

25 ans **CiSM_eF**

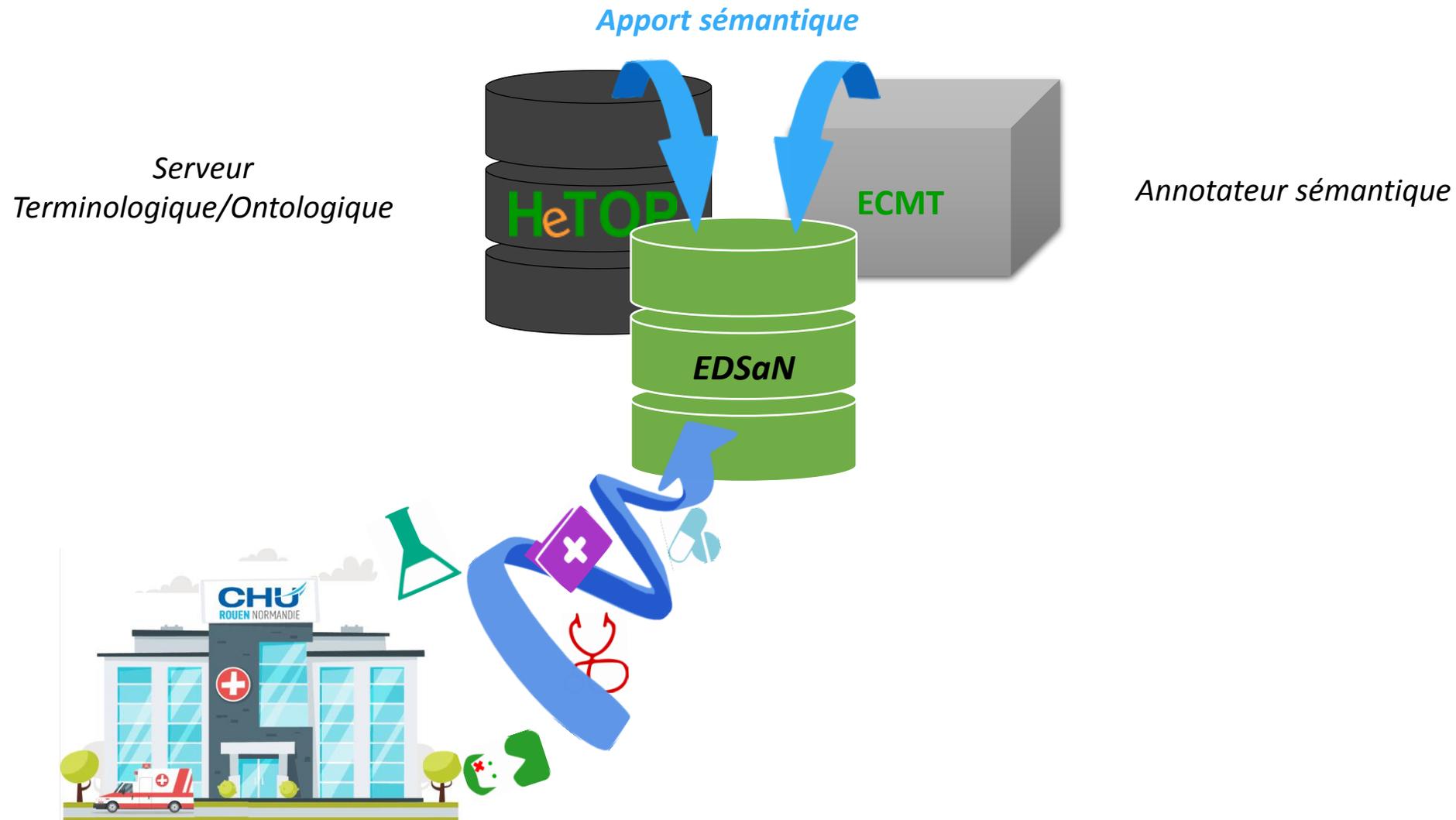
Moteur de Recherche Sémantique

4 février 2020

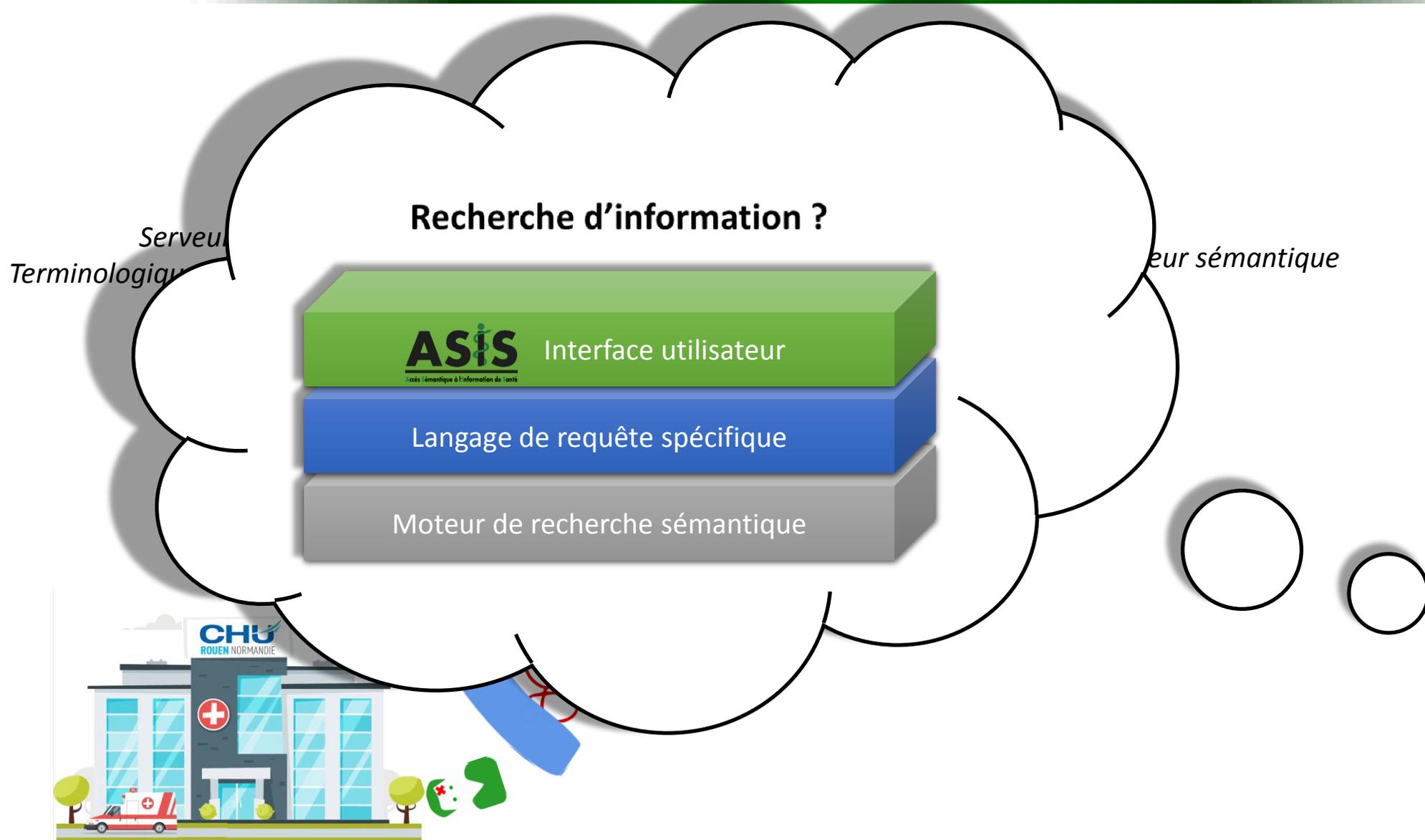
