

Exploring open government data and big data from a quality perspective.

Exploración sobre datos gubernamentales abiertos y *big data* desde una perspectiva de calidad

Gina Maestre-Góngora¹ , Adriana Rangel-Carrillo² ,
Mariutsi Alexandra Osorio-Sanabria³ 

¹ PhD. En Ingeniería de Sistemas y Computación, Universidad Cooperativa de Colombia, Medellín, Colombia,

² Mg. En Gestión de Tecnologías de Información, Universidad Cooperativa de Colombia, Bucaramanga, Colombia

³ PhD. (c) en Ingeniería, Fundación Universitaria María Cano, Medellín, Colombia

gina.maestre@campusucc.edu.co, adriana.rangel@campusucc.edu.co, mariutsialexandraosoriosanabria@fumc.edu.co

(Received: 13 February 2022; accepted: 27 November 2022; Published online: 30 June 2023)

Abstract. Objective: To establish the key elements in the relationship between Big data and open government data, from a quality perspective. Methodology: The authors conducted an exploratory literature review to determine the lines of quality relationship between big data and open government data, emphasizing Colombia's case. Results: The concept of quality is a common factor for big data and open government data, establishing the benefits, such as innovation, transparency, and economic growth, as well as the challenges related to quality aspects, such as guaranteeing the reliability of the origin of the data, facilitating the understanding of the data, and establishing quality standards. Conclusions: The relationship between big data and open government data from the perspective of data quality allows to exploit the potential immersed in the data, contribute significantly to the construction of knowledge, and thus provide answers to different problems or phenomena of society.

Keywords: Data Quality, Open Government Data, Quality Assurance, Information Technology, Open Government, Smart City.

Resumen. Objetivo: Establecer los elementos clave en la relación entre el Big data y los datos abiertos del gobierno, desde una perspectiva de calidad. Metodología: Los autores realizaron una revisión bibliográfica exploratoria para determinar las líneas de relación de calidad entre Big data y datos de gobierno abierto, haciendo énfasis en el caso de Colombia. Resultados: El concepto de calidad es un factor común para Big data y los datos abiertos de gobierno, estableciendo los beneficios, como la innovación, la transparencia y el crecimiento económico, así como los retos relacionados con los aspectos de calidad, como garantizar la fiabilidad del origen de los datos, facilitar la comprensión de los mismos y establecer estándares de calidad. Conclusiones: La relación entre el Big data y los datos abiertos de gobierno desde la perspectiva de la calidad de los datos permite aprovechar el potencial inmerso en los datos, contribuir significativamente a la construcción de conocimiento, y así dar respuestas a diferentes problemas o fenómenos de la sociedad.

Palabras clave: Calidad de los datos, datos del gobierno abierto, garantía de calidad, tecnología de la información, gobierno abierto, ciudad inteligente.

Paper Type: Review Paper.

1 Introduction

In the last decade, open government policies have encouraged national and local governments to provide data in an open format to generate social, political, or economic value (Osorio-Sanabria et al., 2020, p. 157). This type of data is called open government data (OGD), which can be published through repositories or web portals, facilitating access and reuse by citizens in different contexts (Maestre-Gongora et al., El valor

de los datos abiertos de gobierno: un enfoque desde la evaluación de calidad, 2021a; Mahecha Moyano *et al.*, 2017).

On the other hand, through digital government policies and technological advances of the Fourth Industrial Revolution (4IR), the development of Big Data strategies has been proposed to the public sector to improve and increase the use and exploitation of data generated by public and private organizations, citizens, academia and digital devices. Considering Big Data like a type of data that refers to data sets that have characteristics such as volume, variety, velocity, viscosity, volatility, and veracity (Rodríguez *et al.*, 2017).

To ensure the proper use, combination, and exploitation of OGD and Big Data, it is necessary to establish whether the data present quality problems. Sadiq and Indulska (2017) affirm that the data quality uncertainty limits the value that can be obtained from such data. Open data, on the other hand, is frequently unfamiliar to the user and may lack metadata (Sadiq & Indulska, 2017). It is necessary to consider that low-quality data provision increases the cost (in its wider meaning) of accessing and interpreting data (Vetrò *et al.*, 2016). To this end, their evaluation must be conducted from the perspective of confidence and usefulness. As a result, data quality allows us to take a large, perhaps shapeless, set of data and put them in proper order, enabling users to discover patterns, new ideas, and trends (BSA, 2022).

However, the crisis related to the quality of open data for big data analysis is a factor that generates resistance to its use. Identifying, finding, and correcting errors in a large amount and variety of data and tracking down confirmatory sources for unreliable data account for about 50% of the time of spending on the big data process and 60% of time cleaning and organizing. Furthermore, the total cost corresponds to 75% of the budget planned for data analysis (Redman, 2016). Against this backdrop, identifying relevant factors and how they are related to the evaluation of the OGD quality is crucial. This will facilitate big data analysis as a strategic initiative to enhance data capacity and satisfy the needs according to the specific conditions of use (ISO/IEC, 2014). But data quality problems are becoming more acute. To this end, focusing efforts on assessing and ensuring quality to maximize the impact on data use and exploitation is essential.

In this scenario, the Colombian Government, within the framework of the national data exploitation policy (Big Data), raised as one of its objectives to increase the availability and use of OGD for the generation of social and economic value, given the low levels of open data (DNP, MinTIC, SIC, 2018). Additionally, the diagnosis and planning of the action plan for the opening of data (Gobierno de la República de Colombia, World Bank Group, 2016) identified that there are barriers that affect the publication of OGD, such as the absence of a culture of data in the institutions and of systematic processes to determine the quality of the data, as well as, the low technical preparation of those responsible for the data to carry out the said evaluation.

Currently, the use and combination of OGD and Big Data have the potential to support the public sector in "improving decision-making, knowledge generation, and process optimization" (Vélez *et al.*, 2022). For this reason, studies have been conducted that address the relationship between OGD and Big Data from different perspectives. Fraefel, Haller, and Gschwend (2017) discuss through use cases in the cities of Amsterdam, Sapporo, and Tokyo the challenge of the quality of OGD obtained from various sources and subsequently analyzed through Big Data. Faini and Palmirani (2016) describe the scopes and differences between open data and Big Data and how they are approached from Italian Law, furthermore, they analyze how these types of data can be used and combined to promote open government in Italy. Nafis *et al.* (2016) propose an architecture that allows the use of Big Data aligned with open data standards to improve the performance of public services.

Although these studies focus on aspects associated with OGD and Big Data, a study that establishes key aspects of data quality between these types of data has not been conducted. It is because of the above that, in this research, a literature review and content analysis of research that will present theories and case studies, regulations, and technical documents related to OGD, Big Data and data quality during the period from 2010 to 2020 was conducted. This analysis identified the economic, strategic, and technical effects of low data quality in Big Data, the quality attributes applicable to these two types of data, the challenges and benefits of quality assessment, and the implications for theory and practice in this area.

This paper is structured as follows: Section 2 presents the methodological approach. Section 3 reports and describes and 4 discusses some relevant findings, emphasizing the following three questions: What are the critical elements of the relationship among big data, OGD, and data quality assurance? What primary data quality attributes apply to OGD and big data? What are the benefits and challenges of OGD quality assessments? The section. In section 4, we propose a conceptual model, which reflects findings that show the importance of ensuring data quality for big data and OGD through principles, guidelines, and

interrelated attributes, as a point of convergence. Section Finally, Sections 5 present conclusions and future work.

2 Methodology

There is an increase in research related to data quality assurance, OGD, and Big Data to conceptualize relevant aspects, therefore, it becomes essential to present an overview of the relationships between these types of data, the data quality attributes, and challenges and benefits of data quality of OGD posed in those proposals. The existing literature is analyzed through the systematic literature review method with the objective of answering three research questions. The review process was conducted in three stages: (1) Defining the scope and planning of the review, (2) selection of documents, and (3) analysis and synthesis of the results. The phases of the study are described below.

