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

１　日　時　　平成２５年１０月１７日（木）１３：００～１５：３０

２　場　所　　中央合同庁舎第４号館　１２階　全省庁共有１２０８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）次回以降の予定について

（４）閉会

４　＜配布資料＞

【資料１】個人識別できない匿名データは作成できるか（高橋構成員提出資料）

【資料２】匿名化レベルの分類について（菊池構成員提出資料）

【資料３】参考になる事案（位置情報以外）の整理案（佐久間構成員提出資料）

【資料４】医療等分野におけるパーソナルデータの利活用の類型及び考察（松本構成員提出資料）

【資料５】政府統計における匿名化ミクロデータの特徴（伊藤構成員提出資料）

【資料６】実施可能な匿名化（佐藤構成員提出資料）

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

５　出席者

佐藤（一）主査、森主査代理、伊藤構成員、岡村構成員、菊池構成員、
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、向井政府副ＣＩＯ、二宮参事官、濱島参事官、瓜生参事官、宮田補佐、

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

○親委員会に当たるパーソナルデータに関する検討会の第２回が 10 月３日に開かれた。その際に、基本的には第三者機関をつくる、匿名化に関しては 100％識別不可能な匿名化の技術というのは、なかなか難しいということを述べた。また、事業者が匿名化をしたとしても、ちゃんと匿名化をしているということのエビデンスを出すことも難しいし、逆に匿名化されていないのではないかということの証明もなかなか難しいということを補足した。

高橋構成員から資料１の説明があり、以下の発言があった。

○言葉をこの後も正確に使うという観点で、資料１のスライド 12、少なくとも現行の個人情報保護法の考え方では一番上のポチのところは「特定個人識別」と呼んでいると思う。

２つ目のポチのところは、そういう意味では、書き直すならば「ｋ＝１は個人識別を意味するけれども、必ずしも特定個人識別を意味しない」ということになるので、この後もその用語で進められればと思う。特定個人識別と個人識別を混ぜることはよくない。

○法律上、ｋ＝１であることイコール個人識別で識別性があるということになるのかどうなのか自体、まだはっきりしないところがある。このような区別をするということが、ある意味ではちょっと議論を省略しているところもあるかと思う。

菊池構成員から資料２の説明、高橋構成員が代理で資料３の説明、松本構成員から資料４の説明、伊藤構成員から資料５の説明、佐藤構成員から資料６の説明があり、以下の発言があった。

○資料２の 18 枚目のスライドで「99.9％行削除する必要性」ということは、例えば２駅の場合は信濃町と昭島よりもマイナーな駅が選ぶ分には特定できるということは、これはその人がｋ＝１になってしまうということか。

○そのとおり。

○すなわち、これはいわゆる特定ではなくて、個人の識別が可能ということになるのか。

○そのとおり。やがて時間がたち、ほかのデータと照合していくうちに、いずれ本人特定識別まで行ってしまう可能性がある。このレベルで議論しておけば、その後、攻撃者のモ

デルが変わったとしても、同じような議論が上のレベルでもできるという考えである。

○資料２について、個人特定のモデル、特定のやり方というのは、資料に則したような形になるように思う。一方で、これに類しない形としてこのモデルに当てはまらない例というものはあるか。

○資料３にもあったトランザクションのデータ、それから資料４の医療データなどのケースがある。ただ、例えば、医療データでは、病名や治療費というデータについては昭島どころではなくて、データ１個からでも十分特定可能な厳密さがある。分布は違うであろうが、存在可能な再識別性という観点で一様にはかることはできるかと思う。

○そのモデルの場合には、逆に今度はデータへのアクセス容易性が低くなっていくこともあるのでは。

○そのとおりで、そこで守るしかない。資料２では定量的には今回議論せず、攻撃者にとって一番都合のいい状況でリスクを評価した。

○技術検討 WG で期待されている匿名化というのは、当然、守備範囲というものがある。一つの非常にわかりやすい守備範囲としては、そもそもプライバシーとの関係で問題になる。例えば、情報公開法５条には、識別性はないが公にすると権利侵害になるようなものは公開の対象としなくていいと書いてある。それは本当に権利侵害が生じるのか、公開するべきということで裁判を幾つもやっている。そういうところで、これは公表されれば、識別性がなくても権利侵害になるという場合には、あまりｋ＝１かどうかということは全く問題になっていない。問題になっているのは、これが誰かの情報であることは明らかで、それが公表されることはその人にとって権利侵害かどうなのかということを、情報の内容だけを見て判断している。そもそも匿名化で第三者提供しようというようなものの対象にはならないというものも制度上当然出てくる。また、例えば総務省のパーソナルデータ研究会の報告書では、一般パーソナルデータとその慎重な取り扱いを要するものと、センシティブなものと３つ分けている。これを全部匿名化して第三者提供できるようにしなければいけないという話ではない。そもそも対象が限られている。

