technologies

Article

Sovereign Digital Consent through Privacy Impact Quantification and Dynamic Consent

Arno Appenzeller [1,2,*], Marina Hornung [1], Thomas Kadow [1], Erik Krempel [2] and Jürgen Beyerer [1,2]

[1] Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany; hornung.ma@web.de (M.H.); tkadow@web.de (T.K.); juergen.beyerer@iosb.fraunhofer.de (J.B.)
[2] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, 76131 Karlsruhe, Germany; erik.krempel@iosb.fraunhofer.de
* Correspondence: arno.appenzeller@iosb.fraunhofer.de

Abstract: Digitization is becoming more and more important in the medical sector. Through electronic health records and the growing amount of digital data of patients available, big data research finds an increasing amount of use cases. The rising amount of data and the imposing privacy risks can be overwhelming for patients, so they can have the feeling of being out of control of their data. Several previous studies on digital consent have tried to solve this problem and empower the patient. However, there are no complete solution for the arising questions yet. This paper presents the concept of Sovereign Digital Consent by the combination of a consent privacy impact quantification and a technology for proactive sovereign consent. The privacy impact quantification supports the patient to comprehend the potential risk when sharing the data and considers the personal preferences regarding acceptance for a research project. The proactive dynamic consent implementation provides an implementation for fine granular digital consent, using medical data categorization terminology. This gives patients the ability to control their consent decisions dynamically and is research friendly through the automatic enforcement of the patients' consent decision. Both technologies are evaluated and implemented in a prototypical application. With the combination of those technologies, a promising step towards patient empowerment through Sovereign Digital Consent can be made.

Keywords: e-health; digital consent; risk quantification; formal consent model; medical consent; medical data; dynamic consent; broad consent; data sovereignty

Citation: Appenzeller, A.; Hornung, M.; Kadow, T.; Krempel, E.; Beyerer J. Sovereign Digital Consent through Privacy Impact Quantification and Dynamic Consent. *Technologies* **2022**, *10*, 35. https://doi.org/10.3390/technologies10010035

Academic Editor: Fillia Makedon

Received: 14 January 2022
Accepted: 17 February 2022
Published: 21 February 2022

1. Introduction

The use of digital consent seems to be a promising improvement to speed up research projects that use personal health data. The European General Data Protection Regulation (GDPR) considers personal health data as highly sensitive data and sets strict requirements for processing exclusion for them [1]. Article 9 Paragraph 2 a) states that one of the exclusions is the explicit consent of the data subject. Considering a large research project with many participants, the potential overhead of paper-based consent can easily be reduced by using digital consent technologies. While the technology itself becomes more and more usable in practice, from a technical point of view, there are still open questions in terms of usability, privacy, and acceptance of digital consent. This is clearly an interdisciplinary topic that requires an ethical and legal point of view (such as, for example, in the work by Grady [2]), but this paper is limited to technical requirements. Regarding this aspect, we identified two main issues with digital consent in terms of patients' privacy and sovereignty: The lack of a decision support for giving consent and missing realizations of Dynamic Consent. Firstly, pure digital consent can be potentially overwhelming for a patient [3]. While naturally researchers would want as much data as possible, it is difficult to see for patients what value their data have and what risk for the individual privacy sharing of such data holds. In a typical treatment setting, trust would be mandatory, so a patient should

not have any doubts about sharing medical data with a doctor who provides them with medical care. For secondary usage (i.e., research), this is not so obvious. Different factors need to be considered to gain trust and acceptance in a research project [4]. Furthermore, there needs to be a way to assess the potential impact on the individual's privacy when sharing personal health records. This depends on anonymization or privatization used in a project and the general risk of data leakage from a research project. To address this, we introduce a consent privacy impact quantification (CPIQ), which we see as one key part of Sovereign Digital Consent. The other key part is the realization of Dynamic Consent, which has recently become popular in improving participation in research projects [5]. The main idea of Dynamic Consent is that the consent could be altered dynamically. A research project could have changes in its purpose during the research process and so patients could change which data they want to share. In addition, there needs to be way to technically define categories of data so a patient does not need to select every single resource and can also agree to proactive sharing, where data that are not yet created for a certain category (i.e., hearth fitness data) can be shared for future requests. To realize this, we present a Dynamic Consent implementation that uses Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) as data categorization classification. The implementation will use the eXtended Access Control Markup Language (XACML) to implement the patients' consent as access policies. With a combination of those two key properties and the general foundations of digital consent, we define the term Sovereign Digital Consent. The main contributions of this paper are:

- Definition of the term Sovereign Digital Consent;
- Introduction of a Consent Privacy Impact Quantification;
- A technical and conceptual implementation of Dynamic Consent;
- Presentation of a prototype that implements the concepts.

The remainder of this article is structured as follows: Section 2 considers related work in this area. Next, Section 3 introduces and defines the term Sovereign Digital Consent. After this Sections 4 and 5 present the key elements of Sovereign Digital Consent with CPIQ and the implementation of Dynamic Consent. Section 6 discusses the introduced concepts and presents the final result of these, while Section 7 concludes the article.

2. Related Work

There is a multitude of papers that addresses several topics regarding digital consent, privacy quantifications and dynamic consent.

2.1. Digital Consent

The work of Bialke et al. [6] introduces gICS—a generic informed consent system. This tool provides a way to define a consent template for generic consent that can then be printed out, given consent by the individual, and then be processed with a digital documentation of this paper-based consent. While the goal of this paper is to introduce a way to faster process a batch of consent by using a dedicated software, there is still paper-based consent involved. The process of giving and collecting the consent is still an analogue process.

Another publication from Schreiweis and colleagues [7] looks at which techniques can be used to create and enforce digital consent. The authors discuss the usage of Basic Patient Privacy Consent (BPPC) and Advanced Patient Privacy Consent (APPC) which are both standards by the Integrating the Healthcare Enterprise (IHE) initiative. The paper shows that APPC is a promising technique to use in combination with XACML. However, there are no concrete implementations in this paper.

In general, there are many different papers regarding digital consent. None of them focus on a privacy quantification or a patient-centered digital consent implementation such as ours.

2.2. Privacy Quantification

In terms of the privacy quantification, there are publications considering potential re-identification attacks and the privacy risk of data sharing. In "Browsing Unicity: On the Limits of Anonymizing Web Tracking Data", the authors show how hard it is to make tracking on the internet unlinkable to an individual by using generalization techniques, while still preserving utility of the data [8]. The study shows that generalization alone is not sufficient because to gain a good degree of anonymization, the data have to be so much generalized that they are nearly unusable.

Another re-identification risk quantification is performed by Montjoye et al. [9]. They analyze a dataset cellular phone location of 1.5 million people over 15 months. Even when anonymized, with these data the authors were able to identify any individual with 95% accuracy by using only four data points. Considering that common locations such as home, work or other social behaviors lead to the frequent occurrence of the same locations, this can involve a high privacy risk. The study also underlines that most institutions, and especially the individuals in such datasets, are not aware of the potential risk.

One of the quantification-related papers is by Veeningen et al. [10]. This work looks at an exemplary digital health care infrastructure, where different parties hold different data about an individual. In addition, some parties receive raw data and some—anonymized data. By using coalition graphs that show potential linkages of data, it can be shown which merging of data can potential be harmful. In addition, the coalition graphs of different infrastructures can be directly compared to show which one has a lower risk of potential combinations that could lead to data linkage. In contrast to our work, this paper only considers potential risks for data processors and does not consider an individual patient sharing data.

Another quantification is performed in "Quantifying the Costs and Benefits of Privacy-Preserving Health Data Publishing" [11]. The publication describes a cost model for health data. It considers the anonymization costs and the impact of anonymization in terms of data utility. This is contrasted with the cost of a potential data breach and the expected fine for such a privacy leak. Those factors are considered by different algorithms that calculate the privacy and utility trade-off. In this way, the optimal ratio between the two main factors can be found and a cost-optimal data anonymization method can be chosen. This approach is not suitable for our concept because it also looks at the topic from the data processor/broker point of view and does not consider the individual sharing its data and the impact on it.

Tesfay et al. [12] present an approach to evaluate and quantify privacy policies. A machine learning (ML) model is used to rate the policies on a traffic light scale. While the ML technology looks promising, it is not clear whether it generalizes to the medical domain. Furthermore, personal preferences that could lead to a higher risk tolerance are not considered.

While there are many other publications on the topic of a privacy quantification, there is, to the best of our knowledge, no publication that focuses on the privacy impact of declarations of consent from a patient's point of view.

2.3. Dynamic Consent

The related work on Dynamic Consent is mostly about implementations using blockchain technology.

One example of this is the paper by Mamo et al. [13] where the authors discuss a web portal for the Malta Biobank. This portal provides an overview of ongoing research projects and which individuals participate in which study. An interesting piece of information is that the potential participants must complete a test before they can give consent, to prove that they understand to what they are consenting. The portal also offers comprehensive ways to give consent to participate in a project. In addition, consent choices could be altered at any time. From a technological perspective, the portal works with a blockchain technology to store information about the consent.

The other blockchain-based solution is presented in "Patient Consent Management by a Purpose-Based Consent Model for Electronic Health Record Based on Blockchain Technology" [14]. This method of consent management is based on role-based access control (RBAC) which can also be fully mapped to the XACML approach used in this article. The consent itself consists of a role, IDs of doctors who have access, allowed actions and the purpose as hierarchical structure. The consent is stored in a blockchain where it can be updated or altered. However, it remains an open question whether the blockchain technology is really an advantage for this technology and if not, more traditional approaches with XACML, such as the one in our work, should be considered.

