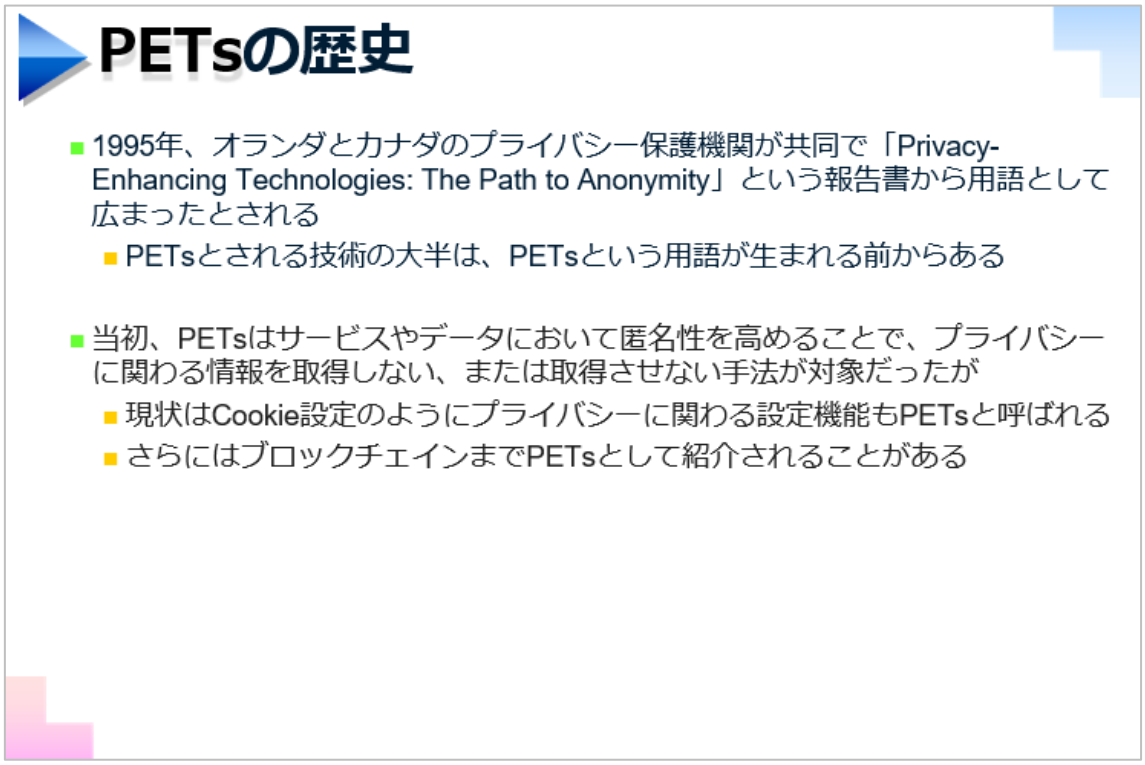
技術講演会第 4 回 Privacy-Enhancing Technologies（PETs）

（国立情報学研究所　佐藤　一郎）

1. Privacy-Enhancing Technologies (PETs)とは

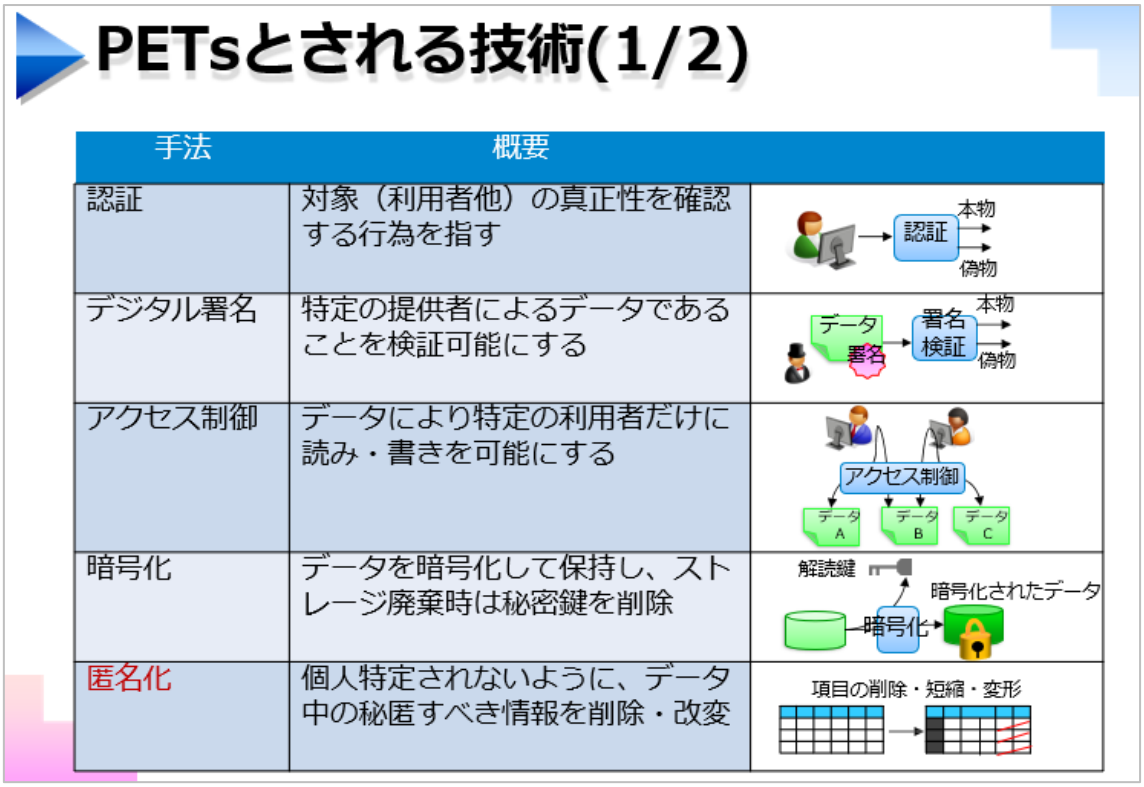
PETs とはプライバシー保護を実現・強化する技術全般を指します。日本語では、「プライバシー強化技術」と訳されることが多くなっていると思います。事業者が自分たちに都合のいいように言っていますし、特に現在は PETs 分野には多くの資金が集まっているため、拡大解釈が多いです。

このため PETs の問題点を指摘すると、スタートアップ企業から反発されてしまいます。



PETs の歴史については、1995 年にオランダとカナダの公的なプライバシー保護機関が共同で出した報告書に初めて登場したと言われています。しかし、PETs と呼ばれる技術自体は古く、100 年以上の歴史を持つものもあれば、ここ数年で登場したものです。この概念は非常に拡大解釈され、ウェブブラウザの Cookie やブロックチェーンも PETs として紹介されることがあり、何が PETs に該当するのかが曖昧である状況です。

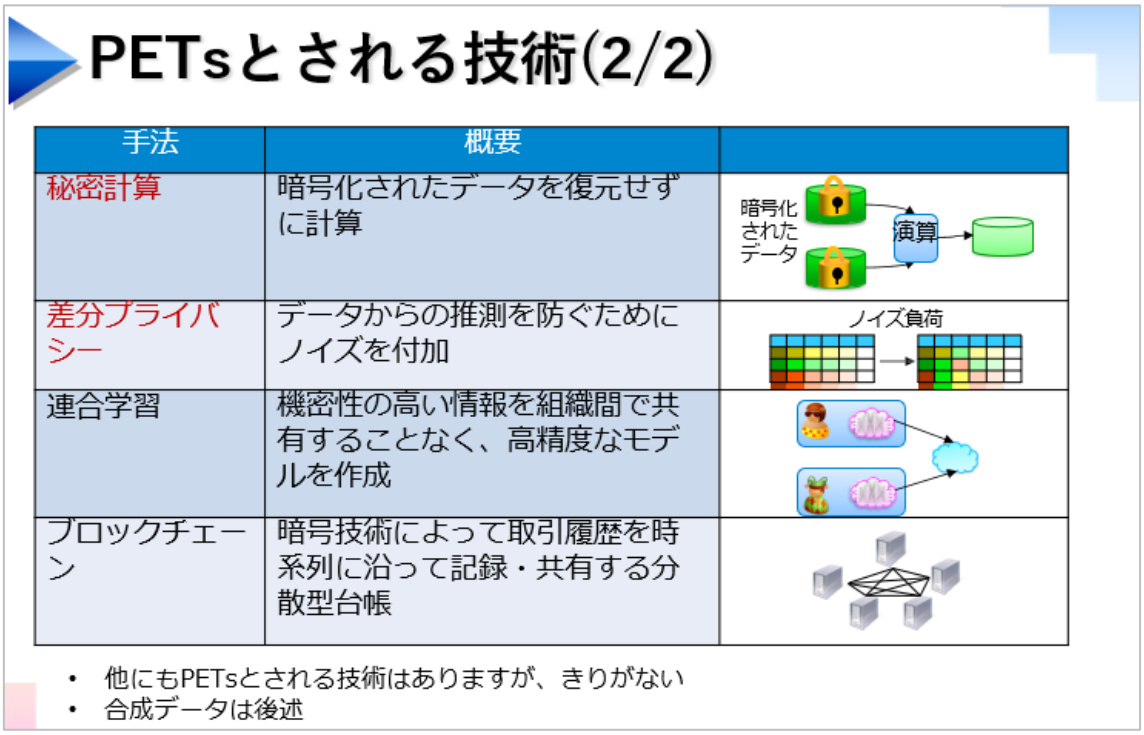
2. PETs とされる技術



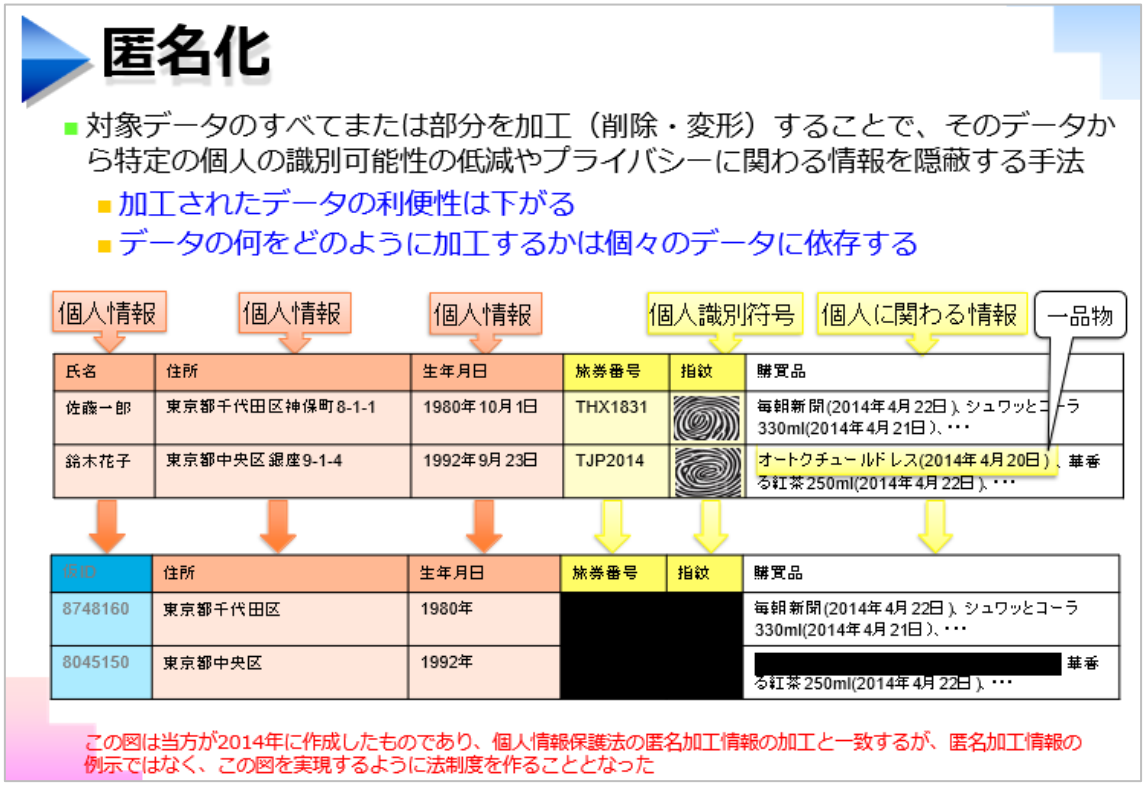
コンピューターにログインするなどの行為やマイナンバーカードによる本人認証、電子透かしやデジタル署名なども PETs と言われることがあります。また、特定のユーザーだけがファイルやデータにアクセスできるアクセス制御も PETs の一例です。

暗号化も PETs の一部です。例えば、ノートパソコン持ち出す際に、セキュリティの観点から SSD やハードディスクを暗号化することがあります。これも PETs に含まれます。理由は、仮にノートパソコンが紛失や盗難に遭ったとしても、暗号化されたストレージが解読されないため、情報漏洩を防げるからです。

