

# Données numériques de santé, partie une

## Introduction

### Flora Bastiani

Bonjour et bienvenue dans le podcast *Penser la santé*. Cet épisode est consacré aux questions posées par les données de santé. Nos données de santé sont au cœur d'un immense mouvement de transformation. Elles nourrissent la recherche médicale, elles alimentent les algorithmes, elles modélisent nos trajectoires de soins. Les données de santé sont désormais de plus en plus enregistrées, transmises, analysées et traitées au format numérique. Elles sont constituées de tous les éléments descriptifs en rapport avec notre santé, des comptes-rendus d'analyses biologiques ou d'imagerie aux dossiers médicaux complets. Les études telles que e3n ou Constance, qui regroupent des cohortes de centaines de milliers de volontaires en France, permettent des études longitudinales pour mieux évaluer les facteurs de risque et les déterminants de santé.

La centralisation numérique des données individuelles proposées par l'outil du gouvernement français Mon espace santé depuis deux-mille-vingt-deux vise la simplification, la coordination des soins et l'*empowerment* du patient, mais il constitue aussi une formidable ressource pour la recherche permettant des collectes de données de grande ampleur. Et pourtant, lorsqu'on cherche à se renseigner sur mon espace santé en ligne, la première question qui apparaît dans Google est : « Pourquoi faut-il refuser Mon espace santé ? »

Alors, y a-t-il un risque à accepter de partager ses données de santé ? Les pratiques de récolte de données de santé ont-elles toujours des objectifs louables ? Jusqu'où ces données peuvent-elles permettre d'améliorer les soins ? Ces questions se situent entre l'importance majeure de la protection des données personnelles et le très fort intérêt pour la communauté scientifique et l'innovation dont nous serons tous bénéficiaires demain.

Pour y réfléchir, j'ai le plaisir d'accueillir Noémie Dubruel qui est juriste en droit de la santé et recherche médicale, doctorante à l'Institut Maurice Hauriou, membre de l'équipe bioéthique Cerpops Inserm, et Fabrice Ferré, médecin anesthésiste réanimateur, praticien hospitalier au sein du CHU (Centre Hospitalier Universitaire) de Toulouse. Bonjour à tous les deux.

## **Risques du recueil de données de santé**

### **Flora Bastiani**

Noémie, en tant que juriste, je vous adresse ma première question. Vous vous intéressez particulièrement à la protection des personnes concernant le recueil et la diffusion des données. On parle souvent des RGPD (Règlement Général sur la Protection des Données). Est-ce que la collecte des données de santé risque de remettre en question les principes de cette protection ? Autrement dit, est-ce que le recueil de données de santé est quelque chose de suffisamment réglementé ? Ou bien toute personne qui, par exemple, utiliserait Mon espace santé prend le risque de voir ses données personnelles utilisées sans son consentement ?

### **Noémie Dubruel**

La collecte des données de santé est quelque chose qui existe depuis très longtemps, même depuis toujours quand on fait de la relation de soin ou de la recherche. Donc ce n'est pas quelque chose qui aujourd'hui vient remettre en cause des principes qui pourraient être posés auparavant. C'est plutôt l'inverse. On avait de la collecte et du traitement de données de santé, et ensuite on a réfléchi à poser des cadres juridiques et éthiques pour mieux préserver la protection des personnes. C'est une protection qui a commencé à être pensée au fur et à mesure de la considération qu'on avait pour les données. Donc, au fur et à mesure, on s'est rendu compte que la donnée de santé était quelque chose d'important, qu'il fallait la protéger et que cela pouvait engendrer des risques pour les personnes de les traiter. Donc on s'est retrouvé à avoir un cadre qui s'est posé.