3. Sovereign Digital Consent

In this section, the concept of Sovereign Digital Consent is introduced. At first, different types of consent are presented; after this, we will define the requirements for Sovereign Digital Consent. Finally, a formal model for this form of consent is introduced which will be later used for our concrete implementations. As stated before, this is a multidisciplinary topic and should be considered from different research areas. This paper focuses on the technical point of view.

3.1. Types of Consent

A recent trend which is often used in the context of medical research or biobanks is the so-called Broad Consent [15]. The main characteristic of this type of consent is that patients grant access to their medical records to research projects that have yet no concrete purpose besides a general one such as "medical research". This should help with the issue that, for example, with Big Data analyses, the concrete purpose cannot be known before the data are collected. Besides the benefits for the researchers, there are concerns about whether such a process is ethical and legal [16]. With the GDPR in the EU, there are specific articles that forbid the processing of personal medical data without an explicit and specific consent (See GDPR Art. 9, Art. 4, and Art. 5), which renders Broad Consent de facto not GDPR compliant. Recital 33 of GDPR acknowledges these issues by allowing consents to certain areas of science that have high ethical standards for the research. However, there are initiatives that are trying to create the legal foundations for Broad Consent, such as the German Medizin Informatik Initiative (MII), which created a patient consent form for research purposes, which is accepted nationwide by the local data protection regulators (https://www.medizininformatik-initiative.de/en/medical-informatics-initiative-given-green-light-germany-wide-patient-consent-template-document (accessed on 11 January 2022)) [17].

Dynamic Consent, on the other hand, aims to strengthen patient involvement in the research process through active participation. The core principle is to alter already given consent in the future [18]. At first, the patient can give a broad access to their personal medical data or only choose to share a smaller subset of this. This consent can be modified at any time by using modern technologies such as dedicated apps or web portals. Through this, the purpose of the research can be dynamically adjusted during the project duration and the patient is able to react. In addition, the patient can also be kept in the loop about the research more easily thanks to modern technology. This is one of the biggest issues with paper-based consent [19]. Several publications describe the process of using dynamic consents for biobanks and see many potential benefits [20,21]. In contrast to Broad Consent, Dynamic Consent can be compliant with the GDPR if it is explicit about its purpose.

While Broad Consent and Dynamic Consent are mostly concepts, they need Digital Consent technologies to be implemented. A paper-based consent would be sufficient for Broad Consent but is not feasible for Dynamic Consent. There are already technologies for digital consent such as BPPC, APPC or Fast Health Care Interoperability Resources (FHIR) Consent, but most of them are not yet widely used.

To summarize, Table 1 compares Broad Consent and Dynamic Consent.

As described before, Broad Consent is a mostly paper-based type of consent. Dynamic Consent, on the other hand, must be a digital consent because it would not be manageable as paper-based process. Autonomy for Broad Consent is rather passive, whereas Dynamic Consent can be an active, empowering, and ongoing participation through the usage of digital technologies. Regarding purpose, Broad Consent has the characteristic that the purpose is hardly restricted and is often for larger research areas rather than a specific purpose. This is one of the big advantages of Dynamic Consent where consent is fine and granularly customizable after the initial decision. The affected individual can review and revise decisions at any part of the process. As for the consent format, the participation for Broad Consent is also mostly paper-based and analogue by using flyers or medical briefings. Here, the digital format of Dynamic Consent also brings new ways of participation with notifications about the research process and on-going questionnaires. This is also the case for informed consent. The information for Broad Consent can be customized depending on the research project or biobank. The digital platform that should be used for Dynamic Consent brings a whole new set of opportunities with specified information for each participant or opportunities to connect with other participants.

Table 1. Comparison of Broad and Dynamic Consent [19].

Feature	Broad Consent	Dynamic Consent
Consent	Paper-based	Digital
Autonomy	Passive	Active, empowered, ongoing
Purpose limitation	Purpose hardly restrictable	Consent customizable after purpose is known
Timing of decision	Linear flow, one-time decisions	Iterative processes supported, ability to review and revise decisions
Research participation	Flyers and leaflets, conversation form consultant to participant	Option to search for participation opportunities, dialogue with researchers, notification of opportunities
Informed Consent	Biobank specific leaflets and resources	Participant-specific information, opportunities to share experiences and questions with community of participants

3.2. Requirements for Sovereign Digital Consent

Based on the legal requirements on the GDPR, the different models of Digital Consent and the technical requirements for a consent enforcement system, we define the requirements for what we call Sovereign Digital Consent.

- **Req. 1:** Controlled by the affected individual:
  The affected person needs to be able to control the sharing of their data at any time. Sovereignty does not only mean a binary choice, i.e., sharing data or not sharing, it should rather be a fine granular choice so that only a subset of data can be shared by consenting. This also resembles the main principle of Dynamic Consent.
- **Req. 2:** Automatic consent enforcement:
  From a technical perspective, it must be possible that a Sovereign Digital Consent can be enforced automatically. Many approaches regarding digital consent only provide a digital consent form which still must be enforced manually. This weakens the impact of the digital consent and can also be a very time-consuming process especially with fine granular consent choices. There are also several related studies, including our own, that describe such mechanisms [22,23].
- **Req. 3:** Informed decision:
  A digital consent decision can be complex, especially if an individual wants to partici-

pate in different research projects from different research institutions. Besides the legal requirements for an informed consent decision, there needs to be a way to evaluate different data sharing requests, considering personal preferences or expected benefits from a research project. Such an approach should also focus on the potential privacy risks data sharing can have. For the user to be sovereign, the user should understand those consequences of a consent decision.

- **Req. 4:** Proactive:
  A sovereign digital consent should also be proactive in a way that the affected person can still comprehend. Proactive means that a patient can give access to whole categories of data. In addition, this can also include future data that still need to be collected. This requirement provides the broad access of Broad Consent, while still being under the user's control by limiting access to certain medical categories.
- **Req. 5:** Research-friendly:
  While the main idea of Sovereign Digital Consent is to empower the affected person, it should still be research-friendly. This can be realized by fulfilling requirements such as automatic enforcement and by the advantages for a researcher of proactive consent. In addition, Sovereign Digital Consent should provide a comprehensive research interface so that requesting data is also an automatic process. While this is rather a technical requirement, this should be still considered for Sovereign Digital Consent.

3.3. Formal Consent Model

For the consent evaluation and the implementation of Dynamic Consent, a formal foundation for Sovereign Digital Consent is needed. To provide this, the requirements from Section 3.2 are combined with the formal consent model of the German "Elektronische Patientenakte" (ePA) [24]. The ePA is the nationwide personal health record in Germany and was introduced in 2021. It acts as central storage for every medical record of a patient. The consent concept of the German ePA considers the treatment scenario where data are shared with a medical professional. To also include our secondary usage scenario, we consider the consent template for research projects created by the German "Medizininformatik-Initiative" (MII) [25]. This template has high relevance since it is accepted and evaluated by the German Federal Commissioner for Data Protection and Freedom of Information [17]. For the description format of the research categories, we used the widely used medical terminology SNOMED CT (https://browser.ihtsdotools.org/? (accessed on 11 January 2022)), which provides a rich and digital usable terminology for medical data. Through a systematic extraction of the most relevant information and combination of those sources, we created this formal consent model for Sovereign Digital Consent.

Table 2 shows the identifiers of the model for Sovereign Digital Consent. Every consent belongs to a subject S_i who can be a patient or a legal guardian. The consent is always related to a so-called authorized party AP_j which can be a researcher or a research project. As described before, the SNOMED CT terminology is used for the consent. A subject defines a set of resources or categories with SNOMED codes to which the authorized party can have access. This will be stored in the set $pRes_j$ for permitted codes. S_i can also restrict access to certain codes with the set $dRes_j$. Those identifiers will then be combined in the consent object $Co_{i,j} = (AP_j, pRes_j, dRes_j)$ which is the consent of subject S_i for the authorized party AP_j. Those identifiers are sufficient for a basic scenario of a research project.

For a sophisticated consent privacy impact quantification, more identifiers are needed to describe a research project in more detail. Every research project should have a specific purpose PU. This alone is required by the GDPR, but a broad consent where the specific goal besides medical research is not yet clear is also possible. Moreover, a research project can have a personal benefit PBE or a social benefit SBE. Those benefits are combined in the benefit identifier $BE = [PBE, SBE]*$. It is also possible that there is more than one promised benefit. The most important part regarding privacy is the degree of anonymization that is used for the different stages of the research project. This is defined by DA_D, which is the degree of anonymization for the data processing stage. A degree of anonymization can be

a value for k-anonymity or l-diversity. Those are common data privatization technologies that are often used for medical research [26,27]. Additionally, the processing security PS also plays a role. We divided this in three categories with $low|medium|high$. This results in $D = (PS, DA_D)$ for the data processing. Another privacy impacting stage is the potential publication of research results. There is also a degree of anonymization DA_{PUB} for the published data. The publication can then be summarized as $PUB = ((false|true), DA_{PUB})$ with the boolean value indicating whether there is a publication or not. The communication of a research project is, besides the publication, a large part of the transparency. Information $I = (false|true)$ indicates whether there is information about the project available. Transparency then is the combination of information and publication $T = (I, PUB)$. All those identifiers can then be summarized in the research information $RI = (PU, BE, D, T)$ to describe the additional research information needed for a consent privacy impact quantification.

Table 2. Identifiers for the Sovereign Digital Consent model.

