Histoire et éthique

Enjeux éthiques du numérique en santé

Comment prendre en compte la réglementation de la recherche sur les données de santé et évolution vers l'*Open Data* dans le Service de santé des armées

C. VERRET[(1)], R. HAUS-CHEYMOL[(2)], S. MARTIN[(3)], A. CALINET[(4)], J. BOUIX[(5)], A. SGARIOTO[(6)], A. MARCEL[(7)]

(1,2,3,5,6) Académie de Santé des Armées, Paris
(4) Institut de Recherche Biomédicale des Armées, Brétigny-Sur-Orge
(7) Direction Centrale du Service de Santé des Armées, Paris
(5) Université Sorbonne Paris Nord, Laboratoire Éducation et Promotion de la santé, Villetaneuse

Auteur correspondant : catherine.verret.pro@outlook.fr

Open access DOI
10.17184/eac.9362

Reçu
le 27 janvier 2025

Accepté
le 1er avril 2025

Publié
le 11 juin 2025

Conflit d'intérêt
Les auteurs ne déclarent pas de conflit d'intérêt.

Résumé :
Les données personnelles de santé en France, générées par divers acteurs (hôpitaux, chercheurs, patients), sont difficiles d'accès et non standardisées. La pandémie de Covid-19 a mis en lumière l'importance du partage rapide de ces données pour faciliter les activités de recherche dans le monde.
Cependant, ces données sont soumises à une réglementation complexe qui entrave la science ouverte.
Cet article explore les réglementations concernant la recherche sur les données de santé, en particulier dans le contexte du Service de santé des armées (SSA). Les données personnelles de santé sont protégées par le code de la santé publique, le Règlement général sur la protection des données (RGPD) et la loi « informatique et libertés », en plus des règlements européens dans le cadre de recherches sur le médicament ou le dispositif médical. Les données suivent un cycle de vie en plusieurs étapes, de la planification à la réutilisation, et doivent respecter des principes de confidentialité et de sécurité. Les recherches utilisant ces données sont soumises à des règles spécifiques, qui varient selon le type de recherche (interventionnelle ou observationnelle), et doivent garantir la protection des personnes concernées. La science ouverte, qui consiste à rendre les données accessibles et réutilisables, a été encouragée en France, mais l'accès des données de santé doit être réalisé dans le respect de la vie privée, de la réglementation, de la sécurité des données et de la protection du secret. Les données de santé peuvent être partagées à condition d'être anonymisées et sécurisées, tout en garan-

tissant leur qualité. La documentation des données doit être anticipée dès la conception de la recherche, en respectant des principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) afin de permettre une réutilisation efficace.
En conclusion, bien que la science ouverte soit encouragée, la gestion des données de santé reste un domaine spécifique où l'éthique, la réglementation, la sécurité des données et la protection du secret doivent être scrupuleusement respectées.

Mots-clés : Confidentialité. Données personnelles de santé. Gestion des données. Réglementation. Sécurité des données. Médecine d'armée.

Abstract:
DIGITAL HEALTH AND ETHICAL ISSUES. COMBINING RESEARCH REGULATION AND PRE-EXISITING HEALTH DATA WHILE MOVING TOWARDS OPEN DATA WITHIN THE ARMED FORCES HEALTH SERVICE
Personal health data in France, generated by various actors, e.g. hospitals, researchers or patients, are difficult to access and non-standardized. The Covid-19 pandemic has highlighted the importance of rapid sharing of these data to facilitate global research. However, these data are subject to complex regulations that limit open science.
This article explores the regulations concerning research on health data within the French Army Health Service–SSA. Personal health data are protected by the Public Health Code, the General Data Protection Regulation – GDPR, and the French data protection act, in addition to European regulations dealing with drug or medical device research. Data follow a multi-stage life cycle, from planning to reuse.
It is required that confidentiality and security principles be respected. Under the aforementioned legal framework and specific regulations, the various types of research, from interventional to observational, will be conducted while guaranteeing the protection of participants. Open science, which consists in making data accessible and reusable, has been encouraged in France, but health data accessibility requires full respect of privacy and regulations issues. Health data can be shared while respecting anonymized and secured data, while guaranteeing its quality. Data documentation must be anticipated at the start of the research design, respecting FAIR principles – easy to Find, Accessible, Interoperable, Reusable – in order to allow effective use of pre-existing data for research purposes.
In conclusion, although open science is encouraged, health data management remains difficult, where ethics, regulations and security must be scrupulously respected.

Keywords: Confidentiality. Personal health data. Data management. Regulation. Data security. Military medicine.

