

QUELLE VALORISATION POUR L'USAGE SECONDAIRE DES DONNÉES DE SANTÉ ?

Note détaillée

1 LA VALEUR LIÉE À L'USAGE SECONDAIRE DES DONNÉES DE SANTÉ DEVRAIT FORTEMENT AUGMENTER AU COURS DES PROCHAINES ANNÉES

Il existe un accord dans la littérature sur l'intérêt économique grandissant de l'usage secondaire des données de santé. Cependant, à date, peu d'études quantifiées sont disponibles :

- La Commission estime qu'au niveau européen, la valeur d'usage des données secondaires va passer de 25 Md€ en 2020 à 43 Md€ en 2028¹, soit une hausse de 70 %. En reprenant la part de la France dans le PIB Européen (en 2019²), on peut estimer un gain de **7,3 Md€** pour la France.
- Pour le Royaume-Uni, E&Y³ estime que les gains seront de 5,0 Md£ pour le NHS et 4,6 Md£ pour les patients. La Commission a repris ce chiffrage pour l'ensemble des pays européens⁴. Pour la France, la Commission estime les gains à **4,6 Md€** par an pour le gouvernement, et **4,2 Md€** pour les patients à long terme.

Il ressort de plusieurs entretiens que **ces estimations seront un minorant des gains futurs**. En effet, elles ne portent que sur les usages connus et anticipables des données de santé dans des secteurs balisés comme l'industrie pharmaceutique. Elles manquent toutes les innovations non anticipables à date qui devraient augmenter les gains globaux. Ces innovations radicales sont complexes à estimer. Par ailleurs, il convient non seulement de prendre en compte les créations d'entreprises qui se vont se développer sur l'usage secondaire de la donnée de santé mais aussi les entreprises qui ont ou vont pivoter dans ce secteur.

Comme mentionné plus haut, les analyses quantifiées au niveau des différents secteurs sont **surtout disponibles pour le secteur de la pharmacie**. IQVIA⁵ estime que sur la période 2021-2023, les gains liés aux données de vie réelle seront de 4 Md\$ pour l'industrie pharmaceutique dans le monde, sur l'ensemble de la chaîne de valeur.

- Dans la phase d'innovation⁶, les gains seraient de 1,75 Md\$ (dont 0,5 Md\$ pour la réduction des coûts). La phase d'innovation est donc là où se situe l'essentiel des gains ;
- Dans la phase d'enregistrement, les gains seraient de 0,6 Md\$ (dont la moitié des gains pour la réduction des coûts) ;
- Sur la phase d'accès et de couverture, les gains de 0,3 Md\$ sont concentrés sur les revenus ;
- C'est également le cas pour la phase de diagnostic et de traitement (0,75 Md\$ de gains) ;
- Enfin, pour la phase de commercialisation, les gains sont estimés à 0,6 Md\$ (dont 0,1 Md\$ pour la baisse de coût).

Les différents entretiens convergent sur le fait qu'il existe peu d'estimations dans le domaine des MedTech. Néanmoins, **il est indéniable que les données jouent un rôle de plus en plus grand également dans ce**

¹ Impact Assessment report Accompanying the document "PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space" (1/4)

² Source : Eurostat

³ E&Y, 2019, "Realising the value of health care data: a framework for the future"

⁴ Lupiáñez-Villanueva, F., Gunderson, L., Vitiello, S., Febrer, N., Folkvord, F., Chabanier, L., Filali, N., Hamonic, R., Achard, E., Couret, H., Arredondo, M. T., Cabrera, M. F., García, R., López, L., Merino, B., Fico, G. (2022). Study on Health Data, Digital Health and Artificial Intelligence in Healthcare, Publications Office of the European Union.

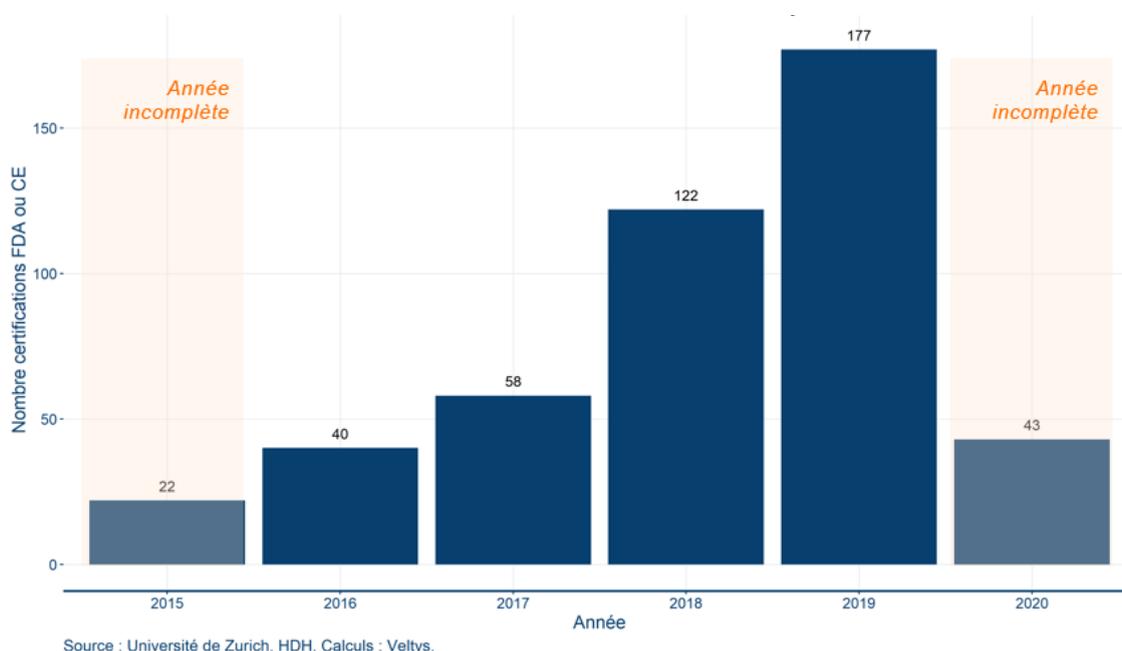
⁵ Gores M. and E. Powers, 2021, *Capturing Value at Scale: The \$4 billion RWE Imperative*

⁶ Le document d'IQVIA distingue les phases d'innovation (R&D et découverte du médicament), d'enregistrement (validation auprès des autorités de santé), d'accès-couverture (favoriser un accès rapide et pour toutes les populations), de diagnostic (médecine de précision notamment) et commercialisation (modèle de vente, marketing)



secteur. Par exemple, le nombre de dispositif MedTech certifiés par un marquage FDA ou CE et incorporant de l'intelligence artificielle a été multiplié par plus de 4 entre 2016 et 2019⁷.

