

Digital Data and Methods as Extension of Qualitative Research Processes: Challenges and Potentials Coming From Digital Humanities and Computational Social Sciences

Lina Franken

Key words: digital humanities; computational social sciences; qualitative social research; research process; digital methods; digital data; sociology of knowledge approach to discourse

Abstract: Integrating digital data and methods can extend the scope of qualitative social research. However, to date, this has hardly ever happened. Computer and data scientists have already developed tools that could be used as a starting point, but they must be adjusted to the special needs of qualitative research. To explore the limits and possibilities for qualitative social research, it is not only essential to critically analyze digital data and the corresponding methods. Researchers should also reflect on their own perspective. In my contribution, I point to connections to the digital humanities and the computational social sciences. Based on the example of a discourse analysis that followed a sociology of knowledge approach, I model the qualitative research process and demonstrate which steps can be extended with digital methods. Subsequently, I discuss key challenges and point to strengths and possible pathways to show how digital data and methods can be further developed. I lay a foundation for creatively integrating new digital elements into the qualitative research process. I thus shed light on the transformation of research questions, fields, methods, and epistemologies.

Table of Contents

- [1. Digital Data and Methods and How They Are Connected to Qualitative Social Research](#)
- [2. Digital Data and Methods for Qualitative Research in Cultural Studies and Social Sciences—a Contradiction?](#)
 - [2.1 Digital humanities, computational social sciences and digital methods](#)
 - [2.2 Points of departure in cultural studies and social sciences: data for qualitative research](#)
- [3. Modeling Qualitative Research Processes Using a Discourse Analysis](#)
- [4. Extending the Research Process in Terms of Data Selection, Collection, Processing and Analysis](#)
- [5. Challenges and Potentials](#)
 - [5.1 Challenges](#)
 - [5.2 Potentials for further development](#)
- [6. Conclusion: Digital Data and Methods Extend Qualitative Research Processes](#)

[Acknowledgments](#)

[References](#)

[Author](#)

[Citation](#)

1. Digital Data and Methods and How They Are Connected to Qualitative Social Research

People are increasingly producing data in their everyday lives (KITCHIN, 2014a; LUPTON, 2015). For example, it has been estimated that worldwide in 2020, around 190 million e-mails were sent per minute, 1.3 million logins were made to Facebook, and over a million U.S. dollars were spent in online shopping (LEWIS, 2020). However, it is not only the production of data that is rapidly increasing as a result of digitization. Human activity is also influenced by the collection, analysis, and utilization of data; the analog and the digital are closely interrelated (BECK, 2019 [2015]; BOELLSTORFF, 2016). At the same time, it is postulated that there is a "Metric Society" (MAU, 2019) which is said to be more quantifiable than ever before. Accordingly, in scientific discourse, there is a powerful narrative that in the age of the Internet, an almost infinite amount of text, image, audio, and video data is available for potential analysis. At the same time, access to large amounts of digital data are presented as simplified, facilitating a quick and precise analysis with digital methods (DOURISH & GÓMEZ CRUZ, 2018; MAYER-SCHÖNBERGER & CUKIER, 2013). In addition, existing qualitative research practices are challenged, as in many cases they are no longer considered sufficient for the analysis of digitally permeated societies. Numerous challenges concerning the use and further development of digital methods and data would need to be addressed in order to realize the potential that undoubtedly exists for cultural studies and the social sciences (MOSCONI et al., 2019; PAULUS, JACKSON & DAVIDSON, 2017; RUPPERT, LAW & SAVAGE, 2013). In this paper, I assume that qualitative research is both changed and extended when digital methods are applied and digital data are processed. Hence, I explore on the one hand which concrete data and methods qualitative researchers could use in which settings. On the other hand, I question how qualitative research designs change when individual steps are realized in a digital way. [1]

There is potential for qualitative research practices in using and developing digital methods. It goes beyond digital or virtual ethnographers development of ethnographic methods to collect data in digital research fields (HINE, 2015; KNOX & NAFUS, 2018; PINK et al., 2016), and also does not focus on the use of a single tool for digital data analysis, for instance in the context of *computer-assisted qualitative data analysis software*, CAQDAS for short (EVERS, CAPRIOLI, NÖST & WIEDEMANN, 2020; GIBBS, FRIESE & MANGABEIRA, 2002; MacMILLAN, 2005). Rather, I argue for an understanding of digital methods that sets out to remove the rigid boundaries between qualitative and quantitative analyses of society. The aim in doing so is not to abandon or subordinate qualitative perspectives. Instead, I ask how methodological approaches and digital possibilities are intertwined and conditioned by each other, which digital data and methods can be used purposefully in qualitative research perspectives, and how they need to be further developed. [2]

In the following, I will first explain what is meant by *digital humanities* (DH) and *computational social sciences* (CSS), and how the two differ or overlap (Section 2.1). I then outline qualitative social research from a cultural studies perspective

and address how digital data extends the scope of sources (Section 2.2). Using the example of my own study of digitization in the health care sector that follows KELLER's sociology of knowledge approach to discourse (2011 [2005]), I model qualitative research processes (Section 3) and demonstrate how they can be extended through digital data and approaches (Section 4). I then systematize key challenges and identify potentials of digital methods in qualitative research (Section 5). Finally, I draw conclusions with regard to using digital data and methods and their further development and adaption to the needs of cultural studies and the social sciences (Section 6). [3]

2. Digital Data and Methods for Qualitative Research in Cultural Studies and Social Sciences—a Contradiction?

2.1 Digital humanities, computational social sciences and digital methods

Recently, in humanities, cultural studies, and the social sciences, digital methods are in development with different foci. In the humanities they are referred to as DH and in the social sciences as CSS. Both are considered bridges that link subjects (SAHLE 2015, Chapter 2) or transformational sciences (JANNIDIS, KOHLE & REHBEIN, 2017a, p.XI), which are expected to be particularly useful in analyzing large amounts of data. For this purpose, new, digital methods of investigation are used in addition to digital data. In this context, quantitative approaches are primarily used because they are compatible with methods from information technology. One difference between DH and CSS concerns the types of data used. While DH work predominantly with historical, retro-digitized material (such as historical documents or paintings), researchers in CSS focus on contemporary digital data (such as posts on Twitter or media coverage) (FRANKEN, 2022a). The methodological developments and considerations, however, are comparable. In both cases, the potential goes far beyond bringing together the digital paradigm (in the sense of computer and information sciences) and humanities or social sciences (in the broad sense of humanities, cultural studies or social sciences). [4]

Digital methods push the boundaries of established qualitative methods that include interviewing, observations, or analyzing audiovisual data, as we will see in Section 4. This changes the way data are generated and analyzed and the way in which findings are presented (BERKELAAR & FRANCISCO-REVILLA, 2018, p.17). As a result, new research opportunities are emerging. SAHLE (2015, Chapter 2) differentiated between "low end" and "high end" methods. However, this classification implicitly devalues the use of generic tools because "high end" DH is in fact not only about developing specific new tools and methods, as he put it (ibid.). Extending one's own methodology with DH perspectives also goes far beyond what would be called "low." SVENSSON (2016, p.16) distinguished between minimal and maximal readings of DH, and thus between "application of computer technology to traditional scholarly work (a minimalist reading) and changing the substance of humanistic matter (a maximalist reading)." Neither, however, clarified clearly enough that fundamentally different activities can be captured with digital methods. In the following, I will therefore distinguish between

an extended methodological toolbox on the one hand and, on the other hand, a methodological experiment kit through which something new can be created. Both DH and CSS are to be understood as toolboxes because existing basic approaches from computer science are used for simple operations. With them, digital methods, procedures, tools and standards are established and further developed, hence facilitating work. For example, commonly used terms in texts can be easily identified ("n-grams," BUBENHOFER, 2017) or large amounts of data from the web can be stored ("crawling," BARBARESI, 2019, p.30). From the perspective of an experiment kit, new things can emerge in DH and CSS through collaboration and openness: experiments that are only made possible by extending the scope of techniques and combining them for new research questions. Through machine learning, for example, researchers can extract implicit statements from historical sources (NELSON, 2020) or visualize the descriptive language used to express specific periods from a large set of images by arranging them by color or brightness (MANOVICH, 2020). However, these experiments become possible only after digital processing has been adapted, modified, and combined to suit possible research interests. Through this differentiation into toolboxes and experiment kits, it becomes clear that digital methods capture a wide range of different activities. [5]

Researchers in DH and CSS ask how research questions in the humanities and social sciences can be used and adapted with digital technology and concepts from computer science. Such methods, procedures, and models are referred to as computational because they use computer programs or scripts to study (not only) digitized societies by collecting, processing, and analyzing digital data. For the humanities, cultural studies, and social sciences, this is an opportunity to rethink how digital processes can and should be combined and enhanced to answer emerging questions. [6]

Both transformational sciences are in the process of becoming institutionalized. DH have been increasingly developing into a stable research network which has come to include research centers, service points in libraries, professorships, and degree programs in recent years; on the blog of their German-language association, more than 100 professorships have been reported in this context (SAHLE, 2019). Introductory textbooks have been published in German (JANNIDIS, KOHLE & REHBEIN, 2017b) and English (SCHREIBMAN, UNSWORTH & SIEMENS, 2004) for more than 15 years, with "Humanities Computing" (McCARTY, 2005) considered to be a precursor since the 1980s. In contrast, there is no German-speaking professional society for CSS (yet), and only few conferences, professorships and degree programs exist. The international research community is more diverse and more strongly shaped by the natural sciences, with central research personalities coming from physics and mathematics (CIOFFI-REVILLA, 2018; CONTE et al. 2012). In recent years, there has been an increase in contributions from political science (BLÄTTE, BEHNKE, SCHNAPP & WAGEMANN, 2018) and sociology (LUPTON, 2015; MARRES, 2017). Furthermore, there are competing terms such as "Digital Methods" (ROGERS, 2019), "Social Research in the Digital Age" (SALGANIK, 2018), "Digital Methods for Social Science" (SNEE, HINE, MOREY, ROBERTS &