2.1 Defining the scope and planning of the review

This study proposes a taxonomy established for literature reviews presented by (Cooper, 1988) in which the objective, the organization of the information, perspective assumed, audience, and coverage of the review are defined. Common conceptual aspects based on guiding questions were considered to organize information, focusing on theories and case studies. This review's target audience is general and academics with interest in the topic of OGD, as shown in Table 1.

Table 1. Taxonomy of Literature Reviews

| Characteristic | Categories | | | |
|----------------|------------------------|--------------------------------|---|----------------------------------|
| | Results | Research Methods | Theories | Practices or Applications |
| Objective | Integration | Critical | | Identification of Central Issues |
| Organization | Chronological | Conceptual | | Conceptual |
| Perspective | Neutral Representation | | Espousal of Position: Data Quality | |
| Audience | General Academic | Practitioners or Policy Makers | | General Public |
| Coverage | Comprehensive | Comprehensive and Selective | Representative | Central |
| | | | Chosen for review | |

Source: Cooper (1988)

This review aims to answer three research questions that fall under three criteria:

- What are the critical elements of the relationship between big data, OGD, and data quality assurance?
- What primary data quality attributes apply to OGD and big data?
- What are the benefits and challenges of OGD quality assessments?

2.2 Document Selection

In the selection of research published in the indexed scientific press, technical documents, or regulations to collect data and answer the research questions, a review protocol was defined that contemplated the search strategy, the search terms in each data source, and the inclusion and exclusion criteria.

2.2.1 Search strategy

The search strategy considered the review of data needed to answer each research question. In order to collect data as completely as possible, a search string was constructed with an appropriate set of keywords,

the most widely used electronic sources, and search filters were established. Table 2 shows the sources consulted, the keywords, and the filters applied.

Table 2. Search Strategy

| Sources | Keywords | Filters |
|---|--|---|
| IEEE Xplore Digital Library Springer Link Google Scholar Proquest Computing Database Scopus | Open government data quality Big data quality Open government data Quality methodology Methodology | The authors applied the search string to the “Abstract,” “Title,” and “Keywords” metadata in the selected sources. Period: 2010 to 2020. |

Source: Authors’ own elaboration.

2.2.2 Inclusion and exclusion criteria

Inclusion and exclusion criteria were defined to filter relevant studies to answer the research questions. We considered the following inclusion criteria:

- Includes OGD quality, Big Data quality, and quality methodology Conceptualization: The abstract refers to the quality of OGD or Big Data or describes a methodology for assessing data quality.
- Documents published such as document type: journal, technical reports, and conferences.
- Documents published between 2010 and 2020.

We considered exclusion criteria, as applied to the titles and abstracts of the documents, were:

- Duplicate studies.
- Article in a language other than English and Spanish.
- Research focus unrelated to OGD quality or Big Data quality.
- Research focus related to OGD quality or Big Data quality, but insufficient information.

2.2.3 Studies Selection

The search results were integrated into a spreadsheet shared on Google Drive. This file was sorted by title to identify duplicate studies and choose the study to be found mainly in the SCOPUS or SpringerLink database, given that the authors have access to these databases. Subsequently, the exclusion criteria were applied and the title and abstract were reviewed to identify contributions to the research. The identification of duplicates and the review and selection of studies were performed by three reviewers, who reported their concept of inclusion or not of each document in the spreadsheet shared online. If all reviewers had the same decision on whether or not to include a document, it was considered definitive. In cases where there was no agreement, the document was read again to give a new concept. The study selection process is shown in Figure 1. The final list of the reviewers' consensus is presented in Table 3, which included 42 documents focusing on Big Data and OGD. We include a dataset to information about the documents and the analysis categories (Maestre-Gongora *et al.*, 2023).

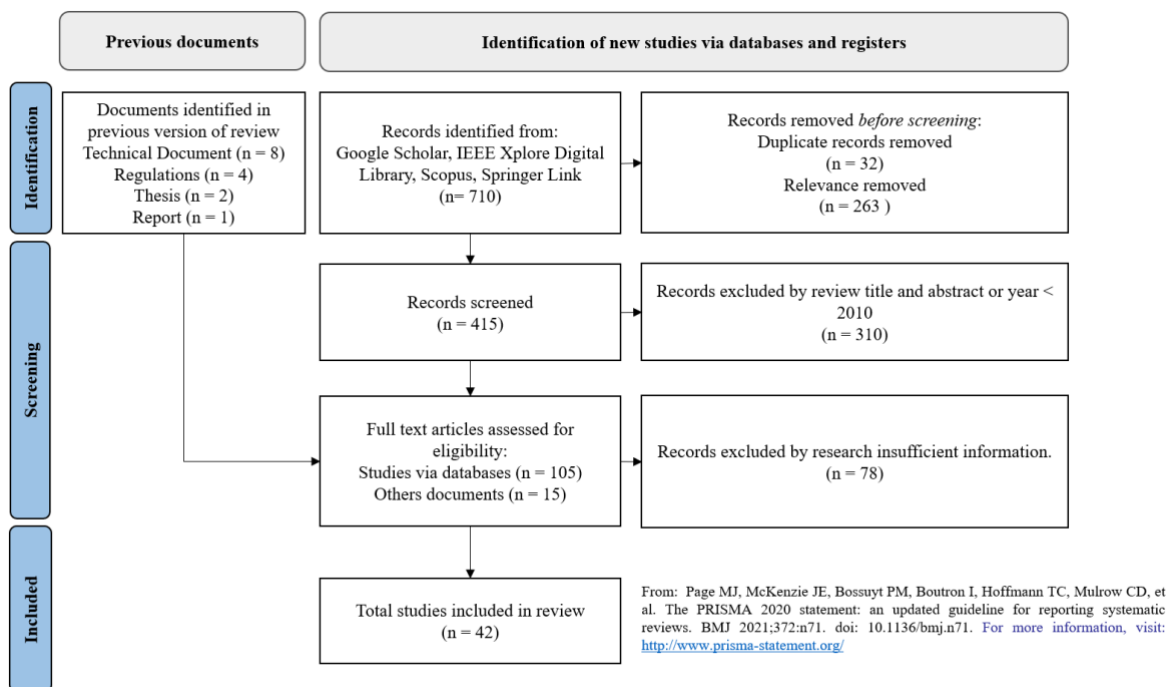


Figure 1. Flow diagram PRISMA methodology (Page et al., 2021)

Table 3. Documents selected for review.

| Document | Type | Year | Language | Country | Approach | |
|--|--------------------|------|----------|-----------|----------|-----|
| | | | | | BD | ODG |
| A systematic review of open government data initiatives | Article | 2015 | English | Germany | | X |
| Value Creation on Open Government Data | Article | 2016 | English | Germany | | X |
| Open Data Kit | Technical Document | 2016 | Spanish | Argentina | | X |
| Data Quality Management for Industry 4.0: A Survey | Article | 2020 | English | ASQ | X | |
| Predictive analytics: driving improvement from data | Article | 2016 | Spanish | IDB | X | X |
| Acceptance, Quality, and Trust Factors - Conceptual Model for Open Government Data Potential Use | Article | 2019 | English | Malaysia | X | X |
| Open data in Latin America and the Caribbean | Article | 2018 | Spanish | IDB | | X |
| From information to intelligence: How to adapt institutions for data analysis in government? | Article | 2019 | Spanish | IDB | | X |
| The Challenges of Data Quality and Data Quality Assessment in the Big Data Era | Article | 2015 | English | China | X | |
| Assessing data quality in open data: A case study | Article | 2018 | English | Colombia | | X |
| Big data in the public sector | Article | 2016 | English | Spain | X | X |
| Law 1712 of 2014 | Regulations | 2014 | Spanish | Colombia | | X |
| Guide to Open Data in Colombia | Technical Document | 2016 | Spanish | Colombia | | X |
| Guide to Quality and Interoperability standards for open data of the Government of Colombia. | Technical Document | 2016 | Spanish | Colombia | | X |

