SYSTEMATIC REVIEW

Open Access

Open government data in health sector: a systematic literature review open government data in health sector

Patrícia Nascimento Silva[1] and Sarah Nascimento Silva[2*]

Abstract

Background Open government data (OGD) in the health sector consolidates transparency, access to information, and collaboration between the government and different sectors of society. It is an essential instrument for health systems and researchers to generate initiatives, drive innovations, and qualify decision-making, whether in health emergencies or supporting the creation of more effective public policies. This review aimed to identify OGD initiatives in healthcare and their possible applications.

Methods A systematic search was conducted in the MEDLINE (PubMed), Web of Science, Scopus, and Scielo databases up to February 12, 2024. Eligibility criteria included original studies that addressed OGD initiatives in public health. Studies using open data to prepare other analyses, reviews, and opinion articles were excluded. The quality methodological of related studies was based on the JBI Critical Appraisal tool. Data extraction included the location of the application, description of the initiative/use of data, themes addressed, data quality, and respective evaluation tools.

Results Of the 3,176 records initially identified, six studies were included describing initiatives to make available datasets or portals that publish OGD in the healthcare area between 2020 and 2023. Most publications evaluated quality and usability requirements, with five related to the COVID-19 pandemics. There is significant heterogeneity in the data formats available, obtaining, on average, a 3-star rating in the Berners-Lee evaluation model. The availability of open data presents geographic differences, with regions being more developed than others in both the most comprehensive and the most regionalized studies.

Conclusions Few studies using OGD in health were identified in the scientific literature. The heterogeneity of open data formats, the lack of standards, and the quality of the data point to limitations and the need for improvements to enable the use and reliable reuse of this data.

Trial registration PROSPERO CRD42024498446.

Keywords Open government data, Open health data, Government health data, Open government health data, Public Health, Systematic review

*Correspondence:
Sarah Nascimento Silva
sarah.nascimento@fiocruz.br
[1]Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[2]Núcleo de Avaliação de Tecnologias em Saúde, Grupo de Pesquisa Clínica e Políticas Públicas em Doenças Infecciosas e Parasitárias, Instituto René Rachou, Fundação Oswaldo Cruz, Avenue Augusto de lima, 1715, Barro Preto, Belo Horizonte, MG 30190-009, Brazil

Background

Open government health data are accessible and transparent, offering potential insights that facilitate monitoring health trends, risk factors, and healthcare data. These insights support the development of programs, policies, and decision-making processes [1]. The scope of open government health encompasses various domains such as environmental health, medical devices, healthcare and social services, community health, mental health, and substance abuse [2]. In public health emergencies, open government health data provide numerous opportunities, ranging from generating early warnings for outbreaks and pandemics to delivering personalized medicine and enhancing health system management [3].

Open Government Data (OGD) are defined by a set of eight principles ensuring that such data are complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary, and license-free. Governments make these data available via the internet for free use, reuse, and sharing under the same licenses that credit the original source [4, 5]. Open government is a concept linked to transparency in governmental activities, social participation, and access to information. This intersects with the term open data, meaning data that are freely available for use and reuse by anyone. This concept led to the creation of OGD data in 2007 in Sebastopol, California, by a group of researchers and representatives from civil society organizations [5, 6]. The discussion on open government gained momentum in 2009 with the issuance of the Open Government Declaration in the United States [7]. It was further propelled in 2011 with the launch of the Open Government Partnership, an international initiative to encourage open government practices globally [8]. Since then, open data policies have been implemented in various countries to regulate the dissemination of OGD.

As open data publication has become more widespread, there has been a massive increase in the data sets available, alongside advancements in communication and information technologies. This has led to observed segmentations within open data, as identified in the health sector. Health data are a distinct category of OGD [9], and new terms have been noted in the literature to designate this segmentation, with some variations: open government health, open health data, government health data, and open government health data.

The availability of open data in the health context enhances the concept of information access and citizen collaboration, aiding patients, healthcare providers, and researchers in decision-making and driving innovations. However, it also necessitates crucial initiatives such as legal guidance, data protection, financial resources, cultural change, and sustained leadership [10]. The dissemination of this data type allows governments and health system institutions to fully utilize their potential to formulate more effective policies. To achieve this, various strategies must be employed to manage open data properly, aimed at strengthening data quality, confidentiality, and security [11, 12]. In 2016 a review found that although there is recognition of the importance of using OGD in health, many countries still do not have guidelines for collecting and managing this data [12]. Given this gap, the study is justified by the importance of detecting whether governments actively and continuously publish OGD on public health, generating data that can support decision-making. The objective of this review to identify, map and describe initiatives developed to date with their potential and limitations.

Method

A systematic literature review was conducted to gather evidence on initiatives and potential applications for using OGD in the health sector. This review aimed to answer the following question: What initiatives involve OGD in public health? The report focuses on the locations where these initiatives are being developed and the proposed applications for their use.

The review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) under the number CRD42024498446, and the review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [13], incorporating some adaptations specific to the proposed theme (Supplementary Table 1).