Romain LELONG, PhD

Centre Hospitalier Universitaire de Rouen & LIMICS, INSERM U1142

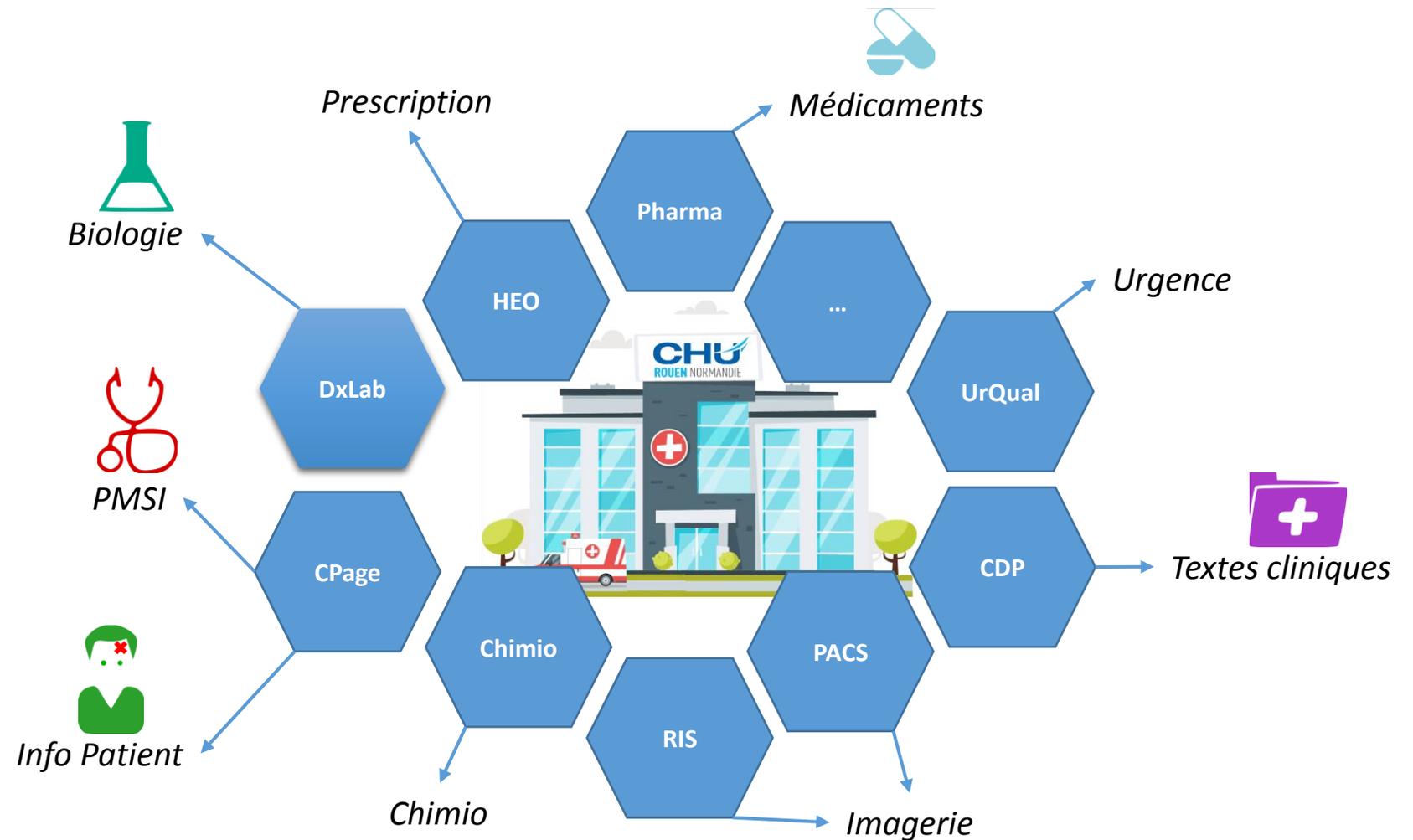
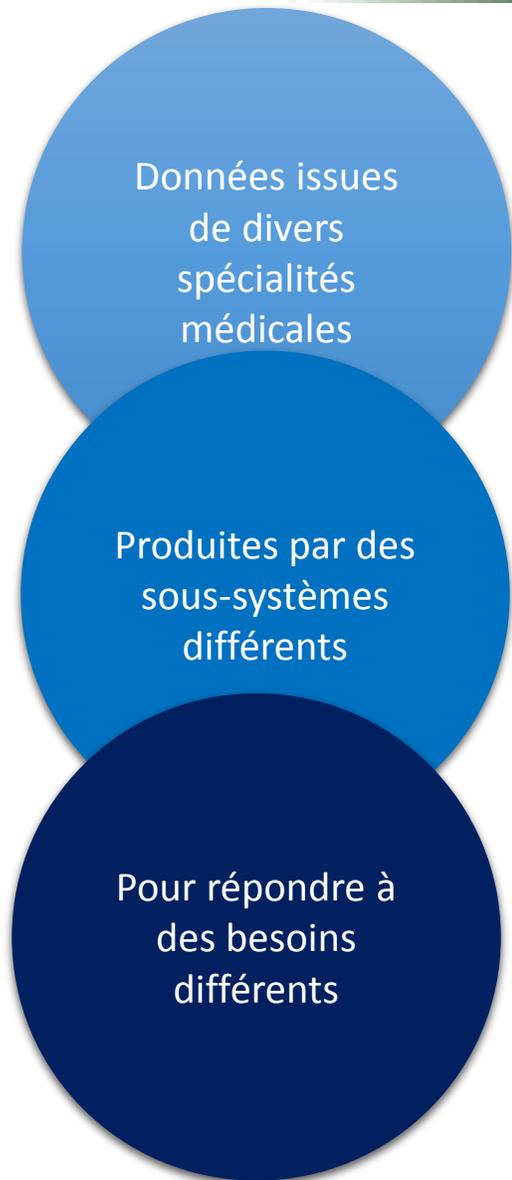