| Document | Type | Year | Language | Country | Approach | |
|---|--------------------|------|------------|---|----------|-----|
| | | | | | BD | ODG |
| National Data Exploitation Policy (Big Data) | Regulations | 2018 | Spanish | Colombia | X | X |
| Open Data - Open Government | Technical Document | 2020 | Spanish | Costa Rica | | X |
| Open Data Colombia | Technical Document | 2019 | Spanish | Colombia | | X |
| Guide for the use and exploitation of Open Data in Colombia | Technical Document | 2019 | Spanish | Colombia | | X |
| Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh. | Article | 2019 | English | China | | X |
| Open Data Quality Requirements | Technical Document | 2019 | Spanish | Colombia | | X |
| Exploring the quality of government open data Comparison study of the UK, the USA, and Korea | Article | 2018 | English | South Korea | | X |
| A Suggested Framework for the Quality of Big Data Deliverables of the UNECE Big Data Quality Task Team | Report | 2014 | English | United Nations Economic Commission for Europe | X | |
| A Data Quality in Use model for Big Data | Article | 2016 | English | Spain | X | |
| Why is data so important? | Article | 2017 | Spanish | United States | X | |
| Open government data: A stage model | Article | 2011 | English | Greece | X | X |
| Participation and Data Quality in Open Data use: Open Data Infrastructures Evaluated | Article | 2015 | English | Netherlands | | X |
| Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators | Article | 2016 | English | Netherlands | | X |
| A Process Pattern Model for Tackling and Improving Big Data Quality | Article | 2018 | English | England/The Netherlands | X | |
| Big data curation | Article | 2016 | English | Ireland | X | |
| Big Data Quality: A Roadmap for Open Data | Article | 2016 | English | Italy | X | |
| Government Open Data Portals: A Measurement of Data Veracity Coverage | Article | 2019 | English | Malaysia | X | X |
| Quality Improvement in Open Data sets based on the application of Logical Consistency Metrics. | Thesis | 2017 | Spanish | Colombia | | X |
| Open Data: Barriers, Risks, and Opportunities | Article | 2012 | English | Luxembourg | | X |
| International Open Data Charter (2015) | Regulations | 2015 | Spanish | Mexico | | X |
| Open Data Policy Implementation Guide | Technical Document | 2017 | Spanish | Mexico | | X |
| Open government data: Facilitating and motivating factors for coping with potential barriers in the Brazilian context | Article | 2014 | English | Brazil | | X |
| Challenges to Use Open Government Data through the Citizens' Lens: A Systematic Review. | Article | 2019 | English | Pakistan | | X |
| Open Data Plan 2020–2021 | Regulations | 2020 | Portuguese | Brazil | | X |
| Barriers to using open government data | Article | 2019 | English | Poland | X | X |
| Proposal for the Evaluation of Open Data Portals | Article | 2019 | English | Colombia | | X |
| Proposing a framework of barriers to opening government data in China A critical literature review | Article | 2017 | English | China | | X |

| Document | Type | Year | Language | Country | Approach | |
|--|---------|------|----------|-------------------------|----------|-----|
| | | | | | BD | ODG |
| Prototype software for open data quality assessment. | Thesis | 2017 | Spanish | Colombia | X | X |
| Publishing open data considering quality criteria | Article | 2019 | Spanish | Spain/Ecuador | | X |
| Quality of Open Research Data: Values, Convergences, and Governance. | Article | 2020 | English | Hungary | X | |
| Benefits and Risks of Open Government Data | Article | 2014 | English | Prague - Czech Republic | X | X |
| Open Government Data Publication Methodology | Article | 2015 | English | Prague - Czech Republic | | X |
| Innovation in government services: The case of open data | Article | 2013 | English | Singapore | | X |

Source: Authors' own elaboration.

2.3 Analysis and Synthesis

The content analysis of the studies for data collection was performed using a manual procedure, which included:

- The review of each paper by the three authors to extract data and information separately from each article, which were reported in the online file.
- Two of the authors independently evaluated the extracted data to build through a matrix classification of the aspects associated with the research, by the emphasis of the study (OGD, Big Data, Open data), the application sector (public, private, research, statistics), by quality effects (economic, technical or strategic) and by characteristics, attributes, and data quality criteria.
- Three authors classified the extracted data by marking with an X the relationship between the study and the aspects to be classified.
- The authors compared their responses to reach a consensus on the contributions of each study.

Of the selected studies (Table 3) we observed that are Article 68,09%, Regulations 8,51%, Report 2,13%, Technical Document, 17,02%, and Thesis 4,26%, considering documents from countries such as Germany, China, Spain, Greece, Holland, England, Ireland, Italy, Malaysia, Pakistan, Poland, the Czech Republic, and Singapore. Colombian publications accounted for 23% of the total number of documents and served as a reference. The authors have dedicated considerable time to conducting research into emerging information technologies. While for the study of the conceptual lines on Big Data and Open Government Data, 90% of the selected literature is focused on OGD, 21% reveals the relationship of big data with OGD, and 42% concentrates only on big data.

Additionally, 10% of the records corresponded to publications released by multilateral organizations, such as the Inter-American Development Bank, United Nations Economic Commission for Europe (UNECE), and American Society for Quality. The following sections present the most relevant ideas and conclusions that the authors drew from the review.

3 Results and Discussion

3.1 Relationship between Big Data, OGD, and Data Quality

The opportunities that the information age and the Internet have generated technological advances in artificial intelligence, the Internet of Things, robotics, nanotechnology, biotechnology, and cloud computing. The above technologies have triggered an exponential increase in data diversity and structure (or lack of structure) and data accumulation at astounding speeds.

The concept of quality has evolved, starting from the inspection applied to products to the quality control for processes, to quality assurance in systems, to total quality, to the total quality management or excellence (Torres Saumeth et al., 2012). As a result, data quality faces multiple challenges due to its particular

characteristics; these challenges encompass the following: (i) Volume and speed: The volume of data is too large and exponentially grows in a short time (PowerData, 2017). (ii) Variety: There are many data sources, types, and structures. About 20% of data are structured; this means that working with 80% of unstructured data is necessary (Cai & Zhu, 2015). (iii) Veracity and value challenges: The accelerated change of data renders data validity and value opportunities very limited. For example, customer data usually lose validity by 2% per month and 25% per year (Cai & Zhu, 2015). (iv) High reuse rate: The same data can have different meanings depending on the context they are used, raising questions about their validity and consistency (Loshin, 2014). (v) Balance monitoring: Data consumers are faced with balancing two undesirable aspects of data monitoring, namely, correcting the data and maintaining consistency with the source at the expense of its quality for use (Loshin, 2014). (vi) Data rejuvenation: Continuously aligning old data with newer datasets is necessary; in this sense, increasing the level of data validation and governance capacity is essential (Loshin, 2014).

Large amounts of time and money are invested in customizations and traditional methods to put out fires and dampen crises. In the short term, crises arise due to lack of data quality. However, there is no investment in addressing long-term problems, and most of the budget allocated to data analysis (75%) is used to adapt data (Redman, 2016).

In the realm of data, some are public; that is, anyone can freely use and reuse them. These data are called open data and are considered key factors for global transformation, enabling governments, private and public companies, and citizens to have clear and reliable information to make decisions (MinTIC, 2021). The objective of the OGD is to open up the data, allowing the academic community and business sectors to reuse them. Such openness facilitates cooperation between public administration, politics, industry, and citizens, strengthening transparency, democracy, participation, and collaborative work (Kucera, 2015). OGD are based on a set of principles that support their use, access, and reuse. Thus, data must be machine-processable, accessible, complete, obtained from the source of origin, nonproprietary, timely, up-to-date, openly licensed, and available to anyone without discrimination (Kucera, 2015).

The value of OGD can be harnessed if their quality is guaranteed as OGD must facilitate analysis for decision-making, obtain contextual responses, and contribute to innovation trends through its qualities, attributes, and predictive potential (BSA, 2022). To embrace big data as a renewable resource and ensure its quality produces economic and social benefits, the government must obtain more relevant results and generate responses that could not be foreseen at the time of data creation (BSA, 2022).

