APPENDIX Service for the Pseudonymization of Electronic Healthcare Records Based on ISO/EN 13606 for the Secondary Use of Information

Roberto Somolinos, Adolfo Muñoz, M Elena Hernando, Mario Pascual, Jesús Cáceres, Ricardo Sánchez-de-Madariaga, Juan A Fragua, Pablo Serrano, Carlos H Salvador

IEEE J Biomed Health Inform. 2015 Nov;19(6):1937-44.

APPENDIX

Pseudonymization is a complex process that is made up of several possible phases and tasks that are carried out or not depending on the content of the extract to be pseudonymized, the information stored previously in the demographic server and the degrees of specification of the quasi-identifiers. The working of a pseudonymizing system in the pseudonymization process in 6 different cases is detailed below. Initially the associated demographic server contains the data of three demographic entities stored in it. Each of these entities may be referenced by one or two type *II* identifiers defined by their *root* and *extension* fields. The following table shows the initial state of the demographic server:

				Id1	Id1	Id2	Id2
Name	Surname	Birthday	ZIP	Root	Ext	Root	Ext
Jane	Doe	01/01/1911	01234	HUPH	d0123	ISCIII	123456
Paula	Poe	02/02/1922	77777	HUPH	p0342	ISCIII	547002
John	Smith	03/03/1933	33333	HUPH	t2121		

In each of the examples, the aim is to pseudonymize an extract that contains the demographic information of a single entity. The following table shows the data of the demographic entities that appear in each of the examples:

					Id1	Id1	Id2	Id2
Example	Name	Surname	Birthday	ZIP	Root	Ext	Root	Ext
1	Richard	Roe	04/04/1944	45678	HUPH	g5404		
2	Jane	Doe	01/01/1911	01234	HUPH	d0123		
3	Paula	Poe	02/02/1922	77777	BIOING	fdf894	HUPH	p0342
4	John	Smith	03/03/1933	33333	HUPH	t2121	CEPA	wert894
5	Harry	Hoe	05/05/1955	55555	GBT	010207		
6	Richard	Roe	04/04/1944	45678	HUPH	g5404		

Example 1. Extract with non-previously registered SoC

The extract sent to be pseudonymized contains information of a demographic entity not registered in the demographic server. This entity is *subject_of_care* and is represented in the extract by means of an identifier with a *root*/*extension* pair (”HUPH”/”g5404” in this case). The patient’s name is “Richard Roe” as can be seen in the table. The extract contains the demographic data of this entity in its *demographic_extract* field:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>g5404</extension>
                <root>
                        <oid>HUPH</oid>
                </root>
        </subject_of_care>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <id>
                        <extension>g5404</extension>
```