Vous avez parlé du RGPD, c'est le règlement européen qui est venu un peu bouleverser tout l'écosystème juridique en Europe. Forcément, en France, ce règlement est venu s'appliquer automatiquement. Mais en France, on était quand même pionnier en la matière. On avait une loi, la loi Informatique et libertés depuis mille-neuf-cent-soixante-dix-huit, qui avait déjà pensé ces enjeux et qui était déjà fondatrice d'un certain cadre important qui reste applicable et prioritaire. On avait cette loi qui a été adaptée aux exigences du RGPD. Dans le RGPD, la nouveauté est une définition qui est arrivée de la donnée de santé. Avant, on n'avait pas forcément de définition précise, on parlait de données de manière générale, on parlait d'informations de manière beaucoup plus générale.

### **Flora Bastiani**

D'accord.

## **Noémie Dubruel**

Le RGPD a apporté une définition de la donnée de santé en la caractérisant comme une donnée qui est particulièrement sensible parce que liée à quelque chose d'encore plus particulier pour l'individu, un domaine considéré comme assez sensible. Un principe d'interdiction de collecte et de traitement des données de santé découle de ce règlement.

Évidemment, si on s'arrête là, on se rend compte qu'on ne peut plus rien faire. Cela ne marchait pas du tout. L'idée du RGPD a été de poser ce principe d'interdiction par la protection, et ensuite de venir dire qu'il y a évidemment des exceptions : notamment dans la prise en charge des personnes la recherche, donc l'intérêt légitime, le consentement, parce que si toute personne consent directement au traitement des données de santé, on peut les traiter. Sous un cadre bien plus précis de protection, en affirmant des droits pour les personnes, notamment le droit à l'information. J'ai parlé du consentement, le droit à l'opposition aussi à ce traitement, et en définissant des finalités précises du traitement des données de santé, on a finalement tout ce cadre qui vient se poser.

Aujourd'hui, le traitement des données de santé est quand même bien encadré. L'accélération de la collecte des données avec le déploiement du numérique et la création de ces espaces comme Mon espace santé ne viennent pas remettre en cause ces cadres puisqu'ils sont construits en écho avec cela, donc ce n'est pas du tout contradictoire. Pour préciser, Mon espace santé va s'intégrer dans une autre motivation de l'Union européenne qui vient d'adopter un règlement pour créer l'espace européen des données de santé, un espace très large qui permettrait d'harmoniser le traitement des données de santé de tous les citoyens européens. Mon espace santé, qui est national, va s'intégrer dans cet écosystème européen, donc une gouvernance européenne qui devrait être plus protectrice encore pour les personnes.

## **Flora Bastiani**

« Plus protectrices » : ce que je comprends, c'est que les données que je peux partager avec mon médecin, par exemple sur cet espace, ou que mon médecin dépose sur cet espace, ne peuvent pas être traitées ou collectées sans mon consentement.

## **Noémie Dubruel**

Pour ces espaces, on n'est pas sur du consentement parce que c'est la finalité de l'intérêt légitime de traiter ces données qui priment. On est sur un principe

Penser la santé : Données numériques de santé, partie une

d'information et de non-opposition. D'où la fameuse information que vous avez dit en introduction sur la présence, quand on va sur Mon espace santé, de cette fameuse phrase : « Pourquoi dois-je refuser ? » On est vraiment obligé d'aller vers les personnes pour leur transmettre un maximum d'informations. Et c'est à la personne d'être proactive et de s'opposer au traitement de ces données.

**Flora Bastiani**

Donc à aller sur son espace santé et de dire : « Je refuse ».

**Noémie Dubruel**

Voilà, c'est cela.

**Flora Bastiani**

Si je comprends bien, dans votre réponse, on a d'abord commencé par collecter les données et ensuite on s'est dit qu'il fallait un cadre réglementaire. La pratique est beaucoup plus ancienne que le règlement lui-même, c'est cela ?

**Noémie Dubruel**

Oui.

## **Usage des données de santé**

**Flora Bastiani**

Fabrice, en tant que médecin, pourriez-vous nous expliquer comment ces données de santé peuvent être utilisées ? Là, on parle du versant juridique, du partage, de la récolte des données. Qu'est-ce qu'on en fait ?

