Examining the Impact of Transparency Portals on Media Coverage: Insights from a High-Profile Case of Public Procurement Irregularities in Spain

Jesús García-García

ORCID Nr: 0000-0002-5120-8851
School of Economics and Business, University of Oviedo, Spain, jesgar@uniovi.es

María Isabel Alonso-Magdaleno

ORCID Nr: 0000-0002-4865-1561
School of Economics and Business, University of Oviedo, Spain, ialonso@uniovi.es

Abstract: This study examined the use of data from a transparency portal in media coverage of a high-profile case of alleged public procurement irregularities in Spain. Access to Twitter API was used to identify relevant URLs related to this issue. It was found that direct links to a portal were of low relevance, and most of the linked documents did not even mention the availability of data from a portal. Qualitative analysis revealed that the most frequent topics were the use of portal data as an authoritative argument to endorse information, the statement that the portal did not contain sufficient information for journalistic purposes, and the absence of data on third parties involved in public procurement. It is recommended that governments promote the existence of Portals and make media outlets aware that providing links to original data is beneficial for their reporting. In addition, linked open data should be used to ensure accuracy and transparency.

Keywords: Transparency, accountability, open data, Twitter, media

1. Introduction

The open government paradigm has established the need for continuous and universal transparency in public management, through the release and opening of data, to exercise accountability, in accordance with the needs and desires of a citizenry immersed in a digital world. It is based on closing the gap generated by the disaffection between governments and citizens that arose during the global financial crisis that began in 2007-2008 and which resulted in a parallel crisis of the representative model derived from the distrust of citizens in the intermediary agents of the democratic system, mainly political parties and mass media (Arpino and Obydenkova, 2020; Bértoa and Rama, 2020; Rico Motos, 2019). This would be a matter of enriching classic horizontal accountability models with

new models of a vertical nature, as defined by O'Donnell (1994, 1998). In the horizontal model, administrations set their own control conditions, which are implemented by their bodies. Their design, which is not oriented towards communication with citizens but rather towards mere technical control by administrative experts, makes their information distant and difficult to understand for society, which ultimately receives it through intermediaries outside governments and administrations, such as the media and data activist groups (Fuchs, 2010; Gutiérrez, 2018).

In this way, digital tools such as transparency portals, open data portals, or even classic institutional websites (hereinafter, all bundled under generic 'Portals'), would serve as a permanently open and freely accessible communication channel that would reduce information asymmetries between the government and the public (Janssen et al., 2017; Matheus and Janssen, 2029). The latter, either directly or through data intermediaries, would be able to access the information on public administration necessary to assess the actions of those in power or to check the information offered by different media outlets. Ideally, these media would link their information to the source of the data used to prepare it. Thus, not only would transparency and accountability be boosted, but also, the information provided by media outlets and data intermediaries would gain in accuracy as the data could be easily cross-checked. Moreover, large technology companies, mainly in the social media field, act as proxies for the distribution of this information, from media intermediaries to citizens. If businesses, such as Google, Facebook, or Twitter do not want to do so, citizens may not have access to relevant information to know the reality and demand transparency and accountability (Rudschies, 2022; Zuboff, 2019).

In this field, many studies focus simply on the degree of data and information disclosure by governments and public administrations to assess transparency and accountability. However, looking only at the supply side can overlook the fact that accountability is nothing more than a communicative process that requires a provider but also a receiver, and that is affected by the issues previously described. In this process, the media acts as a communication channel between the disseminators and final recipients of accountability. Therefore, we decided to study the case from a communicative perspective on digital platforms and social networks and not on the volume of data and information disclosed in the Portals, because in the end, it is this perspective that conditions citizens' satisfaction with transparency and accountability.

This leads us to ask the following research questions, with the aim of determining how the information contained in Portals is actually used in situations of public debate, which we intend to answer in this article:

RQ1: What is the level of direct reference to Portals in public debate?

RQ2: What are the key features of the main documents linked to in public debate?

RQ3: What are the main themes associated with links or citations to Portals as a source of data and in what context do they occur?

The remainder of this paper is organised as follows. Section 2 presents the literature and issues relevant to the analysis, specifically related to the socio-technological context of the research questions. Section 3 presents the methodology followed for the case study and the data collection process.

This is followed by section 4, which presents the results. Finally, section 5 concludes the paper with a discussion of the results and final remarks.

2. Background

This section describes the main issues associated with the disclosure of public information on Portals, which can be used as a source of information for media outlets and later disseminated through media outlets.

Since its inception, the open government paradigm has not been immune to criticism, particularly on its operations. Heald (2012) pointed to unfulfilled expectations about the actual use of the information made available to citizens, due to the low number of users accessing it. Meijer (2013) remarked on the confusing relationship between transparency and accountability, in that, while there appears to be a close relationship between the two, academic literature and case studies have not been able to provide a comprehensive assessment of their effects. Worthy (2015) noted the need to facilitate the accessibility and contextualization of data, for ordinary citizens in order for transparency and open policies to be effective. Gray (2015) pointed out the lack of rigorous analysis of the demand for data and the lack of knowledge about how, for what purpose, and for what purpose it is used. According to Winsvold (2007), any information disclosed on a governmental website, influences the public political agenda by being a source for the press, but it is uncertain whether they contribute to the pluralism of the public debate. In this regard, Lourenço (2016) and García-García and Alonso-Magdaleno (2022) suggest that Portals have been used to support argumentation in public policy debates. Similarly, Crepaz and Kneafsey (2021) stated that despite abundant evidence that media are major users of Portals, their role in data access and dissemination has not been studied. In contrast, Díez-Garrido and Campos-Domínguez (2018) pointed out their scarce use in media. This contradictory result can be reconciled with the statements of the authors of both studies regarding the limited usefulness of the Portal contents for the production of journalistic information.

In this regard, with particular reference to Spain, the scenario of the case study to be addressed, García-García and Curto-Rodríguez (2018), Beltrán-Orenes and Rodríguez-Mateos (2020) and Curto-Rodríguez (2020) point to the wide range and expansion of transparency and open data portals in different administrations, but, at the same time, questions their real information content. The phenomenon described as *openwashing* by Heimstädt (2017) is manifested as the disclosure of data to convey the appearance of modernity, but, as data with no apparent value, incomplete or partial, which would avoid the rigorous substantiation of research, turning the Portal into a usable source for the media, but without great intrinsic worth adding value to the journalistic piece.

Nevertheless, the emergence of fake news and disinformation has challenged and called into question the ability of open data to guide the evaluation of public policies in the general citizen debate. Disinformation includes any form of falsehood, inaccuracy, or misleading content designed, presented and disseminated to harm or create profit (European Commission. Directorate General for Communications Networks, Content and Technology, 2018). The concept of fake news is simply a subset of this information, which refers to the falsity of the content produced. Today, more open views identify the problem of disinformation as a process, referred to as, information disorder

(Wardle & Derakhshan, 2017), and not as an information product. Numerous studies have highlighted the importance of understanding technical and social changes that allow disinformation to act more effectively (Andrejevic, 2020; Aricat, 2018; Gaufman, 2018; Lewandowsky et al., 2017, 2020; Young, 2021).

