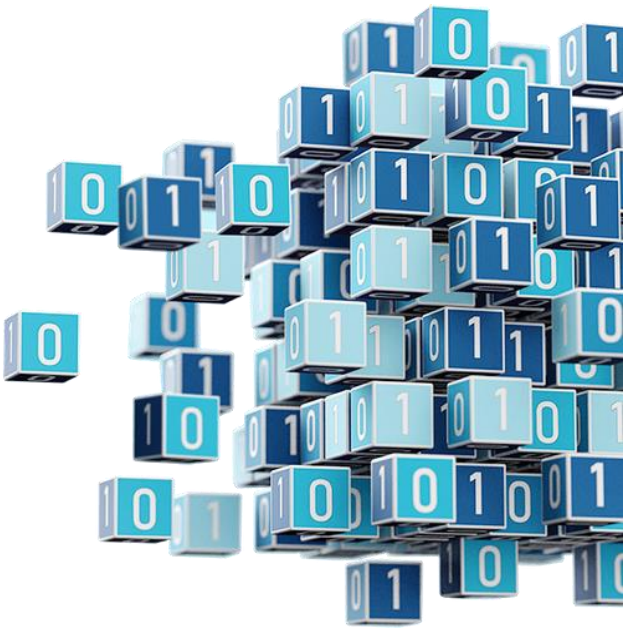
FeatureCloud

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Privacy preserving federated machine learning and blockchaining for reduced cyber risks in a world of distributed healthcare



Deliverable D6.3
Selected smart contract mechanism featuring user rights management

---

Work Package
WP6 Blockchains and user right management

Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078. Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

Document information

Grant Agreement Number: 826078			**Acronym: FeatureCloud**	
Full title	Privacy preserving federated machine learning and blockchaining for reduced cyber risks in a world of distributed healthcare			
Topic	Toolkit for assessing and reducing cyber risks in hospitals and care centres to protect privacy/data/infrastructures			
Funding scheme	RIA - Research and Innovation action			
Start Date	1 January 2019	**Duration**	60 months	
Project URL	https://featurecloud.eu/			
EU Project Officer	Christos MARAMIS, Health and Digital Executive Agency (HaDEA) - Established by the European Commission, Unit HaDEA.A.3 – Health Research			
Project Coordinator	Jan BAUMBACH, UNIVERSITY OF HAMBURG (UHAM)			
Deliverable	D6.3 Selected smart contract mechanism featuring user rights management			
Work Package	WP6 Blockchains and user right management			
Date of Delivery	**Contractual**	31/12/2021	**Actual**	16/12/2021
Nature	Report	**Dissemination Level**	Public	
Lead Beneficiary	05 SBA			
Responsible Author(s)	Aljosha Judmayer, Nicholas Stifter, Walid Fdhila, Fenghong Zhang (SBA Research)			
Keywords	Federated machine learning, blockchain, DLT, consent management, user rights			

History of changes

Version	Date	Contributions	Contributors (name and institution)
V0.1	10/11/2021	Requirements collection	Aljosha Judmayer, Walid Fdhila & Nicholas Stifter (SBA); Julian Matschinske (UHAM); Christof Tschohl, Walter Hötzendorfer & Markus Kastelitz (RI)
V0.2	26/11/2021	First draft	Aljosha Judmayer, Walid Fdhila, Nicholas Stifter & Fenghong Zhang (SBA)
V0.3	03/12/2021	Comments	Walter Hötzendorfer & Markus Kastelitz (RI); Philipp Schindler, Andreas Kern & Simin Ghesmati (SBA)
V0.4	10/12/2021	Draft	Aljosha Judmayer, Walid Fdhila & Nicholas Stifter (SBA)
V1	15/12/2021	Final version	Aljosha Judmayer, Walid Fdhila & Nicholas Stifter (SBA)
V1	16/12/2021	Final approval	Jan Baumbach (UHAM)
V1	16/12/2021	Submission	Miriam Simon (concentris)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Table of Content

1 Objectives of the deliverable based on the Description of Action (DoA) 5

2 Executive Summary 5

3 Introduction (Challenge) 6

4 Methodology 8

5 State of the Art 8

6 Requirements for Consent Management 10

7 System Design 16
7.1 High Level Process for Consent Management 16
7.2 Major challenges and research questions addressed 21
7.3 Implementation of the Process (Decentralized vs Centralized) 24

8 Conclusion 26

9 References 27

10 Table of acronyms and definitions 29

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

1 Objectives of the deliverable based on the Description of Action (DoA)

The objective of this deliverable is based on the description of action, which incorporates mostly aspects from Objective 2 and 3 and thus are part of the corresponding tasks 2 and 3 in work package 6. The corresponding task in this work package is Task 2: "User-rights management in Blockchains (SBA, TUM, MUG). This task integrates *user rights management*[1] into blockchain mechanisms.
Our selected process and design are based on the agreed technical and legal requirements, including required (meta-)information (MUG, TUM). This also includes research on the required and available information granularity (MUG) and, a study on usability with typical participants, including expert partners in hospitals, as well as patients (TUM).

2 Executive Summary

Deliverable (D6.3) defines and settles on the overall process and architecture for the consent management (previously called "user-rights management") and thus is essential for the future roadmap of WP6.

Note: In the process of working on the respective tasks associated with this work package, the term "*user rights management*" occasionally was misinterpreted to account for classical access permissions in computer systems. Therefore, to avoid further confusion, we have decided to adjust the terminology in respect to the proposal and adopt the term *"consent management*" from now on in the project and thus also in all further deliverables. This term more accurately captures the problem that needs to be tackled.

Moreover, we recapitulate on the detailed technical and legal requirements for consent management which have been heavily discussed with the legal experts within the FeatureCloud consortium and have just yet been finalized.
Given this set of requirements, we argue why they can only be fulfilled by posterior **detective measures**, i.e., an auditing mechanism with penalty (legal consequences), and why the violation of certain requirements cannot be prevented solely through automated measures, i.e., on a technical level, in the first place (see the introduction).
By enforcing an information technology supported auditing process, which functions as a detective measure to ensure that only consented data items are used, the chosen design approach also improves the overall integrity of the FeatureCloud framework. Hereby commitments on an immutable ledger serve as a time-ordered audit trail which acts as a deterrent for malfeasance, and thus also help tackle hard problems in the context of federated machine learning in general, e.g., *data poisoning*, *backdoor, or model poisoning attacks*, by rendering participants accountable for all of their actions.

[1] Now more accurately called **consent management,** as outlined in the executive summary.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

3 Introduction (Challenge)

A core design goal of FeatureCloud and fundamental characteristic in federated machine learning is the ability to train models without having to rely on a central data repository. Federated learning allows data providers to retain complete control over their data, and, assuming sufficiently effective measures to ensure that the locally trained models cannot leak private information, e.g., through applying differential privacy techniques, it would appear that the pressing privacy challenges that accompany machine learning on sensitive healthcare data can be largely addressed.

However, while such a federated learning approach may be able to ensure privacy at the level of the data provider and prevent leaking of sensitive information to other study participants or external observers, it does not address the issue whether the access and utilization of data by the data provider, as part of the federated learning, was legitimate, i.e., the necessary patient consent was satisfied for all data, or not.

Indeed, the very properties that seek to preserve privacy in federated learning appear to present fundamental challenges when it comes to ensuring accountability of the involved actors. In the context of scientific research regarding federated machine learning, this type of misbehavior of such data providers, i.e., clients, can be seen as a form of Byzantine failure. However, unlike some forms of adversarial actions, such as *data poisoning*, where the output trained model may retain traces of such manipulation, the unauthorized utilization or intentional omission of data based on some external property that is not inherent, or attached to the data itself, can generally not be inferred from the output model. In a similar manner, recent attacks such as *model poisoning* were shown to be a fundamental challenge in federated learning if the training data is non-i.i.d., rendering standard approaches for dealing with Byzantine behavior ineffective (Bagdasaryan et al., 2020).

