Intégration données massives en santé Cas SNIIR-AM / DMH

Pr Marc CUGGIA
LTSI – Equipe projet Données Massives en Santé
Université Rennes1 - CHU Rennes,

Laboratoire Traitement du Signal et de l'Image LTSI

Instituts thématiques Inserm Institut national de la santé et de la recherche médicale

gircigrandouest LA RECHERCHE CLINIQUE

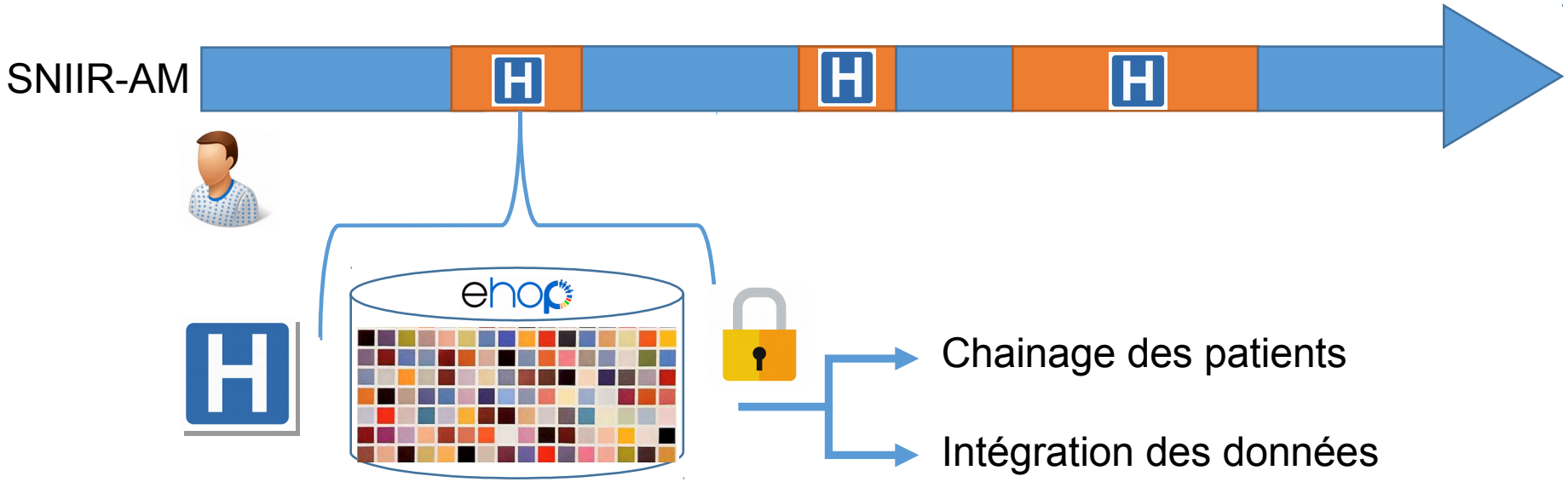
UNIVERSITÉ DE RENNES 1

CHU Rennes

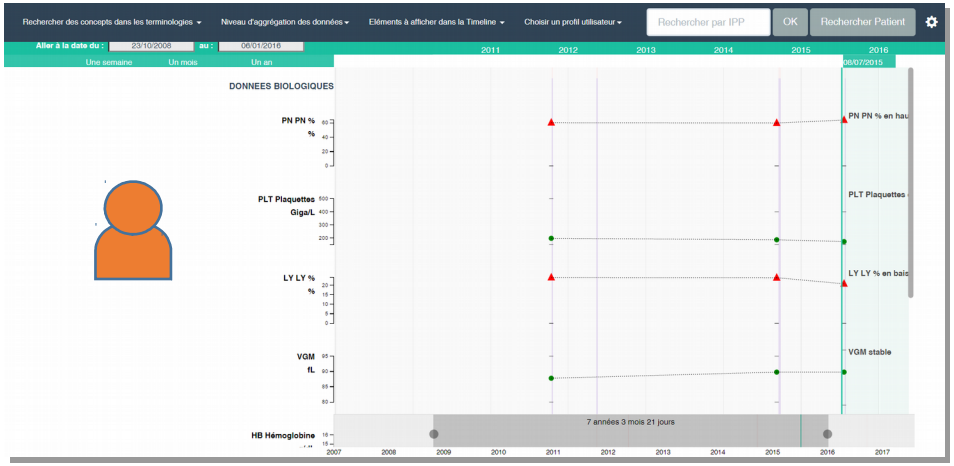
Cas de PEPS : Caractérisation de la trajectoire de santé des patients

