



# Open government data and the infodemic: a missing link in the fight against misinformation during COVID-19 in Italy

Federico Pilati<sup>1</sup> · Sonia Stefanizzi<sup>1</sup>

Received: 21 January 2025 / Accepted: 2 November 2025  
© The Author(s) 2025

## Abstract

This paper investigates the role of Open Government Data (OGD) in mitigating the “infodemic” that accompanied the COVID-19 pandemic in Italy. Focusing on Twitter discussions, we analyze how the public engaged with and debated pandemic-related data, exploring the impact of OGD availability on public understanding and the spread of misinformation. Our findings suggest that while OGD initiatives have the potential to empower citizens and combat misinformation, their effectiveness is contingent upon factors such as public awareness, data accessibility, and user-friendliness. By examining the Italian case, this research contributes to a deeper understanding of the interplay between OGD, public discourse, and misinformation resilience during public health crises.

**Keywords** Open government data · COVID-19 · Infodemic · Misinformation · Public debate

## 1 Introduction

The COVID-19 pandemic not only presented an unprecedented global health crisis but also sparked what the World Health Organization has termed an “infodemic”—an overabundance of information, both accurate and inaccurate, that makes it difficult for people to find trustworthy sources and reliable guidance. While considerable scholarly attention has focused on the spread of misinformation during the pandemic, less emphasis has been placed on examining the structural conditions that may facilitate or hinder the public's ability to distinguish reliable information from false claims. This paper argues that Open Government Data (OGD) represents a critical yet understudied factor in the fight against misinformation during health crises.

---

✉ Sonia Stefanizzi  
sonia.stefanizzi@unimib.it  
Federico Pilati  
federico.pilati@unimib.it

<sup>1</sup> Department of Sociology and Social Research, University of Milano-Bicocca, Milan, Italy

Italy presents a particularly compelling case study for examining the relationship between OGD and the infodemic. As one of the first Western countries severely impacted by COVID-19, Italy's experience in managing both the health crisis and the accompanying information challenges offers valuable insights. The Italian context is especially relevant because it combines several crucial elements: a significant impact of the pandemic's first wave, an active civic debate about pandemic-related data, and an evolving framework for government data transparency.

While existing research has extensively documented the spread of misinformation during the pandemic and various fact-checking initiatives, we argue that these approaches often overlook a fundamental aspect: the role of accessible, comprehensive, and well-documented government data in enabling public understanding and critical evaluation of health-related claims. The availability—or lack thereof—of OGD can significantly influence the public's ability to verify claims, understand trends, and participate in informed discourse about public health measures.

Our study examines this crucial intersection through an analysis of Twitter discussions about pandemic-related data in Italy. This approach allows us to understand how the public engaged with, questioned, and debated official data during the crisis. By analyzing these social media discussions, we can identify patterns in how the availability or absence of OGD influenced public discourse and the spread of misinformation.

Given the lack of widespread awareness and understanding of OGD among the general public, it is possible that discussions around data shifted towards a broader discourse, with many tweets referring to data sources without explicitly framing this issue in terms of OGD. This highlights the need to explore how the public perceives and engages with pandemic related public data, even if they may not be explicitly aware of OGD as a specific concept.

This research contributes to both theoretical and practical understandings of information ecosystems during health crises. Theoretically, it expands the infodemic framework by incorporating the role of OGD as a structural element that can either support or hinder public understanding. Practically, it offers insights into how governments might better leverage open data initiatives to combat misinformation during public health emergencies.

By focusing on Italy's experience, we aim to demonstrate how the absence of readily accessible and comprehensible government data created an information vacuum that may have contributed to the proliferation of misinformation. This case study provides valuable lessons for understanding the relationship between OGD, public knowledge, and misinformation resilience in crisis situations.

Building on this premise, our paper offers a twofold contribution to existing research. First, while previous studies have highlighted the technical and organizational barriers to effective OGD deployment, we focus specifically on how these limitations become critical under conditions of crisis, such as the COVID-19 pandemic. Second, by mapping the public discourse on Twitter, we empirically demonstrate how the absence of coherent data-sharing policies and communication frameworks can foster distrust and misinformation—extending the current understanding of OGD from a purely infrastructural issue to a socio-political one. Thus, this study bridges the gap between policy intentions and public perceptions, offering evidence that poor OGD design does not merely reduce effectiveness but actively contributes to the erosion of public trust.

## 2 The notion of open government data

The notion of Open Government Data is fundamentally rooted in the principles of transparency, participation, and collaboration, which form the bedrock of the broader open government movement. Janssen et al. (2012) define OGD as information collected, produced, or commissioned by government bodies that is made freely available for public use, reuse, and redistribution without restrictions. This definition encapsulates the essence of OGD, but to truly understand its implications, we must delve deeper into its characteristics and the philosophy behind it.

The Open Knowledge Foundation provides a more precise understanding of what constitutes truly open government data. They argue that for data to be considered open, it must meet several crucial criteria. First and foremost, the data should be readily available and accessible. This means that the entire dataset should be obtainable, preferably through internet download, at no more than a reasonable reproduction cost. The emphasis here is on removing barriers to access, ensuring that citizens, researchers, and businesses can easily obtain the information they need (Ubaldi 2013).

However, mere availability is not sufficient. The terms under which the data is provided are equally important. Open government data must permit reuse and redistribution. This includes allowing users to intermix the data with other datasets, fostering innovation and cross-sector insights. The philosophy here is that government data, funded by public money, should benefit the public in as many ways as possible.

Universal participation is another cornerstone of OGD. The data must be open to all, without discrimination against fields of endeavor, persons, or groups. This principle ensures that the benefits of open data are not limited to specific sectors or individuals but are available to society as a whole. It democratizes access to information, leveling the playing field and fostering inclusive innovation and participation.

