

## SOMMAIRE

### LE PARTAGE DES DONNÉES DE SANTÉ SERT-IL LE PATIENT ?

- **La relation médecin-patient toujours au cœur de la confiance**

**Nadir Ammour**, Global Lead Clinical Innovation & Partnership Sanofi

- **Les industriels du DM et les données de santé**

**Armelle Graciet**, directrice des Affaires industrielles du Snitem

- **Introduire un principe de « garantie humaine »**

**David Gruson**, membre du comité de direction de la chaire santé de Sciences Po Paris et créateur d'Ethik-IA

- **Partage des données de santé : l'expérience du diabète Lab de la FFD**

**Caroline Guillot**, sociologue, responsable du diabète Lab de la Fédération française des diabétiques (FFD)

- **Le chemin n'est pas si simple de l'individu au patient...**

**Alain Baleydier**, animateur du groupe de travail AFCROs « Les patients au cœur de la recherche clinique » chargé des relations avec les associations de patients

- **Des garanties publiques de sécurité et de protection**

**Christian Saout**, président de la Commission d'évaluation économique et de santé publique de la HAS

### EXPLOITATION DES DONNÉES DU SNDS : MÉTHODES INNOVANTES ET POSITIONNEMENT AU NIVEAU INTERNATIONAL

- **Une solide conviction de l'intérêt du SNDS**

**Emmanuelle Menke**, responsable des prix et de l'accès au marché chez Novartis

- **Un apport bien intégré**

**Fleur Levrat**, responsable de l'accès au marché pour la division diabète d'Abbot

- **Les possibilités d'appariements entre les cohortes et le SNDS**

**Mehdi Gabbas**, (DEMEX – Cnamts)

- **Algorithmes optimisés pour le traitement des données du SNDS**

**Antoine Pariente**, (Inserm)

- **Qu'est ce que l'IA ? Contexte et finalités – Exemple d'applications**

**Olivier De Fresnoye**, Epidemium



## Est-ce que j'accepte de partager mes données de santé ?

Tout au long de l'année, notre groupe de travail *Real World Data* (RWD) a travaillé à la construction d'un baromètre sur le ressenti de la population française à propos du partage des données de santé, le baromètre ACCORDS (Acceptabilité et Communication sur ses Données de Santé), dont nous avons présenté les premiers résultats lors de notre 11<sup>ème</sup> colloque sur les Données de Santé en Vie Réelle\*. Ceux-ci nous montrent qu'une grande majorité des Français (61 %) se juge plutôt mal informée sur le partage des données de santé. Leur confiance dans les professionnels de santé se situe à un niveau élevé : plus de la moitié d'entre eux sont prêts à partager l'intégralité de leur dossier médical avec eux et avec leur mutuelle. La défiance est en revanche de mise pour près de 72 % des personnes interrogées lorsqu'il s'agit de partager ces données avec des sociétés proposant des objets connectés en rapport avec la santé. Les Français affichent aussi une attitude paradoxale, voire schizophrène, vis-à-vis des réseaux sociaux. Près de deux tiers d'entre eux considèrent qu'une information sur leur santé partagée sur un réseau social ou sur un forum de discussion en ligne n'est pas une donnée de santé. Parallèlement, ils sont tout aussi nombreux à ne pas leur faire confiance. Ils restent aussi partagés (50/50) sur la possibilité de monnayer l'accès à leurs données de santé, mais y semblent prêts si la rémunération est suffisante. Des résultats qu'il nous faudra prendre en compte pour nous garantir sans cesse que les patients reçoivent une information fiable et de qualité sur l'utilisation de leurs données de santé (objectifs, promoteur, garanties d'anonymat et de sécurité...).

Le groupe de travail RWD

\* Cette étude a été réalisée grâce au soutien du groupe VYV et à la participation de B3-TSI et de BVA Santé.

## Le partage des données de santé sert-il le patient ?

### La relation médecin-patient toujours au cœur de la confiance



**Nadir Ammour,**  
Global Lead Clinical  
Innovation & Partnership  
Sanofi

Le développement clinique est très réglementé. Aussi la collecte des données inhérente au développement du médicament et à la recherche biomédicale est-elle très encadrée, ce qui nous offre la possibilité d'avoir des supports pour expliquer au patient pourquoi et dans quel but ses données sont utilisées et pour obtenir son consentement. Dans le cas des données disponibles dans le dossier médical informatisé ou dans la base du SNDS, la problématique à laquelle on va être de plus en plus confrontée réside dans le fait que ces patients n'ont pas nécessairement consenti à une réutilisation de leurs données pour des programmes de recherche. Dans le cas des études cliniques, le consentement est généralement assez facile à obtenir car l'objectif est très précis et les patients en comprennent les tenants et les aboutissants. On n'observe pratiquement pas de rétractation de droit à utiliser leurs données. Mais ici, la relation de confiance entre le médecin et le patient joue un rôle déterminant. C'est elle qui va réellement conditionner le consentement du patient à participer à un essai. Dans le cadre de nos études, si on propose aux patients d'installer des applications sur leur propre smartphone, ils préfèrent souvent disposer d'un appareil séparé. Ils ont bien compris que tout est sécurisé dans nos téléphones et ne veulent pas que leurs informations puissent tomber dans d'autres mains. Il existe donc clairement une dualité dans leur attitude par rapport à l'accès à leurs données et à leur utilisation.

