



SEMINAR REPORT

パーソナルデータ安全活用のためのプライバシー保護技術の開発と制度設計への貢献



日本電信電話（株）
セキュアプラットフォーム研究所
主席研究員／チーフ・セキュリティ・サイエンティスト

高橋 克巳 氏

皆様、こんにちは。NTTの高橋克巳です。今 SCAT の新宿にある会議室にお伺いしているのですが、広い、ちょうど正方形っぽい部屋の中に、本日講演をされる田中さんと SCAT の方のちょうど4人で広い部屋を四角く囲んで巨大な雀卓のようなところにいます。完全に3密も避けられていて、そもそもここにたどり着くまですごく御苦労があったと思いますが、開催にこぎつけられて、誠にありがとうございました。

改めまして、1月に SCAT の会長大賞を受賞しました。ありがとうございました。SCAT の皆様、それからこの仕事は数多くの方々の御支援、御鞭撻、また一緒に仕事をしてくださったことによるものですので、こちらから失礼いたしますが、心より御礼申し上げます。

私はNTTの研究所でずっと研究開発に従事しています。今回の業績はここ15年近くの仕事に対するものですので、自己紹介的なものも話の中でお伝えしたいと思っております。

パーソナルデータを活用するとは

それでは本題に入る前に、パーソナルデータを活用することについて考えてみたいと思います。皆さんがお考えのパーソナルデータの活用は何でしょうか。例えばコンビニで買い物するとポイントがたまります。そういった自分に対していろいろ便益があるというのがひとつの基本です。さらに進んでくると、例えば自分が帰宅しているいろいろしゃべると自分のことを分かってくれていて、電気がついたりテレビがついたりという話があるかと思いますが、これも自分のために役立つ例です。

次に、今は新型コロナによるパンデミックですから、世の中のことが気になります。どう行動しているのか分からないというのが皆さんたくさんあるのではないのでしょうか。コロナ関連のパーソナルデータ活用で有名なものに、人出の空間統計、新宿の人出は昨年比80%でしたとか、こういったことがニュース等で報道されると思いますが、これはまさにパーソナルデータの活用、携帯電話の位置情報を使ってそれが可視化されています。このことで世の中の動きが分かります。

先ほど自分の情報で自分の便益というような話をしましたが、コロナの場合は他人のことが気になるのかもしれない。3密という話がありますが、おとなしく散歩あるいはお出かけをしたいとき、人に迷惑をかけたくない、あるいは人混みのところに行きたくないようなときは、他人がどこにいるか、混んでいるところが分かれば、それが避けられるわけです。

そうすると普段、我々はあまり他人の行動に興味がないとしても、コロナの場合は、例えば密を避けるというようなことで、他人の行動を知り自分の行動を変え、戦略的に振る舞いを決められる可能性があります。行動再開で経済を回すみたいなこともあります。何が本当に良いのかなかなか分からない話です。気がついたら馴染みの飲食店が潰れてしまったという御経験も皆さんもあるかもしれません。しかし応援しようにもどういう振る舞いがいいのかということがなかなか分からないと思います。

パーソナルデータを活用するとは、自分が便利になるだけではなく、コロナの環境下においては世の中全体で他人の振る舞いも知りながら密を避けたり、あるいは助け合ったりというような活用が期待されていると考えます。

本日、お話しすることは2つです。テーマAは、技術についての話です。テーマBは、技術が法制度の一部となりました。匿名化技術ですが、その経緯についてお話をしていきたいと思っております。それではいきなりまとめです。このまとめ2枚で本日の話はご理解いただくことができます。

テーマA まとめ

データ活用のためのプライバシー保護技術

目的

- データを使う時に、そのデータから誰か個人のプライバシーに関わる情報ができるだけもれないようにすること

技術

- 匿名化と暗号化に代表される
 - 匿名化：見られてもわからないようにする
 - 暗号化：見られないようにする

NTT 

5

図1 テーマA まとめ

まずテーマA、「データ活用のためのプライバシー保護技術」をまとめるとこのようになります(図1)。データを使う時にそのデータから誰か個人のプライバシーに関わる情報ができるだけもれないようにすること、これがデータ活用のためのプライバシー保護技術だと考えています。そしてその技術は匿名化と暗号化に代表されると考えています。匿名化というのは、そのデータを見られても分からないようにする。一方、暗号化は、御存じのとおり見られないようにするということになります。

テーマB まとめ

匿名化技術が法制度の一部となった経緯

- 平成27年改正個人情報保護法で導入された(2015)
 - 匿名加工情報
 - 一定のルールの下で、本人の同意を得ることなくデータを活用できる
- 先立つ、IT総合戦略本部に設置された検討会で検討された
 - パーソナルデータに関する検討会(2013年9月-2013年12月)
 - ↳ 技術検討ワーキンググループ(2013年9月-2013年11月)
- さらに先立つ、経産省の研究開発プロジェクトで技術検討された
 - 情報大航海プロジェクト(2007-2009年度)

NTT 

6

図2 テーマB まとめ

続いてテーマBです(図2)。匿名化技術が2015年に改正された個人情報保護法で法制度の一部となりました。これはどういう経緯で作られたかという、先立つ内閣官房のIT総合戦略本部に設置された検討会で、私も参加させていただいて検討されました。ここに書いてある技術検討ワーキンググループはもう7年も前になってしまったのかという感じですが、非常に面白い議論をたくさんさせていただきました。さらに先立つ経産省の情報大航海プロジェクトで技術検討されたという流れであると考えています。

データ活用のためのプライバシー保護技術とは何か

データ活用のためのプライバシー保護技術とは、「データを使う時に、そのデータから誰かのプライバシーに関わる情報がで

きるだけもれないようにする技術」と述べました。ここをもう少し細かく見てみます。

「プライバシーに関わる情報」、これは何か。なかなか定義が難しいものですが、「その人が他人に知られたくない情報」と考えることが多いです。

ただし、それがどこからプライバシーかということに決まりがあるわけではなく、個人差もありますし、感じ方は非常にいろいろあります。我々事業者目線で言うと、多様なプライバシー感に対応する必要がある、ここまで踏み込んでしまってプライバシーに迷惑をかけているのではないかと疑心暗鬼になって悩んでいるところもあります。

それから「情報ができるだけもれない」ことが大事です。できるだけというところがポイントですが、実はパーソナルデータは何かの役務に付帯して渡されるものですから、例えば誰かに住所をもらさなければ荷物の1つも届かないことになりす。何ももらさないわけにはいかないというので、「利用目的に応じた必要最小限」というような言い方がされます。できるだけ伝えない、できるだけもらさないということが目標になります。

誰が使う技術？

- 企業・組織が、お客様に迷惑をかけないため
 - データを落としても、被害を起こさないようにしたい
 - 中身が誰のことかわからないデータにし、自由に活用したい
- 個人が、自衛のため
 - サービスを受ける時の情報登録をできるだけ少なくしたい
 - できることなら利用履歴を把握されたくない
 - そもそも他人に何か知られたくない
- これまで主に、企業・組織のための技術として発展してきた

NTT 

10

図3 誰が使う技術？

続いて、誰が使う技術かを考えます(図3)。まず企業が使うことが中心です。「お客様に迷惑をかけないため」と書いてありますが、お客様に迷惑をかけてしまう最たるものは情報漏洩です。企業は当然防ぎたい。企業のもう一つのモチベーションはデータ活用です。個人単位のデータ活用が必要でない場合は、誰のことかわからないデータにして自由に活用したいというニーズがあります。

次に、プライバシー保護技術は個人が自衛のために使うことができると考えられます。例えばサービスを受けるときに情報登録をできるだけ少なくしたいとか、そもそも他人に何か知られたくない等いろいろなモチベーションがあって、自衛に使うことができます。これまでプライバシー保護技術は主に企業向けに発達してきた経緯がありますがこちらも重要ですね。

プライバシー保護技術は匿名化と暗号化の2つに代表されま。匿名化とは、データを見られても分からないようにするということです。個人が分かるデータも、「20代・女性」であれば誰だか分かりません。他方、暗号化とは鍵を渡した人にしか見られないので、ほかの人には何も分からないという効果があります。

匿名化と暗号化

それではこの匿名化と暗号化を1つずつ、少し細かいところを御紹介していきたいと思います。

匿名化とは

一般的な「匿名化」のイメージ

かんたんそう

氏名	住所	性別	年齢	趣味
佐藤	東京都新宿区	男	45歳	マラソン
鈴木	東京都三鷹市	男	41歳	マラソン
山本	千葉県船橋市	男	32歳	映画

匿名化

氏名	住所	性別	年齢	趣味
**	東京都	男	40代	マラソン
**	東京都	男	40代	マラソン
**	千葉県	男	30代	映画

- 実際に研究してみると
 - 正しい使い方が理解されていない
 - 技術が半完成
 - そもそも名前がよくない

NTT 

13

図4 匿名化とは

上に描いてある図が一般的な匿名化のイメージです(図4)。佐藤さん、鈴木さん、山本さんといった名前を消して、住所等々も加工していくと匿名化ができます。私も簡単そうだと思います。しかし実際に研究してみると、技術は簡単なものも確かにありますが、実は正しい使い方が理解されていない。