The format of the data is also crucial. Open government data should be machine-readable, structured for automated processing. This requirement speaks to the practical use of the data in the digital age. Machine-readable formats allow for efficient analysis, visualization, and integration into various applications and services. It enables developers, data scientists, and researchers to work with the data at scale, uncovering patterns and insights that might be impossible to discern manually.

Timeliness is another critical aspect of OGD. The value of data often diminishes over time, particularly in fast-moving fields such as finance, healthcare, or environmental monitoring. Therefore, open government data should be made available as quickly as necessary to preserve its value. This requirement challenges governments to streamline their data publication processes, moving towards real-time or near-real-time data releases where appropriate.

Lastly, the principle of completeness rounds out the key characteristics of OGD. All public data should be made available, subject only to valid privacy, security, or privilege limitations. This principle pushes for comprehensive data release, avoiding selective publication that might skew public understanding or limit the potential uses of the data.

These characteristics, taken together, paint a picture of OGD as a powerful tool for fostering transparency, driving innovation, and enhancing public participation in governance. However, the journey from these ideals to practical implementation is complex and fraught with challenges.

Building on these principles, the concept of Open Linked Data (OLD) represents an evolution of OGD, aiming not only to make data accessible, but also interconnected and meaningful. By structuring datasets using semantic web technologies (e.g., RDF, URIs, SPARQL), OLD allows different datasets—across institutions, sectors, and countries—to be linked and queried in an interoperable way. This enhances machine-readability and enables web agents and AI systems to interpret, connect, and reason over data autonomously. In this sense, OLD strengthens the foundations of transparency and participation by turning isolated data points into a network of actionable knowledge, opening new avenues for civic innovation, policy analysis, and intelligent public services.

### 3 Open government data during the COVID-19 crisis

During the COVID-19 pandemic, the principle of Open Government Data (OGD) has shown significant gaps and limitations. Several studies have highlighted how inadequate data infrastructures and the poor quality of information released by governments have compromised the public health response and hindered effective management of the pandemic. First, the pandemic has highlighted issues of data inaccuracy and reporting delays, which have negatively impacted modeling and monitoring the spread of the virus. Delays in collecting and publishing public dataset have limited the ability of OGD systems to provide timely and accurate data, which is critical during a global health crisis like COVID-19. This has compromised the ability of governments to adopt timely and effective measures, contributing to worsening the pandemic situation rather than mitigating it. Furthermore, heterogeneous information management and differences between various government agencies have created further complications. The diversity and inhomogeneity of data shared by governments has often led to suboptimal results, rendering many of the public responses to the pandemic ineffective. These problems have been exacerbated by the failure of some governments to adequately manage and share information on cases and testing, making it more difficult for citizens and policymakers to make informed decisions. Another relevant criticism of open data during the pandemic concerns the use of incomplete or selective information, which contributed to a distorted perception of the reality of the health crisis. This has reduced the effectiveness of transparency tools and undermined public trust in governments and institutions. Delays, inhomogeneity and incompleteness of the data published by governments have compromised the effectiveness of the health response and have called into question the real usefulness of the OGD principles in crisis contexts. The COVID-19 pandemic has exposed deep weaknesses in the management of OGD infrastructures, raising questions about the effectiveness of the transparency principle these systems are supposed to promote. The OGD ideal is based on three main pillars: transparency, participation and collaboration. However, the global health crisis has shown that technical, organizational and legal limitations can seriously compromise the effectiveness of this approach.

In recent years, the academic literature on Open Government Data (OGD) has highlighted both the democratic potential and the structural challenges associated with their implementation. Studies conducted prior to the COVID-19 pandemic emphasized the role of OGD in promoting transparency, innovation, and civic participation (Janssen et al. 2012). However, these same authors also pointed to significant barriers to adoption, including poor interoperability, lack of standardized metadata, and the limited capacity of public admin-

istrations to effectively foster data reuse. Charalabidis et al. (2018) offers an authoritative synthesis of theoretical and practical approaches to OGD, emphasizing the importance of a governance ecosystem based on multilevel coordination, interoperable tools, and collaborative practices. Complementarily, Kaasenbrood et al. (2015) show that in the private sector, the propensity to adopt OGD is influenced by factors such as perceived data quality, regulatory clarity, frequency of updates, and trust in the data provider. The onset of the pandemic acted as a real stress test for public data systems. Arvisais-Anhalt et al. (2021) underline how the lack of accurate and timely data hindered the ability of health authorities to respond to the crisis in a coordinated way. Similarly, Zhang et al. (2023) draw attention to institutional fragmentation and the absence of shared infrastructure as key causes of inefficiencies in emergency management. These studies suggest that mere data openness is not sufficient to generate public value. Critical literature has emphasized the risk of symbolic “open-washing” practices (Longo 2011), in which declared transparency does not correspond to actual accessibility or usefulness of the data. Zuiderwijk et al. (2014) argue that the value of OGD depends not only on their availability, but also on their intelligibility, reliability, and ability to support participatory decision-making processes. The pandemic further exacerbated these dynamics. On the one hand, it incentivized the publication of epidemiological data in open formats; on the other, it revealed how disorganized public communication and uncoordinated data infrastructure can fuel uncertainty, distrust, and misinformation (Pilati and Anselmi 2023, 2024). Therefore, the use of OGD must be evaluated not only from a technical perspective, but within a broader framework of information governance, public communication, and institutional trust. These are precisely the dimensions to which this study seeks to offer an original contribution. During the COVID-19 crisis, Italy adopted a fragmented and reactive approach to data-sharing policies. The primary responsibility for collecting and disseminating epidemiological data was assigned to the Istituto Superiore di Sanità (ISS), while the Protezione Civile (Civil Protection Department) played a central role in aggregating and publishing national-level data. However, these datasets were often disseminated through static PDFs or machine-unreadable formats in the initial stages of the pandemic, limiting both real-time access and civic reusability (Miconi et al. 2024). Only later, under increasing public pressure and international comparison, did Italian authorities make partial improvements, such as the publication of the national COVID-19 dataset on the public data portal [dati.gov.it](<https://www.dati.gov.it>) and via GitHub repositories (notably by the Protezione Civile). Despite this, issues of delayed updates, missing metadata, lack of standardization, and unclear data provenance persisted, contributing to public confusion and undermining institutional credibility. Furthermore, the Italian open data strategy lacked integration with broader communication strategies. Open Government Data were often presented without adequate framing or contextual explanations, making them accessible in form but not in function. In this context, the failure was not merely technological but institutional, revealing a deeper misalignment between the principles of open government and their actual implementation during emergency governance.

