D6.8

Design of a mobile service for data anonymization and aggregation

ABSTRACT

This deliverable refers to the data anonymization and aggregation services provided by the InteropEHRate framework. These services aim at anonymizing or pseudonymizing a citizen's health data before being sent to researchers for research purposes. The deliverable points to the basic concepts related to personal data, and to the measures which will be implemented in order to protect this data. In addition, this document provides a description of the scenario that the libraries for anonymization and pseudonymization will be utilized, and finally, the design and implementation of the library for the anonymization and pseudonymization of the health data on the mobile.

Delivery Date	30th July, 2021
Work Package	WP6
Task	T6.3
Dissemination Level	Public
Type of Deliverable	Report
Lead partner	BYTE

InteropEHRate project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826106.

This document has been produced in the context of the InteropEHRate Project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826106. All information provided in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose.

CONTRIBUTORS

	Name	Partner
Contributors	Dimopoulou Stella	BYTE
Contributors	Symvoulidis Chrysostomos	BYTE
Contributors	Menesidou Sofianna	UBITECH
Reviewers	Giannetsos Athanasios	UBITECH
Reviewers	Zacharia Theodora	ISA

LOG TABLE

Version	Date	Change	Author	Partner
0.1	2021-04-26	Table of Contents	Dimopoulou Stella	BYTE
0.2	2021-06-01	Theoretical Background (Definitions, Data Anonymization, Data Pseudonymization, Anonymization vs Pseudonymization)	Dimopoulou Stella	BYTE
0.3	2021-06-04	Abstract, Introduction (Scope of the document, Structure of the document), Design and Implementation of Privacy Mechanisms (Scenario Description, Data Anonymization and Aggregation User Requirements, Design of the Library, Library functionalities, Data Pseudonymization, 1st Variant: Pseudo-identities, Data Anonymization)	Dimopoulou Stella	BYTE
0.4	2021-06-06	Abstract, Contributors, Acronyms, Introduction (Scope of the document, Structure of the document) Design and Implementation of Privacy	Dimopoulou Stella, Menesidou Sofianna	BYTE, UBITECH

		Mechanisms (2nd Variant: Pseudonyms)		
0.5	2021-06-07	List of figures, List of tables, Conclusions and Next Steps, References	Dimopoulou Stella	BYTE
0.6	2021-06-08	Contributors, Log Table, Introduction (Intended audience), Theoretical Background (Anonymization and Pseudonymization on the Mobile)	Dimopoulou Stella	BYTE
0.7	2021-06-09	Updates on Design and Implementation of Privacy Mechanisms	Symvoulidis Chrysostomos	BYTE
0.8	2021-06-10	Updates on Acronyms, Conclusions and Next Steps, References	Dimopoulou Stella	BYTE
0.9	2021-06-29	Updates on Design and Implementation of Privacy Mechanisms (Design of the Library)	Dimopoulou Stella	BYTE
1.0	2021-07-20	Quality Check	Argyro Mavrogiorgou	UPRC
VFinal	2021-07-30	Final revision and submission	Laura Pucci	ENG

ACRONYMS

Acronym	Term and definition
DLP	Discrete Logarithm Problem
GDPR	General Data Protection Regulation
RDSAnonI	RDS Anonymization/Pseudonymization Interface
S-EHR App	Smart EHR Mobile Application
TTP	Trusted Third Party
RDD	Research Definition Document
RRC	Reference Research Center
RDS	Research Data Sharing
RDSI	RDS Interface
PI	Principal Investigator
PII	Personally Identifiable Information
PP	Pseudonym Provider
FHIR	Fast Healthcare Interoperability Resources
JSON	JavaScript Object Notation
DICOM	Digital Imaging and Communications in Medicine
VR	Value Representation

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Scope of the document

In the Research Scenario, both data anonymization and data pseudonymization will be implemented in order to prevent the identity of the citizens from being disclosed when their data is sent to the research centers. When a researcher wants to conduct a specific research, the citizen's data will be gathered and processed — anonymized or pseudonymized — locally on his/her phone before being sent back to the researcher. For this purpose, a library named RDSAnonI is implemented in order to anonymize or pseudonymize the citizen's health data on the phone before it is shared with the researchers.

This deliverable aims to clarify some basic concepts related to personal data. In addition, it mentions the privacy mechanisms which should be implemented when the personal data is processed and used when conducting research studies, as well as it points out their main differences. In the context of InteropEHRate, the scenario regarding the data anonymization and pseudonymization mechanisms that will be implemented on the health data is described, while both the design of the library and the functionalities that the library offers in the mobile application (S-EHR App) are mentioned.

1.2. Intended audience

The audience targeted by this deliverable is mostly developers and researchers in the context of InteropEHRate. The developers may want to understand the design of the library which anonymizes and pseudonymizes the citizens' health data. Researchers because both these two privacy mechanisms – data anonymization and data pseudonymization – are highly recommended for the researchers when processing health data in a research study.

1.3. Structure of the document

The structure of this deliverable is the following.

- **Section 1 — Introduction:** States a brief description of the contents and purpose of the deliverable, along with its structure.
- **Section 2 — Theoretical Background:** Defines some basic concepts according to ISO 29100:2011 and the General Data Protection Regulation, and refers to the anonymization and pseudonymization of the data.
- **Section 3 — Design and Implementation of Privacy Mechanisms:** Specifies the scenario in which the privacy mechanisms — data anonymization and data pseudonymization – are implemented, the design of the RDSAnonI library and its functionalities.
- **Section 4 — Conclusions and Next Steps:** Sets out the conclusions of the deliverable and the next steps.

1.4. Updates with respect to previous version (if any)

Not applicable.

2. THEORETICAL BACKGROUND

2.1. Definitions

This section lists some basic concepts and definitions according to ISO 29100:2011 and the General Data Protection Regulation (GDPR), in order to clarify the relevant terminology, which is used throughout the deliverable regarding the anonymization and pseudonymization of data.

Personal data: Any information which may lead, directly or indirectly, to the identification of a natural person, for example the name, the surname, the identity card number etc. **[GDPR]**. According to ISO 29100:2011 **[ISO 29100]** this data is also named as personally identifiable information (PII).

Data subject: A natural person becomes a data subject when it is associated with some personal data (personally identifiable information) **[GDPR], [ISO 29100]**. The data subjects provide their personal data to the data controllers and data processors **[ISO 29100]**.

Data controller: Any natural or legal person, agency, body, or public authority, who is responsible for determining the purpose and the manner of the processing of the personal data **[GDPR]**. The purpose of processing, as well as the means of this processing, should be legal and prescribed by the legislation **[GDPR]**. The data controller is the one who will decide who can process the personal data **[ISO 29100]**. The processing can be delegated either in whole or in part to one or more privacy stakeholders **[ISO 29100]**.

Data processor: Any natural or legal person, agency, body, or public authority, who processes the personal data of the data subject on behalf of the data controller **[GDPR]**. In other words, the data processor acts on behalf of the data controller and processes the personal data in accordance with the instructions given by him **[GDPR]**.

Sensitive data: The special categories of personal data, also known as sensitive data, is a data category which may cause significant impact on the data subjects if they are disclosed. Sensitive data should be kept confidential and protected against unauthorized access **[ISO 29100]**. Examples of such data are the religious beliefs, the political opinions, and the health data **[ISO 29100]**. The sensitive data should be kept separately from the personal data, and more protection is required when processing this data, as there is a high risk of trespassing the human rights and the freedom of the individuals **[GDPR], [ISO 29100]**.

Processing: Any operation which can be applied to personal data, such as collection, storage, transmission, deletion, etc., with or without the usage of automated means **[GDPR]**.

2.2. Privacy Protection Mechanisms

2.2.1. Data Anonymization

Data anonymization is a process, in which all personally identifiable information is deleted or modified, so that all data remained of this process cannot lead to the identity of the data subject **[PIWIK PRO]**. The anonymized data that will occur is no longer associated with a natural person and the data controller will not be able to link this data with an individual, either directly or indirectly **[PIWIK PRO]**.

Data anonymization is a one-way process, which means that when this process is applied to a dataset, no one will be able to identify the data subject and find out who is the owner of the anonymized data [DATA PRIVACY MANAGER]. If someone can be led to the disclosure of a natural person, then the anonymization process was not implemented properly, which is a violation/data breach of the personal data of the data subject.