Nous devons aussi tous progresser sur la traçabilité et le retour fait aux patients sur l'utilisation de leurs données. S'ils consentent à ce que leurs données soient utilisées dans un programme de recherche, ils veulent savoir à quoi elles ont servi et si elles ont permis de faire évoluer le traitement ou la compréhension de la maladie.

### Les industriels du DM et les données de santé



**Armelle Graciet,**  
directrice des Affaires  
industrielles du Snitem

L'industrie du dispositif médical (DM), comme l'industrie pharmaceutique, utilise les données de santé pour la conception, le développement et la validation de ses nouveaux produits. Ces données servent aussi à obtenir une meilleure connaissance des pathologies et à sélectionner les profils de patients les plus susceptibles de bénéficier d'une technologie. En termes d'innovation, les logiciels d'intelligence artificielle qui entrent dans la catégorie des dispositifs médicaux constituent une catégorie de produits pour lesquels une grande quantité de données est nécessaire. Il faudra donc pouvoir croiser les données issues de différentes sources. Pour les industriels du dispositif médical, tout comme pour ceux de l'industrie pharmaceutique, les conditions de récolte et l'utilisation des données de santé sont extrêmement réglementées. Nous parlons ici de points qui sont bien sûr communs à tous les industriels impliqués dans la recherche et le développement. Mais de nombreux dispositifs médicaux transforment aussi le parcours et l'organisation des soins. Il est donc parfois nécessaire pour les industriels du DM de pouvoir accéder à des données de santé afin de comparer un parcours de soins « classique » au parcours de soins tel qu'il deviendra avec l'arrivée de leur nouveau dispositif et d'une nouvelle organisation.

Enfin, les industriels du dispositif médical ont aussi besoin d'utiliser les données de santé pour répondre aux demandes des autorités pour l'évaluation des produits en amont d'une première demande de remboursement et pour les études post inscription. Pour ces dernières études en vie réelle, les autorités demandent des données exhaustives. Or l'expérience montre qu'on arrive à suivre seulement 10 ou 15 % des patients en vie réelle. L'accès aux bases de données du SNDS qui ouvrent un accès à toute la population concernée permet une exhaustivité importante. L'utilisation de ces données permet aussi aux autorités d'avoir une veille sanitaire beaucoup plus précise sur l'utilisation et sur la vigilance des médicaments ou des dispositifs médicaux.

## Introduire un principe de « garantie humaine »



**David Gruson,**  
membre du comité de  
direction de la chaire  
santé de Sciences Po  
Paris et créateur  
d'Ethik-IA

Les résultats du baromètre ACCORDS sont très cohérents avec ceux du baromètre suivi tous les trimestres par la Chaire Santé de Sciences Po Paris, avec l'Institut Odoxa. Il y a trois ans, lorsqu'on demandait aux Français s'ils préféreraient l'efficacité du service médical rendu ou la protection de leur vie privée, les deux tiers d'entre eux mettaient en avant la protection de leur vie privée. Aujourd'hui, la situation s'est inversée et les deux tiers des Français privilégient maintenant l'efficacité du service médical rendu. C'est une vraie bascule que nous avons essayé d'analyser dans le dernier rapport du Comité consultatif national d'éthique (CCNE)<sup>1</sup>. Le message principal du CCNE est assez inattendu et disruptif : en fait le principal problème éthique du pilotage par les données et l'intelligence artificielle est qu'il n'y en a pas assez. Autrement dit, il ne serait pas éthique de continuer à supporter des non-qualités et des sous-efficacités dans le système alors que celles-ci pourraient être évitées avec plus de pilotage par les données et plus de data management dans un contexte où le niveau de protection juridique des données de santé est plus fort en France qu'ailleurs.

### Eviter de sur-réglementer

Le CCNE estime aussi qu'il y a un risque éthique associé à une sur-réglementation. Dans ce contexte déjà juridiquement protégé, l'ajout de crans réglementaires complémentaires va encourager patients et professionnels de santé français à recourir à des solutions de médecine algorithmique qui se développent hors du territoire du RGPD. Or la pratique est déjà présente : la FDA a labellisé le logiciel IDx-DR pour le diagnostic de la rétinopathie diabétique il y a plus d'un an. Si on n'arrive pas à incuber suffisamment d'innovations dans le territoire du RGPD, nous allons vers un mouvement d'exfiltration de nos données de santé sous la pression de l'usage et donc de nos concitoyens. C'est en quelque sorte le message que le CCNE a voulu apporter. Il propose l'insertion d'un principe de garantie humaine du numérique en santé dans la loi de bioéthique, recommandation que la mission d'information de l'assemblée nationale sur la révision de la loi relative à la bioéthique a d'ailleurs suivie. Cette notion de garantie humaine du pilotage par les données de santé paraît abstraite mais elle est très concrète et c'est précisément ce que vous faites dans chacune des CRO et à l'AF-CROs. Le CCNE propose de développer ce principe sous la forme de référentiel adapté avec, par exemple la mise en place d'outils de suivi et la mise en place d'une capacité d'exercice d'un deuxième regard médical humain pour vérifier le comportement et l'efficacité de l'algorithme. Face à la bascule en cours, il n'est pas possible d'envisager les questions de pilotage de nos données de santé dans un vase clos parce que le numérique rend les systèmes beaucoup plus poreux entre eux au niveau mondial.

