資料論文

プローブ情報システムにおける IDプライバシー問題の現状と考察

Identity and privacy problems with Big Data from Probe Vehicle Systems.

和泉 順子*
Michiko IZUMI

Email: mizumi@hosei.ac.jp

情報通信技術が普及し、さらに高度化・複雑化しているが、同時に広く一般の人々が様々な情報端末を用いてICTサービスを利活用している。その結果、SNS（Social Network Services）上やWebメールや検索の利用、ネットショッピングを日々利活用している情報サービスを始め、多種多様なデジタルデータが情報流通基盤上で収集・管理され、ビッグデータとして利活用されるようになってきた。しかしこのような「人間の行動履歴・属性情報」や「モノの状態」には個人情報やプライバシーに関する懸念が強く残る。本論文では、ビッグデータを用いたシステムの一つであるプローブ情報システムの今までの取り組みと方向性を紹介し、ビッグデータ関連システムのIDやプライバシー問題を考察する。

Probe vehicle information systems as one of information service based on "Big Data", which gather various sensor data, called probe data, by some sensors or equipments on vehicles can innovate new methods to build not only the fine-grained social information services, but also some valuable business around the ITS (Intelligent Transport Systems) area. Probe data for wide area communications were standardized internationally, however, it's noted that there are diversification of threats to privacy because that the collected probe data would include the spatiotemporal data when the vehicle obtained the information.

In this paper, I discuss the provision for privacy issue on the information services based on Big Data from probe vehicle information systems. There exists a tradeoff between data quality (or its utility) and the degree of privacy; however, it should carry the weight of versatility for some existing services and scalability for the future.

*: 法政大学 国際文化学部

1. はじめに

インターネットは今や情報通信基盤として広く社会一般に浸透している。また、科学技術だけでなく行政や企業活動、教育、医療、福祉、交通、運輸、農林漁業、文化普及など多種多様な分野におけるデータやサービスがデジタル化・ICT 化してきており、インターネット上で膨大なデータの流通やサービス展開・制御などが行われている。

近年では特に、情報通信技術自体は洗練・淘汰され高度化・複雑化しているが、サービス利用者としてはスマートフォンをはじめとする携帯情報端末を使いこなし、構造や仕組みが分からない状態でも新しい情報サービスを積極的に利活用することに抵抗を感じない人々が増えてきた。これは、新しく魅力的な情報端末を直感的に利用出来るためのユーザインターフェイスやサービス側の工夫と研究開発の成果であると言える。これにより、広く一般の人々が、違和感や問題を感じることなく様々な情報端末を用いて ICT サービスを利活用し、その結果、情報流通基盤上、特に SNS（Social Network Services）上や Web メールや検索の利用、ネットショッピングを日々利活用している。同時に、これらのサービスを介していつの間にか自分の位置情報やプライバシーに関わる様な行動情報などを含む様々なデジタルデータがインターネット上に流通・蓄積・管理される様になった。つまり、「人間の行動履歴」や「属性」に基づく情報がデジタルデータとして大量に集積される状態になっている。

一方、人間の行動履歴や属性情報以外にも、多種多様なモノがネットワーク化された世界（IoT（Internet of Things））では、エネルギー、医療・ヘルスケア、自動車やロボット等の製造業、農業等、あらゆる産業分野において膨大なデータ（電力使用情報、医療・健康情報、位置情報、気象・土壌情報等）をいかに有効に利活用するかが重要になってきている。センサ・ネットワークをはじめとする「モノの状態」を把握するためのデジタルデータも大量に観測・集積されるようになった。

この「人間の行動履歴・属性情報」や「モノの状態」によるデジタルデータが膨大に収集・管理される状況を捉えて、ビッグデータ、およびオープンデータの利活用が注目を浴びている。ビッグデータに関して言えば、本質的には、データ量の多寡や種類に関する研究開発よりも、膨大なデータからいかに速くまたは正確に「新しい価値」を生み出し、新しい産業の創成や社会課題の解決に繋げるかが議論の中心となっている。しかし、データを利活用することに重きを置いている一方で、徐々に問題として認識されつつある個人情報やプライバシーに関する懸念も徐々に大きくなってきている。近年では、関連する問題事例が報道されることで企業における個人情報やプライバシーの取り扱いに一時的に注目が集まり、その結果、利用者はインターネットをはじめとする情報流通基盤上での自身のプライバシーや個人情報に対して不安や懸念を抱く様になってきた。