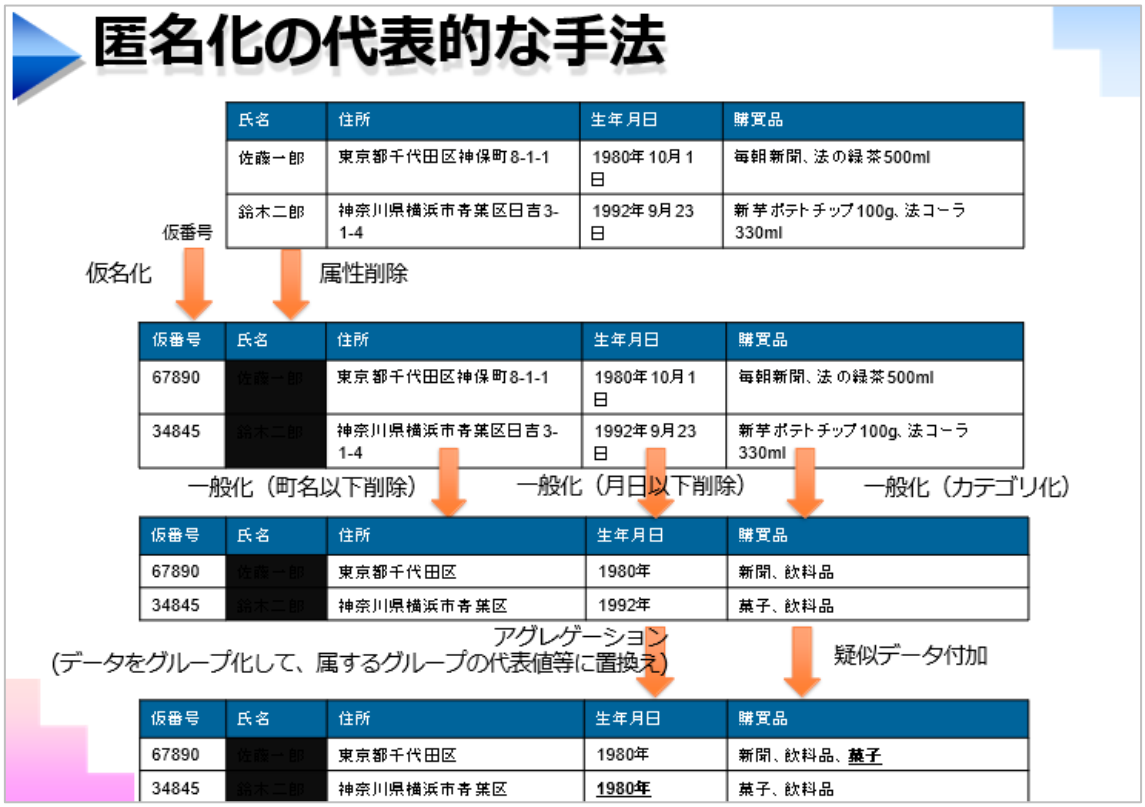
また、匿名化技術も PETs の一部です。例えば、個人情報の観点から特定の個人を識別する情報をデータから取り除くことや、事業者側が重要だと判断するプライバシー情報を秘匿化することを指します。



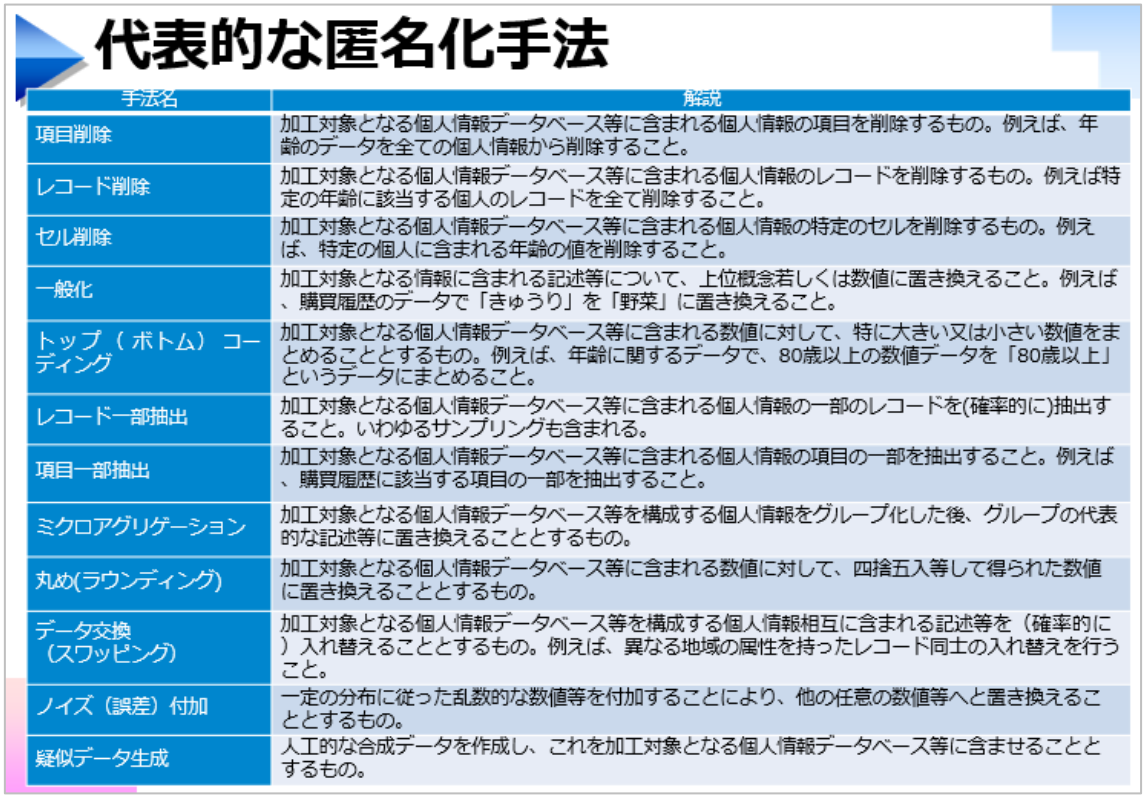
古い意味での PETs の代表例は、秘匿計算、差分プライバシーや連合学習などありますこれらは、秘匿計算や評価方法として利用されています。その他にも様々な手法が存在し、多くの立場から議論されています。また、PETs に含めるかどうかは議論の余地がありますが、合成データも実データを隠しつつ、データ分析を行えるようにするという観点では、PETs として扱われることがあります。



匿名化は、2015 年の個人情報保護法改正で導入された匿名加工情報の概念に基づいています。匿名加工情報とは、氏名、住所、生年月日、旅券番号、指紋、購買品で特定の個人を識別できる情報を加工して取り除くものです。

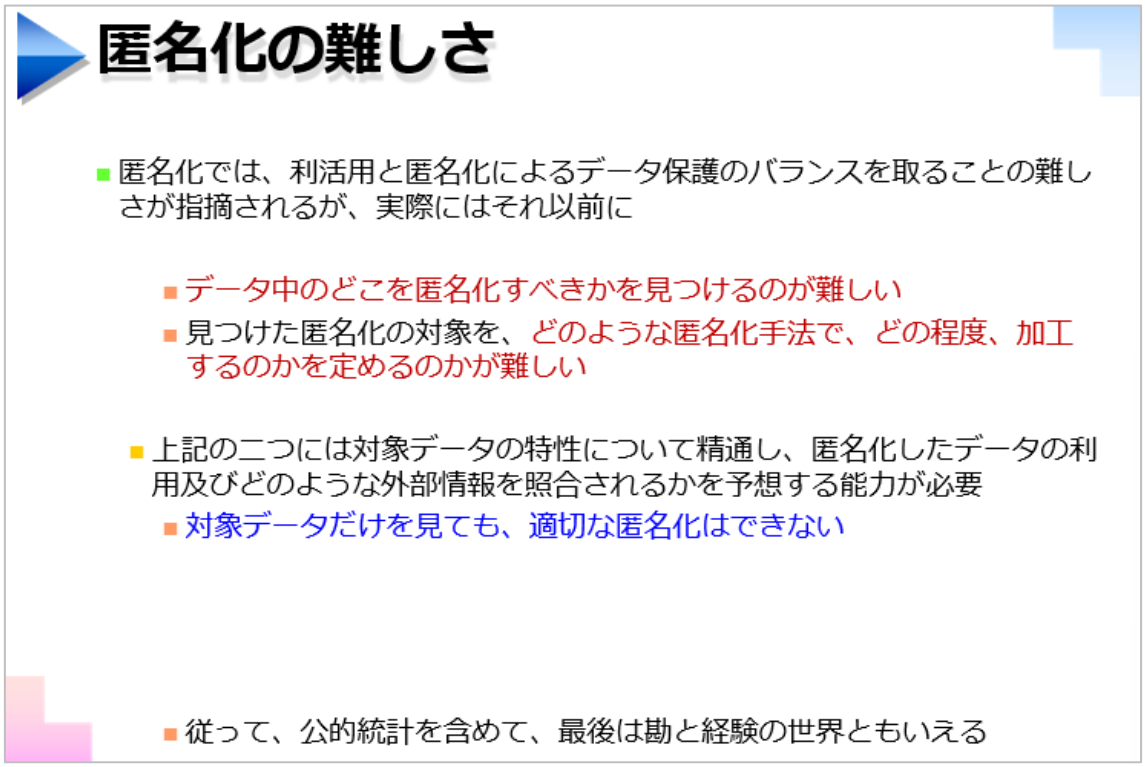


このように特定の個人を識別する情報やプライバシーに関わる情報を取り除く手法が匿名化と呼ばれます。



個々のデータ項目に対する対処方法には様々あります。例えば、特定の項目を削除したり、情報の一部を丸めたり、行や列ごと消去する方法があります。また、トップコーディングは、例えば 100 歳以上の年齢を「100 歳以上」と一括りにします。ボトムコーディングはその逆です。

また、よく使うテクニックとしては、データ交換という手法があります。データを入れ替える限り、平均値などの統計的な影響はあまり出ないため、データの入れ替え、ノイズの付加やダミーのデータで置き換えなど、多様な手法を用いて匿名化を実現しています。

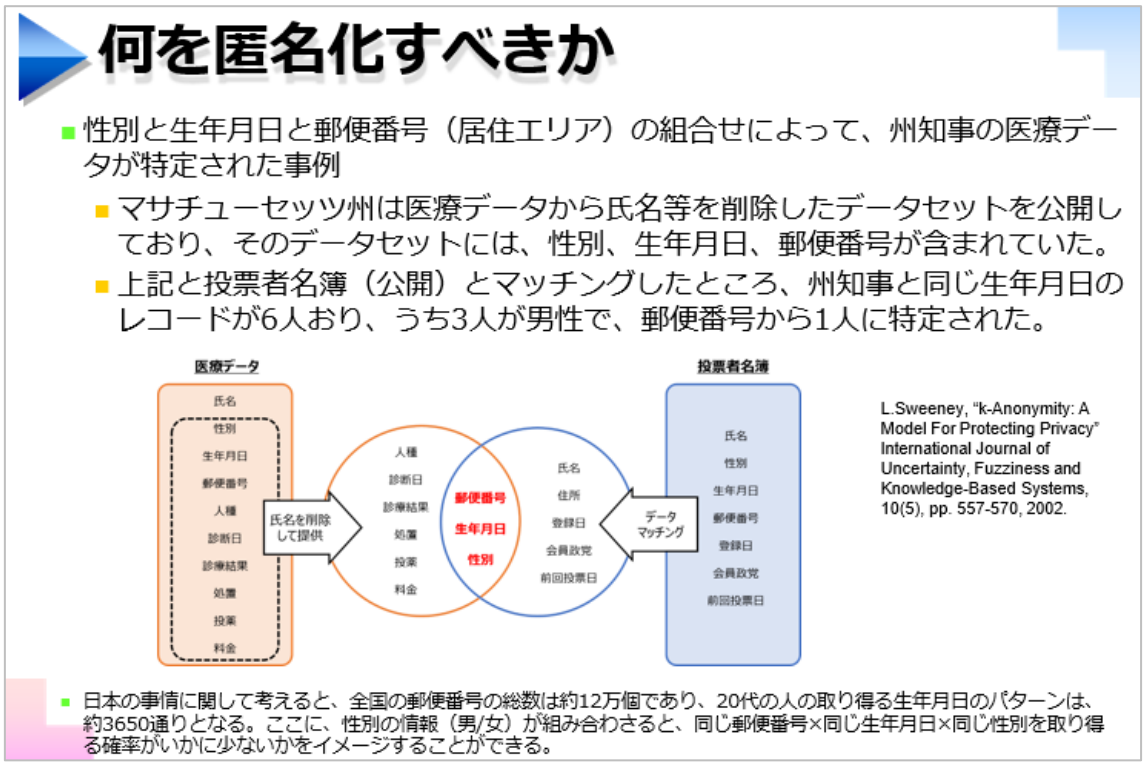


匿名化の難しさは、匿名化を行うと情報は落ちる方向にいくため、利活用が難しくなります。技術者にとっては当然のことですが、匿名化の議論や匿名加工情報の導入前、法学者の中には、匿名化を行っても情報量が減少せず、個人情報の問題が解決すると誤解して、法改正を不要とした方々もおられ、それが個人情報保護法の改正が遅れる一因となったといえます。

実際に匿名化を行う際の難しさは、情報が落ちるだけではなく、名簿や購買履歴のような表形式のデータがある場合、どの部分をどの手法で匿名化するかを決定することもあります。さらに、どの程度匿名化するかは、その情報だけでは判断できません。個人の特定性に関しては、対象の情報やデータ以外の情報、またはデータとの照合によって、特定の個人が識別される可能性があるためです。

特定の個人が識別されるのは、その人が誰かが分かることです。これは氏名だけではなく、他の情報と組み合わせることで識別が可能になる場合も含まれます。そのた

め、どのような外部情報が利用可能であるかを理解した上で行う必要があります。適切な匿名化を実施するためには、高度な専門知識が求められます。単に、利活用をあまり考慮しない匿名化であれば、情報をどんどん削除することは簡単が、利活用を考慮しながら匿名化するには、高度な専門性が必要です。また、AI を利用すれば匿名化が簡単にできるというものでもありません。