In this process of information disorder, one of the main objectives is not the dissemination of disinformation per se, but the destabilisation of democratic institutions through the discrediting of messages that are systematically called into question, thus, affecting democratic environments and citizen decision-making and generating significant adverse consequences that drive citizen disaffection (Lewandowsky et al., 2017). Most of these processes are executed by agents that mimic the content and professional appearance of news media, but without their editorial standards and internal processes to ensure the veracity and accuracy of information, parasitising the credibility of legitimate media (Lazer et al., 2018). In this vein and in an environment of information saturation, many disinformation strategies and tactics can involve *data-driven bullshit* (Bergstrom & West, 2020), where information is accompanied by decontextualised data or graphics, because they are generally given a higher objective value than words, that help to convince the receiver of the message and to take advantage of the phenomenon called *anchor effect*; a cognitive bias that describes the common human tendency to rely too heavily on the first piece of information offered when making decisions (Kahneman, 2012). This phenomenon can be amplified by how the public deals with the complex media ecosystem; on the one hand, they look for media and interpersonal sources of authority to verify information, on the other hand, they avoid certain sources or consciously seek entertainment (Valera-Ordaz et al., 2022; Wenzel, 2019). We highlight this last aspect, since humour on the Web, whether in the form of text or meme, can constitute a symbolic micro-action that attracts the attention of users and the media, making a substantial contribution to shaping public opinion (Mina, 2019). Moreover, even if users are aware of the falsehood of a news item or piece of information, its playful nature can act as an incentive for redistribution (Madrid-Morales et al., 2021). For these reasons, the simple fact in a headline or image disseminated through a social network can be a perfect vehicle for disinformation.

Several studies attribute benefits to social networks, such as being a source of information diversity and helping increase levels of political participation, understanding, and expression (Boulianne, 2019; Fletcher & Nielsen, 2018; Yang et al., 2020). At the same time, there are also many who attribute shortcomings to them, such as attempting to modify the public agenda, being a vehicle for disinformation campaigns, or driving hate campaigns against groups (Allcott & Gentzkow, 2017; Grinberg et al., 2019; Matamoros-Fernández & Farkas, 2021; Stella et al., 2018; Vargo et al., 2018). Social media consumption is an absorbing activity that generates significant emotional reactions, with many of the characteristics of addictive activities and, in its most extreme form, has been described as a behavioural addiction (Andreassen, 2015; Kuss & Griffiths, 2011). Moreover, disinformation is incentivised by lucrative business in which large players in the digital sector indirectly fund disinformation by paying advertising revenue to clickbait site operators (Hao, 2021). This conjunction between disinformation and advertising interests also points to the phenomenon of *micro-targeting*, which, as the Cambridge Analytica scandal demonstrated in the 2016 US presidential election, allows a political campaign to exploit emotions and heuristic biases to play on the weaknesses and

fears of the groups most prone to information manipulation (Berghel, 2018). This effect would extend to the entire population regardless of age, because while it is the younger sectors of the population that give greater credibility to social networks, older sectors form credibility based on trust criteria of those who receive information through second-generation networks of a more private and reduced character, such as WhatsApp or Telegram; they would therefore, be more susceptible to being victims of disinformation that their circle of trust has found on social networks and disseminated through messaging services (Valera-Ordaz et al., 2022). Consequently, disclosure on social media should be understood as an opportunity and a problem for the population as a whole, not only for the older sectors. The transparency and accountability of governments and public administrations are not unaffected by all these considerations, since, in one way or another, a large part of the population uses social networks and digital platforms to discover current affairs affecting their country, region, or municipality.

Our understanding of the scale, scope, and nature of issues affecting public debates on digital platforms and social media remains very limited (European Commission, 2020; Majó-Vázquez et al., 2021). Although media outlets use social media to generate traffic on their websites, very little is known about how a social media conversation can impact the viewing of a news story (Gabielkov et al., 2016). Digital platforms and social media are different places and it is difficult to treat them as a single entity. People who access a social network are sociologically very different from those who access another social network and there will also be differences at the geographical level. Combining their study with a single sample can be complicated. For this reason, we considered that using links to news and other websites that are disseminated on them may be a solution for analysing the use of Portals in public debate. Moreover, in a platform with a high media impact, such as Twitter, which would not only have an incidence on the opinion of users but also its volume of traffic would condition editorial decisions linked to the advertising impact that a news item could achieve in the future (Cagé et al., 2022).

3. Method and data collection

Wardle & Derakhsan (2017) identified three basic elements of information clutter: agent, message, and interpreter. Agents are individuals or entities that create, produce and distribute pieces of information, according to their own motivations. A message is information transmitted in a particular format and features. The interpreter is the receiver of the message and, according to its interpretation, it will take certain actions. Along with these elements, they identified three phases: creation, production and distribution. In the creation phase, the message is elaborated, then in the production phase, it is turned into a media product and finally in the distribution phase, it is circulated or made public. Elements and phases can be combined in multiple ways, for example, the agent that creates the content is often different from the agent who produces it. Our research focuses on message and production, that is, how the data contained in the transparency portal are included in news reports and other documents that may be disseminated by Twitter.

Twitter, a free microblogging service launched in 2006, enables users to communicate through short messages that contain up to 280 characters. This platform facilitates rapid access to information

for individuals possessing Twitter accounts and Internet connectivity through smart devices or computers. Twitter has emerged as a prominent force in the dissemination of information and opinions owing to its extensive user base and connectivity. Its popularity stems from a distinctive feature that allows users to gain insight into other users without establishing a formal connection. By following an individual on Twitter, users can promptly access that person's profile without necessitating permission, follow-back, or even awareness. Furthermore, unlike other social networks such as Facebook, numerous Twitter pages are openly accessible and do not mandate sign-up processes. This attribute has significantly contributed to Twitter's popularity, establishing it as a highly influential platform for distributing information and opinions.

The utilisation of Twitter as a valuable source of information has already yielded successful outcomes in the examination of political discussions (Majó-Vázquez et al., 2021; Stella et al., 2018; Bastos and Mercer, 2017; Majó-Vázquez et al., 2017), the analysis of networked citizen politics and political participation (Peña-López et al., 2014), the exploration of fake news distribution, exposure and engagement (Grimberg et al., 2019; Vosoughi et al, 2018), the identification of influential actors in disinformation campaigns (Smith et al., 2021) and the investigation of the dissemination of political content beyond text, such as images (Martínez-Rolán and Piñeiro-Otero, 2016) and links (Gabielkov et al., 2016). Considering this background, we deemed it appropriate to design our research based on a case study methodology, given that our primary research questions revolve around the utilisation of information, we lack control over behavioural events and our focus centres on a contemporary phenomenon (Yin, 2018, p. 32 ff.).

The study is based on the analysis of a case of procurement of face masks by the Community of Madrid, one of the 17 autonomous communities of Spain and the third most populous, with almost 7 million inhabitants mostly concentrated in the metropolitan area of Madrid. It involved the collection of brokerage commissions by the brother of the president of the Community of Madrid, Isabel Díaz-Ayuso. The commissions came from contracts to import face masks from China for €1.5 million between March and April 2020. Owing to the special circumstances produced by the COVID19 pandemic, contracts could be established by handpicked single-source procurement between the company Priviet Sportive SL and the government of the Community of Madrid. The case under study achieved outstanding notoriety in the Spanish media in the third week of February 2022 because of its strong political implications at the regional and national levels and was based on evidence obtained from the transparency and procurement Portals of the Community of Madrid. This notoriety and its containment in time, together with the existence of documentation related to the issue in official Portals, turn the case into an opportunity to analyse the use of the data contained in the Portals. This is the focus of our study, which, in no way, judges alleged irregularities and responsibilities.