```
                <root>
                    <oid>HUPH</oid>
                </root>
            </id>
            <name>
                <name_part>
                    <entity_part_name>Richard</entity_part_name>
                    <name_part_qualifier>
                        <codeValue>BR</codeValue>
                    </name_part_qualifier>
                    <name_part_type>
                        <codeValue>GIV</codeValue>
                    </name_part_type>
                </name_part>
                <name_part>
                    <entity_part_name>Roe</entity_part_name>
                    <name_part_qualifier>
                        <codeValue>BR</codeValue>
                    </name_part_qualifier>
                    <name_part_type>
                        <codeValue>FAM</codeValue>
                    </name_part_type>
                </name_part>
            </name>
            <addr>
                <addr_part>
                    <address_line>45678</address_line>
                    <address_line_type>
                        <codeValue>ZIP</codeValue>
                    </address_line_type>
                </addr_part>
            </addr>
            <administrative_gender_code>
                <codeValue>male</codeValue>
            </administrative_gender_code>
            <birth_time>
                <time>1944-04-04T00:00:00</time>
            </birth_time>
        </demographic_extract>
</EHR_EXTRACT>
```

To initiate the pseudonymization process the *anonymizeExtract* function of the pseudonymizer module is invoked by sending the extract to be pseudonymized, “RSC” chain as *rootProject* and the degree of specification selected for the quasi- identifiers (gender: included, time of birth: day, place of residence: removed) as arguments. That is, it is wished to pseudonymize the extract including the gender and the complete date of birth and that the *root* field of all of the new identifiers that appear in the pseudonymized extract are the same as “RSC”.

Step 1: Storage of the demographic information included in the extract

The first step in the pseudonymization process is to examine all of the objects of the *IDENTIFIED_ENTITY* type stored in the *demographic_extract* field of the extract and send them for their registry and storage in the demographic server if necessary. In this case only one object appears, that corresponds to the subject of care of the extract. All of the identifiers of the said entity are searched for and only one is found (”HUPH”/”g5404”). The *existII* function of the demographic server is checked to see if there is any entity with the said identifier

already stored in the server. The reply received from the demographic server is negative, which is why this demographic entity is stored in the demographic server by means of the *registerIdentifiedEntity* function.

Step 2: Substitution of the identifiers of the entities of the extract

The following step is to search the extract for all of the identifiers susceptible to being linked to demographic data, and therefore must be substituted. These identifiers appear in the *subject_of_care* (*EHR_EXTRACT* class), *party* (*RELATED_PARTY* class) and *performer* (*FUNCTIONAL_ROLE* class) fields. In this first case, only one substitution of identifiers has to be made, specifically that corresponding to the *subject_of_care* field. To carry out each of the substitutions the private *anonymizeII* function of the pseudonymized module is called. This function receives the old identifier as input arguments and the value of the *rootProject*, which in this case is "RSC", and it returns a new type *II* identifier whose value of the *root* field is that indicated by *rootProject* ("RSC") and which represents the same demographic entity as the old identifier. Finally, it is only necessary to assign the new identifier returned by the function to the *subject_of_care* field of the extract.

The *anonymizeII* method searches, by means of the *existII* function if there is any entity with the said identifier stored in the associated demographic server. As in this case a positive response is received, given that in the previous step the demographic entity of the *demographic_extract* field had been stored, the demographic server is consulted by means of the *equivalentExtension* function to see if there is any identifier *II* stored with a *root* value equal to that indicated in the *rootProject* ("RSC") and which refers to the same demographic entity as the identifier used. In this case the response of the server is negative, therefore a new identifier *II* is generated whose *root* value is the same as the *rootProject* ("RSC") and with a value of the *extension* field assigned in such a way as to guarantee its unicity and that there are no duplications. Before the *anonymizeII* function returns the identifier generated as a response, the list of identifiers that point at the demographic entities in the demographic server is updated, since it may associate the new identifier generated with its corresponding entity. This action is carried out through the *updateSetId* function of the demographic server.

Step 3: Suppression of the demographic information included in the extract

The next step is the suppression of the demographic data that appear in the extract. The data relative to the subject of care related to the quasi-identifiers are handled in the following way:

- Gender: this data is searched for and included in the pseudonymized extract
- Time of birth: this data is searched for and included in the pseudonymized extract
- Place of residence: no data relative to the place of residence is included

Step 4: Removal of key data in free-text fields

The search and substitution mechanism has not found any key word that has to be handled in this extract.

The extract resulting of the pseudonymization is as follows:

<EHR_EXTRACT xmlns="CEN/13606/RM">

