# Appendix

# Appendix I: Methodology Annotations

## I.I Annotations for Section 4.2: Identifying, Analysing, and Selecting Online Critical Data Literacy Resources

### I.I.I Section 4.2.2: Conducting the Content Analysis – Sources and Sampling

*Detailed description of how the sampling for the content analysis was conducted:*
As outlined in the methods chapter, the content analysis aimed for a sample that is as-broad, diverse and comprehensive as possible. As no list of online critical data literacy resources exists, no sampling frame was available and a representative sampling not possible. Instead, a broad and diverse sample was reached by identifying resources through a number of different sources. As a first step to find resources for my sample, all online resources identified in my prior work were examined. This included my 2018 snowball sampling (2020c), which led to 31 resources; an informal mapping and analysis of German-language data literacy resources I undertook in 2020 (see Sander 2020b), which included 41 resources, and a guidebook on critical data literacy tools I co-developed with Jess Brand (2020) with 15 resources. As a next step, I examined all recommendations by the project partners, the *Data Justice Lab* and the NGO *Privacy International*. Eight resource recommendations emerged during the *Data Justice Conference 2021*, and *PI* recommended 24 resources as particularly relevant for my content analysis.

After this, I examined resources that were included in the online database by the Critical Big Data and Algorithmic Literacy Network (CBDALN) as of 07.05.2021 ([no date]). CBDALN is an international network of critical data literacy scholars and practitioners I co-founded in 2020 and that at the time of writing has 22 members. The database collects online resources that critically educate about datafication and that were recommended by network members. The database's criteria for inclusion are similar to my study's selection criteria for critical data literacy resources. The resources should:

- Raise awareness about the collection, analysis and use of our data.
- Emphasise the long-term social implications of such data systems and of this "datafication" on our societies and on individual citizens.
- Critically examine the technological basis of data systems and highlight influences on human interactions and social systems.
- Foster critical thinking, support the ability to form one's own opinion and critically reflect the collection and analysis of personal data.
- Be available (at least) in English or German (Critical Big Data and Algorithmic Literacy Network 2022).

The database can be filtered by language, format, target audience, and license of the resources. I examined all resources that are available in English or German and that were categorised as "introduction", indicating that there is no prior knowledge required. At the time of sampling, this included 90 resources. While I co-developed the database and have added some resources myself, I do not have any influence on the resources that are recommended by other network members. The database thus constitutes a valuable collection of resources that experienced critical data literacy scholars and practitioners view as 'good' literacy tools, therefore making it an ideal source for my sample. In light of the similar inclusion criteria, it is not surprising that the database emerged as the most 'successful' source, leading to the identification of 39 resources that fit my study's criteria for critical data literacy resources.

Finally, these sources were complemented by a list of potential critical data literacy resources I had maintained for two years prior to starting the content analysis. It includes all online resources that I came across or that others had recommended and that, on first glance, seemed to be educating about datafication. All 66 resources on this list were examined, but not all qualified for the sample. Moreover, nine additional resources were identified via other, already included, resources and were examined because they seemed particular relevant for my study. In total, these eight sources led to the identification of 250 resources. Several of those were recommended by two, three or even four of the different sources. For reasons of transparency, I recorded the source and which selection criteria were met for each resource that was examined.

### I.I.II Section 4.2.2: Conducting the Content Analysis – Codebook and Coding Sheet

*Detailed description of the coding sheet and methodological considerations behind its design:*
As detailed in the methods chapter, the coding sheet consisted of four sections with a total of 19 variables. Of the 19 variables in total, ten identified the unit of analysis and provided information on the first two research questions outlined before. Nine further variables aimed at answering the third and fourth research question on

the resource's similarity to the preliminary theoretical framework for critical datafication literacy. The following paragraphs present the variables and methodological considerations behind the design of the coding sheet and codebook in more detail.

The first three variables (1. Title; 2. URL; 3. Origin of resource) make up section A of the codebook (see appendix III), and record which resource is being coded and from which source it originated. This aimed to ensure transparency and traceability. The next three variables constitute the qualifying criteria for this sample (section B): 4. Education about datafication, 5. Critical reflection, 6. No prior knowledge. If a resource met all three criteria, it was included in the sample and the next categories were coded. If a resource did not meet *all three* of these basic criteria, it was excluded from further analysis.

For every resource that met the qualifying criteria, the next seven variables (section C) were coded, aiming at describing the resource and its creator: 7. Creator name; 8. Creator background; 9. Country of origin; 10. Format; 11. Language(s); 12. Publication date; and 13. Target group. These variables were either open fields (e.g., creator name) or included several answer options that were selected based on my prior research (e.g., creator background). Moreover, these variables always included an "other" option if no predefined option was suitable. In this case, the full answer was entered in the field. All answer options for all variables as well as detailed descriptions of each variable can be found in the codebook (see appendix III). Section C further checked for the age of the resource. If no publication date could be identified – as often the case with online resources – this field was omitted. However, if a resource that was older than 2015 was identified, it was excluded from further coding. Due to the fast pace of technology development and the profound changes related to the datafication of our societies in the last years, resources older than 2015 will likely not include current issues of datafication. Also the field '13. Target group' could be omitted if the target audience of a resource was not obvious from using the resource.

The last section D (variables 14–19) examined to what extent a resource was in line with the preliminary framework for critical datafication literacy and thus tested for research questions three and four, including the variables: 14. Interactive design; 15. Different audiences; 16. Practical advice; 17. Individual responsibility; 18. Real-life examples; and 19. Special category. As some of these variables require a certain amount of interpretation or judgement from the coder, I followed Hansen's recommendation of laying down "very clear interpretation guidelines" for these variables (1998, p. 115). The codebook included not only extensive definitions for these variables, but also provided examples of how each variable has been implemented in critical data literacy resources that were identified in my prior study (see appendix III). The variables 14–19 describe the resource's content as well as format. All categories (except for variable 16: practical advice, and 19: special category) were represented as Yes/No fields that were ticked if the resource fit the category. During pi-

loting, these fields were further differentiated (see section I.I.III). The more criteria a resource met, the better it fit my literacy framework. However, none of these criteria were compulsory, but they were only intended to guide the selection of expert interviewees and provide more information about the analysed resources.

