Vol.2014-DCC-7 No.11
Vol.2014-EIP-64 No.11
2014/5/15

ビッグデータ利用における個人データ保護における課題

中川裕志[†1]

ビッグデータのうち特に利用価値が高い個人データの利用にあたっては，プライバシー保護が必要になる．k-匿名化がプライバシー保護の技術として期待されてきた．しかし，個人データが行動履歴などのように巨大化すると，必ずしも有効ではない．まず，匿名化ないし k-匿名化が効果的である場合を明確化する．次に，他人に知られたくないプライバシー情報の定義しにくさについて述べる．これらの問題を法制度の観点から見直す動きについて紹介し，自己情報コントロールの実現による解決法の模索と問題点について議論する．

Problems of Personal Data Protection in Big Data Utilization

HIROSHI NAKAGAWA[†1]

Privacy protection is inevitable when using personal data which is especially higher value in big data. K-anonymity has been expected as a good candidate of privacy protection technology. It is, however, not practical for big personal data such as long geographical traces. In this paper, we clarify in what kinds of cases anonymization, especially k-anonynmization is effective. Then we explain how difficult it is to define so called private information. We also describe the various activities to deal with these problems from the legal viewpoint. Finally, we discuss the possibility of realize self-data-control to solve these problems.

1. 背景

ビッグデータとりわけパーソナルデータの活用が 2013 年 6 月に政府指針として打ち出され[閣議 2013]，政府のパーソナルデータに関する検討会が行われ，技術検討ワーキング WG 報告[佐藤 2013](以下では「報告書」と略記する．)が同年 12 月 10 日に公表された．

この報告書は技術的レベルの高い内容だが，そこで示された方向性に対して IT 業界から，パーソナルデータを含むビッグデータの扱うビジネスを萎縮させるとして反対論があがっている[大豆生田 2014]．

一方，パーソナルデータに関連する法制度に関しては，日本は不十分であるとして，EU からはゲノム情報などの有用な情報の輸入を禁止されているという状況を考慮すると，それを改善する法整備は喫緊と課題といえる[ジュリスト 2013]．

本論文では，この報告を念頭において，匿名化を現実社会で使うにあたっての技術課題，制度設計について述べる．

2. 匿名化における基本概念

個人データの発生源である人をデータ源の個人，パーソナルデータを何らかの手段で多数の個人から集め，それを使った事業を行う人や組織をデータ事業者とする．データ事業者がさらに別の人や組織に自らが収集したデータを渡す場合，受け手をデータ受領者とする．

個人データの収集において，データ源の個人は，データ事業者が示すデータ利用の許諾に関する文書に同意すれば，許諾文の範囲でデータ事業者はデータ利用ができる．ただし，許諾文で想定されるすべての利用法を網羅的に記述することは実質的に不可能である[Schoenberger2011]．

収集される個人データは通常以下の要素からなる．

I. 個人 ID(氏名)
II. 疑似 ID(性別，住所，年齢，国籍，データベースに格納されていることが知られている行動履歴など)
III. その他のデータ
IV. プライバシー情報：III.のその他の情報のうち，他人に知られたくない情報．いわゆるセンシティブ情報(人種，宗教，病名，収入など)が中心だが，場合によっては購買履歴や移動履歴も含むので，拡大した概念を表すためにここではプライバシー情報と呼ぶ．ただし，実はその定義が難しい．これについては5節で触れる．

ここで疑似IDに関する次のこと注意してほしい．個人の行動履歴などがデータベースに格納されていることが知られている場合を「**外部可知**」と呼ぶことにする．外部可知の情報が集積すると個人を特定する可能性があるため，疑似 ID となる．外部可知でない場合を「**外部不可知**」と呼ぶことにする．例を示す．

例1：病院である患者にある検査をしたことは，通常病院関係者以外には外部不可知．
例2：在宅ヘルスケアで，センサーから計測した心拍数などの健康情報を無線 LAN などで担当病院に送信する．これは，同居家族ででもなければ外部不可知だし，データの値自体は本人ですら知らないこともありえる．
例3：カードでの購買履歴はカード会社内部からしか分からないので外部不可知．
例4：滞在位置情報，移動履歴，コンビニでの購買履歴などは物理的な動きを伴うので他人から観察できる．さらに Suica の乗降履歴は特定の人物に目をつけているストーカーなどから見れば，その人物が Suica 履歴データベースに格納されていることが分かるので外部可知である．
例5：公共の場所や店舗に設置された監視カメラの映像に写っているかどうかも第三者から観察でき，監視カメラ映像データベースに格納されていることが分かるので外部可知である．