**Fabrice Ferré**

Merci beaucoup Noémie, parce qu'à chaque fois que tu intervies, cela m'éclaire. Effectivement, cela fait très longtemps qu'on collecte des données. Il y a à peu près un millier d'années, Avicenna était un médecin et scientifique perse qui suggérerait déjà de tester des nouveaux traitements sur un groupe de patients tout en gardant un groupe contrôle. C'était les premiers tests cliniques, et cela nécessitait intrinsèquement la collection des données de santé. C'est vrai que les choses évoluent terriblement. Notamment, maintenant nous disposons d'une masse incommensurable de données. Finalement, ce que ces données de santé

massives permettent, c'est d'aller au-delà de l'expérience de la mémoire d'un médecin en particulier et de comparer des nouvelles observations, des nouvelles situations à l'immensité de la mémoire collective, et enfin d'essayer grâce à cela, d'essayer de deviner l'évolution probable d'un patient. C'est simplement de la médecine guidée par les données, on parle de *data driven medicine*.

On est dans une situation très particulière, qui est un alignement des planètes, avec une médecine qu'on veut de plus en plus focalisée sur ce qu'on appelle les cinq P, c'est-à-dire : préventive, prédictive, personnalisée, fondée sur des preuves et participatives. Cet alignement des planètes rend une médecine des cinq P, l'accessibilité à des données massives parce qu'on est capable d'avoir accès à des données de biologie, aux données des dossiers médicaux, des dossiers patients informatisés, mais également le développement récent des capacités de modélisation et de simulation avec l'utilisation de l'apprentissage machine et des algorithmes d'intelligence artificielle. C'est probablement pour cela qu'on est là aujourd'hui, c'est qu'on est dans cet alignement des planètes qui nous permet de traiter une quantité massive de données sensibles de santé, qui permet de personnaliser la médecine et d'essayer de deviner les trajectoires d'un patient lors de son parcours de soins.

### **Flora Bastiani**

Ce que je comprends, c'est qu'y a deux aspects. Il y a d'abord dépasser l'expérience personnelle du médecin, c'est-à-dire réunir un grand nombre de données. J'ai l'impression que cela a une visée de collecte aussi, d'étendre une expérience beaucoup plus largement que ce qu'un praticien va connaître dans sa carrière, par exemple.

### **Fabrice Ferré**

Absolument, d'étendre une expérience, d'étendre une expertise, d'être en capacité de fournir un diagnostic plus précis, plus fiable, d'être prédictif et au bénéfice des patients.

### **Flora Bastiani**

Et d'un autre côté, il y a cet aspect d'adapter toutes ces connaissances très massives à une situation singulière. On part d'un moment singulier où un médecin, dans une rencontre avec un patient, va programmer un examen en particulier. Ces données vont être à grande échelle, récupérées, pour en faire des synthèses, pour en tirer des connaissances et ensuite cela va à nouveau revenir à l'espace singulier de la rencontre avec le patient.

### **Fabrice Ferré**

Effectivement, à l'échelle de l'individu. Et c'est bien là l'enjeu. Jusqu'à présent, on avait cette grande difficulté qui était la variabilité interindividuelle, la variabilité entre les individus. Lorsqu'on réalisait des études cliniques, des tests cliniques assez classiques, on essayait à partir d'un échantillon de population, de tirer des conclusions et ensuite de les appliquer à l'échelle d'un individu. On le verra peut-être avec les jumeaux numériques, on peut maintenant véritablement identifier la trajectoire d'un patient, la trajectoire d'un individu, celle qui lui correspond le mieux pour personnaliser son traitement et améliorer son pronostic, notamment.

### **Flora Bastiani**

D'accord. Donc c'est très axé sur la personnalisation aussi. C'est ce dont on ne se rend pas forcément compte quand on voit l'inquiétude des personnes qui se demandent ce qu'on va faire de leurs données. C'est très peu présent dans l'esprit collectif de dire que ces données ne sont pas captées à des fins qui ne leur serviront pas.