それから、もう一つは、提供先で例えば先ほどの立川×浜松町×田町に１人いる、信濃町×昭島に１人いるということは、それ単独ではそれほど権利の侵害がある感じがしない。これを提供先でいろいろな情報をつけて一意になったときにその先にいろいろな情報、購入履歴とかそういうものを提供先に渡し集積すると、それはそれで問題が生じる。そういう集積する方向での分析というか、データの利用に関する規制というのは、これは別の観点で見るべき。

それを個人情報保護法に入れられるのかどうかわからないが、匿名化した結果としてｋ＝１になってしまい、それが識別情報として利用されて、そこにいろいろなものがくっついてくるから危険だということは、それはそうなのだけれども、ｋ＝１になった後のことというのは、提供先の規制として別に考えるべきかもしれない。

○資料２のグラフについて伺いたい。仮に例えば米国の FTC３要件的に、そのデータを合理的に非識別化するといった場合に、匿名化措置をとったエビデンスについて用意すべきものがでてきて、例えば統計的にこういう何％でこういったグラフというものを用意したほうがいいとなるような場合、これを現実問題として事業者ができるものか。専門知識のない人たちが匿名化をする状況で運用するということになったら、ある意味で頭を使わなくてもやれるような形にしないといけない。

○心配されるところはよくわかる。まずは、技術的な難易度があるかどうかという問題点が一つと、もう一つが公平に行われているかどうかということを第三者に証明するのは難しく、例えば、資料２の見積もりも簡単なモデルのもとでの話であり、仮定をちょっと変えるだけで随分も結果も変わってくる。

資料５の「disproportionate effort」が、まさにこういう数値をあらわすことに該当するのではないかと思うが、それが実際に再識別確率を求めたりとか、再識別されないことを証明したりとか、そういうところまで求めているかどうかにもよるかと思う。

○先ほど報告した「disproportionate effort」ルールというのは、ガイドラインというよりもむしろ秘匿に対する考え方ということができる。一方で、例えば資料２で計算されたような再識別の確率を実際のデータを使って検証することは行われている。人口センサスの例で言うと、人口センサスのデータと外部のデータとのマッチングを行う、あるいは人口センサスの個票データから個体が識別される確率を算出するために、例えばｋ＝１のｋ-匿名性とある意味で符合すると思うが、母集団一意がどれぐらい見つかるのかということを具体的に数字で出して確率をはじき出すとか、あるいはそれ以外にも個体識別に影響を及ぼすであろう様々な要因に基づいて条件的確率を計算するなど、そういったような形で確率を算出して、その結果、数字が非常に低いという検証結果を出しているということがある。

○もう一点、データオーナーでないとできないことがある。つまり、個別の個票・レコードを持っている人であればその具体的なｋ＝１はｘが何位のとき、という計算ができることになる。資料２作成時、公開された情報のみで作成したが、データを持っている人ならば必ず厳密な評価ができて、またその結果を証明することはできるだろう。

○現実的にはデータも日々変わる。乗車履歴のようなものは比較的年間を通して安定していると思われるが、日々変わるようなデータに関しては、その時々、そのデータの特性ごとに再検証しなければいけないのでなかなか難しい。

○特定個人識別と識別の話について、法律上違う要素が入っていると感じるのは、例えば、法律のほうでは「周知の変名」のような論点である。周知の変名というのは、みんなが知っているペンネームや芸名、ハンドルネームといったものなのだが、それをみんなが知っていれば識別性があるというような話になっており、結局あまりｋ＝１の話に尽きていない。何となくその人だとわかるあるコミュニティ内部での話のようなことが語られており、そういうところがｋ＝１の問題以外に入っていると感じている。それをどう考えるか教えて欲しい。

○言葉を分けないと混乱するということで分けたほうがいいというのが基本的な意見だったが、識別という話と特定という話は階層が違うもので、特定というのは、誰がというのが、名前なり、先ほどあったニックネームであれ、この人だということがわかるところまで行くこと。識別は、例えばＡさんかもしれないけれども、誰かはわからないというところまではたどり着いてしまうというところが識別というところで、個人情報保護法は、それが特定個人を識別できる情報を個人情報と呼んでいる。そのため、簡単に言うと特定されなければ、識別までであれば、厳密には法律上の個人情報とは解釈されない場合もあるというところがちょっとややこしくなっているのだろうと思う。

