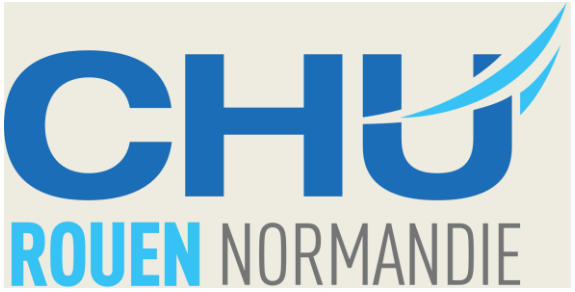
Entrepôt de données de santé



Département d'informatique et d'information médicale, CHU de Rouen
LIMICS U1142 INSERM, Sorbonne Université

Médecins

Stéfan J. Darmoni

Mikaël Dusenne

Jean-Philippe Leroy

Thibaut Pressat-Laffouilhère

Pharmaciens

Laura Gosselin

Romain Leguillon

Catherine Letord

Ingénieurs

Badisse Dahamna

Flavien Disson

Kévin Eyer

Julien Grosjean

Ivan Kergourlay

Romain Lelong

Objectifs pédagogiques

- Connaitre la définition d’un EDS
- Connaitre les objectifs principaux des EDS
- Appréhender la façon de construire et d’exploiter un EDS
- Appréhender les avantages et limites des EDS

EDS - Qu'est ce que c'est

- Les Entrepôts de Données de Santé (EDS) sont des outils informatiques permettant la collection, l'intégration puis le traitement des données de santé provenant d'un grand nombre de sources d'information clinique (dossier patient informatisé, système d'information des laboratoires et d'imagerie, prescription informatisée, dossier infirmier…).

- Agrégation d'un **maximum** d'informations disponibles sur les patients quelle que soit l'application source

➔ Permet de croiser ces informations et donc de sélectionner finement des patients/données

Objectifs des EDS

- Améliorer la recherche interventionnelle grâce aux études de faisabilité d'essais cliniques et l'optimisation des inclusions (*pré-screening*) ;
- Améliorer la recherche observationnelle (registres et cohortes) ;

- Détecter des profils de santé particuliers pour la prévention (primaire ou secondaire) ;
- Créer et maintenir des outils de détection d'événements liés à la vigilance (par exemple, les infections nosocomiales, pharmacovigilance…) ;
- Développer des indicateurs d'activité plus précis et variés ;
- Renforcer la prise en compte des indicateurs de qualité ;
- Améliorer le codage PMSI, par la détection semi-automatique d'atypie entre le codage réalisé et les données issues de l'EDS ;
- Participer à la pédagogie médicale ;

- Développer et évaluer les algorithmes d'aide à la décision.

Aujourd'hui dans les CH

- Large utilisation du Programme de Médicalisation des Systèmes d'Information (PMSI) via CIM-10 / CCAM
- Pour la recherche : « usage détourné » d'un système médico-économique
- Très pratique et fiable dans certains cas…
- …mais beaucoup de biais et approximations

État des lieux des EDS

Dans le monde :

- i2b2 (Harvard MS)
- …

En France :

- ehop (CHU de Rennes)
- Dr Warehouse (Necker, Foch – Paris)
- ConSoRe Continuum Soins Recherche (Centre de Lutte Contre le Cancer)
- EDSaN …

Approche rouennaise pour l'EDS

- Sémantique :
  - Apport d'un lexique en français extrêmement riche (HeTOP)
  - RI améliorée (subsomption/explosion)
- Technologique :
  - Temps de réponses plus bas possible pour une utilisation en temps réel/interactive
    - Annotateur sémantique ECMT (BOW)
      - 6 ms/doc de santé vs. 134 ms/doc Quick UMLS vs. plusieurs secondes pour Metamap
      - 13 M de documents de santé en 22 h => possibilité de lancer ECMT chaque semaine
    - Moteur de recherche
      - I2B2 assez lent ; pas du tout temps réel
      - Doc'EDS ou ASIS : temps de réponse < 2s ; possibilité de tester de nombreuses hypothèses de requêtes lors d'une session de 2 heures

EDS Rouen

- Projet diligenté par la Direction Générale du CHU en octobre 2017
- Première version « draft » en mars 2018
- Premières demandes des cliniciens dès mars 2018 (tests)
- 3 outils = **EDSaN (Entrepôt de Données de Santé Normand)**
- **Accord CNIL : octobre 2020**
- Publication princeps :

Pressat-Laffouilhère, Thibaut; Balayé, Pierre; Dahamna, Badisse; Lelong, Romain; Billey, Kévin; Darmoni, Stéfan J. & Grosjean, Julien. Evaluation of Doc'EDS: a French semantic search tool to query health documents from a clinical data warehouse. **BMC medical informatics and decision making** , February, Volume 22, Number 1, Pages 34, 2022.

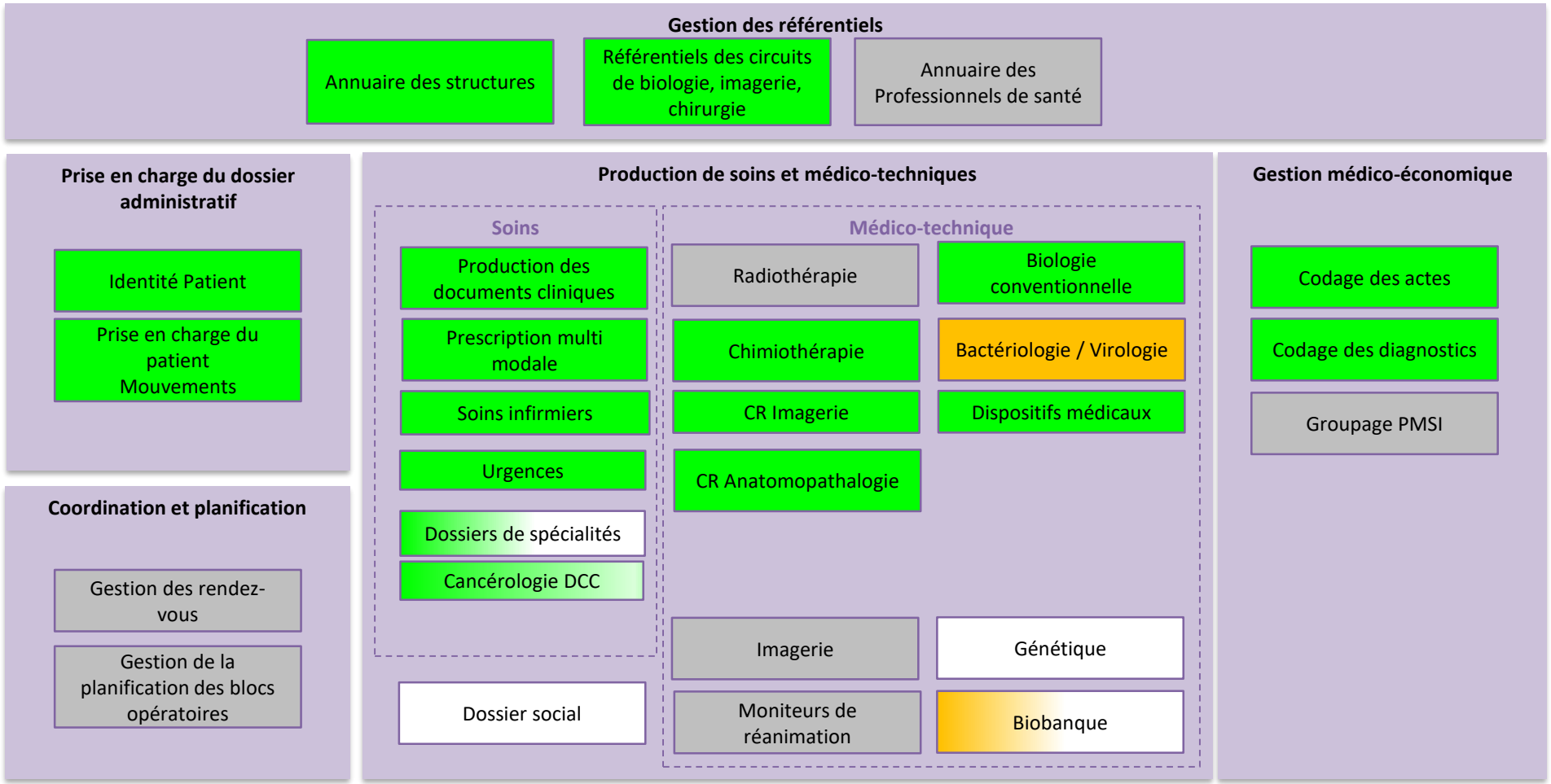
EDSaN au CHU de Rouen

- Volumétrie

Patients	2 millions
Séjours (hospitalisations, séances, consultations)	22,3 millions
Documents (CR, notes infirmières, lettres, ordonnances...)	21,4 millions
Administrations médicamenteuses (en hospitalisation)	1,8 millions
Résultats de biologie (hématologie, biochimie...)	176 millions
Dispositifs médicaux (prothèses...)	116 000
Diagnostics (CIM-10)	11,6 millions
Actes (imagerie, chirurgie...) (CCAM)	10,7 millions

- Antériorité : patients venus au moins une fois au CHU de Rouen les 20 dernières années

EDS Rouen – couverture fonctionnelle



Domaines métiers présents dans EDS

Domaines métier à intégrer en 2022

Domaines métiers hors périmètres EDS

Schéma adapté de l'APHP

Aspects méthodologiques et techniques

EDSaN

Système d'Information Hospitalier

- DPI
- Imagerie
- PMSI
- Biologie
- Documents
- ...