Search strategy

The literature search was completed on February 12, 2024, utilizing databases MEDLINE (PubMed), Web of Science, Scopus, and Scielo. The search strategy employed descriptors tailored to each database, focusing on core themes: open government data, open health data, public health, and their variations (Additional File 1).

Eligibility criteria

The review included original articles, conference proceedings, and books that discussed initiatives to publish and make available OGD data in health sector. We excluded documents that: (i) used open data as a data source for any other study or analysis, including data mining, machine learning, and knowledge discovery, among other techniques; (ii) discussed OGD initiatives for private databases or applications, as well as initiatives by research groups or individuals; (iii) published in languages other than Portuguese, English, and Spanish; (IV) secondary studies such as systematic reviews, and opinion articles.

Study selection

Initially, the titles and abstracts of studies retrieved from the initial search were screened for duplicates and selected by two independent reviewers (PNS and SNS) using the Rayyan application [14]. Subsequently, the full texts of potentially relevant articles were assessed against the selection criteria, maintaining independent review and resolving any discrepancies regarding inclusion or exclusion by consensus.

Data extraction and analysis

Data from the articles were extracted by two independent reviewers (PNS and SNS) and compiled into an Excel spreadsheet. The extracted data included the author, year of publication, location of application, description of the initiative/use of data, themes addressed, data quality of databases, and respective evaluation tools. The data synthesis was presented qualitatively, combining individual findings from the main points identified in the literature, results, and available analysis measures, followed by a global analysis from conclusions of the identified data.

Methodological quality assessment

The methodological quality of the articles was evaluated according to the Joanna Briggs Institute (JBI) manual for systematic reviews using methodological criteria. Given the nature of publications on OGD, which are typically aimed at describing data sets, actions, and initiatives rather than adhering to specific methodological standards and designs, this review uses the JBI Critical Appraisal Checklist for Textual Evidence: Narrative [15]. This tool comprises six questions that can be answered as "yes," "no," "uncertain," or "not applicable" for each assessed domain. The methodological quality of the articles was categorized by the authors of this review as low quality (up to 50% "yes" scores), moderate quality (between 50% and 80% "yes" scores), and high quality (above 80% "yes" scores).

Results

Systematic search

The initial selection of studies from the databases identified 3,176 records, and after the removal of duplicates, 1,922 studies were screened by titles and abstracts. Subsequently, 47 articles remained for full-text review, of which 6 were included (Fig. 1). Among the 41 excluded articles, the reasons were opinion articles or interviews ($n=12$); studies integrating data from various sources ($n=11$); data analysis or data treatments that included analysis of interfaces or publication standards/formats ($n=6$); articles without open access ($n=4$); non-governmental initiatives ($n=3$); languages other than those specified in the protocol ($n=3$); general data not restricted to health ($n=1$); and the type of article—literature review ($n=1$). More detailed information about excluded studies are presented in the Additional File 2.
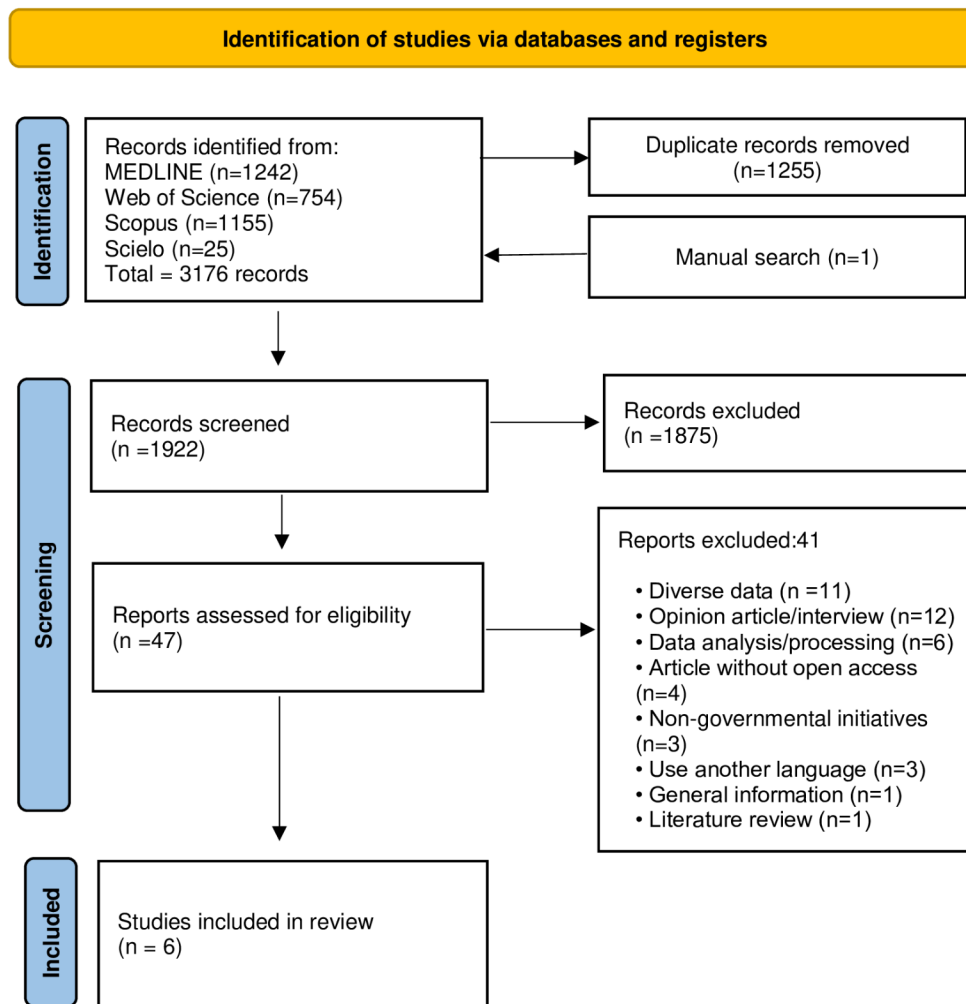
Characteristics of the included studies