Finally, the last variable, "19. Special category" requires some further explanation. This variable was intended to highlight resources that take an unusual approach (content or format), fit the project focus extremely well or were especially recommended by established practitioners. This variable was included for a number of reasons. First of all, content analysis focuses on frequencies and can sometimes neglect "the rare and the absent" (Bauer 2000, p. 148). As online critical data literacy resources are still very under-researched, this open variable helped to avoid the risk of asking the 'wrong' questions in the content analysis and thus overlooking relevant but unexpected aspects. Related to this, a "special category" variable further helped to avoid a 'self-fulfilling prophecy' in which resources are analysed in respect to a certain framework, and then, indeed, fit this framework. As the intention was to learn from the content analysis for further developing the preliminary theoretical framework, it was important to record and highlight unusual or unexpected approaches. Finally, variable 19 helped to stay open-minded in the final selection of interviewees because it allowed to highlight unusual resources but also those that came highly recommended from project partners and established practitioners. This enabled an open-minded final selection of resource creators for the expert interviews, not only rigidly following predefined criteria, but opening up the possibility to think and research outside the box. Moreover, a "comments" category was included, in which first impressions of each resource or further explanations in relation to the coding of the resource could be noted in a logbook-style.

Coding itself took place using an Excel table that included all variables. To save time and due to the large number of resources, I decided against creating individual coding sheets for each resource. Using an Excel table further allowed me to quickly enter responses to Yes/No variables and also record open text for other variables. To simplify coding and prevent errors, the columns in the Excel table were colour-coded according to the sections A-D, and initials for different answer options were used rather than numbering them (e.g., N for "NGO/Civil Society" in variable 8). As recommended in the literature, I also kept a logbook during coding to make notes and in case I "encounter certain examples that do not fit neatly within [my] pre-designed categories" (Deacon et al. 2007, p. 130). In this case, once I decided on a coding solution, I noted it and could be consistent in my coding in similar cases in the future. In these situations and generally throughout the content analysis, being the only coder helped minimise reliability challenges such as inter and intra-coder reliability (see Bauer 2000, p. 143f.).

## I.I.III Section 4.2.2: Conducting the Content Analysis – Piloting

*Detailed description of changes made during and after the content analysis pilot:*
A number of changes were made to the coding sheet based on the pilot study. The first finding of the piloting was that coding directly into the excel table was a very time-efficient method and worked well. Furthermore, some changes to the order of variables were made, moving the qualifying criteria to the front so that I would not have to collect a lot of information on a resource that then ended up being excluded from the sample. Moreover, in an effort for more transparency and traceability, I decided to add the variable "3. Origin of Resource" as it may be useful to know how a resource ended up in my sample and also which resources were identified through more than one source. Piloting further confirmed using initials instead of numbers to signify different answer options as time-efficient and error-reducing. Moreover, I decided to highlight in bold font if a resource fit a category extremely well and by "(X)" if a resource only somewhat fit a variable. For example, it might be useful to differentiate between resources that include one interactive element, like a short quiz, in contrast to those that take a fundamentally interactive approach, such as "Do Not Track" that stops playing without input from the user.

I further came across a number of problematic issues in my pilot study. First, a number of resources recommended further resources about datafication, and I was uncertain whether I should follow these recommendations, leading to a snowballing effect. Given the large number of resources already included in my sample's sources, I decided against examining these recommendations immediately. Instead, I decided to highlight these cases as "C – collection of resources" (variable 10. Format). In case my existing sources would not have led to a sufficient sample, I could have used these to find further resources. Only one exception was made: I identified several highly suitable and very unusual resources via a website created by Tijmen Schep for the Project Sherpa. These nine resources were analysed and four fit my criteria and were fully coded.

I further realised that a number of resources were too extensive and time-consuming to analyse in whole, such as an online course containing five one-hour lessons or a podcast with possibly hundreds of episodes. In these cases, I could not always be entirely certain if the resource, for example, included real-life examples hidden in the many hours of content. To record this uncertainty, these cases were coded as "??". Moreover, the coding revealed the volatility of online resources that are prone to change and develop over time, making it difficult to determine which version of a website I coded in my content analysis. To prevent this at least to a certain extent, I decided to save a local copy of each resource's home page on my computer. While this does not allow for further browsing in the resource, it at least records a first impression of each resource in the sample.

After receiving feedback on the variables and pilot study findings from the project supervisors and the project partner Gus Hosein from Privacy International, it was further decided to differentiate several answer options for variable 16. "practical advice" in order to collect more in-depth data and to allow for cross-comparisons.

## I.II Annotations for section 4.3: Learning from the Experts

### I.II.I Section 4.3.2: The Sampling Process and the Final Sample – Sample Selection Process

*Detailed description of the sample selection process*

In order to narrow down my initial sample of 75 resources that were analysed in the content analysis to a final sample of 10 resources for the expert interviews, several selection steps needed to be taken. As the next paragraphs will outline, some selection steps aimed at identifying the 'best' resources based on criteria informed by theoretical and empirical research, while others aimed at a high *diversity* in the final sample. As argued in chapter 2.2, research on critical data literacy suggests that there is no "one-size-fits-all" approach when it comes to literacy, and different audiences with different capacities as well as different learner types require different approaches to learning about datafication (e.g., Aßmann et al. 2016; Carmi et al. 2020; Pinney 2020). Thus, the final sample should include resources with different formats, thematical foci, that originated in different countries, and that were developed in different creation backgrounds. It was further important to me to keep an open mind during this selection process and to put special emphasis on unusual approaches and resources, hoping to broaden my horizon and gain new perspectives for my theoretical framework.

