

# Panorama international de l'open data en santé : principaux enseignements

L'objectif de ce panorama des initiatives internationales en Open Data dans le domaine de la santé est d'éclairer la Commission pluridisciplinaire sur le type de données mises à disposition du public dans d'autres pays, les modalités et le contexte de leur ouverture et de souligner des cas de réutilisation remarquables de ces données.

Après avoir revu une quinzaine d'initiatives nationales sur l'Open Data en santé, nous avons sélectionné cinq pays illustrant la variété des démarches engagées : le Royaume-Uni, les Etats-Unis, le Canada, le Danemark et Singapour. La diversité des approches observées relève entre autres du type de données mis à disposition, des possibilités offertes pour réutiliser et visualiser les données, et du type d'acteurs associés (publics ou privés). Les différents visages de l'Open Data en santé s'expliquent en partie par la structuration locale des systèmes et des données de santé ; le cadre juridique, politique et l'histoire du mouvement Open Data dans chaque pays semble également jouer un rôle différenciant non négligeable.

## **I - Trois caractéristiques communes aux pays étudiés favorisent l'émergence de mouvements Open Data nationaux**

Les pays étudiés ont en commun trois caractéristiques ayant contribué à l'ampleur du mouvement Open Data en matière de santé.

### ■ Une **volonté politique forte**

Dans les cinq pays étudiés, une initiative Open Data gouvernementale a permis de porter le débat sur l'ouverture des données au niveau national, d'accompagner sa mise en œuvre et d'encourager l'utilisation des données.

Au Canada et au Danemark, l'engagement gouvernemental pour l'Open Data (hors santé) est relativement ancien : le gouvernement canadien a annoncé sa première *Open Data Initiative* dès 1990. Au Danemark, le droit d'accès pour tous aux données administratives est entré dans la loi dès les années 1980 ; en 2009, la *National IT and Telecom Agency*<sup>1</sup> lance le programme *Open Data Innovation Strategy* visant à uniformiser l'accès aux données publiques pour favoriser l'innovation et la croissance économique, rendre l'action publique plus efficace et transparente.

Aux Etats-Unis, si le gouvernement fédéral a dès 1966 passé une loi sur l'ouverture des données fédérales au public (*Freedom of Information Act*), le champ des données de santé n'a véritablement été ouvert qu'à partir de 2009 avec le soutien de l'*Open Government Initiative* puis de l'*Affordable Care Act* de 2010, autorisant l'ouverture de nouvelles données

<sup>1</sup> Agence rattachée au Ministère des Sciences, des Technologies et de l'Innovation.

par le Ministère de la Santé américain. Le gouvernement propose maintenant un site dédié à l'Open Data en santé, *healthdata.gov*.

#### ■ **L'existence de bases de données structurées**

L'organisation des systèmes de collecte, de numérisation et d'agrégation des données de santé a un impact sur la couverture, la profondeur et les types de données ouvertes.

L'existence d'institutions publiques chargées de centraliser des données de santé (appelés ci-après « agrégateurs publics »<sup>2</sup>) a favorisé la constitution de bases de données au périmètre large, disposant, pour les plus anciennes institutions, d'une profondeur<sup>3</sup> importante. Par exemple, le Danemark a constitué dès la fin des années 1970 un registre national des patients, où les médecins devaient documenter chaque visite médicale. Ce registre a permis la constitution d'une base de donnée très complète et appariale à d'autres bases grâce au système d'identifiant personnel unique des patients. Le risque de ré-identification associé à ce système est pris en compte au niveau légal.<sup>4</sup> Les chercheurs ont la possibilité d'accéder à des données détaillées (« niveau granulaire »), et le gouvernement met à disposition du grand public des données agrégées.

A l'inverse, aux Etats-Unis, il existe une multitude d'agrégateurs publics et privés, avec des possibilités limitées pour appairer les bases de données : les données sont ainsi plus fragmentées, limitant la couverture des données mises à disposition du grand public.