The included publications cover initiatives for making available data sets or portals that publish OGD in the health sector from 2020 to 2023. The publications present data from individual countries, such as Spain/Autonomous Communities [16, 17] and Italy [18]; comparisons involving the USA, European Union (EU), and China [2]; South America (Argentina, Colombia, Mexico, Paraguay, Peru, and Uruguay) [19]; and a global scope involving 35 countries with a high number of confirmed COVID-19 cases [20] (Table 1). All of these studies addressed analyses related to the context of the COVID-19 pandemic, except one that carried out a general analysis of healthcare datasets on open data portals [16].

The themes explored in the articles emphasize the exploration of open health data sets on governmental portals, focusing primarily on quality assessment criteria. An exception is one study that developed a framework for evaluating the quality and usability of open government health data in the context of the COVID-19 pandemic [2]. A significant finding across the studies is the data heterogeneity and the absence of standardized protocols for publishing and making data available, necessitating the use of various tools to evaluate open data quality from differing perspectives.

Data quality described

The five-star evaluation model by Berners-Lee [21], a primary reference for assessing open data quality, was used to evaluate COVID-19 data sets in studies by Martin Fernandez et al., 2021 [17] and Tang et al., 2023 [20]. Additionally, the research by Quesada et al., 2020 [16] employed this model, integrating the UNE 178,301 standard [22] for health data sets. The comparative assessment using the five-star model across three studies is depicted in Fig. 2, showing that most data sets received a three-star rating, followed by four stars.

One study utilized the Open Data Quality Validation Tool (HEVDA) [19], to compare Latin American countries (Table 2). The HEVDA enables the validation of various metrics for open data sets in CSV (Comma Separated Variable) formats. In the HEVDA tool, metrics 1 to 4 are considered critical and 5 to 8 are not critical as defined by the authors [19]. These articles, which identified and described the data sets, highlighted significant disparities in the availability of open data, with some regions being more developed than others in both comprehensive [19, 20], and more regionally focused studies [16, 17]. The results of the Martinez study in Latin America show that in the global analysis, the most unmet critical the most unfulfilled critical metric is Metric 4 with

Identification of studies via databases and registers

Identification

Records identified from:
MEDLINE (n=1242)
Web of Science (n=754)
Scopus (n=1155)
Scielo (n=25)
Total = 3176 records

→ Duplicate records removed
(n=1255)

← Manual search (n=1)

Screening

Records screened
(n =1922)

→ Records excluded
(n =1875)

Reports assessed for eligibility
(n =47)

→ Reports excluded:41

• Diverse data (n =11)
• Opinion article/interview (n=12)
• Data analysis/processing (n=6)
• Article without open access
(n=4)
• Non-governmental initiatives
(n=3)
• Use another language (n=3)
• General information (n=1)
• Literature review (n=1)

Included

Studies included in review
(n = 6)

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

Fig. 1 PRISMA flow diagram

41.47%, followed by Metric 3 with 34.12%, then Metric 1 with 18.24% and finally Metric 2 with 6.18%. The most unfulfilled non-critical metric is Metric 5 with 43.67%, followed by Metric 6 with 20.82%, then Metric 7 with 18.16% and finally Metric 8 with 17.35% [19].

Two studies leveraged the COVID-19 pandemic as an opportunity to utilize specific tools or examples to enhance the application of open data. The study by Maione et al., 2022 [18] assessed the patterns of OGD in Italy and drew significant political and managerial implications for the future development of open data as a mechanism for accountability during exceptional times. By defining intrinsic characteristics (visibility, granularity, accessibility, and linkability) and extrinsic characteristics (completeness, timeliness, clarity, comparability, relevance, and accuracy), their overall analysis concluded that the data provide an unsatisfactory level of transparency for public accountability, due both to the partial publication of open data on institutional portals and to the characteristics they exhibit. Consequently, one-third of the portals analyzed did not contain any data related to the pandemic, despite all regional governments claiming to present them.

