applied sciences

MDPI

Article

An Open-Source Software Tool to Facilitate Data Protection Impact Assessments

Layla Tabea Riemann [1,*], Felicia P. S. Hähner [1], Ann-Kathrin Schmitz [1], Maximilian Ataian [1], Matthias Jaster [2] and Frank Ückert [1]

[1] Institute for Applied Medical Informatics, Centrum for Experimental Medicine, University Hospital Hamburg-Eppendorf, 20251 Hamburg, Germany
[2] Deanery of the University Hospital Hamburg-Eppendorf, 20251 Hamburg, Germany
* Correspondence: l.riemann@uke.de; Tel.: +49-(0)40-7410-57491

Featured Application: We present an open-source "plug-and-play" tool that significantly accelerates and facilitates the generation of data protection impact assessment on a local, national, and potentially international scale.

Abstract: In the realm of medical research, preserving patient privacy while facilitating effective research and collaborations poses a significant challenge. Data protection impact assessments (DPIAs) and associated methodologies have emerged as a response to this dual imperative. DPIAs necessitate expertise across diverse domains, resulting in a complex procedural landscape. To address this, we present "DPIA click&go", a user-friendly tool designed to streamline the DPIA process in a plug-and-play manner. This tool enables users to semi-automatically select risks from predefined categories, construct evaluation matrices, access risk-mitigating measures, and re-evaluate risks after the application of mitigation strategies. Tailoring risks and measures to each institute's needs is essential, facilitated by the provided data model, considerably simplifying DPIA creation at an institutional level. The efficacy of the DPIA click&go tool was validated with a real-world project, comparing its performance with a manually created DPIA in terms of risk coverage and mitigation strategies. The promising outcomes of this tool underscore its potential within the national data protection landscape, suggesting its possible foundational role in harmonizing data protection practices at a larger, potentially, European or global, scale.

Keywords: data protection impact assessments (DPIAs); General Data Protection Regulation (GDPR); health care data management; clinical data security support; open source; medical IT infrastructure

Citation: Riemann, L.T.; Hähner, F.P.S.; Schmitz, A.-K.; Ataian, M.; Jaster, M.; Ückert, F. An Open-Source Software Tool to Facilitate Data Protection Impact Assessments. *Appl. Sci.* **2023**, *13*, 11230. https://doi.org/10.3390/app132011230

Academic Editor: Petr Dzurenda

Received: 5 September 2023
Revised: 27 September 2023
Accepted: 5 October 2023
Published: 12 October 2023

1. Introduction

1.1. Challenges in Data Protection and Research

Medical research is increasingly ingesting large, structured data, which is ideal for research due to the proliferation of standardized data models and increasingly powerful processors [1]. This development brings increased attention to dualism at both national and international levels: On one hand, there is a desire to make extensive medical data accessible to researchers and utilize it for research collaborations to improve the quality of medical care. On the other hand, it is crucial to protect sensitive patient data effectively [2]. To ensure that both technically and organizationally everything is performed to safeguard patient data while still enabling research, data protection impact assessments (DPIAs) and related concepts have been established in recent years in Germany. Although data protection has been an important topic in Germany for quite some time now [3], data protection in its technical and organizational implications has gained increased attention in all European countries due to the General Data Protection Regulation (GDPR) which was passed in 2018 and aims to harmonize and standardize European data protection [4].

In that context, more standardized data protection concepts will evolve to harmonize this heterogeneous field [5]. As European legislation's data protection implications are being integrated across the European Union, the significance of this subject is anticipated to extend to the global stage, particularly regarding the United States [6,7].

1.2. Navigating the Complex DPIA Landscape

A typical DPIA thereby comprises at least six main steps: firstly, a description of processing activity; secondly, an assessment of necessity and proportionality; thirdly, the identification and description of potential risks; fourthly, an assessment of these risks concerning the expected harm and the probability of their occurrence; and fifthly, the description of technical and organizational measures (TOMs) aimed at minimizing the probability and/or potential harm of identified risks. Finally, the remaining residual risk is evaluated after the implementation of the TOMs [8–10].

Creating an individual DPIA is not only highly demanding due to the technical and multifaceted nature of the task but also because it necessitates bringing together a diverse range of expertise, including legal, organizational, medical, and especially, informatics aspects [11]. The collaborative effort required to consolidate this respective expertise and integrate different points of view constructively and coherently can be inherently difficult and time-consuming. Nevertheless, this amalgamation of knowledge and perspectives is crucial to ensure a comprehensive and robust assessment of potential risks and data protection measures.

1.3. Introducing DPIA Click&Go

Despite the challenges involved in the DPIA process, the existence of commonalities among many DPIAs provides an opportunity for optimization. Recognizing that multiple DPIAs often encounter similar risks and comparable risk-mitigating strategies, we identified the potential to streamline the process and enhance efficiency, as described in Article 35(1), sentence 2 of the GDPR. As a result, we dedicated our efforts to develop an innovative and user-friendly solution to facilitate the generation of a DPIA: the open-source "plug-and-play" tool, DPIA click&go (in German *DSFA click&go*, translated from the German word *Datenschutzfolgenabschätzung—DSFA*).

The DPIA click&go tool empowers users to navigate the intricacies of conducting a DPIA by offering a systematic approach. In a user-friendly interface, individuals can conveniently select risks from a pre-defined, institute-specific pool of overarching risks and their corresponding sub-risks (Step 1). The pool also contains risks that are project-type-specific, i.e., different use cases can be selected and combined. Subsequently, the tool automatically generates an evaluation matrix to assess the probability of occurrence and the severity of harm because the risk assessment in this regard is initially regarded as comparable in every project (Step 2). This matrix serves as a foundation for further analysis and decision-making.

To address the identified risks effectively, DPIA click&go also provides users their institute-specific, curated risk-mitigating measures, commonly known as technical and organizational measures (TOMs, Step 3). These suggested TOMs, based on industry best practices and relevant institutional regulations, give users a starting point for designing their own data protection strategies.

As the individual DPIA progresses, users can expand on the selected risks and associated TOMs, tailoring them to suit the specific characteristics of their project. The tool allows users to conveniently insert these customized elements as text blocks into the tool, which then generates a Word document, thereby significantly reducing the time and effort required to draft the DPIA report.

A comprehensive literature review on PubMed and Google Scholar, spanning from 2017 to the present, explored GDPR implementation in various institutes and tools for DPIA generation. Our search encompassed terms like GDPR, data protection impact assessment, and DPIA, revealing works addressing GDPR's technical requirements [12], evaluating

privacy and ethics in health research networks [13], formal descriptions [14], mathematical abstractions [15], application-specific frameworks [16], evaluation frameworks for existing software [17], and case studies for step-by-step DPIA generation [18]. Notably, no existing tool offers a versatile, semi-automated approach applicable across diverse healthcare projects, as proposed here.

