

# Un análisis bibliométrico de evaluaciones de calidad de datos en sistemas de gestión académica

*A bibliometric analysis of data quality assessments in academic management systems*

Erika Reina-Guaña<sup>1</sup> 

<sup>1</sup>Instituto Superior Tecnológico Luis Tello, Esmeraldas, Ecuador, 080116

\*Correspondencia: epreina@insluistello.edu.ec

**Citación:** Reina-Guaña, E., (2024). Un análisis bibliométrico de evaluaciones de calidad de datos en sistemas de gestión académica. *Novasinerugia*. 7(1). 163-179.

<https://doi.org/10.37135/ns.01.13.10>

Recibido: 28 julio 2023

Aceptado: 26 octubre 2023

Publicado: 10 enero 2024

Novasinerugia  
ISSN: 2631-2654



**Copyright:** 2024 derechos otorgados por los autores a Novasinerugia.

Este es un artículo de acceso abierto distribuido bajo los términos y condiciones de una licencia de Creative Commons Attribution (CC BY NC). (<http://creativecommons.org/licenses/by/4.0/>).

**Resumen:** La calidad de los datos es un ámbito multidimensional, que refiere al manejo de información interdisciplinaria; para la gestión, operación y toma de decisiones en las instituciones. De cara a lo antes mencionado, el objetivo del presente estudio es explorar la evolución y las tendencias en la investigación sobre evaluaciones de calidad de datos en sistemas de gestión académica, entre 1995 y 2023, a través de un análisis bibliométrico. La metodología está estructurada por dos fases; la primera fase abarcó la extracción y procesamiento de los datos utilizando la base de datos Scopus y el Software VOSviewer y Bibliometrix. La segunda fase comprende el análisis bibliométrico a partir de los indicadores bibliométricos establecidos (cantidad, calidad y estructura). Como resultado se logró identificar un crecimiento progresivo del tema entre 2006 y 2022. También se realizó un análisis de los indicadores bibliométricos como: producción de documentos y autores, distribución de frecuencias de productividad, análisis de países e instituciones, calidad de revistas y coocurrencia de palabras clave. Se concluye que son más de 100 autores y 50 instituciones influyentes en el campo. Además, se determinó que el 80% de las revistas vinculadas al tema se encuentra entre el Cuartil 1 y 4, lo que destaca un alto nivel de prestigio en las investigaciones. Aunque se ha obtenido una visión valiosa de la producción científica sobre calidad de datos en diversos temas, existen escasos trabajos centrados en la calidad de datos en sistemas de gestión académica; esto indica oportunidades de investigación para abordar esta área poco explorada.

**Palabras clave:** Bibliometría, Bibliometrix, Estudio bibliométrico, Evaluación de calidad de datos, Sistema de gestión académica.

**Abstract:** *Data quality is a multidimensional field, which refers to the management of interdisciplinary information; for management, operation and decision making in institutions. In view of the above, the aim of this study is to explore the evolution and research trend on data quality assessments in academic management systems, between 1995 and 2023, through a bibliometric analysis. The methodology is structured by two phases; the first phase comprised data extraction and processing using Scopus database and VOSviewer and Bibliometrix software. The second phase comprised the bibliometric analysis based on the established bibliometric indicators (quantity, quality and structure). As a result, it was possible to identify a progressive growth of the subject between 2006 and 2022. An analysis of bibliometric indicators such as: production of documents and authors, distribution of productivity frequencies, analysis of countries and institutions, quality of journals and co-occurrence of keywords was also carried out. It is concluded that there are more than 100 authors and 50 influential institutions in the field. In addition, it was determined that 80% of the journals linked to the subject are between Quartile 1 and 4, which highlights a high level of prestige in research. Although valuable insight has been gained into the scientific production on data quality in various topics, there are few papers focused on data quality in academic management systems; this indicates research opportunities to address this underexplored area.*

**Keywords:** *Bibliometrics, Bibliometrix, Bibliometric study, Data quality assessment, Academic management system.*

## 1. Introducción

La era digital en la que estamos inmersos ha revolucionado la forma en que una masiva cantidad de datos se recopilan, almacenan y procesan. Es así como, las herramientas tecnológicas disponibles en las instituciones se han convertido en estrategias para la consecución de sus objetivos. El crecimiento exponencial de los datos manejados por los sistemas de información es un aspecto crítico para satisfacer las demandas de los usuarios, ya que estos datos constituyen un recurso esencial en una amplia gama de actividades de la gestión pública (Manrique de la Cuadra, 2018; Martínez, 2022).

Desde esta perspectiva, la evaluación de la calidad de los datos se erige como un componente clave para comprender la situación actual del manejo de los datos en los sistemas de gestión y para validar su influencia en la generación de información efectiva y eficiente. En este contexto, asegurar la calidad de los datos implica una serie de validaciones relacionadas con los requisitos y características de calidad, garantizando que sean precisos, fundamentados, actualizados y precisos (Peralta et al., 2021).

Por otro lado, la innovación de los datos estructurados en las instituciones está experimentando un nuevo enfoque paradigmático, impulsado por la rapidez en la generación de datos, su difusión y la incorporación de tecnologías actuales para su obtención, como dispositivos móviles, redes de datos y la interacción humana. Desde esta perspectiva, investigaciones recientes (Laaz & López, 2020; Gallegos et al., 2021; Patiño, Caicedo & Reina, 2019; Yacelga, 2023) han evidenciado que las instituciones de educación superior enfrentan desafíos significativos, como superar barreras de accesibilidad, garantizar la seguridad y cumplir con normativas vigentes en relación con los datos en los sistemas existentes, así como mejorar la gestión de los procesos académicos. Además, según Gualo et al. (2022), la falta de datos confiables puede conducir a errores en la gestión, operación y toma de decisiones, pérdida de competitividad y daños en la imagen corporativa en caso de una filtración de información personal y confidencial.

Varios autores han contextualizado la importancia de abordar las inconsistencias de los datos en los sistemas, destacando, por ejemplo, el modelo propuesto por Peralta (2021) para evaluar la calidad de los datos basado en el estándar ISO/IEC 25012. Este modelo se implementó utilizando herramientas automáticas y reveló problemas en características de calidad de datos como precisión, consistencia, conformidad y actualidad. Además, autores como Fritz et al. (2017) y Calabrese et al. (2019) han subrayado la necesidad de evaluar la calidad de datos en sistemas de gestión académica y su impacto en la integridad de los datos.

En línea con los estudios previos, este artículo responde a la necesidad de analizar las principales tendencias en las investigaciones sobre evaluaciones de calidad de datos, a través de la aproximación bibliométrica del mapeado científico, se obtiene a partir del análisis de la estructura conceptual, social e intelectual del campo de conocimiento y tiene como objetivo identificar temas clave que deben abordarse en futuras investigaciones, profundizando en la diversidad de sistemas de gestión académica. Para orientar este estudio, se ha establecido una serie de indicadores bibliométricos que se centran en el análisis de la producción de artículos, autores, países e instituciones con una alta producción

documental, redes de coautoría, calidad de revistas publicadas y coocurrencia de palabras clave.

