

Article

ProPublica’s Data Journalism: How Multidisciplinary Teams and Hybrid Profiles Create Impactful Data Stories

Mathias-Felipe de-Lima-Santos

School of Communication, University of Navarra, Spain; mdelimas@unav.es

Submitted: 28 April 2021 | Accepted: 27 July 2021 | Published: 20 January 2022

Abstract

Despite growing interest in the emergence of technologies in journalistic practices, especially from the production perspective, there is still very little research on organizational structures and professional culture in relation to the deployment of these technologies. Drawing on six interviews and observation in staff meetings, this study aims to explore the nuances behind the professional roles of data journalists and how these relate to structural aspects of news organizations. The study focuses on the case of ProPublica, a news organization internationally renowned for its global excellence in data stories. This work considers boundary-making in the context of journalism and focuses on new professional roles in the news industry to produce a hybrid ethnography study based on qualitative data collected immediately before the Covid-19 pandemic hit the United States. The findings reveal the importance of hybrid profiles at ProPublica. While some journalists have had to expand their knowledge to learn more about new areas, such as coding and design, some non-journalistic professionals have had to develop writing skills, and this blurring of traditional boundaries forms an important aspect of ProPublica’s professional culture. The structure of the organization, divided into two teams engaged in cross-sector activities, helps to promote data skills and collaboration with other journalists, which also serves to mitigate any individual lack of experience on certain topics. The article concludes by suggesting that the growing importance of these new professional roles has broader implications for the development of data skills in the newsroom, and also discusses the limitations that can arise from the increasing overlap between journalistic and non-journalistic roles.

Keywords

data journalism; hybrid profile; journalism; multidisciplinary teams; news nerds; ProPublica

Issue

This article is part of the issue “New Forms of Media Work and Its Organizational and Institutional Conditions” edited by Salla-Maaria Laaksonen (University of Helsinki) and Mikko Villi (University of Jyväskylä).

© 2022 by the author(s); licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

There has been much discussion about how the increasing pervasiveness of technology has resulted in new dynamics in digital communication at a theoretical and conceptual level. In this respect, emerging technologies are increasingly shaping newsroom culture, as journalists find themselves taking on new roles (Carlson & Lewis, 2015) and having to adapt to novel values, norms, and practices (Tandoc, 2018). Thus, technology can intensify work pressures and expand the boundaries of work to include a broader range of activities that were not previously seen as an obligatory dimension of journalism, blurring normative boundary distinc-

tions. Conversely, boundary maintenance work establishes discursive claims about who belongs to the field, creating divisions between the work of core and peripheral actors (Eldridge, 2017). As non-traditional journalistic formats are accounting for an ever-growing portion of work in the field, there is a need to better understand these new peripheral actors and the ways they may be transforming the organizational culture (Schapals et al., 2019).

In the current scenario where the journalism industry is changing by leaps and bounds, some organizations are more successful than others in bridging the strategy–implementation gap between their day-to-day operations and their long-term goals by attracting exceptional

professionals with myriad skills (Tandoc, 2018). This is the case of ProPublica, a nonprofit digital-native media organization based in the United States, that defines its work on the use of investigative data journalism as a way to expose “abuses of power and betrayals of the public trust by government, business, and other institutions” (ProPublica, 2021a, para. 1). The organization received initial funding of 30 million dollars from the Sandler Foundation in 2008, which was used to produce real-world investigative stories principally aimed at enthusiasts of watchdog reporting.

Prior research described organizations like ProPublica as social news enterprises, which are public service media organizations with “a strong commitment to social goals and the need to justify their work in terms of impact” (Sparviero, 2020, p. 796). Consequently, social news enterprises are extending the original idea of public journalism to offer content investigative and explanatory journalism rather than attempting to compete with existing news organizations, aiming to increase levels of trust and, consequently, engagement with the public. Thus, these news outlets exemplify the dynamic development of social and economic initiatives through their news products (Usher, 2017). To achieve this, ProPublica produces quality investigative journalism by bringing together a team of journalists and programmers who go beyond traditional forms of journalism to create interactive stories built around data and multimedia. These stories helped to attract the additional capital the organization needed to continue operating. ProPublica’s reputation is seen through its numerous recognitions and awards, such as six Pulitzer Prizes and a myriad of online journalism awards (ProPublica, 2021b).

Data journalism is therefore aligned with the company’s goals, as it provides the ability to uncover stories that were not previously being discussed or reported, a highly significant contribution to today’s media landscape (Coddington, 2015). These goals, among other factors, have influenced ways of working and organizational dynamics in media outlets, putting data journalists, in some cases, at the center of the news production process (Appelgren & Lindén, 2020). Where it previously attracted niche and segmented audiences, data journalism is now an industry trend seeking wider legitimation in a sector “where fluidity is a defining element in journalistic processes, practices, positions, and products” (Hermida & Young, 2019, p. 33). Although ProPublica’s model cannot be easily replicable by competitors due to the lack of funds, it is an important case to examine because of its characteristics that stimulate the production of data stories.

Despite growing interest in the emergence of technologies in journalistic practices, especially from the production perspective, there is still very little research on organizational structures and professional culture in data journalism practice. To fill this gap, this study focuses on the case of ProPublica, internationally renowned for its global excellence in data stories.

Drawing on theories of boundary-making concerning journalism and focusing on new professional roles emerging in the industry, this article highlights the importance of hybrid profiles and multidisciplinary teams at ProPublica, enriching and expanding the literature of the role of these practitioners in the production of high-impact journalism. The study takes a two-stage approach. First, it considers the importance of multidisciplinary teams as an aspect of ProPublica’s professional culture. These professionals have emerging roles that extend beyond traditional journalistic practices, and their backgrounds include a wide range of knowledge fields. Second, it discusses the organization’s dynamics and technology-driven culture, which supports team members while also fostering internal and external collaboration. The article proposes the following research questions:

RQ1: How do ProPublica’s data team characteristics and competencies help to spur its investigative data journalism work, while these new actors inscribe themselves in the field?

