

13

DIGITAL SOCIAL RESEARCH

Key questions

- What are the new challenges for social research in digital society?
- What new types of data are available today, that were unavailable to social researchers in pre-digital society?
- How can the divide between 'qualitative' and 'quantitative' methods be bridged through 'methodological bricolage'?
- What ethical principles should guide digital social research?

Key concepts

Digital social research * the data environment * methodological pragmatism * methodological bricolage * research ethics

Much as the emergence and development of digitally networked tools and platforms has changed the parameters for social interaction, digital society also changes how we think about research methods. In fact, because of its relative newness, social scientific research about the internet and digital media is a key area of methodological development. Routine research labour is rapidly transformed when one tries to capture the fast-evolving patterns of sociality online and through digital tools. Because

of its transformative character, research on digital media still — some years into the ‘information age’ — gives rise to new methods, as well as new challenges and opportunities for the analysis of society and human behaviour.

In this part of the book, consisting of this and three following chapters, I will introduce a framework for carrying out *digital social research*. Of course, as the internet is such a big and ever-present part of today’s societies, there is no one way to define what ‘digital social research’ is. It could be any type of study using any kind of existing and established research method to say something about life in digital society.

The choice of a method for research relates to the aims of the study, the type of data to be analysed, personal preferences of the researcher, and so on. I will, however, present a suggestion as to what digital social research can be. This suggestion will be presented more in terms of a framework or toolkit than a fixed and ready-made universally applicable model. My specific version of digital social research will be, by necessity, shaped by the types of studies that I myself have been doing and engaging with, as well as by my background and position as a mixed-methods sociologist.

In Chapter 15, I will argue that *ethnography* — a ‘qualitative’ approach based largely on interpretations of rich data from interviews and observations — is a very useful method to capture the complexities of digital sociality and how digital society works. Ethnography conventionally relies on collecting research data through participant observation and interviews, aiming to generate close and detailed descriptions and interpretations of social life as it happens in context. I think that the ethnographic approach, with its interpretative stance, provides a solid foundation to undertake research, with potential to generate sociologically relevant knowledge about the complexities of digital society. As anthropologist Gabriella Coleman (2010: 488–489) states:

To grasp more fully the broader significance of digital media, its study must involve various frames of analysis, attention to history, and the local contexts and lived experiences of digital media — a task well suited to the ethnographic enterprise.

Sometimes ethnography alone can be more than enough as a research method, depending on what one wants to find out. But as I will argue in this chapter, the best strategy is often to combine ethnography with other methods. This is a consequence of the changing *data environment* and the increased social complexities which follow from the networked characteristics of digital society. Mixed-methods approaches entail, among other things, venturing beyond the longstanding and well-established divide between so-called qualitative and so-called quantitative methods, as well as crossing

boundaries between conventional academic disciplines. Doing digital social research, still a couple of decades into the 21st century, means entering new terrains and facing new challenges. I will address some of these challenges — for example those relating to *research ethics* — in this chapter.

In Chapter 16, I will discuss some methods for exploring, mapping, and mining digital society that can be useful to expand on the ethnographic foundation. I pay particular attention to social network analysis, but also deal with text mining — more or less automated techniques, developed in the fields of computer science and linguistics, for analysing large collections of documents. I also discuss what Richard Rogers (2013) calls ‘methods of the medium’. Before these two specialised chapters, Chapter 14 will offer some hands-on guidelines for the navigation of the research process — how to formulate a good research question, how to frame your field of study, as well as how to collect and analyse data. But first, in this present chapter, I will address a number of challenges and general strategies when undertaking digital social research.

THE DATA ENVIRONMENT

First, let’s think about what types of information we may be dealing with. As you will remember from the previous chapter, ‘big data’ has become a buzzword that is repeatedly used to name and characterise some of the new types of data that have emerged in digital society. In reality, however, the emergence of big data is only one of many transformations in our data environment, which affects opportunities as well as challenges when doing digital social research. For example, Kingsley Purdam, an expert in research methods, and his data scientist colleague Mark Elliot aptly point out that what is commonly known as ‘big’ data is in fact data defined by several other things, rather than just its large size: it registers things as they happen in real time, it offers new possibilities to combine and compare datasets, and so on. Furthermore, Purdam and Elliot believe that even these characterisations are still not sufficient. This is because those definitions still seem to assume that data is ‘something we have’, when in fact ‘the reality and scale of the data transformation is that data is now something we are becoming immersed and embedded in’ (Purdam & Elliot 2015: 26).