2.2.2. Data Pseudonymization

Data pseudonymization is a process, in which all personally identifiable information is replaced with pseudonyms. The data occurred of this process cannot lead to the identification of the data subject unless there is some additional information. This information should be maintained separately from the pseudonymized data. Furthermore, access to this information should only have the data controller or the person who is responsible and authorized to reverse the procedure if this is the case (e.g., in case of an emergency scenario in healthcare data) [PIWIK PRO]. The above-mentioned additional information is essentially a mapping table, which retains the personal data of the data subjects and the corresponding pseudonyms. Pseudonyms are unique identifiers, and each data subject should have a different pseudonym [i-SCOOP]. These pseudonyms are considered as personal data since they are related with natural persons and can be led to the identity of the data subject [PIWIK PRO].

Data pseudonymization is implemented in order to limit the number of the people who can access the data, and identify the data subjects from their personal data, when they are processing them to fulfil a specific purpose [i-SCOOP]. Through pseudonymization, both the data controllers and the data processors will not be able to know the data subject to which the dataset belongs, as all the identifiable information will have been replaced by a pseudonym. This kind of processing is very useful both in the processing of personal data and in the processing of sensitive data, since an indirect relation between the natural person and his personal data will be maintained [DATA PRIVACY MANAGER]. In addition, only the data controller, who will preserve the mapping table, can identify the data subject if needed.

Data pseudonymization is considered to be implemented correctly, when the remaining attributes/data, with the corresponding pseudonyms, cannot be linked to a natural person, and the procedure cannot be reversed unless the additional information is used (the mapping table) [i-SCOOP]. In any other case, there is a breach of the personal data of the data subject.

2.2.3. Anonymization vs. Pseudonymization

Data anonymization is an irreversible process, as after it is implemented, there is no longer any association between the anonymized data and the natural person. The General Data Protection Regulation is applied only in the processing of personal data, where the personal data is related to a natural person and can lead, directly or indirectly, to the identification of the data subject. When the data is anonymized, it cannot lead to the disclosure of a natural person, since the data is not associated with it [PIWIK PRO]. Thus, the General Data Protection Regulation is not applied to this data [i-SCOOP]. Based on the above mentioned, anonymized data can be shared with third parties, without the owner's consent [PIWIK PRO]. This happens, as once the personal data is fully and properly anonymized, it is no longer considered as personal data, and the General Data Protection Regulation does not apply to subsequent uses.

On the other hand, data pseudonymization is a reversible process, since the data controller is able to find the identity of the natural person using the mapping table [PIWIK PRO]. Since pseudonymized data is considered as personal data, due to the indirect identification of the natural person through the mapping table, the General Data Protection Regulation applies to them [i-SCOOP]. It is worth noting that unauthorized reversal is a data breach, while any reversal occurred, should be prescribed by the law or the purposes of the processing of personal data.

2.2.4. Anonymization and Pseudonymization on the Mobile

Nowadays more and more wearable devices, which collect the individuals' personal data, are implemented. Since these devices gather private information about their users, it is very important to protect the privacy of the individuals [MALEKZADEH 2019]. To achieve this, there are many privacy measures which can be implemented. However, in this deliverable two privacy mechanisms were presented, which are data anonymization and data pseudonymization. In this way, the users' data can be collected and processed, without disclosing their identity [MALEKZADEH 2019].

Even though both data anonymization and data pseudonymization are state of the art techniques, due to the increasing number of the devices which collect data through sensors [MALEKZADEH 2019], there are not many publicly known implementations that support these two privacy mechanisms on the mobile phone.

Data anonymization can also be applied in a variety of location-based services (LSBs), since they have become very popular due to the fact that GPS devices and mobile information technologies are in rapid progress. This kind of services can answer queries such as "what is the nearest cafeteria from my current location" and provide a visual map with all cafeterias which were found around the target point (the point of the user who made the query). Although location-based services are widely used and useful, the privacy of the users is at stake since there are risks with regard to the user's data. The location of the user is considered as personal data, and this is why this information is important with terms of privacy. Since the current location of a user must be shared with the service provider in order for the location-based service to be used, the data which contain information about the citizen's location is considered as a source of privacy information leakage. With the assumption that the service provider combines the user's location information with other personal information, the provider may be able to extract information about the user per se, which is considered a data breach [MANO 2010]. This personal information, which is related to the user, might be information such as political affiliations and medical problems and may get exposed to an adversary. By extension, even though a data subject does not provide any information regarding its identity when sharing his location, an adversary may be able to obtain personal information through location tracking or space and time correlation inference. In this way there may be an exposure of the complete movements of the user [GEDIK 2005]. To address this issue, a technology regarding the anonymization of the user's location has occurred [MANO 2010].

A widely known technique with regard to location anonymity is k-anonymity technique, which is used in order to anonymize the precise location of the mobile users [MANO 2010], [GEDIK 2005]. In k-anonymity, a data subject is considered as k-anonymous if its location information, which is sent via his mobile phone, cannot be distinguished from the location information of at least k – 1 data subjects within a specific region **[GEDIK 2005]**. More specifically, when a user requests a location-based service via his mobile, a trusted third party (TTP), which is placed between the user's phone and the application provider, will find the k – 1 users nearby the target user. After that, the TTP will send the request to a cloaking region, which encloses the locations of k users and is treated as an approximate anonymized location for the target user. In this way, the service provider will only know that the target user – the user who requested that specific location – is located within the given cloaking region. The advantage of this technique is that even if the service provider can be led to the identity of all k users within a region, it cannot distinguish the target user from the k users who were identified **[MANO 2010]**.

3. DESIGN AND IMPLEMENTATION OF PRIVACY MECHANISMS

3.1. Scenario Description

This section refers to the usage of data anonymization and data pseudonymization, and their implementation in the research scenario. More specifically, either data anonymization or data pseudonymization mechanism is applied to the citizen's health data, when it is going to be shared for satisfying specific research purposes. These privacy mechanisms are used in order to modify the data locally on the citizen's phone, so that the remaining data cannot lead to the identification of its owner. In this way, the researchers can process the health data without knowing the identity of the citizen. In some cases, for example in case some abnormal or unusual values appear in the citizen's data, which means that the citizen might have a health problem, the researcher may have the ability to trace back the citizen and inform him/her about this issue. Since data anonymization is an irreversible process, this is not feasible, and this is why data pseudonymization is implemented in this scenario too. The decision regarding the privacy mechanism that will be used in each research study is stated inside the Research Definition Document (RDD). What depends also on the study is whether a pseudo-identity or a pseudonym is generated in data pseudonymization.

According to the deliverable D4.8 **[D4.8]** the health data of the citizen should fulfil the enrolment criteria. If the citizen wants to proceed in the processing of his/her data and provides his/her consent, then in case of pseudonymization either a pseudo-identity or a pseudonym will be generated. These two values — the pseudo-identity and the pseudonym — are two variants that will be used only in case of data pseudonymization in order to re-identify the citizen. In addition, in the data retrieval phase, the citizen's health data is gathered from his/her phone, where it should be anonymized or pseudonymized locally before being sent to the Reference Research Center (RRC).

3.2. Data Anonymization and Aggregation User Requirements

In this section, there will be presented the user requirements regarding the data anonymization and data pseudonymization. More specifically, there are two main non-functional user requirements. The first user requirement refers to the anonymization and pseudonymization privacy mechanisms that will be implemented in order to anonymize or pseudonymize the citizen's data, respectively. The second user requirement concerns the pseudo-identities and refers to the uniqueness of them. Each citizen should have a different pseudo-identity for a different research study, and two citizens cannot have the same pseudo-identity for the same study. After the pseudo-identity is generated, it will be used in order to pseudonymize the data which will be sent to the RRC.

Table 1 depicts the two user requirements **[D2.3]** that relate to the Data Anonymization and Aggregation mechanisms.

#	Title	Main actor	Description
91	Automatic anonymization and sharing of citizen's health data for research.	-	After a citizen accepted an invitation to a research study and the study is started according to the specified protocol, the S-EHR automatically queries its content for the data required by the study, once or periodically, depending on the study, and automatically sends the matching data to the InteropEHRate Research Network. The S-EHR anonymizes the data before sharing it, if required by the protocol.
130	Pseudoidentity restricted to single research protocol.	-	When a citizen gives a digital consent to participate to a research protocol, a specific pseudo-id for that patient will be generated, to be used only for the pseudonymization of data shared within that specific research protocol.
217	Anonymization of structured content in DICOM studies	Data scientist	The healthcare organisation can anonymise structured content of DICOM studies

Table 1 – User Requirements

3.3. Design of the Library

The design of the data anonymization and pseudonymization library is described in this section.

