

Le Health Data Hub (suite)

Pourquoi ? Comment ?

Marcel Goldberg, Marie Zins

► Dans le monde de la recherche et de la santé publique, un consensus existe pour considérer que les données de santé constituent une ressource extrêmement précieuse pour de multiples usages, et qu'il convient d'en faciliter l'accès et le partage. Dans ce domaine, la France dispose de nombreux atouts, notamment de bases de données d'ampleur et de richesse sans doute uniques. Depuis quelques années, les pouvoirs publics ont pris conscience des enjeux autour de ces données et ont mis en place un dispositif technique, légal et réglementaire pour y faire face : le Système national des données de santé (SNDS) accompagné de la Plateforme des données de santé (PDS), plus communément appelée *Health Data Hub* (HDH). Cette plateforme est une infrastructure officiellement créée par un arrêté ministériel du 30 novembre 2019, destinée à faciliter l'accès et l'utilisation des données de santé afin de favoriser la recherche. On ne peut qu'applaudir une telle initiative qui constitue un progrès majeur et ouvre de nombreuses perspectives pour la recherche et la santé publique. Cependant, tel qu'il est conçu, le HDH pose divers problèmes qui amènent à questionner ses orientations actuelles. ◀



UFR de médecine,
Université de Paris,
16 avenue Paul-Vaillant-Couturier,
F-94800 Villejuif, France.
marcel.goldberg@inserm.fr

maladie (SNIIRAM), les causes de décès, les données médico-sociales des Maisons départementales des personnes handicapées (MDPH) et un échantillon représentatif des données de remboursement transmises par des organismes d'Assurance maladie complémentaire. Cet ensemble constitue le SNDS « historique ». Mais, depuis, et très discrètement, le périmètre du HDH s'est étendu à des données de nature et d'origine diverses : les données donnant lieu à la prise en charge des frais de santé en matière de maladie ou de maternité, d'accidents du travail et de maladies professionnelles, les données relatives à la perte d'autonomie, celles à caractère personnel des enquêtes dans le domaine de la santé lorsqu'elles sont appariées avec des données du SNIIRAM, les données recueillies lors des visites médicales et de dépistage obligatoires scolaires et universitaires, les données recueillies par les services de protection maternelle et infantile, les données de santé recueillies lors des visites d'information et de prévention de santé au travail.

Il faut noter qu'on trouve dans ce vaste catalogue, les enquêtes recueillant éventuellement des données sensibles auprès des personnes, absentes des sources d'origine administrative comme, par exemple, les orientations et pratiques sexuelles ou la consommation de substances illicites. À plus long terme, toute donnée collectée dans le cadre d'un acte remboursé par l'Assurance maladie (des données des hôpitaux en passant par celles du dossier médical partagé ou encore celles des logiciels professionnels utilisés par les médecins et les pharmaciens) devra être intégrée au HDH. Très récemment, dans le contexte de l'épidémie de Covid-19, un arrêté précise que s'y ajoutent également les informations collectées dans deux fichiers, tous deux

Le Health Data Hub : de quoi s'agit-il ?

Un Décret du 26 décembre 2016 fixe les modalités de mise en œuvre du Système national des données de santé (SNDS). Le *Health Data Hub* (HDH), créé quant à lui par un arrêté ministériel du 30 novembre 2019, est chargé de gérer la mise à disposition des données rassemblées. Dans les premiers textes, les données qui doivent être intégrées au HDH étaient limitées aux données administratives du SNDS qui intégrait le Système national d'information inter-régimes de l'Assurance

Vignette (Photo © Inserm-Michel Depardieu).



actés dans la loi de prolongation de l'état d'urgence sanitaire, qui serviront au traçage du Covid-19 : le *SI-DEP* (système d'Informations de dépistage), qui regroupe des données de laboratoire, et *Contact Covid*, un grand fichier de personnes potentiellement contaminées par le SARS-Cov-2 [1].

Sur le plan de l'infrastructure informatique, alors que le rapport de préfiguration de la plateforme [2] envisageait une structure en réseau avec des nœuds périphériques, le HDH a fait le choix de la construction d'une plateforme unique, centralisant toutes ces données. Il doit donc constituer une gigantesque infrastructure informatique de centralisation de données de santé très diverses provenant de sources multiples qui concernent certains des aspects les plus intimes de la vie privée des personnes. Le HDH a, par ailleurs, fait le choix d'héberger cette infrastructure dans le nuage (ou *cloud*) AZURE, géré par *Microsoft*, ce qui pose des problèmes spécifiques que nous détaillerons ultérieurement.

