

Kati Mozygemba & Betina Hollstein

In collaboration with Jan-Ocko Heuer, Elisabeth Huber, and Hai Ha Nguyen

Guidance on the anonymization and pseudonymization of qualitative text-based research data

Qualiservice, University of Bremen

Translated and adapted version of the Qualiservice Working
Paper QS-WP-5-2023, status: 2023

Suggested citation:

Mozygemba, Kati & Betina Hollstein (2026): Guidance on the anonymization and pseudonymization of qualitative text-based research data. QS-WP-9-2026, translated and adapted version of the Qualiservice Working Paper QS-WP-5-2023, status: 2023, Bremen,
<https://doi.org/10.26092/elib/4029>

© Qualiservice RDC, 2026



Imprint

Qualiservice Research Data Center (RDC)
University of Bremen
SOCIUM – Research Center on Inequality and Social Policy
Mary-Somerville-Str. 7
UNICOM Building
D-28359 Bremen
Germany



Website: <https://www.qualiservice.org/en/>
Email: info@qualiservice.org

Contents

1 Introduction.....	4
1.1 Background.....	4
1.2 Target group and areas of application.....	6
1.3 Clarification of terms.....	6
2 Weighing research ethics, data protection, and research interest.....	8
2.1 Complying with data protection.....	8
2.2 Considering research ethics.....	9
2.3 Research in the face of conflicting demands between data protection, research ethics, research interest, and good scientific practice.....	10
3 Qualiservice's anonymization concept.....	11
3.1 Informed consent as the basis for data processing.....	11
3.2 Purpose of anonymization.....	11
3.3 Anonymization as part of the research process and a cooperative task.....	11
3.4 The strategy of "flexible anonymization".....	12
3.5 Deletion of personal information.....	13
3.6 List usage and partial standardization.....	13
3.7 Further components of Qualiservice's data protection concept.....	14
4 Replace personal information step by step.....	15
5 Key categories for the replacement of personal information.....	21
5.1 Replacement of names.....	24
5.2 Replacement of time specifications.....	26
5.3 Replacement of location data.....	27
5.4 Replacement of information on training and occupation.....	31
5.5 Replacement of organizations and institutions.....	34
5.6 Information on economic sectors to specify professions and organizations.....	35
5.7 Replacement of health information.....	36
5.8 Replacement of positions and roles.....	37
5.9 Further information that may contain a personal reference.....	38
5.10 Accumulation of different categories of personal information.....	39
6 QualiAnon – The Qualiservice tool for anonymization or pseudonymization of text-based research materials.....	40
7 Summary and outlook.....	44
8 References.....	46
9 Appendix: Examples of the replacement of personal qualitative text data.....	49

1 Introduction

1.1 Background

Anonymization involves changing personal information about research participants, researchers, and any third parties that is contained in the research materials in such a way that it is no longer possible to link the information to individuals and the participants can no longer be identified (or only by employing a disproportionate amount of resources, which is equivalent to factual anonymization). Anonymization aims to protect the personal rights of those involved in qualitative research. The legal basis for anonymizing research data derives from the European General Data Protection Regulation (GDPR; European Union [EU], 2018).¹ Once the data has been anonymized, it no longer falls within the scope of the GDPR. In addition, anonymization represents a protection strategy that is part of the research-ethical requirements of good scientific practice (RatSWD, 2017b) and has been included in various subject-specific ethics codes and position papers on the handling of research data (e.g., German Research Foundation [DFG], 2019; German Association for Folklore Studies [DGV], 2018; German Sociological Association [DGS], 2017; German Association for Educational Science [GERA], 2005, 2016). At the same time, the anonymization of research materials invariably involves a loss of information. It is therefore important to develop strategies that protect the personal rights of both the researchers and the research subjects and at the same time preserve the scientific reuse value of their data. The specific purpose of anonymization (cf. point 3.2) is of particular importance for the development of a suitable anonymization concept (Medjedović, 2011; Opitz & Mauer, 2005).

The scientific reuse of data from qualitative social research is one such purpose and places particular demands on the development of an anonymization strategy. When one considers the example of biographical narratives or ethnographic field observations, it is easy to see that qualitative research, with its aim of comprehending the complexity of the phenomena studied, aims to probe in particular depth the phenomenon under investigation and gain a proper understanding (von Unger, 2014, translated by the authors). Furthermore, in most cases, the research interest is directed precisely at sensitive topics, which are also reconstructed from the subjective perspective of the interviewees (Kretzer, 2013).² The openness of the research process, the high density of information, and the contextual nature of the data generally result in qualitative research materials containing a large amount of interwoven personal information.

¹ If the data is anonymized, it no longer falls within the scope of the GDPR. As long as it is still possible to assign replacements to original information, this is referred to as pseudonymized data. Pseudonymized data is personal data (see point 1.3).

² When collecting personal data, one should always consider whether anonymity can actually be promised to the participants when conducting a study or whether the limits of the procedure should not also be made clear (Kelly 2009). This is an issue that has also been considered, for example, by Stein (2010) – a researcher who herself has had experience with the re-identification of research participants.

If qualitative data is to be reused scientifically, the context (of origin) of the statements that is relevant to social-scientific research and is to be anonymized must be preserved to the greatest possible extent. Certain procedures, such as abstracting information (at different levels of abstraction), are better suited to facilitate secondary analysis than, for example, simply deleting information (Medjedović, 2014).³ The requirements for anonymizing qualitative research materials will vary from study to study (e.g., depending on the sensitivity of research topics) or from dataset to dataset within a study (e.g., if data of varying complexity is obtained through the use of different kinds of interviews such as narrative-biographical interviews versus more structured ones). The anonymization concept must be developed or adapted for each study; a “one size fits all” approach (Stam & Kleiner, 2020) would (unnecessarily) reduce the reuse value and/or possibly increase the risk of re-identification.

Developing a good data-specific protection concept for the reuse of qualitative data requires one to consider questions of procedure and the depth of abstractions as well as the need for additional organizational and technical protective measures. One example of this is the restriction of reuse to specially protected workstations at the research data center (RDC) (onsite use). In addition to technical and organizational measures, another option could be to contact the research participants again and involve them in considerations and decision-making in matters of data protection (von Unger, 2014).

To date, qualitative researchers have found little practical guidance on how to develop an anonymization procedure that is appropriate for their project. Some information on the procedure can be found in texts that deal rather generically with the special protective measures and needs of qualitative data (e.g., Bishop, 2005; Corti, Day, & Backhouse, 2000). Some infrastructural institutions, such as the Finnish Social Science Data Archive (2021), the German Research Data Centre for Education (Meyermann & Porzelt, 2014), or the UK Data Service (2021), also offer short guidance papers or examples of how to anonymize qualitative research data. In addition, there are reports by researchers reflecting on their experiences with the anonymization of qualitative research data (e.g., Richter et al., 2021; Laudel & Bielick, 2019; Saunders et al., 2015a, 2015b; Thomson et al., 2005). Nevertheless, our experience from the consulting practice at the Qualiservice RDC shows that many unanswered questions remain regarding the anonymization of qualitative research materials that lead researchers to seek support.

With this guidance document, we would like to provide answers to these questions and make suggestions for preparing – that is, anonymizing or pseudonymizing – qualitative text-based research materials in a way that protects the personal rights of the participants while enabling the scientific (re-)use of the data. We consider ethical and data protection issues and provide

³ If the intention is to make research data available for secondary scientific use, pseudonymization and anonymization issues should be considered from the outset, not least to plan for the necessary time and personnel resources, in order to comply with data protection and research ethics requirements and to preserve the reuse potential of a dataset to the greatest possible extent.

practical instructions and application examples for the anonymization and/or pseudonymization of qualitative text-based research materials. We show different ways of replacing personal information⁴ and promote a sensitive approach that encourages the greatest possible preservation of the reuse value of a dataset and prevents under- or over-anonymization. To implement our approach efficiently, Qualiservice has developed the QualiAnon anonymization tool for text-based research data (Nicolai et al., 2025) together with other researchers, which can be used free of charge (see point 6). Various application examples can be found in the appendix (see point 9).

1.2 Target group and areas of application

This guidance targets researchers who wish to anonymize or pseudonymize qualitative text-based data and research materials and make them available for archiving and further scientific use (data sharing). The focus is on conceptual considerations that were primarily tested on qualitative interview studies but can also be applied to other text-based research data such as transcripts of focus groups or observation fieldnotes.⁵

1.3 Clarification of terms

The terms “personal data”, “pseudonymization”, and “anonymization” are central to this guide. These are briefly explained below.

Personal data: Article 4 (1) GDPR refers to information as “personal data [...] relating to an identified or identifiable natural person”; this information includes, for example, names, identification numbers or location data as well as specific characteristics that are an expression of “the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (...)” (ibid.).

Special categories of personal data: Art. 9 in the GDPR highlights special categories of personal data that are especially protected: This pertains to the “[p]rocessing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation.” In contrast to the requirements for processing personal data in general, individuals must explicitly consent to the processing of this particular data, that

⁴ When we speak of the “replacement of personal information” in this guide, this refers to both the anonymization and pseudonymization of research data. Anonymization means substituting information such that people cannot be re-identified or could be re-identified only with disproportionate effort. Pseudonymization refers to documenting the original information and its replacements so that it can still be used for the re-identification of persons for research purposes, for instance, for longitudinal studies.

⁵ This guide on anonymizing qualitative research data is part of a collection of other guides and sample templates that Qualiservice makes available free of charge via its website (<https://www.qualiservice.org/en/the-helpdesk/guidelines.html>). These include a guide on the contextualization of qualitative research data (Heuer et al., 2025) as well as a guide along with legally reviewed sample templates for obtaining informed consent for primary research, archiving, and secondary use (Kretzer et al., 2020, templates in English: DOI: [10.26092/elib/3789](https://doi.org/10.26092/elib/3789) & DOI: [10.26092/elib/3790](https://doi.org/10.26092/elib/3790)).

is, the special categories listed in Art. 9 GDPR must be stated explicitly in a declaration of consent if they are to be processed (reservation of authorization).

Pseudonymization: Art. 4 (5) GDPR defines pseudonymization as “the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.” Pseudonymized data is still personal data and falls within the scope of the GDPR. The extent to which personal data can be retained is generally determined by the agreements made between researchers and the research participants consenting to participate in the research.⁶ In individual cases, a balancing of interests pursuant to Art. 6 (1) lit. f GDPR can also be used to determine the permissibility of processing personal data for research purposes without consent. This requires a corresponding balancing of interests.

Anonymization: Anonymization means the “erasure of any personal references” (RatSWD, 2020, p. 18). The GDPR does not offer its own definition of anonymization, yet it does refer to the concept of anonymization (cf. *ibid.*). Recital 26 of the GDPR makes it clear that the GDPR does not apply to anonymized data. Accordingly, the GDPR also lacks further explanations of anonymization (cf. *ibid.*).

The previous German Federal Data Protection Act distinguished between three variants of anonymization that describe different degrees of anonymization: formal, de facto or factual, and absolute anonymization. The current regulation makes no such distinction but mentions only anonymization in general. Yet in practice, it is still common to distinguish between these three variants. *Formal anonymization* involves changing direct identifiers such as names and addresses. In the context of qualitative data, formal anonymization does not fully erase person related data. *Absolute anonymization*, which intends to make re-identification impossible even when investing large amounts of resources, works by means of paraphrasing, masking, or deleting personally identifiable information and accepts strongly reducing its scientific reuse potential to achieve this (Kretzer, 2013). In *de facto or factual anonymization* (anonymization within the meaning of Section 3 [6] Federal Data Protection Act (BDSG), old version), personal data is altered such that “information concerning personal or material circumstances cannot be attributed to an identified or identifiable natural person or that such attribution would require a disproportionate amount of time, expense, and effort” (RatSWD, 2020, p. 18). Factual anonymization offers the possibility of balancing data protection and reuse potential. It is the most prominent form of anonymizing qualitative data.

The German Data Forum (RatSWD) states in its data protection guidelines: “The data protection legislation contains special requirements for anonymization in research. Art. 89 para. 1

⁶ Qualiservice provides information on and legally verified sample templates for informed consent to research, archiving, and the subsequent use of qualitative research data in its guide on obtaining informed consent for primary research, archiving, and secondary use (Kretzer et al., 2020). <https://www.qualiservice.org/en/the-helpdesk/data-protection.html> (accessed: 18/11/2024).

sentence 4 GDPR stipulates that, if possible, any further processing of data for research purposes must be carried out in such a way as to ensure that data subjects cannot or can no longer be identified. This must be interpreted as a requirement to pseudonymize or anonymize data processed for research purposes, insofar as this is feasible (Golla, 2019: 658 f.). Sec. 27 para. 3 sentence 1 GDPR also contains an anonymization requirement for special categories of personal data” (RatSWD, 2020, p. 18.). The RatSWD considers data as being anonymous if it is de facto anonymized, that is, “if personal references can only be restored with disproportionate effort” (ibid).

The anonymization of data itself constitutes a processing of personal data and therefore requires a legal basis.⁷ Such a legal foundation can be derived, for example, from the permissible instances of data processing laid down in Art. 6 GDPR. In this context, however, Art. 5 para. 1 GDPR is also important, which defines exceptions, for example, for academic research.

2 Weighing research ethics, data protection, and research interest

2.1 Complying with data protection

Data protection is about the protection of personal rights and, in particular, about protections that derive from the right to informational self-determination – that is, the right of each and every individual to maintain control over their own personal data and its usage. The legal basis for the protection of this data is the GDPR. It was put into force by the EU on May 25, 2018, and is a directly binding regulation in all member states of the European Union, that is to say, it does not have to be transposed into national law first. National regulations such as the German Federal Data Protection Act (BDSG new), which was reformulated in 2018, and any sub-national regulations such as the data protection laws of the German federal states only supplement the GDPR to the extent that the latter contains corresponding opening clauses.

The GDPR is the binding legal basis for researchers from EU countries. In accordance with it, any processing of personal data requires a legal basis – regardless of whether it involves the collection, storage, analysis, or processing of personal data for subsequent scientific use. The lawfulness of processing is regulated in Art. 6 GDPR. It states that personal data may be processed if, for example, the person has consented to the processing.