### **Fabrice Ferré**

Tout à fait. Lorsqu'on collecte massivement des données de santé, il y a d'abord une dépersonnalisation d'une certaine façon. Il faut en faire la preuve également, mais c'est probablement pour ensuite se recentrer sur un individu en particulier et lui proposer la meilleure prise en charge.

## **Données de santé générées dans l'environnement du patient**

### **Flora Bastiani**

Les données de santé sont parfois générées par les patients eux-mêmes. Là, on parle des examens médicaux, des imageries, etc. Mais il y a aussi des données saisies dans les applications de santé, les suivis de cycle, de régime, de sommeil, issues des montres connectées, des pèse-personnes connectés, des glucomètres connectés... Tous les trackers d'activité physique génèrent des données quotidiennes. Et finalement, nous sommes sans cesse en train de transmettre des données concernant notre santé, mais sans forcément y penser. À qui appartiennent toutes ces données et est-ce qu'elles peuvent faire l'objet d'un recueil pour la recherche ? C'est un autre cadre qui ne semble pas du tout organisé, en tout cas vu de l'extérieur.

## **Fabrice Ferré**

Peut-être que Noémie donnera la réponse qu'il faut. J'aimerais juste dire une chose, c'est qu'on est probablement face à un nouveau mode d'exercice professionnel. En tout cas pour nous, professionnels de santé. On a maintenant la capacité de recueillir des données à très haute fréquence. On prend par exemple ces notions de montres connectées qui sont capables de donner une fréquence cardiaque et même une variabilité de la fréquence cardiaque chez des patients qui sont finalement dans leur environnement habituel et qui donc sont écologiquement intégrés dans leur milieu. Jusqu'à présent, on n'était pas capables de recueillir cela, mais nous ne savions pas comment analyser ce genre de données. Maintenant qu'elles existent, c'est probablement la source de futures découvertes en médecine. Plutôt qu'un patient qui vient ponctuellement tous les six mois à la consultation de son médecin généraliste à qui on prend la tension artérielle ou la fréquence cardiaque qui serait probablement biaisée par le contexte, on est capable de recueillir ces données chez des patients qui sont écologiquement intégrés dans leur environnement. C'est probablement une porte qui s'ouvre vers une nouvelle médecine, une médecine qui pourrait presque être continue dans la mesure où on peut collecter ces données à très haute fréquence. Et encore une fois chez des patients qui sont par exemple à leur domicile ou qui sont en train de se balader à l'extérieur.

## **Flora Bastiani**

Cela permet en quelque sorte de dépasser l'effet blouse blanche d'une tension qui augmente, par exemple.

## **Fabrice Ferré**

Absolument.

## **Noémie Dubruel**

Dépasser l'effet blouse blanche et avoir, comme tu le disais Fabrice, ce contexte de données qui va être bien représentatif de la personne dans son environnement donné et dans son mode de vie. Avant, on ne savait pas comment les collecter et les considérer, mais il y avait aussi cette notion de données qui sortaient du contexte clinique, notamment pour la recherche qui n'a pas tout à fait encore de valeur, qui n'est pas produite dans un cadre méthodologique tel qu'il est entendu dans le domaine de la recherche. Donc cela n'a pas encore beaucoup de valeur, mais c'est en train d'être considéré. On

parle notamment des données de vie réelle qui sont en train d'être considérées par l'Agence européenne du médicament pour être utilisées en recherche.

**Flora Bastiani**

D'accord.

**Noémie Dubruel**

Ce qui est intéressant aussi, c'est que dans le cadre de mes recherches, je suis retombé sur un document de la CNIL (Commission nationale de l'informatique et des libertés), qui avait déjà publié en deux-mille-quatorze un rapport sur le corps hyper connecté des personnes par l'utilisation d'outils connectés. Elle considérait que c'était une nouvelle façon de produire de la donnée qui était massive et qui dépassait un peu l'individu lui-même, alors qu'il est lui-même acteur de la production de ces données. C'est assez intéressant de voir que la CNIL pensait déjà à ces sujets en deux-mille-quatorze.

