Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Le Conseil fédéral

Berne, 4 mai 2022

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Rapport du Conseil fédéral donnant suite au postulat 15.4225 Humbel du 18.12.2015

Table des matières

Liste des abréviations 3
Liste des tableaux et des figures 5
Liste des annexes 5
Résumé 6
Remarque liminaire 8
1 Contexte 9
1.1 Postulat 9
1.2 Contexte 10
1.2.1 Exigences du postulat 10
1.2.2 Interventions parlementaires connexes 10
1.2.3 Précisions conceptuelles 11
1.3 Position dans le contexte des développements actuels 15
1.3.1 Politique en matière de données et numérisation 15
1.3.2 Projets dans le domaine du traitement des données médicales 17
1.4 Objectif du rapport 22
2 Potentiel de la réutilisation et de l’appariement de données médicales pour le domaine de la santé 23
2.1 Surmonter les « distances » existantes 23
2.2 Amélioration de l’efficience 25
3 Conditions pour la réutilisation et l’appariement de données médicales 25
3.1 Développement de l’espace de données dans le secteur de la santé 25
3.2 Conditions-cadres optimales pour l’application des principes FAIR 26
3.2.1 Facilité à trouver des données réutilisables à l’aide de catalogues de métadonnées 26
3.2.2 Respect des exigences qualitatives posées aux données qui peuvent être réutilisées et appariées 27
3.2.3 Permettre l’accès aux données 28
3.2.4 Bases légales pour l’accès aux données et l’utilisation des données 29
4 Proposition de mise en œuvre d’un « système national de réutilisation et d’appariement des données médicales à des fins de recherche » 34
4.1 Mise en place de structures et de processus sur le principe de modularité 36
4.2 Conception fonctionnelle du système 39
4.2.1 Créer les conditions pour la fonctionnalité du système 39
4.2.2 Assurer une conception formellement juridique 43
5 Conclusions du Conseil fédéral et prochaines étapes 43
6 Bibliographie 45
7 Glossaire 49
8 Annexes 53

Liste des abréviations

AOS	Assurance obligatoire des soins
ASSM	Académie suisse des sciences médicales
BSH	Biosurveillance humaine
CCE	Commission centrale d’éthique de l’ASSM
CdC	Centrale de compensation
CG	Consentement général
CNE	Commission nationale d’éthique dans le domaine de la médecine humaine
CSEC-N	Commission de la science, de l’éducation et de la culture du Conseil national
DCC	SPHN Data Coordination Center
DEFR	Département fédéral de l’économie, de la formation et de la recherche
DEP	Dossier électronique du patient
DFI	Département fédéral de l’intérieur
DFJP	Département fédéral de justice et police
DTUA	Data Transfer and Use Agreement
e-ID	Identité électronique
ELSIag	Ethical, Legal and Social Implications advisory group
EPFL	École polytechnique fédérale de Lausanne
EPFZ	École polytechnique fédérale de Zurich
EPR-SPID	Electronic Patient Record – Sectorial Person Identification Number, caractéristique d’identification du patient au sens de l’art. 4 LDEP
FAIR	findable, accessible, interoperable, re-usable
FNS	Fonds national suisse de la recherche scientifique
FRI	Formation, recherche et innovation
I14Y	Interopérabilité
ISB	Institut suisse de bioinformatique
ITSL	Center for Information Technology, Society, and Law (Université de Zurich)
LAMal	Loi fédérale sur l’assurance-maladie ; RS 832.10
LAVS	Loi fédérale sur l’assurance-vieillesse et survivants ; RS 831.10
LDEP	Loi fédérale sur le dossier électronique du patient ; RS 816.1
LE	Logiciel d’enregistrement pour les registres des tumeurs
LEMO	Loi fédérale sur l’enregistrement des maladies oncologiques ; RS 818.33
LEp	Loi fédérale sur la lutte contre les maladies transmissibles de l’homme (loi sur les épidémies) ; RS 818.101
LERI	Loi sur l’encouragement de la recherche et de l’innovation ; RS 420.1
LOGA	Loi sur l’organisation du gouvernement et de l’administration ; RS 172.010
LPD (nLPD)	Loi fédérale sur la protection des données ; RS 235.1 (révision totale de la LPD)
LRH	Loi fédérale relative à la recherche sur l’être humain ; RS 810.30
LSF	Loi fédérale sur la statistique fédérale ; RS 431.01
Mo	Motion
NaDB	Programme de gestion nationale des données
NAVS13	Numéro AVS (assurance-vieillesse et survivants) à 13 chiffres
Obsan	Observatoire suisse de la santé
OCDE	Organisation de coopération et de développement économiques
OFCOM	Office fédéral de la communication
OFJ	Office fédéral de la justice
OFS	Office fédéral de la statistique
OFSP	Office fédéral de la santé publique
OGD	Open Government Data
ONCD	Organe national de coordination des données

Mieux utiliser les données médicales pour assurer l’efficience et la qualité des soins

ONEC	Organe national d’enregistrement du cancer
ORD	Open Research Data
ORH	Ordonnance relative à la recherche sur l’être humain à l’exception des essais cliniques (ordonnance relative à la recherche sur l’être humain) ; RS 810.301
OSAV	Office fédéral de la sécurité alimentaire et des affaires vétérinaires
PFPDT	Préposé fédéral à la protection des données et à la transparence
PHRT	Personalized Health and Related Technologies
PIMS	Personal Information Management Systems
PNR	Programme national de recherche
Po	Postulat
RGPD	Règlement général sur la protection des données
SBP	Swiss Biobanking Plattform
SDSC	Swiss Data Science Center
SEFRI	Secrétariat d’État à la formation, à la recherche et à l’innovation
SLHS	Swiss Learning Health System
SNPC	Stratégie nationale de protection de la Suisse contre les cyberrisques
SPHN	Swiss Personalized Health Network
SpiGes	Secteur stationnaire hospitalier (projet pilote dans le cadre du programme NaDB)
swissethics	Association suisse des commissions d’éthique de la recherche
TIC	Technologies de l’information et de la communication
UE	Union européenne
UniNE	Université de Neuchâtel
UPIC	Unité de pilotage informatique de la Confédération

Liste des tableaux et des figures

Tableau 1 Explication de l’utilisation de données médicales à des fins de recherche et génération de connaissances 13

Tableau 2 Partenaires potentiels dont les unités fonctionnelles et spécialisées peuvent contribuer à la mise en place de l’ONCD 38

Figure 1 Distances surmontables grâce à la réutilisation et à l’appariement des données 23

Figure 2 Exemples d’application pour les distances surmontables grâce à la réutilisation et à l’appariement des données 24

Figure 3 Conception de l’espace de données dans le secteur de la santé 26

Figure 4 Système national de réutilisation et d’appariement des données médicales à des fins de recherche 35

Figure 5 Esquisse pour un ancrage du cadre juridique 43

Liste des annexes

Annexe 1 Réutilisation de données médicales par des tiers 53

Annexe 2 Ancrage des projets visant une meilleure utilisation des données dans l’administration fédérale 57

Annexe 3 Projets d’appariement : défis et potentiel de solutions 58

Annexe 4 Bases légales pour la réutilisation et l’appariement de données 64

Annexe 5 Formes du consentement en Suisse : état actuel et besoins de développement 67

Annexe 6 Exemples de transmission de données avec un changement de finalité et la nécessité d’un consentement 71

Annexe 7 Esquisse de processus : utilisation secondaire des données dans le système de réutilisation et d’appariement des données médicales à des fins de recherche 71

Mieux utiliser les données médicales pour assurer l’efficience et la qualité des soins

Résumé

Mieux utiliser les données médicales pour assurer l’efficience et la qualité des soins ; Rapport du Conseil fédéral donnant suite au postulat 15.4225 Humbel du 18 décembre 2015

Le présent rapport a été élaboré sous la direction de l’Office fédéral de la santé publique (OFSP), en collaboration avec l’Office fédéral de la statistique (OFS) et le Secrétariat d’État à la formation, à la recherche et à l’innovation (SEFRI).

Contexte

Le postulat 15.4225 demande la création de conditions permettant de mieux utiliser les données médicales. L’efficacité, l’adéquation et le caractère économique des soins de santé ne peuvent être prouvés, contrôlés et optimisés que si les données nécessaires sont disponibles et que leur potentiel peut être exploité de manière étendue.

Plusieurs acteurs de la recherche, de la pratique, de la politique et de l’économie se sont intéressés à une utilisation multiple et efficace des données médicales à des fins de recherche, qui sont entretemps nombreuses et souvent collectées à plusieurs reprises. La réutilisation et l’exploitation appariée, c’est-à-dire conjointe, de données sur les mêmes personnes à partir de plusieurs jeux de données doivent se faire dans le cadre de mesures nécessaires en matière de protection et de sécurité des données, mais elles sont souvent semées d’obstacles, comme l’a récemment mis en évidence la pandémie de COVID-19.

Ancrage et objectif

Le rapport s’inscrit dans l’environnement dynamique de la politique des données et de la numérisation et tient compte des interfaces avec des projets publics et privés en cours en matière de gestion des données, en particulier dans le domaine de la santé. La mise en réseau avec des acteurs a eu lieu. La cohérence avec les principes de la gestion des données dans le domaine de la santé, que le Conseil fédéral a fixés le 12 janvier 2022, est garantie.

Dans le but de présenter une proposition pratique sur les possibilités d’une utilisation commune et d’une exploitation appariée des données médicales de différentes sources de tiers à des fins de recherche, le rapport creuse trois questions centrales :

- Quel est le potentiel d’une réutilisation et d’une exploitation appariée des données médicales pour le domaine de la santé ?
- Quelles conditions doivent être remplies pour que les données médicales puissent et soient autorisées à être réutilisées et traitées par différents groupes d’utilisateurs ?
- Quels développements techniques, organisationnels, et juridiques faut-il initier pour remplir ces conditions dans la pratique ?

Constatations

Une réutilisation et un appariement efficaces des données permettent d’acquérir des connaissances socialement précieuses à partir des mêmes sources de données de manière répétée. L’utilisation multiple des données médicales implique un gain d’efficacité grâce à une flexibilité accrue. L’investissement des ressources peut être optimisée pour tous les acteurs des données : il en résulte une plus-value sous la forme de connaissances avec des charges réduites (temps, personnel, finances).

La réutilisation et l’appariement des données médicales présupposent un espace de données médicales fonctionnel qui garantit que les jeux de données sont faciles à trouver, accessibles, compatibles et réutilisables (principes FAIR).

Des développements sont nécessaires non seulement pour les aspects techniques, opérationnels et juridiques, mais aussi pour la mise en place d’une culture commune de réutilisation des données. Pour que la collaboration entre les acteurs puisse fonctionner de manière ordonnée et efficace, les conditions suivantes doivent être remplies concernant les données et les processus qui s’y rapportent :

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

- les données réutilisables sont faciles à trouver par le biais des catalogues de métadonnées ;
- les exigences qualitatives posées aux données réutilisables et appariables sont remplies, à savoir :
  - des normes communes sont respectées pour l'échange interopérable de données (interopérabilité syntaxique et sémantique) ;
  - un identificateur personnel univoque peut être utilisé comme variable d'appariement ;
- l'accès aux données est accordé, les producteurs donnent aux utilisateurs secondaires l'accès à leurs données ;
- les bases légales permettent un accès aux données en fonction des applications en vue d'une utilisation secondaire des données (réutilisation et appariement de données médicales).

Pour chacun de ces aspects, il existe actuellement des lacunes, et les réglementations légales en vigueur posent des conditions très exigeantes à la réutilisation et en particulier à l'appariement de données médicales, voire les rendent impossibles.

Proposition de mise en œuvre

Le rapport parvient à la conclusion qu'il est nécessaire de créer un « système national de réutilisation et d'appariement des données médicales » afin que les acteurs des données disposent de conditions-cadres acceptables et intéressantes pour une utilisation multiple des données médicales à des fins de recherche. Le système esquissé dans le rapport doit être piloté à l'aide d'un cadre juridique. Un organe national de coordination des données s'assurerait que l'échange des données et le traitement qui en découle seraient conformes au droit et sûrs.

Pour que la transmission et la réutilisation des données avec ou sans appariement préalable à des fins de recherche soient illimitées et efficaces, il est nécessaire de créer les conditions permettant la mise en œuvre des quatre points suivants :

(1) une forme adéquate du consentement des personnes concernées à l'utilisation de leurs données (gestion du consentement),
(2) un identificateur personnel univoque comme variable d'appariement,
(3) une gestion efficace de données de qualité conformément aux principes FAIR,
(4) la préservation de la protection et de la sécurité des données grâce à un concept étendu de protection des données.

Pour chacun de ces quatre points, le rapport présente des propositions qui doivent être examinées par des experts d'un point de vue technique, organisationnel et juridique en tenant compte des parties prenantes, puis concrétisées en vue de leur mise en œuvre dans la pratique. Les travaux, projets ou modèles existants ou en cours doivent être pris en compte de manière ciblée pour établir les structures et les processus de gestion des données et la conception fonctionnelle du système.

Conclusion du Conseil fédéral

Le développement de la gestion des données dans le domaine de la santé, qui constitue autant un défi pour les autorités, les fournisseurs de prestations, les assureurs et la recherche, doit avancer de manière bien coordonnée.

C'est pourquoi le Conseil fédéral a l'intention – dans le cadre des mesures visant à améliorer la gestion des données dans le domaine de la santé qu'il a adoptées le 12 janvier 2022 – de concrétiser la proposition de mise en œuvre d'un système national de réutilisation et d'appariement des données médicales à des fins de recherche, telle qu'elle est présentée dans le présent rapport.

De concert avec les organes spécialisés existants de l'administration fédérale, le cas échéant en associant d'autres experts externes compétents et les acteurs concernés, le DFI est chargé de préciser les processus et les structures du système proposé quant à leur plausibilité, leur faisabilité et sous l'angle de la prise en compte internationale, puis de clarifier quelles bases légales doivent être adaptées ou créées pour la mise en œuvre et comment évaluer les conséquences en matière de protection des données. Il est nécessaire de garantir la collaboration avec les unités administratives concernées de la ChF, du DFF, du DFJP et du DEFR.

Remarque liminaire

Le présent rapport fait suite au postulat « Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins » (15.4225), transmis par le Conseil national le 16 juin 2016. Il comprend des propositions visant à développer la gestion des données dans le domaine de la santé, en mettant l'accent sur l'utilisation multiple et l'appariement des données liées à la santé à des fins de recherche. Il s'agit de l'un des nombreux projets menés dans l'environnement dynamique de la transformation numérique du système de santé.

La proposition de mise en œuvre soumise dans le rapport se fonde sur les principes de projets en cours (cf. chap. 1.3.2). Cette proposition concorde surtout avec les sept principes de la gestion des données dans le domaine de la santé que le Conseil fédéral a fixés dans le cadre de l'adoption du « Rapport concernant l'amélioration de la gestion des données dans le domaine de la santé » le 12 janvier 2022[1], car le thème « Données pour les chercheurs » fait explicitement partie de l'approche globale de la gestion des données dans le système de santé.

La prise en compte de ces principes se présente pour l'essentiel comme suit.

Concrètement, le rapport propose un système de réutilisation et d'appariement des données médicales et esquisse des approches techniques, sémantiques, organisationnelles et juridiques pour permettre aux chercheurs d'utiliser les données existantes de manière sûre et conforme à la protection des données (cf. chap. 4). Le rapport n'a explicitement pas pour objet un examen étayé des questions juridiques (relevant de la protection des données) qui découlent de la réutilisation et de l'exploitation appariée des données médicales (voir à cet égard la remarque liminaire sur l'orientation des explications juridiques au chap. 3.2.4).

Les structures et les processus existants ou en cours de constitution qui servent aux échanges de données doivent être intégrés dans le système pour que les flux de données existants soient utilisés dans l'espace de données dans le secteur de la santé (principe 6). La création d'un cadre juridique doit rendre les processus et les structures du système proposé largement applicables. La ou les propositions partent du principe que l'utilisation secondaire des données à des fins de recherche doit être possible indépendamment du but de la collecte de données primaires (principe 2). Une « *consent management* » (gestion du consentement des personnes concernées) est proposée pour permettre le flux de données entre les producteurs/détenteurs de données initiaux et d'autres utilisateurs de données intéressés (principe 6). L'introduction d'un identificateur univoque, disponible comme variable d'appariement, est prévue dans le but de concrétiser la réutilisation des données (principe 5).

La concrétisation de la proposition d'un « système national de réutilisation et d'appariement des données médicales », nécessaire en vue de la mise en œuvre, doit se faire de concert avec les organes spécialisés existants, le cas échéant en tenant compte d'autres experts importants et des acteurs concernés (principe 7). Les tâches sont réalisées dans le cadre des mesures jugées importantes pour la recherche dans le but d'améliorer la gestion des données dans le domaine de la santé et la transformation numérique de l'OFSP.

[1] Cf. « Rapport concernant l'amélioration de la gestion des données dans le domaine de la santé », chap. 6.2

Mieux utiliser les données médicales pour assurer l’efficience et la qualité des soins

1 Contexte

1.1 Postulat

Le 18 décembre 2015, la conseillère nationale Ruth Humbel a déposé le postulat « Mieux utiliser les données médicales pour assurer l’efficience et la qualité des soins » (15.4225).

Texte déposé

Le Conseil fédéral est chargé de présenter dans un rapport de quelle manière les données contenues dans les registres spécifiques à certaines maladies ou dans les études médicales pourraient être reliées afin d’être mieux exploitées. Il fera des propositions sur la manière d’atteindre ce but en garantissant les aspects suivants notamment :

1. la comparabilité des données ;

2. leur claire attribution à la même personne (p. ex. grâce au numéro AVS) ;

3. la sauvegarde de la protection des données (p. ex. en prévoyant un chiffrement approprié des informations permettant d’identifier les personnes et en réglant juridiquement des procédures clairement définies pour relier les données).

Développement

La loi sur l’enregistrement des maladies oncologiques, sur laquelle le Conseil national a statué en conseil prioritaire, vise à améliorer la prévention, le dépistage précoce et le traitement des maladies oncologiques. Nul ne conteste le fait que de meilleures données médicales facilitent la planification et renforcent la garantie de la qualité des soins ; elles contribuent également à améliorer la pose de l’indication (pour les décisions sur les tests diagnostiques et sur les thérapies) et les thérapies. Seules des données médicales ciblées et représentatives de la population permettent de se faire une image plus précise de la situation et donc de mieux piloter le système des soins. Aussi est-il indispensable de relier les différents registres spécifiques à certaines maladies et de faire en sorte qu’on puisse les utiliser.

Avis du Conseil fédéral du 17.2.2016

L’appariement des données ainsi que l’accès des chercheurs aux données du registre (des cancers) ont été abordés à plusieurs reprises dans le cadre des délibérations parlementaires concernant la loi fédérale sur l’enregistrement des maladies oncologiques (LEMO ; RS *14.074*). Il s’avère, entre autres, que les chercheurs estiment qu’il n’est pas toujours facile, dans certains cas, de déterminer si c’est la loi relative à la recherche sur l’être humain (LRH ; RS *810.30*) ou la loi sur la statistique fédérale (LSF ; RS *431.01*) qui s’applique. Pour les raisons susmentionnées, il semble judicieux de présenter brièvement dans un rapport les possibilités qui existent actuellement en matière d’appariement des données et d’accès aux données. Il y a lieu également d’intégrer à ce rapport les premiers constats découlant des demandes des chercheurs dans le cadre de la mise en œuvre du programme national de recherche 74 « Système de santé », notamment concernant l’appariement des données et une meilleure exploitation des sources de données existantes (registre, données relatives aux études, relevés statistiques).

Le postulat (Po 15.4225) a été adopté par le Conseil national le 16 juin 2016. Le Département fédéral de l’intérieur (DFI), compétent en l’espèce, a chargé l’Office fédéral de la santé publique (OFSP) d’établir le présent rapport en réponse au postulat.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

1.2 Contexte

1.2.1 Exigences du postulat

Le postulat 15.4225 demande, dans l'intérêt de la population et des acteurs du domaine de la santé, d'améliorer constamment l'utilisation et la qualité des données relatives à la santé – en bref, les données médicales [1, 2]. Concrètement, il s'agit de créer les conditions afin que les données médicales offrent une bonne base de planification pour des soins de santé de qualité et en constante évolution pour la population suisse. Au sens de l'efficacité, de l'adéquation et du caractère économique des interventions réalisées, il est nécessaire de démontrer et de contrôler constamment la preuve des soins. La qualité et les coûts des soins doivent être analysés et optimisés. Pour ce faire, les données déterminantes doivent être disponibles et leur potentiel doit pouvoir être largement exploité par des institutions aussi bien privées que publiques. La réutilisation et l'exploitation le cas échéant appariée des données existantes permettent d'éviter de nouvelles collectes de données et dépenses.

Le postulat et l'avis du Conseil fédéral du 17 février 2016 relèvent par conséquent la nécessité de présenter des possibilités et d'élaborer des propositions concrètes visant à ce que les chercheurs puissent mieux utiliser les sources de données existantes et exploiter les données appariées. Les propositions doivent garantir :

(1) la comparabilité des données,

(2) leur claire attribution à la même personne, par exemple grâce au numéro d'assurance sociale NAVS13,

(3) la sauvegarde de la protection des données (p. ex. grâce à un chiffrement),

(4) des procédures clairement définies et juridiquement réglées pour apparier les données.

1.2.2 Interventions parlementaires connexes

La réponse au postulat 15.4225 se fait d'un point de vue actuel. Cela implique la prise en compte des développements des six dernières années (2016 à 2021) et des exigences pratiques actuelles, en particulier de la part des chercheurs.

Le présent rapport peut donc aussi aborder certains points des interventions parlementaires suivantes, consacrées à une meilleure utilisation multiple des données médicales et déposées entre 2019 et 2021 :

- Po 19.4193 Nouvelle loi-cadre nationale sur la santé ; déposé par Graf Maya le 26.9.2019, repris par Wettstein Felix le 5.12.2019. Liquidé. Rejeté au Conseil national le 22.9.2021.
- Question 20.1034 Données anonymisées des patients utilisées à des fins de recherche. Tirer les enseignements de la lutte contre le coronavirus ; déposée par Eymann Christoph le 19.6.2020. Liquidé.
- Po 20.3700 Examiner la faisabilité d'un don volontaire de données personnelles anonymisées, en vue d'un usage d'intérêt public ; déposé par Bellaïche Judith le 17.6.2020. Non encore traité au conseil.
- Mo 20.4672 Établir un calendrier contraignant pour la transformation numérique dans le système de santé ; déposée par Humbel Ruth le 18.12.2020. Non encore traitée au conseil.
- Mo 20.4717 Souveraineté numérique sur les données médicales ; déposée par Flach Beat le 18.12.2020. Non encore traitée au conseil.
- Mo 21.3021 Création d'un écosystème intégré de données médicales. Une plus-value pour la recherche et pour la société ; déposée par la CSEC-N le 18.2.2021. Adoption au Conseil national le 16.6.2021, motion suspendue au 2e conseil le 26.10.2021.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

- Po 21.3195 Tirer les leçons de la pandémie pour renforcer la place scientifique suisse ; déposé par Dittli Josef le 16.3.2021. Adopté au Conseil des États (9 points sur 12) le 7.6.2021.
- Mo 21.3957 Transformation numérique dans le système de santé. Rattraper enfin notre retard ! Déposée par Ettlin Erich le 18.6.2021. Adoptée au Conseil des États le 27.9.2021. Adoptée au Conseil national le 17.3.2022.

1.2.3 Précisions conceptuelles

C'est **dans une perspective d'avenir et une approche pratique** que le rapport doit déterminer comment les chercheurs peuvent et doivent réutiliser et exploiter les données personnelles pertinentes pour la santé et les soins de santé sous une forme appariée en toute sécurité (cf. chap. 1.2.1). C'est pourquoi, pour répondre au postulat, nous avons pris pour base la conception suivante des trois composantes principales de cette question :

(1) Données médicales

Les données relatives à la santé sont l'objet de la présente considération. Du fait de l'application de nouvelles méthodes de collecte des données – par exemple par le biais de capteurs, d'appareils mobiles, etc. –, il y a toujours plus de données qui ont ou pourraient avoir un rapport avec la santé d'une personne. En même temps, les possibilités techniques et analytiques permettant d'obtenir, d'enregistrer et d'exploiter des données sont de plus en plus nombreuses. Les acteurs publics et privés manifestent ainsi un intérêt croissant pour **une utilisation et une exploitation appariée d'un très large spectre de données pour répondre à des questions liées à la santé**.

C'est pourquoi le présent rapport, **d'un point de vue technique et pratique**, part d'une conception large du terme « données médicales », qui comprend les types de données suivants :

a) **Les données qui décrivent la santé et l'état de santé d'une personne.**
Exemples : données de routine des soins stationnaires et ambulatoires, données des assurés, données génétiques, données des projets de recherche clinique et épidémiologique, renseignements personnels sur la santé ou les maladies dans les enquêtes ou les auto-mesures.

b) **Les données qui décrivent les déterminants de la santé[2] d'une personne.**
Exemples [3] : informations sur les comportements (facteurs liés au mode de vie tels que l'alimentation, la consommation de tabac, etc.), sur les conditions de vie et de travail et, par conséquent, sur l'intégration sociale ainsi que sur les conditions-cadres socio-économiques, culturelles, structurelles et écologiques dans lesquelles une personne vit et agit ; l'accès aux soins médicaux en fait également partie.

c) **Les données qui décrivent le système de santé et livrent ainsi des informations sur les offres de prestations liées à la santé et leur utilisation.**
Exemples : informations sur la structure des prestataires, les coûts et le financement[3].

Il s'agit donc d'informations qui se rapportent à des personnes physiques (a et b) et à des établissements de soins (c).

Le spectre des **sources de données**[4], c'est-à-dire les **endroits où ces données médicales sont produites et administrées**, est aussi large. Ces données sont en partie produites à plusieurs endroits et donc plusieurs fois, ou elles sont produites une seule fois et stockées à plusieurs endroits.
Exemples : les établissements de santé (examens stationnaires, ambulatoires ou mobiles et traitement/prise en charge), les laboratoires médicaux, les biobanques, les registres, les hautes écoles ainsi que les établissements de recherche publics et privés (projets, études de cohorte, études d'intervention et autres études scientifiques, p. ex. mesures environnementales), les autorités (relevés et évaluations statistiques et administratifs, systèmes de surveillance de la santé et des maladies), les

[2] Facteurs qui peuvent avoir directement ou indirectement une influence positive ou négative sur la santé d'une personne physique.
[3] Voir l'ordonnance sur l'assurance-maladie (OAMal ; RS *832.102*), art. 28 et 30
[4] Les sources de données primaires sont le lieu de provenance où les données ont été générées. De là, des données choisies peuvent être transmises à des sources secondaires, par exemple à des registres, des assurances ou des systèmes de surveillance et de monitorage. Les résultats des évaluations de données, tels qu'ils sont publiés dans la recherche ou disponibles dans l'administration et auprès des partenaires tarifaires, sont des sources de données tertiaires. Pour des exemples, voir [4–7].

assureurs-maladie, les applications sur la saisie mobile de données au quotidien (télémonitoriage par le biais des technologies portables et des applications de santé), etc.

Lorsque le présent rapport décrit les **exigences légales** en vigueur, les **définitions selon la loi fédérale** sur la protection des données (LPD ; RS *235.1*) ou la loi fédérale relative à la recherche sur l'être humain (LRH ; RS *810.30*) sont applicables (cf. chap. 3.2.4).

Les considérations de fond détaillées dans le présent rapport concernent principalement les données de personnes physiques.

(2) Réutilisation de données par des tiers

Si des données ou des jeux de données sont transmis à d'autres personnes et institutions, c'est-à-dire partagés avec d'autres pour une utilisation commune, on emploie aussi le terme d'échange de données (*data sharing*)[5]. Les destinataires peuvent utiliser les données reçues avec ou sans appariement.

a) Réutilisation sans appariement

Il y a deux possibilités : (1) un jeu de données est traité par le destinataire dans la forme fournie pour répondre à une problématique spécifique et directement utilisé pour les évaluations en question. (2) Le destinataire combine le jeu de données fourni avec un ou plusieurs autres jeux de données comparables. Un nouveau jeu de données est ainsi créé, qui contient des informations sur un plus grand nombre de personnes.

b) Réutilisation avec appariement

Si des jeux de données sont partagés et reliés avec d'autres jeux de données pour lesquels il est question en totalité ou en partie de données sur les mêmes personnes, les jeux peuvent être appariés. En cas d'appariement de données (*record linkage*), les données individuelles sur les mêmes personnes de deux jeux au moins sont fusionnées à l'aide d'une ou plusieurs variables d'appariement. Ces variables permettent d'établir s'il s'agit bel et bien de jeux de données sur les mêmes personnes ou sur des personnes différentes. L'appariement génère la création d'un nouveau jeu de données qui, pour chaque personne, couvre davantage de thèmes (étendue de l'information) et/ou contient des informations détaillées (profondeur de l'information).

Des caractéristiques personnelles servent de **variables d'appariement**. Il y a des caractéristiques qui identifient clairement une personne et sont qualifiées d'« identificateur ». Le numéro d'assurance sociale NAVS13 en est un exemple. D'autres caractéristiques telles que le nom, la date de naissance, la nationalité, le numéro postal d'acheminement ou le domicile ne permettent d'identifier une personne que partiellement. Elles peuvent correspondre à plusieurs personnes ou subir des changements, mais aussi être saisies de manière incomplète ou écrites de manière erronée, etc. Dans l'idéal, les jeux de données à apparier contiennent la même caractéristique univoque – un identificateur qui permet un appariement aussi simple que possible sur le plan méthodologique et avec le moins d'erreurs possible.

Deux méthodes d'appariement sont appliquées dans la pratique, seules ou combinées.

L'**appariement déterministe** se fonde sur l'établissement de la concordance univoque ou non entre des caractéristiques/variables qui sont communes aux jeux de données source à apparier. Il peut par exemple s'agir du prénom et du nom ; un identificateur univoque comme le NAVS13 permet toutefois d'établir la concordance des jeux de données individuels de manière nettement plus claire et avec une marge de doute plus faible.