Programmes ETL / ELT (*Extract Transform Load, Extract Load Transform*)

Normalisation dé-identification

Données dé-identifiés

Applications Web / *Datamarts*

Outil de recherche modulaire (1 module = 1 type de données)

- ✓ Documents
- ✓ PMSI
- ✓ Prescriptions
- ✓ Biologie
- ✓ Microbiologie
- ✓ DM

Outil de files actives (recherche & prévention)

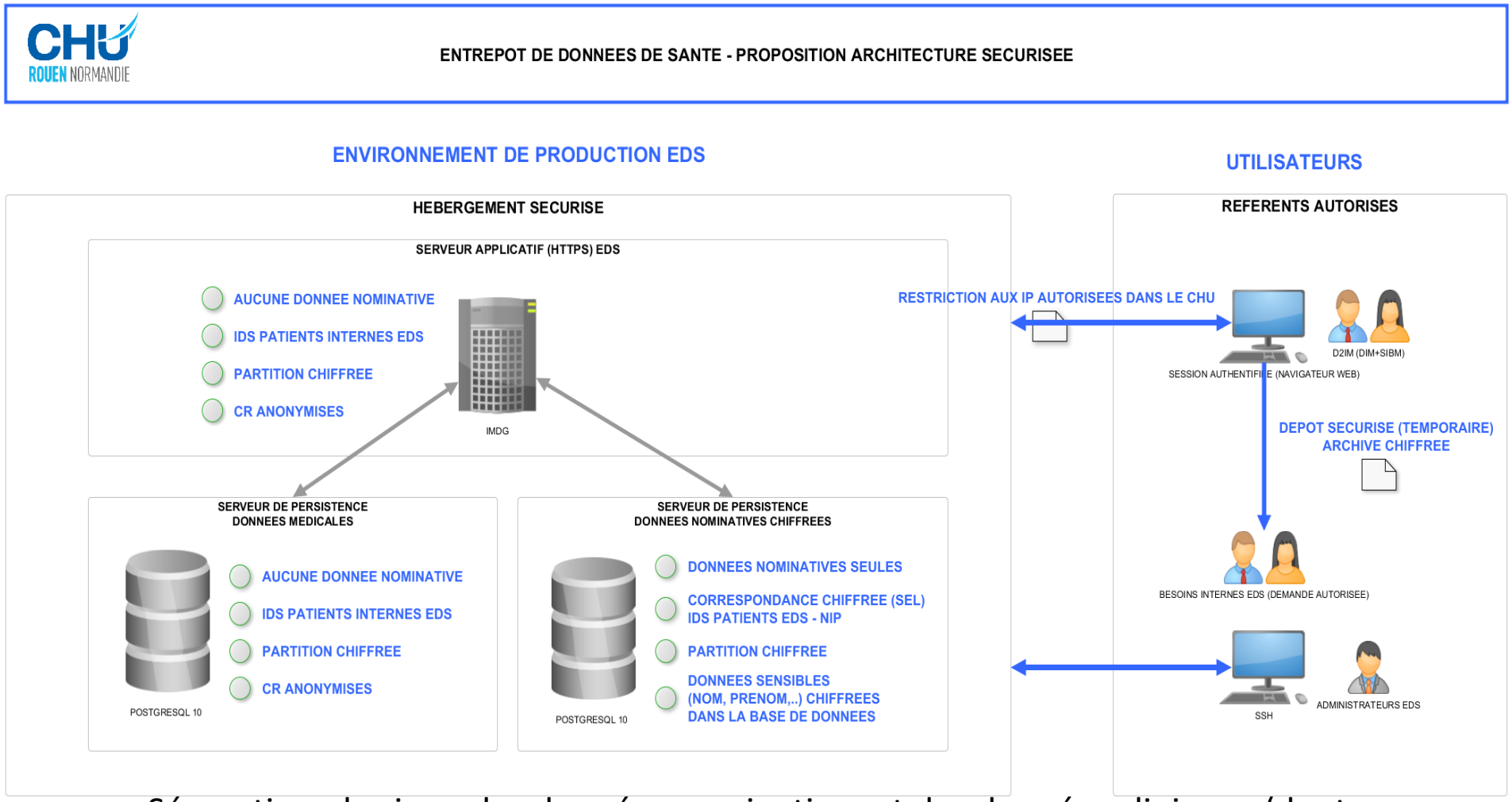
Outil de pilotage (indicateurs de qualité, activité...)

Traitement automatique de la langue

HeTOP **Traitement et apports sémantiques** ***ECMT***

NoSQL

Architecture technique EDSaN



- Séparation physique des données nominatives et des données cliniques (dont documents textuels dé-identifiés)
- Chiffrement des données (tables de correspondances identifiant « public » / identifiant « privé ») ; réversibilité via procédure particulière et par le personnel agréé

Politique d'accès à EDSaN

https://edsan.chu-rouen.fr/edsan/acces-aux-donnees/procedures/

1. Étude de chaque demande par le Comité Scientifique et Éthique de l'établissement
2. Rendez-vous « physique » avec un membre du D2IM pour créer la (ou les) requête(s)
   - Raisons réglementaire et éthique : l'outil de requête donne accès à toutes les données
   - Raison scientifique : biais des données, biais des outils...
   - Accompagnement pour « délester » le chercheur
3. Accès aux données délimitées par la (ou les) requête(s) dans un outil dédié (accès nominatif)
4. Exports possibles par le Comité d'Export (personnels agréés du D2IM)

Résultats & Outils

Plateforme EDSaN

EDSaN

D2IM Département d'Informatique et d'Information Médicales

CHU ROUEN NORMANDIE

Consulter vos jeux de données

Cet environnement est consacré à la sélection des données ou des patients afin de créer des cohortes ou des tableaux de données à partir de l'EDSaN. Pour cela, une ou plusieurs requêtes doivent avoir été élaborées et exécutées pour pouvoir rendre accessible les données à explorer.

Accéder à votre espace personnel

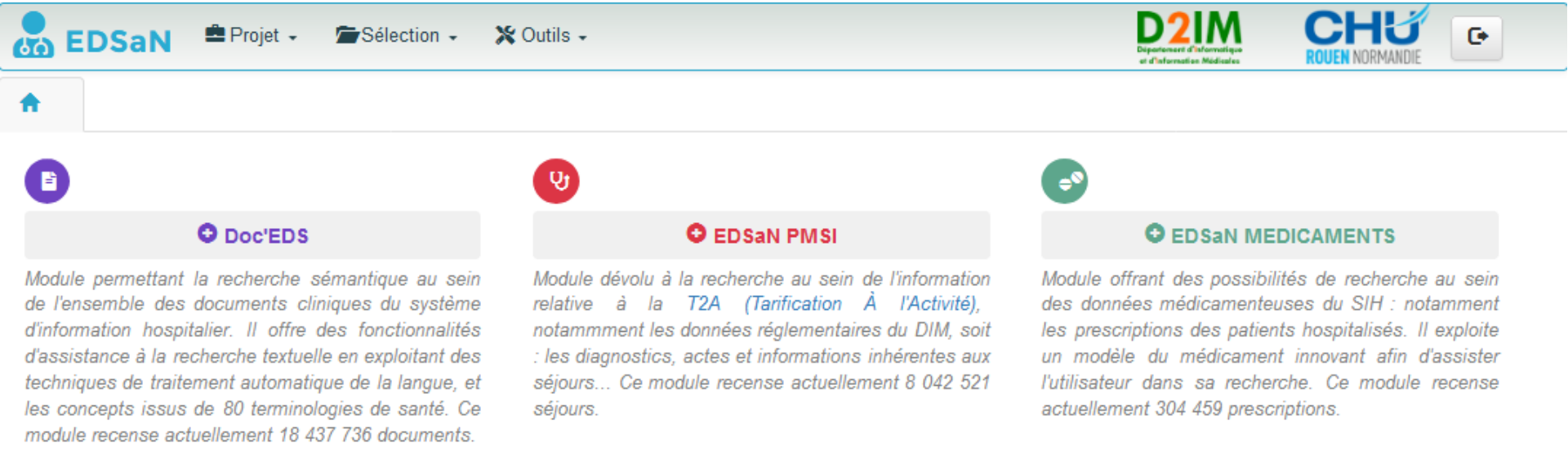
Créer ou ouvrir des requêtes

Uniquement accessible par les administrateurs EDSaN, cet environnement permet de créer des requêtes pour quelles soient notamment disponibles à la consultation par les chercheurs.