The RDS-Anonymization library (RDSAnonI) is a library that communicates with some of the Research Data Sharing (RDS) protocol components. These components are the RDS-Logic library, the Pseudonym Provider and the RDS Interface. According to the deliverable **[D4.10]**, the RDS-Logic library is the main library of the RDS protocol. It communicates with all the other libraries in the S-EHR App and implements most functionalities of the RDS protocol. In addition, the Pseudonym Provider (PP) is responsible for generating and managing the pseudonyms, and the RDS Interface (RDSI) is a remote API which allows the S-HER App to share the citizen's health data with the RRC (as well as other functionalities referred in the deliverable **[D4.8]**).

The libraries constructing the RDS protocol are depicted in Figure 1.

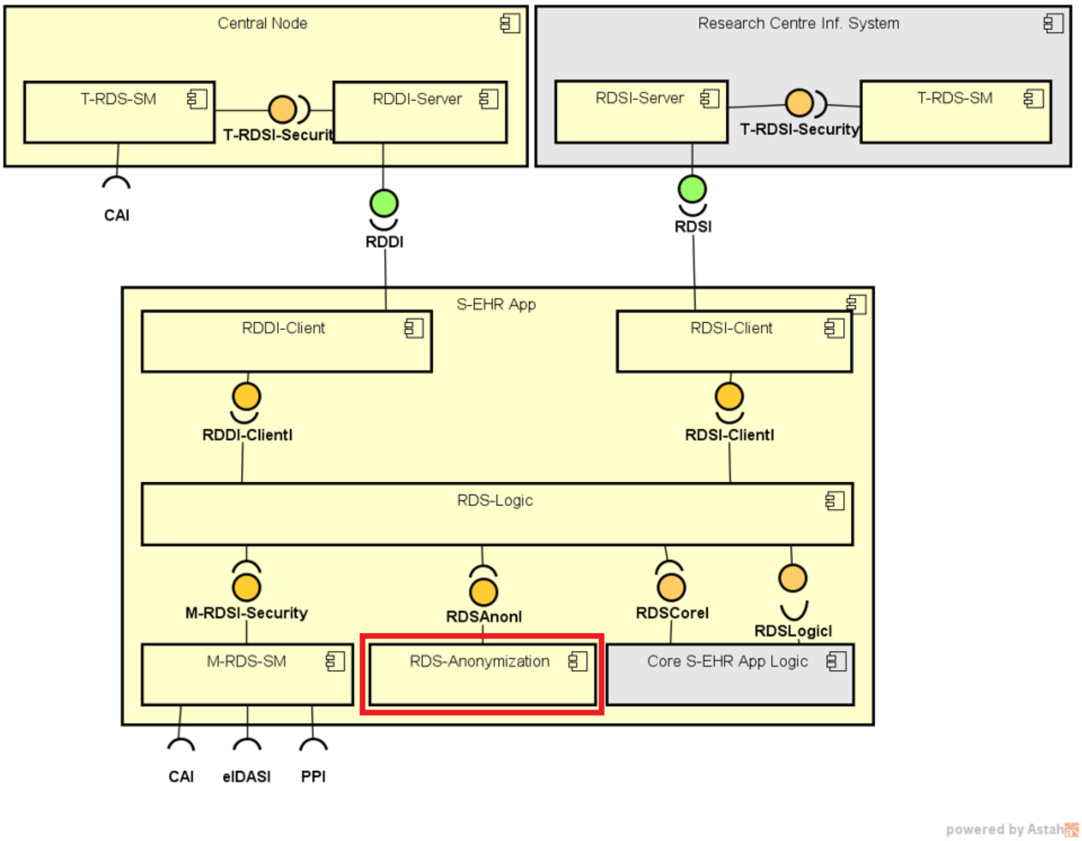

Figure 1 - Research Data Sharing Components

The RDS-Anonymization library (RDSAnonI) — enclosed in a red frame — is analyzed in the current and in the following section. Briefly, this library will be used in the Research Data Sharing (RDS) protocol in order to anonymize or pseudonymize the citizen's health data for a research study. With regard to the structured data, the citizen's data will be anonymized or pseudonymized within his/her phone, whereas the unstructured data will be pre-anonymized and uploaded on the phone.

The following table depicts all five public methods which will be implemented inside the RDSAnonI library, and which will be used in order to achieve data anonymization and pseudonymization.

	Description	**Input**	**Output**
setPseudo	In case of pseudonymization, set the pseudo-identity or the pseudonym.	pseudoType, pseudo, studyID	void
getPseudo	Get the pseudo-identity or pseudonym, which was set.	studyID	pseudo

anonymizeData	Anonymize structured data.	data, fileType	anonymizedData
pseudonymizeData	Pseudonymize structured data.	data, fileType, studyID	pseudonymized Data
retrievePseudonym	Retrieve a pseudonym from the Pseudonym Provider.	anonymous SAML assertion token	pseudonym

Table 2 – Design of RDSAnonl library

The arguments of each method are described below. More specifically,

- The pseudoType defines whether the pseudo variable is a pseudo-identity or a pseudonym.
- The pseudo is either a pseudo-identity or a pseudonym.
- The study ID is an identifier which describes the study.
- The data is an FHIR bundle, which will be either anonymized or pseudonymized.
- The anonymous SAML assertion token is a token which is stored in the citizen's phone, already acquired by eIDAS authentication, and it will be used in order to authenticate to the PP for pseudonym generation.

The above-mentioned methods are represented in more detail as follows.

Name	**setPseudo**
Description	Depending on the requirements of the study, set either a pseudo-identity or a pseudonym for the pseudonymization of a citizen's health data.
Caller	RDS-Logic
Arguments	● pseudoType: indicates whether the library should use a pseudo-identity or a pseudonym ● pseudo: sets the pseudo-identity/pseudonym which will substitute a citizen's personal information ● studyID: the research study to which the pseudo-id applies
Return Value	void
Exceptions	● Thrown if any of the input values is null.
Preconditions	● none

Table 3 – Method: setPseudo

Name	**getPseudo**
Description	Get the pseudo associated to the study indicated as input.
Caller	RDS-Logic
Arguments	● studyID: the research study to which the pseudo-id applies
Return Value	● pseudo: gets the pseudo-identity/pseudonym which was assigned for the given studyID
Exceptions	● InvalidStudyException: if the study ID in input does not have a pseudo associated.
Preconditions	● none

Table 4 – Method: getPseudo

Name	**anonymizeData**
Description	Anonymize structured data.
Caller	RDS-Logic
Arguments	● data: a FHIR bundle containing FHIR resources, attributes, and values, that need to be anonymized ● fileType: the type of the data file
Return Value	The same FHIR bundle as in the input, except that identifying data provided in the input is removed.
Exceptions	● FHIRParsingException: in case the input is not parseable as a FHIR bundle conform to the InteropEHRate Interoperability Profiles.
Preconditions	● none

Table 5 – Method: anonymizeData

Name	**pseudonymizeData**
Description	Pseudonymize structured data.
Caller	RDS-Logic
Arguments	● data: a FHIR bundle containing FHIR resources, attributes, and values, that need to be pseudonymized

	● fileType: the type of the data file
Return Value	The same FHIR bundle as in the input, except that identifying data provided in the input is replaced with the pseudo-id/pseudonym.
Exceptions	● FHIRParsingException: in case the input is not parseable as a FHIR bundle conformant to the InteropEHRate Interoperability Profiles. ● PseudoException: in case the pseudo-identity or the pseudonym is not set.
Preconditions	● The pseudo-id/pseudonym must be set.

Table 6 – Method: pseudonymizeData

Only for pseudonym-based studies:

Name	**retrievePseudonym**
Description	Allows a S-EHR App to receive a pseudonym from a trusted third party that acts as a PP. This trusted third party could also be the RRC or any other entity.
Caller	The RDS-Logic library running on the S-EHR App.
Arguments	● anAssertion: Anonymous assertion token
Return Value	A string containing the pseudonym generated.
Exceptions	● Invalid content (study ID).
Preconditions	● The S-EHR App has already been authenticated by an eIDAS node.

Table 7 – Method: retrievePseudonym

As can be seen in the figure below (Fig. 2), the library provides two privacy mechanisms which can be implemented on the citizens' data.

In the case of data pseudonymization there are two variants which can be used in order to replace the citizens' personal data. The first variant is the pseudo-identity. The RDS-Logic **[D4.8]** retrieves the pseudo-identity from the RDSI (RDS Interface **[D4.8]**) and then the RDSI-Logic library sets this pseudo-identity to the RDSAnonI library. The second variant is the pseudonym. The RDS-Logic library retrieves the pseudonym from the Pseudonym Provider (PP) through the RDSAnonI, and then sets the pseudonym to the RDSAnonI library. After the pseudo-identity or the pseudonym is set, data pseudonymization is achieved in the RDSAnonI library, and the pseudonymized dataset is returned to the RDS-Logic in order to be sent to the Reference Research Center (RRC).