## 2. Metodología

Este trabajo se basó en un estudio Bibliométrico, de acuerdo con Patiño-Toro et al. (2022) la bibliometría es una técnica cuantitativa ampliamente utilizada para analizar el impacto de la producción científica sobre un tema; mapear límites, identificar actores clave y generar nuevas tendencias temáticas en aras de oportunidades de investigación. Esta bibliometría se apoya a partir de la información indexada en la base de datos Scopus, entre el periodo de 2003 y 2023. Para el desarrollo del ejercicio Bibliométrico se establecieron dos fases con base en (Cabuya et al., 2022; Lizano & Sánchez, 2020; Patiño-Toro et al., 2022). En la primera fase se realizó la extracción y procesamiento de los datos que incluye selección de la base de datos, definición de la ecuación de búsqueda, extracción de datos en Scopus y procesamiento de datos en Vosviewer y Bibliometrix. La segunda fase establece el ejercicio bibliométrico de la metadata obtenida a partir de los indicadores bibliométricos establecidos.

### 2.1. Extracción y procesamiento de los datos

#### 2.1.1. Selección de la base de datos

Se ha seleccionado Scopus como la base de datos principal para el análisis bibliométrico debido a que se destaca como una de las bases de datos más grandes y ampliamente utilizadas en la comunidad científica. Además, se ha demostrado que Scopus sigue rigurosos estándares de selección, lo que garantiza la confiabilidad, calidad y alta cobertura interdisciplinaria de los documentos que almacena. Su amplia cobertura integral de revistas revisadas por pares y otros tipos de publicaciones, como libros y actas de congresos, permite obtener una visión más completa de la investigación en diversos campos científicos. Asimismo, Scopus es conocido por su capacidad para proporcionar datos de citas con alta cobertura interdisciplinaria, lo que facilita el estudio y la comparación de diferentes áreas científicas. Por tanto, se considera una base de datos adecuada para el presente estudio (Valencia-Arias et al., 2020).

#### 2.1.2. Ecuación de Búsqueda y extracción de datos en Scopus

La selección de palabras clave se realizó de acuerdo a diversas pruebas de ensayo y error en el motor de búsqueda Scopus seleccionando variantes con palabras compuestas y utilizando la experiencia de otros investigadores sobre el tema (Cabuya et al., 2022). Fueron excluidos los términos "data" y "software" debido a que presentaban resultados relacionados al área de la salud que no se ajustaban a la temática. Para la elaboración de la ecuación de búsqueda se consideró en primera instancia las palabras clave en español y en inglés: "estandar", "evaluación", "sistema de gestion academica" y "software product quality", utilizando el operador booleano AND para que se incluyeran en la búsqueda. De igual forma se utilizó el operador OR para agrupar las palabras claves sinónimas. En combinación, la ecuación de búsqueda necesaria para el análisis es:

TITLE-ABS-KEY ( ( ( ( ( "standard" OR "estandar" OR "ISO" ) AND ( "academic system" OR "sistema de gestion academica" OR "management system" ) OR ( "evaluacion" OR "evaluation" ) ) AND ( "quality" OR "quality of academic software" OR "calidad de datos" OR "data quality" OR "data product quality" ) AND "software product quality" ) ) ) AND PUBYEAR > 1994 AND PUBYEAR < 2024 AND PUBYEAR > 1994 AND PUBYEAR < 2024

La ecuación de búsqueda definida permitió extraer 127 documentos en Scopus considerando título, palabras clave y resumen. El periodo inicial de producción documental analizado se estableció en 1995, año en que se publican investigaciones relevantes y continuas sobre el tema. Cabe indicar que en 1981 se presentó una conferencia en que se incluyeron 49 ponencias y 4 resúmenes; los temas incidieron en la productividad, modelos matemáticos, sistemas de bases de datos, evaluación de herramientas de gestión, aseguramiento de la calidad, calidad de producto de software, hardware y software y educación. Sin embargo, las actas del congreso no obtuvieron citas hasta la actualidad. Por esta razón se considera en el análisis como año inicial 1995 en el que se publicó un nuevo artículo enfocado en evaluaciones de calidad de producto de datos, dejando un sesgo de aproximadamente 13 años sin publicaciones sobre el tema. La fecha de corte se realizó el 14 de junio de 2023.

### 2.1.3. *Procesamiento de datos en Vosviewer y Bibliometrix*

Posterior a la extracción de los datos, se continuó con la carga de los documentos generados en Scopus en el programa RStudio y la aplicación web Biblioshiny del paquete Bibliometrix; encargada de la medición de los indicadores establecidos (Aria & Cuccurullo, 2017). Además, se utilizó el software de código abierto VOSviewer para complementar el mapeo científico de los datos analizados, vinculados a los indicadores de productividad de países, citas por revista, coautoría y coocurrencia (Nandiyanto, Al Husaeni & Al Husaeni, 2021). De esta forma se analiza e interpreta la data capturada dando paso al análisis bibliométrico expresado en el siguiente capítulo.

## 3. **Resultados**

### 3.1. *Indicadores de cantidad y calidad*

En respuesta a los indicadores bibliométricos de cantidad y calidad definidos en el diseño metodológico, a continuación, se exponen los resultados obtenidos. Referidos a cantidad en cuanto a la productividad de las publicaciones y a calidad con base en el desempeño en función de las citas (Patiño-Toro et al., 2022).

#### 3.1.1. *Productividad de artículos por año de publicación*

Considerando la productividad científica analizada en términos de autores, revistas, instituciones y países; se analizan los indicadores de cantidad (Patiño-Toro et al., 2022), como se visualiza en la figura 1 la producción anual entre 1995 y 2023. Los años de mayor producción son 2009 y 2016 con 12 documentos y 2012 con 11 documentos respectivamente,

representando el 18 % del total de la producción. Por el contrario, 1997, 1998, 2000 y 2001 son años que no presentan producción documental. A partir de 2006 se presenta un crecimiento significativo y continuo hasta 2022, cabe indicar que, hasta la fecha de corte de la presente investigación, 2023 no presenta trabajos relacionados al tema. Las primeras investigaciones se enfocaron en el desarrollo de modelos y herramientas dirigidas a evaluar la calidad de productos software, utilizando técnicas de inferencia basadas en reglas difusas para capturar los requerimientos del usuario y los atributos de calidad definidos en la Norma ISO-9126 para automatizar piezas de código que permitan detectar fallas en los sistemas (Antonakopoulos et al., 1995; Dromey, 1995). Estos hallazgos resaltan la creciente importancia de la calidad de datos en sistemas académicos, reflejando cómo la comunidad científica ha ampliado su enfoque en esta área, lo que beneficia a instituciones educativas al mejorar la toma de decisiones y la eficiencia en la gestión académica, dada la influencia de los mismos en la calidad de la educación.