RQ2: How does ProPublica’s organizational dynamic with internal and external collaborations support the production of data stories?

2. Theoretical Grounding

2.1. *Boundaries of Journalism Between the Work of Core and Peripheral Actors*

The concept of “boundary-work” was developed by the sociologist Thomas Gieryn to discuss the difficulties involved in defining what should and should not be considered “science,” and how understandings of what a field is can confer authority on certain social actors and limit the ways actors inscribe themselves in particular spaces. In other words, boundaries are established in a certain field to prevent or contest the emergence of new players attempting to enter from outside the sector and transform it (Gieryn, 1983).

In striving to meet the news industry’s changing needs triggered by social and technological change, the defined boundaries of journalism have become increasingly blurred (Eldridge, 2017), bringing to “the fore questions of what journalism is and what it should be” (Carlson, 2018, p. 1). Boundaries are established to exclude actors, practices, norms, and values that are not considered legitimate in the realm of journalism. However, this understanding treats journalism as a stable object, whereas a series of recent studies has shown that journalism is expanding to encapsulate other activities, such as blogging, social media content, fact-checking, etc. (Carlson & Lewis, 2015).

In part, attempts by journalists to set boundaries are linked to survival, as “social boundaries...yield greater cultural and material resources for insiders” (Lewis, 2012,

p. 841). In trying to exclude others, journalists use the need for objectivity, ethics, and impartiality in their work to contest the admittance of other actors. New professionals are described as “too emotional, too opinionated, too activist, or as relying too much on hearsay” (Schapals et al., 2019, p. 20) to adapt to institutional norms and conventions. Conversely, these new actors show that journalism is not a solid discipline, and is in a constant state of flux, especially in the digital era.

Nevertheless, by establishing these boundaries, journalists are legitimating their profession and structuring their social world to maintain control. Studies have shown that newsrooms are divided into three types of specialized actors: journalists, technologists, and businesspeople. Traditionally, these professionals work within their “silos,” creating social distance between the groups (Kosterich, 2021). This epistemic authority aims to bring power and prestige, and also provides some material benefits (Carlson & Lewis, 2015). For example, journalists may have easier access to high-profile figures or confidential sources using their press pass, and have legal protection if they publish leaked information (Eldridge, 2017).

In the digital media sector, new actors are less constrained by the field’s normative pressures, allowing them to experiment with unconventional ideas and solutions. In doing so, these new entrants are trying to distinguish themselves by preserving their dominant vision and influencing journalists’ work (Tandoc, 2018). This is the case for pioneer journalists, a group of actors dedicated to incorporating new organizational forms, relying on experimentation as a way to redefine the field and its structural foundations (Hepp & Loosen, 2021). A recent study on intra-organizational collaboration points out that IT departments have become a key factor impacting news outlets’ innovative capabilities (Westlund et al., 2021).

Building on these insights, much attention has been paid to the emergence of these new professionals and the new type of empowerment they enjoy in journalism. Eldridge (2017) describes these professionals who openly claim to belong to the journalistic field as “interlopers,” while Ferrucci and Vos (2017) call them “peripheral actors.” Common sense suggests that these professionals are challenging news organizations by extending the boundaries of what journalism is today. In this sense, Belair-Gagnon and Holton (2018) saw a need to expand the concept of “interlopers” due to nuances regarding their roles in the news cycle, and this resulted in three different levels being defined: explicit interlopers, implicit interlopers, and intralopers. The first category includes non-traditional journalism actors who work on the periphery of the profession, contributing to the creation of products and services. Implicit interlopers are less clearly aligned toward journalism, but do not reject journalistic objectivity or impartiality. Intralopers are distinct in that their activities are not journalism-oriented, but they work inside news organizations, using

their expertise to improve news production processes (Belair-Gagnon & Holton, 2018).

In light of these differences, I argue that data journalists can be perceived as peripheral actors, as they resemble implicit interlopers. In comparison to other areas, data journalism has been granted a certain level of acceptance in the news industry through adopting some of its established values, norms, and routines. Simultaneously, these practitioners try to distinguish themselves from other peripheral actors by promoting their values, such as an open data culture and collaborative mindset (Lewis & Usher, 2014; Stalph, 2020). Thus, these professionals are engaged in forms of hybrid journalism, drawing on their structurally diverse backgrounds to merge media skills with other areas of expertise. To better understand this phenomenon, it seems sensible to analyze organizations where data teams are well established and aligned with the goals of the newsrooms, such as the case of ProPublica.

On the other hand, the entanglements between data journalism and other forms of data work create new dependencies and synergies that introduce new actors working in novel forms of collaboration with non-journalistic institutions (Baack, 2018). For example, the role of civic tech organizations in developing data journalism across Africa shows that the boundaries of journalism are being transformed as evolutionary pressures are imposed on the system. Journalists are engaging in constant interaction with non-news workers, changing the patterns of interaction between these actors, their environments and habits, and prompting the question of how these transformations reflect on surrounding institutions and their business practices (Splendore & Brambilla, 2021). Thus, it is important to understand how these hybrid profiles are creating a new cultural logic in newsrooms.

