

## **Grau en Enginyeria Informàtica de Gestió i Sistemes d'Informació**

### **Anonimització de dades clíniques i proves diagnòstiques en un hospital**

### **Memòria**

**Xavier Sanchez de la Corte**

**TUTOR: Alfons Palacios**

2021-2022



## **Abstract**

This project is focused on developing an application that converts DICOM images and medical records that contains private information of the patients in documents free of personal information to make it possible to do research and study cases.

## **Resum**

Aquest treball se centra en el desenvolupament d'una aplicació que converteix les imatges DICOM i les històries clíniques que contenen informació privada dels pacients en documents lliures d'informació personal per tal de poder fer recerca i casos d'estudi.

## **Resumen**

Este trabajo se centra en el desarrollo de una aplicación que convierte las imágenes DICOM y las historias clínicas que contienen información privada de los pacientes en documentos libres de información personal con el objetivo de poder ser usados en investigación y casos de estudio para la docencia.

## Índex

Índex Figures	VI
Índex Taules	VIII
<b>1. Introducció</b>	<b>1</b>
<b>1.1 Motivació.....</b>	<b>1</b>
<b>2. Marc teòric</b>	<b>3</b>
<b>2.1 Context.....</b>	<b>3</b>
<b>3. Objectius i abast</b>	<b>5</b>
<b>3.1 Objectius del producte:.....</b>	<b>5</b>
<b>3.2 Objectius del client: .....</b>	<b>5</b>
<b>3.3 Target:.....</b>	<b>5</b>
<b>4. Anàlisi de referents</b>	<b>7</b>
<b>4.1 Eines i llibreries .....</b>	<b>7</b>
<b>4.1.1 ARX</b>	<b>7</b>
<b>4.1.2 AMNESIA</b>	<b>8</b>
<b>4.1.3 PYDICOM</b>	<b>8</b>
<b>4.1.4 PyPDF2</b>	<b>8</b>
<b>4.1.5 NLP</b>	<b>8</b>
<b>4.2 Estàndards .....</b>	<b>9</b>
<b>4.2.1 OMOP</b>	<b>9</b>
<b>4.2.2 SNOMED CT</b>	<b>9</b>
<b>4.2.3 FHIR</b>	<b>10</b>
<b>4.3 Serveis .....</b>	<b>10</b>
<b>4.3.1 NYMIZ</b>	<b>10</b>
<b>5. Metodologia</b>	<b>11</b>
<b>5.1 Definició del projecte .....</b>	<b>11</b>
<b>5.2 Adquisició de coneixements .....</b>	<b>11</b>
<b>5.3 Producció .....</b>	<b>12</b>
<b>5.4 Testing.....</b>	<b>13</b>
<b>6. Desenvolupament</b>	<b>15</b>
<b>6.1 Anàlisi i definició de requeriments.....</b>	<b>15</b>
<b>6.2 Casos d'ús del sistema.....</b>	<b>16</b>
<b>6.3 Disseny del software.....</b>	<b>26</b>
<b>6.4 Llibreries utilitzades .....</b>	<b>27</b>
<b>6.4.1 TKINTER</b>	<b>27</b>

---

<b>6.4.2 PDFPLUMBER</b>	<b>27</b>
<b>6.4.3 RE</b>	<b>27</b>
<b>6.4.4 ASPOSE.WORDS</b>	<b>27</b>
<b>6.4.5 OS</b>	<b>28</b>
<b>6.4.6 UUID</b>	<b>28</b>
<b>6.4.7 PYDICOM</b>	<b>28</b>
<b>6.5 Documentació i escalabilitat del codi.....</b>	<b>29</b>
<b>6.6 Solució creada .....</b>	<b>32</b>
<b>6.6.1 Pseudo-anonimitzar expedient</b>	<b>33</b>
<b>6.6.2 Anonimitzar DICOM</b>	<b>38</b>
<b>6.6.3 Anonimitzar carpeta d'imatges DICOM</b>	<b>42</b>
<b>6.6.4 Anonimitzar història clínica</b>	<b>45</b>
<b>6.6.5 Anonimitzar carpeta d'històries clíniques</b>	<b>49</b>
<b>7. Conclusió</b>	<b>53</b>
<b>7.1 Anàlisi de resultats.....</b>	<b>53</b>
<b>7.2 Aspectes a millorar .....</b>	<b>53</b>
<b>8. Bibliografia</b>	<b>55</b>



# Índex Figures

Figura 6.2.1. BPMN Anonimització DICOM	17
Figura 6.2.2. BPMN Anonimització història clínica	19
Figura 6.2.3. BPMN Pseudo-anonimització expedient	21
Figura 6.2.4. BPMN Anonimització carpeta imatges DICOM	23
Figura 6.2.5. BPMN Anonimització carpeta històries clíniques	25
Figura 6.3.1. Interfície grafica 1	26
Figura 6.5.1. Codi de Patró PDF	29
Figura 6.5.2. Codi de no Patró PDF	30
Figura 6.5.3. Funció mare Codi	31
Figura 6.6.. Interfície grafica 2	32
Figura 6.6.1.1. Interfície grafica 3	33
Figura 6.6.1.2 Escollir expedient	34
Figura 6.6.1.3 Alerta èxit pseudo-anonimització expedient	35
Figura 6.6.1.4 Alerta error pseudo-anonimització expedient	35
Figura 6.6.1.5 Resultat pseudo-anonimització expedient	36
Figura 6.6.1.6 Resultat PDF pre pseudo-anonimització	36
Figura 6.6.1.7 Resultat PDF post pseudo-anonimització	37
Figura 6.6.1.8 Resultat DICOM pre pseudo-anonimització	37
Figura 6.6.1.9 Resultat DICOM pre pseudo-anonimització	38
Figura 6.6.2.1 Interfície grafica 4	38
Figura 6.6.2.2 Escollir imatge DICOM	39
Figura 6.6.2.3 Alerta èxit anonimització imatge DICOM	40
Figura 6.6.2.4 Resultat anonimització imatge DICOM	40
Figura 6.6.3.1. Interfície grafica 5	42
Figura 6.6.3.2.. Escollir carpeta d'imatges DICOM	42
Figura 6.6.3.3.. Alerta èxit anonimització carpeta d'imatges DICOM	43
Figura 6.6.3.4 Alerta error anonimització carpeta d'imatges DICOM	44
Figura 6.6.3.5.. Resultat operació anonimització carpeta d'imatges DICOM	44
Figura 6.6.4.1.. Interfície grafica 6	45
Figura 6.6.4.2 Escollir història clínica	45
Figura 6.6.4.3. Alerta èxit anonimització història clínica	46
Figura 6.6.4.4. Resultat operació anonimització història clínica	47
Figura 6.6.4.5. Història clínica pre anonimització	47
Figura 6.6.4.6. Història clínica post anonimització	48
Figura 6.6.5.1. Interfície grafica 7	49
Figura 6.6.5.2. Escollir carpeta d'històries clíniques	49
Figura 6.6.5.3. Alerta èxit anonimització carpeta històries clíniques	50
Figura 6.6.5.4. Resultat operació anonimització carpeta històries clíniques	51





---

## ÍNDIX DE TAULES

Taula 6.2.1.....	16
Taula 6.2.2.....	18
Taula 6.2.3.....	20
Taula 6.2.4.....	22
Taula 6.2.5.....	24

# 1. Introducció

Aquest TFG proposa una solució a la utilització d'històries clíniques dels pacients i imatges de format DICOM, un estàndard de transmissió d'imatges mèdiques i dades entre hardware de propòsit mèdic, per usar-les com a casos de recerca.

A causa que hi ha dades personals sensibles que no es poden fer públiques, ja que es vulnera la Llei General de Protecció de Dades [\[2\]](#), es decideix fer uns processos d'anonimització o pseudo-anonimització en aquestes dades.

## 1.1 Motivació

Aquest treball vol millorar la recerca i la preparació dels professionals del món sanitari a partir de l'anonimització d'imatges DICOM i d'històries clíniques per tal que no es vegin afectats per la Llei Orgànica de Protecció de Dades [\[2\]](#)



## 2. Marc teòric

### 2.1 Context

Als hospitals es genera una gran quantitat d'informes clínics i proves diagnòstiques que són molt interessants per la recerca. En aquests informes usualment hi han dades personals sensibles. Aquesta informació ha de passar per uns processos per poder usar-la en recerca.

Aquests processos són l'anonimització o pseudo-anonimització de les dades.

Les dades sensibles són anonimitzades o pseudo-anonimitzades perquè no hi hagi cap forma de relacionar aquesta informació amb les dades de cap persona.

La definició d'anonimitzar i pseudo-anonimització ja és assenyalada per Carles San José Amat[1]: “L'anonimització és fer anònimes unes dades per impedir la identificació personal”.

La definició de pseudo-anonimització ja és assenyalada per Carles San José Amat[1]:

“Tècnica de confidencialitat de les dades personals que consisteix a substituir les informacions que identifiquen una persona per un pseudònim.”.

Tot això que s'ha explicat prèviament és vital per entendre i dur a terme una solució, ja que tot el procés d'anonimització o pseudo-anonimització de les dades passa directament per tal de no vulnerar la Llei de Protecció de Dades[2].

L'Hospital de Mataró proposa aquest TFG per trobar una forma d'anonimitzar o pseudo-anonimitzar històries clíniques i imatges DICOM perquè siguin usades a classes o en recerca.

Prèviament a la reunió amb el Director d'Informàtica i Sistemes d'Informació de l'Hospital de Mataró, en Francesc Moya, el dia 1 de febrer de 2022, la idea que es tenia del projecte era que es volia anonimitzar o pseudo-anonimitzar informació directament des de les bases de dades de l'hospital això implica que l'aspecte legal sigui bastant més complex, per la dificultat tècnica, en canvi, baixa dràsticament, ja que és molt més simple identificar dades sensibles en dades estructurades que en documents i imatges.

## **3. Objectius i abast**

### **3.1 Objectius del producte:**

- Anonimitzar o pseudo-anonimitzar històries clíniques en format pdf.
- Anonimitzar o pseudo-anonimitzar imatges DICOM.

### **3.2 Objectius del client:**

- Desenvolupar un producte que en rebre una història clínica en format PDF, retorni aquest mateix document però anonimitzat o pseudo-anonimitzat.
- Desenvolupar un producte que en rebre una imatge DICOM, retorni aquesta mateixa imatge anonimitzada.
- Desenvolupar un producte que en rebre un expedient retorni tots els ítems de la carpeta pseudo-anonimitzats.
- Desenvolupar un producte que en rebre una carpeta d'imatges DICOM, retorni tots els ítems de la carpeta anonimitzats.
- Desenvolupar un producte que en rebre una carpeta d'històries clíniques, retorni tots els ítems de la carpeta anonimitzats.

### **3.3 Target:**

El TFG està orientat cap als professionals, usuaris o institucions que volen fer recerca de casos reals i pràctics mèdics o fer anàlisis concrets de certes malalties o problemes i per fer-ho necessiten que no hi hagi dades sensibles i personals dels pacients.



## 4. Anàlisi de referents

Dins del sector sanitari és difícil trobar una solució que englobi tot el problema que proposa aquest projecte, però sí que s'ha pogut trobar diferents aplicacions o eines que es poden complementar.

### 4.1 Eines i llibreries

#### 4.1.1 ARX

És una eina d'anonimització de dades sensibles d'ús gratuït compatible amb: SQL, fitxers CSV, MS Excel, MS SQL, DB2, MySQL i PostgreSQL. Aquesta aplicació es pot utilitzar des d'un entorn gràfic amb una interfície molt amigable i fàcil d'entendre. També es pot usar com una llibreria en un entorn de Java i fent crides a la seva pròpia API.

ARX implementa una avaluació de riscos i això és un factor determinant, ja que permet saber a l'usuari si les dades que han estat anonimitzades són de qualitat o és relativament fàcil identificar a qui pertanyen les dades anonimitzades.

L'únic requeriment que té aquesta eina és que es necessita una llicència d'Apache 2.0 la qual és totalment gratuïta. ARX pot ser utilitzat en els sistemes operatius més usats avui en dia: Windows, Mac OS X i Linux.



### **4.1.2 AMNESIA**

Amnesia és una eina d'anonimització de dades flexible que esborra la informació d'identificació de les dades. Aquesta eina és compatible tan sols amb fitxers CSV.

Amnesia pot ser usada a través d'una interfície gràfica molt amigable o a través de trucades REST API.

Aquesta eina va ser utilitzada en la Unió Europea en el projecte “ My Health My Data” per poder fer intercanvis d'informació sanitària de forma segura.

Amnesia està disponible en Windows i Linux.

### **4.1.3 PYDICOM**

Pydicom és una llibreria de Python que permet la lectura, l'edició i l'escriptura en documents DICOM.

### **4.1.4 PyPDF2**

PyPDF2 és una llibreria de Python que permet llegir i editar documents en format PDF.

### **4.1.5 NLP**

NLP o Natural Language Processing se centra en com les màquines entenen, interpreten i processen el llenguatge humà. A través d'aquesta àrea de la intel·ligència artificial s'han creat diverses eines i solucions per poder classificar i dividir a partir dels diferents tipus de paraules que componen una oració.

## **4.2 Estàndards**

### **4.2.1 OMOP**

OMOP CDM (Common Data Model) és un estàndard que permet l'anàlisi de diverses bases de dades a partir de transformar les dades d'aquestes en un mateix format (data model) amb l'objectiu de fer diferents anàlisis. El model de dades OMOP tracta d'una representació conceptual d'una taula amb informació del pacient on s'inclouen totes les dades observacionals rellevants per la possible identificació d'intervencions de salut i resultats de proves.

### **4.2.2 SNOMED CT**

SNOMED ( Systematized Nomenclature of Medicine - Clinical Terms) es va crear per tal d'unificar sota un mateix estàndard la comunicació d'informació de salut ja siguin diagnòstics d'un pacient, medicaments o resultats de proves. És una col·lecció que conté diferents termes mèdics, codis, sinònims i definicions usades en històries clíniques amb la finalitat de facilitar el registre clínic i l'atenció al client.

### **4.2.3 FHIR**

FHIR (Fast HealthCare Interoperability Resources) és un estàndard que es va crear per unificar sota un estàndard el format de les dades i dels recursos per l'intercanvi de registres sanitaris electrònics. Els recursos són la unitat bàsica d'interoperabilitat, representen conceptes del món sanitari: un metge, un pacient, un diagnòstic.

## **4.3 Serveis**

### **4.3.1 NYMIZ**

En Nymiz es pot trobar tant servei com producte i deixa també escollir entre anonimització i pseudo-anonimització. Aquesta eina permet l'anonimització de fitxers com .doc, .docx, .pdf, .txt, power point i també l'anonimització o pseudo-anonimització de bases de dades.