(1) Le CCNE a publié récemment une série de documents relatifs au numérique et à la santé. Il s'agit :

- de l'avis 129 « Contribution du CCNE à la révision de la loi de bioéthique » émis dans le cadre de la révision de la loi de bioéthique ([https://www.ccne-ethique.fr/sites/default/files/publications/avis\\_129\\_vf.pdf](https://www.ccne-ethique.fr/sites/default/files/publications/avis_129_vf.pdf))

- de l'avis 130 « Données massives (big data) et santé : une nouvelle approche des enjeux éthiques »

<https://www.ccne-ethique.fr/fr/publications/donnees-massives-et-sante-etat-des-lieux-prospective-et-nouvelles-questions-ethiques>

- du rapport « Numérique & santé : quels enjeux éthiques pour quelles régulations ? » réalisé par le groupe de travail commandé par le comité consultatif national d'éthique pour les sciences de la vie et de la santé (CCNE) avec le concours de la commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene (CERNA). 19 novembre 2018

[https://www.allistene.fr/files/2018/11/rapport\\_numerique\\_et\\_sante\\_19112018.pdf](https://www.allistene.fr/files/2018/11/rapport_numerique_et_sante_19112018.pdf)

## Partage des données de santé : l'expérience du diabète Lab de la FFD



**Caroline Guillot,**  
sociologue, responsable  
du diabète Lab de la  
Fédération française des  
diabétiques (FFD)

Le Diabète LAB est un service de la Fédération Française des Diabétiques qui existe depuis 2015. Nous y recueillons l'expérience patient, avec deux objectifs essentiels. Le premier relève du marketing social et vise à rendre compte de ce que représente l'expérience du patient. C'est aussi un moyen de permettre aux patients de se comparer entre eux, ce qui peut susciter un stimulus pour faire évoluer leur comportement. Le second objectif est plus politique et vise à soutenir nos plaidoyers. Nous allons par exemple évaluer le service rendu d'un dispositif médical pour participer aux réflexions sur son remboursement. Nous contribuons aussi à la prise en compte de l'expérience patients lors de l'évaluation des pratiques et des critères qualité dans l'organisation d'un système de soin.

Concernant le rapport à la donnée de la santé, je considère qu'il peut s'envisager selon trois scénarios. Le premier, que l'on pourrait qualifier de scénario RGPD, pourrait s'apparenter au modèle DMP, où les données sont fournies en échange d'un service. Ce dispositif a des intérêts mais sa limite est de transformer les individus en machines à consentir. Le deuxième scénario réside dans la monétisation et s'il a beaucoup de limites, il présente au moins l'intérêt de faire prendre conscience au patient de la valeur de sa donnée de santé. Le troisième scénario envisage quant à lui la donnée comme un bien commun pour la production de connaissance. Ce dernier modèle fait sens pour une association de patients, mais il ne me semble pas qu'il y ait de bon ou de mauvais modèle.

Enfin, il faut envisager une segmentation des patients. Les études que nous avons pu mener au Diabète Lab mettent déjà en évidence deux types de patients parmi les diabétiques. Un premier type de patients va accepter de laisser ses données si on lui prouve que le dispositif en question va permettre au médecin de mieux le soigner. Leur rapport à la donnée n'est pas passif, mais ces patients restent dans une relation où ils s'en remettent au médecin et veulent qu'on leur prouve que l'apport de leurs données va réellement servir au praticien. Le deuxième type de patient se place dans l'empowerment : il accepte de laisser ses données si on lui prouve que le dispositif va permettre d'améliorer la connaissance de sa pathologie et lui donne la possibilité d'être davantage acteur de sa maladie. Appréhender les situations sous ces deux angles peut déjà permettre de répondre à de nombreuses questions, notamment sur la façon de créer et de faire évoluer les dispositifs médicaux ou l'ensemble des outils numériques qui recueillent de la donnée.

## Le chemin n'est pas si simple de l'individu au patient...



**Alain Baleyrier,**  
animateur du groupe  
de travail AFCROs  
« Les patients au cœur  
de la recherche clinique »  
chargé des relations  
avec les associations de  
patients

Le groupe de travail « les patients au cœur de la recherche clinique » a pour but de favoriser leur participation active à l'élaboration des programmes de recherche, et ce, à travers les associations de patients. 36 d'entre elles ont ainsi répondu favorablement à la proposition de l'AFCROs de participer à cette initiative qui regroupe aujourd'hui une dizaine d'associations de toutes tailles (AFM, FFD, France Vasculaire...), des membres de l'AFCROs, quelques industriels ainsi qu'une avocate membre d'un comité de protection des personnes (CPP). Nous sommes en train de finaliser un livre blanc afin de poser les bases éthiques nous permettant de travailler ensemble. Nous avons aussi la volonté de rédiger un questionnaire pour mettre en avant l'implication réelle des associations et décrire toute leur participation à la conception d'un protocole, et non pas à la seule préparation de la notice d'information. Des discussions sont en cours pour que ce questionnaire puisse accompagner les dossiers transmis aux CPP, mais plusieurs instances, dont la Commission Nationale des Recherches Impliquant la Personne Humaine (CHRIPH), expriment des réticences.

