

THE NEED TO REFLECT: DATA JOURNALISM AS AN ASPECT OF DISRUPTED PRACTICE IN DIGITAL JOURNALISM AND IN JOURNALISM EDUCATION

Lars Rinsdorf and Raoul Boers

University of Applied Sciences for Media, Stuttgart, Germany

rinsdorf@hdm-stuttgart.de

Open data is both an opportunity and a challenge for journalism. Our paper describes how newsroom quality management has to be readjusted to provide accurate news in data driven journalism and how these changes affect the pedagogy and learning environments of journalism education. We discuss qualitative changes in newsgathering caused by the availability of open data sets and big data sources and their consequences for quality management against the background of structuration theory. Although there is no clear evidence signaling disruptive change in quality management, two developments cause a need for normative realignment, expansion of individual skill sets and an inventory of novel resources.

1 KEY TERMS

Data journalism and the sharing of open data are vibrant and emerging practices described with a great variety of terms in academia. Our understanding of the terms is outlined in this section. According to the definition of the Open Data Initiative we define open data as data that anyone can access, use or share. Data exists on a spectrum from closed, to shared, to open. Open data is characterized by its diversity and rawness. Raw data indicates the sense that it is unadulterated by its publisher and left open for analysis by whoever downloads or links to it as a resource. Open data can ideally be linked to for processing and repurposing and is sustainably supported in an open ecosystem, structured according to open standards and methods.

Open data is not synonymous to big data. In public discourse these two terms are often used interchangeably, leading to confusing debates and an obfuscated understanding of the two. Big data are not necessarily open, nor are open data necessarily big. Indeed, most big data sets are not open as they are valuable commodities for application in business ventures. Big data analysis is an ever-increasing determinant for enterprise strategy.

Following Hammond (2015) and partly Coddington (2015) we define data journalism as a qualitatively new way of reporting which gains insights about relevant societal trends by analyzing open data sets using (semi-) automatized methods to detect meaningful patterns in data structure. Although we can't observe an established set of professional routines (Ausserhofer 2015), data journalism isn't just a technology-driven phenomenon, but is characterized by a post-objective professional mindset stressing the transparency of editorial processes, user integration and the qualitatively new epistemological value of big data sets. Transparency is a strategy of quality management by sharing your approaches of information gathering with users. User integration means to publish a story at a certain point and continue to work on this story afterwards embedding and reacting on the users' feedback on it. The new epistemological value of big data sets is, that knowledge is gained inductively by exploring data, discussing preliminary findings and detect meaningful data patters aided by algorithms.

Comparable to Rogers' (2013) understanding of digital methodology in the social sciences, data journalism is not focused on the web as a virtual world but as a means to decode general societal phenomena that are embedded in online data. This phenomenon is also described as mediatization. To put it another words. We no longer analyze data sets, which are describing social interaction by collecting indicators. Instead of this we look at data, users are providing by interacting in their daily life. Data journalism could be described as the interaction of human and non-human actors, but nevertheless we distinguish it from algorithm journalism, where news making as a whole is more or less done automatically by using artificial intelligence, also known as robot journalism.

When discussing consequences of data journalism for newsroom quality management we refer to a wider understanding of news quality, integrating professional, normative, and market perspectives (Arnold 2009). This allows us to reflect the hybrid character of news and its different perception from a societal ("citizen value") and an economical ("consumer value") point of view (Neuberger 2011). Hence we differentiate between journalism with its specific societal functions

and professional rules and its organizational context provided by media companies (Altmeyden 2007). Following this general understanding of newsroom management, quality management could be defined in particular as set of reflexive practices to control news organizations using allocative and authoritative resources (Wyss 2013). Especially when it comes to the social responsibility of journalism dealing with data and statistics, these perspective is stressing the fact, that one might make demands on journalisms, but that you have to analyze very precisely, in which way news organizations and individual actors in journalism integrate these demands in their institutional environment.