The study by Dan Wu et al., 2022 [2] utilized open government health data related to COVID-19 from the USA, EU, and China to validate a framework for assessing the

Table 1 Characterization of the articles included

Author, year	Country/region	Type of study	Analysis period	Setting (context)	Main results	Quality data	Conclusions
Bertos Quesada et al., 2020	Spain/ autonomous regions	Descriptive	01/15/2020 to 03/01/2020	Analyze the publishing status of healthcare datasets on open data portals	-Identified 485 healthcare datasets (739 formats) - Data in format reutilizable: 92,6% - Predominant data is health statistics,	Variability in data quality by province - Most data were reusable (3 stars): 62.4%	Existence of some autonomous regions more advanced than others in the publication of this type of data.
Dan Wu et al., 2022	EUA EU China (provincial capital cities)	Descriptive (Proposal for a framework)	08/10/2020 to 08/20/2020	Propose a more targeted open government data quality evaluation framework and apply it to open datasets about COVID-19.	-Open government health data in the US, EU and China lacks the necessary metadata. -richness of content and timeliness of open datasets need to be improved	Quality difference of the current open government health data (Metrics assessment: Dataset Quality, Descriptive Data Quality, Metadata Quality)	Points out the need for a targeted framework to evaluate the quality of open government data to measure data quality dimensions with a fine-grained measurement approach.
Maione et al., 2021	Italian	Descriptive (exploratory research)	March/2020 to 05/13/2021	Web content analysis of portals in 20 provinces. Describe characteristics of data opened, the following ones have been contemplated: completeness; timeliness; clarity; comparability; relevance; and accuracy.	- A third of the portals analyzed does not contain any pandemic-related data, despite all the regional governments declaring their commitment to opening up data for transparency and accountability purposes - A large amount of data organized into an aggregated manner, not linked to other data sets and provided in formats that prevent direct or autonomous processing and analysis	-Inconsistencies and shows an unsatisfactory level of transparency for public accountability Assessment intrinsic characteristics (visibility, granularity, accessibility, and linkability) and extrinsic characteristics (completeness, timeliness, clarity, comparability, relevance, and accuracy)	Open data does not enable to properly address the accountability concerns of a demanding forum at times of exception.
Martin Fernandez et al., 2021	Spain	Descriptive	july/2020 to january/2021	Open data related to COVID-19 published by the regions of Spain and the Ministry of Health	-Data in 15 regions and Ministry of Health - Many format to data - Format CSV – 86% das províncias - COVID-19 panel is currently presented geographical information and indicators of evolution of the pandemic	Assessment of 15 regions Four stars: 2 Three stars: 10 Two stars: 1 One star: 2	Enormous diversity among the regions in terms of the publication of COVID data.

Table 1 (continued)

Author, year	Country/region	Type of study	Analysis period	Setting (context)	Main results	Quality data	Conclusions
Martinez et al., 2023	Argentina, Colombia, Mexico, Peru, Paraguay and Uruguay	Descriptive	Not Report	Validate the quality of open data sets from Latin American countries, focusing on the pandemic using HEVDA (Open Data Quality Validation Tool) focused on data quality metrics.	-Locks in the verified files: Paraguay (93.33%) followed by Mexico (73.33%) -High percentage of COVID-19 datasets that do not meet the basic objective of interoperability so that they can be reused -Paraguay has the highest number of blocking data sets, while Argentina has the highest number of no-blocking data sets.	- Better quality in the open public data: Argentina and Uruguay. -Lower quality, Mexico and Paraguay. -The metric with the lowest compliance is metric 5 (redundancy for the domain of a column)	-There are several cases of datasets that cannot be analyzed due to blocking problems, hampering interoperability and reuse -some open data portals have little open information; that is, although they have dashboards (monitoring) regarding COVID-19.
Tang et al., 2023	Many countries of world	Descriptive	January/2021	The authors investigated more than 1500 COVID-19 datasets appearing on 35 Open Government Data (OGD) sites around the world. In addition to examining the number of datasets, data formats, star levels, and data levels, they also investigated the topical focus of individual datasets.	-In 35 OGD sites, there were a total of 1514 COVID-19 datasets -the highest number of COVID-19 datasets was in Canada ($n = 434$, 28.67%), followed by the United States ($n = 190$, 12.55%) and Spain ($n = 183$, 12.09%); -There were 41 different types of data formats within these COVID-19 datasets. The most common data format was CSV ($n = 30$, 85.71%); - The country/area that had the highest number of formats was France (15). Austria, Canada, and Spain all had 14 data formats; - The COVID-19 datasets contained data from six governmental levels, plus an "other" category. The most common level was the central government ($n = 33$, 94.29%), followed by state/provincial ($n = 19$, 54.29%), city/municipal ($n = 15$, 42.86%).	- The most common star level was 4 ($n = 21$), followed by level 3 ($n = 6$). Level 5 has four countries: the United States, Italy, Belgium, and Colombia	Significant differences among the world regions were found in both the number of datasets and the number of data formats. Regional differences were also found among various star levels. In addition, regional differences were found in the topic contributions to two specific clusters.

quality and usability of such data. The framework was developed based on a literature analysis of existing data quality assessment methods and the Delphi approach. Accordingly, evaluation metrics and their weights were determined by 15 health experts using the Delphi method and the analytic hierarchy process, employing the Yaahp software to calculate the metrics' weights and verify their consistency. Their findings underscored deficiencies in the quality of the data sets and metadata across the three countries or regions, indicating a need for improvement.