さらに、米国国家機密情報の告発に伴い、米国内では一般ユーザの個人的な通信内容を含めた多くの情報がビッグデータとして傍受・監視されていることが公となった。この PRISM のような情報収集システムは、それを適正に稼働させることで国家安全保障に役立てるという重要な目的がある一方、Gmail や twitter をはじめとする広く一般に普及しているサービスを利用することで流通するデータが公的機関に収集・監視されていることが判明したことで、一般ユーザの心理的抵抗も大きかった。

このように、日々の生活で利活用している情報サービスの多くはインターネット等を用いて情報を流通・蓄積・管理しており、それらの膨大なデータから新しい価値を見つけるビッグデータの取り組みが注目されている。本論文では、ビッグデータに紐づいているプライバシーや ID の問題を考察するために、ビッグデータを用いたシステムの一つであるプローブ情報システムの今までの取り組みと方向性を紹介し、ビッグデータ関連システムの ID やプライバシー問題を考察することを目的とする。

2. ビッグデータとプライバシー

2.1 ビッグデータとは

総務省の HP[1]記載の定義によると、ビッグデータを構成するデータ群は、ソーシャルメディアデータ、ウェブサイトデータだけでなく、センサデータ、オフィスデータ、ログデータ、オペレーションデータなど出所が多種多様であり、各データを連携させることでさらなる付加価値の創出が期待されている（用語定義・説明は付録 6.1 を参照のこと）。

また、同 HP では、ビッグデータ、つまり「多量の」データが必要な理由として、データ利用者側からビッグデータを捉える場合、事業に役立つ有用な知見を導出する観点から「**高解像**（事象を構成する個々の要素に分解し、把握・対応することを可能とするデータ）」、「**高頻度**（リアルタイムデータ等、取得・生成頻度の時間的な解像度が高いデータ）」、「**多様性**（各種センサーからのデータ等、非構造なものも含む多種多様なデータ）」の 3 つの特徴を満たすため結果的に「多量」のデータが必要となる、としている。つまり、ビッグデータとは、事業に役立つような有用かつ新しい価値を創出するようなデータの組み合わせとして高解像度でフレッシュな多種多様なデータが多量に必要であるとし、それらをソーシャルメディアやウェブサイト、センサ・ネットワーク、各種 ICT サービスのログやオペレーションデータを組み合わせることで得られる「人間の行動履歴・属性情報」や「モノの状態」のデジタルデータで実現していると言える。

しかし、他の情報と組み合わせることで、新しい何か（価値）を創出することがビッグデータの議論の中心である一方で、その新しい何かが個人を特定できる情報であったり、プライベートな情報となったりする

[1] http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc121410.html (2014 年 8 月現在)

場合がある。これは、個々のデータ自体は個人情報ではなく単なる位置情報や購入商品番号などのデータであったとしても、他の情報やデータベースと組み合わせたりまとめたりすることで、いわゆる「名寄せ」や「モザイク効果」が発生し、個人を特定することが可能な情報（パーソナルデータ）が見えてしまうことに由来する。

通常は、ビッグデータの内の個々のデータは、デジタル化し収集・管理される前に「データを収集すること」に関する規約を事前に約款等に記述・説明しており、それに同意した利用者だけがサービスを利用できる。サービスを利用した場合の利用者の属性情報や行動履歴、モノのデータ等はそのサービス提供者に送信・管理され、ショッピングサイトの「おすすめ」情報や提示される広告等の様にサービス向上やマーケティング分析、他サービスへの誘導等に活用される。この事前の説明や利用者の同意なしに、他サービスのデータを流用したり他社に提供したりした場合はいわば「ルール違反」であり、大きな社会問題につながることもしばしば発生している。例えば、JR 東日本が交通系 IC カード Suica の乗降履歴を日立製作所に提供していた件（2013 年 7 月）をはじめ、佐賀県武雄市図書館での T ポイントカード情報の利活用（2013 年 4 月）、ベネッセコーポレーションにおける大規模な個人情報の流出事件（2014 年 7 月）、オムロン関連会社による熱海駅乗降客の無断撮影（2014 年 7 月）や京都大学による商業施設のカメラ情報の解析「環境適応型で実用的な人物照合システム」（2014 年 8 月）など、利用者への事前説明や断りなしに「人間の行動履歴」や「モノの状態」をマーケティング分析や研究開発に利活用して問題となった事例は多い。