さて匿名化の方法ですが、必ずしも決まっておりません。公的統計の分野でも勘と経験に頼る部分があります。

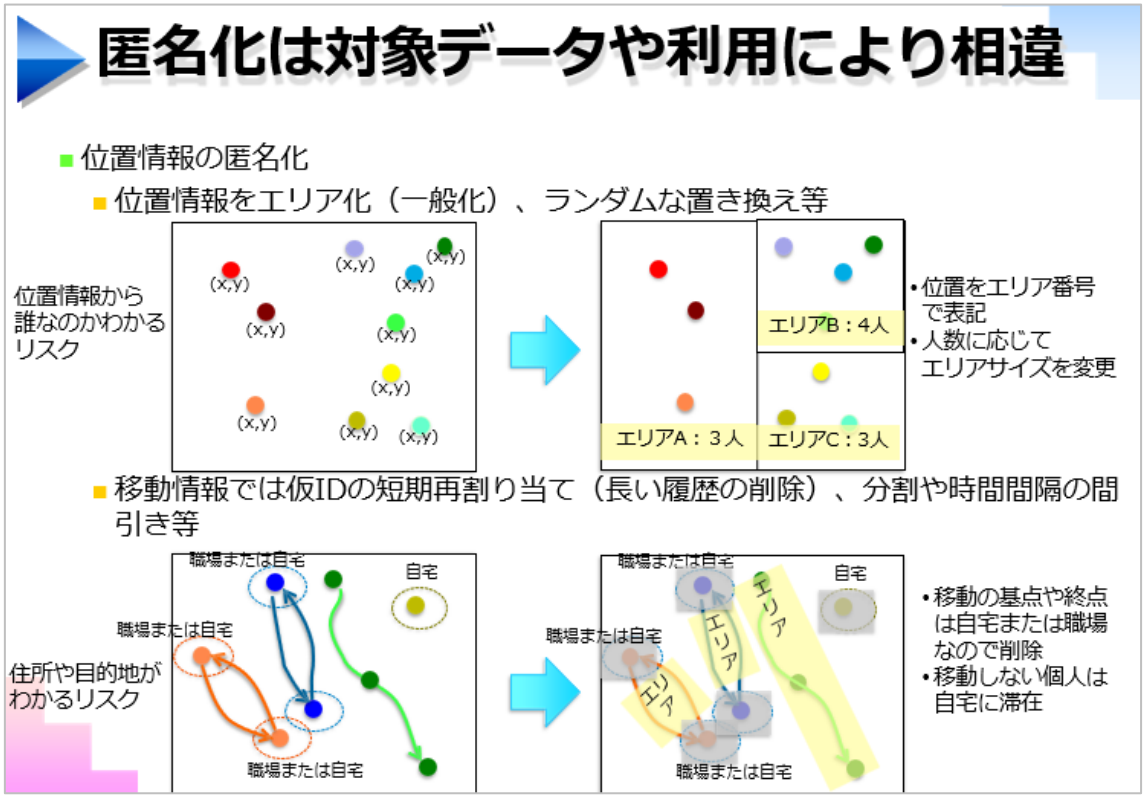
匿名化の難しさが一般に知られたのは、2000 年頃に有名になった論文に、マサチューセッツ州の医療データが公表され、氏名や住所などは削除されているものの、性別、生年月日、郵便番号が含まれています。同時に投票者名簿も公表され、これには名前や生年月日などが含まれています。この二つのデータをマッチングさせることで、多

くの人の医療データ名前を結びつけられる可能性があります。そのため、何をどう組み合わせるかを考慮して匿名化する必要があるのが難しいです。

会員番号	住所	生年月日	年齢	好きな番組	購買品
10001	東京都千代田区神保町･･･	1980年10月1日	33	報道番組	ビール、新聞
10002	神奈川県横浜市緑区鴨居･･･	1992年9月23日	21	バラエティ	ポテトチップ、炭酸飲料
10003	東京都渋谷区神南･･･	1979年10月1日	35	ドラマ	コーヒー、タバコ
10004	東京都江東区牡丹･･･	1990年3月12日	24	ドラマ	化粧品、ヨーグルト
10005	神奈川県川崎市宮前区･･･	1985年7月28日	28	バラエティ	ポテトチップ、炭酸飲料
10006	神奈川県横浜市港北区日吉･･･	1987年4月18日	27	報道番組	コーヒー、新聞
10007	東京都港区白金･･･	1978年12月31日	34	ドラマ	おでん、ヨーグルト
10008	東京都品川区北品川･･･	1985年3月3日	26	バラエティ	おにぎり、お茶
10009	東京都文京区大塚･･･	1989年12月1日	23	ドラマ	コーヒー、おにぎり

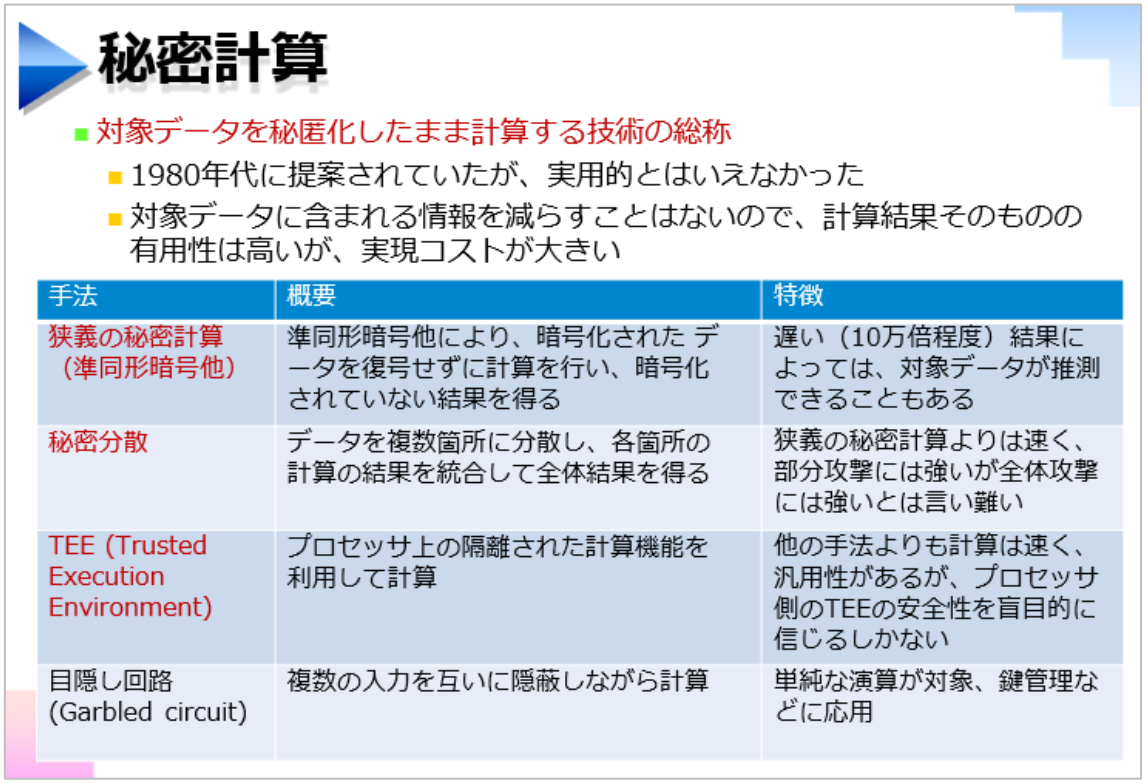
会員番号	住所	生年月日	年齢	好きな番組	購買品
10001	東京都	1980年10月1日	30代	報道番組	ビール、新聞
10003	東京都	1979年10月1日	30代	ドラマ	コーヒー、タバコ
10007	東京都	1978年12月31日	30代	ドラマ	おでん、ヨーグルト
10004	東京都	1990年3月12日	20代	ドラマ	化粧品、ヨーグルト
10008	東京都	1985年3月3日	20代	バラエティ	おにぎり、お茶
10009	東京都	1989年12月1日	20代	ドラマ	コーヒー、おにぎり
10002	神奈川県	1992年9月23日	20代	バラエティ	ポテトチップ、炭酸飲料
10005	神奈川県	1985年7月28日	20代	バラエティ	ポテトチップ、炭酸飲料
10006	神奈川県	1987年4月18日	20代	報道番組	コーヒー、新聞

匿名化の際には、データを丸めたり、削除したりして、複数人が該当するようにすることが望ましいです。これを「K 匿名性」と呼び、匿名化の目標になりします。

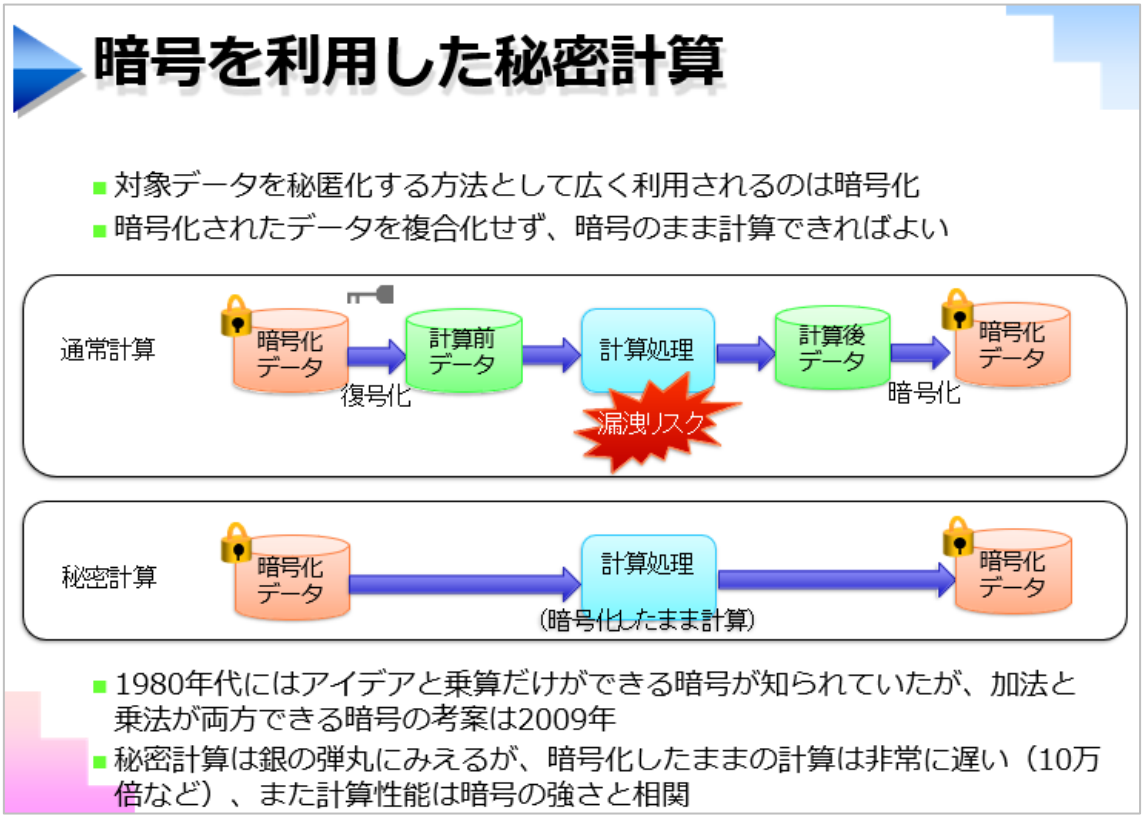


匿名化の方法はデータの種類によって異なります。例えば、位置情報の場合、エリアをメッシュに分けて、同じメッシュ内の人を同じ場所にいるとみなす方法があります。また、移動の情報では、プライバシー保護のために、出発点と到着点の情報を削除する方法などが提案されています。

匿名化には個人の特定以外にもいくつかの問題があります。例えば、非常に潔癖な人がいます。この人は、正確な位置情報が公表される限り、パチンコ店や怪しいエリアを避けることができます。しかし、正確な位置情報はプライバシーに影響することから、位置情報を大まかに匿名化すると、この人があたかもそうした怪しいエリアと同じ場所に扱われてしまい、そうしたエリアに出入りしているかのように見える可能性があります。その結果、誤解を避けるために、この人はさらに距離を置く行動を取り、逆に社会の分断を助長する可能性もあります。匿名化必ずしも万能ではありません、匿名化が社会に影響をあたえる可能性を示しています。

