パーソナルデータに関する検討会　第３回技術検討ワーキンググループ　議事要旨

１　日　時　　平成２５年１１月　１日（金）１５：００～１７：３０

２　場　所　　中央合同庁舎第４号館　１２階　全省庁共有１２１４特別会議室

３　議　事

（１）開会

（２）各構成員の作業項目について

（３）技術検討ワーキンググループのアウトプットについて

（４）閉会

４　＜配布資料＞

【資料１】非定型パーソナルデータに関する分類と匿名化考察（岡村構成員提出資料）

【資料２】匿名データ活用に向けた現行法の解釈について（森主査代理提出資料）

【資料３】匿名データ活用に向けた現行法に基づく手続きについて（佐藤構成員提出資料）

【資料４】※非公開資料（席上配布のみ）

（参考資料１）パーソナルデータに関する検討会　第２回技術検討ワーキンググループ　議事要旨（未定稿版）

（参考資料２）※非公開資料（席上配布のみ）

（参考資料３）※非公開資料（席上配布のみ）

５　出席者

佐藤（一）主査、森主査代理、岡村構成員、佐藤（慶）構成員、高橋構成員、
松本構成員

総務省 総合通信基盤局 消費者行政課
経済産業省 商務情報政策局 情報経済課
消費者庁 消費者制度課

内閣官房情報通信技術（ＩＴ）総合戦略室
遠藤政府ＣＩＯ、二宮参事官、濱島参事官、瓜生参事官、村上調査官、

神成政府ＣＩＯ補佐官、楠政府ＣＩＯ補佐官、満塩政府ＣＩＯ補佐官

６　概要

岡村構成員から資料１の説明、森主査代理から資料２の説明、佐藤（慶）構成員から資料３の説明があり、以下の発言があった。

○これまでに使われている用語を使うことを止めるというのは、私も賛成である。一般での認識違いについては用語のもたらした悲劇だったところがあるので、匿名化は匿名化として、世の中に存在するが、それとここで言っている現行の個人情報保護法とか、あるいは将来改正されるような個人情報保護法の適用対象になるということは、違うものであり得るので、用語を変えて、若干そこに敷居があるほうがいいという気がする。そういう意味では、一部で使われている非識別化というのは、よいと思うし、非識別化というと、再識別化禁止とも整合するし、識別性があるとか、ないというのも整合するしいいと個人的には思っている。

○同じく、企業の中で、仮にガイドラインなどが出た際、周知するときに、１つには、一般用語を定義されると、非常に教育がやりにくくなる。例えば現行法の共同利用においては、「共同に利用することだと誤解しないでください。これは法律が定めた特殊な利用条件を課した利用形態のことを共同利用と言っているので、共同利用という言葉を使わないでください」ということから始めるので、これは非常に誤解を生むし、手間もかかる。

１つのアイデアとしては、用語がある意味聞き慣れない用語であれば、聞いた人は「これは何か？」と聞くと思う。先の例では共同利用と定義してしまったので、これは何かと聞かずに、「共同で利用することを共同利用」だと、誤解しまっていることが多いので、このような提言をさせてもらった。

○今回、資料１のスライドの８枚目には、匿名化できるかどうかということが書いてあるが、ここで言う匿名化というのは、どういう意味か。

○匿名化というのは、実際に大量のデータを再識別化不可能にできるかと考えてもらってよい。

○再識別化できないというのは、誰か１人に結び付けることができないということか。

○そのとおり。簡単に一言でいうと、先ほどの英語の de-identify というのは、非常に難しいということになってしまう。今、技術が進んでいるので、逆にいうと、匿名化もしく

は de-identify をする技術よりも、identify する技術の方がこの 10 年くらいでかなり進んで追い抜いている状況。結果的に、相当な条件をつけないと、それぞれの形であっても、再認識するとか、de-identify することから、identify されたところに戻すということ、逆に identify された状態から de-identify することについては、特定ができないということ。

では、大量のデータを匿名化できるか、という話になると、システムの総合能力で可能ということなのだが、大量のデータを実際に匿名化するためには、大量にこれを抽出して、それに対して、大量に全部１つづつに変更をかけていかなければならない。外から集めて、修正して、その全部をまた外に戻さなければならないことになり、双方向の作業が必要。

