2. Literature Review

2.1 Why Do We Need Education about Datafication?

2.1.1 Introduction into Critical Data Studies

The Datafication of Our Societies

In recent years, our societies have become increasingly governed by data. With the growing collection of data and use of data analytics, today's governments and corporations have a staggering amount and range of data available – more than any other institution in history before them (Ruppert et al. 2017, p. 1). Even more disruptive, however, are the uses of these data masses. In enabling and powering a large variety of algorithmic systems, data not only pose risks around "privacy, surveillance, social discrimination, personalisation and control" (Kennedy and Moss 2015, p. 5), they also lead to new forms of capitalism (Zuboff 2019) and "a power shift" (Hintz 2022, p. 86), creating "new forms of power relations and politics" (Ruppert et al. 2017, p. 2), and altering the "nature of citizenship" (Graham et al. 2019, para.4). In our "data-driven world" (Kitchin 2021, p. 219), citizens are increasingly "profiled, categorized, scored and assessed" (Dencik et al. 2022, p. 8), but they "lack agency over how their data is used" (Ada Lovelace Institute 2020, p. 7). Considering these wide-reaching implications of data systems on how we make sense of our world, how we govern our societies and organisations, and how democracy is understood (Kitchin 2021; Dencik et al. 2022), many scholars speak of the *datafication* of our societies.

The term datafication was coined by Mayer-Schönberger and Cukier, who explained that datafying a phenomenon means to "put it in a quantified format so it can be tabulated and analysed" (2013, p. 78). They argue that this datafication is happening to our entire societal systems, as more and more information about human behaviour and people's attitudes and sentiments is turned into data points that can be analysed and used for making decisions and predicting future behaviour (ibid.). Mayer-Schönberger and Cukier see datafication not only as the "back bone of many of the Web's social media companies", but also describe it as a "great infrastructure project" that our societies are undertaking (2013, p. 91; 96). Taking a more critical perspective, van Dijck highlights that this datafication of our lives is not an objective

process, but rather "staked in ideological assumptions, which are, in turn, rooted in prevailing social norms", and thus requires critical interrogation (2014, p. 200). In a more recent publication, Hepp, Jarke and Kramp further outline "the double character of datafication's processuality" (2022, p. 5). They argue that datafication describes both the individual translations of social processes into data, which are not as neutral and objective as they are often portrayed, and the overall transformation of society: "how society changes when 'online quantified data' become increasingly widespread" (ibid., p. 6).

In order to understand this datafication in more detail and to outline why more education about datafication is needed, several additional key terms need to be clarified. At their core, data systems consist of *data* and *algorithms*. Neither of these are neutral or benign measures. Data are representations of the world that are "never raw", but always shaped by decisions made in the collection, handling and analysis of the data (Kitchin 2021, p. 5). Algorithms are broadly understood as "a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome" (Diakopoulos 2015, p. 400). This can, for example, be mathematic formulae in the form of equations, or computer programmes that consist of many algorithms that together "form large, often complex, recursive decision trees" (Kitchin 2017, p. 17). In the context of datafication, algorithms usually describe software systems that process data in order to – autonomously or with human involvement – take decisions or apply measures that have impacts on the individual or collective level (Directorate-General for Communications Networks, Content and Technology 2018, p.ii).

In the past decade, new, more complex and extensive "*big data*" systems emerged, that are often defined by their huge volume, high velocity and diverse variety (3V definition, Laney 2001) as well as their exhaustiveness, their fine-grained resolution and their relational and flexible nature (Kitchin 2014). Yet, other scholars have argued that big data is "less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets" (boyd and Crawford 2012, p. 663), and have highlighted big data's sociotechnical nature (e.g., Ruppert et al. 2015; Zuboff 2015). Such critical approaches emphasise that big data is not a technology or a technology effect, but that it "originates in the social, and it is there that we must find it and know it" (Zuboff 2015, p. 75). Academic research further speaks of the "sociotechnical assemblage" that makes up (big) data (Kitchin 2014, p. 24). The notion of *data assemblage* describes the combination of "technological, political, social and economic apparatuses and elements that constitutes and frames the generation, circulation and deployment of data" (Kitchin and Lauriault 2014, p. 1).

All these terminologies aim to describe how today's data systems differ from earlier digital technologies, and how they affect our lives and societies. One type of data systems distinctly illustrates these changes: *automated decision-making systems* (ADMs). These systems consist of a "decision-making model, an algorithm that

translates this model into computable code, the data this code uses as an input […] and the entire political and economic environment surrounding its use" (Algorithm Watch 2019, p. 9). In other words, ADMs use data and algorithms to make decisions that can affect many aspects of our lives and societies, such as who is deemed creditworthy, what social services a person is eligible for and how likely it is that they will commit a crime (e.g., Directorate-General for Communications Networks, Content and Technology 2018, p. 10f; Algorithm Watch 2019).

Critical Data Studies – Research on Risks around Datafication

While many data systems are developed and implemented with good intentions (such as increased efficiency, less subjective decision-making or saving resources), the outcome often does not align with these original intentions (Redden et al. 2022). As the following section will highlight, the datafication of our societies comes with considerable risk and can lead to substantial individual and collective harm. The underlying cause for many of these problematic issues is the misleadingly assumed objectivity and accuracy of data that was already outlined above. Numerous scholars have highlighted that big data is "never a neutral tool" (Dalton and Thatcher 2014), that numbers do not speak for themselves (boyd and Crawford 2012, p. 670), and that claims of big data's extraordinary accuracy need to be treated with caution as "all datasets are necessarily limited representations of the world" (Dalton and Thatcher 2014).

Research from the field of *Critical Data Studies* has played an important role in exploring and developing such arguments. The original call for the field was made by Dalton and Thatcher, asking "what does a critical data studies look like, and why do we care?" (2014). The term "critical data studies" was taken up by Kitchin and Lauriault in the same year (2014) and has since developed into a growing, transdisciplinary research field. Critical Data Studies can be seen as "a response to the increasing spread of digital data and data infrastructures for decision- and meaning-making" (Hepp et al. 2022, p. 6), as the field examines the "unique cultural, ethical, and critical challenges" of data systems (Iliadis and Russo 2016, p. 1) and reflects on how "corporations, institutions and individuals collect and use 'big' data – and what alternatives to existing approaches could look like" (Richterich 2018, p. 8). In its essence, the field is "concerned with the significance (and power) of digital data in contemporary society and how it relates to societal transformation" (Hepp et al. 2022, p. 6).

Residing at the "intersection of science and technology studies, social science, policy and legal fields, and the humanities" (Neff et al. 2017, p. 86), Critical Data Studies is inherently inclusive and "open to self-critique and dialog" (Iliadis and Russo 2016, p. 2). In its transdisciplinary nature, Critical Data Studies further has many overlaps with related research fields, such as *Science and Technology Studies*, which has already existed for several decades, and aims for "understanding and critiquing the

role of technology in our lives" (Moats and Seaver 2019, p. 2). Another closely related field is *Critical Algorithm Studies*, which argues for a need for more insight into algorithms (Gran et al. 2021, p. 1779), and challenges and critiques the many claims of data technologies' "objectivity and legitimacy", instead emphasising that "data science is politics by other means" (Moats and Seaver 2019, p. 2). Many overlaps exist in particular between Critical Algorithm Studies and Critical Data Studies, and the terms are sometimes used interchangeably (e.g., Yeung 2018).

In the past years, critical scholars have highlighted many ways in which datafication impacts our lives, our politics, economics and societies as a whole. Key concerns can be summarised as: "less privacy, more surveillance and social discrimination, and a new means of controlling how publics come to be represented and so understood" (Kennedy and Moss 2015, p. 2). Privacy – considered a key human right in many countries (e.g., Germany, see Deutscher Bundestag 1949) – is endangered through the ubiquitous collection and analysis of citizens' data, which increasingly leads to a "surveillance culture" (Lyon 2017), or "dataveillance – the monitoring of citizens on the basis of their online data" (Van Dijck 2014, p. 205). Datafication transforms and increases surveillance, allowing for "user-generated surveillance" (Lyon 2018, p. 5) through social media platforms and even leading to a "new economic order", as Zuboff argues (2019, no page nr.). This new *surveillance capitalism*, according to Zuboff, "claims human experience as free raw material for hidden commercial practices of extraction, prediction and sales" and constitutes a "profoundly antidemocratic social force" (ibid., p. 513).

Dencik, Hintz, Redden and Treré speak of datafication as a "political-economic regime" and highlight how datafication changes power relations, with citizens being increasingly monitored and profiled, while those who govern have access to new datafied management and governance techniques that citizens have "no knowledge about" (Dencik et al. 2022, p. 7; Hintz 2022, p. 81). Because of this "asymmetric relationship between those who collect, store, and mine large quantities of data, and those whom data collection targets" (Andrejevic 2014, p. 1673), it has been argued that datafication poses "severe challenges for democracy" (Hintz et al. 2022, p. 81) and limits citizens' agency (Baack 2015; Milioni and Papa 2019; Dencik et al. 2022). Data-based surveillance has even been called an "authoritarian and illiberal practice in the digital realm" (Hintz and Milan 2018, p. 3939).

Another implication of the increased surveillance through datafication is the "chilling effect" (e.g., Zuboff 2015, p. 82; Hintz and Milan 2018, p. 3949), which describes "people's sense of being subject to digital dataveillance" that causes them to "restrict their digital communication behavior", for example in the form of self-censorship (Büchi et al. 2022, p. 1). Equally, microtargeting – targeting small groups of online users with specific content – and "hypernudging" constitute datafied processes that can have problematic impacts. By showing each internet user the message that could persuade this individual to make a certain decision, hypernudging

aims to "shape individual decision-making to serve the interests of commercial Big Data barons" (Yeung 2017, p. 119). Such attempts at "behaviour modification" (Zuboff 2019, p. 19) are becoming increasingly common in commercial as well as political advertising, as can be illustrated by the 2018 Cambridge Analytica scandal (see e.g., Cadwalladr and Graham-Harrison 2018), and pose significant risks for democratic processes.