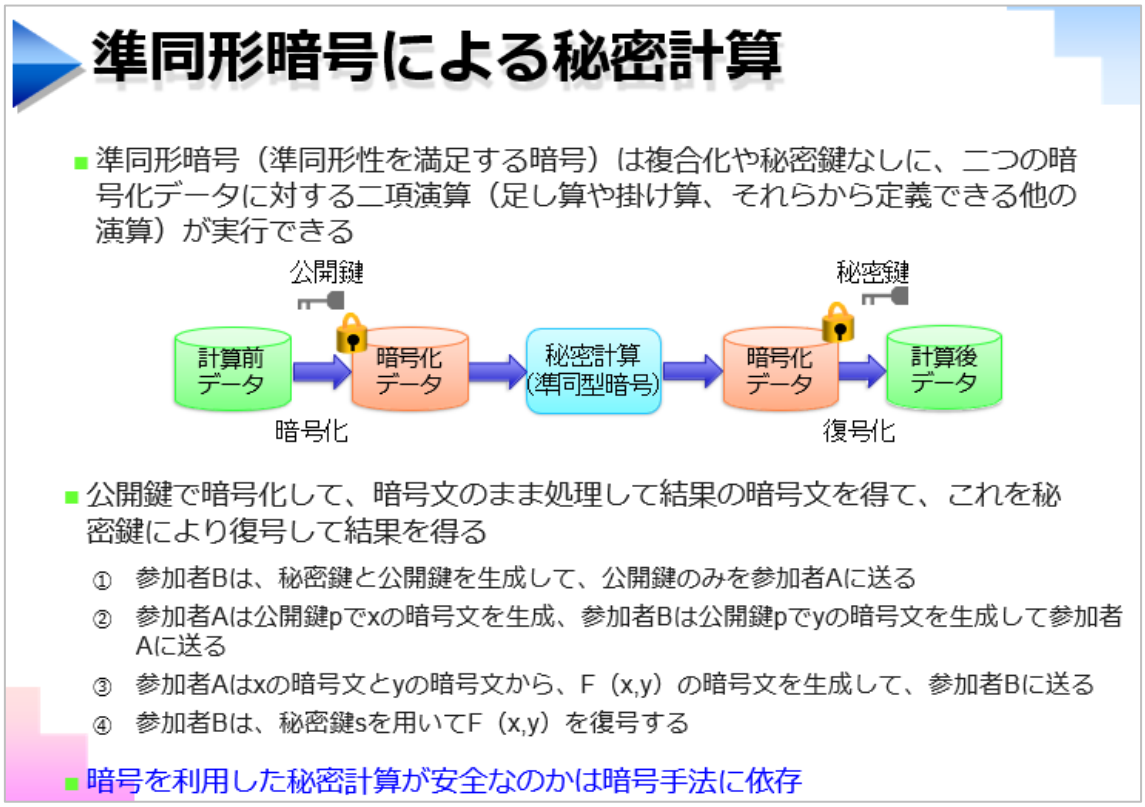


数年前までは PETs といえば、「秘密計算」が主流でした。その秘密計算と呼ばれる技術が多様化しており、代表的なものは四つあります。



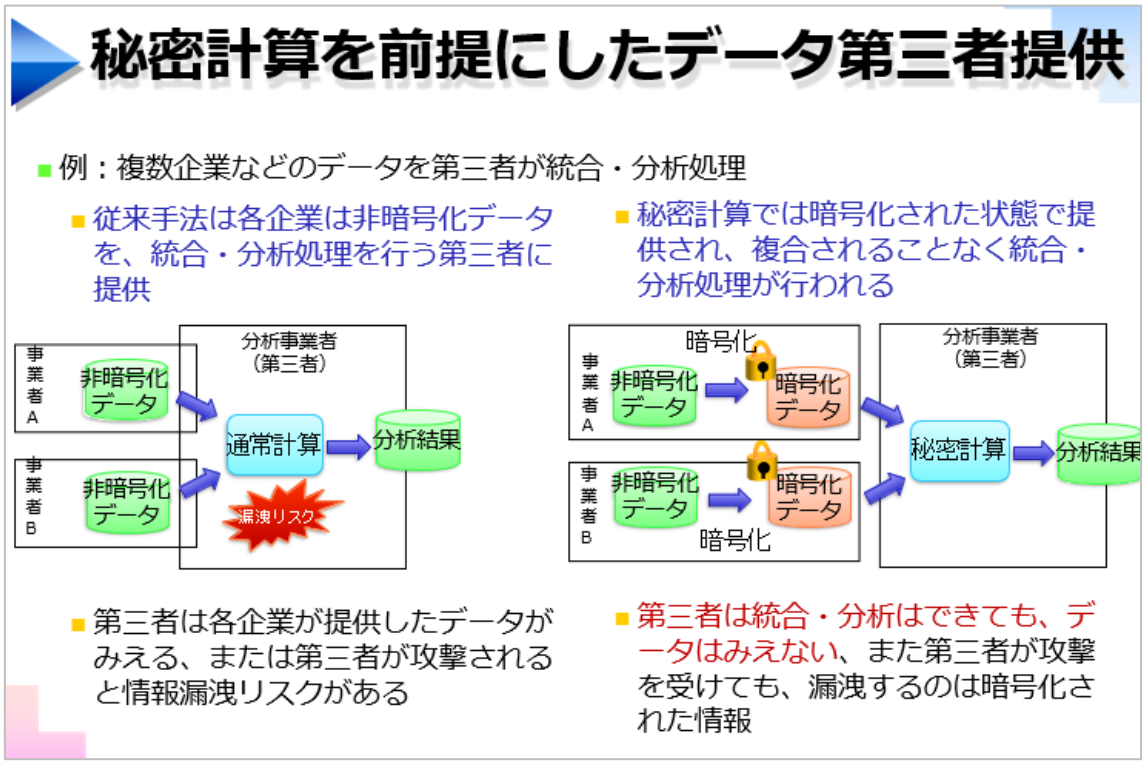
一つ目は、昔ながらの暗号を使った秘密計算です。原理はデータを分析する際に、データそのものを分析するのが理想的だが、データの中身が見えてしまいます。そこで、暗号化したままでデータの分析や処理を行うことで、結果が平文でも暗号文でも、元のデータの秘匿性を保つことは秘密計算の考え方です。

このアイデアは、1980 年代に提唱されたが、実用化されたのは 2000 年代に入ってからです。暗号化されたデータで足し算や掛け算ができるようになったのは 2009 年です。そのため、四則演算が可能になったのは比較的に最近のことです。



暗号手法はさまざまな種類があるが、秘密計算に適しているとされるのが準同形暗号です。準同型暗号とは、暗号化されたデータのまま、掛け算や足し算を行うことができる性質を持つ暗号のことです。準同型暗号の代表例として、RSA があります。公開鍵暗号方式の一つで、暗号化されたデータ同士の掛け算は可能であるものの、足し算はできません。暗号によって特性が異なり、秘密計算に使う場合、足し算と掛け算の両方ができるものが望ましいが、長い間見つかりませんでした。

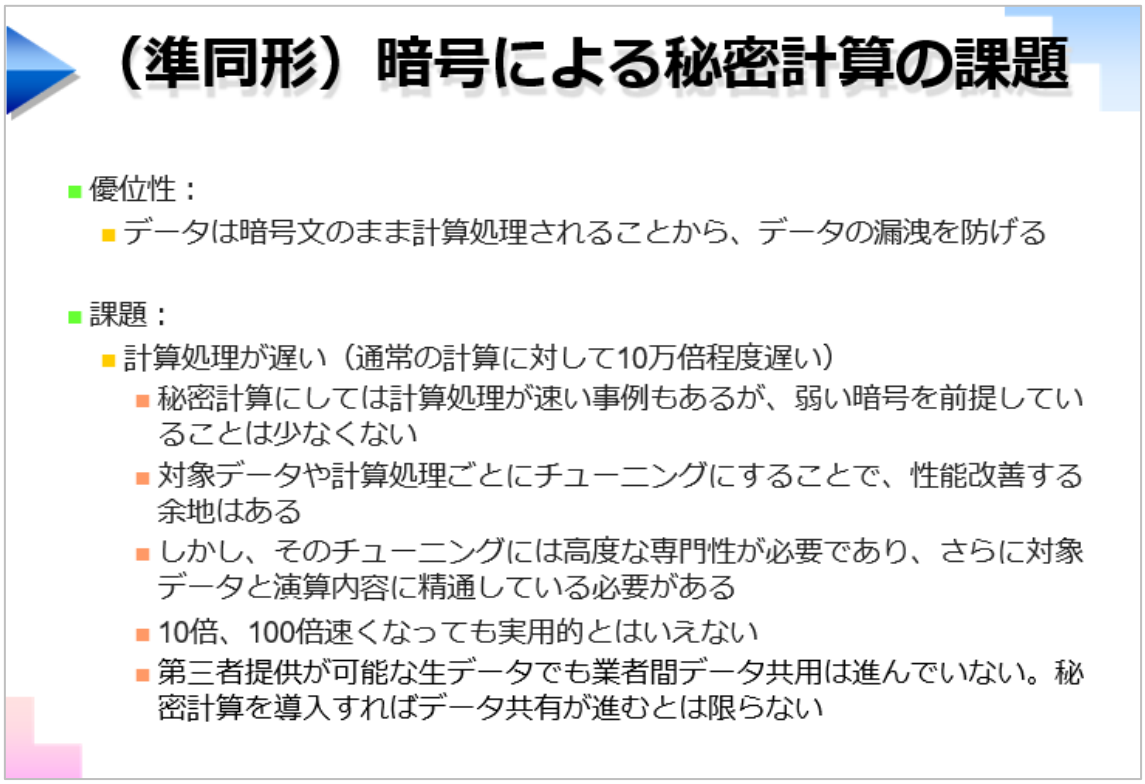
例えば RSA を前提とした場合、元のデータを暗号化したまま計算を行い、結果も暗号化された状態で受け取ります。その後、結果を複号化して、平文として利用する方法です。



二つ図の共通点は、事業者 A と事業者 B がそれぞれ自社データを持っています。このデータは、他社には知られたくないものであり、データ共有が必要になったときのことです。従来の方法では、分析処理を第三者に依頼する際、データは暗号化されていない状態で渡され、第三者はそのデータを分析して結果を返します。

守秘性があっても、事業者は自社の秘密データを第三者に渡すということに抵抗があります。情報漏えいのリスクもあります。しかし、秘密計算を組み合わせると、事業者は暗号化したデータを第三者に渡し、第三者は暗号化されたまま計算を行います。結果は暗号化のままか平文かは方法によりますが、安全にデータを共有できます。分析事業者には暗号化されたデータが渡るが、暗号を解読しない限り、元のデータは分かりません。暗号が強固で鍵が適切に管理されていれば、元のデータは漏れません。

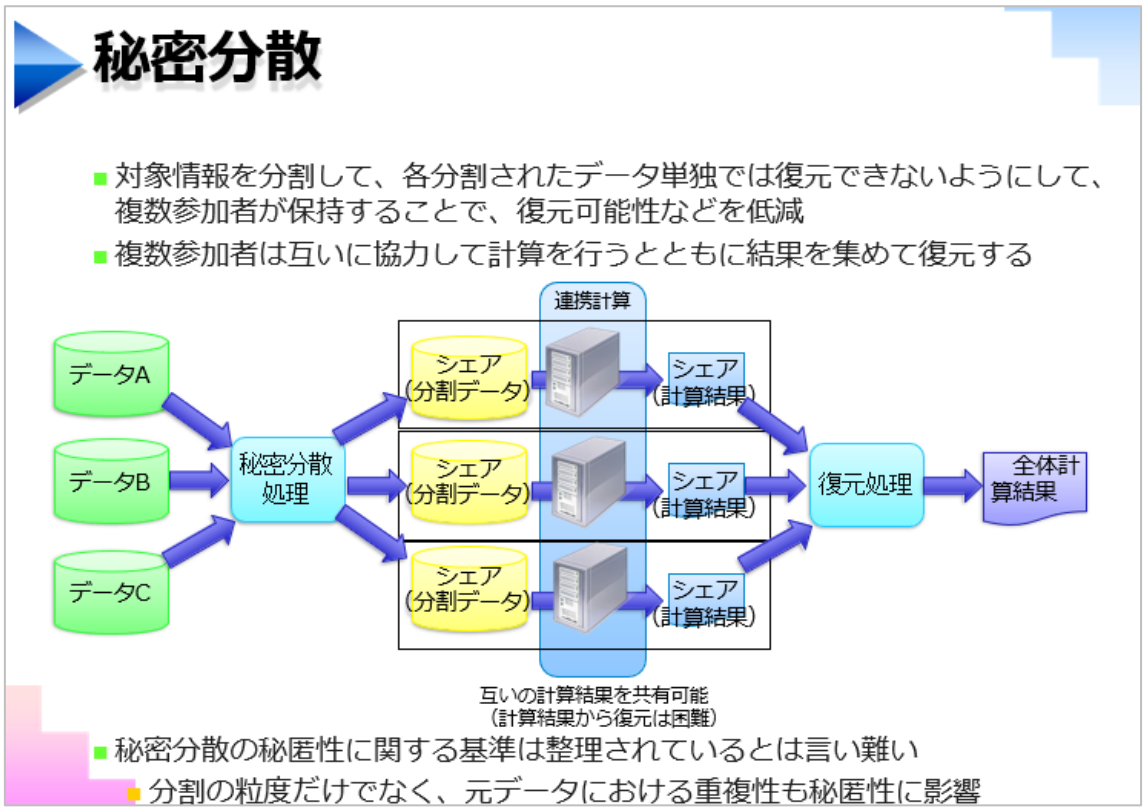
この方法により、事業者間で安全にデータ共有が可能だが、データ処理の速度が遅いという課題があります。約 7~8 年前の状況では、約 10 万倍遅いという評価が一般的でした。