While one may argue that a machine learning model trained on correct data for which there merely was no consent could retain its validity, it is impossible to know if the offending party had intentionally included this data because it was able to observe that a model trained without said data deviated further from the desired outcome, or not. Furthermore, from an ethical standpoint it appears questionable if one were to label a result as valid if it is based on illegitimate and unlawful practices. Hence, it would appear not only desirable but necessary to be able to verify the legitimacy of utilized training data in federated learning.
The same is true if analyzed from a data protection law perspective: Article 5 (2) General Data Protection Regulation (OJ 2016 L 119, p. 1; 'the GDPR') stipulates that each controller (e.g., each hospital) shall be responsible for, and be able to demonstrate compliance with, the principles of data processing set out in Article 5 (1) ('principle of accountability'). More precisely Article 5 (1) (a) GDPR mandates that personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject ('principle of lawfulness, fairness and transparency').

The key challenge that one faces in this regard is introducing a mechanism for proving that **only legitimate data was used** in federated learning without compromising the privacy gained through this technique. If the necessary consent for data were to be provided naively along with the trained model to justify its legitimacy, it would not only leak additional metadata, but also bring into question how an external verifier could be convinced that the claimed consented data matches the actual inputs used during the learning of the model.

In addition to these challenges, current consent management for clinical trials remains a very error prone and manual activity, which heavily relies on **consents given in paper forms** containing handwritten signatures of patients. For example, the United States Food and Drug Administration reports that clinical investigator inspections reveal almost 10% exhibit deficient recording of informed consent (Schüler and Buckley, 2014). As it is unrealistic that all involved actors and hospitals update

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

and digitize their associated processes in the foreseeable future, such legacy systems and processes have to be supported as well, accounting for the different pace at which certain hospitals will update their systems.

Another challenge is that most federated ML studies currently are **not** carried out in a **fully deterministic manner**, i.e., due to parallelization and other sources of non-determinism in the execution, given the exact same input and training parameters, they do not produce the exact same output model. This further complicates the ability to programmatically, or automatically verify the reproducibility of a study and thus that only consented input data has been used.

To address these challenges, currently the only readily available and practically feasible solution can consist of **detective measures**, as verifiable computing or Trusted-Execution-Environment (TEE) technology that would be applicable for the particular use case, (i.e., that offers the required functionality and scalability, as well as the ability to assess the authenticity of manually signed paper consents), are not yet sufficiently established and researched (Brundage et al., 2020).
Moreover, different data and storage formats might be used at different hospitals, further hampering automated solutions employing TEEs which would require uniform data formats across all hospitals as input.

Therefore, we propose an IT supported **auditing process** for consent management within the FeatureCloud framework which takes all these considerations into account and employs manual steps whenever needed. On a high level, the hospitals commit to the used input data together with the appropriate consents and the resulting output. Then a subset of hospitals is sampled provably random and a semi manual audit is carried out at these hospitals on-site. Hereby, the degree of automatization is dependent on the degree to which the processes at the respective hospital are digitized.

The advantage of our process is that also the integrity of the federated ML study can be audited as well, thus increasing the overall confidence in the outcome of the study as a whole. Incidentally the herein proposed approach also helps tackle a fundamental problem in the context of federated learning, namely data poisoning, in the context of Byzantine actors, which is deemed a fundamentally difficult problem to solve, especially in settings where data protection is an issue, and outliers cannot be easily detected and dismissed. For example, a recent SoK on advances and open problems in federated learning[2] mentions that for data poisoning, there is a possibility that the Byzantine threat model is too strong. Especially in the context of FeatureCloud where we could not expect independent and identically distributed data (i.i.d.) detecting anomalies in the models can be considered probably impossible (Bagdasaryan et al., 2020):

> *"It is provably impossible to detect anomalies in models submitted by participants in federated learning, unless the secure aggregation protocol incorporates anomaly detection into aggregation. [...] Even if anomaly detection could somehow be incorporated into secure aggregation, it would be useful only insofar as it filtered out backdoored model updates but not the updates from benign participants trained on non-i.i.d. data."*

[2] Peter Kairouz et al., *Advances and Open Problems in Federated Learning*, 2019

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

4 Methodology

To cope with the evolving requirements of FeatureCloud, (e.g., technical, legal, and privacy), and in order to enhance the different aspects and properties of the solution design proposed in previous deliverables of WP6, an incremental iterative methodology was adopted. This approach gives us the possibility to adjust, refine and review the auditing process as well as the consent management model, thereby addressing unresolved issues in previous iterations, and improving the integrity of the federated machine learning process as well as its compliance with the new requirements, e.g., GDPR.
This deliverable D6.3 builds upon i) the threat model defined in D6.2, ii) the basic design proposed in D6.2, and iii) the technical challenges and research questions identified in D6.1, to propose a better design that also incorporates the new requirements related to consent management.

5 State of the Art

Consent Management Software solutions

Consent management in healthcare is a mechanism, which enables patients to control who can access their data, for how long, and for what purpose (Agbo et.al., 2020). Furthermore, it should ensure that any processing of the medical data is auditable and transparent. In general, consent management includes three main elements: i) the collection of consents, ii) the storage of consents, and iii) the use of collected consent and data (Kakarlapudi et al. 2021). However, in practice, many challenges need to be addressed. First, there is no standardized way for collecting consents, which can be gathered via a secure web interface, mobile application, as a handwritten and signed paper form, or by telephone. In practice explicit consent (meaning an express statement of consent) is required under Article 9 (2) (a) GDPR when processing of special categories of data, such as health data, takes place. This requirement does not rule out oral statements; however, it may be difficult to prove for the controller that all conditions for valid explicit consent were met when the statement was made.[3]
Moreover, identity management and anonymization across data lakes present themselves as an important step to prevent linkability, patient deanonymization, and other privacy related issues. Additionally, guidelines for consent templates are required to ensure that they comply with applicable data protection requirements (e.g., GDPR).
In Germany, for example, the Medical Informatics Initiative (MII)[4], which is a consortium that regroups different German universities, hospitals and research institutions, focuses on providing and improving compliant healthcare processes by establishing mechanisms that define how to use, access and manage patient data across different sites. This includes methods for homogenizing data, increasing interoperability, and defining standardized access rules and procedures to both data and consents.

In particular, the Technology, Methods, and Infrastructure for networked Medical research (TMF) (Pommerening et al., 2014) led efforts to provide guidelines for defining, structuring and capturing informed consents, with a particular focus on data protection and ethics. These efforts also resulted in a model proposal based on the Trusted Third Party (TTP) (Bialke et al., 2015) that is legally

[3] European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679 Version 1.0, adopted on 4 May 2020, margin note 94.
[4] https://www.medizininformatik-initiative.de/index.php/en

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

independent, and which provides three main components related to i) identity management, ii) consent management, and iii) pseudonymization.

In line with this TTP model, the University Medicine Greifswald (UMG) developed tools that reflect the three aforementioned components, namely i) Enterprise Identifier Cross-Referencing service E-PIX (Hampf et al., 2020), ii) generic Informed Consent Service gICS (Rau et al., 2020), and, iii) generic Pseudonym Administration Service gPAS (Bialke et al., 2015). Most importantly, the gICS provides a modular approach for digitally recording and managing informed consents. It also includes consent templates, which may combine a large number of policies, and supports workflows for both paper-based and purely digital consents. gICS is currently employed as an essential part for managing consents in a multitude of projects in the healthcare domain in Germany. Other improvements to the gICS tool were also proposed (Bialke et al., 2018), which provide a structured exchange format for modular and printable consent templates, thereby supporting paper-based consent processes. The TTP acts mainly as a central entity that facilitates consent management across multiple sites, which want to collaboratively conduct research studies, thereby sharing healthcare data. In contrast to FeatureCloud, in this TTP-based model, the anonymized data has to be gathered in one location.

