

# EXEC SUM

## **OUVRIRE PUIS REFERMER : LES DYNAMIQUES D'ACCÈS AUX DONNÉES DE SANTÉ EN FRANCE DEPUIS LES ANNÉES 1990**

Chloé Bérut, postdoctorante, Chaire PARI

CHAIRE  
**PARI**

PROGRAMME DE RECHERCHE  
SUR L'APPRÉHENSION DES RISQUES  
ET DES INCERTITUDES

## Résumé

Aujourd'hui, l'exploitation des données numériques liées à la santé est un enjeu crucial de la scène politique, comme en témoigne la récente création du 'Health Data Hub' en France ainsi que les travaux européens menés sur l'«Espace européen des données de santé». Si l'intérêt pour ce type de données s'est accru au cours de la dernière décennie, notamment à la faveur de l'émergence des techniques de Big Data et d'Intelligence Artificielle, la création des bases de données de santé en France remonte en réalité aux années 1990. Aussi les règles et dispositifs actuels de régulation de l'accès aux données de santé se sont-ils construits à partir d'une série d'évolutions politiques que cet Executive Summary se propose de résumer (voir WP #30 et WP #31 pour une information détaillée). À partir des résultats d'une enquête qualitative menée en 2022, ce document revient sur les différentes séquences de la régulation de l'accès aux données de santé en France. Ce faisant, une analyse de la création des bases de données et des dynamiques d'ouverture opérées au cours des années 2000 est proposée, ainsi qu'un compte-rendu de la restructuration plus récente de l'accès aux données de santé suite à la réforme de 2016, qui marque une dynamique de fermeture de l'accès pour de nombreux acteurs non-étatiques.

OUVRIR PUIS REFERMER :  
LES DYNAMIQUES D'ACCÈS AUX DONNÉES  
DE SANTÉ EN FRANCE DEPUIS LES ANNÉES 1990

### L'enquête menée par la chaire PARI sur la régulation de l'accès aux données de santé

Afin de comprendre les enjeux actuels de la régulation de l'accès aux données de santé, la chaire PARI a financé une recherche d'un an et demi dans le cadre d'un contrat postdoctoral. Cette recherche en science politique avait pour objectif de retracer et d'expliquer les différentes séquences de l'ouverture et de la fermeture de l'accès aux données de santé en France depuis la création des bases données de santé centralisées dans les années 1990.

Le matériau empirique de cette enquête qualitative a principalement été récolté entre janvier et septembre 2022. La collecte de données a mobilisé les stratégies classiques de l'analyse qualitative des politiques publiques : récolte et analyse documentaire (lois, rapports, documents de politique publique, articles de presse), et réalisation d'entretiens semi-directifs avec une vingtaine d'acteurs de la politique de régulation de l'accès aux données de santé (Ministère de la Santé, de la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés — CNAMTS, cabinets ministériels, agences spécialisées dans la régulation de l'accès aux données de santé). Les questions visaient à comprendre les représentations des acteurs en matière de données de santé (bénéfices et/ou risques associés à l'ouverture, positionnement), mais également à recueillir des récits relatifs au déroulé des événements sur une période donnée.

EXEC  
SUM

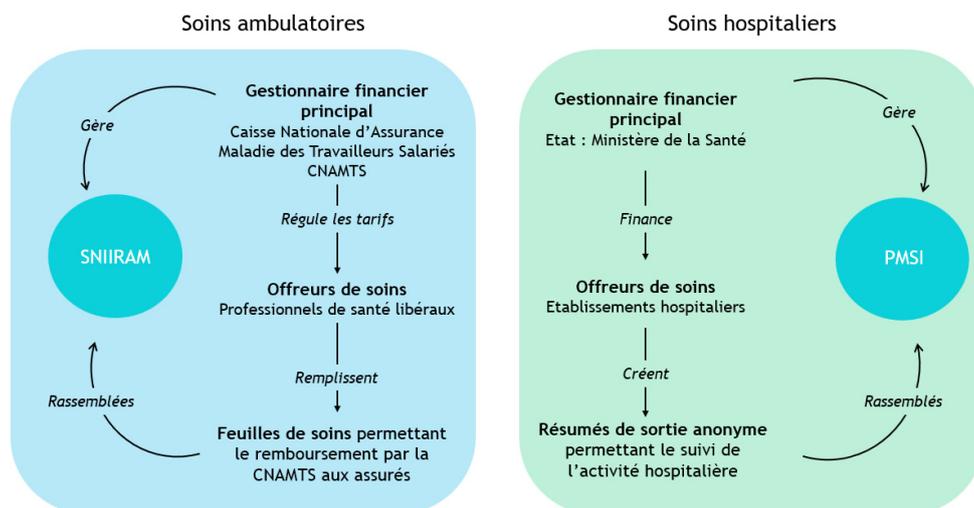
## L'ORGANISATION DU SYSTÈME DE SANTÉ ET LA CONSTRUCTION DES BASES DE DONNÉES DANS LES ANNÉES 1990