### 3.1 Inaccuracy and delays in reporting

One of the major issues that emerged during the pandemic was the inaccuracy of data and significant delays in reporting it. Simone Arvisais-Anhalt et al. (2021) highlighted how the pandemic revealed a notable deficiency in the public health data infrastructure, which lim-

ited the ability to correctly monitor and model the spread of the virus. Late case reporting and a lack of standardization across local and national governments have hampered the ability of public health authorities to respond in a timely and accurate manner. This has made it difficult to predict peaks of infection and efficiently distribute healthcare resources.

In this study, Arvisais-Anhalt et al. describe how data quality and timeliness issues compromised epidemiological modeling during the early stages of the pandemic. This has not only made planning interventions more difficult, but has also undermined public trust in the data provided by governments, with serious repercussions on the credibility of COVID-19 management policies. Delays in data publication have also made the use of pandemic tracking and monitoring technologies less effective, tools that, with more accurate data, could have been used to contain the spread of the virus.

### **3.2 Inhomogeneity and lack of interoperability of data**

The lack of consistency between data published by different government bodies has been another significant obstacle during the pandemic. An investigation conducted by Xiaoqi Zhang et al. (2023) highlighted how the diversity of information and its non-uniform management have complicated intervention decisions, often leading to suboptimal results. For example, the lack of interoperability between various data systems has made it impossible to obtain a clear and integrated view of the pandemic situation at national and international levels. The result has been a fragmented and poorly coordinated management of the crisis, with some regions able to respond better than others simply thanks to a greater availability of updated data.

The study by Zhang et al. highlighted how, in the face of the pandemic, the diversity of approaches and the absence of a coherent regulatory framework for data sharing between local and national authorities have hindered an effective global response to the crisis. The consequences of this inhomogeneity have manifested themselves in a reduced ability to contain the spread of the virus in areas where data was not sufficiently timely or accurate.

### **3.3 Data manipulation and selectivity**

Another critical aspect of OGD management during the pandemic has been the selective or manipulative use of data by some governments. This has led to a distortion of public perception of the severity of the pandemic in certain contexts. Limitations in access to complete data, as noted by several researchers, have compromised transparency, partly defeating the very objectives of the OGD. When health data is incomplete or manipulated to support certain political narratives, the public's ability to properly evaluate response measures is severely compromised.

### **3.4 Criticisms of the effectiveness of OGD in health emergencies**

Several studies have highlighted that, although OGDs are often promoted as tools for accountability and transparency, their actual ability to improve the management of health emergencies has been significantly limited during the pandemic. Arvisais-Anhalt et al. (2021) and Zhang et al. (2023) agree that fragmented data architecture and organizational barriers have negatively impacted the ability to make informed and rapid decisions.

Although the concept of Open Government Data represents a progressive idea, promoting transparency and participation, the COVID-19 pandemic has exposed its serious shortcomings. The difficulties related to the quality of the data, the delays in their publication and the lack of interoperability between the different levels of government have demonstrated that the OGD principle, during a global crisis such as the pandemic, has not been able to guarantee the expected benefits. In particular, poor data quality has compromised the effectiveness of the public health response, raising doubts about the ability of OGDs to play a key role in managing future emergencies.

## 4 Research design

In the vein of Actor-Network Theory (ANT), we choose a quali-quantitative research design to empirically map the Twitter discussion related to Open Data during the pandemic in Italy (Venturini 2010; Venturini and Munk 2021). Starting from the encounter between ANT tradition and the discovery of digital methods, issue mapping has in fact made use of the digital traces and computational tools at its disposal in an innovative way (Venturini 2012; Marres 2015). distancing itself from a prescriptive model in the use of data science techniques, issue mapping seeks to use big data to recreate the complexity of interactional micro-processes among the involved actors, while at the same time it relies on the richness of ethnographic material left behind to interpret it (Venturini and Latour 2010).

We used an established data collection in issue mapping via digital methods (Marres and Weltevrede 2013; Marres and Moats 2015); namely, to collect all the tweets that contain words or hashtags related to conversations about data (i.e., *dato*, *dati*, *data*, etc. in Italian) and use Italian as their primary language. This way, the dataset collected through the official Twitter V2 search API amounts to 112,357 tweets, published between February 1, 2020, and July 31, 2020.

The choice of Twitter as the data resource for our mapping stems from three different motivations. The first reason is preemptively methodological and concerns the possibility of finding all the content circulated on the platform related to the topic of our interest. The second, on the other hand, concerns the desire to capture both top-down and bottom-up discussion, thus consequently we need to rely on social media in which interactions are based predominantly on this logic. Finally, we decided to capture tweets as opposed to Facebook posts because Twitter had grown considerably during the initial phase of the pandemic (e.g., +34% as for 2020 official report), mainly due to the influx of new users seeking an online arena where to discuss Covid-related topics (see Kwak et al. 2010 for an insightful description of Twitter as a news driven social media).

Initially to get a first glimpse on the general trends that the debate has taken in Twitter, we focused on its temporal evolution. To do this, we counted and plotted the number of tweets issued daily: in this way, on the one hand it was possible to observe the general pace of the debate (e.g., did the debate develop consistently over-time or did it advance by means of extemporal peaks?) while on the other hand it was possible to understand its interaction with the advancement of the OGD (e.g., did the introduction of new dataset trigger the debate?).

