3-2-3　プライバシー保護技術の理解促進・普及に向けた試み
3-2-3　Toward the Dissemination of Privacy-enhancing Technologies to Non-experts

小野 元　金森 祥子
ONO Hajime and KANAMORI Sachiko

安全なデータ利活用を実現する技術が世界中で日々開発されている。しかし、開発された技術が社会に普及するためには、それら技術が非専門家たちにも安心して受容される必要がある。安心を実現するための準備として、我々は「プライバシーポリシーで使用される技術用語の理解度とプライバシーポリシーへの同意の調査」「日本人のプライバシーに関わる法制度に対する理解の調査」「差分プライバシー技術を非専門家に説明するにあたり予想される課題の検討」を実施した。これらの調査結果は専門家と非専門家のプライバシー保護技術の理解の間には大きなギャップが存在することを示唆し、安心の実現に向けた課題を浮き彫りにした。

Despite the rapid development of privacy-enhancing technologies globally, widespread adoption among the general public, including non-experts, remains a challenge. To address this, we conducted surveys on the public's understanding of privacy terms and policies, and their familiarity with privacy-related legal frameworks in Japan. We also investigated the challenges associated with explaining differential privacy technologies to non-experts. Our findings indicate a significant disconnect between the understanding of privacy protection technologies among experts and non-experts, underscoring the difficulty of achieving broader societal acceptance.

1　まえがき

セキュリティ基盤研究室の主な目標の１つはユーザのプライバシーを保護しながらユーザに関わるデータを利活用する技術の開発である。サービス事業者にとっては、新たなビジネスの検討、顧客のニーズの把握、サービス向上等、ビジネスを優位に進めるために、ユーザの情報（データ）を収集し、利活用することは不可欠である。一方で、データ提供者であるユーザのプライバシーを保護する必要もある。事業者が収集した情報の中には、ユーザのプライバシーに関わる情報も多く、データ利活用とユーザのプライバシー保護の両立は、現代社会における大きな課題の一つである。この課題の解決のために、数学理論に裏打ちされた厳格な安全性と社会のニーズを満足するデータの有用性を両立できる技術の実現を目指している。

しかし、プライバシーを保護したデータ利活用技術は開発しただけでは社会に広く普及しない。プライバシー技術の非専門家であるデータ提供者にも理解してもらい、その技術を用いたデータ利活用が社会に受け入れられる必要がある。近年、サービス事業者が個人からデータを取得する際には、データの利用目的や保護の方法について明らかにした上で、データ提供者に事前にデータ提供の同意を得ることが常識となりつつある。この常識は法制度として社会に実装されている。

日本では、改正個人情報保護法第 18 条、第 27 条 [1] に基づき、事業者が個人情報を取得する場合は、プライバシーポリシーの作成が義務付けられている。プライバシーポリシーとは、その事業者の個人情報取り扱いに関する基本方針を示したものであり、個人情報の定義、個人情報の取得方法、個人情報の利用目的が定義されている。また、個人データ（個人情報データベース等を構成する個人情報）を安全に管理するための措置、個人データの共同利用、第三者提供、開示・訂正等の手続き、個人情報の取り扱いに関する相談や苦情の連絡先も併せて示されている。

さらに、EU 域内の個人データ保護を規定する法である General Data Protection Regulation（GDPR）は、「それらの個人データの取扱いと関連する情報及びコミュニケーションに容易にアクセスできること及び容易に理解できること、また、明確かつ平易な文言が用いられること」を求めており [2][3]、専門家だけでなく非専門家でも理解できるようにプライバシーポリシーを作成しなくてはならない。

これら法制度が透明性の確保を目指しているのに反して、実際のサービス利用時に提示されるプライバシーポリシーは非専門家にとっては難解である。アメリカで2021年に行われたプライバシーポリシーで使われる技術用語の理解度に関する研究では、当該技術用語が非専門家にとっては理解が難しく、誤解されていることを明らかにした[4]。この研究ではプライバシーポリシーに頻出する技術用語20語についての正しい定義を4つの選択肢の中から参加者に選択してもらうというアンケート形式の調査を実施したところ20用語中15用語で正答率が50％以下であった。また、この研究では技術用語を用いたポリシーと技術用語を用いない説明的文章を用いたポリシーを参加者に提示するアンケート調査も実施し、技術用語を用いたポリシーの方が技術用語を用いず書いたポリシーよりも同意率が上がる場合や、またその逆の場合もあることを示した。これは技術用語の意味を理解していれば判断が変わった人もいたことを示唆しており、難解な技術用語の使用は判断を誤らせることがわかる。本来、プライバシーポリシーはデータ提供者であるユーザがその内容を理解した上で同意(あるいは拒否)することで自己情報コントロールを実現する手段であるはずである。しかし、事業者は法を順守することに重点を置いた難解なプライバシーポリシーをユーザに提示して、本来の目的を果たしていないと上述の研究は暗に示している。