まとめれば，医療情報(病名など)，健康状態のセンサーデータ，財産，金融資産状況などは，データ収集者である病院，金融機関の外部者には知られないので，外部不可知である． そのように守秘義務のある組織が守っているデータ以外は外部可知でありうる．

†1 東京大学
The University of Tokyo

従来，個人情報保護法では，これらのうち個人を特定する個人情報とは，「個人ID」と生年月日などの若干の情報だけを意味しているとされてきた．だが，データ源の個人を特定できうる情報として疑似 ID も個人情報と見なせる．

報告書[佐藤 2013]の主張のひとつは識別と特定を以下のように精密に定義したことである．

- 「特定」とは，「ある情報が誰の情報であるかが分かること」
- 「識別」とは，「ある情報が誰か一人の情報であることが分かること」

この定義により匿名化の処理範囲を明確化し，技術的な検討がしやすくなった点で大きな前進である．

3. 完全な匿名化の不可能性

上記の報告書をまとめるにあたっての規制改革会議からの要請は「データ源の個人が同意しなくてもパーソナルデータを転売も含めて自由に使えるための匿名化の基準作り」であった．

しかし，報告書では，比較的簡単でよく知られている k-匿名化を基本に置くとすると，規制改革会議の要求に沿えるような完全な匿名化は不可能であるとした．以下で少し詳しく説明する．

k-匿名化：個人 ID を消去ないし仮名化したうえで，疑似 ID の情報の一部を消去あるいは精度を落とす技術であり，依然として匿名化の主要技術と位置づけられている[Fun 2010]．例えば，住所の記述から番地を削除するなどして，疑似 ID が同じである人が k 人以上存在するようにデータベースを変更する k-匿名化が報告書では念頭におかれた．つまり，データベースを k-匿名化すれば，データ源の個人は疑似 ID から一意的な識別ができない．当然，匿名化されていない外部のデータベースなどの外部情報と突き併せても個人を識別も特定もできない．

しかし，現実には住所，年齢，性別など少数の情報に疑似 ID が固定されているわけではない．例えば，データ業者 A のデータベースは疑似IDが k-匿名化されているが，個人を特定はできなくても識別できる購買履歴も含まれていたとしよう．一方別のデータ業者Bは購買履歴と，移動履歴（通勤などの乗降駅）からなるデータベースを持っていたとする．上記の議論により，これらはいずれも疑似IDになりうる．すると，データ業者Aのデータベースをデータ業者Bが入手すれば，購買履歴によって個人を一意に識別でき，その個人の移動履歴を知ることができる．したがって，突き合わせに使う外部データベースを予見しきれない以上，識別を防ぐにはデータ業者は 2 節II で述べた広い定義に基づく疑似IDを k-匿名化しなければならない．しかし，そうするとデータベースの精度は悪化し，データの価値は激減する．よって k-匿名化は実質的に不可能ということになる．

個人情報の保護の目的のためには，データ受領者からさらに別のデータ受領者への提供においても匿名性を担保しなければならない．匿名化できない場合はさらなる提供はできないことを法制化が必要である．報告書では，これを以下に記す米国の FTC3 要件をベースに検討している．

ＦＴＣ3 要件

1. データ事業者はそのデータの非識別化を確保するために合理的な措置を講ずるべきである．
2. データ事業者は，そのデータを非識別化された形態で保有及び利用し，そのデータの再識別化を試みないことを，公に約束すべきである．
3. データ事業者が非識別化されたデータを他の事業者に提供する場合には，それがサービス提供事業者であろうとその他の第三者であろうと，その事業者がデータの再識別化を試みることを契約で禁止する．

※個人を識別可能なデータと，ここで説明した非識別化のための措置を講じたデータの双方を保有及び利用する場合には，これらのデータは別々に保管すべきである．

注意しなければならないのは，この要件においてはデータ受領者がデータ事業者となって，他のデータ受領者へのデータの移管を認めていることである．よって，上記の k-匿名化の説明で述べたように，データ受領者＝データ事業者が使う外部データベースを予見することがますます難しくなってくる．かくして，どのような危険性が存在するかを事前に把握しきれない．この状況においては，データ源の個人から同意をとることは難しくなってくると思われる．

その場合でもなお可能なのは，いわゆる統計データである．ただし，ある集合中の個人が別人として識別されたり，実世界でのリアルな個人として特定されることができないような統計データの明確な定義を与える必要がある．実際，統計法では外部データベースとの突き合わせも勘案して以下ように匿名データを規定している．