The database of tweets for the analysis was created using a Boolean search that combined subjects, objects and language in a manner that allowed for high precision in the selection of tweets. Figure 1 outlines the search criteria in Spanish. To improve the search, in the first block corresponding to subjects, we added the terms 'CAM', an abbreviation of 'Comunidad Autónoma de Madrid' widely used in written language, and 'Sol', referring to the location of the seat of government of the

Community of Madrid in the iconic Puerta del Sol building in Madrid, and which is often used to refer to the regional government (e.g., the government of Sol).

Figure 1: Twitter query
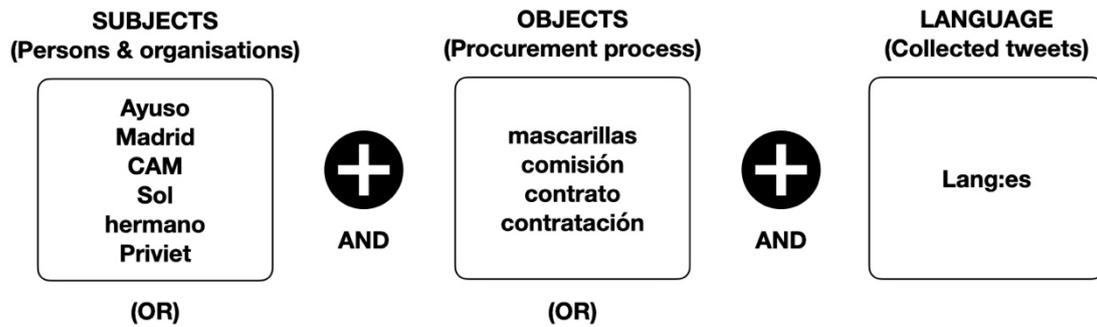
Translated into English, the search criteria are expressed as:

'(Ayuso OR Madrid OR cam OR Sol OR brother OR Priviet) AND (facemasks OR commission OR contract OR procurement) AND (lang:es)'

Through academic access to version 2 of Twitter's Application Programming Interface (API), and using the Twarc2 software, 815,527 tweets were extracted referring to the case study and dated between 16 February and 15 March 2022, corresponding to 148,852 unique user accounts. Unlike studies such as those by Majó-Vázquez et al. (2017, 2021) and Peña-López et al. (2014), we opted to access the Twitter historical archive (Gnip) and not to use the Streaming API, which accesses tweets published in real time, to improve and stabilise the definition of the search criteria for the entire time period. One month was considered a suitable period, as it allowed capturing the initial explosion of the event and its consequences in the public debate, without being affected by later evolutions. The entire process took place prior to Elon Musk's Twitter acquisition. Figure 2 shows the number of daily tweets obtained and how a large majority of the publications and users were concentrated in the first days of the period under study.

*Figure 2: Number of tweets and unique users per day*
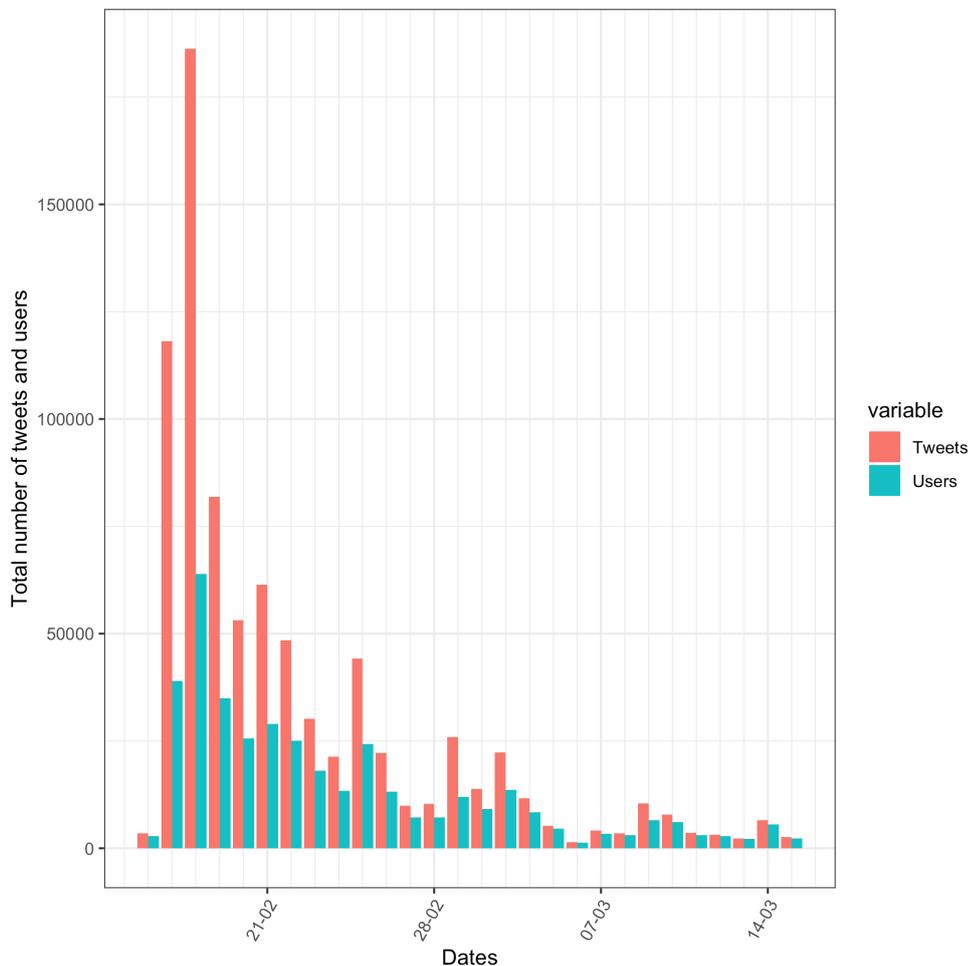
Twitter is a social network where daily conversations involve a wide range of topics. Therefore, it is extremely difficult to restrict the search results to a specific topic. Therefore, to ensure the appropriateness of the data collected, random tests were carried out to check that the extracted tweets and links referred to the case study, especially those that had a higher impact in terms of repercussion. The results showed that, for both tweets and links, the adequacy was above 99%.

From the information stored in the database, the field defined as the URL was extracted for each record, which stores the web link that a tweet may contain. Of the 815,527 tweets, 55,991 contained links to content beyond the text of the tweet itself. Therefore, approximately 7% of the tweets included, or provided access to, information beyond the tweet author's own words. Figure 3 includes links to external sources and internal links to Twitter posts (RT and retweets). However, a large proportion of the remaining tweets is part of the conversation around an initial tweet that provides a link. The number of unique links was 13,357, of which the vast majority were links with minimal impact; for example, 10,735 of them appeared only once. For the analysis, only the 100 links with the highest number of occurrences (top-100) were considered, which generated a range between 2,817 and 56 occurrences with a coverage close to 60% of the total number of links extracted (30,711 total occurrences). To compute the occurrences and avoid splitting them between different versions of the same link, we automatically extended the URLs that used some kind of shortener service and

considered only the part of the URL before the question mark that precedes the UTM tracking parameters.

