49. **Data-Driven Gold Standards: What the Field Values as Award-Worthy Data Journalism**

*Wiebke Loosen*

**Abstract**

This chapter explores the relationship between the datafication of society and a datafied journalism and introduces awards as a means to study the evolution of data journalism.

**Keywords:** Data Journalism Awards, datafication, datafied journalism, data society, journalism research, co-creation

**Introduction: Journalism’s Response to the Datafication of Society**

Perhaps better than in the early days of data journalism, we can understand the emergence of this new reporting style today as one journalistic response to the datafication of society (Loosen, 2018). Datafication refers to the ever-growing availability of data that has its roots in the digitalization of our (media) environment and the digital traces and big data that accrue with living in such an environment (Dijck, 2014). This process turns many aspects of our social life into computerized data—data that is to various ends aggregated and processed algorithmically. Datafication leads to a variety of consequences and manifests itself in different ways in politics, for instance, than it does in the financial world or in the realm of education. However, what all social domains have in common is that we can assume that they will increasingly rely on an ever more diverse range and greater amount of data in their (self-)sense-making processes.

Situating the datafication of journalism in relation to the datafication of wider society helps us to look beyond data journalism, to recognize
it as “only” one occurrence of datafication in journalism, and to better understand journalism’s transformation towards a more and more data-based, algorithmicized, metrics-driven or even automated practice (Loosen, 2018). In particular, this includes the objects and topics that journalism is supposed to cover, or, put differently, journalism’s function as an observer of society. The more the fields and social domains that journalism is supposed to cover are themselves “datafied,” the more journalism itself needs to be able to make sense of and produce data to fulfil its societal role. It is this relationship that is reflected in contemporary data journalism which relies on precisely this increased availability of data to expand the repertoire of sources for journalistic research and for identifying and telling stories.

Awards: A Means to Study What Is Defined and Valued as Data Journalism

One way of tracing the evolution of data journalism as a reporting style is to look at its output. While the first studies in journalism research tended to focus more on the actors involved in its production and were mainly based on interviews, more and more studies have recently been using content analysis to better understand data journalism on the basis of its products (Ausserhofer et al., 2020). Journalism awards are a good empirical access point for this purpose for several reasons. Firstly, award submissions have already proved to be useful objects for the analysis of genres and aspects of storytelling (e.g., Wahl-Jorgensen, 2013). Secondly, data journalism is a diffuse object of study that makes it not only difficult, but, rather, preconditional, to identify respective pieces for a content analysis. The sampling of award nominees, in turn, avoids starting with either a too narrow or too broad definition—this strategy is essentially a means of observing self-observation in journalism, as such pieces represent what the field itself regards as data journalism and believes are significant examples of this reporting style. Thirdly, nominations for internationally oriented awards are likely to influence the development of the field as a whole as they are highly recognized, are considered to be a kind of gold standard and, as such, also have a cross-border impact. In addition, looking at international awards allows us to investigate a sample that covers a broad geographical and temporal range.

However, it is also important to keep in mind that studying (journalism) awards brings with it different biases. The study we are drawing from here is based on an analysis of 225 nominated pieces (including 39 award-winning
pieces) for the Data Journalism Awards (DJA)—a prize annually awarded by the Global Editors Network\(^1\)—in the years 2013 to 2016 (Loosen et al., 2020). This means that our sample is subject to a double selection bias: At first it is self-selective, since journalists have to submit their contributions themselves in order to be nominated at all. In the second step, a more or less annually changing jury of experts will decide which entries will actually be nominated. In addition, prizes and awards represent a particular form of “cultural capital,” which is why award-winning projects can have a certain signal effect for the field as a whole and serve as a model for subsequent projects (English, 2002). This also means that awards not only represent the field (according to certain standards), but also constitute it. That is, in our case, by labelling content as data journalism, the awards play a role in gathering together different practices, actors, conventions and values. They may be considered, then, to have not just an *award-making function* but also a *field-making function*. This means that award-worthy pieces are always the result of a kind of “co-construction” by applicants and jurors and their mutually shaped expectations. Such effects are likely to be particularly influential in the case of data journalism as it is still a relatively new reporting style with which all actors in the field are more or less experimenting.

