

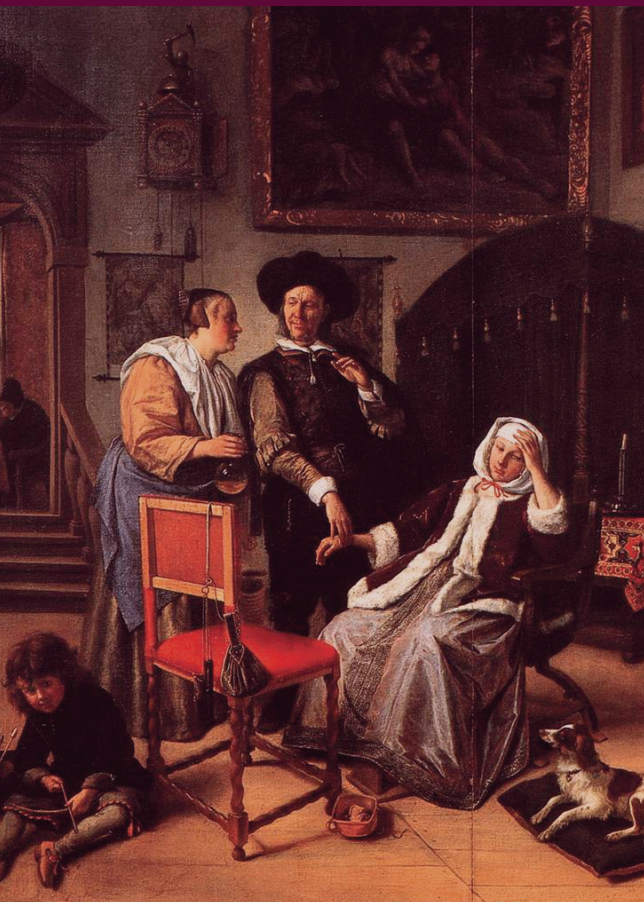
Working paper

31

L'accès aux données de santé en France : ouverture en trompe l'oeil et industrialisation de l'usage des données (période 2010-2020)

Chloé BÉRUT, chaire PARI

Mars 2023



PARI

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

L'accès aux données de santé en France : ouverture en trompe l'œil et industrialisation de l'usage des données (période 2010-2020)

Chloé Bérut, chaire PARI, chercheuse associée au laboratoire Pacte
chloe.berut@umrpacte.fr

Résumé

Ce second Working paper a pour objectif d'analyser les transformations de la régulation de l'accès aux données de santé qui s'opèrent à partir du début des années 2010. Le début de cette décennie est en effet marqué par une forme de politisation de l'enjeu des données de santé, qui intéresse un nombre toujours plus grand d'acteurs et gagne en visibilité dans l'espace public. Suite à plusieurs 'affaires' (et notamment l'affaire Mediator des laboratoires Servier), l'État entame un travail de formulation d'une doctrine plus claire en matière d'accès aux données de santé. À l'issue de ce processus, un article sur les données de santé est adopté dans la loi de modernisation de notre système de santé de 2016. Cet article uniformise les procédures d'accès aux différentes bases de données, et renforce les conditions d'accès aux données de santé pour de nombreux acteurs du secteur privé. En parallèle, il facilite l'accès aux données de santé pour les acteurs du secteur public. Depuis la loi de 2016, l'essentiel des transformations opérées dans ce secteur sont d'ordre technique, avec la création d'une 'plateforme des données de santé' (Health Data Hub) qui vise à permettre un usage industriel des données de santé.

Sommaire

Introduction.....	2
I. ‘Ouvrir les données en commençant par les refermer’, ou la reprise en main du politique (début et milieu des années 2010).....	7
a) La mise sur agenda de l’enjeu des données de santé	8
L’affaire du Mediator®	8
De multiples ‘affaires’ liées à l’usage des données de santé.....	11
L’initiative ‘Libérez nos données’ et la pression médiatique.....	12
Le rôle de la communauté scientifique.....	13
b) Le processus de formulation de l’article 193 de la loi de modernisation de notre système de santé (2016).....	14
Ouvrir ou sécuriser les données de santé : le cadrage ambigu des travaux préparatoires	14
Un accord entre la CNAMTS, l’État, et la communauté scientifique pour sécuriser l’accès aux données de santé	16
c) Les transformations de l’accès aux données de santé suite à la loi de 2016	19
Les règles d’accès : restriction d’accès pour les acteurs ‘privés’, extension des accès de droit pour les acteurs publics.....	20
La technicisation de la gouvernance opérationnelle.....	25
La gouvernance stratégique : reprise en main par l’État.....	27
Les enjeux pratiques : des dynamiques ambivalentes.....	28
II. Vers une industrialisation de l’usage des données de santé ? (fin des années 2010, début des années 2020).....	31
a) L’enjeu de l’intelligence artificielle : une nouvelle fenêtre d’opportunité pour les données de santé.....	31
Du Big Data à l’Intelligence artificielle, une continuité dans le cadrage politique des données de santé.....	31
De l’innovation de garage’ à la DREES à la création du Health Data Hub.....	31
Les acteurs ‘privés’ dans la création du Health Data Hub	35
b) Le Health Data Hub : la montée en puissance des ‘enjeux pratiques’ de l’accès aux données de santé.....	36
Industrialiser l’usage des données de santé via une plateforme technique.....	36
La transformation de l’IDS en plateforme technique : l’éradication définitive des ambiguïtés institutionnelles des années 2000	37
‘L’affaire Microsoft-Azure’	38
Conclusion	42
Bibliographie.....	44
Annexe : Liste anonymisée des entretiens réalisés.....	45

Introduction

L'objectif de ce document de travail est de présenter les premiers résultats d'une enquête menée en 2022-2023 sur la régulation de l'accès aux données de santé en France. Plus précisément, il s'agit d'interroger la construction de cette régulation, en particulier en ce qui concerne l'accès aux bases de données de santé centralisées créées dans le cadre du remboursement de soins et du pilotage financier de l'Assurance Maladie et des hôpitaux. Ainsi, les deux grandes bases centrales analysées dans cette recherche sont le Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM), qui rassemble les données pseudonymisées des remboursements des prestations aux assurés, et le Programme de médicalisation des systèmes d'information (PMSI) qui est constitué par les données relatives aux séjours en hôpital. Depuis 2016, ces deux bases sont rassemblées (et appareillées) au sein du Système National des Données de Santé (SNDS), qui contient également les données sur les causes médicales de décès (CépiDc).

Ces bases de données constituent une source d'informations très importante pour de nombreux acteurs, bien que les objectifs d'origine de ces bases soient essentiellement de nature gestionnaire. Avec le processus de digitalisation de la santé et le développement de l'intelligence artificielle, la question de l'usage secondaire¹ des données de santé devient de plus en plus prégnante. Parfois qualifiées 'd'or noir'² du XXIème siècle, les données de santé peuvent être utilisées à de nombreuses fins : améliorer le parcours du patient, aider à la formulation des politiques publiques, ou encore réaliser des recherches scientifiques, notamment en santé publique (Skovgaard *et al.* 2019). Pour certains acteurs privés, ces données peuvent présenter un intérêt commercial non négligeable.

En parallèle, la crise de la COVID-19 a contribué à renforcer l'attrait pour les données de santé, désormais perçues comme un outil crucial de lutte contre les épidémies. Ainsi, les plateformes nationales d'accès aux données de santé, pour la plupart mises en place très récemment, ont toutes 'accélééré les procédures d'accès aux données pour les recherches relatives à la COVID-19'³. Au niveau européen également, les acteurs s'organisent pour rendre les données des États membres compatibles et comparables ('interopérables') dans un 'Espace européen des données de santé' (*European health data space*, EHDS), dont l'objectif serait, entre autres, d'améliorer la réponse européenne en cas de crise sanitaire.⁴

¹ L'usage secondaire des données correspond à un traitement des données différent de celui qui était envisagé au moment de la récolte des données. Par exemple, les données de la sécurité sociale sont récoltées à des fins administratives et de remboursement, et leur usage dans le cadre d'une recherche en santé publique correspond à un usage secondaire.

² Lors des entretiens menés pendant cette recherche, le vocabulaire lié à l'or est très souvent utilisé pour désigner les données de santé : 'trésor', 'mine d'or', 'coffre-fort'...

³ Health Data Hub (2021). *Les structures d'accès et de gouvernance des données de santé. Restitution du benchmark international 2020*. Paris : publications du Health Data Hub, p.34

⁴ Martins Henrique (2021). *EU health data centre and a common data strategy for public health*. Bruxelles : publications du Parlement européen.

Au vu des outils technologiques dont nous disposons (ou commençons à disposer) aujourd'hui, la question de l'accès à ces bases de données devient donc un enjeu majeur sur la scène politique, et c'est ainsi souvent avec le prisme de l'ouverture ou de la fermeture de cet accès que nous avons tendance à aborder le sujet de la régulation des bases de données de santé. Cela étant posé, répondre de manière simple à cette question (les bases de données sont-elles plutôt 'ouvertes' ou 'fermées' ?) n'a rien d'évident, et ce pour différentes raisons. En premier lieu, et comme nous le verrons dans ce document de travail, l'accès à ces bases de données de santé va être conditionné par un ensemble d'éléments, bien plus complexe que la simple question des règles d'accès. C'est ainsi tout un ensemble de 'blocs' interagissant les uns avec les autres, qui vont participer à la création de conditions d'ouverture ou de fermeture de l'accès aux bases de données de santé : les règles d'accès, mais également la nature de la gouvernance opérationnelle et de la gouvernance stratégique, ainsi d'autres d'enjeux plus 'techniques' ou 'pratiques' :

- **Les règles d'accès** : Les règles d'accès constituent sans doute l'aspect le plus visible de l'accès aux données de santé pour un observateur extérieur. Il est ainsi généralement possible de se faire une première idée des dynamiques d'ouverture et de fermeture d'une politique d'accès aux données en analysant la palette d'acteurs pouvant prétendre y avoir accès, et en observant les conditions dans lesquelles cet accès est susceptible d'être réalisé (objectifs à atteindre, procédure à suivre, moyens à mettre en œuvre etc.). En plus des règles établies sur un plan général, il s'agit également de s'intéresser à l'ensemble des règles d'exceptions qui sont très nombreuses dans ce secteur. Ces règles peuvent soit donner des accès différenciés selon les jeux de données considérés (on peut par exemple, créer plusieurs jeux de données thématiques de données à partir du SNIIRAM, au sein desquels les données peuvent être plus ou moins agrégées), soit établir des procédures d'accès spécifiques à certains acteurs. Aussi les règles d'accès, qui peuvent sembler relativement simples à appréhender aux premiers abords, sont-elles en réalité complexifiées du fait de l'existence d'une multitude d'exceptions.
- **La gouvernance opérationnelle** : Un autre aspect relativement visible de la gouvernance de l'accès aux données de santé pour un observateur externe est relatif à la mise en œuvre opérationnelle de la régulation. Cette gouvernance opérationnelle désigne les acteurs et organisations en charge de mettre en œuvre les règles d'accès adoptées au niveau national, c'est-à-dire dans le cas des données de santé le ou les organisme(s) participant à la délivrance d'autorisations d'accès aux données du SNIIRAM, du PMSI, ou à partir de 2016 du Système National des Données de Santé (SNDS).
- **La gouvernance stratégique** : Un autre type d'instrument central pour comprendre la régulation de l'accès aux données de santé a trait à la 'gouvernance stratégique' de ce secteur. Par 'gouvernance stratégique', nous entendons les dispositifs visant à donner des orientations spécifiques à la politique d'accès aux données de santé. Peu visible, cette gouvernance est pourtant cruciale ; la question est de savoir qui fait partie de cette gouvernance stratégique, dans la mesure où elle est susceptible d'influencer les Ministres et la rédaction des textes réglementaires en matière de données de santé.

- **Les enjeux pratiques** : Un des derniers aspects à prendre en compte dans l'analyse de l'ouverture ou de la fermeture des bases de données de santé est celui des enjeux proprement 'pratiques' ou 'techniques' liés à l'accès et à la réalisation des extractions. En effet, l'octroi d'une autorisation n'est que la première étape de l'accès, il s'agit ensuite de permettre aux demandeurs d'accéder aux données qu'ils souhaitent traiter. Dans ce cadre, plusieurs caractéristiques ou dispositifs peuvent contribuer à faciliter ou au contraire rendre plus difficile l'accès aux données de santé. Les bases de données telles que le SNIIRAM et le PMSI, créées avec un objectif comptable et d'administration, sont en effet particulièrement difficiles à manipuler et à interpréter. Elles requièrent l'acquisition d'outils techniques spécifiques pour pouvoir les explorer, mais également une maîtrise approfondie de la réglementation afin d'éviter les erreurs d'interprétation.⁵ Une des façons de faciliter cet accès est de créer des outils plus faciles à appréhender techniquement, grâce auxquels les données sont déjà en partie traitées (c'est-à-dire triées et lissées). Un autre aspect important à prendre en compte est celui des moyens mis en œuvre pour réaliser les extractions. Si le nombre de demandes est disproportionné par rapport aux moyens alloués, l'instruction des demandes peut alors se dérouler sur un temps extrêmement long, peu compatible avec les objectifs des demandeurs. Ce temps peut encore être allongé lorsque les données demandées font appel à des appariements entre plusieurs jeux de données, appariement qui peuvent être faits de manière systématique (objectif du SNDS ou du *Health Data Hub*). Enfin, le support de l'extraction lui-même peut constituer un enjeu important de l'accès : faut-il se rendre sur place pour analyser les données ? Ou est-il possible de créer une 'bulle' d'accès sécurisée à distance, qui permet de tracer les opérations réalisées sur les données ?

Du fait de ces multiples 'blocs', l'accès aux données de santé se caractérise la plupart du temps par son ambivalence plutôt que par son degré absolu d'ouverture ou de fermeture. Cette ambivalence se manifeste également par la diversité des acteurs à prendre en considération pour comprendre l'ouverture ou la fermeture des bases de données : à chaque période analysée, certains types d'acteurs vont connaître des restrictions d'accès, tandis que d'autres connaîtront dans le même temps une ouverture de leur accès aux données de santé. Pour comprendre l'ouverture et la fermeture des bases de données de santé en France, il faut ainsi comprendre *qui* a accès aux bases de données, à quel *moment* et dans quelles *conditions*.

L'ambiguïté de ce secteur s'explique en grande partie par le fait que l'enjeu de l'accès aux données de santé s'inscrit dans des luttes entre les acteurs historiques du système de santé : assurance maladie, Ministères, professionnels de santé, mais également industries pharmaceutiques, représentants de patients et assureurs complémentaires. Plus encore, la thématique des données de santé a la particularité

⁵ Pour prendre un exemple très simple, une décroissance abrupte de l'utilisation d'une spécialité pharmaceutique peut n'avoir aucun rapport avec la diminution de la fréquence d'une maladie, mais simplement être le reflet de retrait de cette spécialité de la liste des médicaments remboursés.

de rassembler l'ensemble de ces acteurs autour d'un même enjeu, alors que leurs activités tendent habituellement à se concentrer sur des points d'entrée différenciés : autorisation de mise sur le marché et fixation des prix pour les industries pharmaceutiques, négociations conventionnelles (ou question plus générale de la tarification des prestations) pour les professions libérales du secteur de la santé, construction de la 'démocratie sanitaire' pour les organisations de patients (ou 'représentants d'usagers'), ou encore organisation et répartition de la couverture du risque santé pour les assurances maladie obligatoire et complémentaires.

Encadré. Les acteurs du système de santé français

En France, le système de santé (dans sa version après-guerre) a été construit sur un modèle Bismarckien, dans lequel les professionnels de santé exercent majoritairement en libéral. Dans les systèmes dits Bismarckiens, le financement de la protection sociale (et dans notre cas, du risque santé) est principalement assuré par le biais de cotisations salariales et patronales. C'est ainsi le fait de travailler qui 'donne droit' à une protection sociale. En termes organisationnel, cela signifie que les caisses qui récoltent ces cotisations (en France l'Assurance maladie, ou la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés – CNAMTS) sont des acteurs centraux du système de santé, car ce sont elles qui assurent toute la gestion financière du système, en particulier en ce qui concerne les soins ambulatoires (médecine en dehors de l'hôpital). Ce sont ainsi les négociations entre la CNAMTS et les professionnels de santé libéraux qui sont à l'origine de la fixation des prix des consultations et autres actes médicaux. Par contraste, le secteur hospitalier (en dehors du secteur privé) est géré par l'État, et les professionnels de santé qui y travaillent sont généralement des salariés ('agents publics'). L'État a par ailleurs la charge d'assurer la protection sociale des personnes qui ne peuvent travailler. Cette protection est financée par l'impôt, et repose sur un principe de solidarité nationale (et non sur une solidarité entre 'cotisants').

Si, historiquement, la CNAMTS et l'État sont deux acteurs nettement distincts (les décisions à la CNAMTS étant prises par les représentants syndicaux des travailleurs et des employeurs), les évolutions de ces dernières années tendent à renforcer le contrôle de l'État sur la CNAMTS, tandis que le financement de l'assurance santé passe de plus en plus par l'impôt. Malgré ces évolutions, la CNAMTS et l'État (et en particulier le Ministère de la Santé) restent deux 'poids lourds' du système de santé, dont les relations, plus ou moins conflictuelles selon les périodes et les enjeux, sont centrales dans la prise de décision en matière de politique de santé.

