



Rodrigues, C., (2023) "Evidence and transparency in Open Data Journalism: a case study on British and Brazilian news agencies", *communication +1* 10(1).  
doi: <https://doi.org/10.7275/cpo.1880>



## Evidence and transparency in Open Data Journalism: a case study on British and Brazilian news agencies

**Claudia Rodrigues**, [claudiar63@gmail.com](mailto:claudiar63@gmail.com)

The authority of professionalism has been undermined in an environment of growing and proactive disinformation, driven by algorithms that disrupt the trust in established media. We cannot ignore the effect of traditional forms of media bias, but the spread of news on social media makes reliability even more nebulous. The open data journalism (ODJ) model brings together notions of proof of evidence and precision and is identified as a manner to rebuild trust in the media by authors who research this field such as Hermida and Young, Parasie e Dagiral, Borges-Rey and Fink and Anderson. Access to backstage elements of reporting process through links to spreadsheets, google docs and digital documents provides a cornerstone of ODJ. This dimension was embraced by projects such as The Panama Papers and it is a common ground between news agencies established in the midst of a new centrality in the journalistic field. It is relevant to point out that credibility in journalism is linked to aspects related to practices, expertise, and norms that legitimize professional knowledge. The evidence embedded in open data journalism routines conveys levels of transparency that work as a strategic ritual. An experimental and original methodology to draw inferences on the extent to which three outlets deliver transparency served as a parameter for assessing efforts to build reliable narratives. The resulting content analysis method served to measure evidence and accuracy and highlight levels of transparency. To carry out the study of transparency, 135 reportings from The British The Bureau of Investigative Journalism and the Brazilians Gênero e Número and Agência Pública were analyzed and sixteen interviews were conducted. The result revealed high transparency for Agência Pública, medium transparency for Gênero e Número and low transparency for TBIJ. In the interviews, journalists raised concerns about the importance of accuracy and consistency in data collection and the functionality of hyperlinks. ODJ is perceived as a way of reinforcing documental value in newsgathering.

**Keywords:** open data journalism, verification, fact-checking, news agencies, investigative journalism

*communication +1* is a peer-reviewed open-access journal. This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC-BY-SA 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited and any remixes, transformations, or adaptations are distributed with the same license.

## Introduction

The abundance of information on digital platforms has disrupted the centrality of the established mainstream media as a pillar in the mediation process. The emergence of the internet has prompted the expansion of digital journalism and propagated changes in the paradigm of the media business model. This scenario has led to the introduction of terms such as “crisis” and precariousness within the news industry where a drastic reduction of professionals has impacted newsrooms in recent decades.<sup>1</sup> Worldwide, the reconfiguration brought about by post-industrial journalism has affected the legitimacy and especially the boundaries of the media.<sup>2</sup> In a social context entrenched in the modern era, expertise and authority are contested.

The dispersal of the audience over a variety of different social media networks – such as Facebook, Twitter, Instagram, WhatsApp, and Telegram, that have become new sources – has increased the decline of the media’s social role. Discrediting of the media has been remarkably aggravated by the actions of populist figures. Former Brazilian president Jair Bolsonaro is an emblematic case. He maintained news channels on Facebook that disqualified the press.<sup>3</sup>

Amid this scenario, distinct production forms have been intensified and spread by new actors that have gained access to the media realm.<sup>4</sup> Not-for-profit organizations with investigative DNA seek to consolidate themselves as quality media. Among these outlets, some practice open data journalism (ODJ). This journalistic model brings together notions of objectivity, evidence, and precision and is seen to rebuild trust in the media.<sup>5</sup> Hermida and Young propose that the epistemology and expertise of data journalism establish a sense of authority and professional identity.<sup>6</sup>

Open data journalism comes under the “umbrella” of data journalism. However, output tagged as DJ or data-driven journalism (DDJ) doesn’t necessarily constitutes OD. Proposed as a specific nomenclature,<sup>7</sup> ODJ combines information

---

<sup>1</sup> Mark Deuze, and Tâmara Witschge, “O que o jornalismo está se tornando,” *Parágrafo* 4, no. 2, (2016): 8–11, <https://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/478>.

<sup>2</sup> Christopher Anderson, Emily Bell, and Clay Shirky, *Post-Industrial Journalism: Adapting to the Present* (Nova York: Tow Center for Digital Journalism, 2012), <https://doi.org/10.7916/D8NoiS7>.

<sup>3</sup> Leonardo Cazes, Mark Deuze, João Paulo Saconi, and Juliana Dal Piva. “Facebook derruba rede de páginas coordenada por funcionários da Presidência e dos gabinetes de Flávio e Eduardo Bolsonaro. <https://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/478>.

<sup>4</sup> Graham Meikle, *Future Active: media and activism in the internet* (New York: Routledge, 2003); Dan Gillmor, *We the Media: Grassroots Journalism by the People, for the people* (Sebastopol: O’Reilly, 2004).

<sup>5</sup> Alfred Hermida, and Mary-Linn Young, *Data Journalism and the Regeneration of the News* (Routledge, 2019).

<sup>6</sup> Hermida, and Young, *Data Journalism*, 9.

<sup>7</sup> Data journalist doesn’t imply access to the backstage such as ODJ. Many Data Journalism

and *reader-friendly fact-checking*. Verification is paramount to quality journalism. Access to the raw material of the report through hypertextuality is the main feature of open data journalism. The dimension and volume fostered by data visualization are endorsed by the access to spreadsheets and Google Docs.<sup>8</sup>

The growing datafication of society and the open data movement that have fostered the creation of transparency portals in several countries are intertwined with the data culture that influences the practice of Open data journalism (ODJ).<sup>9</sup> As a watershed moment, coverage such as *Wikileaks*<sup>10</sup> and the *Panama Papers*<sup>11</sup> – that offer access to a massive volume of data – helped to consolidate the latter specialty and its collaborative paradigm.

An experimental content analysis methodology, anchored in an original index, has been developed with the aim of demonstrating the extent to which three outlets allow readers to verify the evidence of the content they publish on their websites. The not-for-profit news agencies the *Bureau of Investigative Journalism* (TBIJ) in the UK, and *Género e Número* (GN) and *Agência Pública* in Brazil in an exemplar and expository way adopted ODJ and, therefore, were chosen as corpus.<sup>12</sup> The disclosure paradigm is in their DNA. Both TBIJ and *Agência Pública* – founded respectively in 2010 and 2011 – where platforms where *Wikileaks* reporting were broadcast. In its mission, *Género e Número* aims to boost the debate over human rights, gender, and race through data.<sup>13</sup> GN has a section called Open Data. *Agência Pública* and TBIJ offer ODJ output spread in several sections. Natália Vianna, editor of *Agência Pública* reinforces how transparency is a core in the outlet.

We have been partners with Wikileaks since the beginning. And the principle of Wikileaks, the principle of leakage, that was very strong in 2011. What you need is to achieve reliability realizing that the reader is also an actor, he has every right to question you and every right to investigate what you are publishing. This makes your process and the bases that led you to that conclusion more

---

output focused on visualization without links to the newsgathering process. Therefore, the author proposes Open Data Journalism as one specialty that brings a distinct epistemology compared to Data Driven Journalism or Data Journalism.

<sup>8</sup> Helen Kennedy, Rosemary Lucy, Giorgia Aiello, and William Allen. “The work that visualisation conventions do.” *Information, Communication and Society* 19, no. 6, (2016): 723, <https://doi.org/10.1080/1369118X.2016.1153126>.

<sup>9</sup> Viktor Mayer-Schonberger, and Kenneth Cukier, *Big Data* (London: John Murray 2013).

<sup>10</sup> The leaking of secret data on the wars in Iraq and Afghanistan, as well as American diplomatic documents, exposed violations of human rights.

<sup>11</sup> 11.5 million leaked documents that revealed illicit operations in tax havens.

<sup>12</sup> The Bureau of Investigative Journalism is available at <https://www.thebureauinvestigates.com/> *Género e Número* is available at <https://www.generonumero.media/> and *Agência Pública* is available at <https://apublica.org/>.

<sup>13</sup> In 2021, GN created the *Hub Género e Número* which offers analysis and through reports supported by data-driven approach. <https://www.generonumero.media/hub/en/>.

transparent. With this, you achieve two things. First, the reader, because, if he wants, he can contradict, question and attack. But also, you allow other journalists, researchers, politicians to have an even greater vision about what you published.<sup>14</sup>

A mixed method approach combines content analysis and sixteen in-depth interviews with data journalists such as Natália Vianna. The study shows that ODJ facilitates reader-friendly fact-checking. Nonetheless, a constant issue of concern expressed by the interviewees is the relevance of the rigor of the data gathering and how inoperative hyperlinks act as disruptive elements in the path toward evidence exhibit and transparency.

### **Genealogy and new practices**

The culture of hyperlinks enables navigation of old archives, documents, and official transcriptions, “allowing the reader to trace back the newsgathering process.”<sup>15</sup> De Maeyer highlights that these attributes bolster credibility and transparency, due to the reader being directed to web pages that contain extra information about the sources.<sup>16</sup> The author suggests that “newsgathering practices such as finding and selecting valuable sources are no longer completely hidden.” *The Guardian’s* environmental editor, Jonathan Watts, has observed that the use of hyperlinks operates as a “barometer,” pointing toward the level of adherence to digital journalism, and notes how digital native organizations like TBIJ are associated with the hyperlink model.<sup>17</sup> The British newspaper works in partnership with TBIJ and embraces the same pattern.