統計法第2条12項　この法律において「匿名データ」とは，一般の利用に供することを目的として調査票情報を特定の個人又は法人その他の団体の識別（他の情報との照合による識別を含む．）ができないように加工したものをいう．

この条文中の「識別ができないように加工」に関して「匿名データの作成・提供に係るガイドライン」において，
1）識別情報の削除，2）匿名データの再ソート（配列順の並べ替え），3）識別情報のトップ（ボトム）・コーディング，4）識別情報のグルーピング（リコーディング），5）リサンプリング，6）スワッピング ，7）誤差の導入
などの処理が列挙されているが，匿名化の基準については，

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．このため，提供機関は，匿名化する統計調査ごとにその特性を勘案し，一橋大学における匿名標本データの試行的提供の事例及び諸外国の統計機関における同様の提供の事例等を参考に匿名化の基準となる値，例えば，最小値が 2 件以下とならない等を定める．

としており，ケースバイケースでの処理をデータ事業者に委ねている．よって，匿名化の基準については我々自身が説明責任を果たせるものを提示しなければならない．

4. ケースバイケースの匿名化の展望

報告書では 3 節で述べたように，一般的なデータに対して完全な匿名化ができないとしたが，同時に，個別のデータベースと個別応用によっては匿名化ができる可能性があるので，検討するとしている．ただし，報告書では具体策，具体例を提示していないので，以下で検討する．

3 節の議論により，個人情報すべてが疑似IDになりうることと，突き合わせる外部データベースの予見不可能性が k-匿名化を妨げているので，この条件を回避できる個別ケースでは匿名化の可能性がある．したがって，匿名化の要件は

Vol.2014-DCC-7 No.11
Vol.2014-EIP-64 No.11
2014/5/15

a. 疑似ID(住所, 年齢, 性別, 当該データベースに格納されていることが知られている, すなわち外部可知の個人に関する情報(行動履歴など))の有無

b. 外部可知／不可知： III. の「それ以外の情報」が収集されデータベースに格納されていることが外部の第三者に知られうるか／否か

となる.

上記 a.と b.の組み合わせは表1. に示す各ケースとなる.

表1. 場合分け

III. それ以外の情報	疑似ID無	疑似ID有
外部不可知	不可知 & 疑ID無	不可知 & 疑ID有
	個人IDの匿名化だけでよい	疑似IDの k-匿名化が有効
外部可知	ー	可知 & 疑ID有
		疑似IDの k-匿名化がデータの価値を大きく損なう

以下で表1の各ケースについて検討する.

- 外部不可知 & 疑ID無： データベースに格納されているか否かも知られず, かつ疑似 ID もないとなると, 仮にデータが公開されても本人特定は原理的に不可能である. よって, 個人IDを消去する匿名化だけで, 個人の特定はできない.

  ただし, 本人のデータ自体が万人周知で一意的である場合, 例えば10億円の宝石を購入したなどは外部可知である. この場合は, トップコーディングのような既存の手法で不可知化しなければならない.
- 外部不可知 & 疑ID有： データベースへの格納の有無は知られていないので, 識別, 特定の手がかりは疑似 ID だけである. この場合は, 疑似 ID から識別, 特定されなければよいので, 同じ疑似 ID の人が k 人以上いるように疑似 ID の精度を落とす k-匿名化が有効である.
- 外部可知 & 疑ID無： 外部可知になると同時に疑似IDとみなせるので, この場合に相当するものはない.
- 外部可知 & 疑ID有： データベースへの格納が知られており, データ収集事象を外部から観察できると, 疑似IDになってしまう. 疑似IDのデータが入手できれば, データと観察日時などから本人特定が可能である. では, データ自体を k-匿名化すればよいのではないかというとそれも難しい. なぜなら, 長期にわたって収集されたデータが大きくなると, データ自体の個別性が高まり k-匿名化が困難になる. つまり, k-匿名化するにはデータの精度を大幅に落とさなければならないが, そうなるとデータの価値自体が大きく下がってしまう. また, 個人 ID を仮名化し, その仮名化を 1 日単位など頻繁に取り替えることは有力であるが, 同一の個人の行動履歴ではなくなるため, やはりデータの価値は下がってしまう.

以上をまとめる.

●データベースの個人データが格納されていることが外部可知の場合は, 外部からデータ収集していることを観察あるいは他のデータベースとの突き合わせによって, 識別ないし特定が可能なので, 疑似IDとデータ自体を併せて k-匿名化しなければいけないため, データの価値は大きく減少する. よって, 完全な匿名化手法はない.