Maintenant que les missions de notre groupe de travail ont été présentées, les résultats du baromètre ACCORDS m'amènent une première réflexion. Ces derniers portent sur la population générale. Ce sont donc des citoyens, des individus qui se sont exprimés, or il faut qu'une personne fasse le chemin de devenir un patient et de se dire qu'il a besoin de soins. L'enquête que j'ai menée au sein de notre groupe de travail montre des convergences et des divergences avec le baromètre, qui, selon moi, proviennent de cette sectorisation entre patient et individu. Pour les patients présents dans notre groupe de travail, les données de santé sont bien identifiées. Ce sont évidemment les données qu'ils partagent avec le corps médical, mais aussi celles qu'ils partagent de manière inconsciente avec la sécurité sociale, les mutuelles, le système hospitalier ou encore la pharmacie et celles issues de la mesure de leurs activités physiques. Tout comme les individus interrogés dans le cadre du baromètre, ils ne considèrent pas les informations partagées sur les réseaux sociaux comme des données de santé. En revanche, ces patients qui ont fait la démarche de se fédérer dans une association jugent la notion de consentement très importante. Ils veulent aussi savoir à qui et pour quoi ils transmettent leurs données de santé. La question de la sécurité est aussi considérée comme un point important. Mais il existe un biais, car ces patients connaissent l'existence du règlement général sur la protection des données (RGPD) et sont même capables de citer des hébergeurs agréés de données de santé. Les discussions sur la sécurité doivent donc tenir compte de la diversité des strates de la population. Enfin, l'utilisation commerciale des données de santé est quasiment un tabou, leur objectif étant avant de tout de faire progresser la connaissance sur une maladie.

## Des garanties publiques de sécurité et de protection



**Christian Saout,**  
président de la  
Commission d'évaluation  
économique et de santé  
publique de la HAS

Je veux d'abord évoquer la place du citoyen dans la collecte des données de vie réelle et sa compréhension des enjeux de cette collecte. Il y a d'abord un problème d'accès au numérique en termes de fracture territoriale et sociale et en termes d'usage. La collecte des données de santé implique souvent la participation du patient. Aussi il faut penser à la médiation numérique et à l'accompagnement des personnes pour leur permettre d'utiliser ces outils. Il faut aussi les associer à l'ensemble du processus de construction des outils de collecte, y compris dans la restitution des données, pour qu'ils comprennent l'intérêt de participer à cette récolte. Il faut aussi faire attention à ne pas faire déferler trop d'applications sur les patients. Si, à un moment donné, on ne domestique pas le recours à ces outils, un phénomène de lassitude peut s'installer.

Je veux aussi insister sur les moyens de la collecte des données. L'objectif est d'améliorer la qualité et la sécurité des soins et il est donc important d'associer aussi les professionnels de santé à la conception et à la construction de ces outils de collecte parce qu'il ne s'agit pas d'évaluer un morceau du parcours de soins, mais son ensemble et de repérer les aléas dans ce parcours de soins. Je suis aussi attentif à l'augmentation des moyens de la collecte. De nombreux registres sont aujourd'hui en difficulté alors qu'ils rendent des services importants. On entend beaucoup parler des PREMS (Patient Reported Experience Measures) et des PROMS (Patient-Reported Outcome Measures) mais leurs méthodologies d'élaboration sont assez peu disponibles dans l'espace français. Il faut aussi augmenter les compétences pour l'exploitation des données de vie réelle.

Il est aussi nécessaire de clarifier ce que l'on évalue. S'agit-il d'évaluer des expériences, des résultats, des niveaux de satisfactions, d'évaluer au niveau local, au niveau national... Évalue-t-on un point aigu du parcours ou bien le suivi d'une pathologie ? Tous ces éléments demandent à être clarifiés car il faudra probablement des outils distincts pour des buts différenciés.

Enfin, je voudrais réinsister sur les garanties publiques permettant la protection des données de santé. Nombre de nos concitoyens peuvent penser au RGPD, mais l'ensemble des garanties élaborées par le droit et des systèmes de protection des données doit être rendu lisible et visible aux yeux de l'opinion pour être connu et respecté de tous. Il faut aussi être sensible à la réévaluation régulière des risques de réidentification consécutifs à des croisements de fichiers. On doit aussi penser aux risques de sélection des populations qui peuvent entacher la solidarité universelle ou les solidarités catégorielles. Il sera bien évidemment tentant pour les complémentaires de recourir à ces données pour stratifier des populations et des risques et donc de faire varier les coûts de leur protection. Une unification de ce droit souple est probablement à envisager, peut-être sous l'égide de l'UE. Enfin, il faut veiller à la question de l'explicabilité des algorithmes utilisés dans ces outils de collecte des données de santé. Ces points figurent notamment dans le rapport d'analyse prospective de la HAS sur le numérique en santé<sup>1</sup>.

