Vol.2018-CSEC-81 No.10
Vol.2018-IOT-41 No.10
2018/5/17

リスクベースアプローチに基づく 実用的な個人データの第三者提供ルール策定方法の検討 －米国HIPAA法を例にして－

井口 誠[1,a)]　植松 太郎[1,b)]　藤井 達朗[1,c)]

概要：昨今，ビッグデータ利活用促進のための取組みとして，個人データの流通が重要な役割を果たすようになっている．この流れを受け，日本でも 2017 年 5 月に改正個人情報保護法が全面施行された．結果，匿名加工情報制度の創設など，プライバシー保護を図りつつ個人データを第三者に提供可能とする環境整備が整いつつある．しかし，環境整備は整ったものの，具体的な個人データの第三者提供ルールはまだ策定されていない状態にある．一例として匿名加工情報を見た場合，一律の基準は存在せず具体的な加工方法は明確にはなっていない．本稿では，リスクベースアプローチを用いて実用的な個人データの第三者提供ルールを策定する手法を検討する．特に米国の HIPAA (Health Insurance Portability and Accountability Act of 1996) 法を例に分析を行い，リスクベースアプローチが直感的な個人データの第三者提供ルールの策定に活用されていることを示す．また分析結果を踏まえ，他分野における第三者提供ルール策定へ向けた方法論を考察する．

キーワード：リスクベースアプローチ，プライバシー，個人データ，第三者提供，HIPAA 法

Leveraging Risk-based Approach for Constructing Systematic Standards for Privacy-conscious Data Sharing: Lessons Learned from the HIPAA Privacy Rule

Iguchi Makoto[1,a)]　Uematsu Taro[1,b)]　Fujii Tatsuro[1,c)]

1. はじめに

デジタルエコノミーの発展において，データ流通は必要不可欠な要因である．従来，データ流通は同一組織内に閉じて行われてきたが，近年では異なる組織間でのデータ流通のケースが多くなっている．さらには，オープンデータなどに代表されるように，データを積極的に一般公開してデータ利活用を促進するケースも増えている．

一方で，個人データの流通を行う場合にはプライバシー侵害のリスクの考慮が必要である．プライバシー侵害防止のためには，流通するデータの匿名化やデータ提供先との契約締結などといった適切な対策を講じる必要がある．しかし，多くの例において適切な安全策は必ずしも明確ではない．例えば，諸国の法規制においてデータの匿名化が推奨されているが，実際にデータを匿名化する手段について明確に定義したものは少ない．さらなるデータ流通と利活用の促進のためには，データ提供に必要な対策を明確かつ実用的な形で定義した規則が必要になることは想像に難くない．

本稿では，プライバシーを保護しつつデータを第三者提供するためのルールを策定する際の足がかりとして，リスクベースアプローチが有効であることを示す．まずは既存

[1] Kii 株式会社
Kii Corporation, Minato-ku, Tokyo 107–0052, Japan
[a)] makoto.iguchi@kii.com
[b)] taro.uematsu@kii.com
[c)] tatsuro.fujii@kii.com

のリスクベースアプローチ手法を基に，プライバシーリスク評価用のフレームワークを定義する．このフレームワークを使い，米国 HIPAA (Health Insurance Portability and Accountability Act of 1996) Privacy Rule を分析し，リスクベースアプローチが直感的なルール策定に大きく寄与することを示す．さらに分析結果より，他分野における個人データの第三者提供ルール策定に活用可能な要因を考察する．

2. 関連研究

データのセキュリティ・プライバシー保護の分野にリスクベースアプローチを適用した先行事例はいくつか存在する．データセキュリティの三要素である機密性・完全性・可用性に対するリスクを評価する手法としては，例えば国際標準化機構 (ISO) がまとめた指針 [1] や，国際的オンラインコミュニティである Open Web Application Security Project (OWASP) が提唱する評価手法 [2]，アメリカ国立標準技術研究所がまとめたガイダンス [3] などが代表的である．

異なる組織間でのデータ流通が盛んになるにつれ，データの第三者提供に伴うプライバリーリスクをリスクベースアプローチを用いて評価する手法が提案され始めた [4]．提案された手法の一例としては，フランスの情報処理および自由に関する国家委員会 (Commission Nationale de l'Informatique et des Libertés: CNIL) によるフレームワーク [5] や，イギリスの個人情報保護監督機関 (ICO) による実務指針 [6]，ISO によるガイドライン [7]，カナダオンタリオ州の情報監督機関によるガイダンス [8]， El Eman らによる提案手法 [9] などが挙げられる．