2 OPEN DATA AS AN OPPORTUNITY FOR JOURNALISM

Although datafication is not the main driving force behind data journalism, the ubiquitous availability of data offers tremendous opportunities for journalism as long as this data is handled with respect for its qualitatively new character. The evolution of web technology from a document-based to a data-based network topology has changed the way individuals and organizations source their information and hence the way they structure and disseminate their knowledge. Using semantic web standards, data can be identified and linked, receiving meaning through its contextualization as a result. When datasets are made available as openly approachable objects with their own identifiers as linkable URLs, remarkable new opportunities for analysis and insight are created (Antoniou & van Harmelen, 2008). These opportunities require researchers to approach their research with a new mindset and with the skills to match, which traditional analysts - specifically in the field of media and journalistic reporting- lack.

The qualitatively novel situation for journalists is that their inquiry is no longer restricted to structured datasets with predefined variables, but can include massive sets of unstructured data that require an expert view for analysis. This is required, as the relevant variables must be reconstructed from implicit patterns in the data and its structure. Because of this development, a key asset to contemporary newsrooms is the data visualizer. Data visualization is an important practice for both analyzing numerical data as well as presenting said data in a palatable format for the general audience. Mancosu (2005) asserts that there is a renaissance of visual thinking in mathematics and logic. It is emphasized that visual models of complex mathematical concepts do not only serve heuristic and pedagogical purposes. Visualization can be a legitimate element of mathematical proof. Extending such notions to the realm of journalism and critical analysis of – say- open government data, visualization becomes a means of critical discovery of truths in an epistemologically independent way. In other words: a competent data visualization expert can debunk myths and decipher half-truths and represent the factual undercurrents to audiences that hitherto did not have access to this underlying reality, as that reality lay hidden in the untold numbers of complex data.

As we have asserted, the availability of big data sets is changing the character of data in editorial processes. In traditional journalism, data is mainly used for fact checking to improve the accuracy of news stories. In data journalism, data has a more prominent position at the start of the process: Exploring available data leads to news stories that are enriched by information from analog sources. This matches perfectly to a general trend in modern journalism towards user integration and a turn from products to processes. A well-known example for this is the Guardian's story about the spending of British MPs, where users supported the journalists analyzing thousands of files (Guardian, 2011). This development may be seen as part of a general switch in journalism from a focus on news and scoops to background information and the explanation of current trends. The major steps in that new editorial process are shown in Figure 1.

