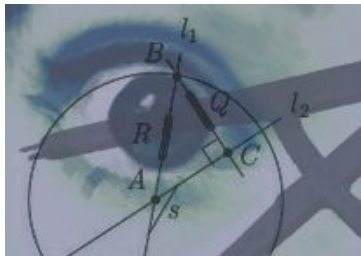


# Le rôle des entrepôts de données biomédicales pour la recherche

## *Exemple de l'entrepôt eHOP*

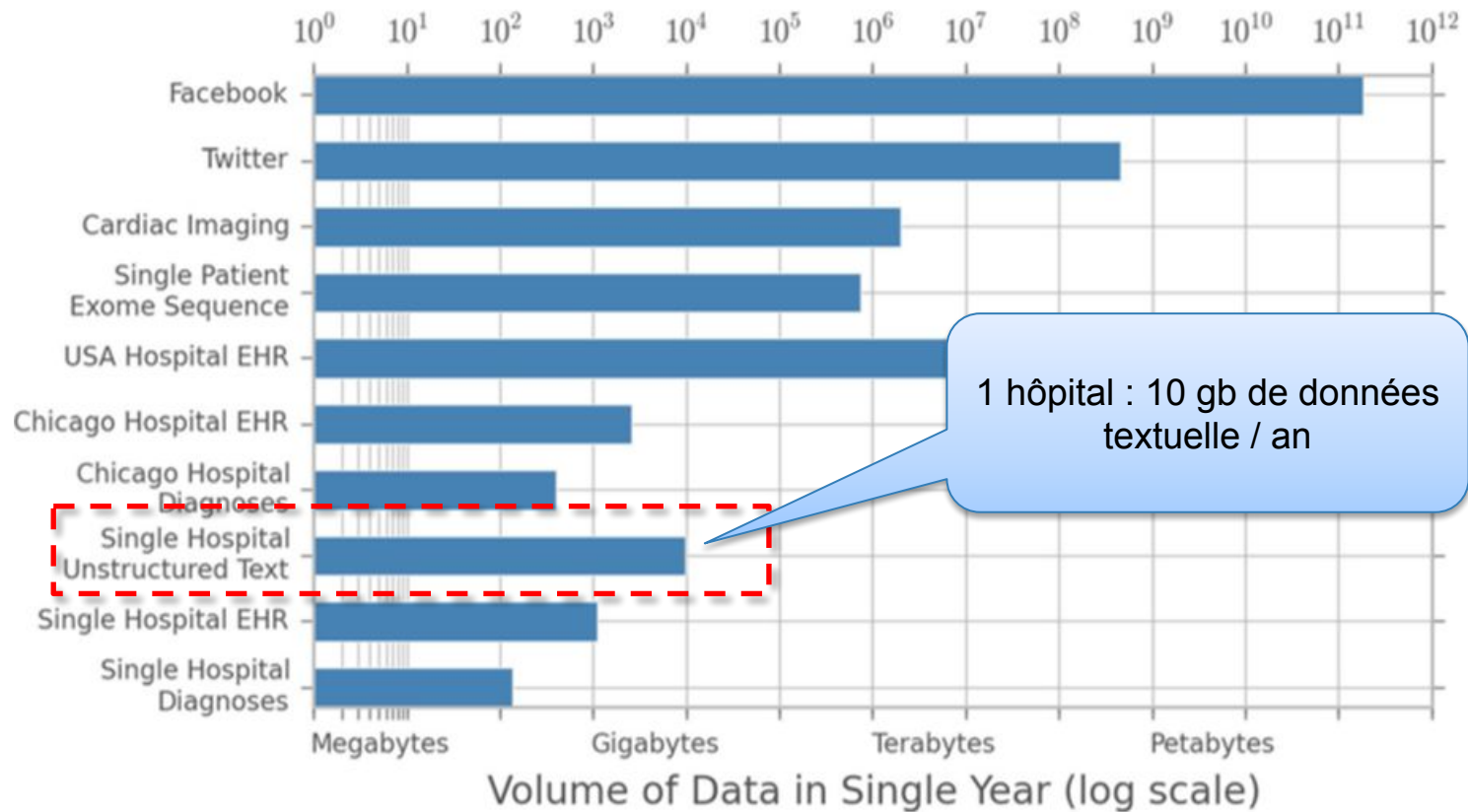
**Dr Guillaume Bouzillé, Pr Marc Cuggia**

*Equipe projet Données Massives en Santé*  
Unité Inserm 1099 Laboratoire Traitement du Signal et de l'Image  
Centres des Données Cliniques, CHU de Rennes

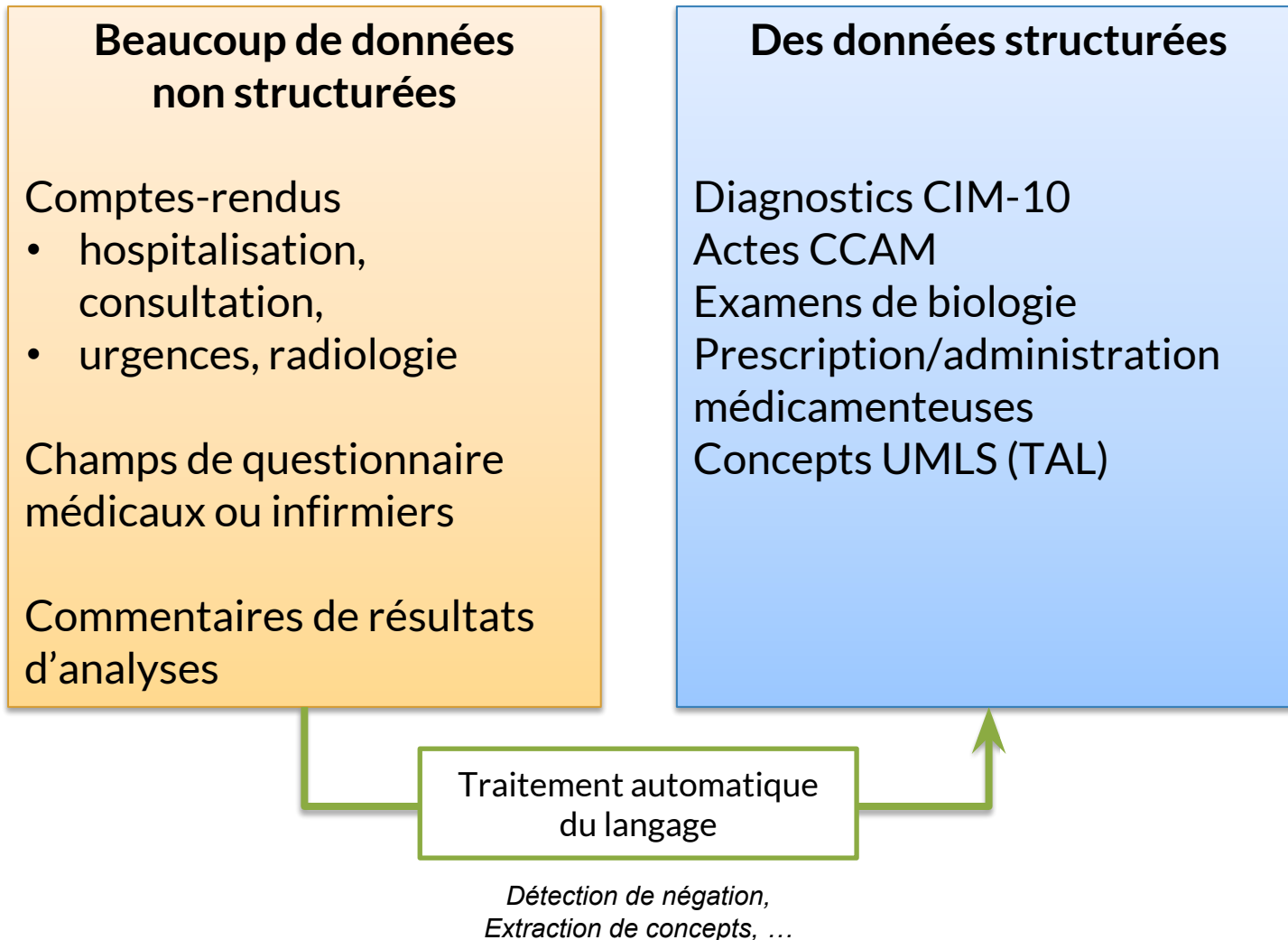


**10ème journée  
du réseau COGITER**

# Une explosion des données numériques



# Des données hospitalières hétérogènes



# Des données sensibles

Une anonymisation irréversible altère les données d'une façon incompatible avec les besoins d'analyse