2.2. Hybrid Profiles: The Advent of New Professionals and Their Role in Data Journalism

Technological developments have brought transformation to the media industry, which has slowly taken a more active role in these innovative processes. This has meant adopting new organizational structures and changes to capacity and resources. At the same time, there is a need to deal with advanced computational capabilities, which previously most journalists did not have the skills to approach. In this context, innovation has taken a pivotal role in reconfiguring newsrooms. This has resulted in new ways of producing and telling stories in influential formats that cross media boundaries. Thus, not only have the boundaries of journalism been expanded, but new professionals have also become part of newsrooms, challenging the traditional logic of journalism (Parasie & Dagiral, 2013). These new actors incorporate knowledge from other areas, mainly business and technology, including product management, data, analytics, and programming (Kosterich, 2021). Described by

Kosterich (2020, p. 52) as “news nerds,” they represent “new forms of professional journalists working in jobs at the intersection of traditional journalist positions and technologically-intensive positions.” These actors, who in most cases have hybrid backgrounds or more specialized skills, are typically more willing than traditional journalists to adapt to change. As a result, they are responsible for introducing new practices, norms, and roles into the industry, and thus play a paramount role in the media industry’s shift toward innovative organizational structures (Malmelin & Villi, 2017).

These hybrid roles are gradually becoming more valued for their contributions to journalistic work (Westlund et al., 2021). In data journalism, these actors have been described in the scholarly literature as journalist-programmers, programmer-journalists, journo-devs, and journo-coders (Hannaford, 2015; Parasie & Dagiral, 2013). With hybrid profiles, demand for these practitioners has grown in newsrooms across the world in an environment where technology shapes journalism and vice versa (Splendore & Brambilla, 2021). Studies suggest that bringing programmers into the newsroom challenges some journalistic principles, reshaping how news is produced and, by extension, distributed (Hermida & Young, 2019). These new professionals act as a “driving force to produce more effective and efficient news by harnessing the power of technological advancements” (Kosterich, 2020, p. 52). Although coders and technologists possess knowledge and skills enabling them to navigate through technical complexity and exercise decision-making power in newsrooms, these are not the only news nerds to have emerged in the data journalism industry. Previous studies have shown that roles vary greatly from culture to culture and newsroom to newsroom (Young et al., 2018), and it has been reported that some of these practitioners do not consider themselves to be journalistic actors, even though their tasks overlap with various forms of journalism (Baack, 2018).

In Europe, scholars have revealed a different attitude to connections between technology and journalism. In the United Kingdom, much less has been written about hybrid profiles, but more about multidisciplinary teams. In a study by Hannaford (2015), two major legacy news organizations were studied, the BBC and the Financial Times. In both organizations, teams composed of programmers, journalists, and designers worked closely together to produce interactive stories. Indeed, multidisciplinary teams that share a common organizational goal tend to foster innovative thinking and promote innovation in newsrooms (Westlund et al., 2021). Conversely, another study has suggested that the multidisciplinary teams vs. hybrid profiles dichotomy has more to do with the size of the news outlet than with cultural aspects. Smaller news organizations are more likely to have one hybrid practitioner due to the “lack of advanced computational skills and a technological infrastructure” (Borges-Rey, 2016, p. 837), while large,

elite news organizations can afford multidisciplinary teams that involve a division of labor (Fink & Anderson, 2015). This approach has been widely adopted in other newsrooms around the world, for example in Australia (de-Lima-Santos et al., 2021).

Newsrooms are struggling to adapt to and embrace these changes in the journalism profession and the resulting new norms and practices. The collaborative mind, for instance, is a break from the traditional mentality of journalism, from highly competitive single newsroom environments where journalists looking for a scoop are reluctant to collaborate, to a “new model” of multiple news outlets and organizations sharing information to expose wrongdoing on a global scale (Carson & Farhall, 2018). Thus, the capacity for innovation is increasingly being developed collectively, and leading news organizations are placing a higher value on cooperative efforts as a key mode of governance (de-Lima-Santos & Mesquita, 2021), impacted by internal and external forces (Westlund et al., 2021). This collaborative mindset can be applied to distinct business units, functions (intra-organizational), and organizations at national (inter-organizational) and international (transnational) levels (Heft et al., 2019). However, the challenge is to maintain and reinforce these collaborative principles over the longer term.

In the highly competitive environment created by the media industry, news nerds can provide a competitive advantage to publishers in the process of digitalization and help them adapt to a datafied world. This process of change is iterative and usually aligned to the external environment. In this sense, newsrooms are strongly influenced by award-winning news organizations, which define the cultural capital, i.e., constitute modes of production and levels of development for subsequent projects (Hermida & Young, 2019). To adapt more quickly to these advances, newsrooms are required to modify their internal structures and processes (Kosterich, 2020). However, the historical organizational structures of elite newsrooms create a normative conflict that either consigns data teams to the margins or fully incorporates them into newsroom culture (Stalph, 2020). Thus, this article addresses the importance of multidisciplinary teams in association with hybrid profiles in the industry, a focus so far lacking in the scientific literature.

3. Methodology

Through a qualitative approach, this study investigates the importance of data journalism in ProPublica’s organizational dynamic. The work embraces a hybrid ethnographic approach by combining observations and in-depth interviews based on qualitative data collected during fieldwork at ProPublica. This approach has been widely used by other researchers, demonstrating that it is an effective way to explore the distinct norms and routines adopted by news organizations (Hermida & Young, 2019).

The first stage of the project involved newsroom observation. This fieldwork aims to capture the participants' point of view, and all details observed while conducting the study are noted, including details of newsroom norms and the routines adopted on the data desk. This is important because it enables the research to elicit a comprehensive picture by seeing through the eyes of the key actors involved in a process. Additionally, the research involved observing an editorial agenda meeting and team meeting to understand the rules and processes involving other units in the newsroom (Bryman, 2012).