More recently, a federated version of the TTP has been proposed, i.e., federated trusted third party (fTTP) (Bahls et al., 2021). In this setting, the identifying data and the corresponding consents remain at the controller site. Therefore, the record linkage as well as the necessary aspects of consent management are implemented in a federated way across the different sites. This allows data to be selected locally in a “federated” manner and the anonymized results are merged across multiple facilities. In order to enable a privacy-preserving record linkage over the merged data, the fTTP implements Bloom filters. This is particularly useful when a study requires linking data of the same patient across multiple sites (e.g., different treatments).

Note that in this case, while consent management is federated, the study itself is conducted on a central repository that includes the merged and anonymized data, contrary to FeatureCloud where the study is also federated. Therefore, the TTP acts as the middleman and ensures traceability and compliance with the data protection guidelines through a centralized and integrated audit and trail mechanisms. In FeatureCloud, data and consents remain on site with the participating controller, and consequently, additional processes and mechanisms need to be put in place to enable trusted federated audits, e.g., using immutable commitments. It is worth noting that some of the processes of the gICS module (e.g., consent template definition) can be used in the context of FeatureCloud, but **locally**, and complemented with secure mechanisms that enable trusted and verifiable audit trails of healthcare data use, without having to rely on a centralized TTP infrastructure.

The consent management suite (COMs) (Heinze et al., 2011) is yet another centralized opt-in solution that was implemented by the university hospital of Heidelberg based on the features of the IHE profile basic privacy patient consent (BPPC). It mainly provides services for electronically defining and storing consent documents and processing queries for digital consents. The current implementation, however, lacks secure logging, authentication and authorization mechanisms. Additionally, it is not clear how consent withdrawals are handled within the system.
The SPECIAL Usage Policy Language[5] is an unofficial specification draft that can be used to express consent and data usage policies in formal and computerized terms, which enable automatic compliance checks of data usage to the specified consents.

[5] https://ai.wu.ac.at/policies/policylanguage/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078.

Consent2share[6] is another open-source tool for consent management and data segmentation, which primarily enables patients to selectively and electronically share their protected healthcare data. The software relies on two main modules; i) patient consent management, and ii) access control services, and integrates with existing healthcare record (EHR) and health information exchange (HIE) systems using interoperability standards. The tool also supports electronic signatures and revocation. In a report by the MITRE corporation (Mitre Report, 2014), different HIO architectures for consent management systems were discussed, clustered into centralized and decentralized models. The report also shows the current technical and regulatory challenges within both implementations. There also exist a multitude of commercial software solutions[7,8] for managing electronic consents that can be used within one single data provider. Moreover, initiatives in Australia led to the implementation of different e-consent management systems such as HealthConnect and emedical book (Win and Fulcher, 2007), where consents are modeled as security profiles with policies that define access to patient data.
Google Cloud Healthcare Consent Management is a U.S. HIPAA compliant framework that helps patients track, modify, and revoke their consent. The consent records are hosted in the cloud and managed by the data provider, while the actual data is stored locally within the hospital database.

PrivacyeSuite v5.0 is another centralized consent management solution with interoperable privacy capabilities developed by HIPAAT International Inc, a company based in the USA and Canada. It provides near-real-time modification of access controls and removal of the adjudication of access to PHI, allowance of multi-state/ organization interaction, and support of privacy by design principles.
However, such tools developed outside of the EU/EEA under a different regulatory regime will need extra scrutiny for compatibility with the de facto “gold standard”, i.e., GDPR, especially as regards consent requirements, data subjects’ rights as well as transfers of personal data to third countries.

All these tools can be integrated with the existing processes of the data providers, and can be used to capture consents in exchangeable formats, however, they do not provide mechanisms for verifiable, and immutable logging of operations on consents and data. This means that although access rules can be defined, they remain locally controlled by the hospitals or biobanks, which in turn have the ability to change the history of the operations. These tools, for example, do not prevent a hospital from hiding consent updates or revocations to an external auditor, and do not offer verifiable mechanisms for linking consents with studies. In relation to FeatureCloud, these tools can be complemented with mechanisms that prevent retrospective manipulation and make it discoverable. Thus, securing such processes with additional verifiable functions may help improve auditability and address these deficiencies, e.g., using distributed ledger technology for committing to consent operations and use within a study.

6 Requirements for Consent Management

In previous deliverables D6.1 and D6.2, we conducted the first iterations on gathering the FeatureCloud requirements related to WP6, elaborated a threat model, and proposed mitigation mechanisms. Based on the collected data, an initial model was provided, which reflects the initial requirements, shows the feasibility of our approach, and evaluates how blockchain-mechanisms can be used to benefit the overall FC architecture and workflow. This Section iterates over the requirements, but with a particular focus on consent management from different perspectives such as legal, technical and privacy/ethical.

[6] https://bhits.github.io/consent2share/
[7] https://www.castoredc.com/econsent/
[8] https://icw-global.com/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078.

Different techniques were employed for the purpose of gathering requirements, ranging from brainstorming sessions to short online workshops that include members from the different consortium partners. In the following, we summarize the relevant technical, privacy and legal requirements which have been gathered. We also want to thank all consortium members who contributed to the identification and clarification of these requirements.

Number: **#1**
Title: **Obtain binary consent for a tuple of study type, data type and identity**
Type: **Technical**
Prio: **1**
Description:
For each piece of used data, FeatureCloud requires a link to the consent of a patient to whom the data belongs (e.g., 'patient 34323'), the type of the data (e.g., gene expression, EHR) and the type of study it will be used for (e.g., cancer research, commercial, …).
The consent system should provide at least one of the following functionalities:

- For a system in which consent can be given in digital form, the system should be able to say 'yes' or 'no' for a given data item/type so that the respective piece of data can be included or excluded from a FeatureCloud study.
- For a system in which consent can be given in paper form, the local project manager needs to include or
- exclude the respective data items according to the available consent forms.

Number: **#2**
Title: **No central consent repository**
Type: **Privacy**
Prio: **1**
Description:
Patient consent should not be stored in one central FeatureCloud-wide repository for security reasons. Since one of FeatureCloud's key security and privacy features is a decentralized approach through federation, the consent repository should not contravene this approach by FeatureCloud-wide centralization.

Number: **#3**
Title: **Proof of consent must be possible**
Type: **Legal**
Prio: **1**
Description:
Patients should provide their consent in such a way that it can be proven during an audit but also during a procedure initiated by a competent authority (e.g., Data Protection Authority or court) that explicit consent has been granted before (e.g., either digitally by signing a consent with a private key, or with a handwritten signature). Hospitals have to be able to prove retrospectively that they were entitled to use the data when the study was performed.

As proof of consent is required under GDPR, the logging of electronic or written consent has to include:
- The identity of the person giving consent,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

- the action taken by the individual, reflecting the act of consent,
- the purposes as well as the processing operations to which consent was given,
- the information provided that led to informed consent (see also Article 13 GDPR in this context - which details the information that needs to be provided),
- the storage period (and thus the validity period of the consent),
- the reference made to the possibility of revoking consent at any time for future processing operations, as well as
- the date the consent was given.

Note: See Article 7 (1) GDPR: Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.
See also Article 7 (3) GDPR: If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

Number: **#4**
Title: **Revoking / modifying of consent must be possible for future FeatureCloud studies**
Type: **Legal**
Prio: **1**
Conflicting: #**3**
Description:
Patients must be able to modify their consent, especially revoke it, ideally digitally using a mobile phone app, but in addition also by other means (see requirement #17). Any modification or revocation of consents affects any future FeatureCloud study. The patient can revoke partial consent, e.g., use of data not for all purposes, but for some kind of analysis.