こうした社会情勢や先行研究の結果を受けて、セキュリティ基盤研究室ではプライバシーポリシーや保護技術を社会に広く普及させるためにプライバシー技術の透明性確保に向けた調査や研究を実施している。それらの研究を3つに分けて紹介する。1つ目の研究は上述の技術用語の理解やプライバシーポリシーへの同意についての日本版の調査の実施[5][6]である。技術用語の理解や技術に対する期待は居住国[7]や属する文化圏[8][9]によって異なることが示唆されており、上述の結果と同じ傾向が我が国で見られるかは自明ではない。2つ目は日本人のプライバシーに関わる法制度への理解度の調査[5][6]である。我が国には改正個人情報保護法などで事業者や行政による個人に関する情報の取り扱い方法が規定されているが、法律や技術の専門家ではない一般の国民はそれらの法によって何がどの程度保護されているのか正しく理解しているのかを調査した。3つ目は近年急速に社会実装が進んでいる差分プライバシーが技術者研究者以外にはどの程度理解できるのかの検討[10]である。

2　日本人の技術用語に対する理解とプライバシーポリシーへの同意の関係の調査

セキュリティ基盤研究室では「まえがき」でも言及した米国で行われたアンケート調査[4]と同様の調査を日本で実施し、国民性や環境によって技術用語に対する理解や、プライバシーポリシーに対する同意に違いはあるか調査した。

2.1　調査対象

調査は日本で幅広く活用されているクラウドソーシングサイトを利用して2023年4月に参加者募集を実施した。参加者を募集する際には、プライバシーに関心の高い参加者ばかりが集まることを避けるため、募集タイトルに「プライバシー」等の語句は含めず、「オンラインサービスの利用に関するアンケート調査」とした。アンケートの冒頭において、研究の目的や研究データの取扱い等について参加者に伝え、自由意志による参加同意を得た。なお、回答のパフォーマンスによって報酬が変わることがないことも伝えた。当研究室における事前調査をもとに、回答完了時間を20分程度と想定し、時給換算で日本の最低賃金を優に超える400円を報酬額に設定した。募集する曜日や時間帯によって参加者の属性に偏りが生じる可能性を考慮し、曜日及び時間帯が異なる3回に分けて募集した。362名の参加者のデータに対して、解析を行った。なお、事前にNICT内部のパーソナルデータ取扱い審議委員会のアセスメントを受けている。

2.2　技術用語に対する理解

日本人はプライバシー保護技術に関する技術用語をどの程度理解しているかを調査するために、アメリカでの調査[4]のうち16用語[*1]に関して、それらを個人情報保護委員会のGDPR仮日本語訳[11]などを参考に翻訳したアンケートを我々は実施した。例えば、「プライバシーポリシー」という語に対しては、参加者に次の5つの選択肢を提示して、正しいと思う1つを選択してもらった：

- ユーザのデータを収集・使用する方法を説明する法的文書(正解)
- ユーザのデータを保護する方法を説明するもの(不正解)
- 企業がユーザについて収集した情報の機密性を保持する方法を説明するもの(不正解)

＊1　オリジナル論文[10]では20語の技術用語を調査しているが、「PII」、「personal information」、「anonymized information」等は、改正個人情報保護法とGDPRで定義が異なるため、分析しないこととした。