本稿は，これらの既存手法を参考にプライバシーリスク評価用のフレームワークを定義している．本稿の新規性は，定義したフレームワークを用いて HIPAA Privacy Rule をリスクベースアプローチの観点から再評価した点にある．

3. リスクベースに基づくプライバシーリスク評価フレームワーク

本節では，データのプライバシーリスクをリスクベースアプローチに基づいて評価するフレームワークを提案する．ここで提案したフレームワークは，第 4 節において HIPAA Privacu Rule を解析する際に利用する．

第 2 節で述べたように，我々は既存手法を参考に簡素化したプライバシーリスク評価フレームワークを構築した．このフレームワークでは，データのプライバシーリスクを以下の手順で評価する．

(1) プライバシーリスクを特定する．
(2) リスクの「重大性 (Severity Level)」を評価する．
(3) リスクの「発生可能性 (Likelihood Level)」を評価する．
(4) 重大性と発生可能性を基に，データの総合的なプライバシーリスクを評価する．

以下，それぞれの手順について解説する．

3.1 プライバシーリスクの特定

初めに，評価の対象となるプライバシーリスク自体を明確にする必要がある．リスクは，第三者提供されたデータが「どのように」悪用され得るかと，「だれが」悪用し得るかを分析することによって明確化される．

「どのように」悪用され得るかは，攻撃者がデータを悪用しデータ主体のプライバシーを侵害するシナリオを考察して分析する．通常，データ悪用シナリオは 1) 他データとの照合，2) 再識別によるデータ主体の特定，3) データの悪用によるデータ主体への損害というステップを踏む．ここでの分析結果は，後ほどリスクの重大性を評価する際の判断基準となる．

「だれが」悪用し得るかを分析するには，データ提供先の特性やデータ提供方法などを評価する．ここでの分析結果は，後ほどリスクの発生可能性を評価する際の判断材料となる．

3.2 リスクの重大性の評価

重大性は，データが悪用された場合に引き起こされる影響の大きさを示す指標である．あるデータの重大性は，このデータの「損害レベル (Prejudicial Level)」と「識別可能性 (Identification Level)」によって決定される．

3.2.1 損害レベル (Prejudicial Level)

損害レベルは，データが悪用された場合に発生する損害の度合いを示す指標である．例えば，顧客のクレジットカード番号とセキュリティコードを含むデータの損害レベルは，顧客のメールアドレスのみを含むデータより高めに査定する．

発生しうる損害を評価する際には，人身損害 (Physical damage)，精神的損害 (Moral damage)，財産的損害 (Material damage) など，複数の種類の損害について検討する [5]．

本稿では，データの損害レベルを以下の通りに 3 段階査定する．

- 高レベル: データ主体が回復不能な損害を被る可能性がある（例：人命の喪失）．
- 中レベル: データ主体が回復可能な損害を被る可能性がある（例：金銭的損失）．
- 低レベル: データ主体は影響を受けないか，回復が極めて容易な損害を被る可能性がある（例：SPAM メッセージの受信）．

3.2.2 識別可能性 (Identification Level)

識別可能性は，データ主体が再識別される可能性を示す

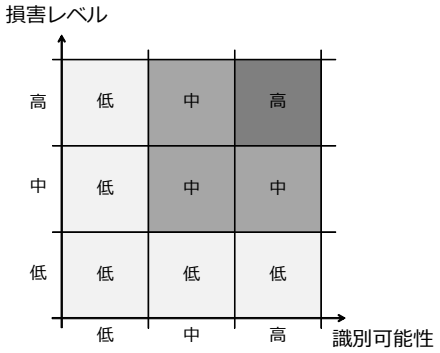

図 1　重大性評価マトリクス
Fig. 1　Severity Assessment Matrix

指標である．例えば，多くの個人識別可能情報 (Personally identifiable information: PII) を含むデータは，少ない PII を含むデータと比較して，データ主体が再識別される可能性が高いといえる．このため，前者のデータの識別可能性は，後者より高めに査定する．

既存技術としてデータのランダム化や一般化，仮名化などといった，データの再識別リスクを軽減するためのデータ加工手法が提案されている [10], [11]．もしこれらのデータ加工が施されている場合は，加工度合いに応じてデータの識別可能性を低めに査定する．

損害レベル同様，本稿ではデータの識別可能性を以下の通りに 3 段階査定する．

- 高レベル: データ主体の再識別は極めて容易である．
- 中レベル: データ主体の再識別は容易ではないが，不可能ではない．
- 低レベル: データ主体の再識別はほぼ不可能である．

