Union internationale des télécommunications

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

X.1148

(09/2020)

SÉRIE X: RÉSEAUX DE DONNÉES, COMMUNICATION ENTRE SYSTÈMES OUVERTS ET SÉCURITÉ

Applications et services sécurisés (1) – Sécurité de la toile

Cadre du processus de désidentification pour les fournisseurs de services de télécommunication

Recommandation UIT-T X.1148

Union internationale des télécommunications

RECOMMANDATIONS UIT-T DE LA SÉRIE X

RÉSEAUX DE DONNÉES, COMMUNICATION ENTRE SYSTÈMES OUVERTS ET SÉCURITÉ

RÉSEAUX PUBLICS DE DONNÉES	X.1–X.199
INTERCONNEXION DES SYSTÈMES OUVERTS	X.200–X.299
INTERFONCTIONNEMENT DES RÉSEAUX	X.300–X.399
SYSTÈMES DE MESSAGERIE	X.400–X.499
ANNUAIRE	X.500–X.599
RÉSEAUTAGE OSI ET ASPECTS SYSTÈMES	X.600–X.699
GESTION OSI	X.700–X.799
SÉCURITÉ	X.800–X.849
APPLICATIONS OSI	X.850–X.899
TRAITEMENT RÉPARTI OUVERT	X.900–X.999
SÉCURITÉ DE L'INFORMATION ET DES RÉSEAUX	
Aspects généraux de la sécurité	X.1000–X.1029
Sécurité des réseaux	X.1030–X.1049
Gestion de la sécurité	X.1050–X.1069
Télébiométrie	X.1080–X.1099
APPLICATIONS ET SERVICES SÉCURISÉS (1)	
Sécurité en multidiffusion	X.1100–X.1109
Sécurité des réseaux domestiques	X.1110–X.1119
Sécurité des télécommunications mobiles	X.1120–X.1139
Sécurité de la toile	**X.1140–X.1149**
Protocoles de sécurité (1)	X.1150–X.1159
Sécurité d'homologue à homologue	X.1160–X.1169
Sécurité des identificateurs en réseau	X.1170–X.1179
Sécurité de la télévision par réseau IP	X.1180–X.1199
SÉCURITÉ DU CYBERESPACE	
Cybersécurité	X.1200–X.1229
Lutte contre le spam	X.1230–X.1249
Gestion des identités	X.1250–X.1279
APPLICATIONS ET SERVICES SÉCURISÉS (2)	
Communications d'urgence	X.1300–X.1309
Sécurité des réseaux de capteurs ubiquitaires	X.1310–X.1319
Sécurité des réseaux électriques intelligents	X.1330–X.1339
Courrier certifié	X.1340–X.1349
Sécurité de l'Internet des objets (IoT)	X.1360–X.1369
Sécurité des systèmes de transport intelligents	X.1370–X.1389
Sécurité de la technologie des registres distribués	X.1400–X.1449
Protocoles de sécurité (2)	X.1450–X.1459
ECHANGE D'INFORMATIONS SUR LA CYBERSÉCURITÉ	
Aperçu général de la cybersécurité	X.1500–X.1519
Echange concernant les vulnérabilités/les états	X.1520–X.1539
Echange concernant les événements/les incidents/l'heuristique	X.1540–X.1549
Echange de politiques	X.1550–X.1559
Heuristique et demande d'informations	X.1560–X.1569
Identification et découverte	X.1570–X.1579
Echange garanti	X.1580–X.1589
SÉCURITÉ DE L'INFORMATIQUE EN NUAGE	
Aperçu de la sécurité de l'informatique en nuage	X.1600–X.1601
Conception de la sécurité de l'informatique en nuage	X.1602–X.1639
Bonnes pratiques et lignes directrices concernant la sécurité de l'informatique en nuage	X.1640–X.1659
Mise en œuvre de la sécurité de l'informatique en nuage	X.1660–X.1679
Sécurité de l'informatique en nuage (autres)	X.1680–X.1699
COMMUNICATIONS QUANTIQUES	
Terminologie	X.1700–X.1701
Générateur quantique de nombres aléatoires	X.1702–X.1709
Cadre de sécurité pour les réseaux QKDN	X.1710–X.1711
Conception de la sécurité pour les réseaux QKDN	X.1712–X.1719
Techniques de sécurité pour les réseaux QKDN	X.1720–X.1729
SÉCURITÉ DES DONNÉES	
Sécurité des mégadonnées	X.1750–X.1759
SÉCURITÉ DE LA 5G	X.1800–X.1819

Pour plus de détails, voir la Liste des Recommandations de l'UIT-T.

Recommandation UIT-T X.1148

Cadre du processus de désidentification pour les fournisseurs de services de télécommunication

Résumé

Les organisations de télécommunication collectent, gèrent, utilisent et partagent des données sur les individus, y compris des informations d'identification personnelle. Aussi utilisent-elles des techniques de désidentification des données pour protéger ces informations de nature privée. La Recommandation UIT-T X.1148 décrit un cadre relatif au processus de désidentification, incluant les étapes opérationnelles, et spécifie les modèles de version de données de même que les types de données qui interviennent dans ce processus pour les fournisseurs de services de télécommunication, sur la base du modèle de cycle de vie des données et les rôles des parties prenantes.

Historique

Edition	Recommandation	Approbation	Commission d'études	Identifiant unique*
1.0	UIT-T X.1148	03-09-2020	17	11.1002/1000/14249

Mots clés

Entité principale des données, désidentification, processus de désidentification, k-anonymat, l-diversité, protection des données PII, modèles de version, t-proximité.

* Pour accéder à la Recommandation, reporter cet URL http://handle.itu.int/ dans votre navigateur Web, suivi de l'identifiant unique, par exemple http://handle.itu.int/11.1002/1000/11830-en.

AVANT-PROPOS

L'Union internationale des télécommunications (UIT) est une institution spécialisée des Nations Unies dans le domaine des télécommunications et des technologies de l'information et de la communication (ICT). Le Secteur de la normalisation des télécommunications (UIT-T) est un organe permanent de l'UIT. Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

L'Assemblée mondiale de normalisation des télécommunications (AMNT), qui se réunit tous les quatre ans, détermine les thèmes d'étude à traiter par les Commissions d'études de l'UIT-T, lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution 1 de l'AMNT.

Dans certains secteurs des technologies de l'information qui correspondent à la sphère de compétence de l'UIT-T, les normes nécessaires se préparent en collaboration avec l'ISO et la CEI.

NOTE

Dans la présente Recommandation, l'expression "Administration" est utilisée pour désigner de façon abrégée aussi bien une administration de télécommunications qu'une exploitation reconnue.

Le respect de cette Recommandation se fait à titre volontaire. Cependant, il se peut que la Recommandation contienne certaines dispositions obligatoires (pour assurer, par exemple, l'interopérabilité et l'applicabilité) et on considère que la Recommandation est respectée lorsque toutes ces dispositions sont observées. Le futur d'obligation et les autres moyens d'expression de l'obligation comme le verbe "devoir" ainsi que leurs formes négatives servent à énoncer des prescriptions. L'utilisation de ces formes ne signifie pas qu'il est obligatoire de respecter la Recommandation.

DROITS DE PROPRIÉTÉ INTELLECTUELLE

L'UIT attire l'attention sur la possibilité que l'application ou la mise en œuvre de la présente Recommandation puisse donner lieu à l'utilisation d'un droit de propriété intellectuelle. L'UIT ne prend pas position en ce qui concerne l'existence, la validité ou l'applicabilité des droits de propriété intellectuelle, qu'ils soient revendiqués par un membre de l'UIT ou par une tierce partie étrangère à la procédure d'élaboration des Recommandations.

A la date d'approbation de la présente Recommandation, l'UIT n'avait pas été avisée de l'existence d'une propriété intellectuelle protégée par des brevets à acquérir pour mettre en œuvre la présente Recommandation. Toutefois, comme il ne s'agit peut-être pas de renseignements les plus récents, il est vivement recommandé aux développeurs de consulter la base de données des brevets du TSB sous http://www.itu.int/ITU-T/ipr/.

TABLE DES MATIÈRES

		Page
1	Domaine d'application	1
2	Références	1
3	Définitions	1
3.1	Termes définis ailleurs	1
3.2	Termes définis dans la présente Recommandation	3
4	Abréviations et acronymes	3
5	Conventions	4
6	Aperçu du processus de désidentification	4
6.1	Modèle de cycle de vie des données et phase de désidentification	4
6.2	Considérations relatives à la désidentification	5
7	Cadre applicable au processus de désidentification	7
7.1	Étape 1 – Examen préliminaire	8
7.2	Étape 2 – Désidentification	9
7.3	Étape 3 – Évaluation de l'adéquation pour le processus de désidentification	9
7.4	Étape 4 – Gestion du suivi	10
8	Utilité des données désidentifiées	11
8.1	Niveaux de désidentification des données	11
8.2	Modèles de diffusion des données	12
8.3	Relation entre le modèle de diffusion de données et le niveau de données	15
Annexe A – Procédures d'évaluation de l'adéquation		16
A.1	Élaboration des documents de base	17
A.2	Constitution d'un groupe d'évaluation	17
A.3	Réalisation de l'évaluation	17
A.4	Mesures de désidentification supplémentaires	18
A.5	Utilisation des données	18
Annexe B – Approches de désidentification des données non structurées		19
Appendice I – Exemples de techniques de désidentification types		21
I.1	Outils statistiques des techniques de désidentification	21
I.2	Outils cryptographiques des techniques de désidentification	21
I.3	Techniques de suppression	21
I.4	Techniques de pseudonymisation	21
I.5	Techniques de généralisation	22
I.6	Techniques de randomisation	22
I.7	Données synthétiques	22
Appendice II – Modèles de processus de désidentification		23
II.1	Modèle de désidentification centré sur les données	23

Page

II.2 Modèle de désidentification centré sur les rôles..... 24

Bibliographie..... 27

Introduction

Compte tenu de l'évolution rapide des services et des technologies de l'information et de la communication basés sur l'Internet, de grandes quantités de données sont générées, transmises et stockées selon un rythme exponentiel. Les données générées proviennent de nombreuses sources: non seulement des capteurs, des caméras ou des équipements de réseau, mais aussi des pages Web, des systèmes de messagerie ou des réseaux sociaux, et bien d'autres. Les ensembles de données sont aujourd'hui si volumineux et complexes et transitent à des vitesses telles que les méthodes et les outils traditionnels de traitement des données ne sont plus adéquats. Il devient de plus en plus difficile de procéder à des analyses de données efficaces dans un délai acceptable. L'analyse des mégadonnées est un paradigme mis en place pour remédier à cette situation.