Pour répondre à la question de la propriété des données, c'est une question très importante pour les juristes notamment. Je travaille beaucoup avec des médecins dans le cadre de la recherche, et c'est assez difficile d'avoir ces discussions-là. Ces médecins qui font de la recherche, qui travaillent sur ces données, considèrent que ce sont leurs données alors que ce n'est pas vrai. On entend beaucoup cette notion de propriété des données, notamment celles de santé. Alors, ce qu'il faut savoir, c'est que la donnée est vraiment rattachée au privé de la personne. C'est considéré comme un attribut du personnel.

**Flora Bastiani**

D'accord.

**Noémie Dubruel**

Pour autant, on n'a pas de propriété sur nos propres données. Personne n'est propriétaire des données pour être clair. Il y a vraiment cette logique que cela fait quand même partie de la vie privée, d'une information qu'on veut formuler et rendre publique, ou pas. Donc on a cette logique de protection de la vie privée qui s'applique, de respect de la dignité humaine, derrière tout cela. Donc on ne peut pas vendre en France du moins, et en Europe, ces données, contrairement à d'autres pays. On ne peut pas s'approprier des données.

Par contre, ce qui est un peu ambigu, c'est que le travail qui est réalisé sur des



Penser la santé : Données numériques de santé, partie une

bases de données, le fait de les agréger, de les travailler, de les nettoyer, etc. Ce que font des médecins, des chercheurs, cela peut avoir une valeur monétaire. Par exemple, quand on fait de la recherche, on va faire des contrats de partage de données qui peuvent avoir une valeur financière, mais c'est la base donnée en soi, ce n'est pas la donnée qu'on partage et qu'on vend, c'est le travail qui a été réalisé sur ces données. C'est le fait de pouvoir les exploiter.

**Flora Bastiani**

C'est une sorte de propriété intellectuelle ?

**Noémie Dubruel**

C'est cela, c'est le code qui s'applique là-dessus.

**Fabrice Ferré**

J'ai une question pour toi, Noémie. Si je dis que les données appartiennent aux patients et qu'ils les mettent à disposition en l'absence d'opposition, est-ce que j'ai raison de l'expliquer ainsi à mes patients lorsqu'ils me posent la question ?

**Noémie Dubruel**

Tu peux très bien le formuler comme cela. Juridiquement parlant, il n'y a pas de propriété, mais clairement, cela fait partie d'eux. Donc je pense que tu peux très bien le formuler comme cela sans problème. Après c'est des droits d'opposition ou de consentement sur ce que je veux divulguer de ma vie, ce que je ne veux pas divulguer, ce que je veux partager...

**Flora Bastiani**

Après, comme toutes les études, il y a une possibilité d'irréversibilité : une fois que j'ai accepté, comme les données sont anonymisées, je ne peux pas les retirer puisqu'elles sont prises dans un flot de données.

**Noémie Dubruel**

Absolument. Depuis le début, on parle de données à caractère personnel, qui sont toujours rattachées à un individu. Quand on pseudonymise une liste des données, on est encore dans de la donnée à caractère personnel. Par contre, quand on anonymise des données, on a des critères qui sont assez clairs pour faire cela, ce n'est vraiment pas facile, surtout dans le domaine de la santé, on

Penser la santé : Données numériques de santé, partie une

se détache totalement de ces enjeux et la donnée devient un objet commercial sans aucun problème. Enfin, on n'a plus du tout le lien avec la personne.

### **Fabrice Ferré**

On s'est assuré qu'on était incapables de ré-identifier les patients à partir de leurs données anonymisées.

### **Flora Bastiani**

Merci beaucoup Noémie Dubruel et Fabrice Ferré. Nous nous retrouverons dans une deuxième partie pour poursuivre ces échanges. C'était Flora Bastiani pour *Penser la santé*, le podcast du réseau de recherche Penser la santé de l'Université Toulouse Jean-Jaurès.