data

MDPI

Article

Expert Experiences in Anonymizing Personal Data and Its Use as Open Data: Qualitative Insights

Norbert Lichtenauer [1,*], Johann Guggumos [2], Matthias Kampmann [3], Juliane Kis [3], Florian Laumer [4], Elena März [2], Florian Wahl [5] and Sebastian Wilhelm [5]

[1] Deggendorf Institute of Technology, Health Campus Bad Kötzting, 93444 Bad Kötzting, Germany
[2] Faculty of Law, University of Augsburg, 86159 Augsburg, Germany; johann.guggumos@jura.uni-augsburg.de (J.G.); elena.maerz@jura.uni-augsburg.de (E.M.)
[3] IT-Sicherheitscluster e.V., 93053 Regensburg, Germany; matthias.kampmann@it-sicherheitscluster.de (M.K.); juliane.kis@it-sicherheitscluster.de (J.K.)
[4] PASSION4IT GmbH, 94234 Viechtach, Germany; florian.laumer@passion4it.de
[5] Deggendorf Institute of Technology, Technology Campus Grafenau, 94481 Grafenau, Germany; florian.wahl@th-deg.de (F.W.); sebastian.wilhelm@th-deg.de (S.W.)
* Correspondence: norbert.lichtenauer@th-deg.de; Tel.: +49-991-3615-8392

Abstract

Introduction: The effective and meaningful use of anonymized personal data, including open data, is globally significant across various sectors. Enhancing data utilization aims to generate substantial societal benefits and added value through innovations, products, and services. However, several legal, ethical, and technical challenges currently hinder the development and broader adoption of open data. Furthermore, the availability of technical support tools with high usability is especially desirable to facilitate the anonymization process effectively. ***Methods***: As part of the EAsyAnon research project, preliminary insights were gathered through a scoping review that identified factors promoting or impeding the anonymization and use of personal data. Based on these findings, a structured interview guide was developed. Following a pretest, 19 interviews were conducted with diverse stakeholders from healthcare institutions, research organizations, public authorities, and private companies. The collected data were analyzed using Kuckartz's structural content analysis methodology, supported by qualitative analysis software. ***Results***: The content analysis yielded five overarching categories and 21 subcategories. These encompassed stakeholder experiences related to anonymization and open data processes, the various types and formats of personal data, identified barriers and enabling factors, support services, and the ethical and legal considerations associated with anonymization. ***Discussion***: The findings highlight significant uncertainty among stakeholders regarding the anonymization of personal data. Although the importance and potential applications of open data for innovation and continuous improvement are widely acknowledged and supported, numerous challenges persist at both the macro and micro levels. The results emphasize a clear need for targeted support measures to address these challenges effectively.

Keywords: open data; anonymization; personal data; data use; data value creation

check for updates

Academic Editor: Donghee Shin

Received: 6 March 2025
Revised: 19 May 2025
Accepted: 19 June 2025
Published: 1 July 2025

Citation: Lichtenauer, N.; Guggumos, J.; Kampmann, M.; Kis, J.; Laumer, F.; März, E.; Wahl, F.; Wilhelm, S. Expert Experiences in Anonymizing Personal Data and Its Use as Open Data: Qualitative Insights. *Data* **2025**, *10*, 105. https://doi.org/10.3390/data10070105

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Open data (OD) initiatives have increasingly influenced a wide range of industries and sectors globally [1]. The release of anonymized, in the sense of Recital 26 of the GDPR, previously personal data (PD), within the meaning of Article 4 of the GDPR, is often

perceived as a significant challenge, necessitating a balance between protecting individual privacy and maximizing the utility of anonymized data [2].

Advances in OD and big data technologies have exponentially expanded the potential for fully automated data collection and analysis in recent years. As a result, the management of such data has become a critical aspect of contemporary societal transformation, with far-reaching implications for all facets of human life [1,3]. OD applications are regarded as vital drivers of progress across society, business, science, and medicine [4], prompting ongoing efforts to develop innovative applications for OD [5,6].

While there is a societal demand for fostering an open data culture, increasingly stringent laws designed to protect PD present practical barriers to its reuse [7]. Currently, the protection of personal data remains one of the most pressing challenges, particularly concerning health data [4]. The potential for PD to be repurposed beyond its originally intended scope underscores the importance of adhering to legal frameworks governing data usage [8]. Ensuring compliance with these laws is imperative [9], given the already stringent ethical and data protection standards for PD utilization [10]. The General Data Protection Regulation (GDPR), which facilitates the harmonization and portability of personal data across Europe, has made significant strides in this area. However, challenges in implementing OD processes persist [8]. In parallel, European initiatives such as Gaia-X aim to establish a European data infrastructure, enabling digital sovereignty, interoperability, and the successful adoption of open-source principles through collaborative efforts from business, academia, and government representatives [11].

The main contribution of this article is to qualitatively explore the practical experiences, challenges, and requirements encountered by stakeholders in diverse sectors in Germany during PD anonymization processes intended for OD publication. Although the existing literature highlights numerous technical, legal, and ethical barriers to the anonymization and reuse of PD as OD, detailed expert perspectives from practice remain underrepresented. Most international studies focus on general organizational, legal, and ethical barriers without considering the specific regional and national implementation challenges [12].

To address this gap and build an empirical foundation for practical anonymization approaches in Germany, we conducted a dedicated regional baseline study on expert experiences related to anonymization and data publication. Specifically, this article provides insights into the following key areas: (1) the diversity and complexity of PD handled across sectors, (2) stakeholders' actual experiences and misconceptions regarding PD anonymization (e.g., confusing anonymization with pseudonymization or underestimating re-identification risks), (3) the identification of economic, technical, legal, ethical, and organizational barriers, and (4) the critical support mechanisms needed to effectively anonymize PD (i.e., ensuring that the risk of re-identification is minimized and compliance with the relevant legal standards, such as those outlined in GDPR Recital 26, is achieved) and promote its broader utilization as OD.

2. Relevance and Problem Definition

In research, alongside the advancements in OD, the concept of Open Science emphasizes free access to scientific publications and raw research data [13]. As early as 2021, UNESCO (United Nations Educational, Scientific, and Cultural Organization) identified Open Science as a critical tool for enhancing the quality of scientific outcomes and processes [14]. In particular, the research sector anticipates substantial benefits from big data analyses [13,15]. To support this vision, the European Commission introduced the European Open Science Cloud (EOSC), a strategy aimed at enabling data exchange and advanced analyses of publicly funded research data while conserving resources [14]. Simi-

larly, Germany is implementing its National Research Data Infrastructure (NFDI) as part of its digital strategy to make existing research data more findable and interoperable [16].

The global COVID-19 pandemic has underscored significant challenges in data exchange within the **healthcare sector**, as PD had to be collected globally, integrated, and made accessible to researchers [17,18]. The demand for international collaboration in data exchange has grown steadily, with calls over recent years to fully leverage the opportunities provided by artificial intelligence (AI) and big data in medicine [19]. Complementing the EOSC, there are increasing demands at the European level for cross-border data exchange in electronic health services to optimize care pathways [8,19]. This initiative is politically supported by the European Health Data Space (EHDS). For years, experts have advocated for a stronger emphasis on technology and big data utilization in healthcare [20]. The application of big data in healthcare offers vast potential: it can enhance prognosis and diagnostics, enable innovative prevention strategies, improve treatment quality, and increase efficiency [3,19–21]. However, training AI models in medicine often requires substantial volumes of specific data [22]. Access to scientific health data is critical for advancing scientific progress and fostering innovation [23]. A key component of medical research involves the ability to link individual data sets, which is viewed as central to unlocking new insights [24]. Nonetheless, this capability raises social and ethical dilemmas, as it necessitates balancing individual data protection with potential societal benefits [2]. Moreover, linking health-related data from diverse sources can provide profound insights into highly intimate aspects of individuals' lives, posing significant ethical risks [1]. As a result, much health data remain stored in isolated "data silos", rendering them inaccessible for scientific purposes [25].