Les organisations de télécommunication collectent, gèrent, utilisent et partagent des données sur les individus, y compris des informations d'identification personnelle. Aussi utilisent-elles des techniques de désidentification des données pour protéger ces informations de nature privée. Les relations entre les parties impliquées dans le flux de données pour l'échange de données déterminent si le processus de désidentification doit intervenir avant la collecte, après la collecte mais avant le stockage, ou uniquement avant le partage des données avec la partie suivante dans l'échange de données. En conséquence, les fournisseurs de services de télécommunication sont chargés d'assurer la désidentification en tant que service fourni aux clients de données de manière opportune, efficace et sûre.

Recommandation UIT-T X.1148

Cadre du processus de désidentification pour les fournisseurs de services de télécommunication

1 Domaine d'application

La présente Recommandation donne un aperçu du processus de désidentification des données basé sur le modèle de cycle de vie des données et décrit un cadre relatif au processus de désidentification incluant les étapes opérationnelles de même que les rôles des parties prenantes dans le processus. Elle traite également des modèles de version de données et des types de données qui interviennent dans ce processus et fournit différentes approches et exemples de désidentification dans ses annexes et appendices.

La présente Recommandation ne traite pas des questions relatives à la réglementation.

2 Références

Aucune.

3 Définitions

3.1 Termes définis ailleurs

La présente Recommandation utilise les termes suivants définis ailleurs:

3.1.1 données agrégées [b-ISO/CEI 20889]: données représentant un groupe d'entités principales de données, tel qu'une collection de propriétés statistiques de ce groupe.

3.1.2 anonymisation [b-ISO/CEI 29100]: processus par lequel des informations d'identification personnelle (IIP) sont modifiées de manière irréversible de sorte que l'entité principale des IIP ne puisse plus être identifiée directement ou indirectement, soit par le contrôleur des PII agissant seul, soit avec la collaboration de toute autre partie.

3.1.3 attribut [b-ISO/CEI 20889]: caractéristique intrinsèque.

3.1.4 ensemble de données [b-ISO/CEI 20889]: collection de données.

3.1.5 désidentification [b-ISO 25237]: terme générique désignant tout processus de réduction de l'association entre un ensemble de données permettant l'identification et le sujet des données (voir § 3.2.4).

3.1.6 processus de désidentification [b-ISO/CEI 20889]: opération consistant à supprimer l'association entre un ensemble de données permettant l'identification et l'entité à qui appartiennent ces données.

3.1.7 technique de désidentification [b-ISO/CEI 20889]: méthode qui consiste à transformer un ensemble de données dans le but de réduire la mesure dans laquelle les informations peuvent être associées aux entités principales des données individuelles.

3.1.8 ensemble de données désidentifiées [b-ISO/CEI 20889]: ensemble de données résultant de l'application d'un processus de désidentification.

3.1.9 information désidentifiée [b-NISTIR 8053]: enregistrement sur lequel suffisamment d'informations PII ont été supprimées ou masquées de sorte que les informations restantes ne permettent pas d'identifier un individu et qu'il n'y ait aucun motif raisonnable de croire que les informations peuvent être utilisées pour identifier un individu.

3.1.10 confidentialité différentielle [b-ISO/CEI 20889]: formalisation de la mesure de la confidentialité, qui garantit que la distribution de probabilité des résultats d'une analyse statistique diffère au plus d'une valeur spécifiée, qu'une entité principale des données spécifique soit ou non représentée dans l'ensemble de données initial.

NOTE – Plus spécifiquement, la confidentialité différentielle inclut:

a) une définition mathématique de la confidentialité, qui postule que, pour que le résultat d'une analyse statistique soit considéré comme préservant la confidentialité, les résultats de l'analyse de l'ensemble de données d'origine ne peuvent être distingués de ceux obtenus si une entité principale de données est ajoutée ou soustraite de l'ensemble de données; et

b) une mesure de la confidentialité, qui permet de surveiller la perte cumulative de confidentialité et de fixer une limite supérieure ("ou budget") pour la limite de perte. Une définition formelle est la suivante: soit ε un nombre réel positif, et M un algorithme randomisé qui prend pour entrée un ensemble de données. L'algorithme M est dit ε-différentiellement confidentiel si pour tous les ensembles de données D1 et D2 qui diffèrent d'un seul élément (les données pour une seule entité principale de données) et pour tout sous-ensemble S de la plage de M, mml_m1, où la probabilité est fondée sur l'aléa introduit par l'algorithme.

3.1.11 identificateur [b-ISO/CEI 20889]: ensemble d'attributs dans un ensemble de données, qui permet l'identification unique d'une entité principale de données dans un contexte opérationnel spécifique.

NOTE – Voir l'Annexe B pour une discussion sur la façon dont cette définition se rapporte à celles données dans d'autres normes.

3.1.12 attribut d'identification [b-ISO/CEI 20889]: attribut d'un ensemble de données, pouvant contribuer à identifier de manière unique une entité principale de données dans un contexte opérationnel spécifique.

3.1.13 parties intervenant dans le traitement des données personnelles [b-ISO/CEI 29100]: personne physique ou morale, autorité publique, agence ou tout autre organisme susceptible d'affecter, d'être affecté par ou de se percevoir comme étant affecté par une décision ou une activité liée au traitement des informations d'identification personnelles (PII).

3.1.14 pseudonymisation [b-ISO/CEI 20889]: technique de désidentification qui remplace un identificateur (ou les identificateurs) d'une entité principale de données par un pseudonyme afin de cacher l'identité de cette dernière.

3.1.15 quasi-identificateur [b-ISO/CEI 20889]: attribut d'un ensemble de données qui, lorsqu'il est mis en rapport avec d'autres attributs de l'ensemble de données, singularise une entité principale de données.

3.1.16 enregistrement [b-ISO/CEI 20889]: ensemble d'attributs concernant une entité principale de données unique.

3.1.17 réidentification [b-ISO/CEI 20889]: processus d'association des données dans un ensemble de données désidentifiées avec l'entité principale des données d'origine.

NOTE – Un processus établissant la présence d'une entité principale des données spécifique dans un ensemble de données est inclus dans cette définition.

3.1.18 singulariser [b-ISO/CEI 20889]: isoler les enregistrements appartenant à une entité principale des données dans l'ensemble de données, en observant un ensemble de caractéristiques connues pour identifier de manière unique cette entité principale des données.

3.1.19 tiers [b-ISO/CEI 29100]: partie intervenant dans la protection de la vie privée autre que l'entité principale des informations d'identification personnelle (PII), le responsable du contrôle des données PII et le responsable du traitement des données PII, et personnes physiques autorisées à traiter les données sous l'autorité directe du responsable du contrôle des données PII ou du responsable du traitement des données PII.

3.1.20 tiers de confiance [b-ISO/CEI 18014-1]: autorité de sécurité, ou son agent, à laquelle d'autres entités font confiance au regard des activités liées à la sécurité.

3.1.21 k-anonymat [b-ISO/CEI 20889]: modèle formel de mesure de la confidentialité garantissant pour chaque identificateur d'un ensemble de données une classe d'équivalence correspondante contenant au moins K enregistrements.

3.1.22 l-diversité [b-ISO/CEI 20889]: modèle formel de mesure de la confidentialité garantissant que, pour un attribut sélectionné, chaque classe d'équivalence comporte au moins L valeurs bien représentées.

NOTE – La L-diversité est une propriété d'un ensemble de données qui fournit une limite inférieure garantie, L, sur la diversité des valeurs partagées par une classe d'équivalence pour un attribut sélectionné.

3.1.23 t-proximité [b-ISO/CEI 20889]: modèle officiel de mesure de la confidentialité garantissant que la distance entre la distribution d'un attribut sélectionné dans une classe d'équivalence et la distribution de cet attribut dans toute la table ne dépasse un seuil T.

NOTE – Une table respecte la t-proximité à l'égard d'un attribut sélectionné si toutes les classes d'équivalence qui contiennent cet attribut respectent la t-proximité.

3.2 Termes définis dans la présente Recommandation

La présente Recommandation définit les termes suivants:

3.2.1 responsable du contrôle des données: partie (ou partie intervenant dans la protection de la vie privée) qui détermine les finalités et les moyens du traitement des données, autre que les personnes physiques qui utilisent les données à des fins personnelles.

3.2.2 responsable du traitement des données: partie qui traite les données au nom et sur instruction d'un responsable du contrôle des données.

3.2.3 délégué à la protection des données: personne nommée par le responsable du contrôle des informations d'identification personnelle (PII) pour veiller, en toute indépendance, au respect des exigences légales ou réglementaires en matière de vie privée.

NOTE – "responsable du contrôle des données PII" est synonyme de "responsable des données".

3.2.4 sujet des données: entité à laquelle se rapportent les données.

NOTE – "sujet des données" est synonyme d'"entité principale des données PII" et d'"entité principale des données".

3.2.5 processus: en rapport avec des informations ou données, il est ici question de l'obtention, de l'enregistrement ou de la détention d'informations ou de données ou de l'exécution d'une opération ou d'une série d'opérations sur les informations ou données, telles que:

– l'organisation, l'adaptation ou l'altération des informations ou données;
– la récupération, la consultation ou l'utilisation des informations ou données;
– la divulgation des informations ou données par transmission, dissémination ou toute autre méthode de diffusion; ou
– l'alignement, la combinaison, le blocage, l'effacement ou la destruction des informations ou données.

4 Abréviations et acronymes

La présente Recommandation utilise les abréviations et acronymes suivants:

DP confidentialité différentielle (*differential privacy*)

DPO délégué à la protection des données (*data protection officer*)

PII information d'identification personnelle *(personally identifiable information)*

TTP tiers de confiance (*trusted third party*)

5 Conventions

Aucune.

6 Aperçu du processus de désidentification

Le processus de désidentification a pour but de protéger la confidentialité des données des sujets. Étant donné que ces données peuvent inclure des informations d'identification personnelle, avant et après l'analyse des données dans le but d'extraire des informations pertinentes, les analystes de données doivent tenir compte de considérations relatives à la sécurité.

Ce paragraphe présente les environnements d'analyse de données, le modèle de cycle de vie des données, les rôles des entités dans le processus de désidentification et d'autres considérations relatives à la désidentification.

6.1 Modèle de cycle de vie des données et phase de désidentification

De façon générale, l'organisation fixe des objectifs de désidentification à des fins de confidentialité et de sécurité. Le présent paragraphe fournit un exemple de cycle de vie des données et décrit à quel moment il convient d'envisager un processus de désidentification basé sur ce modèle de cycle de vie des données.

Le concept de cycle de vie des données est utilisé pour sélectionner les contrôles appropriés en fonction de l'analyse de la possibilité de réidentification. Un exemple de cycle de vie des données est présenté aux paragraphes 6.1.1 à 6.1.5.

La Figure 1 fournit une vue d'ensemble du processus de désidentification dans le modèle de cycle de vie des données.