そして技術が半完成、そもそも名前がよくないのではないかなど、いろいろなことに気づきました。

研究してみると：正しい使い方が理解されていない

- 「名前を削除で大丈夫だ」と考えられていた時代があった
 - 名前を削除して公開されたデータを、他のデータの生年月日・性別・郵便番号とマッチングしたら再識別できてしまった
 - ・ マサチューセッツ州 2002年の事例
 - ・ 再識別：匿名化されたデータから元の個人が誰かを特定すること
 - これを防ぐには匿名化の程度を上げる必要がある
 - ・ 例) 生年月日を「年」まで上げる
- 匿名化の程度の認識
 - 匿名化が必要な程度の認識には個人差がある
 - ・ 例) 自身のことと考えると「ゆるい匿名化では気持ち悪い」
 - 匿名化の程度の現実的な必要性は状況依存
 - ・ 例) 高い安全管理があれば、高い匿名化でなくともよいことがある

NTT 

14

図5 研究してみると：正しい使い方が理解されていない

順番に見ていきます(図5)。まず、「正しい使い方問題」ですが、つい最近まで、「名前を削除しておけば大丈夫だ」という考え方が実際にありました。マサチューセッツ州で20年近く前にあった事例として、名前を削除して医療データを公開したケースがありました。別の名前が削除されていないデータを持ってきて、そのデータの生年月日、性別、郵便番号とマッチングしたら誰かが分かってしまったという事故になりました。これを防ぐためには匿名化の程度を上げる必要があります。程度というのは、例えば生年月日を「生年」にするというようなことです。

匿名化の程度の認識は非常に差があります。例えばデータを使われる個人の気持ちで考えると、「ゆるい匿名化では気持ちが悪い」と考えますし、一方でデータを使う側があまり考えずに「名前を削除しているだけで大丈夫だ」となどという間違った認識が起きる可能性があるわけです。さらに、匿名化をどの程度行うかということもあります。匿名化の程度の現実的な必要性は結構、状況依存です。例えばデータ自体に高い安全管理が施されていれば、匿名化を一生懸命やらなくてもよい場合があると技術的には考えられています。

研究してみると：技術が半完成

加工と評価

- 加工手法はいったん確立していた
 - 「名前を削除」や「年齢を四捨五入」や「特異値を削除」などをする
 - これらの一部は体験的に知られたもの
 - ・ 例) 顔を隠して、年齢のサバをよめば・・・
- 評価手法は未確立だった
 - 匿名化してデータがどれくらい「安全か」を評価する
 - この一部は公的統計分野では歴史的に蓄積があった
 - ・ 「k-匿名性」とは統計分野の知見の一般化と考えられる
 - いわゆるパーソナルデータでは蓄積なし
- (参考) いわゆるIoT時代で、新たな個人識別の問題が見つかった
 - 例) 名前も住所もない履歴データから個人がわかることがあるが、これまでの蓄積が通用しない
 - 個人識別：データから対象の個人が特定されること

NTT 

15

図6 研究してみると：技術が半完成

続いて、技術が半完成という話をしたいと思います(図6)。半完成とはどういうことかということ、実は匿名化は加工すること、それがうまくできているか評価することの両方がないとだめなのです。ざっくり言うとその加工の技術はほぼ確立していますが、評価のほうあまり確立していないという状況でした。

加工手法は普通に確立していました。例えば名前を削除して年齢を四捨五入して云々ということが体験的に知られていました。

一方、評価手法はまだ確立していませんでした。評価とは匿名化したデータがどの程度安全かということを確認することです。これはなかなか難しい問題です。この考え方は公的統計、いわゆる統計局で歴史的に蓄積がありました。これくらいの塊にして統計を公開すれば問題が起きないだろう、といったものです。

後ほど説明しますが、k-匿名性というものがあります。この考え方は実は統計分野で蓄積されてきた知見の一般化と考えられます。ただし、統計的な形式のデータに対しては知見がありましたが、いわゆるパーソナルデータ、履歴データといった複雑で細かいものに関して言えば、蓄積がありませんでした。さらに現在はIoTで新たな個人の識別が起きる問題も認識されています。

3つ目、名前がよくないという話です(図7)。「匿名化しました、いやできてないだろう」という落語のような話があります。匿名化は行為なのか結果なのかということ冷静に見てみる必要があります。世の中で「匿名化」と呼んでいるものの多くは単に加工行為だけのことなので、加工行為だけで匿名化と言っても匿名化されていないよ、ということが起きてしまいます。

研究してみると：そもそも名前がよくない？

「匿名化しました！」←「してねえだろう」

- 匿名化とは、行為なのか？結果なのか？
 - 世の中の匿名化の多くは加工のこと
 - 加工行為だけで匿名化といっても「匿名化されていない」かもしれない
 - 加工と評価がセットの匿名化ならば、行為と結果が一致する
- 加工と評価が一体となった匿名化技術はあまりない
 - 「匿名加工情報」は加工技術と規則で定められた評価のハイブリッド
 - 加工と評価が一体となったツールはある
 - 加工と評価が一体となった技術（あとで紹介）
 - k-匿名性、差分プライバシー



16

図7 研究してみると：そもそも名前がよくない？

他方、加工と評価がセットで匿名化を行うのであれば、行為と結果が一致するということが言えます。しかし、加工と評価が一体となった匿名化技術はあまりないです。「匿名加工情報」という制度については後半説明しますが、これはこれまでに知られていた加工技術に評価を規則ではめ込むというハイブリッド型になっていると考えられます。なお、加工と評価が一体となったツールはあります。NTTでも開発して販売しています。また、加工と評価が一体となった技術があります。これは面白いもので、後ほど紹介してまいります。

匿名化：加工

ではその加工のところをもう少し説明していきたいと思えます。加工とはデータを元のものに変えて誰が分からなくする、正確に言うとプライバシー情報を得る手がかりをなくすというようなことです。

匿名化：加工手法

- データの値を元のものに変えて、誰が分からなくする*
 - *より正確には「プライバシー情報を得る手がかりをなくす」
- 削除・一般化・模造（・放置）のどれかをする
 - 一般化は値をまるめることで、他の人も該当するようにする
 - 模造は値を実際と異なるものにし、他の人だと思わせる

氏名	住所	性別	年齢	趣味
佐藤	東京都新宿区	男	45歳	マラソン
鈴木	東京都三鷹市	男	41歳	マラソン
山本	千葉県船橋市	男	32歳	宇宙旅行

氏名	住所	性別	年齢	趣味
削除	一般化	放置	一般化	個別対応
**	東京都	男	40代	マラソン
**	東京都	男	40代	散歩
**	千葉県	男	30代	模造



18

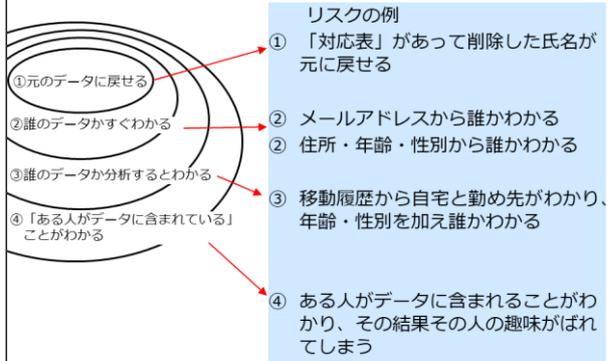
図8 匿名化：加工方法

このようなデータの項目(図8)があったときに、削除、一般化、そして模造と書いてありますが、模造というのは例えば「マラソン」を勝手に「散歩」と変えることです。マラソンと散歩で個人の識別性が変わるかどうか微妙だとは思いますが、嘘をつくことでその人が誰か足がつきにくくするテクニックがあります。あとは匿名化といっても全部の値をいじるわけではなく、そのまま放置しておくといったこともあります。加工はこれだけです。

匿名化：評価

評価は先ほど言ったように、どの程度安全であるか、というようなことです。安全性を確かめるためにはどういったリスク、どういう状況で困ったことが起きてしまうのかということを知する必要があります。

匿名化データからプライバシー情報がもれるリスク



21

図9 匿名化データからプライバシー情報がもれるリスク

匿名化データの安全性評価

- プライバシー情報が漏れるリスクがないこと
- これまでの匿名化の評価手法研究は、リスク②対応が軸
 - これを一般化した考えが「k-匿名性」
- ビッグデータ・IoT時代で、リスク③対応が必要に
 - 一般化が困難
 - この問題に答える期待が「差分プライバシー」に（④もOK）

	リスクの例	リスク対応の評価方法
①元のデータに戻せる	「対応表」があって削除した氏名が元に戻せる	「対応表」がないこと（※数値評価外）
②誰のデータかすぐわかる	メールアドレスから誰かわかる 住所・年齢・性別から誰かわかる	この対応・評価方法の代表が「k-匿名性」
③誰のデータか分析するとわかる	移動履歴から自宅と勤め先がわかり、年齢・性別を加え誰かわかる	※数値評価対象だが一般化困難
④「ある人がデータに含まれている」ことがわかる	ある人がデータに含まれることがわかり、その結果その人の趣味がばれてしまう	この対応・評価方法の代表が「差分プライバシー」



22

図10 匿名化データの安全性評価

ここに「匿名化データからプライバシー情報がもれるリスク」を書いてみました(図9)。リスクを便宜上4つに分けてみます。まずは①、真ん中に書いてありますが、元の個人データにすぐ戻せるものです。これはあつという間にプライバシー情報がもれてしまいます。

②は元のデータにすぐには戻せないものの、メールアドレスなどがあれば誰のデータかすぐ分かってしまうものがあります。

③はすぐには分からないが一生懸命分析をすると分かるというものです。例えば移動履歴を調べていくと自宅と勤め先が分かってしまうので、あとは年齢・性別が分かると誰か分かってしまうかもしれないというシナリオです。

