

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

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

Silvia Alejandra Molina España¹   Sandra Lorena Buitrón Ruiz¹  Francisco José Pino Correa¹ 

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

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.

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.

¿Cómo citar?

Molina SA, Buitrón SL, Pino FJ. Medición de la comprensibilidad de datos en analítica de negocio: el framework B-DataCom. Ingeniería y Competitividad, 2026, 28(2)e-21215823

<https://doi.org/10.25100/iyv.v28i2.15823>

Recibido: 29/04/26

Revisado: 6/07/26

Aceptado: 28/07/26

Online: 26/08/26

Correspondencia

silviamolina@unicauca.edu.co



¿Por qué se realizó este estudio?

Para abordar la falta de un enfoque estructurado que permita medir la comprensibilidad de los datos empresariales en organizaciones de desarrollo de software, con el objetivo de apoyar el análisis de datos y la toma de decisiones organizacionales.

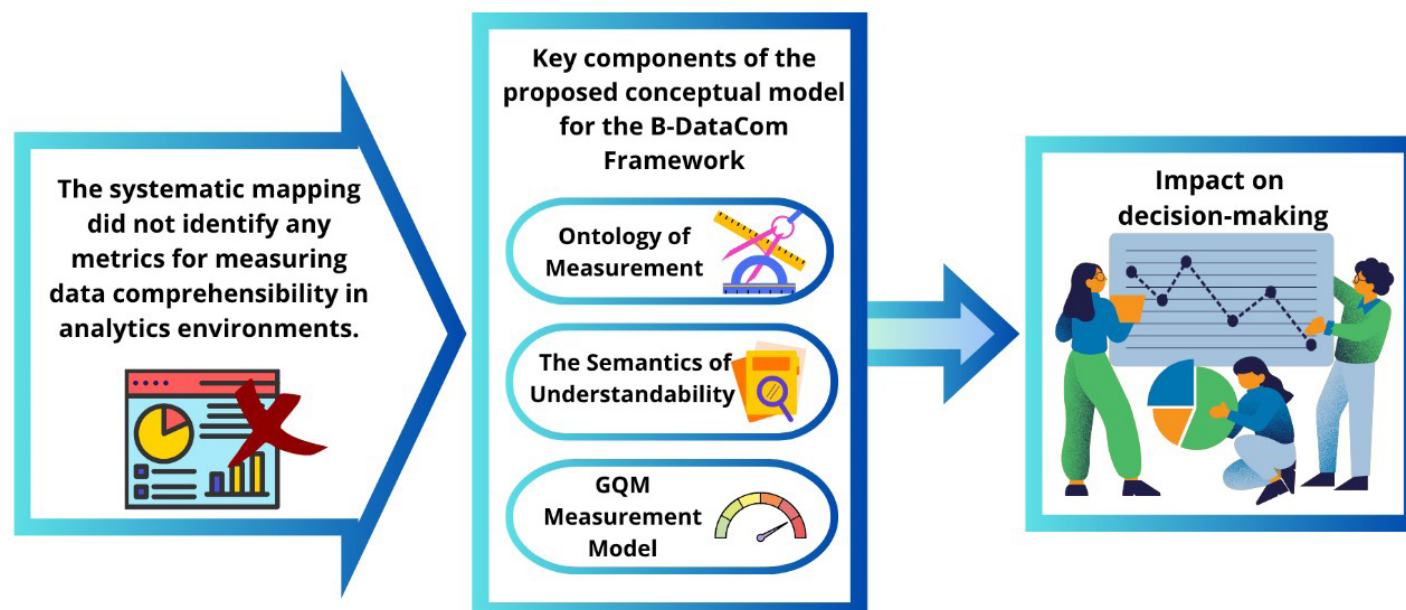
¿Cuáles fueron los hallazgos más significativos?

El mapeo sistemático permitió identificar los principales aspectos descriptivos y factores asociados que influyen en la comprensibilidad de los datos, puso de manifiesto la ausencia de propuestas específicas para su medición y, a partir de estos resultados, propuso el marco B-DataCom como un primer enfoque conceptual para su evaluación.

¿Qué aportan estos resultados?

Los resultados proporcionan una base conceptual para medir la comprensibilidad de los datos mediante la integración de una ontología de medición, los hallazgos de la investigación (aspectos descriptivos y factores que afectan a la comprensibilidad) y el modelo GQM, con el propósito de fortalecer futuras estrategias de medición y apoyar la toma de decisiones en organizaciones de desarrollo de software.

Graphical Abstract



Introducción

En la industria del software, las organizaciones deben garantizar que la calidad de los datos empresariales se mantenga en todos sus procesos, de modo que los datos sigan siendo fiables y precisos para su uso previsto, permitiendo así que esta información apoye una toma de decisiones empresarial eficaz (1). La integración de enfoques como la ciencia de datos, tecnologías como el Big Data y prácticas como el análisis de datos ha transformado significativamente la industria de la calidad de la información (2). El Big Data aborda los complejos desafíos que plantean la gran variedad y el rápido crecimiento de grandes volúmenes de información, así como el procesamiento y análisis de múltiples fuentes y formatos, con el objetivo de extraer verdadero valor en diferentes niveles organizativos. El proceso de análisis exploratorio de datos (EDA) también aporta valiosas ideas dentro del mismo contexto organizativo (3). Esto obliga a las organizaciones a avanzar en sus herramientas de análisis de datos para abordar desafíos cada vez más complejos derivados de la abundante y dinámica información presente en los procesos empresariales (4).