Une telle accumulation de données à caractère personnel particulièrement sensibles dans un système informatique unique n'a pas, à notre connaissance, d'équivalent au monde à l'échelle de toute la population d'un pays de 67 millions d'habitants. On imagine qu'un tel dispositif réunissant toutes les données de santé de la nation répond à des besoins impérieux pour la santé de la population et pour la recherche, qui justifient de tels risques pour la vie privée des personnes (mais aussi pour la nation) en cas de divulgation, et un coût élevé, et que ces besoins ne peuvent être satisfaits d'une autre façon.

À quoi le HDH doit-il servir ? Les objectifs annoncés

Le site du HDH¹ précise ainsi les enjeux : « Le 31 mars 2018, le Président de la République a lancé un plan 'Intelligence artificielle' pour que la France devienne un leader dans ce domaine. La France a déjà de nombreux atouts : une recherche en médecine et en mathématiques appliquées parmi les meilleures du monde, une base de données médico-administratives exceptionnelle, de nombreuses cohortes, registres et données hospitalières, ainsi qu'un écosystème de start-ups très actif dans ces domaines. Il manquait néanmoins une infrastructure clef, un système de base de données et de services liés : c'est l'enjeu du Health Data Hub. Il devrait ainsi permettre de croiser les bases de données de santé dont nous disposons et de faciliter leurs utilisations par les nombreuses équipes de recherche et de développement ».

Les principales raisons qui sont avancées pour justifier le HDH sont donc de faciliter le développement de la recherche et de l'innovation, notamment les applications d'intelligence artificielle (IA) en santé, et, pour cela, faciliter l'accès et l'utilisation des données de santé.

Les promesses du *big data* en santé et le développement de l'intelligence artificielle en santé

L'ensemble des données de santé existantes constitue le *big data* en santé, dont voici la définition [3]. Le *big data* en santé est caractérisé par quatre dimensions (les « 4 V ») : la volumétrie des données ;

la variété des données, dans la mesure où tous les domaines de la santé sont concernés et génèrent des données de nature très différente ; la véricité des données, qui fait référence à leurs qualités ; et, enfin, la vélocité, car ces données produites en continu ont une croissance exponentielle.

Le *big data* en santé porte de grandes promesses en termes d'innovation et de recherche, mais la collecte, la gestion et l'exploitation de données massives posent de nombreux défis qui ont fait émerger ces dernières années la discipline des Sciences des données en santé². Celle-là n'est autre, en réalité, que le croisement de compétences médicales, informatiques, mathématiques et statistiques. Son objectif est de produire de nouvelles connaissances à partir des données massives en « vie réelle », cela grâce à un ensemble d'outils méthodologiques et techniques pour répondre aux besoins et aux cas d'usages en santé.

À côté des méthodes statistiques traditionnelles, les méthodes d'IA ont fait l'objet d'améliorations spectaculaires dans la période récente, et le nombre de leurs applications dans le domaine médical est exponentiel. Parmi ces méthodes, celles fondées sur l'apprentissage automatique (ou *machine learning*) ont bénéficié des avancées les plus significatives avec, en particulier, les forêts aléatoires et les réseaux de neurones convolutifs [4]. Leurs bénéfices potentiels pour la qualité et l'organisation des soins et pour la recherche en santé sont considérables, tout comme leurs enjeux industriels et économiques. Les méthodes d'IA actuelles s'appuient sur l'utilisation de données massives. Leurs applications améliorent généralement leurs performances avec le nombre de données disponibles pour leur mise au point. Il faut cependant se garder de penser que la taille des bases de données disponibles pour l'apprentissage des algorithmes d'IA est une condition suffisante. De nombreux problèmes méthodologiques et techniques doivent être résolus. En particulier, les méthodes supervisées actuellement les plus performantes nécessitent des données « annotées » par des experts, c'est à dire qu'elles sont classifiées pour leur attribuer un label (un diagnostic, par exemple), pour entraîner un modèle d'apprentissage supervisé, afin que les algorithmes puissent « apprendre » à partir d'exemples. Or, ces tâches d'annotation sont complexes et coûteuses. Elles représentent aujourd'hui un véritable frein à l'essor de l'IA en santé. Les méthodes non-supervisées, qui ne requièrent pas d'annotations, n'en demeurent pas moins sensibles à la qualité des données utilisées. Ainsi, l'idée que l'accès à un jeu de données massives

¹ <https://gnius.esante.gouv.fr/acteurs/fiches-acteur/health-data-hub-hdh>

² <https://www.ncbi.nlm.nih.gov/mesh/?term=Data+science>



intégrant plusieurs sources, sur lequel les *data scientists* pourraient facilement appliquer des algorithmes d'IA pour « faire parler » les données est un fantasme, actuellement partagé par beaucoup de ceux qui ne connaissent pas les données de santé.