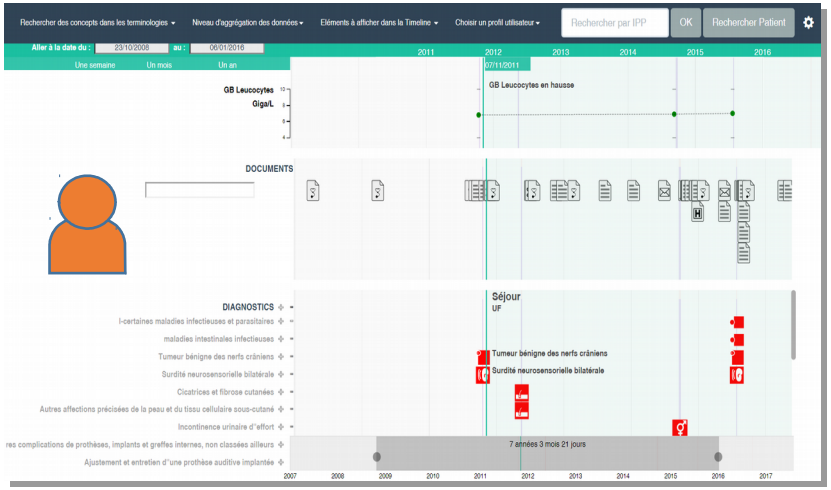
- **SNIIR-AM / DCIR :**
  - Apport :
    - transversalité de l'information
  - Limites :
    - Véracité des données
    - (ex : dispensation=/= Administration)
    - Faible granularité des données
      - Données Hospitalière sommaires (PMSI)
      - Zone d'ombre
    - Qualité des données
      - PMSI => upCoding T2A biais ++

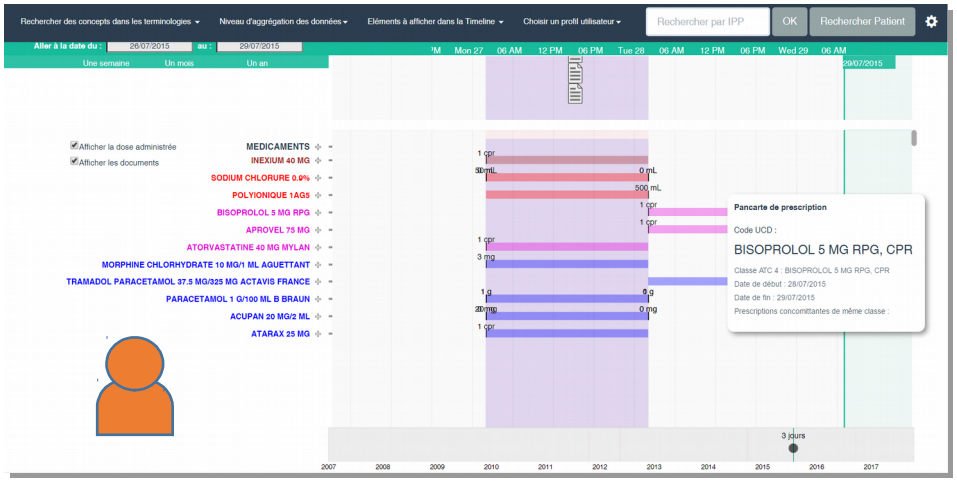
- **Entrepôt de données biomédicales:**
  - Apport :
    - Extrême richesse des données
    - Exhaustivité
    - Granularité
    - Retour au contexte de prise en charge (cliniciens voir patients)
  - Limites :
    - Ne concerne que la séquence d'hospitalisation et de consultation