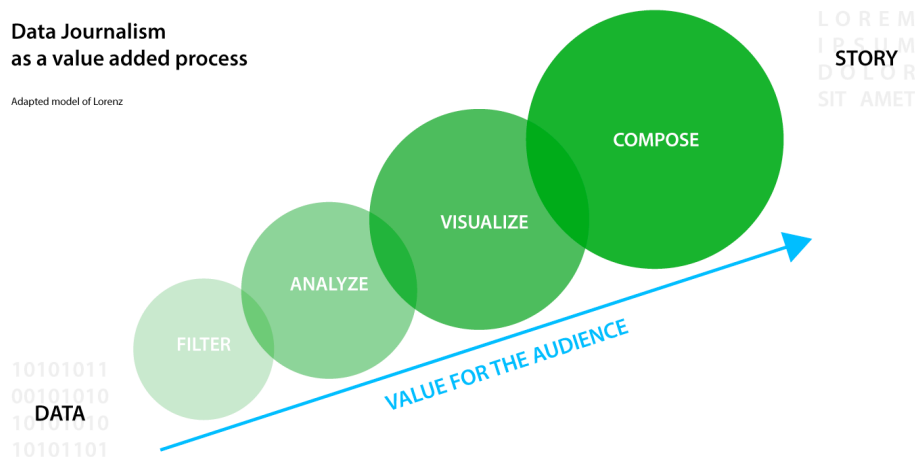


Figure 1: The Editorial Process of Data Journalism

Customers have a more important role to play in value networks in a digital economy. Particularly in journalism this has led to the emancipation of an active audience that influences news value by rating, commenting, or producing and sharing content amongst themselves (Bruns 2008). This is challenging the institutional role of journalism on the one hand, on the other hand journalism can improve the quality of stories through such co-creation and collaboration. In data journalism this could be observed in two different ways: News websites increase the user experience by offering interactive visualization which allow users to customize their view on data. In addition, journalists embed users in research in data stories, which they couldn't realize without the support of a crowd (Gray/Bounegru/Chambers 2012).

Data journalism is an example, not a driving force for change in professional norms of journalism or in the self-perception of journalists. Twentieth century journalism was focused on publishing final versions of accurately researched news stories at a certain deadline. In a digitalized and mediatized environment this institutional arrangement gets under pressure. The more the professional role of journalists switches from gate keepers to gate watchers (Neuberger 2009), publishing a story is no longer the end, but the beginning of a process of continuous improvement in interaction between journalists and users (Jarvis 2010). Under these circumstances journalists are losing their unique position in news production. Instead they are managing news processes as interactions between data, producers, coders, and news bots (Lewis 2015).

3 OPEN DATA AS A CHALLENGE FOR JOURNALISM

Although there are clear advantages for journalism in datafication, we shouldn't ignore the challenges. Generating news out of data requires a new skill set for journalists. Additionally, matters concerning data ownership and transparency should be taken into consideration. The practice of analyzing datasets requires a slew of skills novel to the traditionally trained journalist. Enriching data by recombining datasets with third party data, thereby maximizing their potential yield of complexly layered results, demands a level of skill too daunting for most journalists today.

The complexity of the contemporary journalistic trade and the ever more sought after skills of the versatile newsroom editor, require new levels of competence in numeracy and data manipulation. A recent report of the TOW Center for Digital Journalism (2015) asserts that contemporary journalism schools should teach at least basic mathematical skills and test their students' numerical capabilities. Being able to appreciate and evaluate the quality of data sets and being able to cleanse and structure data for analytical repurposing and enrichment are traditionally thought of as being part and parcel of ICT education and computer science programs. Yet -in part-journalism curricula at university level should incorporate these competences explicitly, or at the very least in minor programs. In addition, basic coding skills are advised.

Detecting relevant patterns in open data sets is a daunting challenge as data is available in a diverse array of structures and contents. More so, matters of trust, reliability and validity of data are confounding issues. Though there is no denying the participative potential of the web, we would like to stress that both the technological infrastructure and the content patterns are strongly influenced by powerful actors from the fields of government and the market economy. Dealing with big data has become a crucial part in maintaining political power and in maintaining the powerful ecosystems of a digitalized economy. Therefore the determining factors for selection and publication of data are often driven by strategic objectives of individual actors. Such dynamics can be problematic for independent journalism in news companies, as they face serious obstacles when transferring to digital business models. In worst-case scenarios journalism finds itself pinned between the stakes of its stakeholders. In addition, commercial actors, like for example an insurance company, might strategically use their data to tell corporate news stories within the scope of their own corporate media platforms.

4 CHANGES IN NEWSROOM QUALITY MANAGEMENT

The trends in digital journalism as described influence newsroom quality management. We can describe these influences from an empirical and a more theoretical perspective. From empirical findings in the data journalism field, the impact of the potential game-changer data journalism on current newsrooms isn't as high as one might expect. In a more general study on technology driven change in newsroom processes, Reich (2013) showed, that well-established newsroom routines are surprisingly robust against technological changes. Ekdale, Singer, Tully & Harmsen (2015) found out in their case-study on newsroom innovation, that journalists incorporate technology quite fast as long as it provides useful tools for news-gathering, but avoid it when it comes to change their professional culture and self-perception on a fundamental level – what data journalism does without a doubt. Ausserhofer (2015) proved this finding in his meta-review of international case studies, and figured out, that although some prominently featured data projects exist, the majority of data initiatives in journalism lack resources and appropriate skills. In his qualitative research on data journalism he provided evidence that some specific professional practices related to data journalism have already emerged (as we have sketched in Figure 2), but no professional routines. A qualitative leap is therefore required. Additionally a shared professional mindset about what data journalism is and how it is conducted could be seen as an indicator for the establishment of data journalism as a relevant sub-domain of journalism.