Figure 1 : certification FDA ou CE avec de l'intelligence artificielle



2 UN FORT DÉVELOPPEMENT DES ENTREPRISES MOBILISANT DES DONNÉES DE SANTÉ EST EN COURS DANS LE MONDE

De nombreuses startups centrées sur l'usage secondaire des données de santé se sont développées. **Ainsi, on compte au moins à date 6 licornes⁸ s'appuyant sur ce type d'usage** (Innovaccer, Komodo Health, Evidation Health, K Health, Clarify Health Solutions, Olive). Toutes ces licornes sont actuellement situées aux Etats-Unis⁹. De la même façon, on compte également au moins une quinzaine d'entreprises ayant réussi des levées de fonds pour plus de 100 M€ dans ce domaine. Peu de ces entreprises sont européennes (à date, seules Exscientia et BenevolentAI qui sont britanniques, ont été identifiées). Les montants engagés témoignent de l'intérêt des différents acteurs pour ces solutions. Ainsi, plusieurs rachats importants ont déjà eu lieu (Flatiron a ainsi été racheté par Roche en 2018 pour 1,8 Md€).

Ces startups ont les caractéristiques suivantes :

- Beaucoup de ces entreprises sont des plateformes, avec plusieurs modèles. Certaines mettent à disposition des données structurées et anonymisées (notamment *via* des abonnements) ainsi que des outils d'IA entraînés sur ces données. C'est notamment le cas de Flatiron Health ou de Komodo Health (valorisée à plus de 3 Md\$). D'autres plateformes se chargent de la structuration des données de leurs clients, et le restituent, avec mise à disposition d'outils. C'est le cas d'Olive, dont la valorisation dépasse 3,8 Md\$. Ce système pose la question de l'intermédiation des producteurs de données, la valeur étant alors captée par la plateforme.

⁷ Urs J Muehlemaier, Paola Daniore, Kerstin N Vokinger, 2021, "Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis", *Lancet Digital Health*

⁸ Les licornes sont des startups valorisées à plus d'un milliard de dollars.

⁹ Babylon Health et BenevolentAI, deux entreprises britanniques ne sont plus considérées comme des licornes depuis leur introduction en bourse.



- Les données le plus souvent utilisées dans ces entreprises sont les *electronic medical records* (EMR). Néanmoins, les données génomiques semblent prendre de plus en plus d'importance. A titre d'exemple, une société comme Sema4, qui possède plus de 500 000 profils génomiques, est aujourd'hui valorisée à plus de 2,9 Md €.
- Concernant le secteur pharmaceutique, les startups identifiées comme étant les plus valorisées se situent sur la partie découverte de nouveaux médicaments, ce qui est conforme à l'analyse d'IQVIA. Par exemple, la société hongkongaise InSilico Medecine a ainsi développé grâce à de l'IA, un médicament candidat en moins de 18 mois pour un coût de 2,6 M \$ ce qui est environ 3 fois moins long et plus de 100 fois moins onéreux qu'un développement classique.
- Concernant les *Medical Devices* utilisant de l'IA, la radiologie est la spécialité médicale qui a récolté le plus de certifications FDA et CE¹⁰ ces dernières années sur ces sujets. Beaucoup d'investissements se font dans des acteurs de taille très importante (comme Siemens ou Philips). Il existe néanmoins des entreprises de taille plus réduite qui déposent de nombreux brevets sur ces sujets (par exemple, l'entreprise israélienne Aidoc Medical a développé 15 solutions cliniques approuvées par la FDA en seulement 6 ans).

De ce point de vue, la France présente un retard important, avec aucune entreprise ayant levé des fonds pour plus de 100 M€ dans le domaine. Cardiologs est une des entreprises les plus en pointe sur le sujet, mais elle ne cumule que 23 M€. Des entreprises comme Withings ou Dreem¹¹ présentent des levées plus importantes (entre 50 et 100 M€), mais leur activité n'est pas centrée principalement sur l'usage secondaire des données de santé.

3 DES PAYS AVEC NIVEAU DE MATURITÉ TRÈS DIFFÉRENTS

3.1 QUELLE DYNAMIQUE GLOBALE ?

L'OCDE a proposé une approche globale des enjeux liés à l'exploitation des données de santé, centrée autour de deux concepts :

- **La disponibilité, la maturité et l'usage dans l'utilisation des données**, calculés sur 13 bases représentatives de l'ensemble des données de santé (comme les données d'hospitalisation, d'urgence, de prescriptions de médicament, de diabète etc.).
- **La gouvernance**, soit l'ensemble des organisations et procédures mises en place dans le pays afin d'encadrer la collecte des données et leur utilisation (automatisation des demandes, durée entre la demande et l'accès aux données ...).

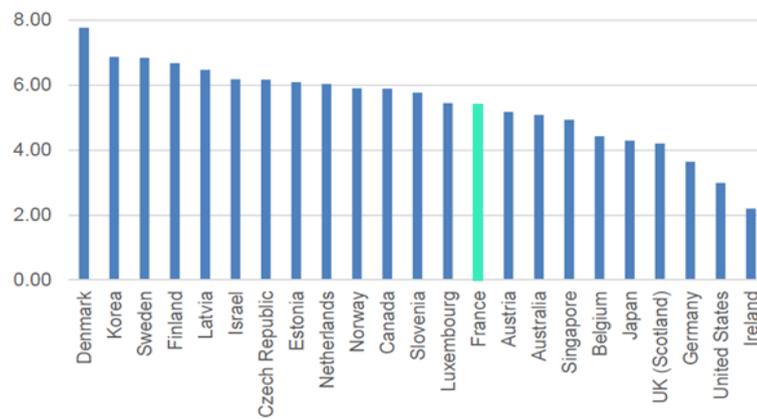
Cette approche permet de classer l'ensemble des pays, en distinguant les pays avec un haut niveau de maturité (des bases interconnectées, couvrant l'essentiel des habitants et des pathologies, et établies à partir de données médicales et non pas administratives ou financières) et une gouvernance renforcée. C'est notamment le cas de la Finlande, du Danemark et de la Corée du Sud.

¹⁰ Urs J Muehlematter, Paola Daniore, Kerstin N Vokinger, 2021, "Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis", *Lancet Digital Health*

¹¹ Withings commercialise des objets connectés (montre, balance, thermomètre etc.) et des applications associées pour visualiser les données de ces objets. Dreem fournit des produits visant à améliorer le sommeil, avec là-aussi une dimension objet connectée.