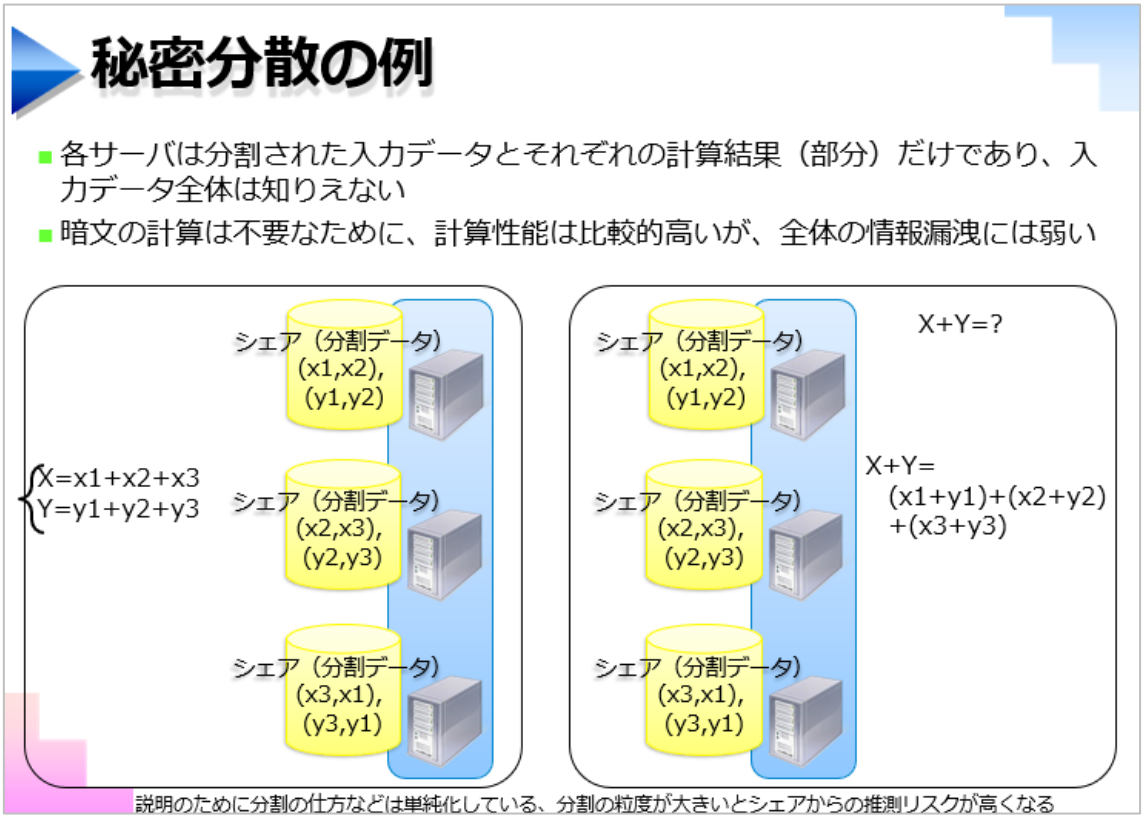
実際には、安全性を犠牲にして速さを追求した暗号が使われることも多かったです。また、単純に暗号化すれば良いではなく、表形式のデータの列方向か行方向で分析するのかによって、暗号化の方法が異なります。分析の方向に合わせて暗号化をチューニングすることで、計算速度を向上させることができます。

しかし、チューニングを行っても速度はせいぜい 10 倍から 100 倍程度の向上で、依然として遅いです。このチューニングには秘密にしておきたい対象データの特性を把握していることが前提になりますし、演算に精通している必要があり、守秘性のあるデータを知り得て、高度な技術者が必要です。

PETs を扱うスタートアップは増えています。その技術者がトイプロブレム的に性能を向上させても、実サービスで実データに対応できるかは不確定です。その上、実データは日々変わるため、最初は良くても段々遅くなることがあります。そのため、秘密計算を導入しているスタートアップは多いものの、個々の事例ごと、さらに頻繁にカスタマイズや最適化が必要となってしまい、多様な事例に適応できるという点ではスケールしない技術となっており、収益化が難しいのが実情です。



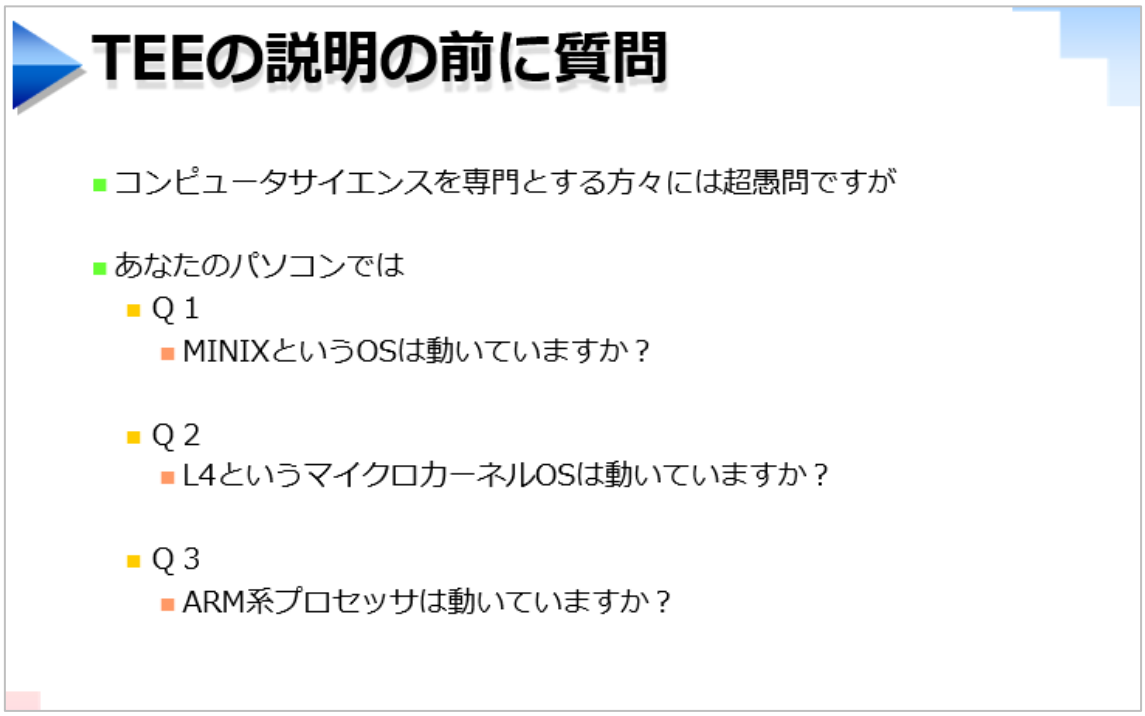
また、データの保持方法を工夫して秘匿性を高める「秘密分散」という方法があります。資料に、データ A、B、C と書いてあり、その対象データを細かく分割します。表形式のデータであれば、セルごとに分割するだけでなく、セル内も特殊な方法で分割します。分割したデータを散らしてストレージに保存し、散らした状態で計算を行います。その結果を集めて全体の計算や分析結果を作り出すのが秘密分散の方法です。



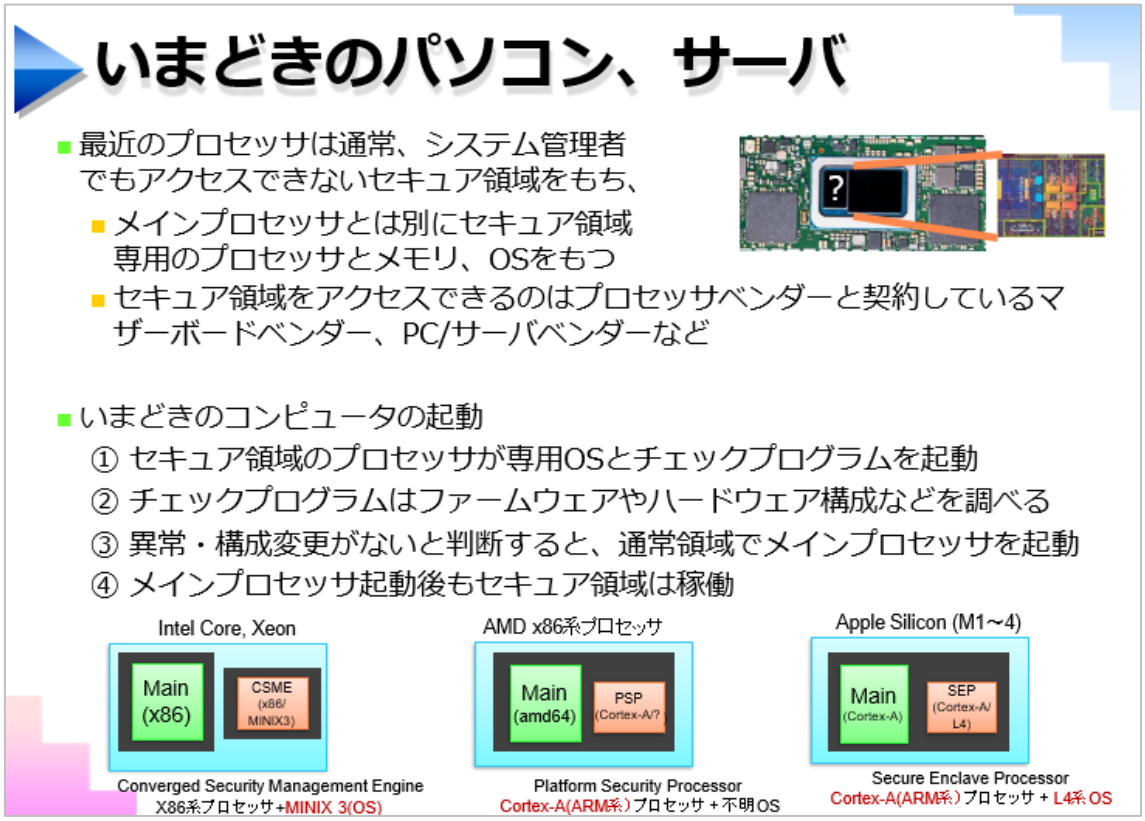
秘密分散を説明するために、以下のような例を用います。

大きなデータ X と Y があり、これを分散すると、x1＋x2＋x3 に、Y を y1、y2、y3 に分割し、それぞれを複数のストレージや事業者に配置します。これにより、各ストレージからは、元の X や Y を復元できません。しかし、例えば X+Y を計算する場合、x1＋x2、y1＋y2、x2＋x3、y2＋y3 を全部計算して、その結果を足し合わせることで X＋Y の結果が得られます。これが秘密分散の方法です。

実際には、秘密分散はもっと高度な方法で行われます。分割したデータを置く場所や計算場所が少ないと安全性が低下します。また、表形式のデータでは縦方向か横方向に合計するのかによって、分割方法が異なります。そのため、工夫が要り、秘密分散の場合にはデータのばらけ方に乱数などを組み合わせることもあります。このように、秘密分散は高度な工夫と技術が必要です。



TEE の説明に入る前にコンピューターの基礎に関わることを復習します。基本的に、Intel 系のプロセッサの TEE は、メインのプロセッサと同じ x86 系のセキュリティ管理用プロセッサ上で MINIX という教育用 OS で管理されています。Apple の場合、Secure L4 と呼ばれるマイクロカーネル OS が管理しています。余談ですが、Macintosh には、セキュア L4 と呼ばれるテンポラルロジックなどを駆使した検証手法を用いて、L4 マイクロカーネルからバグを取りました研究があります。AMD のプロセッサの場合、x86 ですが、セキュリティ管理用に ARM プロセッサが使われていますが、OS は非公開なので、不明です。



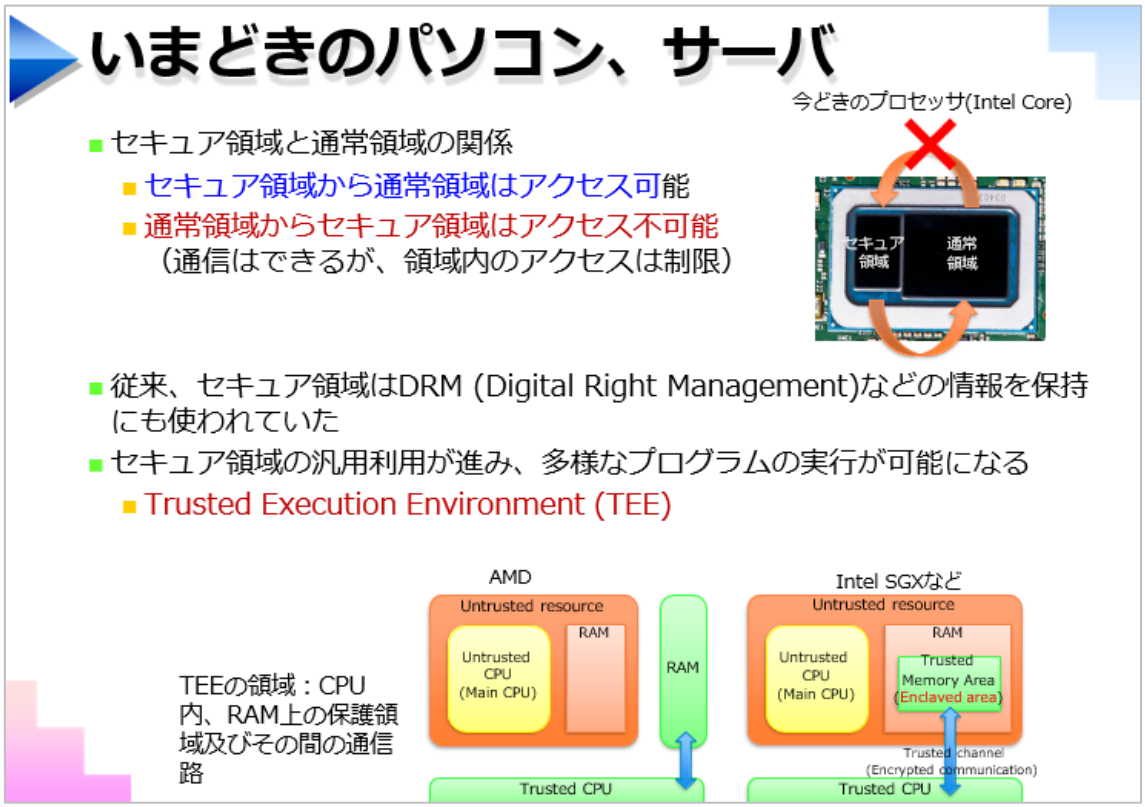
現在のプロセッサは二つのプロセッサが入っています。例えば、Intel の Xeon や Core には、x86 系のプロセッサです。もう一つはセキュアプロセッサで、別の領域に配置されており、TEE を管理しているものです。このセキュアプロセッサもコンピューターであり、メモリがあり、MINIX という OS が動作しています。そのため、現在の Intel プロセッサを搭載したパソコンでは、MINIX が動いているのです。