3.2.3　損害レベルと識別可能性に基づく重大性の評価

データリスクの重大性は，損害レベルと識別可能性の評価結果を基に査定する．査定には 図 **1** の重大性評価マトリクスを用いる．

図 1 で示されているように，損害レベルまたは識別可能性が単独で高レベルであっても，重大性の査定は高レベルにはならない．一例として，「日本のどこかに在住中のとある女性は，アルコールを摂取した後に薬物 X を飲むと死亡する」というデータの損害レベルは高レベルだが，このデータの識別可能性は低レベルのため，最終的な重大性は低レベルと査定される．同様に「○○県××市に住む情報太郎さん（△歳）は，毎日コップ一杯の水を飲む」というデータの識別可能性は高レベルだが，このデータの損害レベルは低レベルのため，最終的な重大性は低レベルと査定される．

ここで査定した重大性のレベルは，データの総合的なプライバシーリスクを評価するためのパラメータになる．

3.3　リスクの発生可能性の評価

発生可能性は，データが悪用される確率を示す指標である．あるデータの発生可能性は，このデータの「脅威レベル (Threat Level)」と「脆弱性レベル (Vulnerability Level)」によって決定される．

3.3.1　脅威レベル (Threat Level)

脅威レベルは，データ提供先が共有されたデータの悪用を試みる確率を示す指標である．脅威レベルの査定は，データ提供先の以下の属性などを考慮して行う．

提供先のサイズ
　データが多数の人間に共有される場合，少人数に限定して共有される場合と比較して，データ悪用が試みられる確率は高いとみなし脅威レベルを高く査定する．

動機
　データ提供先が共有データの悪用を試みる何らかの動機を有している場合は，脅威レベルを高く査定する．

信頼性
　何らかの理由でデータ提供先の信頼性に疑問がある場合は，脅威レベルを高く査定する．

データの脅威レベルは，以下の通りに 3 段階査定する．

- 高レベル: データが悪用される可能性は高い．
- 中レベル: データが悪用される可能性は中程度．
- 低レベル: データが悪用される可能性は低い．

3.3.2　脆弱性レベル (Vulnerability Level)

脆弱性レベルは，データの弱さを示す指標である．言い換えると，データ悪用を防ぐための対策の実施度合いを示す指標である．対策の実施状況が十分である場合は脆弱性レベルを低く査定し，不十分である場合は高く査定する．

一般的に，セキュリティ対策は以下の 3 つに分類される．脆弱性レベルの査定時には，これら全てを判断基準とする．

人的対策 (Administrative safeguards)
　データ悪用防止のためのポリシーや手順（例：手順書，教育，契約書締結など）．

物理的対策 (Physical safeguards)
　共有データを物理的に保護する施策（例：入退管理）．

技術的対策 (Technical safeguards)
　データ悪用防止のための技術的対策（例：認証，暗号化，ログ監査など）．

他の指標同様，脆弱性レベルは以下の通りに 3 段階査定する．

- 高レベル: 現在の対策状況は不十分である
- 中レベル: 現在の対策状況は有効だが，改善の余地がある．
- 低レベル: 現在の対策状況は十分である．

3.3.3　脅威性と脆弱性に基づく発生可能性の評価

データリスクの発生可能性は，脅威性と脆弱性の評価結果を基に査定する．査定には 図 **2** の発生可能性評価マトリクスを用いる．

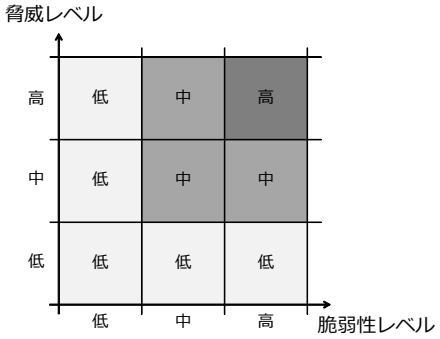

図 2 発生可能性評価マトリクス
Fig. 2 Likelihood Assessment Matrix

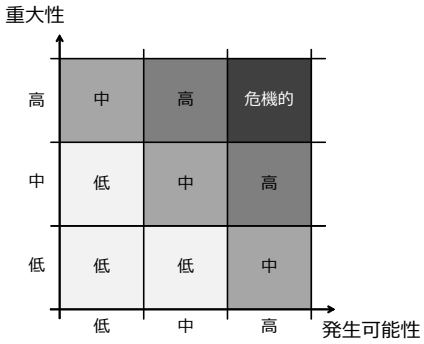

図 3 総合的なプライバシーリスクの評価マトリクス
Fig. 3 Overall Privacy Risk Assessment Matrix

図 2 で示されているように，脅威レベルまたは脆弱性レベルが単独で高レベルであっても，発生可能性レベルの査定は高レベルにはならない．例えば，データ提供先の信頼性に疑問がある（脅威レベルが高い）状況であっても，十分なセキュリティ対策を施す（脆弱性レベルを低く抑える）ことにより，リスクの発生可能性を低レベルに抑制することができる．同様に，仮にデータ提供先が 100% 信頼できる（脅威レベルが低い）場合は，セキュリティ対策は最低限（脆弱性レベルが高）でも問題ないと考えることができる．