④は誰が分からなくてもある人がそのデータに含まれていることが分かってしまうことによるリスクがあります。これはどういう都合が悪いかというと、趣味がばれてしまうと例に書きましたが、例えばある特定の狭い地域でコロナが500人発生したということが分かると、そこに住んでいる人は高い確率でコロナにかかっているのではないかとということが、誰のデータか

分からなくても知られる可能性があることが危惧されます。

では、その安全性評価はどうしているかというと、基本的にはこういったリスクがないようになっていることが安全性評価ということになります(図 10)。これまでの匿名化の手法の研究はリスク②の対応が中心でした。すなわち、誰のデータかすぐ分からないように加工するというものです。ビッグデータやIoT時代になってきて、リスク③の対応が必要になってきたと考えられます。

例えばたかさんの履歴データがあれば、そこから何かいろいろなことが分かってしまうので、それらに対応する必要があります。しかし、その対応方法を一般化するのはかなり難しい問題だと考えられています。この問題に答える技術の候補として「差分プライバシー」が期待されています。

リスク①対応の評価

- リスク①：元のデータに戻せる
- 評価：対応表がないこと

対応表の例

- 例A：データ作成時に○行目は誰であるか記録しておいた
 - この表があれば、どのような加工をしても元に戻せる
- 例B：一般に流通している何か名簿があった
 - 住所や年齢をより加工していれば、マッチングできないことも→リスク②

氏名	性別	年齢	住所	購入品
例A	男	36	東京都中央区	ガム
	男	22	東京都港区	水

匿名化データ

氏名	性別	年齢	住所
1行目			A山一郎
2行目			B川二郎

対応表 (データ作成時に残した場合)

氏名	性別	年齢	住所
例B	男	36	東京都中央区
	男	22	東京都港区

匿名化データ

氏名	性別	年齢	住所
B川二郎	男	22	東京都港区
C野三郎	男	45	東京都豊島区

対応表 (何か名簿があった場合)

NTT 23

図 11 リスク①対応の評価

それでは少し細かいところを説明してまいります(図 11)。まず、元のデータに戻せることへの対応。これは例のように、匿名化したデータと対応表をマッチングさせると元にもどりますので対応表がないようにします。

リスク②対応の評価

- リスク②：誰のデータかすぐわかる
- 評価：誰のデータか「すぐわからない」ようにデータが加工されている

- k-匿名性がその代表評価指標
 - データを加工して、同じような人がk人いるような状況を作る
 - kはパラメータ (大きい程安全、1ではだめなことがある)
- k-匿名化 (方法)
 - 単体で「すぐわかる」項目を削除 (識別子と呼ぶことがある)
 - 組み合わせて「すぐわかる」項目のデータを「k人になるように」加工 (一般化や模造) (準識別子と呼ぶことがある)
 - そのまま放置する項目も容認する (そこから個人がわからないという仮定の下)

NTT 24

図 12 リスク②対応の評価

続いて2番目ですが、突き合せでぱっと分からないにしても、誰のデータか住所と年齢ですぐに分かってしまうケースです(図 12)。これに関しては、どのように評価していくかというと、誰のデータかすぐ分からないようにデータが加工されているかを見ます。それを実現するものに k-匿名性というものがあ

ります。これが代表的な評価指標になります。

これは何かというと、データを加工して同じような人がk人いる状態を作ります。kというのはパラメータで、5や10や100といったことを想定します。大きいほど安全です。「同じ人が100人いれば」ですので、おっさんが100人いればおっさんはその中に紛れて誰か分からないであろうというロジックになります。

k-匿名化の方法はここに書いてあるとおりで、例えば住所と年齢のようなものを組み合わせて、1人にならないように、何人かになるようにデータを一般化していくことが肝になります。

k-匿名性：同じような人がk人以上いるようにする

氏名	住所	年齢	購入品
A	東京都中央区	34	パン、ガム、新聞...
B	神奈川県横浜市	26	鉛筆、弁当...
C	東京都渋谷区	38	ガム、アイス、チョコ...
D	神奈川県鎌倉市	22	書籍、新聞、電池、宝石...
E	埼玉県川越市	17	化粧品、あめ、アイス...
F	神奈川県厚木市	23	時刻表、鉄道模型、カメラ...
G	埼玉県浦和市	19	ネジ、ビス、ハンマー...
H	埼玉県大宮市	9	肉まん、ガム、新聞...
I	東京都練馬区	30	コーラ、弁当、雑誌...
J	埼玉県与野市	18	ガム、水、ドリンク類...

削除 加工 放置

氏名	住所	年齢	購入品	k-匿名性 (k=3) を満たした状態
東京都	東京都	30代	パン、ガム、新聞...	3 😊
東京都	東京都	30代	ガム、アイス、チョコ...	
東京都	東京都	30代	コーラ、弁当、雑誌...	
神奈川県	神奈川県	20代	鉛筆、弁当、清菜...	3 😊
神奈川県	神奈川県	20代	書籍、新聞、電池、宝石...	
神奈川県	神奈川県	20代	時刻表、鉄道模型、カメラ...	
埼玉県	埼玉県	未成年	化粧品、あめ、アイス...	4 😊
埼玉県	埼玉県	未成年	ネジ、ビス、ハンマー...	
埼玉県	埼玉県	未成年	肉まん、ガム、新聞...	
埼玉県	埼玉県	未成年	ガム、水、ドリンク類...	

NTT 25

※ 削除/加工/放置の選び方により決まるわけではなく、は加工量が決められる必要がある

図 13 k-匿名性

これが k-匿名化を説明した図になります(図 13)。住所と年齢を加工してそれぞれ3人、3人、4人というようなグループができました。したがって最低でも3人未満には絞込むことができないので、このデータは3匿名性が達成されていることになります。

リスク③対応の評価

- リスク③：誰のデータか分析するとわかる
- 評価：誰のデータか「データを分析されてもわからない」ようにデータが加工されている

- これはなかなか評価が困難
 - 攻撃者 (再識別を試みるもの) のスキルと知識に依存するため
 - スキルや知識に一定の前提を置いて対策を施す
- 参考：匿名加工情報は本リスク対応を求めている
 - 作成規則第4号第5号
 - (4) 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。)
 - (5) 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。

NTT 26

図 14 リスク③対応の評価

続いて3番目のビッグデータの的なものに対する「データを分析されても分からないようにする」という課題への対応を考えます(図 14)。これはなかなか評価が困難です。困難な理由は自明で、攻撃者—そのデータは誰のものか暴こうとする人—のスキルと知識に依存するため一般化が困難です。したがって対応する場合はスキルや知識に「こういう攻撃者を想定しましょう」という一定の前提を置いて、実際に加工の程度を決めてい

きます。なかなか厄介なのですが、匿名加工情報はこのリスクへの対応も求めています。

リスク④対応の評価

- リスク④：「ある人がデータに含まれている」ことがわかる
- 評価：データが「ある人のデータであってもなくても見分けがつかない」ように加工されている
- 差分プライバシー (Differential Privacy) が代表的な指標
 - データを模造 (ランダムなノイズを乗せる) して嘘をつくことで守る
 - 元の値とほとんど見分けがつかないようにする
 - ϵ というパラメータで「正直に答える確率」を決める (ϵ が0に近いと何もれず、 ∞ のとき真正直の答えになる)
 - この方法は、実は個別のリスク対応でなく、普遍性のある対応になる
 - ・ いたちごっこでなく、任意のスキルと能力を持った攻撃への対応
- 差分プライバシーはデータベースへの問い合わせ結果 (統計等) への適用がわかりやすいので、以降その説明をする
 - いわゆる個票データの匿名化へに役立つかは研究段階といえる

NTT 27

図 15 リスク④対応の評価

続いて最後の「ある人がデータに含まれている」ことが分かる状態が避けられているかということに対応していく技術について、少し説明したいと思います(図 15)。これはどういう評価をするかというとなかなか難しいのですが、文字通りある人がそのデータに含まれているかどうかということの評価するものです。この手法は十数年前に考え出されました。対象のデータがある人に対して、その人が含まれていても含まれていなくても見分けがつかないように加工されている場合、そのデータはその人が含まれていることに耐性があるだろうという考え方です。この考え方を定式化したものが差分プライバシーという指標です。

これはある意味、嘘のデータを作って、そのデータのプライバシーを守ろうという仕組みです。元の値とほとんど見分けがつかないようにすることがテクニックになります。この中で先ほどの k -匿名化は k というパラメータがありましたが、差分プライバシーには ϵ (イプシロン) というパラメータがあって、 ϵ の大小で正直に答える確率を決めていきます。学術的には定義がスパッと言い切れているので研究のしがいがあり、非常に研究が進んでいます。

この技術の非常に良いところは、実は個別のリスク対応ではなくて、そのデータとある人の関係がないということが言えてしまうので、攻撃者の知識やスキルに依存した対応といういたちごっこから解放されて普遍性のある対応になると考えられるところです。

差分プライバシーは統計結果への適用がとても分かりやすいので、その例を中心にして次のページから説明をしたいと思います。

ここでは差分プライバシーを理解するために、統計結果からプライバシーが守られるかという問題を考えてみます(図 16)。

「テストの落第人数を先生が発表しました」という問題を考えてみたいと思います。以下の人数が落第であった—男子5人、女子5人—ということで、ここから誰か分かってはいけないという問題です。

差分プライバシーとは (1/3)

NTT 28

図 16 差分プライバシーとは(1/3)

この問題は一見大丈夫そうな気がしますが、実はこのクラスは女子は35人いますが、男子は5人しかいないという状態で、そのことがもし分かると、落第した男子が分かってしまうという不都合が起きます。