AMD のセキュアプロセッサは、ARM 系を使用しています。Apple の場合、メインの Apple Silicon も ARM 系です。セキュアプロセッサ上で動いているのが L4 です。

キュアプロセッサは、コンピューターがサーバーを含めて立ち上がる際に重要な役割を果たします。Intel のプロセッサでも、最初に立ち上がるのは x86 のメインのプロセッサではなく、セキュアプロセッサです。それが、さまざまな状態をチェックします。具体的には、BIOS やファームウェアが書き換えられていないか、新しいハードウ

ェアが追加されていないか、または欠如していないかを確認します。問題がなければ、次のメインのプロセッサを立ち上げます。

セキュアプロセッサは、メインのプロセッサが立ち上がった後も動作を続けます。以前は、このセキュアプロセッサのメモリに、DRM、Digital Right Management という、ゲームやビデオコンテンツの情報が保存していました。これは、メインのプロセッサからはセキュアな領域、プロセッサの領域にはアクセスができないようにするためです。逆はできるため、現在のコンピューターでは、Windows や Linux を Intel 系プロセッサで使用する場合、それは実は教育用 OS である MINIX の監視下で動いているのです。



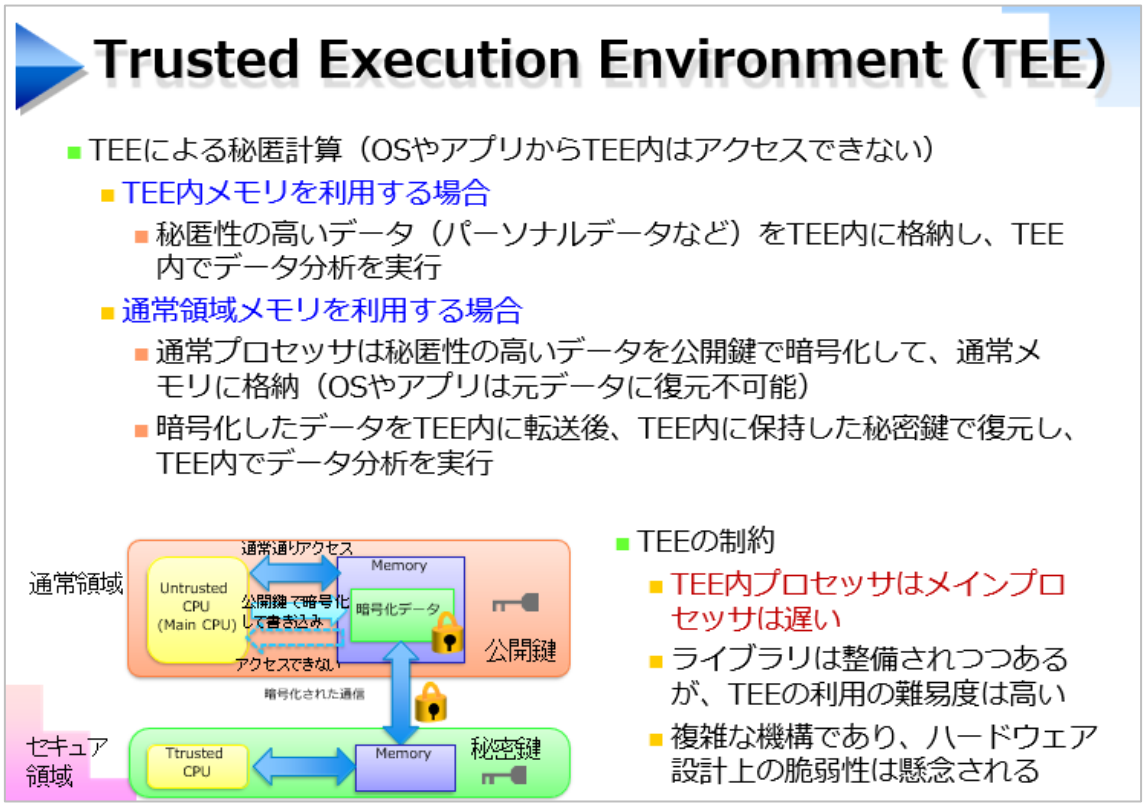
通常の領域からはセキュア領域にアクセスできないため、コンテンツのライセンスの情報をセキュア領域に保管することで、複製対策に利用されていました。以前は、セキュア領域は起動時のチェックとライセンスの管理に限定されているが、Intel や

AMD などが拡張しています。現在では、Intel のメモリ以外にセキュアな領域のプロセッサが動いている領域があり、通常のメモリ内にもセキュア領域からのみアクセス可能な領域が作られています。以前はセキュア領域のサイズが小さかったが、現在は大容量メモリも使えるようになり、実データ処理が可能になります。

セキュア領域ですが、クラウドの問題に対応するためにも使われています。クラウドサーバーがクラウド事業者によって運営されている場合、ユーザーの情報が事業者に見えてしまうことがあります。これを防ぐため、セキュア領域を活用し、クラウド事業者も見ることのできない領域を作ることで、ユーザーが安心してクラウドを利用できるようにしました。このような領域を Trusted Execution Environment（TEE）と呼びます。

TEE が秘密計算と関わる理由について説明します。例えば、プライバシーに関わる情報を全てセキュア領域に入れることで、メインのプロセッサからその情報を見られないようにします。セキュア領域にプライバシー情報を安全に保管できれば、メイン領域からのぞくことは不可能です。

セキュアな領域には汎用のプロセッサがあり、データ分析が可能です。統計的な分析結果のみを取得すれば、プライバシーリスクを低減できます。これが、TEE が秘密計算に関連する理由の一つです。



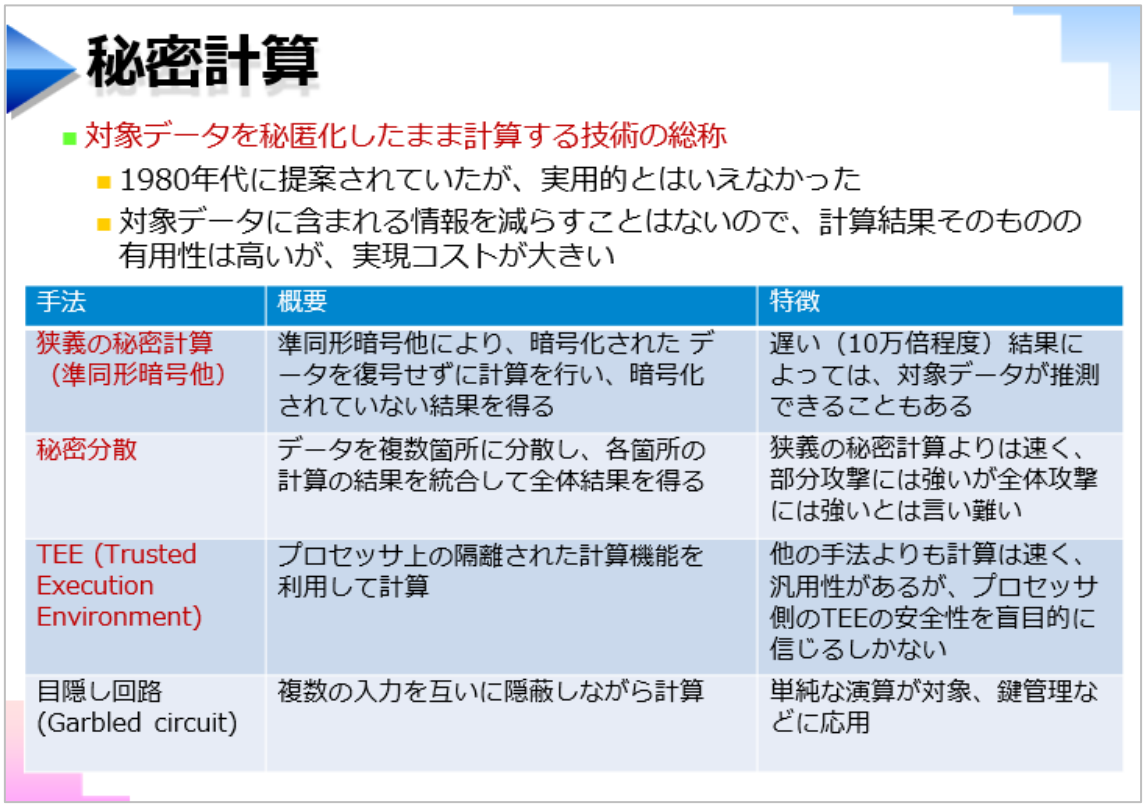
現在の TEE を秘密計算的に使う際には、データを通常のメモリに暗号化して保存します。例えば公開暗号系を使う場合、データは公開鍵で暗号化されており、秘密鍵がないと解読できません。メインプロセッサは暗号化されたデータにアクセスできず、必要な計算を行いながら秘匿すべきデータを暗号化します。

暗号化されたデータの復元鍵はセキュアな領域に保存されているため、暗号化されたデータはセキュアな領域に送信されます。現代のプロセッサは秘匿性の通信機能が組み込まれており、その中で秘密鍵を使って暗号を解読し、計算を行い、結果だけを返す仕組みです。TEE を使うメリットとしては、計算時にデータが平文に戻されるため、暗号化されたまま計算する秘密計算よりも速度が速い点が挙げられます。

しかし、いくつか問題点もあります。セキュア領域のプロセッサは、かなりサブセットに近いです。例えば 2018 年の Xeon プロセッサに搭載されていたセキュアプロセッサは、古い Pentium 程度の性能しかありません。そのため、高速な計算が難しく、

技術的にも扱いが難しいです。最近はライブラリが登場しているものの、アクセスや処理がうまくいかず、挫折する人も多いです。

TEE は複雑な技術であり、脆弱性が多く存在する可能性ありますが、利用者が少ないために発覚していないので、大きな問題になっていない可能性があります。TEE を使った秘匿計算は主にクラウドのデータセンターで行われるため、まずクラウドのセキュリティを突破する必要があり、ハッキングのハードルが高いです。そのため、脆弱性は残っている可能性があるものの、現時点では問題として露呈していないため、ひとまず安心とされています。

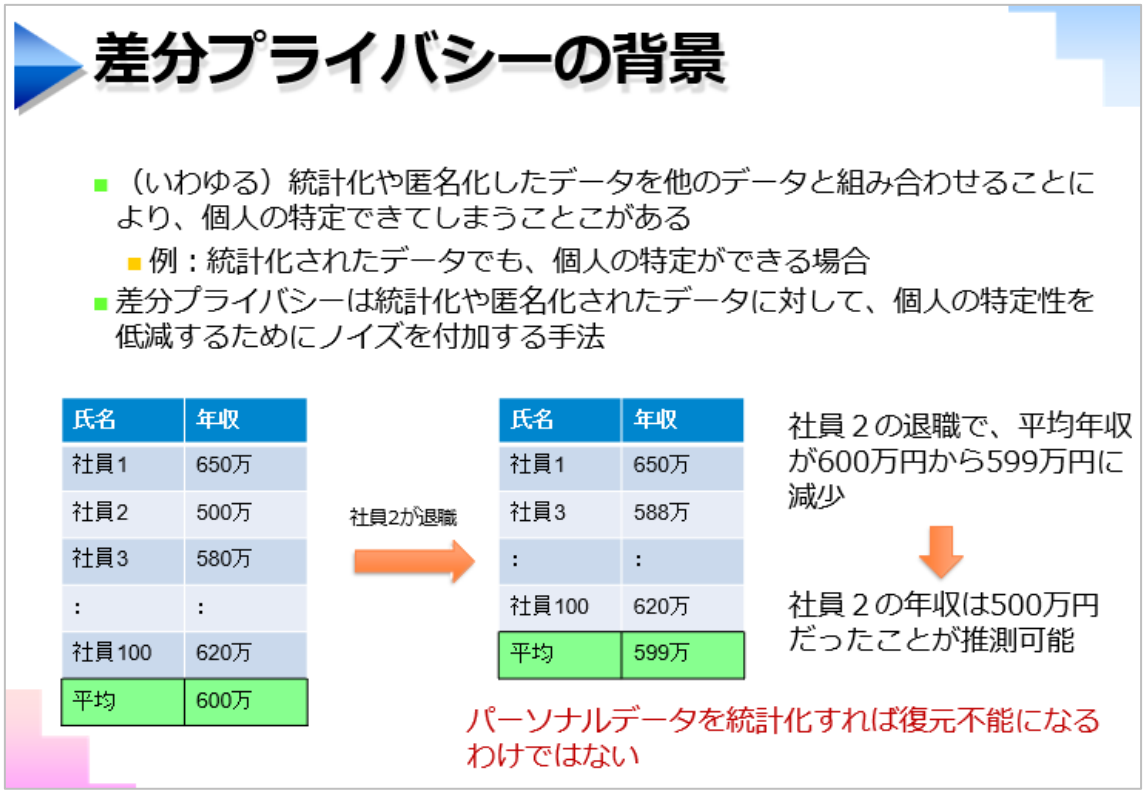


手法	概要	特徴
狭義の秘密計算（準同形暗号他）	準同形暗号他により、暗号化されたデータを復号せずに計算を行い、暗号化されていない結果を得る	遅い（10万倍程度）結果によっては、対象データが推測できることもある
秘密分散	データを複数箇所に分散し、各箇所の計算の結果を統合して全体結果を得る	狭義の秘密計算よりは速く、部分攻撃には強いが全体攻撃には強いとは言い難い
TEE (Trusted Execution Environment)	プロセッサ上の隔離された計算機能を利用して計算	他の手法よりも計算は速く、汎用性があるが、プロセッサ側のTEEの安全性を盲目的に信じるしかない
目隠し回路 (Garbled circuit)	複数の入力を互いに隠蔽しながら計算	単純な演算が対象、鍵管理などに応用