Desde la perspectiva de la calidad de los datos, se pueden considerar diversas características medibles, incluidas las definidas por la norma ISO/IEC 25012 (5). Sin embargo, aunque existe información sobre el modelo de calidad de datos mencionado, sigue existiendo una laguna en este contexto respecto a cómo Comprensible Los datos están destinados a apoyar y fortalecer las estrategias competitivas para la toma de decisiones organizativas (6,7).

Uno de los retos a los que se enfrentan las empresas de desarrollo de software en cuanto a la comprensibilidad de los datos es el rápido crecimiento del volumen de datos, junto con la multiplicidad de formatos y fuentes como Excel, TXT, PDF o archivos CSV. Esto se agrava por la alta probabilidad de contener datos inconsistentes, incompletos o erróneos (8). Estas cuestiones subrayan además la necesidad de datos comprensibles, fiables y útiles que puedan integrarse en herramientas que apoyen la evaluación de la calidad de los datos, beneficiando así la toma de decisiones empresariales (9).

Este artículo presenta los avances en el proyecto de investigación titulado “Marco para medir la comprensibilidad de los datos empresariales en organizaciones de desarrollo de software”, en el que se ha realizado un estudio sistemático de mapeo sobre la comprensibilidad de datos. El objetivo es examinar el papel de esta característica en el contexto empresarial del desarrollo de software y analizar cómo se ha abordado la comprensibilidad de los datos en investigaciones previas relacionadas con procesos de análisis de datos. Además, este progreso en la investigación incluye la definición de un modelo conceptual inicial para el Marco B-DataCom, cuyo propósito es medir el grado de comprensión de los datos empresariales en el contexto especificado, sirviendo así como mecanismo para la evaluación y mejora de la calidad de los datos empresariales en apoyo a una toma de decisiones organizativa mejorada. En consecuencia, el documento está organizado de la siguiente manera: La Sección Dos presenta la metodología utilizada para el estudio sistemático de cartografía; La Sección Tres describe los resultados obtenidos a partir del mapeo. Finalmente, se introduce la versión inicial del Marco B-DataCom, junto con sus limitaciones y trabajos futuros. La Sección Cuatro presenta las conclusiones preliminares de esta investigación.

Metodología

Desarrollar el estudio sistemático de mapeo sobre la comprensibilidad de datos en el contexto de empresas de desarrollo de software, y considerando su impacto en la toma de decisiones organizativas, el protocolo propuesto por (10) fue adoptado. Este protocolo está estructurado en las siguientes cuatro etapas: pregunta de investigación, palabras clave, estrategia de búsqueda de información y evaluación de la calidad de los estudios primarios. Cada una de estas etapas se detalla a continuación.

Pregunta de investigación

¿Qué propuestas existen sobre la comprensibilidad de los datos que apoyen la toma de decisiones empresariales en empresas de desarrollo de software?

Claves clave

Para construir la cadena de búsqueda, se utilizaron las siguientes palabras clave: Análisis de Datos, Desarrollo de Software, Comprensibilidad, Calidad de los Datos, Toma de Decisiones. A partir de estos, la cadena de búsqueda se definió de la siguiente manera:

((Análisis de datos) O (Comprensibilidad)) Y (Desarrollo de Software)

Estrategia de búsqueda de información

Para el proceso de búsqueda de información se utilizó la base de datos Scopus, ya que es una plataforma con amplia cobertura de literatura científica revisada por pares e indexa las principales fuentes de publicación en informática e ingeniería. Esto permitió obtener una visión representativa del estado del arte en la literatura científica relevante para el tema en cuestión. Solo se consideraron artículos de investigación en las áreas de informática e ingeniería, acceso abierto, y publicados entre 2016 y 2025. Al ejecutar la estrategia de búsqueda de información, se encontraron un total de 263 artículos. El proceso de selección de estos estudios fue llevado a cabo por los dos investigadores principales del proyecto, quienes aplicaron los criterios establecidos de inclusión y exclusión y discutieron cualquier discrepancia en la selección de artículos hasta alcanzar un consenso. El proceso de selección para los estudios se llevó a cabo de acuerdo con la metodología sistemática de mapeo adoptada para esta investigación. Posteriormente, los resultados se filtraron según los siguientes criterios de inclusión y exclusión:

Inclusión: Artículos cuyo título, resumen y conclusiones hacen referencia a la analítica de datos o calidad de datos empresariales en términos relacionados con las palabras clave, definidas manualmente según el enfoque del estudio: análisis de datos, Big Data, software, análisis predictivo, minería de datos, toma de decisiones, análisis visual, metadatos, ciencia de datos, análisis de datos, calidad de datos, gestión de datos, adquisición de datos, basado en datos. Tras aplicar estos criterios, se identificaron un total de 20 documentos relevantes .

Exclusión: Artículos que no abordan explícitamente, dentro de su ámbito de estudio, factores e impactos relacionados con la comprensibilidad de los datos en el contexto del desarrollo de software. A partir de estas exclusiones, se obtuvo un conjunto final de 12 artículos primarios .

El proceso de identificación, selección, evaluación de elegibilidad e inclusión de estudios se presenta en la Figura 1.