Figure 3: Number of appearances for each unique URL
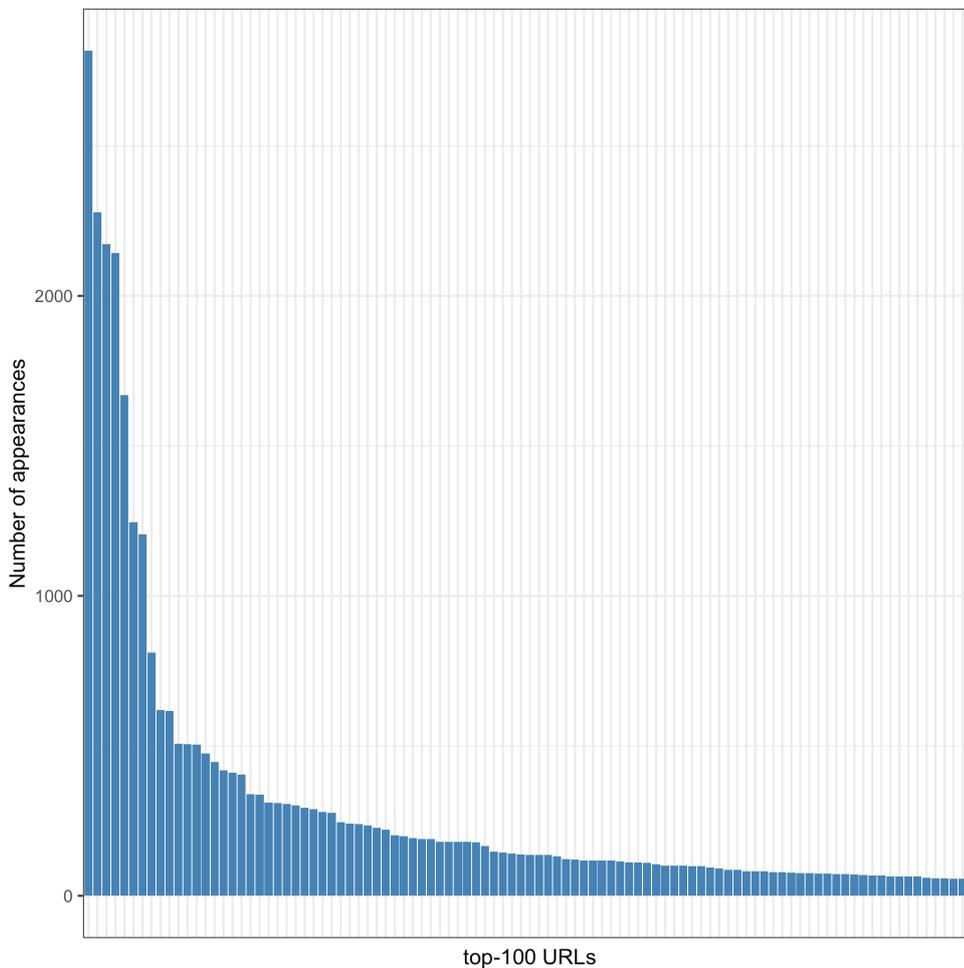
As figure 3 shows, the difference in the number of occurrences between the first top-100 links and the rest is quite remarkable. The top ten links account for approximately 50% of the total number of appearances in the top-100 and approximately 27% of the total number of appearances in the database. Therefore, the impact of the former on public debate is much greater than that of the latter. Nevertheless, we consider that this part of the top-100 with the lowest impact has a sufficient volume of relevance, at the same time as it serves as a sample approximation to the long tail of links that have a lower number of appearances, and in particular, to those closest to this top-100.

The links were classified according to the following criteria:

- Portal: direct link to an open data portal, transparency portal, or institutional website.

- News: informative text on a current public interest event disseminated through a social media outlet.

- Opinion: written expressions from a personal point of view in the media.

- Video: interview, intervention, or news item in audio-visual format (can incorporate on-screen text).

- Image: Screenshot of headline in the media.

- Humour: Humorous content in any format (text, image, meme, etc.).

- Previous: Link to a source before the occurrence of the case study.

Links to tweets are not considered because they are already included in the RT (the links are still counted).

The link analysis is carried out at three levels, differentiating whether the original link is to the Portal category or to another, in the order of relevance:

- Type A: Link directly contained in Portal category tweet.

- Type B: Another category of links in the tweet and the linked content, in turn, links to a transparency portal, open data, or institutional website.

- Type C: Another category of links in the tweet where the linked content mentions the existence of relevant content in a transparency portal, open data, or institutional website but without linking to it.

- Type 0: No links or mentions of Portals in the document linked by the tweet.

For types B and C, an analysis of the citation to the Portal and its context will be conducted. Citation analysis consists of determining the specific topic in which the citation occurs. Context analysis examines the topics covered in the lines and paragraphs immediately preceding and following the quotation. In addition, the position of the quotation in the document will also be analysed (initial part, middle part, or final part) to determine whether the piece bases its claim on the information contained in the Portal.

4. Results

Table 1 provides the results of categories and types of links by frequency of appearance in tweets in the working database and in brackets, by a simple link count. Only 3 percent of the links in the top-100 are type A, specifically links to the Autonomous Community's transparency portal, 5 percent are type B and 17 percent are type C. The vast majority (75 percent) of the contents linked in the tweets did not include any links or references to Portals. Considering the dissemination of the tweets, the impact of type A links decreases to slightly less than 1 percent of the content of the database, whereas in types B and C, it increases to just over 11 and 21 percent respectively, to the detriment of the impact of type 0, which decreases to 67 percent.

Table 1: Classification and types of links by total and single frequencies of occurrence.

	A	B	C	0	Sum
Portal	293 (3)				**293 (3)**
News		3309 (4)	5466 (13)	11196 (51)	**19971 (68)**
Opinion			405 (1)	316 (3)	**721 (4)**
Video				3489 (5)	**3489 (5)**
Image				1143 (8)	**1143 (8)**
Humour				3244 (7)	**3244 (7)**
Previous		111 (1)	606 (3)	1246 (1)	**1963 (5)**
Sum	**293 (3)**	**3420 (5)**	**6477 (17)**	**20634 (75)**	**30824 (100)**

Regarding the types, the majority were news sources, 68% in individual terms and almost the same percentage when considering the dissemination of tweets. The remaining sources all obtained low figures, both in individual terms and in terms of the dissemination of tweets. Among them the percentages of tweet diffusion stand out at around 11 percent for video and humour sources and a weight of 5 percent in individual terms, for links to sources prior to the occurrence of the case. It is also relevant that the most cited link on an individual basis corresponds to a satirical publication. Finally, by definition, Portal sources coincide with link type A, with weights of 3 per cent and 1 per cent, respectively, on individual links and on the set of tweets, respectively.

From the above linked content, 22 media articles containing web links or citations to Portals were extracted. Despite dealing with the case study, 78 articles did not contain any links or references to Portals. Of these 22 articles, 196 paragraphs were found to be relevant to the case study, excluding those that dealt with other topics within the news piece. The paragraphs or sets of paragraphs that were considered relevant were those that were associated with a narrative about the case under study and that were linked in the narrative either to a link to the Portal or to an expressed statement that the information could be found on the Portal, even without including a URL link, in both cases, regardless of whether or not an analysis of the content was provided. Of these, 313 pieces of content relevant to the case were tagged. The following topics emerge from their analysis, listed below, ordered by the frequency of appearance: in a similar way to the study by Lourenço (2016), although their order is considered relevant, we do not believe that the exact number of frequencies of appearance is noteworthy, as this is a qualitative identification at the discretion of the researchers. Figure 4 shows the importance of the topics, according to their frequency of appearance, for both types, B and C news content and also breaking down between the first quartile (Q1) and the rest of the distribution (QX).