L'IA en santé ne pourra se développer à large échelle que si les données de santé respectent certains principes, qui sont énoncés par les critères FAIR (*Findability, Accessibility, Interoperability and Reusability*)³ [5,6]. Ces principes, largement reconnus, sont les suivants : les données doivent être parfaitement décrites par des « métadonnées » normalisées et identifiables automatiquement au travers de catalogues électroniques ; elles doivent être accessibles et interopérables, c'est-à-dire homogènes en termes de signification, de structure et de format, ce qui est rarement le cas pour les données de santé, comme nous le détaillerons plus loin ; le dernier critère FAIR concerne la réutilisabilité des données : il s'agit de s'assurer que les données produites initialement pour une finalité donnée puissent être réutilisables pour de nouveaux objectifs.

Ajoutons que, au-delà des données proprement dites, se pose la comparabilité des populations au sein desquelles elles ont été recueillies. Des algorithmes d'IA, construits par apprentissage à partir de données d'imagerie particulièrement bien harmonisées pour établir des diagnostics, très performants dans le contexte où ils ont été développés, se sont ainsi avérés peu efficaces quand ils ont été utilisés dans des populations aux caractéristiques différentes. Il faut aussi tenir compte de l'évolution des pratiques médicales et de l'utilisation de nouveaux médicaments, de nouvelles procédures d'imagerie, etc. qui peuvent rendre rapidement obsolètes ces algorithmes d'IA. C'est pourquoi, par exemple, les outils d'aide au diagnostic, fondés sur des algorithmes d'IA, font l'objet d'évaluations de long cours, dans le cadre d'une « techno-vigilance » [7,8].

Un accès facilité aux données de santé

Comme nous l'avons rappelé dans notre premier article publié dans le précédent numéro (*m/s* n° 2, février 2020) [15] (→), la loi du 26 janvier 2016 de Modernisation de notre système de santé et le décret du 26 décembre 2016, qui en précise les modalités d'application, ont levé les principaux obstacles législatifs qui limitaient l'accès aux données du SNIIRAM-SNDS. Cependant, lorsqu'il s'agit de données issues du SNDS, les contraintes de sécurité applicables impliquent des moyens informatiques et organisationnels d'un niveau que la plupart des équipes de recherche n'ont pas les moyens de réunir : l'analyse de très vastes ensembles de données par des méthodes très gourmandes en espace et en puissance de calcul requiert également des ressources informatiques considérables. Le HDH a l'ambition de supprimer cet obstacle, en réunissant toutes les données en un dispositif centralisé, en offrant un guichet unique pour les demandes d'accès, et en mettant à disposition des ressources informatiques puissantes de haut niveau.

(→) Voir le Repères de M. Zins *et al.*, *m/s* n°2, février 2020, page 179

Pourquoi le Health Data Hub ne pourra pas atteindre les objectifs annoncés

Les promoteurs du HDH annoncent un progrès majeur dans les possibilités de partage de données de santé. On ne peut que souscrire au principe du partage de données et aux objectifs de facilitation de l'innovation et de développement de l'IA en santé, et se féliciter de la volonté politique d'avancer dans cette direction et de fournir à cette fin des moyens conséquents. L'idée de développer un dispositif consacré à apporter des solutions opérationnelles est incontestablement un progrès décisif, et cette initiative a été très bien accueillie par la plupart des professionnels des données de santé, qui ont été, dans l'ensemble, favorables aux orientations initiales du projet. Cependant, la mise en œuvre de ce vaste chantier s'est progressivement accompagnée d'évolutions qui ne correspondent pas toutes aux attentes suscitées, et il est légitime, aujourd'hui, de s'interroger sur l'adéquation de l'actuel projet du HDH avec les objectifs annoncés. Pour répondre à ces questions, il faut examiner les différentes fonctions du HDH, telles qu'elles ont été annoncées.