```
        <subject_of_care>
                <extension>ANON_SERV_RSC:0000000001</extension>
                <root>
                        <oid>RSC</oid>
                </root>
        </subject_of_care>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <administrative_gender_code>
                        <codeValue>male</codeValue>
                </administrative_gender_code>
                <birth_time>
                        <time>1944-04-04T00:00:00</time>
                </birth_time>
        </demographic_extract>
</EHR_EXTRACT>
```

The state, in the demographic server, of the new registered demographic entity is as follows:

Name:	Richard
Surname:	Roe
Date of birth:	04/04/1944
ZIP:	45678
Root (Id1):	HUPH
Extension(Id1):	g5404
Root (Id2):	RSC
Extension(Id2):	ANON_SERV_RSC:0000000001

Example 2. Extract with previously registered SoC

The entire pseudonymization process has been widely described in the previous example. In the following examples the changes in the processes followed are highlighted in respect to the first example and only the new procedures will be described.

The extract to be pseudonymized in the second case contains information on an entity already registered in the demographic server. This entity is the subject of attention of the extract and is referenced from the extract through the *subject_of_care* field, whose value is ”HUPH”/”d0123”. The name of this entity is “Jane Doe” and her data is in the *demographic_extract* field of the extract. The pseudonymization is initiated through a call to the *anonymizeExtract* function with the *rootProject* input argument with a “RSC” value and the following values for the quasi-identifiers gender: removed, time of birth: year and place of residence: all included. The extract to be pseudonymized in this example is as follows:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>d0123</extension>
                <root>
                        <oid>HUPH</oid>
                </root>
        </subject_of_care>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
```

```
            <id>
                <extension>d0123</extension>
                <root>
                    <oid>HUPH</oid>
                </root>
            </id>
            <name>
                <name_part>
                    <entity_part_name>Jane</entity_part_name>
                    <name_part_qualifier>
                        <codeValue>BR</codeValue>
                    </name_part_qualifier>
                    <name_part_type>
                        <codeValue>GIV</codeValue>
                    </name_part_type>
                </name_part>
                <name_part>
                    <entity_part_name>Doe</entity_part_name>
                    <name_part_qualifier>
                        <codeValue>BR</codeValue>
                    </name_part_qualifier>
                    <name_part_type>
                        <codeValue>FAM</codeValue>
                    </name_part_type>
                </name_part>
            </name>
            <addr>
                <addr_part>
                    <address_line>01234</address_line>
                    <address_line_type>
                        <codeValue>ZIP</codeValue>
                    </address_line_type>
                </addr_part>
            </addr>
            <administrative_gender_code>
                <codeValue>female</codeValue>
            </administrative_gender_code>
            <birth_time>
                <time>1911-01-01T00:00:00</time>
            </birth_time>
        </demographic_extract>
</EHR_EXTRACT>
```

Step 1: Storage of the demographic information included in the extract

In this case, there is also a single entity within the *demographic_extract* field of the extract. The said entity has a single identifier (”HUPH”/”d0123”). The *existII* function of the demographic server is checked to see if there is any entity with the said identifier already stored in the server. In this case, the response received from the demographic server is affirmative, since this entity is already previously stored. For this reason the demographic entity is no longer sent for its storage in the server, but it is checked to see if there are other identifiers that refer to this entity in order to send them to the demographic server and update them. As this entity appears with a single identifier, in this case it is not necessary to make any additional call to the demographic server.

Step 2: Substitution of the identifiers of the entities of the extract

In this extract only the identifier of the *subject_of_care* field is substituted, just as in the previous case. The *anonymizeII* method works by following the same work flow as shown in the first example.

Step 3: Suppression of the demographic information included in the extract

All of the demographic information in the original extract is eliminated. The demographic data of the subject of care is recovered and checked to see whether it has to be added to the pseudonymized extract or not according to the degrees specified for each quasi-identifier:

- Gender: this datum is not incorporated into the new extract
- Time of birth: this datum is searched for and the more specific information on the year (month and day) is eliminated allowing it to be included in the pseudonymized extract in the following way: 1911-00-00T00:00:00
- Place or residence: all of the data relative to the place of residence is included. In this case, only the zip code appears, which is added to pseudonymized extract