The processing of personal data for academic research is particularly privileged in the GDPR. This can be seen in several places (RatSWD, 2020, p. 8), such as in the exemptions from the

⁷ MLS Legal (2020). *Rechtsfragen zur Nutzung von Daten – Forschung, Archivierung, Sekundärnutzung*. <https://www.youtube.com/watch?v=vC8ucG9MJSw&t=1s> (accessed: 11/03/2026).

principles of purpose limitation and earmarking of processing in Art. 5 or the possibility of processing special categories of personal data for research purposes (ibid.).⁸

The GDPR determines the time of anonymization by reference to the purpose for which the information is processed (Art. 5 GDPR). As long as the processing of personal data is necessary to achieve the purpose of processing (e.g., the research objective), it does not have to be anonymized but must be specially protected – for instance, by means of pseudonymization and/or encryption.

2.2 Considering research ethics

Research ethics encompasses the ethical principles and rules that determine more or less bindingly and consensually “how the relationships between the researchers on the one hand and the people involved in social science studies on the other are to be shaped” (Hopf, 2004, pp. 589-590; translated into English by the authors).⁹ Taking research ethics into account involves reflection on and negotiation of the values inherent in research as a social activity. “A research ethics perspective consists of ‘critically reflecting on the extent to which certain ethical principles apply to research activities and can be realized in practice’” (von Unger et al., 2014: 2; translated into English).

Various professional societies have adopted codes of ethics that guide their actions and those of their members,¹⁰ and research funding bodies such as the German Research Foundation (DFG) also address ethical aspects of research activities in guidelines such as those on good scientific practice (DFG, 2022). Preserving the anonymity of those involved in research can be found in these codes as a principle of research ethics alongside others.

Research ethics can simply comply with these data protection regulations for research activities. However, it can also go beyond this, as the following example of making qualitative data available for reuse via Qualiservice shows: In the case in question, the researchers had decided to archive their interview data and make it available for reuse. The research participants had consented to the archiving and further processing for this purpose. This would have been enough to satisfy data protection obligations. However, the research team decided to validate once again the participants’ consent to their personal data being transferred for archiving and

⁸ The RatSWD continues: “Academic research is a legitimate interest in itself and must be weighed against the interests of the data subjects. Considerations on the part of the data subject that must be taken into account include the sensitivity of the data [...], whether the data originate from publicly accessible information or from other sources, and whether the data subject anticipates that the data would be processed (further). The relationship between the researcher processing the data and the data subject must also be considered. For the research interests, it is relevant to consider how important the processing of specific personal data is for the realisation of a research project” (RatSWD, 2020; emphasis in the original deleted).

⁹ The working paper *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften* (RatSWD, 2017a), published by the German Data Forum (RatSWD), provides orientation in matters of research ethics that are relevant to social science studies. The work of Hella von Unger (2014, 2020) also offers helpful tips and information that provides an introduction to the topic. In addition, the RatSWD offers a best practice collection on this issue: <https://www.konsortswd.de/en/topics/best-practices-research-ethics/> (accessed: 11/03/2026)

¹⁰ For instance, the German Sociological Association (DGS) (2017) or the German Association for Educational Science (GERA) (2005, 2016).

subsequent use by presenting to the participants the respective interview transcript, which had been prepared for this purpose, to confirm their decision before the data was finally transferred to Qualiservice. Furthermore, the research team decided to protect the non-anonymized interview transcripts through further technical and organizational measures for reasons of research ethics and make them accessible exclusively in the Qualiservice's guest researchers' room on site even though, from a purely data-protection point of view, non-anonymized further use as a scientific use file (SUF) for external download would have been acceptable in this case.

Dealing with aspects of research ethics is always a study- and dataset-related process. The consideration of possible protective measures concerns the methodological design of a study as well as its implementation, publication, and preparation for subsequent use. In a different context, the decision to restrict subsequent access to the highly protected and thus more strictly regulated conditions for accessing data (possibly only for individual interviews of a dataset) in the guest researcher's room on site to protect the participants could be seen as problematic or paternalistic – for instance, in an oral-history study based on eyewitness interviews (Thomson et al., 2005) or in participatory research (von Unger, 2014) where research participants may in fact want to be named.

2.3 Research in the face of conflicting demands between data protection, research ethics, research interest, and good scientific practice

Data protection and research ethics provide a framework for scientific research. If data is to be reused for scientific purposes, data protection and research ethics requirements must be met. At the same time, scientific (re-)use must be made possible within the framework thus defined. For this reason, the GDPR contains special privileges for research in handling personal data (see point 2.1). The balance between data protection, research ethics, and research interest should be weighed on a study-specific and, if necessary, case-specific basis. The balancing process will often extend over the entire research process and also include possible subsequent use scenarios. With regard to the anonymization process, it is therefore important to develop a secure yet flexible procedure that preserves information relevant to social science research without violating data protection and ethical principles. The solutions that the Qualiservice RDC has developed for this are described in the following.

3 The Qualiservice' anonymization concept

3.1 Informed consent as the basis for data processing

As a rule, the informed consent of the research participants forms the basis for data processing at Qualiservice (cf. footnote 11). The declaration of consent regulates whether and, if so, in what form research data/materials may be used and passed on to third parties – for instance, whether personal information may be retained in the materials to be archived, whether they must be anonymized, or even deleted.¹¹

3.2 Purpose of anonymization

Apart from meeting data-protection and research-ethical requirements, the necessary degree of abstraction also depends on what the anonymized materials will be used for. For example, publishing research results in scientific journals places different demands on the procedure than preparing data for scientific reuse or academic teaching: For publication in scientific journals, interview excerpts are usually selected from a dataset to illustrate the results of analysis and make them comprehensible. This limited use of excerpts only considerably reduces the risk of re-identification. By contrast, the risk of re-identification is greater and anonymization becomes much more complex when an entire dataset is reused scientifically, as the complete interview texts may contain a plethora of contextual references.

3.3 Anonymization as part of the research process and a cooperative task

To ensure that the research materials to be archived are prepared to the best possible extent for secondary use, Qualiservice cooperates closely with the data providers in preparing the materials accordingly (Mozygemba & Kretzer, 2022), advises and provides guidelines, conducts workshops and webinars, and has made the QualiAnon anonymization tool available. However, anonymization or pseudonymization is carried out by the researchers of the respective project. They have detailed knowledge of the data and know where the materials contain sensitive information. Once the pseudonymized or anonymized data is submitted for archiving, Qualiservice checks it and, if necessary, takes further organizational and technical measures to protect it or to complete anonymization.

In general, it is advisable for researchers to involve the RDC where the data is to be archived from the outset when preparing research materials for subsequent scientific use. Qualiservice offers comprehensive advice and support during the research process and on the requirements that are particularly important for archiving, such as informed consent and anonymization. This enables researchers to prepare the data step by step for archiving while they are

¹¹ Information on the templates for informed consent can be found on the Qualiservice website (www.qualiservice.org) or in the Qualiservice guide on informed consent (Kretzer et al., 2020). The consent of the research participants should be documented, ideally in writing. However, other forms of consent are also conceivable. The Qualiservice working paper by Huber, E. & Imeri, S. (2021), for example, addresses alternative options to the written consent.

conducting the research process. This approach also prevents an accumulation of tasks at or after the end of the project, when employees are often no longer working on it or publications are being finalized. Conversely, the integration of data preparation into the research process often also benefits the quality and transparency of the research documentation and thus the research itself (cf. Fielding, 2004). In addition, the RDC can provide support in the application process when it comes to the human and financial resources required for the preparation of the data (cf. point 6).

Qualiservice recommends a step-by-step procedure for anonymization, which initially involves the pseudonymization or anonymization of direct identifiers only, such as names and addresses (formal anonymization), and postpones the anonymization of further information to a later point in time when it better suits the research purpose. This enables the research to include the original information in the analysis. A gradual replacement process also helps to reassess the necessity of replacements: Experience has shown that there is an initial tendency towards over-anonymization, which subsides when researchers are better able to assess where there are risks of re-identification in the data during the course of the project or the anonymization process. The standardized replacement of usually predefined information, such as the general replacement of all place names in an interview, then gives way, for example, to a more nuanced approach that considers whether or not location information contains any personal reference in the respective context. For example, Berlin might be replaced as an interviewee's place of birth but could be retained as the destination of a vacation trip (see point 5.3).

3.4 The strategy of “Flexible Anonymization”

To maintain maximum reuse value, Qualiservice recommends replacing sensitive information with information relevant to social science research (Kretzer, 2013). For example, if the profession of a paramedic is to be replaced, “an emergency medical service worker” could be used instead, and further information, such as that the job involves outpatient or inpatient care, might be added to provide helpful information for reuse. The replacement of information can also be made with different levels of abstraction: For example, a “heart attack” could become an “acute coronary syndrome” or, at a higher level of abstraction, a “cardiovascular disease” or, even more abstractly, an unspecified “disease” (cf. 5.7).

Replacements that abstract from the original information to varying degrees – possibly even retaining the original information (pseudonymization) – form the basis of what Kretzer (2013) was the first to describe as “flexible anonymization.” Flexible anonymization enables the creation of scientific use files with varying degrees of anonymization for subsequent use by expanding and collapsing information. Certain information can be revealed while other information is concealed depending on the research purpose.

Kretzer's description (2013) refers to the use of pseudonymized data where original information is available. However, the approach of applying different levels of information can also be applied to data that has been anonymized – that is, data for which the original information

is no longer available. If, as in the example of a “heart attack”, Qualiservice is provided alternative substitutions that abstract from the original information to varying degrees, these alternatives can be used to create different versions of an anonymized dataset to increase the reuse value for specific research questions. Providing varying degrees of abstraction – with or without reference to the original information (see step 6 on page 18) – thus serves as the basis for flexible anonymization.

3.5 Deletion of personal information

Qualiservice recommends refraining from deleting as well as from blackening and changing information (e.g., Berlin becomes Hamburg or beautician becomes hairdresser). If data has to be deleted because a piece of information could possibly identify a person, the deletion should be clearly indicated in the text. For example, in the case of single words, letters or special characters can be used to provide information about the number of words that have been deleted. At the very least, this provides information about the extent of the deletion, that is, whether only individual words or entire paragraphs are missing.

3.6 List usage and partial standardization

While quantitative social research often relies on standardized replacements for anonymization, qualitative social research generally uses study-specific anonymization concepts to replace personal information. If these replacement concepts are shared, researchers in a subject area can further build on them and thus gradually develop differentiated ones that retain social-scientifically relevant information at different levels of abstraction for specific subject areas. The use of standardized substitutions of categories such as places, names, professions, dates, and so on is likely to lead to less differentiated substitution schemes that abstract information to a rather high degree. The use of standardized classification systems can nevertheless be beneficial in certain subject areas. In addition to, for example, time divisions into months, years, annual periods, days, and so forth, standardized lists such as the International Classification of Occupations (ISCO), which is used in quantitative research to code occupations (e.g., Züll, 2015), can also be used for this purpose. If standardized classifications such as the ISCO are organized hierarchically, they can also help to map different levels of abstraction (cf. point 5.4). Qualiservice has had positive experiences with this approach.¹² Whether the use of a list has advantage for preserving information relevant to social science research when abstracting personal information must be decided on a study-by-study basis.

Four such classifications that can be used to replace personal information are presented below. These classification schemes are freely accessible and regularly updated. They include the aforementioned ISCO (cf. 5.4), the International Classification of Diseases (ICD) (see point 5.7), the Classification of Economic Activities (see point 5.6), and the lists of municipalities

¹² Researchers can also import classifications into the QualiAnon anonymization tool and export, reuse, further develop, and share their (topic-, discipline-specific) replacement categories and thus also publish them as project output (see point 6).

provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR) (cf. 5.3). The BBSR's list even offers the option of linking other lists via a code number and thus using further location-specific information such as details on infrastructure or unemployment in the region for replacements. With the exception of the BBSR's list of municipalities, these lists are structured hierarchically. How the information from the respective list is included in the replacement – for example, whether the associated list code for an economic sector in which a company operates is documented only once so as not to impair the readability of the transcript or whether it is included in every replacement – must be decided at the beginning of the task and should then be handled consistently across the entire study or dataset.

3.7 Further components of Qualiservice's data protection concept

Before we describe the individual steps that Qualiservice takes when replacing personal data, we would like to point out that anonymization is only one element of Qualiservice's data protection concept alongside other technical and organizational protective measures (see, e.g., Heuer, Hollstein, & Mozygemba, 2021). Examples of further measures include the separate storage of identification data and research materials, the restriction of subsequent use to scientific purposes, controlled access to research data, and contractual agreements for the use of research data. The latter prohibit, among other things, re-identification attempts or the publication of entire interviews.¹³ Researchers can also take further individual protective measures such as time embargoes or set conditions for citation.

¹³ Qualiservice's terms of use can be found under the link https://www.qualiservice.org/files/contao-theme/public/documents/downloads/Qualiservice_Vereinbarung_Datennutzung_EN-0821_v2.pdf or can be provided upon request at info@qualiservice.org.

4 Replace personal information step by step

Figure 1 shows an overview of the steps required to prepare, carry out, and document the replacement of personal information so that the procedure can be easily understood by potential reusers and the material still offers sufficient potential for reuse.

