



Sciences Economiques & Sociales de la Santé  
& Traitement de l'Information Médicale

sesstim.univ-amu.fr

**Marc CUGGIA**

*Pr,MD,PhD Inserm LTSI - Données Massives en santé  
Centre de données cliniques - CHU de Rennes*

**Big data en santé :**

**Réutilisation des données massives en santé, définition, exemples de cas  
d'usage, infrastructures et méthodes.**

avril 2017



**Cliquez ici pour voir l'intégralité des ressources associées à ce document**

# Exploitation des données massives en santé pour la recherche médicale : méthodes, outils et cas d'utilisation



Pr Marc CUGGIA

Département d'information médicale

Equipe projet données massives en santé

(Inserm UMR 1099 - LTSI)

Unité Support Fouille de données

(Inserm CIC - 1414 - CHU Rennes)



# Plan de l'exposé

- 1 Données massives en santé
- 2 Champs d'utilisations
- 3 Les centres de données clinique et eHOP
- 4 Les réseaux de données de santés
- 5 Les projets de fouilles sur données massive en santé
- 6 Conclusion

# Jusqu' à hier



Quels sont les patients éligibles pour ce projet de recherche ?

Revenez dans 1 mois ...



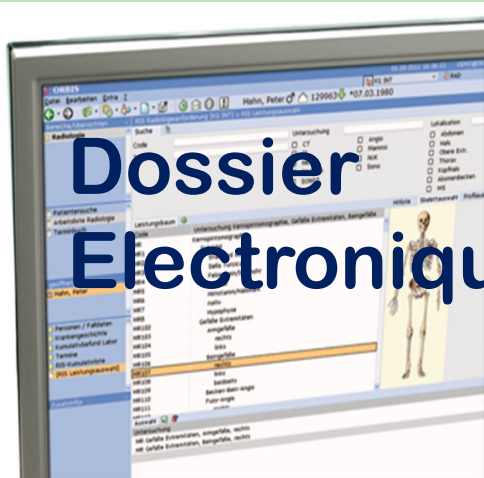
# Jusqu' à aujourd' hui

Un Système d' Information très riche

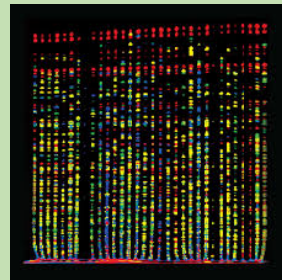
Mais peu exploitable en l' état



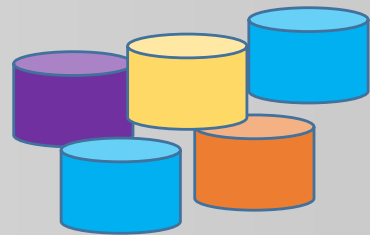
# Dossier Electronique



Imagerie



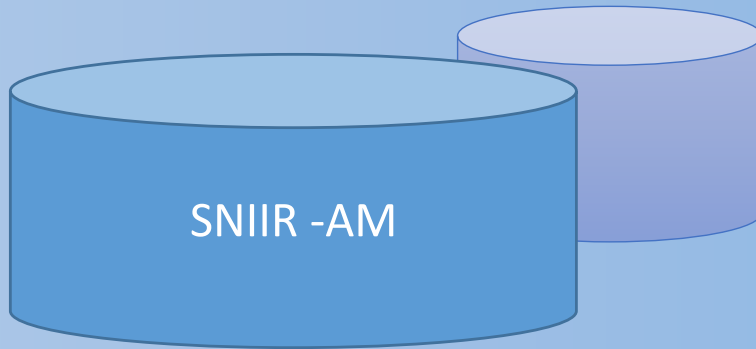
Omic



Base recherche  
Registre (Small data)



SNIIR -AM



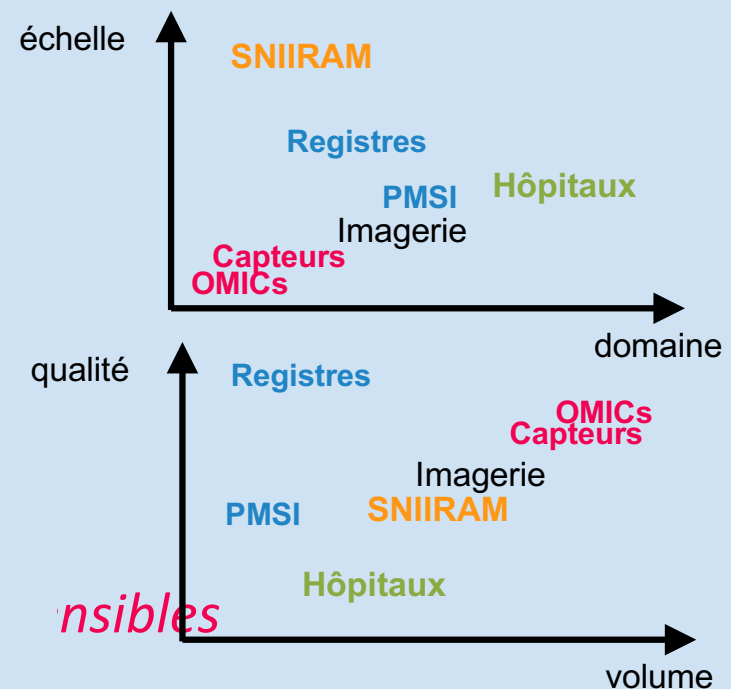
# e-Santé



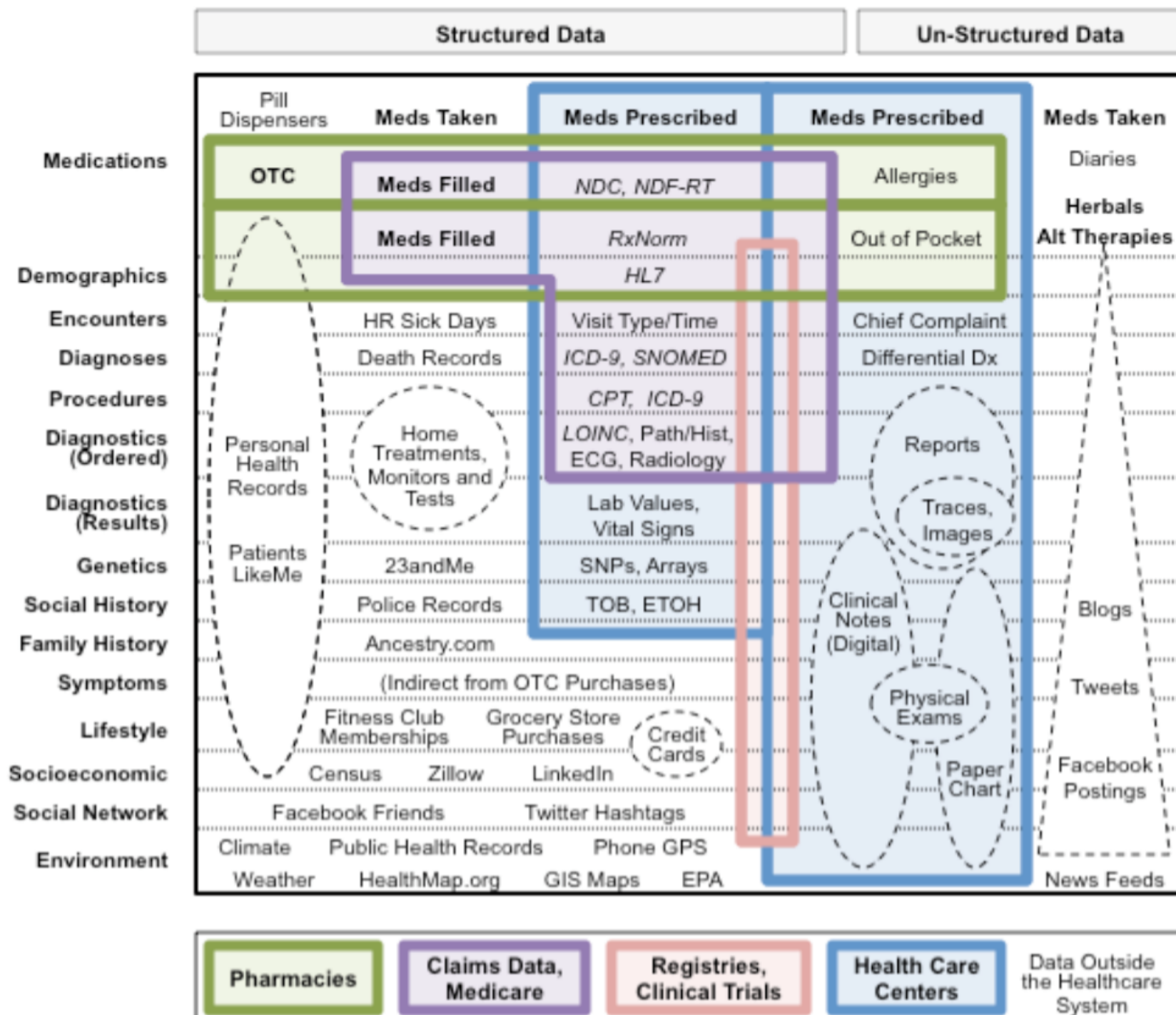
## Dématérialisation des données de santé

### Les “4 V” du Big Data :

- ▷ **Volume :**  
Données clinico-biologiques, OMICs, objets connectés, images, signaux
- ▷ **Variabilité :**  
Des données hétérogènes, multi-domaines multi-échelles
- ▷ **Véracité :**  
Des données de qualité variables
- ▷ **Vélocité :**  
Des données produites au fil de l'eau  
→ réutilisation en temps réel



# “Big Data” Map of Patient Health





# Problématique

## Comment exploiter de façon efficiente ce gisement de données ?

Développer les méthodes et les outils pour :

- Intégrer ces données multi-domaines, multi-échelles
- Tirer de l'information de ces données massives

Dans un cadre compatible avec le caractère sensible des données



# Plan de l'exposé

- 1 Données massives en santé
- 2 Champs d'utilisations
- 3 Les centres de données clinique et eHOP
- 4 Les réseaux de données de santés
- 5 Les projets de fouilles sur données massive en santé
- 6 Se former et travailler au CDC

# Recherche clinique :

## Etude de faisabilité et prescreening

### Criteria

#### Inclusion Criteria:

- 1 Male subjects, over 50 years of age at the time of enrollment.
- 2 Subjects referred to urology for BPH leading to permanent indwelling bladder catheters and are considered poor surgical candidates
- 3 Written informed consent to participate in the study.
- 4 Ability to comply with the requirements of the study procedures

#### Exclusion Criteria:

- 1 Previous surgical treatment for BPH
- 2 Presence of bladder stones
- 3 History of prostate cancer
- 4 Prostate size > 150 g
- 5 History of urethral stenosis or its management
- 6 Known or suspected neurogenic bladder
- 7 History of recent hematochezia in the last 3 months
- 8 Contraindication to intravascular iodinated contrast such as allergies or significant elevated creatinine/renal failure
- 9 Uncorrected coagulopathy
- 1 Subjects who participated in an active stage of any drug, intervention or treatment trial within 30 days of enrollment.
- 1 Subjects with preexisting conditions, which, in the opinion of the investigator, interfere with the conduct of the study.
- 1 Subjects who are uncooperative or cannot follow instructions.
- 1 Mental state that may preclude completion of the study procedure or obtention of infor



MAY 27, 2013

The IRS Mess / Syria's YouTube War / The End of Alimony

TIME



THE  
**ANGELINA**  
EFFECT

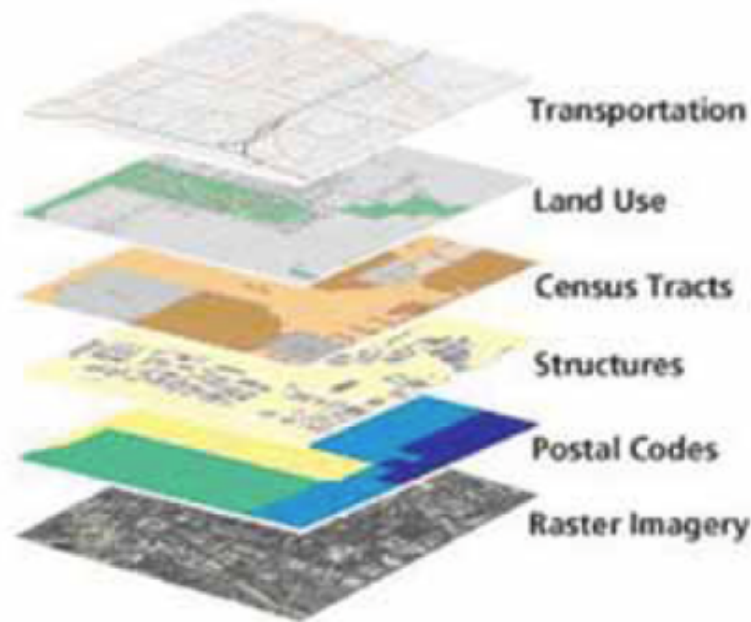
Angelina Jolie's double mastectomy puts genetic testing in the spotlight. What her choice reveals about calculating risk, cost and peace of mind