The *first and second selection step* were informed by existing research on critical data literacy as well as my prior experience in this field. *First*, I decided to exclude all resources that did not include any practical advice for citizens.[1] The importance of constructive advice on how to protect one's data, use alternative tools, make one's voice heard or other constructive 'next steps' has been highlighted in my prior research as well as in other studies (Pangrazio and Sefton-Green 2020, p. 218; Sander 2020c, p. 13; Bilstrup et al. 2022, p. 234). The rationale behind this is that constructive next steps can encourage learners to take action, thus giving them an (albeit limited)

1 For some resources, this could not be determined with certainty, mainly due to their extensiveness (see also I.I.III). Given the importance of constructive advice, these were also excluded in this step.

sense of agency, as well as helping to prevent resignation in light of new knowledge about the risks of datafication.

As a *second step*, all resources that placed particular emphasis on the responsibility of individuals to protect their data were excluded. During coding the resources as part of the content analysis, I noticed that some resources convey the impression that it should be up to the individuals to protect their data in order to avoid negative consequences of datafication. As highlighted in the theoretical framework chapters, this notion is highly problematic. A shift of responsibility to individuals can overwhelm users and potentially lead to discouragement and resignation. Moreover, citizens' agency to protect their data is limited and data collection often takes place in contexts in which citizens do not have a choice to opt out (Aßmann et al. 2016; Mihailidis 2018; Pangrazio and Selwyn 2019; Carmi et al. 2020).

These first two selection steps narrowed down my sample to 55 highly suitable resources. In order to select ten final resources from these 55, I took four different approaches:

a) Examining which resources fit *all* content analysis criteria (7 resources);
b) Identifying which resources were highlighted as *unusual, particularly interesting or very suitable* through the "special category" and / or the comments section of my content analysis (11 resources);
c) Identifying *outliers* (17 resources);[2] and
d) Taking *practical considerations* into account (5 resources).[3]

Through these parallel approaches, a total of 27 resources were identified, including several that were selected in more than one of the four approaches. In order to narrow down to a final sample of ten resources, I examined which resources were *selected in several of these four approaches*. Moreover, I reassessed the resources in light

2 As I aimed for diversity in my final sample and was interested in learning from different and diverse resources, I decided to place particular focus on unusual resource characteristics that stood out from the content analysis findings. Such outliers included, for example, an unusual creator background (e.g., museum); country (e.g., Brazil); format (e.g., email, app or yoga instructions); or target group (e.g., activists, parents, or technologists and policymakers). Moreover, resources that were offered in more than four languages or that applied more than four different formats were considered outliers.

3 As outlined in chapter four, several preliminary expert interviews informed the preparation and conduct of the final interviews. When these preliminary interviews took place, the final sample of my study was not yet determined. However, four resource creators that were interviewed in these preliminary interviews and who had agreed to a follow-up interview at a later point in time emerged in the sample at this point in the selection process. Selecting these came with the advantage that they had already agreed to be interviewed, thus reducing the risk of being rejected and having to find an alternative interviewee.

of my *project focus* and compared key characteristics such as format, country of origin, and creator background, aiming for an equal distribution of *diverse perspectives* in my final sample. This led to a *preselection of 'finalists'*, which were then discussed with the project collaborator Privacy International (represented by Gus Hosein) in detail, and a sample of ten resources was decided.

Unfortunately, out of the ten resource creators selected in this process, only six agreed to be interviewed. Some others expressed their interest or even agreed, but then could not be reached anymore. Therefore, four other resources had to be selected. Here, the same criteria as described above were considered and I tried to replace each of the four resources with a similar one in order to maintain the diversity of different characteristics in the sample.

### I.II.II Section 4.3.4: Preparing and Conducting the Interviews

*Privacy and methodological considerations around the use of virtual face-to-face interviews:*
As outlined in the methods chapter, all interviewees chose to be interviewed via video call. Using virtual face-to-face interviews came with several specific benefits, such as lifting geographical restrictions, but also "*ease and flexibility of scheduling*", including saving money and time spent travelling; "*virtual and visual interaction*", allowing for synchronous visual interaction; "*ease of data capture*" by using audio recording software; "'*public' places and 'private' spaces*", as video calls from home offer privacy from the public and also from the researcher; and "*greater control for participants*" when they are invited to choose their preferred interview format – and, in my case, platform (Hanna and Mwale 2017, p. 259ff, emphasis in original). While giving participants the choice of platform provided more convenience for them, it required flexibility and a certain privacy pragmatism from my side. Considering my research topic, I would have preferred to use privacy-sensitive tools such as Jitsi and Big Blue Button. Unfortunately, these often do not work as smoothly as other platforms, and many people are not as familiar with them. Thus, adapting to my interviewees' wishes and aiming for a smooth conversation without technical disruptions meant using less privacy-sensitive tools.

However, I did not use the built-in recording options of the video calling platforms because the data flows of the recorded data and the companies' access to this data was not transparent and safe enough in order for me to ensure the protection of my participants' data. Instead, the interviews were recorded through software on my local device, with participants giving consent in written form as well as verbally. Audio-recording qualitative interviews helps to gather accurate data while also allowing the researcher to "focus on the flow of the interview, to think about follow up questions, […] and build up rapport with the interviewee" (van Audenhove and Donders 2019, p. 192).