差分プライバシーとは (2/3)

NTT 29

図 17 差分プライバシーとは(2/3)

差分プライバシーの基本的な考えは、分析結果にランダムなノイズを乗せて嘘を作って分からなくしようというものです。

右に例が書いてありますが(図 17)、男子が6人ぐらい、女子は4人ぐらい落第したという嘘を作るわけです。これで本当に大丈夫かどうかというと、やはり先ほどの「男子が5人」ということを知っている、これでも大丈夫かなというふうに思うわけです。この技術の目標は、ノイズを乗せて攻撃者がどういう知識を持っていても分からなくしようというものです。

この上の図の説明を少ししておきますと、データベース D を作ったときに、それに対する答えが $Q(D)$ としましょう。そこにノイズメカニズム M を提供してノイズを乗せて6人とか4人を作るわけですが、そうしたときにその $M(D)$ を評価しているわけです。

このページ(図 18)が差分プライバシーの定義となります。差分プライバシーは誰がやっても結局、数式を使わなければ説明ができないようなところがあって、これはいい技術ですが、数式に落ちてしまうから普及はできるかな、大丈夫かなとずっと考えています。このページだけは数式で説明をします。

差分プライバシーとは(3/3)

誰か一人 (Aさん) を置き換えたデータD'を作った時、分析結果にMでノイズを乗せたもの M(D) と M(D') の見分けがつかないならば、M(D) は Aさんのプライバシーを守る。
 任意のAさんに関して成り立つならば、Mは安全。
 # M(D) と M(D') の差が一定範囲内であれば有用 (εがパラメータ)

定義 任意のD, D'の組み、任意のS (SはMの出力空間) に対して、下式が成り立つ時 M はε-差分プライバシーを満たす。

$$\frac{\Pr[M(D) \in S]}{\Pr[M(D') \in S]} \leq e^\epsilon$$

NTT 30

図 18 差分プライバシーとは(3/3)

Aさんを含んだデータDとAさんを除いたデータD'というデータを2つ作ってみました。それに関して両方にノイズを乗せたデータを作ったとき、「ノイズを乗せた AさんありのデータとAさんなしのデータがこのような不等式で押さえられるときにε差分プライバシーを満たす」が定義になります。こういう技術が今、流行ろうとしています。

こちらの性質は先ほど言ったように普遍的な対応ができてプライバシー保護のいちごっこから解放されるという利点があります。問題点は少しとらえにくい所ですね。k-匿名性は分かりやすいですね。なんだかんだ言っても同じような人、おっさんを9人寄せれば何とかそこに紛らわせるだろうという話ですが、これはデータにノイズを乗せて、そのノイズの乗せ方を評価してあげて、かくかくしかじかでどう安全である、と評価するようなものなので、ちょっと直感的にはいかんとも説明しがたいところはあります。いずれにせよこういう技術が出てきていて、私たちはこの研究もしています。

NTTの匿名化研究

- 匿名化に新しい安全性「Pk-匿名性」を加えた*1
 - ランダム化のアプローチにおいて「特定個人のデータを1/k以下の確率でしか推定できない」ことを保証する指標
 - 数値に対してノイズを足し合わせて値を歪ませるノイズ付加
 - カテゴリ値に対して確率的にランダムな値に置き換えるランダム化
- 「匿名加工情報作成支援ソフトウェア」を開発
 - NTT(2014)*2
 - 情報通信白書(2020)(右図)*3
 - NTTグループ企業より販売

*1 k-匿名性の強制的な拡張とその適用例 コンピュータセキュリティシンポジウム2009
 *2 ヒソクデータ株式会社による匿名化システム(匿名化システム)を開発
<https://www.ntt.co.jp/news/2014/1402/140207b.html> (2014)
 *3 令和2年版 情報通信白書 第1部 P.257, 総務省 (2020)

NTT 31

図 19 NTT の匿名化研究

匿名化技術の最後のパートですが、NTTでも匿名化の研究をしています(図19)。

暗号化

続いて暗号の話です。暗号は通信の秘匿などにいろいろ使われるわけですが、私はプライバシー保護のために暗号を使うと

いう、ちょっと変わったシチュエーションでの話をします。

プライバシー保護のための暗号化

データを普段は見られないようにしておく

- これまで暗号は、データを利用できる人を制限することに役立っていた
- 暗号研究の発展で、プライバシー保護への貢献の考え方を拡張できる (いろいろ役立つ)
 - 利用できる人を制限 → データの利用条件や利用内容を制限する

NTT 33

図 20 プライバシー保護のための暗号化

暗号というのは、データを普段見られないようにしておく、という考えです。これまでデータは鍵の管理も含めて考えると、データを利用できる人を制限することに役立っていたと考えられます。データ、個人情報を暗号化しておきましょう。鍵をあげると開けて中を見ることができる、鍵がないと見られないという単純な話です。

最近の暗号研究の発展で、いろいろな拡張ができるようになりました。利用できる人を制限するところから、データの利用条件や利用内容を制限するというようなことができるようになってきました。(図20)

暗号のプライバシー保護への貢献

従来の暗号の考え方		進化した暗号の考え方
データを利用できる人を制限	拡張 →	データの利用条件を制限 データの利用内容を制限

- データの利用条件を制限の例
 - ある会社の人がある時間帯ある場所だけで利用でき、この条件以外では利用できない
- データの利用内容を制限の例
 - ある計算だけで、他のことができない
 - データ閲覧自体できない

NTT 34

図 21 暗号のプライバシー保護への貢献

ここに先ほど言ったことを表現しました(図21)。条件や内容というふうに、もっと細かいいろいろなことが暗号によってできるようになりました。

データの利用条件を制限するとプライバシー保護的のどのようになれるかという、例えばこのデータはある会社に所属している社員が、ある時間帯、会社の建物の中だけで利用できて、それ以外では利用できないというようなことができます。

次に利用内容の制限の例では、「このデータを使ってある計算だけでほかのことはできない」、「データの閲覧自体できない」ということができます。これが非常に面白い性質ですから、プライバシー保護に生かすいろいろな実例が作られていますし、もっともっと使われてほしいと考えています。

「データ最小化」と暗号の関係

- データ最小化 (Data Minimization)
 - 個人データの取り扱い原則のひとつ
 - JIS X 9250:2017 プライバシーフレームワーク
 - 処理される個人データを必要最小にし、個人データに触れる人を必要最小にする
- 進化した暗号の2性質はデータ最小化に役立つ
 - データの利用条件を制限
 - データの利用内容を制限

NTT

35

図22 「データ最小化」と暗号の関係

ここで Data Minimization という考え方を少し説明しておきたいと思います。プライバシーフレームワークという標準にもなっているデータのプライバシー保護の原則があります。Data Minimization とはその原則の一つで、処理されるデータを必要最小限にして、かつそれに触れる人も必要最小限にするという考え方です。これが非常にプライバシー保護に効くと言われていいます。(図22)

データの利用条件の制限

例) データ x が「ある会社の人がある時間帯ある場所で」だけ利用できる



- 利用条件が満たされた時のみ復号できる暗号が作れる
 - 暗号の仕組みで $f(x)$ を実行できる
 - 例えば $f(x)$ が真の場合 $f(x)=x$ が得られ、偽の場合何も返さない
 - これを関数型暗号という
- この性質を用いて、データの利用条件の制限ができる

NTT

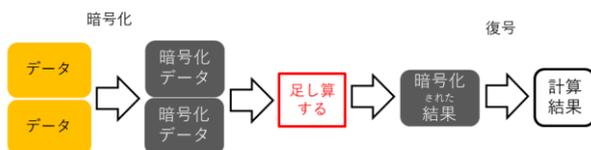
36

図23 データの利用条件の制限

では、利用条件の制限の話をしたしたいと思います(図23)。例えばデータが「ある会社のある時間帯だけ」利用できるといようなものは、関数型暗号というものを使って実現することができます。

データの利用内容の制限

例) 利用を足し算だけに制限する



- 暗号化データ同士の「足し算」をする
 - 一般に暗号化データだけで演算ができる性質を持つ暗号がある
 - これを暗号の準同型性という
- この性質を用いて、データの利用内容の制限ができる
 - ある計算だけで、データの閲覧を含む他のことができない

NTT

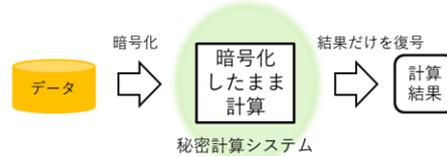
37

図24 データの利用内容の制限

また、利用内容の制限ということ(図24)、実は暗号の一部は暗号化データ同士の「足し算」をするようなことができます。これを暗号の準同型性といいます。この準同型性を使うと暗号化したまま計算ができますし、それをうまく使うことで利用内容の制限ができるようになります。

秘密計算とは

データを暗号化したまま計算までさせ、プライバシーを守る (利用内容の制限)



	暗号化したまま行える範囲	
一般的な暗号	データの通信・保存	-
秘密計算	データの通信・保存	データの計算

NTT

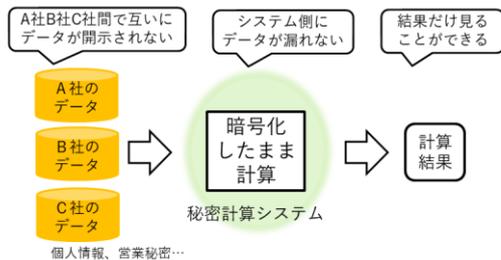
38

図25 秘密計算とは

秘密計算という技術を紹介します。秘密計算というのは暗号化のアプローチの1つで、利用内容の制限をするということに役に立つことができます。秘密計算は何ができるかというと、ここに書いているように暗号化したまま計算して、その結果だけを元に戻すということができます。(図25)

