



Données massive en santé, entrepôts de données de santé et traitements longitudinaux

Pr Marc CUGGIA
PUPH Informatique médicale

Equipe projet Données Massives en Santé Unité Inserm
1099 Laboratoire Traitement du Signal et de l'Image

Réseau des Centres de Données Cliniques
CHU RENNES

Quels usages ?

▷ L'individu

- Médecine personnalisée : Médecine 4P

▷ La recherche

- Épidémiologique : détection et alimentation de cohortes ou de registres
- Clinique : études de faisabilité, recherche de patients éligibles

▷ Les collectivités

- Gestion des risques, qualité des soins
- Vigilances, veille sanitaire

▷ L'organisation raisonnée des soins

- Pilotage des activités
- Analyse des trajectoire de santé

▷ L'enseignement

- Ex : la simulation et les patients virtuels

Dématérialisation des données de santé

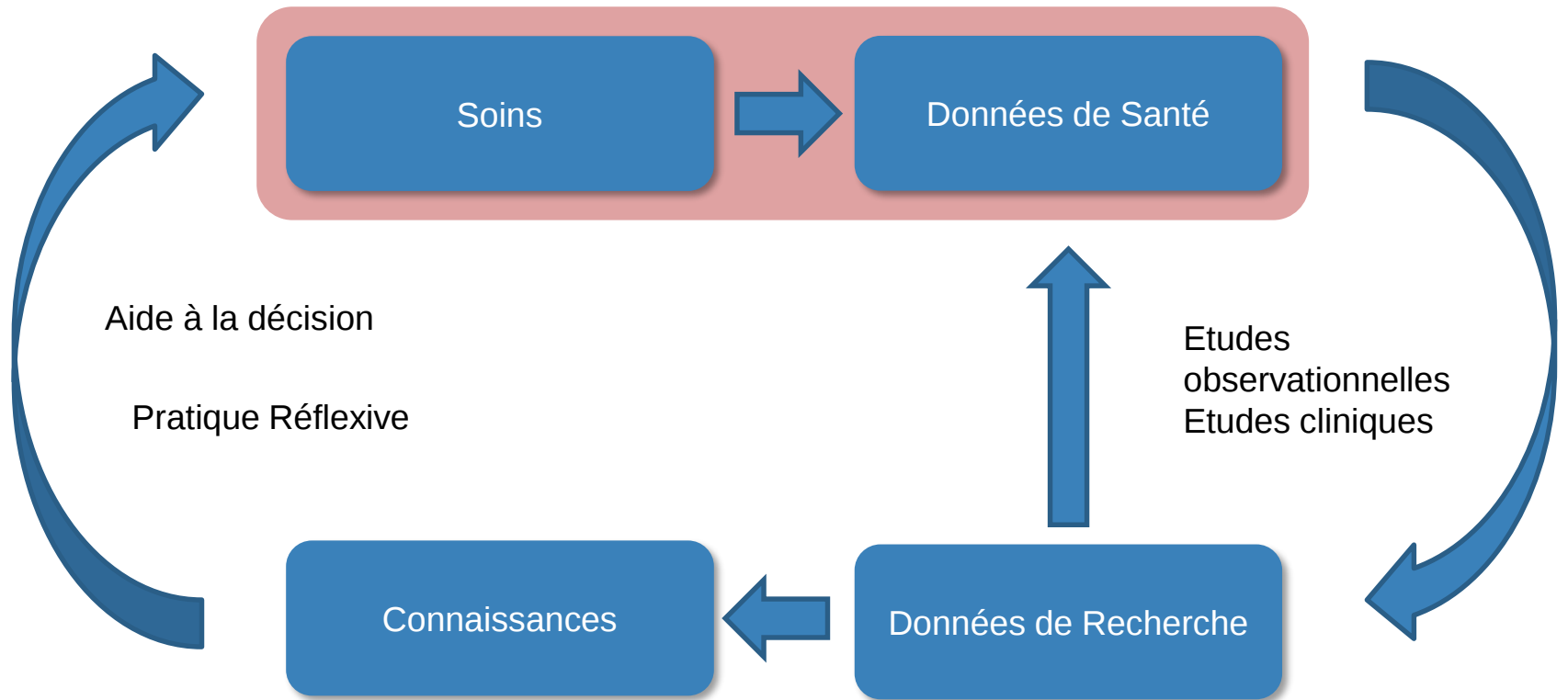
Le soin



La recherche



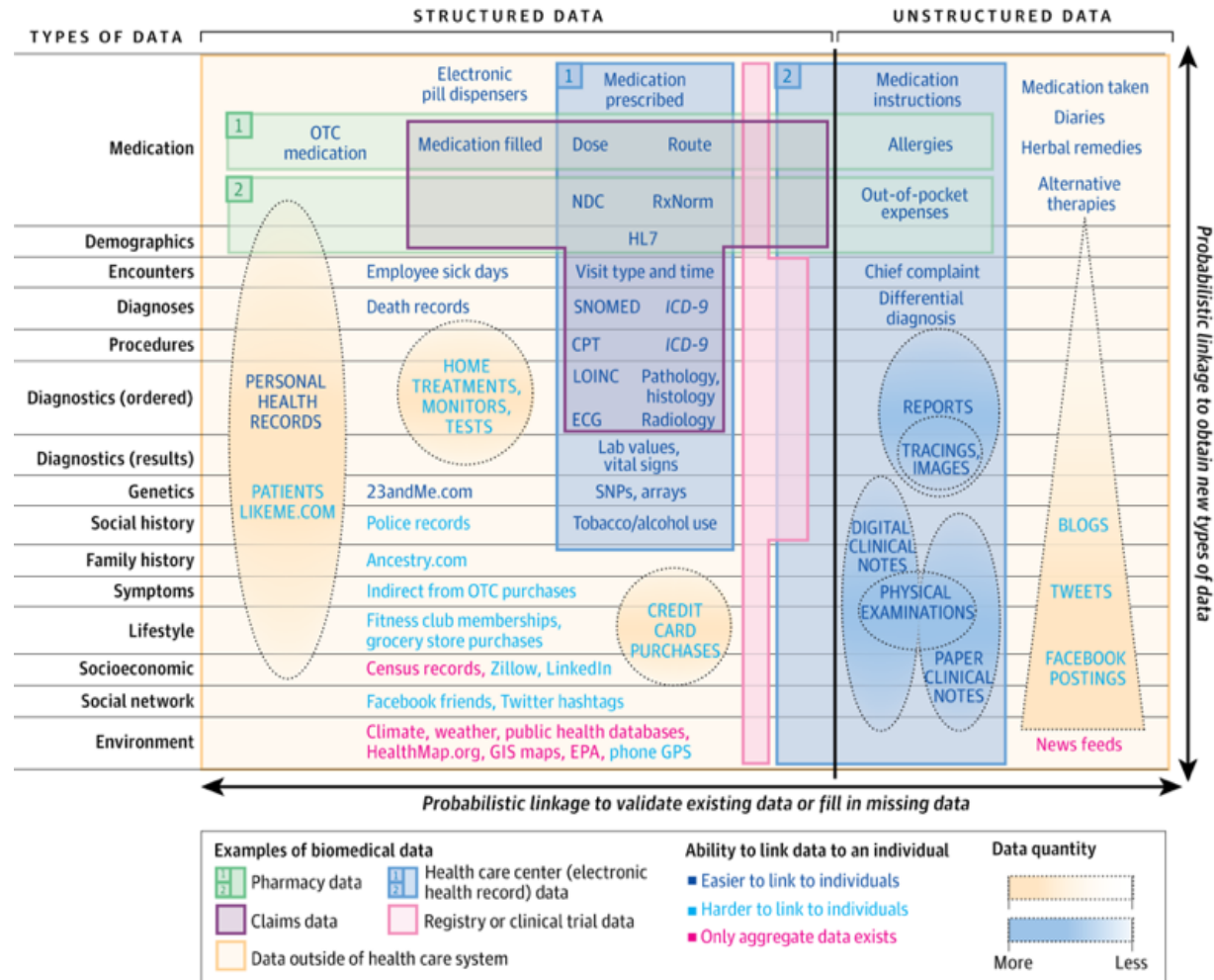
Systeme de sante apprenant



inspiré de Ethier J-F, Friedman C (2015), Westfall (2007)