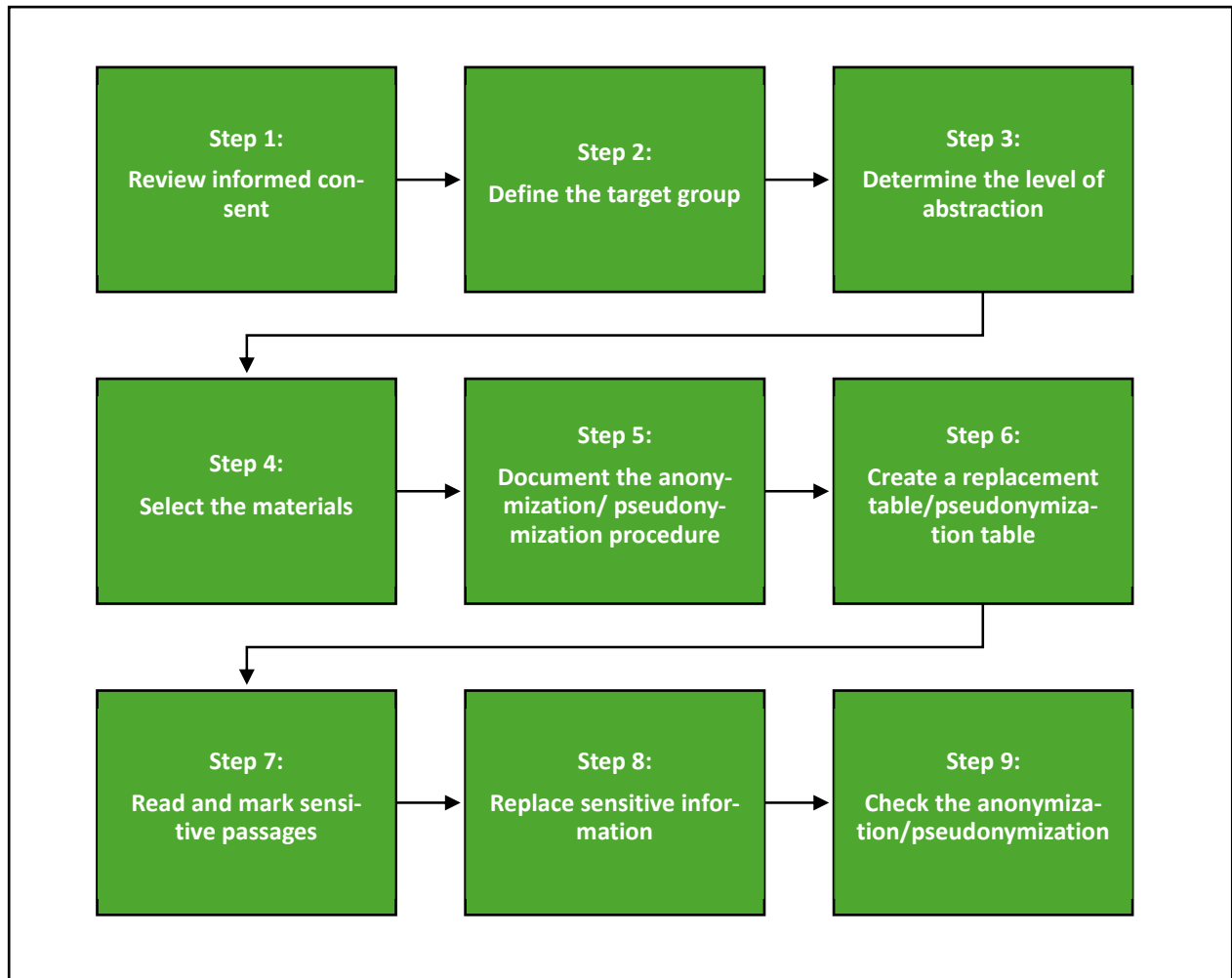


Figure 1: Overview of the procedure

Step 1: Review the Informed Consent

In the informed consent research participants explain how their personal data is to be handled in the research process (cf. point 3.1). The Informed Consent usually is the base for decisions regarding pseudonymization and anonymization as part of data preparation.

Step 2: Define the target group for whom the data is to be anonymized

The question of for whom the data is being anonymized – that is, who will potentially have access to the anonymized data – is important for choosing the level of abstraction of the replacements. Good friends, neighbors, work colleagues, or researchers usually have different

contextual knowledge about the individuals being researched that could be used for re-identification. If parts of an interview are to be published in a book, for example, which can be read by anyone, the level of abstraction must be far higher than if – as in the case of Qualiservice – only researchers with a legitimate research interest and after signing a user agreement can access the data.

Step 3: Determine the level of abstraction (envisaged anonymization level)

Deciding on a certain level of abstraction requires weighing data protection, research ethics, and scientific reuse potential against each other (cf. point 2.3). The decision for a specific level of abstraction can be made at the study level but also at the level of individual data objects, such as individual interviews. The decision is influenced by the target group of subsequent use, study specifics (such as a particular sensitivity of the topic or the presentation), as well as technical and organizational measures (such as highly secure access via the guest researcher's workstation [onsite use]). The decision for a certain level of abstraction must be made on a study-specific and, in some instances, case-specific basis.¹⁴

Step 4: Select the materials to be anonymized or pseudonymized

In addition to the research data in the narrower sense, it may also be necessary to include other materials such as postscripts, socio-demographic data, and – if enclosed to contextualize the dataset – project reports, project publications, and so forth in the planning of anonymization or pseudonymization. These materials may also contain personal information, such as the name of the interviewee or interviewer in the header of the postscript or in published information; for example, such reports often name the locations where the research took place. This information must be taken into account when planning anonymization.

Step 5: Document the anonymization or pseudonymization procedure

The anonymization or pseudonymization procedure must be documented so that secondary users can understand how the research material was changed for subsequent use. At the same time, this documentation supports maintaining a consistent procedure if work is carried out over a longer period of time or if several people process the data. The documentation should contain information on the following points:

- Why was the data anonymized or pseudonymized?
- Will the research materials be submitted in anonymized or pseudonymized form?
- Will the pseudonymization list (see step 6) remain with the researcher or also be transferred to the RDC?
- How was abstraction applied? For example, were different levels of abstraction used?

¹⁴ As in the case of the project described by Laudel & Bielick (2019), this assessment can also lead to the conclusion that data should not be anonymized if its reuse value is to be retained. In such a case, given that the respective research subjects have consented to the disclosure of personal information, one could consider whether certain technical and organizational precautions, such as confining access to the onsite guest researcher workstation, could enable subsequent use. It may also be appropriate to weigh interests in accordance with Art. 6 (1) GDPR. This must be decided on a case-by-case basis.

- Was the anonymization or pseudonymization carried out before the data analysis or was it done step by step? In other words, were the direct identifiers removed first (basic anonymization) and indirect identifiers only abstracted in a later step (e.g., after the analysis was completed)?
- Were the research participants in the field asked to omit names, places, and other possibly identifying information?
- Was an anonymization tool used for anonymization or was this done manually? If the former, which tool was used?
- How are substitutions indicated in the material? (If special characters such as square brackets, pipes, or color highlighting were used for this purpose, care should be taken to ensure that these are not also part of the schema used in interview transcription.)
- Did you work with line or paragraph numbers? (Replacements may change the line numbers of the original material but provide a more precise means of orientation in the text; paragraph numbers will not change.)
- Were replacements made on a case-specific or study-specific basis? For instance, was the city of Berlin only replaced in interview 1 with “place 1, large city in Germany” (i.e., case-specific), or was Berlin masked in this way across all interviews/materials of the study (i.e., study-specific)?
- What information will be replaced (replacement categories such as places, diseases, etc. [cf. point 5.10])? What social-scientifically relevant information should be retained when replacing information?
- Were (controlled) lists (see point 5) used for the replacements? If so, which lists (for controlled lists, the exact citation including the version date of the respective list must be documented, as these lists are regularly updated by the responsible bodies)?
- Are replacements within the same category assigned numbers (e.g., are city names listed as place 1, 2, etc.)? (Assigning numbers has advantages for the preservation of the reuse potential. The replacement of “Berlin” with “place 1” mentioned above is a case in point. If the assigned number were omitted, it would not be possible to determine whether the text was referring to the same or a different location. Assigning numerals, by contrast, allows secondary users to distinguish “place 1, large city in Germany” [Berlin] from “place 2” [as another large city in Germany, e.g., Hamburg].) It is important to document the system of assigning numbers (or other symbols, such as letters) in the anonymization concept and to remain consistent in applying that system.
- Was information deleted? If so, how are deletions indicated (cf. point 3.5)?¹⁵
- Did several people carry out the anonymization or pseudonymization? If so, how was consistency ensured in the joint procedure (e.g., was there an exemplary test run on an interview)?

¹⁵ To preserve the potential for reuse, Qualiservice recommends working with abstractions at different levels and with socially relevant information when replacing data, and deleting data only in particularly sensitive contexts.

Step 6: Create a replacement table/pseudonymization list

The substitution of personal information should be documented in a replacement table (cf. Table 1). This table ought to be stored separately from the data in a secure location during the research project. The header of a replacement list should contain the interview number, the file name, the line or paragraph number of the original information, the replacement categories, as well as the replacements at different levels of abstraction and, if applicable, the original text along with space for comments as column headings. If pseudonymized data is used, the original information should also be retained and documented in the replacement table. The column containing the original information is then removed to anonymize the data. However, substitutions with varying degrees of abstraction are retained along with other information relevant to the social sciences and can be used for flexible anonymization (cf. 3.4).

Step 7: Read for meaning and mark sensitive passages

The next step involves carefully reading the materials to be processed and marking information that could potentially allow conclusions to be drawn about individuals. The material must be scanned for any information that could identify third parties (interviewers, people discussed in the interview).¹⁶

Consultations with researchers and the curation of research materials at the RDC have shown that researchers are inclined to choose very strong abstractions or even delete information, especially early in the anonymization process, out of concern that they might not meet data protection requirements. More information is then replaced than data protection regulations or research ethical considerations would require. Over-anonymization potentially compromises the reuse value of the materials. To avoid it, it is advisable to first mark information that is considered sensitive in several interviews before starting the replacement process. Marking helps to get a feeling for the sensitive passages and facilitates selecting the information that might enable re-identification; if necessary, specific questionable passages can then be discussed in the research team. Marking selected text examples supports joint decision-making in the event of uncertainty regarding the sensitivity of the information and helps to ensure a consistent approach.

¹⁶ Even if research participants have consented to the processing and disclosure of their personal data and it therefore does not have to be anonymized, personal information pertaining to third parties must still be anonymized.

Table 1: Example of a replacement list with different levels of abstraction

Interview no.	Interview title						
1	Interview-ID_1.0						
Line no.	Replacement category	Original text	Replacement, level 1	Replacement, level 2	Replacement, level 3	Replacement, level 4	List usage
5	Name	Elisabeth Gerken	Person 1	Person 1, name of interviewee 1	Person 1, name of interviewee 1, old German first name		
54	Profession	Geriatric nurse	Occupation in the service and sales sector, ISCO08-5	Occupation in the field of care professions ISCO08-53	Occupation in the field of health care professions ISCO08-532	Nursing assistant ISCO-08 5321	ISCO-08
56	Location	Bremen	Location 1, large city	Location 1, large city in Northern Germany	Location 1, large city in Northern Germany with a harbor		BBSR
62	Company	Company Schmidt	Company B, company in the human health and social work activities – WZ Q	Company B, company in social work activities without accommodation – WZ Q88	Company B, business providing social work activities for the elderly and disabled – WZ Q88.1		Economic sectors

Step 8: Replace sensitive information: pseudonymization or anonymization

Qualiservice recommends a step-by-step procedure in which direct identifiers such as names or addresses of individuals are replaced first in a basic anonymization procedure (formal anonymization) or the data is pseudonymized. The anonymization of further information is then carried out in the course of the analysis or after its completion (as soon as the purpose of the research permits). The question of whether certain information can be linked to a person should be a constant consideration when working with the material – especially if information has already been abstracted in the text: For example, information about an interviewee's parents may already lose its personal identifiability when masking the interviewee him- or herself. Across-the-board replacements of certain categories (excluding direct identifiers such as names and addresses) can also lead to over-anonymization.¹⁷

The selected level of abstraction (see step 3) should replace personal information such that as much information relevant to scientific reuse is retained as possible while preventing the re-identification of a person. To maintain the highest reuse value possible, it is advisable to make replacements at different levels of abstraction (e.g., place, large city, large city in Northern Germany, large city in Schleswig-Holstein), so that, for example, the QualiAnon anonymization tool can be used to easily create differently abstracted versions of a text (cf. point 3.4 & 6). Replacements should be based on what is necessary to protect the personal rights of those involved. In addition to pseudonymization or anonymization, technical and organizational measures should be taken into consideration to counteract the risk of re-identification.

The question is often raised of whether possible subject-specific replacement categories already exist that can be used for a specific research topic, what is often not the case. However, using specific lists/ classifications can help here (see point 3.6). Re-using substitution schemes already developed by researchers in similar thematic contexts and develop them further could facilitate the anonymization process. Once a replacement scheme has been created, QualiAnon enables its publication, further use, and differentiation (cf. point 6).

If substitutions are made, these should be indicated in the interview, for instance, by underlining, special characters, or otherwise applying symbols that are not part of the transcription schema. To give an example, this could look as follows: [~ text text text ~]. The substitutions are documented in a replacement table (cf. Table 1). It may make sense to proceed in a step-wise manner, for instance, because information is still required for the analysis or because a final decision has not yet been made as to whether information needs to be anonymized or not. The replacements are then carried out step by step.

Substitutions often diminish the readability of a transcript. To improve readability, the replacements could be adapted grammatically. However, it is important to carefully consider to what extent the original text should be edited for readability.

¹⁷ The general marking of certain information, on the other hand, can facilitate making the decision on the need for replacement (which may also have to be made later) and systematize the procedure.

The “search and replace” function of word-processing software (or comparable functions) can help to find words to be replaced, but should be used with caution, as spelling mistakes, broken words, and dialect terms for the word being searched may not be found in the transcript and therefore might remain in the original. Moreover, automated substitutions may lead to changes in words that contain the same sequence of letters: Let’s say I am looking for Michael and want to replace this with “Person 1, friend of the interviewee.” Using the “replace all” function, two things might go wrong here: First, I might miss the use of the short form “Micha” in the text, thus leaving it unaltered. Second, I might unintentionally change Michaela, the name of the interviewee’s wife, to “Person 1, friend of the interviewee a”, where the wife becomes the interviewee and the “a” might be interpreted as a typo.