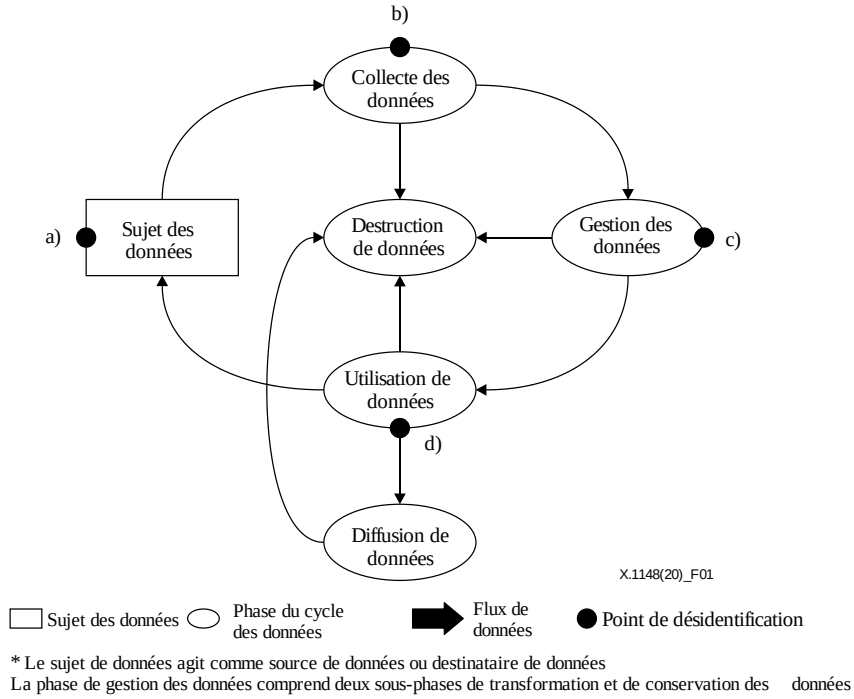

Figure 1 – Processus de désidentification dans le modèle de cycle de vie des données

6.1.1 Phase de collecte des données

Les données sont collectées auprès des sujets de données, qui sont les personnes auxquelles les données se réfèrent. L'ensemble de données ainsi collectées peut inclure des informations PII. La

désidentification crée un nouvel ensemble de données duquel toutes les informations PII ont été supprimées. Il est recommandé que les ensembles de données désidentifiées soient, si possible, utilisés en interne par une organisation au lieu et place de l'ensemble de données d'origine.

Sur la base de ce modèle, la désidentification peut intervenir soit:

– pendant la phase de collecte des données, (b) dans la Figure 1; ou
– lorsque des données ont été collectées mais que l'identificateur n'était pas réellement nécessaire, (a) dans la Figure 1.

Les identificateurs qui ne sont pas nécessaires à la gestion des données (transformation et conservation des données) ne devraient pas être collectés.

6.1.2 Phase de gestion des données

Pour éviter que l'identificateur ne soit archivé, la désidentification doit intervenir après la transformation des données et avant leur conservation, (c) dans la Figure 1. Il est recommandé aux organisations de tenir compte du potentiel de réidentification et de mettre en place des contrôles d'accès clairs, des limites maximales de conservation et des politiques de suppression des données qui réduisent au maximum le potentiel de liaison entre les données désidentifiées. Il est recommandé aux organisations d'envisager des techniques d'anonymisation telles que l'agrégation des données lorsque cela est adapté à l'objectif d'utilisation prévu.

6.1.3 Phase d'utilisation des données

Si des données d'identification personnelle s'avèrent nécessaires au sein d'une organisation à des fins de gestion, il est recommandé de désidentifier les données avant de les publier comme ensemble de données destinées au partage, (d) dans la Figure 1.

6.1.4 Phase de diffusion des données

Les données peuvent être partagées avec des tiers qui sont liés par des contrôles administratifs supplémentaires tels que des "accords sur le partage des données". Des ensembles de données désidentifiées peuvent également être publiés. La procédure de diffusion se décline en trois modèles: public, semi-public et non public. Le degré de désidentification requis peut varier en fonction du modèle de diffusion choisi.

6.1.5 Phase de destruction des données

La destruction des données peut intervenir à n'importe quelle étape, qu'il s'agisse de la collecte, de la gestion, de l'utilisation ou de la diffusion des données. Les données devraient être détruites sur la base de procédures vérifiables de façon à éviter la récupération des données. La destruction des données devrait notamment être envisagée lors de la détection d'une possibilité de réidentification.

6.2 Considérations relatives à la désidentification

L'application de la désidentification tout au long du cycle de vie des données augmente son efficacité. Cependant, la nature des relations entre les parties participant au flux de données détermine la nécessité ou non de procéder à la désidentification des données avant leur collecte, (a) dans la Figure 1, après leur collecte, (b) dans la Figure 1, mais avant leur conservation, (c) dans la Figure 1, ou seulement avant qu'elles ne soient partagées avec la partie suivante dans le flux de données, (d) sur la Figure 1. Cette décision, à son tour, influe sur la capacité des mesures de sécurité et des autres mesures organisationnelles à améliorer l'efficacité d'une technique de désidentification particulière dans chaque cas d'utilisation. Même si la désidentification s'avère utile pour protéger la confidentialité des données des sujets lorsque l'objectif d'utilisation ne vient pas corroborer les techniques d'anonymisation, la technique n'est pas suffisante en soi pour protéger les données des sujets et doit être considérée comme faisant partie d'un cadre global pour la protection des données. Ce paragraphe décrit les caractéristiques et les considérations associées à chacune des phases.

6.2.1 Collecte des données

La désidentification locale (ou désidentification à la source), qui permet à un individu (ou à un responsable du traitement des données pour un individu) de supprimer toutes les informations PII avant la diffusion des données pour analyse, est l'approche la plus couramment utilisée.

L'un des aspects de la désidentification directement lié à la phase de collecte est la minimisation des données. Chaque responsable recueillant les données des sujets doit définir avec précision quelles sont les données strictement nécessaires à l'objectif d'utilisation et limiter la collecte de ces données aux seuls paramètres définis.

Des processus spécifiques devraient être mis en place pour exclure les informations d'identification personnelle inutiles de la collecte et du transfert des données, de manière à réduire le champ des données.

Un autre aspect de la désidentification est l'agrégation des données. Les responsables du contrôle sont tenus d'envisager l'agrégation des données dans tous les cas où l'objectif d'utilisation ne nécessite pas strictement de distinguer les différents sujets.

6.2.2 Gestion des données

6.2.2.1 Transformation des données

La phase de transformation des données peut inclure des techniques de désidentification (agrégation, limitations en matière divulgation des statistiques, chiffrement, etc.). La transformation des données peut se faire sur une ou plusieurs étapes, y compris directement après la collecte et avant la conservation à long terme, après une période substantielle de conservation et avant l'accès, ou de manière intégrée à l'accès.

En principe, les données peuvent être remaniées ou agrégées à tout moment après la collecte et jusqu'à la diffusion. Si la transformation intervient immédiatement après la collecte, l'intervention peut limiter le préjudice potentiel pour les sujets en cas de violation des données, mais elle peut également réduire la possibilité de lier, de fusionner ou de mettre à jour les données après le remaniement.

Le choix de la méthode de transformation des données devrait être effectué après un examen minutieux du préjudice potentiel de l'exposition pour les sujets de données. La décision de transformation devrait également tenir compte des analyses qui seront ultérieurement étayées par l'objectif d'utilisation, dans la mesure où les techniques utilisées pour réduire les risques de divulgation peuvent affecter le potentiel en vue d'utilisations et d'analyses futures.

6.2.2.2 Conservation des données

La conservation des données est définie comme le processus de stockage des données, y compris des informations d'identification personnelle, vers toute forme de stockage non volatile par un responsable du contrôle des données ou une partie agissant sous sa direction. Les contrôles relatifs à la sécurité et à la confidentialité des informations portent déjà majoritairement sur la phase de conservation des données; aussi le présent paragraphe présente-t-il sommairement ces contrôles sans fournir d'informations détaillées [b-ISO/CEI 27001]. Un certain nombre de contrôles de sécurité et de confidentialité sont couramment rencontrés au stade de la conservation, tels que le contrôle d'accès, la maintenance, les évaluations de sécurité, les procédures d'authentification, la surveillance et la réponse aux incidents et les audits.

En particulier, les organisations devraient appliquer des principes de durée de conservation maximale et de suppression des données, visant à garantir que les données ne seront pas conservées plus longtemps que ce qui est strictement nécessaire pour l'objectif d'utilisation prévu et que ces dernières seront entièrement détruites au terme de la durée de conservation maximale. Ainsi, les accords sur le partage des données fixent souvent une durée de conservation maximale au terme de laquelle le destinataire devra détruire les données, par exemple un an après leur réception; les lois peuvent également exiger une disposition contractuelle en ce sens.

6.2.3 Utilisation des données

Les données désidentifiées peuvent être collectées, stockées ou partagées pour diverses finalités et applications, reposant chacune sur la préservation de certaines propriétés de données après la désidentification. L'une des principales raisons de la diffusion d'ensembles de données désidentifiées est de fournir à d'autres la possibilité d'étudier les valeurs et les propriétés des données brutes à des fins de recherche [b-ISO/CEI 20889]. La désidentification viserait par conséquent à préserver autant que possible l'utilité des informations, tout en protégeant la vie privée des individus. Ce double objectif en fait une approche importante à considérer pour une utilisation dans un certain nombre de contextes, incluant les modèles de diffusion des données.

Lors de la diffusion des données désidentifiées, une organisation doit prendre la décision, généralement via un comité d'experts comprenant un large éventail de parties prenantes, d'examiner les impacts potentiels sur les sujets concernés par la diffusion. Les évaluations des risques et les listes de contrôle sont souvent utilisées pour guider cette évaluation et définir un mécanisme de diffusion approprié pour atténuer le risque de réidentification.

Le choix des techniques de désidentification dépend du degré de leur applicabilité ou de leur "utilité" dans un cas d'utilisation particulier.

7 Cadre applicable au processus de désidentification

Ce paragraphe décrit un cadre applicable au processus de désidentification visant à fournir des données PII désidentifiées. Ce cadre se présente en quatre étapes, comme le montre la Figure 2 [b-KOREA].

Étape 1 – Examen préliminaire

L'étape 1 consiste à vérifier si les données visées contiennent ou non des informations d'identification personnelle. Si tel est le cas, passer à l'étape 2. Une désidentification est requise.

Étape 2 – Désidentification

L'étape 2 consiste à désidentifier les données pour empêcher toute déduction des informations individuelles spécifiques depuis l'ensemble de données visé. Cette étape implique des mesures de suppression ou de transformation des éléments PII en totalité ou en partie. Les éléments PII comprennent les identificateurs, les quasi-identificateurs et les attributs sensibles.