The method proposed by Emerson et al. (2011) was used to analyze the data gathered during the observations, following a three-step process: (a) close observations are conducted and systematic notes are taken of what is observed; (b) two-step qualitative analytical coding of the fieldnotes is carried out—open coding, in which the researcher reads the fieldnotes line-by-line and notes all the themes that emerge, followed by focused coding for fine-grained analysis, reducing the number of topics from the previous stage; and (c) the findings are described in the form of narrative “tales,” combining the themes that emerged in stage b to create “a thematic narrative that is fieldnote-centered” (Emerson et al., 2011, p. 202).

Observation data is then triangulated with in-depth, semi-structured interviews to obtain additional information about settings and patterns not captured during the observations. On average, the interviews lasted 45 minutes. These semi-structured interviews covered topics that emerged from the observations, including biography, involvement with the data journalism community, conceptions concerning data journalism, information about routines, organizational relationships, external collaboration, and projects carried out. This part of the study was designed to ensure gender balance among the respondents (see Table 1), with the hope of gaining access to a diverse mix of ideas, priorities, and methods, and avoiding the homophily trap, that is, focusing on similar people who share similar perspectives and provide similar information. By considering gender diversity, this study aimed to obtain an assessment from practitioners as complete and precise as possible, presenting their different views of the underlying situation and approaches to dealing with it. However, it is worth mentioning that gender diversity was not a level of analysis in this study, and, importantly, that only team members

who were in the newsroom at the time of the visits were interviewed.

The observation and interviews were conducted over two days in the second week of March 2020 and were constrained by the Covid-19 crisis, as it happened immediately before the pandemic hit the United States. However, these steps provided sufficient data with which to answer the research questions proposed by this study.

4. Findings

4.1. Data Journalism at the Center: Two Data Teams Working Together

It was the beginning of spring 2020, a year marked by a tragedy that changed the entire world. On arriving in New York, I went to the ProPublica office, located on the 13th Floor of a building on Manhattan's Avenue of the Americas, popularly known as Sixth Avenue. This is an affluent neighborhood filled with the offices of bankers and high-growth companies, proof that the funding ProPublica has received has had a significant impact on its business model. In part, this is thanks to generous donors since its foundation in 2008 (Sparviero, 2020). “Especially in the beginning when we’re just starting to gain credibility, and winning a Pulitzer early on at ProPublica certainly helped us to establish” (R5). Investigative journalism, therefore, became a form of strategic value, helping to attract interest in funding the outlet.

In 2019, ProPublica raised about 4.7 million dollars in online donations, while gifts and major grants above 50 million consolidated more than 19.9 million dollars (ProPublica, 2020). These funds allowed the organization to build a highly qualified team, despite its small newsroom, in discordance with findings of Fink and Anderson (2015) who reported “some fairly profound differences between the way that data journalism was practiced at larger, more resource-rich news organizations” (p. 470). Although ProPublica is equipped with significant resources, when compared with its peers, the news outlet is still small.

The organization began producing data journalism soon after its foundation, and this is an important long-term investment for the company. Data journalism at ProPublica is “split into two separate teams [the data team and the news app team] but, in many ways, there’s a lot of overlap” (R4). Both are led by ProPublica’s deputy

Table 1. Interviewees.

Code	Position	Gender
R1	News Applications Developer	Female
R2	News Applications Developer	Male
R3	News Applications Developer	Female
R4	Data Reporter	Male
R5	Editor	Female
R6	Editor	Male

managing editor. There are many similarities in their tasks and, “in theory, a lot of the people in either can do both tasks” (R5). In terms of their specific roles, the news app team is dedicated to more technically complex tasks and those that require interactive and in-house, scalable solutions (Usher, 2017), while data reporters are responsible for requesting, collecting, and analyzing data, and collaborating with reporters on data stories. “It’s mostly going to be a reporting aid and not something that the public will ever see....So, we often work side-by-side with a news app developer,” explained R5, referring to how the teams are organized in the newsroom.

The units are composed of more than a dozen professionals. This is an indicator of their importance in the newsroom, which is also demonstrated by how the units have grown over time:

Our [data] team now consists of me and seven data reporters, which is huge because when I was hired [in October 2013] I was the only data reporter. In fact, even when I took over the team, we had two data reporters. So, since June–July of 2017, we have nearly tripled in size. (R5)

In total, the data team consists of four reporters responsible for general data reporting for ProPublica’s national operation, one dedicated to covering the tech industry and algorithms (referred to as computational journalists), one data reporter who works at ProPublica’s Illinois office, and one dedicated to the local reporting network project. Similarly, the news app team is made up of “six people. We have three people who are in the office right now and we have three people who are remote. It will soon be two and four” (R6). By having part of the unit working remotely, the news app team is used to dividing up responsibilities and relies on Slack, an online collaborative software tool. This corresponds with the findings of a study by Moran (2020), who describes the growing popularity of online collaborative softwares in virtual newsrooms, which are increasingly coexisting with physical newsrooms, creating new forms of sociability, collectivity, and control within newsmaking. Thus, the news app team’s tasks are distributed between coverage of the Federal Government, based in the Washington DC office, a broad swath of the Midwest, based in the Illinois office, and several other projects, such as Electionland, Visual Evidence, and the Local Reporting Network. Both teams, along with the other newsroom members, use Slack to exchange information and ideas, hold meetings, and resolve any issues that occur.