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Introduction

Les données de santé en France sont produites par de nombreux acteurs (services de soins, organismes de tutelle, structures de recherches, patients, etc.), avec des finalités différentes (données de prise en charge, de remboursement de soins, de recherche, etc.) (1). Ces données sont d'accès et d'utilisation difficiles en raison des multiples systèmes d'information concernés, non standardisés, ni coordonnés (1). Lors de la pandémie de Covid-19, le partage rapide d'informations massives et diversifiées et la collaboration à l'échelle mondiale a favorisé une recherche plus ouverte en santé (1–3). La capitalisation, l'utilisation et la diffusion des données de santé font l'objet d'une stratégie interministérielle récente[1] suite à la publication du règlement européen relatif à l'espace européen des données de santé (4), dont les axes sont de favoriser la transparence et la confiance des citoyens, de constituer des bases de données d'intérêt réutilisables, de réunir les conditions nécessaires au partage et à la mise à disposition des données de santé et de faciliter et simplifier l'utilisation des données.

Cependant, les données de santé sont soumises à un corpus réglementaire complexe, peu favorable à la science ouverte (5).

L'objectif de cet article est de présenter les aspects réglementaires des recherches utilisant les données de santé et d'appréhender le contexte de science ouverte (*Open Science*) appliquée au Service de santé des armées.

1 La recherche sur les données de santé

1.1 Les données personnelles de santé

Les données à caractère personnel concernant la santé sont définies par le Règlement général sur la protection des données (RGPD) comme les données relatives à la santé physique ou mentale, passée, présente ou future d'une personne physique identifiée ou identifiable (6).

Cette définition comprend les informations d'identification directe (nom, prénom) ou indirecte de la personne (numéro de sécurité sociale, adresse, voix, image, etc.), les informations obtenues lors d'un test ou examen et celles relatives à une maladie, un handicap ou un état physiologique. Certaines données de mesures peuvent être considérées comme des données de santé si elles permettent de déduire une information sur l'état de santé de la personne. Trois catégories de données peuvent être considérées :

- les données de santé par nature (antécédents médicaux, maladies, résultats d'examen, etc.) ;
- les données croisées avec d'autres permettent de conclure à un état de santé (mesure de poids et apport calorique, par exemple) ;
- les données utilisées en tant que telles sur le plan médical.

Cette définition exclut les données à partir desquelles aucune conséquence ne peut être tirée sur l'état de santé de la personne (ex. : application collectant un nombre de pas sans croisement de ces données avec d'autres informations).

[1] Stratégie interministérielle pour construire notre patrimoine national des données de santé 2025–2028, Ministère du Travail, de la Santé et des Solidarités, septembre 2024

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Le traitement de données consiste en toute opération appliquée à des données (collecte, enregistrement, organisation, conservation, utilisation, diffusion, rapprochement ou interconnexion, destruction, etc.).

Le responsable de traitement est la personne morale ou physique qui détermine les finalités et les moyens d'un traitement, c'est-à-dire l'objectif et la façon de le réaliser.

Dans le SSA, le responsable de traitement est le Directeur central du SSA (7).
Le représentant du responsable de traitement (Pôle des sécurités de la Direction centrale du SSA) est chargé de la mise en application du RGPD et de la loi informatique et libertés dans le SSA. Le Département promotion et accompagnement des recherches de l'Académie de santé des armées (ACASAN) assure la promotion des recherches impliquant la personne humaine promues par le Directeur central du SSA ; à ce titre, il garantit le respect des réglementations pour les données de santé issues de ces recherches.

1.2 Le cycle de vie des données dans un projet de recherche

Lors d'une recherche, les données vont suivre un cycle de vie qui s'organise le plus souvent en six étapes, allant de la planification du projet à la réutilisation des données (figure 1).

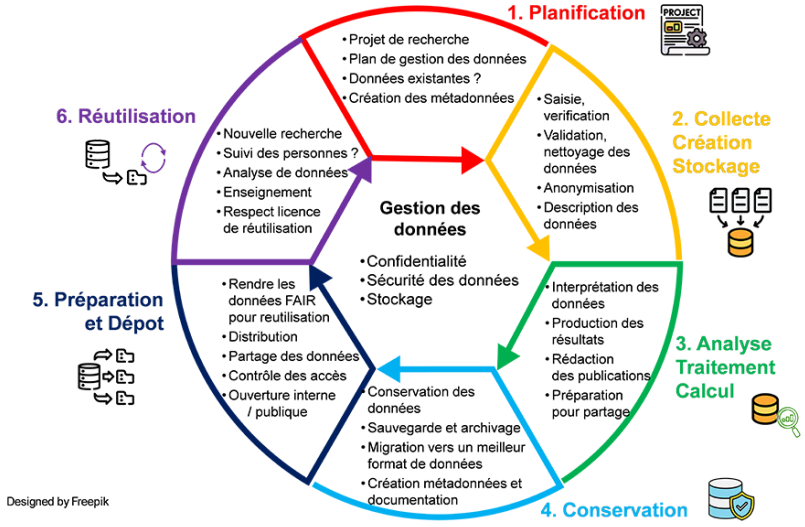


Fig. 1 : Cycle de vie des données adapté de *UK Data Archive*

La planification comprend la conception du projet de recherche et du plan de gestion des données[2] (existantes ou à acquérir), en lien avec les différents acteurs de la recherche. Les démarches réglementaires sont réalisées dès cette étape et avant toute acquisition de données.