### Des bases de données qui reflètent le caractère dual de l'organisation des soins

Depuis 1945, le système de santé français est majoritairement constitué par deux pôles qui chacun assurent le fonctionnement d'une partie différente du système de soins : la Caisse nationale d'assurance maladie (CNAMTS) pour les soins ambulatoires (c'est-à-dire réalisés en dehors de l'hôpital directement au sein des cabinets des professionnels de santé), et l'État pour les soins hospitaliers. Côté soins ambulatoires, la CNAMTS gère le financement du système via la récolte des cotisations sociales des travailleurs salariés, le remboursement des assurés, et les négociations avec les professionnels libéraux pour la fixation des tarifs. Du côté du secteur hospitalier, c'est le Ministère de la Santé qui finance directement les établissements hospitaliers publics.

Cette dualité organisationnelle se retrouve dans la construction des bases de données de santé qui interviennent dans les années 1990. Si, dans les deux cas, il s'agit d'assurer un meilleur pilotage financier du système de santé, les deux bases de données de santé qui sont créées sont gérées par des organismes différents et ne sont pas structurées à partir du même type de données. Ainsi, la CNAMTS crée à la fin des années 1990 le Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM), qui rassemble les données issues des feuilles de soins, tandis que le Ministère de la Santé met en place le Programme de Médicalisation des Systèmes d'Information (PMSI) qui rassemble les données issues de l'activité hospitalière.

### L'organisation du système de santé français et les bases de données de santé



### Des bases conçues pour la gestion financière du système de santé

#### Le SNIIRAM

Dès 1977, la CNAMTS crée une base de données de santé centralisée appelée le Système National Inter-régime (SNIR) dans le but d'agréger les données associées aux remboursements effectués dans le cadre de la pratique libérale. À cette époque cependant, le codage ne permet pas d'avoir une vue détaillée des actes réalisés par les médecins. Ce n'est qu'à partir des années 1990 que la captation d'informations devient suffisamment fine pour constituer des bases de données riches et exploitables. À partir de cette période, l'ensemble des éléments présents sur une feuille de soins est codé afin de donner une information précise sur les actes médicaux pratiqués, ainsi que sur les spécialités pharmaceutiques prescrites et remboursées. La centralisation des bases de données des différents régimes de l'assurance maladie dans un unique entrepôt, le Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM), intervient à la fin des années 1990 avec l'adoption des Ordonnances Juppé, dans un contexte où les pouvoirs

publics cherchent à mettre en place une meilleure maîtrise des dépenses. Depuis, la base de données originelle du SNIIRAM demeure gérée par la CNAMTS.

#### Le PMSI

Le Programme de Médicalisation des Systèmes d'Information (PMSI) est une base de données mise en place dans le cadre de la refonte du financement des hôpitaux. Dans le système hospitalier d'après-guerre, il n'existe en effet pas de données fiables relatives à l'activité hospitalière, ce qui rend le pilotage du financement des hôpitaux particulièrement délicat. Initialement présenté comme un simple enrichissement des données disponibles sur le système hospitalier en 1983, le PMSI est ensuite généralisé suite à l'obligation pour les établissements hospitaliers de procéder à l'évaluation et l'analyse de leur activité en 1991. Aujourd'hui, ce sont les données du PMSI qui sont utilisées dans le cadre de la tarification à l'activité (T2A) des hôpitaux. Le PMSI contient des données produites à partir d'un document administratif appelé 'Résumé de sortie standardisé', qui, une fois, anonymisé (c'est-à-dire dépouillé de ses informations directement identifiantes : nom, numéro de sécurité

