NORME INTERNATIONALE

ISO 25237

Première édition
2017-01

Informatique de santé — Pseudonymisation

Health informatics — Pseudonymization

Numéro de référence
ISO 25237:2017(F)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Sommaire

Page

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (IEC) en ce qui concerne la normalisation électrotechnique.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives).

L'attention est appelée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'OMC concernant les obstacles techniques au commerce (OTC), voir le lien suivant: Avant-propos — Informations supplémentaires.

Le comité chargé de l'élaboration du présent document est l'ISO/TC 215, *Informatique de santé*.

Introduction

La pseudonymisation est reconnue comme une méthode importante de protection des informations de santé à caractère personnel. Les services associés peuvent être utilisés aussi bien au plan national que pour la communication transfrontière.

Les domaines d'application concernent, sans s'y limiter, les secteurs suivants:

— l'utilisation indirecte des données cliniques (par exemple recherche);

— les essais cliniques et la surveillance post-marketing;

— les soins pseudonymes;

— les systèmes d'identification des patients;

— la surveillance et l'évaluation de la santé publique;

— les dossiers confidentiels sur la sécurité des patients (par exemple effets indésirables d'un médicament);

— les rapports comparatifs fondés sur des indicateurs de qualité;

— le contrôle par les pairs;

— les groupes de consommateurs;

— l'assistance technique.

Le présent document fournit un modèle conceptuel des aspects en jeu, des exigences en matière de pratiques fiables, ainsi que des spécifications pour la planification et la mise en œuvre des services de pseudonymisation.

La spécification d'un workflow général, associé à une politique de fiabilisation des opérations, servira de guide général pour la mise en œuvre ainsi que pour l'assurance qualité, et aidera l'utilisateur des services de pseudonymisation à déterminer la confiance qu'il peut accorder aux prestations assurées. Ce guide permettra de former les organismes afin qu'ils puissent assurer eux-mêmes les services de pseudonymisation, avec des compétences suffisantes pour atteindre le degré voulu de qualité et de réduction des risques.

Informatique de santé — Pseudonymisation

1 Domaine d'application

Le présent document établit un certain nombre de principes et d'exigences visant à garantir la protection de la vie privée, grâce à des services de pseudonymisation ayant pour objet de protéger les informations de santé à caractère personnel. Le présent document est applicable aux organismes qui souhaitent s'engager dans des processus de pseudonymisation pour eux-mêmes et aux organismes qui se déclarent dignes de confiance pour engager des opérations dans des services de pseudonymisation.

Le présent document:

— définit un concept de base pour la pseudonymisation (voir Article 5);

— définit une méthodologie de base pour les services de pseudonymisation, y compris au niveau des aspects organisationnels et techniques (voir Article 6);

— spécifie un cadre politique et des exigences minimales pour la ré-identification contrôlée (voir Article 7);

— donne une vue d'ensemble des différents cas d'utilisation où l'opération de pseudonymisation peut être réversible ou irréversible (voir Annexe A);

— fournit un guide pour l'évaluation des risques en cas de ré-identification (voir Annexe B);

— donne un exemple de système qui utilise la désidentification (voir Annexe C);

— fournit des exigences informatives pour l'interopérabilité des services de pseudonymisation (voir Annexe D); et

— spécifie un cadre politique et des exigences minimales favorisant des pratiques fiables pour un service de pseudonymisation (voir Annexe E).

2 Références normatives

Les documents suivants cités dans le texte constituent, pour tout ou partie de leur contenu, des exigences du présent document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document de référence s'applique (y compris les éventuels amendements).

ISO 27799, *Informatique de santé — Management de la sécurité de l'information relative à la santé en utilisant l'ISO/IEC 27002*

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions suivants s'appliquent.

L'ISO et l'IEC tiennent à jour des bases de données terminologiques destinées à être utilisées en normalisation, consultables aux adresses suivantes:

— IEC Electropedia: disponible à l'adresse http://www.electropedia.org/.

— ISO Online browsing platform: disponible à l'adresse http://www.iso.org/obp.