その意味では識別までだったらいいではないかというのが一考あるが、識別された状態というのは、非常に危険な状態であり、Ａさんだということはわかっていると、そこにいわゆる属性情報がたくさん連なることになるので、その中の一番弱い１カ所が特定に結びつくと芋づる式に一瞬にして全部がつながってしまう。資料２の例で言うと、信濃町と立川に乗っているとかというのは、ＡさんだということはわかってもＡさんが誰かということはすぐには特定されない。ただ、すぐに特定されはしないが、そのときに、信濃町から昭島に行っている人は膵臓がんだったということがわかり、それがＡさんと結びついてしまうと、信濃町から立川に行っているのが誰かがわかった瞬間に裏でその人は膵臓がんだということがわかってしまうので、特定はされていないが識別された状態のリスクをどう考えるかというところ。

本来は、特定までされなければ、被害とかには及ばないのだが、識別の状態が非常に危険なリスクの高い状態で放置された状態だと考えれば、そこをやはり禁止するのかどうかが論点となる。資料６はそういう意味で、個人の再識別なのか、特定個人の再識別なのかは敢えて書かずに、両方のところにおいて、再識別という言葉を使っている。

○本日は、かなり識別の話が徹底してなされたわけだが、特定の要件というのはあるか。

○特定のところはまさにリンクの問題や、いわゆる連結可能の問題のところで、連結先のところに特定個人の情報が入ったときが初めてそこがつながるところなので、ここの部分は列を消すというところで本来は消去できるところだと思う。

○資料６について、米国 FTC３要件は、私も国内で適用する場合にかなり問題があると思っている。指摘のあった点以外にも、少し危惧していることは、例えば２番目の「再識別化しないことを公に約束する」ということについて、実際に再識別化をしていないことをどう証明するか、かなり悪魔の証明に近い状況になっていて、現実には、匿名化した情報以外の情報から識別化や特定ができてしまうという状況がある、やっていないのにやったと言われたときに、どう証明するのかというのが、おそらく実運用的には問題になってくると思っている。

それともう一つは、３番目の部分、危惧しているのがいわゆるポイントを使ったカードで、情報を共同利用モデルとして事業者間で共有化するモデルがある。そのときに、責任は誰にあるのか、Ａという会社の情報とＢという会社の情報を持つことによって、識別ができてしまうような場合､どちらに責任があるのかとすごく微妙な問題が出てきてしまい、おそらくそこを少し補足しないと実は難しいという認識である。

○質問のとおりで、現状で米国 FTC に関しては、そこの部分の解決はこれから図るというのが現状だと認識している。米国 FTC は強制力がないことが問題であったので、今回、強制力を持たせるのだというところで解決すると言われることがあるが、むしろ強制力の問題ではなくて、やはり①のところの具体的な指示がないというところの状態において、たとえ強制力を持たせても、再識別化していないということの判断基準がない状態では、何をもって伝家の宝刀を下ろすのかというところが不明になっている。

アメリカの産業界からも現状で既に FTC に対しては①をちゃんと示せ、基準を示せと言っている。

産業界からすると、サインさえすればいいのだなという解釈になってしまっている。そのため、自分がそれで適切だと思ったレベルにしたと言い張ればいいのかという話になってしまう。むしろ健全な企業が FTC に基準を示せと言っていて、語弊があるがある意味ちょっと健全でない企業でサインさえすればいいのだと思っているところはとりあえずサインをして、何らかの第三者、例えば顧問弁護士に審査を受けるのでもよいと勝手に考え回しているというのが現状である。

そのため、今回 FTC が強制力を持った場合にも、ではそこの①のところをどうするのかというのは、まだ公開されていないし、もしかしたらこのままなのかもしれないという危惧はされているという状況である。

○識別と特定の問題について、金融機関の子会社に対する情報提供については個人情報保護法の法律の範疇でやられているはずなので、その行為自身は違法ではないと思う。聞きたいのは、個人情報保護法単体だけではおそらくユースケースとしてこの検討会で犯罪の悪用に関して考えた際、他の法律に抵触する形、例えば脅しのケースなどは刑法であり、それから実際のパーソナルデータを商売に悪用するなど、そういうペアリングのような、法律としてはどういうペアが考えられるのか教えて欲しい。我々が考える上で、ユースケースを一つ一つ挙げるのは難しいが、悪用される側から見たときにはどういう法律のペアリングというか、抱き合わせのケースを考えなければならないのか。