El perquè s'ha descartat Nymiz son varius motius:

- Necessitat de connectivitat a internet i el risc de tindre un atac o filtració de dades sensibles dels pacients.
- Límit d'usuaris que poden utilitzar el servei.
- El cost mensual és elevat.
- L'Hospital de Mataró vol un producte propi, no contractar un servei.

## **5. Metodologia**

En aquest projecte s'ha decidit utilitzar la metodologia Waterfall, ja que en ser tan sols una persona és la més pràctica i simple per dur a terme tota la planificació del projecte.

### **5.1 Definició del projecte**

Primerament, es defineix tot el projecte a partir d'una anàlisi de les solucions existents i el context en què se situa.

Una vegada es fa aquesta anàlisi es crea una llista de funcionalitats que té el producte i es fa una planificació repartint totes aquestes tasques definides en el temps.

Es du a terme un estudi de viabilitat per tal de veure si tot aquest projecte és viable econòmicament, tecnològicament i mediambientalment.

### **5.2 Adquisició de coneixements**

En aquesta etapa s'estudia el com usar les tecnologies i les eines necessàries per a dur a terme el producte.

En aquest cas s'estudia el format DICOM i el com fer l'anàlisi d'imatge per poder anonimitzar o pseudo-anonimitzar les dades sensibles que no es volen mostrar.

També s'estudia totes les maneres possibles per poder anonimitzar o pseudo-anonimitzar les històries clíniques en format pdf i amb dades desestructurades.

## 5.3 Producció

Una vegada es tenen tots els coneixements necessaris tant de l'entorn com de les tecnologies existents és porta a terme el desenvolupament del producte final.

Primerament es desenvolupa l'anonimització o pseudo-anonimització d'imatges DICOM a partir de scripts en Python analitzant cadascuna de les capçaleres de les imatges que han de passar pel procés i protegint totes les dades sensibles necessàries.

Mentre es du a terme aquesta funcionalitat també es crea un script en Python per fer anàlisis d'un pdf amb l'objectiu de trobar totes les dades sensibles que hi ha en el document per així poder anonimitzar-les o pseudo-anonimitzar-les a elecció de l'usuari.

Una vegada es fa l'anàlisi del document i troba totes les dades sensibles que hi conté es passa a desenvolupar una funcionalitat perquè anonimitzi o pseudo-anonimitzi totes aquestes dades depenent de l'opció que esculli l'usuari.

## 5.4 Testing

En aquesta etapa es comprova si les funcionalitats desenvolupades funcionen correctament sense cap mena d'error o bug que pugui causar un resultat imprevist.

El testing que s'utilitza, mentre el prototip està a l'etapa de desenvolupament, és manual. Es basa en petits mock ups i veure el resultat en els fitxers una vegada la funció que es vol testejar s'executa.

A l'hora de trobar els camps patronitzables en la història clínica, per veure si el patró Regex és correcte el que es fa és crear un petit mock up per veure si el detecta o directament es comprova que en el fitxer que el camp que es vol anonimitzar el detecti i l'anonimitzi.

S'ha plantejat fer dos tipus de script per fer test de forma automàtica.

Un script que per cada vegada que s'anonimitzi o és pseudo-anonimitzi una imatge DICOM s'executa un script de test per veure que tots els camps han sigut modificats o anonimitzats correctament.

Per altra banda un script per l'anonimització d'històries clíniques on mira per cada camp de la història clínica que no hi hagi text excepte en els camps de l'edat i sexe.



## **6. Desenvolupament**

### **6.1 Anàlisi i definició de requeriments**

Els requisits funcionals del producte són:

- Anonimitzar històries clíniques en format PDF.
- Anonimitzar imatges en format DICOM.
- Pseudo-anonimitzar un expedient complet
- Anonimitzar carpeta d'històries clíniques en format PDF.
- Anonimitzar carpeta d'imatges en format PDF.

En l'àmbit tecnològic es necessita disposar d'un ordinador per poder instal·lar i executar l'aplicació.



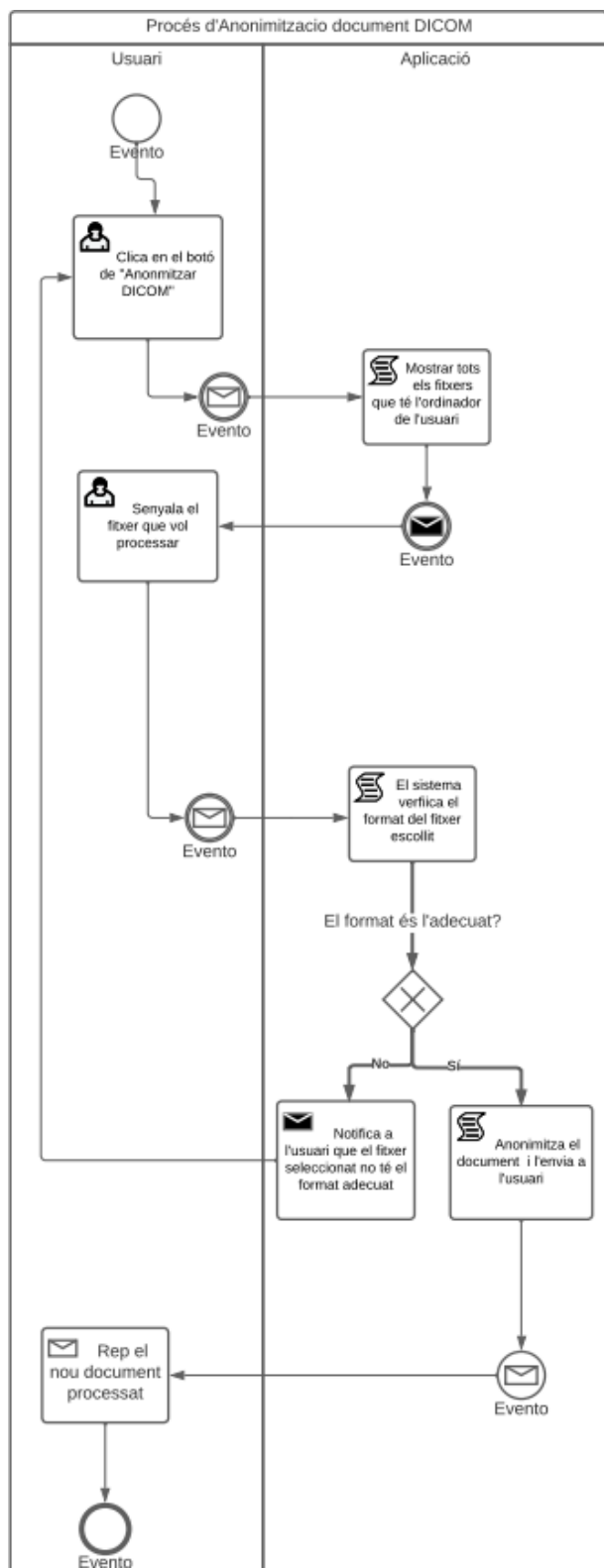
## 6.2 Casos d'ús del sistema

### CA\_U.1. Anonimitzar imatge DICOM

<b>Nom</b>	Anonimitzar document DICOM
<b>Descripció</b>	L'usuari ha de poder escollir el fitxer de format DICOM i obtindre-ho anonimitzat
<b>Precondicions</b>	L'usuari ha d'haver obert l'aplicació.
<b>Postcondicions</b>	L'usuari obté el fitxer amb el mateix format però anonimitzat
<b>Flux normal d'execució</b>	100. L'usuari clica el botó de "Anonimitzar imatge DICOM". 200. L'usuari assenyala el fitxer que vol processar. 300. El sistema processa el document i verifica el format. 400. El sistema anonimitza el document i el retorna a l'usuari.
<b>Flux alternatiu</b>	400. En cas d'haver-hi un error, mostra un missatge d'error a l'usuari.

*Taula 1*

Figura 1. BPMN Anonimització DICOM

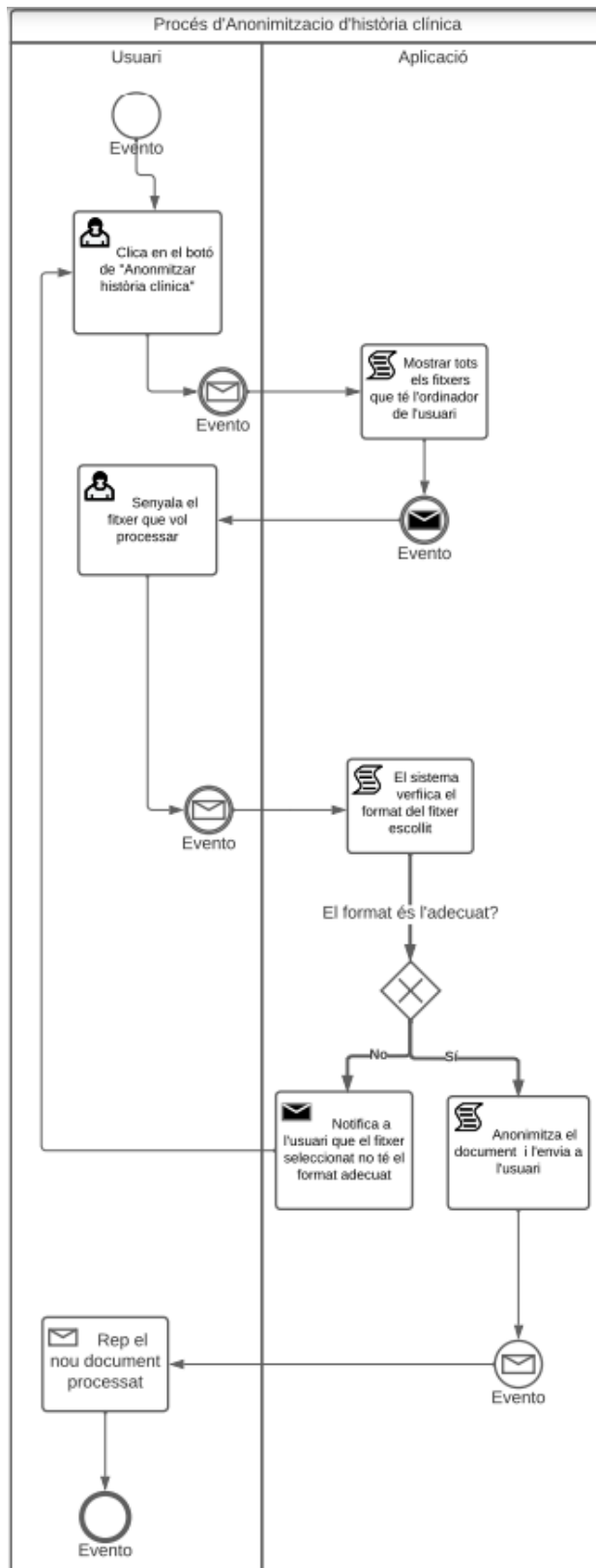


## CA\_U.2. Anonimitzar història clínica

<b>Nom</b>	Anonimitzar història clínica
<b>Descripció</b>	L'usuari ha de poder escollir el fitxer en format PDF que vol processar i l'obté anonimitzat
<b>Precondicions</b>	L'usuari ha d'haver obert l'aplicació.
<b>Postcondicions</b>	L'usuari obté el fitxer amb el mateix format però anonimitzat
<b>Flux normal d'execució</b>	100. L'usuari clica el botó de "Anonimitzar història clínica". 200. L'usuari assenyala el fitxer que vol processar. 300. El sistema processa el document i verifica el format. 400. El sistema anonimitza el document i el retorna a l'usuari.
<b>Flux alternatiu</b>	400. En cas d'haver-hi un error, mostra un missatge d'error a l'usuari.

Taula 2

Figura 2. BPMN Anonimització història clínica

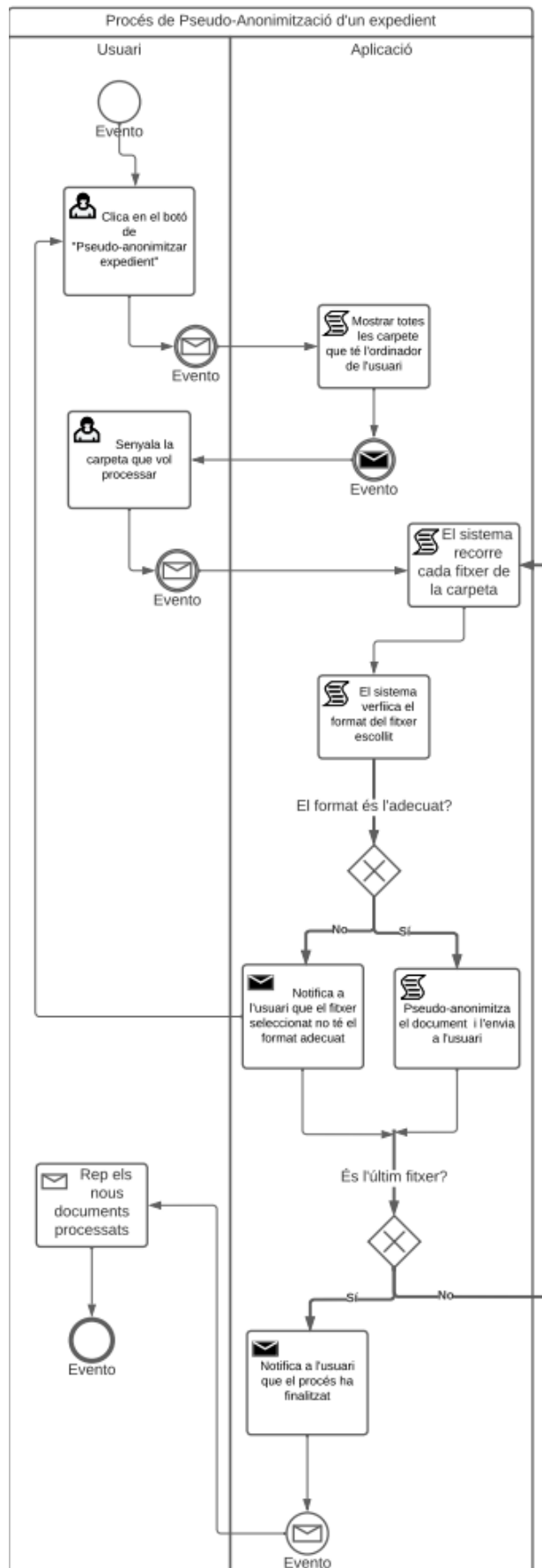


## CA\_U.3. Pseudo-anonimitzar expedient

<b>Nom</b>	Pseudo-anonimitzar expedient
<b>Descripció</b>	L'usuari ha de poder escollir una carpeta on hi hagi, una o varies imatges DICOM i una o varies històries clíniques del mateix pacient per pseudo-anonimitzar-la.
<b>Precondicions</b>	L'usuari ha d'haver obert l'aplicació.
<b>Postcondicions</b>	L'usuari obté dins la carpeta escollida tots els fitxer anonimitzats.
<b>Flux normal d'execució</b>	100. L'usuari clica el botó per pseudo-anonimitzar un expedient 200. L'usuari senyala la carpeta que vol processar. 300. El sistema recorre cada fitxer de la carpeta. 400. El sistema pseudo-anonimitza cada document i el guarda dins de la carpeta escollida
<b>Flux alternatiu</b>	400. En cas d'haver un error, mostra un missatge d'error a l'usuari.