Identifier	Explanation
S_i = Patient | Legal Guardian	Subject
AP_j = [Researcher]	Authorized Party
$pRes_j$ = [SNOMED CT code]	Set of all SNOMED CT codes to which AP_j has access
$dRes_j$ = [SNOMED CT code]	Set of all SNOMED CT codes to which AP_j can have no access
$Co_{i,j} = (AP_j, pRes_j, dRes_j)$	Consent of Subject i for Authorized Party j
PU = [Purpose] | Broad Consent	Research purpose
PBE = [Personal Benefit]*	Personal Benefit
SBE = [Social Benefit]*	Social Benefit
BE = [PBE | SBE]*	Benefit
DA_D = (k-Anonymity, l-Diversity)	Degree of anonymization for D
PS = Low | Medium | High	Processing security
D = (PS, DA_D)	Data processing
DA_{PUB} = (k-Anonymity, l-Diversity)	Degree of anonymization for P
I = ($false$ | $true$)	Information
PUB = (($false$ | $true$), DA_{PUB})	Publication
T = (I, PUB)	Transparency
RI = (PU, BE, D, T)	Research information

4. Consent Privacy Impact Quantification

To fulfill **Req. 3** from Section 3.2, the formal consent model introduced in Section 3.3 is used to provide a consent privacy impact quantification. This can help the affected individual to make an informed consent decision based on personal preferences in terms of acceptance and risk. The combination of those factors provides the quantification for which we will also present a prototypical implementation and an evaluation of the approach.

4.1. Quantification

The quantification is divided in two main factors: risk and acceptance. Risk provides a probability which will be weighted according to the acceptance factors, resulting in the consent privacy impact quantification.

4.1.1. Risk Probability

To understand the risk, an attacker model is needed to describe the potential risk for the shared data.

Definition 1 (Attacker model)**.** *The attacker as a third party can access all publicly available data of a patient. The attacker has knowledge that a patient is participating in a research project. The attacker does not combine knowledge from a potential publication with a data leakage. The attacker tries to learn about a person's sensitive data. This can be reached through re-identification attacks*

on the dataset by linking sensitive data. If the attacker completes the correct assignment, they are successful. The likelihood of a correct linkage is the likelihood of potential damage for the victim.

The possibility for the attacker to gain access to the data depends on two vectors. One way is that the attacker obtains the data through a data leakage. The probability of such a leakage depends on the processing security PS, which can have three different levels in our model. While this does not provide an accurate or continuous scale for data leakage, CPIQ is designed to provide a simplified but exhaustive model. Therefore, we considered three levels of processing security, which can be high, medium or low. It remains to be noted that a deeper examination needs to be conducted to correctly evaluate such a value.

Definition 2 (Data leakage probability DLP)**.** *DLP provides the probability that a data leakage can occur depending on the processing security PS.*

$$DLP = P(Dataleakage) = \begin{cases} 0.75 \; if \; PS = low \\ 0.5 \; if \; PS = medium \\ 0.25 \; if \; PS = high \end{cases}$$

It remains to be noted that even a low processing security results not in 100% risk probability, but rather in a very high probability such as 75%. The same applies to high processing security. The other way for the attacker to gain access to potential sensitive data is through a potential publication. As described in the model, the publication factor is a binary value that results in a 0% or a 100% probability.

Definition 3 (Publication factor PF)**.** *PF provides a binary value depending on whether there is a publication of results or not.*

$$PF = P(Publication) = \begin{cases} 1 \; if \; PUB = true \\ 0 \; if \; PUB = false \end{cases}$$

The attacker's re-identification attempts start if they have gained access to the data. How successful those attacks are depends on the degree of anonymization of the gained data. This can be either DA_D or DA_{PUB}. For CPIQ, only l-diversity will be considered as anonymization technology, since it has a stronger definition than k-anonymity. In addition, the number of resources an individual has shared also plays a role, as this will weaken the anonymization through l-diversity when there are more than one resources. The total relinking probability will be defined by the sensitive attribute exposure probability SAEP.

Definition 4 (Sensitive Attribute Exposure Probability $SAEP$)**.** *$SAEP$ is a worst-case estimation on the attribute linkage probability with regard to the used anonymization degree and the shared resources of an individual.*

$$SAEP = Min(1, \frac{|R|}{l})$$

where $|R|$ is the number of resources of a victim in the dataset and l is the level of l-diversity of the dataset.

Combining the previous mentioned probabilities with the corresponding $SAEP$ results in the following definitions.

Definition 5 (Re-identification probability due to data leakage, RPD)**.** *RPD calculates the data leakage probability potentially caused by a data leakage.*

$$RPD = P(Damage_{Data\ leakage}) = DLP * SAEP_{DP}$$

Definition 6 (Re-identification probability due to publication, *RPP*)**.** *RPD calculates the data leakage probability potentially caused by publication of data.*

$$RPP = P(Damage_{Publication}) = PF * SAEP_{PUB}$$

CPIQ assumes that *RPD* and *RPP* are independent events because a research project would potentially use different anonymization degrees for a publication than for internal data storage. However, in a real-world scenario, this independence could vanish if an attacker combines knowledge of a previous data publication. For the sake of simplicity, Definition 2 excludes this case.

In addition to the risk probabilities *RPD* and *RPP*, CPIQ requires the GDPR as legal basis for the data processing. If data are processed outside the European Union, they should have at least the same protection level as the GDPR. If this is not the case, this will be considered as a 100% risk.

Definition 7 (GDPR Non-Compliance Probability Factor *GNF*)**.** *GNF considers the risk probability if data is processed without regulations such as GDPR.*

$$GNF = \begin{cases} 1 \textit{ if processing location does not have a GDPR} \\ \hookrightarrow \textit{equivalent regulation.} \\ 0 \textit{ else} \end{cases}$$

Those probabilities are combined by using their complementary probabilities of occurrence to treat them as individual elements. This results in the Total Re-Identification Risk Probability *TRRP*.

Definition 8 (Total Re-Identification Risk Probability *TRRP*)**.** *TRRP is the total risk which considers all previously introduced definitions.*

$$TRRP = P(Damage_{Total}) = 1 - ((1 - RPD) * (1 - RPP) * (1 - GNF))$$

4.1.2. Acceptance Factors

In addition to considering the privacy risk, CPIQ also wants to review acceptance factors that could balance the risk factors of a research project. A major factor for this is the purpose of the project. We consider a specific purpose to lead to higher acceptance than using a more general broad consent purpose, which means that a research project has not yet defined a specific purpose besides a general one such as “Medical research”.

Definition 9 (Purpose *PU*)**.** *PU indicates whether a research project has a specific purpose.*

$$PU = \begin{cases} 0 \textit{ if the project uses data within a broad consent} \\ 1 \textit{ else} \end{cases}$$

Another major factor depends on whether the data sharing individual expects personal benefits in terms of their specific condition. In addition to this, there can be also a benefit for the general society. Both factors are represented as binary values for our model.

Definition 10 (Social Benefit *SBE*)**.** *The acceptance factor SBE indicates whether a project has at least one potential social benefit.*

$$SBE = \begin{cases} 1 \textit{ if the project has one or more social benefits} \\ 0 \textit{ else} \end{cases}$$

Definition 11 (Personal Benefit PBE)**.** *PBE indicates whether a research project could have potential personal benefits.*

$$PBE = \begin{cases} 1 \textit{ if the project has one or more personal benefits} \\ 0 \textit{ else} \end{cases}$$

Additionally, the consent model described transparency T as combination out of information I and the publication factor PUB. Both are considered binary acceptance factors. Finally, trust TR is the main foundation for data sharing to a research project. CPIQ identifies trust in the research institute as a strong requirement so it should be always 1. All those factors are combined in the so-called acceptance vector $\overrightarrow{AV}$.

Definition 12 (Acceptance vector $\overrightarrow{AV}$)**.** *The acceptance vector $\overrightarrow{AV}$ is an instance of all previous defined acceptance factors. This vector can be generalized and extended with more or other acceptance factors.*

$$\overrightarrow{AV} = \begin{pmatrix} Purpose \\ Personal\ Benefit \\ Social\ Benefit \\ Information \\ Publication \\ Trust \end{pmatrix}$$

However, for all these factors, the personal weighting is highly subjective and could depend on personal preferences. This weighting is defined by the relevance level Rel.

Definition 13 (Relevance Level Rel)**.** *Rel weights the impact of an acceptance factor on a three-level scale.*

$$Rel = low \mid medium \mid high$$

The relevance of each factor is reflected by the weights in the personal preference vector $\overrightarrow{PRV}$.

Definition 14 (Personal relevance vector $\overrightarrow{PRV}$)**.** *$\overrightarrow{PRV}$ indicates the relevance level Rel for any element of the acceptance vector $\overrightarrow{AV}$.*

$$\overrightarrow{PRV} = \begin{pmatrix} w_{Purpose} \\ w_{Personal\ Benefit} \\ w_{Social\ Benefit} \\ w_{Information} \\ w_{Publication} \\ w_{Trust} \end{pmatrix}$$

Sincw the result for acceptance as dot product of $\overrightarrow{PRV}$ and $\overrightarrow{AV}$ would exceed the range of the risk factor, the $\overrightarrow{PRV}$ needs to be normalized to a value between 0 an 1 to calculate the maximum reachable acceptance value $MRAV$. This can be performed by calculating the dot product of $\overrightarrow{PRV}$ and the eigenvector $\overrightarrow{e}$. Without this normalization, the acceptance and risk could not be put into relation in the final evaluation.