(1) « Numérique : quelle (r)évolution ? ».

[https://www.has-sante.fr/portail/upload/docs/application/pdf/2019-06/rapport\\_analyse\\_prospective\\_2019.pdf](https://www.has-sante.fr/portail/upload/docs/application/pdf/2019-06/rapport_analyse_prospective_2019.pdf)

## Exploitation des données du SNDS : Méthodes innovantes et positionnement au niveau international

Fleur Levrat (Abbott) et Emmanuelle Menke (Novartis) partagent avec nous leur retour d'expérience sur l'impact de l'ouverture des données du SNDS et sur l'intérêt apporté par leurs structures corporate à l'exploitation de ces données pour leurs différentes études.

### Une solide conviction de l'intérêt du SNDS



**Emmanuelle Menke,**  
responsable des prix et  
de l'accès au marché  
chez Novartis

Chez Novartis Pharma, les trois grandes fonctions qui travaillent ensemble sur les données de santé en vie réelle sont le médical, l'accès au marché et les opérations cliniques. La création de l'INDS et la mise en place du SNDS, avec la possibilité d'accéder à toutes ses bases, sont une grande chance pour notre pays et une vraie opportunité. Comment alors nous mettre en ordre de marche pour utiliser au mieux ces données capables de nous faire gagner du temps et d'améliorer la qualité de nos dossiers ? Nous avons informé très vite nos confrères du « global » basés en Suisse qu'il y avait peut-être là des pistes pour des études épidémiologiques ou pour alimenter les « global value dossiers » qu'on peut assimiler à des notes d'intérêt thérapeutique à l'échelle mondiale. La filiale France a travaillé avec le juridique sur le sujet du RGPD et des formations ont été réalisées en interne pour expliquer aussi bien les possibilités d'utilisations des données du SNDS que les finalités interdites à l'ensemble des collaborateurs. A chaque fois que nous rencontrons nos confrères du médical à propos des études post-inscription demandées par la HAS, nous demandons s'il était possible d'utiliser le SNDS pour réaliser des études de coût de santé ou de fardeau de la maladie. En parallèle, nous avons réalisé différentes interventions dans des colloques et cénacles pour expliquer l'intérêt de ces données auprès de nos confrères du global. Ceux-ci ont été particulièrement intéressés par la possibilité de disposer d'une base unique et commune portant sur quelque 66 millions d'assurés sociaux suivis depuis plus d'une dizaine d'années. Ce continuum est une grande force du SNDS. Un projet est maintenant en cours à la demande des équipes du « global » sur une pathologie rhinopharyngée où le « global value dossier » sera basé sur une étude réalisée à partir du SNDS. Nous sommes aussi en concurrence avec d'autres bases européennes dont la Clinical Practice Research Datalink (CRPD) britannique jugée plus facile d'accès mais nous défendons le SNDS et son utilisation devrait aller crescendo pour nos dossiers de transparence mais aussi pour des études post-inscription ou éventuellement des accords avec le CEPS.

## Un apport bien intégré



**Fleur Levrat,**  
responsable de l'accès  
au marché pour la  
division diabète d'Abbott

Chez Abbott, quel est l'intérêt du global pour le SNDS ? J'ai deux biais dans cette réponse. Premier biais : je suis française et je connais très bien le SNDS dont j'attendais l'ouverture avec impatience car je suis convaincue de l'intérêt et de la qualité des données qu'elle permet de recueillir. Deuxième biais : Abbott fonctionne par silo. Au sein d'Abbott, je représente la division diabète qui pendant longtemps a eu pour produits des lecteurs de glycémie et des bandelettes de glycémie capillaires. Il n'y a donc pas de culture d'essais cliniques et encore moins d'études en vie réelle. Mon expérience sera donc complètement différente de celle d'un autre grand groupe plus habitué au médicament, aux essais cliniques et aux études en vie réelle.