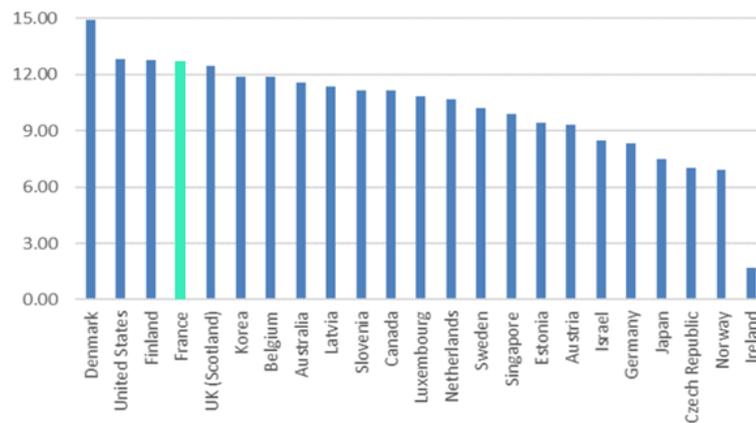


Figure 2 : Classement en termes de maturité et d'usage



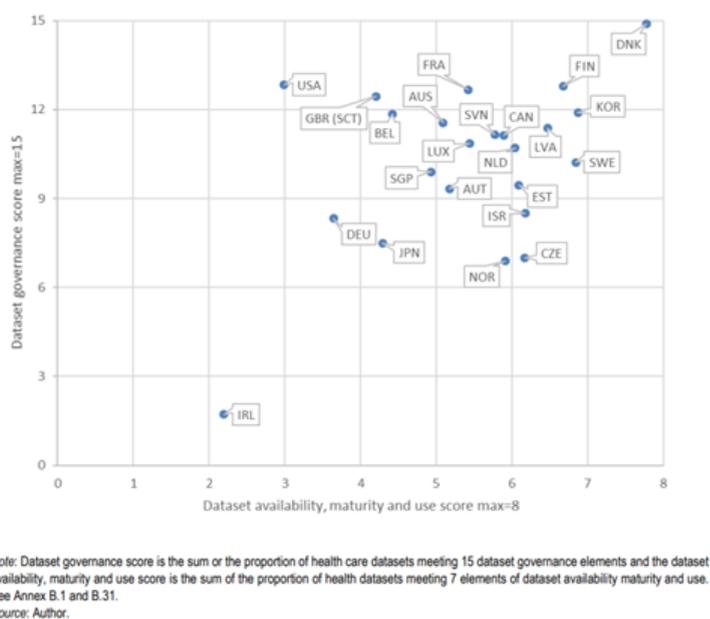
Source : **Survey results : National health data infrastructure et governance**, OECD Health Working Papers No.127, J.Oderkirk, 2021

Figure 3 : Classement en termes de gouvernance



Source : **Survey results : National health data infrastructure et governance**, OECD Health Working Papers No.127, J.Oderkirk, 2021

Figure 4 : Classement global de l'OCDE sur la maturité



Source : **Survey results : National health data infrastructure et governance**, OECD Health Working Papers No.127, J.Oderkirk, 2021

Lors d'un entretien avec l'auteur de l'étude, celui-ci a souligné **que la crise du Covid avait renforcé les écarts entre les meilleurs pays** (Danemark, Finlande et Corée du Sud notamment) et les autres, même si tous les pays avaient progressé.

Ces différences se retrouvent également dans les dépôts de brevet. En effet, l'analyse des brevets dans les BioTech¹², MedTech et Pharma sur la période 2005-2017 permet de dresser les constats suivants :

- Le nombre de brevets a augmenté **pour les MedTech (+39 %), les BioTech (+19 %)**, et a légèrement diminué pour la pharmacie (-8 %).
- Les Etats-Unis rassemblent le plus de brevets dans chacun des domaines, avec à chaque fois plus de 30 % des brevets mondiaux.
- Certains pays présentent un volume de brevets très important par rapport à la taille du pays (notamment Israël et la Corée du Sud). **Ces pays sont notamment les pays les plus en pointe en termes de données, notamment dans la classification OCDE.**
- La Chine a connu une croissance très importante au cours des dernières années.

Table 1 : Evolution des dépôts de brevet en MedTech

Pays	Rang mondial	Nombre brevets de MedTech en 2017	Evolution depuis 2005	Nombre brevets/hab de (pour 100 000 habitants)	Part de brevet mondial	Part brevet de l'OCDE
Etats-Unis	1er	5 624	+ 17 %	1,73	31 %	34 %
Japon	2 ^{ème}	3 650	+ 61%	2,89	20 %	22 %
Allemagne	3 ^{ème}	1 622	- 13 %	1,95	9 %	10 %
Chine	4 ^{ème}	1 041	+ 491 %	0,75	5,5 %	Non inclus dans l'OCDE

¹² Les chiffres portent sur l'ensemble des brevets, pas uniquement ceux liés à l'usage secondaire des données de santé



Corée du Sud	5 ^{ème}	970	+ 199 %	3,88	5,4 %	5,9 %
Pays-Bas	6 ^{ème}	639	+ 114 %	3,76	3,4 %	3,9 %
Royaume-Uni	7 ^{ème}	598	+ 17 %	0,91	3,2 %	3,7 %
France	8 ^{ème}	543	+ 22 %	0,81	2,9 %	3,3 %
Israël	10 ^{ème}	457	+ 49 %	5,08	2,4 %	2,8 %
Suisse	12 ^{ème}	314	+ 2 %	3,93	1,7 %	1,9 %

Source : OCDE

Table 2 : Evolution des dépôts de brevet en Pharmacie

Pays	Rang mondial	Nombre de brevets Pharma en 2017	Niveau d'évolution depuis 2005	Nombre de brevets/hab (pour 100 000 habitants)	Part de brevet mondial	Part de brevet de l'OCDE
Etats-Unis	1er	4 312	- 4 %	1,32	39 %	46 %
Chine	2 ^{ème}	1 101	+ 414 %	0,08	10 %	<i>Non inclus dans l'OCDE</i>
Japon	3 ^{ème}	930	- 33 %	0,74	8 %	10 %
Corée du Sud	4 ^{ème}	617	+ 123 %	2,47	5,6 %	6,7 %
Allemagne	5 ^{ème}	594	- 47 %	0,72	5,4 %	6,4 %
Royaume-Uni	6 ^{ème}	425	- 39 %	0,64	3,9 %	4,6 %
France	7 ^{ème}	417	- 25 %	0,62	3,8 %	4,5 %
Suisse	8 ^{ème}	296	+ 3 %	3,70	2,7 %	3,2 %
Israël	13 ^{ème}	168	- 22 %	1,87	1,5 %	1,8 %
Pays-Bas	14 ^{ème}	136	- 26 %	0,80	1,2 %	1,5 %