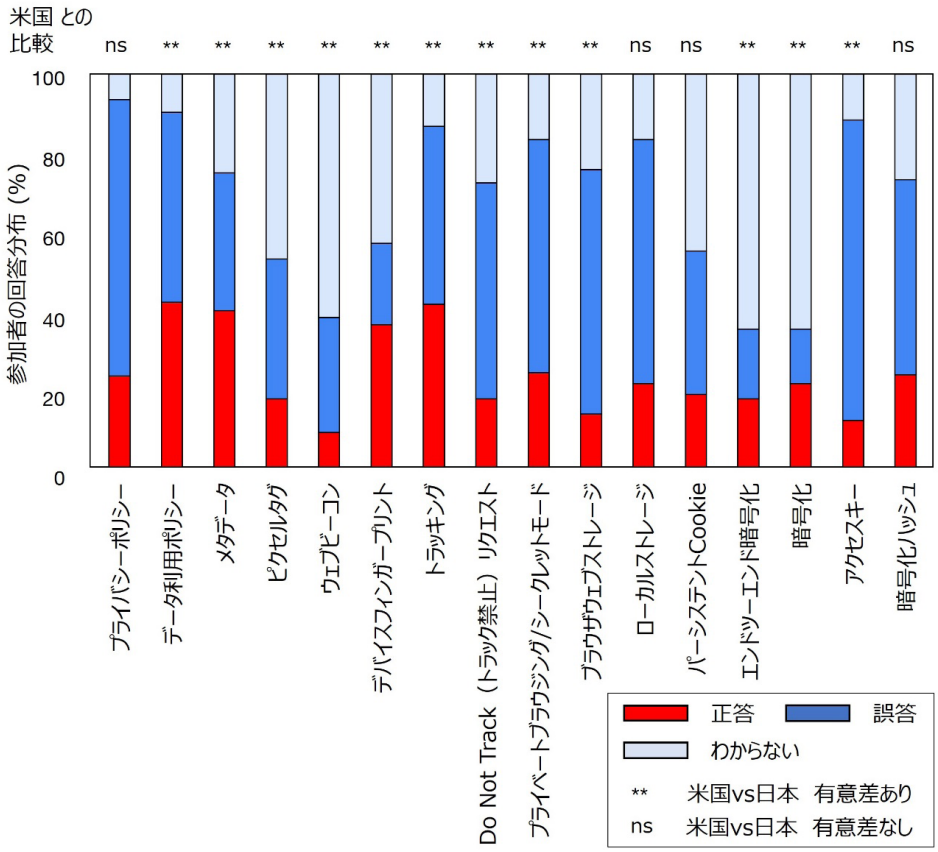


図 1　技術用語の定義の正答率

- 企業は許可なくユーザのデータを他のサイトや企業と共有しないということを述べるもの（不正解）
- わからない

図1に調査対象にした16用語とそれらに対する回答者の正答率を示す。16 用語中 12 語は米国での調査よりも有意に正答率が低かった。特に、「アクセスキー」「プライバシーポリシー」は誤答率が 7 割以上であり、誤った理解が浸透していることが窺える。プライバシーポリシーの定義に関する最も回答率の高い誤答は、「企業がユーザについて収集した情報の機密性を保持する方法を説明するもの」（37.3 %）であり、実際よりも安全性が高い選択肢が選ばれていた。また、セキュリティ基盤研究室で技術的な研究を行っている「エンドツーエンド暗号化」「暗号化」に対して、大半の回答者が「わからない」と回答していることはセキュリティ基盤研究室にとっては興味深い結果である。16 用語全体に関して、既に広まっている日本語表記を採用したにもかかわらず、日本では米国よりも技術用語の定義を誤解している参加者が明確に多いことを示唆する結果となった。

結果を素朴に解釈すると、「日本人は米国人よりもセキュリティ・プライバシースキルが低い」とも読み取れるが、必ずしもそうとは限らない。日本語のプライバシーポリシーを多数調査した結果、非英語圏特有の2つの問題が日本人の技術用語理解を妨げている可能性があることを発見した。1つ目の問題は英単語のカタカナ表記が用語の意味の推論を難しくしていることである。本研究で調査の対象とした16の技術用語のうち 15 語は英単語あるいはそれをカタカナ表記したものが含まれる。（例えば、「プライベートブラウジング」は英熟語 “private browsing” の音をカタカナで表したものである。）英語話者ならば用語全体の意味を知らなくともそれを構成する各単語から意味を推測可能だが、日本人がカタカナ表記された英単語を見てそういった推測をするのは容易ではない。（“private browsing” の意味は “private” と “browsing” の意味から英語話者ならおおむね推測できるだろうが、「プライベートブラウジング」はどこで区切るのかすらわからない日本人もいるだろう。）この仮説を補強する事実として、文脈によって意味が異なる “local” という単語を含む「ローカルストレージ」という用語は、正答は「自分に関する情報が自分のマシンまたはデバイス上に保存される。」であるが、米国での調査では物理的な意味での “local”（企業の敷地内にあるマシンまたはデバイス、ユーザの地域にあるマシンまたはデバイス上に保存される。）と誤答する回答者が日本での調査よりも多かったことが挙げられる。つまり、用語の各構成単語から全体の意味の推論が難しい技術用語は日本人に限

らず意味を誤解しやすい可能性がある。2つ目の問題は技術用語の表記揺れが多いことである。例えば、「Do Not Track Request」という技術用語は「Do Not Track（トラック禁止）リクエスト」「トラッキング拒否」「DNT」「ピクセルタグ」「トラッキングピクセル」などいくつもの日本語表記があることを確認している。こうした表記揺れは、同じ用語を別の用語だと誤認させて理解を妨げている可能性がある。これら非英語圏特有の問題を解消する補助をするだけでも日本人の技術用語理解度を引き上げられるかもしれない。補助の方法としては、技術用語の日本語表記を統一するよう専門家が社会に働きかけるなどが考えられる。また、非英語圏に限らず技術用語理解を助ける方法として、技術用語の表記箇所にその説明を記載したページをリンクすることなどが提唱されている [4]。

2.3　プライバシーポリシーへの同意

「技術用語に対する理解」の調査と並行して、技術用語を用いて書かれたプライバシーポリシーと平易な言葉だけを用いて書かれたプライバシーポリシーで同意度に差があるかの調査も実施した。半数の回答者には技術用語（全6語：ウェブビーコン、トラッキング、セッション cookie、パーシステント cookie、エンドツーエンド暗号化、暗号化）を使用したポリシーを各用語ごとに計6種類、もう半数の参加者には技術用語を使用せずに説明した同じ内容のポリシー6種類を提示して、各ポリシーに同意する可能性を「とても低い」から「とても高い」の5段階で回答してもらった。

