



# Media and Communication Studies in the Age of Digitalization and Datafication: How Practical Factors and Research Interests Determine Methodological Choices

*Dennis Nguyen*

## INTRODUCTION

The terms “digital humanities” and “computational turn” have been a buzz in academia for well over a decade. Their premise is to fundamentally transform how research is conducted in the wider field of the humanities and social sciences, including the interdisciplinary area of media and communication studies. With the rise of a data-focused research mindset that sees digital technologies not only as subjects but more increasingly as research tools, one may assume that there will be an accelerated blending of the “soft” humanities and social sciences with the “hard” data-driven natural sciences. Thanks to more powerful computing devices that can process large amounts of—seemingly—widely available data in the form of, for example, cultural analytics and computational social sciences

---

D. Nguyen (✉)

HU University of Applied Sciences Utrecht, Utrecht, The Netherlands

e-mail: [dennis.nguyen@hu.nl](mailto:dennis.nguyen@hu.nl)

© The Author(s) 2020

D. Nguyen et al. (eds.), *Understanding Media  
and Society in the Age of Digitalisation*,

[https://doi.org/10.1007/978-3-030-38577-4\\_2](https://doi.org/10.1007/978-3-030-38577-4_2)

(Manovich 2018), scholars could overcome the biases and limitations of small samples and move beyond theorization. Eventually, the humanities could explain social reality in almost the same way as the natural sciences do for physical reality.

On one hand, there is an unprecedented volume of data available about cultural activity and social interactions; the never-ending streams of images, videos, memes, and comments that exploded with the emergence of the Web 2.0 one and a half decades ago speak for themselves. Global digital society experiences an era of “digital plentitude” (Bolter 2019): the emergence of multi-focal media culture that defies modernist, normative assumptions about the same and asks critical observers to rethink concepts of culture, public value, pluralism, and democracy. On the other hand, with the emergence of digital culture and digital society run on digital technology, both inevitably became complex assemblages of data that can be subjected to software tools designed for specific research purposes (Rogers 2019). It is widely recognized that global digital society experiences a continuing and accelerating datafication that seems to reach into every corner of private and public spheres. While the cultural, social, political, and legal contexts of national spaces matter in the manifestation of these transformations, they indeed happen on a global scale as they are triggered by tech-entrepreneurs from Silicon Valley, USA, to Shenzhen, China.

The surge of data generation over the past decade is the most obvious indicator for this development. The enormous volume of cultural and social data paired with computational methods may allow academics to overcome the gap between the quantitative and qualitative research traditions and completely change the outlook on how researchers study culture—not least since they have now access to cultural and social data about interactions, relationships, and artefacts that may have been invisible in the past (Manovich 2018). However, whether this potential can be fully realized and, more fundamentally, if it is desirable to turn the qualitative into the quantitative in the context of datafication and digitalization are open to critical debate.

Not to mention that vast volumes of these data are in the hands of private platforms.

The present chapter reviews the discussion about computerizing and digitalizing media and communication studies. It provides an overview of relevant developments, what these developments mean for research practices in a transforming academic environment, and the very practical

implications that arise. The author's main goals are to (a) briefly outline the rise of data-driven approaches as a foundation of digital humanities and computational social sciences; (b) identify what the main challenges are for actually implementing computational methods in the field of media and communication studies (this includes a brief review of the quantitative-qualitative dilemma plaguing, especially, the interdisciplinary area of media and communication studies); and (c) examine what these challenges mean for practising media and communication studies on the fundamentally subjective human experiences in the age of digitalization. The main questions are: how should researchers in media and communication studies justify their methodological choices and how will traditional and new methods coexist? What practical actions can they take to underline the relevance and validity of their approaches?

Here, *media and communication studies* are defined as the collection of disciplines with a focus on media technology, media formats, media consumption, media content, media effects, media economy, media culture, media ecology, media recognition, media politics and political communication, and media psychology. They have their roots in cultural studies, including film and theatre studies, the social sciences, cognitive psychology, and linguistics (Hansen and Machin 2013). Media and communication studies have diverse roots and research interests as well as different preferences when it comes to what are commonly designated as quantitative and qualitative research methods. While media studies is usually more “qualitative” due to their strong links to the traditional humanities, communication studies more willingly embrace the “quantitative” methods due to their links to the empirical social sciences. However, both areas focus on similar research subjects from different angles and have a propensity to stimulate work within mixed methods research designs, not least due to the obscure borders between media and communication studies.

In considering the relationship between humanities and datafication, Manovich basically reframes both humanities and social sciences, seen here as the parent disciplines of media and communication studies, by separating digital humanities, represented by media studies, from social computing, represented by communication studies (Manovich 2018). The broader field of media and communications studies is profoundly affected by the rise of computerized methods and thus is a primary example for trends and challenges in the emerging fields of social computing and digital humanities. Furthermore, to ensure consistency, the term “method”

designates actual tools and instruments for data collection and analysis, while “methodology” defines more of a philosophical stance for how to view the world, what data say about it, and how to create knowledge.

