



# Contre le cancer, la puissance des data

Grâce aux progrès informatiques, il est possible de générer, stocker et analyser des volumes considérables de données. De quoi permettre à la cancérologie d'entrer dans une nouvelle ère, celle de la « médecine augmentée ».

Par Émilie Gillet







L'analyse des données associée à l'expertise des médecins est prometteuse, particulièrement dans la lutte contre le cancer, où les recherches sont nombreuses.

« **N**ous vivons une révolution numérique tirée non seulement par l'abondance de données, mais également par notre capacité à collecter, stocker et analyser les informations »,

déclare Xosé Fernández, Directeur des data à l'Institut Curie. Avec l'avènement du numérique et les progrès informatiques, le monde de la santé est entré dans l'ère des big data. L'analyse de ces données par des programmes informatiques (ou algorithmes) associée à l'expertise des médecins est prometteuse, particulièrement dans la lutte contre le cancer, où les recherches sont nombreuses. « Les capacités des ordinateurs sont quasi illimitées : ils peuvent analyser des milliards d'informations, à une vitesse considérable et en comparant une multitude de paramètres, alors que le cer-

veau humain ne peut guère gérer plus de 5 ou 7 variables en même temps », résume Irène Buvat, directrice d'une unité de recherche au Service hospitalier Frédéric Joliot (CEA, Orsay).

### LES DONNÉES, UNE SOURCE INTARRISSABLE

En France, les données de santé bénéficient d'un écosystème unique au monde. En 2016, la loi de modernisation du système de santé a en effet créé le Système national des données de santé (SNDS). Il compile les données publiques (Assurance maladie, hôpitaux, causes médicales de décès et données relatives au handicap), couvrant ainsi plus de 90 % de la population française. Le SNDS permet notamment de retracer les parcours de soins des patients jusqu'à leur décès, de 2006 à aujourd'hui. L'objectif est de mettre à disposition ces données, sous le contrôle de la Commission nationale de l'informatique et des libertés (Cnil), afin de favoriser les études, recherches ou évaluations présentant un caractère d'intérêt public.

À ces vastes bases s'ajoutent depuis quelques années des données cliniques et biologiques très détaillées, issues des dossiers médicaux des patients pris en charge dans les établissements de santé : informations personnelles, résultats d'examens médicaux, comptes rendus chirurgicaux, protocoles thérapeutiques, etc. Il peut aussi s'agir de données issues de programmes de recherche. Leur recueil est encadré par le règlement général sur la protection des données (RGPD), un texte européen très protecteur transposé dans la loi française en 2018. Il stipule notamment que les personnes dont les données sont recueillies



et utilisées à des fins de recherche doivent en être clairement informées et qu'elles ont un droit d'opposition à leur utilisation pour cette recherche. Seuls les médecins en contact avec les patients ont accès aux données personnelles, tous les autres ne peuvent travailler que sur des données anonymisées. Enfin, toutes ces informations étant sensibles, elles sont stockées dans des conditions strictes de sécurité par des hébergeurs (ou datacenters) certifiés par le ministère de la Santé.

### STRUCTURER ET ANALYSER LES DONNÉES

Générer et enregistrer chaque jour des millions de données de santé est une chose. Mais encore faut-il s'assurer qu'elles soient de bonne qualité et structurées. À quoi servirait en effet une bibliothèque si les livres étaient incomplets ou rédigés en langues inconnues ? « *Aujourd'hui, les progrès informatiques nous permettent de remplir les bases indistinctement avec des données médicales structurées, des textes bruts, des résultats d'analyses, des données génomiques, peut-être un jour des images...* », raconte Emmanuel Reyrat, Directeur du département des systèmes d'informations à Unicancer, qui réunit l'ensemble des centres de lutte contre le cancer. *Mais pour pouvoir ensuite travailler dessus, les bases doivent être correctement organisées.* » C'est l'objectif du projet ConSoRe mis en place par Unicancer. « *Il s'agit d'un système qui collecte, analyse et structure les données, associé à un moteur de recherche multicritère. Cela permet notamment à des médecins d'identifier des patients répondant à des critères de recherche précis, de visualiser graphiquement l'évolution de leur maladie et leurs traitements, de localiser des dossiers de cancers rares, de savoir si des cas similaires ont déjà été traités ailleurs. L'objectif de ConSoRe est de faciliter la recherche d'informations et leur partage entre médecins et chercheurs de différents centres de lutte contre le cancer.* » Du fait de leur volume et de leur fragmentation, analyser les données de santé n'est pas