The normalisation of video calling for work and for staying in touch with loved ones during the COVID-19 pandemic likely helped to foster a natural conversation flow in my interviews. However, video interviewing is not without limitations. In 2021, many people – especially those working from home – experienced a certain frustration, or "Zoom fatigue", in light of too many video calls (cf. Brown Epstein 2020; Bailenson 2021). Moreover, technical problems and poor internet connection can have an impact on the quality of conversation and on "rapport and ultimately on the quality of data collected" (Hanna and Mwale 2017, p. 267). Apart from this, participants need appropriate equipment for video calls (ibid., p. 261). Yet, in my study with creators of online resources, this latter aspect was less problematic. Overall, conducting virtual face-to-face interviews was thus not only a workaround, but in fact a suitable method for my study's goals and sample. Furthermore, already having conducted several expert interviews in person, via telephone and through video calling in research projects in the past helped me navigate the process and its challenges.

## I.II.III Section 4.3.5: Analysing the Interviews

*Detailed description of how the different steps of the thematic analysis combined deductive and inductive elements:*
The first step after transcribing the interviews was to create a list of deductive nodes and categories based on the theoretical framework for critical datafication literacy and the research questions for the interviews. Three categories and nine nodes were developed. However, as will be outlined below, the analysis later showed that only some of the deductive nodes fit the codes derived from the data and an inductive, 'bottom-up' approach worked better to capture key themes and patterns in the data. Thus, the majority of themes in the final thematic framework were identified inductively. The three deductive categories "understanding of literacy", "strategies and principles on how to foster this literacy" and "challenges and practical considerations", however, proved helpful in sorting and gaining an overview of the large number of initial codes (see below).

The next steps in the analysis followed the recommendations for conducting a thematic analysis of Meuser and Nagel (2009) in combination with Braun and Clarke (2006). In the first phase of the analysis, I familiarised myself with the data by reading it several times (Braun and Clarke 2006, p. 16f), and I paraphrased the entire interview transcripts.[4] The paraphrase followed the "unfolding of the conversation" and aimed to "give account of the interviewee's opinions" (Meuser and Nagel 2009,

4 Key information on the resource creator's background and on the resource itself, such as position or training background of the creator or year of publication of the resource were not paraphrased or coded but saved in a separate document.

p. 35). Additionally, I collected particularly catchy quotes that might be used in the findings chapter in a separate document so that they are not lost in the large number of codes later on. In the second phase, the initial codes were created in NVivo. This process condenses the material and "order[s] the paraphrased passages thematically", while keeping close to the text and, if possible, adopting the language of the interviewee (Meuser and Nagel 2009, p. 36). In line with Braun and Clarke's recommendations, the goal was to create as many codes as possible in order to capture all data extracts and all potentially interesting aspects of the data (2006, p. 19). Thus, the initial list of codes consisted of 332 codes, which were used 1–13 times in the interviews.

In the next analytical phase, this long list of codes was revised, and the initial themes were identified. A first attempt to sort the initial codes into the deductive nodes quickly showed that several of these nodes did not fit the codes and others were too broad, not leading to any meaningful themes. Thus, I switched to a bottom-up, inductive approach of identifying themes, which was more productive and 'natural' – sticking close to the data. However, the three deductive categories outlined above proved helpful in providing a first overview of the large number of initial codes. Thus, in a first step, all 332 initial codes were sorted into one of the three categories, with only few codes remaining in an "other" category. After this, I checked and revised all codes in each of the categories, combining similar codes, differentiating too broad codes and in some cases deleting codes that were irrelevant for the research questions (see Meuser and Nagel 2009, p. 36). At the same time, I searched for patterns in the codes, considering how "different codes may combine to form an overarching theme" (Braun and Clarke 2006, p. 19). Also here, a bottom-up approach proved most valuable – first identifying *subthemes* and later examining which broader patterns can be found among these subthemes, leading to the *main themes* of the analysis. Throughout this process, codes and themes were continuously revised, merged with others and sometimes moved between the categories.

This led to an initial framework of 14 main themes and 61 subthemes. As outlined in chapter four, this was visualised in the form of a mind map. In the next, fourth phase of the analysis, all themes (including their total of 680 extracts as well as notes made during initial paraphrasing and coding) were reviewed and edited in several rounds of revision. The goal was to ensure that each theme formed a coherent pattern; was distinct; that the themes overall "'accurately' portrayed the meanings evident in the data set as a whole" (Braun and Clarke 2006, p. 19); and that they were relevant for the research questions.

Through this extensive revision, a final thematic framework with 12 main themes and 45 subthemes was developed. The next analytical phase then consisted of finalising the names of the themes and categories. Based on Braun and Clarke's recommendations, the names should capture the "essence" of each theme and ideally be "concise, punchy, and immediately give the reader a sense of what the theme is

about" (2006, pp. 22; 23). After this, the mind map visualising the thematic framework was updated to represent the final themes and theme names. Moreover, the key quotes collected at the beginning of the analysis were complemented and sorted corresponding to the final thematic framework. Finally, as outlined in chapter four, the themes were analysed according to the "story" they and the data overall tell (ibid., p. 22), and connections to academic literature, prior findings and my study's theoretical framework for critical datafication literacy were drawn (Meuser and Nagel 2009, p. 35f).

# I.III Annotations for section 4.4: Learning from the Educators

## I.III.I Section 4.4.2 – The Survey Sample – Finding the Sample

*Detailed description and reflection on the process of reaching a diverse survey sample:*
As outlined in the methods chapter, the population of interest for my study's qualitative online survey consisted of educators from Germany, the United Kingdom and other places in Europe, who are interested in teaching about digital technologies and datafication. As most relevant groups, I identified teachers, student teachers, teacher trainers, higher education lecturers, adult educators, media education centres, and trainers in civil society.