The already pseudonymized extract is as follows:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>ANON_SERV_RSC:0000000002</extension>
                <root>
                        <oid>RSC</oid>
                </root>
        </subject_of_care>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <addr>
                        <addr_part>
                                <address_line>01234</address_line>
                                <address_line_type>
                                        <codeValue>ZIP</codeValue>
                                </address_line_type>
                        </addr_part>
                </addr>
                <birth_time>
                        <time>1911-00-00T00:00:00</time>
                </birth_time>
        </demographic_extract>
</EHR_EXTRACT>
```

The demographic entity of this extract remains registered in the demographic server in the following way:

Name:	Jane
Surname:	Doe
Date of birth:	01/01/1911
ZIP:	01234
Root (Id1):	HUPH
Extension(Id1):	d0123
Root (Id2):	ISCIII

Extension(Id2): 123456
Root (Id3): RSC
Extension(Id3): ANON_SERV_RSC:0000000002

Example 3. Extract with already registered SoC with an identifier coincident with the *rootProject*

The third extract contains a demographic entity that is referenced from the *subject_of_care* field of the extract, whose content is "BIOING"/"fdf894". This entity is registered in the demographic server, but not in the "BIOING"/"fdf894" identifier. Just as in the previous cases, the extract contains the patient data in its *demographic_extract* field. On this occasion, the pseudonymization in implemented with "ISCIII" as a value of the *rootProject* argument and the values for the degrees of the quasi-identifiers are gender: included, time of birth: groups of 10 years and place of residence: removed. The extract of this example is shown below:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>fdf894</extension>
                <root>
                        <oid>BIOING</oid>
                </root>
        </subject_of_care>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <id>
                        <extension>fdf894</extension>
                        <root>
                                <oid>BIOING</oid>
                        </root>
                </id>
                <id>
                        <extension>p0342</extension>
                        <root>
                                <oid>HUPH</oid>
                        </root>
                </id>
                <name>
                        <name_part>
                                <entity_part_name>Paula</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>GIV</codeValue>
                                </name_part_type>
                        </name_part>
                        <name_part>
                                <entity_part_name>Poe</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>FAM</codeValue>
                                </name_part_type>
                        </name_part>
```

```
                </name>
                <addr>
                        <addr_part>
                                <address_line>77777</address_line>
                                <address_line_type>
                                        <codeValue>ZIP</codeValue>
                                </address_line_type>
                        </addr_part>
                </addr>
                <administrative_gender_code>
                        <codeValue>female</codeValue>
                </administrative_gender_code>
                <birth_time>
                        <time>1922-02-02T00:00:00</time>
                </birth_time>
        </demographic_extract>
</EHR_EXTRACT>
```

Step 1: Storage of the demographic information included in the extract

In the process of sending the entities to the demographic server it is detected that there is a single entity in the *demographic_extract* field. The said entity has two associated type *II* identifiers. Both identifiers are searched to see if they are already registered in the demographic server through the *existII* function. The server replies that one of them is already registered ("HUPH"/"p0342"), while the other one is not ("BIOING"/"fdf894"). For this reason the demographic entity is not sent to be registered, but it is necessary to update the list of identifiers that refer to this entity in the demographic server with the ("BIOING"/"fdf894") identifier. This action is carried out through a call to the *updateSetId* function of the demographic server.

Step 2: Substitution of the identifiers of the entites of the extract

Just as in the other two cases, only the substitution of identifiers in the *subject_of_care* field is carried out in this extract. In this case, the pseudonymization of the identifiers is carried out in respect to the "ISCIII" value of the *rootProject* field. In the first place, the demographic server is searched to see if there is a ("BIOING"/"fdf894") identifier. The response of the server is affirmative, since the list of identifiers in the previous step has been updated. Then it is checked to see if there is any other identifier whose value of the *root* field is "ISCIII" and which points at the same demographic entity as ("BIOING"/"fdf894"). On this occasion, the demographic server finds it and returns the value of the *extension* field of the said identifier which is "547002". With these data, an ("ISCIII"/"547002") identifier is formed, returned and assigned to the *subject_of_care* field of the extract.

Step 3: Suppression of the demographic information included in the extract

After eliminating the demographic information from the extract, the following data, related to its quasi-identifiers and allowed by the degrees of specification, are added:

- Gender: this datum is searched for and incorporated into the pseudonymized extract
- Time of birth: as the *birth_time* field of the demographic package does not allow the inclusion of age ranges, an *ENTRY* object has been created to indicate the range of 10

years in which the date of birth of the subject of care is included in agreement with the archetype shown in figure 6