The notion of a ‘data environment’ underlines that people today are at the same time generators of, but also generated by, this new environment. ‘Instead of people being researched’, Purdam and Elliot (2015: 26) say, ‘they are the research’. Their point, more concretely, is that new data types have emerged — and are constantly emerging — that demand new flexible approaches. Doing digital social research, therefore, often entails discovering and experimenting with challenges and possibilities of ever-new types and combinations of information.

Different types of data

In trying to describe the ever-changing data environment, Purdam and Elliot (2015: 28–29) outline an eight-point typology of different data types based on how the data in question has been generated:

- 1 **Orthodox intentional data:** Data collected and used with the respondent's explicit agreement. All so-called orthodox social science data (e.g. survey, focus group or interview data and also data collected via observation) would come into this category. New orthodox methods continue to be developed.
- 2 **Participative intentional data:** In this category data are collected through some interactive process. This includes some new data forms such as crowdsourced data [...].
- 3 **Consequential data:** Information that is collected as a necessary transaction that is secondary to some (other) interaction (e.g. administrative records, electronic health records, commercial transaction data and data from online game playing all come into this category).
- 4 **Self-published data:** Data deliberately self-recorded and published that can potentially be used for social science research either with or without explicit permission, given the information has been made public (e.g. long-form blogs, CVs and profiles).
- 5 **Social media data:** Data generated through some public, social process that can potentially be used for social science research either with or without permission (e.g. micro-blogging platforms such as Twitter and Facebook, and, perhaps, online game data).
- 6 **Data traces:** Data that is 'left' (possibly unknowingly) through digital encounters, such as online search histories and purchasing, which can be used for social science research either by default use agreements or with explicit permission.
- 7 **Found data:** Data that is available in the public domain, such as observations of public spaces, which can include covert research methods.
- 8 **Synthetic data:** Where data has been simulated, imputed or synthesized. This can be derived from, or combined with, other data types.

The most important point here is that while social research traditionally relies on orthodox intentional data (1), such as surveys and interviews, digital society

has enabled much more far-reaching registration and collection of participative intentional data (2), consequential data (3), self-published data (4), and found data (7). These are types of data that indeed existed before digitally networked tools and platforms but which have been expanded and accentuated. The remaining types — social media data (5), data traces (6), and, at least chiefly, synthetic data (8) — are specific to digital society. Therefore, researchers who analyse this society face dramatically altered conditions for the generation and gathering of data about social processes and interactions.

REVEALING THE MESSY DETAILS

In today's world, large amounts of social data are registered and aggregated independently of initiatives from researchers. This is illustrated by work such as that of computational sociologists Scott Golder and Michael Macy (2011). Their research mapped people's affective states throughout the day as expressed via Twitter posts, in 84 countries, generating results of high interest to its subject-area, but using a research design that was, by necessity, dictated by the availability and character of the timestamped and text-based social media data. Examples of similar studies exist in several other fields where, while the issues dealt with are of high relevance, it is nonetheless the case that the researchers have confronted data that were largely already at hand and constituted in certain ways. Researchers of digital society are often left to dealing with the data generated through the platforms to be analysed, rather than having the opportunity to elicit data in conventional ways controlled by the researcher. While choosing an approach — for instance, opting for a survey or for in-depth interviews — will have continued relevance in some contexts, scholars are now increasingly also facing the challenge of thinking up and constructing some of their 'methods' after the fact.

One of Purdam and Elliot's (2015) main points in the presentation of their typology, discussed in the previous section, is the argument that the complexity of today's data environment forces researchers to constantly think about the highly variable characteristics of data that they encounter or seek out. And one of the key challenges when entering this type of terrain is the need to constantly try out new methods for data gathering. In order to know that the data we elicit or download, as well as the strategies we choose to make sense of it, are appropriate, we may test our strategy to see whether it produces good research results. However, the dilemma is that in order to know that the results are good, we must already have developed the appropriate method. Because of this constant — and potentially endless — need for experimentation and discovery, investigations drawing on new tools and approaches risk becoming stuck and intellectually unproductive very quickly.