WATSON, 2016), or "Digital Culture" (UNIVERSITY OF KLAGENFURT, 2021) that are not always clear-cut in terms of methods, data base, and objects of study. If these individual developments are considered as a whole, they are far from being marginal or existing only as a small discipline (ARBEITSSTELLE KLEINE FÄCHER, 2018). They move into the middle of the humanities, cultural studies, and social sciences. Researchers in individual disciplines continue to further differentiate them. This multifaceted field of research is heterogeneous, and researchers elaborate and experiment simultaneously. [7]

DH and CSS are still in their early stages in many areas. Developments are often project-driven, as SAHLE (2015) noted. There is an unmanageable number of tools and data that are not systematically linked with each other and have often only been created in the form of prototypes. Currently, research is moving between digital manual work and *big data* analyses: In this context, I define digital manual work as working with generic tools, i.e., software used on an everyday basis, such as word processors or Internet browsers. In addition, there is manual work on data sets that is realized with little or no computation, for instance in literature databases, lists of data, or even computer-assisted annotation in CAQDAS environments (FRANKEN, 2020a). *Big data* analyses, on the other hand, promise easy access, but often require enormous effort, and sometimes the results are frustrating. There is also a great lack of knowledge about digital methods across the board in the humanities, cultural studies, and social sciences. When digital methods are already used in broader settings, researchers usually rely on individual tools. NELSON (2020), for example, conducted her analysis based on *topic modeling* (see Section 4 for the procedure), but there was no manual lecture or annotation of the respective texts. This goes hand in hand with complexity reduction, which must not turn into pragmatism. At present, a process of negotiation is underway to explore where digital methods and where digital manual work are necessary and useful. To grasp the character of this process, one can see it as a "liminal space or contact zone" (SVENSSON, 2016, p.5). Liminality, drawing on research on rituals (TURNER, 2005 [1969]; VAN GENNEP, 1999 [1909]), has been understood as a phase of transition in which something new is purposefully created. In a classical understanding, people consciously shape liminality to actively commit to and make visible the liminal phase between old and new when it comes to major changes in their lives, such as starting school or getting married. In terms of digital data and methods, this is the transition from established methods in the humanities, social sciences, and cultural studies to computationally enhanced approaches. If qualitative research perspectives are to be part of this evolution, researchers must actively engage in the liminal phase and visibly shape the radical change to find a place in the newly emerging sphere. [8]

2.2 Points of departure in cultural studies and social sciences: data for qualitative research

From a cultural studies perspective, qualitative research is about contextualized meaning and the socially constructed character of culture (LINDNER, 2003; RECKWITZ, 2004). I argue from the viewpoint of empirical cultural studies as part of the qualitative social research discourse. In such a wide, relational understanding of culture, understanding meaning is central. Speaking with Clifford GEERTZ (1999 [1973]), "man is an animal suspended in webs of significance he himself has spun" (p.5). The present has become historical and is in flux, and therefore cultural heritage and memory practices are of great importance. However, the research interest of qualitative research focuses on the present and recent past and concrete practices—i.e., social action in its associated narratives and knowledge representations—as well as power structures and cultural codes as normative settings which are made explicit through cultural analysis (BECK, 2009; LEIMGRUBER, 2013). RECKWITZ (2008, p.201) has conceptualized these connections as formations of practice and discourse in which both symbolic structures and subjective interpretations are considered. [9]

In this understanding, researchers focus on relations and dynamics between actors. Their interaction with the world is seen as eventful. Non-human actors also have agency (BARAD, 2003). This is especially true in the digital realm, where technical devices and algorithms actively co-construct meaning (REICHERT & RICHTERICH, 2015). The reading of cultural studies as qualitative social research also includes a clear positioning in critical cultural analysis through which the representation of self and other is questioned. Postcolonial and feminist theoretical approaches especially in the context of *science and technology studies* (BECK, NIEWÖHNER & SØRENSEN, 2012; SISMONDO, 2010) are central to this research. Cultural studies thus aim "towards new cultural insights, as well as towards new ways of thinking about, practicing and presenting ethnographic analyses" (FORTUN et al., 2014, p.635). Culture becomes evident in language, knowledge, symbols, materialities, and actions. Accordingly, researchers in cultural studies use a variety of data, and with digital data this scope is extended. [10]

Following BAUR and GRAEFF (2021), I distinguish between research-induced and process-produced data for the humanities, cultural studies, and social sciences. Research-induced data are produced through scientific data collection: Surveys, interviews, photographs, videos, and observation protocols. Process-produced data (BAUERNSCHMIDT, 2014) are existing corpora such as historical archives, administrative data, objects or buildings and, last but not least, products of popular culture such as novels, films, comics, or songs. Some of these data have already been processed as *digital heritage* (SMITH, 2013) or in *digital editions* (KLUG, 2021). In addition, there are new types of process-produced data, especially *trace* data and data from social media. Trace data are data such as a location transmitted by a smartphone or search terms entered on the Internet. They are process-produced, arising rather incidentally and sometimes

without the knowledge or reflection of those who generate them. *Social media* data are consciously produced by users, as contributions such as text or photos are actively shared with others on a platform, but the producers of the data are not necessarily aware of the long-term storage and the sharing with third parties such as platform operators or researchers. Even though there is an increasing awareness of the data produced by the use of digital platforms, many people remain unaware of the extent to which data can be exploited (KROPF, 2019; PEETZ, 2021). Not being generated by researchers, they are also part of process-produced data, as Katharina KINDER-KURLANDA (2020) noted. The category of new types of process-produced data also includes various internet data that may result from a remediation from other process-produced data (such as media reports), but also blogs or forums as born digital data. Not all this data can be processed directly with digital methods. Rather, it is necessary to distinguish by data types. Text data may be found for instance as media reports or texts from the Internet, but can also constitute of interview transcripts or posts on social media. Image data can be extracted from social media, but can also be generated in surveys or stored in archives to document objects and buildings. Finally, audiovisual data may be collected as part of social media or Internet data, such as those embedded in a website, or generated as part of an empirical data collection. In addition, numerical data is becoming more important, as for instance in *trace* data. [11]

Researchers in the humanities, cultural studies, and social sciences usually draw on a multimodal data base, which is especially true for empirical cultural studies. Even without digital methods, qualitative researchers have to select from this variety of data depending on their research question (KOCH & FRANKEN, 2020) in order to extract a meaningful and at the same time manageable set of data. They often experiment methodologically in order to be able to grasp the phenomena under investigation in their complexity. This is evident in digital ethnography, where ethnographic methods are transferred and adapted to digital environments. Another example would be the "walkthrough method" (LIGHT, BURGESS & DUGUAY, 2017), a way of researching apps by systematically documenting the steps of registration and usage. For realizing digital methods, this selection currently means distinguishing by data type. Thus, in my own discourse analysis referenced in this paper, I focused on textual data because there is already a wide range of techniques for collecting and analyzing them with different digital methods. While being pragmatic in the research process, this operationalization should not be understood in a generalizing way, because in principle multimodal data bases are of course also to be considered for digital data and methods. As is the case when the research field is mapped out, the design of a study is partly motivated by access and inaccessibility (MARCUS, 2009; RIEKER, HARTMANN SCHAEELLI & JAKOB, 2020). However, in contrast to other methods, technical possibilities are far more important for the implementation of digital methods because with them, access to the field is restructured. Moreover, they require pragmatic solutions in order to make the data accessible for interpretation. This brings me back to the initial observation of data that only seem to be available and which need to be systematically

restructured for use in qualitative perspectives, especially in the digital sphere. A specific research program is necessary to address challenges and potentials. [12]

3. Modeling Qualitative Research Processes Using a Discourse Analysis

In order to use digital methods, the steps in the research process must first be operationalized. Modeling the research process is helpful in this endeavor (FRANKEN, 2020b). I agree with McCARTY who interpreted modeling as an abstraction: "A 'model' I take to be either a representation of something for purposes of study, or a design for realizing something new" (2004, p.255). McCARTY followed Clifford GEERTZ (1999 [1973], p.93) by distinguishing a "model of" and a "model for." Even though GEERTZ emphasized precisely how the two levels for models of culture are interwoven, it is useful to separate them analytically. For example, the DH literary scholar JANNIDIS did not distinguish between "of" and "for," but simply stated an "of" function. For him, a model is a representation of something, a thing, a term, in another medium, for instance language, image or sound (2017, p.100). FLANDERS and JANNIDIS (2016, p.230) differentiated process modeling from data modeling, the latter being at the level of precisely defining data formats, as was the case in the above-mentioned general distinction for the data spectrum in the humanities, cultural studies, and social sciences. Process modeling as a "model of" is more common in some disciplines than in others. In cultural studies and the social sciences, in which researchers both work empirically and historically, research steps are visualized by numerous introductory textbooks in a linear sequence or circular way (for example, in FLICK, 2007, p.128; MURI, 2014, p.463; REHBEIN, 2017, p.342; WITT, 2001, §15). Modeling is also ubiquitous in computational thinking (DENNING & TEDRE, 2019). This is because in order to create an algorithm, the individual steps must first be broken down into parts in order to precisely determine the method and sequence. Modeling helps in a double sense: through it, one becomes aware of one's own actions and at the same time, it can be used to specifically work out where digital data can be integrated into qualitative research and where and how they can be analyzed with digital methods. [13]