En parallèle, les professionnels de santé, partenaires centraux du système de santé, sont également un acteur crucial des réformes, de même que les laboratoires pharmaceutiques, qui sont en négociation constante avec la CNAMTS et l'État dans le cadre de la fixation des prix et des remboursements des médicaments.

Plus récemment, des initiatives ont été prises afin d'institutionnaliser la participation d'autres acteurs à la gouvernance du système de santé ; c'est le cas par exemple des assureurs complémentaires avec la création de l'Union nationale des organismes complémentaires d'assurance maladie (UNOCAM) en 2004, ou encore

des organisations de patients avec la création en 2016 de France Assos Santé (Union Nationale des Associations Agréées du Système de Santé).

Pour comprendre la régulation de l'accès aux bases de données de santé en France et ses évolutions, c'est précisément ces luttes d'acteurs qu'il s'agit d'analyser, car ce sont elles qui structurent les résultats de la prise de décision. En parallèle, ces luttes de pouvoir s'insèrent dans des développements plus généraux, qui vont directement affecter la capacité des acteurs à influencer le processus décisionnel : c'est le cas par exemple de l'imbrication de l'enjeu des données de santé au sein de problématiques telles que la réduction des dépenses de santé, le développement de l'*open data* ou de l'intelligence artificielle. Sur certaines périodes, les transformations du secteur de l'accès aux données de santé sont ainsi fortement influencées par les 'fenêtres d'opportunité' constituées par la mise sur agenda de ces enjeux plus globaux.

Dans ce second Working paper, ces dynamiques seront explorées sur la période 2010-2022. Après une première période de structuration de l'enjeu des données de santé au cours des années 1990-2010 (WP #30), le début des années 2010 marque un tournant dans la régulation de ce secteur. En effet, le début des années 2010 signe une première forme 'd'autonomisation' de l'enjeu des données de santé : alors que ce secteur était auparavant fortement imbriqué au sein d'autres enjeux plus globaux du système de santé, il devient une problématique à part entière. Ce mouvement d'autonomisation a alors plusieurs conséquences, la première étant un investissement direct du politique (Ministres et cabinets ministériels) dans la création d'une 'doctrine' en matière d'accès aux données de santé. Ensuite, et alors que ce sont principalement des dynamiques institutionnelles qui jouent sur la régulation de l'accès aux données de santé dans les années 1990-2010 (Working paper n°1), l'investissement du politique se traduit par une publicisation plus large de cet enjeu et par l'inclusion d'un plus grand nombre d'acteurs dans le processus décisionnel. Cependant, comme nous le verrons, cette 'politisation' de l'enjeu des données de santé ne résulte pas dans une dynamique d'ouverture des données de santé ; au contraire, la loi de 2016 témoigne plutôt d'une fermeture de l'accès aux données de santé pour un grand nombre d'acteurs externes au secteur public.

Pour explorer ce paradoxe, ce document de travail revient sur les dynamiques et processus ayant donné lieu à des transformations de l'accès aux données de santé en France au cours des années 2010 et jusqu'en 2022. Pour ce faire, ces transformations seront décrites à l'aide de la typologie présentée ci-dessus (règles d'accès, gouvernance opérationnelle, gouvernance stratégique, enjeux pratiques). Du fait de la complexité de ce secteur, l'objectif de ce document de travail est ainsi d'identifier des 'mouvements' ou des dynamiques d'ouverture ou de fermeture, plutôt que de produire un constat, dans l'absolu, sur le 'degré d'ouverture' des bases de données en France.

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. Pour identifier les enquêtés,

nous avons d’abord utilisé la méthode positionnelle (identifier les personnes ayant occupé une fonction institutionnelle ou politique en lien avec la régulation de l’accès aux données de santé), complétée occasionnellement par une méthode réputationnelle (personnes pertinentes à contacter selon les enquêtes). En tout, 21 entretiens semi-directifs ont été réalisés, dont la moitié en présentiel.⁶ Les personnes rencontrées occupaient généralement des fonctions au sein du Ministère de la Santé, de la Caisse Nationale d’Assurance Maladie des Travailleurs Salariés (CNAMTS), des cabinets ministériels ou au sein d’agences spécialisées dans la régulation de l’accès aux données de santé (comme l’Institut des données de santé par exemple). 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. Les informations, en particulier celles qui occupent un statut central vis-à-vis de l’administration de la preuve, ont été ‘triangulées’ autant que possible, c’est-à-dire comparées et confirmées à l’aide de différentes sources.

Dans les deux parties suivantes, les dynamiques de la régulation de l’accès aux données de santé sont décrites autour de deux périodes spécifiques : le début et le milieu des années 2010, puis la fin des années 2010 et le début des années 2020. Dans une première partie, nous nous intéresserons ainsi à la mise sur agenda de la problématique des données de santé en tant que secteur autonome d’action publique, dans un contexte marqué par l’émergence des notions de ‘transparence’ et de ‘Big Data’. Cette partie sera également l’occasion de démêler les processus décisionnels qui aboutissent à la rédaction de l’article 193 de la loi de modernisation du système de santé (Loi Touraine, 2016), qui se traduit par une nette dynamique de fermeture de l’accès aux données de santé pour les acteurs privés, certains parvenant à s’accommoder plus facilement que d’autres aux nouvelles exigences de l’accès aux données de santé. Enfin, une seconde partie traitera des derniers développements en matière d’accès aux données de santé, principalement à partir de l’élection de la République en Marche en 2017. S’il existe une forte continuité entre cette période et la période précédente, la fin des années 2010 marque un tournant plus net vers une volonté d’industrialisation de l’usage des données de santé, notamment avec la création du *Health Data Hub* (ou ‘Plateforme des données de santé’).

I. ‘Ouvrir les données en commençant par les refermer’, ou la reprise en main du politique (début et milieu des années 2010)

Malgré le mouvement d’ouverture de l’accès aux données de santé opéré dans les années 2000, cet accès est encore jugé insuffisant par de nombreux acteurs, qui en dénoncent entre autres la complexité et la lenteur. Le sujet de l’ouverture tend ainsi à prendre de l’ampleur chez un nombre toujours plus grand d’acteurs, ce qui conduit, au début des années 2010, à la mise à l’agenda politique de ce problème. En d’autres mots, Marisol Touraine, qui devient Ministre de la Santé en 2012 suite

⁶ Certains de ces entretiens ont été réalisés antérieurement à cette étude dans le cadre de la préparation d’une thèse de doctorat en science politique sur les politiques d’e-santé.

aux élections présidentielles, se voit saisie de cette question, à laquelle les acteurs souhaitent que l'État apporte une réponse. Commence alors un processus de concertation avec les acteurs de ce secteur qui donne lieu par la suite à un ensemble de réformes vis-à-vis de l'ouverture des données de santé, rassemblées dans l'article 193 de la loi de modernisation de notre système de santé de 2016. Comme nous le verrons, contrairement aux attentes, ce processus décisionnel abouti à une dynamique de fermeture de l'accès aux données de santé, en particulier pour les acteurs ne relevant pas du 'service public'. Ces résultats s'expliquent par la congruence des positionnements de l'État, de la CNAMTS, et de certains représentants de la communauté scientifique pour une meilleure sécurisation de l'accès aux données de santé.

a) La mise sur agenda de l'enjeu des données de santé

Pour commencer, lorsque l'on s'intéresse aux différents éléments ayant conduit à la mise sur agenda du problème des données de santé au niveau politique, c'est bien une multitude d'évènements et de facteurs différents qui semblent avoir contribué à ce phénomène, certains soulignant la nécessité d'une plus grande ouverture des données, d'autres mettant en lumière les risques associés à une telle ouverture. Bien souvent, ce sont d'ailleurs les mêmes évènements qui génèrent à la fois des discours prônant l'ouverture, et d'autres prônant une plus grande sécurisation de l'accès aux données de santé.

L'affaire du Mediator®

En premier lieu, l'affaire du Mediator®, qui éclate en 2010, va contribuer à mettre en exergue l'enjeu des données de santé sur l'agenda politique, en soulignant à la fois les bénéfices d'un meilleur usage des données de santé, mais également les risques qui y sont associés.

L'affaire Mediator®, du nom de la spécialité pharmaceutique produite par les laboratoires Servier, est une affaire emblématique qui met en exergue les failles du système sanitaire français. Ce médicament aurait en effet été responsable d'environ 2000 décès, ainsi que de plusieurs milliers d'hospitalisations et de chirurgies directement attribuables à ses effets secondaires.⁷ Véritable scandale, qui donne même lieu à une production cinématographique,⁸ cette affaire a une très forte répercussion sur l'opinion publique, et conduit à une profonde réforme du système d'autorisation de mise sur le marché des médicaments.

En 1960, les laboratoires Servier mettent au point une molécule appelée benfluorex (commercialisée dans les années 1970 sous le nom de Mediator®) dont l'objectif est de produire un effet 'coupe-faim'. Le benfluorex est créé à partir des amphétamines, l'idée étant de conserver le caractère anorexigène de ces molécules tout en éliminant les effets secondaires graves, notamment cardiaques. Le marché visé est alors le marché du surpoids aux États-Unis (Morelle and Padis 2011). Très vite, les laboratoires Servier œuvrent auprès des agences sanitaires françaises afin

⁷ Favereau Eric (2019). 'Laboratoires Servier, anatomie d'un système', *Libération*, 20 septembre 2019, Disponible sur : https://www.liberation.fr/france/2019/09/20/laboratoires-servier-anatomie-d-un-systeme_1752695/

⁸ 'La fille de Brest', film réalisé par Emmanuelle Bercot sorti en 2016.

d'élargir la palette d'indications associées au benfluorex, dont la prescription est alors également autorisée dans le cadre du traitement du diabète et du cholestérol. Dès les années 1990, des alertes remontées par le système français de pharmacovigilance⁹ soulignent la dangerosité du médicament, qui semble être associé à un risque accru d'hypertension et de valvulopathie cardiaque. Ces premières alertes ne provoquent toutefois pas de réaction du côté des pouvoirs publics.

L'enjeu de l'accès aux données de santé dans le cadre de l'enquête

Ce sont finalement des canaux d'information 'hors circuit'¹⁰ qui permettent de révéler l'affaire au grand jour à la fin des années 2000. En particulier, Irène Frachon, pneumologue au CHU de Brest, joue un véritable rôle de 'lanceuse d'alerte' : suite au décès de certains de ces patients, elle entreprend de mener une enquête sur la dangerosité du Mediator®, enquête tout d'abord limitée aux données dont elle dispose, c'est-à-dire celles de ses propres patients. Cette première étude souffre néanmoins d'un défaut de représentativité, et ce sont finalement ses contacts à la CNAMTS qui lui permettent de démontrer sans doute possible les effets néfastes du Mediator®. Pour ce faire, ce sont précisément les données du SNIIRAM qui sont utilisées, sur la base d'une requête relativement informelle qu'Irène Frachon adresse à Alain Weil, alors médecin de santé public à la CNAMTS. En effet, se serait en le croisant 'quasiment dans un couloir' (Entretien, ex-agent CNAMTS, 22/09/2022 ; Entretien, fonctionnaire et agent CNAMTS, 20/07/2022) qu'Irène Frachon aurait demandé de vérifier ses hypothèses sur les données du SNIIRAM. Ces données permettent alors d'objectiver statistiquement les effets secondaires du Mediator® ; alors que les preuves relevaient auparavant de 'signaux faibles', la dangerosité de ce médicament 'crève les yeux' une fois les données du SNIIRAM mobilisées (Entretien, ex-agent CNAMTS, 22/09/2022).

L'affaire Mediator® contribue alors à mettre sur l'agenda l'enjeu des données de santé, en soulignant 'le côté vraiment extravagant de la fermeture des données' (Entretien, fonctionnaire et agent CNAMTS, 20/07/2022). En effet, si Irène Frachon était passée par une procédure habituelle de demande d'accès au SNIIRAM, l'enquête aurait sans aucun doute été bien plus longue, alors même que le Mediator® continuait à être prescrit. Par ailleurs, l'étude qui permet de démontrer l'existence d'effets secondaires graves associés à la prise du médicament Mediator® utilise en réalité des données qui auraient dû, selon les exigences de la CNIL, être effacées à ce moment-là, la durée de conservation des données ne devant pas excéder deux années (Entretien, fonctionnaire et agent CNAMTS, 20/07/2022 ; Entretien, ex-agent CNAMTS, 22/09/2022).¹¹

⁹ En France, le système de pharmacovigilance vise à faire remonter et à centraliser, en permanence, toute suspicion d'effets néfastes des médicaments. Les professionnels de santé sont ainsi invités à notifier tout effet secondaire potentiellement attribuable aux médicaments qu'ils prescrivent

¹⁰ Bensadon Anne-Carole, Marie Étienne, Morelle Aquilino (2011). *Enquête sur le MEDIATOR®. Synthèse*. Inspection Générale des Affaires sociales, Fiche disponible au format PDF sur : https://www.igas.gouv.fr/IMG/pdf/Synthese_MEDIATOR.pdf

¹¹ L'affaire du Mediator® permet par la suite d'allonger cette durée de conservation.

Les conséquences du Mediator® sur la gouvernance de l'accès aux données de santé

Cette affaire a contribué à plusieurs évolutions dans le secteur de l'accès aux données de santé. Tout d'abord, c'est à la suite du Mediator® que les données du SNIIRAM seront ouvertes plus largement aux agences de sécurité sanitaire (Entretien, agent CNAMTS, 21/04/2022 ; Entretien, fonctionnaire et agent CNAMTS, 20/07/2022). Ensuite, cette affaire met un coup de projecteur sur les acteurs de l'ouverture des données de santé, et en particulier sur l'Institut des Données de Santé, dont les représentants 'vont à l'Elysée' rencontrer les proches conseillers du Président pour défendre l'ouverture de l'accès aux données de santé. Malgré les apparences cependant, des dynamiques contradictoires vont contribuer à contrarier l'agenda de l'IDS. En effet, comme nous l'avons vu précédemment (Working paper n°1), il existe des antagonismes forts entre la CNAMTS et l'IDS, la CNAMTS défendant une position de prudence vis-à-vis de l'ouverture des données de santé (en particulier celles du SNIIRAM), tandis que l'IDS promeut une vision beaucoup plus libérale de cet accès. Aussi, au lieu de laisser l'IDS renforcer sa position institutionnelle, la CNAMTS entreprend-elle de monter un nouveau pôle d'accès aux données de santé, qui rassemblerait (en plus de la CNAMTS et l'État) les institutions de sécurité sanitaire (la Haute Autorité de Santé, l'Institut de Veille Sanitaire et l'Agence Nationale de Sécurité du Médicament) autour d'un Groupement d'Intérêt Public (GIP) dédié à la mise à disposition des données de santé dans le cadre d'études sur les spécialités pharmaceutiques.¹² Ce GIP est alors très clairement perçu comme une mise en concurrence par l'IDS, qui considère que les missions qui lui sont attribuées relèvent déjà de sa charge.¹³ L'IDS bénéficie finalement de soutiens politiques à l'Assemblée Nationale qui lui permettent de conforter sa position (Entretien, agent CNAMTS, 21/04/2022), tandis que la mise en œuvre du nouveau GIP est abandonnée (Entretien, fonctionnaire et agent CNAMTS, 20/07/2022).¹⁴

Bien que l'IDS ait finalement gagné cette bataille institutionnelle, l'affaire Mediator® contribue à faire émerger des arguments remettant en cause sa légitimité, y compris par des acteurs ayant soutenu l'IDS dans les années 2000. En effet, l'ouverture des perspectives d'usage des données de santé met l'IDS dans une position délicate, car celui-ci est exclusivement perçu comme un produit de l'accord passé avec les assurances complémentaires (Entretien, agent CNAMTS, 21/04/2022 ; Entretien, représentant, association d'usagers, 30/06/2022). Ainsi, l'inclusion de nouveaux acteurs et enjeux dans la problématique de l'accès aux données de santé tend à diminuer progressivement la légitimité perçue de l'IDS.

¹² Article 33 de la loi n° 2011-2012 du 29 décembre 2011 relative au renforcement de la sécurité sanitaire du médicament et des produits de santé

¹³ Institut des données de santé (2012). *Rapport au Parlement 2012. Progrès et projets*. Charenton-le-Pont : publications de l'IDS, pp.76-77

¹⁴ Néanmoins, la CNAMTS et l'Agence Nationale de Sécurité du Médicament et des produits pharmaceutiques s'associent en 2018 pour créer un Groupement d'Intérêt Scientifique (GIS) intitulé EPI-PHARE, qui regroupe les équipes d'épidémiologie des produits de santé des deux établissements. L'objectif est alors de faciliter la production d'études en matière de produit de santé et de sécurité sanitaire à partir des données du Système National des Données de Santé.

La double facette de l'affaire du Mediator® : bénéfiques et risques associés à l'usage des données de santé

Par ailleurs, l'affaire du Mediator® souligne un autre aspect lié au traitement des données de santé, à savoir le risque politique et juridique associé leur (non) usage. D'un côté, l'État et les agences de sécurité sanitaire courent le risque d'être mis en cause pour ne pas avoir pris des mesures pour analyser les effets du Mediator®, alors que les données permettant de le faire étaient en réalité disponibles (Entretien, secteur des assurances complémentaires, 15/06/2022 ; Entretien, fonctionnaire et agent CNAMTS, 20/07/2022 ; Entretien, ex-agent CNAMTS, 22/09/2022).

'[...] l'État avait son intérêt de l'État ou de la CNAMTS, avaient leurs intérêts de pouvoir, mais ils avaient aussi la **crainte** que si l'extérieur avait des données et bien ça entraînerait, ça risquait d'entraîner des **mises en cause de la responsabilité pénale**, parce qu'en matière de médicaments notamment, mais aussi dans d'autres affaires, par exemple en matière d'épidémie, hein, on voit bien si les données sont exploitées, bah on peut rechercher si on le veut, qu'est-ce qui s'est passé à tout moment.

