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
D2.3 "Working PAML-Layer with low distortion"

Work Package
WP2 "Cyber risk assessment and mitigation"

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

© FeatureCloud Consortium, 2021
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

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	D2.3 “Working PAML-Layer with low distortion”			
Work Package	WP2 “Cyber risk assessment and mitigation”			
Date of Delivery	**Contractual**	M32 (31.08.2021)	**Actual**	M33 (04.09.2021)
Nature	Report	**Dissemination Level**	Public	
Lead Beneficiary	05 SBA Research			
Responsible Author(s)	Rudolf Mayer (SBA)			
Keywords	Anonymisation, non-consented data, privacy-aware machine learning			

Table of Content

1 Objectives of the deliverable based on the Description of Action (DoA) 4

2 Executive Summary 4

3 Introduction (Challenge) 4

4 Background 5

Definition of anonymised data 5

5 Threats to Privacy 6

6 Privacy-preserving Data Mining Techniques 7

Definitions 7

PPDM Techniques 7

Pseudonymisation 7

Syntactic Anonymisation 8

Synthetic data generation 8

Differential Privacy 9

Local vs. Global Differential Privacy 9

Differential Privacy - Input perturbation 10

Differential Privacy - Objective (functional) and output perturbation 10

Utility estimation 10

Utility and Method comparisons 10

Utility Metrics 11

7 FeatureCloud Integration 12

Fusion Approaches 12

Anonymisation parameter settings 14

Locality of training 14

8 Available Implementations 15

Syntactic Anonymisation 15

Synthetic Data Generation 15

Differential Privacy 16

9 Conclusion 16

10 References 17

11 Table of acronyms and definitions 21

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 main objective of WP2 is the definition of a security and privacy architecture **based on the cyber risk assessment and legal requirements**.

“While the approach of the FeatureCloud project by design mitigates most major concerns regarding security and privacy, the underlying foundations of the platform to be developed need to be secured thoroughly, especially considering the diversity of the local execution platforms on the hospital sites. Important attacker goals include the theft of data on a local level, as well as the theft and manipulation of results, with the inclusion of possible insider attacks.”

Objective 1 of this work package thus aims ***“to develop an additional layer for adding suitable anonymization to the local execution of the algorithms, in case specific consent could not be gathered”***.

The corresponding task in this work package is ***Task 3: Adding a “Privacy-Aware-Machine-Learning Layer***:
“While it is certainly beneficial to the overall result to calculate the feature vectors inside the local execution platforms with as granular data as possible, this processing of sensitive data typically requires consent with respect to the GDPR and NISD, which sometimes might be impossible to achieve. In order to still be able to use this data source, the sensitive information needs to be anonymized beforehand. In this task MUG, SBA and TUM will develop methods for anonymizing the data as carefully as possible in order to preserve as much inherent value as possible, while obeying to all required privacy regulations. Furthermore, this Privacy-Aware-Machine-Learning (PAML) Layer needs to be easily integrateable and requires additional risk analysis and mitigation strategies.”

2 Executive Summary

This deliverable first examines the landscape of potential anonymisation approaches that could be utilised for achieving this objective. Identifying syntactic anonymisation, synthetic data generation and differential privacy as candidates, we introduce each category of approaches and describe the main variants in each category. Then, we analyse their suitability in terms of application in the FeatureCloud platform, based on criteria such as

- the achieved level of privacy
- the ease of use of the anonymised version of the dataset, e.g. whether the representation of the results of the anonymised version is easily combinable with results from federated learning
- the availability of practical, reliable and tested implementations of the methods.

We identify fitting implementations that can accelerate their integration into the platform. Certain aspects (which exact method with which parameters, which form of combination of results, ..) are depending on e.g. the datasets used, and are therefore to be set study-specific by the local and global study coordinators.

3 Introduction (Challenge)

While the objective is clear in the purpose by wanting to utilise also data for which the individuals (patients) have not explicitly given consent, there are several challenges to be solved, including

- Which types of anonymisation are in general suitable?

- Which types of anonymisation are most effective in retaining a high utility (inherent value) of the original data; this is generally dependent on the actual data and machine learning method at hand.
- How can the best fitting method be estimated for a given data set?
- How are these data samples to be integrated into the overall FeatureCloud platform?
- Are there any residual risks from this anonymised data?

4 Background

Definition of anonymised data

We first investigate what “anonymised” means in our context. As the objective states that the aim is to utilise data that has not been given consent too, this implies that we need to consider what the EU General Data Protection Regulation (Regulation (EU) 2016/679, “GDPR”) defines in terms of anonymised data.

We want to refer the reader to Deliverable D2.1 “Risk assessment methodology”, which covers an analysis of anonymisation in regard to GDPR. It states that “*recital 26 of the GDPR is of particular importance:*
“The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.”

and then concludes:

“Firstly, it [recital 26 of the GDPR] clarifies that **pseudonymised data are personal data**, as long as the additional information which makes it possible to attribute the data to a natural person is available, and that **anonymous data are not personal data**. Secondly, it makes another very important clarification: To determine whether a natural person is identifiable, not every theoretical possibility to identify the person must be taken into account but only means reasonably likely to be used to do so. To ascertain whether **means are reasonably likely** to be used, all objective factors should be taken into account, such as the costs of and the amount of time required for identification, the available technology at the time of the processing, and technological developments (Esayas 2015). “

Relevant for the analysis of residual risks in anonymised data is the following conclusion from D2.1:
“From this, it can be concluded that in order to determine whether data is personal data under GDPR a practical, not a theoretical standpoint must be taken. Means reasonably likely to be used, are means that not only exist theoretically but that would be used practically.
In the context of this risk assessment methodology, this means that a practical assessment must be carried out: From the attack vectors on the anonymity of the data found in this document only those are legally relevant that are reasonably likely to be used by an actual attacker in practice. This must

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

be assessed on the basis of objective factors such as the costs and the amount of time required, the required skills, the potential gain and the available technology but also possible technological developments in the future.
In order to assess whether an attack vector is relevant from a legal perspective, attacks that are reasonably unlikely can be ignored. An attack can be considered being reasonably unlikely if it cannot be imagined that it will happen in practice in the given context because the attacker will shy away from the effort.”

With this background, we then consider all methods that do not, with means reasonably likely to be used, allow an attacker to identify an individual, to be acceptable as an anonymization method.

In the following, we detail frequently discussed and utilised anonymization methods.

5 Threats to Privacy

The threats to the privacy of individuals considered in the FeatureCloud system mostly originate from participating in a dataset collecting microdata, i.e. data where one record corresponds to one individual. The specifics of the disclosure risks from which a dataset is to be protected can be derived from categorising the attributes of the dataset into different types.

Direct identifiers (personally identifiable information, PII)) are attributes that can uniquely identify a record in the dataset (*Privacy enhancing data de-identification terminology and classification of techniques*, 2018), such as social security numbers, telephone numbers, email addresses, etc. Their presence generally allows direct association of records to a specific individual.