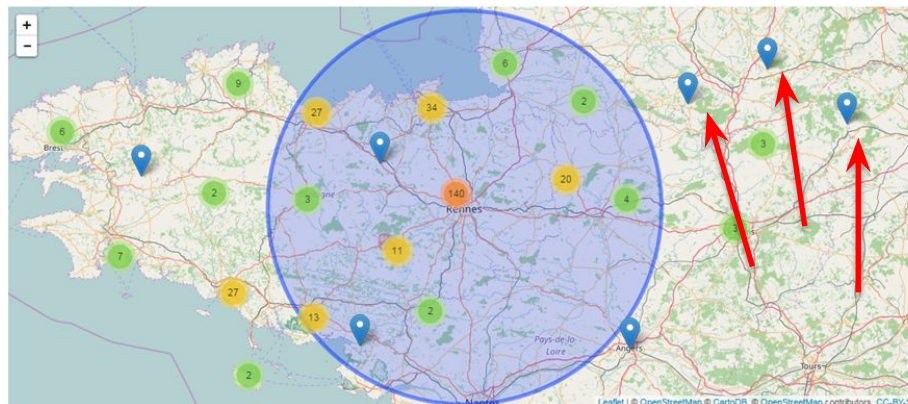
La pseudonymisation est la règle mais :

