

Les bases de données de l'assurance-maladie : un potentiel pour l'amélioration du système de santé et pour la recherche



Claude Gissot et Dominique Polton

Service des statistiques et de l'évaluation économique
Caisse nationale d'assurance-maladie des travailleurs salariés

Les bases de données de l'assurance-maladie sont rassemblées dans un système national qui ouvre des possibilités remarquables en santé publique. Ce système a en effet des utilisations multiples : il permet d'examiner la dépense de soins selon des critères médicaux, d'analyser des parcours de soins, de connaître l'efficacité de traitements en vie réelle, et peut contribuer à surveiller la sécurité des médicaments. Ce système est de plus en plus utilisé, notamment par les chercheurs, et la Caisse nationale d'assurance-maladie s'est organisée pour faciliter ces utilisations. Un débat a été lancé récemment sur l'intérêt et les enjeux d'une ouverture plus large des données, et les modalités d'accès pourraient être amenées à évoluer.

La création d'un Système National Inter-Régimes d'Assurance Maladie (SNIIRAM) a été prévue par la loi de financement de la sécurité sociale pour 1999. Après plusieurs années de travail technique et l'accord de la CNIL, un entrepôt de données a été constitué à partir de 2003, puis complété et enrichi au fil des années, constituant aujourd'hui une source d'information très riche sur la santé de la population et le fonctionnement du système de soins.

Ce que contient le SNIIRAM

Le SNIIRAM est une base de données qui décrit les soins fournis à chaque individu, examens diagnostiques, interventions médicales, médicaments, soins infirmiers,..., à l'aide des informations contenues dans les feuilles de soins ou les factures des cliniques. Les bénéficiaires des soins sont repérés par un identifiant anonyme pérenne commun avec le PMSI (programme de médicalisation des systèmes d'information) sur les séjours hospitaliers : il est ainsi possible de reconstituer le parcours de soins des patients et la chronologie des événements de santé qu'ils reflètent. Les professionnels et établissements qui délivrent ces soins sont identifiés dans la base de données, qui comporte également des informations concernant la personne bénéficiaire des soins, notamment celles qui sont en rapport avec le remboursement, comme le bénéfice d'une affection de longue durée, ALD (ouvrant droit à l'exonération du ticket modérateur) ou de la CMUC (couverture maladie universelle complémentaire).

On dispose ainsi, pour l'intégralité de la population française, de renseignements sur les pathologies traitées (diagnostics précis en cas d'hospitalisation ou en cas d'ALD) et sur les soins fournis, avec un grand degré de détail (codes des médicaments prescrits, descriptions des actes médicaux réalisés, praticien ayant réalisé l'acte,...). Ces données peuvent donner lieu à de multiples utilisations, que nous illustrerons plus loin par quelques exemples.

La puissance statistique, l'absence de perdus de vue ou de biais d'enregistrement, la précision de la chronologie sont des points forts de cette base de données. Il y a bien sûr aussi des limites pour certaines études, car elles ne comprennent pas de données cliniques (exhaustivité des

diagnostics, données staturales et index de masse corporelle (IMC), niveau tensionnel quantifié,...), paracliniques (résultats d'examen), ni d'informations sur les antécédents ou facteurs de risque (tabac, alcool,...), et peu de données sociales. Néanmoins ces limites peuvent être dépassées, notamment par le biais d'appariement avec d'autres sources. Actuellement un appariement exhaustif est réalisé en routine pour intégrer des données sur les séjours hospitaliers (venant du PMSI) et les dates de décès ; un pilote a été réalisé, avec des résultats très positifs, pour l'appariement avec les causes de décès, et de nombreux appariements ponctuels ou réguliers sont réalisés à la demande d'équipes de recherche avec des enquêtes, cohortes, registres,...

Des utilisations multiples

Les usages que l'on peut faire de ces données sont potentiellement très nombreux. Elles peuvent être utilisées pour l'analyse des pratiques de soins, la régulation du système de santé, la surveillance des effets des médicaments, la recherche épidémiologique,...

Quelques types d'utilisations sont illustrés ci-dessous par des exemples issus des travaux que l'assurance maladie réalise elle-même. Ce ne sont pas les seuls, loin de là, car un nombre croissant d'administrations et de chercheurs se sont investis dans les années récentes pour pouvoir les utiliser pour leurs propres besoins, comme nous le verrons en conclusion.