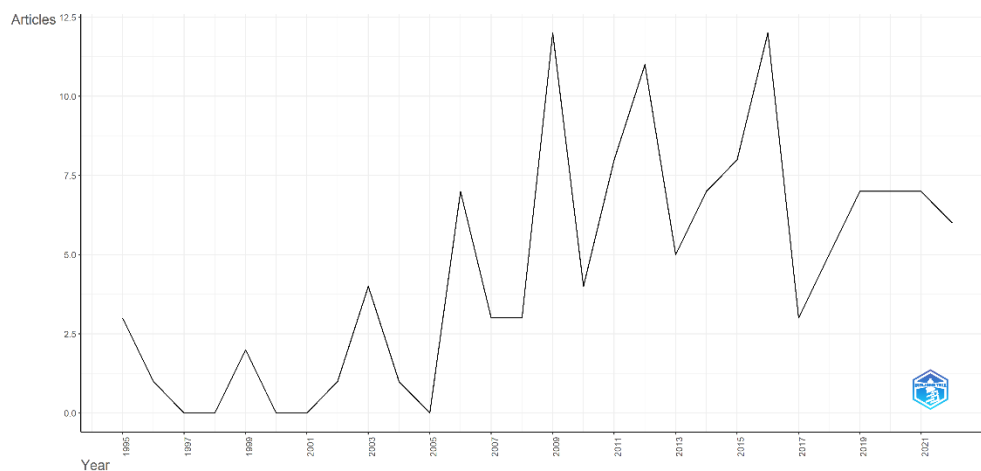


Figura 1. Producción de documentos por año

### 3.1.2. Productividad de autores

Con respecto a los autores más productivos, se puede observar en la figura 2, los diez investigadores con el mayor número de documentos, el autor Idri A. encabeza la lista con un total de 8 artículos publicados; sus investigaciones se centran en la recopilación y medición de requisitos de calidad de aplicaciones móviles y sistemas de apoyo orientadas al área de la salud como: servicio al cliente, donación de sangre y seguimiento del embarazo; aplicando la ISO 25010, 25023 y 25030 a través de la identificación de las características de calidad del producto software en listas de verificación y comprobación. Como resultados de las investigaciones se evidenció que la interacción del usuario y los requerimientos funcionales son los que inciden en las subcaracterísticas externas del modelo de calidad del producto de software (Carrizo, et al., 2021). Además, en uno de los trabajos se realizó un mapeo científico sobre la predicción de la calidad de los productos de software y por último una investigación dirigida a establecer una guía que permite evaluar la calidad de tecnologías móviles mediante la ISO 9126 (Moumane & Idri, 2014). Seguidamente, los autores Abran, A., Al-Qutaish, R., Piattini M. y Visser, J. con un total de 5 investigaciones; sus trabajos están dirigidos en definir, medir, evaluar y lograr la calidad adecuada de los productos de software, la mantenibilidad y la sostenibilidad en el contexto del software de

aplicaciones centrados en el usuario, aplicando manuales como TL9000 y normas ISO 9126, 25000 y 15939. Además, proporcionan a los investigadores una forma de explorar la vinculación de causa y efecto mediante el estudio de los campos de repositorios de datos de proyectos de software relacionados con la calidad que pueden resultar útiles en estudios empíricos y de evaluación comparativa. Asimismo, se proponen enfoques y modelos para evaluar la calidad técnica del código fuente, interpretar grupos de código fuente y estimar el esfuerzo de mantenimiento de los sistemas de software. Los estudios se basan en mediciones y experimentos realizados en diferentes contextos.

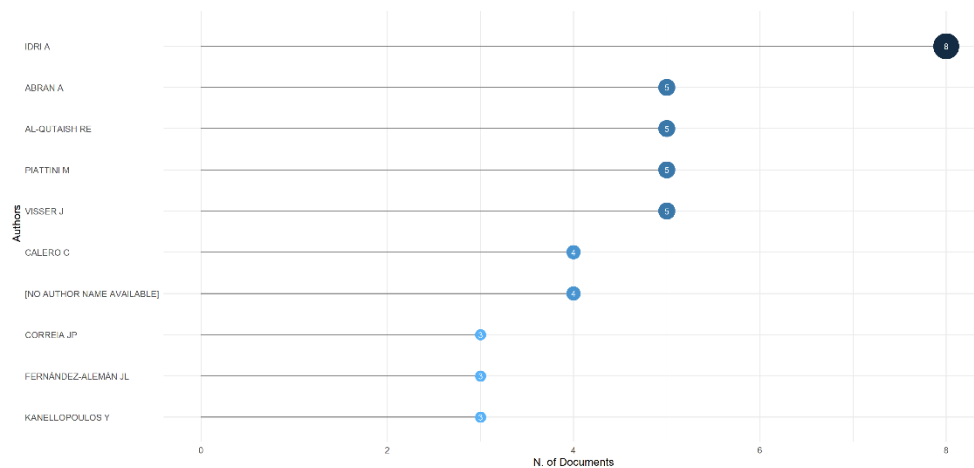


Figura 2. Autores más productivos

En la figura 3 se presenta al indicador de calidad, citas por autor, con un total de 293 autores. De esta forma destacan cinco investigadores más relevantes, con al menos un documento que ha sido citado. El autor Visser J. lidera la lista con 345 menciones, donde se exponen aspectos para el análisis y consultoría de calidad del código, basado en la norma ISO/IEC 9126 y métricas de código fuente. Además, se propone un modelo de calidad técnica que amplía la jerarquía definida por ISO/IEC 9126 y que estudia la relación entre las propiedades del sistema y las características de calidad. También se describe la creación de un repositorio de referencia para comparar sistemas y se propone un método para interpretar grupos de código fuente utilizando el modelo de calidad de ISO/IEC 9126. Por último, se examina el esfuerzo necesario para mantener un sistema y se discuten los problemas y requisitos de un modelo de mantenibilidad, presentando un nuevo enfoque. Seguidamente el autor Dromey (1995), con 336, en el que se proponen modelos para la calidad del producto software asociados a propiedades de calidad utilizadas en la definición de declaraciones y componentes de lenguajes de programación, además el modelo permite definir estándares de codificación específicos del lenguaje, clasificar sistemáticamente defectos de calidad y desarrollar auditores automáticos de código para detectar defectos en el software. También, Heitlager I. con 223, Kuipers T. con 206, donde se aborda la relación entre la calidad técnica del código fuente y el esfuerzo necesario para mantener un sistema de software. Se examina el modelo ISO 9126 y se identifican problemas con el Índice de Mantenibilidad (MI). Se propone un nuevo modelo de mantenibilidad que resuelve estos problemas y se discuten las experiencias en consultoría de gestión de TI. Por último, Bevan N. con 171 en el que se integra un nuevo marco para la calidad del producto de software basado en ISO/IEC 9126-1. El marco incluye la calidad interna (propiedades estáticas del

código), la calidad externa (comportamiento del software durante la ejecución) y la calidad en uso (el grado en que el software satisface las necesidades del usuario). La calidad en uso se relaciona con la usabilidad en términos ergonómicos.

Es notable indicar que estos resultados con base en Garcés-Giraldo et al. (2022) no garantizan la calidad de los documentos publicados, pero permite comprender su impacto en el ámbito académico y científico. Considerando la clasificación anterior, se evidencia una relación relativa entorno a los autores con mayor producción documental de frente al número de citas generadas; esto es debido a que predominan algunos de los autores destacados en ambos indicadores, como es el caso de Visser J. con 345 citaciones, Idri A. con 97 citaciones y Al-Qutaish, R., con 55 citaciones. Por el contrario, el autor Bevan, Nigel cuenta no cumple con esta afirmación ya que solo cuenta con 1 documento y forma parte del ranking de autores más citados.