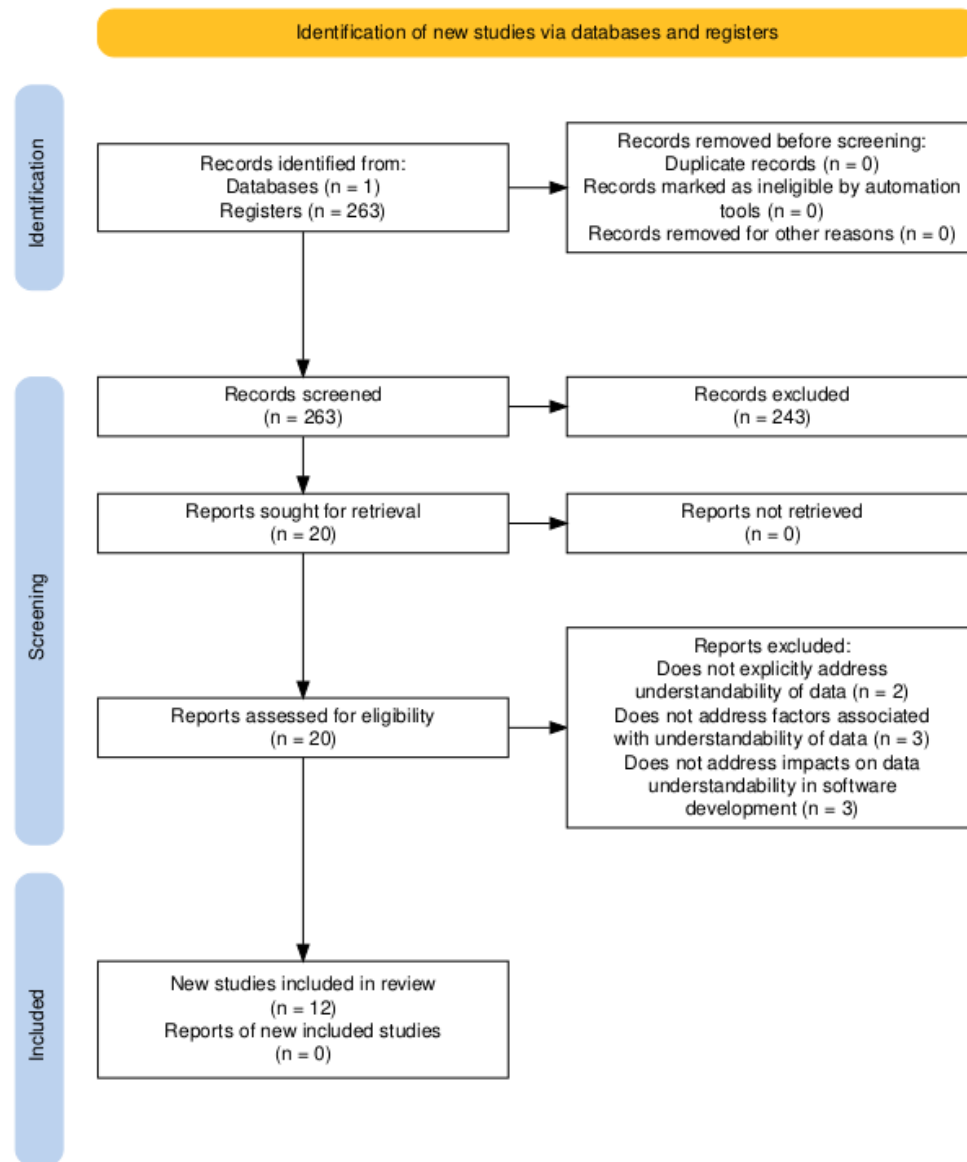


Figura 1. Diagrama de flujo del proceso de selección del estudio (adaptado de PRISMA 2020) (11).

Fuente: Elaboración propia usando la herramienta PRISMA2020 Shiny App (12)

Evaluación de la calidad de los estudios primarios

Determinar el rigor y la relevancia de los estudios primarios (10), se realizó una evaluación de calidad sobre las 12 propuestas. Para este fin, se aplicaron las siguientes tres preguntas de investigación: **(i)** ¿Incluyen técnicas de medición de calidad de datos? **(ii)** ¿Cuál es la importancia de la calidad de los datos para el análisis empresarial según los autores? **(iii)** ¿Contienen aspectos relacionados con la comprensibilidad que permiten su descripción?

Para cada una de las preguntas planteadas, se evaluó el grado de contribución de calidad (GAC) al trabajo de investigación. La evaluación fue realizada de forma independiente por los dos

investigadores principales del proyecto, quienes asignaron las valoraciones correspondientes a cada estudio mediante una escala de Likert (13), que clasifica el nivel de contribución de la siguiente manera: 1 para aportación baja, 2 para contribución media y 3 para aportación alta. El resultado de este proceso de evaluación de calidad se presenta en la Tabla 1, en función del modo de las calificaciones.

Tabla 1. Modo por propuesta en la evaluación de la GCA

Árbitro.	Calificación de la escala de Likert por pregunta			Modo
	(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

Como se muestra en la Tabla 1, las valoraciones más frecuentes se encuentran en las escalas alta y media. Se observaron distribuciones bimodales, indicando una percepción equilibrada con estas mismas escalas. Esto aporta mayor fiabilidad a la investigación al demostrar coherencia en la evaluación de la contribución de los estudios. En los casos en que las calificaciones mostraron una distribución bimodal, ambos valores se mantuvieron en la tabla, respetando la naturaleza del modo obtenido y asegurando transparencia en la presentación de los resultados.

Del análisis se obtuvieron las siguientes respuestas a las preguntas planteadas:

(i) ¿Qué aspectos se miden, en relación con la calidad de los datos, en las propuestas? Se identificó que el 100% de los artículos (12) abordan aspectos de calidad de datos. Estudios (4) y (6) se basan en mediciones subjetivas que evalúan la percepción y la facilidad de comprensión de los datos, como la escala de Likert. Además, los estudios (9), (15), (20), (21), y (22) se basan en mediciones cuantitativas directas como el número de errores, el tiempo y la precisión, con el objetivo de evaluar la eficiencia de los datos, afectando indirectamente su comprensibilidad. Por otro lado, estudios (14), (17), (16), y (18) emplean modelos teóricos que estructuran la evaluación mediante atributos como accesibilidad e interoperabilidad. Por último, estudia (19) implica una medición técnica centrada en el procesamiento y estandarización de datos, con el objetivo de reducir la ambigüedad en la comprensibilidad. Esto demuestra que, aunque los estudios abordan ciertos aspectos de la calidad de los datos, no incorporan técnicas explícitas para su medición.