Étape 3 – Évaluation de l'adéquation

L'étape 3 consiste à évaluer l'adéquation de l'ensemble de données désidentifiées incluant les éléments PII. Parmi les facteurs pertinents à considérer se pose la question de savoir si l'ensemble de données visé contient encore des informations d'identification personnelle, s'il existe une possibilité directe de réidentification ou une possibilité de mise en lien pouvant mener à la réidentification.

Étape 4 – Gestion du suivi

L'étape 4 consiste à mesurer le niveau de sécurité managériale et technique pour prévenir la réidentification.

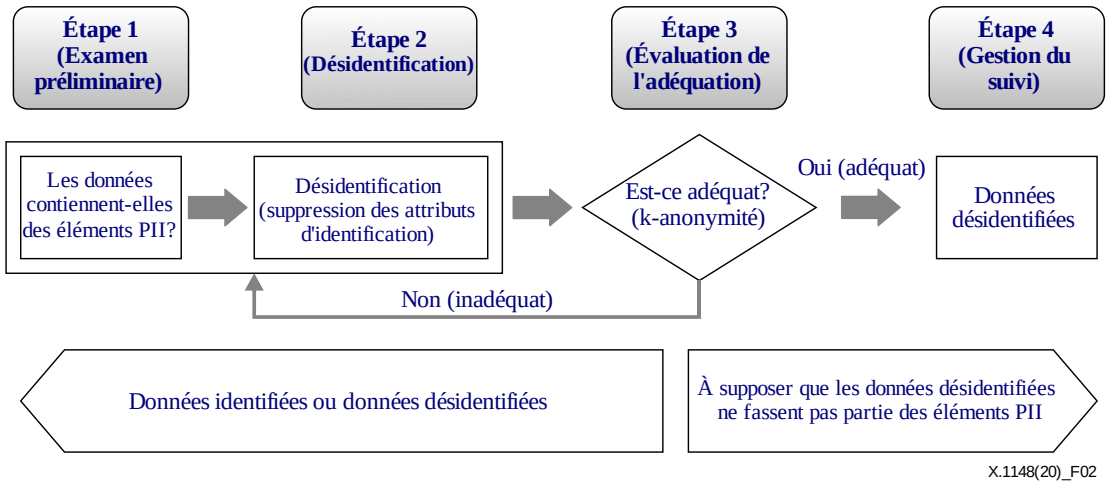

Figure 2 – Processus de désidentification

Chacune de ces étapes est détaillée aux paragraphes 7.1 à 7.4.

7.1 Étape 1 – Examen préliminaire

Les organisations qui ont l'intention d'utiliser ou de fournir des données à diverses fins devraient tout d'abord déterminer leurs politiques et normes. Il est recommandé que les politiques et normes incluent les éléments suivants:

– Quel est le but et l'usage prévu des informations désidentifiées?
– De quel type d'attributs de données les données désidentifiées sont-elles composées?
– Quelles sont les techniques utilisées pour la désidentification?
– Quels sont les niveaux de risque et les effets pervers de la réidentification?
– Quelles sont les solutions disponibles si un individu spécifique est réidentifié?
– Comment le niveau de réidentification est-il évalué?
– Comment la main d'œuvre et les coûts de désidentification sont-ils définis?

Les considérations spécifiques qui constituent l'examen préliminaire peuvent varier selon le type de données et l'utilisation prévue. Cependant, il est recommandé d'établir un ensemble de normes.

Les organisations qui ont l'intention de traiter des données à des fins diverses devront se référer aux normes en vigueur pour vérifier si les données spécifiques contiennent ou non des éléments PII. Même si ce n'est pas le cas, l'organisation est tenue d'envisager toute possibilité de mise en lien des données disponibles avec les mesures appropriées de réduction des risques. En présence d'éléments PII, l'étape de désidentification est requise.

Les critères de jugement applicables aux éléments PII sont notamment les suivants:

– Il n'y a pas de limitations spéciales sur les données concernant leur type, leur forme, leurs caractéristiques et leur format.
– Si un responsable du contrôle des données peut identifier un individu à l'aide de données, ces données sont considérées comme des informations d'identification personnelle.
– Les données doivent se rapporter à un individu. La valeur statistique d'un groupe composé de plusieurs individus n'est pas une information d'identification personnelle.
– Les données qui peuvent identifier un individu via une combinaison avec des informations supplémentaires sont considérées comme des informations d'identification personnelle. Les informations supplémentaires font normalement référence à des informations accessibles au public/facilement accessibles.

7.2 Étape 2 – Désidentification

7.2.1 Désidentification pour les identificateurs

Un "identificateur" est une donnée, telle qu'une valeur ou un nom, qui est assignée de manière unique à un individu ou à une chose liée à un individu. Il convient, en règle générale, de minimiser l'ensemble des "identificateurs", mais les identificateurs inclus dans les ensembles de données ne devraient pas être supprimés.

Un identificateur qui est strictement nécessaire pour l'objectif poursuivi peut cependant contenir certaines données, telles que:

– identificateur unique (numéro de résident, numéro de sécurité sociale, numéro de passeport, numéro d'identification d'étranger, numéro de permis de conduire, etc.);
– nom (en caractères chinois, nom anglais, etc.);
– adresse détaillée (numéro de maison, numéro de rue, etc.);
– date (date d'anniversaire (mariage, etc.), date de certificat, etc.);
– numéro de téléphone (mobile, domicile, télécopie, etc.);
– numéro de dossier médical, numéro national d'assurance maladie, numéro de destinataire de l'aide sociale, etc.;
– numéro de compte en banque, numéro de carte de crédit, etc.;
– photos (image fixe, vidéo de télévision en circuit fermé (CCTV), etc.);
– données biométriques (empreinte digitale, vocale, de l'iris, etc.);
– adresse de courriel, adresse IP, adresse de commande d'accès au support (MAC), localisateur uniforme de ressource (URL) de la page d'accueil, etc.;
– code d'identification (numéro d'employé, numéro de client, etc.);
– autre numéro d'identification unique (numéro de matricule militaire, numéro d'enregistrement de l'entreprise, etc.).

7.2.2 Désidentification pour les quasi-identificateurs et attributs hautement identifiables

De manière générale, les quasi-identificateurs inclus dans les ensembles de données devraient être supprimés s'ils sont sans rapport avec l'objectif d'utilisation prévu pour les données. Des techniques de désidentification telles que la pseudonymisation et l'agrégation devraient être appliquées dans les cas où un quasi-identificateur lié à l'utilisation des données comporte des éléments identifiables.

Les données à haut potentiel d'identification, à l'instar des données comportementales, nécessitent l'utilisation de techniques de désidentification et, lorsque cela est possible, d'anonymisation.

7.2.3 Techniques de désidentification

Diverses techniques, telles que la pseudonymisation, l'agrégation, la suppression des données et le masquage des données, peuvent être utilisées individuellement ou en association. Appliquée seule, la technique de la pseudonymisation peut ne pas être suffisante.

Il existe différents types de techniques, toutes aisément accessibles. Il convient de sélectionner la technique la plus adaptée et de l'appliquer sur la base de l'objectif d'utilisation prévu, en tenant compte des forces et des faiblesses propres à chacune de ces techniques. Après la désidentification, on peut passer à l'étape suivante.

7.3 Étape 3 – Évaluation de l'adéquation pour le processus de désidentification

Si la désidentification ne suffit pas, il peut être possible d'identifier un individu en combinant d'autres données ou en utilisant plusieurs techniques d'inférence.

Afin de réduire le risque de réidentification, il est nécessaire de procéder à une évaluation de l'adéquation des données désidentifiées avant utilisation. Les questions ci-après notamment doivent être posées:

– Quel est le but de cette demande de désidentification?
– Quels sont les types d'attributs de données impliqués dans la désidentification (incluant les identificateurs ou non)?
– Quel est le niveau adéquat de désidentification?

Cette évaluation de l'adéquation pourrait être réalisée par un délégué à la protection des données (DPO), un tiers de confiance (TTP) mandaté ou un panel d'évaluation externe.

Il peut être fait appel à des modèles de protection de la vie privée, dont le modèle de la k-anonymité, pour procéder à cette évaluation de l'adéquation. Le modèle de la k-anonymité est l'outil de base en matière d'évaluation. D'autres modèles (l-diversité, t-proximité, confidentialité différentielle (DP), etc.) peuvent être appliqués au besoin.

Se référer à l'Annexe A pour plus de détails sur l'évaluation de l'adéquation.

7.4 Étape 4 – Gestion du suivi

7.4.1 Mesures de protection pour les données désidentifiées

Des mesures de protection sont mises en œuvre pour éviter que les données désidentifiées ne puissent être réidentifiées en cas de fuite et/ou de combinaison avec d'autres données. Il peut s'agir notamment des mesures suivantes:

– mesures de protection managériales: désignation du responsable des fichiers de données désidentifiées, détermination du partage des données désidentifiées et destruction des données une fois l'objectif d'utilisation atteint;
– mesures de protection techniques: restriction de l'accès aux fichiers de données désidentifiées, gestion des enregistrements d'accès, et installation et exploitation des programmes de sécurité.

Par ailleurs, les mesures de sécurité comprennent également des mesures de protection à prendre en cas de fuite de données désidentifiées. Il peut s'agir notamment des mesures suivantes:

– analyse des motifs de fuite et mise en œuvre de mesures de sécurité managériales et techniques pour éviter les fuites supplémentaires;
– retrait et destruction des données en cas de fuite.

7.4.2 Surveillance des possibilités de réidentification

Le responsable du contrôle des données qui a l'intention d'utiliser des données désidentifiées ou de les fournir à un tiers surveillera régulièrement les possibilités de réidentification.

Si une possibilité de réidentification est détectée, il est demandé au responsable du contrôle des données qui a reçu les données désidentifiées d'interrompre le traitement des données, de les retirer et de les détruire.

7.4.3 Exigences concernant les contrats avec un tiers

Tout contrat avec un tiers visant à fournir ou à attribuer des données désidentifiées à des fins d'utilisation doit tenir compte des risques de réidentification. La gestion de ces risques inclut:

– la notification aux sujets de la divulgation de données à des parties tierces;
– la fourniture de données anonymisées à des parties tierces à chaque fois que cela est possible;

– l'interdiction de réidentification: le contrat stipule que le responsable du contrôle chargé de traiter les données de désidentification n'a pas le droit de réidentifier les données en les combinant avec d'autres données;
– restrictions sur les nouvelles fournitures ou attributions: le contrat stipule les limites contractuelles autorisées en matière de fourniture ou d'attribution du processus de traitement des données désidentifiées;
– notification des risques de réidentification: le contrat stipule l'obligation de stopper le traitement des données et d'informer l'expéditeur et le destinataire sur le thème de la réidentification lorsque les données sont réidentifiées ou que les possibilités de réidentification sont plus grandes.