3.1
contrôle d'accès
ensemble des moyens garantissant que seules les entités autorisées peuvent accéder aux ressources d'un système informatique, et seulement d'une manière autorisée

[SOURCE: ISO/IEC 2382:2015, 2126294]

3.2
anonymisation
processus par lequel des *données à caractère personnel* (3.37) sont altérées irréversiblement, de telle façon que la personne concernée ne puisse plus être identifiée, directement ou indirectement, par le responsable du traitement des données, seul ou en collaboration avec une autre partie

Note 1 à l'article: Ce concept est absolu et peut, dans la pratique, être difficile à mettre en œuvre.

[SOURCE: ISO/IEC 29100:2011, 2.2, modifiée]

3.3
données anonymisées
données (3.14) de sortie produites par un processus d'*anonymisation* (3.2)

[SOURCE: ISO/IEC 29100:2011, 2.3, modifiée]

3.4
identifiant anonyme
identifiant (3.27) d'une personne ne permettant pas l'*identification* (3.26) de la *personne physique* (3.34)

3.5
authentification
établissement de la validité de l'identité déclarée

3.6
attaquant
personne exploitant délibérément les vulnérabilités des contrôles de sécurité techniques et non techniques, afin de piller ou de compromettre les réseaux et les systèmes d'information ou d'empêcher les utilisateurs légitimes d'accéder aux ressources de ces réseaux et systèmes

[SOURCE: ISO/IEC 27033-1:2015, 3.3]

3.7
cryptogramme, texte chiffré
données (3.14) résultant d'un chiffrement et dont le contenu sémantique n'est pas disponible sans recours à des techniques cryptographiques

[SOURCE: ISO/IEC 2382:2015, 2126285]

3.8
confidentialité
propriété d'une *information* (3.29) qui n'est ni disponible, ni divulguée aux personnes, entités ou processus non autorisés

[SOURCE: ISO 7498-2:1989, 3.3.16]

3.9
clé de chiffrement de contenu
clé cryptographique utilisée pour chiffrer le contenu d'une communication

3.10
responsable du traitement
personne physique ou morale, autorité publique, service ou tout autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du *traitement de données à caractère personnel* (3.40)

3.11
cryptographie
discipline incluant les principes, moyens et méthodes de transformation des *données* (3.14), dans le but de cacher leur contenu, d'empêcher que leur modification passe inaperçue et/ou d'empêcher leur utilisation non autorisée

[SOURCE: ISO 7498-2:1989, 3.3.20]

3.12
algorithme cryptographique
<chiffre> méthode permettant de transformer des *données* (3.14), dans le but de cacher leur contenu, d'empêcher que leur modification passe inaperçue et/ou d'empêcher leur utilisation non autorisée

3.13
gestion de clés cryptographiques
gestion de clés
production, stockage, distribution, suppression, archivage et application de *clés* (3.31) conformément à la *politique de sécurité* (3.46)

[SOURCE: ISO 7498-2:1989, 3.3.33]

3.14
donnée
représentation réinterprétable d'une *information* (3.29) sous une forme conventionnelle convenant à la communication, à l'interprétation ou au traitement

Note 1 à l'article: Les données peuvent être traitées par des moyens humains ou automatiques.

[SOURCE: ISO/IEC 2382:2015, 2121272]

3.15
intégrité des données
propriété assurant que des *données* (3.14) n'ont pas été modifiées ou détruites de façon non autorisée

[SOURCE: ISO 7498-2:1989, 3.3.21]

3.16
liage de données
appariement et combinaison de *données* (3.14) issues de plusieurs bases de données

3.17
protection des données
organisation technique et sociale permettant de négocier, gérer et garantir le caractère *privé* (3.39) et la sécurité des informations

3.18
personne concernée
personne à laquelle se rapportent les *données* (3.14)

3.19
déchiffrement
processus visant à convertir des *données* (3.14) chiffrées pour les ramener à leur forme d'origine, de manière à pouvoir les comprendre

3.20
désidentification
terme général qui désigne tout processus réduisant l'association entre un ensemble de *données* (3.14) d'identification et la *personne concernée* (3.18)

3.21
données d'identification directe
données (3.14) qui identifient directement un individu

Note 1 à l'article: Les identifiants directs sont les données qui peuvent être utilisées pour identifier une personne sans informations supplémentaires ou par recoupement avec d'autres informations du domaine public.