Quasi-identifying attributes (Dalenius, 1986), (*Privacy enhancing data de-identification terminology and classification of techniques*, 2018) are attributes that by themselves are not uniquely identifying an individual, but do so in combination with other quasi-identifiers, at least for a (large) portion of the participants in the dataset. Frequent examples are birth date, post (ZIP) codes, and sex.

Sensitive attributes (*Privacy enhancing data de-identification terminology and classification of techniques*, 2018) generally contain information about individuals that they are not willing to share, e.g. a medical diagnosis or their salary.
Disclosure of these might do certain harm to the individuals. In contrast to quasi-identifiers, they can generally not be used for identifying, even if used in combination, but are frequently the target of disclosure attacks.

Other attributes are generally considered **insensitive**.

Threats to privacy can be categorised as follows.
Identity disclosure is generally considered the strongest form of disclosure, and means that an attacker can associate an individual to a specific record. This is also referred to as *re-identification*. It is often achieved via a record-linkage attack, where the target dataset is correlated to other data available to the attacker (public, or private). Such disclosure in many jurisdictions has direct legal consequences for the data controllers.

Attribute disclosure (Elliot, 2005) means that an attacker can learn (exactly, or approximately) the value of one (or more) attributes of an individual that are contained in the targeted database. For example, an attacker might learn the approximate salary or the medical diagnosis of an individual in a database, but knowing just some of the quasi-identifying attributes. It might be achieved even without uniquely associating an individual to a specific record in a dataset, if the set of records an attacker might narrow a match down to still has all (or most) records containing the same (or similar) sensitive values.

Membership disclosure is generally considered to be the weakest form of disclosure. Here, an attacker can, e.g. via linking data from multiple sources, infer whether or not an individual is contained in a dataset. Unlike the other disclosure settings, it does not directly release sensitive attributes from the dataset, i.e. it does not directly unveil information from the dataset itself. However, an attacker might infer meta-information from the membership of an individual in a dataset, e.g. if this dataset is a medical dataset containing information about patients carrying a certain disease.

6 Privacy-preserving Data Mining Techniques

Definitions

In (Mendes and Vilela, 2017), Privacy-preserving Data Mining (PPDM) techniques are classified into four main categories: (i) data collection privacy, (ii) Privacy-Preserving Data Publishing (PPDP), (iii) Data Mining Output Privacy (DMOP) and (iv) distributed privacy.
Data collection privacy relates to the data randomisation strategies before it is sent to a data collector. PPDP includes techniques such as k-anonymity, *l*-diversity, *t*-closeness, personalised privacy and ε-differential privacy.

DMOP relies on ensuring that the computation, which is performed on original, unabridged data, does not require the exchange of input data, but releases only the outcome of the computation to the data analyst. It encompasses association rule hiding, among other subcategories such as downgrading classifier effectiveness, query auditing and inference control, and distributed privacy includes approaches that provide privacy over partitioned data. Federated Learning, the main concept FeatureCloud relies on, would be subsumed in the DMOP category, though some intermediate form of data is exchanged.

PPDM Techniques

Here, we briefly describe established approaches for performing Privacy-Preserving Data Mining (PPDM).

Pseudonymisation

Even though generally not considered a privacy-preserving approach, pseudonymisation is still frequently used when analysing sensitive data. In essence, pseudonymisation is a de-identification approach that replaces all **direct identifiers** such as a social security number with a pseudonym, i.e. an artificial identifier (*Privacy framework, Amendment 1*, 2018).
It is the "process of removing the association between a set of identifying data and the data subject and adding an association between one or more pseudonyms and a particular set of characteristics"[1].
Pseudonymised data is highly vulnerable to inference attacks, by record linkage on common quasi-identifiers with other datasets that still carry the identifiers.
Failed approaches of releasing such partially sanitised data sets came notable e.g. in the form of the AOL search data logs released in 2006, where some individuals could be re-identified[2], or the Netflix price data, which could be linked to public profiles on the Internet Movie Database (IMDB) (Narayanan and Shmatikov, 2006).
The attributes on which such record-linkage attacks rely are the above-mentioned quasi-identifiers, which are generally attributes that per-se alone are not uniquely identifying, but they might become

[1] https://www.iso.org/standard/63553.html

[2] https://www.nytimes.com/2006/08/09/technology/09aol.html

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

so when considered together, e.g. a person's birth date in combination with the postcode of their residence can identify a large number of the population uniquely.

Syntactic Anonymisation

According to ISO, "Anonymisation is a process by which personal data is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party. The concept is absolute, and in practice, it may be difficult to obtain."[3]. This includes a family of techniques that typically generalise or suppress records until a specified syntactic condition is met (Clifton and Tassa, 2013).

As the best-known representative, *k*-anonymity (Samarati and Sweeney, 1998) aims at achieving anonymity of the individuals (data subjects) contained in the data such that for each individual it holds that the record cannot be distinguished on the set of quasi-identifying attributes from at least k-1 others in the data set. Such a group of *k* (or more) records is called an *equivalence class*, or *k-group*. With *k*-anonymity, identity disclosure (or re-identification) is thus not possible anymore. This is generally achieved by generalising values to a higher-level semantic concept, or eventually suppressing certain values altogether. As there are generally multiple ways to achieve the same level of *k* in a certain dataset by generalising different attributes to a different level (or suppressing individual values), these solutions are generally measured by some preserved data quality, e.g. the number of steps needed to take to generalise. Finding an anonymisation that fulfills the desired level of *k*, and is optimally in terms of this data quality, is generally a NP-hard problem (Bonizzoni et al., 2009), and thus is generally solved by a heuristic, for example the *Flash* algorithm (Kohlmayer et al., 2012).

k-anonymity is a syntactic anonymisation method, and the ancestor to a range of further techniques addressing some of the disclosure attacks *k*-anonymity is still vulnerable to, e.g. l-diversity (Machanavajjhala et al., 2006), or <α,k>-anonymity (Chi-Wing et al., 2006). l-diversity addresses the issue that disclosure can still happen even if the identity itself is not disclosed. For example, if a certain individual for which an attacker wants to disclose information can only be identified to be one of multiple samples from one group of *k* records, if all of these records do share the same value for the sensitive attribute the attacker wants to learn (e.g. the income, or a disease), the attacker will learn this information, even if it remains unclear which specific record belong to the target person. l-diversity thus ensures that among the sensitive values within one equivalence class have at least *l* different values. Again, multiple flavours of l-diversity exist, such as distinct l-diversity (at least *l* distinct values exist in each equivalence class), or the more advanced *entropy l-diversity*, which also considers the distribution of these distinct values (Aggarwal and Yu, 2008).