The aim of modeling is to abstract processes which means that essential aspects are considered while deviations from the ideal-typical course are left out (FLANDERS & JANNIDIS, 2016, p.229; THALLER, 2017, p.16). In principle, research processes in the humanities, cultural studies, and social sciences are comparable (see Figure 1): There is a problem that is described in such a way that a manageable question can be derived from it. On this basis, researchers choose a methodological setting and concrete methods before they collect data and analyze the data. Finally, the researchers write down their results from which new problems can be derived, either by themselves or other researchers. Such a process is then not linear, but iterative and cyclic. Examples include adjustments due to pretest results or when it becomes apparent that relevant data are still missing. In DH, these steps were soon referred to as part of the "Scholarly Primitives" (UNSWORTH, 2000), defined as "basic functions common to

scholarly activity across disciplines, over time, and independent of theoretical orientation" (§2). [14]

In the following, an exemplary study will be examined more closely to clarify these idealized processes. The focus was on the question of how the acceptance of digital health care in general and of telemedicine in particular is negotiated and which problems are identified by relevant actors in the discourse. The discourse analysis was realized in the research network "Automated Modeling of Hermeneutic Processes (hermA)" (GAIDYS et al., 2017) and was characterized by the collaboration of scholars from the humanities, cultural studies, and social sciences with computer scientists and computer linguists. Transferring this study to the model outlined above, first, the problem was acknowledged that telemedicine is a topic discussed in society. From this step, the question arose which actors are involved in this discourse and how acceptance becomes a subject of discourse. Methodologically, I chose a sociology of knowledge approach to discourse (KELLER, 2011 [2005]) with sampling and coding according to the grounded theory methodology (GLASER & STRAUSS, 2010 [1967]; CHARMAZ, 2014). In terms of data collection and analysis, I iteratively reduced the potentially relevant multimodal data to digitally available text in order to be able to use different digital methods. Theoretical sampling (STRÜBING, 2019) and inductive methods as well as abductive inferences (REICHERTZ, 2013) were essential for compiling the corpus. Manual annotation as digital manual work (FRANKEN, 2020a; FRANKEN, KOCH & ZINSMEISTER, 2020) was central to understanding contexts and carving out discourse positions and actors during the multiple loops of data collection and analysis. This was followed by writing down central results, such as the fact that patients had disappeared from the discourse (FRANKEN, 2022b). Additionally, further problems were articulated, such as the discursive settings around data protection. Figure 1 shows the idealized research process and a selection of digital extensions that were integrated. Using this approach, around 13,000 texts were collected, later reduced to approximately 9,000 relevant ones. Using the different digital methods explained below, 87 interesting texts were preliminarily selected to undergo qualitative analyses.

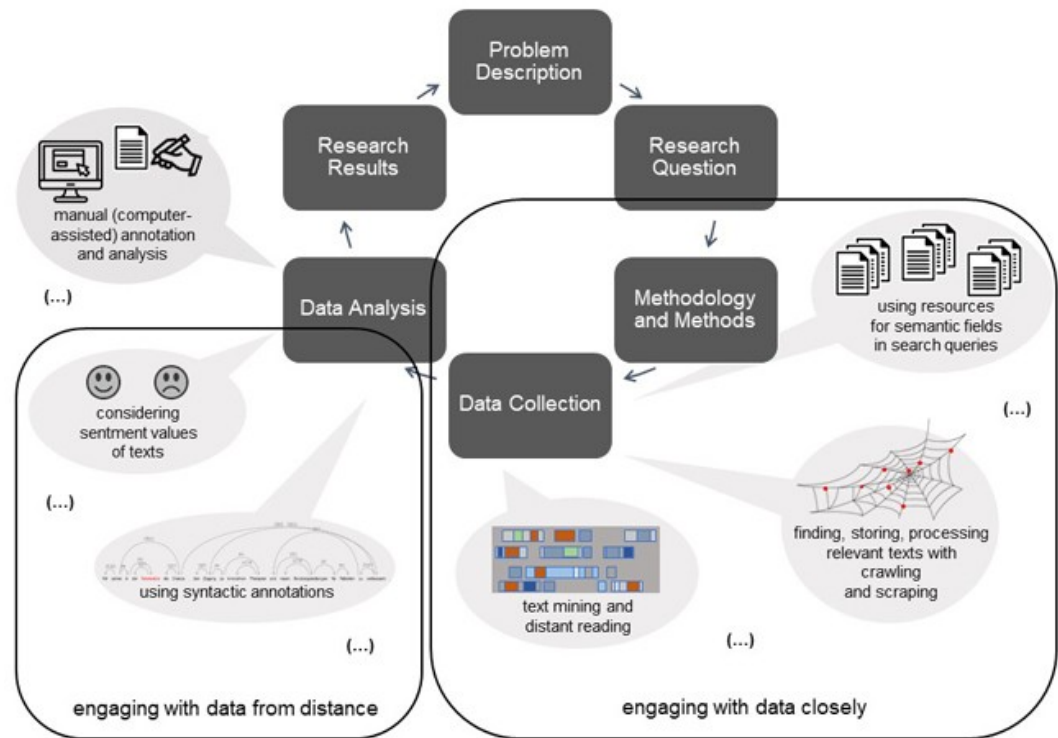


Figure 1: The process of qualitative research, extended by digital methods [15]

4. Extending the Research Process in Terms of Data Selection, Collection, Processing and Analysis

Modeling the discourse analysis in Section 3 showed that data selection, collection and analysis are suitable for digital methods. With these methods, researchers change and extend the previously described research process. In a closer look at these steps, the first thing to note is data selection, which is a preparatory step in data collection. This requires prior knowledge about the field of inquiry, which is key for conducting an informed search. Using resources such as controlled vocabularies and norm data turns out to be particularly helpful. Semantic fields can be formed based on the norm data and then extended with the results of a manual review of some of the data at an early stage (for the method, see ADELMANN, FRANKEN, GIUS, KRÜGER & VAUTH, 2019). With semantic fields, linguistic connections are conceptualized as semantic units, even though they are not clear-cut (TRIER, 1973 [1931]). By creating them, we no longer search for individual terms, but with a set of terms. Creating this set is in large parts digital manual work. This includes reviewing vocabularies and other norm data for thematically suitable categories as well as reading some parts of the data to identify relevant search terms. [16]

My semantic field consisted of all hyponyms that I deemed relevant, drawing from the German National Library's Integrated Authority File (WIECHMANN, 2012). Depending on the topic, several hundred hyponyms are compiled that are impossible to fully reflect upon. If, for example, one creates a semantic field on

food and drink (STIEMER & VAUTH, 2020), extensive lists of foods quickly become available on this data basis. For more specific cases such as telemedicine, terms tend to stand out that would otherwise not have been brought to the researcher's attention, such as telecardiology. These terms are read out from the vocabularies automatically via an interface which issues output in standard formats and is made available by the German National Library. Combining different authority data and manual research at an early stage of research is key to building a comprehensive semantic field on the topic under investigation. In the course of dealing with the topic, semantic fields can grow and become more precise. Especially in large, unknown data sets, researchers can thus identify material that would not be found by using only a few search terms (KOCH & FRANKEN, 2020). [17]

Digital methods are also useful for data collection. In order to extract data sets, researchers need *crawling* and *scraping* processes, i.e., finding, storing and processing relevant Internet data (BARBARESI, 2019). In this step, digital manual work also initially predominates, and the main interest shifts towards generating *big data* computationally at a later stage. To reflexively define starting points of *crawling* in the network, so-called *seed* URLs, individual researchers' knowledge and ability for manual searching and compilation is again necessary. Next in the process is scaling. It is mainly done by an algorithm, that stores text based on the starting points, extracts links from this text, follows the links and again stores the text. In order to conduct thematic *crawling* in a focused manner, it is essential to prepare semantic fields in advance and manually create a list of web addresses that serve as starting points (ADELMANN & FRANKEN, 2020). This method quickly produces large sets of data. Of course, data can also be generated in other ways, for example by accessing the interfaces of social media, so-called *application programming interfaces* (APIs), and storing *tweets* on individual *hashtags*, for example. *Scraping* stores the data that are selected for further processing. This method closely links data selection and collection. In the exemplary study discussed here, websites of health insurance companies, doctors' associations and patients' associations were used as seed URLs. From this, customer journals, scientific studies and political strategy papers, but also technical instructions for doctors' practices and information texts were found. Identifying relevant discourse actors and their positions is central for a sociology of knowledge approach to discourse. Consequently, thematic *crawling* is also suitable for extracting a broad spectrum of positions—provided that previous knowledge has informed the semantic fields and groups of actors that the analysis is based on. However, in the results, the data are no longer or only in small parts organized around the previously identified discourse actors because the links on the web pages do not adhere to the boundaries drawn by researchers. This necessitates new perspectives, for example on how to attribute them, and shows how important digital manual work is during and after computational steps. The obtained data can be further supplemented by data such as tweets or even qualitative interviews. [18]