暗号化をしながら計算は現実的には難しいです。秘密分散は、暗号化をあまり行わないため速度は出るが、多くの台数が必要で、分割の技術も難しいです。TEE はハードウェアの支援を受けるため性能は良好で、特に秘密計算と比べると高速です。しか

し、TEE そのものにバックドアが存在しないという保証がなく、プロセッサベンダーを信頼するしかない問題があります。

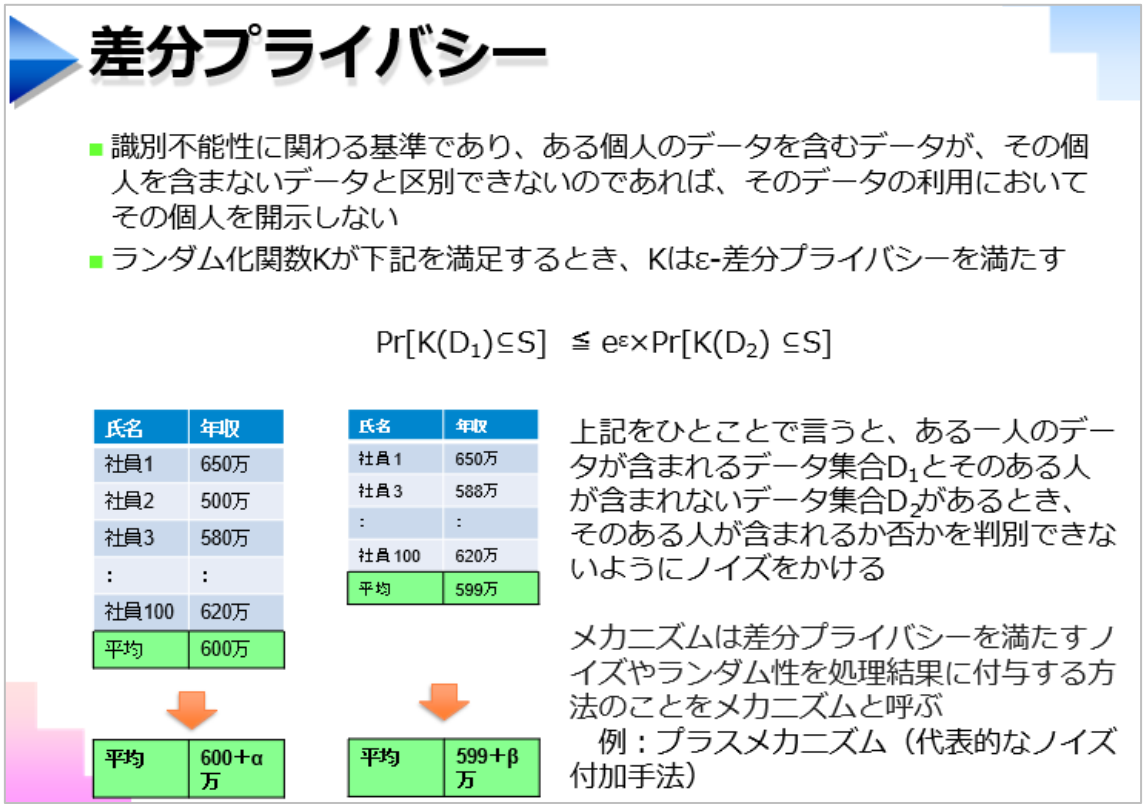
そのため、秘密計算系が各国の政府の法制度に取り入れられにくいのは、技術の詳細が理解されていないことが大きな要因です。経済界の人々は技術が少しでも使えると大騒ぎされるため、法制度を作るのは簡単ではないのが実情です。



PETs とよく話題になるのが差分プライバシーです。差分プライバシーとは、秘匿化の手法ではなく、情報の秘匿基準を示すものです。具体的には、統計処理や匿名化されたデータから個人を識別できるかどうかの基準を定めるものです。

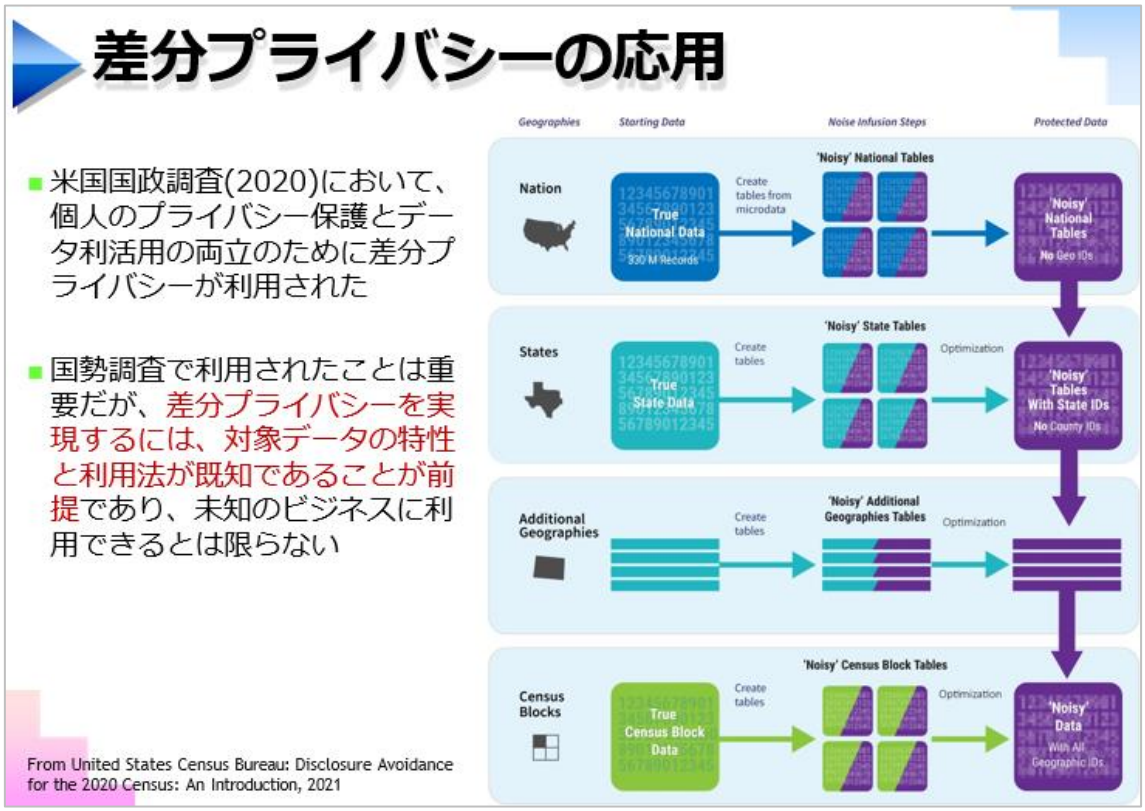
統計化されたデータだから個人の識別はできないと主張する人がいますが、実際にはプライバシーが見えます。例えば社員の給与を示す表で、平均給与が公開されている場合、社員 2 が退職すると平均値が変わり、社員 2 の給与が推測されてしまいます。

このような問題は統計化されたデータでも頻繁に発生します。こうした問題を防ぐために、統計化や匿名化されたデータに対して加工が必要です。その加工の基準が差分プライバシーです。



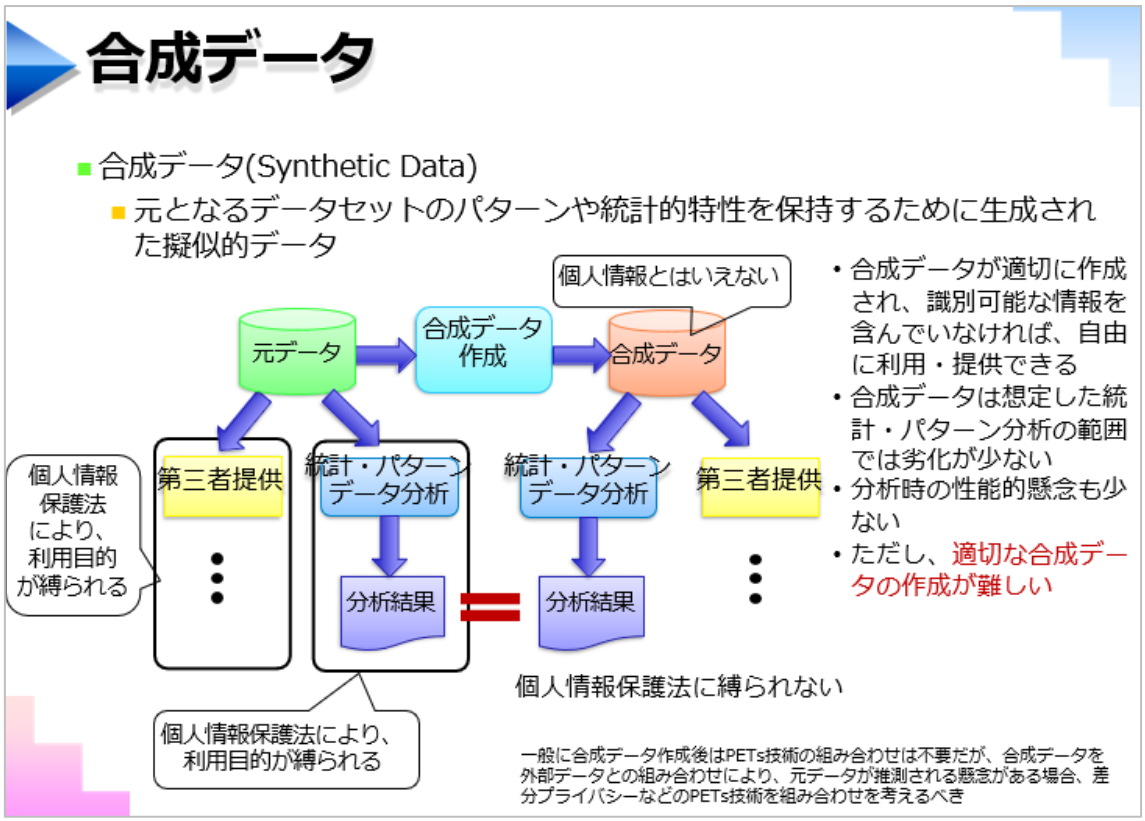
差分プライバシーは、二つのデータ集合のうち一方に含まれる対象がもう一方には含まれない場合、その対象の情報を復元できる確率を定式化したものです。

この PETs データにノイズを加えることで秘匿化を行います。例えば、データを少しずらすことでノイズを加え、特定の要素を取り除く前後の統計データや匿名化データから、その要素を推定できるかどうかを判断します。ノイズをかけることを「メカニズム」と呼ばれ、いくつかの種類があります。それらがうまく機能するかを評価するのが差分プライバシーの役割です。

