SITUATING **OPEN DATA**

GLOBAL TRENDS IN LOCAL CONTEXTS

Edited by Danny Lämmerhirt, Ana Brandusescu, Natalia Domagala & Patrick Enaholo

Situating Open Data

Global Trends in Local Contexts

Edited by
Danny Lämmerhirt, Ana Brandusescu,
Natalia Domagala & Patrick Enaholo

AFRICAN
MINDS

Published in 2020 by African Minds
4 Eccleston Place, Somerset West 7130, Cape Town, South Africa
info@africanminds.org.za
www.africanminds.org.za

This work is published under a Creative Commons Attribution 4.0 International License (CC-BY).

ISBN Paper 978-1-928502-12-8
ISBN eBook 978-1-928502-13-5
ISBN ePub 978-1-928502-14-2

Orders:
African Minds
4 Eccleston Place, Somerset West 7130, Cape Town, South Africa
info@africanminds.org.za
www.africanminds.org.za

For orders from outside South Africa:
African Books Collective
PO Box 721, Oxford OX1 9EN, UK
orders@africanbookscollective.com
www.africanbookscollective.com

2.

Journalists and the intermediation of open data: A Nigerian perspective

Patrick Enaholo & Doyinsola Dina

Over the years the open data movement has reached a remarkable and ever-growing number of countries. Several groups and actors have embraced the campaign towards the liberation and democratisation of data in ways that would allow it to be interpreted and shared through networked platforms such as the internet (Davies 2012). Although the origins of the movement can be traced to specific political and social activities that began in the 1990s, most notably in the UK and the US, it was the initial attempts at formulating a definition of *openness* that catalysed a wider appreciation of the concept among more diverse segments of society (Heimstädt et al. 2014). Many of these definitions pointed beyond emphasising the need for a steady supply of data (mostly from governments) to their use, reuse and redistribution (by citizens) with little or no restrictions (Open Knowledge Foundation n.d.). According to the World Bank data toolkit, the two dimensions of data openness are legal and technical openness. To be legally open means data must be placed in the public domain or under liberal terms of use with minimal restrictions. On the other hand, data must be technically open, that is, it must be published in electronic formats that are machine readable and non-proprietary, so that anyone can access and use the data with common, freely available software tools. Data must also be publicly available and accessible on a public server, without password or firewall restrictions. Thus, the availability and accessibility of data served as a precondition for its openness along with its machine-readability – a quality that enables data to be re-ordered, re-shaped and re-transmitted using computer-based algorithms.

However, the fact that data is open does not necessarily imply that it is useful. The usefulness of data requires that some underlying meaning can be drawn

from it in order to generate new knowledge or enrich existing knowledge for the benefit of its end-user. While open data may be understood as content that can be freely used, modified and shared by anyone for any purpose (Open Knowledge Foundation n.d.), it is often generated and made accessible by a supplier and processed and packaged in the most appropriate format that an end-user can benefit from. Thus, within an open data ecosystem, ‘end-users’ are the final recipients in the spectrum of data transfer that begins with the data supplier. In the context of open government data (OGD), governments are the suppliers or providers of data and citizens are the end-users. Along this spectrum of data transfer, studies have pointed to the existence of stumbling blocks posed on the side of data providers as well as data users. Data provider-centric barriers include the reluctance to openness as a result of: a) the effort and cost required to convert closed into open data; b) the potentially high cost of facilitating the uptake of complex datasets; c) the poor quality of datasets; d) the absence of legal and policy frameworks; e) a lack of capacity to implement and sustain open data practices; and f) resistance by data custodians to opening up datasets. Data user-centric barriers include lack of access, low levels of data literacy, lack of human, social and financial capital to effectively use open data, and open up several datasets that can potentially create greater value when combined together (Janssen & Zuidervijk 2014; Magalhães et al. 2013; Gurstein 2011). González-Zapata and Heeks (2015) summarise the main barriers of OGD into four categories: data absence, lack of data provision, lack of data quality and the digital divide. The last item, according to the authors, is a ‘recurrent obstacle in developing countries’ (2015: 4).

These challenges in the transfer of open data have warranted the need for a group of actors who perform crucial functions of sustaining the data supply chain through advocacy, interpretation and communication of open datasets for easy digestibility. These actors are referred to as open data intermediaries because of their role as linkages between open data suppliers and users (Van Schalkwyk et al. 2016). Open data intermediaries work with the data suppliers and/or the end-users to overcome the barriers associated with the provision or utilisation of open data (González-Zapata & Heeks 2015).

One of the early studies of the role of intermediaries in open data ecosystems was carried out by Magalhães et al. (2013) who identified three terms commonly adopted in academic literature and communities of practice: civic start-ups, open data services and infomediaries. With developments in the study and practice of open government data ecosystems and intermediaries, these overlapping categories have been overtaken by a wider set of groups with boundaries that are more clear-cut and distinct. For example, Johnson and Green (2017) use the term ‘infomediaries’ more broadly and synonymously with ‘intermediaries’ to encompass those that make open [government] data more accessible and useful for end-users. Among the categories of infomediaries in their classification are government agencies, the private sector, non-governmental organisations and

community groups (including civil society organisations (CSOs)), academics and researchers and the media.