Using digital methods for analysis requires data to be processed first, especially when textual data is of interest. It is important that large data sets are named and

stored in a structured manner. Depending on the further steps one must also determine which units of the text should be stored separately. Linguistic processing is indispensable for achieving precise results of digital analysis. It is used to identify individual words (*tokenization*), to revert them into their basic form (*lemmatization*) and to classify the words used (*part-of-speech tagging*) or to break them down according to syntactic relations in the sentences (*parsing*). Such *natural language processing* methods originate from computational linguistics and are widely used in DH and CSS (for an introduction see ANDRESEN & ZINSMEISTER, 2019; PUSTEJOVSKY & STUBBS, 2012). For example, word frequencies are only of limited significance without *lemmatization*. In other contexts, it is more interesting to break down textual data by parts of speech or by entities. For instance, this is the case if researchers want to keep traceable that "Jens Spahn" and "Federal Minister of Health" were synonymous with each other—at least for a specific time period. In data processing, unstructured text (from a computational perspective) becomes structured text that is machine-readable and can thus be processed with digital methods. While there are tools for some processing steps that can be easily reused, depending on the question and the data structure it may be necessary to plan for additional efforts for these steps. Digital manual work should not be neglected here either, because researchers make a large number of decisions about methods and implementations in detail during processing, they work with different versions of the data and with corresponding file naming, data filing and overview lists. [19]

Once data that are potentially relevant for the research question have been appropriately prepared, the initial phase of sifting and analysis begins. It can be supported by *text mining* and *distant reading*, i.e., filtering within large amounts of text for relevant passages (IGNATOW & MIHALCEA, 2017; KOCH & FRANKEN, 2020; LEMKE & WIEDEMANN, 2016). Concrete methods include statistical and computer linguistic analyses, ranging from simple frequency distributions to *collocations*. The latter is used to examine words that frequently occur together. Terms from the previously created semantic fields can again be used to work out specific connections. *Topic modeling* (BLEI, 2012; JACOBS & TSCHÖTSCHEL, 2019) can also be used to include terms that commonly occur together. If this is frequently the case, this *machine learning* method assumes that terms belong to a topic. Since this technique is not based on a set of predefined terms, it is particularly suitable for supporting inductive and abductive research approaches (NELSON, 2020), because the result of a *topic modeling* is a grouped list of frequently co-occurring terms without any annotation. This is fundamentally different from statistical methods, where predefining terms are common practice. After using a *topic modeling* tool such as the command-line based "Mallet" ("Machine Learning for Language Toolkit," McCALLUM, 2002), which has become a standard in the DH and CSS, again a substantial amount of digital manual work needs to be undertaken, mainly reviewing and labeling lists. It takes the tool only a few minutes to create the actual *topic term* matrix, i.e., a list of terms assigned to a *topic*, from a text that has been appropriately prepared. What does, however, require a lot of work after the data has been processed is the elaboration and use of the meanings contained in the matrix. In a large amount of unknown texts, *topic modeling* reveals those texts that strongly represent

individual *topics*. Based on the corresponding *topic* lists, researchers know that, for example, terms such as "patient," "data," "consent," "processing," and "health data" often appear together, but do not know in detail what the texts contain. In addition, analytically relevant details such as gendered forms of words are missing. By reducing down to *topics* the terms of which are (or could be) actually related to the relevant research question, this method can be used to identify meaningful texts or parts of them within the data. Relevant texts can subsequently be sifted manually in order to learn more about the thematic focus. *Topic modeling* can thus be used in different dimensions for filtering within textual data. However, as this method works with tables rather than the original texts, all the contexts are missing. [20]

Syntactic annotations are helpful for retracing the contexts (ANDRESEN et al., 2020). Here, linguistic structures are evaluated to determine the concrete context of word usage. This allows for a different view of the (still unknown or not yet intensively examined) data, such as when attributions to already known concepts are made visible. For example, in the sample material, telemedicine was ascribed to terms like "possibility," "enrich" and "enable." However, many of the attributions could not be interpreted without considering the context. In the *web crawling* corpus, telemedicine was attributed to both "ensured" and "endangered." At first sight and without knowing the passage in the texts, this revealed the complexity of the collected data, which obviously contained different discourse positions. Researchers are again directed here to potentially relevant texts and text passages. Syntax profiling allows for a different view of the material, and the results can be used as filters. These techniques deliver substantially more informative results than the mere frequencies that CAQDAS packages help provide under the name of *mixed-methods* approaches to qualitative methods. [21]

Another method for data analysis is considering sentiment values of texts, i.e., classifying word meanings as positive or negative (D'ANDREA, FERRI, GRIFONI & GUZZO, 2015; LIU, 2015). This involves searching for terms with positive and negative connotations in parts of texts, either based on dictionaries established in advance (often by third parties) or on machine learning. Exploratory experiments conducted as part of the discourse analysis that is presented in this paper quickly revealed that in both alternatives there were (too) many ambiguities when it came to categorizing terms as either positive or negative. Hence, computationally assigned meanings yielded limited insight. Only with digital manual work was it possible to come to reliable conclusions. Manually examining the individual computational analyses is the only way to decide whether the use of words in a specific context is actually positive or negative. Similar to *topic modeling*, this raises new methodological questions of documentation and also of quality criteria (STRÜBING, HIRSCHAUER, AYAß, KRÄHNKE & SCHEFFER, 2018) for corresponding analytic practices. [22]

Manual annotation, i.e., highlighting central parts of the text and then developing categories inductively (FRANKEN, KOCH & ZINSMEISTER, 2020) is essential to analysis and computers can only offer limited support. While annotations can be carried out computationally at the word level, as in the example given in this

section in which all mentions of "Jens Spahn" and "Federal Minister of Health" as well as further synonyms for him can be marked automatically, both brainwork and digital manual work are required during and after digital filtration practices to analyze structures of meaning. Annotation tools digitally support the step that is otherwise realized with pen and paper. For manual annotation, there is still no *open-source* tool that caters well to the needs of the humanities, cultural studies, and social sciences. Consequently, proprietary solutions from CAQDAS are generally used here, even though this contradicts the principles of *open science*. [23]

Digital methods must be differentiated according to their field of application. *Topic modeling*, for example, is well suited to identifying broad sub-topics. *Sentiment analysis*, on the other hand, is related to the specific text and can therefore be used to detect special features. Methods that engage with data in a more distant manner are suitable for early stages of analysis, while methods that do so more closely are a better fit for late phases. The methods presented as examples show how data selection, collection and analysis can be supported and extended by digital methods. Qualitative researchers have to be keen to switch between methods and data in order to arrive at well-founded conclusions. Possible findings are heavily reliant on the selected approaches and how they are arranged. A priori assumptions are present in all steps and often already inscribed in tools, methods and approaches—they have to be made transparent and questioned again and again. [24]

There are of course other methods, such as in the context of *natural language processing* (as in the application of BAKER [2012] or NELSON [2020]). Visualizing data sets can also be used as a basis for interpretation (DRUCKER, 2020; SVENSSON, 2016). The context of the particular research question is crucial for assessing the applicability to research interests in the humanities, cultural studies, and social sciences. For other sorts of data, one could set up a model that is comparable to the one for texts described in this paper. The analysis of images, for example, is already well developed, but it is even more difficult for audiovisual data (attempts can be found, for example in LEGEWIE & NASSAUER [2018]). Currently, a multimodally combined data base usually means that different analyses have to be conducted simultaneously because there are hardly any methods that could, for example, computationally process text-image combinations. [25]

5. Challenges and Potentials

What are the challenges, but also the potentials, of extending research processes in the humanities, cultural studies and social sciences with digital data and methods? As has been shown, not all steps of the research process can be extended with digital methods. The parts that are determined by epistemology still need to be processed manually by the researchers. [26]

5.1 Challenges

I see five key challenges in using digital data and digital methods as a toolbox, as well as developing them further in the sense of an experiment kit. First, corpus generation is not yet satisfactory for all research areas. DH and CSS often work with process-produced data (Section 2.2). Studies are usually limited to one type of data. However, the strength of qualitative empirical disciplines, as well as history or geography, lies precisely in combining different types of data. The step of corpus creation thus faces specific challenges in finding and storing the different data as well as in the inclusion of diverse and linked research-induced data. [27]

Second, this is connected to the fact that the DH research community in particular is highly centered on written language. For CSS, data from social media play an important role, but they rarely consider connections, for instance between text and images. Research is severely hampered as long as image or audio data are still analyzed separately from text. Heterogeneous and multimodal data need to be included to a larger extent than they are now. [28]

Third, it is central to ask about the DH methods that are applied. The choice of approaches and their order determine possible findings and are a challenge in the modeling and realization of the research process. It is usually not sufficient to limit oneself to one method. In fact, research perspectives are extended and broadened by combining different approaches that are engaging with data more closely as well as in a more distant manner. Research questions change in the course of projects, and methodological challenges and possibilities often only become apparent along the way. [29]

Fourth, using digital methods changes the research process modeled above in terms of data selection, collection, processing, and analysis. Instead of an individual arrangement of the data, iteratively adjusted and increasingly overlooked in their complexity, a non-linear, granular perspective on the material is adopted. Insights into manually unmanageable data settings are structured primarily by the outputs of algorithmic methods. This is another reason why it is always necessary to manually sift through data, for example to read linear text, in order not to lose sight of the overall view and depth of it. BAKER and LEVON (2015) also noted that results of digital methods are still too often seen as neutral and as a countermeasure to the alleged *cherry-picking* of qualitative social research. [30]