The second step in our research was related to a wide reconnaissance of the debate structure, to do this we decided to rely on two different strategies. A first metric concerns the average daily percentage of retweets and replies out of the total number of tweets: in this

way we had a rough measure on how much the conversation was based on the general production of original content or was instead driven by a few tweets. Secondly, we calculated the Gini coefficient on the concentration of retweets in order to understand the verticality or horizontality of the debate with respect to prominent influencers (Bracciale et al. 2018). Indeed, in our case the Gini coefficient measures the extent to which the distribution of retweets within our dataset deviates from a perfectly equal distribution. A coefficient of 0 expresses perfect equality where everyone has the same retweets, while a coefficient of 1 expresses full inequality where only one person has all the retweets.

Starting from the insights gained through the time series analysis we then moved our focus to community detection. In order to map the communities of users involved we have employed a pure retweet network (i.e., excluding mentions and comments): this entails an assumption, namely it presumes that retweeting something means, most of the times, an endorsement of the original tweet; while some Twitter users routinely state that 'RT is not endorsement' we do have substantial empirical evidence of the contrary, at least when it comes to extrapolate wider user communities (Metaxas et al. 2015).

Since a mapping of meaningful interactions requires different time windows (Venturini and Munk 2021), we decided to consider three different moments. The first corresponds to the days preceding and following the official introduction of the lockdown (from February 1st to March 31st), the second moment is the one of the reinforcement of restrictive measures during easter's holidays and their subsequent slow legislative weakening (from March 1st to May 31st), and the last one refers to the final abolition of the most restrictive measures for everyone (from June 1st to July 31st). This selection of time frames thus allowed us both to have three insights into different crucial moments in the evolution of the debate, and at the same time to follow in a balanced way its development over the six-month period.

For each of the selected time periods we extrapolated from our datasets all the retweets and we built three different directed networks. In these retweets' network, each node represents a user while an arc between two nodes, whose value is unitary, indicates a retweet of a specific tweet made by the user to whom the arc is directed. To identify the communities within the retweet networks we used the community detection algorithm called 'Louvain' (Blondel et al. 2008). This algorithm optimizes the modularity function, which measures the density of internal arcs of a single community compared to that of external arcs. To measure the permeability between each cluster of users identified, we relied instead on a similar parameter called 'E-I Index' (Krackhardt and Stern 1988), that allows to measure the openness or closure of each most prominent cluster. Indeed, the E-I parameter measures the portion of internal arcs, i.e., directed from one member to another of the same community, compared to the number of external arcs, i.e., directed from a member of a community to an external member. Using this measure, a node whose arcs connect only to nodes outside the community will have an E-I equal to +1 while, for a node whose arcs connect only to nodes inside the community, the E-I will be equal to -1. Since these extreme values are very rare in a real-world context, within the spectrum of values between -1 and +1, the E-I becomes a reliable measure of a node's tendency to connect to nodes within its community. From these measures, it is possible to convert the total number of internal and external arcs into a normalized index which, defined in this way, can represent a good estimate of how closed the communities are, i.e., how their structure is similar to an echo-chamber.

The operations described so far allowed us to map the macro and meso-structures of the interactions among the actors and users involved in the debate on Twitter, but at the

same time they still cannot describe in detail the opinions and positions of the very same. To make up for this lack, the last two choices made in our research design are focused on visual network analysis and qualitative content analysis. Let us anticipate that both choices are consequential again from the previous research outcomes: indeed, as we will illustrate in more detail in the results section, the open data debate developed in an extremely hierarchical way, with only a few and very influential accounts that centralize the information flow and consequently also few tweets able to dictate the overall agenda. Firstly, to have a more ethnographic kind of information related to the actors' positioning, we decided to filter our retweets networks by keeping only the most influential accounts. To visualize the reduced retweet networks, we used the open-source software Gephi (Bastian et al. 2009), a widely adopted tool in network analysis that allows for the interactive exploration and spatial layout of graph data. Gephi supports various layout algorithms and metrics for detecting patterns of interaction, centrality, and modularity in complex social networks. It proved especially useful for revealing the structural relationships and influence hierarchies among Twitter users in our dataset. The images shown in the results section were obtained by applying firstly the visualization algorithm Force Atlas 2 to the entire retweet network (Jacomy et al. 2014) and then by removing the nodes whose sum of outgoing arc weights is less than 20 (that is the suggested threshold to maximize the reading of our network maps without losing relevant nodes spatialization). By means of these reduced samples, and by keeping node labels only for public Twitter accounts, we were therefore able to navigate and describe in detail the role and relation of politicians, journalists, health experts, and legacy media. Finally, to gain a deeper insight into the viral contents that drove the debate, we extrapolated and analyzed the content of the 2,500 most retweeted tweets.

## 5 Results

The temporal analysis of Twitter discussions reveals a declining trend in engagement over time. After an initial surge of interest in data-related issues, the debate gradually diminished, with only occasional spikes in tweet activity coinciding with legislative decrees in May and June that lifted restrictions. Regarding the concentration of the debate, the analysis confirms a typical Twitter pattern where conversations are dominated by a small number of voices. This concentration is quantitatively demonstrated by the Gini coefficient for retweets, which reaches a high value of 0.77, indicating that the majority of retweets were concentrated among a select few users, while most participants received minimal engagement (Fig. 1).