7.4.4 Contremesures applicables à la réidentification

Dans le cas où les données désidentifiées sont réidentifiées, il convient de stopper le traitement des données et de prendre les mesures nécessaires pour éviter les fuites de données PII.

Les données réidentifiées doivent être immédiatement détruites.

8 Utilité des données désidentifiées

8.1 Niveaux de désidentification des données

Ce paragraphe décrit les étapes de la désidentification des données, qui correspondent chacune à un type de données permettant, à des degrés divers, d'identifier directement un individu ou de montrer dans quelle mesure l'individu est associé aux caractéristiques (ou attributs) propres aux données. La spécification des données dans le contexte de l'utilisation ou du traitement des données devrait non seulement inclure le type de données, mais également décrire dans quelle mesure les données peuvent permettre d'identifier un individu ou d'associer un individu à un ensemble de caractéristiques propres aux données.

La Figure 3 présente les différentes étapes du processus de désidentification, depuis les données identifiées jusqu'aux données désidentifiées. Le risque de réidentification varie en fonction de l'étape concernée. À chaque type de données correspondent des étapes spécifiques qui doivent s'appliquer à l'ensemble de données au fur et à mesure du processus de désidentification.

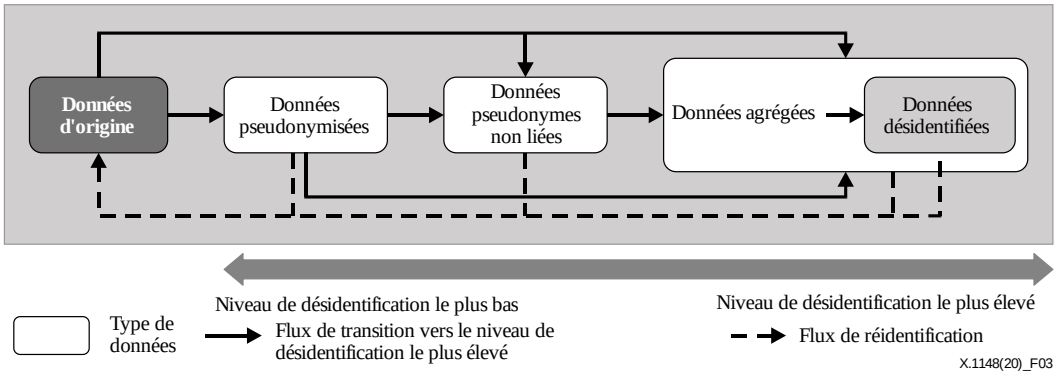

Figure 3 – Niveaux de désidentification des données

Comme le montre la Figure 3, toutes les données sont présentes au stade de la désidentification. On trouve à droite (au niveau de désidentification le plus élevé), les données désidentifiées qui ne sont pas liées à des individus (par exemple, des relevés météorologiques historiques) et pour lesquelles il n'existe aucun risque de confidentialité. À l'extrême gauche (au niveau de désidentification le plus bas), on a les données identifiées qui sont directement liées à des individus spécifiques. Entre ces deux extrêmes, se trouvent des données qui peuvent être liées moyennant un effort, qui ne peuvent

être liées qu'à des groupes de personnes, et qui sont basées sur des individus mais qui ne peuvent pas être reliées à eux. En général, les processus de désidentification sont conçus pour pousser les données vers la droite tout en conservant l'utilité souhaitée, ce qui réduit le risque de fournir des données désidentifiées à une population plus large ou au grand public.

8.1.1 Niveau des données d'origine

Au stade initial des données identifiées, les données peuvent être associées sans ambiguïté à une personne spécifique, dans la mesure où un individu est observable dans les informations. On trouvera des conseils sur ce qui peut être considéré comme des identificateurs dans le § 4.4.1 de [b-ISO/CEI 29100].

8.1.2 Niveau des données pseudonymisées

Au stade des données pseudonymisées, les données ne peuvent pas être raisonnablement modifiées autrement que par la partie qui a procédé à l'assignation d'alias, car tous les identificateurs sont remplacés par des alias. Toutefois, les données pseudonymisées peuvent encore être réidentifiées grâce aux liens existants avec d'autres données.

Voir le terme "pseudonymisation" défini au § 3.1.14.

8.1.3 Niveau des données pseudonymes non liées

Au stade des données pseudonymes non liées, tous les identificateurs sont supprimés ou remplacés par des alias, pour lesquels la fonction d'assignation est effacée ou irréversible, de sorte qu'un lien ne puisse être raisonnablement rétabli par quiconque, notamment la partie à l'origine des assignations. Toutefois, les données pseudonymes non liées peuvent encore être réidentifiées grâce aux liens existants avec d'autres données.

8.1.4 Niveau des données agrégées

À ce stade, les données fournissent des informations sur un nombre suffisant de personnes de façon à ce que les attributs au niveau individuel ne puissent être déduits en tant que données statistiques combinées et ne contenant pas d'entrées au niveau individuel. Si l'on utilise des techniques d'agrégation, les données agrégées n'atteignent pas le degré d'identifiabilité en dessous d'un seuil si la taille des cellules, pour un niveau donné de certaines combinaisons de variables, peut conduire quelqu'un à identifier un individu particulier.

Voir le terme "données agrégées" défini au § 3.1.1.

8.1.5 Niveau des données désidentifiées

Au stade des données désidentifiées, les données sont non liées et les attributs sont modifiés (par exemple, les valeurs des attributs sont randomisées ou généralisées) de telle sorte qu'il existe un niveau de confiance raisonnable qu'une personne ne puisse pas être identifiée, directement ou indirectement, par les données seules ou en combinaison avec d'autres données.

8.2 Modèles de diffusion des données

Le modèle de diffusion des données désidentifiées se décline en trois versions selon la situation des contextes d'analyse de données [b-UKAN].

Il existe donc trois modèles différents pour diffuser les données désidentifiées: public, semi-public et non public.

Chaque modèle fournit différents niveaux de disponibilité et de protection des informations. La pertinence d'un système par rapport à un autre peut varier en fonction de la finalité de la diffusion des données et/ou des exigences législatives y relatives. Le modèle de diffusion joue un rôle important dans le processus de désidentification, sachant que le degré de désidentification peut varier selon le modèle de diffusion sélectionné.

Chacun des trois modèles est présenté aux paragraphes 8.2.1 à 8.2.3.

8.2.1 Modèle public diffusion des données

Dans la version publique traditionnelle, n'importe qui peut accéder aux données sans inscription et sans condition. Des exemples de ce modèle incluent les données publiquement accessibles des organisations et les données postées vers un répertoire des données en accès libre tel qu'un portail web. Les organisations publient de manière proactive des ensembles de données et les mettent gratuitement à la disposition de quiconque pour les utiliser et les republier.

Lors de la diffusion publique de données, il est courant d'imposer aussi peu de restrictions que possible en termes d'informations, y compris concernant qui peut y accéder et comment. Dans ces conditions, si les personnes qui téléchargent l'ensemble de données ne peuvent pas être identifiées, ces divulgations devraient être traitées comme des diffusions publiques de données.

Même en cas d'accès aux demandes d'information mentionnées au § 8.2.2, cela devrait être traité comme des diffusions publiques de données dans les cas qui ne nécessitent pas que la personne demandant des informations accepte les termes ou conditions concernant le traitement, la confidentialité ou la sécurité des informations.

8.2.2 Modèle semi-public de diffusion des données

Le modèle semi-public de partage des données est plus restrictif que le modèle public et se justifie s'il existe une procédure formelle de demande et d'approbation pour accéder aux données. Dans ce cas, le destinataire des données peut accepter certaines conditions d'utilisation ou signer un "contrat de clic". Les contrats de clic sont des conditions d'utilisation en ligne qui peuvent imposer des restrictions sur ce qui peut être fait avec les données et la façon dont les données sont traitées. Cependant, n'importe qui peut télécharger ces données.

La désidentification peut également être utile en ce qui concerne l'accès aux demandes d'information pour les ensembles de données. En utilisant la désidentification, les organisations peuvent répondre aux demandes de manière à protéger la vie privée tout en préservant l'utilité des informations. Les organisations peuvent utiliser des contrôles d'accès pour certaines restrictions lors du partage de données via un système d'information, par exemple:

– demander à ce que tous les utilisateurs s'enregistrent et fournissent leurs coordonnées avant d'accéder aux données;
– utiliser des protocoles d'authentification pour vérifier l'identité d'une personne;
– utiliser des systèmes d'accès hiérarchisé permettant d'accorder différents niveaux d'accès à différentes parties en fonction, par exemple, des affiliations ou des informations d'identification de la personne.

Avec de tels systèmes d'information, un système d'interrogation interactif pourrait être mis à la disposition d'une communauté de chercheurs, et des données brutes pourraient être mises à la disposition d'un petit nombre d'analystes agréés par une procédure minutieuse de sélection.

Se présente toutefois le cas de l'accès des données ne nécessitant aucun partage de données lorsque les analystes demandent à ce que le responsable du contrôle des données effectue une analyse en leur nom. Par conséquent, ce cas pourrait ne pas impliquer le partage de données par l'organisation.

8.2.3 Modèle non public de diffusion des données

Les ensembles de données contenant des informations d'identification personnelle peuvent être partagés au sein des organisations et entre ces dernières uniquement si la divulgation est autorisée par les directives réglementaires du pays. Si la divulgation n'est pas autorisée et que les institutions souhaitent toujours partager des ensembles de données, les éventuelles données PII doivent être supprimées. Les diffusions de données non publiques présentent une moindre disponibilité, mais offrent une protection plus élevée, nécessitant un plus faible niveau de désidentification.

Lors du partage d'informations entre organisations, l'accès à l'ensemble de données étant limité à l'organisation, les exigences relatives à la confidentialité et à la sécurité des informations peuvent être définies et appliquées par le biais d'un accord de partage de données. Pour qu'une divulgation de données soit considérée comme non publique, il doit y avoir un accord de partage de données entre les parties. L'accord de partage de données est un élément important de la stratégie d'atténuation des risques pour ces versions et inclut certaines clauses courantes concernant:

– la spécification de ces accès autorisés (contrôles du destinataire);
– les exigences spécifiques à la sécurité des données (contrôles de l'infrastructure);
– les restrictions d'utilisation, en particulier l'interdiction de lien avec d'autres fichiers et de réidentification délibérée (autres contrôles de données et de gouvernance);
– les exigences de destruction des données une fois l'utilisation terminée (contrôles de gouvernance).

L'accord de partage de données poursuit un triple objectif:

– Il établit une distinction claire entre les personnes ou organisations auxquelles le responsable du contrôle des données fait confiance et celles auxquelles il ne fait pas confiance.
– Il pose un cadre spécifiant les conditions de l'accès.
– Il peut prévoir des sanctions ou des pénalités si les individus/organisations viennent à transgresser ces conditions d'accès.

