

個人情報保護に関する法律とビッグデータのための非識別化技術

慎 祥揆^{†1}

Act on the Protection of Personal Information and De-identification Processing Technology for Big Data

by

Sanggyu Shin

(Received on Oct. 23, 2023 & accepted on Dec. 22, 2023)

Abstract

This paper introduces the law definitions of de-identification, re-identification, anonymization, and pseudonymization based on the amendment act of the Act on the Protection of Personal Information to be enforced from April 1, 2022. Furthermore, this paper the current international standardization trends in the de-identification field, including the standardized frameworks ISO/IEC 27559 and ITU-T X.1148 (Framework of de-identification process for telecommunication service providers), ISO/IEC 20889 and ISO/IEC 29100. Personal data de-identified through anonymization or pseudonymization must be correctly de-identified before it is used as part of publicly available big data sets. Many companies are implementing Big Data projects that require a sound legal understanding and development according to international standards to remain compliant with the ever-increasing regulatory risk requirements. Dealing with Big Data and sensitive personal data requires knowledge and technical competence to maintain the appropriateness of that data.

Keywords: Act on the Protection of Personal Information, de-identification, anonymization, pseudonymization, ISO/IEC 20889:2018, ISO/IEC 27559:2022, ITU-TSG17 X.1148

1 はじめに

最近の ChatGPT や生成 AI のブームは、巨大なデータの収集と分析を基盤にしている。このように、最近 5G 時代を迎え、人工知能 (AI)、クラウド、モノのインターネット (IoT) などの技術の必須要素でありながら、第 4 次産業革命において重要な位置づけにある「データ」産業による経済の活性化が期待されている。しかしながら、企業と機関がデータを収集し活用する過程で、個人のプライバシーを侵害する可能性もまた増加している。このように、収集されたビッグデータはデータ科学の分野だけでなく人文科学の分野においても加速度的に使用され、技術発展に貢献している反

面、それに伴い収集されたデータの個人情報の問題や、著作権の問題など、様々な技術的、法律的な問題を生じてしまっている。これらのビッグデータは、一般にアクセス可能なデータを適切に普及および匿名化した場合にのみ使用することができるよう法律で定められている [1]~[3]。

法律の範囲内でその活用を促進させるためには、ビッグデータの分野に関わる開発のフレームワークの導入及びその標準化を促進し、これを管理するための、非識別化 (De-identification)、再識別化 (Re-identification)、匿名化 (Anonymization)、仮名化 (Pseudonymization) 技術の理解と新たな国際的な基準の提案が必要となる。

Sweeney の研究によると、アメリカの人口の 87% をその性別、郵便番号、生年月日をもとに、一意に識別できることが立証された [4]。他には、2006 年に EFF

^{†1} 東海大学 情報理工学部 コンピュータ応用工学科
Dept. of Applied Computer Engineering, School of Information Science and Technology, Tokai University

(Electronic Frontier Foundation) 研究チームによって匿名の AOL 検索クエリを収集・分析を行われ、非識別化された Netflix の利用データから個人が特定できることが確認されている。いずれの組織も研究の名目でデータを公開していた [5]。これは、匿名化によるデータ加工は実用的ではあるが、更なる研究と相当な注意が必要であることを示している。

匿名化で代表される非識別化処理は、識別可能なデータを修正して特定の個人を識別できないように処理するプロセスであり、非識別分野では非識別処理を行う際にデータの有用性を損なうことなく非識別水準を高めることに対する技術的方法論に関する研究が活発に進められている。また、非識別処理環境と処理目的に沿った適切な水準の非識別化の可否を判断するための、データの非識別化を保障することに関する要求事項を開発する必要性が生じている。

本稿では、2022 年 4 月 1 日から施行された改正個人情報保護法による非識別情報の法律的定义と、現在非識別分野で標準化されたフレームワークとデータ非識別評価要求事項の提示の必要性のために開発された国際標準及び標準化推進動向を紹介する。まず、第 2 章では個人情報の保護に関する法律について、第 3 章で、非識別化処理の概念と法律に基づいた各情報の定義と違いを、非識別化プロセスに関連する国際標準化の動きを第 4 章で述べ、最後にまとめる。

2 個人情報の保護に関する法律

最近、ChatGPT の個人情報収集問題やインターネット広告のプライバシー保護に関連する問題など、個人情報収や知的財産に関するトラブルがよく話題になっている。インターネット上のプライバシー保護については、EU やアメリカが先行して取り組んでいたが、日本でも 2022 年 4 月 1 日から、2020 年公布の個人情報の保護に関する法律（改正個人情報保護法）が施行された [1]。個人情報保護法の改正の方向は、個人の権利利益の拡大するための規制強化の方向であるが、個人データの利活用の促進のための仮名加工情報の創設、情報漏洩時の報告の義務化など、情報を扱う企業に直接影響を与える新たな規律の他、個人情報の範囲を広げるような試みも見られる。

