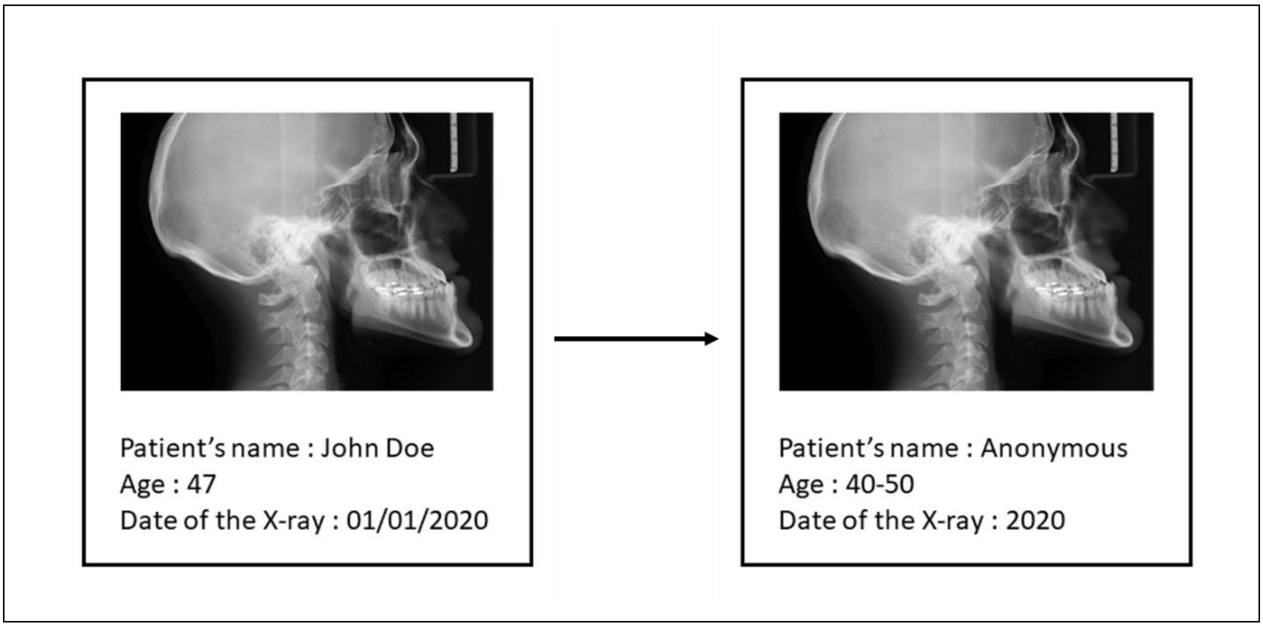
Hes·so VALAIS WALLIS
Haute Ecole de Gestion
Hochschule für Wirtschaft

Filière Informatique de gestion
Studiengang Wirtschaftsinformatik

Travail de Bachelor 2020

Interactive web interface for the configuration of imaging and patient data anonymization



Patient's name : John Doe
Age : 47
Date of the X-ray : 01/01/2020

Patient's name : Anonymous
Age : 40-50
Date of the X-ray : 2020

Etudiant-e : Audrey Viriot

Professeur : Adrien Depeursinge

Travail rendu le : 27/07/2020

HES-SO Valais-Wallis • rue de la Plaine 2 • 3960 Sierre
+41 58 606 89 11 • info@hevs.ch • www.hevs.ch

ABSTRACT

Karnak is a tool with normalization and de-identification capabilities for medical images in DICOM format. It is used in medical research in order to anonymize all information that can be sensitive. Algorithms of normalization and de-identification are under development by the Karnak team, but no user interface exists. Karnak is destined to be in a research pipeline. The purpose of this pipeline is to conduct medical images from hospital to an online DICOM library of anonymized medical images.

For user interface, the Karnak team would like an easy-to-use web application developed in recent technologies, allowing users to create anonymization profiles based on DICOM standard profile. However, before starting the development of a frontend application, the Karnak team would like to verify the feasibility of a tool like this through a prototype. The objective of this thesis is to determine if it is possible to create a web interface allowing the creation of anonymization profiles based on DICOM standard profile.

After studying DICOM standard and the different DICOM anonymization tools existing on the market, a working prototype has been created as prove of concept. The prototype is well based on DICOM standard profile and provides anonymization profiles created by users. This kind of application is feasible. However, if the dataset is larger, performance problems may be encountered. In addition, some improvements must be taken into accounts to guarantee the patient privacy.

Key words: Karnak, user interface, DICOM, anonymization, medical imaging, web application

FOREWORD

This thesis has been realized as part of final work for a bachelor's degree in business information technology at HES-SO Valais-Wallis of Sierre. The work has been done from the 15th May 2020 to the 27th July 2020.

The goal of this thesis is to propose a solution capable of configuring the anonymization of DICOM files used in medical research.

ACKNOWLEDGEMENTS

I would like to express my special thanks to Adrien DEPEURSINGE, and Roger SCHAER for guiding me all along this project. Their advice has been a great help to me.

I would also like to thank Nicolas RODUIT and his team for allowing me to collaborate on this project.

Helen SONN, thanks for correcting my work and your help.

Finally, thanks to my friends and family for their support throughout this bachelor's degree.

TABLE OF CONTENTS

LIST OF FIGURES **1**

LIST OF TABLES **3**

ABBREVIATIONS **4**

INTRODUCTION **5**

1. OVERALL CONTEXT **6**

1.1. General information 6

1.2. Karnak 7

1.3. Thesis objective 8

2. MEDICAL IMAGING **8**

2.1. Definition 8

2.2. Use for research purposes 9

2.3. DICOM format 10

2.4. Anonymization VS de-identification 13

3. EXISTING ANONYMIZATION TOOLS **13**

3.1. Software for healthcare professionals 14

3.2. Independant software 14

3.3. Software included in a research pipeline 16

3.4. Conclusion 17

4. DEFINITION OF USE CASES **17**

4.1. Users 17

4.2. Use cases 17

4.3. Mock-ups 18

5. CHOICE OF TECHNOLOGIES **20**

5.1. Frontend framework 20

5.1.1. Evaluation criteria 20

5.1.2. React 21

5.1.3. Angular 21

5.1.4. Vue.js 21

5.1.5. Conclusion 22

5.2. Main tools and libraries used 22

5.2.1. Material-UI 22

5.2.2. Github 23

5.2.3. Firebase Hosting ... 23

6. PROTOTYPE ... 23

6.1. ARCHITECTURE ... 23

6.2. DATA ... 24

6.2.1. confidentiality_profile_attributes.json ... 25

6.2.2. module_to_attribute.json ... 26

6.2.3. attributes_modules_linked.json ... 26

6.2.4. rules_colors.json ... 27

6.3. INTERFACES AND FUNCTIONNALITIES ... 28

6.3.1. General appearance of the application ... 28

6.3.2. Setting up anonymization profiles with advanced options ... 29

6.3.3. Displaying attributes ... 30

6.3.4. Settings ... 30

6.3.5. Apply a custom rule for an attribute ... 31

6.3.6. Search by module, id or attribute name ... 31

6.3.7. Exportation of a profile ... 31

6.3.8. Importation of an existing profile ... 32

6.3.9. Reset the application ... 33

6.4. DEPLOYMENT ... 33

7. PROJECT MANAGEMENT ... 33

7.1. GENERAL INFORMATION ... 33

7.2. SPRINT 0 ... 34

7.3. SPRINT 1 ... 35

7.4. SPRINT 2 ... 36

7.5. SPRINT 3 ... 37

7.6. SPRINT 4 ... 38

8. CONCLUSION ... 40

8.1. RESULTS ... 40

8.2. RECOMMENDATION FOR FURTHER ACTION ... 40

8.3. LOOKING TO THE FUTURE ... 40

9. DECLARATION ON HONOUR ... 40

10. REFERENCES ... 42

12. ANNEXES ... 43

I. DETAILS OF THE COMPARATIVE STUDY OF FRONTEND FRAMEWORK ... 43

II. JAVASCRIPT FUNCTION TO MERGE TWO JSON FILES TOGETHER ... 43

III. FUNCTION TO DETECT CONFLICT IN ATTRIBUTES........ 43

IV. FUNCTION 'HANDLECHANGE' TO MANAGE BUGS........ 44

V. PRODUCT BACKLOG........ 45

VI. BURN DOWN CHART SPRINT 0........ 45

VII. BURN DOWN CHART SPRINT 1........ 46

VIII. BURN DOWN CHART SPRINT 2........ 46

IX. BURN DOWN CHART SPRINT 3........ 47

X. BURN DOWN CHART SPRINT 4........ 47

LIST OF FIGURES

Figure 1 - Structure of IMAGINE Project 6
Figure 2 - Karnak operation 7
Figure 3 - Workflow of Karnak (OsiriX-Foundation, 2020) 7
Figure 4 - Thesis objective in Karnak workflow (OsiriX-Foundation, 2020) 8
Figure 5 - Workflow for clinical research 9
Figure 6 - Representation of a patient in DICOM world 10
Figure 7 - Format for writing an attribute 11
Figure 8 - Visualization of a DICOM file 11
Figure 9 - How anonymization works 13
Figure 10 - How depersonalization works 13
Figure 11 - Screenshot of DICOM browser interface 16
Figure 12 - Anonymzed DICOM image 16
Figure 13 - Use cases diagram 18
Figure 14 - Mock-up of the home page 19
Figure 15 - Mock-up of the edit page 19
Figure 16 - React logo 20
Figure 17 - Angular logo 20
Figure 18 - View.js logo 20
Figure 19 - Graph of percentage of Stack Overflow questions vs. technology per month 22
Figure 20 - Architecture of the prototype (inputs/outputs) 23
Figure 21 – Example of an architecture with a backend 24
Figure 22 - Structure of an attribute in JSON 25
Figure 23 - Structure of an attributes in JSON with additional properties 26
Figure 24 - Object in module_to_attribute.json 26
Figure 25 - Structure of attributes_modules_linked.json 26
Figure 26 - Content of rules_colors.json 27
Figure 27 - Home page of the application 28
Figure 28 - Edit page of the application 29
Figure 29 - Creation of a new profile 29
Figure 30 - Management of advanced conflicts options 30
Figure 31 - Displaying attributes in treeview 30
Figure 32 - Settings box 30
Figure 33 - Custom attributes 31
Figure 34 - Search bar 31
Figure 35 - Profile export form 32
Figure 36 - Structure of an attribute in exported JSON file 32
Figure 37 - Profile import form 33
Figure 38 - Cross to reset the application 33

Figure 39 - List of tasks for sprint 0 35
Figure 40 - List of tasks for sprint 1 36
Figure 41 - List of tasks for sprint 2 37
Figure 42 - List of tasks for sprint 3 38
Figure 43 - List of tasks for sprint 4 39

LIST OF TABLES

Table 1 - Explanations of the different rules for changing data in DICOM............................ 12

Table 2 - Excerpt of DICOM basic profile 12

Table 3 - Anonymization of DICOM anonymization software 14

Table 4 - Features available in tested software... 15

Table 5 - Selection criteria for frontend technology 21

Table 6 - Comparison of different JSON file sizes 27

Table 7 - Color code used 28

Table 9 - User stories for sprint 0.. 34

Table 10 - User stories for sprint 1 35

Table 11 - User stories for sprint 2 36

Table 12 - User stories for sprint 3 37

Table 13 - User stories for sprint 4 38

ABBREVIATIONS

AI	Artificial Intelligence
CHUV	Centre Hospitalier Universitaire Vaudois
CSS	Cascading Style Sheet
DICOM	Digital Imaging and Communication in Medicine
IODs	Information Object Definitions
JSON	JavaScript Object Notation
KB	Kilobit
MRI	Magnetic Resonance Imaging
NRG	Neuroinformatics Research Group
PACS	Pictures Archiving and Communication System
SPHN	Swiss Personalized Health Network
UI	User Interface
VR	Value Representative
WASC	Web Application Security Consortium

INTRODUCTION

Nowadays, medical imaging enables to collect a lot of data in addition to taking pictures of the human body. These photos and data allow health care professionals to establish a diagnosis to treat patients. However medical imaging is also essential for medical research. In order to realize their studies, researchers access patient images and data. To protect patients from an invasion of their privacy, patient's resources used in medical research must be anonymized. The needs for anonymization can be diverse in medical research.

In order to facilitate the anonymization process of medical images, Karnak (currently under development) is "a DICOM Gateway with normalization and de-identification capabilities" (OsiriX-Foundation, 2020). The final goal is to have a pipeline to send DICOM files from the Picture Archiving and Communication System (PACS) of hospitals to Kheops (an online library of anonymized DICOM). To meet the needs of each project, Karnak must allow users to configure what should be anonymized and what should not be anonymized.

The objective of this work is to propose a web interface for the configuration of anonymization of imaging and patient data anonymization.

Firstly, the general context of this project will be explained. Secondly, medical imaging will be defined and more specifically medical imaging in the context of research. Thirdly, anonymization tools available on the market will be studied then the use cases of the prototype will be defined. Finally, after a comparison of the different technologies to develop the web application, development and project management will be explained.

1. Overall context

1.1. General information

This work is part of a Swiss Personalized Health Network (SPHN) project called "IMAGINE," whose goal is to implement an image-based, personalized medicine infrastructure at the national level. It uses Artificial Intelligence (AI) to get the best of modern radiological and molecular imaging to predict optimal treatment options and prognosis based on personal patient characteristics. The structure of the project is explained in the figure 1 below. IMAGINE is composed of three parts:

- Standardization: after the collection of data by medical imaging, data are curated and stored, ready to use for the research.
- Analysis: different methods are used to analyse medical images.
- Integration: results of analysis are integrated to existing tool and usable by all researchers in Switzerland.

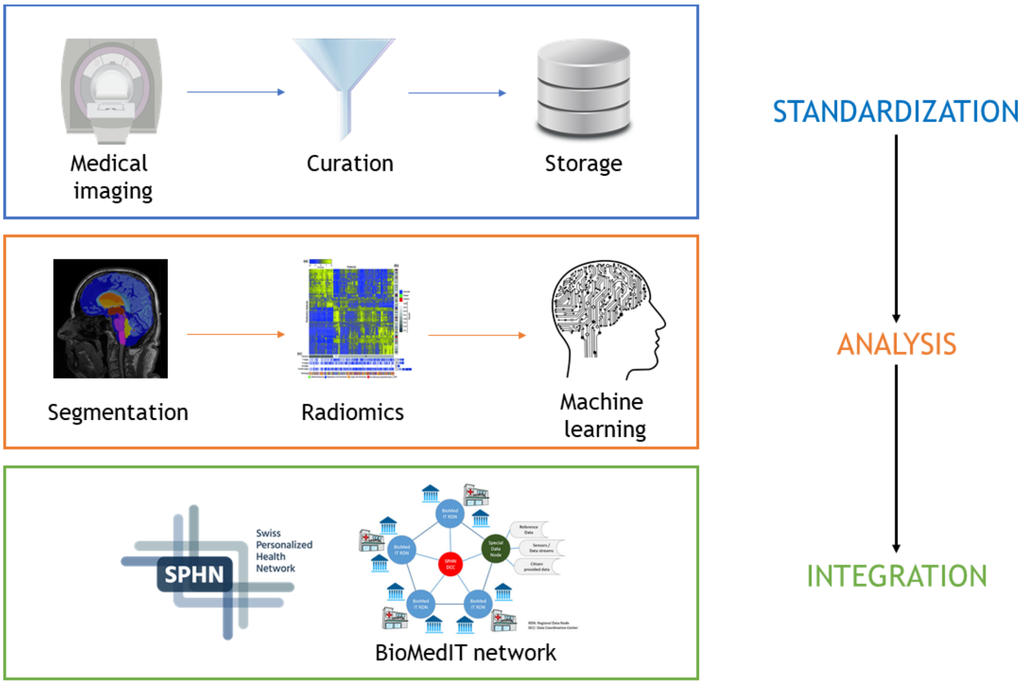


Figure 1 - Structure of IMAGINE Project

Source: Based on (Medigift, n.d.) with images from https://link.springer.com/chapter/10.1007/978-3-319-68843-5_11, https://sphn.ch/fr/home/, https://trends.medicalexpo.fr/project-432097.html

This thesis will concentrate on the standardization aspect, and more precisely on the curation part. Curation can be defined as follows “Data curation includes all the processes needed for principled and controlled data creation, maintenance, and management, together with the capacity to add value to data” (Miller, 2014). In the case of this work, curation involves the anonymization or de-identification of medical data in order to remove all sensitive information about the patient.

1.2. Karnak

Karnak is a tool with normalization and de-identification capabilities. A medical image like a radiography contains sensitive data like the patient's name or the patient's age. It is impossible to use images and data like this in a research context. With Karnak, sensitive information is anonymized and, makes the medical data usable by researchers. For example, in figure 2 the radio is John Doe's, he is 47 and he passed the X-ray exam the 1st of January 2020. After Karnak treatment, sensitive information is removed or replaced so that the patient cannot be recognized.

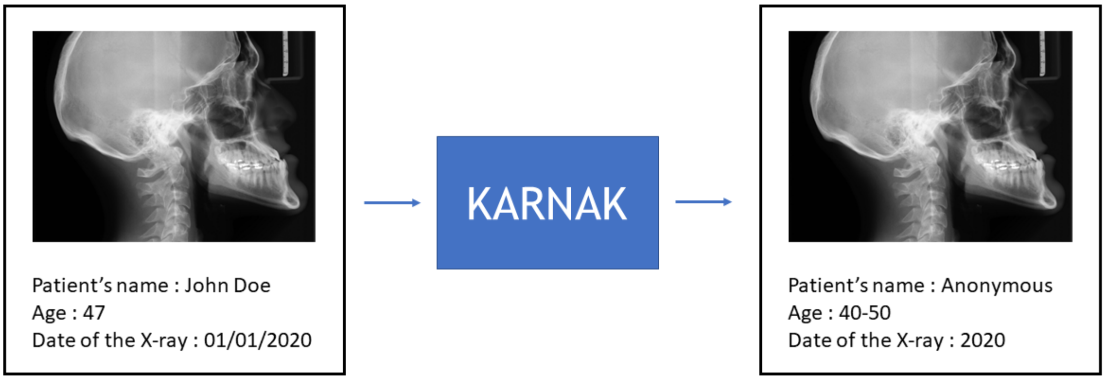

Patient's name : John Doe
Age : 47
Date of the X-ray : 01/01/2020
KARNAK
Patient's name : Anonymous
Age : 40-50
Date of the X-ray : 2020

Figure 2 - Karnak operation

Source: author's source

Karnak is part of a research pipeline modelized in figure 3. Medical imaging equipment like Magnetic Resonance Imaging (MRI) or a user from a workstation can send data to Karnak. After the transformation in Karnak, data can be sent to Kheops (an online application to store, share and display medical images).