Créer une requête

Outil modulaire de recherche

- 1 module = 1 type de données
- Disponible à ce jour :
  - Documents textuels (80% information clinique pertinente[1])
  - PMSI (mouvements, codages diagnostics et actes médicaux)
  - Prescriptions/administrations médicamenteuses dans l'établissement
  - Dispositifs médicaux
  - Biologie « conventionnelle »



[1] Raghavan P, Chen JL, Fosler-Lussier E, Lai AM. How essential are unstructured clinical narratives and information fusion to clinical trial recruitment? AMIA Jt Summits Transl Sci Proc. 2014 Apr 7;2014:218-23. PMID: 25717416; PMCID: PMC4333685.

Les fonctions techniques des EDS

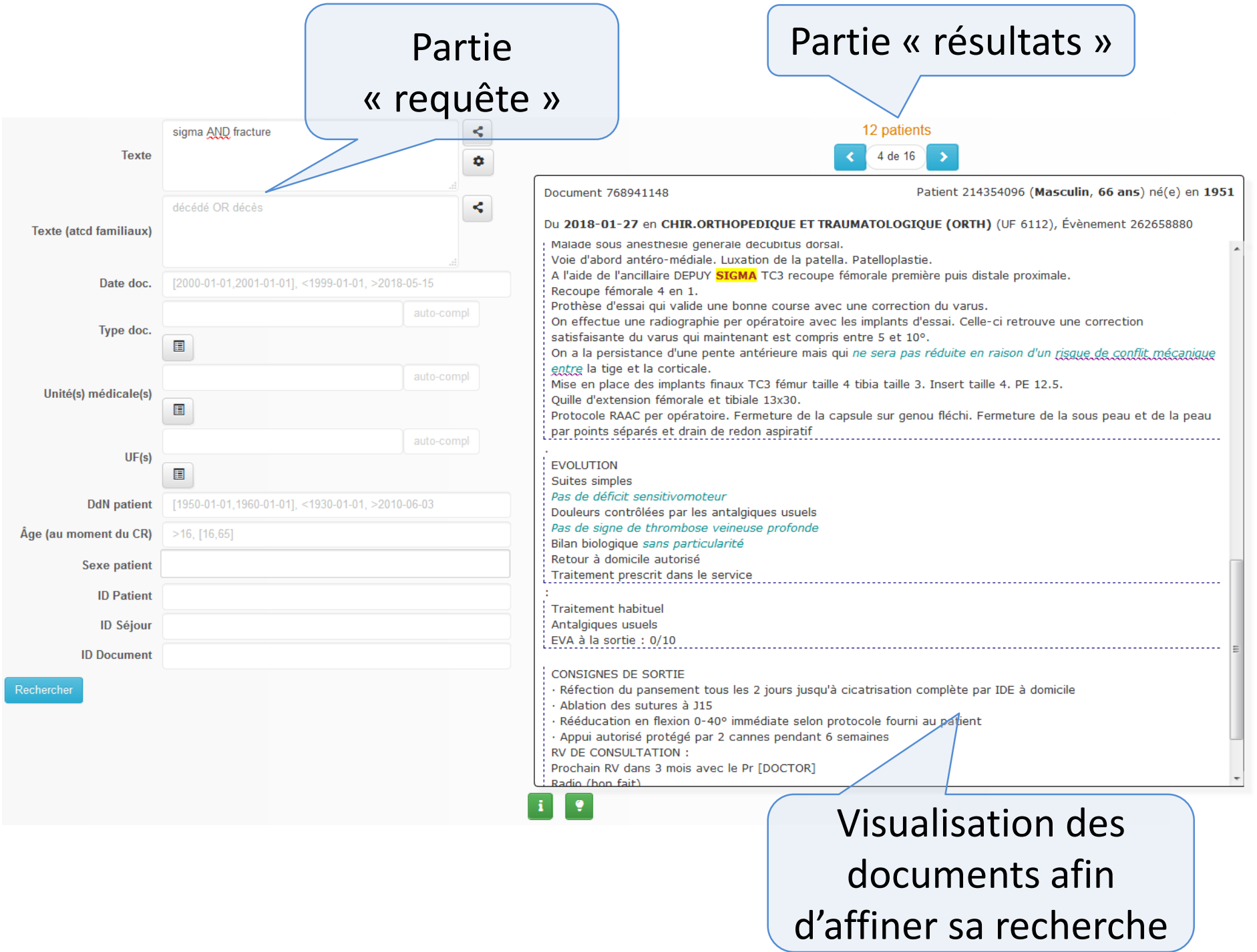
- 3 outils complémentaires :
  - Effectuer des requêtes
  - Consulter et trier des dossiers patients
  - Tableaux de bord

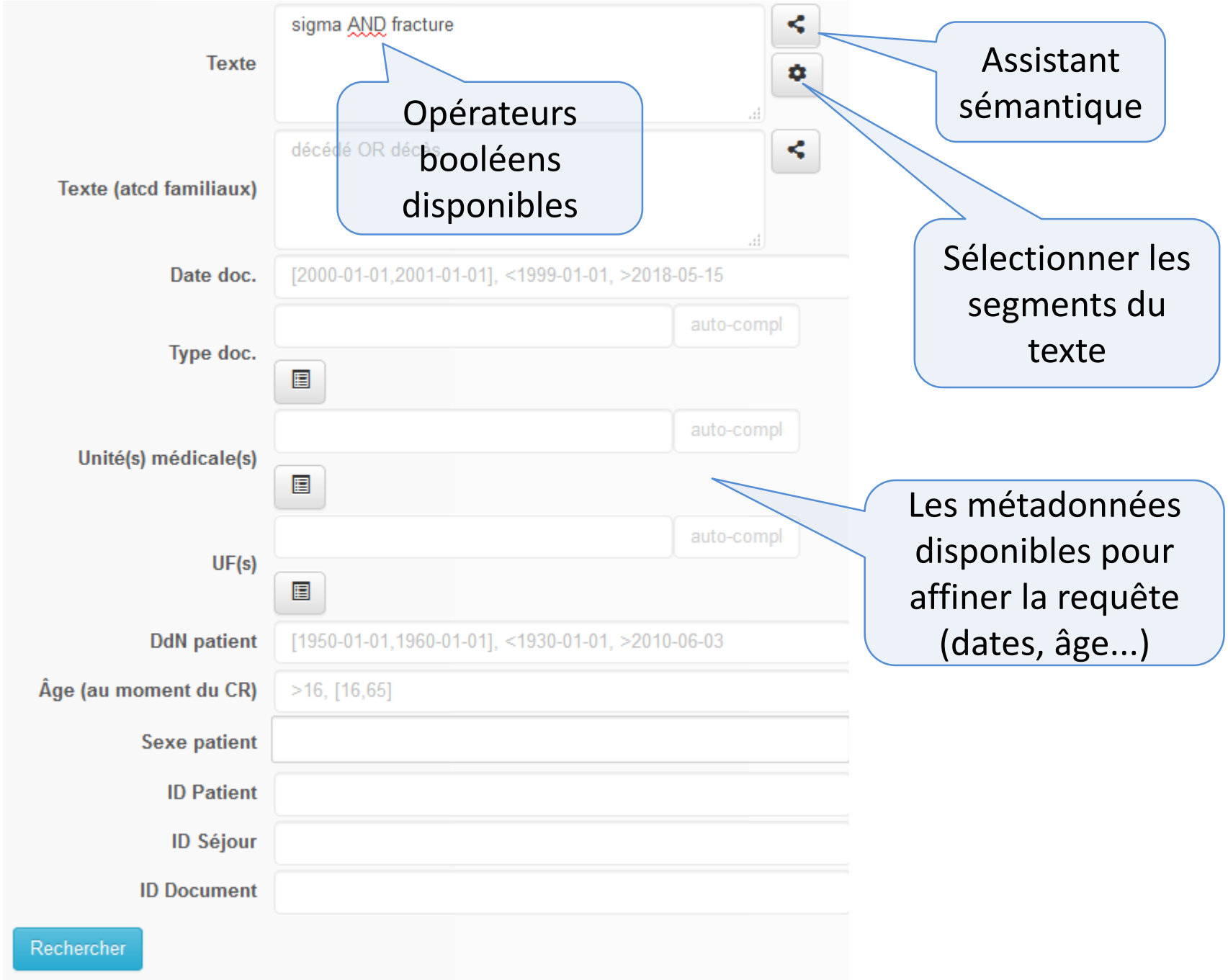
- Utilisation d'assistants sémantiques (outil d'amont)
- Statistiques / exports / visualisations / *data mining /* extraction de valeurs *(feature extraction)* (outils d'aval)