Definition 15 (Maximum reachable acceptance value $MRAV$)**.** *To normalize the acceptance based on the personal preferences, $MRAV$ is used.*

$$MRAV = \langle \overrightarrow{PRV}, \overrightarrow{e} \rangle$$

Finally, the total acceptance is a result by division of the dot product of $\overrightarrow{PRV}$ and $\overrightarrow{AV}$ with $MRAV$. $MRAV$ is used to normalize the result to a value between 0 and 1, as previously mentioned.

Definition 16 (Acceptance)**.** *The final acceptance score is defined using the personal and acceptance vectors and the MRAV.*

$$Acceptance = \frac{\langle \overrightarrow{PRV}, \overrightarrow{AV} \rangle}{MRAV}$$

4.1.3. CPIQ Score

Combining the Acceptance Factors with the Risk Probability results in the CPIQ quantification, which weights the different properties of a consent for research projects and the shared data.

Definition 17 (Consent-Privacy-Impact-Quantification (CPIQ))**.**

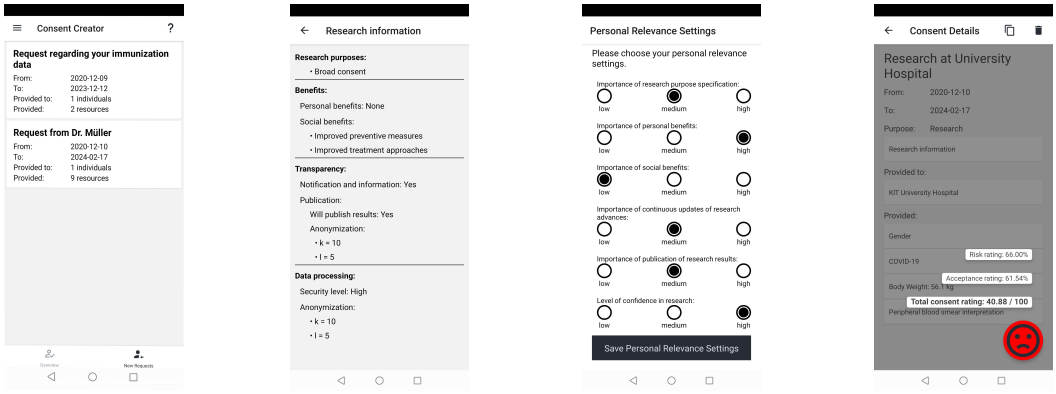
$$CPIQ = Acceptance * (\frac{L}{2} * (1 - \frac{1}{s})) + (1 - Risk) * (\frac{L}{2} * (1 + \frac{1}{s}))$$

where $s \geq 1$ *and* L *have maximum value.*

The higher the CPIQ value is, the higher will be the recommendation to consent to the research project. It remains to be noted that L and s are modifiable values that can be set according to general preferences. L is the maximum score CPIQ can reach. A good value for L would be 100 so that CPIQ has a percentage scale. s describes the ratio between risk and acceptance. The higher s is the higher is the weight of the acceptance. A $s = 1$ means that the acceptance has no impact on the evaluation, so a $s > 1$ is recommended. Furthermore, the maximum impact of acceptance is limited by $\frac{L}{2}$. In this case, 100% acceptance and 100% of risk led to a total of 50%. For our evaluations and tests, we used a ratio of 1:3 for acceptance and risk with $s = 2$.

4.2. Implementation

To evaluate CPIQ in a prototypical scenario, it was implemented in an existing consent management system developed by Fraunhofer IOSB. The consent management system provides a smartphone application called "Patient App" that lets users access their medical health records. They can see every piece of data that is recorded of them and is part of the system. In addition, they can also manage data sharing through consent management. The consent management considers the scenario of consent for data sharing to a specific party, such as a doctor for treatment, and the scenario of consent requests for an example of a research project. Figure 1a shows a screenshot of the consent requests view. CPIQ requires every consent request to provide the research information RI defined in Section 3.3. This information is then part of the consent format and can be displayed to the user, as the screenshot in Figure 1b shows. To fill the personal relevance vector $\overrightarrow{PRV}$ and give weight to each acceptance factor, the app asks the user for their preferences. The screen where the settings are requested is shown in Figure 1c. With the research information and the acceptance weights, CPIQ can now be calculated for consent. Figure 1d shows a CPIQ rating for an exemplary consent. This consent has a rather bad CPIQ score with 40 and so it is not recommended to share data with this project. In addition to the score, CPIQ uses a simple traffic light system indicated by the red smiley. This traffic light system shows a value of red from 0–50, yellow from 50–75 and green for scores above 75. It remains to be noted that further research needs to be done to find an optimal user interface element that supports the decision of an individual. Furthermore, CPIQ also shows a detailed overview of the risk and acceptance rating so the user can comprehend the evaluation and make an informed decision.



(**a**) (**b**) (**c**) (**d**)

Figure 1. Screenshots of the CPIQ implementation in the "Patient App" (Source: [28]). (**a**) Consent requests view. (**b**) Research information view. (**c**) Personal relevance setting view. (**d**) Consent privacy impact rating view.

4.3. Evaluation

To evaluate CPIQ, an edge case analysis on the acceptance factors is conducted, the risk probabilities and its assumptions are discussed, and the complete formula is evaluated. For the acceptance factors, different test scenarios were created that should be edge cases for CPIQ. Table 3 orovides an overview of the scenarios.

Table 3. Overview of different acceptance scenarios for edge cases.

	1		**2**		**3**		**4**		**5**		**6**	
Factor	$\overrightarrow{AV}$	$\overrightarrow{PRV}$	$\overrightarrow{AV}$	$\overrightarrow{PRV}$	$\overrightarrow{AV}$	$\overrightarrow{PRV}$	$\overrightarrow{AV}$	$\overrightarrow{PRV}$	$\overrightarrow{AV}$	$\overrightarrow{PRV}$	$\overrightarrow{AV}$	$\overrightarrow{PRV}$
PU	×	3	✓	3	✓	1	×	3	×	2	×	1
PBE	×	3	✓	3	✓	1	✓	3	✓	2	✓	1
SBE	×	3	✓	3	✓	1	✓	1	✓	1	✓	1
I	×	3	✓	3	×	3	✓	1	✓	1	✓	1
PUB	×	3	✓	3	✓	1	✓	1	✓	1	✓	1
TR	✓	1	✓	3	✓	1	✓	1	✓	1	✓	1
Rating		6%		100%		63%		70%		75%		83%

The first scenario shows the worst case, where each acceptance factors has the highest relevance, but no acceptance factors are provided by the research project. Trust is still required for a consent to be even considered. However, this leads to the expected low rating of 6%. In the second scenario, each factor has a high relevance weight and is also provided by the research project. As expected, this leads to a 100% acceptance rating. Scenario 3 shows how large the impact of a missing acceptance factor with a high weight can be. When every other factor is provided, but is only rated with a lower relevance, the highest-rated is missing. This leads to a mixed acceptance rating of 63%. In the fourth scenario, the impact of the relevance weight is shown by altering the relevance of an accepted factor from 1 to 3. This leads to a higher total rating of 70%. Scenario 2 shows that generally lower ratings help the total score to rise because the same provided acceptance factors from the previous scenario are only rated2 instead of 3, which leads to a 75% rating. Finally, the impact of a low overall preference can be shown in scenario 6.

For the risks factors, it remains to be noted that the l-diversity estimation provides a lower bound risk estimation. In the real world, it is not possible to claim the l-diversity value as a realistic damage occurrence probability. Many assumptions need to be made, for example, that the attacker really is certain that their victim is in the dataset or that a potential data leakage leaks the complete dataset and not only parts of it. However, the goal of CPIQ is to provide a model that rather overestimates potential risk than underestimating it.

The final formula of CPIQ weights the risk probabilities with the acceptance factors.

As a high acceptance should not completely neutralize a high risk, a weighting factor s is introduced. $s = 1$ means that only risk contributes to the final CPIQ score. The higher s is, the higher the weight of the acceptance will be. The maximum share of acceptance can be 50%. Figure 2 features a visualization with $s = 2$ and $L = 100$. This means that $CPIQ = 25 * Acceptance + 75 * (1 - Risk)$ and risk and acceptance are balanced 1:3. The dotted line shows the recommendation boundaries with everything above the upper line resulting in a good recommendation, everything between the first and the second line in a medium recommendation and everything below the second line in no recommendation. This visualization shows that no acceptance rating can outbalance a risk rating higher than 66%, which always results in no recommendation. With this desired effect, it is underlined that acceptance always plays a subordinate role in comparison to the risk.

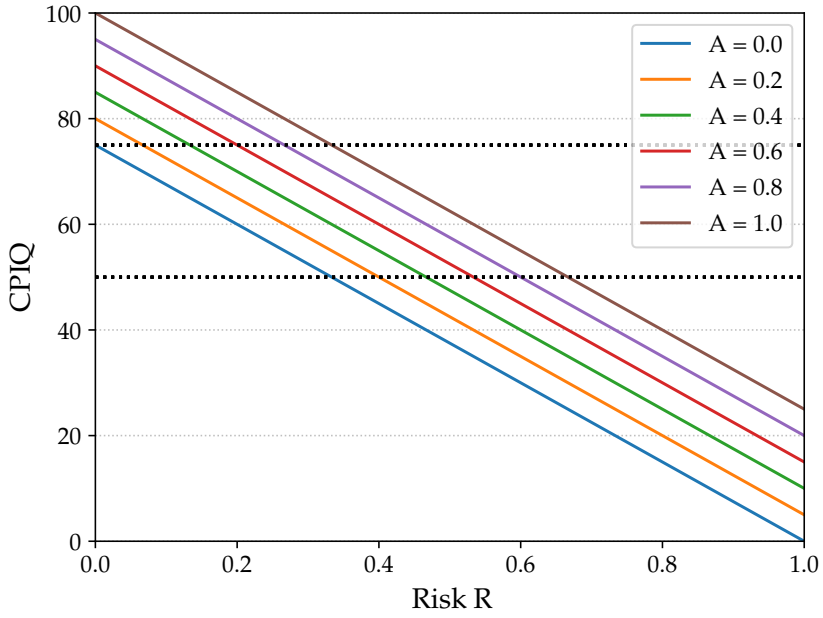


Figure 2. Consent privacy impact rating visualisation (Source: [28]).