Scholars have also highlighted the complementary roles played by open data intermediaries. González-Zapata and Heeks (2015) classify them into five different types of agents on the basis of their interaction with data. Thus, intermediaries can act as demanders, producers, validators, developers and communicators of open government data. While the first three roles are directly related to the production of data, the latter two are linked to the manipulation of data to create information. They also note that, in practice, these roles are not mutually exclusive and intermediaries may perform multiple tasks at the same time. In line with this argument, Enaholo (2017) highlighted research on the role of CSOs in the OGD ecosystem: either as advocates for the provision of government data (demanders) or as grassroots advocates for citizen participation in the use of data (communicators) or both.

To perform any of these roles, open data intermediaries must possess the skills and competencies required to transform data into formats that can be beneficial to end-users. On the basis of their varying competencies within the open data ecosystem, Van Schalkwyk et al. (2016) apply Bourdieu's field theory nomenclature of 'capital'. The authors distinguish the financial, technical, social, cultural and symbolic capital as part of the repertoire that enable open data intermediaries to perform their roles within the ecosystem. The authors further argue that while one advocacy group may possess the symbolic or cultural capital required to apply pressure on government to release open data (as demanders), they may lack the technical or social capital required to convert the data into information that can be effectively communicated to end-users. Inversely, intermediaries with technical or social capital may lack the requisite symbolic or cultural capital to facilitate the supply of data from the government. Crucially, Van Schalkwyk et al. (2016: 15) note that intermediaries do not typically operate as single agents but rely on the capitals possessed by other intermediaries to 'unlock the full value of the transaction between the provider and the user in each of the fields in play'.

This complementarity of roles can be demonstrated using the list of intermediaries from Johnson and Greene (2017) highlighted above. Government agencies can act as demanders of data since their proximity to the originating source of data invests them with some symbolic capital. As highlighted by Van Schalkwyk et al. (2016), this role can also be played by civil society organisations with close ties to the government (perhaps through the agencies) as a result of their social capital. The media, academics and researchers may possess the required technical capital to interpret the data as information to be effectively communicated to end-users by other intermediaries with creative or social capital. The latter can include the media, non-governmental or civil society groups.

Our reason for specifying these roles along with their concomitant competencies (or capital) is to draw attention to the fact that those competencies often emerge from the skills and expertise usually associated with the professional

groups to which open data intermediaries belong. When those skills are lacking, open data intermediaries may try to obtain them by improving those knowledge areas within their profession or acquiring new expertise from elsewhere. For example, Enaholo (2017) discusses how civil society workers in Nigeria are ‘metamorphosing into a community of open data enthusiasts, perhaps in the hope that, through open data, the effectiveness of their role as advocates for good governance would be enhanced’. This transformation, according to him, ‘is driven by the proliferation of CSOs with the skills and knowledge of web-based open data systems and tools’ (Enaholo 2017: 95).

In this study, we have taken a different approach by examining the work of open data intermediation by media professionals, specifically, journalists. This approach is based on the need to recognise the distinctions that exist among the roles played by intermediaries based on their professional biases and competencies. To this end, we draw attention to the role of journalists and demonstrate how their work of open data intermediation involves competencies that appear to tilt towards end-users. Thus, in the terminology used by González-Zapata and Heeks (2015), journalists can be recognised primarily as ‘communicators’ of information (generated from data) but also, to some extent, as developers. While the former necessitates the possession of creative competencies, the latter focuses more on technical ones. From this perspective, we examine the intricacies of this intermediation – the processes and barriers – and how it contributes to or hinders their end-users’ ability to assimilate open data.

In their study of the context of open data in Nigeria, Mejabi et al. (2014) examine the motivations, capacities and sustainability of intermediaries’ role. According to the authors, CSOs analyse and re-present budget information while media professionals interpret the budget data to drive civic engagement and participation. In our research, we extend the findings of Mejabi et al. (2014) by: first, examining in some detail the processes by which journalists (as media professionals) carry out this role; second, unearthing those factors that limit and undermine their ability to do so; and, third, determining how these factors affect the broad spectrum of intermediation within the open data ecosystem. In doing this, we also contribute to the body of literature on the nature and diversity of roles and capacities of open data intermediaries in developing contexts.

Conceptual framework

Journalism and the ‘quantitative’ turn

The journalism profession is one that is characterised by the activity of reporting factual events in society to an audience. Journalists engage in gathering, assessing, creating and presenting news and information for public consumption. They disseminate the narratives generated from this process through a variety of media channels, most notably print and online newspapers, but also radio

and television. As members of a profession, journalists differ from bloggers and those people who employ the press tools they now have in their possession to disseminate information (known as citizen journalists) by the fact that they hold themselves to a higher standard. Besides gathering and disseminating information, journalists try to assess the accuracy of the information, take care to avoid inadvertent error, identify sources (and their motives) whenever feasible, and ensure that promotional materials, and photos, video, audio, graphics, sound bites, and quotations do not misrepresent the information (Davis 2010).

Coddington (2015) notes that although journalism has historically been built around textual and visual presentation of reportage, the use of numbers was traditionally less prominent until the emergence of computer-assisted reporting (CAR) in the early 1970s – the period which marked, as Petre (2013) suggests, the advent of journalism's 'quantitative' turn. Computer-assisted reporting is considered the first approach to the systematic use of computers for data gathering and statistical analysis in journalism (Gray et al. 2012) and was a direct descendant of precision journalism, a term that was famously formulated by Philip Meyer. Since then, computer-assisted reporting has brought about an assortment of computer-based journalism practices including programmer-journalism, open-source journalism, data journalism, computational journalism, among others (Coddington 2015).