(ii) ¿Cuál es la importancia de la calidad de los datos para el análisis empresarial según los autores? Se encontró que el 100% de los artículos (12) destacan el papel e importancia de los datos en la analítica dentro del contexto empresarial del desarrollo de software. En cuanto a la calidad de

los datos, se identificó que, de estos 12 artículos, 8 (66,7%) aportan una contribución elevada al describir las contribuciones a la mejora de la calidad de los datos de manera concisa y directa, lo que facilita la comprensión de los datos y los procesos analíticos para la toma de decisiones en el contexto empresarial del desarrollo de software. Los cuatro artículos restantes (33,3%) aportan una contribución media, ya que sus contribuciones a la calidad de los datos son indirectas, dado que su enfoque principal no aborda específicamente este aspecto.

(iii) ¿Contienen aspectos asociados a la comprensibilidad que permiten su descripción? De los 12 artículos (100%), siete (58,3%) aportan una contribución media, ya que incluyen aspectos como la facilidad, claridad e inteligibilidad que pueden asociarse con la comprensibilidad, aunque no se abordan explícitamente como un atributo directo en su estudio. Los cinco artículos restantes (41,7%) aportan una contribución baja, ya que estos aspectos asociados con la comprensibilidad se abordan de forma mínima e indirecta en los estudios. Esto indica la necesidad de identificar aspectos que permitan la descripción de la comprensibilidad de los datos en el contexto del desarrollo de software.

A partir de estos resultados, se decidió continuar con los 12 trabajos de investigación en el proceso de Mapeo Sistemático.

Resultados y discusión

Para organizar la información encontrada sobre la comprensibilidad de los datos en el contexto de las organizaciones de desarrollo de software, se estableció un esquema de clasificación que define tres perspectivas de análisis: (i) estudios relevantes, (ii) la perspectiva en el ámbito de la metodología y (iii) la perspectiva en el ámbito de la comprensibilidad de datos. Dentro de cada perspectiva, se identificaron temas para analizar. Cada uno de estos resultados se describe a continuación.

Estudios relevantes

A continuación se presenta un análisis de la investigación relevante identificada durante el proceso sistemático de mapeo, agrupada en tres núcleos temáticos: (i) comprensibilidad y calidad de los datos; (ii) uso e integración de metadatos; (iii) análisis de datos, Big Data y toma de decisiones.

Comprensibilidad y calidad de los datos

Estudio (9) Presenta las opiniones de profesionales y empresarios sobre cómo la comprensibilidad, utilidad y fiabilidad apoyan los modelos de calidad para herramientas de análisis de software. De manera similar, (23) busca profundizar en la comprensión de los datos para generar mayor valor en servicios y productos, contribuyendo con un marco a través de su investigación. Aunque ambos estudios abordan la comprensibilidad de forma indirecta, no se presenta como un atributo directamente medible mediante métricas específicas.

Uso e integración de metadatos

Estudio (15) demuestra cómo los metadatos ayudan en el análisis y la comprensión de archivos de datos públicos de proteómica a través del formato MAGE-TAB. Igualmente, (17) analiza metadatos

establecidos que soportan flujos de trabajo computacionales FAIR guiando y registrando su procedencia, mientras que (24) explica cómo los metadatos de Git contribuyen a la comprensión de proyectos de software mediante la exploración interactiva. Por su parte, (25) se centra en los metadatos como elemento esencial en la búsqueda de servicios y repositorios, y (26) destaca su importancia en la computación reproducible (RCS). Por último, (27) utiliza metadatos para diseñar plataformas de análisis de datos y mejora de requisitos, integradas en la infraestructura del marco DASE. Aunque estos estudios reconocen el papel de los metadatos en la mejora de la comprensibilidad de los datos, no proponen métricas específicas para evaluar esta característica en apoyo de procesos analíticos dentro del contexto empresarial del desarrollo de software.

Análisis de datos, Big Data y toma de decisiones

Estudios (3), (4), y (6) buscan responder cómo la analítica de Big Data impacta y genera valor empresarial en las pymes, centrándose en herramientas analíticas y paneles estratégicos de QaSD para mejorar la toma de decisiones de alto nivel. De manera similar, los estudios (14), (16), y (18) destacan cómo los marcos, metodologías y procesos de digitalización de Q-Rapids en las pymes guían y fortalecen las organizaciones para mejorar los procesos analíticos, la calidad de los datos y la toma de decisiones empresariales. A su vez, (19), (28), y (29) Analizar la mejor manera de invertir en estrategias de análisis de negocio en proyectos de desarrollo de software integrando el Modelado de Información de Edificios (BIM) según las capacidades organizativas, con el objetivo de lograr ventajas competitivas y permitir la toma de decisiones basada en datos. Igualmente, (30), (31), y (32) Propone la integración de gemelos digitales y marcos para optimizar la gestión y el análisis de datos en contextos empresariales. No obstante, aunque estos estudios establecen la importancia del análisis de datos, el Big Data y la toma de decisiones en el desarrollo de software, no consideran directamente la comprensibilidad como un factor potencialmente relevante dentro de estos procesos.