Figure 4: Relevance of the topics
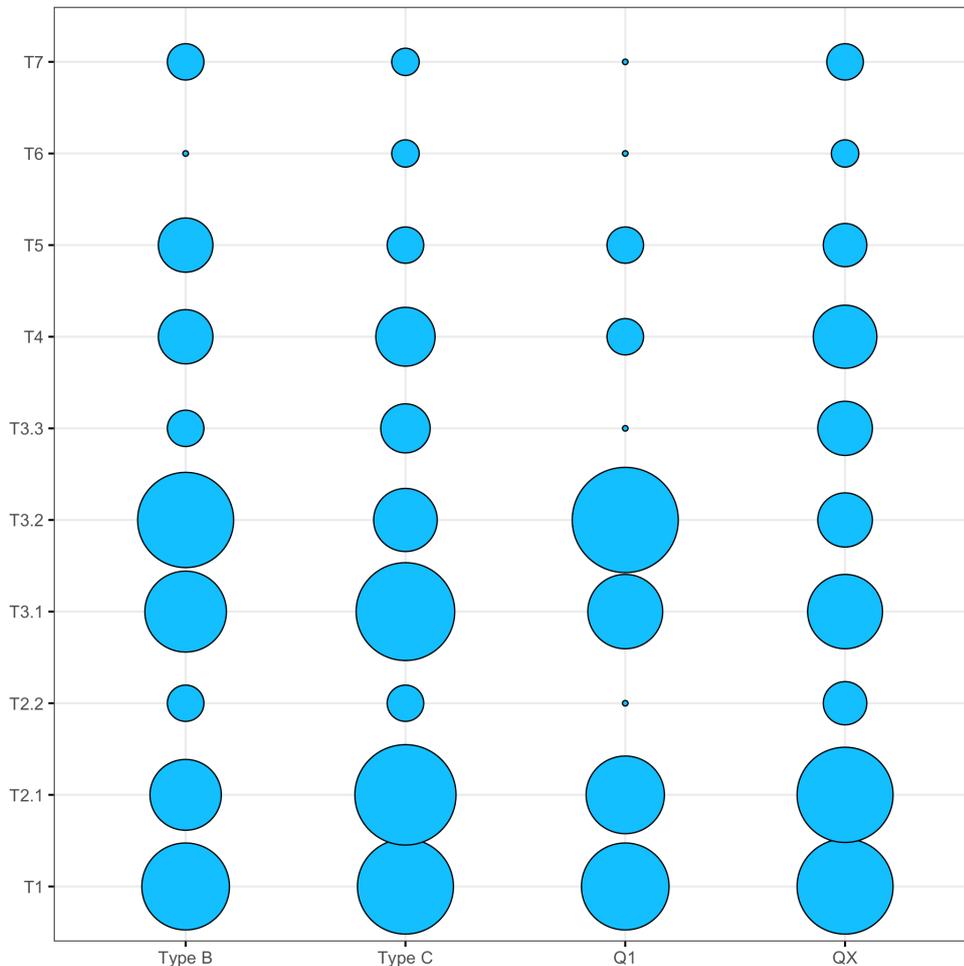
Overall, the topics ordered by cumulative frequency of occurrence were as follows:

- **T1 - Reproduction of data published on the Portal of an economic or technical nature on the transaction or on the persons involved.** This corresponds to factual data extracted from published documents, referring to monetary expressions of expenditure and technical characteristics of the goods purchased or natural or legal persons involved in the transaction.

- **T2 - Information on third parties:**
  - **T2.1 - Specific additional information about third parties involved in the transaction that is not found in the data on the Portal.** In general, this is information of a commercial nature on the activities of companies, entrepreneurs, members of corporate governance bodies, or cases of corporate management and their relations with other companies or entrepreneurs who may also have had commercial relations with Public Administration.
  - **T2.2 - Additional information and data obtained on third parties at other official public access Portals.** Mainly Companies Register but also Public Procurement Portal. This topic is complementary to the previous one, but the news item expressly states its source through official Portals.

- **T3 - Lack of data and information:**

- **T3.1 - Explicit manifestation of lack of data of interest for the journalistic piece.** Mainly, references to the absence of a detailed breakdown of the goods acquired and their characteristics of information on different actions in the contracting phase, such as award and formalisation documents or valuations of the intervention of intermediaries or of potential conflicts of interest with members of the government or public administration.

- **T3.2 - Additional information on the operation provided by the journalist, either by their own means or by statements from third parties, which are not found on the Portal.** This mainly concerns information on the third parties involved in the operation and their relationship with previous operations, entities other than the successful bidder or public officials. This issue would complement the previous one, being actually the response offered in the journalistic piece, to the absence of information on the Portal.

- **T3.3 - Explicit manifestation in the news item of the absence of information on the operation that has been tried to obtain from sources related to Public Administration, but these have refused to respond or have not provided it.** Mainly, relations with third parties or aspects of the award process. Again, this issue complements the two previous ones.

- **T4 - Use of data for purposes beyond simple reporting.** Mainly to show inefficiencies or ineffectiveness in spending, contrast statements by the government or its officials, or compare with similar operations of another or the same administration.

- **T5 - Mentions of the Portal as an object of interest in its own right, usually in pronouncements.** Mainly of a defensive nature by the government or its representatives defending transparency because of the availability of data on the Portal, but also of an offensive nature by the opposition endorsing the use of data as having been obtained from the Portal.

- **T6 - References to the integrity of information contained in the Portal.** Its modification of the original content, its appearance at a later date than press reports on the operation, or its removal after availability.

- **T7 - Emotional narrative around the information.** Stylistic figures with no effect on content, with the aim of constructing a narrative.

Given the low level of type B references (documents that include an explicit link to a Portal), the results do not show a conclusive link use pattern. In general, they have been used as an argument of authority, linking to the original source of the data (T1) and serving to construct argumentations, mainly in the context of the lack of data on third parties involved in the operation, which the journalist considers to be of interest to the story (T3). It is particularly noteworthy that in none of the cases is there a need for additional data on natural or legal persons, solved by means of a link to another type of appropriate official data source, such as the Companies Register.

The results are more diverse when we only consider citations to the Portal without including a link to it, but simply a citation as a source of information. In this instance, we have also analysed both the sentence or set of sentences in which the quotation is found and the preceding and following context, understood as sentences and/or paragraphs, immediately before and after the quotation.