Hyperlinks are, in this sense, a factor that adds a documental value in the digital world and their concept embraces principles such as databases and collective memory.<sup>18</sup> Manovich observes that digital journalism can be seen as a database system, with the reader being the one who operates it through using hyperlinks.<sup>19</sup>

Databases in the newsgathering process gained in popularity with the rise of computer-assisted reporting (CAR), introduced by Philip Meyer.<sup>20</sup> Meyer took

---

<sup>14</sup> Natália Vianna in interview with the author on March 4, 2021.

<sup>15</sup> Mark Deuze, “Journalism and the web: An analysis of Skills and Standards in an online Environment,” *Gazette* 61, no. 5, (1999): 383, <https://journals.sagepub.com/doi/abs/10.1177/0016549299061005002>.

<sup>16</sup> Juliette De Maeyer, “The journalistic Hyperlink. Prescriptive discourse about linking in online news,” *Journalism Practice* 6, no. 5–6 (2012): 163, <https://www.tandfonline.com/doi/abs/10.1080/17512786.2012.667273>.

<sup>17</sup> Interview with the author on July 12, 2021 via google meet.

<sup>18</sup> Lev Manovich, *The Language of New Media* (Cambridge: MIT, 2001), 181; Leonel Azevedo de Aguiar, “A validade dos critérios de noticiabilidade no jornalismo digital,” in *Jornalismo online: modos de fazer*, ed. Carla Rodrigues (Rio de Janeiro: PUC Rio, 2009), 170.

<sup>19</sup> Manovich, *The Language of New Media*, 181.

<sup>20</sup> Philip Meyer, *Precision Journalism: A Reporter’s Introduction to Social Sciences Methods* (Maryland:

advantage of the potential for gathering data and analyzing it through a new method he introduced on computer in 1967 when he investigated conflicts that took place in Detroit, United States. The aim was to identify the motivations behind the protests and what was the profile of the protesters. The report, published on the Detroit Free Press, revealed that African Americans from the south of the USA were not the majority, as was being claimed in editorials. Also, Meyer found out that shortage of jobs, precarious living conditions, and police brutality were culminating in outrage and indignation.

If Meyer could already foresee the potential of databases in terms of the systemization of content, this process has now been made much easier with the support of tools such as Excel. The recognition of patterns and documental evidence, alongside the use of interviews as a complement, constitutes the model created by Meyer.<sup>21</sup> Schudson notes the influence of scientific objectivity in the emergence of the reporter era, in the late nineteenth century, long before the invention of the computer and when journalism was influenced by positivism.<sup>22</sup> Anderson underlines that Meyer claims epistemological rigor through precision journalism, the adoption of “transparent process drawn from the social science canon.”<sup>23</sup> As the latter author points out the four editions of the book *Precision Journalism*, by Meyer, apply a variety of social science techniques to the journalistic work. Data-oriented contextual journalism was raising in the 1970’s and 1980’s and CAR was associated to elite journalistic practices.<sup>24</sup>

The evolution toward an increasingly digital environment fostered a friendly scenario for data collection and prepared the ground for data journalism practices.<sup>25</sup> Wikileaks would not have been viable outside of this framework. About 391,000 documents reporting incidents in the Iraq War were identified between 2004 and 2009. This coverage was a collaboration between *The Guardian*, *Der Spiegel*, *The New York Times*, *El País*, *Le Monde* and Wikileaks and began with

---

Roman & Littlefield, 1973).

<sup>21</sup> Christopher W. Anderson, *Apostles of Certainty: Data Journalism and politics of doubt*. (Oxford/Uk: Oxford University Press, 2018).

<sup>22</sup> Michael Schudson, *Descobrimos as notícias: uma história social dos jornais nos Estados Unidos* (Petrópolis: Vozes, 2010).

<sup>23</sup> Anderson, *Apostles of Certainty*, 120.

<sup>24</sup> Anderson, *Apostles of Certainty*, 125–126.

<sup>25</sup> Alexander Howard, *The art and Science of Data-Driven Journalism* (New York: Tow Center or Digital Journalism, 2014); Jonathan Gray, Liliana Bounegru, and Lucy Chambers, *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News* (Sebastopol, C: O’Reilly, 2012); Sylvain Parasié and Eric Dagiral, “Data-driven Journalism and the Public good. Computer-Assisted Reporters and Programmers-journalist in Chicago,” *New Media and Society* 15, no. 6, (September, 2013): 853–871, <https://journals.sagepub.com/doi/10.1177/1461444812463345>; Terry Flew, Christina Spurgeon, Anna Daniel, and Adam Swift, “The promise of computational Journalism,” *Journalism Practice* 6, no. 2, (October 2011): 157–171, <https://www.tandfonline.com/doi/abs/10.1080/17512786.2011.616655>.

one question: “Are you good with spreadsheets?”<sup>26</sup> Twenty years ago, Bardoel and Deuze predicted an explosive growth in information on a worldwide scale.<sup>27</sup> Events like Wikileaks, the Panama Papers and the Pandora Papers are examples of the progression in the use of data.

Data journalism requires a mindset centered in a process that consists in a fusion of journalism and specific methods. But in its essence, DJ is deeply rooted in the routines of investigation which focus on precision, accuracy, and evidence.<sup>28</sup> This ethos is intrinsically intertwined with the ideal of sharing data which is the core of the open-source movement and the hacker culture. Drawing from Karlsson and Plaisance,<sup>29</sup> Coddington suggests that the increasing discredit of journalism underscores the significant journalistic value of transparency, creating a backdrop that favors the strengthening of the media’s credibility.<sup>30</sup> The surge of ODJ comes side by side with practices related to open data culture and transparency.<sup>31</sup> The adoption of this sort of performative ritual is not common sense between the mainstream media.

### The transparency of data

Transparency is one of the epistemological dimensions of data journalism,<sup>32</sup> which suggests the possibility of revealing the processes used in the construction of news. This is a resource endorsed by the agencies the Bureau of Investigative Journalism, *Agência Pública* and *Gênero e Número* using hyperlinks that give access to documents and databases organized in spreadsheets by reporters. In other words, these

---

<sup>26</sup> Simon Rogers, *Facts are Sacred* (London: Faber and Faber, 2013), 70.

<sup>27</sup> Bardoel, Jo and Deuze, Mark, “Network journalism: converging competences of old and new media professionals,” *Australian Journalism Review* 23, no. 2 (December, 2001): 95, <https://hdl.handle.net/2022/3201>.

<sup>28</sup> James Ettema, and Theodore Glasser, *Custodians of Conscience: Investigative Journalism and Public Virtue* (New York: Columbia University, 1998).

<sup>29</sup> Michael Karlsson, “Rituals of Transparency. Evaluating Online News Outlet’s Uses of Transparency Rituals in the United States, United Kingdom and Sweden,” *Journalism Studies* 11, no. 4, (July 2010): 535–545, <https://www.tandfonline.com/doi/abs/10.1080/14616701003638400>; Patrick L. Plaisance, “Transparency: an assessment of the Kantian roots of a key element in media ethics practice,” *Journal of Mass Media Ethics* 22, no. 2–3, (December, 2007): 187–207, <https://www.tandfonline.com/doi/abs/10.1080/08900520701315855>.

<sup>30</sup> Mark Coddington. “Clarifying journalism’s quantitative turn: A typology for evaluating data journalism, computational journalism and computer-assisted reporting,” *Digital Journalism* 3, no. 3 (November 2015): 340, <https://www.tandfonline.com/doi/full/10.1080/21670811.2014.976400>.

<sup>31</sup> Seth Lewis and Nikki Usher, “Open source and journalism: Toward new frameworks for imagining news innovation,” *Media, Culture & Society* 35, no. 5, (June 2013): 602–619, <https://journals.sagepub.com/doi/abs/10.1177/0163443713485494>; Sylvain Parasie, “Data Journalism and the promise of transparency,” in *The Routledge companion to media and scandals*, ed. Howard Tumber, and Silvio Waisbord (London: Routledge, 2019), 263–272.

<sup>32</sup> Coddington, “Clarifying,” 340.

organizations open the sources of information to the reader. This practice conducts facticity.<sup>33</sup> The ritual of transparency is a performative act to prove true narratives.<sup>34</sup> Thus, it merges with notions of verification, objectivity, and authority.<sup>35</sup>

Karlsson observes that this praxis allows the establishment of a “common faith” between those who produce the news and those who consume it.<sup>36</sup> It is a counterpoint to the idea of journalistic writing as a “black box,”<sup>37</sup> whose functioning is somewhat impenetrable. It is, therefore, a transparency that is related to production processes and how these encompass newsgathering and the ability of verification by the audience. But reception research conducted by Karlsson, Clerwall and Nord published in 2014 found that only the combination of internal and external links – to a PDF document – was significant for the credibility.<sup>38</sup>

The consolidation of public transparency,<sup>39</sup> advocated by open data movements, has greatly contributed to the spread of data journalism – in association with the technical expertise developed to process, collect, and analyze

<sup>33</sup> Juliette De Maeyer, and Avery Holton, “Why linking matters: A metajournalistic discourse analysis,” *Journalism* 17, no. 6 (August, 2016): 78, <https://journals.sagepub.com/doi/10.1177/1464884915579330>.