Furthermore, the disclosure of **government data**, so-called Open Government Data (OGD), has also become an important topic in OD worldwide [26–28]. OGD is recognized for its transformative potential [28], and governments across the globe are striving to establish OGD ecosystems that are expected to deliver substantial cultural and institutional benefits [26]. Integrating the extended use of data into political decision-making processes is particularly emphasized at the local level [29]. Many OGD initiatives stem from inclusive governance philosophies that promote citizen participation, positioning them as co-producers with access to official information [30]. Moreover, the publication of government data can foster service innovation and enhance the transparency of public authorities [31]. While public authorities already release significant volumes of data, the analysis and utilization of these data present considerable untapped potential [28].

In the **economic sector**, there is broad consensus across industries that big data and OD will play a pivotal role in the future, necessitating the development of employee capacities and expertise in these areas [26,32]. Specifically, the healthcare and pharmaceutical industries view data analysis as a promising avenue for securing or enhancing competitive advantages [32]. Furthermore, studies have demonstrated a positive correlation between OD and economic growth [26]. The continued reuse and processing of data are therefore regarded as critical drivers of future economic development and value creation [27]. Facilitating big data analyses is considered a collective responsibility of all stakeholders [6]. However, achieving this requires the development of numerous supporting tools and frameworks, many of which are currently unavailable [6].

The EAsyAnon research project (Recommendation and Audit System for the Anonymization of Data) addresses this issue, highlighting the dual challenge faced by many industries: while OD offers substantial benefits and added value, there is a notable lack of effective tools for the anonymization of PD. Funded by the German Federal Ministry of Education and Research (BMBF) and the European Union's Next Generation Programme, this initiative is being developed at the Deggendorf Institute of Technology. The project

aims to provide a secure and user-friendly solution for anonymizing PD while maximizing its utility. The proposed system comprises three components: an intelligent recommendation system that suggests appropriate anonymization techniques for specific data sets while considering legal and ethical implications; an audit system that evaluates the risk of data re-identification; and a trust service that transparently certifies the confidentiality of the anonymization process.

3. Methods

3.1. Research Questions and Objectives

The present qualitative results address the following research questions:

1. What types and forms of personal data under the GDPR are collected, processed and stored by healthcare organizations, research institutions, public authorities, and companies in Germany?
2. What experience has been gained so far by healthcare institutions, research institutions, authorities, and companies in anonymizing the personal data they collect voluntarily and publishing it as OD?
3. What barriers and promoting factors do healthcare institutions, research institutions, authorities, and companies mention to anonymize the PD they collect and subsequently publish it as OD?
4. What necessary support services are mentioned by healthcare institutions, research institutions, authorities and companies in order for them to anonymize the PD they collect voluntarily and publish it as OD?
5. What specific legal and ethical aspects do healthcare institutions, research institutions, public authorities and companies consider necessary to anonymize the PD they collect voluntarily and publish it as OD?

The aim is to develop recommendations for the further establishment and dissemination of PD anonymization and its further use as OD.

3.2. Research Design

The EAsyAnon project employs a mixed-methods approach (Figure 1), integrating qualitative and quantitative research methodologies [33,34]. The research questions are addressed through the triangulation of a systematic literature analysis, a qualitative interview study, and a final quantitative questionnaire survey. This approach facilitates a comprehensive investigation of the research object by incorporating diverse perspectives and methodological frameworks [34,35]. Within this mixed-methods framework, an exploratory sequential design was adopted. In this design, the two sub-studies are conducted sequentially, with the results of the first sub-study informing and influencing the subsequent survey [35].

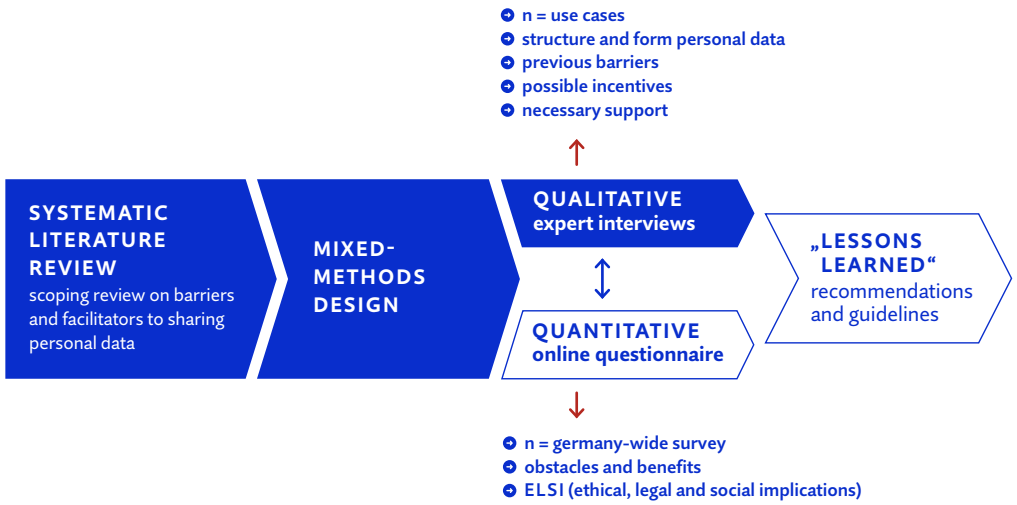

Figure 1. Empirical research process (own illustration). The blue-filled points indicate completed research steps. The systematic review has been previously published in [12]. This article presents the qualitative study, which forms the basis for a subsequent quantitative survey. Finally, the lessons learned and recommendations for action will be derived.

3.3. Method of Data Collection

For the qualitative sub-study, episodic interviews were selected as the method of data collection. This approach generates open, narrative accounts of participants' experiences while enabling the definition and correlation of key terms [36]. The interviews were conducted using a structured guideline developed based on the literature research and a systematic approach to questionnaire creation [37]. A cognitive pretest of the interview questions was conducted beforehand to assess the linguistic clarity and content comprehension of the items, ensure their relevance to the underlying research questions, and identify any redundancies [38]. Five pretests were carried out with relevant experts, after which the guidelines were revised accordingly. The finalized interview guide is provided in the Supplemental Material.

3.4. Sample and Implementation

After defining the inclusion criteria for the interviews (Table 1), experts were recruited through a sampling process based on these predefined criteria. Relevant networks focusing on anonymization and open data were contacted within the research group. Additionally, snowball sampling was employed, with interviewees asked to identify other experts with relevant experience at the end of their interviews.

Table 1. Inclusion criteria sample ($n = 19$).

Criteria for Participation in the Interviews
Age of majority
Capacity to consent
Sufficient knowledge of German to be able to conduct interviews in German, both in written and spoken form
Anyone with relevant experience in open data and anonymization, particularly data protection officers and other experts, could take part
Position in the institution associated with, e.g., open data/anonymization or processing of data/data protection/data management/data administration (DPO, ISB, data administration, etc.)

Comprehensive clarification was provided through informed consent procedures [39], and a reflection period was observed prior to participation. Recommendations for conducting guided expert interviews were followed, with all interviewers receiving prior training [40]. Data collection spanned from October 2023 to January 2024, and only participants based in Germany were interviewed. The entire data collection process was managed by the project partner PASSION4IT due to their thematic proximity and shared expertise in the research subject. Before conducting the interviews, all involved personnel were extensively trained on the interview process, covering the preliminary discussion, introductions, interview structure, specialized techniques, and potential challenges. PASSION4IT coordinated with the recruited participants to schedule online appointments. The interviewer and interviewees were unfamiliar with one another prior to the study. All interviews were conducted online, with the audio recorded on a dictation device and subsequently transcribed using the F4 software in accordance with established transcription rules [41]. No interviews were canceled or prematurely terminated. The interview transcripts were not reviewed by the participants. Following the principle of theoretical saturation [41], interviews were conducted until data redundancy was achieved so that it can be assumed that a certain generalizability of the statements is given.

3.5. Ethics and Data Protection

To proactively address ethical and data protection considerations for the research project [42], an application was submitted to the Joint Ethics Committee of Bavarian Universities (GEHBa) prior to the survey. The application received approval (No. GEHBa-202309-V-124). A comprehensive data protection concept, aligned with legal requirements and principles of good scientific practice, was presented to the ethics committee. In addition, written data protection procedures were developed in collaboration with the responsible data protection officers.