A ma arrivée au poste de responsable de l'accès au marché il y a un peu plus d'un an, ma première mission a été de convaincre ma hiérarchie de l'importance de la mise en place d'études en vie réelle, notamment en France où nous disposons de cette merveilleuse base de données du SNDS. Des articles tels que celui que vient de faire l'AFCROs sur le SNDS sont très utiles et il est important d'avoir aussi des publications en anglais pour nous aider à démontrer l'intérêt de ces bases de données auprès du corporate. Je m'étais appuyée sur un article en anglais publié par l'équipe de Bordeaux. L'approbation a été très rapide et unanime, mais il y a eu des contraintes de mise en place notamment pour décider en interne qui gèrera ce type de projet, qui contractualisera, qui sera le responsable auprès de la CNIL pour les projets France. Ces mêmes questions se posent aussi dans d'autres pays et il était très important pour nous de mettre en œuvre dès le départ ce processus de validation des projets en intégrant toutes les équipes et en incluant dès l'amont les data privacy officers (délégués à la protection des données) pour anticiper d'éventuelles demandes d'étude d'analyse d'impact de la protection des données. Les contraintes ont aussi été administratives. Toutes les démarches et textes de la CNIL ainsi que les conventions avec la Cnamts sont rédigées en français. Or toute ma hiérarchie s'exprime en anglais et je devais donc tout traduire avant qu'ils signent. Si on veut une plus grande implication du global, il faudrait donc avoir des modèles de convention en anglais pour faciliter les démarches administratives. Au final, l'attente du global par rapport au SNDS est très importante. Abbott Diabetes Care a bien compris l'intérêt de cette base de données, de sa puissance, de sa qualité et de sa représentativité quasi parfaite. Nous espérons notamment démontrer avec l'analyse des données du SNDS que les coûts d'hospitalisation épargnés justifient le coût d'acquisition plus élevé de notre dispositif FreeStyle Libre. Par ailleurs, le FreeStyle Life change les pratiques médicales en diabétologie. Au lieu de se focaliser exclusivement sur l'hémoglobine glyquée, ce dispositif permet de connaître le temps passé en hypoglycémie ou en hyperglycémie et de suivre entre autres la variabilité glycémique. Ces données étant enregistrées dans le cloud, notre rêve serait de pouvoir les apparier avec des bases de données de santé et pourquoi pas y intégrer aussi les avis de patients. La démarche est initiée aux Etats-Unis avec le lien de notre dispositif avec des bases de données de remboursement mais il reste des barrières de confidentialité des données à surmonter.

## Les possibilités d'appariements entre les cohortes et le SNDS



**Mehdi Gabbas,**  
(DEMEX – Cnamts)

**A**u sein du pôle DEMEX de la Cnamts, une équipe de cinq personnes répond à l'ensemble des demandes reçues via l'INDS. Ces demandes se répartissent entre quatre types principaux. Le premier correspond à des cohortes sur critères qui représentent plus de 90 % des demandes. Un deuxième type de demande est ensuite arrivé régulièrement et progressivement et correspond à des appariements probabilistes. Dans ce cas, on cherche à rapprocher et à faire matcher deux sources de données présentant des données communes mais dont l'une d'entre elles n'inclut ni le NIR, ni d'informations nominatives mais contient des données médico-administratives telles qu'hospitalisations, consommations médicamenteuses, date de naissance... A titre d'exemple, la Société Française de Cardiologie nous a sollicités il y a un an pour appairer les données du SNDS avec les données de leur registre recensant l'ensemble des patients bénéficiaires d'un TAVI (transcatheter aortic valve implantation - implantation transcathéter d'une valve aortique) afin d'améliorer la connaissance de cette technologie au point de vue médical, économique et organisationnel. Les fichiers ont pu être formatés et leur jonction réalisée à partir d'un algorithme défini par les responsables de la cohorte, ce qui a suscité de nombreux échanges avec eux sur les critères principaux et les types de données à suivre. Nous avons ainsi pu récupérer 84 % des patients.

Le troisième type de demande concerne des appariements directs entre le SNDS et des cohortes ou des registres disposant du NIR. Les procédures FOIN (Fonction d'Occultation des Identifiants Nominatifs) 1 et 2 permettent ensuite de récupérer les identifiants Sniiram à partir du NIR, la jonction est alors directe et les données extraites. En termes de démarche, le porteur de projets ou son tiers de confiance, doit prendre contact avec le responsable de la cellule DEMEX qui va lui communiquer la personne habilitée à transmettre le fichier des NIR. Un ticket est ensuite ouvert en interne à destination du CPAM Versailles/CRIP afin d'ouvrir l'accès à la plateforme sécurisée de transfert de données SAFE à la personne habilitée. Le CRIP envoie un mail à la personne habilitée à transmettre les NIR et lui communique le mot de passe par téléphone. Le porteur de projets peut ensuite se connecter à SAFE et transmettre/déposer le fichier des NIR. Une fois le fichier envoyé via SAFE, le traitement se fait en automatique (Cryptage FOIN1 à Toulouse, Cryptage FOIN2 à Evreux et mise à disposition du fichier Foin2 au niveau du portail SNDS). Après validation du fichier FOIN2 par la cellule DEMEX, les données sont extraites et mises à disposition au porteur de projets. Le quatrième type de demande correspond à des tableaux de bord agrégés n'incluant pas de données sensibles à l'image, par exemple du nombre de diabétiques par département.

## Algorithmes optimisés pour le traitement des données du SNDS



**Antoine Pariente,**  
(Inserm U1219 PH  
- Equipe Pharmaco-  
épidémiologie –  
Plateforme nationale  
de Pharmaco-  
épidémiologie DRUGS-  
SAFE – Université de  
Bordeaux)

Qu'est-ce qu'un algorithme ? Pour Wikipedia, c'est « une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre une classe de problèmes ». Une fois posée cette définition, que peut-on opposer entre algorithme et modèle explicatif sachant que les deux peuvent parfois utiliser les mêmes techniques ? Dans les modèles explicatifs, le modèle connaît la réponse et lui cherche une explication, alors qu'on attend de l'algorithme qu'il donne une réponse, soit en identifiant un état présent (algorithme dit déterministe) soit en prédisant des états à venir (algorithme dit prédictif).