Taula 3

Figura 3. BPMN Pseudo-anonimització expedient

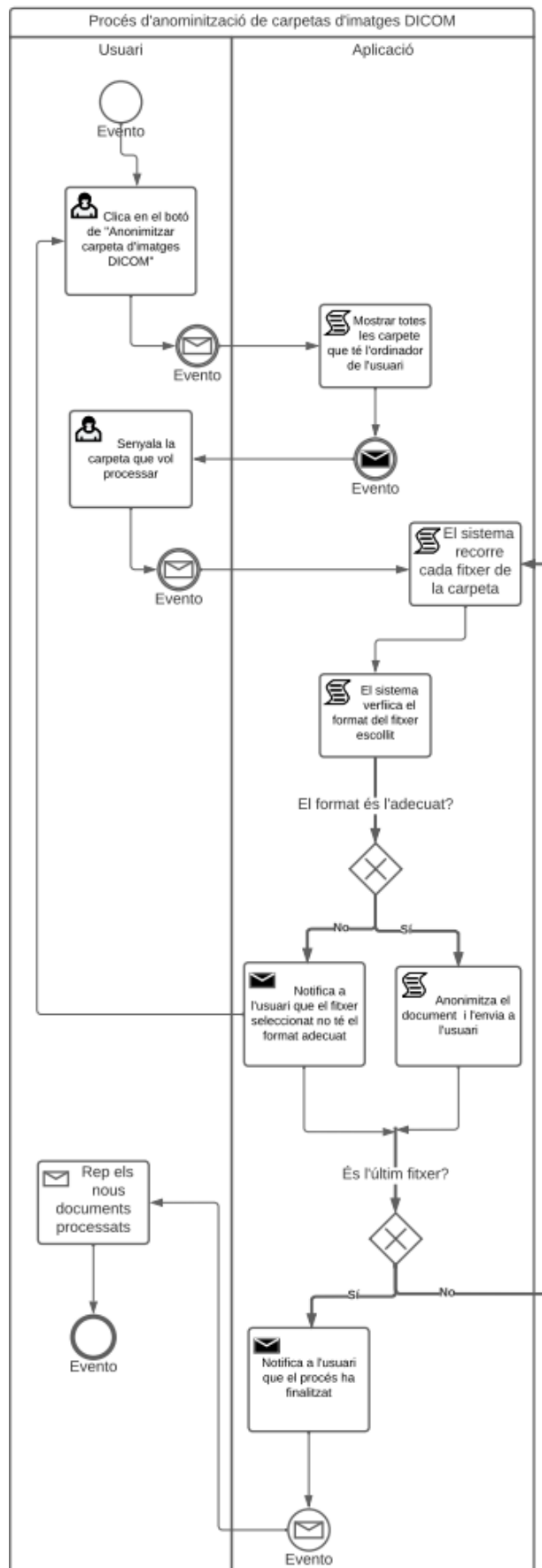


## CA\_U.4. Anonimitzar carpeta d'històries clíniques

<b>Nom</b>	Anonimitzar carpeta d'històries clíniques
<b>Descripció</b>	L'usuari ha de poder escollir una carpeta on hi hagi una o múltiples històries clíniques de diferents pacients.
<b>Pre-condicions</b>	L'usuari ha d'haver obert l'aplicació.
<b>Post-condicions</b>	L'usuari obté els fitxers de la carpeta amb el mateix format però anonimitzats.
<b>Flux normal d'execució</b>	100. L'usuari clica el botó per anonimitzar carpeta d'històries clíniques. 200. L'usuari assenyala la carpeta que vol processar. 300. El sistema recorre cada fitxer de la carpeta. 400. El sistema anonimitza cada document i el guarda dins de la carpeta escollida.
<b>Flux alternatiu</b>	400. En cas d'haver-hi un error, mostra un missatge d'error a l'usuari.

Taula 4

Figura 4. BPMN Anonimització carpeta imatges DICOM



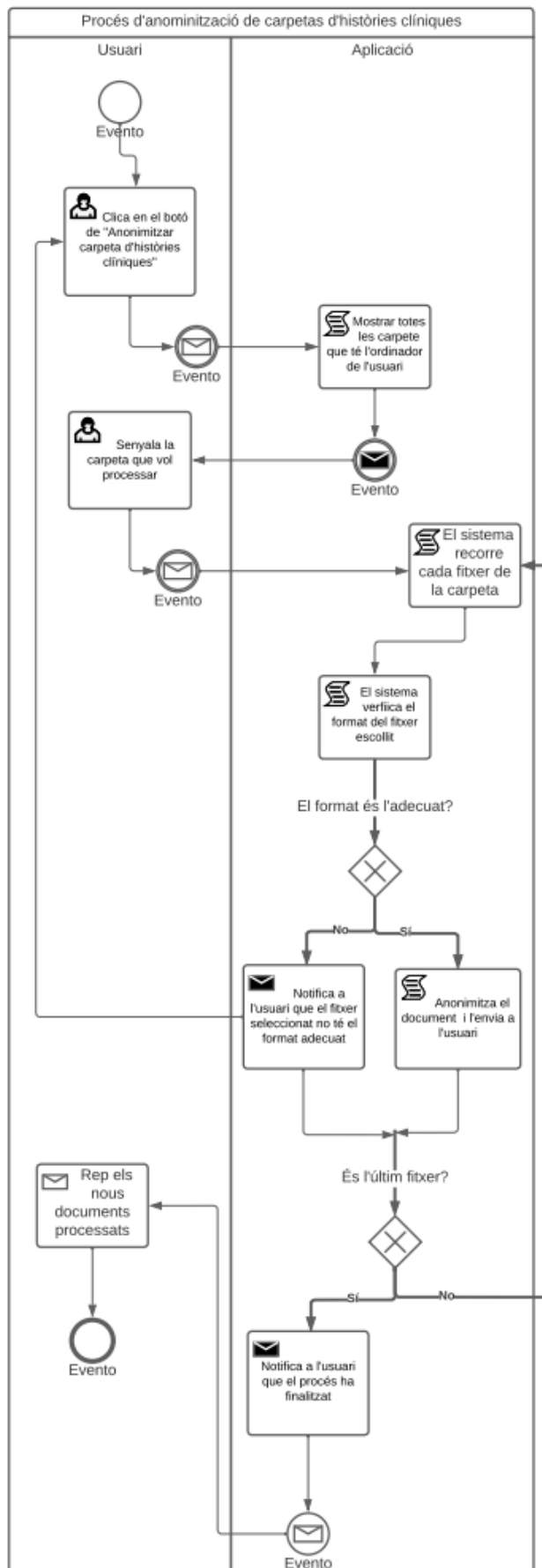


## CA\_U.5. Anonimitzar carpeta d'imatges DICOM

<b>Nom</b>	Anonimitzar carpeta d'imatges DICOM
<b>Descripció</b>	L'usuari ha de poder escollir una carpeta on hi hagi una o varies imatges DICOM de diferents pacients.
<b>Pre-condicions</b>	L'usuari ha d'haver obert l'aplicació.
<b>Post-condicions</b>	L'usuari obté els fitxers de la carpeta amb el mateix format però anonimitzats.
<b>Flux normal d'execució</b>	100. L'usuari clica el botó per anonimitzar carpeta d'històries clíniques. 200. L'usuari senyala la carpeta que vol processar. 300. El sistema recorre cada fitxer de la carpeta. 400. El sistema anonimitza cada document i el guarda dins de la carpeta escollida.
<b>Flux alternatiu</b>	400. En cas d'haver un error, mostra un missatge d'error al usuari.

Taula 5

Figura 5. BPMN Anonimització carpeta històries clíniques



### 6.3 Disseny del software

Pel software s'ha decidit utilitzar Python com a llenguatge, per la gran quantitat de llibreries que té, per la facilitat a l'hora d'utilitzar IA i els components de processament de llenguatge i finalment per la rapidesa i efectivitat amb l'edició i creació de documents en format PDF.

L'arquitectura que es fa servir és una mena de MVC on el controlador i la vista són el mateix script, anomenat `gui_controller.py`.

Aquest controlador s'encarrega de decidir a quin script ha de cridar segons l'opció que l'usuari escull.

Els scripts, que fan de models, que pot cridar són:

- Pdf\_Anonymizer.py: aquest script gestiona tota l'anonimització i pseudo-anonimització d'una història clínica.
- Dicom-Anonymizer.py: aquest script gestiona tota l'anonimització i pseudo-anonimització d'una imatge DICOM.
- Expedient\_Anonymizer.py: aquest script gestiona la pseudo-anonimització d'un expedient i l'anonimització de carpetes a partir de recórrer tots els ítems d'aquests i fent crides a Pdf\_Anonymizer.py i Dicom-Anonymizer.py dependent del format del ítem.

Figura 6. Interfície grafica 1



## 6.4 Llibreries utilitzades

### 6.4.1 TKINTER

És la llibreria per defecte de la interfície de Python. Aquesta llibreria té molts mòduls i opcions diferents per poder personalitzar el màxim possible la teva interfície gràfica.

### 6.4.2 PDFPLUMBER

És una llibreria que permet llegir, extreure text i informació detallada de cada rectangle i línia. Després de tindre problemes amb diferents tipus de OCR (són les aplicacions que a partir d'una imatge poden extreure el text d'aquesta) ja que extreuen el text d'una forma que no és possible llegir o fer algoritmes a partir d'aquesta. Es decideix provar la llibreria PDFPLUMBER amb la que es va poder extreure el text del fitxer d'una forma perfecte i llegible sense necessitat de processar el text amb l'intenció de fer-ho llegible.

### 6.4.3 RE

És un mòdul de Python que permet buscar en strings patrons Regex. Aquesta llibreria s'utilitza per trobar dintre del text totes aquelles dades sensibles que es poden trobar a partir de patrons, com per exemple el DNI, un número de telèfon, etc.

### 6.4.4 ASPOSE.WORDS

És una llibreria utilitzada per processar tasques en documents, ja sigui modificar, crear, convertir, renderitzar o imprimir documents per així no dependre d'altres aplicacions o automatitzacions.

En aquest cas s'ha utilitzat aquesta llibreria per transformar el fitxer PDF escollit en un Word on poder modificar el text que es vulgui.

També aquesta llibreria té una funció de substitució de text a partir d'un patró Regex que ha sigut molt útil per tal de fer la cerca de totes aquestes dades sensibles patronitzables i anonimitzar-les directament.

### **6.4.5 OS**

Aquest mòdul de Python ho proveeix a través de la seva llibreria estàndard. El mòdul permet accedir a funcionalitats dependents del sistema operatiu com per exemple: manipular directoris, canviar i alterar les direccions a directoris o fitxers i recórrer tots els fitxers dins d'un directori.

### **6.4.6 UUID**

Aquest mòdul permet crear objectes UUID per atribuir a cada pacient un identificador únic per relacionar la seva història clínica i la imatge DICOM del mateix pacient.

### **6.4.7 PYDICOM**

Aquesta llibreria permet llegir i modificar imatges mèdiques. Amb aquesta llibreria s'ha pogut anonimitzar tots els camps que contenen dades sensibles o introduir dades per poder fer la pseudo-anonimització

## 6.5 Documentació i escalabilitat del codi

El projecte creat no és un producte final, és un prototip i té molts aspectes a millorar, casos no coberts, etc.

Però el codi s'ha desenvolupat d'una manera en la que afegir patrons, o en el cas que s'afegeix un camp al format d'història clínica, sigui molt fàcil d'implementar i escalar el projecte per tal de cobrir tots els casos possibles.

En el cas que sigui necessari afegir un nou camp en el format de la història clínica de l'hospital que sigui patronitzable, és a dir que segueix un patró o una estructura com un telèfon o un DNI.