3.6. Qualitative Data Analysis

The interview data were analyzed using the software MAXQDA 2022, following the inductive–deductive approach to structuring content analysis [43]. Initially, deductive main and subcategories were derived from the interview guide during textual pre-analysis. These categories were further refined and supplemented through inductive coding based on an initial analysis of six interviews. Subsequently, the revised category system was applied to all interviews [43]. To ensure intersubjective comprehensibility, selected data passages were collaboratively analyzed by multiple project participants [44]. The category system was further revised following these collaborative evaluations.

3.7. Quality Criteria

The quality of the research process was guided by four core criteria in qualitative research: trustworthiness, transferability, reliability, and confirmability [45]. The trustworthiness of the results is based on the data collected and is achieved by making the data available in an open repository (see Data Availability Statement). The transferability of the results was made clear by the heterogeneous sample and the existing redundancies in the interviews, regardless of the specific institution or organization. The transparent procedure ensured the reliability of the qualitative work according to the COREQ criteria [46] and the confirmability of the results was achieved by the involvement of several researchers in the data analysis and the implementation of internal validations of the category system. Additionally, characteristics such as the appropriateness of the subject, empirical saturation, theoretical depth, textual performance, and originality were carefully considered [47]. Key elements of qualitative empiricism—including comprehensibility, methodological transparency, empirical anchoring, study limitations, reflexive subjec-

tivity, interpretive coherence, and relevance—were integrated throughout the research process [45]. International guidelines for qualitative research, including COREQ [46], were followed to ensure methodological rigor and adherence to best practices.

4. Results

4.1. Socio-Demographic Data

A total of 19 expert interviews were conducted, with 79% of the participants being male ($n = 15$) and 21 % female ($n = 4$). The majority of interviewees were from the business sector and identified themselves as representing companies ($n = 6$). Additionally, participants were drawn from the research and healthcare sectors ($n = 5$ each) and from government organizations and authorities ($n = 3$) (see Figure 2).

The interviewees were based in various federal states, including Bavaria ($n = 11$), Berlin ($n = 3$), Baden-Württemberg ($n = 2$), Saxony ($n = 1$), Lower Saxony ($n = 1$), and North Rhine-Westphalia ($n = 1$). The average age of respondents was 40.7 years (range: 28 years; median: 41 years), and their average professional experience in data management was 15.2 years (range: 36 years; median: 14 years). The interviews had an average duration of 30.11 min (range: 20 min; median: 32 min). In total, 572 min of audio material were collected, all of which were fully included in the analysis.

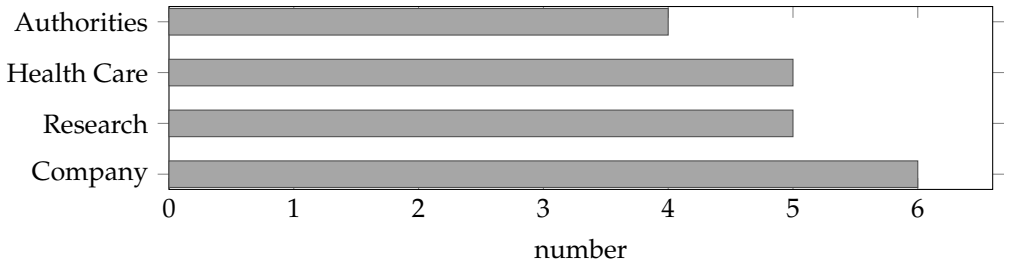

Figure 2. Origin of the interviewees ($n = 19$).

4.2. Content Analysis

Using the questions from the structured interview guide (Supplemental Material), the initial deductive main categories were developed a priori [43]. The entire interview material was coded into these deductive categories. In a subsequent analysis, all codes were inductively refined, leading to the creation of corresponding subcategories as well as definitions for the categories and codes. Three researchers participated in the coding and analysis process, fostering extensive discussions and exchanges of information. Ultimately, five main categories and 21 subcategories were established (Table 2). The interviewees did not evaluate the results.

Table 2. Main and subcategories of the interview analysis.

Main Category	Subcategory
Types and forms of PD	Types of PD in data collection
	Forms of data collection
	Data origin
	Purpose of data collection
	Structure of the data
	Format of the data
	Data processing

Table 2. *Cont.*

Main Category	Subcategory
Experience with anonymization and OD	Understanding OD Understanding PD Assessment of the sensitivity of PD Experience with OD Experience with anonymization
Barriers and facilitators	Promotion factors Obstacles
Support services	Systemic support Personnel support Technical support
Ethical and legal implications	Ethical aspects Legal aspects Institutions

4.3. Types and Forms of Personal Data

All the institutions and companies surveyed stored PD, such as customer data, consumption data, participant data, research data, protocols, or even sensitive health data and socio-demographic data of customers and employees. Depending on the sector and focus of activities, there was enormous heterogeneity in the scope and level of detail of existing PD. A quantitative comparison was not the objective; the focus was on capturing the diversity of data types and formats to inform the design of a broadly applicable anonymization tool.

Furthermore, there was great diversity in the form of data collection, with a combination of handwritten and electronic data collection usually taking place and purely electronic collection of PD hardly being established (*"You cannot do it completely without paper yet"*; transcription IP4, pos. 37). In some cases, handwritten data were subsequently digitized in a resource-intensive process.

As a rule, the PD in health, companies, and authorities were constantly collected by the person responsible for the process. Particularly in the area of research, and to some extent in the area of health, several people often collected the data.

The purposes of data collection were also very different and industry-specific. PD was frequently used for internal purposes and statistics due to legal obligations or for external services, as well as for research, training, and marketing, where marketing purposes, in particular, were seen as very critical in the sample. In some cases, too much PD was also collected in the opinion of the interviewees (*"Yes, much data is collected, and nobody looks at it anymore"*; transcription IP4, pos. 129). Other uses of PD included passing it on to customers, registers or archives following specific agreements, and for training purposes or quality assurance. In some cases, PD was even passed on in non-anonymized form or used for purposes that were not originally intended (*"They are not only processed for the respective purposes"*; transcription IP18, item 57, and *"we would like to work with it, we know that we are not allowed to do so, we are now in a grey area, but we also do not want to address anyone, because if we address them, we are not allowed to do anything with the data"*; transcription IP17, item 119). The involvement of ethics committees was only mentioned in the area of research.

Furthermore, many specific data formats were found across all sectors, particularly in the healthcare system with the specific Hospital Information System (HIS), and for image and audio formats. Microsoft Office formats, CSV, and PDF formats were named as generally standard in the sample. The underlying data structure of the raw data was often described as unstructured, with a lot of handwritten data, which was processed in a resource-intensive manner (*"Yes, semi-structured, so we have different sources, we have interview*

data [...] transaction data [...] ultimately it is always manual work to bring them together"; transcription IP19, item 48). Structured data, often collected by machine, were scarce and often due to legal requirements.

In addition to the many data formats, many software systems were reported in data processing, primarily from Microsoft and SAP. However, internal data management systems and special software for machines, statistics, and internal purposes were also used. In addition to internal infrastructure, cloud solutions were also used for data processing. In some cases, the software could only be used via access regulations and in compliance with security concepts (*"the data protection concept always makes it clear who the data processing centres are"*; transcription IP19, item 62).

Furthermore, the data were processed, stored, and deleted for the specific purposes in compliance with internal deletion concepts and legal requirements, although there were isolated reports of incomplete data cleansing (*"So you always have a project start and a project end, [...] okay, which data have we collected, which can we delete now, which not, and that is often the case, that it takes place with a bit of a delay[...], and I find that extremely difficult [...] to do all this management, from where the data is now"*; transcription IP17, pos. 89). In some cases, limitations in data processing became apparent due to the poor quality of the raw data. The data were digitally and physically stored, mainly at the data collector's premises and on other internal employee devices or external structures (cloud and data center). External storage service providers, in particular, were presented to support small institutions in compliance with specifications and standards.