sociale...), est rebaptisé 'résumé de sortie anonyme' (RSA). Les RSA contiennent les informations médicales et administratives des patients de passage dans un service hospitalier. Les diagnostics sont classés à l'aide de la Classification internationale des maladies (CIM) 10, une classification produite par l'Organisation Mondiale de la Santé. Les actes médicaux sont quant à eux référencés à partir de la Classification commune des actes médicaux (CCAM). Ces données permettent ensuite de classer les patients par groupe homogène de malades (GHM), qui, à partir de données administratives (comme l'âge) et des diagnostics permettent de déterminer une rémunération forfaitaire pour la prise en charge par les hôpitaux. Dès 2000, la gestion de la base et de ses accès est réalisée par une agence publique qui rassemble les anciennes missions ministérielles dédiées à la mise en place du PMSI : l'Agence technique de l'information sur l'hospitalisation (ATI).

### Le risque de réidentification et la mise en place de l'autorisation CNIL

Lors de la création du SNIIRAM et du PMSI, la question de l'accès et de l'usage des données de santé par des acteurs tiers n'est pas envisagée. Il faut attendre les premières requêtes d'accès aux PMSI par des journalistes du magazine Sciences et Avenir en 1998 pour que se pose la question de la régulation de l'accès à ces bases. Les journalistes souhaitent alors utiliser le PMSI pour réaliser le premier 'Palmarès des hôpitaux'. La publication de ce palmarès pousse le législateur à adopter des mécanismes de protection de ces données, qui, bien qu'anonymisées, demeurent néanmoins sensibles. En effet, les données du PMSI et du SNIIRAM, très détaillées, peuvent donner lieu à des pratiques de réidentification dès lors qu'un acteur tiers possède déjà quelques informations sur une personne (date de naissance, date des soins et lieu d'hospitalisation par exemple). Ainsi, à partir de 1998, l'accès aux données du PMSI ou du SNIIRAM devient conditionné par l'obtention d'une autorisation préalable de la Commission Nationale de l'Informatique et des Liberté (CNIL).

## UNE OUVERTURE PROGRESSIVE AU COURS DES ANNÉES 2000 AVEC UNE FORTE DIFFÉRENCIATION DES RÈGLES D'ACCÈS

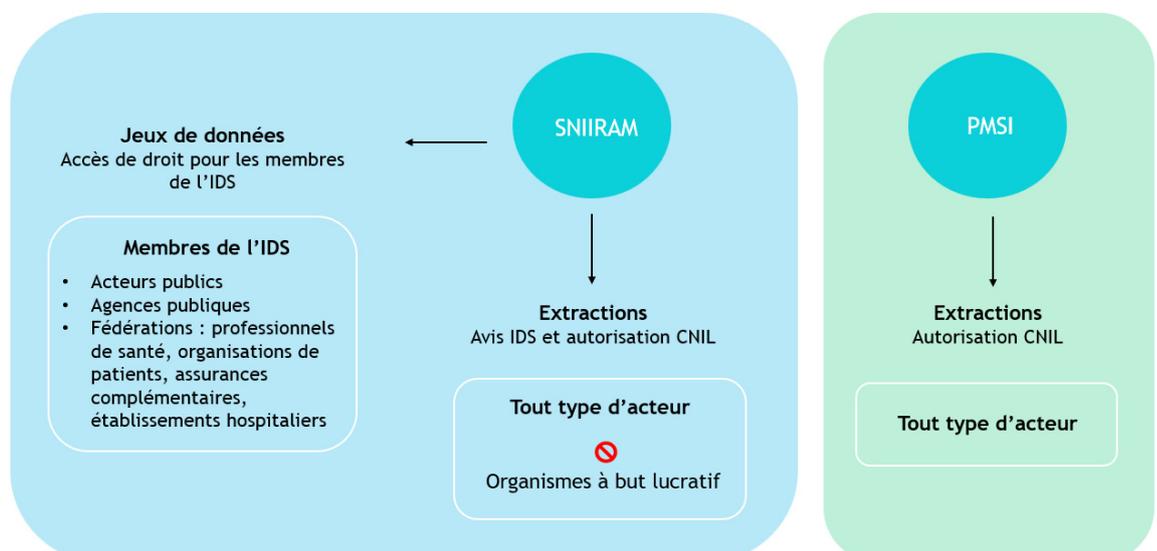
### La différenciation des règles d'accès entre le SNIIRAM et le PMSI