●データベースの個人データが格納されていることが不可知の場合は, 疑似IDがなければ k-匿名化は不要, 疑似 ID があれば疑似 ID を対象にした k-匿名化が有効となる.

これらを表1の各場合のセルの下段に記した. 一方, 外部可知／不可知の中間の場合については, 以下の課題がある.

●収集した全パーソナルデータからサンプリングしてデータベースを作る方法もある. この場合, 個人毎にはサンプリングされたかどうか分からないので, データベース上にある人情報が格納されているかは, 確率的に不可知／可知である. よって, プライバシーの安全性は確率モデルを作って評価する必要があり, 今後の課題である.

医療情報, ゲノム情報, 個人のセンサーから収集される健康情報, カードでのネット通販による購入, 個人の金融資産状況などは, 通常は外部観察できないので, 匿名化した上での活用ができそうである. 一方, 外部可知な移動履歴, 購買履歴などは, 匿名化が困難であり, 活用が難しい. これは, 一見, 重要度の低そうな行動の履歴データがかえって活用できないという矛盾した結論にみえる. しかし, 外部不可知なデータというのは, 元来が個人の物理的あるいは法制度的にプライベートな場面で収集されるので, プライバシーという意味では最初から堅守されている. 一方, 行動の履歴はプライバシー情報として堅守されていない. よって, 常識に沿った結論になっている.

5. プライバシー情報

II. の疑似IDおよび III. のその他のデータのうち, 個人にとって他人に知られると不都合なデータを2節でプライバシー情報と定義した. しかし,「不都合」とは何かを精密に定義していない. 加えて, 何がプライバシー情報かは個人ごとに異なる. この節ではこの問題を扱う.

● コアなプライバシー情報:

誰にとっても他人に知られたくない情報をコアなプライバシー情報とする. ゲノム情報, 病気などの生体情報ないし健康情報, 財産, 債務, 学業成績, 親族などがあげられるが, 何を選ぶかは社会常識によるしかない. 逆に言えば, その定義には社会常識程度の安定性はある. コアなプライバシー情報はおおよそいわゆるセンシティブ情報に対応する.

ところで, EU では滞在場所の情報は氏名と同じレベルの個人IDと見なす Data Protection Directive[LIBE2012] が昨年の欧州議会で可決されている. 日本では, 滞在場所, 移動履歴がどの個人IDなのかセンシティブ情報なのかの議論すら進んでいない状況である. ゲノム情報は個人IDに準じるとする考えがでてきている.

●状況依存プライバシー情報

上記の滞在場所や移動履歴がプライバシー情報かどうかは個人ごとに異なる. 例えば, ストーカー行為を受けている人にとっては, 相手に知られたくない情報なので, プライバシー情報であろう. しかし, 他人につきまとわれることのない人であればプライバシー情報とみなさなくてもよい. 議論を簡単にするためには EU のように個人 ID としてしまうのもひとつの策である. ただし, 滞在場所や移動履歴はビジネスに役立つ情報なので, できれば活用したいものである.

購買履歴も個人ないし状況依存である. たとえば, 薬剤の購入は場合によってはプライバシー情報になりうる.

宗教, 政治信条, 友人関係, 親類関係も状況依存性が高い. 友人関係は, 本人だけではなく, その友人に累が及ぶ可能性

があるので，プライバシー情報になりやすい．たとえば，ある売り込み業者が自分の名前をかたって友人に売り込みをすると，友人関係が悪くなる可能性がある．

このように状況依存のプライバシー情報は一律な扱いが困難である．プライバシー情報であっても 1) 外部不可知であり，2) トップコーディングなどで個人の特定ができない状態になっており，3) さらに疑似IDも存在しない，ないし k-匿名化されているなら，データ事業者が第三者に再識別や特定をしないという条件で提供しても危険性はない．それ以外の場合だと，第三者提供するにはデータ収集時に本人同意が必要であろう．だが，データ収集時に，そのデータの利用方法をすべて列挙することは不可能である．一方，データ源の個人にとっても，収集されたデータが後になってプライバシー情報になる，あるいはプライバシー情報だと気づくかもしれない．こういった事態に対応については後に述べる．

6. k-匿名化が誘発する濡れ衣

まず，以下の表2のデータベースの例について考えてみよう．

表 2 滞在場所のデータベース例

名前	年	性	住所	N 月 M 日 P 時の所在
一郎	35	男	文 京 区 本 郷	K 消費者金融店舗
次郎	30	男	文 京 区 湯 島	T 大学
三子	33	男	文京区弥生 ZZ	T 大学