Dans l'**appariement probabiliste**, les jeux de données source sont comparés à l'aide de plusieurs caractéristiques/variables personnelles, par exemple le prénom, le nom, le sexe et la date de naissance. La méthode est donc basée sur les probabilités car elle s'appuie, pour chacune des caractéristiques, sur le calcul de la probabilité d'observer une concordance. Aussi bien la discrimination des caractéristiques personnelles disponibles que la qualité des données à apparier et la conception de la

[5] Les termes qui sont traduits sont souvent utilisés en anglais dans la littérature spécialisée et dans le langage courant.

stratégie d'appariement sont déterminantes pour le succès de l'appariement. Cette méthode est très exigeante, exige beaucoup de temps et présuppose une grande expérience [8, 9].

Pour de plus amples explications, voir le chapitre 8, annexe 1.

(3) Utilisation des données à des fins de recherche

Dans la discussion sur les possibilités d'une utilisation multiple et d'une exploitation appariée des données médicales, il est en principe nécessaire de faire la distinction entre l'utilisation des données aux fins envisagées initialement – l'utilisation primaire – et l'utilisation à des fins qui en divergent – l'utilisation secondaire.

L'**utilisation secondaire des données à des fins de recherche** est au cœur du présent rapport, contrairement aux fins purement statistiques ou administratives. L'utilisation secondaire et le traitement de données à des fins relevant du droit de la surveillance ou de la politique de la santé ne font explicitement pas l'objet de ce rapport. Les données collectées initialement à des fins de surveillance doivent toutefois pouvoir être réutilisées comme des sources de données à des fins de recherche.

Pour la définition de la « finalité[6] » de la recherche, le présent rapport se fonde directement sur l'**utilisation multiple de données médicales par des chercheurs/particuliers dans la pratique**. L'utilisation multiple de données médicales concerne et encourage la recherche académique et industrielle, la recherche de l'administration fédérale, mais aussi l'obtention/la recherche d'informations pour le développement de la politique de la santé en général.

La vue d'ensemble ci-après (tableau 1 ; pour les définitions, voir le chap. 7, Glossaire) donne davantage d'explications. Dans la pratique, il n'est pas toujours possible d'attribuer clairement une utilisation primaire ou secondaire des données ou un projet à l'un des « domaines à des fins de recherche ».

Tableau 1 Explication de l'utilisation de données médicales à des fins de recherche et génération de connaissances

Utilisation de données aux fins de la recherche académique
Il s'agit de la recherche méthodologique de nouvelles connaissances. Elle comprend : – la recherche fondamentale, dont l'objectif premier est l'acquisition de connaissances ; – la recherche appliquée, dont l'objectif premier est de contribuer à la résolution de problèmes pratiques. Elle génère des connaissances servant à définir des actions et des orientations. Cercle d'utilisateurs des données : les universités/hautes écoles et les instituts de recherche extra-universitaires, privés et publics. Caractéristiques : – Questions scientifiques en rapport avec la santé de la population – Génération de nouvelles connaissances dans différentes disciplines scientifiques – Fourniture de bases scientifiques pour l'innovation, p. ex. du système de santé – Collaboration nationale et internationale – Soutien en grande partie par les pouvoirs publics par le biais des organes de la Confédération chargés d'encourager la recherche Exemples : – Recherche biomédicale et recherche clinique orientée vers les patients – Recherche sur les services de santé (*health service research*) – Recherche en santé publique avec les domaines suivants : prévention et promotion de la santé, épidémiologie des maladies transmissibles et non transmissibles (surveillance, monitorage), recherche d'intervention et d'évaluation, médecine du travail et de l'environnement (biosurveillance, réduction de l'exposition).

[6] En opposition à la « définition » de la recherche telle qu'elle est utilisée dans la LRH.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Utilisation de données aux fins de la recherche du secteur privé / industrielle[7]

Il s'agit de l'étude scientifique de questions pratiques qui découlent de l'activité commerciale ou des objectifs économiques d'une entreprise. Elle comprend :

- la quête du progrès technologique ;
- la recherche de nouvelles connaissances applicables dans la pratique, induite par le but poursuivi.

Cercle d'utilisateurs des données : économie privée (industrie)

Caractéristiques :

- Questions scientifiques en rapport avec la santé de la population, p. ex. diagnostic, traitement
- Innovation basée sur la connaissance, p. ex. développement/introduction de nouveaux produits, procédés de fabrication, etc.
- Collaboration nationale et internationale
- Collaboration avec la recherche académique ou les universités/les hautes écoles ; instituts de recherche universitaire financés par l'industrie
- Intérêt commercial

Exemples :

- Recherche biomédicale et recherche clinique orientée vers les patients, p. ex. études pharmacologiques
- Développement et test de technologies biomédicales

Utilisation de données aux fins de la recherche de l'administration

Il s'agit de la recherche initiée par l'administration fédérale, parce que celle-ci a besoin des résultats de cette recherche pour accomplir ses tâches.[8]

Cercle d'utilisateurs des données : établissements publics et privés actifs dans le domaine scientifique, mandatés ou soutenus par l'administration fédérale.

Caractéristiques :

- Axée sur la politique, pratique, souvent interdisciplinaire et réactive
- Génération de nouvelles connaissances, p. ex. pour le domaine politique de la santé
- Sert à l'acquisition de connaissances scientifiques ou technologiques et à la formation de l'opinion

Exemples :

- Octroi de mandats de recherche à des tiers (recherche contractuelle), p. ex. dans le domaine des maladies transmissibles et non transmissibles, des maladies professionnelles, de l'alimentation, de la biomédecine, de l'évaluation des conséquences de techniques médicales, etc.
- Soutien des établissements de recherche universitaire pour la réalisation de programmes de recherche
- Réalisation de ses programmes de recherche, notamment en collaboration avec les établissements de recherche universitaire, les institutions de promotion de la recherche et les organes d'encouragement

Utilisation de données à des fins d'information du système de santé

Il s'agit de la mise à disposition de bases techniques élaborées scientifiquement pour la représentation des soins de santé et le développement de la politique de la santé en tant que telle[9]**.** En fait partie l'utilisation de données médicales pour :

- la statistique des soins de santé ;
- la planification des besoins en matière de soins (structures, ressources) ;
- le pilotage et le soutien dans l'exécution des processus ;
- la garantie et l'amélioration de la qualité ;
- l'amélioration de l'efficacité des prestations médicales (efficacité, adéquation, économicité).

Cercle d'utilisateurs des données : autorités sanitaires (Confédération, cantons et communes), fournisseurs de prestations (fédérations), assureurs et autres acteurs du système de santé

Caractéristiques :

- Utilisation des données dans l'accomplissement des tâches légales, p. ex. rapports de santé, gestion de crise

[7] Cf. Pharmaforschung im 20. Jahrhundert - Research Collection (ethz.ch), travail à la frontière entre haute école et industrie.
[8] Cf. art. 16 RS *420.1* – Loi fédérale du 14 décembre 2012 sur l'encouragement de la recherche et de l'innovation (LERI) (admin.ch) et La recherche à l'OFSP (admin.ch)
[9] Conformément à l'art. 57 RS 172.010 - Loi du 21 mars 1997 sur l'organisation du gouvernement et de l'administration (LOGA) (admin.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

- Caractère de routine
- Orientation nationale

1.3 Position dans le contexte des développements actuels

Plusieurs acteurs de la recherche, de la pratique, de la politique et de l'économie sont intéressés à une réutilisation efficace et à un appariement des données médicales, accompagnés de mesures de protection et de sécurité des données. Avec la pandémie de COVID-19, le besoin et donc l'intérêt pour trouver une solution en ce sens ont fortement augmenté[10]. Cette exigence s'inscrit sur un plan thématique dans des stratégies et des mesures actuelles de la Confédération et présente des interfaces avec des projets publics et privés dans le domaine de l'utilisation des données, en particulier des données médicales.

1.3.1 Politique en matière de données et numérisation

En comparaison internationale, la Suisse présente encore des besoins de développement dans la numérisation. L'étude #SmartHealthSystems[11] a constaté un décalage entre une activité politique intense et une utilisation effective des données nettement moins importante, qui se présente comme suit et qu'il s'agit de supprimer [10].

Le 22 mars 2017, le Conseil fédéral a fixé, dans la **politique des données de la Confédération**[12], les objectifs fondamentaux d'une politique des données en Suisse qui soit cohérente et tournée vers l'avenir, et reconnu que les données ouvertes sont la matière première d'une société et d'une économie numériques. Dans ce cadre, il réclame des bases légales modernes pour la gestion des données. Le 9 mai 2018, il a posé les premiers jalons de sa politique des données et prescrit des **mesures relatives aux données ouvertes (*Open Data*) et à la portabilité des données**. La mise en œuvre de la stratégie « Open Government Data (OGD) » 2019-2023[13] et la réalisation d'une stratégie « Open Research Data (ORD) » 2021-2024[14] en font partie.

Par ailleurs, la **Stratégie Suisse numérique** adoptée par le Conseil fédéral le 5 septembre 2018 précise qu'il est nécessaire d'exploiter les opportunités de la numérisation en Suisse et que la Suisse doit ainsi notamment se positionner comme un espace de vie attractif et un pôle économique et scientifique innovant [11]. Le 11 septembre 2020, le Conseil fédéral a adopté la **stratégie actualisée Suisse numérique**[15] pour accorder une plus grande place aux aspects liés aux données dans la politique numérique de la Confédération [12]. Les expériences acquises sur les vulnérabilités de la numérisation durant la pandémie de COVID-19 sont un facteur important. Le Conseil fédéral a désormais fixé une priorité stratégique concernant les **espaces de données fiables** (pour une définition, voir le chap. 7, Glossaire), qui doivent permettre une utilisation des données au-delà de leur destination initiale et des nouvelles formes d'utilisation commune. À cette fin, il est indispensable de soutenir en Suisse l'accès à des espaces de données correspondants pour les personnes, les entreprises et le secteur public, et que tous les acteurs exercent la plus grande autodétermination numérique possible sur leurs données. Il en va de même pour l'espace de données « Santé » et le domaine des soins de santé[16].

Ainsi, la Stratégie **Santé 2030**[17], adoptée par le Conseil fédéral le 6 décembre 2019, se concentre notamment sur les défis liés à l'évolution technologique numérique et à l'influence positive des déterminants de la santé. Mieux exploiter le potentiel existant des données médicales et encourager la numérisation dans le système de santé de manière coordonnée comptent parmi les priorités stratégiques (1.1).

[10] Les mégatendances du système de santé suisse (nfp74.ch)
[11] #SmartHealthSystems (bertelsmann-stiftung.de)
[12] Jalons d'une politique des données en Suisse (admin.ch)
[13] Open Government Data (OGD) | Office fédéral de la statistique (admin.ch)
[14] Cf. Open Science (admin.ch) avec des informations supplémentaires, par exemple, sur la convention de 2020 entre le SEFRI, swissuniversities, le FNS et les deux EPF ; Open Research Data - swissuniversities; Swiss_National_ORD_Strategy_fr.pdf (swissuniversities.ch)
[15] Nouvelles priorités pour la Suisse numérique (admin.ch)
[16] Cf. Stratégie Suisse numérique [12] : 4.7.5 ; 4.8 ; 4.8.2 ; 4.9.2.
[17] Politique de la santé : stratégie du Conseil fédéral 2020–2030 (admin.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

La quantité, la diversité et la complexité (p. ex. en ce qui concerne la qualité des données) des données médicales de sources privées et publiques augmentent avec la **numérisation du domaine de la santé**. Il devient ainsi possible d'**analyser les problématiques dans l'ensemble du domaine des soins de manière plus large et approfondie** et d'y **répondre plus rapidement**. Une nouvelle ère s'est ouverte pour la recherche biomédicale : les frontières entre la recherche – le gain de connaissances scientifiques – et les soins de santé pratiques s'estompent de plus en plus avec l'utilisation de données dans ces deux buts [13]. Le nouveau **plan directeur « Mesures de la Confédération afin de renforcer la recherche et la technologie biomédicales**[18] **» 2020-2025**, élaboré sur mandat du Conseil fédéral du 14 décembre 2018 par le Département fédéral de l'intérieur (DFI) en collaboration avec le Département fédéral de l'économie, de la formation et de la recherche (DEFR) et le Département fédéral de justice et police (DFJP), soutient également cette évolution. La meilleure utilisation des données doit permettre une amélioration durable de la qualité des résultats des soins de santé dans les structures appropriées et les processus tout au long de la chaîne thérapeutique. Les nouvelles données, informations, connaissances et prestations génèrent une plus-value pour tout un chacun. Cela correspond à l'objectif d'un système de santé apprenant, qui doit tenir compte des opportunités et des risques à la suite de la numérisation. Il faut également s'attendre à des changements qui ne sont pas directement visibles dans le domaine socioculturel. De nouvelles hiérarchies peuvent se créer et un accroissement des inégalités sociales est imaginable du fait de la discrimination, de l'optimisation personnelle et de l'individualisation à la place de la solidarité [14].

Cela montre la nécessité de piloter la gestion des données disponibles électroniquement de manière contrôlée. La **recommandation de l'OCDE sur la gouvernance des données de santé** (17 janvier 2017)[19] mentionne douze principes qui doivent guider les États membres dans l'élaboration d'un « cadre national de gouvernance des données de santé ». Sous le thème **« Health in the 21st Century : Putting Data to Work for Stronger Health Systems »**[20], les États membres de l'OCDE – dont la Suisse – se sont en outre prononcés en 2019 pour le partage des données médicales existantes avec des tiers. Ils ont aussi discuté de la nécessité de réutiliser les jeux de données à d'autres fins que la destination initiale (utilisation secondaire) et de les exploiter de manière appariée [15]. À l'heure actuelle, peu d'États membres satisfont pleinement aux exigences recommandées en matière d'utilisation des données [16].

Si la Suisse entend mieux soutenir la comparaison internationale dans le cadre de la numérisation du système de santé, il est non seulement indispensable de rendre les données médicales disponibles pour une utilisation répétée à différentes fins, mais aussi que les données soient disponibles sous forme électronique et comparables. Dans le cadre de la **Stratégie Cybersanté Suisse 2.0**[21] pour les années 2018 à 2022, adoptée par le Conseil fédéral le 14 décembre 2018, la Confédération et les cantons s'engagent, dans le champ d'action « Coordonner la numérisation », pour une utilisation multiple des données/informations aussi bien médicales qu'administratives et des infrastructures (notamment les moyens d'identification) (objectifs B1 et B2). Pour ce faire, il est nécessaire de garantir l'interopérabilité technique et sémantique des systèmes informatiques (objectif B3, Stratégie d'interopérabilité) et de désigner un organe d'entretien des normes sémantiques (objectif B4). Le 8 avril 2021, eHealth Suisse a publié un **document de base pour des processus interopérables dans le système de soins de santé** [17]. Il est indispensable d'encourager la compréhension et les efforts des acteurs du système de santé pour renforcer l'échange de données harmonisées dans les processus thérapeutiques. D'un point de vue stratégique, il s'agit ainsi de contribuer à une utilisation multiple plus fréquente des données et des infrastructures, en tenant compte des points communs avec les processus de l'administration publique et de la recherche[22].

[18] Renforcer la recherche et la technologie biomédicales (admin.ch)
[19] OECD Recommendation on Health Data Governance - OCDE. Les principes concernent la coordination au sein du gouvernement et avec les organisations importantes, l'implication et la participation des différents acteurs, la transparence, l'information et les mécanismes de consentement pour les individus, le contrôle des systèmes de données médicales publics et l'encouragement du développement technologique pour l'utilisation des données, les procédures d'autorisation pour la recherche, la formation concernant la gestion de ces données et les mesures de contrôle.
[20] Health in the 21st Century : Putting Data to Work for Stronger Health Systems | en | OECD
[21] Stratégie Cybersanté Suisse 2.0 (admin.ch) et Stratégie Cybersanté Suisse 2.0 – 2018–2022 (e-health-suisse.ch)
[22] Rapport_intermediaire_strategie_eHealth_Suisse_2.0.pdf (e-health-suisse.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

La **Stratégie informatique de la Confédération 2020-2023**[23] adoptée par le Conseil fédéral le 3 avril 2020 se concentre sur les changements nécessaires pour adapter l'informatique de la Confédération aux besoins opérationnels futurs et pour soutenir la transformation numérique dans l'administration. Elle poursuit l'objectif d'**améliorer la gestion des informations, des données et des processus** (axe stratégique A) et prévoit notamment d'établir, à l'échelon de l'administration fédérale, un **réseau de compétences en science des données** qui développe systématiquement dans toute l'administration fédérale la gestion en matière de science des données[24] (*data science management*). Ce réseau cherche à collaborer avec les autorités opérant dans ce domaine et entretient des liens de coopération avec l'enseignement, la recherche et l'industrie. L'**Office fédéral de la statistique (OFS)** crée un **centre de compétences en science des données**[25] faisant partie du réseau (objectif A2-1 de la stratégie informatique), qui propose d'abord des services internes à l'administration en tant que prestataire, au sens de « *data science as a service* ». Il s'agit par exemple du suivi méthodologique dans la réalisation de projets appliqués externes ou internes des offices. À cet égard, l'OFS a renforcé la collaboration avec l'Université de Neuchâtel (UniNE)[26] début 2021 et conclu un partenariat stratégique avec le « Swiss Data Science Center » des deux écoles polytechniques fédérales (EPFZ et EPFL)[27].

En 2019, le **Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI)** a rédigé, à l'attention du Conseil fédéral, un rapport qui, sous le titre de **« Feuille de route suisse pour les infrastructures de recherche**[28] **»**, constitue une base de planification pour l'élaboration du message FRI 2021-2024. Sont également considérées comme des infrastructures de recherche les infrastructures scientifiques numériques d'information et de services, à l'instar des banques de données avec de très grandes quantités de données et des analyses de données complexes. La feuille de route tient concrètement compte de la planification d'**infrastructures et de services de données** bien coordonnés et facilement accessibles **pour un système scientifique ouvert**. Pour le **Fonds national suisse (FNS)**[29], l'amélioration de la qualité et de l'exploitabilité des données et des résultats de la recherche est également un objectif stratégique important qui est concrétisé dans sa politique en matière de science ouverte, avec les recommandations *Open Access to Publications*[30] et *Open Research Data*[31].

Le SEFRI s'engage en outre en faveur d'une **amélioration des conditions-cadres de la recherche clinique en Suisse**. Sur mandat du SEFRI, l'Académie suisse des sciences médicales (ASSM) a analysé les défis de la recherche clinique et formulé, dans un livre blanc publié en 2021, des champs d'action et des objectifs pour faire avancer le positionnement international de la recherche clinique centrée sur les patients [18]. Des mesures concrètes sont recommandées pour la période 2022-2024 dans le but de réduire la complexité des processus réglementaires et relatifs aux données, pour une mise en œuvre plus efficace et plus rapide de la recherche clinique.

1.3.2 Projets dans le domaine du traitement des données médicales

Au cours de la dernière décennie, le nombre de projets publics et privés dans le domaine de l'utilisation des données médicales a fortement augmenté. Pour le présent rapport, il est nécessaire de distinguer deux catégories de projets sur le fond. D'une part, les projets qui s'intéressent concrètement à l'exigence du postulat visant une meilleure utilisation des données et qui fournissent des constats pratiques importants pour répondre au postulat. S'agissant de ces projets directement pertinents, les projets de la Confédération et les projets de recherche financés par les pouvoirs publics sont considérés ci-après séparément. D'autre part, les projets qui abordent des questions connexes et dont les enseignements ne sont pas directement pertinents pour le présent rapport. La vue d'ensemble ci-dessous n'a aucune prétention à l'exhaustivité.

[23] Stratégie informatique de la Confédération 2020-2023 (admin.ch)
[24] Combinaison de la science des données et de l'économie/gestion d'entreprise : analyser / exploiter des données, concevoir et introduire des solutions numériques intelligentes pour les processus d'affaires
[25] Centre de compétences en science des données (DSCC) | Office fédéral de la statistique (admin.ch)
[26] Méthodes statistiques, science des données et intelligence artificielle : l'OFS et l'UniNE intensifient leur collaboration - Centre de compétences en science des données | Communiqué de presse | Office fédéral de la statistique (admin.ch)
[27] Les Écoles polytechniques fédérales et l'OFS alliés en science des données et intelligence artificielle pour le bien commun - Centre de compétences en science des données : nouvelle collaboration | Communiqué de presse | Office fédéral de la statistique (admin.ch)
[28] Feuille de route suisse pour les infrastructures de recherche 2019 (admin.ch)
[29] Stratégie (snf.ch) et http://www.snf.ch/SiteCollectionDocuments/mehrjahresprogramm_2021_2024_f.pdf
[30] Open Access to Publications (snf.ch)
[31] Open Research Data (snf.ch) ; Open Science (admin.ch) et vereinbarung-open-research-data-sign_d.pdf (en allemand)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

(1) Projets de la Confédération

Dans le cadre des mesures de maîtrise des coûts recommandées par un groupe d'experts dans le but de décharger l'assurance-maladie obligatoire[32], le Pr Ch. Lovis des Hôpitaux universitaires de Genève, mandaté par le DFI/OFSP en 2018, a élaboré avec un groupe de travail le **rapport Lovis** [19], dont le Conseil fédéral a pris connaissance en décembre 2019. Le rapport montre comment la gestion des données multi-dimensionnelles de l'assurance-maladie obligatoire devrait être améliorée (interopérabilité ; transmission des données et utilisation multiple à la place de relevés multiples) pour augmenter la transparence dans le système de santé et pouvoir mieux accomplir les tâches légales tout en réduisant les charges. Cela présuppose aussi des développements structurels pour une meilleure gestion des données.

En complément, le rapport donnant suite au **postulat CSSS-E 18.4102 « Stratégie de données cohérente pour le domaine de la santé »** s'intéresse aux possibilités d'une collecte de données adaptée aux besoins et optimisée en termes de charge de travail, dans le but de parvenir à une meilleure transparence dans le domaine de l'assurance obligatoire des soins (AOS).

Le rapport Lovis et les travaux relatifs au Po 18.4102 soutiennent par conséquent aussi la **motion 16.4011 « Numérisation. Éviter les récoltes de données en parallèle »** du groupe libéral-radical du 14 décembre 2016[33], acceptée par le Conseil fédéral le 22 février 2017. Le **« principe *once-only* »** (« une fois pour toutes ») vise à garantir que les entreprises ne doivent pas fournir les mêmes données et informations à plusieurs autorités (saisie unique, utilisation multiple). La coordination de l'OFS, des différents offices fédéraux, des cantons et des communes doit conduire à une baisse considérable de la charge de travail liée aux questions et aux contrôles pour les entreprises.

Dans le prolongement de sa séance spéciale du 20 juin 2018, le Conseil fédéral a chargé l'OFS de **développer le programme de gestion nationale des données (NaDB)**[34]. Celui-ci porte principalement sur les données administratives (Confédération, cantons et communes) et sur des questions relatives à la qualité et à la documentation de nouvelles sources de données non exploitées, qui sont souvent produites par des institutions privées. Une infrastructure coordinatrice et technique – la **plateforme d'interopérabilité I14Y** – doit améliorer les flux de données en tenant compte de la qualité, de la comparabilité et des possibilités d'appariement des données.

Pour le domaine statistique de la santé, l'OFS a lancé le **projet de mise en œuvre pilote du NaDB sur la gestion des données dans le secteur stationnaire hospitalier (SpiGES)** [20]. Ce projet tient compte de la mise en pratique des considérations du rapport Lovis.

L'OFSP et l'OFS s'engagent conjointement pour une **gestion coordonnée des données dans le domaine de la santé**[35]. L'annexe 2 (chap. 8) montre graphiquement la constellation des projets stratégiques précités et des mesures visant une meilleure utilisation des données dans l'administration fédérale.

Le 12 janvier 2022, le Conseil fédéral a pris connaissance du **« Rapport concernant l'amélioration de la gestion des données dans le domaine de la santé »** [21] élaboré par l'OFSP et chargé les départements concernés de lui rendre rapport sur les prochaines étapes jusqu'en juillet 2022[36]. Le rapport suit une approche globale de la gestion des données dans le domaine de la santé et présente **sept principes de la gestion des données** qui font également autorité pour le domaine de la recherche :

1) Il convient de séparer les discussions concernant les enjeux technico-sémantiques et les discussions concernant les enjeux politico-sociétaux.
2) Les solutions liées aux flux et à la conservation des données sont à considérer séparément de la finalité de leur utilisation.
3) Les travaux doivent être amorcés à partir d'un modèle commun.

---

[32] Nouvelles mesures visant à contenir les coûts de la santé (admin.ch) et fiche-d-information-rapport-du-groupe-d-experts-sur-la-maitrise-de-la-hausse-des-couts.pdf
[33] 16.4011 | Numérisation. Éviter les récoltes de données en parallèle | Objet | Le Parlement suisse
[34] Gestion nationale des données NaDB | Office fédéral de la statistique (admin.ch)
[35] Projet « Feuille de route Gestion des données dans le domaine de la santé dans le contexte global de la numérisation ».
[36] La gestion des données dans le domaine de la santé doit être améliorée (admin.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

4) Les définitions des différents acteurs sont à uniformiser (regroupements). Le recours à des définitions uniformes contribue à la mise en œuvre du principe *once only* et à l'utilisation multiple des données.
5) Des identifiants communs doivent être introduits.
6) Les données et les flux de données existants doivent être utilisés en priorité.
7) La mise en œuvre doit être réalisée par étapes individuelles concrètes.

Ces principes découlent des recommandations formulées dans le rapport Lovis [19] sur la gestion des données pour parvenir à plus de transparence dans le système de santé ainsi que des enseignements tirés de la mise en œuvre des mesures visant à surmonter la pandémie de COVID-19.

D'autres actions à l'échelle fédérale fournissent des informations fondamentales pour répondre au postulat :

- Le **rapport du groupe d'experts concernant le traitement et la sécurité des données**[37] du 17 août 2018, en application de la motion 13.3841[38] du 26 septembre 2013, avec des références à la gestion des aspects relevant du droit de la protection des données.
- Le **rapport d'évaluation** du Pr David Basin du 27 septembre 2017 **sur l'évaluation des risques d'une utilisation du numéro d'assurance sociale suisse NAVS13 dans différents domaines**, établi sur mandat de l'Office fédéral de la justice (OFJ) et du Préposé fédéral à la protection des données et à la transparence (PFPDT) [22].
- À la suite du **postulat 17.3968 « Concept de sécurité pour les identifiants des personnes**[39] **»**, le DFI a présenté un projet mis en consultation sur des modifications des dispositions de la loi fédérale sur l'assurance-vieillesse et survivants (LAVS ; RS *831.10*) qui règlent l'**utilisation systématique du numéro AVS dans les échanges avec les autorités** dans des conditions clairement définies[40]. La révision de la LAVS a été décidée par le Parlement le 18 décembre 2020[41]. Le domaine de la recherche n'est pas pris en considération.
- Sur mandat de l'OFJ et de l'Office fédéral de la communication (OFCOM), le Center for Information Technology, Society, and Law (ITSL) de l'Université de Zurich a procédé à un examen en juillet 2017 pour déterminer si le législateur doit agir pour encourager la **réutilisation des données personnelles sous le contrôle des personnes concernées**, en particulier dans le cadre des **systèmes de gestion des informations personnelles (PIMS)**[42] [23]. Le mandat est lié aux développements en Europe[43] et en Suisse dans le domaine de la protection des données. La **gestion du consentement**, c'est-à-dire la gestion de l'utilisation et des préférences de validation des données personnelles, fait partie des fonctions élémentaires du PIMS [24].

(2) Projets de recherche financés par les pouvoirs publics

En 2017, la Confédération (SEFRI) a lancé l'**Initiative nationale d'encouragement de la médecine personnalisée**[44] et chargé l'ASSM de sa mise en œuvre. L'initiative a pour but de permettre aux chercheurs l'accès à des données médicales cliniques de qualité. L'une des missions du **Swiss Personalized Health Network (SPHN)**[45] est de montrer comment l'infrastructure de données mise en place peut être exploitée durablement à l'échéance de l'initiative en 2024. Le SPHN s'engage, en tenant compte des hôpitaux universitaires et des hautes écoles, en particulier en coordination avec le champ d'action stratégique **Personalized Health and Related Technologies (PHRT)**[46] du domaine des

[37] Rapport du groupe d'experts concernant le traitement et la sécurité des données (admin.ch)
[38] 13.3841 | Commission d'experts pour l'avenir du traitement et de la sécurité des données | Objet | Le Parlement suisse
[39] 17.3968 | Concept de sécurité pour les identifiants des personnes | Objet | Le Parlement suisse
[40] 19.057 | LAVS. Modification (Utilisation systématique du numéro AVS par les autorités) | Objet | Le Parlement suisse et FF **2019** 7359 - Message relatif à la modification de la loi fédérale sur l'assurance-vieillesse et survivants (Utilisation systématique du numéro AVS par les autorités) (admin.ch).
[41] FF **2020** 9951 - Loi fédérale sur l'assurance-vieillesse et survivants (LAVS) (Utilisation systématique du numéro AVS par les autorités) (admin.ch)
[42] Exemple d'un PIMS dans le domaine de la santé en Suisse : MIDATA.coop
[43] Système de gestion des informations personnelles | European Data Protection Supervisor (europa.eu)
[44] Initiative nationale d'encouragement de la médecine personnalisée (admin.ch)
[45] SPHN - Swiss Personalized Health Network (SPHN)
[46] PHRT - Personalized Health and Related Technologies (sfa-phrt.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

EPF, pour une application de normes uniformes dans l'organisation des données, dans le but de rendre les différentes données compatibles (interopérables) et réutilisables pour des tiers. L'Institut suisse de bioinformatique (ISB) s'est vu confier la responsabilité du « centre national de coordination des données » avec l'infrastructure de soutien « BioMedIT », qui coordonne le développement et la mise en œuvre de la stratégie d'interopérabilité SPHN et soutient un échange de données cliniques conforme à la protection des données et correct d'un point de vue réglementaire. Comme il s'agit de données médicales sensibles, des recommandations sur des aspects éthiques, réglementaires et sociaux sont également formulées dans le cadre de l'initiative d'encouragement[47]. Le SPHN fournit ainsi des bases, des connaissances et des expériences importantes qui sont applicables à un large spectre de données médicales (cf. chap. 1.2.3).