[...]

Simplement sur l'histoire du Médiateur®, si vous voulez au contraire ça, déterminé les gens qui ne voulaient pas de l'ouverture, ça les a déterminés à être encore plus contre l'ouverture, puisqu'ils ont dit, mais il y a l'affaire du Médiateur®, **plus jamais l'affaire du Médiateur®** et le meilleur moyen pour eux qu'il n'y ait plus l'affaire du Médiateur® c'était pas du tout d'être plus transparent, plus rigoureux, plus strict, c'était **d'empêcher l'accès aux données** pour que de nouvelles Irène Frachon ne viennent pas les embêter et leur envoyer les juges d'instruction et les services d'enquête.' (Entretien, fonctionnaire et agent CNAMTS, 20/07/2022)

'[...] ce à quoi ces données ouvrent, c'est le **risque permanent qu'on vous dise que c'était dans les données** et qu'on n'est pas allé le chercher.' (Entretien, ex-agent CNAMTS, 22/09/2022)

Les usages des bases de données par des acteurs externes à l'appareil étatique comportent, de toute évidence, un fort potentiel de déstabilisation politique. Au regard de l'usage des données de santé, l'affaire Mediator® a donc sans doute joué à double tranchant, soulignant à la fois l'intérêt des bases de données dans le cadre de la poursuite de l'intérêt collectif, mais également les risques associés à leur existence.

De multiples 'affaires' liées à l'usage des données de santé

Une dichotomie similaire apparaît dans plusieurs 'affaires' au début des années 2010. On note par exemple le cas de l'entreprise Fourmisanté, qui récupérait les données disponibles sur le site de la CNAMTS à propos des professionnels de santé pour les intégrer à un logiciel de géolocalisation, permettant aux utilisateurs de s'orienter vers le professionnel de santé de leur choix en fonction de différents critères, notamment de remboursement. Alors qu'Etalab, l'administration en

charge de favoriser la transparence et l'ouverture des données publiques,¹⁵ décerne à Fourmisanté le prix grand public de son premier concours Dataconnexions^{16, 17}, la CNAMTS s'oppose fermement à cet usage et adresse une mise en demeure à la start-up, au motif qu'elle ne disposerait pas de l'autorisation de transférer une base de données contenant des informations spécifiques sur les professionnels de santé (Entretien, représentant, secteur des assurances complémentaires, 16/06/2022).

De même, la création, en 2013, d'un Observatoire citoyen du reste-à-charge en santé par le Collectif Interassociatif en Santé (CISS), Santéclair et 60 millions de consommateurs provoque immédiatement des levées de bouclier du côté des professionnels de santé, qui en dénoncent les limites méthodologiques ainsi que le caractère 'vindicatif' (Entretien, ex-représentant des professionnels de santé, 16/06/2022).¹⁸ Extraites du SNIIRAM, les données qui servent à alimenter cet observatoire sont mise à disposition du CISS via la création et l'alimentation d'un tableau de bord par l'IDS.¹⁹

Avec les débuts de l'exploitation du potentiel des bases de données de santé, ce sont donc tout autant l'intérêt et les difficultés d'accès aux données que les risques politiques et juridiques associés à leur usage qui sont soulignés.

L'initiative 'Libérez nos données' et la pression médiatique

En parallèle, certains acteurs issus de la 'société civile' vont contribuer à mettre ce sujet sur l'agenda politique en militant activement pour une plus grande ouverture des données. Ainsi, en janvier 2013 est lancé un mouvement intitulé 'Libérez nos données' rassemblant des acteurs relativement hétéroclites²⁰ : représentants d'utilisateurs du système de santé, chercheurs, assureurs, entrepreneurs (comme Celtipharm, devenu par la suite OpenHealth, entreprise de collecte de données de santé), consultants et journalistes.²¹ Parmi ces acteurs, on compte des représentants d'utilisateurs (notamment le CISS), l'UFCC-Que-Choisir, 60 millions consommateurs, Santéclair, Fourmi Santé, ou encore l'ESSEC Business School. Ces acteurs, très

¹⁵ Créée en 2011, Etalab est une administration publique française dont le but est de favoriser la transparence et l'ouverture des données produites au sein des administrations publiques. *In fine*, l'objectif est que ces données soient réutilisées de manière 'innovante' par les acteurs de la société civile. Pour plus d'information sur le mouvement 'Open data' porté par le gouvernement voir : Cluzel-Métayer 2018.

¹⁶ Les Dataconnexions désignent des concours (6 éditions à ce jour) dont l'objectif est de distinguer les initiatives basées sur une réutilisation des données des administrations publiques.

¹⁷ Bourboulon François (2012). 'Six projets innovants distingués par le concours Dataconnexions'. *Les Echos*, publié le 1 juin 2012 [en ligne], Disponible sur : <https://www.lesechos.fr/2012/06/six-projets-innovants-distingues-par-le-concours-dataconnexions-357639>

¹⁸ CSFM. 'Observatoire citoyen des restes à charge : un tribunal inquisitorial'. Tribune du 27/05/2013, Disponible sur : <https://toute-la.veille-acteurs-sante.fr/23257/observatoire-citoyen-des-restes-a-charge-un-tribunal-inquisitorial/>

¹⁹ Institut des données de santé (2015). *Rapport au Parlement 2015. Ouverture, qualité, partage : des avancées*. Charenton-le-Pont : publications de l'IDS, p.31

²⁰ Cette absence d'homogénéité n'est pas sans soulever des interrogations à l'époque, même par les acteurs participant à ce mouvement. Les suites de l'affaire Mediator créent par ailleurs des dissensions importantes entre les acteurs, qui conduisent certains à quitter le collectif (Entretien, représentant, association d'utilisateurs, 30/06/2022).

²¹ Cette initiative donne lieu à l'ouverture d'un site dédié dont sont extraites les informations relayées dans cette partie : <http://www.opendatasante.com> [Site web archivé]

divers, semblent avoir en commun une expérience de refus d'accès à certaines données du SNIIRAM par la CNAMTS (Entretien, représentant, association d'usagers, 30/06/2022). Au vu de cette variété, les motifs et les potentiels usages des données de santé par les différents acteurs sont très hétérogènes ; cependant, le plaidoyer en faveur de l'ouverture de cet accès (qui circule sous la forme d'une pétition) souligne les bénéfices pour les patients, par exemple pour prévenir les scandales sanitaires, comparer la qualité des hôpitaux ou des professionnels de santé, pour mesurer les dépassements d'honoraires, ou tout simplement pour assurer plus de transparence dans la gestion du système de santé. L'accès aux données de santé est cadré comme une problématique de démocratie sanitaire et de transparence, avec l'idée que les citoyens devraient disposer de la meilleure information possible sur les soins qui leur sont proposés. Cette initiative vise à générer un engouement médiatique afin de s'assurer que la question de l'accès aux données de santé soit traitée politiquement, et ce mouvement aurait ainsi contribué 'à secouer le cocotier' (Entretien, ex-agent CNAMTS, 22/09/2022), c'est-à-dire à interpeller la Ministre sur ce sujet (Entretien, fonctionnaire, 17/06/2022 ; Entretien, agent CNAMTS, 21/04/2022).

Le mouvement Libérez nos données se double en parallèle et à une période à peu près similaire de plusieurs interventions dans les médias de membres de l'IDS, notamment son Président Christian Babusiaux, qui se positionne ouvertement en faveur d'une plus grande ouverture de l'accès aux données de santé, par exemple dans le cadre d'un entretien avec un journaliste de *L'Argus de l'Assurance* en mars 2012.²² D'autres acteurs se réfèrent aux interventions du Président du comité des experts de l'IDS, Didier Sicard, qui aurait été interviewé dans un article 'au vitriol' (Entretien, ex-agent CNAMTS, 22/09/2022). Le tout aurait ainsi créé une 'pression médiatique' sur les administrations (Entretien, représentant, secteur des assurances complémentaires, 16/06/2022).

Le rôle de la communauté scientifique

D'un autre côté, plusieurs acteurs de la communauté scientifique, ayant pris conscience de l'existence et du potentiel d'une base de données telle que le SNIIRAM au cours des années 2000, soulignent la nécessité de rendre cette base plus disponible et accessible aux chercheurs (Entretien, ex-agent CNAMTS, 19/04/2022).²³ En particulier, le Haut Conseil de la santé publique (HCSP) créé en 2007 avec une mission d'expertise auprès des pouvoirs publics, déplore les limites d'accès aux données de santé qui pourraient lui permettre de réaliser des évaluations plus fines des politiques de santé nationales.²⁴ Parmi les membres du groupe de travail en charge de cette question au HCSP, on retrouve en chef de file Marcel Goldberg, épidémiologiste à l'INSERM, connu pour son travail sur les

²² Limoge François (2012). 'Christian Babusiaux, président de l'Institut des données de santé (IDS): « Chacun a intérêt à pouvoir disposer des données des autres »'. *L'Argus de l'Assurance* [en ligne]. Disponible sur : <https://www.argusdelassurance.com/interviews/christian-babusiaux-president-de-l-institut-des-donnees-de-sante-ids-chacun-a-interet-a-pouvoir-disposer-des-donnees-des-autres.54610>

²³ DREES (2015). *Données de santé : anonymat et risque de ré-identification*. Dossiers Solidarité et Santé, n° 64, p.38 ; Conférence Nationale de la Santé (2010). *Avis sur les données de santé informatisées*. Paris : publications de la Conférence Nationale de Santé, p.5

²⁴ HCSP (2009). *Les systèmes d'information pour la santé publique*. Paris : La Documentation française

cohortes Gazel et Constances et sur les possibilités d'appariement avec le SNIIRAM (Goldberg *et al.* 2016). Il semblerait ainsi que certains chercheurs aient également contribué à la mise sur agenda le problème de l'accès aux données de santé, sans doute aidés, au moins partiellement, par le canal institutionnel du HCSP, dont le rapport aurait poussé la CNAMTS à travailler sur un ensemble de solutions techniques afin de faciliter l'accès aux données du SNIIRAM (Entretien, ex-agent CNAMTS, 19/04/2022).

b) Le processus de formulation de l'article 193 de la loi de modernisation de notre système de santé (2016)

Ouvrir ou sécuriser les données de santé : le cadrage ambigu des travaux préparatoires

Au début des années 2010, un ensemble d'évènements, soulignant à la fois le potentiel (en grande partie inexploité) des données de santé, de même que les risques associés à leur usage, contribuent ainsi à mettre le sujet sur l'agenda politique de la Ministre Marisol Touraine fraîchement élue. Elle décide alors de commander un rapport ce sujet, dont la rédaction est confiée à Pierre-Louis Bras (Inspecteur Général des Affaires Sociales, IGAS) et André Loth (Direction de la Recherche, des Études, de l'Évaluation et des Statistiques, DREES).²⁵ Rédigé en quelques mois et publié en septembre 2013, ce rapport de 128 pages souligne les difficultés et la complexité du système d'accès aux données de santé, mais met également en avant les risques de réidentification et de 'mésusage' des données de santé. Il est par ailleurs intéressant de noter que dans la partie dédiée aux constats à propos des accès aux bases de données, la première sous-partie traite des risques de réidentification, avant de détailler les difficultés, les problèmes d'appariement, la complexité et les 'délais excessifs' d'accès aux bases de données. Ce rapport traite donc autant des risques que de la nécessité de simplifier l'accès aux bases de données,²⁶ qui doit être assuré seulement aux acteurs pouvant démontrer l'intérêt public associé à leur démarche et présentant un certain nombre de garanties.

À la suite de ce rapport, c'est cependant le cadrage axé sur l'ouverture des données de santé qui prédomine dans les discours officiels (Entretien, secteur des assurances complémentaires, 15/06/2022). À ce titre est par exemple instaurée la Commission 'Open data santé', qui rassemble les parties-prenantes du système de santé en vue de préparer une réforme de l'accès aux données de santé dans le cadre de la loi de modernisation de notre système de santé adoptée en 2016. En plus de répondre aux demandes de nombreux acteurs vis-à-vis de l'ouverture des données de santé (voir sous-section précédente), le cadrage de cette Commission s'aligne avec des préoccupations plus générales du gouvernement en matière d'*open data* et de transparence des administrations, comme en témoigne l'investissement de la mission Etalab dans cette initiative.

²⁵ Bras, Pierre-Louis et André Loth (2013). *Rapport sur la gouvernance et l'utilisation des données de santé*. Paris : publications de la DREES.

²⁶ En ce sens, le contenu du rapport reflète la lettre de mission adressée par la Ministre, dans laquelle ces deux thématiques sont également présentes. Source : Bras, Pierre-Louis et André Loth (2013). *Rapport sur la gouvernance et l'utilisation des données de santé*. Paris : publications de la DREES, p.3

La Commission Open data santé réunit 43 membres et mène de nombreux travaux l'accès aux données de santé entre novembre 2013 et mai 2014.²⁷ Elle est animée par le Ministère de la Santé, et plus précisément Philippe Burnel (Délégation à la stratégie des systèmes d'information en santé, DSSIS) et Franck von Lennepe (directeur de la DREES). En plus du recueil des contributions des différents acteurs de l'écosystème de santé (Syndicat National de l'Industrie des Technologies Médicales SNITEM, Fédération de l'Hospitalisation Privée FHP, Les Entreprises du Médicament LEEM, Union Nationale des Professionnels de Santé UNPS, Syntec Numérique), la Commission organise une 'consultation publique' en ligne qui recueille environ 200 réponses. Le rapport final, rendu en juillet 2014, contient également des éléments de *benchmark* et est supposé servir à alimenter le projet de loi de modernisation du système de santé.²⁸

Ce rapport préconise une très grande ouverture des données anonymes en *open data*, soutenant l'idée que tout ce qui ne nuit pas aux individus doit être partagé le plus largement possible, sans toutefois devenir trop coûteux. Le rapport propose ainsi plusieurs solutions techniques qui permettraient d'ouvrir au maximum au public les données de santé anonymes existantes. Le rapport recommande également de faciliter l'accès aux données de santé comportant un fort risque de réidentification dans certains cas spécifiques (en particulier dans le cadre de la recherche scientifique). Suite à la publication de ce rapport, Marisol Touraine élargit son équipe en recrutant Anne-Gaëlle Javelle, qui devient conseillère 'open data et e-santé' d'octobre 2014 à février 2017. À ce moment-là, le discours officiel semble donc favoriser un cadrage en faveur de l'ouverture des données de santé, comme en attestent les intitulés de la commission 'Open data' et des missions de la conseillère en santé digitale de Marisol Touraine, ainsi que le titre de l'article de loi qui sera issu de ces travaux 'Créer les conditions d'un accès ouvert aux données de santé' (article 193 de la loi de 2016).

Cependant, comme nous le verrons lors de la description des instruments de la régulation de l'accès aux données de santé issus de la loi de modernisation de notre système de santé (section II.c), l'ouverture des données de santé est alors au mieux ambivalente, sinon clairement axée sur une dynamique de fermeture. D'ailleurs, malgré l'orientation annoncée de la Commission Open data santé, certains acteurs témoignent de l'infléchissement très rapide des débats vers la sécurisation des données de santé (Entretien, représentant, association d'usagers, 19/07/2022 ; Entretien, secteur des assurances complémentaires, 15/06/2022) :

'voilà je pense que c'était ça l'idée derrière et que au fond la position française n'a jamais été malgré son nom, d'ailleurs je crois que je l'avais dit d'ailleurs, que on aurait mieux fait d'appeler ça la **Commission « Closed data »**' (Entretien, représentant, association d'usagers, 19/07/2022)

²⁷ Si cette Commission a réalisé une 'consultation publique', on note qu'il n'y aucun représentant de patients au sein de la Commission, essentiellement composée de fonctionnaires, de professionnels de santé et d'industriels.

²⁸ Commission Open Data en santé (2014). *Rapport remis à Madame Marisol Touraine, Ministre des Affaires sociales et de la Santé, le 9 juillet 2014*. Paris : publications de la DREES.

‘Suite à cette commission, enfin, pendant cette commission, il y a eu beaucoup de débats autour... **des risques de mésusages**, des données de santé. Et c'est à ce moment-là qu'a été formalisée ce qu'on appelle les finalités interdites d'utilisation du SNDS, qui visent deux grandes catégories d'acteurs’ (Entretien, secteur des assurances complémentaires, 15/06/2022)

Un accord entre la CNAMTS, l'État, et la communauté scientifique pour sécuriser l'accès aux données de santé

Malgré un cadrage officiel axé sur l'ouverture de l'accès aux données de santé, la dynamique qui se joue alors dans la prise de décision favorise une sélection des instruments orientée vers la fermeture. En effet, si les partisans de l'ouverture sont relativement nombreux (représentants de patients/consommateurs, industriels, complémentaires santé, IDS), les partisans d'une meilleure régulation de l'accès trouvent trois alliés de poids, à savoir l'État, la CNAMTS, et le monde de la recherche scientifique (Entretien, fonctionnaire et agent CNAMTS, 20/07/2022 ; Entretien, agent CNAMTS, 21/04/2022). Si ces trois acteurs sont loin d'être d'accord sur l'ensemble de la politique des données de santé,²⁹ on trouve cependant une résonance forte entre leurs arguments en faveur d'une meilleure régulation.