**Evolving but Not Revolutionizing: Some Trends in (Award-Worthy) Data Journalism**

Studies that analyze data-driven pieces generally demonstrate that the evolution of data journalism is by no means a revolution in news work. As a result, they challenge the widespread belief that data-driven journalism is revolutionizing journalism by replacing traditional methods of news discovery and reporting. Our own study broadly concurs with what other empirical analyses of “daily” data journalism samples have found (Loosen et al., 2020). These only represent fairly limited data collections, but they do at least allow us to trace some developments and perhaps, above all, some degree of consistency in data journalism output.

In terms of who is producing data-driven journalism on an award-worthy level, results show that the “gold standard” for data journalism, that is, worthy of peer recognition, is dominated by newspapers and their online departments. Over the four years we analyzed, they represent by far the largest group among all nominees as well as among award-winners (total:

\(^1\) https://www.globaleditorsnetwork.org/about-us/, https://www.datajournalismawards.org/
43.1%; DJA awarded: 37.8%). The only other prominent grouping comprises organizations involved in investigative journalism such as ProPublica and the International Consortium of Investigative Journalists (ICIJ), which were awarded significantly more often than not. This might reflect the awards’ inherent bias towards established, high-profile actors, echoing findings from other research that data journalism above a certain level appears to be an undertaking for larger organizations that have the resources and editorial commitment to invest in cross-disciplinary teams made up of writers, programmers and graphic designers (Borges-Rey, 2020; Young et al., 2018). This is also reflected in the team sizes. Of the 192 projects in our sample that had a byline, they named on average just over five individuals as authors or contributors, and about a third of projects were completed in collaboration with external partners who either contributed to the analysis or designed visualizations. This seems particularly true for award-winning projects that our analysis found were produced by larger teams than those only nominated (M = 6.31, SD = 4.7 vs M = 4.75, SD = 3.8).

With regards to the geographies of data journalism that receives recognition in this competition, we can see that the United States dominates: Nearly half of the nominees come from the United States (47.6%), followed at a distance by Great Britain (12.9%) and Germany (6.2%). However, data journalism appears to be an increasingly global phenomenon, as the number of countries represented by the nominees grew with each year, amounting to 33 countries from all five continents in 2016.

Data journalism’s reliance on certain sources influences the topics it may or may not cover. As a result, data journalism can neglect those social domains for which data is not regularly produced or accessible. In terms of topics covered, DJA nominees are characterized by an invariable focus on political, societal and economic issues, with almost half the analyzed pieces (48.2%) covering a political topic. The small share of stories on education, culture and sports—in line with other studies—might be unrepresentative of data journalism in general and instead result from a bias towards “serious” topics inherent in industry awards. However, this may also reflect the availability or unavailability of data sources for different domains and topics or, in the case of our sample, the applicants’ self-selection biases informed by what they consider worthy of submission and what they expect jurors to appreciate. In order to gain more reliable knowledge on this point of crucial importance, an international comparative study that relates data availability and accessibility to topics covered by data reporting in different countries would be required. Such a study is still absent from the literature but could shed light on which social domains and topics are covered by which
analytical methods and based on which data sources. Such an approach would also provide valuable insight to the other side of this coin: The blind spots in data-driven coverage due to a lack of (available) data sources.

One recurring finding in content-related research on data journalism is that it exhibits a “dependency on pre-processed public data” from statistical offices and other governmental institutions (Borges-Rey, 2020; Tabary et al., 2016; Young et al., 2018). This is also true of data-driven pieces at an award-worthy level: We observed a dependence on data from official institutions (almost 70% of data sources) and other non-commercial organizations such as research institutes and NGOs, as well as data that are publicly available, at least on request (almost 45%). This illustrates, on the one hand, that data journalism is making sense of the increased availability of data sources, but, on the other, that it also relies heavily on this availability: The share of self-collected, scraped, leaked and requested data is substantially smaller. Nonetheless, data journalism has been continually linked to investigative reporting, which has “led to something of a perception that data journalism is all about massive data sets, acquired through acts of journalistic bravery and derring-do” (Knight, 2015; Parasie, 2015; Royal & Blasingame, 2015). Recent cases such as the Panama Papers have contributed to that perception. However, what this case also shows is that some complex issues of global importance are embedded in data that require transnational cooperation between different media organizations. Furthermore, it is likely that we will see more of these cases as soon as routines can be further developed to continuously monitor international data flows, for example, in finance, not merely as a service, but also as deeper and investigative background stories. That could stimulate a new kind of investigative data-based real-time journalism, which constantly monitors certain finance data streams, for example, and searches for anomalies.