A big part of the problem lies in the *opacity* of data practices. As the algorithmic models behind these systems are usually proprietary, their workings are frequently black-boxed and only rarely "open to public scrutiny and supervision" (Kennedy and Moss 2015, p. 2; Kitchin 2017; Redden 2018b). Thus, most of the time we cannot be certain how these data systems make their decisions or what data they base them on. This is particularly problematic when it comes to flawed and biased data systems. Not only can algorithms "make mistakes and operate with biases" (Diakopoulos 2015, p. 398), but also data that is "drawn from a world that is unequal will reflect that inequality" (Redden 2018b). Moreover, datafication reinforces existing discriminations, for example through search engines that produce racist results (Noble 2018), in consumer finance, where "risky groups" are denied credit, or in social policy, where "specific communities associated with particular behaviours can be targeted for increased surveillance" (Ruppert et al. 2015, p. 2). Scholars have argued that this algorithmic discrimination is the result of the "epistemological foundation" of big data systems: "data fundamentalism, post-explanatory anticipatory pragmatics, and anti-political solutionism" (Prietl 2019, p. 2); and that it tends to disproportionally affect communities that are already disadvantaged (Eubanks 2018), and places where "people, laws, and human rights are the most fragile", such as in the global South (Milan and Treré 2019, p. 319). In light of these social justice concerns that arise through datafication, scholars have called for improved "data justice" (Dencik et al. 2016; Dencik et al. 2022).

However, datafication is not only associated with risks that *may* negatively affect people, but people "are *already* being negatively affected by algorithmic systems" (Redden 2022, p. 60, emphasis in original). Redden has collected such harms and has developed a taxonomy of the negative influences that data systems have on people's lives (ibid., p. 60ff). She found that people have been targeted and exploited based on "inferred vulnerabilities"; that discriminatory social sorting practices have disproportionally disadvantaged people based on their "ethnicity, gender, sexuality and income"; that the "manipulation of information and targeting of voters" has undermined democratic processes and sowed social division; that the automation of government systems has prevented people "from accessing the necessities of life"; and that work surveillance and data breaches have led to physical and mental harms (ibid., p. 71f).

2.1.2   Citizens' Lacking Knowledge and Resignation

Informed Consent?

Another significant problem with the datafication of our societies is people's lack of knowledge and understanding of data collection and algorithmic practices as well as many people's dissatisfaction with these practices when learning about them. Although data collectors like to argue that their users "consent" to these practices, research increasingly shows that this is often not the case. Studies have found that up to 69% of British users accept terms and conditions without reading them (Ofcom 2019, p. 8; see also Meier et al. 2020; Larsson et al. 2021), and that 51% do not understand them when they try to read them (Doteveryone 2018, p. 17). This is not surprising given that most terms and conditions and privacy policies are written in legal jargon and, already in 2008, it would have taken a user around 25 days a year if they read every privacy policy they encountered online (McDonald and Cranor 2008). In fact, Draper and Turow argue that privacy policies deliberately use rhetorical practices of obfuscation to provide a misleading sense of the data that is collected (2019, pp. 7; 9; see below). While the General Data Protection Regulation (GDPR) that came into effect in May 2018 may have led to more transparency on the web, it has not been very effective in fostering users' consent (Degeling et al. 2019). Rather, scholars argue that it may have induced "a false sense of privacy and security for users" as few websites offer their users an actual choice regarding cookie-based tracking (ibid., p. 14), and users tend to assume that the mere existence of a policy protects their privacy (Turow et al. 2018). Overall, the notion of "*meaningful* consent" (Johnson et al. 2017, p. 37, emphasis in original), that is explicit, informed and freely given has been seriously questioned due to internet users' lacking knowledge; their acceptance of cookies because they "feel they have no choice" (Larsson et al. 2021, p. 117); and due to the "dark patterns" of internet services that nudge users toward the "least privacy friendly options" (Forbrukerrådet 2018, p. 4) and implement new "features that cannot be turned off, or settings that are invisible to the user" (Shade and Chan 2020, p. 332).

Lacking Knowledge

Moreover, recent research has increasingly found internet users' knowledge and understanding of data practices to be fragmentary at most. These findings span country borders and various aspects of data collections and analysis: For example, although 94% of British adults have internet access at home and they spend an average of nearly four hours daily online (Ofcom 2022), 24% do not know how technology companies make money, with more than two-thirds (70%) unaware of how free-to-use apps, social media (62%) and search engines (57%) make money from data (Miller et al. 2018, pp. 6; 21). Similarly, less than half of Europe's population knows what an algorithm is or is aware that algorithms are already being used in many areas of life

(Grzymek and Puntschuh 2019, p. 10), and Americans are found to not have "the basic knowledge to make informed cost-benefit choices" about how their data is used by marketers and often base their choices on incorrect information (Turow et al. 2015, p. 16f). Similarly, people's understanding of social media platforms appears to be limited, with 84% of Facebook users not knowing that Facebook categorises their interests (Hitlin and Rainie 2019, p. 8), and many being unaware that an algorithm curates their News Feed (Eslami et al. 2015, p. 153).

These numbers have not increased much over time, with newer studies identifying some increases in people's awareness of data collection and the use of their data between 2018 and 2020, but no significant increase in their understanding of technology companies' business models (Miller et al. 2020). Thus, also in 2020, the British people's understanding "remains shallow" and their ability to shape their online experiences based on their knowledge is poor (ibid., p. 15). Further, a German study found that while people's knowledge has increased in some areas between 2018 and 2022, the number of people who understand the influence of algorithms and artificial intelligence on their lives remains very small (merely increased from 27% to 29%) (Overdiek and Petersen 2022, p. 9). Similarly, Yates et al. described British citizens' knowledge as "patchy", and they found a "very limited understanding" of how technologies work and the "economics and technical ecosystems" that underpin these platforms (2021, pp.xiii, xii). The underlying narrative of these findings seems to be that while people may be aware that their data is being collected online, their understanding of the business models of the internet platforms they use daily remains limited (see also Johnson et al. 2017). This correlates with findings of a recent large-scale international study that found that the majority of internet users do not know how Google or Facebook are free of charge to use (Akman 2022, p. 22). Moreover, the study demonstrated that only a third of respondents knew how a search engine ranks its results and nearly a quarter had "no idea" how this takes place (ibid., p. 26).

Some scholars even speculate that this lacking knowledge and understanding, or "the inability to anticipate the potential uses of such data", may not be "an artifact of laziness or ignorance" but could rather be intentional, constituting a "defining attribute of data-mining processes" (Andrejevic 2014, pp. 1673; 1685). Similarly, Draper and Turow argue that "capitalist systems benefit from the cultivation of resignation as a strategy to neutralize critical or political action" and therefore engage in "obfuscatory communication practices" to hamper people's understanding of data mining processes (2019, pp. 5f; 7; see also below). However, despite the opacity of online platforms, people experience, make sense of and try to imagine how algorithms function, which has been described as people's "algorithmic imaginary" (Bucher 2017, p. 40). These feelings towards data practices have often been found to be negative and uneasy ones, as will be outlined in the next section.

Internet Users Feel Uncomfortable

Contrary to the common argument that users feel like they have 'nothing to hide', research from recent years has repeatedly found that people are concerned about their data, how it is used and how they can protect it (Miller et al. 2018; Selwyn and Pangrazio 2018; Miller et al. 2020; Müller-Peters 2020; Kennedy et al. 2021b). This concern is not dependent on people's knowledge about data practices: studies have found that people's concerns "correlate with both high and low levels of understanding" (Ada Lovelace Institute 2022, p. 20), and even without a good understanding of data practices, people still have strong emotional responses that they are able to articulate clearly (Kennedy et al. 2021a). In fact, many people feel uncomfortable with current data practices, articulating that targeted advertising and state surveillance make them feel "uncomfortable" and "uneasy" (Bucher 2017, p. 39; Dencik and Cable 2017, p. 771), that they find it "intrusive or creepy" (Akman 2022, p. 25), think it is "unacceptable" after learning more about it (Worledge and Bamford 2019, p. 5), and that they (nearly half of participants) feel negatively about receiving it (Miller et al. 2018, p. 15). Looking at Facebook in particular, studies found that people reacted with "surprise and anger" when becoming aware of the curation of their News Feeds through an algorithm (Eslami et al. 2015, p. 153), and that they feel "not very or not at all comfortable" with the way Facebook analyses their interests and traits (Hitlin and Rainie 2019, p. 8).

Moreover, people are unhappy about not being able to better protect their own data online (Miller et al. 2018, p. 16) and they want clearer information about data practices, better regulations and more ethical and responsible technologies (Ada Lovelace Institute 2022). Several studies have identified the creative tactics users apply to resist data collection and profiling, to mislead algorithms and thus to enact small forms of agency (Bucher 2017; Gangneux 2020; Kennedy et al. 2020; Larsson et al. 2021). Besides these clear findings on people's wishes for more understanding and control of their data, studies have also identified the complex and often conflicted and contradictory feelings that people have about how their data is used. People often show "mixed feelings" towards algorithms and data practices, appreciating benefits such as time savings, quality of services, convenience or personalisation, but at the same time voicing their concern, mistrust, anxiety and fears of a "risk of manipulation" (Grzymek and Puntschuh 2019, p. 10; Doteveryone 2018; Ofcom 2019; Overdiek and Petersen 2022; Shade and Shepherd 2013). These mixed feelings may be explained by the different contexts of data practices, as the context of data collection and use, the actors behind these practices and many other factors maintain or undermine trust in data practices (Kennedy et al. 2020). However, in some studies, respondents have both articulated that they have "nothing to hide", but also their "strong concerns about their privacy" (Demertzis et al. 2021, p. 131), or that they dislike being targeted with advertisement, but still would not be willing to pay for the online services they use in exchange for their data (Akman 2022).

Resignation towards Data Practices

This complex picture of people's attitudes towards data practices is further compli-cated when taking into consideration that many people seem to feel resigned to-wards data collection. Various studies have found evidence for such a "resignation" (Turow et al. 2015); "digital resignation" (Draper and Turow 2019); "surveillance real-ism" (Dencik and Cable 2017); "online apathy" (Hargittai and Marwick 2016) or "pri-vacy cynicism" (Hoffmann et al. 2016). Despite some differences in their foci, sam-ples and methods; in their core, these studies examine the same phenomenon. Dig-ital resignation describes a situation where users believe that any efforts to protect their data are futile because data collection – and therefore the possibility of surveil-lance – is inevitable, and they assume that they can never gain control over their data. Importantly, these users do not consent to datafication, nor are they indif-ferent about it (see e.g., Kennedy et al. 2020, pp. 24; 48) or see data collection as a "tradeoff for benefits they receive" (Turow et al. 2015, p. 3), but they rather feel *pow-erless*. Because resigned users can show privacy behaviours that are similar to those who feel indifferent about big data practices, "resignation can obscure signals that people care deeply about privacy" (Draper and Turow 2019, p. 11). Instead, data col-lection is seen as negative and undesirable *because* people still value their privacy and are concerned about surveillance. These findings correspond with other studies on privacy and data collection that emphasise people's concerns around privacy and a "lack of control" (see above; European Commission et al. 2015, p. 115).