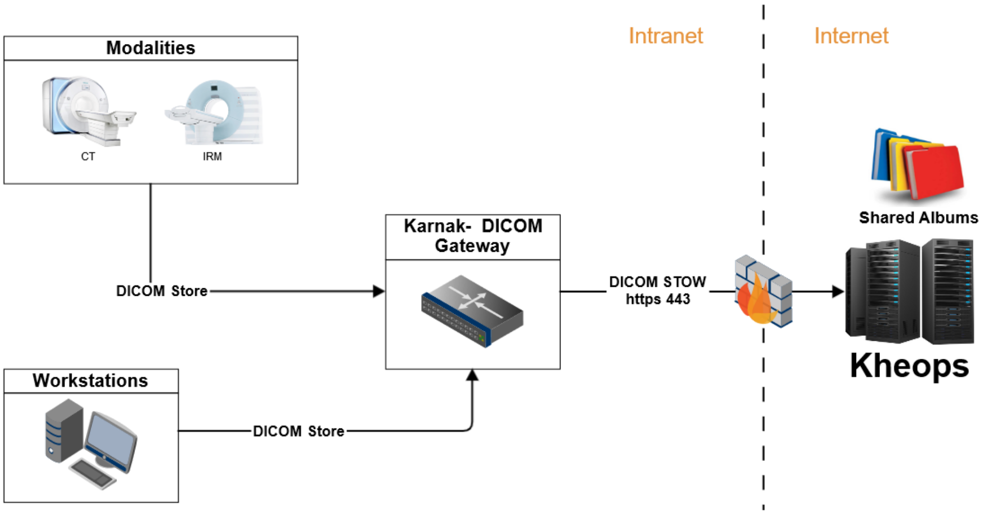


Figure 3 - Workflow of Karnak (OsiriX-Foundation, 2020)

Source : https://github.com/OsiriX-Foundation/Karnak

1.3. Thesis objective

Some research projects demand that certain sensitive data be retained. Each project has different demands in data anonymization. The long-term goal is to propose a prototype of an interactive web interface for Karnak, to allow creation of anonymization profiles. In figure 4, the web interface is represented in the red square.

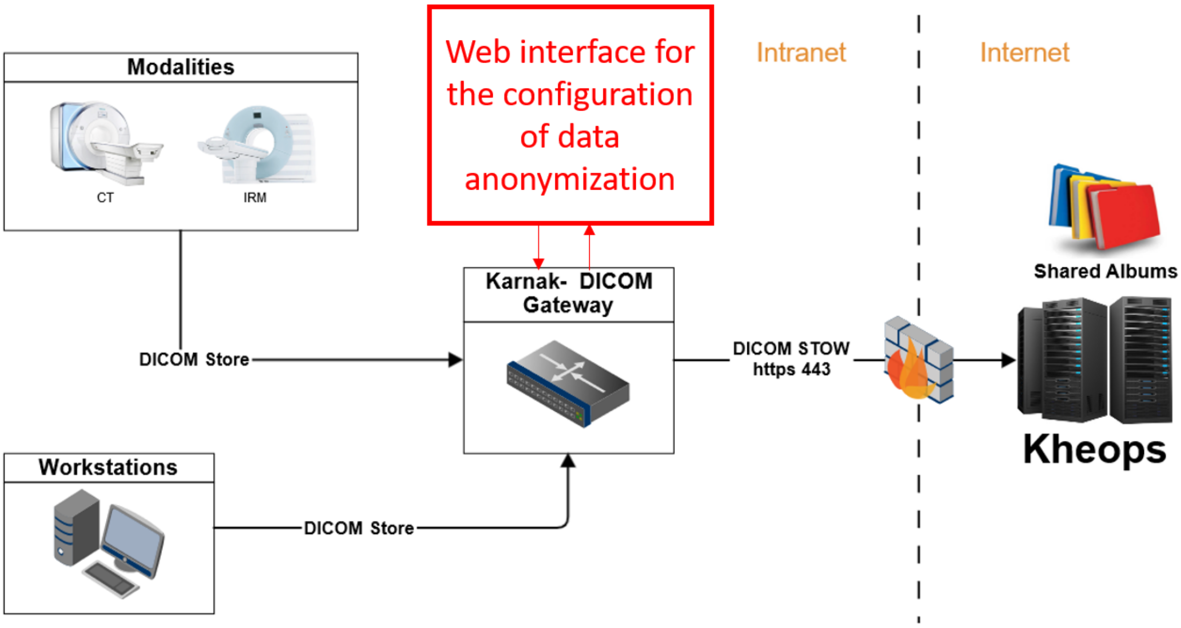


Figure 4 - Thesis objective in Karnak workflow (OsiriX-Foundation, 2020)

Source : author's source based on https://github.com/OsiriX-Foundation/Karnak

However, for this thesis the goal is to propose a web interface to create and configure anonymization profiles. The application will be completely independent of Karnak backend currently under development (https://github.com/OsiriX-Foundation/Karnak).

2. Medical imaging

2.1. Definition

Medical imaging can be defined as follows: "Medical imaging refers to several different technologies that are used to view the human body in order to diagnose, monitor, or treat medical conditions. Each type of technology gives different information about the area of the body being studied or treated, related to possible disease, injury, or the effectiveness of medical treatment." (U.S Food and Drug Administration, 2018).

There are various imaging techniques such as:

- Endoscopy
- Elastography

- Fluoroscopy
- Medical ultrasonography or ultrasound
- MRI
- Etcetera …

2.2. Use for research purposes

The use of medical images in other areas than health and care is increasingly used. "The use of medical images beyond the institutional border, for instance, to build phenotype-specific databases, for teaching and even for research purposes has been increasing." (Monteiro, Costa, & Oliveira, 2017). In this sense it is important to control the use of this data, medical data contains personal information about patients. It's why procedures and conditions must be respected to have the right to use it. Figure 5 describes the workflow for clinical research:

- First an investigator asks to do a research project. The ethics committee decides if the demand is valid or not.
- If the demand is accepted, the investigator has the right to extract data.
- Before analysis, the investigator can validate his cohort by accessing the unencrypted data and add extra data from the patient record.
- After the validation, the investigator is obliged to encode data.
- Data is ready to use in research project.

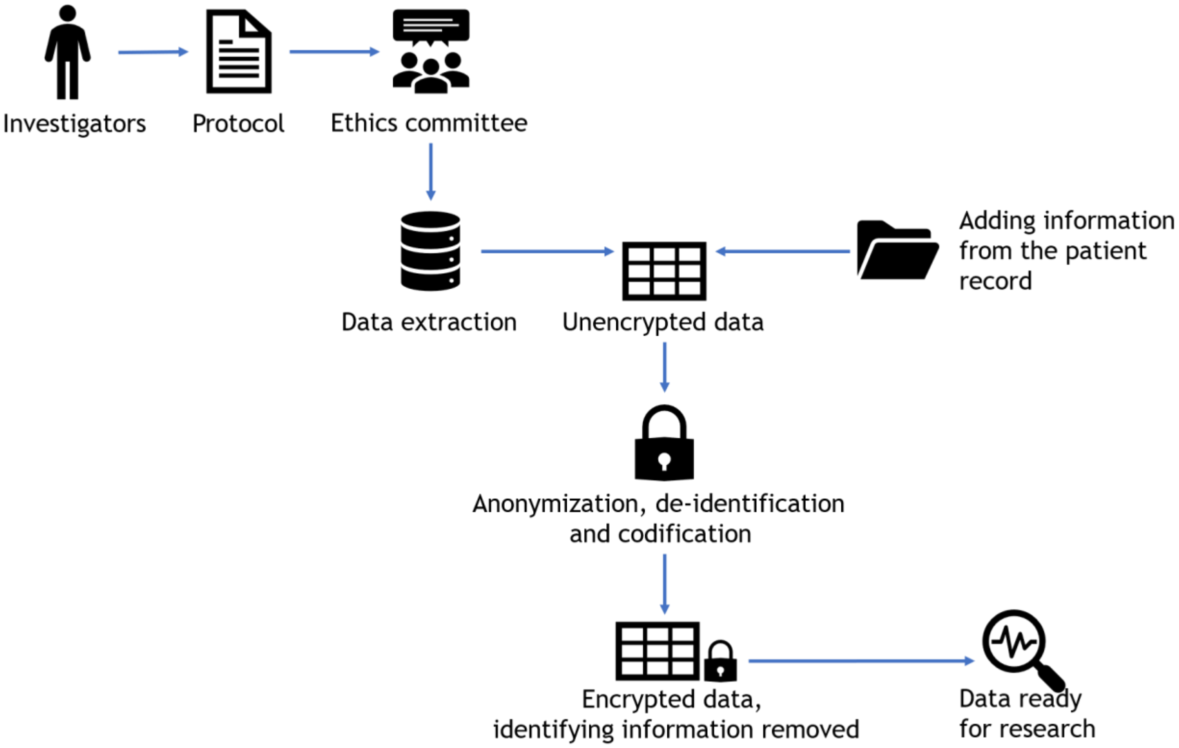

Investigators
Protocol
Ethics committee
Data extraction
Unencrypted data
Adding information from the patient record
Anonymization, de-identification and codification
Encrypted data, identifying information removed
Data ready for research

Figure 5 - Workflow for clinical research

Source: Based on (Unité de Data Science & Recherche (CHUV))

Some official texts govern the use of data for research purposes. The federal act on research involving human beings explains, "Biological material and genetic data may be anonymized for research purposes if the person concerned or the legal representative or next of kin have been informed in advance and have not dissented to anonymization." (Swiss federal authorities, 2014). Whereas to use data in uncoded or coded form the consent of the person concerned is needed. Anonymization is the best solution to collect as much consistent data as possible.

2.3. DICOM format

All medical imaging techniques provide various images and data, from the type of device used for the examination to the private information of the patient. All this data is stored in DICOM (Digital Imaging and Communication in Medicine) format. According to the official website of DICOM standard : "DICOM is the **international standard** to transmit, store, retrieve, print, process, and display **medical imaging** information" (DICOM, 2020).

Real-world objects are represented by IODs (Information Object Definitions) in DICOM format. Oleg S. Pianykh says: "Think about IODs as collections of attributes, describing each particular data object" (Pianykh, 2008). In figure 5, the patient is represented by a table of attributes (Name, ID …).

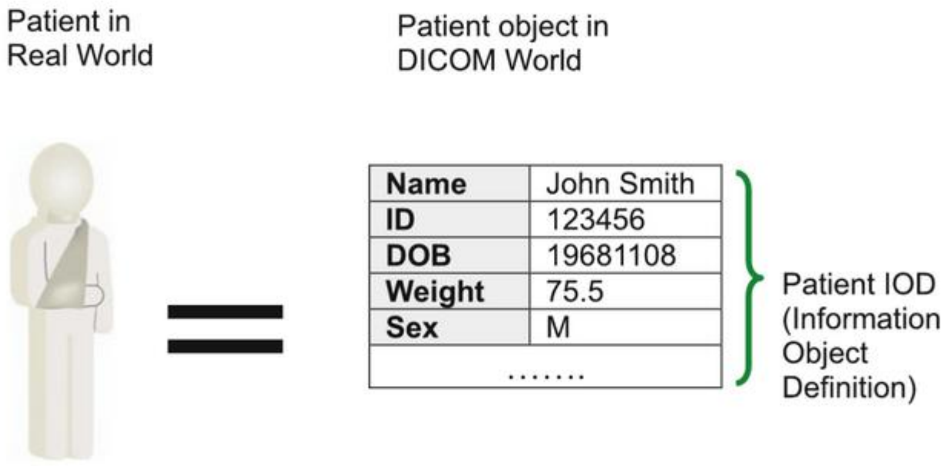


Figure 6 - Representation of a patient in DICOM world

Source: (Pianykh, 2008)

Attributes are composed of a name, a tag and a VR (Value Representative).

- A name, for example "Treatment date".
- A tag is composed of two numbers. The first is the ID of a group of attributes, the second is the ID of the attributes in the group. For example, "(3008,0250)".
- A VR is the data type and format of the attribute's value. For example, DA, is a string of characters of the format "yyyymmdd".

In general, the attributes are represented by their name and tag as shown in figure 7 below.

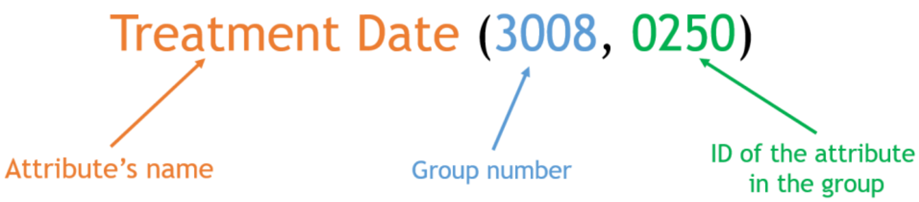
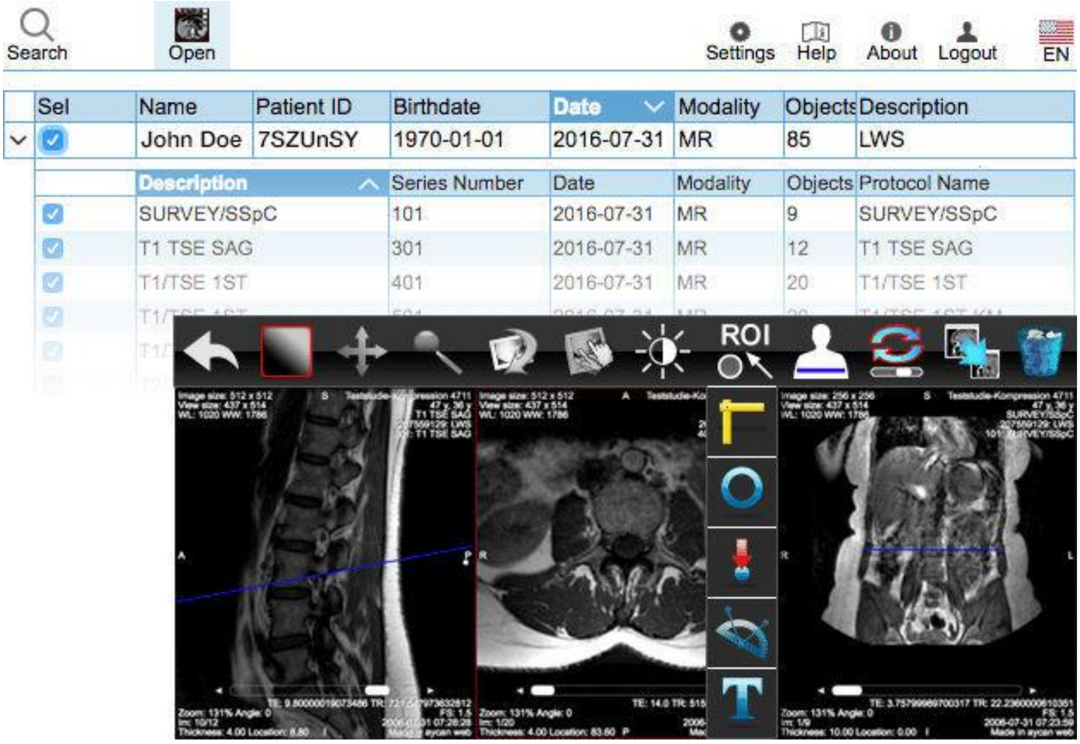


Figure 7 - Format for writing an attribute

Source: author's source

In an application of visualization, a DICOM image is displayed as shown in figure 8, pictures of the patient's body are related to information about him (name, patient ID, birthdate …).



Sel	Name	Patient ID	Birthdate	Date	Modality	Objects	Description
☑	John Doe	7SZUnSY	1970-01-01	2016-07-31	MR	85	LWS

Description	Series Number	Date	Modality	Objects	Protocol Name
☑ SURVEY/SSpC	101	2016-07-31	MR	9	SURVEY/SSpC
☑ T1 TSE SAG	301	2016-07-31	MR	12	T1 TSE SAG
☑ T1/TSE 1ST	401	2016-07-31	MR	20	T1/TSE 1ST

Figure 8 - Visualization of a DICOM file

Source : https://www.medicalexpo.fr/prod/aycan-medical-systems/product-100474-676514.html

This data may be sensitive and some of it may need to be modified or removed from the file to be used for research purposes. To solve this problem, DICOM uses profiles. "Profiles are provided to address the balance between the removal of information and the need to retain information so that the Datasets remain useful for their intended purpose." (DICOM, 2020). The goal of these profiles is to protect the patient and the organization involved by deleting or replacing data to avoid potential leakage of personal identifiable information. But also to allow researchers to keep some important information for their project.

Different rules applicable to attributes exists, they are described in table 1 below.

CODE	DESCRIPTION	EXAMPLE
D	Replaced by a dummy value.	“John Smith” replaced by “XXX”.
Z	Replaced by a null value.	“John Smith” replaced by “null”.
X	Remove.	The field “firstname” is removed.
K	Keep.	“John Smith”.
C	Clean, replaced by value with similar meaning.	“Monday 1st June” replaced by “Monday”.
U	Replace by a dummy UID.	“12345AB” replaced by “98765YZ”.

Table 1 - Explanations of the different rules for changing data in DICOM

Source: (DICOM, 2020)

Rules are part of the profile, in table 2 an excerpt of DICOM basic profile (a safe profile that avoids any leakages). For example, the date of admission of the patient must be removed according to this profile. The patient’s birth date must be replaced by a null value and the person’s name must be replaced by a dummy value.

ATTRIBUTE NAME	BASIC PROFILE
Date of admission	X
Patient's birth date	Z
Person’s name	D

Table 2 - Excerpt of DICOM basic profile

Source: (DICOM, 2020)

Profiles are customizable, everyone can adapt the anonymization rules according to their needs and the instruction from an institution. Advanced options exist and can be applied to certain attributes:

- Retain safe private option.
- Retain UIDs option.
- Retain device identity option.
- Retain patient characteristics option.
- Retain long full dates option.
- Retain long modified dates option.
- Clean description option.

- Clean structured continued option.
- Clean graph option.

These options make it possible to assign a rule to certain attributes. (DICOM, 2020)

2.4. Anonymization VS de-identification

It is important to note that there are two ways of anonymizing data: anonymization or depersonalization/pseudonymization.

Anonymization is an irreversible separation of the link between the dataset and the identity of the dataset contributors in order to prevent future re-identification under any conditions, including for study organizers (Unité de Data Science & Recherche (CHUV)).

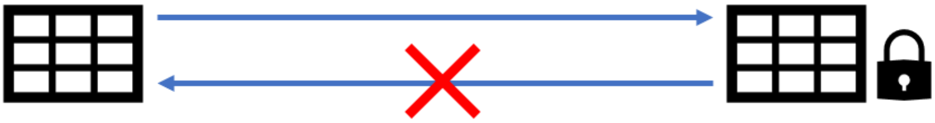


Figure 9 - How anonymization works

Source: author’s source

Unlike anonymization, the codification/depersonalization reversibly breaks the link between the dataset and the identity of the dataset contributors. Re-identification, by a trusted third party and under certain conditions, is then made possible by means of correspondence tables between real identities and coded identities (Unité de Data Science & Recherche (CHUV)).

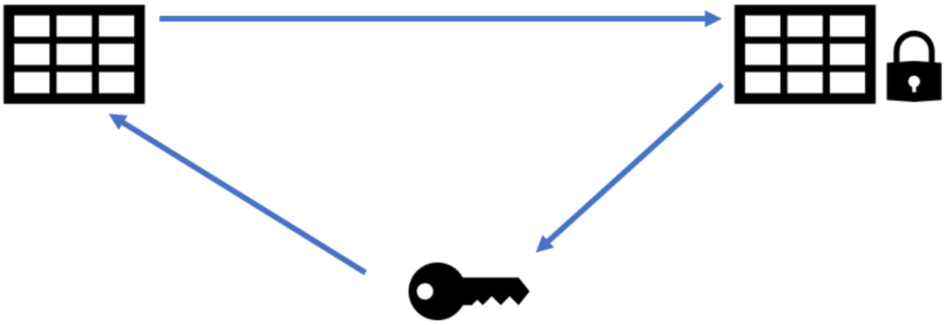


Figure 10 - How depersonalization works

Source: author’s source

3. Existing anonymization tools

The aim of the following section is to list the DICOM anonymization tools available on the market. Anonymization DICOM software can be divided into three categories:

- Software sold with medical imaging machines (MRI, scanner …) used by healthcare professionals.
- Software available on internet.
- Software included in a research pipeline.