Cartographier les pathologies et médicaliser l'analyse de la dépense

La CNAMTS a entrepris de développer des algorithmes permettant d'identifier les pathologies ou des facteurs de risque dont souffre la population, à partir des traitements qui sont prodigués : diagnostics posés au cours d'une hospitalisation ou enregistrés par les médecins conseil à l'occasion d'une mise en ALD, médicaments traceurs, etc.

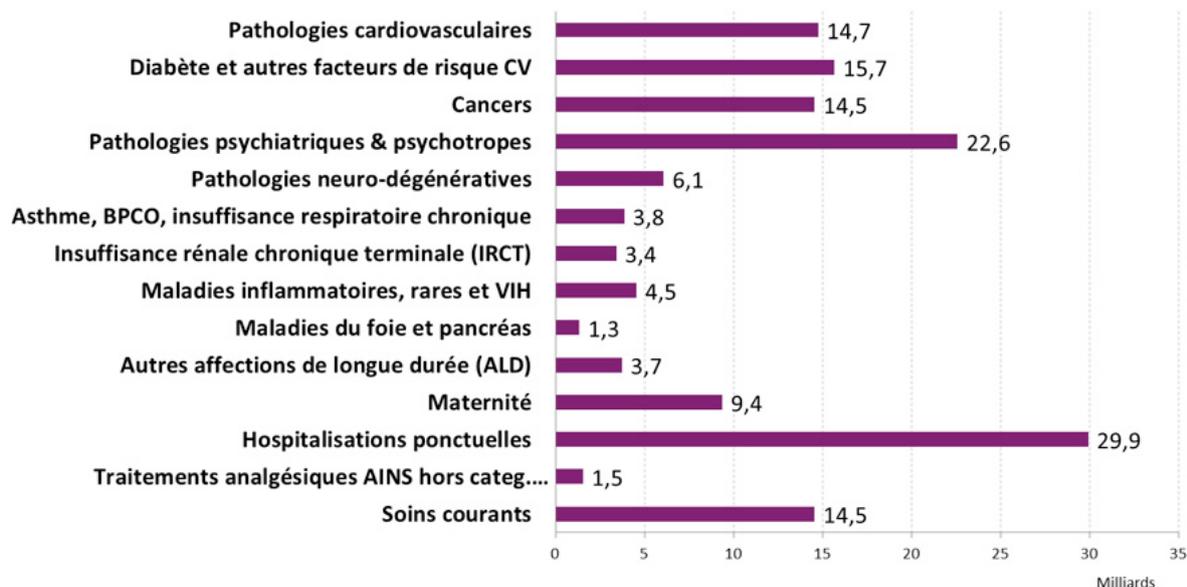
Ces données n'ont pas, bien sûr, une valeur épidémiologique identique à celle d'enquêtes ou de registres qui permettent d'évaluer les prévalences de manière précise : elles ne repèrent, par construction, que les patients pris en charge pour ces pathologies, avec les biais que cela peut entraîner ; elles sont dépendantes de règles administratives qui peuvent évoluer. Mais à l'expérience, dans beaucoup de domaines, la photographie de l'état de santé et les tendances d'évolution qu'elle permet d'établir corroborent les données épidémiologiques disponibles, et elles peuvent être utilisées pour beaucoup d'analyses avec un coût de collecte et de traitement très avantageux.

Pour que ces algorithmes puissent être discutés, améliorés et partagés au sein de la communauté des chercheurs, mais aussi des administrations et agences de santé, plusieurs initiatives ont été prises :

- un réseau informel, plutôt piloté par l'Inserm, rassemble les organismes qui ont vocation à utiliser de tels algorithmes pour leurs besoins propres (équipes de recherche, agences, CNAMTS, Haut conseil de santé publique,...), avec l'objectif de mutualiser les expertises et de mettre en commun les requêtes statistiques élaborées ;
- un travail d'analyse critique du premier jeu d'algorithmes construit par la CNAMTS a été demandé à un département d'information médicale (DIM) d'un centre hospitalier, les conclusions de ce travail pouvant constituer une base pour le travail collectif évoqué ci-dessus.