Source : OCDE

Table 3 : Evolution des dépôts de brevet en BioTech

Pays	Rang mondial	Nombre de brevets BioTech en 2017	Niveau d'évolution depuis 2005	Nombre de brevets/hab (pour 100 000 habitants)	Part de brevet mondial	Part de brevet de l'OCDE
Etats-Unis	1er	4 381	+ 26 %	1,35	39 %	45 %
Japon	2 ^{ème}	1 317	- 11 %	1,04	11,7 %	13 %
Chine	3 ^{ème}	920	+ 576 %	0,09	8,2 %	<i>Non inclus dans l'OCDE</i>
Corée	4 ^{ème}	687	+ 101 %	2,75	6,1 %	7,0 %
Allemagne	5 ^{ème}	660	- 20 %	0,80	5,9 %	6,7 %



France	6 ^{ème}	467	+ 26 %	0,70	4,2 %	4,8 %
Royaume-Uni	7 ^{ème}	438	- 8 %	0,66	3,9 %	4,5 %
Suisse	10 ^{ème}	218	+ 39 %	2,73	1,9 %	2,2 %
Pays-Bas	11 ^{ème}	186	- 18 %	1,09	1,7 %	1,9 %
Israël	12 ^{ème}	164	+ 7 %	1,82	1,5 %	1,7 %

Source : OCDE

3.2 ANALYSE SPÉCIFIQUE PAR PAYS

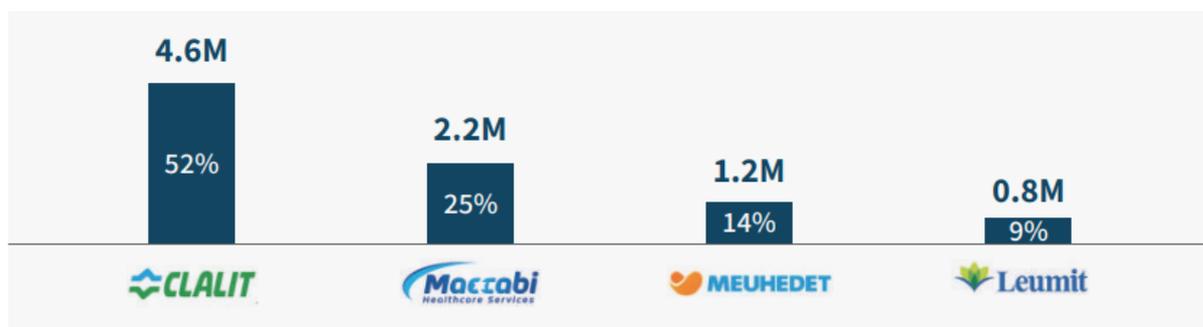
Nous avons retenus 5 pays :

- Israël, la Corée du Sud et le Danemark. Ces pays sont en effet très bien situés dans les classements d'accès aux données.
- Les Etats-Unis, qui présentent les plus forts volumes d'investissement dans ce domaine.
- Le Royaume-Uni, qui est comparable à la France dans son mode d'organisation du système de santé

3.2.1 Israël : un réseau de startups très développé

Doté d'un système de soins de santé parmi les meilleurs au monde, **Israël exploite un ensemble de données de santé numériques avec une profondeur de 25 ans**. Ces données proviennent des 4 organisations officielles d'assurance maladie auxquelles la loi israélienne oblige tous les citoyens du pays à adhérer.

Figure 5 : Répartition de la population israélienne dans les différentes HMO

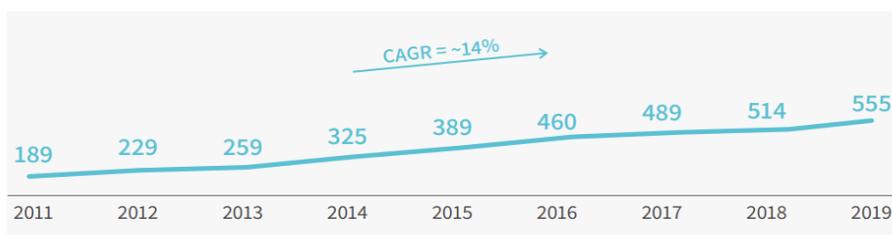


Source : *Digital Health : The Israeli Promise*, Ministry of Economy and Industry, State of Israel, March 2020

Chaque HMO israélienne rassemble les données des patients dans des *EMR*, ce qui permet d'obtenir un ensemble de données numériques complet sur sa population. D'après l'étude de l'OCDE, en 2021, les données de santé couvraient **88 % des domaines de santé et 100 % de ces données étaient extraites des EMR**.

Selon le 2019 Global Skills Index, Israël se classe au 1er rang mondial pour les compétences en science des données. En conséquence, plus de 1 200 entreprises d'IA ont été créées en Israël depuis 2010. Cette forte orientation d'Israël vers l'IA a fortement influencé le secteur de la santé numérique : en 2018, le pays accueillait plus de 550 start-ups de santé numérique (+ 193 % par rapport à 2011) .

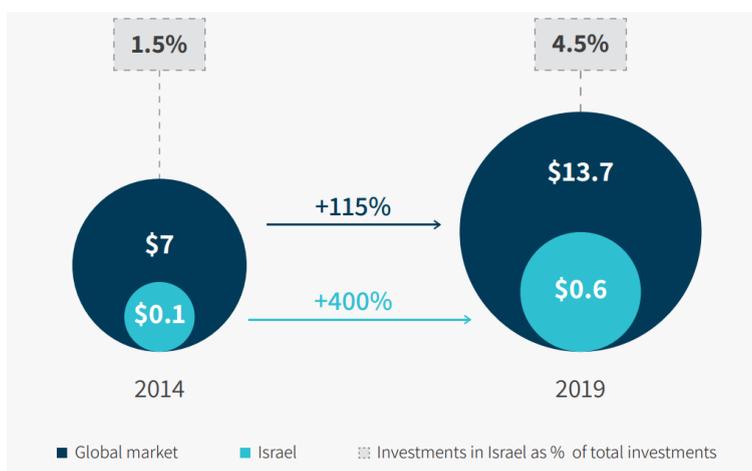
Figure 6 : Evolution du nombre d'entreprises de santé numérique en Israël (2011-2018)



Source : **Digital Health : The Israeli Promise**, Ministry of Economy and Industry, State of Israel, March 2020

Par ailleurs, la part d'investissement dans ce domaine a considérablement augmenté sur les dernières années, passant de 1,5 % des investissements mondiaux dans la santé numérique à 4,5 % entre 2014 et 2019, alors que la population du pays représente environ 0,1% de la population mondiale.