[2]Le plan de gestion des données comprend la description des données et la collecte ou la réutilisation des données existantes, la documentation et la qualité des données, les exigences légales et éthiques, la description du traitement et de l'analyse des données, le stockage et la sauvegarde des données pendant le processus de recherche, le partage et les modalités de conservation des données à long terme (exemple : https://www.science-ouverte.cnrs.fr/actualite/plan-gestion-donnees-cnrs/).

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Les données sont ensuite collectées ou acquises, voire anonymisées si nécessaire. Puis, elles sont validées et analysées, selon le plan de gestion des données initialement déterminé.

Une fois le traitement analytique terminé, les données doivent être préparées pour leur conservation : les durées de conservation dépendent du type d'étude. Devraient être conservées les données brutes (i.e. initialement recueillies), les données analysées, les programmes d'analyse de données, ainsi que les métadonnées du projet permettant de décrire la recherche dans le respect de la réglementation (i.e. protocole, note d'information, support de recueil de données, preuve du respect de la réglementation [autorisation, avis, etc.], etc.).

Il est préférable que les données conservées suivent les principes de la science ouverte détaillés ci-dessous (8).

La gestion des données impose, à toutes les étapes de leur cycle de vie, des notions de confidentialité (du projet et des données utilisées), de sécurité des données, d'intégrité et de stockage.

2 La réglementation des recherches utilisant des données de santé

2.1 Un corpus réglementaire complexe, basé sur le risque pour la personne

Par principe, le traitement des données personnelles de santé est interdit, quel que soit le contexte, mais demeure toutefois possible dans certaines conditions dérogatoires[3]. Pour mémoire, le cadre réglementaire s'applique tout le long du cycle de vie de la donnée.
De nombreux textes encadrent la recherche en santé (5). Les études interventionnelles portant sur les médicaments ou les dispositifs médicaux relèvent de règlements européens (9–11). Les autres recherches utilisant des données de santé relèvent du code de la santé publique (en particulier de l'article L.1121-1) et du RGPD (6).
Les Recherches impliquant la personne humaine (RIPH) correspondent aux recherches interventionnelles ou observationnelles « en vue du développement des connaissances biologiques ou médicales », « ayant pour finalité l'efficacité ou la sécurité d'actes, de produits dans un but diagnostique ou de traitement ou de prévention »[4]. Elles nécessitent la participation active des participants (collecte prospective de données non habituelles dans le soin). En fonction du risque, ces études sont soumises à autorisation [Agence européenne du médicament (EMA), Agence nationale de sécurité du médicament et des produits de santé (ANSM)] et avis de Comité de protection des personnes (CPP). Les recherches n'impliquant pas la personne humaine (RNIPH) ne nécessitent pas la participation active des participants (réutilisation de données du soin, évaluation de pratiques des professionnels ou études en sciences humaines et sociales, études de satisfaction, etc.)
Quel que soit le type de recherche, les traitements de données de santé sont soumis au RGPD et à la loi informatique et libertés (6,12) et doivent se conformer à des Méthodologies de références (MR) publiées par la Commission nationale de l'informatique et des libertés ou faire l'objet d'une autorisation de la CNIL.

[3] article 9.2 du RGPD et articles 6 et 44 de la loi informatique et libertés
[4] Articles L.1121-1 et R.1121-1 du code de la santé publique

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Ces traitements de données doivent se justifier en raison d'un motif d'intérêt public (base légale du traitement). Les participants doivent être informés du traitement, de leurs droits (d'accès, de rectification, d'opposition, d'effacement, etc.) et ne pas s'opposer à l'utilisation de leurs données. Les données ainsi recueillies doivent se limiter au strict besoin de la recherche (principe de minimisation) et n'être accessibles qu'aux équipes autorisées avant toute mise en œuvre.
Dans le domaine des recherches en santé, il est très rare qu'une donnée ne puisse être rattachée à une personne (5), car elle provient toujours d'un individu. Les données sont le plus souvent pseudonymisées, en remplaçant les données directement identifiantes (nom, prénoms, numéro de dossier médical, etc.) par des données indirectement identifiantes (alias ou numéro séquentiel relié à une liste nominative). Il s'agit d'une opération réversible (il est possible de retrouver l'identité d'une personne si l'on dispose d'informations supplémentaires).
Les données qui ne permettent plus la réidentification d'une personne (données anonymes), selon les principes d'individualisation, de corrélation et d'inférence, ne sont pas soumises à la règlementation RGPD, de même que les données statistiques ou agrégées.

2.2 Démarches réglementaires des recherches sur données de santé

Le chercheur doit élaborer un projet de recherche dans le domaine de la santé, mentionnant la nécessité d'utiliser des données de santé. Les démarches relatives aux registres et aux entrepôts de données de santé ne sont pas abordées dans cet article (plusieurs projets ou projets non définis). Il doit lister les données nécessaires et suffisantes pour répondre à la question de recherche. Ces données peuvent avoir été déjà collectées (pour le soin ou pour une autre recherche) ; la collecte initiale doit alors avoir été faite dans un cadre légal à documenter sous peine de ne pouvoir les réutiliser dans le cadre d'une recherche. Ces données peuvent être déjà agrégées ou individuelles. Une fois le projet rédigé, la faisabilité est à questionner (données existantes, mais limitées dans le temps ou l'espace, étude prétest ou étude pilote) avant de conduire le projet complet. La sécurisation des données, la conservation de leur intégrité et la gestion de la confidentialité doivent être décrites dans cette étape de conception et être respectées tout au long de la recherche.