4.4. Experiences with Open Data and Anonymization

The understanding of the term OD was very heterogeneous, even within the sectors. Depending on the institution and company, open access or certain restrictions on use were associated with it (*"I know open data from the context that you want to make government data or research data [...] generally accessible, often with the idea that this is taxpayer-funded data and that it should flow back somehow"*; transcription IP1, pos. 9). Aspects of the anonymity of PD, consent procedures, and data protection were also discussed differently. While open access was formulated in research and public authorities, the focus in business and the healthcare system was on restricted access. OD was generally seen as an opportunity for digital participation.

In contrast, there was a homogeneous and GDPR-compliant understanding of PD, with only very isolated conceptual difficulties. All characteristics that make a person identifiable and enable statements to be made at an individual level were considered personal data. There was a high level of awareness of PD, particularly in health and research, although cooperation with ethics committees was reported only in the area of research.

When PD was published, it was typically in aggregated form. However, in some cases, data aggregation was rejected due to concerns about reduced data quality or was only feasible to a limited extent.

It also became clear that the sample had little experience in the use or provision of OD from PD while being highly interested in the topic. The few people with a background of experience with OD were predominantly from the field of research. Here, positive experiences with new perspectives through OD were described, as well as a high willingness to provide data and a high demand for OD, as well as to increase the quality and transparency of research. Negative experiences with OD processes across all areas involved included a lack of recognizable benefits, insufficient data quality, and unresolved ethical and legal issues. Internal guidelines, external obligations of journals and clients, and legal regulations were described as reasons for the provision and anonymization of PD.

Overall, the sample had little experience with processes for anonymizing PD. Anonymization concepts were often unknown or the data type was not considered suitable. In some cases, an imprecise separation of pseudonymization and anonymization was evident.

In the few known cases of anonymization of PD, anonymization was mainly carried out using aggregation. However, limitations were also discussed here (*"Aggregations can be made, but in some places it is not possible because the exact statement is important, for example in the case of an indication it is important that it is not just an abnormality in the neurological area, but that it is really the indication multiple sclerosis or Alzheimer's"*; transcription IP2, item 57). The anonymization of PD was described both manually and using software. In addition to managers, specialists from data protection and law were primarily involved in the anonymization process, as were some cooperation partners, customers, and external service providers. In addition, it was occasionally described that non-anonymized PD was passed on to clients or customers based on contracts and obligations.

If people had no experience with anonymization and OD, this was often justified by the fact that OD was not relevant for their area, that the data was not suitable due to its high sensitivity, or that economic disadvantages would result from competitive situations (*"because it makes us vulnerable [. . .], we are very interested in these open data scenarios; I think they are absolutely right, but our data is simply excluded due to its structure and content"*; transcription IP19, pos. 70).

4.5. Barriers and Support Factors

Several barriers and facilitating factors were mentioned in the interviews that can occur during the collection, anonymization, and publication of PD.

Barriers to further use of PD and OD were described in terms of economic, personal, technical, data, legal, and ethical aspects.

Economic aspects, in particular, were seen as a hurdle for OD, with possible costs in material resources and personnel, safeguarding, and liability for anonymity in the case of commercial use being mentioned. The companies also mentioned the unwanted transfer of knowledge through OD as an obstacle, which could mean increasing economic competition and losing innovation.

Personnel barriers were also identified due to a lack of interest and sensitivity to the topic of OD, with a lack of trust in anonymization techniques also becoming apparent.

Furthermore, inherent data aspects could represent a hurdle for OD, for example, if data were classified as too sensitive, content and structure were considered unsuitable, or there were uncertainties regarding possible data manipulation. In addition, some respondents reported a lack of overview of their data situation and few conceivable uses for OD.

A lack of technical infrastructure and anonymization techniques perceived as inadequate were described as technical hurdles, as well as a general lack of specialist staff and expertise on anonymization and OD.

Challenges were also described on unclear and unknown legal bases, particularly about processes for publishing PD and consent procedures. In some cases, PD was also seen as the personal property of the person collecting the data. In addition, legal obligations can have a negative effect on the use of anonymized data and are partly favored by the federal legal structure in Germany (*"In this area, federalism is annoying because every federal state has its own data protection officer, its own data protection law, and now in the healthcare sector we still have the state hospital laws"*; transcription IP24, pos. 77).

Ethical hurdles were seen in OD's irreversibility and lack of control options. Further obstacles included a lack of transparency about what happens to the data, possible action against the interests of the data donor, and the presentation of undesirable results.

In addition, many uncertainties regarding liability and responsibility for potential damage in the event of cyberattacks, re-identification or data merging became clear. A lack of ethical guidelines and structural problems on the part of the state and authorities were cited as obstacles to innovation for institutions and companies in the OD spectrum. At the same time, the interviewees addressed an unfair reciprocity principle of OD, as organizations that do not donate OD themselves can use the OD of others. It was also discussed that although there have already been many positive political commitments to OD, these have not yet led to binding regulations.

Economic, personnel, technical, institutional, ethical, and legal aspects and a positive expectation of the outcome were described as **promoting factors**.

Economic incentives from the state, support for the publication of OD, and sufficient resources available in the institution or company were considered beneficial. Incentive systems for data donors were also described as beneficial.

In the area of personal aspects, positive experiences and attitudes towards OD, especially about recognizing the benefits and willingness to donate data, were considered beneficial (*"I would say [. . .] knowledge is the only good that increases when you share it, I would also make use of that with open data"*; transcription IP25, pos. 68). In the area of research in particular, an increased reach of data was also associated with an increase in reputation.

In technical terms, an existing infrastructure, the high usability of existing software, and the data collection and structure standards were beneficial. The use of special techniques that facilitate anonymization (e.g., differential privacy) was also described as beneficial.

In institutional support factors, an active and strategic decision on OD processes at the management level, with a vision for the publication and usability of OD, supplemented by internal support about implementation and equipment.

Ethically conducive factors included regulated accessibility and usability of OD, transparency at all levels, and proactive reporting of findings to data donors.

A clear legal framework regarding the implementation of anonymization, access to data, the transfer of liability issues, and commercial protection of innovations were also described as beneficial (*"Especially for a medium-sized company, it can be an issue that funding might be something because then you have to build up infrastructure first [. . .] Legal certainty is the most important thing [. . .] otherwise nobody would do it"*; transcription IP1, pos. 97). Liability in particular was highlighted as important across all sectors and should be transferable externally.

The expected positive outcome of OD and existing best practice was considered particularly beneficial across the entire sample.

For example, OD could promote overarching social goals at a macro level, such as through innovations and increased research, which are expected through OD, or improve internal institutional processes and reduce bureaucracy. Billing data, movement data, and health data in particular were seen as having great potential.

4.6. Support Services

In addition, questions were asked about possible and necessary support services for the further establishment of OD.

State support was requested through the establishment of standards, particularly in the processing of health data, as well as through the expansion of guidelines and checklists and the presentation of best practice examples.

Financial resources for the process as well as the establishment of an infrastructure and the assumption of legal liability risk, for example in the form of legal expenses insurance, were also described as necessary and important support services (*"especially for a medium-sized company, it can be an issue that funding may be something, because you have to build up infrastructure first"*; transcription IP1, pos. 97).

Furthermore, personnel support from specialists and departments for anonymization was considered important (*"So an employee position that takes over and then takes care of it, I think that would be pretty good [. . .] a specific human contact person, and if this person then uses software for this, then that is fine too, but I would always prefer to correspond with a person"*; transcription IP23, pos. 151–153), whereby support from external service providers was also mentioned.

Furthermore, technical support in the form of software was requested above all. On the one hand, this should carry out anonymization securely and following the guidelines and at the same time provide information on OD aspects and have maximum interoperability (*"so definitely better via software, so the process that is currently being used is, of course, very labour-intensive for people, and what I think is the biggest obstacle for us is that there should be software where personal data can be entered"* Transcription IP2, pos. 93). An open-source solution that allows specific settings was discussed here. In addition, data processing from OD portals was requested to facilitate research.

4.7. Ethical and Legal Implications

In addition, questions were asked about specific ethical and legal aspects in the participating sectors that play a role in the voluntary anonymization of PD and its publication as OD.