同じように、多種多様なデータもそうで、多種多様のものを集めて、双方向に戻して、書き直して、全部匿名化しなければいけないということになる。

一度インターネット、もしくは企業の中のデータベースに入った場合、中をのぞくことは可能なため、例えば前回、前々回のワーキンググループでもあったとおり、男性、女性で分かれていて、立川と駅の情報だけであっても、他の大量のデータを集めていくと、またもとに戻して推測ができる。推測はデータの数が多くなればなるほど、だんだん当たってくるので、もとに戻せてしまうことになる。そういうことで、結論的には、技術的には非常に難しいと思う。

したがって、皆さんが言うように、ある程度の法律もしくは罰則という形を整理しておかないと、例えば技術的に匿名化手法を公表して対応というのは、難しい。技術的にはできてしまう。

○森主査代理のスライド３で、パーソナルデータの利用・流通に関する研究会報告書の説明をされたときに、規制の中身というのは、第三者提供時における同意以外のことも考えているのかといったところが気になっている。

規制改革会議の問題意識で、個人情報には該当しない旨を明確にすべきということに関して、規制を逃れたいというのは、第三者提供時における同意以外にもあるのか。ここに関しては、鈴木委員は、第三者提供するデータは、匿名性のレベルによるのかもしれないが、基本的には個人情報を提供すると考え、第三者委員会が係わることによって、同意を不要にするという解釈だった。その辺りの親会の見解はどうか。

○今回、匿名化の目的の１つの考え方としては、ＦＴＣのレポートのように、そもそも規制の対象外にする、個人情報ではなくしてしまうという考え方がある。もう一つは、第三者提供の規制の適用をなくす。我々の目の前にある問題意識は、第三者提供に関する規制の適用外にすることに集中していると思う。

○再識別化という用語の定義と、禁止事項に対する正しい概念は何なのかを合意

して、それをわかりやすい文章で書く必要がある。

○再識別化については外国の先例をなぞれないと思ったほうがいいと思う。少なくとも英語でこれを区別している例を知らないのでどちらかというと、日本のこの議論は先行しているので、日本で答えを出さないといけない。

資料の「再識別化」については、従来の経緯として、再識別化と言っていたので「再」を使っているところがある。ＷＧの最後には、固まった図に、それぞれの要件を勘案して、用語の名前をつけるというのも、進め方としてはあると思う。

最後、わかりやすさをどう普及するかは、図を基に補足する文章を追記して定義するのがよいと思う。図で説明したことを法制化する人の御苦労は大変だと思うが、このＷＧの審議としては、いいのではないかと思って、御提案した。

○先ほどの御質問の対象となった禁止、再識別化のところだが、世の中にあるということだとどういうものがあるのか

○特にあるということではない。Ｙに関しては禁止というふうに、この矢印を位置づけているということだけを書いている。だから、Ｙのところだけが、内容を書いているという状態である。

そういう意味では、不可能とは言えないと、統計の専門家の方からは言われてしまったので、Ｚをどうするのか。ただ、Ｚについては、禁止はどういう禁止で担保できるのかということがあるので、Ｚはどういう位置づけにして、さらにどういう呼び方にするのか、悩ましいと思ったので、「？」をつけている。

○私も再識別化の「再」は取りたい。なぜかというと、最初の段階で情報を識別できない情報があり得るので、「再」を入れると識別化の定義を変に限定してしまい、識別化の全てを網羅出来ない可能性があるので、「再」は抜いたほうが、誤解がないという認識でいる。

あと１点、顔認識のところで、国家の安全保障の問題もあり、例えばスーパーの万引き防止のレベルもある。意見があれば、伺いたい。

○実際には多くの問題を含んでいて、運用に至っている。例えば、顔情報に肖像権があると言われかねない場合は、それぞれごとに特別な契約を結んで縛るのと、もう一つは、かなり曖昧にして運用している。

ただ、犯罪捜査の場合に限っていうと、そこだけは特例で、それぞれの国であったり、日本でもそうだが、特殊な例外事項を設けて運用している、運用で逃げている。

○顔認証の問題は技術検討ＷＧで議論する範囲を越えるが、今後、考えていく上では、例外的に扱うのか、対策を立てる必要があると理解している。

○実際のところは、個人情報、犯罪ではなくてつながる。画像、顔の情報は、個人認識しやすい情報であり、何らかの措置が要ると思う。

○再識別化の「再」を取ったらどうかというところは、まさにそこの部分は範囲にかかわると思っている。第三者提供の文脈の中においては、第三者提供として提供された情報をもとに、もとの状態に戻すというところに限った文脈であれば、「再」はあってもいいと思う。