Les démarches réglementaires sont guidées par l'objectif de l'étude (finalité), l'origine et le type de données nécessaires à la recherche, la possibilité d'informer les participants et l'impact sur ces derniers (figure 2).

Pour les recherches portant sur le médicament (étude clinique, essai clinique), portant sur un dispositif médical (investigation clinique), un dispositif médical de diagnostic *in vitro* (étude de performance) ou une recherche impliquant la personne humaine, les démarches réglementaires sont réalisées par le chercheur et le Département promotion et accompagnement des recherches de l'ACASAN pour le compte du Directeur central du SSA (promoteur de l'étude) qui couvrent les autorisations (EMA, ANSM) et les avis du CPP en lien avec le pôle des sécurités de la DCSSA pour les démarches RGPD/CNIL (conformité à une MR, ou autorisation CNIL le cas échéant).

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

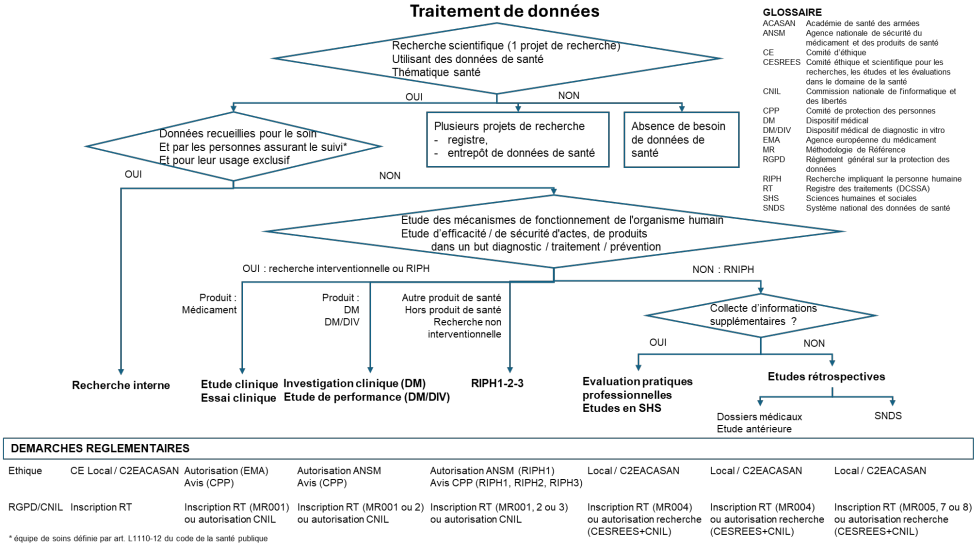


Fig. 2 : Démarches réglementaires des recherches en santé (janvier 2025)

Les autres recherches relèvent du pôle des sécurités de la DCSSA.
Un avis d'un Comité d'éthique de la recherche (CER) est recommandé (local, Comité d'évaluation éthique de l'ACASAN [CE2 ACASAN], comité spécialisé de société savante, etc.). Les RNIPH internes sont réalisées à partir de données recueillies dans le cadre du suivi des patients dans un seul service/hôpital, par les personnels assurant ce suivi et pour leur usage exclusif. Ce type de recherche peut cependant faire l'objet de publication, mais ne peut pas être enregistré par un comité d'éthique (attribution d'un numéro IRB – *institutional review board*).

Les autres RNIPH peuvent être conduites par des investigateurs appartenant ou non à l'équipe de soins, avec intervention de partenaires publics ou privés signataires d'un protocole de collaboration scientifique (forme de convention). Les patients concernés peuvent avoir été pris en charge par un même service ou plusieurs services d'un même établissement. Si les patients ont été pris en charge par plusieurs établissements, sont exclus de l'étude les mêmes patients pris en charge par plusieurs de ces établissements afin de ne pas reconstituer un parcours de soins à leur insu. Les patients s'étant opposés ou s'opposant à la réutilisation de données de santé les concernant sont exclus de l'étude. Une analyse de l'impact sur la vie privée des personnes concernées résulte d'une analyse conjointe avec le pôle des sécurités de la DCSSA, qui transmet la notice d'information pour diffusion aux participants. Une recherche conforme à la MR004 doit être enregistrée auprès de la Plateforme des données de santé (*Health Data Hub – HDH).*

Toute modification de la recherche peut entraîner des démarches réglementaires supplémentaires, voire une requalification de la recherche en RIPH.

À l'issue de la recherche, les données sont réglementairement conservées pour une durée variable selon le type d'étude.