The analysis of the retweet network from 1st February to 31st March, visualized in Fig. 2, reveals a notable fragmentation in the public discourse, with the Louvain algorithm identifying four distinct communities of conversation with a negative E-I ranging from  $-0.01$  to  $-0.34$ . While these clusters partially align with the traditional right vs. left wings political affiliations, the underlying narrative transcends purely partisan divisions, converging predominantly on the persistent lack of reliable data in Italy's pandemic response. This frustration regarding data reliability emerges as a unifying theme across clusters, as evidenced by the most frequently shared content. A comparative international perspective resonates strongly in the network, with one highly circulated tweet lamenting: "Looking at how other countries manage their COVID data makes me wonder what we're doing in Italy. Without reliable numbers, we're navigating blindly." This sentiment reflects a broader concern about

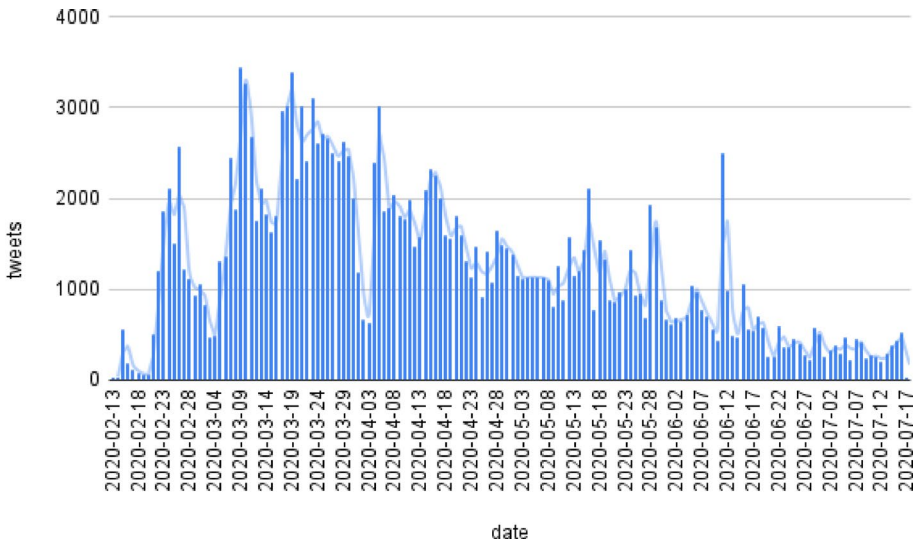


Fig. 1 Numbers of tweets per day

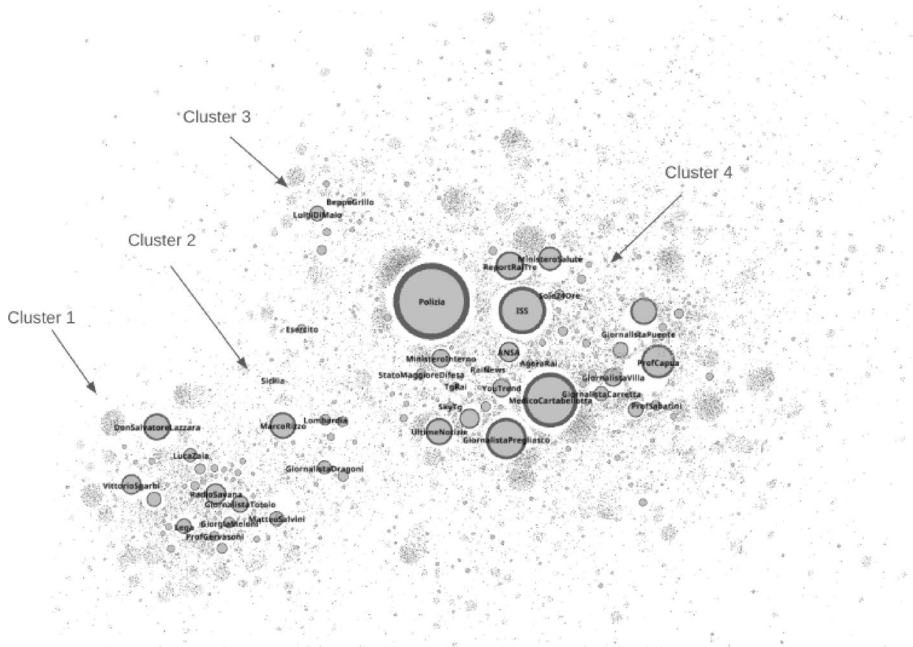


Fig. 2 retweets network visualization from February 1st to March 31st

Italy's efforts in the global context of pandemic management. The analysis also highlights significant engagement with tweets criticizing institutional transparency. A particularly influential voice in the network observes: "The management of public health data in Italy is symptomatic of a broader problem: opacity in public institutions. We can't make informed decisions if we don't even know what data we have and how it's being collected." This tweet underscores how the data issue connects to deeper structural concerns about institutional accountability. Further reinforcing these themes, another widely shared tweet emphasizes the practical implications of data deficiency: "Other countries are using data to manage their pandemic response effectively. Meanwhile, in Italy, we're still struggling with basic data collection and sharing. How can we expect to control the spread if we can't even measure it properly?" This perspective highlights the crucial link between data quality and effective public health response. The network structure thus reveals how, despite fragmentation, the Italian public discourse coalesces around a shared critique of data management practices, emphasizing the fundamental importance of reliable data for effective pandemic response and policy-making.

The analysis of the retweet network from early April to the end of May 2020, visualized in Fig. 3, reveals a significant shift in the public debate, with the emergence of two distinct communities characterized by growing distrust in government data. While Open Government Data platforms were starting to gain traction, their impact on public discourse remained limited, overshadowed by the overwhelming volume of other pandemic-related content. This period witnessed a rise in narratives of data manipulation and withholding by government entities, particularly within clusters dominated by right-wing and conservative users. A prominent tweet circulating within these clusters stated: "They're hiding