However, it has been shown that in practice, most extension schemes to k-anonymity have a high cost in forms of utility loss. For example, (Brickell and Shmatikov, 2008) showed that applying a *3*-diversity to a dataset was worse than ensuring a 100-anonymity - and both had low utility compared to the original dataset.

Synthetic data generation

Synthetic data is artificially generated data containing records that are similar to the original ones, while preserving the high-level, global relationships within the data (i.e. similar statistical moments as the original data, and similar correlation), without actually disclosing real, single data points that contain sensitive information. One of the earliest usages of synthetic data was in the partial synthetic data approach by (Rubin, 1987), where certain attributes (columns) are generated synthetically. Fully synthetic data creates complete data samples, without any attributes given.

[3] https://www.iso.org/obp/ui/#iso:std:iso:25237:ed-1:v1:en:term:3.2

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

Synthetic data generally involves two steps: (i) learning a model representing the data of an original, source data set, and then (ii) instrumenting it to generate data points forming a target dataset, which is similar in overall characteristics to the source dataset. This is illustrated in Figure 1.

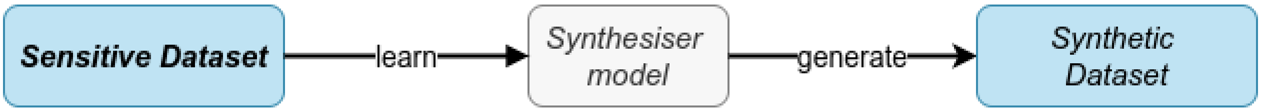

Figure 1: Synthetic Data generation process

Several approaches for obtaining a synthesiser model have been published, mostly differing on the complexity and expressiveness of the model learned, e.g. from a simple covariance matrix to generative adversarial networks (GANs) all in the Synthetic Data Vault (SDV) (Patki et al., 2016), Bayesian Networks in the DataSynthesizer (Ping et al., 2017), or Decision Trees in synthpop (Nowok et al., 2016), to more advanced models such as Generative Adversarial Networks (GANs), also provided e.g. in newer versions of the Synthetic Data Vault.
An overview and utility analysis is performed e.g. in (Hittmeir et al., 2019a), showing that synthetic data can achieve effectiveness at the same or very similar level as the original data.

Regarding regulatory compliance, synthetic data is likely the least explored of all mentioned methods in this deliverable. As synthetic samples are not related to original entries in terms of a 1-to-1 correspondence, the possibility of identity disclosure in the form of record linkage is often considered not to be meaningful, at least for fully synthetic data (see (Bellovin et al., 2019) and (Elliot, 2014)). Attribute disclosure was for example discussed in (Taub et al., 2018) and later generalised in (Hittmeir et al., 2020), with the conclusion that synthetic data generally lowers the attribute disclosure risk, but that this reduction also correlates with a reduction in data utility, especially if the learned target attribute is a sensitive attribute and thus among those to be protected.

Differential Privacy

Differential Privacy (DP) (Dwork, 2006) is a technique that adds noise to data, according to a defined level of privacy guarantee that is controlled by a parameter ε.
In contrast to syntactic anonymisation techniques, privacy is guaranteed by the data analysis **process**, which offers plausible deniability. In principle, the concept states that a participant should not be affected by being part of a study. That is, for any two adjacent datasets D_1 and D_2 that differ in only one record, then a process *A* is ε-differentially private if $P[A(D_1)=O] \leq e^{\varepsilon} \cdot P[A(D_2)=O]$, where $P[A(D_1)=O]$ is the probability that O is the output of running A on the database D_1.
A relaxed form of differential privacy is the so-called (ε,δ) differential privacy, which relaxes the original concept of differential privacy against unlikely events with probability ≤ δ. A review of more extensions of the basic differential privacy concept is provided e.g. by (Desfontaines and Pejó, 2020). Differential privacy can be applied at various stages in the data analysis process - to the input, to the analysis function itself, or to the output of the function.

Local vs. Global Differential Privacy

Literature also sometimes distinguishes between local and global (or central) differential privacy (Dwork and Roth, 2013). **Local** differential privacy is especially used in settings when (many) clients send their data to a central service, where e.g. a statistics is computed from these inputs. Local differential privacy in this context means that a differentially private mechanism is applied to input data before it is sent to the coordinator. **Global** (or central) means that the aggregator applies the differentially private mechanism after the data was collected. Local differential privacy is thus beneficial if the clients do not trust the coordinator with their raw, unabridged data. Local differential

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

privacy relies on the fact that data cannot be accurately estimated from individual privatised data, thereby providing privacy for the data collection process.
Local differential privacy has its origins in the Randomised Response mechanisms (Warner, 1965), which offers plausible deniability to a data collection process for surveys, where the true (initially binary, later ordinal) answer is randomly perturbed, so that there is an uncertainty over the correctness of a response of an individual to a specific question, but in a way that the overall distribution of answers stays nearly the same. Recent implementations include e.g. RAPPOR (Erlingsson et al., 2014), developed by Google and utilised in some of their end-user products, such as the Chrome Browser, or Apple's usage of differential privacy for, in a similar fashion, analysing usage behaviour of end-users (Differential Privacy Team Apple., 2017), though many details about this implementation are unclear (Tang et al., 2017).

Differential Privacy - Input perturbation

One option, as mentioned above, is to apply the differentially private process to the input data to release a new version of said data, i.e. to perform an **input perturbation**. Here, the input data is modified by adding noise to the values, before publishing the data to the data analysis step (Fukuchi et al., 2017). While the above mentioned local differential privacy is primarily meant to be applied to the data **before** it is pooled, it can in principle also be applied to already aggregated data for the purpose of anonymising individual records.
However, at the moment, most of the works in literature focuses on simple examples where single-attribute data is to be perturbed (e.g. (Kairouz et al., 2014)), and there are no vetted implementations available for approaches that consider multiple attributes at the same time, or there exist only theoretical extensions to approaches such as RAPPOR that can be used for perturbing multiple attributes (Murakami and Kawamoto, 2019).

Differential Privacy - Objective (functional) and output perturbation

Differential privacy can further be applied in two later stages of the data analysis process, which both affect the final output, and are thus to be considered a DMOP approach, where the input data is never released to the data analyst.
Output perturbation adds noise to the output of the algorithm (Chaudhuri et al., 2011) (Dwork and Roth, 2013), and is thus generally agnostic of the specific algorithm, though the amount of noise is dependent on the type of processing performed. It is therefore important to estimate the **sensitivity** of the algorithm that is run on the data, either the global sensitivity which is independent of the dataset, or the local sensitivity for a specific, given dataset.

Objective (functional) perturbation adds noise at the level of model internals, for example, the objective function (Chaudhuri et al., 2011) (Kifer et al., 2012) of a learning algorithm.
Recent prominent techniques are developed for iterative optimisation processes, where noise is not injected to the results, but to the coefficients of the polynomial representation of the loss function. One example is the differentially private stochastic gradient descent (DPSGD) algorithm (Abadi et al., 2016), the extended DPSGD algorithm (eDPSGD) (Yu et al., 2019) or the adaptive Laplace Mechanism algorithm (Phan et al., 2017), which all target stochastic gradient descent based learning algorithms, and can thus be used to e.g. train Neural Networks.