La mise en place d'une **cohorte nationale de personnes en bonne santé à long terme avec biosurveillance humaine (BSH)**[48] progresse sous la houlette de l'OFSP. Grâce à l'appariement de données sur des marqueurs mesurés à plusieurs reprises dans des échantillons biologiques (p. ex. des polluants ou des oligoéléments) avec d'autres données de paramètres de santé mesurés ou demandés en même temps, ce projet doit contribuer à l'identification des déterminants de la santé – par exemple des facteurs environnementaux – qui influent ensemble ou séparément sur la santé de la population. En mai 2017, le Conseil fédéral a approuvé la réalisation, coordonnée par l'OFSP, d'une **étude pilote (2019-2021) pour la préparation d'une cohorte nationale de personnes en bonne santé**, sous le nom d'**« Étude suisse sur la santé »**[49]. La Swiss Biobanking Platform (SBP), une organisation nationale de coordination des biobanques humaines et non humaines, Unisanté Lausanne et des instituts de recherche à Berne et à Bâle sont les partenaires de l'OFSP. À partir de 2023, l'étude suisse sur la santé sera réalisée à l'échelle nationale.

Par ailleurs, la Confédération (SEFRI) soutient aussi le **projet de coopération et d'innovation Swiss Learning Health System (SLHS)**[50] de 2017 à 2024. Ce projet veut favoriser le dialogue entre les différents acteurs importants du système de santé afin de développer et d'intégrer constamment des solutions fondées sur des bases factuelles sur les défis actuels et futurs du système de santé. Parmi ces défis, il y a la **gestion d'informations standardisées sur la santé** avec les éléments suivants : la standardisation des informations de santé, la banque de données pour les informations de santé (*data warehouse*) et le laboratoire du système de santé (*health system lab*)[51]. Dans ce cadre, les partenaires de l'Université de Neuchâtel ont élaboré des *SLHS-Policy Briefs* sur les principes de la création d'archives centrales de métadonnées [25] et sur des recommandations de mise en œuvre correspondantes [26].

Le **programme national de recherche PNR 74 « Système de santé »** (2015-2023)[52] veut promouvoir des recherches innovantes dans le domaine des services de santé qui contribuent à faire face aux défis concrets dans la prise en charge des malades chroniques en Suisse. C'est pourquoi il poursuit également l'objectif d'optimiser à plus long terme **la disponibilité, l'accessibilité et l'appariement des données médicales**[53]. Plusieurs des projets soutenus ont initié des regroupements inédits de données. Ces projets livrent des indications pratiques sur les limites de la faisabilité technique et juridique dans le cadre de la réutilisation et de l'appariement de données, à l'instar de l'étude SIHOS[54] [27, 28]. Un projet évalue à quelles conditions il serait possible de regrouper des données médicales de différentes sources, par exemple des établissements de soins, des assureurs et des autorités[55]. Le **besoin de coordination de la pratique de la recherche et du droit** est mis en évidence pour pouvoir faciliter la réutilisation des données médicales tout en satisfaisant aux exigences éthiques et légales (thème du consentement) [29].

Les expériences pratiques concernant l'accès, le regroupement et l'appariement de données de différentes sources, notamment dans les projets du PNR 74, ont conforté un réseau d'institutions et

[47] ELSI Advisory Group (ELSIag) - SPHN
[48] Projets de biosurveillance humaine en Suisse (admin.ch)
[49] Étude | Étude suisse sur la santé (etude-sur-la-sante.ch)
[50] Platform for health systems and services research | SLHS, financé par le SEFRI dans le cadre des contributions liées à des projets au sens de la LEHE
[51] The Project | Swiss Learning Health System (slhs.ch)
[52] Accueil (nfp74.ch)
[53] Plus de transparence et de qualité grâce à de meilleures données de santé (nfp74.ch)
[54] 04 Les inégalités sociales dans les soins stationnaires en Suisse (nfp74.ch)
[55] 10 Promouvoir la compilation de données médicales en Suisse (nfp74.ch)

d'équipes de recherche, sous la direction de la fondation FORS (Centre de compétences suisse en sciences sociales), dans leur décision de créer l'**initiative stratégique linkhub.ch**[56] fin 2019. Le but est de faciliter l'accès et l'**appariement de données, principalement à des fins de recherche**, tout en préservant les principes juridiques et scientifiques [30].

Cette initiative pourrait gagner en importance de le cadre des nouveaux **programmes nationaux de recherche PNR 78 « COVID-19 »** (2020-2023)[57] et **PNR 80 « COVID-19 et société »** (dès 2021)[58], dans la mesure où il est nécessaire de compléter des données individuelles de projets à orientation biomédicale et clinique par des données de projets en sciences sociales et humaines.

(3) Évolutions techniques et sociales dans la gestion des données

Les projets recensés ci-dessous ne sont pas directement pertinents pour répondre au postulat. Ils fournissent toutefois des informations intéressantes dans le contexte de la numérisation en cours et donc pour les développements qui découlent de ce rapport.

- Le **logiciel d'enregistrement (LE) pour les registres des tumeurs** a des interfaces avec la Centrale de compensation (CdC), l'OFS, les services cantonaux des habitants et l'Organe national d'enregistrement du cancer (ONEC). Il est aujourd'hui exclusivement développé et utilisé pour l'enregistrement du cancer, mais pourrait être développé comme **« modèle de collecte de données »** pour d'autres applications. Un tel modèle pourrait être utilisé dès qu'il est nécessaire de déclarer et d'enregistrer des données structurées de différentes sources de manière centralisée ou nationale.
- Les projets de recherche menés dans le cadre du **programme national de recherche PNR 75 « Big Data »** (2015 à 2023)[59] doivent fournir les bases scientifiques pour une utilisation efficace et appropriée de grandes quantités de données. Ils étudient des questions liées à **la technologie de l'information et ses incidences sociales** dans des domaines d'application concrets. Le projet « BEHALF - Bigdata-Ethics-HeaLth Framework »[60] s'intéresse de manière prospective aux défis qui apparaîtront avec l'augmentation des **tendances au *big data* dans le domaine de la santé** [31]. L'idée de « *big (biomedical) data* » y prend peu à peu forme, notamment parce que les données médicales bénéficient du statut spécial de données sensibles [32, 33].
- Le **programme national de recherche PNR 77 « Transformation numérique »**[61] (2018-2026) étudie les interactions et les conséquences concrètes de la transformation numérique en Suisse dans plusieurs domaines. Il s'agit de générer des connaissances sur les **opportunités, risques, défis et solutions liés à la numérisation dans le contexte suisse**. La mise en œuvre se fait par le biais de trois modules : (1) formation, apprentissage et tournant numérique, (2) **éthique, fiabilité et gouvernance**, (3) économie numérique et marché du travail.
- La **Stratégie suisse de cyberadministration 2020-2030**[62] précise notamment que la disponibilité d'une identité électronique reconnue par l'État (e-ID) est essentielle pour le succès d'une cyberadministration. De manière générale, une preuve e-ID permet aux habitants de la Suisse de justifier de leur identité en ligne au moyen d'une preuve numérique. À la suite du rejet de la loi sur les services d'identification électronique (LSIE ; affaire du CF 18.049)[63] en votation populaire le 7 mars 2021, une **consultation publique** a été menée à l'automne 2021 **sur une future solution d'e-ID suisse**, à laquelle ont également pris part la science/recherche et la société civile. La question de **l'utilité, des cas d'application et des exigences posées à une e-ID officielle** a

[56] linkhub.ch
[57] Portrait – PNR 78 – Covid-19
[58] Nouveau programme national de recherche « COVID-19 et société » (admin.ch)
[59] Accueil (nfp75.ch)
[60] Un cadre éthique pour l'utilisation de mégadonnées présentant un intérêt dans le domaine de la santé (nfp75.ch)
[61] Portrait – PNR 77
[62] Stratégie suisse de cyberadministration - www.egovernment.ch
[63] 18.049 | Loi sur les services d'identification électronique | Objet | Le Parlement suisse et Identité électronique : la loi sur l'e-ID (admin.ch) ; E-ID étatique (admin.ch) et Décision de principe du Conseil fédéral sur l'e-ID (admin.ch)

été débattue. Le 17 décembre 2021, le Conseil fédéral a pris sur cette base une décision de principe sur l'e-ID et fixé les principes pour la conception d'une future identification officielle. La procédure de consultation relative à la nouvelle loi sur l'e-ID devrait être lancée à l'été 2022.

- Le 18 avril 2018, le Conseil fédéral a adopté la **stratégie nationale de protection de la Suisse contre les cyberrisques (SNPC) 2018-2022**[64] (sous la responsabilité de l'Unité de pilotage informatique de la Confédération [UPIC]), qui repose sur les travaux réalisés dans le cadre de la première SNPC (2012-2017). Élaborée en collaboration avec les milieux économiques, les cantons et les hautes écoles, cette stratégie constitue la base permettant les efforts communs requis afin de réduire les cyberrisques. La dépendance croissante vis-à-vis des technologies de l'information et de la communication (TIC), qui découle de la numérisation, rend la Suisse vulnérable aux défaillances, dysfonctionnements et abus de ces technologies. Les **menaces en matière de protection des données** se sont aggravées depuis quelques années en raison des évolutions technologiques et sociales en cours. La **protection des infrastructures nécessaires**, dont les réseaux et les services de base **qui servent au traitement des données**, relève de l'intérêt national de la Suisse.

1.4 Objectif du rapport

Ce rapport a pour objet principal de présenter une proposition pratique sur les possibilités d'une utilisation commune et d'une exploitation appariée des données médicales de différentes sources de tiers à des fins de recherche.

Pour parvenir à cet objectif, il creuse les questions centrales suivantes aux chapitres 2 à 4 :

- Quel est le potentiel d'une réutilisation et d'une exploitation appariée des données médicales pour le domaine de la santé ? (motivation)
- Quelles conditions doivent être remplies pour que les données médicales puissent et soient autorisées à être réutilisées et traitées par différents groupes d'utilisateurs ? (contexte, besoins)
- Quels développements techniques, sémantiques, organisationnels et juridiques faut-il initier pour remplir ces conditions dans la pratique ? (solution)

Le rapport livre donc aussi des réponses aux interventions parlementaires connexes des années 2019 à 2021 mentionnées au chap. 1.2.2.

Bases du présent rapport

Les informations suivantes ont été utilisées pour apporter une réponse pratique à ces questions :

- Développements actuels et connaissances des projets et initiatives en cours (cf. chap. 1.3) et des publications spécifiques.
- Les résultats d'une enquête en ligne réalisée par l'OFSP en collaboration avec l'OFS sur le thème de la réutilisation des données, avec un intérêt particulier pour l'appariement. Les producteurs de données/responsables[65] et les utilisateurs de données interrogés ont fourni des connaissances sur les défis et les solutions possibles à partir de projets d'appariement partiellement ou entièrement mis en œuvre en Suisse [34].

[64] https://www.isb.admin.ch/isb/fr/home/ikt-vorgaben/strategien-teilstrategien/sn002-nationale_strategie_schutz_schweiz_cyber-risiken_ncs.html
[65] Cf. FF **2020** 7397 - Loi fédérale sur la protection des données (LPD) (admin.ch), art. 5, let. j. *Responsable du traitement* : la personne privée ou l'organe fédéral qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement de données personnelles. Let. d. *Traitement* : toute opération relative à des données personnelles, quels que soient les moyens et procédés utilisés, notamment la collecte, l'enregistrement, la conservation, l'utilisation, la modification, la communication, l'archivage, l'effacement ou la destruction de données

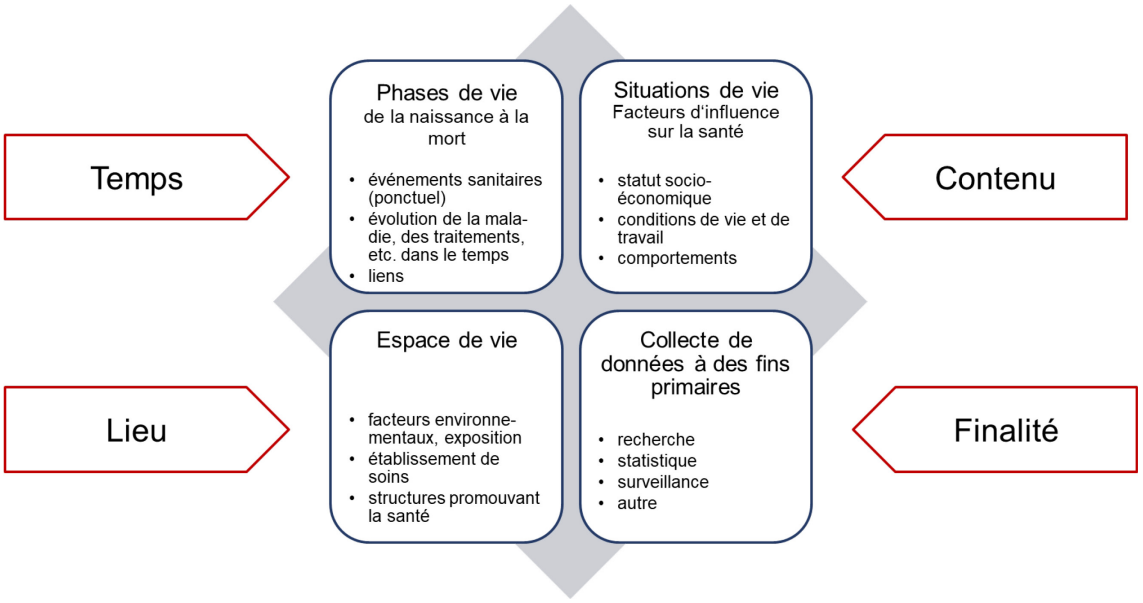
2 Potentiel de la réutilisation et de l'appariement de données médicales pour le domaine de la santé

Comme le rapport l'a expliqué en introduction, la collecte, la réutilisation et l'exploitation appariée des données médicales fournissent les bases de décision pour les optimisations et les innovations nécessaires pour des soins de santé efficients et de qualité. L'utilisation multiple de données médicales à des fins de recherche (voir les précisions conceptuelles au chap. 1.2.3) est judicieuse, car les mêmes données peuvent être utilisées le plus souvent pour répondre à plusieurs questions relatives à la santé. Les nombreuses données disponibles ont été collectées sur différents thèmes très spécifiques, à des moments différents, concernant différents lieux/espaces de vie et à des fins différentes. Si ces « distances » existantes entre les données/collectes de données sont surmontées et si l'utilisation commune des données est conçue plus efficacement, il est possible d'obtenir de manière répétée, à partir des mêmes sources de données, davantage de connaissances précieuses pour la société. Les connaissances résultant de l'utilisation multiple des données permettent d'aider les gens à vivre plus sainement et plus longtemps.

2.1 Surmonter les « distances » existantes

La réutilisation et l'exploitation appariée des données médicales permettent de surmonter les distances des données et des collectes de données en termes de temps, de contenu, de lieu et de finalité. La figure 1 le montre de manière générale, la figure 2 à l'aide d'exemples d'application.

Figure 1 Distances surmontables grâce à la réutilisation et à l'appariement des données



Temps

La vie de chaque personne est marquée par des événements sanitaires isolés ou répétés. Des maladies peuvent évoluer différemment sur le plan individuel au fil du temps, nécessiter différentes méthodes thérapeutiques qui peuvent être prescrites, choisies ou sollicitées. Les événements sanitaires au cours de la vie peuvent être directement ou indirectement liés. L'exploitation commune des données se rapportant à des personnes, collectées à différents moments et en différents lieux (sources des données), permet de compléter efficacement des études à long terme complexes et onéreuses.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

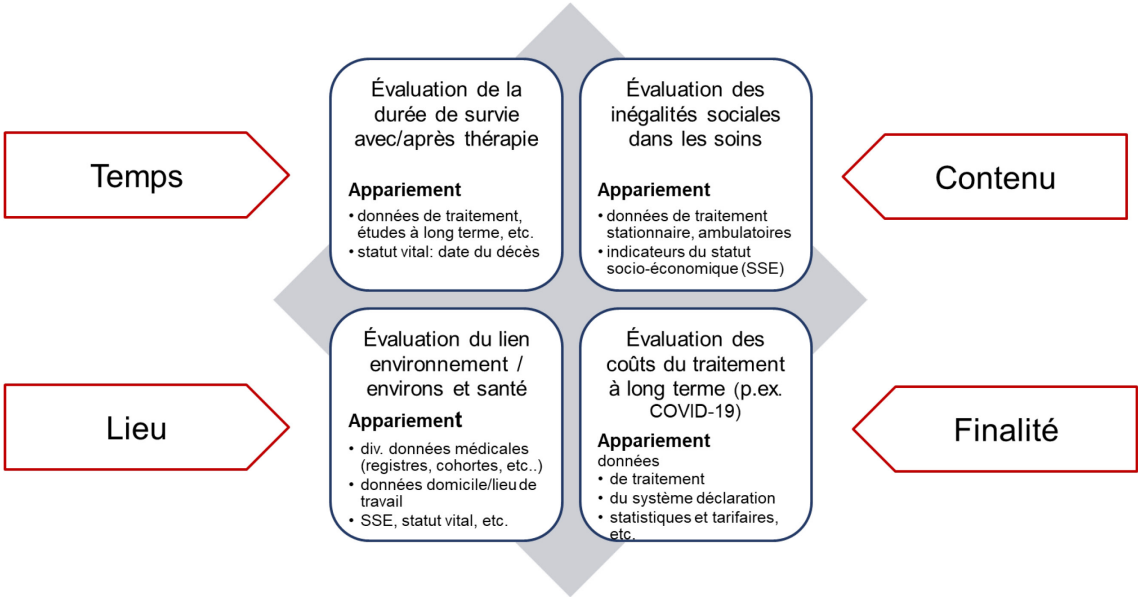
Contenu et lieu

L'évolution de l'état de santé dans le temps peut être influencé et expliqué par plusieurs facteurs qui concernent les situations de vie (p. ex. niveau de formation, rapport de travail, situation financière, consommation de tabac) et les espaces de vie (facteurs environnementaux/exposition au domicile et au travail, proximité d'établissements/offres de soins, offres de sport, offres de restauration, etc.). Ces facteurs peuvent changer pour l'individu au cours de sa vie et être différents selon les groupes de personnes. La prise en compte de ces informations provenant de sources de données existantes aide à surmonter de telles distances en termes de contenu et de lieu et permet de comprendre l'impact de facteurs spécifiques sur la santé et de pouvoir prendre des décisions relatives à la politique de la santé qui sont axées sur la pratique (les soins). Le gain de connaissances bénéficie à la population grâce à de meilleures offres de soins (prévention, diagnostic et traitement).

Finalité

La possibilité de s'appuyer sur un large spectre de données est garantie en permettant un écart contrôlé par rapport à la finalité des données (finalité primaire) dans le cadre de l'utilisation secondaire des données. Comme le montre la figure 2 à titre d'exemple, les données collectées initialement à des fins médicales, statistiques ou de surveillance peuvent être utilisées de manière appariée à titre secondaire à des fins de recherche pour évaluer les coûts d'un traitement à long terme des patients atteints du COVID-19.

Figure 2 Exemples d'application pour les distances surmontables grâce à la réutilisation et à l'appariement des données



À l'aide d'exemples d'application formulés de manière générale, la figure 2 explique le potentiel de l'utilisation multiple des données existantes se rapportant à des personnes, collectées de manière systématique et étendue par différents services, c'est-à-dire par des experts de qualité. Seules les données absolument nécessaires qui n'existent pas encore devraient être collectées à l'avenir à titre de complément. Cette exigence se justifie aussi par l'urgence des questions de la recherche motivées par la politique de la santé et pour des questions de ressources.

2.2 Amélioration de l’efficience

Augmenter la flexibilité

L’utilisation multiple des données médicales implique aussi un gain d’efficacité grâce à une flexibilité accrue. L’évaluation de l’adéquation, de l’efficacité et de l’économicité du système de santé est complexe et soulève toujours des questions de recherche aussi bien prévisibles qu’inattendues. La réutilisation et l’exploitation appariée des données médicales existantes permettent de fournir l’étendue et la profondeur nécessaires de l’information dans un délai utile et de reproduire également les résultats des études. En même temps, cette approche garantit une réactivité telle qu’elle est nécessaire en situation de crise (p. ex. pandémie de COVID-19). En fonction de la situation, il est nécessaire de pouvoir réagir à des questions inédites urgentes ou qu’en partie prévisibles.

Optimiser l’utilisation des ressources

L’utilisation multiple des données médicales génère aussi une plus-value grâce à la préservation des ressources.

D’une part, les personnes concernées (fournisseurs de données) ne sont pas sollicitées à plusieurs reprises par différents acteurs pour la collecte de la même information. Mais elles ont connaissance de l’utilisation multiple de leurs données collectées une seule fois et bénéficient du savoir et des progrès en matière de santé qui en ont découlé.

D’autre part, les producteurs de données/responsables exploitent mieux la valeur informative des données collectées le cas échéant à des fins et des moments différents en partageant les données avec des tiers. Ils permettent de répondre à de nouvelles questions avec des charges réduites (temps, personnel et finances).

Les utilisateurs secondaires de données évitent la collecte de données primaires, coûteuse et pénible pour les personnes concernées. Ils fournissent de nouvelles connaissances sur la base d’un ou plusieurs jeux de données appariés qui peuvent être réutilisés à leur tour.

Une réorientation dans l’utilisation des ressources financières par les producteurs de données et les utilisateurs secondaires de données génère aussi une plus-value. L’utilisation multiple des données existantes implique une diminution des moyens nécessaires pour les nouvelles collectes de données.

3 Conditions pour la réutilisation et l’appariement de données médicales

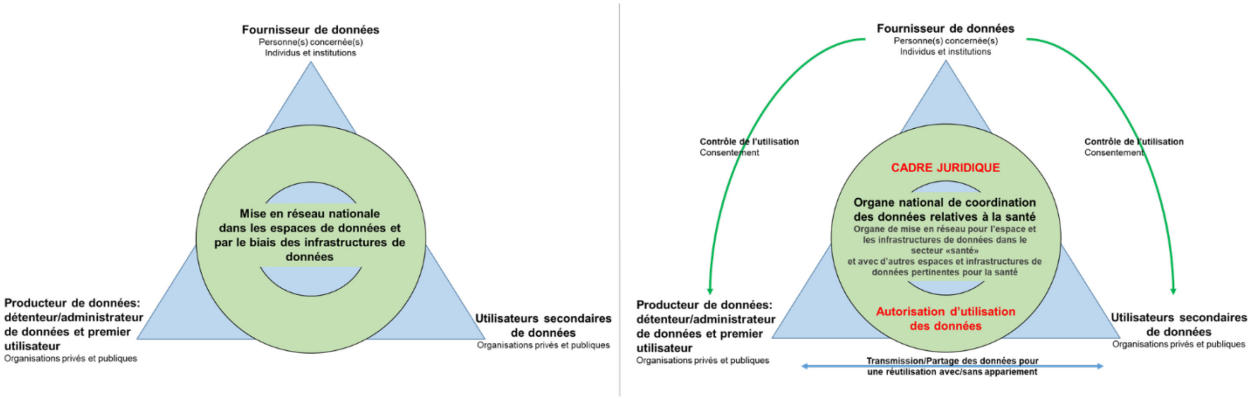
3.1 Développement de l’espace de données dans le secteur de la santé

Un espace de données se définit par la structure des relations entre les acteurs des données, c’est-à-dire les personnes concernées, les producteurs de données/responsables et les utilisateurs secondaires de données (voir la figure 3, partie gauche). Dans un espace de données, les trois groupes peuvent collaborer de manière équitable et autonome sur la base de droits en matière de données prévus par la loi et de principes éthiques, de processus et d’infrastructures fiables [35].

L’espace de données dans le secteur de la santé (désigné ci-après comme espace de données médicales) doit être développé (cf. chap. 1.3.1, Stratégie Suisse numérique [12]).

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Figure 3 Conception de l'espace de données dans le secteur de la santé



Cela concerne des aspects techniques, sémantiques, organisationnels et juridiques, mais aussi l'émergence d'une culture commune de la réutilisation des données, où les charges et les revenus devraient être répartis également entre les trois groupes [33, 36]. Dans ce contexte, il est nécessaire de clarifier non seulement les liens entre les acteurs des données, mais aussi leurs liens avec les données. Tant que la question se pose de savoir s'il y a une « propriété des données » ou pour qui, et ce que cela implique exactement, les aspects du contrôle ou des droits de contrôle et des droits d'utilisation seront au cœur de la discussion [37–40]. Une possibilité serait de comprendre les données médicales comme une propriété partagée (bien collectif) dans l'espace de données, après le consentement de la personne concernée, et de coordonner l'échange de données à l'échelle nationale en suivant un cadre juridique (autorisation d'utiliser les données) (voir la figure 3, partie droite) [37].

Pour que la collaboration entre les acteurs puisse fonctionner de manière ordonnée et efficace dans l'espace de données médicales, des conditions importantes doivent être remplies concernant les données et les processus qui s'y rapportent. Dès la production, l'enregistrement et l'archivage des données, il se décide si les données médicales se prêtent à une réutilisation et à un appariement avec d'autres données, si elles sont disponibles et si l'accès sera possible. De manière générale, plus une réutilisation et un appariement de données sont coûteux et exigeants, plus les organisations privées et publiques sont incitées à produire de nouvelles données au lieu de recourir aux données existantes.

3.2 Conditions-cadres optimales pour l'application des principes FAIR

Les principes FAIR[66] précisent que les jeux de données devraient être faciles à trouver (*findable*), accessibles (*accessible*), compatibles (*interoperable*) et réutilisables (*re-usable*). Ils s'appliquent aux données, mais doivent aussi être applicables dans la mise en œuvre pratique de toutes les stations du cycle de vie des données[67]. Cela n'est possible qu'en garantissant les conditions-cadres nécessaires en l'espèce.

3.2.1 Facilité à trouver des données réutilisables à l'aide de catalogues de métadonnées

Le paysage des données médicales en Suisse est complexe (cf. chap. 1.2.3) et caractérisé par un excédent de données. Les données sont produites en permanence sur les individus malades ou en bonne santé, les ménages, les entreprises et les charges/coûts. Il convient d'éviter des collectes répé-

[66] FAIR Principles - GO FAIR (go-fair.org), Les principes de données FAIR | KRAFT11 (force11.org) et Die FAIR Data Prinzipien für Forschungsdaten - TIB-Blog sur l'importance et les rôles pour les scientifiques et les répertoires
[67] Le cycle de vie des données comprend toutes les stations par lesquelles les données peuvent passer : production des données > traitement des données > analyse des données > archivage des données > accès aux données > réutilisation des données (utilisation multiple) > production des données (p. ex. à la suite de l'appariement avec d'autres jeux de données)

tées et en partie hétérogènes des mêmes informations auprès des personnes physiques et des institutions concernées dans le domaine de la santé [19]. C'est pourquoi il est essentiel de connaître l'existence des sources de données, y compris les informations sur leur composition, leur qualité et leurs possibilités d'utilisation.

Pour que les données soient faciles à trouver, il faut des catalogues de métadonnées où les informations sur les collections de données médicales en Suisse sont systématiquement rassemblées et documentées de manière standardisée. Les métadonnées cataloguées offrent aux utilisateurs une base de décision en les renseignant sur le contexte et la qualité des collections de données existantes. Les utilisateurs de données peuvent ainsi déterminer si des données existantes sont adéquates pour répondre à leur question spécifique et quelles dispositions d'accès de nature juridique, éthique et technique sont applicables. Un grand défi dans la mise en place des catalogues de métadonnées consiste par conséquent à rendre compte des multiples exigences posées aux collections de données en ce qui concerne l'utilisation, les buts ou les utilisateurs.

3.2.2 Respect des exigences qualitatives posées aux données qui peuvent être réutilisées et appariées

Par qualité, on entend les propriétés et les caractéristiques d'une collection de données qui signalent son adéquation pour une réutilisation ciblée et le cas échéant un appariement avec d'autres collections de données [41]. La qualité des données contenues dans une collection est jugée différemment du point de vue de l'utilisateur individuel. Les exigences peuvent varier en fonction de l'utilisation envisagée. Si les exigences posées par les utilisateurs à la qualité des données disponibles ne sont pas ou pas suffisamment remplies [13, 34], certains utilisateurs renoncent à réutiliser les données. En lieu et place, il faut s'attendre à la collecte multiple de données similaires.

La crédibilité[68], l'utilité[69] et l'interprétabilité[70] des données sont au cœur de l'évaluation [41]. En ce qui concerne ce dernier critère, l'enquête de l'OFSP a révélé le fait suivant : les utilisateurs de données ne peuvent pas être certains que les jeux de données qui leur sont fournis sont affichés de manière constamment uniforme et clairement compréhensible pour les utilisateurs spécialisés [34]. D'après leur estimation, les utilisateurs de données consacrent près de 80 % du travail et du temps à l'épuration et à la préparation des collections de données en cas de réutilisation des données avec appariement préalable [34]. Il est donc nécessaire de promouvoir l'harmonisation des collections de données. Toutes les mesures doivent être prises pour que les données de différentes collections de données soient comparables, c'est-à-dire qu'elles soient saisies selon les mêmes normes et que leur combinaison et leur appariement soient ainsi plus efficaces.

Normes pour l'échange interopérable de données

La réutilisation durable de données n'est possible que si les données parviennent effectivement et efficacement du producteur de données/responsable à l'utilisateur secondaire. Cela présuppose une interopérabilité[71]. Il s'agit de la capacité, pour des systèmes informatiques indépendants, de travailler ensemble autant que possible sans rupture médiatique pour mettre à la disposition de l'utilisateur des données ou des informations de manière électroniquement exploitable, sans nécessité d'accord spécial entre les systèmes. Pour cela, il est important de respecter des normes communes. Elles sont la base de l'échange interopérable de données, car elles permettent à l'émetteur et au destinataire de se comprendre sous l'angle de la technique et du contenu. Sur le plan technique, les interopérabilités syntaxique et sémantique sont déterminantes.