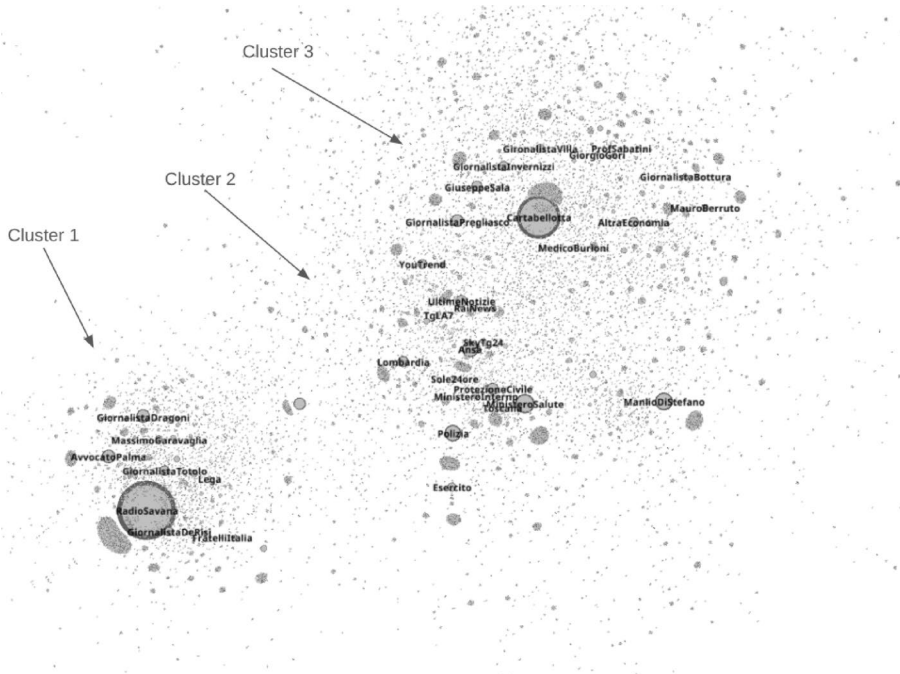
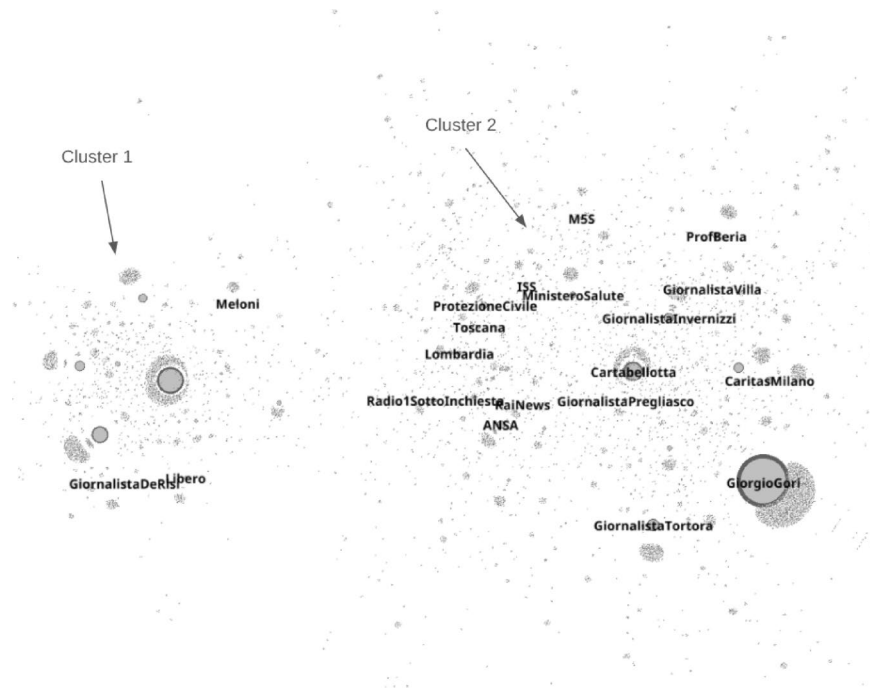


Fig. 3 retweets network visualization from April 1st to May 31st



**Fig. 4** retweets network visualization from June 1st to July 31st

the real numbers. This is not about public health, it's about controlling the people". The network analysis highlights the emergence of a larger, more cohesive group amplifying these theories. Hashtags such as #NewWorldOrder became central to these discussions, framing the lack of transparency as a deliberate attempt to control the population. A highly influential tweet within these networks asserted: "The data is being manipulated to justify their draconian measures. We need to demand transparency and accountability". This statement underscores the erosion of trust in government institutions. Furthermore, content analysis reveals a significant focus on the political motivations also on the opposite larger counterpart made of journalists and health experts clusters. A widely shared tweet within these clusters claimed: "They're using the pandemic to push their own agenda. Conspiracies about data are just a piece of the puzzle." This perspective emphasizes the belief that right-wing parties and politicians were exploiting the crisis to advance their own political objectives, further fueling polarization. This sense, the network structure—with a growing negative E-I index in all the three communities detected by the Louvain algorithm ranging from -0.23 to -0.67—reveals how, despite the initial promise of Open Government Data platforms, the discourse during this phase became increasingly polarized and characterized by a deep-seated distrust in government data and institutions. This erosion of trust had significant implications for public health efforts, undermining the effectiveness of public health guidelines and contributing to a climate of fear and uncertainty.

Finally, the analysis of the retweet network from June 2020 onwards, visualized in Fig. 4, reveals a notable shift in the discourse, with a gradual increase in the availability of Open Government Data. While public institutions and social organizations, such as the civil pro-

tection, started to release more structured datasets, the impact on public discourse remained relatively limited. Despite the increased availability of data, discussions around OGD remained on the periphery of the broader conversation surrounding the pandemic. A prominent tweet circulating within the network observed: "We need better tools and resources to make sense of it all (i.e., using data to understand the pandemic)". Furthermore, the emergence of alternative conspiratorial narratives, such as those surrounding vaccine development and distribution, further marginalized the discourse on OGD, particularly within the conservative and right-wing spheres. The polarized network structure made up of two communities with a high negative E-I index—namely -0.44 and -0.54—thus reveals that despite increased efforts to improve data availability, the discourse on OGD remained relatively limited in scope and impact.

## 6 Discussion

Our analysis of Twitter discourse throughout the COVID-19 pandemic reveals a clear temporal evolution in the conversation surrounding Open Government Data, particularly in relation to the broader infodemic. Through the retweet network analysis and the correspondent content analysis of the most popular tweets within the identified communities, we observed three distinct phases in the public discussion between February 2020 and the end of July 2020. These phases corresponded with shifts in both the types of data being discussed and the ways in which Open Government Data became entangled with misinformation and conspiracy theories.