El que s'hauria de fer és:

1. Crear una funció al script de Pdf\_Anonymizer que segueix l'estructura d'aquesta mateixa funció.
2. Trobar el patró Regex que cobreix el cas que es vol anonimitzar.
3. Veure quants són els espais necessaris per fer la substitució de les paraules trobades.
4. Cridar la funció creada a la funció mare

Figura 7. Codi de Patró PDF

```
def anonymize_NIF(text, doc):
    options = aw.replacing.FindReplaceOptions()
    nif_numbers=re.findall(r"[0-9]{8}[A-Z]", text)
    for nif_number in nif_numbers:
        doc.range.replace(nif_number, "          \n", options);
```

En el cas sigui necessari afegir un nou camp a la història clínica que no sigui patronitzable o que tingui un patró molt més difícil de trobar el que s'ha de fer per cobrir aquest nou cas és:

1. Afegir una condició if on la paraula start sigui la paraula que ve abans del que es vol anonimitzar i que la paraula end sigui el que va després del contingut que es vol anonimitzar
2. Veure quants són els espais necessaris perquè el document quedi amb el mateix estil previ a l'anonimització

Figura 8. Codi de no Patró PDF

```
def anonymize_not_patronable_fields(start, end, text, doc, patient_id=""):
    #We get what is between the two words
    words=text[text.find(start)+len(start):text.rfind(end)]
    #Delete the white spaces from the left and the right of the words
    words=words.strip()
    if start=="Cognoms:" and end=="Adreça:":
        doc.range.replace(words, "\n")
    if start=="Adreça:" and end=="Nom:":
        doc.range.replace(words, "\n")
    if start=="Nom:" and end=="Edat:":
        if patient_id!="":
            doc.range.replace(words, patient_id)
        else:
            doc.range.replace(words, "\n")
    if start=="Metge/essa:" and end=="Servei:":
        doc.range.replace(words, ")")
    if start=="Servei:" and end=="Data ingrés:":
        doc.range.replace(words, ")")
    if start=="Núm. Llit:" and end=="DIAGNÒSTICS":
        doc.range.replace(words, "")
```

Figura 9. Funció mare Codi

```
def anonymize_ER(path,patient_id=""):
    result=""
    try:
        text=getting_PDF_text(path);
        doc=PDF_to_Word(path)
        anonymize_not_patronable_fields("Cognoms:", "Adreça:",text,doc);
        anonymize_not_patronable_fields("Adreça:", "Nom:",text,doc);
        anonymize_not_patronable_fields("Nom:", "Edat:",text,doc,patient_id);
        anonymize_not_patronable_fields("Metge/essa:", "Servei:",text,doc);
        anonymize_not_patronable_fields("Servei:", "Data ingrés:",text,doc);
        anonymize_not_patronable_fields("Núm. Llit:", "DIAGNÒSTICS",text,doc);
        anonymize_NIF(text,doc);
        anonymize_NASS(text,doc);
        anonymize_dates(text,doc);
        anonymize_CIP(text,doc);
        anonymize_phone_numbers(text,doc);
        anonymize_NHC(text,doc,path);
        save_anonimized_ER(path, doc)
        delete_word_file(path)
        result="OK"
    except:
        result="ERROR"
    return result;
```

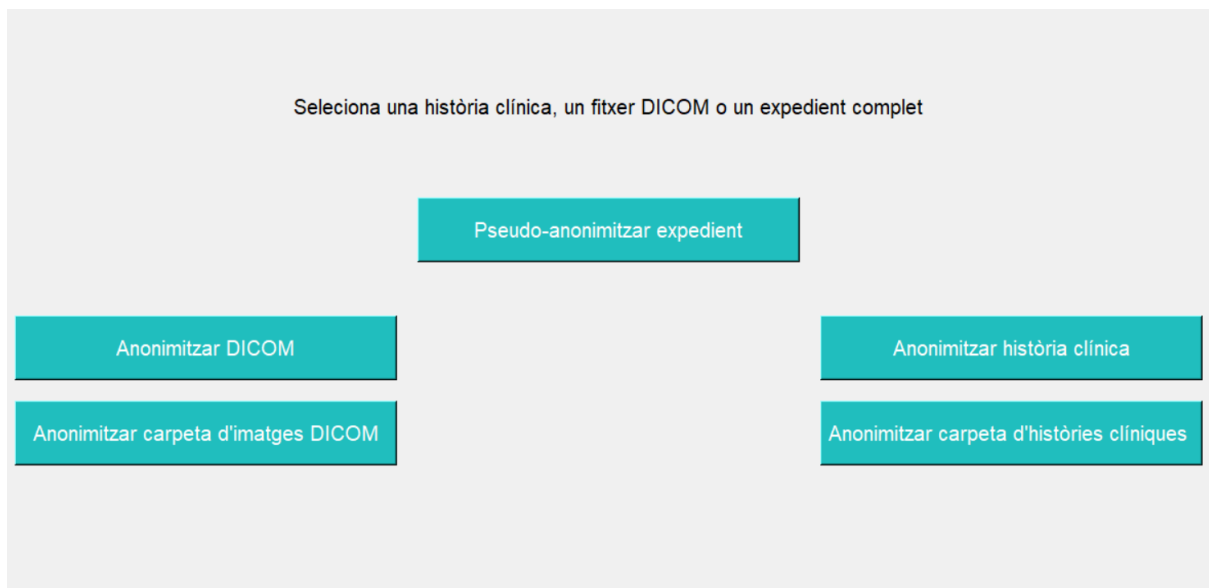


## 6.6 Solució creada

L'aplicació d'escriptori creada és intuïtiva i fàcil d'utilitzar per a l'usuari i també molt fàcil d'evolucionar i escalar pels desenvolupadors.

Una vegada s'executa el programa la primera pantalla que apareix és la del menú principal on hi ha totes les opcions possibles.

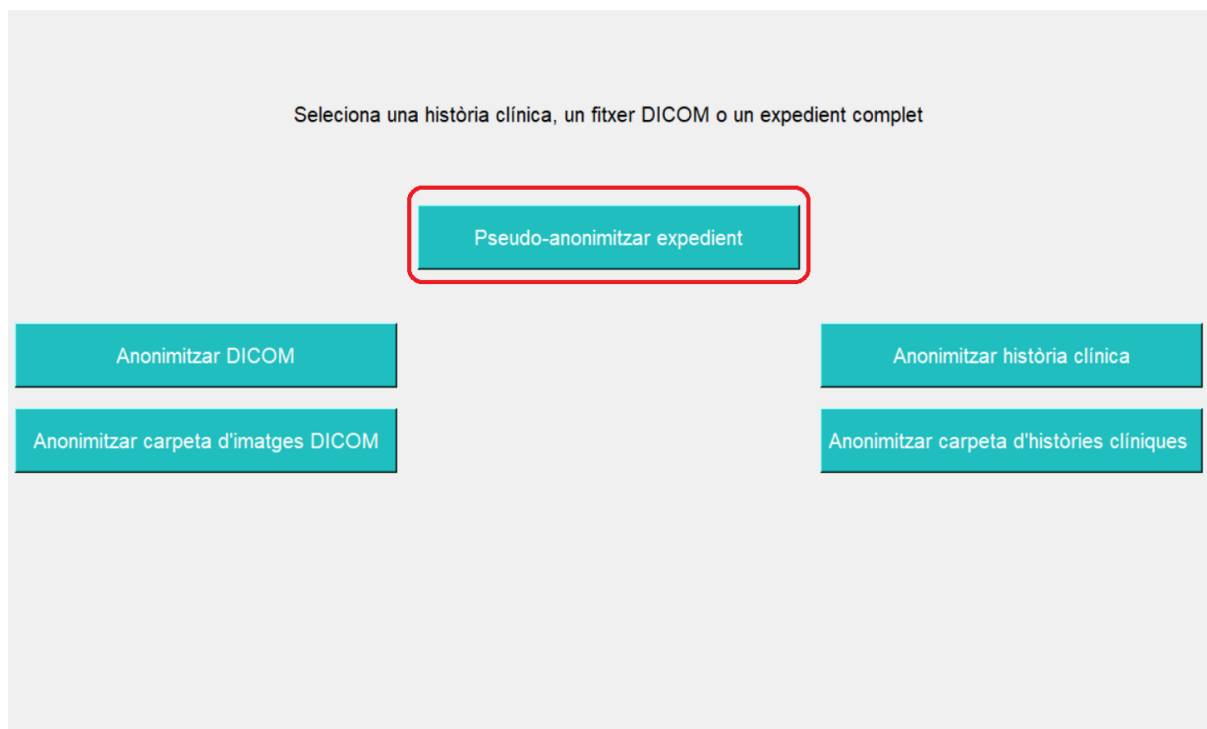
Figura 10. Interfície grafica 2



### 6.6.1 Pseudo-anonimitzar expedient

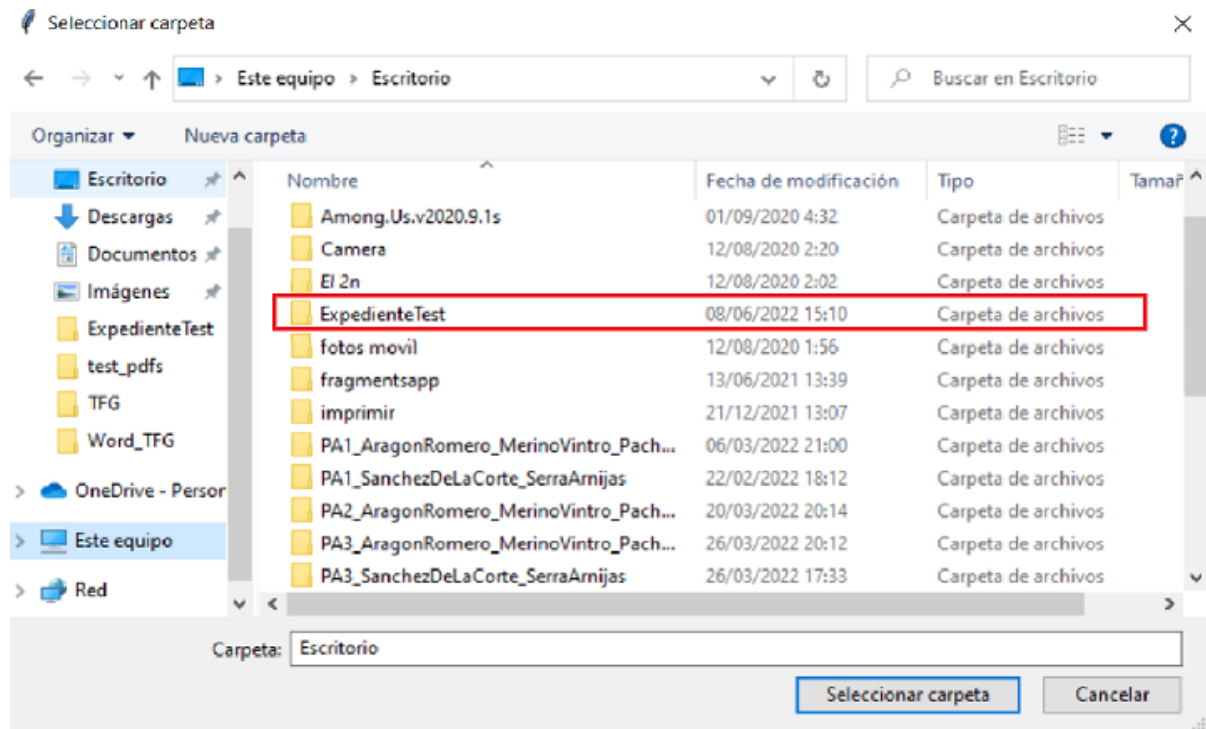
Aquest cas permet pseudo-anonimitzar un expedient complet d'un pacient, és a dir, aquesta opció permet escollir una carpeta on hi hagi una o més d'una història clínica o imatge DICOM del mateix pacient per tal de relacionar-les amb un ID creat específicament per aquell pacient.

Figura 11. Interfície grafica 3



Una vegada l'usuari ha clicat el botó s'obre una finestra per poder escollir l'expedient que l'usuari vol pseudo-anonimitzar.

Figura 12. Escollir expedient



Una vegada l'usuari ha seleccionat el directori, l'aplicació processa des de la interfície gràfica, que fa de controlador, a quin script ha de cridar.

En aquest cas es crida al script de Expedient\_Anonymizer.py que aquest crida als scripts de Pdf\_Anonymizer.py i al script de Dicom\_Anonymizer.py per pseudo-anonimitzar tots els ítems del directori.

Una vegada l'operació ha acabat, l'aplicació mostra una alerta amb el resultat d'aquesta.

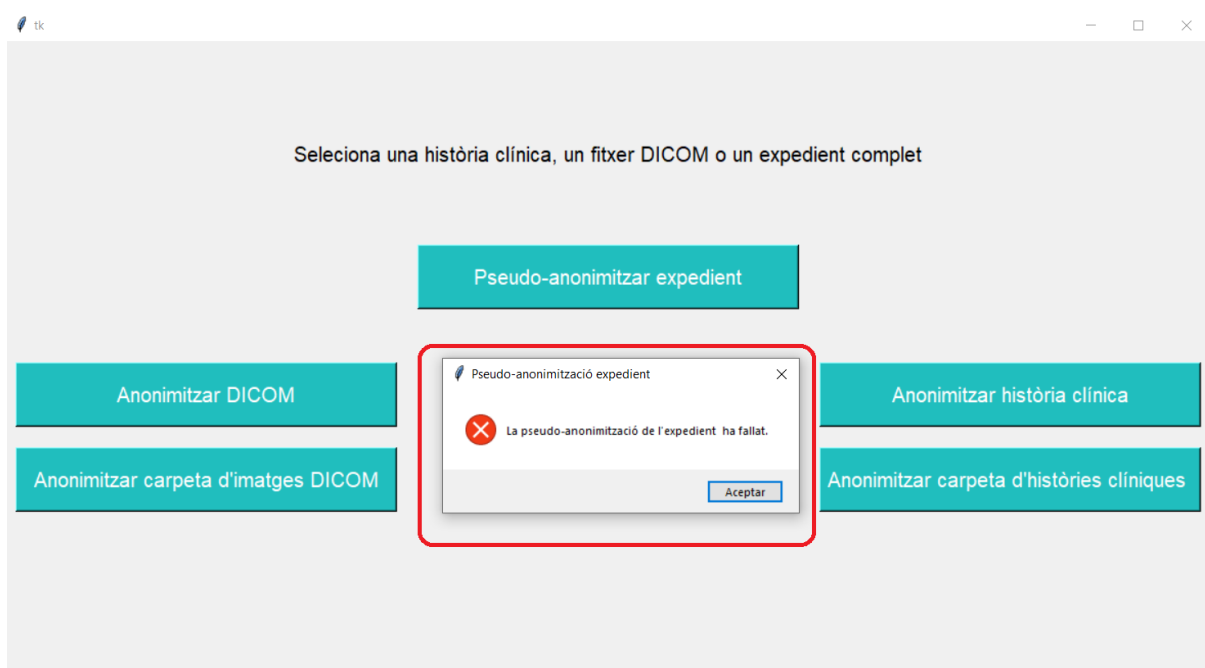
En cas que l'operació sigui un èxit mostra aquesta alerta:

Figura 13. Alerta èxit pseudo-anonimització expedient



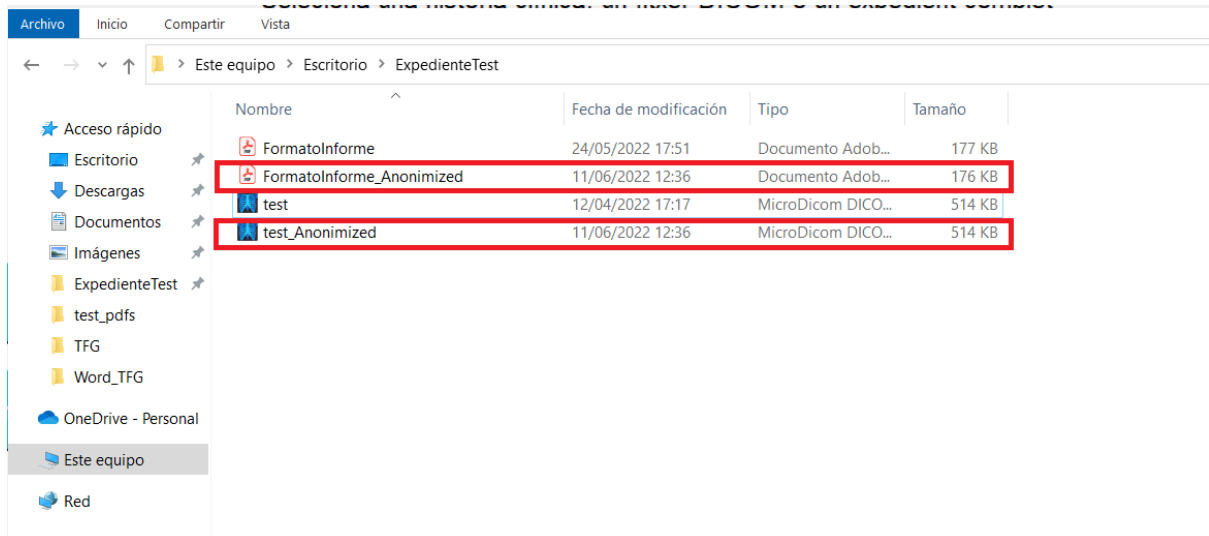
En el cas que per algun motiu l'operació de pseudo-anonimització d'un expedient falli es mostra aquesta alerta:

Figura 14. Alerta error pseudo-anonimització expedient



Finalment, si l'operació ha sigut un èxit dintre, del directori que l'usuari ha seleccionat, pot trobar els fitxers pseudo-anonimitzats.