[68] correct/peu d'erreurs, cohérent, fiable/compréhensible
[69] complet, exact, sans redondances, actuel, pertinent
[70] uniforme, clair, compréhensible
[71] Il s'agit d'un dispositif multidimensionnel dans lequel sont impliqués toujours plus d'organisations/personnes, leurs systèmes d'information (TIC) et applications logicielles, une infrastructure d'échange de données et d'informations ainsi que l'information qui doit être transmise. On distingue les dimensions de l'interopérabilité politique, organisationnelle, technique, syntaxique et sémantique. Pour de plus amples explications, voir Glossaire - eHealth Suisse (e-health-suisse.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

- L'**interopérabilité syntaxique** garantit que les données puissent être transmises et traitées correctement entre les différents systèmes. Les normes syntaxiques définissent la structure de l'information, comme par exemple les normes XML (Extensible Markup Language), CDA (Clinical Document Architecture) ou FHIR® (Fast Healthcare Interoperability Resources).
- L'**interopérabilité sémantique** est la capacité à transmettre la signification technique des informations. Des normes sémantiques sont utilisées pour que les systèmes récepteurs reconnaissent et comprennent la teneur des données transmises, en utilisant une terminologie (ou classification) étendue relative à la santé, et rendues faciles à trouver et accessibles par le biais de catalogues de métadonnées.
  Voici des exemples dans le domaine médical : les classifications CIM-10 (Classification statistique internationale des maladies et des problèmes de santé connexes 10), LOINC (Logical Observation Identifiers Names and Codes, noms et identificateurs pour désigner les résultats d'analyses et de tests de laboratoire et de clinique) et SNOMED/SNOMED CT (Systematized Nomenclature of Human and Veterinary Medicine – Clinical Terms).

Identificateur personnel comme variable d'appariement

Un appariement de données médicales au niveau de l'individu (cf. chap. 1.2.3 et annexe 1, chap. 8) n'est réalisable qu'en partie au prix de grands efforts, faute de caractéristiques précises pour identifier les personnes [34]. L'annexe 3 (cf. chap. 8) explique les défis auxquels sont confrontés les utilisateurs secondaires de données à l'aide d'exemples de projets.

D'une part, il manque en Suisse une base légale pour un identificateur univoque librement utilisable dans le cadre des appariements. D'autre part, seuls les jeux de données anonymisés peuvent parfois être transmis du fait des réglementations légales en vigueur (cf. chap. 3.2.4). Même s'il est possible de transmettre des données codées, un appariement déterministe avec une attribution univoque de plusieurs jeux de données à une personne est pratiquement exclu, car les producteurs de données/responsables n'utilisent généralement pas le même pseudonyme – au sens d'un identificateur univoque individualisé. Pour un appariement probabiliste, réalisé à l'aide de plusieurs caractéristiques personnelles, le codage devrait être réversible, mais avec une violation de la protection de la personnalité.

Pour mieux pouvoir exploiter le potentiel des données existantes dans le contexte de l'utilisation multiple des données et de l'appariement, les producteurs de données/responsables privés et publics ont besoin d'un identificateur univoque.

3.2.3 Permettre l'accès aux données

Les utilisateurs de données sont en principe confrontés à des obstacles dans la planification et la réalisation de projets d'appariement en raison de la difficulté ou de l'absence d'accès à d'autres sources de données. En plus des aspects systémiques ainsi que techniques et organisationnels, une insécurité concernant les conditions d'accès et d'utilisation (cf. chap. 3.2.4) contribue également à cette appréciation [13, 33, 34].

Dans un paysage suisse de la recherche caractérisé par la concurrence, il est nécessaire d'examiner d'autres bases et mécanismes d'incitation pour que les producteurs de données/responsables soient disposés à accorder l'accès aux données qu'ils administrent à d'autres utilisateurs. Dans le cadre du PNR74 « Système de santé » (cf. chap. 1.3.2), les parties prenantes ont mentionné plusieurs approches en lien avec les mesures existantes en matière de données ouvertes et le développement de la gestion des données dans le domaine de la santé. Il s'agit par exemple [13, 33] :

- de régler le contrôle scientifique ou juridique des données produites,
- de définir des prescriptions détaillées pour la transmission et la réutilisation, y compris l'appariement de données médicales.

Dès que le rapport de synthèse du PNR74 sera disponible, il sera possible d'examiner l'applicabilité de ses recommandations générales.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

La mise en œuvre des systèmes et des stratégies d'*Open Government Data* (OGD)[72] et d'*Open Research Data* (ORD)[73] dans le domaine de l'administration et de la recherche vise une plus grande ouverture en ce qui concerne les données et les contenus, caractérisée par un accès libre, la réutilisation et la diffusion des données. L'idée des données ouvertes ne porte toutefois *pas* sur des données se rapportant à des personnes ou des données qui contiennent des secrets d'entreprises ou des secrets commerciaux. À cet égard, les conditions pour une réutilisation avec un appariement des données (voir aussi le chap. 3.2.2) ne sont pas pleinement réunies.

D'un point de vue technique et organisationnel, il est nécessaire d'améliorer les conditions pour que les producteurs de données/responsables puissent donner à d'autres personnes ou organisations l'accès à leurs collections de données en vue d'une réutilisation et d'une exploitation appariée. Pour ce faire, il faut un modèle praticable d'enregistrement, d'archivage et de transfert des données en toute sécurité. Jusqu'à présent, de nombreux producteurs individuels de données stockent et analysent leurs collections de données dans des silos. Seuls des personnes ou des groupes d'utilisateurs définis ont accès à ces supports de données isolés au sein d'une organisation.

3.2.4 Bases légales pour l'accès aux données et l'utilisation des données

Remarque liminaire : objet des explications juridiques ci-dessous

Le présent rapport ne comprend aucune appréciation juridique définitive des approches esquissées. Il montre plutôt des orientations et des solutions possibles, de nature factuelle ou technique, qui décrivent le processus de réutilisation et d'appariement de données médicales. Il est évident que de nombreuses questions juridiques (en matière de protection des données) se posent dans ce cadre, qui ne pourront être clarifiées qu'au cours des prochaines étapes de la concrétisation, au regard de l'intégration dans l'actuelle structure de normes et d'éventuelles adaptations. Il s'agit aussi de définir les responsabilités en matière de protection des données. C'est pourquoi les conditions-cadres du droit de la protection des données et de la recherche ne sont notamment présentées ci-après que sous une forme générale et fragmentaire, sans émettre d'appréciations concrètes et juridiquement définitives ; celles-ci seront l'objet de clarifications juridiques approfondies dans le cadre des prochaines étapes du projet en ce qui concerne les options à concrétiser.

Problématique

Les règles juridiques à respecter en cas d'accès et d'utilisation multiple de données médicales sont étroitement liées au thème de l'accès aux données. La transmission des données du producteur/responsable à un utilisateur secondaire ne peut se faire – en observant des normes de sécurité reconnues – que si les deux parties se sont entendues sur l'ensemble des aspects éthiques et des dispositions légales en lien avec le traitement des données médicales qui se rapportent en grande partie à des personnes. Dans le contexte actuel, il est nécessaire de fixer les conditions d'accès et d'utilisation dans un accord pour le transfert et l'utilisation des données (*Data Transfer and Use Agreement [DTUA]*) [42].

L'importance des aspects éthiques et juridiques ou des obstacles correspondants pour un traitement multiple et apparié des données est souvent sous-estimée [13]. Les bases légales historiques s'avèrent complexes avec de nombreux éléments individuels [29]. Il est donc évident que le projet décrit dans le cadre du présent rapport nécessiterait des adaptations juridiques à plusieurs niveaux.

Les bases légales et principes en vigueur façonnent l'utilisation secondaire des données médicales très exigeant voire impossible pour les producteurs de données/responsables et les utilisateurs de données des domaines de la recherche et de l'administration, en particulier lorsque la réutilisation présuppose un appariement des données [29, 33, 34].

L'annexe 4 (chap. 8) esquisse plus en détail les bases légales en vigueur pour la réutilisation et l'appariement des données. Il est essentiel que le traitement des données liées à la santé se rapporte au

[72] Stratégie « Open Government Data » 2019–2023 | Office fédéral de la statistique (admin.ch) et opendata.swiss
[73] Open Research Data - swissuniversities, Open Research Data (snf.ch) et Open Science (admin.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

droit fondamental à la protection de la sphère privée et à l'autodétermination informationnelle des personnes concernées.

Les législations fédérales et cantonales sur la protection des données définissent les règles de base pour le traitement des données par les autorités et les particuliers. La LPD s'applique au traitement de données personnelles par des personnes privées et des organes fédéraux (art. 2, al. 1, LPD), tandis que les lois cantonales sur la protection des données sont notamment applicables lorsque des données sont traitées par des autorités cantonales. Dans certains cas, il existe encore d'autres règles concernant le traitement des données dans d'autres lois.

Définitions selon la loi sur la protection des données et la loi relative à la recherche sur l'être humain

Comme le précise le chap. 1.2.3, les définitions selon la LPD (RS *235.1*) et la LRH (RS *810.30*) sont applicables en examinant les dispositions légales en vigueur.

Dans la législation fédérale, sont considérées comme des **données personnelles** au sens de la **LPD** toutes les informations qui se rapportent **à une personne identifiée ou identifiable** (art. 3, let. a). La révision totale de la loi sur la protection des données (nLPD[74]), qui entrera en vigueur au troisième trimestre 2023, reprend cette définition (cf. art. 5, let. a, nLPD) ; désormais, les données des personnes morales n'entrent toutefois plus dans son champ d'application.

- Une personne est identifiée lorsqu'il découle clairement des informations disponibles à qui les données se réfèrent.
- La personne est identifiable lorsqu'elle n'est pas directement identifiée par les informations disponibles, mais que les circonstances ou le contexte permettent de l'identifier.

L'élément déterminant pour évaluer la question de l'identifiabilité est de savoir si l'identité de la personne découle directement des données ou du contexte des données ou d'un appariement avec d'autres données[75].Dans ce dernier cas, la charge nécessaire pour l'appariement est déterminante. Elle est influencée par les moyens de la technologie moderne. Elle peut être aussi bien réduite (p. ex. par le biais d'outils de recherche et d'appariement) qu'augmentée (p. ex. par l'utilisation de puissantes méthodes de codage). Si le lien à une personne est supprimé de telle sorte à ne pouvoir être rétabli qu'au prix d'efforts disproportionnés, les données peuvent être considérées comme anonymisées et ne relèvent plus du champ d'application de la LPD.

Sont considérées comme des **données personnelles liées à la santé** au sens de l'art. 3, let. f, **LRH**, les « informations concernant une personne déterminée ou déterminable qui ont un lien avec son état de santé ou sa maladie, données génétiques comprises ». Il s'agit d'informations sur une personne qui ont un lien avec une maladie physique ou psychique ou qui donnent des renseignements sur la structure et le fonctionnement de la personne concernée. La notion de données liées à la santé correspond à celle des données sur la santé de la législation sur la protection des données (cf. art. 3, let. c, ch. 2, LPD)[76].

Les données sur la santé entrent dans la catégorie des données personnelles sensibles (art. 3, let. c, ch. 2, LPD). Mais le terme de « données médicales » n'est pas défini dans la LPD. D'aucuns suggèrent toutefois, ce qui est facilement concevable, que ce terme doit être compris aujourd'hui – notamment au regard des développements en matière de saisie, de réutilisation et d'appariement des données – plus largement qu'au moment de l'entrée en vigueur de la LPD en 1992[77].

---

[74] Nouvelle loi fédérale sur la protection des données (nLPD), adoptée le 25 septembre 2020, publiée le 6 octobre 2020 FF **2020** 7397 - Loi fédérale sur la protection des données (LPD) (admin.ch)
[75] Rudin Beat, Kommentierung Art. 3, in : Baeriswyl Bruno/Pärli Kurt (éd.), Stämpflis Handkommentar Datenschutzgesetz, Bern 2015, N 10 zu Art. 3.
[76] Cf. Message sur la LRH, 7310 FF **2009** 7259 8045 - Message sur la loi fédérale relative à la recherche sur l'être humain (admin.ch)
[77] Cf. Sprecher Franziska, Datenschutz und Big Data im Allgemeinen und im Besonderen im Gesundheitsrecht, ZBJV 154/2018 p. 482, 494. Cf. en outre la définition des données médicales à l'art. 4 du règlement général sur la protection des données (RGPD) de l'Union européenne (UE) du 27 avril 2016 (EUR-Lex - 32016R0679 - EN - EUR-Lex (europa.eu)) : « les données à caractère personnel relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne »

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Risques en matière de protection des données

La réutilisation, mais en particulier l'appariement de données médicales sont assortis de divers risques en matière de protection des données. Si plusieurs jeux de données sont appariés, la valeur informative sur chaque individu augmente. Il peut en outre arriver que la collection de données résultant d'un appariement de plusieurs jeux contienne en fin de compte les données de moins d'individus que celle des jeux initiaux. On crée ainsi une empreinte plus forte basée sur les données et une réidentification des personnes concernées qui fournissent les données est potentiellement facilitée grâce aux caractéristiques connues – le profil de la personnalité [43]. L'utilisation de techniques de divulgation déductives[78] est un défi constant du fait des possibilités de traitement des données qui sont sans cesse meilleures et plus rapides. Des études ont montré que les personnes concernées sont conscientes de ces risques et souhaitent des informations transparentes sur la procédure et les mesures de protection avant de consentir à une réutilisation des données avec appariement [40].

Un appariement de données médicales concerne en particulier les dispositions générales en matière de protection des données (art. 4 ss LPD) à plusieurs égards. Dans le présent contexte de l'appariement de données médicales, il est d'abord nécessaire de clarifier de manière étayée comment il est possible de trouver une solution qui est en même temps compatible avec les principes de base du droit de la protection des données.

En ce qui concerne les « données personnelles sensibles », il existe des dispositions en partie plus restrictives (cf. p. ex. l'art. 4, al. 5, LPD, art. 6, al. 7, let. a, nLPD).

Le traitement automatique des données pourrait aussi constituer un « profilage » ou un « profilage à risque élevé ». La nLPD (art. 5, let. f et g) définit ces termes comme suit :

- **profilage** : toute forme de traitement automatisé de données personnelles consistant à utiliser ces données pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique ;
- **profilage à risque élevé** : tout profilage entraînant un risque élevé pour la personnalité ou les droits fondamentaux de la personne concernée, parce qu'il conduit à un appariement de données qui permet d'apprécier les caractéristiques essentielles de la personnalité d'une personne physique.

Ces traitements des données sont soumis à des prescriptions en partie plus strictes que les données personnelles ordinaires (cf. p. ex. art. 6, al. 7, let. a et b, nLPD). La LPD en vigueur connaît déjà la notion de « profil de la personnalité », décrite comme « un assemblage de données qui permet d'apprécier les caractéristiques essentielles de la personnalité d'une personne physique » (art. 3, let. d, LPD) ; des dispositions en partie plus strictes s'appliquent aussi à cet égard (cf. art. 4, al. 5, LPD).

Avec l'entrée en vigueur de la nLPD, il y aura en principe une obligation de procéder à une analyse d'impact relative à la protection des données personnelles (art. 22 nLPD).

La réutilisation et l'exploitation appariée de données médicales ne peuvent évidemment être concrétisées que si les conditions-cadres légales correspondantes sont créées ou que les normes existantes sont adaptées. La conformité avec les principes du droit de la protection des données n'est pas évaluée de manière définitive dans le cas présent au regard de l'intérêt public en faveur d'une utilisation multiple des données médicales, sur lequel se concentre le rapport. Pour tenir compte de l'exigence d'une utilisation accrue des données médicales de la part des milieux scientifiques et de la société, des clarifications juridiques approfondies sont nécessaires ; la pesée des intérêts entre les demandes précitées et l'intérêt pour une protection des données aussi étendue que possible doit être effectuée par les instances compétentes. Il convient également de décider s'il est nécessaire de formuler des exceptions dans des lois spéciales pour certains domaines, et le cas échéant comment.

[78] Méthodes de divulgation du contenu d'un thème, ici de la personne ; la logique est déductive, c'est-à-dire qu'on conclut les aspects concrets ou spéciaux à partir des aspects généraux.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Réutilisation et appariement des données statistiques liées à la santé

Les conditions pour une réutilisation des données médicales à partir de la statistique fédérale publique (« données statistiques ») sont précisées dans la loi sur la statistique fédérale (LSF ; RS *431.01*). Pour des raisons de protection des données, ces données médicales ne peuvent être transmises que sous une forme entièrement anonymisée et exclusivement à des fins ne se rapportant pas à des personnes, comme la recherche, la planification et la statistique.

En outre, des données de la statistique publique de la Confédération peuvent être appariées par l'OFS sur la base de la LSF, de l'ordonnance concernant l'exécution des relevés statistiques fédéraux (ordonnance sur les relevés statistiques ; RS *431.012.1*)[79] et de l'ordonnance du DFI concernant l'appariement de données statistiques (RS *431.012.13*)[80]. L'OFS peut aussi effectuer des appariements de données sur mandat de tiers dans le cadre d'un contrat de protection des données, où les appariements sont soumis aux mêmes exigences en matière de protection des données que les appariements internes à l'OFS. Les données appariées par l'OFS ne peuvent pas être transmises durablement et doivent être restituées à l'OFS une fois les travaux terminés et, dans le cas des données médicales sensibles, détruites dans tous les cas.

Réutilisation et appariement des données personnelles liées à la santé

La réutilisation des données personnelles liées à la santé à des fins de recherche est précisée dans la LRH (art. 32 ss) et dans l'ordonnance relative à la recherche sur l'être humain à l'exception des essais cliniques (ordonnance relative à la recherche sur l'être humain, ORH ; RS *810.301*) (en particulier les art. 28 ss).

Les mêmes conditions-cadres juridiques sont applicables en ce qui concerne les appariements de données qui ne sont réalisables qu'en utilisant des caractéristiques personnelles telles que le nom, le prénom et la date de naissance comme variables d'appariement, car il manque en Suisse une base légale pour un identificateur univoque librement utilisable.

Les données collectées de manière anonymisée ou les données anonymisées ou désidentifiées liées à la santé (voir la définition à l'art. 3, let. i, LRH) n'entrent pas dans le champ d'application de la LRH (art. 2, al. 2, let. c). Il convient toutefois de tenir compte de la disposition spéciale concernant les données personnelles génétiques (art. 32 LRH, art. 30 ORH).

Concept du consentement

La LRH est à la base du concept du consentement. Cela signifie qu'indépendamment de certains cas exceptionnels définis dans la loi ou l'ordonnance, la personne concernée doit donner son consentement (*informed consent*) à la réutilisation de données personnelles à des fins de recherche.

Le concept de l'opposition est en revanche appliqué dans la loi fédérale sur l'enregistrement des maladies oncologiques (LEMO ; RS *818.33*). Les données personnelles sont enregistrées par le registre des tumeurs compétent seulement si le patient ne s'y est pas opposé après avoir été informé de manière circonstanciée (cf. art. 6 LEMO).

L'annexe 5 (cf. chap. 8) donne une vue d'ensemble des formes de consentement applicables selon le droit en vigueur. Ces formes de consentement sont soutenues par swissethics, l'Association suisse des commissions d'éthique de la recherche, et par les hôpitaux universitaires suisses dans le cadre d'un modèle de consentement général (CG)[81]. La vue d'ensemble tient également compte des possibilités mises en œuvre dans la pratique pour une transmission ou une réutilisation des données.

Défis pour les producteurs et les utilisateurs de données

La pratique du consentement en vigueur en Suisse limite les possibilités, pour les producteurs de données/responsables, de satisfaire les demandes des utilisateurs intéressés visant à accéder à leurs

[79] https://www.admin.ch/opc/fr/classified-compilation/19930224/index.html
[80] https://www.admin.ch/opc/fr/classified-compilation/20122208/index.html
[81] Les hôpitaux universitaires suisses se sont entendus sur l'utilisation du modèle, cf. 20200331_Regulatory_Affairs_Watch_unimedsuisse_Nienhaus_Agnes.pdf

données médicales en vue d'une réutilisation et – en fonction de la demande – d'une exploitation appariée conformément à la loi.

Les exigences posées au consentement dépendent principalement de la nature des données (degré de sensibilité), du rapport avec la personne (degré d'anonymisation) et du risque d'identification non autorisée de la personne concernée. Moins une **atteinte à la personnalité** peut être exclue **lors du traitement des données**, plus les exigences posées au consentement (ou à la justification) sont sévères et plus la destination autorisée est limitée.

Si l'accès aux données est accordé à des tiers, cela signifie dans la plupart des cas aussi un **détachement du but initial de la collecte ou de l'utilisation des données** (pour des exemples, voir l'annexe 6, chap. 8).

En vertu du droit en vigueur, la transmission des données est en principe possible en cas de consentement général, de données sous forme codée et d'approbation du projet par la commission d'éthique. Si un **appariement** de plusieurs jeux et sources de données précède la **réutilisation**, il faut absolument **tenir compte** des points suivants :

- Un appariement réussi présuppose l'existence d'une caractéristique univoque ou de plusieurs caractéristiques permettant d'identifier une personne, raison pour laquelle le consentement de la personne concernée est généralement indispensable.
- Pour les données provenant de différentes sources, des dispositions différentes voire contradictoires peuvent s'appliquer en matière de consentement.
- Toutes les données du nouveau jeu apparié sont soumises au niveau de protection le plus élevé, c'est-à-dire celui des données personnelles liées à la santé.

Exigences posées aux personnes concernées

Avec l'encouragement ciblé d'une utilisation multiple des données médicales, la nécessité d'obtenir de manière prospective ou rétrospective le consentement des personnes concernées qui fournissent les données augmente par conséquent. C'est l'élément central qui initie le processus des données (première collecte) et qui le maintient (réutilisation). En donnant leur consentement, les personnes concernées doivent non seulement donner avec confiance leur accord à l'utilisation de leurs données dans des infrastructures sûres. Elles doivent aussi pouvoir se considérer comme des acteurs agréés dans l'espace de données médicales qui contribuent, en leur faveur et en celle d'autrui, au développement de la recherche – pas seulement médicale ou clinique – relative à la santé et au développement du système de santé (cf. chap. 3.1, figure 3 à droite).

Cela correspond au **concept de la souveraineté des données**[82], qui gagne de plus en plus en importance dans les débats politiques et juridiques. Avec l'élargissement du spectre de données personnelles et non personnelles qui sont collectées dans l'espace de données médicales ou qui sont disponibles et doivent être réutilisées, il est toujours plus difficile d'organiser l'accès et l'utilisation de ces différentes données.

Besoin de développement

Il est indispensable que les utilisateurs secondaires de données puissent demander le libre consentement pour la réutilisation spécifique et l'exploitation appariée des données auprès du producteur de données/responsable ou auprès des personnes concernées.

Pour cela, il est nécessaire de **moderniser la pratique du consentement** en Suisse. Indépendamment de la forme du consentement, il existe d'abord un besoin de rattrapage technique dans la mise en œuvre d'un consentement virtuel ou électronique. Cela peut signifier au minimum le passage du consentement éclairé du format papier avec signature manuscrite au format électronique, comme le

[82] Cf. Glossar der digitalen Transformation | bidt > Glossar: Datensouveränität | bidt (en allemand) pour les notions de « souveraineté des données » et « souveraineté numérique ».

thématise un projet du SPHN[83] et comme l'évaluation de la LRH[84] l'identifie comme une mesure essentielle à prendre [44, 45]. Mais l'objectif devrait porter sur une mise en œuvre plus étendue dans l'environnement numérique, qui présuppose la création d'un **concept-cadre éthique pour la souveraineté numérique** et pose des défis en ce qui concerne la transmission et la compréhension de l'information, l'acceptation et la confiance [46, 47].

En même temps, il faut un **« cadre juridique »** clairement structuré dans lequel les services compétents prennent des décisions cohérentes, par le biais de méthodes de mise en œuvre efficaces, en ce qui concerne la réutilisation et l'appariement de données pour différentes finalités (autorisation d'utiliser les données).

Il s'agit de trouver une solution commune qui satisfait tous les acteurs des données, qui garantit la protection des données de manière adéquate et qui permet d'encourager et de mener une recherche importante pour le domaine de la santé et la santé de la population.

4 Proposition de mise en œuvre d'un « système national de réutilisation et d'appariement des données médicales à des fins de recherche »

L'enquête de l'OFSP auprès des producteurs de données/responsables et des utilisateurs de données [34] a montré qu'une réutilisation multiple et appariée des données médicales à des fins de recherche est possible aujourd'hui, mais souvent difficile et inefficace, car semée d'obstacles.

Les obstacles corrélés résultent ou existent dans la production, l'acquisition et l'enregistrement, la transmission ou l'appariement des données. Ils concernent l'ensemble du cycle de vie des données et des principes FAIR prérequis (cf. chap. 3.2).

La plupart des utilisateurs de données concernés ont pu gérer leur problème spécifique avec certaines pertes (p. ex. concernant la pertinence des résultats) en tenant compte d'autres acteurs/experts, c'est-à-dire trouver une solution spécifique à chaque cas [34]. Il manque toutefois un cadre structurant qui permet la mise en œuvre systématique et durable de tels projets de recherche.

En résumé, une **solution adaptée aux besoins** doit répondre aux **exigences** suivantes :

(1) Il y a le consentement à l'utilisation primaire et secondaire des données ou une autre forme de prise en compte adéquate de la personne concernée.

(2) Une référence définie aux personnes est possible, car une réutilisation présupposant un appariement est exclue de fait avec des données anonymisées.

Un identificateur univoque est disponible comme variable d'appariement.

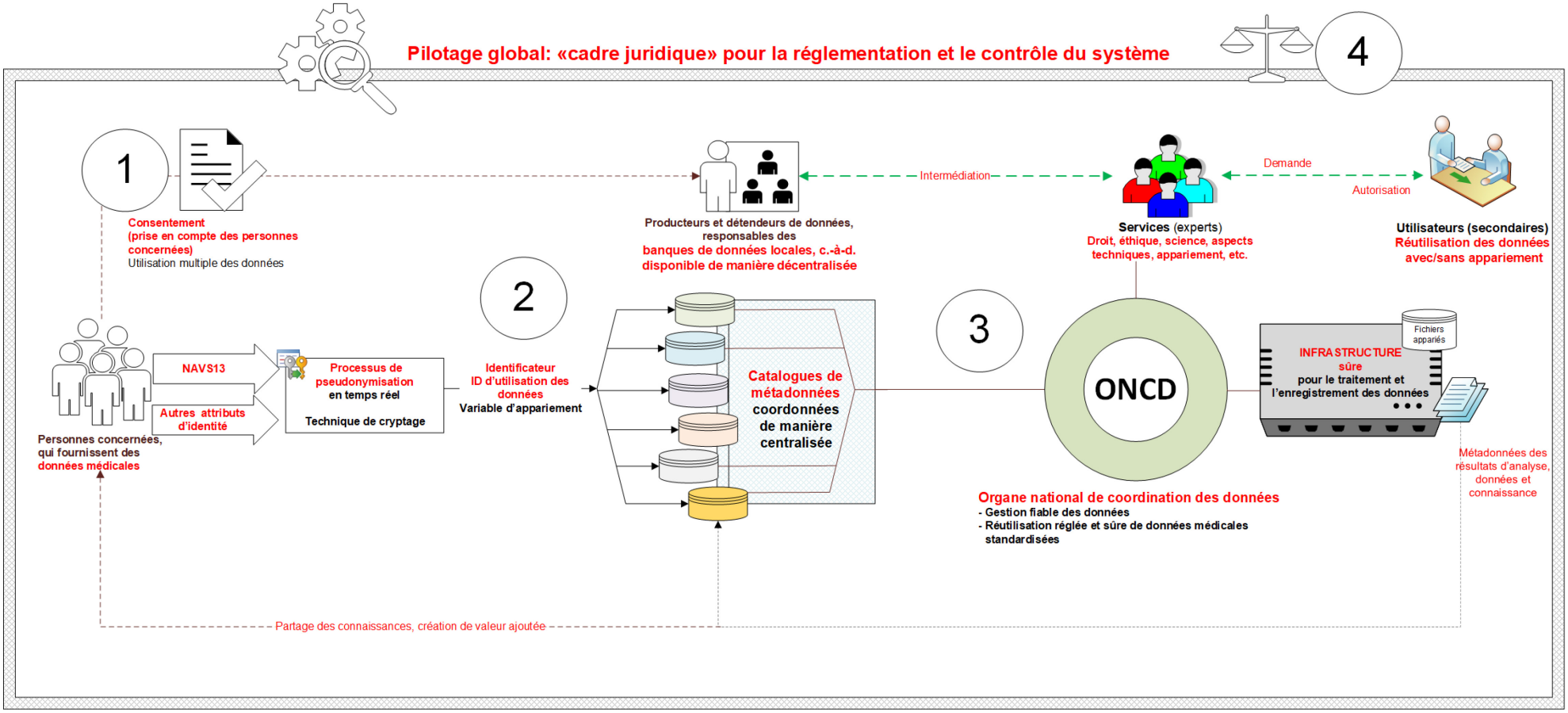
(3) Les exigences qualitatives posées aux données et à leur gestion sont systématiquement remplies.

(4) Un concept étendu de protection des données est défini et garantit l'application cohérente des mesures de protection des données et de sécurité des données tout au long du cycle de vie des données.

Le **« système de réutilisation et d'appariement des données médicales à des fins de recherche »** (désigné ci-après comme système) présenté à la figure 4 (voir p. 35) se réfère à ces quatre points, symbolisés par les chiffres ① à ④.