指摘の部分は、まさにＺの問題だと思う。そのような、もともと公表情報みたいなものから、無理やり１とか２をつくるというのは、社会的課題としてはあると思うが、その意味で、４のところは「逆」なのか「再」なのかはわからないが、第三者提供の文脈の場合にはあってもいいと思う。ただ、私自身も余りこだわりはなくて、とにかく命名さえされればいい。ただ「識別化」だけにすると、誤解が出たり、あと、識別化を制約するというと、世界中他に類を見ないほど厳しい制約のように聞こえてしまう。それが気になる。

○ここに関しては、あまり明確な答えはないが、「再」をつけると、誤解する人もいるということを言いたい。

佐藤主査及び事務局から資料４の説明があり、以下の発言があった。

○顔を認識するのは、非常に簡単な技術でできる。顔の認識と定型情報というのは、どこかで必ずつながっているので、例えば顔の情報は出すことを禁止することによって、かなりの確率で再認識化を防げるということが１つある。

そしてデジカメで撮った場合には、位置情報が全部写真の裏にメタ情報として入っているから、顔の検索をするときに、メタ情報の文字で検索しているケースが非常に多い。「どこに、いつごろいた、この人たち」と検索したら、顔が出てくることもあるので、画像情報については特定して、これを出してはいけない、もしくはある程度契約で結ぶということだけは、特例で要るのではないかと思う。

また、年収は個人を特定しやすいと思う。

○簡単な匿名化であれば、こちらで危険なものを定性的に決めてしまえばいいのではないかと思う。

私が今の御説明で聞きたたかったのは、データを保有する事業者と分析等をする事業者（第三者）及び分析結果を利用する事業者間について、データを保有する事業者が分析等をする事業者（第三者）へ第三者提供することは、自分では全然利用しないということか。そうであれば、現行法だと委託ということか。

○そうである。

○資料３のスライド２の１番目は、レベル１とレベル２の両方をやるのかという意味でいえば、レベル１に今回は範囲を制限したほうがいいのではないかと考えている。
スライド２の右側の図で見るとわかるが、レベル１と禁止を組み合わせて保護というところで、本来レベル２は曖昧なままになるが、これが今回の報告書の限界だと、個人的には思っている。
資料３のスライド３を見ると、５の問題を取り扱わないということで、５が消えると、Ｃ問題、Ｚ問題も全部対象外になる。レベル２には気づいたけれども、ＷＧではレベル１をやる。
レベル１を今回はやって、レベル２に関しては、課題として書いてもよいのではないか。統計も、ここの部分は不可能とは言い切れないので、レベル２はこのまま課題に戻す。その状態は、現状でも世の中では統計みたいなことを暗黙に許しているわけであるし、諸外国でも許しているので、現状から悪くなるわけでもない。国際的に見ても、低くなるものでもないので、国際的な課題は、そのまま日本でも課題として残すという割り切りができれば、レベル１だけにするというのでよいのではないかと思う。

○親会のほうは、レベル１で止められるかというと、多分そうでもない。

○親会の話が出たけれども、結局、第三者提供の話である。これは契約が前提であり、まさに相手方が誰というのは決まっている。それから、親会の鈴木先生の提案も契約を前提にしているおり、レベル１で回答しても、親会からの諮問の内容には反していないと思う。

○佐藤主査が言われている第三者提供の類型Ｃが、一番ターゲットになるというか、ある意味では、こういったビジネスが成り立つようなことを目指すべきだと考えている。
医療の場合は、１次データホルダーの医療機関や健保自身はそれほど大きな存在ではなくて、それを集積する組織などが必要になる。さらにいえば、レセプトと健診情報といった異なる目的の情報を結合している事例もある。集積したり、結合したりしている組織が、データブローカー的な役割を果たし、データブローカーが、それぞれの目的に合った匿名化処置をやって更に下流の第三者に提供するというのが、有用性から見た目指すべき目標だと思う。　ビックデータビジネスをやりたい第三者提供先がビジネスをできるようにす

ることが重要だと考えている。こうしたことを可能にするためには、何らかの制約を受けるべきは、データブローカー的役割のサービスプロバイダーであり、そこに何らかのお墨つきを与えることによって、第三者提供時の同意を不要にするというのが、私自身の考えになる。