BY JEFFREY KLUGER & ALICE PARK

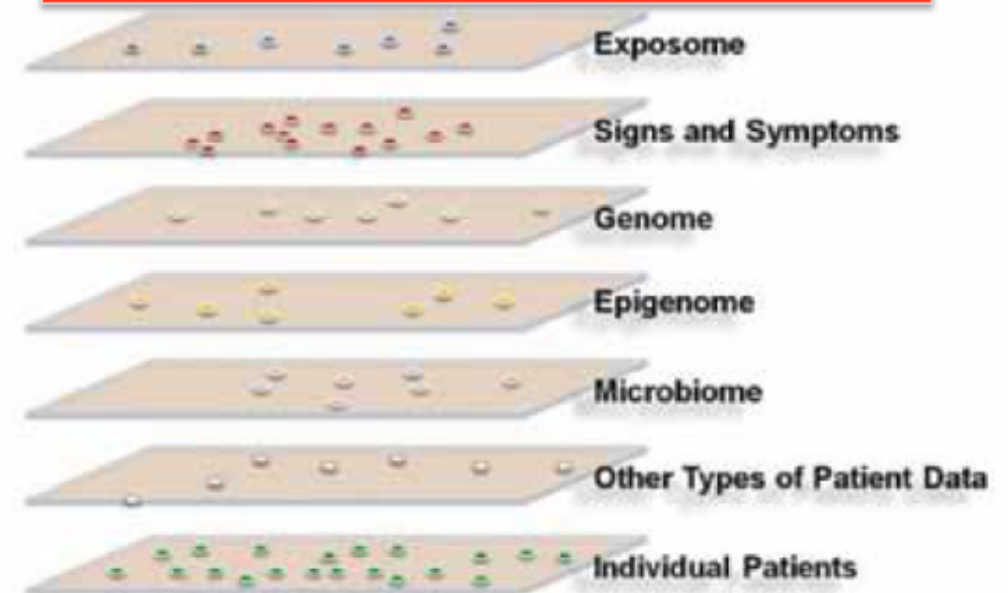
time.com

# Intégration des données pour une Médecine translationnelle, prédictive et personnalisée

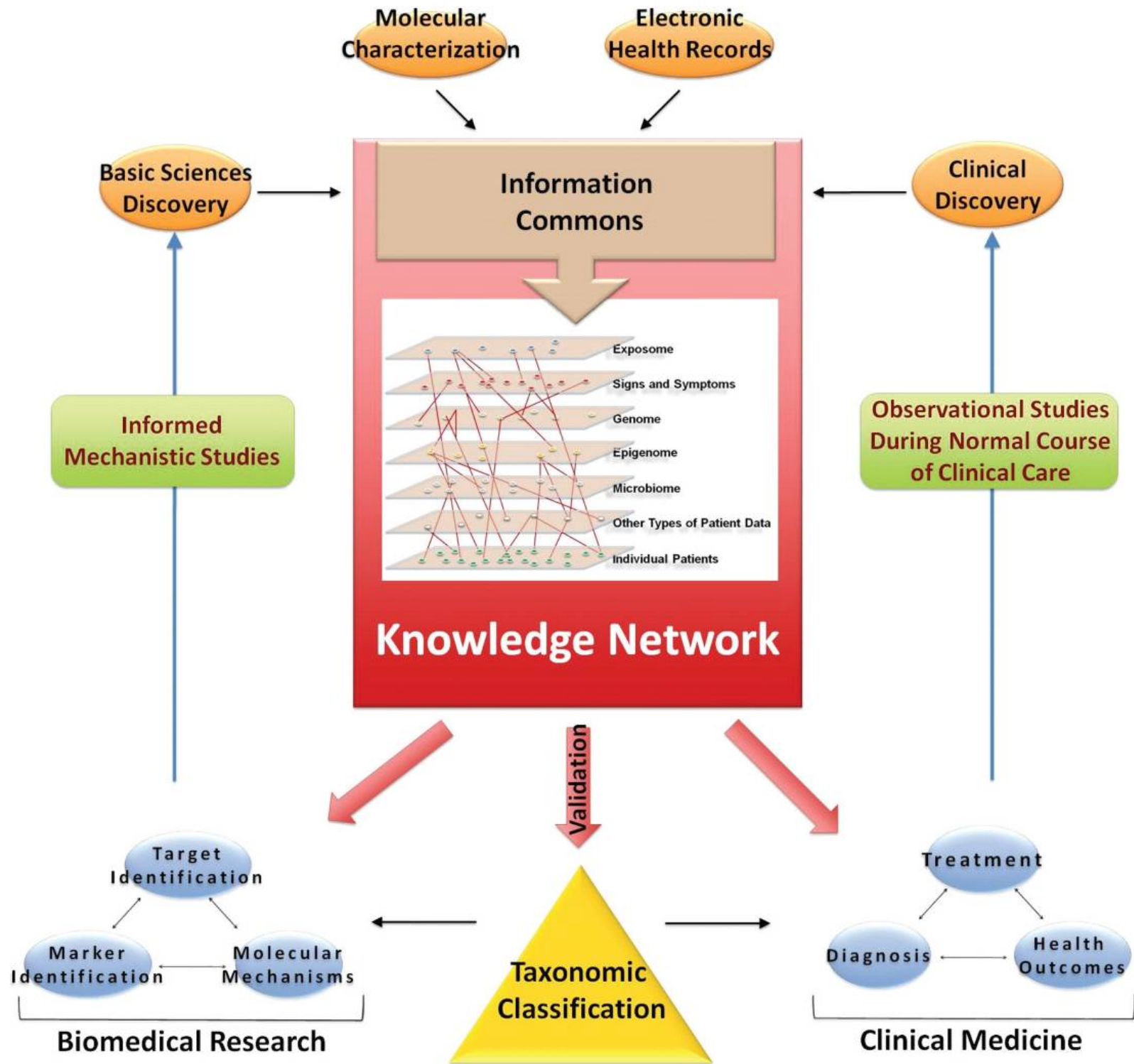
Google Maps: GIS layers  
Organized by Geographical Positioning



Information Commons  
Organized Around Individual Patients



Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease  
Report from National academy of science, USA, 2011



# Quels usages ?

- **Pour l'individu**

- Médecine ciblée et personnalisée

- **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 des trajectoires de santé

- **Pour l'enseignement**

- Professionnels de santé, éducation, information des citoyens

# Plan de l'exposé

- 1 Données massives en santé
- 2 Champs d'utilisations
- 3 Les centres de données clinique et eHOP
- 4 Les réseaux de données de santés
- 5 Les projets de fouilles sur données massive en santé
- 6 Se former et travailler au CDC

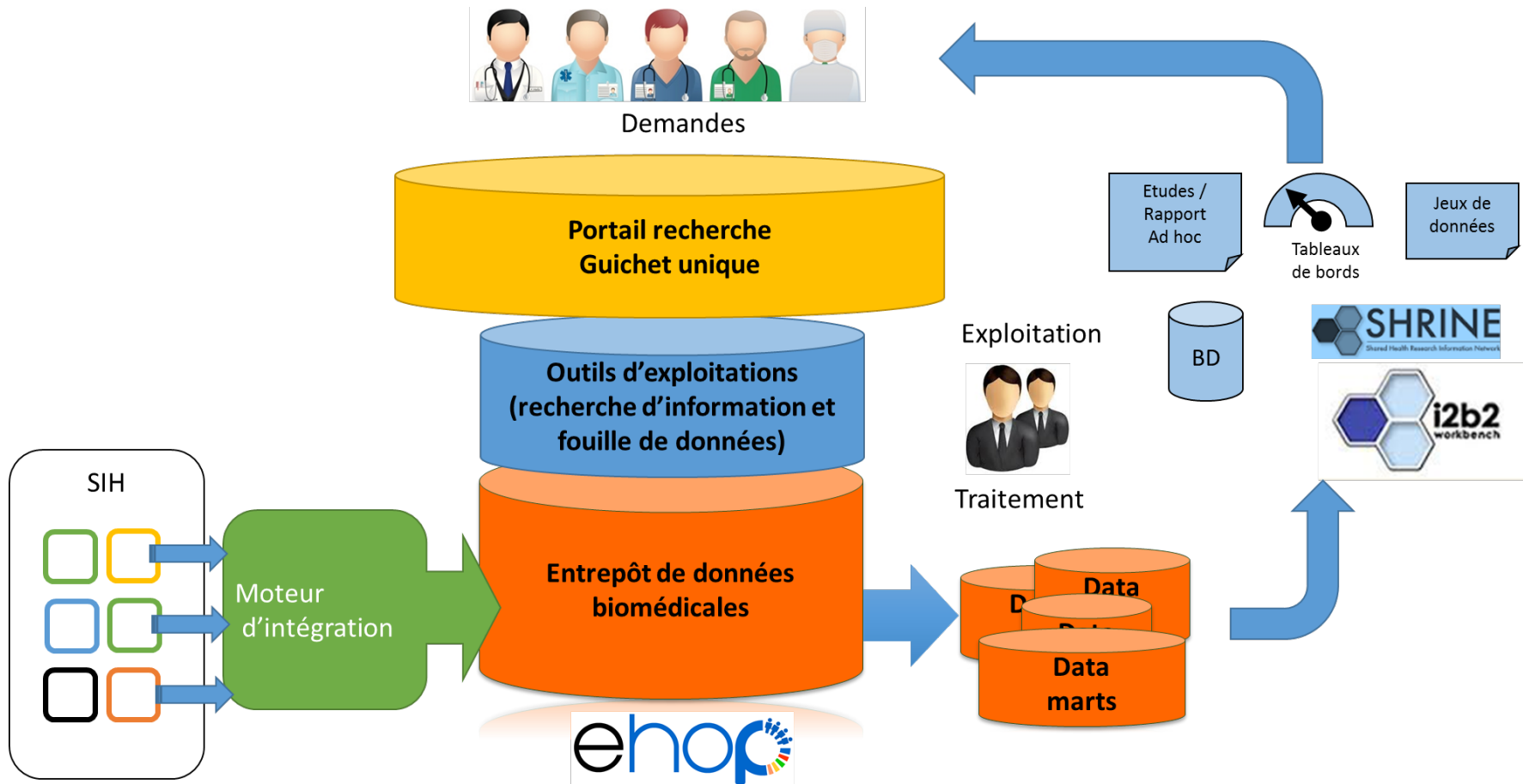


# Centre de Données Clinique CDC

- Organisation intra-hospitalière « **guichet unique** »
- Propose une offre de service au sein de chaque établissement pour **exploiter les données du SIH grâce à l'entrepôt de données et des outils de fouille de données**
- Rassemble des **compétences Informatiques (DSI), réglementaires et méthodologiques (DIM), Métiers** (Recherche clinique, épidémiologie).
- **Structure de confiance**
  - **Accède à l'exhaustivité des données qui lui sont confiées**
  - **Effectue les traitements et en assure la traçabilité**
  - **Respect strict du cadre éthique, juridique et réglementaire**
- Met en œuvre des actions pour l'amélioration de la qualité des données de l'entrepôt. (Ex : optimisation du codage)

# Une infrastructure big data

## Le système eHop : traitement des demandes

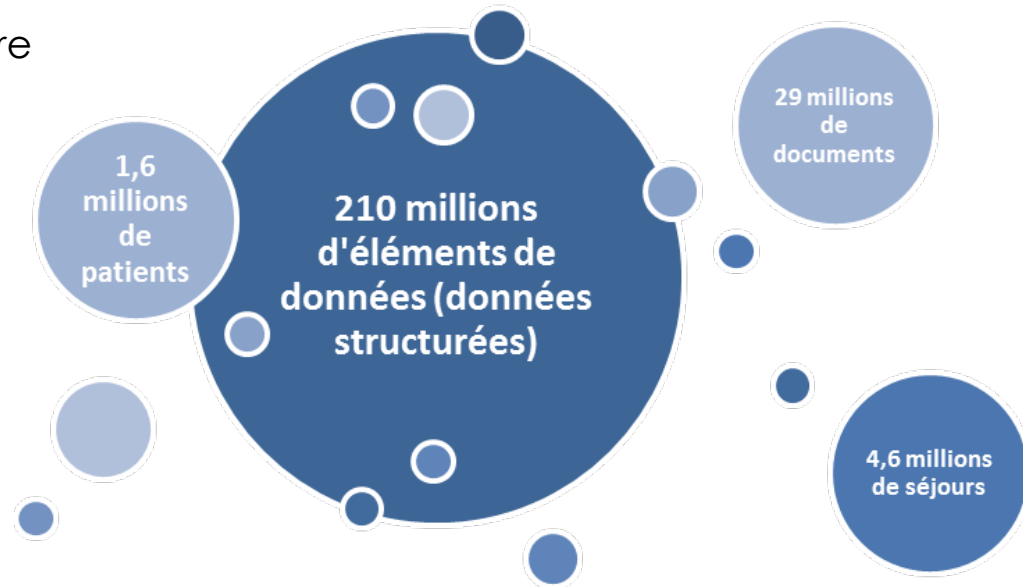


## Exemple du CDC de Rennes

- Sources de données

- Données identité/séjour/structure
- Comptes rendus Hospitaliers
- Prescription/administration
- Bio banque (CRB)
- Dossiers Urgence
- Données de Laboratoires
- Dossiers infirmier (projet)
- PMSI
- Anatomopathologie
- Imagerie

- Entrepôt de données



60 Etudes cliniques et 10 projets de recherche (depuis la MEP)

# Entrepôts de données biomédicales

## Des données structurées :

- ▷ Biologie, diagnostic et actes par le PMSI, Prescription/Administrations médicamenteuses
- ▷ Décrites avec terminologies (standards+++): CIM-10, CCAM, ATC

## Beaucoup de données non structurées :

- ▷ Comptes-rendus (hospitalisation, consultations, urgences, radiologie, etc)
- ▷ Champs de questionnaires médicaux ou infirmiers
- ▷ Zones de commentaires de résultats d'analyses



# Entrepôts de données biomédicales

## Données structurées

The screenshot displays the eHOP (Electronic Health Observation Platform) interface. The top navigation bar includes the eHOP logo, a 'Changer d'étude' button, a 'Un problème?' dropdown menu, an 'Administration' link, the user name 'Pierre ANO', and a 'Se déconnecter' button.

The main content area is divided into two panels. The left panel shows a hierarchical tree of medical categories with expandable/collapsible icons. The selected category is 'I10 hypertension essentielle (primitive) (11 patients)'. Below the tree, there are controls for 'Type' (set to 'Inclure'), 'Choix' (set to 'Toutes'), and buttons for '+Ajouter' and 'Annuler'. Other categories visible include 'ACTES (CCAM)', 'Appareil-Organe (ADICAP)', 'Classification ATC', 'Cytopathologies (ADICAP)', 'Diagnostics (CIM-10)', and various sub-categories like 'I-certains maladies infectieuses et parasitaires (8 patients)', 'Xx-causes externes de morbidite et de mortalite (22 patients)', etc.

The right panel shows the search results for the query 'Recherche de I10 hypertension essentielle (primitive) dans Diagnostics (CIM-10) (0 patient(s))'. It includes tabs for 'Requête', 'Résultats', and 'Statistiques'. The 'Requête' tab is active, showing the search context and a list of structured criteria. Below the criteria, there are sections for 'SQL (2 requêtes exécutées en 0.05s)' and 'JSON'. A message at the bottom states 'ⓘ Cette requête n'est pas enregistrée.' At the bottom right, there are three buttons: '☆ Nouvelle requête', '📄 Enregistrer', and '▶ Exécuter'.

# Entrepôts de données biomédicales

## Données structurées

The screenshot displays the eHOP (Electronic Health Observation Platform) interface. The top navigation bar includes the eHOP logo, the user name 'Pierre ANO', and a 'Se déconnecter' button. The main content area is divided into a left sidebar for navigation and a right pane for search results.