5. Proactive Dynamic Consent

In this section, our implementation of proactive dynamic consent is introduced. With proactive dynamic consent, the requirements: **Req. 1** (patient controlled) **Req. 4** (proactive) and **Req. 5** (research-friendly) from the requirements for Sovereign Digital Consent in Section 3.2 are fulfilled. This section describes the medical data categorization that is needed for our approach and presents a Dynamic Consent implementation in XACML. To conclude a real-world prototype, it is shown that it is also used for an evaluation.

5.1. Medical Data Categorization

To summarize medical findings or assign a certain observation to a body region, medical data categorization is needed. This can be helpful for a researcher or doctor to describe sets of data needed for the research or treatment instead of defining specific observations. For the patient medical data categorization offers, there is a way to hierarchically sort data in a more comprehensible way rather than a list of all observation. While many medical data categorizations offer ways to locate findings to body regions, it is still an expert system that requires basic medical knowledge. This may not be the optimal solution for every patient, so more research is needed in the field for an optimal patient interface to work with medical data. However, data categorization is needed for Dynamic Consent to provide the dynamic part of only allowing specific data access, and for the proactive

part to enable proactive access to a whole category of data such as, for example, heart disease-related observation.

The two most common types of data categorization are International Statistical Classification of Diseases and Related Health Problems (ICD) by the World Health Organization (WHO) and SNOMED CT. The current version of ICD is ICD-10 (https://icd.who.int/browse10/2019/en (accessed on 11 January 2022)), which was introduced in 1994. ICD is a classification system for medical diagnoses and health problems. ICD-10 is revised regularly and was recently extended with classifications related to the COVID-19 pandemic. It categorizes diseases and health problems in 22 chapters and contains around 14,000 different codes.

Figure 3 shows an excerpt of chapter 22 which is also used for the emergency listing of COVID-19. The finest division of an ICD-10 code is a combination of one capital letter and 3 digits (e.g., U07.1 for COVID-19). There are country-specific extensions of ICD and the WHO provides an application programming interface (API) for automated queries on the classification.

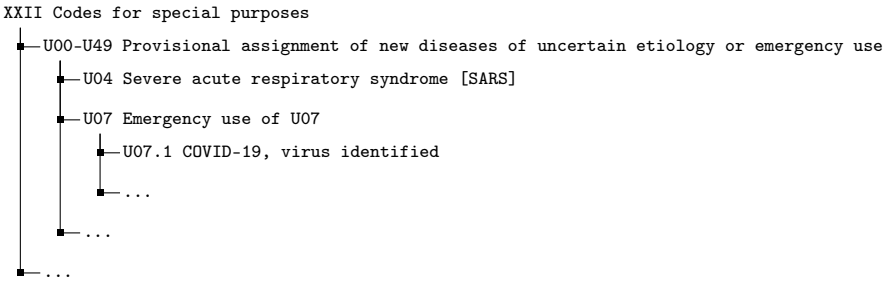


Figure 3. Example structure of chapter 22 of ICD-10.

Another type of categorization is SNOMED CT which has the aim to represent clinical contents as unambiguously and as precisely as possible. It is the most comprehensive, multilingual clinical healthcare technology in the world, with more than 350,000 concepts. As for ICD, there are country- and language-specific versions of SNOMED CT. The terminology consists of three core components: concepts, descriptions and relationships. A concept is a numerical code identifying a clinical term organized in hierarchies. These codes contain textual descriptions of the specific term. There exist different types of relationships between concept codes such as “associated”, “due to”, “has focus” or “is a”.

Figure 4 presents a visualization of the hierarchical structure for COVID-19 in the SNOMED CT terminology. It remains to be noted that each hierarchy also has other elements (e.g., influenza under viral diseases). SNOMED CT also provides a rich API through the Expression Constraint Language (ECL) to define queries for SNOMED CT codes and classifications.

Table 4 presents a comparison of the advantages and disadvantages of SNOMED CT and ICD-10. ICD-10 provides a relatively small categorization with a low hierarchy depth. Therefore, it seems easier to understand and handle. Additionally, it is available in printed and electronic versions. On the downside, it is not complete and has a fixed granularity. The hierarchy does not follow any logical aspects and is mono-hierarchical. There are also no cross-references, e.g., there is no connection between a disease and symptoms. SNOMED CT seems large and complex. The hierarchies can be very deep, so it must be said that SNOMED is not human-readable without digital technology. On the positive side, it is the most comprehensive medical terminology available. It has a variable granularity and offers relationships and cross references between concepts. This is also offered in a multilevel structure. Furthermore, it also includes the ECL query language for a powerful search in SNOMED CT. Following this comparison and the potential advantages of SNOMED CT compared to ICD-10, SNOMED CT is used for the implementation of Dynamic Consent in this paper.

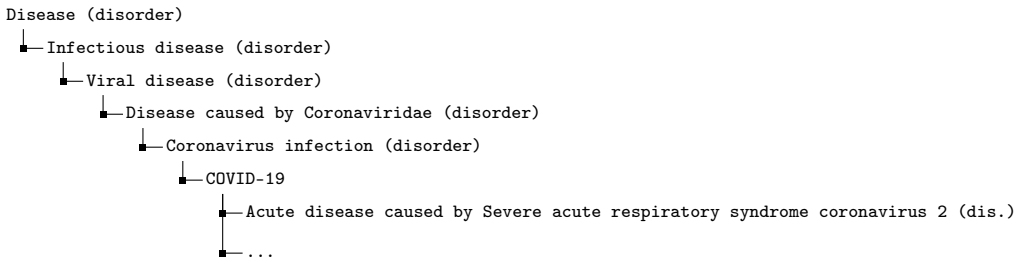


Figure 4. Example hierarchy in SNOMED-CT.

Table 4. Comparison of medical categorization procedures.

	ICD-10	SNOMED CT
Advantages	• Small • Low hierarchy depth • Easier to understand and to handle • Printed and electronic manuals available	• Most comprehensive medical terminology worldwide • Variable granularity • Relationships and cross-references between concepts • Multilevel structure • Mapping to ICD-10 • ECL for easy search through SNOMED CT
Disadvantages	• Not all diseases included • Fixed granularity • Outline does not correspond to medical or practical aspects • Mono-hierarchical structure • No connection between disease and symptom • Tendency of information loss • No clear distinction between diagnoses and symptoms	• Very large and complex • High hierarchy depth • Many terms not needed • Cannot be used without digital technology

5.2. Dynamic Consent with XACML

As foundation for the Dynamic Consent XACML enforcement, a formal enforcement model is needed. To enable hierarchical requests to categories, functions are required to search through the SNOMED CT hierarchy. The functions *descendantOf*, *ancestorOf* and *parentOf* are based on actual ECL functions. With $snoDesc(\cdot)$ and $snoAnc(\cdot)$, a list of all descendants or ancestors can be created for a SNOMED CT code. Additionally, the function $snoParent(\cdot)$ lists the parent of a SNOMED CT code. It remains to be noted that a SNOMED CT code can have more than one parent.

Definition 18 (Hierarchical Functions)**.** *The hierarchical functions reflect the parent children relationship between different SNOMED CT codes.*

$$\begin{aligned} snoDesc : SNO &\rightarrow SNO^A \\ sno_\alpha &\mapsto \{sno_{\alpha+1}, sno_{\alpha+2}, \ldots, sno_{\alpha+A}\} \end{aligned}$$

$$\begin{aligned} snoAnc : SNO &\rightarrow SNO^B \\ sno_\beta &\mapsto \{sno_{\beta-1}, sno_{\beta-2}, \ldots, sno_{\beta-B}\} \end{aligned}$$

$$\begin{aligned} snoParent : SNO &\rightarrow SNO, \\ & snoParent(sno_\gamma) = sno_{\gamma-1}, \\ & where \;\; sno_{\gamma-1} \in snoAnc(sno_\gamma) \cap snoDesc(sno_{root}) \cap snoDesc(sno_{req}) \end{aligned}$$

The consent decision is made with the function $conDec(\cdot)$, which returns *permit* or *deny* for a given SNOMED CT code. For the evaluation of the consent decisions, a root element must be set as SNOMED CT code (sno_{root}). This can be the highest SNOMED CT concept or any other code. In general, there are three cases.

1. No policy exists for the given code, and it is the root element then *deny* is returned;
2. A policy exists for the given code, the available policy is evaluated and the corresponding consent decision ($con(\cdot)$) is returned.
3. Otherwise, $conDec(\cdot)$ will be called recursively for the parent of the given code.

The function $pol(\cdot)$ checks whether a policy exists for the given code by checking the defined *permit* set ($pRes_j$) or the *deny* set ($dRes_j$).