The fifth challenge lies in the pluralities of meaning that are inherent to data from the humanities, cultural studies, and social sciences and their interpretation, but which to date can hardly be captured digitally. Approaches to qualitative research are particularly characterized by the fact that no unambiguous answers exist, and in many cases, these are not even sought. Formal categories, which are often central in digital methods, are not always (easy) to find for these questions. Accordingly, there remains a great need for further development in the experiment kit of digital methods. [31]

The reflexive level is transverse to these challenges, which is emphasized in DH (GEIGER & PFEIFFER, 2020) but still needs more development. Qualitative researchers are well prepared for these issues with their reflections on methods and theory building. The turn to digital methods requires specific critical perspectives on data and code structures, on their conditionality and operationalizability, on limitations of the corresponding research and possibilities of its expansion which may even be prevented by a lack of reflection. In this regard, it is necessary to ask about the digitized hermeneutic production of knowledge that has been altered by the application of digital methods and the sections of reality that are addressed with it. [32]

5.2 Potentials for further development

In Sections 4 and 5.1, I have already implicitly mentioned potentials because the challenges described there are also possible avenues for further development. First, by using standards and vocabularies, corpora can be compiled in a more informed way. Data thus become easier to find, more comparable, and more qualitatively accessible. A particularly promising approach is annotating: It can serve as a link between processing and analyzing a corpus. Many sub-steps can already be automated, for example through *named entity recognition* in which entities such as persons or places are automatically annotated (IGNATOW & MIHALCEA, 2017, p.61). Computational annotations of, for instance, *collocations* can be used to make visible structures that evoke manual annotation and thus interpretation. [33]

Second, further developing semantic search technologies and image recognition for combined analyses has great potential with regard to analyzing complex relational phenomena and decoding meanings. Using semi-automatic methods that actively include *humans in the loop* lends itself to be used for multimodal data bases. Thus, in the sense of a human-computer interaction, humans and computers repeatedly interact with each other in loops (WOLETZ, 2016). Multimodal data bases, among other factors, drive such perspectives and contribute to what I called a methodological experiment kit above: Their purpose is to bring new challenges and the advancements resulting from them to digitally working cross-sectional disciplines. [34]

Third, using digital methods is advantageous if it is reflected upon and applied in appropriate steps of the research process. This occurs by drawing on the methodical toolbox. The use of integrative work environments could further

support the process of iteratively jumping back and forth between data and methods in order not to get lost between the different tools and data sets. Despite the need for a flexible and user-friendly tool that allows for free combination and comprehensible usage of different methods, the development of such environments is still pending. Provided that complexity has been appropriately reduced and that the research process remains traceable, it can be assumed that such a tool would lower the threshold to use digital methods. In this way, existing digital methods approaches could become a part of a wider field of qualitative research. In addition, developing such a software could close the gap between digital data and their processing in research data management since at present, the long-term storage of (possibly) re-usable research data is still too disconnected from the actual steps of data selection, collection, and analysis. [35]

Fourth, changing the process of analysis can lead to new perspectives on the data. Digital methods allow for more targeted sampling and potentially different insights because granular perspectives that are both closely and distantly engaging with data can be used to make more informed decisions about which parts of the research field should undergo deeper analysis. At the same time, analyses using digital methods extend the scope of potentially relevant data and provide insight into data sets that would be unmanageable when working in a purely qualitative manner. Even without an integrated work environment there are already many tools that can perform individual work steps or offer support for them. [36]

Fifth, the most comprehensive dimension in which the methodological experiment kit can be used and further developed for qualitative research is the pluralities of meaning. Initial approaches already exist in this area. An example for this would be *embeddings*, i.e., vectors that represent neighboring terms in their specific context. They can capture polysemy, i.e., diversity of meanings, and linguistic synonyms based on how they are used in the specific sentence (WIEDEMANN, REMUS, CHAWLA & BIEMANN, 2019). This can be an approach to digitally recognize meanings and support manual retrieval of similar utterances by means of the contexts in which they are used. [37]

Again, the reflexive level plays an important role in all five of these potentials. It is important to acknowledge what may not be found with digital methods and what can perhaps still only be found with digital manual work, with the researchers' contextual knowledge and their specific view of the data represented in relations and contexts. Data are not neutral, nor should they be a black box. They are incomplete and remain messy, they are alive, and in many cases, they are also broken or uninterpretable (PINK, RUCKENSTEIN, WILLIM & DUQUE, 2018). Data are embedded in a web of meanings and different agency and therefore can only be understood in their discursive constructions and materialities. Besides the data, the choice of analytical methods and tools determines which questions qualitative researchers can ask of the material. If, for example, as in the discourse analysis presented above, the data are not elaborately prepared for a network analysis, this perspective on the data remains hidden. It follows that: "To some extent, tools and infrastructure shape the questions we can ask"

(SVENSSON, 2016, p.16). For the analysis of such boundaries, *science and technology studies* offer points of connection especially with *critical data and code studies*, in which data and algorithms are critically examined for their efficacy (BOYD & CRAWFORD, 2012; KITCHIN & LAURIAULT, 2018). Greater integration of these perspectives into the practice of digital methods development and usage can be beneficial for both sides. [38]

6. Conclusion: Digital Data and Methods Extend Qualitative Research Processes

What does the extension of qualitative research processes look like in terms of digital data and methods that are currently subsumed under the terms DH or CSS? It is necessary to use concrete methods from the methodic toolbox and to develop them further to cater to the needs of qualitative researchers with the methodological experiment kit. In doing so, the limits of digital methods have to be kept in mind and it has to be reflected what is possible and productive. [39]

The use and further development of digital methods and tools extend qualitative research processes. Creatively integrating new elements into the research process is particularly important for problem solving from and in DH and CSS. These new approaches are more than service providers. It must be clear, however, that most questions cannot be fully addressed by digital approaches. There is nevertheless the potential to not only support analyses, but to develop them further. Even if the intensive pervasion of the data in a hermeneutic understanding as the core step of qualitative analyses can be built on digital methods more frequently in the future, the focus will continue to be on digital manual work. Digital methods can facilitate this in many places. At the same time, integrating digital methods can provide for the expansion of qualitative research, especially with the principles of sorting out uninteresting and finding particularly relevant material. [40]

There is also untapped potential in the better combination of quantitative and qualitative analysis. Much of this is already common practice in quantitative research because it uses other types and amounts of data. Therefore, it is worth taking a fresh look at old discussions around *mixed methods* (MORSE & NIEHAUS, 2016). Quantitative and qualitative research are at different stages of a shift in the research processes towards datafication in a society that is increasingly mediated or made possible by data (DOURISH & GÓMEZ, 2018). In quantitative research, the stronger change lies in the data basis towards *trace data* and the reuse of data sets that are created unintentionally; in qualitative research, it is more the access to the data itself that is changing. With digital methods, large amounts of data can be processed, which changes data collection in particular. Qualitative methods and digital manual work can then be applied to data that turned out to be particularly relevant during previous analyses. This extension can be seen as a further development of *mixed methods* approaches. I assent to the view of geographer and big data researcher Rob KITCHIN (2014b) who proposes a *data-driven science* that combines deductive, inductive, and abductive elements. Such a combination makes it possible to apply the methods

described above to large amounts of data in order to generate new hypotheses from the material—i.e., in a *data-driven* way. Subsequently, these then need to be further validated or falsified. [41]

Only if the existing possibilities and limits currently offered by digital data and methods to qualitative research are specifically explored can a change in the methodological approach be advanced. Researchers must be aware that a substantial amount of digital manual work is involved in these steps - and will probably continue to be involved. The promise of *big data* analyses providing a well-founded analysis of data at the push of a button contradicts the quality criteria of the humanities, cultural studies, and social sciences. However, digital manual work must not be perceived as a burden to be minimized, but as a necessary and purposeful engagement with one's own data. Digital manual work is just as much a part of scientific work as the conceptual and intellectual discussions and cannot or should not be done by an algorithm. The inseparable connection of the reflexive level with the data and methods becomes apparent here. [42]

Qualitative research comes with special requirements for DH and CSS that are coined not only by the way qualitative research is defined but also the specific questions, data, and methods as well as the researchers' needs. Qualitative researchers can contribute to interdisciplinary contexts in multiple ways: They bring new perspectives on multimodal data and other uses for established methods, and they reflect on how knowledge is produced under new digital circumstances (FRANKEN, 2022a). As digital data and methods become increasingly important in research fields as well as in science policy, the potentials of digital methods should be brought into the breadth of qualitative discussions. Only if qualitative researchers actively shape these processes will these be further developed in a suitable way and according to the accompanying research logics. Qualitative researchers across the board must therefore ask themselves what using digital data and methods means for their research processes and methods, their research fields and objects, and their research questions and epistemologies. [43]

Acknowledgments

For constructive discussions and the support regarding the technical implementation of the exemplary study, I would like to particularly thank Benedikt ADELMANN, Melanie ANDRESEN, Evelyn GIUS, Wolfgang MENZEL, Michael VAUTH and Heike ZINSMEISTER. Darina HASHEM and Susanne HOCHMANN actively supported the work on the data basis of the study. Anne DIPPEL, Moritz FÜRNEISEN, Gunther HIRSCHFELDER, Isabella KÖLZ, Hannah ROTTHAUS, Sarah THANNER and Libuše Hannah VEPREK as well as the discussants of the colloquium at the Department of Sociology at the LMU Munich contributed valuable food for thought and enriching comments on earlier drafts. I would like to thank the anonymous reviewers and the editors of *FQS* for their numerous helpful comments and questions. Nils EGGER, Katharina LILLICH and Florian SCHMID provided valuable support in finalizing the textual version.