Chaînage des données

Données OMICs ++

Maladies Rares

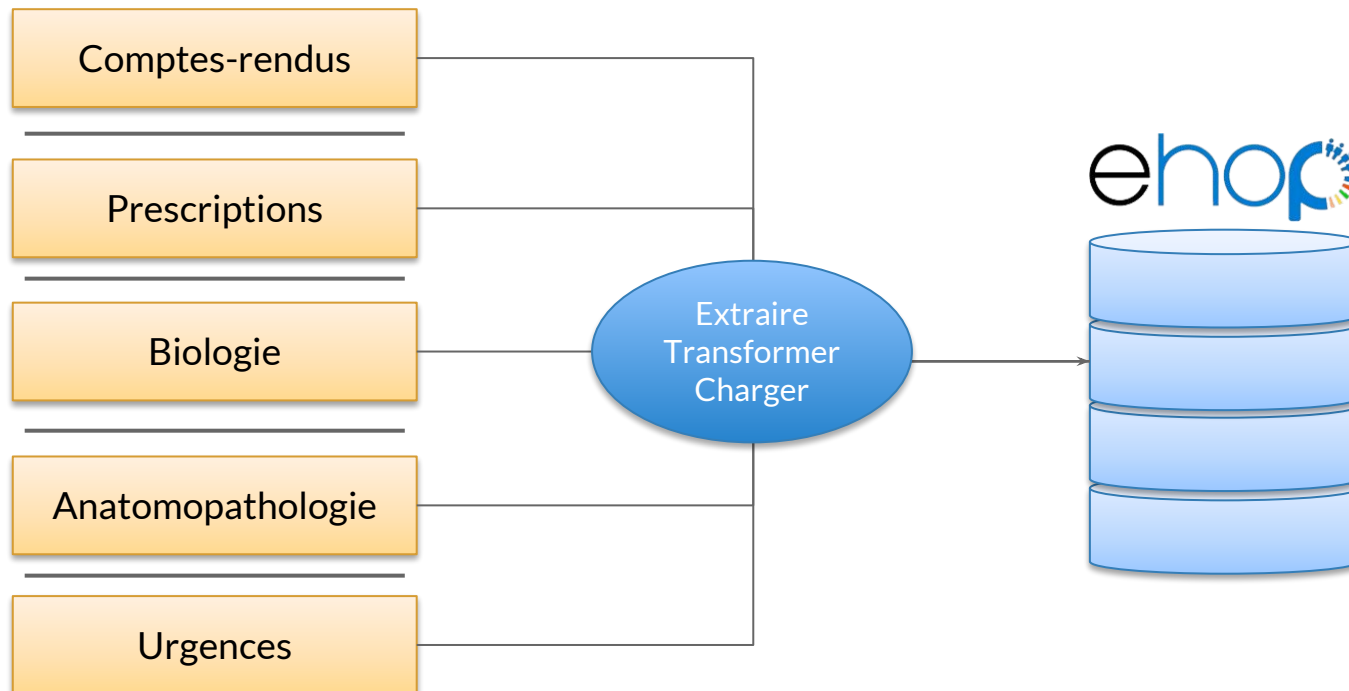
⇒ **Risque de réindentification**



# Un besoin de décloisonner les données

Des données produites pour le soin et cloisonnées

Une base de données orientée recherche

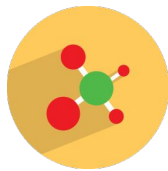


# Entrepôts de données biomédicales

## Exemple du CHU de Rennes :



1,2+ Millions  
de Patients



38+ Millions  
de Documents



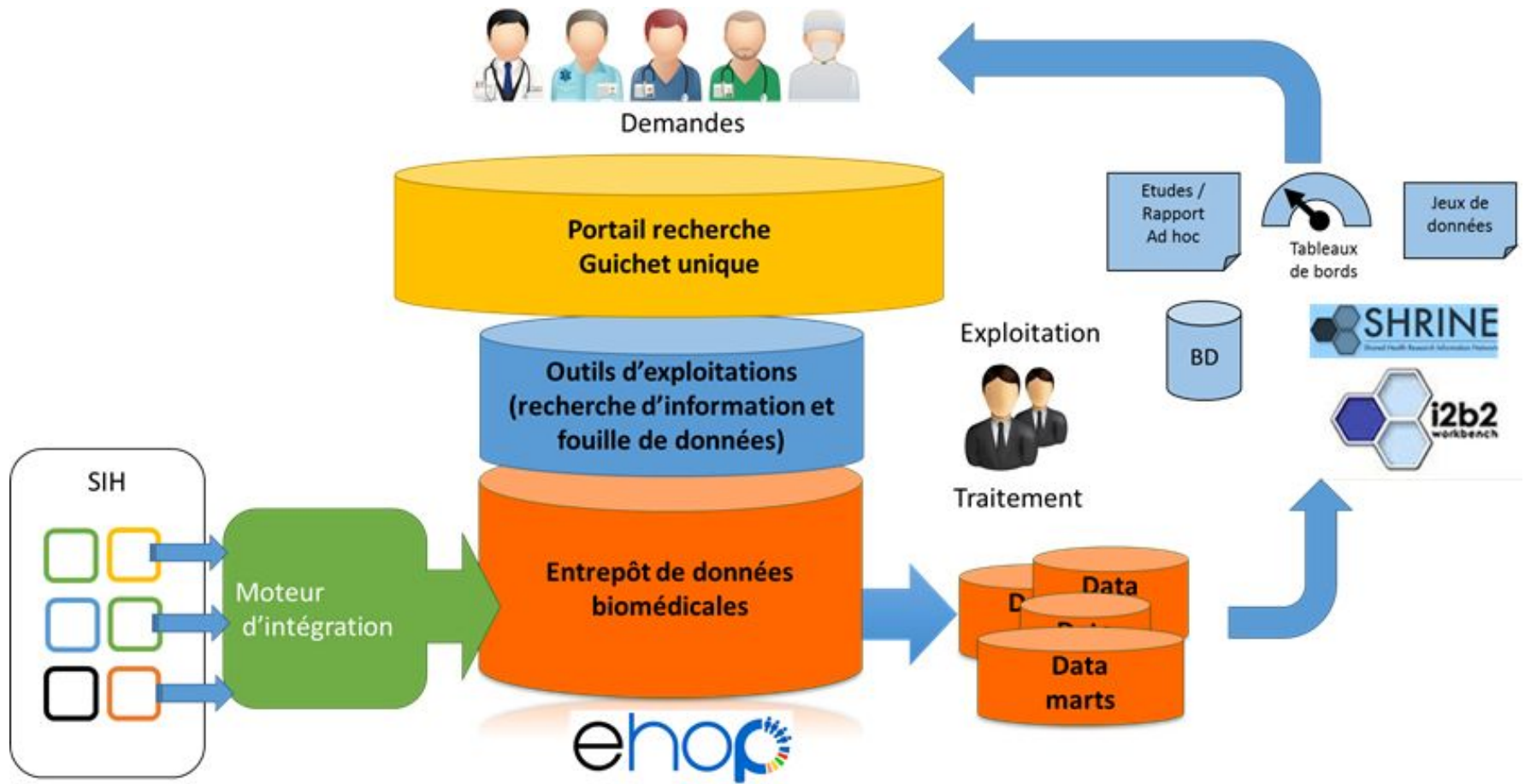
5+ Millions  
de Séjours



400+ Millions  
de Concepts

Données identité/séjour/structure  
Comptes rendus Hospitaliers  
Comptes-rendus de Consultation  
Prescription/administration  
Bio banque (CRB)  
Dossiers Urgence  
Données de Laboratoires  
Dossiers infirmier (projet)  
PMSI  
Anatomo-pathologie  
CR Imagerie  
Dossier Odontologie

# Les Centres de Données Cliniques (CDC)



# eHOP : application à la recherche clinique

The screenshot displays the eHOP application interface. At the top, there is a navigation bar with the eHOP logo, the user name 'Guillaume BOUZILLE', and a 'Se déconnecter' button. Below this, a secondary navigation bar shows 'Requête\*' (selected), 'Résultats', 'Statistiques', and 'Patients (1666927)'. The main interface is divided into several sections:

- Left sidebar:** Contains navigation options like 'Requêtes enregistrées (25 éléments)', 'Recherche avancée', 'Recherche Structurée', 'Elements', 'Gestion des paniers', and 'Gestion des datamarts'.
- Search criteria panel (left):** Titled 'Recherche avancée', it includes sections for 'Critères généraux', 'Recherche Textuelle', and 'Recherche Structurée'. The 'Recherche Textuelle' section is active, showing options for 'Type' (Inclure/Exclure), 'Requête' (Texte), 'Document' (Sélectionner type ...), 'Unité(s) Fonctionnelle(s)' (Unité(s) Fonctionnelle(s)), and 'Périphrases négatives' (Inclure).
- Main content area:**
  - Contexte:** A box stating 'La recherche s'effectue sur l'ensemble de l'étude.'
  - Critères Généraux (Nombre de patients: 9 279):** A yellow box containing an 'Actif' status, a date range 'Les documents datés du 01/01/2015 au 31/12/2015', functional units 'Unités fonctionnelles: 5401,5403,5405,5409', and 'Intersection des critères sur le patient'.
  - Liste des critères textuels:** A blue box containing a list of criteria. The first criterion is highlighted with a red dashed border and a red box labeled 'Recherche par mot-clé':
    - #1 Inclure** les patients dont les documents contiennent l'expression "neurinome OR schwannome" (Nombre de patients: 4 160)
    - Type:  Inclure  Exclure
    - Requête: neurinome OR schwannome
    - Document: Sélectionner type ...
    - Unité(s) Fonctionnelle(s): Unité(s) Fonctionnelle(s)
    - Périphrases négatives:  Inclure



# eHOP : application à la recherche clinique

The screenshot displays the eHOP application interface. At the top, the header includes the eHOP logo, navigation links like 'Changer d'étude', 'Un problème?', and 'Administration', the user name 'Guillaume BOUZILLE', and a 'Se déconnecter' button.

The main interface is divided into several sections:

- Left sidebar:** Contains navigation options: 'Requêtes enregistrées (25 éléments)', 'Recherche avancée' (with sub-sections for 'Critères généraux', 'Recherche Textuelle', 'Recherche Structurée', and 'Elements'), 'Gestion des paniers', and 'Gestion des datamarts'.
- Top navigation:** Includes 'Requête\*', 'Résultats (346)', 'Statistiques', and 'Patients (1666927)'.
- Search filters (Recherche avancée):**
  - Critères généraux:** Includes 'Type' (Inclure/Exclure), 'Requête' (Texte), 'Document' (Sélectionner type ...), 'Unité(s) Fonctionnelle(s)' (Unité(s) Fonctionnelle(s)), and 'Périphrases négatives' (Inclure).
  - Buttons:** '+Ajout critère'.
- Results area:**
  - Status:** 'Actuellement affiché: 50 sur 346 patient(s), 86 sur 714 document(s). Requête exécutée en 3.36s.' (highlighted with a red dashed box).
  - Concepts:** A yellow bar with a '+Ajout critère' button.
  - Table:** A table with columns: ID, Age, S, CP, Actions, and Document(s). It lists patient records with their details and associated documents.

ID	Age	S	CP	Actions	Document(s)
3126892614	85 ans	M	35230		Compte Rendu de Consultation Lettre au patient
3126892585	37 ans	F	35360		Compte Rendu de Consultation
3126856448	84 ans	F	35220		Compte Rendu de Consultation Compte Rendu de Consultation
3128931312	29 ans	F	35750		Compte Rendu de Consultation
3128903684	57 ans	M	56380		Compte Rendu de Consultation
3128904233	55 ans	F	50320		Courrier Dr LENA Hervé Compte Rendu de Consultation
3128889427	68 ans	F	35830		ORL CRC 20 11 2015

# eHOP : application à la recherche clinique

The screenshot displays the eHOP application interface. At the top, there is a navigation bar with the eHOP logo, a search bar, and several menu items: "Changer d'étude", "Un problème? -", "Administration", "Guillaume BOUZILLE", and "Se déconnecter". The main content area is titled "Compte Rendu de Consultation" and features three tabs: "Document", "Autres documents", and "Journal d'accès". The "Document" tab is active, showing a list of concepts: "schwannome (2)". Below this, there is an "Informations" section containing the following text: "PRENOM NOM\_3128931312, agée de 27 ans à la date du document 23/04/2015 - CR ou fiche de consultation ou de visite IEP UF : 5409-CE ORL PL". The main body of the report contains the following text: "Rennes, le 17/04/2015", "Consultation du 13/04/2015", "Cher Olivier,", "Je te remercie de m'avoir adressé Madame [PRENOM] [NOM\_3128931312] née le 29/05/1988 qui présente un très volumineux schwannome vestibulaire droit. Cette patiente n'a pas d'antécédent en dehors d'une adénoïdectomie dans l'enfance. Elle a eu un petit garçon il y a 6 mois. Elle est enseignante, professeur des écoles. Elle a remarqué depuis quelques années une légère instabilité sans handicap particulier. Elle a ressenti une baisse auditive peu après son accouchement. Cliniquement, il existe un léger syndrome vestibulaire droit avec un test de Fukuda latéralisé et un petit nystagmus au Head Shaking Test. La motricité et la sensibilité faciale sont normales. Il n'y a pas de syndrome cérébelleux."