#### ■ **Un cadre réglementaire**

Sauf à Singapour, les pays étudiés disposent tous de textes de loi relatifs au droit à l'information. Ces textes, selon des modalités variant en fonction des pays, stipulent que les données publiques doivent être mises à disposition des citoyens sur demande. Au Danemark par exemple, la loi de 1985 sur les documents administratifs définit les modalités de requête du citoyen aux données publiques administratives et fixe les délais obligatoires de réponses des administrations.

En outre, certains gouvernements ont adopté des lois, règlements ou circulaires obligeant les administrations et opérateurs publics à ouvrir leurs données. A titre illustratif, au Canada, l'*Access to Information Act* de 2010 demande aux entités publiques régionales de publier la liste des données dont ils disposent, et d'ouvrir les données demandées en format exploitable.<sup>5</sup> De même, dans le domaine de la santé aux Etats-Unis, l'enrichissement du *Health Insurance Portability and Accountability Act* dans les années 1990 et 2000 a favorisé la mise à disposition des bases de données des programmes *Medicare* et *Medicaid*. Cette loi

<sup>2</sup> Parmi les agrégateurs publics, on peut citer *NHS England* au Royaume-Uni, la *Danish Health and Medicine Authority* au Danemark ou la *Food and Drug Administration* aux Etats-Unis.

<sup>3</sup> Le terme « profondeur » fait référence ici au nombre d'années sur lesquelles les données sont disponibles.

<sup>4</sup> L'accord préalable des patients est légalement requis avant toute récolte ou communication de leurs données de santé

<sup>5</sup> Le terme « exploitable » désigne ici des formats de fichiers facilitant la réutilisation automatisée des données. Par exemple : fichiers Excel ou « .csv » plutôt que « .pdf »

facilite également l'utilisation qui est faite des données en généralisant l'identifiant unique NPI que l'on trouve dans ces bases.

Autre aspect du cadre réglementaire, les mentions légales des sites internet jouant le rôle de portails d'accès aux données de santé ont également évolué. Dans de nombreux cas, la création de sites et plateformes officiels d'Open Data ont marqué une avancée significative, puisque les sites internet des agrégateurs de données de santé portent généralement des conditions générales d'utilisation restrictives. Par exemple, le site du *Canadian Institute of Health Information*, qui centralise l'accès à toutes les bases de données nationales canadiennes, interdit explicitement l'utilisation des données à des fins commerciales, ce qui n'est pas le cas de la plateforme officielle d'Open Data *data.gc.ca*.

## **II – Dans le domaine de la santé, on observe trois approches différentes de l'Open Data**

L'étude a permis d'identifier trois principales approches de l'Open Data en santé se traduisant par des choix différents en matière de types de données mises à disposition, de modalités d'accès et de réutilisation des données et de services proposés.

- **Le Danemark et le Canada mettent l'accent sur l'information du patient sur le système de soins**

Ces pays mettent à disposition des données présentant une large couverture, une bonne profondeur et dans des volumes conséquents sur la santé publique et l'offre de soins. Grâce à l'existence de bases patients nationales, ces pays sont également les seuls ayant ouvert des données relatives à la consommation de soins – certes à un niveau agrégé, mais avec de multiples critères d'entrée. Par exemple, au Danemark, il est possible de consulter le *National Patient Register* contenant les données relatives à toutes les interactions de chaque patient avec le système de santé dans le secteur hospitalier en choisissant des critères d'agrégation – par exemple l'âge, le sexe, l'année – puis de télécharger les tableaux ainsi obtenus.

Ces pays ont également beaucoup investi pour faciliter la visualisation des données ouvertes par les citoyens. Au Canada notamment, le site *ourhealthsystem.ca* utilise les données des bases du *Canadian Institute of Health Information* pour présenter au grand public les chiffres clés sur le système de santé canadien selon 5 axes : l'accessibilité, la qualité de prise en charge, les dépenses, la prévention des maladies et l'évolution de l'état de santé de la population. Les différents modules informatifs de ce site sont didactiques et intuitifs afin de faciliter l'accès et la compréhension des données par tous.

En revanche, les possibilités de réutilisation de ces données sont contraintes par les mentions légales des principaux portails d'accès. C'est le cas au Danemark où les conditions d'utilisation du site de l'organisme qui rassemble la majorité des données de santé, le *Serum Statens Institute*, limitent l'utilisation à l'usage personnel.