[83] Projet de développement de l'infrastructure SPHN « E-General Consent: Development and Implementation of a Nationwide Harmonized Interactive Electronic General Consent », cf. 2017DEV04_Lay_summaries_20170117 (sphn.ch)
[84] Loi relative à la recherche sur l'être humain : évaluation (admin.ch)

Figure 4 Système national de réutilisation et d'appariement des données médicales à des fins de recherche



Explications : Les **points ① à ④** montrent où, dans le système, et dans quelle phase du cycle de vie des données il est nécessaire de satisfaire aux exigences relatives à une solution adaptée aux besoins pour la réutilisation et l'appariement de données médicales.
La **moitié droite de la figure** concerne l'enjeu principal de l'échange de données entre les producteurs et les utilisateurs secondaires des données.
La **moitié gauche de la figure** représente les conditions de base qui doivent être remplies pour que la transmission et la réutilisation des données avec appariement préalable soient illimitées et efficaces.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

On peut comprendre ce système comme un **écosystème numérique (multi-)sectoriel**[85] (pour la définition, voir le chap. 7, Glossaire), comme le demande la motion CSEC-N 21.3021 « Création d'un écosystème intégré de données médicales. Une plus-value pour la recherche et pour la société »[86].

En **initiant la mise en place d'un tel système**, la **Confédération** peut créer des conditions-cadres acceptables et intéressantes pour tous les acteurs impliqués pour l'utilisation multiple des données médicales à des fins de recherche [48, 49]. Le système prévoit, pour l'ensemble du cycle de vie des données, les structures et les processus nécessaires dans lesquels les personnes concernées, les producteurs de données/responsables et les utilisateurs de données médicales peuvent agir conformément au droit en respectant des exigences et des normes de qualité et de sécurité.

L'activité principale du système

En principe, la Confédération proposerait un **service numérique** qui permet la création d'une valeur ajoutée aux données médicales (cf. chap. 2).

Ce service consiste **à établir et à régler l'échange de données médicales de qualité** pour une réutilisation avec ou sans appariement préalable entre les producteurs et les utilisateurs de données en tenant compte des personnes concernées. Un **organe national de coordination des données** (ONCD) assumerait cette **mission fiduciaire**[87] au nom de la Confédération (cf. figure 4).

Structure générale

Pour fournir ce service et permettre la collaboration, l'ONCD a besoin d'éléments techniques, à savoir une plateforme numérique. En plus de la réalisation pratique et technique de la transmission et de l'utilisation des données, il doit proposer/remplir, en tant que « prestataire de services du système », des fonctions techniques complémentaires qui concernent des aspects tels que l'éthique, le droit, le domaine des contrats, la qualité, etc.

Par conséquent, la mise en place du système doit suivre deux approches d'importance équivalente. Il est d'abord nécessaire de clarifier la construction d'un point de vue technique, organisationnel et commercial, puis de contrôler et de garantir le respect des exigences juridiques, en particulier en matière de protection des données, et éthiques dans le cadre de l'application des structures et des processus.

4.1 Mise en place de structures et de processus sur le principe de modularité

L'application du principe de modularité permet de parvenir à la meilleure fonctionnalité[88] possible du système en économisant du temps. La construction du système est simplifiée grâce à des éléments « préconçus ». Avec l'expérience, le système peut être transformé, développé ou optimisé par le biais de nouveaux éléments (extension de la modularité).

La participation de partenaires[89] et d'experts potentiels au processus de mise en place, en particulier du domaine du droit et de la législation, revêt une importance capitale. Une prise en compte précoce des acteurs, en particulier des acteurs des données, simplifiera la mise en œuvre définitive du système et des nouveaux processus [49].

Par conséquent, les institutions et les experts en Suisse doivent participer, avec leurs structures ou unités fonctionnelles (applications, solutions techniques, prescriptions, etc.) existantes – éprouvées et fonctionnelles – ou en cours de constitution, à la mise en place et à l'exploitation du système. Les

[85] L'application générale du système est envisageable pour la réutilisation de données personnelles sensibles et d'autres données sensibles.
[86] 21.3021 | Création d'un écosystème intégré de données médicales. Une plus-value pour la recherche et pour la société | Objet | Le Parlement suisse
[87] Voir le terme de fiduciaire des données ou, en anglais, *Data Trustee* et son dérivé, le *Trust Center*
[88] Degré de ciblage du produit ; le système est « fonctionnel » dès qu'il fournit à l'utilisateur du produit utilement des fonctions avec des propriétés définies pour une utilisation donnée.
[89] Une distinction est faite entre les partenaires de (l'éco-) système et les participants à (l'éco-)système. Les (entreprises) partenaires ont des interfaces techniques et pratiques directes avec la plateforme numérique, elles fournissent des services/applications, alors que les participants n'ont pas d'interfaces directes avec la plateforme numérique, mais ce sont des acteurs importants qui interagissent indirectement.

rôles, compétences et dépendances respectifs, également de nature financière, doivent être définis et documentés dans le contexte global du système.

Établir des structures et des processus pour la gestion des données

Comme le rapport l’a déjà esquissé (voir le chap. 4, Système/Activité principale), l’**organe national de coordination des données** (ONCD) devrait assumer la fonction centralisée de la gestion des données tout en servant de point de référence technique.

Pour encourager l’utilisation multiple des données, il faut en particulier proposer des incitations aux producteurs et aux utilisateurs de données, les soutenir, mais aussi les contrôler. C’est pourquoi la gestion des données devrait être simple et flexible, de sorte à pouvoir prendre efficacement et appliquer systématiquement les décisions nécessaires (au cas par cas) pour une transmission et une réutilisation appariée en tenant compte des dispositions légales.

Pour l’acquisition, la transmission, l’appariement, l’exploitation et l’enregistrement des données, l’ONCD proposerait et maintiendrait un environnement formel et informatique sûr. Les structures et les processus devraient être conçus de telle sorte à pouvoir garantir la sécurité et la confiance aux personnes concernées, aux producteurs de données/responsables et aux utilisateurs de données.

L’annexe 7 (chap. 8) donne une description simple des processus envisageables pour un traitement secondaire des données, sans anticiper la concrétisation nécessaire des processus exacts, des compétences, etc.

Pour la mise en place des structures et des processus de l’ONCD, il serait opportun de recourir à des « éléments de construction » existants. Le tableau 2 (voir p. 38 ; non exhaustif) montre quels partenaires potentiels pourraient contribuer par des solutions partielles – pour la plupart techniques et organisationnelles (unités fonctionnelles, infrastructures/interfaces techniques) – et par l’expertise (unités spécialisées) dans des domaines spécifiques de la gestion des données.

Tableau 2 Partenaires potentiels dont les unités fonctionnelles et spécialisées peuvent contribuer à la mise en place de l'ONCD

Domaine fonctionnel / spécialisé	**Institution, organisation ou projet** (sans prétendre à l'exhaustivité)	
Infrastructures et processus pour l'échange et le traitement des données	Swiss Personalized Health Network (SPHN) - BioMedIT Network - SPHN Data Coordination Center (DCC)	SPHN - Swiss Personalized Health Network (SPHN) Projet BioMedIT - SPHN ; Data Coordination Center - SPHN
	Domaine des EPF - Personalized Health and Related Technology (PHRT) - Swiss Data Science Center (SDSC)	PHRT - Personalized Health and Related Technologies (sfa-phrt.ch) ; SDSC - Swiss Data Science Center
	Office fédéral de la statistique (OFS), programme de gestion nationale des données (NaDB), projet pilote Secteur stationnaire hospitalier (SpiGes)	Gestion nationale des données NaDB | Office fédéral de la statistique (admin.ch)
	OFS, Interopérabilité et registre (IOR), service sedex	sedex (www.sedex.ch) | Office fédéral de la statistique (admin.ch)
	SWITCH The Research Data Connectome project (open research data)	The Research Data Connectome - SWITCH
Infrastructures et processus pour la réalisation d'appariements de données	OFS, appariement de données	Appariement de données | Office fédéral de la statistique (admin.ch)
	FORS linkhub.ch	Mandat | linkhub.ch
Service coordinateur d'interopérabilité [17] : définition d'exigences sur l'établissement, l'application et la surveillance des normes	eHealth Suisse	Interopérabilité sémantique - eHealth Suisse (e-health-suisse.ch) ; aide à la mise en œuvre (e-health-suisse.ch)
	SPHN Data Coordination Center (DCC)	Data Coordination Center - SPHN
	OFS Centre de compétences en science des données (DSCC) – services de base OFS – plateforme d'interopérabilité NaDB (Swiss Data Steward)	Centre de compétences en science des données (DSCC) | Office fédéral de la statistique (admin.ch) ; Gestion nationale des données NaDB | Office fédéral de la statistique (admin.ch)
Examen et conseils pour les questions éthiques et juridiques	Commission nationale d'éthique dans le domaine de la médecine humaine (CNE c/o Office fédéral de la santé publique)	[Mission] Commission nationale d'éthique dans le domaine de la médecine humaine CNE (admin.ch)
	Swissethics	swissethics | site Internet de l'Association suisse des commissions d'éthique de la recherche
	Commission centrale d'éthique (CCE) de l'Académie suisse des sciences médicales (ASSM)	Commission centrale d'éthique (assm.ch)
	OFSP, division Biomédecine, section Recherche sur l'être humain	Recherche sur l'être humain (admin.ch)
	Swiss Personalized Health Network (SPHN) - ELSI – ethical, legal and social issues – Advisory Group (ELSIag) - SPHN Data Coordination Center (DCC) : ELSI Helpdesk	ELSI Advisory Group (ELSIag) - SPHN Data Coordination Center - SPHN

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

4.2 Conception fonctionnelle du système

Seule la conception fonctionnelle de l'« édifice » rend les structures et les processus créés formellement applicables sur le plan juridique. Pour ce faire, les conditions exposées au chap. 3.2 pour la réutilisation et l'appariement de données médicales doivent être remplies.

La conception s'appuie sur la finalité (cf. chap. 1.2.3 pour la finalité de la recherche) et le champ d'application des données. On peut en principe partir de l'idée que les structures et les processus proposés ne sont pas seulement importants pour le traitement de données médicales, mais par exemple aussi pour le traitement de données issues du domaine des assurances sociales ou des domaines de la migration et de l'économie.

En se concentrant sur l'espace de données médicales, il s'agit de créer un cadre juridique, dont un concept étendu de protection des données, qui permettrait une exploitation de l'édifice en toute sécurité et en toute confiance de même que le pilotage et le contrôle de l'ensemble du système.

4.2.1 Créer les conditions pour la fonctionnalité du système

En Suisse, il y a des obstacles à l'application des principes de données FAIR. Les conditions pour le fonctionnement du système de réutilisation et d'exploitation appariée des données médicales à des fins de recherche ne sont pas pleinement réunies.

Pour les défis suivants, il s'agit – comme le rapport l'a déjà mentionné – de trouver des solutions réalisables :

- le consentement ou la prise en compte adéquate des personnes concernées pour l'utilisation secondaire des données,
- la référence nécessaire à des personnes dans les données pour les appariements (variable d'appariement),
- la connaissance de la disponibilité, de l'accès et de la qualité des données (catalogues de métadonnées, normes),
- le plein respect de la protection et de la sécurité des données.

Les défis sont localisés dans la représentation graphique du système (figure 4, voir p. 35) aux points ①, ②, ③ et ④. Plusieurs solutions sont envisageables.

Les propositions esquissées ci-après sont une première base, mais elles doivent être examinées par des experts en tenant compte des acteurs, puis concrétisées en vue de leur mise en œuvre.

① Consentement des personnes concernées à la réutilisation des données

Une modernisation de la pratique du consentement doit favoriser et soutenir la réutilisation et l'exploitation appariée du large spectre des données médicales par la recherche. Une gestion numérique des consentements et des autorisations pour/par les personnes concernées – appelée gestion du consentement (*consent management*) – devrait pouvoir être appliquée à toutes les collectes de données ou études qui comprennent des informations personnelles pertinentes pour la santé.

Dans l'intérêt d'observer et d'optimiser constamment et durablement la situation et la couverture sanitaire de la population (tout au long de la vie), deux options sont envisageables pour une nouvelle procédure en matière de consentement et méritent un examen approfondi.

(1) Dans une **procédure dynamique et interactive**, les personnes concernées (fournisseurs de données) donnent un **consentement électronique à la première collecte et à l'utilisation** des données médicales (modèle *opt in* à défaut d'une autre réglementation dans la loi) de manière spécifique et ainsi destiné à des fins déterminées après avoir été informées en conséquence.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Un **droit d'opposition** doit s'appliquer à l'**utilisation secondaire** des données. Les personnes concernées consentent, dans le cadre de la première collecte, à ce que leurs données puissent être fournies ou transmises dans le temps pour une utilisation secondaire avec/sans appariement. Elles reçoivent des informations sur les projets correspondants **par voie électronique**, ce qui leur permet de faire opposition au cas par cas (modèle *opt out*).

(2) Comme alternative, la personne concernée peut **volontairement donner** des données nouvellement collectées, actuelles ou futures pour une **réutilisation libre non spécifiée sans révocation**, pour autant qu'aucun intérêt commercial ne soit poursuivi. Il convient d'examiner si la diffusion des données – en particulier des futures données – devrait être une nouvelle fois confirmée par la personne concernée après un délai minimal fixé dans la loi (p. ex. 5 ans) ou un délai plus long à choisir librement et comment procéder en l'absence de décision.

Il s'agit d'un **nouveau développement à partir du principe de base du consentement général** qui se concentre jusqu'ici sur l'utilisation multiple des données personnelles des patients aux fins de la recherche médicale. L'application devrait désormais se référer de manière universelle au spectre complet des données personnelles liées à la santé et permettre une utilisation à des fins qui dépassent la recherche purement médicale.

Pour les deux options, la question de la réutilisation des données après la mort et, ainsi, du don de données posthume sur une base volontaire (cf. chap. 8, annexe 5) devrait être explicitement examinée et clarifiée. L'examen de la question du don de données correspond à l'exigence du postulat 20.3700 Bellaiche Judith « Examiner la faisabilité d'un don volontaire de données personnelles anonymisées, en vue d'un usage d'intérêt public[90] ».

La modernisation de la pratique du consentement affecte aussi sa réglementation juridique et son ancrage dans la loi fédérale (cf. chap. 3.2.4 et 4.2.2), qui devrait aussi prévoir une possibilité de punir les abus.

Pour le développement organisationnel et technique de la *gestion du consentement* en Suisse, il existe déjà un premier concept[91] et le domaine des services numériques propose des incitations pour la mise en œuvre pratique par le biais de *Privacy Cockpits* [50].

② Identificateur personnel univoque comme variable d'appariement

La réutilisation et l'appariement de données ne sont efficaces que si un identificateur personnel univoque peut être appliqué en lien avec la procédure de consentement.

À l'heure actuelle, il est interdit, en l'absence de base légale (LAVS[92]), d'utiliser directement le **NAVS13** comme identificateur univoque et variable d'appariement à des fins de recherche. Pour des raisons de protection des données, une modification de loi en ce sens a été reportée[93] jusqu'ici, mais pourrait être considérée comme une option.

Pour le système (cf. figure 4), l'alternative proposée est un processus de pseudonymisation cryptologique sûr qui est applicable pour la production de données primaires. Une technique de codage appliquée de manière centralisée doit permettre de générer en temps réel un **identificateur d'utilisation des données (ID d'utilisation des données)** univoque non signifiant **sur la base du NAVS13 et d'autres caractéristiques personnelles collectées de manière traditionnelle** pour la personne concernée et de l'attribuer au jeu de données concerné. Cela signifie que l'ID d'utilisation des données non signifiant est enregistré avec les données médicales auprès du producteur de données/responsable et permet en principe un appariement déterministe efficace.

Un **service de pseudonymisation indépendant** des producteurs et utilisateurs de données privés et publics sur les plans administratif et organisationnel devrait assurer la génération des ID d'utilisation des données. Seul ce service de confiance doit pouvoir identifier la personne concernée, et à des fins

[90] 20.3700 | Examiner la faisabilité d'un don volontaire de données personnelles anonymisées, en vue d'un usage d'intérêt public | Objet | Le Parlement suisse
[91] Projet de développement de l'infrastructure du SPHN « C3-Study Citizen Centered Consent », cf. 2017DEV15_Lovis_Lay_summaries_20180129 (sphn.ch)
[92] RS *831.10* - Loi fédérale du 20 décembre 1946 sur l'assurance-vieillesse et survivants (LAVS) (admin.ch) ; FF **2020** 9643 - Loi fédérale sur l'assurance-vieillesse et survivants (LAVS) (Utilisation systématique du numéro AVS par les autorités) (admin.ch)
[93] Cf. FF **2013** 4747 - Message concernant la loi fédérale sur le dossier électronique du patient (LDEIP) (admin.ch), p. 4782

bien définies (p. ex. en lien avec la *gestion du consentement* pour la réutilisation des données ; si la personne concernée souhaite être informée en cas de risque pour la santé ; en cas de révocation du consentement, etc.).

Une telle **utilisation systématique du NAVS13** en vertu de l'**art. 153*b* LAVS** doit être réglée dans une loi spéciale pour le système ou pour une unité fonctionnelle spécifique de l'ONCD [94].

La **procédure de pseudonymisation** précise doit être développée ou définie et intégrée dans le système comme une unité fonctionnelle centralisée. Ce faisant, il est également nécessaire de clarifier le couplage avec d'autres fonctions dans le système, en particulier le processus de consentement et la réutilisation sans ou avec appariement préalable.

Les procédures d'ores et déjà appliquées aujourd'hui ou les identificateurs qui en résultent doivent être utilisés comme des **modèles** ou évalués comme des options par les experts. Il y a notamment :

- L'attribution centralisée du numéro d'identification du patient du DEP, qui pourrait se développer en identificateur personnel pour le secteur de la santé, selon le message[95] du Conseil fédéral concernant la LDEP[96]. La Centrale de compensation (CdC) attribue et gère le numéro d'identification du patient au sens de la LDEP (EPR-SPID), qui n'est apparié avec le numéro AVS qu'à l'interne de la CDC.
- La procédure de hachage de l'OFS, appliquée jusqu'ici de manière décentralisée, pour générer le code de liaison anonyme, qui devrait faire l'objet d'un examen quant à sa pertinence hors du domaine statistique.

Il est nécessaire d'estimer les conséquences pratiques et juridiques de chaque option dans le contexte de la conception étendue des structures et des processus du système.

Les solutions décentralisées (voir paragraphe suivant) peuvent permettre l'appariement de données entre elles au moyen de l'identificateur personnel, tant que les solutions garantissent séparément et conjointement que la protection des données est assurée pour les données de sortie.

③ Gestion efficace de données de qualité (principes FAIR)

Les producteurs de données/responsables suivent jusqu'ici leurs propres règles de gestion des données. Grâce à des structures sûres et à des processus clairs, le système vise à renforcer la confiance entre les producteurs et les utilisateurs de données. Une solution globale durable en matière de gestion des données aura pour corollaire que les données médicales existantes de qualité seront disponibles de manière générale, accessibles, interopérables et utilisables pour différentes tâches moyennant des garde-fous appropriés.

Une solution de conservation des données centralisée, où toutes les données médicales disponibles en Suisse sont stockées ensemble sur un support unique, est jugée irréalisable au regard de la quantité et de la complexité des données, pour des raisons de sécurité, mais aussi sociales [13, 51]. C'est pourquoi il convient de soutenir une solution de **conservation des données décentralisée ayant une portée locale**, ce qui confère aux producteurs de données publics et privés une coresponsabilité dans le fonctionnement du système.

Une utilisation efficace des données au sein d'une organisation, mais surtout la **mise à disposition** de précieuses données **pour des tiers**, impliquent que les différentes collections de données de la même organisation soient réunies et enregistrées dans un système[97] local en tenant compte des dispositions applicables en matière de protection et de sécurité des données. En tant que responsables, les producteurs de données doivent être soutenus dans la génération de **données de la meilleure qualité possible**.

---

[94] Cf. RS 831.10 - Loi fédérale du 20 décembre 1946 sur l'assurance-vieillesse et survivants (LAVS) (admin.ch) Quatrième partie : Utilisation systématique du numéro AVS en dehors de l'AVS
[95] Cf. FF **2013** 4747- Message concernant la loi fédérale sur le dossier électronique du patient (LDEIP) (admin.ch), chap. 2, p. 4803-04, commentaire de l'art. 6 « Autres utilisations du numéro d'identification du patient »
[96] RS *816.1* - Loi fédérale du 19 juin 2015 sur le dossier électronique du patient (LDEP) (admin.ch)
[97] « Data Warehouse » ou « Data Lake », voir le glossaire au chapitre 7.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Pour réutiliser consciemment les données médicales en Suisse, les utilisateurs de données ont besoin des **métadonnées correspondantes**, c'est-à-dire des descriptions structurées et standardisées d'ensembles et de contenus de données dans des catalogues. Ainsi, les tiers peuvent notamment identifier l'accessibilité et la convivialité des stockages de données locaux en clarifiant les droits d'accès aux données et les conditions d'utilisation, mais trouvent surtout des informations sur la qualité des données (cf. chap. 3.2.1 et 3.2.2)[98].

Par conséquent, il convient d'envisager comme solution la **conservation de métadonnées coordonnée de manière centralisée avec une conservation de données décentralisée** (*local data custodian*). Pour une gestion efficace des données, l'ONCD (cf. chap. 4.2.1) dépend de l'existence et de la **collaboration avec les catalogues de métadonnées**, qui sont librement consultables pour les utilisateurs de données intéressés de Suisse et de l'étranger. Pour ce faire, il est nécessaire de constituer des catalogues de métadonnées indispensables et/ou de mettre en réseau les archives existantes en Suisse, de même que la documentation standardisée des collections de données conformément aux exigences définies.

Les **travaux commencés** doivent être poursuivis de manière ciblée et les **structures et/ou processus intégrés dans le système**.

La mise en œuvre du programme NaDB et, dans le domaine de la santé, du projet pilote SpiGes par l'OFS [52] devrait être appliquée à l'ensemble du spectre des données médicales, c'est-à-dire globalement à l'espace de données médicales et, par conséquent, soutenir la collecte unique et l'utilisation multiple des données médicales dans le secteur public et privé.

La plateforme d'interopérabilité I14Y est essentielle et doit permettre de rendre les métadonnées faciles à trouver, accessibles, interopérables et réutilisables à l'aide de catalogues de données pour tous les utilisateurs de données et services intéressés. L'OFS assume aujourd'hui déjà le rôle de l'administrateur des données suisse (*Swiss data steward*).

L'initiative Swiss Learning Health System (SLHS) a livré des connaissances complémentaires sur le développement d'archives de métadonnées suisses centralisées sur la santé à prendre en compte [26].

④ Concept étendu de protection des données

Les structures et processus à la base du système doivent être sûrs et l'interaction des acteurs impliqués dans les processus doit être organisée et contrôlée efficacement.

Le « 5-Safes Framework » est adéquat pour évaluer les risques liés à l'échange de données et piloter le développement et la mise en œuvre des structures et des processus, de telle sorte que les sûretés nécessaires[99] sont garanties en tout temps pour les personnes, les projets, les paramètres, les données et les résultats [53].

À cet égard, il est nécessaire d'établir et de mettre en œuvre un concept étendu de protection des données au sens de l'art. 7 nLPD. Le concept fixe la conception des processus, de la technique ainsi que des mesures organisationnelles et juridiques (« *privacy by design* ») et définit des paramètres par défaut, en particulier le respect standard des principes d'affectation et de minimisation des données (« *privacy by default* »).

[98] Exemple : archivage des métadonnées de cinq études de cohorte suisses dans le catalogue international Maelstrom, soutenu par le programme d'encouragement du SPHN. Cf. Maelstrom Catalogue - Maelstrom Research (maelstrom-research.org)

[99] Cf. www.fivesafes.org « Safe people » : peut-on compter sur le fait que les utilisateurs utilisent les données de manière appropriée ? « Safe projects » : cette utilisation des données est-elle appropriée, légale, éthiquement défendable et judicieuse ? « Safe settings » : la possibilité d'accès limite-t-elle une utilisation non autorisée ou les erreurs ? « Safe data » : les données comportent-elles suffisamment d'informations pour violer la confidentialité ? « Safe outputs » : la confidentialité est-elle préservée pour les résultats du processus de gestion des données ?

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

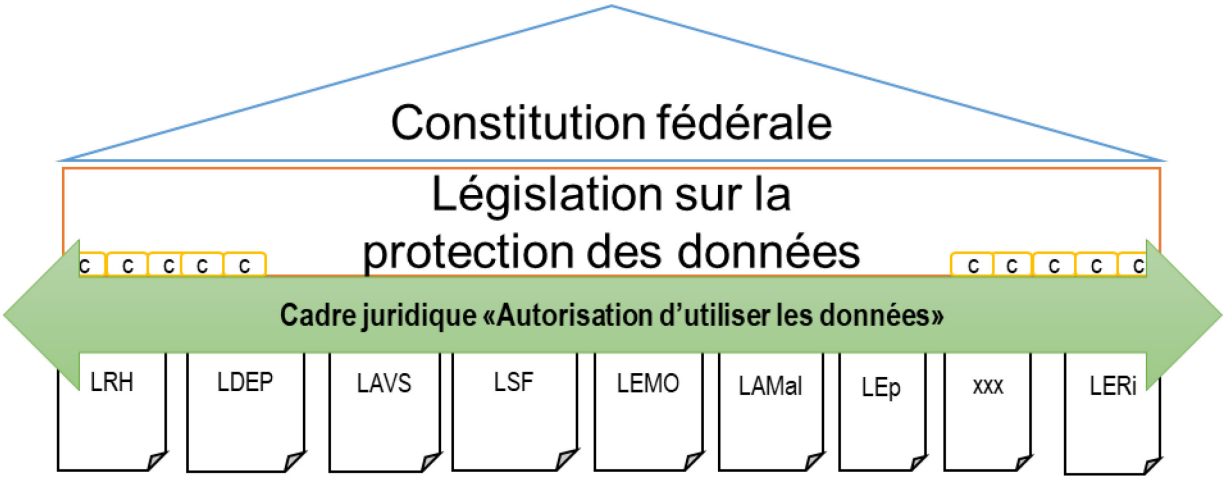
4.2.2 Assurer une conception formellement juridique

Autorisation d'utiliser les données

Pour que les structures et les processus à la base du système soient largement applicables, des adaptations légales seront nécessaires[100]. Les personnes concernées, les producteurs de données/responsables privés et publics et les utilisateurs de collections de données doivent disposer d'un cadre de confiance et de sécurité juridique dans lequel la transmission de données médicales à des tiers de même que leur utilisation répétée et leur exploitation appariée sont permises à différentes fins.

Il faut des règles du jeu qui puissent être comprises, acceptées et systématiquement appliquées par l'ensemble des acteurs dans l'espace de données médicales. C'est pourquoi il est proposé que le **système soit piloté à l'aide d'un cadre juridique** qui associe logiquement des principes applicables régis par les lois spéciales, y compris les dispositions d'exécution, et garantisse la mise en œuvre d'un concept étendu de protection des données (cf. figure 5). Pendant la mise en place du système, il est indispensable de discuter des questions juridiques dans le cadre des développements et concrétisations nécessaires ainsi que de clarifier et d'initier aux éventuels besoins d'adaptation spécifiques de bases légales existantes, par exemple la LRH, la LAVS, etc.

Figure 5 Esquisse pour un ancrage du cadre juridique



5 Conclusions du Conseil fédéral et prochaines étapes

Conclusions

Il est urgent de prendre des mesures dans le domaine de la numérisation du système de santé en Suisse. Il s'agit aussi de développer la gestion des données dans le domaine de la santé, qui pose un défi aussi bien aux autorités (Confédération, cantons) qu'aux fournisseurs de prestations, aux assureurs et à la recherche et doit progresser de manière coordonnée. C'est la seule façon d'utiliser les ressources disponibles de manière efficace et effective pour de meilleurs soins de santé.

C'est pourquoi le Conseil fédéral a l'intention – dans le cadre des mesures visant à améliorer la gestion des données dans le domaine de la santé – de concrétiser la proposition de mise en œuvre d'un système national de réutilisation et d'appariement des données médicales à des fins de recherche, telle qu'elle est présentée dans ce rapport. Le système esquissé a le potentiel de créer les conditions

[100] Voir la prise de position du conseiller fédéral Alain Berset du 22 septembre 2021 sur l'intervention parlementaire 19.4193 | Nouvelle loi-cadre nationale | Bulletin officiel | Le Parlement suisse

de base techniques, organisationnelles et commerciales nécessaires qui permettront une utilisation des données sensibles sûre et conforme à la protection des données. Par cette décision, le Conseil fédéral tient compte de la motion 21.3021[101] déposée par la CSEC-N, qui exige la création d'un écosystème intégré de données médicales, tout comme les demandes de plusieurs autres interventions parlementaires[102]. Il réagit ainsi également à la nécessité, apparue clairement à la suite de la pandémie de COVID-19, d'une utilisation des données flexible et à court terme, et reconnaît les enseignements et les progrès des projets publics et privés visant à améliorer l'utilisation des données médicales (p. ex. programme national de gestion des données NaDB, Swiss Personalized Health Network SPHN, linkhub.ch, PNR74 Système de santé, eHealth Suisse, etc.). Il est nécessaire d'exploiter la dynamique existante de manière ciblée et de bien coordonner, poursuivre ou fusionner rapidement les projets en cours.

Prochaines étapes

Le développement d'un système global cohérent comprend des structures appropriées et des réglementations légales qui soutiennent durablement les producteurs et les utilisateurs de données dans le processus de transmission, de réutilisation et d'appariement de données médicales précises tout en gagnant la confiance des personnes concernées et de la population. Ce faisant, il est également nécessaire de clarifier les rôles, compétences et responsabilités institutionnelles, par exemple lors de l'octroi d'un partage de données pour une réutilisation avec/sans appariement et le respect des obligations de rendre compte.