Analyses textuels (module Doc'EDS)

Doc'EDS (module document)

- 80% de l'information clinique se trouve dans les documents textuels (comptes rendus, lettres, transmissions infirmières...)
- Il faut prendre en compte la structure du document et le discours (Traitement Automatique du Langage Naturel ou *NLP*) :
  - Négation
  - Condition/suspicion/hypothèse/futur
  - Antécédents familiaux
  - Segmentation du document


Partie « requête »
Partie « résultats »
Texte
sigma AND fracture
Texte (atcd familiaux)
décédé OR décès
Date doc.
[2000-01-01,2001-01-01], <1999-01-01, >2018-05-15
Type doc.
auto-compl
Unité(s) médicale(s)
auto-compl
UF(s)
auto-compl
DdN patient
[1950-01-01,1960-01-01], <1930-01-01, >2010-06-03
Âge (au moment du CR)
>16, [16,65]
Sexe patient
ID Patient
ID Séjour
ID Document
Rechercher
12 patients
4 de 16
Document 768941148
Patient 214354096 (Masculin, 66 ans) né(e) en 1951
Du 2018-01-27 en CHIR.ORTHOPEDIQUE ET TRAUMATOLOGIQUE (ORTH) (UF 6112), Évènement 262658880
Malade sous anesthesie generale decubitus dorsal.
Voie d'abord antéro-médiale. Luxation de la patella. Patelloplastie.
A l'aide de l'ancillaire DEPUY SIGMA TC3 recoupe fémorale première puis distale proximale.
Recoupe fémorale 4 en 1.
Prothèse d'essai qui valide une bonne course avec une correction du varus.
On effectue une radiographie per opératoire avec les implants d'essai. Celle-ci retrouve une correction satisfaisante du varus qui maintenant est compris entre 5 et 10°.
On a la persistance d'une pente antérieure mais qui ne sera pas réduite en raison d'un risque de conflit mécanique entre la tige et la corticale.
Mise en place des implants finaux TC3 fémur taille 4 tibia taille 3. Insert taille 4. PE 12.5.
Quille d'extension fémorale et tibiale 13x30.
Protocole RAAC per opératoire. Fermeture de la capsule sur genou fléchi. Fermeture de la sous peau et de la peau par points séparés et drain de redon aspiratif
.
EVOLUTION
Suites simples
Pas de déficit sensitivomoteur
Douleurs contrôlées par les antalgiques usuels
Pas de signe de thrombose veineuse profonde
Bilan biologique sans particularité
Retour à domicile autorisé
Traitement prescrit dans le service
:
Traitement habituel
Antalgiques usuels
EVA à la sortie : 0/10
CONSIGNES DE SORTIE
· Réfection du pansement tous les 2 jours jusqu'à cicatrisation complète par IDE à domicile
· Ablation des sutures à J15
· Rééducation en flexion 0-40° immédiate selon protocole fourni au patient
· Appui autorisé protégé par 2 cannes pendant 6 semaines
RV DE CONSULTATION :
Prochain RV dans 3 mois avec le Pr [DOCTOR]
Radio (bon fait)
Visualisation des documents afin d'affiner sa recherche


Texte
sigma AND fracture
Texte (atcd familiaux)
décédé OR décès
Date doc.
[2000-01-01,2001-01-01], <1999-01-01, >2018-05-15
Type doc.
auto-compl
Unité(s) médicale(s)
auto-compl
UF(s)
auto-compl
DdN patient
[1950-01-01,1960-01-01], <1930-01-01, >2010-06-03
Âge (au moment du CR)
>16, [16,65]
Sexe patient
ID Patient
ID Séjour
ID Document
Rechercher
Opérateurs booléens disponibles
Assistant sémantique
Sélectionner les segments du texte
Les métadonnées disponibles pour affiner la requête (dates, âge...)

RAMI[DOCTOR] 5 mg, un par jour
FUROSEMIDE 20 mg, un par jour
SELOKEN 200 mg, 1 par jour
SINVASTATINE 40 mg, un par jour
INNOHEP 14000, une injection par jour
IMOVANE 7.5 mg, un au coucher
VESICARE 5 mg, un par jour
UROREC 8 mg, un par jour

MODE DE VIE
Marié (femme aidante)

HISTOIRE DE LA MALADIE :
Patient de [redacted] ans hospitalisé le matin [redacted]/04 à Becquerel pour une cure de radiothérapie dans le cadre de la prise en charge d'un carcinome épidermoide de **l'oesophage** suivit au CHU en gastro-entérologie (médecin référent : [DOCTOR]).
Le patient décrit des expectorations mêlées de sang survenues en deuxième moitié de nuit (nuit du [redacted] au [redacted]/04). Un second épisode de crachats sanglants après effort de toux au cours de l'évaluation pré radiothérapie. Dans le contexte de **tumeur** de **l'oesophage**, le patient est transféré en unité de soins intensifs de gastro-entérologie au CHU pour une probable hématémèse.

EXAMEN CLINIQUE :
Taille : 176 cm : 76 kg. IMC : 24.5 kg/m². Température : 36.4°C Pouls : 77 par mn. TA : 109/70. *EVA : 0.* Hb : 12.6 g/Dl. Bonne tolérance hémodynamique de l'hémorragie, *pas de déglobulisation*
Dénutrition. OMS 2-3.
Conjonctives colorées. *Absence d'ictère.*
Cardiovasculaire : BDC irrégulières. *Pas de souffle. Pas d'OMI Pas de Turgescence des jugulaires.*
Respiratoire : Râles crépitants à la base pulmonaire gauche. Toux importantes suivies d'expectorations teintées de sang.
Digestif : présence d'une sonde naso-gastrique d'alimentation entérale. Abdomen souple, indolore. BHA présents et normaux. *Absence d'argument en faveur d'une ascite clinique.* Toucher rectal : selles dures de couleur normale (pas de méléna *ni de rectorragies*). Prostate de taille augmentée avec disparition du sillon médian.
Neurologique : bonne orientation temporo-spatiale. *Pas de déficit sensitivo-moteur.*

BIOLOGIE :
Hémoglobine à 12,3 g/dl. .Sérologie et antigénurie aspergillaire en cours.

RADIOLOGIE :
Radio de thorax :
TDM du [redacted]02/20[redacted] : Stabilité de l'épaississement circonférentiel oesophagien avec multiples adénomégalies médiastinales.Apparition de plusieurs nodules lobaires inférieures gauches :

Anonymisation (patients et médecins)

Détection des suspicions/hypothèses/doutes/futur

Détection des négations/infirmations

SCHEMA DE TRAITEMENT : Cacis (Campo + Cisplatine) + radiothérapie.

CIBLES MESURABLES :
Tumeur primitive : oesophage.
Adénopathies : oui.

ANAMNESE (depuis la précédente hospitalisation) :

Antécédents : chirurgie sur le ménisque droit. Chirurgie sur varices superficielles des membres inférieurs. Terrain atopique : eczéma.
Frère décédé d'un cancer du poumon il y a 30 ans.
Oncle décédé d'un cancer de l'oesophage il y a 16 ans.

EXAMEN CLINIQUE :
Indice de performance : grade OMS 0. Poids : 89 kg. Surf. corp.: 2 m2.

EXAMENS COMPLEMENTAIRES :

Biologie : GB : 6600 G/L PN : 3890 G/L Plaquettes : 272000 G/L Hb : 13,5 g/dl. TP : 99 %.
Urée : 0,24. Créat : 7,3 mg/l. TCA : 1,18.

TRAITEMENT RECU :
Chimiothérapie :
- Campo 60 mg/m2/j soit 120 mg/j (soit 100 % de la dose standard).
- Cisplatine 30 mg/m2/j soit 60 mg/j (soit 100 % de la dose standard).
- Traitement anti-émétique préventif par Zophren.
- Traitement anti-émétique curatif par Solumédrol.