EDSaN : Entrepôt de Données de Santé Normand



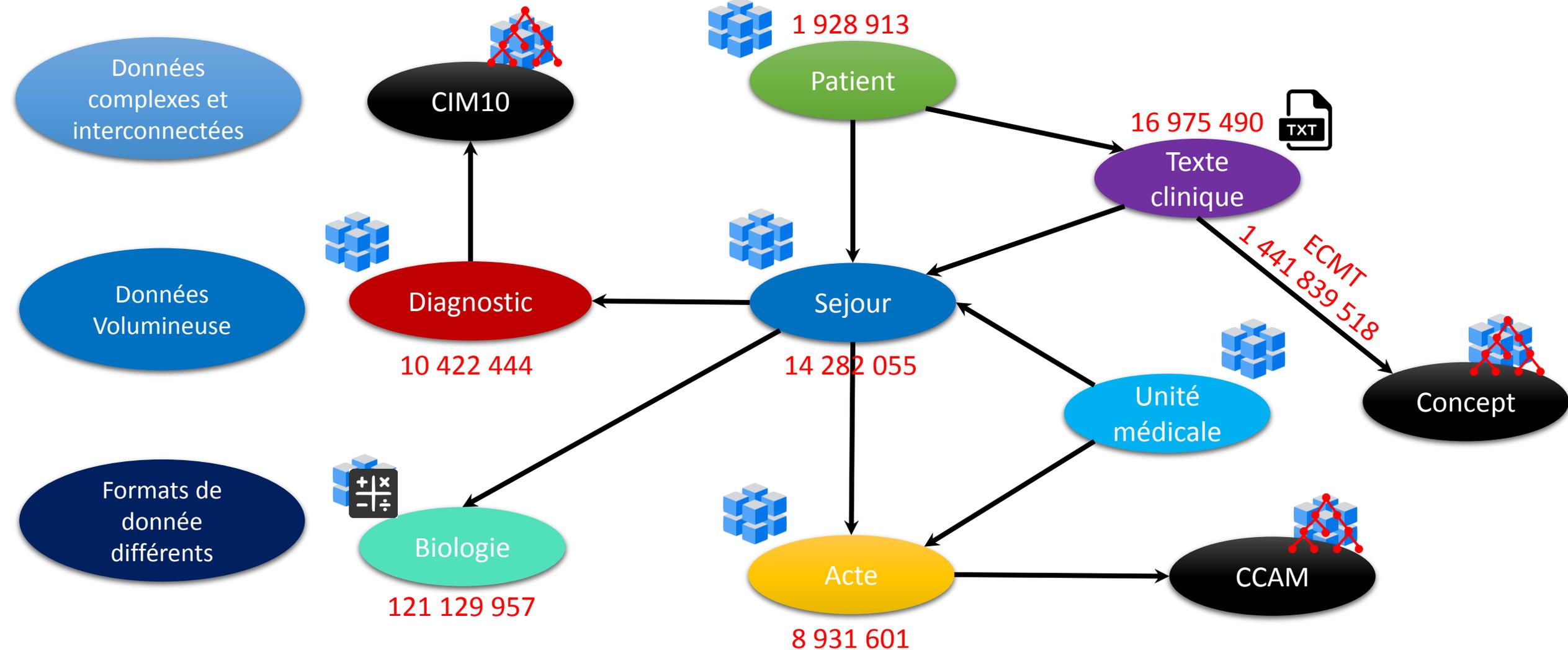
EDSaN : Entrepôt de Données de Santé Normand