Utility estimation

Utility and Method comparisons

(Mendes and Vilela, 2017) describe privacy and data quality metrics for different approaches. PPDP methods are compared in detail in (Chen et al., 2009) (Fung et al., 2010). Other surveys also propose data quality and utility metrics (Bertino et al., 2008) (Bertino and Fovino, 2005) (Fletcher and Islam, 2015), while a few mention the trade-off between privacy and utility of PPDM (Malik et al., 2012) (Verykios et al., 2004).

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

A framework for evaluating Privacy-Preserving Data Mining (PPDM) techniques is proposed in (Bertino et al., 2005). The work considers efficiency, scalability, data quality, hiding failure and privacy level. However, only one specific type of PPDM, namely private *association rule hiding*, where data is perturbed in order to avoid mining sensitive rules, is considered.

The surveys in (Chen et al., 2009) (Fung et al., 2010) focus on PPDP methods, and mention utility aspects. Furthermore, utility comparison of a few PPDM techniques is made in (Sattar et al., 2013). Several surveys of PPDM techniques provide descriptions and high level comparison of the approaches (Aggarwal, 2015) (Shah and Gulati, 2016) (Aldeen et al., 2015) (Tran and Hu, 2019). (Tran and Hu, 2019) performs a comparison of PPDM techniques, listing common scenarios in Big data analytics

Utility Metrics

Data utility metrics are used to quantify the quality or utility of data perturbed by a privacy-preserving mechanism. There are two main types of metrics (Fung et al., 2010): (i) metrics that directly measure information loss of the perturbed data and (we refer to that as “data-centric”; (Fung et al., 2010) calls this as “general purpose”) (ii) metrics that quantify the loss in quality of the statistical analysis carried out on the perturbed data (we refer to this as “task-centric”; (Fung et al., 2010) calls this a “special purpose metric”).

Data-centric metrics include, for example, normalized loss, discernibility, generalization counting, record-level squared error, non-uniform entropy, etc. (Chen et al., 2009)

Measuring the effectiveness of private data via task-centric metrics is investigated in (Wimmer and Powell, 2014), (Malle et al., 2017) and (Slijepčević et al., 2021) where the authors evaluate the performance of predictive machine learning models trained on *k*-anonymised data. In these analyses the most common metrics for evaluating machine learning (ML) model performance, such as classification accuracy, precision, recall F1 score and confusion matrix are used to describe the difference in effectiveness between a model trained with original data versus a model trained on private (modified) data.

While both data- and task-centric metrics are beneficial in determining quality of the private data, the latter holds the advantage in scenarios where the usage of data can be anticipated, which is a common scenario for data publishing, i.e. medical institution shares the data with data analyst to predict some clear target such as a disease of a patient. However, this type of metric requires more effort and time to estimate compared to the simple information loss metrics, therefore trading accuracy in data effectiveness estimation for time and effort might be prefered.

Being able to estimate the task effectiveness from data centric methods would be beneficial. In (Šarčević et al., 2020) we study the correlations between the two groups of metrics while evaluating *k*-anonymous data. However, in most cases, no clear trend and correlations can be found.

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

7 FeatureCloud Integration

Fusion Approaches

In the FeatureCloud project, the information learned from the data to which no explicit consent has been given to be used via federated learning, could be utilised in multiple ways, as compared to or combined with the outcome of the federated learning, where we have direct consent and can access the raw data for the federated learning purpose.

- In the simplest case, the outcome from the PAML layer is used independently from any federated learning outcome, and thus there are no specific requirements and limitations.

- On the other hand, it might be considered to combine the outputs from both the federated learning and the PAML layer. In this case, it is important to consider how such a combination shall be achieved. This can be either by integrating (merging, combining) the learned models, or combining the outputs of the learned models. This is comparable to the problem of early and late fusion in multi-modal analysis (Baltrusaitis et al., 2019) (D’mello and Kory, 2015) in general, where one can decide to combine (merge) features (early fusion, (Snoek et al., 2005)) or classifier outputs (late fusion, (Snoek et al., 2005)). Some related work from multi-model analysis also defines the term “earliest fusion” (Seeland et al., 2017). Along this categorisation, we can define the following three approaches:
    - If the results shall be directly integrated, one can integrate at the feature level, i.e. combine the fully-consented, unabridged data with the anonymised data. We can consider this the “earliest fusion” equivalent. Depending on the output of the anonymisation step, if this output is in a modified representation, this basically requires the unabridged data to be anonymised in the same way.
    - If the models shall be integrated, this could be seen as an early fusion equivalent. This requires that we merge the model parameters of the federated model learnt on the unabridged data, and the model learned on the anonymised data. Again as above, if the feature representation is different for the models learnt, so will the inputs to these models change, and thus there will be semantic gaps in the two models. In such cases, an integration might not be achievable or meaningful.
    - If the model outputs are to be combined, this is roughly equivalent to late fusion. This is likely the easiest setting, as for such a combination, one can rely on proven ensemble learning techniques (Sagi and Rokach, 2018) to reach a final prediction. This approach is independent of whether the data representation has been changed, or not.

These options of fusion are illustrated in Figure 2.