To reach this population, the most appropriate communication channel(s) for each educational profession in each of the three (supra)national contexts had to be identified. This selection was not only informed by my prior experience in the field of critical data literacy research, but also by several experts I contacted. These included, among others, researchers directly working with educators from different backgrounds, and current and former civil society activists in the field of critical data literacy. In discussion with these experts, the project partner Privacy International, and the project supervisors, a list of communication channels was assembled through which the survey would aim to reach a broad and diverse sample of educators from different fields (see appendix VII). Attention was paid to aiming at as much diversity as possible, yet ensuring consistency throughout the different contexts. In line with Clark et al. (2021, p. 186), the list included mailing lists, social media channels, and individual contacts who would be able to distribute the questionnaire even further, aiming to reach as many members of my target population – educators from different backgrounds interested in educating about datafication – as possible. Moreover, every contact was invited to share the survey information and link with other educators they know who might be interested in the survey's topic.

Clark et al. further highlight the typically low response rates for online surveys (2021, p. 186). This proved to be a limitation of my study as well, although I took considerable efforts to reach a big sample. As represented in appendix VII, a large number of educational organisations, groups, mailing lists and interested individuals

were contacted. However, a limitation of this sampling technique is that it is unknown how many and which people the survey invitation reached, as it is not always possible to determine whether an organisation or individual shared the information about the study within their organisation or mailing list. Yet, in some cases, I received confirmation. Therefore, it can be said with certainty that the survey link was circulated:

- In the members newsletter of the German society for media pedagogy scholars (GMK);
- In the mailing list by the British Media, Communication and Cultural Studies Association (MeCCSA);
- In the Weekly Digest mailing list by the European Communication Research and Education Association (ECREA);
- In a German newsletter for educators interested in learning more about digital educational technologies (UNBLACK THE BOX);
- By the social media channels of several educational organisations, such as The Media Education Association and the Association for Citizenship Teaching;
- Among several relevant German and European educational research projects;
- In a German teacher training centre;
- By several teachers in Germany and the UK;
- And by several higher education colleagues in different European countries, who shared the survey in their personal networks.

Already through these channels – particularly through the large mailing lists by different associations – it is likely that the survey information reached several thousand people. Additionally, the promotion via Twitter was successful, with many individuals as well as organisations retweeting the survey invitation, including the Association for Teacher Education in Europe, the British Media Education Association, an Irish educational research journal, a European association for media and learning, a German media pedagogy association, a popular German guide to teaching material (bildungsserver.de), a research institute and several researchers and media pedagogues. In total, the main English-language tweet reached 6,907 impressions and generated 49 link clicks, and the main German tweet reached 1,683 impressions with 26 link clicks.

Moreover, I sent several rounds of invitations in order to reach a sample that is as diverse and equally balanced as possible. In a first round of invitations, I contacted 46 groups and individuals via email, website contact forms and sometimes Twitter in the end of November 2021. A first examination of the collected data by the beginning of January 2022 showed that 38 people had completed the full questionnaire, but there was a predominance of German participants and those from formal education fields. I tried to address this imbalance, in a second round of invitations.

This included sending several reminders to those groups and individuals from a UK and EU background who had not responded yet, and sending invitations to 29 new contacts from British or European background, including many from the non-formal education sector. This helped to balance the sample and was complemented by a third round of invitations in the end of January 2022, with several reminders as well as invitations to four new groups and individuals, again aiming for a diverse and balanced sample. Through these steps, a diverse and fairly balanced sample was achieved (see also chapter 5.3.1).

## I.III.II Section 4.4.3 – Developing the Questionnaire – Pilot Study

*Detailed description of changes made to the questionnaire after the pilot study:*
As outlined in the methods chapter, the survey pilot testers provided detailed and extensive feedback and constructive suggestions. Most comments related to small changes in wording, which were all implemented. There were several issues related to translation, mainly regarding the different German terms that can be used for "educator" and "teaching". In the German language, there are no common broad terms that summarise all kinds of teaching and being an educator. Instead, the German language tends to differentiate for example between teaching in schools ("unterrichten") and being a school teacher ("Lehrer/in") in contrast to teaching in university ("lehren") and being a lecturer ("Dozent/in" or "Lehrende"). It was thus difficult to find a wording that included all areas of education and I decided to use more than one term in some instances (e.g., by writing "unterrichten/lehren").

The pilot testers further recommended separating several of the open questions because they were asking about too many aspects in one question. In several places, the pilot testers suggested adding examples in order to clarify the question or the answer options, or they suggested smaller changes in question or answer option wording. These suggestions were implemented in all instances. For example, several answer options in question five were renamed: such as "non-formal education (e.g., media education centre)" instead of "media education". In question 17, the first statement "Interactive approaches work well and are popular with learners" was changed to "Interactive approaches are a great way to engage learners" (see appendix VIII). This captures my intended meaning similarly well without – as criticised by pilot testers – asking about two different things in one question. Moreover, more details on my research project were added to the landing page.

## I.III.III Section 4.4.3 – Developing the Questionnaire – Final Questionnaire

*Detailed information on the methodological considerations behind the questionnaire design:*
Apart from recommendations from the methodological literature, the development of the questionnaire was strongly influenced by my study's previous theoretical and empirical findings. The survey aimed to investigate if these findings corresponded with educators' daily lived experience and thus aimed to examine the following questions:

a) To what extent are topics related to datafication already being covered and critical data education already fostered by educators?
b) Does the predominance of practical digital and data skills over critical reflection of the societal implications of datafication that I found in the theoretical literature as well as in many resources I analysed correspond with the topics that educators predominantly covered in their teaching?
c) How well-equipped do educators feel to teach about these different dimensions of critical data literacy?
d) Do educators know about the resources I analysed in my content analysis and if yes, what do they think of them?
e) Does the popularity of certain formats I found in my prior study (Sander 2020c) correspond with the formats educators prefer?
f) Do educators – based on their practical experience – agree with key findings on 'best-practice' approaches to educate about datafication that have emerged from my own and other scholars' research on critical data literacy?
g) And finally: What do educators need in order to better educate about datafication and what are their wishes for future critical data literacy resources?

