

医療等情報にかかる民間データベースの 現状と連携上の課題

一般社団法人医療データベース協会 ディレクター
情報処理安全確保支援士・弁護士・弁理士
足立 昌聰

2024.3 医療等情報の二次利用に関するワーキンググループ 技術作業部会資料

医療データベース協会（AMDJ）について

- 医療、介護及び健康等に関連し民間事業者によって運営されるデータベースが安心して利活用されるために、意見形成や環境づくりを行い、また、その活動を通して、**医療データベース業界を健全に発展**させ、患者を含む生活者のQOL向上へと貢献することを目的とし、2014年4月に設立。
- 2017年2月より、**認定個人情報保護団体**として、会員企業の個人情報の適正な取扱いを支援。**医療データベース事業者の”trust”(信任)の確保**のための活動を展開。

“安心・安全な医療DBをすべての人に”

近年、日本における医療データの利活用への社会的要請が益々高まっております。しかし、国際的な疫学研究にも通用するデータ品質の保証、患者やデータ提供者のプライバシーに十分配慮したデータベース利活用のあり方など、医療・公共団体・企業などの関係各界を横断して解決されるべき課題が散見されます。

当協会では、各界の関係者が安心して医療DBを活用でき、ひいては医療DB産業が医療の質の向上や公衆衛生の発展に貢献できるための環境整備について、民間のデータベース実務家としての立場から意見形成を行ってまいります。

- 2024年3月1日現在、正会員9社、賛助会員2社の計11社が加盟。
- 日本薬剤疫学会が公表している「日本における臨床疫学・薬剤疫学に応用可能なデータベース*」で紹介されている営利企業11社中8社が加盟。

* https://www.jspe.jp/mt-static/FileUpload/files/JSPE_DB_TF_J.pdf

民間の医療データベース

【データの由来による分類】

- 民間の医療データベースは、概ね、データ生成時の個人情報取扱事業者・行政機関等の種類に大別して、①医療機関由来、②保険者由来、③保険薬局由来の3種の系統が存在する。

【データの法的性質による分類】

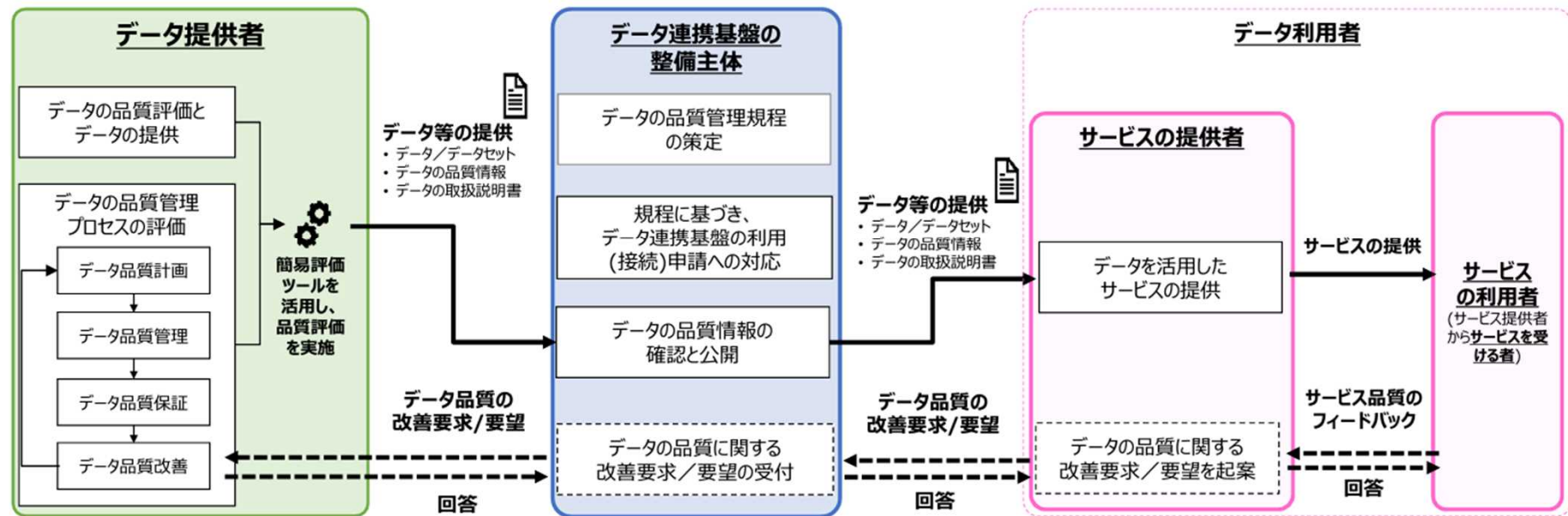
- 2024年3月現在、民間の大規模医療データベースはすべて、**個人情報保護法に基づく匿名加工情報（行政機関等匿名加工情報を含む。）**で構成されている。本人同意に基づく個人データの取得を介するケースは、PHR/CMP等を介した前向き臨床研究など限定的。
- 2024年3月現在、次世代医療基盤法に基づく匿名加工医療情報の提供は認定された3スキームにとどまり、**営利企業が認定匿名加工医療情報作成事業者になった例はない。**

