

# 大規模データの匿名加工情報における課題

菊池 浩明  
Hiroaki Kikuchi

明治大学, Meiji University

## 1 はじめに

2017年5月に全面施行した改正個人情報保護法や行政機関個人情報保護法等改正法によって、民間や行政機関や保有するパーソナルデータに対する匿名加工情報(非識別加工情報)が本格的に活用されようとしている。匿名加工情報とは、「(定められた措置を講じて)特定の個人を識別することができないよう個人情報を加工して得られる個人に関する情報であって、当該個人情報を復元することができないようにしたもの(第二条9項)」と定められている。匿名加工情報の作成の方法に関する基準(第十九条)は、

1. 個人を識別する記述等の全部又は一部を削除する。
2. 個人識別符号の全部を削除する。
3. .. 個人情報を連結する符号を削除する。
4. 特異な記述等を削除する。
5. ... 他の個人情報に含まれる記述等の差異その他の当該個人情報データベース等の性質を勘案して、... 適切な措置を講じること。

と提案されている。業種ごとに定められた認定個人情報保護団体が、消費者や関係者の意見を聴いて、その対象事業に適した個人情報保護指針を整備していくことになっている。しかしながら、再識別のリスクの評価や加工の方法には、高度な専門性やその分野に特化した背景知識の両方が必要であり、誰に対しても納得できる信頼できる基準には至っていない。とりわけ、大規模で長期間の履歴情報から生じるレコード再識別のリスクについては危険性が認識されながらも、その正確な評価が困難である。それゆえ、これらの情報の活用を計画している事業者も、長期間にわたった履歴データの匿名加工については躊躇しているのが現状である。

そこで、英国に現存する無店舗型オンラインショッピングサイトにおける2010年からの1年間の購買履歴であるOnline Retail Data Setを用いて、匿名加工の基本定義と有用性と安全性の評価指標を検討する。考慮した不正行為と対象としたリスクについて述べ、サンプルとするいくつかの再識別アルゴリズムを仮定する。