For instance, you are researching some aspect of social interaction on a platform like YouTube, and have decided that an analysis of user comments on videos seems to be the data collection method of choice. Now, if this had been survey responses, or interview transcriptions, you could rely on an entire canon of literature on methods and well-established research practices in order to understand how to work with such data. Even though you might want to undertake new approaches or challenge the conventional ways of going about the research, you would at least have a sort of baseline or common practice to relate to and argue with. But in the case of YouTube comments, you would have to do a lot more groundwork. First, for example, you would have to find a way of collecting the comments. If the number of comments was large enough for it to be inconvenient to manually copy and paste them — which is often the case — you would have to find some tool or another to automatically capture and download them. This risks the use of trial and error as you work your way through a variety of browser plugins, scripts, or applications, none of which may eventually do what you want them to do. This process can be very time-consuming and it is not uncommon that the researcher becomes so engaged with this very quest for a tool that he or she — instead of doing the social research that was initially intended — starts to devote a lot of time searching for ever ‘better’ tools or learning how to code their own tools. And this is only the first step out of several subsequent ones, where other challenges may throw you off track.

Once the comments are collected and ordered, there are wide ranges of issues as regards to how knowledge of the comments should be achieved as well as ethical issues to address. What are the comments actually? Are they individual comments or conversations? How should you, if at all, take the likes and dislikes of the comments into consideration? Do all of the comments relate to the YouTube video in question, or can the comment threads take on lives of their own, to become forums for the discussion of issues other than those instigated by the video? How can you, ethically, use these data for research? Do you need the informed consent of all the people who have posted in the thread? And so on, *ad infinitum*. In sum, because of the inherent multidimensional complexity and unresolved questions, research on digital society must embrace research methods as a creative act. Instead of relying on previous work, copying and pasting run-of-the-mill methods sections into our papers, researchers must ‘reveal the messy details of what they are actually doing, aiming toward mutual reflection, creativity, and learning that advances the state of the art’ (Sandvig & Hargittai 2015: 5).

METHODOLOGICAL BRICOLAGE

Nearly twenty years ago, in the preface to a book about researching the internet, Steve Jones (1999: x) wrote that ‘we are still coming to grips with the changes that we feel are brought about by networked communication of the type so prominently

made visible by the Internet’. And this is still the case. Research on digital society has continued to be a trading zone between conventional academic disciplines — it is truly transdisciplinary. In their book about ‘internet inquiry’, Annette Markham and Nancy Baym (2009: xiv) explain that:

While most disciplines have awakened to an understanding of the importance of the internet in their fields, most do not have a richly developed core of scholars who agree on methodological approaches or standards. This absence of disciplinary boundaries keeps internet studies both desirable and frustrating.

This frustration, they argue, makes researchers of digital society push the boundaries of ‘disciplinary belonging’ in ways that most academic research would benefit from doing more of. Furthermore, they write that as very few internet researchers have been specifically trained in how to do it well, one is by necessity forced to actively and critically navigate a landscape of old and new methods in order to seek out ways of engaging with data that suit one’s particular project. It is seldom workable to just apply previously existing theories and methods when studying digital society. Some perspectives and approaches can most likely be, and have also to some extent been, repurposed for digital media research — for example, survey methods and interviews. But one must remember that the internet, and its networked social tools and platforms, is in many ways a different research context, possessing an ‘essential changeability’ that demands a conscious shift of focus and method (Jones 1998b: xi).

Because of this, researching digital society often demands that the person carrying out the data collection and analysis is even more critical, and more reflective, than what is already demanded by scholarship in general. The specific challenges of doing digital social research have, Markham and Baym (2009: vii–viii) argue, ‘prompted its researchers to confront, head-on, numerous questions that lurk less visibly in traditional research contexts’. One such issue is the urgent need to address the longstanding dispute in social science between ‘qualitative’ and ‘quantitative’ methodological approaches, which has persisted, apparently unresolvable, for more than a century. Among researchers, there are still traces of a battle between case-oriented interpretative perspectives, on the one hand, and variable-oriented approaches focused on testing hypotheses on the other. Scholars who prefer case-oriented methods will argue that in-depth understandings of a smaller set of observations are crucial for grasping the complexities of reality, and those who prefer variable-oriented approaches will argue that only the highly systematised analysis of larger numbers of cases will allow scholars to make reliable statements about the ‘true’ order of things.