【データの属性による分類】

- 電子カルテ等の**アウトカム**（診療行為を実施した結果に関する情報）を含むものと、レセプト等の**インプット**（診療行為の実施に関する情報）しか含まないものに大別される。
- 次世代医療基盤法では、アウトカム（公的DBですでにカバーされているものを除く。）を含まないデータベースは認定されない。

データの品質管理に対する考え方

- データの品質管理については、ISO/TS 8000で国際標準化されており、データ提供者からデータ利用者に至るプロセス*について、品質管理のPDCAサイクルを回すことが適切。
- レギュラトリー領域などでは、プロセスに関連するデータ特性（追跡可能性等）も重要。



* 2023.9内閣府地方創生推進事務局「データ連携基盤を通して提供されるデータの品質管理ガイドブック」より
https://www.chisou.go.jp/tiiki/kokusentoc/supercity/pdf/supercity_230926_guidebook_honsi.pdf

データ自体の品質特性

- データ品質の国際標準規格であるISO/IEC 25012では、**15種の品質特性**が定義されている。
- 正確性、完全性及び一貫性（基礎的品質特性）のほか、**最新性や効率性などの商業的な付加価値となり得る品質特性（付加的品質特性）**により市場競争が行われている。

正確性 (Accuracy)	データに誤りがなく、信頼性がある状態か
完全性 (Completeness)	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか
一貫性 (Consistency)	データに矛盾はないか、データ間で整合性が取れている状態か
信憑性 (Credibility)	出所や更新日が明示されているか
最新性 (Correctness)	更新期間や最終更新日が明示されているか
アクセシビリティ (Accessibility)	環境依存文字等が使用されていないことが明示されているか
標準適合性 (Compliance)	標準に準拠したデータ形式であることが明示されているか
機密性 (Confidentiality)	アクセス制限等の対策が講じられていることが明示されているか
効率性 (Efficiency)	効率化のため、コードの割り当て等を行っていることが明示されているか
精度 (Precision)	データの精度について明示されているか
追跡可能性 (Traceability)	データの変更履歴（変更者／変更日等）が明示されているか
理解性 (Understandability)	データ全体や項目ごとにメタデータが明示されているか
可用性 (Availability)	利用時間帯等について明示されているか
移植性 (Portability)	標準的な形式でデータをエクスポートできることが明示されているか
回復性 (Recoverability)	バックアップ等の障害時の対応方針が明示されているか

品質特性における協調領域

- **基礎的品質特性、アクセシビリティ及び追跡可能性**については、データの加工段階ではもはや**データ欠損が回復できないし、文字コード置換等の作業は付加価値となりづらく、上流の段階でこれらの品質特性が担保されることが強く期待される。**

正確性 (Accuracy)	データに誤りがなく、信頼性がある状態か
完全性 (Completeness)	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか
一貫性 (Consistency)	データに矛盾はないか、データ間で整合性が取れている状態か
信憑性 (Credibility)	出所や更新日が明示されているか
最新性 (Correctness)	更新期間や最終更新日が明示されているか
アクセシビリティ (Accessibility)	環境依存文字等が使用されていないことが明示されているか
標準適合性 (Compliance)	標準に準拠したデータ形式であることが明示されているか
機密性 (Confidentiality)	アクセス制限等の対策が講じられていることが明示されているか
効率性 (Efficiency)	効率化のため、コードの割り当て等を行っていることが明示されているか
精度 (Precision)	データの精度について明示されているか
追跡可能性 (Traceability)	データの変更履歴（変更者／変更日等）が明示されているか
理解性 (Understandability)	データ全体や項目ごとにメタデータが明示されているか
可用性 (Availability)	利用時間帯等について明示されているか
移植性 (Portability)	標準的な形式でデータをエクスポートできることが明示されているか
回復性 (Recoverability)	バックアップ等の障害時の対応方針が明示されているか