Definition 19 (Consent Decision)**.** *The consent decision returns the access decision for a given SNOMED-CT code.*

$$conDec : SNO \rightarrow \{permit, deny\},$$

$$conDec(z) = \begin{cases} deny & , if \ \neg pol(z) \wedge z = sno_{root} \\ con(z) & , if \ pol(z) \\ conDec(snoParent(z)) & , otherwise \end{cases}$$

$$pol : SNO \rightarrow \{true, false\}, \quad pol(y) = y \in pRes_j \oplus y \in dRes_j$$

$$con : SNO \rightarrow \{permit, deny\}$$

The request function gets a tuple with the researcher and the requested categories sno_{req} and returns a set of different tuples. These tuples are the findings for which a patient has given consent to sharing.

Definition 20 (Request)**.** *The request function models a researcher request for data.*

$$Req : R \times SNO \rightarrow SNO^{L+1} \times \{permit, deny\}$$
$$(AP_j, sno_{req}) \mapsto snoDesc(sno_{req}) \cup \{sno_{req}\} \times con(sno_{req})$$

$sno_{req} :=$ *Requested category as SNOMED CT code*

The combination of those functions results in the Dynamic Consent enforcement.

Definition 21 (Dynamic Consent Enforcement)**.**

$$Let \ F = \{f_0, \ldots, f_h\} = Find_i \cap snoDesc(sno_{req}).$$

$$Req(AP_j, sno_{req}) = \begin{cases} \{(sno_{req}, conDec(sno_{req}))\} & , F = \emptyset \\ \{(f_0, conDec(f_0)), \ldots, (f_h, conDec(f_h))\} & , otherwise \end{cases}$$

The first case describes when no finding exists under the requested category yet. Therefore, the function $conDec(\cdot)$ is only executed for the requested category. Otherwise, the function $conDec(\cdot)$ is executed for all findings of the given category. To summarize the cases where permission gets granted, the following is defined:

$$Req(AP_j, sno_{req}) = \begin{cases} \{(sno_{req}, permit)\} & , conDec(sno_{req}) \in pRes_j \ \wedge \ F = \emptyset \\ \{(f_0, permit), \ldots, (f_h, permit)\} & , \forall f \in F : conDec(f) \in pRes_j \end{cases}$$

Permission is granted, either for the case that there exist no findings under the requested category, and the first policy of a predecessor of the requested category is a *permit* policy, or the category itself has a *permit* policy (Case 1). Otherwise, the first policies of a predecessor of all findings are all *permit* policies or they have a *permit* policy themselves (Case 2).

With this formal consent enforcement model XACML policies can be created. XACML is an attribute access control language where access decision is made according to the given attributes of the requested resource and the requesting subject [29]. For more details on XACML, this paper refers to the original sources. In addition, the here-presented policies are written in the XACML dialect Abbreviated Language For Authorization (ALFA), which is used to write shorter and more comprehensible XACML policies [30]. Listing 1 shows the policy structure.

Listing 1. Final policy structure in ALFA syntax.

```
namespace policyStructure {
        import policyStructure.attributes.*

        policyset patient {
                target clause parameters.patientId == "$patient_id$"
                apply denyUnlessPermit
                research_1
        }

        policyset research_1 {
                target clause parameters.researchId == "$research_1$"
                apply denyUnlessPermit

                policy _permitAccess {
                        target clause parameters.snomedId == "$snomed_id_1$"
                        apply denyUnlessPermit
                        rule permitAccess {
                                permit
                        }
                }

                policy _denyAccess {
                        target clause parameters.snomedId == "$snomed_id_2$"
                        apply denyUnlessPermit
                        rule denyAccess {
                                deny
                        }
                }
        }

}
```

Since the considerations are made per patient, a patient is set as a root `policyset`. With `target clause`, the targets get checked. This includes the patient, researcher or the SNOMED CT code. According to $Req(\cdot,\cdot)$, the permission is initially denied, so all *rule-* and *policy-combining algorithms* are set to `denyUnlessPermit`. Since a consent has separate sets for *permit* and *deny* per researcher, every researcher is set as their own `policyset` and the separate sets as a `policy`. A `policy` contains the SNOMED CT codes (the identifiers in Listing 1 act as placeholders for SNOMED codes in a real policy) as targets, for which the respective consent applies. Finally, the consent decisions are evaluated through the `rule` which is set in the `policy`. The necessary targets are defined as attributes in an additional file.

5.3. Implementation

As with CPIQ, the Dynamic Consent system is implemented in the prototypical research interface. Through the usage of SNOMED CT and XACML, a more sophisticated architecture is needed, which is shown in Figure 5.

There are two parties in the scenario: the patient and the researcher. All systems that store data or manage the access are located at a research data center which is shown by the dotted line. The prototypical research interface is the access point for the researcher where they can post requests and get the corresponding findings. As data server, a HAPI FHIR (https://hapifhir.io/hapi-fhir/ (accessed on 11 January 2022)) server is used which stores the patient data. HAPI FHIR is a state-of-the-art server implementation for the medical data format FHIR [31]. As a tool, the patient used the previously mentioned "Patient App" which is extended for use of Dynamic Consent. With this app, the data on the FHIR Server can be viewed and consent policies for research usage can be created. Those policies can be stored on the AuthzForce (https://authzforce.ow2.org/ (accessed on 11 January 2022)) server which is an open-source implementation of a XACML enforcement system. To add a new consent decision, medical data or categories as SNOMED, CT codes can be selected in the "Patients App" by explicitly setting the decision to *permit* or *deny* for the given code. To obtain the medical categories and hierarchical information for the findings of a patient the SNOMED CT terminology server Snowstorm (https://github.com/IHTSDO/

snowstorm(visitedon11/03/2021) (accessed on 11 January 2022)) is used, which can be used with ECL (https://confluence.ihtsdotools.org/display/DOCECL (accessed on 11 January 2022)) queries. The corresponding consent decision will then be stored as XACML policy on the AuthzForce server. A researcher can now make research requests via the prototypical research interface, to request medical data from patients with the specification of a SNOMED CT code. The research interface now uses the FHIR server, AuthzForce and Snowstorm to obtain data which the patient has consented to share. The researcher receives a list of the patients and their corresponding findings as result.

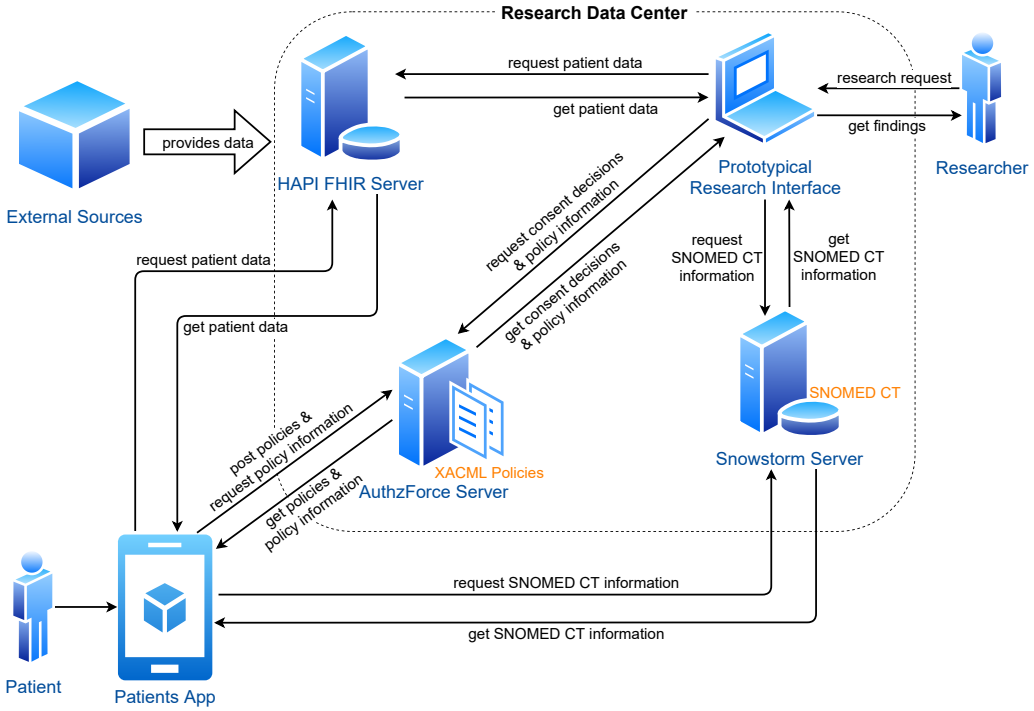
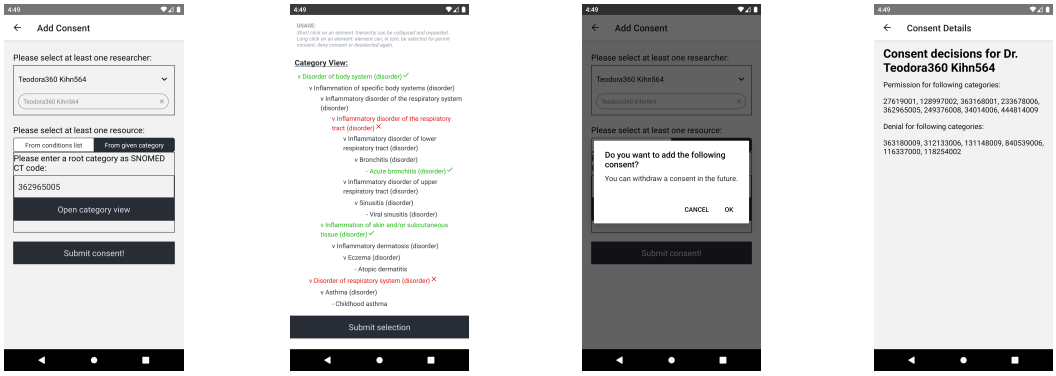


Figure 5. System architecture of the prototypical Dynamic Consent implementation.