Among the main challenges and difficulties reported by the studies, the heterogeneity of formats and the lack of standardization for data set publication stand out. This diversity is evident even within data from the same country or region, as indicated in analyses in Spain [19] and Spain and its Autonomous Communities [16]. Furthermore, the correct representation of data through adequate and standardized metadata will be essential for interoperability with other systems and its long-term digital preservation.
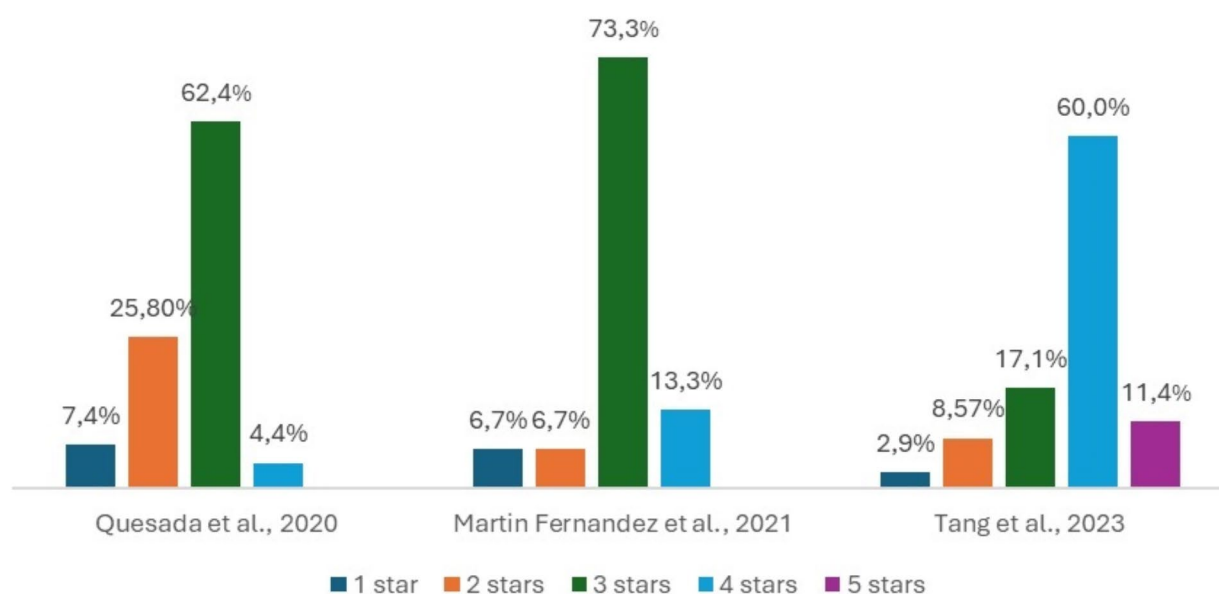
Fig. 2 Evaluation of datasets by 5-star model - Tim Berners-Lee

Table 2 Compare quality study of open government in latin American countries (Martinez et al., 2023)

Open Data Quality Validation Tool (HEVDA)	Argentina	Colômbia	México	Paraguay	Peru	Uruguay
Metric 1 - Decimal numbers	73,0%	60,0%	27,0%	7,0%	47,0%	47,0%
Metric 2 - Duplicated records	87,0%	67,0%	27,0%	7,0%	33,0%	80,0%
Metric 3 - Incomplete data detection	73,0%	53,0%	27,0%	7,0%	53,0%	87,0%
Metric 4 – Invalid Characters	53,0%	53,0%	13,0%	7,0%	47,0%	80,0%
Metric 5 – Redundancy in the domain of a column	40,0%	27,0%	0,0%	7,0%	20,0%	40,0%
Metric 6 - Redundancy between values of columns of the same record	60,0%	53,0%	27,0%	7,0%	33,0%	87,0%
Metric 7 - ID estimation	80,0%	67,0%	27,0%	7,0%	53,0%	73,0%
Metric 8 – Possible trivial fields	80,0%	53,0%	20,0%	7,0%	53,0%	73,0%
Overall	68,3%	54,1%	21,0%	7,0%	42,4%	70,9%

Methodological quality assessment

The quality of the reports was assessed using the JBI Critical Appraisal Checklist for Textual Evidence, resulting in a high-quality classification (Table 3). The results show that, from a methodological perspective, all studies presented an adequate narrative of the results. This tool has 6 evaluation criteria which aim to identify elements related to establishing the legitimacy of the narrator as part of assessing the degree of authenticity of the narrative; the relationship between the characters and the place where the event takes place; and assessing the authenticity of the narrative to distinguish between the emotional persuasiveness of the 'story' and the objective accuracy of the narrative.

