PWSCUP2017
個人情報保護委員会規則第19条の解釈と考察

M-OND-A

東京大学大学院中川研究室

門田 将徳

PWSCUPの意義

急速な情報通信技術の発達

より便利な生活への**期待**

個人情報の漏洩**リスク**の増加

- 個人情報保護法の改正[1] ・「**匿名加工情報**」が定義

匿名加工・再識別コンテストPWSCUP
- 共通のデータセットに対して，匿名加工技術の客観的な評価を行う
- 匿名加工情報の有効活用に向けた課題について考察

- 本人同意がなくとも第三者提供ができる
  → 有効に活用すれば，**安全な産業・サービスの発展が期待**できる
- しかし，具体的かつ統一的な作成手法に関する検討は進んでいない

共通データセット

共通データセットは，英国に実存するオンラインショッピングサイトにおける購買履歴データをもとに，PWSCUP実行委員会によって作成された500人の顧客に関するデータである．
PWSCUP2017では，**トランザクションデータのみが匿名加工対象**である．

	UserID	Sex	Birthday	Nation
1	12993	m	1970/01/01	United Kingdom
2	14243	f	1960/01/01	United Kingdom
3	14315	f	1950/01/01	Germany
4	13098	f	1970/01/01	United Kingdom
5	12748	m	1970/01/01	United Kingdom
6	12412	f	1960/01/01	Others

マスターデータM

	UserID	InvoiceID	Date	Time	ItemID	Price	Quantity
1	12993	0	2011/08/11	12:01	22561	1.65	12
2	14243	0	2010/12/11	15:21	23285	0.85	10
3	14243	0	2011/06/16	14:22	23285	0.85	8
4	14315	0	2011/03/18	12:59	48129	7.95	2
5	13098	0	2010/12/15	10:38	23049	8.25	6
6	13098	0	2011/01/06	08:54	23285	0.85	24
7	13098	0	2011/11/17	22:45	23049	8.25	6
8	12748	0	2011/09/28	15:21	47504H	0.79	1

トランザクションデータT

匿名加工

再識別フェーズ

匿名加工フェーズ

UserID	12	1	3	6	8	9	11
12993	DEL	DEL	DEL	DEL	A1	DEL	DEL
14243	B1	DEL	DEL	DEL	DEL	DEL	DEL
14315	DEL	DEL	C1	DEL	DEL	DEL	DEL
13098	DEL	D1	DEL	DEL	DEL	DEL	D2
12748	DEL	DEL	DEL	DEL	DEL	E1	DEL

推定仮名表$\hat{F}$

再識別攻撃

	UserID	InvoiceID	Date	Time	ItemID	Price	Quantity
4	C1	0	2011/03/17	0:00	48129	7.95	4
1	A1	0	2011/08/01	0:00	22534	1.60	11
7	D2	0	2011/11/18	0:00	23049	8.00	5
2	B1	0	2010/12/14	0:00	23285	0.85	9
8	E1	0	2011/09/30	0:00	47504H	0.79	2
6	D1	0	2011/01/06	0:00	23285	0.85	23