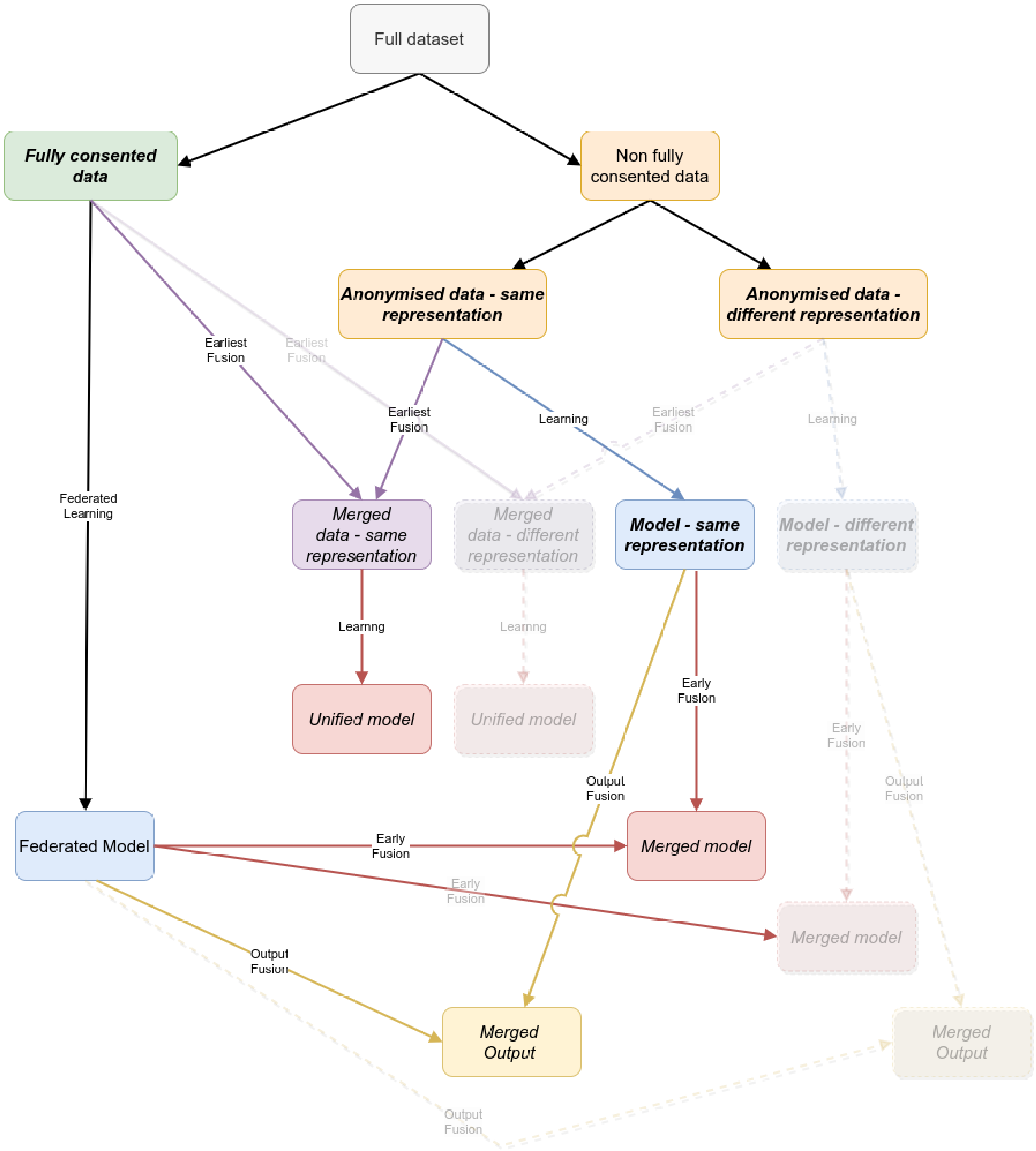

Figure 2: Different fusion techniques for combining the original, unabridged with the anonymised data. *N.b.: anonymised data can result in either the same or a different representation. Therefore, each option is duplicated, depending on the input representation; for clarity, one of the options is displayed less prominent.*

- If the users are willing to give consent for output or objective (functional) differential privacy, but not for the federated learning (for whatever reason, e.g. because they expect a lower privacy from federated learning as compared to the mathematical guarantees from differential privacy), then a combination of these results could be easier. In fact, the primary option is a merged output approach according to the terminology given above, and as illustrated in Figure 3.

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

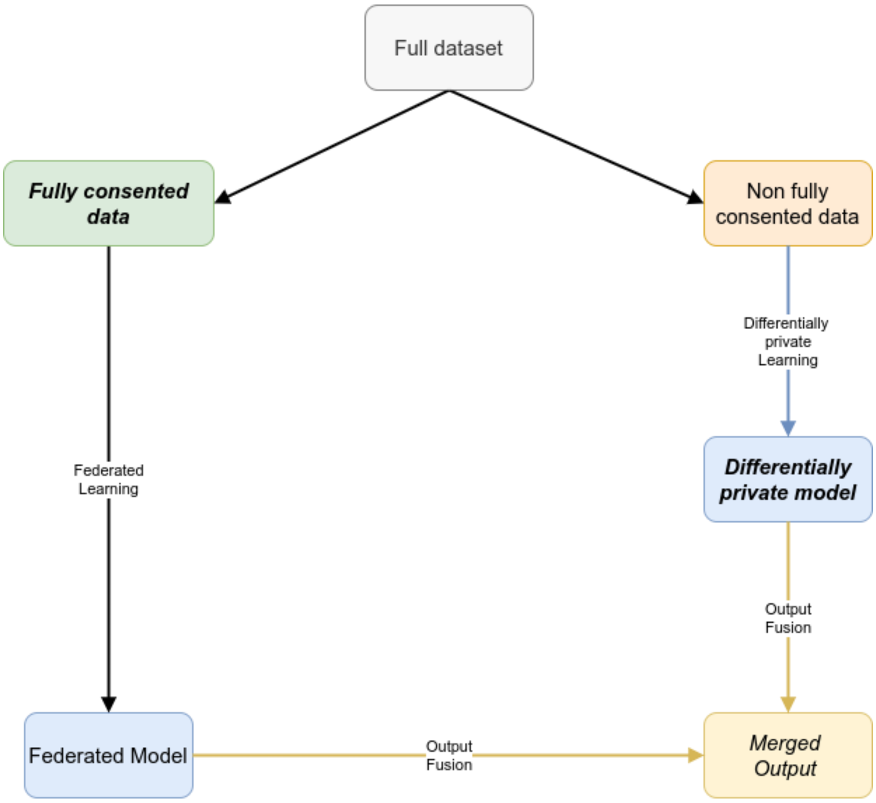

Figure 3: Fusion techniques for combining the original, unabridged with the differentially private model.

Anonymisation parameter settings

Regulatory requirements for anonymisation are not describing concrete details that would allow deducing specific parameter settings for the anonymisation settings. Thus, which level of e.g. *k* for *k*-anonymity or which value of ε for the ε-differential privacy shall be set is not globally determinable. Thus, the most reasonable approach is that the local study coordinators involved in the learning process decide on useful parameters.

Locality of training

One other aspect to consider is where / how to train models from this anonymised data. One option is to replicate the setup for the “regular” (unabridged) federated learning, i.e. each learning on the anonymised data happens as part of a federated learning. Another option is that the locally anonymised data is then aggregated and learned from in a central location.
Both approaches entail that the (potential) modifications applied to the data representation (e.g. the generalisations for syntactic anonymisation) are the same among all clients, otherwise the approach for aggregating the global model needs to be modified as well, resp. the data aggregation becomes more difficult.

Overall, it seems that these decisions (the type of anonymisation, the type of combination, the parameters, and locality of training) are best set by each individual study, as they are partially dependent on the study participants, or the types of data involved.

8 Available Implementations

In this section, we describe tools and frameworks that provide implementations of various anonymisation approaches, and are thus candidates for integration into the FeatureCloud system. As mistakes in anonymisation could have drastic consequences such as data leaks, it is important to rely on approaches and implementations that are mature. We thus prefer relying on third-party solutions that exist for a longer time period, are actively maintained, and show a proper software development process. Therefore, also open-source solutions are preferred (or a requirement), as that enhances openness and clarity in the code, and reduces the risk of malicious code and hidden bugs. Thus, the list below is already mostly focussed on implementations fulfilling these requirements.