○質問は、パーソナルデータが悪用された際にどんな害が生じるかということを考えて、それにそのどういう法律の適用があり得るかということだと認識した。
例えば、資料３で実際にサンドイッチを食べているのを見て、それから Facebook をチェックしたらライブに行ったと書いていた。したがって、その人はポルノ DVD を借りたというのがあったが、そのように攻撃者が追跡するような場合、例えば住所がわかったり間取りがわかったりして窃盗に利用するなどは禁止されるが、それをどこまで悪用されるかということは当然想定して考えなければいけない。それを個人情報保護法の中に取り込むとしたら、それはある種のセンシティブさというか、先ほど３分類として話をしたとおり、その慎重な取り扱いを要する、センシティブな情報か、などのやり方で悪用可能な情報を取り込んでいけるのかもしれない。
　ペアリングということでも、プライバシーの裁判所の事案で裁判所が心配しているのは、将来どの法に触れるかということではなく、やはり権利侵害があるかないかの判断になるので、法に触れるかという観点から言っているわけではなく権利侵害がどういう場合に発生するのかということ。それについて裁判所が懸念を示しているのは、集積することによって、個人の人格に密接にかかわるような情報になるということがあり得るということ。
　この判断においては、考えるべきは収集される情報の性質ということが一つと、もう一つは収集したり利用したりすることの目的がどうなのか、正当な目的かどうか、それで収集したり利用したりすることの方法及び対応がどうか、ということ。そのことを総合的に判断して、その権利を侵害したかどうかを考えていて、あまり、将来どういうルール違反につながるかということは、それほど明らかにされていないし、やはり危険なものをより重大な身体・生命にかかわるようなものに使われるものは、センシティブ性のほうで評価するのではないかと思う。

○すなわち、権利侵害のところまできちんと定義をして、その先に例えば犯罪であるかの判断ということで、今回の定義の中ではそこまではまずは考えないということと認識した。

○情報単体としての危険性と、我々の一番重視すべき基準は権利侵害ということだと思う。

○先ほどの例で銀行の例、犯罪の例が挙げられたが、匿名化をすることによって識別性を下げるということはできるが、同時に匿名化をすることによるマイナス点もしくは問題点を我々議論で考えておかなければならない。例えば銀行の例で言うと、ローンの返済に滞った人が１人そのｋ-匿名性のグループに含まれていたときに、下手するとそこのグループに入っている人みんなが疑われる可能性がある。情報によっては、ネガティブな人が１人でもグループにいたら、グループ全体を否定したほうがいいというような事例。これは避けて通れない問題だと思っていて技術で何か補完をするのか、規律によってそれを何らかの形でリカバーするのかというのは、考えておかなければいけないことかなと思っている。つまり、ｋ＝１にならないからこそ問題になるということ。ｋ＝１になればむしろその人に特定できるので、周りの人は巻き添えにならない。

○例えば経産省ガイドラインは、本人の氏名というものを個人情報の例として明確に挙げているわけだが、裁判所はそのケースによっては、氏名だけだと同姓同名の人があり得るから権利侵害があるとは必ずしも言えないみたいなことを言っているケースもある。そういう意味で、特定というものも非常にわかりにくい。

○あるクラスにカテゴリーされることによる不利益な点というのはいわゆる１-多様性の話がある。資料３のサンドイッチの件はまさにそうで、コンサートに行ってサンドイッチを食べたらポルノ DVD を買ってしまったみたいなことが特定されてしまったり、先ほどの例で言うと、信濃町に行くと癌であることがばれてしまったりとか、そういう属性の一部がわかってしまう問題点というのは確かにある。

○何でこういうことを言ったかというと、匿名化をすれば大丈夫で、何でもハッピーになると皆が思っているかもしれないので、問題点というのも認識していただいたほうがいいということで、あえて発言した。

○特定するかどうかは、先ほど列の削除だという話が出たが、それはそのとおりだと思うので、特定するかどうかは属性が右に伸びるかどうかの問題で、識別されるかどうかというのは、当該行がｋ＝１かどうかという、基本的にはそういう考え方でもよいのだろうか。