Step 9: Check the anonymization/pseudonymization

Finally, the anonymized/pseudonymized research material must be checked again. This check targets information that might still be contained in the text and could possibly re-identify individuals. In the process, the replacements should also be checked since word remnants or entire markings may have been left in, a replacement might turn out to be unnecessary, or additional information might be missing. Once the researchers have completed this check, they can submit the materials to Qualiservice. Qualiservice checks the results of the anonymization or pseudonymization and supplements them if necessary.¹⁸

5 Key categories for the replacement of personal information

Below, we present an overview of the main replacement categories. It is advisable to check each substitution to see whether it is necessary to reduce the risk of re-identification of a person. The key is to find a procedure that does justice to the respective study and the available research materials and thus suits the topic and data well. The aforementioned step-by-step procedure (see step 8) can be helpful here. It is not advisable to simply routinely replace the categories mentioned. Whether a replacement is necessary should be checked for each individual case: Often the replacement of one or a few categories already changes the informational content in an interview in such a way that re-identification becomes less likely. If, for example, teachers in a large city are interviewed, this group will be much larger than, say, the population of prison guards in the same city. Depending on the contextual information available in the interview, in the first case the professional activity and the name of the city in which the

¹⁸ It is important to note that not all RDCs check the anonymization or pseudonymization of the transmitted data like Qualiservice does. Some RDCs leave this to the data providers. If you are interested in data sharing, you should therefore inform yourself about the procedure at the selected RDC and any potential liability that might be associated with anonymization.

research was conducted could possibly be retained or an abstraction at the regional level might be sufficient to eliminate information that could identify individuals from the text.¹⁹

The categories that often play a role in substitution include names, dates, locations, educational and occupational information, names of organizations and institutions, individual characteristics, and positions and roles (cf. Table 2). Replacements can be made and differentiated along these (and other) categories. In the following, we will discuss the replacement categories compiled here, drawing on examples and experiences from Qualiservice's guidance services. The replacement categories presented here form the basis for a category scheme that can be further developed at the category level and on a topic-specific basis and can be fed into and edited in Qualiservice's QualiAnon anonymization tool (cf. point 6).

¹⁹ The Finnish Social Science Data Archive (FSD) also presents an instructive example of this: "We were never really very religious even though my aunt was one of the first women to be ordained as priest in Finland." The authors discuss whether the information that the interviewee's aunt was one of the first female pastors in Finland could allow conclusions to be drawn about the interviewee. However, knowing that 94 women were ordained as pastors at the first ordination puts any concerns in this respect into perspective (<https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymization-and-identifiers/>; accessed: 09/10/2023).

Table 2: Overview of key categories for the replacement of personal information (summarized from Qualiservice's consulting and curation practice)

	Replacement category	Possible subcategories	Original information	Possible replacements without differentiated degrees of abstraction
1	Characteristics of the person	Name, characteristics (e.g., function for the respondent), number of children, gender	Jutta	Name of wife
2	Time information	Date, time period, month, respondent's age	Aug 2002–Oct 2005	Summer 2002 to fall 2005
3	Location information	Country, region, federal state, association of municipalities, city, district, street names, bodies of water, mountains, rural areas, place of residence, place of work, other	City of Bremen	City A, large city in Northern Germany
4	Education	Educational qualification, field of qualification, type of school, further education, type of training	Geriatric nursing training	Training as health associate professionals (ISCO08 32)
5	Professional and comparable activities	Functions, profession, activity, voluntary commitments, positions and roles, other	Paramedic	Ambulance worker (ISCO08 3258)
6	Organizations and institutions	Companies, schools, training centers, authorities/administration, economic sectors, others	Horch interview service	Enterprise A – other professional, scientific, and technical activities (WZ M74)
7	Distinct individual features	Illnesses, awards, sports activities, accidents, other	Heart attack	Disease of the circulatory system (ICD11 11)
8	Other	Languages, nationality, religion	Italian	Southern European language, the national language, his first language, her mother tongue

5.1 Replacement of names

Names identify people relatively clearly. They are direct identifiers.²⁰ If individuals have not consented to the use and disclosure of their personal data,²¹ names (e.g., of research participants, researchers, or other so-called third parties) must be anonymized. Various options are available for this (cf. Table 3).

If protecting a person requires a high level of abstraction, a replacement category such as “name of a person” or “person” should be selected. However, more information can be retained by simply adding a number (“name of person 1” or “person 1”). The numbering serves to identify the replaced person 1 as one and the same person throughout the text or the entire study. Further added value is achieved if information is retained, for instance, about the role of the name bearer in the interview (e.g., “name of the interviewee”) or the type of relationship between the name bearers (e.g., “name of the wife of person 1”, “mother of the interviewee”). Further information such as the gender of the name bearer or the origin of a replaced name (e.g., German family name, Arabic family name, boy’s name, girl’s name) offer additional possibilities to increase the reuse value.

Another option is to replace names with other (first) names – so-called pseudonyms.²² Saunders, Kitzinger, and Kitzinger (2015a), for example, describe their experiences with this procedure. The choice of pseudonyms from a similar cultural and linguistic background and/or the choice of pseudonyms with a similar meaning can provide additional information. However, it should be borne in mind that the choice of name may be linked to cultural and socio-structural references, for instance, to a person’s social background or age (Gerhards, 2005, 2010).²³ This potential information could be lost if the name is replaced with another name or, in the worst case, changed in a way that lays false trails in the material. Against this background, the choice of a different name must be carefully considered and should be documented (cf. point 4 step 5). If you decide on this form of replacement, name lists can help. Examples of officially published lists can be found at the *Gesellschaft für Deutsche Sprache* (Society for the German Language), which has been publishing annual lists of the most common first names in Germany since 1977, or, with regard to German surnames, information from *Deutscher Familienamenatlas* (the German Surname Atlas) can be helpful, which also provides information on the geographical distribution of names.

²⁰ An overview of direct identifiers is presented in Watteler & Ebel (2019).

²¹ In the context of open science, eyewitness interviews, or participatory research, it is often the case that names and other information should be preserved – also at the request of the participants.

²² The term pseudonym is also used in this context if the original name is not documented.

²³ Gerhards describes how – in addition to a couple’s individual preferences – their social status and cultural milieu influence their choice of a first name for their child and thus the identity of the child bearing that name while also ensuring distinction from other cultural milieus (Gerhards, 2005, 2010). This becomes clear in the description of a couple’s selection process: “The repertoire of potential first names was thus confined to those acceptable to the parents’ social milieu. Names from films, television, and sports were judged to be vulgar, since they were often used by the lower classes. Old German names like Wilhelm, Uta or Otto were rejected on grounds of being too traditional and otherwise conservative. Jewish names like Sarah and Daniel would have been considered if they hadn’t been used a decade earlier by the same milieu; they were shopworn and did not satisfy the parents’ need to christen their child with something distinctive and individual” (Gerhards, 2005, 1).

Table 3: Examples of the replacement of names

Original	Possible replacements	Notes
Gerda	Helga, interviewee 1 (old High German first name), name of interviewee, person 1	
My cousin Berta	Name of the interviewee's cousin, name of the interviewee's maternal cousin	
Tim Schulze	Interviewer 1 (male)	Information about interviewers should also be documented, as this can provide important information about the interview process or the conduct of the interview. For example, it makes a difference whether a student employed at a company or a researcher from outside that company interviews the staff council or whether a midwife or sociologist conducting research asks a pregnant woman questions about pregnancy care.

If statements refer to persons in the public eye and their duties – for instance, the Federal Chancellor – or are related to the public interest, the names can be retained provided that the person is speaking as a public official and there is no risk of re-identification of other third parties who are not in the public eye. However, if the statements contain information that has nothing to do with the office held but concerns the person's private sphere, for example, these must also be changed.

5.2 Replacement of time specifications

Temporal information can also identify a person. This substitution category includes dates, time spans, information on the course of time, and the beginning or end of certain phases of life. When abstracting temporal data, it is important to carefully consider whether the information to be replaced can help to re-identify a person. A person's date of birth generally can identify that individual. This can be different for information on the duration of an employment relationship, where the interview context will determine how sensitive the information is – especially if other information such as the place of employment or information on the company has already been abstracted. Table 4 shows examples of substitutions.

Table 4: Examples of the replacement of temporal specifications

	Original text	Replacement options	Notes
Date specification	01/04/2020	In spring 2020, in April 2020, 2020, [specific date]	The exact periods of spring, summer, fall, and winter should be documented; e.g., March 21 to June 20 could become spring.
Month	01/04/2020	Early April 2020, in April, month xy	The time spans constituting the beginning, middle, and end of a month, year, decade, century, etc. should be documented. For months, e.g., you could choose the beginning of the month to be the 1st–10th day of a month; the middle of the month might be the 11th–20th day and the end of the month the 21st–31st day.
Year	2002	In the early 2000s, at the start of the 21st century	
Age information	I was 8 years old.	In my childhood, before I was 10 years old, [exact age]	This requires documenting how you define the age intervals. For instance, 40–43 = early 40s, 44–46 = mid-40s, 47–49 = late 40s. However, the informational content is also important for the choice of abstraction. This results from the interview or study context. For example, “the age of 25” can easily be replaced with “mid-20s.” However, if someone is talking about the age of 65, the information “start of retirement,” e.g., could also be important for understanding the narrative and should possibly be reflected in the abstraction. More precise information can be added to the date in brackets.
	Before my 18th birthday	When I came of age, as a young adult	
Time span	From mid-April 2004 to the end of December 2004	From spring 2004 to winter 2010; from month A to month J (10 months), for a while in 2004, several months in 2004, several months in the early 2000s, from Easter 2004 to Christmas 2004	When replacing the 12 months with month A – L or 1 – 12, you should be aware that the order of the letters and numbers may suggest a sequence of months. The information about the duration of a period of time can also be lost by replacing it with month A to month B, which is perhaps better added in brackets (10 months, several months, etc.). Abstraction into seasons may help.

5.3 Replacement of location data

The variety of place names is huge. The spectrum ranges from information on countries, federal states, regions, municipalities and associations of municipalities, city names, names of city districts to street names, names of water bodies, names of mountains, areas of land, and so on (an overview of frequently occurring examples can be found in Table 5).

While location information can identify a person, such information is also often found in interviews without any specific reference to a person. For example, a business trip to Berlin or a vacation at the Baltic Sea is not personal information per se. Or in a study on teachers in Hamburg, the city itself can be named under certain circumstances because the risk of re-identification associated with this information is low, as the teacher population is large in this major city. Saunders et al. (2015a) describe it similarly: In their study, naming the location of a clinic would pose a re-identification risk for the interviewees because only a very small number of clinics treat the rare disease under study.

A general replacement of all locations without checking for personally identifiable information would generally limit the reuse potential of a dataset. The context in which research participants embed their accounts is just as decisive for the replacement of location information as the context of the study itself. It is also important to check whether the location of the data collection has already been published elsewhere (e.g., articles, project descriptions).

Replacements of place names can involve varying degrees of abstraction. For example, the function that a place has for the interviewee could be relevant to social science (e.g., place of birth, place of residence, place of work). In addition, details on the size, location, or infrastructure of a place provide important information for subsequent use. For example, if an interviewee says that she works in Bremen and this information is considered sensitive, information such as “a large city in Northern Germany” could provide important clues for subsequent use despite the need to replace the place “Bremen”. At what level (municipality, state, federal state, region) and in what breadth information is preserved must be decided on a case-by-case basis.

Table 5: Examples of locations to be anonymized and options for replacing them

	Original text	Replacement options	Notes
Cities/districts	Bremen	Location, location 1, city A, city 1	You can simply replace location information indicating that a location was replaced. However, the informational content of this replacement is low. It is greater if it is assigned a letter or number. This makes it possible to check whether the substituted locations are the same.
	Bremen	City 1 (large city), city 1 (large city in Northern Germany)	Reference is made here to the BBSR classification. Further information such as a more precise definition of the location or information on the infrastructure can be added.
	Bremen	Large city A, place of residence 1	More information can be retained by mentioning the function of the place – e.g., place of residence, place of work, place for relaxation. Adding a numeral or a letter can make it possible to distinguish between different locations. This could be relevant, e.g., when talking about work-related commuting in an interview. In this case, distances between place of work 1 and place of work 2 could be indicated in brackets.
	Downtown Berlin	District A (or 1) of the large city A, district A (or 1) of city A	Here, too, information on the infrastructure or socio-structural and political characteristics of the district can be provided, e.g., in brackets. Examples include information on unemployment, the population with migration background, or income and education structure. Depending on the research project, information such as the presence of green spaces, transportation connections, etc. can also be interesting.
	Rome	Capital city of a (southern) European country + number	Cities abroad can be abstracted in the same way as the replacements for cities described above. Lists such as those from the EU's statistical office (EUROSTAT) ²⁴ could be helpful here.
	X-city	Unidentifiable place + letter	An interviewee might also mention a non-identifiable place in an interview. Replacing a city that may have remained incomprehensible in the interview with "x-city" could raise questions in subsequent use. Reference to the fact that the place is not identifiable or remained incomprehensible in the interview can be significant for subsequent use.

²⁴ <https://ec.europa.eu/eurostat/web/nuts/overview> (accessed: 17/11/2025)

Street names	Mehr-mannstraße	Street + number	An address mentioned by an interviewee can refer to the function of that place, e.g., the place where he/she lives, works, goes to school, etc. Adding a number or a letter can provide the information that different addresses or the same addresses are involved.
	Mehr-mannstraße	Residential address (with or without adding a numeral/letter)	
Countries	Germany	European country	Whether a numeral or letter is added or not, it could still be beneficial to add information that characterizes the country in more detail.
Waters	Chiemsee	Lake	The size or location of the lake can be further defined. Adding a numeral/letter may also be useful if several lakes are presented in the text.
Region	East Prussia	European region	
Concrete places	Bremen Airport	Transportation hub 1 in large city A	Here, too, further information can be provided which, depending on the sensitivity of the data, can abstract more or less from the place of origin – in this case, the airport.
	Saftschuppen (store name)	Warehouse near the train station	Replacing the place with its function for the interviewee (e.g., place of work, meeting place) can – even without adding a numeral or letter – provide social-scientifically relevant information. This form of replacement can be beneficial when replacing stores, discotheques, fitness studios, etc.

Experience has shown that it is helpful if the replacement of place names is done consistently across an entire study: A place name should be replaced in the same way in all of the interviews in a study. This facilitates maintaining the coherence of the dataset, and the interviews can be related to each other. For example, in a study on experiences of change that was conducted in two places in the 1990s, the experiences of people in place 1 (small rural community, strong religiosity, high unemployment) could be compared with the experiences of people in place 2 (medium-sized small town, university town, good infrastructure), although the places were abstracted because of a possible risk of re-identification of the interviewees.