References

- Adelmann, Benedikt & Franken, Lina (2020). Thematic web crawling and scraping as a way to form focussed web archives. In Sharon Healy, Michael Kurzmeier, Helena La Pina & Patricia Duffe (Eds.), *Book of abstracts. Engaging with web archives: "Opportunities, Challenges and Potentialities" (#EWAVirtual)*, Maynooth University Arts and Humanities Institute, Kildare, Ireland, September 21-22 (pp.35-37), <https://doi.org/10.5281/zenodo.4058013> [Accessed: January 13, 2023].
- Adelmann, Benedikt; Franken, Lina; Gius, Evelyn; Krüger, Katharina & Vauth, Michael (2019). Die Generierung von Wortfeldern und ihre Nutzung als Findeheuristik. Ein Erfahrungsbericht zum Wortfeld "medizinisches Personal". In Patrick Sahle (Ed.), *Book of abstracts. 6th conference Digital Humanities im deutschsprachigen Raum e.V. (DHd 2019) "Digital Humanities: multimedial & multimodal"*, Universität Mainz/Universität Frankfurt/M., Germany, March 25-29 (pp.114-16), <https://zenodo.org/record/4622122> [Accessed: October 25, 2021].
- Andresen, Melanie & Zinsmeister, Heike (2019). *Korpuslinguistik*. Tübingen: Narr Francke Attempto.
- Andresen, Melanie; Begerow, Anke; Franken, Lina; Gaidys, Uta; Koch, Gertraud & Zinsmeister, Heike (2020). Syntaktische Profile für Interpretationen jenseits der Textoberfläche. In Christoph Schöch (Ed.), *Konferenzabstracts. 7th conference Digital Humanities im deutschsprachigen Raum e.V. (DHd 2020) "Spielräume. Digital Humanities zwischen Modellierung und Interpretation"*, Universität Paderborn, Germany, March 2-6 (pp.219-223), <https://zenodo.org/record/4621932> [Accessed: October 25, 2021].
- Arbeitsstelle Kleine Fächer (2018). *Was ist ein kleines Fach?*, <https://www.kleinefaecher.de/kartierung/was-ist-ein-kleines-fach.html> [Accessed: October 25, 2021].
- Baker, Paul (2012). Acceptable bias? Using corpus linguistics methods with critical discourse analysis. *Critical Discourse Studies*, 9(3), 247-256.
- Baker, Paul & Levon, Erez (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication*, 9(2), 221-236, <https://doi.org/10.1177/1750481314568542> [Accessed: October 25, 2021].
- Barad, Karen (2003). Posthumanist performativity. Toward an understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society*, 28(3), 801-831.
- Barbaresi, Adrien (2019). The vast and the focused. On the need for thematic web and blog corpora. *Proceedings, Workshop "Challenges in the Management of Large Corpora" (CMLC 7)*, Cardiff, UK, July 22, <https://doi.org/10.14618/ids-pub-9025> [Accessed: October 25, 2021].
- Bauernschmidt, Stefan (2014). Kulturwissenschaftliche Inhaltsanalyse prozessgenerierter Daten. In Christine Bischoff, Karoline Oehme-Jüngling & Walter Leimgruber (Eds.), *Methoden der Kulturanthropologie* (pp.415-430). Bern: UTB.
- Baur, Nina & Graeff, Peter (2021). Datenqualität und Selektivitäten digitaler Daten. Alte und neue digitale und analoge Datensorten im Vergleich. In Birgit Blätzel-Mink (Ed.), *Gesellschaft unter Spannung. Verhandlungen des 40. Kongresses der Deutschen Gesellschaft für Soziologie 2020*, Berlin/digital, Germany, September 14-24, https://publikationen.sozioogie.de/index.php/kongressband_2020/article/view/1362 [Accessed: March 13, 2022].
- Beck, Stefan (2009). Vergesst Kultur – wenigstens für einen Augenblick! Oder: Zur Vermeidbarkeit der kulturtheoretischen Engführung ethnologischen Forschens. In Sonja Windmüller, Beate Binder & Thomas Hengartner (Eds.), *Kultur – Forschung. Zum Profil einer volkskundlichen Kulturwissenschaft* (pp.48-68). Berlin: LIT.
- Beck, Stefan (2019 [2015]). Von Praxistheorie 1.0 zu 3.0. Oder: wie analoge und digitale Praxen relationiert werden sollten. *Berliner Blätter. Ethnographische und ethnologische Beiträge*, 81, 9-27.
- Beck, Stefan; Niewöhner, Jörg & Sørensen, Estrid (Eds.) (2012). *Science and Technology Studies. Eine sozialanthropologische Einführung*. Bielefeld: transcript.
- Berkelaar, Brenda L. & Francisco-Revilla, Luis (2018). Motivation, evidence and computation. A research framework for expanding computational social science participation and design. In Cathleen Stützer, Martin Welker & Marc Egger (Eds.), *Computational social science in the age of big data. Concepts, methodologies, tools, and applications* (pp.16-62). Köln: Herbert von Halem.
- Blätte, Andreas; Behnke, Joachim; Schnapp, Kai-Uwe & Wagemann, Claudius (Eds.) (2018). *Computational Social Science. Die Analyse von Big Data*. Baden-Baden: Nomos.

- Blei, David M. (2012). Probabilistic topic models. Surveying a suite of algorithms that offer a solution to managing large document archives. *Communications of the ACM*, 55, 77-84, <http://www.cs.columbia.edu/~blei/papers/Blei2012.pdf> [Accessed: October 25, 2021].
- Boellstorff, Tom (2016). For whom the ontology turns. Theorizing the digital real. *Current Anthropology*, 57, 387-406.
- boyd, danah & Crawford, Kate (2012). Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- Bubenhofer, Noah (2017). Kollokationen, n-Gramme, Mehrworteinheiten. In Kersten Sven Roth, Martin Wengeler & Alexander Ziem (Eds.), *Handbuch Sprache in Politik und Gesellschaft* (pp.69-93). Berlin: De Gruyter.
- Charmaz, Kathy (2014). *Constructing grounded theory. Introducing qualitative methods*. Los Angeles, CA: Sage.
- Cioffi-Revilla, Claudio (2018). *Introduction to computational social science. Principles and applications* (2nd ed.). Cham: Springer.
- Conte, Rosaria; Gilbert, Nigel; Bonelli, Giulia; Cioffi-Revilla, Claudio; Deffuant, Guillaume; Kertesz, János; Loreto, Vittorio; Moat, Suzy; Nadal, Jean-Pierre; Sánchez, Angel; Nowak, Andrzej; Flache, Andreas; San Miguel, Maxi & Helbing, Dirk (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1), 325-346, <https://doi.org/10.1140/epjst/e2012-01697-8> [Accessed: October 25, 2021].
- D'Andrea, Alessia; Ferri, Fernando; Grifoni, Patrizia & Guzzo, Tiziana (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 26-33, <http://dx.doi.org/10.5120/ijca2015905866> [Accessed: October 25, 2021].
- Denning, Peter J. & Tedre, Matti (2019). *Computational thinking*. Cambridge, MA: The MIT Press.
- Dourish, Paul & Gómez Cruz, Edgar (2018). Datafication and data fiction. Narrating data and narrating with data. *Big Data & Society*, 5(2), 1-10, <https://doi.org/10.1177/2053951718784083> [Accessed: October 25, 2021].
- Drucker, Johanna (2020). *Visualization and interpretation. Humanistic approaches to display*. Cambridge, MA: The MIT Press.
- Evers, Jeanine; Caprioli, Mauro Ugo; Nöst, Stefan & Wiedemann, Gregor (2020). What is the REFI-QDA standard. Experimenting with the transfer of analyzed research projects between QDA software. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 21(2), Art. 22, <http://dx.doi.org/10.17169/fqs-21.2.3439> [Accessed: October 25, 2021].
- Flanders, Julia & Jannidis, Fotis (2016). Data modeling. In Susan Schreibman, Raymond George Siemens & John Unsworth (Eds.), *A new companion to digital humanities* (pp.229-237). Chichester: Wiley Blackwell.
- [Flick, Uwe](#) (2007). *Qualitative Sozialforschung. Eine Einführung*. Reinbek: Rowohlt.
- Fortun, Kim; Fortun, Mike; Bigras, Erik; Saheb, Tahereh; Costelloe-Kuehn, Brandon; Crowder, Jerome; Price, Daniel & Kenner, Alison (2014). Experimental ethnography online. *Cultural Studies*, 28(4), 632-642.
- Franken, Lina (2020a). Kulturwissenschaftliches digitales Arbeiten. Qualitative Forschung als "digitale Handarbeit"? *Berliner Blätter. Ethnographische und ethnologische Beiträge*, 82, 107-118, <https://www2.hu-berlin.de/ifeeoj/index.php/blaetter/article/view/1069/16> [Accessed: October 25, 2021].
- Franken, Lina (2020b). Methodologie der Zukunft? Automatisierungspotentiale in kulturwissenschaftlicher Forschung. In Dagmar Hänel, Ove Sutter, Ruth Dorothea Eggel, Fabio Freiberg, Andrea Graf, Victoria Huszka & Kerstin Wolff (Eds.), *Planen. Hoffen. Fürchten. Zur Gegenwart der Zukunft im Alltag* (pp.217-233). Münster: Waxmann.
- Franken, Lina (2022a). Erweiterungen der Digital Humanities durch kulturwissenschaftliche Perspektiven. In Michaela Geierhos (Ed.), *Konferenzabstracts. 8th conference Digital Humanities im deutschsprachigen Raum e.V. (DHd 2022) "Kulturen des digitalen Gedächtnisses"*, Potsdam/digital, Germany, March 7-11 (pp.101-104), <https://zenodo.org/record/6327985> [Accessed: May 02, 2022].
- Franken, Lina (2022b). Invisible patients? Patients' agency within the discourse on telemedicine. *Curare. Zeitschrift für Medizinethnologie*, 45(2), 101-113.