In the case of data anonymization, the RDS-Logic sends the original dataset to the RDSAnonI library, the RDSAnonI anonymizes the dataset and returns it to the RDS-Logic, again in order to send it to the Reference Research Center.

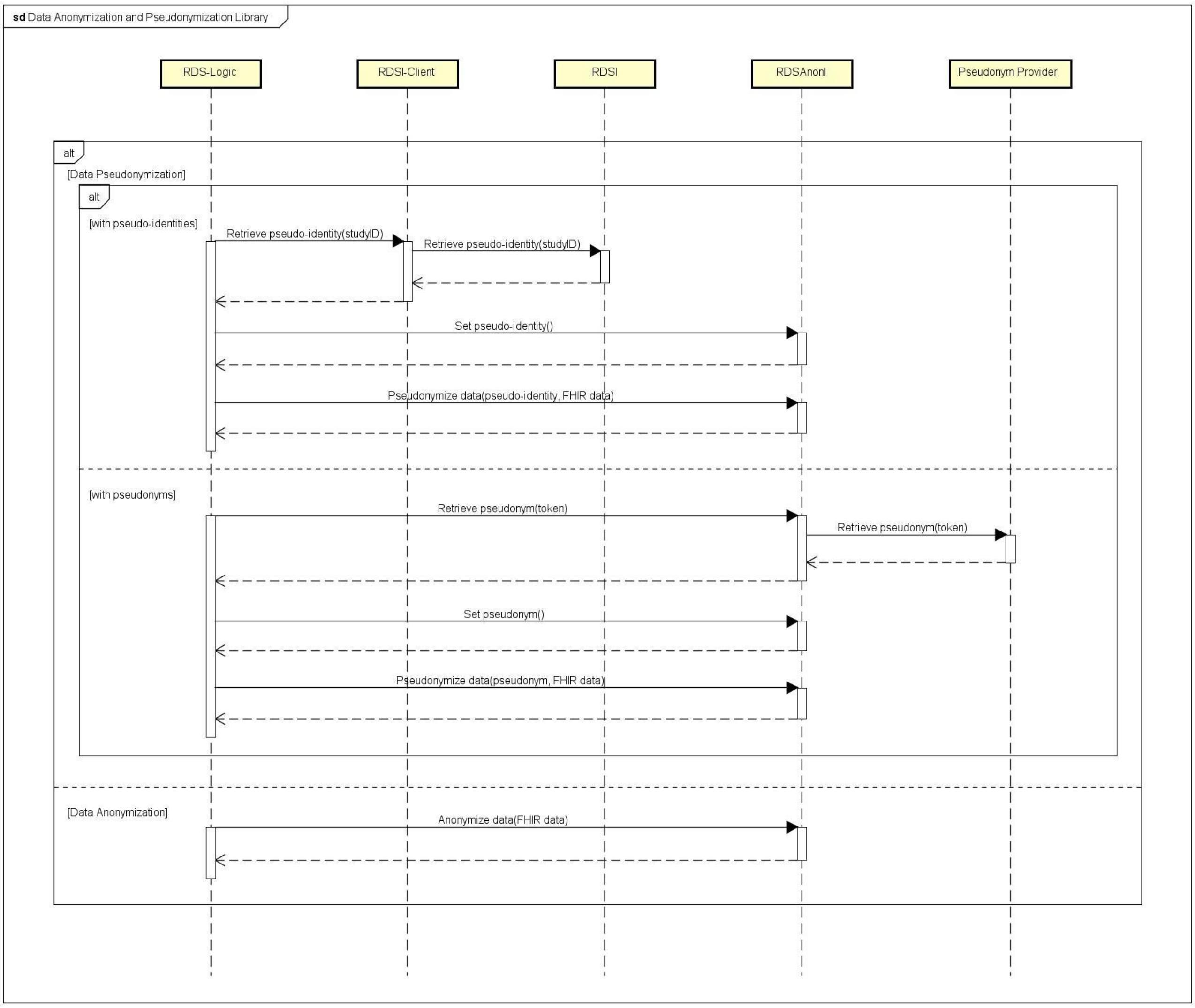
sd Data Anonymization and Pseudonymization Library
RDS-Logic
RDSI-Client
RDSI
RDSAnonI
Pseudonym Provider
alt
[Data Pseudonymization]
alt
[with pseudo-identities]
Retrieve pseudo-identity(studyID)
Retrieve pseudo-identity(studyID)
Set pseudo-identity()
Pseudonymize data(pseudo-identity, FHIR data)
[with pseudonyms]
Retrieve pseudonym(token)
Retrieve pseudonym(token)
Set pseudonym()
Pseudonymize data(pseudonym, FHIR data)
[Data Anonymization]
Anonymize data(FHIR data)

Figure 2 - Data Anonymization and Pseudonymization Library

3.4. Library functionalities

As it was already mentioned in the Scenario Description, the health data is either anonymized or pseudonymized on the citizen's phone, and it is sent to the RRC in order to be used for research purposes. In this section, both data anonymization and data pseudonymization mechanisms will be analyzed, as well as how they are integrated in the context of InteropEHRate Framework.

To begin with, the RRC is responsible for choosing the privacy mechanism which is going to be used in the research. This mechanism is defined within the RDD. If it is desirable to identify a citizen, then data pseudonymization should be implemented. In this case, there are two variants which can be used in order to retain an association between the data and its owner (i.e. the citizen). The first variant is the pseudo-identity and the second one is the pseudonym. On the other hand, if it is not needed to be led to the identity of the citizen, and no relation should be kept, then data anonymization should be implemented.

In both data pseudonymization and data anonymization the data to be anonymized or pseudonymized is in FHIR format. The files which are supported by the RDSAnonI library – which get pseudonymized or anonymized – are JSON files. A sample of a FHIR resource is depicted in the following snippet:

```
"identifier" : [
  {
      "system" : "http://interopEHRate.eu/fhir-resource/",
     "value" : "Patient/MS01"
  }
],
"name" : [
  {
      "family" : "Smith",
     "given" : [
           "Markus"
     ]
  }
],
"gender" : "male",
"birthDate" : "2013-12-05",
"address" : [
  {
     "use" : "home",
     "type" : "physical",
     "city" : "Rome",
     "country" : "IT"
  }
]
```

Table 8 – Original JSON file

3.4.1. Data Pseudonymization

As it was already mentioned, the RDD includes the privacy mechanism which will be implemented on the dataset. In addition, it contains policies and information about the research and the necessary data which is needed for conducting this research. Depending on the research study, there are two alternatives/variants for achieving data pseudonymization. The first variant is the pseudo-identity, which is generated at the RRC if the citizen gives his/her consent to participating in the study. The second variant is the pseudonym which is generated by the Pseudonym Provider, a third-party trusted service.

The reason why there are two variants in case of data pseudonymization is the following. First of all, the pseudo-identities are human-readable which makes them also more practical, and both the hospitals and the researchers prefer them as opposed to the pseudonyms. On the other hand, the pseudonyms are more secure due to their higher degree of randomness, in contrast with the pseudo-identities which have limited randomness. Consequently, in order to satisfy the above-mentioned requirements, there will be two alternatives implemented.