3.22
communication
fait de révéler des *données* (3.14) ou d'y donner accès

Note 1 à l'article: Le fait que le destinataire regarde réellement les données, les transforme en connaissances ou les conserve, est sans importance vis-à-vis de la réalisation de la communication.

3.23
chiffrement
processus de conversion d'*informations* (3.29) ou de *données* (3.14) en un bloc chiffré ou un code

3.24
identifiant de soins de santé
identifiant du sujet des soins
identifiant (3.27) d'une personne, destiné à être principalement utilisé par un système de soins de santé

3.25
personne identifiable
personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale

[SOURCE: Directive 95/46/EC]

3.26
identification
processus consistant à utiliser des attributs déclarés ou observés d'une entité pour distinguer cette entité parmi d'autres dans un ensemble d'identités

Note 1 à l'article: L'identification d'une entité dans un contexte déterminé permet à une autre entité de différencier les entités avec lesquelles elle interagit.

3.27
identifiant
information (3.29) utilisée pour déclarer une identité, avant corroboration potentielle par un authentifiant correspondant

[SOURCE: ENV 13608-1:2000, 3.44]

3.28
données d'identification indirecte
données (3.14) qui ne peuvent identifier une personne que lorsqu'elles sont utilisées conjointement avec d'autres données d'identification indirecte

Note 1 à l'article: Les identifiants indirects peuvent réduire à un individu la population à laquelle la personne appartient, s'ils sont utilisés en combinaison.

EXEMPLE Code postal, sexe, âge, date de naissance.

3.29
information
connaissance concernant un objet qui, dans un contexte déterminé, a une signification particulière

[SOURCE: ISO/IEC 2382:2015, 2121271, modifiée]

3.30
irréversibilité
pour toute transformation d'identifiable en pseudonyme, situation dans laquelle il est informatiquement irréalisable de remonter à l'*identifiant* (3.27) d'origine en partant du *pseudonyme* (3.43)

3.31
clé
série de symboles commandant les opérations de *chiffrement* (3.23) et de *déchiffrement* (3.19)

[SOURCE: ISO 7498-2:1989, 3.3.32]

3.32
liage d'objets d'information
processus permettant d'établir une association logique entre différents objets d'information

3.33
dossier personnel de santé longitudinal ou à vie
dossier coordonné permanent regroupant des informations importantes, classées par ordre chronologique

Note 1 à l'article: Ce dossier peut inclure toutes les données historiques collectées ou être extrait sous forme de synthèse définie par l'utilisateur, regroupant des faits et des événements démographiques, génétiques, cliniques et environnementaux marquants, conservés dans un système automatisé.

[SOURCE: ISO/TR 21089:2004, 3.61, modifiée]

3.34
personne physique
être humain par opposition à personne morale, laquelle peut être un organisme privé ou public

3.35
identification de personne
processus établissant une association entre un objet d'information et une personne physique

3.36
identifiant personnel
information permettant d'identifier une même et unique personne dans un contexte donné

3.37
données à caractère personnel
information concernant une *personne physique* (3.34) identifiée ou identifiable («personne concernée»)

[SOURCE: Directive 95/46/EC]

3.38
principale utilisation des données à caractère personnel
utilisations et *communications* (3.22) prévues pour les *données* (3.14) collectées

3.39
respect de la vie privée
garantie de l'absence d'intrusion dans la vie privée ou les affaires d'un individu dans la mesure où cette intrusion résulte de la collecte et de l'utilisation illégales et non fondées de *données* (3.14) relatives à cet individu

[SOURCE: ISO/IEC 2382:2015, 2126263]

3.40
traitement de données à caractère personnel
opération ou ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des *données à caractère personnel* (3.37), telles que la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, ainsi que le verrouillage, l'effacement ou la destruction

[SOURCE: Directive 95/46/EC]

3.41
sous-traitement
personne physique ou morale, autorité publique, service ou tout autre organisme qui traite des *données à caractère personnel* (3.37) pour le compte du *responsable du traitement* (3.10)