In essence, disciplines associated with the humanities and social sciences in the early twenty-first century find themselves at multiple cross-roads concerning their ontological and epistemological foundations, their perceived value by the scientific community and public at large, and a potential fragmentation into very different research programmes within these general fields. This fragmentation is primarily driven by political and economic factors that determine how research, and teaching, is conducted at many academic institutions. Furthermore, the Internet and the Web are both subjects in virtually all academic disciplines and not restricted to just one and studying them demands an inherently interdisciplinary outlook. Since both change the way society produces, uses, and consumes media, questions related to studying them are of heightened relevance for the field of media and communication studies. Schäfer and van Es argue that:

Datafication and computerization will come to affect all research agendas and inform the skill sets of students and scholars alike. We predict that the term ‘digital humanities’ will sound increasingly pleonastic and will eventually disappear—it will lead not to the replacement of established methods in the humanities, but rather to an expansion in the curricula we study and the methods we use. (2017, 15)

Methods associated with this trend rely mainly on digital data native to the Web (Rogers 2019) and the development and application of software tools to computerize data collection and analysis, that is, using computer code on digital data about human social and cultural activity. This chapter looks at some of the implications of these trends for media and communication studies. The goal is not to discredit the value of digital data and digital methods; to the contrary, it is the author’s firm belief that digital data-focused approaches offer new perspectives beyond traditional research methods, both qualitative and quantitative, and can broaden our understanding of culture and society. However, it is an essential and legitimate question to ask what is actually doable for many researchers in the area and when computational methods based on innovative digital research strategies and/or large data sets hit their limitations for understanding how humans interact with each other, and with technology, in contexts relevant for media and communication studies.

The discussion begins with a review of the rise of data-driven, digital research strategies before the chapter addresses the main challenges in implementing them. It closes with some general recommendations for pragmatic research agendas in media and communication studies.

### RISE OF DATA-DRIVEN APPROACHES

The “digital” has always been an important subject of media and communication studies. Early critical research in the late 1980s and 1990s explored what the rise of networked computer-based communication could mean for the very nature of democracy and societal organization. Media studies scholars theorized and investigated how digital communication technology may transform the performance of identity and the formation of communities, change the power play of politics, and reconfigure how society “does” culture.

Examples are plentiful: a key text in early cyberculture studies is Donna Haraway’s *Cyborg Manifesto* (1984), which continues to influence views on human-tech interaction to this day. Critical researchers in the field of political communication built robust theoretical frameworks to explain cyberconflicts (Karatzogianni 2006) based on previous scholarship on power, networks, and social organization in sociology, philosophy, and cultural studies. Derrida, Foucault, Deleuze, Guattari, and the like retain their influence and relevance in the still somewhat nascent and evolving field of Internet Studies. For example, Benjamin Bratton’s *The Stack* (2015) is virtually inaccessible without solid knowledge about previous work on power, structure, and sovereignty as conceptualized by thinkers such as Foucault. Jodi Dean (2010) and Geert Lovink (2008, 2019) provided some of the most profound critiques on digital culture and continue to offer equally insightful and provocative analyses. Henry Jenkins (2006) and Jenkins et al. (2015) built a career on theorizing and analysing how digital technology led to the emergence of conversion culture and participatory culture. Research also explored the digital transformation of journalism and public discourse (Papacharissi 2014; Bruns 2017; Nguyen 2017; Ferra 2019). Other studies explore how trends in decentralization, artificial intelligence, and increasing datafication change our culture and society (e.g. Bunz and Meikle 2018; McStay 2018).

Since digital culture relies on digital media, it was only logical that anything related to the Internet and Web became a research subject from a media and communication studies perspective. When conducting their

studies, researchers usually applied what they were familiar with and many continue to do so: for example, the established qualitative and quantitative methods from media studies, such as close reading, ethnography, or critical discourse analysis, versus the more social science-oriented communication studies, such as surveys, policy analyses, or “quantitative” content analyses (Krippendorff 2019; Hansen and Machin 2013), as well as methods associated with media psychology (Giles 2009; Dill 2013) such as experiments. They all yield valuable insights into how technology shapes processes of communication and thus forms of socialization, media consumption, negotiating and performing culture, representation, and participation.

What has changed over the past decade is not that digital issues and trends suddenly emerged as interesting research subjects. However, there is an intensive, ongoing debate about methods and methodologies—that is, what data should be considered for relevant analyses, and how to conduct such analyses in media and communication studies. While the discussion is not entirely new and closely connected to the perceived gap and difficult relationship between quantitative and qualitative research methodologies, it regained topicality with the increasing datafication of society and culture in the post-Web 2.0 world. In short, proponents of a computational turn in the field of media and communication studies argue that with digital data the study of society and culture can uncover previously inaccessible territory in the process of knowledge production. With the digitalization of media formats and contents as well as of virtually all communication processes, the sheer volume of what can be potentially discovered from large swaths of digital data about social, cultural, political, and economic interactions holds promise for new types of truly data-focused analyses also in previously rather confined areas of qualitative research and critical reflection.