Figura 15. Resultat pseudo-anonimització expedient



El fitxer PDF abans de ser pseudo-anonimitzat:

Figura 16. Resultat PDF pre pseudo-anonimització

HOSPITAL DE MATARÓ  
 COMPLEJIO SANITARIO DEL MEDITERRANEO  
 Carretera de Cleres, s/n - 08380 Mataró  
 Tel. 937 417 700 Fax 937 417 733

1/1

NO NO, NO

**Cognoms:** Sanchez de la Corte      **Adreça:** C. de la Tramuntana 12-14 baixos 1a  
**Nom:** Xavier  
**Edat:** 22      **Sexe:** M      **CIP:** ABC1234567894      **NIF:** 43584128W  
**Data naixement:** 29/05/2000      **NHC:** 123456      **NASS:** 08/1123564789  
**Telèfon / Mòbil:** 629631477 / 0

---

**INFORME D'ASSISTÈNCIA**

**Metge/essa:** Juan Fernandez Ortega      **Servet:** Medicina interna  
**Data ingress:** 14/05/2014      **Data Àlta:** 14/05/2015  
**S'informa a:** Metge de família      **Núm. lliç:** 903 - 70

---

**DIAGNÒSTICS:**  
 1.

**PROCEDIMENTS:**  
 1.

---

**ANTECEDENTS PATOLÒGICS:**  
 Tractament habitual:

**MOTIU D'INGRÉS:**

**EXPLORACIÓ FÍSICA:**

**EXPLORACIONS COMPLEMENTÀRIES:**  
 Anàlítica:  
 ECG:  
 Rx de tòrax:

**EVOLUCIÓ:**  
**TRACTAMENT:**

MEDICACIÓ	ESMORZAR	DINAR	SOPAR	NIT

**CONTROL:**

El fitxer PDF després de ser pseudo-anonimitzat:

Figura 17. Resultat PDF post pseudo-anonimització

Evaluation Only. Created with Aspose.Words. Copyright 2003-2022 Aspose Pty Ltd.

**HOSPITAL DE MATARÓ**  
CONSEJO SANITARIO DEL MARISSIMU  
 Carrer de la Creu, s/n - 08094 Mataró  
 Tel. 937 417 700 Fax 937 417 733

1 / 1  
NO NO, NO

Cognoms: \_\_\_\_\_ Adreça: \_\_\_\_\_  
 Nom: b56a77d0-2e33-48aa-9d42-2b  
 Edat: 22 Sexe: M CIP: \_\_\_\_\_ NIF: \_\_\_\_\_  
 Data naixement: \_\_\_\_\_ NHC: \_\_\_\_\_ NASS: \_\_\_\_\_  
 Telèfon / Mòbil: / 0

**INFORME D'ASSISTÈNCIA**

Metge/essa: \_\_\_\_\_ Servei: \_\_\_\_\_  
 Data ingress: \_\_\_\_\_ Data Alta: \_\_\_\_\_  
 S'informa a: Metge de família Núm. Illt: \_\_\_\_\_

**DIAGNÒSTICS:**

**PROCEDIMENTS:**

**ANTECEDENTS PATOLÒGICS:**

**MOTIU D'INGRÉS:**

**EXPLORACIÓ FÍSICA:**

**EXPLORACIONS COMPLEMENTÀRIES:**

Analítica:

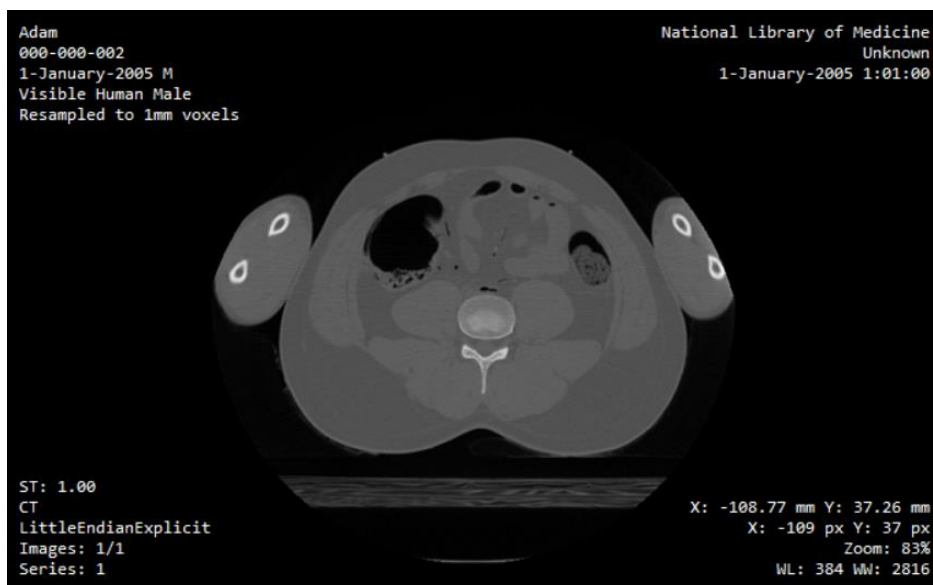
**TRACTAMENT:**

MEDICACIÓ	ESMORZAR	DINAR	SOPAR	NIT

Created with an evaluation copy of Aspose.Words. To discover the full versions of our APIs please visit: <https://products.aspose.com/words/>

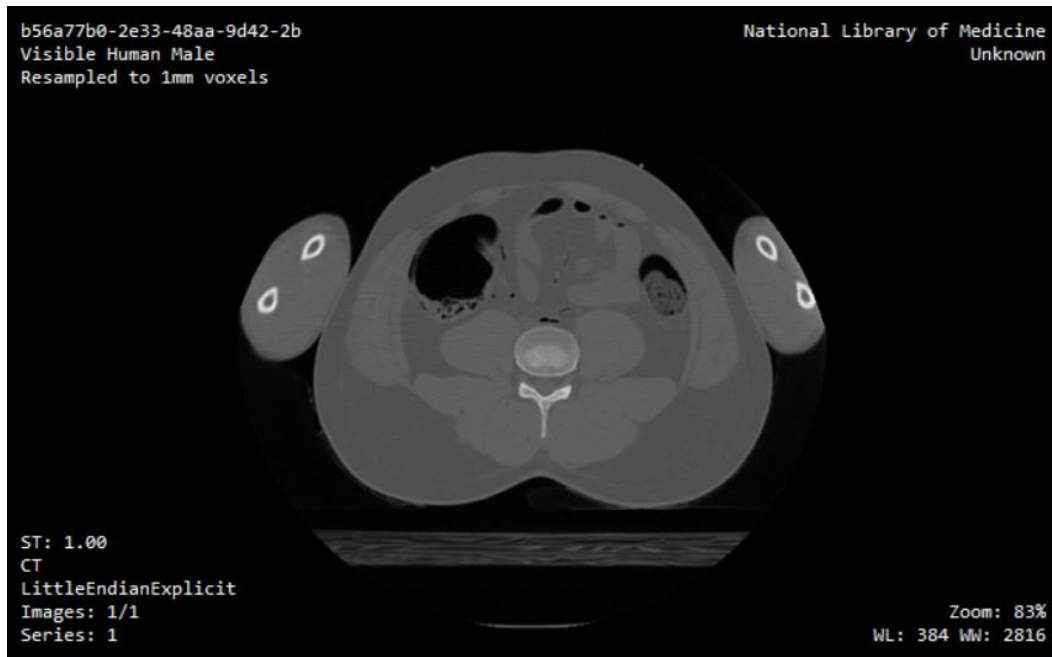
El fitxer DICOM abans de ser pseudo-anonimitzat:

Figura 18. Resultat DICOM pre pseudo-anonimització



El fitxer DICOM després de ser pseudo-anonimitzat:

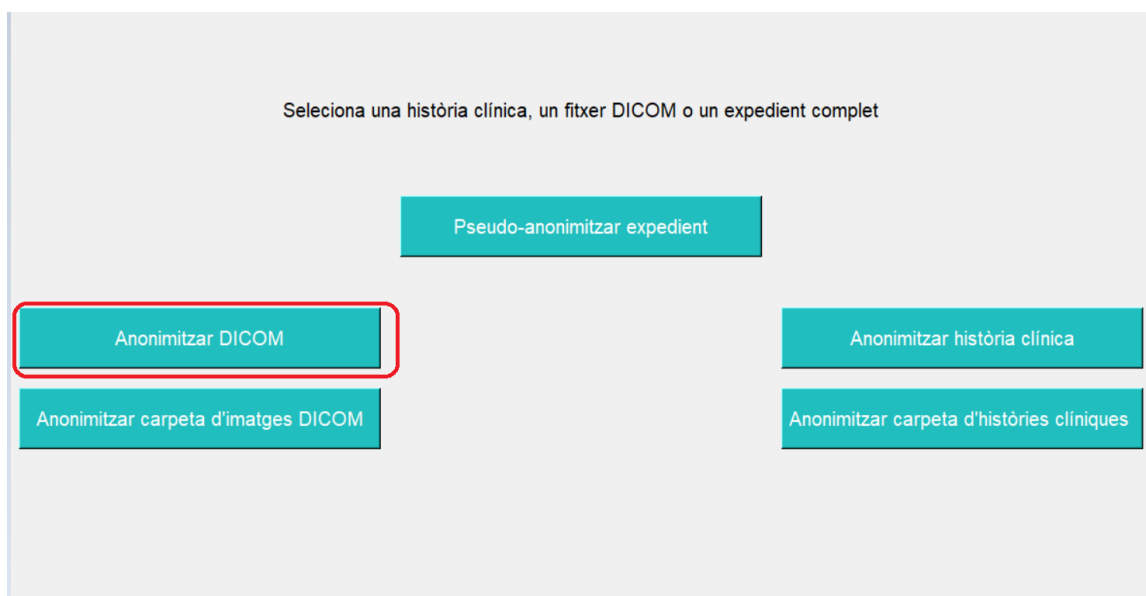
Figura 19. Resultat DICOM pre pseudo-anonimització



## 6.6.2 Anonimitzar DICOM

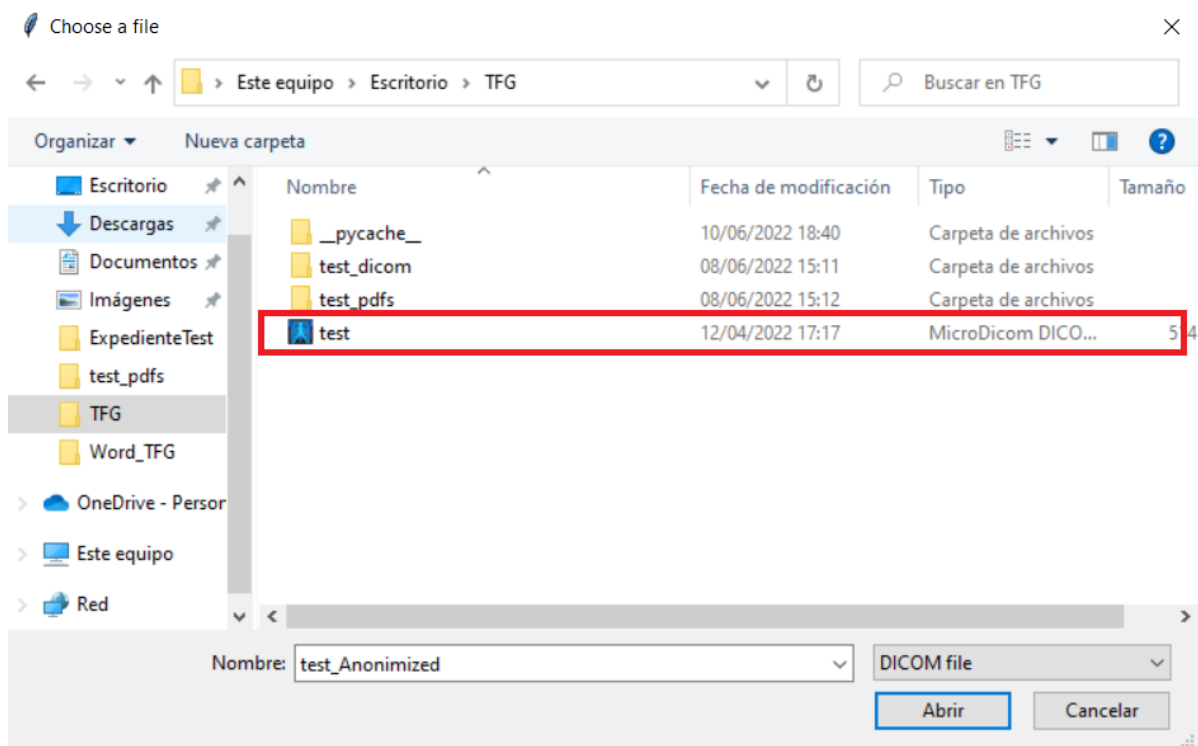
Aquest cas permet anonimitzar una imatge en format DICOM d'un pacient.

Figura 20. Interfície grafica 4



Una vegada l'usuari ha clicat el botó s'obre una finestra per poder escollir la imatge DICOM que l'usuari vol anonimitzar.

Figura 21. Escollir imatge DICOM



Una vegada l'usuari ha seleccionat el fitxer, l'aplicació processa des de la interfície gràfica ,que fa de controlador, a quin script ha de cridar.