Today, however, there is an increasingly widespread consensus that the employment of combinations of ‘qualitative’ and ‘quantitative’ methods is a valid and recommended

strategy, which allows researchers to benefit from their various strengths and balance their respective weaknesses. The ‘qualitative’ tradition is seen as the more inductively oriented interpretative study of a small number of observations, while the ‘quantitative’ tradition is characterised by the deductively oriented statistical study of large numbers of cases. This has given rise to the common notion that ‘qualitative’ research produces detailed accounts through close readings of social processes, while ‘quantitative’ research renders more limited, but controllable and generalisable, information about causal relations and regularities of the social and cultural fabric.

I think that the best strategy is *methodological pragmatism*, focusing on the problem to be researched, and on what type of knowledge is sought. Instead of methodological positioning within the existing field of methods literature, one can instead, as methodologists Norman Denzin and Yvonna Lincoln (2000) have suggested, conceive one’s research strategy as a form of *bricolage*. ‘Bricolage’ is a French term — popularised by cultural anthropologist Claude Lévi-Strauss (1966) — which refers to the process of improvising and putting pre-existing things together in new and adaptive ways. From that perspective, our research method is not fully chosen beforehand, but rather emerges as a patchwork of solutions — old or new — to problems faced while carrying out the research. As critical pedagogy researcher Joe Kincheloe (2005: 324–325) observes: ‘We actively construct our research methods from the tools at hand rather than passively receiving the “correct,” universally applicable methodologies’, and we ‘steer clear of pre-existing guidelines and checklists developed outside the specific demands of the inquiry at hand’. So, developing your method as a bricolage means placing your specific research task at the centre of your considerations, and allowing your particular combination and application of methods take shape in relation to the needs that characterise the given task.

THINKING ABOUT WHAT WE CAPTURE

The previously discussed demand for reflexivity on behalf of the digital social researcher operates on several different levels. In a similar vein to the bricolage approach described above, Markham and Baym also argue that research design is an ongoing process, and that it is to be expected that any study will be reframed continuously throughout the process of research. They write:

Different questions occur at different stages of a research process, and the same questions reappear at different points. Second, the constitution of data is the result of a series of decisions at critical junctures in the design and conduct of a study. [...] We must constantly and thoroughly evaluate what will count as data and how we are distinguishing side issues from key sources of information. (Markham & Baym 2009: xvii)

As Jones (1999) emphasises, when researching the specificities of the internet, it is important to remember that its uses — as discussed in Chapter 1 — are always contextualised. Research subjects, both human and non-human actors in the sense of actor-network theory, as mentioned elsewhere in this book, are part of physical space as much as they are part of ‘cyberspace’. This means, Jones (1999: xii) says, that ‘[n]ot only is it important to be aware of and attuned to the diversity of online experience, it is important to recognize that online experience is at all times tethered in some fashion to offline experience’.

So, while it is exciting to study the internet and digital society, it is also especially challenging. New platforms, concepts, and social practices emerge fast enough for making the ‘internet’ in itself into a compelling area of inquiry. The field, Markham and Baym (2009: xviii–xix) write, has a ‘self-replenishing novelty [that] always holds out the promise for unique intellectual spaces’. But, as discussed above, new terrains of research brings with them new challenges and difficulties. First, there is a need for constant reflection about the role of the self in research. Processes of digital social research highlight that researchers are actually co-creators of the field of study. Our choices are made in contexts where there are no standard rules for research design and practice, and this makes such choices more meaningful. Furthermore, the often-disembodied character of digital social settings makes it important to think a little deeper about the relationship between the researcher and the researched:

What decisions are we making to seek consent; what counts as an authentic self-representation? How are we conceptualizing the embodied persons we study? How are we framing our own embodied sensibilities? Do we approach what we are studying as traces left in public spaces or as embodied activities by people situated in rich offline contexts? We must consider how to interpret other people’s selves and how to represent ourselves to the people we study, especially when we may not be meeting them in person. (Markham & Baym 2009: xviii–xix)

Researchers and their subjects, Purdam and Elliot (2015: 47) say, increasingly bleed into one another. This is because ‘as the proportion of our lives spent online grows, so the boundary between data and subject becomes less distinct’. In the same sense that offline identities of people are partially coming together in the minds and memories of others, our online selves are partially constructed in our intentional or unintentional data footprints.