**Navigation Menu (Left Sidebar):**

- ▶ ACTES (CCAM)
- ▶ Appareil-Organe (ADICAP)
- ▶ Classification ATC
- ▶ Cytopathologies (ADICAP)
- ▼ Diagnostics (CIM-10)
  - ▶ I-certains maladies infectieuses et parasitaires (8 patients) +
  - ▶ Xx-causes externes de morbidite et de mortalite (22 patients) +
  - ▶ Xxi-facteurs influant sur l'etat de sante et motifs de recours aux services de sante (24 patients) +
  - ▶ Iii-maladies du sang et des organes hematopoiétiques et certains troubles du systeme immunitaire (19 patients) +
  - ▶ Iv-maladies endocriniennes, nutritionnelles et metaboliques (20 patients) +
  - ▶ Xxii-codes d'utilisation particuliere (2 patients) +
  - ▶ V-troubles mentaux et du comportement (11 patients) +
  - ▶ Vi-maladies du systeme nerveux (10 patients) +
  - ▶ Vii-maladies de l'oeil et de ses annexes (1 patient) +
  - ▶ Viii-maladies de l'oreille et de l'apophyse mastoïde (1 patient) +
  - ▼ Ix-maladies de l'appareil circulatoire (26 patients) +
    - ▼ Maladies hypertensives (12 patients) +
      - I10 hypertension essentielle (primitive) (11 patients) -
  - Type: Inclure ▼
  - Choix: Toutes ▼
  - +Ajouter | Annuler
  - ▶ I11 cardiopathie hypertensive (1 patient) +
  - ▶ I12 nephropathie hypertensive (1 patient) +
  - ▶ Cardiopathies ischémiques (3 patients) +
  - ▶ Affections cardiopulmonaires et maladies de la circulation pulmonaire (1 patient) +
  - ▶ Autres formes de cardiopathies (10 patients) +

**Search Results (Right Pane):**

Requête | Résultats | Statistiques

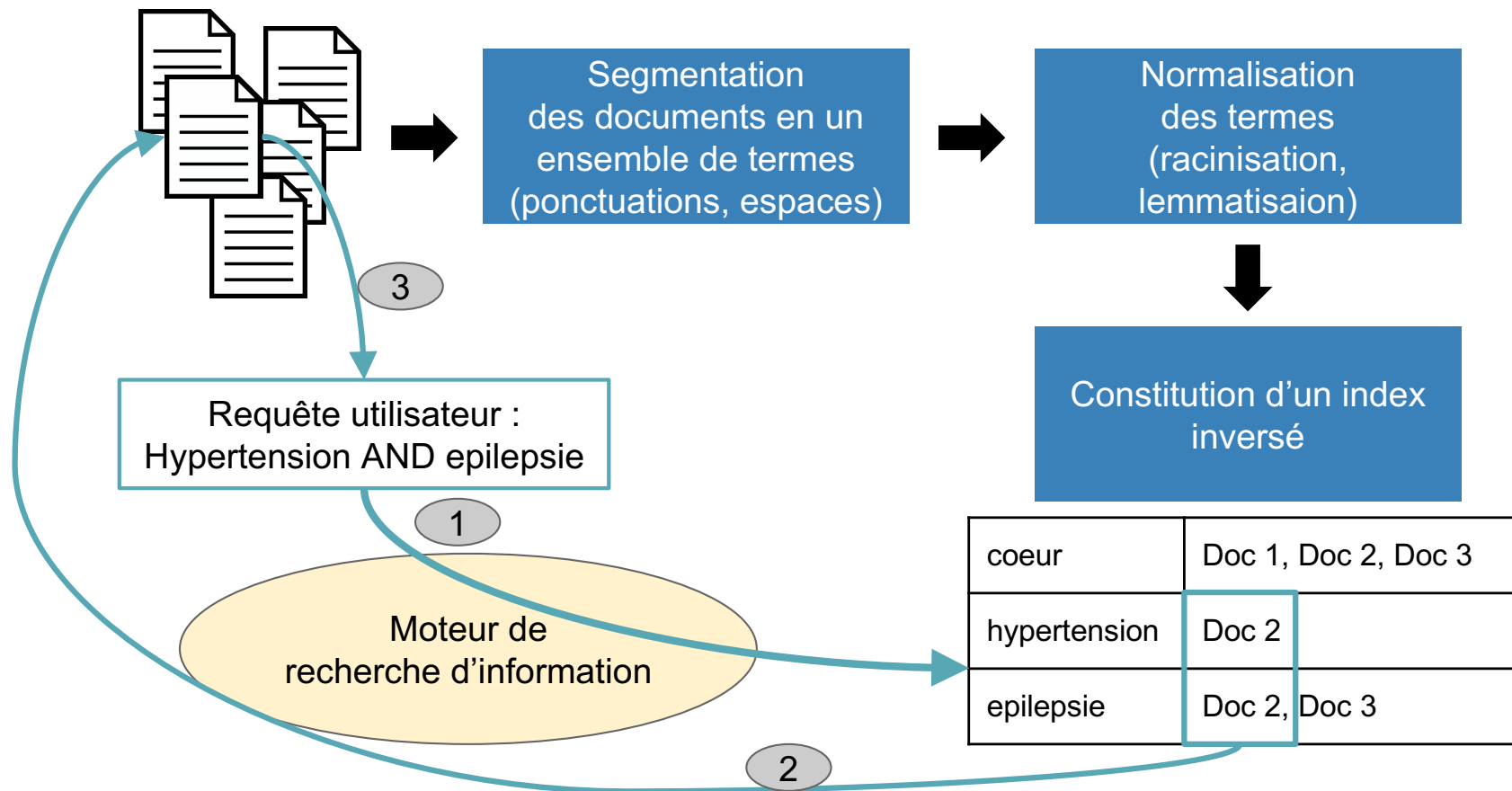
Actuellement affiché: 11 sur 11 patient(s), 14 sur 14 document(s). Requête exécutée en 0.05s.

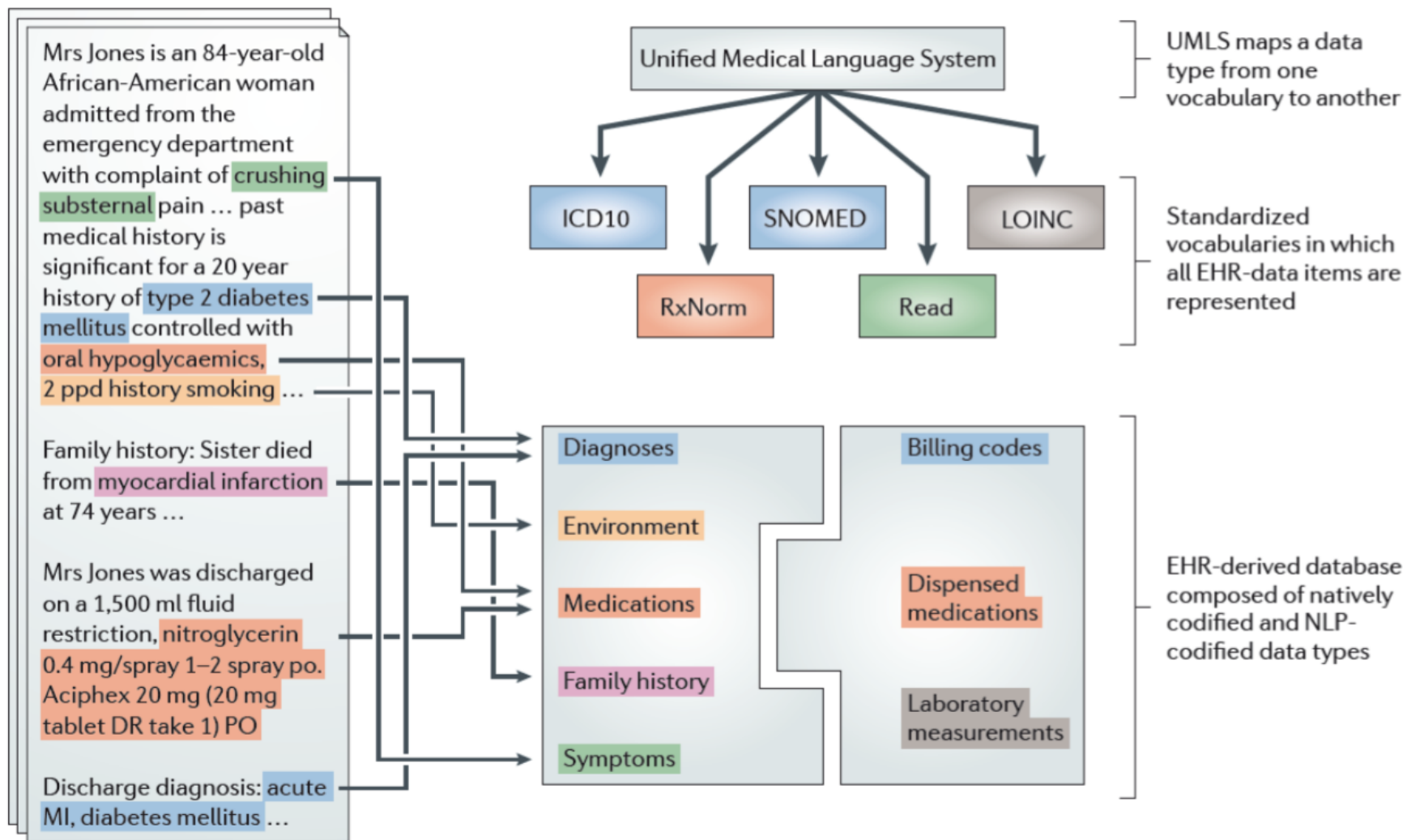
Concepts +

IPP	Nom	Prénom	Date naissance	S	CP	Actions	Document(s)
10001	LOULLE_ZZ	Robin_	11/02/1972	M	-		Diagnostics 04/ PMSI
10022	VENDELEE_ZZ	Pierre_	18/01/1932	M	-		Diagnostics 01/ PMSI Diagnostics 14/ PMSI Diagnostics 24/ PMSI
10024	TERNAY_ZZ	Jean_	11/10/1943	M	-		Diagnostics 20/ PMSI
10028	HILAIRE_ZZ	Michele_	01/09/1928	F	-		Diagnostics 28/ PMSI

# Entrepôts de données biomédicales

Données non structurées : recherche d'information





Kohane I, Nature Review Gen. 2011



# Entrepôts de données biomédicales

## Données non structurées : traitement automatique du langage

Rennes le 25/01/2016

Cher Confrère,

Votre patient, **Monsieur Martin Julien** (né le 21/06/1945), a été hospitalisé du 3 au 15/01/2016 en secteur unité du service de Médecine Interne pour **chute avec anémie et hypercalcémie**.

Antécédents :

- Médicaux :
  - HTA, Hypertrophie bénigne de prostate
  - 2012 : plusieurs épisodes de coma hypoglycémique et découverte d'un **insulinôme**
- Chirurgicaux :
  - adénomectomie transvésicale,
  - exérèse insulinôme (pT1N0M0).

Histoire de la maladie :

- Depuis environ 2 mois le patient décrit une altération de l'état général (anorexie asthénie) avec perte d'autonomie. Ces éléments sont concomitants à l'apparition de rachialgies d'horaire mixte mais invalidantes selon le patient.
- Le 03/01/2016, fait une chute d'allure mécanique, sans perte de connaissance, en voulant ramasser un prospectus au sol. Le patient n'a pas réussi à se relever et les voisins ont alerté les secours.
- Aux urgences, on constate des signes d'incurie ainsi qu'une pancytopenie, une hypercalcémie et des lacunes osseuses au scanner cérébral.

Clinique à l'entrée :

- signes généraux :
  - OMS 2, TA 105/66, FC 91, Sat 97 % AA, 36.8°C
  - altération de l'état général avec perte de poids (59kg pour 70 habituellement)
  - [...]
- absence de troubles digestifs, de confusion en lien avec l'hypercalcémie
- adénopathie centimétrique inguinale gauche
- absence de signes hémorragiques
- reste de l'examen normal.

Examens complémentaires à l'entrée :

- Anémie à 8 g/dL, macrocytaire **VGM 106 fL**, arégénérative (réticulocytes 44G/L) thrombopénie à 107 g/L

Extraction de concepts : diagnostics, actes

Contextualisation : antécédents

Contextualisation : chronologie

Contextualisation : négation, incertitude

Extraction de concepts : valeurs biologiques, prescription et posologies

Triplet : concept / valeur / unité

# Entrepôts de données biomédicales

## Approche type recherche d'information

The screenshot displays the eHOP web application interface. The top navigation bar includes the eHOP logo, 'Changer d'étude', 'Un problème?', 'Administration', the user name 'Pierre ANO', and a 'Se déconnecter' button. The main content area is divided into several sections:

- Portée de l'étude:** Shows 'Nombre de patients 38' and 'Nombre de documents 15773'.
- Requêtes enregistrées (1 élément):** A section for saved queries.
- Recherche avancée:** A section for advanced search with the following criteria:
  - Critères généraux:** Expanded to show 'Recherche Textuelle'.
  - Type:** Radio buttons for 'Inclure' (selected) and 'Exclure'.
  - Document:** A dropdown menu.
  - Unité(s) Fonctionnelle(s):** A text input field containing 'Unité(s) Fonctionnelle(s)'.
  - Requête:** A text input field containing 'fracture', which is highlighted with a red box.
  - Périphrases négatives:** A checked checkbox for 'Inclure'.
  - Enrichissement sémantique:** An unchecked checkbox for 'Etendre'.
- Statistiques:** A box indicating 'Actuellement affiché: 13 sur 13 patient(s), 21 sur 21 document(s)' (highlighted with a red box) and 'Requête exécutée en 0.06s.'.
- Concepts:** A section for concepts.
- Tableau de résultats:** A table with columns: IPP, Nom, Prénom, Date naissance, S, CP, Actions, and Document(s). The table lists five patient records with their respective medical documents.

IPP	Nom	Prénom	Date naissance	S	CP	Actions	Document(s)
10020	SALENCY_ZZ	Noémie_	01/04/1925	F	-		CR hospitalier (séjour) 12,
10024	TERNAY_ZZ	Jean_	11/10/1943	M	-		CR d'imagerie médicale 06, CR hospitalier (séjour) 18,
10025	ECHALAS_ZZ	Noela_	21/11/1927	F	-		CR hospitalier (séjour) 28,
10031	DOLE_ZZ	Michel_	05/02/1942	M	-		CR hospitalier (séjour) 28,

eHOP [Changer d'étude](#) [Un problème? ▾](#) [Administration](#) [Export](#) Guillaume BOUZILLE [Se déconnecter](#)

---

**Recherche rapide**

Mots-clés ...

---

**Contexte: Ensemble de l'étude**

Portée de l'étude

Nombre de patients **1609308**

Nombre de documents **27595479**

---

**Requêtes enregistrées (1 élément)**

6236015	retacrit	delamarr	rév. 0	23/08/2016 14:46:14	<input type="checkbox"/>
---------	----------	----------	--------	---------------------	--------------------------

---

**Recherche avancée**

Critères généraux

Recherche Textuelle

Type  Inclure  Exclure

Requête

Document

Unité(s) Fonctionnelle(s)

Périphrases négatives  Inclure

Enrichissement sémantique  Etendre

---

**Requête\*** [Résultats](#) [Statistiques](#)

Contexte

La recherche s'effectue sur **l'ensemble de l'étude**.

Critères Généraux

Les documents datés du 01/01/2014 à aujourd'hui  
Unités fonctionnelles: 5401,5403,5405,5409  
Intersection des critères sur le patient

Liste des critères textuels

**Inclure** les patients dont les documents contiennent l'expression **"neurinome or schwannome"**

JSON

Cette requête n'est pas enregistrée.