<sup>34</sup> Karlsson, “Rituals of Transparency,” 536; Plaisance, “Transparency”; Klaus Meier, “Transparency in Journalism: Credibility and trustworthiness in the digital future,” *Conference Future of Journalism*, Cardiff, (September 2009); Arthur Hayes, Jane Singer, and Jerry Ceppos, “Shifting Roles, Enduring Values: The Credible Journalist in a Digital Age,” *Journal of Mass Media Ethics* 22, no. 4 (December, 2007): 277, <https://www.tandfonline.com/doi/abs/10.1080/08900520701583545>; Verica Rugar, “How Did You Find That Out? Transparency of the Newsgathering Process and the Meaning of News,” *Journalism Studies* 7, no. 1 (August, 2006): 127–143, <https://www.tandfonline.com/doi/abs/10.1080/14616700500450426>; David S. Allen. “The trouble of transparency,” *Journalism Studies* 9, no. 3 (April, 2008): 323–340, <https://www.tandfonline.com/doi/abs/10.1080/14616700801997224>.

<sup>35</sup> Gaye Tuchman, “Objectivity as Strategic Ritual: An Examination of Newsmen’s Notion of Objectivity,” *American Journal of Sociology* 77, no. 4 (January, 1972): 660–679, <https://www.journals.uchicago.edu/doi/abs/10.1086/225193>; Bill Kovach, and Tom Rosenstiel, *Elementos do jornalismo: o que os jornalistas devem saber e o público exigir* (São Paulo: Geração Editorial, 2003), 126–129.

<sup>36</sup> Karlsson, “Rituals of Transparency,” 537.

<sup>37</sup> Jane Singer, “The Political J-Blogger: ‘Normalizing’ a New Media Form to fit Old Norms and Practices,” *Journalism* 6, no. 2 (May, 2005): 181, <https://journals.sagepub.com/doi/10.1177/1464884905051009>.

<sup>38</sup> Michael Karlsson, Christer Clerwall, and Lard Nord, “You ain’t seen nothing yet. Transparency’s (lack of) Effect on Source and Message credibility,” in *The Future of Journalism: In an age of digital media and economic uncertainty*, ed. Bob Franklin (New York, Routledge, 2017), 456–459.

<sup>39</sup> In the research conducted with the news agencies, the concept of transparency is related to information from primary sources. There is another meaning of the word that is not covered by this study that is related to delivering transparency about the editorial line.

big data – which, ultimately, is in line with democratic ideals.<sup>40</sup> The institution of the *Freedom of Information Act* (FOIA), in 1966, in the USA – which obliges the provision of data that is outside the public domain – further expanded the “cycle” of transparency. In Brazil, a similar law – the *Access to Information Law* (LAI) – was enacted in 2011, during the government of Dilma Rousseff of the *Brazilian Labour Party*.

Transparency versus opacity is one of the typologies created by Coddington while reflecting on the epistemology that surrounds the professional field that encloses computational journalism, data journalism and data-driven journalism, and open data journalism.<sup>41</sup> On the grounds that ODJ exercises the choice of an open-source routine, however, nuances are erased between those specialties. Transparency of methods consists in revealing the way of doing things – or of newsgathering, which is the process of collecting news, with links to the drafts of the investigation. When analyzing the effects of this type of practice, Meier invokes Neuberger to state that, in this way, transparency allows the audience to assess the quality of the report, which upraises the credibility of what is being published.<sup>42</sup>

Without directly referring to ODJ, studies on transparency raise questions related to the credibility of journalism in the digital age.<sup>43</sup> Hyperlinks (or simply links) are seen as a resource that materializes this transparency. As Coddington discusses the theme of transparency,<sup>44</sup> the reflections of the authors of such studies can apply to ODJ practices. An example is the work of Singer,<sup>45</sup> who observes the reality of blogs that provide links to original articles – primary sources – and to

---

<sup>40</sup> Sylvain Parasié, and Eric Dagiral, “Data-Driven Revelation? Epistemological tensions in investigative journalism in the age of big data,” *Digital Journalism* 3, no. 3 (November 2014): 364–380, <https://www.tandfonline.com/doi/abs/10.1080/21670811.2014.976408>; Sylvain Parasié, “Data Journalism and the promise of transparency, in: *The Routledge companion to media and scandals*, ed. Howard Tumber, and Silvio Waisbord (London: Routledge, 2019); Lewis and Usher, “Open Source,” 602–619.

<sup>41</sup> Coddington, “Clarifying,” 340.

<sup>42</sup> Klaus Meier, “Transparency,” *Conference Proceedings of the Future of Journalism* (Cardiff, 2015); Christoph Neuberger, “Objektivität,” In *Handbuch Journalismus und Medien*, ed. Siegfried Weischenberg, Hans Kleinsteuber, and Bernhard Pörksen (Konstanz: UKV, 2005), 325–328.

<sup>43</sup> Rupal, “How did you find that out?,” 127–143; Singer, “The Political J-Blogger,” 173–198; Hayes, Singer, and Ceppos, “Shifting Roles, Enduring Values,” 262–279; Allen, “The Trouble,” 323–340; Meier, “Transparency”; Karlsson, “Rituals of Transparency,” 535–545; Michael Karlsson, “The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority,” *Journalism* 2, no. 3 (April, 2011): 279–295, <https://doi.org/10.1177/1464884910388223>; Maeyer, and Holton, “Why linking matters,” 776–792.

<sup>44</sup> Mark Coddington. “Defining and mapping data journalism and computational journalism.” In *The Handbook of Developments in Journalism Studies*, ed. Scott Eldridge, and Bob Franklin (London: Routledge, 2018), 225–236.

<sup>45</sup> Singer, “The Political J-Blogger,” 173–198.

verifiable information such as press releases. While she considers how some press releases are not very illuminating, Singer notes that the study points to the redemption of authority. In a more recent article, De Maeyer and Holton address the value of links and define categories such as good links:

One of the prominent discourses associated with the ‘good’ link was that of the ability to show sources. Links are useful for journalists, as our sample pointed out, because they allow them to directly point to the original material that they used to build their story. But there are two underlying purposes to this argument. On the one hand, links to sources produce credibility. Conversely, links to sources ensure credit is given where credit is due. [...] Links help journalists to ensure credibility when they point to original material such as documents, data, and other primary sources. In this case, links are used as a demonstration of facticity.<sup>46</sup>

By providing the sources of information, Meier<sup>47</sup> understands that this form of transparency gives journalism the reliability of the scientific method, as established by Neuberger.<sup>48</sup> However, transparency is not a panacea, and the questioning comes from the professionals involved in the production of ODJ. The design director at the agency *Género e Número*, Marília Ferrari, considers that even the choice of words and the visualization of the data act as filters, in a process to make the data understandable for those who access it:

So, there are several filters to access the database. There are several steps to maintain this transparency. There’s no way to avoid framing. Making data available is a very big contribution to transparency, due to offering real access to that. In the portals, even if the data is accessible, it is not accessible in a very coherent way. This is a job that we try to do, to clean the bases and organize it for the readers.<sup>49</sup>

Ferrari’s testimony confirms that the role of transparency in the context of journalism involves complexities, as estimated by Allen.<sup>50</sup> If, on the one hand, it establishes legitimacy, on the other hand, it offers more information about processes and methods. Therefore, it is ammunition for possible challenges to authority.

---

<sup>46</sup> De Maeyer and Holton, “Why linking matters,” 781.

<sup>47</sup> Meier, “Transparency.”

<sup>48</sup> Neuberger, “Objektivität,” 327.

<sup>49</sup> Marília Ferrari in interview with the author on February 9, 2021.

<sup>50</sup> David S. Allen, “The Trouble,” 323–340.

## Methodology

The theoretical-methodological framework of this research is based on the “rule of transparency” forged by Kovach and Rosenstiel, which establishes the principle of verification.<sup>51</sup> Based on the paradigm of the scientific method, it is grounded on the assumption that experimental proof brings objectivity and reliability on a scientific basis. This concept coalesces into the idea of journalism as a discipline of verification. Kovach and Rosenstiel’s axiom were inspired by the precision journalism introduced by Philip Meyer in the 1970s.

The proposal of verification underpins studies on transparency.<sup>52</sup> Karlsson sustains that the ‘rituals’ of Tuchman provides discernible elements for public scrutiny which are related to standardized routines.<sup>53</sup> The practices of open data journalism (ODJ) evoke an intersection between Meier’s notions of “disclosure transparency”<sup>54</sup> and Tuchman’s conception of a particular form of journalistic strategic performance. The latter defines supporting evidence and discernible procedures – such as quotation marks – as strategic rituals handled by journalists to convey objectivity to their routines.<sup>55</sup>

The process of deconstructing a story was taken in account in order to trace discernible elements that indicates good practices in Open Data Journalism. Several researchers served as an important guide for the creation of an experimental methodology.<sup>56</sup> An extensive bibliographical review intertwined to

---

<sup>51</sup> Bill Kovach, and Tom Rosenstiel, *Os elementos*, 128.

<sup>52</sup> Meier, “Transparency”; Neuberger, “Objektivität”; Karlsson, “Rituals”; Singer, “The Political J-Blogger”; Allen, “The Trouble.”

<sup>53</sup> Karlsson, “Rituals,” 536.

<sup>54</sup> Meier, “Transparency.”

<sup>55</sup> Tuchman, “Objectivity,” 119.