○列の削除というところは最低限担保が必要だと思う。
　諸外国はどうしているのかというと、匿名の詳細をここまで詳細に検討している国は、知る限り日本以外にない。諸外国では匿名は匿名、統計についても統計と言っているだけで、この匿名がｋ-匿名性の手法で評価して、１なのか２なのかというところで匿名と言っていいのか悪いのか、などの議論は少なくとも日本以外にはない。

だから、そういう意味だと、ちょっと乱暴な表現ではあるものの、諸外国は特定さえされなければいいというところに割り切っている可能性がある。ただ、そのときに識別から特定に行くところのリスクヘッジをどうするのかというのは、それぞれ何らかのことはやっている場合もあると思う。例えば、FTC においても、特定はされていなくても、勧告を出している場合もあるわけなので、そういう意味では、そこが特定さえされなければいいというところだとは単純には言えないが、ただ、法律の条文とかそういうレベルでは、特定のところまでの話しかしていなくて、この識別性に関しての部分に関して言えば、匿名や統計はいわゆる匿名、いわゆる統計をしていればいいという考え方にしか、今、なっていないのが現状だと思う。それでよいということを提案しているわけでは全くなく、日本はある意味では逆にここを乗り越えると他国に例のない非常に正確にそこの部分をちゃんとやった例に進んでいくのだろうとは認識している。

○組み合わせで識別がおきなくても属性がわかってしまう問題を潰すというのは、技術的にはかなり難しい問題。どこにどういう不都合な事実が潜んでいるかを判別するのは難しい。たとえば統計的な情報を公表する場合、公表する情報の数が有限であれば、１個１個を吟味することはできるのではないか。しかし無数の組み合わせを含んだ情報の中に、不都合な事実が本当に入っていないかという確かめが非常に技術的には難しい。

○12～13 年前の某検索エンジンの会社について、ペンタゴンの中の地下の地図を一般公開していたので、検索エンジンでそれを載せた。ところが、ペンタゴンが３日か４日後に非公開にした。一方、検索エンジンでアーカイブボタンを押すとまだ出てくる 50 億ドルの罰金を取ったという大きな問題があった。先ほど消去の話が出たが、消去が本当にできるのかというところが、技術的に確認する必要がある。実際には、いろいろな情報がリンクしてという話がある中、データベース上では全部存在していて、それを 10 列ある中で、２列しか公開しなくても、必ずリンクは残っているし、消した情報も基本的には残るので、技術的には無理だと思うのだが、何らかの条文もしくは契約文書か何かで完全にそれが見られないようにしておくことなど、IT 的な一つの縛りを入れておかないと、実際にはデータは消えない。ここのところは逆に言うと、非常に無理であるということもあるし、本当にもう一つの言い方をすると、それは何か技術的に縛る条項を入れないといけない。ディスクの中身を消去しても全部残っていて、大型コンピュータになればなるほど、消すというのは物理的には無理になってくるので、それもどこかで一文入れるようにしなさいというガイドを入れるのか、IT 技術的な何か一つボーダーというのが必要なのではないかと思う。

○今回の匿名化は、やはり第三者提供の文脈の中での匿名化というところを考えるべき。
第三者提供の際にその列を含めないという話なので、もとの持ち主が消去する話は全く

別の問題であり、渡すときにその列を含めないというのが本来正しい表現だったかなと思う。だから、削除ではなくて、渡すときに含めないという意味である。

そういう意味では、氏名、例えば今後、再識別化を禁止するというような運用管理基準をつくったときに、それはしてはいけないということになっているので、名前がくっついていてもいいのではないかというのはやはり乱暴で、名前はやはり最低限消してほしい。コンタクトインフォメーションと言われているが、単体で個人まで行き着いてしまうようなものは一旦全部消すことになると思う。例えば、メールアドレス、電話番号、住所等々みたいなものは消すということを、先ほどの私の資料で言う最低基準として定めることはそんなに難しくはないと思っている。

そのため、十分条件にすると大変なのだが、必要条件はピンポイントで列記して、これとこれとこれは第三者提供するときに渡してはいけない、そのほかのもらったものをコンピュータで処理して無理やりもとに戻すようなことは運用上やってはいけない、という組み合わせにすれば、双方が非常に現実的な形で基準化できるのではないか。

○今話をしたのは渡すときという行為そのもののこと。現状、クラウドのサービスがどんどん増えている中で、単体でファイルを渡しているように見えていても、実際には大型のデータベースのコピーを一部見られるようにするという行為を渡すとすることもあるし、本当に切り取って渡すケースもあって、そこのところをきちんと渡すと定義しておけば、問題ないと思うけれども、今のクラウドサービスではなかなかファイルの一部分を切り取って本当にアクセスできない状態で渡すというふうになっているかというと、ハッキングすれば戻っていけるわけであり、そこは私も懸念しているところ。