Assigning numbers to clearly identify places in spite of abstraction allows one to specify an unlimited number of places. The use of letters (A to Z), by contrast, quickly leads to a confusing and often error-prone series of letters (e.g., after using the substitution “city Z” you have to start with “city AA” or similar). On the other hand, the use of letters has the advantage of preventing any assumption of a prioritization of individually numbered places that was not intended by the interviewee. For instance, if different places of work or residence of a person are presented in the text, the numbering might suggest that place of residence 1 is the prioritized place of residence.

For the partially standardized replacement of places in Germany, researchers can draw on the overview of city and municipality types issued by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). This list contains the names of municipalities, the federal state where they are located, the name and number of the association of municipalities of which each municipality is part, information on the area and the number of inhabitants for each municipality, as well as more and less differentiated information on the municipality types.²⁵ The name of a municipality can be searched for in the list and, depending on the chosen procedure, abstracted by replacing it with associated information at a higher level of abstraction such as the federal state or the type of municipality.

The types of municipalities used by the BBSR are categorized into four main types (big city, medium-sized towns, small town, rural community). The legend can be used to provide information about regions. The types of municipalities provide information on area and population and also on how many municipalities and associations of municipalities can be assigned to the respective category.

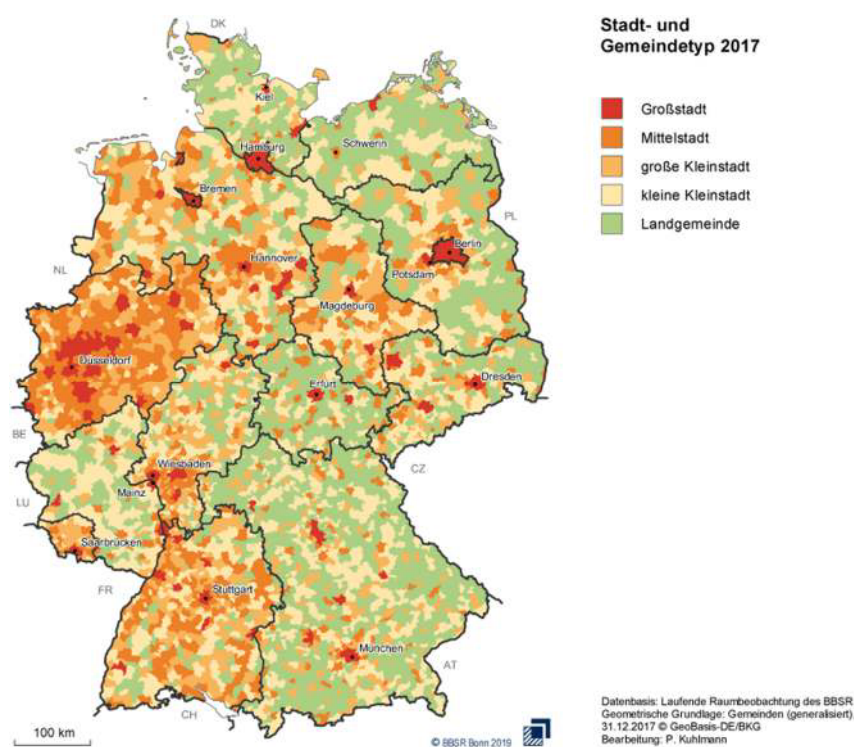


Figure 2: City and municipality types of the BBSR (classified into large cities, medium-sized towns, large small towns, small towns, and rural communities); source: <https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/gemeinden/StadtGemeindetyp/StadtGemeindetyp.html>, accessed: 11/03/2026; no English version available.

²⁵ <https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/gemeinden/StadtGemeindetyp/StadtGemeindetyp.html> (accessed: 17/11/2025)

The municipality code number can be used to combine information on the respective locations drawn from other BBSR lists. This list provides information, for example on the infrastructure of the location and district-related information at the regional level (e.g., regional unemployment, the regional housing market) or on its (aggregated) population and unemployment figures. A replacement could look like this: “Location 1, large city in central Germany, low unemployment.”

Further links to information at the European level can be provided via the NUTS (Nomenclature of Territorial Units for Statistics) regions²⁶ and the European types of regional settlement structures. It is also possible to navigate at the NUTS regional level using the municipality code. Further information on the description of specific places and regions at the European level is available, for example, from the European Commission’s statistics.²⁷ In addition, the United Nations (UN) provides a list that also includes country names.²⁸ If you decide to use lists, these can be fed into the QualiAnon-tool and used to create replacements (see point 6).

5.4 Replacement of information on training and occupation

Information on occupations, professional activities, training, and so on may contain personal data. In this category, too, it is important to weigh replacements carefully to preserve the reuse potential of the material, as abstraction is always associated with a loss of information and reduces the reuse potential of the material. With regard to education and occupation, Thomson et al. (2005), for example, discuss the difficulties associated with the replacement of “nurse” and “physician” with “health professional” in an interdisciplinary research project.²⁸ In some cases, it may also be advisable to leave the job titles untouched because the replacement of other information might be sufficient to prevent the re-identification of persons or because the number of people working in an occupation is sufficiently large that mentioning the occupation will not disclose any personal information (see the example of the priestess in footnote 19) or it is not (see the example of prison guards on page 21). How best to proceed should therefore be considered on a case-by-case basis.

Occupations can be abstracted in different ways. Examples of substitutions can be found in (Table 6). To avoid “false trails,” Qualiservice recommends abstracting occupations and professional activities to varying degrees and not replacing them with other occupations. For example, replacing “physiotherapist” with “masseur” would mix two professions that differ at least in terms of training and entry requirements. It would be better to abstract “physiotherapist” and “masseur” and replace them with a collective category like “personal service profession in the healthcare sector,” which could be supplemented by further information such as “focus on the

²⁶ <https://ec.europa.eu/eurostat/web/regions-and-cities/> (accessed: 11/03/2026)

²⁷ <https://ec.europa.eu/eurostat/web/nuts/overview> (accessed: 11/03/2026)

²⁸ For example, in interdisciplinary discussions between researchers in business and nursing, the business researchers proposed replacing the words ‘nurse’ and ‘physician’ with the broader category of ‘health care professional’ for the sake of preserving the individual professionals’ anonymity. From an organizational analysis perspective, concerned with the professions as a group, this was a possibly acceptable solution. Our nursing colleague objected, however, saying that a participant’s identity as a nurse or physician was crucial to data analysis” (Thomson et al., 2005).

musculoskeletal system” for physiotherapists. Although information would then also be lost (cf. Thomson et al., 2005), no information would be included in the text that is not contained in the original.

Hierarchically structured standardized lists can help in the search for adequate replacements with different levels of abstraction. The International Classification of Occupations (ISCO)²⁹ of the International Labor Organization (ILO) is suitable for occupations. This classification is an established instrument for anonymizing information on occupations and professional activities in quantitative datasets (Ebel, n.d.; Züll, 2015). ISCO distinguishes between major occupational groups, sub-major groups, minor groups, and unit groups in a hierarchically structured classification system, which contains information on the respective occupations and occupational activities in varying degrees of abstraction.³⁰ The classification of qualification levels follows the International Standard Classification of Education (ISCED).³¹

The ISCO classification features a hierarchical structure. This makes it possible to abstract information to varying degrees without laying false trails. The higher the level of abstraction selected, the greater the number of occupations that fall into the respective category, thus reducing the risk of re-identification. Accordingly, depending on the level of abstraction, the occupation of geriatric nurse can become a “service and sales workers occupation” (ISCO 08 code 5), “care occupation” (ISCO 08 code 53), or a “personal care workers in health services” (ISCO 08 code 532). Additional information such as the target group being cared for or the care sector (outpatient vs. inpatient) can be added as further important information for subsequent use. In the example, the replacement could look as follows:

“Ms. A.: (1) Well, I, (1) I actually started out differently professionally. I was [name of occupation, group: building and housekeeping supervisors ISCO: 515 000], then got married, had two children, stopped working, was at home for about 10 years, and then did an internship in [a health care facility for older people]³² via a [training course in health care] and was then immediately asked if I could stand in as a [name of occupation, personal care workers in health services (ISCO 532 000)] (2). That’s what I did, and then I just stuck with it. The one time I stepped in turned into [several, >5] years of work, and then I thought, well, if I’m already doing this work, then I’ll do the training too. And then I did the [training as a health associate professional (ISCO 3200)].”

Information on training can be personal. The complexity and diversity and therefore also the lack of comparability of training and education systems presents a particular challenge. If a high degree of abstraction is sufficient, UNESCO’s ISCED can be useful. ISCED aims to ensure the comparability of educational institutions at the European level via the “skill levels” defined in ISCO-08. This classification system contains ten qualification levels, which range

²⁹ <https://ilostat.ilo.org/methods/concepts-and-definitions/classification-occupation/> (accessed: 11/03/2026)

³⁰ In addition to ISCO-08, the (German) Classification of Occupations (KldB) is also used in the social sciences. Occupational titles are compiled in KldB, 2010. It provides a complete list of job titles in specific sectors. If the focus of the information is not on occupations but on activities, it may be more appropriate to use ISCO-08

³¹ <https://www.uis.unesco.org/en/methods-and-tools/isced> (accessed 11/03/2026)

³² This could be further supplemented with socially relevant information such as *establishment in a rural area with less than 100 people in care, large employer of women, ...*

from early childhood education in kindergarten to completed doctorates at universities. However, a look at the complex and diversely ramified German education system makes it clear that the use of the ISCED quickly reaches its limits if one seeks to retain information on the internal differentiation, variance, and specifics of the German education system in the dataset. Whether the use of lists is beneficial must be decided on a study-specific and, above all, subject-specific basis.

Table 6: Examples of the replacement of information on education and occupation

Original	Possible replacements	Remark
I study social policy.	I study [specification of subject]. I study [a social science subject]	
I achieved my university entrance qualification (<i>Abitur</i>) at this school.	I graduated from this school.	
I am a geriatric nurse.	I am [name of profession]. I am a [nurse] I am [name of occupation, service and sales workers occupation, ISCO-08: Code 05:] I am [name of occupation, personal care workers in health services, ISCO08 code 532:]. I am [name of occupation, health care professions, home-based personal care workers, ISCO08 code 5322:]	Other, freely created substitutions or other specifications of the field of activity (non-clinical, inpatient, outpatient) can be used.
My specialty is statistics.	My specialty is [specification of area of expertise]. My specialty is [specification of the mathematical subject area].	
Then I completed a one-year practical instruction program (<i>Praxisklasse</i>).	Then I completed [a school year in a model project]. Then I completed [a school year in a model project in cooperation with companies]. Then I completed [a school year in a model project for students with a secondary school-leaving certificate and learning and performance deficits].	The practical instruction model is implemented in a specific regional area. Similar type of projects exist regionally and nationally. However, depending on the context, the specific name could enable re-identification, as such models are limited to certain regions of different sizes.

5.5 Replacement of organizations and institutions

Information about organizations and institutions can also be personal. This includes information about companies where people work, theaters and museums they go to, projects they are involved in, as well as schools, clinics, and authorities provided these can be linked to a person and are relevant in the interview context. An overview of examples for this topic area can be found in Table 7. To avoid over-anonymization and preserve the potential for reuse, it is also necessary to decide on a case-by-case basis whether a replacement should be made and, if so, how much abstraction is required or what additional information such as size, orientation, or special characteristics may be helpful for secondary use. The size of the company, the number of its employees, or its national or international orientation can be important for characterizing companies (e.g., to differentiate a McDonald's restaurant from the small fast-food restaurant around the corner).

Table 7: Examples and replacement suggestions for information on institutions and organizations

Example	Possible replacements	Comment
Companies		
Karstadt	Department store chain, department store chain 1, department store chain A	
Baker Neumann	Company 1 (small family business), company 1 food sector (small family business), a grocery store	The rules for numbering should be documented, as should, e.g., the differentiation between small, medium-sized and large companies
We employ about 25 people	We are a medium-sized company with 20–30 employees	Company sizes should be defined and documented.
Burger King	System catering – more than 1,000 employees	Information such as “internationally active” could be added.
Two-person operation	Very small company	
Three apprentices	3 - 5 apprentices, under 5 apprentices	
Schools/training centers		
Adult education center	Training institution + number	
Müller School	Advanced track upper-secondary education + number, name of the school	
Kolping	Denominational association providing social assistance	
Authorities/administration		
Chamber of Commerce	A membership-based corporation under public law in the retail sector	
At the tax office	At a state authority, state office	
Other		
Radio Bremen	Public radio broadcaster	
Überseemuseum	Museum	

5.6 Information on economic sectors to specify professions and organizations

Information on economic sectors is rarely part of a replacement. Much more frequently, they are mentioned to supplement information on occupations, professional activities, and organizations. To include the economic sector in the replacement, you can research the company named in the text and provide information on this basis. However, standardized lists such as commercial registers or classifications of economic sectors (e.g., the German Classification of Economic Activities, Edition 2008, WZ) can also be helpful here. This classification is published regularly by Germany's Federal Statistical Office and is available free of charge.^{33,34} It is divided into 21 sections (A – U), which are further subdivided into 88 divisions, 272 groups, 615 classes, and 839 subclasses. Examples of how it can be applied can be found in Table 8.

Table 8: Examples of the application of WZ08

Original	Level 1	Level 2	Level 3	Level 4
Siemens	Manufacturing industry (WZ 2008 para. C)	Manufacture of computer, electronic, and optical products (WZ08.26)	Manufacture of optical instruments and photographic equipment (WZ08, 26.70.0)	
Charité	Establishment in the Human health and social work activities sector (WZ08, Q)	Human health activities (WZ08.86)	Hospital activities (WZ08, 86.1)	Hospital activities (excluding university hospitals, preventive care and rehabilitation centres) (WZ08, 86.10.01)

If an interviewee reports that they were employed by the Charité (university hospital in Berlin), this can be abstracted to varying degrees in the anonymized version by stating that they were employed in an institution in the healthcare and social services sector, in a healthcare facility, or in a hospital or university clinic. As the degree of abstraction increases, the degree of the informational content provided decreases. When using WZ2008 to replace the company name of Siemens, this could take the following form:

³³ <https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-2008-englisch.html> (accessed: 11/03/2026)

³⁴ "The classification of economic activities, 2008 edition (WZ 2008), was created with the intensive participation of data users and data producers in administration, business, research, and society. It takes into account the specifications of the statistical classification of economic activities in the European Community (Nomenclature statistique des activités économiques dans la Communauté européenne [NACE] Revision 2), which were published by Regulation (European Community [EC]) No. 1893/2006 of the European Parliament and of the Council of 20 December 2006 (Official Journal of the European Communities No. List 393 page 1). The European Commission has given its consent in accordance with Article 4, paragraph 3, of the above-mentioned Regulation." (<https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html>) (accessed 11/03/2026; translated into English).