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Si une partie de la recherche a lieu dans un pays étranger, il est nécessaire de prendre en compte également la réglementation en vigueur dans le pays concerné ainsi que le transfert de ces données, partie intégrante de leur traitement.

Le non-respect du cadre réglementaire expose à des sanctions disciplinaires et pénales.

3 Données de santé ouvertes

3.1 Contexte de la Science ouverte

Évoquée dès les années 1970, le concept de science ouverte a été développé dans les dix dernières années et montré ses applications notamment en Santé publique (13).

Les administrations publiques françaises ont l'obligation de publier leurs données d'intérêt général, dans des formats ouverts et réutilisables depuis 2016, permettant ainsi la réutilisation des données publiques (3). Le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation promeut une stratégie de diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique, par la généralisation de l'accès ouvert aux publications, la structuration et l'ouverture des données de la recherche et des codes sources dans le respect du paradigme "aussi ouvert que possible, aussi fermé que nécessaire". Par ailleurs, certaines revues scientifiques demandent aussi l'accès et l'archivage de données (souvent sous forme agrégée) dans leur espace.
De nombreuses ressources à destination des chercheurs (guides, plateformes de données ou de publications ouvertes) ont été mises en place par le Comité pour la science ouverte ou les institutions (Centre national de la recherche scientifique – CNRS, HDH, Agence nationale de recherches sur le sida et les hépatites virales – ANRS, etc.) afin d'encourager l'ouverture de la science[5].
Enfin, le règlement sur l'espace européen des données de santé (4) prévoit notamment, à l'horizon 2030, une obligation de mettre à disposition les données de santé électronique détenues par tout organisme (de soins ou de recherche) pour une utilisation secondaire (sous condition), et une procédure européenne harmonisée encadrant la réutilisation des données de santé, incluant une transparence de réutilisation (publication des résultats, etc.).

3.2 Utilisation des données de santé ouvertes

Pour concevoir une recherche, il est possible d'utiliser des ressources librement accessibles, qu'elles soient en accès ouvert (disponibilité directe de la ressource) ou ouvertes à une collaboration (accès ouverts et réutilisables, en fonction de leur licence de diffusion).
Ces ressources comprennent des mises à disposition de moyens (entrepôts de données, forges logicielles) ou de résultats (plateformes de revues en accès ouvert, de prépublication, bases de thèses ou travaux académiques, etc.).
En France, les sites *data.gouv.fr* et *recherche.data.gouv.fr* sont des plateformes ouvertes et communautaires qui visent à centraliser et structurer les données ouvertes en France, en particulier issues du ministère des Solidarités et de la Santé, la Caisse nationale de l'Assurance Maladie et Santé publique France.

[5] https://www.ouvrirlascience.fr/accueil/, https://www.science-ouverte.cnrs.fr/

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Les données ouvertes peuvent être disponibles sous réserve d'une licence, document qui fixe les droits et obligations du concédant du jeu de données et du réutilisateur du jeu de données (exclusivité ou non, gratuité, durée, etc.) (14). La rediffusion de données initialement diffusées en « open data » constitue un traitement de données à caractère personnel (14). Ainsi, la recherche réutilisant les données devra être licite, poursuivre une finalité déterminée, explicite et légitime. Elle devra porter sur des données pertinentes et proportionnées, mises à jour, effectuées en toute transparence, en respectant les droits des personnes. La conservation de ces données doit être limitée et la sécurité devra être garantie tout au long de ce nouveau traitement de données (14).
Les données peuvent être réutilisées dans une collaboration scientifique avec le détenteur des données (exemple : base de données d'enquête du portail de l'Alliance nationale pour les sciences de la vie et de la santé – AVIESAN).
Pour mémoire, le fait qu'une ressource soit en accès ouvert ne constitue pas une garantie de qualité. Comme toute ressource documentaire, elle doit faire l'objet d'une évaluation critique avant d'être utilisée.
Quelles que soient les formes de réutilisation des données, un projet de recherche et un plan de gestion des données doivent être élaborés, décrivant la sécurité et la confidentialité des données. Une licence ou une convention doit encadrer le transfert des données et les mentions à faire figurer dans les futures publications afférentes (origine des données, remerciements, auteurs, etc.).

3.3 Ouverture des résultats d'une recherche

La publication des résultats d'une recherche est une obligation légale[6]. Un manuscrit peut être publié directement dans une revue de science ouverte ou diffusé six à douze mois après la publication initiale pour des recherches ayant bénéficié majoritairement de fonds publics.