In this work, we will demonstrate using an example project how the here-introduced DPIA click&go tool can be highly effective in the local data protection context of our institute to significantly simplify the creation of a DPIA. Furthermore, the DPIA click&go tool can serve as a guiding example for other national institutes, as well as for other countries facing similar challenges.

Emphasizing the challenges, rationale, and contributions of our work, this introduction sets the stage for a comprehensive exploration of the DPIA click&go tool and its implications in the realm of data protection and medical research.

2. Materials and Methods

2.1. Purpose and Overview of DPIAs

In the context of the General Data Protection Regulation (GDPR), the European data protection legislation, the conduction of a DPIA is mandated if the data use poses a high risk to the rights and freedoms of natural persons due to its nature, scope, context, and purposes. These criteria are usually fulfilled for both research endeavors and clinical routine projects encompassing the collection, transmission, and/or storage of patient-related information but are not restricted to these kinds of projects [19]. The primary objective of a DPIA resides in the thorough evaluation of potential risks arising from data processing, transmission, and storage, coupled with the identification of TOMs to curtail these risks. It is noteworthy that a DPIA and data protection considerations in general are oriented toward risks that hold the potential to compromise the rights of patients (= data subject rights), while not encompassing risks to researchers seeking data utilization. For instance, the presence of data corruption upon reaching the researcher may not necessitate risk mitigation strategies within the DPIA. However, if patient data is housed within a platform that experiences corruption, rendering it inaccessible, such a risk impacts the patient's interests and mandates consideration within the DPIA framework. Another facet warranting attention within a DPIA pertains to the rights of physicians or other professionals involved. These professionals also hold data protection rights, underscoring the need for the development of TOMs to address potential issues, such as the retrospective reassignment of patients to their respective physicians once the data have been pseudonymized. The DPIA thus serves as a pivotal instrument in safeguarding the multifaceted rights and interests of patients, researchers, and medical practitioners alike.

Before commencing any project or undertaking a DPIA, it is crucial to clarify the scope and extent of the DPIA's application, including determining when other project partners must conduct their own DPIA. Moreover, before the actual initiation of data flows, the generation of a DPIA is imperative.

2.2. Decomposing a DPIA

Looking more into the details of DPIAs: German data protection supervisory authorities refer to the "Standard Data Protection Model" (SDM) as a method to ensure a consistent privacy consulting and auditing practice, specifically for determining technical and organizational measures by the GDPR [20]. The essential component of the SDM consists of a concept called "seven elementary warranty objectives", which are anchored in the principles of Article 5 of the GDPR [21]. These warranty objectives include ensuring confidentiality, integrity, availability, transparency, intervenability, and non-linkability of personal data processes, along with the requirement of "data minimization". These warranty objectives represent the core of each DPIA.

In addition to the aforementioned warranty objectives, there exist subgroups of risks that are attributable to specific data processing stages, thus reflecting the data lifecycle.

For instance, the subgroup includes aspects such as patient consent/notification, documentation, utilization, erasure, and so forth. A comprehensive listing of these subgroups can be found in Table 1. The data life cycle, which is shown in Table 1, represents the chronologically structured steps for every project concerning data handling in general. The cycle thus facilitates the assessment of risks regarding completeness. To associate individual risks with their respective subgroups, it is crucial, unless already undertaken as part of the project, to generate a data flow diagram and systematically scrutinize it from inception to completion to identify potential sources of risk. Consequently, each identified risk is given a risk source code, e.g., Doc01, Usa03, and Era04 for documentation, usage, and erasure, respectively. These codes are used to describe the risk in greater detail and to make a risk table within the DPIA, depicting the risk code and a short explanation of the risk at hand. Both the subgroups as well as the individual risks are crucial parts of a DPIA.

Table 1. The processing operations stipulated by legal parameters (Art 4 No 2 GDPR) are consolidated into larger clusters of processing activities, which are subsequently aggregated to encompass the higher echelon of the data lifecycle. It is worth noting that supplementary categories, specifically patient consent, and project management, hold relevance in their alignment with processing operations on a broader spectrum, demanding equal consideration. This tabular framework serves as a tool to facilitate the methodical compilation of risks systematically arranged according to process stages. It thereby also serves as an aid to assign the identified risks to their subgroup and a risk code. These risk codes are utilized to generate a table composed of the risk code and a short explanation of the risk at hand. Later, these codes are used to match the associated TOMs.

Basic Data Processing Operations	Categories	Phase of Data Lifecycle	Comments
1. Data collection 2. Capture data	1. Input	Collection	Collection of personal (and medical) data. The collector then is in charge of the sensitive data.
3. Organisation 4. Arrange 5. Storage	2. Redcating 3. Storage	Availablility	The data is stored in an organized manner and is available in a processable state.
6. Adaption 7. Selection 8. Query 9. Utilization 10. Disclosure 11. Equalisation 12. Data sharing 13. Data transfer 14. Distribution 15. Limitation	4. Editing 5. Usage 6. Distribution 7. Merge 8. Restriction	Usage	The Data is accessible for lawful and appropriate processing, potentially event to unauthorized third parties. The data can be linked with other processes and access can be restricted.
16. Destruction	9. Deletion	Deletion	Data will be irreversibly removed or physically deleted

Subsequently, to attribute the ascertained risk to an appropriate warranty objective, a determination must be made regarding the primacy of the specific risk concerning the given warranty objectives, and whether any concurrent warranty objective is of negligible relevance. For instance, in cases where a consent withdrawal is executed but fails to reach the IT administration, resulting in the persistence of patient data within the system for research or analogous purposes, both the confidentiality and intervenability warranty objectives are contravened. Yet, it merits highlighting that the breach of confidentiality is to be perceived as substantially more hazardous than the intervenability violation. Alternatively, when a singular risk yields non-negligible consequences in divergent dimensions, it could potentially implicate two distinct warranty objectives. For instance, in the event of the theft of the data server, there exists the possibility of the pseudonymized data falling into unauthorized hands, thereby transgressing the confidentiality warranty objective. Conversely, the availability warranty objective may also be compromised, rendering the data inaccessible. Consequently, a single risk source can engender divergent potential consequences. Simultaneously, distinct risks can result in the same harm; for instance, an attacker might use phishing tactics to acquire login credentials, enabling access to software handling patient-related data, or a user might inadvertently neglect to log out after utilizing

said software. The resultant harm in both cases would involve unauthorized parties gaining access to sensitive patient-related information.