ここで査定した発生可能性は，データの総合的なプライバシーリスクを評価するためのパラメータになる．

3.4 データの総合的なプライバシーリスクの評価

最後に，査定した重大性と発生可能性を基に，データの総合的なプライバシーリスクを査定する．査定は **図 3** の評価マトリクスを用いて行う．

評価結果は次のように解釈する．

- 低レベルの場合，現状は理想的な状態である．
- 中レベルの場合，現状は許容可能な状態である．プライバシーリスクを低レベルに改善する必要がある場合は，現時点において中レベル以上と評価されている重大性または発生可能性を改善する．
- 高レベルまたは危機的レベルの場合，現状は許容できない状態である．現時点において高レベルと評価されている重大性または発生可能性を改善する必要がある．

重大性の改善は，データの識別可能性を下げることで可能である．例えば，データの再識別リスクを下げるための加工をデータに施すなどの対策を検討する．

発生可能性の改善は，データの脆弱性レベルを下げることで可能である．例えば，データ悪用を防ぐために，強度のより高い暗号方式の導入やログ監査の強化などの対策を検討する．

4. リスクベースアプローチによる HIPAA Privacy Rule の分析

本節では，前節で解説したプライバシーリスク評価フレームワークを用いて HIPAA Privacy Rule を分析し，リスクベースアプローチが実用的なデータ共有ルールの策定に有効であることを示す．まず初めに HIPAA Privacy Rule の概要と，Privacy Rule において定義されている 3 つのデータ共有方式を解説する．次に，プライバシーリスク評価フレームワークに基づいてこれらのデータ共有方式を評価した結果を示す．

4.1 HIPAA の概要

HIPAA [12] は Health Insurance Portability and Accountability Act of 1996 の略称であり，米国における保護対象保健情報 (Protected Health Information: PHI) の取り扱いを安全かつ効率的に行うために制定された法律のことを指す．2009 年には Health Information Technology for Economic and Clinical Health (HITECH) Act により対象範囲が拡張され，PHI のセキュリティおよびプライバシーを保護するための国家基準を規定している．

HIPPA は，Covered Entity と Business Associate に対して適用される (45 CFR §160.103)．

Covered Entity (CE)
ヘルスケアプロバイダー (例 : 医師，薬局)，ヘルスプラン（例 : 医療保険関係者)，ヘルスケア情報センター．

Business Associate (BA)
Covered Entity の代理で PHI を処理・管理するサービスを提供する事業提携者．一例として，Covered Entity の代わりに PHI の管理を行うクラウドサービスプロバイダーは Business Associate に該当する [13]．

HIPAA は， PHI の機密性・完全性・可用性の保護へ向けた国家基準を定める Security Rule と，PHI のデータ主体のプライバシー保護へ向けた国家基準を定める Privacy Rule により構成されている．第三者へのデータ共有方式は Privacy Rule においてカバーされているため，本稿ではこちらにフォーカスする．

4.2 HIPAA Privacy Rule が定義する PHI 共有方式

HIPAA Privacy Rule では，以下の 3 つのデータ共有方式が定義されている．いずれの方式についても，共有対象となるデータセットの仕様やデータ共有先に適用すべき安全策が明確に定義されている [14].

4.2.1 加工せずにそのまま共有

1 つ目の共有方式は，PHI をそのまま開示する方式である．この方式は，CE がヘルスケア情報処理の一部を BA に委託する場合に用いられる[*1]．例えば，クラウドサービスプロバイダーが CE からの委託を受けて PHI の管理を行う場合，通常 PHI はそのまま BA たるクラウドサービスプロバイダーに共有される．

この方式で PHI を開示する場合，BA が PHI を HIPAA Security Rule に準拠した形で処理するよう監督する義務が CE には課せられる．この確証を得るために，CE は BA との間に Business Associate Agreement (BAA) を締結することが義務づけられている．

4.2.2 匿名化データセットとして共有

2 つ目の共有方式は，PHI を匿名化データセット (de-identified data set) に加工して開示する方式である．匿名化データセットはデータ主体との関連付けが完全に絶たれるため，PHI とは見なされなくなる．よって，CE はこのデータを任意の人間に対して，任意の方法で開示することができる．

