

Methods for Datafication, Datafication of Methods: Introduction to the special issue

By Stine Lomborg, Lina Dencik and Hallvard Moe

Over the past decades digital media have become a central infrastructure in society, and people around the globe rely on digital media every day to get things done: coordinate social activities, work, shop, handle financial transactions, participate in politics, consult with health professionals, or be entertained. Digital media enable processes of datafication (Mayer-Schönberger & Cukier, 2013): users' online activities leave digital traces that are transformed into data points in databases, kept by service providers and other private and public organisations, and repurposed for commercial exploitation, business innovation, surveillance – and research. Increasingly, this also extends to sensors and recognition technologies that turn homes and cities, as well as our own bodies, into data points to be collected and analysed (Hintz, Dencik and Wahl-Jorgensen 2018). So-called 'traditional' media industries, too, including public service broadcasting, have been datafied, tracking and profiling audiences, algorithmically processing data for greater 'personalisation' as a way to compete with new players and streaming services (Van den Bulck and Moe 2018). Datafication both raises new research questions and brings about new avenues, and an array of tools, for empirical research. This special issue is dedicated to exploring these, linking them to broader historical trajectories of social science methodologies as well as to central concerns and perspectives in media and communication research.

'Datafication', and its companion 'big data', was for a while the subject of a certain hype in public debate and research across disciplines. As a phenomenon and area of study, 'datafication' is typically qualified in relation to a specific domain: datafication of health, work, the public sector, the city, or everyday life. This further means that the study of datafication is cross-disciplinary in nature, and potentially opens up spaces for interdisciplinary collaboration. Media and communication research has a lot to contribute to other disciplines in this respect, both in regards to understanding and explaining the infrastructural and political-economic underpinnings of the platforms and services that drive datafication, and to exploring and theorizing how users, citizens and consumers engage with and are affected by datafication. At the same time, datafication is also a central concern for media and communication research itself, creating different ways of approaching the study of mediated phenomena. The technical ability to generate and analyse digital data traces has fueled a rich field of digital methods and advanced the discipline of computational social science. Studies of social media, in particular, have often incorporated methodological innovations that rely on algorithmically processing large amounts of data, whilst more established methods of content analysis of media messages have also been reconfigured with the advent of big data (Rogers 2013; Zamith and Lewis 2015). With this special issue, we weave together these streams by asking two questions: on the one hand, what research questions media and communication scholars should ask of datafication, and, on the other hand, how new and existing methods enable us to answer them. This special issue of *European Journal of Communication*, is based on revised versions of papers first presented at a symposium convened by the journal and the issue editors, and held at the University of Copenhagen in September 2019.

The study of datafication itself from within media and communication research warrants theoretical clarification about the nature of our field and its objects of study.

First, in digital media the object of study of media and communication is not clearly delimited to a relatively stable set of media institutions, genres or forms of production and reception. As the discipline of media studies developed over the 20th century, at least in the pre-digital era, the object of media studies was in a sense commonly accepted to be print, radio and television, whereas interpersonal communication was studied across adjacent disciplines from rhetoric to social psychology and linguistics. In his seminal essay on "the great divide", Everett Rogers (1999) noted that this was reflected in institutional divisions across universities in the western world where the study of mass and interpersonal communication was separated into

different departments. The same distinction can be found in the editorial of the first issue of the present journal: The "communication" in *European Journal of Communication* referred to "processes of public communication within and between societies and thus primarily to do with mass media and mass communication" (JGB, DM, KER, 1986, 3).

It would be difficult to find a journal introduction in our field that does not claim new, exciting or ominous technological change. The 1986 editorial quoted above made readers aware of a "so-called 'communications revolution'" that not only pressured institutional structures and practices of communication, but also gave rise to "an increase and diversification of the intellectual problems for communication research" (JGB, DM, KER, 1986, 4). Still, the advent of digital media has fundamentally challenged both the divide between mass and interpersonal media, and the nature of the domain of media and communication scholarship itself. What counts as media and thus relevant objects of study for media and communication scholarship in the digital age? Across personal devices, whether laptops or smartphones, websites or apps for banking, shopping, childcare and so on sit side by side with digital news sites and streaming services. All of these share the same material basis and to varying degrees distribute content, facilitate interaction, engage in data extraction and thus call for our attention. From a loose definition of media in terms of materiality, anything digital can be media. In a recent book on the internet of things, Bunz & Meikle (2018), for instance, suggest we may think of smart sensor-based objects as networked media that collect data and communicate with and about users. If anything digital can be media, it fundamentally challenges taken-for-granted notions in our field: What are the media institutions to be studied when the networked media can be anything from babysitter apps to rifles? And where are the "texts" in these media that audience studies still hold as a defining criteria (Livingstone 2004, Picone 2017)?