そこで、日本では、高度情報通信ネットワーク社会推進戦略本部（以下「IT 総合戦略本部」）が、高度情報通信ネットワーク社会形成基本法（平成 12 年法律第 144 号）第 36 条の規定に基づき、平成 25 年 12 月 20 日に「パーソナルデータの利活用に関する制度見直し方針（以下「制度見直し方針」）から制度改正の内容を検討している。IT 総合戦略本部において平成 26 年 6 月 24 日に「パーソナルデータの利活用に関する制度改正大綱」が出され、パブリックコメントに付されたが（当該パブリックコメントは、平成 26 年 7 月 24 日に締め切られた）[2]、この基本的な枠組みとして、「1 本人の同意がなくてもデータの利活用を可能とする枠組みの導入等」「2 基本的な制度の枠組みとこれを補完する民間の自主的な取組の活用」「3 第三者機関の体制整備等による実効性ある制度執行の確保」となっている。この「本人の同意がなくてもデータの利活用を可能とする枠組みの導入」が実現すれば、上記事例は大きな問題にはならなくなる。この「本人の同意がなくてもデータの利活用を可能とする枠組み」が導入された場合、個人を特定する情報（ID 等）やプライバシー問題はどのように扱われるべきか、技術だけでなく教育・啓蒙や社会・法整備、情報倫理等の多方面から検討が必要であると考えられる。本論文では、この問題について、ビッグデータシステムの一つでもあるプローブ情報システムの今までの取り組みを、主に技術面から考察する。

2.2 プライバシーとは

「プライバシー」という言葉はよく使われる用語ではある（用語定義・説明は付録 6.2 を参照のこと）が、分野によってその定義や認識は異なる未整理の概念である。現在は、1880 年にルイス・フランダイスが宣言した「もっとも包括的な権利であり、かつ文明人によって最も価値を置かれている権利」だけに留まらず、思想や言論、監視からの自由、自己情報コントロールの権利、自分自身の評判や検索からの保護等も含まれており、プライバシーという概念の定義はより困難になっている。

通信をはじめとする情報技術分野では、1990 年代頃までは個人情報や利用者のプライバシー問題のみならず通信やシステムのセキュリティ全般も、現在程は注目されていなかった様に思われる。例えば、技術論文や IETF における RFC（Request for Comment）等のような技術の国際標準に関する文書では提案システムの設計や仕様、実験、評価などは詳細に記述されているが、セキュリティやプライバシーに関する議論や考察は論文の最後、「今後の課題」のような章で、"security considerations"として多くの場合「今後検討すべき」程度に留まっていた。しかし、近年では情報サービスに関する個人情報やプライバシーへの関心が高まっており、特に日本では、個人情報保護法の成立前後あたりから、一般の利用者にも少しずつ、インターネットや情報サービスは便利なだけではなく、「安心・安全」も必要であるという意識が発現し、プライバシーや個人情報に関する意識のみならず、情報リテラシやメディア教育等の必要性も問われ始めている。

なお、本論文では、主にビッグデータの ID やプライバシーに関して考察することを目的とするため、ここでは情報リテラシ（または情報デリカシー）やメディア教育の情報セキュリティやプライバシーに対する重要性については言及しない。

2.3 個人情報と個人に関する情報（パーソナルデータ）の違い

情報流通としての個人情報やプライバシー保護については、1980 年にまとめられた「プライバシー保護と個人データの国際流通についてのガイドラインに関する OECD 理事会勧告（OECD プライバシーガイドライン）」[3]が各国地域で参照されているが、日本でも同ガイ

[2] パーソナルデータの利活用に関する 制度改正大綱
http://www.kantei.go.jp/jp/singi/it2/info/h260625_siryou2.pdf
なお、この大綱には、継続的な検討課題として「1 新たな紛争処理体制の在り方」「2 プロファイリング」「3 プライバシー影響評価(PIA)」「4 名簿屋」に関する問題提起もなされている。

[3] OECD, Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (1980).