# eHOP : application à la recherche clinique

**Recherche Structurée**

Elements +

**Gestion des paniers**

Liste des paniers ( 4 élément(s) ) +

Titre	Personne	# Inclus	# Exclus	# Docs	Actions
Hypertension	demo	2	4	301	
Neurinomes	demo	339	5	26132	
test2	bouzille	36	0	4261	
test	bouzille	1	4	124	

**Panier sélectionné: Hypertension ( 2 inclus, 4 exclus )**

Liste des patients inclus / Liste des patients exclus

ID	Age	Sexe	# Documents	Actions
1621612874	85 ans	F	124	
1621576291	38 ans	F	177	

+ Nouveau

**Gestion des datamarts** +

**Gestion des exports** +

**Requête** | **Résultats (346)** | **Statistiques** | **Patients (346)**

Actuellement affiché: 50 sur 346 patient(s), 85 sur 710 document(s). Requête exécutée en 0.23s.

**Ajouter l'ensemble des patients au panier**

**Concepts** +

ID	Age	S	CP	Actions	Panier	Document(s)
1621576204	85 ans	M	35230		<input type="checkbox"/> <input checked="" type="checkbox"/>	Compte Rendu de Consultation 23/03/2015  Afficher Lettre au patient 29/06/2015  Afficher
1621576291	37 ans	F	35360		<input checked="" type="checkbox"/> <input type="checkbox"/>	Compte Rendu de Consultation 17/03/2015  Afficher
1621612874	84 ans	F	35220		<input checked="" type="checkbox"/> <input type="checkbox"/>	Compte Rendu de Consultation 17/07/2015  Afficher Compte Rendu de Consultation 21/08/2015  Afficher
1622696378	29 ans	F	35750		<input type="checkbox"/> <input checked="" type="checkbox"/>	Compte Rendu de Consultation 23/04/2015  Afficher
1622710862	57 ans	M	56380		<input type="checkbox"/> <input checked="" type="checkbox"/>	Compte Rendu de Consultation 24/04/2015  Afficher
1622710371	55 ans	F	50320		<input type="checkbox"/> <input checked="" type="checkbox"/>	Courrier Dr LENA Hervé 18/03/2015  Afficher Compte Rendu de 18/03/2015  Afficher

# eHOP : application à la recherche clinique

**eHOP**    Changer d'étude    Un problème?    Administration    Guillaume BOUZILLE    Se déconnecter

Recherche Textuelle

Recherche Structurée

Veuillez sélectionner une terminologie:

Actes (CCAM)

Rechercher ...    sur:    Libellé

▼ SYSTÈME NERVEUX CENTRAL, PÉRIPHÉRIQUE ET AUTONOME (216 patients, 750 docs)

- ▶ ACTES DIAGNOSTIQUES SUR LE SYSTÈME NERVEUX (187 patients, 551 docs)
- ▼ ACTES THÉRAPEUTIQUES SUR LE SYSTÈME NERVEUX CENTRAL INTRACRÂNIEN (74 patients, 143 docs)
  - ▶ Actes thérapeutiques sur le cerveau (1 patient, 2 docs)
  - ▶ Actes thérapeutiques sur les méninges, les ventricules et le liquide cérébrospinal [LCS] intracrâniens (6 patients, 13 docs)
  - ▼ Exérèse de tumeur intracrânienne extraencéphalique (74 patients, 128 docs)
    - Exérèse de tumeur de la pointe du rocher, par abord translabyrinthique (2 patients, 4 docs)
    - Exérèse de tumeur de l'angle pontocérébelleux et\_ou du méat acoustique interne [conduit auditif interne], par abord translabyrinthique (56 patients, 99 docs)

Type: Inclure

+Ajouter    Annuler

- Exérèse de tumeur de la pointe du rocher, par abord suprapétreux (1 patient, 1 doc)
- Exérèse de tumeur de l'angle pontocérébelleux et\_ou du méat acoustique interne [conduit auditif interne], par abord infraoccipital rétrosigmoidien (3 patients, 5 docs)
- Exérèse de tumeur de l'angle pontocérébelleux et\_ou du méat acoustique interne [conduit auditif interne], par abord suprapétreux (12 patients, 19 docs)

- ▶ ACTES THÉRAPEUTIQUES SUR LE SYSTÈME NERVEUX CENTRAL SPINAL [RACHIDIEN] (3 patients, 5 docs)
- ▶ STIMULATION DU SYSTÈME NERVEUX CENTRAL (1 patient, 14 docs)
- ▶ ACTES THÉRAPEUTIQUES SUR LES NERFS CRÂNIENS ET LES NERFS SPINAUX (18 patients, 35 docs)

Les documents datés du 01/01/2015 au 31/12/2015  
Unités fonctionnelles: 5401,5403,5405,5409  
Intersection des critères sur le patient

Liste des critères textuels

- #1 Inclure les patients dont les documents contiennent l'expression "neurinome or schwannome" (Nombre de patients: 346)


Liste des critères structurés

- #1 Recherche de Exérèse de tumeur de l'angle pontocérébelleux et\_ou du méat acoustique interne [conduit auditif interne], par abord translabyrinthique dans : Actes (CCAM) ▶ SYSTÈME NERVEUX CENTRAL, PÉRIPHÉRIQUE ET AUTONOME ▶ ACTES THÉRAPEUTIQUES SUR LE SYSTÈME NERVEUX CENTRAL INTRACRÂNIEN ▶ Exérèse de tumeur intracrânienne extraencéphalique ▶ Exérèse de tumeur de l'angle pontocérébelleux et\_ou du méat acoustique interne [conduit auditif interne], par abord translabyrinthique (Nombre de patients: 56)

{ } JSON

☆ Nouvelle requête    Réinitialiser    Enregistrer    Exécuter

# eHOP : application à la recherche clinique

 **eHOP** Changer d'étude Un problème? ▾ Administration Guillaume BOUZILLE Se déconnecter

- DATHB date de prélèvement (1 patient, 1 doc) +
- ▼ HORMONO (41 patients, 288 docs) +
  - ▼ SHBG (1 patient, 2 docs) +
    - SHBG (1 patient, 2 docs) +
  - ▼ GDS (29 patients, 825 docs) +
    - ▼ GDS2+HbCO+MetHb+Re (19 patients, 729 docs) +
      - HbO2 (10 patients, 80 docs) +
      - HBT Hb totale (10 patients, 80 docs) +
      - MetHb (4 patients, 7 docs) +
    - ▼ carboxyhémoglobine (2 patients, 2 docs) +
      - HbCO (2 patients, 2 docs) +
    - ▶ ph Be foetal (2 patients, 2 docs) +
  - ▼ Hémato (178 patients, 18 717 docs) +
    - ▼ L.Hyperbaso (1 patient, 2 docs) +
      - LHB L.Hyperbaso (1 patient, 2 docs) +
    - ▼ Num Hématies F (178 patients, 16 580 docs) +
      - HB Hémoglobine (178 patients, 883 docs) -

Type: Exclure ▾  
Choisir un intervalle ou une valeur hors intervalle  
Intervalle de valeurs : de  à  12  
Ou X fois la normale supérieure :  fois la normale sup  
Ou valeur anormale basse :   
Ou valeur anormale haute :   
Choix: Tous  
annulé^  
0.2 (N)  
Annulé