The struggles over delimitations of our field and changing of objects of study has led prominent scholars to suggest communication, rather than media, as the basic object of study (e.g. Jensen, 2010; Castells, 2007). In his book on media convergence, for instance, Jensen (2010) proposes to focus on the prototypical, relatively stable communication practices (one-to-one, one-to-many and many-to-many) that have historically traveled well across media, in what he labels a 'three-step flow of communication', nodding to Katz & Lazarsfeld's (1955) seminal study of the relation between mass and interpersonal media in the formation of public opinion. In their status of the field of communication scholarship in the wake of datafication research, Turow & Couldry (2018) advance a related move away from content-centric approaches to media towards seeing media as technologies for data extraction. This move implies joining forces with perspectives outside of our field to illuminate the "larger arena of knowledge production that we must also learn to associate with media" (p. 415). While Turow and Couldry are particularly referring to the intersection of media and communication scholarship with critical analysis of personal targeting, discrimination etc as treated in discussions of surveillance capitalism (Zuboff, 2018) and data justice (Dencik et al. 2016), we take their intervention to invite a broader engagement with data traces and questions for datafication - from entertainment media to smart cities and digital public sector initiatives - in our field. The 'loosening' of the boundaries around our object of study not only implies that our domain is broadening in scope, but also that other disciplines increasingly inhabit it. We should take this development as an opportunity to advance, synthesize and test our historically grounded theories and empirical knowledge of media and communication in society in relation to emergent research about digital media to better understand the economic, social and cultural implications of datafication.

Second, the relationship between communication and datafication must be unravelled to elaborate the possible contribution of our field to the study of datafication. One useful point of departure for such an endeavour is to advance the basic conceptualizations of 'data' and data-driven technologies in terms of media and communication. As critical data studies scholars have noted, data are neither "raw" (Gitelman, 2013) or neutral, nor do they make sense in and of themselves. In making data meaningful and understanding how they come to matter, communication theory holds crucial insights. Firstly, communication scholarship reminds anyone interested in data and knowledge production that data do not solely communicate by way of the

transfer of information from individuals to digital systems and back. We must also interrogate digital data in terms of ritualized communication that help sustain a certain social order. In that vein, lending inspiration from Lomborg & Frandsen (2016), we can understand digital communication to comprise communication with the system, the self and the social world (peers in social networks, but also, commercial entities and societal structures). Hence, datafication is premised on each and every micro-instance of communication with digital media and contributes to the ongoing configuration of digital systems (Jensen & Helles, 2017), the constitution of the self and social relationships. Jensen (2012) explores the role of meta-data as new forms of meta-communication, taking meta-data to refer to codes and contexts in time and space that may guide meaning-making with communication, both for the parties involved, for those making value out of data, and for scholarship. Digital meta-data can be retrieved and made analysis-ready to an extent that is historically unprecedented and warrants our empirical engagement. These two theoretical clarifications - regarding our object of study and the communicative framing of data - point to new questions: Do we have the methods for studying datafication? And how should datafication be used to study communication?

Studying (with) datafication

We seek to ground the methodological understanding and application of new methods in longstanding practices of empirical research. New methods do not emerge in a vacuum – there are historical continuities and predecessors, and old methods are still relevant. And yet, perhaps datafication enables us to deal with lines of inquiry that were difficult if not impossible to pursue before. We are today facing radically more and more easily available traces, used for more and more purposes. Making sense of this development calls for an interdisciplinary approach that draws on different entry-points. Much research focused on datafication has concerned itself with the technologies themselves, finding innovative ways to explore how data is algorithmically processed and transforming information environments through forms of ‘reverse engineering’ or ‘audits’ as a way to highlight new forms of gate-keepers and agenda-setters (e.g. Bucher 2012; Diakopoulos 2015; Rogers, 2013; Sandvig et al. 2014). This has advanced understandings of datafication as decision-making systems that shape the terms of mediation, knowledge production and social exchange. At the same time, the danger of ‘algorithmic fetishism’ (Monahan 2018) that drives a focus on opening up the ‘black-box’ of data systems as a way to make sense of digital infrastructures and social relations has led to a call for media scholars to more actively insert their long-standing engagement with the hermeneutic and action space between production and consumption into the study of datafication (Livingstone 2018), (re)claiming audience agency and everyday practices (Kennedy 2018), and emphasising the situated, contextual aspects of data as a way to understand dynamics of power (Dencik 2019). This suggests a critical engagement with the ways established methods in media and communication research, familiar in audience and production studies and beyond, remain relevant and crucial for making sense of data as more than its technical infrastructure.