8.2.4 Comparaison des modèles de diffusion des données

Dans un environnement de flux de données, l'un des moyens pour limiter les possibilités de réidentification est de contrôler la façon dont les données peuvent être obtenues et utilisées. Ces contrôles peuvent être classés selon différents modèles de diffusion des données, chacun présentant des avantages et des risques spécifiques. Les organisations peuvent également choisir d'adopter une approche d'accès à plusieurs niveaux, combinant plusieurs modèles pour répondre à une variété de cas d'utilisation et de menaces à la vie privée. De plus, les modèles de diffusion des données devraient considérer la possibilité de diffusions multiples ou périodiques. Les modèles nommés vont de "zéro restriction" à des "restrictions strictes". Le Tableau 2 présente une comparaison des modèles de diffusion des données.

Tableau 2 – Comparaison des modèles de diffusion des données

	Modèle public	**Modèle semi-public**	**Modèle non public**
Droits d'accès	• Tout le monde a librement accès aux données publiées	• L'accès aux données publiées (ou sous-ensemble) est restreint à des individus ou organisations	• L'accès aux données publiées est restreint à un sous-ensemble d'individus ou d'organisations
Cas d'utilisation	• Accès aux données sans restriction via un portail web (en libre accès pour tous)	• Cadre sécuritaire en ligne • Accès fourni • Accès virtuel à distance • Accès via un serveur d'analyse	• Partage entre les organisations et au sein des organisations
Droits	• Droits illimités de réutilisation et de redistribution des données	• Mise à disposition d'individus et organisations autorisés	• Interdiction de réutilisation, de rediffusion et de distribution des données

Tableau 2 – Comparaison des modèles de diffusion des données

	Modèle public	Modèle semi-public	Modèle non public
Attaque de réidentification	• Attaque de démonstration pour publicité	• Attaque d'initié délibérée • Reconnaissance fortuite d'un individu dans l'ensemble de données par une connaissance • Fuite de données	

8.3 Relation entre le modèle de diffusion de données et le niveau de données

8.3.1 Modèle non public de diffusion des données

Lors du partage de données depuis une source de données vers le modèle de diffusion non public, les données doivent être désidentifiées. Dans des circonstances normales, en dehors de ce modèle, il serait possible d'utiliser des données pseudonymisées non liées et des données de désidentification de niveau supérieur. Dans le cas présent, des outils de désidentification tels que la pseudonymisation, la cryptographie, les données synthétisées, la suppression, etc., peuvent être utilisés.

Toutefois, s'il existe un contrat ou une loi spéciale régissant les relations entre les deux parties, les données pseudonymisées pourraient être utilisées pour analyser et stocker des données pendant cette phase.

8.3.2 Modèle semi-public de diffusion des données

Lors du partage de données depuis une source de données vers le modèle de diffusion semi-public, un niveau de désidentification plus élevé qu'avec le modèle non public sera requis. Il y a lieu de procéder à un traitement statistique pour empêcher la réidentification. À l'issue, des données agrégées et des données de désidentification de niveau supérieur pourraient être diffusées à destination du modèle semi-public. Plus spécifiquement, des outils de désidentification tels que les statistiques, la randomisation, etc., peuvent être utilisés.

Comme le montre le Tableau 2, un niveau de désidentification relativement moins élevé qu'avec le modèle de diffusion public peut être admis, dans la mesure où seules certaines personnes ou organisations peuvent accéder aux données.

8.3.3 Modèle public de diffusion des données

Lors du partage de données depuis une source de données vers le modèle de diffusion public, un niveau de désidentification plus élevé qu'avec le modèle semi-public est requis. Il y a lieu d'exécuter le processus pour obtenir des données désidentifiées, après quoi les résultats peuvent être utilisés pour le modèle de diffusion public comme le montre le Tableau 2.

Annexe A

Procédures d'évaluation de l'adéquation

(Cette annexe fait partie intégrante de la présente Recommandation.)

La présente annexe fournit un modèle applicable aux procédures d'évaluation de l'adéquation [b-KOREA]. Voir la Figure A.1.

Les étapes de la procédure d'évaluation de l'adéquation sont les suivantes:

– Élaboration des documents de base. Un responsable du contrôle des données prépare les documents de base nécessaires à l'évaluation de l'adéquation tels qu'un énoncé de données, le statut de désidentification et le niveau de gestion des organisations d'utilisateurs. On entend par "organisation d'utilisateurs" une organisation qui a l'intention d'utiliser les données désidentifiées à l'issue du processus de désidentification.

– Constitution d'un groupe d'évaluation. Un agent responsable de la confidentialité peut mettre sur pied un groupe d'évaluation ou appeler un DPO ou TTP pour procéder à l'évaluation.

– Réalisation de l'évaluation. Le groupe d'évaluation apprécie le niveau d'adéquation de la désidentification en utilisant les documents de base préparés par le gestionnaire PII.

– Mise en œuvre d'une désidentification supplémentaire. Le responsable du contrôle des données procède à une désidentification supplémentaire tenant compte des opinions des participants à l'évaluation si le résultat de l'évaluation est inadéquat.

– Utilisation des données. Les données peuvent être utilisées ou fournies à des fins telles que l'analyse des mégadonnées si l'évaluation de la désidentification s'avère adéquate.

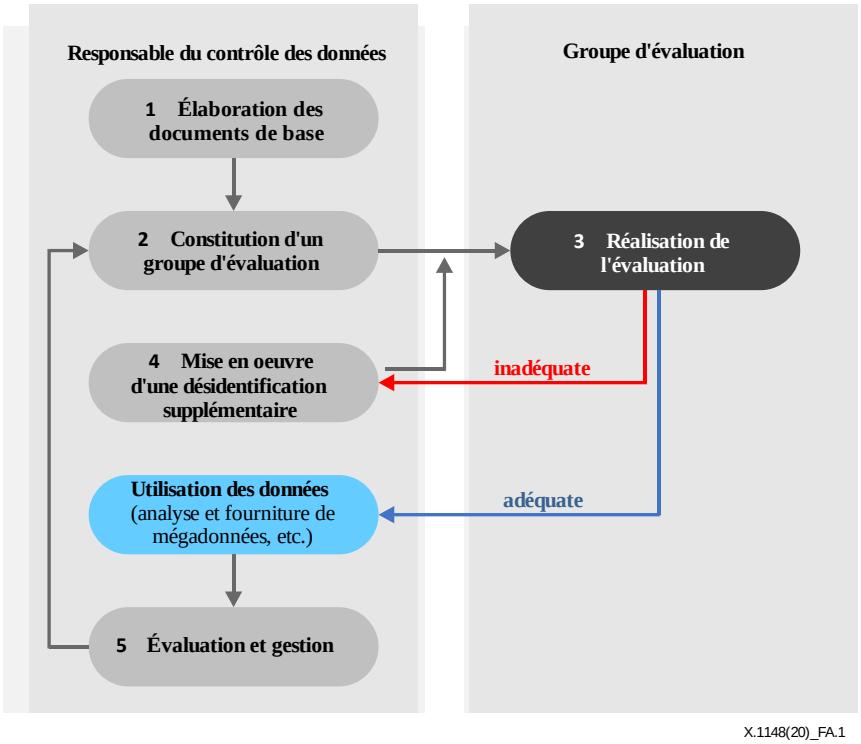

Figure A.1 – Procédure d'évaluation de l'adéquation en cas de désidentification

A.1 Élaboration des documents de base

Un responsable du contrôle des données prépare les documents de base nécessaires à l'évaluation de l'adéquation tels qu'un énoncé de données, le statut de désidentification et le niveau de gestion des organisations d'utilisateurs.

A.2 Constitution d'un groupe d'évaluation

Un agent responsable de la confidentialité peut mettre sur pied un groupe d'évaluation. Il peut nommer plus d'un expert juridique dans le domaine de la désidentification depuis un pool d'experts géré par des agences spécialisées dans chaque domaine et faire appel à des professionnels externes.

Le groupe d'évaluation se compose de membres qui n'ont pas d'intérêt direct dans le but de l'utilisation des données.

A.3 Réalisation de l'évaluation

Le groupe d'évaluation apprécie l'adéquation de la désidentification en utilisant les documents de base et le modèle de k-anonymité.

– Examen préliminaire. Révision des documents de base préparés par le responsable du contrôle des données et vérification, dans le cadre d'un entretien, de la présence d'éléments d'identification personnelle dans un ensemble de données et du caractère approprié du but d'utilisation et des techniques de désidentification.
– Analyse des éventuelles tentatives de réidentification. Analyse des possibles tentatives de réidentification incluant l'intention, le niveau de protection des éléments PII et la capacité du responsable du contrôle des données qui utilise ou reçoit les données.
– Analyse de l'impact de la réidentification. Évaluation de l'impact potentiel sur un sujet de données lorsque les données sont réidentifiées de manière intentionnelle ou non intentionnelle.
– Analyse quantitative. Vérification de l'exactitude d'une valeur K fournie par un responsable du contrôle des données.
– Détermination de la valeur de référence pour l'évaluation. Le groupe d'évaluation définit globalement la valeur de référence pour l'évaluation en tenant compte des possibilités de réidentification, de l'impact de la réidentification, des résultats de l'analyse quantitative et du but de l'utilisation des données.
– Évaluation de l'adéquation. Estimation de l'adéquation de la désidentification en comparant les valeurs calculées issues de la valeur moyenne de référence avec l'analyse quantitative.

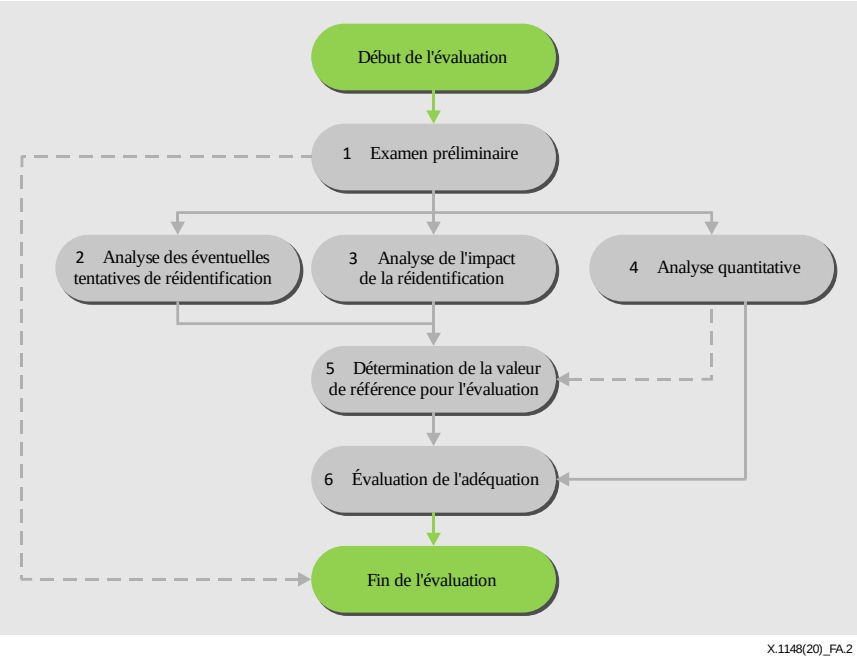

Figure A.2 – Procédure d'évaluation de l'adéquation

A.4 Mesures de désidentification supplémentaires

– Un responsable du contrôle des données mettra en place des mesures de désidentification supplémentaires, sur la base des informations obtenues auprès du groupe d'évaluation, si les résultats de l'évaluation sont inadéquats.