登録制は実質的には許認可制で、何らかの許認可を第三者機関がやる。そのときに、データを受ける第三者提供先がどの程度の安全管理措置をやっているかということも加味して認可すれば、良いのではないかと考えている。

○最後のところだけコメントをしておきたいのだが、前回、FTC の３要件を出したときに、匿名化の措置について、届け出制なのか、登録制にするかという議論があって、非常にいい考えだと思う一方で、それを受け入れる機関はどういうふうに処理するのか。現実問題として、登録制で中を審査するようなものは無理というか、いろんなものが起きたときにできないと考える。

またそれについては、元データを見ないとわからないので、多分言えないと思う。だから、現実には、ノーティスだけ渡して、何か問題が起きたときに、それをベースに証拠としては使えると思うが、それ以上のことは難しいと思っている。

一方で、届け出制にしないと日本はいわゆるプライバシー団体みたいなところもないので、実効性があるかどうかというのは、難しいところだと思っている。

ただ、技術検討ＷＧで、技術的に見てどう違いがあるのかということは言っておくべきだが、FTC の３要件の形にする、届け出制にする、登録制にする点には踏み込まなくてもいいと思っている。

○レベル１は相当低くてもいいのではないか。そのうえで、禁止の措置で主として担保するということにすれば、レベル１は常識の範囲レベルまで、たとえば、列削除程度のところまでに落とし込んだ上で、どの列を削除したのかということは公表する。

どの属性情報が提供されたのかというところを最低限管理できれば、気づく機会を制度上担保することはできるのではないか。

届け出の方法として複雑な処理方法等を入れると、おっしゃるように、この届け出は、見ても誰もわからないことになるので、そこはトレードオフなのだろうと思う。

結論ではないけれども、レベル１は低いところから始めて、気づいた追加条件があれば、順次、追加していくこともあると思う。

○問題としては、フレームワークの対象外とするかどうかというところの話なので、要件の①、合理的な手段を講じなければならない。要件の②が、データの再識別化を試みないことを公的に約束する。ここまでは提供元の話。提供元というか、ここまで提供先の話は全然出てこない。③に提供する場合にはということで、場合の話になるので、①と②で終

わっていれば、ＦＴＣ５条を背景にそもそも個人情報ではないという扱いをする。提供する場合、③が出てきて、この場合には、当然相手を特定しているからこそ、契約ができるわけなので、誰かわからない人に公表という話ではない。その限度では３要件も相手が決まっている。第三者提供の話である。結局、日本側では、３要件を全部まとめて皆さんが引用しているから、我々としては、レベル１がテーマであるということで、いいと思う。

○今、議論にならなかったことで、加えておかなければいけないことが２つある。

１つは、識別化の要件ないし匿名化の要件に関して、具体的な処理を書くのか、意見を伺いたい。

あと、医療データに関しては、医療データと一般のデータはそんなに違いがないことも事実であるが、ただ、１つ違いがあるとすると、いわゆる連結可能性というか、もとの個人を特定しなければいけない場合が出てくる。そういった対応表を持っているか、持っていないかということが出てくる。それは医療データに限らないが、対応表を持っているデータに関しては、別に扱ったほうがいいと思う。報告書にそれを記載するにしても、全体の流れからは、分けて考えたほうがいいと思っている。

なるべくターゲットの問題を単純化して、例外と言ってはいけないのだが、コアな話とは違うところは別に説明したほうが、読み手にとってはわかりやすいと思っている。皆さんに想像していただきたいが、そこに連結可能性を含めるデータを入れると、そもそも匿名化の意図も誤解されかねないので、分けて考えたほうがいいと思う。

○ＦＴＣのナンバー３にあるような、契約で縛るということが、現実だと思う。理由は私のところで述べましたけれども、データの処理が可能だということと、先ほど佐藤構成員が言われたレベル１がいいと思う。レベル１で、かつＦＴＣに限りなく近くなるが、契約で縛らないと、一般の人にはわからないと思う。最終的に考えると、公表という場合、方法論については公開しない､規制しないというところが､落としどころではないかと思う。

○私自身、そこの部分は必ずしも契約でいいと思ってはいない。先ほど法律で違法にすればいいのではないかとあったが、それは結構わかりやすいと個人的には思う。契約については、相手が特定するということははっきりしたけれども、それ以外の安全性の担保として、契約でいいのかということはある。