L'archive ouverte HAL (Hyper Article en Ligne)[7] est une archive ouverte multidisciplinaire permettant de partager en libre accès les résultats de recherche, publiés ou non, au service des chercheurs académiques.
La publication dans les revues en *open access* (basé sur la notion d'auteur-payeur) sélectionne des équipes capables de payer des coûts de publication qui peuvent être élevés (2 000 euros en moyenne en 2020, allant de 500 à plus de 6 000 euros pour un article), favorisant ainsi les publications issues des « pays du Nord ».
En 2022, 69 % des articles de revue étaient en accès ouvert, variable selon la thématique (79 % en mathématiques, 75 % en biologie fondamentale) et 22% mentionnaient le partage de leurs données[8]. Seuls 52 % des essais cliniques terminés en 2020 avaient posté ou publié des résultats dans les trois ans après la fin de l'essai (19 % pour les études observationnelles)[7]. Pour le SSA, le recensement exhaustif des publications n'est pas réalisé. Cependant, parmi les articles référencés dans Pubmed en 2022 et dont au moins un auteur servait dans le SSA, 55 % étaient publiés dans une revue en accès ouvert[9]. Par ailleurs, les universités diffusent en accès libre certains travaux universitaires (thèses de médecine, notamment).

[6] Articles L.1121-15 et L.1121-16-1 du code de la santé publique
[7] https://hal.science/
[8] Source : Baromètre de la science ouverte (https://barometredelascienceouverte.esr.gouv.fr/
[9] Interrogation Pubmed avec mots clés : armées et France en Affiliation, pour l'année 2022

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

3.4 Ouverture des données de santé d'une recherche

Dès la conception d'une recherche, le chercheur doit prévoir le devenir des données et des résultats et leur ouverture potentielle.
Dans le domaine de la santé, le principe « aussi ouvert que possible, partage restreint autant que nécessaire » apporte une réponse à l'obligation d'ouverture des données tout en respectant la protection des données personnelles.

3.4.1 Anonymisation et principe FAIR des données

Pour garantir une ouverture sécurisée des données de santé, les données doivent être de qualité, sécurisées et anonymisées.
L'anonymisation doit être adaptée aux usages prévus et reposer sur les trois critères suivants : l'individualisation (est-il toujours possible d'isoler un individu ?), la corrélation (est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?) et l'inférence (peut-on déduire de l'information sur un individu ?). Un ensemble de données pour lequel au moins un des trois critères n'est pas respecté ne pourra être considéré comme anonyme qu'à la suite d'une analyse détaillée des risques de réidentification (14).
En pratique, l'anonymisation doit consister en une suppression des éléments d'identification directe et indirecte (données non structurées, etc.), ou généralisation de certaines données (géolocalisation, classes d'âge, etc.), en ne gardant que les données strictement utiles au regard des finalités en vertu du principe de minimisation des données.
Pour être réutilisables, les protocoles de recherche, les matériaux, le code et les données doivent être organisés de manière à ce que les chercheurs puissent les trouver, y accéder, interagir avec eux et les comprendre (15).
Pour faciliter la découverte et l'accès aux données, leur interopérabilité et leur réutilisation, les données doivent suivre les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable – *Findable, Accessible, Interoperable, Reusable*) (8).
Le centre de ressources DoRANum[10] décrit les étapes suivantes afin de respecter ces principes :

- choisir un entrepôt de données adapté et vérifier qu'il attribue au jeu de données un identifiant pérenne ;
- documenter un maximum les données et utiliser une terminologie adaptée à sa discipline ;
- définir clairement les droits d'accès aux données au moment du partage (licence) ;
- privilégier des formats de fichiers ouverts, non-propriétaires ;
- attribuer une licence de réutilisation à ses jeux de données ;
- mentionner les productions liées aux données (publications, thèses, autres jeux de données, codes et logiciels, etc.).

Des ressources en ligne sont disponibles sur le site de la science ouverte (passeport pour la science ouverte) ou sur la plateforme de formation DoRANum (jeux sérieux pour acquérir des bonnes pratiques de gestion et de partage des données de la recherche *GopenDoRe* et *Faut pas s'en FAIR*).

[10] DoRANum est une plateforme de formation en ligne sur la gestion et le partage des données de la recherche selon les principes FAIR, réalisée par l'Inist-CNRS et le GIS « Réseau Urfist » depuis 2015. https://doranum.fr/enjeux-benefices/le-minimum-a-fair__10__13143__p4wf-4j63/

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

3.4.2 Aspects réglementaires et intégrité des données

Tout au long de sa recherche, le chercheur doit se conformer aux dimensions éthiques de la gestion des données : respecter leur qualité, leur intégrité, la propriété intellectuelle et la vie privée. L'anonymat des données est incontournable. Ce n'est qu'à cette condition que de nouvelles finalités sont envisageables. Pour mémoire, si les données sont anonymes, leur utilisation sort du cadre d'application du RGPD. La réutilisation de données de recherche non anonymes reste possible sans nouvelle information du patient sous réserve de finalités identiques et avant la période de fin de durée de conservation pour les études sous MR tandis que la réutilisation de données de recherche non anonymes pour d'autres finalités nécessite une démarche de mise en conformité avec nouvelle information des patients.
Il est également tenu à une recherche scientifique intègre, en assurant la traçabilité de sa recherche (cahiers de laboratoire, par exemple). Si la diffusion du jeu de données est possible, une licence de diffusion doit lui être attribuée. La licence est choisie par le chercheur et défini ce que le réutilisateur est autorisé à faire avec les données. Le réutilisateur doit respecter l'intégrité des données et doit mentionner la filiation de l'information.