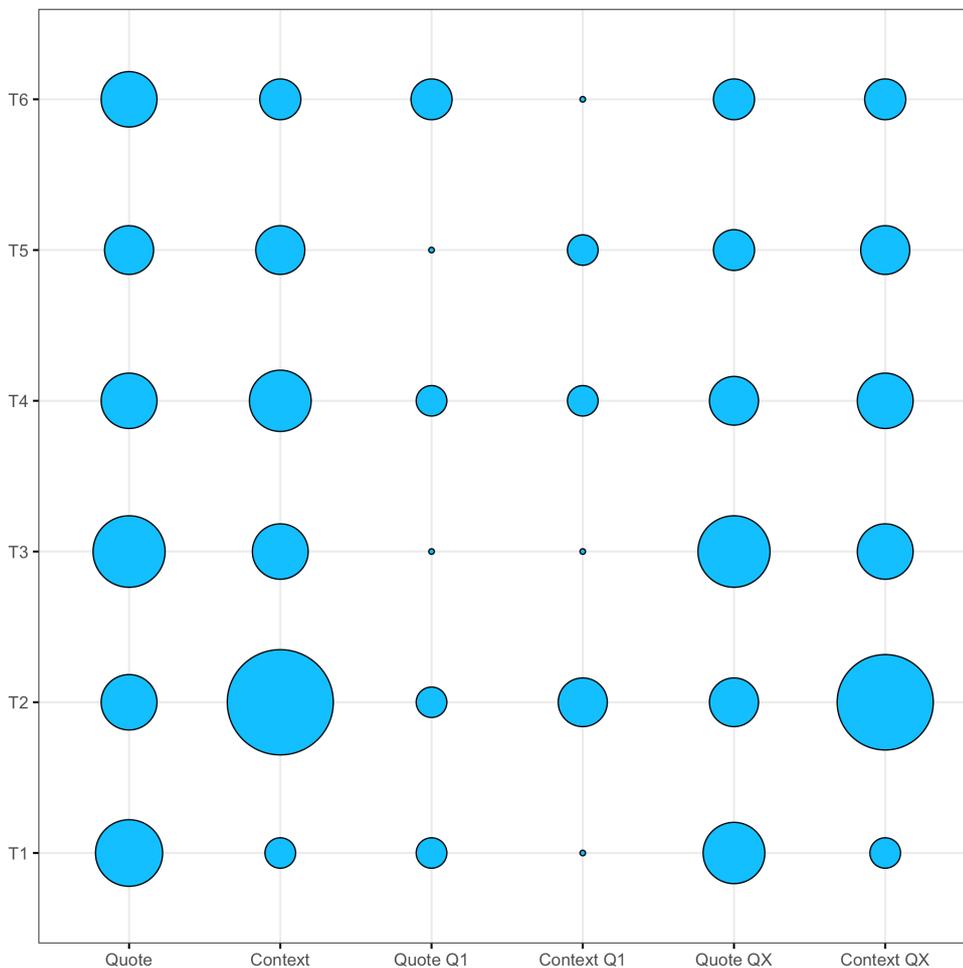*Figure 5: Analysis of Portal quotes and their contexts*

Figure 5 shows the results obtained in citations and context for the overall type C articles for the first quartile (Q1) and for the rest of the quartiles (QX). In this case, we grouped T2 and T3 topics without any breakdown. Again, we do not consider the exact frequencies of occurrence of the topics as relevant but simply their approximate weights. It should be noted that either a quotation or context can be assigned to more than one topic, especially in the context, because it is a larger volume of information.

No emotional reference was found in either the quote or context (T7). Overall, the citations to the Portal are more or less evenly distributed among all the topics (with the exception of T7), with a slightly greater relevance for topics T1 and T3, endorsing the specific data offered by means of the authoritative argument of coming from the Portal (T1) and expressing the insufficiency of the information contained in the Portal to elaborate the news (T3), respectively. This uniformity is not so clear when only the news items in the first quartile are taken, with significantly fewer citations referring to the issues of insufficient information (T3) and defensive use of the Portal (T5). For the other quartiles as a whole, the results were similar to those for the total population.

Regarding the contexts, the results show a greater weight of the topic of relations with third parties (T2) at all levels, although higher in quartiles 2 to 4 (QX), to the detriment of the topics of providing specific data on the Portal (T1) and the manifestation of insufficient information (T3), which were found more frequently in the quotations. This lower importance is particularly significant for the first quartile (Q1).

The results show that the highest number of quotes to the Portal is set in a context in which third parties involved in the transaction are being disclosed. We also found that the topic of showing specific Portal data (T1) appears substantially less in the context. This is because, in the development of the news piece, factual data are framed in the context of relations with third parties illustrating what the object of the acquisition or service was and it is common here that the citation to the Portal is offered as an argument of authority and objectivity.

Finally, regarding the place where quotations are found, the middle parts of the documents are predominant and slightly more than 60% of them are to be found. The initial and final parts of the documents amounted to almost 20% of the total. It is noteworthy that when only Q1 items were considered, almost all citations were found in the middle part.

5. Discussion and conclusions

This paper has carried out a case study on the use of information contained in open data portals, transparency portals and other institutional websites in the media's treatment of information. For this purpose, a case of particular relevance in Spain was selected, associated with an alleged irregularity in public procurement that had a high media impact, as it triggered a crisis in one of the main political parties in Spain in the first quarter of 2022.

Conversations on the social network Twitter were used as a measure of the impact of different news items on public opinion. Using access to the Twitter API, we extracted all tweets that matched a fairly restrictive selection criterion to ensure that the results were not affected by other conversations on the social network, particularly those related to the leadership crisis that originated in the political party. In this way, a database of 815,527 tweets was created, from which 13,357 unique URLs were extracted, once they had been extended, if they came from a shortener and the UTM information had been removed. The number of occurrences of each unique URL was considered a measure of its relevance to public debate. In view of the frequency distribution, it was decided to study only the top-100 URLs, with special attention paid to those contained in the first quartile of the distribution as those with the highest impact.

The URLs were classified according to their nature and whether they were a direct link to a Portal, included a link to a Portal in the document to which they were linked, or whether the linked document merely contained a mention of information available on the Portal. The results showed the low relevance of direct links to a Portal, the high weight of news in the press media, and that most of the linked documents did not even contain a mention of the information available on a Portal.

Among those documents that provided a link to the Portal or mentioned that the data were available on the Portal, a selection was made of the relevant paragraphs for the case study. A qualitative

analysis was carried out to identify the main topics that were addressed when the link or mention of the Portal appeared and within the textual context surrounding this link or quotation. The results revealed that the most frequent topics were quotations or links to the Portal as an authoritative argument to endorse the information shown, the statement that the Portal did not contain the necessary data on the subject of the contract for journalistic information, and the absence of information on third parties involved in procurement with the Public Administration. This last issue was the most repeated, both in terms of citations and context.

The case analysis reveals that the information contained in the data Portals has been used mainly by the media, while the overall number of users has been low. This raises the question of whether the lack of expectations regarding its use, as suggested by Heald (2021), is met. In addition, there is evidence of potential demand for other data, especially those held in Companies Register, which present barriers to access, such as paywalls. This result is in line with Worthy (2015), who highlighted the importance of contextualisation in accessing other resources and contributes to closing the gap mentioned by Gray (2015) on the necessary analysis of the demand for data.

Our results also support the role of media in the dissemination of Portal data, both in terms of how it is used and the informational context provided. Crepaz and Kneafsey (2021) argue that media plays a key role in data dissemination and our findings confirm this claim. At the same time, our results also agree with Diez-Garrido and Campos-Domínguez (2018) that the media's use of data is limited. We speculate that this is because of the lack of a comprehensive view of all relevant individuals or legal entities in the disclosed data. This limitation leads us to the need to open up business registers and employ higher-level data models, such as Linked Open Data, to achieve a more complete perspective.

Although it cannot be stated categorically, as the data offered have proven useful, this lack of real context means that we could also find ourselves faced with the phenomenon described as *openwashing* and expressed by various authors, particularly in Spain, but probably applicable to other areas, that Portals may have more value in themselves as a technological artefact than for the actual content they offer.

