



T-20-01
認定匿名加工
医療情報
作成事業者

参考資料2

論点整理に関する 認定事業者からの情報提供



Japan Medical Association Medical Information Management Organization

令和4年4月20日

一般財団法人 日本医師会医療情報管理機構

1-1. 利活用を推進する観点からの匿名加工基準の在り方

■ 希少疾患、希少薬について

- J-MIMOでは、疾患名、薬剤名について、提供対象テーブルに含まれる件数が3未満 ($k < 3$) のものを「その他」として秘匿しています。全国数十万人規模のDBにおいて、かなりの希少疾患、希少薬でも3件以上は存在しており、利活用者の皆様が危惧されるほど情報量は損なわれていません。
- $K < 3$ の疾患や薬剤に着目した研究開発がどうしても必要な場合も、相談に応じることが可能です。その場合には、注目する疾患、薬剤との関連が少ない他の項目の粒度を粗くするなど、加工の方法を工夫します。
- ただし、治験薬については、そのコードから企業名や患者IDが類推できる場合があるため、網羅的に「その他」として秘匿する方法を検討中です。

■ 一般化について

- 医療機関名、住所、生年月日等について、それぞれ地方ブロック、市区町村、生年月等に一般化しています。
- 検査値等も、要望に応じて区間分割を行います。要望がなければ実数値のまま提供します。
- 医療機関名については、次世代医療基盤法ガイドラインではなく、提供元医療機関の要望により秘匿しています。

■ トップ・ボトムコーディングについて

- 生年月日、診療日等の日付について、古い日付の下位3件～10件程度をボトムコーディング（特に小さい値をまとめること）しています。
- 身長等の静的属性（変化しにくい項目）について、上位3件程度をトップコーディング（特に大きい値をまとめること）しています。

1-1. 利活用を推進する観点からの匿名加工基準の在り方

■ ノイズ付加について

- 生年月日、診療日等の日付について、±数日の範囲で患者ごとに一律のノイズ付加（乱数を付加して値をずらすこと）を行います。患者ごとに見れば日付が平行移動することとなり、日付の間隔は不変です。
- ノイズ付加については、入院日等による再識別リスクを抑制するための措置ですが、ノイズを付加した結果、月がずれてしまうことによる矛盾や、上市前に処方されているように見えるなど、課題も多いと考えています。
- ノイズ付加による再識別リスクの低下と、データの正確性の低下とのバランスについて、一認定事業者だけでは判断しかねると考えており、WGでの議論をお願いします。

■ 仮IDについて

- J-MIMOでは、提供する度に異なる仮IDを付与し、シード/ソルト値（ランダムな仮IDを生成する際の元になる値）についても、納品・検収後一定期間を経過したら削除することとしています。
- このため、認定事業者自身でも、匿名加工された情報を再識別することはできません。
- 現在、同一の匿名加工医療情報取扱事業者に繰り返しデータを提供する場合でも、毎回全データについて匿名加工処理と仮IDの再生成を実施しています。
- しかしながら、データ件数が1億件を超えると、毎回全データに対して匿名加工を実施することは、システムリソースの逼迫やメモリ障害の原因にもなります。J-MIMOでも、1億件を超えた頃からメモリ障害の件数も増えています。
- 第4回次世代医療基盤法検討WGでも議論されたように、次世代医療基盤法に基づく匿名加工は、個人情報保護法に基づく匿名加工より厳格に管理されています。このため、認定事業者が設置する審査委員会が承認した場合には、同一の匿名加工医療情報取扱事業者に対して、同一の仮IDでの差分データの提供を認めるなど、ガイドラインの緩和をお願いします。