3.5 *Open data* et intégrité scientifique

L'intégrité scientifique contribue à garantir l'impartialité des recherches et l'objectivité de leurs résultats (article L.211-2 du Code de la recherche).
L'accessibilité aux publications scientifiques et à leurs données est considérée comme l'une des conditions garantissant une intégrité scientifique aussi bien des travaux que des résultats de la recherche (16). Cependant, les manuscrits peuvent être publiés rapidement pour des travaux non aboutis, ou avec un contrôle par des pairs très limité, voire dans des revues prédatrices (17). Ce phénomène a été particulièrement visible dans un contexte d'urgence de santé publique (Covid-19). L'accès ouvert ne garantit pas l'absence de problèmes méthodologiques ou éthiques dans les recherches, ainsi que des défauts de qualité ou de présentation incomplète des résultats (18).
L'accessibilité des publications peut supprimer les obstacles à la lecture et à l'utilisation des résultats de la recherche. Il s'agit d'une composante nécessaire de la science ouverte et reproductible, au même titre que le partage des données, du code et des matériaux, mais il ne suffit pas à lui seul à garantir la qualité de la recherche (18).

3.6 SSA et science ouverte

3.6.1 Mise en œuvre de la science ouverte dans le SSA

À la date de publication de cet article, le SSA ne dispose pas d'une zone de dépôt permettant la conservation des données d'une recherche et de ses métadonnées. Dans un premier temps, un catalogue pourrait être mis en place, documentant les recherches menées et dont les données seraient accessibles à l'image du portail Épidémiologie–France (*health databases*) de l'Institut national de la santé et de la recherche médicale (INSERM)[11].

[11] Le portail Épidémiologie–France, sous la direction d'AVIESAN (Alliance nationale pour les sciences de la vie et de la santé), propose un catalogue en ligne des principales bases de données en santé de source française qui peuvent être utiles au développement de la recherche et de l'expertise en santé publique (https://epidemiologie-france.aviesan.fr/)

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Les données de la recherche pourraient ainsi être réutilisées pour de nouvelles finalités, en garantissant le respect de la réglementation. Trois études sont identifiées sur portail Épidémiologie–France décrivant leurs bases de données dans un but de collaboration scientifique. Les auteurs n'ont pas connaissance d'autres jeux de données partagées, issues de recherches dans le SSA.
Comme dans toute administration, les données et les résultats des recherches n'appartiennent pas au chercheur du SSA mais à l'Institution. Le chercheur souhaitant publier une recherche portant sur une thématique sensible, une population spécifique ou qui mentionnerait des résultats pouvant entraîner des conséquences pour les armées doit recueillir l'accord de la DCSSA ou son représentant dès l'initiation de la recherche et avant toute diffusion.
L'ouverture des données d'une recherche menée dans le SSA est soumise à l'accord du Directeur central du SSA (promoteur de la recherche / responsable de traitement). En plus du cadre réglementaire décrit, il est rappelé que le traitement de données identifiant le statut de militaire est restreint aux seules personnes autorisées (article R. 4123-47 du code de la défense).

3.6.2 Avantages et limites de la science ouverte dans le SSA

Les avantages de la science ouverte pour le SSA résident dans l'amélioration de la recherche, en réduisant les coûts. L'accès libre aux données et résultats des recherches est un élément facilitant les collaborations scientifiques civilo-militaires, nationales et internationales, et les partenariats public-privé, en particulier dans les domaines d'expertise du SSA. Les moyens de recherche peuvent être mutualisés entre équipes (accès à des infrastructures et expertises partagées, élargissement des populations d'étude, etc.), permettant une réduction des coûts (en particulier de collecte de l'analyse de données) et une optimisation des financements publics. La diffusion des données et des résultats des recherches rigoureuses renforce la confiance, la transparence et l'adhésion des populations aux politiques de santé.

L'ouverture des résultats et des données des recherches est soumise à des contraintes réglementaires fortes. Ces données ou résultats peuvent concerner des populations sensibles (militaires en activité, anciens combattants, blessés de guerre) et/ou peuvent contenir des informations stratégiques. Leur divulgation non maîtrisée pourrait entraîner des risques pour la sécurité nationale et un risque d'exploitation par des acteurs hostiles, visant à obtenir des renseignements sur l'état de santé des troupes.

Le SSA, en lien avec le ministère, doit adopter une approche mesurée conciliant ouverture scientifique et protection des données sensibles, en tenant compte de la création de l'espace européen des données de santé (4). Une stratégie d'ouverture et un cadre doivent être établis, en développant des plateformes sécurisées de partage des données, accessibles uniquement à des chercheurs autorisés, mettre en place des procédures strictes d'anonymisation, favoriser des collaborations encadrées avec les acteurs civils. Un cadre spécifique doit permettre de définir quelles données peuvent être rendues publiques sans risque pour la sécurité nationale.