Cette analyse par pathologie permet déjà, par exemple, une approche nouvelle de la dépense de santé et d'assurance maladie, avec une grille de lecture qui renvoie à sa finalité première – prendre en charge les problèmes de santé de la population – plutôt qu'au raisonnement habituel par catégorie d'offres de soins (soins hospitaliers, soins de ville, médecins, dentistes, infirmiers, médicament...). On peut l'illustrer dans le graphique ci-dessous, qui montre la répartition de l'objectif national des dépenses d'assurance maladie¹ par groupe de pathologie ou traitement.

1. Hors dépenses médico-social et dotations forfaitaires



Graphique 1. Répartition des dépenses de l'assurance-maladie en 2011 (146 milliards d'euros)

Au-delà de cette approche très "macro", il est possible d'analyser les parcours de soins de patients pour une pathologie donnée, leur conformité aux recommandations, leur variabilité sur le territoire. Un exemple, parmi beaucoup d'autres, en est donné ci-dessous.

Analyser les parcours de soins : exemple des thyroïdectomies pour cancer ou nodule bénin

Environ 40 000 thyroïdectomies sont réalisées chaque année en France. Il s'agit de femmes dans 80% des cas, âgées en moyenne de 51 ans.

L'ablation de la thyroïde est en principe réalisée devant la découverte d'un cancer, d'un nodule de la thyroïde pour lequel les résultats de la ponction préalable à l'intervention sont douteux (probabilité de cancer faible mais non nulle), devant certaines formes de goître et d'hyperthyroïdie.

Selon les recommandations de la Haute autorité de santé, l'intervention n'est pas justifiée en cas de nodule bénin. Pour les cancers de très petite taille, l'intervention est très discutée, car elle n'est pas sans conséquences, alors que le cancer de la thyroïde est un cancer le plus souvent peu invasif.

Avant une thyroïdectomie pour nodule, les patients devraient idéalement avoir subi une échographie et un examen biologique (TSH), puis dans la plupart des cas une cyto-ponction (afin de porter l'indication opératoire ou de se livrer à de la simple surveillance) en cas de nodule suspect à l'échographie.

Les recommandations s'accordent sur le fait que la cyto-ponction fait partie des examens qui doivent orienter la décision thérapeutique. Les données du SNIIRAM confirment en effet que la chirurgie pour nodule bénin n'est pas anodine :

- tous les patients ayant une thyroïdectomie totale doivent avoir un traitement par hormones thyroïdiennes à vie, et 44% de ceux ayant une thyroïdectomie partielle en ont également un. Ce traitement nécessite un suivi régulier et peut avoir parfois des répercussions désagréables sur la vie quotidienne en cas de déséquilibre hormonal (troubles de l'humeur, fatigue, frilosité...),

- 4% des patients ont des répercussions sur le fonctionnement d'une corde vocale qui peut nécessiter ensuite des séances d'orthophonie et 1% une atteinte définitive des glandes parathyroïdes qui requiert un traitement par calcium à vie,
- sans compter d'autres complications possibles (cicatrices).

Or l'analyse des données de l'assurance maladie montre qu'aujourd'hui 69% des patient opérés n'ont pas subi de cyto-ponction avant intervention chirurgicale. A l'inverse, les dosages hormonaux sont fréquents, mais pas tous adaptés. La pratique de la cyto-ponction, insuffisante en moyenne, varie par ailleurs fortement d'une région à l'autre. Si globalement en France aujourd'hui, pour 4 cancers on opère 5 nodules bénins, ce ratio est également très différent d'une région à l'autre.

Ce type d'analyse est aujourd'hui reproduit sur de nombreux traitement et épisodes de soins, et permet de mettre au point des programmes d'action pour homogénéiser les pratiques et améliorer le respect des recommandations médicales.

Connaître l'efficacité des traitements en vie réelle : l'exemple des statines²

Les statines sont des molécules utilisées pour faire baisser les taux de certains cholestérols. A l'inverse des autres pays européens où les prescriptions de statines se concentrent sur la simvastatine, en France, la rosuvastatine non génériquée occupe une place importante dans les prescriptions.