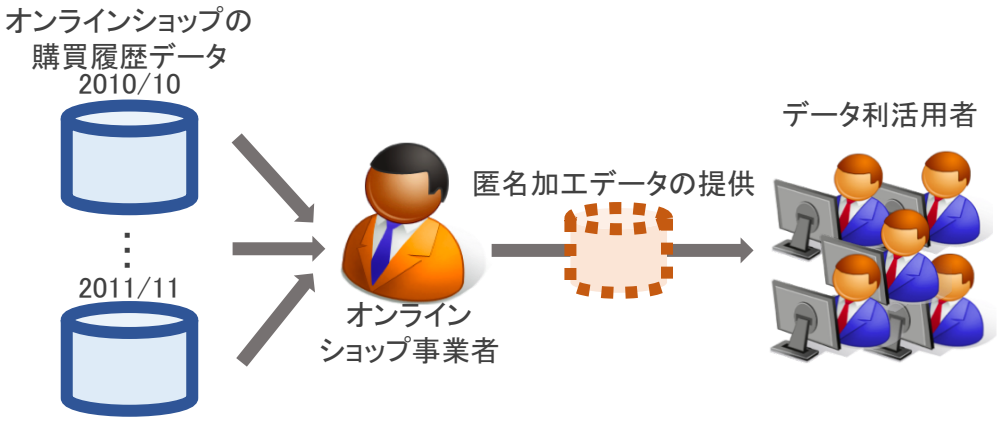
匿名加工データS

匿名加工フェーズ

参加者は，共通データセットに対して匿名加工を施し，
作成された匿名加工データを評価システムに提出する．

□ 想定されたユースケース

- オンラインショップ事業者は，収集されたデータに対して，**月ごとに匿名加工および有用性の評価**を行い，匿名加工データとしてデータ利活用者に提供
- データ利活用者は，匿名加工データの利用目的や必要とする有用性，もしくはそれを計測するための有用性指標を，匿名加工情報の作成条件として，事前にオンラインショップ事業者に伝えているものとする

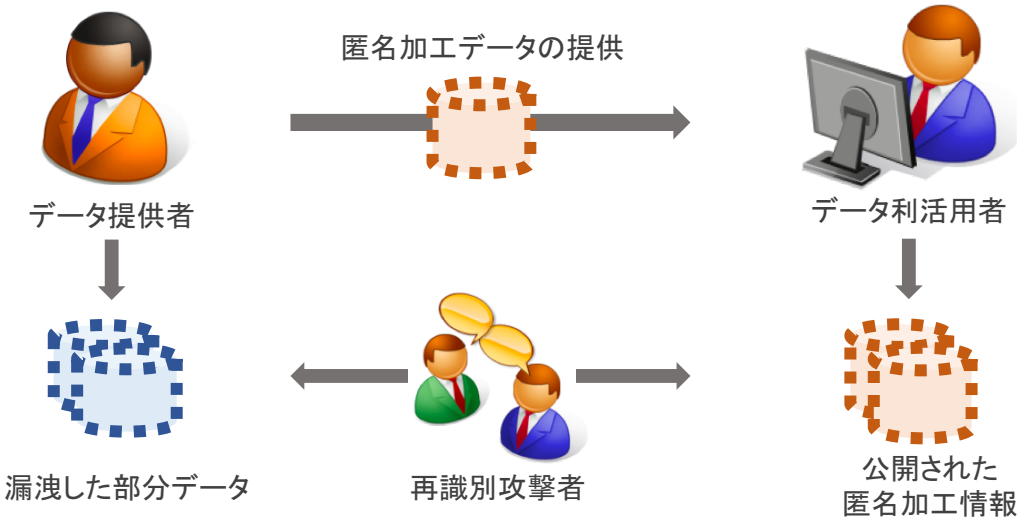


再識別フェーズ

参加者は，他の参加者が作成した匿名加工データに対して，再識別攻撃を行い，
推定仮名表を評価システムに提出する．

□ 想定されたユースケース

- オリジナルデータを所有するオンラインショップ事業者が，ミスやハッキングなどの理由により，**オリジナルデータの一部を流出**させてしまったという状況
- 再識別攻撃者は，匿名加工データに含まれる**仮名を**，
  オリジナルデータに含まれる**UserIDと接続**すること（再識別攻撃）が目的である



再識別の定義

推定仮名表によって，全ての月の仮名を当てられた場合，
その顧客が再識別されたとする

⇒ 1つでも正しく当てられなかった月があると，再識別とはみなされない

□仮名表

顧客	12	1	2	3	4	5	6	7	8	9	10	11
A	A1	DEL	DEL	A1	DEL	A2	DEL	A2	DEL	A2	A2	DEL
B	B1	DEL	B1	B1	B2	B3	DEL	DEL	DEL	B3	B3	DEL
C	DEL	DEL	C1	DEL	C1	C1	DEL	DEL	DEL	C2	DEL	DEL

□推定仮名表

顧客	12	1	2	3	4	5	6	7	8	9	10	11	
A	A1	DEL	DEL	A1	DEL	A2	DEL	A2	DEL	A2	A2	DEL	→ 再識別成功
B	B1	DEL	F2	B1	B2	B3	DEL	DEL	DEL	B3	B3	DEL	→ 再識別失敗
C	DEL	DEL	DEL	DEL	C1	C1	DEL	DEL	DEL	C2	DEL	G5	→ 再識別失敗

匿名加工情報の作成例(1)

一般財団法人日本情報経済社会推進協会(以下，JIPDEC)は，匿名加工情報の事例集[6]を公開している．事例3では購買履歴データの匿名加工事例の紹介を行っている．