品質特性における競争領域

- 効率性や可用性といった付加的な品質特性は求められる水準が需用者により異なり、基礎的品質特性を超えた「使いやすさ」は競争領域であり、プロセスも複雑になりやすい。
- 標準適合性は、需用者により適合すべき標準の範囲が異なる。

正確性 (Accuracy)	データに誤りがなく、信頼性がある状態か
完全性 (Completeness)	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか
一貫性 (Consistency)	データに矛盾はないか、データ間で整合性が取れている状態か
信憑性 (Credibility)	出所や更新日が明示されているか
最新性 (Correctness)	更新期間や最終更新日が明示されているか
アクセシビリティ (Accessibility)	環境依存文字等が使用されていないことが明示されているか
標準適合性 (Compliance)	標準に準拠したデータ形式であることが明示されているか
機密性 (Confidentiality)	アクセス制限等の対策が講じられていることが明示されているか
効率性 (Efficiency)	効率化のため、コードの割り当て等を行っていることが明示されているか
精度 (Precision)	データの精度について明示されているか
追跡可能性 (Traceability)	データの変更履歴（変更者／変更日等）が明示されているか
理解性 (Understandability)	データ全体や項目ごとにメタデータが明示されているか
可用性 (Availability)	利用時間帯等について明示されているか
移植性 (Portability)	標準的な形式でデータをエクスポートできることが明示されているか
回復性 (Recoverability)	バックアップ等の障害時の対応方針が明示されているか

クレンジング、コード化、標準化等の実際例（一部）

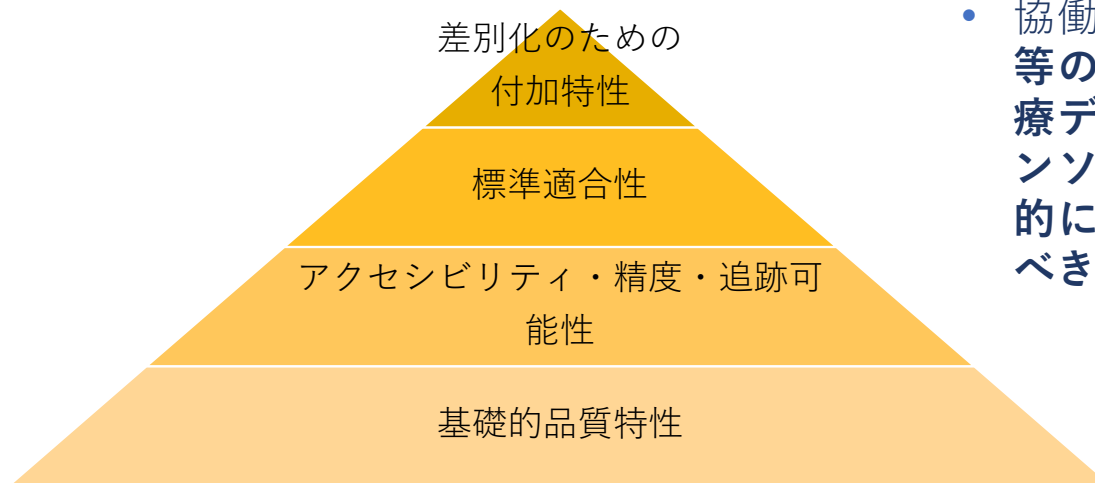


- 取込データの最初のバリデーション（基礎的品質特性、精度の確認等）やクレンジングでは、検査値、投与単位等の学会公式基準等への変換や、いわゆる「その他のX」系データのコード化などを行う。
- **コードマッチしない項目のコード化は、各社が蓄積した辞書や自然言語処理に基づいて行われている（競争領域）。**
- 標準適合性のためのマスタのメンテナンスは、各マスタについて月1回以上であり、**診療報酬情報提供サービスや薬価基準収載医薬品等の公的な情報については、改定の都度マスタを更新し、過去分との統合的な解析を確保している。**
- 医薬品毎の添付文書に含まれる情報等の付加や解析ニーズの高いコードの付加、多言語化等は、個別の需用者のニーズを踏まえ、匿名加工の前後で実施される。
- 全行程の工程数としては、3,000ステップを超えている* ところもある。