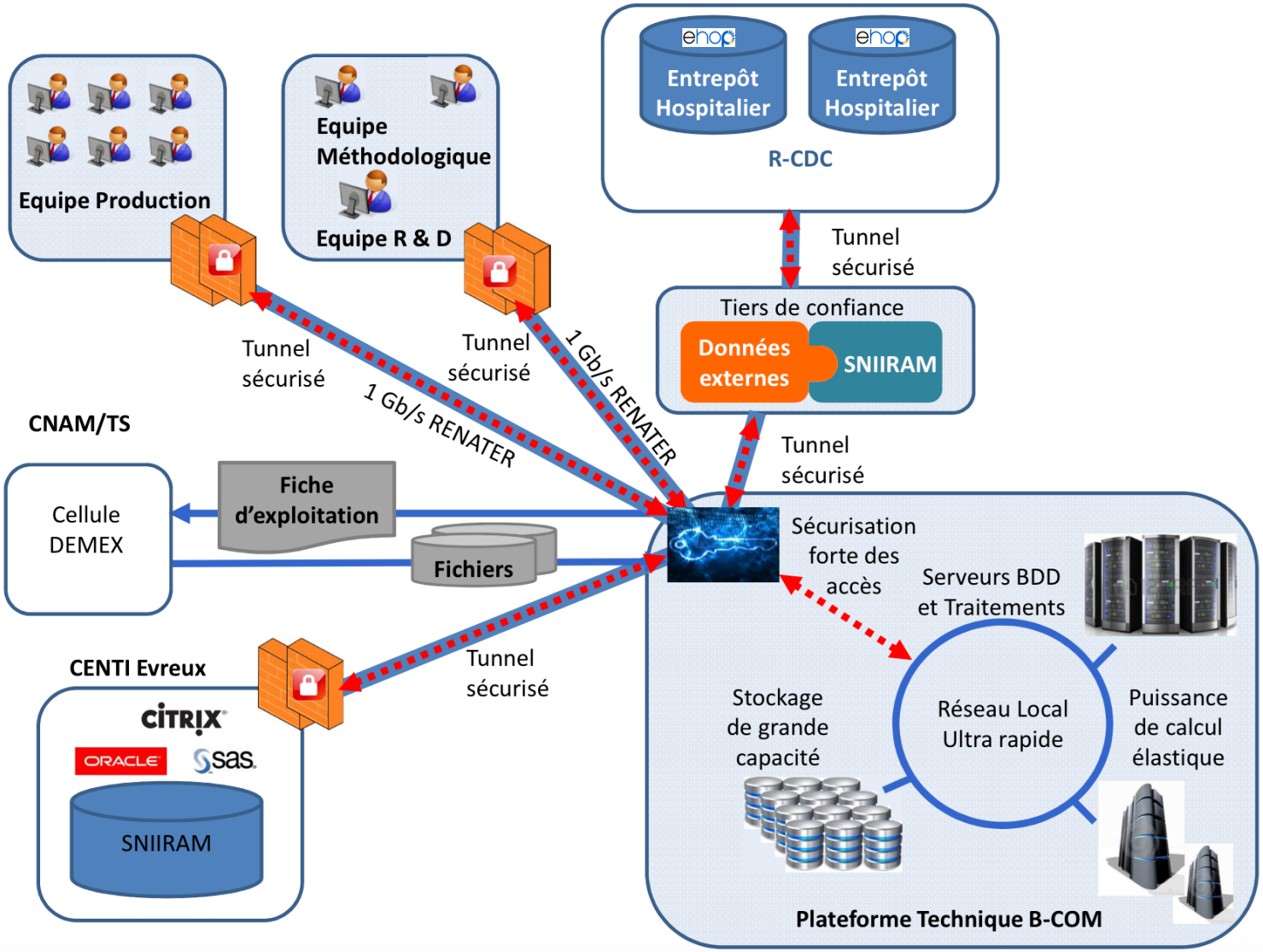


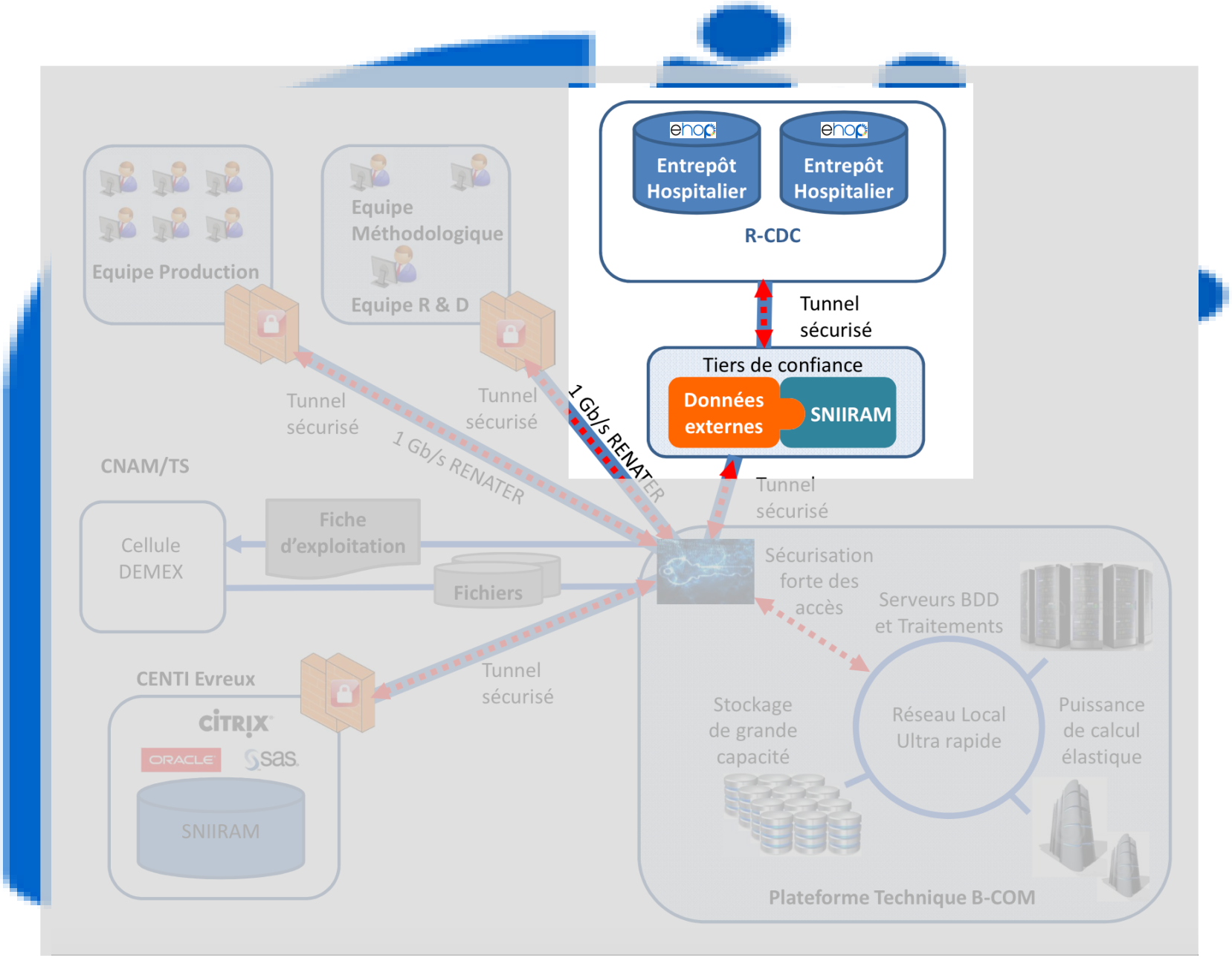
Visualisation des trajectoires patients







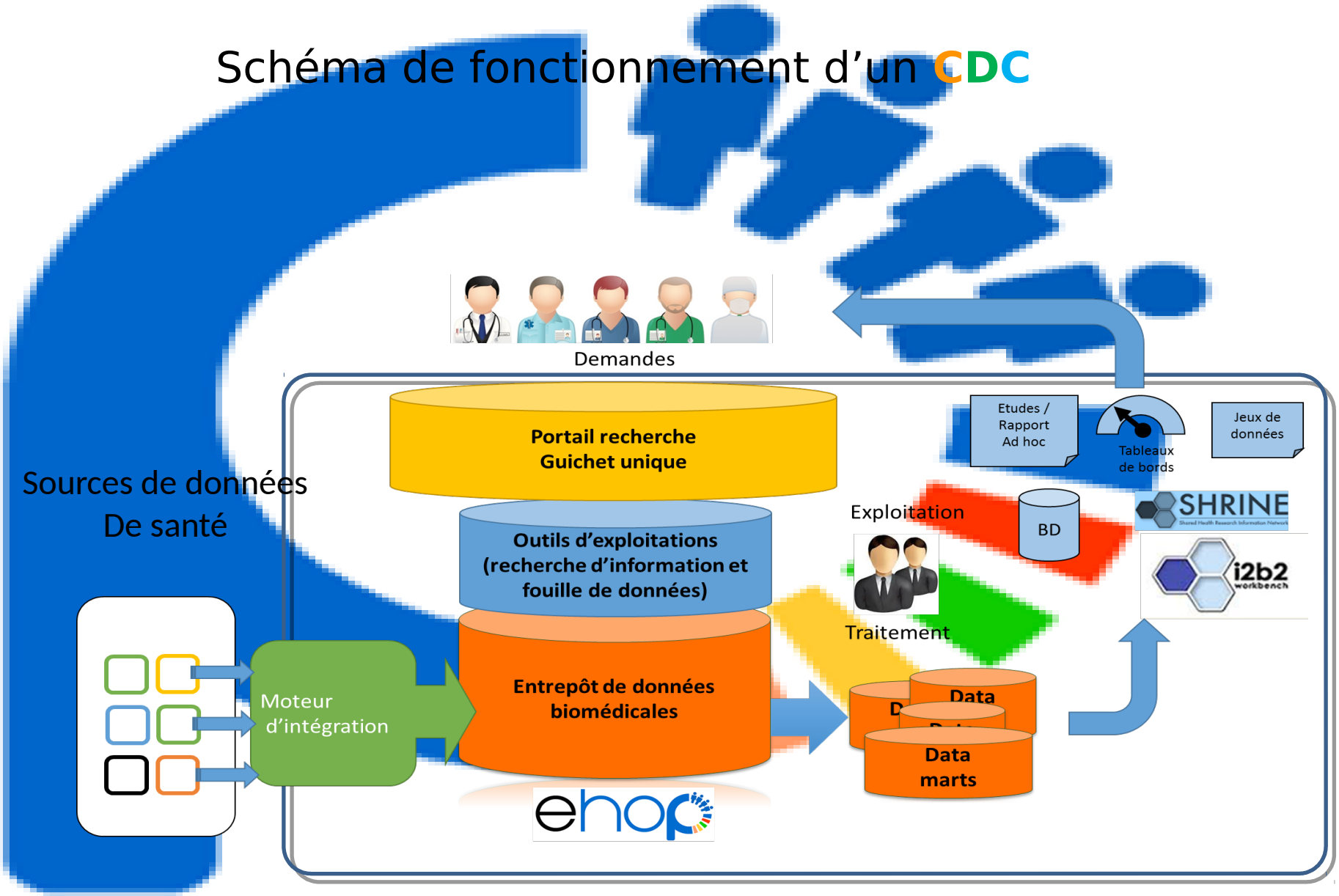

Entrepôt Hospitalier
Entrepôt Hospitalier
R-CDC
Equipe Production
Equipe Méthodologique
Equipe R & D
Tunnel sécurisé
Tunnel sécurisé
Tunnel sécurisé
1 Gb/s RENATER
1 Gb/s RENATER
Tiers de confiance
Données externes
SNIIRAM
Tunnel sécurisé
CNAM/TS
Cellule DEMEX
Fiche d'exploitation
Fichiers
Sécurisation forte des accès
Serveurs BDD et Traitements
Tunnel sécurisé
CENTI Evreux
CITRIX
ORACLE
sas
SNIIRAM
Stockage de grande capacité
Réseau Local Ultra rapide
Puissance de calcul élastique
Plateforme Technique B-COM


Entrepôt Hospitalier
Entrepôt Hospitalier
R-CDC
Tunnel sécurisé
Tiers de confiance
Données externes
SNIIRAM
1 Gb/s RENATER
Tunnel sécurisé
Equipe Production
Equipe Méthodologique
Equipe R & D
Tunnel sécurisé
Tunnel sécurisé
1 Gb/s RENATER
CNAM/TS
Cellule DEMEX
Fiche d'exploitation
Fichiers
Sécurisation forte des accès
Serveurs BDD et Traitements
Tunnel sécurisé
CENTI Evreux
CITRIX
ORACLE
sas
SNIIRAM
Stockage de grande capacité
Réseau Local Ultra rapide
Puissance de calcul élastique
Plateforme Technique B-COM

les Centres de Données Cliniques (CDC)