With regards to the second question we pose - how datafication enables new and reconfigures existing methods for studying communication - we enter into an important, critical discussion in our field. The internet and especially social media made available to communication researchers for the first time, a whole new dimension of human communication. The potential for fresh insights seemed to greatly expand with the advent of ‘big data’, promising quick, even real-time, diagnosis of whole populations, thus ending the need for selections and theory-driven analysis. Especially when researchers started to utilize opportunities to access data through different software Application Programming Interfaces (API’s), we saw a blossoming of studies of different platforms, most dominantly Facebook and Twitter (e.g. Stoyceff et al, 2017). More recently, access to API’s were drastically limited following especially the Facebook-Cambridge Analytica privacy scandal, which revealed in early 2018 how consultancy firm Cambridge Analytica, through a researcher, had purchased personal data harvested from Facebook users

without their consent for political analysis and marketing purposes. Referred to as the “APIcalypse” (Bruns, 2019), the move actualized questions about researchers’ collaborations with social media platforms (Puschmann, 2019). Beyond problems with data access, more fundamental questions concern how we can use different kinds of data to understand processes of communication. To what extent does datafication not just offer, but require, us to think anew when it comes to the methods we use, and to what extent should we rely on the tried and tested approaches in our field? While digital infrastructures afford opportunities for studying communication in new ways, appropriating data for research does not exist outside of theory. The turn to digital methods and computational processes for analysing social phenomena is one that is embedded in particular epistemologies that also extend previously established understandings of the world familiar to the field of media and communication (boyd and Crawford 2012). In the datafication of methods, we need critical interventions questioning data quality, validity, ethics etc. (e.g. Lomborg & Bechmann, 2014) to invite continuous systematic reflections on how to conduct empirical research in our field.

With this special issue we would like to contribute to what in the past decade or so has been a growing methodological interest in the data “out there” – both in terms of how methods enable specific ways of seeing the social world and specific forms of knowledge production, ethical implications of using digital data for research, and methodological innovations that tap into vast data resources and the infrastructures of datafication as such. Rather than pursuing a purely “how to” methods-focused framing, we have developed this special issue together with the contributors and the journal editors as one that links new tools and procedures to underlying methodological discussions as well as the ontological and epistemological assumptions that undergird our fields. We have thus curated a special issue that will be asking what media and communication scholarship is bringing to our understanding of these issues, and how key pillars of communication research are or can be applied to the study of datafication.

Questions for Media and Communication Research

As such, this special issue grapples with approaches to empirical research that interlink questions of methods and tools with epistemology and practice. It discusses the datafication of methods, as well as methods for studying datafication.

Building on his work with digital methods, the first article by Richard Rogers examines the phenomenon of deplatforming, the banning of certain users (typically far-right actors, some of whom have gained internet celebrity status) from mainstream social media platforms such as Facebook, Twitter, Instagram and YouTube. Rogers asks if deplatforming is an effective strategy for regulating hate speech and extremism online, and if so - for who? As deplatformed users are known to find alternative platforms for their communicative endeavors, deplatforming may simply push the trouble to darker corners of the web. Telegram, a Russian-founded, alternative social media platform offering message encryption, protected speech and broadcast to masses of followers free from government interference offers a case in point (the platform itself is banned in Russia). Rogers deploys digital methods rationales and tools such as hyperlink analysis to study migration of the deplatformed from mainstream to alternative social media, and the newly built ‘Telegram scraper’ to extract via the platform’s API, content from the deplatformed celebrities’ public Telegram channels to study extremist discourses. While communicative activity remains, Rogers finds that audiences are thinning and language use has become milder.