Note 1 à l'article: Voir la Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

3.42
pseudonymisation
type particulier de *désidentification* (3.20) qui, à la fois, supprime la corrélation avec la personne concernée et ajoute une association entre un ensemble donné de caractéristiques ayant trait à la personne concernée et un ou plusieurs *pseudonymes* (3.43)

3.43
pseudonyme
identifiant personnel (3.36) différent de l'identifiant personnel normalement utilisé et employé avec des données pseudonymisées pour assurer la cohérence de l'ensemble de données, reliant ainsi toutes les informations relatives à une personne concernée, sans communiquer la véritable identité de la personne

Note 1 à l'article: Le pseudonyme peut être dérivé de l'identifiant personnel normalement utilisé, de manière réversible ou irréversible, ou n'avoir aucun rapport avec ce dernier.

Note 2 à l'article: Le terme de pseudonyme se limite généralement à désigner un identifiant qui ne permet pas la dérivation directe de l'identifiant personnel normal. De telles informations pseudonymes sont donc fonctionnellement anonymes. Une tierce partie de confiance peut être en mesure d'obtenir l'identifiant personnel normal à partir du pseudonyme.

3.44
destinataire
personne physique ou morale, autorité publique, service ou tout autre organisme qui reçoit communication de *données* (3.14)

3.45
utilisation secondaire des données à caractère personnel
utilisations et *communications* (3.22) qui divergent de l'usage initialement prévu pour les *données* (3.14) collectées

3.46
politique de sécurité
plan ou programme d'action adopté pour assurer la sécurité informatique

[SOURCE: ISO/IEC 2382:2015, 2126246]

3.47
tierce partie de confiance
autorité de sécurité, ou son mandataire, à qui d'autres entités accordent leur confiance pour des activités en rapport avec la sécurité

[SOURCE: ISO/IEC 18014-1:2008, 3.20]

4 Abréviations

DICOM Imagerie numérique et communication en médecine [Digital Imaging and Communication in Medicine]

HIPAA Loi américaine sur la transférabilité des régimes d'assurance-maladie et l'imputabilité des données sensibles [Health Insurance Portability and Accountability Act]

SIS Système d'information de santé

VIH Virus de l'immunodéficience humaine

IP Protocole Internet [Internet Protocol]

VoV Victime de violence

5 Exigences concernant la protection du caractère privé des identités dans le domaine de la santé

5.1 Objectifs de la protection de la vie privée

L'objectif de la protection de la vie privée, en tant qu'objectif de sécurité visant à garantir la confidentialité, est d'empêcher la communication non autorisée ou non souhaitée d'informations sur une personne, lesquelles peuvent en outre influer sur des facteurs de risque juridiques, organisationnels et financiers. La protection de la vie privée est un sous-domaine du droit générique au respect de la vie privée qui, par définition, englobe d'autres entités sensibles au respect du caractère privé des données, telles que les organismes. Étant donné que l'aspect «respect de la vie privée» est le mieux réglementé et le plus répandu, ce modèle conceptuel met l'accent sur le respect du caractère privé des données. Des solutions de protection conçues pour le respect de la vie privée peuvent également être transposées pour la protection des données sensibles d'autres entités. Cela peut être utile dans les pays où le caractère privé des données des entités ou des organismes est réglementé par la loi.

La protection des données à caractère personnel a deux objectifs: l'un qui consiste à protéger les données à caractère personnel interagissant avec des applications en ligne (par exemple, navigation Web) et l'autre qui consiste à protéger les données à caractère personnel collectées dans des bases de données. Le présent document se limitera à ce dernier objectif.

Les données peuvent être extraites de bases de données. L'objectif est de réduire le risque que l'identité des personnes concernées soit communiquée. Les chercheurs travaillent sur des «cas», antécédents longitudinaux de patients, collectés avec le temps et/ou provenant de différentes sources. Pour la compilation des différents éléments de données constitutifs des cas, il est, toutefois, nécessaire d'utiliser une technique qui permette d'agréger les données des personnes concernées, sans nuire à leur vie privée. Pour y parvenir, il est possible de recourir à la pseudonymisation des données.

La désidentification est utilisée pour réduire les risques liés au respect de la vie privée dans un large éventail de situations.