Manovich argues that computational, data-driven research designs in the digital humanities can broaden our understanding of who actually performs culture beyond an elitist definition: ‘Cultural Analytics is interested in everything created by everybody. In this, we are approaching culture the way linguists study languages or biologists who study the life on earth. Ideally, we want to look at every cultural manifestation, rather than selective samples’ (Manovich 2015, 7). While this aspiration is not new—especially in media studies, which already over two decades ago included a broad spectrum of genres of music, film, and TV into a society’s cultural output—in the age of digital methods this reaches a new dimension. In

the 1990s, researchers argued why the Simpsons are as much an insightful cultural artefact to study as is Goethe's Faust; now we have to add the output of millions of "prosumers" (Bruns 2008), that is, user-generated content in the form of, for example, memes, mods, fanfiction, and trending hashtags. Jay Bolter explains that:

[d]igital media did not by any means cause the decline of elite culture, which began before the computer had developed as a medium—that is, as a widely shared platform for expression and communication. But digital media now provide an ideal environment for our flattened, or perhaps we should say lumpy, media culture in which there are many focal points but no single center. This multiplicity, this loss of center, is not a "problem" to be solved. It is simply the condition of our culture today. (Bolter 2019, 2)

Leading advocates for a digital turn in the humanities such as Rogers (2019) and Manovich (2018) point to the potentials of analysing large volumes of data, or at least of tapping into new types of digital data, with the help of software for gathering new insights on previously inaccessible social and cultural phenomena. There is little to dispute about their general logic as a plethora of studies illustrate the inherent creativity and productivity of computerized research in the humanities and social sciences. Digital methods as proposed by, for example, the "Digital Methods Initiative" try to make use of this aspect of the broader digital transformation that modern global society is experiencing with new research strategies and tools that analyse data native to Web-based, inherently digital human activity (Rogers 2019). This needs to include a critical reflection on how the different platforms, apps, and sites shape that activity through their designs and what data they actually collect for what purposes with the algorithms they deploy (Rieder et al. 2018).

With larger amounts of digital data about behaviour and cultural outputs relevant for media and communication studies, the "traditional" methodologies and blueprints for data analysis may no longer be solely suitable for the task of deriving meaningful conclusions. While it cannot be claimed that the catalogue of established methods in, for example, the traditionally more qualitative media studies has become completely obsolete, there is a growing trend of applying computational methods, which themselves are in no small extent firmly grounded in quantitative research methods, to uncover patterns in data sets about cultural phenomena that can have millions of data points (Manovich 2018).

The rapidly growing volume of data opens up new opportunities to conduct research. For some, it holds the promise to overcome the limitations and subjectivity of soft disciplines such as sociology, anthropology, and cultural studies, all of which influenced the foundations of modern media and communication studies. The emergence of big data inevitably includes cultural and social phenomena, as the ubiquity of social networking media shows. Especially social networking media's popularity inspired and stimulated advances in embracing digital data and methods into the toolkit across almost all disciplines associated with the humanities and social sciences. During the hype of application programming interfaces (APIs) that started around the early 2010s, researchers experimented with different, often self-programmed, purpose-specific data collection and data analysis tools that harvested data from Facebook, Instagram, Twitter, and so on, and visualized their usually descriptive findings. Word clouds, timelines, heat maps, network graphs, and the like that process digital data made their entrance into the field of media and communication studies as well, with the promise to move beyond the limitations of small samples and human coding.

Marketing businesses and the big tech platforms themselves exploit data from social networking media with algorithmic, partly autonomous research methods on big data sets, which are in essence forms of content analysis and experiments for specific commercial goals. Using methods from the social sciences in commercial contexts is nothing new, as surveys, interviews, focus groups, and media resonance analyses have a firm place in business-oriented, applied research. However, in the current platform economy, the gap between data-haves and have-nots has grown larger and the resulting hierarchy of access to insights has become static. More recently, machine learning and narrow artificial intelligence have emerged as another set of computational methods that can profoundly change how data are collected, analysed, and interpreted in the field, while smart devices related to the Internet of Things continue to expand the volume and variety of data. Media and communication researchers deal with an increasing spectrum of communication technologies that go far beyond modern iterations of broadcasting models and multimedia but include a growing number of smart technologies that use sensors to see, hear, feel, and interpret human activities (Bunz and Meikle 2018).

Yet despite the tendencies of digitalizing, quantifying, and therefore "scientifying" traditional academia, there are various political and practical obstacles that put clear and strong limitations to computerization and



digitalization of the humanities and social sciences. Due to the huge economic interests tied to data, access is often fairly limited from the outset; a lot of cultural and social activity takes place on private platforms that have limited interest in openly sharing the data they collect in the current framework of the global data economy. Tech companies such as Facebook, Amazon, and Apple collect data primarily for product and service development as well as a commodity sold to companies in the marketing and advertising sector (Zuboff 2019).

Additionally, frequent scandals concerning privacy violations and perceived misuse of data for commercial and political goals, as seen in the Cambridge Analytica episode involving Facebook, put further limitations on access to data. Tools that only a few years ago allowed to collect and analyse data from social networking media have either lost access to previously much richer data or are basically defunct today (e.g. NetVizz). Even more fundamentally, it is also a contentious question how representative and accurate the data on many of the prevailing online platforms are, that is, what their quality truly is. Not to mention, not all researchers in the media and communication studies display the same openness towards digital methods. This can be for different personal, practical, and research interest-related reasons.