Au cours des années 2000, les règles d'accès aux données de santé vont connaître une double différenciation : tandis que le SNIIRAM se voit doté de règles spécifiques par rapport au PMSI (dont les règles d'accès demeurent stables avec une autorisation de la CNIL), son accès demeure très hétérogène selon le type d'acteur considéré. Une des nouvelles règles d'accès au SNIIRAM concerne par exemple l'interdiction aux organismes à but lucratif, afin de prévenir tout risque de mésusage des données de santé (réidentification, utilisation à des fins commerciales). Cette différenciation se poursuit avec la création de l'Institut des données de Santé (IDS) en 2004. Cet institut, créé par la loi du 13 août 2004 relative à l'Assurance Maladie, fait partie d'un paquet plus large de réformes visant à renforcer le rôle des assureurs complémentaires dans la gestion du risque maladie. L'IDS est alors conçu comme un moyen de faciliter cette 'cogestion' du risque maladie : la mission principale de l'institut est d'assurer le partage des données de santé entre les différents acteurs du système de santé : État, assureurs publics et privés (mutuelles), professionnels de santé.

### L'Institut des données de santé et l'ouverture progressive des données du SNIIRAM

À partir du milieu des années 2000, l'IDS met à disposition de ses membres tout un ensemble d'outils permettant d'accéder aux données du SNIIRAM dans un format principalement agrégé ou sous forme d'échantillon (l'Echantillon Généraliste de Bénéficiaires). Ces formats permettent de limiter les risques de réidentification des citoyens. Dans ce contexte, les acteurs membres de l'IDS bénéficient d'un accès facilité aux données de santé, qui sont à l'occasion retravaillées par l'IDS dans le cadre de la publication de 'tableaux de bords' thématiques. Les membres de l'IDS sont principalement des organisations représentant les divers acteurs du système de santé : acteurs et agences publiques, professionnels de santé, organisations de patients, fédérations d'assureurs complémentaires, fédérations hospitalières.