Le Conseil fédéral charge le DFI de concrétiser les processus et les structures du système proposé de même que les adaptations nécessaires des bases légales, en particulier en ce qui concerne la conformité avec la protection des données. À cet égard, il est nécessaire d'approfondir les propositions présentées dans le présent rapport quant à leur plausibilité, leur faisabilité et sous l'angle de la prise en compte internationale, en collaboration avec les organes spécialisés existants, le cas échéant en tenant compte d'autres experts importants et acteurs concernés. Un aspect du mandat consiste à intégrer directement les éléments existants et en devenir dans le système. Il s'agit en particulier, mais pas exclusivement, du développement de la plateforme d'interopérabilité I14Y découlant du programme NaDB pour une conservation des métadonnées coordonnée de manière centralisée et la prise en compte des enseignements de l'initiative SPHN. Si des clarifications spécifiques s'avèrent nécessaires, par exemple sur des questions liées à l'acceptation d'un don de données posthume dans la population, il est indispensable d'impliquer les acteurs concernés.

Les travaux doivent intervenir sous la direction du DFI (OFSP et OFS) et en étroite collaboration avec la ChF (DTI), le PFPDT, le DFF (CdC, OFIT), le DFJP (OFJ) et le DEFR (SEFRI), et s'inscrire stratégiquement dans la feuille de route pour la mise en œuvre de la transformation numérique de l'OFSP.

[101] 21.3021 | Création d'un écosystème intégré de données médicales. Une plus-value pour la recherche et pour la société | Objet | Le Parlement suisse
[102] Cf. liste au chap. 1.2.2

6 Bibliographie

1. Académie Suisses des Sciences Médicales (2019) Développement durable du système de santé. Feuille de route de l’ASSM. Swiss academies communications. https://doi.org/10.5281/zenodo.2222811
2. Santé publique Suisse (2013) Manifeste des données de meilleures qualité pour augmenter l’efficacité du système de santé. https://gesundheitsmanifest.ch/fr/gesundheitsdaten/. Accédé: 4 avril 2022
3. Hurrelmann K, Richter M (2018) Determinanten von Gesundheit. In: Bundeszentrale für gesundheitliche Aufklärung BZgA, Richter M, Hurrelmann K (Hrsg) Leitbegriffe. BZGA - Federal Centre for Health Education, https://doi.org/10.17623/BZGA:224-i008-1.0
4. Office fédéral de la santé publique OFSP Système de monitorage suisse des Addictions et des Maladies non transmissibles (MonAM). https://www.obsan.admin.ch/fr/MonAM. Accédé: 4 avril 2022
5. kofam Koordinationsstelle Forschung am Menschen Portail SNCTP (Swiss National Clinical Trials Portal). https://www.kofam.ch/fr/portail-snctp/rechercher-des-essais-cliniques/. Accédé: 4 avril 2022
6. ONEC · Organe national d’enregistrement du cancer Données. https://www.onec.ch/fr/donnees/. Accédé: 4. Avril 2022
7. NICER · Institut National pour l'Epidémiologie et l'Enregistrement du Cancer Publications. https://www.nicer.org/fr/publications/. Accédé: 4 avril 2022
8. Schmidlin K, Clough-Gorr KM, Spoerri A (2015) Privacy preserving probabilistic record linkage (P3RL): a novel method for linking existing health-related data and maintaining participant confidentiality. BMC Med Res Methodol 15:46. https://doi.org/10.1186/s12874-015-0038-6
9. Steck N, Spoerri A, Egger M (2015) Verknüpfte Gesundheitsdaten und Datenschutz in der Forschung: (k)ein Widerspruch. Schweiz Ärzteztg 96 (5051). https://doi.org/10.4414/saez.2015.04207
10. Thiel R, Deimel L, Schmidtmann D et al. (2018) #SmartHealthSystems. Digitalisierungsstrategien im internationalen Vergleich. Studienbericht. Bertelsmann Stiftung, Gütersloh
11. Office fédéral de la communication OFCOM. Direction opérationnelle Suisse numérique (2018) Stratégie Suisse numérique. OFCOM, Biel/Bienne
12. Office fédéral de la communication OFCOM. Direction opérationnelle Suisse numérique (2020) Stratégie Suisse numérique. https://www.bakom.admin.ch/dam/bakom/fr/dokumente/informationsgesellschaft/strategie/strategie_digitale_schweiz.pdf.download.pdf/Strategie-DS-2020-FR.pdf. Accédé: 4 avril 2022
13. Geneviève LD, Martani A, Mallet MC, Wangmo T, Elger BS (2019) Factors influencing harmonized health data collection, sharing and linkage in Denmark and Switzerland: A systematic review. PLoS ONE 14:e0226015. https://doi.org/10.1371/journal.pone.0226015
14. Specker M (2019) Die Seele im digitalen Zeitalter. In: Specker Manuela (Hrsg) Sozialalmanach 2019 Digitalisierung – und wo bleibt der Mensch? Caritas Verlag, Luzern, S 53–72
15. OECD (2019) Health in the 21st Century: Putting Data to Work for Stronger Health Systems. OECD Health Policy Studies. OECD Publishing, Paris, https://doi.org/10.1787/e3b23f8e-en
16. Oderkirk J (2021) Survey results: National health data infrastructure and governance. OECD Health Working Papers, No. 127. OECD Publishing, Paris, https://doi.org/10.1787/55d24b5d-en
17. eHealth Suisse, Centre de compétences et de coordination de la Confédération et des cantons (2021) Bases pour des processus interopérables dans le système de soins de santé. Contribution aux travaux stratégiques de la Confédération et des cantons. https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/F/rapport_interoperabilite.pdf. Accédé: 4. Avril 2022
18. Swiss Academy of Medical Sciences (2021) White Paper: Clinical Research. Swiss academies communications Vol. 16 (N°4). https://doi.org/10.5281/zenodo.4772725

19. Lovis C (2019) Stratégie de transparence dans le domaine des coûts et prestations de santé [Transparenzstrategie im Bereich der Gesundheitskosten- und Leistungen]. Rapport commandé par l’Office fédéral de la santé publique (nur auf Französisch erhältlich), Genève/Berne
20. Office Fédéral de la Statistique OFS, Division Santé et Affaires Sociales (inédit) Projet services de santé hospitaliers stationnaires (SpiGes). Programme Gestion national des données (NaDB). Factsheet, Avril 2021
21. Office fédéral de la santé publique OFSP (2022) Rapport concernant l’amélioration de la gestion des données dans le domaine de la santé. État de la mise en œuvre du mandat 8 du rapport concernant l’évaluation de la gestion de crise pendant la pandémie de COVID-19 (1re phase, février 2020 à août 2020), Berne. https://www.bag.admin.ch/bag/fr/home/das-bag/publikationen/bundesratsberichte.html
22. Basin D (2017) Risk Analysis on Different Usages of the Swiss AHV Number. Evaluation on behalf of the Federal Office of Justice and the Federal Data Protection and Information Commissioner. https://www.bj.admin.ch/bj/de/home/publiservice/publikationen/externe/2017-09-27.html. Accédé: 18 novembre 2021
23. Weber RH, Thouvenin F (2017) Gutachten zur Möglichkeit der Einführung eines Datenportabilitätsrechts im schweizerischen Recht und zur Rechtslage bei Personal Information Management Systems (PIMS), zuhanden des BAKOM, Center for Information Technology Society and Law (ITSL), Universität Zürich. https://www.aramis.admin.ch/Default?DocumentID=50028&Load=true
24. Technology and Privacy unit of the European Data Protection Supervisor (2020) Personal information management systems. EDPS TechDispatch Issue 3. https://doi.org/10.2804/11274
25. Maalouf E, De Santo A, Cotofrei P, Stoffel K (2020) Design principles of a central metadata repository as a key element of an integrated health nformation system. Swiss Learning Health System, Policy Brief. https://www.slhs.ch/en/policy-briefs-stakeholder-dialogues/our-topics/design-principles-of-a-central-metadata-repository/. Accédé: 15 mars 2022
26. De Santo A, Maalouf E, Cotofrei P, Stoffel K (forthcoming) Swiss Health Metadata Repository: implementation recommendations; Requirements, recommended architecture and prototype implementation. Swiss Learning Health System, Policy Brief
27. Bayer-Oglesby L, Bachmann N, Zumbrunn A et al. Linkage of Hospital and Census data to analyse the Social Dimension of Hospitalisations for Chronic Conditions: Methods and first results of the SIHOS-study on social inequalities in utilization and outcome of in-patient health care in Switzerland. Conference on Equity in Healthcare: Measure - document - understand, Bern, 15. October 2019
28. Bayer-Oglesby L, Bachmann N, Zumbrunn A (2020) Situation sociale et hospitalisations pour maladies chroniques (Obsan Rapport 11/2020). Observatoire suisse de la santé (Obsan), Neuchâtel. https://www.obsan.admin.ch/sites/default/files/2021-08/obsan_11_2020_rapport-corr.pdf
29. Martani A, Egli P, Widmer M, Elger B (2020) Data protection and biomedical research in Switzerland: setting the record straight. Swiss Med Wkly 150:w20332. https://doi.org/10.4414/smw.2020.20332
30. FORS and linkhub.ch (2020) Accessing and linking data for research in Switzerland. Report commissioned by the Swiss Academy for Humanities and Social Sciences (SAGW) and the Swiss Academies of Arts and Sciences. https://linkhub.ch/report/. Accédé: 18 november 2021
31. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E (2018) Considerations for ethics review of big data health research: A scoping review. PLoS ONE 13:e0204937. https://doi.org/10.1371/journal.pone.0204937
32. Vayena E, Dzenowagis J, Brownstein JS, Sheikh A (2018) Policy implications of big data in the health sector. Bull World Health Organ 96:66–68. https://doi.org/10.2471/BLT.17.197426
33. Geneviève LD, Martani A, Perneger T, Wangmo T, Elger BS (2021) Systemic Fairness for Sharing Health Data: Perspectives From Swiss Stakeholders. Front Public Health 9:669463. https://doi.org/10.3389/fpubh.2021.669463

34. Beer-Borst S (2020) Situationsanalyse zur Nutzung von Gesundheitsdaten in der Schweiz. Bericht über eine Online-Umfrage des Bundesamts für Gesundheit BAG (November 2019 bis Januar 2020), Bern (disponible uniquement en allemand)
35. Golliez A Was sind Datenräume und wozu brauchen wir sie? Netzwoche Nr.17/2020. https://www.netzwoche.ch/meinungen/2020-12-07/was-sind-datenraeume-und-wozu-brauchen-wir-sie. Accédé: 18 novembre 2021
36. Brall C, Berlin C, Zwahlen M, Ormond KE, Egger M, Vayena E (2021) Public willingness to participate in personalized health research and biobanking: A large-scale Swiss survey. PLoS ONE 16:e0249141. https://doi.org/10.1371/journal.pone.0249141
37. Martani A, Geneviève LD, Elger B, Wangmo T (2021) 'It’s not something you can take in your hands'. Swiss experts’ perspectives on health data ownership: an interview-based study. BMJ Open 11:e045717. https://doi.org/10.1136/bmjopen-2020-045717
38. Hutchings E, Loomes M, Butow P, Boyle FM (2020) A systematic literature review of researchers' and healthcare professionals' attitudes towards the secondary use and sharing of health administrative and clinical trial data. Syst Rev 9:240. https://doi.org/10.1186/s13643-020-01485-5
39. Rivas Velarde MC, Tsantoulis P, Burton-Jeangros C, Aceti M, Chappuis P, Hurst-Majno S (2021) Citizens' views on sharing their health data: the role of competence, reliability and pursuing the common good. BMC Med Ethics 22:62. https://doi.org/10.1186/s12910-021-00633-3
40. Hutchings E, Loomes M, Butow P, Boyle FM (2020) A systematic literature review of health consumer attitudes towards secondary use and sharing of health administrative and clinical trial data: a focus on privacy, trust, and transparency. Syst Rev 9:235. https://doi.org/10.1186/s13643-020-01481-9
41. Apel D, Behme W, Eberlein R, Merighi C (2015) Datenqualität erfolgreich steuern. Praxislösungen für Business-Intelligence-Projekte, 3., überarb. und erw. Aufl. Safari Tech Books Online. dpunkt.verl., Heidelberg
42. Ethical, Legal and Social Implications advisory group (2018) Ethical Framework for Responsible Data Processing in Personalized Health Research. Version 2. https://sphn.ch/wp-content/uploads/2019/11/Ethical_Framework_20180507_SPHN.pdf. Zugegriffen: 18. November 2021
43. Mooney SJ, Pejaver V (2018) Big Data in Public Health: Terminology, Machine Learning, and Privacy. Annu Rev Public Health 39:95–112. https://doi.org/10.1146/annurev-publhealth-040617-014208
44. Widmer T, Frey K, Eberli D, Schläpfer B, Rickenbacher J (2019) Evaluation des Humanforschungsgesetzes (HFG). Schlussbericht (uniquement disponible en allemand), Institut für Politikwissenschaft der Universität Zürich und KEK-CDC Consultants, Zürich
45. Enderle M (2019) Évaluation de la loi relative à la recherche sur l’être humain (LRH) : position de l’Office fédéral de la santé publique. https://www.bag.admin.ch/dam/bag/fr/dokumente/biomed/forschung-am-menschen/evaluationhfg/evaluation-hfg-stellungnahme-bag.pdf.download.pdf/StellungnahmeBAG_20191114_def_FR.pdf. Accédé: 4 avril 2022
46. Frauenhofer-Institut für Software- und Systemtechnik ISST (2021) ViCon – Virtueller Consent Assistent. Mensch-Technik-Interaktion für digitale Souveränität. Sammlung der Projektsteckbriefe zum Förderschwerpunkt des BMBF. https://www.isst.fraunhofer.de/de/news/news/ViCon.html. Accédé: 18 novembre 2021
47. Geneviève LD, Martani A, Wangmo T et al. (2019) Participatory Disease Surveillance Systems: Ethical Framework. J Med Internet Res 21:e12273. https://doi.org/10.2196/12273
48. Trapp M, Naab M, Rost D, Nass C, Koch M, Rauch B (2020) Digitale Ökosysteme und Plattformökonomie: Was ist das und was sind die Chancen? https://www.informatik-aktuell.de/management-und-recht/digitalisierung/digitale-oekosysteme-und-plattformoekonomie.html. Accédé: 18 novembre 2021

49. Trapp M, Naab M, Rost D, Nass C, Koch M, Rauch B (2020) Digitale Ökosysteme: Welche Herausforderungen stellt der Aufbau und wie gelingt er? https://www.informatik-aktuell.de/management-und-recht/digitalisierung/digitale-oekosysteme-welche-herausforderungen-stellt-der-aufbau-und-wie-gelingt-er.html. Accédé: 18 novembre 2021
50. Frauenhofer-Institut für Experimentelles Software Engineering IESE (2018) Die DSGVO und was sie für Digitale Dienste bedeutet: Digitale Ökosysteme und Plattformökonomie – Datensouveränität in der Praxis. Blog-Serie des Frauenhofer IESE zur DSGVO, Teil 4. https://www.iese.fraunhofer.de/blog/digitale-oekosysteme-und-plattformoekonomie-datensouveraenitaet-in-der-praxis/. Accédé: 18 novembre 2021
51. Geissbuhler A (2013) Lessons learned implementing a regional health information exchange in Geneva as a pilot for the Swiss national eHealth strategy. Int J Med Inform 82:e118-24. https://doi.org/10.1016/j.ijmedinf.2012.11.002
52. Office fédéral de la statistique OFS (2021) Gestion national des données. NaDB Information détaillées. https://www.bfs.admin.ch/bfs/fr/home/nadb/nadb.assetdetail.18004288.html. Accédé: 4. Avril 2022
53. Desai T, Ritchie F, Welpton R (2016) Five Safes: designing data access for research. University of the West of England (Faculty of Business and Law) Economics Working Paper Series 1601. http://rsss.cass.anu.edu.au/sites/default/files/rsss/Ritchie_5safes.pdf
54. Académie Suisses des Sciences Médicales (ASSM) et Commissions d’éthique suisses relative à la recherche sur l’être humain Factsheet : Modèle pour le Consentement général version 1/2017. https://swissethics.ch/doc/swissethics/general_consent/Factsheet_GK_f.pdf. Accédé: 4 avril 2022
55. Krutzinna J, Taddeo M, Floridi L (2019) Enabling Posthumous Medical Data Donation: An Appeal for the Ethical Utilisation of Personal Health Data. Sci Eng Ethics 25:1357–1387. https://doi.org/10.1007/s11948-018-0067-8
56. Krutzinna J, Floridi L (Hrsg) (2019) The Ethics of Medical Data Donation. Philosophical Studies Series, Band 137. Springer International Publishing, Cham

7 Glossaire

Terme	Explication	Source (référence)
Administration des données (*data governance*)	La gestion étendue de la disponibilité, de l'utilisation (utilité), de l'intégrité et de la sécurité des données utilisées dans un système. L'administration des données constitue donc le *cadre* pour une gestion efficace des données dans les « processus inhérents aux entreprises », qui prescrit des règles et des lignes directrices pour la gestion des données.	What Is Data Governance and Why Does It Matter? (techtarget.com) Krotova A, Eppelheimer J. Was bedeutet Data Governance? Eine Clusteranalyse der wissenschaftlichen Literatur zu Data Governance
Assurance-qualité et amélioration/encouragement de la qualité **(Développement de la qualité)**	L'assurance-qualité est le volet de la gestion de la qualité qui vise à créer la confiance pour que les exigences posées à la qualité (ici la qualité des soins) soient remplies. L'amélioration de la qualité (encouragement de la qualité, ici des prestations) est le volet de la gestion de la qualité qui vise à renforcer la capacité à remplir les exigences posées à la qualité. La *qualité des soins* se réfère pour l'essentiel à la *qualité des prestations fournies* dans le système de santé. Sont concernés les domaines du dépistage et du diagnostic, de la thérapie/traitement (médecine curative), de la prise en charge et des soins, de la prévention et de la promotion de la santé. Les prestations remboursées par l'assurance-maladie doivent être efficaces, appropriées et économiques (EAE). À la suite de l'entrée en vigueur de la modification de la LAMal le 1er avril 2021, le Conseil fédéral a pour tâche de fixer tous les quatre ans les objectifs en matière de garantie et d'encouragement de la qualité des prestations (développement de la qualité) (art. 58 LAMal). Pour réaliser ses objectifs dans le domaine du développement de la qualité, il institue une commission (Commission fédérale pour la qualité). Les fédérations des fournisseurs de prestations et des assureurs concluent des conventions relatives au développement de la qualité (conventions de qualité) valables pour l'ensemble du territoire suisse. Des données sont collectées et des indicateurs de qualité fixés pour mesurer la qualité.	DIN EN ISO 9000:2015 3.3.6 DIN EN ISO 9000:2000 Développement de la qualité en Suisse (admin.ch) OFSP, Stratégie qualité de la Confédération dans le système de santé suisse (en allemand), 09.10.2009 Loi fédérale sur l'assurance-maladie (LAMal ; RS 832.10), art. 58, 58*a* à 58*c*
Biobanque	Institution organisée avec une structure et une réglementation de direction (gouvernance) qui est chargée de la gestion des ressources biologiques. Les biobanques sont non seulement destinées à la recherche, mais également à des fins diagnostiques et thérapeutiques. À cet égard, elles contiennent des échantillons d'origine humaine et non humaine, par exemple des liquides, des tissus, des bactéries ou d'autres types d'échantillons. Ces échantillons sont reliés à des données qui décrivent l'historique des échantillons, la méthode de prélèvement, le traitement des échantillons et les résultats d'analyse. Ils sont utilisés dans le cadre de projets de recherche pour le développement de nouveaux traitements ou de nouvelles technologies.	What is a Biobank? – Swiss Biobanking Platform
Biomédecine	Domaine interdisciplinaire à la limite de la médecine et de la biologie. La biomédecine combine des contenus et des questions de la médecine expérimentale avec des méthodes de la biologie moléculaire (génétique) et de la biologie cellulaire.	Biomedizin - Lexikon der Biologie (spektrum.de) Biomédecine – Wikipedia
Cancer d'intervalle	Cancer du sein découvert dans l'intervalle de deux ans qui sépare les mammographies.	https://www.swisscancerscreening.ch
Data Lake	Système local centralisé d'enregistrement des données dans une organisation. Dans un *Data Lake*, les données brutes sont stockées dans le format d'origine et de manière non structurée, sans but d'utilisation prévue.	Luber S, Litzel N. Definition: Was ist ein Data Lake? BigData Insider 15.2.2018 Was ist ein Data Lake? (bigdata-insider.de)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Terme	Explication	Source (référence)
Data Warehouse	Système local centralisé d'enregistrement des données dans une organisation. Un *Data Warehouse* archive des données qui sont préparées dans des formats et des structures pour certaines applications ou analyses.	Luber S, Litzel N. Definition: Was ist ein Data Lake? BigData Insider 15.2.2018 (comparaison Data Lake et Data Warehouse)
Écosystème, numérique	Il s'agit d'un système sociotechnique qui comprend des bases techniques sous la forme d'une plateforme numérique ainsi que les organisations et les personnes impliquées dans le système et leurs liens.	https://www.informatik-aktuell.de/management-und-recht/digitalisierung/digitale-oekosysteme-und-plattformoekonomie.html
Espace de données	Un espace de données permet aux acteurs des données (fournisseurs de données, utilisateurs de données et personnes concernées) de collaborer simplement et équitablement. Cela se fait sur la base de droits en matière de données et de principes éthiques inscrits dans la loi, de processus et d'infrastructures fiables. Les espaces de données sont envisageables à différents niveaux géographiques, du local à l'international, mais aussi possibles dans certains secteurs économiques, par exemple celui de la santé, ou par-delà plusieurs secteurs économiques.	Golliez A. Netzwoche Mo 07.12.2020
Étendue de l'information	Quantité de thèmes	
Gestion des données	Il est procédé avec les ressources de données existantes au sens d'un bien économique. Le traitement des données est planifié et piloté en tenant compte de la quantité et de la qualité des données, des aspects spatiaux et, par conséquent, des aspects de la sécurité dans le cadre de l'accès, de la transmission, du traitement et du stockage des données.	
Infrastructure de recherche	Critères remplis par une infrastructure de recherche : – l'infrastructure de recherche fournit une contribution importante au développement d'un domaine de recherche (« plus-value » scientifique); – l'infrastructure de recherche est utilisée intensivement par les chercheurs en Suisse (importance nationale) ; – l'accès à l'infrastructure de recherche est réglé pour les communautés de chercheurs nationales et internationales ; – l'infrastructure de recherche peut se situer dans un seul emplacement ou s'organiser au sein d'un réseau comprenant plusieurs sites et une structure de gestion centralisée ; – une infrastructure de recherche ne poursuit pas en premier lieu une recherche autonome mais est à disposition des chercheurs pour leurs projets.	Feuille de route du SEFRI pour les infrastructures de recherche 2019.
Pilotage et soutien aux processus du système de santé	Il s'agit de la réglementation de l'offre de services de santé et du marché des assurances. Pour ce faire, la Confédération (Secrétariat général DFI, OFSP, OFS et obsan) et les cantons (CDS) collaborent au sein du dialogue « Politique nationale de la santé » (PNS).	Académie Suisse des Sciences Médicales (2019) Développement durable du système de santé. https://doi.org/10.5281/zenodo.2222811 Dialogue « Politique nationale de la santé » : plateforme de la Confédération et des cantons (admin.ch)
Profondeur de l'information	Degré de détail de l'information	
Protection des données (*privacy*)	Il s'agit de la protection de la personnalité. Les lois et ordonnances règlent la gestion des données se rapportant à des personnes. Il s'agit d'une protection « théorique » des données personnelles contre les abus.	Protection des données (admin.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Terme	Explication	Source (référence)
Protection des données dès la conception (*Privay by Design*)	Cf. art. 25, par. 1, RGPD et art. 7, al. 1 à 2, nLPD Les données personnelles des personnes concernées doivent être ainsi directement protégées dès le début. Lors du développement d'un logiciel ou d'un matériel qui peut traiter des données, il est nécessaire de tenir compte de l'aspect de la protection des données dès la conception. Cela concerne surtout le développement de logiciels. La protection des données dès la conception peut par exemple être appliquée dès le stade du développement par des mesures techniques et organisationnelles correspondantes telles que la pseudonymisation ou l'anonymisation.	Art. 25 EU-DSGVO: Datenschutz durch Technikgestaltung (datenschutz-experte.de) Privacy by Design & Privacy by Default (datenschutzexperte.de) FF **2020** 7397 - Loi fédérale sur la protection des données (LPD) (admin.ch) Glossaire « Suisse numérique » et note de discussion e-ID
Protection des données par défaut (*Privay by Default*)	Cf. art. 25, par. 2, RGPD et art. 7, al. 1 et 3, nLPD Les données personnelles des personnes concernées doivent être ainsi directement protégées dès le début. Il s'agit de la *protection des données comme réglage par défaut*. Les réglages de service, des systèmes ou des appareils par défaut (réglages par défaut) doivent être appliqués de telle sorte à veiller au mieux à la protection des données. Il s'agit aussi de protéger la sphère privée et les données personnelles des utilisateurs (personnes concernées) qui ne sont pas très technophiles et ne peuvent le cas échéant pas changer les paramètres de confidentialité à leur convenance. Les réglages par défaut peuvent être adaptés individuellement par les utilisateurs (fonction *opt in* / de consentement), l'accord devant être demandé activement par le prestataire.	Art. 25 EU-DSGVO: Datenschutz durch Technikgestaltung (datenschutz-experte.de) Privacy by Design & Privacy by Default (datenschutzexperte.de) FF **2020** 7397 - Loi fédérale sur la protection des données (LPD) (admin.ch) Glossaire « Suisse numérique » et note de discussion e-ID
Qualité dans le domaine de la santé	La qualité est une mesure pour déterminer dans quelle mesure les prestations du système de santé pour les individus et les groupes de population a) augmentent la probabilité des résultats souhaités en matière de santé et b) correspondent aux connaissances spécialisées actuelles.	Définition de l'Institute of Medicine (IOM), utilisée dans : OFSP, la Stratégie qualité de la Confédération dans le système de santé suisse (en allemand), 09.10.2009
Recherche biomédicale	La recherche biomédicale regroupe : - l'étude de maladies et de pathologies (physiques ou mentales) données, y compris le dépistage, l'étiologie, la prévention, le traitement et la réadaptation des personnes ; - la conception de méthodes, de médicaments ou de dispositifs permettant de poser un diagnostic, d'apporter une aide ou de soulager l'individu pendant et après le traitement de la maladie ou de l'affection donnée ; - la recherche scientifique centrée sur la compréhension des processus vitaux impliqués dans les maladies et le bien-être humain, y compris des domaines spécialisés sur les origines cellulaires et moléculaires des pathologies, la génétique ou l'immunologie. Une liste complète de telles activités inclut les essais cliniques et les examens de laboratoire, l'étude des problèmes engendrés par des polluants environnementaux et différents comportements à risque.	Le rapport sur les mesures de la Confédération afin de renforcer la recherche et la technologie biomédicales (EDI 18.12.2013) utilise la définition de l'OCDE : The OECD Glossary of Statistical Terms
Recherche clinique	Elle joue un rôle clé dans la prise en compte de nouvelles méthodes de thérapie et de diagnostic dans le quotidien des soins médicaux. Pour ce faire, elle travaille directement sur les patients. Les études cliniques sont le principal instrument visant à établir un lien entre la recherche fondamentale biomédicale et le quotidien des soins. Elles doivent démontrer scientifiquement l'efficacité et la sécurité des nouvelles méthodes.	https://www.gesundheitsforschung-bmbf.de/de/klinische-forschung-6456.php

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Terme	Explication	Source (référence)
Recherche épidémiologique	Elle porte sur la fréquence, la répartition et les causes des maladies ou des troubles de la santé dans des populations définies dans des conditions environnementales réelles. Elle étudie les conditions d'une bonne santé et les différents facteurs d'influence sur la genèse, l'évolution et les conséquences de maladies ainsi que les possibilités de prévention. Les types d'études sont les études d'observation (études transversales, études de cohorte, études cas-témoins, études écologiques) et les études d'intervention.	Composé à partir de différentes sources
Recherche sur les services de santé (*health services research*)	Elle analyse les structures, les processus et l'efficacité des soins dans les conditions de la vie de tous les jours et cherche ainsi des solutions pour un système de santé de qualité. Elle s'adresse aux patients, mais aussi à la population en bonne santé (prévention, promotion de la santé). Les domaines principaux sont l'accès aux services de santé, leur adéquation et leurs coûts, l'orientation vers le patient et la qualité des résultats. La recherche sur les services de santé fournit ainsi des informations scientifiques et indépendantes d'intérêts particuliers qui servent aux citoyens, aux fournisseurs de prestations et aux décideurs de la politique et de l'économie.	Recherche sur les services de santé (assm.ch) : voir crameri_masterarbeit_mph.pdf
Sécurité des données (*security*)	Il s'agit de la protection des moyens et de l'utilisation des données de toute nature, c'est-à-dire pas uniquement des données se rapportant à des personnes. L'intégrité, la disponibilité, la confidentialité et l'authenticité des données sont assurées par des solutions techniques ou des mesures organisationnelles, personnelles ou physiques et locales. Il s'agit d'une protection « pratique » des données de toute nature.	Unterschied Datenschutz Datensicherheit | Datenschutzexperte Datensicherheit: Was bedeutet das? | Datenschutz 2021
Statistique, statistique des services de santé	Elle concerne l'exploitation à des fins statistiques de données administratives, de registres et de données fournies par les réseaux d'observations et de mesures. La statistique des services de santé comprend notamment des informations sur les structures des fournisseurs de prestations (structures d'exploitation, infrastructure, personnel), sur les diagnostics, les prestations fournies, le recours à certaines prestations, les itinéraires cliniques des patients, les coûts d'exploitation ainsi que la formation et la formation continue du personnel. Il s'agit en outre d'observer le développement du système de santé suisse et de fournir les bases pour la planification des soins, la surveillance épidémiologique de la population, la science et l'information du public.	Ordonnance concernant l'organisation de la statistique fédérale, RS *431.011*, art. 3, al. 2, let. d OFS, Statistique des services de santé – La statistique des soins ambulatoires dans le cadre du projet MARS, V6.2, 2016.
Systèmes de gestion des informations personnelles (*Personal Information Management System* PIMS)	Ce sont des systèmes de collecte et de gestion centralisées des propres données personnelles qui sont axés sur une future utilisation de ces données. Il peut s'agir d'exploitations de données à des fins personnelles, ou des personnes concernées mettent par exemple leurs données médicales à la disposition de la recherche scientifique ou de fournisseurs de services commerciaux. Parmi les éléments conceptuels, il est question de permettre aux personnes concernées le contrôle et le suivi de l'accès.	Weber RH, Thouvenin F (2017) Gutachten zur Möglichkeit der Einführung eines Datenportabilitätsrechts im schweizerischen Recht und zur Rechtslage bei PIMS, zuhanden des BAKOM, Center for Information Technology Society and Law (ITSL), Universität Zürich.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

8 Annexes

Annexe 1 Réutilisation de données médicales par des tiers

Si des données ou des jeux de données sont transmis à d'autres personnes et institutions, c'est-à-dire partagées avec d'autres pour une utilisation commune, on utilise également le terme d'échange de données (*data sharing*). Le partage de données avec des tiers présuppose en général que (a) les tiers connaissent déjà des collections de données existantes pour une utilisation commune et peuvent y avoir accès, (b) que les collections de données répondent à des exigences de qualité et que les jeux de données sont échangeables sur le plan formel et technique, et (c) que la fusion avec d'autres jeux de données peut se faire avantageusement dans le respect des réglementations légales en vigueur. Les destinataires des données peuvent réutiliser les données médicales reçues avec ou sans appariement.