3.1. Software for healthcare professionals

Many manufacturers offer software with their imaging medical devices. This software has features such as:

- Display images.
- Edit images and metadata.
- Save images on hard disk or in patient's folder.
- Anonymize images (remove patient's information on the picture).

For example, Siemens, Philips or Samsung provide this kind of software with their devices. However, these tools are reserved for health professionals, it is difficult to obtain more precise information.

3.2. Independant software

There is different open source software available on the internet. This is disconnected from any machine or system. A lot of independent or amateur developers make their work available on platforms, however for the following part the selected software has been designed by research groups or companies.

NAME OF SOFTWARE	COMPANY/RESEARCH GROUP	WEBSITES
DICOM Browser	Neuroinformatics Research Group (NRG) of the Washington University School of Medecine	https://nrg.wustl.edu/software/dicom-browser/
DICOM Anonymizer	Rubo Medical Imaging, a company specialized in DICOM software	http://www.rubomedical.com/dicom_anonymizer/
Dicomanonymizer	Neologica, a company specialized in medical imaging solution	http://www.dclunie.com/pixelmed/software/webstart/DicomCleanerUsage.html

Table 3 - Anonymization of DICOM anonymization software

Source: author's source

After a test of each of the above software options, the main features are as follows ('X' means that these features are proposed by the software, if there is nothing, it means these features are not proposed by the software):

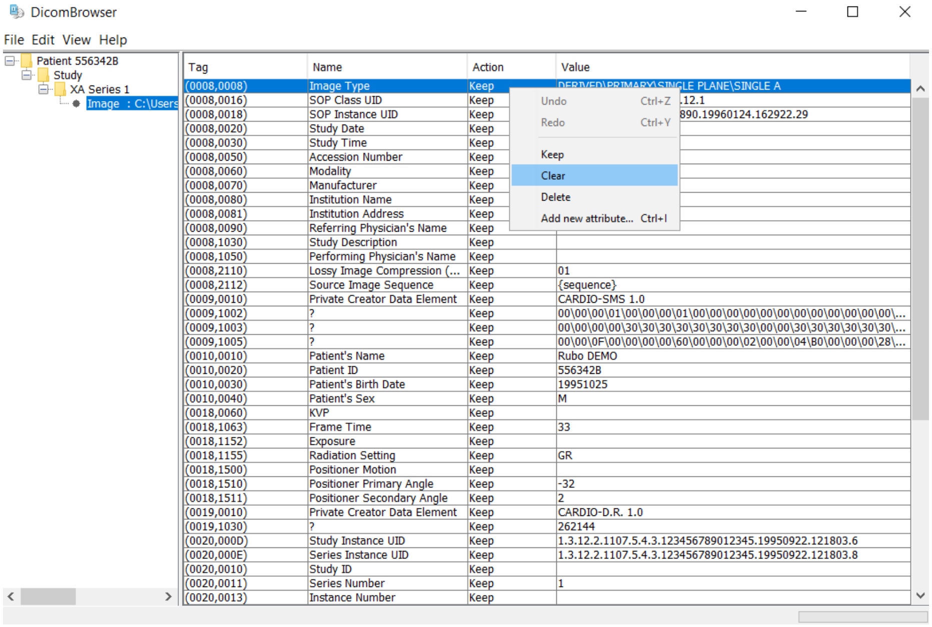
	DICOM Browser	DICOM Anonymizer	Dicomanonymizer
Import DICOM files from the disk	X	X	X
Display DICOM attributes	X		
Change anonymization rules	X		
Anonymize few images with the same profile		X	X
Save file to disk	X	X	X
Set custom values for attributes			X
Edit all the attributes	X		

Table 4 - Features available in tested software

Source: author's source

All tools offer to work on an existing DICOM file and to save it on the disk after anonymization. However only one software tool proposes to modify the anonymization rules on all attributes, the other applies anonymization only on a few attributes like patient UID. The interfaces are basic and not always easy to use. This software must be downloaded and installed on the user's computer, none of them are a web application.

For example, in DICOMBrowser after importing a DICOM file, in order to change rules of anonymization (keep, clear or delete), it's necessary to right click on the line of the attribute and select (figure 11). The default anonymization rule is 'Keep'. It's impossible to apply a rule to several attributes. It is possible to modify an attribute directly by clicking on its value, no verification is done. The interface is basic, there is no user guide or help.



Tag	Name	Action	Value
(0008,0008)	Image Type	Keep	DERIVED\PRIMARY\SINGLE PLANE\SINGLE A
(0008,0016)	SOP Class UID	Keep	.12.1
(0008,0018)	SOP Instance UID	Keep	890.19960124.162922.29
(0008,0020)	Study Date	Keep	
(0008,0030)	Study Time	Keep	
(0008,0050)	Accession Number	Keep	
(0008,0060)	Modality	Keep	
(0008,0070)	Manufacturer	Keep	
(0008,0080)	Institution Name	Keep	
(0008,0081)	Institution Address	Keep	
(0008,0090)	Referring Physician's Name	Keep	
(0008,1030)	Study Description	Keep	
(0008,1050)	Performing Physician's Name	Keep	
(0008,2110)	Lossy Image Compression (...	Keep	01
(0008,2112)	Source Image Sequence	Keep	{sequence}
(0009,0010)	Private Creator Data Element	Keep	CARDIO-SMS 1.0
(0009,1002)	?	Keep	00\00\00\01\00\00\00\01\00\00\00\00\00\00\00\00\00\00\00\...
(0009,1003)	?	Keep	00\00\00\00\30\30\30\30\30\30\30\00\00\30\30\30\30\30\30\...
(0009,1005)	?	Keep	00\00\0F\00\00\00\00\60\00\00\00\02\00\00\04\B0\00\00\00\28\...
(0010,0010)	Patient's Name	Keep	Rubo DEMO
(0010,0020)	Patient ID	Keep	556342B
(0010,0030)	Patient's Birth Date	Keep	19951025
(0010,0040)	Patient's Sex	Keep	M
(0018,0060)	KVP	Keep	
(0018,1063)	Frame Time	Keep	33
(0018,1152)	Exposure	Keep	
(0018,1155)	Radiation Setting	Keep	GR
(0018,1500)	Positioner Motion	Keep	
(0018,1510)	Positioner Primary Angle	Keep	-32
(0018,1511)	Positioner Secondary Angle	Keep	2
(0019,0010)	Private Creator Data Element	Keep	CARDIO-D.R. 1.0
(0019,1030)	?	Keep	262144
(0020,000D)	Study Instance UID	Keep	1.3.12.2.1107.5.4.3.123456789012345.19950922.121803.6
(0020,000E)	Series Instance UID	Keep	1.3.12.2.1107.5.4.3.123456789012345.19950922.121803.8
(0020,0010)	Study ID	Keep	
(0020,0011)	Series Number	Keep	1
(0020,0013)	Instance Number	Keep	

Figure 11 - Screenshot of DICOM browser interface

Source: Screenshot of DicomBrowser

3.3. Software included in a research pipeline

A study was conducted in 2016 to create “a de-identification pipeline for ultrasound medical images in DICOM format” (Monteiro, Costa, & Oliveira, 2017). The goal is to recognize and then hide information on images as shown in figure 12 below with the white square.

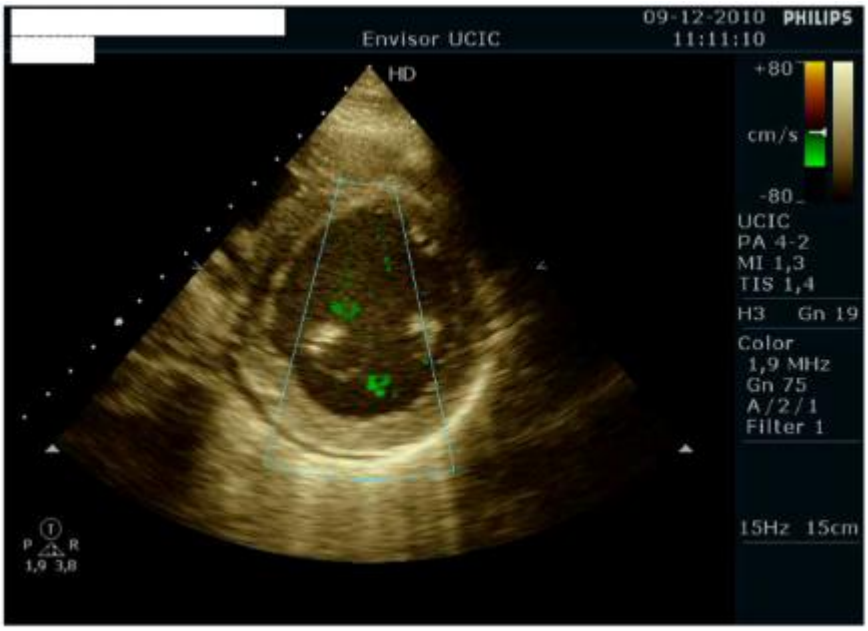


Figure 12 - Anonymzed DICOM image

Source: https://link.springer.com/content/pdf/10.1007/s10916-017-0736-1.pdf

The study was a success, de-identification in pixel data is possible but there are limitations. Certain factors such as image resolution can influence the result. However, this tool does not deal with the anonymization of metadata, only metadata on the images. No configuration is possible, the goal is to anonymize all the sensitive data present on the image.

3.4. Conclusion

Of the various existing tools mentioned above, none of them is a web application, they all require installation. The notion of profile is not discussed, and the work is done on DICOM files where the attributes to be anonymized are generally the basic ones (identity of the patient ...).

4. Definition of use cases

After a study of existing tools, this section will focus on the use cases definition for the application. The goal is to build an easy-to-use web interface to create anonymization profile for DICOM files.

4.1. Users

For this application two types of users can be defined:

- Researchers, someone who wants to create an anonymization profile for his research project and share it with his colleagues.
- Institutions wishing to establish an institutional profile. For example, a hospital wishing to establish its own profile of anonymization and share it with his collaborators. The person responsible for creating this profile does not necessarily have the technical knowledge in computer science and medical imaging.

4.2. Use cases

After discussion with Nicolas Roduit computer scientist at the University Hospitals of Geneva, it was decided that through the application, the user must be able to:

- Apply anonymization rules to DICOM attributes.
- Search for attributes by module, tags or name.
- Apply an advanced option to attributes.
- Import an existing profile of anonymization and edit it.
- Export the profile of anonymization in JavaScript Object Notation (JSON) format.

From that, the use cases diagram was constructed:

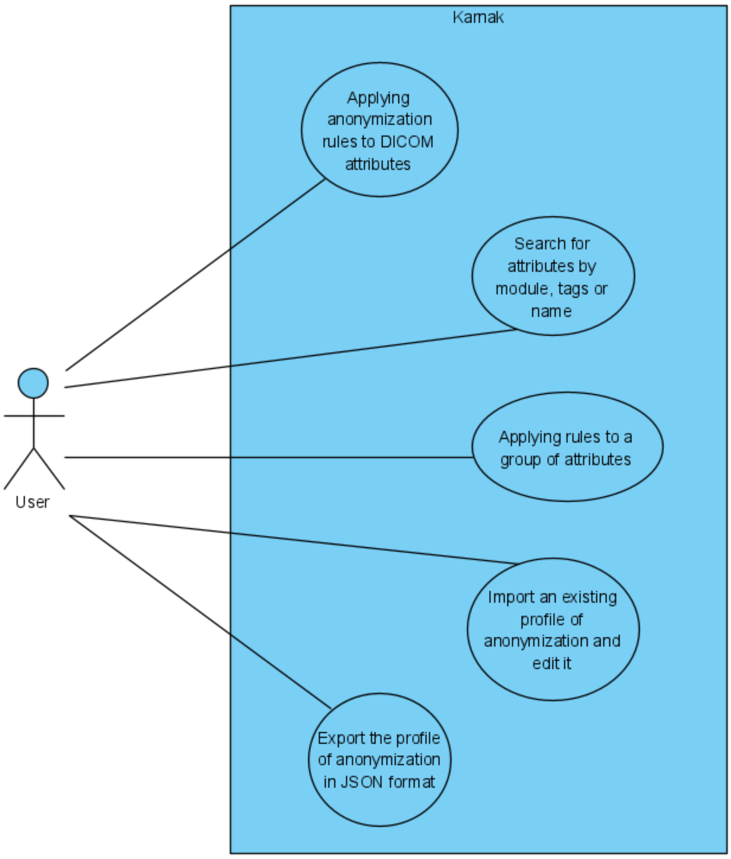

Karnak
Applying anonymization rules to DICOM attributes
Search for attributes by module, tags or name
Applying rules to a group of attributes
Import an existing profile of anonymization and edit it
Export the profile of anonymization in JSON format
User

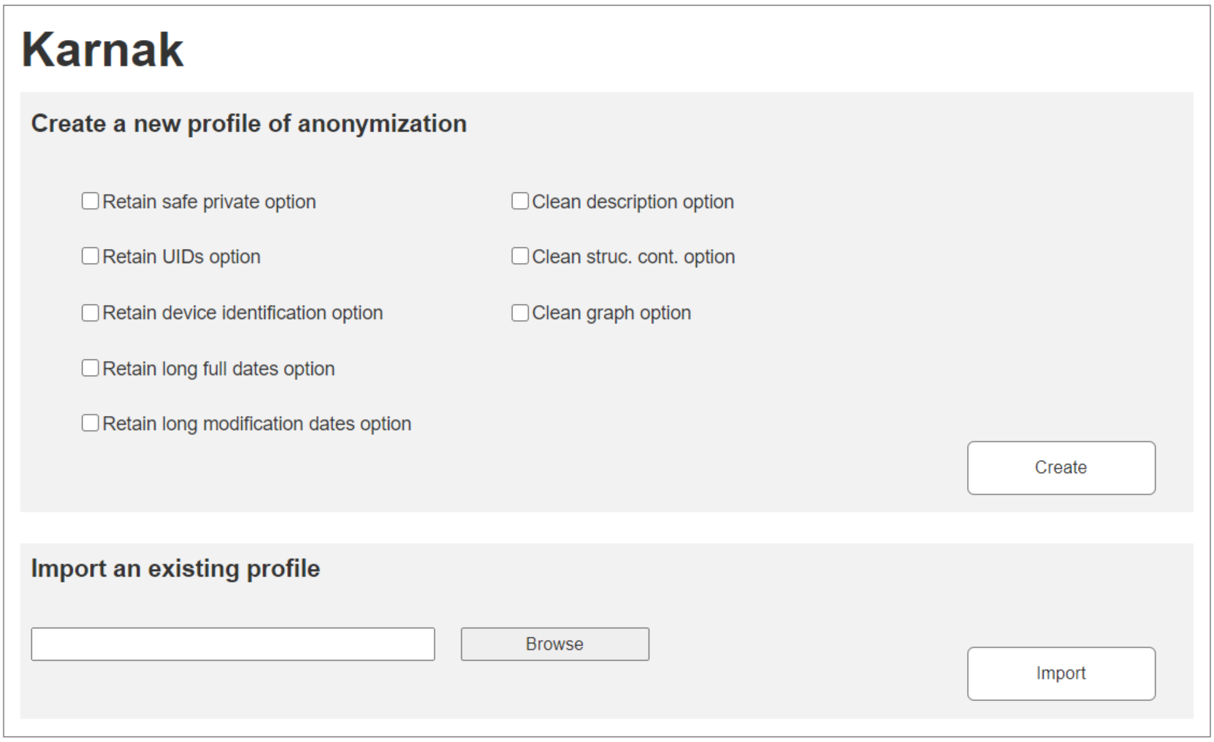
Figure 13 - Use cases diagram

Source: author's source

The interface must be user friendly; users must quickly understand how to use this application.

4.3. Mock-ups

Different mock-ups have been made from the established use cases (figure 14 and 15).



Karnak

Create a new profile of anonymization

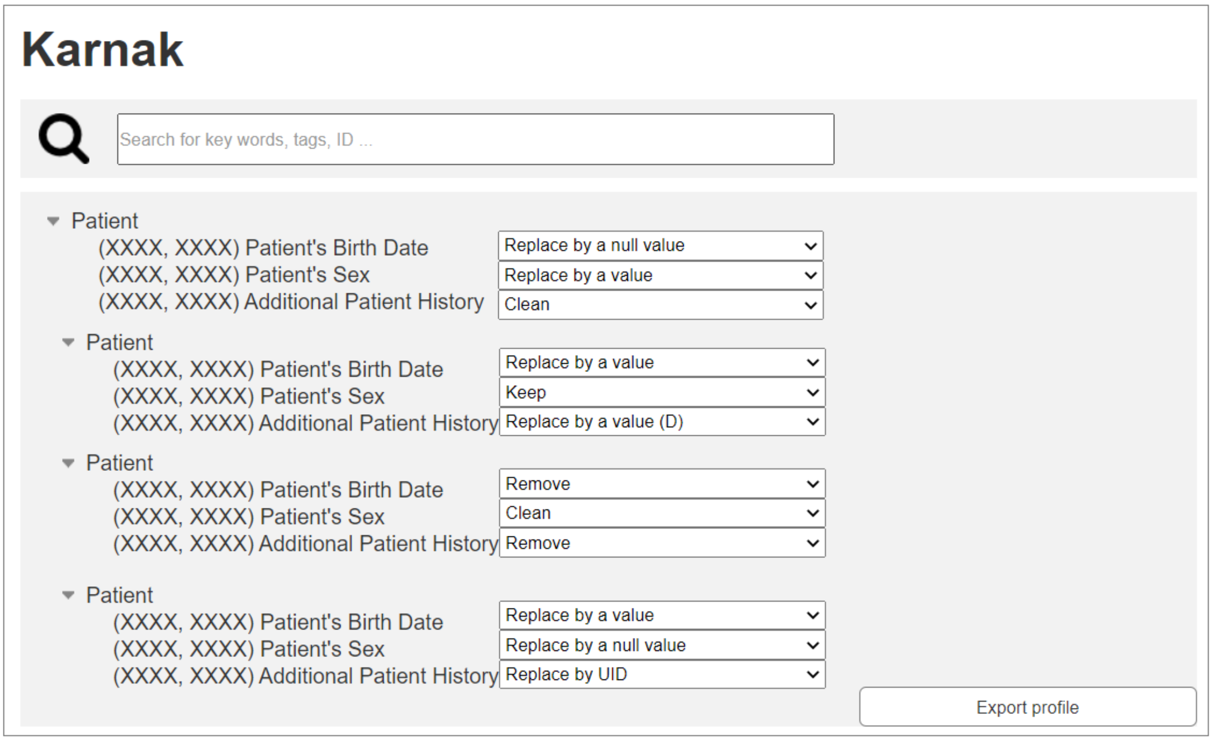
□ Retain safe private option
□ Retain UIDs option
□ Retain device identification option
□ Retain long full dates option
□ Retain long modification dates option

□ Clean description option
□ Clean struc. cont. option
□ Clean graph option

Import an existing profile

Figure 14 - Mock-up of the home page

Source: author's source



Karnak

Search for key words, tags, ID ...