### Les règles d'accès au SNIIRAM et au PMSI dans les années 2000



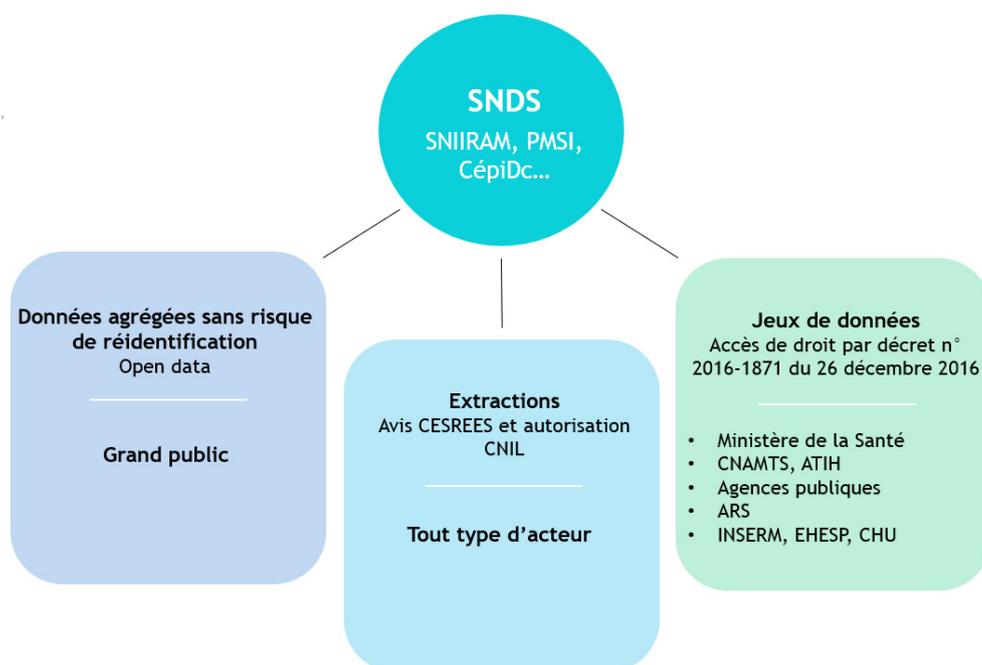
## LA REFONTE DU SYSTÈME D'ACCÈS AUX DONNÉES DE SANTÉ DEPUIS 2016

### Une harmonisation des règles et des bases de données pour protéger des mésusages des données de santé

À la fin des années 2000 et au début des années 2010, différents événements contribuent à remettre l'enjeu de l'accès aux données de santé sur l'agenda politique. C'est notamment le cas de l'affaire Mediator, qui souligne à la fois le potentiel et le manque d'exploitation des bases de données de santé pour la santé publique. C'est en effet grâce aux données du SNIIRAM qu'Irène Frachon, pneumologue au CHU de Brest, démontre les effets néfastes (et parfois mortels) du Mediator, spécialité pharmaceutique pourtant commercialisée depuis plusieurs décennies par les laboratoires Servier. En parallèle, de nombreux acteurs de la société civile se mobilisent en 2013 au sein de l'initiative 'Libérez nos données', qui rassemble représentants d'usagers du système de santé, chercheurs, assureurs, entrepreneurs, consultants et journalistes. Ces acteurs revendiquent une plus grande ouverture de l'accès aux données de santé, considérée comme une véritable problématique de transparence et de démocratie sanitaire.

Une refonte du système d'accès aux données de santé est alors entreprise par Marisol Touraine, Ministre de la Santé de 2012 à 2017. Cette refonte conduit à harmoniser les règles d'accès des différentes bases pour les différents acteurs. Ainsi, le PMSI et le SNIIRAM sont rassemblés au sein d'un même système appelé 'Système National des Données de Santé' (SNDS). Ce système a vocation à rassembler d'autres bases de données en lien avec la santé comme le CépiDc de l'INSERM, qui rassemble les données relatives aux causes médicales de décès. Avec la création du SNDS, les règles d'accès s'harmonisent et deviennent similaires pour l'ensemble des bases de données. Ces transformations traduisent une volonté de reprise en main par l'État, qui souhaite rationaliser un processus d'accès aux données de santé jusqu'alors très fragmenté. En parallèle, le législateur se montre soucieux de protéger les données de santé des possibles mésusages dont elles pourraient faire l'objet (utilisation à des fins commerciales par les laboratoires pharmaceutiques ou pour des pratiques de sélection du risque par les assureurs).

### Les règles d'accès au Système National des Données de Santé à partir de 2016



### Des dynamiques différenciées selon les acteurs

Désormais, l'ensemble des acteurs (y compris les organismes à but lucratif) doit se soumettre à une même procédure pour accéder aux données de santé. Cette procédure inclut l'obtention de l'avis d'un comité scientifique (le Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé ■ CESREES) et l'autorisation de la CNIL. Les accès de droit dont bénéficiaient les membres de l'IDS sont ainsi révoqués, sauf pour les acteurs de la sphère publique qui sont désormais plus nombreux à pouvoir accéder au SNDS sans passer par une autorisation CNIL. En parallèle, les données agrégées sans risque de réidentification sont supposées faire l'objet d'une

mise en ligne pour le grand public dans le cadre d'une politique d'*open data*.

*In fine*, cette refonte se traduit par des dynamiques ambivalentes selon le type d'acteur considéré. Pour les acteurs de la sphère publique, la réforme de 2016 constitue une nette ouverture. À l'inverse, les anciens membres de l'IDS ont vu leurs prérogatives d'accès limitées, puisqu'ils doivent désormais passer par une procédure classique d'autorisation d'accès (CESREES et CNIL). Paradoxalement, les organismes à but lucratif (comme les laboratoires pharmaceutiques) qui ne faisaient pas partie de l'IDS bénéficient de ces nouvelles mesures, puisqu'ils peuvent désormais accéder au SNIIRAM sans restriction liée à leur statut.

### Dynamiques d'ouverture et de fermeture

de l'accès aux données de santé induites par la réforme de 2016

Ministères et services publics	●
Agences publiques	●
Acteurs de la recherche en santé	●
Professionnels de santé	●
Organisations de patients	●
Assureurs complémentaires	●
Laboratoires pharmaceutiques	●

Dynamique d'ouverture ●

Dynamique de fermeture ●

### Le Health Data Hub et la volonté d'industrialiser l'usage des données de santé

Après les élections de 2017, le secteur des données de santé connaît de nouvelles transformations avec la création du 'Health Data Hub' (ou plateforme des données de santé) en 2019. Cette plateforme vise à faciliter l'usage des données de santé à grande échelle. Il s'agit de rassembler le plus grand nombre possible de bases de données de santé et d'en stocker une copie directement au HDH afin d'accélérer les procédures d'appariement (chaînage des bases de données entre elles) et d'extraction. Ces transformations n'impliquent pas de changement en termes de règles d'accès, mais témoignent de la volonté d'industrialiser l'usage des données de santé dans un contexte politique qui favorise les projets en lien avec le développement de l'Intelligence Artificielle. La mise en œuvre du HDH rencontre cependant de nombreuses difficultés et est aujourd'hui toujours en cours.