Según el análisis de estudios relevantes, se pone una concentración significativa en el análisis de datos, el Big Data y la toma de decisiones, destacando su uso e importancia en el contexto empresarial del desarrollo de software desde la perspectiva de infraestructuras analíticas, modelos predictivos y marcos tecnológicos, al tiempo que se pasa por alto la calidad semántica de los datos. Además, se ha demostrado que los metadatos ayudan a la comprensión, interpretación, contextualización y análisis de la información. De manera similar, existe una carencia de investigación respecto a la integración y evaluación de la comprensibilidad y calidad de los datos en contextos de desarrollo de software, lo que justifica la necesidad de un marco que apoye y fomente la toma de decisiones empresariales.

Perspectiva en el ámbito metodológico

Desde la perspectiva metodológica del dominio, los estudios primarios identifican las dimensiones: tipos de contribución y tipos de validación. Cada uno se describe a continuación:

Dimensión del tipo de contribución: Los tipos de contribución encontrados en los estudios primarios se describen en la Tabla 2.

Tabla 2. Tipos de contribución (basados en (33))

Contribución	Definición
Estudio (ES)	Trabajos de investigación que, desde una perspectiva específica, resumen, analizan y generan conclusiones que contribuyen desde el contexto estudiado a otros esfuerzos de investigación. Incluye: revisión, clasificación y compendio.
Enfoque (EN)	Trabajos de investigación que describen conceptualmente un tema específico en un proceso para resolver problemas y preguntas. Incluye: enfoque conceptual.
Propuesta metodológica (PM)	Trabajos de investigación que, para lograr una función u objetivo específico, incluyen etapas, procedimientos o pasos. Incluye: método, técnica, metodología.
Herramienta (H)	Investigaciones que crean herramientas para apoyar, comprender, asistir y llevar a cabo un proceso. Incluye: modelo, marco y marco de trabajo.

Dimensión del tipo de validación realizada: Los tipos de validación encontrados en los estudios primarios, entendidos como tipos que definen la fiabilidad y precisión de la investigación publicada, se describen en la Tabla 3.

Tabla 3. Tipos de validación (basados en (34))

Validación	Definición
Teórica (T)	Documentos que sean relevantes y mejoren la práctica. Suelen ser metodológicas y pueden presentar algoritmos. Incluye revisiones y estudios de literatura.
Encuestas (E)	Documentos que presentan respuestas a preguntas planteadas a una audiencia con conocimientos generales sobre el campo. Proporcionan las perspectivas de los resultados a futuros investigadores.
Estudios de Caso (EC)	Documentos que ofrecen una descripción detallada de cómo y por qué se aplicó la práctica. Estos son analíticos y descriptivos con una perspectiva objetiva.

Experimentos (EX)	Documentos que presentan validación mediante un estudio empírico. El investigador tiene control sobre el experimento, incluyendo la población objetivo, la duración, los tipos de tareas y las áreas.
-------------------	---

La Figura 2 presenta el gráfico de burbujas que ilustra la intersección entre las dimensiones TC y TV de los estudios, mostrando así qué tipos se priorizan y guían la investigación.

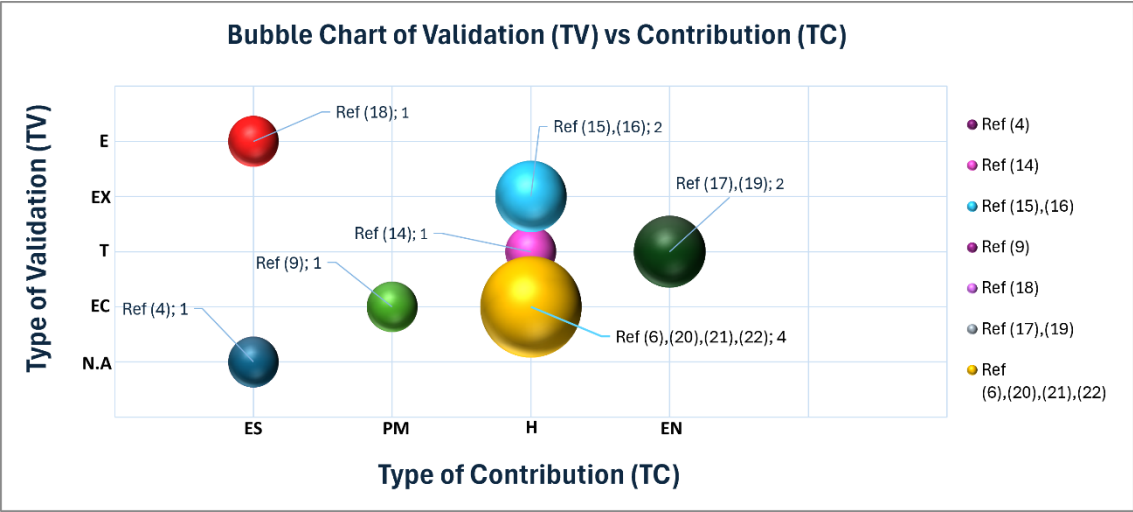


Figura 2. Gráfico de Validación (TV) frente a Contribución (TC). Fuente: Explicación propia.

Basándose en el análisis desde la perspectiva metodológica, se identificó que el 33,33% (4/12) de las propuestas utilizaban Estudio de Caso (CE) como tipo de validación y Herramienta (H) como tipo de contribución; El 16,67% (2/12) corresponden a estudios donde el tipo de validación es Experimento (EX) con el mismo tipo de contribución; El 16,67% (2/12) son estudios donde el tipo de validación es Teórico (T) y el tipo de contribución es Enfoque (EN); 8,33% (1/12) es un estudio con validación teórica (T) y contribución de la herramienta (H); 8,33% (1/12) es un estudio con validación de Estudio de Caso (CE) y contribución a Propuestas Metodológicas (PM); 8,33% (1/12) es un estudio con validación de encuestas (E) y contribución al estudio (ES); y el 8,33% (1/12) no presenta el tipo de validación (N.A) ni el tipo de estudio (ES) como tipo de contribución.