Internet users' resignation toward the collection of their data has further been identified in a number of empirical studies. For example, Worledge and Bamford found that a key reason why nearly nine in ten participants did not click on advert preference messages was that they felt not "able to do anything about it" (2019, p. 12). Other studies have found that people "accept cookies even though they do not want to" because they feel that they "have no choice" (Larsson et al. 2021, p. 117), and that people feel "resigned but also 'uneasy'" about data tracking (Yates et al. 2021, p.xii). Resignation might even increase over time, with Doteveryone's 2018 report finding that 25% of British internet users felt like "there's no point" in changing their privacy settings as "companies will get round them anyway" (Miller et al. 2018, p. 17), and their 2020 report identifying "a strong feeling of resignation" in the British public, with two-thirds feeling that they "don't have any say in what technology companies do" (Miller et al. 2020, p. 24). Digital resignation scholars argue that internet users' resignation derives from a perceived necessity of using the internet and social me-dia (Hargittai and Marwick 2016, p. 3751) and fear of "significant social and economic penalties" in case of opting out (Turow et al. 2015, p. 9) in combination with uncer-tainty of how data systems work (Dencik and Cable 2017, p. 770), concerns about privacy violations (Hargittai and Marwick 2016, p. 3749), and lacking control over their data (Turow et al. 2015, p. 20).

Resignation as a Reasonable Response?

Many scholars argue that this reasoning is in fact "not paradoxical", but "a pragmatic response to the contemporary networked social environment" (Hargittai and Marwick 2016, p. 3752f). Thus, resigned or cynical attitudes are seen as a "coping mechanism" that is developed in order "to avoid cognitive dissonance" (Hoffmann et al. 2016, p. 2), and that constitutes a "rational response to a seemingly inevitable outcome" – consumer surveillance (Draper and Turow 2019, p. 5). Another explanatory model explains resignation with the power of imagination, arguing that the "sheer ubiquity of surveillance infrastructures" makes it difficult to think they can be challenged, and that our "hampered imagination" lets datafication and surveillance be seen as inevitable (Dencik 2018, p. 37f). In short, resignation scholars seem to agree that this attitude constitutes a reasonable and pragmatic response to the apparent inevitability of intrusive data practices.

Some authors even go further and argue that resignation not only constitutes a rational response to data practices, but that companies actively foster resignation among internet users through their obfuscatory practices (Draper and Turow 2019, p. 7). Examples of this "placation, diversion, jargon, and misnaming" are incomprehensible privacy policies and incomplete transparency initiatives (ibid., p. 9). This reiterates Andrejevic's argument that internet users' lacking knowledge constitutes a defining attribute of data mining processes and supports the notion that it might be in the digital platforms' interests to keep users' knowledge of data practices limited (2014; see above). However, the last years have seen a shift in technology companies' approach to privacy and some big technology firms now use data privacy as a "strategic competitive advantage", as examined in a recent Master's Thesis (Trindade 2020). Moreover, it is uncertain whether resignation leads to more data disclosure based on the belief of its inevitability and perhaps a perceived lack of technical ability (as suggested by, among others, Hoffmann et al. 2016; Dencik and Cable 2017), or rather to active resistance against data collection by engaging in privacy-protecting behaviour, avoiding certain internet services, and applying "a variety of imperfect, but creative, social strategies" to maintain at least a certain degree of control over one's data (Hargittai and Marwick 2016, p. 3753; see also Draper and Turow 2019; Yates et al. 2021).

2.1.3   Working towards a More Informed Citizenry

Calls for More Literacy and Public Involvement

In light of these problematic findings on internet users' resignation as well as the "major understanding gap around technologies" (Doteveryone 2018, p. 5) that was outlined throughout this chapter, many calls for more *education* about data and more *societal and public involvement* have been made (e.g., Zuboff 2015; O'Neil 2016; Marwick and Hargittai 2018; Miller et al. 2018; Müller-Peters 2020). On the one hand, scholars

have argued that our democratic institutions need to respond to datafication by "developing a means to make black-boxed processes and the implications of changing systems open to wider debate and intervention" (Redden 2018a, p. 10). On the other hand, the need for a "greater public awareness" that might "open up discussion of policy solutions to regulate such [data] practices" has been emphasised (Marwick and Hargittai 2018, p. 14). One way to work toward such greater awareness is *data activism*.

Data Activism's Role in Challenging Datafication and Raising Awareness

The term data activism is used to describe "new forms of civic engagement and political action" that take a critical approach towards datafication (Milan and van der Velden 2016, p. 58). Data activism constitutes a "theoretical construct grounded in empirical observations" (Milan and Gutiérrez 2015, p. 121). Scholars have highlighted that data activists play an important role not only in working toward more socially just datafication processes, but also in acting as "mediators" between data and ordinary citizens, as "literacy promotion" is seen as one of the key tasks of data activists (Beraldo and Milan 2019, p. 5). Examples for data activism practices include the use of open data to foster citizens' understandings and their agency (e.g., Baack 2015; Meng and DiSalvo 2018); engaging citizens in political matters related to data (e.g., Baack 2018); and projects that raise awareness about data practices (e.g., Milioni and Papa 2019), also through unusual approaches such as using "theatre and performance as instruments of critical data studies" (Windeyer 2019, p. 316).

These examples already hint at an important distinction in data activism projects: data activism that *applies data practices* and activism *about big data* (which does not necessarily take place digitally, see Windeyer 2019). The first type, "proactive" data activism, uses data and data technologies from a bottom-up perspective to "provoke social change", to offer alternative narratives and empower citizens (Milan and van der Velden 2016, pp. 66; 67). The second type, "re-active" data activism, perceives massive data collection by state and industry as a threat to our "values, freedoms and activities" and encourages practices of resistance, such as "self-defence, civil disobedience and disruption" as well as using "technical means like encryption or anonymity networks" to prevent monitoring from state or corporation (Milan and van der Velden 2016, p. 67).

Despite their different approaches, both types of data activism conflate to various degrees and constitute "two facets of the same phenomenon, which has data and information at its core" (Milan and Gutiérrez 2015, p. 123). Moreover, both contain an explicit *pedagogical dimension* and often aim to foster citizens' agency (Baack 2018, p. 51f), critical reflection and "critical imagination" of alternative data narratives (Milan 2017). This emphasises data activism's relevance for my study. Moreover, one of the most similar empirical studies to my study comes from the field of data activism: Milioni and Papa's analysis of eight online data activism case studies combines data

activism with affordance theory (2019). The authors highlight how these online resources provide "hidden affordances", "new affordances", "meta-affordances", and "anti-affordances" that aim to empower users and foster their critical understanding (ibid., pp. 7–10). Several of the analysed tools also fall into my study's definition of critical data literacy resources. Moreover, Milioni and Papa warn against shifting "responsibility from data industries and regulatory agencies to the 'enlightened' user" (ibid., p. 11f) – a highly relevant point that will be further discussed in the next chapter.

Calls for More Literacy about Datafication

A second key approach of working towards greater public awareness of data practices and more societal and public involvement is to *promote education about datafication*. The hope behind this is that more "reflexive, active and knowing publics" (Kennedy and Moss 2015, p. 1) might not only empower citizens, but also lead to a much-needed broader public discussion about "the transparency, accountability and oversight of data systems required for protecting citizens' rights" (Redden 2018b). When outlining ways to better educate the public about data practices, scholars have suggested a number of ideas and concepts, which will be analysed in detail in the next chapter. Many highlight the need to support "digital education" and "digital understanding" to foster public engagement (Miller et al. 2018, p. 6; Müller-Peters 2020, p. 128) and people's "data sovereignty" (Kitchin 2021, p. 223). Scholars have also argued that digital literacy programs should move away from individualistic approaches and should rather "find ways to highlight the societal implications" of the structural levels of datafication (Turow et al. 2018, p. 475). Already in 2015, Turow, Hennessy and Draper succinctly summarised what is needed:

> We need initiatives that will give members of the public the right and ability to learn what companies know about them, how they profile them, and what data lead to what personalized offers. We also need to get people excited about using that right and ability. (2015, p. 21)

In a similar vein, others have called for a "broad campaign of popular capacity-building" that builds competences for an informed debate about datafication by enabling "people to form considered opinions and debate the issue in a factually informed way" (Grzymek and Puntschuh 2019, p. 11). These authors further suggest a "Bundeszentrale für algorithmische Kompetenz" (German for: Federal Agency for Algorithmic Literacy), a public institution that works towards informing and educating citizens about algorithmic systems (ibid., p. 32). Moreover, Iliadis and Russo emphasise that educating people about data practices in fact constitutes one of three basic principles of Critical Data Studies: "the identification of social data problems, the design of critical frameworks for addressing social data problems, and the ap-

plication of social solutions to increase data literacy" (2016, p. 5). Based on Aristotle's Nicomachean Ethics, they refer to the notion of "education for the common good", which, they argue, deeply informs Critical Data Studies (ibid.). In their view, this entails that scholars from this field should contribute to data literacy efforts and provide individuals "with the necessary tools for becoming more informed and the ability to organize efforts around data justice issues" (ibid.). Thus, it can be said that responsibilities to foster better understanding of data practices are seen to lie not only with governments and public institutions, but also with academia.

More Knowledge = More Concern?

In recent years, a growing number of studies has examined the relation of people's knowledge about datafication and their concern about data practices, in other words: does more knowledge and understanding entail a more concerned attitude? Some studies have compared the levels of concerns of more and less knowledgeable internet users and have found that those who are more knowledgeable are "more likely to have negative attitudes" towards data uses (Kennedy et al. 2021b, p. 9) and demonstrate "higher privacy concerns and were more likely to take privacy protection strategies on Facebook" (Xie and Karan 2019, p. 10). This suggests that "change is needed, because when people know about data uses, they don't like what they know" (Kennedy et al. 2021b, p. 9). These findings are strengthened by several before-and-after studies, which found that the *mere mention* of privacy in newspaper articles led to less disclosure of personal information, arguing that people's privacy concerns may be dormant and only manifest when asked to think about privacy (Marreiros et al. 2017, pp. 3; 10); or that Facebook users who learned about the platform's curation algorithm showed lasting changes in their usage patterns, an increased "algorithmic awareness", and felt overall more in control (Eslami et al. 2015, p. 153).

Some studies have further identified diverse effects of educating about data practices. My own previous small-scale longitudinal study examined the effect of online critical data literacy resources on university students' privacy concern and behaviour (Sander 2020a). The findings demonstrated a nearly unambiguous increased concern for privacy and distinctly more privacy-sensitive internet usage one week after the intervention, and diverse developments after eight months, with some participants showing concern only about *some* aspects of privacy, one 'defaulting back' to their original attitude and behaviour, and two others showing a *persistent and even growing* increase in privacy concern and behaviour (ibid., p. 5ff). Pangrazio and Selwyn's study with young social media users, on the other hand, found diverse attitudes of the participants regarding different areas of data use (2018). While participants developed "clearer understandings" and new concerns on how their data is used, they were less concerned about text and image analyses and their attitudes to targeted advertising hardly changed (ibid., p. 5f). This last finding

is somewhat contrary to a British nationally representative study by Worledge and Bamford, who observed distinct changes in people's perceptions of online advertising and its strategies, with 43% viewing these practices as unacceptable after learning more about how they work (14% before intervention) (2019, p. 19).