## MAIN CHALLENGES OF COMPUTERIZING AND DIGITALIZING MEDIA AND COMMUNICATION STUDIES

The major challenges related to trends of computerizing and digitalizing research in the media and communication studies mainly derive from practical, ontological, epistemological, and ethical issues. In sum, these factors have a profound effect on shaping the actual application of research methods in studies on media and communication subjects.

### *Practical Restriction in Skills and Resources*

Starting with one of the most common and obvious roadblocks: acquiring the necessary skills in developing and applying tools for digital data collection and analysis is a time-consuming process. It is simply not feasible for all researchers in the humanities and social sciences to expand their skillset in this respect. Learning how to code even a rather simple API-based tool for collecting and visualizing data is not something that can be easily

achieved on the fly. While relatively accessible programming languages such as R or Python have become more popular in recent years, it takes more than just an introductory course before coding skills can supplement and enhance the toolset of methods and methodological outlook of individual researchers and departments. It costs time, money, and human resources that may not always be available in sufficient quantity. A misplaced emphasis on computational methods may also undermine existing expertise and falsely position coding skills as an omnipotent solution to eventually all research problems. Efforts to increase data literacy among all scholars need to be intensified in this respect and room for acquiring and experimenting with new methods to develop new methodologies should be made—while acknowledging the value of non-digital research methods where it is suitable.

Nevertheless, all of this becomes a rather difficult issue in light of the pressures of the neoliberal university. Another major challenge for the field, in general, is what could be described as a potentially growing imbalance between relevant sub-disciplines and departments: some may embrace digital methods quicker and more efficiently than others; some departments are in a better position, in terms of staff, outlook, and resources, to train researchers and/or recruit experts. This may further drift research and educational institutes away from each other in terms of research output, innovation, and thus influence and power. In extreme cases this can also lead to seemingly unrelated disciplines contesting the traditional areas of expertise of others, for example, computer scientists challenging scholars of literature (e.g. Congjing et al. 2017). Does that mean non-digital research activities are inferior?

Often, researchers do not have any other choice than to use traditional methods for practical reasons related to resources and access. For individual researchers it is recommendable to keep an eye on the output of computer science-related disciplines that may conduct studies in their domains and forge networks for collaborations across research cultures. Investing time in learning coding is another recommendation, not so much to only build tools but to be able to scrutinize digital phenomena and research conducted from a strong computational angle. On an institutional level, research coordinators should increase efforts in stimulating cross-departmental cooperation for the same effect from a more strategic angle to delimit intellectual isolation and silo-forming. While not every researcher can and should become a coder, a basic understanding of coding for media and communications researchers, and computer engineers

learning about domain specific theories, methods, and empirical insights, may be due.

### *Practical Restrictions to Access*

This is one of the most apparent limitations and one of the fundamental contradictions of the age of digitalization and datafication: while data are created on an unprecedented scale, access to the same has become increasingly limited for various business-related and legal reasons. While it was possible to create networks of social relations from, for example, Facebook data just a few years ago, increased competition between data companies and necessary regulation put considerable limitations on how much data can be retrieved from the major social networking media platforms. As a consequence, private companies and some wealthy universities with sufficient financial means not only can conduct research on a wider scale but often have exclusive access to rich data due to the fact that they either provide the services and devices that collect data about human behaviours, including social and cultural activities, or can purchase large data sets from data-generating companies.

The political economy of the Internet and the profound consequences of platformization have been critically analysed in various recent studies (Smyrnaioi 2018; Zuboff 2019; Couldry and Mejias 2019; Srincek 2017). Restrictions to APIs and limited access to the big platform's data domains may form one of the main reasons for a return of "classic" research methods and a rethinking of digital methods beyond "scraping" large data volumes, researchers may rather explore how they can exploit the functionality/modus operandi of online platforms within the user-centric interfaces instead of getting the data via means of coding as in programming per se.

### *Challenges Deriving from the "Quantitative-Qualitative Divide"*

Additionally, some scholars' interests and research problems may not fit a heavily data-driven approach; there are questions that go beyond descriptive and correlational analyses with large data sets. The still unresolved tension between qualitative and quantitative research methods further affects how media and communication studies are being performed. The motivations *for* and problems *of* trying to make the humanities more empirical are not new but regain urgency in the discussion about digital humanities:

Instead, they approach their objects of study from interpretive and critical perspectives, acting in the assumption that in doing so they necessarily also pre-constitute them. However, with the introduction of digital research tools, and tools for data research specifically, humanistic scholarship seems to be increasingly indebted to positivist traditions. For one, this is because those tools, more often than not, are borrowed from disciplines centred on the analysis of empirical, usually quantitative data. Inevitably, then, they incorporate the epistemic traditions they derive from. (Masson 2017, 25)

The credo that data, quantification, and measurement are absolutely necessary for reaching genuine understanding and knowledge has become something of a new dogma in global digital society. The processes behind creating knowledge have always been subject to trends and discussions in the media and communication studies. More often than not, this involved the tensions and questions of compatibility between what are commonly referred to as quantitative and qualitative research methodologies. The discussion between quantitative research and qualitative research traditions has a long history, with the most extreme clashes between positivist scientists and postmodern scholars even making their way into public media discourses in the 1990s, as seen in the infamous “science wars” episode. In less spectacular but equally profound respect, German academia experienced its own dispute over methods, or *Methodenstreit*, in the 1960s (Rieder and Röhle 2017).