toujours simple. Pour cela, les chercheurs utilisent notamment l'intelligence artificielle (IA). Ce terme regroupe différentes méthodes de programmation informatique qui imitent des modes de raisonnement propres aux humains. On distingue l'IA « faible », qui repose sur des algorithmes mis au point par l'homme et permettant à une machine d'exécuter une tâche comme par exemple établir un diagnostic ou proposer une thérapeutique en fonction de critères prédéfinis. L'IA « forte » a pour but de créer des programmes permettant aux machines d'apprendre par elles-mêmes le bon algorithme : on parle de *machine learning*. Plus la machine s'entraîne et plus elle est capable de reproduire le raisonnement humain, à condition de lui avoir fourni au départ les bons exercices avec les corrigés. L'objectif est que les algorithmes fassent des rapprochements, par exemple des similitudes moléculaires ou génétiques entre des cancers touchant des organes très différents, auxquels un médecin n'aurait pas forcément pensé.

### DES PERSPECTIVES POUR LES MALADES

De nombreux laboratoires de recherche se sont emparés de ces nouveaux outils d'IA. Récemment, plusieurs équipes internationales ont ainsi mis au point des algorithmes capables de diagnostiquer un mélanome. Les programmes ont d'abord appris à reconnaître un mélanome en examinant et en comparant des dizaines de milliers d'images de lésions cancéreuses et de simples grains de beauté, puis ont été capables d'identifier des tumeurs atypiques avec plus de réussite que des médecins. « *IRM, mammographie, TEP scan... Les images sont des données extrêmement intéressantes aujourd'hui pour l'IA, car elles contiennent énormément d'informations. Si elles sont de qualité, et associées à des dossiers cliniques bien renseignés, c'est une excellente base de travail. Qui plus est, les algorithmes permettent de guider les variations d'interprétation qui existent entre deux médecins,*

### Un partenariat Institut Curie-Owkin pour mieux analyser les dossiers cliniques

En 2017, l'Institut Curie a signé un partenariat inédit avec la start-up française Owkin, qui développe des services d'intelligence artificielle. Objectif : « *Valoriser le potentiel de données accumulées depuis plus de quinze ans à l'Institut Curie : extraire de la connaissance, faire avancer la recherche et développer ainsi de nouveaux outils pour mieux traiter nos patients* », résume le Dr Alain Liwartowski, Directeur adjoint des data de l'Institut Curie. Owkin développe en effet des technologies de *machine learning* : il s'agit de logiciels qui, en examinant des volumes considérables de données, apprennent seuls à en faire ressortir des informations pertinentes, par exemple des similitudes. L'une des applications pourrait être par exemple la découverte de nouveaux éléments permettant de prédire la toxicité et la réponse à différents traitements de chimiothérapie. Point fort de ces technologies, elles sont capables d'analyser des données en langage naturel ou des images qui n'ont pas été codées informatiquement.



lStock

# Les données de santé à l'Institut Curie



## Les données de santé, qu'est-ce que c'est ?

- **Les informations médicales des patients traités à l'Institut Curie :** antécédents, résultats des examens biologiques, génétiques et d'imagerie médicale, comptes rendus d'opérations, traitements et effets secondaires...
- **Les données issues des recherches :** analyses génétiques, moléculaires, questionnaires de qualité de vie...

L'Institut Curie traite 54000 patients par an, dont 9000 nouveaux patients.

Plus de 5 pétaoctets de données (5 millions de Go), soit environ 500000 dossiers médicaux.