- Place of residence: no data is included relative to this quasi-identifier

The result of the pseudonymization is shown below:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>547002</extension>
                <root>
                        <oid>ISCIII</oid>
                </root>
        </subject_of_care>
        <all_compositions>
                <name xsi:type="SIMPLE_TEXT"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                        <originalText>Other demographic data</originalText>
                </name>
                <synthesised>false</synthesised>
                <content xsi:type="ENTRY" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
                        <name xsi:type="SIMPLE_TEXT">
                                <originalText>Birthtime range</originalText>
                        </name>
                        <synthesised>false</synthesised>
                        <uncertainty_expressed>false</uncertainty_expressed>
                        <items xsi:type="ELEMENT">
                                <synthesised>false</synthesised>
                                <value xsi:type="IVLTS">
                                        <low>
                                                <time>1920-00-00T00:00:00</time>
                                        </low>
                                        <high>
                                                <time>1929-00-00T00:00:00</time>
                                        </high>
                                </value>
                        </items>
                </content>
        </all_compositions>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <administrative_gender_code>
                        <codeValue>female</codeValue>
                </administrative_gender_code>
        </demographic_extract>
</EHR_EXTRACT>
```

In this example, the entity corresponding to the subject of attention is registered in the demographic server in the following way:

Name:	Paula
Surname:	Poe
Date of birth:	02/02/1922
ZIP:	77777
Root (Id1):	HUPH
Extension(Id1):	p0342
Root (Id2):	ISCIII

Extension(Id2): 547002
Root (Id3): BIOING
Extension(Id3): fdf894

Example 4. Extract with SoC with different identifiers to those already registered

The extract to be pseudonymized contains data on an entity whose information is also collected in the demographic server. This entity is the subject of attention of the extract and is referenced from the *subject_of_care* field with the "HUPH"/"t2121" identifier. The demographic data of the patient is collected in the *demographic_extract* field of the extract. The pseudonymization process is made in respect to the "RSC" value of the *rootProject* field and with the following degrees of specification gender: included, time of birth: removed and place of residence: post or zip code. The xml code of this extract is as follows:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>t2121</extension>
                <root>
                        <oid>HUPH</oid>
                </root>
        </subject_of_care>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <id>
                        <extension>t2121</extension>
                        <root>
                                <oid>HUPH</oid>
                        </root>
                </id>
                <id>
                        <extension>wert894</extension>
                        <root>
                                <oid>CEPA</oid>
                        </root>
                </id>
                <name>
                        <name_part>
                                <entity_part_name>John</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>GIV</codeValue>
                                </name_part_type>
                        </name_part>
                        <name_part>
                                <entity_part_name>Smith</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>FAM</codeValue>
                                </name_part_type>
                        </name_part>
                </name>
```

```
            <addr>
                <addr_part>
                    <address_line>33333</address_line>
                    <address_line_type>
                        <codeValue>ZIP</codeValue>
                    </address_line_type>
                </addr_part>
            </addr>
            <administrative_gender_code>
                <codeValue>male</codeValue>
            </administrative_gender_code>
            <birth_time>
                <time>1933-03-03T00:00:00</time>
            </birth_time>
        </demographic_extract>
</EHR_EXTRACT>
```

Step 1: Storage of the demographic information included in the extract

The registering and updating process of demographic entities is totally analogous to that of example 3, since in both cases there is a single demographic entity with two identifiers, one already known by the demographic server and the other one still not.

Step 2: Substitution of the identifiers of the entities of the extract

As in all of the previous examples, only one substitution of identifiers is carried out in this context, specifically that of the *subject_of_care* field. In the first place, the demographic server is searched to see if there is a ("HUPH"/"t2121") identifier, the response of the server is positive. There is then another search carried out to see if there is any other identifier whose value of the *root* field is "RSC" and which points at the same demographic entity as ("HUPH"/"t2121"). As on this occasion no identifier is found that fulfils these conditions, the pseudonymizer generates a new identifier *II* with "RSC" as the *root* value and with a value of the *extension* field which ensures its unicity. This identifier will be assigned to the *subject_of_care* field of the extract, prior to the updating of the list of identifiers of the demographic entity stored in the demographic server through the *updateSetId* function.

Step 3: Suppression of the demographic information included in the extract

The handling of the demographic data of the subject of care is as follows:

- Gender: it is searched for and included in the pseudonymized extract
- Time of birth: not included in the new extract
- Place of residence: all of the data the same as or more general than the zip code (city, state, country) must be included in the pseudonymized extract, and the more specific (street, number) must be eliminated. This example only contains the zip code which is why it is added to the pseudonymized extract.