Faciliter l'instruction des dossiers de demande d'accès aux données de santé

Les récentes évolutions des textes ont allégé considérablement les contraintes réglementaires pour accéder aux données de santé ou procéder à des appariements. Il reste qu'il n'est pas toujours facile de savoir, selon les situations, quels textes s'appliquent, et de les comprendre, d'autant plus qu'ils évoluent rapidement. Le HDH offre un guichet unique pour les demandes et propose une aide pour la préparation des dossiers, ce qui serait utile pour de nombreux demandeurs. Subsisteront les obstacles liés à la diversité et la longueur de l'instruction des demandes. Bien que les textes prévoient des délais de réponse rapides des organismes concernés (le CESREES [Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé], la CNIL [Commission nationale de l'informatique et des libertés], les CPP [Comités de protection des personnes]), encore faut-il que les dossiers de demande leur soient présentés. Pour les structures publiques de recherche, le véritable goulot d'étranglement se situe bien souvent en amont, au sein des tutelles des équipes demandeuses, du fait de l'insuffisance des moyens pour préparer les dossiers. Il arrive ainsi que des demandes restent en souffrance au sein des organismes, pendant de longues périodes, faute de personnel qualifié en nombre suffisant. Le HDH ne peut évidemment pas apporter d'amélioration à ce problème.

³ <https://www.go-fair.org/fair-principles/>

Réunir, rendre accessibles et exploitables des données de santé de sources diverses

Le progrès majeur, revendiqué par la création du HDH, est le rassemblement en un système unique, de données recueillies et gérées initialement dans des lieux multiples, et selon des méthodes et des procédures hétérogènes, ce qui constitue un obstacle à leur accès et à leur exploitation. L'idée qui a motivé et guidé les promoteurs du HDH est celle d'un immense entrepôt dans lequel les différents utilisateurs pourraient facilement naviguer dans le vaste ensemble de données provenant de plusieurs sources pour appliquer aisément des algorithmes, tirant avantage de la diversité des données ainsi réunies. Alors qu'à l'origine le HDH a été essentiellement conçu pour faciliter l'utilisation du SNIIRAM (le SNDS historique), c'est cet objectif qui a conduit à élargir le périmètre des données constituant le SNDS (donc du HDH) pour constituer une longue liste de sources de données différentes qui s'allonge constamment. Or, cette ambition est en fait un fantasme inatteignable pour diverses raisons.

La première tient à la qualité des différentes bases de données concernées. Construites dans des buts, des circonstances et avec des méthodes qui, pour la plupart, n'ont rien à voir entre elles, leur qualité et leur validité sont extrêmement variables : *big data* n'est pas synonyme de *good data*, et la plupart d'entre elles sont loin d'être gérées selon les principes FAIR (voir ci-dessus). Or les algorithmes d'IA ont besoin de données valides. Avant d'utiliser une base de données, un examen minutieux de ses caractéristiques et de sa qualité, impliquant ceux qui l'ont construite, est indispensable, sans quoi son intégration dans le HDH est inutile.

Une autre raison, majeure, est que les données réunies sont rarement « interopérables ». Du point de vue de l'exploitation des données, la réunion dans une plateforme unique de données de différentes sources n'a d'intérêt que si elles sont interopérables, c'est-à-dire que, d'une façon ou d'une autre, elles sont rendues homogènes en termes de signification, de structure et de format. Cela est nécessaire pour appliquer des programmes d'analyse statistique ou des algorithmes d'IA. L'interopérabilité offre la possibilité d'échanger des données de différentes origines et de réaliser des analyses sur des données mises en commun. Elle a donc une composante sémantique (les données concernées doivent avoir la même signification dans toutes les sources) et une composante technique (les variables correspondantes doivent avoir des structures et des formats compatibles).

Lorsqu'il s'agit d'enrichir par appariement des données de santé proprement dites avec des données d'une autre nature (données sociales, économiques ou d'expositions environnementales par exemple), le problème de l'interopérabilité ne se pose pas, car ces données ne concernent pas les mêmes phénomènes : analyser le risque de telle maladie selon le revenu, par exemple, est une opération courante qui ne pose pas de problème particulier (hormis bien sûr la qualité des données utilisées). Le problème se pose lorsqu'il s'agit d'analyser simultanément des données de sources différentes qui concernent des variables supposées refléter le même phénomène, comme une maladie ou une exposition à un facteur de risque, par exemple. Hormis quelques variables simples, comme l'âge et le genre, ou les données

codifiées selon des nomenclatures officielles, comme la *Classification internationale des maladies*⁴, les données de santé ne sont pas « naturellement » interopérables.