As depicted in the table above (Table 8), the example of the FHIR JSON file consists of some keys and values. In order to pseudonymize the data, all citizen’s personal data — all the keys that corresponds to personal information — should be replaced either with a pseudo-identity or a pseudonym. More specifically, the value of one key should be replaced with a pseudo-id/pseudonym and the remaining values should be deleted. In the table below (Table 9), there is an example of the implementation of pseudonymization with pseudo-identities. Considering that the pseudo-identity is the “12521ef1d71643fa87d1f35778139cfadfe073580fb50b38fae6c1df678fa783c285c”, the key “given” has been replaced with this pseudo-identity and the other values (system, value, family, birthDate, city, and country), which are personal data, have been deleted.

```
"identifier" : [
  {
      "system" : "",
     "value" : ""
  }
],
"name" : [
  {
      "family" : "",
     "given" : [

     "12521ef1d71643fa87d1f35778139cfadfe073580fb50b38fae6c1df678fa783c
285c"
     ]
  }
],
"gender" : "male",
"birthDate" : "",
"address" : [
  {
     "use" : "home",
     "type" : "physical",
     "city" : "",
     "country" : ""
  }
]
```

Table 9 – Pseudonymized JSON file (with pseudo-identity)

3.4.1.1. 1st Variant: Pseudo-identities

The pseudo-identity is an alphanumeric sequence which is generated by the pseudo-identity generation service and consists of three parts, the prefix, an incremental number and the suffix. In order to generate the service these pseudo-identities, the study ID should be defined. The study ID is essentially a unique identifier which describes the study, and it is the first part — the prefix — of the pseudo-identity. Every time that a pseudo-identity is generated for a citizen for a specific study, the incremental number increases by one. This happens in order to create different pseudo-identities per study and per citizen. To retain the incremental number for each study, an SQL database is used. This database keeps the study ID and the current incremental number. The third part of the pseudo-identity is the suffix. The suffix is the result of a hash function — SHA-256 **[THE SSL STORE]** — which takes as input the prefix along with the incremental number and returns the hash value, which is a random value. In more detail, the pseudo-identity is constructed as follows:

```
pseudo-identity = prefix + incremental_number_for_current_study + suffix
```

Where:

- `prefix`: the ID of the study
- `incremental_number`: an incrementally increasing integer
- `suffix`: result of SHA-256 hash function

After the aforementioned values are defined and the pseudo-identity is generated, it should be sent both to the citizen (S-EHR App) and to the Principal Investigator (PI) who is responsible for keeping the mapping table that retains an association between the citizen and his/her data. Only the PI should have access to the mapping table and re-identify the citizen if needed.

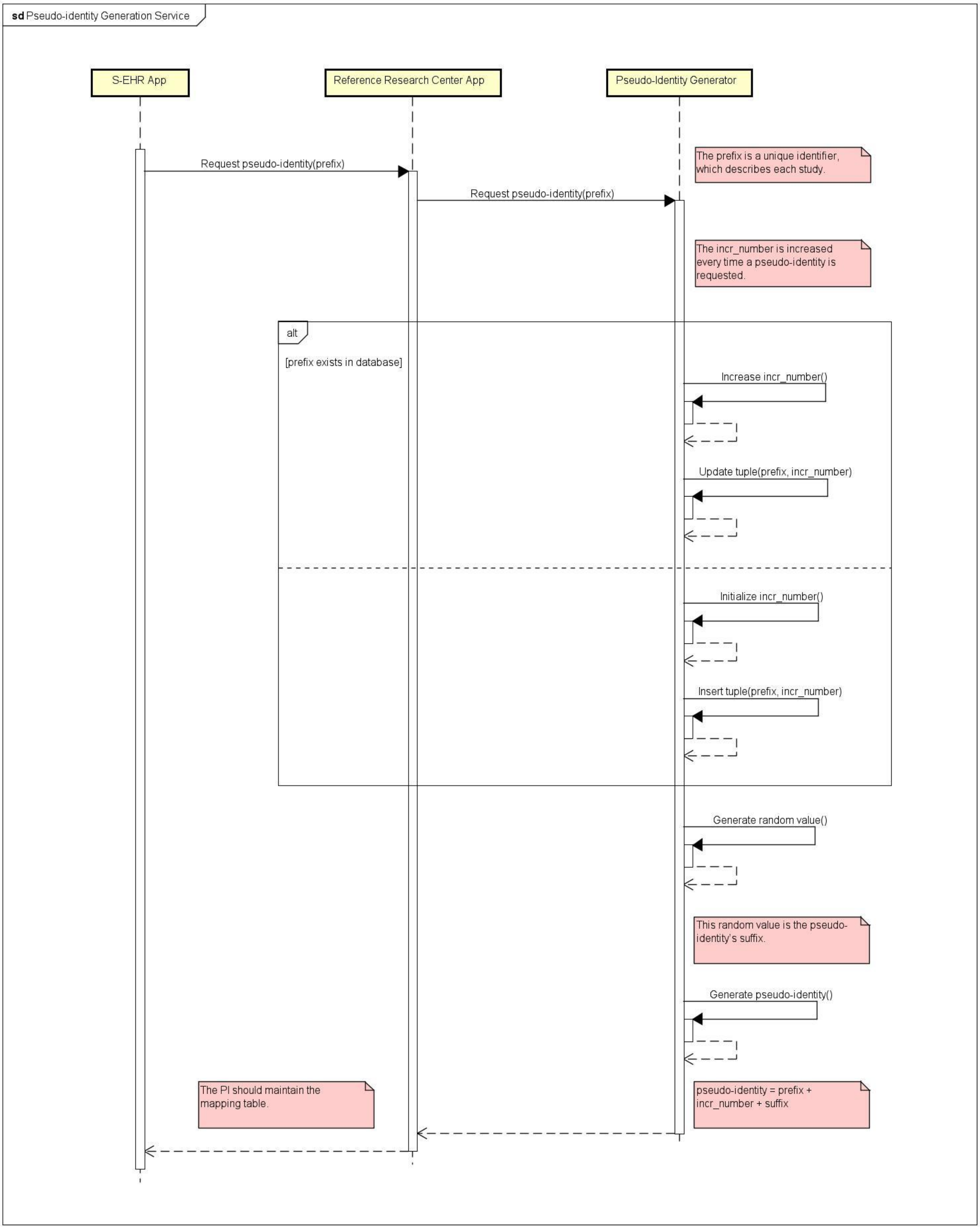
sd Pseudo-identity Generation Service
S-EHR App
Reference Research Center App
Pseudo-Identity Generator
Request pseudo-identity(prefix)
Request pseudo-identity(prefix)
The prefix is a unique identifier, which describes each study.
The incr_number is increased every time a pseudo-identity is requested.
alt
[prefix exists in database]
Increase incr_number()
Update tuple(prefix, incr_number)
Initialize incr_number()
Insert tuple(prefix, incr_number)
Generate random value()
This random value is the pseudo-identity's suffix.
Generate pseudo-identity()
The PI should maintain the mapping table.
pseudo-identity = prefix + incr_number + suffix

Figure 3 - Pseudo-identity Generation Service

After the pseudo-identity is generated by the service running at the RRC, then the RDS-Logic library retrieves the pre-referred pseudo-identity and sets it to the RDSAnonI library. Using this pseudo-identity, the RDSAnonI library pseudonymizes the citizen's dataset and sends it to the RRC through the RDS-Logic.

3.4.1.2. 2nd Variant: Pseudonyms

Pseudonym is an anonymous short term credential generated by a trusted Pseudonym Provider (PP). A PP is a trusted organisation responsible for the pseudonym management (according to the IEEE 1609.2-2013 specification), and is used for the anonymous communication of the citizen's health data to a (Research) Reference Centre (RRC). The need for a PP per country stems from the specific privacy requirements and healthcare data protection legislation that differs significantly between countries. The anonymous credential system or pseudonym system initially introduced in **[Chaum 1985]**, where blind signatures based on RSA signatures and based on the discrete logarithm problem (DLP) were adopted for the pseudonym construction. This work generalized and improved in both the DLP and strong RSA assumption settings **[Brands 2000]**. Another similar anonymous credential concept, also based on a proof of knowledge technique, was introduced in **[Camenisch 2002]** and uses group signatures, with the advantage of multi-show unlinkable credentials from application's point of view. In the second variant, the construction of the pseudonym will be based on the more advanced crypto primitives of group-based signatures introduced in **[Camenisch 2002].** Such a scheme provides higher user privacy levels even in the complex scenario where users move around different domains (i.e., countries or member states) and they need to acquire pseudonyms without revealing personal information regarding their country of origin. More details regarding the second variant can be found in the D3.4.

3.4.2. Data Anonymization

The second privacy mechanism is data anonymization. In this case, after the data is anonymized, then it cannot lead to the citizen's identity, since the process is irreversible, and the re-identification is not feasible anymore.

The dataset which will be used for each research study can be classified in two main categories, namely structured and unstructured data. The structured data will be anonymized locally on the citizen's phone before being sent to the RRC. More specifically, the RDS-Logic library will send the dataset to the RDSAnonI library, and the RDSAnonI library will return the anonymized dataset back to the RRC through the RDS-Logic library.