+Ajouter | -Annuler

  - Hémoglobine (4 patients, 7 docs) +
  - ▼ Hémato spé (9 patients, 161 docs) +
    - ▼ MYELOGRAMME (4 patients, 116 docs) +
      - 0HB1 Hémoblaste (2 patients, 2 docs) +

patients: 340

Liste des critères structurés

  - #1 Recherche de Exérèse de tumeur de l'angle pontocérébelleux et\_ou du méat acoustique interne [conduit auditif interne], par abord translabyrinthique  
dans : Actes (CCAM) ▶ SYSTÈME NERVEUX CENTRAL, PÉRIPHÉRIQUE ET AUTONOME ▶ ACTES THÉRAPEUTIQUES SUR LE SYSTÈME NERVEUX CENTRAL INTRACRÂNIEN ▶ Exérèse de tumeur intracrânienne extraencéphalique ▶ Exérèse de tumeur de l'angle pontocérébelleux et\_ou du méat acoustique interne [conduit auditif interne], par abord translabyrinthique (Nombre de patients: 56)
  - #2 Recherche de HB Hémoglobine  
dans : labo ▶ Hémato ▶ Num Hématies F ▶ HB Hémoglobine  
avec en options: [min: 12] (Nombre de patients: 169)  
Type:  Inclure  Exclure  
Choisir un intervalle ou une valeur hors intervalle  
Intervalle de valeurs : de 12 à :   
Ou X fois la normale supérieure :  fois la normale sup  
Ou valeur anormale basse :   
Ou valeur anormale haute :   
Choix: Tous  
annulé^  
0.2 (N)  
Annulé

# eHOP : application à la recherche clinique

eHOP    Changer d'étude    Un problème?    Administration    Guillaume BOUZILLE    Se déconnecter

## Résultats examens

Document    Autres documents    Journal d'accès

Concepts : **hemoglobine (1)**

Informations

PRENOM NOM\_1622993136, agée de 32 ans à la date du document  
05/01/2015 - Résultats examen de biologie  
IEP  
UF : 5409-CE ORL PL

RESULTATS DES EXAMENS

N° de demande : 5010585360    Date de demande : 05/01/2015 17:46    Demandeur : 5409 C. EXT. ORL PL  
Date prélèvement : 05/01/2015 17:46    Réalisé par : 3200 MC HITC P

Priorité : S

Libellés	Valeur	Anor	Normes	Etat
Hémato				
Num Hématies F				
appareil	(Impédance-cytométrie en flux, Automate DxH 800 Beckman Coulter)	N		F
9NGR				F
Hématies	5.10 Tera/L	N	4.00 - 5.20	F
<b>Hémoglobine</b>	14.5 g/dL	N	11.5 - 16.0	F
Hématocrite	44.9 %	N	37.0 - 47.0	F
VGM	88 fL	N	79 - 97	F
TCMH	28 pg	N	27 - 32	F
CCMH	32 g/dL	N	30 - 36	F
RDW	11.1	N	11.5 - 14.5	F

# Les Centres de Données Cliniques (CDC)

## Structure de confiance

.....

*Accès l'exhaustivité des données qui lui sont confiées*

*Effectue les traitement et en assure la traçabilité*

*Respect du cadre éthique, juridique et réglementaire*

*Actions pour l'amélioration de la qualité des données*

## Organisation intra-hospitalière Guichet unique

.....

*Compétences multidisciplinaires*

- informatique, réglementaire,
- biostatistiques, machine learning,
- recherche clinique, épidémiologie

*Expertise méthodologique*

*Offre de services*

**Autour de l'exploitation des données santé**

## Centre de Recherche et développement en sciences des données

.....

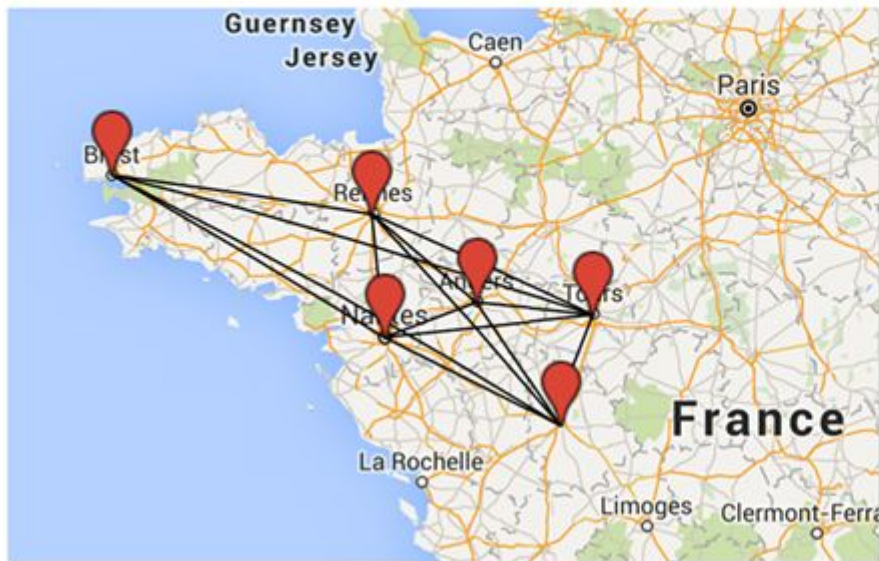
*Rapprocher l'expertise auprès des données, des praticiens et des patients*

*Dynamique de recherche autour de problématiques concrètes et évaluation en vie réelle*



# Réseau interrégional des CDC

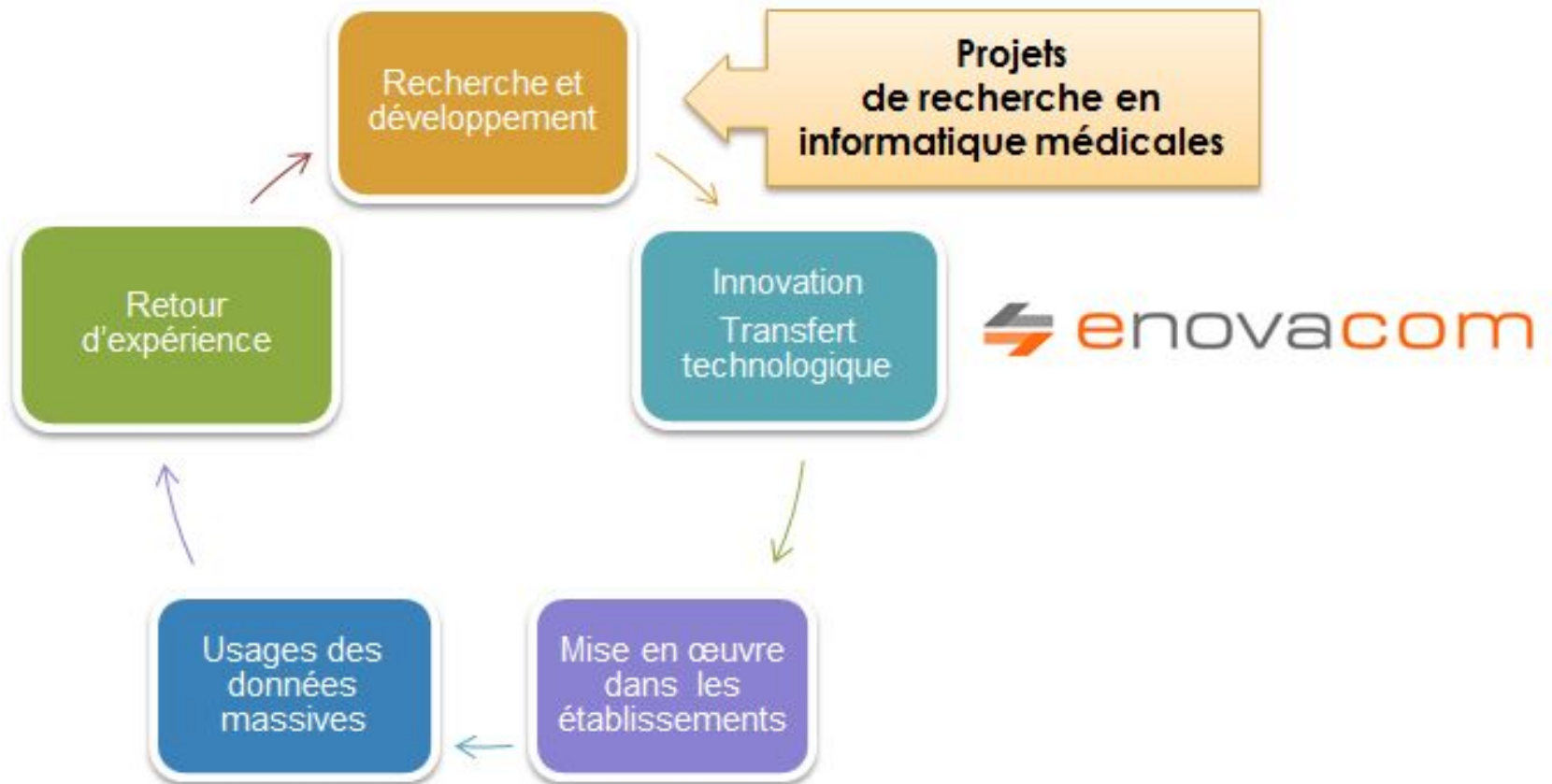
*Vers une exploitation synergique du big data en santé*