個人データを加工して利用する情報としては、これまでも統計情報と匿名加工情報があったが、これに新たな加工情報が加わっている。他にも様々な新設があるが、その他、新たに定められたものは様々あるが、企業のデジタルマーケティングにおいて実務上重要な内容としては個人関連情報があげられる。まず、個人情報や個人関連情報、匿名加工情報、仮名加工情報など

を法律に準じながら、その違いと定義を述べる。

2.1 法律による各法律用語の定義

個人情報とは、生存する個人に関する情報のことで、氏名、生年月日などのデータによって特定の個人を識別できる情報、または個人識別符号¹を含む情報のことを指している [1][2][6]。

個人関連情報とは、個人情報保護法上の定義によると「生存する個人に関する情報であって、個人情報、仮名加工情報及び匿名加工情報のいずれにも該当しないものをいう」というような消極的な定義がされている。一般的には、氏名など個人情報と結びついていないネット閲覧履歴、位置情報、Cookie 情報等が該当する。Cookie によって取得される情報は、それ単体では個人情報ではないが、今回の改正法では個人関連情報と扱われるようになった [1][2]。

統計情報は、複数人の情報から共通要素に係る項目を抽出して同じ分類ごとに集計等して得られる情報であり、一般に、特定の個人との対応関係が排斥されているため、個人情報に該当しないものである [7][8]。統計情報は個人情報保護法で特段の規制はないため、利用目的も第三者への提供も自由である。従って、顧客や利用者から個人データの第三者提供の同意を取得すれば、そのままの個人データから一定の加工を施したものまで提供が可能になる。しかし、現実的にはあらためて第三者提供への同意を再度取得するのは難しい [9]。

2.2 個人情報保護法の改正必要性

EU では、インターネット上におけるプライバシーへの懸念から、2018 年に GDPR (General Data Protection Regulation: 一般データ保護規則) が施行された [10][11]。この施行によって、従来の EU データ保護指令よりもさらに厳格化されることになり、IP アドレスや Cookie などのオンライン識別子もパーソナルデータに含まれるように規定された。また、企業側には、ユーザーから明示的な同意を取得することが求められることとなった。

その GDPR が 1 つのきっかけになり、世界各国でプライバシーに関する議論がさらに活発化し、2020 年にはアメリカのカリフォルニア州では、CCPA (California Consumer Privacy Act: カリフォルニア州消費者プライバシー法) という独自の法律が施行された

¹ 身体の一部の特徴をコンピュータで利用するために変換したもの（たとえば、DNA や顔など）と、公的なサービスの利用のためにサービス利用者に割り振られる番号（たとえば、免許証の番号やマイナンバーなど）

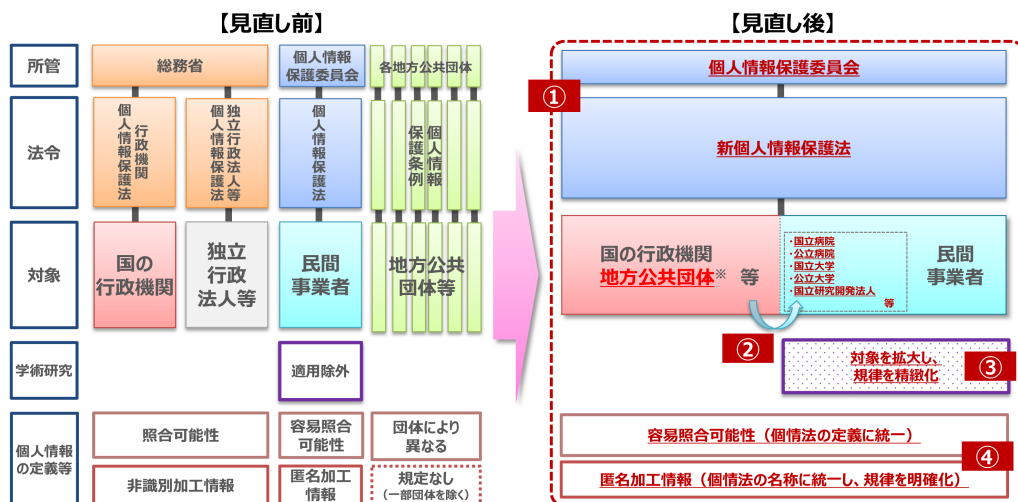


図1 個人情報保護制度見直しの全体像 [13]

[12]. その他、韓国²、ブラジルやタイなど世界各国においても同様の法律が相次いで検討・施行されている。

産業界の動きでも、Apple 社や Google 社による 3rd Party Cookie の利用制限など、プラットフォームによる技術的な規制も同時に進んでいる。