The possibility to modify consent (other than revocation) is not legally required and is therefore an optional feature primarily in the interest of FeatureCloud in order to avoid full revocation of consent and to reduce the hurdle of obtaining consent.

Note: See Article 7 (3) GDPR: The data subject shall have the right to withdraw his or her consent **at any time**. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.

Number: **#5**
Title: **Consent must be deleted if revoked**
Type: **Privacy, Legal**
Prio: **1**
Description:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

The information that a patient has given consent to usage of their data must be deleted when revoked. However, a certain term of storage of consent after revocation of consent is permissible if needed for proof that consent has been collected. To that end, the controller may keep a record of consent statements received, so he can show how consent was obtained, when consent was obtained and the information provided to the data subject at the time shall be demonstrable. The controller shall also be able to show that the data subject was informed and the controller´s workflow met all relevant criteria for a valid consent (EDPB, Guidelines 05/2020).

Number: **#6**
Title: **Site A should not be able to access the content of consents of Site B**
Type: **Privacy, Legal**
Prio: **1**
Description:
To ensure compliance with data protection requirements, a site should only be able to access the content of their own patient consents, even though they might participate in a collaborative study with other sites.

Number: **#7**
Title: **Delegation of rights to give consents**
Type: **Legal, Technical**
Prio: **3**
Description:
There should exist the possibility to delegate the right to give consent to other persons (legal guardians, e.g., children cannot provide consents and therefore parents do it on their behalf or inversely, old people give their rights to their descendants). However, this option has to be individually assessed by each institution and/or local DPO (Data Protection Officer) in accordance with national legislation and jurisprudence.

Number: **#8**
Title: **Consent unlinkability across sites**
Type: **Technical, Privacy**
Prio: **1**
Description:
Consents of a patient on site A should not be linkable to consents of the same patient given on site B. This heavily depends on how identity is defined (e.g., one global identifier such as insurance number or hospital specific identifier).

Number: **#9**
Title: **Commitment to used data items (and the related consents)**
Type: **Technical, Privacy, Procedural**
Prio: **1**
Description:
The hospitals (different sites) need to commit to the used data items and the related collected consents. In the event of an Audit, this is necessary to verify that the data which has been used for a study is backed by an appropriate consent.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Number: **#10**
Title: **Commitment or confirmation of the coordinator/study operator that he/she has collected all the results/models and performed the aggregation.**
Type: **Technical, Procedural**
Prio: **1**
Conflicting:
Depends on: #9
Description:
After the coordinator has aggregated all the results/models, he/she commits to the aggregated result. This triggers the next step of the process. At a later point in time the coordinator can then prove that he/she has aggregated all the results correctly by publishing all the results/models and thereby allowing for reproducible results.

Number: **#11**
Title: **Random selection of sites/hospitals for audit**
Type: **Technical, Procedural**
Prio: **2**
Conflicting:
Depends on: #10
Description:
After the results/models which operated on the on-site data, have been collected, a subset of sites/hospitals is selected (provably) at random. The results are published such that all sites can verify if they have been correctly selected.

Number: **#12**
Title: **Audit that the committed data items are correct and have matching consents**
Type: **Technical, Procedural**
Prio: **1**
Depends on: #11
Description:
For every selected site/hospital, one or multiple selected auditors verify that the previously committed data really has matching consents. Hereby, the consents can be in electronic or paper format.

Number: **#13**
Title: **Commitment or confirmation of the auditor(s) that they have finished their audits**
Type: **Technical, Procedural**
Prio: **1**
Depends on: #12
Description:
The auditor needs to inform all sites/hospitals that she has finished the audit period. After this event, all hospitals are allowed to delete the used data and results (if they have already been collected by the study operator).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Number: **#14**
Title: **The used commitments must not reveal the original input data (hiding)**
Type: **Privacy, Technical**
Prio: **1**
Description:
The used commitments must not reveal or disclose the original input data to which they commit to. Cryptographically speaking, the commitment has to fulfill the hiding property, i.e., for any attacker it is not computationally feasible (not in polynomial time) to derive the original input (preimage) used to create the commitment.

Number: **#15**
Title: **The provided commitments/confirmations have to be cryptographically signed by the respective parties**
Type: **Technical**
Prio: **1**
Description:
The provided (digital) commitments have to be cryptographically signed by the respective parties. This is necessary to attribute these actions to the respective party (non-repudiation). This requirement implies that the public cryptographic credentials (e.g., public keys) of the relevant actors (auditors, coordinators and hospitals) need to be known/registered within the system.

Number: **#16**
Title: **The actors must agree on the order as well as the authenticity of the published commitments**
Type: **Technical**
Prio: **1**
Description:
All involved actors (like auditor, coordinator and hospitals) have to agree on the history of all published commitments, their order, and the generated randomness. This is necessary, to perform the correct steps at the correct point in time (e.g., publish commitments or delete unnecessary data later).

Number: **#17**
Title: **Data subjects unable to use a digital device must be able to give and revoke consent**
Type: **Ethical, Legal**
Prio: **1**
Description:
People who are unable to use a smartphone or computer, either because of their medical condition or because they are not proficient enough to use such devices, must also have the opportunity to give and revoke their consent. The possibility to modify consent (other than revocation) is not legally required and is therefore an optional feature primarily in the interest of FeatureCloud in order to avoid full revocation of consent and to reduce the hurdle of obtaining consent. Therefore it is optional whether modification of consent (other than revocation) is possible exclusively in an app or also in an alternative form for people unable to use an app.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Number: **#18**
Title: **Revocation unobservability**
Type: **Ethical, Legal**
Prio: **2**
Description:
The attending physician/doctor treating the data subject/doctor who recruited the data subject for a study should not be able to know that the data subject revoked consent. Patients could feel a pressure to give their consent and not to revoke it later in order to improve their relationship with their treating physician, on whom their life may depend. This should be avoided by this requirement both because of the legal requirement that consent must be freely given and also because of the practical aspect that the hurdle of obtaining consent in the first place should be as low as possible. This was discussed as a key non-functional requirement from the beginning of FeatureCloud.

Number: **#19**
Title: **If required by local regulations, the consent forms have to pass the review of an Ethics Committee prior to their use**
Type: **Ethical**
Prio: **1**
Description:
In accordance with the law of the Member State concerned, an ethical review may be performed by an ethics committee prior to the start of a research project. The review by the local ethics committee may also encompass consent forms.

7 System Design

First, we describe the high-level **audit process for consent management** that aims to fulfill all given requirements. Then, we sketch out **two concrete implementations** of this process; i) one using a *decentralized approach,* which utilizes a consensus algorithm and a blockchain, and ii) the other is a *centralized approach* and relies on the auditor acting as a central trusted-third-party, which is honest-but-curious. This means the central auditor has access to everything except for the raw data of all hospitals. At the end, we compare the different trade-offs and security guarantees that can be achieved within the respective approaches.

7.1 High Level Process for Consent Management

This section sketches the process of executing a FeatureCloud study, with a clear focus on how consents are handled and managed within this process to detect misbehavior of participants, i.e., using non-consented data, and as a byproduct also enhance the overall integrity and reproducibility of a FeatureCloud study.

Actors/Roles:

- ***coordinator***:
  The coordinator prepares and announces the project, and aggregates the results i.e., creates a FeatureCloud study using a published docker image from the FeatureCloud AI store, invites collaborators, and aggregates the federated study results.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Assumption: It is assumed that the docker container provided by the coordinator is trusted. This means that the docker container and the included software does not leak data and is designed to execute as reproducible as possible, i.e., ideally the execution of the docker container is fully deterministic such that running it with the same inputs results in the exact same output. In FeatureCloud, the docker image of a federated algorithm (FeatureCloud app) is published by a developer, and then certified by FeatureCloud after checking its privacy and performance.