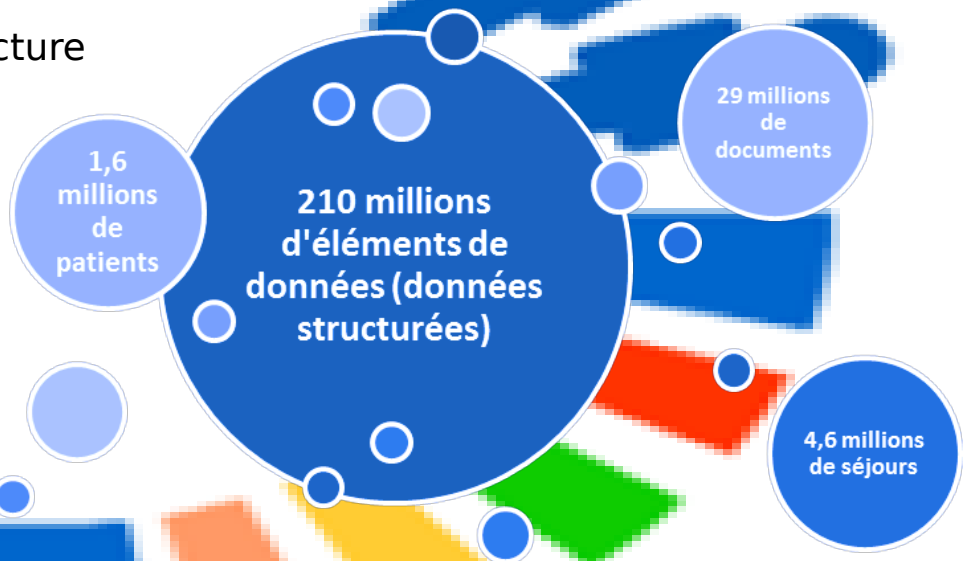
- **Organisation intra-hospitalière « guichet unique »**
- Propose une **Expertise Méthodologique**
- **Offre de services :** exploiter les données de Santé au sein de l'établissement
- Rassemble des compétences multidisciplinaires : **Informatiques réglementaires** et **Biostatistiques** , **Métiers** (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
- **Centres de recherche et développement en data sciences**