After I created a first draft questionnaire, it was revised and condensed several times based on feedback from the project collaborator Privacy International as well as based on recommendations from methodological literature (see e.g., Reja et al. 2003; Braun et al. 2021; Clark et al. 2021). The final questionnaire used in my survey consisted of a landing page that provided information about the study; a short demographics section (1); a section on educators' experience with topics around digital technologies, (big) data and datafication (2); a section on educational resources about datafication, including a final open question asking for additional comments (3); and a last page that thanked for the participation in my study and asked participants if they were interested in testing and providing feedback for the resource developed with PI (for whole final questionnaire, see appendix VIII).

The landing page contained information on my research project and its goals, details on the survey and data handling as well as contact information for participants who had questions or wanted to express their interest in the study's find-

ings. As outlined in the methods chapter, this also included information necessary for *informed consent* (see e.g., Regmi et al. 2017, p. 642; Braun et al. 2021, p. 8f). Moreover, the landing page specified the study's target population and stated the estimated completion time for the questionnaire, giving a realistic estimation, as recommended in the literature (Clark et al. 2021, p. 187).

The first section on demographics then asked for a small number of basic personal details (age in groups; gender, providing sufficient options and including the option to self-describe; country of residence; and nationality) that would help to situate findings within these contexts. Moreover, one question asked about the area of education that participants worked in (multi-select question, providing nine options and an "other" field) and another invited them to describe their individual position or role in an open text field. These questions aimed at understanding in which educational contexts aspects of datafication might already be covered, but also discovering if the survey was successful in reaching diverse educators from different fields and countries, as had been intended.

The second section asked about the educators' experiences with teaching about topics such as digital technologies, (big) data systems and their implications on society. This section included five questions. The first two used a rating scale to examine how *well-equipped* and how *experienced* educators felt in teaching about different aspects of digital technologies and (big) data. Here, and in all other rating questions, a 5-point rating scale was used. This gave participants the opportunity to indicate if they felt neutrally about a certain option. A "N.A." option further allowed to indicate if a question was not applicable to participants. Adding a visual scale above the numbers and using the same 5-point scale for all rating questions further aimed at a user-friendly design.

For both questions (7 and 8), the broad thematical field of digital and data technologies was separated into four key topics: Digital technologies in general; data security; (big) data systems and algorithms; and the way digital media and (big) data affect society. These four topics were informed by findings that highlight the prevalence of concerns around digital technologies in general and data security questions in contrast to less known issues around algorithmic systems and the way these systems affect society (see literature review as well as Sander 2020). I was curious to see if the same prevalence could be found in educators' knowledge and experiences. In order to clarify each topic, pilot testers suggested to add example questions for each topic, which were displayed in smaller font below each topic. Both rating questions were placed on the same page for ease of use. Moreover, an explanatory sentence was added for each question based on pilot tester feedback, emphasising what was meant by "well-equipped" and "experienced".

Subsequently, three open questions (9–11) asked about more details on the educators' experience with teaching about digital and data technologies, such as the exact topics covered, the key skills and understanding aimed for, and inviting partic-

ipants to add further information on the context, methods, successes and challenges they encountered in teaching about these topics.

The third section focussed on educational resources about datafication. This section included four open and three rating questions and investigated how educators inform themselves on topics around digital technologies and data (12), how they find teaching material on these topics (and if there are any challenges) (13), and how satisfied they are with their access to information and teaching material about digital technologies and (big) data (14). In addition, educators were asked to rate the usefulness of 12 different design formats of educational resources from "not at all useful" to "extremely useful" (15). The selection of formats was made based on findings from my prior research and this study's analysis of online critical data literacy resources (see chapter 5.1). Examples were provided for formats that might be unclear. The rating question was followed-up by an open question (16) inviting participants to name examples of useful resources they have used and to provide more details on what makes a resource useful for them.

The next and penultimate question of the survey (17) was likely the most complex question to design. It asked participants to indicate to what extent they agree with ten statements about how best to educate about digital technologies and data systems. In order to develop these ten statements, I reviewed my prior empirical findings (Sander 2020c), findings from this study's literature review and theoretical framework, further theoretical literature on how best to implement critical data literacy as well as first findings from this study's content analysis of online critical data literacy resources. In a second step, I tried to condense all these findings into simple statements that reflected academic findings on critical data literacy as well as ongoing academic debates, for example on whether or not practical data skills are necessary for critical data literacy. I deliberately included controversial issues, such as the question of "individual responsibility" or "shocking learners", in these statements. The statements were revised, condensed and simplified several times to make them as concise and clear as possible. Key terms for each questions were underlined for further clarification and ease of reading. Feedback from pilot testers about this question was consistently very positive, indicating that efforts for simplification and clarification were successful.

Finally, as recommended in the literature, the questionnaire ended with a final open question, inviting participants to share final remarks and additional comments that may have not been covered by the questionnaire thus far. As Braun et al highlight, this can often generate "unanticipated and useful data" (2021, p. 8). In addition to this open invitation, several suggestions were made on what participants could comment on: for example, on what they would need to be better able to educate about digital technologies and data systems, whether there are any particular kinds of resources they need but can't find, or on their overall wishes for educational resources. Through these suggestions, I hoped to inspire and nudge participants to

elaborate more freely on their needs and wishes when it comes to critical data literacy resources.