Data stories produced by these teams are not merely daily news stories, but aim to produce a greater impact on society, echoing Sparvierio’s (2020) definition of social news enterprises. In general, these data stories require a lot of time, effort, and dedication from the practitioners who create them. “I think it’s a luxury in the news industry. We’ll spend the time that we need to dig into a story. And many news organizations don’t have the

staff or the budget to do that” (R5). This does not mean that all the projects are complex, although I observed a high level of complexity in some analysis and visualization, requiring a certain knowledge of statistics and coding for team members to implement their solutions, for instance, the Polluter’s Paradise series. On the other hand, Dollars for Docs is an example of a project that did not demand complex analysis and became one of ProPublica’s best-known projects. Dollars for Docs simply combined several data sets and made them available on one portal. One result of this project was that this data has now been released by the government, “but before they did not, it was on 20 different pharmaceutical websites” (R5). This project means that the public can access the data to check whether any doctor or teaching hospital has received money from pharmaceutical or medical device companies. These are specific characteristics of ProPublica, which differ from those previously reported in the data journalism scholarship, as organizations suffer from limited resources including “time, tools, manpower, and the financial means and expertise” to produce data stories (Fink & Anderson, 2015, p. 470).

4.2. Hybrid Schemes to Support the Production of Data Stories

Another important aspect of the organization is that news apps and data teams work as knowledge and collaboration hubs for the newsroom. The relationship between these teams and other journalists in the newsrooms expands beyond their work tasks, as they have lunches and coffee breaks together, which reinforces their relationships. As a result, many journalists working internally at ProPublica value data very highly. These journalists typically pitch ideas to their teams that use datasets or pitch the datasets themselves. Conversely, practitioners on the data or news app teams may find an interesting dataset or story and contact a journalist who can help them to get access to the sources.

In particular, the structure of the organization, composed of two data teams acting across sectors, helps to promote data skills and collaboration with other journalists. This internal collaboration is important to the success of data stories, as these professionals can leverage their combined potential to the fullest, serving to overcome any individual lack of knowledge on particular topics. However, some interviewees bemoaned the fact that data journalists are sometimes seen as a service desk. “I am always cognizant, we don’t want to become a service desk. We have to be treated as co-collaborators and not just the data folks who do an analysis and give you your results, neither make any calls nor do any part of reporting. Fortunately, that has happened very rarely at ProPublica” (R5).

In part, this concern relates to the makeup of the teams behind the projects. They come from a variety of backgrounds, including computer science, design, and law, among others. For example, R2 went to

design school, where he also studied computer science. However, he always believed that “design is a field that is good when you combine it with another field....[In my previous job,] I was getting interested in the combination of computer science, design, and journalism” (R2). Consequently, he began doing some side projects in journalism, which brought him into contact with a community of journalists, educators, entrepreneurs, and advocates passionate about journalism at the Newsgeist conference, allowing him to get his first job in the news industry (Lewis & Usher, 2014).

Similarly, R3 studied computer science, but she “always had a natural interest in journalism.” Thus, data journalism seemed to her an opportunity to “combine my interests, computer science, and journalism. [To gain experience in journalism,] I worked for the student paper in college....Then, I realized that data visualization was my way in. So, I decided to learn DataViz (data visualization) in my spare time” (R3). Shortly afterward, she was offered a Google News Lab Fellowship at ProPublica, which was her first step toward data journalism. However, the fellowship only lasted two months, meaning that she had to look for new job opportunities. After a placement at the MIT, she got a job at New York Public Radio. As a local reporter, R3 did shoe-leather reporting, going out onto the streets looking for people to interview:

I just took it upon myself to pad the pavement and go to community board meetings. I did things that traditional reporters do when they’re learning how to report, because I wanted to learn how to be a good reporter. I needed to understand what this is like because I also felt I wasn’t gonna get another opportunity to do that and I was probably right. As a data journalist, you’re typically not going to community board meetings and interviewing angry tenants. So, I learned by doing and also by reading and listening to the people around me....I think it can be challenging for data journalists to sometimes just understand what traditional reporting means and looks like because you don’t get much practice in it. Unless that’s where you started and then you transitioned into data journalism, but a lot of people I know have only ever done data journalism. So, I think that you’ll have a nebulous idea of what it’s like to be a traditional reporter but having an on-the-ground experience that I had, it was really helpful. (R3)

4.3. *The Increased Specialization of Knowledge to Delimit Boundaries*

Although these professionals have some experience that allows them to write stories, the majority bemoaned the specialization that typically ends up limiting their writing to “annotation in infographics” (R1). “Most of the writing that I do happens inside the graphic, like an explainer text” (R3). This echoes the findings of Stalph (2020),

who found that data journalists are either consigned to the margins or fully incorporated into newsroom culture. Similarly, this is a way to establish journalism boundaries to prevent or contest the emergence of new players in the field of journalism (Carlson & Lewis, 2015).

The journalists, meanwhile, had had to learn coding skills, showing that the definition of journalism is expanding to encapsulate other activities (Belair-Gagnon & Holton, 2018). “I didn’t have anywhere close to what a typical data journalist would have. I knew and I was very good at Excel. But I didn’t know any of the programming languages....I had to learn R and Python so that I can do more complex, ambitious work.” (R4). R6, meanwhile, “started building side projects like maps that updated crime numbers,” while he was studying journalism. He took “whatever digital classes existed, which were not that many at that time” (R6). This allowed him to get an internship at the *Los Angeles Times* on their data desk, where he further developed these skills. Similarly, R5 also “took a bunch of data journalism classes. Moreover, I worked at the data library for NICAR. So, when I graduated in early 2012, I was hired full time for a nonprofit investigative reporting center, much like ProPublica.” The fact that she was immediately hired in a role like this shows the importance of this training to develop her data skills.