最左列は人名だが，これは匿名化されなければならない．2，3，4 列は，疑似ＩＤで，１～5 列が総合されると，就活や婚活中に人にとっては，最右列の所在地に消費者金融が記載されていることは芳しくない．そこで，名前を A，B，C と仮名化し，疑似ＩＤの情報を粗いものに変更して表 3 のように改変する．こうすると疑似 IDは 3 人とも同じになるので，3-匿名化が実現でき，消費者金融に行った人を特定できない．ところが，疑似ＩＤでは 3 人を区別できないので消費者金融に行っていない残りの 2 人も消費者金融に行ったことを疑われる．これを k-匿名化が誘発する濡れ衣と呼ぶ[中川 2013]．濡れ衣を防ぐには２つの方法がある．

表３　3-匿名化したデータベース

仮名	年齢	性別	住所	N 月 M 日 P 時の所在
A	30 代	男	文京区	K 消費者金融店舗
B	30 代	男	文京区	T 大学
C	30 代	男	文京区	T 大学

第１の方法は，k-匿名化の k を大きくすることである．例えば，表 2，3 のような例で，k=30 であるなら，消費者金融店舗に出入りしたのが１名であると，わざわざ他の 29 名を疑う労力は骨折りであるという心理が働くであろう．[中川 2013]では，このことをコストの観点から分析している．ただし，k を大きくすると，データの精度が下がり価値が低下する．

第２の方法は，k-匿名化をしないことである．疑似ＩＤを変化させて，一致する人を増やす操作をしないので，通常は消費者金融店舗に出入りした人は１人と識別される．よって，濡れ衣を疑われる人はいない．ただし，他のデータベースと突き合わせると本人の特定がしやすくなる．

濡れ衣は無実の罪という側面が強いので，7節で述べる自己情報コントロールの問題として扱うのが適当であろう．

7. 自己情報コントロール

7.1 問題点の洗い出し

以上の考察から，k-匿名化のような識別の曖昧化を狙った匿名化では，実社会での応用場面をカバーしきれないことが分かった．

表 1 における外部不可知 & 疑ＩＤ有，外部不可知 & 疑ＩＤ無の場合には，ひとたび流出したら取り返しがつかないゲノム，健康情報が当てはまる．データ処理を病院などの収集した組織内に限定し，他のデータ事業者への提供は禁止すべきである．医療情報においては，これは現状と同じレベルの情報管理と考えられる．他の医療機関や研究所との協力に際しては，提供は直接提供する機関までとし，データ受領者である組織からさらに他の組織に提供することは禁止すべきである．これは，データの流通を追跡が容易に可能な範囲に抑えるための処方である．

表１の場合分けのうち，k-匿名化が有効に作用する可能性は，外部不可知 & 疑ＩＤ有　の場合だけであり，パーソナルデータの利活用において k-匿名化だけに頼ることは難しい．移動履歴や購買履歴などビジネス的価値が高いデータは，外部可知 & 疑ＩＤ有であり，k-匿名化には本質的に馴染まない．

この状況での現実的方策を考えてみよう．

- 個人 ID を消去ないし仮名化すること．さらに仮名の変更を頻繁に行うこと．この基礎的方策により，簡単には識別や特定ができなくなるので，必須である．
- 疑似ＩＤはデータベース内に含ませないことをデフォールトとする．疑似ＩＤも必要な場合は，それだけをデータベースから分離して別のデータベースとして，仮名化されている個人ＩＤとの対応テーブルは暗号化などでさらに管理を厳重化する．疑似ＩＤが存在しなければ，個人の特定は難度が高い．
- 第三者へ再配布できるのは疑似ＩＤなしのデータのみとする．ただし，移動履歴などはそれ自体が疑似ＩＤなので第三者配布はできなくなってしまう．

上で述べたように，個人データを k-匿名化などの技術でデータ源の個人が特定あるいは識別できないようにし，同意なしに個人データを利活用しようとする試みは隘路に陥っている．ここでデータ源の個人の「同意」という問題の原点に戻ってみる．個人ＩＤを消去するだけの匿名化処理をするだけで，広範なデータ利用に同意してくれるならパーソナルデータ利活用の障壁は下がる．だが，データ源の個人が容易にデータ利用に同意してくれるか，またどのような条件なら同意してくれやすいかが問題として浮上する．この問題を考える題材として EU Data Protection 改正案などと自己情報コントロールについて説明する．

