資料８

山本構成員提出資料

匿名化に関する考察

2012年6月20日

東京大学大学院情報学環　山本隆一

匿名化すれば個人情報ではない
では匿名化とは何か？

- 医療・介護事業者における個人情報の適切な取り扱いのためのガイドライン
  当該個人情報から、当該情報に含まれる氏名、生年月日、住所等、個人を識別する情報を取り除くことで、特定の個人を識別できないようにすることをいう。

- 疫学研究に関する倫理指針
  個人情報から個人を識別することができる情報の全部又は一部を取り除き、代わりにその人と関わりのない符号又は番号を付すことをいう。資料に付随する情報のうち、ある情報だけでは特定の人を識別できない情報であっても、各種の名簿等の他で入手できる情報と組み合わせることにより、その人を識別できる場合には、組合せに必要な情報の全部又は一部を取り除いて、その人が識別できないようにすることをいう。

- 匿名データの作成・提供に係るガイドライン（統計法）
  匿名化の基準：
  調査票情報の特性は統計調査ごとに異なることから、各統計調査について一律に匿名化の基準を設定することは困難である。

匿名化すれば個人情報ではない
では匿名化とは何か？

- HIPAA Privacy rule (U.S.) § 164.514
  次のいずれかの場合、個人が識別できないとして良い。

  (1)一般的に受け入れられ、統計的かつ科学的な原則及び方法に関して適切な知識及び経験を持った者が、情報の利用を行うものが、単独または合理的に入手可能な情報と照らし合わせることで、個人が特定されるリスクを評価し、リスクが十分低いことを判断した分析の経過および結果を文書化した場合。

  (2)以下のあげる19項目を本人、親類、雇用者、世帯員に関する情報から除いた場合。
  名前、州より小さい範囲の住所、2万人以下に限定される郵便番号の上位3桁(2万人以下に限定される場合は000に)、すべての日付(年は除く)と89歳以上の年齢、電話番号、FAX番号、電子メールアドレス、SSN、カルテ番号、保険番号、口座番号、免許証番号、車両番号およびシリアル番号、装置の識別番号およびシリアル番号、URLs、IPアドレス、生体識別情報(指紋等)、顔写真を含む識別可能な写真、連結可能または不可能匿名化のために付与した番号を除く固有の数字・特徴またはコード

利活用試行例　レセプト等情報データベース(NRDB)

- ○氏名○生年月日の「日」○保険医療機関の所在地及び名称○カルテ番号等○国民健康保険一部負担金減額、免除、徴収猶予証明書の証明書番号○被保険者証（手帳）等の記号・番号○公費受給者番号は削除、保険医療機関番号は保持
- 保険者番号・記号番号・生年月日・性別からハッシュ値①を生成、氏名・生年月日・性別からハッシュ値②を生成して格納している。
- 疫学研究指針で言う連結不可能匿名化とは考えていない。
- 公益性の確認、目的外利用の禁止、地域情報、医療機関情報の原則提供禁止、安全管理の確認、公表形式の審査（患者の場合は10人未満に特定されてはいけない、等）

匿名化と仮名化 (Anonymization and Pseudonymization)

- 仮名化：個人を識別できる情報を元に戻れない変換(一方向変換)で、一意の識別子に置き換えた状態。対応表が存在し、それを使うことで、個人の情報に戻ることができる状態。

  対応表が安全に管理されていれば、個人が識別される恐れは低い。

  連結可能匿名化と同じ意味で、対応表を破棄すれば、あるいは初めから対応表を作らなければ連結不可能匿名化と呼ぶ。

  ISO/TS25374 Health Informatics – Pseudonymization (2008)

  連結不可能匿名化は一般に言う匿名化と同じ意味。
  つまり匿名化とは何かという問は残る。

医療・介護・福祉情報における匿名化（二次利用）

- 多くの場合、HIPAA Privacy Rule §164.514の(2)番目の規則のように個人の識別につながる項目を明確にした上で、除去すれば問題はない。
- しかし・・・・
- 精度が落ちすぎて公益性の高い調査さえできないこともある。（日付、地理情報・・・）
  リスクを十分下げれば利用可能では？
- 稀少疾患、頻度が低い医療行為、少数しか用いられない薬品など、個人の識別につながる可能性を排除できない。
  特に身近な人の場合、関連情報が入手しやすくリスクが増える。
- リスクの評価が必要な場合がある。（HIPAA Privacy Rule §164.514（1））
  完全匿名化が自明でない場合はリスク評価を行うべきか。

匿名化のリスク評価を導入すべきではないか

- 最悪の場合、何人に特定されるか？

  最小特定人数（BIN） L. Sweeney, 1998, Medinfo Proceeding
  K-匿名性（k-anonymity） L. Sweeney, 2002
  もっとも多く用いられており、ツールも豊富。
  ただし計算量は比較的大きい。

- その項目はどの程度多様か？

  l-多様性（l-diversity） A. Machanavajjhala et.al., 2006

- 分布に著しい偏りがないか？

  t-近似性（t-closeness） N. Li et.al., 2007

リスクが0ではない匿名化は真の匿名化か？

- リスクは目的に応じて評価される。
  公益性
  緊急性
  ・・・・

- 目的外利用はリスクの再評価が必要
  第三者提供も制限されるべき
  ↓
  開示や訂正要求への対応は別途考える必要があるが
  個人情報保護法制の対象に含めるべきではないか。

リスク評価の審査はどこで行うべきか？

- 番号法で言う第三者委員会？

  相当の審査数があり、対応可能か？

- 倫理委員会？

  現状で審査可能か？

  健康被害の可能性が、まずは起こらない一方で、
  プライバシー侵害に関する可能性の審査が中心、
  匿名性のリスク評価も含まれる。