Overall, these studies strengthen the notion that critical data literacy efforts can be effective in educating internet users, and that for many internet users, more knowledge about data practices leads to more concern and feelings of uneasiness. While the findings regarding effects on people's *concerns* differ slightly in the different studies, the results on participants' altered *behaviour* – for some even in the longer term – seem consistent. Although more research on the relation between knowledge, concern about data and behaviour is needed, these first studies clearly support the calls for more education on datafication outlined above and reemphasise the importance of literacy in this field.

2.1.4   Conclusion

This chapter reviewed key literature from the Critical Data Studies field, which outlines the societal transformation processes and risks related to datafication. Moreover, studies that highlight citizens' lacking knowledge on these processes and risks were presented along with initial findings that suggest that more knowledge on data practices can lead to more concerns and a more empowered internet usage. In light of this, many scholars have called for more and better education about data technologies in recent years, and many new data literacy conceptualisations have been suggested.

2.2 Educating about Data: (Critical) Data Literacies

This chapter reviews existing approaches to (critical) data literacy and offers a tentative categorisation of such approaches. Calls for more education about data practices are increasing and educational responses have even been put forward as not only the dominant, but also the "most plausible and successful strategy to combat the challenges of datafication" (Pangrazio and Sefton-Green 2020, p. 212). Scholars further highlight that education about datafication should be seen as a key component or even prerequisite of legal and tactical responses to datafication (ibid., p. 218).

These claims correlate with an upsurge of data literacy conceptualisations, initiatives, projects, networks and curricula in recent years.[1] However, while the term data literacy may seem self-evident, a closer examination of the data literacy discourse reveals great disparities regarding how data literacy is understood and

---

1   For examples, see introduction chapter.

what educational objectives are being followed, highlighting that the term "rests on competing definitions" (Pangrazio and Sefton-Green 2020, p. 212). In light of this, some researchers have called for a unified terminology of data literacy (Koltay 2015), whereas others have argued that finding a single definition of literacy is "in fact [...] not desirable" in light of the multiple different skills and competences that are required (Fotopoulou 2020, p. 4). An examination of current literacy discourses quickly shows that the majority of data literacy concepts originate from data science tradition and follow an instrumental, skills-based approach, aiming at fostering people's technical skills to use and manage data (as highlighted by, e.g., Raffaghelli and Stewart 2020; Seymoens et al. 2020; Pangrazio and Sefton-Green 2022). Raffaghelli and Stewart, for example, conducted an extensive literature analysis of data literacy approaches with a focus on higher education faculty development, and found that the majority of literacy definitions "lean towards an instrumental framing", focussing mainly on the development of technical skills (2020, p. 447). The authors argue that while some analysed approaches aim for learners' empowerment, most do not support such "holistic idea of data literacy" and lack "acknowledgement of the emerging debates on data and ethics, minorities, privacy, agency and empowerment" in datafied societies (ibid., p. 444; 447).

In line with these findings, my study's review of the literature found that critical data literacy conceptualisations that aim for understanding and critical reflection of data systems are far less common than instrumental approaches. Yet, during the time of my engagement with the data literacy field, critical approaches to data literacy have increased significantly. When first reviewing research on data literacy as part of my Master's dissertation in early 2018, critical perspectives were rare and the few that could be identified almost all took a combined practical and critical approach. As I continued my research in the field during a research fellowship in 2019 and prepared two publications on "critical big data literacy", several new critical conceptualisations had emerged, including some that placed a strong focus on a broader reflection of datafication (see Sander 2020c; Sander 2020a). This trend continued while conducting my initial literature review for this study in early 2020 and was observed again when revising this chapter in 2023. Particularly in these two years, a large increase of critical data literacy publications was identified, signifying the growing importance of this field. This can also be illustrated by the texts referenced in this chapter: while the first version from early 2020 reviewed only 33 (critical) data literacy conceptualisations, this final version of the chapter provides an overview of a total of 75 publications.

This chapter thus draws on insights from my ongoing reviews of data literacy literature throughout the last five years and aims to provide a comprehensive – yet not exhaustive – overview of the diverging definitions of (critical) data literacy. Moreover, reflecting on the findings of these ongoing literature reviews, the chapter suggests a tentative categorisation of literacy approaches. According to this, data liter-

acy concepts can be broadly categorised into three overlapping categories that are located on a spectrum between practical-instrumental and critical-reflective literacy understandings (see fig. 1). The 'traditional' understanding of data literacy as data science skills as well as creative approaches to using data constitute the first category of *practical-instrumental literacy understandings*. The second category aims to *foster critical perspectives through the active use of data* and includes concepts that place a varying focus on practical data usage skills while also promoting critical reflection. Finally, some authors suggest that critical data literacy should focus on supporting *understanding and critical reflection of the wider, structural-level implications* that come with the datafication of our societies. Although all approaches to data literacy can offer useful insights, the concepts included in this last category are of particular relevance for the theorisation and further conceptualisation of critical datafication literacy in my study as well as for informing my empirical research.

Figure 1: Categorisation of (critical) data literacy concepts along a spectrum from practical-instrumental to critical-reflective approaches.
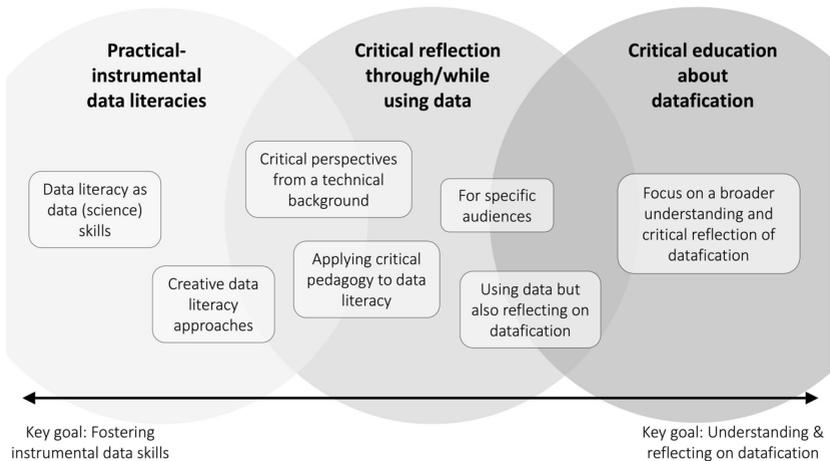
Yet, it should be clarified that this categorisation constitutes a tentative order, aiming for a structured overview of the field of data literacy concepts rather than a fully developed typology of data literacy approaches. The categories are not distinct or exhaustive in nature and many of the presented concepts include aspects that fit in more than one of these proposed categories. Moreover, as mentioned above, the (critical) data literacy research field evolved significantly during the course of my study, and in 2019, very few literature analyses of data literacy concepts had been published. One early example is Maybee and Zilinksi's 2015 study, which identified

eight key frameworks for data literacy in a literature review. Similar to my categorisation, they found predominantly data usage skills, but also "awareness: understanding data and its role in society" (2015, p. 2).

Between 2020 and 2022, several new categorisations of (critical) data literacy conceptualisations that make similar distinctions as mine were published. For example, Yousef et al.'s analysis of data literacy as a "multidimensional concept" found practical uses of data such as research data management or learning analytics as key areas of data literacy concepts, but also identified goals such as an informed citizenry, awareness of datafication and citizen participation (2021, pp. 26f, 30). In a German-language examination of critical data literacy approaches, Dander differentiated critical data literacy concepts by eight levels, such as their audiences and educational contexts, but also (relating to the same differentiation in data activism projects) their pro- or reactive nature (2021). Finally, Raffaghelli analysed nine higher education data literacy initiatives and identified four dimensions, which range from using data in pedagogical or open data settings to fostering understanding of the role data play in society and a postcolonial perspective on data (2020, p. 15). While using different terminologies and developing different categorisations of data literacy concepts, what these studies have in common is that they identify very diverse goals of data literacy approaches, ranging from active data usage skills to awareness and reflection. This chapter details the diverse objectives and characteristics of existing (critical) data literacy concepts, outlining parallels and key lines of argumentation in the field as well as highlighting differences and gaps of research.

2.2.1   How Is Data Literacy 'Traditionally' Understood?

Data Literacy as Data (Science) Skills

As outlined above, the most 'traditional' understanding of data literacy can be broadly summarised as a practical, skills-based approach that aims to teach how to *use data and datasets*. In their overview of previous data literacy research, Gray et al. summarise this understanding as: "being able to access, analyse, use, interpret, manipulate and argue with datasets in response to the ubiquity of (digital) data in different fields" (2018, p. 2). A similar definition is suggested by Wolff et al., who define data literacy as "the ability to ask and answer real-world questions from large and small data sets through an inquiry process", which includes the abilities to "select, clean, analyse, visualise, critique and interpret data" as well as using data to communicate stories and in design processes (2016, p. 23). Schüller et al.'s "data literacy charter" emphasises data literacy's role as a "key competence for all people in the 21st century" as it is "key to systematically transforming data into knowledge and actions" (2021, p. 2). This highlights the core objective of many 'traditional', instrumental data literacy understandings: to enable and motivate people to actively use data, deal confidently with their own and other people's data and to "use

new drivers and technologies such as Big Data, Artificial Intelligence or Internet of Things" (ibid.).

Frank et al. follow a similar understanding and add that they view data literacy as a "recent addition to a growing band of literacies such as numerical literacy, statistical literacy and IT literacy" (2016, p. 5). Others equally highlight the overlap or connection between these concepts, arguing that, given the "greatly amplified role that data now play in our lives", statistical literacy should be updated to include data literacy (Gould 2017). Other suggestions state data literacy should be incorporated into information literacy programmes (Calzada and Marzal 2013) or propose a combined "data information literacy" (Carlson et al. 2011). Similar to the data literacy discourse, also statistical literacy concepts have broadened in recent years, developing from a narrow technical understanding to "a broader meaning including critical, ethical-political aspects" (François and Monteiro 2018, p. 3). The "Data Revolution Group", a United Nations Secretary-General's Independent Expert Advisory Group, similarly called for a "global data literacy" that includes statistical literacy but also aims to enable citizens to "fully participate in initiatives to foster citizenship in the information age" (2014, p. 20). Other approaches to data literacy focus on business and employability perspectives through data literacy (Gummer and Mandinach 2015; Wolff and Kortuem 2015). Thus, even among instrumental, practical data literacy concepts, a variety of goals and approaches can be identified, and no single definition has emerged so far. Yet, the one defining characteristic of these approaches is that they aim to encourage learners to actively use data (analytics) for their purposes.