7.2 EU の Data Protection 改正案と FTC５条

2014年3月12日欧州議会でEUのData Protection 改正案[LIBE2012]が可決された．この改正案はさらに理事会の可決を経なければ成立しないが，EUの個人情報保護に対する姿勢を示すものとして注目する必要がある．ここでは，1) A right to be forgotten，2) Easier access to your own data，3) Putting you in control，4) Data protection first, not an afterthoughtのような個人が自己情報に対するアクセスや訂正，消去を請求できる権利，個人データ保護をシステム設

計時から盛り込むべき，など強い個人情報保護の姿勢が打ち出されている．EUではプライバシー保護は人権の一部と考えられていることによるのであろう．

EUより個人情報保護規制が緩いとされる米国でも前記のFTC３要件があり，さらにそれらに関して「不公正又は欺瞞的行為又は慣行」に該当した場合はFTCによって差し止めや排除命令，民事制裁金を課すことができる．米国においてはさらに2012年にインターネット上で消費者のプライバシーを守る目的でPrivacy Bill of Rights [Whitehouse 2012] が公表された．そこでもAccess and Accuracyという項目でデータ源の個人が自己のデータを収集したデータ事業者に対し，個人が扱える様式でアクセスおよび訂正の要求を出す権利を持つとしている．

残念なことに日本の個人情報保護法やEUや米国の動きに追いついて入らず，個人情報の扱いを所掌する組織は整備されていない．結果として，EUからはデータ保護が不十分な国とみなされており，EU域内のデータを日本に持ち込むことができない．だが，EUのような規制の強い法制を敷くことにはデータを扱う業界から抵抗がある．このような状況なので，個人情報保護法の改正が計画されているが，その内容については2014年初旬の論文執筆時点では模索が続いている．

7.3 同意と説明責任

報告書[佐藤 2013]に記載された技術検討ワーキング WG への規制改革会議からの要請を見ると，個人データを利用するにあたっては，データ源の個人が同意さえすれば自由に使えると思われているようだが，これは必ずしも正しくない．つまり，同意の内容で無制限な個人データ利用を可能だと書くと，多くの人々から同意が得られなくなる恐れがある．[Schörnberger2013]の 9 章には，ビッグデータの利用法は収集の前には予め列挙できないので，利用法を指定しての同意取得は実効性がないと述べられている．同様にデータ提供先を同意時点で列挙しきれないであろう．

さらにショーンベルガーは[Cate2013]において，2008 年の調査では，このような契約文書のプライバシー・ポリシーを全文読むと，年間 244 時間が必要になり，同意はプライバシー・ポリシーを読まずに同意のクリックだけをする形式的な操作になっており，形骸化していると指摘している．そこで，実質的なデータ源の個人の個人情報保護をデータ事業者側に説明責任を課すことによって確保する方向を提案している．そのためには説明責任を担保する法整備を要請し，OECD のデータ保護のためのガイドラインの改正版[Cate2013]を提案している．具体的にはデータ源の個人の自己情報コントロール（開示，訂正と消去）ができることをガイドラインとしている．

これに対して，プライバシー・バイ・デザインの提唱者カブキアンは反論を展開している[Cavoukian2014]．つまり，同意を軽視しデータ業者の説明責任に個人情報保護を委ねるのは危険であるとし，むしろ個人データの管理をデータ源の個人が行うことを提案している．この反論は非常に強硬であり，その強硬さの理由が理解しにくかった．その理由について以下に考察する．カブキアンが提案し推進しているのは BigPrivacy というアイデアを基礎とするトラストフレームワーク(Trust Framework) [Cavoukian 2013]という民間企業の共同プロジェクトである．これは，サールズの Vender Relation Management:VRM [Searls 2012]に近いシステムとなる．ただし，新規の目的に対してデータ収集システムから設計する場合は可能性があるにしても，既に企業が収集しているデータ，ないしは事業化されているデータ収集と利用に適用することは困難である．このような民間ベースのシステムが機能するためには，個人情報保護法制上の制約から自由でなければならない．そのためには参加企業とデータ源の個人との契約に基づく必要があり，同意は必須になる．よって，同意を軽視するショーンベルガーに激しく反論したものを思われる．

[IPA2014]では，データ源の個人が自分の個人情報をデータ業者が利用することを同意する基礎となるデータ業者への信頼感醸成に寄与する要因に関する 1103 人への調査結果として，1)第三者機関（たぶん公的な個人情報保護機関）の保証，2)データ業者の良い評判，3)自己情報へのアクセス記録の閲覧，が列挙されている．1),2)は技術的課題の枠外である．3)は自己情報コントロールの機能の一種であり，技術的課題を含む．