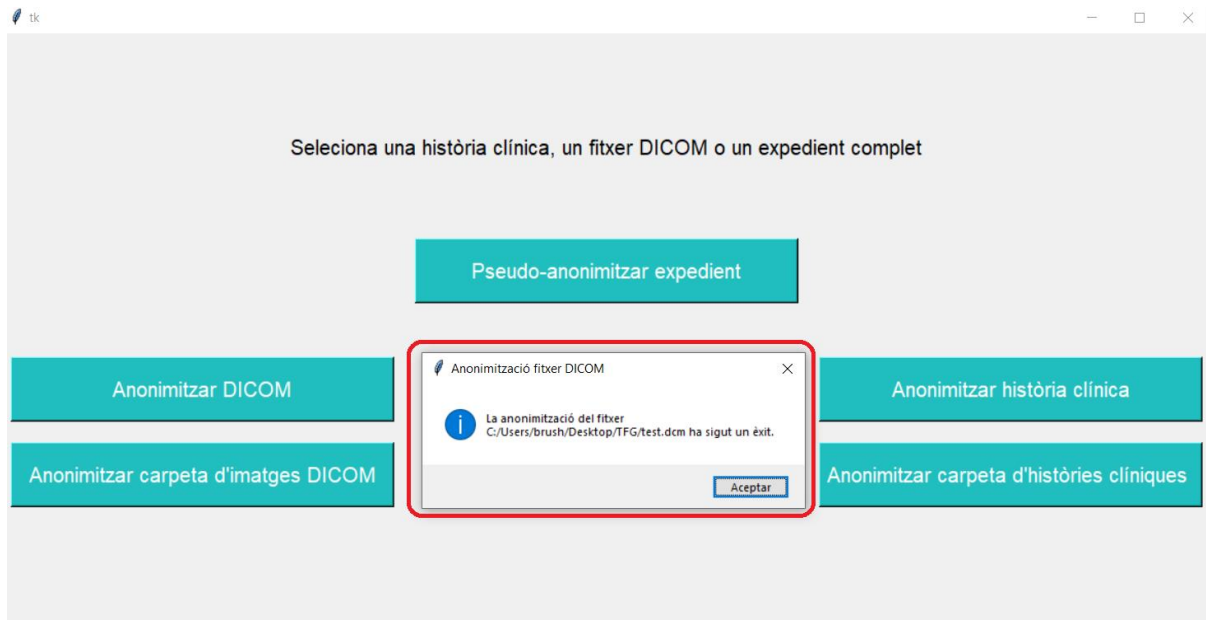
En aquest cas es crida al script de Dicom\_Anonymizer.py per anonimitzar la imatge en format DICOM.



Una vegada l'operació ha acabat, l'aplicació mostra una alerta amb el resultat d'aquesta.

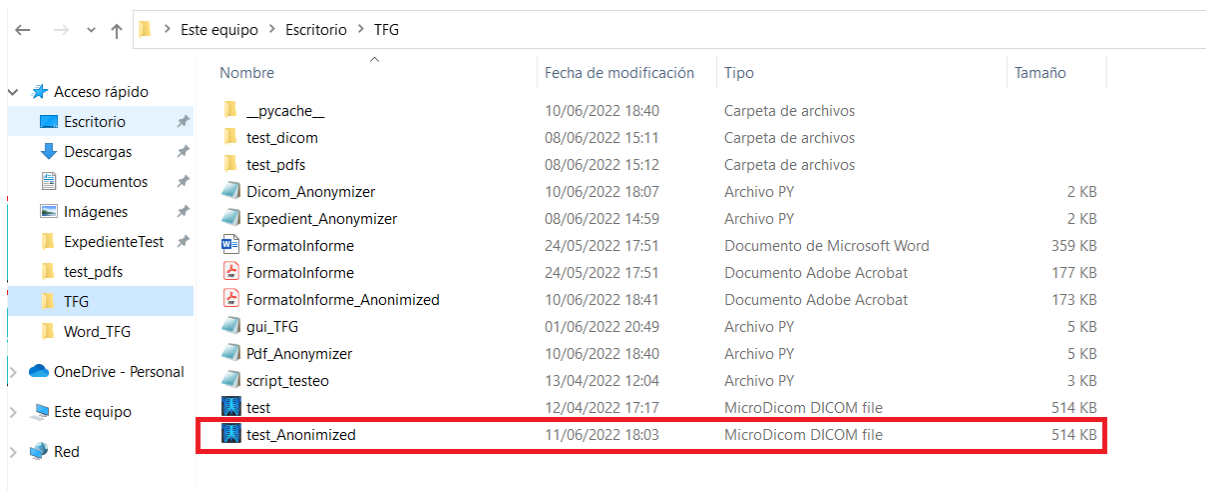
En cas que l'operació sigui un èxit mostra aquesta alerta:

Figura 22. Alerta èxit anonimització imatge DICOM



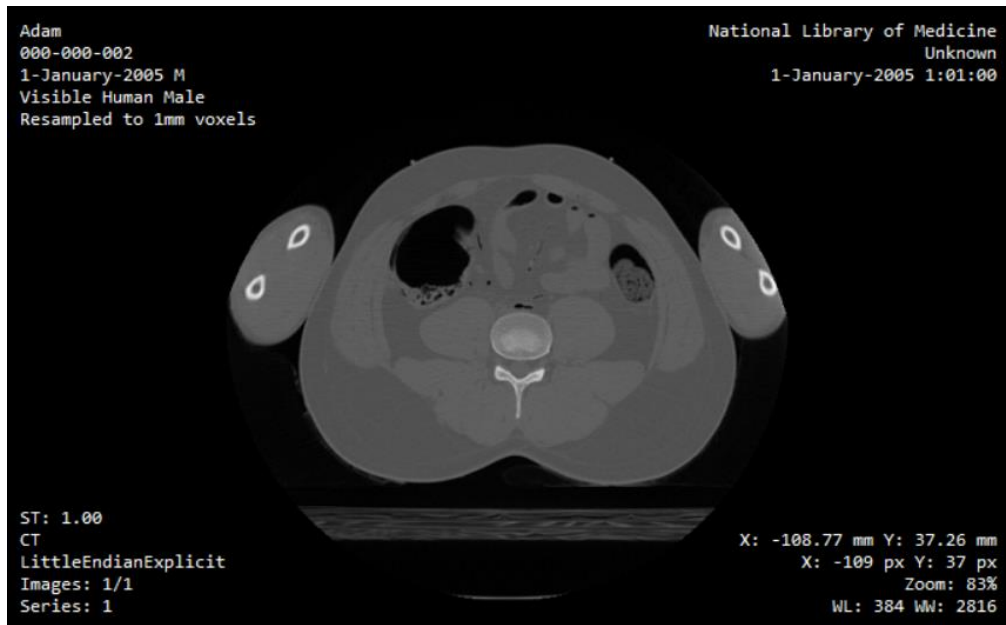
Finalment si l'operació ha sigut un èxit, dintre del directori on es troba el fitxer, l'usuari pot trobar els fitxers DICOM anonimitzats.

Figura 23. Resultat anonimització imatge DICOM



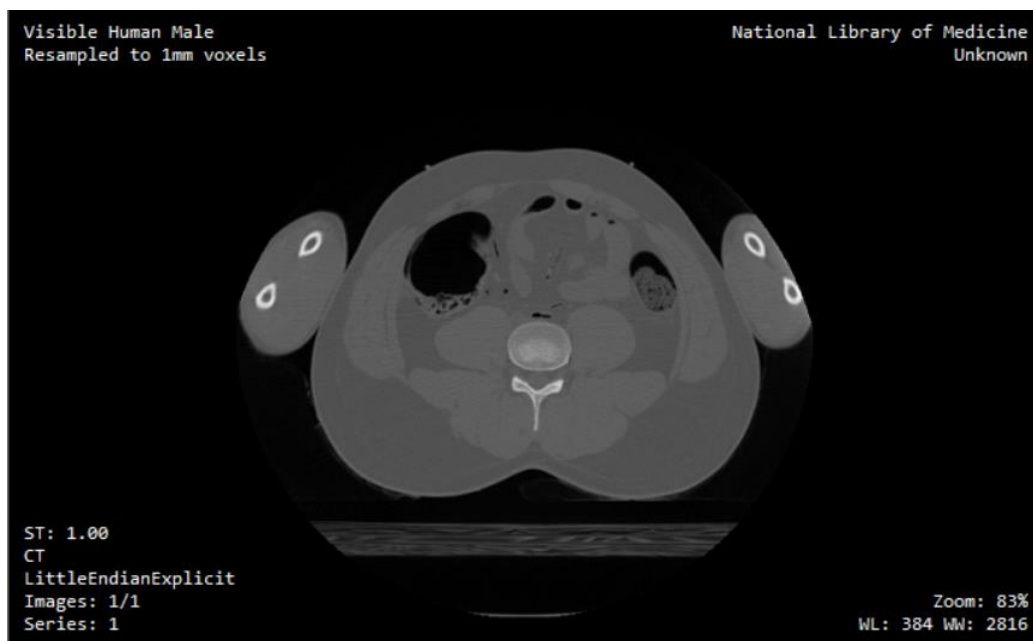
El fitxer DICOM abans de ser anonimitzat:

Resultat DICOM 1



El fitxer DICOM després de ser anonimitzat:

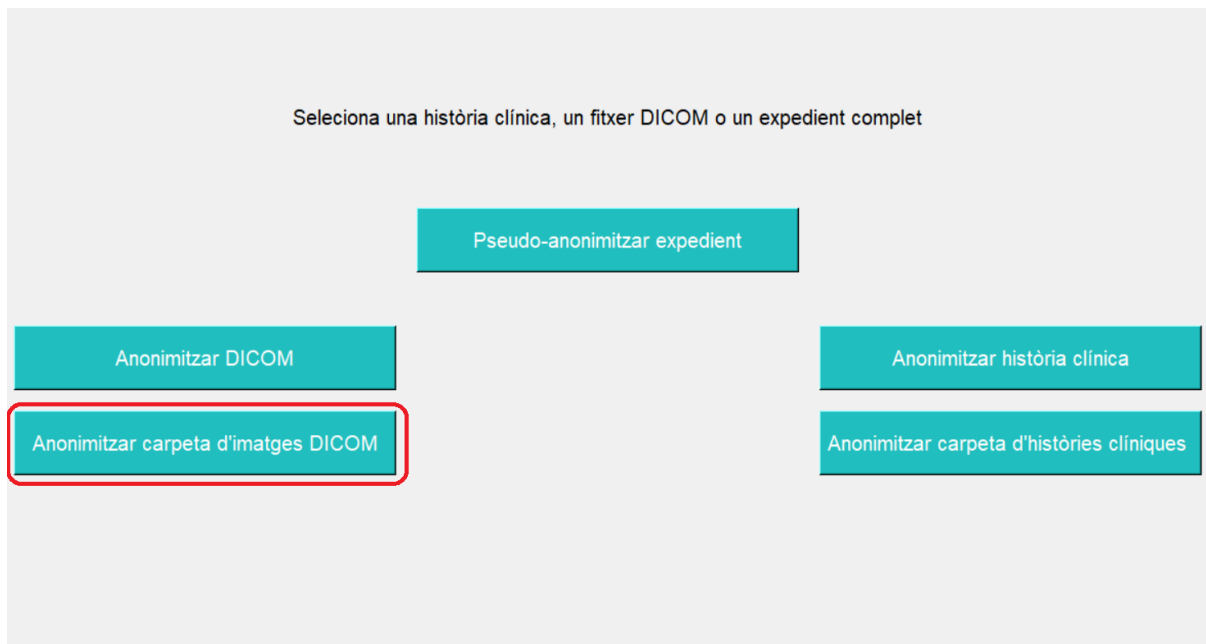
Resultat DICOM 2



### 6.6.3 Anonimitzar carpeta d'imatges DICOM

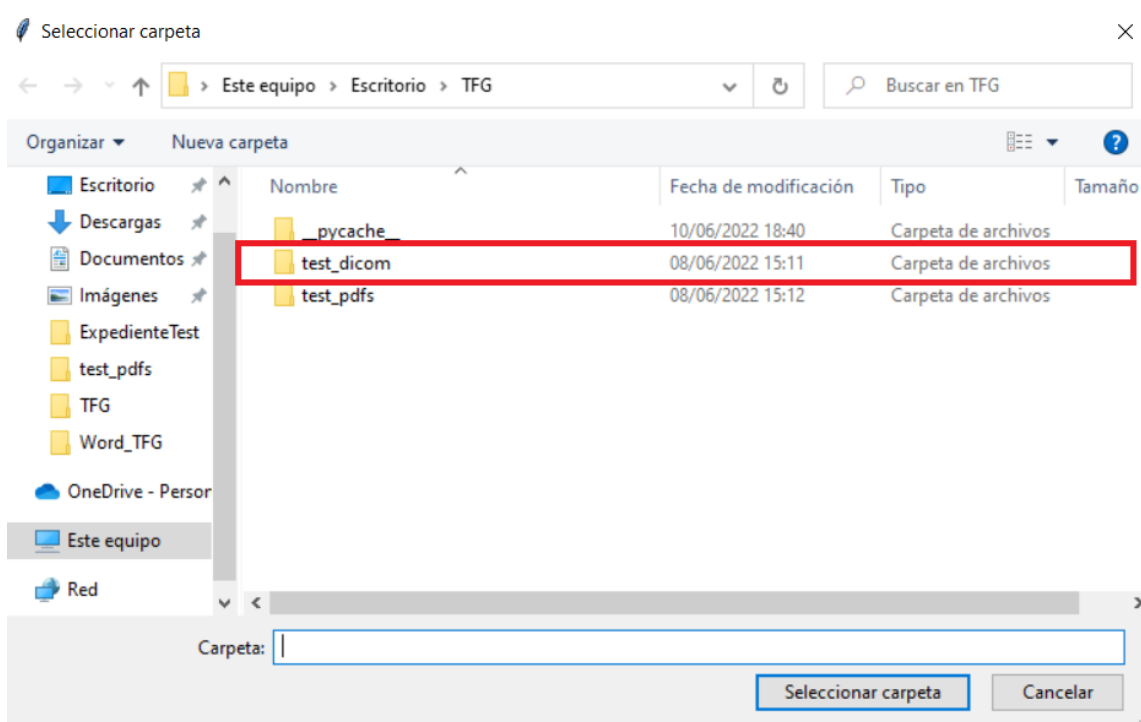
Aquest cas permet anonimitzar una carpeta sencera de fitxers de format DICOM.

Figura 24. Interficie grafica 5



Una vegada l'usuari ha clicat el botó s'obre una finestra per poder escollir el directori amb imatges DICOM que l'usuari vol anonimitzar.

Figura 25. Escollir carpeta d'imatges DICOM



Una vegada l'usuari ha seleccionat el directori, l'aplicació processa des de la interfície gràfica ,que fa de controlador, a quin script ha de cridar.