▾ Patient
(XXXX, XXXX) Patient's Birth Date — Replace by a null value
(XXXX, XXXX) Patient's Sex — Replace by a value
(XXXX, XXXX) Additional Patient History — Clean

▾ Patient
(XXXX, XXXX) Patient's Birth Date — Replace by a value
(XXXX, XXXX) Patient's Sex — Keep
(XXXX, XXXX) Additional Patient History — Replace by a value (D)

▾ Patient
(XXXX, XXXX) Patient's Birth Date — Remove
(XXXX, XXXX) Patient's Sex — Clean
(XXXX, XXXX) Additional Patient History — Remove

▾ Patient
(XXXX, XXXX) Patient's Birth Date — Replace by a value
(XXXX, XXXX) Patient's Sex — Replace by a null value
(XXXX, XXXX) Additional Patient History — Replace by UID

Figure 15 - Mock-up of the edit page

Source: author's source

The home page (figure 14) is divided in two parts. The first section is for creation of new anonymization profile. The user can check some advanced options to apply rules of anonymization on to a group of attributes before the creation of the profile. The second part is for the importation of an existing profile.

In the edit page (figure 15) all attributes representing a threat to data anonymity are displayed (an extract of the file confidentiality_profile_attributes.json containing all these attributes can be found in annex). The user can change the anonymization rules according to these needs. A search bar is also available to search for an attribute based on its name, module name or tag. Once completed, the user can export his profile in Json format.

5. Choice of technologies

5.1. Frontend framework

In this section, the technology used for the development of the prototype will be determined. The following frameworks and libraries have been pre-selected:

React

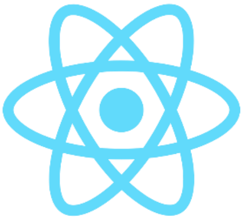


Figure 16 - React logo

Source: https://fr.reactjs.org/

Angular

Figure 17 - Angular logo

Source : https://angular.io/

Vue.js

Figure 18 - View.js logo

Source : https://vuejs.org/

Angular, React and Vue.js are the most appreciated and modern web technologies (Benitte, Greif, & Rambeau, 2019). That's why they have been pre-selected among all the other frontend frameworks.

5.1.1. Evaluation criteria

In the table below, different comparison criteria for the choice of technology.

CRITERIA	DESCRIPTION	POINTS
Documentation/Community	Is there quality documentation online? Is there a large community posting tutorials, questions on the forums? It is the main criteria, if available resources are varied and of quality, they can	10

	influence development from the installation of the environment to the deployment.	
Web interface requirements	Does the framework or library fit the needs of the application?	5
Libraries/frameworks available for User Interface (UI)	Are there libraries/frameworks for UI? Does a library allow a large choice of components including (tree view, list, button ...)?	5
Installation/creation of the web application	Is it easy to set up the development environment? Is the technology easy to use?	5

Table 5 - Selection criteria for frontend technology

Source: author's source

5.1.2. React

React is an open source JavaScript library for designing user interfaces, developed by Facebook in 2013. This library benefits from high-quality and regularly updated documentation. React fits the needs of the application, it's a library created for building user interfaces. In addition, the changes of anonymization rules in our interface, will impact UI, these changes can be managed easily with state. React benefits from many libraries/frameworks that can be added. For example, Material-UI or Ant-Design has all the necessary components. About the installation, the command 'create-react-app' makes it possible to set up an environment quickly and easily.

Grade: 24/25.

5.1.3. Angular

Angular is a TypeScript open source web application framework, developed by Google in 2016. Documentation is updated regularly; a lot of resources are available on the website. Angular needs clean code structure and is not very flexible (Wohlgethan, 2018). In the case of the application, flexible tool is preferable. There are many libraries for UI like Angular Material Design and has all the necessary components. Installation is quick and easy, it is necessary to know TypeScript in addition to JavaScript, html and CSS.

Grade: 22/25.

5.1.4. Vue.js

Vue.js is an open source JavaScript framework to construct user interfaces, created by Evan You in 2014. The documentation is complete; however, his community is less important than React and Angular.

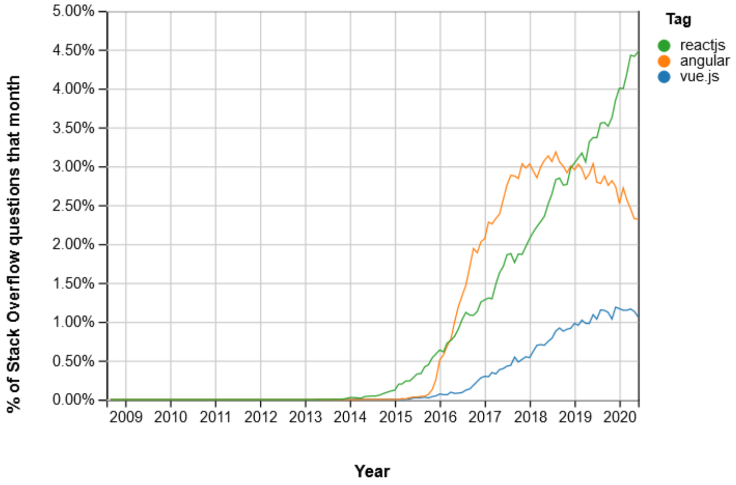


Figure 19 - Graph of percentage of Stack Overflow questions vs. technology per month

Source: (Stackoverflow, 2020)

Vue.js is flexible and can meet the needs of the application. There are many libraries for UI like Vuetify, this library contains all the components necessary for the design. Setting up the environment is not difficult, it’s necessary to know JavaScript, HTML and CSS.

Grade 23/25.

5.1.5. Conclusion

These three technologies are similar, the documentation is regularly updated and complete. These three technologies are component based, however the structure needed for angular and the smaller community of Vue.js can be problematic during the development. React seems to be the best technology to build the application.

For more details the comparison table is available in the appendix (annex I).

5.2. Main tools and libraries used

5.2.1. Material-UI

For UI, Material-UI will be use. It is a library created in 2014 “to unify React and Material Design” (Material-UI, 2020). Materia Design created by Google, is a design for unification of interfaces between devices (phones, tablets, televisions, websites …) (Zeste de savoir, 2019).

Material-UI has a large catalog of open source components and icons. Using this library has some advantages:

- Style management: no stylesheet is needed. Material-UI offers directly function to edit CSS (Cascading Style Sheet).

- Consistency of components (colors, fonts are the same for all components), it saves time and allow to build a user-friendly web interface faster.
- A strong community uses this library and clear documentation is online.

5.2.2. Github

"GitHub is a collaborative code hosting site built on top of the git version control system." (Kalliamvakou, et al., 2014). In order to save the code incrementally during the development, a repository has been created and is available (https://github.com/audreyviriot/webInterface_Karnak). Karnak's team can access to this online repository.

5.2.3. Firebase Hosting

Firebase is a group of services offered by Google. Services are for example:

- Cloud
- Database
- Authentication
- Hosting
- …

To facilitate deployment, Firebase Hosting will be use. With only one command, web application can be online (Firbase, 2020).

6. Prototype

6.1. Architecture

Web application will be built as follows (figure 20):

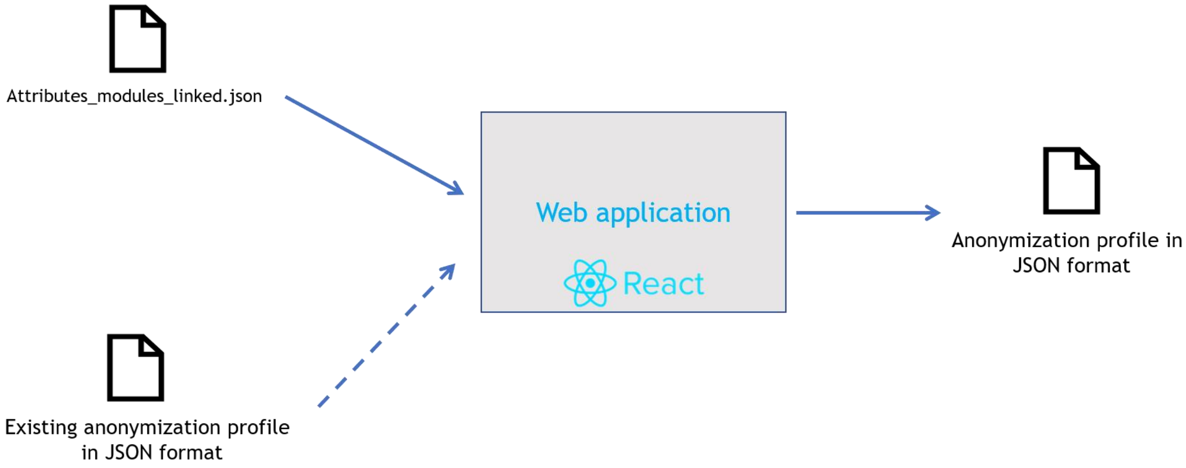

Attributes_modules_linked.json
Web application
React
Anonymization profile in JSON format
Existing anonymization profile in JSON format

Figure 20 - Architecture of the prototype (inputs/outputs)

Source: author’s source

Data will be provided by the JSON “attributes_modules_linked.json” built from two JSON files: ‘confidentiality_profile_attributes.json’ and ‘module_to_attributes.json’, . User will also be able to provide an existing anonymization profile in JSON format. The result will be an anonymization profile in JSON format. As part of the prototype, the output (anonymization profile in JSON format) will be saved on user’s computer.

With this structure, web application can be adapted to any backend. Data contained in the output file will provide all information needed for anonymization functions. In the figure below, an example of architecture with a backend.

- The presentation layer is the UI.
- API layer allows to exchange data between the presentation layer and the backend.
- Business layer transforms data, that’s where anonymization happens. With information sent by frontend, business layer knows which anonymization rule applies to which attribute.
- Persistence layer facilitates access to database.

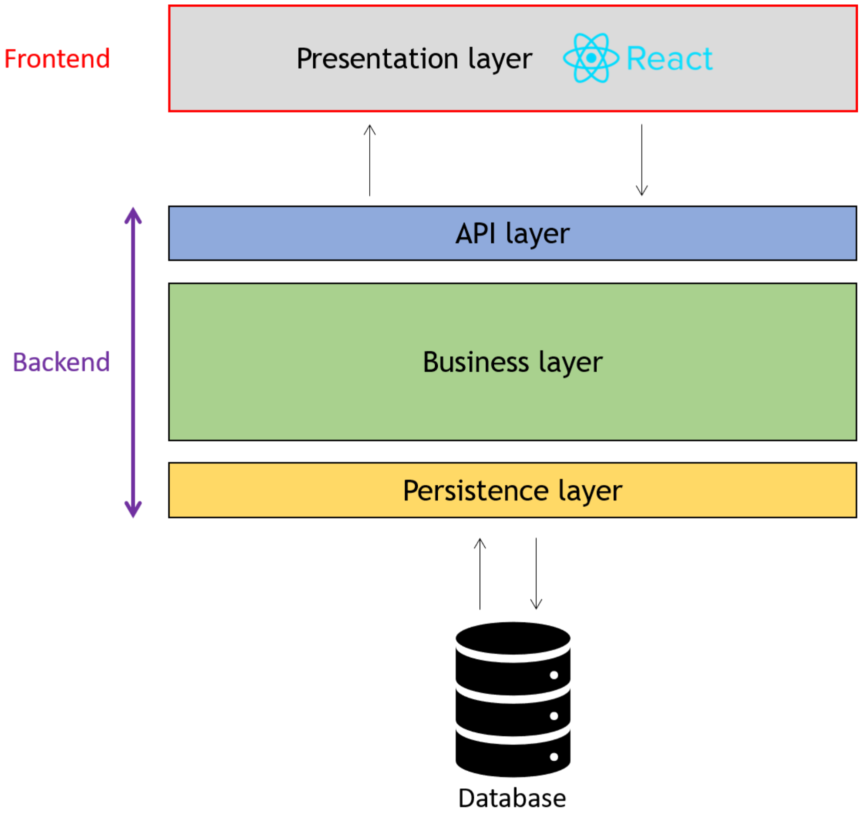


Figure 21 - Example of an architecture with a backend

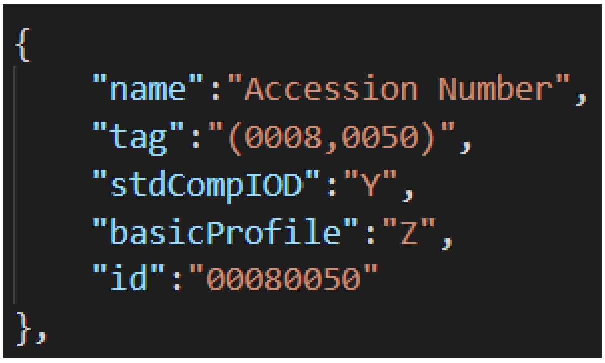
Source: author’s source

6.2. Data

The following files (confidentiality_profile_attributes.json and module_to_attribute.json) are from DICOM Standard Parser. “This program parses the web version of the DICOM standard into human and machine-friendly JSON files.” (Innolitics, 2020). Data in these files will be used in the application

6.2.1. confidentiality_profile_attributes.json

For this prototype, only attributes representing a potential problem for patient confidentiality are considered. All this data can be found in the file ‘confidentiality_profile_attributes.json’. The file is structured as follows:



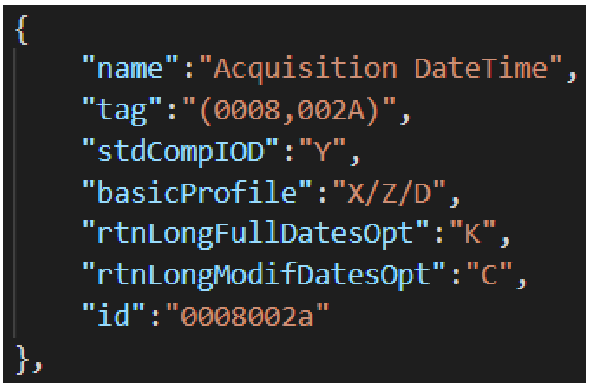
```
{
    "name":"Accession Number",
    "tag":"(0008,0050)",
    "stdCompIOD":"Y",
    "basicProfile":"Z",
    "id":"00080050"
},
```

Figure 22 - Structure of an attribute in JSON

Source: (Innolitics, 2020)

Each tag has name, tag, is part of standard composite IODs (yes Y or no N), have a rule of anonymization for the basic profile and an id. Each field will be useful except ‘stdCompIOD’, this information will not be used. Some attributes may have other anonymization rules as explained in part 2.3. The following properties can be found in concerned objects:

- rtnSafePrivOpt
- rtnUIDsOpt
- rtnDevIdOpt
- rtnPatCharsOpt
- rtnLongFullDatesOpt
- rtnLongModifDatesOpt
- cleanDescOpt
- cleanStructContOpt
- cleanGraphOpt"



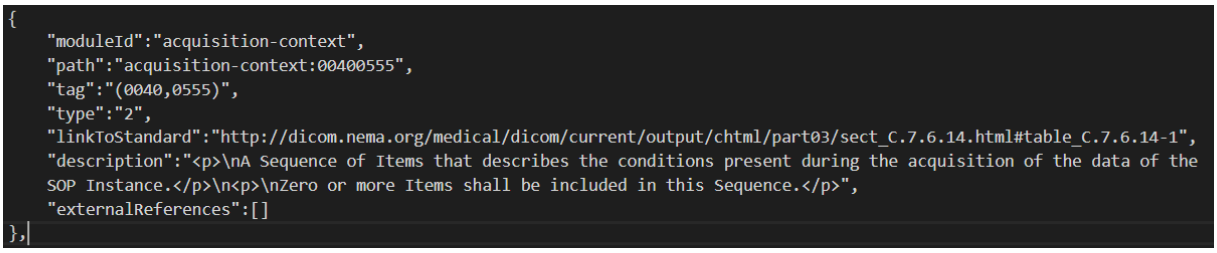
```
{
    "name":"Acquisition DateTime",
    "tag":"(0008,002A)",
    "stdCompIOD":"Y",
    "basicProfile":"X/Z/D",
    "rtnLongFullDatesOpt":"K",
    "rtnLongModifDatesOpt":"C",
    "id":"0008002a"
},
```

Figure 23 - Structure of an attributes in JSON with additional properties

Source: (Innolitics, 2020)

6.2.2. module_to_attribute.json

In order to group the attributes, a second JSON file was used: 'module_to_attribute.json'.



```
{
    "moduleId":"acquisition-context",
    "path":"acquisition-context:00400555",
    "tag":"(0040,0555)",
    "type":"2",
    "linkToStandard":"http://dicom.nema.org/medical/dicom/current/output/chtml/part03/sect_C.7.6.14.html#table_C.7.6.14-1",
    "description":"<p>\nA Sequence of Items that describes the conditions present during the acquisition of the data of the
    SOP Instance.</p>\n<p>\nZero or more Items shall be included in this Sequence.</p>",
    "externalReferences":[]
},
```

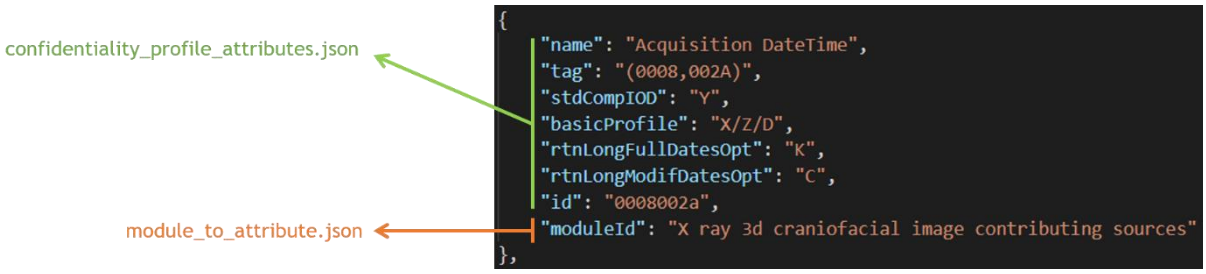
Figure 24 - Object in module_to_attribute.json

Source: (Innolitics, 2020)

This file contains valuable information about attributes. The field "moduleId" will be used to group by module the attributes in the web interfaces.

6.2.3. attributes_modules_linked.json

This file was created using a JavaScript function (annex II) to merge the two files presented above. It is constructed in this way:



```
{
    "name": "Acquisition DateTime",
    "tag": "(0008,002A)",
    "stdCompIOD": "Y",
    "basicProfile": "X/Z/D",
    "rtnLongFullDatesOpt": "K",
    "rtnLongModifDatesOpt": "C",
    "id": "0008002a",
    "moduleId": "X ray 3d craniofacial image contributing sources"
},
```

Figure 25 - Structure of attributes_modules_linked.json

Source: author's source

All the information from “confidentiality_profile_attributes.json” has been kept and the module ID of “module_to_attribute.json” was added. The name of the module has been reformatted so that it is directly usable in frontend.

This file was created so as not to overload the application with too large files. In addition, only one file is required for the application.

