



Ministère des finances et des comptes publics
Ministère des affaires sociales, de la santé et des droits
des femmes
Ministère du travail, de l'emploi, de la formation professionnelle
et du dialogue social

santé

protection sociale

drees

recherche, études, évaluation et statistiques

Données de santé

André Loth

Typologies des données de santé : celle d'Etalab... et d'autres

Données socio-démographiques	Santé publique et épidémiologie	<ul style="list-style-type: none">▪ Etat de santé▪ Epidémiologie▪ Habitudes de vie▪ Environnement
	Offre de soins	<ul style="list-style-type: none">▪ Infrastructures▪ Services proposés▪ Personnel▪ Tarifs et honoraires▪ Infrastructure de recherche▪ Informations médicales
	Consommation de soins	<ul style="list-style-type: none">▪ Activité des établissements de santé▪ Consultations▪ Biologie▪ Transport▪ Médicaments et dispositifs médicaux▪ Indemnités et accidents de travail▪ Activité des établissements médico-sociaux▪ Autres
	Performance et opérations	<ul style="list-style-type: none">▪ Performance financière▪ Performance opérationnelle▪ Qualité
	Autres dépenses	<ul style="list-style-type: none">▪ Dépenses de soin et de biens médicaux▪ Dépenses de soin de longue durée▪ Dépenses de prévention▪ Autres dépenses soins▪ Dépenses administratives▪ Modalités/sources de financement

➤ Recensement des données disponibles par Etalab

➤ Sources : voir le rapport 2009 du HCSP sur les données de santé

➤ Mission IGAS sur les informations à la disposition du public (offre de soins, indicateurs de qualité, conduites à tenir...)

➤ Distinguer données individuelles et autres données (qualité de l'air, de l'eau, offre de soins...)