Interactivity counts as a quality criterion in data journalism, but interactivity is usually implemented with a relatively clear set of features—here our results are also in harmony with other studies and what is often described as a “lack of sophistication” in data-related interactivity (Young et al., 2018). Zoomable maps and filter functions are most common, perhaps because of a tendency to apply easy-to-use and/or freely available software solutions, which results in less sophisticated visualizations and interactive features. However, award-winning projects are more likely to provide at least one interactive feature and integrate a higher number of different visualizations. The trend towards rather limited interactive options might also reflect

2 https://panamapapers.icij.org/
Journalists’ experiences with low audience interest in sophisticated interactivity (such as gamified interactivity opportunities or personalization tools that make it possible to tailor a piece with customized data). At the same time, however, interactive functions as well as visualizations should at best support the storytelling and the explanatory function of an article—and this requires solutions adapted to each data-driven piece.

A summary of the developmental trends over the years shows a somewhat mixed pattern, as the shares and average numbers of the categories under study were mostly stable over time or, if they did change, they did not increase or decrease in a linear fashion. Rather, we found erratic peaks and lows in individual years, suggesting the trial-and-error evolution one would expect in a still emerging field such as data journalism. As such, we found few consistent developments over the years: A significantly growing share of business pieces, a consistently and significantly increasing average number of different kinds of visualizations, and a (not statistically significant, but) constantly growing portion of pieces that included criticism (e.g., on the police’s wrongful confiscation methods) or even calls for public intervention (e.g., with respect to carbon emissions). This share grew consistently over the four years (2013: 46.4% vs 2016: 63.0%) and was considerably higher among award winners (62.2% vs 50.0%). We can interpret this as an indication of the high appreciation of the investigative and watchdog potential of (data) journalism and, perhaps, as a way of legitimizing this emerging field.

From Data Journalism to Datafied Journalism—and Its Role in the Data Society

Data journalism represents the emergence of a new journalistic sub-field that is co-evolving in parallel with the datafication of society—a logical step in journalism’s adaptation to the increasing availability of data. However, data journalism is no longer a burgeoning phenomenon; it has, in fact, firmly positioned itself within mainstream practice. A noteworthy indicator of this can again be found when looking at the Data Journalism Awards. The 2018 competition introduced a new category called “innovation in data journalism,” which suggests that data journalism is no longer regarded as an innovative field in and of itself, but is looking for novel approaches in contemporary practice.3

3  https://www.datajournalismawards.org/categories/
We can expect data journalism’s relevance and proliferation to co-evolve alongside the increasing datafication of society as a whole—a society in which sense making, decisions and all kinds of social actions increasingly rely on data. Against this background, it is not too difficult to see that the term “data journalism” will become superfluous in the not too distant future because journalism as a whole, as well as the environment of which it is part, is becoming increasingly datafied. Whether this prognosis is confirmed or not: The term “data journalism,” just as the term “data society,” still sensitzes us to fundamental transformation processes in journalism and beyond. This includes how and by what means journalism observes and covers (the datafied) society, how it self-monitors its performance, how it controls its reach and audience participation, and how it (automatically) produces and distributes content. In other words, contemporary journalism is characterized by its transformation towards a more data-based, algorithmicized, metric-driven or even automated practice.

However, data is not a “raw material”; it does not allow direct, objective or otherwise privileged access to the social world (Borgman, 2015). This way of understanding data is all the more important for a responsible data journalism as the process of society’s datafication advances. Advancing datafication and data-driven journalism’s growing relevance may also set incentives for other social domains to produce or make more data available (to journalists), and we are likely to see the co-evolution of a “data PR,” that is, data-driven public relations produced and released to influence public communications for its own purposes. This means that routines for checking the quality, origin and significance of data are becoming increasingly important for (data) journalism, and raise the question of why there may be no data available on certain facts or developments.