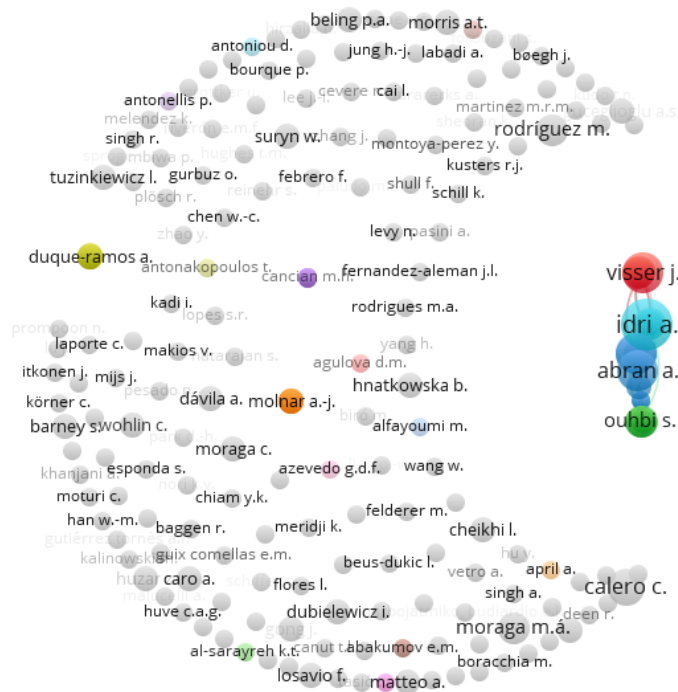


Figura 3. Autores con más citaciones

Además, con base en la estimación del coeficiente de la Ley de Lotka se presenta un modelo sobre la forma en que la producción científica se concentra de acuerdo con los autores en un campo específico. Evidencia que el menor número de autores publica la mayor parte de las investigaciones, por el contrario, la mayoría de los autores tiene una producción científica inferior. Desde esta perspectiva variante, los investigadores profundizan la dinámica de la comunidad científica y la competitividad en los trabajos publicados (Cabuya et al., 2022). En la figura 4, se puede concluir que la literatura científica sobre evaluaciones de calidad de producto de datos en sistemas de gestión académica tiene una correlación significativa. Se destacan en el análisis 247 autores con un solo documento escrito, 35 con dos estudios, 7 con tres trabajos, mientras que 2 han publicado en cuatro ocasiones, 4 autores con 5 documentos y tan solo 1 dispone un total de 8 artículos científicos. Estos hallazgos ayudan a identificar a los líderes en el campo y sus contribuciones clave. Esto permite a otros investigadores



acceder a conocimientos relevantes y entender la dinámica de la comunidad científica, impulsando avances en la calidad del software y la gestión académica.

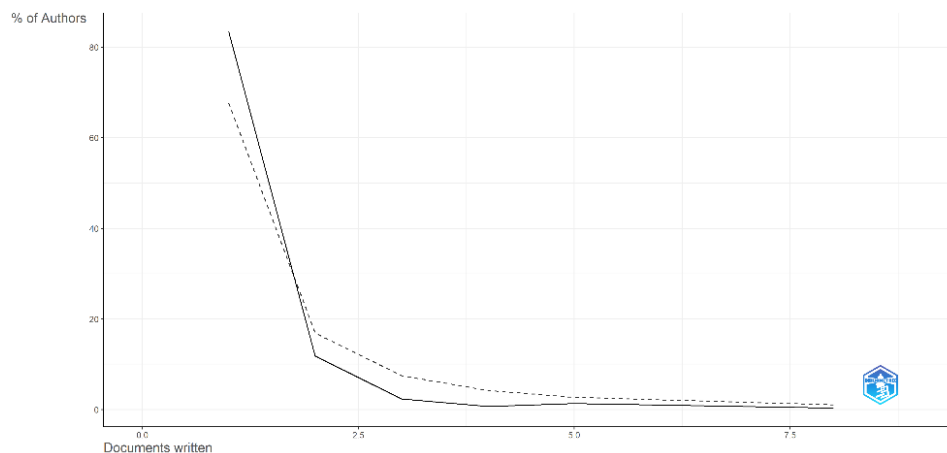


Figura 4. Productividad de países

### 3.1.3. Productividad de las países e instituciones

Se identifican 45 países que trabajan en el tema, organizados en 25 clústeres como se visualiza en la figura 5, siendo España, Países Bajos y Canadá los países con mayor producción documental, con 18, 8 y 7 documentos respectivamente. En coordinación a la calidad e impacto de los temas, España cuenta con 293 menciones, siendo el segundo país más citado, después de Países Bajos con 390. Esto indica que ambos países son los mayores productores de conocimiento con un alto nivel de impacto para la comunidad científica. En este sentido, no sorprende que España al ser la sede de AQClab, el primer laboratorio del mundo en obtener la acreditación ISO/IEC 17025 para la evaluación de la calidad de los productos de software basada en la serie de normas ISO/IEC 25000, sea el país más productor y citado. Se intuye, por tanto, que tiene una fuerte presencia en el ámbito de la investigación y desarrollo de tecnología educativa, así como una destacada participación en estándares internacionales (Verdugo, Rodríguez & Piattini, 2014). Por último, cabe mencionar que el trabajo más citado de España es "Evaluation of the OQuaRE framework for ontology quality" con 44 citas, sin embargo, no está relacionado al tema. Por el contrario, el segundo documento más citado es "Interactions between environmental sustainability goals and software product quality: A mapping study" con 37 citaciones que ofrece una visión general de los enfoques encontrados en la literatura para abordar las interacciones entre la calidad del producto de software y la sostenibilidad en el contexto del software de aplicación. El estudio examina los métodos y estrategias propuestos para integrar consideraciones de sostenibilidad en la evaluación y mejora de la calidad del software. El análisis de estas aproximaciones busca promover una mayor conciencia sobre la importancia de abordar conjuntamente la calidad del software y la sostenibilidad en el desarrollo y mantenimiento de aplicaciones.



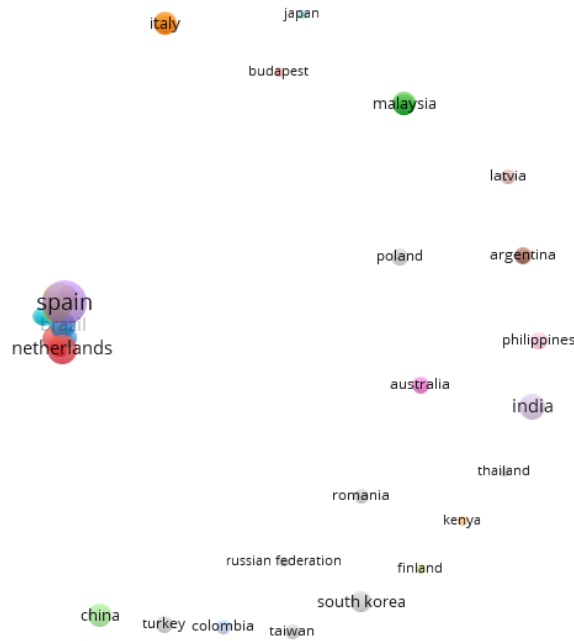


Figura 5. Productividad de países

Frente a la productividad de las instituciones vinculadas, en la figura 6 se presentan las 10 con mayor producción respecto a un total de 160. La Universidad Mohammed V University in Rabat y Ecole Nationale Supérieure d'Informatique ubicadas en Marruecos, lideran la lista con 10 publicaciones cada una, en segundo lugar, Rabat Information Technology Center también situada en Marruecos con 8 publicaciones y en tercer lugar la Universidad de Murcia y la Universidad de Castilla-La Mancha ubicadas en España con 7 productos respectivamente. Se evidencia una amplia difusión del conocimiento sobre el tema, mientras que por otra parte se afirma que dos de las instituciones con mayor número de producciones pertenecen a España que es uno de los países con la mayor cantidad de documentos publicados en Scopus. El análisis de la productividad de países e instituciones vinculados es esencial para identificar líderes en investigación y facilitar la colaboración. Además, ayuda a los investigadores a seleccionar instituciones para posibles colaboraciones y acceder a recursos de calidad en el campo de estudio.