After participants submitted their answers, the survey ended with a final thank-you-page, which also invited the participants to contact me via email if they were interested in testing and providing feedback on a new educational resource that was being developed with Privacy International as part of this project. In order to ensure anonymity in the collected data, I asked the educators to email me rather than collecting their email addresses through the survey. This likely led to fewer (8) responses from potential testers, but anonymity was ensured.

# Appendix II: Content Analysis Sample

| Nr. | Title | URL (all: analysed in April/May 2021, last accessed: March 2024) |
|---|---|---|
| R1 | Your data matters | https://ico.org.uk/your-data-matters/ |
| R2 | My data and privacy online. A toolkit for young people | http://www.lse.ac.uk/my-privacy-uk |
| R3 | Breaking the Black Box | https://www.propublica.org/article/breaking-the-black-box-what-facebook-knows-about-you |
| R4 | The Privacy Paradox. Note to Self | Main: https://project.wnyc.org/privacy-paradox/ Also (newer): https://www.wnycstudios.org/podcasts/notetoself |
| R5 | Advocacy Assembly – Courses about Privacy | https://advocacyassembly.org/en/courses/ – Topic: Privacy |
| R6 | Chupadados the Datasucker. The hidden faces of our beloved technologies | https://chupadados.codingrights.org/en/ |
| R7 | Youngdata | https://www.youngdata.de/ |
| R8 | Digital Shred. Privacy Literacy Toolkit | https://sites.psu.edu/digitalshred/ |
| R9 | Do Not Track | https://donottrack-doc.com/en/ |
| R10 | Me and My Shadow | https://myshadow.org/ |
| R11 | Privacy International | https://privacyinternational.org/ |
| R12 | Surveillance Self-Defense, EFF | https://ssd.eff.org/ |
| R13 | A data-day | https://tacticaltech.org/news/a-data-day/ |
| R14 | The Power of Privacy — Documentary | https://www.youtube.com/watch?v=BvQ6I9xrEu0 |
| R15 | Free your data | https://freeyourdata.org/de/ |

| | | |
|---|---|---|
| R16 | Daten, Daten, Daten | https://www.politische-bildung.nrw.de/themen/big-data-story |
| R17 | Anna. Das vernetzte Leben | https://www.annasleben.de/ |
| R18 | Lernparcours Big Data | http://bigdata.jfc.info/lernparcours.html |
| R19 | Deine Daten Deine Rechte / Your Data Your Rights | https://deinedatendeinerechte.de/ |
| R20 | DATASELFIE | https://dataselfie.jnw-sdm.ch/ |
| R21 | How to Selbstauskunft. Kampf um meine Daten | https://www.br.de/puls/themen/netz/selbstauskunft-kampf-um-meine-daten-100.html |
| R22 | Selbstdatenschutz und digitale Selbstverteidigung | https://www.selbstdatenschutz.info/ [Not available anymore. Analysed version: https://web.archive.org/web/20210417193332/https://www.selbstdatenschutz.info/ ] |
| R23 | Privacy-Handbuch | https://www.awxcnx.de/ |
| R24 | Selbstdatenschutz! Tipps, Tricks und Klicks | https://www.blm.de/files/pdf1/blm-selbstschutz.pdf |
| R25 | Dein Algorithmus – meine Meinung! Algorithmen und ihre Bedeutung für Meinungsbildung und Demokratie | https://www.mabb.de/files/content/document/UEBER%20DIE%20MABB/Download-Center/Publikationen/Broschueren/Algo-Broschur_20170403_mabb_neu.pdf |
| R26 | Big Data und politische Bildung | https://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/bigdata/ [Not available anymore. Analysed version: https://web.archive.org/web/20210507040640/https://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/bigdata/ ] |
| R27 | Dein Netz | https://dein-netz.org/hi/ [Not available anymore. Analysed version: https://web.archive.org/web/20210517035140/https://dein-netz.org/hi/ ] |
| R28 | Klicksafe | http://www.klicksafe.de/ |
| R29 | Dein Tag in Daten | http://www.watchyourweb.de/assets/watchyourweb/mdb/2/Infografik_Datentag.png |
| R30 | Watch your web: Big Data | http://www.watchyourweb.de/p2304197173_564.html#fd8ee25326f243a432e5cefe8ae8ac55 |
| R31 | Digital Defense Playbook. Community Power Tools for Reclaiming Data | https://www.odbproject.org/wp-content/uploads/2019/03/ODB_DDP_HighRes_Single.pdf |

| | | |
|---|---|---|
| R32 | Unbias Fairness Toolkit and Youth Jury resource pack | https://unbias.wp.horizon.ac.uk/fairness-toolkit/ And: https://uyj.wp.horizon.ac.uk/overview/ |
| R33 | Automating NYC and (en)coding inequality? | https://automating.nyc/ |
| R34 | #Nachgeschaut – Big Data Erklärvideo | https://www.mediasmart.de/2019/03/nachgeschaut-big-data-erklaervideo/ |
| R35 | Behind the Buzzwords – Big Data | https://www.bbc.co.uk/sounds/play/m000lghg |
| R36 | Von der flächendeckenden Überwachung zur flächendeckenden Lenkung der Bürger. Big Data als Lenkungswerkzeug | https://media.ccc.de/v/pw17-93-von_der_flachendeckenden_uberwachung_zur_flachendeckenden_lenkung_der_burger#t=3 |
| R37 | Big Data verstehen. Lehrer*innen-Leitfaden: Big Data | https://www.erlebe-it.de/Unterrichtsmaterialien/Big-Data-verstehen |
| R38 | Big Data-Forscher Dirk Helbing über den „perfekten Sturm" der Digitalisierung | https://www.youtube.com/watch?v=83cEm-kPEeg |
| R39 | Lehrmittel Big Data – Museum für Kommunikation | https://www.mfk.ch/bigdata/ |
| R40 | Big Up 4 Big Data | https://medienfachberatung.de/wp-content/uploads/2017/03/Spielanleitung-Big-Up-4-Big-Data_MFB-Schwaben.pdf |
| R41 | Datak – Spiel um deine Daten | https://www.datak.ch/#/start |
| R42 | Daten, Algorithmen, Kontrolle der Zukunft | https://www.surveillance-studies.org/daten-algorithmen-kontrolle-der-zukunft/ |
| R43 | Digital Literacy Best Practices | https://www.netliteracy.org/digital-literacy/ |
| R44 | Unterrichtsprojekt Digitalisierung – Leben in einer digitalen Welt | https://www.bpb.de/lernen/grafstat/304602/digitalisierung |
| R45 | Zukunftswerkstatt: Digitopia | https://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/bigdata/253169/zukunftswerkstatt |
| R46 | Future Influencer – Die smarte Schule!? | https://future-influencer-prototype.jimdofree.com/ |