Technically speaking, the coordinator is assumed to be a *covert adversary* (Aumann and Lindell, 2007), meaning they act honestly if they are afraid of being caught cheating, but if they would know that there is no way that they can be caught cheating, they would act maliciously.

- ***participant / hospital / local project manager****:*
  Participants in a FeatureCloud study are hospitals. The local administration and participation are managed by the *local project manager*. Therefore, the terms *participant*, *hospital* and *local project manager* can be used interchangeably.

  Assumption: It is assumed that the participants can fetch study information from the coordinator securely, i.e., coordinators are authenticated and the integrity of the transferred data (including the docker images) is ensured.
  Further, it is assumed that patient data stored at the hospital is secured by state-of-the-art IT security measures. This means that hospitals do not actively leak or sell any patient data, and that the majority of hospitals are not compromised so that attackers cannot exfiltrate patient data. Therefore, the data of most patients can be assumed to securely reside within a hospital.

  Like the coordinator, the participants are assumed to be *covert adversaries* (Aumann and Lindell, 2007), meaning they act like honest participants if they are afraid of being caught cheating, but if they would know that there is no way that they can be caught cheating, they would act maliciously.

- ***patient****:*
  The patient gives consent for the processing of his/her medical data for certain purposes/studies.
  This could happen ***upfront*** (consent is given before an appropriate federated study exists), or ***on demand*** (a patient is asked to consent for a specific study). Regardless of these two types, we differentiate between consents given in ***paper form*** and consents given in ***digital form*** directly by the patient. We designed the consent management process to support both types of consents, such that the overall process is backward compatible and can thus be practically established more easily in the current healthcare infrastructure, which heavily relies on consents in paper form.

  Assumption: It is assumed that the patients provide informed consent and are not lured into giving consent inconsiderately.

- ***auditor:***
  The auditor is the legal entity (e.g., data protection officer) that checks the proper execution of a study at randomly selected hospitals (on site) and approves that only properly consented data was used during the study. Furthermore, she ensures that the gathered results are reproducible. This is done by rerunning the federated study on her own hardware given the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

used data of the participant which she is supposed to audit. If the auditor detects misbehavior by the participant, then this will be reported.

Assumption: The auditor is assumed to faithfully execute her auditing duties, i.e., she only approves the proper execution of a study if this really was the case. Although the auditor is expected not to leak, or misuse data provided by the participant during an audit, the auditors will by design never be able to observe all data used within a federated FeatureCloud study, as this would defeat the general purpose of federation in the first place.

Preparation of Study (coordinator):
The study was prepared by the coordinator and published in the FeatureCloud ecosystem with the state set to be *announced*. This already includes the docker container and a description of all required data items, so that potential participants can check if they are eligible to participate in the study, i.e., have the required data in sufficient quantity and with the proper consent of the respective patients. Note that it is also possible for a coordinator to define a workflow that involves multiple applications (i.e., a composition of multiple apps).

As soon as a new FeatureCloud study is announced, it can be fetched in a secure way by potential participants of the study including all relevant metadata. Therefore, potential participants poll for new studies regularly (see below).

Preparation for Study (participant):
To prepare for a study, participants have to check for newly announced studies or invitations, assess whether or not they can contribute, and possibly also collect further consents in the process:

- Suitable study announced:
  Either the local project manager polls for newly announced studies, or he is notified/invited by the coordinator to participate in a newly announced study. Either way, the participant has to ensure that sufficient eligible data is available to participate in the study, and that the required consents of patients are available. If the required threshold is reached, the participant informs the coordinator that his hospital can contribute to the study. In this process the participant transmits a range for the number of patients for which he has consented to data required for this specific study.
  Note that the participant is assumed to not submit any data if the number of eligible patients who have given consent is too low, such that a deanonymization of those patients or the reconstruction of their data would be possible during the study. Moreover, participants can accept or decline an invitation.

- Collect consents:
  We differentiate between two ways in which a hospital can collect consents to use patient data in federated FeatureCloud studies: i) in ***paper form***, and ii) in ***digital form***.
  Currently consents are mostly given in paper form, i.e., containing a handwritten signature of the patient. In this case, the local project manager is responsible for digitizing the consent information, and keeping the original document for auditing purposes. If patients are capable of giving consents in digital form directly (e.g., using a qualified electronic signature), this digitalization step of the local project manager can be omitted.

  Furthermore, we distinguish two cases with regard to the time at which consent was given:
  a. *upfront:*
     As it is now, patient consent is collected by hospitals upfront in a generic way that is not tailored towards a specific feature cloud study. As such, a consent is necessarily

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

not tailored towards a specific study, but they are usually broader[9] and therefore, might be eligible for multiple studies. Here, legal criteria regarding consents in the healthcare system (with regard to the processing of genetic data, biometric data or data concerning health, see Article 9 (4) GDPR) have to be taken into account that might be different for each EU member state.

b. *on demand:*
After a feature cloud study has been announced, the hospital actively asks matching patients for consent to use their data in this specific feature cloud study.
In a scenario where patients are capable of giving digital consents (e.g., using a qualified electronic signature), it becomes easier to ask patients, on the fly, for their consent to a specific study.
They can then decide if they permit access to certain data items, for a particular study on demand. If they agree, they issue a digital signature under their ID and send it to the hospital.

Execute Study (coordinator, participants & auditor):
If the coordinator has received enough confirmations from participants with enough data to participate in an announced study, then the study is executed.

- *Set FeatureCloud Study to state execution (coordinator):*
The state of the FeatureCloud study is changed from **announced** to **execution** in the FeatureCloud ecosystem by the coordinator. This signals all participants to execute the study.

- *Execute the federated study and commit to input, output and used consents (participants):*
As soon as the study is in the execution phase, each participant executes the docker container with the required and consented input data.
To facilitate auditability and fulfill **requirement #9**, the participants commit to the patient data used as input to the study, the associated consents that allow this data to be used in the respective study, as well as the result of the study. It has to be ensured that the used commitments fulfill the **requirements #14, #15** and **#16** and thus are cryptographically hiding, signed and published on a secure bulletin board accessible by all involved parties, or made available to them in another secure way.

  Ideally, the study is deterministic and thus fully reproducible, which means that the same input data produces the exactly same output data. In this case, the commitment can be a cryptographic hash over input, output, as well as associated consents for the used inputs.
  If we are dealing with consents in paper form, the local project manager is responsible for digitizing the content of the consents such that she can create a cryptographic hash over those consents. To ensure auditability, the original documents also need to be kept such that a spot check of this digitalization step can be performed by the auditor. If the consent was directly given in digital form, the auditor can simply check the digital signature of the patient and the scope of the machine-readable consent. If all data structures are standardized this process allows for further automation.

- *Aggregate all submitted results and signal successful aggregation (coordinator):*
When all participants have submitted their trained models, i.e., partial results, the coordinator performs the aggregation of the submitted results. After all results have been successfully aggregated and the final output computed, the coordinator changes the state of the study to

[9] Recital 33 of the GDPR allows as an exception that the purpose may be described at a more general level.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

aggregated. Therefore, the coordinator has to commit to the collected results, as well as to the outcome of the final aggregation (this should fulfill **requirement #10**).
If the results of the study can be public right after this step, they can readily be released, but there might be cases where the final results should be further processed and contextualized in a scientific paper before release of the raw output data.
Again, ideally the study is designed to be deterministic, and thus the aggregation of the same partial inputs yields exactly the same overall result. In any case, the commitment can be a cryptographic hash of used inputs and outputs. If the study is not designed to be deterministic the auditor has to assess the plausibility of the self-computed result given the original input and output (see the description of the audit for more details).

