Use cases of Import and Anonymization

## Abbreviations

**SHS** Shanoir Server

**SU** ShanoirUploader

**---------------------------------------------------------------------------------------------------------------------------------**

**PID** Patient Identifier

**PID-SHS** Patient Identifier in the Shanoir Server database

**Note**: It corresponds to the *Common Name* field on the Shanoir interface and to the *Name* column of the *Subject* table.

**PID-SU** Patient Identifier from ShanoirUploader

**Note**: For all the other cases except OFSEP, it is the value of the input field “*NewPatientIDTF*” and this is transferred in the element *patientID* in the *upload-job.xml* file.

For OFSEP, this field is empty because the input field is not activated.

**---------------------------------------------------------------------------------------------------------------------------------**

**PHK** Patient's Hash Key

**PHK-SHS** Patient's Hash Key in the Shanoir Server database

**Note:** Hash with the length 14: Hash (first name + last name + birth date)

It corresponds to the *SubjectIdentifier* column of the *Subject* table.

**PHK-SHS-SU** Patient's Hash Key from ShanoirUploader (all the other cases except OFSEP)

**Note:** Hash with the length 14: Hash (first name + last name + birth date)

For OFSEP, see **PHK-OFSEP**.

**PHK-OFSEP** Patient's Hash Key for OFSEP

**Note:** Hash of SHA-256 with the length 64 calculated by SU.

Hash (first name + birth name + birth date) : Pseudonymus is not used

**Note**: **PHK-SHS-SU** and **PHK-OFSEP** are transferred in the *upload-job.xml* in the element *subjectIdentifier*.

# Context

This technical specification collects all use cases regarding the data import in Shanoir platform and the anonymization either with ShanoirUploader or with Shanoir platform. It contains the data import either from a DICOM ZIP file or from the ShanoirUploader. It presents especially the features of an import in Shanoir OFSEP platform, which has several differences from the Shanoir Neurinfo platform. It helps to summary the already existing import and anonymization and to list the future developments.

## Import and Anonymization

During the data import process from a ZIP file, a local anonymization will be done when a new subject is created. A subject identifier (PHK) will be automatically generated and saved in the database at the end.

In case of the data import from ShanoirUploader, since an anonymization is already accomplished by the uploader, no more local anonymization will be done during the import in Shanoir. When a new subject is created, the “*Is the subject already anonymized?*” option will not be displayed and internally pre-selected.

For the OFSEP project, the patient’s hash key (PHK) is calculated based on the patient’s first name, birth name and his birth date. It has a length of 64.

*PHK = SHA256(hashP1(first\_name)|| hashP1(birth\_name)|| hashP1(birth\_date))*

*«SHA256 »: SHA\_256 bits,*

*«hashP1 » : Pseudonymus hash with soundex « 0 »,*

*« || »:  concatenation symbol.*

For the other cases, the patient’s hash key has a length of 14 and is calculated with the first name, the last name and the birth date.

*PHK = SHA(first\_name || last\_name || birth\_date)«SHA»: SHA\_algorithm 160 bits \*,*

*« || »:  concatenation symbol.*

*\** <https://www.cs.clemson.edu/course/cpsc424/material/Cryptography/Message%20Digest%20Functions.pdf>

https://fr.wikipedia.org/wiki/SHA-0

There are also some features about the patient identifier (PID) for the OFSEP project. In the other cases except OFSEP, an input field “*NewPatientIDTF”* is present in ShanoirUploader interface to let the user enter a patient’s identifier. This PID-SU is then transferred to the SHS. In case of the OFSEP project, the input field “*NewPatientIDTF*” is disable in the SU, and the PID-SU is calculated automatically in Shanaoir platform based on the Id of the center to which the subject is attached. The PIS is a string of seven characters. The first tree characters represent the Id of the center formatted to a string of three characters. The last four characters are incremental numbers increased when we add a new subject to a study which study card is attached to the given center.

*PID= XXXYYYY*

*«XXX»: Center Id*

*«YYYY»: Incremental number*

Regarding all these features and differences, the following table lists all of the use cases of the import and the anonymization. Each use case is explained in detail below.

|  |  |  |
| --- | --- | --- |
|  | **Import From SU** | **Import From ZIP** |
| **Neurinfo** | |  |  |  | | --- | --- | --- | | **PHK-SU = PHK-SHS** | **PHK-SU ≠ PHK-SHS** | | | Case #1 | **PID-SU = PID-SHS** | **PID-SU ≠ PID-SHS** | | Case #2 | Case #3 |   Case A | Case B |
| **OFSEP** | Case C   |  |  | | --- | --- | | **PHK-SU = PHK-SHS** | **PHK-SU ≠ PHK-SHS** | | Case #1 | Case #2 | | Case D |

* **Case A: Import from ShanoirUploader with the Shanoir Neurinfo platform**

Four cases may be present here:

|  |  |  |
| --- | --- | --- |
|  | **PHK-SU = PHK-SHS** | **PHK-SU ≠ PHK-SHS** |
| **PID-SU = PID-SHS** | Case #1   the patient correspond with the one in the database  | Case #2    Warning   (two different patients?) |
| **PID-SU ≠ PID-SHS** | Case #3   the patient is new  |

1. **Case #1**: PHK-SU = PHK-SHS

The patient could be found within Shanoir database using the PHK. The existing patient (subject) will be preselected on Shanoir or Shanoir uplader interface. The user will not be able to change the subject or create a new subject. No local anonymization in SHS is required as the patient is already correctly anonymized by the SU.