Figure 7 : Investissement annuel en santé numérique (israélien VS mondial)

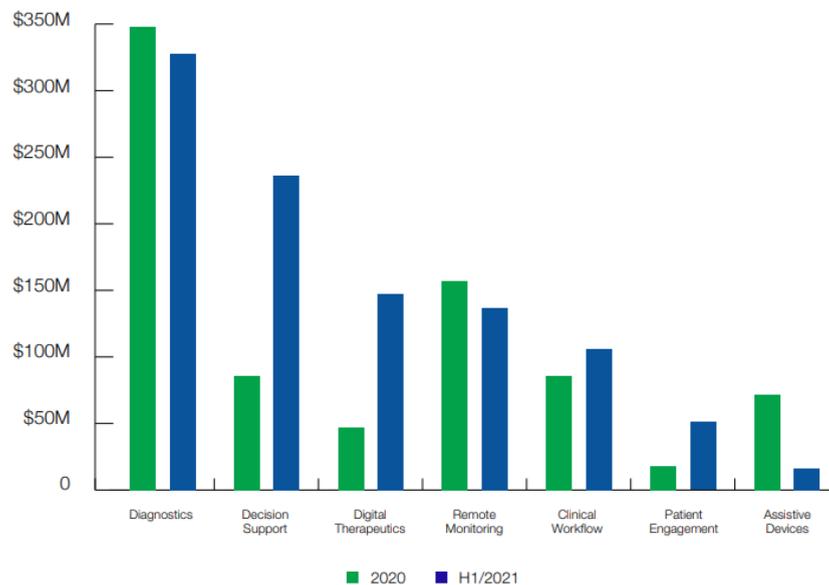


Source : **Digital Health : The Israeli Promise**, Ministry of Economy and Industry, State of Israel, March 2020

En 2021, les entreprises israéliennes de santé numérique ont levé un total de plus d'1 Md\$, contre 577 M€ pour les entreprises françaises¹³ Les sous-secteurs médicaux de télémédecine, d'aide à la décision et surtout de diagnostics ont rassemblé 53 % de ce montant.

¹³ <https://frenchhealthcare.fr/fr/la-e-sante-francaise-a-leve-577-millions-deuros-en-2021/>

Figure 8 : Financement par sous-secteurs médicaux (2020 VS 2021)



Source : *Israel's digital health sector*, Start-Up Nation Central, Lena Rogovin, 2021

Les HMO, qui détiennent des bases de données importantes et permanentes, ont mis en place des centres de recherche et d'autres initiatives de recherche visant à développer des solutions de médecine digitale. Clalit et Maccabi ont notamment développé chacun un centre de recherche. Meuhedet et Leumit, les plus petites HMO, sont également en train d'améliorer leurs capacités de recherche et leur accès aux données.

Clalit a en effet mis en place la *Clalit Research Institute* qui vise à transformer les données cliniques extraites des *EMR* en véritables informations décisionnelles. **Le HMO a ainsi investi 287M \$ en 2018** dans l'agrégation et l'anonymisation de ses données. Il est à noter que la société détient plus de 10 ans de données cliniques, phénotypiques, génomiques et sociodémographiques sur ses 4,7 M d'adhérents. Clalit a ainsi permis le développement de nombreuses start-ups telles que *Zebra Medical Vision*, une société d'analyse d'imagerie médicale qui a été rachetée pour 200M\$.

Parallèlement, Maccabi a mis en place Maccabitech, une branche dédiée à la recherche épidémiologie et clinique. L'organisation a fourni ses données à deux start-ups et notamment à la licorne américaine *K Health* qui met à disposition une plateforme de diagnostic clinique et qui est aujourd'hui valorisée à plus de 1,5 Md\$.

3.2.2 La Corée du Sud : le rôle des gros établissements

Le système de santé public en Corée du Sud est le Service National d'Assurance Maladie (NHIS) et est conduit par le ministère de la santé et du bien-être. Il est universel et financé par le gouvernement coréen et couvre plus de 51 millions de personnes (95 % de la population). Le secteur est notamment dominé par cinq hôpitaux, le Big 5 (*Asan Medical Center, Samsung Medical Center, Seoul National University Hospital, Severance Hospital* et *St Mary's Hospital*) qui exploitent des réseaux d'hôpitaux secondaires. **La Corée du Sud est connue pour avoir un système de santé par les meilleurs du monde notamment concernant la qualité des soins de santé¹⁴.**

Quant à sa gestion des données de santé, la Corée a été classée par l'OCDE en 2021 parmi les trois pays ayant déclaré les politiques et pratiques les plus solides en matière de disponibilité, de maturité et d'utilisation des données de santé et de gouvernance nationale. Le système d'analyse à distance des données médicales est géré par le service de révision et d'évaluation de l'assurance maladie (HIRA).

¹⁴ Par exemple, d'après le rapport *Health at a Glance* publié en 2019 par l'OCDE, la Corée du Sud détenait le taux de survie au cancer le plus élevé des pays de l'OCDE



La Corée dispose de données de santé en temps réel qui lui permettent par exemple, de rendre compte de l'état quotidien des ressources médicales et de leur utilisation. La numérisation de la santé en Corée du Sud est en grande partie le résultat de l'innovation menée par l'État. **En effet, en 2021, le budget national de recherche et développement sur la santé numérique était de 850M \$.**

Le pays annonçait en 2021 un taux d'adoption **de plus de 90 % d'EMR dans les établissements médicaux**. Ces EMR contiennent les notes du médecin et autres informations médicales (diagnostic, prescriptions médicales...), les résultats de laboratoire et de radiologie, les résultats des tests pathologiques et les dossiers infirmiers.

Quant à l'utilisation secondaire des données médicales, en 2020¹⁵, **plus de 50 % des hôpitaux tertiaires (300 lits et +) avaient mis en place un entrepôt de données cliniques ou un modèle de données partagées**. Le *Seoul National University Hospital (SNUH)* constitue à ce titre un bon exemple. Ouvert en 1885, le SNUH est le plus vieil hôpital de Corée. Il compte 1700 lits et 2000 patients soignés chaque jour. Le SNUH était aussi le premier hôpital de la région Asie-Pacifique à passer au 100% numérique

L'hôpital fournit ses données à différentes sociétés de santé numérique. Pour citer quelques exemples :

- La société coréenne *JLK inspection* implantée en Corée, au Japon et aux Etats-Unis, spécialisée dans la fabrication de solutions d'imagerie médicale basées sur l'IA. L'hôpital fournit à cette entreprise, des données d'imagerie et des dossiers médicaux. L'entreprise a obtenu 7 certifications CE en tout.
- La société américaine *Synapse*, évaluée à plus de 100 M\$ en 2020, qui vise à améliorer la qualité des soins de santé d'oncologie grâce à l'IA. Ce partenariat cherche à développer la principale plateforme de médecine de précision en Corée du Sud et dans la région Asie-Pacifique : les deux entreprises construisent ensemble le premier réseau de partage de données dans le monde réel de la Corée du Sud pour les principaux hôpitaux d'oncologie de précision de la région. Le SNUH met notamment à disposition les données cliniques, moléculaires et les EMR.