HIPAA Privacy Rule では，PHI の匿名加工方法として「専門家による判定 (45 CFP §164.514(b)(1))」と「Safe Harbor 法 (45 CFR §164.514(b)(2))」を定義している．

専門家による判定の場合，統計的かつ科学的な方法に精通した専門家が，データの性質などに合わせて PHI の匿名加工方法を決定する．この際，HIPAA Privacy Rule は，選択した加工方法が再識別リスクを「極めて小さく」することを要求している．

Safe Harbor 法は，より直感的なルールを定めている．この方法は，匿名化のために削除すべき 18 個の識別子を定義している（**表 1**）．PHI よりこれらの識別子が削除されると，データは匿名化されたとみなされる．

4.2.3 限定されたデータセットとして共有

3 つ目の共有方式は，PHI を限定されたデータセット (limited data set) に加工して開示する方式である (45 CFR §164.514(e))．HIPAA Privacy Rule は，PHI を限定されたデータセットに加工するために削除すべき 16 個の識別子を定義している（表 1）．表に示されているように，Safe Harbor 法では削除が義務付けられている 2 つの識別子が，限定されたデータセットにおいては許可されている．

[*1] 例えば改正個人情報保護法においては「委託」は「第三者提供」には当てはまらないとされているが，本稿では広義の意味での「第三者」に対するデータ提供であると解釈し，この方式を他の第三者に対するデータ共有方式と同列に扱う．

表 1 匿名化データセットと限定されたデータセットの比較
Table 1 Comparison of de-identified and limited data set

識別子	匿名化データセット (Safe Harbor 法)	限定されたデータセット
名前	完全削除	完全削除
地理情報	州名より細かい情報を削除．人口が 20,000 以上の地域については 3 桁の ZIP コードを許可	町名，市名，州名，ZIP コード以外の住所情報を削除
日付	完全削除	許可
電話番号	完全削除	完全削除
FAX 番号	完全削除	完全削除
メールアドレス	完全削除	完全削除
社会保障番号	完全削除	完全削除
カルテ番号	完全削除	完全削除
保険番号	完全削除	完全削除
口座番号	完全削除	完全削除
免許書番号	完全削除	完全削除
車両番号	完全削除	完全削除
装置の識別番号	完全削除	完全削除
URL	完全削除	完全削除
IP アドレス	完全削除	完全削除
生体情報	完全削除	完全削除
顔写真	完全削除	完全削除
上記以外の識別子	完全削除	許可

限定されたデータセットは，研究目的，公衆衛生，またはヘルスケアオプションのための利用に使用用途が限定される．また，データの共有に先立ち，データ提供先と Data Use Agreement (DUA) を締結する必要がある (45 CFR §164.514(e)(4))．DUA では，以下を含む内容が規定される．

- 限定されたデータセットの利用および開示目的として許容される内容の定義．
- 上記の許容内容を逸脱する利用や開示を防止するために施すべき対策の定義．
- データの再識別，およびデータ主体への連絡の禁止．

一般的に，DUA は BAA より弱い法的拘束を行うものと言える．

4.3 フレームワークによる PHI 共有方式の評価

HIPAA Privacy Rule において，リスクベースアプローチに関する明確な言及はなされていない[*2]．唯一の言及は HIPAA のガイドライン [14] においてなされており，専門家による判定により PHI の匿名加工を行う際にはデータリスクの評価を行わなければならないと記されている．

明確に主張こそされていないが，HIPAA Privacy Rule を解析すると，定義されている PHI 共有方式の全てがリスクベースアプローチに基いて設計されていることが見て取れる．以下に，第 3 節で解説したプライバシーリスク評価フレームワークに当てはめて，それぞれの PHI 共有方式を評価した結果をまとめる．

- 加工せずにそのまま共有する場合

[*2] HIPAA Security Rule においては，PHI のセキュリティリスクを評価することが義務付けられている (45 CFR §164.308(a)(1)(ii)(A))

– 損害レベル: データは PHI なので，高レベルと査定．
– 識別可能性: 無加工のままデータ共有するため，高レベルと査定．
– 脅威レベル: 信頼できる BA を選択することが義務付けられているため，低レベルと査定．
– 脆弱性レベル: BAA の締結など十分な安全策を施すことが義務付けられているため，低レベルと査定．