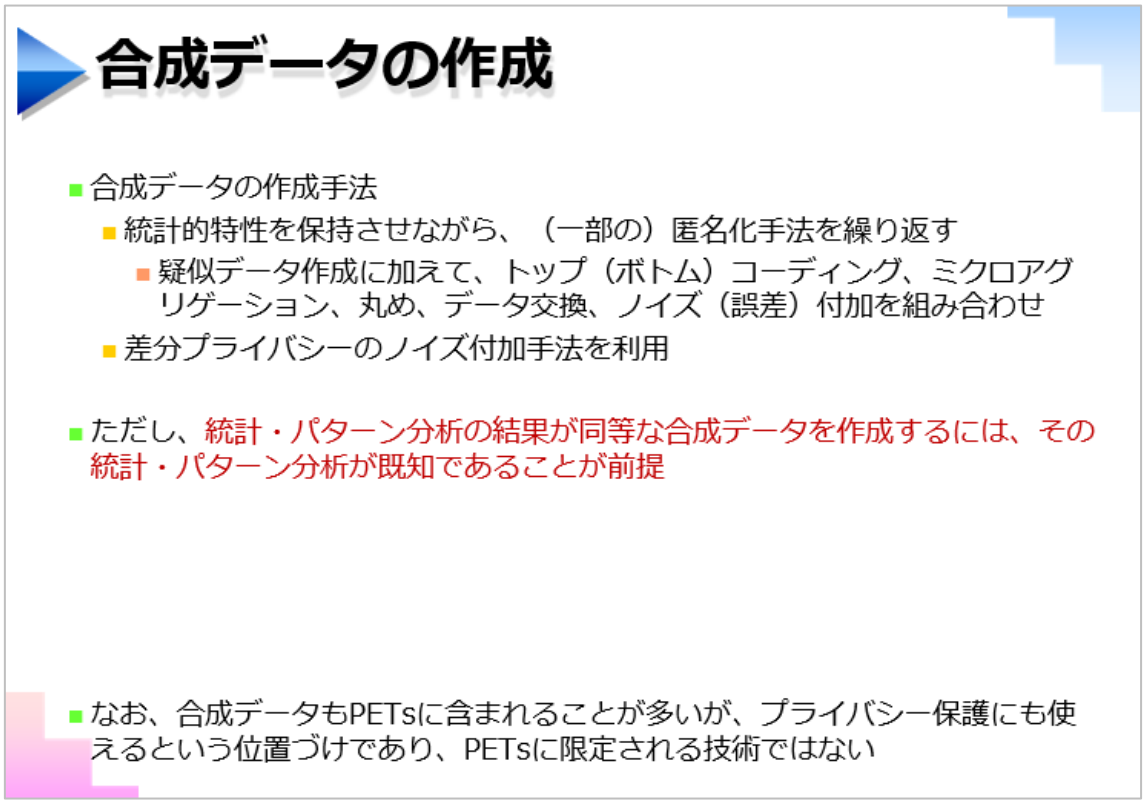


差分プライバシーは 2020 年のアメリカの国勢調査で国単位や州単位など様々な単位で利用され、個人の特定防ぐために詳細なデータを提供する際に使用されました。このことから差分プライバシーは注目を集めました。しかし、差分プライバシーは対象データの特性が相当程度わかっていることが前提となるため、データの特性が予測できる国勢調査は有効であっても、それ以外の状況、例えば対象の特性がよくわからないデータ（例えば新規案件のデータなど）では差分プライバシーの利用は難しいと考えるべきでしょう。

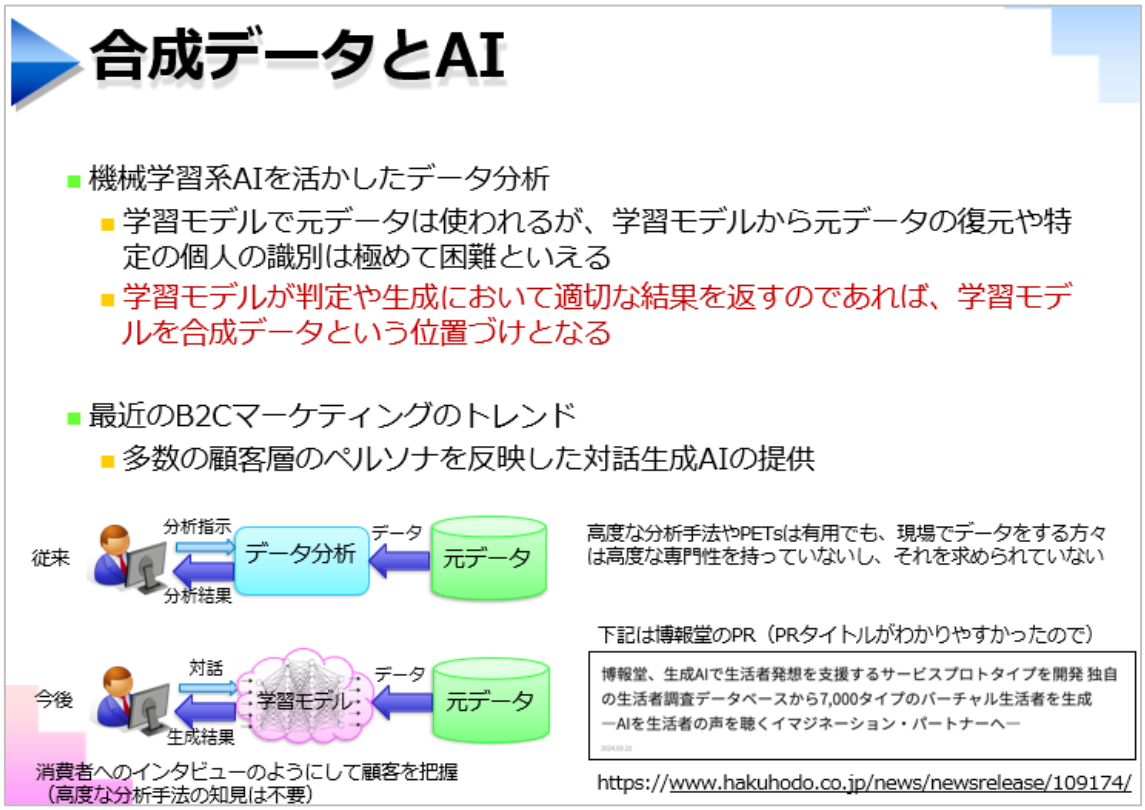
3. 合成データ



現在、データ分析でよく使われている手法の一つに「合成データ」があります。これは、元となるテータセットの統計的特性やパターン分析を保持するために生成された擬似データです。合成データを分析するのが、最近のトレンドだと思います。

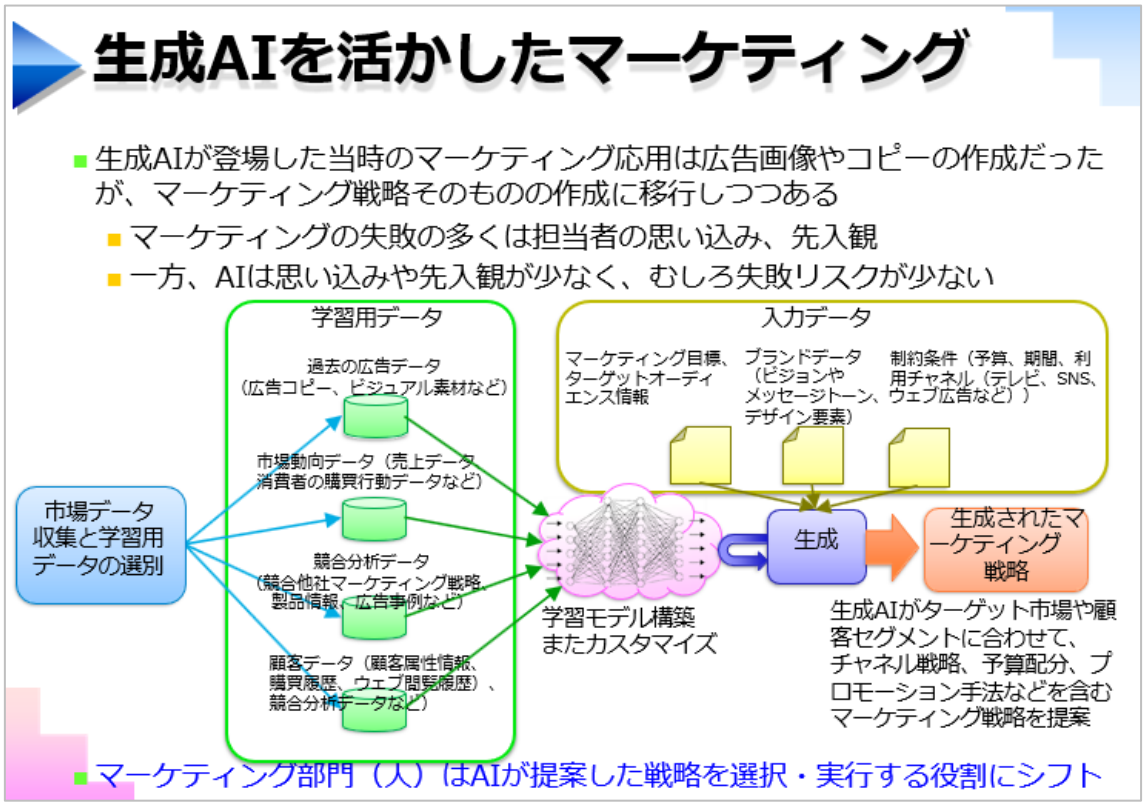


合成データの作り方には、匿名加工の繰り返し、差分プライバシーのチェック、生成ネットワークや生成 AI の利用などがあります。これらの技術の進歩により、合成データの作成は比較的容易になり、利用が増えているように感じます。

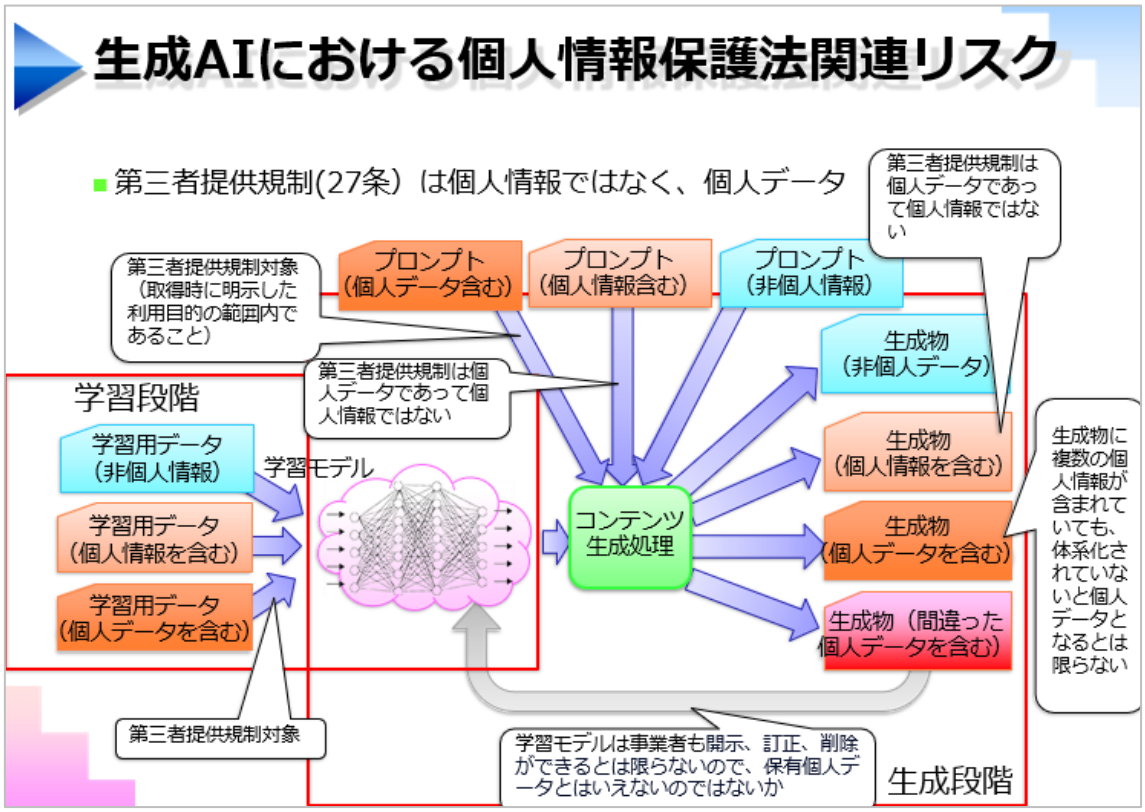


マーケティングでは、ペルソナを使った分析が一般的だったが、現在では直接パーソナルデータを利用することは少なくなっています。断片的なデータから人間がペルソナを作り、そのペルソナを基に分析を行うため、直接個人情報を使用しません。

さらに、生成 AI や他の AI 技術を利用して、ダミーのペルソナを作成できるようになり、パーソナルデータを直接使う必要がありません。ペルソナは人間の想像や恣意的な要素が含まれているため、それを利用するケースが増えています。



マーケティングについても、データ分析をせずに AI に依頼することが増え、この傾向は今後さらに強まるのではないかと感じています。



4. AI と PETs

生成AI向けのPETs

- 生成AIへのプロンプト内容が生成AI事業者に漏洩することが懸念
  - プロンプトの内容を学習モデルに反映すると、その内容が他の利用者に生成物に含まれる可能性がある
  - 一部の生成AI事業者（例：OpenAI）はプロンプトを学習モデルに反映させないと表明しており、それを信じるしかない
    - 学習モデルの構築以外には利用される可能性はある
- 生成AI向けの情報漏洩対策ソリューション
  - 例えばプロンプト中の人名などを識別子に置き換えてから生成ＡＩに入力し、生成後、出力中の識別子を元の人名などに戻すなど
    - 置換対象は人名や住所、電話番号などの明らかに個人情報（相当）とわかる情報のみ
    - それ以外はそのまま生成AIに渡されると考えるべき
  - RAGを組み合わせた生成AI利用で、RAGが参照するデータベースに含まれるパーソナルデータには要注意

また、生成 AI も個人情報保護法上の問題に対応するため PETs も登場します。例えば、企業がプロンプトに個人情報を入力する際に、個人情報を識別子に置き換えて、出力時にその識別子を元の個人情報に戻すサービスが多く存在します。しかし、プロンプト内の個人情報を完全に特定することは難しく、漏れが発生するため、100%とは保証できません。

日本は依然として LLM が全盛で、現在生成 AI が最も多く活用されているのは、自動車の自動運転の走行データの合成です。生成 AI のアプリが、他の機械学習系 AI の学習用データを作るために利用するのがトレンドになっていません。

一時期、顔認識においてマイノリティーの方が不利になる問題があったが、これはデータが不足していたことが原因です。そのデータを生成 AI で補完することが出てきています。そのため、顔認識の導入に関しては、警察などが利用することに対して以前よりも寛容な見方が広がる可能性があります。