**Etude:** Etude neurinome (version démo) (#1996) +

Recherche rapide -

Mots-clés ... 🔍

Contexte: Ensemble de l'étude +

Requêtes enregistrées (2 éléments) -

#	Titre	👤	Rév.	Date	Actions
6236089	intraveineuse	admin	rév. 0	29/09/2016 11:27:39	📄
6236051	Requête initiale	bouzille	●	31/08/2016 12:43:41	📄

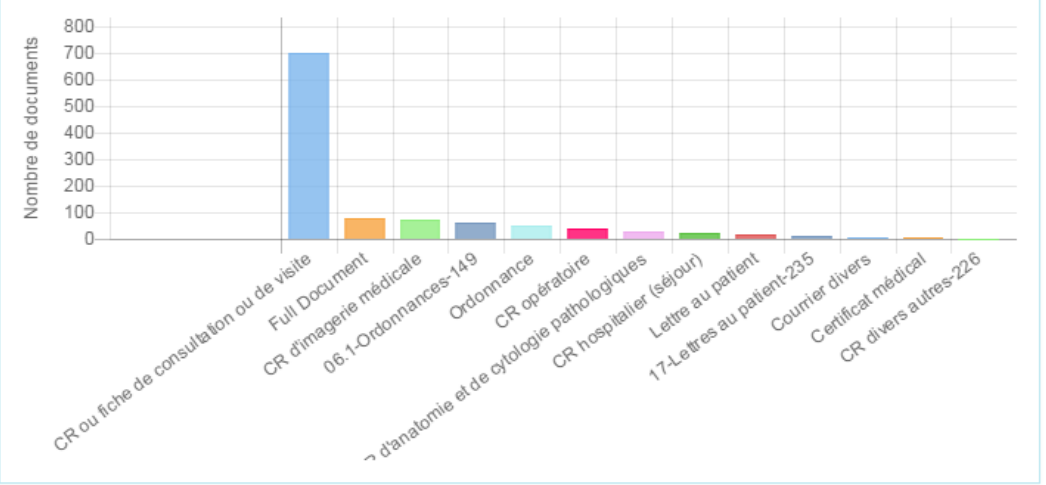
Recherche avancée -

- Critères généraux +
- Recherche Textuelle +
- Recherche Structurée +
- Elements +

Gestion des paniers +

Gestion des datamarts +

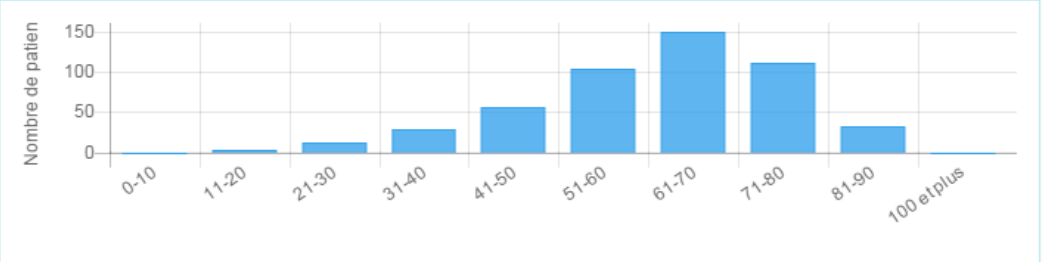
Répartition par type de document -



Répartition des documents par UF/UM +

Nombre de séjours par UF/UM +

Age du patient maintenant -



Etude: Etude neurinome (version démo) (#1996) +

Recherche rapide -

Mots-clés ... 🔍

Contexte: Ensemble de l'étude +

Requêtes enregistrées (2 éléments) -

#	Titre	👤	Rév.	Date	Actions
6236089	intraveineuse	admin	rév. 0	29/09/2016 11:27:39	📄
6236051	Requête initiale	bouzille	●	31/08/2016 12:43:41	📄

Recherche avancée -

- Critères généraux +
- Recherche Textuelle +
- Recherche Structurée +
- Elements +

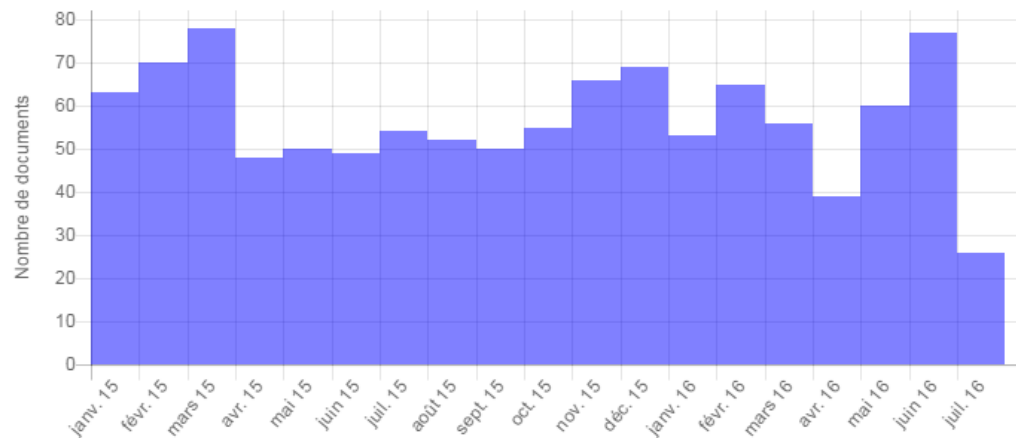
Gestion des paniers +

Gestion des datamarts +

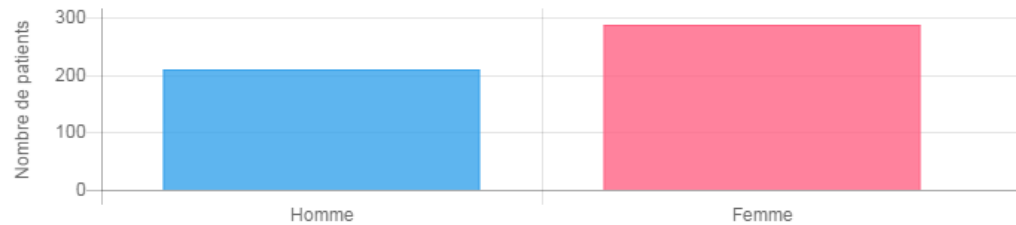
### Répartition des documents au cours du temps -

Groupement par :

Mois ▾



### Répartition par sexe -



### Répartition par type de document +

**Etude:** Etude neurinome (version démo) (#1996) +

**Recherche rapide** -

Mots-clés ... 🔍

**Contexte:** Ensemble de l'étude -

Portée de l'étude

Nombre de patients 498

Nombre de documents 32 348

**Requêtes enregistrées (2 éléments)** -

#	Titre	👤	Rév.	Date	Actions
6236089	intraveineuse	admin	rév. 0	29/09/2016 11:27:39	📄
6236051	Requête initiale	bouzille	●	31/08/2016 12:43:41	📄

**Recherche avancée** -

📄 Critères généraux +

🔍 Recherche Textuelle +

📄 Recherche Structurée +

📄 Elements +

**Gestion des paniers** +

**Requête\*** **Résultats (498)** **Statistiques**

Actuellement affiché: 50 sur 498 patient(s), 84 sur 1 080 document(s). Requête exécutée en 2.38s.

**Concepts** +

IPP	Nom	Prénom	Age	S	CP	Actions	Document(s)	
IPP	NOM_PATIENT1	PRENOM	74 ans	F	35120	📄 📄 📄	Compte Rendu de Consultation	12/02/2015 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT2	PRENOM	87 ans	M	50600	📄 📄 📄	ORL_CRC 02.05.2016	02/06/2016 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT3	PRENOM	53 ans	M	53410	📄 📄 📄	ORL_CRC 26.01.2016	17/02/2016 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT4	PRENOM	52 ans	F	35170	📄 📄 📄	Compte Rendu de Consultation	26/01/2015 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT5	PRENOM	60 ans	F	35230	📄 📄 📄	Compte Rendu de Consultation	12/08/2015 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT6	PRENOM	41 ans	M	35200	📄 📄 📄	ORL_CRC le 02.11.2015	05/11/2015 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT7	PRENOM	72 ans	F	35200	📄 📄 📄	Compte Rendu de Consultation	18/02/2015 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT8	PRENOM	63 ans	F	35520	📄 📄 📄	ORL_CRC 21.03.2016	31/03/2016 <span style="float: right;">👁️ Afficher</span>
IPP	NOM_PATIENT9	PRENOM	66 ans	M	35200	📄 📄 📄	Compte Rendu de Consultation	20/01/2015 <span style="float: right;">👁️ Afficher</span>

eHOP
ORL\_CRC 10.05.2016
Se déconnecter

Patients décédés à l'hôpital

Intersection des critères

Recherche Textuelle

Type

Requête

Document

Unité(s) Fonctionnelle(s)

Périphrases négatives

Enrichissement sémantique

Recherche Structurée

Masquer noeuds

- ▶ Actes (CCAM)
- ▶ Appareil-Organe
- ▶ Classification ATC
- ▶ Cytopathologies
- ▶ DXCARE CONC
- ▶ DXCARE QUEST
- ▶ Diagnostics (CIM)
- ▶ Pathologies Tumorales
- ▶ Pathologies non Tumorales
- ▶ Prélèvement-Technique
- ▶ labo

Gestion des paniers

Document
Autres documents
Journal d'accès

\* Concepts positifs: schwannome

Informations

PRENOM NOM\_PATIENT3, âgée de 65 ans à la date du document  
02/06/2016 - CR ou fiche de consultation ou de visite  
IPP, IEP  
UF : 5409-CE ORL PL

Docteur Bernard CADRE

Chirurgien Maxillo-Facial

Chirurgien des Hôpitaux

e-mail : bernard.cadre@chu-rennes.fr

DEBERGE Sarah

Interne

Docteur HELOURY ANTOINE

24 RUE DE LA GARE

22340 MAEL CARHAIX

Rennes, le jeudi 2 juin 2016

Réf. : FN / SD

Consultation du 10/05/2016

Cher Confrère,

Je vois ce jour en consultation Madame [PRENOM] [NOM\_PATIENT3] née le [Agé(e) de 65] dans le cadre de mon mémoire sur la qualité de vie des petits neurinomes suivis au CHU de Rennes.

Madame [NOM\_PATIENT3] a été opérée d'un schwannome vestibulaire droit par voie translabyrinthique en 2007. La dernière IRM remonte en 2012 et met en évidence une exérèse complète sans récidive.

Les acouphènes sont permanents au niveau de son oreille droite mais ils ne la gênent pas.

Il n'y a pas de syndrome vestibulaire. Elle a seulement une paralysie faciale lorsqu'elle se fatigue, sinon il n'y en a pas.

L'otoscopie est normale. L'audiogramme met en évidence une audition tout à fait stable et normale

cutée en 0.88s.

3/2015 [Afficher](#)

5/2015 [Afficher](#)

3/2015 [Afficher](#)

2/2016 [Afficher](#)

02/2016 [Afficher](#)

02/2016 [Afficher](#)

05/2016 [Afficher](#)

05/2016 [Afficher](#)

7/2015 [Afficher](#)

3/2015 [Afficher](#)

2/2016 [Afficher](#)

3/2016 [Afficher](#)

ORL\_CRC 10.05.2016

Se déconnecter

Document    Autres documents    Journal d'accès

Afficher 50 éléments    Rechercher :

Date	Séjour	UF	Titre	Type
02/06/2016	174470030	5409	ORL_CRC 10.05.2016	CR ou fiche de consultation ou de visite
30/03/2016	172685422	4541	PMSI_ACTE	Actes CCAM
15/06/2012	150278196	6309	RÃ©sult. IRM 06.2012	CR ou fiche de consultation ou de visite
06/06/2012	150278196	6309	crc du 24.05.12	CR ou fiche de consultation ou de visite
14/02/2012	148613185	5409	Compte Rendu de Consultation	CR ou fiche de consultation ou de visite
10/01/2012	147952516	6309	crc du 3.01.12	CR ou fiche de consultation ou de visite
16/11/2011	146943881	6301	cro du 4.11.11	CR opératoire
11/11/2011	146943881		PMSI_DIAG	Diagnostics PMSI
11/11/2011	146943881	6301	Résultats examens	Résultats examen de biologie
04/11/2011	146943881	6301	PMSI_ACTE	Actes CCAM
03/11/2011	146943881	6301	Résultats examens	Résultats examen de biologie
20/10/2011	146824689	6555	PMSI_ACTE	Actes CCAM
23/09/2011	146285960	6309	crc du 19.09.11	CR ou fiche de consultation ou de visite
11/02/2011	137430660	5409	Ordonnance	Ordonnance
11/03/2010	137430660	5409	Compte Rendu de Consultation	CR ou fiche de consultation ou de visite
05/03/2010	137430660	5409	PMSI_ACTE	Actes CCAM
29/01/2008	126255332	5409	Compte Rendu de Consultation	CR ou fiche de consultation ou de visite
05/03/2007	122436528	5409	Compte Rendu de Consultation	CR ou fiche de consultation ou de visite
13/02/2007	121933390	5401	Compte Rendu d'Hospitalisation	CR hospitalier (séjour)
07/02/2007	121933390	5401	Compte-rendu opératoire	CR opératoire
07/02/2007	121933390		PMSI_DIAG	Diagnostics PMSI

cutée en 0.88s.

3/2015    Afficher

5/2015    Afficher

3/2015    Afficher

7/2016    Afficher

02/2016    Afficher

02/2016    Afficher

05/2016    Afficher

05/2016    Afficher

7/2016    Afficher

7/2015    Afficher

3/2015    Afficher

2/2016    Afficher

3/2016    Afficher

7/2016    Afficher

Recherche Textuelle

Type

Requête

Document

Unité(s)

Fonctionnelle(s)

Périphrases négatives

Enrichissement sémantique


Recherche Structurée

Masquer noeuds

- ▶ Actes (CCAM)
- ▶ Appareil-Organes
- ▶ Classification ATC
- ▶ Cytopathologies (Cytologie)
- ▶ DXCARE CONCLUSIONS
- ▶ DXCARE QUESTIONS
- ▶ Diagnostics (CIM)
- ▶ Pathologies Tumorales
- ▶ Pathologies non Tumorales
- ▶ Prélèvement-Techniques
- ▶ labo

Gestion des paniers




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

---

**Etude:** Etude neurinome (version démo) (#1996) +

Recherche rapide -

Mots-clés ... 🔍

Contexte: Ensemble de l'étude +

Requêtes enregistrées (1 élément) -

#	Titre		Rév.	Date	Actions
6236051	Requête initiale	bouzille	●	31/08/2016 12:43:41	📧

Recherche avancée -

☰ Critères généraux +

📄 Recherche Textuelle +

🔗 Recherche Structurée +

📁 Elements -

+  
 Substance  
Active

+  
 Spécialité

Gestion des paniers +

Liste des éléments -

**Inclure** les patients dont les documents contiennent les médicaments: ATORVASTATINE ACT 10MG CPR,ATORVASTATINE ACT 20MG CPR,ATORVASTATINE ACT 40MG CPR ✎ ✖

Type:  Inclure  Exclure

ATORVASTATINE CALCIQUE ANHYDRE 🔍

Inclure les spécialités n'étant plus commercialisées

Liste des spécialités associées: Tout sélectionner

Afficher 10 éléments Rechercher:

Nom	Action(s) ⚡
ATORVASTATINE ACT 10MG CPR	☑
ATORVASTATINE ACT 20MG CPR	☑
ATORVASTATINE ACT 40MG CPR	☑
ATORVASTATINE ALM 10MG CPR	☐
ATORVASTATINE ALM 20MG CPR	☐
ATORVASTATINE ALM 40MG CPR	☐
ATORVASTATINE ALM 80MG CPR	☐
ATORVASTATINE ARW 10MG CPR	☐
ATORVASTATINE ARW 20MG CPR	☐
ATORVASTATINE ARW 40MG CPR	☐

Prescription

Document    Autres documents    Journal d'accès

Informations

PRENOM NOM\_PATIENT2, âgée de 53 ans à la date du document  
 17/06/2015 - Prescription de médicaments  
 IPP, IEP  
 UF : 4328-HC SC SURV.CONTINUE PNEUMOLOGIE PL

Afficher 50 éléments    Rechercher :

Classe ATC	Libellés	Voie	Début	Fin	Posologie	Commentaire	Administration
ATORVASTATINE	TAHOR 40 MG, CPR	Voie orale	15/06/2015 18:00:00	17/06/2015 14:13:00			16/06/2015 18:00:00 1.00 cpr 15/06/2015 18:00:00 1.00 cpr
ALPRAZOLAM	ALPRAZOLAM 0.5 MG MYLAN, CPR SÉC	Voie orale	16/06/2015 07:00:00	16/06/2015 09:26:00		15/06/2015 17:23 : 2h avant [MS0019]	16/06/2015 07:00:00 1.00 cpr
PARACETAMOL	DOLIPRANE 500 MG, GÉLULE	Voie orale	15/06/2015 16:00:00	17/06/2015 14:13:00			15/06/2015 18:45:00 2.00 gel