Among these variants, data (or data-driven) journalism has been identified as the most prominent successor of CAR. It was born out of the apparent glut in the availability of data in recent times (Borges-Rey 2016) and has emerged as a more rounded form of journalism than its predecessor. Some have argued that data journalism should be seen as the result of continuity in the profession's foray into the realm of computers rather than an outright change in approach. However, the development of CAR was in a context of information paucity that prevented journalists from finding answers to pressing questions and contributed to the need to devote time to gathering and analysing data using computer-based methods (Gray et al. 2012). Computer-assisted reporting was therefore a technique for augmenting reportage rather than a core part of journalism.

On the other hand, besides using data as a means to enhance stories, data journalism places data within the journalistic workflow. Data is at the heart of the data journalist's task of transmitting news and information to the public. It presents a broader understanding of news reporting: from a narrow conception of 'news events' coverage to the broader notion of 'situational reporting' in such a way that journalists move 'beyond the reporting of specific and isolated events to providing a context which gives them meaning' (Gray et al. 2012: n.p). Indeed, as Schrock and Shaffer (2017: 4) argue, 'data can be approached as a medium because it is inscribed with meanings, transmitted, decoded, and interpreted through specific practices.' Data journalism may therefore be described as a subset of processes involved in using data as a tool in the task of uncovering the hitherto obscure meanings behind events and phenomena that occur in society.

The practice of data journalism

Essentially, the work of the data journalist is to explore datasets to unearth new stories or to support existing ones. Since these stories are often concealed in large datasets, journalists require a specific set of specialised skills in addition to the domain-specific expertise to fully understand the data and its potential impact (Rogers et al. 2017). Data journalism is therefore perceived as a technique to extract relevant information from such datasets with the aid of statistical, visualisation and interactive methods for analysing, clustering and presenting data (Aitamurto et al. 2011). For proponents of the continuity school of thought, data journalism also inherits the expertise of data mining and collection from computer-assisted reporting.

With increased availability and better tools, data journalism is driven by the need to explain medium to large datasets. In developing stories and reports, the data journalist becomes ‘the gatekeeper to reducing complexity and providing meaning to the data’ (Rogers et al. 2017). Accordingly:

[data journalism] represents a new role for journalists as a bridge and guide between those in power who have the data and the public who desperately want to understand the data and access it but need help [in doing so]. (Rogers et al. 2017)

For this, data journalists must augment their journalistic skills with those that would enable them to work effectively with data for the benefit of their audiences.

In general agreement with other schools of thought, we identify three categories of skills required by data journalists: 1) the treatment of data as a source to be gathered, verified and validated, 2) the application of statistics to interrogate it, and 3) visualisations to present it. In addition to these can be added the skill of reporting which is already inherent in journalism practice. For this study, we adopt these categories but we break them into five individual competencies for the data journalist: a) data collection (or mining); b) data validation (or cleaning); c) data analysis; d) data visualisation; e) data reporting. These competencies do not cover all the steps taken by journalists in carrying out their work. Rather, we focus specifically on those ones that are directly related to journalists’ manipulation of data in drawing out news stories. For example, some scholars (such as Bradshaw 2014) have suggested data verification as an important skill or process in journalism that supports the ethical standards that are key characteristics of the profession. This comprises the protection of data sources, the effective handling of data leaks and privacy. However, since these processes do not necessarily involve working with data, we do not include them among our list of competencies. We discuss each of these in turn.

Data mining

Even with the abundance of data in contemporary society, data journalists often need to search for what is relevant and accessible to them. This requires the ability to explore a variety of options such as freedom of information (FOI) requests, internet searches, or crowdsourcing techniques. Each of these processes demands a sufficient level of mastery to yield the desired outcome. For example, obtaining data from the internet requires a knowledge of modern search techniques, the ability to use web-scraping tools or to write simple code that can pull data from the internet. In addition, a basic understanding of the prevailing laws will be needed to take advantage of FOI requests.

Data validation

This is also referred to as data cleaning. At their raw state, datasets typically require cleaning and validation in order to be useful for any form of analysis. When data journalists acquire data from various sources, they often need to combine them using software such as spreadsheets or databases. For this, they need to be familiar with any of the available tools that perform this function. Occasionally, they also require a basic knowledge of statistical procedures in order to prepare the data for more robust analysis. This step is important because it helps guarantee the accuracy of findings from the data.

Data analysis

Analysing datasets is central to the data journalism process (Aitamurto et al. 2011). According to Doig (2012), it is akin to interviewing a live source: 'you ask questions of the data and get it to reveal the answers.' Data analysis, also referred to as data interpretation (Knight 2015), is the differentiating factor between CAR and data journalism, with the latter laying more emphasis on a more inductive and exploratory approach that does not necessarily depend on how the data was collected (Coddington 2015). To carry out data analysis, journalists need a basic understanding of data structures as well as working knowledge of numerical and statistical principles.