## D'où viennent-elles ?

- **Des dossiers médicaux** intégralement numériques depuis 2004
- **Des données de recherche**
- **Des données personnelles** déclarées par les patients, via l'application MyCurie depuis fin 2018.



## Où sont-elles stockées ?

**Dans les entrepôts de données (datacenters)** de l'Institut Curie, situés à Paris, Orsay et Saint-Cloud. Les données sont sauvegardées en plusieurs lieux pour plus de sécurité.



## Quelles sont les missions de la Direction des data de l'Institut Curie ?

- Structurer et organiser leurs archivages
- Développer des outils pour extraire les données et les analyser afin que la recherche progresse.
- Contrôler la qualité des données récoltées

L'Institut Curie est le premier centre anticancer à s'être doté d'une telle structure.

## Quels enjeux pour les chercheurs ?

- Disposer d'une quantité considérable de données et des outils pour les analyser, les comparer, et identifier de nouvelles pistes de recherche.
- Améliorer les techniques de diagnostic, de pronostic et de décisions thérapeutiques.
- Ouvrir une nouvelle ère, celle de la « médecine augmentée » où les médecins travailleront en s'appuyant sur les données de santé et l'intelligence artificielle.





Les images, notamment issues des IRM, sont des données extrêmement intéressantes pour l'intelligence artificielle.



explique Irène Buvat. *Les programmes apprennent seuls à faire des rapprochements. Cela permet de défricher énormément de données et de découvrir de nouvelles pistes de recherche.* » Avec la même approche, une équipe Inserm-Université Paris Sud a ainsi suggéré qu'il pourrait être envisageable de prédire la réponse à l'immunothérapie de certaines tumeurs solides à partir d'une simple image scanner. Une autre, au Centre de recherche des Cordeliers, à Paris, a développé un algorithme qui pourrait déterminer à l'avance les patients ayant une réponse complète à la radiochimiothérapie afin de leur éviter une opération pour un cancer du rectum. De telles avancées devraient aussi être possibles à partir de données moléculaires ou génétiques. Certaines équipes utilisent encore l'IA pour analyser des milliers de génomes dans le cadre de la recherche sur les cancers rares ou pour déterminer de nouvelles cibles thérapeutiques par exemple.

### ÉTHIQUE ET GARANTIE HUMAINE

« En théorie, data et intelligence artificielle permettent d'ouvrir de nombreuses voies de



**« L'implication de patients dans les comités d'éthique qui valident les protocoles de recherche est une meilleure garantie pour la protection de leurs données. »**

-

**Anne Schweighofer, 49 ans, formée à l'Université des Patients et créatrice de l'entreprise Patient Conseil**

*recherche, déclare Irène Buvat. Mais nous devons être très vigilants quant à la rigueur de notre travail si nous ne voulons pas faire naître de faux espoirs ! Par ailleurs, l'objectif n'est surtout pas de remplacer les médecins mais de les aider, de faire en quelque sorte de la "médecine augmentée".* »

En novembre dernier, le Comité consultatif national d'éthique (CCNE) s'est exprimé à propos de l'intelligence artificielle dans le domaine de la santé. Selon son rapport, « le risque principal serait de ne pas avoir suffisamment recours à l'IA, explique David Gruson, codirecteur du groupe



GRÂCE  
À VOUS

### La campagne MC21 pour le projet big data

Lancée en novembre 2017, la campagne MC21 (Marie Curie XXI<sup>e</sup> siècle) est une campagne de collecte destinée à soutenir le financement de quatre projets essentiels et innovants portés par l'Institut Curie. Parmi ces projets, le développement du big data figure en bonne place avec un besoin de financement à hauteur de 10 millions d'euros d'ici à 2021. Porté par un comité de campagne composé de 12 membres très impliqués tant personnellement qu'à travers leurs réseaux, la campagne a déjà permis de collecter près de 4 millions d'euros ([vaincrelecaner.curie.fr](http://vaincrelecaner.curie.fr)).