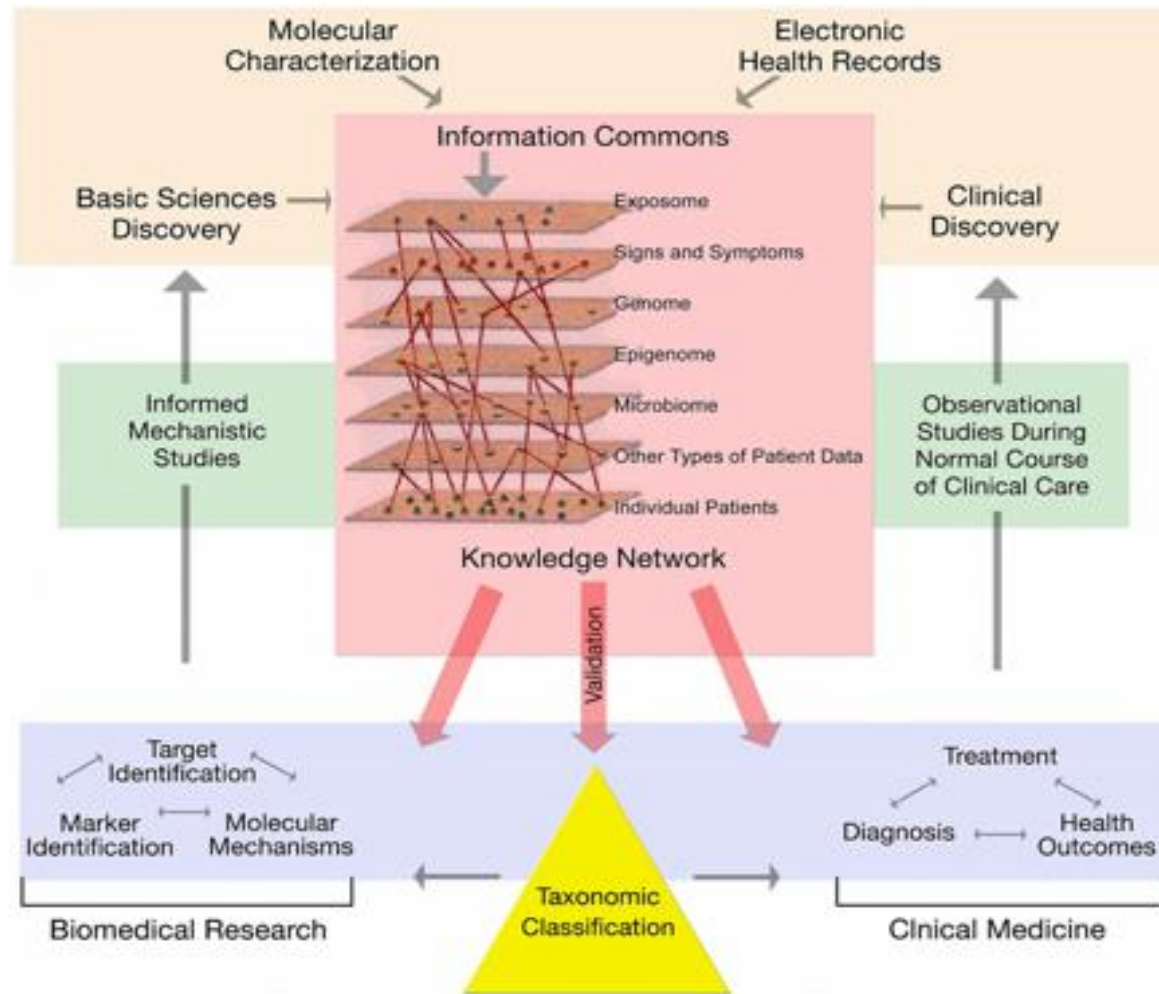
Interopérabilité



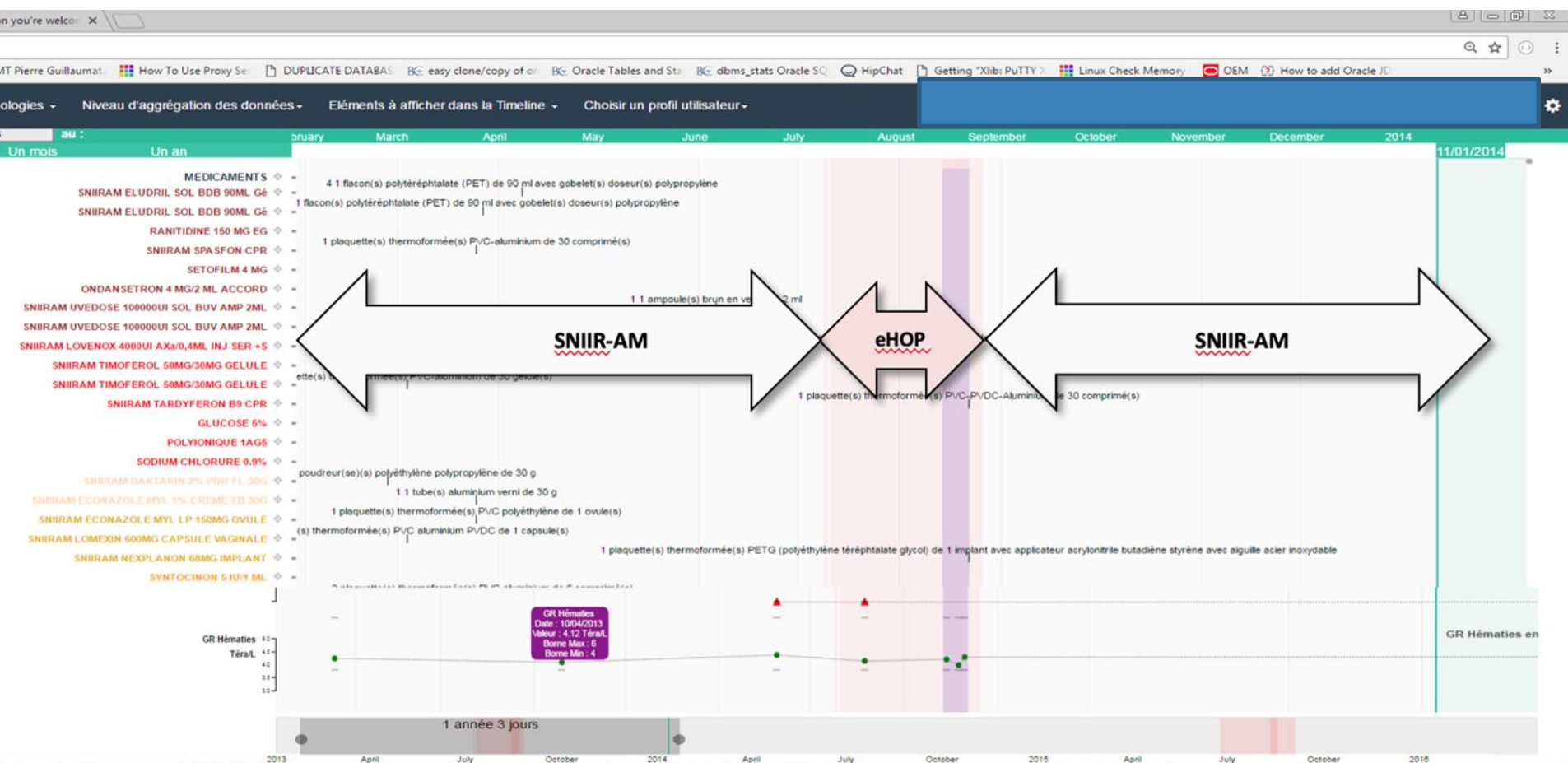
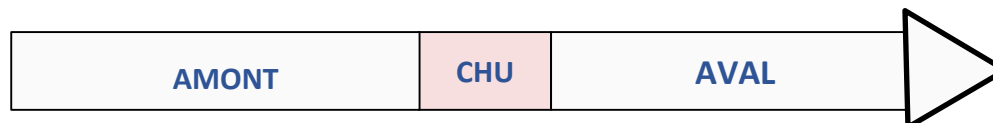
G. M. Weber, K. D. Mandl, and I. S. Kohane, "Finding the missing link for big biomedical data," *JAMA*, vol. 311, no. 24, pp. 2479–2480, Jun. 2014.