The pseudonymized extract resulting from this example is shown below:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
            <extension>ANON_SERV_RSC:0000000003</extension>
            <root>
```

```
                <oid>RSC</oid>
            </root>
        </subject_of_care>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
            <administrative_gender_code>
                <codeValue>male</codeValue>
            </administrative_gender_code>
            <addr>
                <addr_part>
                    <address_line>33333</address_line>
                    <address_line_type>
                        <codeValue>ZIP</codeValue>
                    </address_line_type>
                </addr_part>
            </addr>
        </demographic_extract>
</EHR_EXTRACT>
```

The entity pseudonymized in this example is stored and registered in the demographic server in the following way:

Name:	John
Surname:	Smith
Date of birth:	03/03/1933
ZIP:	33333
Root (Id1):	HUPH
Extension(Id1):	t2121
Root (Id2):	CEPA
Extension(Id2):	wert894
Root (Id3):	RSC
Extension(Id3):	ANON_SERV_RSC:0000000003

Example 5. Extract with non-registered SoC, performers and party

The extract contains only the demographic data of the entity corresponding to the *subject_of_care* field, whose identifier is “GBT”/”010207”. This extract also references other demographic entities from its *performer* and *party* fields, although it does not contain demographic data on it. The pseudonymization process is made in respect to the “RSC” value of the *rootProject* argument and the following values for the quasi-identifiers gender: included, time of birth: month and place of residence: country. The extract to be pseudonymized in this example is as follows:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
            <extension>010207</extension>
            <root>
                <oid>GBT</oid>
            </root>
        </subject_of_care>
        <all_compositions>
```

```
                <synthesised>false</synthesised>
                <composer>
                        <performer>
                                <extension>010208</extension>
                                <root>
                                        <oid>GBT</oid>
                                </root>
                        </performer>
                </composer>
                <content xsi:type="ENTRY" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
                        <synthesised>false</synthesised>
                        <uncertainty_expressed>true</uncertainty_expressed>
                        <subject_of_information>
                                <party>
                                        <extension>010210</extension>
                                        <root>
                                                <oid>GBT</oid>
                                        </root>
                                </party>
                        </subject_of_information>
                        <other_participations>
                                <performer>
                                        <extension>010209</extension>
                                        <root>
                                                <oid>GBT</oid>
                                        </root>
                                </performer>
                        </other_participations>
                </content>
        </all_compositions>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <id>
                        <extension>010207</extension>
                        <root>
                                <oid>GBT</oid>
                        </root>
                </id>
                <name>
                        <name_part>
                                <entity_part_name>Harry</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>GIV</codeValue>
                                </name_part_type>
                        </name_part>
                        <name_part>
                                <entity_part_name>Hoe</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>FAM</codeValue>
                                </name_part_type>
                        </name_part>
                </name>
                <addr>
```

```
                    <addr_part>
                        <address_line>55555</address_line>
                        <address_line_type>
                            <codeValue>ZIP</codeValue>
                        </address_line_type>
                    </addr_part>
                </addr>
                <administrative_gender_code>
                    <codeValue>male</codeValue>
                </administrative_gender_code>
                <birth_time>
                    <time>1955-05-05T00:00:00</time>
                </birth_time>
            </demographic_extract>
</EHR_EXTRACT>
```

Step 1: Storage of the demographic information included in the extract

Since this is an extract on a patient who is not registered in the demographic server, the sending and updating procedure of the demographic entities and the substitution of the identifier ("GBT"/"010207") of the *subject_of_care* field is the same as that of example 1. But this extract, also contains the type *II* identifiers in the clinical part which must be substituted by new identifiers. The ("GBT"/"010208") and ("GBT"/"010209") identifiers refer to *performer* fields and the ("GBT"/"010210") identifier links a *party* field with its corresponding demographic entity.