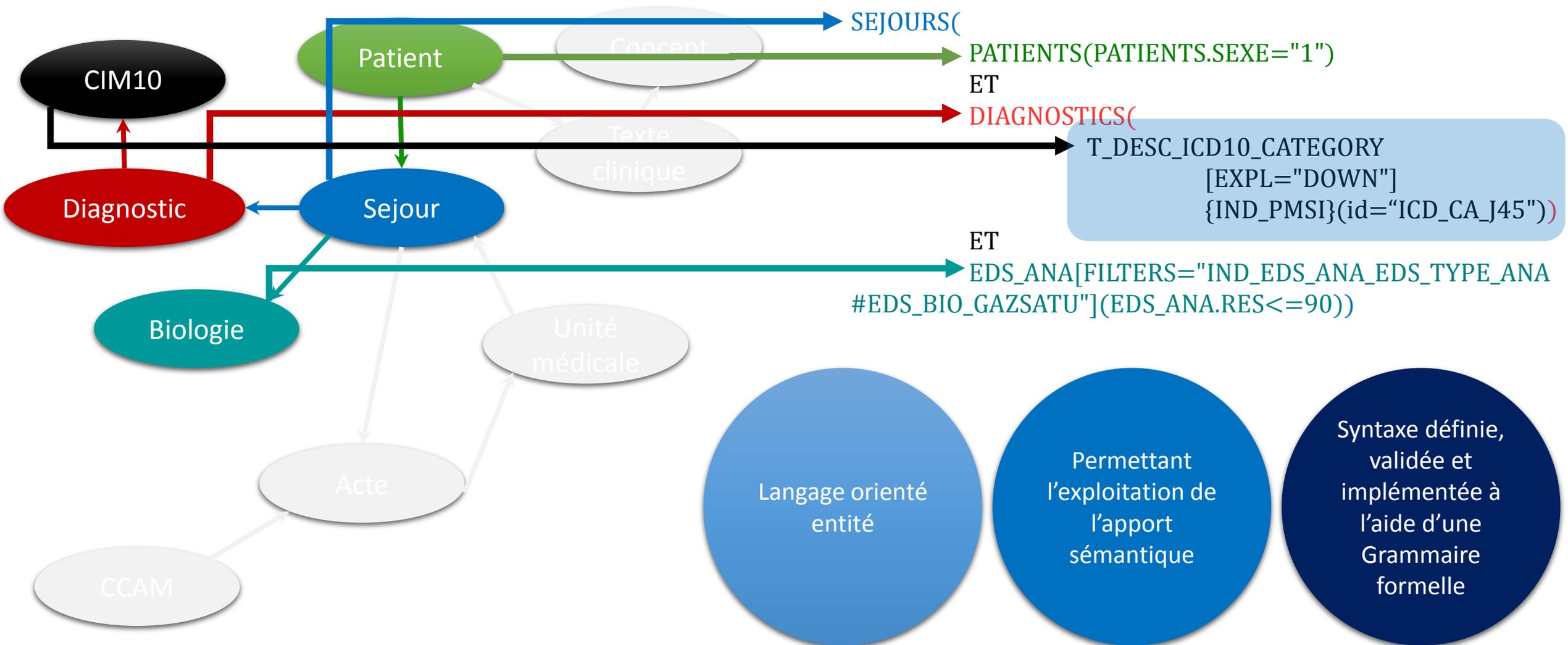
Problématique : Des données cibles variées