Creative Data Literacy Approaches

Several scholars have built on such skills-based data literacy approaches but have added a focus on creativity. D'Ignazio, for example, has suggested a "creative data literacy" as an approach for "non-technical learners" who may need "an alternative to the traditional quantitative approach to working with data" (2017, pp. 6; 7). Tactics for this include working with "community-centered data" or making data "messy" (ibid., p. 8, 11). Together with Bhargava, D'Ignazio has further proposed a "data mindset" that supports the "ability to think both creatively and critically" about opportunities to use data to answer questions and gain insights (2018). Here, they also clearly differentiate this "critical and creative mindset" from the "technical skills, statistical techniques, or particular software packages" that instrumental, skills-based definitions of data literacy such as the ones introduced above focus on. Thus, although they also aim for practical use of data, D'Ignazio and Bhargava clearly regard data literacy as an approach that should foster critical perspectives and work towards *empowerment* (e.g., D'Ignazio 2017, p. 6) This becomes even clearer in their 2015 concept, which will be outlined in more detail in section 2.2.3.

A similarly creative and critical data literacy understanding is suggested by Ahlborn et al. who suggest fostering data literacy through "critical making" and through using data, coding and artificial intelligence in creative play scenarios (2021, pp. 21; 32). Despite this focus on practically using data, their goal for data literacy is for learners to gain new perspectives on how social, cultural, political and economic processes are affected by data and to become empowered to participate in datafied societies in an informed manner (ibid., p. 34).

These critical and creative approaches to data literacy already show strong overlaps with the next category: concepts that aim to *foster critical perspectives through or while using data*. While purely instrumental, skills-based data literacy concepts constitute the most common and established understanding of data literacy, they are only of limited relevance for the conceptualisation of critical datafication literacy in my study. Although statistical and data analysis skills can also support better understanding of datafication, these competences can shed little light on questions such as "how are my data used to profile and score citizens" or "what is the influence of algorithms on my daily life?". Therefore, data literacy needs a "more complex" approach "than simply learning to access, use, analyze, or create data through statistics or comprehending data visualizations and data calculations" (Markham 2020, p. 230). Moreover, research has found that fostering technical skills alone, without also promoting understanding and critical reflection, is a "problematic strategy" and can lead, "directly or indirectly, to more exposure [of children] to online risks" (Livingstone et al. 2021a, pp. 21; 22).

In line with this, it has been argued that the uncritical use of data, such as everyday data-related practices in education, can lead to "'data (il)literacy' – an uncritical, one-dimensional understanding of data and datafication" (Mertala 2020, p. 1). According to Mertala, when learners are introduced to data as "direct measurements" instead of the "imperfect proxies and indicators" they are, a subconscious outcome – or "hidden curriculum" – of this could be that they "overestimate the accuracy of data and build excessive trust in datafied systems" (ibid., p. 14). For these reasons, critical education about data is required.

2.2.2   Fostering Critical Perspectives Through the Use of Data

When examining the history of different literacy concepts, a common trend emerges: often, initial understandings of literacy are of instrumental, narrow nature, describing predominantly practical skills. Over time, broader and more critical and reflective literacy understandings emerge. This "gradual expansion" has been traced in digital literacy concepts (Pötzsch 2019, p. 235; see also chapter 3.1) as well as in concepts of statistical and mathematical literacy (François and Monteiro 2018, p. 3). The data literacy field shows a similar development: while the instrumental, skills-focussed data literacy concepts outlined above still constitute

the most established understanding of data literacy in many contexts in- and out-side academia, calls to include more critical perspectives have emerged in recent years. Many new data literacy conceptualisations that include critical perspectives aim to *combine data usage skills*, for example analysing and visualising data, with *critical and ethical reflections* for example on the "protection of privacy, surveillance, epistemic manipulation and exploitation of data-related labour" (Špiranec et al. 2019). This combined approach has also been described as connecting "data literacy goals", such as collecting, analysing, examining and communicating with data, with "critical literacy goals", for example uncovering social inequalities with data, addressing data ethics or envisioning "new public spaces with data" (Louie 2022, pp. 4–6). The following sections provide an overview of different combined practical and critical data literacy approaches.

Critical Perspectives Originating from a Technical Background

Despite the dominance of instrumental, skills-based data literacy understandings, suggestions for critical data literacy approaches are increasing and have even come from technical backgrounds. One example is the already mentioned article by François and Monteiro whose concept originates from a statistical and mathematical literacy background and constitutes a very practical data literacy approach that nevertheless is "explicitly connected with ethical and political aspects" and aims for "constructive, concerned and reflective citizens" (2018, p. 1). Despite the technical background of their literacy concept, the authors clearly emphasise that today's citizens "must be able to understand the processes behind and to value the powers and limitations of Big Data" in order to achieve active citizenship and be prepared for the "data-driven society in which they live" (ibid., p. 5).

Similarly, Koltay et al. build on very practical definitions of data literacy that aim for learning to access, manage, handle and understand data, yet at the same time they emphasise the importance of a "critical assessment of data" and call for a "profound understanding of the big data phenomenon" (2015, pp. 403; 407). Although the authors place a great emphasis on practical skills such as learning to read graphs or drawing conclusions from data, they also aim to support learners in recognising when data are used misleadingly or inappropriately, and they see critical thinking as "one of the most important goals of data literacy education" (ibid., p. 407). A more recent publication by Verständig further describes critical data literacy as best located at the intersection of computer science and media education, and reports on a higher education course project, in which problem-based learning was applied to foster both students' data science skills and critical reflection of data phenomena (2021).

Applying Critical Pedagogy to Data Literacy Approaches

Another popular approach is to combine skills-based data literacy conceptualisations with the Freirean critical pedagogy that aims to empower learners and address questions of power in relation to data technologies. Indeed, the field of critical pedagogy with its pioneer Paulo Freire provides a highly relevant approach for educating about data that many critical data literacy conceptualisations have built on (for overview, see Špiranec et al. 2019) and that will be examined in more detail in chapter 3.3. An early example of a practical data literacy approach with a strong critical pedagogy perspective is Philip et al.'s "Big Data Literacy" (2013). Besides a pronounced focus on data science related objectives, their framework also includes two distinctly critical objectives. It aims to enable students to understand that "data are never neutral or objective" but always premised on certain assumptions and that they (re)produce forms of power in society, and aims to encourage students to see the generation, analysis and work with data as "powerful ways to understand and address societal issues as well as issues that affect their lives and the lives of people in their communities" (ibid., p. 114f). Philip et al. ground their framework within sociocultural theories of learning and critical pedagogy perspectives that "situate learning within its larger historical, social, political, and economic contexts" (p. 112).

Moreover, Dasgupta and Hill's concept of "critical algorithmic literacies" is rooted in Freire's critical pedagogy (2021, p. 5). Through practically engaging and experimenting with real-world datasets, they aim to encourage children to not only understand technical concepts around algorithms but also to "understand and critique the algorithmic systems that affect their lives" and evaluate their societal implications (ibid., p. 5; 2). Tygel and Kirsch place an even greater emphasis on Freire's influence and develop a "Freirean Inspired Critical Data Literacy" (2016, p. 113). Besides data processing, data communication and data production, their literacy concept promotes the skill of "data reading", which "starts with understanding how the data was generated [and] also includes understanding who produced it, in which context and why", and positions data as "the output of a social process" rather than an objective fact (ibid., p. 117).

Practical and Critical Approaches for Specific Audiences

As the empirical findings in this study will demonstrate, it is important to consider the audience that a (critical) data literacy intervention addresses. In line with this, some data literacy conceptualisations are developed for specific audiences, such as high school students or teachers, whereas others address the general public. Koltay's already mentioned data literacy concept, for example, addresses students but also academics or "data librarians" and aims to fulfil "the mission of the contemporary academic library" (2015, p. 401). Fontichiaro et al. equally target "librarians and their classroom colleagues" in their practical guidebook for "data literacy in the real world", which aims to foster practical data usage skills as well as awareness of

data collection and reflection of ethical questions and "unintended consequences" of big data (2017, pp.ii, 255).

Hautea et al. focus on working with youths and suggest "possible strategies for engaging youth in critically thinking about the role of data in their lives and in their world" (2017, p. 920). In their understanding of critical data literacy, the authors follow an instrumental skills-based data literacy definition by the Oceans of Data Institute but also emphasise that "the term 'critically' is central" in their critical data literacy approach, building – again – on Paulo Freire's already mentioned critical pedagogy (ibid.). Another data literacy concept that addresses a very specific audience is Fotopoulou's "critical data literacies for civil society organisations" (2020). While she also considers "citizens more generally", the author focusses on collective actors: civil society organisations that need to "make sense of data, communicat[e] in ways that are relevant to broad audiences, and us[e] data for the social good" (ibid., p. 1). Fotopoulou builds on D'Ignazio and Bhargava's 2015 definition of big data literacy (see 2.2.3), but argues for "literacies in the plural", framing data literacies as "agentic, contextual, critical, multiple, and inherently social" (ibid., p. 4). Particularly relevant is the author's emphasis on criticality, stressing the "importance of raising awareness about the ideological and power aspects of data" (ibid.). Similar to other combined practical and critical approaches, Fotopoulou argues that this critical awareness "can be developed gradually and in parallel to developing certain technical competences" (ibid, p. 15).

Concepts That Begin to Consider Datafication as Part of Active and Practical Approaches

Finally, the literature review identified several examples of data literacy concepts that take a practical and skills-based approach while at the same time aiming for *critical reflection of how data and datafied systems change our lives and our society.* While still advocating for practical data skills and fostering critical understanding *through using data*, these concepts place a greater emphasis on the systemic and societal implications of datafication. Nevertheless, the lines between this tentative sub-category and the previous and following sub-categories are blurry and overlaps may occur.

What the concepts presented in this section have in common is that they, similar to those presented in the first section, aim to foster various practical skills that can be summarised as "reading and working with datasets" (Gray et al. 2018, p. 8), for example "data collection, data cleaning and rendering and visualisation" (Davies 2022, p. 158); interpreting, navigating, collecting and presenting data (Seymoens et al. 2020, p. 103); or giving people access to their own data through apps or hackathons (Pybus et al. 2015). However, these data literacies do not stop at this practical level, but also aim for critical reflection both of the data analysis and of the way data practices affect our lives and societies. For example, Davies suggests that learners should gain a critical perspective on how flawed and misleading data collection, analysis and repre-

sentation can be, aiming for "social scientific critical analysis of strengths and weaknesses of quantitative data" that should be connected to goals of data justice (2022, p. 158). Seymoens et al., relating to a Data Literacy Competence Model that was developed by the Flemish Knowledge Centre for Digital and Media Literacy,[2] similarly highlight that besides *using* data, individuals should also *understand* data, for example by observing how it is used, analysing individual and social consequences of data practices, and evaluating and reflecting on the impacts of these consequences (2020, p. 103). Thus, this framework is oriented towards learners "critically and consciously understanding the role of data in society [and] personal life" (ibid., p. 104).