La position du Ministère de la Santé

En premier lieu, les représentants de l'État, que ce soit au cabinet ministériel ou à la DREES, se montrent particulièrement précautionneux vis-à-vis de l'ouverture des données de santé, qui est alors associée à tout un ensemble de risques et génère une certaine forme de ‘réticence politique’ (Entretien, ex-fonctionnaire, 06/04/2022 ; Entretien, fonctionnaire, 17/06/2022 ; Entretien, représentant, association d'usagers, 30/06/2022 ; Entretien, secteur des assurances complémentaires, 15/06/2022).

‘[...] mais je pense qu'il avait pour mission, de la part de Marisol Touraine, de faire en sorte que quand la puissance publique collecte des données et qu'elle les traite, elle ait un niveau de garantie suffisant dans l'accès à ces données, pour qu'elle ne soit **jamais poursuivie pénalement pour infraction à la sécurité des données**, ou pour avoir causé un risque à la réputation des personnes, ou avoir causé un risque à la sécurité des personnes. À la sécurité sanitaire.’ (Entretien, représentant, association d'usagers, 19/07/2022)

‘[...] la **principale crainte c'était de dire si on va trop loin dans l'ouverture, on peut avoir des fuites, et si demain il y a des fuites, enfin un accès inapproprié aux données, et bien c'est toute la crédibilité du système qui sera remis en cause**. Enfin moi le souvenir que j'en ai c'est plus ça qu'on avait au niveau du cabinet de la Ministre comme réticences à gérer c'est-à-dire voilà, n'allons pas trop

²⁹ En effet, l'idée de confier la gestion du SNIIRAM à la DREES est pendant un temps évoquée (et très mal reçue par la CNAMTS, Entretien, agent CNAMTS, 21/04/2022 ; Entretien, fonctionnaire et agent CNAMTS, 20/07/2022), tandis que les chercheurs prennent soin d'établir des distinctions claires entre les données du SNIIRAM et les données de leurs propres cohortes, qu'ils jugent beaucoup plus complexes à manipuler et qui donc mériteraient une régulation plus stricte que celles du SNIIRAM (Entretien, juriste spécialiste des données de santé, 15/06/2022).

loin trop vite parce que ça fragilise le système et si le système est fragilisé et bien on n'aura plus rien, parce que on aura à ce moment-là la société civile va se retourner contre nous, et là où aujourd'hui on a des gens qui demandent de l'ouverture on aura des gens qui demanderont la fermeture.’ (Entretien, ex-fonctionnaire, 06/04/2022)

Il semblerait que ce positionnement ait également été facilité par la structuration du débat interministériel sur le sujet des données de santé. En effet, l’attribution de ce portefeuille aurait créé des dissensions (une ‘bataille considérable’, Entretien, ex-conseiller de Marisol Touraine, 19/07/2022) entre le Ministère de l’Économie et des Finances, le Ministère de la Recherche et le Ministère de la Santé, avec d’un côté les promoteurs d’une valorisation maximale industrielle et/ou scientifique de ces données considérées comme une ‘mine d’or’, et de l’autre les défenseurs du droit de la personne et de la vie privée au Ministère de la Santé, dont l’objectif était de s’assurer que les données ne tombent pas entre des mains ‘malveillantes’ (Entretien, ex-conseiller de Marisol Touraine, 19/07/2022).³⁰ Le Ministère de la santé prend alors le rôle du dernier ‘rempart’ (Entretien, représentant, association d’usagers, 30/06/2022) contre un usage industriel et déraisonné des données, notamment par les géants du numérique :

‘Oh bah oui vous savez comment tout ça est parti quand j’étais au cabinet, c’est parti suite à une rencontre du **patron de Google**, avec la Ministre, patron de Google euh, voulant voir la Ministre, justement pour voir comment pouvoir accéder aux données de santé françaises. Compte-tenu de leur caractère unique, des données de santé, et du système mis en place par la France, et notamment par l’Assurance maladie depuis tant d’années. Voilà, comment, et donc il y a eu une prise de conscience, à ce moment-là, de l’enjeu considérable des données de santé. [...] Et donc **soit les autorités publiques s’en saisissaient, soit ça allait leur échapper définitivement**. C’était vraiment ça, les discussions qu’on avait, parce qu’on se disait OK, là on ne va pas lui donner les clés d’y accéder mais, Google va y accéder. Sauf si, on, on on construit quelque chose, qui ferme pas la porte, mais qui l’encadre’ (Entretien, ex-conseiller de Marisol Touraine, 19/07/2022)

Du côté de l’administration, cette période est également associée à une forme de ‘découverte’ et de clarification conceptuelle entre ‘anonymisation’ et ‘pseudonymisation’ des données de santé, dont les subtilités étaient jusqu’alors reléguées au champ des spécialistes de ce secteur.

[...] bon je pense que tout le monde l’admet mais ça a été **très long à faire admettre** à ce moment-là, c’est que des **données anonymisées, ne sont pas anonymes**, parce que ce sont des données pseudonymisées en fait et non pas anonymisées, donc pseudonymisées et que ces données pseudonymisées de santé avec des données très fines, à travers les parcours, révèlent, peuvent révéler des choses. Et je me souviens que à ce moment-là [M.X, CNAMTS], m’avait dit en

³⁰ L’arbitrage en faveur du Ministère de la Santé se fera finalement suite à une décision du cabinet du Président de la République.

rigolant qu'il avait fait passer le message, en très haut lieu, donc je pense à l'Elysée, que bah **sur les disques du PMSI on pouvait retrouver n'importe quel Président de la République dès lors qu'il avait été hospitalisé**, parce qu'il suffisait d'avoir deux trois informations publiques, sur où ils étaient hospitalisés et à peu près à quelle date, ce que en fait on sait, et voilà et vous le retrouvez dedans, et une fois que vous l'avez retrouvé, vous pouvez à travers ça retrouver son parcours avant après et avoir des informations en santé. Et donc c'est vraiment à ce moment-là et moi ça m'a frappé **je n'avais vraiment jamais vraiment réalisé ça**, et il y avait une personne qui en fait avait été lanceur d'alertes depuis plusieurs années, dont j'ai oublié le nom mais vous allez me le dire. Sur les risques du PMSI.'

'Le **cœur de débat** il était vraiment sur qu'est-ce qu'une donnée **pseudonymisée** et, ou une donnée **anonyme**. Et ça les gens avaient du mal à comprendre. Ou ne voulaient pas comprendre.'

(Entretien, fonctionnaire, 17/06/2022)

En accord avec ce 'cadrage', la DREES met en place, durant cette période, toute une série de travaux pour travailler sur l'anonymisation, la pseudonymisation et le risque de ré-identification :

'Où c'était un peu nouveau parce que on avait monté, je pense que c'était après le rapport de Pierre-Louis Bras et André Loth, soit c'est pendant le rapport soit c'est juste après mais on avait monté un **séminaire de recherche**, donc ça tout ça doit être en ligne aussi parce que à ce moment-là on les avait publiés sur le site de la DREES, donc organisé par M.X, mais on avait fait venir des gens, y compris des chercheurs alors il y avait des gens de l'INSEE des gens d'ailleurs enfin pour expliquer **l'état des travaux** de recherche sur la **pseudonymisation** et le **risque de réidentification**, et voilà on avait fait un **rapport** là-dessus, X avait fait un rapport. Et on voit bien en effet que c'est vraiment *work in progress* hein.' (Entretien, fonctionnaire, 17/06/2022)

Le positionnement de la CNAMTS

D'un autre côté, la CNAMTS défend sa position habituelle de sécurisation des données de santé, confortant ainsi sa position historique de prudence vis-à-vis de l'ouverture des données de santé (Entretien, ex-agent CNAMTS, 22/09/2022, voir également Working paper n°1).

'Et côté CNAMTS euh, bah la CNAMTS a été, avait plutôt une posture, en tant que gestionnaire historique du coffre-fort, de **prudence**, de prudence euh, en disant attention ce sont des données très sensibles il faut **faire très attention**, de un c'était une question de prudence [...]. Mais dans un premier temps, ils étaient plutôt des côtés des oui, c'est ça comme n'importe quel détenteur de coffre-fort en disant attention attention, ceci est très précieux il faut faire attention.' (Entretien, ex-fonctionnaire, 06/04/2022)

Un des arguments récurrents est ici le risque de mésusage des données, notamment du fait du risque d'interprétation erronée des résultats. En effet, l'exploitation de

la base de données de la CNAMTS nécessite une connaissance fine de la réglementation afin d'interpréter correctement les résultats qui en sont extraits.

Le positionnement des chercheurs de la Commission Open data

Il est intéressant de noter qu'à ce moment-là ce type d'arguments entre en résonance forte avec le positionnement de certains chercheurs scientifiques participant à la Commission Open data, dans la mesure où il fait écho à des craintes que les chercheurs associent eux-mêmes à la création de leurs propres bases de données, qui auraient par ailleurs vocation, *in fine*, à alimenter le Système National des Données de Santé (SNDS) de manière similaire au SNIIRAM. Ainsi, si les chercheurs militent pour une plus grande ouverture des données de santé à la recherche scientifique, ils semblent également exprimer des réticences vis-à-vis des risques liés à l'interprétation des résultats, et se positionnent en faveur d'une sécurisation accrue de l'accès aux données de santé selon des protocoles scientifiques (Entretien, fonctionnaire, 17/06/2022 ; Entretien, représentant, association d'usagers, 30/06/2022).

[...] les **chercheurs** ce qui les intéressait, c'était de s'assurer que l'ouverture des données allait se faire **au bénéfice de la recherche**, et que donc les conditions d'accès aux données permettraient aux chercheurs d'y accéder, alors que les parties prenantes du secteur de la santé, c'est-à-dire pour simplifier les professionnels de santé les industriels du médicament et les organismes complémentaires, eux ce qui les intéressait c'était d'accéder aux données pour eux pour leurs membres, et donc pas du tout pour les mêmes sujets que les chercheurs en fait. Et eux étaient beaucoup moins demandeurs d'avoir des conditions d'accès avec des **comités scientifiques** et un encadrement de tout ça, parce qu'ils avaient peur que ça se fasse à leur détriment, alors qu'ils avaient l'impression qu'à l'IDS c'était facile.' (Entretien, fonctionnaire, 17/06/2022)

Au-delà des chercheurs de la Commission Open data, l'INSERM, qui constitue un des acteurs majeurs de la communauté scientifique, aurait également défendu une ligne relativement dure afin de s'assurer de l'exclusion des laboratoires pharmaceutiques de l'accès aux données de santé (Entretien, agent CNAMTS, 21/04/2022).

Alors que depuis 2004, l'IDS avait réussi à trouver des soutiens institutionnels au sein de l'État pour avancer son agenda d'ouverture, l'investissement politique du Ministère de la Santé se traduit au début des années 2010 par une volonté de sécurisation des données de santé, davantage en accord avec les orientations de la CNAMTS. Ce positionnement est par ailleurs soutenu par certains chercheurs, et, à ce moment-là, par les professionnels de santé (Entretien, représentant, association d'usagers, 30/06/2022 ; Entretien, ex-agent CNAMTS, 22/09/2022).

c) Les transformations de l'accès aux données de santé suite à la loi de 2016

Dans ce contexte, la sélection des instruments associée à cette période de formulation de la politique publique est marquée par une nette dynamique de

fermeture, en tout cas en ce qui concerne l'accès aux données de santé par des acteurs 'privés'.

'[...] on dit qu'on ouvre et en fait **on commence par refermer**' (Entretien, fonctionnaire, 17/06/2022)

'[...] après le **paradoxe**, on a dit bon, il faut faire de **l'open data**, il faut ouvrir l'accès aux données, il faut favoriser l'utilisation des données dans les projets de recherche et tous ces trucs-là. Et ben on est aussi arrivés dans une gouvernance où, **paradoxalement on a refermé** finalement, le partage' (Entretien, secteur des assurances complémentaires, 15/06/2022)

'Le problème de la loi Touraine, c'est qu'elle voulait, elle avait **deux finalités** en fait, assez **contradictaires**. Elle voulait, elle avait pour ambition d'ouvrir beaucoup les données du SNDS [...] Le problème, c'est qu'on a réécrit dans cette même loi le chapitre de la loi Informatique et libertés dédié à la recherche. Et dans tout ça, dans l'esprit RGPD il y avait toute une logique de protection des données. Donc en fait vous poursuivez **une finalité d'ouverture et de protection**, et on voyait bien les enjeux politiques qu'on ne pouvait pas avancer dans l'ouverture, sans avancer dans la protection. Je me suis souvent demandé si s'il y avait eu un *timing* différent et qu'on n'avait pas dû travailler dans cette même dans une même loi, sur l'ouverture et les régimes de de de la loi Informatique et libertés si on aurait maintenu un **régime aussi lourd**, mais il y a eu quand même des avancées, mais ça restait **très lourd**.' (Entretien, juriste spécialiste des données de santé, 15/06/2022)

'[...] là effectivement ça a été restreint, on a refermé, on a **resserré les boulons** là hein.' (Entretien, représentant professionnels de santé, 23/05/2022)

'[...] ça change rien, et y a **pas eu plus d'ouverture**, et contrairement à ce qui est écrit dans la loi on veut ouvrir au privé le privé a eu zéro, zéro [Rires], zéro !' (Entretien, agent CNAMTS, 21/04/2022)

Les règles d'accès : restriction d'accès pour les acteurs 'privés', extension des accès de droit pour les acteurs publics

PMSI et SNIIRAM : une harmonisation des règles d'accès à l'occasion de la création du Système National des Données de Santé (SNDS)

Un des grands changements de la loi de 2016 est relatif aux transformations des règles d'accès au PMSI, qui étaient alors plus libérales et surtout moins sécurisées que celles du SNIIRAM. En particulier, la question de la copie des données annuelles du PMSI sur une disquette transférée ensuite aux demandeurs apparaît comme particulièrement problématique, car elle ne permet pas de tracer l'usage des données. À partir de 2016 donc, le PMSI, désormais incorporé dans le Système National des Données de Santé (SNDS) qui vise à appareiller de manière permanente différentes bases de données de santé, doit faire l'objet d'un accès sécurisé, autorisé non plus seulement par la CNIL mais également par le Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de

la Santé (CEREES, par la suite rapidement renommé CESREES, Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé), et, le cas échéant, par l'Institut National des Données de Santé (INDS, nouvelle mouture de l'IDS, voir sous-section sur la gouvernance opérationnelle pour plus de détails). Pour les journalistes et d'autres acteurs industriels habitués à utiliser ces données dans le cadre de leurs missions, l'article 193 de la loi de 2016 est ainsi perçu comme un nouveau verrou posé à l'accès aux données de santé (Malye and Vincent 2015 ; Entretien, industriel du numérique en santé, 22/07/2022).

Un autre changement important concerne plus directement les données extraites du SNIIRAM. Ainsi, jusqu'en 2016 et comme nous l'avons vu dans la partie précédente, les membres de l'IDS disposent d'un accès de droit à certains jeux de données, tels que les *datamarts* (données agrégées) et l'Echantillon Généraliste de Bénéficiaires (EGB). Or, la loi de 2016 révoque ces accès pour les membres de l'IDS, qui doivent désormais passer par des procédures d'autorisation classiques (éventuellement simplifiées et mutualisées – c'est-à-dire octroyées pour une période donnée sur un même type de traitement) afin d'accéder à ces données. Ces transformations concernent notamment les fédérations de professionnels de santé, les représentants de patients et les fédérations d'assureurs complémentaires. Pour l'ensemble de ces acteurs, la loi de 2016 marque ainsi une régression par rapport aux droits d'accès dont ils pouvaient se prévaloir jusqu'alors.

Les accès 'de droit' ne sont cependant pas révoqués pour tout le monde. En effet, il est prévu qu'une liste de 'services publics' se voie autorisée, par décret, un accès permanent à certaines données du SNDS.³¹ Cette liste inclut les directions générales du Ministère de la Santé (et en premier lieu la DREES, qui devait auparavant faire des demandes d'accès pour chaque projet)³², les Agences Régionales de Santé, la CNAMTS, l'ATIH, ainsi que de nombreuses agences publiques (la Haute Autorité de Santé, l'Agence nationale de sécurité du médicament et des produits de santé, l'Institut National du Cancer, l'Institut de recherche et documentation en économie de la santé, le Haut Conseil pour l'avenir de l'assurance maladie...). On y trouve également les acteurs de la communauté scientifique : les équipes de recherche de l'INSERM, les équipes de recherche des centres hospitaliers universitaires et des centres de lutte contre le cancer, les équipes de recherche et de formation de l'École des hautes études en santé publique.

'Et donc une des grandes **avancées** de la loi même si ça se voit moins de l'extérieur c'est les **accès permanents pour toute la sphère publique**, agences de santé etc., tous ceux qui sont vraiment au titre de **l'intérêt général**. [...] il y avait des accès permanents, et ça, pour tous les acteurs publics alors peut-être pas du point de vue des assureurs mais en tout cas du point de vue, ni des industriels, du point de vue de tous les acteurs publics qui sont nombreux et qui sont quand même les premiers à porter l'intérêt général c'était **une énorme évolution** ce décret avec les accès permanents, vraiment énorme, et si

³¹ Décret no 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « système national des données de santé »

³² Entretien, fonctionnaire, 17/06/2022 ; Entretien, représentant, secteur des assurances complémentaires, 16/06/2022

vous demandez à l'ANSM, à la HAS, donc je vous ai cité la DREES mais y a pas que la DREES y en a plein, si vous demandez à l'IRDES enfin y a plein d'institutions avec des chercheurs, qui sont des chercheurs de la sphère publique, qui ont des accès permanents et qui publient beaucoup maintenant , donc ça c'était une **énorme avancée** quand même.' (Entretien, fonctionnaire, 17/06/2022)

Enfin, alors que l'accès au PMSI par les organismes à but lucratif était autorisé pour le PMSI et non pour le SNIIRAM, la loi de 2016 harmonise les règles et autorise l'accès à ces bases pour tout type d'acteurs, du moment qu'il y a poursuite d'un 'intérêt public'. Cela signifie que les laboratoires pharmaceutiques peuvent désormais utiliser les données du SNIIRAM, à condition de passer par un intermédiaire (voir ci-dessous).