<sup>56</sup> James Ettema, and Theodor Glasser, “On the epistemology of Investigative Journalism,” *Communication* 8 (August, 1984): 183-206; Manuel Pinto, “Fontes jornalísticas: contributos para o mapeamento do campo,” *Comunicação e Sociedade* 14, no. 1-2 (2000): 277-294, [https://repositorium.sdum.uminho.pt/bitstream/1822/5512/1/CS\\_vol2\\_mpinto\\_p277-294.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/5512/1/CS_vol2_mpinto_p277-294.pdf); Joaquim Fidalgo, “A questão das fontes nos códigos deontológicos dos jornalistas,” *Comunicação e Sociedade 2 – Série Comunicação Cadernos do Noroeste* 14, no. 1-2 (October 2000): 319-337, [https://repositorium.sdum.uminho.pt/bitstream/1822/5513/1/CS\\_vol2\\_jfidalgo\\_p319-337.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/5513/1/CS_vol2_jfidalgo_p319-337.pdf); Bill Kovach, and Tom Rosenstiel, *Os elementos*, 127-129; Josenildo Guerra, “Ranking Q-Avalia da qualidade jornalística no Brasil e em Portugal 2018: uma avaliação experimental,” *Estudos em Jornalismo e Mídia* 17, no. 1 (June 2020): 54-74, <https://periodicos.ufsc.br/index.php/jornalismo/article/view/1984-6924.2020v17n1p54>; Rogério Christofoletti, “Indicadores de Qualidade no Jornalismo: políticas, padrões e Preocupações de jornais e revistas brasileiros,” *Série Debates CI (Unesco)* 3 (2010): 1-61, <https://unesdoc.unesco.org/search/22058382-ed11-4c5f-8e4c-59f5be8dffed>; Luis Augusto Egyto Cerqueira, “Qualidade Jornalística: ensaio para uma matriz de indicadores,” *Comunicação e Informação/ Série Debates CI*, no. 6 (Novembro, 2010): 1-33, <https://unesdoc.unesco.org/ark:/48223/pf0000189918>; Stephen Lacy, and Tom Rosenstiel, “Defining and Measuring Quality Journalism,” *Rutgers School of Communication and*

newsgathering routines knowledge allowed the identification of elements that disclose transparency through ODJ. The methodology is grounded in original indexes and subindexes. The indexes *method* and *evidence* reveal competences adopted in ODJ practices. Together, the indexes are the parameters toward the verification of excellence procedures on the journalistic routines of ODJ – and therefore attest the level of transparency on each piece of reporting. The subindexes operate as clues that lead to the macro indexes evaluation.

The literature review aligned with previous reading of part of the research corpus helped the deduction of the indexes and subindexes. The corpus is formed by 153 outputs from *Género e Número*, *Agência Pública*, and the Bureau of Investigative Journalism.<sup>57</sup> The sample covers output published on the three agencies' websites from their early years until 2021. It encompasses pieces from the section *Dados Abertos (Open Data)* by *Género e Número*<sup>58</sup> and the section *Dados (Data)* by *Agência Pública*. As TBIJ had ODJ spread throughout the site, the criteria were awarded pieces and content from the section Bureau Local which offers access to all its data resources.<sup>59</sup> Publica discontinued *Dados* in 2021 and subsequently publishes ODJ in all sections of the site.

Each subindex has its own criteria, as a starting point in the content analysis process. They function as unities of analysis that are related to journalistic concepts what provides consistency to the experimental methodology originally developed for the research in focus.<sup>60</sup> They are *themes* that circumscribe the step-by-step of the reporting concoction.<sup>61</sup> Some reveals practices that have arose under de emergency of digital journalism or the open data journalism; others are part of traditional routines. The subindexes are: *methodology* and *expertise* (both information accessible to readers on the websites), *open data* (when there is real access to backstage material),<sup>62</sup> *data sources* (the data origin and its volume), *usual sources* (the complementary sources),<sup>63</sup> *off-the-record sources* (a negative element

---

*Information* (March, 2015): 1–66, <https://www.issuefab.org/resources/31212/31212.pdf>; Tom Rosenstiel, Marion Just, Todd Belt, Atiba Pertilla, Walter Dean, and Dante Chinni, *We Interrupt this Newscast: How to improve local News and win ratings too* (New York: Cambridge University Press, 2007), 194; Adegboyega Ojo, and Bahareh Heravi, "Patterns in Award Winning Data Storytelling," *Digital Journalism* 6, no. 6 (November, 2017): 693–718, <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1403291>; Wiebke Loosen, Julius Reimer, and Fenja Schmidt, "When data becomes News: A content analysis of data journalism pieces," *Conference Proceedings of the Future of Journalism* (Cardiff, 2015); Joyce Stooibrant, "Finding the News and mapping the links: a case study of hypertextuality in Dutch-language healthy News website," *Information, Communication & Society* 22, no. 14 (Jun 2018): 2128–2155, <https://doi.org/10.1080/1369118x.2018.1477971>; Trust Project: <https://thetrustproject.org/trust-indicators/>.

<sup>57</sup> The corpus consists of 53 reports from *Agência Pública*, 47 reports from *Género e Número*, and 35 from TBIJ.

<sup>58</sup> <https://www.generonumero.media/dados-abertos/>.

<sup>59</sup> <https://www.thebureauinvestigates.com/local/open-resources>.

<sup>60</sup> Earl Babbie, *The Practice of Social Research*. (Belmont: Ca: Wadsworth, 1989), 124–125.

<sup>61</sup> Robert P. Weber, *Basic Content Analysis* (Newbury Park/Ca: Sage, 1990), 18.

<sup>62</sup> As pointed before, not all data journalism offer access to reporting backstage being this procedure the main difference that explains the proposed ODJ specialty.

<sup>63</sup> Official, Institutional, Commercial, Specialists, Activist/NGO, off-the-record, social media

considered lack of accuracy)<sup>64</sup> *visualization* (how much it increases the information), *balanced point of view* (fairness), and *hyperlink effectiveness* (how much the links to the open data actually work).

The subindex *data sources*, for example, was measured through type of sources of information: official, NGOs; personal data (created by the news agency's teams); *Freedom of Information* (FOI)/ *Lei de Acesso à informação* (LAI),<sup>65</sup> human data (*voices scraped from data*), and digital documents. Two studies support the creation of this subindex: a content analysis conducted by Loosen, Reiner and Schmidt on data journalism awarded pieces, and the research on narrative patterns led by Ojo and Heravi.<sup>66</sup>

The previous analysis of the reports showed that data journalism from the three agencies uses sources that do not originate from scraped data – coined as usual sources or complementary ones – such as face to face, telephone or email interviews with authorities, specialists, and characters. The number and type of sources are parameters for accuracy and quality observed by Lacy and Rosenstiel in the Project for Excellence in Journalism; scholars understand that sources and documents point out the rigor in the newsgathering method.<sup>67</sup>

The content analysis was organized in Excel. Each agency has its own Excel book. After a painstaking reading piece by piece, the data collection in absolute numbers was registered in each correspondent agency Excel book in nine spreadsheets created exclusively for each subindex.

The sum obtained for each subindex represent specific categories. For each category there is a corresponding weight. The methodology on Excel allows crossovers and findings on a percentual criterion. For data sources, for example, the ranks are *very strong*, *strong*, *moderate*, and *weak* (Figure 1). The example of weight for the subindex data sources, following criteria suggested by Lacy and Rosenstiel; Manuel Pinto, Joaquim Fidalgo, Ettema & Glasser, Cerqueira, Ojo and Heravi and Loosen.<sup>68</sup>

---

and character.

<sup>64</sup> Kovack and Rosenstiel delineate that if the source is broadly informed, the audience can judge the reliability (Kovack, and Rosenstiel, *Os elementos*, 140); Fidalgo, in turn, highlights how the journalistic deontological codes define the source identification as a rule of transparency although anonymity is a legitimate right (Fidalgo, “A questão das fontes,” 326).

<sup>65</sup> LAI (*Lei de Acesso à Informação*) corresponds to *Freedom Of Information* in Brazil and was established in November 2011 to force institutions to open data that is not publicly available.

<sup>66</sup> Wiebke Loosen, Julius Reimer, and Fenja Schmidt, “When data,” *Conference Proceedings of the Future of Journalism* (Cardiff, 2015); Adegboyega Ojo, and Bahareh Heravi, “Pattern,” 693–718.

<sup>67</sup> James Ettema, and Theodor Glasser, “On the epistemology,” 183–206; Manuel Pinto, “Fontes jornalísticas,” 277–294; Joaquim Fidalgo, “A questão das fontes,” 326; Rogério Christofolletti, “Indicadores,” 1–61.

<sup>68</sup> Lacy, and Rosenstiel, “Defining and Measuring Quality Journalism,” 1–66; Pinto, “Fontes jornalísticas,” 277–294; Cerqueira, “Qualidade Jornalística,” 1–31; Ojo, and Heravi, “Patterns,”

	Minimum	Maximum	Weight
Weak	1	1	1
Moderate	2	3	2
Strong	3	5	3
Very strong	Above 5		4

Figure 1 – The weight of the subindex data sources comes from the quantitative content analysis that collates the number of appearance of each type of source of data in each reporting (official, NGOs; personal research data; *Freedom of Information* (FOI)/ Lei de Acesso à informação (LAI),<sup>69</sup> human data (voices), and digital documents). The same process was used for all other subindexes (collation of absolute number converted in weightages).

For visualization the categories are *no elements*, *low elements*, *some elements*, and *many elements*. The subindex *visualization* encompasses tables, graphs, animation, interactive, and forms (reports, letters, press releases, etc.). Open data is 0 if there is no data accessible and 1 for data available via links. Hyperlink effectiveness is ranked as extremely broken if only 40% work, highly broken if between 40% or 70% of the links work, faulty if between 70% and 85% of the links work and ok if more than 85% of the links work. For the content analysis, only *genuine* links as described were considered – they lead the readers to digital documents, spreadsheets or googledocs and, therefore, offer evidence of the reporting backstage.<sup>70</sup> Giuliana Bianconi – editor of *Gênero e Número* – proclaims that

(...) data brings precision and allows us to visualize fragments from a source that is not declaratory. (...) This is our context through a more scientific method and processes. Then you ask: what about rigor? We try to bring this consistency in the analysis. You need to validate the data yourself. It is a constant process of validating the data source. And we validate it based on the credibility of the organization that create the data, reading the methodology that was used (...) I understand that we create an additional layer when we do open data journalism. Transparency is part of the process.<sup>71</sup>

The subindexes that measure the index *evidence* are a crossover of data sources, usual sources, off-the-record sources, visualization, and hyperlink effectiveness in *Figure 2*. Together, those subindexes bring evidence of the newsgathering.