• **L'interopérabilité sémantique** est un non-sens scientifique dans le cas de la plupart des données qui doivent être réunies dans le HDH, pour la simple raison que les données recueillies dans chaque source n'ont généralement pas la même signification. Par exemple, si on s'intéresse à l'insuffisance cardiaque, on peut en effet trouver des données sur cette maladie dans différentes sources que le HDH veut réunir : un dossier de service hospitalier de cardiologie, un code PMSI (programme de médicalisation des systèmes d'information qui produit des résumés de séjours hospitaliers)⁵, un diagnostic de médecin généraliste ou de cardiologue en ville, une prise en charge par l'Assurance maladie au titre des Affections de longue durée (ALD), la déclaration d'un sujet dans une enquête, sur les réseaux sociaux, etc. Mais, selon la source, le terme « insuffisance cardiaque » n'aura pas du tout la même signification : elle peut s'appuyer sur des investigations paracliniques approfondies, sur un examen plus superficiel, sur un code de résumé de séjour hospitalier, voire sur une simple déclaration (dont on sait, dans cet exemple, qu'elle sera souvent erronée). On ne parle donc pas de la même chose dans chaque source. Même quand une variable semble identique entre deux sources, l'expérience montre que, la plupart du temps, ce n'est pas le cas. Par exemple, la plupart des cohortes recueillent des données sur le tabagisme. Mais selon les objectifs scientifiques de chacune, il peut s'agir simplement de la notion de fumeur ou d'ex-fumeur, alors que dans d'autres cas, on s'intéressera au type de tabac, à la quantité, aux modalités de consommation, à la durée, à l'âge de début, etc.

Lorsqu'on veut analyser un ensemble de données issues de deux ou plusieurs sources concernant le même phénomène, la première opération qu'il est nécessaire d'effectuer est l'harmonisation de ces données (c'est-à-dire leur donner un format identique, ce qui, évidemment, n'a de sens que si les données censées représenter le même phénomène ont la même signification). Harmoniser des données signifie les modifier pour les rendre identiques quelle qu'en soit la source, ce qui

⁴ La Classification internationale des maladies (CIM) est la classification médicale permettant le codage en morbi-mortalité proposée et recommandée par l'Organisation mondiale de la santé (OMS).

⁵ Le PMSI est un outil de description et de mesure médico-économique de l'activité hospitalière, introduit en France dans le milieu des années 1980 comme un outil épidémiologique. Obligatoire depuis la loi du 31 juillet 1991 qui oblige les établissements de santé à procéder à l'évaluation et à l'analyse de leur activité, il a été généralisé dans le secteur hospitalier public en 1994 et dans le secteur hospitalier privé en 1996.



nécessite des choix qui sont dictés par les objectifs scientifiques des analyses envisagées. En poursuivant l'exemple du tabac, on voit bien que l'harmonisation des données de plusieurs cohortes aboutira à des résultats très différents, selon qu'il s'agit simplement de prendre en compte le statut tabagique des sujets pour des analyses concernant l'obésité, ou si l'objectif de l'étude est de comparer la nocivité de différents types de tabac. Une longue expérience de participation à des consortiums internationaux de cohortes, dont l'objectif est précisément de réunir des données de chaque cohorte pour des projets et des analyses communs, montre que chaque projet implique un long travail de concertation, de comparaison et de définition de données en vue de leur harmonisation, qui ne peut être réalisé que par les responsables des données concernées, et qui nécessite une expertise et une connaissance approfondie des données, des conditions de leur recueil, des modalités de nettoyage, de validation, etc., ainsi que la définition des objectifs scientifiques communs de leur analyse. Ce travail préalable, essentiel, qui doit être recommencé à chaque nouveau projet d'analyse commune, ne nécessite évidemment pas que les données à harmoniser soient gérées dans un système informatique commun car il ne s'agit pas d'un travail de type informatique, mais de nature scientifique.