In ethics, the high importance of individual voluntariness in participating in OD was considered a priority. In addition, data protection principles in accordance with Art. 5 of the GDPR, such as data minimization or ensuring data integrity, should be fully considered. OD can also avoid unreasonable duplication of data collection and the associated duplication of resources and burdens for individuals.

The ethical aspects of technical progress were also discussed, which requires a broad social consensus, as risks always accompany progressive developments. The dangers posed by OD were described across all sectors, for example, through stigmatization or inherent potential for abuse. The great importance of trust and transparency at all process levels and, above all, to reduce fears among data donors was emphasized, as was a broad social awareness of OD and the existence of an ethical code for OD.

In order to build trust, the anonymization process should be carried out with the greatest possible transparency regarding methods and results. As a residual risk of re-identification will always remain—especially given future technological developments—it is essential to implement anonymization processes with maximum transparency regarding methods and limitations. Only in this way can the remaining risk be minimized. At the same time, some level of risk acceptance is necessary to enable progress. If you want progress, you must take a particular risk (*"I always think that if you want medicine to be advanced [. . .] and if you have nothing better to do than spend all day trying to make sure that your data cannot be decrypted, then you are wrong because any anonymised data can also be decrypted with the right tools"*; transcription IP24, pos. 81).

OD was also presented as an instrument of power, and it was explained that OD should, as far as possible, benefit society as a whole and not be used for purely economic purposes, which is why ethical issues and objectives must be consistently taken into account when using OD. An ethical code including defined access and usage restrictions for OD was described as important, as insecure anonymization poses a high potential

for abuse, especially in the case of critical infrastructure. Companies emphasized that OD enables digital participation in the data ecosystem and that monopoly positions of large corporations could be avoided, which should be in society's interest. In addition, the German mentality in particular was characterized as very cautious, and that complete security will never be possible with OD. In addition, financial burdens for the solidarity community were described as a result of the establishment and use of OD, such as license fees for software or necessary certifications.

Regarding specific legal aspects, the federal German structure, with many additional data protection laws at the state level, was described as complicated. In particular, federal data protection regulations were perceived as a hindrance in the healthcare sector. In addition, there are sometimes conflicting regulations, for example, when a general right to data erasure and the legal obligation to provide data come together. Furthermore, Europe-wide regulatory provisions on the use of AI (EU AI Act) were seen as a way of preventing the threat of misuse and the risk of re-identification. Possible risks due to liability issues should be externalized in OD processes, and the software should assume potential damages from external service providers or the state.

On a personal level, education and self-determination for data release, personal responsibility for self-protection among data donors, consent to OD, and the clarification of ownership claims were emphasized as important legal foundations. The use and integrity of data trustees were also discussed in the areas of health and research.

Furthermore, many personal uncertainties regarding legal understanding and anonymization became clear. Across all sectors, the sample revealed great uncertainty as to whether their anonymization processes are legally compliant and, at the same time, little legal knowledge was reported.

From a legal and technical perspective, it was emphasized for the anonymization processes that the raw data should not leave the place of origin. Additionally, it was noted that small data sets can pose specific re-identification risks, especially when they contain unique or rare attribute combinations. Smaller data sets often require higher levels of generalization to ensure privacy, which in turn can reduce the utility and interpretability of the data. In addition, regular technical checks were called for without creating additional bureaucracy, which are based on EU law and include checks using the latest anonymization techniques. A renewed security check of OD was seen as resource-intensive and technically difficult across all sectors, especially due to future technical developments. In this context, a certain period of validity of anonymization and OD was also addressed, whereby contradictions became clear, as OD in circulation was considered no longer controllable. Synthetic data sets in certain areas and an increased use of differential privacy would therefore be more suitable.

In addition, state certifications and seals based on known DIN ISO standards and procedures were considered important across the entire sample to demonstrate quality standards in processing and control access to OD. At the same time, the demand for a clear overview of certifications and seals was emphasized. In addition, the use of digital identity in Germany for OD purposes was described.

Regarding the institutions responsible for checking anonymization, both the state and companies were named as suitable. Verification by neutral bodies was described as confidence-building. The state was often described as more trustworthy and reliable in the areas of health, research, and authorities, as there is no profit motive from OD. In the case of a review of anonymization by a state-commissioned body, it was stated as positive that responsibilities and procedures would be better known and that the profit motive of potential data users could be monitored more closely.

However, the companies surveyed argued that state institutions are less competent and less flexible compared to the digital agility of neutral bodies or companies (*"Well, if something should going ahead, then it should not be the state"*; transcription IP22, pos. 63). For this reason, companies were sometimes seen as more suitable, as they have more competencies and aspects relating to international liability.

5. Discussion

In this discussion, the results are reflected concerning the research questions.

Research Question 1: The first question addressed the types and forms of PD collected, processed, and stored by the groups involved in Germany, in accordance with the GDPR requirements.

The analysis revealed a considerable heterogeneity of existing PD in all sectors, specifically depending on the activities and processes of the individual sectors. Synonymously, a significant variation in the current data structure was revealed, whereby a combination of handwritten and electronically collected, semi-structured data was predominantly reported across all sectors. Even within individual sectors, a wide range of structures and formats of PD were indicated, especially in the research and health sectors. Furthermore, a wide range of uses became apparent, primarily for internal purposes or external obligations, whereby legal omissions in the disclosure of PD also became apparent. The data were processed using many different software systems, particularly in the healthcare sector, with different software being used for PD in all sectors. Internal infrastructures and cloud systems were predominantly used for processing and storing data, although fewer cloud services were used in the sample, particularly in the healthcare sector. Internal security and deletion concepts for handling data were reported in some cases, but not in a sector-specific manner. For smaller companies in particular, an advantage in data management was mentioned in cooperation with external service providers.

Research Question 2: Here, the experiences of the sectors involved in anonymizing PD and publishing it as OD were to be investigated.

While individual experiences with anonymization techniques for PD and publication as OD were predominantly reported in the research sector, there was little or hardly any previous experience in health, public authorities, and companies. At the same time, there was a lack of knowledge regarding anonymization concepts and clarity regarding the terms anonymization and pseudonymization, which was often evident among the interviewees. If PD was anonymized, it was usually performed with aggregation, both physically and with software support. Internal specialists from data protection and law were predominantly involved, but external service providers were also mentioned. Furthermore, the understanding of OD, anonymization processes, and consent processes was unclear and indifferent among many respondents. In health and companies, OD tended to be discussed with access restrictions in the sample. In contrast, in the areas of research and public authorities, open access was understood. Concerning the assessment of PD, there was a GDPR-compliant understanding, as was the assessment of the sensitivity of PD, especially in health and research. In addition, it became clear that, except in research, almost no experience in the use of OD was reported in the sample. Although the sample indicated a great interest in OD, there was often a lack of specific and personal recognition of the benefits of using OD, or there were clear uncertainties regarding possible disadvantages or the existing data quality.

Research Question 3: In the sample, possible barriers and facilitators regarding the anonymization of PD and its subsequent publication as OD were to be identified.

The barriers identified were described in terms of economic, personal, technical, data, legal, and ethical aspects. Using money and personnel resources for the necessary infras-

tructure was identified as a significant barrier across all sectors. Possible compensation payments in the event of liability claims were also a barrier, with the benefits of OD being unclear in some cases. In addition, fears of a possible loss of innovation due to OD were reported, particularly in the corporate sector. A lack of interest, a lack of expertise in anonymization techniques, and a lack of personnel expertise about OD were described as personnel barriers across the sample. In addition, challenges regarding the necessary technical infrastructure and the underlying data quality were identified across all sectors. Further challenges arose due to unknown or unclear legal bases and Germany's federal data protection law structure. Ethical hurdles were formulated primarily due to the irreversibility of OD, a lack of control options, possible misuse, and future re-identification through technical possibilities. In addition, a lack of ethical guidelines and insufficient political efforts regarding OD were discussed across all sectors. Economic, personnel, technical, institutional, ethical, and legal aspects, as well as a positive expectation of the outcome, were described as support factors. Economic incentive systems of the state for the establishment of OD and data donation were described as important support factors in the entire sample. Previous personal experience and a positive attitude towards OD were also identified as important personal support factors. In addition, an existing infrastructure and high usability of the systems used were described as supportive, as were established OD portals and platforms that support data discovery. Furthermore, institutional focus and support at all levels across all sectors were important factors in establishing OD processes. Ethical clarity and transparency regarding access and use, and a clear legal framework that regulates anonymization and protects innovation were also considered beneficial. The regulation of liability issues was described as particularly beneficial across all sectors. In addition, a proven positive outcome and the existence of best-practice examples in using OD generated a conducive ecosystem for further OD use.