日本においても、2005 年に個人情報保護法（正式には「個人情報の保護に関する法律」）が施行されたが、最近のインターネットの発展とそれに伴う企業のデータ活用がますます広がっており、個人情報の扱う環境にも大きく変化している。これらの環境変化に応じるため、法律も時代に合わせてその内容を大きく変化させなければならなくなった。

それに従い、個人情報保護法では、その改正法附則において 3 年ごとの見直しが求められている。個人情報保護に関する国際的な動向や、情報通信技術の進展、それに伴う新たな産業の創出などを勘案して必要性が認められた場合に見直しが行われる。従い、2017 年には第 1 回の個人情報保護法改正が、2020 年 6 月に第 2 回の改正個人情報保護法が公布されて 2022 年 4 月から施行されている。

2020 年の改正内容が全面施行され、2021 年の改正が一部施行となり、2022 年 4 月施行の個人情報保護法は、以下の観点から内容が見直された。

- 個人の権利利益の保護
- 技術革新の成果による保護と活用の強化

- 国際的な制度調和と連携
- 越境データの流通増大に伴う新たなリスクへの対応
- AI・ビッグデータ時代への対応

また、2023 年 4 月施行の個人情報保護法の改正では、国の行政機関、独立行政法人等、民間事業者及び地方公共団体等においてこれまで別々の法律や条例により運用されてきた個人情報の取扱いが同一の法（個人情報保護法）によって運用されることとなり、国の機関である個人情報保護委員会が全体を所管することとなった。

法の改正について、個人情報保護委員会のホームページからの説明を参考として追加する（図 1 とその説明）[13]。

図 1 の①：個人情報保護法、行政機関個人情報保護法、独立行政法人等個人情報保護法の 3 本の法律を 1 本の法律に統合するとともに、地方公共団体の個人情報保護制度についても統合後の法律において全国的な共通ルールを規定し、全体の所管を個人情報保護委員会に一元化。

図 1 の②：医療分野・学術分野の規制を統一するため、国公立の病院、大学等には原則として民間の病院、大学等と同等の規律を適用。

図 1 の③：学術研究分野を含めた GDPR の十分性認定への対応を目指し、学術研究に係る適用除外規定について、一律の適用除外ではなく、義務ごとの例外規定として精緻化。

図 1 の④：個人情報の定義等を国・民間・地方で統一するとともに、行政機関等での匿名加工情報の取扱いに関する規律を明確化。

²2023 年 9 月 15 日に施行された韓国の改正個人情報保護法には、情報主体である国民の権利は実質的に保障しながら、オンライン・オフラインに二重化されている個人情報処理基準をデジタル環境に合わせて一元化するなど、これまで各界で議論されてきた多様な内容が含まれている。

3 非識別化処理

改正個人情報保護法にて個人データの利活用の促進に向けたものとして新たに仮名加工情報が定義付けられたが、それぞれどのような場合に使えるのか、特に匿名加工情報と仮名加工情報の違いを理解する必要がある。

3.1 非識別化の概念

匿名加工情報と仮名加工情報とも、非識別化と関連するものである。

非識別加工情報とは、個人情報に含まれる記述等の一部を削除すること又は個人識別符号の全部を削除すること等により、「特定の個人を識別することができないように個人情報を加工して得られる個人に関する情報であって、当該個人情報を復元することができないようにしたもの」と定義されている [14]。

非識別処理とは、「運用データから氏名や電話番号、生年月日などの識別情報を取り除く処理であって、識別情報を構成するデータの削除、数値の丸め込み、不可逆符号への変換などを含むもの」をいう [15]。

3.2 匿名処理

2015 年度の個人情報保護法の改正に伴って創設された、匿名加工情報の定義は以下の通りである。

次の各号に掲げる個人情報の区分に応じて当該各号に定める措置を講じて特定の個人を識別することができないように個人情報を加工して得られる個人に関する情報であって、当該個人情報を復元することができないようにしたものをいう [1][2]。

- 一 第一項第一号に該当する個人情報 当該個人情報に含まれる記述等の一部を削除すること（当該一部の記述等を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。
- 二 第一項第二号に該当する個人情報 当該個人情報に含まれる個人識別符号の全部を削除すること（当該個人識別符号を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

要約すると、匿名加工情報とは「特定の個人を識別できないように個人情報を加工して得られる個人に関する情報であって、当該個人情報を復元できないようにしたもの」である。

従って、匿名処理（Anonymisation）は、合理的に予想されるあらゆる手段（費用、時間、技術発展）を考慮しても、個人を識別できないように処理することを意

味する。匿名情報は個人情報ではないため、個人情報保護法の適用対象ではなく、自由に活用できる [16][17]。

しかし、個人情報データを匿名化しても、英 ICO が独自の「匿名化規範ガイド」で概説しているように、自分に関して他にどのようなデータが出回っているのかわからないし、それが誰かしらによって匿名データセットとどのようにマッピングされるかも、決してわからないという問題が残る [18]。