- Franken, Lina; Koch, Gertraud & Zinsmeister, Heike (2020). Annotationen als Instrument der Strukturierung. In Julia Nantke & Frederik Schlupkoth (Eds.), *Annotation in scholarly editions and research* (pp.89-108). Berlin: De Gruyter, <https://doi.org/10.1515/9783110689112-005> [Accessed: October 25, 2021].
- Gaidys, Uta; Gius, Evelyn; Jarchow, Margarete; Koch, Gertraud; Menzel, Wolfgang; Orth, Dominik & Zinsmeister, Heike (2017). Project description—hermA: Automated modelling of hermeneutic processes. *Hamburger Journal für Kulturanthropologie*, 7, 119-123, <https://journals.sub.uni-hamburg.de/hjk/article/view/1213> [Accessed: October 25, 2021].
- Geertz, Clifford (1999 [1973]). *The interpretation of cultures. Selected essays*. New York, NY: Basic Books.
- Geiger, Jonathan & Pfeiffer, Jasmin (2020). Spielplätze der Theoriebildung in den Digital Humanities. In Christoph Schöch (Ed.), *Konferenzabstracts. 7th conference Digital Humanities im deutschsprachigen Raum e.V. (DHd 2020) "Spielräume. Digital Humanities zwischen Modellierung und Interpretation"*, Universität Paderborn, Germany, March 2-6 (pp.48-51), <https://zenodo.org/record/4621980> [Accessed: October 25, 2021].
- Gibbs, Graham R.; Friese, Susanne & Mangabeira, Wilma C. (2002). Technischeinsatz im qualitativen Forschungsprozess. Einführung zu FQS Band 3(2). *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 3(2), Art. 8, <http://dx.doi.org/10.17169/fqs-3.2.847> [Accessed: October 25, 2021].
- Glaser, Barney G. & Strauss, Anselm L. (2010 [1967]). *Grounded Theory. Strategien qualitativer Forschung*. Bern: Huber.
- Hine, Christine (2015). *Ethnography for the internet. Embedded, embodied and everyday*. London: Bloomsbury Academic.
- Ignatow, Gabe & Mihalcea, Rada F. (2017). *Text mining. A guidebook for the social sciences*. Los Angeles, CA: Sage.
- Jacobs, Thomas & Tschötschel, Robin (2019). Topic models meet discourse analysis. A quantitative tool for a qualitative approach. *International Journal of Social Research Methodology*, 22(5), 469-485.
- Jannidis, Fotis (2017). Grundlagen der Datenmodellierung. In Fotis Jannidis, Hubertus Kohle & Malte Rehbein (Eds.), *Digital Humanities. Eine Einführung* (pp.99-108). Stuttgart: J.B. Metzler.
- Jannidis, Fotis; Kohle, Hubertus & Rehbein, Malte (2017a). Warum ein Lehrbuch für Digital Humanities?. In Fotis Jannidis, Hubertus Kohle & Malte Rehbein (Eds.), *Digital Humanities. Eine Einführung* (pp.XI-XIII). Stuttgart: J.B. Metzler.
- Jannidis, Fotis; Kohle, Hubertus & Rehbein, Malte (Eds.) (2017b), *Digital Humanities. Eine Einführung*. Stuttgart: J.B. Metzler.
- [Keller, Reiner](#) (2011 [2005]). *Wissenssoziologische Diskursanalyse. Grundlegung eines Forschungsprogramms* (3rd ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kinder-Kurlanda, Katharina E. (2020). Big Social Media Data als epistemologische Herausforderung für die Soziologie. In Sabine Maasen & Jan-Hendrik Passoth (Eds.), *Soziologie des Digitalen – Digitale Soziologie?* (pp.109-133). Baden-Baden: Nomos.
- Kitchin, Rob (2014a). *The data revolution. Big data, open data, data infrastructures & their consequences*. Los Angeles, CA: Sage.
- Kitchin, Rob (2014b). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1-12, <https://doi.org/10.1177%2F2053951714528481> [Accessed: October 25, 2021].
- Kitchin, Rob & Lauriault, Tracey P. (2018). Toward critical data studies: Charting and unpacking data assemblages and their work. In Jim Thatcher, Andrew Shears & Josef Eckert (Eds.), *Thinking big data in geography. New regimes, new research* (pp.3-20). Lincoln, NE: University of Nebraska Press.
- Koch, Gertraud & Franken, Lina (2020). Filtern als digitales Verfahren in der wissenssoziologischen Diskursanalyse. Potentiale und Herausforderungen der Automatisierung im Kontext der Grounded Theory. In Samuel Breidenbach, Peter Klimczak & Christer Petersen (Eds.), *Soziale Medien. Interdisziplinäre Zugänge zur Onlinekommunikation* (pp.121-138). Wiesbaden: Springer Vieweg.
- Klug, Helmut W. (Ed.) (2021). *KONDE Weißbuch. Kompetenznetzwerk Digitale Edition*, <https://www.digitale-edition.at/> [Accessed: March 13, 2022].
- Knox, Hannah & Nafus, Dawn (Ed.) (2018). *Ethnography for a data-saturated world*. Manchester: Manchester University Press.

- Kropf, Jonathan (2019). Recommender Systems in der populären Musik. Kritik und Gestaltungsoptionen. In Jonathan Kropf & Stefan Laser (Eds.), *Digitale Bewertungspraktiken. Für eine Bewertungssoziologie des Digitalen* (pp.127-163). Wiesbaden: Springer VS.
- Legewie, Nicolas & Nassauer, Anne (2018). YouTube, Google, Facebook: 21st century online video research and research ethics. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 19(3), Art. 32, <https://doi.org/10.17169/fqs-19.3.3130> [Accessed: October 25, 2021].
- Leimgruber, Walter (2013). Entgrenzungen. Kultur – empirisch. In Reinhard Johler, Christian Marchetti, Bernhard Tschöfen & Carmen Weith (Eds.), *Kultur_Kultur. Denken. Forschen. Darstellen* (pp.71-85). Münster: Waxmann.
- Lemke, Matthias & Wiedemann, Gregor (Eds.) (2016). *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*. Wiesbaden: Springer VS.
- Lewis, Lori (2020). Infographic: What happens in an internet minute 2020. *AllAccess.com*, March 10, <https://www.allaccess.com/merge/archive/31294/infographic-what-happens-in-an-internet-minute> [Accessed: October 25, 2021].
- Light, Ben; Burgess, Jean & Duguay, Stefanie (2017). The walkthrough method. An approach to the study of apps. *New Media & Society*, 20(3), 881-900.
- Lindner, Rolf (2003). Vom Wesen der Kulturanalyse. *Zeitschrift für Volkskunde*, 99, 177-188.
- Liu, Bing (2015). *Sentiment analysis. Mining opinions, sentiments, and emotions*. Cambridge: Cambridge University Press.
- Lupton, Deborah (2015). *Digital sociology*. London: Routledge Taylor & Francis Group.
- MacMillan, Katie (2005). More than just coding? Evaluating CAQDAS in a discourse analysis of news texts. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(3), Art. 25, <https://doi.org/10.17169/fqs-6.3.28> [Accessed: October 25, 2021].
- Manovich, Lev (2020). *Cultural analytics*. Cambridge, MA: The MIT Press.
- Marcus, George E. (2009). Introduction. Notes toward an ethnographic memoir of supervising graduate research through anthropology's decades of transformation. In James D. Faubion & George E. Marcus (Eds.), *Fieldwork is not what it used to be. Learning anthropology's method in a time of transition* (pp.1-34). Ithaca, NY: Cornell University Press.
- Marres, Noortje (2017). *Digital sociology. The reinvention of social research*. Cambridge: Polity.
- Mau, Steffen (2019). *The metric society: On the quantification of the social*. Cambridge, MA: Polity.
- Mayer-Schönberger, Viktor & Cukier, Kenneth (2013). *Big data. A revolution that will transform how we live, work and think*. London: Muray.
- McCallum, Andrew Kachites (2002). *MALLET. A machine learning for language toolkit*, <http://mallet.cs.umass.edu/> [Accessed: October 25, 2021].
- McCarty, Willard (2004). Modeling. A study in words and meanings. In Susan Schreibman, John Unsworth & Raymond George Siemens (Eds.), *A companion to digital humanities* (pp.254-270). Malden, MA: Blackwell.
- McCarty, Willard (2005). *Humanities computing*. Basingstoke: Palgrave Macmillan.
- Morse, Janice M. & Niehaus, Linda (2016). *Mixed method design. Principles and procedures*. London: Taylor and Francis.
- Mosconi, Gaia; Li, Qinyu; Randall, Dave; Karasti, Helena; Tolmie, Peter; Barutzky, Jana; Korn, Matthias & Pipek, Volkmar (2019). Three gaps in opening science. *Computer Supported Cooperative Work (CSCW)*, 28(3-4), 749-789.
- Muri, Gabriela (2014). Triangulationsverfahren im Forschungsprozess. In Christine Bischoff, Karoline Oehme-Jüngling & Walter Leimgruber (Eds.), *Methoden der Kulturanthropologie* (pp.459-473). Bern: UTB.
- Nelson, Laura K. (2020). Computational grounded theory. A methodological framework. *Sociological Methods & Research*, 49, 3-42, <https://doi.org/10.1177/0049124117729703> [Accessed: October 25, 2021].
- Paulus, Trena M.; Jackson, Kristi & Davidson, Judith (2017). Digital tools for qualitative research: Disruptions and entanglements. *Qualitative Inquiry*, 23(10), 751-756.
- Petz, Thorsten (2021). Digitalisierte intime Bewertung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 73, 425-450.