| | | |
|---|---|---|
| R47 | Online-Konferenz Weiterbilden: Ich habe doch nichts zu verbergen! – für den verantwortungsvollen Umgang mit Daten sensibilisieren | https://www.gutes-aufwachsen-mit-medien.de/informieren/themen/news-detail/detail/online-konferenz-weiterbilden-ich-habe-doch-nichts-zu-verbergen-fuer-den-verantwortungsvollen-umgang-mit-daten-sensibilisieren |
| R48 | Big Data – inform@21 | https://inform21.ch/de/big-data/ |
| R49 | Big Data \| Medienportal der Siemens Stiftung | https://medienportal.siemens-stiftung.org/de/big-data-111955 |
| R50 | nano-Magazin Wissen: Big Data | https://www.3sat.de/wissen/nano/big-data-104.html [Not available anymore. Analysed version: https://web.archive.org/web/20210127145203/https://www.3sat.de/wissen/nano/big-data-104.html ] |
| R51 | nano-Magazin Wissen: Datenspuren | https://www.3sat.de/wissen/nano/datenspuren-100.html [Not available anymore. Analysed version: https://web.archive.org/web/20210120035638/https://www.3sat.de/wissen/nano/datenspuren-100.html ] |
| R52 | Panel Discussion "Big Data" | https://digitallearninglab.de/unterrichtsbausteine/panel-discussion-big-data |
| R53 | Podcasts zu Digitalisierung und Privatheit | https://www.leopoldina.org/themen/digitalisierung-und-demokratie/big-data-podcasts-1/ |
| R54 | Stadt \| Land \| DatenFluss | https://ki-campus.org/datenfluss |
| R55 | Überwacht und verkauft | https://projekte.sueddeutsche.de/artikel/digital/digitale-privatsphaere-ueberwacht-und-verkauft-e784992/ |
| R56 | Understanding predictive privacy harms | https://www.journalismfestival.com/programme/2016/understanding-predictive-privacy-harms |
| R57 | Unterrichtsmaterial: Rund um Daten | https://appcamps.de/unterrichtsmaterial/unterrichtsmaterial-zu-datenschutz/ |
| R58 | Watching You | https://www.studioimnetz.de/projekte/watchingyou/ |
| R59 | Data Literacy in the Real World: Conversations & Case Studies | http://datalit.sites.uofmhosting.net/wp-content/uploads/2017/08/data_literacy_in_the_real_world.pdf |
| R60 | The Internet of Everything | https://www.arte.tv/de/videos/RC-019504/internet-of-everything/ [Not available anymore. Analysed version: https://web.archive.org/web/20210620001719/https://www.arte.tv/de/videos/RC-019504/internet-of-everything/ ] |
| R61 | Mijente #TakeBackTech Curriculum Materials | https://drive.google.com/drive/folders/147CCSTLyN8OsfwFpUtw7mU-VCnJN58SX Background: https://notechforice.com/ |

| | | |
|---|---|---|
| R62 | Center for Humane Technology | https://www.humanetech.com/ |
| R63 | Screening Surveillance \| The Surveillance Studies Centre | https://www.sscqueens.org/projects/screening-surveillance |
| R64 | Erklärfilm: Der Weg der Daten | https://www.bpb.de/mediathek/318379/erklaerfilm-der-weg-der-daten |
| R65 | The emotion business: How companies are learning to read your mind | https://ig.ft.com/emotion-recognition/ |
| R66 | How normal am I? | https://www.hownormalami.eu/ |
| R67 | Project Sherpa – Videos | https://www.project-sherpa.eu/category/videos/ |
| R68 | What is mathwashing? | https://www.mathwashing.com/ |
| R69 | Social Cooling – big data's unintended side effect | https://www.socialcooling.com/ |
| R70 | STEALING UR FEELINGS | https://www.stealingurfeelin.gs/ |
| R71 | Datensammlung – Big Data (Modul) | https://de.digitale-lernwerkstatt.com/lehrer/sammlung/5f0c222cef587a00113e1748 |
| R72 | Schule macht Daten | https://ki-campus.org/courses/datenschule-jt2021 |
| R73 | Daten- und Algorithmenethik | https://ki-campus.org/courses/daethik |
| R74 | Coveillance: Watching the watchers | https://coveillance.org/ |
| R75 | Clear Your Tracks | https://www.clearyourtracks.org [Not available anymore. Analysed version: https://web.archive.org/web/20210511094939/http://clearyourtracks.org/ ] |

# Appendices III – IX

The remaining appendices can be found online:

https://tvl.ink/9783839473788

These include:
Appendix III: Content Analysis Codebook
Appendix IV: Participant Information Sheet
Appendix V: Participant Consent Form
Appendix VI: Expert Interview Guide
Appendix VII: Educator Survey Sampling
Appendix VIII: Educator Survey Questionnaire
Appendix IX: Report on Collaboration with Privacy International