However, there is a consensus that you cannot take “every person working for a tech firm in Silicon Valley and put them in a newsroom and expect that they’re going to automatically be a great data journalist” (R5). Data journalists need to demonstrate that they have the necessary skills. This shows that there is a certain level of acceptance in the news industry through adopting some of its established journalistic standards (Belair-Gagnon & Holton, 2018; Carlson & Lewis, 2015). For example, the computational journalist was hired after publishing a post on Medium in which he reported his findings on the Federal Communications Commission’s comments on net neutrality using machine learning algorithms. Having a law and computer science degree, he demonstrated that he also could do journalism by producing this analysis (R5). By combining these distinct skills and expanding the boundaries of journalism (Carlson & Lewis, 2015), while turning news nerds into journalists (Kosterich, 2020) and putting both together in multidisciplinary teams (Hannaford, 2015), ProPublica makes use of multiple internal resources in its quest to produce high-impact journalism, in line with the ideas of Borges-Rey (2016) who found that “the best stories they have produced were those where data journalists collaborated with specialised correspondents” (p. 838).

4.4. *Making an Impact Locally*

In striving to make the greatest possible impact nationally and internationally, ProPublica also relies on external collaborations and partnerships. “We are not as ubiquitous as the *New York Times*, the *Wall Street Journal*,

or the *Washington Post*. What I think is interesting is that a lot of our stories end up running in the *Times* or the *Post*....[Our website] is not a destination website, especially because we publish once a day,” explained R5. In this context, the news outlet found in collaboration, mainly at the national level, a key vehicle to reach a wider public and exercise an important influence over government decisions and policies. This is in line with previous research showing that data journalists tie together to reach the common goal of providing a quality product efficiently (de-Lima-Santos & Mesquita, 2021; de-Lima-Santos & Salaverría, 2021; Heft et al., 2019).

By 2019, ProPublica had 228 publishing partners (ProPublica, 2020). According to the interviewees, the news outlet always has an eye on maximizing the impact of its stories, and collaborative alliances are an important strategy for reaching this goal. Collaboration has become a way for the organization to achieve its long-term financial goals, by showing donors the importance of ProPublica’s journalism (Sparviero, 2020), particularly investigative data journalism. The news app team even includes a member whose role is entirely dedicated to establishing partnerships and collaborations.

Among these projects, the Local Reporting Network is one of the most important cooperative programs. “At this point, it’s about a little over 20 newsrooms all across the US that we work with. They pitch us for a year-long project and ProPublica pays their salary and they get access to our research team, news app data, and engagement” (R6). In this way, ProPublica promotes data journalism beyond its newsroom and shares its know-how and experience with local newsrooms. One data journalist from each team is responsible for working with these newsrooms, and this is “the same job we all do, but working with those local reporters instead of ProPublica’s reporters” (R3).

Importantly, not every project in the local reporting network has a data component. In those that do, there is a higher level of involvement. For example, “in the case of Hawaii, that’s as far deep as we’ll go. Fully co-reporters with our [local] reporter, and we’ll be doing much of the work. Then, whenever it comes time to write the story, we’ll be helping to write the story and all the other stuff” (R4). Otherwise, the collaboration is limited to “a consultation on something, if those local journalists want to request data from an agency or they got this data back and they are trying to figure out what to do with it” (R4).

In one example, ProPublica partnered with The Texas Tribune, where they “co-hired 10 people who will live in Texas to report on Texas-based issues” (R6). This entailed a five-year commitment of 5.75 million dollars (ProPublica, 2020). One ProPublica project involved building a map of toxic areas in the US, and there were several severely affected areas in Texas. This team helped “to build a graphical walkthrough of the Texas areas” (R6). In this process, the news app team rarely “[does] graphics with other news organizations. In fact, it’s usually hard

to share graphics across news organizations” (R4). Thus, the collaborations focus on the sources and knowledge on the ground that these local journalists have, which, “in some cases, we just won’t be able to do without them” (R6). The idea behind this is to build local impact through these partnerships, more than developing these skills together. “We get the ability to distribute journalism with a local organization. So, if we’re gonna be writing about a place and we want local impact then it’s best to publish with the local news organization” (R6). This echoes the findings of Heft et al. (2019), who demonstrated that transnational collaboration among organizations from different European countries was important in creating impact across Europe.

5. Conclusions

This research aimed to explore the role of multidisciplinary teams and hybrid profiles at ProPublica and how these professionals and their modes of work contribute to significant data projects. In its quest to produce high-impact investigative data journalism, ProPublica has adopted a two-team approach (RQ1). While the data team is responsible for requesting, collecting, and analyzing data, the news app team is more focused on creating visuals and apps, elevating the “computational prowess and product development” (Usher, 2017, p. 1128). The latter is also in charge of more complex data analysis. While some journalists have had to expand their knowledge and learn more about other areas, such as coding and design, some non-journalistic professionals have also had to develop writing skills. These hybrid roles are an important aspect of ProPublica’s professional culture. Although the organization’s work involves specialized components that may require a high level of specific expertise, the professionals working in these units have hybrid backgrounds that make it difficult to categorize them as either journalists or technologists (Splendore & Brambilla, 2021).

Additionally, ProPublica’s data teams comprise multidisciplinary teams so that different skill sets complement each other. These findings demonstrate that ProPublica became a data journalism powerhouse by combining not only multidisciplinary teams, but also practitioners with hybrid backgrounds. In data journalism scholarship, it is common to treat these two approaches separately, which expands the dichotomy of journalist-programmers (Hannaford, 2015; Parasie & Dagiral, 2013) versus multidisciplinary teams (de-Lima-Santos et al., 2021; Fink & Anderson, 2015). This study suggests that this model of collaboration between journalists and non-journalists has broad implications for the development of data skills in the newsroom, such as the case of ProPublica. This extends the literature by showing that the combination of hybrid profiles in multidisciplinary teams is a strong strategic fit to create and develop data teams in newsrooms, complementing each other’s skillsets and ensuring coherence in the overall approach to the matter.