# Les Centres de Données Cliniques (CDC)

Une R&D centrée sur les besoins et les usages



# Réseau interrégional des CDC



Développer les usages et la valorisation des données de santé



Exploiter les plateformes eHOP de façon **multicentrique**



Partager les **compétences**, les **méthodes** et les **outils**



Etablir des **partenariats scientifiques, industriels et territoriaux**



Coordonner un **programme de R&D** et des enseignements adaptés



Former aux nouveaux métiers **Master SDS**

# Quels usages ?



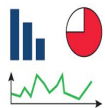
- Pour l'individu
  - Médecine 4P



- Pour les collectivités
  - Vigilances, veille sanitaire



- La recherche
  - Épidémiologique : Détection de cohortes
  - Clinique : Études de faisabilité, Recherche de patients éligibles



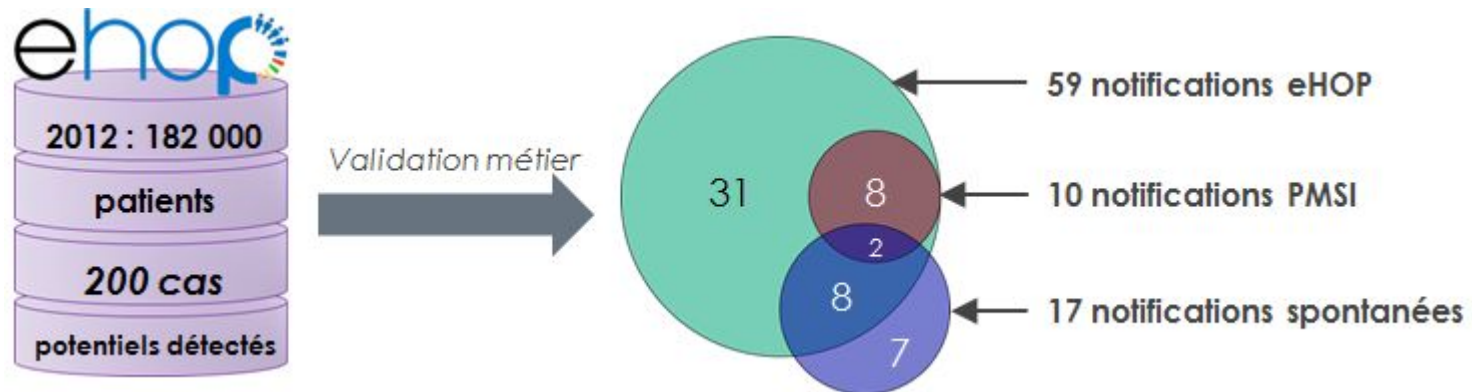
- Médico-administratif
  - organisation raisonnée des soins, pilotage des activités, analyse de trajectoires de santé