### 6.1 Phase 1: February–March 2020: the absence of data

In the initial phase, spanning from February to the end of March 2020, the conversation around COVID-19 was largely characterized by a widespread absence of reliable data. The public discourse on Twitter during this period was dominated by anxieties related to the lack of official information. Tweets primarily focused on the absence or inconsistency of official data, with many accounts expressing frustration at the insufficient or unclear data being made available by government and health authorities. The conversation was mostly centered around a sense of disarray and opacity in governmental responses.

Within this context, the discourse was highly fragmented, with multiple small clusters of users amplifying concerns about the perceived lack of transparency. Popular tweets during this phase typically expressed confusion about the virus's spread and the lack of clear, reliable metrics to understand the scale of the crisis. While some users questioned the reliability of official data, there was no organized discussion of Open Government Data per se; rather, the issue was framed as a broader question of government failure to provide meaningful and accessible data, if not in the form of its daily bulletin.

Moreover, this period saw the first, early emergence of various populist and fringe voices, who used these accusations to fuel conspiracy theories about the virus's origin, its spread, and the role of international organizations like the World Economic Forum.

## 6.2 Phase 2: April–May 2020: the rise of conspiracy theories

The second phase, from early April to the end of May 2020, marked a significant shift in the Twitter conversation. Although Open Government Data platforms were starting to gain traction, particularly in academic and policy circles, the general public had not yet fully embraced these resources. The engagement with OGD was also constrained by the overwhelming volume of other COVID-19-related content, which continued to dominate social media.

During this period, in the cluster led by right-wing and conservative users, attention increasingly turned towards the perceived manipulation or withholding of data by governments, accompanied by rising concerns over political interference. Notably, discussions around data became more polarized, with a growing narrative of distrust in official statistics and data releases. This period saw an explosion of conspiracy theories linking the lack of data to deliberate governmental actions or political agendas.

Our network analysis in this phase highlighted the emergence of larger, more cohesive networks of users who amplified these theories. Hashtags such as #NewWorldOrder became central to these discussions. The conversation was often framed as a deliberate attempt by governments to obscure the true scope of the pandemic, either for political gain or to control the narrative around the virus. Within these networks, content analysis revealed that many of the most retweeted tweets were not just about the absence of data but also accused governments of manipulating the available data to suit their own interests, undermining public trust in health guidelines and lockdown measures. The result was a profound erosion of trust in both the data and the institutions that were supposed to be responsible for public health management.

## 6.3 Phase 3: June–July 2020: low-volume engagement

By June 2020, the availability of more comprehensive Open Government Data began to increase, as public institutions and social organizations, such as the civil protection, started to release more structured datasets in an attempt to provide greater transparency. This period saw the first substantive ex-post emergency discussions related to the availability of COVID-19 case data, testing statistics, and recovery rates. However, despite the increasing availability of data, the conversation around OGD remained relatively limited in scope and volume.

The network analysis during this phase revealed that while there were still active discussions around the role of data in combating the pandemic, these conversations remained on the periphery compared to broader debates about the virus and its management.

The shift to alternative conspiratorial narratives, such as the one on vaccine development and distribution, meant that the discourse on OGD struggled to gain a foothold also within the conservative and right-wing's area.

## 7 Summary

The analysis highlights a critical paradox in the COVID-19 pandemic's relationship with Open Government Data: the very absence of a coherent discourse on OGD itself ultimately contributed to the mythology of OGD as a panacea for the infodemic. At the outset of the pandemic, the absence of reliable and accessible data became a central theme of public discourse. The lack of transparency was initially framed as a failure of governmental institutions to provide adequate information, which seeded the first layer of distrust and confusion. This absence of data was not merely a logistical issue, but a symbolic void that allowed misinformation and conspiracy theories to take root, further fracturing public understanding of the pandemic.

In the second phase, as accusations of data manipulation and political interference escalated, the narrative around data shifted into the realm of speculation and distrust. Rather than a call for more open and accessible data, public discourse morphed into demands for transparency that were framed through the lens of political agendas and the alleged deliberate suppression of information. Here, OGD was no longer an idealistic solution to the information crisis; instead, it became part of a larger narrative of governmental control, aligned with conspiracy theories that further fueled the infodemic.

Finally, by June 2020, when formal Open Government Data platforms began to consolidate, the conversation around their possible use lacked the widespread engagement necessary to counterbalance the pervasive spread of misinformation. The initial momentum for Open Government Data as a solution to the infodemic had been stymied by the preceding months of distrust and conspiracy, rendering the discourse on OGD almost an afterthought in the broader pandemic narrative.

These findings suggest that the absence of a clear and coherent public conversation on Open Government Data, coupled with its instrumentalization in broader political and conspiratorial discourses, ultimately contributed to the mythologization of OGD as a cure-all for the infodemic. OGD, far from being an antidote to the crisis of misinformation, were subsumed into the very narratives that fueled the pandemic's information crisis. In this sense, the discourse on OGD itself became an integral part of the infodemic, highlighting the complex relationship between data transparency, public trust, and misinformation.

## 8 Conclusion

The COVID-19 pandemic has highlighted the structural and functional limitations of Open Government Data, exposing a central paradox: while OGD is designed to promote transparency, participation, and collaboration between citizens and institutions, its implementation during the health crisis often yielded opposite effects, fueling misinformation and distrust.

This paradox stems not only from technical shortcomings, such as delays or a lack of standardization in publication, but also from ineffective management that failed to transform data into comprehensible and useful tools for the public. One of the most evident issues was the connection between data quality and the infodemic—the massive spread of unverified or misleading information.