秘密計算の利点

- 計算結果以外は誰にも見えないデータ運用
 - 大切なために流通できなかったデータの新しい統合分析が可能に



NTT

39

図26 秘密計算の利点

秘密計算の利点にはいろいろなことがあります。とにかく暗号化したままデータを集めることができるので、例えば複数の違う会社のデータを集めることもできます。A社・B社・C社で暗号化してデータを持ち込むので、互いにデータがばれてしまうことはありません。また、計算するシステムで、例えばクラウドの中で計算するとしても、クラウド側にデータがもれることが想定されないという利点もあります。

そして結果だけ見ることができるとい性質を持っています。例えば何か統計を取りたいという目標がはっきりと決まっているのであれば、途中の個人情報は誰も見る必要はないはずですから、結果だけ見ることができるといのは望ましい性質ではないかと考えられます。(図26)

NTTの秘密計算の研究

- 秘密分散に基づくマルチパーティ秘密計算システムを開発
 - 統計計算の例

氏名	住所	性別	年齢	趣味	暗号化
佐藤	東京都新宿区	男	45歳	マラソン	暗号化したまま計算
鈴木	東京都三鷹市	男	44歳	マラソン	
山本	千葉県船橋市	男	31歳	映画	

結果だけを復号

東京の40代男性でマラソン趣味 100人

秘密計算

- 歴史
 - 秘密計算の実用性を確認した基礎研究 (NTT 2005, 2011)*1*2
 - 秘密計算の世界トップレコードを確認した基礎研究 (NTT 2017)*3
 - 秘密計算の世界初の実証 (NTT x JALSG 2012)*4
 - 秘密計算の公的統計への応用 (NTT x 総務省統計研修所)*5
 - 秘密計算でディープラーニングを実現 (NTT 2019)*6

NTT 40

図 27 NTT の秘密計算の研究

少し細かい話になりますが、私たちは秘密分散に基づくマルチパーティ秘密計算システムという秘密計算を作っています(図27)。このように個人データを入れると、暗号化したまま計算して結果だけ、「東京でマラソンが趣味の人は 100 人いる」というようなことがぽこんと出てきます。下半分はNTTの業績になりますが、NTT 研究所では 15 年以上前、この技術があまり知られる前から継続して秘密計算の研究を行なっています。

匿名化技術が法制度の一部となった経緯

それではテーマBで匿名加工情報に関する話をしていきたいと思います。

匿名加工情報とは

- 匿名加工情報とは、特定の個人を識別することができないように個人情報を加工し、当該個人情報を復元できないようにした情報のことをいいます。
- また、匿名加工情報は、一定のルールの下で、本人同意を得ることなく、事業者間におけるデータ取引やデータ連携を含むパーソナルデータの利活用を促進することを目的に個人情報保護法の改正により新たに導入されました。

【例】医療情報の例

お客様 → 病院・調剤薬局 → 第三者 → 医療技術の発展 / AI活用 / 創薬研究

NTT 個人情報保護委員会より https://www.ppc.go.jp/personalinfo/tokumeikakouinfo/ 42

図 28 匿名加工技術とは

匿名加工情報とは何かというと、図(図28)の一番下を御覧ください。例えば医療情報が集まったとき、第三者と書いてありますが、いろいろな医療研究をしたり薬を作ったりというときにデータを集めたいので、それを匿名化して自由に使う仕組みはないかというニーズに応じて作られた制度になります。匿名加工情報の内容の説明はしませんが、個人情報を加工して誰か分からないようにするというのが基本的な考え方です。

もう少し法律の話をします(図 29)。匿名加工情報制度には、匿名加工情報を作る事業者に対して実際に何々をしなさいという4つの義務があります。

匿名加工情報制度での事業者義務 (主に4つ)

- ① 適切な加工
- ② 安全管理措置
- ③ 公表義務
- ④ 識別行為の禁止

特にここが技術者の貢献があったポイントです
詳しくはこちら

今日は「経緯」を中心に

- 適切な加工に関するもの
 - (1) 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(置換を含む。以下同じ。)すること。
 - 例→氏名は削除
 - (2) 個人識別符号の全部を削除すること
 - 例→顔画像、指紋等
 - 個人識別符号は全て法令に定められています
 - (3) 個人情報と他の情報とを連結する符号を削除すること
 - 例→事業者内で個人情報を分散管理してデータベース等を相互に連結するために割り当てられているID等は削除する。
 - (4) 特異な記述等を削除すること
 - 例→年齢16歳のように、国内で数人しかいない場合など。
 - (5) 上記のほか、個人情報とデータベース内の他の個人情報との差異等の性質を勘案し、適切な措置を講ずること

NTT 個人情報保護委員会より https://www.ppc.go.jp/personalinfo/tokumeikakouinfo/ 43

図 29 匿名加工情報制度での事業者義務

①は適切な加工をしてくださいということですが、この「適切な加工」が私たち技術者の貢献があったポイントです。技術者が法律の改正に貢献できるというのはなかなかない機会でしたので、非常に得がたい経験をさせていただきました。①の内容は下の(1)~(5)になります。これは法律からのコピーに近いものですが、この中身ではなく経緯の話を今日はしていきたいと思えます。

匿名加工情報が導入された経緯

- 2015 平成27年改正個人情報保護法で導入
 - 匿名加工情報
 - 一定のルールの下で、本人の同意を得ることなくデータを活用できる
- 2013 IT総合戦略本部の検討会で検討
 - パーソナルデータに関する検討会
 - 技術検討ワーキンググループ
 - 直前に 交通系ICカード履歴事案あり
- 2007ごろ 経産省の研究開発プロジェクトで匿名化技術を検討
 - 情報大航海プロジェクト

本資料で説明する経緯は、報告者の体験に基づくもので、抜け・偏りがあることをご容赦願います。

NTT 44

図 30 匿名加工情報が導入された経緯

ざっくり言って 2013 年の「パーソナルデータに関する検討会」が中心にあり、その前にも色々な検討があったというような話をしています。この経緯についてはいろいろと調べてそれなりに客観性がある情報の提供を心がけましたが、所詮私の体験に基づくもので、賞をいただいた受賞講演だから自分为中心でもいいよねみたいな甘えがありますので、抜け、解釈の違いがあることを御容赦いただければと思います。

このページ(図 30)で大事なことは、2013年に政府が本格的に検討を始めたということ、さらにその前にはビッグデータを使おうという世の中の動きがあったということです。後者は役所的には 2007 年の経産省の研究開発プロジェクトがあったのですが、2010 年代前半にビッグデータ活用の取り組みが色々行われました。

その中で、交通系 IC カードで個人情報保護の是非が問われた問題が起きたということもありました。語弊があるかもしれませんが、こういったことも制度を作る議論を深め、広い意味で

貢献しました。この交通系 IC カードの件は、その後の事業社の皆さんの真摯な努力もあって、歴史的に悪い意味ではなくポジティブに振り返る必要がある事案になったと思います。

こういった流れで導入された匿名加工情報の制度をこれから残りの時間を使って順番に説明してまいります。

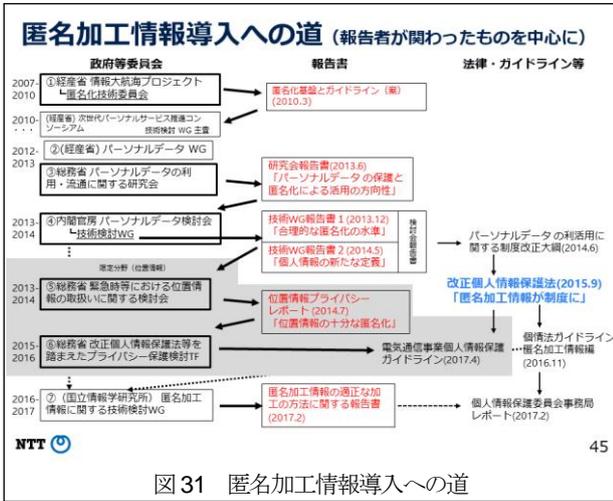


図 31 匿名加工情報導入への道

この図「匿名加工情報導入への道」が本日のメインコンテンツです(図 31)。これは恐ろしく時間がかかりました(笑)。なんの意味があるのかというと、SCAT 大賞をいただいた業績を振り返る場ですので、自分では非常に楽しかったのですが、謎の地図になってしまいました。

見方ですが、右側の青字「2015年の個人情報保護法改正」がマイルストーンで、ここで「匿名加工情報」という技術込みの制度ができた流れを表現しています。そのことを遡ってどうやって技術が制度になったかを正確に評価するのはなかなか難しいところで、それこそ「諸説あります」になります。しかし私が関わらせていただいたものを中心にとというエクスキューズを入れて、こういった流れで捉えていくととても面白いと思います。左側の①～⑦までそれぞれ何が起きたかということを残りの時間で紹介していきます。