* 構成員個人の所属企業のユースケースに基づく

品質確保における官民協調と標準化の考え方*

- 上流でなければ担保できない特性（基礎的品質特性、アクセシビリティ等）は官民で課題感を共有し、商用・学術利用を問わずに求められる標準化を進めるべき。
- 標準適合性のためにマスタのメンテナンス等で民間事業者が相当の資本を投下している知的財産を官民の連携プラットフォームのために利用する必要がある場合は、標準規格必須特許（SEP）におけるFRAND*の考え方を適用し、官民構成員の当該マスタの利用機会の確保と同時に、相当の利益回収の機会が確保されるべき。



- 協働領域・競争領域を含め、学術、製薬企業等の医療等データの二次利用者に加えて、医療データベース事業者を交えた連携基盤（コンソーシアム等）で、品質確保のための商業的に持続可能な仕組みの創出を官民で目指すべき。

*FRAND (Fair, Reasonable, And Non-Discriminatory terms and conditions): ライセンスに際しての金額や条件が公正、合理的かつ非差別的であること

* 構成員個人の私見であり、AMDJ加盟企業の意見を代表するものではない。

二次利用基盤全体に関する技術的論点

【データプライバシーに関する論点】

- 医療等データによる**個人識別性の問題（データプライバシー）**と、本人の社会的実在の特定（いわゆる身バレ）の問題を**区別して議論する必要がある**。
- k匿名性等の統計的な匿名性（データプライバシー）を議論する場合は、**攻撃者の具体的な想定を置かないまま議論をすると統計的な定性的・定量的な評価手法の選択ができない**。
- **標準適合性の確保の前後で匿名性が変わる**（同値のsubject数が変わる）ため、**どの段階で仮名化・匿名化を実施するのか**を議論する必要がある。解析環境に解析者が持ち込んだマスタでも匿名性は変化することにも留意する必要がある。
- **アドホックな解析条件で匿名加工の粒度を可変させるのか、定期的に静的な匿名加工情報のデータベースを構築するのか**（前者の方が幅広のニーズに対応するが、設計が難しい）。

【本人関与に関する論点】

- 母集団への取込へ本人関与について、オプトイン・オプトアウトで技術的な難しさには大差がない。それよりも、**本人関与の機会の有無のフラグ管理をネットワーク全体で確保する必要**があり、次世代医療基盤法のインシデントもここに問題が集中している。
- 死者の情報は個人情報ではないが、生前に本人関与の機会がないと取り込めない制度設計の場合には、**随時死亡情報が連携されない限り、取込可否の判定ができない**（本人関与を必要とする理念的背景が技術的設計に影響する。）。

民間の医療データベースとの連携に関する論点

【匿名加工情報との連携の法的スキーム】

- 次世代医療基盤法上の仮名加工医療情報と異なり、個人情報保護法上の匿名加工情報（行政機関等匿名加工情報を含む。）は、**匿名加工情報取扱事業者に提供された時点で、個人識別可能な情報を含まない。**
- **連結可能とするために被保険者番号を保持した時点で、その提供は個人データ・保有個人情報の第三者提供に当たり、連結は新たな個人情報の取得となるため、個人情報取扱事業者との間で連結可能な状態での提供を行うことは（特別法があっても）困難。**

【連携基盤上での個人データの取扱主体】

- 民間DBと連結したデータの解析者は現行法上は仮名加工情報の取扱いの主体となるため、安全管理措置を確保すべき主体となる。**安全管理措置の内容と責任分界点を適切に設定される必要がある。**

【商用データベースのライセンス管理】

- 連携された民間データベース由来のデータはHIC等の解析基盤上で解析されることとなると思われるが、その場合に**解析者の実際の利用形態**（解析者固有のデータとして解析環境への持ち込みを行いのか、そもそもカタログ収載されるのか等）により、データの出力手法が異なる。

Keywords: 医療等情報, 医療データベース協会, AMDJ, 匿名加工情報, 仮名加工情報, 匿名加工医療情報, 仮名加工医療情報, 行政機関等匿名加工情報, 標準化, 基礎的品質特性, 付加的品質特性, データプライバシー

AMDJ 一般社団法人 医療データベース協会

ディレクター
情報処理安全確保支援士・弁護士・弁理士
足立 昌聰

本資料に関するご質問等: [madachi\[アット\]jmdc.co.jp](mailto:madachi@jmdc.co.jp)

2024.3 医療等情報の二次利用に関するワーキンググループ 技術作業部会資料