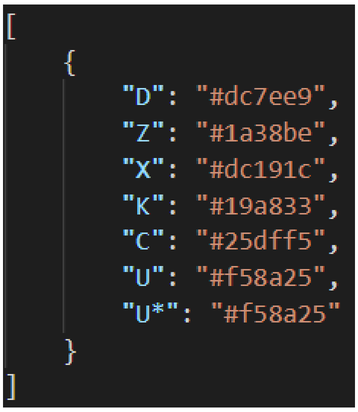
FILE NAME	SIZE	TOTAL
confidentiality_profile_attributes.json	82 KB (Kilobit)	57864 KB
module_to_attribute.json	57782 KB	
attributes_modules_linked.json	104 KB	104 KB

Table 6 - Comparison of different JSON file sizes

Source: author’s source

6.2.4. rules_colors.json

A simple Json file was created to link a rule of anonymization to a color. It is easier to access color from anywhere in the application. It is also convenient to change a color if needed. There is no distinction between U and U*.



```
[
    {
        "D": "#dc7ee9",
        "Z": "#1a38be",
        "X": "#dc191c",
        "K": "#19a833",
        "C": "#25dff5",
        "U": "#f58a25",
        "U*": "#f58a25"
    }
]
```

Figure 26 - Content of rules_colors.json

Source: author’s source

Here is the color code used.

RULE OF ANONYMIZATION	COLOR
D - Replaced by a dummy value	🟪
Z - Replaced by a null value	🟦

X - Remove	🟥
K - Keep	🟩
C - Clean, replaced by value with similar meaning	🟦
U - Replace by a dummy UID	🟧

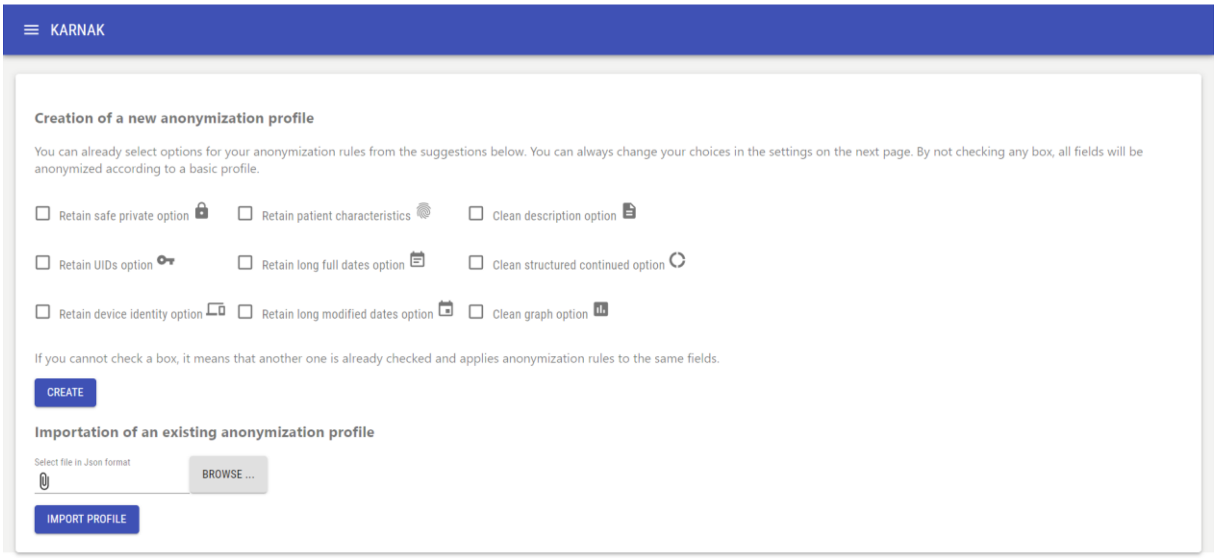
Table 7 - Color code used

Source: author's source

6.3. Interfaces and functionnalities

6.3.1. General appearance of the application

The web application is composed of two pages: the home page and the edit page. The first one allows users to adjust some settings before creating a new profile of anonymization or import an existing anonymization profile.



Creation of a new anonymization profile

You can already select options for your anonymization rules from the suggestions below. You can always change your choices in the settings on the next page. By not checking any box, all fields will be anonymized according to a basic profile.

□ Retain safe private option
□ Retain patient characteristics
□ Clean description option
□ Retain UIDs option
□ Retain long full dates option
□ Clean structured continued option
□ Retain device identity option
□ Retain long modified dates option
□ Clean graph option

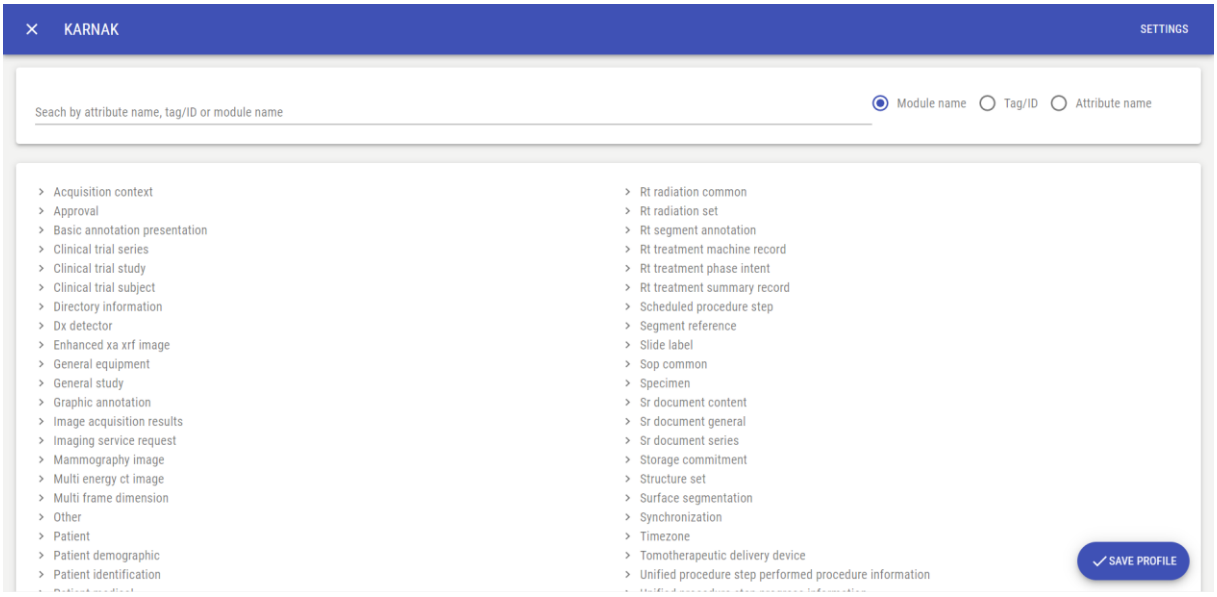
If you cannot check a box, it means that another one is already checked and applies anonymization rules to the same fields.

Importation of an existing anonymization profile

Figure 27 - Home page of the application

Source: author's source

The edit page displays all the tags and allows users to edit anonymization rules of attributes.



Seach by attribute name, tag/ID or module name

● Module name ○ Tag/ID ○ Attribute name

- Acquisition context
- Approval
- Basic annotation presentation
- Clinical trial series
- Clinical trial study
- Clinical trial subject
- Directory information
- Dx detector
- Enhanced xa xrf image
- General equipment
- General study
- Graphic annotation
- Image acquisition results
- Imaging service request
- Mammography image
- Multi energy ct image
- Multi frame dimension
- Other
- Patient
- Patient demographic
- Patient identification
- Patient medical
- Rt radiation common
- Rt radiation set
- Rt segment annotation
- Rt treatment machine record
- Rt treatment phase intent
- Rt treatment summary record
- Scheduled procedure step
- Segment reference
- Slide label
- Sop common
- Specimen
- Sr document content
- Sr document general
- Sr document series
- Storage commitment
- Structure set
- Surface segmentation
- Synchronization
- Timezone
- Tomotherapeutic delivery device
- Unified procedure step performed procedure information
- Unified procedure step progress information

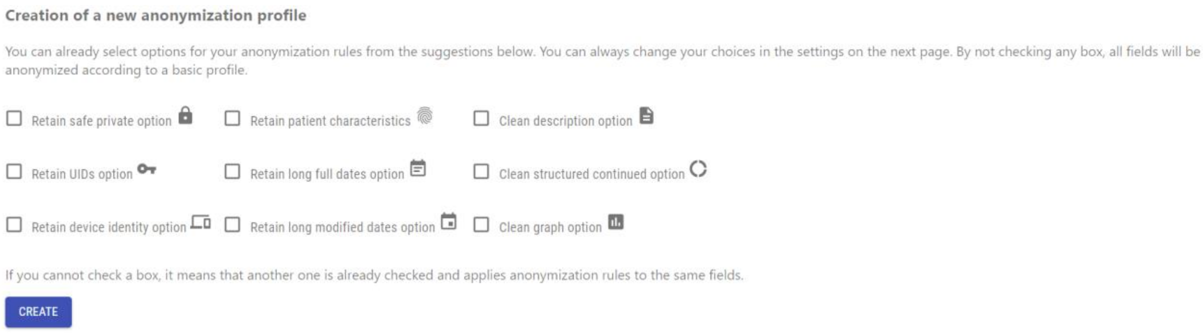
Figure 28 - Edit page of the application

Source: author’s source

Functionalities will be detailed in the following parts.

6.3.2. Setting up anonymization profiles with advanced options

On the first page, users can create a new profile. Depending on their needs they can check advanced options. Once ticked (or not), the user can click on button “Create”.



Creation of a new anonymization profile

You can already select options for your anonymization rules from the suggestions below. You can always change your choices in the settings on the next page. By not checking any box, all fields will be anonymized according to a basic profile.

□ Retain safe private option □ Retain patient characteristics □ Clean description option

□ Retain UIDs option □ Retain long full dates option □ Clean structured continued option

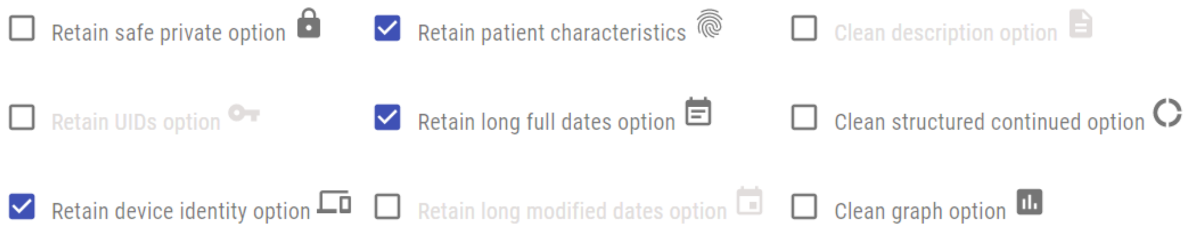
□ Retain device identity option □ Retain long modified dates option □ Clean graph option

If you cannot check a box, it means that another one is already checked and applies anonymization rules to the same fields.

Figure 29 - Creation of a new profile

Source: author’s source

Some attributes have several anonymization rules possible among the advanced options. To avoid conflicts, if one option is checked and can potentially affect the same field as another, the second is disabled.



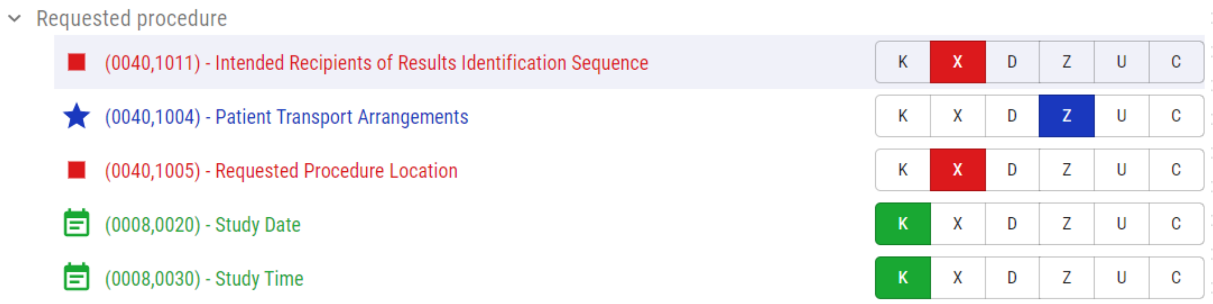
☐ Retain safe private option
☑ Retain patient characteristics
☐ Clean description option
☐ Retain UIDs option
☑ Retain long full dates option
☐ Clean structured continued option
☑ Retain device identity option
☐ Retain long modified dates option
☐ Clean graph option

Figure 30 - Management of advanced conflicts options

Source: author's source

6.3.3. Displaying attributes

All attributes are then displayed in tree view, they are grouped by modules. All have an icon corresponding to the applied profile (standard, advanced options or custom) and a color corresponding to a rule of anonymization.



Requested procedure

🟥 (0040,1011) - Intended Recipients of Results Identification Sequence	K	X	D	Z	U	C
★ (0040,1004) - Patient Transport Arrangements	K	X	D	Z	U	C
🟥 (0040,1005) - Requested Procedure Location	K	X	D	Z	U	C
(0008,0020) - Study Date	K	X	D	Z	U	C
(0008,0030) - Study Time	K	X	D	Z	U	C

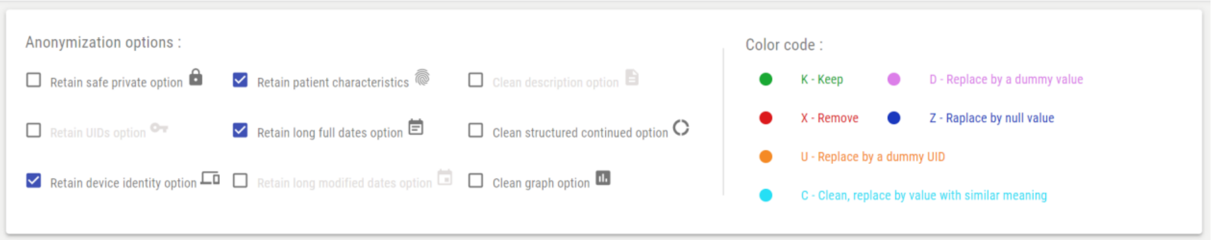
Figure 31 - Displaying attributes in treeview

Source: author's source

For example, in figure 26, the attribute (0040,1011) will be deleted (red color) and this rule correspond to the basic profile (square icon). The attribute (0040, 1004) will be replaced by a null value (blue color) and this rule correspond to a customization of a user. The attribute (0008,0020) will be keep (red color) and this rule correspond to "Retain long full dates option" (calendar icon).

6.3.4. Settings

To help the user it is possible to display a settings section. Users can change advanced options, see icons corresponding to rules and see which colors match with which rules.



Anonymization options :

☐ Retain safe private option
☑ Retain patient characteristics
☐ Clean description option
☐ Retain UIDs option
☑ Retain long full dates option
☐ Clean structured continued option
☑ Retain device identity option
☐ Retain long modified dates option
☐ Clean graph option

Color code :

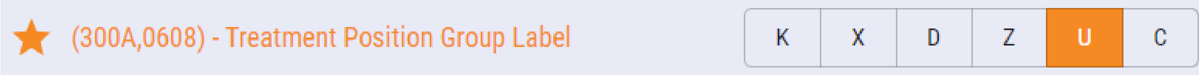
● K - Keep
● D - Replace by a dummy value
● X - Remove
● Z - Replace by null value
● U - Replace by a dummy UID
● C - Clean, replace by value with similar meaning

Figure 32 - Settings box

Source: author's source

6.3.5. Apply a custom rule for an attribute

Each attribute is customizable with a buttons group. Clicking on one of them will automatically change the anonymization rule.



★ (300A,0608) - Treatment Position Group Label

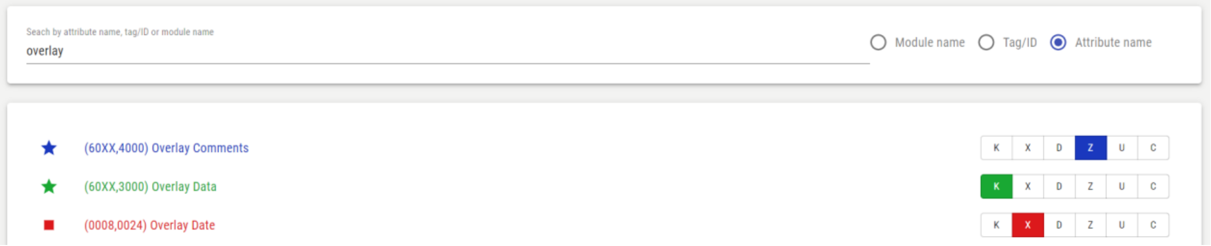
Figure 33 - Custom attributes

Source: author's source

For example, in figure 28, the field will be replaced by a dummy UID.

6.3.6. Search by module, id or attribute name

With a search bar, users can search an attribute by module, tag or name. Once found, users can apply custom rules to this attribute.



Seach by attribute name, tag/ID or module name: overlay

○ Module name ○ Tag/ID ● Attribute name

★ (60XX,4000) Overlay Comments

★ (60XX,3000) Overlay Data

■ (0008,0024) Overlay Date

Figure 34 - Search bar

Source: author's source

6.3.7. Exportation of a profile

In order to save anonymization profile, users can export it in JSON format. A name must be given, and the profile is saved on the user's computer.

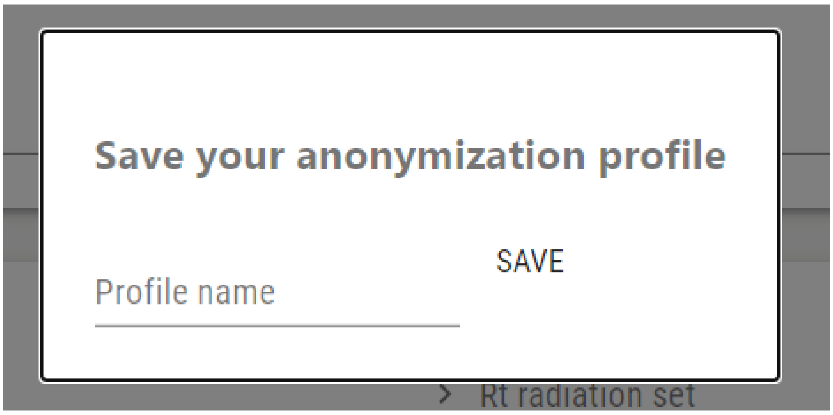
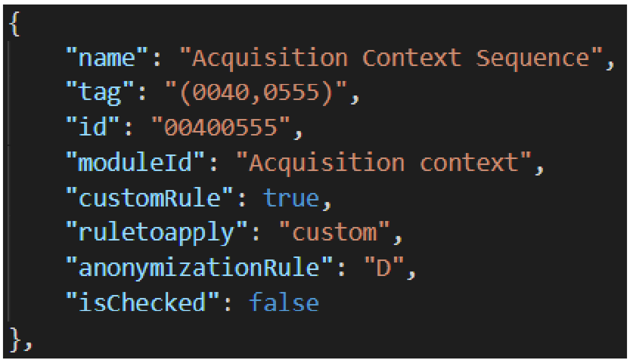


Figure 35 - Profile export form

Source: author's source

The exported JSON file is on the following form:



```
{
    "name": "Acquisition Context Sequence",
    "tag": "(0040,0555)",
    "id": "00400555",
    "moduleId": "Acquisition context",
    "customRule": true,
    "ruletoapply": "custom",
    "anonymizationRule": "D",
    "isChecked": false
},
```

Figure 36 - Structure of an attribute in exported JSON file

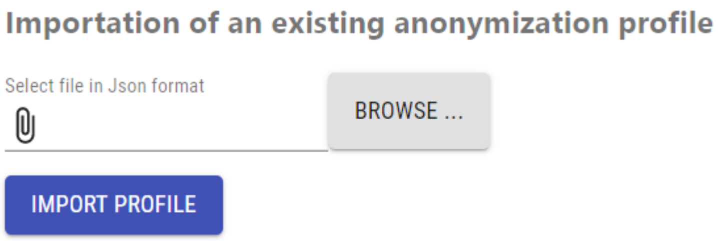
Source: author's source

In the exported file an attribute has some additional properties:

- The module ID is directly included in the object. No need to get the information in 'module_to_attribute.json' file.
- 'customRule' to specify if the rule is customized by the user.
- 'ruleToApply' to specify the name of the profile applied.
- 'anonymizationRule' to specify the rule applied.
- 'isChecked' to specify if the rule applied comes from advanced option or not.