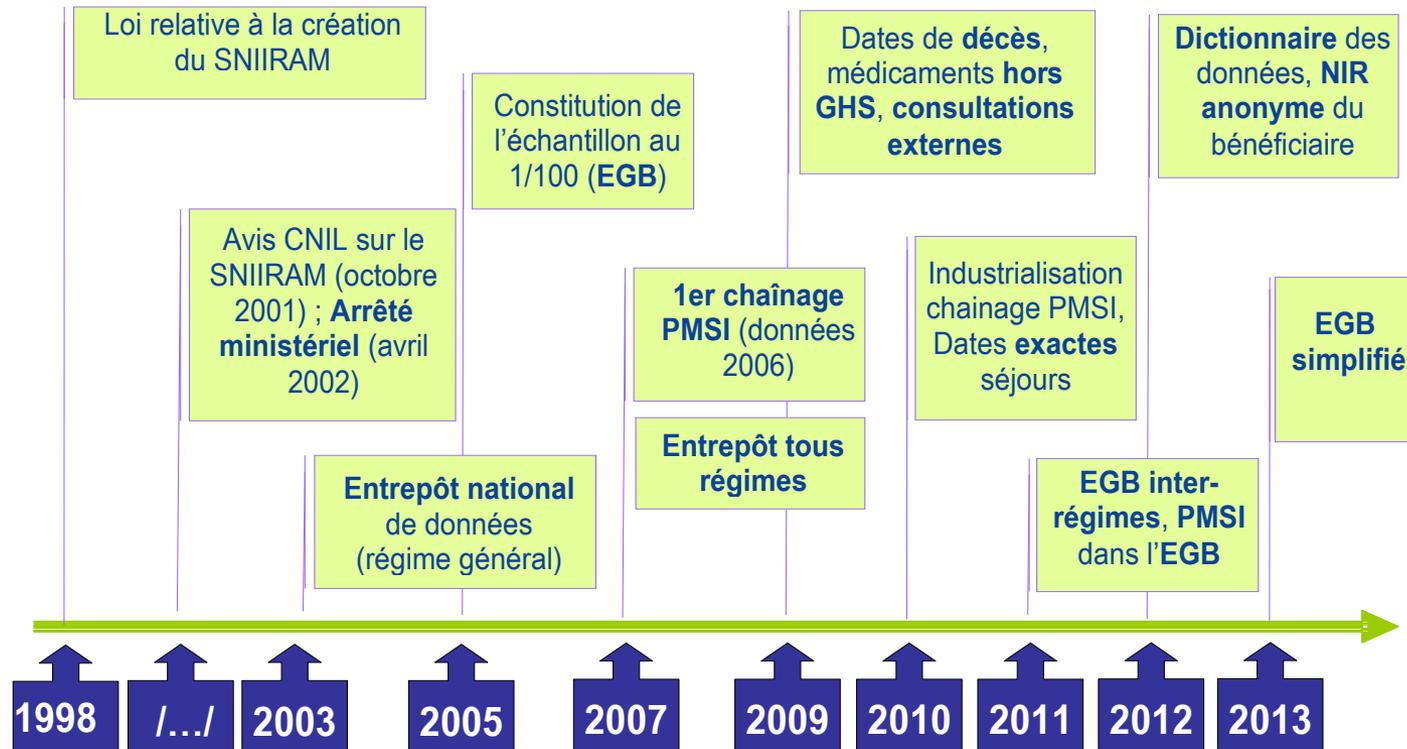
Données de santé : vue d'ensemble

- **Données médicales à visée individuelle** (pour soigner) : dossiers médicaux (compte-rendus, prescriptions, images, résultats d'examens, prélèvements, données génétiques...)
- **Données médico-administratives à visée « métier »** (pour administrer, gérer, contrôler, autoriser, payer...) : PMSI, feuilles de soins...
- **Données à visée de connaissance** : collectées directement (enquêtes, panels, cohortes, recherche biomédicale...) ou obtenues par recyclage :
 - Réutilisation (par les chercheurs) de dossiers médicaux, de registres de maladies, de biobanques...
 - Réutilisation de données administratives « déidentifiées » sous diverses formes (données granulaires, données agrégées, indicateurs etc.)
 - Par les administrations elles-mêmes (SI décisionnel ou de pilotage...)
 - Et par d'autres (chercheurs, institutions, entreprises, journalistes...)

Données médico-administratives et SNIIR-AM

- Données médico-administratives à vocation exhaustives:
 - Décomptes assurance maladie (SNIIRAM)
 - Résumés de sortie hospitaliers (PMSI)
 - Certificats de décès (CépiDc)
 - Certificats de santé des enfants...
 - Autorisations d'activité (référentiels : FINESS, RPPS...)
 - etc.
- SNIIRAM étendu:
 - Toute la population
 - Age, sexe, commune de résidence, date de décès
 - Soins de ville (y c. médicaments)
 - Soins hospitaliers
 - Arrêts de travail...
 - ALD, CMU...
 - **Chaînage (suivi longitudinal par individu)**
 - prestataires de soins

SNIIRAM-PMSI : une base récente



Une richesse sous-utilisée et un débat public

- Enjeux démocratiques, sanitaires, économiques...
- Obstacles juridiques (complexité, NIR...) et obstacles techniques
- Critiques des chercheurs
- Critiques de l'Institut des données de santé
- Polémique début 2013
- Rapport Bras sur la gouvernance et l'utilisation des données de santé (octobre 2013)
- « Commission open data » et projet de loi de santé

Utilisations du SNDS pour études et recherches

- Qualité de la prise en charge dans notre système de soins
- Impact des médicaments en vie réelle (pharmaco-épidémiologie)
- Parcours de soins
- Cartographie des dépenses par pathologie et analyse médicalisée des dépenses (diabète, cancers, dépression...)
- Etudes de patientèle ou de performance des établissements de santé, suivi de la diffusion des nouveaux traitements, production de tables de mortalité (segmentation du risque) etc.
- on n'a pas besoin d'une base exhaustive sur le plan statistique. C'est souvent vrai, mais :
 - études d'événements rares => La France peut être le principal fournisseur de données de qualité au niveau mondial
 - études régionales (disparités de prise en charge, inégalités territoriales, études de parcours...)
- A quoi s'ajoutent les études faites par appariements avec d'autres jeux de données :
 - cohortes et recueils complémentaires auprès des personnes (exemple 200 000 personnes dans CONSTANCES à terme...)
 - enquêtes (ex : enquête de la DREES auprès des personnes âgées dépendantes en 2015-2016, enquêtes santé de l'Irdes et de la DREES...)
 - - ou appariements avec autres bases administratives
 - - le système national de gestion des carrières de la CNAV : impact des professions sur la santé...)
 - - les bases de la CNAF ou de la DGFIP (impact de la composition et des revenus du ménage...)