Radiothérapie : à prévoir.

« Segmentation » du document

Détection des antécédents familiaux

Différents bilans

Statistiques & bilans

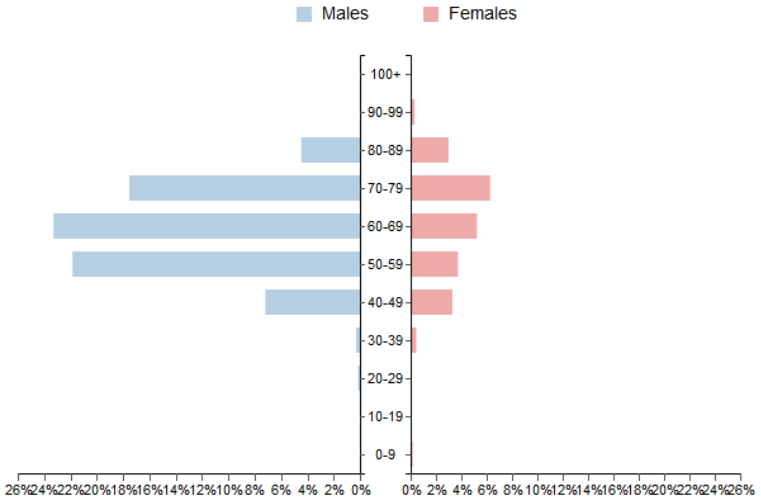
- Stats démographiques
- Date doc.
- Type doc.
- Unité(s) médicale(s)
- UF(s)
- DdN patient
- Age (au moment du CR)
- Sexe patient
- Code(s) acte(s)
- Code(s) diag(s)

150 femmes (22,4%)

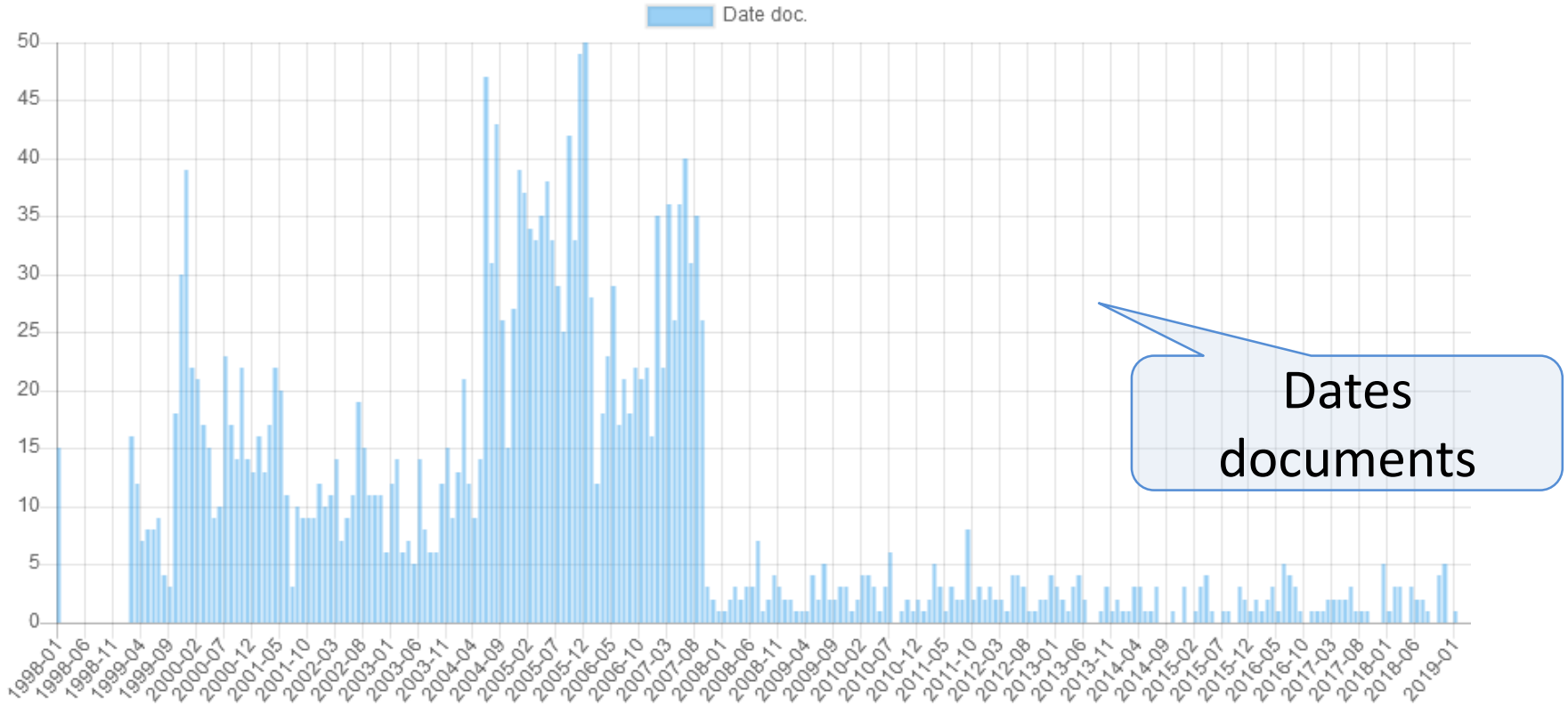
520 hommes (77,6%)

Statistiques des âges (au moment du CR)