Dès 2020, cette plateforme fait notamment face à d'importantes polémiques vis-à-vis du choix de l'entreprise américaine Microsoft-Azure comme hébergeur des copies des bases de données de santé. La question de la nationalité n'est ici pas anodine : en effet, selon l'origine de l'entreprise, la réglementation en matière de protection des données personnelles peut largement différer. C'est d'ailleurs pourquoi la Commission européenne établit des accords avec des pays tiers afin d'assurer un niveau de protection similaire à celui de l'Union européenne (UE) en cas de transfert de données. Dans le cas de l'entreprise Microsoft-Azure, le problème vient du fait que les données peuvent potentiellement se retrouver sous juridiction américaine. Or, les accords relatifs à la protection des données personnelles entre l'UE et les États-Unis sont particulièrement instables et ont été annulés plusieurs fois par la Cour de Justice de l'UE, qui considère qu'ils ne permettent pas d'assurer un niveau de protection suffisant des droits fondamentaux des citoyens. Le choix d'un hébergeur américain a ainsi créé d'importantes polémiques parmi les acteurs du système de santé. Avec l'approche des élections présidentielles de 2022, le dossier est officiellement mis en attente. Fin 2022 cependant, la directrice du HDH annonce que l'hébergement demeurerait chez Microsoft-Azure au moins jusqu'en 2025, le temps de développer une 'solution souveraine'.

## CONCLUSION

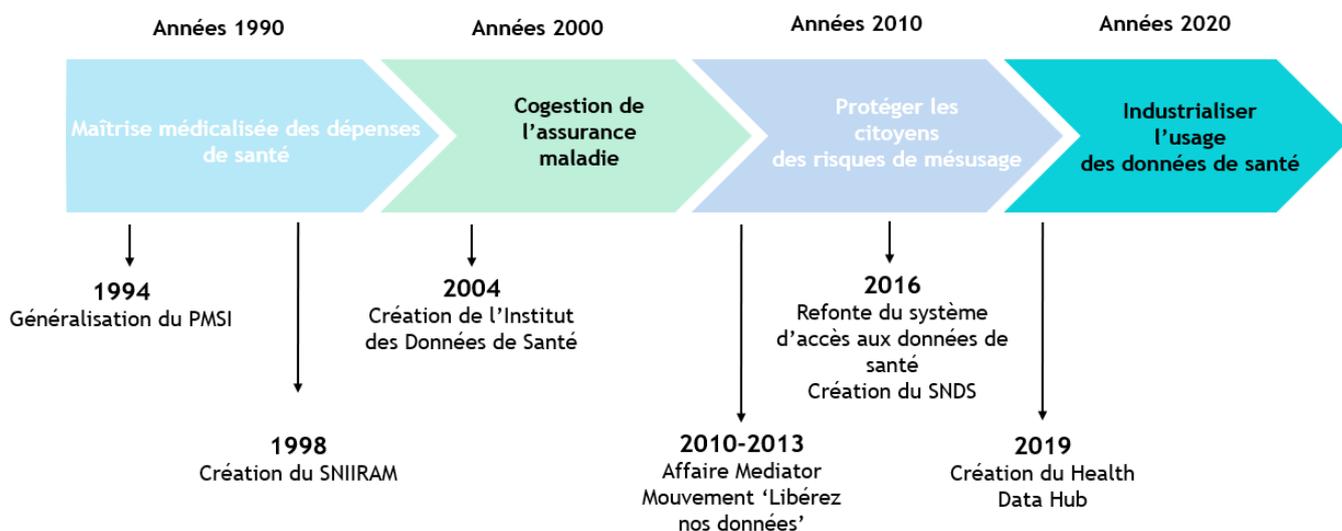
Cette recherche permet d'identifier trois grandes périodes dans la régulation de l'accès aux bases de données de santé. La première est relative à la mise en place des bases de données de santé dans les années 1990 dans un contexte d'amélioration du pilotage financier du système de santé. La création de deux bases de données distinctes (le SNIIRAM et le PMSI) reflète alors l'organisation duale du système de santé entre la caisse nationale d'assurance maladie et l'État. Les usages des données de santé par des acteurs tiers ne sont pas envisagés dans un premier temps, et il faut attendre 1998 pour qu'une première couche de régulation se mette en place avec l'introduction d'une autorisation de la CNIL.