Second, Purdam and Elliot argue, ‘the activities of others also contribute to constructing these footprints, for example, a photograph of a person might be in the public domain as a result of being posted online by someone else’. Additionally, that photograph might also have been shared, tagged, liked, or remixed by somebody else, and it may contain ‘meta-identity information’ (2015: 47). So, if a ‘researcher’ analyses this

photo, posted by a ‘research subject’, then who or what is actually being analysed? Things are further complicated in the movement from orthodox intentional datasets to various types of data streams or synthetisations, which blurs the distinction between data and analysis.

Third, and finally, it is important to think about the quality of the data used in research. Conventional social science has a set of established mechanisms for quality control, which assess things such as the reliability, validity, and generalisability of research results. The introduction of new types of data, and new modes of data gathering, demands that we ask ourselves questions about rigorous and robust methods of going about our research in order to avoid unnecessary errors or biases. When analysing different platforms, such as a discussion forum or Twitter, and making claims about society, we must remain critical to whose views — whose society — are being expressed on the platform in question, and in our particular sample. Generally, however, conventional and established ways for thinking about such things can’t be easily transferred to studies based on many of the new data types. The criterion of validity, for example, is about evaluating to what degree one is actually studying what one purports to study. Giving an example based on Twitter, Purdam and Elliot (2015: 48) posit:

For example, a tweet might be generated for fun, to provide information or to persuade or mislead; the motivation obviously affects the meaning of the tweet. With survey data and even, to some extent, administrative data, the impact of respondent motivations is, at least in principle, structured by (or perhaps mediated by) the data collection instrument itself. Thus, a well-designed social science research instrument can constrain motivational impact. But this is not so with Twitter data; here people’s motivations are given full rein – a tweet might be designed to manipulate or obfuscate, to attract truth or to repel it. It might be designed to fantasize or ‘try out an opinion’, to provoke a response or simply to create controversy.

So, here we can choose different pathways: Do we want to find verification techniques with which to check the ‘quality’ of these data — for example, looking at a user’s tweets over time to see whether a tweet is characteristic or not — or is it more feasible to argue that we are not studying the person, but something else. Society? Culture? The medium?

DIGITAL RESEARCH ETHICS

It has been argued throughout this chapter that research about digital society demands continuous critical reflection. This is true to an even larger degree than in many other types of research, since this is a field without, as yet, an established methodological

tradition. The negotiations are ongoing and, as illustrated above, it is important to maintain an ongoing discussion about methods, and for researchers to give others insight into their research processes, even though they may be messy or sometimes even may feel like one is ‘cheating’ or cutting corners. The subject of digital research ethics is an especially urgent strand of this discussion. Here, too, there is an ongoing discussion about best practice. How the researcher navigates issues of research ethics will differ, depending on how the data in question have been generated and collected. Looking at the typology presented by Purdam and Elliot (2015), new ethical challenges arise especially in relation to self-published data, social media data, data traces, and found as well as synthetic data. New data types, and new ways of accessing and gathering data, demand that the researcher constantly navigates the data environment and makes choices in a critically reflective way.

Principles of research ethics, and how persons that are researched should be ethically treated, are codified in a number of documents and policies throughout the academic community. Most of these codifications work best in relation to what Purdam and Elliot call ‘intentional data’. And although ethical principles about maximising the benefits and minimising the harm of research are a good starting point — as is the obvious need to respect fundamental rights of human autonomy, dignity, and safety — digital research demands consistent reflection, as discussed in the previous section, about what information one is really capturing. Like the other methodological considerations discussed in this chapter, issues that relate to how one should deal with the ethical treatment of data are highly context-sensitive. Markham and Baym (2009: xviii) emphasise that context-specific uses of the internet demand that the researcher continuously and carefully reconsiders notions such as privacy, consent, trust, and authenticity.