– Le groupe d'évaluation procédera à la réévaluation dès que le responsable du contrôle des données aura mis en œuvre les mesures de désidentification supplémentaires.

A.5 Utilisation des données

– Utilisation des données désidentifiées dans l'analyse des mégadonnées ou fourniture de ces données à un tiers si la désidentification est évaluée (ou réévaluée) adéquate.

– Il est en principe interdit de fournir ou de divulguer des données au public ou à des utilisateurs de données qui ne sont pas sous contrat en l'absence d'une bonne stratégie d'atténuation des risques pour les modèles de diffusion des données, compte tenu du risque élevé de réidentification.

– Destruction des données une fois que l'objectif d'utilisation est atteint ou si celles-ci ne sont plus nécessaires.

– Les étapes de gestion du suivi devraient être observées au cours du processus d'utilisation des données, pour une utilisation efficace sous forme de données désidentifiées.

Annexe B

Approches de désidentification des données non structurées

(Cette annexe fait partie intégrante de la présente Recommandation.)

À l'inverse de la désidentification des données structurées, la désidentification des données non structurées consiste à appliquer des méthodes de désidentification aux données brutes, au lieu et place des champs de données structurées. La procédure consiste par exemple, sur la photo ci-dessous, à supprimer les visages ou à les remplacer par d'autres comme le montre la Figure B.1.

Figure B.1 – Un exemple de désidentification faciale

On recense quatre types de données non structurées:

1) les données texte non structurées: données web, rapport, blog, news, etc.;
2) les données vidéo non structurées: toutes les données vidéo sont non structurées et certaines informations d'étiquette fournissent des données régularisées;
3) les données audio non structurées: toutes les données audio sont non structurées et certaines informations d'étiquette ou audio reconnues sont converties en données texte;
4) les données de journaux non structurées: les données de journaux générées par les machines sont non structurées, mais comportent souvent un schéma et peuvent être converties en une forme structurée.

Tout système de traitement de désidentification devrait, de manière à pouvoir représenter les informations syntaxiques pour les données non structurées (texte, voix, image, vidéo), comporter les trois unités suivantes:

1) une unité de détection des informations multimédias, pour détecter les méta-informations texte depuis les données multimédias saisies:
   - incluant un détecteur de parole, qui convertit l'entrée vocale en texte pour suivre un objet ou une activité inclus dans l'entrée vocale;
   - incluant un détecteur de reconnaissance optique de caractères, qui extrait les caractères depuis une entrée d'image;
   - incluant un détecteur visuel, qui extrait un objet ou une activité – inclus dans une entrée d'image ou d'image animée – depuis l'entrée d'image ou d'image animée;
   - incluant un détecteur visuel de phrase, qui extrait une phrase de texte depuis une entrée d'image ou d'image animée;

2) une unité de mise en forme basée sur la connaissance, qui divise les méta-informations texte et les informations de contexte en éléments syntaxiques dans une configuration extrinsèque et en éléments sémantiques dans une configuration intrinsèque:
   - les informations syntaxiques comprennent des informations sur la source générant les données multimédias, des informations de données multimédias générées par la source et des informations de détection d'objet extraites d'une région d'importance;
   - les informations sémantiques comprennent des informations sur l'événement inclues dans la région d'importance pour la configuration des données multimédias et des informations de contexte;
3) une unité de désidentification, qui supprime les éléments PII identifiables de la base de connaissances et des méta-informations texte.

Le processus de désidentification des données non structurées devrait définir les exigences et niveaux de sécurité en la matière comme suit:

- Cible de la désidentification: identifier l'objet cible qui devrait être protégé pour l'application ou les services en ligne?
- Modalités de la désidentification: identifier la méthode qui devrait être utilisée pour la désidentification. Quel est le niveau de désidentification (par exemple, boîte noire, pixélisation, floutage)?
- Désidentification et réidentification: il convient de déterminer s'il est nécessaire de récupérer ou de réidentifier les données. Si une politique requiert une photo originale pour une enquête criminelle, la photo désidentifiée peut-elle être récupérée?

Appendice I

Exemples de techniques de désidentification types

(Cet appendice ne fait pas partie intégrante de la présente Recommandation.)

Cet appendice contient des exemples et des descriptions de techniques de désidentification types.

I.1 Outils statistiques des techniques de désidentification

– Échantillonnage: processus consistant à diffuser un échantillon d'un ensemble de données et non l'ensemble de données dans son intégralité. En cas de diffusion d'un sous-échantillon, la probabilité de réidentification est plus faible.

– Agrégation: ensemble de fonction statistiques permettant d'obtenir la valeur représentée d'un ensemble de données entier.

I.2 Outils cryptographiques des techniques de désidentification

– Chiffrement déterministe [b-ISO/CEI 11770]: schéma de chiffrement qui produit toujours le même texte chiffré pour un texte en clair et une clé donnés sur des exécutions distinctes de l'algorithme de chiffrement.

– Chiffrement conservant l'ordre des fréquences [b-AGRAWAL]: schéma de chiffrement dans lequel l'ordre numérique des textes en clair est préservé.

– Chiffrement homomorphe [b-ISO/CEI 18033-6]: schéma de chiffrement qui permet d'effectuer des calculs sur du texte chiffré, générant ainsi un résultat chiffré qui correspond au résultat des opérations à effectuer sur le texte en clair, une fois déchiffré.

– Chiffrement préservant le format [b-NIST 800-38G]: un schéma de chiffrement dans lequel le texte chiffré est au même format que le texte en clair.

– Partage secret homomorphe [b-ISO/CEI 18033-6]: type d'algorithme de partage de secret dans lequel le secret est chiffré à l'aide d'un chiffrement homomorphe.

I.3 Techniques de suppression

– Masquage: processus qui consiste à remplacer un champ par une valeur ou à le supprimer. Il peut s'agir par exemple de remplacer un numéro de téléphone par des astérisques ou un pseudonyme généré aléatoirement.

– Suppression locale: processus visant à supprimer ou à enlever des valeurs spécifiques d'attributs des enregistrements sélectionnés. Le fait d'enlever les données renforce la protection de la confidentialité, mais peut diminuer l'utilité de l'ensemble de données.

– Suppression d'enregistrement: processus qui implique la suppression d'un ou de plusieurs enregistrements d'un ensemble de données.

I.4 Techniques de pseudonymisation

Processus qui supprime l'association avec un sujet de données et ajoute une association entre un ensemble particulier de caractéristiques concernant le sujet de données et un ou plusieurs pseudonymes. En règle générale, la pseudonymisation consiste à remplacer des identificateurs directs par un pseudonyme, tel qu'une valeur générée de manière aléatoire. Les identificateurs directs peuvent être des noms, des adresses e-mail et des numéros émis par le gouvernement. Tous les identificateurs directs et attributs d'identification supplémentaires ou restants potentiel sont remplacés par un pseudonyme.

I.5 Techniques de généralisation

– Lissage: processus qui consiste à remplacer une valeur numérique par une autre valeur qui est approximativement égale, mais qui a une représentation plus courte, plus simple ou plus explicite.

– Codage haut et bas: processus par lequel les attributs dont les valeurs sont supérieures à une limite supérieure (ou limite inférieure) sont définis comme un seuil sur la plus grande (ou la plus petite) valeur possible.

I.6 Techniques de randomisation

– Ajout de bruit: processus dans lequel une valeur aléatoire qui ne peut être prédite est ajoutée à un attribut sélectionné d'un ensemble de données.

– Permutation: processus d'échange des valeurs d'un attribut sélectionné entre les enregistrements d'un ensemble de données sans modification.

– Micro-agrégation: processus par lequel toutes les valeurs d'attributs continus sont remplacées par leurs moyennes calculées d'une certaine manière algorithmique.

I.7 Données synthétiques

Les données synthétiques sont une approche qui génère artificiellement des microdonnées pour représenter un modèle de données statistiques prédéfini. Par définition, les ensembles de données synthétiques ne contiennent pas de données collectées auprès des sujets de données existants, mais ils semblent réels pour l'usage auquel ils sont destinés.

Appendice II

Modèles de processus de désidentification

(Cet appendice ne fait pas partie intégrante de la présente Recommandation.)

Cet appendice contient des exemples et des descriptions de modèles de processus de désidentification.

II.1 Modèle de désidentification centré sur les données

Sachant que les techniques de désidentification modifient les données d'origine pour empêcher la divulgation des informations d'identification personnelle, privilégier l'utilité ou la confidentialité est clairement un dilemme qui se pose. L'enjeu consiste à protéger la confidentialité avec une perte de précision minimale: idéalement, les utilisateurs de données devraient exécuter leurs analyses sur les données désidentifiées sans perdre en précision par rapport aux résultats de ces analyses lorsqu'elles sont exécutées sur les données d'origine.

Dans la pratique, la désidentification parfaite est difficile à obtenir sans compromettre l'utilité de l'ensemble de données. Avec les mégadonnées, cela est encore plus compliqué en raison de la quantité et de la variété des données. D'un côté, un faible niveau de désidentification (par exemple, lorsque la désidentification ne supprime que les identifiants directs) n'est souvent pas suffisant pour garantir la non-identifiabilité. De l'autre côté, un niveau trop élevé de désidentification peut empêcher de lier des données sur le même individu (ou sur des individus similaires) provenant de différentes sources et, par conséquent, contrecarrer bon nombre des avantages potentiels des mégadonnées.

Ce paragraphe présente deux modèles de désidentification axée sur les données, visant à répondre au dilemme existant entre utilité et confidentialité. Des mesures d'utilité spécifiques et génériques relatives à l'utilisation des données peuvent être utilisées pour déterminer comment mesurer l'utilité d'un ensemble de données publié une fois désidentifiées.

II.1.1 Modèle de désidentification privilégiant l'utilité

Dans le cas des mégadonnées, les informations sur un individu sont souvent recueillies auprès de plusieurs sources indépendantes. Par conséquent, la possibilité de lier des enregistrements qui appartiennent à la même personne (ou à une personne similaire/du même type) est essentielle dans la création de mégadonnées.