Réutilisation sans appariement

Une réutilisation de données sans appariement est possible de deux manières. D'une part, un jeu de données peut être préparé *dans la forme transmise* pour répondre à une problématique spécifique et directement utilisé pour les exploitations correspondantes (cf. figure A1-1).

Figure A1-1 Réutilisation de données sans appariement : un jeu de données (représentation schématique).

Jeu de données A (de tiers)

Personne	Variable				
Numéro	**A**	**…**	**X**	**Y**	**Z**
NA1	18	…	11	98	25
NA2	67	…	23	76	109
…	…	…	…	…	…
NAn	25	…	24	87	12

→

Jeu de données (partiel) A pour analyse

Personne	Variable		
Numéro	**A**	**X**	**Y**
NA1	18	11	98
NA2	67	23	76
…	…	…	…
NAn	25	24	87

Exemples :

Un étudiant utilise le jeu de données du Registre suisse du cancer de l'enfant, collecté pour le monitorage national des cancers chez l'enfant, pour analyser l'occurrence d'un certain type de cancer dans plusieurs régions de Suisse.

Un groupe de recherche utilise le jeu de données anonymisé de l'enquête nationale sur l'alimentation, menuCH[103], réalisée sur mandat de l'Office fédéral de la sécurité alimentaire et des affaires vétérinaires (OSAV), pour étudier la consommation de viande de la population.

D'autre part, le destinataire peut fusionner/combiner le jeu de données qui lui est transmis avec un ou plusieurs autres jeux comparables (cf. figure A1-2). Un nouveau jeu de données est ainsi créé, qui comporte des informations sur un plus grand nombre d'individus.

Figure A1-2 Réutilisation de données sans appariement : combinaison de deux jeux de données (représentation schématique).

Jeu de données A (de tiers)

Personne	Variable	
Numéro	**X**	**Y**
A1	11	98
A2	23	76

+

Jeu de données B (propre jeu ou jeu de tiers)

Personne	Variable	
Numéro	**X**	**Y**
B1	45	37
B2	65	---

→

Jeu combiné (pour la réutilisation)

Personne	Variable	
Numéro	**X**	**Y**
A1	11	98
A2	23	76
B1	45	37
B2	65	---

[103] Résultats concernant la consommation alimentaire (admin.ch) et Switzerland - National Nutrition Survey menuCH 2014-2015 (unisante.ch)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Les exploitations de plus grands échantillons fournissent des résultats plus solides et, en cas de diffusion des résultats, il est compliqué, dans l'idéal impossible, d'identifier les personnes. Cette démarche est intéressante quand il n'y a certaines informations de santé que sur quelques individus. Le nouveau jeu de données, plus important, permet une utilisation plus large sur le plan thématique et divergente du but initial de la collecte.

Exemples : création d'un jeu de données à partir de plusieurs petits jeux de différentes cliniques ou centres spécialisés.

Un consortium de recherche fusionne des données de personnes souffrant d'une maladie rare (p. ex. syndrome de Dravet), collectées initialement à des fins de traitement stationnaire et ambulatoire dans plusieurs institutions. La transmission permet de réutiliser les données dans leur globalité pour répondre à des questions sur le diagnostic (occurrence) et le traitement de cette maladie.

Dans le cadre de la Stratégie Sel de la Confédération (OSAV), on s'intéresse à une méthode simple mais valide pour déterminer la consommation de sel de la population. L'excrétion de sodium dans l'urine de 24 heures sert de méthode de référence (très fastidieuse), l'excrétion de natrium dans l'urine spot comme alternative. Il s'agit de déterminer si les deux méthodes livrent des résultats comparables. On veut éviter une étude multicentrique compliquée et on cherche donc des données du plus de personnes possible dans différents centres de soins et de recherche en Suisse qui, dans le cadre d'études, ont appliqué en parallèle les deux méthodes de collecte selon un protocole standard. Les résultats de laboratoire (méthodes d'analyse standardisées) et d'autres données nécessaires sur le sexe, l'âge, la région de domicile, la médication, les maladies, etc. peuvent être compilés de manière centralisée pour une étude comparative des méthodes (étude de validation).

Certains assureurs proposent à leurs assurés une participation à prix réduit à des programmes sportifs spéciaux. Un groupe de recherche dans le domaine de la prévention et de la promotion de la santé s'intéresse à la question de savoir si la participation diffère selon les régions, les saisons, l'âge et le sexe. À cet effet, il demande les données nécessaires à plusieurs assureurs pour les évaluer en commun.

Mais la démarche est également intéressante pour étudier l'occurrence de maladies transmissibles et chroniques à l'échelle suprarégionale/nationale et internationale.

Exemple : comparaison internationale

Les registres des tumeurs de plusieurs pays, dont la Suisse, transmettent certaines de leurs données au « Global Cancer Observatory » du Centre international de recherche sur le cancer (CIRC/OMS)[104] à des fins de réutilisation.

Réutilisation avec appariement

Lorsque des jeux de données sont partagés et fusionnés avec d'autres jeux pour lesquels il s'agit en tout ou en partie de données sur les mêmes personnes, ils peuvent être appariés. Le terme d'appariement de données (*record linkage*) est utilisé lorsque les *données individuelles des mêmes personnes* contenues dans deux jeux de données au moins sont fusionnées à l'aide d'une ou plusieurs variables d'appariement (*linkage variables*). Il en résulte un nouveau jeu de données augmenté dans l'étendue et/ou la profondeur de l'information par personne (*linked data set*).
Les caractéristiques personnelles servent de variables d'appariement. Il y a des caractéristiques qui identifient clairement une personne, à l'instar du numéro d'assurance sociale NAVS13. De tels numéros univoques sont appelés « identificateurs ». D'autres caractéristiques ne permettent d'identifier une personne qu'en partie. Elles peuvent s'appliquer à plusieurs personnes (p. ex. nom, date de naissance, nationalité), changer avec le temps (p. ex. numéro postal d'acheminement, domicile), permettre différentes graphies, être écrites de manière erronée ou incomplète (p. ex. seule l'année de naissance). Ces facteurs peuvent influer considérablement sur le résultat de l'appariement des données.

[104] Global Cancer Observatory (iarc.fr)

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

La figure A1-3 montre de manière schématique le principe d'un appariement de données. Des caractéristiques personnelles permettent d'établir s'il s'agit effectivement de jeux de données de la même personne ou de personnes différentes. Plus la ou les caractéristiques personnelles utilisées sont univoques et peu sujettes aux erreurs, plus les jeux des mêmes personnes peuvent être associés et appariés sur une base incontestable et efficace. Dans l'idéal, les jeux de données à apparier contiennent la même caractéristique univoque – un identificateur. Cela permettrait une utilisation de routine de l'appariement déterministe, qui est plus simple d'un point de vue méthodologique (pour la méthodologie, voir ci-dessous).

Figure A1-3 Transmission de données avec appariement (représentation schématique)

Jeu de données A (de tiers)

Personne			Variable
Numéro	**M*1**	**M2**	**X**
NA1	xxxa	12.01.34	98
NA2	yyyy	05.07.56	76
….	….	….	….
NA8	aaaa	15.04.67	53
NA9	bbbb	06.08.79	45
NAn	zzzz	01.09.60	17

∞

Jeu de données B (propre jeu ou jeu de tiers)

Personne			Variable
Numéro	**M1**	**M2**	**Y**
NB1	zzzz	09.01.60	1
NB2	aaaa	15.04.67	2
….	….	….	….
NB20	xxxx	12.01.34	2
NB21	bbbb	06.08.79	1
NBn	hhhh	10.10.65	3

↓

Jeu de données apparié avec une concordance incontestable

Personne			Variable	
Numéro	**M1**	**M2**	**X**	**Y**
V1	aaaa	15.04.67	53	2
V2	bbbb	06.08.79	45	1
….	….	….	….	….

***M** = caractéristique personnelle, variable(s) d'appariement

Exemples :

Mortalité à 30 jours suivant l'intervention à l'hôpital (indicateur clé pour la qualité des établissements traitants) : pour calculer les taux de mortalité dans les 30 jours suivant une intervention, il est nécessaire de tenir compte des décès *pendant* mais aussi *après* les hospitalisations. Les décès pendant le traitement stationnaire sont enregistrés dans les hôpitaux. Un appariement des données de la Statistique médicale des hôpitaux (MS) avec les données sur les décès dans la population (Statistique du mouvement naturel de la population BEVNAT) permettrait de fusionner simplement les informations nécessaires sur les décès après la sortie de l'hôpital. Seule la BEVNAT comprend toutefois le NAVS13, qui identifie une personne de manière univoque. L'appariement des deux jeux de données doit donc se faire au prix d'une méthode alternative coûteuse, avec pour corollaire que l'appariement ne réussit que dans près de 60 % des cas. Ainsi, les résultats ne permettent pas de conclusions fiables. Le même défi existe du reste pour toutes les études à long terme ou les registres qui dépendent d'informations sur le statut vital. Dans ces cas, l'appariement avec la date de décès est même impossible la plupart du temps, car il manque l'identificateur univoque nécessaire en l'espèce.

Pour l'évaluation d'un programme cantonal de dépistage du cancer du sein, les données des patientes avec un diagnostic de cancer du sein enregistrées dans le registre cantonal des tumeurs sont appariées avec le jeu de données concernant la participation au programme de dépistage. En l'absence d'un identificateur univoque, les caractéristiques personnelles (prénom, nom, date de naissance, etc.) sont utilisées comme variables d'appariement. Parmi les participantes au programme de dépistage, il est ainsi possible d'identifier les patientes qui ont développé un cancer d'intervalle. Il est aussi possible d'identifier les patientes pour lesquelles un cancer du sein a été diagnostiqué dans le cadre du dépistage, mais qui n'étaient jusqu'ici pas connues du registre des tumeurs.

Mieux utiliser les données médicales pour assurer l’efficience et la qualité des soins

Deux méthodes d’appariement sont appliquées dans la pratique, seules ou combinées :

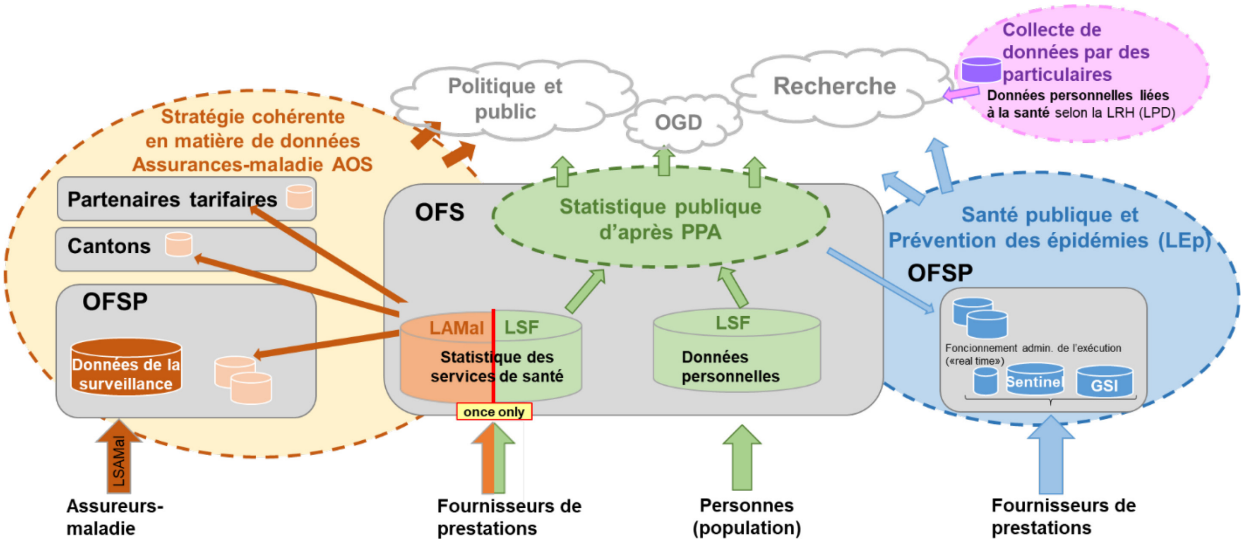
(1) L’**appariement déterministe**, également qualifié sans ambiguïté ou de hiérarchique. Cette méthode se fonde sur l’*établissement de la concordance univoque ou de l’absence de concordance* entre des caractéristiques/variables qui sont communes aux jeux de données source à apparier. Il peut par exemple s’agir du prénom et du nom. En raison du risque d’erreur décrit pour ce type de variables, il faut toutefois s’attendre à ce que la part de concordance erronée soit élevée et que le jeu de données apparié qui en résulte soit plutôt petit (pertes). Si les jeux de données source contiennent une caractéristique qui identifie clairement la personne, à l’instar du NAVS13, la concordance des jeux de données individuels peut être établie de manière nettement plus claire et avec une marge de doute presque nulle. Le nouveau jeu de données ainsi apparié sera plus complet.

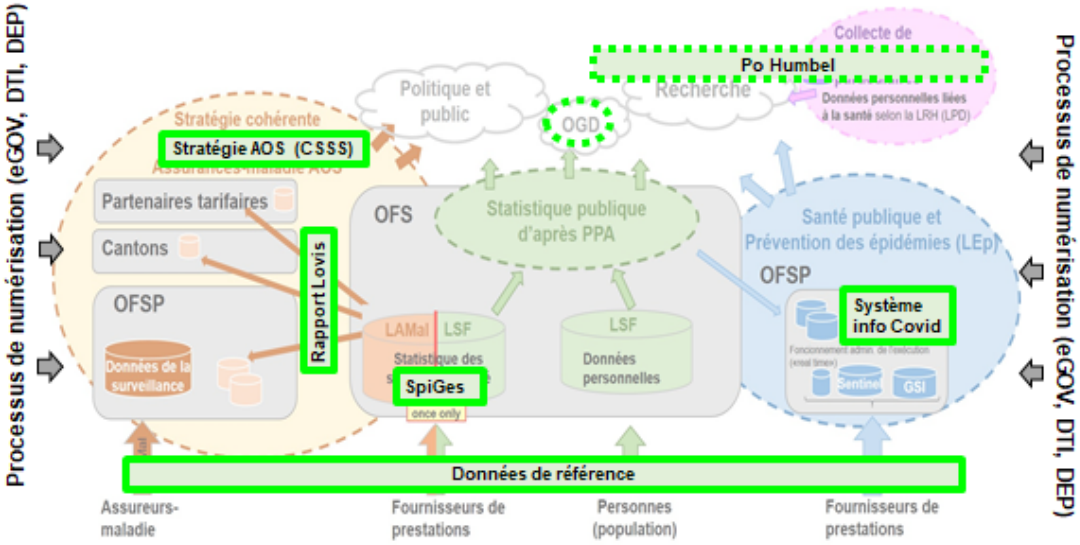
(2) L’**appariement probabiliste**. Dans cette méthode, les jeux de données source sont comparés à l’aide de plusieurs caractéristiques/variables personnelles, par exemple le prénom, le nom, le sexe et la date de naissance. Le type de concordance (totale, partielle, nulle) est établi et l’étendue de la concordance est calculée. En cas de concordance lacunaire, le meilleur appariement possible des jeux de données individuels est établi à l’aide des probabilités, calculées pour les différentes caractéristiques, d’observer une concordance. L’exactitude et le succès de l’appariement dépendent donc de la discrimination des caractéristiques personnelles disponibles, mais surtout de la qualité des données à apparier et de la conception de la stratégie d’appariement. Cette méthode est très exigeante et présuppose une grande expérience.

S’il faut s’attendre à ce qu’il n’y ait pas d’identificateur univoque (p. ex. NAVS13) pour toutes les personnes, la stratégie serait dans un premier temps de réaliser un appariement déterministe à l’aide de l’identificateur. La méthode probabiliste peut être appliquée dans un second temps, comme nous l’avons décrit ci-dessus, pour toutes les personnes pour lesquelles aucune concordance n’a été trouvée. Cela permet ainsi d’optimiser le résultat de l’appariement.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Annexe 2 Ancrage des projets visant une meilleure utilisation des données dans l'administration fédérale





Légende

Po Humbel	Postulat 15.4225 « Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins ». Concerne la réutilisation et l'exploitation appariée de données médicales par des tiers.
SpiGes	Projet pilote Secteur stationnaire hospitalier, réalisé dans le cadre du programme de gestion nationale des données (NaDB). Concerne la structure et les données sur les soins (stationnaires) des hôpitaux.
Rapport Lovis	Rapport du Pr Ch. Lovis sur mandat du DFI/OFSP, décembre 2019 [19]. Concerne les données de l'assurance-maladie au sens de la LAMal.
OGD	Statégie *Open Government Data* (OGD) 2019-2023
Stratégie AOS	Postulat CSSS-E 18.4102 « Stratégie de données cohérente pour le domaine de la santé », avec un accent sur l'assurance obligatoire des soins
eGOV, DTI, DEP	Administration numérique, transformation numérique, dossier électronique du patient

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Annexe 3 Projets d'appariement : défis et potentiel de solutions

Pour répondre aux questions de la politique sanitaire et sociale, les institutions privées et publiques dépendent de la réutilisation et de l'exploitation appariée de données. Une enquête de l'OFSP auprès des utilisateurs de données l'a confirmé [34]. Les exemples suivants montrent les défis de différents projets d'appariement.

A3-EXEMPLE 1 :

Projet	**Révision de la LAMal, financement hospitalier : projet partiel Qualité des prestations hospitalières stationnaires.**
Objectif du projet	Calcul des taux de mortalité dans les 30 jours suivant l'intervention. Calcul des taux de ré-hospitalisation et analyse des interfaces et de la durée d'hospitalisation (itinéraire du patient).
But de l'utilisation secondaire des données*	Information du système de santé
Sources des données appariées	Statistique médicale des hôpitaux (MS), statistique du mouvement naturel de la population / décès (BEVNAT) et statistique des hôpitaux (KS). Statistique médicale (MS) sur plusieurs années pour suivre l'itinéraire des patients dans le temps.
Défis	*Absence d'identificateur univoque commun (variable d'appariement)* : Le calcul des « véritables » taux de mortalité à 30 jours tient compte des décès pendant et après les hospitalisations. Pour cela, il fallait un appariement de la MS avec la BEVNAT sur la base du code de liaison anonyme CLA (OFS), car seule la BEVNAT comprend le NAVS13. La génération du CLA pour la BEVNAT était grevée d'erreurs. Une concordance/appariement ne pouvait être réalisée que pour 60 % des cas. Insuffisant pour un calcul ou des conclusions fiables. Le CLA était utilisé pour analyser le déroulement du traitement de certains patients dans la MS (p. ex. transferts des soins somatiques aigus à la réadaptation). Le CLA établi par les hôpitaux et transmis à l'OFS était généré de manière en partie erronée. Il a fallu choisir de coûteuses solutions de remplacement pour fournir des résultats d'une fiabilité appropriée.
Remarques **Potentiel de solutions**	Le statut vital est un indicateur clé pour considérer la santé de la population. L'appariement déficient ou impossible de la MS et d'autres sources de données (p. ex. données des registres, études à long terme) avec la date de décès est un problème fondamental que le système proposé permettrait de résoudre. La nécessité de reproduire l'itinéraire des patients est grande, par exemple pour les rapports de santé sur les maladies oncologiques ou dans la recherche sur le suicide. L'appariement nécessaire des données des enquêtes transversales pluriannuelles au niveau de l'individu pourrait se faire de manière systématique dans un système de réutilisation et d'exploitation appariée des données médicales.

*d'après le chap. 1.2.3, tableau 1

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

A3-EXEMPLE 2

Projet	**PNR74, étude SIHOS : Les inégalités sociales dans les soins stationnaires en Suisse.**
Objectif du projet	Étudier les rapports des caractéristiques du traitement stationnaire et du suivi en cas de maladies chroniques avec la situation sociale et l'appartenance à un groupe vulnérable. Pour ce faire, fusionner et apparier les données médicales et socio-économiques existantes dans une nouvelle banque de données de manière anonymisée.
But de l'utilisation secondaire des données	Recherche académique/universitaire
Sources des données appariées	Statistique médicale des hôpitaux (MS), statistique des institutions médico-sociales (SOMED), statistique du mouvement naturel de la population (BEVNAT), relevé structurel (RS). Également nécessaire : statistique de la population et des ménages (STATPOP) et banque de données UPI de la CdC plus cadre/registre d'échantillonnage.
Défis	*Absence d'identificateur univoque commun (variable d'appariement) et questions juridiques :* Il est nécessaire de choisir une « voie détournée pour l'identificateur », qui nécessite des ressources et présuppose une solution spéciale concernant l'utilisation du NAVS13. Des problèmes techniques, liés au CLA comme dans l'exemple 1, entraînent 40 % de pertes de données et compromettent la pertinence des résultats. *Pérennité du nouveau jeu de données* : La nouvelle collection de données appariée pourrait être poursuivie sous forme de « cohorte », mais elle doit être détruite.
Remarques **Potentiel de solutions**	Une demande de recherche a été acceptée, même si la faisabilité est incertaine. La clarification juridique et technique préalable avec des experts du futur organe national pour la coordination des données doit éviter de tels retards/interruptions de projets. De plus, la coordination centralisée des acteurs permettrait de soutenir les chercheurs et d'optimiser l'utilisation des ressources. Le potentiel des anciennes données et des données appariées/nouvelles serait exploité plus durablement.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

A3- EXEMPLE 3

Projet	**Étude PETROL* : Childhood cancer and vicinity of residence to petrol stations and major roads : a census-based nationwide cohort study.**
Objectif du projet	Clarification du lien entre les facteurs environnementaux à proximité du domicile et les cancers de l'enfant et de l'adolescent : exposition aux stations-service, aux garages automobiles et aux grandes routes.
But de l'utilisation secondaire des données	Recherche de l'administration
Sources des données appariées	Données du registre du cancer de l'enfant (enfants <16 ans), données du recensement (1990 et 2000) ou Swiss National Cohort avec des informations géocodées sur le domicile, statistique du mouvement naturel de la population (BEVNAT) – naissances vivantes.
Défis	Le groupe de recherche de l'ISPM de Berne est connu pour son expertise dans les appariements de données probabilistes (sans identificateur univoque). Le projet a donc pu être mis en œuvre sans problème majeur. À prendre en compte : L'appariement s'est fait avec plusieurs caractéristiques personnelles. Le sexe, la date de naissance, les dates de naissance des parents, le domicile géocodé au moment du recensement, la commune de domicile lors du recensement et, lors de la naissance, la nationalité. Un consentement à l'utilisation spécifique des données était nécessaire. La réalisation d'un suivi de longue durée de ces personnes exige d'autres considérations organisationnelles / juridiques.
Remarques **Potentiel de solutions**	Dans un système de réutilisation et d'exploitation appariée des données médicales, il est nécessaire d'impliquer les experts en Suisse et de leur offrir une organisation efficace et une infrastructure de travail sûre. L'utilisation de données se rapportant à des personnes pourrait être minimisée et l'appariement se faire avec un plus faible risque d'erreurs grâce à un identificateur univoque.

* L'étude réalisée pour l'OFSP indique que les enfants vivant à moins de 100 mètres d'une autoroute ont un risque accru de développer une leucémie. Ce lien était le plus fort pour les enfants de moins de cinq ans et les enfants qui n'ont pas changé de domicile.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

A3- EXEMPLE 4

Projet	**Informative indicators on healthcare equity in Switzerland (2020-2021)**
Objectif du projet	Créer des bases pour garantir l'égalité des chances dans le système de santé et lors du traitement. Identifier un gradient social potentiel à l'échelle nationale et cantonale.
But de l'utilisation secondaire des données	Recherche de l'administration
Sources des données appariées	Statistique médicale des hôpitaux (MS) avec des données agrégées au niveau MedStat sur la base de la statistique de la population et des ménages (STATPOP ; statut de séjour), relevé structurel (RS), statistique de l'aide sociale, statistique du chômage et autres indicateurs créés par MicroGIS.
Défis	*Absence d'identificateur univoque commun (variable d'appariement) et dispositions de la protection des données* Un appariement au niveau individuel n'était pas faisable, ou que de manière limitée. Pour des raisons de protection des données, il aurait fallu renoncer à l'utilisation des données relatives au domicile. Mais le gradient social devait aussi être démontré dans les différents cantons. Sans informations sur le domicile, aucune connaissance sur les hospitalisations évitables (problèmes lors du traitement) n'aurait pu être générée dans certaines localités. Pour pouvoir conserver les informations sur le domicile, il était nécessaire d'utiliser les données agrégées au niveau des régions MedStat à la place des données individuelles dans les analyses. Il y a un risque de conclusions erronées. Sur la base de données agrégées, il n'est pas possible de conclure à la distribution d'une caractéristique chez les individus avec une certitude absolue. On ignore si les hospitalisations évitables concernent surtout les plus défavorisés dans les régions à forte exclusion sociale, car on ne connaît que la corrélation au niveau des régions MedStat. Les analyses étaient suffisantes pour se faire une première idée de l'égalité des chances dans les cantons. Des études plus approfondies sont initiées après coup dans les cantons. Il s'agit de déterminer dans quelle mesure on arrive à des réponses sur le plan quantitatif et s'il est nécessaire d'utiliser des méthodes qualitatives en complément.
Remarques **Potentiel de solutions**	Le travail fourni a permis de créer une bonne base pour des études de suivi. La recherche des problèmes dans le système de santé est toutefois démesurément longue et coûteuse pour représenter la situation chez les patients dans une certaine mesure. Il faut un système avec des structures protégées et la possibilité d'utiliser un identificateur univoque pour pouvoir cerner les points faibles dans le traitement des plus défavorisés en ménageant les ressources, initier ainsi des changements et en vérifier les effets à plusieurs reprises.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

A3- EXEMPLE 5

Projet	**Diverses questions liées au COVID-19**
Objectif du projet	Études sur les liens entre les antécédents médicaux, le lieu de l'infection, le diagnostic/moment du diagnostic, le traitement, les conséquences à long terme, y compris la mort, et la situation sociale
But de l'utilisation secondaire des données	Information du système de santé
Sources des données appariées	Données du système de déclaration, données du Contact Tracing, données sur la vaccination, données sur les tests Statistique du mouvement naturel de la population / décès (BEVNAT) et statistique des hôpitaux (KS). Statistique médicale des hôpitaux (MS), statistique de la population et des ménages (STATPOP), relevé structurel (RS), registre de la Centrale de compensation CdC (Finances et Registres centraux FRC), statistique de l'aide sociale, statistique du chômage, autres Pour ne nombreux chercheurs, le risque de conclusions erronées existe. Les responsables du projet tiennent compte des informations sur le voisinage/NPA, car ils n'ont pas de données disponibles/appariables au niveau individuel sur la précarité socio-économique.
Défis	*Absence d'identificateur univoque commun (variable d'appariement), questions liées à la protection des données, absence de consentement, qualité des données*
Remarques **Potentiel de solutions**	Analogue aux exemples 1, 2 et 4

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

A3- EXEMPLE 6

Projet	**Projet FNS BioLink de l'ISPM, Université de Berne**
Objectif du projet	Développement d'un instrument qui consulte, dans les sources de données, si, pour des personnes, des données de toutes les banques de données (ou en tout cas plus d'une) sont disponibles avec certains critères d'inclusion.
But de l'utilisation secondaire des données	Recherche académique/universitaire
Sources des données appariées*	Banque des tumeurs des pédiatres oncologues (à Zurich), Germline DNA Bank à Genève, Swiss Childhood cancer Survivor Study et registre du cancer de l'enfant
Défis	*Disponibilité des données, absence de consentement, absence d'identificateur univoque commun (variable d'appariement)* La réalisation de certaines études requiert un nombre minimal de jeux de données individuels. Lors d'appariements de données médicales de plusieurs sources, le degré d'exploitation – en particulier en cas d'appariement probabiliste sur la base de plusieurs caractéristiques permettant d'identifier des personnes – est souvent insuffisant pour pouvoir réaliser une étude avec la puissance statistique nécessaire. Les projets avec une procédure de demande complexe, y compris une demande d'autorisation d'accès aux banques de données par la ou les commissions d'éthique et, le cas échéant, une demande de consentement, ne peuvent alors pas être réalisés.
Remarques **Potentiel de solutions**	Le projet fournit un élément d'infrastructure/instrument potentiel pour un système de réutilisation et d'appariement des données médicales

* Pour la banque des tumeurs et la banque d'ADN, il y a un consentement signé à l'appariement de données.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Annexe 4 Bases légales pour la réutilisation et l'appariement de données

D'un point de vue juridique, la distinction entre les données personnelles et les données non personnelles est essentielle.
Dans la mesure où les **données médicales** sont considérées comme des **données personnelles**, leur traitement concerne le droit fondamental à la protection de la sphère privée et de l'autodétermination informationnelle des personnes concernées.