After pointing out all potential risk factors and matching them to subgroups and warranty objectives within the DPIA, it is necessary to evaluate these risks. With the application of a rating mechanism that assesses both the likelihood of occurrence and the extent of potential harm, each rated on a scale ranging from 1 to 4, a composite risk score is computed. This cumulative score facilitates the categorization of risks into three discrete levels: low-, medium-, and high-risk, as depicted in Figure 1. Within the DPIA framework, both the enumeration and elucidation of individual risk scores, alongside the aggregation of risk scores for each identified risk, constitute essential components. Should a risk surpass the low-risk category in Figure 1, the imperative arises to proffer TOMs aimed at mitigating either the likelihood of incidence, the magnitude of potential harm, or optimally, both factors. Furthermore, if TOMs are viable for risks falling below this threshold, their implementation is recommended. The overarching objective remains the attainment of the utmost risk reduction to ensure the robust safeguarding of patient data. Lastly, a risk evaluation after the application of TOMs must be performed. If the risk score is not high anymore, the project can be conducted. To match the TOM to the right risk, the risk codes, which were described above, are utilized. The main process to create a DPIA is depicted in Figure 2. The parts that can be facilitated by the DPIA click&go tool are shown in Figure 2B–D; the data flow diagram is still a requirement as a starting point for each individual project (Figure 2A). An exemplary DPIA snippet including a specific risk, one of the seven objectives, a risk evaluation, tailored TOMs, as well as the risk evaluation after the application of the TOMs is displayed in Figure 3. Figures 2 and 3 also aim to give brief examples of the types of risks that are addressed using DPIAs.

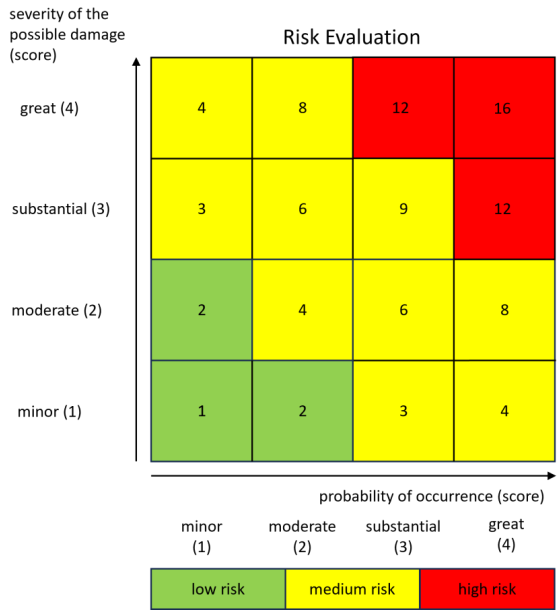

Figure 1. The proposed matrix allows a comprehensive risk assessment for the data protection impact assessment (DPIA) by utilizing a rating system that considers both the probability of occurrence and the severity of possible damage, each rated on a scale from 1 to 4. By multiplying these ratings, a combined risk score is derived, which enables the classification of risks into three distinct categories: low-, medium-, and high-risk. Risks falling into the low-risk category necessitate the implementation of technical and organizational measures (TOMs) to address and mitigate their impact. If, even after the implementation of TOMs, a risk remains classified as high (indicated by the red zone), it indicates an unacceptably high level of risk. Consequently, proceeding with the project would be impossible from a data protection perspective and requires further analysis and adjustments to ensure compliance with established risk thresholds. The figure was oriented by [22].