Les réaction des acteurs 'privés' au processus de formulation de la loi de 2016

Alors que l'État, la CNAMTS et certains représentants de la communauté scientifique s'accordent pour sécuriser davantage l'accès aux données de santé (voir partie précédente), les autres acteurs 'privés' ne tardent pas à réagir négativement aux nouvelles règles d'accès de l'article 193.

Ainsi, les représentants de patients ainsi que les fédérations d'assureurs complémentaires expriment clairement leur opposition à l'article de loi dans une lettre ouverte à la Ministre datée du 04 février 2015.³³ Selon ces organisations, cet article ne respecterait pas les propositions faites par la Commission Open data santé, et adopterait des mesures qui n'auraient pas été discutées au préalable par les acteurs de l'écosystème de la santé. Certains acteurs dénoncent ainsi un 'simulacre de concertation' (Entretien, fonctionnaire et agent CNAMTS, 20/07/2022). La lettre souligne enfin que le 'projet de loi n'affirme pas de droits nouveaux en matière d'accès aux données pour tous ceux qui gèrent le système de santé ou qui concourent à une meilleure connaissance en santé.' Ces acteurs dénoncent un manque de simplicité, d'efficacité et de lisibilité du régime d'accès aux données de santé. La lettre est signée par Danièle Desclerc Dulac (Présidente du Collectif Interassociatif en Santé), Etienne Caniard (Président de la FNMF) et Didier Sicard (Président du Comité d'experts de l'Institut des données de santé).

Du côté des professionnels de santé, les oppositions à l'article 193 interviennent un peu plus tard, avec le décret d'application de la loi publié en décembre 2016 qui liste les 'services publics' bénéficiant d'un accès permanent aux données du SNDS.³⁴ Pour rappel, les professionnels de santé étaient plutôt en accord avec la CNAMTS, l'État et les chercheurs pour œuvrer à une plus grande sécurisation des données de santé, en tout cas envers les laboratoires pharmaceutiques et les assureurs complémentaires (voir partie précédente). Cependant, les fédérations de représentants des professionnels de santé s'attendaient à faire partie de la liste de 'services publics' autorisés à bénéficier d'un accès permanent aux données du

³³ CISS, FNMF. Lettre ouverte à la Ministre. Objet : Accès aux données de santé. Paris, 04 février 2015.

³⁴ Décret no 2016-1871 du 26 décembre 2016 relatif au traitement de données à caractère personnel dénommé « système national des données de santé »

SNDS, se considérant comme les principaux ‘producteurs de données’ et comme assurant une mission de service public. Or, ils n’obtiennent pas cet accès permanent et doivent donc passer par les mêmes procédures d’accès que les représentants d’usagers et que représentants des complémentaires santé. Au moment de la découverte du décret, les professionnels de santé expriment finalement un positionnement proche des autres acteurs ex-membres de l’IDS :

‘Nous nous avons été très surpris surtout qu’on a été, on s’était dit, parce que il faut pas oublier non plus que c’est **nous**, par nos facturations par le soin, en soignant les gens on fait les facturations, **qui alimentons tout le système**. C’est quand même **aberrant**, que nous soyons obligés, soumis à des **obligations phénoménales**, qu’on ne nous reconnaît pas comme un organisme ayant disons, rendant un service public ça c’est quand même **aberrant**. Et donc on a vu tout se refermer et c’est là qu’on s’est retrouvé à devoir mettre un argumentaire pour pouvoir accéder aux données de santé.’

‘– Pierre-Louis Bras nous avait auditionné à l’époque, alors on a dit nos besoins, nos problématiques, elle n’a pas été prise en compte et quand **le décret est sorti** et qu’il a listé les gens qui avaient un accès permanent **nous n’étions pas dedans**.

– Et ça vous a surpris.

– J’ai été **surpris** au-delà d’être **mécontent** on était vraiment très surpris de voir qu’effectivement l’UNPS n’était pas dedans. Alors que c’est une institution créée par la loi, là **on n’a pas compris**.’

(Entretien, représentant professionnels de santé, 23/05/2022)

Le cas des assureurs complémentaires et des laboratoires pharmaceutiques

On note enfin que la loi de 2016 insiste de manière particulièrement visible sur l’exclusion de l’accès aux données de santé de deux types d’acteurs (‘structurellement méchants’, Entretien, ex-fonctionnaire, 06/04/2022) : les laboratoires pharmaceutiques et les assureurs complémentaires, qui doivent tout deux apporter la preuve qu’ils ne feront pas usage des données de santé pour les ‘finalités interdites’, également inscrites dans la loi. Ces finalités interdites concernent d’une part la sélection du risque, et d’autre part la promotion commerciale de produits pharmaceutiques. Ces deux acteurs peuvent néanmoins passer par des intermédiaires pour réaliser les études en leur nom, ce qui de fait n’impacte pas tellement les pratiques des industries pharmaceutiques habituées à recourir aux *Contract Research Organisations* (CRO) pour mener leurs études post-Autorisation de Mise sur le Marché (Entretien, représentant, industries pharmaceutiques, 21/07/2022 ; Entretien, secteur des assurances complémentaires, 15/06/2022 ; Entretien, fonctionnaire, 17/06/2022). Ainsi, pour certains représentants de l’industrie pharmaceutique, l’article 193 de la loi de 2016 relève quasiment du ‘non-événement’ :

‘– Ah oui la loi de modernisation du système de santé français, je pensais à Ma santé 2022. Alors **je sais plus** ce qu’il y avait là-dedans, sur la partie data.’

‘Mais là les données dont on parle effectivement nous **on ne s’est pas battu du tout**, alors je dis, nous industrie pharma, moi j’étais assez impliqué, hein sur tous ces sujets-là par ailleurs’

‘**Pas de souvenirs**, alors ce sujet-là nous a, **on l’a suivi** hein mais on n’a pas été, ce n’est pas des choses sur lesquelles on avait des amendements particuliers on s’est plutôt, on avait plutôt essayé de faire que ce cadre-là, qu’on puisse opérer dans ce cadre-là, mais on n’était pas... On avait peut-être d’autres sujets sur cette loi de Marisol qui nous ont heurtés, mais là rien ne résonne. La Ministre Touraine, devrais-je dire, nous ont heurtés mais **je n’ai pas souvenir**, non, non non.’

‘ça OK, très bien, nous, **ça nous va très bien** hein.’

(Entretien, représentant, industries pharmaceutiques, 21/07/2022)

‘[...] on dit que pour accéder aux données, d’abord il faut que ça soit avec un intérêt de son public, qu’il n’y ait pas une finalité interdite qui sont listés dans la loi. Donc ça, **c’est le cadre, c’est normal, c’est très bien, enfin il n’y a pas de problème**, et la 2ème chose c’est qu’on dit que soit ils doivent passer par un bureau d’études hein, soit si ils font un accès direct, ils doivent démontrer, ils doivent démontrer qu’en fait ils seront en capacité à tout moment de respecter la finalité interdite. Voilà mais ça c’est les conditions et les modalités d’application, **ce n’est pas gênant !** vous avez un nombre très important de demandes qui sont faites et et donc, je peux vous dire que **ça ne gêne pas et ça ne bloque absolument pas**, ni la recherche, ni l’utilisation. Il est normal et il est sain d’avoir un encadrement et des modalités d’accès. [...] Ce n’est pas parce qu’il y a un code de la route quand vous interdit de conduire ! C’est important.’

(Entretien, industrie pharmaceutique, 24/08/2022)

Pour certains, cette loi constitue même une avancée ‘évidente’ pour les industries pharmaceutiques (Entretien, industrie pharmaceutique, 24/08/2022, voir également : Entretien, secteur des assurances complémentaires, 15/06/2022), au sens où les organismes à but lucratif ne sont désormais plus exclus par principe de l’accès aux données du SNIIRAM.

Par contraste, la réception de la loi de 2016 pour les assureurs complémentaires est beaucoup plus difficile, d’une part parce que cette loi institue une régression par rapport aux précédentes régulations, mais également parce que contrairement aux laboratoires pharmaceutiques, les assureurs complémentaires ont peu d’alternatives en termes de sources de données de santé.³⁵ De ce fait, l’article 193 de la loi de

³⁵ En ce qui concerne l’aspect commercial de leurs activités, les industries pharmaceutiques peuvent en effet compter sur les données du Groupement d’Intérêt Economique (GIE) GERS – Groupement pour l’élaboration et la réalisation de statistiques (Entretien, industriel du numérique en santé, 22/07/2022 ; Entretien, représentant, industries pharmaceutiques, 21/07/2022)

2016 est perçu comme très problématique du côté des assurances complémentaires :

‘Moi je trouve ça quand même effectivement **incroyable** que la loi de 2016 ait **restreint** l'utilisation de de de, des données à des organismes tels qu'aussi bien les fédérations hospitalières que les Fédérations complémentaires hein, que de de France Assos Santé ou de l'UNPS, c'est quand même pour le pour moi ça c'est vraiment une **aberration complète.**’ (Entretien, secteur des assurances complémentaires, 15/06/2022)

‘[...] enfin en tout cas du point de vue de la complémentaire santé vraiment **procès** [...] **d'intention** sur, on va, on va interdire, on va interdire des usages dans la loi sur, pour l'industrie pharmaceutique sur les aspects marketing et pour les complémentaires santé sur la sélection du risque. Et là pour le coup des **articles purement politiques** parce que, c'était une façon de rappeler finalement les interdits mais qui étaient déjà interdits par la loi.’ (Entretien, représentant, secteur des assurances complémentaires, 16/06/2022)

‘Et là donc dans les questions qu'on s'est posées au moment où on écrivait l'article 193, donc gros débat autour des méthodologies de référence, procédures simplifiées, qu'on a beaucoup travaillées avec la CNIL, ça a été très dur avec la CNIL pour dire que quand même on mettait dans l'article 193 qu'on pouvait faire des accès simplifiés, validés par la CNIL, sur certaines données pas sur toutes mais quand il y avait un niveau de risque moindre, et par ailleurs, débat sur qui va quand même accéder aux données en direct, et où le sujet d'avoir des intermédiaires. Et donc là ça a été une décision politique quand même globalement prise par [...], de dire mais **on ne va pas faire les accès directs pour les industriels et pour les complémentaires.** Ça a été **extrêmement discuté** à l'époque, **pas trop par les industriels**, qui honnêtement ont l'habitude de travailler avec des intermédiaires, comment ils les appellent leurs boîtes d'intermédiaires, les CRO [Contract Research Organizations] non ? ça doit être ça peut-être, donc oui ils ont leur boîte voilà d'études, qui leur font les études donc pour eux c'était pas vraiment un problème, **le problème c'était les complémentaires** qui avaient je crois pris l'habitude en interne de traiter les données de l'EGB, et pas que les données des *datamarts* vraiment les données individuelles et qui là pour qui ça a été une régression, et qui l'ont vécu comme vous nous faites pas confiance, il y a un procès d'intention sur le fait que, ils nous disaient c'est pas vrai jamais on utilise ces données pour faire de la sélection du risque.[...] Et bon voilà mais enfin **avec les industriels franchement ça a pas été très compliqué.** Parce que c'est bien la preuve qu'on changeait peu leur modèle économique.’ (Entretien, fonctionnaire, 17/06/2022)

La technicisation de la gouvernance opérationnelle

Au regard de la gouvernance opérationnelle de l'accès aux données de santé, on note une certaine forme de technicisation de l'enjeu. L'accès au SNIIRAM était ainsi auparavant autorisé par la CNIL suite à un avis de l'IDS et de ses membres. Au moment de la rédaction de la loi de 2016, le rôle des membres de l'IDS dans

L'octroi des autorisations est perçu de manière négative pour les rédacteurs de la loi, qui insistent sur manque de 'compétences' des membres de l'IDS et sur l'aspect 'peu sérieux' de l'examen des demandes (Entretien, fonctionnaire, 17/06/2022 ; Entretien, ex-fonctionnaire, 06/04/2022).³⁶ Ainsi, avec la loi de 2016, le rôle de l'I(N)DS dans la gouvernance opérationnelle devient relativement marginal, puisqu'il n'est censé intervenir que dans le cadre où le projet d'accès ne rentrerait pas dans les standards scientifiques habituels en matière de méthodologie (Entretien, ex-fonctionnaire, 06/04/2022), ce qui concerne en l'espèce un nombre minoritaire de cas (Entretien, ex-agent CNAMTS, 22/09/2022). L'idée est donc davantage d'obtenir un 'regard technique, professionnel' (Entretien, fonctionnaire, 17/06/2022) et une 'procédure plus claire' (Entretien, ex-fonctionnaire, 06/04/2022) d'autorisation d'accès aux données de santé. Le principal comité chargé d'émettre un avis sur l'accès aux données de santé est en effet le CEREES (comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé), qui juge les demandes d'accès avec une approche essentiellement scientifique (Entretien, secteur des assurances complémentaires, 15/06/2022 ; Entretien, ex-agent CNAMTS, 22/09/2022).³⁷ Les membres de ce comité sont sélectionnés sur la base de leur 'compétences en matière de recherche dans les domaines de la santé, de l'épidémiologie, de la génétique, de la biostatistique et des sciences humaines et sociales et en matière de traitement des données à caractère personnel'³⁸.

Cette technicisation scientifique du comité chargé d'instruire les demandes a des répercussions très concrètes, par exemple chez les représentants de professionnels de santé qui, en plus de se voir refuser l'accès de droit aux données du SNDS, doivent produire un type d'argumentaire qui leur est peu familier :

'Par exemple l'autre sujet qu'on a eu c'était avec les **professionnels libéraux**, ça a été un peu **catastrophique**. Ils ont déposé un dossier, si vous voulez le problème qu'on a c'est qu'on a fait un comité, le CEREES qui est scientifique et donc ils font comme s'ils faisaient un appel d'offres de recherche, mais les professionnels libéraux quand ils demandent à avoir accès aux données **ils peuvent pas faire un dossier de recherche**, donc ils écrivent des choses, peut-être un peu [...] du genre 'Monsieur machin' fait ça etc et les gens ils tombent là-dessus [...] Moi j'ai mouillé la chemise énormément pour eux parce que si vous voulez ils sont, alors là pour eux c'était des injustices totales, c'est-à-dire qu'on avait fabriqué un machin qui finalement

³⁶ Du fait de ce manque de 'compétences', les rédacteurs de la loi perçoivent l'IDS comme une organisation qui octroie des autorisations de manière relativement artisanale et à un cercle limité de connaissances (partage des données 'entre copains', Entretien, fonctionnaire, 17/06/2022). Pour rappel, la procédure de validation d'une demande à l'IDS est la suivante : étude du dossier par un 'comité d'expert' composé de personnalités scientifiques, puis vote par le bureau de l'IDS (représentant des différents membres de l'IDS).

³⁷ Voir également : Morange Pierre (2017). *Rapport d'information sur les données médicales personnelles inter-régimes détenues par l'assurance maladie, versées au SNIIRAM puis au Système national des données de santé (SNDS)*. Paris : publications de l'Assemblée Nationale, p.105

³⁸ Décret n° 2016-1872 du 26 décembre 2016 modifiant le décret n° 2005-1309 du 20 octobre 2005 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés

contribuait à les exclure eux, [...] le problème c'est que les **laboratoires pharmaceutiques eux ont aucun problème avec ça, c'est nickel chrome**, eux ça passe. **Mais les gens qui ne savent pas faire des dossiers un peu propres, administratifs**, c'est un peu ça la médecine de ville c'est vrai pour tout, c'est vrai pour l'accès aux données mais c'est aussi vrai quand ils répondent à des appels à candidature pour des appels à projet et tout ça, c'est **qu'ils n'ont pas tellement les codes**, [?] Donc voilà donc par exemple avec l'UNPS l'union nationale des professionnels de santé, on a eu chaud hein ! Ils étaient proches d'être rejetés, ça le CEREES voulait les virer.'

'Donc on a dit bah oui [?] donc ils ont fait leur dossier, le dossier était [Rires] comment dire, assez **naïf** quoi, **ils savaient pas trop comment faire un dossier**, des petits verbatims de gens qui avaient utilisés les données, je pense que le **CEREES ils sont tombés de leur chaise**, et personne n'avait fait non plus de formulaire pour dire voilà comment on présente un dossier hein, [?] formaté recherche, ils attendaient des choses de voir des protocoles [?] moi j'ai toujours trouvé ça pas bien, parce que en fait ça veut dire qu'on barrait, d'une certaine manière on voulait ouvert, mais **on barrait l'accès** par des choses assez subtiles comme ça genre la composition d'un comité, **des gens qui étaient pas du milieu**, du sérail.'