3.3 仮名処理

2020 年度の個人情報保護法の改正に伴って提言された、仮名加工情報の定義は以下の通りである。

次の各号に掲げる個人情報の区分に応じて当該各号に定める措置を講じて他の情報と照合しない限り特定の個人を識別することができないように個人情報を加工して得られる個人に関する情報をいう [1][2]。

- 一 第一項一号に該当する個人情報 当該個人情報に含まれる記述等の一部を削除すること（当該一部の記述等を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。
- 二 第一項二号に該当する個人情報 当該個人情報に含まれる個人識別符号の全部を削除すること（当該個人識別符号を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

つまり、仮名加工情報は「他の情報と照合しない限り特定の個人を識別できないように個人情報を加工して得られる個人に関する情報」である [20]。

従来、匿名加工情報が存在していたにもかかわらず、2020 年 3 月に閣議決定された個人情報保護法改正案によって新たに仮名加工情報が新設された理由として、個人情報保護法制度改正大綱では以下の理由が挙げられている [19]。

- 事業者の中で、「仮名化」と呼ばれる加工を施した上で個人情報を活用する事例がみられたため。
- EU において、個人情報としての取り扱いを前提としつつ、若干緩やかな取り扱いを認める「仮名化」が規定され、国際的にもその活用が進みつつあるため。即ち、一定の安全性を確保しつつ、さらなる利活用を促進する観点から、他の情報と照合しなければ特定の個人を識別することができないように加工された情報である、仮名加工情報が導入されたことである。

従って、仮名処理（Pseudonymisation）は、個人情報を追加情報なしには特定個人を識別できないように処理することである。すなわち、データセット内にある識別子を暗号化・ハッシュなどの手法を用いて他の

表1 各情報間の違い

	個人情報	統計情報	匿名加工情報	仮名加工情報
第三者提供	○ 同意必要	○ 同意不要	○ 同意不要	×
再識別	-	不可能	禁止	禁止
利用目的の変更	-	○ 同意不要	○ 同意不要	○ 同意不要
開示・利用停止等の請求対応	必要	不要	不要	不要
漏えい等の報告	必要	不要	不要	不要

属性に置き換える処理である。仮名情報は仮名化された情報であり、個人情報であるため個人情報保護法の適用を受け、制限された目的の下で技術的・管理的に安全措置を行なった上で活用できる。

追加情報とは仮名化過程で生成される情報で、このような追加情報は原本情報とは別に分離して保管しなければならない、これに対する技術的・管理的な安全措置が必要である。

3.4 各情報の違い

仮名加工情報も匿名加工情報も、両方とも個人情報や個人識別符号を加工して作成することは共通している。端的な違いは、第三者提供できるか否かである。本人に同意を得ることなく第三者に提供したい場合は、匿名加工情報か統計情報を、仮名加工情報はいかなる場合も第三者提供は禁止である。

本人に同意を得ることなく利用目的を変更したい場合は、仮名加工情報、匿名加工情報、統計情報が利用できるが、仮名加工情報は同一社内、委託先、共同利用の場合のみで、元の個人にアクセスする利用は禁止である。各情報間の違いを表1に示す[21]。

3.5 再識別問題

再識別化に関する懸念は、イギリスの NHS デジタル（保健・社会医療情報センター）が2021年、「計画・研究のための一般診療データ」という構想のもと、イギリス国民の健康データの大規模収集を推し進めたことで表面化した。この計画では、イングランド全住民の GP 医療記録を中央の研究機関に伝送・集約することになっており、国民にはオプトアウト³するのに短い猶予しか与えられていなかった[22]。

日本国内では、JR 東日本のプライバシー関連の話題で、2013年のユーザーの利用履歴を第三者である日立製作所へ提供していた件がある[23]。

再識別問題を打開する理論的な方法のひとつは、個人の身元を明らかにするようなデータをひたすら取り除き続ける手法がある。しかし、データを削除する

たびに、データセットの価値が下がってしまう。

匿名性と有用性を両立させる方法のひとつとして、ディファレンシャルプライバシーがある。パラメータを微妙に変化させることでデータに統計的なノイズを加える手法である。たとえば、ある人の年齢や郵便番号をわずかにずらして、その相関を取りにくくする手法などである。しかし、研究者やハッカダちはデータベースに繰り返しアクセスする手法などで、そのノイズを除外することができる。対策として、ディファレンシャルプライバシーのもう一つの要素として、データにアクセスできる回数に制限が設ける方策が採られている。