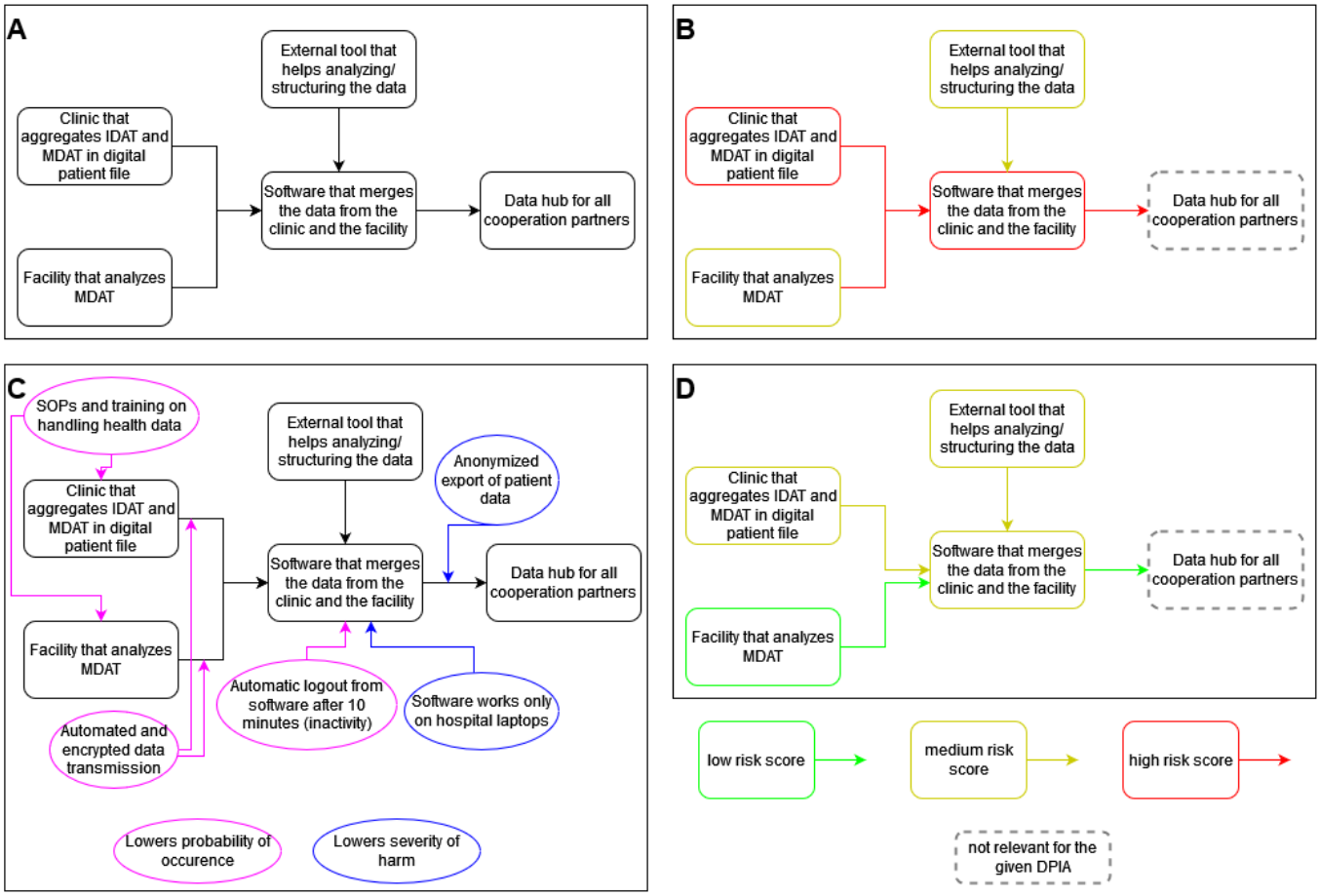

Figure 2. The process can be divided into several key steps (simplified) consisting of different project parts that are interconnected, i.e., where patient-related data are transferred (here, indicated by arrows): (**A**) The creation of a data flow diagram, such as the one used in the here-presented example project—the tumor board project with multiple consortium partners. (**B**) The analysis and categorization of potential risks are based on the matrix presented in Figure 1. In this step, specific TOMs are required for the red parts of the matrix, and depending on their risk score, some of the yellow parts may also necessitate TOMs. Implementing TOMs for parts below the threshold is advisable to achieve a lower risk score, which is always the objective. The dotted line implies that the DPIA solely covers the transmission part from the merging software to the data hub, not the data hub itself, as the latter requires a separate DPIA. (**C**) A preliminary selection of TOMs aimed at reducing the probability of occurrence (marked in pink) or the severity of harm (marked in blue). It is important to acknowledge that the same TOM may apply to different risk sources, multiple TOMs may mitigate the same risk, and some TOMs can simultaneously lower both the probability of occurrence and the severity of harm. (**D**) The risk evaluation after the implementation of the TOMs. The absence of red parts indicates that the project can proceed from a data protection standpoint. Note, that the DPIA click&go tool only offers suggestions for (**B**–**D**), but not for (**A**). The creation of the data flow diagram (**A**) is obligatory and unique to each individual project. Note, that "external tool" hereby refers to a tool that was not developed by our institute but by a different academic institute. IDAT = identifying data, MDAT = medical data, SOP = standard operating procedure.

1. Damage events "breach of the warranty objective of confidentiality"

a. In the context of registration/logging in

Risk source: Reg01

The password is discovered by "try and error".

Consequence:

The confidentiality of the data would be directly affected by an unauthorized determination of the password. Attackers would have access (at least) to the plain names and dates of birth of all patients created in the software. The amount of data that would be accessible to attackers depends on the user rights of the hacked person. In addition, the attackers would have access to the data of the physicians and the assignment to the patients they treat.

Evaluation matrix:

Probability of occurrence: 2 – The awareness of all users for the use of secure passwords is high.

Severity of harm: 4 – The damage caused by unauthorized access to the software must be classified as high.

Risk score: 8

... (all other risks in context to their warranty objective)

Damage event due to non-granting of confidentiality:

Risk source: Reg01

The password is discovered by "try and error".

The following technical and organizational measures are defined for containment:

- The xxx institute password guidelines (version xx of xx.xx.20xx, paragraphs x-x) apply to all employees.
- The guideline for information security and data protection (version xx of xx.xx.20xx, paragraphs x-x) applies to all employees.
- Only a designated system administrator has access to the usage rights and can assign passwords and access data.
- The German Federal Office for Information Security (BSI) provides recommendations for creating secure passwords to ensure IT security: change preset password, password should be easy to remember, password should be at least 8 characters long, all available characters can be used, at least three character types should be used, security through multi-factor authentic, the same password should not be used for multiple accounts (Status August 2023).

Renewed risk assessment under Consideration of the measure taken.

Probability of occurrence: 1 – Due to the defined TOMs, the probability of passwords being determined by "try and error" is almost eliminated.

Severity of harm: 4 – as before.

Risk score: 4

Figure 3. Example of a chosen risk source (registration and logging in) and a warranty objective (confidentiality) accompanied by the associated evaluation matrix, TOMs, and subsequent risk re-evaluation. The risk source code (Reg01 in this case) is used to generate a risk table, in which the code is associated with a short explanation of the risk at hand. Moreover, these codes are used later to match the risk to the related TOMs. Note that the specific format is not mandatory and can differ for other institutes.

In summary, a DPIA—and thus, the DPIA click&go tool—encompasses warranty objectives, individual risks categorized under data lifecycle subgroups, associated risk codes, a tabulated presentation of these risk codes with concise explanations, risk scoring involving the assessment of the probability of occurrence and the severity of harm, amalgamated risk scoring, TOMs, where deemed necessary, and a subsequent reassessment of risk scores encompassing both the two individual as well as the cumulative perspectives post-TOMs application. DPIA click&go is a semi-automated tool that incorporates the aforementioned steps and provides best-practice components as default content that can be extended and adapted individually.

Due to the well-defined warranty objectives and their subgroups, a DPIA can be effectively deconstructed, allowing for the seamless creation of individual text snippets that can be flexibly combined as needed. This fact is the foundation for the DPIA click&go tool.

2.3. TOMs

Given that TOMs form a pivotal cornerstone of a DPIA, the subsequent paragraph provides an in-depth elaboration on this crucial aspect. As implied by the nomenclature, *technical and organizational measures* encompass a blend of technical and organizational considerations that serve to elevate the protective mantle enshrouding patient-related data. Some of these measures possess universal applicability, transcending project-specific boundaries. Nevertheless, it is important to acknowledge that many risks warrant meticulous tailoring to the distinctive contours of each individual institute. Thus, the DPIA click&go tool has to be fed with the institute's specific DPIA snippets, especially regarding the TOMs, as the published version of the tool is solely intended to provide a semi-automated inter-institutional streamlining of the whole DPIA formulation process. Note that the ensuing descriptions of TOMs do not encapsulate the full range of possible TOMs but aim to give a broad overview of potential categories for mitigating risk strategies.

Starting with the technical aspects to consider: Adequate time and expertise are essential for mapping the data flow and establishing strong technical safeguards for a project's secure launch. Regarding software coding, only trained IT experts should handle health software, and implementing a double-check process is advisable.