○今の話の問題というのは、もう一つ、消すということ以外に、いわゆるデータの修正の問題もあって、第三者提供をしたときに、提供をしたデータに何か問題があったときに、それを修正するといったときの問題も出てくるのだと思っている。

○今後の進め方の話として、資料４の説明にもあったとおり、医療関係は全く分けて整理を進める形にしたほうがよい。医療に合わせると、民間のマーケティング情報は厳し過ぎる基準になるし、逆にマーケティングのほうを想定していれば、医療は全然緩いということになるので、ちょっと医療のところは特出しにして分けてこの基準の整理、あるいは検討の整理というのはしていかないといけないのではないかと思う。

資料４にもあるとおり、では何で医療だけなのだというと、一つは、まず、やはり同意を必ずとるということで問題ないか、ということ。公益性の問題として同意をとらなくても統計化して利用するということは､許してもらえないと困るという部分が一つあるのと、あとは先ほどの特定の推定のところに行きやすい属性情報というものが実は非常に多く存在しているというところがあり、少し医療は特出しと思っている。ただ、それに該当する

ほかのものがあれば、同じくそこに投げ込んで特出し化していくということもあると思うので、これをちょっと一緒にして汎用的な基準にしてしまうと、両者にとってどっちつかずのものになってしまうので、基本的には分けて考えたほうがいいのかなと思う。

あと、逆に医療のところはそう考えていくと、諸外国のところで、医療に特化した形での基準というのは先例があるので、それを参考にするということはあり得るかもしれない。

○医療データの取り扱いに関して、基本的には、今回の枠組みの中に乗せたいというような話を関係当局の方から伺っている。なので、一つの考え方としては、パーソナルデータという枠の中に入れて、医療データだけにはこうこうこういう制限があるという形で外に出すようなやり方というのが一つのやりかた、それ以外にもちろん医療データは別物と分けるという考え方もあると思い、それは議論だと思っている。

○表現が悪かったが、特出しと言っているのは外に出すという意味はなくて、区別して議論をするという意味。この検討会の射程の中に入っていることは別に構わないし、それを区別して考えるというところが必要だという趣旨である。

○そういう意味で言うと、金融の与信データなどは、医療と同じぐらいセンシティブなデータが含まれているので、やはり特出しするのであれば、金融も入ってくるため、これもこれもとどんどん増えていくことになるかもしれない。

○ただ、先ほどのとおりセンシティビティーで分ける必要はないと思っている。属性の特性の問題、それからいわゆる同意をとれないという問題という話。金融に関しては、例えば預金残高を何か統計に使うか使わないかというのは、これはやはりもしそれに同意が必要だったら同意をとればいいし、不同意のものは使えないということになっても仕方ないと思う。医療は、それでは不同意となったときに、それの医療研究ができなくなってもいいのかというところのいわゆる公益性の問題というところでのポイントが一つ。それから、例えば、医薬品みたいなものに関して、金額だけから医薬品名がわかって、それで医薬品名がわかると病気までわかるという問題も医療分野は持っていて、そういうようなものは他のものにあまりない。値段だけから、その人のクリティカルなセンシティブ情報まで一気にわかってしまうということがないので、そういう意味で医療はやはり特殊であるし、やはり海外でも医療を特殊化するのはそういう部分ではないかなと思って発言した。センシティビティーの問題で分ける必要はないと思う。

○あと、医療データに関する要件として、いわゆる再結合可能性というものを担保しなければいけないのが取り扱いの違いだと思うのだけれども、それに関して何か補足はあるか。

○少し細かいが、医療等分野について昨年度の厚生労働省での検討会で一番もめたのは「医療等分野」の「等」の範囲。ここにそれぞれの構成員の意見、思惑の違いがあった。検討会での議論は、医療分野の個人情報保護法の特別法というよりも、医療等分野での個人番号である医療等 ID の適用される範囲の話になるが、要は医療等の「等」が、介護や福祉まで含まれるかといった話になる。こうした議論は、増大する社会保障の対応にも関係していて、この社会保障分野を効率化するためにはどうしたらいいかという話になる。そのためには、介護や福祉分野まで含んだ医療等分野のパーソナルデータを利活用し、社会保障分野全体、そこをいかに最適化、効率化するかといった話になる。そうすると、今度は狭義の医療ではとどまらなくなる。そういったところが、多分、今すぐにではないかもしれないが、やはり次の課題としてある。医療分野が特別だと言ってしまうと、今度は、狭義の医療だけで閉じたパーソナルデータの利活用したできなくなる可能性がある。現在、求められているのは健康な人を病気にしなくするといったことでありが、そのためには、狭義の医療分野のパーソナルデータだけでは十分に利活用可能な統計情報等が作れない。
そういった課題があるので、もちろん現時点では特殊な話かもしれないが、これからは取り組むべき課題であり、医療分野が、特別とばかりは、言えないではないかと思う。