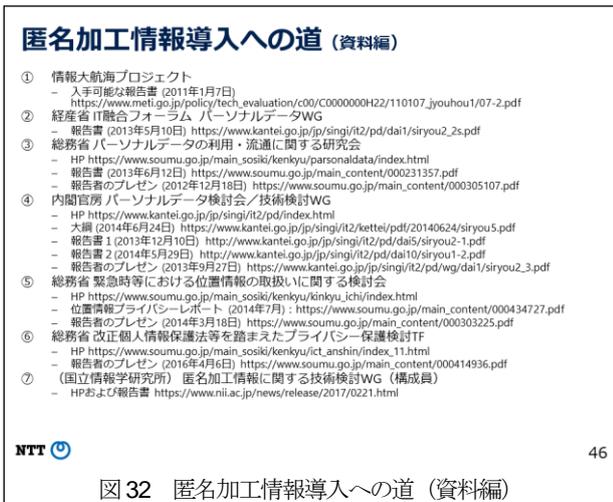


図 32 匿名加工情報導入への道 (資料編)

これは自分の備忘録のようなもので、それぞれの報告書をチェックし直して現在ウェブで取れるものをあげています。(図 32)

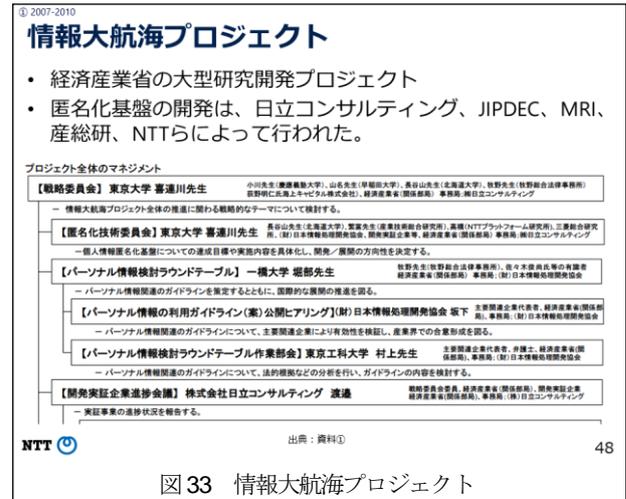


図 33 情報大航海プロジェクト

では時系列に、時代の古いところから入っていききたいと思います。

まず経産省の情報大航海プロジェクトです(図 33)。先の一覧(図 31)の大半はいわゆる国の政策を決めるための委員会ですが、一番上の情報大航海プロジェクトは俗に国プロという、研究開発をするものでした。ただし、研究開発の中で制度のことも一緒に議論していました。匿名化が注目されて期待の対象になった流れはここからスタートしているのではないかと考えています。

匿名化に関して言えば、ここに書いてある皆様と一緒に仕事をさせていただきました。プロジェクト全体は今回 SCAT 会長大賞を受賞された東大の喜連川先生がヘッドとなつて行われたものであります。



図 34 匿名化基盤の開発

いろいろ書いてありますが、とにかくプロジェクト全体の目的はビッグデータを扱えるようにするということでしたので、その基盤となる匿名化技術を開発しました。(図 34)

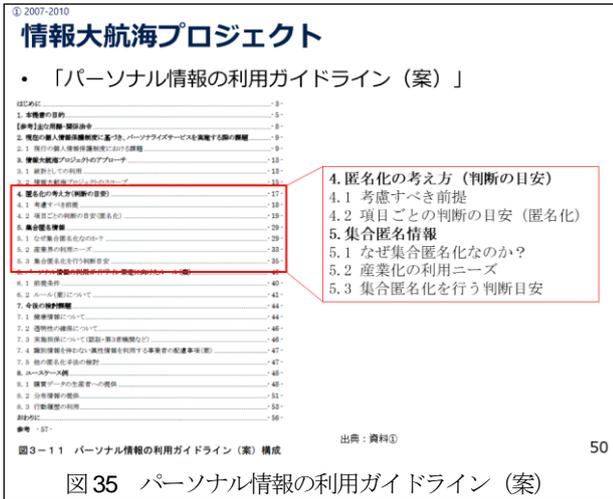


図 35 パーソナル情報の利用ガイドライン (案)

同時にパーソナル情報の利用ガイドラインというものを作られていました(図35)。その中には匿名化の考え方や集合匿名化がいいということが打ち出されています。

集合匿名化というのは別名 k-匿名化というのですが、集合匿名化をすることでデータの安全な活用ができるという方針が打ち出されました。したがって、ここが後々非常に影響を与えたと思います。集合匿名化という考え方が、匿名加工情報の考え方に影響を与えています。

ただし、このガイドライン (案) の (案) のところが重要で、このガイドラインは私の記憶の限り、結局日の目を浴びることはなく未完成のまま終わってしまいました。ですので、後続のいろいろな検討会に部分的に拾われていく、というような形になりました。

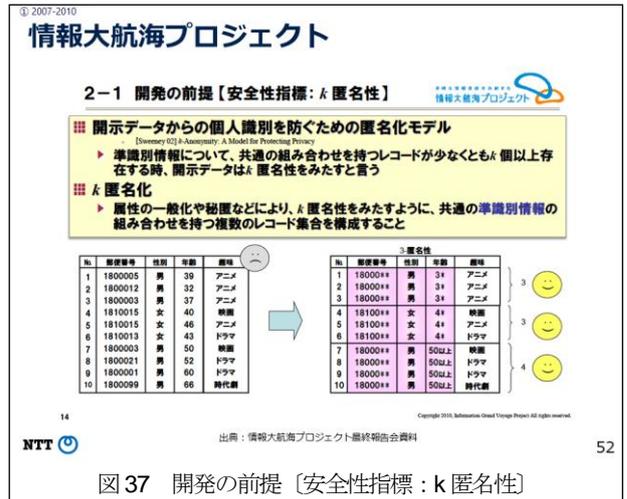


図 37 開発の前提 (安全性指標: k匿名性)

ここで理論的な基盤として k-匿名性に着目したのは正しかったと思っていますが、後々これに苦しめられることにもなります。

続いて(図31)②、経産省のパーソナルデータのワーキンググループがありました。そこに至るまで2年ぐらゐ時間がたっています。その間、実は情報大航海はコンソーシアムを立ち上げて、私もその検討ワーキングの主査などをさせていただきましたが、結構苦戦していました。

そんな中、2012年に総務省と経産省がほぼ同時に検討会を始めました。これが個人情報保護法改正のための内閣官房の検討につながっていきました。

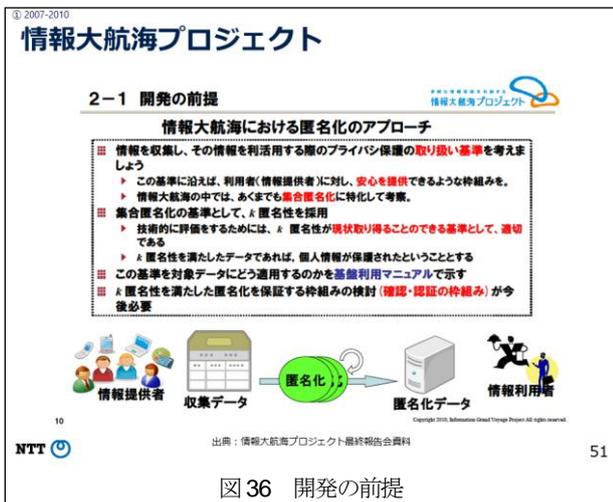


図 36 開発の前提

これはその当時の古い資料を発掘してまいりましたが、このような形で匿名化でデータを活用しようという話です。(図36)

このk-匿名性の説明の図(図37)は10年以上前に描いた図ですが、これはめっちゃくちゃコピーされて人様に使ってもらっているのを私も何度も見たことがあります。これは私が描いたものです。この図は非常に有名になりまして、「k-匿名性推し」みたいなものがここで確立しました。

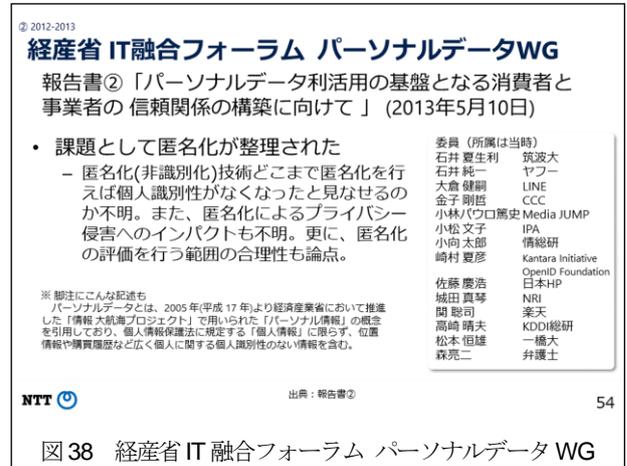


図 38 経産省 IT 融合フォーラム パーソナルデータ WG

まず(図31)②の経産省のほうですが、こういう報告書が出ていました(図38)。こちらの先生方が委員をやっている、これには私は参加していません。課題として匿名化が整理されていて、匿名化について議論する必要があると、はっきりと書かれています。

次に、同時に開催された総務省の研究会(図31)③の話します。

こちら(図39)のメンバー表に私が入っています。これは役所の研究会のデビュー戦で40代でしたが、ほとんど大先生ばかりで、あまり技術者もいないというアウェー感満々というところでやってきました。

© 2012-2013
総務省 パーソナルデータの利用・流通に関する研究会
 報告書③「パーソナルデータの適正な利用・流通の促進に向けた方策」(2013年6月12日)

・ **パーソナルデータ保護のための技術の活用**

- 匿名化技術について、国の統計情報など再識別化を不可能又は十分に困難にしたものは、実質的個人識別性はないといえ、自由に利活用できると考えられる。
- 他の情報との連結等により再識別化の可能性がある匿名化されたパーソナルデータは、米国FTCにおける考え方を踏まえ、次のような条件をすべて満たす場合は、本人の同意を得なくても、利活用を行うことが可能と整理できると考えられる。
 - ① 適切な匿名化措置を施していること。
 - ② 匿名化したデータを再識別化しないことを約束・公表すること。
 - ③ 匿名化したデータを第三者に提供する場合は、提供先が再識別化することを契約で禁止すること。