During project planning, two critical IT areas should ideally be focused on: (1). Setting up secure networks with robust firewalls to thwart malware and prevent security vulnerabilities that could lead to data breaches. (2). Utilizing secure computers, approved by IT professionals, ideally without ports such as USB that might allow entry of malicious data or malware capable of compromising software or databases. Further general technical considerations can involve encrypted servers, automatic logouts for inactive software, routine backups, and controlled access to software overseen by a limited group of trained IT administrators. To prevent "man-in-the-middle" attacks at interfaces, it is recommended to restrict database access to specific IP addresses. When applicable, exploring additional security measures like an audit trail or a "demilitarized zone" (DMZ) might be considered. Ensuring that patient data-handling software includes functional email communication enables users to connect with experts in case of issues. Importantly, documenting all technical steps taken in a dedicated plan or equivalent documentation is imperative.

In addition to the mentioned technical measures, there are organizational measures that should be considered: TOMs primarily involve clear standard operating procedures (SOPs) that give specific and legally binding action instructions to everyone involved in data handling and related activities within the project. Each part of the institutes involved should have its own SOP. Moreover, all hospitals should provide official guidelines, e.g., for password management, handling technical devices, or general data protection and security. Regular training (e.g., (bi-)annually) is a must to educate employees about data protection and the measures of the institute. Access to hospital buildings should be tightly controlled, and security measures like cameras and security staff should be in place to prevent unauthorized access to computers with patient data or servers.

Another potential organizational measure stems from the ability to classify patient-related data into two distinct categories: (1). data facilitating direct patient identification, such as full name, address, birth date, health insurance number, and patient ID, referred to as identifying data (IDAT); and (2). the patient's medical data (MDAT), encompassing diagnoses, lab values, medical images, and treatment plans, among others. Segregating the storage of MDAT and IDAT, with distinct pseudonym assignments for each, provides an enhanced level of data security [23,24]. This strategic separation restricts potential attackers' access to only one dataset, thereby curtailing the utility of the information they could potentially glean. Recent attention has been directed toward trustee services to ensure the segregation and non-merging of IDAT and MDAT, particularly their respective pseudonyms. The concept of trustee services involves an independent institute or entity entrusted with managing the assignment lists, thereby exclusively possessing the pseudonyms that enable the unique linking of IDAT and MDAT. Consequently, the trustee services, that maintain the lists enabling the reallocation of IDAT and MDAT, must provide the highest data security standards.

It is important to highlight that a single TOM can be pertinent to various risk sources, as exemplified by the role permission concept. This practice contributes to the accurate management of distinct software components, each presenting distinct risk sources. Consequently, this singular TOM effectively mitigates the likelihood of occurrence across multiple risk scenarios. Furthermore, it is also possible that the same risk requires multiple TOMs, not only to minimize both the probability of occurrence and the severity of harm but also if there are multiple measures to lower one of the scores, as depicted in Figure 3. To show the potential of the DPIA click&go tool, especially in terms of TOMs, we used it to create a DPIA for one example project.

2.4. The Example Project

We chose a project that aims to optimize the existing process of a tumor board as an example. One objective of this project is to achieve standardization and enhancement of prevailing clinical procedures. The project encompasses a broad scope, as the project consortium comprises over 20 additional university hospitals. However, data protection has emerged as a complex challenge due to several factors. This includes the intricacies involved in primary data processing and its subsequent augmentation with identifiable information. The process entails the pseudonymized extraction of data, channeling it to a multi-consortium data hub. Notably, the project also involves the integration of the optimized tumor board into clinical routines. This integration necessitates concurrent operation while accommodating adaptations aligned with the optimized procedures. This project was used to schematically describe and visualize the process of a DPIA generation, as depicted in Figure 2. It should be noted that the generated DPIA of the chosen project only covers the part that involves our institute and does not include the data protection considerations of the other 20 cooperation partners.

3. Results

The text snippets within the DIPA click&go tool were carefully developed for our institute by incorporating the knowledge of legal experts, IT specialists, clinical project partners, as well as existing DPIAs. This enabled us to create a tool for our institute in which different use cases (e.g., multicentric registers, biomaterial usage, data warehouses, and tumor boards) are included. An exemplary snippet of the DPIA click&go tool to illustrate its function is shown in Figure 4. The complete description of the mockup can be found in the Supplementary File S1. It is important to acknowledge that neither the comprehensive publication of the DPIAs for the projects nor the publication of the utilized DPIA click&go text snippets is feasible due to internal data protection constraints within the hospital setting.