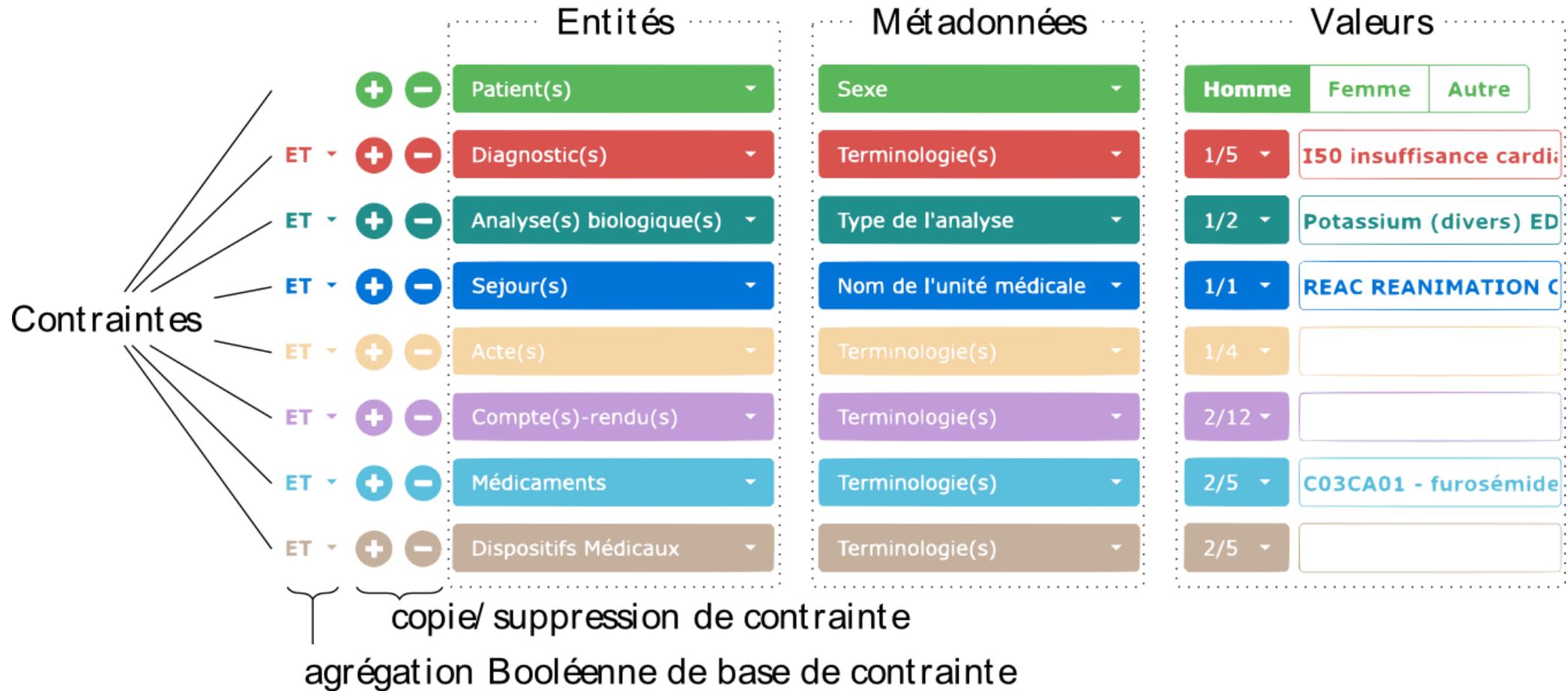


Et donc, un modèle conceptuel de données « complexe »

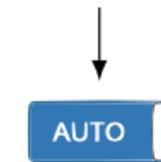


Un moteur de recherche muni d'un langage de requête logique

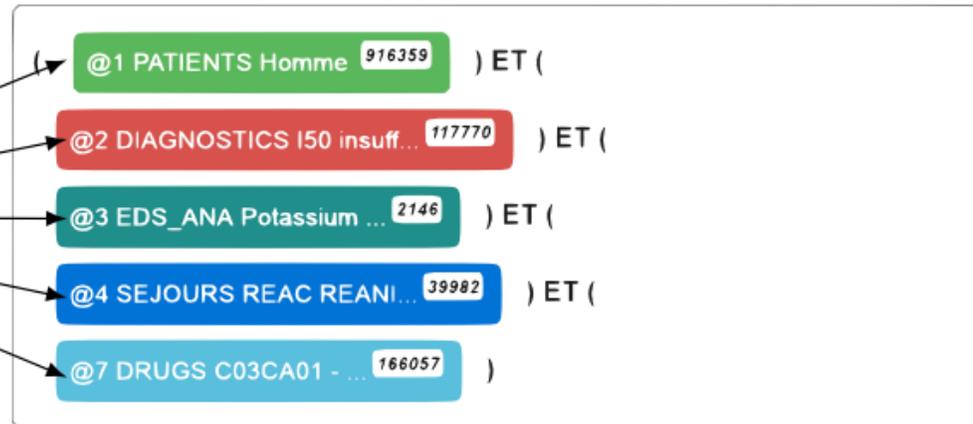




Constitution automatique
de la contrainte globale



Contraintes unitaires
définies en étape n° 1



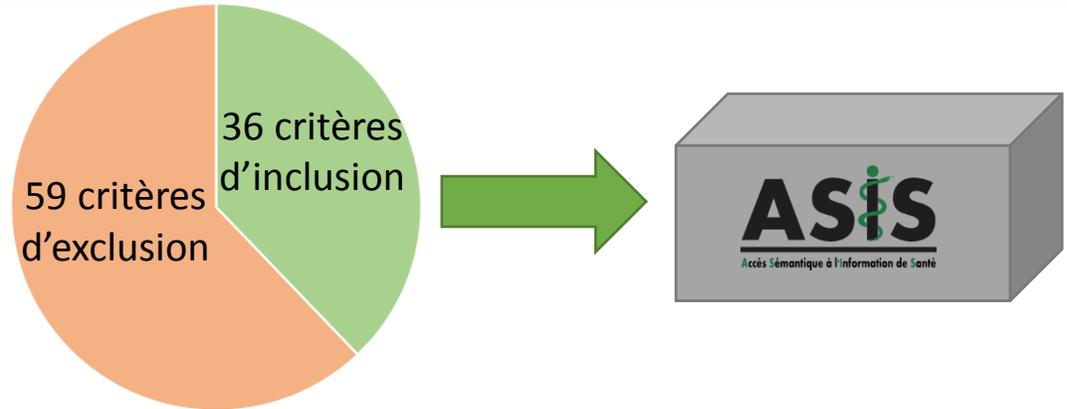
zone éditable permet-
tant la constitution
d'une contrainte Boo-
lèenne globale