Plus globalement, on trouve actuellement 616 startups¹⁶ dans le domaine de la HealthTech en Corée du Sud. L'entreprise Lunis par exemple, est une start up coréenne qui avait amassé en 2021, plus de 130 M\$ de financement. La société est spécialisée dans le développement de logiciels d'analyse et d'interprétation de radiographies : elle a notamment reçu plusieurs certifications CE.

3.2.3 Le Danemark : un volume de données important, mais peu exploité

Le système de santé danois est basé sur une couverture universelle gratuite et l'égalité d'accès au système de santé. Le système de santé danois est délégué aux 5 différentes régions.

Le portail danois de la santé en ligne *Sundhed.dk* est la plate-forme nationale en matière de soins de santé. **Elle contient les données personnelles et générales sur les soins de santé de tous les Danois**. Les Danois ont accès en temps réel, par exemple, à leur dossier médical, au renouvellement de leurs ordonnances ou à leurs résultats de laboratoire. De plus, la plateforme propose des consultations vidéo en ligne avec des professionnels de santé.

En termes de gestion des données de santé, l'OCDE classait le Danemark comme le « **pays faisant état des politiques et pratiques nationales les plus solides en matière de disponibilité, de maturité et d'utilisation des données sur la santé et de gouvernance des ensembles de données sur la santé** » en 2021. En effet, Le système danois de services de santé publique dispose de nombreux registres et bases de données administratifs et médicaux complets qui sont collectés depuis plus d'un demi-siècle : des informations sur les causes de décès ont notamment été recueillies depuis 1875 et l'incidence du cancer a été enregistrée pour l'ensemble du pays depuis 1943. Le registre national danois des patients conserve des dossiers qui datent de 30 à 40 ans, ce qui en fait l'un des plus anciens registres hospitaliers nationaux au monde. Les données de santé danoise recouvrent plus de 5,5 millions de personnes soit presque l'entièreté de la population.

Plusieurs projets ont été lancés par l'Etat danois concernant ses données de santé :

¹⁵ Kyehwa Lee, Libga Seo, Dukyong Yoon, Kwangmo Yang, Jae-Eun Yi, Yoomi Kim, Jae-Ho Lee, 2022, "Digital Health Profile of South Korea : A Cross Sectional Study", *International Journal of Environmental Research and Public Health*

¹⁶ HealthTech Startups in South Korea | Tracxn



- Le **Health Tech Hub Copenhagen**, une organisation qui regroupe les différents acteurs des secteurs de santé pour permettre l'adoption de la technologie des soins de santé à grande échelle. En 2021, les membres du *Health Tech Hub Copenhagen* avaient levé un demi-milliard de DKK (~67M €).
- Le **Danish National Genome Center**, une agence gouvernementale qui a été créée pour mettre en œuvre une stratégie nationale de médecine personnalisée du gouvernement danois. Sa mission est de développer un diagnostic plus précis, un traitement ciblé et de renforcer la recherche au sein du système de santé danois à travers les données génomiques. D'ici 2024, le Centre national danois du génome et ses collaborateurs comptent séquencer les génomes entiers de 60 000 patients diagnostiqués avec un cancer, des maladies auto-immunes et des maladies rares.

Malgré l'abondance de ces données, aucune mesure n'a été prise concernant l'utilisation secondaire des données de santé, ce qui limite les possibilités pour les entreprises dans le domaine. On peut cependant citer quelques exemples :

- La société **Radiobotics** a réussi à récupérer des données publiques pour développer son algorithme d'IA. Radiobotics développe des algorithmes d'apprentissage automatique afin d'automatiser la lecture des radiographies des os et des articulations et aider à prendre des décisions médicales. La société utilise les données de l'hôpital danois Bispebjerg. La société a obtenu la certification CE pour son algorithme en 2019.
- Une autre société danoise, **Visiopharm**, a cumulé les certifications CE entre 2015 et 2022. Elle en a obtenu 5 grâce au développement de logiciels d'analyse d'images histopathologiques pour le diagnostic et la recherche sur le cancer grâce à l'IA.

3.2.4 Les Etats-Unis : des investissements privés très élevés

Les Etats-Unis, à la différence de la majorité des pays de l'OCDE, ne possèdent pas de système de santé universel. Leur système de santé est fondé sur les assurances privées. Les *National Institutes of Health* (NIH) sont les acteurs qui s'occupent majoritairement de la centralisation des données de santé dans le pays. Les NIH conduisent de nombreux programmes de recherche destinés à améliorer la santé des citoyens américains.

Le rapport de l'OCDE concernant les infrastructures nationales de données de santé rend compte du **décalage entre la puissance technologique du pays et la faible implication de l'Etat américain**¹⁷ dans la mise à disposition de ses données. L'OCDE note notamment une absence de normes de données, une incohérence des formats de données et une inexactitude des données démographiques dans différents ensembles de données ce qui entrave la précision des couplages des bases de données. Par ailleurs, parmi les différents ensembles de données demandés par l'OCDE, les Etats-Unis n'en possédaient que 54 % (contre 92 % pour la Corée et la France). De plus, parmi ces bases, seuls 33 % couvraient 80 % ou plus de la population, ce qui constitue le chiffre le plus faible de l'OCDE.

Pourtant, parallèlement, les Etats-Unis se classent parmi les pays avec l'une des meilleures gouvernances de l'OCDE en termes de données de santé avec notamment un véritable effort dans la mise à disposition de ces données à la recherche.

Les entreprises américaines de santé sont aussi les mieux valorisées au monde. On comptait en mai 2022, 93¹⁸ licornes de HealthTech dans le monde. Parmi celles-ci, 68 % étaient américaines. Les startups de santé numérique basées aux États-Unis ont rapporté près de 30 milliards de dollars en 2021, soit presque le double de l'investissement total de l'année précédente.

Les organismes de soins de santé et les hôpitaux des États-Unis détiennent d'importantes quantités de données sur la santé des patients stockées sous forme de dossiers médicaux électroniques. Grâce à certaines dispositions de la loi sur la portabilité et la responsabilité en matière d'assurance maladie (HIPAA), **les organismes de santé sont désormais en mesure d'exploiter ces données**. Tant qu'ils dépersonnalisent les dossiers, ils peuvent donner ou vendre les données à des partenaires pour la recherche. Ils n'ont pas besoin d'obtenir le consentement des patients pour le faire, ni même de les en informer. Classée meilleur hôpital du monde en 2021, la *Mayo Clinic* a bénéficié de cette disposition pour mettre à disposition ses données de santé. En 1986, elle fonde la *Mayo Clinic*

¹⁷ *Survey results : National health data infrastructure et governance*, OECD Health Working Papers No.127, J.Oderkirk, 2021

¹⁸ *The shifting digital health investment landscape in 2022 | Healthcare Dive*



Venture, fonds d'investissement qui offre des participations financières aux entreprises. L'entreprise aide la *Mayo Clinic* à commercialiser des informations dérivées des données des patients et des recherches de ses médecins.