Moyenne	63,3
Écart type	11,5
Minimum	7
Maximum	93
Q1	54
Q2	63
Q3	72

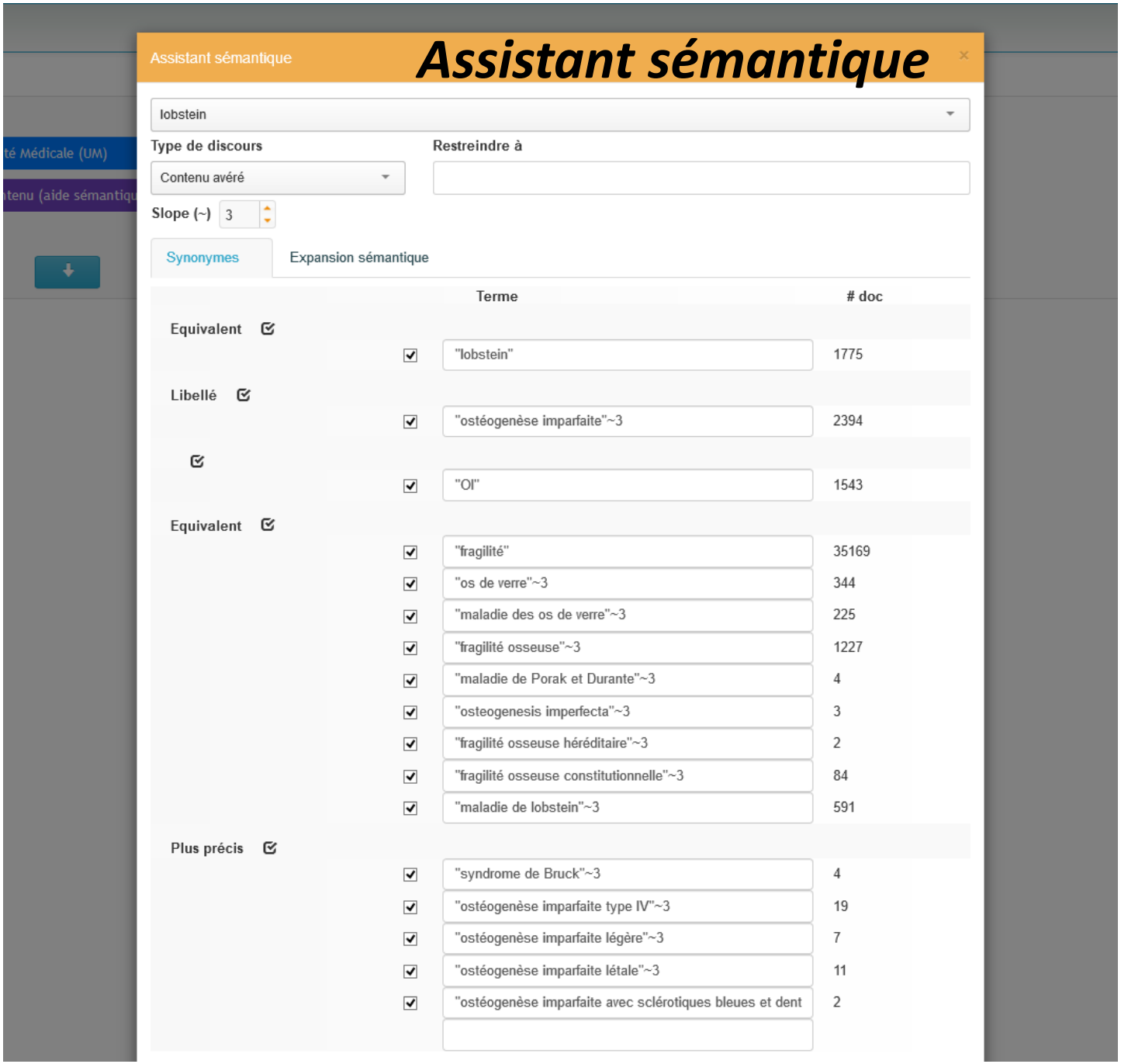


Population

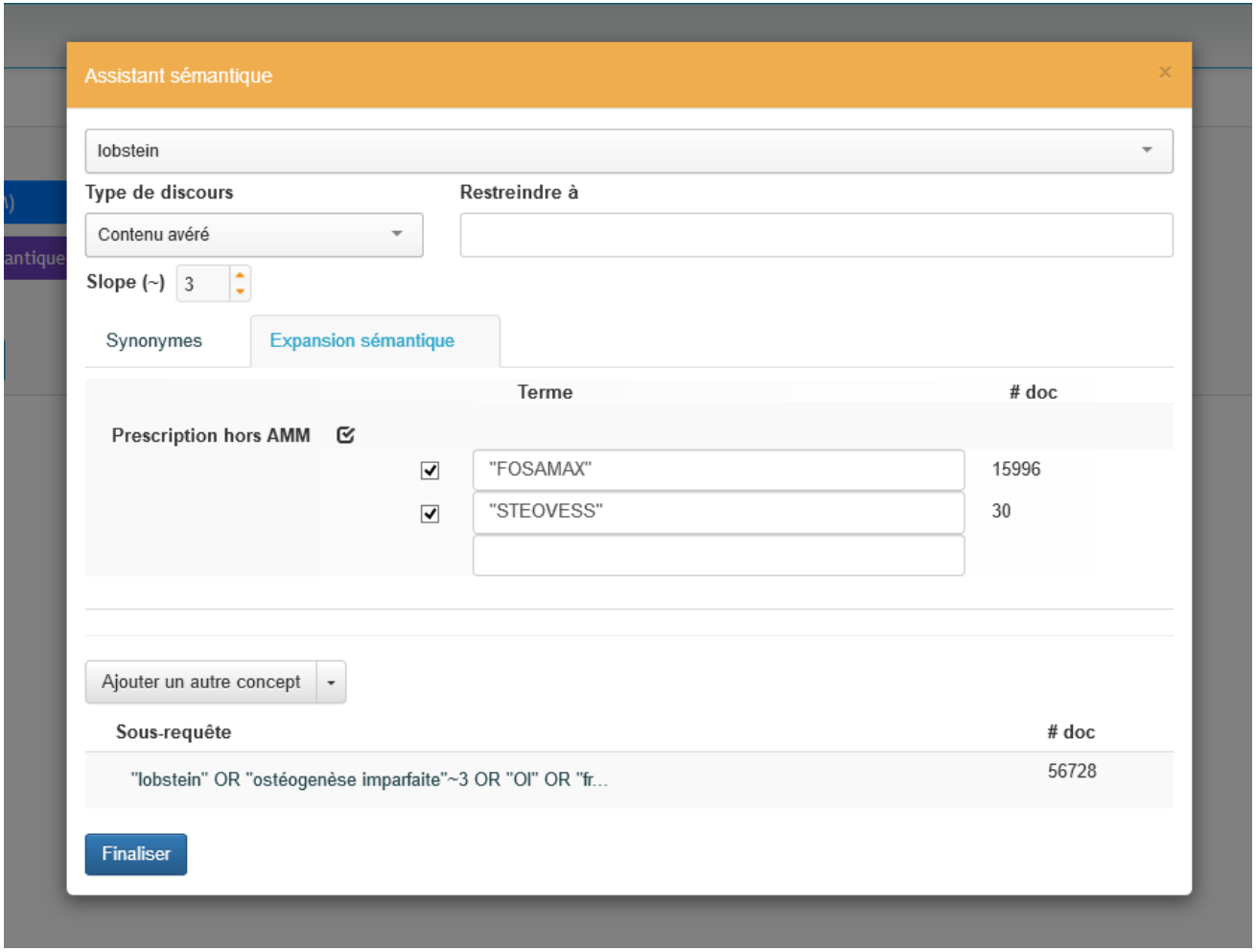


SEJUM	#	%
DIGE HEPATO GASTRO ENTEROLOGIE NUTRITION	1685	61,9%
CGCD CHIRURGIE GENERALE ET DIGESTIVE	235	8,6%
RADI IMAGERIE CENTRALE	149	5,5%
URGE URGENCES	128	4,7%
PHIE PHARMACIE	93	3,4%
ORLO O.R.L ADULTES	61	2,2%
PNM1 CLINIQUE PNEUMOLOGIQUE HCN	59	2,2%
PHYS PHYSIOLOGIE DIGESTIVE	34	1,2%
CARD CARDIOLOGIE	25	0,9%
REAC REANIMATION CHIRURGICALE	23	0,8%

Répartition dans les unités médicales / services

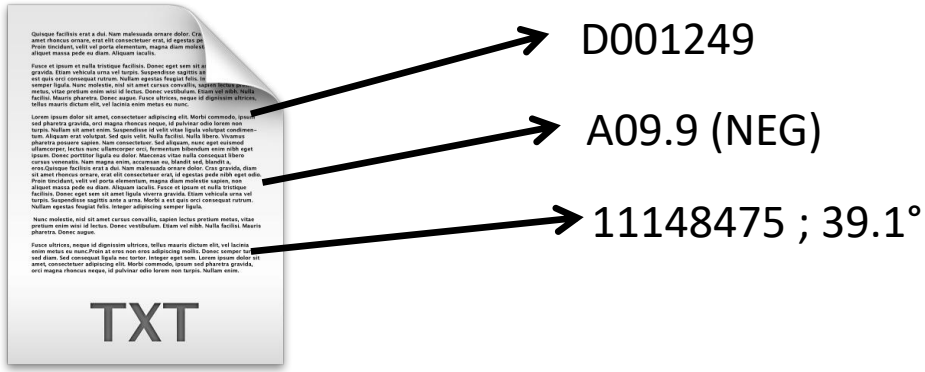

Assistant sémantique
Assistant sémantique
lobstein
Type de discours
Contenu avéré
Restreindre à
Slope (~) 3
Synonymes
Expansion sémantique
Terme
doc
Equivalent
"lobstein" 1775
Libellé
"ostéogenèse imparfaite"~3 2394
"OI" 1543
Equivalent
"fragilité" 35169
"os de verre"~3 344
"maladie des os de verre"~3 225
"fragilité osseuse"~3 1227
"maladie de Porak et Durante"~3 4
"osteogenesis imperfecta"~3 3
"fragilité osseuse héréditaire"~3 2
"fragilité osseuse constitutionnelle"~3 84
"maladie de lobstein"~3 591
Plus précis
"syndrome de Bruck"~3 4
"ostéogenèse imparfaite type IV"~3 19
"ostéogenèse imparfaite légère"~3 7
"ostéogenèse imparfaite létale"~3 11
"ostéogenèse imparfaite avec sclérotiques bleues et dent 2

Assistant sémantique



ECMT : annotation automatique des documents textuels

- Extracteur de Concepts Multi-Terminologique



- Utilisation en aval (documents d'un groupe de patients)
- Utile pour le *data mining*, observer des tendances ou structurer des informations précises (médicaments...)