Step 2: Substitution of the identifiers of the entities of the extract

The substitution of these identifiers is carried out following the same procedure, which is why only the substitution of the ("GBT"/"010208") identifier is described. The *anonymizeII* function is called with the objective of obtaining an identifier equivalent to ("GBT"/"010208") and whose value of the *root* field is the same as the *rootProject* ("RSC"). Initially, it looks to see if there is a ("GBT"/"010208") identifier in the demographic server. The response of the server is negative, since on this occasion there are no demographic data in the server on the entity to which the identifier refers. In spite of this, the identifier must be substituted to ensure a correct pseudonymization. A new identifier is created with a "RSC" value in its *root* field, in the same way as has been done in previous cases, which will be the identifier that substitutes ("GBT"/"010208") in the pseudonymized extract. Before finalising, and although there are no demographic data, it is registered in the demographic server that these two identifiers refer to the same entity. For this reason a new entity is created with no data and both identifiers are added in its *id* field. The said new entity is sent to the demographic server for its registry and storage through the *registerIdentifiedEntity* function.

Step 3: Suppression of the demographic information included in the extract

The demographic information on the subject of care related to its quasi-identifiers is handled in the following way:

- Gender: it is included in the pseudonymized extract

- Time of birth: the datum is searched for and the more precise information on the month is eliminated, that is, the day of birth is eliminated, but the month and year maintained
- Place of residence: all of the data the same as or more general than the country must be included in the in the pseudonymized extract, and the rest must be eliminated. This example only contains the zip code which is why it is must not be included in the pseudonymized extract as it is more specific than the country.

The following result of the pseudonymization of this example is obtained:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>ANON_SERV_RSC:0000000004</extension>
                <root>
                        <oid>RSC</oid>
                </root>
        </subject_of_care>
        <all_compositions>
                <synthesised>false</synthesised>
                <composer>
                        <performer>
                                <extension>ANON_SERV_RSC:0000000005</extension>
                                <root>
                                        <oid>RSC</oid>
                                </root>
                        </performer>
                </composer>
                <content xsi:type="ENTRY" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
                        <synthesised>false</synthesised>
                        <uncertainty_expressed>true</uncertainty_expressed>
                        <subject_of_information>
                                <party>

        <extension>ANON_SERV_RSC:0000000007</extension>
                                        <root>
                                                <oid>RSC</oid>
                                        </root>
                                </party>
                        </subject_of_information>
                        <other_participations>
                                <performer>

        <extension>ANON_SERV_RSC:0000000006</extension>
                                        <root>
                                                <oid>RSC</oid>
                                        </root>
                                </performer>
                        </other_participations>
                </content>
        </all_compositions>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <administrative_gender_code>
                        <codeValue>male</codeValue>
                </administrative_gender_code>
                <birth_time>
                        <time>1955-05-00T00:00:00</time>
                </birth_time>
```

```
        </demographic_extract>
</EHR_EXTRACT>
```

In this example there are several participating entities to be pseudonymized. How each of them is registered in the demographic server is shown below:

Name:	Harry
Surname:	Hoe
Date of birth:	05/05/1955
ZIP:	55555
Root (Id1):	GBT
Extension(Id1):	010207
Root (Id2):	RSC
Extension(Id2):	ANON_SERV_RSC:0000000004

Root (Id1):	GBT
Extension(Id1):	010208
Root (Id2):	RSC
Extension(Id2):	ANON_SERV_RSC:0000000005

Root (Id1):	GBT
Extension(Id1):	010209
Root (Id2):	RSC
Extension(Id2):	ANON_SERV_RSC:0000000006

Root (Id1):	GBT
Extension(Id1):	010210
Root (Id2):	RSC
Extension(Id2):	ANON_SERV_RSC:0000000007

Example 6. Extract with identifier data in other fields

This last example is practically identical to the first one, with the exception that the extract includes a composition with a field text. This field is *name*, whose usual function is to store the name of the composition; however in this example it includes data with which demographic information on the entities participating in the extract can be obtained. Specifically it includes the value of the *extension* field of the subject of attention. The degrees of specification with which the pseudonymization has been carried out are gender: removed, time of birth: groups of five years and place of residence: removed. The pseudonymization took place under *rootProject* "RSC". The code of this extract is shown below:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>g5404</extension>
                <root>
                        <oid>HUPH</oid>
                </root>
```

```
        </subject_of_care>
        <all_compositions>
                <name xsi:type="SIMPLE_TEXT"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                        <originalText>This patient g5404 has the code g5404</originalText>
                </name>
                <synthesised>false</synthesised>
        </all_compositions>
        <demographic_extract xsi:type="SUBJECT_OF_CARE_PERSON_IDENTIFICATION"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                <id>
                        <extension>g5404</extension>
                        <root>
                                <oid>HUPH</oid>
                        </root>
                </id>
                <name>
                        <name_part>
                                <entity_part_name>Richard</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>GIV</codeValue>
                                </name_part_type>
                        </name_part>
                        <name_part>
                                <entity_part_name>Roe</entity_part_name>
                                <name_part_qualifier>
                                        <codeValue>BR</codeValue>
                                </name_part_qualifier>
                                <name_part_type>
                                        <codeValue>FAM</codeValue>
                                </name_part_type>
                        </name_part>
                </name>
                <addr>
                        <addr_part>
                                <address_line>45678</address_line>
                                <address_line_type>
                                        <codeValue>ZIP</codeValue>
                                </address_line_type>
                        </addr_part>
                </addr>
                <administrative_gender_code>
                        <codeValue>male</codeValue>
                </administrative_gender_code>
                <birth_time>
                        <time>1944-04-04T00:00:00</time>
                </birth_time>
        </demographic_extract>
</EHR_EXTRACT>
```