En aquest cas es crida al script de `Expedient_Anonymizer.py` concretament a la funció de `anonymize_folder()` , aquesta funció recorre tots els ítems del directori seleccionat i depèn del format del fitxer crida a un script o un altre, en aquest cas crida per cada ítem el script de `Dicom_Anonymizer.py` .

Una vegada l'operació ha acabat, l'aplicació mostra una alerta amb el resultat d'aquesta.

En cas que l'operació sigui un èxit mostra aquesta alerta:

*Figura 26. Alerta èxit anonimització carpeta d'imatges DICOM*



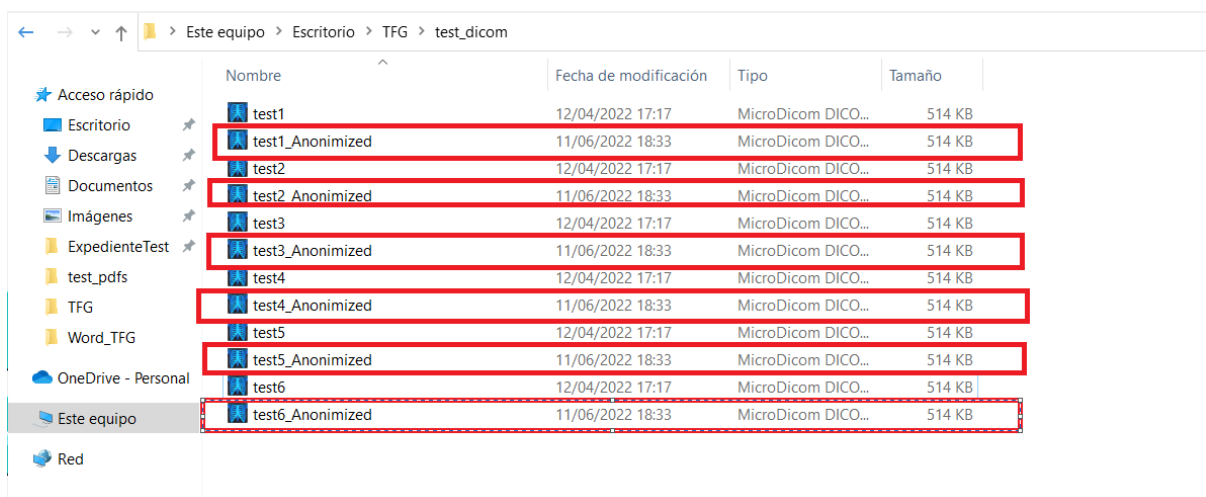
En el cas que per algun motiu l'operació d'anonimització d'un dels ítems falli es mostra aquesta alerta:

Figura 27. Alerta error anonimització carpeta d'imatges DICOM



Finalment si l'operació ha sigut un èxit, dintre del directori seleccionat per l'usuari, pot trobar tots els fitxers DICOM anonimitzats.

Figura 28. Resultat operació anonimització carpeta d'imatges DICOM



Un de les millores que s'ha plantejat per l'aplicació és fer una subcarpeta on agrupar tots els fitxers anonimitzats, ja que facilitaria el possible desplaçament de tots els fitxers

### 6.6.4 Anonimitzar història clínica

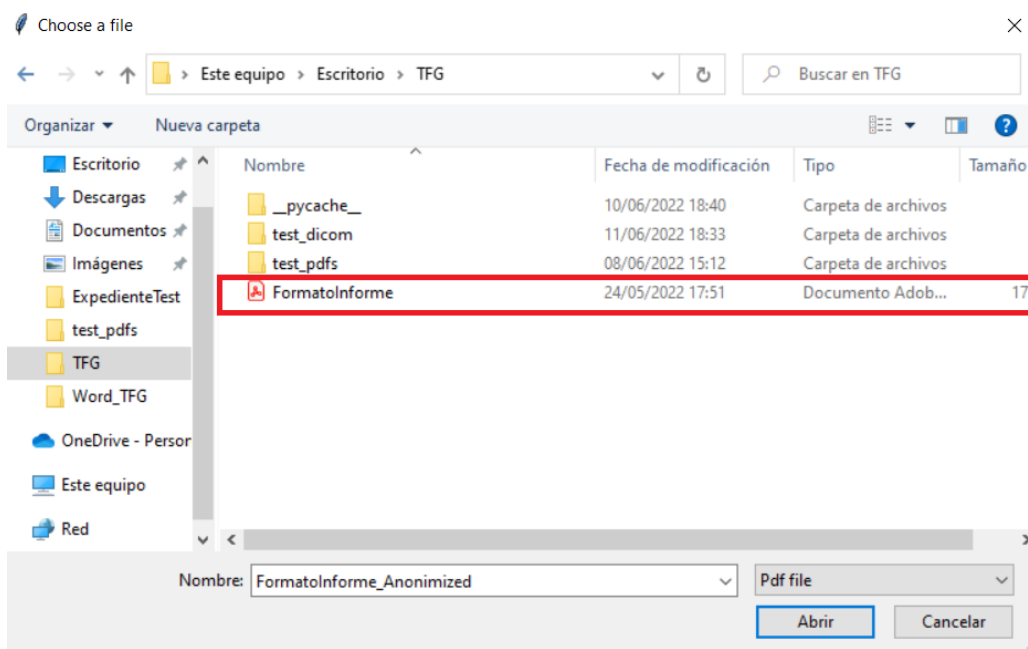
Aquest cas permet anonimitzar una història clínica d'un pacient de l'Hospital de Mataró en format PDF.

Figura 29. Interfície grafica 6



Una vegada l'usuari ha clicat el botó s'obre una finestra per poder escollir l'història clínica que l'usuari vol anonimitzar.

Figura 30. Escollir història clínica



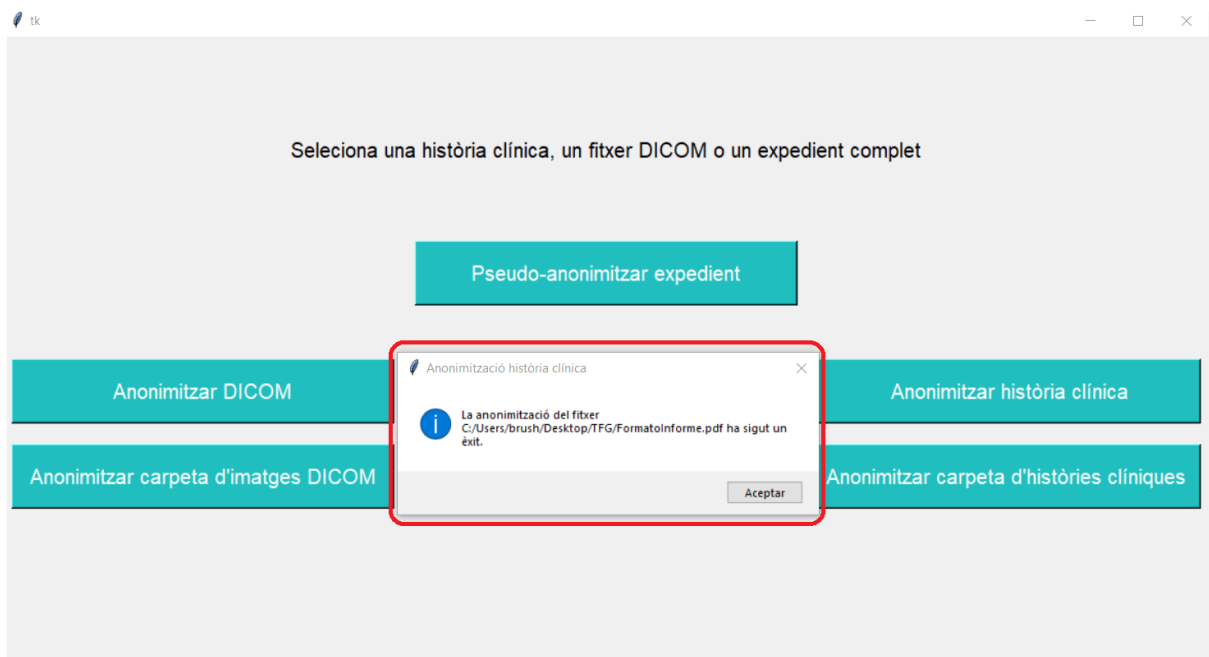
Una vegada l'usuari ha seleccionat el fitxer, l'aplicació processa des de la interfície gràfica ,que fa de controlador, a quin script ha de cridar.

En aquest cas es crida al script de Pdf\_Anonymizer.py per anonimitzar la història clínica en format PDF.

Una vegada l'operació ha acabat, l'aplicació mostra una alerta amb el resultat d'aquesta.

En cas que l'operació sigui un èxit mostra aquesta alerta:

Figura 31. Alerta èxit anonimització història clínica



Finalment si l'operació ha sigut un èxit, dintre del directori on es troba el fitxer, l'usuari pot trobar l'història clínica anonimitzada.

Figura 32. Resultat operació anonimització història clínica

Nombre	Fecha de modificación	Tipo	Tamaño
__pycache__	10/06/2022 18:40	Carpeta de archivos	
test_dicom	11/06/2022 18:33	Carpeta de archivos	
test_pdfs	08/06/2022 15:12	Carpeta de archivos	
Dicom_Anonymizer	10/06/2022 18:07	Archivo PY	2 KB
Expedient_Anonymizer	08/06/2022 14:59	Archivo PY	2 KB
FormatoInforme	24/05/2022 17:51	Documento de Microsoft Word	359 KB
FormatoInforme	24/05/2022 17:51	Documento Adobe Acrobat	177 KB
<b>FormatoInforme_Anonimized</b>	<b>11/06/2022 19:08</b>	<b>Documento Adobe Acrobat</b>	<b>173 KB</b>
gui_TFG	01/06/2022 20:49	Archivo PY	5 KB
Pdf_Anonymizer	10/06/2022 18:40	Archivo PY	5 KB
script_testeo	13/04/2022 12:04	Archivo PY	3 KB
test	12/04/2022 17:17	MicroDicom DICOM file	514 KB
test_Anonimized	11/06/2022 18:03	MicroDicom DICOM file	514 KB

El fitxer PDF abans de ser anonimitzat:

Figura 33. Història clínica pre anonimització

1 / 1

**HOSPITAL DE MATARÓ**  
CONSORCI SANITARI DEL MARESME  
Carretera de Clotxa, 36 - 08301 Mataró  
Tel: 937 417 700 Fax: 937 417 730

NO NO, NO

**Cognoms:** Sanchez de la Corte      **Adreça:** C. de la Tramuntana 12-14 baixos 1a  
**Nom:** Xavier  
**Edat:** 22      **Sexe:** M      **CIP:** ABC1234567894      **NIF:** 43584128W  
**Data naixement:** 29/05/2000      **NHC:** 123456      **NASS:** 08/1123564789  
**Telèfon / Mòbil:** 629631477 / 0

---

**INFORME D'ASSISTÈNCIA**

**Metge/essa:** Juan Fernandez Ortega      **Servei:** Medicina interna  
**Data ingress:** 14/05/2014      **Data Alta:** 14/05/2015  
**S'informa a:** Metge de família      **Núm. lliç:** 903 - 70

---

**DIAGNÒSTICS:**  
1.

**PROCEDIMENTS:**  
1.

---

**ANTECEDENTS PATOLÒGICS:**  
Tractament habitual:

**MOTIU D'INGRÉS:**

**EXPLORACIÓ FÍSICA:**

**EXPLORACIONS COMPLEMENTÀRIES:**  
 Anàlítica:  
 ECG:  
 Rx de tòrax:

**EVOLUCIÓ:**  
**TRACTAMENT:**

MEDICACIÓ	ESMORZAR	DINAR	SOPAR	NIT


**CONTROL:**



El fitxer PDF després de ser anonimitzat:

Figura 34. Història clínica post anonimització

Evaluation Only. Created with Aspose.Words. Copyright 2003-2022 Aspose Pty Ltd.

 **HOSPITAL DE MATARÓ**  
C/DOCTOR MARTEL 181. 08431 MATARÓ  
 Carretera de Clua, s/n - 08494 MATARÓ  
 T. 937 417 700 Fax 937 417 733

1/1  
NO NO, NO

Cognoms: \_\_\_\_\_ Adreça: \_\_\_\_\_  
 Nom: \_\_\_\_\_  
 Edat: 22 Sexe: M CIP: \_\_\_\_\_ NIF: \_\_\_\_\_  
 Data naixement: \_\_\_\_\_ NHC: \_\_\_\_\_ NASS: \_\_\_\_\_  
 Telèfon / Mòbil: / 0

**INFORME D'ASSISTÈNCIA**

Metge/essa: \_\_\_\_\_ Servei: \_\_\_\_\_  
 Data ingress: \_\_\_\_\_ Data Alta: \_\_\_\_\_  
 S'informa a: Metge de família Núm. lli: \_\_\_\_\_

**DIAGNÒSTICS:**

**PROCEDIMENTS:**

**ANTECEDENTS PATOLÒGICS:**

**MOTIU D'INGRÉS:**

**EXPLORACIÓ FÍSICA:**

**EXPLORACIONS COMPLEMENTÀRIES:**

Anàlítica:

**TRACTAMENT:**

MEDICACIÓ	ESMORZAR	DINAR	SOPAR	NIT

Created with an evaluation copy of Aspose.Words. To discover the full versions of our APIs please visit: <https://products.aspose.com/words/>

## 6.6.5 Anonimitzar carpeta d'històries clíniques

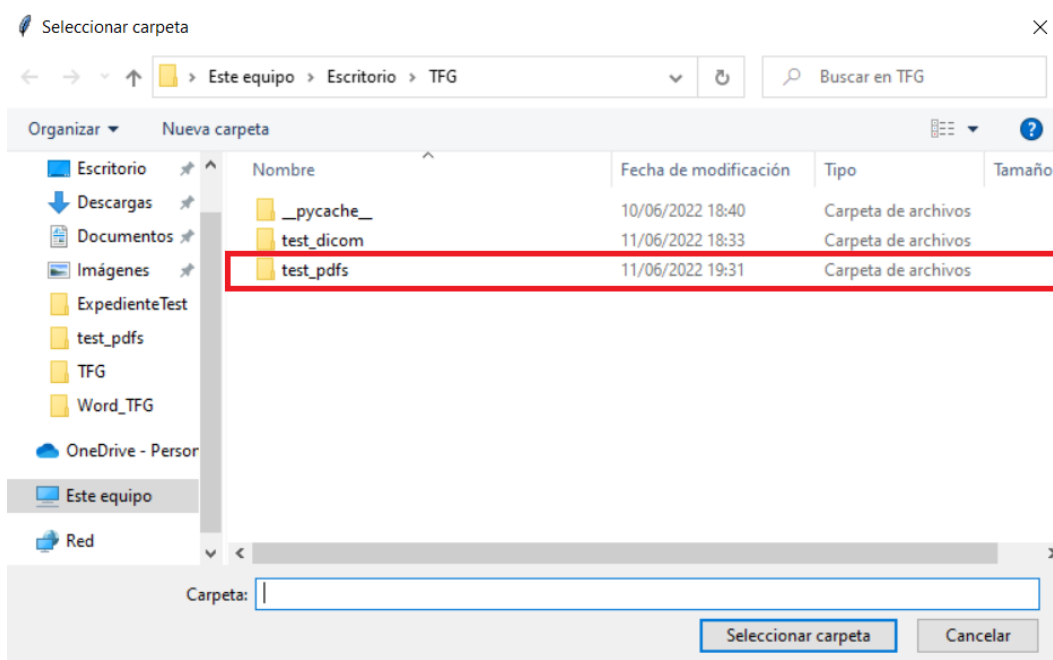
Aquest cas permet anonimitzar una carpeta sencera d'històries clíniques de l'Hospital de Mataró de format PDF.