The emergent character of the field of digital social research, a field which is in a perpetual ‘beta state’, makes it impossible to escape questions about ethical decisions. Such questions must be posed and responded to iteratively. Even though one might wish there were clear rules, issues like these must always be navigated inductively. In light of these concerns, the Association of Internet Researchers (AoIR) has put together an Ethics Working Committee, composed of internet researchers from a variety of regions and countries. The committee argues that ethical issues are complex and that they can rarely be handled in any binary way. There is in fact ‘much grey area’ (Buchanan & Markham 2012: 5). The transdisciplinary character of digital social research means that researchers and institutions confront many contradictions and tensions that are impossible to resolve completely. Instead, ‘many competing interests must be negotiated by researchers, ethics review boards, and institutions’ (2012: 6). For the individual researcher it is recommended, the committee writes, that ethical decision-making is approached as a process, dealing with the issues in a contextualised fashion throughout the

research process. This is because different issues will be pertinent at different stages. This approach is in line with the perspective of research method as a bricolage. It is not only the methodological choices, but also the ethical considerations that are emergent and unwinding.

Ethical principles for digital social research

The AoIR Ethics Working Committee (Buchanan & Markham 2012: 4–5) has arrived at the following key guiding principles:

The greater the vulnerability of the community/author/participant, the greater the obligation of the researcher to protect the community/author/participant.

Because ‘harm’ is defined contextually, ethical principles are more likely to be understood inductively rather than applied universally. That is, rather than one-size-fits-all pronouncements, ethical decision-making is best approached through the application of practical judgment attentive to the specific context [...]

Because all digital information at some point involves individual persons, consideration of principles related to research on human subjects may be necessary even if it is not immediately apparent how and where persons are involved in the research data.

When making ethical decisions, researchers must balance the rights of subjects (as authors, as research participants, as people) with the social benefits of research and researchers’ rights to conduct research. In different contexts the rights of subjects may outweigh the benefits of research.

Ethical issues may arise and need to be addressed during all steps of the research process, from planning, research conduct, publication, and dissemination.

Ethical decision-making is a deliberative process, and researchers should consult as many people and resources as possible in this process, including fellow researchers, people participating in or familiar with contexts/sites being studied, research review boards, ethics guidelines, published scholarship (within one’s discipline but also in other disciplines), and, where applicable, legal precedent.

While discussions about the concept of the ‘human subject’ in digital social research, about definitions of public versus private, about data protection and ownership, and of several other pertinent dimensions must be continuously reviewed, the AoIR committee presents the above points as general principles for researchers to turn to as a starting point. The guidelines described in the quote above state that:

- The vulnerability of research subjects should decide how careful the researcher is.
- The rights of research subjects to be protected should be balanced against the importance of conducting the research.
- Research ethics must be continuously discussed among researchers and other relevant actors.

To this list, we can add some more important things to keep in mind. Anthropologist Tom Boellstorff and colleagues (2012: 129–149) suggest the following:

- The principle of care. Taking good care of informants and making sure that they gain something from their participation.
- Informed consent. Make sure that informants know about the nature and purpose of the study.
- Mitigating legal risk. Being aware of relevant laws that govern one’s research.
- Anonymity. Avoiding the inappropriate revelation of the identities of informants — or any sort of confidential details or otherwise — that might lead to their identification.
- Deception. Don’t pretend to be something you are not, and don’t use ‘fly on the wall’ practices to study sensitive topics.
- Empathy. Try to forge a ‘sympathetic depiction of informants’ lives, even when discussing aspects of informants’ lives that some might find troubling’. This does not have to mean that the researcher ‘agrees’ with any actions or beliefs of the informants, but one must labour to ‘grasp informants’ own visions of their worlds’.

FURTHER READING

Brady, Henry E., & Collier, David (Eds.) (2010). *Rethinking Social Inquiry*. Lanham, MD: Rowman & Littlefield.

With this edited volume, Brady and Collier engage in discussions about deficiencies of ‘qualitative’ versus ‘quantitative’ approaches, and direct focus towards how to navigate and

(Continued)