On implementing this example after that of example 1, a specific identifier has already been assigned to the entity “Richard Roe” whose value of the *root* field is “RSC”, which is why, on carrying out this pseudonymization, the same identifier is detected and assigned, this maintaining the coherence between identifiers within the same project.

Step 3: Suppression of the demographic information included in the extract

The demographic data on the subject of care are handled depending on the degree of each quasi-identifier:

- Gender: this datum is not included in the new extract
- Time of birth: this datum is included by means of a *COMPOSITION* object similar to that of example 3, but in this case with groups of five years
- Place of residence: no data is added relative to the place of residence

Step 4: Removal of key data in free-text fields

Once all of the steps in the pseudonymization are finished, it is checked to see whether the pseudonymized extract includes any relevant data, such as identifiers of the entities participating in the extract, which may permit a link to demographic information. In this case, it is found that within a textual field the value of the *extension* field of one of the identifiers of the original extract appears several times. This is why the said value is substituted by the value of the *extension* field of the identifier used in its place in the pseudonymized extract.

The pseudonymized extract, after the final validation, is shown below:

```
<EHR_EXTRACT xmlns="CEN/13606/RM">
        <subject_of_care>
                <extension>ANON_SERV_RSC:0000000001</extension>
                <root>
                        <oid>RSC</oid>
                </root>
        </subject_of_care>
        <all_compositions>
                <name xsi:type="SIMPLE_TEXT"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                        <originalText>This patient ANON_SERV_RSC:0000000001 has the
code ANON_SERV_RSC:0000000001</originalText>
                </name>
                <synthesised>false</synthesised>
        </all_compositions>
        <all_compositions>
                <name xsi:type="SIMPLE_TEXT"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
                        <originalText>Other demographic data</originalText>
                </name>
                <synthesised>false</synthesised>
                <content xsi:type="ENTRY" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance">
                        <name xsi:type="SIMPLE_TEXT">
                                <originalText>Birthtime range</originalText>
                        </name>
                        <synthesised>false</synthesised>
                        <uncertainty_expressed>false</uncertainty_expressed>
                        <items xsi:type="ELEMENT">
                                <synthesised>false</synthesised>
                                <value xsi:type="IVLTS">
                                        <low>
                                                <time>1940-00-00T00:00:00</time>
                                        </low>
                                        <high>
                                                <time>1944-00-00T00:00:00</time>
                                        </high>
                                </value>
```

```
                    </items>
                </content>
            </all_compositions>
</EHR_EXTRACT>
```

This example does not generate changes in the state of the demographic server, since the participating demographic entity is already registered with all of its data in the server. The state of the server as regards this entity does not vary and is as follows:

Name:	Richard
Surname:	Roe
Date of birth:	04/04/1944
ZIP:	45678
Root (Id1):	HUPH
Extension(Id1):	g5404
Root (Id2):	RSC
Extension(Id2):	ANON_SERV_RSC:0000000001