Discussion

The pursuit of data openness in the health sector is a necessary action, emphasized in various publications, with its importance highlighted during the early health outbreaks of the 21st century, including the Ebola epidemic and the emergence of the Zika virus [23]. The COVID-19 pandemic, emerging after the open government movement initiated in the 2000s, was the foremost public health emergency and was identified as a priority theme in this review, offering an overview of OGD for use in public health contexts. The heterogeneity of open data formats, the absence of standards, and the low data quality underscore the various limitations and the need for improvements to ensure the reliable use and reuse of these data.

The term "open government health" has been recently adopted and carries a meaning significantly different from merely providing open health data, as it requires adherence to a standard format that facilitates data interoperability for reuse. Interoperability is the ability of different systems and organisations to work together to improve communication, co-operation and co-ordination to ensure that people, organisations and computer systems interact to exchange information effectively and efficiently without creating a technological dependency

Table 3 Assessment of the quality of the report by adapting the tool critical appraisal checklist for textual evidence narrative

Author, year	Is the generator of the narrative a credible or appropriate source?	Is the relationship between the text and its context explained?	Does the narrative present the events using a logical sequence so the reader or listener can understand how it unfolds?	Do you, as reader or listener of the narrative, arrive at similar conclusions to those drawn by the narrator?	Do the conclusions flow from the narrative account?	Do the conclusions flow from the narrative account?	Classification* (low <50%; moderate 50–80%; Hight >80%)
Bertos Quesada et al., 2020	+	+	+	+	+	+	High
Dan Wu et al., 2022	+	+	+	+	+	+	High
Maione et al., 2021	+	+	+	+	+	+	High
Martin Fernandez et al., 2021	+	+	+	+	+	+	High
Martinez et al.,2023	+	+	+	+	+	+	High
Tang et al., 2023	+	+	+	+	+	+	High

+ Yes/ −no

*Classification proposed by authors

between them. Interoperability includes organisational requirements related to work processes, semantic requirements related to the correct interpretation of meaning, and technical requirements related to standards for the presentation, collection, exchange, processing and transport of this data [24]. As a result, the lack of standards for open government data on public health is another barrier to the reuse of these data by governments themselves and also by society.

Another important point to note is that the principles of open government data concern active transparency, i.e. data published on the initiative of governments, and not passive transparency, where citizens request information. Therefore, our review did not request information from government organisations to determine the existence and availability of this data. In addition, although some countries are working on policies and guidelines to establish standards for OGDs, the movements are still local. However, these definitions need to be discussed at a global level in order to link and compare data across countries.

During the pandemic, many data sets were released on websites and portals, combined with other open health data or initiatives, without clear policies or rules for data handling, resulting in incomplete data with limited reusability [25]. However, many of the available data were related to public health statistics, records on health and the environment, and healthcare facilities—vital data for societal reuse that, therefore, require standards to ensure their appropriate use [26]. The use of open health data in the context of Covid-19 may be an indication of the importance of this data format for communicating information to the public in the context of health emergencies, and therefore the importance of producing these data on an ongoing basis. Although the OGD principles were created in 2007, the first publications of health initiatives appeared after 2020. This may be a reflection of low levels of compliance, or difficulties for governments in planning and communicating data on this data standard.

The findings of this systematic review corroborate those of earlier publications on the topic, indicating the low availability and uneven distribution of OGD repositories in public health worldwide [27]. Moreover, the most significant issue relates to the lack of standards for these data. A review study on the use of open data in health highlighted the primary gap as the absence of coordinated guidelines and policies for collecting and managing public health data in many countries, particularly in Latin America [12]. This context is similar to this systematic review, which, although it found many records on the topic, identified only a few articles that met the criteria of this study, namely those that identified initiatives and possible applications with the open data sets on public health produced by governments.

Despite publishing health data on their own platforms and within their healthcare systems [28], many countries do not present data in an open format. Other works feature private initiatives, which are not necessarily governmental [29, 30]. To broaden the dissemination of OGD in the health sector, some publications have discussed actions and proposals to guide governments in this initiative, such as a checklist developed to instruct health agencies on making health data available, aiming to contribute essential standards for offering these data [31]. Another study developed a health data governance strategy with tools, concepts, and recommendations to enable countries to generate open, high-quality, reliable, and secure data [11]. Dissemination of these guidelines will be important for data standards and for the improvement of OGDs being produced by countries. For example, Martinez's study, which evaluated data from Latin America, found that the critical metrics of the HEVDA tool - 'invalid characters' and 'incomplete data' - were not met by 41.7% and 34.12% of the datasets analysed, respectively [19]. This demonstrates the fragility of the OGDs produced in many coutries.