Syntactic Anonymisation

- The ARX toolkit[4] (Prasser et al., 2014) provides many forms of syntactic data anonymisation, such as k-anonymity, l-diversity, t-Closeness (Li et al., 2007), etc. It provides both a graphical user interface, and an API in the Java programming language.
- Crowds[5] provides a Python implementation of the Optimal Lattice Anonymisation algorithm (El Emam et al., 2009).
- The UTD (University of Texas at Dallas) Anonymization ToolBox[6] offers multiple algorithms for k-anonymity and l-diversity, such as Datafly (Sweeney, 1997), Mondrian (LeFevre et al., 2006), or Incognito (LeFevre et al., 2005), all via a Java library.
- Mondrian is also provided as Python implementation[7] for k-anonymity, but the general maturity of this project is not clear.

As the most mature and best-maintained project, we decided to integrate the ARX toolkit.

Synthetic Data Generation

- Synthpop[8] (Nowok et al., 2016) provides an implementation for the R statistical computing language. A recent port to Python was also published[9].
- The Synthetic Data Vault[10] (Patki et al., 2016) provides a Python library providing both simple (Gaussian Copula) and more advanced (Generative Adversarial Networks) methods to generate synthetic data. It is provided as Python library.
- The Data Synthesizer[11] (Ping et al., 2017) is a Python library that provides Bayesian-Network based methods.

Given our previous experiences from evaluating the synthetic data generation tools on various types of tasks (classification, regression, as well as supervised, unsupervised and semi-supervised anomaly detection) ((Hittmeir et al., 2019a) (Hittmeir et al., 2020) (Hittmeir et al., 2019b) (Mayer et al., 2020)), both data synthesizer and synthpop are well suited in terms of achieved data utility. In terms of computational speed, synthpop is more efficient when the number of attributes is large. We thus chose this package for integration.

[4] https://arx.deidentifier.org/
[5] https://pypi.org/project/crowds/
[6] http://www.cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php
[7] https://github.com/qiyuangong/Mondrian
[8] https://www.synthpop.org.uk/get-started.html
[9] https://github.com/hazy/synthpop
[10] https://sdv.dev/
[11] https://github.com/DataResponsibly/DataSynthesizer

Differential Privacy

- The IBM Diffprivlib[12] is a framework for differential privacy written in the Python programming language, and offers also objective (functional) differential privacy, e.g. for linear regression, logistic regression, Naive Bayes
- diffpriv[13] (Rubinstein and Aldà, 2017) is a library for the R statistical computing language. It provides output differential privacy, with additional functionality to estimate the sensitivity.
- Tensorflow-privacy[14] allows training models with objective (functional) differential privacy, e.g. logistic regression
- PySyft[15] is a framework primarily focussed on providing federated learning, but similar to tensorflow-privacy contains models with objective differential privacy.

Given that we need to maintain control over exact parameter setting and thus also the sensitivity analysis, we selected to use the IBM Diffprivlib, as it provides methods for both output as well as objective (functional) differential privacy.

9 Conclusion

This deliverable analysed the potential landscape of solutions to “develop an additional layer for adding suitable anonymization to the local execution of the algorithms, in case specific consent could not be gathered”. Starting from a review on existing approaches that are subsumed under anonymisation techniques, we identified methods for syntactic anonymisation such as k-anonymity and its extensions, synthetic data generation, as well as multiple flavours of differential privacy. For each of these methods, we analysed advantages and shortcomings for achieving this objective.

Several aspects cannot be globally determined, e.g. what specific consent the users might be willing to give (e.g. for allowing analysis that guarantees differentially private output), the locality of the analysis of the anonymised data, which exact parameters to set to control the strength of the achieved data protection, and how results from the original, federated, and the anonymised analysis shall be integrated. The exact choice of a method further likely depends on the type of data and the dataset.

We thus opted for an approach where the PAML layer provides methods from all categories of approaches that can be used depending on the requirements of a specific study. Thus, we chose to highlight multiple different options a local and global study coordinator can choose from, given their setting. We further selected fitting, existing implementations of the anonymisation methods for their integration into the overall FeatureCloud platform.

[12] https://github.com/IBM/differential-privacy-library

[13] https://github.com/brubinstein/diffpriv

[14] https://github.com/tensorflow/privacy

[15] https://github.com/OpenMined/PySyft

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

10 References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L., 2016. Deep Learning with Differential Privacy, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS ’16. ACM, New York, NY, USA, pp. 308–318. https://doi.org/10.1145/2976749.2978318

Aggarwal, C.C., 2015. Data mining: the textbook. Springer.

Aggarwal, C.C., Yu, P.S. (Eds.), 2008. Privacy-preserving data mining: models and algorithms, Advances in database systems. Springer, New York, NY.

Aldeen, Y.A.A.S., Salleh, M., Razzaque, M.A., 2015. A comprehensive review on privacy preserving data mining. SpringerPlus 4.

Baltrusaitis, T., Ahuja, C., Morency, L.-P., 2019. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. 41, 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Bellovin, S.M., Dutta, P.K., Reitinger, N., 2019. Privacy and synthetic datasets. Stan Tech Rev 22, 1.

Bertino, E., Fovino, I.N., 2005. Information driven evaluation of data hiding algorithms, in: Int. Conf. on Data Warehousing and Knowledge Discovery. Springer.

Bertino, E., Fovino, I.N., Provenza, L.P., 2005. A framework for evaluating privacy preserving data mining algorithms. Data Min. Knowl. Discov. 11.

Bertino, E., Lin, D., Jiang, W., 2008. A survey of quantification of privacy preserving data mining algorithms, in: Privacy-Preserving Data Mining. Springer.

Bonizzoni, P., Della Vedova, G., Dondi, R., 2009. The k-Anonymity Problem Is Hard, in: Kutyłowski, M., Charatonik, W., Gębala, M. (Eds.), Fundamentals of Computation Theory, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 26–37. https://doi.org/10.1007/978-3-642-03409-1_4

Brickell, J., Shmatikov, V., 2008. The cost of privacy: destruction of data-mining utility in anonymized data publishing, in: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 08. Presented at the the 14th ACM SIGKDD international conference, ACM Press, Las Vegas, Nevada, USA, p. 70. https://doi.org/10.1145/1401890.1401904

Chaudhuri, K., Monteleoni, C., Sarwate, A.D., 2011. Differentially private empirical risk minimization. J. Mach. Learn. Res. 12.

Chen, B.-C., Kifer, D., LeFevre, K., Machanavajjhala, A., 2009. Privacy-Preserving Data Publishing. Found. Trends Databases 2.