Intégration



Comprendre les trajectoires patients : Chainage & Integration Données Cliniques - SNIIRAM

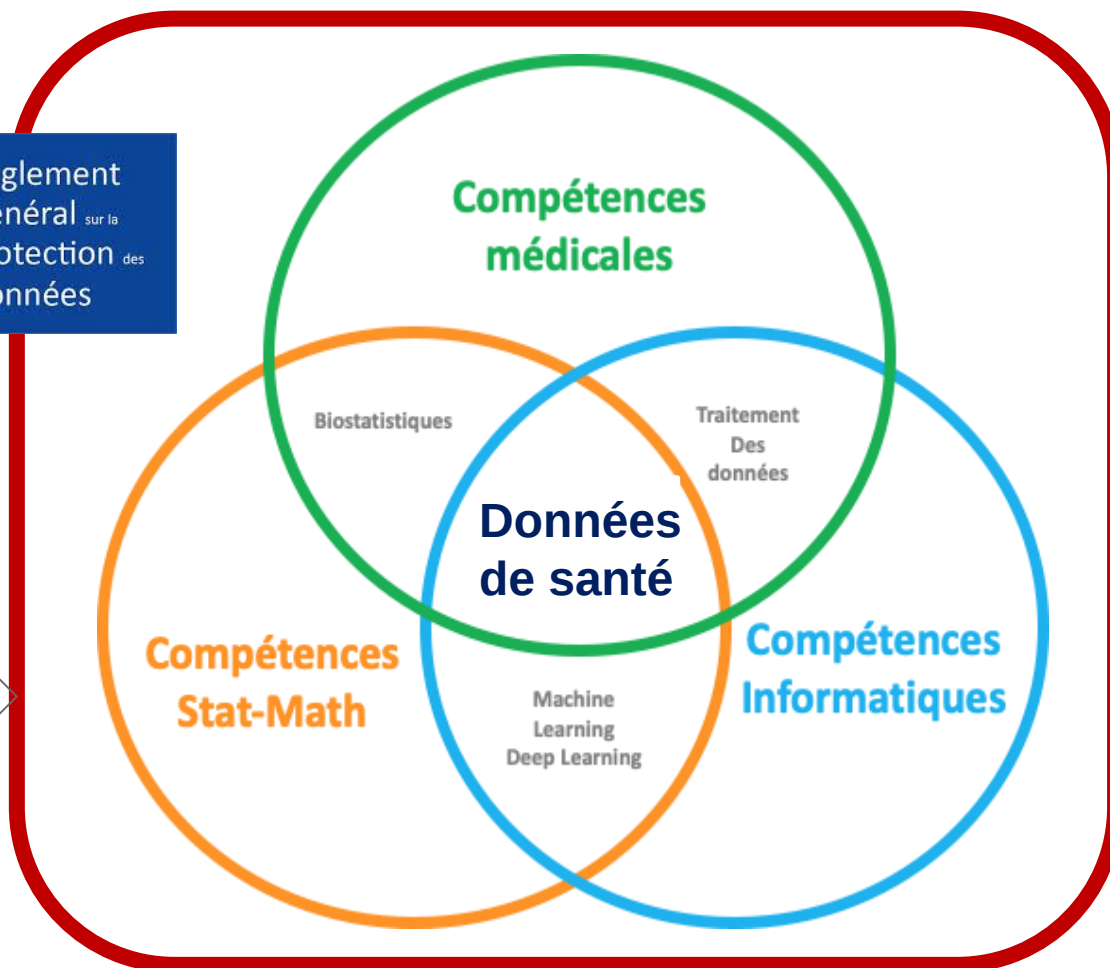




Exploitation sécurisée



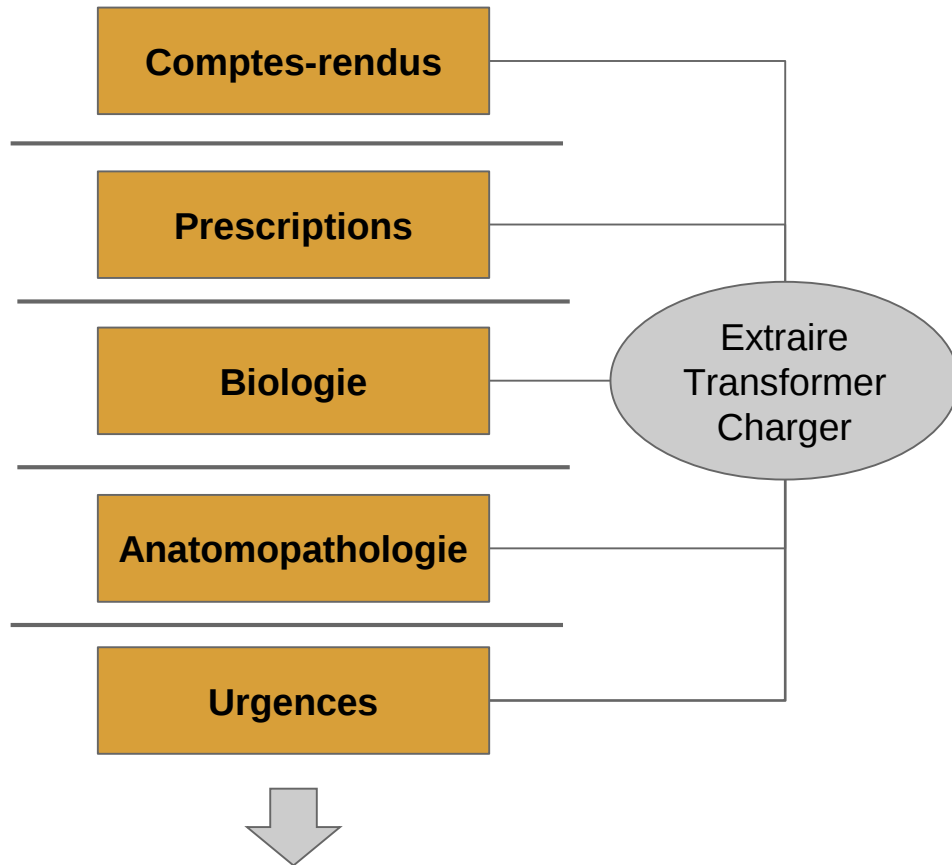
DEIDENTIFICATION
CRYPTO-TATOUAGE
AVATARDISATION



Ohno-Machado, L. (2018). Data science and artificial intelligence to improve clinical practice and research. *Journal of the American Medical Informatics Association*, 25(10), 1273-1273.

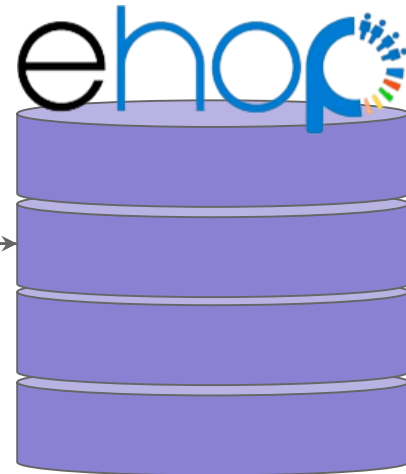
Les entrepôts de données biomédicales

Des données produites pour le soin



Des données cloisonnées

Une base de données
orientée pour la
réutilisation des données



Exemple du CHU de Rennes :

- 1,7 millions de patients
- 52 millions de documents
- 450 millions d'éléments de données

Accompagner et encadrer l'exploitation des données en santé

Les Centres de Données Cliniques

Plateau forme hospitalo-Universitaire “guichet unique” :

- ▷ Propose une **expertise** méthodologique et des **services** pour exploiter les données de santé
 - ▷ S'appuie sur une infrastructure technologique : EDS + Outils d'exploitation
 - ▷ Adossée à une structure de recherche : INSERM LTSI
- ▷ Rassemble des compétences **multidisciplinaires**
 - Informatique médicale, Biostatistiques, TAL
 - Juridiques, Réglementaires, Éthiques
 - Cliniciens, épidémiologistes
- ▷ 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 oeuvre des actions pour l'amélioration de la qualité des données
- ▷ Centre de **recherche et développement** en “data sciences”
 - Labcom LITIS, partenariats avec des startups (VITADX, Enovacom, SanCare ...)

Données hospitalières :

Des données structurées :

- ▷ Biologie, diagnostic et actes par le PMSI, Prescription/Administrations médicamenteuses
- ▷ Décrites avec terminologies
 - Standards : CIM-10, CCAM, ATC
 - Ou locales... : biologie, formulaires médicaux

Beaucoup de données non structurées : (80 %)

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

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, il 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 *pancytopénie*, 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

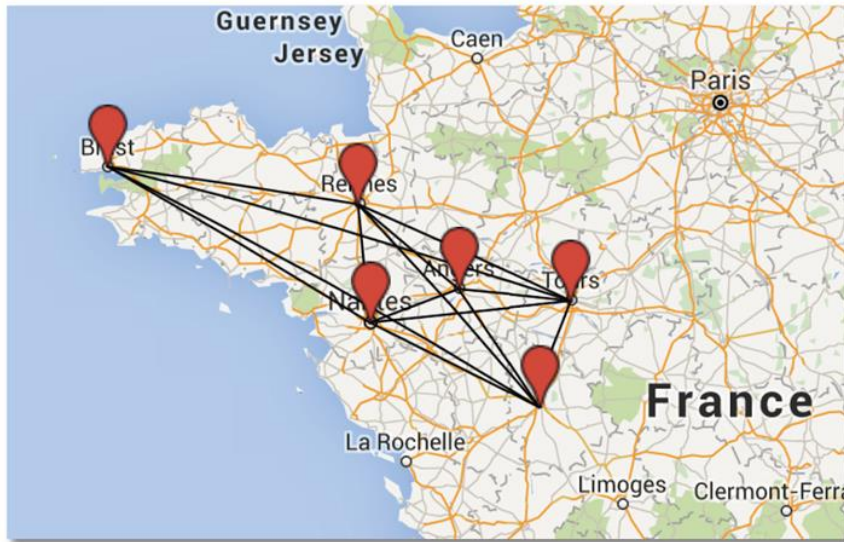
Contextualisation : chronologie

Contextualisation : négation, incertitude

Extraction de concepts : valeurs biologiques, prescription et posologies

Triplet : concept / valeur / unité

Réseau interrégionnal des Centres de Données Cliniques



4,6 millions
patients



112 millions
documents
cliniques



5+ Millions
de Séjours



1,3 milliard
Données
structurées

Données identité/séjour/structure

Comptes rendus Hospitaliers

Comptes-rendus de Consultation

Prescription/administration

Bio banque (CRB)