(Entretien, ex-agent CNAMTS, 22/09/2022)

'[...] quand on a dû faire **l'argumentaire** ça a **pas été facile** [...]'
(Entretien, représentant professionnels de santé, 23/05/2022)

Par contraste, les organisations qui disposent d'experts scientifiques dans l'usage des données de santé ont plus de facilité à produire des dossiers qui correspondent aux normes implicitement attendues (Entretien, secteur des assurances complémentaires, 15/06/2022). À ce titre, le cas des laboratoires pharmaceutiques constitue un paradoxe particulièrement intéressant : alors que la loi visait précisément à exclure ce type d'acteur de l'accès aux données de santé (et ce malgré une ouverture du SNDS aux organismes à but lucratif), la congruence entre les règles adoptées et le mode de fonctionnement de ces industries (recours à des organismes tiers pour la réalisation de leurs projets de recherche) conduit à propulser les laboratoires pharmaceutiques au premier rang des utilisateurs des données de santé (Entretien, industrie pharmaceutique, 24/08/2022 ; Entretien, secteur des assurances complémentaires, 15/06/2022).

La gouvernance stratégique : reprise en main par l'État

Avant la réforme de 2016, la gouvernance stratégique était, formellement, assurée par la COPIIR, un comité rassemblant la CNAMTS, l'État et les professionnels de santé (voir Working paper n°1). Cependant, au cours des années 2000, l'Institut des Données de Santé (dont les membres représentent divers acteurs tels les patients, les fédérations hospitalières, les assureurs complémentaires) vient se poser en concurrent de ce COPIIR, en militant activement pour l'ouverture des données de santé.

La réforme de 2016 remet à plat cette gouvernance stratégique, et déplace son centre de gravité au sein de l'État, et plus précisément au sein du Ministère de la

Santé (Entretien, représentant, secteur des assurances complémentaires, 16/06/2022 ; Entretien, agent CNAMTS, 21/04/2022). Un nouveau comité de pilotage resserré est ainsi mis en place autour des directions du Ministère de la Santé et des gestionnaires (ou ‘producteurs’) des bases de données : direction de la sécurité sociale, direction générale de l’offre de soins, direction générale de la santé, délégation à la stratégie des systèmes d’information de santé, CNAMTS, ATIH, INSERM-CepiDC, Haut fonctionnaire de défense et de sécurité, direction générale de la cohésion sociale, direction générale pour la recherche et l’innovation, et présidence de l’IDS (puis de l’INDS). Par ailleurs, le secrétariat et le pilotage de ce comité est directement assuré par la mission de mise en œuvre de l’article 193 dirigée par Mylène Girard à la DREES.³⁹ C’est donc désormais la DREES, ‘reprenant les compétences du COPIIR’ qui définit ‘les priorités stratégiques en matière de donnée de santé, informe des différentes évolutions et de l’avancée des décisions et arbitre les sujets de débat’.⁴⁰

Ce faisant, ces transformations mettent fin à l’ambiguïté institutionnelle de l’IDS, qui évoluait à la frontière de la gouvernance opérationnelle et stratégique. Opposée à cette réforme, la direction de l’Institut des Données de Santé est entièrement renouvelée à l’occasion de la transformation de cette organisation en ‘Institut National des Données de Santé’ (Entretien, fonctionnaire, 14/04/2022 ; Entretien, représentant, secteur des assurances complémentaires, 16/06/2022).

Les enjeux pratiques : des dynamiques ambivalentes

Enfin, les enjeux pratiques d’accès aux données de santé sont marqués par plusieurs mouvements ambivalents. La transformation de l’Institut des Données de Santé (IDS) en Institut National des Données de Santé (INDS) s’accompagne ainsi d’un arrêt de la publication des outils de facilitation d’accès aux données de santé destinés à ses membres (notamment les tableaux de bord : Entretien, agent CNAMTS, 21/04/2022 ; Entretien, représentant, association d’usagers, 30/06/2022 ; Entretien, représentant professionnels de santé, 23/05/2022). En parallèle, les données agrégées ne comportant pas de risque de réidentification sont présentées comme devant être accessibles en *open data*, objectif à ce jour seulement partiellement réalisé (Entretien, industriel du numérique en santé, 22/07/2022). L’accès aux données du PMSI, davantage sécurisé, est *de facto* plus difficile que lorsque les données étaient simplement copiées sur disquette. De plus, il semblerait que certaines données avec un niveau assez fin de granularité, auparavant disponibles sur le site de l’ATIH, aient été remplacées par des données plus agrégées, présentant une profondeur historique plus réduite, avec des règles de floutage en cas de résultats comportant un nombre de cas inférieur à 11 (Entretien, industriel du numérique en santé, 22/07/2022).

D’un autre côté, plusieurs mesures sont prises pour enrichir les données de santé, notamment en facilitant l’utilisation du NIR (Numéro de Sécurité Sociale) pour les

³⁹ Morange Pierre (2017). *Rapport d’information sur les données médicales personnelles inter-régimes détenues par l’assurance maladie, versées au SNIIRAM puis au Système national des données de santé (SNDS)*. Paris : publications de l’Assemblée Nationale, p.97

⁴⁰ Morange Pierre (2017). *Rapport d’information sur les données médicales personnelles inter-régimes détenues par l’assurance maladie, versées au SNIIRAM puis au Système national des données de santé (SNDS)*. Paris : publications de l’Assemblée Nationale, p.97

appariements et également grâce à la création du SNDS, qui contient alors trois bases de données appareillées de manière permanente : le SNIIRAM, le PMSI, et les données du CépiDc (Centre d'épidémiologie sur les causes médicales de décès – INSERM).

En conclusion, le début et le milieu des années 2010 semblent marqués, à première vue, par un paradoxe : alors que l'enjeu des données de santé est mis à l'agenda sous l'angle de l'ouverture, c'est plutôt une dynamique de fermeture qui caractérise les transformations adoptées dans la loi de 2016 de modernisation de notre système de santé. Ces dynamiques de fermeture s'appliquent en particulier aux anciens membres de l'Institut des Données de Santé (représentants des professionnels de santé, des assureurs complémentaires et des patients), qui voient leurs accès de droit au SNIIRAM révoqués. Désormais, ces acteurs doivent passer par une procédure standard d'accès aux données de santé, au même titre que tout acteur privé. *In fine*, les acteurs privés qui parviennent à faire un plus grand usage des données de santé sont ceux dont les pratiques se révèlent congruentes avec la réglementation en vigueur, ce qui est, par exemple, le cas des laboratoires pharmaceutiques. En parallèle, la loi de 2016 ouvre des accès de droit pour tout un ensemble de 'services publics', notamment les directions générales du Ministère de la Santé et de nombreuses agences publiques.

Ces transformations s'expliquent par le processus d'autonomisation et de politisation du secteur de l'accès aux données de santé. Dans les années 1990 et 2000, l'enjeu des données de santé est davantage traité dans le cadre de problématiques plus générales, et ne fait pas vraiment l'objet d'interventions fortes du politique, la principale dynamique décisionnelle étant alors guidée par les luttes institutionnelles entre l'IDS, la CNAMTS et l'État. Au début des années 2010 se pose la question de la nature de la 'doctrine de l'État' en matière d'accès aux données de santé. Il s'agit alors pour le Ministère de la Santé d'adopter une posture globale face à des revendications de nombreux acteurs de la société civile. Au niveau du Ministère de la Santé, les acteurs se positionnent en faveur d'une plus grande sécurisation de l'accès aux données de santé, ce qui fait écho à la position traditionnelle de prudence de la CNAMTS et aux besoins de certains représentants de la communauté scientifique. Face à ces trois 'poids lourds', les autres acteurs (notamment les représentants des professionnels de santé, des patients et des assureurs complémentaires) ne parviennent pas à obtenir gain de cause, et subissent une régression dans les accès de droit qui leur étaient auparavant octroyés.

Tableau. La régulation de l'accès aux données de santé dans les années 2010

Règles d'accès	Gouvernance opérationnelle	Gouvernance stratégique	Enjeux pratiques
Evolution par rapport à la période précédente			
Fermeture/Ouverture	Fermeture	Fermeture	Ouverture/Fermeture
Fermeture : alignement des règles du PMSI sur celles du SNIIRAM, fin de	Technicisation de l'enjeu par la création du CE(S)REES,	Recentralisation du pilotage stratégique au sein de l'État,	Ouverture : Facilitation de l'utilisation du NIR pour les

l'accès de droit pour les membres de l'IDS, exclusion de l'accès direct pour les assureurs complémentaires et les industries du médicament Ouverture : Extension du nombre d'organismes publics avec accès de droit, fin de la dichotomie acteurs à but lucratif/non lucratif pour le SNIIRAM	comité technique qui donne un avis scientifique préalable à l'autorisation de la CNIL. L'INDS peut donner un avis dans la mesure où la demande ne respecte pas les standards scientifiques habituels	la DREES occupe désormais les missions d'animation et de pilotage de la gouvernance stratégique	appariements, création du SNDS, les jeux de données anonymes sont supposés être accessibles en open data Fermeture : fin de la publication des outils de facilitation de l'accès aux données de santé produits par l'IDS (tableaux de bord), fin de la copie des données du PMSI sur disquette
--	--	---	--

Tableau. Les accès aux bases de données de santé selon les types d'acteurs (début – milieu des années 2010)

	SNDS : SNIIRAM + PMSI + CépiDC (procédure : CESREES/INDS + CNIL)	SNDS accès de droit (jeux de données)	SNDS accès de droit (totalité)
Acteurs publics (Ministères et services)	Oui	Oui (acteurs listés dans le décret n° 2016-1871 du 26 décembre 2016)	Oui (acteurs listés dans le décret n° 2016-1871 du 26 décembre 2016)
Agences publiques	Oui	Oui (acteurs listés dans le décret n° 2016-1871 du 26 décembre 2016)	Oui (acteurs listés dans le décret n° 2016-1871 du 26 décembre 2016)
Acteurs de la recherche : INSERM, CHU, EHESP	Oui	Oui (acteurs listés dans le décret n° 2016-1871 du 26 décembre 2016)	Non
Professionnels de santé	Oui	Non	Non
Organisations de patients	Oui	Non	Non
Assurances complémentaires	Oui	Non	Non
Laboratoires pharmaceutiques	Oui	Non	Non

II. Vers une industrialisation de l'usage des données de santé ? (fin des années 2010, début des années 2020)

Après l'adoption de la loi de modernisation de notre système de santé en 2016 et le changement de gouvernement en 2017, le secteur des données de santé connaît de nouvelles transformations, principalement d'ordre 'technique'. En effet, en 2019 l'Institut National des Données de Santé est transformé en Plateforme des Données de Santé, (ou Health Data Hub, HDH), plateforme dont le but est de permettre une réelle industrialisation de l'usage des données de santé. Ces changements sont notamment permis par la survenue d'une nouvelle 'fenêtre d'opportunité' avec la mise à l'agenda de l'enjeu de l'intelligence artificielle au niveau gouvernemental. 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 pour héberger les données de santé.

a) L'enjeu de l'intelligence artificielle : une nouvelle fenêtre d'opportunité pour les données de santé

Du Big Data à l'Intelligence artificielle, une continuité dans le cadrage politique des données de santé

Suite à l'élection d'Emmanuel Macron, le discours politique en matière de données de santé poursuit la sur la lancée 'open data' portée par Marisol Touraine, avec un focus davantage orienté vers l'industrialisation de l'usage des données de santé. Si au début des années 2010 le débat sur les données était marqué par l'idée de 'Big Data' et 'd'open data', la fin de la décennie voit émerger des discours de plus en plus nombreux sur l'Intelligence Artificielle (IA) et sur les potentialités offertes par un usage et une exploration intensifs des entrepôts de données. Dans cette optique, la France souhaite se positionner en tant que 'leader européen' ; en atteste son attitude proactive vis-à-vis de la création d'un 'Espace européen des données de santé'.⁴¹ En termes de cadrage, on assiste donc à une forte continuité entre la manière dont est posé le problème au début des années 2010 et la fin de cette même décennie.

De 'l'innovation de garage' à la DREES à la création du Health Data Hub

Au-delà de cette continuité en termes de cadrage, plusieurs événements concomitants contribuent à donner un nouveau souffle à la question des enjeux pratiques de l'accès aux données de santé, ce qui donne lieu en 2019 à la création du Health Data Hub (HDH).

Une première partie de ces éléments est liée aux travaux de la DREES réalisés dans suite directe de l'adoption de la loi de 2016 de modernisation de notre système de

⁴¹ Suite à l'initiative TEHDAS (Towards the European Health Data Space), la France et la Finlande ont signé un Memorandum of Understanding afin de favoriser la coopération entre les plateformes de données de santé françaises et finlandaises. L'objectif est *in fine* de travailler sur la mutualisation de l'accès aux données de santé et de préfigurer l'Espace européen des données de santé. Source : Health Data Hub, 'Le Health Data Hub signe son premier accord de collaboration transfrontalier', Site du Health Data Hub, publié le 24/11/2021 ; Disponible sur : <https://www.health-data-hub.fr/actualites/parteneriat-findata>

santé. Franck Von Lennep, alors directeur de la DREES, missionne une petite équipe afin de faciliter l'accès de ses propres services au SNDS. Il apparaît en effet que techniquement, cet accès est rendu très difficile, notamment au regard des appariements. Par ailleurs, le portail de la CNAMTS ne permettrait pas de réaliser certains traitements innovants (Entretien, fonctionnaire, Health Data Hub, 24/08/2022). Cette mission met ainsi en place un système de portail en ligne pour faciliter l'accès des membres de la DREES au SNDS.

De manière plus générale, les difficultés d'accès sont soulignées par de nombreux acteurs, ce qui contribue à mettre sur l'agenda l'enjeu pratique de l'accès aux données du SNIIRAM.

'Mais d'un autre côté ça entraîne beaucoup de problématiques parce que il n'y a **pas forcément beaucoup de cohérence entre les délais** pour l'obtention des données, en passant par ce processus réglementaire et en suivant toutes les étapes et les demandes faites par ailleurs par les pouvoirs publics ou les instances de régulation qui demandent justement d'avoir, de disposer de données pour valider leur décision de renouvellement d'inscription de produits, de remboursement de... je ne sais pas d'accès en fait au marché sur une population cible, enfin voilà en fait et ça, c'est très problématique parce qu'en plus le **délai d'accès a tendance à s'allonger.**' (Entretien, industriel du numérique en santé, 22/07/2022)

'Du point de vue de la sphère privée, effectivement on n'a pas eu, sur le papier c'était théoriquement possible mais en réalité la procédure d'analyse à l'usage il me semble qu'elle n'a **pas permis quand même beaucoup d'ouverture** et puis surtout elle n'a pas réglé la **question de l'engorgement des structures d'instruction** en fait. D'où l'idée après du Health Data Hub de dire mettons des moyens pour l'instruction.' (Entretien, représentant, secteur des assurances complémentaires, 16/06/2022)

'Donc ce dont souffrent les entreprises du médicament aujourd'hui, c'est c'est des **délais de d'accès à ces données**, même si c'est de l'accès indirect via une structure tierce, je veux dire encore une fois ce qui pour nous est important, c'est plus le résultat d'exploitation des données, mais c'est quand même compliqué d'accéder effectivement... car il y a des **questions de délais** qui encore une fois doivent se regarder en compétition par rapport à d'autres pays, hein. Je veux dire quand certains pays, il suffit de d'envoyer une demande et 2 semaines plus tard, donc j'exagère mais, vous avez droit à cet accès à un set de données qui vous permet de faire vos travaux, ici ça va prendre 9 mois.' (Entretien, représentant, industries pharmaceutiques, 21/07/2022)

La CNAMTS quant à elle, semble manquer de ressources pour répondre aux demandes d'accès aux données du SNDS (Entretien, fonctionnaire, 17/06/2022 ; Entretien, représentant, secteur des assurances complémentaires, 16/06/2022 ; Entretien, industrie pharmaceutique, 24/08/2022 ; Entretien, fonctionnaire, Health Data Hub, 24/08/2022).

Alors que l'objet initial de la mission lancée par Franck Von Lennep est de faciliter l'utilisation des données du SNDS pour la DREES uniquement, le projet prend

finalement de l'ampleur à la faveur d'un changement de direction de la DREES. En 2017, Jean-Marc Aubert remplace en effet Franck Von Lennepe à la tête de cette direction, et décide de déposer une candidature pour ce projet au fonds pour la transformation de l'action publique.⁴²

'Et à et à peu près à ce moment-là, Jean-Marc Aubert est arrivé, en tant que directeur donc c'était fin... 2017 et **il a dit bah en fait ce serait super de voir plus grand** dans ce projet d'utilisations innovantes des données du SNDS et du coup, d'écrire un projet pour **candidater au fond de transformation de l'action publique**, et donc c'est ce qui a été fait à l'époque' (Entretien, fonctionnaire, 24/08/2022 ; voir également : Entretien, fonctionnaire, 17/06/2022)

Jean-Marc Aubert vient du secteur privé où il occupait depuis 2013 de hautes fonctions au sein d'Iqvia, entreprise américaine issue de la fusion entre IMS Health et Quintiles, spécialistes de l'exploitation de données de santé (notamment pour le compte des industries pharmaceutiques). Du fait de son parcours, Jean-Marc Aubert est rapidement soupçonné de conflits d'intérêt et est accusé de 'pantouflage' lorsqu'il retourne à Iqvia pour prendre les fonctions de Président de la filiale française en décembre 2019.⁴³

Parallèlement à ces développements est publié en 2017 le rapport rédigé par Cédric Villani intitulé 'Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne'.⁴⁴ Très concrètement, ce rapport va permettre de mettre en lumière les projets contribuant au développement des usages de l'intelligence artificielle, et c'est ainsi que le projet initial de la DREES va être repris et généralisé pour devenir le Health Data Hub, qui obtient les financements du fonds de transformation de l'action publique (Entretien, fonctionnaire, Health Data Hub, 24/08/2022). On retrouve, comme dans les années 1990 avec les ordonnances Juppé, un réel effet de 'fenêtre d'opportunité' qui met sur le devant de la scène des projets relativement techniques entrepris au sein des administrations.