1. **Case #2**: PHK-SU ≠ PHK-SHS and PID-SU = PID-SHS

Since the subject’s PHK does not exist in SHS, t During the subject creation, the system detects that the patient could be found within Shanoir database using the PID, but the PHK is different. We display a warning that the PID is already used: “Are you sure the patient identifier is correct? There exists already one in the database, please change the common name”. So, the user has to change the *Common Name* since the same one exists already for another subject. If the user change other fields like the birth date, another PHK will be calculated and stored in the SHS. In any case an anonymization will have to be done since the common name is changed.

**Case #4**: PID-SU ≠ PID-SHS and PHK-SU ≠ PHK-SHS

The patient could not be found in Shanoir database using the PHK. A new subject creation page will be proposed. The *Birth Date* field will not be editable; The *Common Name* and the *Sex* fields will be editable. The common name introduced by the user is not used for another subject, so the new subject could be created and saved in the database of SHS

* **Case B: Import from ZIP File with the Shanoir Neurinfo platform**

Both the subject selection and the new subject creation tabs will be proposed to the user. Nothing else to develop.

* **Case C: Import from ShanoirUploader with the Shanoir OFSEP platform**

Since the PID-SU for OFSEP is empty and the PHK-OFSEP is already calculated by the SU, only two cases will be present here:

1. **Case #1:** PHK-OPSEP = PHK-SHS: the subject exists already in SHS

The subject will be preselected and the creation tab will be disable.

1. **Case #2:** PHK-OFSEP ≠ PHK-SHS: the subject does not exist before

The subject selection tab will be disabled and a new subject creation page will be open. The *Sex* and the *Birth Date* fields will be prefilled with the elements in the *upload-job.xml* file and are not modifiable. PID-SHS will be calculated automatically as described earlier. The PHK-OFSEP and the 10 hash values will be saved in the database of the SHS.

Both of these two cases need no more local anonymization in SHS as the patient was already correctly anonymized by the SU.

* **Case D: Import from ZIP File with the Shanoir OFSEP platform**

Both the subject selection and the new subject creation tabs will be proposed to the user, as for the Case B. The differences will be present in the creation of the subject’s hash key and in the anonymization: OFSEP requires the birth name, the first name and the birth date to generate the patient’s hash key. During the anonymization, the new hash key for OFSEP will be added into the *PatientID Dicom field*.

On the Shanoir interface, the user choice “*yes*” or ”no” for the “*Is the subject already anonymized?*” will be deleted, as in any case the anonymization will have to be done with the birth name, the first name and the birth date (asked by OFSEP).

On the user interface, there will be a *First Name*, a *Birth Name* and the *Birth Date* fields to be proposed to calculated the PHK.

The 10 hash values have to be created with Pseudonymus. Therefore pseudonymus will have to be integrated into Shanoir server. The 10 hash values and the PHK will have to be calculated and stored during import.

**To-do List:**

* In ShanoirUploader, for OFSEP, the hash 64 (e.g. 111fa0500b66b34d04172c6006bfe1485a5bf63409886416a482e259434e74d3) is actually saved in the *patientID* element in the upload-job.xml file. It may cause confusion. This hash 64 will now be saved in the *subjectIdentifier* element, which stores the patient’s hash key in other cases. The *patientID* element becomes empty because it’s not active for OFSEP. (done)
* In Shanoir, Add a *is.OFSEP* Tag in Shanoir.properties to distinguish between the Shanoir OFSEP project and the Shanoir Neurinfo project. (done)
* Modify the size of the SUBJECT\_IDENTIFIER column of the SUBJECT table from 50 to 64. (SUBJECT\_IDENTIFIER is a String)
* For the import from SU and for OFSEP, since the *PatientIdentifier* element is empty, display the *SubjectIdentifer* element of the upload-job.xml at the place of the Patient Identifier at the beginning of import. (done)
* Case subject exists already in the database

(The Case #1 of the Case A, the Case #1 of the Case C)

* Display the subject and disable the new subject creation tab;
* Disable the local anonymization in the SHS.
* Case creation of a new subject for OFSEP

(The Case #2 of the Case C, ShUp Ofsep)

* Disable the subject selection tab and show the subject creation page;
* Prefill the *Birth Date* and the *Sex* field with the *patientBirthDate* element and the *patientSex* element in the upload-job.xml file. These two fields are not modifiable. Display the common name calculated for this subject. Ask user to select the imaged object category, the language hemispheric dominance, the manual hemispheric dominance and enter a personal comment. Imaged object category is mandatory but there is no obligation to fill the other fields.
* Disable the local anonymization and store the PHK-OFSEP in the database of the SHS;
* Store all 10 other hash values in the upload-job.xml in the database: PseudonymusHashValues table.
* Case creation of a new subject for Neurinfo

(The Case #3 of the Case A, ShUp)

* Disable the subject selection tab and show the subject creation page;
* Preselect internally the “Is the subject already anonymized?” option as “yes” and do not display it;
* Disable the *Birth Date* edition;
* Enable the *Common name* and the *Sex* edition;
* Ask user to define a new common name
* Store the PHK calculated by the SU in the database of the SHS.
* For the Case #2 of the Case A:
* Disable the subject selection tab and show the subject creation page;
  + Preselect internally the “Is the subject already anonymized?” option as “yes” and do not display it;
  + Let the user change the *Common Name*;
* For the Case D:
  + Disable the “Is the subject already anonymized?” choice;
  + Add the *Birth Name* and the *First Name* fields on the interface;
  + Calculate the PHK with the birth name, the first name and the birth date;
  + Store the PHK and anonymize the Dicom files with the PHK;
  + Integrate Pseudonymus in the SHS and calculate the 10 hash values;
  + Store the 10 hash values in the PseudonymusHashValues table.
* At the beginning of the import, check if the is.OFSEP Boolean is true, if the PHK-OFSEP/ the PHK-SU/ the PID-SU exists already in the database of the SHS to determine in which case we are.