Dossiers Urgence

Données de Laboratoires

Dossiers infirmier

PMSI

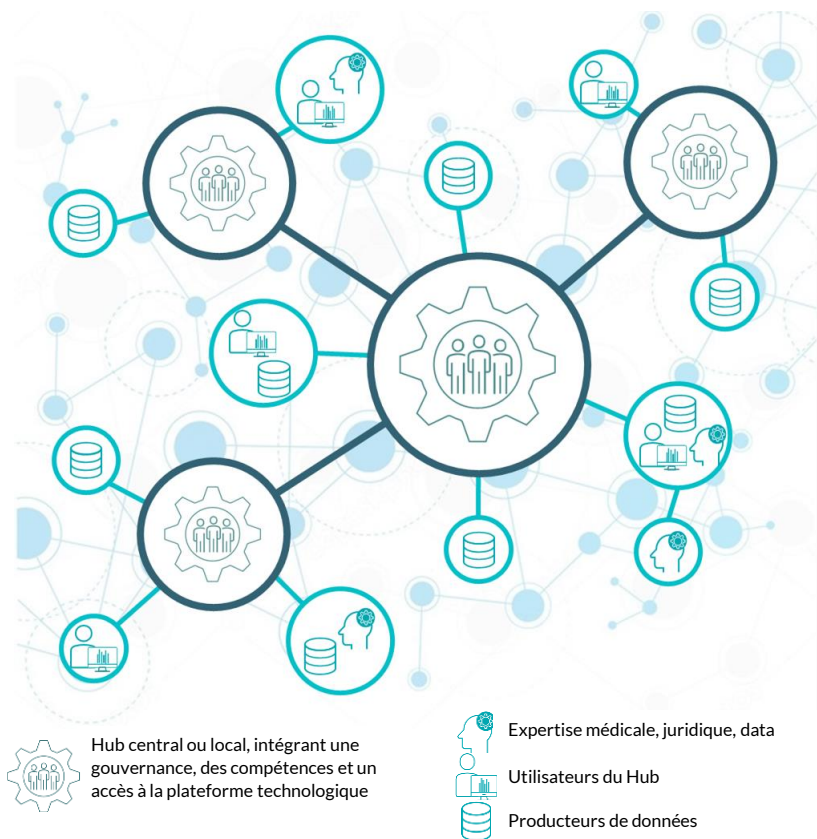
Anatomo-pathologie

CR Imagerie

Dossier Odontologie

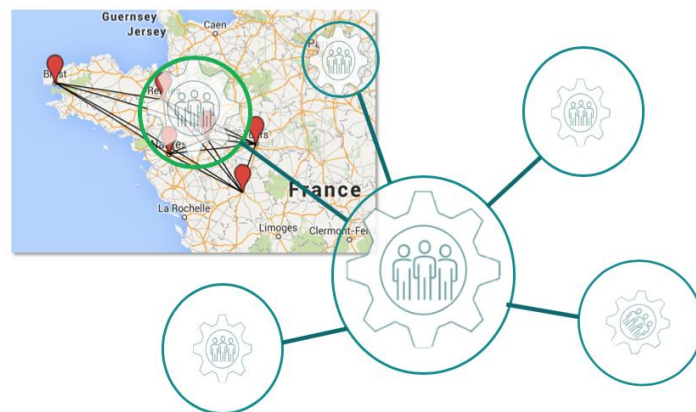
Le Health Data Hub : une approche déconcentrée s'appuyant sur les initiatives existantes

Organisation en réseau



Principes clé de l'organisation du Health Data Hub

- Un réseau constitué d'un Hub central et de Hubs locaux
- Une gouvernance, une offre de service, des processus et un engagement de résultats communs
- Une mutualisation des investissements technologiques à l'état de l'art et à haut niveau de sécurité



Pharmacovigilance

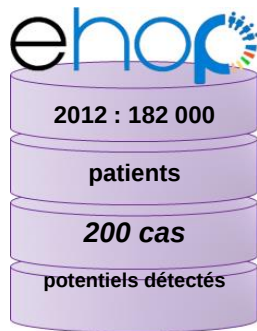
Contexte : Sous notification des effets indésirables médicamenteux

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

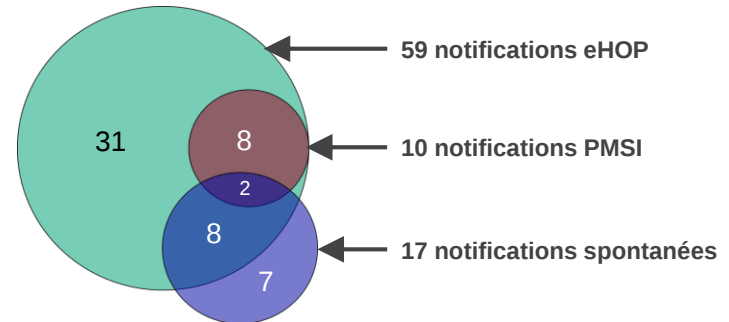
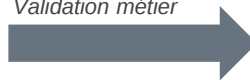
Méthodes :

- ▷ Recherche d'information et traitement automatique du langage

Requête : (choc or collapsus) and ("ige spécifique" or anaphy% or "bilan allergeo\-anesth%") or « **Tryptase%** » not ((alimentaire% NOT near((alimentaire%, medicamenteu%), 3)) or farine or seigle or blé) not (venin% or guepe% or hymenoptere% or piqure% or abeille% or bourdon% or insecte%) not near((antécédent%, choc% or collapsus), 50, true)



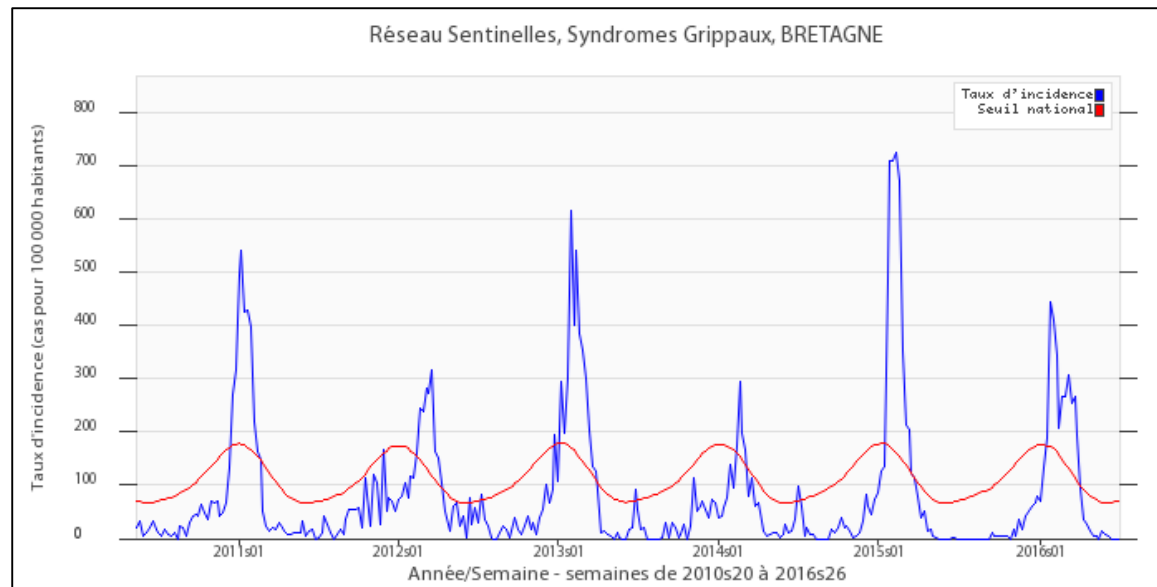
Validation métier



- ▷ 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)