法律的には非識別化に関する更なる法整備を求める方向もあるが、管理面の対策であり、技術的な対策にはならない。

4 国際標準化の動き

4.1 非識別技術標準

ISO/IEC 20889:2018 標準は ISO/IEC JTC 1 で開発した非識別分野の最初の国際標準[24]であり、ISO/IEC 29100 の個人情報保護原則[25]に従って標準化された用語を用いて既存の非識別化技術を分類し、再識別化のリスクを低減するための基盤技術及び各技術の適用可能性等の特性について説明する標準である。

標準の主な内容としては、再識別攻撃の種類、非識別化技術の種類、一般プライバシー測定モデル、非識別化技術適用のための一般的な原則、非識別化ツール・技術・モデルの特性紹介がある。一方、ISO/IEC 20889で提示された技術はテーブル形式に変換できるデータセットに適用可能な内容であり、自由形式のテキストやイメージ、オーディオ、ビデオを含む複雑なデータセットには適用できない。標準の最大の特徴は、複数の技術と定量的な観点からのリスク等を考慮した上で、同一の攻撃リスクに相反する助けになる非識別技術があれば、有用性の観点から最適な技術を選択することを提示することである。

ISO/IEC 20889 は上記した通り、ISO/IEC 29100 の個人情報保護原則に基づき、個人情報保護措置の説明・設計に使用される個人情報保護強化データ非識別化技術の説明を提供している。

³ユーザーに許可を取ることなく、宣伝広告をメールなどの方法で送りつけること。また、宣伝広告の受け取りを、ユーザーが拒否する意思を示すこと。

表2 非識別フレームワークの構成

構成要素	主な内容
文脈 (Context)	攻撃者が活用できる外部情報を決定するために、データ受信者がデータを活用できるようになる環境及び状況を評価。これは、文脈（Context）を考慮したコントロール（受信者の IT セキュリティコントロール、書面契約に記述された義務、ポリシー及びガバナンス処置）を通じてリスクを管理できることを意味する。
データ (Data)	攻撃者が活用可能な追加情報が PII を表したり明らかにしたりすることに使用できる方法を決定するためにデータ自体を評価。リスクは、データ変換によってどのようなデータが活用可能となり、どのようなデータ形式が今後活用可能となるかを制限することによって管理できる。
識別可能性 (Identifiability)	識別可能性を評価する方法は、文脈リスク（攻撃確率）とデータリスク（攻撃がある場合の公開確率）によって決定。識別可能性が事前定義された許容誤差レベル未満になるようにするには、適切な許容誤差を定義する必要がある。
ガバナンス (Governance)	管理者が現在及び将来において上記事項が一貫的で、効率的に完了するようにするために文書化される手順及び慣行、非識別化されたデータの可用性が確保される前、確保される過程のうち、確保された後に必要な準備を意味する。

ISO/IEC 29100:2011 では、次のような個人情報セキュリティフレームワークを規定している。

- 一般的なプライバシー用語を指定
- 個人識別情報 (PII) を処理する行為者とその役割を定義
- プライバシーの考慮事項について説明。そして、
- 情報技術について既知のプライバシー原則への参照を規定

ISO/IEC 29100 は、PII 処理のために個人情報保護統制が必要な情報及び通信技術システムやサービスの指定、調達、設計、開発、テスト、維持管理、管理及び運営に関わる自然人及び組織に適用される。

4.2 非識別化フレームワーク

国際標準 ISO/IEC 27559:2022 は、国際標準化機構である ISO と IEC 合同技術委員会 (JTC 1) SC 27 WG5 から現在 DIS (Draft International Standard) バージョンまで進行された標準であり [26]、2018 年 11 月に制定された国際標準である ISO/IEC 20889[24] プライバシー強化データ非識別化用語および技術分類の後続標準として制定中である。

ISO/IEC 27559 標準は、非識別化されたデータのライフサイクルに関連するリスクを把握し、緩和するためのフレームワークを提供する。この文書では、プライバシーの向上に寄与する非識別化とはどういうことを考慮しておこなわなければならないかということに記載している。

ISO/IEC 27551:2021 が非連結化（匿名化など）ということはどういうことかという概念を整理しているのに対して [27]、この規格では、具体的にどういうことをすればよいかということを整理している。このために、この文書では、

- データアセスメント（攻撃者が得ることができると

思われる対象個人の性質、データの形式、個別属性の性質、データセットの性質などを評価）

- 攻撃モデリング
- 非識別化可能性評価
- 対策実施
- バナンスに関しての説明

を提供している [28]。表 2 に、27551 にて非識別フレームワークの構成を示す。図 2 は、データ非識別プロセスを示す。

ITU-TSG17 X.1148 非識別プロセスフレームワーク [29] は、韓国の金融保安院 (FSI: Financial Security Institute) と KISA (Korea Internet & Security Agency: 韓国インターネット振興院) が ITU-T SG17 (情報保護) で開発した「非識別処理フレームワーク標準」であり、2016 年 SG 17 定期会議で金融保安院と韓国電子通信研究院 (ETRI: Electronics and Telecommunications Research Institute)、韓国の順天郷大学校 (Soon Chun Hyang Univ.)、中国のチャイナモバイル (China Mobile) が共同エディターとして参加し、新たに進行可能課題として採択された新規課題である。2016 年 8 月に韓国インターネット振興院が非識別措置支援センターに正式指定されることになり、12 月から本標準の共同エディタとして参加し、現在は両機関で主導的に最終標準制定のための手続きを進めている [30]。