Schéma de fonctionnement d'un CDC



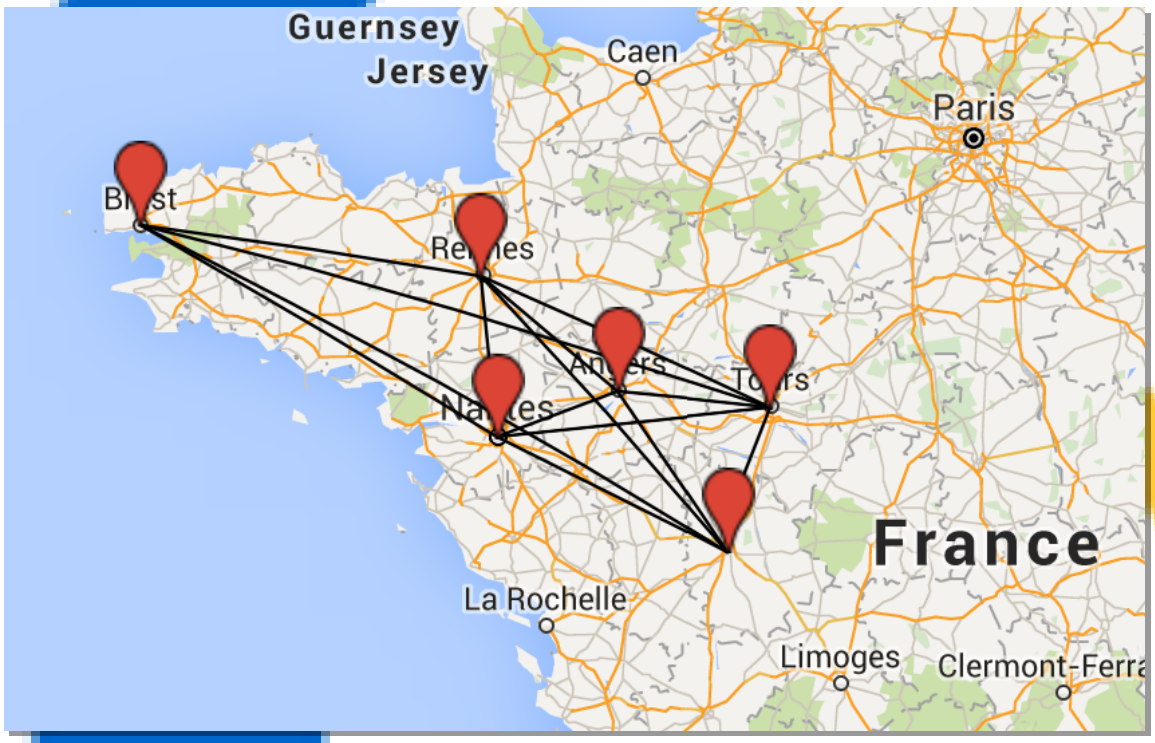
Exemple du CDC de Rennes

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



140 Etudes cliniques et 15 projets de recherche (depuis la MEP 2016)

Le réseau interrégional des CDC



Méthode : Chainage et intégration
Critères de qualité des données

- **Exactitude (Accuracy)+++**

  Comment les données ont elles été saisies ?
  Vérification de la saisie des données avec des tables de référence ?

- Complétude (Completeness)

  Taux de données manquantes ?
  Sait-on pourquoi certaines données sont manquantes ?

- **Uniformité (Consistency)+++**

  Données sont-elles au même format dans les 2 sources ?
  Doublons repérés ?

- Accessibilité (Accessibility)

  A-t-on accès à l'ensemble des données disponibles ?
  Quelles sont les potentielles variables d'appariement qui ne sont pas disponibles ?

- Actualité (Timeliness)

  Sait-on à quel moment les données ont été saisies ?
  Ages+++ Adresses+++

- Plausibilité (Believability)

  Est-on sûr que nos données sont vraies ?
  Valeurs fausses ? Impossibles ?

Méthode : Appariement d'identités

4 étapes principales

- 1 : Nettoyage des données et standardisation
- 2 : Indexation
- 3 : Calcul des mesures de similarité
- 4 : Classement des paires potentielles

Variables SNIIRAM (étude SACHA)

Effectif : 25 327 patients

- Année de naissance
- Sexe
- N° INSEE de la commune de naissance
- Séjours MCO (2011 à 2013)
- Etablissement
- Date de début (2011 à 2013)
- Date de fin (2012 à 2013)
- Diagnostics (Principal, Secondaire, Associés)

Variables eHOP CHU de Rennes

Effectif : 1 600 000 patients

- Date de naissance
- Sexe
- N° INSEE de la commune de naissance
- Séjours MCO depuis 1995
- Etablissement (CHU de Rennes)
- Date de début
- Date de fin
- Diagnostics (Principal, Secondaire, Associés)

Principes d'appariement d'un patient SNIIRAM avec un patient EHOP

- Pour un patient du SNIIRAM, **n** est le nombre de patients EHOP pour lesquels les critères d'appariement sont vérifiés dans au moins 1 séjour MCO du SNIIRAM

- **Si n=0** : Appariement en échec
- **Si n=1** : Appariement réussi
- **Si n>1** : Appariement à affiner

Critère retenu (test n° 1)

Correspondance exacte :

- Année de naissance
- Sexe
- Date de début de séjour
- Diagnostic principal

Critère retenu (test n° 2)

Correspondance exacte :

- Année de naissance
- Sexe
- Date de début de séjour
- Date de fin de séjour
- Diagnostic principal

Résultats

- Nombre de patients SNIIRAM : 25 327 patients
- Temps d'appariement : 18 secondes

	Test 1 Temps=18 s n=25 327	Test 2 Temps=18 s n=25 327
n=0 (pas de correspondance de patients)	25 (0,1 %)	711 (2,8 %)
n>0 (au moins une correspondance de patients)	25 302 (99,9%)	24 616 (97,2 %)
n>1 (plus d'une correspondance de patients)	**781 (3,1 %)**	**221 (0,9 %)**
n=1 (une correspondance unique de patients)	**24 521 (96,8 %)**	**24 395 (96,3 %)**

Discussion et perspectives

Correspondances non trouvées

- Séjours manquants dans EHOP ?