Another widely discussed issue in the health sector involves the ethical considerations for data sharing. Data sharing among the scientific community is an emerging trend, and while the extensive benefits of this practice have been recognized [32, 33], numerous challenges exist to ensure that sharing is conducted correctly and within ethical guidelines [34, 35]. Various legislations must ensure that sensitive data receive appropriate treatment for correct reuse in open data repositories [36]. Undoubtedly, one of the greatest challenges involves providing data that is adequately anonymized yet still capable of preserving useful and reliable information in the health field [37, 38].

The use of OGD should be encouraged in health sector organizations, given its potential to generate social and economic value, providing important support for clinical and epidemiological decision-making in public health systems. However, various challenges and improvements are necessary to achieve this goal. Considering the need to establish better practices or international standards for health data sets, several improvements are proposed for the use of open government health data: (1) increase the number of data sets; (2) improve the data format to make it accessible, interoperable, and reusable; (3) incorporate data sets from various governmental levels; and (4) enhance the relevance of data set content by including topics of specific interest to a particular location [20]. Therefore, it is important to raise awareness among the public and public organizations about the quality of published data and interoperability standards, enabling data reuse in decision-making to promote economic development, transparency, and social participation [19].

Among the various limitations of this study, the issues inherent in the search method stand out. Although broad and standardized terms from the literature were used, they may not have been sufficient to retrieve studies that were inadequately indexed or published in other formats. It is also noteworthy that the retrieval of only a few articles, particularly those related to the pandemic, may be related to the challenges posed by the sudden and unexpected diffusion of this data format during a health emergency. Our search strategy only covered four scientific databases; other repositories were not included and could be a source of data for OGDs. The eligibility criteria also limited the search to publications in the literature, making it impossible to include other primary data sources that could be chosen by governments to publicize OGDs. Furthermore, it was not possible to investigate ethical topics.

The publication of open government health data may not have been the chosen option during the waves of the pandemic for disseminating data, given the popularity of other communication vehicles to convey pandemic information to citizens. Many studies combined and modeled data to obtain projections and track trends, extending beyond the role of OGD. Finally, the vast amount of data generated currently and the quality limitations already identified in the literature may be reasons for governments' retention and more cautious analysis, choosing to delay the release of these data on open data platforms. Nonetheless, this is the first systematic search that has managed to gather studies dedicated to the health sector, presenting an overview of these data. As future work, an evaluation of the quality of the data and the standards used for publication is essential to assess the potential for data re-use, interoperability with other systems and global analysis.

Conclusions

The lack of standards, the low utilization and dissemination of OGD, and the absence of guidelines to coordinate their release in public policies result in any data produced being considered fragile and not representative of the actual scenario. Consequently, this undermines its primary objective of guiding decision-making in health. The full potential of these data can only be tapped when standards that ensure their quality and representativeness to reflect health situations accurately are established. With reliable open health data endorsed by government authorities, patients, professionals, and managers can enhance their decision-making, identify shortcomings, seize opportunities, and drive more effective innovations and interventions in public health. Therefore, in order to promote global interoperability and reuse of OGDs in the health sector, consideration should be given to an international model that defines standards for the

representation, organisation and publication of these data.

Abbreviations

HEVDA	Open Data Quality Validation Tool
JBI	Joanna Briggs Institute
OGD	Open Government Data
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-024-21149-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Data availability