◘ データセット

- 顧客情報（マスターデータ）および，購買履歴（トランザクションデータ）
- 会員番号（連結符号）で紐づけられている

氏名	会員番号	生年月日	性別	住所
甲野太郎	000001	1963/1/1	男	S 区 AA 町 4 丁目 19 番地 1 号
乙山次郎	000002	1974/2/2	男	T 区 BB 町 1 丁目 3 番地 2 号
丙田花子	000003	1952/3/3	女	S 区 CC 町 2 丁目 5 番地 3 号

会員番号	購入店	日時	購入総額	付与ポイント
000001	D 魚店	2016/12/5 11:40	300 円	3
000001	E 豆腐店	2016/12/5 11:35	130 円	1
000001	F 傘屋	2016/12/6 14:15	840 円	84

◘ ユースケース

- 商店街が，新規出店を検討する事業者に対して，匿名加工情報を公開する
- 商店街が保有するマスターデータとトランザクションデータを，匿名加工情報の形で提供することで，当該事業者はその情報を分析し，出店するべきかどうかを判断する
- 当該事業者は，性別，年代，購入時間帯，利用店種数，購入金額をもとに購買履歴を分析

匿名加工情報の作成例(2)

JIPDECでは，マスターデータ，トランザクションデータに対して以下の加工を行っている．

年代	性別	住所
50 代	男	S 区 AA 町 4 丁目
40 代	男	T 区 BB 町 1 丁目
60 代	女	S 区 CC 町 2 丁目

購入店	日時	購入総額
食品	月曜日 11:00-12:00	201-300 円
食品	月曜日 11:00-12:00	101-200 円
服飾品	火曜日 14:00-15:00	700-800 円

属性	再識別リスク	措置	対応号
名前	個人の特定	全部消去	第1号
会員番号	連結符号	全部消去	第1,3号
生年月日	組み合わせによる個人の特定	年代への丸め	第1号
性別	組み合わせによる個人の特定	無加工	—
住所	組み合わせによる個人の特定	丁目以下を消去	第1号
購買店	**提供先において不要な情報**	店種へ置換	第5号
購入日時	組み合わせによる個人の特定	曜日，1時間単位に置換	第5号
購入総額	組み合わせによる個人の特定	100円単位に置換	第5号
付与ポイント	**提供先において不要な情報**	全部消去	—

個人情報保護委員会規則第19条

個人情報保護委員会規則第19条[2]

法第36条第１項の個人情報保護委員会規則で定める基準は、次のとおりとする。

一. 個人情報に含まれる特定の**個人を識別することができる記述**等の全部又は一部を削除すること（当該全部又は一部の記述等を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

二. 個人情報に含まれる**個人識別符号**の全部を削除すること（当該個人識別符号を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

三. 個人情報と当該個人情報に措置を講じて得られる情報とを**連結する符号**（現に個人情報取扱事業者において取り扱う情報を相互に連結する符号に限る。）を削除すること（当該符号を復元することのできる規則性を有しない方法により当該個人情報と当該個人情報に措置を講じて得られる情報を連結することができない符号に置き換えることを含む。）。

四. **特異な記述**等を削除すること（当該特異な記述等を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

五. 前各号に掲げる措置のほか、個人情報に含まれる記述等と当該個人情報を含む個人情報データベース等を構成する**他の個人情報に含まれる記述等と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、その結果を踏まえて適切な措置を講ずること。

個人情報保護委員会規則19条1号

個人情報に含まれる特定の個人を識別することができる記述等の全部又は一部を削除すること（当該全部又は一部の記述等を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

チームの解釈
- 単体で個人を識別できる記述
  → 氏名
- 組合せによって個人を識別できる記述
  → 住所，生年月日，電話番号，性別，クレジットカード番号など

加工手法
- トランザクションデータに上記の記述なし
- (CUP'17では加工対象ではないが，)マスターデータについても，UserID以外の属性値の組合せで，既に2-匿名化を満たしているため，加工の必要なし

個人情報保護委員会規則19条2号

個人情報に含まれる個人識別符号の全部を削除すること（当該個人識別符号を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