These definitions already point at a stronger focus on understanding and critically reflecting the implications of data systems. Apart from that, some combined practical and critical-reflective approaches to data literacy aim to *highlight potential negative consequences* of data systems and *empower* citizens in datafied societies. Nguyen, for example, argues that data literacy "offers paths to holistic strategies" for addressing the challenges of datafication and should be seen as a "tool of empowerment" that is important for "retaining agency and enabling inclusion as well as resistance in the digital society" (2021, p. 212). Pybus et al. suggest a "holistic data literacy" that aims to foster the ability to "question how meaning is constructed and (re)presented from the data", but also encourage learners to "unpack those opaque material processes" and to participate in "the active (re)shaping of data infrastructures" (2015, p. 4). In addressing these goals by giving learners access to their own data, the authors hope to promote "empowerment that comes from gaining access to what we collectively generate" (ibid., p. 8).

A similarly early data literacy conceptualisation by Crusoe agrees that data literacy should emphasise that "algorithms are opaque", "utilize stereotypical user models", may determine what the individual "gain[s] or lose[s]", and can lead to discrimination (2016, p. 40). Besides knowledge of "what data are, how they are collected, analyzed, visualized and shared", Crusoe suggests that data literacy consists of "the understanding of how data are applied for benefit or detriment, within the cultural context of security and privacy" (ibid., p. 38). While equally a combination of practical and critical goals, Gray et al.'s "data infrastructure literacy" goes one step further and aims to

> promote critical inquiry into datafication, into how datasets are created with certain purposes in mind as well as opening up "infrastructural imagination" (Bowker 2014) about how they might be created, used and organised differently (or not at all) – and the tensions that emerge between these two. (2018, p. 3)

---

2   As no original English-language publication of this framework could be found, Seymoens et al.'s text constituted the most suitable source for this framework.

Despite this very broad and critical objective, Gray et al.'s literacy concept is grounded within data, science, technology and human-computer interaction studies and includes a strong emphasis on practical data skills. Similarly, a systematic literature analysis that aimed to identify critical dimensions in data literacy approaches concluded that critical data literacy can be understood as a "critical approach to data and our data realities" that encourages learners to problematise and transform the "oppressing and unjust conditions of life produced by exclusionary, exploitative, invasive and manipulative uses of data" (Špiranec et al. 2019).

Overall, it seems that the combined approach of fostering practical skills as well as critical reflections is a relatively common approach in (critical) data literacy conceptualisations. These concepts provide many relevant insights for a more in-depth conceptualisation of critical datafication literacy in my study, for example by suggesting different understandings of 'critical reflection of data'. Many of these concepts aim for critical reflection that is directly related to data usage, including questions around critically considering the origin and context of datasets, the way they are used and for what purposes, where and how securely data are stored or how they affect individual privacy. However, in light of the increasing complexity of algorithms and data structures, such critical perspectives in the sense of "*what* do the data tell us" are not enough and a critical attitude towards data needs to also reflect on how data powers affect our actions and experiences in life, asking: "*how* do data take effect?" (Leineweber 2021, p. 147f, own translation, emphasis in original). In addition, the majority of these concepts still places the focus on fostering the skills to practically use data. Yet, as van Audenhove et al. put it, "*using data* and *understanding data* are different competences that both need attention", and a broader discussion beyond skill-based data literacy understandings is needed (2020, p. 2, emphasis in original). This reiterates findings from the NGO Doteveryone that argue at there is "a big difference between having skills – knowing how to use the internet – and having understanding – knowing the implications of using the internet" (Miller et al. 2018, p. 5).

While some of the presented concepts already touch on critical reflection of how data and datafied systems change our society, and there is certainly overlap between the different categories that are suggested, what has been missing thus far are critical conceptualisations of data literacy that *place critical thinking and reflection at their core* and focus on promoting a broader critical reflection of the *structural and systemic levels of datafication* that are becoming deeply engrained into today's societies and *how to effectively challenge them*.

2.2.3   Towards Critical Understanding of Datafication

In recent years, a small but growing number of concepts have emerged that focus on exactly this distinctly critical and reflective understanding of data literacy. These

approaches emphasise that "digital understanding is not about being able to code, it's about being able to cope", and focus on improving people's understanding of the "impacts of technologies – how they shape people's lives and society as a whole" (Miller et al. 2018, pp. 5; 8). The terminology, as usual, differs widely from, for example, a "'manifesto' for the 'public understanding of big data'" that connects concerns around big data to the "broad multidisciplinary fields of public understanding of science (PUS) and public engagement with science and technology (PEST)" (Michael and Lupton 2016, p. 105); to different variations of data literacies as well as calls for an "algorithmic literacy" that aims to enable "people to form considered opinions and debate the issue in a factually informed way" (Grzymek and Puntschuh 2019, p. 11). Markham even speaks of "X_literacy, [...] ('X' meaning whatever we want to call it)" (2020, p. 229). What the concepts presented in this next section have in common, however, is that they place scrutiny and critical reflection at their core, aiming for, among others, "deep curiosity" and "critical orientation" that lead to "a level of comprehension and critical awareness that, once we have it, keeps us asking new questions" (Markham 2020, p. 229). In other words, they call for an understanding of "how algorithmic processes can turn otherwise harmless-seeming data into valuable information" and reflection of "the critical perspectives around these practices" through data literacy (Bilstrup et al. 2022, p. 226). The following section details a small number of concepts and frameworks from this final category that are most relevant for my study.

An Extended Definition of Big Data Literacy – D'Ignazio and Bhargava

The earliest concept in this section that has influenced many critical understandings of data literacy that were developed later on (e.g., Fotopoulou 2020) is D'Ignazio and Bhargava's "extended definition of Big Data literacy" (2015). The authors highlight four problematic issues of big data: big data's lack of transparency, extractive collection, technological complexity, and control of impact, and argue that these problems are "ones of 'data literacy'" (ibid., p. 2). To address these issues, D'Ignazio and Bhargava suggest a big data literacy that – again – builds on Paulo Freire's work on "empowerment through literacy education", aiming for technical skills as well as "emancipation [...] through the literacy process" (ibid., p. 2; 5). Besides practical skills of working with data, their extended big data literacy hopes to encourage learners in:

- Identifying when and where data is being passively collected about your actions and interactions.
- Understanding the algorithmic manipulations performed on large sets of data to identify patterns.
- Weighing the real and potential ethical impacts of data-driven decisions for individuals and for society. (ibid., p. 3)

Such distinct critical understanding and reflection is unusual for data literacy concepts from this time. A working paper from Bhargava and others from the same month further argues for a "literacy in the age of data", which aims to enable citizens to "constructively engage in society through and about data", and which can evolve and expand over time in light of the continuously changing requirements of an increasingly datafied society (Data-Pop Alliance and Internews 2015, pp. 26; iv). This flexible nature of data literacy as well as the authors' focus on empowerment and emancipation constitute relevant insights for my study's conceptualisation of critical datafication literacy.

Educator's Data Literacy – Raffaghelli

Again highlighting potential overlap between the tentative categories suggested in this chapter, Raffaghelli's literacy understanding is developed out of a context of *educator's* data literacy (2022). Despite this specific context, her conceptualisation of data literacy provides novel insights and unusual approaches for addressing the wide-reaching implications of datafication on our societies through a literacy-angle. The author's goal is to "support educators' reflection and organic action" and develop a "bigger picture" of data literacy (ibid., p. 82). Based on a review of critical data literacy literature and 12 interviews with different educators, Raffaghelli develops a complex conceptual scheme for data literacy that builds on "the proactive and reactive dispositions towards data" identified by data activism research (see chapter 2.1), aiming to "strike a balance" between these perspectives (ibid.). This scheme presents data literacy as a "complex puzzle" that includes traditional, instrumental understandings of data literacy; "critical and participatory approaches" that aim to empower learners through gaining access to, creating and using their own data; as well as data literacy approaches that call for "activism against datafication and algorithmic biases" by applying media education principles to data (ibid., p. 92). Thus, Raffaghelli's conceptual scheme for data literacy highlights the interplay between data activism, data science and media education in (critical) data literacy approaches, emphasising the variety of skills and understandings that citizens need to navigate datafied societies in an informed and empowered manner.

Critical Digital Literacy – Pötzsch

Another very relevant concept that aims to foster citizens' understanding of datafication is Pötzsch's critical digital literacy (2019). Despite using the term "digital literacy", this approach provides manifold insights into fostering critical awareness and reflection of the impacts of datafied systems. In contrast to many literacy concepts presented thus far, Pötzsch does not consider data skills to be an integral part his literacy understanding. In contrast, he emphasises that neither digital technologies nor expensive software are required for critically educating about digital technologies and datafied systems (ibid., p. 227; see also section 2.2.4). This constitutes a clear

dissociation from combined practical and critical concepts that aim to promote critical awareness through the active use of data. While Pötzsch acknowledges instrumental user skills, he rather places his focus on *critical thinking*, aiming to foster "the ability to critically assess, appropriate, and, if need be, replace, repair, or resist such [digital] technologies" in order to form "a self-determined future citizenry" (ibid., p. 236).

This critical thinking about 'the digital' takes "structural aspects of the technology into account", while using the "widest possible contextualisation of technology", which, importantly, includes issues of "exploitation, commodification, and degradation in digital capitalism" (Pötzsch 2019, p. 221). To achieve such "critical awareness of digital technologies' wider societal, political, economic, cultural, and environmental implications", the author argues that "blackboxed technologies need to be opened up, contextualised, and critically questioned" with a view on "capitalist dynamics, environmental ramifications, and individual empowerment" (ibid., p. 235; 236; 222). Pötzsch's distinct goals of a self-determined citizenry and a wide-angle perspective on the impacts of technology, including structural transformations in society, constitute novel approaches that have informed this study's development of a critical datafication literacy framework.