3.2 Data Quality Attributes Applicable to OGD and Big Data

Data are the raw material required for developing different activities of humanity. Data have attributes and dimensions that can be evaluated to determine the quality level and enable their use depending on the context.

The following principles are used to ensure data quality (Cai & Zhu, 2015): Data represent a unit of value and cost impact for interested customers who view them as a product; data have a quality level, which results from the type of process applied to their generation and processing; several factors influence data quality, such as the purpose of their generation, their potential use, their timeliness and validity, and interested users.

To ensure compliance, standards that involve data quality management and assurance (Caro *et al.*, 2013) have been developed. An example is the standard ISO/IEC 25012 (2022), which defines two relevant points regarding data quality:

- **Inherent Quality:** It highlights the intrinsic potential that data have to satisfy and the implicit and explicit needs of its creation purpose, determining the value of such data when their quality characteristics or dimensions meet the defined requirements: accuracy, completeness, consistency, credibility, accessibility, compliance, confidentiality, efficiency, traceability, and comprehensibility.
- **System-Dependent Quality:** It corresponds to the dependence that data have on the technological domain used: accessibility, compliance, confidentiality, efficiency, traceability, comprehensibility, availability, portability, and recoverability.

Table 4 classifies the quality characteristics or dimensions for the two points of view:

Table 4. ISO/IEC 25012 Data Quality Model.

| Features | Data Quality Viewpoints | |
|-------------------|-------------------------|------------------|
| | Inherent | System-Dependent |
| Accuracy | X | |
| Completeness | X | |
| Consistency | X | |
| Credibility | X | |
| Timeliness | X | |
| Accessibility | X | X |
| Compliance | X | X |
| Confidentiality | X | X |
| Efficiency | X | X |
| Accuracy | X | X |
| Traceability | X | X |
| Comprehensibility | X | X |
| Availability | | X |
| Portability | | X |
| Retrievability | | X |

Source: Standard ISO/IEC 25012 (2022)

The characteristics or dimensions of quality include additional factors of big data related to their processing criteria, called “V” factors, namely, volume, speed, variety, value, variability, veracity, and visibility (Mukherjee & Shaw, 2016).

These factors provide characteristics, such as size and structure, that generally exceed the capabilities of acquisition, storage, and processing of data on time, through traditional methods and tools, and also exceed the perception capacity of human beings (Miloslavskaya & Tolstoy, 2016).

To use and exploit open data, in the Colombian case (MinTIC, 2019a) the government defines processes for managing open data, emphasizing quality monitoring through quality, usage, and impact measurements (Osorio-Sanabria et al., 2020). Colombia updated the quality requirements for OGD to seal data quality in public institutions, aligned with the life cycle of data. The guide mentioned sets the stage for data preparation, access, use, evaluation, and improvement that are necessary to leverage the value that open data hold (MinTIC, 2019a). At each stage, a set of quality requirements is verified to obtain the “Seal of Excellence” for open data, such as identifying the dataset to be published, prioritizing the data to be published, ensuring the completeness of the data, ensuring data availability, ensuring data timeliness, and documenting the data through metadata (MinTIC, 2019b).

Although the country has defined guidelines to advance in the opening and use of open data, these are assumed in an isolated manner and there are few integrated methodologies that allow the IT offices of public entities to approach the process by clearly identifying tools and instruments to generate quality open data that facilitate the implementation process of open government data towards a quality approach (Maestre-Góngora et al., 2021b).

To guarantee the quality of OGD in Colombia, the government has established 16 quality criteria as standards (Table 5): confidentiality, relevance, up-to-date, traceability, compliance, accuracy, completeness, consistency, accuracy, portability, credibility, comprehensibility, accessibility, efficiency, recoverability, and availability (MinTIC, 2023).

Table 5. Quality Criteria Standard in Colombia.

| Criteria | Interoperability Guidance |
|-----------------|---|
| Confidentiality | Data should only be accessed by authorized persons to protect reserved and classified information. |
| Relevance | Published data must be useful for users. In this concept, data have more or less relevance according to the case of analysis. |
| Currency | Currency and timeliness of the published data. |
| Traceability | History of the available dataset: dates of creation, publication, and updates. |

| Criteria | Interoperability Guidance |
|-------------------|--|
| Compliance | Compliance with current guidelines and standards such as DCAT (Data Catalog), for the description of the Metadata. |
| Accuracy | Correctly filled out data. |
| Completeness | Complete information, complete data for all columns |
| Consistency | Coherent data, free of contradictions. |
| Accuracy | Level of disaggregation at which the data are published concerning the level of disaggregation at which they were generated |
| Portability | Unrestricted formats for data reuse |
| Credibility | Truthful and reliable information for users. |
| Comprehensibility | Characteristics that allow the user to read and interpret the data. |
| Accessibility | Technological tools that guarantee access to the user who requires it |
| Efficiency | Ability of the platform to perform data analysis and downloading with expected performance levels and times. |
| Recoverability | Attributes that allow maintaining and preserving a specific level of operations and quality. |
| Availability | Ensures that authorized users have access to information and other associated information assets at the required place and time. |

Source: Ministerio de Tecnologías de la Información y las Comunicaciones (2023)

Public offices produce and manage a large amount of data, in various formats, at high velocity and variety, which are published to facilitate access to the general public and their reuse in different contexts. In Ciancarini et al. (2016) it is defined that open data is a subdomain of Big Data, and that data quality issues are aggravated in Big Data as entities are having a growth in data generation and management; therefore, it becomes essential to focus efforts on assessing and ensuring the quality of this representative subdomain. Table 6 shows the intersection between big data and OGD quality dimensions. According to the literature review, the authors conducted studies into quality attributes.

Table 6. Quality attributes for big data and OGD.

| Quality Attributes | Open Data | Big Data | Authors |
|--------------------|-----------|----------|---|
| Accessibility | X | X | Attard et al. (2015); Cai & Zhu (2015); Koltay (2020); Merino et al. (2016); MinTIC (2018); United Nations Economic Commission for Europe [UNECE] (2015); Wahyudi et al. (2018); Williams & Tang (2020) |
| Update | X | | MinTIC (2018); Yi (2019) |
| Aptitude | | X | Cai & Zhu (2015); Merino et al. (2016); MinTIC (2018); Yi (2019) |
| Coherence | | X | Attard et al. (2015); Merino et al. (2016); MinTIC (2018); UNECE (2015) |
| Completeness | X | X | Cai & Zhu (2015); Koltay (2020); Merino et al. (2016); Wahyudi et al. (2018); Williams & Tang (2020); UNECE (2015) |
| Understandable | X | X | Koltay (2020); MinTIC (2018); Wahyudi et al. (2018); Yi (2019) |
| Concise | | X | Koltay (2020); Wahyudi et al. (2018) |
| Compliance | X | | MinTIC (2018); Yi (2019) |
| Consistency | X | X | Cai & Zhu (2015); Koltay (2020); Talukder et al. (2019); Wahyudi et al. (2018); Williams & Tang (2020) |
| Credibility | | X | Cai & Zhu (2015); Koltay (2020); Merino et al. (2016); Wahyudi et al. (2018) |
| Availability | | X | Cai & Zhu (2015); Merino et al. (2016) |
| Accuracy | X | X | Attard et al. (2015); Cai & Zhu (2015); Koltay (2020); Merino et al. (2016); MinTIC (2018); UNECE (2015); Wahyudi et al. (2018); Yi (2019) |
| Integrity | X | X | Cai & Zhu (2015); Talukder et al. (2019); Williams & Tang (2020); Yi (2019) |
| Interpretable | | X | Koltay (2020); Wahyudi et al. (2018) |
| Readable | | X | Cai & Zhu (2015); Williams & Tang (2020) |
| Objectivity | | X | Koltay (2020); Wahyudi et al. (2018) |

| Quality Attributes | Open Data | Big Data | Authors |
|--------------------|-----------|----------|---|
| Opportunity | | X | Cai & Zhu (2015); Koltay (2020); Merino et al. (2016); UNECE (2015); Wahyudi et al. (2018); Williams & Tang (2020) |
| Portability | | X | Merino et al. (2016); MinTIC (2018); Talukder et al. (2019); Williams & Tang (2020) |
| Accuracy | X | X | Merino et al. (2016); Williams & Tang (2020) |
| Security | | X | Cai & Zhu (2015); Koltay (2020); Merino et al. (2016); UNECE (2015); Wahyudi et al. (2018) |
| Relevance | X | X | Cai & Zhu (2015); Koltay (2020); Merino et al. (2016); MinTIC (2018); Talukder et al. (2019); UNECE (2015); Wahyudi et al. (2018) |
| Reputation | | X | Wahyudi et al. (2018); Yi (2019) |
| Traceability | X | | MinTIC (2018); Yi (2019) |
| Usability | | X | Cai & Zhu (2015); UNECE (2015) |
| Added Value | | X | Koltay (2020); Wahyudi et al. (2018) |