1-1. 利活用を推進する観点からの匿名加工基準の在り方

■ テキスト情報について

- J-MIMOでは電子カルテSS-MIX2標準ストレージ（電子カルテの厚生労働省標準規格）を取得していますが、SS-MIX2標準ストレージの中にも、医療機関独自のコード（管理番号）とフリーテキストが入っている場合があります。
- このテキストの中に、医療従事者名や医療機関名、日付が混入しており、クレンジングに多大な労力を要しています。
- 現在、3 σ （99.5%）程度の精度でエラーを検出できていますが、100.00%の精度を技術的に保証することは不可能です。
- このため、意図しない情報がデータに混入していた場合を想定し、認定事業者と匿名加工医療情報取扱事業者との機密保持に加え、仕様と異なるデータを発見した場合、直ちに報告し、対応を協議する旨を契約書で規定しています。

■ マスタについて

- 現在、J-MIMOでは、有償の商用マスタ（施設、薬剤等）を購入するだけの資金的余裕がなく、ノーマスタでの運用を強いられています。
- 認定事業者が使用するマスタについては、社会インフラとして無償で使用できる環境を国が整備して頂きたいと思います。

■ クレンジングについて

- リアルワールドデータの特徴として、入力ミスや欠損はごく普通に発生します。
- 極端に古い日付（例：90年前）や3桁分大きいまたは小さい値（単位のミス?）、欠損に由来する0と正しい計測値としての0の峻別等、データクレンジング（入力ミス等の修正作業）については、J-MIMOはまだほとんど対応できていないのが現状です。
- 利活用者の皆様には、クレンジングが終わるまで利用を控えるのではなく、クレンジングの体制づくりに積極的にご協力頂きたいと思います。

1-1. 利活用を推進する観点からの匿名加工基準の在り方

■ 仮名加工について

- 仮名加工情報に対する過剰な期待や誤解が広がることを危惧しています。
- 仮名加工情報の第三者提供は禁止であること、仮名加工情報の作成主体であっても再識別は禁止であることについて、誤解が広がらないよう、周知をお願いします。
- 仮名加工情報の共同利用についても、個人情報保護法ガイドライン通則編3-6-3(3)において、共同利用とは、「社会通念上、共同して利用する者の範囲や利用目的等が」「本人が通常予期しうると」「客観的に」認められる範囲内であり、事例はグループ企業・親子兄弟会社と労使関係に限られています。
- 現在のところ、本人の治療に必要な地域医療連携でも、共同利用は認められておらず、開示側で黙示の同意、参照側で口頭同意が必要とされています。本人の治療に必要な活動に共同利用が認められていないにも関わらず、本人の治療に不要な活動に共同利用が認められることは、本人の権利利益の保護の観点から不当ではないでしょうか。
- 仮名加工の共同利用について、その範囲を厳格に限定すべきと考えます。
- 一方、安全管理措置として識別子や識別符号を削除し、その処理を委託するケースは産官学を問わず広く行われています。データの実態としては仮名加工情報によく似ていますが、これを「仮名加工情報」と呼んでしまうと、委託元での再識別行為が法違反となってしまいます。安全管理措置として識別子や識別符号を削除したデータを、私達は何と呼べばよいのでしょうか？ 個人関連情報として位置付けることも可能とは思いますが、正しい理解に向けた周知をお願いします。

2-3. 名寄せが可能である次世代法DBの長所を発揮した、NDBなどの公的DBや既存の民間DBとの連携

- NDB, MID-NET, マイナポータル, LIFE, KDB等、国又は公的機関が整備するDBを認定事業者のDBと連結可能として頂きたいと思います。
- また、学会や大学が整備する疾患レジストリや疫学研究DBを認定事業者のDBに統合し、認定事業者にDBの運営を委託することで、研究者は煩雑なDBの管理から解放され、本来の研究に専念することが可能となります。この際の移行・統合に要する費用を文部科学省やAMED等でご検討頂きたいと思います。
- 民間DBについては、まず個人情報保護法の遵守状況について丁寧な監査が必要と考えます。個人情報保護法の遵守状況を確認できた段階で、認定事業者のDBとの連結に大きな価値が生じます。