表1は「とても低い」と1点「とても高い」を5点として各ポリシーに対する全回答者の回答の平均値を評価したものである。結果として、6語中3語で技術用語を用いたプライバシーポリシーと技術用語を用いず平易な文章で説明したプライバシーポリシーの同意度に統計的に有意な差が見られた。これにより、プライバシーポリシーに技術用語を使用することが日本のユーザの同意度に影響し、技術用語の使用が真の同意取得の壁となっていることが確認された。「ウェブビーコン」と「トラッキング」では、技術用語を利用したポリシーの方が同意率は高く、技術用語を用いないポリシーとの差が統計的に有意であった。この結果と「技術用語に対する理解」の結果を総合すると、一部の日本人は技術に対して実態よりも高いデータ保護能力を期待しており、かつ、その誤った期待を基にデータ所有者でありデータ提供者である本人が望まない（意図しない）データ提供を行っている可能性を示唆している。ただし、暗号化に関しては、その定義が「わからない」という回答率が高いが、その説明を平易な言葉で聞けばポリシーの同意に肯定的になるという、他の技術用語とはやや異なる受け止められ方をしていることは特筆に値する。

表1　技術用語を使用したポリシーと使用しないポリシーの平均同意可能性の比較

	技術用語 使用（平均同意可能性）	技術用語 不使用（平均同意可能性）
ウェブビーコン	**2.58**	1.91
トラッキング	**2.62**	2.35
セッションcookie	3.30	3.25
パーシステントcookie	3.05	3.07
エンドツーエンド暗号化	3.72	3.76
暗号化	3.18	**3.86**

このような結果になった理由はこの調査結果だけからはわからないため、今後の課題の一つである。例えば、「暗号化」という技術用語のように、「わからない」を選択した回答者が多い場合、わからないものをどの程度快適（安全）に感じるかそうでないかは国の文化的側面や国民の性格特性が影響すると考えられる。同意率やその差の方向性については更なる調査の下で結論づける必要があると考える。こうした調査は、社会のプライバシー技術への期待の理解につながり、他の研究にも良い影響を与えると予想される。

3　日本人のプライバシーに関わる法制度に対する理解の調査

技術用語やプライバシーポリシー以外で非専門家にとって理解が難しいものに我が国のプライバシーに関わる法制度が挙げられる。日本国の法律は最初から日本語で記述されているため「技術用語に対する理解」で指摘した表記揺れ等の問題は生じないが、日常的には用いない用語などが含まれておりその理解は容易ではない。プライバシーに関わる我が国の法制度を日本人がどの程度理解できているか我々は調査した。本研究は、「日本人の技術用語に対する理解とプライバシーポリシーへの同意の関係の調査」とは異なり、オリジナルの研究である。

この調査では以下の2点について日本人が理解しているかをアンケートで調査した：

(i) 個人情報保護法の下で実施可能なデータの取り扱いに関する質問

(ii) 10種の個人情報が要配慮個人情報（改正個人情報保護法が定める個人情報の中でも「不当な差別、偏見その他の不利益が生じないようにその取扱いに特に配慮を要するもの」）に該当するか否かに関する質問

(i) の理解度を調査するために、特定の種類のデータについての取り扱い方法が正しい選択肢を選んでも

らう 4 択問題を実施した。例えば、匿名加工情報の取り扱いについての選択肢は次の 4 つである：

- 原則として他の企業への提供は禁止（不正解）
- ユーザの同意を得れば他の企業への提供可能（不正解）
- ユーザの同意を得ずに他の企業への提供可能（正解）
- わからない

表 2　個人情報保護法下で実施可能なデータの取り扱いに対する期待（アンケートの質問と結果）

仮名加工情報の提供	
原則として他の企業への提供は禁止	**18.0%**
ユーザの同意を得れば他の企業へ提供可能	42.5%
ユーザの同意を得ずに他の企業へ提供可能	10.5%
わからない	29.0%
匿名加工情報の提供	
原則として他の企業へ提供は禁止	16.6%
ユーザの同意を得れば他の企業へ提供可能	36.5%
ユーザの同意を得ずに他の企業へ提供可能	**24.3%**
わからない	22.7%
外国にある第三者への個人データの提供	
提供禁止	13.5%
国が安全性を認めていない外国へは提供禁止	13.5%
ユーザの同意を得れば提供可能	**49.2%**
ユーザの同意を得ずに提供可能	1.9%
わからない	21.8%
アカウントのないユーザのデータ収集・データ利用	
同意を得ていないため，収集できない	9.4%
収集可能だが，同意を得ていないため利用できない	40.6%
収集可能であり，ポリシーに記載の通り利用可能	**34.3%**
わからない	15.7%

太字は法で定められた内容と合致する選択肢（正答）を示す．

(ii) の理解度を調査するための調査として、銀行口座、国籍、病歴など様々な種類の情報を回答者に提示して、それぞれに対して「要配慮個人情報に該当すると思う」「要配慮個人情報に該当しないと思う」「わからない」の 3 つの選択肢から一つを選んでもらった。なお、これら調査の回答者は「日本人の技術用語に対する理解とプライバシーポリシーへの同意の関係の調査」の回答者と同一である。

