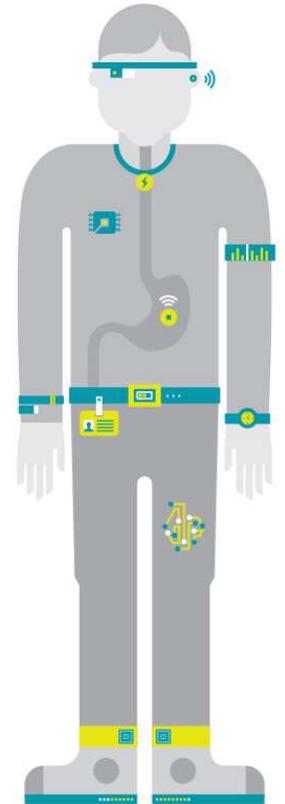




# Big Data et Open data :



## QUELS ACCÈS POUR LA RECHERCHE ?



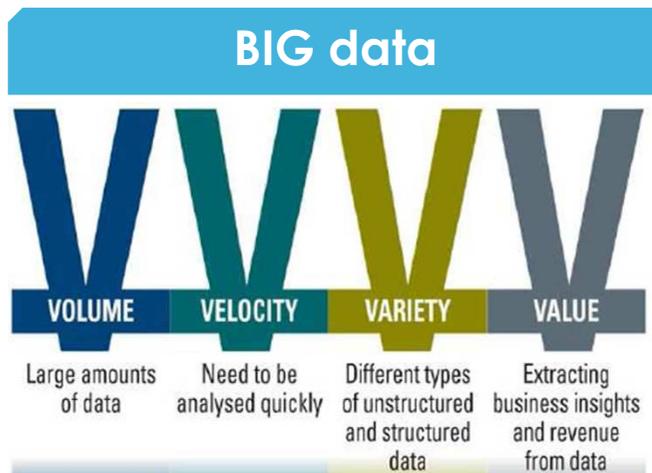
**Gilles Chatellier, Vincent Varlet, Corinne Blachier-Poisson  
et les participants à la table ronde N° 6 de Giens 2015 :**

Nathalie Beslay, Jehan-Michel Behier, David Braunstein, Mireille Caralp, Brigitte Cognard-Chassol, Isabelle Diaz, Laure Fournier, Anne Jossieran, Philippe Lechat, Cinira Lefevre, Franck von Lennep, Karine Levesque, Philippe Maugendre, Guillaume Marchand, Didier Mennecier, Nicholas Moore, Sophie Ravoire, Christine Riou





# 2 | Introduction et définitions



Le monde double son stockage de données tous les ans

25 Hexa Octets ( $10^{18}$ ) de données de santé en France (stock)

Les analyses et algorithmes nécessitent des professionnels pointus

Les enjeux scientifiques et industriels dépassent nos frontières le domaine étant très compétitif

La confrontation entre la garantie de confidentialité et le dynamisme et l'ouverture au service :

- du patient
- de la société (intérêt collectif ou intérêt industriel)
- de la science

Le principe de l'OPEN data doit avant tout être ressenti comme BENEFIQUE

Cinq caractéristiques majeures semblent se détacher mais il reste de nombreuses problématiques, telles :

- Mise en œuvre informatique
- Propriété de la donnée ou de l'analyse

## OPEN data

**PARTAGE (intérêt public)**

**COÛT**

**ACCESSIBILITÉ**

**STANDARDISATION FORMAT**

**RÉ-UTILISATION**



# 3 | Entrepôts de données (bio)médicales

## Clinical Data Warehouse

Des questions juridiques, politiques, techniques..

Des interrogations sur la qualité de la collecte des données

De l'anonymisation à la dé-identification et à la ré-identification

Une disponibilité insuffisante en France

## CNIL, Art.47, SNDS

La France a des atouts ! Ne pas prendre de retard !

Source des données: très variée!

- SI des établissements universitaires (soins)
- SI Médico-économiques (PMSI, prescription → SNIIRAM)
- Surveillance organisée des décès : CepiDC
- Et aussi:
  - Cohortes,
  - Registres, réseaux patients ...
  - Maladies rares
  - Données d'essais contrôlés
  - ....

Les **entrepôts** sont encore insuffisamment développés en France en raison de leur coût, du manque de personnels qualifiés, de la crainte des aspects réglementaires, mais les outils existent (**I2B2**)

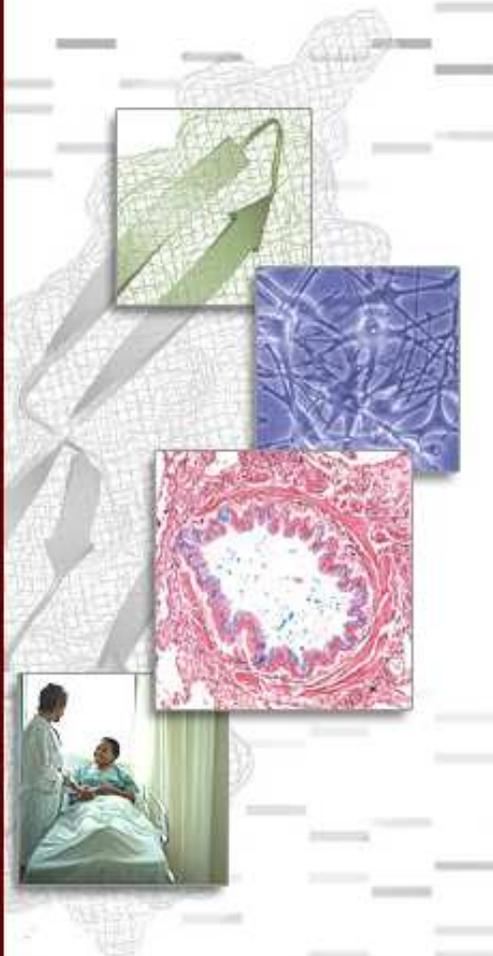
De manière générale:

- Pilotage stratégique et gouvernance en devenir
- Textes législatifs encore en devenir
- Obstacles aux accès (ou limitations fortes sous conditions)
- La crainte de stigmatisation et d'interprétation de la finalité
- Délais d'accès décourageants dans un secteur dominé par la pression du temps (universitaires, industriels)

Cependant des éléments de progrès :

- L'identifiant unique, sésame, séparé des données de santé
- La traçabilité des accès (demain le blockchain?)
- Des méthodologies de référence





### MISSION

i2b2 (Informatics for Integrating Biology and the Bedside) is an NIH-funded National Center for Biomedical Computing based at Partners HealthCare System. The i2b2 Center is developing a scalable informatics framework that will enable clinical researchers to use existing clinical data for discovery research and, when combined with IRB-approved genomic data, facilitate the design of targeted therapies for individual patients with diseases having genetic origins. This platform currently enjoys wide international adoption by the CTSA network, academic health centers, and industry. i2b2 is funded as a cooperative agreement with the National Institutes of Health.

### DRIVING BIOLOGY PROJECTS

- :: Overview
- :: Current DBPs
  - :: Autoimmune/CV Diseases
  - :: Diabetes/CV Diseases
- :: Past DBPs
  - :: Airways Diseases
  - :: Hypertension
  - :: Type 2 Diabetes Mellitus
  - :: Huntington's Disease
  - :: Major Depressive Disorder
  - :: Rheumatoid Arthritis
  - :: Obesity

### RESOURCES

- :: Overview
- :: Computational Tools
- :: Software
- :: NLP Research Data Sets
- :: NLP Shared Tasks
- :: Academic Users' Group
- :: Publication Data

### SOFTWARE



# 4 | Les objets connectés et la e-santé

## 3 catégories d'usage



### Dispositifs médicaux connectés de télémédecine (HPST) :

- Recueil, traitement et transmission de données médicales
- Partie intégrante du diagnostic et/ou suivi

### Dispositifs connectables de médecine :

- Plus « anciens », mais connectables à d'autres systèmes
- Qualité des données : pertinent, reproductible

### Objets connectés participant au bien-être

- Non CE médical, non DM (sans évaluation des capteurs)
- La nécessité d'accompagner et qualifier les constructeurs s'ils veulent participer à l'effort vers le patient

## Collecte, sécurité, CGU

### Une vigilance « hors F et EU »



### La qualité des flux (entre équipements) et la sécurité

### L'information et le respect de la vie privée (agrégation)

### Les analyses statistiques et l'utilisation



# 5 | Les Recommandations (France / EU)

**Faciliter une organisation  
afin de promouvoir  
l'accès et la qualité**

**Au sein de l'UE, la France n'est pas en déficit qualitatif ou quantitatif des bases disponibles**

- Faciliter le travail des chercheurs
- Rendre la recherche en France compétitive et attractive

**Alléger /automatiser les évaluations de la CNIL afin de réduire les délais (6-8 mois) :  
méthodologie (demandeurs) et ressources (CNIL)**

**Institut National des Données : une gouvernance claire, un cadre opérationnels partagés avec TOUS les  
acteurs (« études d'intérêt public) et des ressources adaptées**

**Médicalisation des données : création d'un groupe de travail à la recherche d'interfaçage entres les données  
nationales et les « entrepôts de données »**

**Nomenclatures à mettre à jour (classification des actes médicaux)**

**Éduquer et transformer les métiers. Former et recruter des nouvelles compétences : « data scientist »**

**Faire en sorte que le Big soit Open (base SNIIRAM), en le rendant par exemple plus simple et plus accessible.  
(exemple : échantillons sans besoin de la CNIL)**

**Apparier les bases (simplification de la procédure, réduction délais, utilisation NIR, tiers de confiance)**

**Standardiser un mode opératoire pour l'open data pérenne des bases (exemple : INSERM)**



**Merci à ce groupe si tonique!**



# Recommandations

## Quel accès pour la recherche ? -1-

### Constats

1. Délais
2. Institut National des Données de Santé
3. « Médicalisation » des données
4. Nomenclatures
5. Ressources méthodologiques insuffisantes

### Améliorations

1. Respect des délais!  
Méthodologies de référence
2. Le créer, et le faire fonctionner, moyens disponibles, valoriser les experts (valable pour la CNIL)
3. Côté de la pathologie, classification NYHA (modification de code) : groupe utilisateur (chercheurs, industriels)
4. Mise a jour rapide
5. Coursus « Big data » (recenser, créer ..)

# Recommandations

## Quel accès pour la recherche ? -2-

### Constats

1. Modèle de données SNIRAM
2. Documentation SNIRAM
3. Appariement des bases
4. Open Data : cohortes, essais industriels, académiques, extraction big data
5. Intérêt public à définir

### Améliorations

1. Améliorer le modèle
2. WIKI-SNIRAM
3. Simplification procédure /délais de mise en œuvre
4. Contractualiser l'accès ultérieur aux données et sa pérennisation (modèle INSERM)
5. Pratique décisionnelle élaborée par les acteurs