Préconisations du rapport Bras : deux types d'accès

- Pour les données anonymes : Open data (le développer !)
 - Cas particulier des données nominatives sur l'activité des professionnels (autoriser la réutilisation des données déjà publiées)

- Pour les données indirectement nominatives : accès restreint, projets autorisés par la CNIL, traçabilité (pas de copies), réservé à la recherche et aux études d'intérêt public (demandeurs publics ou privés)
 - Cas particulier des organismes dont la mission requiert un accès permanent

Quatre chantiers à la suite du rapport Bras dans la perspective de la loi de santé

- « Commission open data », concertation avec l'appui d'Etalab : de la polémique au consensus...
- Etude technique sur les risques de réidentification dans les bases de données médico-administratives (et sur la production de jeux de données appauvris ou d'échantillons anonymes)
- Etude de cadrage (avec l'appui du SGMAP) : quels services, quels moyens ? La CNAMTS opérateur de la mise à disposition de données (traduction dans la COG)
- Travaux juridiques et projet de loi

L'ouverture des données de santé dans le projet de loi

- Création d'un système national des données de santé (SNDS = SNIIRAM + PMSI + causes de décès + données médico-sociales + données d'assurance maladie complémentaire)
- Règles d'accès et gouvernance (gestion confiée à la CNAMTS, open data et accès restreint, représentation des acteurs dans un Institut national des données de santé qui est aussi un guichet unique, pilotage par l'Etat)
- En finir avec la contrainte du décret en CE pour les appariements nécessitant le NIR (en vue d'une recherche ou étude en santé)
- Fusion des chapitres IX et X de la loi informatique et libertés : unification et simplification (autant que possible)

Données personnelles, données « déidentifiées » ou « pseudonymisées », données anonymes

- Feuilles de soins et dossiers médicaux sont évidemment nominatifs à l'origine
- Risque pour la vie privée... Données de santé (= maladies) = données sensibles
- Les données du SNIIRAM-PMSI sont « dé-identifiées » (pas de nom + NIR chiffré de manière irréversible) mais pas forcément anonymes :
 - Si les données sont précises et exhaustives, on y reconnaît des personnes sur lesquelles on dispose d'informations par ailleurs (proches, célébrités, employés, clients...)
 - Si par ex. on connaît hôpital, mois de sortie, durée du séjour, date (m/a) de naissance, sexe, code géographique, alors 89% des hospitalisés sont identifiables dans la base PMSI (Blum D., Trouessin G., 2011) et 100 % des personnes hospitalisées plus d'une fois dans l'année ou décédées à l'hôpital
- Les données du PMSI et les extractions du SNIIRAM sont aujourd'hui en accès restreint mais peu contrôlé et sans traçabilité (ni des accès ni des requêtes) :
 - Plus il y a de personnes, dans le secret, moins il y a de secret...
- Mesure du risque (cf. *Rapport sur les risques de réidentification*) :
 - Nombre de personnes présentant des caractéristiques uniques dans la base (dénombrement)
 - Notoriété des variables (classement) : date de naissance ou de décès > lieu et dates d'hospitalisation > date de consultation
 - Efficacité des mesures de protection ou de compensation, probabilité des attaques (évaluation plus subjective)

Risque de réidentification : comment s'en prémunir ?

- La loi protège
- La traçabilité aussi...
- Donc pas d'open data pour les données *personnelles* de santé...

Ce n'est pas un prétexte : droit à la vie privée...

- L'objectif c'est l'ouverture la plus large possible (maximum de données, maximum d'utilisations) compatible avec la protection de la vie privée

Gouvernance : quatre questions

- Où se discutent les orientations, les résultats (doctrine, critiques, demandes, évaluation...) : un nouvel Institut des données de santé
- Qui donne l'autorisation d'accès à des données personnelles ?
 - CNIL (ou « comité du secret » pour la statistique publique et les données fiscales)
 - avis en amont : comités de protection des personnes (les CPP et leur comité national d'harmonisation), le CCTIRS (comité d'expertise scientifique), le futur Institut national des données de santé (?)
 - Comment les porteurs de projet s'y retrouvent-ils ?
- Qui gère concrètement les données ? CNAMTS, ATIH... (un rôle pour le CASD ?)
- Pilotage par un Etat plus présent et plus attentif aux demandes

- A lire** (site de la DREES ou Documentation française...):
- *Rapport sur la gouvernance et l'utilisation des données de santé (« rapport Bras »)*
 - *Rapport de la Commission open data en santé (y c., en annexe 9, rapport sur les risques de réidentification)*
 - *Rapport du Conseil d'Etat 2014 sur le numérique et les droits fondamentaux*
 - *article 47 du projet de loi sur la santé (et son exposé des motifs)*

Merci de votre attention