加工手法

- トランザクションデータ，マスターデータ共に，記述なし

個人情報保護委員会規則19条3号

個人情報と当該個人情報に措置を講じて得られる情報とを連結する符号（現に個人情報取扱事業者において取り扱う情報を相互に連結する符号に限る。）を削除すること（当該符号を復元することのできる規則性を有しない方法により当該個人情報と当該個人情報に措置を講じて得られる情報を連結することができない符号に置き換えることを含む。）。

チームの解釈

- UserIDは，上記記述に該当する
  （マスターデータとトランザクションデータを紐づけている）

加工手法

1. 規則第19条4号5号の解釈を元に，各顧客について
   分割するべき月と消去するべきレコードを決定
2. 分割する期間ごとに，UserIDに適当な文字列を追加
   例）UserID “12345” を，3つの期間に分割する場合，各期間内のUserIDを，
   “12345AAA”，“12345BBB”，“12345CCC”などとする
3. 消去する必要のあるレコードはUserIDを”DEL”に変更
4. “DEL”を除く全てのUserIDをハッシュ関数に投入
   （pythonのhashlib.md5を用いて128bitの適当な値に変換）

個人情報保護委員会規則19条4号

特異な記述等を削除すること（当該特異な記述等を復元することのできる規則性を有しない方法により他の記述等に置き換えることを含む。）。

チームの解釈

- レコードに含まれる1つ1つの値のうち，特異と考えられる記述と解釈した
  例）1人にしか購入されていない特殊な商品
- 個人情報保護法ガイドライン(匿名加工情報編)[3]では，第4号が指す特異性とは「一般的なあらゆる場面において**特異であると社会通念上認められる記述**」とされている
- 共通データセットでは，商品がItemIDのみで管理されており，ItemIDのみから各商品の「社会通念上認められうる特異性」を判断することはできないため，ItemIDは該当しないと判断した

加工手法

- 該当する記述なし
- 4号に該当する記述の有無は，元データからのサンプリングの仕方にも依存する

個人情報保護委員会規則19条5号

前各号に掲げる措置のほか、個人情報に含まれる記述等と当該個人情報を含む個人情報データベース等を構成する他の個人情報に含まれる記述等と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、その結果を踏まえて適切な措置を講ずること。

チームの解釈

- それぞれの顧客のレコード集合に現れる特異性と解釈した
  例）購入月パターン，ある月の購入アイテム集合など
- 個人情報保護法ガイドライン(匿名加工情報編)[3]では，第5号の解釈と購買履歴の加工について，「蓄積されたこと等によって特定の個人の識別又は元の個人情報の復元につながるおそれがある部分については，適切な加工を行わなければならない」とされている

加工手法

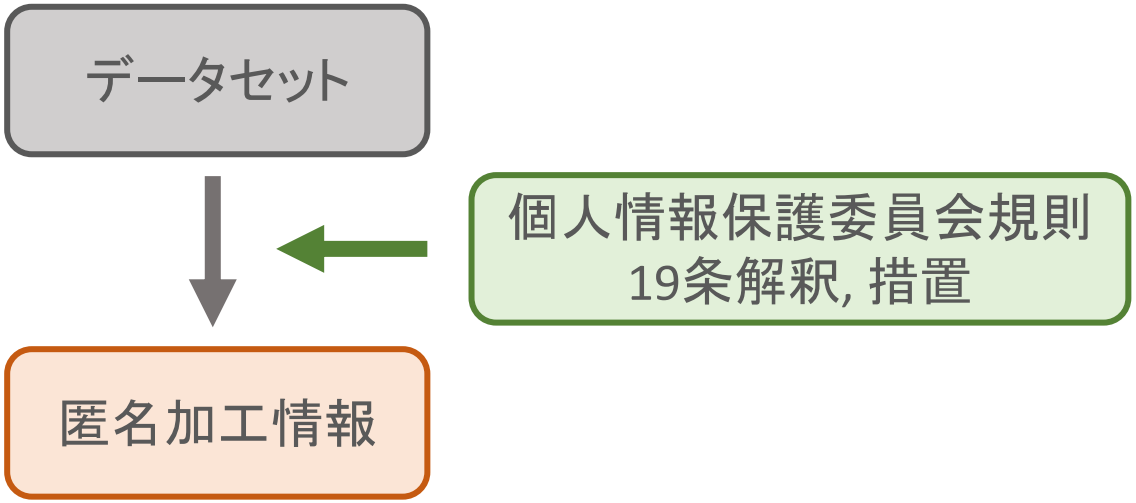
- 購入月パターンだけで1人の顧客に特定される顧客は500人中154人存在する
- それらの顧客は分割もしくは，ある月のレコードを丸ごと消去した
- 有用性指標に与える影響が最も小さい月を分割する