Etude : Objectif et Méthode

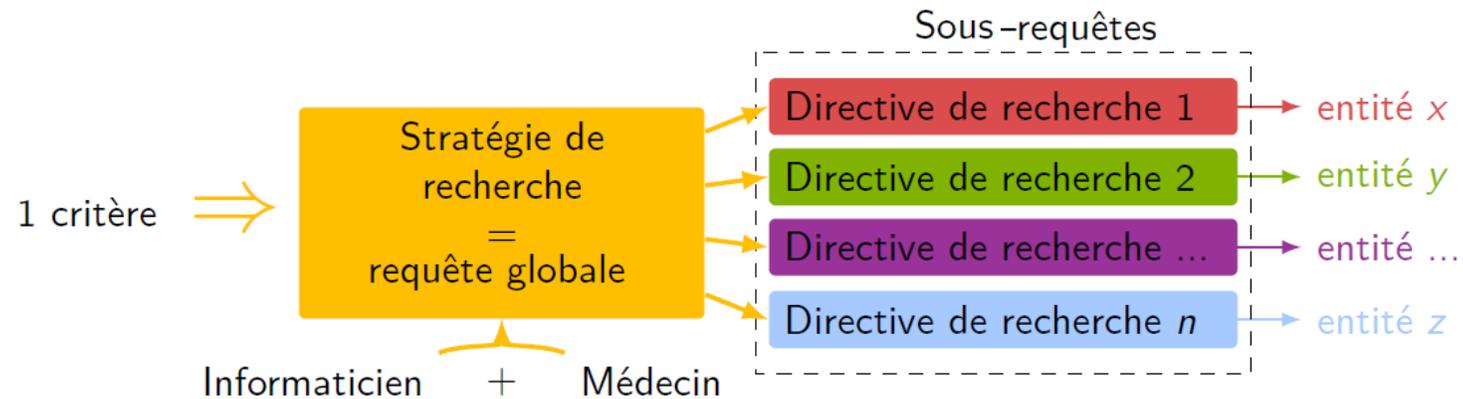
Objectif

Evaluer la capacité du système à répondre à 95 critères issus de 5 études cliniques du CHU de Rouen.

Étude exploratoire : évaluer les « forces et faiblesses » du système.



Méthode



Etude : Méthodologie

Méthodologie

Observation empirique de trois propriétés :

Niveau de prise
en charge du
critère

Total
Précis
Partiel
Imprécis
Inopérant
Non-Applicable (N/A)

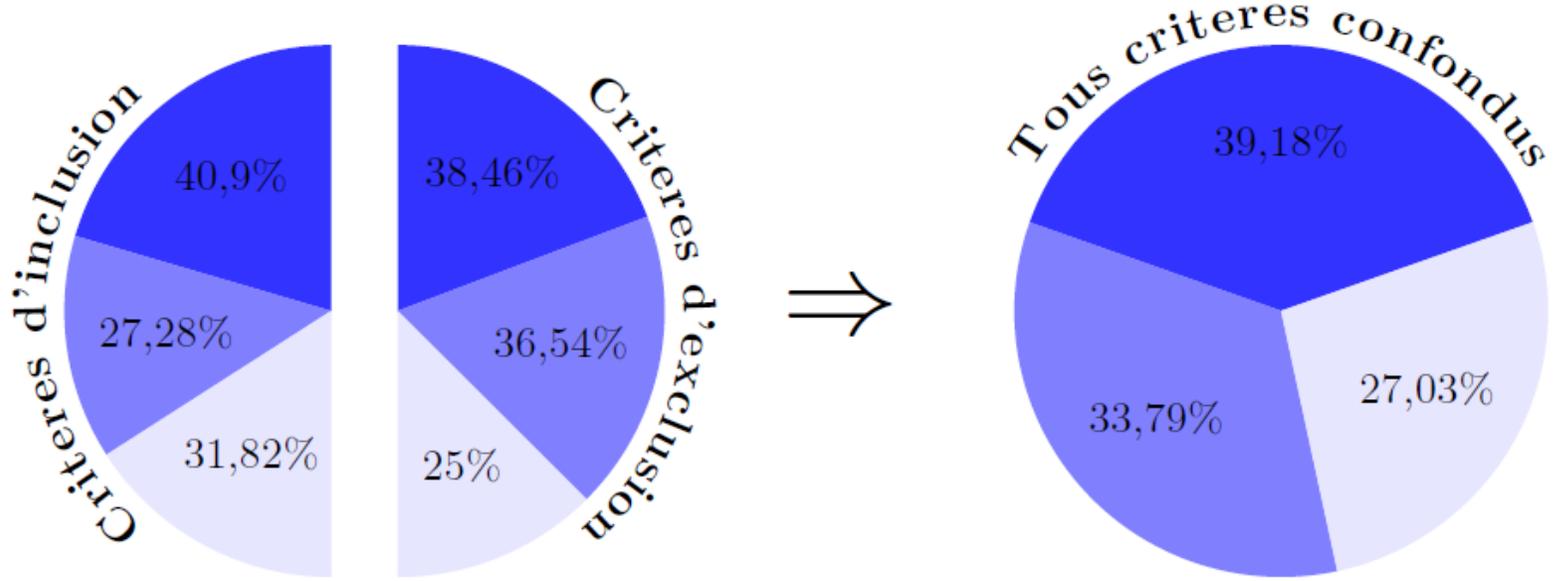
Les sources
d'informations
ciblées

 Données patients
 Données PMSI
 Données de séjours
 Données biologique
 Textes cliniques
 Informations externes