- *Random selection of participants for audit (auditors & participants)*
  After the state of the study has changed to aggregated, the random selection and assignment of auditors to participants takes place. To select the subset of participants that is subject to an audit, a (distributed) random number generation algorithm is used (this addresses **requirement #11**). We refer to Schindler et al. (2020) for a recent comparison of available solutions in this domain.
  For now we assume that a suitable algorithm has been executed successfully and a set of auditors has been assigned to a subset of participants at random. If this is the case the state is changed to **audit** releasing a publicly verifiable random number in the process.

- *The audit is performed (auditors & participants):*
  In this step, the on-site audits are performed by the auditors (**requirement #12**). Therefore, the entire federated study is recalculated locally on hardware of the auditor, using the patient data provided by the selected participant. The result is compared to the commitment (cryptographic hash) that was previously submitted by the respective participant. Additionally, the associated consents are presented to the auditor, who checks if they allow for this type of data usage. Also spot-checks on the digitized consents are performed if the digitalization step was done by the local project manager. Therefore, the handwritten signature of the paper consent form is checked and maybe even former patients are contacted and asked if they have given consent. If the consent was directly given in digital form, the auditor checks the provided signatures associated with the consents.

  If the federated study was not designed to be completely deterministic, the entire original input as well as the output that has been submitted to the coordinator has to be kept to allow for reproducibility during the audit. In such a case the auditor also re-runs the federated study on his hardware and then checks if the previous output is sufficiently close to his output, provided the same input data.

  If everything was in order, the auditor acknowledges the successful outcome of the audit (**requirement #13**). If all audits have been completed the state of the study changes to **postprocessing**.

  If problems have been detected this is also made transparent by the auditor. If they are severe, the results from this particular participant have to be discarded and the study switches back to the **execution** phase.

  Note that the auditor does not keep any data collected from a participant on-site. Even if the auditor would be malicious, she would only be able to exfiltrate data from one participant.

- *Postprocessing and cleanup of data that is no longer needed (participants):*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

After all auditors have confirmed the proper execution of the federated Feature Cloud study, the state of the study is changed to **postprocessing**. This signals every participant that they can now delete all data that is no longer needed for auditing purposes.
This includes, the used input data as prepared for the Feature Cloud study, as well as the created output data, if the study was not designed to be fully deterministic.
Moreover, all consents which have been revoked, or deprecated during the execution of the study can now be deleted as well.

To fulfill **requirement #10** and ensure the integrity of the Feature Cloud study, by proving that the coordinator has correctly aggregated all received results of a federated study, the latter has to publish all collected results to allow for reproducibility and verification of the aggregation. Since we are in a federated model, the different models submitted by the participants should not pose a privacy concern. If the aggregation was proven to be correct, the change of the study is changed to **finished**.

Note: With respect to **Requirement #18** on revocation unobservability, it is indeed possible to address it on an organization level, i.e., by separating duties or roles within the system. For example, an access control mechanism (e.g., RBAC role-based access control) that provides restricted access to operations on consents would be sufficient. In this case, consents are managed by different entities/roles than the doctors within the same hospital. This prevents bad treatments by doctors for patients who did not give their consents to use their data. For example, in Austria, the Austrian Data Protection Authority requires hospital operators to have RBACs in place.

7.2 Major challenges and research questions addressed

We identified the following major challenges and research questions which are either already addressed by the suggested approaches, or will be the topic of further research within the FeatureCloud project.

1) **RQ1: How to check if non-consented data has been used within FeatureCloud studies?**

   First of all, the local controller, e.g., the hospital, is legally responsible for and has to be able to demonstrate compliance with the core principles of the GDPR, including that the processing of patients' data is being done lawfully, fairly and in a transparent manner.

   The core challenge that has to be tackled by the consent management system, is to check if non-consented data has been used in a FeatureCloud study.
   Given the requirements, the only practically feasible solution can consist of **detective measures**, as currently no form of scalable verifiable computing or Trusted-Execution-Environment (TEE) technology exists for the particular use case, that offers the required functionality, as well as the ability to assess the authenticity of manually signed paper consents. Moreover, different data and storage formats might be used at different hospitals, further hampering automated solutions employing TEEs which would require uniform data formats across all hospitals as input.

2) **RQ2: How can a correct link between identities and the corresponding consents and data be ascertained?**

   To solve this problem, FeatureCloud has to rely on an external identity provider, which can ensure that the used identities in the interaction with the hospital are genuine and that this

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

fact can also later be verified by an auditor. The creation of fake identities and fake consents should not be possible, because this would in turn allow the use of fake data. Fake data may be used to conduct a poison attack, where the poisoned data set will result in a biased output model that does not reflect the original/legitimate data.

3) **RQ3: How to ensure that a participant (hospital) uses the most up-to-date consent for a study?**

A patient, and in accordance with GDPR regulations, has the possibility to iteratively update or revoke previous consents she has given (e.g., expiry data, scope, studies). These updates (either paper or digital) are locally stored within the hospital infrastructure. Of course, the patient also gets a proof of consent update or revocation, but she is not involved in the auditing process, as the auditor will only have access to the hospital data.
Therefore, it becomes possible for the hospital to simply hide consent updates/revocations and continue to use the old ones, e.g., with broader scope. During an audit, the hospital will only show the consent that suits the desired outcome.
By doing this, the hospital can **include** previously consented data as input to a study by hiding either a revocation of a consent, or a consent update that for example narrows its scope or exclude specific study types.

How this question is best addressed, depends on the specific scenario:

Paper consents: For paper consents, the original document corresponding to the old consent can be destroyed or marked as revoked under the supervision of the patient. Thereby, the old document cannot be presented to an auditor during an audit.

Digital- and paper consents: For both paper and digital consents, one way around this problem would be to solely use on-demand consents that are only valid for one specific study[10]. Thereby, old consents cannot be reused in newer studies. Alternatively, also the validity period of a consent can be defined to be very low to minimize the threat.

Digital consent: Information regarding currently valid, or invalidated consents is published, or communicated to the auditor in a privacy preserving manner. This necessarily requires that the patient (as the only actor besides the hospital that knows when a consent is updated) gets more actively involved in the process of either ensuring that such an update is somehow communicated to the auditor, or by communicating such an update to the auditor directly. Therefore, this question relates to RQ5, which questions the role of the patient in the auditing process. In the implementation section, we discuss several approaches on how this could be achieved.

However, it needs to be stressed again that the local controller, e.g., the hospital, is also legally responsible for and has to be able to demonstrate compliance with the core principles of the GDPR, including that the processing of patients' data is being done lawfully, fairly and in a transparent manner. Additionally, the possibility of the imposition of an administrative fine in accordance with Article 83 (5) GDPR should motivate each participating site to have effective measures in place as regards consent management.

[10] https://profiles.ihe.net/ITI/TF/Volume1/ch-19.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

4) **RQ4: How to prevent the hospital from excluding consented data from a study?**

Similar to the problem of data poisoning through injection of fake data as input to machine learning, the hospital can also exclude a subset of eligible and consented data from a study[11]. Although legally it might not be considered as a problem, ethically it might be questionable as the training outcome, with and without the exclusion, may deviate, and therefore it will not be possible to know whether or not retaining the data was intentional to achieve a desired output.

Technically, the problem of the hospital excluding eligible data cannot be solved without the involvement of the patient as this is the only actor besides the hospital who knows about his treatment and an appropriate consent issued by her. Therefore, this question also relates to RQ5 regarding the role of the patient in the auditing process.
As the process of gathering consents is also carried out by the hospitals themselves, they are always in the position to selectively not ask for consent of specific patients, or selectively omit the recording of necessary data for a specific study to bias the output of study. Thus, this attack vector can never be entirely prevented without changing our underlying actor model and the associated requirements.

Under the assumption that the auditor is able to get access to all available consents, it would be possible for him to check if consented data has not been used in a study, although it would have been eligible to use this data. This, of course, cannot solve the issue that hospitals might not record relevant data in the first place, or do not ask specific patients for their consent.