3.4.2.1. Anonymization of FHIR Resources on S-EHR App

Anonymization of structured data, and more specifically of FHIR json files, can be achieved by removing all the values which represent personal information. Based on this assumption, all personal data of the pre-referred json files should be deleted and only the necessary information should be maintained. For example, the anonymized version of the json file of Table 8 is depicted in Table 10 and the values which have been deleted are the following: system, value, family, given, birthDate, city, and country.

```
"identifier" : [
  {
      "system" : "",
     "value" : ""
  }
],
"name" : [
  {
      "family" : "",
     "given" : [
          ""
     ]
  }
],
"gender" : "male",
"birthDate" : "",
"address" : [
  {
     "use" : "home",
     "type" : "physical",
     "city" : "",
     "country" : ""
  }
]
```

Table 10 – Anonymized JSON file

3.4.2.2. Anonymization of Structured content in DICOM studies

The electronic health data provided by the healthcare organization to the citizen are often in unstructured format and more specifically, in cases of images, in DICOM format. DICOM (Digital Imaging and Communications in Medicine) is a standard protocol which is used in the health sector in order to manage and transmit the medical images **[TECHTARGET]**. The anonymization of the DICOM images is a very challenging process since personal data is included in both structured format – in the attributes within the DICOM dataset — and in unstructured format — in pixel level on the image. With regard to the structured content of the DICOM images, the process of the anonymization can be achieved by removing or replacing the attribute values which may lead to the identification of the citizen in order to achieve both the anonymization and maintain the usefulness of the image. The reason why the attributes are not only removed but also replaced with random values is that the removal of the attributes may render the DICOM image unusable **[IMAIOS]**.

On the other hand, the anonymization of the unstructured content of the DICOM images is very difficult since many factors may affect and differentiate the whole process. For example, a specific device used by a specific healthcare organization in order to generate the DICOM images may store the personal data on the image in a different position compared to another device of another healthcare organization.

In the context of InteropEHRate, the requirement with regard to the anonymization of the structured content of the DICOM images will be satisfied through a service running on the side of the healthcare organization. On the contrary, the anonymization of the unstructured content of the DICOM images will not be implemented, since it is out of scope, and it will be achieved with tools chosen by the hospitals.

As it was already mentioned, the anonymization operation of the structured content of the DICOM images will be implemented via a service in each healthcare organization. The reason why this process is not implemented on the mobile phone – via a library in the S-EHR App – is that it is time-consuming, and the computational resources needed in order to process the images are beyond the capabilities of the current mobile phones, especially in the cases where several medical images need to be anonymized. The sequence of steps to be followed in order to anonymize the attributes, which contain personal data, is as depicted in the figure below (Fig. 4).

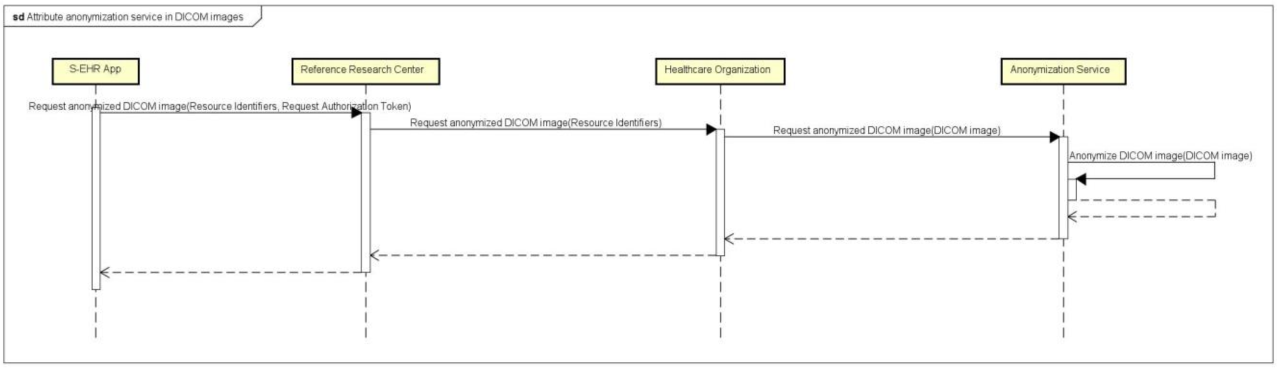
sd Attribute anonymization service in DICOM images
S-EHR App
Reference Research Center
Healthcare Organization
Anonymization Service
Request anonymized DICOM image(Resource Identifiers, Request Authorization Token)
Request anonymized DICOM image(Resource Identifiers)
Request anonymized DICOM image(DICOM image)
Anonymize DICOM image(DICOM image)

Figure 4 - Attribute anonymization service in DICOM images

1. As it was already mentioned in the Scenario Description, before the data is being retrieved from the citizen's smartphone, in order to be sent to the RRC, it should get anonymized or pseudonymized. In case of the DICOM images, the S-EHR App should provide some resource identifiers which allow the data to be retrieved by the source/origin healthcare organization. In addition, the S-EHR App should also provide a Request Authorization Token for each DICOM image, which ensures that the citizen has given his/her consent to the RRC in order to retrieve a DICOM image from its corresponding healthcare organization **[D4.9]**.
2. After the S-EHR App has provided the identifiers and the tokens to the RRC, the RRC requests the anonymized versions of the DICOM images – the anonymized versions of the structured content of the DICOM images – from the healthcare organization.
3. The service which is responsible for the anonymization operation of the structured content of the DICOM images receives a request along with a DICOM image from the RRC.
4. The service deletes or modifies the necessary attributes, placed within the DICOM dataset, as referred to the standard **[NEMA]**, **[ANNEX]** with regard to the DICOM images format, in order to generate the anonymized version of the structured content of the DICOM image.
5. Finally, the service returns the anonymized version of the structured content of the DICOM image to the healthcare organization and by extension to the RRC.

4. CONCLUSIONS AND NEXT STEPS

This deliverable had the aim of specifying the privacy mechanisms which will be used in the Research Scenario when the citizens' data is shared for specific research studies. More specifically, the design of the data anonymization and pseudonymization library was analyzed as well as how this library works. In case of data anonymization, all personal data is modified so that it cannot lead to the identity of the citizen, whereas in case of data pseudonymization all personal data, which is linked with a citizen, is replaced either with a pseudo-identity or with a pseudonym. Both privacy mechanisms are implemented on structured data on the phone, whereas their implementation on unstructured data seems to be very challenging and it is not performed in the S-EHR App. However, for the structured content of the DICOM images, a service will be implemented as well. This service will be used on the side of the healthcare organization in order to anonymize the attribute values placed within the DICOM dataset.

In March 2022, the deliverable D6.13 will be also released as a follow-up to the current deliverable. The deliverable D6.13 is a demonstration report which refers to the implementation and integration of the data anonymization and pseudonymization library, as well as to the service with regard to the structured content of the DICOM images. More specifically, it will present additional information about the aforementioned privacy mechanisms as well as how they are implemented in the context of InteropEHRate for satisfying the goals set in the Research Scenario.

REFERENCES

[Brands 2000] Brands, Stefan A. (2000). Rethinking public key infrastructures and digital certificates. MIT Press. ISBN 978-0-262-02491-4.

[Camenisch 2002] Camenisch, Jan; Lysyanskaya, Anna (2002). "Dynamic Accumulators and Application to Efficient Revocation of Anonymous Credentials". In Yung, Moti (ed.). Advances in Cryptology — CRYPTO 2002. Lecture Notes in Computer Science. 2442. Springer. pp. 101–120. doi:10.1007/3-540-45708-9_5. ISBN 978-3-540-44050-5.

[Chaum 1985] Chaum, David (October 1985). "Security without identification: transaction systems to make big brother obsolete". Communications of the ACM. 28 (10): 1030–1044. CiteSeerX 10.1.1.319.3690. doi:10.1145/4372.4373. S2CID 15340054.

[D2.3] InteropEHRate Consortium. D2.3: User Requirements for cross-border HR integration V3, 2021. www.interopehrate.eu/resources

[D4.8] InteropEHRate Consortium. D4.8: Specification of protocol and APIs for research health data sharing - V1, 2021. www.interopehrate.eu/resources

[D4.9] InteropEHRate Consortium. D4.9: Specification of protocol and APIs for research health data sharing - V2, 2021. www.interopehrate.eu/resources

[D4.10] InteropEHRate Consortium. D4.10: Design of the libraries for health data sharing for research - V1, 2021. www.interopehrate.eu/resources

[DATA PRIVACY MANAGER] Data Privacy Manager. Website: https://dataprivacymanager.net/pseudonymization-according-to-the-gdpr/

[GDPR] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

[GEDIK 2005] Gedik Bugra, Ling Liu. "Location Privacy in Mobile Systems: A Personalized Anonymization Model". 25th IEEE International Conference on Distributed Computing Systems (ICDCS'05) (2005): 620-629.