表 2 は調査 (i) でのアンケートの質問と選ばれた選択肢の割合である。

- 仮名加工情報は法律で原則として他の企業への提供が禁止されているが、それを正しく答えられた回答者はわずか 18 % であった。42.5 % もの回答者がユーザの同意を得れば他企業へ提供できると誤って回答した。
- 匿名加工情報はユーザの同意を得ずに他の企業へ提供可能であるが、その選択肢を正しく選べた回答者は 24.3 % で、半数以上の回答者が法律で定められている取り扱いよりも厳格な取り扱いの選択肢を選んでいる。
- 外国にある第三者への個人データの提供はユーザの同意があれば可能であると法律で定められており、回答者の約半数はその正しい選択肢を選ぶことができた。
- アカウントのないユーザのデータ収集・データ利用は利用可能かつプライバシーポリシーに書いてある範囲での利用が法律で認められているが、その選択肢を正しく選べた回答者は 34.3 % にとどまる。それよりも多い 40.6 % の回答者は「収集は可能だが、利用はできない」と誤って回答した。

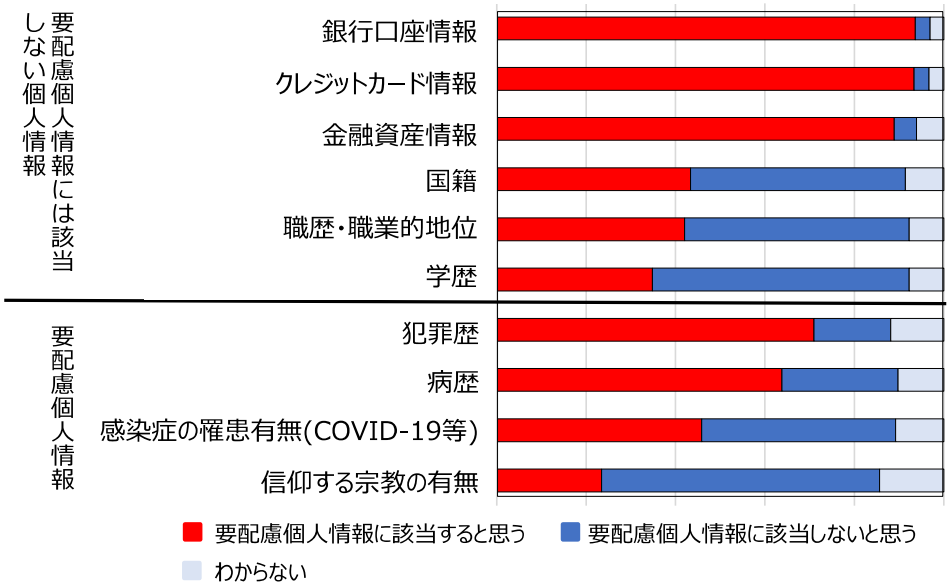


図 2　要配慮個人情報への該当／非該当アンケート調査の結果

4つの質問全てで正答率が50％を下回り、かつ、「外国にある第三者への個人データの提供」以外の質問では不正解の選択肢が選ばれた割合が正解の選択肢が選ばれた割合よりも高かった。この結果から日本国の法律で定められているデータの取り扱い方と、日本人が期待するデータの取り扱い方の間に小さくないギャップが存在すると推測される。なぜこのような結果になったかの理由の調査はこれからの課題であるが、いくつかの仮説が立てられる。仮名加工情報と匿名加工情報の質問では、一番多く選ばれた項目が一致しているため、そもそも仮名加工情報と匿名加工情報の違いを多くの回答者が理解できなかった可能性がある。また、アカウントのないユーザのデータについては、GDPRなどより厳しいデータ保護を要求する日本国外の法律と混同している可能性が考えられる。

図2は(ii)の調査で対象とした個人情報10種と全回答者の回答の割合である。銀行口座情報、クレジットカード情報、金融資産情報は法律上の要配慮個人情報ではないにも関わらず、回答者の8割以上が要配慮個人情報であると誤答した。反対に、感染症の罹患無と信仰する宗教の有無は法律上の要配慮情報であるのに、要配慮個人情報だと回答した回答者は半数以下であった。法律上の要配慮個人情報とアンケート回答者が考える要配慮個人情報の間に一部で大きなギャップがあることが窺われる。このようなギャップが生じる理由として、回答者の多くが要配慮個人情報とは、その情報が漏えいした場合の経済的損失が大きい情報だと誤解しているからだと考えられる。要配慮個人情報であると考える参加者が多かった銀行口座、クレジットカード、金融資産情報は漏えいした場合、他者に自らの資産を不正に使用されてしまう可能性があり、犯罪歴、病歴も就職において不利な扱いを受ける可能性がある。

上述した(i)(ii)の調査結果はいずれも法律が規定するデータの取り扱い方と日本人が期待するデータの取り扱い方の間に大きなギャップがあることを示唆している。このギャップは事業者、ユーザ双方に不利益をもたらすため望ましくない。例えば、ユーザが法律によって厳格な保護が義務付けられていると期待していたにも関わらず実際には保護されない場合、ユーザは間違った理解を基に望まないデータ提供を誤って行う可能性がある。このようなユーザの望まないデータ提供はユーザの自己情報コントロール権及びプライバシー権を侵害する。反対に、ユーザが厳格には保護されないと思っているにも関わらず実際には法律によって厳格に保護されている場合、ユーザは必要以上にデータ提供に慎重になり事業者はデータ取得の機会を失う。