個人情報保護委員会規則19条5号

この文字列がユニークな顧客については，分割もしくは月ごとの消去が必要

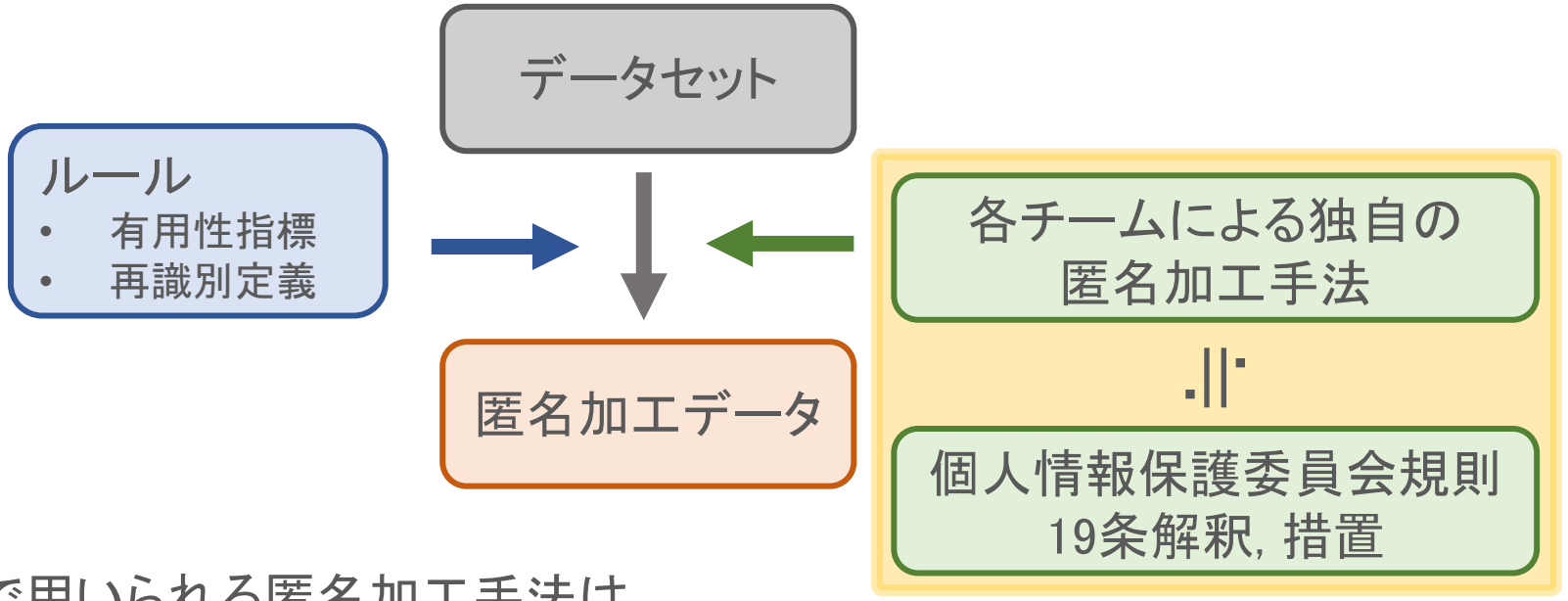
顧客ID	12	1	2	3	4	5	6	7	8	9	10	11	購入月パターン
12365	0	0	22	0	0	0	0	0	0	0	0	0	001000000000
12373	0	0	14	0	0	0	0	0	0	0	0	0	001000000000
12378	0	0	0	0	0	0	0	0	219	0	0	0	000000001000
12384	0	0	0	0	0	0	0	0	14	0	0	13	000000001001
12393	0	22	0	0	17	0	0	0	0	25	0	0	010010000100

データ流通市場における匿名加工



個人情報保護法委員会規則第19条は,
匿名加工情報を成形するための,
必要十分条件としての役割を担う

PWSCUPにおける匿名加工



PWSCUPで用いられる匿名加工手法は,
ルールで定められる有用性指標と
再識別定義の制約の中で, いかにして高得点を取るかを
考えて作られた匿名加工手法であり,
各チームによる19条の解釈も反映されているものの,
第三者に提供できる状態に加工されているとは限らない