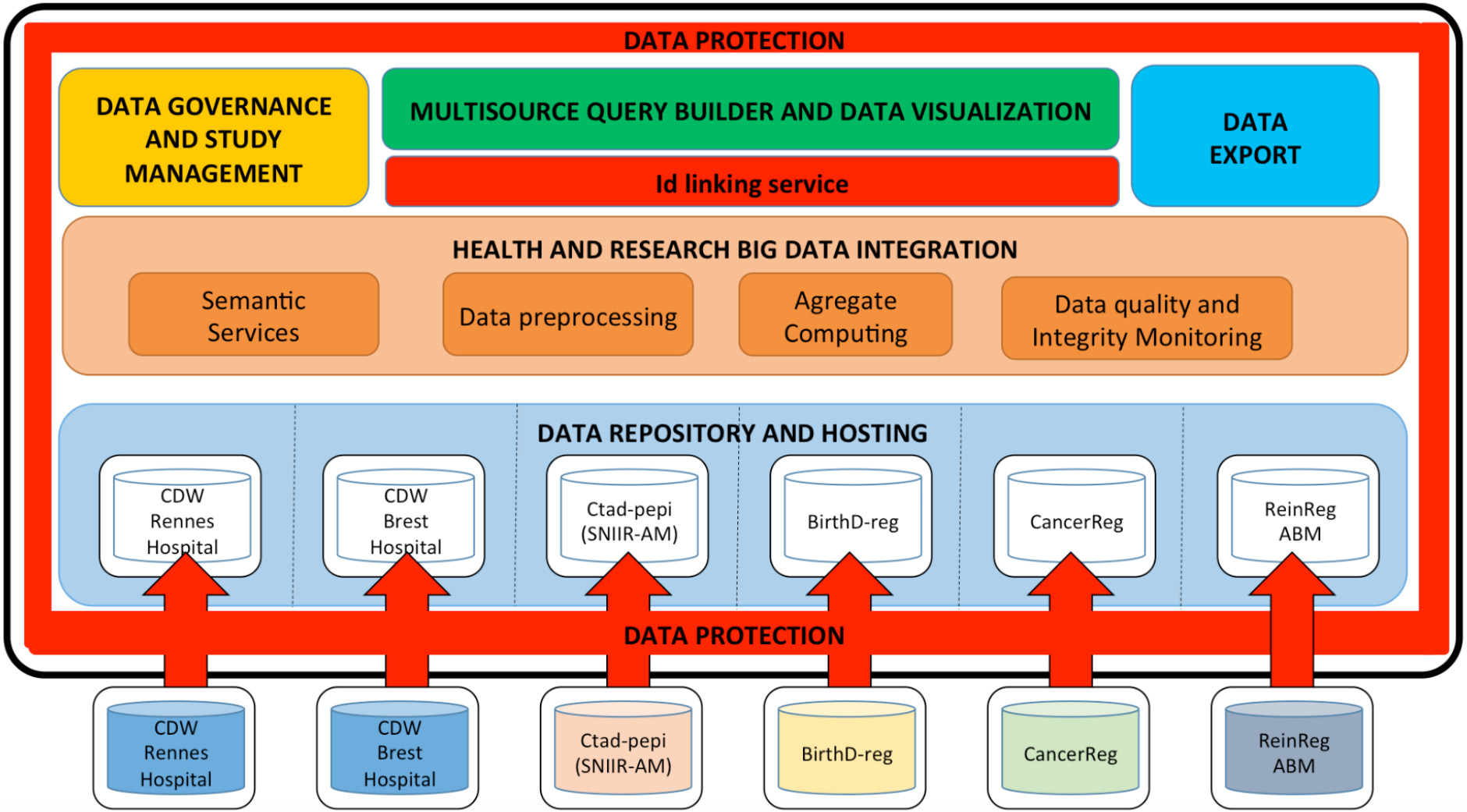
Correspondances multiples

- Utiliser le diagnostic secondaire, les diagnostics associés, les actes…

Perspectives : Intégration d'autres sources de données

- **Projet INSHARE** : Intégrating and Sharing Health Big Data for Research
- ANR 2016-2018

INSHaRE
Integrating and Sharing
Health bigdata for REsearch



Conclusion

- Synergie et complémentarité des sources de données
  - Big et small data
  - GHT, Soin primaire, EPAD, santé connectée …
- Verrous :
  - Méthodologiques : chainage et intégration des données
  - Opérationnels
  - Ethiques, juridiques et … administratif
- Approche centrée territoire de populations
  - Proche des problématiques de terrain, des expertises médicales et des patients

Merci pour votre attention

Contact : marc.cuggia@univ-rennes1.fr