Grâce à la campagne MC21, l'une des premières entreprises à s'être associée à la campagne est le groupe Bessé-Conseil en assurances, qui s'est engagé dans le soutien du projet big data sur plusieurs années. « L'objectif de notre engagement auprès de l'Institut Curie est de rendre un service immédiat aux patients atteints de cancer et de se préoccuper de l'avenir en poussant plus loin le progrès scientifique au bénéfice du plus grand nombre », déclare Pierre Bessé, président du groupe.

En savoir plus :

[vaincrelecaner.curie.fr](http://vaincrelecaner.curie.fr)



PAROLE  
D'EXPERT

**XOSÉ FERNÁNDEZ**  
**DIRECTEUR DES DATA**  
**À L'INSTITUT CURIE**



Uniel Chametaine / Institut Curie

### Quelle est l'importance des data à l'Institut Curie ?

Dès l'an 2000, l'Institut Curie a commencé à numériser

les dossiers médicaux de ses patients, et en 2004 nous avons complètement abandonné le papier : données personnelles, comptes rendus d'examen, d'opérations, analyses génétiques... mais aussi toutes les données issues de la recherche, tout est informatisé ! Nous avons été le premier centre français de lutte contre le cancer à adopter cette stratégie du "tout numérique", puis à créer une Direction des data.

### Vous dirigez cette Direction des data. Quels sont ses rôles ?

L'objectif principal est de tisser des liens étroits entre la clinique et la recherche,

*c'est-à-dire permettre aux chercheurs d'utiliser ces données au mieux pour découvrir de nouvelles pistes thérapeutiques, par exemple, ou optimiser le diagnostic des cancers. Pour cela, nous devons réfléchir à la meilleure façon de classer ces données, de les structurer, afin qu'elles soient accessibles le plus facilement possible.*

### Vous réfléchissez aussi aux outils à développer pour les analyser...

*Oui, grâce à des partenariats avec d'autres équipes de recherche mais aussi avec des entreprises privées, nous mettons au point des outils d'intelligence artificielle pour exploiter au mieux ce volume considérable d'informations. En parallèle, nous travaillons à la rédaction d'une charte qui définira précisément la protection de ces données et les utilisations que nous pouvons en faire. Grâce aux data et à l'intelligence artificielle, nous sommes à la veille d'une grande révolution dans la prise en charge des cancers.*

### Croiser les données publiques avec celles de l'Institut Curie

Le Système national des données de santé (SNDS) regroupe les données individualisées et anonymes de l'intégralité des remboursements de dépenses de santé pour les sujets affiliés à l'un des régimes obligatoires de Sécurité sociale, les données des hôpitaux et les causes médicales de décès. Cela représente plus de 60 millions de Français et 450 téraoctets de données ! Au sein de l'équipe data de l'Institut Curie, dirigée par Xosé Fernández, ces données de santé publiques sont croisées avec celles de l'Institut Curie. Concrètement, cela permet de connaître de façon exhaustive le parcours de soins des patients. Par exemple, il est possible d'identifier les grossesses des patientes traitées pour un cancer du sein à l'Institut Curie, ou celles qui ont reçu tel autre traitement ailleurs pendant leur chimiothérapie. Grâce à ces données de vie réelle, c'est un nouveau champ de recherche qui s'ouvre. *In fine*, l'objectif est de fluidifier et d'améliorer les parcours de soins en cancérologie, et de mieux comprendre leur impact sur la survie des patients.

## 72 ➤

**PROJETS DE RECHERCHE** portant sur les données de santé ont été évalués par l'équipe data de l'Institut Curie.

de travail du CCNE. Nous avons beaucoup de données de santé en France, il ne faudrait pas les sous-exploiter. » Un autre risque important serait de surréglementer ce secteur. « Le RGPD est déjà suffisamment protecteur. » Le rapport souligne aussi deux autres risques inhérents à l'IA. Celui de la tension entre individus et collectif : « La logique de santé publique, notamment d'un point de vue économique, n'est pas toujours compatible avec celle qui prime à l'échelle de la santé d'une personne. Nous devons anticiper ce risque avec des législations particulières qui obligerait par exemple les médecins à informer leurs patients qu'ils ont

utilisé un algorithme pour établir un diagnostic ou faire un choix thérapeutique. » Enfin il convient d'établir un principe de garantie humaine de l'IA : que le sens critique de l'être humain prédomine toujours par rapport à la machine ! Un enjeu que le mathématicien et député Cédric Villani a bien pris en compte lorsque, rendant son rapport sur le développement de l'IA en France en mars 2018, il a déclaré : « L'enjeu n'est pas la compétition entre deux intelligences mais au contraire leur association ; comment l'intelligence humaine pourra utiliser l'intelligence de synthèse pour se faciliter la tâche. »