De tels traitements des données sont visés par la **législation** générale de la Confédération et des cantons **sur la protection des données** (loi fédérale sur la protection des données, LPD ; RS *235.1*, lois cantonales sur la protection des données). Celle-ci définit les **règles de base pour le traitement des données par les autorités et les particuliers**, mais ne règle pas les traitements concrets dans des domaines particuliers. Ce cas de figure est précisé dans le cadre de législations spéciales, comme par exemple la loi fédérale sur l'enregistrement des maladies oncologiques (LEMO ; RS *818.33*), entrée en vigueur en janvier 2020, dont l'un des buts prévus est notamment le soutien à la recherche[105].

La **réutilisation de données personnelles liées à la santé pour la recherche** est réglée dans la loi fédérale relative à la recherche sur l'être humain (**loi relative à la recherche sur l'être humain**, LRH ; RS *810.30*), aux art. 32 ss, et dans l'ordonnance relative à la recherche sur l'être humain à l'exception des essais cliniques (ordonnance relative à la recherche sur l'être humain, ORH ; RS *810.301*). Le **principe du consentement** est au cœur de la LRH. En d'autres termes, le consentement (*informed consent*) des personnes concernées doit être en principe obtenu pour la réutilisation de données pour un projet de recherche ou à des fins de recherche.

Pour la réutilisation de *données personnelles non génétiques liées à la santé sous une forme codée*, il suffit, en vertu de l'art. 32 ORH, que la personne concernée soit informée des points énumérés dans l'ordonnance et ne se soit pas opposée à la réutilisation. Pour certains cas exceptionnels justifiés, la LRH prévoit une autorisation de la commission d'éthique compétente pour remplacer un défaut de consentement de la personne concernée (art. 34 LRH).

La loi et l'ordonnance règlent en outre le devoir d'information et les modalités de l'opposition (voir en particulier l'art. 10 ORH) et définissent les devoirs de diligence en lien avec l'anonymisation, le codage et la conservation des données personnelles utilisées pour la recherche.

Les *données personnelles génétiques* ne peuvent être anonymisées à des fins de recherche que si la personne concernée a été informée sur l'anonymisation envisagée, son droit d'opposition, les conséquences de l'anonymisation sur les résultats concernant sa santé et la possibilité de transmettre des données à des tiers à des fins de recherche et qu'elle ne s'y est pas opposée (art. 30 ORH).

Les données liées à la santé collectées de manière entièrement anonyme ou anonymisées après la collecte *sont d'entrée exclues du champ d'application de la LRH* (art. 2, al. 2, let. c, LRH). L'anonymat présuppose que le lien entre une personne et ses données soit à ce point interrompu qu'il ne soit plus possible de le rétablir, ou alors seulement avec une charge excessive. Seule cette condition permet de garantir que les données restent hors du champ d'application de la LRH et des dispositions légales générales en matière de protection des données.

Des dispositions particulières s'appliquent à la **transmission de données médicales de la statistique fédérale publique (« données statistiques »)**. En vertu de l'art. 19, al. 2, de la loi sur la statistique fédérale (LSF ; RS *431.01*), celles-ci peuvent être notamment communiquées comme des *données personnelles à des fins ne se rapportant pas à des personnes comme la recherche, la planification et la statistique*, si :

- ces données sont rendues anonymes dès que le but du traitement le permet,
- le destinataire ne communique ces données à des tiers qu'avec l'accord de l'organe qui les a produites,
- la forme choisie par le destinataire pour communiquer les résultats ne permet pas d'identifier les personnes concernées et

[105] Cf. RS *818.33* - Loi fédérale du 18 mars 2016 sur l'enregistrement des maladies oncologiques (LEMO) (admin.ch), art. 2 et 23

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

- tout porte à croire que le destinataire respectera le secret statistique et les autres dispositions relatives à la protection des données.

Les directives internes de l'OFS relatives à la protection des données en cas de transmission de données individuelles à des tiers précisent cette *disposition potestative* de manière restrictive et relèvent que les données personnelles sensibles ne peuvent jamais être transmises sous une forme non anonymisée. *Les données médicales de la statistique publique ne peuvent par conséquent être transmises que sous une forme entièrement anonymisée et uniquement à des fins ne se rapportant pas à des personnes.*

À noter par ailleurs que les données de l'OFS liées à la santé peuvent aussi être utilisées à d'autres fins, comme le prévoit l'art. 59*a* de la loi fédérale sur l'assurance-maladie (LAMal ; RS *832.10*). Dans ce cadre, les données de l'OFS peuvent être réutilisées si elles sont utilisées pour les buts prévus par la LAMal. Mais l'utilisation et le traitement des données médicales à ces fins relevant du droit de la surveillance ne font explicitement pas l'objet du présent rapport (cf. chap. 1.2.3, Précisions conceptuelles, (3) Utilisation des données à des fins de recherche).

Les **conditions-cadres juridiques** mentionnées en lien avec une réutilisation des données s'appliquent également dans le cadre des **appariements de données**.

En Suisse, il manque une base légale pour un numéro d'identification univoque librement utilisable (« identificateur personnel ») dans le cadre d'appariements déterministes ou probabilistes. Le numéro d'assurance sociale[106] introduit en 2008 ne peut pas être utilisé pour les appariements dans le domaine de la recherche, car sa destination est limitée par la loi. En dehors du domaine des assurances sociales, il ne peut être utilisé qu'en vertu d'une base légale qui définit la destination et le cercle des bénéficiaires[107]. L'*utilisation des noms, prénoms et dates de naissance pour l'appariement*, à défaut, présuppose la *présence d'un* **consentement** de la personne concernée sur la base des conditions-cadres légales en vigueur.

Les **données de la statistique publique de la Confédération** peuvent être appariées par l'OFS sur la base de la LSF, de l'ordonnance concernant l'exécution des relevés statistiques fédéraux (ordonnance sur les relevés statistiques ; RS *431.012.1*)[108] et de l'ordonnance du DFI concernant l'appariement de données statistiques (RS *431.012.13*)[109]. Ces **appariements** peuvent uniquement être réalisés **à des fins ne se rapportant pas à des personnes** telles que la recherche, la planification et la statistique. Un appariement de ces données à des fins administratives est par conséquent exclu (voir le secret statistique inscrit à l'art. 14, al. 1, LSF).

Comme, en vertu de l'art. 14*a*, al. 1, LSF, seul l'OFS a une base légale pour effectuer de tels appariements de données, mais que d'autres services peuvent aussi avoir un intérêt certain pour de tels appariements, le Conseil fédéral a prévu, dans l'ordonnance sur les relevés statistiques, que l'OFS peut effectuer des appariements sur mandat de tiers dans le cadre d'un contrat de protection des données (art. 13*k* de l'ordonnance sur les relevés statistiques). Les appariements sur mandat de tiers sont soumis aux mêmes exigences en matière de protection des données que les appariements internes à l'OFS (pour accomplir ses missions statistiques). Ils ne peuvent donc être effectués que s'ils sont appropriés et nécessaires pour les travaux de recherche. Seules les variables et les caractéristiques qui sont réellement nécessaires et appropriées peuvent être appariées sur la base du principe de proportionnalité.

Les données appariées par l'OFS ne peuvent pas être transmises durablement en vertu de l'art. 13*l* en relation avec l'art. 9 de l'ordonnance sur les relevés statistiques et doivent être restituées à l'OFS

[106] Le numéro AVS (admin.ch)
[107] En décembre 2020, le Conseil national et le Conseil des États ont approuvé une modification de la loi, en vertu de laquelle les autorités fédérales, cantonales et communales peuvent utiliser en général systématiquement le numéro AVS dans le cadre de leurs tâches légales (cf. art. 153*b* ss LAVS ; entré en vigueur au 1er janvier 2022)
[108] https://www.admin.ch/opc/fr/classified-compilation/19930224/index.html
[109] https://www.admin.ch/opc/fr/classified-compilation/20122208/index.html

ou détruites une fois les travaux terminés. Comme les données médicales comptent parmi les données sensibles, les données appariées doivent être effacées dans tous les cas une fois les travaux statistiques d’exploitation terminés (art. 14*a*, al. 1, LSF et art. 13*m* de l’ordonnance sur les relevés statistiques).

En complément, quelques points clés sur la **situation juridique dans l’UE**, qui ne diffère pas fondamentalement de la situation en Suisse.

Sous le règlement général de l’UE sur la protection des données (RGPD), en vigueur depuis mai 2018, les données concernant la santé (art. 4, ch. 15, RGPD) jouissent d’une protection élevée (art. 9 RGPD). En vertu de l’art. 5, par. 1, let. b, RGPD, un *traitement ultérieur à des fins de recherche scientifique* est compatible avec les finalités initiales des données et donc *autorisé si* « des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée » sont prévues (art. 9, par. 2, let. j, et 89 RGPD). En raison de diverses clauses d’ouverture, la concrétisation de ces mesures spécifiques reste de la compétence des États membres. Les réglementations concernant la protection des secrets professionnels ne sont pas affectées (art. 9, par. 3, RGPD). Il faut donc s’attendre à des réglementations différentes à l’échelle nationale au sein de l’UE. Dans les pays scandinaves, par exemple, l’appariement de plusieurs jeux de données par le biais d’un « numéro personnel » est permis, alors que ce n’est pas encore le cas dans d’autres États, si tant est qu’il existe un numéro personnel.

Dans l’UE, l’utilisation de données se rapportant à des personnes pour la recherche est en général possible si elle est autorisée par le consentement des personnes concernées ou, le cas échéant, par une base légale. Seule une anonymisation préalable et sûre des données donne pour l’instant la possibilité de les utiliser à des fins de recherche sans condition et hors du cadre applicable en matière de protection des données.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Annexe 5 Formes du consentement en Suisse : état actuel et besoins de développement

La vue d'ensemble ci-dessous résume les aspects principaux des formes de consentement applicables aujourd'hui en Suisse et concerne l'utilisation primaire et secondaire des données.

<table>
<tr><th></th><th>Consentement spécifique</th><th>Consentement général (CG)</th></tr>
<tr><td>Base légale*</td><td colspan="2">Loi relative à la recherche sur l'être humain (LRH), chapitre 2, section 2 : Information et consentement.
Art. 16 Consentement éclairé
Art. 17 Consentement à la réutilisation de matériel ou de données
LRH, chapitre 4 : Réutilisation de matériel biologique et de données personnelles liées à la santé.
Art. 32 Réutilisation de matériel biologique et de données génétiques (al. 1 et 2)
Art. 33 Réutilisation de données personnelles non génétiques liées à la santé (al. 1)
Ordonnance relative à la recherche sur l'être humain (ORH), chapitre 3 : Réutilisation pour la recherche avec du matériel biologique et des données personnelles liées à la santé.</td></tr>
<tr><td></td><td>ORH, art. 28 Consentement éclairé pour la réutilisation de matériel biologique et de données personnelles génétiques pour un projet de recherche sous une forme non codée</td><td>ORH, art. 29 Consentement éclairé pour la réutilisation de matériel biologique et de données personnelles génétiques à des fins de recherche sous une forme codée (al. 1)
ORH, art. 31 […] données personnelles non génétiques liées à la santé à des fins de recherche sous une forme non codée
ORH, art. 32 […] données personnelles non génétiques liées à la santé à des fins de recherche sous une forme codée</td></tr>
<tr><td>Mise en œuvre formelle dans la pratique</td><td colspan="2">Pour les modèles, voir swissethics | Feuilles d'information / de consentement</td></tr>
<tr><td></td><td>Consentement spécifique</td><td>Consentement général (CG)</td></tr>
<tr><td>Contexte d'application Consentement</td><td>Consentement à la participation aux :
Projets de recherche sur des personnes[1] au sens de la LRH/ORH, chapitre 2
Essais cliniques[2] au sens de la LRH OClin</td><td>Consentement à la réutilisation de données et d'échantillons (génétiques et non génétiques) liés à la santé à des fins de recherche, qui sont collectés/prélevés dans le domaine stationnaire ou ambulatoire.
Application principalement à l'hôpital (clinique), d'après les recommandations de Médecine Universitaire Suisse, 2019[3]</td></tr>
<tr><td>Personnes concernées</td><td>Personnes physiques :
Individus malades et en bonne santé selon des critères d'inclusion et d'exclusion</td><td>Personnes physiques :
Patients
Participants</td></tr>
<tr><td>Durée de validité du consentement</td><td>Durée du projet de recherche, y compris publication</td><td>Illimitée, au-delà du décès</td></tr>
<tr><td>Retrait/révocation</td><td>Possible</td><td>Possible</td></tr>
<tr><td>Type de données (et matériel comme source de données)</td><td>Données personnelles liées à la santé, génétiques et non génétiques
Matériel biologique (échantillons)
(données personnelles et médicales)</td><td>Données personnelles liées à la santé, génétiques et non génétiques
Matériel biologique (échantillons restants)</td></tr>
</table>

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

	Consentement spécifique	Consentement général (CG)
		Données personnelles
Utilisation des données par les producteurs/responsables eux-mêmes		
Affectation Échange de données groupe de recherche	À des fins de recherche pour ce projet, transmission de données possible (aussi à l'étranger) sous une forme codée	Pour les projets multicentriques en collaboration avec des partenaires de recherche (tiers) en Suisse et à l'étranger
Transmission de données (et de matériel biologique) **à des tiers à des fins de réutilisation**		
Faisabilité	Seulement autorisée avec un *consentement séparé*[4] ; optionnel pour répondre ultérieurement à d'autres problématiques. ➲ **correspond à un CG**	Autorisée pour des projets en Suisse et à l'étranger selon les dispositions de la LPD et de la LRH
- Affectation	Recherche médicale (données et échantillons) Projet(s) non précisé(s)	Recherche : à l'hôpital et en collaboration avec des institutions publiques et privées
- Limitation dans le temps de l'utilisation des données/échantillons	Pour une durée indéterminée	Illimité jusqu'à une révocation
- Format des données		
▪ non codé	Illicite	Illicite
▪ codé	Admis	Admis
▪ anonymisé	Admis (aucune révocation possible)	Admis
▪ agrégé	Admis	Non prévu
Appariement	**Abordé dans le formulaire pour la protection des données en cas d'analyses génétiques**	**Pas abordé dans le formulaire**

* RS *810.30* - Loi fédérale du 30 septembre 2011 relative à la recherche sur l'être humain (Loi relative à la recherche sur l'être humain, LRH) (admin.ch)

RS *810.301* - Ordonnance du 20 septembre 2013 relative à la recherche sur l'être humain à l'exception des essais cliniques (Ordonnance relative à la recherche sur l'être humain, ORH) (admin.ch)

RS *810.305* - Ordonnance du 20 septembre 2013 sur les essais cliniques hors essais cliniques de dispositifs médicaux (Ordonnance sur les essais cliniques, OClin) (admin.ch)

[1] Projets de recherche sur l'être humain qui ne sont pas qualifiés d'essais cliniques, par exemple la collecte de données personnelles liées à la santé ou le prélèvement de matériel biologique et leur réutilisation à des fins de recherche.

[2] Un projet de recherche sur des personnes qui sont affectées à une certaine intervention liée à la santé (p. ex. application d'un médicament, d'une mesure de promotion de la santé, etc.) pour étudier l'effet sur la santé ou la structure et le fonctionnement du corps humain.

[3] Deuxième projet de CG harmonisé, élaboré en 2018 par les hôpitaux universitaires/unimedsuisse, approuvé et adopté par swissethics. Cf. https://www.unimedsuisse.ch/de/projekte/generalkonsent et https://www.unimedsuisse.ch/application/files/3215/8385/1516/Recommendations__GC_2019_E_def.pdf. L'ORH permettrait une application plus large, par exemple aussi dans les soins ambulatoires.

[4] Déclaration de consentement à la réutilisation de données (génétiques) et de matériel biologique sous une forme codée (template_studieninformation_f.pdf (swissethics.ch))

Explications complémentaires

D'un point de vue formel, il faut distinguer deux approches (modèles) pour le consentement. Dans le modèle « *opt in* », les personnes volontaires concernées donnent activement leur consentement écrit à l'utilisation ciblée des données après avoir été informées et après un délai de réflexion. Dans le modèle « *opt out* », on part du principe que les personnes concernées consentent en principe après avoir été informées, sauf si elles décident de s'y opposer.

Pour les projets de recherche, le modèle « *opt in* » est d'usage. D'après l'enquête réalisée par l'OFSP [34], le modèle « *opt in* » est appliqué dans la plupart des études de cohorte et des registres, même si

certains registres procèdent d'après le modèle « *opt out* »[110]. Il y a en outre des dispositions légales qui rendent un consentement inutile pour certains registres/services de déclaration (obligation)[111]. Il en va de même pour les registres qui ne collectent pas de données se rapportant à des personnes ou uniquement des données anonymisées/agrégées.

L'évaluation de la LRH a identifié la nécessité de prendre des mesures pour les réglementations concernant la réutilisation de données médicales à des fins de recherche [45]. Cela concerne des problèmes dans l'interprétation de termes, mais aussi un besoin d'harmonisation et d'adaptation pour le consentement général ainsi que pour la mise en œuvre des dispositions différenciées et des exigences formelles de la LRH et des dispositions d'exécution concernant l'information et le consentement.

Les **déclarations de consentement spécifiques aux projets/études** dans le cadre de la mise en œuvre de la LRH et des dispositions d'exécution permettent aux chercheurs de demander en option, à la personne concernée et informée, un consentement à la réutilisation des données ou du matériel biologique en dehors du projet de recherche spécifique ou de l'étude clinique, limité à une utilisation aux fins de la recherche médicale.

Le **consentement général (CG)** est une base importante pour permettre l'utilisation multiple de données médicales et de matériel biologique dans la recherche médicale. Par CG au sens strict, on n'entend ci-dessus que le consentement désigné comme « *opt in* » (art. 32, al. 2, et 33, al. 1, LRH). La « non-opposition » une fois l'information donnée (« *opt out* », art. 33, al. 2, LRH) n'est pas comprise dans cette définition stricte. La mise en œuvre actuelle du modèle de CG (voir la vue d'ensemble ci-dessus, note 3) comprend toutefois aussi la non-opposition. Comme l'utilisation du terme de CG s'est imposée dans pratique aussi bien pour le modèle *opt in* que pour le modèle *opt out*, le terme est utilisé ci-dessous au sens large.

Le **modèle de CG est considéré comme une recommandation et n'est pas appliqué de manière universelle** ; il est surtout utilisé dans les hôpitaux/cliniques qui réutilisent des données relatives aux traitements et le cas échéant aux projets à des fins de recherche[112]. La transmission de données et d'échantillons codés et anonymisés, obtenus avec le CG, est également possible – en cas d'information en ce sens – à l'étranger ; ce faisant, l'art. 42 LRH et les dispositions relatives à la transmission de données à l'étranger doivent être respectés. Il ressort du CG que la transmission de matériel biologique et de données médicales à des fins de recherche à l'étranger se fait exclusivement dans des pays qui disposent d'une réglementation au moins comparable en matière de protection des données et d'un contrôle des projets de recherche par une commission d'éthique.

L'Académie suisse des sciences médicales (ASSM) et swissethics ont déjà relevé, dans le premier modèle de CG (2017), que les droits et les devoirs du destinataire doivent être réglés contractuellement et qu'aucun gain ne doit être visé avec la transmission des données et des échantillons[113] [54]. Un **élément controversé du CG** est son application générale, c'est-à-dire la référence aux données et des échantillons qui sont collectés/prélevés au moment du consentement, mais aussi à un moment ultérieur (p. ex. après cinq ans ou plus en cas de nouveau traitement à l'hôpital). De nouvelles formes flexibles/dynamiques de la pratique du consentement pourraient offrir une solution qui permettrait la concomitance de la production des données et du consentement à leur utilisation.

Les explications sur les utilisateurs de données potentiels se limitent exclusivement à l'utilisation des données dans les projets de recherche médicaux. Cela ne correspond pas à la diversité des questions dans le domaine des soins de santé auxquelles il faudrait répondre par l'utilisation commune d'un large spectre de données personnelles liées à la santé. En outre, une transmission directe et non codée des données se rapportant à des personnes à des sources de données secondaires, notamment les registres, n'est pas prévue dans le cadre du CG. Les registres, qui servent à garantir la qualité

[110] Les registres des tumeurs en Suisse ; SwissDisc Registry – registre suisse pour l'enregistrement des patients opérés d'une hernie discale ; Acute myocardial infarction in Switzerland (AMIS Plus) ; Swiss Registry of Cardiac Arrest (SwissRECA) ; Groupe de travail pour l'assurance qualité dans les disciplines chirurgicales (AQC) ; Juvenile Inflammatory Rheumatism Cohorte (JIR-cohorte).
[111] Mandat de prestations du 31 janvier 2019 dans le domaine MHS sur la participation aux registres concernant la médecine hautement spécialisée ; loi sur les épidémies (LEp ; RS *818.101*) ; loi sur la prévoyance professionnelle vieillesse, survivants et invalidité (LPP ; RS *831.40*) ; loi sur la procréation médicalement assistée (LPMA ; RS *810.11*) ; ordonnance sur la radioprotection (ORaP ; RS *814.501*).
[112] https://swissethics.ch/fr/news/2017/07/03/generalkonsent-eine-wichtige-basis-fuer-die-medizinische-forschung et
[113] https://swissethics.ch/doc/swissethics/general_consent/Factsheet_GK_f.pdf

dans le domaine de la santé, ont besoin – dans la mesure où aucune réglementation spéciale ne s’applique – d’un consentement spécial des personnes concernées pour l’enregistrement des données médicales. Cela génère du travail supplémentaire et implique une charge supplémentaire pour les personnes concernées.

Les informations et déclarations de consentement utilisées à l’heure actuelle n’abordent pas du tout (CG), ou uniquement pour les données génétiques, le risque de détermination de profils (cf. chap. 3.2.4) lié à la réutilisation des données à la suite d’un appariement.

Le rapport d’évaluation de la LRH [44] traite de la **gestion problématique des « anciennes » données**. Dans ce cadre, il faudrait aussi, dans le but de trouver une solution, aborder la question d’un **don de données après la mort** sur une base volontaire (don de données posthume), par analogie à un don d’organes malgré des différences conceptuelles (matérialité, concomitance). Dans le CG, cette question est fixée par le biais de sa durée de validité, mais elle mérite de plus amples clarifications. En l’absence d’un consentement à une réutilisation, les données d’une personne sont bloquées après son décès. Il serait cohérent d’encourager non seulement la réutilisation des données médicales à des fins de recherche du vivant des personnes concernées, mais aussi le don de données posthume, car les motivations sont comparables. Krutzinna et ses collègues [55, 56] fournissent des bases fondées en l’espèce. Une étude suisse [39] a montré que les citoyens accordent une grande importance à un consentement informé et basé sur la confiance et adoptent une attitude différenciée sur la question du don de données (*data donation*), en fonction des destinataires/utilisateurs de données publics ou privés ainsi que de la possibilité de pouvoir revenir sur leur décision.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

Annexe 6 Exemples de transmission de données avec un changement de finalité et la nécessité d'un consentement

Exemple A4-1

Une entreprise de recherche privée s'intéresse à l'effet à long terme d'un « médicament pour le cœur » utilisé dans les soins aux patients, mais veut éviter un essai clinique onéreux. C'est pourquoi les chercheurs aimeraient réaliser une évaluation sur la base des données thérapeutiques existantes dans le cadre des soins ambulatoires et stationnaires en Suisse.

L'accès à des données qui, initialement, ont été uniquement saisies pour le traitement des personnes concernées (changement de finalité) doit être accordé à un acteur privé. Pour pouvoir bien évaluer l'effet du médicament dans le temps, il est nécessaire de suivre l'itinéraire des patients. Des caractéristiques personnelles (appariement de données ; identification personnelle) sont nécessaires en l'espèce. Un consentement des personnes concernées doit être obtenu pour la transmission des données thérapeutiques à l'entreprise de recherche.

Exemple A4-2

Un groupe de recherche veut utiliser le jeu de données de l'enquête nationale sur l'alimentation, menuCH, réalisée sur mandat de l'Office fédéral de la sécurité alimentaire et des affaires vétérinaires (OSAV), pour étudier si les personnes qui consomment beaucoup de viande développent plus souvent un cancer du côlon. Pour ce faire, il faudrait apparier les données des enquêtes personnelles sur l'alimentation avec les données de l'Organe national d'enregistrement du cancer. Ce projet est soumis aux dispositions de la LRH.

Le consentement des participants à l'enquête sur l'alimentation est spécifique au projet (but premier : monitorage de la consommation alimentaire). Un consentement séparé est nécessaire pour une réutilisation des données dans une autre étude qui présuppose une utilisation de caractéristiques personnelles. Les personnes inscrites dans le registre des tumeurs pour un cancer du côlon devraient aussi consentir à une utilisation de leurs données dans cette étude. À l'heure actuelle, cela ne serait réalisable qu'à certaines conditions et au prix d'une charge importante. De plus, la question de l'efficacité de l'appariement se poserait également en l'absence d'un identificateur univoque.

Annexe 7 Esquisse de processus : utilisation secondaire des données dans le système de réutilisation et d'appariement des données médicales à des fins de recherche

Le processus de traitement secondaire des données comprend les étapes partielles ou les aspects suivants : l'acquisition des données, la fourniture des données sans/avec appariement, l'analyse des données fournies et la publication des résultats, ainsi que l'enregistrement et l'archivage des jeux de données appariés. La description suivante donne une idée du déroulement possible du processus représenté schématiquement dans le graphique du système (figure 4, p. 35). Les structures ou unités fonctionnelles et les exigences complexes posées à leur exploitation seront pleinement concrétisées par des spécialistes en ce qui concerne la mise en œuvre du système.

Traitement des demandes concernant l'utilisation secondaire des données (acquisition des données)

Les demandes relatives à l'utilisation de collections de données existantes enregistrées comme étant accessibles dans les catalogues de métadonnées seraient traitées et évaluées de manière centralisée par l'Organe national de coordination des données (ONCD). Sur demande, l'ONCD peut soutenir les chercheurs (les utilisateurs de données) dans la recherche de collections de données appropriées. Il veille à un traitement cohérent et rapide des demandes.

Mieux utiliser les données médicales pour assurer l'efficience et la qualité des soins

L'ONCD aurait pour tâche d'examiner la faisabilité des demandes ou des projets entrants dans une perspective juridique, éthique, scientifique et technique. Il s'agit de vérifier le but de l'utilisation secondaire des données (cf. chap. 1.2.3, tableau 1), les données sollicitées pour répondre à la problématique versus les données absolument nécessaires en l'espèce, ainsi que la faisabilité ou la complexité de l'utilisation et de l'appariement des données (p. ex. le nombre de collections de données demandées et, le cas échéant, de données inconnues fournies par les requérants).

Préparation et mise en œuvre de la transmission des données (fourniture des données)

Après avoir examiné la question de savoir si un consentement des personnes concernées est nécessaire pour la réutilisation prévue des données, l'ONCD lancerait la demande virtuelle de consentement des personnes concernées pour la nouvelle utilisation des données (figure 4, points ④, ① et ②). Dans le cadre du consentement, la génération d'un « jeton de sécurité » serait envisageable, qui servirait de signal pour l'utilisation autorisée de l'identificateur univoque pour une tâche spécifique lors de la réutilisation des données, comme celle de l'appariement de deux jeux de données ou plus.

La clarification et l'exécution de la transmission des données et de l'utilisation secondaire des données par l'ONCD rendent caduque la nécessité d'un accord individuel pour le transfert et l'utilisation des données (*DTUA*) entre le requérant (utilisateur des données, tiers) et le producteur des données/responsable. La transmission des données par le responsable à l'infrastructure de l'ONCD se fait sous une forme codée (cf. figure 4, point ③).

L'ONCD préparerait ensuite les jeux de données transmis par les producteurs de données/responsables pour l'utilisateur secondaire des données. À l'occasion de l'évaluation de la demande d'utilisation secondaire des données, il aurait déjà défini quelles informations personnelles identifiables le ou les jeux de données requis peuvent tout au plus contenir (approche minimale). Les requérants ne doivent en principe recevoir que des jeux de données désidentifiés à traiter afin d'empêcher l'identification des personnes concernées.

Appariement des données pour les utilisateurs secondaires (fourniture des données)

La préparation des collections de données et leur appariement ne doivent se faire que dans un environnement informatique restreint, autonome et protégé qui est rattaché à l'ONCD. Seuls les experts en appariement pourraient utiliser l'ID d'utilisation des données univoque (cf. figure 4, points ② et ③) pour apparier les collections de données originales. Avant la validation de la nouvelle collection de données appariée, l'ID d'utilisation des données et d'autres caractéristiques d'identification non pertinentes pour les utilisateurs des données seraient en outre effacées et remplacées par un numéro d'appariement non signifiant spécifique au projet.

L'enregistrement des jeux de données appariés et désidentifiés doit permettre leur réutilisation en cas de demandes comparables. Un enregistrement des scripts serait autorisé pour permettre aux experts une reconstruction efficace ou une adaptation du processus d'appariement. En revanche, les jeux de données originaux fournis à l'ONCD seraient effacés une fois le projet terminé.

Traitement des données (analyses) et publication par les utilisateurs secondaires

Le traitement de la nouvelle collection de données par l'utilisateur secondaire des données se ferait toujours dans une infrastructure de données sûre et rattachée à l'ONCD. Après un contrôle de sécurité, les utilisateurs des données obtiendraient un droit d'accès personnel et limité dans le temps au domaine d'infrastructure spécifique. Si, lors de la fourniture des données, l'ONCD fixe des directives pour la publication des résultats (p. ex. sur l'agrégation des données), l'utilisateur secondaire des données doit en tenir compte. De plus, la rétroaction des informations techniques concernant les projets aux catalogues de métadonnées se ferait par l'intermédiaire de l'ONCD, et les informations de fond ou les conclusions seraient transmises à la population, et donc aux personnes concernées.