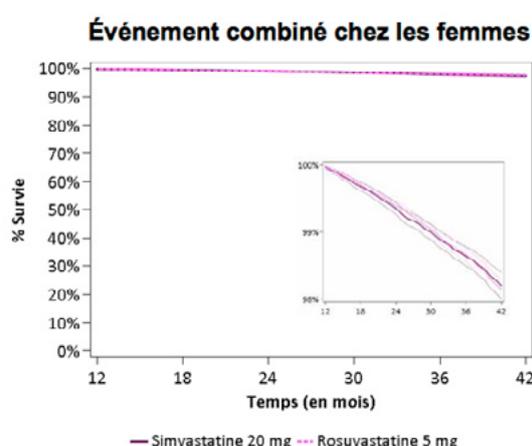
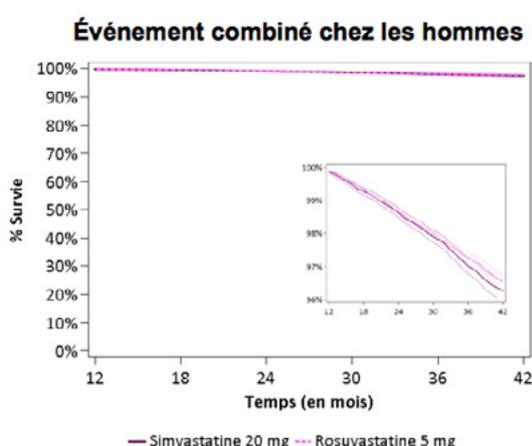
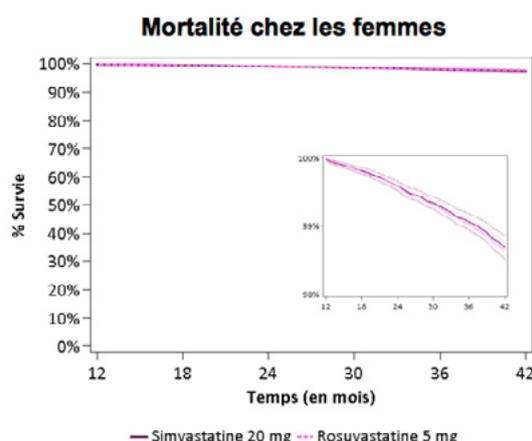
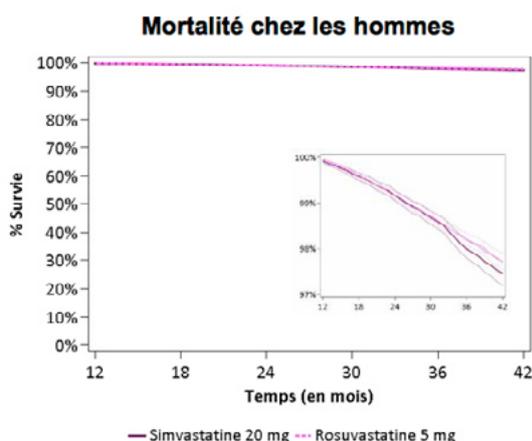
Dans ce contexte, la CNAMTS a souhaité, à partir des données du SNIIRAM, mesurer en population réelle l'efficacité de la rosuvastatine, comparativement à la simvastatine, pour prévenir, chez des patients indemnes de pathologies cardiaques, la survenue d'événements cardiovasculaires graves et de décès ; les dosages étudiés pour les deux composés (rosuvastatine 5 mg et simvastatine 20 mg), ayant montré une capacité comparable à réduire le LDL cholestérol³, marqueur de risque démontré de pathologies cardio- vasculaires.

L'étude a porté sur une cohorte de plus de 160 000 patients dont les deux tiers débutaient un traitement par rosuvastatine 5 mg. Dans aucune des analyses réalisées (ventilées par sexe, événement étudié et type d'analyse), le hazard ratio ajusté⁴ comparant rosuvastatine 5 mg et simvastatine 20 mg n'était significativement différent de un, ce qui indique une absence de différence dans l'efficacité. Les résultats de cette étude en vie réelle n'étaient donc pas en faveur d'une prescription préférentielle de la rosuvastatine par rapport à la simvastatine en prévention primaire.

2. Référence : Neumann A, Maura G, Weill A, Ricordeau P, Alla F, Allemand H. Comparative effectiveness of rosuvastatin versus simvastatin in primary prevention among new users: a cohort study in the French national health insurance database. *Pharmacoepidemiol Drug Saf* 2013. Epub 2013 Dec 2.