Information Literacy in the Digital Age – Polizzi

Like Pötzsch, Polizzi also uses the term "critical digital literacy" to describe a very relevant literacy understanding for this chapter (2020b). Polizzi emphasises that critical digital literacy, or "information literacy in the digital age", needs to go beyond a focus on digital content such as evaluating online information, and rather foster a broader understanding of digital infrastructures (ibid., p. 1). He criticises that research thus far has overlooked the importance of understanding the digital environment for "civic and political engagement" and calls for literacy in the sense of "an ensemble of critical abilities, knowledge and interpretations that are essential in the context of democratic participation and social inclusion in the digital age" (ibid., p. 7; 2). Similar to Pötzsch, Polizzi argues that critical digital literacy should foster understanding of the socio-economic structures of digital technologies and embedded power structures, revealing "democratising potentials" as well as "structural constraints" (ibid., p. 16). To achieve this, Polizzi – similar to my arguments in chapter 3.2 – sees value in connecting literacy perspectives from media studies with insights from political research and democratic theory (ibid., p. 13). This distinguishes this literacy approach and makes it particularly relevant for this study's development of a theoretical framework for critical datafication literacy.

Data Citizenship – Carmi, Yates et al.

The third relevant concept, the "data citizenship" model, is developed in several publications by the "Me and My Big Data" project at the University of Liverpool (Carmi

et al. 2020; Yates et al. 2020; Yates et al. 2021). While the authors' journal paper focusses on the connection of data literacies "to disinformation, misinformation and malinformation" (Carmi et al. 2020, p. 1), this particular context is not specified in the two reports. Nevertheless, this specific context of data literacy should be kept in mind when examining the data citizenship model. The authors aim to address gaps in existing data literacy frameworks by exploring "links between 'data, power, and positionality'" and encouraging citizens to have "critical and active agency" (Yates et al. 2020, p. 10; Carmi et al. 2020, pp. 1; 10).

The data citizenship framework revolves around three areas (Yates et al. 2021, pp. 14–25). First, "Data Doing" addresses 'traditional' data literacy skills around using and managing data but doing so in "an ethical and critical manner" (ibid., p. 19). The second domain, "Data Thinking" takes a more reflective approach (ibid.). It should be noted that the definition of this domain showed distinct changes between the initial and the final research report, from a focus on *using data* to think (critically) about data (see Yates et al. 2020, p. 10) to a strong emphasis on critical understanding of the data economy and critical reflection in the sense of Paulo Freire's "conscientização – or a critical consciousness"[3] as goals for data citizenship (Yates et al. 2021, p. 19f). Particularly this newly strengthened critical focus makes data citizenship a highly relevant model for informing my study.

The last domain, "Data Participation" also takes a distinctly critical perspective to educating about data, as it "seeks to address the problem of citizens' feelings of disempowerment in data-driven environments and the unequal power dynamics" by highlighting the "importance of citizens' proactive engagement" (Yates et al. 2021, p. 21). Building on data activism and data literacy research in combination with "ideas from digital citizenship and democratic education", the authors understand data participation as "civic and/or collective action", encouraging learners to "exercise their digital / data human rights" (ibid., p. 22). This interdisciplinary combination hints at what will be demonstrated in more detail in chapter 3: how much critical data literacy conceptualisations can learn from more established fields such as citizenship education or Paulo Freire's critical pedagogy. However, the connection between the concept of "data citizenship" to data literacy and education scholarship is sometimes blurry, and the context of dis-/mis- and malinformation is not always clear.

Moreover, several suggestions for promoting data citizenship in practice are made. While the authors stress the need for "proactive citizens" instead of "passive consumers", they nevertheless warn against merely shifting the responsibility to the individual, which may "potentially remov[e] the onus from regulatory bodies, institutions or the state" (Carmi et al. 2020, pp. 9; 15). If the responsibility to be data literate lies with the individuals, the authors warn, "we can expect socio-economic

---

3   See chapter 3.3 for a detailed definition.

inequalities to influence their ability to have access and resources to such education programmes" (ibid., p. 12). Furthermore, the authors emphasise the need to consider different populations and "address different social contexts", particularly of marginalised communities, arguing that "a 'universal' one size fits all solution" for data literacy is not possible (ibid., p. 11). Moreover, data literacies should go beyond the individual and instead work with people's "networks of literacy", including the people they engage with and the different media they use (ibid., p. 12). Finally, the authors – similar to Pötzsch (2019) – highlight that data literacy does not need to utilise digital methods but should be expanded "outside the 'digital' realm" and regarded as a "more holistic and networked experience" that includes online and offline spheres (ibid., p. 17). All of these valuable considerations have informed several methodological decisions of the empirical research conducted in my study (see chapter 4).

Critical Digital and Data Literacies – Pangrazio et al.

The work of Pangrazio and co-authors proposes several literacy concepts for promoting critical reflection of the datafication of our societies. This section will outline four key conceptualisations by Pangrazio and others: the concept of "critical digital literacy" (Pangrazio 2016), "personal data literacies" (Pangrazio and Selwyn 2019), research on "the social utility of data literacy" (Pangrazio and Sefton-Green 2020), and on "critical data education" (Pangrazio and Selwyn 2020). Despite their different terms and some variations in authorship, all these approaches provide highly relevant insights for an extended framework for critical datafication literacy.

First, an early publication by Pangrazio suggests a *critical digital literacy* concept that examines "broader issues associated with digital media use", such as how such technologies "reinforce issues of social class, race and gender and what might be done to challenge and overturn exploitation and inequality" (2016, p. 169). This literacy intends to "cultivate a dispassionate, critical disposition in a context that invests deeply in the personal and affective" and work towards a "nuanced understanding of power and ideology within the digital medium" (ibid., p. 168). The goal and challenge for such literacy is to encourage "the individual to move between these mindsets (i.e., critical and technical) as part of their digital practices" (ibid., p. 170). By placing this strong focus on a broad and structural understanding as well as critical reflection, this literacy already distinctly differs from most digital and data literacy conceptualisations at the time.

A key concept that informed my study's theoretical framework for critical datafication literacy and the empirical research that was conducted is Pangrazio and Selwyn's *personal data literacies* (2019). The authors suggest to "see personal data as a 'text' in the New Literacies sense" instead of seeing it as "information to be managed" (ibid., p. 427). In taking this perspective, they go beyond data skills and rather "foreground[...] 'personal data' as a social practice and tool for action" (ibid.). Moreover,

Pangrazio and Selwyn's personal data literacies framework places a strong emphasis on critical thought. It aims to foster understanding of the sociocultural context around data and the ideologies that underpin them, and to cultivate the "metaphorical 'space'" to reflect critically on the implications of data and the "inherently political nature of the broader data assemblage" (ibid., p. 426).

Of the five domains that make up Pangrazio and Selwyn's personal data literacies framework – "(1) Data Identification, (2) Data Understandings, (3) Data Reflexivity, (4) Data Uses, and (5) Data Tactics" (2019, p. 428) – particularly the third and fifth provide valuable new insights for my study. These aim to support citizens' critical understanding of the "profiling and predictions that are made from processed personal data" and their ability to employ "tactics of resistance and obfuscation" (ibid., p. 429; 431). Overall, the authors place a strong emphasis on fostering the individual's "capacity to understand and control one's personal data", which they see as "a crucial part of living in contemporary society" (ibid., p. 419). Yet, Pangrazio and Selwyn nevertheless criticise a shift of "the burden of time and responsibility" to the individual in other literacy approaches and clarify that their focus lies on *supporting* citizens "to engage critically with their personal data, so they have a sense of understanding, control and agency within the data assemblage" (ibid., p. 425; 426). This careful balance between empowerment and responsibility distinguishes this framework's novel approach to data literacy.

Moreover, the authors make suggestions for practically implementing personal data literacies. Similar to Carmi et al. above, they recognise that people have "varied capacities" and therefore "unequal agency" in engaging with data, and argue that such "data inequalities" need to be considered by data literacy frameworks (Pangrazio and Selwyn 2019, 433f). The authors further suggest two "areas of obvious application" for their framework: "public education and academic research" (ibid., p. 431). They highlight that critical data and digital literacies are "sorely missing from education", and that they should be strengthened in both formal education and in "informal educational initiatives more appropriate for adults" (ibid., p. 431; 432). Moreover, they call "IT providers and government agencies" to action (ibid., p. 431). Both the authors' novel framework and these suggestions for a practical implementation provide numerous valuable insights for my study's theoretical and empirical analyses.

A third key publication by Pangrazio and Sefton-Green examines the *social utility of data literacy*, reviewing recent models of data literacy, which are categorised as "formal, personal and folk pedagogies of data" (2020, p. 208). The article criticises that although literacy is often "posited as the ideal response" to challenges of datafication, how such literacy is understood is often vague and "there is certainly an absence of practical models" (ibid., p. 211). The authors suggest that data literacy should foster "critical understandings of the technological infrastructure and the political economy of digital platforms" and should provide citizens with the "skills required

to have agency in a datafied world" and to protect their privacy (ibid., p. 214). Such literacy, they argue, can "protect individuals from being manipulated by datafication processes" and is central to democratic processes, together with an "optimisation of both legal and tactical responses" (ibid., p. 218).

Apart from this relevant definition, Pangrazio and Sefton-Green raise a number of novel considerations that have informed my study. They provide a clear argumentation for the suitability of the term "literacy" for educating about datafication, arguing that the educational dimension of literacy suggests that individuals can have sufficient literacy to participate while at the same time being "continuously involved in processes of self-improvement" (2020, p. 217). Moreover, they differentiate between operational and critical literacy, whereas the latter comes with "a host of 'higher order' questions, including the ability to challenge, ask questions and reflect" (ibid., p. 213). This literacy understanding resembles the German concept of "Bildung" (see chapter 3.2) and strongly informed my study's critical datafication literacy framework (see chapter 3.4). Another significant point the authors raise and that strongly informed my study's empirical approach is that "without viable alternative data practices, critical awareness might actually increase feelings of anxiety and apathy", a problematic issue that also emerged in my prior empirical work (ibid., p. 218; see also Sander 2020a; Sander 2020c). Therefore, data literacy interventions should always include "a set of practical strategies and tactics that individuals can experiment with in order to operationalise their newfound awareness and understanding" (Pangrazio and Sefton-Green 2020, p. 218).

Pangrazio and Selwyn further suggested a *critical data education* concept based on the findings of three empirical research projects (2020). Similar to the literacy frameworks outlined above, this concept aims for a broad and critical understanding of datafication and its social, political and economic implications as well as fostering data protection strategies. However, it places a particular focus on "the social injustices associated with datafication" and on "revealing structural inequalities" that data systems create and strengthen, clearly arguing that a "critical approach" is "the most appropriate way to fully address issues of datafication" (ibid., p. 5). Moreover, this article provides a number of valuable insights for practically implementing critical data education (see section 2.2.4).

Overall, the four publications by Pangrazio and co-authors suggest novel and highly relevant literacy concepts that provide many insights for the critical datafication literacy framework developed in my study and have strongly informed the study's empirical research. In particular, the authors' strong focus on a broad critical perspective on the social, political and economic dimensions of data and the structural changes through the datafication of our society constitute key novel insights.