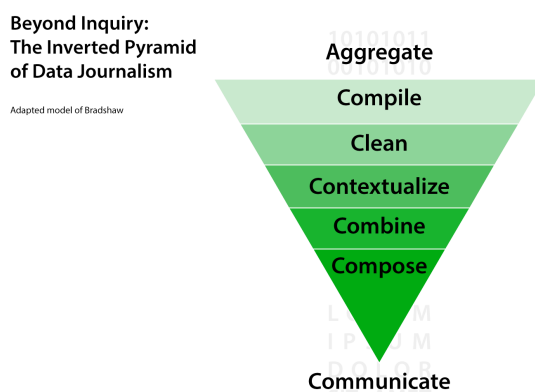


Figure 2: Shared Practices in Data Journalism

Within the framework of structuration theory (Giddens 2007) these findings could be sufficiently explained. News organizations are challenged by significant qualitative change in their organizational environment. On the one hand their institutional role in reducing complexity and societal sense-making (Kaplan 2006) is problematized by new powerful actors like e.g. online social networks. On the other hand established professional routines in part are no longer reproduced in the daily practice of journalists. Hence, news organizations are forced into loops of

reflexive control where they are using allocative and authoritative resources for structural adaptation to the environmental change, by strategic institutionalization (Buschow 2012) on the signification level and by readjustment of professional norms on an organizational level – for instance by connecting high-end journalism with certain media types like daily newspapers or stressing the importance of analytical long forms as prototypical quality journalism. The measure of success for the outcome of such activities depends on the reproduction of this institutional regime in professional practices of journalists, thereby reforming their professional identity.

There are several practical repercussions for quality management in data journalism projects. Firstly source credibility will gain importance and its characteristics will change. Source quality is no longer simply a matter of personal trust, but a matter of data consistency and of the invested strategic interests of data providers. Political motivations might for instance influence the definition of indicators in a dataset on employment. Secondly, in addition to fact checking, process transparency must be provided, allowing experts to evaluate research results and the tools that are used to gather, analyze and visualize data. Thirdly the user experience of interaction with data provided by news companies will emerge as an important criterion for quality from the perspective of marketable content.

Beside these changes in professional norms data journalism requires specific resources which have to be provided by news companies. Of course we're talking about secure and powerful server infrastructure to run interactive data projects. Data projects need well defined processes and key performance indicators to control the interaction of professionals and users involved in the production. We need repositories where the analytic tools developed in a project are stored for use in further projects. But the most important resource is the capability to staff and manage multi-disciplinary teams of journalists, coders, crowd managers etc. which are required to analyze advanced datasets successfully and to produce valuable and engaging content.

Finally, this leads us to also address quality management on a more general level. In data journalism value for news customers is often created in a value network of different actors inside and outside a news company. This creates a situation of uncertainty and market disruption, where neither markets nor hierarchies are an appropriate modus of coordination. Under such circumstances, trust could be an alternative mode of coordination to interact under acceptable transaction costs instead of market interaction or hierarchical integration (Möllering and Sydow, 2005). "Doing trust" (Frevert, 2014: 47) to stand the contingency in your organizational environment, for news organizations means to be aware of their vulnerability, but to still initiate self-energizing trust-building processes and consciously use the reflexive character of trust to maintain value networks in data projects.