5) **RQ5: What role do patients play in the audit process?**

We differentiate between two types of patients:

a) Patients who, to a certain extent, are able/willing to use IT applications/software, e.g., mobile applications, digital signatures.
b) Patients who can't or are not willing to use or interact with any IT infrastructure. Those patients may only rely on written signatures.

The question then, is how both types of patients will be involved in this whole process? How is it possible to enforce certain rules in the FeatureCloud process, that makes it possible for both types of patients to be indirectly or directly involved in the auditing.

There is an inherent security advantage if patients digitally sign consents, as such consents cannot be backdated as the signature of a consent is not known in advance by the hospital and thus cannot be committed to. Whereas, handwritten signatures on consent forms can in theory be collected by the hospital in retrospect after already having used the data of patients (unconsented) for certain studies.

6) **RQ6: How can FeatureCloud benefit from research in the field of self-sovereign identities and verifiable credentials?**

With the inception of blockchain technology new methods for managing identities, which promise better decentralization and self-sovereignty emerged (Fdhila et al., 2021) (Ghesmati et al., 2021). The aim here is to investigate i) whether or not it is possible to take advantage

[11] This would also be the case, if the hospital excludes a consent update that would broaden the scope of the consent and thus allow for data of the respective patient being used in a study which would have not been possible under the original consent.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

of the properties leveraged by such methods, ii) if they can be applied in the context of FeatureCloud, and iii) their limits with respect to current use case. In particular, we will investigate two main technologies:

a) Distributed Identifiers (DIDs)
   A DID is a unique identifier, usually associated with a DID document that specifies cryptographic material, and verification methods essential for proving ownership of the DID and engaging in trustworthy communication with the DID owner. A DID method, on the other hand, defines how a DID can be created, resolved, updated and revoked.
b) Verifiable Credentials (VCs)
   Verifiable credentials are identity attributes and assertions about a specific subject issued by an identity provider. In contrast to traditional credentials, a relying party (third party service) can check the validity of a VC without having to interact with the issuer.

7.3 Implementation of the Process (Decentralized vs Centralized)

In this section, we compare two concrete implementations of an audit process; i) *centralized*, and ii) *decentralized*, which is based on a blockchain-based consensus algorithm.

Centralized Audit Process
In this approach, the auditor acts as a trusted-third-party, which is ***honest-but-curious***.
This means that, although being trusted with the auditing process, the auditor is not allowed to observe patient/input data from all hospitals, as this would defeat the entire purpose of the FeatureCloud project and the federated machine learning approach utilized within it.

Execution phase:
In this approach, the hospitals sign their commitments to the used input, output as well as the consents and send them to the auditor in the execution phase, who forwards these to the coordinator.

To ensure that all consented data that would be eligible for use in a specific study has been used, regular commitments to all available consents has to be sent to the auditor, such that the auditor knows the number and references to all available consents. Thereby, during an on-site audit all these consents can be inspected if some have been omitted.

Aggregated phase:
After the coordinator signals proper aggregation of the results. The auditor samples a set of hospitals for audit and performs the audits.

Post Processing phase:
To verify the integrity of the study the auditor has to publish all received models such that the aggregation can be reproduced and thus verified by interested parties.

Advantages:
- Low overhead of participants as no setup and execution of a consensus protocol is required.
- No data published anywhere. Only the auditor receives all models and commits regarding the used inputs, outputs and consents.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Disadvantages:

- Auditors are required to run their own secure infrastructure for storing and managing the commitments (additional overhead for the auditor). This also introduces a single point of failure (the auditor).
- Overall verifiability/integrity checking of the aggregation process is problematic since this would require consensus over the used inputs. As there is no secure bulletin board, the participants have to rely on the auditor forwarding all messages to the coordinator.
  Consequently, the auditor has to publish all received signed models that have been submitted for aggregation to allow verifiability. Otherwise, the data transferred by a participant to the auditor and the data transferred to the coordinator might not be the same. Thereby, equivocations and double submissions can also be resolved by the auditor.
- If the auditor turns out to be malicious, she can always claim that she has not received messages from a certain hospital she wishes to exclude. Therefore, this approach only works if the auditor is *honest-but-curious*.
- The random selection of the auditor cannot be verified by other participants. Thus, it cannot be ruled out that the auditor is malicious and has informed all participants that they are subject to an audit, and thus would be able to observe all patient data of a study on-site at every participant. Therefore, a distributed random number generation protocol has to be executed between the participants anyway, but this would require agreement on the used inputs and thus require consensus which would again lead to decentralized approach that was not the goal here.
- If the exclusion of eligible data is considered as a threat, the resolution requires recurring commitments to all available consents, as well as inclusion proves to the patients that their (current) consent is included in these commitments. This is more difficult to realize in a centralized environment as the hospitals have to inform the patients that they have committed to their consents, while the patients can only check if this was really the case after the auditor has published the respective received messages.

Decentralized Audit Process

In this approach, a consensus-based system is, or has been established between the coordinator(s), the participants as well as the auditor(s). This system agrees on the current state of a ledger (aka. blockchain) on which state updates, like commitments, are recorded. Thereby, the system works as a secure bulletin board. The main difference to the centralized approach is that the auditor now can be modeled as a *covert adversary*, thus strengthening the overall security of the system.

Execution phase:

In this approach the hospitals sign their commitments to the used input, output as well as the consents and publish them on the blockchain in the execution phase. Thereby, special care has to be taken, that this is done in a privacy preserving way (i.e., merkle tree root hash of all inputs, outputs and consents).

Aggregated phase:

After the coordinator signals proper aggregation of the results the underlying blockchain can be used as a secure bulletin board and thus as a basis to execute a distributed random number generation protocol. The output of which samples a set of hospitals for audit and assigns each one of potentially many auditors at random.

Advantages

- Auditor can be modeled as a *covert adversary*.
- Agreement on submitted data/models that have been used, therefore equivocation is not possible and public verifiability is achieved as a byproduct.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

- If the auditor turns out to be malicious and does not faithfully acknowledge a correctly executed Feature Cloud study, the participant might be able to request another auditor if this is deemed permissible. This possibility, should highlight, that in a case where everything has been committed properly within a decentralized architecture, also the hospitals have certain ways to protect themselves from untrustworthy auditors. Since everything the participant has to prove has been committed already, and therefore does not have the ability to introduce retrospective changes, changing or assigning multiple auditors is a non-issue.

Disadvantages:

- If the exclusion of eligible data is considered as a threat, the resolution requires recurring commitments to all available consents, as well as inclusion proves to the patients that their (current) consent is included in these commitments.

8 Conclusion

In this deliverable, we outlined the need for an auditing process that ensures only consented data is used, and improves the integrity of federated machine learning, where data privacy is of utmost importance. The main reason lies in the fact that, in federated learning, it is *provably impossible* to detect anomalies in models/data submitted by participants, if anomaly detection cannot be used due to the data not being identically distributed data (i.i.d.), which is the case in the FeatureCloud project. Therefore, two alternative approaches present themselves as possible viable solutions: i) relying on Trusted Execution Environments (TEEs), or ii) elaborating an auditing process, which provides detective security measures to identify misbehavior in retrospect. There are a couple of reasons why we chose not to pursue a TEE based design in FeatureCloud. First of all, current TEE designs are lacking functionality and performance that would be required for their adaptation within the Project. Moreover, the availability of different TEE vendors and thus the resulting dependency on them is unsatisfactory, especially in the light of recent EU initiatives that should strengthen the digital sovereignty for Europe[12].
As a result, we opted for a design that enables a reliable post auditing process, which supports two different architectures; *centralized*, and *decentralized*. The latter uses a consensus algorithm and a blockchain as an audit trail, (for hashes with high min-entropy input, signatures of involved parties and meta information). The main difference between those two approaches is that in the decentralized setting, the trust assumptions regarding the auditor can be further reduced, thereby making the overall security of the design more robust.
Both solutions are designed to support paper-based processes, i.e., handwritten signatures on consent forms, but ideally would work in settings where patients are able to provide digital signatures in a secure way. This allows hospitals to participate in FeatureCould studies, even though having different internal processes, as well as different degrees of digitalization within their healthcare infrastructure.