- **Entre 2018 et 2020, 16 entreprises ont eu accès à des données de patients anonymisées issues de la Mayo Clinic.** Sur cette même période, la *Mayo Clinic Venture* a conclu 321 accords de licence : ces accords concernent la commercialisation de dispositifs médicaux traditionnels mais aussi le développement d'appareils et de produits numériques reposant sur l'IA et l'accès aux données. **Ces accords ont généré 80 M\$/an pour la Mayo Clinic.**
- La *Mayo Clinic* fournit ses données de soins de santé à des sociétés telles que *Nference*, valorisée à plus de 100M \$ qui a créé une plateforme qui vise à compiler et à analyser des données pour accélérer les efforts de découverte de nouveaux médicaments et d'autres traitements. **Plus récemment, la Mayo Clinic a signé un partenariat avec la société K Health, valorisée à plus de 1,4 Md \$,** à qui elle fournit des données sur les traitements de ses patients en échange de quoi, la société fournit l'application à ses patients.

3.2.5 Le Royaume-Uni : une stratégie articulée autour du NHS

Le système de santé du Royaume-Uni est basé sur une couverture universelle gratuite et l'égalité d'accès au système de santé. D'un point de vue de la politique nationale de mise à disposition des données de santé, le NHS a mis en place une stratégie à deux niveaux :

- **Gestion de la gouvernance des données nationales :** le *Health Data Research UK (HDR-UK)* qui est un organisme qui gère la gouvernance et la qualité des données au niveau national. Lancé en 2019, le HDR-UK est l'institut britannique des données de santé avec pour but de fédérer près de 86 organisations dans 56 bureaux répartis sur 32 sites au sein de la *UK Health Data Research Alliance*. Le parcours vers l'accès aux données se fait via l'*Innovation Gateway* mis en place par l'institut, qui offre également des catalogues des métadonnées des structures partenaires ;
- **Fourniture d'accès aux données nationales :** plusieurs banques de données nationales fournissent actuellement un accès à des bases de données aux acteurs de l'écosystème à des fins de recherche.
 - o Créé en 2006, **UK Biobank** est une cohorte de 502 536 participants ayant entre 40 et 69 ans. Les données collectées, depuis 2006, sont issues des questionnaires qui sont adressés aux participants, et aussi les données biologiques et échantillons collectés. UK Biobank octroie l'accès à ces données aux chercheurs des secteurs publics et privés pour la recherche d'intérêt général, après s'être enregistrés sur la base de données de UK Biobank (*Access Management System*) ;
 - o Créé en 2007, Le système **SAIL Databank** est une architecture nationale pour la recherche et l'évaluation en matière d'e-santé en ligne, permettant de mettre à disposition des données personnelles pseudonymisées à des fins de recherche. Les données concernent principalement la population galloise. Les chercheurs peuvent accéder à un large éventail de données collectées régulièrement sur une période pouvant aller jusqu'à 20 ans. Les données SAIL sont utilisées dans de nombreux domaines de recherche, comme par exemple pour évaluer l'impact des politiques politiques sanitaires.
 - o Créé en 2011, le **Clinical Practice Research Datalink (CPRD)** est un service pour la recherche qui fait partie du ministère britannique de la santé. Les données CPRD proviennent d'un réseau britannique de plus de 2 000 cabinets de soins primaires et comprennent 50 millions de patients dont 16 millions sont actuellement des patients actifs enregistrés avec au moins 20 ans de suivi pour 25 % des patients. CPRD octroie l'accès aux données aux chercheurs du secteur public et privé du monde entier afin de soutenir la recherche observationnelle en santé.

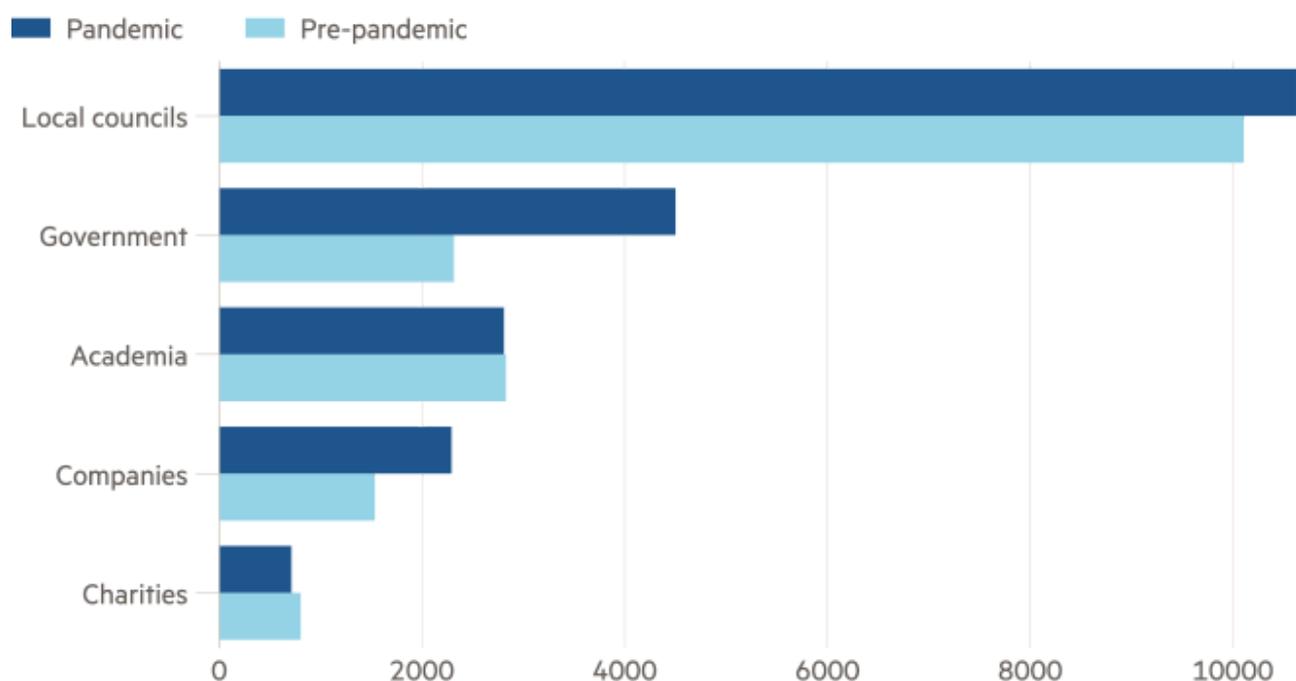
La crise COVID a entraîné une forte accélération de la stratégie nationale de mise à disposition des données. En effet, depuis la crise du COVID, le NHS a considérablement accéléré le partage de données à des fins de recherche pour favoriser et faciliter la découverte d'innovations thérapeutiques à l'échelle nationale. Selon une analyse conduite par le *Financial Times*¹⁹, depuis le COVID, le NHS aurait ainsi partagé plus d'une centaine de bases de données nationales avec l'écosystème, majoritairement auprès du secteur public et aussi auprès de 43 entreprises privées depuis le COVID, parmi lesquelles McKinsey & Company, Novavax ou encore AstraZeneca.