5 CHANGES IN JOURNALISM EDUCATION ON UNIVERSITY LEVEL

Journalism education is struggling to keep up with developments in digital journalism in general, so data journalism specifically is a far off subject for most university-level educators. Indeed, although course subjects related to data science have gained traction, at many journalism schools debates are raging on whether journalists should want to become data scientists. Such debates are unfair simplifications and a more nuanced stance is in order.

Instead of simply tagging data skills and subjects onto traditional journalism programs, existing programs and their pedagogical strategies should be rebuilt from the ground up. Both contents and teaching methods should reflect the complexity of contemporary newsroom practices. Bradshaw (2015) proposes an approach based on problem-based peer learning. In such an approach students work together in an online newsroom simulation, tackling news reporting as a stream of activities that incorporate both classic journalism skills and contemporary digital skills in equal measure. Instead of focusing on producing a story, journalism classes should focus on the iterative network processes of online journalism. Multidisciplinary teamwork with rotating roles and responsibilities should remedy the inherent problem of overspecializing in a given role. Another pitfall to avoid is the practice of resorting to content curating and rehashing existing content instead of deeper penetration of issues through research. It is specifically with this issue that the competence of data journalism presents itself as a valuable and sought after skillset, not least in light of future employment in a market evermore reliant on data analysis.

Adapting or readjusting educational programs requires a realignment of the learning goals along its components: knowledge, skills and competences. The main focus of this process of realignment has been on digital skills related to publishing and editing for convergent media and data journalism skills are a relatively recent addition to that equation (Filak, 2014).

When attempting to include data journalism skills, it is important to set the scope of the learning goals. Data journalism in its current practice consists of two main variants that should be discerned. On the one hand there is investigative data journalism, dealing with dedicated, high cost and highly specialized data enquiry and research. On the other hand there is a more general data journalism that is set to become the de facto method for any high quality news team in their day-to-day operation. Where investigative journalists are given a generous budget in time and finances in order to employ high tech data and coding skills, general data journalists have limited budgets with more modest scopes befitting their means. Although the skillset of the general data journalist is shallower and less ambitious, their focus is more on broadening understanding by providing background analysis and giving context to news stories. In many newsrooms both variants of data journalist will operate in tandem (Uskali & Kuutti, 2015).

It is an open subject for debate whether journalism schools should provide for both variants of data journalism, or whether certain advanced data analysis skills should be sub-contracted to specialists outside the newsroom.

6 CONCLUSION

The emerging field of digital journalism causes disruption. Data journalism is one aspect of this. This disruption is reflected in the state of journalism education. Both are facing similar challenges to cope with the dynamics and complexity of multidisciplinary cooperation and economic pressuring of changing ecosystems in the field of media, from serial chains to multiplexing interdependent networks. And so both the professional and educational fields can use similar strategies to cope with these challenges. It remains a topic of discussion as to what extent educators should specialize or generalize. Journalism and education should embrace the idea that they cannot control the whole process of news making or education, but they have to manage collaborative and transparent networks and reflect on the ethical implications of such practice. It is now more important than ever for the journalism profession to focus on the role as sense-maker and audience guide.