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Conclusion

La gestion des données de santé en France est régie par un cadre réglementaire complexe et relativement récent qui vise à protéger les individus tout en permettant leur utilisation dans des recherches scientifiques. Bien que les données de santé soient essentielles pour faire progresser la médecine et la recherche, elles doivent être utilisées avec une grande rigueur pour assurer la protection de la vie privée des sujets, la sécurité des informations personnelles et les intérêts de la défense nationale.

Le contexte de la science ouverte permet une meilleure collaboration nationale et internationale et une plus grande transparence, mais cela ne doit pas se faire au détriment de la confidentialité et de l'intégrité scientifique. Les principes de transparence et de réutilisation des données, en particulier sous des formats ouverts et réutilisables, offrent des opportunités pour améliorer les connaissances médicales tout en respectant les droits des citoyens. Dans ce cadre, les autorités compétentes et celles du Service de santé des armées (C2E ACASAN, référent RGPD, etc.), jouent un rôle crucial en veillant à ce que les recherches respectent les normes éthiques et légales. L'ouverture des données, accompagnée d'une régulation stricte, peut ainsi contribuer à un progrès scientifique tout en préservant les valeurs fondamentales de protection des données personnelles.

En conclusion, bien que la science ouverte soit encouragée, la gestion des données de santé reste un domaine spécifique où l'éthique, la réglementation, la sécurité des données et la protection du secret doivent être scrupuleusement respectées.

Financements

Ce travail n'a pas bénéficié de financement.

Contributions des auteurs

Conceptualisation, rédaction – version originelle : CV (Réserviste_Acasan).
Rédaction, Révision et correction : RH, SM, AC, JB, AS, AM.

Remerciements au linguiste relecteur : Pierre-Alexandre Clerc.

Références bibliographiques

1. Zins M, Cuggia M, Goldberg M. Les données de santé en France. Abondantes mais complexes. Med Sci MS. 2021 ;37 :179–184. doi :10.1051/medsci/2021001
2. Hervé C, Forest C, Noûs C, Haiech J. Plaidoyer pour une culture scientifique ouverte et socialement responsable. Ethics Med Public Health. 2020 ;14 : 100526. doi :10.1016/j.jemep. 2020.100526
3. Lucas E, Fouquet A, Jezewski-Serra D, Ben Raies J, de Crouy-Chanel P, Alleaume C. Santé Publique France face au défi de l'Open Data. Bull Epidemiol Hebd. 2024 ;475–81.
4. Règlement (UE) 2025/327 du Parlement européen et du Conseil du 11 février 2025 relatif à l'espace européen des données de santé et modifiant la directive 2011/24/UE et le règlement (UE) 2024/2847.
5. Hulier-Ammar E, Chioccarello A, Touche P, Ivasilevitch A, Stoeklé H-C, Hervé C. Recherche sur données : aspects juridiques et éthiques à travers l'expérience de l'hôpital Foch. Médecine Droit. 2022 ;2022 :8–14. doi :https://doi.org/10.1016/j.meddro.2021.06.003
6. Règlement (UE) 2016/679 du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

7. Directive n° 507029/ARM/DCSSA/SDD/SES/NP du 08 juin 2022 relative à l'organisation et au pilotage de la recherche et de l'innovation dans le service de santé des armées.
8. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 ;3 :160018. doi :10.1038/sdata.2016.18
9. Règlement (UE) n° 536/2014 du Parlement européen et du conseil du 16 avril 2014 relatif aux essais cliniques de médicaments à usage humain et abrogeant la directive 2001/20/CE.
10. Règlement (UE) n° 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE.
11. Règlement (UE) n° 2017/746 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic *in vitro* et abrogeant la directive 98/79/CE et la décision 2010/227/UE de la Commission.
12. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dite « loi informatique et libertés » modifiée par la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles.
13. Huston P, Edge VL, Bernier E. Reaping the benefits of Open Data in public health. Can Commun Dis Rep Releve Mal Transm Au Can. 2019 ;45 : 252–256. doi :10.14745/ccdr.v45i10a01
14. Guide pratique de la publication en ligne et de la réutilisation des données publiques ("Open Data"). Commission Nationale Informatique et Libertés, Commission d'accès aux documents administratifs.
15. Meier M, Vinkers CH, Pruessner JC, Sep MSC. Open science needs a standardized data format : Suggestions for the field of psychoneuroendocrinology. Psychoneuroendocrinology. 2024 ;169 :107170. doi :10.1016/j.psyneuen.2024.107170
16. Léger M. Intégrité scientifique : le nouveau défi de la science et de la recherche. Anesth Réanimation. 2023 ;9 : 428–433. doi :10.1016/j.anrea.2023.10.007
17. Beaufils P. Intégrité scientifique et open access. Bull Ortho Francoph. 2022 ;92 :28–32.
18. Logullo P, De Beyer JA, Kirtley S, Schlüssel MM, Collins GS. Open access journal publication in health and medical research and open science : benefits, challenges and limitations. BMJ Evid-Based Med. 2024 ;29 :223–228. doi :10.1136/bmjebm-2022-112126

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816