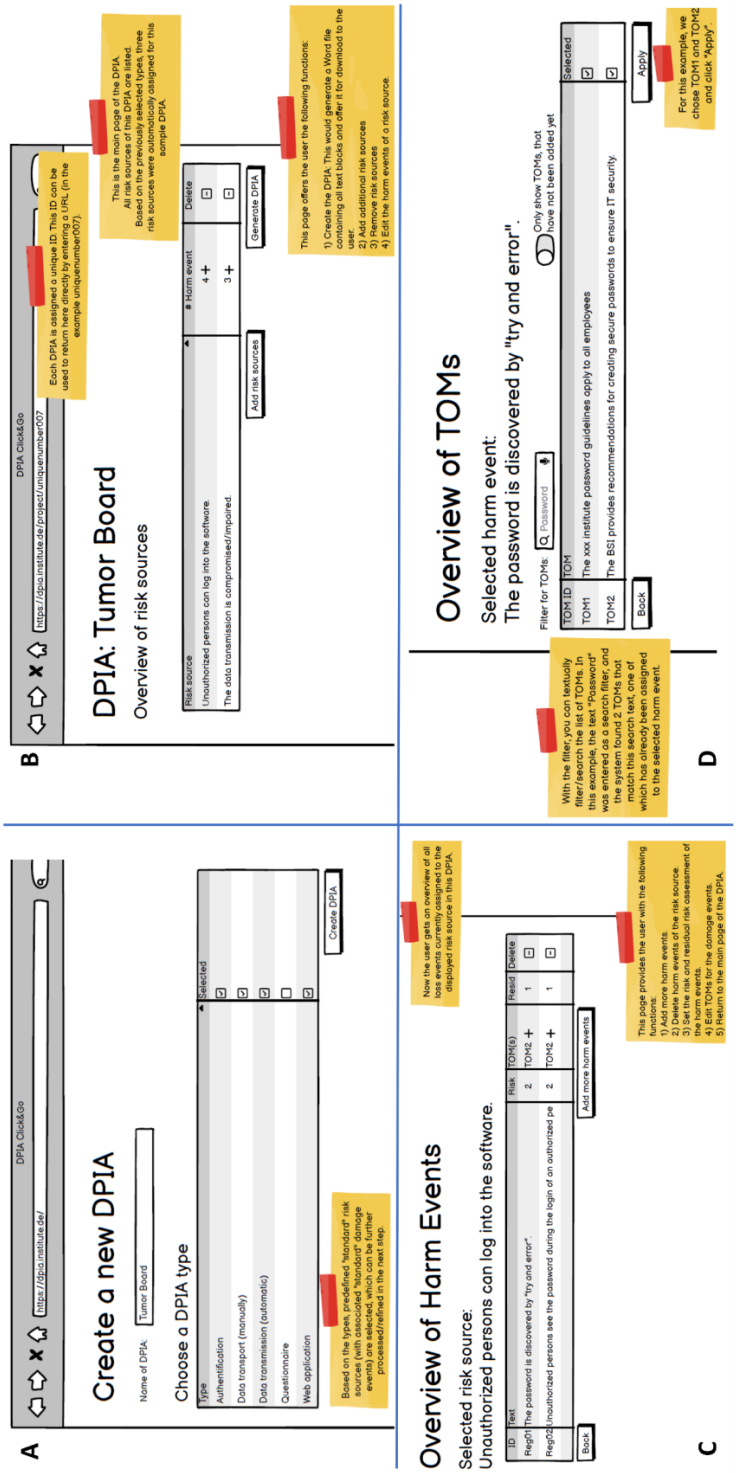

Figure 4. Screenshots of the mockup version to showcase some crucial parts (simplified) of the DPIA click&go tool (original version in German). (**A**) First, one can select the type of DPIA, i.e., which components are involved. (**B**) Then, the corresponding general risk sources and their risk score are suggested. (**C**) Afterward, an overview of the potential harm events related to the selected risk source is depicted. (**D**) TOM(s) that aim to mitigate the probability of occurrence and/or severity of harm for the selected harm event are suggested, thereby enhancing data protection strategies. Note that the content of the text snippets can be changed, new TOMs and risks can be added, and all scores can be adjusted. After the described selection process from (**A**) to (**D**), there is a button that generates the customized DPIA as a Word document. The complete description of the tool can be found in the Supplementary File S1.

The backend of the DPIA click&go tool is coded in Java (OpenJDK Amazon Corretto [25]), whereas the frontend is built in React [26]. The utilized database is PostgreSQL [27]. The selection of Java, React, and PostgreSQL for the backend, frontend, and database technologies, respectively, was driven by their established status in the industry. Java 17 serves as our code standard, while React with TypeScript was chosen to simplify implementation, enhance flexibility, and ensure compatibility with web browsers used in the clinical environment, particularly in the context of corporate policy restrictions. Finally, PostgreSQL was the preferred database choice due to our extensive experience and consistently positive outcomes with the platform.

The code of the "raw" DPIA click&go tool will be publicly available here: https://github.com/LTRiemann/DPIA_click_and_go.git (accessed on 4 October 2023). Note, that the non-tailored, publicly available version of the tool only includes risks and TOMs that are suggested by the German Federal Office for Information Security (in German: *Bundesamt für Sicherheit in der Informationstechnik—BSI*). Each institute has to adapt and fundamentally extend them to their specific requirements and circumstances by using already existing DPIAs. The included risks and TOMs serve solely the purpose of facilitated understanding. Each existing DPIA can be included in the tool in the form of an Excel spreadsheet or directly by using a PostgreSQL database. The corresponding data model and example Excel spreadsheet can be also found on GitHub. Due to the search function within the tool and the possibility of seeing the complete list of TOMs and risks, the curation of the text snippets is facilitated.

In the context of the chosen exemplary tumor board project, a comparative analysis between the DPIA generated using the DPIA click&go tool and the manually crafted DPIA, developed by a cohort of researchers and IT specialists lacking legal expertise, revealed the discovery of 11 additional risks and 14 corresponding TOMs. While the tool reduced the DPIA generation time to 5 h for a team of three, manual generation involved six individuals, each dedicating 18 h. To assess user-friendliness quantitatively, we obtained a System Usability Scale (SUS) [28] including the experiences of five members of our institute in different positions who already utilized the DPIA click&go tool for generating DPIAs. The obtained scores ranged from 72.5 to 82.5 (other scores: 75, 77.5, 77.5).

It is worth highlighting that a DPIA drafted by a multi-professional team consisting of research, IT and legal experts, quality assurance staff, and process managers among others is considered the gold standard, as it incorporates all different perspectives that are crucial to obtain a complete risk and TOM evaluation as each profession is familiar with its specific risk scenarios. Thus, legal expertise is of significant advantage in interpreting and deciphering legal requirements. Within the non-legal expert group, the tool notably facilitated the recognition of risk sources primarily associated with management, patient consent, and data subject rights. These specific risks often fall outside the scope of the conventional data flow diagram, are thus prone to oversight, and are usually included by the legal experts within the DPIA team. Additionally, the tool proved instrumental in identifying relevant TOMs, especially those aligned with existing institutional guidelines or SOPs that frequently lie beyond the traditional purview of research considerations. Even if the suggested guidelines were not directly applicable to the project at hand, they served as valuable prompts, necessitating the formulation of similar or analogous measures tailored to the new project's requirements.

4. Discussion

In this study, we introduced the user-friendly "plug-and-play" tool DPIA click&go to simplify the process of creating DPIAs. This tool streamlines the identification of risk sources, their categorization into subgroups and warranty objectives, the calculation of risk scores based on likelihood and potential harm, the subsequent formulation of TOMs, as well as the reassessment of the risks after the application of TOMs. Since these steps are typically the most labor-intensive and challenging, this semi-automated tool significantly reduces the workload and human power involved.

In the context of our project, the customized DPIA click&go tool demonstrated remarkable efficiency, significantly expediting the DPIA creation process compared with the manual generation of a DPIA. This was evidenced by a comparative analysis of time and personnel requirements, as well as the identification of risks and mitigation strategies. The enhancement in DPIA generation efficiency was attributed to the tool's user-friendly frontend. This aspect was rigorously assessed using an SUS evaluation, which yielded consistently high ratings, all surpassing the threshold of 68. These results firmly establish that the tool indeed excels in delivering a user-friendly interface and experience.

Due to internal data protection considerations, it was not possible to publish the institute's specific risks and TOMs, but the ones published and recommended by the BSI. It should be noted though that these do not reflect a complete data protection set that suffices the requirements of a full DPIA and thus have to be extended. In addition to the existing risks and TOMs in the tool, users have to add their own risks and TOMs from existing DPIAs to tailor them to their specific institutional and project needs and to be able to use the full potential of the tool. This customization part is particularly advantageous for users of one institute planning multiple projects with similar use cases all requiring DPIAs, as the risk and TOM pool is continuously extended. Hence, it is anticipated that the quality of DPIAs formulated by domain specialists lacking legal proficiency is enhanced due to the inherent suggestions of additional risks and corresponding TOMs provided by the tool, addressing potential oversights that might otherwise be overlooked. This, in turn, affords researchers more dedicated time for their primary undertakings while concurrently alleviating the responsibilities of the data protection expert. In our institutional context, this hypothesis was validated within the tumor board DPIA category, comparing DPIAs manually generated with those produced using the DPIA click&go tool.