Source: Authors' elaboration

A total of 22 documents are related to big data, and 11 documents are related to OGD quality; 8 of these attributes (i.e., accessibility, completeness, comprehensibility, consistency, accuracy, completeness, precision, and relevance) are considered in both application domains.

3.3 Data Quality Attributes Applicable to OGD and Big Data

The challenges of OGD are mostly related to the data consumer perspective, such as facilitating data discovery and promoting data use. Predominantly, challenges related to quality aspects include the following: ensuring the data reliability and facilitating data understanding (Attard et al., 2015; Cai & Zhu, 2015; Kalampokis et al., 2011; Merino et al., 2016; Wahyudi et al., 2018), promoting the use of the data with quality standards (Attard et al., 2015; Kucera, 2015; Martin et al., 2013; Talukder et al., 2019; Warraich et al., 2019; Cai & Zhu, 2015), unifying and updating data to improve accuracy for use (MinTIC, 2019a), guaranteeing not only data quality but also a reliable operating platform and emerging technology (Munné, 2016), and ensuring data quality before being published (MinTIC, 2019a).

The benefits related to data quality aspects are the following: creating opportunities in society by increasing data capacity to predict and solve problems (Maestre Góngora & Nieto Bernal, 2015), allowing the reuse of both public and private data in various contexts (Munné, 2016), creating strategic advantages and promoting collaborative work between the state and society to improve government entities' processes and exploit the inherent value and potential of data in the economic, social, political, cultural, environmental, strategic, ideological, and other fields of society (Munné, 2016).

4 Implications for Theory and Practice

Quality assurance is fundamental to leveraging the value of big data because it requires reliable information to make timely decisions through efficient processes that facilitate the conversion of large data volumes agilely. Among the main processes of big data are data management and data analysis. Data management implies a series of steps and involves technologies to acquire, store, prepare, and retrieve data as input to perform data analysis. Data analysis involves techniques for analyzing and extracting intelligence from big data (Gandomi & Haider, 2015). The rule is that the larger the sample of data, the more accurate the statistics and other analysis techniques (Russom, 2011).

The quality of OGD is approached from an interdisciplinary approach, which is based on the suitability for use (Attard et al., 2016). In turn, use is diversified into various possibilities and contexts, from an objective perspective, employing agreed-upon attributes and a subjective perspective linked to data consumers' requirements and experience. The suitability of OGD for reuse is a cornerstone of data quality (Yi, 2019); however, quality can vary widely, leading to inconsistencies in terms of datasets and lack of granularity (Warraich et al., 2019; Zuiderwijk et al., 2016), thus impeding the use of the data and portals where they are published.

Open Data allows to generate public, social, and political value to improve transparency and oversight practices from citizens, thus helping people to take a more active role in the dynamics of the territories, by allowing access and use of information for individual and community decision making (Maestre Góngora & Nieto Bernal, 2015). Alternatively, they facilitate product innovation and the creation of new business models and highlight the need to improve government management when necessary. They allow planning future scenarios for society based on decisions, programs, and defined policies (MinTIC, 2021).

Government data are inherently involved in the analysis process that needs to be conducted before taking objective decisions. They also help obtain contextual answers and contribute to innovation trends through their qualities, attributes, and predictive potential (BSA, 2022). Data have transformed different research fields and benefited daily life (e.g., climate forecasts, vehicle flow, and solving more fundamental challenges). The negative impact of having large amounts of data with low levels of quality extends not only to the use itself but also to its potential use, i.e. the suitability of the data for future use and its potential for response in different contexts and decision making in the medium and long term.

It is important to generate methodologies that focus on assessing the quality of OGD related to big data analysis. Consequently, this fosters new research that broadens its horizons due to the disruptive advance that has originated new technologies and the most marked trends (e.g., artificial intelligence, the Internet of Things, and cloud computing).

Based on the literature review, we present a conceptual model for OGD quality assessment with principles, guidelines, dimensions, and attributes. Three principles and dimensions are identified. The dependency hierarchy is established to detail the guidelines through 12 related quality attributes as shown Figure 2. Overall, this model provides a vision of the required objectives that facilitate the evaluation of OGD from a quality perspective.

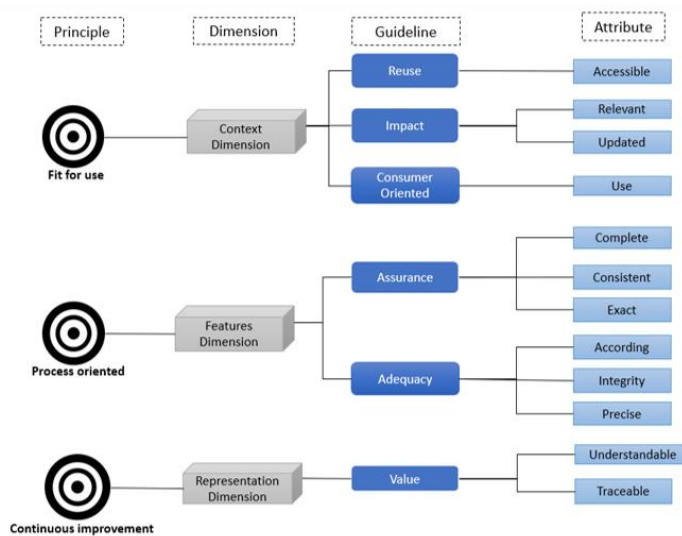


Figure 2. Conceptual Model OGD and Data Quality

In terms of principles, *fitness for use* ensures the data's ability to serve a defined purpose. The *process-oriented* principle manages data quality through processes that effectively implement the quality strategy, with interrelated activities for a systematic execution and *continuous improvement* that ensures data quality through planning, execution of concrete actions, and periodic verification.

About the dimensions, the *data context* corresponds to the characteristics of the origin and use to further understanding and applications; *data characteristics* describes inherent properties, which help define, interpret, and predict their use; and *data representation* corresponds to its context of the origin or in the possible contexts of reuse.

The following guidelines are proposed: *Assurance* focuses on planned and systematic activities that help data inherent properties achieve quality requirements. *Adequacy* refers to conditioning the data to meet quality requirements related to their inherent properties. *Reuse* refers to the potential use of the data in various social, economic, political, and academic contexts. *The impact* is directed toward generating significant and positive changes for consumers, according to data in a specific context, consumer-oriented,

which facilitates the data consumer contextual knowledge and access to data. Finally, *value* ensures that data are represented as a unit of value for consumers.

Within the objectives of open-data initiatives worldwide, reuse helps maximize public value generation, which is enabled by publishing high-quality data and ensuring that they are transformed into products, services, or new applications. The publication of OGD with a quality approach facilitates data exploitation by its consumers, generating a great impact both at the individual, collective, and organizational levels. Consumers play an essential role in reusing data and assessing the impact generated by the data. There is no point in having large amounts of data available if they cannot be exploited, i.e., discovered, and used, and if the value embedded in them cannot be analyzed and discovered (Osorio-Sanabria et al., 2020).