Chi-Wing, R., Li, J., Fu, A.W.-C., Wang, K., 2006. (a, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing, in: 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD). ACM Press, Philadelphia, PA, USA. https://doi.org/10.1145/1150402.1150499

Clifton, C., Tassa, T., 2013. On syntactic anonymity and differential privacy, in: Int. Conf. on Data Engineering Workshops (ICDEW). IEEE, Brisbane, QLD. https://doi.org/10.1109/ICDEW.2013.6547433

Dalenius, T., 1986. Finding a needle in a haystack or identifying anonymous census records. J. Off. Stat. 2.

Desfontaines, D., Pejó, B., 2020. SoK: Differential privacies. Proc. Priv. Enhancing Technol. 2020, 288–313. https://doi.org/10.2478/popets-2020-0028

Differential Privacy Team Apple., 2017. Learning with privacy at scale. Apple.

D’mello, S.K., Kory, J., 2015. A Review and Meta-Analysis of Multimodal Affect Detection Systems. ACM Comput. Surv. 47, 1–36. https://doi.org/10.1145/2682899

Dwork, C., 2006. Differential Privacy, in: 33rd International Colloquium on Automata, Languages and Programming (ICALP). Springer, Venice, Italy.

Dwork, C., Roth, A., 2013. The Algorithmic Foundations of Differential Privacy. Found. Trends® Theor. Comput. Sci. 9, 211–407. https://doi.org/10.1561/0400000042

El Emam, K., Dankar, F.K., Issa, R., Jonker, E., Amyot, D., Cogo, E., Corriveau, J.-P., Walker, M., Chowdhury, S., Vaillancourt, R., Roffey, T., Bottomley, J., 2009. A Globally Optimal k-Anonymity Method for the De-Identification of Health Data. J. Am. Med. Inform. Assoc. 16, 670–682. https://doi.org/10.1197/jamia.M3144

Elliot, M., 2014. Final Report on the Disclosure Risk Associated with the Synthetic Data Produced by the SYLLS Team. University of Manchester.

Elliot, M., 2005. Statistical Disclosure Control, in: Encyclopedia of Social Measurement. Elsevier. https://doi.org/10.1016/B0-12-369398-5/00378-9

Erlingsson, Ú., Pihur, V., Korolova, A., 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response, in: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. Presented at the CCS’14: 2014 ACM SIGSAC Conference on Computer and Communications Security, ACM, Scottsdale Arizona USA, pp. 1054–1067. https://doi.org/10.1145/2660267.2660348

Fletcher, S., Islam, M.Z., 2015. Measuring information quality for privacy preserving data mining. Int. J. Comput. Theory Eng. 7.

Fukuchi, K., Tran, Q.K., Sakuma, J., 2017. Differentially private empirical risk minimization with input perturbation, in: Int. Conf. on Discovery Science. Springer.

Fung, B.C.M., Wang, K., Chen, R., Yu, P.S., 2010. Privacy-preserving Data Publishing: A Survey of Recent Developments. ACM Comput. Surv. 42.

Hittmeir, M., Ekelhart, A., Mayer, R., 2019a. On the Utility of Synthetic Data: An Empirical Evaluation on Machine Learning Tasks, in: Int. Conf. on Availability, Reliability and Security (ARES). ACM Press, Canterbury, CA, United Kingdom. https://doi.org/10.1145/3339252.3339281

Hittmeir, M., Ekelhart, A., Mayer, R., 2019b. Utility and Privacy Assessments of Synthetic Data for Regression Tasks, in: 2019 IEEE International Conference on Big Data (Big Data). Presented at the 2019 IEEE International Conference on Big Data (Big Data), IEEE, Los Angeles, CA, USA, pp. 5763–5772. https://doi.org/10.1109/BigData47090.2019.9005476

Hittmeir, M., Mayer, R., Ekelhart, A., 2020. A baseline for attribute disclosure risk in synthetic data, in: Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy.

Kairouz, P., Oh, S., Viswanath, P., 2014. Extremal Mechanisms for Local Differential Privacy, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Kifer, D., Smith, A., Thakurta, A., 2012. Private convex empirical risk minimization and high-dimensional regression, in: Conference on Learning Theory. JMLR Workshop and Conference Proceedings.

Kohlmayer, F., Prasser, F., Eckert, C., Kemper, A., Kuhn, K.A., 2012. Flash: Efficient, Stable and Optimal K-Anonymity, in: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing. Presented at the 2012 International Conference on Privacy, Security, Risk and Trust (PASSAT), IEEE, Amsterdam, Netherlands, pp. 708–717. https://doi.org/10.1109/SocialCom-PASSAT.2012.52

LeFevre, K., DeWitt, D.J., Ramakrishnan, R., 2006. Mondrian Multidimensional K-Anonymity, in: 22nd International Conference on Data Engineering (ICDE’06). Presented at the 22nd International Conference on Data Engineering (ICDE’06), IEEE, Atlanta, GA, USA, pp. 25–25. https://doi.org/10.1109/ICDE.2006.101

LeFevre, K., DeWitt, D.J., Ramakrishnan, R., 2005. Incognito: efficient full-domain K-anonymity, in: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data - SIGMOD ’05. Presented at the the 2005 ACM SIGMOD international conference, ACM Press, Baltimore, Maryland, p. 49. https://doi.org/10.1145/1066157.1066164

Li, N., Li, T., Venkatasubramanian, S., 2007. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity, in: 2007 IEEE 23rd International Conference on Data Engineering. Presented at the 2007 IEEE 23rd International Conference on Data Engineering, IEEE, Istanbul, pp.

106–115. https://doi.org/10.1109/ICDE.2007.367856

Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M., 2006. L-diversity: privacy beyond k-anonymity, in: 22nd Int. Conf. on Data Engineering (ICDE). https://doi.org/10.1109/ICDE.2006.1

Malik, M.B., Ghazi, M.A., Ali, R., 2012. Privacy preserving data mining techniques: current scenario and future prospects, in: 2012 Third Int. Conf. on Computer and Communication Technology. IEEE.

Malle, B., Kieseberg, P., Holzinger, A., 2017. DO NOT DISTURB? Classifier Behavior on Perturbed Datasets, in: Machine Learning and Knowledge Extraction, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 155–173. https://doi.org/10.1007/978-3-319-66808-6_11

Mayer, R., Hittmeir, M., Ekelhart, A., 2020. Privacy-Preserving Anomaly Detection Using Synthetic Data, in: Singhal, A., Vaidya, J. (Eds.), Data and Applications Security and Privacy XXXIV, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 195–207. https://doi.org/10.1007/978-3-030-49669-2_11

Mendes, R., Vilela, J.P., 2017. Privacy-Preserving Data Mining: Methods, Metrics, and Applications. IEEE Access 5. https://doi.org/10.1109/ACCESS.2017.2706947

Murakami, T., Kawamoto, Y., 2019. Utility-Optimized Local Differential Privacy Mechanisms for Distribution Estimation, in: 28th USENIX Security Symposium (USENIX Security 19). USENIX Association, Santa Clara, CA, pp. 1877–1894.