このような望ましくない誤解を避けることはユーザの権利の担保と事業者の利益の確保につながるため、その方法の研究は今後の課題の一つである。ユーザが誤解しやすいデータの取り扱いをハイライトで表示するなどの工夫が必要であると考えられる。この方法論の研究において、国ごとにデータの取り扱いに関する法律もユーザの意識も異なることが大きな挑戦になる。これらの違いのためにある国でうまく機能した方法が別の国でも適切に機能する保証が全くない。したがって、日本国と日本人のための方法の研究は海外の研究結果を待つだけでなく、日本の研究機関が主体的に実施する必要がある。

4　新たな技術の社会実装に備えて

ここまでで紹介した理解度調査の対象にはならなかったものの急速に社会実装が進んでいる技術として差分プライバシー[12]がある。差分プライバシーは統計(平均値、分散、学習モデルのパラメータ等)を公開する際に、統計に乱数を加えることで、統計を解析しても元のデータを逆算できないことを要求するプライバシー定義である。図3と図4は平均値から特定の1人のデータを逆算する攻撃と、乱数を使えばその攻撃を防げることを端的に表した図である。乱数を加えるメカニズムのプライバシー保護の強度を非負の実数ϵを用いて表現する。図5はそのϵの定義を端的に表した図である。ある人のデータがxかx'かに絞り込めたとして、公開された乱数入りの統計を見て、xとx'のもっともらしさの比(図中のe^{ϵ})が小さいほど、強いプライバシー保護であるとみなす。(ϵが小さいほど、2択の絞り込みが当たる確率は1/2に近づく。2択が1/2でしか当てられないというのは意味のある推測が全くできないことを意味する。)差分プライバシーを使うことで、統計公開時のプライバシー侵害のリスクを定量的に評価及び制御することができる。ただし、より強い保護(より小さいϵ)を達成するためにはより大きな分散をもつ乱数を使う必要があり公開する乱数入り統計の信頼度を大きく損なうため、達成したいプライバシー保護の強度と統計の信頼度のバランスを考えて慎重にϵを決める必要がある。図6は保護の強度と統計の信頼度のトレードオフを表した図である。セキュリティ基盤研究室でも差分プライバシーに基づいてリスクを制御しながら機械学習をする技術を開発しており([13][14]など)、差分プライバシーの概念や利点を社会に浸透させることは、本研究室が開発した技術が社会に貢献する機会を増加させることにつながる。

特に著名な差分プライバシーの社会実装の例としては、2020年のアメリカの国勢調査[15]がある。この国

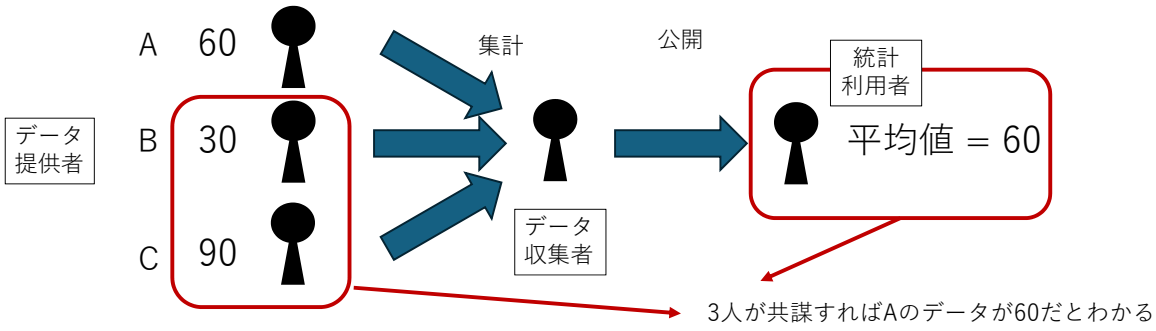


図３　素朴な平均値の公開に対する攻撃

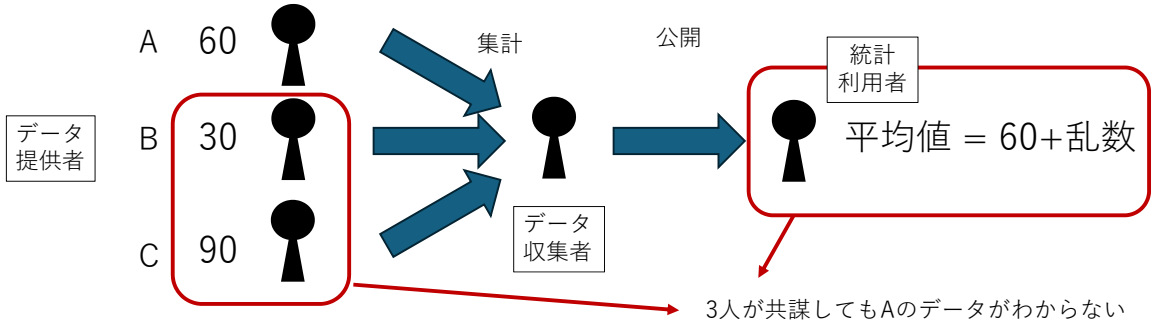
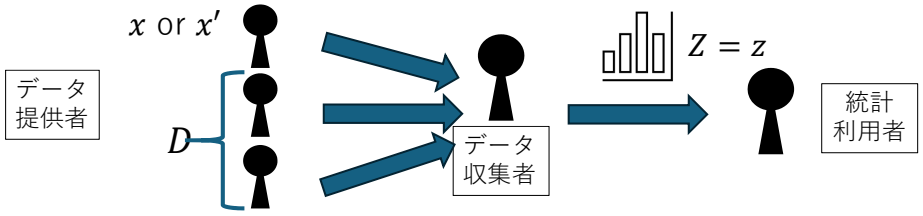