On the other hand, there is a certain level of skepticism about the capabilities of these journalists in terms of writing and reporting data stories, even though some have already worked in other newsrooms. This skepticism aligns with the established journalistic boundaries that work to stratify different groups of people in newsrooms (Carlson & Lewis, 2015). Similarly, specialization is a process of establishing boundaries, as professionals are concentrating on and becoming experts in a particular subject or skill. However, this tends to happen to a much lesser extent at ProPublica, where the focus on data journalism since the organization's foundation has fostered a culture of collaboration with data teams. Furthermore, the data editors are constantly reinforcing the idea that data teams are not a "service desk." This is important because data has inherent biases and may discriminate against certain groups of people, meaning that the involvement of data journalists is vital from the early stages of investigation to ensure a deep level of understanding (Tong & Zuo, 2019). Acknowledging the rise of these hybrid profiles is an important step in reengineering "journalism's longstanding professional boundaries" (Kosterich, 2021, p. 24).

In another vein, partnerships and collaboration are central to news organizations' ability to innovate (RQ2), as stories are increasingly being developed collectively, meaning that collaborative networks are becoming a key mode of governance (de-Lima-Santos & Mesquita, 2021; de-Lima-Santos & Salaverría, 2021; Westlund et al., 2021). Through collaboration, ProPublica gains access to sources and information that they would not otherwise have (Heft et al., 2019). Additionally, this process has a significant impact on local communities, as local newsrooms working with ProPublica have decisively influenced the political agenda on several topics. Thus, these actors transcend existing professional boundaries, leading to new interdependencies and collaborations (de-Lima-Santos & Mesquita, 2021). Their interdependent, long-lasting relationships also impact ProPublica's trajectory. Hence, awards are a result of these joint efforts.

Although not every organization can hope to successfully emulate ProPublica, the American digital news outlet serves as a model for many others, based on the principles of transparency and accountability leading to trust (Howard & Constantaras, 2019). The organization's hybrid business model, as defined by Sparviero (2020), is also becoming a model for others, which can inspire other newsrooms to establish data practices, policies, and capabilities to collect and use on news reporting.

Concerning the limitations of the research, this study is restricted to the observations and views of practitioners who were in the New York newsroom on the days I visited. Unfortunately, this was constrained by the Covid-19 crisis, as it was planned to last longer, allowing me to demonstrate causality and how these practitioners approach unexpected events. Similarly, the newsroom was not working at full capacity. Future research could

explore similar topics from the perspectives of practitioners working in other ProPublica offices, or remotely. Additionally, it would be interesting to consider the perspectives of partner organizations to understand how these partnerships are helping to develop data skills at the local level. It would also be interesting to investigate how other members of the newsroom perceive data journalists and the limitations that arise from collaboration between the two spheres.

In conclusion, I was able, despite the limitations discussed, to highlight some relevant aspects of data journalism practices at ProPublica. This study contributes to the scholarly literature by expanding the concept of journalistic boundaries in the context of data journalism epistemologies, shedding light on the experiences and realities of a specific group that had not previously been discussed at this level.

Acknowledgments

This project was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 765140. The author thanks the ProPublica's teams for sharing their expertise and assistance throughout all aspects of this study.

Conflict of Interests

The author declares no conflict of interests.

References

- Appelgren, E., & Lindén, C. G. (2020). Data journalism as a service: Digital native data journalism expertise and product development. *Media and Communication*, 8(2), 62–72. <https://doi.org/10.17645/mac.v8i2.2757>
- Baack, S. (2018). Practically engaged: The entanglements between data journalism and civic tech. *Digital Journalism*, 6(6), 673–692. <https://doi.org/10.1080/21670811.2017.1375382>
- Belair-Gagnon, V., & Holton, A. E. (2018). Boundary work, interloper media, and analytics in newsrooms. An analysis of the roles of web analytics companies in news production. *Digital Journalism*, 6(4), 492–508. <https://doi.org/10.1080/21670811.2018.1445001>
- 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>
- Bryman, A. (2012). *Social research methods*. Oxford University Press.
- Carlson, M. (2018). Boundary work. In T. P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh, & A. Sehl (Eds.), *The international encyclopedia of journalism studies*. Wiley.
- Carlson, M., & Lewis, S. (2015). *Boundaries of journalism*.