## VOTRE FONDATION

La continuité de la recherche et des soins dans un même lieu – l'Institut Curie – stimule l'innovation, favorise les échanges et les découvertes. Fondation privée reconnue d'utilité publique, l'Institut Curie est habilité à recevoir les dons et legs du public. Notre volonté de progresser est encouragée par le soutien et la générosité de nos donateurs, testateurs et partenaires, que je remercie chaleureusement.

**Pr Thierry Philip**, président du Directoire de l'Institut Curie

## Mettez votre IFI 2019 au service de l'innovation contre le cancer !

**La loi de finances 2018 a apporté des modifications importantes, comme la suppression de l'impôt de solidarité sur la fortune (ISF) et son remplacement par l'impôt sur la fortune immobilière (IFI).**

Cette année, si vous êtes assujetti à l'impôt sur la fortune immobilière (IFI), vous pouvez choisir de mettre la totalité ou une partie de cet impôt au service de la lutte contre le cancer aux côtés de l'Institut Curie. En effectuant un don déductible de votre impôt sur la fortune immobilière 2019, vous contribuez aux grands programmes d'innovation de l'Institut Curie et faites reculer le cancer.

### LES RESSOURCES PRIVÉES, VITALES POUR L'INSTITUT CURIE

Les dons des personnes assujetties à l'IFI sont un formidable accélérateur pour mener à bien nos missions de recherche et de soins. Cette année, ces donateurs contribueront à faire avancer nos programmes de recherche, notamment le projet européen de recherche PEVO Data,

un essai clinique coordonné par l'Institut Curie qui associera épigénétique, immunologie et data pour soigner les patients touchés par un cancer de la peau en récurrence (carcinome épidermoïde cutané). En 2019, grâce à votre appui, cette nouvelle approche multidisciplinaire pourra être mise en œuvre afin de comprendre les mécanismes de résistance des médicaments et d'identifier de nouvelles voies thérapeutiques.

Choisir de faire un don à l'Institut Curie au titre de l'IFI vous permet de décider de l'affectation de votre impôt tout en permettant à nos équipes de chercheurs et médecins de disposer de la totalité de votre don, au service des patients. Pour bénéficier de cette mesure fiscale et en faire profiter nos médecins et chercheurs, vous pouvez, en toute confidentialité, effectuer votre don dès aujourd'hui. Il sera déductible à 75 % de votre IFI dans la limite de 50000 euros par an.

### VOTRE CONTACT

**Elsa Oliveira**  
Institut Curie  
26 rue d'Ulm  
75248 Paris Cedex 05  
Tél. : 01.56.24.55.02  
elsa.oliveira@curie.fr



Uriel Chantraine/Institut Curie

### Les modalités changent :

**La déclaration IFI doit être effectuée sur l'annexe n° 2042-IFI, dans les mêmes délais que la déclaration d'impôt sur le revenu : le 17 mai pour les déclarations papier, entre le 22 mai et le 5 juin pour les déclarations en ligne (la date varie selon votre département d'habitation).  
Détails sur le site [impots.gouv.fr](http://impots.gouv.fr)  
Votre don doit être effectué avant la date de déclaration afin d'être pris en compte.**

### Le saviez-vous?

Si vous n'êtes pas soumis à l'IFI cette année, vous bénéficiez toujours de la possibilité de réduire votre impôt sur le revenu de 66 % du montant de vos dons à l'Institut Curie, dans la limite de 20 % de votre revenu imposable.