Au début des années 2000, une réforme plus globale de la gestion du système de santé donne lieu à la création de l'Institut des Données de Santé, (IDS), qui a vocation à assurer le partage des données de santé entre les principaux acteurs de la 'cogestion' du système de santé : État, CNAMTS, professionnels de santé mais également assureurs complémentaires. Dans le cadre des activités de l'IDS, certains acteurs non-étatiques comme les fédérations d'assureurs complémentaires ou les organisations de patients bénéficient d'un accès facilité aux données du SNIIRAM, tandis que les organismes à but lucratif sont exclus d'un tel accès. En parallèle, l'accès au PMSI est possible pour n'importe quel type d'acteur à partir du moment où celui-ci obtient une autorisation de la CNIL.

Suite à différents événements qui mettent en lumière la nécessité d'une exploitation plus ouverte des données de santé (affaire Mediator et initiative 'Libérez nos données'), la question de l'accès aux données de santé revient sur l'agenda politique au début des années 2010. Le système d'accès aux données fait alors l'objet d'une refonte qui harmonise l'accès aux différentes bases de données de santé. Ces réformes témoignent d'une volonté de reprise en main par l'État, qui souhaite rationaliser l'accès aux bases de données de santé et protéger les citoyens des potentiels mésusages de ces données. La réforme de 2016 se traduit par des dynamiques variées selon les acteurs considérés : alors que de nombreux acteurs publics se voient octroyer un accès facilité aux données de santé dédormais rassemblées dans le Système National des Données de Santé, d'autres (et en particulier les anciens membres de l'IDS) voient leurs possibilités d'accès se limiter.

Enfin, les transformations récentes liées à la création du Health Data Hub sont avant tout d'ordre technique : il s'agit en effet de créer les conditions de l'industrialisation de l'usage des données de santé, dans un contexte marqué par les problématiques liées au développement de l'Intelligence artificielle.

Les règles d'accès au Système National des Données de Santé à partir de 2016



# CHAIRE PARI

PROGRAMME DE RECHERCHE SUR L'APPRÉHENSION DES RISQUES ET DES INCERTITUDES

Retrouvez l'intégralité de l'article sur notre site, [www.chaire-pari.fr](http://www.chaire-pari.fr) rubrique Publications > Working papers

PARI, placé sous l'égide de la Fondation Institut Europlace de Finance en partenariat avec l'ENSAE/Excess et Sciences Po, a une double mission de recherche et de diffusion de connaissances.

Elle s'intéresse aux évolutions du secteur de l'assurance qui fait face à une série de ruptures : financière, réglementaire, technologique. Dans ce nouvel environnement, nos anciens outils d'appréhension des risques seront bientôt obsolètes. PARI a ainsi pour objectifs d'identifier leur champ de pertinence et de comprendre leur émergence et leur utilisation.

**L'impact de ses travaux se concentre sur trois champs :**

- les politiques de régulation prudentielle dans un contexte où Solvabilité 2 bouleverse les mesures de solvabilité et de rentabilité (fin du premier cycle de la chaire);
- les solutions d'assurance, à l'heure où le big data déplace l'assureur vers un rôle préventif, créant des attentes de personnalisation des tarifs et de conseil individualisé ;
- les technologies de data science appliquées à l'assurance, modifiant la conception, l'appréhension et la gestion des risques.

Dans ce cadre, la chaire PARI bénéficie de ressources apportées par Addactis, la CCR, Generali, Groupama, la MGEN et Thémis.

Elle est co-portée par **Pierre François**, chercheur au CNRS, doyen de l'Ecole Doctorale de Sciences Po et **Laurence Barry**, chercheur à Datastorm, la filiale de valorisation de la recherche de l'ENSAE.

**PARTENAIRES**