[i-SCOOP] i-SCOOP. Website: https://www.i-scoop.eu/gdpr/pseudonymization/

[IEEE 1609.2-2013] IEEE Standard for Wireless Access in Vehicular Environments--Security Services for Applications and Management Messages," in IEEE Std 1609.2-2016 (Revision of IEEE Std 1609.2-2013) , vol., no., pp.1-240, 1 March 2016, doi: 10.1109/IEEESTD.2016.7426684. https://standards.ieee.org/standard/1609_2-2016.html

[IMAIOS] IMAIOS. Website: https://www.imaios.com/en/Company/blog/DICOM-Anonymization

[ISO 29100] ISO/IEC 29100:2011 Information technology — Security techniques — Privacy framework

[LEADTOOLS] LEADTOOLS. Website: https://www.leadtools.com/help/sdk/v21/dicom/api/overview-value-representations.html

[MALEKZADEH 2019] Malekzadeh Mohammad, Clegg, Richard, Cavallaro Andrea, Haddadi Hamed. "IoTDI '19: Proceedings of the International Conference on Internet of Things Design and Implementation." ACM Digital Library (2019): 49-58.

[MANO 2010] Mano Masanori, Ishikawa Yoshiharu. " Anonymizing user location and profile information for privacy-aware mobile services ". ResearchGate (2010): 68-75.

[NEMA] DICOM NEMA. Website: http://dicom.nema.org/dicom/2013/output/chtml/part15/chapter_E.html

[PIWIK PRO] Piwik PRO. Website: https://piwik.pro/blog/benefits-data-pseudonymization-anonymization-gdpr/

[TECHTARGET] TechTarget. Website: https://searchhealthit.techtarget.com/definition/DICOM-Digital-Imaging-and-Communications-in-Medicine

[THE SSL STORE] The SSL Store. Website: https://www.thesslstore.com/blog/difference-sha-1-sha-2-sha-256-hash-algorithms/

ANNEX

DICOM Structured Content Anonymization

The service which was referred in section 3.4.2.2 Anonymization of Structured content in DICOM studies aims to anonymize the structured content of the DICOM images as formerly described. More specifically, several attributes will be removed or replaced with random values in order to de-identify the citizen, to whom the image belongs. The attributes which should get modified in order to achieve the anonymization process are presented in the Table below (Table 11). The first column refers to the attribute name, the second column refers to the tag, whereas the third one refers to the basic profile **[NEMA]**. The profiles essentially provide instructions on how the data should get modified in order for a DICOM image to get anonymized **[IMAIOS]**.

There are four types of data elements. These types are as follows **[IMAIOS]**:

- Type 1: The attribute is required and must have a valid value.
- Type 2: The attribute is required, but can contain either the value “unknown” or a zero-length value.
- Type 3: The attribute is optional. This means that the attribute may be included, or not, and can contain a zero-length value.
- Type 1C: The attribute is conditional. This means that if a condition is met, the attribute is of Type 1; otherwise, the tag is not sent.
- Type 2C : The attribute is conditional. This means that if a condition is met, the attribute is of Type 2; otherwise, the tag is not sent.

The action codes, contained in Table 11 below, are the following **[NEMA]**.

- D: The attribute value should be replaced with a non-zero length value. This value should be consistent with the Value Representation (VR). *The Value Representation of a Data Element describes the data type and format of the values of the Data Element **[LEADTOOLS]**.*
- Z: The attribute value should be replaced either with a zero-length value or with a non-zero length value consistent with the VR.
- X: The attribute should be removed.
- K: The attribute value can be either remained unchanged for non-sequence attributes or cleaned for sequences.
- C: The attribute value should be replaced with values consistent with the VR and of similar meaning with the original, without containing identifiable information.
- U: The attribute value should be replaced with a non-zero length UID. This UID should be internally consistent within a set of Instances.
- Z/D: The action code Z is applied (Type 2) unless D is required to maintain conformance with the data object (Type 1).
- X/Z: The action code X is applied (Type 3) unless Z is required to maintain conformance with the data object (Type 2).

- X/D - The action code X is applied (Type 3) unless D is required to maintain conformance with the data object (Type 1).
- X/Z/D: The action code X is applied (Type 3) unless action code Z (Type 2) or D (Type 1) is required to maintain conformance with the data object.
- X/Z/U* - The action code X is applied (Type 3) unless action code Z (Type 2) or replacement of contained instance UIDs (U - Type 1 sequences containing UID references) is required to maintain conformance with the data object.

Attribute Name	Tag	Basic Profile
Accession Number	(0008,0050)	Z
Acquisition Comments	(0018,4000)	X
Acquisition Context Sequence	(0040,0555)	X
Acquisition Date	(0008,0022)	X/Z
Acquisition DateTime	(0008,002A)	X/D
Acquisition Device Processing Description	(0018,1400)	X/D
Acquisition Protocol Description	(0018,9424)	X
Acquisition Time	(0008,0032)	X/Z
Actual Human Performers Sequence	(0040,4035)	X
Additional Patient's History	(0010,21B0)	X
Admission ID	(0038,0010)	X
Admitting Date	(0038,0020)	X
Admitting Diagnoses Code Sequence	(0008,1084)	X

Admitting Diagnoses Description	(0008,1080)	X
Admitting Time	(0038,0021)	X
Affected SOP Instance UID	(0000,1000)	X
Allergies	(0010,2110)	X
Arbitrary	(4000,0010)	X
Author Observer Sequence	(0040,A078)	X
Branch of Service	(0010,1081)	X
Cassette ID	(0018,1007)	X
Comments on the Performed Procedure Step	(0040,0280)	X
Concatenation UID	(0020,9161)	U
Confidentiality Constraint on Patient Data Description	(0040,3001)	X
Content Creator's Name	(0070,0084)	Z
Content Creator's Identification Code Sequence	(0070,0086)	X
Content Date	(0008,0023)	Z/D
Content Sequence	(0040,A730)	X
Content Time	(0008,0033)	Z/D
Context Group Extension Creator UID	(0008,010D)	U

Contrast Bolus Agent	(0018,0010)	Z/D
Contribution Description	(0018,A003)	X
Country of Residence	(0010,2150)	X
Creator Version UID	(0008,9123)	U
Current Patient Location	(0038,0300)	X
Curve Data	(50xx,xxxx)	X
Curve Date	(0008,0025)	X
Curve Time	(0008,0035)	X
Custodial Organization Sequence	(0040,A07C)	X
Data Set Trailing Padding	(FFFC,FFFC)	X
Derivation Description	(0008,2111)	X
Detector ID	(0018,700A)	X/D
Device Serial Number	(0018,1000)	X/Z/D
Device UID	(0018,1002)	U
Digital Signature UID	(0400,0100)	X
Digital Signatures Sequence	(FFFA,FFFA)	X
Dimension Organization UID	(0020,9164)	U

Discharge Diagnosis Description	(0038,0040)	X
Distribution Address	(4008,011A)	X
Distribution Name	(4008,0119)	X
Dose Reference UID	(300A,0013)	U
Ethnic Group	(0010,2160)	X
Failed SOP Instance UID List	(0008,0058)	U
Fiducial UID	(0070,031A)	U
Filler Order Number / Imaging Service Request	(0040,2017)	Z
Frame Comments	(0020,9158)	X
Frame of Reference UID	(0020,0052)	U
Gantry ID	(0018,1008)	X
Generator ID	(0018,1005)	X
Graphic Annotation Sequence	(0070,0001)	D
Human Performers Name	(0040,4037)	X
Human Performers Organization	(0040,4036)	X
Icon Image Sequence(see Note 12)	(0088,0200)	X
Identifying Comments	(0008,4000)	X

Image Comments	(0020,4000)	X
Image Presentation Comments	(0028,4000)	X
Imaging Service Request Comments	(0040,2400)	X
Impressions	(4008,0300)	X
Instance Creator UID	(0008,0014)	U
Institution Address	(0008,0081)	X
Institution Code Sequence	(0008,0082)	X/Z/D
Institution Name	(0008,0080)	X/Z/D
Institutional Department Name	(0008,1040)	X
Insurance Plan Identification	(0010,1050)	X
Intended Recipients of Results Identification Sequence	(0040,1011)	X
Interpretation Approver Sequence	(4008,0111)	X
Interpretation Author	(4008,010C)	X
Interpretation Diagnosis Description	(4008,0115)	X
Interpretation ID Issuer	(4008,0202)	X
Interpretation Recorder	(4008,0102)	X
Interpretation Text	(4008,010B)	X