- Pour l'enseignement
  - Simulation, éducation

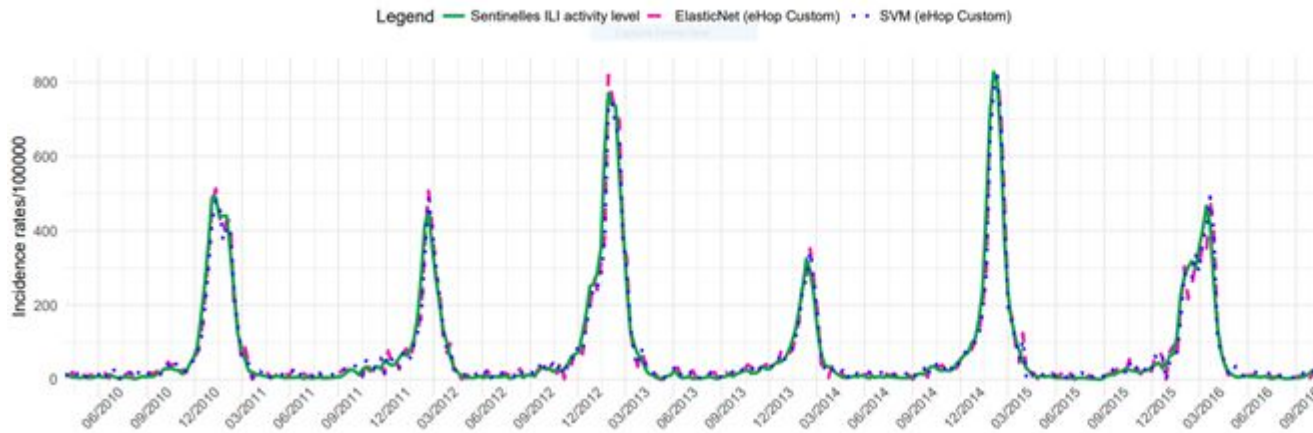
# Exemple pour la pharmacovigilance

## Détection d'effets indésirables médicamenteux par recherche d'information dans les textes.



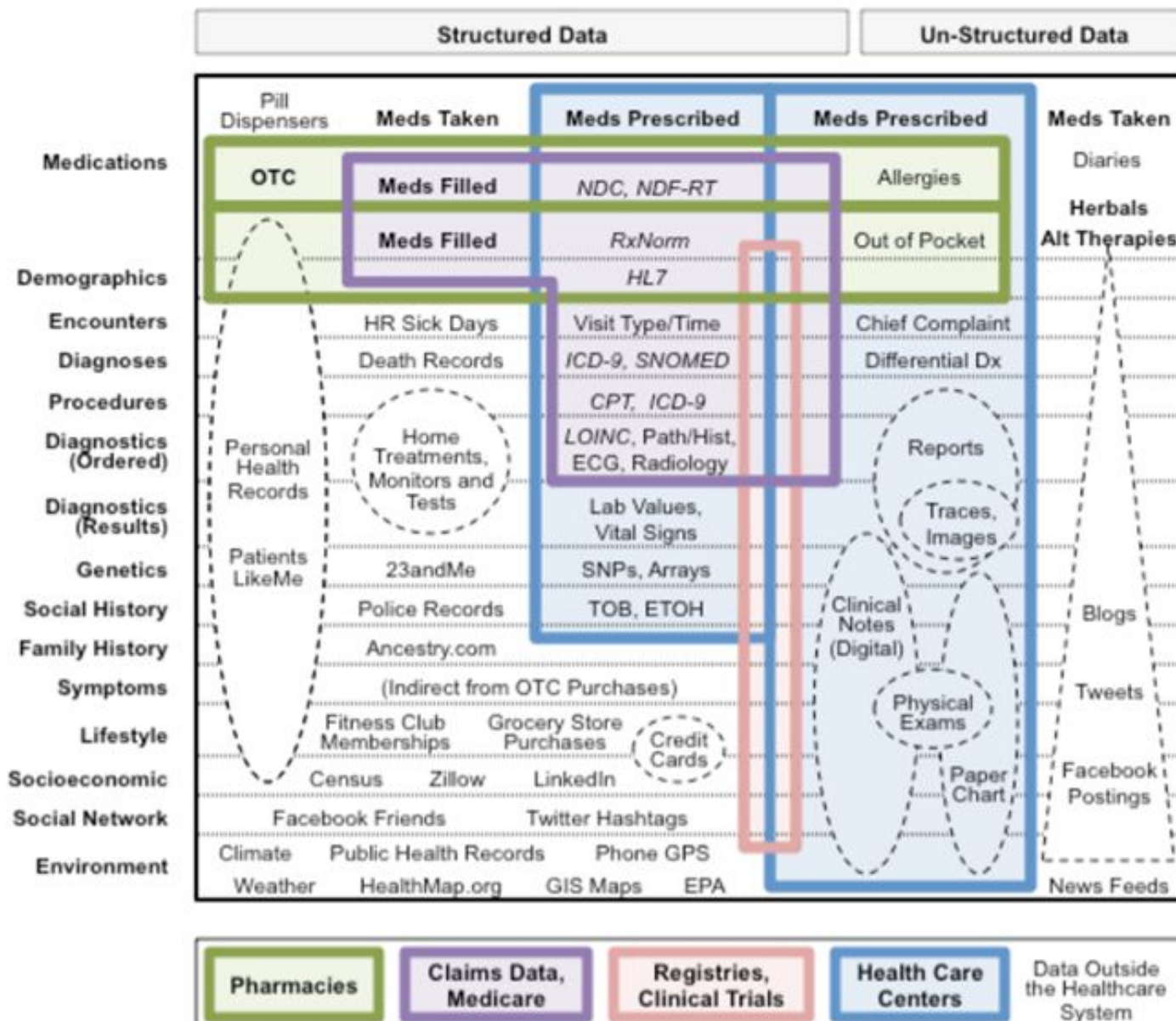
# Exemple pour la surveillance syndromique

Utilisation des données de Sentinelles et de eHOP pour produire des indicateurs temps réels d'activité grippale.



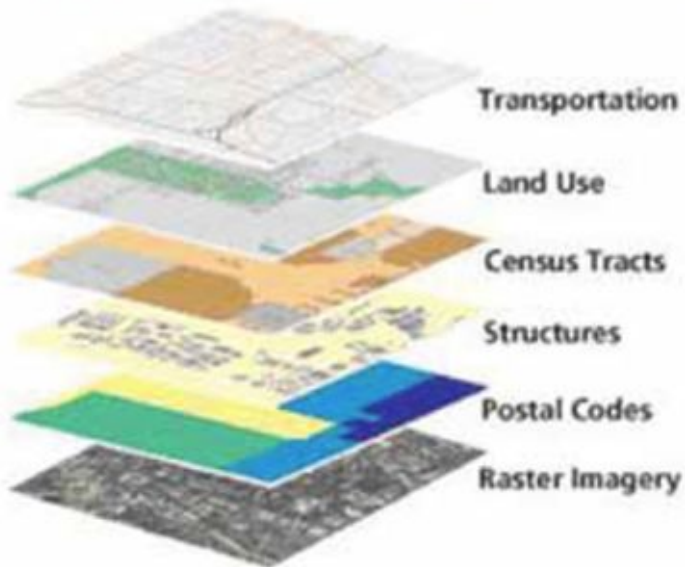
National	Global	
	PCC	MSE
<b>Ehop Custom</b>		
RF	0.947	2292
SVM	<b>0.98</b>	<b>866</b>
Elastic+Arima	0.98	872
<b>Google Custom</b>		
RF	0.937	2607
SVM	0.977	968
Elastic+Arima	0.977	988