¹⁹ <https://www.ft.com/content/6f9f6f1f-e2d1-4646-b5ec-7d704e45149e>

Figure 9 : Données partagées par le NHS

Most NHS data sharing has been with the public sector

Number of data releases received



Pandemic refers to April 2020 to May 2021. Pre-pandemic is the prior 14 months, February 2019 to March 2020. Sources: NHS Digital, FT analysis

© FT

Lorsque l'on analyse les entreprises britanniques qui exploitent des données de santé pour développer leur activité, on observe un niveau de maturité élevé comparé aux autres pays européens. Par exemple, le Royaume-Uni compte plusieurs entreprises avec de forte valorisation qui exploitent la donnée de santé de façon secondaire :

- **Benevolent AI**²⁰ combine l'IA et l'apprentissage automatique avec l'expertise humaine pour découvrir de nouveaux médicaments. Il fournit une plate-forme de R&D computationnelle qui couvre chaque étape du processus de découverte de médicaments, alimentant un pipeline interne de plus de 25 programmes de médicaments, de la découverte précoce aux phases cliniques. BenevolentAI a également découvert un candidat médicament réutilisé pour covid-19. L'entreprise créée en 2013 est valorisée à plus de 1 Md€ ;
- **Babylon Health**²¹ qui propose un service de santé mobile entièrement intégré offrant un accès aux médecins. Avec l'application Babylon, l'utilisateur peut parler à un médecin généraliste par téléphone ou par appel vidéo, poser des questions médicales via un service de texte et surveiller la santé avec un système de suivi. L'entreprise a été créée en 2013 et sa valorisation actuelle sur le marché s'élève à plus de 4 Md€ (estimation Oct. 2021) ;

4 PAR RAPPORT À DES PAYS COMPARABLES, LA FRANCE PRÉSENTE UN RETARD DANS L'UTILISATION SECONDAIRE DES DONNÉES DE SANTÉ

Les travaux de l'OCDE indiquent que la France se situe plutôt à un haut niveau sur les questions de gouvernance, et à un niveau plus moyen sur les questions de maturité et d'usage secondaire des données. Cela s'explique par :

²⁰ Introduit en bourse, plus considéré comme une licorne

²¹ Introduit en bourse, plus considéré comme une licorne



- **L'absence de bases de données sur certaines aires thérapeutiques** (par exemple le cancer, et plus encore le diabète ou les maladies cardio-vasculaires).
- **Le manque d'automatisation dans la collecte des données, notamment cliniques.** En particulier, J. Oderkik a estimé dans son entretien que la valeur était essentiellement portée par les données cliniques, ce qui devrait conduire à terme à ré-évaluer les indicateurs et le classement (et devrait entraîner une régression de la France, qui fournit essentiellement des données financières).
- **Des facteurs plus structurels** (temps pour collecter les données, difficultés d'identification unique entre les bases, faible utilisation pour évaluer la performance).

Concernant les freins à l'usage de l'intelligence artificielle en santé en France, AI for Health²² s'est appuyé sur un questionnaire adressés à 80 acteurs (hospitaliers, pharmacie, biotech etc.) et sur une vingtaine d'entretiens pour identifier les principaux points bloquants. Il en ressort que :

- La disponibilité et la qualité des données sont perçus comme le premier obstacle par les répondants, suivi par le manque de budget (notamment dans les institutions publiques comme les établissements de santé et les centres de recherche). L'enjeu de disponibilité / qualité est perçu comme prioritaire dans tous les secteurs.

Concernant les cas d'usage à fort potentiel, la situation varie en fonction des secteurs :

- Les pharmacies / technologies médicales identifient la découverte de médicament ou le développement de dispositif médicaux (72 %), devant l'épidémiologie (57 %). Les enjeux d'accès au marché ou de production sont moins cités (resp. 35 % et 32 %)
- Pour l'hôpital, l'organisation des soins est citée comme ayant la plus grande valeur potentielle (77 %), devant la chirurgie et les urgences (50 % et 41 % respectivement).

Ces analyses sont confirmées par les travaux de *l'Open Data Institute*, qui a conduit une étude qualitative qui vise à qualifier et caractériser la situation générale des pays européens en termes d'usage secondaire de la donnée de santé. Le rapport se fonde sur une analyse des écosystèmes pays pour classer ces derniers selon plusieurs dimensions (infrastructure, capacités, éthique, innovation, équité, engagement) afin de mesurer leur niveau de maturité générale sur la question.

Dans l'évaluation, l'ODI point du doigt plusieurs freins à cet usage dans le cas français :

- **Enjeux d'infrastructure** : l'ODI montre dans son étude que d'un point de vue des infrastructures actuellement mises en place, la France, sur certains points, connaît un retard important, par exemple l'absence de cadre de collecte et de développement de PROMs ou encore d'open data / open science ;
- **Enjeux de *capability*** : absence de cadre d'évaluation des technologies de santé en France ; parmi les leaders en Europe, la France ne dispose pas aujourd'hui d'un cadre d'évaluation propre des technologies développés dans l'écosystème santé (dimension *Capability* reportée dans l'étude)
- **Enjeux d'innovation** : manque de reconnaissance de l'usage secondaire des données de santé dans les politiques gouvernementales ; l'étude indique à la fois qu'en termes d'investissement dans les dossiers patient électroniques et la priorisation donnée à l'usage secondaire de la donnée de santé, la France souffre d'un certain retard ;
- **Enjeux éthiques** : (i) selon l'étude, les cas d'exploitation secondaire des données ne permettent pas suffisamment aujourd'hui d'atteindre une forme d'équité dans le bénéfice dont pourraient profiter les patients (en particulier certains sous-groupes), (ii) la loi Bioéthique passée en France, si elle donne un cadre légal visant à assurer le fait que les gens rendent des comptes dans l'usage des données, est davantage orientée pour faciliter l'exploitation ayant une visée de services de télémédecine (IA) que de recherche scientifique pure.

Ce retard se retrouve à la fois dans les chiffres globaux, et dans le retour des entretiens :

- Comme indiqué lors de l'analyse des startups, la France ne présente pas d'entreprises reposant sur l'utilisation des données secondaires avec des valorisations importantes (plus de 50 M€)
- De plus en plus d'industriels utilisent des données à l'étranger en particulier aux Etats-Unis et au Canada.

²² AI for Health Book, 2021