Governments with active OGD initiatives must guarantee data reliability, promote collaboration to generate public value (Munte-Kunigami & Serale, 2018), and ensure data availability and use (MinTIC, 2019b). All these can only be achieved by maintaining an adequate level of quality. This should help prepare data for processing and reuse (MinTIC, 2019b). A challenge faced by data producers is to maintain adequate and up-to-date datasets (Munte-Kunigami & Serale, 2018) to facilitate interpretation in the corresponding context (Martin et al., 2013) and enable a proper understanding of the data (Kucera & Chlapek, 2014).

Based on this review, once the key elements for an open-data evaluation model from a quality approach have been identified, methodologies and an accompanying strategy can be proposed to implement an OGD quality evaluation methodology in public entities, based on data use and information exploration with potential value for third parties.

5 Conclusion

The data and the different ways that they are produced and consumed pose quality challenges related to the inherent characteristics of use and value generation. From the perspective of data quality, the relationship between big data and OGD allows to exploit data potential, significantly furthers knowledge, and provides answers to different problems.

Big data and OGD that have an adequate quality level are resources that generate strategic value by generating new knowledge and facilitating decision-making. OGD quality is a component that has evolved from being a publication requirement to becoming a process inherent to the data, through which the benefits of exploiting its properties for impactful uses that add value to consumers are enabled.

An OGD conceptual model provides a perspective to establish strategic data publishing opportunities with quality to support open government policies and motivate open data exploitation by communities, citizens, academia, and businesses. OGD quality assessment is a process that requires identifying and conditioning relevant factors that facilitate Big Data analysis as a strategic initiative, to enhance the ability of the data to meet stated and implied needs according to specific conditions of use.

Future work will focus on validating the proposed model with at least two public entities of the Colombian State to identify strengths and weaknesses. Then, it is proposed to conduct the methodological proposal for evaluating open data modeled in Business Process Model and Notation, which includes four phases: plan and obtain, structure, and publish, promote use, and monitor for their application in a case study.

ORCID iD

Gina Maestre-Góngora  <https://orcid.org/0000-0002-2880-9245>

Adriana Rangel-Carrillo  <https://orcid.org/0000-0002-2206-748X>

Mariutsi Alexandra Osorio-Sanabria  <https://orcid.org/0000-0001-7375-3052>

References

Albano, C. S., & Reinhard, N. (2014). Open Government Data: Facilitating and Motivating Factors for Coping with Potential Barriers in the Brazilian Context. In M. Janssen, H. J. Scholl, M. A. Wimmer, & F. Bannister (Ed.),

- Electronic Government. EGOV 2014. Lecture Notes in Computer Science. 8653*, pp. 181–193. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44426-9_15
- Attard, J., Orlandi, F., & Auer, S. (2016). Value Creation on Open Government Data. *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 2605-2614). Koloa, HI, USA: IEEE. <https://doi.org/10.1109/HICSS.2016.326>
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015, October). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399-418. <https://doi.org/10.1016/j.giq.2015.07.006>
- Bracken, M., Greenway, A., & Angeles, K. (August de 2019). *De la información a la inteligencia: ¿Cómo adaptar las instituciones para el análisis de datos en el gobierno?* <https://doi.org/10.18235/0001850>
- BSA. (2022, July 20). *What's the Big Deal With Data?* BSA.org: https://www.bsa.org/files/reports/bsadatastudy_en.pdf
- Cadena-Vela, S., Fuster-Guilló, A., & Mazón, J.-N. (2019). Publicando datos abiertos considerando criterios de calidad. *Revista Ibérica de Sistemas e Tecnologías de Informação - RISTI(E22)*, 295–308. <https://core.ac.uk/download/pdf/266365329.pdf>
- Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14, 1-10. <https://doi.org/10.5334/dsj-2015-002>
- Caro, A., Fuentes, A. M., & Soto, A. (2013). Desarrollando sistemas de información centrados en la calidad de datos. *Ingeniare. Revista chilena de ingeniería*, 21(1), 54-69. <https://doi.org/10.4067/S0718-33052013000100006>
- Ciancarini, P., Poggi, F., & Russo, D. (2016). Big Data Quality: A Roadmap for Open Data. *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 210-215). Oxford, UK: IEEE. <https://doi.org/10.1109/BigDataService.2016.37>
- Ciancarini, P., Poggi, F., & Russo, D. (2016). Big Data Quality: A Roadmap for Open Data. *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*. Oxford, UK: IEEE. <https://doi.org/10.1109/BigDataService.2016.37>
- Cooper, H. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1,104), 104-126. <https://doi.org/10.1007/BF03177550>
- Datos Abiertos Costa Rica. (2020, May). *Guía Nacional para la publicación de Datos Abiertas*. <https://www.mideplan.go.cr>
- Departamento Nacional de Planeación, Ministerio de Tecnologías de la Información y las Comunicaciones, Superintendencia de Industria y Comercio. (17 de April de 2018). *Política Nacional de Explotación de Datos (Big Data)*. Documento CONPES 3920. <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>
- DNP, MinTIC, SIC. (2018). *CONPES 3920 Política Nacional de Explotación de Datos (Big Data)*. Bogotá D.C., Colombia: Consejo Nacional de Política Económica y Social, República de Colombia, Departamento Nacional de Planeación. POLÍTICA NACIONAL DE EXPLOTACIÓN DE DATOS
- DOF. (2017, December 12). *Guía de Implementación de la Política de Datos Abiertos*. Diario Oficial de la Federación: https://www.dof.gob.mx/nota_detalle.php?codigo=5507476&fecha=12/12/2017#gsc.tab=0
- Faini, F., & Palmirani, M. (2016). Italian Open and Big Data Strategy. In A. Kó, & E. Francesconi (Eds.), *Electronic Government and the Information Systems Perspective. EGOVIS 2016. Lecture Notes in Computer Science* (Vol. 9831, pp. 105-120). Guimarães, Portugal: Springer. https://doi.org/10.1007/978-3-319-44159-7_8
- Fraefel, M., Haller, S., & Gschwend, A. (2017). Big Data in the Public Sector. Linking Cities to Sensors. In M. Janssen, & M. Janssen (Ed.), *Electronic Government. EGOV 2017. Lecture Notes in Computer Science* (Vol. 10428, pp. 276-286). St. Petersburg, Russia: Springer. https://doi.org/10.1007/978-3-319-64677-0_23
- Freitas, A., & Curry, E. (2016). Big Data Curation. In J. M. Cavanillas, E. Curry, & W. Wahl (Eds.), *New Horizons for a Data-Driven Economy. A Roadmap for Usage and Exploitation of Big Data in Europe* (pp. 87–118). Springer, Cham. https://doi.org/10.1007/978-3-319-21569-3_6
- Gandomi, A., & Haider, M. (2015, April). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gobierno de la República de Colombia, World Bank Group. (2016, Mayo 16). *Open Data Readiness Assessment*. Datos.gov.co: https://herramientas.datos.gov.co/sites/default/files/2020-11/articles-9407_evaluacion_apertura_datos.pdf
- Goldsmith, S. (2017, March 7). *Análisis predictivo: Impulsar mejoras mediante el uso de datos (Predictive analytics: driving improvement from data)*. <https://blogs.iadb.org/administracion-publica/es/analisis-predictivo-impulsar-mejoras-mediante-uso-datos/>
- Herrera-Melo, C. A., & González-Sanabria, J. S. (2020). Proposal for the Evaluation of Open Data Portals. *Revista Facultad de Ingeniería*, 29(54), e10194. <https://doi.org/10.19053/01211129.v29.n0.2020.10194>
- Huang, R., Lai, T., & Zhou, L. (2017). Proposing a framework of barriers to opening government data in China: A critical literature review. *Library Hi Tech*, 35(3), 421-438. <https://doi.org/10.1108/LHT-01-2017-0014>
- ISO/IEC. (2014). *ISO/IEC 25000:2014(en)*. ISO.org: <https://www.iso.org/obp/ui#iso:std:iso-iec:25000:ed-2:v1:en>
- ISO25000. (2022). *ISO/IEC 25012*. <https://iso25000.com/index.php/en/iso-25000-standards/iso-25012>
- Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). Open Government Data: A Stage Model. In M. Janssen, H. J. Scholl, M. A. Wimmer, & Y.-h. Tan (Ed.), *International Conference on Electronic Government EGOV 2011: Electronic Government. Lecture Notes in Computer Science. 6846*, pp. 235-246. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22878-0_20

- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7-15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- Koltay, T. (2020). Quality of Open Research Data: Values, Convergences and Governance. *Information*, 11(4), 175. <https://doi.org/10.3390/info11040175>
- Kucera, J. (29 de April de 2015). Open Government Data Publication Methodology. *Journal of Systems Integration*, 6(2), 52-61. <https://doi.org/10.20470/jsi.v6i2.231>
- Kucera, J., & Chlapek, D. (2014). Benefits and Risks of Open Government Data. *Journal of Systems Integration*, 5(1), 30-41. <https://www.proquest.com/openview/dc63783cc24944fd3379676fcad69b89/1?pq-origsite=gscholar&cbl=2032453>
- Ley 1712 de 2014. (2014, March 6). <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=56882>
- Li, Y., Seshadri, N., & Ariyavisitakul, S. (1999, March). Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels. *IEEE Journal on Selected Areas in Communications*, 17(3), 461-471. <https://doi.org/10.1109/49.753731>
- López Beltrán, N. E., & Mahecha Moyano, J. F. (2017). *Prototipo de software para la evaluación de la calidad de los datos abiertos*. Tesis presentada como requisito parcial para optar al título de Ingeniero de Sistemas, Universidad Católica de Colombia, Facultad de Ingeniería, Programa de Ingeniería de Sistemas, Bogotá. <https://repository.ucatolica.edu.co/server/api/core/bitstreams/c87d4a51-bb6c-4f36-be38-bd365a076df7/content>
- Loshin, D. (2014, May 19). *Understanding Big Data Quality for Maximum Information Usability*. [White paper]. TechRepublic: <https://www.techrepublic.com/resource-library/whitepapers/understanding-big-data-quality-for-maximum-information-usability/>
- Maestre Góngora, G. P., & Nieto Bernal, W. (2015, December). Factores Clave en la Gestión de Tecnología de Información para Sistemas de Gobierno Inteligente. *Journal of technology management & innovation*, 10(4). <https://doi.org/10.4067/S0718-27242015000400012>
- Maestre-Gongora, G., Rangel-Carrillo, A., & Osorio-Sanabria, M. (2021a). El valor de los datos abiertos de gobierno: un enfoque desde la evaluación de calidad. *Revista de Investigación, Desarrollo e Innovación*, 11(3), 507-518. <https://doi.org/10.19053/20278306.v11.n3.2021.13348>
- Maestre-Góngora, G., Rangel-Carrillo, A., & Osorio-Sanabria, M. (2021b). Modelo para evaluación de calidad de datos abiertos de gobierno: Caso Colombia. In P. J. Escamilla-Ambrosio, L. Hernández-Callejo, S. Nesmachnow, P. Moreno, & D. Rossit (Ed.), *Proceedings of the IV Ibero-American Conference on Smart Cities (ICSC-CITIES 2021)*, (pp. 1-16). https://drive.google.com/file/d/1KIPzPathh-tisStS8R8E4t_eVBXrRUy3/view
- Maestre-Gongora, G., Rangel-Carrillo, A., & Osorio-Sanabria, M. A. (2023, February 20). Open government data and big data from a quality perspective [Data set]. *En Revista Colombiana de Computación*. Zenodo. <https://doi.org/10.5281/zenodo.7657421>
- Mahecha Moyano, J., López Beltrán, N., & Velandia Vega, J. (2017). Assessing data quality in open data: A case study. *2017 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)* (pp. 1-5). Bogotá, Colombia: IEEE. <https://doi.org/10.1109/CONIITI.2017.8273343>
- Marsh, R. (2005). Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management. *Journal of Database Marketing & Customer Strategy Management*, 12(2), 105-112. <https://doi.org/10.1057/palgrave.dbm.3240247>
- Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). Open Data: Barriers, Risks and Opportunities. *European Conference on e-Government; Kidmore End: Academic Conferences International Limited*, 301-317. <https://www.proquest.com/docview/1400450850?pq-origsite=gscholar&fromopenview=true>
- Méndez Matamoros, J. H. (2017). *Mejoramiento de calidad en conjunto de datos abiertos basado en la aplicación de métricas de consistencia lógica*. Tesis de grado para optar por el título de Magíster en Ciencias de la Información y las Comunicaciones, Universidad Distrital Francisco José de Caldas, Maestría en Ciencias de la Información y las Comunicaciones, énfasis en Sistemas de Información, Bogotá, Colombia. <https://repository.udistrital.edu.co/bitstream/11349/8032/1/MendezMatamorosJorge>
- Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016, October). A Data Quality in Use model for Big Data. *Future Generation Computer Systems*, 63, 123-130. <https://doi.org/10.1016/j.future.2015.11.024>
- Miloslavskaya, N., & Tolstoy, A. (2016). Big Data, Fast Data and Data Lake Concepts. *Procedia Computer Science*, 88, 300-305. <https://doi.org/10.1016/j.procs.2016.07.439>
- Ministério da Mulher, da Família e dos Direitos Humanos. (2020, July 15). *Plano de Dados Abertos 2020-2022*. <https://www.gov.br/mdh/pt-br/acao-a-informacao/dados-abertos/DadosAbertos20202022.pdf>
- Ministerio de Modernización. Presidencia de la Nación. (2016, October 6). *Kit de datos abiertos*. https://www.argentina.gob.ar/sites/default/files/2_kit_de_datos_abiertos.pdf
- MinTIC. (2018, October 1). *Guía de estándares de Calidad e Interoperabilidad de datos abiertos del Gobierno de Colombia*. https://herramientas.datos.gov.co/sites/default/files/2020-11/A_guia_de_estandares_final_0.pdf
- MinTIC. (2019a, September). *Guía para el uso y aprovechamiento de Datos*. <https://herramientas.datos.gov.co/sites/default/files/Guia%20de%20Datos%20Abiertos%20de%20Colombia.pdf>