In the initial stages of the pandemic, the absence of timely and reliable data created an informational void that facilitated the emergence of conspiratorial narratives. This vacuum

contributed to eroding public trust in institutions, widening the gap between citizens and their governments. The technical availability of data, while important, proved insufficient to counter these dynamics. For Open Government Data to become an effective resource, it must be embedded in strategies that enhance its accessibility and social relevance. Technical transparency—defined as the publication of accessible and interoperable data—must evolve into social transparency. This requires an approach that makes data not only available but also interpretable and meaningful for the public.

During the pandemic, the politicization of data underscored the risks of fragmented and instrumental use of information, exacerbating ideological divisions and polarizing public debate. In this context, institutions must go beyond merely providing quality data; they must adopt strategies that promote a shared narrative and build trust. Crafting accessible yet robust narratives around OGD that explains their significance and highlights the related social utility can bridge the gap between citizens and institutions, fostering a culture of participatory transparency.

The Italian case demonstrates that data must be considered part of a broader social and political ecosystem. Their effectiveness lies in their ability to integrate into a system that fosters dialogue between institutions and citizens, creating the conditions for mutual trust. This means rethinking the role of OGD not just as technical tools but as key elements of democratic governance that is genuinely inclusive and participatory. The pandemic experience underscores that the potential of OGD cannot be realized without significant changes in their management and communication. Rather than merely publishing information, institutions should invest in strategies that make data relevant and useful for the public, enhancing their impact on public discourse. This approach requires a long-term vision, where data becomes an integral part of a continuous process of dialogue and trust-building between governments and citizens.

In conclusion, the pandemic has shown that poorly managed data risks contributing to the same dynamics of misinformation it aims to combat. Only by integrating technical transparency, effective communication, and public participation can OGD become a powerful tool to strengthen social resilience and consolidate democratic governance based on transparency and participation. Thus, the crisis serves both as a warning and an opportunity: transforming open data into a cornerstone of democracy demands a collective effort to combine technology, communication, and public engagement into a shared vision of trust and collaboration.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arvisais-Anhalt, S., Lehmann, C.U., Park, J.Y., Araj, E., Holcomb, M., Jamieson, A.R., Basit, M.: What the coronavirus disease 2019 (COVID-19) pandemic has reinforced: the need for accurate data. *Clin. Infect. Dis.* **72**(6), 920–923 (2021)
- Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Web Soc. Media* **3**(1), 361–362 (2009)
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**(10), P10008 (2008)
- Bracciale, R., Martella, A., Visentin, C.: From super-participants to super-echoed: participation in the 2018 Italian electoral Twittersphere. *Partecipazione e Conflitto* **11**(2), 361–393 (2018)
- Jacomy, M., Venturini, T., Heymann, S., Bastian, M.: Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **9**(6), e98679 (2014)
- Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag.* **29**(4), 258–268 (2012)
- Kaasenbrood, M., van Eijk, A.M.G., Janssen, M., de Jong, M., Bharosa, N.: Exploring the factors influencing the adoption of open government data by private organisations. *Int. J. Public Adm. Digit. Age* **2**(2), 75–92 (2015)
- Krackhardt, D., Stern, R.N.: Informal networks and organizational crises: an experimental simulation. *Soc. Psychol. Quart.* (1988). <https://doi.org/10.2307/2786835>
- Kwak, H., Lee, C., Park, H., & Moon, S.: What is Twitter, a social network or a news media?. In *Proceedings of the 19th International Conference on the World Wide Web*, pp. 591–600
- Longo, J.: Open data: digital-era governance thoroughbred or new public management Trojan horse? *Public Policy Forum* (2011)
- Marres, N., Weltevrede, E.: Scraping the social? Issues in live social research. *J. Cult. Econ.* **6**(3), 313–335 (2013)
- Marres, N.: Why map issues? On controversy analysis as a digital method. *Sci. Technol. Hum. Values* **40**(5), 655–686 (2015)
- Marres, N., Moats, D.: Mapping controversies with social media: the case for symmetry. *Social Media+ Society* **1**(2), 2056305115604176 (2015)
- Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M., and Finn, S.: What do retweets indicate? Results from user survey and meta-review of research. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 9(1), pp. 658–661, (2015)
- Miconi, A., Pezzano, S., Pilati, F., Risi, E.: As the infection curve shows». Data-fetishism and reification of the medical discourse on Facebook during the Covid-19 epidemic. *Rass. Ital. Sociol.* **65**(4), 831–862 (2024)
- Pilati, F., Anselmi, G.: The AstraZeneca affair. How the junk news regime affected the public debate on the COVID-19 vaccination controversy in Italy. *Tecnoscienza-Italian J. Sci. Technol. Stud.* **14**(1), 105–122 (2023)
- Pilati, F., Anselmi, G.: From the «rally 'round the flag» to the emergence of a (Counter)«digital movement of opinion». A quali-quantitative analysis of Twitter debate on the COVID-19 «lockdown» in Italy. *Comunicazione Politica* **25**(1), 61–84 (2024)
- Ubaldi, B.: Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, No. 22 (2013)
- Venturini, T.: Diving in magma: how to explore controversies with actor-network theory. *Public Underst. Sci.* **19**(3), 258–273 (2010)
- Venturini, T., and Latour, B.: The Social fabric: digital footprints and quali-quantitative methods. In: *Proceedings of futur en Seine*, pp. 87–101, (2010)
- Venturini, T.: Building on faults: how to represent controversies with digital methods. *Public Underst. Sci.* **21**(7), 796–812 (2012)
- Zhang, X., Fu, J., Hua, S., et al.: Complexity of government response to COVID-19 pandemic: a perspective of coupled dynamics on information heterogeneity and epidemic outbreak. *Nonlinear Dyn.* **111**, 22055–22074 (2023)
- Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., Ferro, E.: *The World of Open Data: Concepts, Methods, Springer, Tools and Experiences* (2018)
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R.: Socio-technical implications of open data: barriers and potential. *Gov. Inf. Q.* **31**(1), 98–108 (2014)