6.3.8. Importation of an existing profile

If the user wants to edit an existing profile, he can import it into the application. The imported file must be in JSON format.



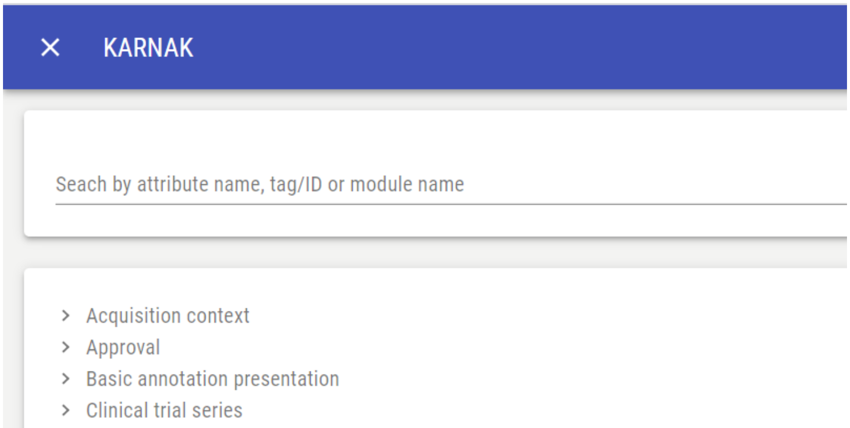
Importation of an existing anonymization profile

Figure 37 - Profile import form

Source: author's source

6.3.9. Reset the application

In the case of users wanting to restart and cancel all changes, they can click on the cross in the upper left corner. He will be able to create a new profile or import an existing one.



×
KARNAK
Seach by attribute name, tag/ID or module name
Acquisition context
Approval
Basic annotation presentation
Clinical trial series

Figure 38 - Cross to reset the application

Source: author's source

6.4. Deployment

The application has been deployed with Firebase. It is accessible with the following link: https://karnak-f9bce.web.app/. A repository named "build" has been created with a production build of the prototype inside. After Firebase project has been created and initialize in the environment. With the command line 'firebase deploy', the prototype has been posted online.

7. Project management

7.1. General information

To facilitate the development of the application the agile framework Scrum was used. Scrum is "A framework within which people can address complex adaptive problems, while productively and creatively delivering products of the highest possible value." (Scrum.org, 2020).

The use of scrum made it possible to divide the development into sprints. In this project, sprint is a period of five days (from Wednesday to next Tuesday without the weekend), have a short period for sprint to enable regular feedback from the client. At the beginning, of each sprint, user stories were selected in the product backlog (annex V). The goal was to have a functional increment of the application at the end of each sprint. Each sprint has an objective defined at its beginning which will allow a step-by-step development.

The product backlog is "an ordered list of everything that is known to be needed in the product" (Scrum.org, 2020). The support for the product backlog was an excel sheet.

There are several roles in Scrum:

- The product owner is responsible for the product backlog.
- The development team develops the product, transforms items from the product backlog in releasable increment.
- The scrum master helps the people involved in Scrum to understand and apply better this framework.

The author embodied each of these roles.

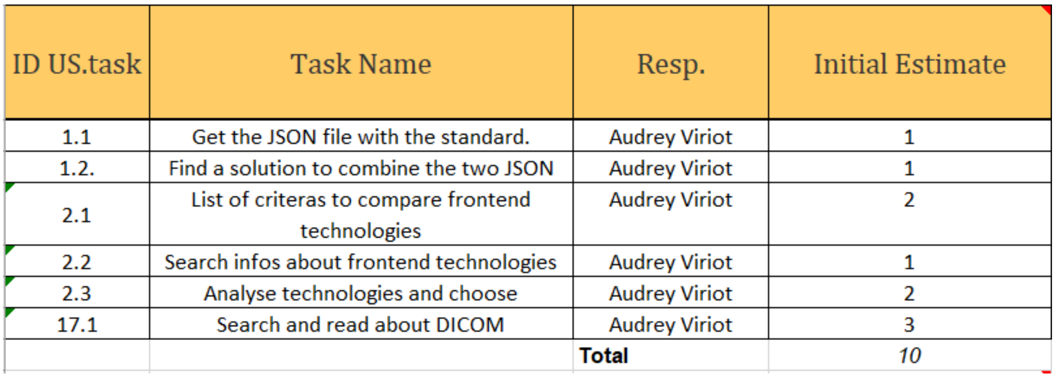
7.2. Sprint 0

The goal of the first sprint was to get all the necessary resources to be ready to code. There was ten story points for this sprint (annex VI).

User story number	User story	Story points
2	As a developer I want to determine which frontend technology I will use for the prototype so that I can code.	5
17	As a developer I want to understand DICOM so that I can develop a suitable prototype.	3
1	As a developer I want to get data or access to backend, so that I can link the frontend with data.	2

Table 8 - User stories for sprint 0

Source: author's source



ID US.task	Task Name	Resp.	Initial Estimate
1.1	Get the JSON file with the standard.	Audrey Viriot	1
1.2.	Find a solution to combine the two JSON	Audrey Viriot	1
2.1	List of criteras to compare frontend technologies	Audrey Viriot	2
2.2	Search infos about frontend technologies	Audrey Viriot	1
2.3	Analyse technologies and choose	Audrey Viriot	2
17.1	Search and read about DICOM	Audrey Viriot	3
		Total	*10*

Figure 39 - List of tasks for sprint 0

Source: author's source

7.3. Sprint 1

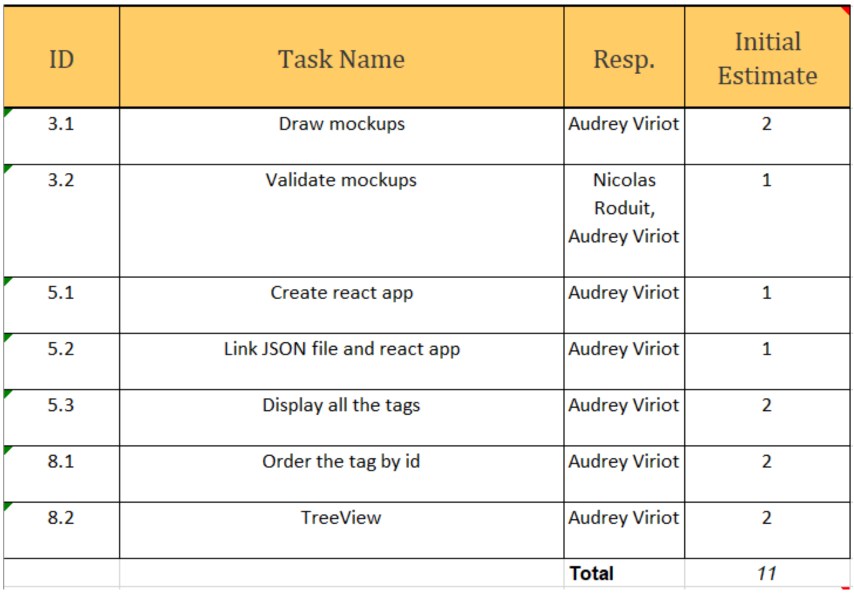
The goal of the second sprint was to get a working development environment with attributes displayed in a list. There were eleven story points for this sprint (annex VII).

The main difficulty in this sprint was to group attributes by module. In a fist time, attributes were grouped by group ID (data group), however "a data group does not convey any semantic meaning beginning with DICOM Version 3.0." (DICOM, 2020). The JSON file 'module_to_attributes.json' was added as a resources to the project to group attributes by modules.

User story number	User storty	Story points
5	As a developer I want to list all the attributes, so that I can see all the attributes I want to anonymize.	3
3	As a developer I want to draw mockups so that I can have a better idea of the UI.	3
8	As a user I want to group attributes by module ID so that I can have a better view of all the attributes	5

Table 9 - User stories for sprint 1

Source: author's source



ID	Task Name	Resp.	Initial Estimate
3.1	Draw mockups	Audrey Viriot	2
3.2	Validate mockups	Nicolas Roduit, Audrey Viriot	1
5.1	Create react app	Audrey Viriot	1
5.2	Link JSON file and react app	Audrey Viriot	1
5.3	Display all the tags	Audrey Viriot	2
8.1	Order the tag by id	Audrey Viriot	2
8.2	TreeView	Audrey Viriot	2
		Total	*11*

Figure 40 - List of tasks for sprint 1

Source: author's source

7.4. Sprint 2

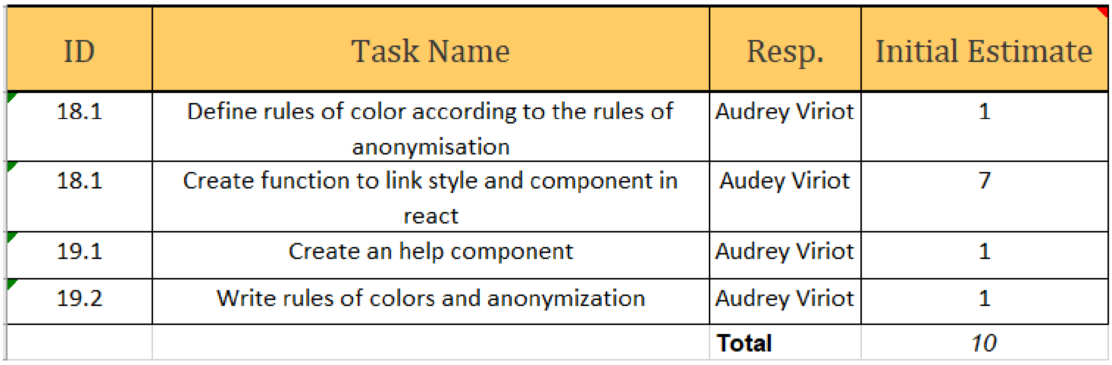
The goal of this sprint was to display attributes in color according to their anonymization rules. There were ten storys point for this sprint (annex VIII).

In order to manage colors in a simple way, the JSON file 'rules_colors.json' was created to centralize the color management. Some anonymization rules have different format than others, for example 'X/Z'. After a meeting with Karnak team and Nicolas Roduit it was decided that in case more than one rule are possible for an attribute, the last one is retained.

User story number	User story	Story points
18	As a user I want to see colored attributes according to anonymization rules so that I can see clearly which tags is concerned by which rules.	8
19	As a user I want to get some help so that I can understand the different rules and colors.	2

Table 10 - User stories for sprint 2

Source: author's source



ID	Task Name	Resp.	Initial Estimate
18.1	Define rules of color according to the rules of anonymisation	Audrey Viriot	1
18.1	Create function to link style and component in react	Audey Viriot	7
19.1	Create an help component	Audrey Viriot	1
19.2	Write rules of colors and anonymization	Audrey Viriot	1
		Total	*10*

Figure 41 - List of tasks for sprint 2

Source: author's source

7.5. Sprint 3

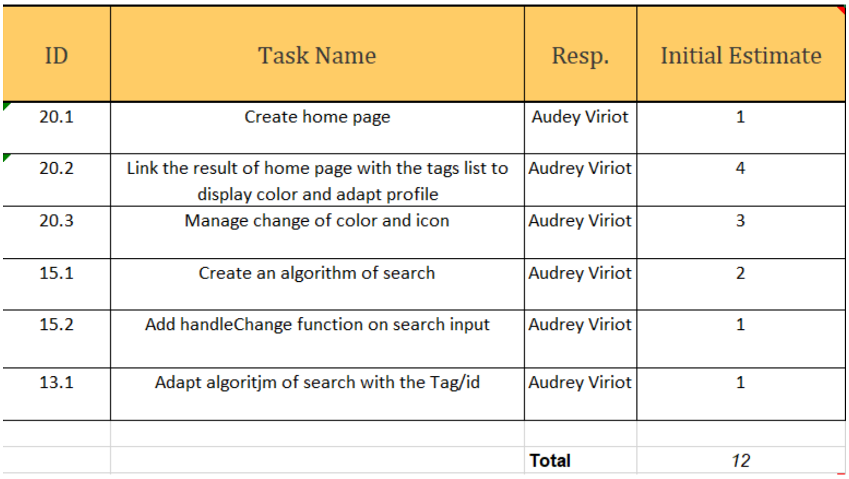
The goal of this sprint was navigating in the application, search for attributes and add advanced options. There were twelve story point for this sprint (annex IX).

Sometimes an attribute can have different advanced options, to avoid problem a function to detect conflict has been created (annex III). If an advanced option is ticked and can create a conflict on at least one attribute, the second advanced option affected by the conflict is disabled.

User story number	User story	Story points
20	As a user I want to choose if I use standard profile or another option so that I can configure my profile as I want.	8
13	As a user I want to search an attribute by his ID so that I can edit it.	3
15	As a user I want to search through a group/category of attributes so that I can see all the tags related to this category.	1

Table 11 - User stories for sprint 3

Source: author's source



ID	Task Name	Resp.	Initial Estimate
20.1	Create home page	Audey Viriot	1
20.2	Link the result of home page with the tags list to display color and adapt profile	Audrey Viriot	4
20.3	Manage change of color and icon	Audrey Viriot	3
15.1	Create an algorithm of search	Audrey Viriot	2
15.2	Add handleChange function on search input	Audrey Viriot	1
13.1	Adapt algoritjm of search with the Tag/id	Audrey Viriot	1
		Total	*12*

Figure 42 - List of tasks for sprint 3

Source: author's source

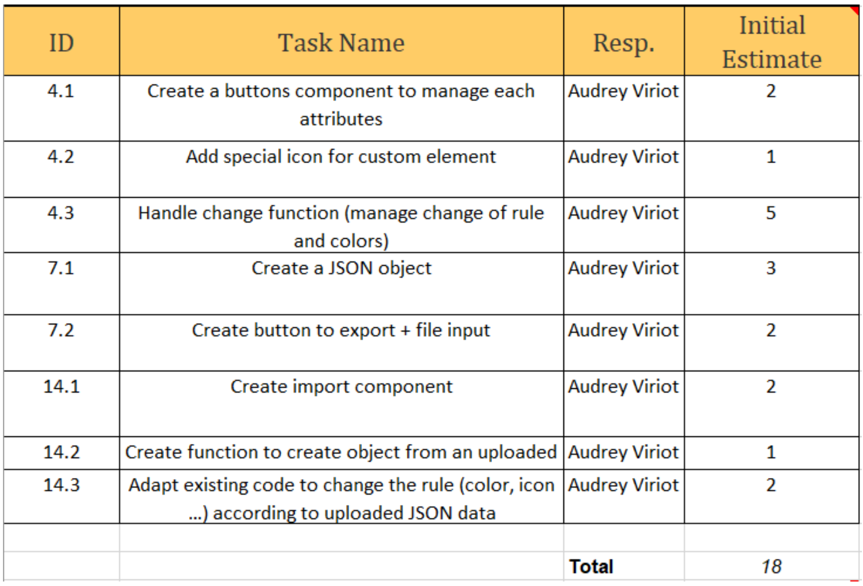
7.6. Sprint 4

The goal of this sprint was to add custom anonymization rules to attribute and export/import a profile of anonymization. There were eighteen story points for this sprint (annex X). The last sprint was a little more consistent, but it was finished in time. The import and export of JSON file have been done quickly but, the customization of attribute's rule has been more complex. Many conditions have been added to avoid bugs (annex IV).

User story number	User story	Story points
4	As a user I want to choose which attribute I want to anonymize so that I can edit the profile like I want.	8
7	As a user I want to export my profile in JSON so that I can save my profile.	5
14	As a user I want to import an existing profile in JSON format so that I can work with an existing profile.	5

Table 12 - User stories for sprint 4

Source: author's source



ID	Task Name	Resp.	Initial Estimate
4.1	Create a buttons component to manage each attributes	Audrey Viriot	2
4.2	Add special icon for custom element	Audrey Viriot	1
4.3	Handle change function (manage change of rule and colors)	Audrey Viriot	5
7.1	Create a JSON object	Audrey Viriot	3
7.2	Create button to export + file input	Audrey Viriot	2
14.1	Create import component	Audrey Viriot	2
14.2	Create function to create object from an uploaded	Audrey Viriot	1
14.3	Adapt existing code to change the rule (color, icon ...) according to uploaded JSON data	Audrey Viriot	2
		Total	*18*

Figure 43 - List of tasks for sprint 4

Source: author's source

8. Conclusion

8.1. Results

This work confirms that a web interface to create anonymization profiles for DICOM files is feasible. A working prototype has been developed; it allows the creation of anonymization profiles based on the DICOM standard. It produces a JSON file including all the necessary data for the anonymization of DICOM files. As a researcher, the use of this tool allows to customize each attribute according to the needs of the medical research project. After, the user can then save his profile and share it with his collaborators.

8.2. Recommendation for further action

For future development, it will be desirable to have access to the backend (for the moment under development). In this way, the web interface will communicate directly with a database and algorithms of anonymization like explained in 6.1.

In the case of a larger dataset (more attributes), the use of asynchronous functions could avoid performance problems (for example long data loading). In addition, it will be necessary to focus on the way the attributes are grouped together. An optimization of the modules could be envisaged, for a smoother and easier navigation.

To ensure the protection of patient's data, attribute like patient's name for example must not be editable, even if this attribute is present in the dataset.

Features can be added, like the possibility to display all the customized attributes or all the attributes affected by an advanced option.

8.3. Looking to the future

Today more and more data are being collected in many areas. In this work the personalization of the anonymization profile concerned data intended for medical research. It is quite possible to adapt this solution for areas of research that do not use a DICOM file.

9. Declaration on honour

I hereby declare that I have carried out this final research project on my own without any help other than the references listed in the list of references and that I have only used the sources mentioned.I will not provide a copy of this paper to a third party without the permission of the department head and of my advisor, including the partner company with which I collaborated on this project, with the exception of those who provided me with information needed to write this paper and whose names follow:

- Adrien DEPEURSINGE
- Roger SCHAER
- Nicolas RODUIT and his team

Audrey VIRIOT (July 2020)

10. References

Benitte, R., Greif, S., & Rambeau, M. (2019). *Front End Frameworks*. Retrieved from StateOfJS: https://2019.stateofjs.com/front-end-frameworks/

Dabbs, M. (2019, July 29). *The Fundamentals of Web Application Architecture*. Retrieved from Reinvently: whttps://reinvently.com/blog/fundamentals-web-application-architecture/

DICOM. (2020). *E Attribute Confidentiality Profiles*. Retrieved from DICOM: http://dicom.nema.org/medical/dicom/current/output/chtml/part15/chapter_E.html

DICOM. (2020). *Home*. Retrieved from DICOM: https://www.dicomstandard.org/

DICOM. (2020). *The Data Set*. Retrieved from DICOM: http://dicom.nema.org/dicom/2013/output/chtml/part05/chapter_7.html

Firbase. (2020). *Products*. Récupéré sur Firebase: https://firebase.google.com/products

Innolitics. (2020). *innolitics/dicom-standard*. Récupéré sur Github: https://github.com/innolitics/dicom-standard

Kalliamvakou, E., Singer, L., Gousios, G., German, D., Blincoe, K., & Daniela, D. (2014). *The promises and perils of mining GitHub*. Retrieved from https://dl.acm.org/doi/pdf/10.1145/2597073.2597074

Material-UI. (2020). *About Us*. Récupéré sur Material-UI: https://material-ui.com/company/about/

Medigift. (n.d.). *IMAGINE*. Retrieved from Medigift: http://medgift.hevs.ch/wordpress/projects/imagine/

Miller, R. (2014). Big Data Curation. *20th International Conference on Management of Data (COMAD)*. Hyderabad, India.