- 匿名化データセットとして共有する場合
  – 損害レベル: データは PHI なので，高レベルと査定．
  – 識別可能性: データは匿名化されているため，低レベルと査定．参考情報として，Benitez と Malin による調査では，Safe Harbor 法によって匿名化された PHI の再識別リスクは 0.01% から 0.25% であったと報告されている [15]．
  – 脅威レベル: データは任意の人間に開示可能なため，高レベルと査定．
  – 脆弱性レベル: データは安全策なしで開示可能なため，高レベルと査定．
- 限定されたデータセットとして共有する場合
  – 損害レベル: データは PHI なので，高レベルと査定．
  – 識別可能性: データは部分的に匿名化されているため，中レベルと査定．前出の Benitez と Malin による調査では，限定されたデータセットに加工された PHI の再識別リスクは 10% から 60% であったと報告されている [15]．
  – 脅威レベル: データの利用目的は，研究目的，公衆衛生，またはヘルスケアオプションに限定されるため，CE は比較的信頼できるデータ共有先を選択する可能性が高い．よって中レベルと査定．
  – 脆弱性レベル: データ共有先は，DUA の内容に従って，限定されたデータセットを適法かつ安全に使用する義務を負う．ただし DUA による拘束力は BAA より弱いため，中レベルと査定．

各 PHI 共有方式のデータプライバシーリスク評価結果を **表 2** にまとめる．HIPAA Privacy Rule で定義された全ての PHI 共有方式が，最終的なプライバシーリスクを許容レベル（中レベル）に抑制していることがわかる．これは，それぞれの方式において共有対象となるデータの内容に応じて適切なレベルの対策適用を義務付けることにより，プライバシーリスクの重大性と発生可能性のバランスをうまく取っているためである．

5. 他分野における実用的な個人データの第三者提供ルール策定へ向けた考察

本節では，第 4 節の HIPAA Privacy Rule 分析結果より得られた知見について議論する．ここでは，以下の 2 点に論点を絞り，特に日本国内で第三者提供ルールを策定するケースを想定した考察を行う．

- 明確なデータ匿名化ルールの策定
- 異なる識別可能性レベルを持つデータセットへの対応

5.1 明確なデータ匿名化ルールの策定

HIPAA Privacy Rule においては，Safe Harbor 法によるデータ匿名化の明確な手順が定義されており，CE にとって実用的なルールになっている．

5.1.1 HIPAA Privacy Rule の解析で得られた知見

HIPAA Privacy Rule においてデータ匿名化ルールが明確に定義できた大きな要因の 1 つに，ルールの適用対象が PHI に限定されていたことが挙げられる．適用対象となるデータを限定化した効果は以下の通りである．

- 対象データの内容が固定化される．固定化により，データが悪用された場合に発生しうる損害内容が絞られるため，データの損害レベルが固定化される．HIPAA の場合，対象データが PHI であるため，損害レベルは HIGH に固定化された．
- 対象データに含まれる PII が固定化されるため，明確なデータ匿名化ルールが定義しやすくなる．HIPAA の場合，どの識別子を削除すべきかを定義した具体的なルールが策定されていた．
- 対象データとデータ匿名化ルールが明確に定まるため，データ匿名化ルールが検証可能になる．HIPAA の場合，Safe Harbor 法により規定された匿名化データセットと限定されたデータセットの妥当性が，ルール策定後に Benitez と Malin [15] により検証された．

仮に適用対象を絞り込まずにルール策定を試みた場合，以下のように系統的なデータ提供ルールの策定は困難になる可能性が高い．

- 様々なデータが適用対象になるため，発生しうる損害内容が多岐にわたる．この結果，実用的な侵害レベルの査定が困難になる．
- データによって含まれる PII にばらつきが生じるため，削除すべき識別子などルールの具体化に必要な情報を決定できない．その結果，データ匿名化ルールは曖昧にならざるを得ない．
- 対象データとデータ匿名化ルールが曖昧なため，後ほどデータ匿名化ルールを検証することが困難になる．

5.1.2 他分野での第三者提供ルール策定へ向けた考察

他分野において個人データの第三者提供ルールを策定する場合も，HIPAA Privacy Rule 同様に適用対象データを限定的にするべきである．この知見自体は，日本においても一定の認知を得ていると思われる．一例として，経済産業省が 2016 年に公開した「匿名加工情報作成マニュアル」[16] では，電力利用データ，購買データ，移動データという 3 つのユースケースが取り上げられており，それぞれのユースケースに閉じた形でデータ匿名化方法が例示されている．また，日本情報経済社会推進協会 (JIPDEC) が

表 2　プライバシーリスク評価フレームワークによる PHI 共有方式の評価結果
Table 2　Evalution of PHI sharing methods with the privacy risk assessment framework

開示方法	識別可能性	損害レベル	重大性	脅威レベル	脆弱性レベル	発生可能性	総合的なプライバシーリスク
無加工のまま開示	高		高	低	低	低	中
匿名化データセットとして開示	低	高	低	高	高	高	中
限定されたデータセットとして開示	中		中	中	中	中	中