While there are various ontological and epistemological arguments on each side to discredit the other, it would be naïve to ignore also political and personal factors; researchers who are adamant that only their understanding of data, methodology, and eventually knowledge generation matters can also have personal motivations, including the unwillingness or inability to actually understand other ontological and epistemological outlooks and make use of the other camps’ methods. In short, there are quantitative researchers who shy away from diving deeper into small samples, for example, in the form of heavily theorizing and critically reading media texts, and there are qualitative researchers who are, in blunt terms, simply afraid of numbers in any form. Although no large, systematic study is available on this, there is often plenty of anecdotal evidence that most researchers can relate to and recognize. Silo-thinking and comfortable seclusion are important factors that maintained the separation between research traditions and thus ensured its continued existence. That does not mean that there are no real arguments for keeping both perspectives relevant and

separate on ontological and epistemological grounds (Sale et al. 2002). Indeed, qualitative research assumes a constructivist worldview while quantitative research is often based on the idea of an objectively accessible reality that can be measured.

Inevitable problems that qualitative researchers face are that their findings might be severely delimited to the few cases they investigated and that they increase bias in the research process through the methods they employ. While quantitative researchers can usually build on a more solid empirical foundation for their observations and subsequent claims, they face their own set of challenges: sampling always remains an issue, even if  $N$  equals thousands, if not hundreds of thousands, of cases. The misconception that numeric data are free of bias has been frequently refuted. Turning things into data is a subjective process, which undermines the alleged objectivity of data—it matters when, how, and in which form something is counted, regardless of the data volume (Kitchin 2014; Loukissas 2019). Quantitative research can also not happen in a vacuum: it should build on strong theory, reflect on and contextualize findings, and inform theory building or improve existing theories.

With the methodological school for mixed methods, there is a third perspective that addresses the perceived gap between quantitative and qualitative research before the popularization of digital methods. Teddlie and Tashakkori define mixed methods as the third research community that is ‘working primarily within the pragmatist paradigm and interested in both narrative and numeric data and their analyses’ (2009, 4). However, it also faces its own challenges and is not easily achieved by merely including different methods within one larger research design without indicating a clear and meaningful connection. Qualitative and quantitative research have always dealt with different ontological and epistemological outlooks on the world that make it difficult to combine the two (Sale et al. 2002). Meaningful integration that respects both paradigms and their underlying assumptions, constructivist versus positivist, can only take the form of complementary, additive research designs in which each method maybe conducted in parallel or sequentially (Sale et al. 2002, 8). The integration of different methodologies needs to take place in carefully designed frameworks and motivated by precise research questions (Plowright 2011). This also means that digital methods and computational research need to allocate themselves between the two camps and justify how they can build bridges between them.

What all research traditions share is that they base their arguments and eventually theories on specific data types, regardless of whether focus is placed on qualitative or quantitative methods. Data can be defined as: ‘the raw material produced by abstracting the world into categories, measures and other representational forms—numbers, characters, symbols, images, sounds, electromagnetic waves, bits—that constitute the building blocks from which information and knowledge are created’ (Kitchin 2014, 1). It might be more accurate then to rather speak of hypothesis-generating, that is, qualitative, and hypothesis-testing, that is quantitative, methods that ideally complement each other (Nguyen 2017, 136, citing Terfrüchte 2011). This reframing may contribute to narrowing the ideological gap between the two schools of thought and the emerging field of digital methods could be the right place to show the productivity of this conceptualization.

Researchers who favour qualitative research methods need to clearly explain how the insights they won can inform further research to test their proposed hypotheses and theories; quantitative researchers, which include those who conduct research on larger digital data sets, need to address the context in which the data they analyse were generated and should not view it as an objective representation of reality but an abstraction of human interactions in specific contexts that are subject to historical contingencies. Media and communications researchers need to acknowledge that media are consumed and used on a large scale, implying a need for quantitative research, but eventually become a very individual, subjective experience, which indicates a need for qualitative research (Baumann and Scherer 2012). Ignoring one side over the other will delimit understanding research subjects relevant for the field.

### *Practical Challenges in Bridging Theory and Digital Empiricism*

Strongly connected to the debate about qualitative, quantitative, and mixed methods research is the question of how to combine new data-driven research approaches with existing theoretical frameworks and insights won with “non”-digital research methods, which can be qualitative and quantitative in nature (or combine both). One issue with the rise of digital methods and data-focused research perspectives is the potential neglect of building strong theoretical frameworks—or even to draw from existing ones (Chandler 2015).

In the efforts to build new tools to tap into new data pools for understanding digital society and culture, theories relevant to specific contexts for social interaction and cultural activity can take a backseat in favour of empirical findings alone. For example, studies that make use of data collection tools to discuss fake news or memes have the strength of numbers but may not always connect it to useful theories and insights from, for example, journalism studies and framing research. While data-driven schools of thought have a strong argument when claiming that their insights are based on considerable data volumes and that their analyses can reveal statistically significant correlations, which are indeed absolutely necessary for deriving empirically grounded generalizations, ignoring already existing strong theory and letting the data “speak for themselves” is at the same time one of their potential weaknesses. Couldry and Mejjas (2019) argue that the current situation shows the need for strong social theory, first and foremost to make sense of how datafication changes society with its detrimental effects for democracy and individual freedom.