Data visualisation

The practice of data journalism places emphasis on graphics and visualisation in the presentation of news and information (Knight 2015; Rogers 2011). The importance given to visualisations is essentially bound up with the data journalist's attentiveness to the audience (Coddington 2015). Visualisations are graphical representations of stories and reports that make it easier to consume news stories and narratives transmitted from data. Effective visualisations are the result of graphic design skills along with the ability to display quantitative information using visual perception and cognitive principles (Tufte 1983). Data journalists make use of a variety of tools and techniques to create visualisations from spreadsheets to graphics editing tools to programming languages.

Data storytelling

This is the final stage of the data journalism process in which journalists use storytelling techniques to transmit findings from the data to their audiences. The writing of news stories is often within the standard skillset of journalists. However, data stories (i.e. stories within the data journalism context) are typically amalgamations of stories in textual form with visualisations in graphics form. Oftentimes, these stories include the raw datasets published for readers to analyse and explore for themselves (Gray et al. 2012).

It is important to note that the order of the stages presented here does not always match what obtains in reality. Indeed, there has been a debate in the practice and research of data journalism about the preferred workflow for journalists (Rogers et al. 2017; Uskali & Kuutti 2015). Aitamurto et al. (2011) note the existence of two primary workflows. In the first, journalists start working with data only after the discovery of a story idea. This is similar to CAR whereby the data simply serve only to complement the story. In the second workflow, data serves as the starting point for the whole story. This matches the order of stages presented above. There is a general agreement in the literature that the former workflow is more common among journalists, although Uskali and Kuutti (2015) opine that the latter will be the future of data journalism.

Data journalism as open data intermediation

From the foregoing, it may be inferred that when journalists operate within the confines of an open data ecosystem, they are more likely to do so as data journalists. Rogers (2011) refers to this practice as ‘open data journalism’, the difference being that, in this instance, journalists work with data that is open and democratised and has the characteristics described earlier, namely to be available, accessible and machine-readable. Since the main interest of this study is open government data, we focus on data journalists who assume the role of open data intermediaries because the data with which they carry out the processes outlined above flow from the government as primary suppliers to citizens as end-users. Our concern then is to understand how journalists – as data journalists – fulfil this role within the ecosystem.

To this end, we formulate a conceptual framework that combines theories around the competencies of open data intermediaries (Van Schalkwyk et al. 2016) with the five stages of the data journalism process highlighted in the previous section. Thus, just as Van Schalkwyk et al. (2016) suggest that multiple intermediaries in an open data ecosystem possess different forms of capital (or competencies), we show how the work of data journalism involves two main competencies – technical and creative – which stem from the five stages of their expected workflow. This is illustrated in Figure 1.

Figure 1. The data journalism workflow within an open data supply chain

In the diagram, it can be seen that technical competencies comprise data collection, validation and analysis. These are the parts of the workflow that require more data-orientated skills from journalists partly because they involve the manipulation of raw data. By technical competencies, we refer specifically to the ability to handle data in a way that is initially removed from the information that they contain. These competencies are not exclusive to journalism. In fact, they can be said to belong primarily to disciplines with roots in statistical and quantitative analysis. However, the availability of tools and software have made data processing skills more accessible to journalists in the task of investigating and unearthing factual stories. With the technical skills required for data manipulation, journalists will be equipped to effectively interrogate datasets and use the answers derived to write reports.

On the other hand, the creative competencies are more attuned to the existing skillset of journalists who deal more with the art of relaying information in a clear and comprehensible way. According to Figure 1, these competencies include data visualisation and storytelling. As with technical competencies, our focus is not the content of the stories but their form. Creative competencies refer to the skills associated with constructing the narrative, including the language, the style and the writing techniques that contribute to the overall readability that makes it easier for audiences to understand what is being conveyed.

Methodology

In the sections above, we have drawn on a variety of sources to identify data journalism as the particular genre of journalism practice that pertains to the role of open data intermediary within the open data value chain. We have also outlined the processes (or stages) involved in the work of data journalists and shown how these can be embedded within the broader context of an open data ecosystem. Since there was a need to examine how this framework exists in practice, we engaged with data journalists in order to find answers to our research questions. Our aim was to identify those factors that enhance, limit and undermine journalists' ability to intermediate open data and how such factors affect the broad spectrum of intermediation within the open data ecosystem. The study is therefore data-centric as it focuses on the effectiveness of journalists' application of data as a tool to support news stories. For this reason, we do not examine the content of the news stories themselves since this would lead us away from the primary focus of the chapter.

Our study comprised three research methods: an online survey, content analysis and focus group discussions. We took the quantitative routes for the first two methods and qualitative for the last method. In total, 127 journalists from Nigerian mainstream media as well as freelancers who write for media houses were selected. Out of these, 94 journalists responded to the online survey using Google forms and 33 journalists joined the focus group discussions. In addition, 20 data stories that were shared by survey respondents were selected for content analysis by the authors.

Survey

The first research instrument used for the collection of responses was an online questionnaire prepared and distributed by email to the respondents. Purposeful sampling technique was adopted. The questionnaire was broadly targeted at journalists and sent specifically through data journalist networks. To further restrict the respondents to those who were more likely to be practising data journalism, we sent the survey as part of a recruitment process for a data journalism workshop. The survey was targeted at journalists based in a Nigerian newsroom or media outlet and this was verified by asking for the name of their media organisation along with links to their previous news stories if available.