○おそらく医療データを使いたいと思っている人はたくさんいる。例えば、食品とかで、医療情報に基づいて食事のメニューを作るようなときなど、いろいろなケースがあるので、なかなか線を引くのは難しいのではという認識でいる。

○ちょうど、今後、処方せんを電子化するといった話があるが、処方せんに関して言えば、処方薬だけではなくて、一般薬とのインターラクションを見たいと言い要求があると考えられるが、そうするとは処方薬ではない薬との販売情報等とのデータの結合が必要になるので、そういったところも視野に入れて考える必要があるのではないかと思う。

○資料１の９ページの下の図が今の医療データの取り扱いともかかわってくる。会員番号や生年月日、住所、年齢というものを匿名化しても、属性情報、この場合は購買品から個人が識別なり、場合によっては特定できてしまうのではないかということなのだが、であれば、購買品も何らかの加工する匿名化の対象にするというと、多分、データの活用的に非常につらくなってくる。

ここの部分というのは、ひとまとめに購買品とか属性情報も何らかの加工をしろというのか、特定の業種ごとにというか、用途ごとに加工方法を規定するのか、おそらく全部のケースに関してこのワーキングで指針をつくることはできないと思うけれども、ただある程度の流れはつくっておかないと今後活用が進まないと思っている。

例えば、カーナビであれば、乗り始めと乗り捨てのデータを落としなさい、捨てなさいというのも一つの考え方。ある程度このワーキングとしては、業種や用途に応じて、属性

にかかる有効利用と購買品に相当するようなデータというものの取り扱いを決めていかなければいけないと思っているけれども、何かその辺補足するようなことがもしあれば発言してほしい。

○医療的な話で言うと、病気を加工してしまい、みんな例えば「内臓疾患」になってしまうとそれは全然役に立たなく、加工されていない正確な情報のままできちんとした管理をするしかないので、当然匿名化して活用の限界はある。

領域やアプリケーションごとに妥当なものというのはあるのかもしれず、資料２の例で出たアプリケーション、鉄道の話で言うと、結局どこの駅からどこの駅に行ったと限定したレベルであれば、それなりに匿名化したデータは作成できそうだという話だったと思うが、そういう切り出し方を探っていくのが現実的なのではと考える。

ただ、それは非常に探索範囲もかなり広くなってしまい、この技術検討 WG で扱える範囲ではないと思うが、結局そのデータを分析する目的に応じて必要なデータを作成することが重要なので、それは万能な匿名データを考えるより意味がある。その分析の目的をしっかり定義して、それに応じて適切なデータが匿名でも役に立つ領域があれば、それなりに役に立つ匿名データが出る可能性があるということかと思う。

○資料６に匿名化よりも「再識別化不可能」という言葉でまとめられていたが、今話のあった病名などがデータの加工ができないということを考えると、でもその分析用の情報は個人を特定・推定し得る情報なので、匿名化というよりは、再識別不可能性または識別不可能性の如何に差をつけずに議論をまとめていったほうがいいのかと思っている。

○禁止ではなくて、不可能という言い方の方がよいということか。

○匿名化される対象以外からも個人を特定できたり、識別できてしまう可能性があるので、広く見ておいたほうがいいのではということ。ただ、この議論は構成員の考え方次第であり、意見を聞いてから考えなければいけないことだと思っている。

○第三者提供ではない一般論という意味であれば主査の言うとおりだと思うが、資料６の文脈は第三者提供をするときの話である。そのときに不可能という言葉よりは禁止というほうがいいのではないかというのが考え。

○この後ワーキングとしての成果物は何か。

一般的に使えるガイドラインを作ろうとしているのか、あるいはこれをしてはだめ、といったような、べからず集のようなものを考えているのか。先ほど医療は特出しするかしないかといった議論はもちろんあったが、それを含めてこれから先、目指すべきアウトプ