Indexations automatiques couvrant plus d'un document

Les concepts médicaux présents dans le corpus

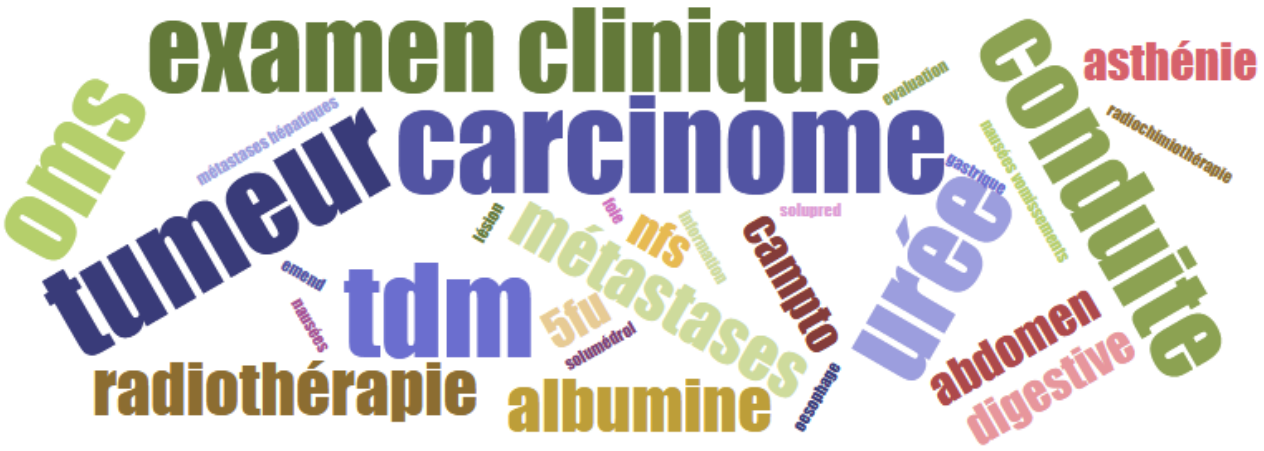
Afficher éléments

Filtrer :

Catégorie	Concept	Identifiant	Terminologie	Nb documents
processus néoplasique;	tumeur	SCT_CO_108369006	SCT	2313
thérapeutique; médicaments; pharmacie; procedure thérapeutique ou préventive;	chimiothérapie	MSH_D_004358	MSH	1856
procedure thérapeutique ou préventive;	chimiothérapie	SCT_CO_363688001	SCT	1856
procedure thérapeutique ou préventive;	chimiothérapie	SCT_CO_367336001	SCT	1856
diagnostic; activité de soins médicaux;	examen clinique	MSH_D_010808	MSH	1801

Affichage de l'élément 1 à 5 sur 3,710 éléments

Précédent 1 2 3 4 5 ... 742 Suivant



Filtrer (maladies, procédures, …) : ici médicaments

Afficher éléments

Filtrer : PHA_RAC

Catégorie	Concept	Identifiant	Terminologie	Nb documents
Substance pharmacologique; produit chimique inorganique;	cisplatine	PHA_RAC_1826	PHA	1150
chimie organique; Substance pharmacologique;	zophren	PHA_RAC_8895	PHA	1015
chimie organique; Substance pharmacologique;	campto	PHA_RAC_1470	PHA	517
chimie organique; Substance pharmacologique;	solumédrol	PHA_RAC_7444	PHA	401
médicament clinique;	solupred	PHA_RAC_7445	PHA	306

Affichage de l'élement 1 à 5 sur 332 éléments (filtré de 3,710 éléments au total)

Exemple COVID long

Nuage de mots des dossiers qui contiennent « covid long » au CHU de Rouen

Très différent de ce qu'on peut voir en ville avec profils neuro-psychiatriques

Plusieurs explications possibles (cumulables) :

- Les patients du CHU viennent pour des problèmes aigus (urgences…)
- Les professionnels du CHU ne s'intéressent pas souvent aux problèmes neuro-psychiatriques sous-jacents
- La mention de COVID long au CHU ne vaut (probablement) pas diagnostic
- Hypothèse supplémentaire : l'outil se base sur l'ECMT et donc les terminologies qui ne sont peut-être pas performantes pour les aspects neuro-psychiatriques… (évaluation en cours)

Tableaux de bord

Exemple Épidémie de bronchiolites aux urgences pédiatriques 2022, CHU de Rouen

EDSaN DashBoard

D2IM Département d'informatique et d'information Médicales

CHU ROUEN NORMANDIE

Cas de bronchiolites aux urgences pédiatriques

Les cas de bronchiolite sont comptabilisés sur la base des codes diagnostics CIM-10 de type J2* ou J15* ou J18* ou J19*, ou encore s'il est fait mention de "bronchio*" dans la conclusion.

Afficher 10 éléments

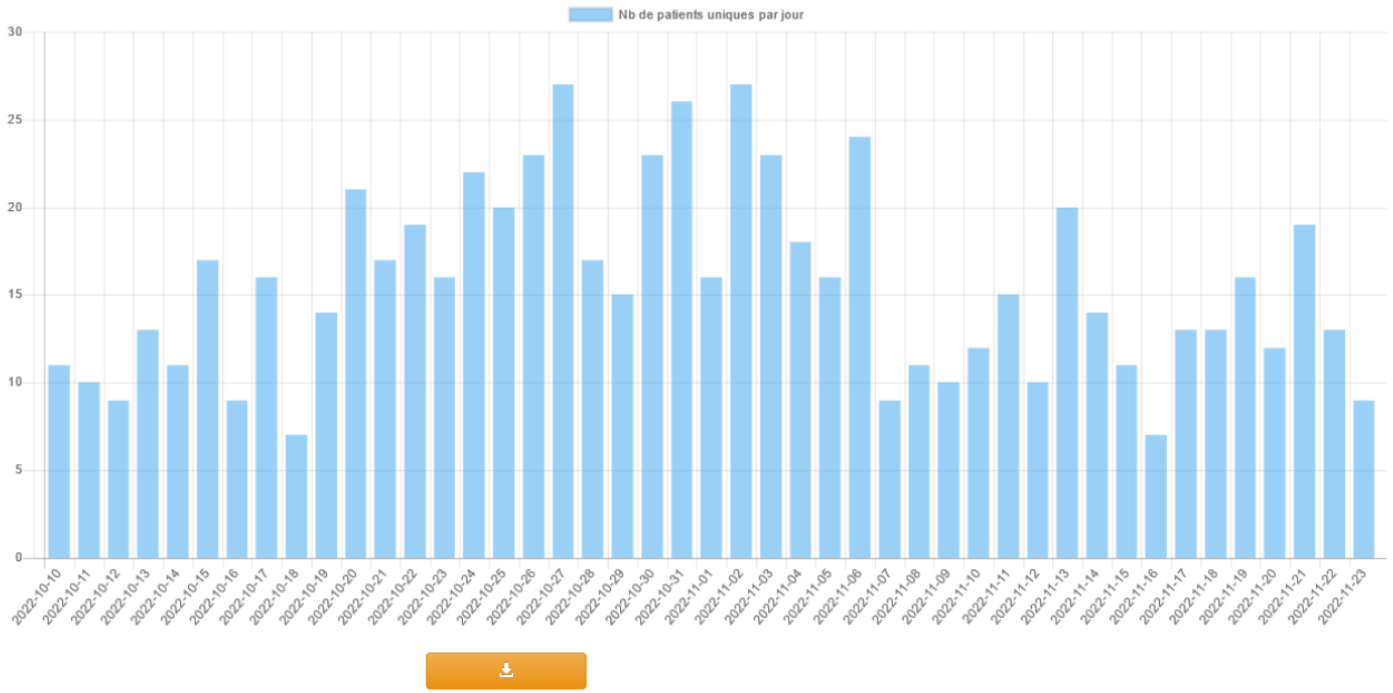
Filtrer :

45 lignes (période : 45 derniers jours)

DATE	NB TOTAL DE PATIENTS	DONT TROUVÉS PAR CODE DIAG	DONT TROUVÉS PAR CONCLUSION
2022-10-10	11	11	9
2022-10-11	10	9	5
2022-10-12	9	9	3
2022-10-13	13	12	7
2022-10-14	11	10	9
2022-10-15	17	14	6
2022-10-16	9	9	4
2022-10-17	16	15	6
2022-10-18	7	6	6
2022-10-19	14	12	8