The overall aim of the survey was to understand the application of data journalism among those who practice it. For this, we sought to find out what data journalists understood by the concept of 'data journalism' in order to determine whether it corresponded to the objectives of open data intermediation (i.e. to bridge the gap between open data suppliers and end-users). We also wanted to ascertain their level of awareness and practice of the processes involved in data journalism and to know the extent to which they applied those processes

when working directly with open government data (i.e. publicly available data disclosed by the government). In addition, the survey included questions about the common workflows that journalists adopted when working with data. This was to enable us to deduce how their preferred workflows affected their role as open data intermediaries.

Content analysis

After the online survey, we analysed the content of 20 data stories which were shared with us by survey respondents. The stories were selected from those that were provided by the respondents and were shortlisted using the following criteria: first, they contained some reference to numeric data; second, they were published in recognised media platforms; and third, they were provided by a journalist who demonstrated a clear understanding of data journalism and had practised it to a reasonable degree.

The choice of this method was informed by the need to verify the claims made by the respondents about their knowledge and practice of data journalism. Our assumption was that, if they did understand and apply the processes of data journalism in their work, it would be evident in their output (i.e. their data stories) – except if they encountered objective challenges and limitations in performing the various tasks involved in data journalism. Thus, we reckoned that the inability to work effectively as data journalists would negatively reflect on their role as open data intermediaries. If this was indeed the case, we wanted to see its manifestation. And to understand why this would be the case, we resorted to a focus group discussion.

Focus group discussion

A total of 33 journalists, most of whom have been working with data for about two to five years were selected for the focus group discussions (FGD). The method specifically helped to provide answers to our research questions dealing with those barriers and limitations that hindered journalists from using data effectively in their role as open data intermediaries. Our interpretation of effectiveness here refers to journalists' ability to undertake the processes required to convert raw data into information that can easily be digested by their audience (i.e. the end-users of the open data spectrum). For this, we specifically asked journalists about their possession of the skills needed at each stage of their work with data. This was in a bid to know whether they actually possessed the required competencies to carry out the various tasks involved and, if not, the challenges they encountered in acquiring or putting them into practice.

The FGD also served as a mechanism for triangulation, whereby the responses from the discussants were used to confirm or better understand answers they gave in the survey. This was important because, as we discuss below, some of the

survey responses did not correspond with those we received during the FGD. For example, respondents' choice of data journalism workflows 'in theory' did not match what they carried out 'in practice'.

In our view, the methodological triangulation we adopted by using three different methods increased the credibility of the result and ensured a deeper understanding of the phenomenon we sought to investigate, more than would have been achieved with a single one.

Findings

From the survey responses we received on the concept of data journalism, we found that most of the respondents understood what it stood for and its relationship with the journalism profession. Many of the answers demonstrated a clear grasp of the notion of data journalism and, in their own words, described the relevance of their data stories to the audience. For example, one of them stated:

My simple concept of data-driven journalism is journalism practice based on Data Analytics. That is, a system where a journalist engages the analysis and filtering of large datasets for the purpose of creating a news story. In doing this, he or she examines datasets in order to draw conclusions about the information the data contains, so as to present a credible news story to his/her readers or audience as the case may be. However, let me add that the importance of 'data-driven news story' cannot be overemphasised as it helps the general public or specific groups or individuals to understand patterns and make decisions based on the findings of the given journalists.

For the purpose of our examination of data journalism through the lens of open data intermediation, this response, like some others given in the survey, is useful because it specifically points to the end-user as the beneficiary of the journalist's attempt to 'draw conclusions about the information [contained in the data]' and presented in the form of 'a credible news story.' This clearly aligns with the role of open data intermediaries in transforming open data into information and packaging it into a digestible form (i.e. the news story) that end-users can benefit from.

The only drawback of this particular answer (and many others) was the lack of reference to the data provider. However, a few did include the data source in their definition and specifically referred to government data.

My understanding of data driven journalism is [that it is] the new aspect of journalism where journalists [look] through figures around government or public funded projects to find the story the figures tell. This is about statistics, numbers help journalists to explain the narrative around a project, budget, sub-sectors, etc.

However, a different question in the survey gave the opportunity for the respondents to select their usual sources of data for writing stories (such as the World Bank, International Monetary Fund (IMF), Nigeria Stock Exchange (NSE), Central Bank of Nigeria and so on). Interestingly, the most common data provider was a government agency, the Nigerian Bureau of Statistics (NBS), which is known to publish datasets from different sectors of the country. Perhaps the omission of the data source in their definitions of data journalism implies that, since government data is the default provider of data, it can be taken for granted.

Beyond their understanding of data journalism as a concept, the survey sought insight into respondents' practice of it. Thus, it included a question about the processes of data journalism in which participants were most involved. Among the 5 skills listed in our framework, the most frequently selected was data storytelling. This was expected since, as mentioned previously, storytelling falls within the regular skillset of journalists as their primary occupation. Other skills included data analysis, data collection and data validation – in this respective order. The final skillset, data visualisation, was the least selected by the journalists. Even though we have classified visualisation under the creative competencies, it requires graphic design skills which can be technical and unfamiliar to journalists. As we discovered afterwards, journalists are often at a loss when trying to choose the most appropriate graphical representation (e.g. charts, infographics) to enhance the textual content of their news stories.