Affichage de l'élément 1 à 3 sur 3 éléments    Précédent    1    Suivant

Fermer

Carte

Afficher le cercle d'éloignement

Centre :

Rennes

Rayon:

100 Km

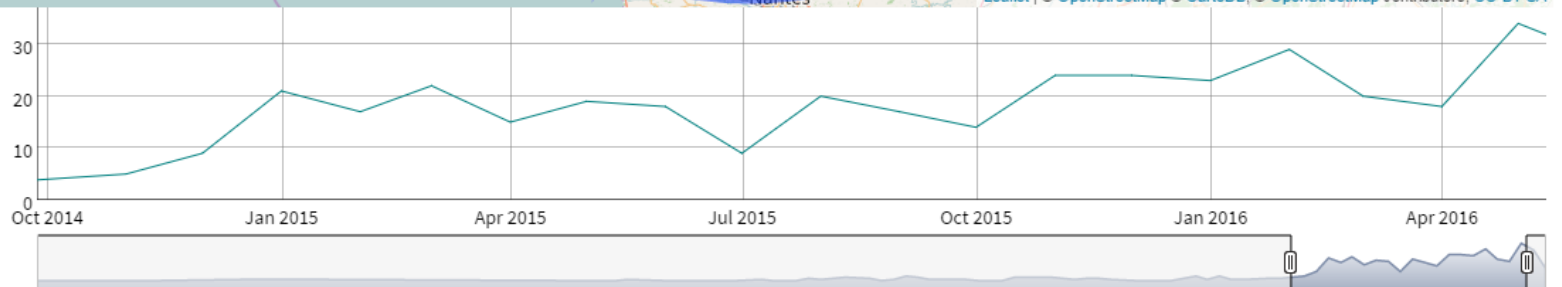
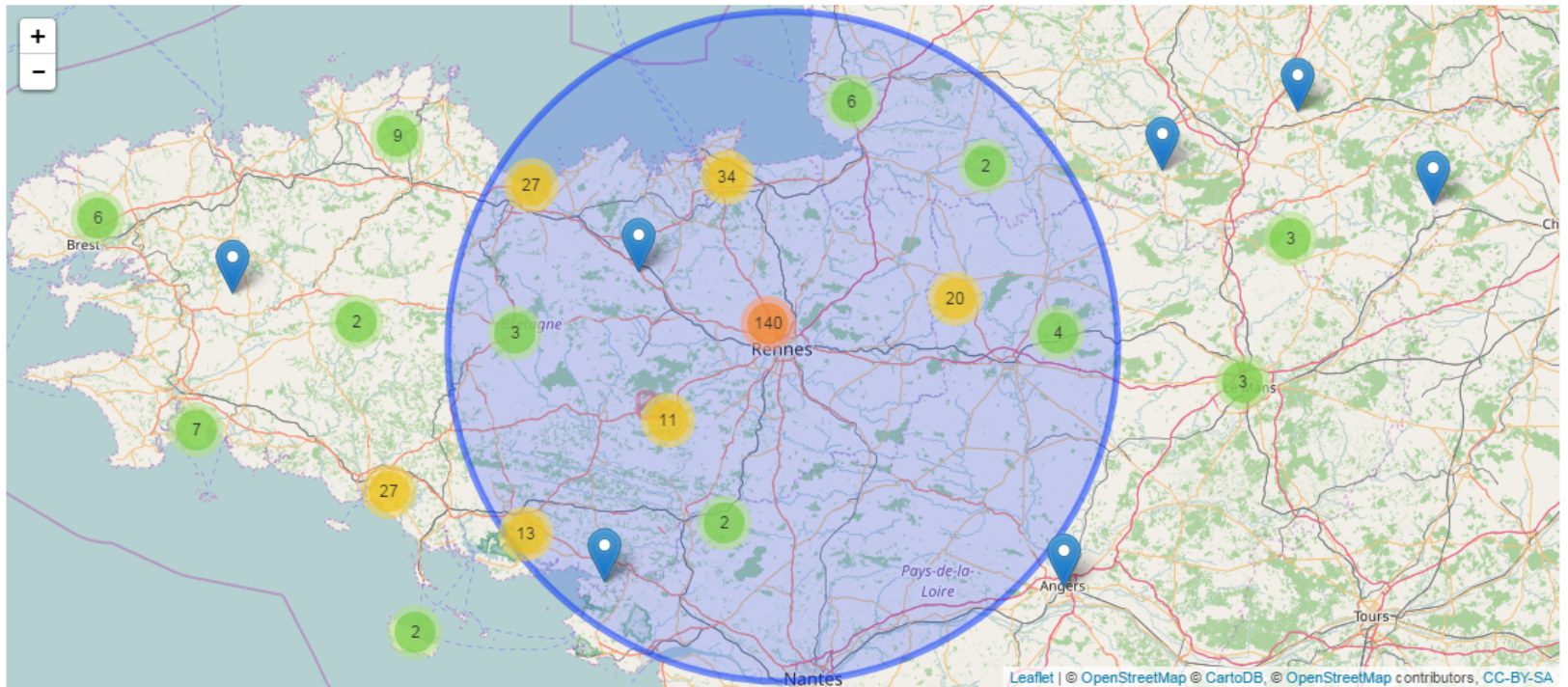
Timeline

Etudes Population Statistiques

Etude

# 259 Patients

sur les 498 de l'étude, pour la période du 26/09/2014 au 11/05/2016 dans un périmètre de 100Km autour de Rennes.



Aller à la date du : 1/1/2010 au : 14/6/2014

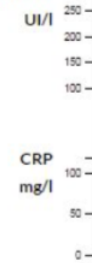
6 Une semaine Un mois Un an

2 April July October 2011 April July October 2012

DONNEES BIOLOGIQUES

PAL Phosphat. alcalines UI/l

CRP mg/l



DOCUMENTS

Search bar for documents



- Afficher la dose administrée
- Afficher les documents

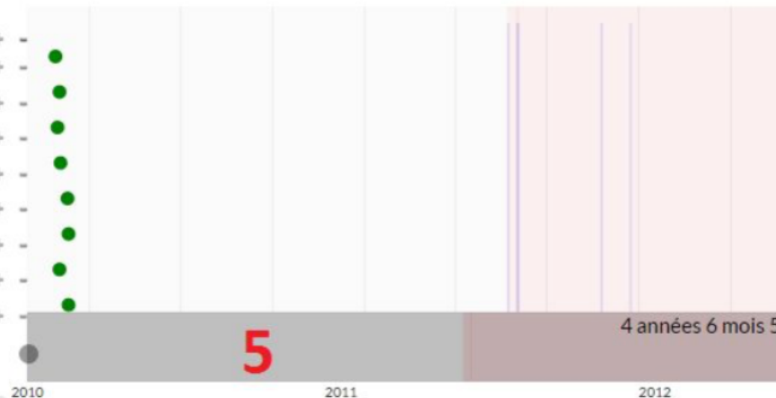
MEDICAMENTS

- LOVENOX 4 000 IU ANTIXA/0.4 ML, SOL INJ, SRG
- GLUCOSE 5%, POCHE 50 ML VIAFLO
- SODIUM CHLORURE 0.9%, SOL PR PERF, POCHE 50 ML VIAFLO
- SODIUM CHLORURE 0.9%, SOL PR PERF, POCHE 100 ML VIAFLO
- POLYIONIQUE 1AG5, POCHE 500 ML VIAFLO
- DEXAMETHASONE 4 MG/1 ML MYLAN, SOL INJ, AMP



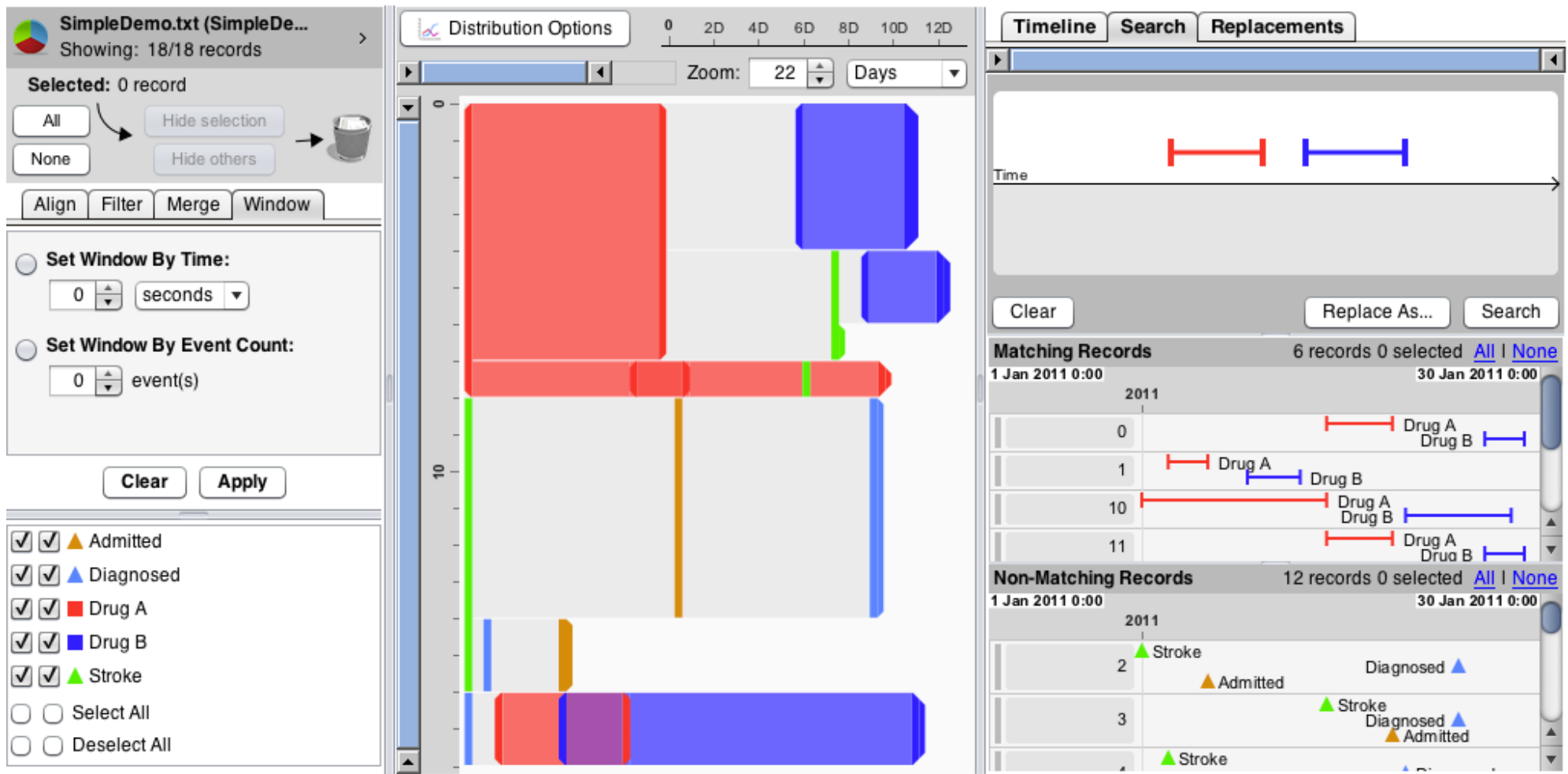
ACTES MEDICAUX

- Scanographie du crâne et de son contenu, sans injection de produit de contraste
- Échographie-doppler transthoracique du coeur et des vaisseaux intrathoraciques
- Échographie-doppler des artères cervico-céphaliques extracrâniennes, sans mesure de l'épaisseur de l'intima-média
- Échographie-doppler des veines des membres inférieurs et des veines iliaques, sans marquage cutané
- Transfusion de produit sanguin labile non érythrocytaire
- Radiographie du bras
- emnographie [IRM] unilatérale ou bilatérale de segment du membre supérieur, sans injection de produit de contraste
- Tumorectomie ostéoarticulaire d'un membre, par voie transcutanée avec guidage scanographique

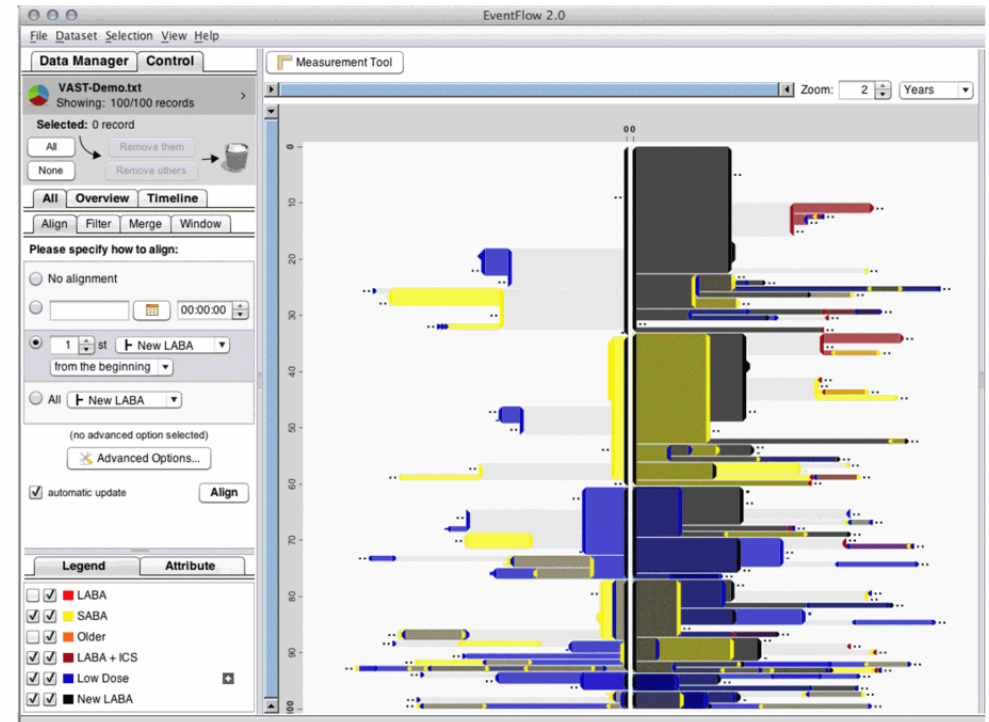
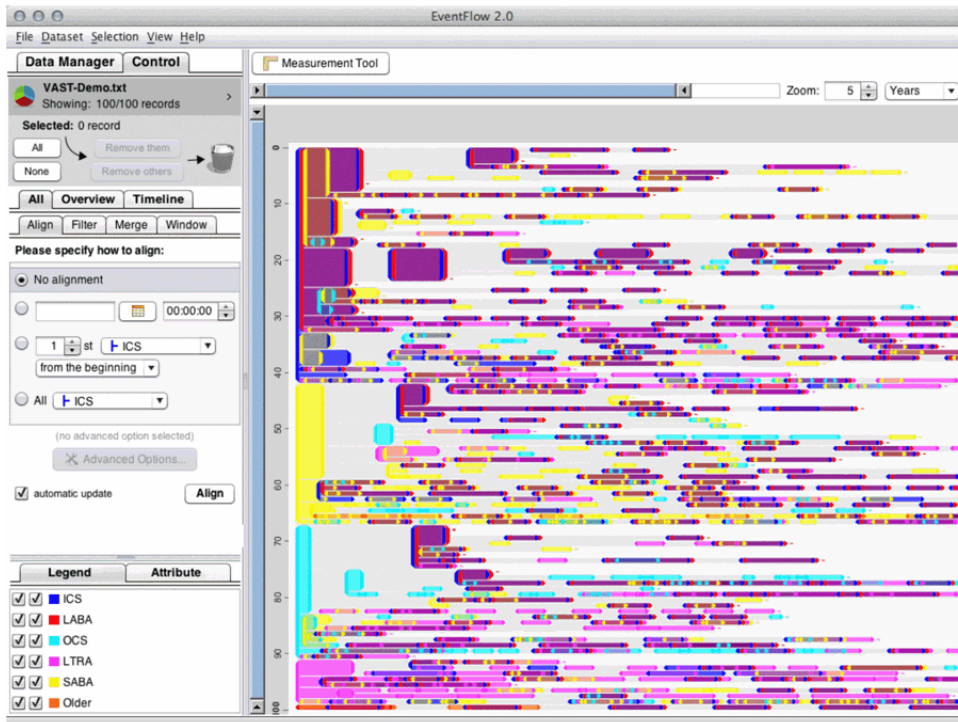
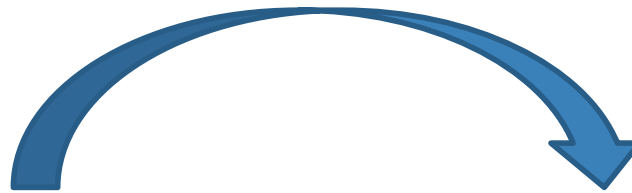