Due to the tool's dockerization and its database in MS Excel or Postgres, extensive coding knowledge is not necessary for its utilization. This makes it accessible and adaptable to various hospitals, institutes, and personnel dealing with patient data, who need to conduct thorough DPIAs.

Since it became apparent that the GDPR would be enacted, and especially since the broad content became known, the topic of DPIAs has engaged other institutions, their researchers, and their legal experts. When situating the DPIA click&go tool within the literature, it is noteworthy that other authors have either focused on the formal description [14] or the mathematical abstraction of DPIAs [15] or presented a framework for a specific multilayer application [16]. Another paper introduces a framework to evaluate and generate DPIAs for existing, clinically utilized, and self-contained software solutions in accordance with GDPR guidelines [17]. Yet another offers a case study with a hypothetical project, guiding the step-by-step generation of a DPIA for a hypothetical healthcare-related project in collaboration with a legal expert and a project specialist [18]. To our knowledge, none of the aforementioned works provide an open-source "plug-and-play" framework that streamlines the DPIA process, especially for complex, networked projects in the healthcare sector.

While the DPIA click&go tool notably streamlines the process, it is essential to emphasize that complete automation is neither achieved nor aspired, even after the incorporation of the existing, institute's internal DPIAs. Human participation remains crucial, involving a careful review of proposed risks, their alignment with the project's specifics, identification of any missing risks, their potential inclusion, and an evaluation of the suggested risk scores and TOMs for their relevance and potential risk reduction. Otherwise, risks, TOMs, and/or logic errors remain uncorrected. Consequently, the tool's propositions demand evaluation, and the DPIA's finetuning necessitates a multidisciplinary team, particularly those well-versed in IT, to ensure comprehensive risk consideration. IT experts provide the required technical understanding of the whole data life cycle and all involved IT-relevant components, which is integral in order to provide a secure application. Despite the possibility of displaying the complete list of TOMs and risks as well as providing an internal

search function, the tool needs careful and regular maintenance to prevent multiple entries describing the exact same risk or TOM.

Moreover, it is important to note that the DPIA click&go tool does not hold any legal binding, serving solely as a resource for suggestions. It enriches the understanding of DPIA-relevant risks, fostering an improved workflow for DPIA formulation. It should also be stressed that the execution of TOMs derived from the DPIA is not provided by the DPIA click&go tool but has to be implemented for each institute individually.

The design of the DPIA click&go tool prioritizes universality, and thus, it does not encompass the project-specific description segment, which inherently varies across institutes and encompasses unique factors such as project summary, acquired data, name of collaboration partners, and their specific functions within the project. This segment typically functions as the preamble to the DPIA.

It is important to acknowledge that the evaluation of the DPIA click&go tool was initially confined to internal hospital projects, with the showcased project serving as a use case in which the authors hold a profound understanding of the risks and intricate technical aspects. This deliberate approach was essential to establish a robust assessment of the tool's performance. Moving forward, our efforts will focus on applying the tool to diverse hospital projects where the authors are not involved, expanding the breadth of assessment.

Our alignment with the GDPR principles and regulations has been paramount in shaping the DPIA click&go tool's universal relevance, although we acknowledge the potential bias introduced by our internal data protection perspective and local legal regulations. Consequently, the testing of the tool on different internal use cases is followed by its introduction in other German hospitals or institutes that operate under distinct local data protection laws and institute-specific guidelines. A subsequent phase involves translation and adaptation of the tool to adhere to legal frameworks in various countries.

Next to the described steps toward greater internal, national, and international dissemination, future endeavors will also be directed toward the expansion of the tool's capabilities with the integration of a natural language processing (NLP) plugin, enabling tailored suggestions for text snippets. The incorporation of an NLP plugin stands to augment our capacity to aid non-legal experts in refining their text snippets by offering nuanced recommendations. Additionally, the utilization of an NLP-based plugin presents an opportunity to introduce a chatbot, thereby further streamlining the DPIA generation process. This enhancement aims to facilitate a more flexible and institution-specific generation of DPIA reports.

5. Conclusions

The presented DPIA click&go tool not only facilitates the process of DPIA creation but also enhances the comprehensiveness of risk assessment for non-legal experts. Beyond its functional benefits, the tool can contribute to bridging knowledge gaps among legal experts, IT specialists, physicians, and other stakeholders within the healthcare domain, ensuring a holistic approach to DPIA formulation. While the tool is designed within the framework of German data protection laws, the overarching risks are largely congruent across jurisdictions due to shared technical infrastructure and implementations. This suggests the adaptability of DPIA click&go to international contexts through translation and customization as per national standards. Ultimately, this tool aims to serve as a catalyst for forming a DPIA community that generates content-aligned and standardized DPIAs.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app132011230/s1, File S1: Complete mockup of the DPIA click&go tool (English).

Author Contributions: Conceptualization, L.T.R., F.Ü., F.P.S.H. and M.A.; methodology, L.T.R., F.P.S.H. and A.-K.S.; software, M.A.; validation, M.J. and L.T.R.; formal analysis, L.T.R. and A.-K.S.; investigation, L.T.R.; resources, F.Ü.; data curation, M.A., F.P.S.H. and A.-K.S.; writing—original draft preparation, L.T.R.; writing—review and editing, L.T.R., M.J. and F.Ü.; visualization, L.T.R. and

A.-K.S.; supervision, F.Ü.; project administration, F.P.S.H., L.T.R. and F.Ü.; Funding Acquisition, F.Ü. All authors have read and agreed to the published version of the manuscript.