本標準は非識別処理手続きに対する国際標準で、全産業分野に関係なく適用が可能である。ただし、非識別処理詳細技術は ISO 20889 と範囲が重複するため、別途取り扱っていない。

本非識別処理手順標準の主な内容は、

- データの流れ（ライフサイクル）による非識別処理が必要な地点定義（図 3）
- 韓国のガイドラインで提示した 4 段階の非識別処理

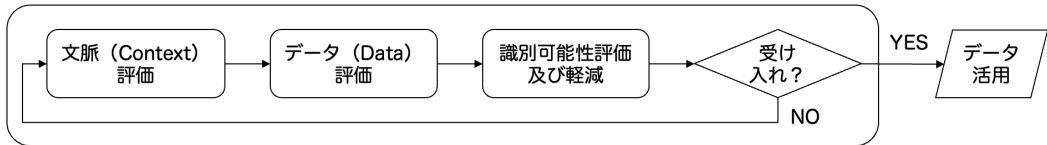


図2 データ非識別プロセス

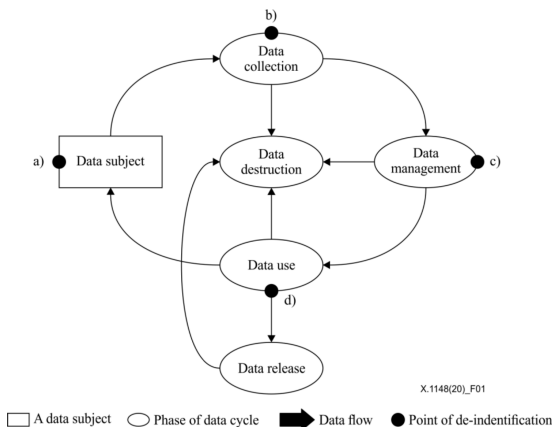


図3 データライフサイクルモデルにおける非識別プロセス [29]

手順 (図4)

- 非識別情報提供モデル別特性と非識別処理水準によるデータ形態 (図5) である。

図3はデータの流れ「生命周期」にともなう非識別処理が必要な3つの地点を定義している。具体的には、ライフサイクルモデルからデータを収集する地点 ((a) または (b))、データ管理する地点 (c)、データ使用する地点 (d) で非識別処理を行うことができる。最初の地点 (a)、(b) はデータが収集されるが識別情報が実際に必要でない場合、非識別処理を行う場合であり、個人情報保護目的のためにデータを管理する段階であるデータ変換及びデータ保存段階で必要でない識別子を収集時点であらかじめ除去する地点である。2番目の地点である (c) 地点は、非識別情報の管理を避けるためにデータ変換後およびデータ保存前に非識別処理を行うことができる地点である。最後の地点 (d) は、業内で完全に識別されたデータが必要な場合、識別情報は分析用途で使用し、データ共有前に非識別処理を進めることができる地点である [32]。

図3のようにデータライフサイクル全般にわたって目的に合った非識別処理を行う場合、個人情報露出リスクを最小限に抑え、データ共有を容易に行えるとい

うメリットがあり、どの時点で非識別処理を行うべきかについての政策決定には、「データ分析のためのデータ活用目的とデータの特性」などが影響を及ぼしかねない。また、この決定は各データ分析および活用目的によってより効率的でデータ有用性を増大させることができる特定非識別技術を選択するのも役立つ [32]。

図4は韓国国内のガイドライン [31] で提示した4段階の非識別処理手順について扱っており、国際標準として技術従属性などにより適合しない部分を修正補完している。

図5は非識別処理過程のデータ形態と非識別水準に対する図で、非識別処理水準によるデータ形態と非識別処理過程で現れるデータ形態を識別データから非識別データまで表現している。図5に表現された各データ形態は、いずれも個人識別の危険が存在し、それぞれ異なる水準を持つ。特に、図5の左側にある「識別データ形態」は非識別処理が行われるにつれて次第に最も高い水準に変換され、右側の端にあるデータ形態は特定の個人を識別できないデータ形態で再識別可能性が非常に低い形態を示し、左端にあるデータ形態は特定の個人と直接連結された識別可能なデータ形態を示す [32]。