For the next deliverables, we will focus more on possible consensus technologies and infrastructures which can be used within the context of FeatureCloud and the integration aspects of our design approaches.

[12] https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

9 References

Bagdasaryan, E., et al. (2018) How to backdoor federated learning. International Conference on Artificial Intelligence and Statistics. PMLR.

Brundage, Miles, et al. "Toward trustworthy AI development: mechanisms for supporting verifiable claims." *arXiv preprint arXiv:2004.07213* (2020).

Schüler, P., and Buckley, B. (Eds.) (2014) Re-Engineering clinical trials: Best practices for streamlining the development process. Academic Press.

Kairouz P. et al. (2019) Advances and Open Problems in Federated Learning.

Schindler, Philipp et al. (2020) Hydrand: Efficient continuous distributed randomness. IEEE Symposium on Security and Privacy (SP). IEEE.

Agbo, C. C., and Mahmoud, Q. H., 2020. “Design and implementation of a blockchain-based e-health consent management framework”. In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 812–817.

Kakarlapudi, P. V., and Mahmoud, Q. H., 2021. “Asystematic review of blockchain for consent management”. Healthcare, 9(2).

Nchinda, N., Cameron, A., Retzepi, K., and Lippman, A., 2019. “Medrec: A network for personal information distribution”. In 2019 International Conference on Computing, Networking and Communications (ICNC), pp. 637–641.

Nchinda, N., Cameron, A., Retzepi, K., and Lippman, A., 2019. “Medrec: A network for personal information distribution”. In 2019 International Conference on Computing, Networking and Communications (ICNC), pp. 637–641.

Genestier, P., Zouarhi, S., Limeux, P., Excoffier, D.,Prola, A., Sandon, S., and Temerson, J. M., 2017. “Blockchain for consent management in the ehealth environment: A nugget for privacy and security challenges”. Journal of the International Society for Telemedicine and eHealth, 5.

Agbo, C. C., and Mahmoud, Q. H., 2019. “Comparison of blockchain frameworks for healthcare applications”. Internet Technology Letters, 2(5), p. e122.

Polge, J., Robert, J., and Le Traon, Y., 2021. “Permissioned blockchain frameworks in the industry: A comparison”. ICT Express, 7(2), pp. 229–233.

Aumann, Y., Lindell, Y., 2007. Security Against Covert Adversaries: Efficient Protocols for Realistic Adversaries. J. Cryptol. 23, 281–343. https://doi.org/10.1007/s00145-009-9040-7

Bahls, T., Hampf, C., Bialke, M., Hoffmann, W., 2021. Lösungsbaustein fTTP (federated Trusted Third Party) als ein Enabler für vernetzte medizinische Forschung mit dezentraler Datenhaltung 4.

Bialke, M., Bahls, T., Geidel, L., Rau, H., Blumentritt, A., Pasewald, S., Wolff, R., Steinmann, J., Bronsch, T., Bergh, B., Tremper, G., Lablans, M., Ückert, F., Lang, S., Idris, T., Hoffmann, W., 2018. MAGIC: once upon a time in consent management-a FHIR® tale. J. Transl. Med. 16, 256. https://doi.org/10.1186/s12967-018-1631-3

Bialke, M., Penndorf, P., Wegner, T., Bahls, T., Havemann, C., Piegsa, J., Hoffmann, W., 2015. A workflow-driven approach to integrate generic software modules in a Trusted Third Party. J. Transl. Med. 13, 176. https://doi.org/10.1186/s12967-015-0545-6

Fdhila, W., Stifter, N., Kostal, K., Saglam, C., Sabadello, M., 2021. Methods for Decentralized Identities: Evaluation and Insights, in: González Enríquez, J., Debois, S., Fettke, P., Plebani, P., van de Weerd, I., Weber, I. (Eds.), Business Process Management: Blockchain and Robotic Process Automation Forum, Lecture Notes in Business Information Processing. Springer International Publishing, Cham, pp. 119–135. https://doi.org/10.1007/978-3-030-85867-4_9

Ghesmati, S., Fdhila, W., Weippl, E., 2021. Studying Bitcoin Privacy Attacks and Their Impact on Bitcoin-Based Identity Methods, in: González Enríquez, J., Debois, S., Fettke, P., Plebani, P., van de Weerd, I., Weber, I. (Eds.), Business Process Management: Blockchain and Robotic Process Automation Forum, Lecture Notes in Business Information Processing.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078.

Springer International Publishing, Cham, pp. 85–101. https://doi.org/10.1007/978-3-030-85867-4_7

Hampf, C., Geidel, L., Zerbe, N., Bialke, M., Stahl, D., Blumentritt, A., Bahls, T., Hufnagl, P., Hoffmann, W., 2020. Assessment of scalability and performance of the record linkage tool E-PIX® in managing multi-million patients in research projects at a large university hospital in Germany. J. Transl. Med. 18, 86. https://doi.org/10.1186/s12967-020-02257-4

Heinze, O., Birkle, M., Köster, L., Bergh, B., 2011. Architecture of a consent management suite and integration into IHE-based regional health information networks. BMC Med. Inform. Decis. Mak. 11, 58. https://doi.org/10.1186/1472-6947-11-58

Mitre Report, 2014. Electronic Consent Management: Landscape Assessment, Challenges, and Technology. Mitre corporation.

Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., 2014. Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0. https://doi.org/10.32745/9783954662951

Rau, H., Geidel, L., Bialke, M., Blumentritt, A., Langanke, M., Liedtke, W., Pasewald, S., Stahl, D., Bahls, T., Maier, C., Prokosch, H.-U., Hoffmann, W., 2020. The generic Informed Consent Service gICS®: implementation and benefits of a modular consent software tool to master the challenge of electronic consent management in research. J. Transl. Med. 18, 287. https://doi.org/10.1186/s12967-020-02457-y

Win, K., Fulcher, J., 2007. Consent Mechanisms for Electronic Health Record Systems: A Simple Yet Unresolved Issue. J. Med. Syst. 31, 91–6. https://doi.org/10.1007/s10916-006-9030-3

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

10 Table of acronyms and definitions

BPPC	Basic Privacy Patient Consent
COMs	Consent Management Suite
concentris	concentris research management GmbH
DiD	Decentralized Identifier
EHR	Electronic Healthcare record
fTTP	Federated Trusted Third Party
GDPR	General Data Protection Regulation
gICS	generic Informed Consent Service
GND	Gnome Design SRL
HIE	Health Information Exchange
HIPAA	The Health Insurance Portability and Accountability Act
LPM	Local Project Manager
MMI	Medical Informatics Initiative
MS	Milestone
MUG	Medizinische Universitaet Graz
Patients	In this deliverable, we use the term “patients” for all research subjects. In FeatureCloud, we will focus on patients, as this is already the most vulnerable case scenario and this is where most primary data is available to us. Admittedly, some research subjects participate in clinical trials but not as patients but as healthy individuals, usually on a voluntary basis and are therefore not dependent on the physicians who care for them. Thus, to increase readability, we simply refer to them as “patients”.
RI	Research Institute AG & Co. KG
SBA	SBA Research Gemeinnutzige GmbH
SDU	Syddansk Universitet
TEE	Trusted Execution Environment
TMF	Technology, Methods, and Infrastructure for networked Medical research
TTP	Trusted Third Party
TUM	Technische Universitaet Muenchen
UHAM	University of Hamburg
UM	Universiteit Maastricht
UMG	University Medicine Greifswald
UMR	Philipps Universitaet Marburg
VC	Verifiable credential
WP	Work package

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078.