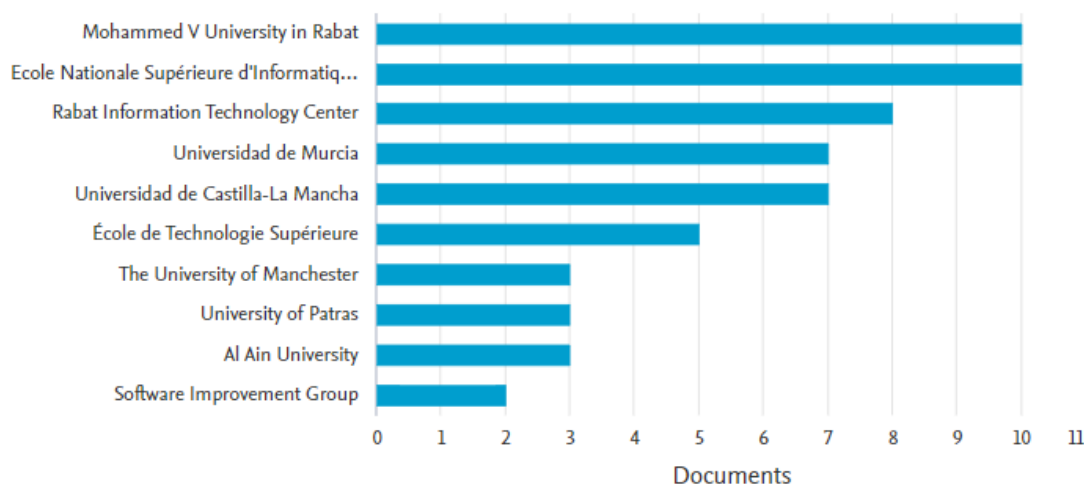


Figura 6. Documentos publicados por afiliación

### 3.1.4. Productividad de revistas

El indicador productividad de revistas es utilizado para explorar y recuperar la información de la evolución que ha tenido la temática desde el análisis de su entorno (Valencia-Arias et al., 2020). Desde este contexto, en la figura 7 se visualiza las diez revistas con el mayor número de publicaciones vinculadas al tema, de un total de 98 analizadas. La revista *Communications in Computer and Information Science* encabeza la lista con 8 documentos producidos, en segundo lugar, figura *ACM International Conference Proceeding Series* y *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* con 5 producciones, luego se posicionan, a una distancia mínima *Lecture Notes in Business Information Processing* y *Software Quality Journal* con 4 producciones cada una. Posteriormente *CEUR Workshop Proceedings* con 3 trabajos y el resto de las revistas cuentan con menos de tres producciones. Con base en el factor de impacto de las 10 revistas más productoras; el 100% están indexadas en el SJR y el 80% se encuentra entre el Cuartil 1 y Cuartil 4, lo que destaca un alto nivel de prestigio y relevancia en las publicaciones sobre el tema. La revista más productiva; *Communications in Computer and Information Science*, está ubicada en el Q4. Las revistas ubicadas en el Q1 corresponden a *Software Quality Journal*, *Computer Standards and Interfaces* e *Information and Software Technology*.

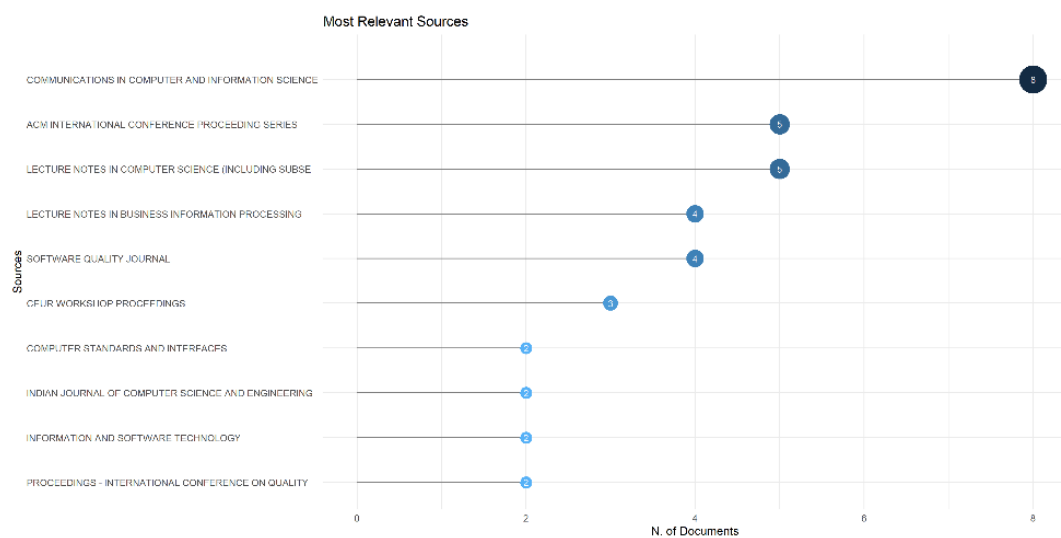


Figura 7. Documentos publicados por revista

En la figura 8, resaltan las 10 revistas con mayor índice de citas por producción, se realizó un filtro de revistas con un mínimo de 2 documentos y al menos una citación vinculada. La revista con el mayor número de citas por divulgación, corresponde a *Software Quality Journal* con 225, en segunda instancia *Quatic 2007 - 6th International Conference on the Quality of Information and Communications Technology* con 214 citas. A continuación, con una diferencia significativa las revistas: *Information and Software Technology*, *Communications in Computer and Information Science*, *Proceedings - International Conference on Quality Software* y *Lecture Notes in Computer Science* con 69, 58, 40 y 39 citas respectivamente. El resto de revistas se sitúan por debajo de las 26 citas.

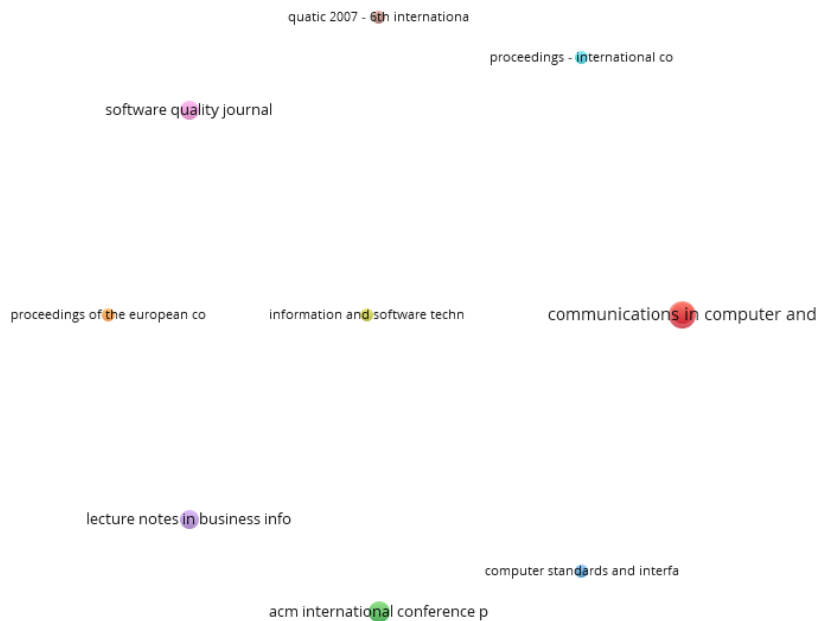


Figura 8. Citaciones por revista

Con respecto a la información obtenida previamente, estos indicadores son relevantes para verificar las revistas que tienen el índice de producción más alto. En este sentido con base en Garcés-Giraldo et al. (2022) se confirma una correlación directa entre el nivel de productividad de una revista y su número de citas; ya que se evidencia que la revista Software Quality Journal siendo la tercera revista más productiva, situada en Q1, posee el mayor número de citaciones (225). Además, cabe mencionar que de las 10 revistas con mayor producción el 80% se encuentran dentro de las 10 con el índice más alto de citaciones. El análisis de la productividad de revistas este campo, proporciona información valiosa sobre las fuentes de conocimiento más destacadas. Además, el hecho de que todas las revistas estén indexadas en SJR y el 80% estén en Cuartiles 1 a 4 refleja un alto prestigio y relevancia en el campo. El conocimiento de estos hallazgos beneficia a los investigadores al orientarlos hacia las fuentes más prolíficas y citadas, lo que puede mejorar la calidad y la visibilidad de sus propias investigaciones en este campo altamente competitivo.