ットの方向性だけでも決まらないと進め方がわからない。技術的には今日いろいろな意見があったけれども、やはり匿名化の手法それぞれにリスクがあり、匿名化した後でも属性の一部が推定されるような危険性があるというところで我々の議論は概ね合意していると思うけれども、ここから先をどうするのか。

○具体的な考えがあるかというと、主査としてのスタンスは基本的に構成員の議論を見て考えるということで、前回と今回の議論で、基本的に匿名化には限界があり、特に汎用的な匿名化の方法はないというのが、おそらく皆さん共通の認識だと思っている。

ただ、そう言ってしまうと、データの活用というのはなかなかできないわけで、どこかで折り合いをつけなければいけない。つまり、個人に係る情報を保護しつつ、そのデータを活用するという折り合いをつける。その手段が匿名化という理解である。

つまり、両方、今言った条件、保護と活用と両方担保する形で結論をまとめられたらと思っている。それをどこまでやるかということになるが、一つには、これは技術検討ワーキングであるので、できないことはできない、できるということはできるということはまず第一のミッションである。

その意味で言うと、匿名化を主にやりなさいと我々はミッションを受けているが、匿名化としてできること、できないことというのをまずきっちり分けたほうがいい。ただ、できない、だけをいうわけにもいかず、どこか落としどころはあると思うので、そこの部分は具体例になるかもしれない。先ほどの話のように業種とか用途ごとになるかもしれないが、列挙するような形でもよいと思っている。

あと、ガイドラインをつくれるかどうかというのは、それは技術ワーキングのミッションではないと認識している。もう少し技術に立って、実現可能性ないし、運用面も含めてできるできないという部分をまとめることが重要なミッションと思っている。

○ガイドラインは作らない、親委員会には技術的な立場からの今できることとできないことを整理すると理解した。

○親委員会にどのような形のものを出すのか。この議事録だけだとかなり議論が発散しているので、構成員間での合意がとれるかどうか。

○どういう形にするのかという問題については、それは例えば作業部会のような形で公表する場合と、親会に対して提言するようなものである場合とで全然中身も違うと思う。

親会側の事情を言えば、資料６の FTC３要件はかなり前面に出てきていて、総務省のパーソナルデータ研究会の報告書でも同じようなことが書かれていたし、親会で前回、鈴木先生の提案もあった。実際には合理的な非識別化というのはアメリカでもわからないということなので、そういうところは親会としては知りたいことだろうと思うし、同じ匿名化

の技術にしても、私がわからないと他の親会委員の方も技術的なことはわからない可能性が高いので易しく説明することが必要になったりする。他方、作業部会として公表するということであれば、それはもう少しプロフェッショナルな内容にするということもあるので、それは決める必要がある。

○医療分野は特出ししたほうがいいという話があったが、資料４の最後のスライドは医療分野だけでなくなるべく汎用的な話として表現できないかと思い書いたもの。一般的に第三者提供をするための条件について記述してみた。

医療分野に関して言えば、「安全に匿名化した状態」を定義したいという話があったが、そもそも医療データで、利活用可能な安全に匿名化したデータというものは非常に難しいと思っていて、結局は、匿名性のレベルとアクセスコントロール性及びその技術的安全性、その組み合わせによりプライバシー侵害を防ぐことになる。これは、先ほど議論のあった削除なども含めてになる。

○今日はどうもありがとうございました。

この親会のほうの趣旨は、何しろデータをいかに個人、社会のために有用に使えるようにするための道を開くかということ。個人的に考えたのは、一社完結型で情報を使う状態に対し、いろいろな人が情報を使う際の活用度合いが同じレベルになるだけでも随分違うと考えた。

そのために、要するに変な使われ方をしないようにするための何か工夫ができれば非常によいということと、一社完結型であってもまだ壁があるというところについては、今度は法律の話をしてもらわないといけないのではないかと思った。

また、今日明確になったのは、特定個人に何かが戻っていってしまうというようなケースについて、医療はもともとそれをすることが目的にしているが、誰、ではなく同一人物であるということだけがわかっていればデータセットが非常に役に立つという分野が当然あるということで、話がわかりやすく進みそうな期待があるなという気がした。

さらに、最初のデータを加工して渡ったものと、全然別のところから来たデータを一緒にしたらもとに戻りやすいというものだけは相当避けないといけないということが非常にわかり、大変有用でした。

素人でもそれくらいはわかる議論があったので、結果としても大変いいまとめができることを期待している。

以上