However, one could also argue for bringing theory back into the discussion about digital humanities and social computing in the field of media and communication studies. It is a curious situation: a lot of research explores how new ways to collect and analyse data can change our horizon in studying culture and society, which is good, as testing previous hypotheses and theories can help with filtering out the non-useful from the useful and to revise and critically check previous assumptions. However, it would be a folly to completely disregard all pre-digital humanities scholarship in favour of digital data-focused research only. In fact, traditional methods may be applied to complement and scrutinize the findings of digital methods and vice versa.

### *Challenges Related to the Quality of Data*

Data are a construct that is open to interpretation and trends and determined by business and political interests; this has always been the reality of data, but it needs renewed emphasis especially in the context of the data economy (Kitchin 2014). As with all statistical research, correlation is not causation and when it comes to ethically difficult issues such as media representation and participation, asking why something is happening for historical and cultural reasons is an important aspect for deriving meaningful conclusions. Researchers who deal with issues of, for example, media representation, identity performance, and digital activism should try to

combine the strengths of computational and digital methods with the theoretical insights relevant for the domain they are investigating.

More data are not necessarily better, and aside from ethical issues discussed below, it is also ontologically and epistemically important to critically scrutinize what a large data set actually represents before any conclusions are drawn from the analysis. Data, especially about humans, are never without context, that is, political, cultural, and social implications. How big data change research has been at the centre of important debates in academia with proponents and critics agreeing over the use of data and the implications for research but disagreeing over the actual extent of value for gaining unbiased knowledge; some claim data speak for themselves from an extremist positivist perspective (Mayer-Schönberger and Cukier 2013) while others point to clear limitations and ethical issues (boyd and Crawford 2012). Critical questions should be raised about what the data actually represent, what is not included, and therefore what the limitations of the insights won from them are. This goes for traditional and new research methodologies in equal measure.

### *Ethical Issues*

The ethical issues or, to be more precise, the intentional or unintentional lack of and violation of ethical considerations are an important issue in the academic discourse about computational and digital methods in research and the private sector: ‘unless used very carefully, data science can actually perpetuate and reinforce prejudice’ (Kelleher and Tierney 2018, 193). For researchers, the ethical questions usually do not concern the commercial exploitation of data but rather relate to questions of consent, access, replicability, and transparency in general. Strong ethical stances question the wisdom of, for example, simply analysing social networking media content, that is, mainly user messages and interactions, due to their seemingly public availability without informing every account behind a post and metric that their data are subjected to analysis for research output.

While these could be seen as an extreme position, it is true that simply anonymizing data may not suffice to guarantee that individuals cannot be traced back through the content of messages that were subjected to a form of, for example, content analysis. How data are collected, aggregated, and depersonalized should be clearly explained in any method section of a research paper. Each media and communications department should have their own ethics guidelines and check where they need an update.



Traditional qualitative and quantitative research methods have strong ethical frameworks that may serve as an inspiration for covering new research frontiers involving digital media technologies that collect data from new angles with new technologies, especially in regard to the emerging Internet of Things and the increasing datafication of daily life (Bunz and Meikle 2018; Couldry and Mejias 2019).

### OUTLOOK: HOW TO PRACTISE MEDIA AND COMMUNICATION STUDIES?

It is a difficult situation: on the one hand, it is hard to deny that there are clear limitations to the insights won with traditional scholarship in the media and communication studies when it comes to overcoming subjectivity, the potential for generalization, and informing policy development. On the other hand, data-driven approaches are not necessarily less subjective and come with their own biases. Furthermore, not all researchers in the field have the skills, tools, or access to the relevant resources needed to conduct genuinely digital, computational research on media and communication research subjects as proposed by highly influential pioneers.

Another conclusion to draw here is that descriptive analyses alone do not suffice to add valuable knowledge to the academic discourse, regardless of how novel or even large the data collection. Ignoring the historical, political, social, and cultural formations that contribute to the emergence of and interaction with a communication technology, the content it carries, and the relationships it enables would be a potentially fatal blind spot in an inflexible data-driven methodology that induces its insights only from the immediate observations it makes from a source of digital data. This point has been repeatedly raised in the discussion about digital humanities (Schäfer and van Es 2017; Rieder and Röhle 2017) but cannot be emphasized enough in an age where the claim that finding correlation beats understanding causation is propagated.

Now, how should researchers in media and communication studies justify their methodological choices and how will traditional and new methods coexist? It is generally not a problem if individual researchers build up a specific expertise in methods that either serve hypothesis-generating (qualitative) or hypothesis-testing (quantitative); it is not an issue per se to have a methodological preference, i.e. to be better either with the critical reading of small samples or with the statistical analysis of larger data sets.