Exemple d'application : surveillance de la grippe

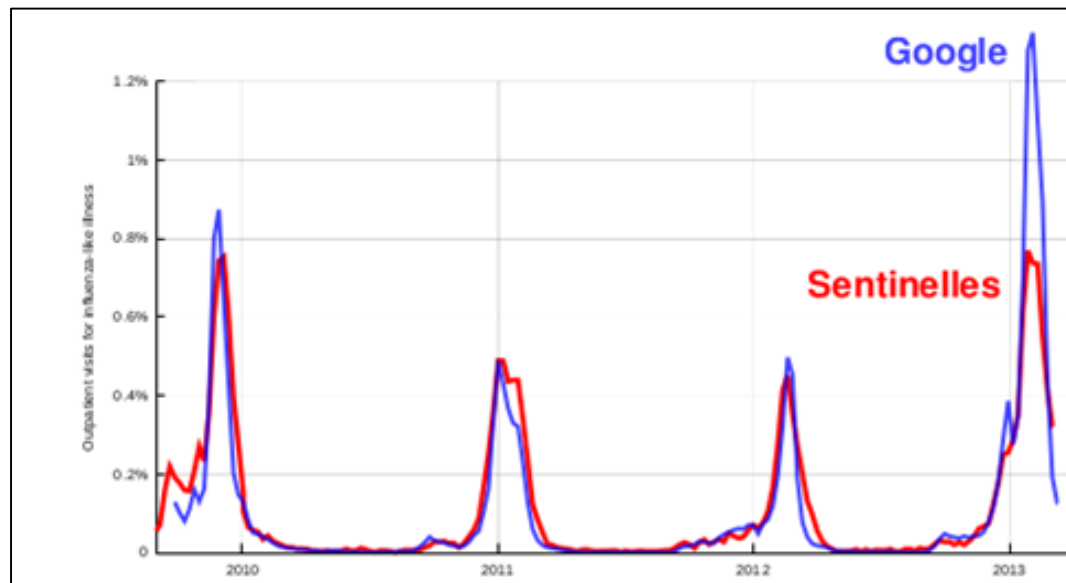
Surveillance épidémique par le réseau Sentinelles



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

Exemple d'application : surveillance de la grippe

Alternatives : surveillance basée sur les données du web

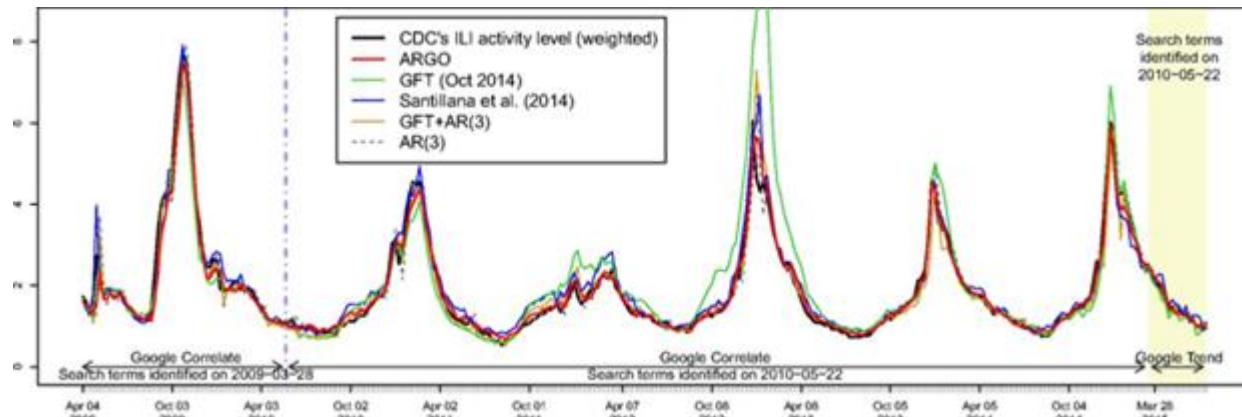


▷ Limite : indicateurs non fiables à plusieurs reprises

Exemple d'application : surveillance de la grippe

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

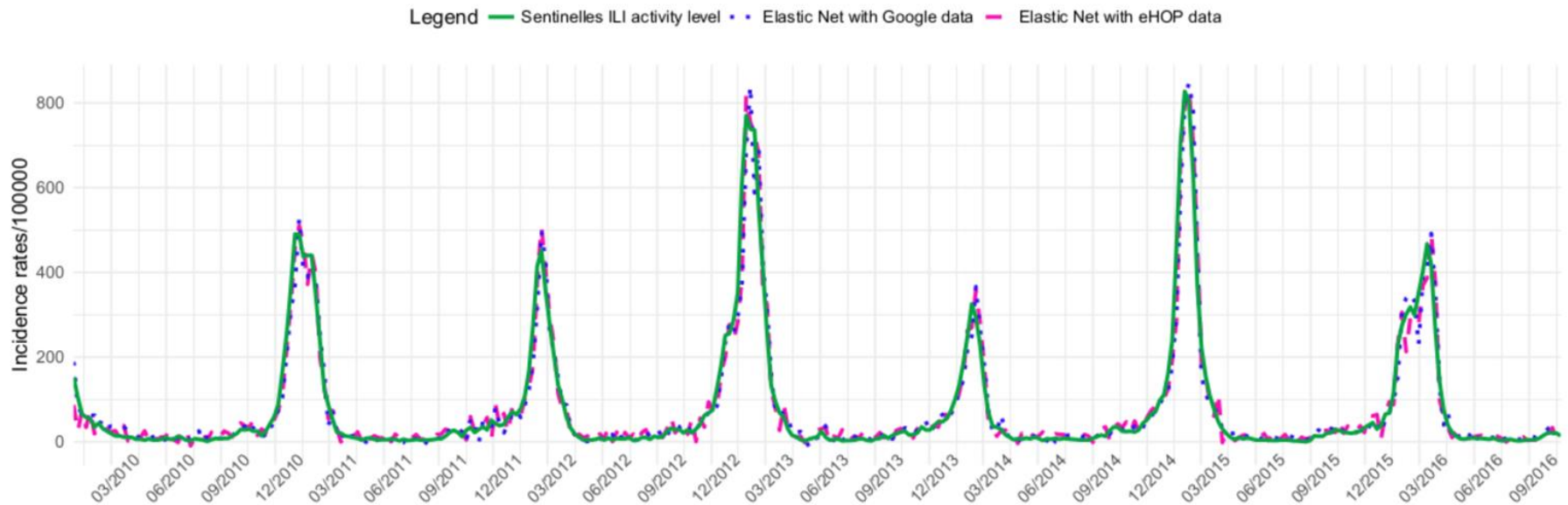
Modèle ARGO : Santillana et al.



- ▷ Surveillance de l'activité épidémique dans des sous populations (nouveau-né, personnes âgées)
- ▷ Surveillance de pathologies ne générant pas d'activité sur le web

surveillance de la grippe

Utilisation des données eHOP : Résultats niveau national :



Moyennes par année	Corrélation de Pearson		MSE		Δ Hauteur pic		Δ retard pic	
	Dataset eHOP	Dataset Google	Dataset eHOP	Dataset Google	Dataset eHOP	Dataset Google	Dataset eHOP	Dataset Google
RF	0.90	0.87	6916	9139	-22	-76	1.33	2
SVM	0.96	0.95	2716	3348	6	21	0.83	0.66
Elastic net	0.96	0.95	2664	3352	26	44	0.66	0.83