2017 年に公開した「匿名加工情報の事例集」 [17] では，所有車データ，顧客データ，購買履歴，移動履歴（人の流れ）という 4 つのユースケースにおけるデータ匿名化の参考例がまとめられている．ただし，いずれの資料もあくまで参考情報として提示されているにすぎず，実際のルールを規定するものではない．

日本の改正個人情報保護法においては，認定個人情報保護団体が，限定されたデータに対する第三者提供ルールの策定にふさわしいポジションにいると言える．認定個人情報保護団体は，安全管理措置や匿名加工情報の作成方法などを定めた個人情報保護指針を，業界の特性などを考慮した上で作成する役割を担っている [18]．このため，各認定個人情報保護団体は，自身が所属する業界に関わるデータに限定した匿名化ルールを策定し，これを個人情報保護委員会に提出することが期待される．しかしながら，現在までに個人情報保護委員会に届け出された個人情報保護指針の中で，具体的な匿名加工の手順や指標の策定まで踏み込んだ内容のものは存在しない．また，届け出された個人情報保護指針の多くが，個人情報委員会のガイドラインとほぼ変わらない内容であるとの報告もある [19]．業界におけるデータ利活用を促進するには，各認定個人情報保護団体が担当業界におけるデータに限定した分析を実施し，明確なデータ匿名化手順を指針としてまとめるように促す必要がある．

5.2　異なる識別可能性レベルを持つデータセットへの対応

HIPAA Privacy Rule では，表 2 で示したように識別可能性の異なる 3 種類のデータセット（無加工データ，匿名化データセット，限定されたデータセット）が定義されている．また，それぞれのデータセットについて系統的な第三者提供ルールが定められている．

5.2.1　HIPAA Privacy Rule の解析で得られた知見

HIPAA Privacy Rule は，どのデータセットを選択した場合についても，そのデータセットの重大性に応じて，適切にリスクの発生可能性をコントロールするデータ提供ルールが定義されている．このため，CE は状況に応じて適切なデータセットを選択し，これに応じた安全策を講じることができる．

3 種のデータセットのうち，注目すべきは識別可能性が中レベルに査定される，限定されたデータセットである．一般的にデータの匿名性と有用性は反比例の関係にあり，データを完全に匿名化すると有用性が許容レベルを超えて劣化してしまうケースがある．このような場合，CE はデータを部分的に匿名化して限定されたデータセットを作成の上，これを共有することができる．限定されたデータセットは，無加工データを提供する場合と比較してより緩やかな安全策のみでデータ提供が行える．このように，データの有用性やデータ共有の利便性を考慮した場合，限定されたデータセットという選択肢の存在は大きい．

5.2.2　他分野での第三者提供ルール策定へ向けた考察

他分野において個人データの第三者提供ルールを策定する場合も，異なる識別可能性レベルを持つデータセットを前提とするべきである．特に，限定されたデータセットに相当する，部分的に匿名化された（識別可能性が中レベルな）データセットをサポートするメリットは大きいと思われる．以下，考えられるメリットを 2 点ほどまとめる．

- 分野によってはデータの完全な匿名化手順の定義がそもそも困難である可能性がある．一例として EU の第 29 条データ保護作業部会が公開した匿名化技術に関する意見書 [10] では，既存の匿名化技術に完璧なものは存在しないと結論づけられており，完全な匿名化が難しいケースがあることが示唆されている．このようなケースにおいても，合理的な水準まで匿名化を施したデータを中レベルの識別可能性を持つデータセットと定義することにより，無加工な（識別可能性が高レベルな）個人データとは異なる第三者提供ルールを策定することが可能になる．
- 準同型暗号などに代表される，個人データに対して暗号化を施すプロトコルを用いたデータ共有を採用する場合を考える．日本を含む多くの国において，暗号化された個人データは，復号鍵を廃棄しない限り完全匿名化されたデータとは見なされない．そこで，暗号化された個人データを識別可能性が中レベルのデータと定義する．この場合，発見可能性を中レベルに抑制する安全策適用を義務付ける共有ルールを策定することになる．例えば準同型暗号を用いたプロトコルの場合，データ提供先がデータに対して実施可能な演算は，プロトコルによって技術的に制限される．これは HIPAA Privacy Rule において，データ提供先がデータに対して行える操作を DUA により法的に制限して

いる状態に類似している．そこで，準同型暗号プロトコルの適用がデータの脆弱性を中レベルに抑制する安全策であると定義の上，これをデータ共有規則に盛り込むことが考えられる．