# Exemple d'approche populationnelle : Eventflow

Simplification of temporal event sequences (Megan Monroe, Rongjian Lan, Catherine Plaisant, Ben Shneiderman)



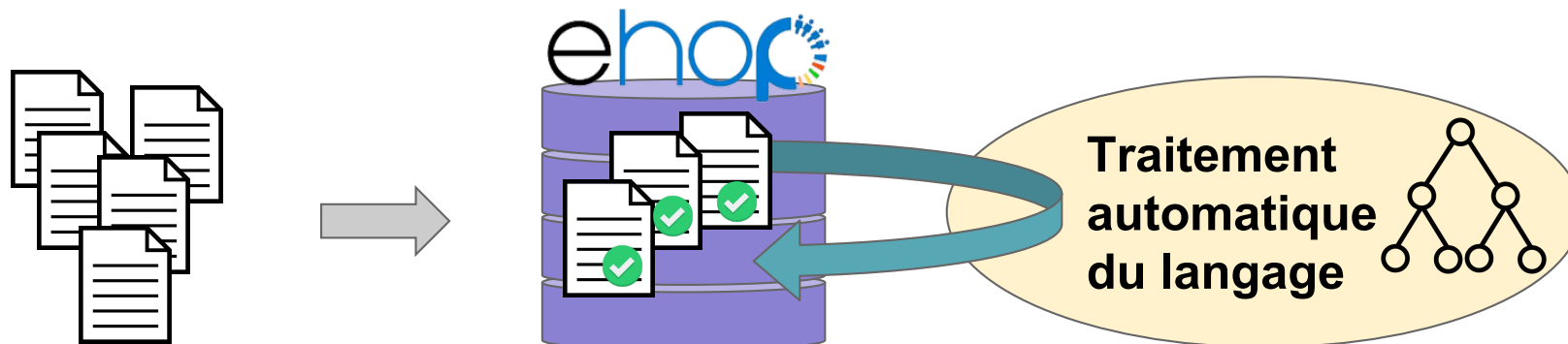
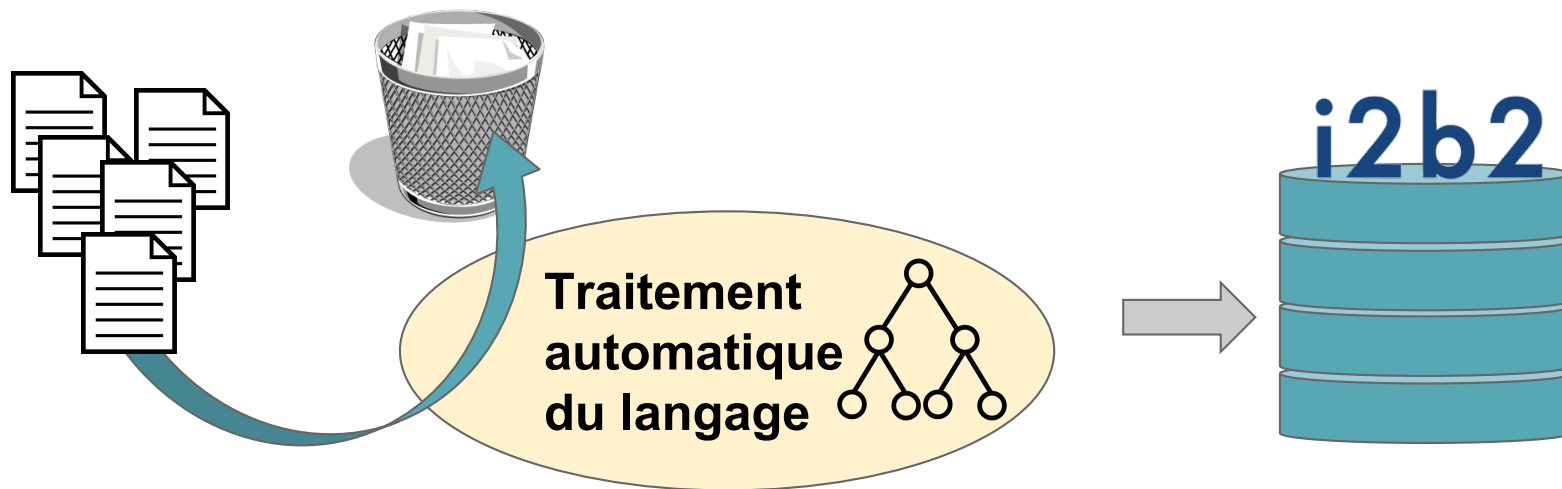
# Reduction de dimension



Shneiderman, B., Plaisant, C.  
[Sharpening Analytic Focus to Cope with Big Data Volume and Variety: Ten strategies for data focusing with temporal event sequences \(2015\)](#)

# Entrepôts de données biomédicales

Données non structurées : traitement automatique du langage



# Plan de l'exposé

- 1 Données massives en santé
- 2 Champs d'utilisations
- 3 Les centres de données clinique et eHOP
- 4 Les réseaux de données de santé
- 5 Les projets de fouilles sur données massive en santé
- 6 Se former et travailler au CDC

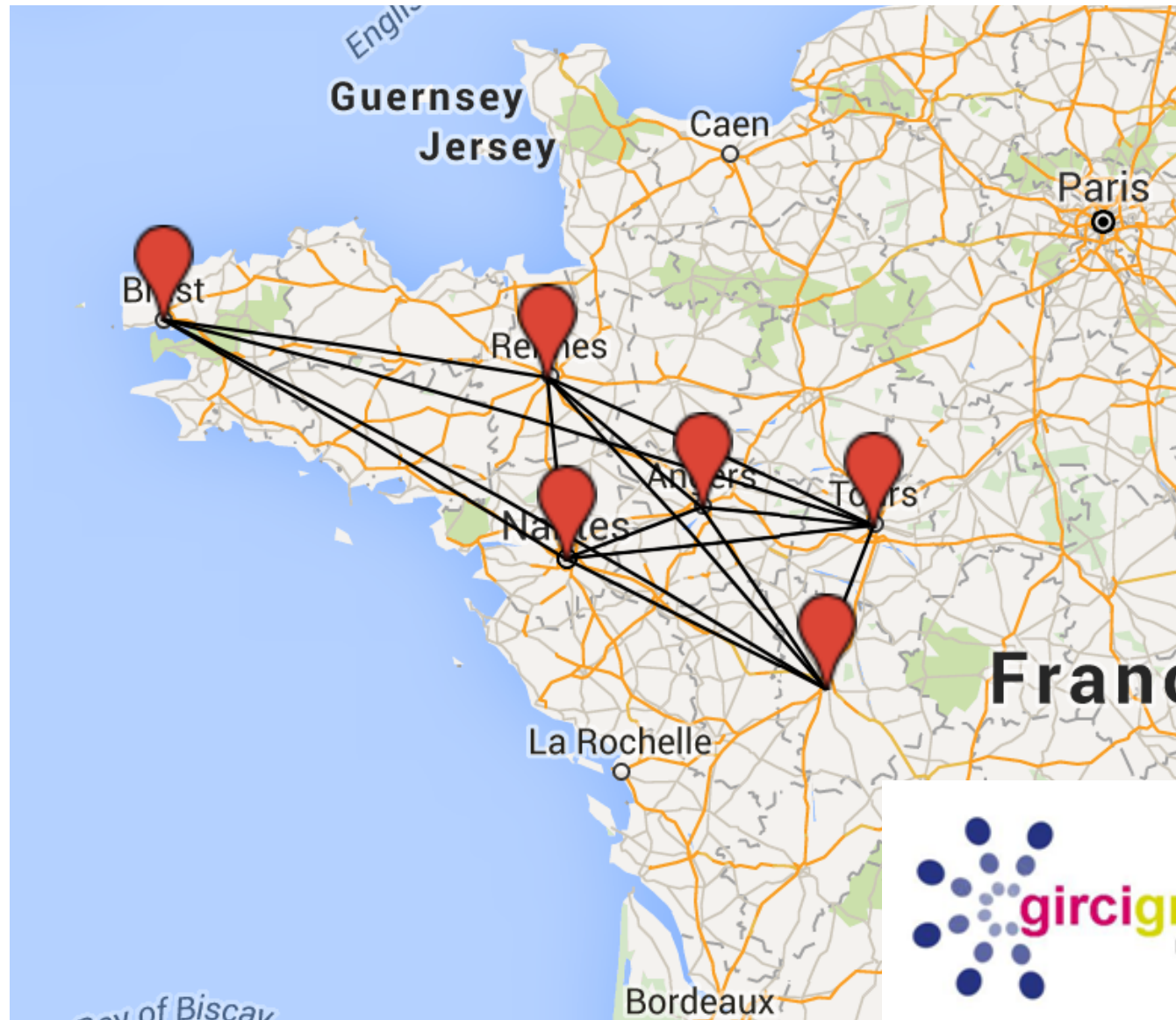


# Réseaux de données de santé Clinical Data Research Networks



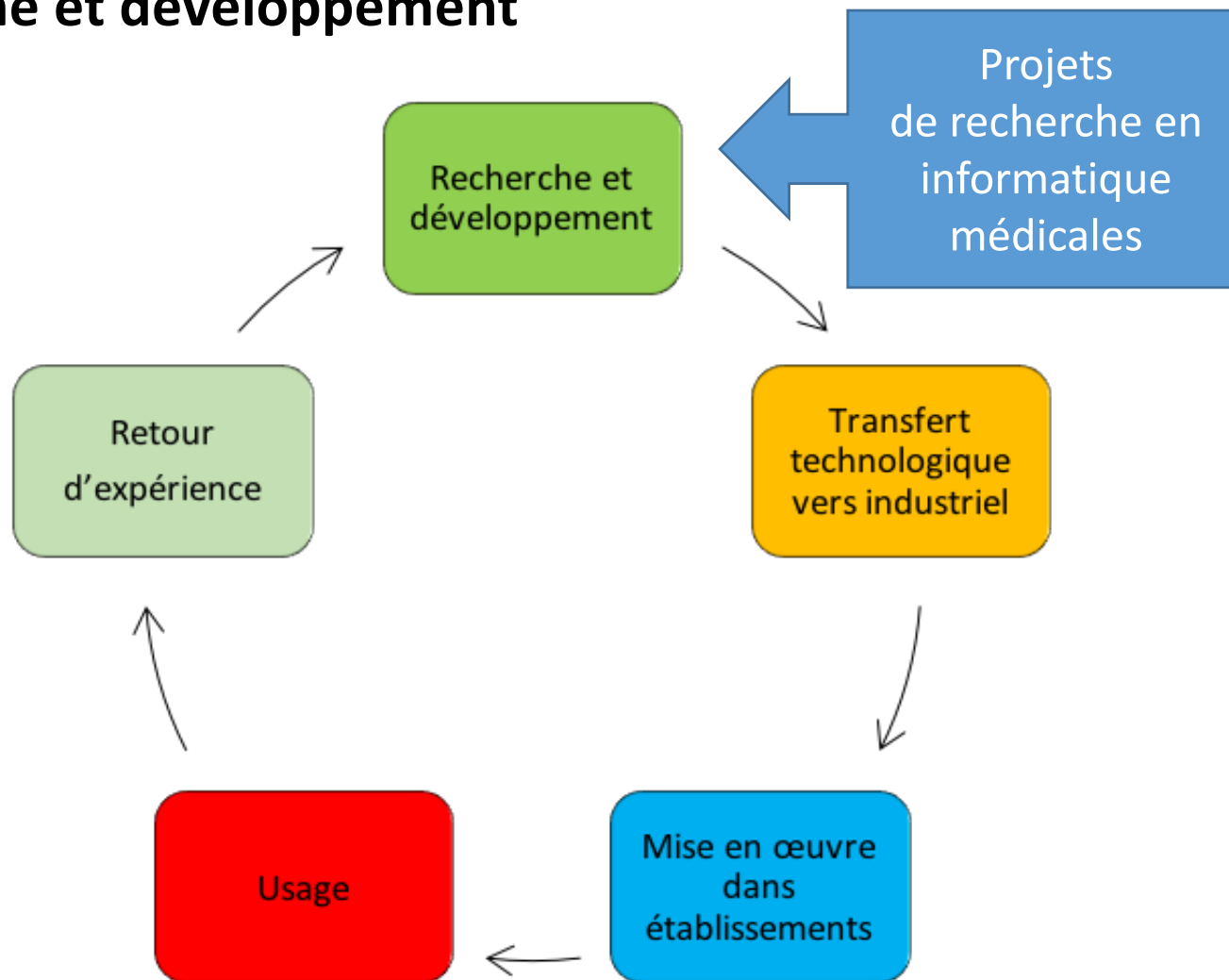


# Réseau inter-régional des CDC



# Ri-CDC : Objectifs

- **Faire évoluer les outils en établissant une démarche vertueuse de recherche et développement**