### 3.2. Indicadores de estructura

Los indicadores bibliométricos de estructura analizan la relación y las conexiones entre los investigadores, por tanto, exploran la coautoría y la coocurrencia de palabras clave en la producción científica. Esto contribuye a la comprensión de las redes científicas y tendencias de investigación en un campo determinado (Garcés-Giraldo et al., 2022).

#### 3.2.1. Red de coautoría

Con respecto a la red de coautoría como indica Garcés-Giraldo et al. (2022) refiere a la vinculación entre dos o más autores para colaborar en la publicación de una producción científica, en la figura 9 se presenta 14 clústeres de coautoría siendo los de color rojo y verde, los de mayor peso en la red. El clúster rojo vincula la coautoría de los investigadores Calero C., Caro A., García F., García-Mireles G., Moraga C., Moraga M., Piattini M. y Rodríguez M.

con trabajos que en su mayoría coinciden con análisis de modelos de calidad de datos y modelado de la confiabilidad del software con base en estándares internacionales específicamente la serie ISO/IEC-25000 (Bambazek et al., 2023). Por otro lado, el clúster verde relaciona a los autores Bachiri M., Fernández-Alemán J., Idri A., Ouhbi S., Sardi I. y Toval A., quienes han basado sus estudios en el uso de los estándares ISO/IEC 9126 e ISO 25010 para la evaluación de la calidad del producto software y su incidencia en software de uso personal (Hasanah, Atikah y Rochimah, 2020; Yanes, Sepúlveda, & Díaz, 2019). A continuación, el clúster azul representado por los autores Correia J., Heitlager I. y Kanellopoulos Y., que han colaborado en trabajos enfocados en temas de evaluación y mejoramiento de la calidad técnica de los productos de software mediante el benchmarking, así como estudios basados en encuestas (Arvanitou et al., 2022). La contribución de este análisis radica en la capacidad de identificar a los expertos en cada subcampo, permitiendo colaboraciones más eficaces y el acceso a conocimientos especializados. Además, proporciona información sobre las tendencias de investigación en calidad del software, lo que puede orientar investigaciones futuras y estrategias de publicación.



Figura 9. Red de coautoría

### 3.2.2. Red de coocurrencia

Con respecto a las palabras clave que forman la red de coocurrencia, refieren a la vinculación directa que existe entre dos conceptos que se derivan del tema en estudio (Garcés-Giraldo et al., 2022). A partir de este contexto, en la figura 10, se identifican 84 palabras en un sistema de conexiones constituido por 7 clústeres. Los clústeres se presentan con un color específico para cada grupo de nodos; de esta forma se determina la red de ocurrencia de los términos que se repiten con mayor frecuencia en los artículos de investigación de la data analizada. Se tomó como unidad de análisis las palabras claves indexadas, con un mínimo de 3 ocurrencias de una palabra, debido a la especialización del tema en calidad de datos en sistemas de gestión académica. Las variables vinculadas más representativas dentro de los grupos son: (1-rojo) computer software selection and quality

control, (2-verde) software quality, (3-azul) ISO 25010, (4-amarillo) ISO standards y quality evaluation, (5-morado) software metrics, (6-celeste) software product quality and quality assurance y (7-naranja) software design.

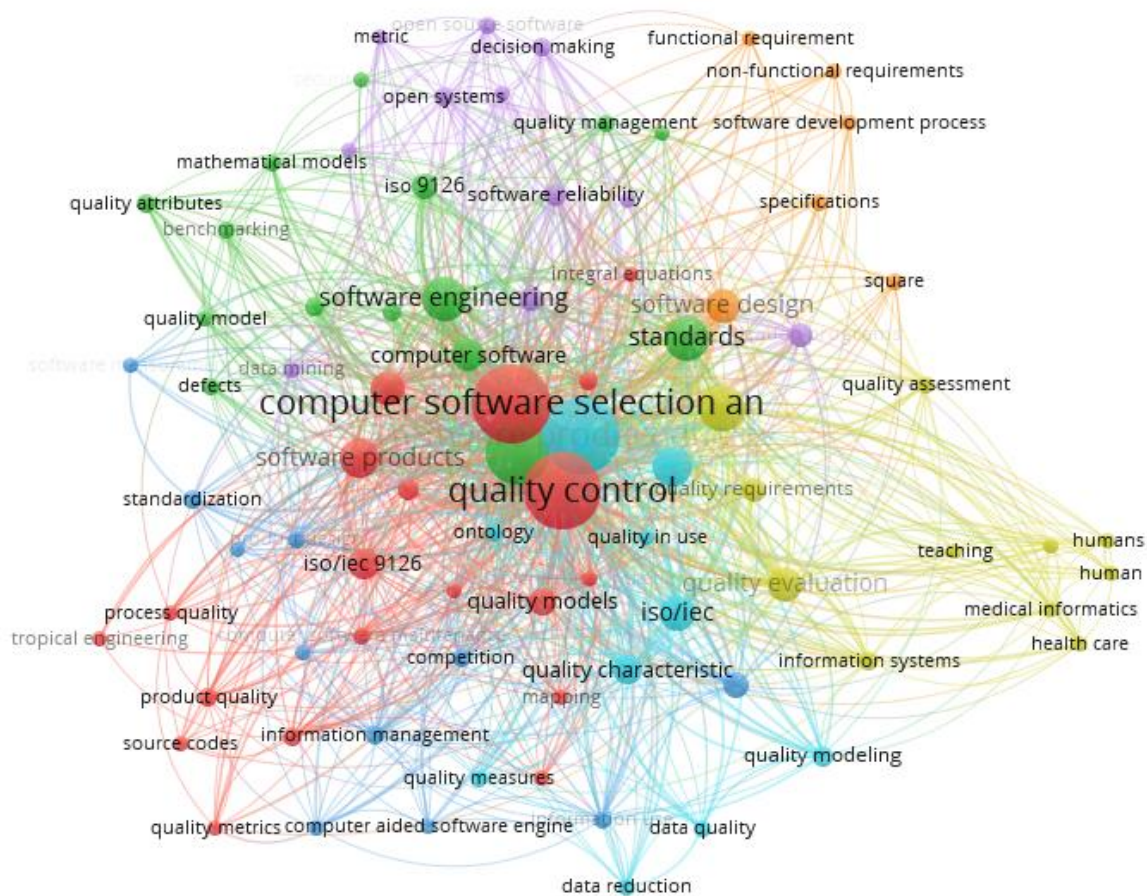


Figura 10. Red de coocurrencia

De acuerdo con la figura 10, en el clúster (1-rojo) los trabajos reflejan aspectos de la calidad del software y la gestión del mantenimiento, mediante el cumplimiento de estándares internacionales como la ISO/IEC 9126, la evaluación de la mantenibilidad, creación de mapas de características del sistema, métricas, modelos y evaluación de la calidad del software en el contexto de las tecnologías de la información. Asimismo, el clúster (2-verde) representan estudios sobre el benchmarking de software aplicando modelos matemáticos y mediciones de calidad para evaluar los atributos de calidad. Es así como, se enfocan en la gestión de calidad, estándares como ISO 9126, ingeniería de software, detección de defectos, seguridad de datos y modelos de calidad de software. En ese mismo contexto, el clúster (3-azul) aborda temas sobre la gestión de la información y su uso en el diseño y la implementación de software; focalizándose en la competencia asistida por software, medición de software, pruebas de software y estándares como ISO/IEC 25010. Igualmente, el clúster (4-amarillo) evidencia la intersección de la atención médica y las tecnologías de la información; se incluyen en los trabajos la aplicación de estándares ISO, evaluación de calidad y requisitos dirigidos a la enseñanza en el ámbito de la informática médica. También en el clúster (5-morado) se resalta un interés por abordar temas centrados en software