3. LDL : « low-density lipoprotein »

4. Le « hazard », ou risque, est la probabilité pour un sujet de développer la maladie à un âge donné, sachant qu'il ne l'a pas présentée jusqu'alors. Cette fonction dépend des facteurs de risque et des traitements éventuels pris. Comparer deux traitements revient à comparer le rapport ou « hazard ratio » de ces risques entre sujets comparables (d'où le terme ajusté) soumis à l'un ou l'autre traitement. C'est un indicateur couramment utilisé en épidémiologie clinique.



Graphique 2. Efficacités comparées de deux médicaments anti-cholestérol : la rosuvastatine 5 mg et la simvastatine 20mg.(instaurations en prévention primaire par un généraliste en 2008-2009 chez les 40-79 ans ; prise régulière pendant la première année). Les graphiques du haut représentent les taux de survie de 12 à 42 mois ; les graphiques du bas représentent les taux de survie sans hospitalisation pour cardiopathie ischémique aiguë ou accident vasculaire cérébral aigu. A gauche : graphiques concernant les hommes ; à droite : graphiques concernant les femmes.

Surveiller la sécurité du médicament : l'exemple de la contraception orale combinée⁵

En décembre 2012, une femme de 25 ans exposée à un contraceptif oral combiné (COC) de troisième génération et victime d'accident vasculaire cérébral avec séquelles portait plainte contre la firme pharmaceutique et les autorités sanitaires. Les medias ont fortement relayé cette information.

Le débat qui a suivi a généré de nombreuses questions de la part des utilisatrices de contraceptifs oraux et des réactions de professionnels de la santé. Le 11 janvier 2013, la ministre des affaires sociales et de la santé annonçait plusieurs mesures visant à limiter, en France, la prescription

5. Référence : « Risque d'embolie pulmonaire, d'accident vasculaire cérébral ischémique et d'infarctus du myocarde chez les femmes sous contraceptif oral combiné en France : une étude de cohorte sur 4 millions de femmes de 15 à 49 ans à partir des données du SNIIRAM et du PMSI ». Rapport final du 26 juin 2013. Accessible en ligne sur Ameli.fr

des pilules de troisième et quatrième génération et demandait qu'une étude rétrospective pharmaco-épidémiologique soit réalisée sur des données françaises pour évaluer la sécurité de ces molécules.

Les résultats de cette étude, conduite par la CNAMTS en partenariat avec l'ANSM⁶, ont été rendus publics quelques mois plus tard, en juin de la même année. Ils étaient similaires à ceux des études observationnelles internationales les plus récentes et les plus puissantes. L'analyse de la cohorte de plus de 4 millions de femmes, résidant en France, et ayant eu des remboursements de COC confirmait ainsi l'existence d'un doublement du risque d'embolie pulmonaire des COC de troisième génération par rapport à ceux de deuxième génération. Le risque d'embolie pulmonaire entre les pilules de deuxième et de troisième génération passait de 25 à 50 pour 100 000 personnes-années.

Mais cette étude a permis également de quantifier les risques, moins connus, associés aux différents dosages d'éthinylestradiol. Au total, il s'avérait que les progestatifs d'ancienne génération comme le lévonorgestrel, combinés à 20 µg d'éthinylestradiol, étaient associés à un moindre risque thromboembolique veineux et artériel ; l'association de 100 µg de lévonogestrel et de 20 µg d'éthinylestradiol étant commercialisée et remboursée en France depuis avril 2010.

L'accès au SNIIRAM

Le SNIIRAM a été d'emblée conçu comme un système ouvert vers des utilisateurs externes : c'est ce qui a guidé les choix d'architecture (portail unique pour tous, utilisation de certains logiciels du marché facilement manipulables par des non experts, etc.). C'est aussi la raison pour laquelle l'accord de la CNIL a été sollicité non seulement sur la constitution de cette base de données, mais aussi dès le début sur son utilisation par des tiers.

De nombreux organismes ont des accès pérennes : Ministère, agences ou autorités sanitaires publiques comme l'ANSM, l'InVS⁷, la HAS⁸ et les ARS⁹, instituts de recherche comme l'INSERM¹⁰ ou l'IRDES¹¹, membres de l'Institut des données de santé comme les unions professionnelles des professions de santé, l'union nationale des organismes d'assurance maladie complémentaire... Ces organismes accèdent aux données à travers une procédure d'habilitation, le périmètre des données accessibles par chaque organisme étant déterminé en fonction de ses missions. Il peut s'agir des bases de données agrégées, de l'échantillon de bénéficiaires au 1/100ème, qui comporte une profondeur historique d'une dizaine d'années, jusqu'à la base de données comportant les données exhaustives détaillées.