# pcornet

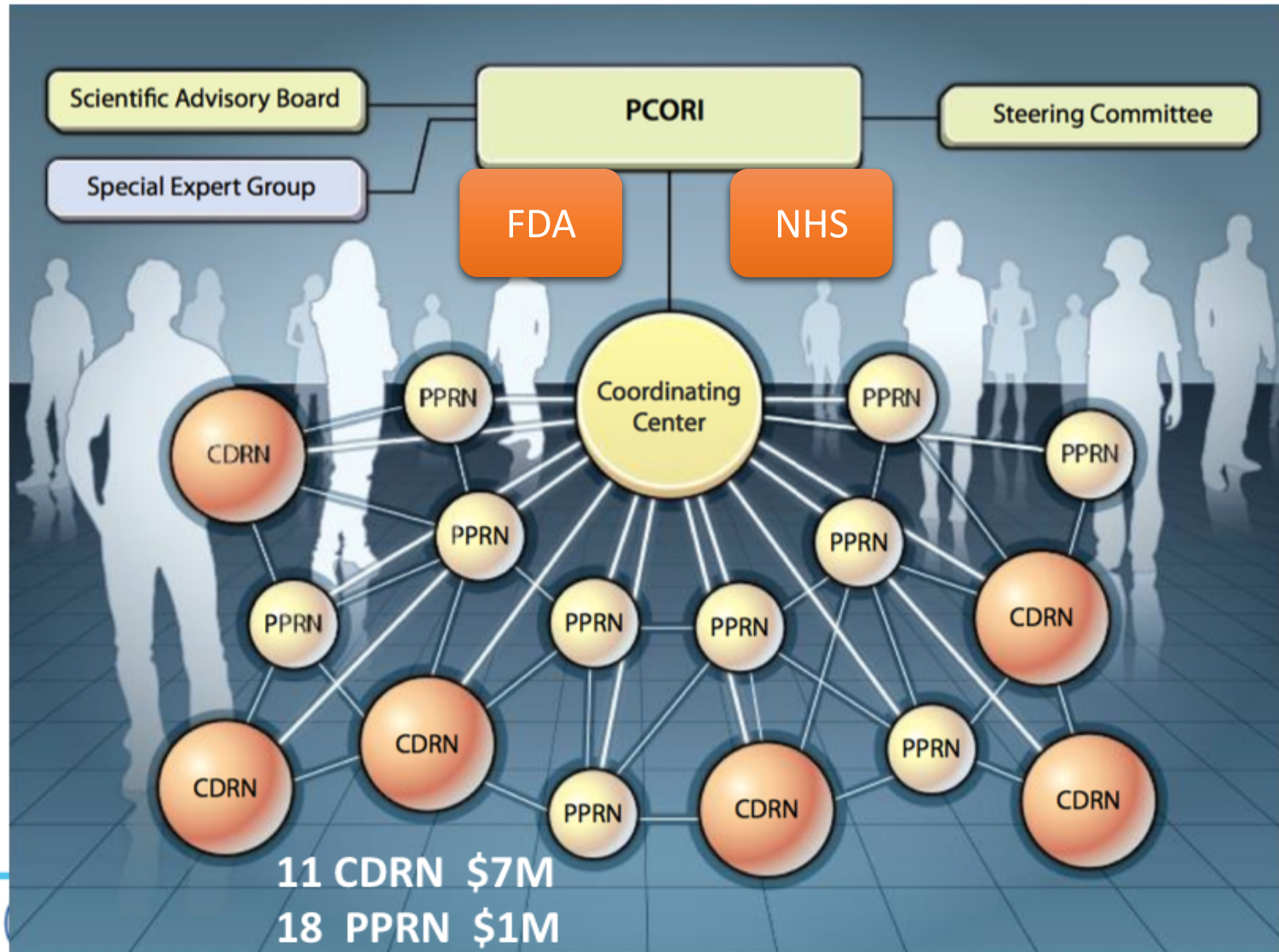
The National Patient-Centered Clinical Research Network

**PCORnet**

**\$95 million**

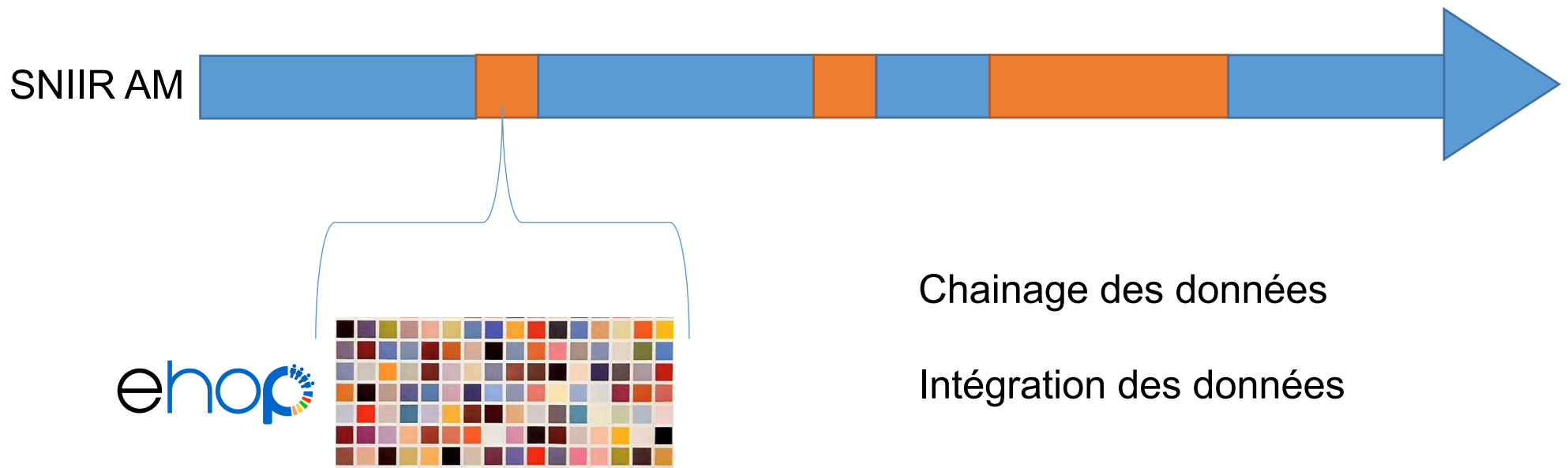
**Start date Jan 2014**

**18 months**



# Dynamique de recherche

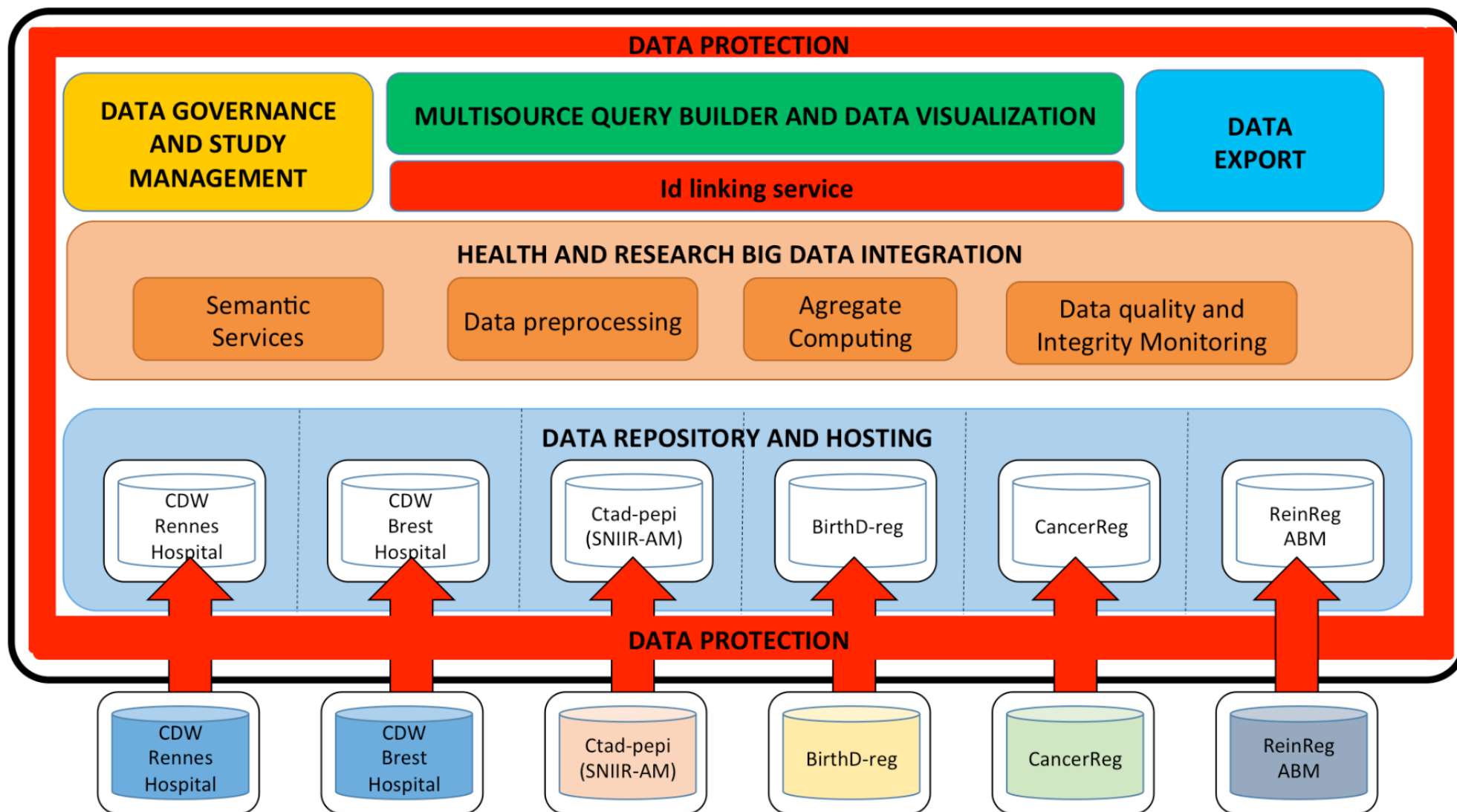
- **Projet PEPS** : Plateforme d'Etude des Produits de Santé
- Utilisation du SNIIR-AM pour réaliser des études de pharmacovigilance
- SNIIR AM : Données de remboursement + PMSI (données hospitalières)
- Caractérisation de la trajectoire de santé des patients





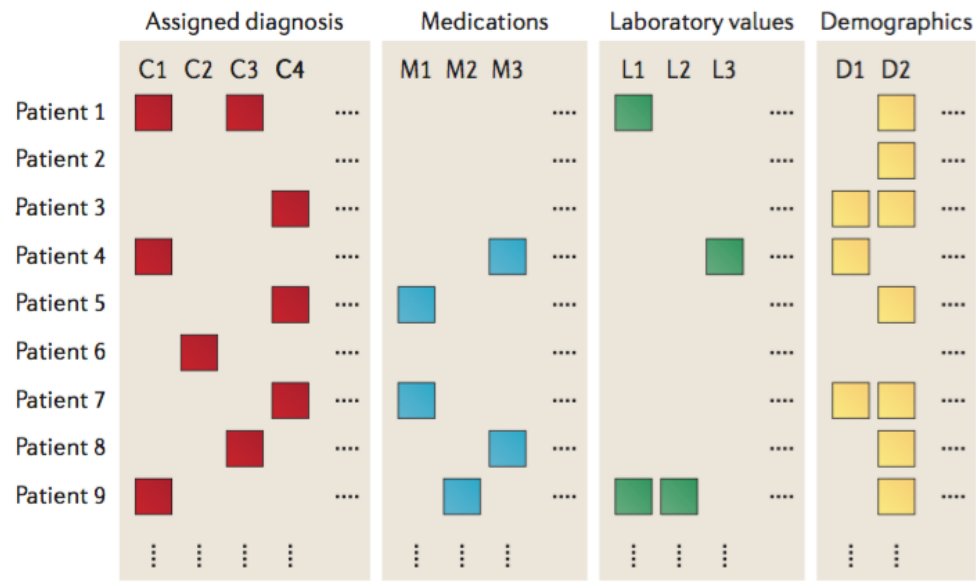
- **Projet INSHARE (AO ANR) (3 ans):**

- Plateforme Cloud d'Intégration et partage des données en santé pour la recherche
- Protection tatouage des données
  - Intégration des données hospitalières, SNIIRAM, et des registres épidémiologiques



# Plan de l'exposé

- 1 Données massives en santé
- 2 Champs d'utilisations
- 3 Les centres de données clinique et eHOP
- 4 Les réseaux de données de santé
- 5 La fouille sur données massive en santé
- 6 Se former et travailler au CDC

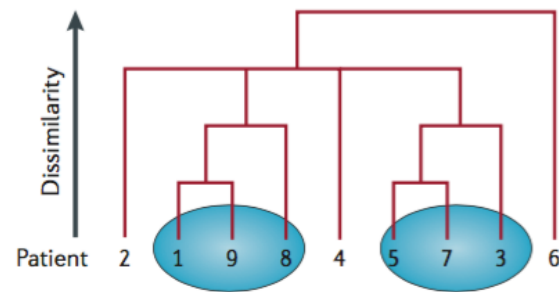


**a Comorbidity**

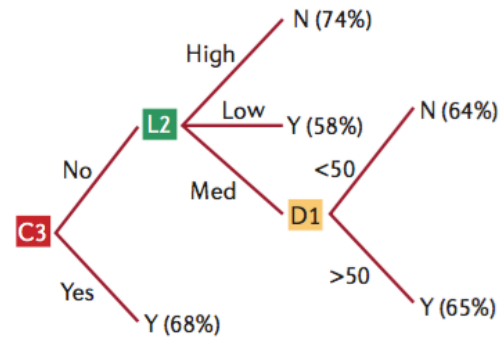
	C4	¬C4	
C2	10	40	50
¬C2	90	860	950
	100	900	1,000

Relative risk = 2

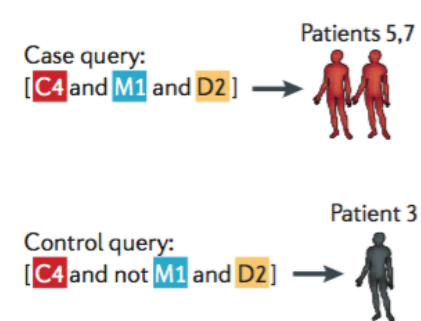
**c Patient clustering**



**b Machine learning**



**d Cohort querying**





# Exemple

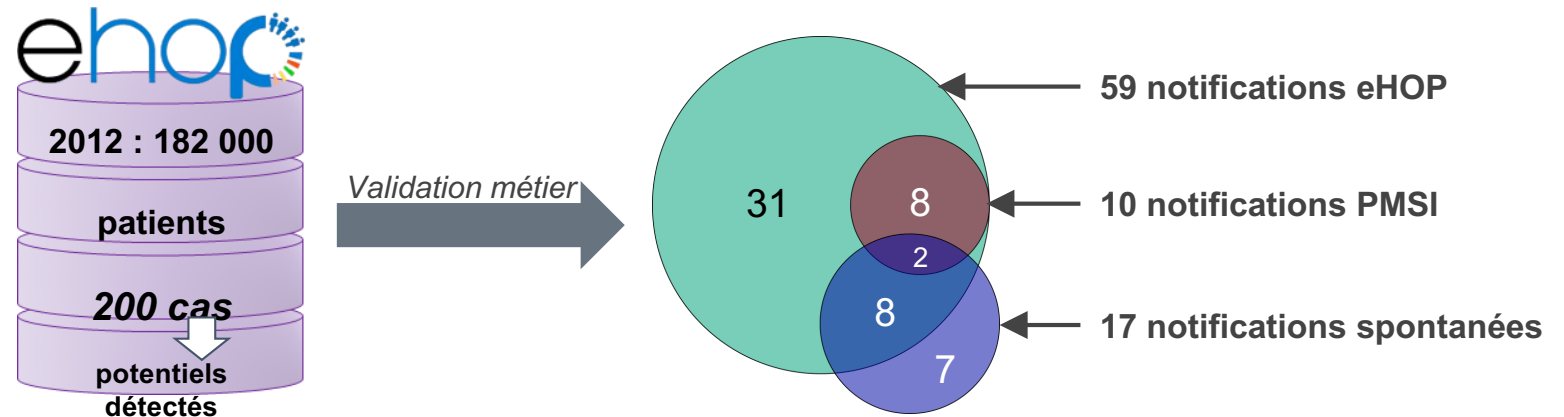
## Pharmacovigilance

### Contexte : Sous notification des effets indésirables médicamenteux

*Hypothèse : détection des chocs anaphylactiques médicamenteux à partir d'un entrepôt de données biomédicales*