693–718; Wiebke Loosen et al. “When data becomes news,” *Conference Proceedings of the Future of Journalism* (Cardiff, 2015).

<sup>69</sup> LAI (Lei de Acesso à Informação) corresponds to *Freedom of Information* in Brazil.

<sup>70</sup> De Maeyer, and Holton, “Why linking matters,” 783; Joyce Stooibrant, “Finding the News,” 2041.

<sup>71</sup> Giuliana Bianconi in interview with the author on January 18, 2021.

*Evidence evaluation*

Subindex	Performance	Weight
Data sources Usual sources	Weak	1
	Moderate	2
	Strong	3
	Very Strong	4
Off-the-record sources	Absence	0
	Poor (50%)	-1
	Very poor (75%)	-2
	Extremely poor (100%)	-3
Visualization	No elements	1
	Low elements	2
	Some elements	3
	Many elements	4
Hyperlink Effectiveness	Extremely broken (<=40%)	1
	Highly broken (>40 to 70%)	2
	Faulty broken (>70 to 85%)	3
	OK (>=85%)	4
Open data	Absence	0
	Existent	1

Figure 2 – Categories for subindexes that measure Evidence and their respective weights.

The categories and weights of the subindexes that measure the index *method*: data sources, usual sources, expertise, methodology and balance point of view and its criteria in Figure 3. Together those subindexes point out the rigor in the newsgathering. The subindex methodology, for example, brings the method used to achieve the information scraped by the teams.

*Method evaluation*

Subindex	Performance	Weight
Data sources Usual sources Expertise	Weak	0
	Moderate	1
	Strong	2
	Very strong	3
Methodology	Non-existent	0
	Existent	1
Balanced point of view	Lack diversity	0
	Low diversity	1
	High diversity	2

Figure 3 – Categories for subindexes that measure Method and their respective weights.

The weightage variables were originally developed for this experimental methodology and lead to automatic results in Excel that consolidated percentual for the indexes *method* and *evidence*. As part of the process, these macro criteria have their own categories. For evidence: *very strong*, *strong*, *moderate*, and *weak*. For method: *incomplete*, *partial*, and *complete*.

If the sum of the quantitative content analysis (in absolute numbers) of the subindexes data sources, usual sources, expertise, methodology and balance point of view is over 14 the reporting is defined as complete for method. Between 13 and 8 is partial and below 8 is incomplete. For *Evidence*, the reporting is ranked as strong if the sum of weights of the subindexes data sources, usual sources, off-the-record sources, visualization, and hyperlink effectiveness is over 10 points. Figure 4 shows the table for *Evidence*.

Table for Evidence

	Inferior	Superior
Very strong - <= 24	14	-
Strong <= 15	10	14
Moderate <= 10	5	10
Weak <=3	0	5

Figure 4 – The weightage for the index Method.

Excel brings all the results automatically and permits a several number of crossovers and comparisons between subindexes and between indexes, especially with the use of filters. As the subindexes are collated, Excel automatically calculates both results on the spreadsheets for method and evidence and through the sum/total obtained by each reporting (issue), it shows which category is defined. The outputs are classified as issues (Figure 5).

	D	E	F	G	H	I	J	K	L
	Data source	Usual Sources	Off-the-record	Visualization	Hyperlink Effectiveness	Open Data	Total	Evidence	
2	Issue_1	4	4	-1	2	4	1	14	Strong
3	Issue_2	4	4	0	1	3	1	13	Strong
4	Issue_3	4	4	-1	2	4	1	14	Strong
5	Issue_4	4	4	-1	1	4	1	13	Strong
6	Issue_5	4	4	-1	1	4	1	13	Strong
7	Issue_6	4	4	-1	1	3	1	12	Strong
8	Issue_7	2	4	-1	1	4	1	11	Strong
9	Issue_8	3	4	-1	1	4	1	12	Strong
10	Issue_9	2	4	-1	1		0	6	Moderate
11	Issue_10	3	3	-1	2	4	1	12	Strong
12	Issue_11	2	4	0	2	2	1	11	Strong
13	Issue_12	2	4	-1	2	4	1	12	Strong

Figure 5 – The automatic results for Evidence through the data crossing of subindexes. This is a fragment of the sample from The Bureau of Investigative Journalism with total of 43 pieces. Each issue corresponds to a reporting.

As Excel allows all types of crossovers, the crossing of results for method and evidence reveals how much transparency the outputs deliver to the readers – the ability of verification. The criteria for transparency are *high*, *medium*, *low*, or *zero* transparency. The methodology allows the filtering and perception of quantitative inferences about each subindex and how these performances interfere with the level of transparency.

The sixteen in-depth interviews with data journalists from the three agencies helped in confronting the results of the content analysis, looking into contradictions, and amplifying conclusions. Analytical categories include (1) *verification as a path to accuracy*, (2) *rigor in data gathering* and (3) *reader-friendly fact-checking*.

### Levels of transparency

This original and experimental proposal was developed from the hypothesis that the open data journalism practiced by *Agência Pública*, *Gênero e Número*, and the Bureau of Investigative Journalism allows the verification of the content of a report and, in this way, contributes to reinforcing the verification of the information. The aim was to apply the “rule of transparency” and observe how evident the effort exerted by professionals is in delivering verifiable content. The macro indexes of method and evidence – that embrace rigor and expert practices – are cornerstones for the evaluation of transparency fabricated through open data journalism. Transparency as a path to share evidence with the audience is crucial for the quality of news.<sup>72</sup>

The exhaustive analysis, recorded on Excel spreadsheets, identified attributes such as volume of data sources and complementary sources that reporters also make use of; the level of diversity (substantial appearance of different points of view); professional experience of the reporter or team; the inclusion of the methodology about the data gathering; how effective the access to the data is (through verification of the functioning of the links); and the contribution provided by visualization. Method and evidence – out of a combination of subindexes – bring a big picture that reveal how each piece of reporting performs in terms of elements of evidence that provide transparency.

These are the results for the *method* (*complete*, *partial*, and *incomplete*) in Figure 6.

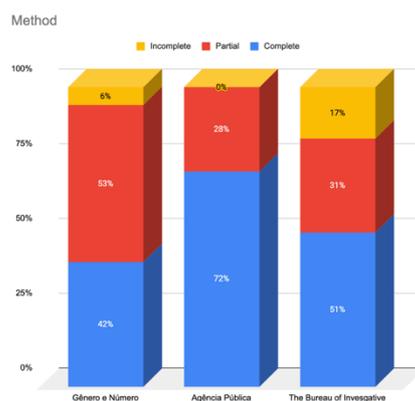


Figure 6 – Results for the index Method with comparison of the three agencies.

<sup>72</sup> Stephen Lacy, and Tom Rosenstiel, “Defining and Measuring Quality Journalism,” 1–66.

These are the results for the index *evidence* (*weak*, *moderate*, *strong*, and *very strong*) in *Figure 7*.

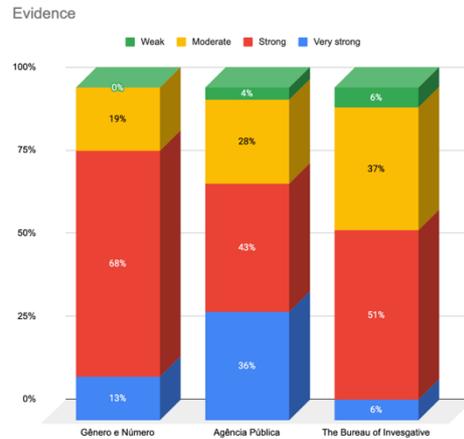


Figure 7 – Results for the index Evidence with comparison of the three agencies.

A procedure was used to determine the level of transparency: the data crossing of the best performance for method (*complete*) and the best performance for evidence (*very strong*). *Agência Pública* was the one that achieved *high transparency*. Fifteen of the forty-seven reports analyzed from the section Dados (Data) were measured with the best performance – *complete method* and *very strong evidence* which led to 25% transparency. GN has six between fifty-three stories ranked with the best performance, which reveals 11% level of transparency. TBIJ has only two out of thirty-five with the highest rating on performance, which brings the result of 5% transparency. These percentages can be better visualized in *Figure 8*.

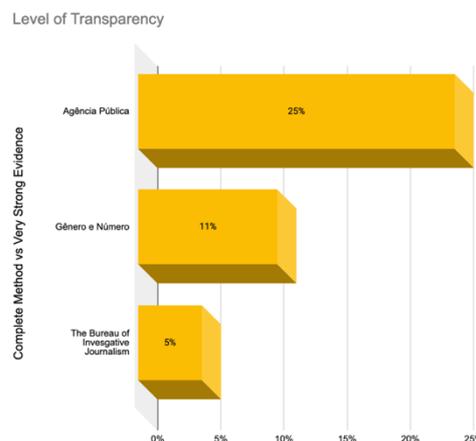


Figure 8 – Results for the index Transparency from a crossover of Method and Evidence.