日本の法制度においては，認定個人情報保護団体が作成する個人情報保護指針において，部分的に匿名化されたデータセットまで踏み込んだ検討が行えるかどうかが 1 つの大きなポイントになる．今後，匿名化データとも無加工データとも異なる第 3 の選択肢を用意の上，これに対応する安全管理措置やデータ加工手順をカバーした個人データ提供ルールを定めた個人情報保護指針が数多くの認定個人情報保護団体より作成されることを期待する．

6. おわりに

本稿では，系統的かつ実用的な個人データの第三者提供ルールの策定に向けた検討を行った．まずリスクベースの考え方を基にしたプライバシーリスク評価フレームワークを定義し，このフレームワークを使って HIPAA Privacy Rule を分析した．分析の結果より，HIPAA Privacy Rule が，リスクの重大性と発生可能性のバランスをうまく取ることにより系統的かつ実用的なデータ共有ルール策定に成功していることが明らかとなった．

さらに，上記分析結果より得た知見を，特に日本において新たな個人データの第三者提供ルールを策定する際にどう活用可能か考察した．今後，認定個人情報保護団体などが第三者に対するデータ共有に関わる指針を作成するにあたり，本稿で考察した内容が何らかの参考になれば幸いである．

参考文献

[1] ISO/IEC 27005:2011: Information technology — Security techniques — Information security risk management, Standard, International Organization for Standardization (ISO) (2011).
[2] Open Web Application Security Project (OWASP): OWASP Risk Rating Methodology, (online), available from ⟨https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology⟩ (accessed 2018-02-01).
[3] Ross, R. S.: NIST SP 800-30 Rev. 1: Guide for Conducting Risk Assessments, Technical report (2012). https://dx.doi.org/10.6028/NIST.SP.800-30r1.
[4] Garfinkel, S. L.: NISTIR 8053: De-Identification of Personal Information, Technical report (2015). https://doi.org/10.6028/nist.ir.8053.
[5] Commission Nationale de l'Informatique et des Libertés (CNIL): *Methodology for Privacy Risk Management, Translation of June 2012 edition* (2012).
[6] Information Commissioner's Office: *Anonymisation: Managing Data Protection Risk, Code of Practice* (2012).
[7] ISO/IEC 29134:2017: Information technology – Security techniques – Guidelines for privacy impact assessment, Standard, International Organization for Standardization (ISO) (2017).
[8] Information and Privacy Commissioner of Ontario: *De-identification Guidelines for Structured Data* (2016).
[9] El Emam, K. and Arbuckle, L.: *Anonymizing Health Data: Case Studies and Methods to Get You Started*, O'Reilly Media, Inc., 1st edition (2013).
[10] European Commission Article 29 Data Protection Working Party: *Opinion 05/2014 on Anonymisation Techniques* (2014).
[11] Mendes, R. and Vilela, J. P.: Privacy-Preserving Data Mining: Methods, Metrics, and Applications, *IEEE Access*, Vol. 5, pp. 10562–10582 (online), DOI: 10.1109/ACCESS.2017.2706947 (2017).
[12] U.S. Department of Health & Human Services, Office for Civil Rights: Health Information Privacy, (online), available from ⟨https://www.hhs.gov/hipaa/index.html⟩ (accessed 2018-02-01).
[13] U.S. Department of Health & Human Services, Office for Civil Rights: Guidance on HIPAA & Cloud Computing, (online), available from ⟨https://www.hhs.gov/hipaa/for-professionals/special-topics/cloud-computing/index.html⟩ (accessed 2018-02-01).
[14] U.S. Department of Health & Human Services, Office for Civil Rights: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, (online), available from ⟨https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html⟩ (accessed 2018-02-01).
[15] Benitez, K. and Malin, B.: Evaluating re-identification risks with respect to the HIPAA privacy rule, *Journal of the American Medical Informatics Association*, Vol. 17, No. 2, pp. 169–177 (online), DOI: 10.1136/jamia.2009.000026 (2010).
[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（「匿名加工情報作成マニュアル」）Ver1.0 (2016).
[17] 日本情報経済社会推進協会 (JIPDEC) 認定個人情報保護団体事務局：匿名加工情報の事例集 (2017). https://www.jipdec.or.jp/protection_org/u71kba00000001hh-att/AOP_006.pdf.
[18] 個人情報保護委員会：認定個人情報保護団体制度の概要，（オンライン），入手先 ⟨https://www.ppc.go.jp/personal/nintei/summary/⟩（参照 2018-04-01）.
[19] 石田　茂：認定個人情報保護団体における匿名加工情報に係る取組み状況について，情報処理学会研究報告，Vol. 2018-EIP-79, No. 10, pp. 1–6 (2018).