一方，[IPA2014]では自己情報を見せる範囲を自分自身で設定する非常に強い自己情報コントロールは信頼感の向上に正の寄与はないという調査結果も示している．これは日本においては未だ個人による選択や制御の概念が浸透していないことを示す．また，将来的な浸透も不透明である．

7.4 自己情報コントロールを中心に据える方向性と課題

この節では，以上の考察から今後の個人情報保護の方向性を提案する．

個人データのデータレコードから個人 ID を分離することと仮名化は個人情報保護において必須と考えられる．その上でデータ源の個人がプライバシー・ポリシーに同意を得やすい，あるいは不同意に雪崩を打たないための抑止力として自己情報コントロールのひとつである自己情報へのアクセス記録の閲覧が説明責任の機能として実装されていることは有力である．[IPA2014]では調査項目に入っていないが，EU の Data Protection 改正案あるいは OECD のデータ保護のためのガイドラインの改正版[Cate2013]に書かれた自己情報コントロール（開示，訂正と消去）が実装されていれば，同意を得やすいことは確実であろう．また，データ事業者において説明責任あるいは自己情報コントロールの整備が義務づけられる法制度の整備が行われ，国際的な標準に近づくことで，日本がデータ保護の法制度が整った国として認識され，前記のデータ禁輸などが解消すれば国益に叶う．法制度の整備とは，個人情報保護の法制はもちろん，個人情報の保護を監督，管理を行う公的な第三者機関の設置も必要であろう．このような機関として 2014 年 1 月 1 日に設置された特定個人情報保護委員会を発展させる方向が考えられている．

なお，従来重視されてきた k-匿名化は，k が 5 以下の小さな数だと，データ源の個人に安心してもらえるかどうか疑問がある．また，k が小さいと 6 節で述べた濡れ衣の被害の可能性も高まる．それよりは，実効性のある自己情報コントロールが確保されていることを示す方が直観的で広範なデータ源の個人に理解を得やすいと思われる．

だが，開示，訂正と消去という自己情報コントロールはその実装となると問題が多い．例えば，自己情報へのアクセス記録の閲覧であるが，複数のデータベースを統合したデータベースへのアクセス，あるいは自分のデータが部分的に使われた場合のアクセスの状況をどのように閲覧させるかは複雑なデータベース検索システムが必要である．さらに，かりに k-匿名化も併用された場合，k-匿名化されたデータベースにおける自己情報へのアクセスとどう定義するか，また，k-匿名化されたデータベースを他のデータベ

ースと組み合わせた場合，など数理モデルが未知数で困難な場合が多い．

このような困難さは，匿名化ないし k-匿名化された個人データが他のデータ事業者に販売や再配布された場合，

1)販売先が併用するデータベースが把握しきれない，
2)販売，再配布がさらに多段に行われると，個人データの行き先の把握すら困難になる，

という状況なので，深刻な事態に陥る．よって，自己情報コントロールは実質的に不可能な状態になる．

加えて，自己情報コントロールを行えるタイミングの問題も大きい．つまり，データ業者が処理されたデータを開示する前から行えるのか，開示後でなければ行えないのかという問題である．カブキアンのプライバシー・バイ・デザイン[Cavoukian 2010]の考え方によれば，開示前から行えなければならない．技術的には開示前で開示後でも同じだが，開示前が原則であるとなると，いつ開示してよいかがはっきりしないため，データ処理の即応性に問題が生じかねない．ただし，これは技術的問題ではなく，各データ事業者が設定するプライバシー・ポリシー，ないしは法制度の問題として解決すべきものと考えられる．

7.5 自己情報管理の代理人

自己情報コントロールはデータ源の個人に安心感を与える効果がある一方，常にデータ事業者のサイトを監視してアクセス履歴閲覧や訂正，消去の要求を個人から発することは，個人負担が大きい．結果として同意の場合と同様に実効性のない枠組みになりかねない．VRM[Searls2012]では，デンター源個人，データ事業者，サードパーティ（クレジットカード会社のような存在．どちらかと言えば，データ事業者に近い存在）に加え，データ源の個人の代理をして個人情報の管理を行うフォースパーティを導入している．VRM ではフォースパーティは，人間よりはプログラムによって実現する方向が提案されている．ただし，個別性が強いデータ源の個人の代理人という意味では，人間のエージェントがデータ源個人から対価をもらって，データ業者に個人データの使われ方を調べ，訂正要求し，場合によっては削除要求とその執行を確認するという役割をしてくれる方法も考えられる．