Research Question 4: A further question addressed the desired support services so PD can be anonymized voluntarily and published as OD. Above all, government support through further standards and guidelines was formulated, as was the provision of financial resources. In particular, the assumption of liability risk was described as an essential form of support across all sectors. Furthermore, personnel support from specialists with the relevant technical expertise was described as important and included technical support. Interoperable software solutions that provide targeted support for anonymization and OD and, if possible, inclusive OD platforms that facilitate data access were particularly desired as technical support.

Research Question 5: A further question addressed specific legal and ethical implications in the sectors involved, regarding the willingness to anonymize PD and publish it as OD.

Ethical implications arose above all about informed consent, particularly the voluntary nature and clarification of data use in the case of data donation. In addition, an ethical necessity for OD was seen on the one hand to minimize social burdens caused by duplicate surveys and the financial resources required for data collection. OD should benefit society and not be driven by monetary motives. To this end, a comprehensive ethical code was called for, especially in the areas of health and research, which enables access and use as well as comprehensive security, while at the same time ensuring the greatest possible digital participation for all. On the other hand, the potential for misuse of OD was identified across all sectors, as progressive technical developments can always go hand in hand with as-yet unforeseeable risks in the future. The authorities, in particular, saw far-reaching dangers here with critical infrastructure. Furthermore, there were cross-industry fears of financial burdens due to license fees and certification costs, which would ultimately have to be borne by the community. Therefore, the consensus of the sample was that a focus should

be placed on confidence-building and transparent processes regarding anonymization and OD to convince data donors.

Legal implications arose throughout the sample due to the federal German legal structure with many additional data protection rules to be observed at the state level, particularly in the healthcare system. In addition, authorities recognized contradictory legal bases when data were to be published on the one hand and an individual right to data deletion was guaranteed on the other. Liability clarification was presented as a priority task to be solved across all sectors, with solutions being hoped for in European regulations such as the AI Act. Furthermore, personal self-determination for data release and the legal basis must remain fully protected. Data trustees from the research and healthcare sectors were also called for. Significant legal uncertainties were identified across all sectors when evaluating OD processes and anonymization. Furthermore, a legal requirement for regular technical control and protection measures was discussed in the sample, whereby companies in particular emphasized that no additional bureaucracy should be created. The process of complying with standards and controls should be as resource-efficient as possible.

In some cases, synthetic data sets and differential privacy were also described as a solution to potential data protection conflicts. Furthermore, government certifications and seals were considered beneficial across all sectors to prove OD's quality, security, and anonymization. An independent institution was generally preferred as the verifying institution, with the health and research sectors considering state control to be more important and companies in favor of private sector responsibility, as they were considered to be more agile than the state.

6. Conclusions

Based on the literature analysis and the evaluation of the interviews, the following guiding principles can be derived from the research questions. The planned recommendations for action will be formulated after the quantitative data collection is completed and an initial implementation of the *EAsyAnon* system in practice using realistic use cases.

The present qualitative survey showed a high variability of existing PD and the associated data structures and processing systems in the sectors involved in the sample. This is already known from other international studies and is therefore not a unique German challenge [6,12,48]. Therefore, the intention to establish OD processes worldwide requires further harmonizing file systems and software. This is seen as an important step in tackling the technically challenging diversity of files. A consistent and politically and institutionally supported establishment of the FAIR principles (findable, accessible, interoperable, reusable) in the steps of data collection and processing therefore seems essential, which has also already been postulated by international studies [12,17].

In addition, access to software should be made as inclusive as possible, and it should ideally be designed as an open-source system to enable participation in OD worldwide.

Furthermore, it became clear in the sample that there is hardly any experience with anonymization techniques in Germany and that OD has hardly been used to date, with a few exceptions in the research sector. This raises eyebrows, as the great importance of OD processes for future value creation has been emphasized politically worldwide for years. At the same time, the political commitments to support and promote OD do not seem to have reached a broad audience. The present German sample confirms a global lack of expertise on the topic [49]. Especially for a large, industrialized country like Germany, which has the third largest economy in the world, OD offers great potential for launching innovations, improving services at all levels, and thus securing future value creation and the associated social prosperity. In addition, the survey revealed many uncertainties regarding OD and

the necessary anonymization, which requires a clear need for further education and training and integration of knowledge on the topic into the curricula of training and studies.

It can be assumed that knowledge about data protection and integrity and the ethically responsible handling of data will continue to grow in the future. Such awareness-raising towards a data-oriented culture has long been called for [9,26,50].

The national barriers and support factors identified in this study are in line with international challenges [12]. They can form the basis for concrete approaches to action at a personal, institutional, and societal policy level. Ways must be found to secure the necessary resources in all sectors and create motivation to participate in OD. This includes concrete solutions for the as-yet-unresolved important issue of liability for potential damage caused by OD and comprehensive protective measures for individuals and institutions involved in OD developments to establish trust and security across the board. Targeted incentive systems are needed to motivate the anonymization of PD and subsequent data donation on the one hand, and a clear, supportive ethical and legal framework that provides the necessary security for all those involved on the other.

The federal structure in Germany is particularly challenging in terms of data protection, as the GDPR contains numerous opening clauses for national regulations. The national Federal Data Protection Act (BDSG) takes these up and contains opening clauses for the legislation of the federal states. This means that different legal requirements may exist in individual federal states, such as hospital data protection regulations. This heterogeneous legal situation makes uniformly implementing data protection measures considerably more difficult. In addition, significant uncertainties regarding legal responsibility and liability in the event of data protection breaches were discussed. Concerns about liability for data protection breaches are not unfounded, as even minor breaches can lead to significant compensation obligations for many affected persons. Companies with many customers, in particular, face potentially far-reaching consequences. Ideally, such regulations should not only apply nationally or throughout Europe, but should also be valid and implemented worldwide, which would do justice to OD's global usage requirements.

Directive (EU) 2019/1024 on OD and the re-use of public sector information has so far only obliged public bodies in the EU to make sure data sets are available as OD, which serves the further re-use of public data and thus creates a level playing field for companies. At the national level, this has been implemented through the Data Use Act, among other things, although far-reaching international regulations are required. Furthermore, support must serve the goal of anchoring OD in society. This also includes individual solutions and services, and an inclusive design of OD platforms to ensure digital participation in the data ecosystem for all interested parties.

This study clarified that discussing and resolving previously unresolved ethical and legal implications represents a key moment for disseminating OD at all levels. Possible benefits and risks need to be discussed at a broad level and brought to a final consensus so that the challenges can be tackled on a secure compromise basis and the use of OD can be further developed.

Consideration should be given here to neutral and trustworthy institutions that can perform a control and security function in the OD ecosystem. There is an imminent potential for abuse, especially due to unforeseeable technical developments in the context of the strong technological acceleration in artificial intelligence and all the associated innovations. Nevertheless, the great potential of OD must not play a lesser role in the argumentation here. The concrete presentation of successes and innovations from OD processes that provide benefits or make things easier for everyone can specifically support the acceptance and understanding of OD developments in society at large.

7. Limitation

This study has several limitations that should be considered when interpreting the results.

Due to the small sample size, the generalizability of the categories and results is limited, even if they overlap with international results. The qualitative interview survey conducted as part of the project is not representative. The aim was not to develop a phenotype, but to prepare for the implementation of a more extensive quantitative survey with a larger sample in line with the chosen mixed-methods approach, which builds the quantitative study on the hypothesis-generating answers from the interviews. The qualitative data collection was carried out until the principle of data saturation was reached and only ended when the interviews revealed many redundancies regarding the research questions across sectors.

There were also challenges in recruiting specialists with the relevant expertise, which does not rule out a selection bias. Due to the use of the snowball method, the recruitment process of some interview participants was not completely transparent for the researchers.