Figure 6 shows the implementation in the "Patient App". The screen in Figure 6a displays the start view of the dynamic consent process. Here, the user can select the researcher for the consent and can open the category view to make their consent decision. This category view is shown in Figure 6b where the user sees their findings in a hierarchical tree. Certain findings or whole categories can be chosen to be shared or explicitly denied from sharing. When the user has made their decision, the consent must be submitted, as shown in Figure 6c. Finally, Figure 6d shows the consent details of the consent for a specific researcher or research project. The numbers are SNOMED CT codes. For a real-world application, those codes should be replaced with the human-readable description of the code.



(**a**) (**b**) (**c**) (**d**)

Figure 6. Screenshots of the Dynamic Consent implementation in the "Patient App". (**a**) Add dynamic consent: Category selection view. (**b**) Category view: Selection example. (**c**) Submit consent. (**d**) Consent details.

5.4. Evaluation

To evaluate the here-presented Dynamic Consent, it is evaluated against requirements that are derived from the properties of Dynamic Consent, as mentioned in Section 3.1. Additionally, the GDPR sets requirements for informed consent which are also examined.

Table 5 shows the properties of Dynamic Consent and the status of implementation in the presented technology. A checkmark (✓) indicates that this property is working properly. A dot (◯) shows that this property is not implemented.

Dynamic Consent requires that it be possible to grant access to a broad sample of data (**DC.1**). This can be allowed using the category selection in the category view, which is shown in Figure 6b. **DC.2** requires that a consent can be modified dynamically. The "Patient App" allows users to modify their consent at any time. This also satisfies **DC.10** because the consent modification occurs in real-time. The changes also come into effect immediately (**DC.11**) because every request goes through the AuthzForce server, which has always up-to-date polices. In addition, a Dynamic Consent requires that a consent can be given (**DC.3**) and can be taken back (**DC.4**) for every single finding individually. The category view also enables this functionality since a single finding can be permitted or denied for a consent. Furthermore, it should be possible to receive an overview of all shared data for certain authorized parties (**DC.5**). As Figure 6d shows, this is also possible with the "Patient App". Other requirements are that the purpose of the research or data usage should be always clear (**DC.6)**, that there should be a transaction log on data usage (**DC.7**), that participants can choose the level of information about the data usage (**DC.8**), and that a researcher has ways to inform patients (**DC.9**). While those are requirements for Dynamic Consent, these are not necessary enforcement specific requirements. Therefore, they were considered out-of-scope. However, with the technology of the "Patient App", most of these properties can be implemented. For example, a dedicated feedback function can fulfill **DC.8** and **DC.9**. Alternatively, a general data usage log can give the functionality required for **DC.6** and **DC.7**. Finally, the "Patient App" and the corresponding technology architecture can be considered as modern technologies so **DC.12** is fulfilled.

Table 6 shows an overview of the GDPR requirements for informed consent.

Article 4 No. 11 GDPR requires that a consent be the freely given wish of the subject's data sharing preference (**Reg.1**). As a patient can only add consent themself, the consent is freely given. Furthermore, with the explicit indication of *permit* and *deny*, the patient signalizes whether the consent is allowed or not. **Reg.2** and **Reg.3** set requirements regarding the withdrawal. It needs to be possible at any time and as easy as giving consent. Through the consent management system in the "Patient App", those requirements can be considered fulfilled. Withdrawal at any time is possible either completely or on a fine granular level. Furthermore, it uses the same system as giving consent, so it should not be

more complicated. In Article 5(1)(b) GDPR, it is required that the purpose of a declaration of consent is limited (**Reg.4**). While this can be favored by the implementation this is not enforced on a technological level. The last GDPR requirement derived from Article 9(2)(a) requires a consent to be explicit for one or more specific purposes (**Reg.5**). As with the previous requirement, this can be favored by our implementation by giving each consent a choice of purpose or a static one, but it is not enforced technologically yet.

Table 5. Evaluation of Dynamic Consent properties.

ID	Property	Status
DC.1	Grant access to a broad sample of data	✓
DC.2	Modify given consent choices dynamically	✓
DC.3	Consent can be given for every single finding individually	✓
DC.4	Consent can be taken back for every single finding individually	✓
DC.5	Overview of all shared data and with whom they are shared	✓
DC.6	Overview of the research uses of data	◯
DC.7	Records of all transactions and interactions in one place	◯
DC.8	Participants can determine frequency and scope of being informed	◯
DC.9	Researcher can inform participants	◯
DC.10	Modify given consent choices in real-time	✓
DC.11	Changes take effect immediately	✓
DC.12	Modify given consent choices with modern technologies	✓

Table 6. Evaluation of compliance with GDPR.

ID	Description	Article	Status
Reg.1	Consent according to GDPR	Article 4 No. 11 GDPR	✓
Reg.2	Withdraw consent at any time	Article 7(3) GDPR	✓
Reg.3	Withdrawal as easy as giving consent	Article 7(3) GDPR	✓
Reg.4	Purpose limitation	Article 5(1)(b) GDPR	◯
Reg.5	Given explicit consent for one or more specified purposes	Article 9(2)(a) GDPR	◯

6. Discussion

In this section, the presented technologies CPIQ and Proactive Dynamic Consent are discussed with a focus on the Sovereign Digital Consent requirements defined in Section 3.2.

Our presented system with the "Patient App" is a combination of those two technologies in one consent management system. As the "Patient App" and the corresponding consent management are under the control by the patient, **Req. 1** of the Sovereign Digital Consent requirements is fulfilled. The app offers a way for patients to access their data and to manage with whom they share the data.

According to **Req. 2**, a consent management system should enforce consent automatically. Considering the architecture in Figure 5, this is fulfilled. The prototypical research interface is the main entry point for a researcher to request data. These data always go through the enforcement workflow, so that only data with a permit consent are shared. This process is performed automatically when a researcher requests data of a certain category. With CPIQ, a consent evaluation in terms of privacy impact is offered. It uses the provided research information which should be included in the consent request of a research project. This helps the user to make an informed decision when sharing their data.

The third requirement (**Req. 3**) for Sovereign Digital Consent is an informed consent decision. With the addition of CPIQ, this is fulfilled.

The fourth requirement for a proactive consent (**Req. 4**) is addressed by the Dynamic Consent Implementation. With the choice of broad categories, a whole set of findings can be permitted for data sharing. Through consent to categories of medical data, findings can be included which will be made in the future and fall in the certain category. However, this comes with a set of questions. It must be clear to the user at any time that a category can include future findings. It cannot be expected that a user always understands which finding falls in which category. Obviously, it is no good solution to ask the user every time a new finding is made if this should be included in the consent. More work in this area is needed to search for ways for the user to always understand their consent decision and to make proactive consent traceable over time.

Finally, **Req. 5** defines that Sovereign Digital Consent should be research-friendly. With the prototypical research interface offered by our implementation, there is a way to directly access data to which a patient has given consent. This is possible at any time for an authorized researcher. Additionally, this is also a growing set of findings since the consent can be proactive. Through the automatic enforcement, there is also no more manual processing of the consent required. However, while the "Patient App" offers the functionality, there is yet no implementation of a consent request through the research interface, however, it should be rather easy to add. We think this functionality renders the here-presented implementation research-friendly.

While all those requirements are fulfilled, it remains to be said that this implementation of Sovereign Digital Consent is not complete. CPIQ has some limitations in terms of quantification and sets strict requirements to what is needed for a consent evaluation by defining a static set of acceptance and risk factors. The Dynamic Consent implementation itself needs further work in terms of user interaction and interfaces. Better ways need to be found for presenting the consent decisions to users and giving them tools to make fine granular consent choices with comprehensible interfaces.

In addition, CPIQ and the presented Dynamic Consent implementation should be more standardized to conform, for example, to the widely used Health Level Seven International (HL7) standards. However, our Dynamic Consent and CPIQ was built on the foundation of FHIR Consent as suggested by Mense et al. [32], yet it still lacks a deeper integration in the HL7 standard. For example, the Composite Security and Privacy Domain could be used in the future, as described by Blobel et al. [33]. Standardization could improve architectures such as the ones discussed by De Meo et al. [34]. With CPIQ, patients could also consider the privacy impact of sharing health care data and Dynamic Consent helps to fine granularly manage the data sharing.

All in all, those technologies also need to be looked at in an interdisciplinary way to gain insights from a legal, ethical and a researcher's perspective.

7. Conclusions

This paper presents the concept of Sovereign Digital Consent which is a patient-empowering and research-friendly implementation of digital consent. To define Sovereign Digital Consent, requirements are shown that can be fulfilled with the combination of a consent privacy impact quantification and an implementation of Proactive Dynamic Consent.

The consent privacy impact quantification CPIQ uses two main factors to calculate the individual privacy risk of sharing data to a research project. Those factors are risk and acceptance. For risk, CPIQ uses a worst-case estimation for data loss through leakage or a publication. Acceptance is based on research information and the personal preferences of an individual. To underline the focus on the privacy risk, acceptance is balanced by a certain factor to the risk. This results in the CPIQ evaluation formula which is also implemented in the "Patient App" prototype to show the real-world feasibility of the technology. In addition to its limitations, CPIQ is a first approach for a transparent and

comprehensible evaluation of declaration of consent by considering personal preferences of the affected person.

