang H, Han S, et al. Auto-Validate by-History: Auto-Program Data Quality Constraints to Validate Recurring Data Pipelines. En: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2023 [citado el 2 de octubre de 2025]. p. 4991-5003. (KDD '23).

https://doi.org/10.1145/3580305.3599776

23. Lee B, Ahmed-Kristensen S. D3 framework: An evidence-based data-driven design framework for new product service development. Comput Ind. el 1 de enero de 2025;164:104206.

https://doi.org/10.1016/j.compind.2024.104206

24. Kim Y, Kim J, Jeon H, Kim YH, Song H, Kim B, et al. Githru: Visual Analytics for Understanding Software Development History Through Git Metadata Analysis. IEEE Trans Vis Comput Graph. febrero de 2021;27(2):656-66

https://doi.org/10.1109/TVCG.2020.3030414

25. □ezník T, Raes L, Stott A, De Lathouwer B, Perego A, Charvát K, et al. Improving the documentation and findability of data services and repositories: A review of (meta)data management approaches. Comput Geosci. el 1 de diciembre de 2022;169:105194.

https://doi.org/10.1016/j.cageo.2022.105194

26. Leipzig J, Nüst D, Hoyt CT, Ram K, Greenberg J. The role of metadata in reproducible computational research. Patterns. el 10 de septiembre de 2021;2(9).

https://doi.org/10.1016/j.patter.2021.100322

27. Bandara M, Rabhi FA, Bano M. A knowledge-driven approach for designing data analytics platforms. Requir Eng. el 1 de junio de 2023;28(2):195-212.

https://doi.org/10.1007/s00766-022-00385-5

28. Ahmad MO, Ahmad I, Rana NP, Khan IS. An Empirical Investigation on Business Analytics in Software and Systems Development Projects. Inf Syst Front. el 1 de abril de 2023;25(2):917-27.

https://doi.org/10.1007/s10796-022-10253-w

29. Sompolgrunk A, Banihashemi S, Golzad H, Le Nguyen K. Strategic alignment of BIM and big data through systematic analysis and model development. Autom Constr. el 1 de diciembre de 2024;168:105801

https://doi.org/10.1016/j.autcon.2024.105801

doi: 10.25100/iyc.v28i2.15823

30. Ryzhakova G, Malykhina O, Pokolenko V, Rubtsova O, Homenko O, Nesterenko I, et al. Construction Project Management with Digital Twin Information System. Int J Emerg Technol Adv Eng. el 1 de octubre de 2022;12(10):19-28.

https://doi.org/10.46338/ijetae1022_03

31. Rique T, Perkusich M, Dantas E, Albuquerque D, Gorgônio K, Almeida H, et al. On Adopting Software Analytics for Managerial Decision-Making: A Practitioner's Perspective. IEEE Access. 2023;11:73145-63.

https://doi.org/10.1109/ACCESS.2023.3294823

32. Zhu J, Zhuang E, Fu J, Baranowski J, Ford A, Shen J. A Framework-Based Approach to Utility Big Data Analytics. c. mayo de 2016;31(3):2455-62.

https://doi.org/10.1109/TPWRS.2015.2462775

33. Buitrón Ruiz SL. Marco de trabajo para la elicitación de requisitos no funcionales basada en la gestión del conocimiento Merlinn [Internet]. 2016 [citado el 10 de octubre de 2025].

http://repositorio.unicauca.edu.co:8080/xmlui/handle/123456789/1291

34. Montesi M, Lago P. Software engineering article types: An analysis of the literature. Journal of Systems and Software, 81(10), 1694-1714. J Syst Softw. el 1 de octubre de 2008;81:1694-714

https://doi.org/10.1016/j.jss.2007.11.723

35. Ferreira M, Garcia F, Ruiz F, Bertoa M, Calero C, Vallecillo A, et al. Medición del Software Ontología y Metamodelo [Informe Técnico]. noviembre de 2006. Informe Técnico UCLM-TSI-001. https://n9.cl/izkla

36. Basili V, Trendowicz A, Kowalczyk M, Heidrich J, Seaman C, Münch J, et al. Aligning Organizations Through Measurement: The GQM+Strategies Approach. Cham, Suiza: Springer; 2014.

https://doi.org/10.1007/978-3-319-05047-8

doi: 10.25100/iyc.v28i2.15823