- MinTIC. (2019b, December 2). *Requisitos de calidad para Datos Abiertos*. Sello de Excelencia: https://sellodeexcelencia.gov.co/documents/UTSF_SDE_Requisitos_de_calidad_para_datos_abiertos_2019_1_2_02_v_2_0.pdf
- MinTIC. (2021, June). *Hoja de Ruta de Datos Abiertos Estratégicos para el Estado Colombiano*. Datos Abiertos Colombia: https://herramientas.datos.gov.co/sites/default/files/2021-07/Hoja%20de%20Ruta%20Datos%20Abiertos%20Estrat%C3%A9gicos%202021_1.pdf
- MinTIC. (2023). *Inicio*. Datos abiertos de Colombia: <https://www.datos.gov.co/>
- Muente-Kunigami, A., & Serale, F. (2018, July). *Los datos abiertos en América Latina y el Caribe*. <https://publications.iadb.org/>: <https://publications.iadb.org/es/los-datos-abiertos-en-america-latina-y-el-caribe>
- Mukherjee, S., & Shaw, R. (2016). Big data-concepts, applications, challenges, and future scope. *International Journal of Advanced Research in Computer and Communication Engineering*, 5(2). <https://doi.org/10.17148/IJARCC.2016.5215>
- Munné, R. (2016). Big Data in the Public Sector. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New Horizons for a Data-Driven Economy* (pp. 195-208). Springer, Cham. <https://doi.org/10.1007/978-3-319-21569-3>
- Muquet, B., Wang, Z., Giannakis, G. B., de Courville, M., & Duhamel, P. (2002). Cyclic prefixing or zero padding for wireless multicarrier transmissions? *IEEE Transactions on Communications*, 50(12), 2136-2148. <https://doi.org/10.1109/TCOMM.2002.806518>
- Nafis, F., Yousfi, S., & Chiadmi, D. (2016). How Big Open Data Can Improve Public Services. In A. El Oualkadi, F. Choubani, & A. El Moussati (Ed.), *Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015. Lecture Notes in Electrical Engineering*. 381, pp. 607–612. Springer, Cham. https://doi.org/10.1007/978-3-319-30298-0_65
- Osorio-Sanabria, M. A., Amaya-Fernández, F., & González-Zabala, M. (2020). Políticas, normas y estrategias que fomentan los datos abiertos en Colombia: un análisis de literatura. *Universidad Católica Del Norte*(62), 155–188. <https://doi.org/10.35575/rvucln.n62a7>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372(71), 1-9. <https://doi.org/10.1136/bmj.n71>
- Paulraj, A. J., Gore, D. A., Nabar, R. U., & Bolcskei, H. (2004, Feb.). An overview of MIMO communications—A key to gigabit wireless. *Proceedings of the IEEE*, 92(2), 198–218. <https://doi.org/10.1109/JPROC.2003.821915>
- PowerData. (2017, April 6). *Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad*. PowerData: <https://www.powerdata.es/big-data>
- Rangel-Carrillo, A. M., Maestre-Góngora, G. P., & Osorio-Sanabria, M. A. (2020, December). Principios, lineamientos, dimensiones y atributos para la evaluación de calidad de Datos Abiertos de Gobierno. *AiBi Revista De Investigación, Administración E Ingeniería*, 8(S1), 54–65. <https://doi.org/10.15649/2346030X.950>
- Redman, T. C. (2016, September 22). Bad Data Costs the U.S. \$3 Trillion Per Year. *Harvard Business Review*, 1-4. Sitio web de HBR: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>
- Rodríguez, P., Palomino, N., & Mondaca, J. (2017, Julio). El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe. *Resumen de políticas del BID*(266). <https://doi.org/10.18235/0000893>
- Russom, P. (2011). *Big Data Analytics. Executive Summary*. [tdwi.org: https://tdwi.org/research/2011/09/~/-/media/TDWI/TDWI/Research/BPR/2011/TDWI_BPReport_Q411_BigData_Analytics_Web/TDWI_BPReport_Q411_Big%20Data_ExecSummary.ashx](https://tdwi.org/research/2011/09/~/-/media/TDWI/TDWI/Research/BPR/2011/TDWI_BPReport_Q411_BigData_Analytics_Web/TDWI_BPReport_Q411_Big%20Data_ExecSummary.ashx)
- Sabri, N. A., Emran, N. A., & Harum, N. (2019, October). Government Open Data Portals: A Measurement of Data Veracity Coverage. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(12), 1975-1983. <https://doi.org/10.35940/ijitee.L2908.1081219>
- Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, 37(3), 150-154. <https://doi.org/10.1016/j.ijinfomgt.2017.01.003>
- Secretaría de la Función Pública, Coordinación de Estrategia Digital Nacional. (2015, October 28). *Adopción de la Carta Internacional de Datos Abiertos*. Carta Internacional de Datos Abiertos: <https://www.gob.mx/sfp/documentos/carta-de-datos-abiertos-nuevo#:~:text=L%20Carta%20Internacional%20de%20Datos,un%20conjunto%20de%20principios%20compartidos>
- Talukder, S., Shen, L., Talukder, F. H., & Bao, Y. (2019). Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh. *Technology in Society*, 56, 147-156. <https://doi.org/10.1016/j.techsoc.2018.09.013>
- TodoBI. (2019, October 25). *11 Consejos sobre Bad Data: el enemigo silencioso en Business Intelligence y Big Data*. <https://www.todobici.com/11-consejos-sobre-bad-data-el-enemigo/>
- Torres Saumeth, K. M., Ruiz Afanador, T. S., Solís Ospino, L., & Martínez Barraza, F. (2012). Calidad y su evolución: una revisión. *Dimensión empresarial*, 10(2), 100-107.
- UNECE. (2015). *A Suggested Framework for National Statistical Offices for assessing the Quality of Big Data*. https://cros-legacy.ec.europa.eu/system/files/Task%20Team%20Big%20Data%20Quality%20Framework_937_unblinded_v1.pdf

- Vélez, M. I., Gómez Santamaría, C., Osorio Sanabria, M. A., & Sánchez Quintero, T. (2022). *Informe 2: Conceptos fundamentales y uso responsable de la inteligencia artificial en el sector público*. (M. C. Rodríguez, Ed.) scioteca.caf.com:
<https://scioteca.caf.com/bitstream/handle/123456789/1921/Conceptos%20fundamentales%20y%20uso%20responsable%20de%20la%20inteligencia%20artificial%20en%20el%20sector%20p%C3%BAblico.pdf?sequence=5&isAllowed=y>
- Vetrò, A., Canova, L., Torchiano, M., Orozco Minotas, C., Lemma, Raimondo, & Morando, Federico. (2016, April). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337. <https://doi.org/10.1016/j.giq.2016.02.001>
- Wahyudi, A., Kuk, G., & Janssen, M. (2018). A Process Pattern Model for Tackling and Improving Big Data Quality. *Information Systems Frontiers*, 20, 457-469. <https://doi.org/10.1007/s10796-017-9822-7>
- Wang, D., Han, B., Zhao, J., Gao, X., & You, X. (2003, September). Channel estimation algorithms for broadband MIMO-OFDM sparse channel. *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003*. 2, pp. 1929-1933. Beijing, China: IEEE. <https://doi.org/10.1109/PIMRC.2003.1260454>
- Warrach, N. F., Rasool, T., & Sajid, M. (2019). Challenges to Use Open Government Data through the Citizens' Lens: A Systematic Review. *Journal of Political Studies*, 26(2), 1-14. http://pu.edu.pk/images/journal/pols/pdf-files/1-v26_2_19.pdf
- Wieczorkowski, J. (2019). Barriers to Using Open Government Data. *ICEEG '19: Proceedings of the 3rd International Conference on E-commerce, E-Business and E-Government* (pp. 15-20). ACM. <https://doi.org/10.1145/3340017.3340022>
- Williams, D., & Tang, H. (February de 2020). Data quality management for industry 4.0: A survey. *Quality Resources*, 22(2), 26-35.
- Yang, Z., & Kankanhalli, A. (2013). Innovation in Government Services: The Case of Open Data. In Y. K. Dwivedi, H. Z. Henriksen, D. Wastell, & R. De' (Ed.), *Grand Successes and Failures in IT. Public and Private Sectors. TDIT 2013. IFIP Advances in Information and Communication Technology*. 402, pp. 644-651. Bangalore, India: Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38862-0_47
- Yi, M. (2019). Exploring the quality of government open data: Comparison study of the UK, the USA and Korea. *The Electronic Library*, 37(1), 35-48. <https://doi.org/10.1108/EL-06-2018-0124>
- Zainal, N. Z., Hussin, H., & Nazri, M. N. (December de 2019). Acceptance, Quality and Trust Factors - Conceptual Model for Open Government Data Potential Use. *International Journal on Perceptive and Cognitive Computing*, 5(2). <https://doi.org/10.31436/ijpcc.v5i2.120>
- Zuiderwijk, A. (2017). Analysing Open Data in Virtual Research Environments: New Collaboration Opportunities to Improve Policy Making. *International Journal of Electronic Government Research (IJEGR)*, 13(4), 17. <https://doi.org/10.4018/IJEGR.2017100105>
- Zuiderwijk, A., & Janssen, M. (2015). Participation and Data Quality in Open Data use: Open Data Infrastructures Evaluated. *European Conference On e-Government*. <https://www.semanticscholar.org/paper/Participation-and-Data-Quality-in-Open-Data-use%3A-Eijk-Janssen/1a2444ae708836712897dc9a1f5b7e373ed136e6>
- Zuiderwijk, A., Janssen, M., & Iryna, S. (2016). Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 116-146. <https://doi.org/10.1080/10919392.2015.1125180>