Monteiro, E., Costa, C., & Oliveira, J. (2017, April 13). A De-Identification Pipeline for Ultrasound Medical Imagesin DICOM Format. *Springer Science+Business Media*.

OsiriX-Foundation. (2020). *OsiriX-Foundation/Karnac*. Retrieved from Github: https://github.com/OsiriX-Foundation/Karnak

Pianykh, O. (2008). *Digital Imaging and Communications in Medecine (DICOM)*. Boston (USA): Springer.

Scrum.org. (2020). *The Scrum Guide*. Retrieved from scrumguides.com: https://www.scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-US.pdf

Stackoverflow. (2020). *Stackoverflow Insights - Trends*. Récupéré sur Stackoverflow: https://insights.stackoverflow.com/trends?tags=vue.js%2Cangular%2Creactjs

Swiss federal authorities. (2014, January 1). Federal Act on Research involving Human Beings. *Chapter 4 Further Use of Biological Material and Health-Related Personal Data for Research*. Bern, Switzerland.

U.S Food and Drug Administration. (2018, 08 28). *Medical Imaging*. Retrieved from U.S Food and Drug Administration: https://www.fda.gov/radiation-emitting-products/radiation-emitting-products-and-procedures/medical-imaging

Unité de Data Science & Recherche (CHUV). (n.d.). *Instructions CHUV : Dépersonnalisation / codification des données patients structurées et textuelles pour les projets de recherche clinique*. Lausanne.

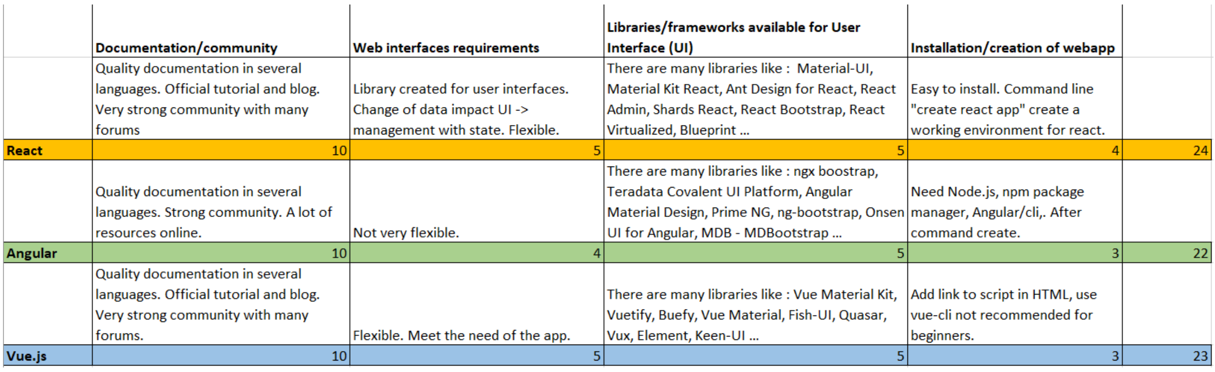
Web Application Security Consortium. (2005). *Glossary*. Retrieved from Web Application Security Consortium: http://www.webappsec.org/projects/glossary/

Wohlgethan, E. (2018, May 15). Supporting Web Development Decisions byComparing Three Major JavaScript Frameworks:Angular, React and Vue.js. Hamburg.

Zeste de savoir. (2019, August 12). Développer et publier une app AndroidMaterial Design.

12. Annexes

I. Details of the comparative study of frontend framework

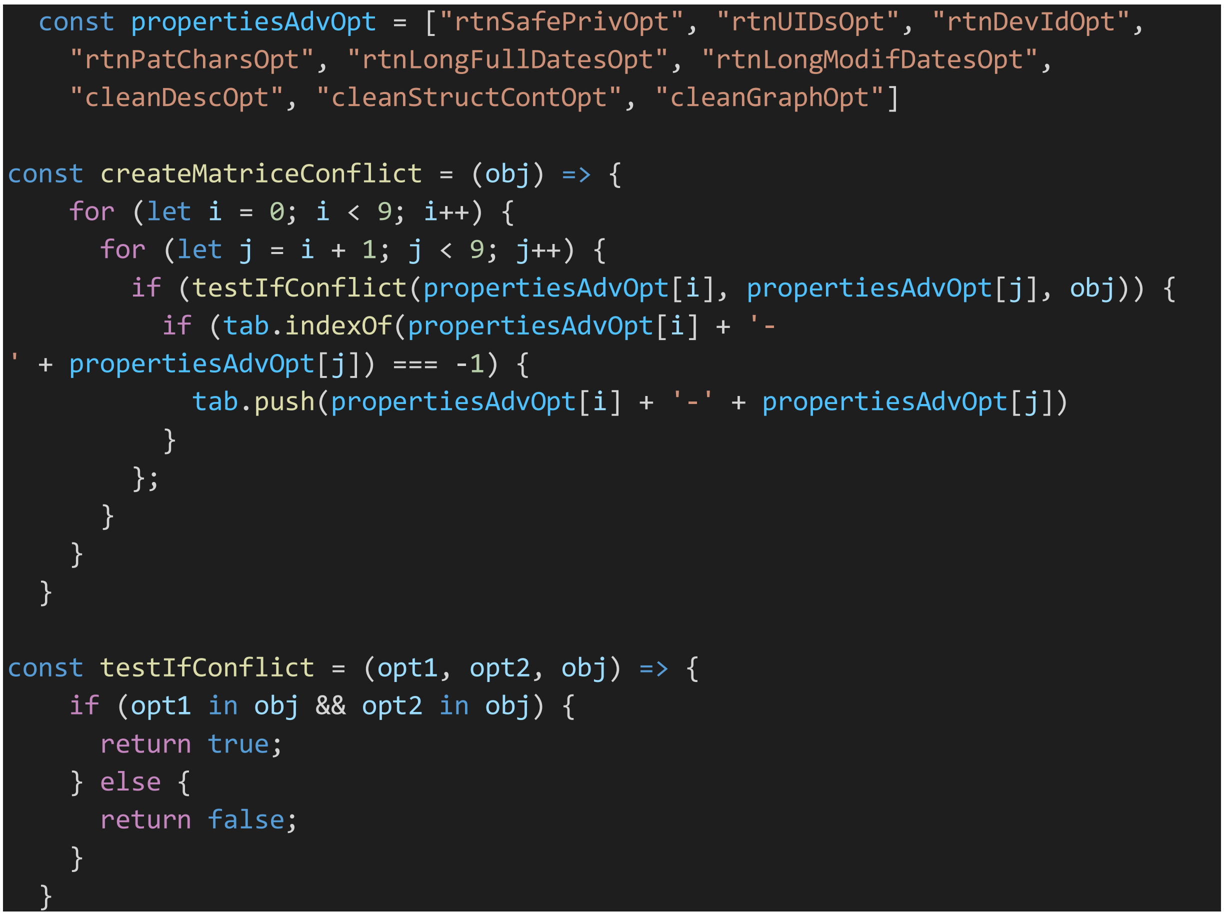


	Documentation/community	Web interfaces requirements	Libraries/frameworks available for User Interface (UI)	Installation/creation of webapp	
	Quality documentation in several languages. Official tutorial and blog. Very strong community with many forums	Library created for user interfaces. Change of data impact UI -> management with state. Flexible.	There are many libraries like : Material-UI, Material Kit React, Ant Design for React, React Admin, Shards React, React Bootstrap, React Virtualized, Blueprint ...	Easy to install. Command line "create react app" create a working environment for react.	
React	10	5	5	4	24
	Quality documentation in several languages. Strong community. A lot of resources online.	Not very flexible.	There are many libraries like : ngx boostrap, Teradata Covalent UI Platform, Angular Material Design, Prime NG, ng-bootstrap, Onsen UI for Angular, MDB - MDBootstrap ...	Need Node.js, npm package manager, Angular/cli,. After command create.	
Angular	10	4	5	3	22
	Quality documentation in several languages. Official tutorial and blog. Very strong community with many forums.	Flexible. Meet the need of the app.	There are many libraries like : Vue Material Kit, Vuetify, Buefy, Vue Material, Fish-UI, Quasar, Vux, Element, Keen-UI ...	Add link to script in HTML, use vue-cli not recommended for beginners.	
Vue.js	10	5	5	3	23

II. JavaScript function to merge two JSON files together

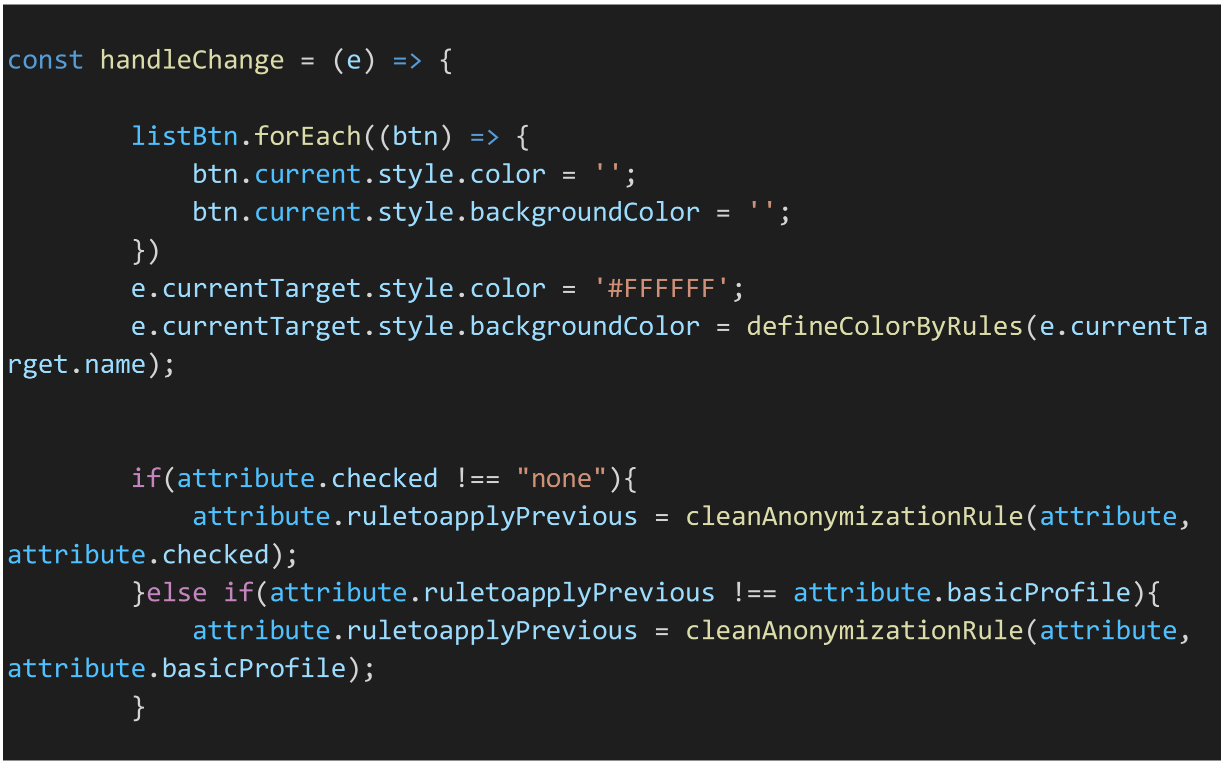
```
const attributes =  require('../Resources/confidentiality_profile_attributes.j
son');
const modules = require('../Resources/module_to_attributes.json');
const fs = require('fs');

attributes.map((attribute, key) => {

    modules.map((module, key) => {
        if (attribute.tag === module.tag) {
            var newStr = module.moduleId.split('-').join(' ');
            attribute.moduleId = newStr.charAt(0).toUpperCase() + newStr.subst
r(1);
        }
        return true;
    })

    if(attribute.moduleId === undefined){
        attribute.moduleId = "Other";
    }

})

var json = JSON.stringify(attributes);
fs.writeFile('../Resources/attributes_modules_linked.json', json, 'utf8', (res
, err) => {
    if(err){
        console.log("Error : "+err);
    }
});
```

III. Function to detect conflict in attributes

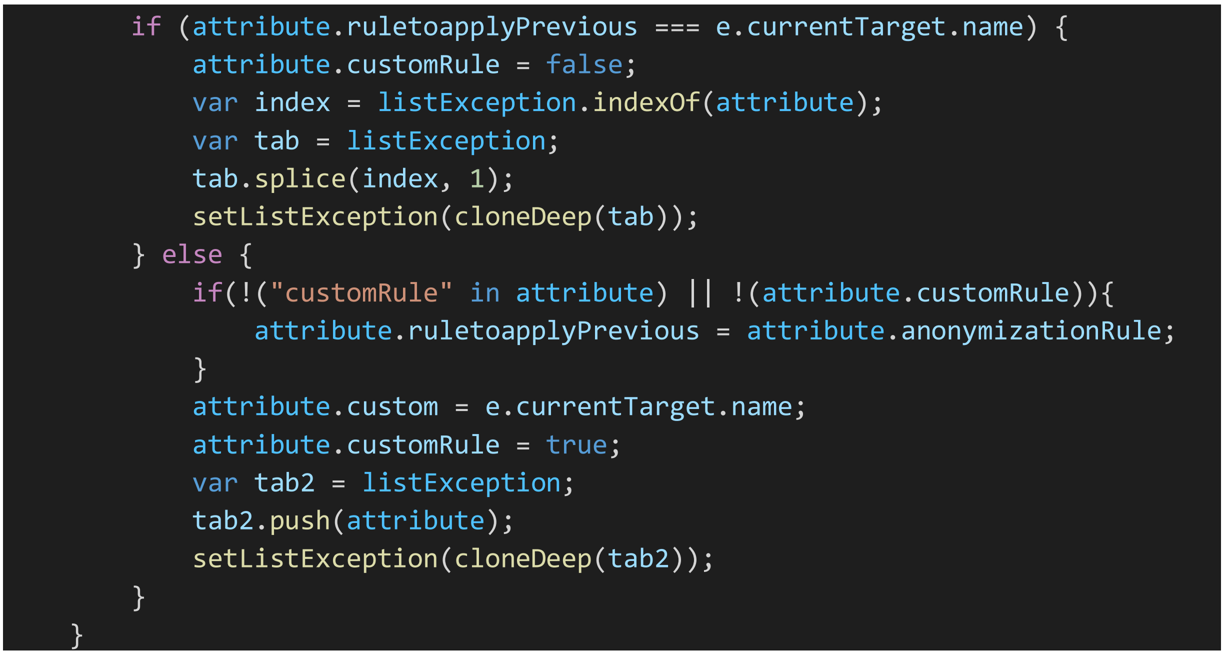


```
 const propertiesAdvOpt = ["rtnSafePrivOpt", "rtnUIDsOpt", "rtnDevIdOpt",
  "rtnPatCharsOpt", "rtnLongFullDatesOpt", "rtnLongModifDatesOpt",
  "cleanDescOpt", "cleanStructContOpt", "cleanGraphOpt"]

const createMatriceConflict = (obj) => {
  for (let i = 0; i < 9; i++) {
   for (let j = i + 1; j < 9; j++) {
    if (testIfConflict(propertiesAdvOpt[i], propertiesAdvOpt[j], obj)) {
     if (tab.indexOf(propertiesAdvOpt[i] + '-
' + propertiesAdvOpt[j]) === -1) {
      tab.push(propertiesAdvOpt[i] + '-' + propertiesAdvOpt[j])
     }
    };
   }
  }
 }

const testIfConflict = (opt1, opt2, obj) => {
  if (opt1 in obj && opt2 in obj) {
   return true;
  } else {
   return false;
  }
 }
```

IV. Function ‘handleChange’ to manage bugs

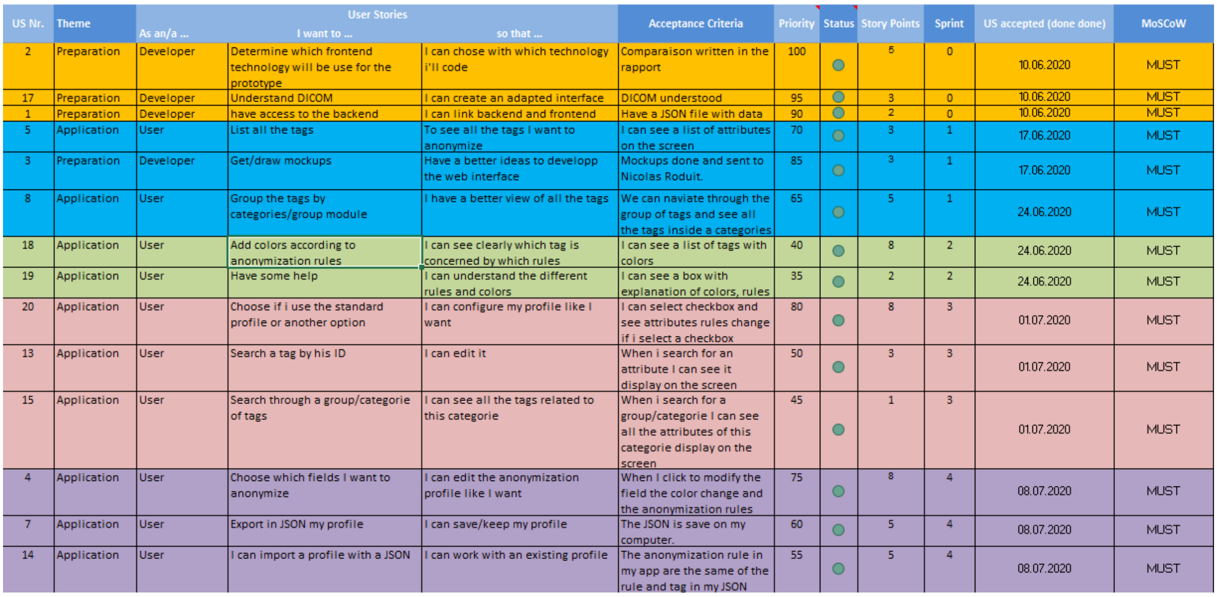


```
const handleChange = (e) => {

    listBtn.forEach((btn) => {
      btn.current.style.color = '';
      btn.current.style.backgroundColor = '';
    })
    e.currentTarget.style.color = '#FFFFFF';
    e.currentTarget.style.backgroundColor = defineColorByRules(e.currentTa
rget.name);


    if(attribute.checked !== "none"){
      attribute.ruletoapplyPrevious = cleanAnonymizationRule(attribute,
attribute.checked);
    }else if(attribute.ruletoapplyPrevious !== attribute.basicProfile){
      attribute.ruletoapplyPrevious = cleanAnonymizationRule(attribute,
attribute.basicProfile);
    }
```