In summary, I can organize our findings according to seven “Cs”—seven challenges and underutilized capacities of data journalism that may also be useful for suggesting modified or alternative practices in the field.

**Collection.** Investigative and critical data journalism must overcome its dependency on publicly accessible data. More effort needs to be made in gaining access to data and collecting them independently.

**Collaboration.** Even if the “everyday” data-driven piece is becoming increasingly easier to produce, more demanding projects are resource- and personnel-intensive, and it is to be expected that the number of globally relevant topics will increase. These will require data-based investigations across borders and media organizations, and, in some cases, collaboration with other fields such as science or data activism.
Crowdsourcing. The real interactive potential of data journalism lies not in increasingly sophisticated interactive features but in crowdsourcing approaches that sincerely involve users or citizens as collectors, categorizers and co-investigators of data (Gray, 2018).

Co-Creation. Co-creation approaches, common in the field of software development, can serve as a model for long-term data-driven projects. In such cases, users are involved in the entire process, from finding a topic to developing one and maintaining it over a longer period.

Competencies. Quality data journalism requires teams with broad skill sets. The role of the journalist remains important, but journalists increasingly need a more sophisticated understanding of data, data structures and analytical methods. Media organizations, in turn, need resources to recruit data analysts who are increasingly desirable in many other industries.

Combination. Increasingly complex data requires increasingly sophisticated analysis. Methods that combine data sources and look at these data from a variety of perspectives could help paint more substantial pictures of social phenomena and strengthen data journalism’s analytical capacity.

Complexity. Complexity includes not only the data itself, but its increasing importance for various social areas and political decision making. In the course of these developments, data journalism will increasingly be confronted with data PR and “fake data.”

What does this mean? Taking into account what we already know about (award-winning) data journalism in terms of what kinds of data journalism are valued, receiving wide public attention and contributing to a general appreciation of journalism, what kinds of data journalism do we really want? In this regard, I would argue that responsible data journalism in the data society is one that: Investigates socially relevant issues and makes the data society understandable and criticizable by its own means; is aware of its own blind spots while asking why there are data deficiencies in certain areas and whether this is a good or a bad sign; actively tries to uncover data manipulation and data abuse; and, finally, keeps in mind, explains and emphasizes the character of data as “human artefacts” that are by no means self-evident collections of facts, but are often collected in relation to very particular conditions and objectives (Krippendorff, 2016).

At the same time, however, this means that data journalism’s peculiarity, its dependency on data, is also its weakness. This limitation concerns the availability of data, its reliability, its quality and its manipulability. A responsible data journalism should be reflexive about its dependency on data—and it should be a core subject in the discussion on ethics in data
journalism. These conditions indicate that data journalism is not only a new style of reporting, but also a means of intervention that challenges and questions the data society, a society loaded with core epistemological questions that confront journalism’s assumptions about what we (can) know and how we know (through data).

These questions become more urgent as more and increasingly diverse data is incorporated at various points in the “circuit of news”: As a means of journalistic observation and investigation, as part of production and distribution routines, and as a means of monitoring the consumption activities of audiences. It is in these ways that datafied journalism is affecting: (a) journalism’s way of observing the world and constructing the news from data, (b) the very core of journalism’s performance in facilitating the automation of content production, (c) the distribution and circulation of journalism’s output within an environment that is shaped by algorithms and their underlying logic to process data, and (d) what is understood as newsworthy to increasingly granularly measured audience segments.

These developments present (data) journalism with three essential responsibilities: To critically observe our development towards a datafied society, to make it understandable through its own means, and to make visible the limits of what can and should be recounted and seen through the lens of data.

Works Cited


---

**About the Author**

**Wiebke Loosen** is a senior journalism researcher at the Leibniz Institute for Media Research/Hans-Bredow-Institut (HBI) in Hamburg, as well as a Professor at the University of Hamburg.