REFERENCES

- Altmeppen, K. (2007). Das Organisationsdispositiv des Journalismus. In: K. Altmeppen (Ed.), *Journalismustheorien: Next Generation. Soziologische Grundlegung und theoretische Innovation*. Wiesbaden: VS, 281-302.
- Arnold, K. (2009). *Qualitätsjournalismus. Die Zeitung und ihr Publikum*. Konstanz: UVK.
- Ausserhofer, J. (2015). „Die Methode liegt im Code“: Routinen und digitale Methoden im Datenjournalismus. In: A. Maireder et al. (Eds.), *Digitale Methoden in der Kommunikationswissenschaft*. 87-111. doi: 10.17174/dcr.v2.5.
- Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and Beyond: From production to produsage*. New York et al.: Peter Lang.
- Coddington, M. (2015). Clarifying Journalism's Quantitative Turn, *Digital Journalism*, 3:3, 331-348, DOI: 10.1080/21670811.2014.976400
- Buschow, C. (2012). *Strategische Institutionalisierung durch Medienorganisationen. Der Fall des Leistungsschutzrechtes*. Köln: von Halem.
- Bradshaw, P., Hickman, J., & Jones, J. (2015). Stories & Streams: A Problem-Based Design for Student-Led Collaboration and Peer-to-Peer Teaching Across Media Practice Modules. *Collaborative Learning in Media Education*, 103.
- Ekdale, B., Singer, J. B., Tully, M., Harmsen, S. (2015). Making Change: Diffusion of Technological, Relational, and Cultural Innovation in the Newsroom. *Journalism & Mass Communication Quarterly*, Vol. 92(4) 938–958.
- Filak, V. F. (2014). *Convergent Journalism: An Introduction: Writing and Producing Across Media*. CRC Press.

- Frevert, U. (2014). Über Vertrauen reden: Historisch-kritische Beobachtungen. In: J. Barberowski, (Ed.), *Was ist Vertrauen? Ein interdisziplinäres Streitgespräch*. Frankfurt, New York: Campus, 31-48.
- Gray, J., Bounegru, L., Chambers, L. (2012). *The Data Journalism Handbook*. <http://datajournalismhandbook.org/1.0/en/> (26.1.2015)
- Guardian (2011). *Reading the riots. Investigating England's summer of disorder*. <http://www.theguardian.com/uk/series/reading-the-riots> (11.02.2016)
- Hammond, P. (2015). From computer-assisted to data-driven: Journalism and Big Data. *Journalism* 1–17 DOI: 10.1177/1464884915620205
- Jarvis, J. (2010). *Here's How We're Going To Revolutionize The News Business: The Cloud*. Online resource available under <http://www.businessinsider.com/revolutionizing-the-news-business-with-the-cloud-2010-4?IR=T> (26.01.2016)
- Kaplan, R.L. (2006). The News About New Institutionalism: Journalism's Ethic of Objectivity and Its Political Origins. *Political Communication*, 23:2, 173-185.
- Mancosu, P. (2005). *Visualization in logic and mathematics* (pp. 13-30). Springer Netherlands.
- C. Lewis, S. C. (2015). Journalism In An Era Of Big Data, *Digital Journalism*, 3:3, 321-330, DOI: 10.1080/21670811.2014.976399
- Möllering, G, Sydow, J. (2005). Kollektiv, kooperativ, selektiv: Vertrauen und Glaubwürdigkeit in Unternehmungen und Unternehmensnetzwerken. In: B. Dernbach & M. Meyer (Eds.), *Vertrauen und Glaubwürdigkeit. Interdisziplinäre Perspektiven*. Wiesbaden: VS, 64-93.
- Neuberger, C. (2011). *Definition und Messung publizistischer Qualität im Internet. Herausforderungen des Drei-Stufen-Tests*. Berlin: Vistas.
- Neuberger, C. (2009). Internet, Journalismus und Öffentlichkeit. Analyse des Medienumbruchs. In: C. Neuberger, C., Nuernbergk, M., Rischke (Eds.), *Journalismus im Internet: Profession – Partizipation – Technisierung*. Wiesbaden: VS, 19-105.
- Reich, Z. (2013). The Impact of Technology on News Reporting: A Longitudinal Perspective, *Journalism & Mass Communication Quarterly* 90(3) 417–434
- Rogers, R. (2013). *Digital Methods*. Cambridge: MIT Press.
- Uskali, T. I., Kuutti, H. (2015). Models and streams of data Journalism. *The Journal of Media Innovations*, 2(1), 77-88.
- Wyss, V. (2013). Qualitätsmanagement in Redaktionen. In: K. Meier (Ed.), *Journalismusforschung: Stand und Perspektiven*. Baden-Baden: Nomos. 89-104.