# “Big Data” Map of Patient Health

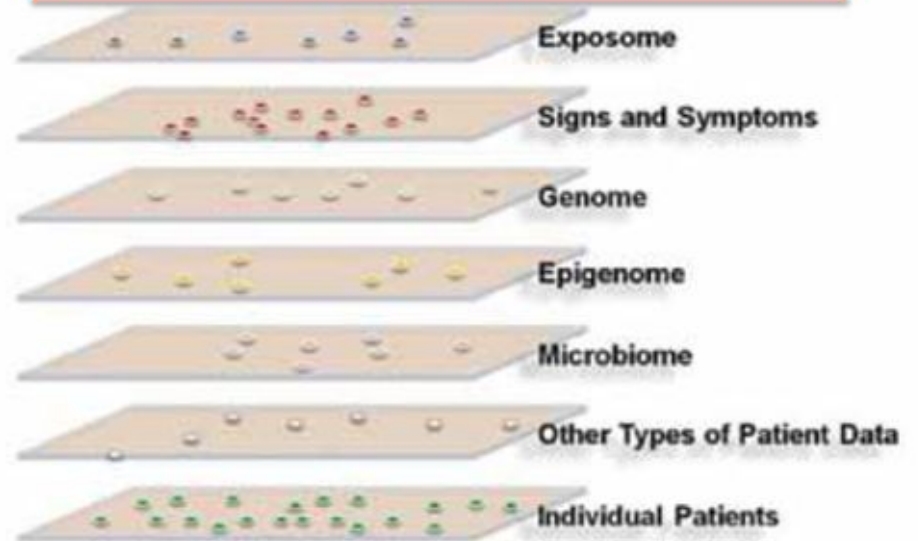


# Un besoin d'intégration des données multi-source multiéchelle

**Google Maps: GIS layers  
Organized by Geographical Positioning**

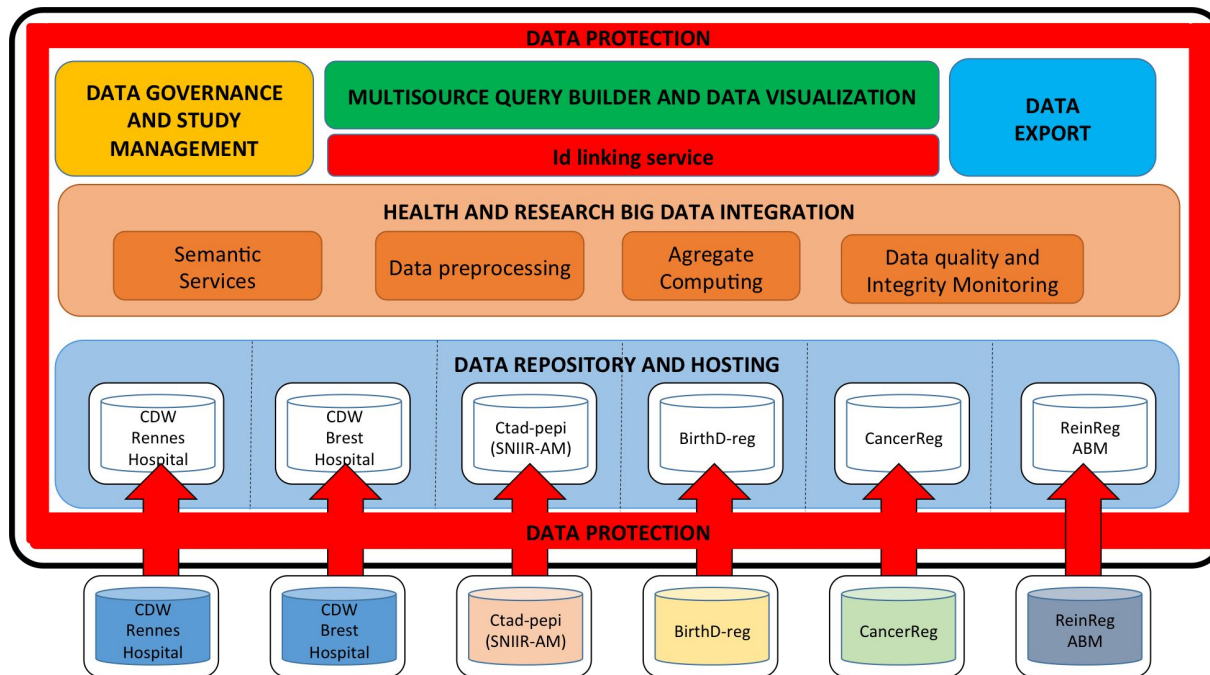


**Information Commons  
Organized Around Individual Patients**





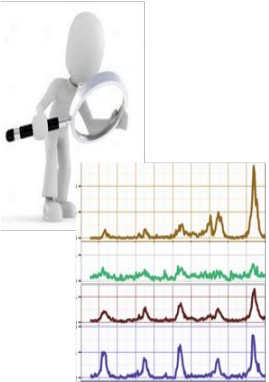
# Le projet Integrating and Sharing Health Data for Research (InSHARE)

Démontrer la faisabilité d'une plateforme d'intégration et de partage de données de santé **multi domaine et multi-échelle**





# Le projet Integrating and Sharing Health Data for Research (InSHARE)

Cas d'usage	Applications	Sources de données utilisées
	<p>Analyse de trajectoire de patients</p> <ul style="list-style-type: none"> <li>• Etude des parcours de soins des patients insuffisants rénaux dialysés en urgence</li> </ul>	<ul style="list-style-type: none"> <li>• Registre Rein</li> <li>• SNIIR-AM</li> </ul>
	<p>Enrichissement de registres épidémiologiques</p> <ul style="list-style-type: none"> <li>• Etude du lien entre diabète et cancer chez les patients en insuffisance rénale terminale dialysés</li> </ul>	<ul style="list-style-type: none"> <li>• Registre Rein</li> <li>• Registre des Cancers de la région Poitou Charentes</li> <li>• SNIIRAM</li> </ul>
	<p>Extractions de signaux à partir du big data en santé.</p> <ul style="list-style-type: none"> <li>• Surveillance syndromique (ex Grippe)</li> <li>• Détection d'effets indésirables liés aux interactions médicamenteuses</li> </ul>	<ul style="list-style-type: none"> <li>• Entrepôts de données biomédicaux hospitaliers</li> <li>• SNIIR-AM</li> <li>• Données publiques du réseau sentinelle</li> <li>• Bases de connaissances médicamenteuses</li> </ul>

# Une collaboration académique et industrielle

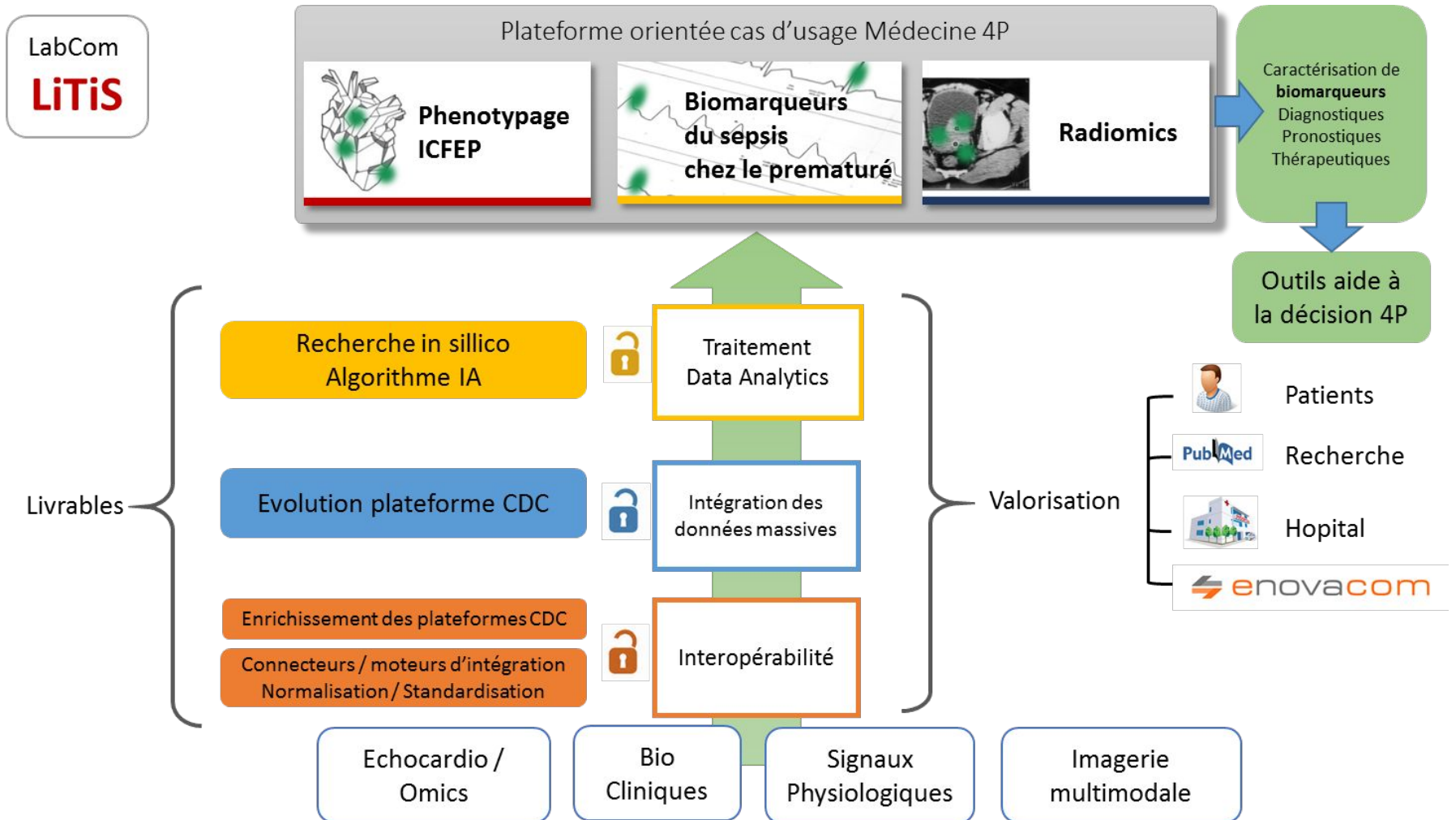
## LabCom LiTiS

**L**aboratoire **I**nteropérabilité  
**T**raitement et **I**ntégration de  
données massives en **S**anté

### **Objectif du LabCom LiTiS :**

Constuire les plateformes big data santé  
orientées Médecine 4P

# Une collaboration académique et industrielle



# Une initiative nationale

Contact : [marc.cuggia@univ-rennes1.fr](mailto:marc.cuggia@univ-rennes1.fr)



## Groupe de travail « Données Massives en Santé »

1. Les **usages** présents et à venir dans l'ensemble des secteurs d'activité des CHU
2. Les **nouveaux métiers** autour du « big data » et leurs places à l'hôpital
3. Les **technologies** à mettre en œuvre, l'enjeu de leurs convergences et leur interopérabilité
4. Le **cadre de gouvernance de protection** des données permettant un développement des usages
5. La question de la démarche **qualité** et la **certification** autour des collections de données.
6. Les modalités de **partage** et d'échange des données, la question de leur **hébergement**, les liens avec le SNIIR-AM et le Système National des Données de Santé.
7. La **valorisation** des données et le **modèle économique** sous-jacent.



*Des questions ?*  
*guillaume.bouzille@univ-rennes1.fr*