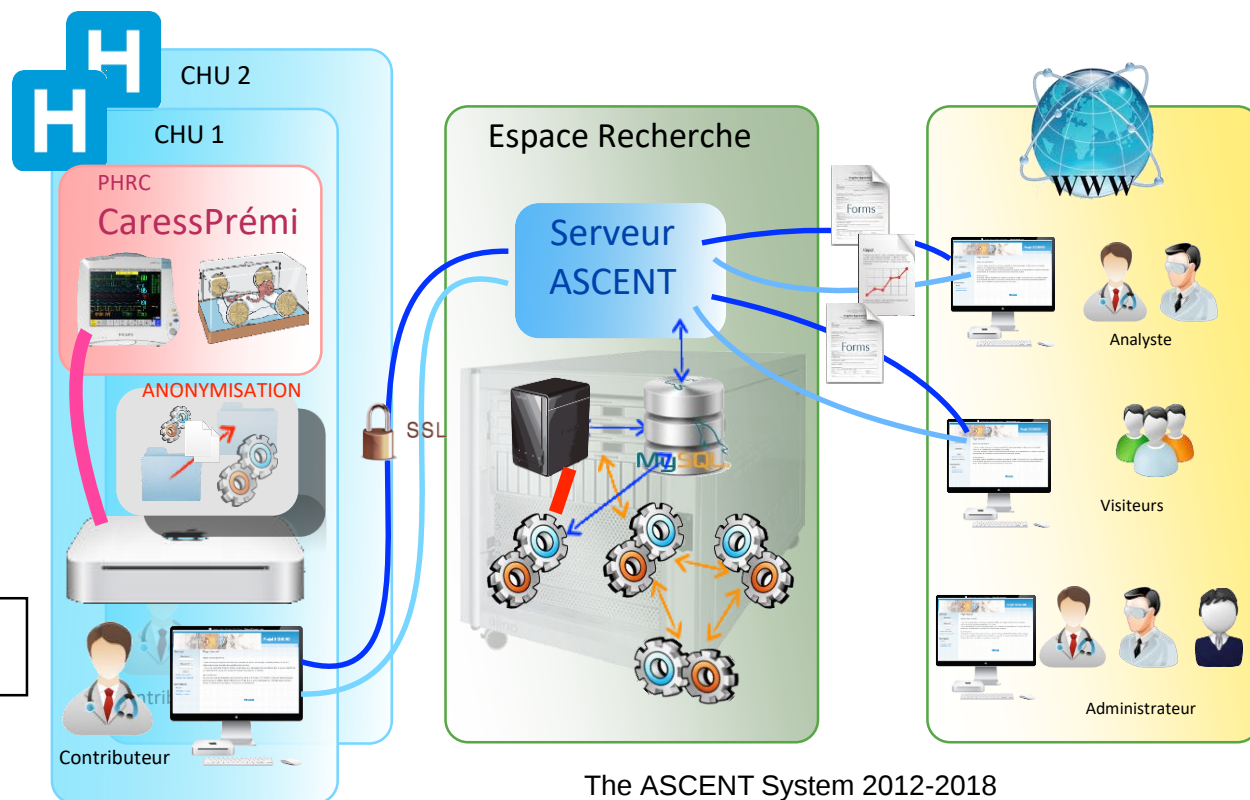
Objectif :

Identification de biomarqueur précoce et non-invasif d'infection chez le prématuré

Données:

- 550 patients inclus, 3 CHUs
- Haute résolution (ECG, resp.,...)
- Basse résolution (clinique-biologie)
- 24/7 x durée hospit.

> 30 années-patients pour chaque donnée avec résolution de 62,5 à 500Hz



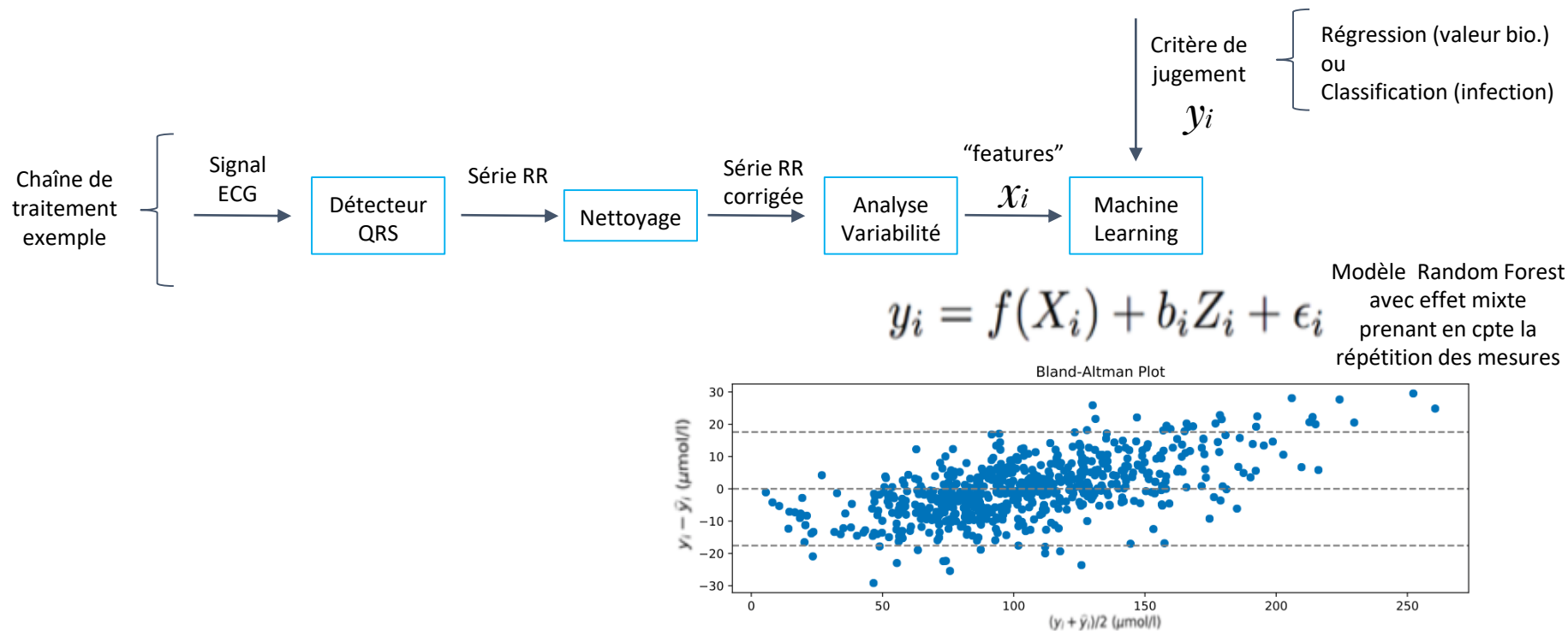
The ASCENT System 2012-2018

<https://ascent-ltsi.univ-rennes1.fr>

IDDN .FR.001.540040.000.S.P .2015.000.10000

PHRC-I CaressPremi

PI: Alain Beuchée



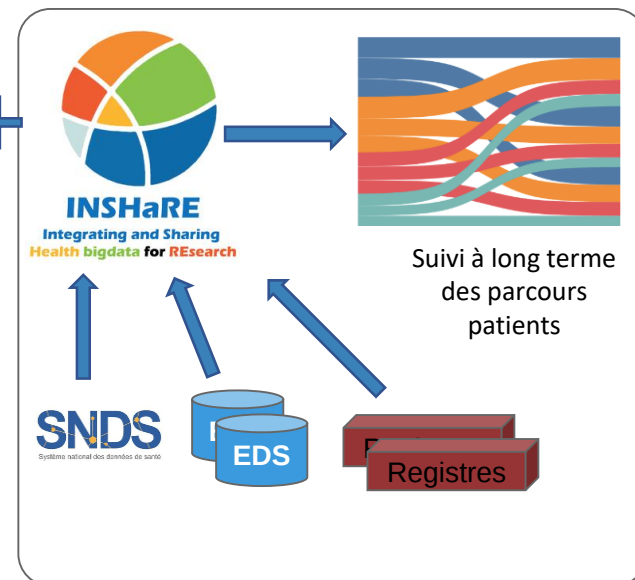
Prédiction court terme permettant une aide à la décision ... bénéfice à long terme?

Cas d'usage : Biomarqueurs du sepsis chez les nouveau-nés prématurés

Evaluer l'apport des
biomarqueurs prédictif des
sepsis sur le devenir à long
terme des nouveaux nés