'Cédric Villani avait essayé **d'embarquer en effet dans son rapport l'ensemble des sujets santé** aussi, et il y avait eu un moment et d'ailleurs il y a eu si je me souviens bien peut-être en 2018 c'était au Collège de France je crois, il y avait eu un **séminaire sur l'intelligence artificielle**, qui avait je pense, **Cédric Villani** était là voilà, et le **Président de la République** était venu, et avait **annoncé la création du Health Data Hub** à ce moment-là, donc en fait le lien a été fait,⁴⁵ [...] c'était quand même un peu dans le, dans le paysage on

⁴² Lancé fin de 2017, ce fonds a pour objectif de financer des 'projets innovants' de transformation de l'action publique 'au service des politiques prioritaires du gouvernement'. Pour plus d'informations, voir : <https://www.modernisation.gouv.fr/transformer-laction-publique/fonds-pour-la-transformation-de-laction-publique>

⁴³ Foucart Stéphane et Stéphane Horel, 'Données de santé : conflit d'intérêts au cœur de la nouvelle plate-forme', *Le Monde*, publié le jeudi 26 décembre 2019, p. ECO14 ; voir également : Entretien, représentant, association d'usagers, 30/06/2022

⁴⁴ Villani Cédric (2018). *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*. Rapport au Premier Ministre, mission parlementaire du 8 septembre 2017 au 8 mars 2018. 235 p.

⁴⁵ Annonce faite par le Président de la République en clôture du 'Sommet sur l'intelligence artificielle' et de la présentation du rapport Villani au Collège de France le 29 mars 2018.

peut dire **dans l'ombrelle générale de Cédric Villani** qu'on peut replacer le Health Data Hub. Mais ça c'est je dirais que c'est quand même un peu le, le, un contexte un peu parce que il y avait Cédric Villani à ce moment-là mais sinon on l'aurait fait je pense, honnêtement je pense qu'on l'aurait fait.' (Entretien, fonctionnaire, 17/06/2022)

'[...] en **parallèle de ça et sans qu'on s'en rende compte** du tout Cédric Villani, le député, lui travaillait sur comment massifier les données dans différents secteurs pour promouvoir l'intelligence artificielle et rendre possible les cas d'usage d'intelligence artificielle. Et je sais que des gens de la DREES et Mme X notamment, avaient été auditionnés par Cédric Villani, mais moi je suis pas du tout rentré en contact avec lui à cette période. Mais quand il a publié son rapport où il disait ce serait super de faire des plateformes de données dans différents secteurs prioritaires, y compris la santé bah **le gouvernement s'est un peu tourné vers l'administration** pour savoir comment interpréter ces recommandations et nous on a dit bah écoutez, nous en fait, on a fait une sorte de **prototype à l'échelle de la DREES** avec un serveur Big data, des données du SNDS et on a même candidaté au fonds de transformation de l'action publique. Et c'est un petit peu comme ça que les trucs se sont, connectés donc ce n'est pas une vision stratégique, politique publique réfléchi par plein de gens [Rires]. C'est c'est plus un **concours de circonstances**, un **alignement de planètes**, mais hyper positif parce que du coup, à la suite de ça, ça s'est vachement accéléré puisque du coup on nous a confié la mise en œuvre bah de la reco [recommandation] de Cédric Villani, donc on a organisé une mission de préfiguration avec des experts en pilotage parce que ce qu'on avait porté à la DREES jusque-là, c'était vraiment de l'innovation de de garage.' (Entretien, fonctionnaire, 24/08/2022)

Aussi le rapport de Cédric Villani est-il présenté comme une référence par plusieurs acteurs du secteur des données de santé (Entretien, secteur des assurances complémentaires, 15/06/2022 ; Entretien, fonctionnaire, 17/06/2022 ; Entretien, agent CNAMTS, 21/04/2022 ; Entretien, fonctionnaire, Health Data Hub, 24/08/2022). Pour d'autres, il existerait par ailleurs une continuité entre les développements du Health Data Hub et le positionnement d'Emmanuel Macron lorsqu'il était Ministre de l'Economie, de l'Industrie et du Numérique (voir débats à propos de l'attribution du portefeuille ministériel lors de la rédaction de la loi de modernisation de notre système de santé de 2016, section I.b) :

'Entre nous vous savez très bien que le Health Data Hub c'est **l'enfant de Macron** quand même c'est l'enfant de Macron [...]' (Entretien, représentant professionnels de santé, 23/05/2022)

'[...] mais **Macron** était déjà sur un logiciel différent [?] il faut pas oublier qu'il était à Bercy, et qu'il **s'était battu** pour que, pour récupérer à Bercy les données de santé hein, le **contrôle**.' (Entretien, ex-conseiller de Marisol Touraine, 19/07/2022)

Les acteurs 'privés' dans la création du Health Data Hub

Lors de la création du Health Data Hub, le rôle de deux acteurs (en plus des administrations) apparaît comme particulièrement prégnant : les chercheurs et les industriels. À ce titre, la composition de la mission de préfiguration du HDH est particulièrement parlante ; sur les 17 membres, cinq sont des chercheurs, quatre des industriels, et six des acteurs publics.

Tableau. Composition de la mission de préfiguration du Health Data Hub

	Nom	Organisation	Type acteur
Pilotes de la mission	Marc Cuggia	CHU de Rennes	Chercheur
	Dominique Polton	INDS	Acteur public
	Gilles Wainrib	Owkin	Industriel
Rapporteure	Stéphanie Combes	DREES	Acteur public
Membres du groupe de travail	Emmanuel Bacry	CNRS, Ecole Polytechnique	Chercheur
	Muriel Barlet	DREES	Acteur public
	Hugues Berry	Inria	Chercheur
	Thomas Borrel	LEEEM	Industriel
	Isabelle Gentil	DSSIS	Acteur public
	Claude Gissot	CNAM	Caisse d'assurance maladie
	Didier Guillemot	Institut Pasteur, l'Inserm, l'Université de Versailles Saint Quentin	Chercheur
	Eric Guittet	DGRI	Acteur public
	Jérôme Kalifa	Lixoft	Industriel
	Yves Levy	Inserm	Chercheur
	Caroline Noublanche	Apricity	Industriel
		Elisa Salamanca	AP-HP
	Henri Verdier	DINSIC	Acteur public

Source : Cuggia, Marc, Dominique Polton, Gilles Wainrib et Stéphanie Combes (2018). *Health Data Hub - Mission de préfiguration*. Paris : Ministère des Solidarités et de la Santé, p.101

Si les chercheurs et les Professeur des universités-praticien hospitalier (PU-PH) soutiennent d'abord très fortement le projet du Health Data Hub, ils finissent cependant par s'y opposer (et pour certains, par créer des solutions alternatives comme le 'Ouest DataHub', plateforme interrégionale des données de santé des Hôpitaux Universitaires du Grand Ouest). Le motif de ce retournement de situation semble principalement lié à la volonté des administrations d'inclure les cohortes de ces chercheurs dans le Health Data Hub. En effet, si les chercheurs soutiennent les initiatives ayant vocation à faciliter l'appariement et l'accès à

d'autres données que les leurs, ils souhaitent néanmoins garder un certain contrôle sur l'accès à leurs propres cohortes, ce qui n'est pas dans l'objectif du HDH.⁴⁶ Aussi le soutien initial des communautés de chercheurs tend-il à s'étioler au cours de la mission de préfiguration du HDH (Entretien, juriste spécialiste des données de santé, 15/06/2022 ; Entretien, fonctionnaire, Health Data Hub, 24/08/2022, voir également les futurs travaux de Laurène Assailly).

Du côté des laboratoires pharmaceutiques, on note également un soutien très fort à la création du Health Data Hub qui lui, ne faiblit pas (bien que cette catégorie d'acteurs témoigne aujourd'hui d'une insatisfaction vis-à-vis de la mise en œuvre encore peu aboutie du HDH)⁴⁷. Ainsi, dès les annonces de création du Health Data Hub en Mars 2018, les industriels se positionnent et témoignent leur soutien lors du huitième Conseil stratégique des industries de santé en juillet 2018,⁴⁸ dont le rapport de synthèse présente le HDH comme l'une de ses 'mesures phares'.⁴⁹ Certains acteurs évoquent ainsi une réelle 'synergie' entre les initiatives du gouvernement et les intérêts de l'industrie pharmaceutique :

'Donc c'était pour vous dire que petit un effectivement sous le quinquennat précédent, dans le quinquennat précédent, il y a un certain nombre d'**impulsions**, plutôt bien coordonnées par le **gouvernement** ou au niveau de l'État, par Macron, et **synergique avec des centres d'intérêt du milieu des entreprises du médicament** puisque porté dans ces conseils stratégiques des industries de santé.' (Entretien, représentant, industries pharmaceutiques, 21/07/2022)

b) Le Health Data Hub : la montée en puissance des 'enjeux pratiques' de l'accès aux données de santé

Industrialiser l'usage des données de santé via une plateforme technique

L'objectif principal de la réforme conduite sous le dernier quinquennat est de transformer l'INDS en plateforme des données de santé, ou Health Data Hub (HDH). Au-delà de la modification de l'acronyme, la création du HDH, qui intervient par arrêté en 2019,⁵⁰ témoigne de la volonté de porter l'usage des

⁴⁶ Une fois copiée dans le HDH, la procédure d'accès aux données de santé est gérée par le CESREES et la CNIL, et non plus par les producteurs de données.

⁴⁷ Entretien, représentant, industries pharmaceutiques, 21/07/2022 ; Entretien, industrie pharmaceutique, 24/08/2022, voir également : AFCROs, 'Lettre ouverte à Monsieur Olivier Véran, Monsieur Thomas Fatôme et Madame Stéphanie Combes', publiée le 15 décembre 2021 sur le site de l'Association française des *Contract Research Organisations* (CROs), disponible sur : <https://www.afcros.com/lettre-ouverte-a-mm-olivier-veran-et-thomas-fatome-et-madame-stephania-combes/>

⁴⁸ Le Conseil stratégique des industries de santé (CSIS), créé en 2004, est lieu d'échange entre les pouvoirs publics et les industries de la santé, rattaché au portefeuille du Premier Ministre. Source : Jégou, Jean-Jacques (2008). *Rapport d'information fait au nom de la commission des Finances, du contrôle budgétaire et des comptes économiques de la Nation sur la taxation de l'industrie du médicament*. Paris : publications du Sénat, p.84

⁴⁹ Conseil stratégique des industries de santé (2018). *Notre ambition pour les industries de santé, 8ème Conseil stratégique des industries de santé (CSIS)*. Paris : Service de presse de Matignon, 43p.

⁵⁰ Arrêté du 29 novembre 2019 portant approbation d'un avenant à la convention constitutive du groupement d'intérêt public « Institut national des données de santé » portant création du groupement d'intérêt public « Plateforme des données de santé »

données de santé à un niveau industriel, c'est-à-dire de créer les conditions d'un usage massif des données de santé. Alors que l'IDS puis l'INDS servaient simplement d'intermédiaires entre le demandeur de données et les gestionnaires des bases de données (la CNAMTS, l'ATIH et le CépiDc principalement), le HDH est supposé au contraire agir comme un réel offreur de données. Pour ce faire, il est prévu que le HDH dispose lui-même d'une copie des bases de données, à partir de laquelle il pourra soit donner l'accès aux demandeurs, soit réaliser lui-même des études. L'objectif du HDH est également de centraliser le plus grand nombre possible de bases de données et de les appareiller de manière permanente (sans avoir à demander une autorisation de la CNIL pour chaque appariement), afin de disposer de données par exemple environnementales qui pourraient être mises en relation avec les données de santé. En somme, il s'agit de construire la 'capacité à avoir d'immenses entrepôts de données d'une très grande richesse' (Entretien, fonctionnaire, 17/06/2022).

De toute évidence, la construction d'une telle plateforme requiert d'importants moyens, aussi le budget annuel du HDH correspond-il à 30 fois celui de l'IDS pendant les années 2000, tandis que son personnel a été multiplié par 10 (une centaine de personnes en comptant les externes).⁵¹ L'ensemble de ces transformations, certes techniques, marquent un (relatif) tournant dans la gestion des bases de données qui prévalait jusqu'alors ; en effet, les gestionnaires historiques des bases de données (comme la CNAMTS) ne sont désormais plus les seuls à disposer de l'intégralité de la base qu'ils alimentent.

La transformation de l'IDS en plateforme technique : l'éradication définitive des ambiguïtés institutionnelles des années 2000

Au niveau de la gouvernance, on peut noter que le HDH associe une palette très large d'acteurs (56 parties prenantes), plus large que celle de l'I(N)DS puisque les industriels sont également représentés dans son 'assemblée générale'.⁵² Cependant, contrairement au rôle des parties prenantes du temps de l'IDS, cette assemblée générale n'a pas de pouvoir ni sur la gouvernance opérationnelle (qui est assurée par le CESREES et la CNIL, voir section I.c) ni sur la gouvernance stratégique (comité piloté par la DREES, voir section I.c)⁵³. Il s'agit plutôt d'avoir un lieu

⁵¹ Le budget actuel est toutefois un budget de lancement temporaire (environ 80 millions d'euros sur 4 ans), et nous ne disposons pas encore d'informations sur le montant du budget de fonctionnement une fois que ce premier budget sera épuisé. Source : Entretien, fonctionnaire, Health Data Hub, 24/08/2022

⁵² Les membres du HDH sont : les différents Ministères intéressés, les caisses nationales d'assurance maladie, les représentants des caisses d'assurance complémentaire, des instituts de recherche (INSERM, CNRS...), les fédérations hospitalières, les représentants des professionnels de santé, France Assos Santé, les agences sanitaires, et les représentants d'industriels (LEEM, AFCROs., SNITEM, Syntec, France Digitale). Source : Site web du Health Data Hub : <https://www.health-data-hub.fr/notre-organisation> ; voir également l'arrêté du 29 novembre 2019 portant approbation d'un avenant à la convention constitutive du groupement d'intérêt public « Institut national des données de santé » portant création du groupement d'intérêt public « Plateforme des données de santé »

⁵³ Un arrêté publié le 29 juin 2021 officialise la création de ce comité ainsi que sa composition : directions générales du Ministère de la Santé, Secrétariat général des ministères chargés des affaires sociales (SGMAS), CNAMTS, HDH (qui en assure le secrétariat), différentes agences publiques ainsi que quatre représentants des établissements hospitaliers et deux représentants d'associations d'usagers du système de santé.

d'expression des intérêts des parties-prenantes, certains acteurs parlent ainsi de lieu de 'diffusion d'information' et d'échange de 'retours d'expérience' (Entretien, industriel du numérique en santé, 22/07/2022). Si pour certains cette inclusion dans la gouvernance constitue une amélioration par rapport à la période précédente (cas des laboratoires pharmaceutiques), pour d'autres (et notamment les anciens membres de l'IDS), il s'agit plutôt d'une régression :

'En tout cas moi comme ça maintenant avec la perception de l'extérieur c'est que, le dialogue qu'on avait réussi à instaurer entre, à travers l'IDS avec la DREES, la CNAMTS, notamment la Mutualité [Fédération Nationale de la Mutualité Française] qui a été la seule à construire une base de données pérenne pour faire **ce dialogue-là en fait il s'est perdu** au moment du Health Data Hub en fait. On n'était plus sur les mêmes, **on n'était plus sur les mêmes... échanges opérationnels**, que ce qu'on avait pu avoir à l'époque.' (Entretien, représentant, secteur des assurances complémentaires, 16/06/2022)

' – C'est le CESREES qui décide, qui après voit avec la CNIL. Nous on est là on gère, on gère quand on est, on a des réunions comme [?] l'assemblée générale, on voit donc toutes les perspectives pour les années en cours, les travaux en cours, et pareil on vote pour un budget, pour, les grands projets mais bon après **on n'a plus le rôle enfin d'analyste et de censeur** [...]

– D'accord. Donc vous êtes moins impliqué finalement que... sous l'IDS et l'INDS.

– Oui tout à fait **on y est moins**. Et puis cette structure est devenue tellement, tellement énorme que des fois comme je dis toujours pour nous de nous y retrouver de l'intérieur, je suis représentant de X là-dedans, c'est vrai qu'il faut arriver à s'y retrouver, c'est pas évident.' (Entretien, représentant professionnels de santé, 23/05/2022)

Avec le HDH, l'IDS a définitivement perdu le statut ambigu qui lui permettait de jouer un rôle au sein de la gouvernance stratégique des données de santé (voir Working paper n°1). Son rôle d'opérateur technique est désormais on ne peut plus clair :

' [...] quand vous êtes un opérateur, vous n'êtes pas un Ministère, et en gros le Ministère c'est vraiment celui qui porte la stratégie et quand il dit des trucs aux acteurs, ben en gros c'est c'est vraiment, c'est, c'est la règle quoi. Et puis l'opérateur, c'est celui qui met en œuvre, c'est hyper intéressant aussi, mais du coup **c'est pas le Hub qui va élaborer la stratégie** en tout cas, il peut y contribuer mais c'est pas nous qui allons imposer quoi que ce soit à l'écosystème.' (Entretien, fonctionnaire, Health Data Hub, 24/08/2022)

L'affaire Microsoft-Azure'

Au-delà de la mise en place de la gouvernance, les transformations techniques du HDH sont encore largement en cours, ce qui peut s'expliquer d'une part par des difficultés techniques, légales et administratives inhérentes à ce projet, mais

également par la survenue de 'l'affaire Microsoft-Azure' qui a rencontré un écho particulier dans les médias.