The lack of knowledge and use of data visualisation were apparent in the data stories analysed. A textual analysis of the stories revealed that, in most of the stories, numbers are discussed within the text, but not represented visually. This made it cumbersome to read them. For some of these, the numbers were represented as visualisations (often in the form of charts or infographics), thus repeating the same ideas conveyed in the text. Sometimes, these visualisations were not legible because they were too small thus rendering them useless since they added little to the reader's comprehension of the story. For some others, there was no clear correlation between the story itself and the visualisation, which had the overall impact of confusing, and perhaps misleading, the reader. An example is shown in Figure 2.

Only one out of all the stories analysed did not have any of the problems outlined above. Rather, it included charts that were legible and that complemented the story even though the information conveyed was sufficient for readers.

The fact that most of the data stories were difficult to read or comprehend goes beyond the poor quality of their visualisations. Gray et al. (2012: 191) argue that, 'unlike other visual media – such as still photography and video – data visualisation is deeply rooted in measurable facts'. This is because it is based less on aesthetics than on the data that it is designed to transmit. Since visualisations are information presented in graphical format using pre-attentive attributes that encode the underlying data (Few 2004), they depend on other

aspects of the work with data such as analysis, validation and, to some degree, data collection and mining. Therefore, creating effective visualisations often requires some knowledge of the other three.

Figure 2. News stories with supporting data visualisations

Source: <https://nationaldailyng.com/efcc-fails-to-nail-single-politician-among-353-convictions-secured-in-10-years/>

Source: <https://www.dailytrust.com.ng/more-foreign-visitors-troop-into-nigeria-data.html>

In the focus group discussions, we interrogated the reasons behind the poor quality of journalist's data stories and discovered that, contrary to claims about how they worked with data, the journalists lacked the expertise to actually manipulate the data using modern tools and software. We found that most of the knowledge they had was simply theoretical but often not practised. We also discovered that many of the journalists did not independently create visualisations to better their stories. Rather, they relied on charts, graphs or infographics created by others within their organisation or from external sources such as civic tech organisations. One of such organisations that featured frequently in our research was Code for Nigeria. Code for Nigeria is a civic tech group that uses data to give citizens hyper-local and hyper-personal information to make better informed decisions about everyday issues. They offer tools that give infomediaries like journalists and civic activists an easy 'plug and play' toolkit for finding and embedding interactive data visualisations into their storytelling. Indeed, some of the data stories analysed included visualisations created by Code for Nigeria.

Our research is not without limitations. For example, our sample size could have been larger in order to accommodate more journalists. We also restricted our focus groups to just one. Clearly, the study would have benefitted from a more ambitious recruitment of respondents even though we were limited by insufficient funding and logistical challenges – since journalists are not the

easiest set of individuals to assemble in one place. In addition, all the data stories analysed were online. The inclusion of news reports published offline may have provided more nuance to the study. Finally, to increase the robustness of our findings, we could have included other actors in the open data value chain such as CSOs and end-users.

Discussion

Despite the findings and limitations, the fact that data journalists can and do operate as open data intermediaries cannot be denied. Therefore, how effectively do data journalists play this role? Based on our study, we hold that, in the Nigerian context, the contribution of data journalists is currently limited since they do not appear to possess the key competencies they need to communicate more effectively through their data stories. As shown above, our sample size of data journalists lacked the technical competences needed to act constructively as intermediaries despite their theoretical knowledge of their field.

While the creative competency of storytelling already forms part of their existing repertoire of skills (with varying degrees of proficiency), we found that journalists' lack of other skills required in their work with data – that is, the technical competencies of data collection, validation and analysis – obstructs their ability to work with open data. Although there exists a growing set of online tools that make it easier for journalists to use data for their stories, we argue that, based on our findings, these may be insufficient without the fundamental skills needed for data manipulation. While their lack of technical skills may be related to the absence of subjects like statistical analysis during their basic journalistic training, it can also be attributed to their lack of interest in moving beyond press releases and flashy headlines as well as structural problems in their media organisations or in society that lead to a lack of motivation or editorial mandate.

Nevertheless, based on the interactions with the journalists in our study, there was a general acknowledgement that their technical competencies were lacking or needed to be improved. Apart from the opportunities for skills acquisition which they frequently try to take advantage of (such as the workshop which was the bait for our research), journalists rely on the technical competencies of other actors in the wider ecosystem to complement their own creative competencies. These other actors can be individuals from within or outside their media organisation such as civic tech groups.

This kind of intervention supports research findings from authors such as Van Schalkwyk et al. (2016) and Enaholo (2017), among others, about the complementary roles played by multiple intermediaries in the open data ecosystem. In this particular scenario, the storytelling skills of data journalists (a key component of their creative competences) is combined with assets (in the form of data visualisations) provided by other actors like Code for Nigeria to compensate for the lack of technical competencies that journalists need in

their work of intermediation. However, in spite of its benefits, our findings suggest that this collaboration is potentially fraught with shortcomings because its results were not optimal. This assessment is based on our observation of the output from the data journalists we researched. The quality of the written article did not often match that of the visualisations that were designed to complement them. For example, one of the stories, a well-written report by a journalist about the shortage of women in Nigeria's political space, included charts prepared by Code for Nigeria that were illegible and unclear, and that failed to effectively play the role of supporting the main article.