Estas estadísticas reflejan que la investigación sobre la comprensibilidad de los datos en el contexto del desarrollo de software es coherente con la naturaleza del tema, ya que los investigadores priorizan soluciones aplicables en lugar de solo hipótesis no probadas en contextos reales. Este tipo de validación permite mejoras mediante la investigación empírica en diferentes contextos y bajo diversos usos de herramientas. Asimismo, también hay evidencia de una baja tendencia a la investigación con contribuciones en Estudios (1/12) y Propuestas Metodológicas (1/12), así como para validaciones orientadas a Encuestas (1/12). Esta brecha puede reforzarse en futuras investigaciones, ya que demuestra deficiencias persistentes en teoría respecto a la calidad y comprensibilidad de los datos en estos mismos contextos.

Perspectiva en el ámbito de la comprensibilidad de los datos

Desde esta perspectiva, los estudios identifican temas que permiten comprender el estado actual de la comprensibilidad de los datos en el contexto del desarrollo de software. Cada uno de ellos se enumera a continuación:

Elementos de Medición de la Calidad de los Datos (EMC): Técnicas, métricas y métodos para medir la calidad de los datos en el contexto del desarrollo de software.

Aspectos descriptivos de la comprensibilidad de datos (ADM): Aspectos descriptivos del término “comprensibilidad” que se encuentran en la literatura. La identificación de estos aspectos se realizó mediante codificación manual de los estudios primarios, extrayendo los términos usados por los autores para describir o asociar la comprensibilidad de los datos, que posteriormente se clasificaron como aspectos descriptivos.

La siguiente Tabla 4 presenta un resumen de los 12 estudios primarios, junto con la clasificación de las perspectivas previamente descritas, en términos de Elementos de Medición de la Calidad de los Datos (EMC) (técnicas, métricas y métodos) y Aspectos Descriptivos de la Comprensibilidad de los Datos (ADM). A continuación se presenta un análisis de esta perspectiva.

Tabla 4. Resumen de los resultados de los estudios primarios y el esquema de clasificación de la información.

Árbitro.	EMC	ADM
(4)	Escala Likert Scale (de 1 a 5 puntos) (escala ordinal para cuantificar respuestas subjetivas)	Fácil
(9)	Número de problemas completamente especificados (un punto de datos directo para mediciones posteriores)	Tranquilo, despejado
(14)	Modelo de calidad Q-Rapids (un modelo)	Fácil, evidente
(15)	KPI (Indicador Clave de Rendimiento): Eficiencia del flujo de trabajo, tasa de cumplimiento, velocidad de procesamiento (TB/hora), precisión en la detección de anomalías.	Fácil
(16)	MAGE-TAB Proteomics (formato estándar de anotación de datos)	Inteligible
(17)	Modelo de calidad Q-Rapids	Fácil
(18)	KPI (Indicador Clave de Rendimiento): Indicadores FAIR (Localizables, Accesibles, Interoperables, Reutilizables)	Fácil, evidente
(6)	Escala de Likert (de 1 a 5 puntos) Actitudes, opiniones y percepciones de los usuarios. Técnica Likert	Fácil

(19)	Coincidencia fiable de nombres multiculturales (enfoque), Estandarización (técnica), Limpieza (técnica), Herramientas de perfilado de datos (herramienta), Emparejamiento (técnica), Técnicas de supervivencia (técnica).	Fácil
(20)	Dimensión de calidad: Precisión, Completitud, Consistencia, Puntualidad y Asimetría.	Fácil
(21)	Dimensión de calidad: Precisión, Completitud, Conformidad, Unicidad, Consistencia y Legibilidad.	Fácil
(22)	Métricas: métrica de distribución única y métrica de distribución doble.	Fácil

Durante el proceso de análisis de la información obtenida a través de los temas, se identificó que el 91,67% (11/12) asocia y caracteriza aspectos de comprensibilidad con el concepto de “Fácil”, el 8,33% (1/12) con el concepto de “Inteligible”, así como el 8,33% (1/12) con el concepto de “Claro” y el 16,67% (2/12) con el concepto de “Evidente”. La palabra “Fácil” es la más recurrente entre los estudios primarios. Según la literatura, los autores buscan implícitamente que la información en el campo de la calidad de los datos pueda ser entendida sin ambigüedad ni dificultad por parte del usuario. Sin embargo, esta interpretación superficial puede dificultar una conceptualización adecuada y objetiva de la comprensibilidad de los datos según las necesidades de las empresas de desarrollo de software. Además, hay evidencia de que no existe una definición formal de lo que es la comprensibilidad de los datos, ni hay aspectos descriptivos directamente relacionados con ella. Por ello, se demuestra la necesidad de caracterizar y definir la comprensibilidad de forma más rigurosa en el contexto organizativo de las empresas de desarrollo de software.

En cuanto a los elementos de medición de la calidad de los datos mostrados en la Tabla 4, se identifica que el 66,67% (8/12) representan Indicadores Clave de Rendimiento (KPI) en técnicas como: número de problemas completamente especificados, modelo de calidad Q-Rapids, precisión, completitud, consistencia, puntualidad, asimetría, precisión, completitud, conformidad, unicidad, consistencia, legibilidad, emparejamiento fiable, perfilado de datos y técnicas de emparejamiento. Esto indica que, aunque en la práctica se realizan mediciones cuantitativas orientadas al rendimiento o la calidad de los datos, aún carecen de formalización, trazabilidad y uniformidad.