But experts for both should work more closely together in truly interdisciplinary research designs to make use of the strengths of both worlds. For example, in research on frames and framing in public discourses, big data-driven, computerized approaches can help with categorizing large sets of data for pre-clustering content from news text and/or social networking media in the form of topic modelling (Hekman 2018) for critical revision and finalizing with human coding. There is work in research that can be easily and usefully automated. Observations in interviews and experiments with respondents about framing effects could then be added to further expand our understanding of how framing in content may or may not influence individual opinions. Each methodology's assumption about what constitutes reality and the value of data needs to be clearly outlined and reconciled with the logic of its counterpart(s) within a mixed methods design. This problem has been repeatedly addressed in discussions about mixed methods and should now include also advances in computerizing and digitalizing media and communications research.

What practical actions can researchers take to underline the relevance of their approaches? The field of media and communication studies will have to accept that traditional methods will stay for the foreseeable future. The reasons for this are diverse: many research questions cannot be addressed with computational and/or digital research designs per se. For example, while flawed and limited in their own ways, both interviews and surveys allow researchers to collect specific data that need to be lured out from respondents as they would otherwise remain inaccessible. Advances in natural language processing may allow to analyse enormous data sets of texts, but complex discursive constructions related to framing of issues and people may rely for now on human interpretation in the data collection and analysis process. Ideally, universities do not only offer training in coding classes to at least inspire researchers to explore their own paths into the field of computational methods in the digital humanities; they also provide the context in which sensible and fruitful collaborations between, for example, computer sciences and social sciences and cultural studies can thrive. On an individual level, researchers need to learn to understand each other while maintaining a strong expertise in specific theories and methods; it is not only helpful but increasingly important if people understand the basics of code, since code shapes the phenomena many researchers are interested in and passionate about.

Research steps currently singled out as associated with the computational turn and digital humanities may ideally become simply part of the

logic of established methods, enhancing and complementing rather than replacing them. However, this poses practical challenges in regard to the extent to which researchers can, should, and are willing to be trained in novel methodologies. Differences in skills and interest may preserve old-fashioned approaches for much longer than some would expect. This does not mean that the respective research output has limited value from the outset but forces researchers to clearly justify their choices, outline the strengths and weaknesses, and explain why other methods would not have been suitable to address the research problem in focus. This means that researchers should at least check if novel, computerized methods could substitute or supplement their research efforts, and if not, why that is so. The point has been made elsewhere eloquently but is worth repeating (Rogers 2019; Manovich 2018; Schäfer and van Es 2017): the methodologies associated with the digital humanities and computational turn in the humanities and social sciences offer great potential to analyse society and culture on a different scale and with less direct intrusion by the researcher but has its own limitations and ethical issues to address; it is indeed very insightful to use, for example, algorithmic solutions to find patterns in cultural products or large social/cultural data sets, but the questions to answer then are related to how the algorithms work and who actually participates in online activities in the first place. As with any other method, it is vital to soberly assess the strengths and weaknesses.

The alleged superiority of a worldview that embraces an uncompromisingly empirical-positivistic outlook and exclusively prioritizes quantitative research with the overall goal of reaching ever more accurate predictive models has long left the confinements of academia and even business and industry but also invaded cultural and social life—not least with the ubiquity of personalized devices and AI-driven software that can be found in billions of people's hands on an almost 24/7 basis. Geert Lovink (2019) reminds us that focusing solely on data that come from private platforms designed to extract profits from users and not to enhance their well-being may distract us from asking fundamental questions: what place media technology actually has in society, what value it should add, and when its application in the currently prevailing economic logic is rather harmful than beneficial.

These critical questions have always been at the centre of media and communication studies regardless of where exactly they allocate themselves on the spectrum between qualitative and quantitative inquiry. While each researcher needs to decide and justify why a specific path was chosen,

the rise of digital methods does not yet initiate an end to traditional research along the qualitative-quantitative spectrum. Much rather, for practical and research-interested based reasons, different methodologies continue to coexist and should explore when and how to complement each other. This also means that questions of reliability, validity, transparency, and ethics stay as relevant as ever.

## REFERENCES

- Baumann, E., & Scherer, H. (2012). Wider das Ideologische. Plädoyer für den Pragmatismus in der Kombination qualitativer und quantitativer Methoden in der Kommunikationswissenschaft. In W. Loose & A. Scholl (Eds.), *Methodenkombinationen in der Kommunikationswissenschaft. Methodologische Herausforderungen und empirische Praxis*. Köln: Herbert von Halem Verlag.
- Bolter, J. (2019). *Digital Plenitude. The Decline of Elite Culture and the Rise of New Media*. Cambridge and London: MIT Press.
- boyd, d., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679.
- Bratton, B. (2015). *The Stack*. Cambridge: MIT Press.
- Bruns, A. (2008). *Blogs, Wikipedia, Second Life and Beyond. From Production to Producersage*. New York: Peter Lang.
- Bruns, A. (2017). *Gatewatching and News Curation. Journalism, Social Media, and the Public Sphere*. New York: Peter Lang.
- Bunz, M., & Meikle, G. (2018). *The Internet of Things*. London: Polity.
- Chandler, D. (2015). A World Without Causation: Big Data and the Coming of Age of Posthumanism. *Millennium: Journal of International Studies*, 43, 833–851.
- Congjing, R., Xinlia, L., & Haiying, H. (2017). ‘Visualization Analysis of Shakespeare Based on Big Data’. Conference Proceedings for *International Conference on Advances in Big Data Analytics* ABDA. Online retrievable at <https://csce.ucmss.com/books/LFS/CSREA2017/ABD3216.pdf> (accessed 16/03/2020).
- Couldry, N., & Mejias, U. A. (2019). *The Costs of Connection. How Data is Colonizing Human Life and Appropriating it for Capitalism*. Stanford: Stanford University Press.
- Dean, J. (2010). *Blog Theory. Feedback and Capture in the Circuits of Drive*. Cambridge and Malden: Polity Press.
- Dill, K. E. (ed) (2013). *The Oxford Handbook of Media Psychology*. Oxford: Oxford University Press.
- Ferra, I. (2019). *Digital Media and the Greek Crisis: Cyberconflicts, Discourses and Networks*. London: Emerald Publishing.