Although visualisations are not the only means of displaying data in news stories, they are most effective in concealing the difficulty usually associated with trying to comprehend the meaning behind a large amount of numbers in a text-based medium. For this reason, along with words, charts and visualisations complete the ensemble of a genuine data-driven story. We argue, therefore, that the poor use of data can be manifested by the low quality of data visualisations in terms of meaning, accuracy and complementarity with the main article. In our view, such cases of poor data visualisations are the outcome of a lack of or insufficient coherence among stages of the data journalism process. We opine that when adjoining stages of the process are disconnected, there is a negative impact on the data story and, hence, on the effective intermediation of the journalist in the open data ecosystem. Thus, when the technical competencies (i.e. the use of data) do not seamlessly support the creative ones (the language, style and flow of the article) in the data journalism process, there is a negative impact on the effectiveness of open data intermediation.

To deal with this, one solution is the upskilling of journalists in order to improve their technical competencies (Rogers et al. 2017). Rather than taking the technically orientated tasks away from journalists, a panacea to the challenge of ineffective intermediation among data journalists would be to upgrade their competencies in their area of deficiency. The need for training and developing data journalism skills has been one of the main challenges facing journalists in different parts of the world (see Appelgren & Nygren 2014 for Sweden and Hannaford 2015 for the UK). This research suggests that the same can be said for the Nigerian context. In our view, improving the technical competencies of journalists who work with data may have the effect of improving the quality of their output as data journalists to the benefit of end-users.

Another recommendation is to diversify the competencies of journalists within newsrooms in order to be able to form data reporting teams made up of differently skilled individuals. In contrast to the introduction of data experts as external actors within media organisations, this strategy involves a closer integration and complementarity of skills and would ensure that data-driven news stories are no longer produced by individuals but by teams of journalists whereby each one contributes on the basis of their core competencies. This approach to data journalism would require a re-configuration of newsrooms, one which would

comprise a variety of journalists with different sets of competencies – technical, creative and so on. The existence of data teams in newsrooms is not novel. However, in the Nigerian context, based on our findings, it is a strategy that needs to be adopted more in order to improve the intermediary potential of data journalists.

Our findings suggest that journalists play a vital role as intermediaries through the interpretation of open data in formats that are more digestible for end-users. Their intervention lies at the intersection between the open data and the principles of data journalism – a burgeoning practice in the journalism space. The combination of both has shaped the process involved in the performance of intermediation by journalists. We describe this process as a five-stage process, namely, open data *collection, validation, analysis, visualisation and storytelling*. As we discussed, these steps are not novel and have been applied in data-driven reporting among journalists for quite a while (Coddington 2015; Gary et al. 2012; Rogers et al. 2017). However, we also show that they introduce a relatively distinct methodology into the work of open data interpretation which can (and indeed has) been adopted by other actors in the ecosystem. While the first three can be categorised under technical competencies or capital (Van Schalkwyk et al. 2016), the final two (especially the last – storytelling) may be classified as ‘creative competencies’ which belong firmly within the domain of journalism.

These competencies require specific skillsets which need to be possessed in order to advance the work of open data intermediation. The creative competencies refer to those skillsets that pertain to journalists’ ability to write stories and, to some extent, visualise data. While this set of competencies was common to journalists in our research, we found that the other set of skills – the technical competencies of data collection, validation and analysis – was deficient. We also found that, to overcome this challenge, journalists rely on the technical competencies of other actors in the ecosystem and complement those with their own creative competencies. However, this solution proved imperfect because it resulted in inadequate and defective information for end-users. In other words, the quality of open data intermediation is lowered. This is because the five-stage process that is specific to data journalism forms a coherent unity in which one stage feeds easily into the next. When there is insufficient coherence between two adjoining steps, the process becomes less effective.

The solutions we propose to absence of technical competencies include, on the one hand, the skilling and upskilling of journalists in order to improve their technical competencies and, on the other hand, a reconfiguration of journalist newsrooms to accommodate a wider variety of competencies within data reporting teams. From the perspective of the overall open data ecosystem, the process of skilling can also be directed at other intermediaries such as CSOs through the improvement of their own creative competencies. However, even if the latter is achieved, the desired outcome could be adversely affected by the fact that CSOs would lack the institutional paraphernalia associated with the journalism profession.