図４　乱数を加えた平均値の公開



$$\frac{P(Z=z|D\cup\{x\})}{P(Z=z|D\cup\{x'\})}\leq e^{\varepsilon}$$ が常に(任意のD,x,x',zに対して)成り立つとき，その保護手法はε-DPであるという

元のデータが$D\cup\{x'\}$のとき，$Z=z$が観測される確率

図５　差分プライバシーの定義の要約

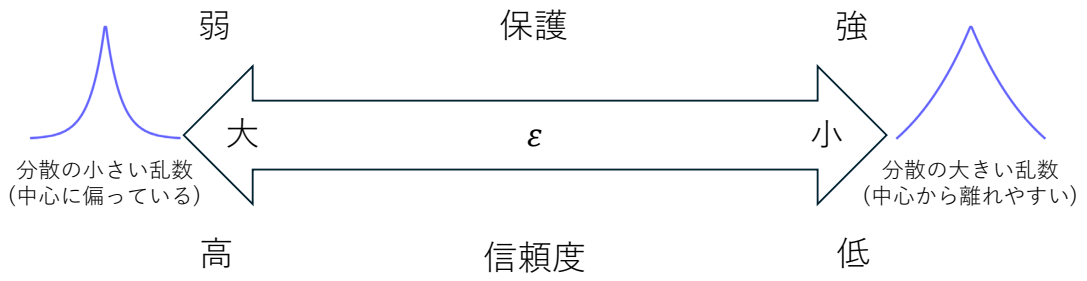


図６　ϵの役割

勢調査は集計結果を差分プライバシーの定義を満たすように乱数を含む加工をしてから公開した。さらに、技術の実装だけでなく国民への情報発信も積極的に行い、平易な言葉で保護のアイディアを説明する web ページや動画も公開している [16]。この広報動画では、2010 年までの国勢調査の結果が他の公開情報と付き合わせることで多くの個人データが特定されてしまうことを批判した上で、差分プライバシーの数学的アイディアに基づいて設計された乱数がいかにプライバシー保護と統計の信頼性を両立するかを直感的に説明している。

乱数を取り入れることで逆算ができなくなることを非専門家にも理解してもらうことにはある程度成功していると考えられる。例えば、図 7 や図 8 に示すような図を使って差分プライバシーの性質を非専門家に理解させる試みがユーザブルセキュリティの有力国際会議で発表されるなど、説明ツールの開発や非専門家の差分プライバシーへの意識調査が進んでいる [17]。

しかし、こういった説明ツールは必ずしも保護の度合い ϵ の意味を説明しない。差分プライバシーは ϵ の大きさによって保護の度合いを表現し、端的に言えば、ϵ が大きいほど元データの逆算に成功する確率が高い。単に「乱数を使って差分プライバシーの定義を満たしている」といっても、その時の逆算成功率が 1 % なのか 99 % なのかが分からなければ無意味である。そのため、将来的にこれらの説明ツールのようなもので ϵ の意味を説明せずにプライバシーポリシーのようにユーザに提示した時には、ユーザは安全の度合いを誤認した状態でデータ提供に同意させられる恐れがある。もしもそのような意図しない同意が多発すれば、差分プライバシー技術の社会への普及の障害になりかねない。

このような意図しない同意を避け、ϵ の大きさも含めて理解されるための方法の開発は我々の今後の目標の一つである。この課題に関しても、先の章で述べたような国ごと文化圏ごとの違いがある可能性があるため、海外で実施された意識調査の結果から日本人にも当てはまる知見が得られるかどうかは定かではない。したがって、海外での実施された調査の日本国（日本語）での追試も実施する必要があり、目標達成までに課題が山積している。

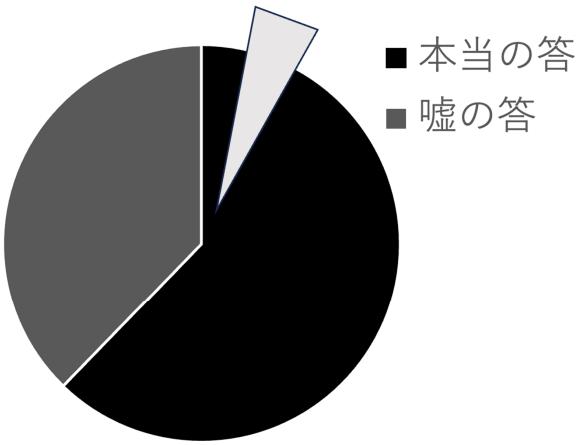


図 7　$\epsilon = 0.5$ のスピナーの例。円盤を回転させて、止まった時に三角の頂点が指している方の行動をとる

5　むすび

「日本人の技術用語に対する理解とプライバシーポリシーへの同意の関係の調査での調査」「日本人のプライバシーに関わる法制度に対する理解の調査」及び先行研究 [4] は、技術用語の理解促進だけではプライバシー保護技術が社会に受容され普及するには不十分であることを示唆している。受容の障害になる理由や受容される方法の調査研究は我々の今後の課題である。それら理由や方法は国や文化圏によって異なる可能性が大いにあるため、日本人のための研究を日本の公的研究機関で実施する意義は大きい。また、伝統的な技術体系だけでなく、「新たな技術の社会実装に備えて」で紹介した差分プライバシーなどの新しいプライバシー保護技術と社会の間を如何に取り持つかも我々の課題である。