- Giles, D. (2009). *Psychology of the Media*. London: Palgrave.
- Hansen, A., & Machin, D. (2013). *Media & Communication Research Methods*. London: Palgrave Macmillan.
- Haraway, D. (1984). *A Cyborg Manifesto*. Cambridge: Macat.
- Hekman, E. (2018). *Public Service Media and Networked Publics on Twitter. A Big Data Approach*. Conference paper presented at RIPE@2018, Madrid, Spain.
- Jenkins, H. (2006). *Convergence Culture. Where Old and New Media Collide*. New York: NYU Press.
- Jenkins, H., Ito, M., & boyd, d. (Eds.). (2015). *Participatory Culture in a Networked Era. A Conversation on Youth, Commerce, and Politics*. John Wiley & Sons.
- Karatzogianni, A. (2006). *The Politics of Cyberconflict*. London: Routledge.
- Kelleher, J. D., & Tierney, B. (2018). *Data Science*. Cambridge, MA: MIT.
- Kitchin, R. (2014). *The Data Revolution. Big Data, Open Data, Data Infrastructure and Their Consequences*. London: Sage.
- Krippendorff, K. (2019). *Content Analysis. An Introduction to Its Methodology* (4th ed.). London: Sage.
- Loukissas, Y. A. (2019). *All Data Are Local. Thinking Critically in a Data-Driven Society*. MIT Press.
- Lovink, G. (2008). *Zero Comments. Blogging and Critical Internet Culture*. London: Routledge.
- Lovink, G. (2019). *Sad by Design. On Platform Nihilism*. Amsterdam: Pluto Press.
- Manovich, L. (2015). 'The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics'. Online retrievable via [http://manovich.net/content/04-projects/088-cultural-analytics-social-computing/cultural\\_analytics\\_article\\_final.pdf](http://manovich.net/content/04-projects/088-cultural-analytics-social-computing/cultural_analytics_article_final.pdf) (accessed 16/03/2020).
- Manovich, L. (2018). 100 Billion Data Rows per Second. Media Analytics in the Early 21st Century. *International Journal of Communication*, 12, 473–488.
- Masson, E. (2017). Humanistic Data. An Encounter between Epistemic Traditions. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society. Studying Culture Through Data*. Amsterdam: Amsterdam University Press.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data. A Revolution That will Transform How We Live, Work and Think*. London: John Murray.
- McStay, A. (2018). *Emotional AI. The Rise of Empathic Media*. London: Sage.
- Nguyen, D. (2017). *Europe, the Crisis, and the Internet. A Web Sphere Analysis*. London: Palgrave Macmillan.
- Papacharissi, Z. (2014). *Affective Publics. Sentiments, Technology, and Politics*. Oxford: Oxford University Press.
- Plowright, D. (2011). *Using Mixed Methods. Frameworks for an Integrated Methodology*. London: Sage.

- Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From Ranking Algorithms to 'Ranking Cultures': Investigating the Modulation of Visibility in YouTube Search Results. *Convergence: The International Journal of Research into New Media Technologies*, 24(1), 50–68.
- Rieder, B., & Röhle, T. (2017). Digital Methods. From Challenges to Bildung. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society: Studying Culture Through Data* (pp. 109–124). Amsterdam: Amsterdam University Press.
- Rogers, R. (2019). *Doing Digital Methods*. London: Sage.
- Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). 'Revisiting the Quantitative-Qualitative Debate. Implications for Mixed-Methods Research', in *Quality & Quantity* 36, pp. 43–53.
- Schäfer, M. T., & van Es, K. (Eds.). (2017). *The Datafied Society. Studying Culture through Data*. Amsterdam: Amsterdam University Press.
- Smyrniotis, N. (2018). *The Internet Oligopoly*. Bingley: Emerald Publishing.
- Srincek, N. (2017). *Platform Capitalism*. Cambridge and Malden: Polity.
- Teddle, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research. Integrating Quantitative and Qualitative Approaches in the Social and Behavioural Sciences*. London: Sage.
- Terfrüchte, L. (2011). *Symbolverwendung im Journalismus. Eine empirische Fallstudie zur öffentlich-rechtlichen Berichterstattung*. Münster: University of Münster.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*. New York: Editions Milan.