Ms. Bach: "... We moved here because [Company B – WZ08,26: Manufacture of data processing equipment, electronic, and optical products] was of course an attractive place to work. And that's about 200km away ..."

At the European and international level, there are other classifications that can contain helpful information. These include, for example, the United Nations' International Standard Industrial Classification of all Economic Activities (ISIC),³⁵ which breaks down economic sectors and branches of industry internationally.

5.7 Replacement of health information

Health information represents an element of "individual characteristics" (cf. Table 2), which describe individual special features and so make a person re-identifiable. In the EU, it is particularly protected by Art. 9 GDPR.

If health information is to be abstracted, depending on the research topic and level of abstraction, the terms "disease," "chronic disease," "diagnosis of the first (or second) disease," or, the name of the disease (e.g., cancer) can be used. The International Classification of Diseases and Related Health Problems (ICD)³⁶ can be helpful in the search for standardized replacements. The ICD is used worldwide to classify diagnoses. It is published and regularly updated by the World Health Organization (WHO).

The ICD is structured hierarchically and contains 26 main groups of diseases at the highest level of abstraction. These include, for example, endocrine, nutritional, and metabolic diseases (group 5), diseases of the circulatory system (group 11), and diseases of the skin (group 14). These main groups branch out into increasingly differentiated subgroups. The hierarchy level can be derived from the respective ICD code number. The ICD code can be broken down into further levels, each of which provides different levels of specialized information on diseases. For example, a heart attack of an interviewee could be replaced as a "disease", as a "disease of the circulatory system (11, ICD11)", or as an "ischaemic heart disease (11, BA6Z ICD11)". From one abstraction level to the next, the number of people suffering from the designated health problem increases, thus reducing personal identifiability.

While taking into account the declaration of consent and any replacements already made, it is important to weigh what exactly needs to be replaced to prevent re-identification. Depending on the study context, a heart attack could be sufficiently unspecific and need not necessarily be further abstracted. Examples of this weighing of the degree of abstraction of replacements are described by Saunders et al. (2015a) in their comments on the anonymization of the research materials of a study with relatives of chronically ill people who were severely restricted in their perception of themselves and their environment:

³⁵ <https://ilostat ilo.org/methods/concepts-and-definitions/classification-economic-activities/> (accessed: 11/03/2026)

³⁶ <https://www.who.int/standards/classifications/classification-of-diseases> (accessed: 11/03/2026)

“Over and above name, place, occupation and religion there are often features in any dataset, which may provide unique (or closely narrowing) identifying information. We have given examples of a range of such features throughout the discussion above – including idiosyncratic details of a case. Depending on the substantive focus of the research there may also be a recurring issue that needs special attention. In our case this was the nature of the original injury. Such information was often interwoven throughout accounts and also has prognostic significance. Brain damage resulting from ‘non-traumatic’ injury (e.g. from oxygen deprivation following cardiac arrest) is more predictably devastating than resulting from traumatic brain injury (e.g. from a blow to the head). Distinguishing between these types of injury is therefore analytically important. However, beyond this we commonly used generic phrases such as ‘road traffic accident’ or ‘sporting injury’, which allowed us to describe a range of events while not revealing more specific information.” (Saunders et al. 2015a, 627)

In addition to the ICD, the WHO offers other classifications that may be of interest for the abstraction of personal health information in specific research projects. These include, for example, the International Classification of Functioning, Disability and Health (ICF)³⁷ or the International Classification of Health Interventions (ICHI).³⁸ The ICD can be fed into the QualiAnon-tool (see chapter 6). If lists are used for replacements, the list-specific terms of use must be followed and the respective version of the classification must be documented and cited (see point 4, step 5).

5.8 Replacement of positions and roles

Individuals can also be identified through information on positions and roles in their family or work contexts, with regard to political engagement or voluntary commitments, or through placings in sporting competitions. A selection of examples can be found in Table 9. Again, one must check whether the information contained in the text would facilitate personal re-identification or whether, for example, the replacement of other information in the text already virtually rules this out, so that the original information might be retained.

³⁷ <https://www.who.int/classifications/icf/en/> (accessed: 11/03/2026)

³⁸ <https://www.who.int/classifications/ichi/en/> (accessed: 11/03/2026)

Table 9: Examples of the replacement of information on positions and roles

Original	Possible replacements	Remark
He was the first chairman of the local marksmen's club.	He [held a leading position] in the local marksmen's club. He was the first chairman of the local [association].	If locations are published, e.g., in descriptions of the research project, the reference to the "local" marksmen's club could also be replaced. If the specifics of the club are replaced, it may be possible to retain the function of the chairman (see second example).
My father was a master blacksmith.	My father was a [master craftsman].	
	My father was a [master craftsman in the manufacturing industry (WZ2008 para. C)] My father was a [master craftsman in the manufacture of fabricated metal products, except machinery and equipment (WZ 2008 para. C, 25)]	Information on economic sectors can be used to provide information of varying degrees of abstraction that is relevant to the social sciences (cf. point 0).
This year I became a consultant to the Minister for Education.	This year I became a consultant [in the state's public administration]. This year, I became a consultant to the Minister [xx (further specification of the governmental department)].	Positions held by others can also contribute to re-identifying a person. In the example, the "Minister for Education" is replaced and the position of consultant is retained. This decision can vary depending on the context of the study and could also lead to the consultant position being abstracted. In the second example, information is omitted. The deletion of the two words is marked with an x for each word.
He regularly had lunch with Helmut Kohl in private.	He regularly had lunch privately with [a high-ranking German politician]. He regularly had lunch privately with [a German chancellor].	The replacement of names of public figures such as Helmut Kohl is a particular issue (see point 5.1). The example assumes that naming Helmut Kohl is associated with a re-identification risk for the research participant.

5.9 Further information that may contain a personal reference

Depending on the context of the text and the research questions, information from other areas, such as on languages, nationalities, denominations, or sports, can enable the re-identification of a person in addition to the replacement categories mentioned. Here, too, it is first necessary to check whether and, if so, how much of the information needs to be replaced to avoid over- or under-anonymization. In particular, information on religious affiliation or sexual orientation, which, like health information receives special attention from by the GDPR, must be examined. The balancing of necessary protective measures and scientific interest in knowledge is a constant consideration in the abstraction of information, as described by Saunders et al. (2015a, 624):

“We compromised through cutting references which would narrow down identification of their religion/cultural background too tightly – this involved losing some very interesting data but seemed a reasonable compromise, albeit one which made the quotations much more anodyne than they had been in the original form. The question we ask ourselves here is whether we have veered too far, and ‘white-washed’ the data, ‘forfeiting much of the richness yielded by the study’ (Parry and Mauthner, 2004: 144) and draining it of meaning?”

Difficulties arise when information on religion or religious acts contains information that is important for the research question. Saunders et al. (2015a, 623) describe, for example, how in a religious act it becomes clear that the consciousness of a patient with a brain injury is greater than estimated by doctors:

“[W]e often replaced religion/culture with similar but unrelated items, or with generalized descriptions (e.g. Jaspal’s Hinduism became ‘Jaspal’s religious faith’). This strategy became problematic when religion, culture or ethnic identity was not simply ‘background information’ but crucial ‘context for deeper and fuller understanding’ (Clark, 2006: 6). After lengthy discussion we decided to retain certain features in some cases. For example, George’s brother, David, had been injured six years prior to the interview, and at the time was diagnosed by medical professionals as having ‘low awareness’. George and his wife, Linda, (interviewed together) were very optimistic about David’s situation. Their religion underpinned their shared views about David’s future, along with a strong sense of family togetherness that they associated with their cultural heritage (from outside the UK). They consistently represented these factors as underpinning their belief in David’s ongoing recovery.”

In some cases, there are also list-like overviews for this category that can help in the search for suitable replacements. However, these must be checked for their suitability (technical quality, accessibility, and currentness) on a case-by-case basis. The website “Ethnologue language name index,”³⁹ for example, could provide helpful information on languages. In addition, official websites such as those of Germany’s Federal Statistical Office⁴⁰, the UN⁴¹, the WHO,⁴² or EUROSTAT⁴³ provide information on countries, their populations, their regional profiles, and other aspects that could support the search for suitable abstractions in terms of social-scientifically relevant information.

5.10 Accumulation of different categories of personal information

The aim of the procedures described here is to comply with data protection requirements while preserving the reuse potential of a dataset. We have described individual replacement categories that often contain personal information and recommend a procedure that examines all information to be potentially replaced in detail for its personal identifiability. When deciding whether or not a piece of personal information should remain in the text, we should bear in

³⁹ <https://www.ethnologue.com/data-consulting> (accessed: 11/03/2026)

⁴⁰ https://www.destatis.de/EN/Themes/_node.html (accessed: 11/03/2026)

⁴¹ <https://unstats.un.org/unsd/classifications/> (accessed: 11/03/2026)

⁴² <https://www.who.int/standards/classifications/other-classifications> (accessed: 11/03/2026)

⁴³ <https://ec.europa.eu/eurostat/web/main/data/statistical-themes> (accessed: 11/03/2026)

mind that replacements made in one category may influence the identifiability of personal information in other replacement categories. For instance, if we abstract not only direct identifiers such as names and addresses but also information on location and time, this may already be sufficient for anonymizing a text. Conversely, mentioning the place of data collection in publications and contextual materials can lead to the need for a much more comprehensive intervention in the interview text to achieve this goal.

The study design also influences the decision for a suitable anonymization concept. When archiving panel studies that research the same people over several years, for example, consistent anonymization should be ensured across all waves. Here it may help to replace certain information such as place names consistently across the study from the outset in the same standardized way – in contrast to what was described above – even without checking for personally identifiable information, to facilitate the consistent handling of location information. In other words, without checking whether, for example, mentioning the town of “Kappeln on the Schlei” in various places in the materials contains personally identifiable information (e.g., being a place of birth or just a vacation destination), all mentions of “Kappeln” would be replaced to ensure pragmatically that this town – defined as ‘place 1, a small medium-sized town near the water and dominated by agriculture’ in wave 1 – corresponds to that same town in wave 5. In this case, we would forego the possible preservation of original information in favor of a consistent anonymization concept that remains the same over several waves. This approach would have the advantage of reducing any time-consuming reworking of interviews from previous waves, which could become necessary due to references between interviews with the same people and the associated risk of re-identification.

6 QualiAnon – The Qualiservice tool for anonymization or pseudonymization of text-based research materials⁴⁴

Abstracting personal information in text-based research materials for the purpose of scientific research is a resource-intensive process. This is particularly true when it comes to preparing the materials for subsequent scientific use. To address this issue, Qualiservice developed in cooperation with the chair of Computational Social Science at the Technical University of Munich and the Munich School of Public Policy and the world data archive PANGAEA an anonymization tool – QualiAnon – that can be used to make and process case- or study-related replacements and ensures that identification data can be stored separately from other information such as anonymized interview transcripts.

QualiAnon is designed so that users always have control over the workings of the software and its processing of information. The tool runs without a connection to the Internet. Automatic

⁴⁴ QualiAnon was developed as an open-source tool by the Qualiservice research data centre at the University of Bremen with funding from the DFG (project number: 398979974). Since 2025, it has been maintained and provided as part of the research data management portfolio of QualidataNet, a service of the Consortium for Social, Behavioural, Educational and Economic Data (KonsortSWD-NFDI4Society) for qualitative research (DFG project number: 42494171) in the context of the National Research Data Infrastructure (NFDI) in Germany.

processes have been deliberately omitted. This means that QualiAnon supports the replacement process but does not make any replacements automatically. The user remains in charge of critically reading and classifying information as sensitive.

The following considerations played a special role in the development: The anonymization tool should be able to flag sensitive information and replace it using different levels of abstraction. The replacement should be documentable and editable and, if desired, the original information should also be preserved (pseudonymization).

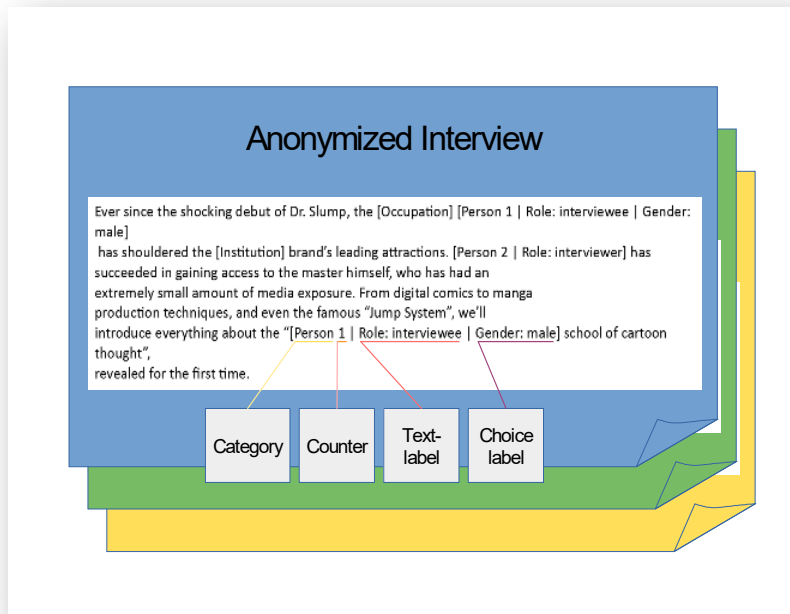


Figure 3: Example of anonymization using QualiAnon (illustration: Tom Nicolai)

Figure 3 shows an example of using QualiAnon for the anonymization of a name in an interview excerpt. For each entity that QualiAnon replaces, a category can be selected from an editable system of replacement categories (e.g., person, profession, location). The information used to replace the name in the example is given in square brackets: “[Person 1 | Role: interviewee | Gender: male].”