Les limitations
observées

 Aucune
 Données inconsistantes
 Recherche imprécise
 Limitation technique
 Critère subjectif/générique
 Communication

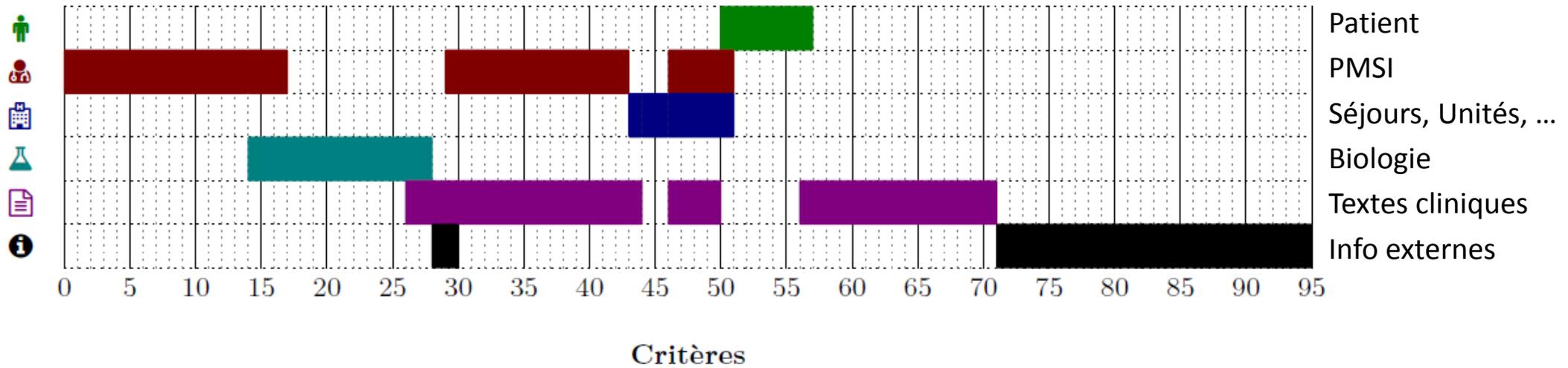
Etude : Résultat (Niveau de prise en charge)



- {Total, Précis} (utile pour la constitution de cohortes)
- {Partiel} (utile comme outils de pre-screening)
- {Imprécis, Inopérant} (exploration de l'information de Santé)

Etude : Résultat (Sources d'information utilisées)