構成員 (所属は当時)
 糸井雅晴 日本IBM
 若下直行 日立
 岡村久通 NII・弁護士
 奥屋滋 NEC
 菊池公男 富士通
 桑子博行 日本データ通信協会
 新保史生 慶応大
 関野司 楽天
 曾我部真裕 京大
 高橋克巳 NTT
 辻井重男 中央大
 土合幸彦 三豊市
 富沢高明 日本マイクロソフト
 中尾康二 NICT
 長田三純 地協連
 新居真吾 KDI
 別所直哉 ヤフー
 堀部政男 一橋大
 安岡寛道 NRI
 吉川尚宏 ATカーニ
 吉田一雄 経団連

NTT 出典: 報告書③ 56

図 39 総務省パーソナルデータの利用・流通に関する研究会

ここで何が議論されたかという、匿名加工情報のスキームのアイデアの卵がここで書かれています。どうということかという、「再識別の可能性がある匿名化されたパーソナルデータ」を「再識別化しないことを約束すること」を条件付きで活用させたいのではないかとというようなことが、この報告書で出てまいります。ここが後々の匿名加工情報への議論につながっていきます。

© 2012-2013
総務省 パーソナルデータの利用・流通に関する研究会
パーソナルデータの分類例と個人識別の関係

直接個人識別できるのは①に限られるが、②から④でも実質上の個人識別ができる場合がある

- ①-1 個人を識別できる属性を別表に貼り出し、後の対応付けを可能とする(連結可能匿名データ)
- ②-2 個人を識別できるデータを削除する(連結不可能匿名データ)
- ③-2-2の属性をさらに保護し個人を識別しにくくする(高度匿名データ)
- ④ 生データから同じ属性を持つ人数を数え上げる(統計表) / ④ 分析の結果

①生データ(実名データ)

氏名	生年月日	勤務地	趣味
鈴木二郎	1973.10.23	東京都千代田区〇町1	野球
三浦数典	1967.02.27	神奈川県横浜市中区〇町2	サッカー

②-1 連結可能匿名データ

識別番号	勤務地	趣味
1	東京都千代田区〇町1	野球
2	神奈川県横浜市中区〇町2	サッカー

②-2 連結不可能匿名データ

識別番号	勤務地	趣味
1	東京都千代田区〇町1	野球
2	神奈川県横浜市中区〇町2	サッカー

③ 高度匿名データ

氏名	生年月日	勤務地	趣味
鈴木二郎	1973.10.23	東京都	野球
三浦数典	1967.02.27	神奈川県	野球

④分析結果

	全体	東京	埼玉
野球	100	60	40
サッカー	41	33	8

NTT 出典: ③での報告書プレゼン 57

図 40 パーソナルデータの分類例と個人識別の関係

© 2012-2013
総務省 パーソナルデータの利用・流通に関する研究会
まとめ

- ・ パーソナルデータを利用するためのプライバシー保護の技術情報の共有を以下の観点から行った
 - 利用プロセスモデル、データ形式、守るべき原則、技術分類
- ・ 匿名化処理の方法を決めるときに、現実にとどのような危険があるかについても考えておく必要がある。統計情報の場合、住所、氏名が流出することはあり得ない。
 (中略)しかし、もし対象を特定するような試みが実際に行われたら、それはミクロデータ提供の危険性、ひいては統計調査の危険性を指すものとして利用されてしまうであろう。ところが、絶対的な匿名性を担保しようとする、ドイツの経験のように提供できる情報が極めて限られてしまう。したがって、この問題は匿名化処理だけで対策を考えるべきではなく、そのような試みを行うこと自体を制限しておくことが必要となる。このため、データを提供するときには、利用目的を限定し、データの管理を適正に行わせることを義務付けておかなければならない。
 (匿名データの作成・提供に係るガイドライン 改正 平成23年3月28日 総務省政策統括官(統計基準担当)決定 別紙1匿名化処理の考え方)

NTT 出典: ③での報告書プレゼン 58

図 41 総務省パーソナルデータの利用・流通に関する研究会
 まとめ

これはその当時、私がおの委員会ではプレゼンさせていただいたものです(図 40)。今はだいぶ慣れてきましたが、テンパってプレゼンしたのを覚えています。書いている内容はどうということはなく、いろいろな技術的整理をたくさん行いました。

これも(図 41)同プレゼンからですがドイツの統計局の言葉を引用して、実際に匿名化という水準を決めることは結構困難だよということを述べています。

そんなこんなで報告書が出ました。この報告書に、2013年6月に先ほどのことが書かれていて、それを受ける形で内閣官房のほうで、総務省だ、経産省ではなくてオールジャパンでこの検討会(図 31 ④)が立ち上がりまして。この話をしてまいります。ここが今日になります。

© 2013-2014
内閣官房 パーソナルデータ検討会-技術検討WG
 大綱④「パーソナルデータの利活用に関する制度改正大綱」(2014年6月24日)

- パーソナルデータの利活用は、目的外利用や第三者提供において大きな効果をもたらすことから、それらを本人の同意がなくても行うことを可能とする枠組みを導入する。「個人の特定性を低減したデータ」への加工と、本人の同意の代わりとしての取扱いを規定。
- ・ IT総合戦略本部
 - ↳ パーソナルデータに関する検討会
 - 構成員 (所属は当時)
 - 伊藤 清彦 経済同友会
 - 宇賀也 東京大
 - 金丸 英文 フューチャーアーキテクト
 - 佐藤 一郎 NII
 - 穴戸 常寿 東京大
 - 新保 史生 慶応大
 - 鈴木 正朝 新潟大
 - 滝久雄 ぐるなび
 - 長田 三紀 地協連
 - 堀部 政男 一橋大
 - 松岡 萬里野 日本消費者協会
 - 柳田 哲史 経団連
 - 森 亮二 弁護士
 - 安岡 寛道 NRI
 - 山本 隆一 東京大
 - 技術検討ワーキンググループ
 - 構成員 (所属は当時)
 - 伊藤 伸介 明海大
 - 岡村 久和 日本IBM
 - 菊池 浩明 明治大
 - 佐久間 淳 筑波大
 - 佐藤 一郎 NII
 - 佐藤 慶浩 日本HP
 - 高橋 克巳 NTT
 - 松本 泰 セコム
 - 森 亮二 弁護士

NTT 出典: ④大綱ならびに検討会資料 60

図 42 内閣官房 パーソナルデータ検討会-技術検討WG 1/5

私は技術検討ワーキンググループに入れていただきました。親会がこのような形であり、佐藤一郎先生、弁護士の森先生が両方に参加して議論していくという形で、この検討は本当に楽しい検討でした。結局、「個人の特定性を低減したデータ」を加工することで何か新しい取り扱いをしておこうということが決まっています。(図 42)

© 2013-2014
内閣官房 パーソナルデータ検討会-技術検討WG
 ④技術検討WG報告書 1 (2013年12月10日)

- 検討テーマ: 合理的な水準の匿名化とは
- 回答: あらゆる情報について、識別非特定情報または非識別非特定情報への加工を実現する汎用的な技術・手法は存在しない。ケースバイケース。
- 個人識別性に関して技術的に整理。

図 1. 匿名化に関する本WGで定義した用語の関係性

NTT 出典: ④報告書 1 61

図 43 内閣官房 パーソナルデータ検討会-技術検討WG 2/5

どうことをやったかということ、技術検討ワーキングに問いかけられた課題は、「合理的な匿名化の水準を決めてください」ということが来たわけです。数年前に k-匿名性を押した

ことで、ここで苦しめられることとなります。「そういういいものだったら合理的な水準を決められるだろう」と言われたわけです。

それに対して私たちWGが返した答えは、「あらゆる情報について汎用的な技術・手法は存在しない」ので、「何か基準は決められない」ゼロ回答をしました。つまり、いい技術があると言いつつも、この分野に関してはそう簡単にはいかないという押し返す、非常に辛い受け答えでした。内容としては下の図のような細かい検討もして、個人識別性に関するいろいろな認識を高めることができました。(図43)

④ 2013-2014
内閣官房 パーソナルデータ検討会-技術検討WG
 ④技術検討WG報告書2 (2014年5月29日)
 - 検討テーマ：個人情報等の定義等について
 - 検討結果
 ・特定の個人が識別されていない情報であって、特定の個人が識別されるおそれのある情報を「(仮称)準個人情報」と整理を提案
 ・特定の個人を識別する蓋然性の高い識別子を選定
 ・「(仮称)個人特定性低減データ」への加工については、最低限の加工方法であっても、データの種類に応じて加工方法が多様であるから一律の基準を示すことは困難

NTT 出典：④報告書2 62
 図44 内閣官房 パーソナルデータ検討会-技術検討WG 3/5

この技術検討ワーキングはもう1つ報告書を書いていて、これは第2弾です(図44)。ここでは「(仮称)準個人情報」という考えを整理しました。

これは最終的に固まった匿名加工情報よりも個人情報に近いもので、これに対する必要な技術や規律の検討が親会より指示されました。最終的に(仮称)準個人情報は採用に至りませんでしたので、「準個(じゅんこ)はどこへ行った」というネタがありました。この検討の蓄積が匿名加工情報の規則類に活かされることになりました。

④ 2013-2014
内閣官房 パーソナルデータ検討会-技術検討WG
「匿名化」と個人情報の範囲
 ・パーソナルデータの状態と個人情報に該当するかの関係の正しい理解が必要
 ・「匿名データ」は大きく3通りに分類できる

個人情報とされる範囲
 ①実名データ
 ②匿名データ
 ③統計データ
 1. 連結可能匿名データ
 2. いわゆる匿名データ
 3. 高度な匿名データ

※「実名データ」「匿名データ」「統計データ」等の名称は議論用のもので技術用語ではない

NTT 出典：④報告書のプレゼン 63
 図45 内閣官房 パーソナルデータ検討会-技術検討WG 4/5

当時、委員会内でもプレゼンをさせていただきましたがどういいう意味があったかという、啓蒙なんです。一筋縄ではないかなというように細かい技術を説明しながらも、制度化は乱暴に行わないでくださいというような議論をしていました。(図45)

④ 2013-2014
内閣官房 パーソナルデータ検討会-技術検討WG
匿名化技術とは何をする技術なのか？
 会員番号、生年月日、住所、年齢、購買品1、購買品2、購買品3、.....