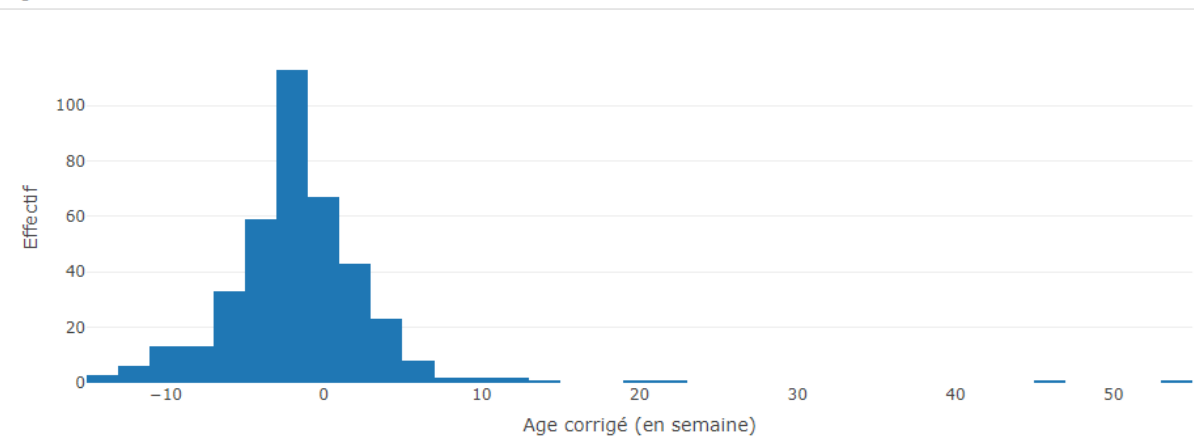


DÉMONSTRATION FAISABILITÉ CHU RENNES

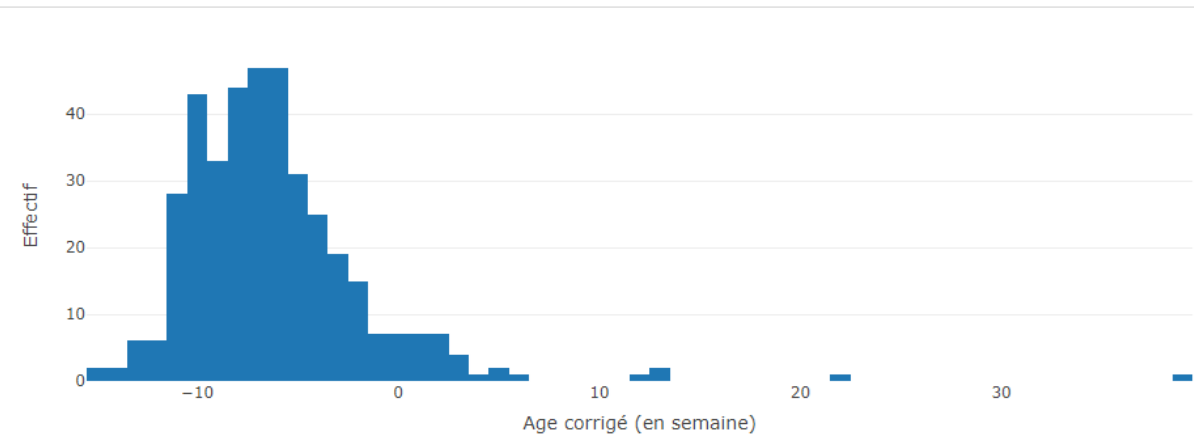


PERSPECTIVES

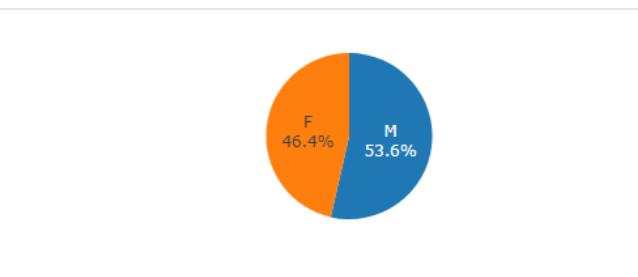
Âge à la sortie



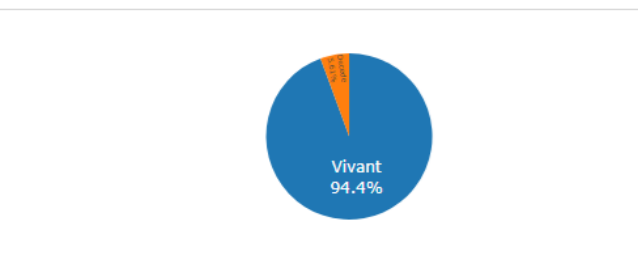
Durée d'oxygénothérapie



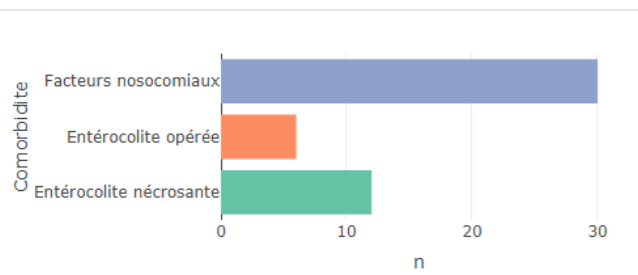
Sexe ratio



Statut vital



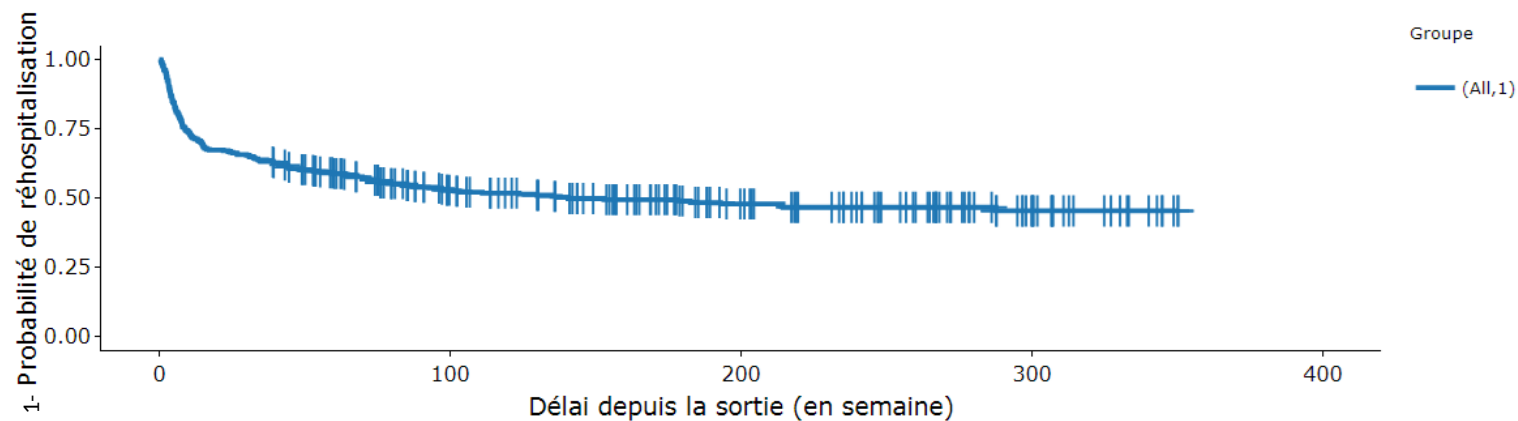
Infections



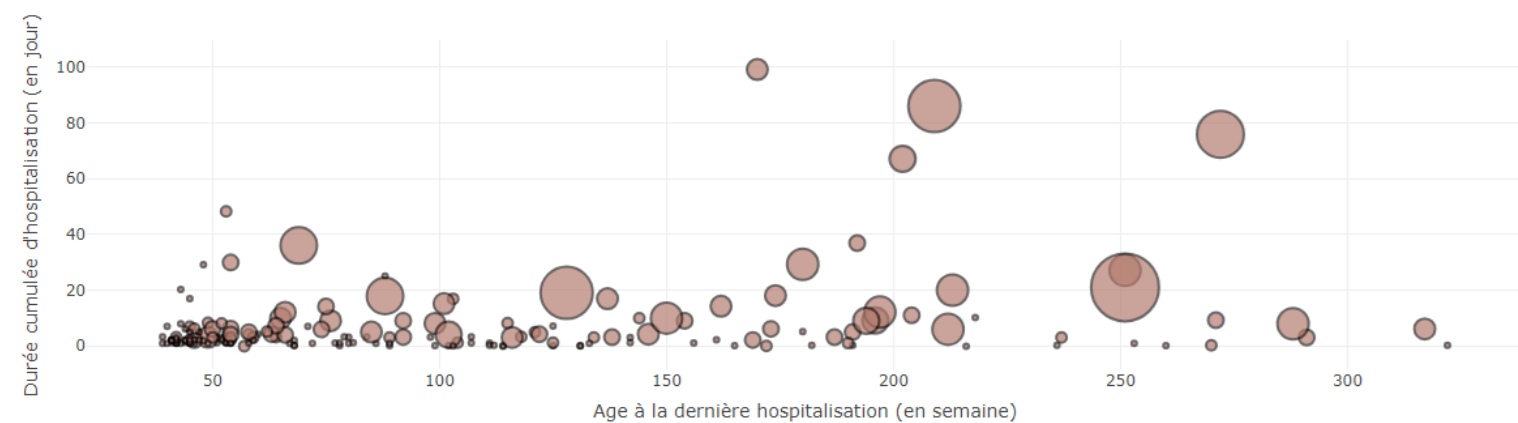
Pathologies

- ☒ Aucune
- ☐ Entérocolite nécrosante
- ☐ Entérocolite opérée
- ☐ Facteurs nosocomiaux
- ☐ Oxygénothérapie