The Proactive Dynamic Consent implementation uses the medical data categorization terminology SNOMED CT to enable fine granular consent management. The patient can select certain medical data categories instead of selecting every finding individually. Additionally, it is possible to grant access to data dynamically and to only share certain data according to the personal preferences. Through the medical categorization, data can also be shared proactively, so that future findings can already be included for long-term research projects. This and the automatic enforcement of the consent make the technology research-friendly. An implementation of the consent with the access control language XACML in a sophisticated consent enforcement architecture is presented. Furthermore, the research interface and the implementation of the patient control in the "Patient App" is shown and successfully evaluated against requirements of the GDPR and requirements derived out of literature for dynamic consent.

While both technologies still have limitations, the combination can provide a solid foundation towards a Sovereign Digital Consent.

Author Contributions: Conceptualization, A.A.; methodology, A.A.; software, T.K. and M.H.; investigation, M.H. and T.K.; writing—original draft preparation, A.A.; writing—review and editing, E.K., M.H.; supervision, E.K. and J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported as a Fraunhofer Lighthouse Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

ALFA	Abbreviated Language For Authorization
APPC	Advanced Patient Privacy Consent
BPPC	Basic Patient Privacy Consent
CPIQ	Consent Privacy Impact Quantification
ePA	Elektronische Patientenakte
FHIR	Fast Health Care Interoperability Resources
GDPR	General Data Protection Regulation
HL7	Health Level Seven International
ICD	International Statistical Classification of Diseases and Related Health Problems
IHE	Integrating the Healthcare Enterprise
MII	Medizininformatik-Initiative
RBAC	Role-Based Access Control
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
WHO	World Health Organization
XACML	eXtended Access Control Markup Language

References

1. Commission, E. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). 2016. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed on 11 January 2022).
2. Grady, C. Enduring and Emerging Challenges of Informed Consent. *N. Engl. J. Med.* **2015**, *372*, 855–862. [CrossRef]
3. Bester, J.; Cole, C.M.; Kodish, E. The limits of informed consent for an overwhelmed patient: Clinicians' role in protecting patients and preventing overwhelm. *AMA J. Ethics* **2016**, *18*, 869–886.

4. Kim, K.K.; Joseph, J.G.; Ohno-Machado, L. Comparison of consumers' views on electronic data sharing for healthcare and research. *J. Am. Med. Inform. Assoc.* **2015**, *22*, 821–830. [CrossRef]
5. Budin-Ljøsne, I.; Teare, H.J.A.; Kaye, J.; Beck, S.; Bentzen, H.B.; Caenazzo, L.; Collett, C.; D'Abramo, F.; Felzmann, H.; Finlay, T.; et al. Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. *BMC Med. Ethics* **2017**, *18*, 4. [CrossRef]
6. Bialke, M.; Penndorf, P.; Wegner, T.; Bahls, T.; Havemann, C.; Piegsa, J.; Hoffmann, W. A workflow-driven approach to integrate generic software modules in a Trusted Third Party. *J. Transl. Med.* **2015**, *13*, 176. [CrossRef]
7. Schreiweis, B.; Bronsch, T.; Merzweiler, A.; Bergh, B. Implementing modular research consents using IHE advanced patient privacy consents. *Stud. Health Technol. Inform.* **2018**, *247*, 840–844. [CrossRef]
8. Deußer, C.; Passmann, S.; Strufe, T. Browsing Unicity: On the Limits of Anonymizing Web Tracking Data. In Proceedings of the 2020 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 18–21 May 2020; pp. 777–790.
9. De Montjoye, Y.A.; Hidalgo, C.A.; Verleysen, M.; Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. *Sci. Rep.* **2013**, *3*, 1376. [CrossRef]
10. Veeningen, M.; de Weger, B.; Zannone, N. Formal Modelling of (De)Pseudonymisation: A Case Study in Health Care Privacy. In *Security and Trust Management: Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 145–160.
11. Khokhar, R.H.; Chen, R.; Fung, B.C.; Lui, S.M. Quantifying the Costs and Benefits of Privacy-Preserving Health Data Publishing. *J. Biomed. Inform.* **2014**, *50*, 107–121. [CrossRef] [PubMed]
12. Tesfay, W.B.; Hofmann, P.; Nakamura, T.; Kiyomoto, S.; Serna, J. PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation. In Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics, Tempe, AZ, USA, 21 March 2018.
13. Mamo, N.; Martin, G.M.; Desira, M.; Ellul, B.; Ebejer, J.P. Dwarna: A blockchain solution for dynamic consent in biobanking. *Eur. J. Hum. Genet.* **2020**, *28*, 609–626. [CrossRef]
14. Tith, D.; Lee, J.S.; Suzuki, H.; Wijesundara, W.M.a.B.; Taira, N.; Obi, T.; Ohyama, N. Patient Consent Management by a Purpose-Based Consent Model for Electronic Health Record Based on Blockchain Technology. *Healthc. Inform. Res.* **2020**, *26*, 265–273. [CrossRef] [PubMed]
15. Caulfield, T.; Kaye, J. Broad Consent in Biobanking: Reflections on Seemingly Insurmountable Dilemmas. *Med. Law Int.* **2009**, *10*, 85–100. [CrossRef]
16. Petrini, C. "Broad" consent, exceptions to consent and the question of using biological samples for research purposes different from the initial collection purpose. *Soc. Sci. Med.* **2010**, *70*, 217–220. [CrossRef]
17. Medizininformatik-Initiative. Medizininformatik-Initiative Erhaelt Gruenes Licht Fuer Bundesweite PATIEnteneinwilligung. 2020. Available online: https://www.medizininformatik-initiative.de/de/medizininformatik-initiative-erhaelt-gruenes-licht-fuer-bundesweite-patienteneinwilligung (accessed on 11 January 2022). (In Germany)
18. Kaye, J.; Whitley, E.A.; Lund, D.; Morrison, M.; Teare, H.; Melham, K. Dynamic consent: A patient interface for twenty-first century research networks. *Eur. J. Hum. Genet.* **2015**, *23*, 141–146. [CrossRef]
19. Teare, H.J.; Morrison, M.; Whitley, E.A.; Kaye, J. Towards 'Engagement 2.0': Insights from a study of dynamic consent with biobank participants. *Digit. Health* **2015**, *1*. [CrossRef] [PubMed]
20. Mont, M.C.; Sharma, V.; Pearson, S. EnCoRe: Dynamic Consent, Policy Enforcement and Accountable Information Sharing within and across Organisations. Available online: https://www.hpl.hp.com/techreports/2012/HPL-2012-36.pdf (accessed on 11 January 2022).
21. Prictor, M.; Lewis, M.A.; Newson, A.J.; Haas, M.; Baba, S.; Kim, H.; Kokado, M.; Minari, J.; Molnar-Gabor, F.; Yamamoto, B.; et al. Dynamic Consent: An Evaluation and Reporting Framework. *J. Empir. Res. Hum. Res. Ethics* **2020**, *15*, 175–186. [CrossRef] [PubMed]
22. Appenzeller, A.; Rode, E.; Krempel, E.; Beyerer, J. Enabling Data Sovereignty for Patients through Digital Consent Enforcement. In Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, 29 June–1 July 2020; Association for Computing Machinery: New York, NY, USA, 2020. [CrossRef]
23. Verreydt, S.; Yskout, K.; Joosen, W. Security and Privacy Requirements for Electronic Consent: A Systematic Literature Review. *ACM Trans. Comput. Healthc.* **2021**, *2*, 1–24. [CrossRef]
24. Gematik. E-Patientenakte. Available online: https://www.gematik.de/anwendungen/e-patientenakte/ (accessed on 11 January 2022). (In Germany)
25. Medical Informatics Initiative. Patient Consent Form Template. Available online: https://www.medizininformatik-initiative.de/sites/default/files/2020-11/MII_WG-Consent_Patient-Consent-Form_v1.6d_engl-version.pdf (accessed on 11 January 2022). (In Germany)
26. Sweeney, L. *K*-Anonymity: A Model for Protecting Privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2002**, *10*, 557–570. [CrossRef]
27. Machanavajjhala, A.; Gehrke, J.; Kifer, D.; Venkitasubramaniam, M. L-diversity: Privacy beyond k-anonymity. In Proceedings of the 22nd International Conference on Data Engineering (ICDE'06), Atlanta, GA, USA, 18–21 May 2006; pp. 24–24. [CrossRef]
28. Appenzeller, A.; Kadow, T.; Krempel, E.; Beyerer, J. CPIQ—A Privacy Impact Quantification for Digital Medical Consent. In Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 29 June 2021–2 July 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 534–543. [CrossRef]

29. Standard, O. eXtensible Access Control Markup Language (XACML) Version 3.0. Available online: http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html (accessed on 11 January 2022). (In Germany)
30. OASIS. Abbreviated Language for Authorization (ALFA) Version 1.0. Available online: https://www.oasis-open.org/committees/download.php/55228/alfa-for-xacml-v1.0-wd01.doc (accessed on 11 January 2022). (In Germany)
31. Braunstein, M.L. *Health Informatics on FHIR: How HL7's New API Is Transforming Healthcare*; Springer International Publishing: Cham, Switzerland, 2018.
32. Mense, A.; Blobel, B. HL7 standards and components to support implementation of the European general data protection regulation. *Eur. J. Biomed. Inform.* **2017**, *13*, 27–33. [CrossRef]
33. Bernd, B.; Ruotsalainen, P.; Lopez, D.; Gonzalez, C. How to Use the HL7 Composite Security and Privacy Domain Analysis Model. *Int. J. Biomed. Healthc.* **2015**, *3*, 12–17.
34. De Meo, P.; Quattrone, G.; Ursino, D. Integration of the HL7 standard in a multiagent system to support personalized access to e-health services. *IEEE Trans. Knowl. Data Eng.* **2010**, *23*, 1244–1260. [CrossRef]