Figura 35. Interficie grafica 7



Una vegada l'usuari ha clicat el botó s'obre una finestra per poder escollir el directori amb històries clíniques que l'usuari vol anonimitzar.

Figura 36. Escollir carpeta d'històries clíniques



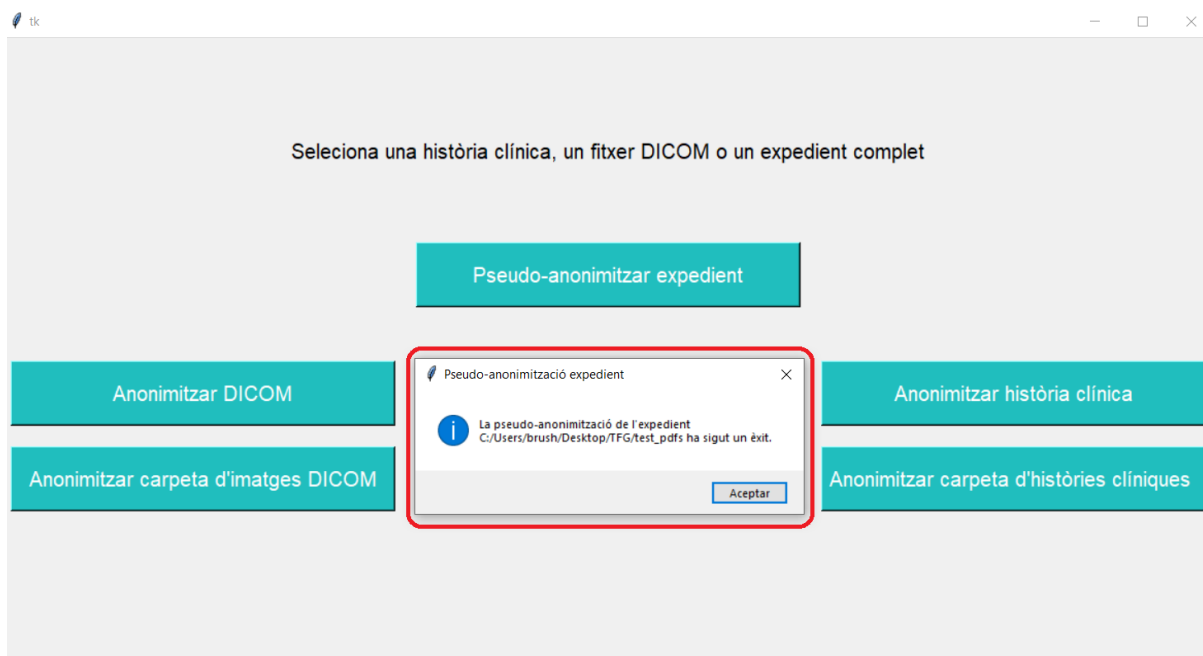
Una vegada l'usuari ha seleccionat el directori, l'aplicació processa des de la interfície gràfica, que fa de controlador, a quin script ha de cridar.

En aquest cas es crida al script de `Expedient_Anonymizer.py` concretament a la funció de `anonymize_folder()`, aquesta funció recorre tots els ítems del directori seleccionat i depèn del format del fitxer crida a un script o un altre, en aquest cas crida per cada ítem el script de `Pdf_Anonymizer.py`.

Una vegada l'operació ha acabat, l'aplicació mostra una alerta amb el resultat d'aquesta.

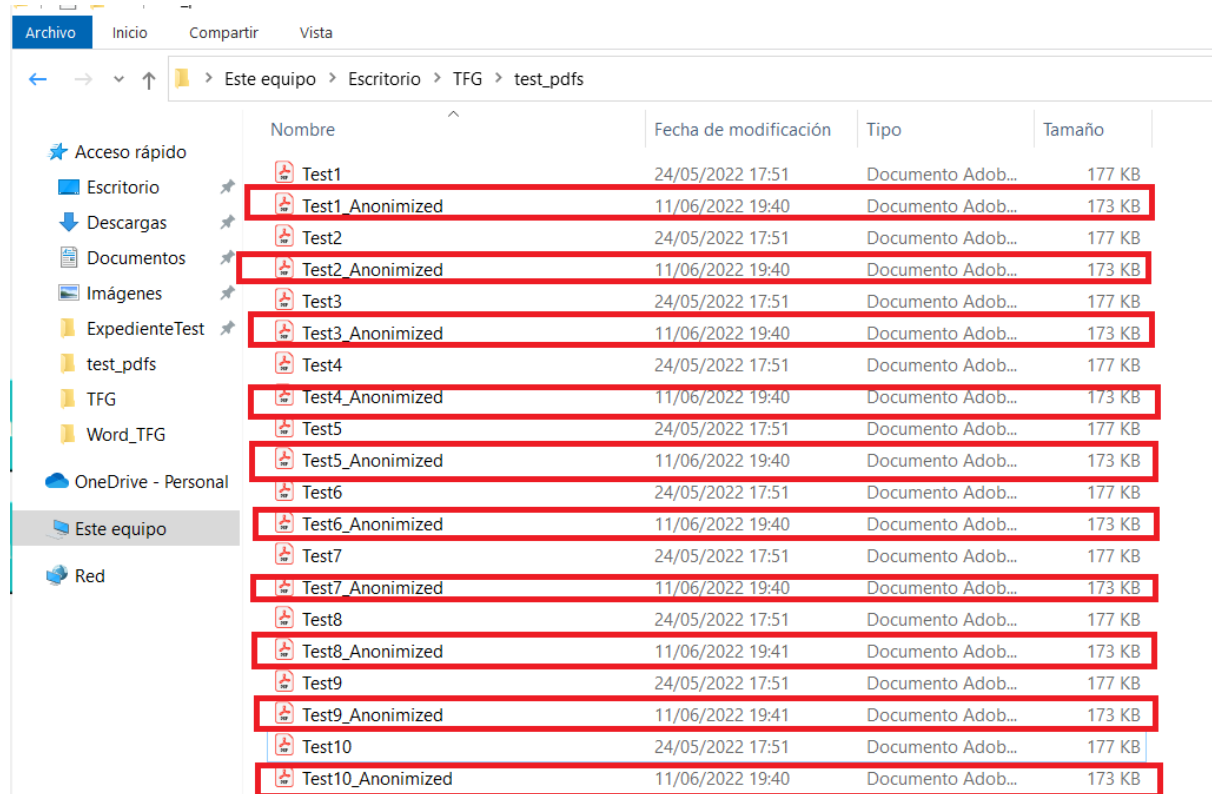
En cas que l'operació sigui un èxit mostra aquesta alerta:

*Figura 37. Alerta èxit anonimització carpeta històries clíniques*



Finalment si l'operació ha sigut un èxit, dintre del directori seleccionat per l'usuari, pot trobar totes les històries clíniques anonimitzades.

Figura 38. Resultat operació anonimització carpeta històries clíniques



Nombre	Fecha de modificación	Tipo	Tamaño
Test1	24/05/2022 17:51	Documento Adob...	177 KB
Test1_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB
Test2	24/05/2022 17:51	Documento Adob...	177 KB
Test2_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB
Test3	24/05/2022 17:51	Documento Adob...	177 KB
Test3_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB
Test4	24/05/2022 17:51	Documento Adob...	177 KB
Test4_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB
Test5	24/05/2022 17:51	Documento Adob...	177 KB
Test5_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB
Test6	24/05/2022 17:51	Documento Adob...	177 KB
Test6_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB
Test7	24/05/2022 17:51	Documento Adob...	177 KB
Test7_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB
Test8	24/05/2022 17:51	Documento Adob...	177 KB
Test8_Anonimized	11/06/2022 19:41	Documento Adob...	173 KB
Test9	24/05/2022 17:51	Documento Adob...	177 KB
Test9_Anonimized	11/06/2022 19:41	Documento Adob...	173 KB
Test10	24/05/2022 17:51	Documento Adob...	177 KB
Test10_Anonimized	11/06/2022 19:40	Documento Adob...	173 KB



## 7. Conclusió

### 7.1 Anàlisi de resultats

Els resultats del projecte han sigut tot un èxit, ja que assoleix tots els objectius que s'han proposat des del començament.

Els encarregats de l'Hospital de Mataró n'estan molt contents i volen seguir amb el projecte per dur a terme una solució final molt més completa i per aconseguir-ho volen comptar amb mi per integrar-lo amb els seus sistemes i acabar de desenvolupar-ho.

### 7.2 Aspectes a millorar

Com s'ha comentat anteriorment aquest projecte no deixa de ser un prototip per tant te bastants aspectes a millorar. Sobretot a l'hora d'intentar cobrir el màxim número de tipus de pacients possibles.

Amb tipus de pacients el que es vol dir és que els patrons Regex no cobreix tots els documents nacionals de tots els països, només els d'Espanya i el mateix passa amb el codi de la targeta sanitària i el número de la seguretat social.

Per tant, hi ha bastants punts on continuar evolucionant el projecte per tindre una efectivitat i seguretat molt més alta en els resultats d'aquest.



## 8. Bibliografia

[1] Carles San José Amat.(Abril, 2019). Anonimització i pseudonimització de les dades [en línia] [consulta: 20 de novembre de 2021 ]. Disponible a <https://formaciooberta.eapc.gencat.cat/espaistematics/gestio-dades/anonimitzacio-i-pseudonimitzacio-de-les-dades.html>

[2] Autoritat Catalana de Protecció de dades.(Abril,2016). Reglament general de protecció de dades. Disponible a: <https://apdcat.gencat.cat/ca/autoritat/normativa/reglament-general-de-proteccio-de-dades/>

[3] eCityclic.(Juliol, 2020). El que has de saber sobre la seudonimització de les dades [en línia] [consulta: 20 de novembre de 2021 ]. Disponible a <https://www.ecityclic.com/ca/noticies/el-que-has-de-saber-sobre-la-seudonimitzacio-de-les-dades>

[4] Khandoker Tanjim Ahammad.(Abril, 2021). Data Anonymization Tools [en línia] [consulta: 22 de novembre de 2021 ]. Disponible a <https://github.com/Leibniz-HBI/Social-Media-Observatory/issues/68>

[5] Francis P., Probst Eide S., Munz R. (2017) Diffix: High-Utility Database Anonymization. In: Schweighofer E., Leitold H., Mitrakas A., Rannenber K. (eds) Privacy Technologies and Policy. APF 2017. Lecture Notes in Computer Science, vol 10518. Springer, Cham. [https://doi.org/10.1007/978-3-319-67280-9\\_8](https://doi.org/10.1007/978-3-319-67280-9_8)

[6] Khaled El Emam, Luk Arbuckle .(December,2013).Anonymizing Health Data. ISBN: 9781449363079.



- [7] Fernando Tablado.(Gener, 2020). K-anonimidad – ¿Qué es? ¿Métodos? ¿En qué consiste? [en línia] [consulta: 24 de noviembre de 2021 ]. Disponible a [K-anonimidad – ¿Qué es? ¿Métodos? ¿En qué consiste?](#)
- [8] Agencia Española de Protección de Datos.(Gener, 2016). Orientaciones y garantías en los procedimientos de Anonimización de datos personales [en línia] [consulta: 25 de noviembre de 2021 ]. Disponible a [AEPD-Orientaciones procedimientos anonimización de datos](#)
- [9] OMOP [en línia] [consulta: 16 de Diciembre de 2021]. Disponible a [OMOP Common Data Model – OHDSI](#)
- [10] SNOMED [en línia] [consulta: 16 de Diciembre de 2021]. Disponible a [SNOMED](#)
- [11] Jaime Domínguez. FHIR, ese gran desconocido [en línia] [consulta: 19 de Diciembre de 2021 ]. Disponible a [FHIR, ese gran desconocido](#)
- [12] Thomas Neubauer, Johannes Heurix,A methodology for the pseudonymization of medical data, International Journal of Medical Informatics, Volume 80, Issue 3, 2011, Pages 190-204, ISSN 1386-5056, <https://doi.org/10.1016/j.ijmedinf.2010.10.016> .
- [13] Kruse, C.S., Smith, B., Vanderlinden, H. *et al.* Security Techniques for the Electronic Health Records. *J Med Syst* **41**, 127 (2017). <https://doi.org/10.1007/s10916-017-0778-4> .
- [14] Xiao-Bai Li, Jialun Qin (2017) Anonymizing and Sharing Medical Text Records. Information Systems Research 28(2):332-352. <https://doi.org/10.1287/isre.2016.0676>
- [15] Matt Wes.(Abril, 2017). Looking to comply with GDPR? Here's a primer on anonymization and pseudonymization [en línia] [consulta: 3 de Gener de 2022 ]. Disponible a <https://iapp.org/news/a/looking-to-comply-with-gdpr-heres-a-primer-on-anonymization-and-pseudonymization/>

- [16]. Advances in Model and Data Engineering in the Digitalization Era MEDI 2021 International Workshops: DETECT, SIAS, CSMML, BIOG, HEDA, Tallinn, Estonia, June 21–23, 2021, Proceedings: MEDI 2021 International Workshops: DETECT, SIAS, CSMML, BIOG, HEDA, Tallinn, Estonia, June 21–23, 2021, Proceedings. DOI:[10.1007/978-3-030-87657-9](https://doi.org/10.1007/978-3-030-87657-9) . ISBN: 978-3-030-87656-2.
- [17] Zaask[en linia][consulta: 5 de Febrer de 2022]. Disponible a <https://www.zaask.es/cuanto-cuesta/programador-php>