

2017 年度課題研究会活動成果報告書

課題研究会名：匿名加工に求められる技術要件の検討会

設置期間：2017 年 7 月 1 日～2020 年 3 月 30 日

代表幹事の氏名・所属：木村 映善 国立保健医療科学院

幹事の氏名・所属：

山本 隆一	一般財団法人医療情報システム開発センター
黒田 知宏	京都大学大学院医学系研究科
菊池 浩明	明治大学総合数理学部
高橋 克巳	NTT セキュアプラットフォーム研究所
美馬 正司	日立コンサルティング
工藤 憲一	日本医師会 ORCA 管理機構（株）

活動成果の概要：

第 1 回は 8 月 17 日に一般財団法人医療情報システム開発センターにて、第 2 回から第 5 回は、9 月 13 日、10 月 5 日、11 月 1 日、15 日に日本医師会 ORCA 管理機構にて、延べ 5 回開催した。第 1 回目は課題研究会発起に関して、コアメンバーによる幹事によって、当時進行していた次世代医療基盤法にもとづく認定事業者のガイドライン、匿名加工プロセス WG、そして本課題研究会の役割分担の確認を行い、本課題研究会は匿名加工処理に関する技術およびリスク評価について学術的に検討する場であることを確認した。第 2 回以降は、具体的な匿名加工に関する検討を進めた。

第 2 回 準識別子の種類と評価方法

個人情報保護法、個人情報保護委員会における個人情報と匿名化にかんする定義をふまえて、「特定」「識別」の定義の確認を行った。個人にかんする情報として、識別子、準識別子、静的属性、半静的属性の定義について確認し、識別子区分の検討では、項目の意味合いに加え、連続・離散値等の特性について配慮が必要であることを確認した。匿名化の安全性の評価について様々な方法が考えられるが、もっとも基本的な評価法として、準識別子を用いた k 匿名性の評価をベースとし、その上で、求める要件にあわせて個別にリスクを考慮していくこととした。データの利活用性に配慮するため、準識別子に指定することは限定的に進めていく一方、準識別子になった項目は

k 匿名性の概念を用いて厳しく保護する方針とし、またリスク考慮においては、提供先基準において評価・検討することとした。これらの議論を踏まえて、健診・糖尿病症例のデータの項目定義を関係者で確認して、識別子、準識別子、静的属性、半静的属性の分類をトライアルした。

第3回 匿名加工に関するユースケースの検討

特定健診受診者をコホートに組み入れるケース、集計表ベースでの分析例、特定健診の microdata を用いた研究「Anonymisation: managing data protection risk code of practice ico. (England)」からのシナリオを紹介・検討した。匿名加工を検討するにあたり、「再識別」「再特定」の用語の整理、一般人原理において、攻撃の意図・契約を遵守しないものは一般人とはみなさないこと、また一般人であっても、NII レポートにあるように、住所、氏名、性別、生年月日のある名簿程度の情報は入手可能であり、それにもとづく照合可能性を配慮する必要性について確認した。医療機関については準識別子、あるいは静的属性として扱うべきかの議論を通して、分析シナリオや提供対象となるデータに応じて提供元の委員会が個別検討をする必要性があること、またその個別検討の成果を共有する仕組み、機構の必要性について議論した。

第4回 特異な値の処理

特異な値の定義について、top/bottom coding に関する論文を紹介した。国によって人口の構成・分布が異なっており、top/bottom coding の閾値を母集団の特性、照合可能性を踏まえて弾力的に設定することの必要性を確認した。

第5回 匿名加工ツールや匿名加工処理フレームワークの紹介

NTT セキュアプラットフォーム研究所が開発した匿名加工ツールの機能として、追加、削除、変換、トップ・ボトムコーディング、一般化、ランダム化、ノイズ付与、外れ値加工、有用性評価、匿名性指標等が紹介された。また、幹事より匿名加工処理プロセスの評価フレームワークについて紹介があり、今後匿名加工処理を通してリスク評価の事例蓄積、評価のプロセスを進めたい旨を確認した。

[学会発表] 計(1)件

木村映善, 藤村明子, 美馬正司, 菊池浩明, 千田浩司, 佐久間淳. パーソナルデータの利活用のための情報処理技術の方向性について. 医療情報学 37(Suppl). 2017:210-1.