謝辞

本研究の一部は、JST CREST JPMJCR21 M1 の助成を受けたものです。

図 8　画像の解像度と ϵ の大小関係を対応づけて、直感的に説明する方法

【参考文献】

1 個人情報の保護に関する法律 https://elaws.e-gov.go.jp/document?lawid=415AC0000000057
2 General data protection regulation (GDPR), Regulation - 2016/679 - EN - gdpr - EUR-Lex (europa.eu), 2016.
3 個人情報保護委員会, "透明性に関するガイドライン (guidelines on transparency 仮日本語訳)," 2018. https://www.ppc.go.jp/files/pdf/toumeisei_guideline.pdf
4 Tang, J., Shoemaker, H., Lerner, A., and Birrell, E. Defining privacy, "How users interpret technical terms in privacy policies," Proceedings on Privacy Enhancing Technologies 2021, 3 (2021), pp.70–94.
5 金森 祥子, 池田 美穂, 亀石 久美子, 長谷川 彩子, "個人情報保護法に対するユーザの理解度の調査," Proceedings of Computer Security Symposium 2023 (2023).
6 Kanamori, S., Ikeda, M., Kameishi, K., and Hasegawa, A., "User comprehension of technical terms in privacy policies and expectations of the privacy protection law in Japan," https://www.usenix.org/conference/soups2023/presentation/kanamori-poster, 2023. This article accepted in the poster session of the Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023).
7 Herbert, F., Becker, S., Schaewitz, L., Hielscher, J., Kowalewski, M., Sasse, A., Acar, Y., and Dürmuth, M., "A world full of privacy and security (mis)conceptions? findings of a representative survey in 12 countries," Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2023), CHI '23, Association for Computing Machinery.
8 Ur, B. and Wang, Y., "A cross-cultural framework for protecting user privacy in online social media," Proceedings of the 22nd International Conference on World Wide Web (New York, NY, USA, 2013), WWW '13 Companion, Association for Computing Machinery, pp.755–762.
9 Li, Y., Kobsa, A., Knijnenburg, B. P., and Nguyen, M. C., "Cross-cultural privacy prediction," Proceedings on Privacy Enhancing Technologies, 2017.
10 小野 元, 紀伊 真昇, "プライバシーバジェット ϵ の決定にデータ提供者の意思を反映させるにはどうすればいいか？：文献調査，問題整理と一提案," Proceedings of 2024 Symposium on Cryptography and Information Security, 2024.
11 個人情報保護委員会, "GDPR (general data protection regulation: 一般データ保護規則)," https://www.ppc.go.jp/enforcement/infoprovision/EU/
12 Dwork, C., McSherry, F., Nissim, K., and Smith, A., "Calibrating noise to sensitivity in private data analysis," Theory of Cryptography (Berlin, Heidelberg, 2006), S. Halevi and T. Rabin, Eds., Springer Berlin Heidelberg, pp.265–284.
13 Phong, L. T., and Phuong, T. T., "Differentially private stochastic gradient descent via compression and memorization," Journal of Systems Architecture 135 (2023), 102819.
14 Nojima, R. and Wang, L., "Differential private (random) decision tree without adding noise," Neural Information Processing (Singapore, 2024), B. Luo, L. Cheng, Z.-G. Wu, H. Li, and C. Li, Eds., Springer Nature Singapore, pp.162–174.
15 Abowd, J. M., "The U.S. census bureau adopts differential privacy," Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (New York, NY, USA, 2018), KDD '18, Association for Computing Machinery, p.2867.
16 United States Census Bureau. Protecting privacy in census bureau statistics, 2021. https://www.census.gov/library/video/2021/protecting-privacy-in-census-bureaustatistics.html
17 Karegar, F., Alaqra, A. S., and Fischer-Hübner, S., "Exploring User-Suitable metaphors for differentially private data analyses," Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022) (Boston, MA, Aug. 2022), USENIX Association, pp.175–193.

小野 元　(おの はじめ)

サイバーセキュリティ研究所
セキュリティ基盤研究室
研究員
博士 (統計科学)
プライバシー、統計、AI セキュリティ

金森 祥子　(かなもり さちこ)

サイバーセキュリティ研究所
セキュリティ基盤研究室
主任研究技術員
プライバシー、ユーザブルセキュリティ
【受賞歴】
2024 年　2024 年度山下記念研究賞
2023 年　CSS2023 優秀論文賞 (兼・UWS 優秀論文賞)