Une désidentification extrême est utilisée pour le matériel pédagogique qui sera largement rendu public, et dont il convient néanmoins qu'il transmette suffisamment de détails pour être utile à des fins d'éducation médicale. (Il existe un profil IHE d'assistance à l'automatisation pour la réalisation de ce type de désidentification. Une grande partie du processus est adaptée au patient et à l'objectif poursuivi en éducation médicale.)

Les organismes de santé publique utilisent des bases de données désidentifiées pour suivre et comprendre les maladies.

Les essais cliniques utilisent la désidentification à la fois pour protéger la vie privée et pour éviter les biais subconscients en supprimant d'autres informations comme, par exemple, l'administration au patient d'un placebo ou d'un médicament expérimental.

Une désidentification légère est utilisée dans de nombreuses études cliniques, où on ne révèle pas aux évaluateurs qui est le médecin traitant, l'hôpital, le patient, etc., à la fois pour réduire les risques liés au respect de la vie privée et pour éliminer les biais subconscients. Ce type de désidentification empêche uniquement la communication fortuite des données aux évaluateurs. Un effort délibéré peut facilement permettre de découvrir l'identité du patient, etc.

Lorsqu'il faut produire des statistiques ou des analyses sur la charge de travail dans les hôpitaux ou dans le cadre de traitements délivrés au titre de contrats avec des administrateurs ou des acheteurs de services de soins de santé, il est nécessaire de pouvoir séparer les différents patients, sans avoir besoin de savoir qui ils sont individuellement. Voici un exemple d'utilisation de la désidentification dans un cadre commercial.

Le processus de stratification des risques (de réhospitalisation, par exemple) peut être engagé à partir des dossiers de services de soins principaux et secondaires des patients. Les dossiers sont désidentifiés pour l'analyse, mais lorsque les patients sont signalés comme étant à haut risque, ces patients peuvent être réidentifiés par un médecin compétent pour permettre des actions de suivi. Pour plus de détails sur la pseudonymisation dans le domaine de la santé, voir l'Annexe A.

5.2 Généralités

La désidentification est le terme général qui désigne tout processus réduisant l'association entre un ensemble de données d'identification et la personne concernée, avec une ou plusieurs utilisations prévues pour l'ensemble de données obtenu. La pseudonymisation est subordonnée à la désidentification. Le pseudonyme est le moyen par lequel des données pseudonymisées sont liées à une même personne ou à des systèmes d'information sans que l'identité de ladite personne soit révélée. La désidentification peut limiter de façon inhérente l'utilité des données résultantes. La pseudonymisation peut être exécutée avec ou sans possibilité de réidentifier la personne concernée (pseudonymisation réversible ou irréversible). Il existe plusieurs scénarios de cas d'utilisation de la pseudonymisation dans le domaine de la santé, notamment dans le traitement électronique sans cesse croissant des données patient, couplé à des attentes en matière de protection de la vie privée toujours plus importantes du côté des patients. Plusieurs exemples sont fournis à l'Annexe A.

Il est important de noter que, tant qu'il y a des données pseudonymisées, il existe un risque de réidentification non autorisée. Ce n'est pas très différent du chiffrement, qui peut être cassé par force brute, mais l'objectif est de rendre l'opération tellement difficile que le coût en soit prohibitif. Du fait que la désidentification est un concept plus récent que le chiffrement, les risques associés ne sont pas aussi bien compris.

5.3 La désidentification en tant que processus de réduction des risques

5.3.1 Généralités

Il convient que le processus de désidentification tienne compte des contrôles de sécurité et de respect de la vie privée qui s'appliqueront à l'ensemble de données obtenu. Il est rare de réduire les risques à un point tel que l'ensemble de données n'ait pas besoin de contrôles de sécurité permanents.