In our age, populist movements are increasingly difficult to study with well-tested methods from media and communication research: these movements are often explicitly opposed to established media, and grasping their stand on issues, revealing their strategies, or understanding the discourses that mobilize supporters are therefore hard challenges. Approaching populist movements where they are – in social media – Cornelius Puschmann, Julian Ausserhofer and Josef Šlerka propose that digital trace data and computational methods offer new potentials for communication research. With a topic modelling analysis of the comments posted on the Facebook page of the Pegida movement in Germany to those of the Alternative for Germany

(AfD), Puschmann, Ausserhofer and Šlerka argue that in addition to making the activities and arguments of such new political actors visible and open for analysis, the resulting analysis brings out the small, gradual changes in their nativist agendas, and thereby document the dynamic nature of these actors.

In her contribution, Anat Ben-David enters the debate on methods for social media research after easy access to platforms' API for data harvesting got shut down. Offering a different approach to digital methods than the previous contributions, she uses archival theory and describes Facebook as a contemporary archon of public records, while at the same time being unarchivable. What we need, argues Ben-David, is counter archives. Counter archiving datafied platforms can make visible the knowledge production, as well as the limitations, of publicly available records, allowing for a critique of the epistemic affordances of the data that Facebook (or other social media) research relies upon. In her article, Ben-David illustrates the strategy of counter archiving by way of two examples: A public archive of the Israeli parliament on Facebook, 2015-2019, and a screenshot archive of political advertising from the 2019 Israeli election. Pointing to implications for the suggested strategy, she also discusses how such practices blurs the boundaries between archivists, activists and scholars.

Whilst Ben-David engages with the affordances of platforms as a way to question datafied knowledge production, Alice Mattoni, in her contribution, turns the lens to the effects of data in our social world by situating data practices in the wider communication repertoires of activists. Focusing on the interplay between big data and anti-corruption activism, Mattoni argues for the value of the long-standing social science tradition of grounded theory as a way to disentangle the complex communicative contexts in which data-enabled activism happens. Grounded theory, she argues, is particularly useful for studying emergent phenomena such as datafication and for selecting case studies in a way that is sensitive to different practices across diverse contexts. It also allows for situational analysis that can grapple with the multitude of human and non-human actors that make up the network surrounding different cases of data-enabled activism. This is particularly pertinent, Mattoni contends, for highlighting the local instances of a global phenomenon, rejecting universalism and instead elevating the particularities of experiences of datafication.

In their article on datawalking as a method for studying datafying infrastructures, Karin van Ess and Michiel de Lange develop a methodological reflection of datawalking as a novel way of producing knowledge about the temporally and spatially situated data-operations by which contemporary media technologies come to shape culture and society. Grounding their discussion in examples of infrastructure and the mediated, or "smart", city, and reaching into the disciplines of critical data studies and urban studies, the authors take datawalking to address the invisibility, decontextualization and (lack of) accessibility of data and infrastructures often assumed in scholarship on datafication. Datawalking, envisioned to work in tandem with other qualitative methods, is said to produce embodied, situated and generative knowledge of how current city-life is orchestrated by way of datafying infrastructures. In doing so, argues van Es and de Lange, datawalking may help groom critical consciousness and alternative narratives about the consequences of datafication and the pervasive presence of big tech in everyday life, including the power of smart technology in shaping human futures.

In the final contribution to this special issue, Rasmus Helles and Jakob Ørmen ask us to critically reflect on the continuities of long-standing approaches to explanation in the context of novel methodological innovations in the form of big data. Pushing back against the notion that big data has led to a new methodological paradigm in research, Helles and Ørmen instead argue that these new analytical techniques associated with big data should be seen as being productively integrated into existing explanatory strategies familiar in media and communication studies. Using neo-positivism, critical realism and interpretivism as illustrative examples of explanatory frameworks they outline different applications of big data techniques for media and communication research. In so doing, the paper pertinently stresses the distinct nature of scholarly pursuits that will need to critically reflect on the epistemic 'surplus' and 'debts' that come with appropriating tools from outside social science in order to formulate the different varieties of standards that we want to

apply to these models. This, Helles and Ørmen conclude, is a call to media and communication research that privileges the historical agility of the field to contend with methodological innovation, even in the form of 'new' big data.

Acknowledgements

The editors wish to thank Fieke Jansen, Lars Nyre, Dag Elgesem, Kristian Ahm, Philippe Maarek and Signe Sophus Lai for contributing book reviews and notes aligned with the theme of the special issue.

Stine Lomborg received funding from the Independent Research Fund Denmark (grant no. 8018-00113B). Lina Dencik is funded by a Starting Grant from the European Research Council (grant no. 759903). Hallvard Moe received funding from The Research Council of Norway (grant no. 247617).