Maintenant concernant le SNDS, nous sommes face à une mine d'or dont le gisement est large mais dont la richesse est néanmoins limitée par une concentration en minerai au moins inconstante. Du point de vue de la recherche et de l'évaluation, il existe plusieurs limites : variables de suivi des patients pour lesquelles on perd l'information, informations cliniques restreintes, en particulier concernant la sévérité des maladies, absence de données sur l'indication des traitements ou sur le mode de vie... Mais heureusement des moyens vont être développés pour pallier ces limites. Le Health Data Hub mais aussi le couplage avec les cohortes existantes ou avec les entrepôts de données de santé hospitaliers ouvrent des perspectives intéressantes. Parallèlement, nous essayons tous que ces données « parlent » plus qu'elles ne le font spontanément en utilisant des algorithmes soit pour identifier des choses soit pour enrichir les données présentes.

Sachant qu'un algorithme est destiné à résoudre des classes de problèmes, on peut identifier deux besoins principaux parmi les réponses à trouver à partir des données du SNDS. Au niveau des algorithmes déterministes, il s'agit de savoir qui est malade et qui ne l'est pas ou éventuellement qui présente un état donné et qui ne le présente pas. Les algorithmes prédictifs visent, quant à eux, à utiliser les données existantes pour faire des prédictions à l'échelle individuelle ou à l'échelle populationnelle. Il est à noter qu'il existe un besoin important d'algorithmes d'identification des pathologies ainsi qu'en témoigne un des appels d'offres lancé par la Cnamts.

Parmi les exemples de développement d'un algorithme pour l'identification d'une maladie, on peut citer le cas d'algorithmes pour l'identification de patients diabétiques construits sur la base de la cohorte Constances. Dans cette cohorte, la connaissance de la maladie était établie et il était donc possible de regarder quelles étaient les meilleures performances pour identifier la présence d'un diabète. Cette approche assez classique est parfaite pour la validation de l'algorithme, mais elle présente une limite, car elle repose sur une cohorte réalisée à un moment donné sur les caractéristiques de population. Mon deuxième exemple porte sur l'étude des déterminants liés aux soins pour l'identification des personnes âgées dépendantes pour les activités de la vie quotidienne. De tels travaux ne peuvent pas être réalisés avec les données du SNDS. Néanmoins il existe des données de cohorte telles que la cohorte AMI réalisée à partir des assurés de la MSA, ce qui permet de disposer aussi en parallèle des données de consommation de soins. Des algorithmes ont pu être mis au point grâce à l'application des techniques de lasso répétées sur un grand nombre d'échantillons. Il s'agissait de construire différents algorithmes en tenant en compte toutes les variables présentes soit dans 70 %, soit dans 80 % soit dans 90 % des échantillons puis de déterminer des scores pronostics capables d'identifier les patients dépendants. Le modèle présentant systématiquement les meilleures performances était celui qui regroupait les variables retrouvées dans 80 % des échantillons et était associé à une sensibilité de 75 % et une spécificité de 96 %. Appliqué à l'EGB pour la population des patients âgés de plus de 65 ans, cet algorithme donne une prévalence de dépendance et un âge du 1<sup>er</sup> trimestre d'entrée en dépendance cohérents avec les résultats de la cohorte AMI. Une fois cette étape réalisée, l'objectif est d'identifier les facteurs associés à l'entrée en dépendance. Des résultats communiqués en juin 2019 montrent ainsi qu'une consommation de benzodiazépines en cours est associée à un plus haut risque de dépendance. Si on réussissait à faire baisser toute la consommation de benzodiazépines chronique après 65 ans à partir de 2020, il serait possible d'économiser environ 100 000 cas de démences. Ces exemples montrent l'intérêt des algorithmes pour compenser les faiblesses ou les limites du SNDS, mais leur nécessité devrait se réduire dans le futur avec les possibilités offertes par le Health Data Hub et les couplages. Attention toutefois à ne pas trop faire confiance aux algorithmes et aux modèles prédictifs qui, s'ils sont des outils puissants, n'en sont pas pour autant dénués de risques. Leurs prédictions sur les évolutions des marchés financiers n'ont pas été sans faille et ont généré de la casse...

## Qu'est ce que l'IA ? Contexte et finalités – Exemple d'applications



**Olivier De Fresnoye,**  
Epidemium

Le projet Human Genome Project (HGP), qui s'était donné pour ambition de séquencer l'ensemble du génome humain, représente à mes yeux un des projets les plus emblématiques de la collaboration internationale en sciences. La raison en est simple : les données générées par d'autres projets internationaux de collaborations scientifiques, tels que la station spatiale internationale ou le CERN, ont une durée de vie limitée qu'on peut estimer à 30 ans alors que les données du HGP pourront encore être exploitées dans 100 ans ou dans 200 ans pour faire de l'archéogénomique ou pour mieux comprendre les pathologies.