先ほどお金の流れが逆だから､パワーがないという話があったけれども､もう一つ別に、より根源的な問題として、契約を締結しているのは、提供元と提供先であるが、そこで利益が守られているのは本人、情報の主体である。提供元としては、自分が知らない人のことだから、これを実現するインセンティブを持たない。

仮に契約の中にその条項が入ったとしても、それはお上が契約の中に条項を入れろと言ったから、入れたけれども第三者の利益なので、動機がそもそも欠けている。これをどう

実現するかという問題があるので、そこまで３要件の３番目が考えているのだったら、別のものに置きかえ、改めたほうがいいと思う。

○１つ補足しておくと、データを保有する事業者と分析等をする事業者（第三者）間を契約でやるのか､何らかの規律でやるのかというのは迷うところで､企業の活動からすれば、契約でいったほうが自由度は高いが、最低限のところは縛るべきなのかもしれない。

これは技術検討ＷＧで議論することではないのかもしれないが､EU とのことを考えると、FTC に近い形になってくると､EU とは離れていく。ただ､EU のデータをもらうときに､FTC よりはきつめにする必要がある部分があって、最低限の部分は、法律で守っているので、EU に対して大丈夫であると言わなければいけない状況が出てくるので､個人的な考え方としては、契約だけでは難しくて、規律をプラスする形のほうがいいと思う。

○順番としては、もし、データを保有する事業者と分析等をする事業者（第三者）間が委託なのであれば、親会の話である委託だと、第三者提供の制限がかからないことについて追い越している感じにはなると思う。レベル１にリソースを傾注したほうがいいのではないかと思う。

○委託の話は除外したほうがいい。

委託の場合にまで匿名化してしまうと、データが丸まり、処理するデータの精度が悪くなるので、事実上、ビジネスとしてよくない。委託というのは、データを保有する事業者が責任を持つという状態があるので、責任所在が明確であれば、その部分に関しては、現行法のとおりでよく、わざわざ匿名化する必要はないと思う。

ただ、分析等をする事業者（第三者）側がやる内容に関して、不必要なものを出すのというのは、純粋にデータ最少化という観点で、安全管理上やればいい話であって、今回の匿名の件とは絡める必要はないと思う。ここは委託ではない場合に限った話のほうが、すっきりするし、もともと規制改革委員会が緩和したいと言っている範囲も、そこの話だと理解している。

○ここの委託に関しては、考慮していない。なぜかというと、データホルダーが自分で分析できないから、委託するというのは、可能性はあるけれども、ただ、現実には、分析する事業者がほかのデータとも組み合わせると思う。その時点で単純な委託とは言い切れないので、分類をどうするかという問題はあるけれども、そこは考えていない。今までの類型でもいけると思っている。

○規制改革会議のお題として与えられている、現行法での対応に関しては、現行法であると、合理的な匿名化の水準であるとか、統計も含めて、体系的に整理することは難しいと

いうのが１つの答え。それを踏まえて、これを改善するには、新たな制度なり法的措置をする必要がある。そこまでは、明確な答えとして、共通認識であろうと思うので、まず１つ押さえておきたいところだと思う。

次に新たな制度を検討するときに、匿名化を適用した情報に対し、これを新たな法制度の中の特別な取り扱いの位置づけにするのか、そういう情報は、その枠外を出るのかという観点も考えなければいけないのではないか。

もう一つ、今は形式的な特定個人の識別性という観点だが、親会の議論だと実質的な個人識別性という観点で､守るべきパーソナルデータを考えているので､そうした観点だと、単純に個人の識別、名前、氏名だけではなくて、位置情報だ、継続的に取得される購買情報、あるいは携帯電話、パーソナルデータの ID なども、守るべきものとして考えていこうということがあるので、単純に匿名されるデータを落とせば十分だというのが、本当に妥当かどうかということも含めて、考えていかなければいけない。

そういった意味では、次の段階で詰めていく新たな法的な措置の部分は、親会の議論ともお互いにフィードバックして議論しないと、ここだけで先行した議論というのは、なかなか難しいのではないかと。

それから、レベル１を議論するにしても、非特定化で十分なのか、非識別化までいかなければ、提供してはいけないとなるのか。レベル１の中でも、レベルがあるのだろうと思っている。公開を前提にしないで第三者提供する匿名情報だとしても、どこまであれするのかということがあると思う。

〇それも結局法的措置との関係があり、最終的には親会の判断だと思うので、余りこちらで細々言えないという問題はある。

以上