References:

Boyd, D. & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5): 662-679, DOI: [10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878)

Bruns, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11): 1544-1566, DOI: [10.1080/1369118X.2019.1637447](https://doi.org/10.1080/1369118X.2019.1637447)

Bucher, T. (2012) Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media and Society*, 14(7): 1164–1180.

Bunz, M. & Meikle, G. (2018). *The internet of things*, Cambridge, Polity Press.

Castells, M. (2007). Communication, Power and Counter-power in the Network Society. *International Journal of Communication*, 1, 238-266.

Diakopoulos, N. (2015) Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3): 398-415.

Dencik, L., Hintz, A. & Cable, J. (2016), Towards Data Justice? The ambiguity of anti-surveillance resistance in political activism, *Big Data & Society*. DOI: 10.1177/2053951716679678

Dencik, L. (2019). Situating practices in datafication - from above and below. In: Stephansen, H. and Treré, E. (eds.) *Citizen Media and Practice*. London and New York: Routledge.

Gitelman, L. (2013). *"Raw Data" Is an Oxymoron*. Cambridge, MA, MIT Press.

Hintz, A., Dencik, L. & Wahl-Jorgensen, K. (2018). *Digital Citizenship in a Datafied Society*. Cambridge: Polity.

Jensen, K. B. (2010). *Media Convergence*. London & New York: Routledge.

Jensen, K. B. (2012). Lost, Found, and Made : Qualitative Data in the Study of Three-Step Flows of Communication (pp. 435-450). In: Volkmer, I. (ed.) : *The Handbook of Global Media Research*. Malden, MA: Wiley-Blackwell.

Jensen, K. B. & Helles, R. (2017). Speaking into the system : Social media and many-to-one communication. *European Journal of Communication*, 32(1): 16-25.

JGB, DM & KER (1986). Editorial. *European Journal of Communication*, 1(1): 3-6.

Katz, E. & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communication*. Glencoe, IL: Free Press.

Kennedy, H. (2018) Living with Data: Aligning Data Studies and Data Activism Through a Focus on Everyday Experiences of Datafication. *Krisis – Journal for Contemporary Philosophy*. Issue 1: 18-30

Mayer-Schonberger, V. & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, Houghton Mifflin Harback Publishing.

Livingstone, S. (2004) The Challenge of Changing Audiences. Or, What is the Audience Researcher to do in the Age of the Internet? *European Journal of Communication*, 19(1): 75-86

Livingstone, S. (2019). Audiences in an Age of Datafication: Critical Questions for Media Research. *Television & New Media*, 20(2): 170-183

Lomborg, S. & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, 30(4): 256-265.

Lomborg, S. & Frandsen, K. (2016). Self-tracking as communication. *Information, Communication & Society*, 9(7):1015-1027.

Mejias, U. A. & Couldry, N. (2019). Datafication. *Internet Policy Review*, 8(4). DOI: 10.14763/2019.4.1428

Picone, I (2017) Conceptualizing media users across media: The case for 'media user/use' as analytical concepts, *Convergence*, 23(4), 378–390. <https://doi.org/10.1177/1354856517700380>

Puschmann, C (2019) An end to the wild west of social media research: a response to Axel Bruns. *Information, Communication & Society*, 22(11): 1582-1589, DOI: 10.1080/1369118X.2019.1646300

Rogers, E. M. (1999). Anatomy of the two subdisciplines of communication study. *Human Communication Research*, 225(4): 618-631.

Rogers, R. (2013) *Digital Methods*. MIT Press.

Sandvig, C, Hamilton, K, Karahalios, K. (2014) Auditing algorithms: Research methods for detecting discrimination on Internet platforms. In: *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, Seattle, USA, 22 May 2014:. 1–23.

Stoycheff E, Lui J, Wibowo KA, et al. (2017) What Have We Learned about Social Media by Studying Facebook? A Decade in Review. *New Media & Society*, 19(6): 968–980.

Turow, J. & Couldry, N. 2018. Media as data extraction: towards a new map of a transformed communications field. *Journal of Communication*, 68: 415–423.

Van den Bulck, H., & Moe, H. (2018). Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. *Media, Culture & Society*, 40(6): 875–892. <https://doi.org/10.1177/0163443717734407>

Zamith, R. and Lewis, S.C. (2015) Content Analysis and the Algorithmic Coder: What Computational Social Science Means for Traditional Modes of Media Analysis. *The ANNALS of the American Academy of Political and Social Science*, 659(1): 307-318

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, Public Affairs.