Modelo conceptual del marco B-DataCom

En cuanto a la definición de una versión inicial del Marco B-DataCom (Marco de Comprensibilidad de Datos Empresariales), esta investigación adopta una perspectiva que permite una medición estructurada de la comprensibilidad de los datos empresariales, basada en conceptos teóricos como la Ontología de Medición (35) y centrado en la semántica de la comprensibilidad, apoyado por el modelo de medición GQM (36). Estos cimientos delimitan el alcance constructivo del marco, como se ilustra en la Figura 3.

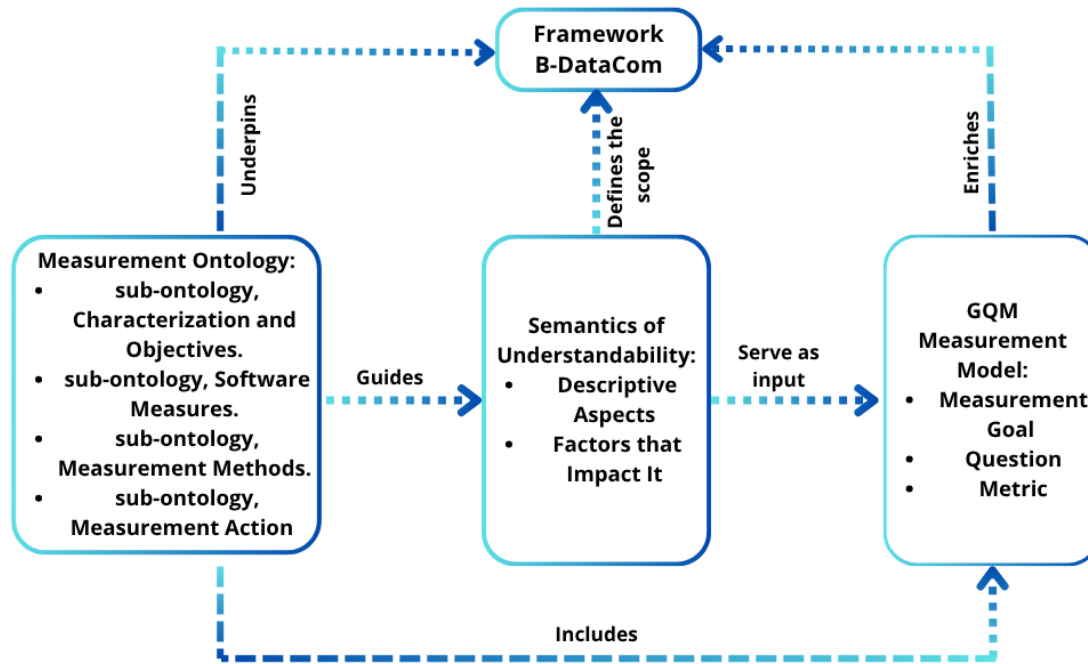


Figura 3. Modelo conceptual del marco B-DataCom. Fuente: Explicación propia.

Uno de los pilares del Marco B-DataCom es la Ontología de Medición, que comprende cuatro subontologías: caracterización y objetivos, medidas de software, métodos de medición y acciones de medición. En conjunto, facilitan la interpretación y representación del contexto de medición. Este pilar sustenta el marco al proporcionar elementos ontológicos que abordan la necesidad de medir la comprensibilidad de los datos en el contexto de las organizaciones de desarrollo de software. También guía el componente semántico de la comprensibilidad e incorpora la teoría del Modelo de Medición GQM.

Por otro lado, la semántica de la comprensibilidad, como segundo pilar conceptual del Marco, establece el alcance del significado de comprensibilidad dentro del Marco y sirve como aporte para el modelo de medición GQM. El objetivo de este componente es definir atributos y factores medibles que afectan la comprensibilidad en el contexto organizacional.

Finalmente, el Marco se enriquece con elementos del Modelo de Medición GQM (Meta-Pregunta-Métrica) (36), permitiéndole obtener indicadores en cuanto al grado de comprensión de los datos. Esto incluye un proceso de medición compuesto por Objetivos de Medida, Preguntas de Medición y Métricas.

El objetivo del Marco B-DataCom es evaluar la comprensibilidad en el contexto de una organización de desarrollo de software. Establece preguntas para identificar datos directos e indirectos relacionados con la comprensibilidad dentro del contexto y, finalmente, métricas que responden a las preguntas establecidas sobre la medición, generando indicadores para la comprensibilidad de los datos. Los tres pilares descritos anteriormente están integrados e interrelacionados, dando lugar al modelo conceptual del Marco B-DataCom que se desarrollará para su implementación. El objetivo es apoyar los procesos de análisis de datos empresariales mediante mediciones estructuradas.

Limitaciones y trabajos futuros: El marco B-DataCom presentado en este artículo corresponde a una versión conceptual inicial. Aún no ha sido implementado ni validado empíricamente en organizaciones de desarrollo de software. Por lo tanto, los resultados se limitan a la base teórica obtenida a partir del mapeo sistemático y a la definición y construcción de los componentes conceptuales que constituyen el Marco B-DataCom. Como resultado de estas limitaciones, la siguiente etapa de esta investigación avanzará en su implementación definiendo métricas GQM asociadas a sus componentes como mecanismos de medición. Posteriormente, será validado mediante un estudio de caso en una empresa de desarrollo de software, evaluando, en un contexto real, la aplicabilidad, utilidad y capacidad del Marco B-DataCom para medir la comprensibilidad de los datos empresariales y apoyar la toma de decisiones organizativas.