Own Initial Conceptualisation: Critical Big Data Literacy

Finally, I have suggested an initial conceptualisation of a so-called *critical big data literacy* in my prior work. This literacy aims for an awareness, understanding and ability to critically reflect upon big data collection practices, data uses and the possible risks and implications that come with these practices, as well as the ability to implement this knowledge for a more empowered internet usage (Sander 2020a; Sander 2020c). Importantly, this critical reflection does not contain a certain set of beliefs or attitudes as its outcome, meaning it does not necessarily entail a negative stance towards data practices. Rather, critical big data literacy aims at empowering citizens to question and scrutinise the socio-technical systems of big data practices and to form considered opinions on current debates around data analytics. This would allow citizens to make informed decisions on personal choices such as which data to share or which services to use, and it would enable them to engage in public debates around datafied systems. Nevertheless, it is important to emphasise that – similar to previous concepts – the goal of this literacy is not to "merely shift responsibility to individuals" (Sander 2020c, p. 5). Instead, providing learners with constructive advice is intended to empower them and to counteract a potential resignation in learners that may arise when learning about the ubiquitous collection and analysis of personal data in our datafied worlds.

Preliminary Conclusion:
Insights and Shortcomings of Existing (Critical) Data Literacies

Overall, this chapter has demonstrated that although the 'traditional' instrumental and skills-based understanding of data literacy still dominates academic discourses, many critical data literacy approaches are emerging. Particularly the last five years have shown a strong increase of critical perspectives on educating about data. While the authors of such critical concepts agree that data literacy needs to go beyond data usage skills and should include critical perspectives, they take different stances on the extent to which technical skills of using, interpreting and manipulating data should still be involved in critical data literacy as well as what the 'critical' in critical data literacy should look like. Some focus on critical reflection during handling data or of digital content, whereas others call for a broader perspective, aiming for critical understanding of political, societal and economic impacts of datafication and data infrastructures. The latter critical-reflective and empowering understanding of data literacy as an approach for critically educating about datafication constitutes the most relevant approach for this study's development of a theoretical framework for critical datafication literacy.

However, some scholars also question the suitability of literacy as a concept for educating about data or highlight the shortcomings of existing (critical) data literacy approaches. Fotopoulou, for example, despite suggesting a new critical data literacies concept, also argues that "the term 'literacy' has been deemed too basic and

indeed counterproductive" (2020, p. 5) and she instead considers "expertise" or "socially distributed expertise" as developed by Bassett (2015). A comprehensive analysis of different alternative terms and concepts to literacy is provided by Pinney, who highlights that literacy as a term can be problematic because it has different meanings and is embedded with different notions of power for different people (2020, p. 229). She therefore concludes that literacy can be useful as a concept but not so much as a term, especially when addressing audiences outside of academia (ibid., 2020).

Moreover, a limitation of many data literacy conceptualisations is that they focus on "individualised rather than collective forms" (Hintz et al. 2022, p. 149). In instrumental literacy concepts, this focus is evident as they foster technical skills and competences for the individual. However, critical data literacy approaches also tend to focus on promoting the skills to better protect one's own data and to avoid data disclosure, thus shifting the responsibility of addressing problematic issues around datafication to the individuals. Similarly, digital privacy is often framed as individual responsibility (Kazansky 2015, p. 193). This "responsibilisation" in which users "enter into a dependent relationship with opaque technologies and are effectively left no choice but to deal with ensuing threats individually" is problematic since notions of informed consent, notice and choice have been called into question (ibid.; see also Forbrukerrådet 2018; Larsson et al. 2021). Thus, recently, calls for more collective approaches have increased, and some critical data literacy scholars have recognised and addressed the tension between an empowering literacy in contrast to a shift of responsibility to citizens (e.g., Pangrazio and Selwyn 2019; Carmi et al. 2020).

Finally, the "seeming[...] 'neutrality'" of existing data literacy concepts has been criticised (Jansen 2021, p. 1). Jansen argues that the "mere act of centring data in a literacy approach is political and value ridden" and that a "(re)politicization of data literacy" is needed (ibid.). This could take place by "politicizing the learner" who should not only learn about risks and opportunities of data but also "gain competencies to be able to influence political processes and challenge existing power structures" (ibid., p. 5). Moreover, Jansen contends that the "practice of data literacy in itself needs to be politicized" and its "blind spots related to structural inequality and power" should be reflected (ibid.). Thus, a (re)politicised data literacy takes a "bottom-up approach to dismantling power structures, understanding inequality and promoting political participation" and should learn from pedagogical approaches that aim for "democratic engagement, dismantling power structures, policy reform and activism" (ibid., p. 1; 8). Several of Jansen's criticisms have been addressed in my study, which learns from political education (chapter 3.2) and critical pedagogy (chapter 3.3) and suggests a critical datafication literacy framework that aims for broad, systemic understanding of datafication as well as enabling citizens to become active (see final framework in chapter 7).

2.2.4   Practically Implementing Critical Data Literacies

Apart from the theoretical conceptualisations of (critical) data literacies that have been outlined in the chapter thus far, academic research has also made many suggestions on how best to implement critical education about data in practice. As will be argued in more detail in chapter 3.4, my study's framework for critical datafication literacy focusses on a theoretical literacy conceptualisation, identifying key goals and objectives for critical education about data. However, my study's empirical research examines one way to practically implement critical data literacy: online educational resources about datafication. For this research, much could be learnt from reviewing recent suggestions for a practical implementation of critical data literacy. The following section aims to provide a concise overview of key advice from other scholars and practitioners on how best to critically educate about datafication.

Scholars have highlighted that it is difficult to implement literacy objectives into a practical model, but that different approaches are suitable: formal education, personal pedagogies of data, folk pedagogies of data – i.e., "open, public programmes [...] raising awareness and critical understanding of data" (Pangrazio and Sefton-Green 2020, p. 217) as well as artworks (Stark and Crawford 2019). As already mentioned above, a first key recommendation by the literature is that there is no one-size-fits-all-approach to literacy but that different audiences need different approaches (Carmi et al. 2020, p. 11). Thus, data literacy promotion should "involve providing multiple pathways" for people with different needs and capacities (Data-Pop Alliance and Internews 2015, p.iv), and educators should be aware that it may not be possible to reuse educational resources outside their intended context (Jansen 2021, p. 8).

Moreover, several suggestions are made in regard to the content of critical data literacy approaches. A common recommendation is to emotionally engage learners by using data or hypothetical data scenarios that are relevant to them and reflect their realities (D'Ignazio and Bhargava 2015; Tygel and Kirsch 2016; Markham 2019; Agesilaou and Kyza 2021; Bilstrup et al. 2022). This could, for example, be achieved by showing learners what data is stored about them by comparing mainstream and alternative services (Pötzsch 2019, p. 233), or by using relevant real-life case studies (Fontichiaro et al. 2017, p.v). Further concrete suggestions include using metaphors to describe data systems (Windeyer 2019, p. 318) or using popular culture (Pötzsch 2019, p. 235) to foster learners' imagination of datafication, which is often limited (Pangrazio and Selwyn 2020, p. 11). Importantly, educators should not try to convince learners that certain data practices are good or bad (Fontichiaro et al. 2017, p.v; Markham 2019, p. 758), but rather act as "problem posers" that encourage students to act as "active agents" (Aliakbari and Faraji 2011, p. 80). Critical reflection should be promoted, for example by prompting learners with good questions (Markham 2020, p. 231f), historicising technology (Pötzsch 2019, p. 232) or by showing learners ways

in which data are used that are undesirable to them (Bilstrup et al. 2022, p. 233). In order to not leave learners feeling frustrated or resigned after gaining such information, it is key to provide "adequate pathways forward" (Jansen 2021, p. 7) and show learners "how they can take the first steps to call for change" (Bilstrup et al. 2022, p. 234). In doing so, it is crucial to not reinforce the responsibilisation critiqued above.

Academic research further makes a number of relevant suggestions on useful formats to foster critical data literacies. Importantly, scholars have highlighted that (critical) data literacy does not necessarily need to be fostered using digital formats but that traditional teaching methods and embodied and tangible approaches may even be better suited to promote critical reflection of datafication (Pötzsch 2019; Windeyer 2019; Bilstrup et al. 2022). Apart from that, applied and participatory approaches are recommended (D'Ignazio and Bhargava 2015; Iliadis and Russo 2016; Markham 2019) as well as using data in creative play scenarios (Ahlborn et al. 2021, p. 34). While some authors also advocate for gamified approaches (Agesilaou and Kyza 2021, p. 464), others found that tools that were "too fun" made it difficult to evoke critical reflection (Pangrazio and Selwyn 2020, p. 11). A clear call was identified to materialise the intangible world of data in order to break down complex issues and foster understanding (Windeyer 2019, p. 329; Pangrazio and Selwyn 2020, p. 9f). This could, for example, be achieved by using visual approaches (D'Ignazio and Bhargava 2015, p. 4), such as data visualisations. These further "have the potential to evoke empathy, pity, sorrow, shame and other emotions" (Kennedy and Hill 2018, p. 843), which reiterates other scholars' emphasis on the role of emotions when discussing data and its implications for learners' own lives (Bilstrup et al. 2022, pp. 226; 233).

2.2.5   Conclusion

This chapter highlighted how many conceptualisations of (critical) data literacy already exist, how diverse they are, and which parallels can be identified. A tentative categorisation of literacy approaches from practical-instrumental to critical-reflective was suggested, which highlighted a strong dominance of instrumental data literacies. Critical approaches that go beyond practical data usage skills or critical reflection during data handling or of digital content are still rare. Yet, a final category of literacies that aim for broad critical understanding and reflection of political, societal and economic impacts of datafication and data infrastructures was identified. These literacy approaches proved most relevant for my study and provided manifold insights for the development of a theoretical framework for critical datafication literacy. Nevertheless, many of the analysed critical data literacy concepts focus on individualised approaches, thus potentially leading to a 'responsibilisation' of individuals, and they often come with little theoretical grounding, which can leave them disconnected from insights of traditional critical education scholarship. Scholars have further argued that data literacy needs to be 're-politicised' to address ques-

tions of power and inequalities and to promote political participation. Particularly in this regard, much can be learnt from traditional education scholarship.

My study builds on the existing critical data literacy conceptualisations that were analysed in this chapter and aims to address gaps in research by developing a theoretically and empirically grounded framework for critical datafication literacy. Chapter 3.4 presents a preliminary outline of this framework and discusses how it has been informed and shaped by insights from this chapter and from the three traditional educational approaches that are analysed in the next chapters. In addition to these theoretical findings, this chapter also highlighted that scholarly research has made many useful suggestions on how to implement critical education about data intro practice, which have informed this study's methodological decisions (see chapter four) and will be discussed in more detail in chapter six.