- Pink, Sarah; Ruckenstein, Minna; Willim, Robert & Duque, Melisa (2018). Broken data. Conceptualising data in an emerging world. *Big Data & Society*, 5(1), <https://doi.org/10.1177/2053951717753228> [Accessed: October 25, 2021].
- Pink, Sarah; Horst, Heather A.; Postill, John; Hjorth, Larissa; Lewis, Tania & Tacchi, Jo (2016). *Digital ethnography. Principles and practice*. Los Angeles, CA: Sage.
- Pustejovsky, James & Stubbs, Amber (2012). *Natural language annotation for machine learning*. Beijing: O'Reilly.
- Reckwitz, Andreas (2004). Die Kontingenzperspektive der "Kultur". Kulturbegriffe, Kulturtheorien und das kulturwissenschaftliche Forschungsprogramm. In Friedrich Jaeger & Jörn Rüsen (Eds.), *Handbuch der Kulturwissenschaften. Vol. 3: Themen und Tendenzen* (pp.1-20). Stuttgart: J.B. Metzler.
- Reckwitz, Andreas (2008). Praktiken und Diskurse. Eine sozialtheoretische und methodologische Relation. In Herbert Kalthoff, Stefan Hirschauer & Gesa Lindemann (Eds.), *Theoretische Empirie. Zur Relevanz qualitativer Forschung* (2nd ed., pp.188-209). Frankfurt/M.: Suhrkamp.
- Rehbein, Malte (2017). Informationsvisualisierung. In Fotis Jannidis, Hubertus Kohle & Malte Rehbein (Eds.), *Digital Humanities. Eine Einführung* (pp.328-342). Stuttgart: J.B. Metzler.
- Reichert, Ramón & Richterich, Annika (2015). Introduction. Digital materialism. *Digital Culture & Society*, 1, 5-17, <https://doi.org/10.25969/mediarep/634> [Accessed: October 25, 2021].
- [Reichert, Jo](#) (2013). Die Abduktion in der qualitativen Sozialforschung. Über die Entdeckung des Neuen (2nd ed.). Wiesbaden: Springer VS.
- Rieker, Peter; Hartmann Schaelli, Giovanna & Jakob, Silke (2020). Zugang ist nicht gleich Zugang. Verläufe, Bedingungen und Ebenen des Feldzugangs in ethnografischen Forschungen. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 21(2), Art. 19, <https://doi.org/10.17169/fqs-21.2.3353> [Accessed: October 25, 2021].
- Rogers, Richard (2019). *Doing digital methods*. London: Sage.
- Ruppert, Evelyn; Law, John & Savage, Mike (2013). Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, 30(4), 22-46.
- Sahle, Patrick (2015). Digital Humanities? Gibt's doch gar nicht!. *Zeitschrift für digitale Geisteswissenschaften*, http://dx.doi.org/10.17175/sb001_004 [Accessed: October 25, 2021].
- Sahle, Patrick (2019). *Professuren für Digital Humanities*, <https://dhd-blog.org/?p=11018> [Accessed: October 25, 2021].
- Salganik, Matthew J. (2018). *Bit by bit. Social research in the digital age*. Princeton, NJ: Princeton University Press.
- Schreibman, Susan; Unsworth, John & Siemens, Raymond George (Eds.) (2004). *A companion to digital humanities*. Malden, MA: Blackwell.
- Sismondo, Sergio (2010). *An introduction to science and technology studies* (2nd ed.). Chichester: Wiley-Blackwell.
- Smith, Rachel Charlotte (2013). Designing heritage for a digital culture. In Wendy Gunn, Ton Otto & Rachel Charlotte Smith (Eds.), *Design anthropology. Theory and practice* (pp.177-138). London: Bloomsbury.
- Snee, Helene; Hine, Christine; Morey, Yvette; Roberts, Steven & Watson, Hayley (Eds.) (2016). *Digital methods for social science. An interdisciplinary guide to research innovation*. Basingstoke: Palgrave Macmillan.
- Stiemer, Haimo & Vauth, Michael (2020). "Mit den Himbeeren stimmte etwas nicht". Die Ernährung in der dystopischen Gegenwartsliteratur. In Jennifer Grünwald & Anja Trittelvitz (Eds.), *Ernährung und Identität* (pp.203-222). Stuttgart: ibidem.
- Strübing, Jörg (2019). Grounded Theory und Theoretical Sampling. In Nina Baur & Jörg Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp.525-544). Wiesbaden: Springer VS.
- Strübing, Jörg; Hirschauer, Stefan; Ayaß, Ruth; Krähnke, Uwe & Scheffer, Thomas (2018). Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. *Zeitschrift für Soziologie*, 47(2), 83-100.
- Svensson, Patrik (2016). *Big digital humanities*. Ann Arbor, MI: University of Michigan Press.

- Thaller, Manfred (2017). Digital Humanities als Wissenschaft. In Fotis Jannidis, Hubertus Kohle & Malte Rehbein (Eds.), *Digital Humanities. Eine Einführung* (pp.13-18). Stuttgart: J.B. Metzler.
- Trier, Jost (1973 [1931]). *Über Wort- und Begriffsfelder*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Turner, Victor (2005 [1969]). *Das Ritual. Struktur und Anti-Struktur*. Frankfurt/M.: Campus.
- Universität Klagenfurt (2021). *Humanwissenschaft des Digitalen*, <https://www.aau.at/digital-age-research-center/humanwissenschaft-des-digitalen/> [Accessed: October 25, 2021].
- Unsworth, John (2000). Scholarly primitives. What methods do humanities researchers have in common, and how might our tools reflect this. *Presentation*, Symposium Humanities Computing: Formal Methods, Experimental Practice, King's College, London, UK, May 13, <https://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html> [Accessed: October 25, 2021].
- van Gennepe, Arnold (1999 [1909]). *Übergangsriten*. Frankfurt/M.: Campus.
- Wiechmann, Brigitte (2012). Die Gemeinsame Normdatei (GND). Rückblick und Ausblick. *Dialog mit Bibliotheken*, 24(2), 20-22, <https://d-nb.info/1038598885/34> [Accessed: October 25, 2021].
- Wiedemann, Gregor; Remus, Steffen; Chawla, Avi & Biemann, Chris (2019). Does BERT make any sense? Interpretable word sense disambiguation with contextualized embeddings. In German Society for Computational Linguistics & Language Technology (Ed.), *Proceedings of the 15th Conference on Natural Language Processing* (pp.161-170), https://corpora.linguistik.uni-erlangen.de/data/konvens/proceedings/papers/KONVENS2019_paper_43.pdf [Accessed: April 28, 2022].
- Witt, Harald (2001). Forschungsstrategien bei quantitativer und qualitativer Sozialforschung. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 2(1), Art. 8, <https://doi.org/10.17169/fqs-2.1.969> [Accessed: October 25, 2021].
- Wolletz, Julie (2016). *Human-Computer Interaction. Kulturanthropologische Perspektiven auf Interfaces*. Darmstadt: BÜCHNER.

Author

Lina FRANKEN is professor of digital humanities in cultural studies at the University of Vechta. Before that, she was interim professor of computational social sciences at the Department of Sociology at LMU Munich, Germany. Before that, she was a project coordinator for the collaborative research project "Automated Modelling of Hermeneutic Processes—The Use of Annotation in Social Research and the Humanities for Analyses on Health" at the University of Hamburg's Institute of Anthropological Studies in Culture and History (2018-2021). From 2013 to 2017 she was project coordinator for the project "Digitales Portal Alltagskulturen im Rheinland" [Digital Portal for Everyday Cultures in the Rhineland] at the Landschaftsverband Rheinland that was supported by the German Research Foundation. She got her doctorate in comparative cultural studies at the University of Regensburg in 2017 after studying folklore, early modern history and media studies. Her research interests consist of computational social sciences and digital humanities, methodology and development of digital methods, mechanization and digitization in everyday life and science, cultures of teaching in school education, immaterial cultural heritage as well as work and food cultures.

Contact:

Prof. Dr. Lina Franken

University of Vechta
Cultural Studies
Driverstr 22, 49377 Vechta, Germany

E-mail: lina.franken@ui-vechta.de

URL: <https://www.uni-vechta.de/kulturwissenschaften/lehrende/franken-lina>

Citation

Franken, Lina (2025). Digital data and methods as extension of qualitative research processes: Challenges and potentials coming from digital humanities and computational social sciences [43 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 23(2), Art. 12, <https://doi.org/10.17169/fqs-22.2.3818>.