Interpretation Transcriber	(4008,010A)	X
Irradiation Event UID	(0008,3010)	U
Issuer of Admission ID	(0038,0011)	X
Issuer of Patient ID	(0010,0021)	X
Issuer of Service Episode ID	(0038,0061)	X
Large Palette Color Lookup Table UID	(0028,1214)	U
Last Menstrual Date	(0010,21D0)	X
MAC	(0400,0404)	X
Media Storage SOP Instance UID	(0002,0003)	U
Medical Alerts	(0010,2000)	X
Medical Record Locator	(0010,1090)	X
Military Rank	(0010,1080)	X
Modified Attributes Sequence	(0400,0550)	X
Modified Image Description	(0020,3406)	X
Modifying Device ID	(0020,3401)	X
Modifying Device Manufacturer	(0020,3404)	X
Name of Physician(s) Reading Study	(0008,1060)	X

Names of Intended Recipient of Results	(0040,1010)	X
Occupation	(0010,2180)	X
Operators' Identification Sequence	(0008,1072)	X/D
Operators' Name	(0008,1070)	X/Z/D
Original Attributes Sequence	(0400,0561)	X
Order Callback Phone Number	(0040,2010)	X
Order Entered By	(0040,2008)	X
Order Enterer Location	(0040,2009)	X
Other Patient IDs	(0010,1000)	X
Other Patient IDs Sequence	(0010,1002)	X
Other Patient Names	(0010,1001)	X
Overlay Comments	(60xx,4000)	X
Overlay Data	(60xx,3000)	X
Overlay Date	(0008,0024)	X
Overlay Time	(0008,0034)	X
Palette Color Lookup Table UID	(0028,1199)	U
Participant Sequence	(0040,A07A)	X

Patient Address	(0010,1040)	X
Patient Comments	(0010,4000)	X
Patient ID	(0010,0020)	Z
Patient Sex Neutered	(0010,2203)	X/Z
Patient State	(0038,0500)	X
Patient Transport Arrangements	(0040,1004)	X
Patient's Age	(0010,1010)	X
Patient's Birth Date	(0010,0030)	Z
Patient's Birth Name	(0010,1005)	X
Patient's Birth Time	(0010,0032)	X
Patient's Institution Residence	(0038,0400)	X
Patient's Insurance Plan Code Sequence	(0010,0050)	X
Patient's Mother's Birth Name	(0010,1060)	X
Patient's Name	(0010,0010)	Z
Patient's Primary Language Code Sequence	(0010,0101)	X
Patient's Primary Language Modifier Code Sequence	(0010,0102)	X
Patient's Religious Preference	(0010,21F0)	X

Patient's Sex	(0010,0040)	Z
Patient's Size	(0010,1020)	X
Patient's Telephone Numbers	(0010,2154)	X
Patient's Weight	(0010,1030)	X
Performed Location	(0040,0243)	X
Performed Procedure Step Description	(0040,0254)	X
Performed Procedure Step End Date	(0040,0250)	X
Performed Procedure Step End Time	(0040,0251)	X
Performed Procedure Step ID	(0040,0253)	X
Performed Procedure Step Start Date	(0040,0244)	X
Performed Procedure Step Start Time	(0040,0245)	X
Performed Station AE Title	(0040,0241)	X
Performed Station Geographic Location Code Sequence	(0040,4030)	X
Performed Station Name	(0040,0242)	X
Performed Station Name Code Sequence	(0040, 4028)	X
Performing Physician Identification Sequence	(0008,1052)	X
Performing Physicians' Name	(0008,1050)	X

Person Address	(0040,1102)	X
Person Identification Code Sequence	(0040,1101)	D
Person Name	(0040,A123)	D
Person Telephone Numbers	(0040,1103)	X
Physician Approving Interpretation	(4008,0114)	X
Physician(s) Reading Study Identification Sequence	(0008,1062)	X
Physician(s) of Record	(0008,1048)	X
Physician(s) of Record Identification Sequence	(0008,1049)	X
Placer Order Number / Imaging Service Request	(0040,2016)	Z
Plate ID	(0018,1004)	X
Pre-Medication	(0040,0012)	X
Pregnancy Status	(0010,21C0)	X
Protocol Name	(0018,1030)	X/D
Reason for the Imaging Service Request	(0040,2001)	X
Reason for Study	(0032,1030)	X
Referenced Digital Signature Sequence	(0400,0402)	X
Referenced Frame of Reference UID	(3006,0024)	U

Referenced General Purpose Scheduled Procedure Step Transaction UID	(0040,4023)	U
Referenced Image Sequence	(0008,1140)	X/Z/U*
Referenced Patient Alias Sequence	(0038, 0004)	X
Referenced Patient Sequence	(0008,1120)	X
Referenced Performed Procedure Step Sequence	(0008,1111)	X/Z/D
Referenced SOP Instance MAC Sequence	(0400,0403)	X
Referenced SOP Instance UID	(0008,1155)	U
Referenced SOP Instance UID in File	(0004,1511)	U
Referenced Study Sequence	(0008,1110)	X/Z
Referring Physician's Address	(0008,0092)	X
Referring Physician's Identification Sequence	(0008,0096)	X
Referring Physician's Name	(0008,0090)	Z
Referring Physician's Telephone Numbers	(0008,0094)	X
Region of Residence	(0010,2152)	X
Related Frame of Reference UID	(3006,00C2)	U
Request Attributes Sequence	(0040,0275)	X
Requested Contrast Agent	(0032,1070)	X

Requested Procedure Comments	(0040,1400)	X
Requested Procedure Description	(0032,1060)	X/Z
Requested Procedure ID	(0040,1001)	X
Requested Procedure Location	(0040,1005)	X
Requested SOP Instance UID	(0000,1001)	U
Requesting Physician	(0032,1032)	X
Requesting Service	(0032,1033)	X
Responsible Organization	(0010,2299)	X
Responsible Person	(0010,2297)	X
Results Comments	(4008,4000)	X
Results Distribution List Sequence	(4008,0118)	X
Results ID Issuer	(4008,0042)	X
Reviewer Name	(300E,0008)	X/Z
Scheduled Human Performers Sequence	(0040,4034)	X
Scheduled Patient Institution Residence	(0038,001E)	X
Scheduled Performing Physician Identification Sequence	(0040,000B)	X
Scheduled Performing Physician Name	(0040,0006)	X

Scheduled Procedure Step End Date	(0040,0004)	X
Scheduled Procedure Step End Time	(0040,0005)	X
Scheduled Procedure Step Description	(0040,0007)	X
Scheduled Procedure Step Location	(0040,0011)	X
Scheduled Procedure Step Start Date	(0040,0002)	X
Scheduled Procedure Step Start Time	(0040,0003)	X
Scheduled Station AE Title	(0040,0001)	X
Scheduled Station Geographic Location Code Sequence	(0040,4027)	X
Scheduled Station Name	(0040,0010)	X
Scheduled Station Name Code Sequence	(0040,4025)	X
Scheduled Study Location	(0032,1020)	X
Scheduled Study Location AE Title	(0032,1021)	X
Series Date	(0008,0021)	X/D
Series Description	(0008,103E)	X
Series Instance UID	(0020,000E)	U
Series Time	(0008,0031)	X/D
Service Episode Description	(0038,0062)	X

Service Episode ID	(0038,0060)	X
Smoking Status	(0010,21A0)	X
SOP Instance UID	(0008,0018)	U
Source Image Sequence	(0008,2112)	X/Z/U*
Special Needs	(0038,0050)	X
Station Name	(0008,1010)	X/Z/D
Storage Media File-set UID	(0088,0140)	U
Study Comments	(0032,4000)	X
Study Date	(0008,0020)	Z
Study Description	(0008,1030)	X
Study ID	(0020,0010)	Z
Study ID Issuer	(0032,0012)	X
Study Instance UID	(0020,000D)	U
Study Time	(0008,0030)	Z
Synchronization Frame of Reference UID	(0020,0200)	U
Template Extension Creator UID	(0040,DB0D)	U
Template Extension Organization UID	(0040,DB0C)	U

Text Comments	(4000,4000)	X
Text String	(2030,0020)	X
Timezone Offset From UTC	(0008,0201)	X
Topic Author	(0088,0910)	X
Topic Keywords	(0088,0912)	X
Topic Subject	(0088,0906)	X
Topic Title	(0088,0904)	X
Transaction UID	(0008,1195)	U
UID	(0040,A124)	U
Verifying Observer Identification Code Sequence	(0040,A088)	Z
Verifying Observer Name	(0040,A075)	D
Verifying Observer Sequence	(0040,A073)	D
Verifying Organization	(0040,A027)	X
Visit Comments	(0038,4000)	X

Table 11 - Attributes of DICOM images which should get modified