Comme nous l'avons vu, une des particularités du HDH est d'héberger des copies des bases de données mères, qui étaient jusque-là uniquement entreposées chez les différents gestionnaires (CNAMTS, PMSI, CépIDc-INSERM etc.). Un des enjeux de la création du HDH est donc relatif à la question technique de l'hébergement des bases de données ; il s'agit en effet de savoir quel organisme sera chargé de stocker ces données éminemment sensibles qui concernent l'ensemble de la population française. Afin de sélectionner l'hébergeur de données, l'État commande une étude au cabinet Cap Gemini, dont les résultats auraient désigné l'entreprise Microsoft-Azure comme unique acteur privé ayant les capacités de réaliser un tel hébergement de données (Entretien, agent CNAMTS, 21/04/2022).⁵⁴ Les autres hébergeurs (notamment le français OVH Cloud) n'auraient pas pu démontrer leur capacité à répondre aux besoins techniques de la plateforme HDH dans une temporalité courte. Sur la base de cette étude, et sans qu'un réel appel d'offre public soit lancé, l'entreprise Microsoft-Azure, d'origine américaine, est sélectionnée pour assurer l'hébergement des copies des données du SNDS.⁵⁵

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'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 (CJUE) qui considère qu'ils ne permettent pas d'assurer un niveau de protection suffisant des droits fondamentaux des citoyens. En 2015, la CJUE invalide ainsi l'accord *Safe Harbor*, qui avait été mis en place au début des années 2000 suite à l'adoption de la Directive européenne 95/46/CE sur la protection des données personnelles. À la suite de cette annulation, un nouvel accord nommé '*Privacy Shield*' est adopté en 2016, puis est de nouveau invalidé par la CJUE en juillet 2020 (arrêt 'Schrems II'). La CJUE considère notamment que le droit américain n'encadre pas assez l'accès aux données personnelles par les autorités américaines lorsqu'elles agissent dans le cadre de la préservation de la sécurité nationale.⁵⁶ L'adoption du *Cloud Act* américain en 2018 autorise en effet

⁵⁴ Voir également : Jacquot, Guillaume. 'Cédric O bousculé au Sénat sur le choix de Microsoft pour héberger le Health Data Hub', *Public Sénat*, Disponible sur : <https://www.publicsenat.fr/article/parlementaire/cedric-o-bouscule-au-senat-sur-le-choix-de-microsoft-pour-heberger-le-health>

⁵⁵ Pollet Mathieu (2021). 'Le choix de Microsoft pour le Health Data Hub français continue d'alimenter les critiques', *Euractiv*, publié le 19 février 2021, Disponible sur : <https://www.euractiv.fr/section/economie/news/le-choix-de-microsoft-pour-le-health-data-hub-francais-continue-dalimenter-les-critiques/>

⁵⁶ CNIL (2020). 'Invalidation du Privacy shield : les premières questions-réponses du CEPD', page web du site de la CNIL, publiée le 31 juillet 2020, disponible sur : <https://www.cnil.fr/invalidation-du-privacy-shield-les-premieres-questions-reponses-du-cepd>

les autorités américaines à accéder aux données détenues par des entreprises américaines, même si ces données sont stockées dans d'autres pays.

L'attribution du contrat d'hébergement des données de santé à Microsoft-Azure intervient précisément quelques mois avant l'invalidation du *Privacy Shield*, dans un contexte où les possibilités d'accès aux données personnelles par les autorités américaines font débat. Ce choix (qualifié par certains 'd'erreur' ou de 'catastrophique', Entretien, ex-conseiller de Marisol Touraine, 19/07/2022 ; Entretien, ex-agent CNAMTS, 22/09/2022 ; Entretien, ex-agent CNAMTS, 19/04/2022) génère alors de vives tensions.

[...] et puis ce qui a pénalisé fortement ça vous le savez c'est que le fait que, comment dirais-je le Health Data Hub héberge les données de santé chez Microsoft ça a quand même **fait grincer des dents**. [...] ça a amené quand même beaucoup de **remue-ménage** [?] ça a pas du tout été apprécié.' (Entretien, représentant professionnels de santé, 23/05/2022)

Suite à l'invalidation du *Privacy Shield* en juillet 2020, des associations du secteur du logiciel libre et des syndicats décident d'adresser une requête en référé-liberté au Conseil d'État afin de faire suspendre les activités du HDH.⁵⁷ Si le Conseil d'État ne juge pas nécessaire une telle suspension, il confirme cependant que le risque de transfert de données aux autorités américaines, bien que faible, ne peut pas être totalement exclu.⁵⁸ La CNIL quant à elle, dans un avis du 11 juin 2020, souligne les risques de transfert des données de santé aux autorités américaines et exprime sa préférence pour le choix d'un hébergeur de données 'relevant exclusivement des juridictions de l'Union européenne'⁵⁹.

Suite à ces polémiques, le gouvernement cherche à temporiser, plusieurs déclarations sont alors faites, certaines annonçant le futur transfert des données de santé vers un hébergeur européen,⁶⁰ d'autres se concentrant plus spécifiquement sur la résolution des incertitudes juridiques.⁶¹ En février 2022, c'est au tour de la

⁵⁷ Les organisations signataires du recours sont les suivantes : Conseil national du logiciel libre (CNLL), l'association Ploss Rhône-Alpes, l'association SoLibre, la société Nexedi, l'association Interhop (les hôpitaux français pour l'interopérabilité et le partage libre des algorithmes), le syndicat national des journalistes (SNJ), le syndicat de la médecine générale (SMG), l'union française pour une médecine libre (UFML), l'union générale des ingénieurs, cadres et techniciens CGT (UGICT-CGT), l'union fédérale médecins, ingénieurs, cadres, techniciens CGT santé et action sociale (UFIMCT - CGT santé et action sociale), l'association Constances, l'association les Actupiennes et l'association française des hémophiles (AFH). Certaines de ces organisations se sont regroupées au sein d'un collectif appelé 'SantéNathon', dont l'objectif est précisément d'empêcher le transfert des données de santé des citoyens français vers les États-Unis. Plus d'informations sur : <https://www.santenathon.org/>

⁵⁸ Ordonnance du Conseil d'État n° 444937 du 13 Octobre 2020

⁵⁹ CNIL (2020). 'La Plateforme des données de santé (Health Data Hub)', page web du site de la CNIL publiée le 11 juin 2020 [archive], disponible sur : <https://www.cnil.fr/fr/la-plateforme-des-donnees-de-sante-health-data-hub>

⁶⁰ Le Monde (2020). 'Données de santé : le gouvernement veut rapatrier le Health Data Hub, hébergé chez Microsoft', publié le 09 octobre 2020, Disponible sur : https://www.lemonde.fr/pixels/article/2020/10/09/donnees-de-sante-le-gouvernement-veut-rapatrier-le-health-data-hub-heberge-chez-microsoft_6055394_4408996.html

⁶¹ <https://www.euractiv.fr/section/economie/news/le-choix-de-microsoft-pour-le-health-data-hub-francais-continue-dalimenter-les-critiques/>

CNAMTS d'exprimer officiellement son opposition à l'hébergement des données de santé par Microsoft.⁶² Avec l'approche des élections présidentielles, le dossier est officiellement mis en attente.⁶³ Fin 2022, la directrice du HDH annonce finalement que l'hébergement demeurerait chez Microsoft-Azure au moins jusqu'en 2025, le temps de développer une 'solution souveraine'.⁶⁴

Malgré ces difficultés et la résonance de cette affaire dans la sphère médiatique, le dossier du HDH bénéficie d'un soutien politique important suite au rapport Villani :

'On a quand même été **suivis par l'Élysée** pendant tout le temps, donc avec des réunions, ce n'était pas tous les 15 jours, hein, c'était plutôt de l'ordre trimestriel, mais, donc donc du coup, on avait un suivi Matignon et un suivi, un suivi santé, donc ça n'a pas que des avantages d'être suivi par tous les cabinets, mais ça a quand même celui que **ils s'assurent qu'il n'y avait pas de problèmes** qui soient trop handicapants et que le projet puisse avancer quoi.' (Entretien, fonctionnaire, Health Data Hub, 24/08/2022)

En conclusion, la fin des années 2010 et le début des années 2020 sont essentiellement marquées par des transformations d'ordre 'technique' ou pratique, visant à permettre un usage plus industriel des données de santé. Ces transformations, toujours en cours, font suite à l'apparition d'une nouvelle 'fenêtre d'opportunité' autour de la notion d'intelligence artificielle : en particulier, le rapport sur ce sujet publié par le député Cédric Villani permet de donner une ampleur inédite au projet de la DREES qui visait à faciliter les accès de ses services au SNDS. Cependant, ces transformations techniques ne modifient en rien les règles d'accès, ou la forme de la gouvernance opérationnelle et stratégique tels qu'ils prévalent depuis la réforme de 2016. Plus encore, la réalisation concrète des promesses du Health Data Hub demeure encore en suspens, ralentie (entre autres) par les polémiques autour de l'attribution de l'hébergement des données de santé à l'entreprise américaine Microsoft-Azure.

⁶² InterHop (2022). 'Avis du Conseil de la CNAM sur les travaux menés par la Commission mixte (COR et CSITN) sur les données de santé', reproduction intégrale publiée le 04/02/2022, disponible sur : <https://interhop.org/2022/02/04/avis-cnam-commission-mixte>

⁶³ Piquard Alexandre (2022). 'Health Data Hub : l'hébergement par Microsoft ne sera pas remis en jeu « avant la présidentielle »', *Le Monde* [en ligne], Disponible sur : https://www.lemonde.fr/economie/article/2022/01/20/health-data-hub-l-hebergement-par-microsoft-ne-sera-pas-remis-en-jeu-avant-la-presidentielle_6110275_3234.html

⁶⁴ Vitard, Alice (2022). 'Microsoft restera l'hébergeur du Health Data Hub jusqu'en 2025', *L'Usine digitale* [en ligne], Disponible sur : <https://www.usine-digitale.fr/article/microsoft-restera-l-hebergeur-du-health-data-hub-jusqu-en-2025.N2043032>

Tableau. La régulation de l'accès aux données de santé à la fin des années 2010

Règles d'accès	Gouvernance opérationnelle	Gouvernance stratégique	Enjeux pratiques
Evolution par rapport à la période précédente			
Statut quo	Statut quo	Statut quo	Ouverture (en cours)
-	CESREES + CNIL (plus d'avis de l'INDS sur les projets qui ne relèveraient pas d'une méthodologie scientifique classique)	-	Création du Health Data Hub : copie des bases de données originales, appariement avec un grand nombre de données, budget multiplié pour le traitement et les extractions

Conclusion

Ce second Working paper sur les évolutions de l'accès aux données de santé en France avait pour objectif de revenir sur les transformations intervenues dans ce secteur au cours des années 2010, jusqu'au début des années 2020. Cette période est marquée par un mouvement d'autonomisation et de politisation de ce domaine de l'action publique. En effet, si l'accès aux données de santé était auparavant traité comme un sous-enjeu de problématiques plus globales (Working paper n°1), le début des années 2010 marque la transformation de ce secteur en catégorie autonome de l'action publique : il s'agit désormais pour l'État de produire sa doctrine en matière d'accès aux données de santé, indépendamment des négociations relatives à d'autres enjeux du système de santé. Cette autonomisation se double d'un processus de politisation : d'une part, le politique (c'est-à-dire la Ministre et son cabinet) s'investit directement sur ce sujet ; d'autre part, le processus de formulation cherche à inclure le plus grand nombre d'acteurs possibles (publics et privés), notamment au sein de la Commission 'Open data santé' créée en 2013. Alors que la plupart des acteurs 'privés' (représentants de patients, fédérations d'assureurs complémentaires, industries) souhaitent ouvrir davantage l'accès aux données de santé, trois acteurs de poids se prononcent au contraire pour une meilleure sécurisation des données de santé : la CNAMTS, l'État, et certains représentants de la communauté scientifique. Ainsi, malgré un cadrage axé sur l'ouverture des données de santé (en accord avec les notions de 'Big data', 'd'open data' et de 'transparence' qui prévalent alors au sein de l'administration), la loi de modernisation de notre système de santé de 2016 contribue plutôt à une dynamique de fermeture de l'accès aux données de santé, en particulier pour les acteurs qui ne relèvent pas d'un 'service public' (beaucoup d'organismes publics se voient au contraire octroyer un meilleur accès aux données du Système National des Données de Santé). Cette dynamique se double par

ailleurs d'une technicisation de la procédure d'autorisation d'accès, et d'un recentrage du pilotage stratégique au sein de l'État.

À la suite des élections présidentielles en 2017, on n'observe pas de revirement radical de la politique d'accès aux données de santé. Cet enjeu est désormais associé à la problématique plus générale du développement de l'intelligence artificielle, ce qui permet d'accélérer certaines transformations techniques visant à industrialiser l'usage des données de santé. C'est précisément l'objectif du Health Data Hub, plateforme technique des données de santé qui remplace l'Institut National des Données de Santé en 2019 avec des moyens financiers multipliés par trente. Le Health Data Hub est censé détenir une copie des bases de données du SNDS afin de permettre l'accès aux demandeurs de manière rapide. Aujourd'hui, la mise en œuvre du HDH est toujours en cours ; au-delà des difficultés inhérentes à ce projet, ce sont également des polémiques liées à l'attribution de l'hébergement des données à l'entreprise américaine Microsoft-Azure qui contribuent à en ralentir le déploiement.

Par contraste avec la première période de développement de l'accès aux données de santé en France (1990-2000), la seconde période est donc marquée par une reprise en main du politique et se traduit par une dynamique de fermeture de l'accès aux données de santé pour plusieurs catégories d'acteurs non-étatiques. Dans le cas français, l'autonomisation de ce secteur d'action publique se traduit donc par des conditions d'accès plus strictes aux données de santé, doublées en parallèle d'une volonté d'industrialisation de l'usage de ces données qui se matérialise essentiellement par des transformations de nature technique.

Bibliographie

- Cluzel-Métayer, Lucie (2018). 'La construction d'un service public de la donnée', *Revue française d'administration publique*, 167:3, 491–500.
- Goldberg, Marcel et al. (2016). 'L'ouverture du Système national d'information inter-régimes de l'assurance maladie (SNIIRAM) : des opportunités et des difficultés. L'expérience des cohortes Gazel et Constances', *Revue d'Epidémiologie et de Santé Publique*, 64, 313–320.
- Malye, François, and Jérôme Vincent (2015). 'Le palmarès des hôpitaux', *Les Tribunes de la santé*, 47:2, 65–74.
- Morelle, Aquilino, and Marc-Olivier Padis (2011). 'Mediator : l'histoire d'une seconde défaite de la santé publique. À propos du rapport de l'Igas', *Esprit*, Mai:5, 71–79.
- Skovgaard, Lea L., Sarah Wadmann, and Klaus Hoeyer (2019). 'A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good', *Health Policy*, 123:6, 564–571.

Annexe : Liste anonymisée des entretiens réalisés

	Fonction anonymisée	Type entretien	Date entretien	Durée entretien
1	Entretien, ex-fonctionnaire, PMSI, DSSIS	Entretien en visioconférence	06/04/2022	01:02:51
2	Entretien, fonctionnaire, ex-IDS	Appel téléphonique	14/04/2022	00:27:32
3	Entretien, ex-agent CNAMTS	Entretien présentiel	19/04/2022	02:02:26
4	Entretien, agent CNAMTS	Entretien présentiel	21/04/2022	01:39:41
5	Entretien, représentant professionnels de santé	Entretien en visioconférence	23/05/2022	0:41:30
6	Entretien, ex-fonctionnaire, PMSI, Sesame-vitale	Utilisation secondaire entretien thèse	24/05/2018 30/05/2018	05:09:50
7	Entretien, agent CNIL	Utilisation secondaire entretien thèse	15/06/2018	01:15:26
8	Entretien, ex-conseillère de Marisol Touraine	Utilisation secondaire entretien thèse	24/11/2016	00:51:50
9	Entretien, secteur des assurances complémentaires	Entretien en visioconférence	15/06/2022	01:06:44
10	Entretien, juriste spécialiste des données de santé	Entretien présentiel	15/06/2022	01:35:40
11	Entretien, ex-représentant des professionnels de santé	Entretien présentiel	16/06/2022	00:40:17
12	Entretien, représentant, secteur des assurances complémentaires	Entretien en présentiel	16/06/2022	01:07:15
13	Entretien, fonctionnaire, DREES	Entretien en présentiel	17/06/2022	00:57:44
14	Entretien, représentant, association d'usagers	Entretien en visioconférence	30/06/2022	00:52:00
15	Entretien, ex-conseiller de Marisol Touraine	Entretien en visioconférence	19/07/2022	0:24:35
16	Entretien, fonctionnaire et agent CNAMTS	Entretien hybride	20/07/2022	01:25:56
17	Entretien, représentant, industries pharmaceutiques	Entretien en présentiel	21/07/2022	00:39:17
18	Entretien, industriel du numérique en santé	Entretien en visioconférence	22/07/2022	00:46:38
19	Entretien, industrie pharmaceutique	Entretien en visioconférence	24/08/2022	00:32:39
20	Entretien, fonctionnaire, Health Data Hub	Entretien en visioconférence	24/08/2022	00:56:24
21	Entretien, ex-agent CNAMTS	Entretien en présentiel	22/09/2022	01:18:57

PARI

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

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