Affichage de l'élément 1 à 10 sur 45 éléments

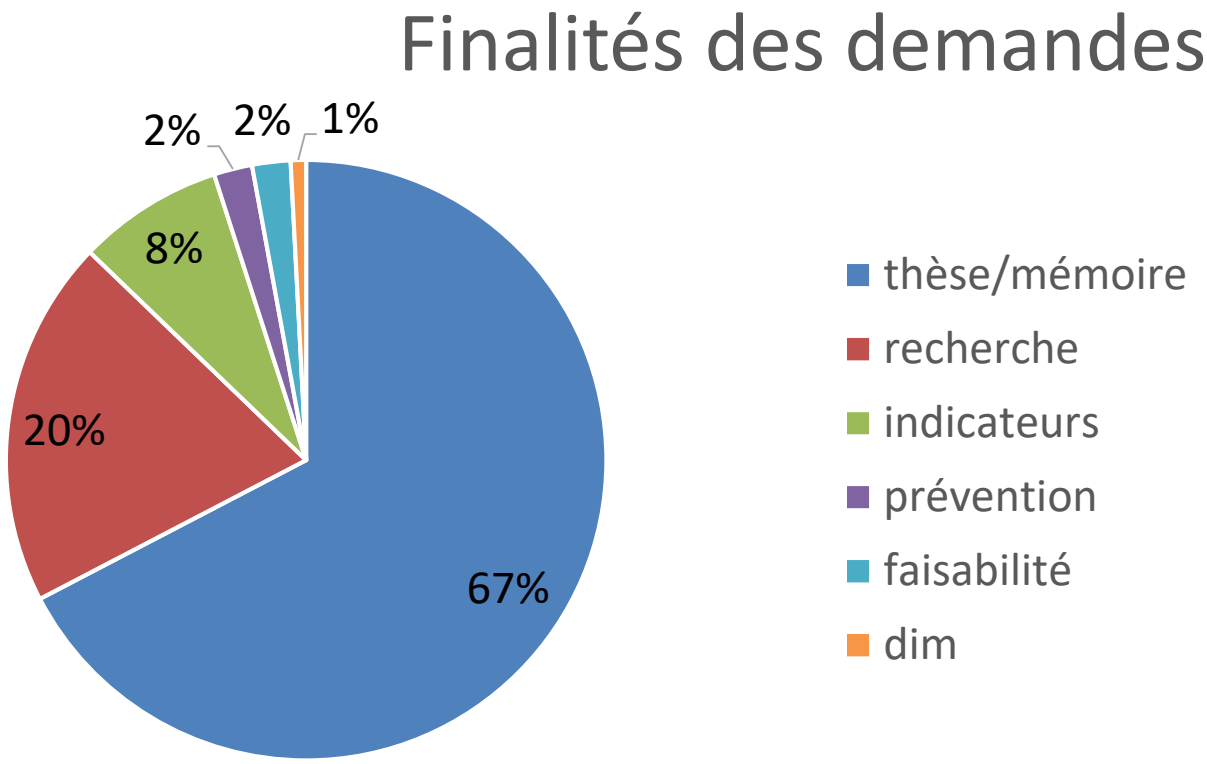
Précédent 1 2 3 4 5 Suivant



Expérience EDSaN : bilan novembre 2022

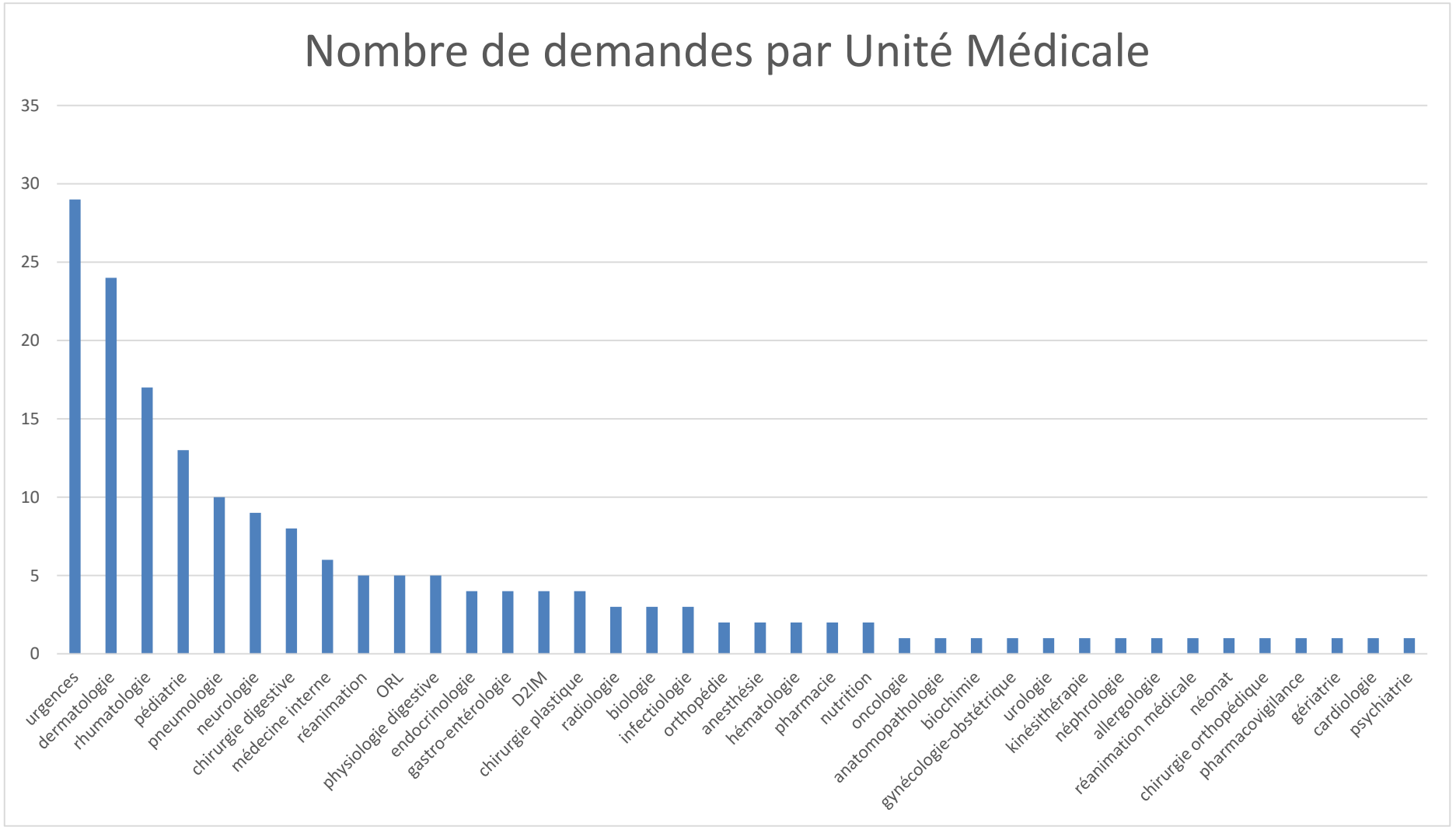
État des lieux

- Opérationnel depuis début 2018 pour des cas de tests et prototypages
- Accord CNIL obtenu en octobre 2020
- Au 23 novembre 2022 : **305 demandes traitées** (recensées dans un registre, conformément à la RGPD)



Cas d'usage COVID-19 :

- fourniture DG/CME des documents produits par semaine depuis pandémie COVID-19 (reflet des activités)
- surveillance de pathologies particulières (AVC, infarctus... pendant les confinements)
- élaboration d'une liste de patients « à très haut risque COVID » pour priorisation de la vaccination


Nombre de demandes par Unité Médicale
35
30
25
20
15
10
5
0
urgences
dermatologie
rhumatologie
pédiatrie
pneumologie
neurologie
chirurgie digestive
médecine interne
réanimation
ORL
physiologie digestive
endocrinologie
gastro-entérologie
D2IM
chirurgie plastique
radiologie
biologie
infectiologie
orthopédie
anesthésie
hématologie
pharmacie
nutrition
oncologie
anatomopathologie
biochimie
gynécologie-obstétrique
urologie
kinésithérapie
néphrologie
allergologie
réanimation médicale
néonat
chirurgie orthopédique
pharmacovigilance
gériatrie
cardiologie
psychiatrie

Les acteurs impliqués

- Coopération forte **D2IM (Santé Publique +/- informatique médicale) - DRCI (Recherche) – DSI (Informatique)**
- Avec la DRCI
  - Mise en place de la gouvernance
  - Demande CNIL +++
- Avec la DSI
  - Partenariat sur les ETL
  - Création d'un lac de données en amont d'EDSaN
- Implication indispensable du/de la Délégué(e) à la Protection des Données (DPD = DPO) et du/de la Responsable de la Sécurité des Systèmes d'Information (RSSI)

Points forts et limites

- Outils puissants : rapides, technologies innovantes
- Maximum de données cliniques agrégées mais pas forcément exhaustif! (papier...)
- Utilisation plus « fine » de la donnée via algorithmes d'IA (NLP, etc.)
- Environnements très sécurisés

- Porte vers la médecine de demain : 5P
- Encore beaucoup de chemin!!

Perspectives/problématiques futures

- Lien EDS hôpital – hôpital (interopérabilité...)
- Lien EDS hôpital et EDS ville (médecine générale)

- Favoriser l'appariement ou la comparaison avec des **bases nationales** (HDH/SNDS)

Merci

stefan.darmoni@chu-rouen.fr

julien.grosjean@chu-rouen.fr