Furthermore, our analysis found no clear evidence of disinformation manoeuvres in the links disseminated through the social network. Almost all the linked references corresponded to established and reputable news outlets. Within the news content, even those paragraphs categorised as emotional do not seem to have a clear campaign objective, at least not systematically. However, our aim was to obtain the links disseminated through the social network but not to examine in detail the conversation as a whole or the social relations around the posts with links and other mentions of the case that did not contain links. This is where misinformation could be found, especially *data-driven bullshit*, which would distort the social conversation and divert attention from the objective data. In addition, we highlight that the most cited link comes from a satirical medium, which raises the possibility that the humorous nature has contributed to the dissemination of information and formation of public opinion, as established in previous studies (Mina, 2019; Madrid-Morales et al., 2021).

We are aware of the limitations of a single case study to draw generalisable conclusions about the use of Portals, especially when an important part of the study is based on qualitative analysis. Nevertheless, we believe that our results illustrate the use that is being made of the data contained in the Portals in the public discussion and, particularly, they highlight the main informational deficits. It is clear that there will always be some data that a media outlet requires for information that is not present on the Portal, which is why passive transparency policies must always be complemented with active transparency methods that allow effective access to public information. However, much of these data are already held by Public Administrations, specifically in registers of declarations of interests, pressure groups, public agendas, and especially in the Companies Register.

An adequate opening of these data, particularly from the Business Register under the still-pending transposition of the European Directive (EU) 2019/1151, would allow the creation of an ecosystem of Linked Open Data (LOD) through semantic web technology under the Resource Description Framework (RDF) defined by the World Wide Web Consortium (W3C) as the standard for describing web resources, specifying metadata and representing information. This facilitates the free linking of data from different sources and institutions for their joint exploration and the creation of new combinations of data that enable their visualisation and show their interrelation for the creation of information. It is worth noting that the independent technical review carried out by the Open Government Partnership on the IV National Plan for Open Government in Spain (2020-2024)[1] evaluates the opening of the Business Register as a recurrent demand of anti-corruption and integrity organisations and qualifies its data as key information to identify private interests that may be directly related to decision-making, the management of public money, or in public procurement processes.

Evidently, our recommendation is not a magic solution. The results showed that 75% of the links did not contain any mention of the existence of data on the Portal. We do not know why this happens: ignorance of the existence of the resource, lack of interest in the mention, desire to provide as simple information as possible, or simply the appearance that the media outlet has obtained all data through its own informative activities. In any case, it is necessary for governments and public authorities to promote the existence of the magnificent resources that these Portals represent for accountability to citizens, albeit through the media channel, and it is necessary to make these intermediaries aware that offering their readers links to the original data is a plus in terms of their rigorous reporting.

Any information pyramid masks a power structure, Portals are a way to permeate the different layers of governments and administrations by making data flow from the top to the bottom where citizens are. Media and data activist organisations are the right intermediaries to transmit information to most of the population. The existence of an appropriate structure of open data linked to various Portals allows these intermediaries to fulfil their roles while ensuring the accuracy of the information. Undoubtedly, all these actions are costly and sometimes very difficult to measure in relation to their monetary benefits beyond their ethical components. But there is no doubt that the costs of transparency are always lower than the costs of opacity.

---

[1]   Available on: https://www.opengovpartnership.org/documents/spain-action-plan-review-2020-2024/

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Andreassen, C. S. (2015). Online Social Network Site Addiction: A Comprehensive Review. *Current Addiction Reports*, *2*(2), 175–184. https://doi.org/10.1007/s40429-015-0056-9

Andrejevic, M. (2020). The political function of fake news: Disorganized propaganda in the era of automated media. In M. Zimdars & K. McLeod (Eds.), *Fake news: Understanding media and misinformation in the digital age* (pp. 19–28). The MIT Press. https://scholar.google.com/scholar_lookup?title=The%20political%20function%20of%20fake%20news&author=M.%20Andrejevic&publication_year=2019

Aricat, R. (2018). Effacing the Dilemma of the Rumouring Subject: A Value-oriented Approach towards Studying Misinformation on Social Media. *Journal of Human Values*, *24*(1), 56–65. https://doi.org/10.1177/0971685817733576

Arpino, B., & Obydenkova, A. V. (2020). Democracy and Political Trust Before and After the Great Recession 2008: The European Union and the United Nations. *Social Indicators Research*, *148*(2), 395-415. https://doi.org/10.1007/s11205-019-02204-x

Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, *37*(1), 38-54. https://doi.org/10.1177/0894439317734157

Beltrán-Orenes, P., & Rodríguez-Mateos, D. (2020). Presence and location of Spanish municipal transparency Portals. *Profesional de la Información*, *29*(6). https://doi.org/10.3145/epi.2020.nov.35

Berghel, H. (2018). Malice Domestic: The Cambridge Analytica Dystopia. *Computer*, *51*(5), 84–89. https://doi.org/10.1109/MC.2018.2381135

Bergstrom, C. T., & West, J. D. (2020). *Calling bullshit: The art of skepticism in a data-driven world* (First edition). Random House.

Bértoa, F. C., & Rama, J. (2020). Party decline or social transformation? Economic, institutional and sociological change and the rise of anti-political-establishment parties in Western Europe. *European Political Science Review*, *12*(4), 503-523. https://doi.org/10.1017/S1755773920000260

Boulianne, S. (2019). Revolution in the making? Social media effects across the globe. *Information, Communication & Society*, *22*(1), 39–54. https://doi.org/10.1080/1369118X.2017.1353641

Cagé, J., Hervé, N., & Mazoyer, B. (2022). *Social Media Influence Mainstream Media: Evidence from Two Billion Tweets* (No. DP17358-2). CEPR Centre for Economic Policy Research. https://portal.cepr.org/discussion-paper/20328

Crepaz, M., & Kneafsey, L. (2021). Usability of transparency portals: Examination of perceptions of journalists as information seekers. *Public Administration*, padm.12777. https://doi.org/10.1111/padm.12777

Curto-Rodríguez, R. (2020). Operational transparency of the Spanish autonomous communities through their open data portals. *El Profesional de la Información*, *29*(1), Article 1. https://doi.org/10.3145/epi.2020.ene.15

Díez Garrido, M., & Campos Domínguez, E. (2018). Journalist´ view of transparency in Spain. *Spanish Journal of Transparency*, 7, 49–69. https://doi.org/10.51915/ret.9

European Commission. (2020). *Technology and democracy: Understanding the influence of online technologies on political behaviour and decision making.* (No. JRC122023). Joint Research Centre. https://data.europa.eu/doi/10.2760/709177

European Commission. Directorate General for Communications Networks, Content and Technology. (2018). *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation.* Publications Office. https://data.europa.eu/doi/10.2759/739290

Fletcher, R., & Nielsen, R. K. (2018). Automated Serendipity. *Digital Journalism*, 6(8), 976–989. https://doi.org/10.1080/21670811.2018.1502045

Fuchs, C. (2010). Alternative Media as Critical Media. *European Journal of Social Theory*, 13(2), 173–192. https://doi.org/10.1177/1368431010362294

Gabielkov, M., Ramachandran, A., Chaintreau, A., & Legout, A. (2016, June 14). *Social Clicks: What and Who Gets Read on Twitter?* ACM SIGMETRICS / IFIP Performance 2016. https://hal.inria.fr/hal-01281190

García-García, J., & Alonso-Magdaleno, M. I. (2022). Deconstructing Government Budgets through Visual Representation Software. *International Journal of Public Administration*, 45(11), 807-818. https://doi.org/10.1080/01900692.2021.1891549

García-García, J., & Curto-Rodríguez, R. (2018). Disclosure of public information of Spanish regional governments (2013-2017): Open data portals, transparency portals and institutional websites. *El Profesional de La Información*, 27(5), 1051. https://doi.org/10.3145/epi.2018.sep.09

Gaufman, E. (2018). The Trump carnival: Popular appeal in the age of misinformation. *International Relations*, 32(4), 410–429. https://doi.org/10.1177/0047117818773130

Gray, J. (2015). *Open Budget Data: Mapping the landscape*. Open Knowledge. http://www.fiscaltransparency.net/eng/resource_open_public.php?IdToOpen=20150902128

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374–378. https://doi.org/10.1126/science.aau2706

Gutiérrez, M. (2018). *Data activism and social change*. Springer Berlin Heidelberg.