Lorsqu'on élargit à l'humain, à sa santé, à son environnement, le volume de données (données cliniques, SNDS, données de remboursement, données structurées, données non structurées, données de compte-rendus médicaux, données biologiques, données d'imagerie...) explose. S'y ajoutent maintenant les données générées par des capteurs, par nos téléphones portables, par les montres connectées... Quand on se penche sur un patient et sur sa pathologie, on s'aperçoit que ses données cliniques et ses données biologiques constituent à peu près un tiers de ce qui pourrait nous intéresser puisqu'on peut aller jusqu'aux données comportementales, aux données sociales et à toutes les données sur son environnement qui nous permettent de mieux le comprendre, de mieux comprendre sa maladie et son évolution et de trouver in fine les bonnes stratégies thérapeutiques. Nous sommes donc débordés de données et ce déferlement a donné lieu aux innovations technologiques qu'on met sous le chapeau de big data et qui nous permettent de traiter et de comprendre tous ces renseignements. Mais que faire de toutes ces données ? On peut faire des analyses mais on peut aussi aller plus loin et utiliser les grandes architectures d'intelligence artificielle pour mêler machine et humain et nous orienter vers une médecine prédictive et préventive.

### De la « force brute » à la blockchain

L'intelligence artificielle remonte à la fin du 19<sup>ème</sup> où on a commencé à vouloir que les machines puissent suppléer, supplanter l'homme. Puis les capacités de calcul arrivées dans les années 50 ont progressé jusqu'aux premiers cas de machine learning dans les années 80, avec des machines dotées de capacités cognitives et maintenant au deep learning avec les réseaux de neurones. On est passé de la machine capable de battre Gary Kasparov, c'est-à-dire de la force brute, de la machine capable d'analyser chaque coup très vite pour essayer de trouver la manière de gagner, à la machine capable d'apprendre et de concevoir de nouvelles stratégies, désarçonnant ainsi le champion du monde de go et l'ensemble de la communauté des joueurs. On peut aller encore un cran au-dessus avec des machines qui apprennent à jouer à des jeux, battent des humains et peuvent aussi apprendre à d'autres machines qui, elles-mêmes, commencent aussi à battre des êtres humains. La capacité d'apprentissage de machines nourries par des données se concrétise aussi en santé, en particulier dans les domaines de l'imagerie où une machine entraînée par la lecture d'accumulations d'images est capable d'identifier la présence d'un mélanome avec une exactitude équivalente voire supérieure à celle de certains dermatologues. Les machines sont donc capables aujourd'hui d'apporter des solutions complémentaires, d'accompagner le professionnel de santé, d'améliorer la disponibilité des solutions de santé dans les zones les plus reculées et sans doute aussi de générer des économies pour le système de santé. ....

.... Aujourd'hui, cette intelligence artificielle, dont les promesses sont gigantesques et les risques et les freins encore bien présents, est permise pour trois raisons :

- les volumes de données qui permettent aux machines d'apprendre et de comprendre ce qu'il y a à l'intérieur de ces données ;
- la capacité de calcul qui se multiplie de façon régulière ;
- l'existence des algorithmes, des langages de programmation.

Dans le domaine de la santé, cet environnement nous permet de répondre à de nombreuses questions et de réaliser une grande série d'actions : détection d'un cancer, classification, segmentation, prédiction, recommandation ou même analyse naturelle du langage pour essayer d'extraire des historiques de patients à partir de textes libres écrits dans des comptes rendus médicaux et des informations pour accompagner les professionnels de santé. Mais attention cette belle promesse de l'IA dans le domaine de la santé est encore semée d'embûches. Il reste difficile d'apparier des données avec des formats différents. Les enjeux éthiques sont forts et nombreux : jusqu'où aller dans la délégation de la décision côté médecin et jusqu'où aller dans la délégation du consentement côté patient ? Des éléments de solution arrivent : il faut essayer de toujours penser à garantir la qualité des données, car si la machine apprend sur des données de mauvaise qualité, elle va mal apprendre. Les architectures de type blockchain s'avèrent intéressantes à ce niveau, car ces réseaux distribués autorisent une traçabilité complète depuis la collecte de la donnée jusqu'à sa réutilisation. Elles permettent un audit quasiment complet et infalsifiable de la provenance de la donnée, assurant ainsi l'intégrité de la donnée dans toutes ses étapes de transformation et dans son passage depuis le patient, le médecin, le chercheur jusqu'au développement de médicaments.

### Le programme Epidemium

Si les données cliniques ou les données fermées sont extrêmement intéressantes, il ne faut pas non plus oublier l'immense quantité de données ouvertes disponible, même si la qualité n'est pas toujours au rendez-vous. En 2015, le programme Epidemium a été lancé dans cette optique d'utilisation des données ouvertes avec des partenaires académiques, hospitaliers et industriels dont le groupe Roche qui accompagne l'initiative depuis ses débuts. Nous regroupons des communautés informelles (patients, chercheurs, professionnels de santé...) pour essayer d'explorer de nouveaux territoires grâce aux données en open data, et pour mettre en œuvre les architectures et les outils algorithmiques d'IA. Notre objectif est d'accélérer et d'améliorer la recherche grâce à la force communautaire. De nombreux projets ont émergé de cette approche nouvelle et nous représentons une communauté de plus de 1500 personnes qui collaborent et interagissent, avec, à la clé, déjà un certain nombre de publications scientifiques démontrant que l'utilisation des données ouvertes et des architectures d'intelligence artificielle couplées à ces communautés hétérogènes peuvent être extrêmement disruptives et innovantes par rapport à des approches plus classiques.