そのまゝ用いていいか
 加工すれば大丈夫か、削除するか

削除 加工(保護) そのまゝ(非保護)

匿名化技術は、取り決めに従って属性に対して、削除、加工、無加工のどれかの操作を行うこと
 ・活用の際、どの属性をどのように扱うのかを取り決めるのが、個人情報取り扱い責任者の責務
 ・非保護の属性の選定には十分な注意が必要である

NTT 出典：④報告書のプレゼン 64
 図46 内閣官房 パーソナルデータ検討会-技術検討WG 5/5

これも匿名化技術の説明でよく引用された図(図46)だと思いますが、匿名化の基本的な考え方の整理をしています。

匿名加工情報導入への道

2007 ①総務省 情報大航海プロジェクト
匿名化技術委員会
匿名化推進とガイドライン(案)(2010.3)

2010 ②(経産省) 次世代パーソナルサービス推進コンソーシアム
③総務省 利用・流通

2012-2013 ④内閣府

2013-2014 ⑤総務省 緊急時等における位置情報の取扱いに関する検討会
位置情報プライバシーレポート(2014.7)
「位置情報の十分な匿名化」

2015 ⑥総務省 改正個人情報保護法等を踏まえたプライバシー保護検討TF
電気通信事業者個人情報保護...匿名加工情報編ガイドライン(2017.4)
匿名加工情報の適正な加工の方法に関する技術検討WG
匿名加工情報の適正な加工の方法に関する報告書(2017.2)

2016 ⑦(国立情報学研究所) 匿名加工情報に関する技術検討WG
個人情報保護委員会事務局レポート(2017.2)

2017

法律・ガイドライン等
改正個人情報保護法(2015.9)「匿名加工情報が制度に」
個人情報保護法(2015.9)「匿名加工情報が制度に」
個人情報保護委員会事務局レポート(2017.2)

NTT 65
 図47 匿名加工情報導入への道

このように技術検討ワーキンググループの貢献もあって、このパーソナルデータの大綱が作られて、そこから1年を経て2015年に匿名加工情報が出来たということになります(図47)。この灰色のところ(図47 ⑤ ⑥)は全体の流れを受けて、早速並行してやっていたということで、総務省的な電気通信事業のところでは、特に位置情報に関するパーソナルデータの分析をさんざんやりました。これは最終的には電気通信事業のガイドラインの改正までつながっています。

④ 2013-2014
総務省 緊急時等における位置情報の取扱いに関する検討会
 報告書⑤位置情報プライバシーレポート(2014年7月)
 ・位置情報の個人特定リスクを整理
 ・加工方法を整理
 ・「十分な匿名化」
 - 再特定化・再識別化が不可能又は極めて困難と言える程度に加工
 ・例えば「全ての属性に対して、同じ位置情報(移動の軌跡を含む)が複数ある状況を作り出す」
 - 通信の秘密に該当する位置情報に関して「十分な匿名化」を行うことを前提に、有効な同意となる場合を検討。

・電気通信事業者個人情報保護ガイドライン(2017.4)へ

NTT 出典：報告書⑤ 67
 図48 総務省 緊急時等における位置情報の取扱いに関する検討会 1/2

これは位置情報の新しい使い方が匿名化に関してできたという画期的なもので、この整理に基づいて現在のコロナに関する人出の「新宿が何人でした」という空間統計が提供されていると言ってよいと思います。(図 48)

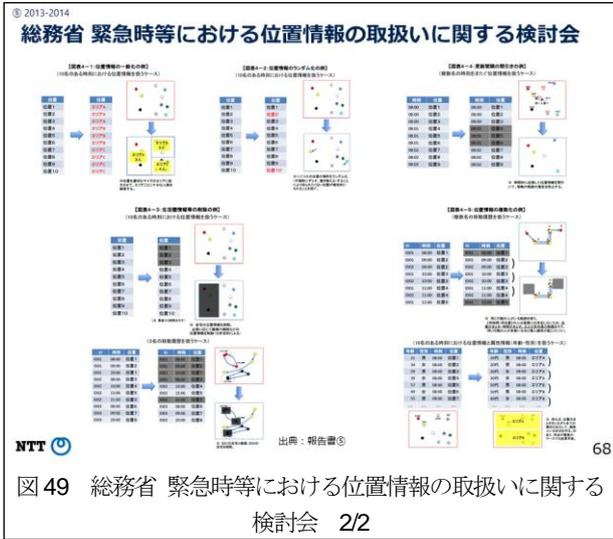


図 49 総務省 緊急時等における位置情報の取扱いに関する検討会 2/2

これはいろいろな図(図 49)が報告書にあって、どういうときにはどういう匿名化が考えられるというバリエーションが山ほどあって、報告書、位置情報プライバシーレポートに描かれています。これは全部私が描いた図です。総務省のエディターの方が本当によく理解している人で、色々なリクエストをくださって苦労したことを覚えています。

ということで、個別分野に関することもめでたく議論でき、こういった物ができてまいりました。

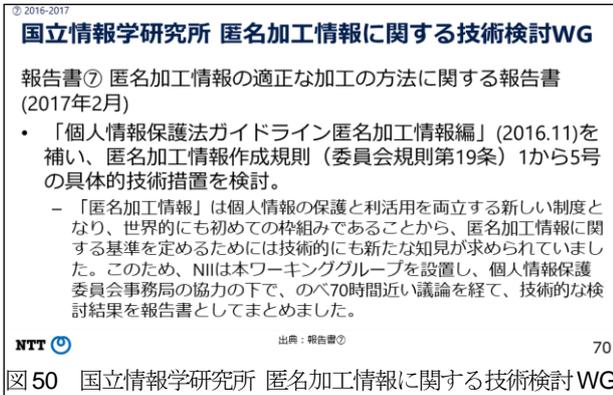


図 50 国立情報学研究所 匿名加工情報に関する技術検討WG

最後に、国立情報学研究所(NII)で検討が行われました(図 50)。これは何をしたかという、個人情報保護法改正に関する匿名加工のルール、ガイドラインができてきましたが、その深掘り、細かいところを埋める作業を個人情報保護委員会の協力のもと、自主的にやりましたという記録です。

まとめと展望

以上で私の話はほぼ終わりです。まとめてまいります。

本講演録は、令和 2 年 10 月 27 日に開催された SCAT 主催「第 108 回テレコム技術情報セミナー」のテーマ、「Society5.0 時代における国民の安心安全を支える研究業績」の講演内容です。

*掲載の記事・写真・イラストなど、すべてのコンテンツの無断複写・転載・公衆送信等を禁じます。

まとめ

- テーマA
データ活用のためのプライバシー保護技術とは何か
 - データを使う時に、そのデータから誰か個人のプライバシーに関わる情報ができるだけもれないようにすること
 - 匿名化と暗号化に代表される
- テーマB
匿名化技術が法制度の **これ** 一部となった経緯 →

NTT

72

図 51 まとめ

2つの話をしました。テーマAは、データ活用のプライバシー保護技術とは何かという話をしました。これは匿名化と暗号化に代表されます。テーマBは経緯ということで、これは一言で言うとこの地図ということになります。(図 51)

今後の展望

- データの側面からのプライバシー保護の研究報告をした
- プロセス側面からの同研究も同様に重要である
 - データが誰から誰にどうやって
- 倫理の側面からの同研究も同様に重要である
 - なんのため
- データをもっと日々の生活や未来の社会のために役立てるため研究を続けます
 - データをもっとすげえく使えるようにしたい
- 技術と制度にまたがる研究の機会と議論や刺激を与えてくださった諸氏に感謝します
- このような研究環境の発展や人材育成に尽力します

NTT

73

図 52 今後の展望

今後の展望です(図 52)。今日はデータの側面からプライバシー保護の研究報告をしました。個人データをどのように加工すればよいかというものです。しかし、プロセスの側面からの研究も重要だと考えています。データを一発で安全に加工して流通できればよいですが、それだけでできることは限られます。より高度なデータ利用やプライバシー保護をするためには、データが誰から誰にどうやっていくかということをちゃんと調べていくことが大事です。それから倫理的な側面も大事だと考えています。データをもっと日々の生活や未来の社会のために役立てるための研究を続けたいと思います。

改めて、技術と制度にまたがる研究の機会と議論や刺激を与えてくださった皆様に感謝して、さらに研究を続けるとともに、これからの研究環境の発展や人材育成に尽力してまいりたいと思います。今日は機会をいただきましてありがとうございました。御清聴ありがとうございました。