再識別の定義

1人の顧客を再識別するために，全期間の仮名を正しく当てなければならない

→ **仮名制御による再識別リスク低下**の効果測定ができた

顧客	12	1	2	3	4	5	6	7	8	9	10	11	
A	A1	DEL	DEL	A1	DEL	A2	DEL	A2	DEL	A2	A2	DEL	→ 再識別成功
B	B1	DEL	F2	B1	B2	B3	DEL	DEL	DEL	B3	B3	DEL	→ 再識別失敗
C	DEL	DEL	DEL	DEL	C1	C1	DEL	DEL	DEL	C2	DEL	G5	→ 再識別失敗

- ただし，PWSCUP2017の再識別の定義の元では，再識別されていないと判定された顧客であっても，実際には多くの月のデータは再識別されていた
- どの情報がその個人にとって機微な情報なのか，どれくらいの量の情報であれば漏洩を許容できるかということは，個人によって異なる
  - 個人の思考に合わせてデータに機微性の重み付けをする研究[9]が存在する
  - 来年度以降のPWSCUPでは，このように**機微性の重み付け**を行い，それに応じて再識別スコアが変動するようなルール　→ より現実的な内容のコンテスト

まとめ

1. 匿名加工情報の利用目的に関する再検討
   - 匿名加工情報は，本人の同意なしに，本来の利用目的外での利用が可能である
   - しかし，匿名加工情報の作成手法は利用目的に大きく依存
   - 利用目的が限定的であるほど，安全な匿名加工手法を適用できる
   - 統計データのみを公開する場合との違いはあるのか

2. 公式な再識別の定義の必要性
   - PWSCUP2017で採用された再識別の定義法では，不十分な点が存在
   - より具体的な議論を行うためには，**統一的な再識別の定義が必要**

3. 様々な状況を想定した匿名加工手法の開発と評価
   - データセットやユースケース，再識別の定義を限定した上で，
     様々なパターンに対して匿名加工手法を確立することが必要

参考文献

[1]“個人情報の保護に関する法律(平成15年法律第57号. 平成27年法律第65号及び平成28年法律第51号により改正, 平成29年5月30日施行)”, 2017.

[2]個人情報保護委員会, “個人情報の保護に関する法律施行規則(平成28年10月5日個人情報保護委員会規則第3号)”, 2017.

[3]個人情報保護委員会, “個人情報の保護に関する法律についてのガイドライン(匿名加工情報編)(平成28年個人情報保護委員会告示第6号ないし第9号)”, (2017).
(https://www.ppc.go.jp/files/pdf/guidelines04.pdf, 2017).

[4]個人情報保護委員会,“「個人情報の保護に関する法律についてのガイドライン」及び「個人データの漏えい等の事案が発生した場合等の対応について」に関するQ&A”,
(https://www.ppc.go.jp/files/pdf/kojouhouQA.pdf, 2016).

[5]個人情報保護委員会,“個人情報保護委員会事務局レポート:匿名加工情報「パーソナルデータの利活用促進と消費者の信頼性確保の両立に向けて」”, (https://www.ppc.go.jp/files/pdf/report_office.pdf, 2017).

[6]一般財団法人日本情報経済社会推進協会,“匿名加工情報の事例集”,
(https://www.jipdec.or.jp/protection_org/u71kba00000001hh-att/AOP_006.pdf, 2017)

[7]菊池 浩明, 小栗 秀暢, 中川 裕志, 野島 良, 波多野 卓磨, 濱田 浩気, 村上 隆夫, 門田 将徳, 山岡 裕司, 山田 明, 渡辺 知恵美, “PWSCUP2017:長期間の履歴データの再識別リスクを競う”,コンピュータセキュリティシンポジウム2017論文集, pp.128-135, (2017).

[8]PWSCUP 実行委員会,“PWSCUP 2017 匿名加工・再識別コンテスト 競技ルール, Ver1.3”,
(https://pwscup.personal-data.biz/web/pws2017/index.php, 2017).

[9]中川 拓麻, 荒井 ひろみ, 中川 裕志, “集合値データに対する個人適応型匿名化手法”, コンピュータセキュリティシンポジウム2017論文集, pp.1549-1556, (2017).