De fait, dans le champ de la santé, il existe des normes et des standards internationaux d'interopérabilité. Certains concernent des nomenclatures, comme la terminologie clinique SNOMED CT⁶ ou LOINC, nomenclature d'observations diagnostiques et cliniques⁷. D'autres standards internationaux portent sur l'interopérabilité « syntaxique » qui concerne la façon dont sont codées et formatées les données, comme OMOP-CMD⁸ qui est un modèle relationnel de bases de données de santé, ou des spécifications techniques pour les échanges informatisés de données comme HL7/FHIR⁹, DICOM¹⁰, CDISC¹¹. Des efforts conséquents, portés par l'Agence du numérique en santé (ANS), sont en cours en France pour imposer un socle d'interopérabilité¹² et obtenir un système numérique en santé cohérent ayant un impact direct sur la qualité et la réutilisabilité des données. Mais pour permettre la réutilisation de données d'origines différentes, il faut qu'elles aient été générées à la source en utilisant la même nomenclature (harmonisation *a priori*) et les mêmes formats de données, ce qui habituellement n'est pas le cas pour les données de santé qui sont produites par des professionnels de santé divers dans des contextes souvent très différents. Les données de soins sont aujourd'hui produites par une multitude de logiciels métiers non harmonisés entre eux. Elles sont captives de ces applications, et leur réutilisation suppose leur « dé-silotage »¹³ afin qu'elles puissent être intégrées et harmonisées. Diverses initiatives existent à cette fin, notamment dans le cadre

des entrepôts de données de santé hospitaliers qui se déploient actuellement à grand rythme¹⁴ [9-11]. Une approche également en développement est l'utilisation de méthodes d'IA pour faciliter la réutilisation des données par phénotypage automatique [12,13] : par exemple, des méthodes d'IA de traitement automatique du langage sont utilisées pour interpréter des données de dossiers médicaux hospitaliers dont le contenu est textuel à près de 80 % en moyenne [14], et en extraire automatiquement des informations telles que les antécédents, les traitements, les diagnostics. Ces algorithmes fournissent également des métriques (ou scores) de confiance concernant les informations extraites qui peuvent être ensuite prises en compte dans les analyses statistiques ultérieures¹⁵. Ces standards sont néanmoins encore très peu utilisés dans la pratique courante et les données que le HDH veut centraliser dans son infrastructure informatique sont actuellement extrêmement hétérogènes et non harmonisées. Rappelons qu'il s'agit « des données donnant lieu à la prise en charge des frais de santé en matière de maladie ou de maternité, d'accidents du travail et de maladies professionnelles, les données relatives à la perte d'autonomie, les données à caractère personnel des enquêtes dans le domaine de la santé lorsque ces données sont appariées avec des données du SNIIRAM, les données recueillies lors des visites médicales et de dépistage obligatoires scolaires et universitaires, les données recueillies par les services de protection maternelle et infantile, les données de santé recueillies lors des visites d'information et de prévention de santé au travail ». On peut remarquer que cette liste à la Prévert, qui va bien au-delà de ce qui était prévu à l'origine du projet par le Décret du 26 décembre 2016, a été établie en toute discrétion par une modification en juin 2019 de l'article L. 1461-1 du code de la santé publique.

• **La composante technique de l'interopérabilité** peut, par ailleurs, être complexe et demander un important travail lorsque des données qui concernent le même objet ont été recueillies selon des procédures et des techniques variées : par exemple, des spirométries ou des électrocardiogrammes mesurés avec des appareils différents, l'utilisation d'échelles de dépression non compatibles, ou de nomenclatures qui ne sont pas « alignées » (c'est-à-dire qu'on ne peut pas passer simplement d'un code à un autre) pour coder les maladies (quand elles le sont !).

⁶ Systematized Nomenclature Of Medicine - Clinical Terms

⁷ Logical Observation Identifiers Names & Codes

⁸ Observational Medical Outcomes Partnership - Common Data Model

⁹ Fast Healthcare Interoperability Resources

¹⁰ Digital imaging and communications in medicine

¹¹ Clinical Data Interchange Standards Consortium

¹² Un cadre d'interopérabilité bientôt opposable. <https://www.dsih.fr/article/3887/un-cadre-d-interoperabilite-bientot-opposable.html>

¹³ Le silotage des données fait référence à un système d'information constitué d'outils et de bases très peu connectés entre eux.

¹⁴ AP-HP - Entrepôt de Données de Santé. <https://eds.aphp.fr/>

¹⁵ N2C2: National NLP Clinical Challenges. <https://n2c2.dbmi.hms.harvard.edu/>