REFERENCES

- Aitamurto T, Sirkkunen E & Lehtonen P (2011) Trends in data journalism. *Hyperlocal*: 1–27. http://virtual.vtt.fi/virtual/.../D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf
- Appelgren E & Nygren G (2014) Data journalism in Sweden: Introducing new methods and genres of journalism into “old” organizations. *Digital Journalism* 2(3): 394–405
- Baack S (2011) A new style of news reporting: Wikileaks and data-driven journalism. *Cyborg Subjects*: 1–10. <https://www.ssoar.info/ssoar/handle/document/40025>
- Borges-Rey E (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice* 10(7): 833–843. <https://doi.org/10.1080/17512786.2016.1159921>
- Bradshaw P (2014) Data journalism. In: L Zion & D Craig (eds) *Ethics for digital journalists: Emerging best practices*. Abingdon-on-Thames: Routledge. pp. 202–219
- Coddington M (2014) Clarifying journalism’s quantitative turn. *Digital Journalism* 3(3): 331–348. <https://doi.org/10.1080/21670811.2014.976400>
- Davies T (2012) How might open data contribute to good governance? *Commonwealth Governance Handbook* 2012/13: 148–150
- Davis M (2010) Why journalism is a profession. In: C Meyers (ed.) *Journalism ethics: A philosophical approach*. USA: Oxford University Press. pp. 91–102
- Dal Zotto C, Schnker Y & Lugmayr A (2015) Data journalism in news media firms – The role of information technology to master challenges and embrace opportunities of data-driven journalism projects. Paper presented at the Twenty-Third European Conference on Information Systems (ECIS), Münster, Germany
- De Maeyer J, Libert M, Domingo D, Heinderyckx F & Le Cam F (2015) Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432–446. <https://doi.org/10.1080/21670811.2014.976415>
- Doig S (2012, n.d) Basic steps in working with data. In: *Data Journalism Handbook* 1. <https://datajournalism.com/read/handbook/one/understanding-data/basic-steps-in-working-with-data>
- Enaholo P (2017) Beyond mere advocacy: CSOs and the role of intermediaries in Nigeria’s open data ecosystem. In: F van Schalkwyk, SG Verhulst, G Magalhães, J Pane & J Walker (eds) *The Social Dynamics of Open Data*. Cape Town: African Minds. pp. 89–108
- Few S (2004) Eenie, meenie, minie, moe: Selecting the right graph for your message. *Intelligent Enterprise* 7: 14–35
- Gill M, Corbett J & Sieber R (2017) Exploring open data perspectives from government providers in western Canada. *Journal of the Urban & Regional Information Systems Association* 28(1): 19–29
- González-Zapata F & Heeks R (2015) Understanding multiple roles of intermediaries in open government data. Paper presented at the 13th International Conference on Social Implications of Computers in Developing Countries. Negombo, Sri Lanka, May 2015
- Gray J, Bounegru L & Chambers L (2012) *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Sebastopol: O’Reilly Media
- Gurstein MB (2011) Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16(2). <https://doi.org/10.5210/fm.v16i2.3316>
- Hannaford L (2015) Computational journalism in the UK newsroom. *Journalism Education* 4(1): 1–21
- Heimstädt M, Saunderson F & Heath T (2014a) Conceptualizing open data ecosystems: A timeline analysis of open data development in the UK. Paper presented at the

- Conference for E-Democracy and Open Government (CeDEM2014), Krems, Austria, 21 May
- Heimstädt M, Saunderson F & Heath T (2014b) From toddler to teen: Growth of an open data ecosystem. *JeDEM-eJournal of eDemocracy and Open Government* 6(2): 123-135
- Janssen M & Zuiderwijk A (2014) Infomediary business models for connecting open data providers and users. *Social Science Computer Review* 32(5): 694-711
- Johnson PA & Greene S (2017) Who are government open data infomediaries? A preliminary scan and classification of open data users and products. *Journal of the Urban & Regional Information Systems Association* 28(1): 9-18
- Knight M (2015) Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice* 16(1): 55-72
- Magalhães G, Roseira C & Strover S (2013) Open government data intermediaries: A terminology framework. Paper presented at ICEGOV '13: Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance, Seoul, Korea, 22 October. pp. 330-333
- Mejabi OV, Azeez AA, Adedoyin A & Oloyede MO (2014) *Case study report on investigation of the use of the online national budget of Nigeria*. Open Data Research Network. <http://www.opendataresearch.org/sites/default/files/publications/Investigation%20of%20Open%20Budget%20Data%20in%20Nigeriaprint.pdf>
- Metta S, Messina A & Italiana RR (2017) An end-to-end approach for delivering data-driven stories. Paper presented at the International Broadcasting Convention (IBC2017), Amsterdam, Netherlands, 14 September
- Petre C (2013, 30 October) A quantitative turn in journalism? Tow Center for Digital Journalism. <http://blog.chartbeat.com/2013/10/31/quantitative-turn-journalism/>
- Rogers S (2011, 28 July) Data journalism at the Guardian: What is it and how do we do it? *The Guardian*. <https://www.theguardian.com/news/datablog/2011/jul/28/data-journalism>
- Rogers S, Schwabish J & Bowers D (2017) Data journalism in 2017: The current state and challenges facing the field today. Google News Lab
- Schrock A & Shaffer G (2017) Data ideologies of an interested public: A study of grassroots open government data intermediaries. *Big Data & Society* 4(1): 1-10
- Stalph F (2017) Classifying data journalism. *Journalism Practice* 12(10): 1332-1350
- Stewart KN (1948) Review of *Journalism in the United States* by Robert W Jones. *The New England Quarterly* 21(2): 279-280
- Tufte ER (1983) *The visual display of quantitative information*. Cheshire: Graphics Press
- Uskali TI & Kuutti H (2015) Models and streams of data journalism. *The Journal of Media Innovations* 2(1): 77-88
- Van Schalkwyk F, Chattapadhyay S, Cañares M & Andrasen A (2016) Open data intermediaries in developing countries. *Journal of Community Informatics* 12(2): 9-25