reutilizable para mejorar la toma de decisiones. Se exploran métricas, confiabilidad y métricas de software en sistemas académicos. Con la misma relevancia el clúster (6-celeste) destaca la calidad y reducción de datos, modelado de calidad, medidas y características de calidad. También se abordan la garantía de calidad, estándares ISO/IEC, ontologías y la calidad en el uso de datos. Finalmente, el clúster (7-naranja) direcciona sus temas hacia los requisitos funcionales y no funcionales en el diseño y desarrollo de software. Se analizan las especificaciones y el proceso de desarrollo, así como la importancia de cumplir con los requisitos específicos para garantizar la calidad del producto en diversos sistemas. Este análisis revela una rica diversidad de enfoques en la investigación de calidad de datos en sistemas de gestión académica. Los clústeres identificados abarcan temas que van desde la calidad del software y la gestión del mantenimiento hasta el benchmarking de software y la seguridad de datos. La intersección con la atención médica y la informática médica también es evidente, al igual que el énfasis en la reutilización del software para la toma de decisiones y la gestión de datos. Además, se destacan aspectos como la modelación de calidad, la garantía de calidad y la importancia de cumplir con los requisitos específicos para garantizar la calidad del producto en diversos sistemas. Este conocimiento permite a los investigadores y profesionales del campo identificar áreas de interés, colaboración y desarrollo, promoviendo un enfoque holístico hacia la mejora de la calidad de datos en sistemas académicos y de información.

De acuerdo con Surjandy (2022), Bronselaer (2021) & Sadiq, Yeganeh & Indulska (2011) la literatura ofrece diversas aristas que clasifican el corpus de publicaciones, entre aspectos interdisciplinarios; teóricos y prácticos, lo que proporciona un medio para comprender la idoneidad de los temas específicos de la calidad de datos; mejorando la transversalización de los mismos, hacia el desarrollo de soluciones holísticas.

#### **4. Conclusiones**

El presente artículo ha utilizado una metodología bibliométrica basada en la información indexada en Scopus para analizar la producción científica sobre evaluaciones de calidad de datos en sistemas de gestión académica desde 1995 hasta 2023. Sin embargo, ha experimentado un crecimiento significativo desde el año 2006 hasta el año 2022, lo que evidencia el interés y relevancia del tema en la comunidad científica. La investigación revela la complejidad de evaluar la calidad de los datos, con diversos enfoques y modelos en la literatura. Destaca el uso predominante de la familia ISO/IEC 25000 como guía, reflejando una creciente atención a la calidad de datos en sistemas de información. Además, se señala el interés en abordar barreras de accesibilidad, seguridad y cumplimiento de normativas en Instituciones de Educación Superior.

Los autores más productivos en esta área de investigación han centrado sus trabajos en el desarrollo de modelos y herramientas para evaluar la calidad de los productos de software orientados a la salud, utilizando estándares internacionales como la serie ISO/IEC 25000 y 9126.

España, Países Bajos y Canadá son los países con mayor producción de artículos en este campo, destacándose España como uno de los países más citados y con una fuerte presencia en el ámbito de la investigación y desarrollo de tecnología educativa.



Las universidades Mohammed V University in Rabat y Ecole Nationale Supérieure d'Informatique en Marruecos, junto con las universidades de Murcia y Castilla-La Mancha en España, son las instituciones más productivas.

La revista "Software Quality Journal" lidera tanto en productividad como en citas, destacándose como una fuente importante de investigación sobre evaluaciones de calidad de datos en sistemas de gestión académica. Además, las revistas con mayor producción de artículos sobre evaluaciones de calidad de datos están indexadas en el SJR y se encuentran mayormente en los cuartiles superiores, lo que indica su prestigio y relevancia en la comunidad científica.

En cuanto a las redes de coautoría y coocurrencia de palabras clave, han permitido visualizar las relaciones entre los investigadores y las áreas temáticas más relevantes. Se han identificado clústeres de autores que han colaborado en trabajos relacionados con la calidad del software, gestión de mantenimiento, el benchmarking de software y la intersección entre la atención médica y las tecnologías de la información.

Los resultados actuales ofrecen una visión valiosa de la producción científica en el tema, pero se destacan desafíos y oportunidades en la investigación académica. La mayoría de documentos se enfocan en calidad de datos en salud y gestión empresarial, con poca atención a sistemas de gestión académica. Esto sugiere un área poco explorada con potencial para futuras investigaciones.

Como trabajo futuro se propone el desarrollo de nuevos enfoques de evaluación de calidad de datos: que vayan de la mano con los estándares ISO/IEC 25012 y 25040. Estos nuevos enfoques podrían considerar aspectos como la heterogeneidad de datos en el ámbito académico, la inclusión de datos no estructurados y la gestión de la privacidad en el contexto de la calidad de datos. De la misma forma enfocar estudios empíricos y casos de uso en instituciones académicas para evaluar la calidad de los datos en los sistemas de gestión que manejan. Esto permitiría una comprensión más profunda de los problemas específicos que enfrentan, así como identificar prácticas efectivas para mejorar la calidad de los datos.

## Conflicto de Interés

La autora declara no existen conflictos de interés.

## Referencias

- Antonakopoulos, T., Agavanakis, K. & Makios, V. (1995) CASE tools evaluation: an automatic process based on fuzzy sets theory. In: *Proceedings: Sixth IEEE International Workshop on Rapid System Prototyping*, R. Lauwereine (ed.), 140 – 146. IEEE Computer Society Press, New York, NY, USA.
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Arvanitou, E., Ampatzoglou, A., Chatzigeorgiou, A., Avgeriou, P., & Tsiridis, N. (2022). A metric for quantifying the ripple effects among requirements. *Software Quality Journal*, 30(3), 853-883. <https://doi.org/10.1007/s11219-021-09581-y>
- Bambazek, P., Groher, I., & Seyff, N. (2023). Requirements engineering for sustainable software systems: a systematic mapping study. *Requirements Engineering*, 28(1), 481–505. <https://doi.org/10.1007/s00766-023-00402-1>