- **Le Royaume-Uni et les Etats-Unis se focalisent plus sur l'autonomisation du patient tout au long de son parcours de soins**

La particularité de ces pays est de mettre à disposition du grand public des données nombreuses et détaillées sur la performance et la qualité des établissements et professionnels de santé. Au Royaume-Uni, des indicateurs de performance tels que les taux de mortalité post-hospitalière sont disponibles par établissement. Aux Etats-Unis, la qualité de la prise en charge dans les hôpitaux fait l'objet d'enquêtes auprès des patients, dont les résultats sont disponibles par spécialité pour chaque établissement.

En revanche, dans ces pays, peu de données de consommation de soins sont disponibles du fait de la couverture limitée des bases publiques de données patients. De plus, l'accès aux données disponibles est souvent payant.

Les services caractéristiques développés dans ces pays sont des outils permettant aux citoyens de choisir des professionnels et établissements de santé. Par exemple, au Royaume-Uni, le site de la *NHS Choices* aide le patient à s'orienter dans ses recherches d'informations médicales et permet même la prise de rendez-vous en ligne. Les différents modules du site permettent entre autres de chercher des informations sur l'un des 500 médicaments ou l'une des 700 pathologies répertoriés à partir de mots-clés ou de symptômes ressentis mais aussi de rechercher un praticien et d'obtenir toutes les informations pratiques le concernant.

- Enfin, Singapour se singularise par un **accompagnement plus poussé dans la réutilisation des données**

A Singapour, un volume important de données de santé au niveau individuel est collecté mais leur ouverture au grand public est limitée à quelques jeux de données agrégées.<sup>6</sup> Seuls les praticiens peuvent ponctuellement accéder aux données détaillées, dans le cadre du traitement d'un patient. Par ailleurs, lorsque des données détaillées (hors données patient) sont accessibles au grand public, ce n'est qu'en consultation en ligne et toute réutilisation est interdite. Par exemple, le *Singapore Medical Council* propose sur son site des données sur les professionnels et les établissements de santé ainsi que les services qu'ils proposent mais leur extraction dans des formats de données exploitables n'est pas proposée. De plus, ces informations ne sont pas réutilisables au regard des conditions d'utilisation du site.

L'Etat encourage activement la réutilisation des données entièrement ouvertes : le site officiel Open Data du gouvernement accueille les applications développées à partir de ces données. Par exemple la plateforme *OneMap* adossée au site officiel est ouverte aux développeurs qui souhaitent y intégrer des fonctionnalités nouvelles ou des visualisations géo-localisées des jeux de données. D'autre part, l'organisme public *Up Singapore* anime des concours, notamment le *Health Up Challenge* sur la santé, qui réunit des développeurs dans le but de créer en quelques jours des prototypes de cas d'utilisation des données.

<sup>6</sup> La base patients constituée depuis 2011 via le *National Electronical Health Record* couvre toute la population et collecte toutes les informations sur les interactions entre un patient et le système de santé mais elle est entièrement fermée au grand public.

### III – Illustration des services développés à l'étranger à partir de l'Open Data en santé

Cette section présente quelques unes des données de santé ouvertes dans les pays étudiés, ainsi que des services développés à partir de ces données.

Le **Danemark** a ouvert le registre complet des professionnels de santé, soit toutes les personnes habilitées à exercer une profession dans le domaine de la santé et nécessitant l'obtention d'un diplôme médical. Ce registre, l'*Authorization Register*, comprend notamment les nom, prénom, date de naissance, lieu de travail, profession et date d'obtention du diplôme de tous les professionnels de santé. Cette base de données a permis l'élaboration d'un outil de recherche sur le site *sundhed.dk* permettant de trouver les données relatives à un professionnel de santé ainsi que les établissements dans lesquels il exerce.