この二つのデータ形態の間には特定個人を識別するための試みを通じて特定個人や個人が含まれたグループを捜し出すことができるデータ形態が存在する。また、特定の個人データに基づくが、元のデータに復旧できないデータ形態も存在する。データの効率的活用のためには、高い水準の非識別処理が進められても要求されるデータの有用性を維持しつつ、図5の右方向の非識別データ形態に変換されなければならない [32]。

本標準は、既存の 20889 国際標準が各技術に対する紹介を主に並べたものとは異なり、全体の非識別処理手順をフレームワークで構成・開発したという点で意義がある。本標準は 2020 年 3 月、ITU-T SG 17 にて定期会議で最終標準として確定された。

5 まとめ

最近、ChatGPT や生成 AI で代表される AI 熱風で、北米、欧州を中心に個人情報侵害の憂慮が提起されている。日本国内にも 2 次改正個人情報保護法が 2022

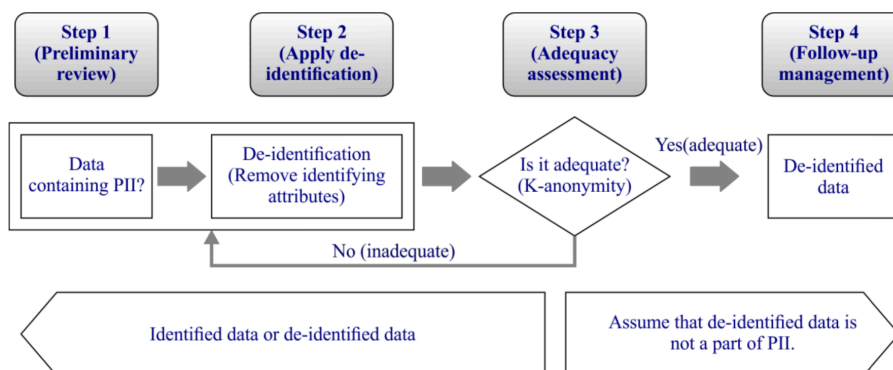


図4 非識別プロセス [29]

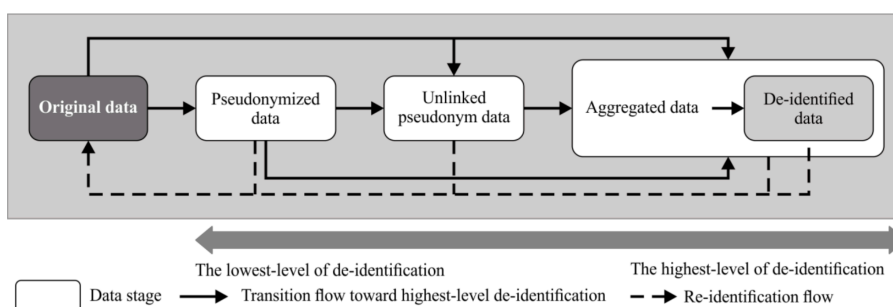


図5 非識別データステージ [29]

年4月1日から施行されているが、まだ法理的理解と非識別化プロセスに関連する支援システムの不備が見られる。ChatGPTや生成AIに利用されるビッグデータ収集が直接的に個人情報を収集するわけではないが、学習のために収集されたデータに個人情報が含まれる可能性があり、また、再識別技術を利用して個人を特定できるリスクが残っているためである。日本も数回の個人情報流出事故を契機に、個人情報を扱う際には匿名化や仮名化、個人情報フィルタリング技術などの適用が義務化されている。今回の個人情報保護法の改正でも仮名化情報に対する法律的根拠が追加されたことで、非識別情報の作成も活用が拡大するものと展望する。音声認識、テキスト生成、写真生成などの生成AI技術、個人化データ生成などデータを活用した経済を急速に成長させるためには、個人情報活用がより重要になり、その必要と共に安定した活用のための非識別化、個人情報保護技術に対する関心と開発が必要になるだろう。特に、非識別化フレームワークの国際標準としてITU-Tで最終採択されたX.1148とISO/IEC 27559などの国際標準の内容と標準化への国際的な動きを調べることで、要求条件に応じた非識別情報を正しく理解し、活用できることを期待し、

データ産業の経済活用化へ役立つ事を図る。

参考文献

- [1] e-GOV | 法令検索: “個人情報の保護に関する法律”, <https://elaws.e-gov.go.jp/document?lawid=415AC0000000057>
- [2] 個人情報保護委員会: “個人情報の保護に関する法律についてのガイドライン”, <https://www.ppc.go.jp/personalinfo/legal/guidelines-tsusoku/>
- [3] 個人情報保護委員会: “令和3年改正個人情報保護法について（官民を通じた個人情報保護制度の見直し）”, <https://www.ppc.go.jp/personalinfo/minaoshi/>
- [4] Sweeney, L.: “Simple Demographics Often Identify People Uniquely”, Data Privacy Working Paper 3, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (2000)
- [5] John O.: “AOL reported to FTC”, https://www.theregister.com/2006/08/16/aol_suffers_ftc_complaint/
- [6] 総務省: “個人情報取扱事業者の責務”, https://www.soumu.go.jp/main_sosiki/joho-tsusin/security/business/executive/05.html
- [7] 個人情報保護委員会: “統計情報と匿名加工情報の違