Hao, K. (2021, November 20). *How Facebook and Google fund global misinformation*. MIT Technology Review. https://www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/

Heald, D. (2012). Why is transparency about public expenditure so elusive? *International Review of Administrative Sciences*, 78(1), 30–49. https://doi.org/10/gcx5x2

Heimstädt, M. (2017). Openwashing: A decoupling perspective on organizational transparency. *Technological Forecasting and Social Change*, 125, 77–86. https://doi.org/10.1016/j.techfore.2017.03.037

Janssen, M., Matheus, R., Longo, J., & Weerakkody, V. (2017). Transparency-by-design as a foundation for open government. *Transforming Government: People, Process and Policy*, 11(1), 2-8. https://doi.org/10.1108/TG-02-2017-0015

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books.

Kuss, D. J., & Griffiths, M. D. (2011). Online Social Networking and Addiction—A Review of the Psychological Literature. *International Journal of Environmental Research and Public Health*, *8*(9), Article 9. https://doi.org/10.3390/ijerph8093528

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Smillie, L., Garcia, D., Hertwig, R., Weatherall, J., Egidy, S., Robertson, R. E., O'Connor, C., Kozyreva, A., & Lorenz-Spreen, P. (2020). *Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making*. Publications Office of the European Union.

Lourenço, R. P. (2016). Evidence of an Open Government Data Portal Impact on the Public Sphere: *International Journal of Electronic Government Research*, *12*(3), 21–36. https://doi.org/10.4018/IJEGR.2016070102

Madrid-Morales, D., Wasserman, H., Gondwe, G., Ndlovu, K., Sikanku, E., Tully, M., Umejei, E., & Uzuegbunam, C. (2021). Comparative Approaches to Mis/Disinformation| Motivations for Sharing Misinformation: A Comparative Study in Six Sub-Saharan African Countries. *International Journal of Communication*, *15*(0), Article 0.

Majó-Vázquez, S., Congosto, M., Nicholls, T., & Nielsen, R. K. (2021). The Role of Suspended Accounts in Political Discussion on Social Media: Analysis of the 2017 French, UK and German Elections. *Social Media + Society*, *7*(3), 1–20. https://doi.org/10.1177/20563051211027202

Majó-Vázquez, S., Zhao, J., & Nielsen, R. K. (2017). *The digital-born and legacy media news media on Twitter during the French presidential elections*. Oxford, UK.

Martínez-Rolán, X., & Piñeiro-Otero, T. (2016). The use of memes in the discourse of political parties at Twitter: Analysis of the 2015 state of the nation debate. *Communication & Society*, 145-159. https://doi.org/10.15581/003.29.1.145-159

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television & New Media*, *22*(2), 205–224. https://doi.org/10.1177/1527476420982230

Matheus, R., & Janssen, M. (2020). A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory. *Public Performance & Management Review*, *43*(3), 503-534. https://doi.org/10.1080/15309576.2019.1691025

Meijer, A. (2013). Understanding the Complex Dynamics of Transparency. *Public Administration Review*, *73*(3), 429–439. https://doi.org/10/f4xbkv

Mina, A. X. (2019). *Memes to movements: How the world's most viral media is changing social protest and power.* Beacon Press.

O'Donnell, G. A. (1994). Delegative Democracy. *Journal of Democracy*, *5*(1), 55–69. https://doi.org/10.1353/jod.1994.0010

O'Donnell, G. A. (1998). Horizontal Accountability in New Democracies. *Journal of Democracy*, *9*(3), 112–126. https://doi.org/10.1353/jod.1998.0051

Peña-López, I., Congosto, M., & Aragón, P. (2014). Spanish Indignados and the evolution of the 15M movement on Twitter: Towards networked para-institutions. *Journal of Spanish Cultural Studies*, *15*(1–2), 189–216. https://doi.org/10.1080/14636204.2014.931678

Rico Motos, C. (2019). 'Let the Citizens Fix This Mess!' Podemos' Claim for Participatory Democracy in Spain. *Politics and Governance*, *7*(2), 187-197. https://doi.org/10.17645/pag.v7i2.1893

Rudschies, C. (2022). Power in the modern 'surveillance society': From theory to methodology. *Information Polity*, *27*(2), 275-289. https://doi.org/10.3233/IP-211536

Smith, S. T., Kao, E. K., Mackin, E. D., Shah, D. C., Simek, O., & Rubin, D. B. (2021). Automatic detection of influential actors in disinformation networks. *Proceedings of the National Academy of Sciences*, *118*(4). https://doi.org/10.1073/pnas.2011216118

Stella, M., Ferrara, E., & Domenico, M. D. (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, *115*(49), 12435–12440. https://doi.org/10.1073/pnas.1803470115

Valera-Ordaz, L., Requena-i-Mora, M., Calvo, D., & López-García, G. (2022). Unraveling disinformation: Notions and discourses from the Spanish population. *Comunicar*, *30*(72), 21–32. https://doi.org/10.3916/C72-2022-02

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, *20*(5), 2028–2049. https://doi.org/10.1177/1461444817712086

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making* (DGI(2017)09). Council of Europe. http://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

Wenzel, A. (2019). To Verify or to Disengage: Coping with "Fake News" and Ambiguity. *International Journal of Communication*, *13*, 1977-1995.

Winsvold, M. (2007). Municipal Websites in the Local Public Debate. *Nordicom Review*, *28*(2), 7-23. https://doi.org/10.1515/nor-2017-0207

Worthy, B. (2015). The Impact of Open Data in the UK: Complex, Unpredictable, and Political. *Public Administration*, *93*(3), 788–805. https://doi.org/10/f7rr53

Yang, T., Majó-Vázquez, S., Nielsen, R. K., & González-Bailón, S. (2020). Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences*, *117*(46), 28678–28683. https://doi.org/10.1073/pnas.2006089117

Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth edition). SAGE.

Young, J. C. (2021). Disinformation as the weaponization of cruel optimism: A critical intervention in misinformation studies. *Emotion, Space and Society*, *38*, 100757. https://doi.org/10.1016/j.emospa.2020.100757

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

About the Authors

Jesús García-García

Jesús García-García works as tenured Associate Professor in the Department of Accounting of the University of Oviedo. His research is focused on examining the interaction between citizens, digital technology, and public administrations, with a particular focus on the role of open data and transparency portals. The aim of this research is to gain a deeper understanding of how the internet has facilitated communication and closer collaboration between civil society and governments.

María Isabel Alonso-Magdaleno

María Isabel Alonso-Magdaleno works as tenured Associate Professor in the Department of Business Administration of the University of Oviedo. Her research is centered on the analysis of transparency processes, open data initiatives, and disclosure practices of public administrations, at both local and national levels. Additionally, she has conducted research on the role of open-source and open-knowledge models as socially responsible activities.