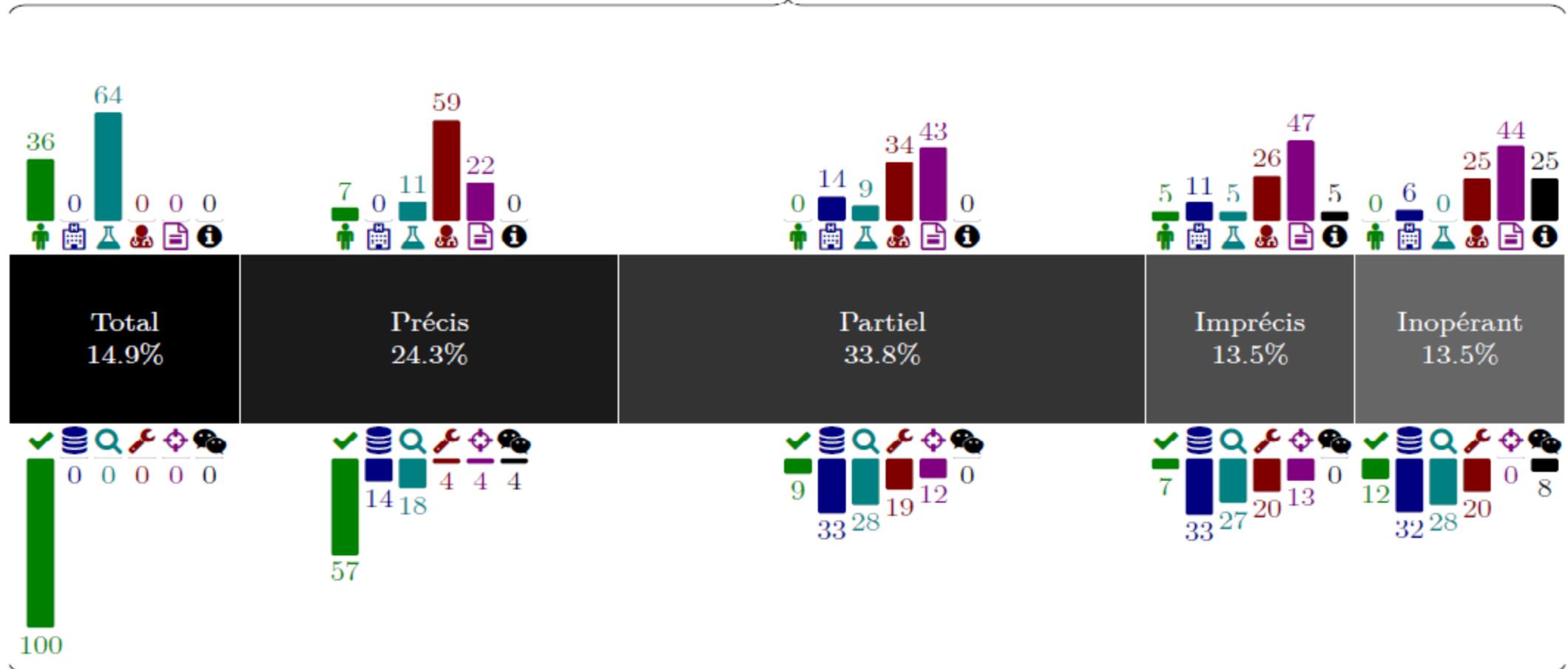
Sources d'information



Source	taux	pourcentage
Textes cliniques	37/95	38,95%
PMSI	36/95	37,89%
Info externes	26/95	27,36%
Biologie	14/95	14,73%
Séjours	8/95	8,42%
Patient	7/95	3,37%

Résultat : Mise en correspondance des trois propriétés

Distribution des sources d'information



Distribution des limites

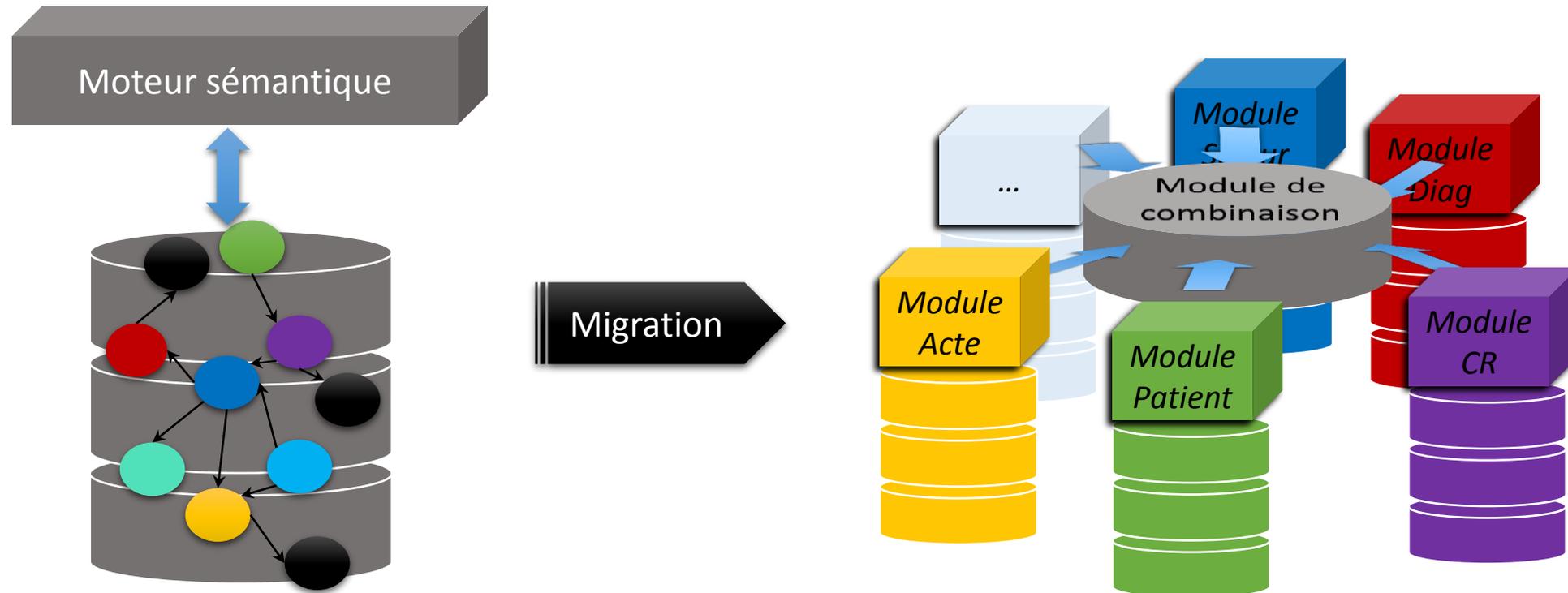
Perspectives : 2020 – 202n

Analyse de la PME Alicante :

Nécessité de deux serveurs puissants pour CHU Rouen : 120 K€ matériel

Si volonté d'être utilisé en CH :

Migration du paradigme « langage formel » vers une application « modulaire » (Labo, médicaments, PSMI, ...) + intégration avec Doc'EDS



Conclusion

Adaptabilité aux cas d'usages :

Langage de requête vs. formulaires « figés ».

Exhaustivité :

Possibilité d'atteindre et de rechercher n'importe quelle élément d'information du modèle conceptuel de donné.

*Globalement de **bon temps de réponse.***



Syntaxe complexe du Langage de requête

(nécessité de connaître le model conceptuel de donnée pour formuler des requêtes)

Architecture matériel requise lourde

(Serveur de 144 cœur + 1 To de RAM)

Les textes cliniques restent :

Une source d'information majeure;

Une limitation majeure en terme de recherche d'information.

25 ans CiSM^eF

Merci pour votre attention

Des questions ?