Narayanan, A., Shmatikov, V., 2006. How To Break Anonymity of the Netflix Prize Dataset. CoRR abs/cs/0610105.

Nowok, B., Raab, G., Dibben, C., 2016. synthpop: Bespoke Creation of Synthetic Data in R. J. Stat. Softw. Artic. 74.

Patki, N., Wedge, R., Veeramachaneni, K., 2016. The Synthetic Data Vault, in: 2016 IEEE Int. Conf. on Data Science and Advanced Analytics (DSAA). https://doi.org/10.1109/DSAA.2016.49

Phan, N., Wu, X., Hu, H., Dou, D., 2017. Adaptive Laplace Mechanism: Differential Privacy Preservation in Deep Learning, in: 2017 IEEE International Conference on Data Mining (ICDM). Presented at the 2017 IEEE International Conference on Data Mining (ICDM), IEEE, New Orleans, LA, pp. 385–394. https://doi.org/10.1109/ICDM.2017.48

Ping, H., Stoyanovich, J., Howe, B., 2017. DataSynthesizer: Privacy-Preserving Synthetic Datasets, in: 29th Int. Conf. on Scientific and Statistical Database Management. Chicago, IL, USA.

Prasser, F., Kohlmayer, F., Lautenschläger, R., Kuhn, K.A., 2014. ARX--A Comprehensive Tool for Anonymizing Biomedical Data. AMIA Annu. Symp. Proc. AMIA Symp. 2014, 984–993.

Privacy enhancing data de-identification terminology and classification of techniques (Standard), 2018. . Int. Organization for Standardization.

Privacy framework, Amendment 1 (Standard), 2018. . Int. Organization for Standardization.

Rubin, D.B. (Ed.), 1987. Multiple Imputation for Nonresponse in Surveys, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/9780470316696

Rubinstein, B.I.P., Aldà, F., 2017. Pain-Free Random Differential Privacy with Sensitivity Sampling, in: Precup, D., Teh, Y.W. (Eds.), Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, pp. 2950–2959.

Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. WIREs Data Min. Knowl. Discov. 8. https://doi.org/10.1002/widm.1249

Samarati, P., Sweeney, L., 1998. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. https://doi.org/10.1184/R1/6625469.v1

Šarčević, T., Molnar, D., Mayer, R., 2020. An Analysis of Different Notions of Effectiveness in k-Anonymity, in: Domingo-Ferrer, J., Muralidhar, K. (Eds.), Privacy in Statistical Databases, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 121–135.

https://doi.org/10.1007/978-3-030-57521-2_9

Sattar, A.S., Li, J., Ding, X., Liu, J., Vincent, M., 2013. A general framework for privacy preserving data publishing. Knowl.-Based Syst. 54, 276–287.

Seeland, M., Rzanny, M., Alaqraa, N., Wäldchen, J., Mäder, P., 2017. Plant species classification using flower images—A comparative study of local feature representations. PLOS ONE 12, e0170629. https://doi.org/10.1371/journal.pone.0170629

Shah, A., Gulati, R., 2016. Privacy preserving data mining: techniques, classification and implications-a survey. Int J Comput Appl 137.

Slijepčević, D., Henzl, M., Klausner, L.D., Dam, T., Kieseberg, P., Zeppelzauer, M., 2021. k-Anonymity in Practice: How Generalisation and Suppression Affect Machine Learning Classifiers. ArXiv210204763 Cs.

Snoek, C.G.M., Worring, M., Smeulders, A.W.M., 2005. Early versus late fusion in semantic video analysis, in: Proceedings of the 13th Annual ACM International Conference on Multimedia - MULTIMEDIA ’05. Presented at the the 13th annual ACM international conference, ACM Press, Hilton, Singapore, p. 399. https://doi.org/10.1145/1101149.1101236

Sweeney, L., 1997. Guaranteeing anonymity when sharing medical data, the Datafly System. Proc. Conf. Am. Med. Inform. Assoc. AMIA Fall Symp. 51–55.

Tang, J., Korolova, A., Bai, X., Wang, Xueqiang, Wang, Xiaofeng, 2017. Privacy Loss in Apple’s Implementation of Differential Privacy on MacOS 10.12.

Taub, J., Elliot, M., Pampaka, M., Smith, D., 2018. Differential Correct Attribution Probability for Synthetic Data: An Exploration, in: Domingo-Ferrer, J., Montes, F. (Eds.), Privacy in Statistical Databases, Lecture Notes in Computer Science. Springer International Publishing, Valencia, Spain, pp. 122–137. https://doi.org/10.1007/978-3-319-99771-1_9

Tran, H.-Y., Hu, J., 2019. Privacy-preserving big data analytics a comprehensive survey. J. Parallel Distrib. Comput. 134.

Verykios, V.S., Bertino, E., Fovino, I.N., Provenza, L.P., Saygin, Y., Theodoridis, Y., 2004. State-of-the-art in privacy preserving data mining. ACM Sigmod Rec. 33.

Warner, S.L., 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. J. Am. Stat. Assoc. 60, 63–69. https://doi.org/10.1080/01621459.1965.10480775

Wimmer, H., Powell, L., 2014. A Comparison of the Effects of K-Anonymity on Machine Learning Algorithms. Int. J. Adv. Comput. Sci. Appl. 5. https://doi.org/10.14569/IJACSA.2014.051126

Yu, L., Liu, L., Pu, C., Gursoy, M.E., Truex, S., 2019. Differentially Private Model Publishing for Deep Learning, in: 2019 IEEE Symposium on Security and Privacy (SP). Presented at the 2019 IEEE Symposium on Security and Privacy (SP), IEEE, San Francisco, CA, USA, pp. 332–349. https://doi.org/10.1109/SP.2019.00019

11 Table of acronyms and definitions

concentris	concentris research management GmbH
DP	Differential Privacy
DMOP	Data Mining Output Privacy
GANs	generative adversarial networks
GDPR	General Data Protection Regulation
DPSGD	differentially private stochastic gradient descent
eDPSGD	extended DPSGD
GND	Gnome Design SRL
ML	Machine Learning
MS	Milestone
MUG	Medizinische Universitaet Graz
NISD	EU Network and Information Security directive
PAML	Privacy-Aware-Machine-Learning
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”.
PII	personally identifiable information
PPDM	Privacy-preserving Data Mining
PPDP	Privacy-preserving Data Publishing
RI	Research Institute AG & Co. KG
SBA	SBA Research gemeinnützige GmbH
SDU	Syddansk Universitet
SDV	Synthetic Data Vault
TUM	Technische Universitaet Muenchen
UHAM	University of Hamburg
UM	Universiteit Maastricht
UMR	Philipps Universitaet Marburg
WP	Work package