Funding: We would like to thank the deanery of the medical faculty of our hospital for their financial support and their legal advice, and we would like to thank the developer team of our institute for helping to implement the tool. We acknowledge financial support from the Open Access Publication Fund of UKE—Universitätsklinikum Hamburg-Eppendorf and DFG—German Research Foundation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The presented DPIA click&go tool (in German) with the risks and TOMs provided by the BSI, as well as the data model in Excel and Postgres, will be openly available on Gitlab https://github.com/LTRiemann/DPIA_click_and_go.git (accessed on 4 October 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mallappallil, M.; Sabu, J.; Gruessner, A. A review of big data and medical research. *SAGE Open Med.* **2020**, *8*. [CrossRef] [PubMed]
2. Tayan, O. Concepts and Tools for Protecting Sensitive Data in the IT Industry: A Review of Trends, Challenges and Mechanisms for Data-Protection. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 46–52. [CrossRef]
3. Der Bundesbeauftragte für Datenschutz und Informationssicherheit. History of German Data Protection. Available online: https://www.bfdi.bund.de/DE/DerBfDI/Inhalte/Datenschutzpfad/Geschichte-Datenschutz.html (accessed on 1 September 2023).
4. Hallinan, D.; Martin, N. Fundamental rights, the normative keystone of DPIA. *Eur. Data Prot. Law Rev.* **2020**, *6*, 178–193. [CrossRef]
5. Bieker, F.; Friedewald, M.; Hansen, M.; Obersteller, H.; Rost, M. A Process for Data Protection Impact Assessment Under the European General Data Protection Regulation. In Proceedings of the Privacy Technologies, Law and Policy—4th Annual Privacy Forum, Frankfurt/Main, Germany, 7–8 September 2016. [CrossRef]
6. Jones, M.L.; Kaminski, M.E. An American's Guide to the GDP. *Denver Law Rev.* **2020**, *98*, 93–128.
7. Li, H.; Yu, L.; He, W. The Impact of GDPR on Global Technology Development. *J. Glob. Inf. Technol. Manag.* **2019**, *22*, 1–6. [CrossRef]
8. Der Bayerische Landesbeauftragte für den Datenschutz (BayLfD). Datenschutz Bayern. Available online: https://www.datenschutz-bayern.de/dsfa/ (accessed on 2 September 2023).
9. Exculpa, "DSFAs Muster". Available online: https://exkulpa.de/datenschutz/datenschutz-folgenabschaetzung-muster-beispiele-notwendigkeit/#t-1678372060290 (accessed on 1 September 2023).
10. Europäische Kommission. Wann Ist Eine Datenschutz-Folgenabschätzung Erforderlich? Available online: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/when-data-protection-impact-assessment-dpia-required_de (accessed on 5 September 2023).
11. Chassang, G. The impact of the EU general data protection regulation on scientific research. *eCancermedicalscience* **2017**, *11*, 709. [CrossRef] [PubMed]
12. Hussein, R.; Wurhofer, D.; Strumegger, E.; Stainer-hochgatterer, A. General data protection regulation (GDPR) toolkit for digital health. In *MEDINFO 2021: One World, One Health—Global Partnership for Digital Innovation*; IOS Press: Amsterdam, The Netherlands, 2022; pp. 222–226. [CrossRef]
13. Di Iorio, C.T.; Carinci, F.; Oderkirk, J.; Smith, D.; Siano, M.; De Marco, D.A.; De Lusignan, S.; Hamalainen, P.; Benedetti, M.M. Assessing data protection and governance in health information systems: A novel methodology of Privacy and Ethics Impact and Performance Assessment (PEIPA). *J. Med. Ethics* **2021**, *47*, E23. [CrossRef] [PubMed]
14. Binns, R. Data protection impact assessments: A meta-regulatory approach. *Int. Data Priv. Law* **2017**, *7*, 22–35. [CrossRef]
15. Mollaeefar, M.; Siena, A.; Ranise, S. Multi-stakeholder cybersecurity risk assessment for data protection. In Proceedings of the ICETE 2020—Proceedings of the 17th International Joint Conference on e-Business and Telecommunications SECRYPT, Paris, France, 8–10 July 2020; Volume 3, pp. 349–356. [CrossRef]
16. Kalloniatis, C.; Lambrinoudakis, C.; Musahl, M.; Kanatas, A.; Gritzalis, S. Incorporating privacy by design in body sensor networks for medical applications: A privacy and data protection framework. *Comput. Sci. Inf. Syst.* **2020**, *18*, 323–350. [CrossRef]
17. Todde, M.; Beltrame, M.; Marceglia, S.; Spagno, C. Methodology and workflow to perform the Data Protection Impact Assessment in healthcare information systems. *Inform. Med. Unlocked* **2020**, *19*, 100361. [CrossRef]
18. Várkonyi, G.G.; Gradišek, A. Data protection impact assessment case study for a research project using artificial intelligence on patient data. *Informatica* **2020**, *44*, 497–505. [CrossRef]
19. Europäische Kommission. Verordnung zum Schutz Natürlicher Personen bei der Verarbeitung Personenbezogener Daten, zum Freien Datenverkehr und zur Aufhebung der Richtlinie 95/ 46/ EG (Datenschutz-Grundverordnung). Verordnung (EU) 2016/ 679 des Europäischen Parlaments und Rates. 2016. Available online: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679 (accessed on 31 August 2023).

20. Rost, M. Das Standard-Datenmodell—Eine Methode zur Datenschutzberatung und -Pruefung auf der Basis Einheitlicher Gewaehrleistungsziele. 2022. Available online: http://www.govdata.de/dl-de/by-3-0 (accessed on 3 September 2023).
21. Hamburg Datenschutzbeauftragte. Short Papers Regarding the GDPR (German). Available online: https://datenschutz-hamburg.de/pages/kurzpapiere-dsgvo/ (accessed on 3 September 2023).
22. Lorop. Datenschutz Berlin. Available online: https://datenschutz-berlin.com/datenschutz-folgenabschaetzung/ (accessed on 20 September 2023).
23. Brinkmann, L.; Klein, A.; Ganslandt, T.; Ückert, F. Implementing a data safety and protection concept for a web-based exchange of variable medical image data. *Int. Congr. Ser.* **2005**, *1281*, 191–195. [CrossRef]
24. Lablans, M.; Borg, A.; Ückert, F. A RESTful interface to pseudonymization services in modern web applications. *BMC Med. Inform. Decis. Mak.* **2015**, *15*, 2. [CrossRef] [PubMed]
25. OpenJDK Amazon Corretto. Available online: https://docs.aws.amazon.com/corretto/latest/corretto-17-ug/what-is-corretto-17.html (accessed on 17 August 2023).
26. React—The Library for Web and Native User Interfaces. Available online: https://react.dev/ (accessed on 17 August 2023).
27. PostgreSQL. Available online: https://www.postgresql.org/ (accessed on 17 August 2023).
28. Brooke, J. SUS: A "Quick and Dirty" Usability Scale. In *Usability Evaluation in Industry*; CRC Press: Boca Raton, FL, USA, 2020; pp. 207–212. [CrossRef]