For an accurate understanding of the results for transparency, a second path was taken. It was necessary to scrutinize the performances of the subindexes that underpin the findings around *method* and *evidence*. The *hyperlink effectiveness* in

83% of *Agência Pública*'s reporting shows that readers manage to access most of the spreadsheets and documents collated by the team. *Balanced point of view* reveals that 72% of the sample has *high diversity*. *Expertise* also has a high score: *very strong* in 43% and *strong* in 57% of the reports. With a clear emphasis on visualization, 28% and 23% of the output are framed respectively on the categories *many elements* and *some elements*. About data sources, 57% of the production is *strong* and in 32% *moderate*. The content analysis reveals that the team searches for usual sources frequently: 55% of the reports are defined as *very strong* and 28% as *strong*. Even without methodology in 68% of the corpus, overall *Agência Pública* still achieves the best performance, hence shows *high transparency*.

Confined to the parameter adopted to infer the degree of transparency, *GN* delivers *medium transparency*, although its performance is only around 6 points ahead of *TBIJ*. Between six and nine subindexes applied on *GN* output perform well in the evaluation of method and evidence. All reporting allows access to data in the section *Dados abertos* (Open data) with a hyperlink effectiveness of 81%. The evaluation pointed to *very strong* and *strong* expertise in, respectively, 34% and 47% of the fifty-three articles analyzed. Data sources were highly rated (42% *very strong* and 25% *strong*), such as usual sources (32% *very strong* and 32% *strong*). But the factors that detract from transparency are the absence of methodology in 55% of the corpus, less visualization (57% with *low elements* and 13% with *none*) and, especially, the low performance in balanced point of view (45% with *low diversity* and 21% *none*).

*TBIJ* shows weak performance in at least five of the nine subindexes used to measure method and evidence, which leads to *low transparency* for the British news agency. An aspect that proved to be preponderant in the result is that balanced point of view is compromised – there is a deficit in 46% and it is *low* in 20% of the output. Visualization does not contribute to the evidence, since 51% of the reports have *low elements* and there are *no elements* in 41%. The lack of uniformity is evident in the assessment of usual sources: while 60% were rated very strong, 11% and 23% were rated, respectively, as moderate, and weak. Only 66% of the reporting includes the methodology. And, if the agency offers open data in 86% of the output verified, the hyperlink effectiveness is moderate in 43% of the production, which highly depreciates the transparency. Expertise is the highlight of the British agency, where among the professionals 40% and 51% were quoted with very strong and strong experience, respectively. It should be noted that the section *Bureau Local* has all links active and offers access to “open resources” Where the reader will find “reporting recipes”<sup>73</sup> in *Figure 9*.

---

<sup>73</sup> The Open Resources section of *Bureau Local* leads the reader to the reporting recipes and data related to the project *Dying Homeless* – a series about death in the streets of the UK: <https://docs.google.com/document/d/1zUHUPn9-UfVEPdzi4qkGsTr7jNtqxTIyUEuQdtiQh3o/edit>.

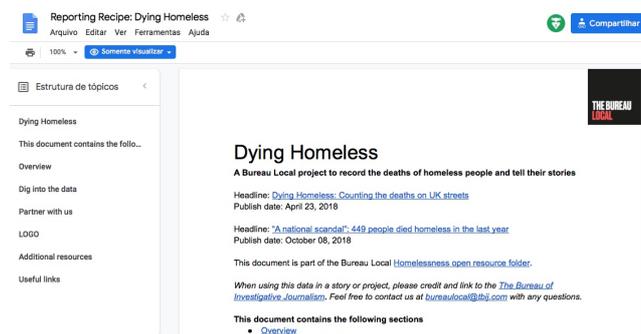


Figure 9 – Reporting Recipe organized by the team of Bureau Local/TBIJ with access to data spreadsheets, Google Docs, digital documents, and advice about sharing content policies.

To broaden inferences, a subjective analysis considered not only the percentage found after the confrontation of *method* and *evidence*, but also, to discover nuances and peculiarities, performances of each subindex were observed.

It is notable how the subindex *balanced point of view* contributes to improving the performance of the index method. *Agência Pública* achieves 72% of *high diversity*. If the endeavor in building quality news becomes evident in the analysis of this agency, the absence of methodology in 35% of the thirty-four reports quoted as *complete* is significant. But the efficiency found in other subindexes (data sources, usual sources, and balanced point of view) indicates how much accuracy is important to produce verifiable content. For example, 85% of the reporting from *Agência Pública* provides access to official or digital documents (in thirty-eight out of forty-seven stories). The number of documents accessible via links determines a *very strong* performance in the subindex data sources for *Agência Pública*, which affects the results for the index method in a decisive way. The presence of documents in a report enhances its credibility.<sup>74</sup>

Effectiveness is a new challenge that arises with the use of hyperlinks, one of the main components of online journalism.<sup>75</sup> It is only possible to take advantage of the unlimited database scale offered on digital platforms if the links published on websites are active. In the in-depth interviews, data journalists recognized the obstacles they face during migration processes that lead to broken links – which disrupt access to data and therefore jeopardize the fact-checking capability of ODJ. Although the TBIJ presents *complete method* in 51% of its analyzed output, in general,

<sup>74</sup> Ettema, and Glasser, *Custodians*, 13–14, 132–151, 200.

<sup>75</sup> Javier Diaz Noci, and Ramon Salaverría, “Hipertexto periodista: teorías y modelos,” in *Manual de redacción ciberperiodista*, ed. Javier Diaz Noci, and Ramon Salaverría (Barcelona: Ariel Comunicación, 2003), 81–139; Marcos Palácios, “Ruptura, continuidade e potencialização no journalism on-line,” in *Modelos de jornalismo digital*, ed. Elias Machado (Salvador: Calandra, 2003), 13–37.

the performance is undermined due to its poor effectiveness. This affects evidence of elements and, consequently, performative transparency. The comparison between agencies for *hyperlink effectiveness* in Figure 10.

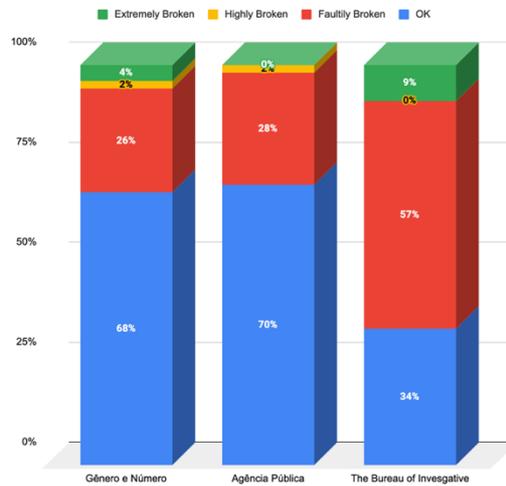


Figure 10 – Results for the subindex Hyperlink Effectiveness.

The analysis shows that *off-the-record* sources (57% in TBIJ, 32% in GN, and 25% in *Agência Pública*) as well as *few visual elements* interfere in the result for the index *evidence*. The absence of methodology did not compromise the performance of *Agência Pública*. The analysis, therefore, points to the importance of a balance between all the subindexes because, together, they grounded the performance of *method* and *evidence* and, hence, transparency. Below is the graph for *balanced point of view*, reveals the agency with a low level of transparency – the Bureau of Investigative Journalism – appears mainly with low diversity (see Figure 11).

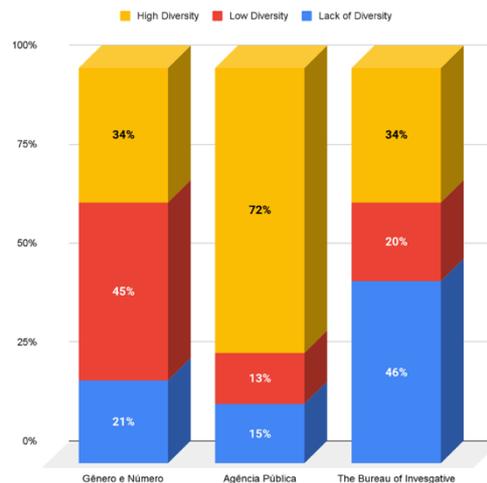


Figure 11 – Results for the subindex *Balanced Point of View*.

Outlined as a key factor in studies that aimed to assess news quality,<sup>76</sup> balanced point of view is a subindex that significantly affects the accomplishment of the index *method*. This fact suggests that elements of traditional newsgathering – such as usual sources beyond the data – are fundamental. This conclusion emerged in the interviews and was pointed out as an intrinsic part of ODJ routines.

The discrediting of journalism and its role according to normative theories of the media weakens its relevance and authority, an issue addressed in the book *The Journalism Manifesto*.<sup>77</sup> The interviews uncovered the concern among professionals about the importance of reliability in the data collection process. Precision involves a search for information originating in reliable databases with the support of statistical procedures. Another point raised is how crucial is the fact-checking of spreadsheets that will be shared with the audience through hyperlinks. Alexandra Heal and Maeve McClenaghan from TBIJ, data analyst Natália Leão and reporter Vitória Régia da Silva from GN, and editor and reporter Bruno Fonseca from *Agência Pública*, converged on the strategy of submitting the scraped data to specialists when there is any contradictory issue. On top of this, verification is an essential step when preparing the visualization. These conclusions led to two analytical categories: *verification as a path to accuracy* and *rigor in data gathering*.