- Routledge.
- Carson, A., & Farhall, K. (2018). Understanding collaborative investigative journalism in a “post-truth” age. *Journalism Studies*, 19(13), 1899–1911. <https://doi.org/10.1080/1461670X.2018.1494515>
- Coddington, M. (2015). Clarifying journalism’s quantitative turn. *Digital Journalism*, 3(3), 331–348. <https://doi.org/10.1080/21670811.2014.976400>
- de-Lima-Santos, M.-F., & Mesquita, L. (2021). Data journalism beyond technological determinism. *Journalism Studies*, 22(11), 1416–1435. <https://doi.org/10.1080/1461670X.2021.1944279>
- de-Lima-Santos, M.-F., & Salaverría, R. (2021). From data journalism to artificial intelligence: Challenges faced by *La Nación* in implementing computer vision in news reporting. *Palabra Clave*, 24(3), 1–40. <https://doi.org/10.5294/pacla.2021.24.3.7>
- de-Lima-Santos, M.-F., Schapals, A. K., & Bruns, A. (2021). Out-of-the-box versus in-house tools: How are they affecting data journalism in Australia? *Media International Australia*, 181(1), 152–166. <https://doi.org/10.1177/1329878X20961569>
- Eldridge, S. A. (2017). *Online journalism from the periphery*. Routledge.
- Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2nd ed.). University of Chicago Press.
- Ferrucci, P., & Vos, T. (2017). Who’s in, who’s out? Constructing the identity of digital journalists. *Digital Journalism*, 5(7), 868–883. <https://doi.org/10.1080/21670811.2016.1208054>
- Fink, K., & Anderson, C. W. (2015). Data journalism in the United States: Beyond the “usual suspects.” *Journalism Studies*, 16(4), 467–481. <https://doi.org/10.1080/1461670X.2014.939852>
- Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, 48(6), 781–795. <https://doi.org/10.2307/2095325>
- Hannaford, L. (2015). Computational journalism in the UK newsroom. *Journalism Education*, 4(1), 6–21.
- Heft, A., Alfter, B., & Pfetsch, B. (2019). Transnational journalism networks as drivers of Europeanisation. *Journalism*, 20(9), 1183–1202. <https://doi.org/10.1177/1464884917707675>
- Hepp, A., & Loosen, W. (2021). Pioneer journalism: Conceptualizing the role of pioneer journalists and pioneer communities in the organizational re-figuration of journalism. *Journalism*, 22(3), 577–595. <https://doi.org/10.1177/1464884919829277>
- Hermida, A., & Young, M. L. (2019). *Data journalism and the regeneration of news*. Routledge. <https://doi.org/10.4324/9781315163895>
- Howard, A., & Constantaras, E. (2019). Journalists and the media. In T. Davies, S. B. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data* (pp. 395–405). African Minds.
- Kosterich, A. (2020). Managing news nerds: Strategizing about institutional change in the news industry. *Journal of Media Business Studies*, 17(1), 51–68. <https://doi.org/10.1080/16522354.2019.1639890>
- Kosterich, A. (2021). Reengineering journalism: Product manager as news industry institutional entrepreneur. *Digital Journalism*. Advance online publication. <https://doi.org/10.1080/21670811.2021.1903959>
- Lewis, S. C. (2012). The tension between professional control and open participation. *Information, Communication & Society*, 15(6), 836–866. <https://doi.org/10.1080/1369118X.2012.674150>
- Lewis, S. C., & Usher, N. (2014). Code, collaboration, and the future of journalism: A case study of the hacks/hackers global network. *Digital Journalism*, 2(3), 383–393. <https://doi.org/10.1080/21670811.2014.895504>
- Malmelin, N., & Villi, M. (2017). Media work in change: Understanding the role of media professionals in times of digital transformation and convergence. *Sociology Compass*, 11(7), 1–13. <https://doi.org/10.1111/soc4.12494>
- Moran, R. E. (2020). Subscribing to transparency: Trust-building within virtual newsrooms on slack. *Journalism Practice*. Advance online publication. <https://doi.org/10.1080/17512786.2020.1778507>
- Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media & Society*, 15(6), 853–871. <https://doi.org/10.1177/1461444812463345>
- ProPublica. (2020). *2020 annual report*. <https://www.propublica.org/reports>
- ProPublica. (2021a). *About us*. <https://www.propublica.org/about>
- ProPublica. (2021b). *Awards*. <https://www.propublica.org/awards>
- Schapals, A. K., Maares, P., & Hanusch, F. (2019). Working on the margins: Comparative perspectives on the roles and motivations of peripheral actors in journalism. *Media and Communication*, 7(4), 19–30. <https://doi.org/10.17645/mac.v7i4.2374>
- Sparviero, S. (2020). Hybrids before nonprofits: Key challenges, institutional logics, and normative rules of behavior of news media dedicated to social welfare. *Journalism & Mass Communication Quarterly*, 97(3), 790–810. <https://doi.org/10.1177/1077699020932564>
- Splendore, S., & Brambilla, M. (2021). The hybrid journalism that we do not recognize (anymore). *Journalism and Media*, 2(1), 51–61. <https://doi.org/10.3390/journalmedia2010004>
- Stalsh, F. (2020). Evolving data teams: Tensions between organisational structure and professional subculture. *Big Data & Society*, 7(1), 1–13. <https://doi.org/10.1177/2053951720919964>
- Tandoc, E. C. (2018). Five ways BuzzFeed is preserving (or transforming) the journalistic field. *Jour-*

nalism, 19(2), 200–216. <https://doi.org/10.1177/1464884917691785>

Tong, J., & Zuo, L. (2019). The inapplicability of objectivity: Understanding the work of data journalism. *Journalism Practice*, 15(2), 1–17. <https://doi.org/10.1080/17512786.2019.1698974>

Usher, N. (2017). Venture-backed news startups and the field of journalism: Challenges, changes, and consistencies. *Digital Journalism*, 5(9), 1116–1133. <https://doi.org/10.1080/21670811.2016.1272064>

Westlund, O., Krumsvik, A. H., & Lewis, S. C. (2021).

Competition, change, and coordination and collaboration: Tracing news executives' perceptions about participation in media innovation. *Journalism Studies*, 22(1), 1–21. <https://doi.org/10.1080/1461670X.2020.1835526>

Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great data journalism?: A content analysis of data journalism awards finalists 2012–2015. *Journalism Practice*, 12(1), 115–135. <https://doi.org/10.1080/17512786.2016.1270171>

About the Author



Mathias-Felipe de-Lima-Santos (PhD) is a researcher at the University of Navarra, Spain, under the JOLT project, a Marie Skłodowska-Curie European Training Network funded by the European Commission's Horizon 2020. Previously, he was a visiting researcher at the Queensland University of Technology, Australia. He is co-editor of the book *Journalism, Data and Technology in Latin America* published by Palgrave Macmillan in 2021. His research interests include the changes of the journalistic practice with a focus on business models, data, and novel technologies.