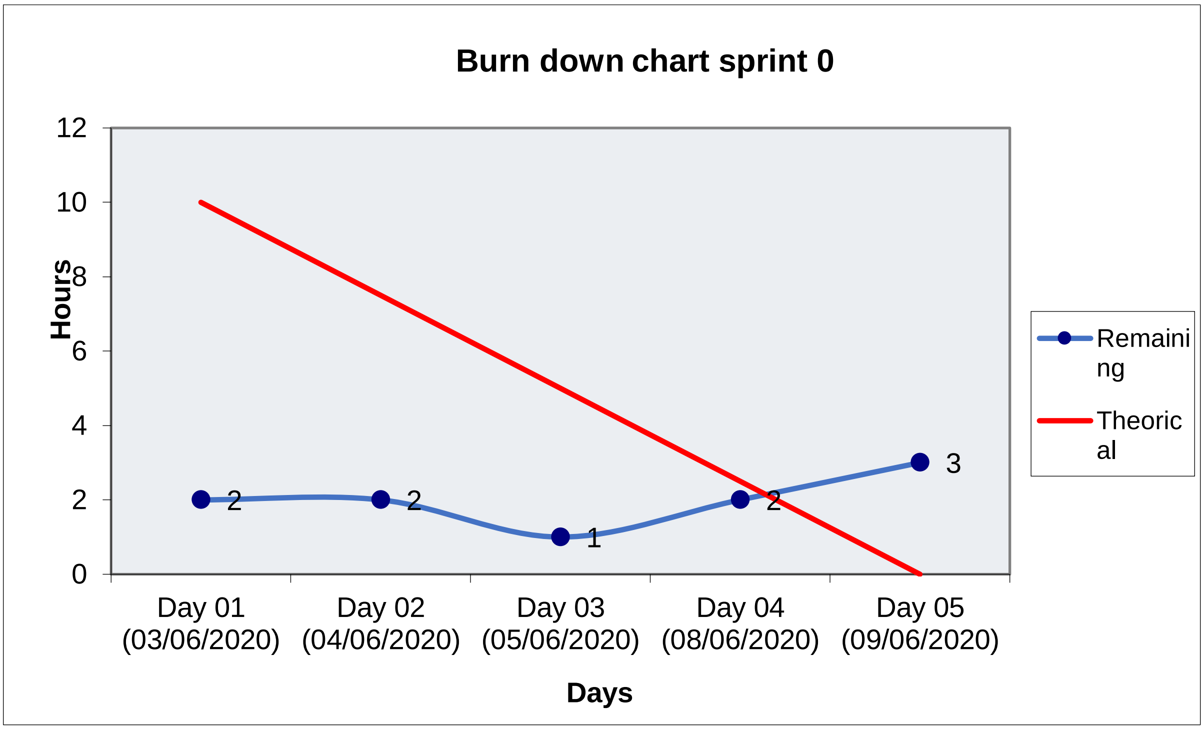
```
    if (attribute.ruletoapplyPrevious === e.currentTarget.name) {
        attribute.customRule = false;
        var index = listException.indexOf(attribute);
        var tab = listException;
        tab.splice(index, 1);
        setListException(cloneDeep(tab));
    } else {
        if(!("customRule" in attribute) || !(attribute.customRule)){
            attribute.ruletoapplyPrevious = attribute.anonymizationRule;
        }
        attribute.custom = e.currentTarget.name;
        attribute.customRule = true;
        var tab2 = listException;
        tab2.push(attribute);
        setListException(cloneDeep(tab2));
    }
}
```

V. Product backlog

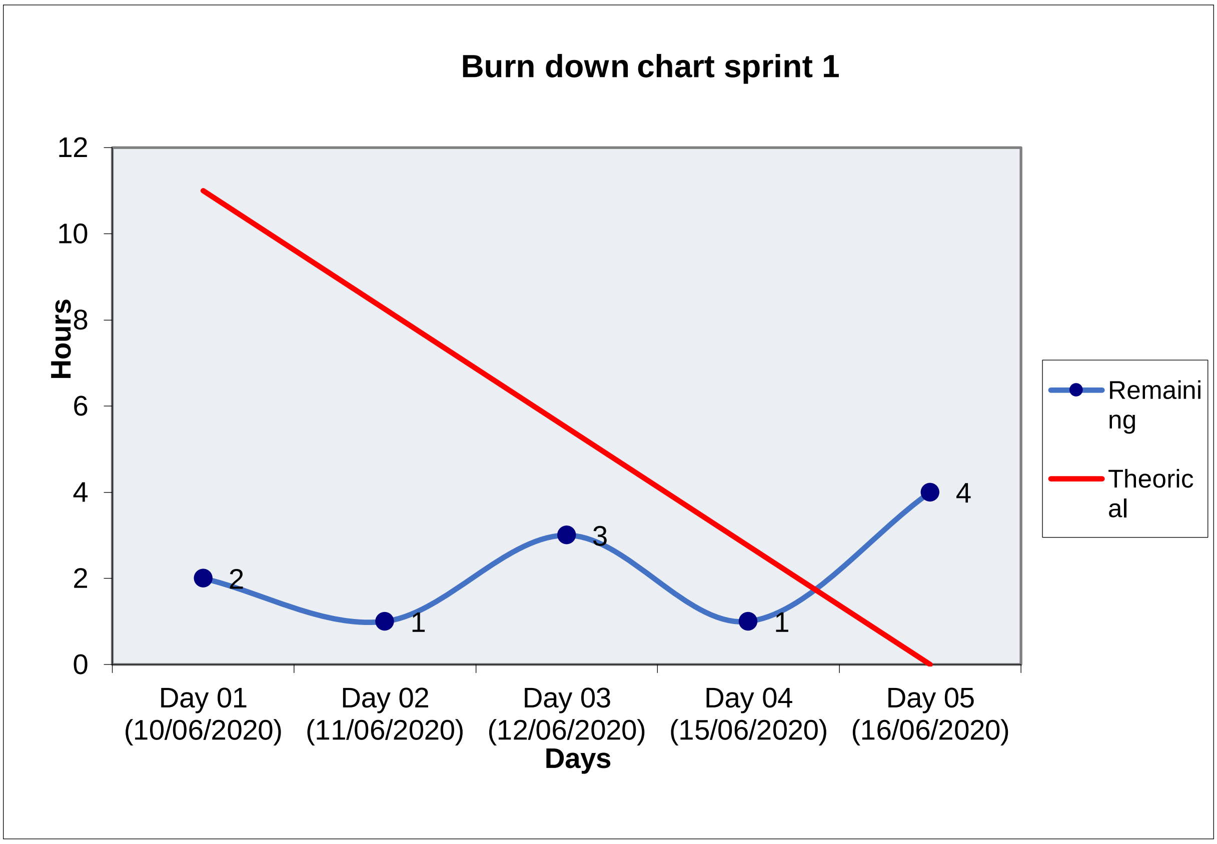


US Nr.	Theme	User Stories			Acceptance Criteria	Priority	Status	Story Points	Sprint	US accepted (done done)	MoSCoW
		As an/a ...	I want to ...	so that ...							
2	Preparation	Developer	Determine which frontend technology will be use for the prototype	I can chose with which technology i'll code	Comparaison written in the rapport	100	●	5	0	10.06.2020	MUST
17	Preparation	Developer	Understand DICOM	I can create an adapted interface	DICOM understood	95	●	3	0	10.06.2020	MUST
1	Preparation	Developer	have access to the backend	I can link backend and frontend	Have a JSON file with data	90	●	2	0	10.06.2020	MUST
5	Application	User	List all the tags	To see all the tags I want to anonymize	I can see a list of attributes on the screen	70	●	3	1	17.06.2020	MUST
3	Preparation	Developer	Get/draw mockups	Have a better ideas to developp the web interface	Mockups done and sent to Nicolas Roduit.	85	●	3	1	17.06.2020	MUST
8	Application	User	Group the tags by categories/group module	I have a better view of all the tags	We can naviate through the group of tags and see all the tags inside a categories	65	●	5	1	24.06.2020	MUST
18	Application	User	Add colors according to anonymization rules	I can see clearly which tag is concerned by which rules	I can see a list of tags with colors	40	●	8	2	24.06.2020	MUST
19	Application	User	Have some help	I can understand the different rules and colors	I can see a box with explanation of colors, rules	35	●	2	2	24.06.2020	MUST
20	Application	User	Choose if i use the standard profile or another option	I can configure my profile like I want	I can select checkbox and see attributes rules change if i select a checkbox	80	●	8	3	01.07.2020	MUST
13	Application	User	Search a tag by his ID	I can edit it	When i search for an attribute I can see it display on the screen	50	●	3	3	01.07.2020	MUST
15	Application	User	Search through a group/categorie of tags	I can see all the tags related to this categorie	When i search for a group/categorie I can see all the attributes of this categorie display on the screen	45	●	1	3	01.07.2020	MUST
4	Application	User	Choose which fields I want to anonymize	I can edit the anonymization profile like I want	When I click to modify the field the color change and the anonymization rules	75	●	8	4	08.07.2020	MUST
7	Application	User	Export in JSON my profile	I can save/keep my profile	The JSON is save on my computer.	60	●	5	4	08.07.2020	MUST
14	Application	User	I can import a profile with a JSON	I can work with an existing profile	The anonymization rule in my app are the same of the rule and tag in my JSON	55	●	5	4	08.07.2020	MUST

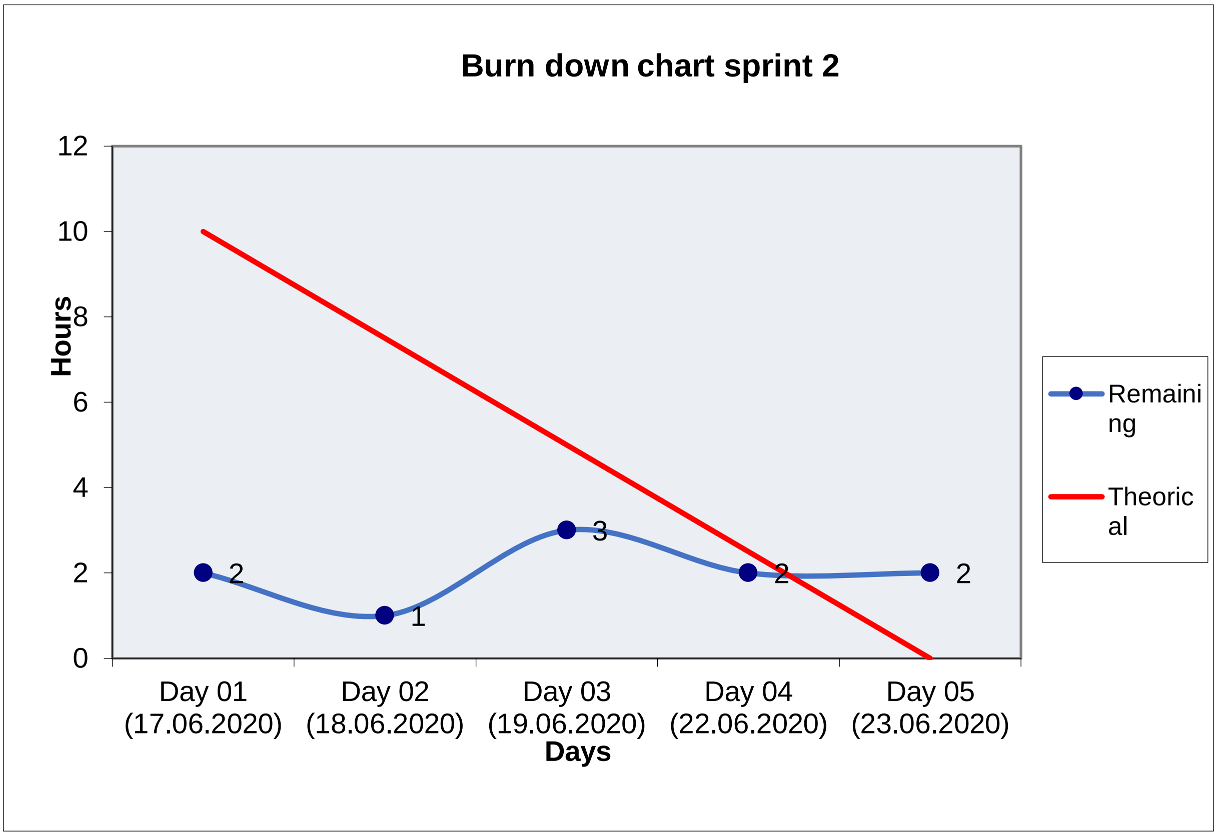
VI. Burn down chart sprint 0



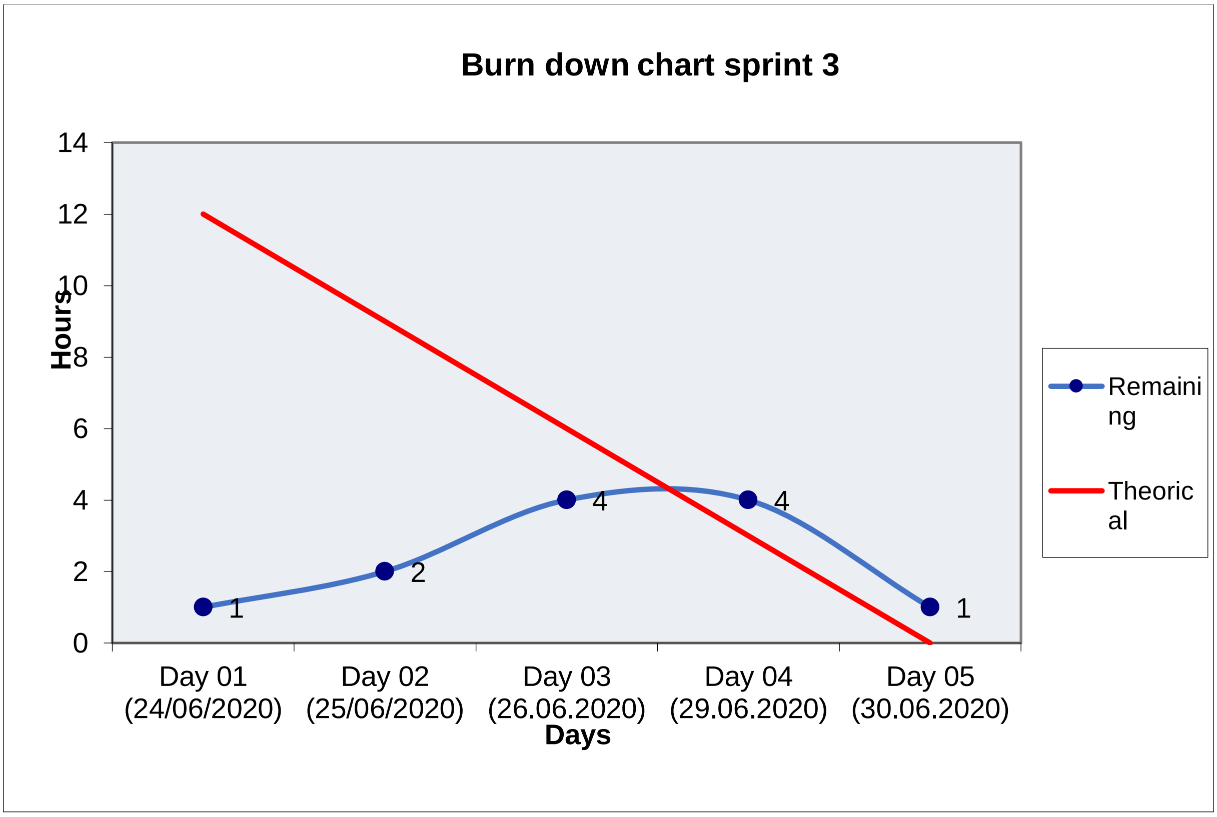
VII. Burn down chart sprint 1



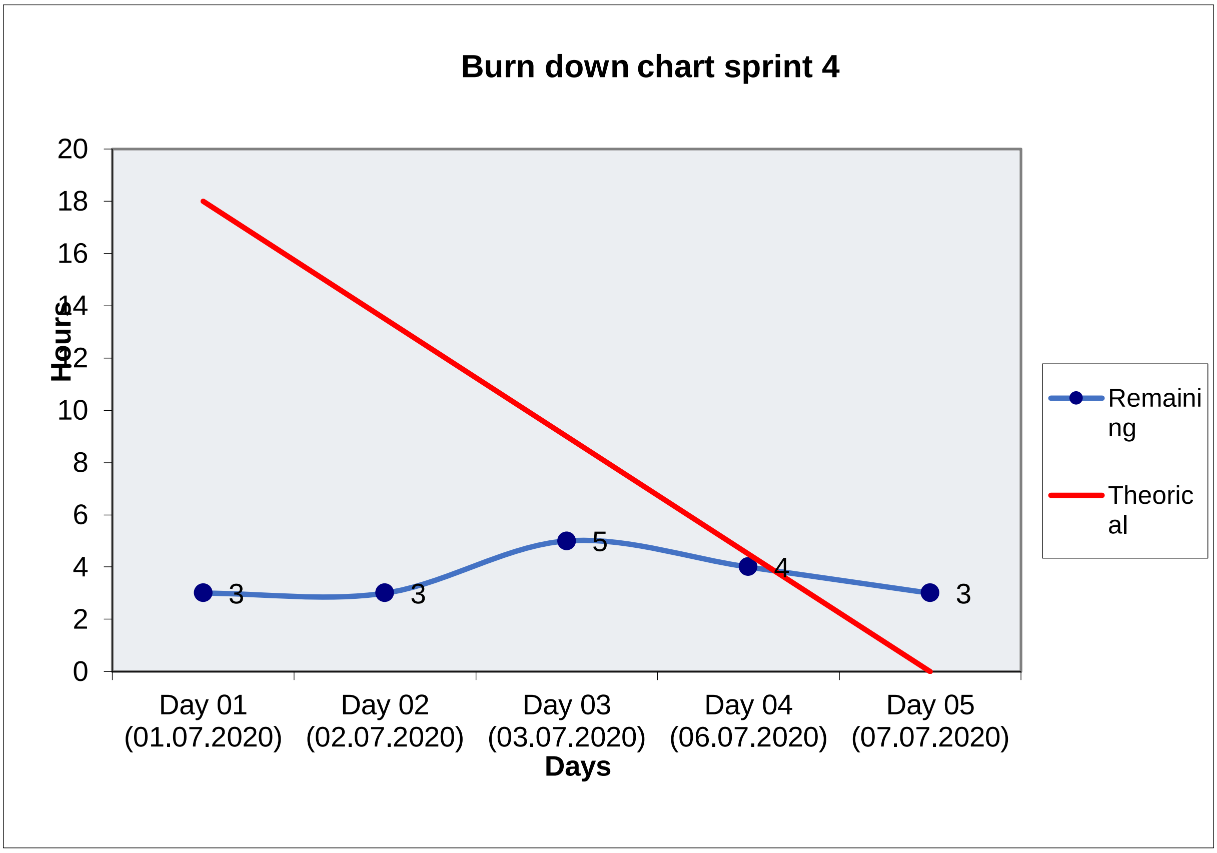
VIII. Burn down chart sprint 2



IX. Burn down chart sprint 3



X. Burn down chart sprint 4


Burn down chart sprint 4
Hours
20
18
16
14
12
10
8
6
4
2
0
3
3
5
4
3
Day 01 (01.07.2020)
Day 02 (02.07.2020)
Day 03 (03.07.2020)
Day 04 (06.07.2020)
Day 05 (07.07.2020)
Days
Remaining
Theorical