SCIS 2022 2022 Symposium on
Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 – 21, 2022
The Institute of Electronics,
Information and Communication Engineers

パーソナルデータの等結合に適した匿名化技術の考察
On Anonymization for Equi-Join of Personal Data

千田 浩司 * 紀伊 真昇 * 市川 敦謙 * 野澤 一真 †
Koji Chida Masanobu Kii Atsunori Ichikawa Kazuma Nozawa
長谷川 慶太 † 堂面 拓也 † 中川 智尋 † 青野 博 †
Keita Hasegawa Takuya Domen Tomohiro Nakagawa Hiroshi Aono
寺田 雅之 †
Masayuki Terada

キーワード 匿名化, プライバシ, 等結合, パーソナルデータ

あらまし

各組織が保有するパーソナルデータを等結合 (連結) して組織横断で利活用し，社会課題の解決やサービスの質向上に資する機運が高まってきている．しかし国内外の現行法制度の下では本人同意なく等結合を行うことは難しく（ただし次世代医療基盤法 [1] や，個人情報保護法における仮名加工情報 [2] 等の例外はある），パーソナルデータの組織横断的な利活用を促進する手段が望まれる．そこで本稿では，匿名性を保ちつつ等結合できる技術的アプローチについて，いくつかの手法を例示し特徴を考察する．

等結合とは，複数の表形式データについてそれぞれ列を指定し (ID 列)、ID 列の値が等しいレコードを連結する処理である．ID 列以外の値をどれだけ個人識別困難となるよう加工しても，ID 列の値から個人を識別できるリスクが残る．各組織が保有する表形式データを等結合する場合，基本的には ID 列を相手の組織に開示する必要があるため，このリスクは無視できない．そこで ID 列を除く個々の表形式データは個人識別困難となるよう加工することを前提として，ID 列の値を不可逆変換する方法 (仮名化) が考えられるが，相手の組織は基本的に同様の変換が可能となるため，ID 列の値を推定し得るという問題が残る．さらには，等結合したデータが個人識別困難かどうかという問題も生じる．

我々は上記の課題に対し，二者間で匿名性を保ちつつ等結合できる技術的アプローチについて考察する．具体的には，

1. セキュア 2 パーティ計算 ([3] 等) と匿名化を組み合わせて用いる，
2. 等結合後も個人識別困難となるよう各者が自身の表形式データを匿名化する，

の 2 通りのアプローチについて，特徴比較や実現に向けた課題整理を行う．

* 日本電信電話株式会社 NTT 社会情報研究所, 東京都武蔵野市緑町 3-9-11, NTT Social Informatics Laboratories, Nippon Telegraph and Telephone Corporation, 3-9-11, Midori-cho Musashino-shi, Tokyo
† 株式会社 NTT ドコモ クロステック開発部, 東京都千代田区永田町 2-11-1, X-Tech Development Department, NTT DOCOMO, Inc., 2-11-1, Chiyoda-ku, Tokyo

参考文献

[1] 内閣府健康・医療戦略推進事務局, 「次世代医療基盤法」とは, 2021 年 9 月, https://www8.cao.go.jp/iryou/gaiyou/pdf/seidonogaiyou1.pdf (2021 年 12 月 9 日参照).

[2] 第 181 回 個人情報保護委員会, 資料 1-8 個人情報の保護に関する法律についてのガイドライン（仮名加工情報・匿名加工情報編）の一部を改正する告示案, 2021 年 8 月 4 日, https://www.ppc.go.jp/files/pdf/210804_shiryou-1-8.pdf (2021 年 12 月 9 日参照).

[3] Yao, A.C.,“Protocols for secure computations,” 23rd Annual Symposium on Foundations of Computer Science (FOCS ’82), pp.160–164.