The content analysis confirmed the hypothesis that the transparency of open data in the routines of ODJ allows the verification of the information published. Below is an example from TBIJ of how ODJ introduces the model of journalism in layers. A spreadsheet from the awarded piece “The Drone Warfare” that tracked US drone strikes in Pakistan, Afghanistan, Yemen, and Somalia from 2010 and 2020 (see *Figure 12*.)<sup>78</sup>

The screenshot shows a Google Spreadsheet with the following content:

Strike ID	Date
B1	01/01/2001
T1	02/02/2002

Notes section:

3 Please see our methodology for more information on our sources: <http://www.thebureauinvestigates.com/2011/08/10/pakistan-drone-strikes-the-methodology2/#sources>

4 The full methodology is available here: <http://www.thebureauinvestigates.com/2011/08/10/pakistan-drone-strikes-the-methodology2>

5 The full HTML datasets can be found here: <http://www.thebureauinvestigates.com/category/projects/drones/drones-graphs/>

6

7 The Bureau's work is licensed under creative commons. You are free to reuse this data but we ask you to please cite the Bureau as the source, and link back to this dataset. It is possible to download this spreadsheet however converting it into a Excel format will cause specific problems with the summary tables and the graphs because there are important differences between the coding of functions in Google drive and Excel.

8 More information on how to use the relevant functions in Excel: <https://support.office.com/en-in/article/SUMIFS-function-9dd8179e-cced-41d4-ac38-08f5b929e5>

9 More on our Creative Commons license: <http://creativecommons.org/licenses/by-nc-nd/3.0/>

10

11 If you have any questions about this spreadsheet, about how we're presenting the data here, or if you have any corrections to our figures, please contact us at the below email address and include "Pakistan drone data" in the subject line.

12 Contact the Bureau: [info@thebureauinvestigates.com](mailto:info@thebureauinvestigates.com)

13

14 The attacks here are listed in chronological order. This is a live database: the Bureau does not consider any record closed. Therefore casualty estimates for any strike could change when new, credible information comes to light that affects our understanding of a particular event.

15 Below are detailed notes explaining the coding, and column headings used in this spreadsheet:

16

17 **Strike ID** A four or five figure alphanumeric code identifying each strike. The letters signify under which President the strike was conducted (B for Bush, Ob for Obama, T for Trump). The code's numeric sequence is chronological: the higher the number, the more recent the strike. It does not signify which strikes were added to the database most recently.

18 **Date** In UK format: dd/mm/yyyy

Figure 12 – One of the many spreadsheets with information about the drone strikes on TBIJ.

<sup>76</sup> Rosenstiel, Just, Belt, Pertilla, Dean, and Chinni, *We Interrupt this Newscast*, 194.

<sup>77</sup> Barbie Zelizer, Pablo Boczkovski, and Christopher W. Anderson, *The Journalism Manifesto* (Cambridge: Polity Press, 2021).

<sup>78</sup> The Drone Warfare: <https://www.thebureauinvestigates.com/projects/drone-war>.

In the final scrutiny of *GN*'s performance, offering open data in 100% of the reporting in the section *Dados abertos* denoted a differential that favored this agency against TBIJ – especially as *GN* reveals higher hyperlink effectiveness than TBIJ. Journalists agreed that the evidence offered through access to spreadsheets and documents enhance the credibility of the reporting. But they are not sure about the engagement of the audience.<sup>79</sup> A reception research – such as the one conducted by Karlsson, Clerwall and Nord<sup>80</sup> – could give a real dimension of the effects of the transparency of ODJ through the critical lenses of the audience. None of the journalists informed the analytics around the access of the data by the readers. However, Rachel Oldroyd, former editor of TBIJ, exposed her point of view.

It doesn't matter. If one person goes to the data, it's because they are the right person. They have gone there for a reason. (...) It's worth to put the effort in if one person gets through. It's about building the trust, actually.<sup>81</sup>

However, the lack of hyperlink effectiveness – with a high record especially in TBIJ – shows how the exercise of verification that exposes good practices can impair transparency in practices of open data journalism. The reality of common broken links discloses one relevant issue: the importance of the preservation of accessible databases at websites. The links are only in the origin of the process of visibility. The research pointed out how data offered by the three agencies became an important and supportive source used in the public domain sphere. *GN* reinforced the documentary value of open data journalism by creating, in August 2021, the *GN Hub* portal, which offers open data research and analysis based on the expertise achieved by the agency through the practice of ODJ.<sup>82</sup>

---

<sup>79</sup> The researcher requested analytics for the website traffic, but they were not received. Rachel Oldroyd, former managing editor of TBIJ, said in the interview that even only one access is enough to justify the transparency policy.

<sup>80</sup> Karlsson, Clerwall, and Nord, "You ain't seen nothing yet," 456–459.

<sup>81</sup> Rachel Oldroyd in interview with the author in the newsroom of TBIJ in London on December 5, 2019.

<sup>82</sup> Available at <https://www.generonumero.media/hub/en/>.

## Bibliography

- Abramson, Jill. *Merchants of Truth: the business of news and the fight for facts*. New York: Simon & Schuster, 2019.
- Aguiar, Leonel Azevedo de. “A validade dos critérios de noticiabilidade no jornalismo digital.” In *Jornalismo on-line: modos de fazer*, ed. Carla Rodrigues (Rio de Janeiro: PUC-Rio, 2009), 163–192.
- Allen, David S. “The trouble of transparency,” *Journalism Studies* 9, no. 3 (April, 2008): 323–340.  
<https://www.tandfonline.com/doi/abs/10.1080/14616700801997224>
- Anderson, Christopher W. *Apostles of Certainty: Data Journalism and Politics of Doubt*. Oxford/Reino Unido: Oxford University Press, 2018.
- Babbie, Earl. *The practice of social research*. Belmont/Ca.: Wadsworth, 1989.
- \_\_\_\_\_, Bell, Emily, and Shirky, Clay. *Post-Industrial Journalism: Adapting to the Present*. Nova York: Tow Center for Digital Journalism, 2012.  
<https://doi.org/10.7916/D8NoiJS7>.
- Bertrand, Claude-Jean. *A deontologia das Mídias*. São Paulo: EDUSC, 1999.
- Borges-Rey, Eddy. “Unravelling data journalism: A study of data journalism practice in British newsrooms.” In *The Future of Journalism: Risks, Threats and opportunities*, edited by Stuart Allan, Cynthia Carter, Stephen Cushion, Lina Dencik, Inaki Garcia-Blanco, Janet Harris, Richard Sambrook, Karin Wahl-Jorgensen, Andy Williams, 833–843. New York: Routledge, 2015.
- Cerqueira, Luis Augusto Egyto. “Qualidade Jornalística: ensaio para uma matriz de indicadores,” *Comunicação e Informação/ Série Debates CI*, no. 6 (Novembro, 2010): 1–31.  
<https://unesdoc.unesco.org/ark:/48223/pf0000189918>.
- Christians, Clifford. G., Theodore L. Glasser, Denis McQuail, Kaarle Nordestreng, and Robert A. White. *Normative Theories of the Media*. Illinois: Illinois University Press, 2009.
- Christofolletti, Rogério. “Indicadores de Qualidade no Jornalismo: políticas, padrões e Preocupações de jornais e revistas brasileiros,” *Série Debates CI (Unesco)* 3 (2010): 1–61.  
<https://unesdoc.unesco.org/search/22058382-ed11-4c5f-8e4c-59f5be8dffed>.
- Coddington, Mark. “Clarifying journalism’s quantitative turn: A typology for evaluating data journalism, computational journalism and computer-assisted reporting,” *Digital Journalism* 3, no. 3 (November 2015): 331–348.  
<https://www.tandfonline.com/doi/full/10.1080/21670811.2014.976400>.
- \_\_\_\_\_. “Defining and mapping data journalism and computational journalism.” In *The*

- Handbook of Developments in Journalism Studies*, edited by Scott Eldridge, and Bob Franklin, 225–236. London: Routledge, 2018.
- Cornu, Daniel. “Da deontologia do jornalismo à ética da informação.” In *20 anos de jornalismo contra a indiferença*, edited by Ana Teresa Peixinho, Ana Teresa; Carlos Camponez; Isabel Vargues and João Coimbra Figueira, 102–119. Imprensa da Universidade de Coimbra, 2015.
- Deuze, Mark; Jo Bardoel. “Network journalism: converging competences of old and new media professionals,” *Australian Journalism Review* 23, no. 2 (December, 2001): 91–103.
- De Maeyer, Juliette. “The journalistic Hyperlink. Prescriptive discourse about linking in online news,” *Journalism Practice* 6, no. 5 (2012): 692–701.  
<https://www.tandfonline.com/doi/abs/10.1080/17512786.2012.667273>.
- \_\_\_\_\_, Juliette, and Avery Holton. “Why linking matters: A metajournalistic discourse analysis,” *Journalism* 17, no. 6 (August, 2016): 776–794.  
<https://journals.sagepub.com/doi/10.1177/1464884915579330>.
- Deuze, Mark, “Journalism and the web: An analysis of Skills and Standards in an online Environment,” *Gazette* 61, no. 5 (1999): 373–390.  
<https://journals.sagepub.com/doi/10.1177/1461444803005002004>.
- \_\_\_\_\_, and Tâmara Witschge. “O que o jornalismo está se tornando,” *Parágrafo* 4, no. 2 (July/December, 2016): 8–20.  
<https://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/478>.
- Ettema, James, and Theodore Glasser. *Custodians of Conscience: Investigative Journalism and Public Virtue*. New York: Columbia University, 1998.
- \_\_\_\_\_. “On the epistemology of Investigative Journalism,” *Communication* 8 (August, 1984): 1–33. <https://eric.ed.gov/?id=ED247585>.
- Fidalgo, Joaquim. “A questão das fontes nos códigos deontológicos dos jornalistas,” *Comunicação e Sociedade 2 – Série Comunicação Cadernos do Noroeste* 14, no. 1–2 (October 1999): 319–337.  
[https://repositorium.sdum.uminho.pt/bitstream/1822/5513/1/CS\\_vol2\\_jfidalg\\_o\\_p319-337.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/5513/1/CS_vol2_jfidalg_o_p319-337.pdf).
- Flew, Terry, Christina Spurgeon, Anna Daniel, and Adam Swift. “The promise of computational Journalism,” *Journalism, Practice* 6, no. 2 (October 2011): 157–171.  
<https://www.tandfonline.com/doi/abs/10.1080/17512786.2011.616655>.
- Gillmor, Dan. *We the Media – grass roots journalism by the people, for the people*. Sebastopol: O’Reilly, 2006.
- Gray, Jonathan, Liliana Bounegru, and Lucy Chambers. *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News* Sebastopol, CA: O’Reilly, 2012.