The applied substitution categories can be numbered consecutively for all interviews in a study so that person 1 is recognizable as the same person throughout the entire interview or in further interviews. In addition, depending on the sensitivity of the information and the desired security category, different levels of abstraction of the replacement can be selected, which facilitates the implementation of the concept of “Flexible Anonymization” (cf. point 3.4). In the example, the replacement is that person 1 is the interviewee and is male. The system of categories created in the replacement process can be exported and shared with other researchers (see the github-manual for the tool). In this way, topic- or area-specific replacement schemes can be created that provide users working on similar topics with qualified suggestions for reference points or even elaborated anonymization concepts for their own project.

Further developments were necessary to enable researchers to use QualiAnon outside of the specially protected curation environment of an RDC. To ensure that research data is handled in compliance with data protection legislation, users should be able to store identification data separately from other information.

The tool works with a system of universally unique identifiers (UUIDs; cf. Figure 4). The anonymized interview (“blackened interview”) contains the marker ID that is set in the original text during editing. The replacement ID refers to the selected translation. The marker IDs and replacement IDs of a study as well as the original information are stored in a separate document.

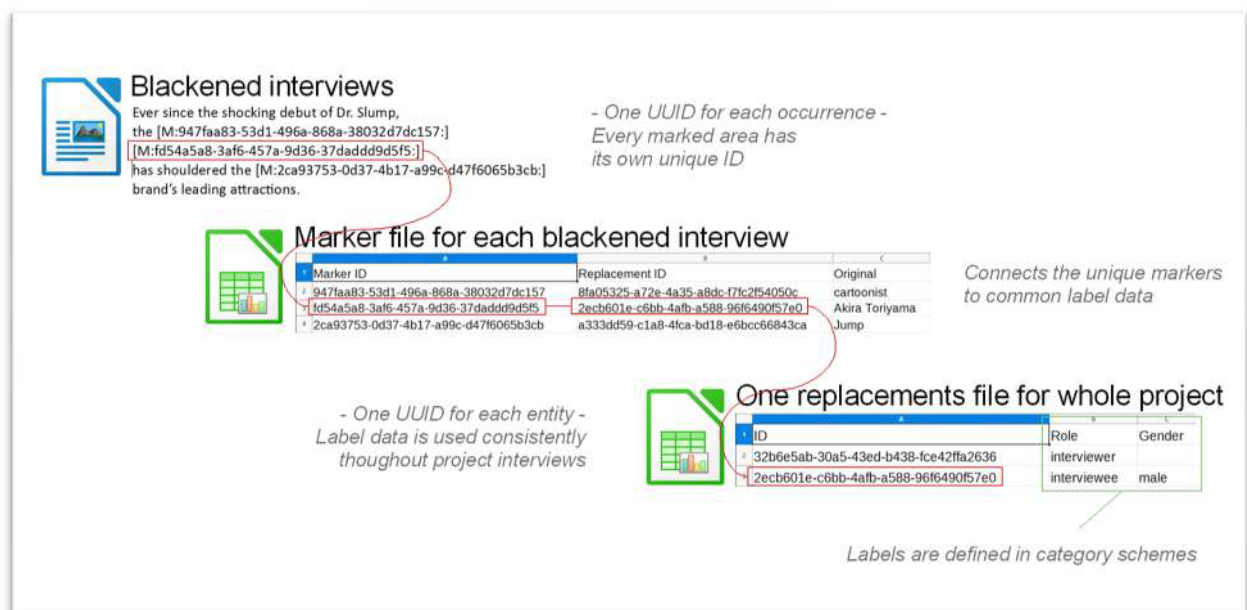


Figure 4: The QualiAnon data model, connections between UUIDs (illustration by Tom Nicolai)

The current version of the QualiAnon tool (Nicolai et al. 2025) was presented in its virtual launch in May 2021 and is being continuously further refined. Figure 5 shows how the individual building blocks of the program can be organized during the processing of a project.

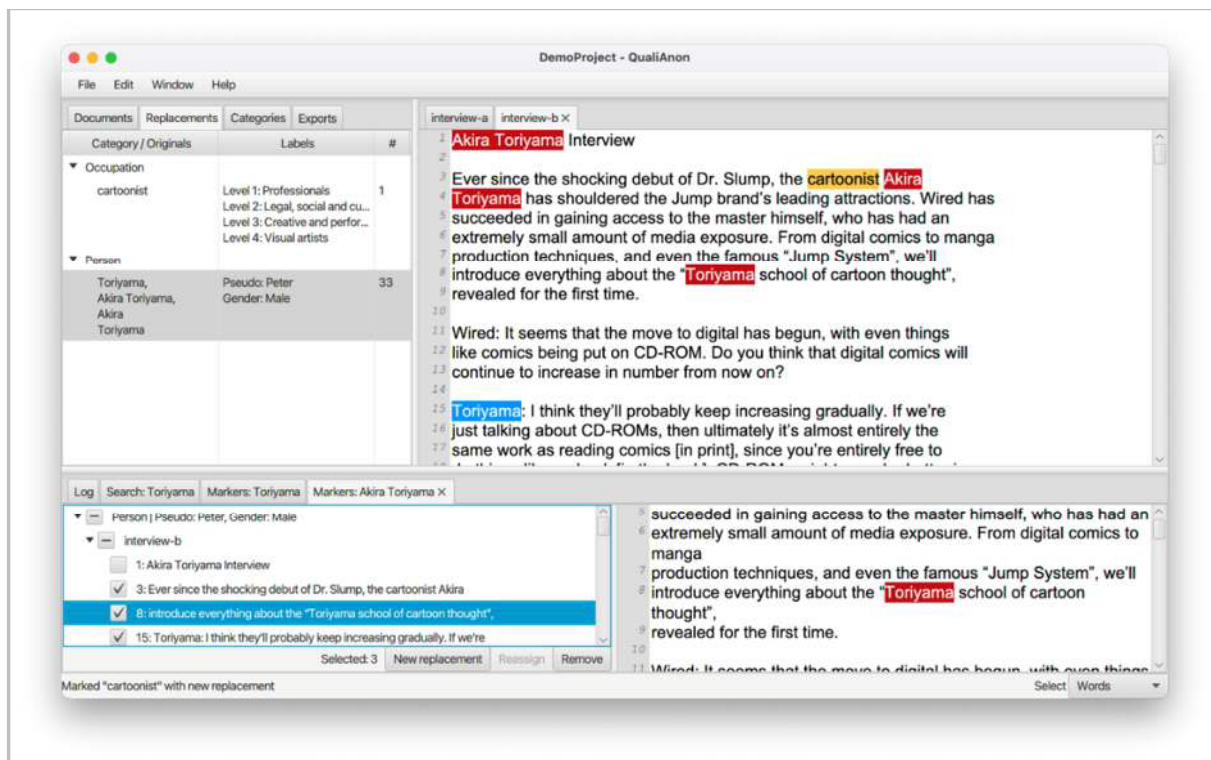


Figure 5: Screenshot of program application (Tom Nicolai)

QualiAnon has been developed as an open-source tool. The code as well as a user manual are published via a Github repository.⁴⁵ Access is available free of charge.⁴⁶ The individual application steps can be found in the detailed manual and the video tutorial.⁴⁷

⁴⁵ <https://github.com/pangaea-data-publisher/qualianon> (accessed: 11/03/2026)

⁴⁶ <https://www.qualiservice.org/en/the-helpdesk/tools.html> (accessed: 11/03/2026)

⁴⁷ <https://www.youtube.com/watch?v=RYQn4DjdKmo> (accessed: 11/03/2026)

7 Summary and outlook

The anonymization of personal data is a strategy for meeting data protection and research ethics requirements and thus also the requirements of good scientific practice (DFG, 2022). In contrast to quantitative social research, qualitative social research lacks established anonymization procedures, which also owes itself to the nature of the research materials, their particular density of information and context, and the high degree of interrelatedness of the information. A further challenge for the anonymization of qualitative text data lies in the fact that the proximity to the original text must be maintained as much as possible to be able to reuse qualitative research materials scientifically. Step by step, from case to case, and often also in relation to the mention of one and the same piece of information in the text, it must be decided whether information in the text harbors a risk of re-identification and must be replaced and which information can be retained in the original wording. At the same time, it is important to look for ways to maintain the readability of the text.

These particular challenges speak against the implementation of a standardized anonymization procedure in qualitative social research. At the same time, owing to limited resources and with a view to preserving the potential for reuse, it makes sense to establish (topic- and discipline-specific) best practices in qualitative social research as well. Establishing a set of best practices makes it possible to further develop strategies and record the advantages and disadvantages of certain procedures. The use of authorized lists, regularly updated and compiled by public institutions, for the replacement of, say, places or diseases can help to develop topic-specific partial standardizations, as can the advancement and sharing of topic-specific schemes of replacement categories. However, it must be considered on a case-by-case basis whether the use of standardized lists has benefits for the preservation of social-scientific information or whether completely different information (which is not recorded in a standardized way) is more relevant to subsequent use.

To plan the anonymization of qualitative research data and estimate the necessary effort, it is advisable to first create and document an anonymization concept. The purpose for which anonymization is carried out is crucial to estimating the necessary effort: Anonymizing an illustrative quote in a scientific journal is far less time-consuming than preparing an entire interview transcript for subsequent use. We must also consider for whom the specific research materials are to be anonymized: Should the anonymized materials be publicly accessible, so that even friends or colleagues of interviewees can view the materials? Or will access be limited to researchers who sign a user agreement, as is the case with Qualiservice? Questions of over- and under-anonymization also arise differently in relation to the respective purpose of use and must be assessed against the background of the planned type of subsequent use. Detailed documentation of the chosen anonymization procedure helps to ensure a consistent approach within the research team and enables secondary users to track the changes made to the text.

In conclusion, it should be noted that absolute anonymity cannot be achieved in a digitized world such as ours while at the same time preserving the scientific reuse value of qualitative

research materials. What is feasible, however, is the possibilities offered by de facto anonymization and pseudonymization in conjunction with other organizational and technical measures such as encryption, password protection, access restrictions, separate storage of identifiers and research materials, onsite use, and so forth. These reduce the risk of possible re-identification of individuals and minimize it to an extent that the qualitative data can continue to be researched or used in the preparation of applications and in academic teaching. For this sustainable approach, Qualiservice is working together with researchers on innovative solutions that are practicable and leave enough room for study- and case-specific flexibility. We hope to be able to support researchers in developing a suitable approach with our guide. Of course, we are happy to answer any questions and welcome suggestions, reports on experiences, and feedback.

8 References

- Bishop, L. (2005). Protecting Respondents and Enabling Data Sharing: Reply to Parry and Mauthner. *Sociology*, 39(2), 333-336.
- Corti, L., Day, A., & Backhouse, G. (2000). Confidentiality and Informed Consent: Issues for Consideration in the Preservation of and Provision of Access to Qualitative Data Archives. *FQS*, 1(3), Art. 7.
- Gerhards, J. (2005). *The Name Game*. Transaction Publishers.
- Gerhards, J. (2010). *Die Moderne und ihre Vornamen*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92353-6>
- Heuer, J.-O., Hollstein, B., & Mozygemba, K. (2021). Data Sharing mit Qualiservice. *Soziologie*, 50 (4), 459–471.
- German Research Foundation/ Deutsche Forschungsgemeinschaft (DFG). (2019). *Handreichung des Fachkollegiums 106 Sozial- und Kulturanthropologie, Außereuropäische Kulturen, Judaistik und Religionswissenschaft zum Umgang mit Forschungsdaten*. <https://www.dfg.de/re-source/blob/173710/045c16e9430426ef321c76dabcbaf51/handreichung-fachkollegium-106-forschungsdaten-data.pdf> (accessed: 11/03/2026)
- German Research Foundation/ Deutsche Forschungsgemeinschaft. (2022). *Guidelines for Safeguarding Good Research Practice. Code of Conduct*. <https://doi.org/10.5281/zenodo.6472827>.
- German Educational Research Association (GERA)/ Deutsche Gesellschaft für Erziehungswissenschaft (DGfE). (2005). *Anonymisierung von Daten in der qualitativen Forschung: Probleme und Empfehlungen*. https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2005.11_Anonymisierung_von_Daten.pdf (accessed 11/03/2026)
- German Educational Research Association (GERA)/ Deutsche Gesellschaft für Erziehungswissenschaft (DGfE). (2016). Ethikkodex der Deutschen Gesellschaft für Erziehungswissenschaft. *Erziehungswissenschaft*, 53 (27).
- German Sociological Association/ Deutsche Gesellschaft für Soziologie (DGS). (2017). *Ethikkodex der deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS)*. https://soziologie.de/fileadmin/user_upload/dokumente/Ethik-Kodex_2017-06-10.pdf (accessed: 11/03/2026)
- German Association for Folklore Studies/ Deutsche Gesellschaft für Volkskunde (DGV) (Ed.). (2018). *Positionspapier zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten*. https://www.d-g-v.de/wp-content/uploads/2019/04/dgv-Positionspapier_FDM.pdf (accessed: 11/03/2026)
- Ebel, T. (n. d.). *Empfehlungen zur Anonymisierung quantitativer Daten*. https://www.pe-docs.de/volltexte/2022/21970/pdf/fdb-informiert_3_Ebel_ua_Hinweise_zur_Anonymisierung_von_quantitativen_Daten_2015_v1-2_A.pdf (accessed: 11/03/2026)
- Fielding, N. (2004). Getting the Most from Archived Qualitative Data: Epistemological, Practical and Professional Obstacles. *Int J Social Research Methodology*, 7(1), 97-104.
- Finnish Social Science Data Archive. (2021). *Anonymization and Personal Data*. <https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymization-and-identifiers/> (accessed: 11/03/2026)
- Heuer, J.-O., Kretzer, S., Mozygemba, K., Huber, E., & Hollstein, B. (2025). *Contextualisation of Qualitative Research Data for Subsequent Use. A Handout for Researchers on Creating a Study Report for the Qualiservice Research Data Centre*. QS-WP-7-2025, translated and adapted version of the Qualiservice Working Paper QS-WP-1-2020, Bremen <https://doi.org/10.26092/elib/3791>
- Hopf, C. (2004). Forschungsethik und qualitative Forschung. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *Qualitative Forschung. Ein Handbuch* (pp. 589-600). Rowohlt.
- Huber, E. & Imeri, S. (2021). Informed Consent in Ethnographic Research: A Common Practice Facing New Challenges. *Qualiservice Working Papers*. Bremen: Forschungsdatenzentrum Qualiservice. <https://doi.org/10.26092/elib/1070>