Au **Canada**, deux jeux de données principaux ont permis le développement de services d'information au patient particulièrement utiles dans le contexte canadien. Le premier contient les durées moyennes d'hospitalisation et les tarifs pratiqués dans les hôpitaux en fonction de la région, de l'âge du patient et du type d'intervention. L'ouverture de ces données a permis le développement d'un outil, le *PatientCost Estimator*, qui permet à quiconque d'estimer le coût et la durée d'hospitalisation moyenne liés à une intervention en fonction de ces critères. Le deuxième jeu de données contient les temps d'attente observés dans les hôpitaux pour 8 types d'interventions principales. Ces données ont servi de base au développement d'un module du site *ourhealthsystem.ca* qui permet d'estimer le délai moyen de prise en charge par type d'intervention et par région et de les confronter aux objectifs nationaux. Ce service est d'autant plus utile aux patients qu'au Canada il est nécessaire d'être inscrit sur la liste d'attente d'un hôpital avant d'y être pris en charge.

Aux **Etats-Unis**, l'outil *Hospital Compare* offre un service équivalent en combinant les données ouvertes des enquêtes de satisfaction des patients et des indicateurs de performance des hôpitaux. Les citoyens peuvent ainsi comparer des hôpitaux sur plusieurs dizaines de critères. Le patient peut ensuite télécharger les données et choisir son hôpital sur des critères comme par exemple le temps moyen pendant lequel les patients ont été sous l'emprise de la douleur suite à un AVC dans un hôpital donné. Ce service prend toute son importance dans un pays où le patient choisit lui-même l'établissement dans lequel il sera hospitalisé.

Au **Royaume-Uni**, le site *NHS Choices* propose une note de satisfaction des patients pour chaque cabinet médical du pays. Ces données aident les citoyens à choisir un cabinet médical et permet même la prise de rendez-vous en ligne. Autre exemple, la prescription et la consommation mensuelle de médicaments par molécule et les coûts associés sont également ouverts au public. L'entreprise *Mastodon C* s'est récemment servi de ces données pour démontrer que la prescription de statines génériques permettait une économie d'environ 200 M€ sur le total des dépenses de santé liées aux statines.

Enfin, à **Singapour**, des jeux de données très précis sont mis à disposition concernant l'épidémiologie et notamment le nombre de cas de dengue et leurs localisations par quartier. La dengue est une préoccupation sanitaire majeure dans ce pays et l'anticipation des épidémies de dengue est une priorité afin de mettre en place les efforts de prévention qui s'imposent et limiter les risques de contagion. C'est précisément ce que permet de faire l'application mobile *DengueLah*, à l'échelle du quartier, en se basant sur l'exploitation algorithmique de données d'épidémiologie et de météorologie du site d'Open Data national.

## Conclusion

L'open data dans le domaine de la santé est une démarche récente, y compris au Danemark, au Canada, aux Etats-Unis et au Royaume-Uni, qui se sont lancés avant les autres. La manière dont chaque système de santé est structuré détermine en partie le type de données produites et la place que la puissance publique occupe dans ces flux de données. Ceci résulte dans des approches de l'open data assez contrastées à ce stade, alors même que les valeurs et objectifs affichés par les gouvernements des divers pays sont relativement homogènes : promouvoir la transparence sur l'offre de soin, aider les patients et citoyens à s'orienter, encourager l'innovation et la création de nouveaux services à la fois par les acteurs du système de santé et par le secteur privé.

L'ouverture des données issues du secteur privé n'a pas été évoquée dans cette étude, du fait de la masse relativement limitée de données disponibles à ce jour. Il faut cependant noter que des industriels et assureurs s'intéressent également à l'open data en santé – pour accéder à des informations pour leur développement commercial, mais également pour ouvrir certaines données dont ils disposent. C'est le cas par exemple du géant de l'industrie pharmaceutique Johnson & Johnson qui a récemment annoncé le partage de ses données cliniques avec les chercheurs du *Yale School of Medicine's Open Data Access Project* permettant ainsi l'exploitation de ces gisements de données jusque-là délaissés.

Par ailleurs, la collecte des données de santé n'est plus aujourd'hui uniquement le fait de professionnels de santé mais bien d'une multitude d'acteurs, à l'instar des quantités d'applications pour Smartphones qui relèvent et structurent déjà les données de santé de millions d'individus (comme les bracelets électroniques mesurant l'activité physique et les cycles de sommeil). On peut ainsi envisager que la société civile et les entrepreneurs joueront un rôle plus important dans l'ouverture des données dans un futur proche.