- Bronselaer, A. (2021). Data quality management: an overview of methods and challenges. In T. Andreassen, G. DeTres, J. Kacprzyk, H. L. Larsen, G. Bordogna, & S. Zadrozny (Eds.), *Flexible query answering systems (FQAS 2021)*, 12871, 127–141. [https://doi.org/10.1007/978-3-030-86967-0\\_10](https://doi.org/10.1007/978-3-030-86967-0_10)
- Cabuya, D., Alvarado, C., Carrascal, R., Escandón, S., Riola, J. & Fajardo-Toro, C. (2022). Ciberseguridad y ciberdefensa marítima: análisis bibliométrico años 1990–2021. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E49), 197-210. Recuperado de: <https://www.proquest.com/openview/67f6e1eb563ceb5e0da444656cb6c800/1.pdf?pq-origsite=gscholar&cbl=1006393>
- Calabrese, J., Esponda, S., Pasini, A. C., Boracchia, M., & Pesado, P. M. (octubre de 2019). Guía para evaluar calidad de datos basada en ISO/IEC 25012. In XXV Congreso Argentino de Ciencias de la Computación CACIC (pp. 694-706), Red de Universidades con Carreras en Informática, Córdoba.
- Carrizo, C., Caro, A., Salgado, C., Sánchez, A., & Peralta, M. (2021). Una Herramienta para la Definición y Especificación Temprana de Requisitos de Calidad de Datos. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E46), 660-674. Recuperado de: <https://www.proquest.com/openview/9835e2c4c8e01a87bbdfd857c1979cae/1?pq-origsite=gscholar&cbl=1006393>
- Dromey, R. (1995). A model for software product quality. *IEEE Transactions on software engineering*, 21(2), 146-162. <https://doi.org/10.1109/32.345830>
- Fritz, E., Montejano, G., & García, P. (2017). Selección de Atributos de Calidad de Datos en Sistemas de Gestión de Aprendizaje bajo la Familia de Normas ISO/IEC 25000. En *XIX Workshop de Investigadores en Ciencias de la Computación WICC* (pp. 627-631), Red de Universidades con Carreras en Informática (RedUNCI). Buenos Aires.
- Gallegos, M., Galarza, J., & Almuiñas, J. (2022). Los sistemas de información como sustento a la gestión de la calidad en las Instituciones de Educación Superior. *Revista San Gregorio*, 1(49), 137-149. <https://doi.org/10.36097/rsan.v0i49.1866>
- Garcés-Giraldo, L., Benjumea-Arias, M., Cardona-Acevedo, S., Valencia-Arias, A., Patiño-Vanegas, C., Padierna, O., & Garay, R. (2022). Análisis bibliométrico en adopción de logística de entrega de mercancías mediante el uso de drones. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E50), 386-397. Recuperado de: <https://www.proquest.com/openview/802b5239fe1d162785aba0e573512898/1.pdf?pq-origsite=gscholar&cbl=1006393>
- Gualo, F., Caballero, I., Piattini, M., Verdugo, J., & Rodríguez, M. (septiembre, 2022). *Medición, evaluación y certificación de calidad de datos en la transformación digital* [Presentación de paper]. En Congreso español de metrología, España.
- Hasanah, N., Atikah, L., & Rochimah, S. (2020, September). Functional Suitability Measurement Based on ISO/IEC 25010 for e-Commerce Website. In *7th International Conference on Information Technology, Computer, and Electrical Engineering ICITACEE 2020 - Proceedings* (pp. 70-75), Semarang, Indonesia. 10.1109/ICITACEE50144.2020.9239194.
- Laaz, I., & López, M. (2020). *Gestión de calidad de datos en los sistemas de información en la Unidad de Tecnología de la ESPAM MFL según la norma ISO 25012* (tesis de pregrado) Escuela Superior Politécnica Agropecuaria de Manabí, Manabí, Ecuador.
- Lizano, H. & Sánchez, P. (2020). Evolución Tecnológica en Turismo: Un Análisis Bibliométrico. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E36), 480-495. Recuperado de: <https://www.proquest.com/openview/851eccdaee34bc546646d07da08ae7e6/1?pq-origsite=gscholar&cbl=1006393>
- Manrique de la Cuadra, A. (2018). *Desarrollo de un catálogo de reglas de negocio referentes a datos, basado en ISO/IEC 25012 y SBVR* (tesis de maestría) Universidad de Castilla – La Mancha, Ciudad Real, España.

- Martínez, M. (2022). *Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales* (tesis doctoral) Universidad Nacional de La Plata, La Plata, Argentina.
- Moumane, K., & Idri, A. (noviembre de 2014). Using ISO 9126 with QoS DiffServ model for evaluating software quality in mobile environments. En *2014 Second World Conference on Complex Systems (WCCS)*. Agadir, Morocco. 10.1109/ICoCS.2014.7060905
- Nandiyanto, A., Al Husaeni, D., & Al Husaeni, D. (2021). A bibliometric analysis of chemical engineering research using vosviewer and its correlation with covid-19 pandemic condition. *Journal of Engineering Science and Technology*, 16(6), 4414-4422. Recuperado de: [https://jestec.taylors.edu.my/Vol%2016%20Issue%206%20December%202021/16\\_6\\_4.pdf](https://jestec.taylors.edu.my/Vol%2016%20Issue%206%20December%202021/16_6_4.pdf)
- Patiño, S., Caicedo, A., & Reina, E. (2019). Modelo de evaluación del Dominio Control de Acceso de la norma ISO 27002 aplicado al proceso de Gestión de Bases de Datos. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E22), 230-241. Recuperado de: <https://www.proquest.com/docview/2317841707>
- Patiño-Toro, O., Rodríguez-Correa, P., Fernández, A., Escorcía-González, J., Valencia-Arias, A. & Jiménez-Guzmán, A. (2022). Los MOOC como Entornos Virtuales Colaborativos: Una aproximación desde un análisis bibliométrico. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E50), 155-168. Recuperado de: <https://www.proquest.com/scholarly-journals/los-mooc-como-entornos-virtuales-colaborativos/docview/2725643374/se-2>
- Peralta, M. (2021). Calidad de Datos en Sistemas de Gestión Académica Universitaria basado en ISO/IEC 25012. *Interfases*, 14(014), 65-88. <https://doi.org/10.26439/interfases2021.n014.5393>
- Peralta, M., Merma, J., Soto C., & Jiménez, W. (2021). Evaluación de la Calidad de Datos en un Sistema de Gestión Académica de una Universidad Peruana Basado en el Estándar ISO/IEC 25000. In *Memorias de la Décima Segunda Conferencia Iberoamericana de Complejidad, Informática y Cibernética en el contexto de the 13th International Multi-Conference on Complexity, Informatics, and Cybernetics, IMCIC 2022*, (pp. 68-73). <https://doi.org/10.54808/CICIC2022.01.68>
- Sadiq, S., Yeganeh, N., & Indulska, M. (2011). 20 years of data quality research: themes, trends and synergies. In *Proceedings of the Twenty-Second Australasian Database Conference-Volume 115*. Australia
- Surjandy, C. (2022). Analysis of Information Quality and Security Factors that Affect the use of Pet Apps during COVID-19. In *2022 2nd International Conference on Information Technology and Education (ICIT & E)*, (49-53). Indonesia. 10.1109/ICITE54466.2022.9759548.
- Valencia-Arias, A., Ocampo-Osorio, C., Quiroz-Fabra, J., Garcés-Giraldo, L., & Valencia, J. (2020). Tendencias investigativas en la aplicación de realidad aumentada en el sector turístico: un análisis bibliométrico. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E36), 229-242. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=8550442>
- Verdugo, J., Rodríguez, M. & Piattini, M. (2014). Using Agile Methods to Implement a Laboratory for Software Product Quality Evaluation. In Cantone, G., Marchesi, M. (eds), *Agile Processes in Software Engineering and Extreme Programming*, XP 2014. Lecture Notes in Business Information Processing, vol 179. Springer, Cham. [https://doi.org/10.1007/978-3-319-06862-6\\_10](https://doi.org/10.1007/978-3-319-06862-6_10).
- Yacelga, N. (2023). *Calidad de datos en la especificación de requerimientos de software* (tesis de maestría) Universidad Técnica de Ambato, Ambato, Ecuador.
- Yanes, J., Sepúlveda, R., & Díaz, H. (2019). La evaluación de la calidad de datos: una aproximación criptográfica. *Computación y Sistemas*, 23(2), 557-568. <https://doi.org/10.13053/cys-23-2-2899>.