- Kelly, A. (2009). In Defence of Anonymity: Rejoining the Criticism. *British Educational Research Journal*, 35(3), 431-445.
- Kretzer, S. (2013). Arbeitspapier zur Konzeptentwicklung der Anonymizierungs-/Pseudonymisierung in Qualiservice. https://wiso.genios.de/document/SSOA_47605 (accessed: 09/10/2024)
- Kretzer, S., Mozygemba, K., Heuer, J.-O., & Huber, E. (2020). Erläuterungen zur Verwendung der von Qualiservice bereitgestellten Vorlagen für die informierte Einwilligung. *Qualiservice Working Papers*, QS-WP-2-2020. <https://doi.org/10.26092/elib/192>
- Laudel, G. & Bielick, J. (2019). Forschungspraktische Probleme bei der Archivierung von leitfadengestützten Interviews. *Forum Qualitative Sozialforschung*, 20(2), 1-32. <https://doi.org/10.17169/fqs-20.2.3077>
- Medjedović, I. (2011). Secondary Analysis of Qualitative Interview Data: Objections and Experiences. Results of a German Feasibility Study. *FQS*, 12(3), Art 10. <https://doi.org/10.17169/fqs-12.3.1742>
- Medjedović, I. (2014). *Qualitative Sekundäranalyse. Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung*. Springer VS.
- Meyermann, A. & Porzelt, M. (2014). *Hinweise zur Anonymisierung qualitativer Daten* (Vol. 1). Deutsches Institut für Internationale Pädagogische Forschung. <http://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-1.pdf> (accessed: 11/03/2026)
- MLS Legal (2020). *Rechtsfragen zur Nutzung von Daten – Forschung, Archivierung, Sekundärnutzung*. <https://www.youtube.com/watch?v=vC8ucG9MJSw&t=1s> (accessed: 11/03/2026).
- Mozygemba, K. & Kretzer, S. (2022). Datenvielfalt im Data-Sharing – eine kooperative Aufgabe von Forschenden und Forschungsdatenzentrum. In C. Lohmeier & T. Wiedemann (Eds.), *Datenvielfalt in kommunikationswissenschaftlichen Forschungskontexten* (pp. 157–178). Springer VS.
- Nicolai, T., Mozygemba, K., Kretzer, S., Hollstein, S. & Gordeev, E. (2025): QualiAnon - Qualiservice tool for anonymizing text data (version 1.5.0). Qualiservice. University of Bremen. Software available at: <https://github.com/pangaea-data-publisher/qualianon>.
- Opitz, D., & Mauer, R. (2005). Erfahrungen mit der Sekundärnutzung von qualitativem Datenmaterial – Erste Ergebnisse einer schriftlichen Befragung im Rahmen der Machbarkeitsstudie zur Archivierung und Sekundärnutzung qualitativer Interviewdaten. *FQS*, 6(1), Art. 13.
- Rat für Sozial- und Wirtschaftsdaten [German Data Forum] (RatSWD). (2017a). *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften*. Bundesministerium für Bildung und Forschung (BMBF), <https://doi.org/10.17620/02671.1>
- RatSWD [German Data Forum] (2020): Data Protection Guide. 2nd fully revised edition. RatSWD Output 8(6). Berlin, German Data Forum (RatSWD). <https://doi.org/10.17620/02671.57>
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), <https://gdpr-info.eu/> (accessed: 11/03/2026)
- Richter, C., Kwelik, N., Müller, M., & Severing, L. (2021). Qualitative Daten anonymisieren und für Sekundäranalysen aufbereiten: Das Bochumer Anonymisierungsmodell (BAM). In C. Richter & K. Mojescik (Eds.), *Qualitative Sekundäranalysen. Daten der Sozialforschung aufbereiten und nachnutzen* (pp. 153–184). Springer VS.
- Saunders, B., Kitzinger, J., & Kitzinger, C. (2015a). Anonymizing Interview Data: Challenges and Compromise in Practice. *Qualitative Research*, 15(5), 616-632. <https://doi.org/10.1177/1468794114550439>
- Saunders, B., Kitzinger, J., & Kitzinger, C. (2015b). Participant Anonymity in the Internet Age: From Theory to Practice. *Qualitative Research in Psychology*, 12(2), 125-137. <https://doi.org/10.1080/14780887.2014.948697>
- Stam, A., Kleiner, B. (2020): Data Anonymisation: legal, ethical, and strategic considerations. FORS Guide No. 11, Version 1.0, Lausanne: Swiss Centre of Expertise in the Social Sciences FORS. doi:10.24449/FG-2020-00011

- Statistisches Bundesamt (Destatis). (2008). *German Classification of Economic Activities*. (Edition 8). https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-2008-englisch.xls?__blob=publicationFile. (accessed: 11/03/2026)
- Stein, A. (2010). Sex, Truths, and Audiotape: Anonymity and the Ethics of Exposure in Public Ethnography. *Journal of Contemporary Ethnography*, 39(5), 554-568.
- Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T., & Estabrooks, C. A. (2005). Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. *FQS*, 6(1), Art. 29.
- UK Data Service. (2021). <https://ukdataservice.ac.uk/events/webinar-how-to-anonymize-qualitative-and-quantitative-data/> (accessed: 11/03/2026)
- University of Leeds. (2008). *Timescapes Anonymization Guidelines. Version in Use 18 Aug 08 1b*. <https://timescapes-archive.leeds.ac.uk/wp-content/uploads/sites/47/2018/04/Timescapes-Anonymization-Guidelines-18Aug08-in-use.pdf> (accessed: 11/03/2026)
- von Unger, H. (2014). Forschungsethik in der qualitativen Forschung: Grundsätze, Debatten und offene Fragen. In H. von Unger, P. Narimani, & R. M'Bayo (Eds.), *Forschungsethik in der qualitativen Forschung: Reflexivität, Perspektiven, Positionen* (pp. 15-39). VS Verlag.
- von Unger, H. (2018). Forschungsethik, digitale und biographische Interviews. In H. Lutz, M. Schiebel, & E. Tuider (Eds.), *Handbuch Biographieforschung* (pp. 681-693). Springer VS.
- von Unger, H. (2020). Archivierung und Nachnutzung qualitativer Daten aus forschungsethischer Perspektive. In RatSWD (Ed.), *Handreichung Datenschutz*. <https://doi.org/10.17620/02671.50>
- Watteler, O. & Ebel, T. (2019). Datenschutz im Forschungsdatenmanagement. In U. Jensen, S. Net-scher, & K. Weller (Eds.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (pp. 57-80). Barbara Budrich. <https://doi.org/10.2307/j.ctvbkk1p8.7>
- Züll, C. (2015). *Berufscodierung* (SDM – Survey Guidelines). GESIS Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/sdm-sg_019

9 Appendix: Examples of the replacement of personal qualitative text data

All original texts were translated from German to English.

9.1 Example 1⁴⁸

Original text:

P08: Well, and then I took part in competitions anyway. It started, I was, well, it started with the club championship; then I became club champion; I became county champion, district champion, state champion. Yes, and then it always continued until I became state champion, four years like that. And then in the fifth year, well, my dream was always to get to the German championship. Yes, and then I made it in 1970, right? Well, first the finals in Rothenburg, there, in the Sven-Hannawald-Gym. That was the state championship, so I shot air rifle and small bore, right? (Clears throat).

IV1: Hm, hm.

P06: Yes, and then I went to Berlin (indistinct) there, in the Olympic stadium, then, the shooting (indistinct), and that had always been my wish; and then that was it.

Marking of possibly sensitive passages in example 1

P08: Well, and then I took part in competitions anyway. It started, I was, well, it started with the club championship; then I became club champion, I became county champion, district champion, state champion. Yes, and then it always went on until I became state champion, four years like that, and then in the fifth year, well, my dream was always to get to the German championship. Yes, and then I made it in 1970, right? Well, first the finals in Rothenburg, there, in the Sven-Hannawald-Halle. That was the state championship, so I shot air rifle and small bore, right? (Clears throat).

IV1: Hm, hm.

P06: Yes, and then I went to Berlin (indistinct) there, in the Olympic stadium, then, the shooting (indistinct), and that had always been my wish; and then that was it.

Anonymization of example 1:

P08: Well, and then I took part in competitions anyway. It started, I was, well, it started with the club championship; then I became club champion, I became county champion, district champion, state champion. Yes, and then, so it always went on until I became state champion, four years like that. And then in the fifth year, well, my dream was always to get to the German championship. Yes, and then I made it in 1970, right? Well, first the finals in location A, rural community, there, in the venue for competitions 1. That was the national championship, so I shot in two disciplines, right? (Clears throat).

IV1: Hm, hm.

P06: Yes, and then I went to place B, big city (indistinct) there, in the venue for competitions 2, then, the shooting (indistinct), and that had always been my wish; and then that was it.

⁴⁸ Different colours represent different replacement categories such as locations, time, diseases, occupations.

9.2 Example 2

Original text:

A: (1) Well, I, (1) I had actually started out differently professionally. I got married, had two children, stopped working, and it was really bad. My father-in-law had a stroke on the first of April 1987. I was at home for about 10 years and then did an internship in an old people's home as part of a nursing course and was immediately asked if I could stand in as a care assistant (2). That's what I did and I just stuck with it. The one time I stepped in turned into seven years of work. And then I thought, well, if I'm already doing this work, then I'll do the training too. And then I did the training to become a nursing assistant and then worked straight away as a manager in the care sector. But of course I'd already been working in this field for seven years, right?

Marking of possibly sensitive passages in example 2

A: (1) Well, I, (1) I had actually started out differently professionally. I got married, had two children, stopped working, and it was really bad. My father-in-law had a stroke on the first of April 1987. I was at home for about 10 years and then did an internship in an old people's home as part of a nursing course and was immediately asked if I could stand in as a care assistant (2). That's what I did and I just stuck with it. The one time I stepped in turned into seven years of work. And then I thought, well, if I'm already doing this work, then I'll do the training too. And then I did the training to become a nursing assistant and then worked straight away as a manager in the care sector. But of course I'd already been working in this field for seven years, right?

Anonymization of example 2:

A: (1) Well, I, (1) I had actually started out differently professionally. I got married, had two children, stopped working, and that was really bad. My father-in-law had a cerebrovascular disease (ICD11-8B2z) at the end of the 1980s. I was at home for about 10 years and then did an internship as part of a training course in healthcare and was immediately asked if I could stand in as a personal care worker in health services (ISCO 532) (2). That's what I did that and I just stuck with it. The one time I stepped in turned into seven years of work. And then I thought, well, if I'm already doing this work, then I'll do the training too. And then I did the training to become a personal care workers in health services (ISCO-08 532) and then worked straight away as a manager in personal care workers in health services (ISCO 532). But of course I'd already been working in this field for seven years, right?

9.3 Example 3

Original text:

- P001: No, my husband was with another company. I don't know, do you know the Hamburg stores? My husband was at Flimmermüller and I was at Hagenschmidt.
- IV1: Flimmermüller, that's familiar.
- P001: And I still have ten years plus/.
- IV1: But that means it was always seasonal work. So whenever Dom⁴⁹ plus/.
- P001: Only Dom then. Because I worked as a salesclerk in Finkenwerder, but only until noon. And then I started at the market at three in the afternoon, until the evening, on top of that.
- IV1: Oh, the sales job, was that here in Finkenwerder? Well, that's a different one.
- P001: Still another job. There in the fruit and vegetable (department), I was a sales assistant for twelve years.

Marking of possibly sensitive passages in example 3

- P001: No, my husband was with another company. I don't know, do you know the **Hamburg** stores? My husband was at **Flimmermüller** and I was at **Hagenschmidt**.
- IV1: **Flimmermüller**, that's familiar.
- P001: And I still have ten years plus/.
- IV1: But that means it was always seasonal work. So whenever **Dom** plus/.
- P001: Only **Dom** then. Because I worked as a salesclerk in **Finkenwerder**, but until only noon. And then I started at the market at three in the afternoon, until the evening, on top of that.
- IV1: Oh, the sales job, was that in **Finkenwerder**? Well, that's a different one.
- P001: Still another job. There in the fruit and vegetable (department), I was a salesclerk for twelve years.

Anonymization example 3:

- P001: No, my husband was with another company. I don't know, do you know the **[local]** stores? My husband worked for **[company A, retail sale of food, beverages and tobacco (in sales rooms) – WZ47.2]** and I worked for **[company B, retail sale of textiles – WZ47.51]**.
- IV1: **[Company A]**, that's familiar.
- P001: And I still have ten years plus/.
- IV1: But that means it was always seasonal work. So whenever there was **[festival]** plus/.
- P001: Only **[festival]** then. Because I worked as a salesclerk in **[district A in city A, big city]**, but only until noon. And then I started at the market at three in the afternoon, until the evening, on top of that.
- IV1: Oh, the sales job, was that in **[district A in city A, big city]**? Well, that's another one.
- P001: Still another job. There in the fruit and vegetable (department), I was a salesclerk for twelve years.

⁴⁹ a funfair

About Qualiservice.

Qualiservice is a research data centre that archives and provides qualitative research data from the various fields of the social sciences for scientific re-use. Our services include personalized and study-specific advice, as well as curation and secure processing of research data for re-use and long-term archiving. We also provide research data and relevant contextual information for scientific re-use in research and teaching. Internationally interoperable metadata ensure that all data sets can be searched and found. Persistent identifiers (i.e., digital object identifiers, DOI) guarantee that data and details about the study remain permanently citable.

Qualiservice was accredited by the German Data Forum (RatSWD) in 2019 and adheres to its quality assurance criteria. Qualiservice is committed to the *DFG Guidelines for Safeguarding Good Scientific Practice* and takes into account the *FAIR Guiding Principles for scientific data management and stewardship* as well as the *OECD Principles and Guidelines for Access to Research Data from Public Funding*.

Qualiservice partner:



Specialised Information Service for Social and Cultural Anthropology (FID)



Specialised Information Service (FID) Political Science -



Specialised Information Service (FID) Criminology



Data portal for Racism and Right-Wing Extremism Research (DP-R|EX)



Specialised Information Service (FID) Sociology - Socio-

Accreditation by

Research Data Centre accredited by the RatSWD 

Funded by