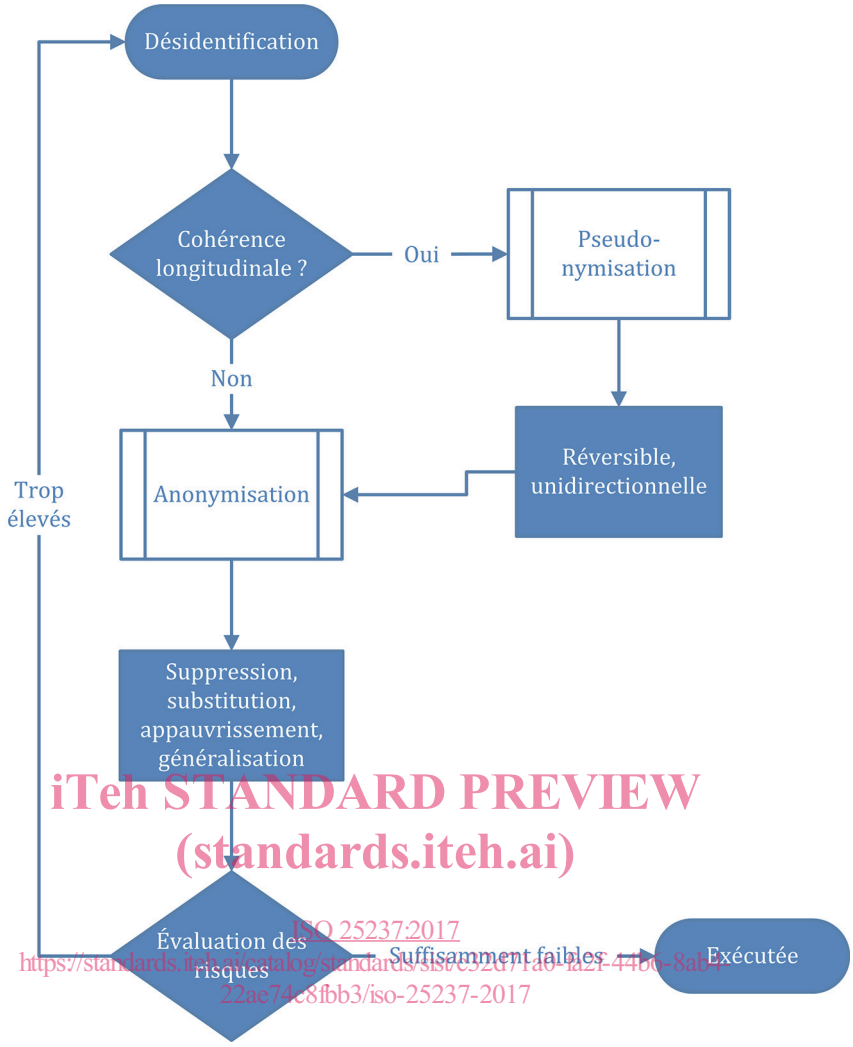

Figure 1 — Représentation graphique du processus de désidentification

La Figure 1 est un schéma informatif représentant le processus de désidentification. Elle révèle que le concept de niveau supérieur est la désidentification en tant que processus. Ce processus utilise des sous-processus: la pseudonymisation et/ou l'anonymisation. Ces sous-processus utilisent différents outils, qui sont spécifiques au type d'élément de données sur lequel ils agissent et à la méthode de réduction des risques utilisée.

L'état de départ correspond à celui où aucune donnée n'est autorisée à traverser le système. Il convient que chaque élément soit justifié par l'utilisation prévue pour l'ensemble de données obtenu. L'utilisation prévue pour l'ensemble de données a une grande incidence sur le processus de désidentification.

5.3.2 Pseudonymisation

La désidentification peut tirer avantage de la pseudonymisation lorsqu'une cohérence longitudinale est nécessaire. Il peut s'agir de conserver un ensemble de dossiers qu'il convient d'associer les uns aux autres et qui, sans cette cohérence longitudinale, pourraient être séparés. Il est ainsi possible de regrouper tous les dossiers d'un patient sous un même pseudonyme. Cela peut également permettre de garantir que, chaque fois que des données sont extraites en un ensemble désidentifié, les nouvelles entrées sont aussi associées au même pseudonyme. En pseudonymisation, l'algorithme utilisé peut être intentionnellement réversible ou intentionnellement non réversible. Un schéma réversible peut être une table de correspondance secrète qui, sous réserve d'autorisation, peut être utilisée pour découvrir l'identité d'origine. Dans un schéma non réversible, une table temporaire peut être utilisée durant le processus, mais elle est détruite dès que le processus est terminé.