Délais de réhospitalisation



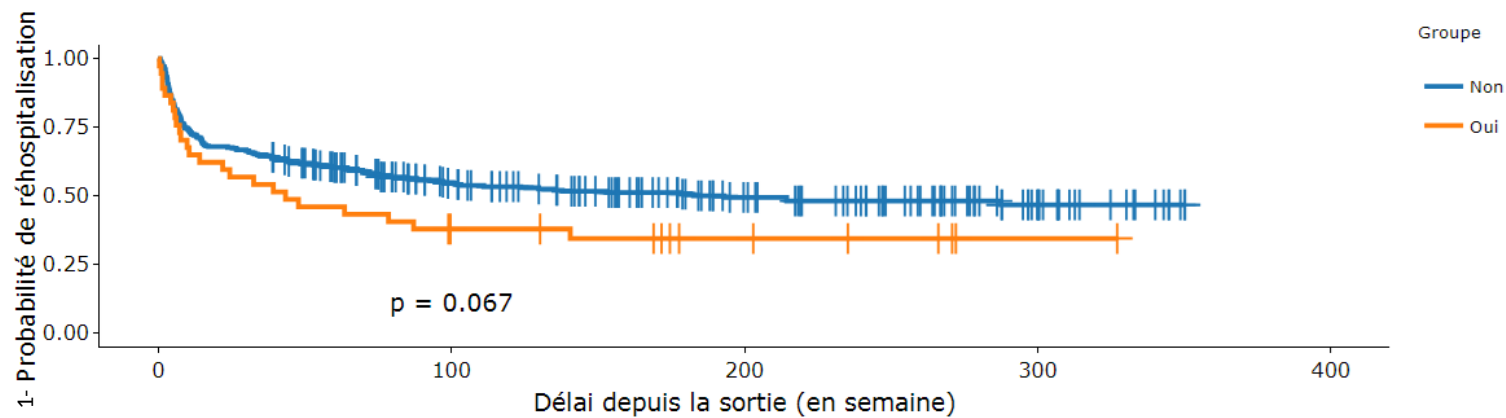
Trajectoires d'hospitalisations



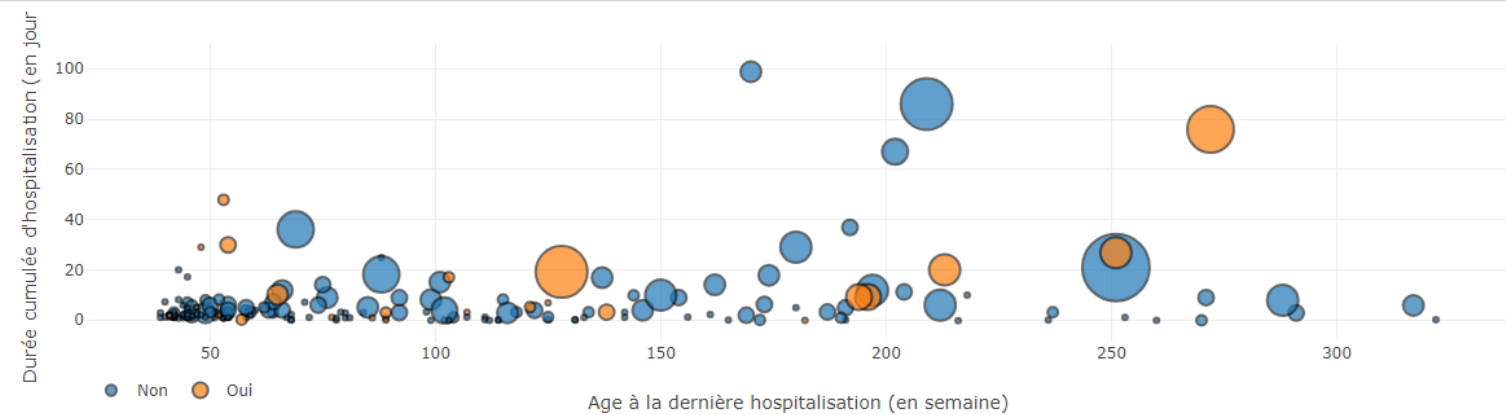
Pathologies

- ☐ Aucune
- ☒ Entérocolite nécrosante
- ☐ Entérocolite opérée
- ☐ Facteurs nosocomiaux
- ☐ Oxygénothérapie

Délais de réhospitalisation



Trajectoires d'hospitalisations



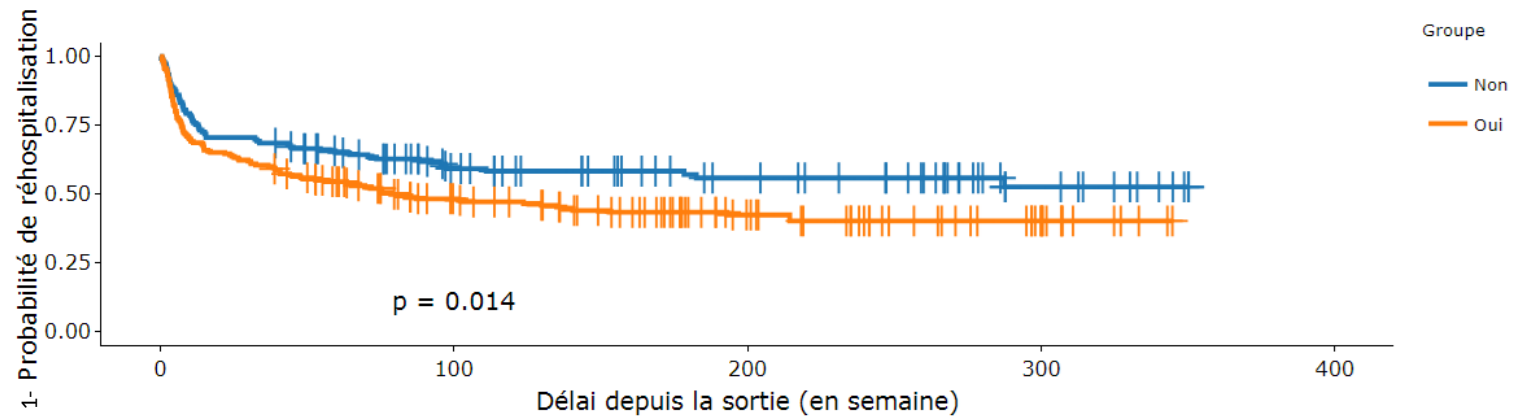
Pathologies

- ☐ Aucune
- ☐ Entérocolite nécrosante
- ☐ Entérocolite opérée
- ☐ Facteurs nosocomiaux
- ☒ Oxygénothérapie

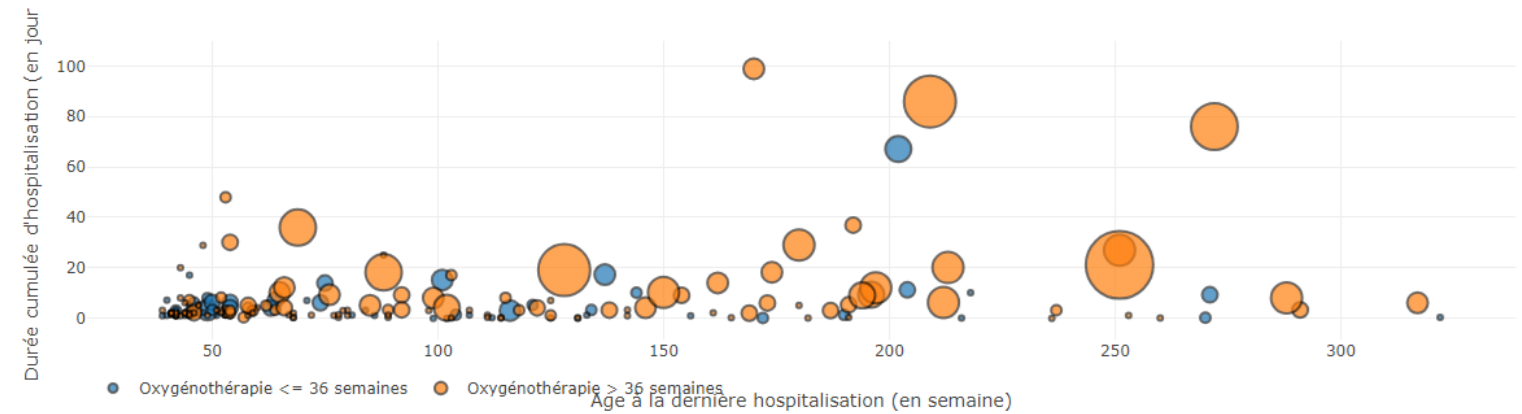
Nombre de jours:



Délais de réhospitalisation



Trajectoires d'hospitalisations



En conclusion



- **NOUVELLES OPPORTUNITES au travers de sources de données devenues exploitables**
 - Nécessite un **Jeu collectif – complémentarité des sources**
 - Rupture **quantitative** dans la masse de données disponibles
 - La qualité des données reste un facteur clef
- **Au-delà des données: l'expertise MULTIDISCIPLINAIRE est indispensable**
 - Jamais “seul face aux données”
 - Emergence de nouveaux métiers
 - En **proximité** l'expertise médicale et des patients
- **Créer les conditions de la CONFIANCE**
 - Usages **ETHIQUES** et **DEONTOLOGIQUES** dans des conditions de **PROTECTION** des données optimales
 - Conduite du changement
- **Ancrage de l'innovation dans les TERRITOIRES : Eviter l'écueil de la centralisation**
 - Catalyser les forces vives , impliquer les acteurs et créer des synergies.