- Guerra, Josenildo. "Ranking Q-Avalia da qualidade jornalística no Brasil e em Portugal 2018: uma avaliação experimental," *Estudos de Jornalismo e Mídia* 17, no.1 (June 2020): 54–74.  
<https://periodicos.ufsc.br/index.php/jornalismo/article/view/1984-6924.2020v17n1p54>.
- Hall, Stuart, Chas Critcher, Tony Jefferson, John Clarke, and Brian Roberts. "A produção social das notícias: o mugging nos media." In *Jornalismo: questões, teorias e "estórias"*, edited by Nelson Traquina, 309–344. Lisboa: Veja, 1999.
- Hayes, Arthur, Jane Singer, and Jerry Ceppos. "Shifting Roles, Enduring Values: The Credible Journalist in a Digital Age," *Journal of Mass Media Ethics* 22, no. 4 (December, 2007): 262–279.  
<https://www.tandfonline.com/doi/abs/10.1080/08900520701583545>
- Herman, Eduard S., and Noam Chomsky. *Manufacturing Consent: the political economy of mass media*. Nova York: Pantheon Books, 2002.
- Hermida, Alfred and Mary Linn Young. *Data Journalism and the Regeneration of the News*. Routledge, 2019.
- Howard, Alexander. *The art and Science of Data-Driven Journalism*. New York: Tow Center or Digital Journalism, 2014.
- Karlsson, Michael. "Rituals of Transparency. Evaluating Online News Outlet's Uses of Transparency Rituals in the United, United Kingdom and Sweden," *Journalism Studies* 11, no. 4 (July, 2010): 535–545.  
<https://www.tandfonline.com/doi/abs/10.1080/14616701003638400>.
- \_\_\_\_\_. "The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority," *Journalism* 2, no. 3 (April, 2011): 279–295. <https://journals.sagepub.com/doi/10.1177/1464884910388223>.
- \_\_\_\_\_, C. Clerwall, and L. Nord. "You ain't seen nothing yet: Transparency's (lack of) Effect on Source and Message Credibility." In *The Future of Journalism: In an age of digital media and economic uncertainty*, edited by Bob Franklin, 456–466. New York, Routledge, 2017.
- Kennedy, Helen, Rosemary Lucy Hill, Giorgia Aiello, and William Allen. "The work that visualisation conventions do. *Information*." *Communication & Society* 19, no. 6 (March, 2016): 715–735. DOI: 10.1080/1369118X.2016.1153126, 2016.
- Kovach, Bill, and Tom Rosenstiel. *Elementos do jornalismo: o que os jornalistas devem saber e o público exigir*. São Paulo: Geração Editorial, 2003).

- Lacy, Stephen, and Tom Rosenstiel. *Defining and Measuring Quality Journalism*. Rutgers School of Communication and Information (March, 2015): 1–66. <https://www.issuelab.org/resources/31212/31212.pdf>.
- Lewis, Seth, and Nikki Usher. “Open source and journalism: Toward new frameworks for imagining news innovation.” *Media, Culture & Society* 35, no. 5 (June 2013): 602–619. <https://journals.sagepub.com/doi/abs/10.1177/0163443713485494>.
- Loosen, Wiebke, Julius Reimer, and Fenja Schmidt. “When data becomes News: A content analysis of data journalism pieces.” In *Conference Proceedings of the Future of Journalism* (Cardiff, 2015).
- Manovich, Lev. *The Language of New Media*. Cambridge: MIT, 2001.
- Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data*. London: John Murray, 2013.
- McQuail, Denis. (1992). *Media performance: Mass communication and the public interest*. London: Sage, 1992.
- McQuail, Denis. *Journalism and society*. London: Sage, 2013.
- Meier, Klaus. “Transparency in Journalism: Credibility and trustworthiness in the digital future.” In *Conference Future of Journalism*, Cardiff, September 2009. <https://edoc.ku.de/id/eprint/13483/> <http://www.caerdydd.ac.uk/jomec/resources/foj2009>.
- Meikle, Graham. *Future Active – media activism and the internet*. Nova York: Routledge, 2003.
- Meyer, Philip. *Precision Journalism: A Reporter’s Introduction to Social Sciences Methods*. Maryland: Roman & Littlefield, 1973).
- Neuberger, Christopher. “Objektivität.” In *Handbuch Journalismus und Medien*, edited by Siegfried Weischenberg, Hans Kleinsteuber, and Bernhard Pörksen, 406–430. Konstanz: UKV, 2015.
- Noci, Javier Diaz, and Ramon Salaverría. “Hipertexto periodista: teorías y modelos.” In *Manual de redacción ciberperiodista*, edited by Javier Diaz Noci, and Ramon Salaverría, Barcelona, 81–137. Ariel Comunicación, 2003.
- Ojo, Adegboyega, and Bahareh Heravi. “Patterns in Award Winning Data Storytelling,” *Digital Journalism* 6, no. 6 (November, 2017): 693–718. <https://www.tandfonline.com/doi/abs/10.1080/21670811.2017.1403291?journalCode=rdij20>.
- Palácios, Marcos. “Ruptura, continuidade e potencialização no jornalismo on-line: o lugar da memória.” In *Modelos de jornalismo digital*, edited by Elias Machado, 13–36. Salvador: Calandra, 2003.

- \_\_\_\_\_. “Jornalismo online, informação e memória: apontamentos para debate.” *Revista PJ: BR Jornalismo Brasileiro*, São Paulo, no. 4, (April, 2004). [https://pjr.eca.usp.br/arquivos/artigos4\\_f.htm](https://pjr.eca.usp.br/arquivos/artigos4_f.htm).
- Parasie, Sylvain. “Data Journalism and the promise of transparency.” In *The Routledge companion to media and scandals*, edited by Tumber Howard, and Silvio Waisbord, 263–272. London: Routledge, 2019.
- \_\_\_\_\_, and Eric Dagiral. “Data-Driven Revelation? Epistemological tensions in investigative journalism in the age of big data,” *Digital Journalism*, Routledge 3, no. 3 (November, 2014): 364–380. <https://www.tandfonline.com/doi/abs/10.1080/21670811.2014.976408>.
- \_\_\_\_\_. Dagiral, Eric. “Data-driven Journalism and the Public Good: Computer-Assisted Reporters and Programmer-Journalists in Chicago,” *New Media and Society* 15, no. 6 (November, 2012): 853–871. <http://journals.sagepub.com/doi/abs/10.1177/1461444812463345>.
- Pinto, Manuel. “Fontes jornalísticas: contributos para o mapeamento do campo,” *Comunicação e Sociedade* 14, no. 1–2 (2000): 277–294. [https://repositorium.sdum.uminho.pt/bitstream/1822/5512/1/CS\\_vol2\\_mpinto\\_p277-294.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/5512/1/CS_vol2_mpinto_p277-294.pdf).
- Plaisance, Patrick L. “Transparency: an assessment of the Kantian roots of a key element in media ethics practice,” *Journal of Mass Media Ethics* 22, no. 2 (December, 2007): 187–207. <https://www.tandfonline.com/doi/abs/10.1080/08900520701315855>.
- Rogers, Simon. *Facts are Sacred*. London: Faber and Faber, 2013.
- Rosen, Jay. *What are journalists for?* New Haven and London: Yale University Press, 1999.
- Rosenstiel, Tom, Marion Just, Todd Belt, Atiba Pertilla, Walter Dean, and Dante Chinni. *We Interrupt this Newscast: How to improve local News and win ratings too*. New York: Cambridge University press, 2007.
- Rupar, Verica. “How Did You Find That Out? Transparency of the Newsgathering Process and the Meaning of News,” *Journalism Studies* 7, no. 1 (August, 2006): 127–143. <https://www.tandfonline.com/doi/abs/10.1080/14616700500450426>.
- Schudson, Michael. *Descobrimos as notícias: uma história social dos jornais nos Estados Unidos*. Petrópolis: Vozes, 2010.
- Singer, Jane. “The Political J-Blogger: ‘Normalizing’ a New Media Form to fit Old Norms and Practices,” *Journalism* 6, no. 2 (May, 2005): 173–198. <https://journals.sagepub.com/doi/10.1177/1464884905051009>.
- Stroobant, Joyce. “Finding the news and mapping the links: a case study of

hypertextuality in Dutch-language health news websites,” *Information, Communication & Society* 22, No. 14, (June 2018): 2138–2155.

<https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1477971>.

Tuchman, Gaye. “Objectivity as Strategic Ritual: An Examination of Newsmen’s Notion of Objectivity,” *American Journal of Sociology* 77, no. 4 (January, 1972): 660–679. <https://www.jstor.org/stable/2776752>.

Weber, Robert P. *Basic content analysis*. Newbury/Ca.: Sage, 1990.

Zelizer, Barbie, Pablo Boczkovski, and Christopher W. Anderson. *The Journalism Manifesto*. Cambridge: Polity Press, 2021.