Le modèle de désidentification privilégiant l'utilité applique une technique de désidentification avec un choix de paramètres heuristiques et des propriétés adéquates en matière de préservation de l'utilité sur le jeu de microdonnées, après quoi le risque de divulgation est mesuré. Cette approche prend du temps et ne comporte pas de garanties officielles en matière de confidentialité. Le risque de réidentification peut par exemple être estimé de manière empirique en tentant de lier les enregistrements entre les ensembles de données d'origine et de données désidentifiées. Si le risque existant est jugé trop élevé, la technique de désidentification sera exécutée une nouvelle fois avec des paramètres de confidentialité plus stricts et vraisemblablement de plus grandes concessions au niveau de l'utilité, en modifiant les paramètres de manière itérative jusqu'à ce que le risque de divulgation empirique soit suffisamment faible, comme c'est le cas dans les statistiques officielles.

La possibilité de liaison est certes souhaitable du point de vue de l'utilité, mais elle n'en demeure pas moins une menace pour la vie privée: la précision des liens devrait être beaucoup moins importante dans les ensembles de données désidentifiées que dans les ensembles de données d'origine. Le niveau de liaison compatible avec une technique de désidentification ou un modèle de désidentification axé sur la confidentialité détermine si et comment un analyste peut lier indépendamment des données désidentifiées (selon cette technique/ce modèle) qui correspondent au même individu.

II.1.2 Modèle de désidentification privilégiant la confidentialité

Le modèle de désidentification privilégiant la confidentialité comporte un paramètre qui garantit une limite supérieure pour le risque de divulgation de réidentification et peut-être aussi pour le risque de divulgation d'attribut. Ce modèle est appliqué en utilisant une technique de désidentification propre au modèle avec des paramètres issus des paramètres du modèle. Les modèles les plus courants qui reposent sur la confidentialité incluent la k-anonymité et ses extensions, ainsi que la confidentialité différentielle ε, qui conduit souvent à un faible niveau d'utilité/de liaison des données.

Selon ce modèle, si l'utilité des données désidentifiées qui en résultent est trop faible, le modèle de confidentialité utilisé devrait être appliqué avec une autre technique de désidentification moins dommageable pour l'utilité, ou un paramètre de confidentialité moins strict devrait être choisi, voire un autre modèle de désidentification privilégiant la confidentialité devrait être utilisé.

II.2 Modèle de désidentification centré sur les rôles

Ce paragraphe présente trois types de modèle en indiquant les rôles et les responsabilités de chacun dans le processus de désidentification. Le modèle centré sur les rôles consiste, dans une large masure, à répondre aux questions suivantes: "Qui?", "Quoi?" et "Où et comment?":

– Qui a accès aux données?

– Quelles analyses peuvent-être ou non réalisées?

– Où l'accès/l'examen des données est-il réalisé et comment l'accès est-il obtenu?

II.2.1 Désidentification centralisée

Le processus de contrôle de la divulgation statistique se concentre sur la désidentification centralisée, laquelle est effectuée par un responsable du contrôle des données qui a accès à l'ensemble de données d'origine. Cette approche centralisée présente certains avantages et inconvénients, comme le montre le Tableau II.1.

Tableau II.1 – Caractéristiques de la désidentification centralisée

	Détails
Avantages	• Les individus n'ont pas besoin de désidentifier les enregistrements de données qu'ils fournissent. On peut s'attendre à ce que le responsable du contrôle des données, qui dispose de plus grandes ressources de calcul et probablement d'une plus grande expertise en matière de désidentification, désidentifie adéquatement tout l'ensemble de données. • Le responsable du contrôle des données détient une vue globale de l'ensemble de données d'origine et est donc le mieux placé pour optimiser le compromis entre l'utilité des données et le risque de divulgation existant.
Inconvénients	• Le responsable du contrôle des données doit bénéficier de la confiance de toutes les parties qui fournissent les données d'origine (car il a accès à toutes les données d'origine). Bien que cela ne soit pas un problème pour les statistiques officielles, où le responsable du contrôle des données est un institut national de statistique, cela peut être un obstacle majeur dans le cas bien connu des mégadonnées, par exemple lorsque le responsable du contrôle des données qui assemble plusieurs sources de données s'avère être une entreprise privée (par exemple, un courtier en données). • Dans le cas des mégadonnées, la désidentification peut être une charge de calcul trop lourde pour un seul responsable du contrôle. • De nombreux responsables du contrôle des données sont impliqués dans un scénario unique de traitement des mégadonnées, rendant ainsi l'approche centralisée ingérable.

Les approches de désidentification locale et de désidentification collaborative complètent les avantages et les inconvénients ci-dessus.

II.2.2 Désidentification locale

La désidentification locale est une approche alternative de limitation de la divulgation qui peut être appliquée dans des scénarios (y compris les mégadonnées) où les individus (sujets de données) ne placent pas leur confiance (ou ne la placent que partiellement) dans le responsable du contrôle des données qui assemble les données. Chaque sujet désidentifie ses propres données avant de les remettre au responsable du contrôle.

Dans un souci de protection de la confidentialité, les données collectées par une source spécifiée devraient être désidentifiées à la source avant d'être mises à disposition. Cependant, la désidentification, si elle est effectuée de manière indépendante par chaque source, entraîne plus de pertes d'informations que la désidentification centralisée, car les sujets désidentifient leurs données sans voir les données des autres sujets. Dans ces conditions, les sujets ne profitent pas d'une vue globale de l'ensemble de données, ce qui complique les choses pour trouver un bon compromis entre la limitation du risque de divulgation atteinte et la perte d'informations encourue.

II.2.3 Désidentification collaborative

Le processus de désidentification collaborative combine la faible perte d'utilité de la désidentification centralisée et le haut niveau de confidentialité de la désidentification locale. L'un des problèmes avec la désidentification centralisée est que, si un sujet de données ne fait pas confiance au responsable du contrôle des données pour utiliser et/ou désidentifier correctement ses données, il peut décider de fournir de fausses données (provoquant ainsi un biais de réponse) ou pas de données du tout (provoquant ainsi un biais de non-réponse). Les sujets pourraient donc collaborer pour définir le risque de divulgation associé à leurs données et appliquer localement sur cette base le niveau de protection adéquat, de manière répartie et collaborative, selon deux grands principes:

– Il n'en résulte pas plus de perte d'informations qu'avec l'ensemble de données qui serait obtenu dans le cadre de l'approche centralisée pour le même niveau de confidentialité. Cette méthode surpasse l'approche locale en ce qu'elle génère moins de pertes d'informations.
– Les sujets de données et le responsable du contrôle des données n'obtiennent pas plus de renseignements sur les attributs confidentiels de tout autre sujet de données spécifique que ceux contenus dans l'ensemble finalisé de données désidentifiées. Cette méthode surpasse l'approche centralisée en offrant également la confidentialité vis-à-vis du collecteur de données.

De plus, l'approche collaborative pourrait mener à des protocoles qui fonctionnent sans heurts et sans mécanismes d'application externes. En ce qui concerne la désidentification des microdonnées, la protection de la confidentialité par un sujet affecte la protection de la confidentialité que les autres obtiennent. Pour améliorer la co-utilité dans l'approche collaborative, une transformation multipartite sécurisée est requise pour les protocoles électroniques qui permettent à deux ou plusieurs parties d'effectuer une transformation impliquant deux de leurs ensembles de données de telle manière qu'aucune partie n'ait besoin de remettre explicitement un ensemble de données à une autre. Parce que la transformation multipartite sécurisée permet de transformer les requêtes sans qu'il soit nécessaire de centraliser l'ensemble des stockages de données, elle réduit les risques de violation de données et permet des calculs entre les parties qui ne se font pas entièrement confiance. Les calculs multipartites peuvent offrir à la fois une meilleure confidentialité et une meilleure utilité dans certains contextes.

Bibliographie

[b-ISO/CEI 11770] ISO/CEI 11770 (toutes les parties), *Technologies de l'information – Techniques de sécurité – Gestion de clés*.

[b-ISO/CEI 18033-6] ISO/CEI 18033-6, *Techniques de sécurité IT – Algorithmes de chiffrement – Partie 6: Chiffrement homomorphe*.

[b-ISO/CEI 20889] ISO/CEI 20889 (2018), *Terminologie et classification des techniques de dé-identification de données pour la protection de la vie privée*.

[b-ISO/CEI 27001] ISO/CEI 27001 (2018), *Technologies de l'information – Techniques de sécurité – Systèmes de gestion de la sécurité de l'information*.

[b-ISO/CEI 29100] ISO/CEI 29100 (2011), *Technologies de l'information – Techniques de sécurité – Cadre privé*.

[b-NIST 800-38G] NIST Special Publication 800-38G (2016), *Recommendation for Block Cipher Modes of Operation: Methods for Format-Preserving Encryption*.

[b-NISTIR 8053] NISTIR 8053 (2015), *De-Identification of Personal Information*.

[b-AGRAWAL] Agrawal, R., Kiernan, J., Srikant, R., et Xu, Y. (2004), *Order preserving encryption for numeric data, SIGMOD '04 Proceedings of the 2004 ACM SIGMOD international conference on Management of data, Paris, France, juin, p. 563-574*.

[b-KOREA] Korean Ministry of the Interior, *Guidelines on De-identification Measures, juin 2016*.
<http://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000821178&fileSn=2&nttId=7187&toolVer=&toolCntKey_>
(consulté pour la dernière fois le 26 juillet 2019)
<https://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000827161&fileSn=0> (English, consulté pour la dernière fois le 12 décembre 2020).

[b-UKAN] UK Anonymization Network, *The anonymisation decision-making framework, 2016* <https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>.

SÉRIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série D	Principes de tarification et de comptabilité et questions de politique générale et d'économie relatives aux télécommunications internationales/TIC
Série E	Exploitation générale du réseau, service téléphonique, exploitation des services et facteurs humains
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission, systèmes et réseaux numériques
Série H	Systèmes audiovisuels et multimédias
Série I	Réseau numérique à intégration de services
Série J	Réseaux câblés et transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Environnement et TIC, changement climatique, déchets d'équipements électriques et électroniques, efficacité énergétique; construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	Gestion des télécommunications y compris le RGT et maintenance des réseaux
Série N	Maintenance: circuits internationaux de transmission radiophonique et télévisuelle
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux
Série Q	Commutation et signalisation et mesures et tests associés
Série R	Transmission télégraphique
Série S	Équipements terminaux de télégraphie
Série T	Terminaux des services télématiques
Série U	Commutation télégraphique
Série V	Communications de données sur le réseau téléphonique
Série X	**Réseaux de données, communication entre systèmes ouverts et sécurité**
Série Y	Infrastructure mondiale de l'information, protocole Internet, réseaux de prochaine génération, Internet des objets et villes intelligentes
Série Z	Langages et aspects généraux logiciels des systèmes de télécommunication

Imprimé en Suisse
Genève, 2021