Tout autre organisme à but non lucratif de recherche, université, école ou autre structure d'enseignement lié à la recherche a la possibilité, pour effectuer une étude en santé publique, d'accéder pour cette étude aux bases de données agrégées ou à l'EGB¹² après approbation de l'Institut des données de santé (IDS) et autorisation de la Commission nationale de l'informatique et des libertés (CNIL). Au-delà de ces produits préformatés, l'ensemble des organismes de recherche ou d'études à but non lucratif peuvent également solliciter une extraction ad hoc des données exhaustives, qui est réalisée à la demande par la CNAMTS après accord des instances ayant compétence à autoriser ces traitements (CNIS¹³, CCTIRS¹⁴, IDS, CNIL) et sous certaines conditions juridiques et techniques.

6. Agence nationale de sécurité du médicament

7. Institut national de veille sanitaire

8. Haute autorité de santé

9. Agences régionales de santé

10. Institut national de la santé et de la recherche médicale

11. Institut de recherche et documentation en économie de la santé

12. Echantillon généraliste des bénéficiaires (au 1/100^e)

13. Conseil national de l'information statistique

14. Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé

Il n'y a en revanche pas d'accès prévus par les textes pour les organismes à but lucratif. Ouvertes à beaucoup d'acteurs du système de santé et de représentants de la société civile, les données du SNIIRAM sont parfois difficiles à exploiter pour les utilisateurs potentiels de par leur technicité et leur complexité. En effet, ces informations ne sont initialement pas recueillies dans un objectif d'études mais dans le but de verser des prestations aux assurés.

Leur architecture et leur contenu intègrent des contraintes de production et évoluent avec la législation. C'est pourquoi la CNAMTS propose des formations aux organismes habilités à accéder au SNIIRAM et met à disposition des utilisateurs des outils d'aide au traitement des données (dictionnaire des concepts et des données, forum d'échanges, centre de support, journées d'information, clubs utilisateurs, requêtes prédéfinies). A titre d'exemple, la CNAMTS a formé environ 160 chercheurs à l'échantillon généraliste des bénéficiaires (EGB) depuis son ouverture en 2007. En 2013, une cinquantaine de chercheurs a interrogé de manière régulière cette base de données. Dans le cas des extractions ad hoc, une demi-journée d'information est organisée par la CNAMTS pour accompagner les équipes de recherche dans l'utilisation des données du SNIIRAM à des fins d'études et de recherche.

L'accès aux données de santé, et notamment aux données du SNIIRAM, est actuellement en cours d'examen dans le cadre de la commission « Open data » mise en place par la Ministre de la santé. Les modalités d'accès pourraient être modifiées suite aux propositions de cette commission, avec l'objectif de faciliter et simplifier l'accès aux données et de développer encore leur utilisation, en particulier pour la mise en œuvre des politiques de santé.

En conclusion

Le SNIIRAM (système national d'information inter-régimes de l'Assurance maladie) est une base de données médico-administratives qui ouvre des possibilités remarquables en santé publique. D'année en année, cette base de données s'est enrichie et améliorée. Pour donner un exemple, ce n'est qu'en 2010 que le chaînage avec le PMSI a pu être opéré en routine, permettant l'analyse des parcours complets de soins des patients. C'est à partir de cette même année que les dates précises d'hospitalisation (au lieu du mois) ont également été disponibles, ce qui est évidemment essentiel pour analyser l'impact de certains traitements. Ce n'est donc que depuis quelques années que cet outil offre un potentiel d'analyse réellement très riche, qui va encore s'améliorer dans l'avenir.

Le SNIIRAM est de plus en plus utilisé, notamment par les chercheurs, et la CNAMTS a accompagné cette dynamique d'ouverture (formations, club utilisateurs, outils d'aide au traitement, soutien financier à une plate-forme pour la recherche,...). Un débat a été lancé récemment sur l'intérêt et les enjeux d'une ouverture plus large des données, et les modalités d'accès pourraient être amenées à évoluer.