As a result, selection and exit bias cannot be ruled out, especially for people with technical expertise. Experts from four different professional fields were included to increase the range of perspectives and achieve a consensus between the disciplines. The sample may have an attrition bias due to a particular affinity and social desirability on the part of people particularly interested in the interview. However, the high correlation of the present results with other studies tends to attenuate any possible attrition bias. In addition, the sample has a gender bias with a significantly higher proportion of male participants, which may further limit the representativeness of the results.

In addition, significant national and international legislative changes were made after the data collection period, which may impact the development and perception of OD processes. In particular, the European Union's Digital Services Act (DSA), which came into force in all EU member states on 17 February 2024, marks a decisive change in regulating the digital ecosystem. Although the study has considered these new regulatory frameworks as far as possible, their impact may not yet be fully reflected in the analysis.

Furthermore, the interviewees used terms such as synthetic data, differential privacy, and anonymization techniques such as data aggregation, although it is unclear how exactly the interviewees' understanding of the terms corresponds to the actual definitions. In general, it should be noted that little specialist expertise has been presented in the study or in international surveys to date. In future studies, an expansion about the requirements and circumstances of specific institutions and facilities in terms of OD and anonymization should continue to be a fixed component.

Some of the colloquial statements were challenging to translate. Therefore, the translated quotes were linguistically smoothed for readability while preserving the intended meaning.

Supplementary Materials: The anonymized transcripts and the interview guide are available at: http://www.doi.org/10.5281/zenodo.14810988 (accessed on 23 June 2025).

Author Contributions: Conceptualization, N.L. and S.W.; Methodology, N.L., Software, S.W.; Validation, M.K., J.K., N.L., E.M. and J.G.; Formal Analysis, N.L.; Investigation, N.L.; Resources, F.L. and N.L.; Data Curation, F.L.; Writing—Original Draft Preparation, N.L.; Writing—Review and Editing, S.W.; Visualization, S.W.; Supervision, F.W. Project Administration, F.W.; Funding Acquisition, S.W. and F.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the publication were funded by the European Union and sponsored by the German Federal Ministry of Research, Technology and Space under grant number 16KISA128K. ("Verbundprojekt: Empfehlungs- und Auditsystem zur Anonymisierung—EAsyAnon").

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Joint Ethics Committee of Bavarian Universities (GEHBa) (protocol code GEHBa-202309-V-124).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: During the preparation of this manuscript/study, the authors used GPT 4.0 for the purposes of minor text polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Data Availability Statement: The original data presented in the study are openly available in [Zenodo] at: http://www.doi.org/10.5281/zenodo.14810988 (accessed on 23 June 2025).

Conflicts of Interest: Author Matthias Kampmann and Juliane Kis were employed by the company IT-Sicherheitscluster e.V. Author Florian Laumer was employed by the company PASSION4IT GmbH. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

OD	Open data
PD	Personal data
OGD	Open Government Data
GDPR	General Data Protection Regulation

References

1. Ethikrat, D. Big Data und Gesundheit—Datensouveränität als Informationelle Freiheitsgestaltung: Stellungnahme. 2017. Available online: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf (accessed on 5 March 2025).
2. Househ, M.; Grainger, R.; Petersen, C.; Bamidis, P.; Merolli, M. Balancing Between Privacy and Patient Needs for Health Information in the Age of Participatory Health and Social Media: A Scoping Review. *Yearb. Med. Inform.* **2018**, *27*, 029–036. [CrossRef] [PubMed]
3. Viberg Johansson, J.; Bentzen, H.B.; Mascalzoni, D. What ethical approaches are used by scientists when sharing health data? An interview study. *BMC Med. Ethics* **2022**, *23*, 41. [CrossRef] [PubMed]
4. Avraam, D.; Jones, E.; Burton, P. A deterministic approach for protecting privacy in sensitive personal data. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 24. [CrossRef]
5. Kamikubo, R.; Lee, K.; Kacorri, H. Contributing to Accessibility Datasets: Reflections on Sharing Study Data by Blind People. In Proceedings of the CHI '23 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023; ACM: New York, NY, USA, 2023; pp. 1–18. [CrossRef]
6. Rehman, A.; Naz, S.; Razzak, I. Leveraging big data analytics in healthcare enhancement: Trends, challenges and opportunities. *Multimed. Syst.* **2021**, *28*, 1339–1371. [CrossRef]
7. Mahomed, S.; Labuschaigne, M.L. The evolving role of research ethics committees in the era of open data. *S. Afr. J. Bioeth. Law* **2023**, *15*, 80–83. [CrossRef]
8. Feeney, O.; Werner-Felmayer, G.; Siipi, H.; Frischhut, M.; Zullo, S.; Barteczko, U.; Øystein Ursin, L.; Linn, S.; Felzmann, H.; Krajnović, D.; et al. European Electronic Personal Health Records initiatives and vulnerable migrants: A need for greater ethical, legal and social safeguards. *Dev. World Bioeth.* **2019**, *20*, 27–37. [CrossRef]
9. van Donge, W.; Bharosa, N.; Janssen, M.F.W.H.A. Future government data strategies: Data-driven enterprise or data steward?: Exploring definitions and challenges for the government as data enterprise. In Proceedings of the dg.o '20 21st Annual International Conference on Digital Government Research, Seoul, Republic of Korea, 15–19 June 2020; ACM: New York, NY, USA, 2020; pp. 196–204. [CrossRef]
10. Nellåker, C.; Alkuraya, F.S.; Baynam, G.; Bernier, R.A.; Bernier, F.P.; Boulanger, V.; Brudno, M.; Brunner, H.G.; Clayton-Smith, J.; Cogné, B.; et al. Enabling Global Clinical Collaborations on Identifiable Patient Data: The Minerva Initiative. *Front. Genet.* **2019**, *10*, 611. [CrossRef] [PubMed]