All data supporting the findings of this study are available within the paper and its Supplementary Information. Information for analysing the data and drawing up the tables in the article was entered into a data repository created by the authors and can be accessed via the link: http://doi.org/10.17632/tbdm4939z5.1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Martin EG, Helbig N, Shah NR. Liberating data to transform health care: New York's open data experience. JAMA. 2014;311:2481–2.
2. Wu D, Xu H, Yongyi W, Zhu H. Quality of government health data in COVID-19: definition and testing of an open government health data quality evaluation framework. Libr Hi Tech. 2022;40:516–34.
3. Laessig M, Jacob B, AbouZahr C. Opening data for global health. The Palgrave Handbook of Global Health Data Methods for Policy and Practice. 2019;:451–68.
4. Agune RM, Gregorio Filho AS, Bolliger SP. Governo aberto SP: disponibilização de bases de dados e informações em formato aberto. 2010.
5. OPEN GOVD. Eight principles of open government data. 2007.
6. OPEN DEFINITION (OD). The Open Definition. Avalilable at: http://opendefinition.org/. Accessed 20 January 2024. 2016.
7. Orszag P. Open government directive. Memorandum for the Heads of Executive Departments and agencies. Washington, DC: Office of Management and Budget; 2009.
8. Open Government Partnership (OGP). About Open Government Partnership. Avalilable at: http://opendefinition.org/ Accessed 20 January 2024. 2011.
9. Martin EG, Law J, Ran W, Helbig N, Birkhead GS. Evaluating the quality and usability of open data for public health research: a systematic review of data offerings on 3 open data platforms. J Public Health Manage Pract. 2017;23:e5–13.
10. Martin EG, Begany GM. Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. J Am Med Inform Assoc. 2017;24:345–51.
11. D'Agostino M, Marti M, Mejia F, de Cosio G, Faba G. Strategy for governance of open health data: a paradigm shift in information systems/Estrategia para la gobernanza de datos abiertos de salud: un cambio de paradigma en Los sistemas de informacion/Estrategia para a governanca de dados abertos de saude: uma mudanca de paradigma nos sistemas de informacao. Rev Panam Salud Publica. 2017;41:NA–NA.
12. D'Agostino M, Samuel NO, Sarol MJ, de Cosio FG, Marti M, Luo T, et al. Open data and public health. Revista Panam De Salud Pública. 2018;42:e66.
13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;:n71–71.
14. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5:210.
15. McArthur A, Klugarova J, Yan H, Florescu S. Chapter 4: systematic reviews of text and opinion. JBI Man Evid Synthesis JBI. 2020;10:134–74.
16. Quesada B, Marco Cuenca G. Análisis del estado actual de apertura de datos en salud a nivel autonómico a través del estudio de los portales de datos abiertos. 2020.
17. Fernández AM, Latorre MA, Sánchez-Eznarriaga BC, García LF. Disponibilidad de datos abiertos relacionados con la Covid-19 en Los portales web de las comunidades autónomas y del Ministerio De Sanidad. Rev Esp Salud Publica. 2021;:64.
18. Maione G, Sorrentino D, Kruja AD. Open data for accountability at times of exception: an exploratory analysis during the COVID-19 pandemic. Transforming Government: People Process Policy. 2022;16:231–43.
19. Martínez R, Pons C, Rodríguez R, Vera P. Quality study of open government data related to COVID-19 in Latin America. Revista Facultad de Ingeniería Universidad de Antioquia; 2023. pp. 18–32.
20. Tang R, Jiang J, Zhang Y, Luo J. Open Government Data (OGD) sites and the sharing of country-specific real-time pandemic information: an investigation into COVID-19 datasets available on worldwide OGDs. Inf Process Manag. 2023;60:103489.
21. Berners-Lee T, Linked data, Cambridge, W3C., 2009. Available from: Available from: https://www w3 org/DesignIssues/LinkedData html Access on: May. 2020;3.
22. Española. N, UNE 178301. ; 2015: Ciudades Inteligentes: datos abiertos (Open Data). 2015.
23. Carney TJ, Weber DJ. Public Health Intelligence: learning from the Ebola Crisis. Am J Public Health. 2015;105:1740–4.
24. Kush RD, Warzel D, Kush MA, Sherman A, Navarro EA, Fitzmartin R, et al. FAIR data sharing: the roles of common data elements and harmonization. J Biomed Inf. 2020;107:103421.
25. Ong SWX, Young BE, Lye DC. Lack of detail in population-level data impedes analysis of SARS-CoV-2 variants of concern and clinical outcomes. Lancet Infect Dis. 2021;21:1195–7.
26. Gong M, Jiao Y, Gong Y, Liu L. Data standards and standardization: the shortest plank of bucket for the COVID-19 containment. Lancet Reg Health–Western Pac. 2022;29.
27. Greenberg CJ, Narang S. World Health Organization member states and open health data: an observational study. Epidemiol Biostat Public Health. 2016;13.
28. Martins DP, Correa-Netto NF, Melo N, Loggetto SR, de Liberal MMC. Overview of lymphoma diagnosis in Brazilian public health system patients: open data analysis for health care planning. Hematol Transfus Cell Ther. 2022;44:40–8.
29. Martos-Benitez FD, Soler-Morejon CD, Garcia-del Barco D. Chronic comorbidities and clinical outcomes in patients with and without COVID-19: a large population-based study using national administrative healthcare open data of Mexico. Intern Emerg Med. 2021;16:1507–17.

30. Branda F, Maruotti A. 2022 Uganda Ebola outbreak: early descriptions and open data. J Med Virol. 2023;95:e28344.
31. Martin EG, Shah NR, Birkhead GS. Unlocking the power of open health data: a checklist to improve value and promote use. J Public Health Manage Pract. 2018;24:81–4.
32. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. Bull World Health Organ. 2010;88:462–6.
33. Walport M, Brest P. Sharing research data to improve public health. Lancet. 2011;377:537–9.
34. Jao I, Kombe F, Mwalukore S, Bull S, Parker M, Kamuya D, et al. Research stakeholders' views on benefits and challenges for public health research data sharing in Kenya: the importance of trust and social relations. PLoS ONE. 2015;10:e0135545.
35. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care. Data ran into trouble. J Med Ethics. 2015;41:404–9.
36. Hoffman S. The promise and perils of open medical data. Hastings Cent Rep. 2016;46:6–7.
37. Bader S, Schumann A, Sigg S, Lecue F, Srivastava B, Nie Z et al. Joint Proceedings of the Workshop on AI Problems and Approaches for Intelligent Environments and Workshop on Semantic Cities. ACM; 2013.
38. Schintler LA, Kulkarni R. Big data for policy analysis: the good, the bad, and the ugly. Rev Policy Res. 2014;31:343–8.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.