Conclusión

El mapeo sistemático identificó 12 estudios primarios que, aunque utilizan métricas de calidad de datos para aspectos ligeramente relacionados con la comprensibilidad, como precisión, completitud y consistencia, no proporcionan una medición estructurada basada en modelos explícitos para la característica de comprensibilidad, considerando su semántica. Del mismo modo, el 91,67% de los estudios primarios se relacionan con la comprensibilidad únicamente con el aspecto descriptivo "Fácil", que evidencia una conceptualización limitada y superficial de esta característica. De manera similar, no existe una conceptualización ni caracterización formal de la misma, lo que ha limitado aún más su operativización en los procesos organizativos, tendiendo a considerarse algo intrínseco a la calidad sin lograr una comprensión profunda de sus implicaciones e impactos en el análisis de datos. Esta brecha dificulta la integración y establecimiento de la comprensibilidad como un atributo medible que apoya el análisis de datos para la toma de decisiones.

Los resultados del mapeo ponen de manifiesto la relevancia de la calidad de los datos para fortalecer la analítica y la toma de decisiones empresariales en el contexto del desarrollo de software, reflejada en el uso predominante de métricas, marcos, modelos y técnicas orientadas a mejorar la calidad y la gestión de datos. Sin embargo, aún no se ha abordado cómo la comprensibilidad influye e impacta en estos procesos. A partir de lo anterior, esta investigación, a través del marco B-DataCom, busca contribuir a reducir esta brecha conceptualizando y midiendo la comprensibilidad como un elemento esencial para el uso efectivo de los datos empresariales en entornos organizacionales, con el objetivo de apoyar los procesos de análisis de datos empresariales basados en la comprensión de los distintos tipos de datos presentes en el contexto del desarrollo de software, como proveedores, costes, desarrollo de productos, clientes, proyectos y datos de entrenamiento, entre otros.

Desde una perspectiva teórica, esta investigación amplía la aplicación de la Ontología de Medición al ámbito de la comprensibilidad de datos, que incluye el modelo de medición GQM, guía la semántica de la comprensibilidad y sustenta el Marco B-DataCom. Así, los datos se reconocen como un activo fundamental y estratégico dentro del análisis de datos empresariales, cuya utilidad depende tanto de su calidad como de su grado de comprensión.

A nivel empresarial y social, el Marco B-DataCom tiene como objetivo apoyar el diagnóstico del grado de comprensión de los datos empresariales, entendido como una característica medible perteneciente al concepto de calidad de datos, con el fin de fortalecer las acciones de mejora para la analítica y la toma de decisiones empresariales. A nivel académico, proporciona una base conceptual y metodológica en esta línea de investigación, utilizando fundamentos de modelos de calidad de datos que siguen siendo válidos y esenciales para crear nuevo conocimiento en tecnologías y enfoques relacionados con Big Data y ciencia de datos.

En este sentido, la investigación demuestra la importancia de articular los fundamentos establecidos en el Marco con la dinámica de las organizaciones, ya que el crecimiento continuo del volumen y la diversidad de datos requiere mecanismos que mejoren su comprensión y uso. Esto permite a las organizaciones integrar conceptos de calidad de datos en sus prácticas de análisis empresarial, fortaleciendo así el conocimiento empresarial.

Como dirección futura de investigación, se propone ampliar el Marco B-DataCom incorporando un nuevo componente para la captura y procesamiento automático de datos empresariales, apoyado por técnicas de inteligencia artificial, de modo que los indicadores de comprensibilidad generados por el Marco se obtengan de manera oportuna y continua para las áreas de gestión organizacional implicadas en la toma de decisiones. Además, se propone validar la aplicabilidad y adaptabilidad del Marco en diferentes contextos organizativos y ampliar su alcance a escenarios relacionados con Big Data y ciencia de datos.

Financiación

Esta investigación está financiada por COLFUTURO para programas de máster y doctorado a través de la convocatoria "Formación de capital humano de alto nivel para promover el liderazgo social y comunitario y abordar la Cuarta Revolución Industrial en el departamento de Cauca (4RI)" en 2024.

Declaración sobre el Uso de la Inteligencia Artificial

Grammarly se utilizó exclusivamente para la traducción y corrección del idioma inglés del manuscrito. Los autores revisaron y verificaron todas las modificaciones realizadas y asumen plena responsabilidad por el contenido científico del artículo.

Declaración de créditos de contribución de autoría

Conceptualización – Ideas: Silvia Molina, Sandra Buitrón, Francisco Pino; **Curación de datos:** Silvia Molina, Sandra Buitrón, Francisco Pino; **Análisis formal:** Silvia Molina, Sandra Buitrón; **Adquisición de financiación:** Silvia Molina, Sandra Buitrón, Francisco Pino; **Investigación:** Silvia Molina, Sandra Buitrón; **Metodología:** Silvia Molina, Sandra Buitrón, Francisco Pino; **Administración del proyecto:** Silvia Molina, Sandra Buitrón, Francisco Pino; **Supervisión:** Silvia Molina, Sandra Buitrón, Francisco Pino; **Validación:** Silvia Molina, Sandra Buitrón, Francisco Pino; **Visualización – Preparación:** Silvia Molina, Sandra Buitrón; **Redacción – borrador original – Elaboración:** Silvia Molina, Sandra Buitrón; **Escritura – revisión y edición – Elaboración:** Silvia Molina, Sandra Buitrón, Francisco Pino;

Implicaciones éticas: ninguna declarada; **Conflictos de intereses:** ninguno declarado.

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.

<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.jiime.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. Pezník 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>

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>