11. für Wirtschaft und Klimaschutz, B.B. Das Gaia-X Ökosystem—Bmwk.de. Available online: https://www.bmwk.de/Redaktion/DE/Dossier/gaia-x.html (accessed on 5 March 2025).
12. Lichtenauer, N.; Schmidbauer, L.; Wilhelm, S.; Wahl, F. A Scoping Review on Analysis of the Barriers and Support Factors of Open Data. *Information* **2023**, *15*, 5. [CrossRef]
13. Dos Santos Rocha, A.; Albrecht, E.; El-Boghdadly, K. Open science should be a pleonasm. *Anaesthesia* **2023**, *78*, 551–556. [CrossRef]
14. Eva, G.; Liese, G.; Stephanie, B.; Petr, H.; Leslie, M.; Roel, V.; Martine, V.; Sergi, B.; Mette, H.; Sarah, J.; et al. Position paper on management of personal data in environment and health research in Europe. *Environ. Int.* **2022**, *165*, 107334. [CrossRef]
15. Alzahrani, A.G.; Alhomoud, A.; Wills, G. A Framework of the Critical Factors for Healthcare Providers to Share Data Securely Using Blockchain. *IEEE Access* **2022**, *10*, 41064–41077. [CrossRef]
16. Forschungsdateninfrastruktur, N. Die Nationale Forschungsdateninfrastruktur (NFDI): For a FAIR Data Future. Available online: https://www.nfdi.de/wp-content/uploads/2024/05/NFDI-Verein-Kurzinfo-v10.pdf (accessed on 5 March 2025).
17. Queralt-Rosinach, N.; Kaliyaperumal, R.; Bernabé, C.H.; Long, Q.; Joosten, S.A.; van der Wijk, H.J.; Flikkenschild, E.L.; Burger, K.; Jacobsen, A.; Mons, B.; et al. Applying the FAIR principles to data in a hospital: Challenges and opportunities in a pandemic. *J. Biomed. Semant.* **2022**, *13*, 12. [CrossRef] [PubMed]
18. Schwalbe, N.; Wahl, B.; Song, J.; Lehtimaki, S. Data Sharing and Global Public Health: Defining What We Mean by Data. *Front. Digit. Health* **2020**, *2*, 612339. [CrossRef]
19. Bentzen, H.B.; Castro, R.; Fears, R.; Griffin, G.; ter Meulen, V.; Ursin, G. Remove obstacles to sharing health data with researchers outside of the European Union. *Nat. Med.* **2021**, *27*, 1329–1333. [CrossRef]
20. Horn, R.; Kerasidou, A. Sharing whilst caring: Solidarity and public trust in a data-driven healthcare system. *BMC Med. Ethics* **2020**, *21*, 110. [CrossRef] [PubMed]
21. Fylan, F.; Fylan, B. Co-creating social licence for sharing health and care data. *Int. J. Med. Inform.* **2021**, *149*, 104439. [CrossRef] [PubMed]
22. McQuate, S. Q&A: How to Train AI When You Don't Have Enough Data— Washington.edu. 2024. Available online: https://www.washington.edu/news/2024/03/28/train-ai-machine-learning-when-you-dont-have-enough-data/ (accessed on 5 March 2025).
23. Deist, T.M.; Dankers, F.J.; Ojha, P.; Scott Marshall, M.; Janssen, T.; Faivre-Finn, C.; Masciocchi, C.; Valentini, V.; Wang, J.; Chen, J.; et al. Distributed learning on 20 000+ lung cancer patients—The Personal Health Train. *Radiother. Oncol.* **2020**, *144*, 189–200. [CrossRef]
24. Baker, D.B.; Knoppers, B.M.; Phillips, M.; van Enckevort, D.; Kaufmann, P.; Lochmuller, H.; Taruscio, D. Privacy-Preserving Linkage of Genomic and Clinical Data Sets. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 1342–1348. [CrossRef]
25. Hallock, H.; Marshall, S.E.; 't Hoen, P.A.C.; Nygård, J.F.; Hoorne, B.; Fox, C.; Alagaratnam, S. Federated Networks for Distributed Analysis of Health Data. *Front. Public Health* **2021**, *9*, 712569. [CrossRef]
26. Mutambik, I.; Nikiforova, A.; Almuqrin, A.; Liu, Y.D.; Floos, A.Y.M.; Omar, T. Benefits of Open Government Data Initiatives in Saudi Arabia and Barriers to Their Implementation. *J. Glob. Inf. Manag.* **2022**, *29*, 1–22. [CrossRef]
27. Seo, J.; Kim, B.; Kwon, H.Y. Open Data Policies Analysis Disputes Mediation Cases in Korea: Based on OUR Data Index and ODB. In Proceedings of the DG.O2021: The 22nd Annual International Conference on Digital Government Research, Omaha, NE, USA, 9–11 June 2021; ACM: New York, NY, USA, 2021; pp. 153–167. [CrossRef]
28. Kawashita, I.; Baptista, A.A.; Soares, D. Open Government Data Use by the Public Sector—An Overview of Its Benefits, Barriers, Drivers, and Enablers. 2022. Available online: http://hdl.handle.net/10125/79648 (accessed on 18 June 2025).
29. Rempel, E.; Barnett, J.; Durrant, H. Contrasting views of public engagement on local government data use in the UK. In Proceedings of the ICEGOV2019 12th International Conference on Theory and Practice of Electronic Governance, Melbourne, VIC, Australia, 3–5 April 2019; ACM: New York, NY, USA, 2019; pp. 118–128. [CrossRef]
30. Smith, G.; Sandberg, J. Barriers to innovating with open government data: Exploring experiences across service phases and user types. *Inf. Polity* **2018**, *23*, 249–265. [CrossRef]
31. Crusoe, J.; Melin, U. Investigating open government data barriers: A literature review and conceptualization. In Proceedings of the Electronic Government: 17th IFIP WG 8.5 International Conference, EGOV 2018, Krems, Austria, 3–5 September 2018; Proceedings 17; Springer: Cham, Switzerland, 2018; pp. 169–183.
32. Pesqueira, A.; Sousa, M.J.; Rocha, A. Big Data Skills Sustainable Development in Healthcare and Pharmaceuticals. *J. Med. Syst.* **2020**, *44*, 197. [CrossRef]
33. Levitt, H.M.; Bamberg, M.; Creswell, J.W.; Frost, D.M.; Josselson, R.; Suárez-Orozco, C. Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *Am. Psychol.* **2018**, *73*, 26–46. [CrossRef] [PubMed]
34. Schoonenboom, J. The Fundamental Difference Between Qualitative and Quantitative Data in Mixed-Methods Research. *Forum Qual. Sozialforschung/Forum Qual. Soc. Res.* **2023**, *24*. [CrossRef]
35. Kuckartz, U. *Mixed-Methods: Methodologie, Forschungsdesigns und Analyseverfahren*; Springer: Berlin/Heidelberg, Germany, 2014.

36. Flick, U. Das Episodische Interview. In *Empirische Forschung und Soziale Arbeit*; VS Verlag für Sozialwissenschaften: Wiesbaden, Germany, 2011; pp. 273–280. [CrossRef]
37. Baur, N.; Blasius, J.; Helfferich, C. Leitfaden-und Experteninterviews. In *Handbuch Methoden der empirischen Sozialforschung*; VS Verlag für Sozialwissenschaften: Wiesbaden, Germany, 2014; pp. 557–559.
38. Buschle, C.; Bethmann, A. Kognitives Pretesting. Zenodo. 2017. [CrossRef]
39. Schröder, A.; Proll, L.; In-Albon, T. Informed Consent in Onlinestudien: Wieviel verstehen Teilnehmende wirklich und lässt sich das ändern? *Z. Klin. Psychol. Psychother.* **2023**, *52*, 38–50. [CrossRef]
40. Bogner, A.; Littig, B.; Menz, W. *Interviews Mit Experten: Eine Praxisorientierte Einführung*; Springer: Berlin/Heidelberg, Germany, 2014.
41. Przyborski, A.; Wohlrab-Sahr, M. *Qualitative Sozialforschung: Ein Arbeitsbuch*; De Gruyter: Oldenburg, Germany, 2013.
42. Forschungsgemeinschaft, D. Guidelines for Safeguarding Good Research Practice. Code of Conduct. Available online: https://www.dfg.de/resource/blob/174052/1a235cb138c77e353789263b8730b1df/kodex-gwp-en-data.pdf (accessed on 6 March 2025).
43. Kuckartz, U. *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*; Beltz Juventa: Weinheim, Germany, 2018.
44. Flick, U.; Kardorff, E.v.; Steinke, I. Qualitative Forschung: Ein Handbuch (14. Auflage, Originalausgabe). In *Reinbek bei Hamburg: Rowohlts enzyklopädie im Rowohlt Taschenbuch Verlag*; Rohwolt Taschenbuchverlag: Reinbek bei Hamburg, Germany, 2022 .
45. Doring, J.B.N. *Forschungsmethoden und Evaluation in den Sozial-und Humanwissenschaften*; Springer: Berlin/Heidelberg, Germany, 2016.
46. Tong, A.; Sainsbury, P.; Craig, J. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *Int. J. Qual. Health Care* **2007**, *19*, 349–357. [CrossRef] [PubMed]
47. Strübing, J.; Hirschauer, S.; Ayaß, R.; Krähnke, U.; Scheffer, T. Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. *Z. Soziol.* **2018**, *47*, 83–100. [CrossRef]
48. Broes, S.; Lacombe, D.; Verlinden, M.; Huys, I. Toward a Tiered Model to Share Clinical Trial Data and Samples in Precision Oncology. *Front. Med.* **2018**, *5*, 6. [CrossRef]
49. Sandoval-Almazan, R.; Valle Gonzalez, L.; Millan Vargas, A. Barriers for Open Government Implementation at Municipal Level: The Case of the State of Mexico. In Proceedings of the DG.O2021: The 22nd Annual International Conference on Digital Government Research, Omaha, NE, USA, 9–11 June 2021; ACM: New York, NY, USA, 2021; pp. 113–122. [CrossRef]
50. Dove, G.; Shanley, J.; Matuk, C.; Nov, O. Open Data Intermediaries: Motivations, Barriers and Facilitators to Engagement. *Proc. ACM Hum.-Comput. Interact.* **2023**, *7*, 1–22. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.