8. 今後の課題

ビッグデータのうちでもビジネス的価値が高いパーソナルデータの利活用の促進が叫ばれているが，パーソナルデータは個人情報を含むため，技術的問題だけではなく法制度も関係してくる問題である．法制度としては個人情報保護法の改正作業が進んでいるが，そこでの主要な技術である匿名化については完全なものが存在しないという報告書[佐藤 2013]や論文[板倉 2014]が発表されており，今後の焦点は個別の場合における議論に移って行かざるをえない．本論文では注目されている k-匿名化が効果を持つ場合と持たない場合を分ける基準を示した．さらに k-匿名化が使えない場合に有力と考えられる方法，1)データ源の個人のデータ収集時における「同意」，2)データ事業者の説明責任を比較検討した．その中で自己情報コントロールが中心的な役割を果たすことを考察した．

今後は，匿名化が有効である場合分けの精密化，自己情報コントロールが実効性を持つかどうかの技術課題の検証を行っていく必要がある．

参考文献

[Cavoukian 2010]A. Cavoukian : 7 Foundational Principles of Privacy by Design. Information and Privacy, 2010. Commissioner/Ontario. http://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/

[Cavoukian2013] A. Cavoukian, Drummond Reed. Big rivacy: Bridging Big Data and the Personal Data Ecosystem Through *Privacy by Design*. www. privacybydesign.ca/. 2013

[Cate2013]F. H. Cate, P. Cullen, V. Mayer-Schonberger: Data Protection Principles for the 21st Century Revising the 1980 OECD Guidelines, http://www.oii.ox.ac.uk/publications/Data_Protection_Principles_for_the_21st_Century.pdf ,2013

[Cavoukian2014]A. Cavoukian, et.al. Big Data Calls for Big Privacy-Not Only Big Promises, www. privacybydesign.ca/index.php/big-data-calls-big-privacy-big-promises/ 2014

[LIBE2012]Committee on Civil Liberties, Justice and Home Affairs. Rapporteur: Dimitrios Droutsas. The European Commission's data protection reform proposals. 2012.12 http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-%2F%2FEP%2F%2FNONSGML%2BCOMPARL%2BPE-501.928%2B02%2BDOC%2BPDF%2BV0%2F%2FEN

[Searls2012]Doc Searls. The Intention Economy:When Customers Take Charge. Harvard University Review Press, Watertown Massachusetts,2012(邦訳： インテンション・エコノミー.翔泳社 2013)

[Schoenberger2011]ショーンベルガー&クキエ(斉藤栄一郎訳). ビッグデータの正体. 講談社.2013. (V.M.Shönberger and K. Cukier. BIG DATA A Revolution That Will Transform How We Live,Work, and Think.Houghton Mifflin Harcourt Publishing Co. 2013.)

[Whitehouse2012] The United States Government: CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY, Feb. 2012

[閣議 2013]世界最先端 IT 国家創造宣言について. 平成 25 年 6 月 14 日閣議 決定. 2013

[佐藤 2013]佐藤一郎，他: 技術検討 WG 報告書，パーソナルデータに関する検討会，2013． http://www.kantei. go.jp /jp/singi/it2/pd/dai5/siryou2-1.pdf

[中川 2013]中川裕志, 角野為耶: 滞在場所の k-匿名化と濡れ衣. 情報処理学会.第 62 電子化知的財産・社会基盤研究発表会(EIP 研究会) Vol.2013-EIP-62, No.12.2013

[大豆生田2014]大豆生田 崇志. 「プライバシーフリーク」発言を検証する．日経ビジネスオンライン IT Pro ,2013. http://business.nikkeibp.co.jp/article/opinion/20140312/260977/?ST=tech&rt=nocnt

[ジュリスト 2013]ジュリスト 3 月号 特集 ビッグデータの利活用に向けた法的課題 ―パーソナルデータ保護法制の展望, 有斐閣, 2014.3

[IPA2014]情報処理推進機構．ＩＰＡテクニカルウォッチ：パーソナルデータを活用したオンラインサービスに有効な個人情報保護対策，2014.3

[板倉 2014]板倉陽一郎，他:「完全な匿名化」幻想を超えて．SCIS 2014 3D1-4 IEICE. 2014.

[Fun 2010]B.C.M.Fun, K.Wang, R.Che, P.S.Yu: Privacy-Preserving Data Publishing: A Survey of Recent Development, ACM Computing Surveys, Vol. 42, No. 4, Article 14, 2010.