#### Méthodes :

- Recherche d'information et traitement automatique du langage



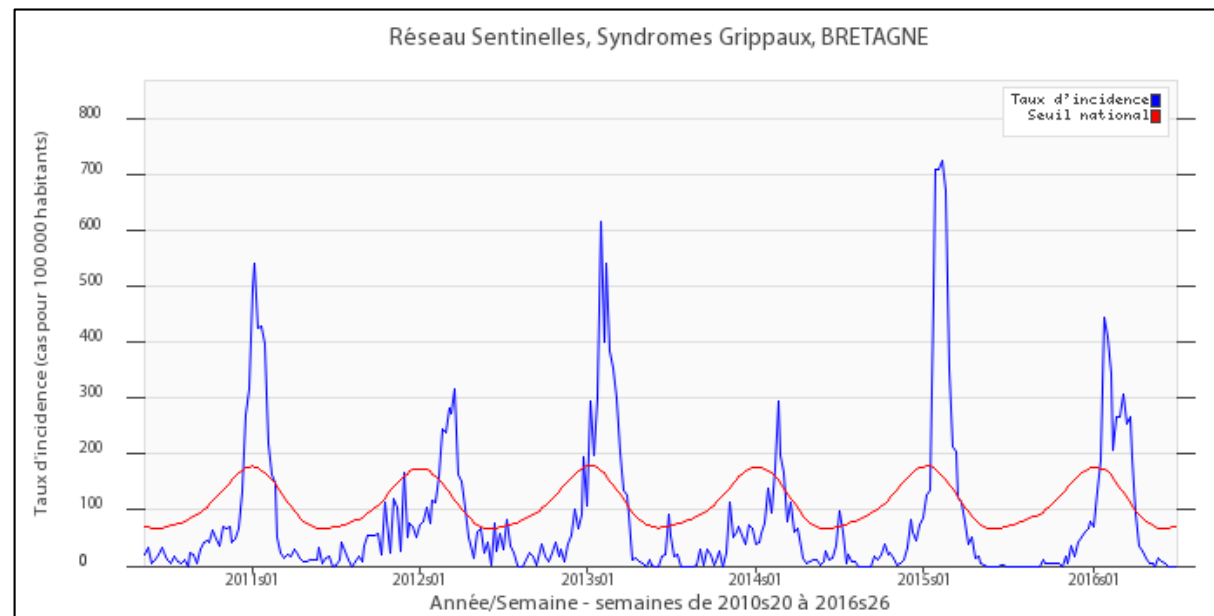
#### Retombées et valorisation :

- Mise en oeuvre en routine au CRPV de Rennes
- *Bouzillé G, Osmont MN et al. Drug safety and big clinical data: detection of drug-induced anaphylactic shocks (en cours)*

# Surveillance syndromique - exemple de la grippe

**Contexte : maladie à fort impact socio-économique**

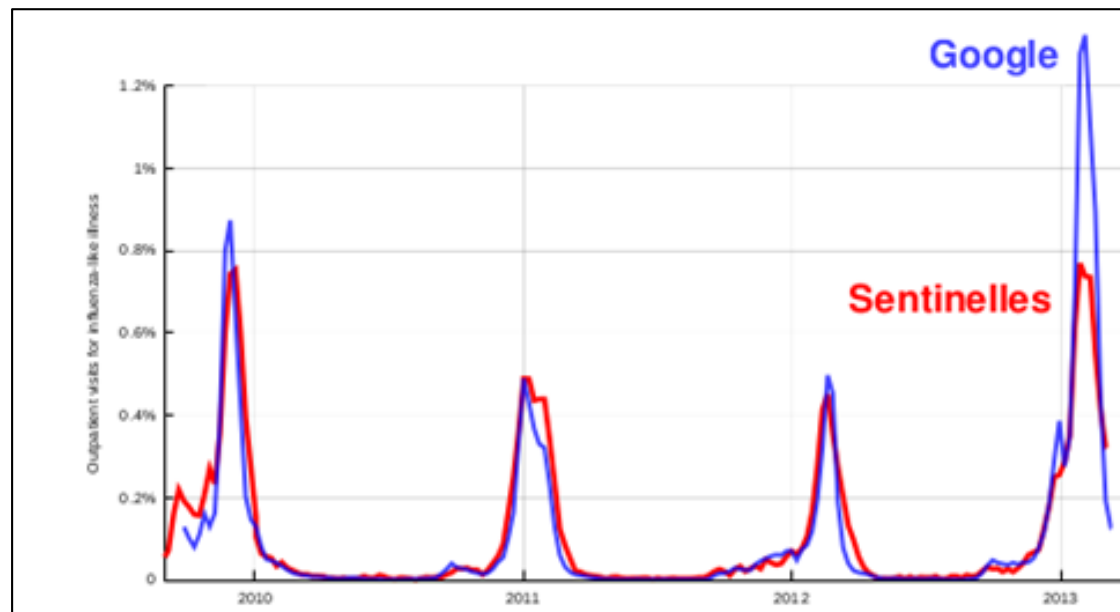
Surveillance épidémique par le réseau Sentinelles



- ▷ Limite : production des indicateurs avec un délai de 1 à 3 semaines

# Surveillance syndromique - exemple de la grippe

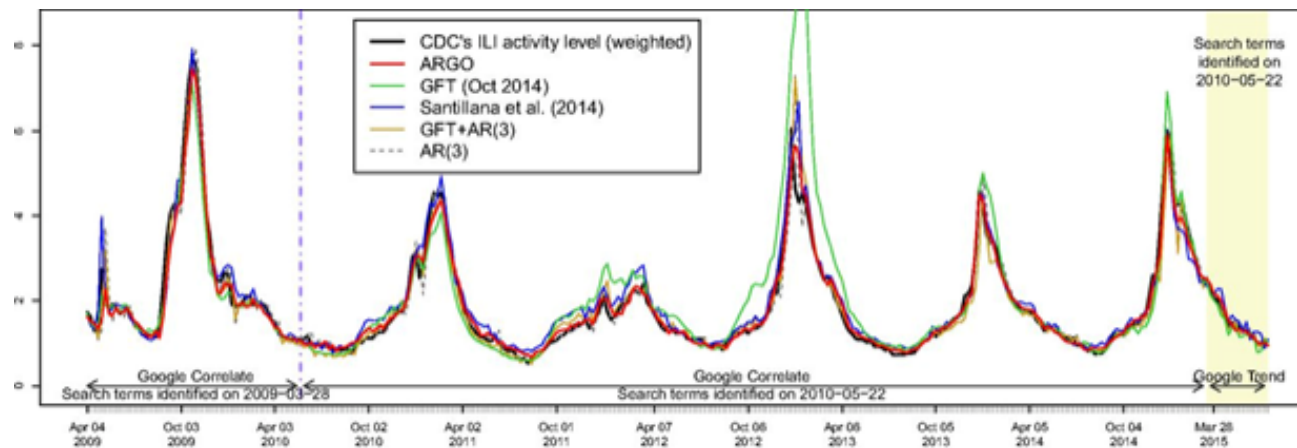
Alternatives : surveillance basée sur les données du web



▷ Limite : indicateurs non fiables à plusieurs reprises

# Surveillance syndromique - exemple de la grippe

Alternatives : surveillance associant données Sentinelles et données du données du web



# Surveillance syndromique - exemple de la grippe

## Hypothèse : associer données de Sentinelles et Big Data Hospitalier

### Méthodes :

#### 1. Extraction d'information depuis eHOP (incidence hebdomadaires) :

- ▷ Utilisation de requêtes par mots-clés :
  - Documents comportant un mot commençant par « gripp »
  - Documents mentionnant des symptômes grippaux
  - Documents de patients des urgences venant pour grippe et repartant chez eux
  
- ▷ Utilisation des données structurées
  - Codes diagnostics se rapportant à la grippe
  - Examens biologiques de recherche de virus grippal par PCR
  
- ▷ Découpage par services, classes d'âge, type de documents.



# Surveillance syndromique - exemple de la grippe

## Hypothèse : associer données de Sentinelles et Big Data Hospitalier

### Méthodes :

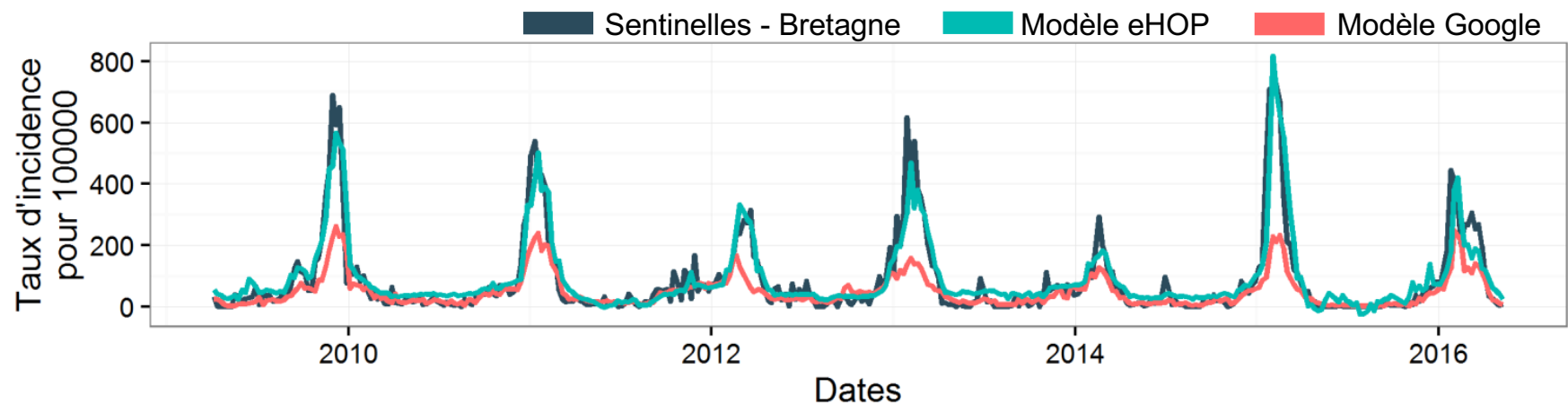
#### 2. Modélisation statistique :

$$y_t = \mu_y + \sum_{j=1}^N \alpha_j y_{t-j} + \sum_{i=1}^K \beta_i X_{i,t} + \epsilon_t$$

- ▷ Une composante basée sur l'historique de Sentinelles
- ▷ Une composante basée sur les données obtenues depuis eHOP
- ▷ Une étape de sélection de variables par régression type lasso
- ▷ Un modèle construit pour chaque semaine à prédire

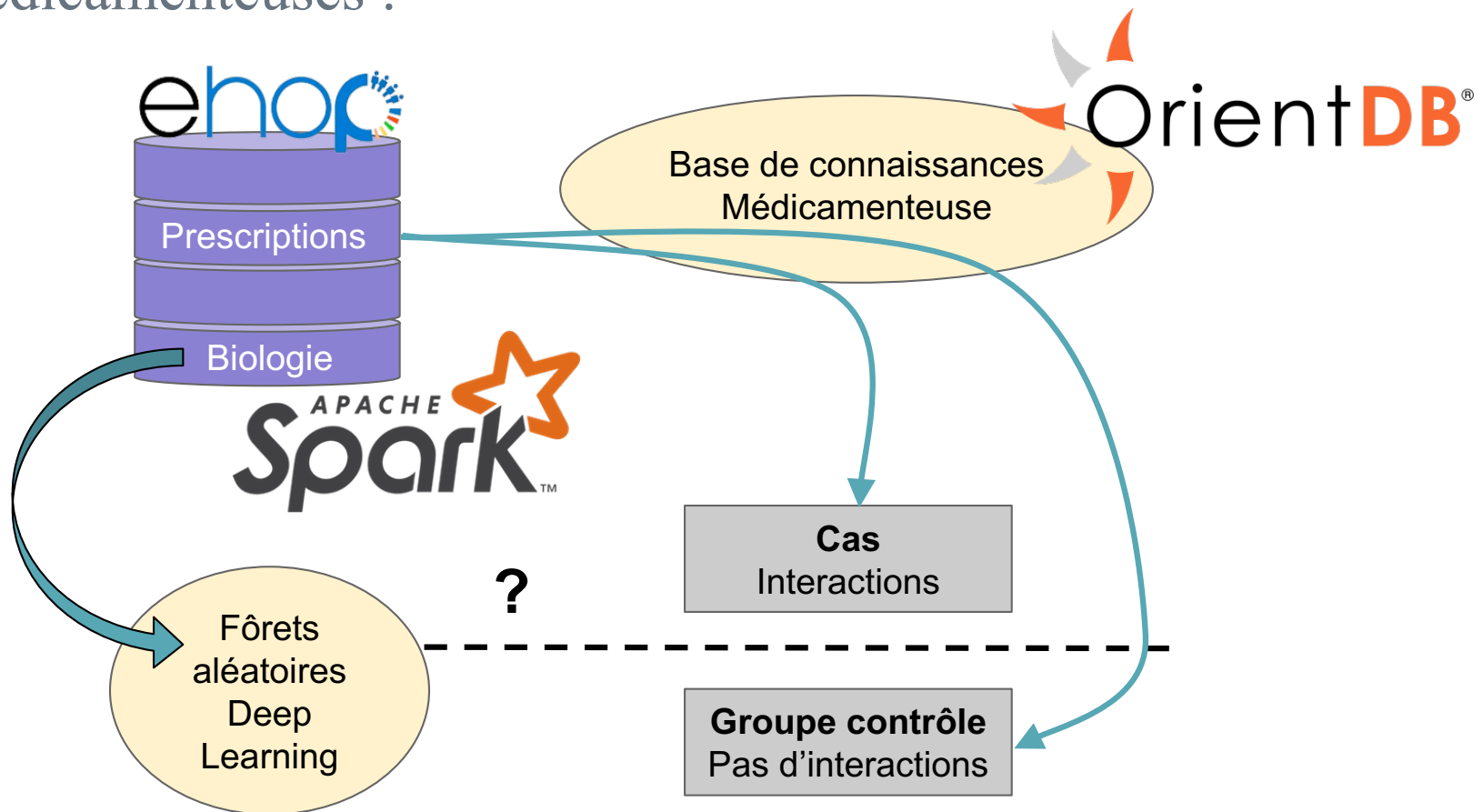
# Surveillance syndromique - exemple de la grippe

## Résultats :



	Coefficient de corrélation		Erreur quadratique moyenne	
	Modèle eHOP	Modèle Google	Modèle eHOP	Modèle Google
Prédictions nationales	0.95	0.97	2563	2931
Prédictions régionales	0.92	0.88	2881	9797

# Détection des effets indésirables dues aux interactions médicamenteuses :





# Que faut il retenir ?

- ✓ **Le Big data** en santé est constitué de données :
  - Hétérogènes
  - Multi-domaines
  - multi-échelles
  - Données sensibles
  
- ✓ Les **entrepôts** sont des **concentrateurs** de données permettant une **exploitation facilitée des données** mais qui nécessitent :
  - Une technicité : nouveaux métiers
  - Des organisations tiers de confiance permettant leur exploitations respectant la protection des données
  
- ✓ Ces technologies peuvent être mises en œuvre dans de **nombreux cas d'utilisation** en recherche (mais aussi au delà)
  - Ils sont **complémentaires** des approches classiques de recueil pro-actif d'information
  
- ✓ La tendance est la constitution de **réseaux de données** pour une exploitation **multicentrique**
  
- ✓ Nous en sommes au **début**