- いは何ですか”,
https://www.ppc.go.jp/all.faq_index/faq1-q15-2/
- [8] 総務省: “個人情報の保護に関する法律についてのガイドライン”, 個人情報保護委員会 (2016)
- [9] 板倉陽一郎: “個人情報保護法改正によるデータ流通への影響 2017 年改正, 2020 年改正及び 2021 年改正法案”, 公正取引委員会 (2021)
- [10] GDPR (General Data Protection Regulation) :
<https://gdpr-info.eu/>
- [11] 個人情報保護委員会: “EU (外国制度) GDPR (General Data Protection Regulation: 一般データ保護規則)”,
<https://www.ppc.go.jp/enforcement/infoprovision/EU/>
- [12] CCPA (California Consumer Privacy Act) :
<https://oag.ca.gov/privacy/ccpa>
- [13] 個人情報保護委員会: “個人情報保護制度の見直しに関するタスクフォース「個人情報保護制度の見直しに関する最終報告 (概要) ”,
https://www.cas.go.jp/jp/seisaku/kojinjyoho_hogo/pdf/r0212saisyuhoukoku_gaiyou.pdf
- [14] 小松 由季: “行政機関等の保有する個人情報の利活用に向けた法整備 ― 行政機関個人情報保護法等改正法案 ―”, 参議院 (2016)
- [15] 総務省: “個人情報保護法・電気通信事業における個人情報保護に関するガイドライン抜粋”,
https://www.soumu.go.jp/main_content/000305049.pdf
- [16] 経済産業省: “事業者が匿名加工情報の具体的な作成方法を検討するにあたっての参考資料”,
https://www.meti.go.jp/policy/it_policy/privacy/downloadfiles/tokumeikakou.pdf
- [17] 菊池浩明: “個人情報保護法の改正とビッグデータ活用”, 映像情報メディア学会誌, Vol.72, No.6, pp.872-878 (2018)
- [18] Information Commissioner’s Office: “Anonymisation: managing data protection risk code of practice”,
<https://ico.org.uk/media/1061/anonymisation-code.pdf>
- [19] 個人情報保護委員会: “個人情報保護法 いわゆる 3 年ごと見直し 制度改正大綱”,
<https://www.ppc.go.jp/files/pdf/seidokaiseitaiko.pdf>
- [20] 菊池浩明: “匿名加工とは何か”, 情報処理, Vol.59, No.5, pp.441-443 (2018)
- [21] 寺田眞治: “改正個人情報保護法の「匿名」と「仮名」, 加工情報の違いを具体例で理解”,
<https://active.nikkeibp.co.jp/atcl/act/19/00350/012100001/?P=3>
- [22] Lindsay C.: “‘Biggest data grab’ in NHS history
 stuffs GP records in a central store for ‘research’ – and the time to opt out is now”,
https://www.theregister.com/2021/05/13/nhs_data_grab/
- [23] プライバシーテック研究所: “どう見る? JR 東日本のデータ活用と個人情報保護法”,
<https://acompany.tech/privacytechlab/jr-face-recognition-personal-information-protection-law>
- [24] ISO/IEC 20889:2018 Privacy enhancing data de-identification terminology and classification of techniques (2018)
- [25] ISO/IEC 29100:2011 Information technology, Security techniques, Privacy framework (2017)
- [26] ISO/IEC 27559:2022 Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework (2022)
- [27] ISO/IEC 27551:2021 Information security, cybersecurity and privacy protection Requirements for attribute-based unlinkable entity authentication (2021)
- [28] 崎村夏彦: “2022 年のプライバシー標準”,
<https://gihyo.jp/lifestyle/column/newyear/2022/privacy-standards>
- [29] International Telecommunication Union: X.1148 – Framework of de-identification process for telecommunication service providers (2020)
- [30] Lee H. and Lim H.: “金融ビッグデータ処理のための非識別処理技術の国際標準化動向”, Korea Information Processing Society Review, Vol.30, No.3, pp.20-26 (2023)
- [31] Ministry of the Interior and Safety(韓国: 行政安全部): “個人情報非識別措置ガイドライン”, (2016)
- [32] Choi J., Lee Y., Oh Y., and Lim H.: “非識別処理分野の国際標準化動向”, Review of KIISC (Korea Institute of Information Security & Cryptology), Vol.29, No.4, pp.13-18 (2019)