ドラインを参考にしつつ、インターネットをはじめとする ICT 基盤によって構築された情報化社会において便益を享受するための法規として、個人情報保護法が平成 15 年 5 月に成立・公布され、平成 17 年 4 月より全面施行されている。

前章で「名寄せ」の問題に触れたが、個人を特定する情報は、氏名や住所、生年月日のような個人識別性を有する、いわゆる「個人情報」だけに留まらない。日本の個人情報保護法は「個人情報」を保護の対象としているが、その定義は以下のようになっている。

> 「生存する個人に関する情報であって、当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む。）（個人情報保護法第 2 条第 1 項）」

つまり、生存している個人に関する情報であるため亡くなった方の情報は保護の対象外である一方で、氏名や生年月日のような情報だけでなく、他の情報と照合等をすることによって特定の個人を識別することが出来るものは対象とされている。個人情報保護法は、その目的（第 1 条）を、「個人情報を取り扱う事業者の遵守すべき義務等を定めることにより、個人情報の有用性に配慮しつつ、個人の権利利益を保護すること」と定義しており（第 1 条全文は付録 6.3 を参照のこと）、個人情報取り扱い事業者等の個人情報の適正な取り扱いと個人の権利利益の保護を規定した法律であって、利用者のプライバシー保護が目的とされているわけではないことが分かる。

では、この「他の情報と照合することによって特定の個人を識別可能」な情報として、IP アドレスやクッキーのような具体的な情報が、この定義内の「特定の個人を識別することができる」、いわゆる個人識別性の要件を満たすか否か、あるいは個人識別性がない情報であっても保護対象とすべきものがあるか否か。これについては議論も多々成されており、このような個人に関する情報（個人識別性を問わない「個人に関する情報」）のことを「**パーソナル情報**」あるいは「**パーソナルデータ**」として概念整理されつつある。パーソナルデータという用語は、2005 年度から推進された経済産業省「情報大航海プロジェクト」での「パーソナル情報」という概念とほぼ同義であると考えられる。情報大航海プロジェクトではこのような情報を以下のように定義している。

> 「個人情報保護法に規定する「個人情報」に限らず、位置情報や購買履歴など広く個人に関する個人識別性のない情報」

また、2012 年から総務省で開催されている「パーソナルデータの利用・流通に関する研究会」においては、同様の概念を「パーソナルデータ」と定義し、用いている。本論文では、個人識別性を有する「個人情報」に限定することなく、広く「個人に関する情報」を「パーソナルデータ」として検討の対象とする。

前述の通り、個人情報だけでなくパーソナルデータの利活用については、ビッグデータを活用したマーケティング分析や動向調査等の側面から多くの可能性が期待されている。例えば、2011 年 3 月 1 日の東日本大震災発生時には、被災地で必要な物資や情報等が twitter による個々のつぶやきから分析されたり、被災地に実際に赴いた自動車のカーナビゲーション（テレマティクスサービス）や運転者の携帯電話の位置情報等から道路状況を解析し、現在通行可能な経路、不通になっている道路などの情報が共有されたりするなど、防災・減災に役立てる湯用な試みも数多く報告された。道路や自動車などに関わる ITS（Intelligent Transport Systems: 高度道路交通システム）分野においては、この震災以降は特に、個人の行動履歴（パーソナルデータ）が災害情報の共有に有用であることに注目を集めており、2013 年に東京で開催された ITS 世界会議では「Sharing of Road and Traffic Information」等のセッションで複数の論文が発表されている[4]。

2.4 ビッグデータを用いたサービス事例としてのプローブ情報システム

プローブ情報システムとは、インターネット等の情報通信基盤を用いて車両に搭載されたセンサーから車両の走行状態や位置情報、時刻や周辺環境情報、運転者や車両個体の情報等に関する情報を収集し、統計処理を施して交通情報や気象情報、車両運行予測、旅行者サポートなどのような有益な情報サービスを生成するためのシステムである（図 1）。広義のプローブ情報システムは、取り扱う情報は車両に限定されず、周辺状況を”プローブ”したデータを収集し、加工・共有・提供するシステムであるが、本論文では車両あるいは ITS 分野におけるプローブ情報について論じる。なお、車両プローブ情報システムと同義で使われる用語として、フローティングカーデータ（Floating Car Data）と呼ばれるサービスも存在する。

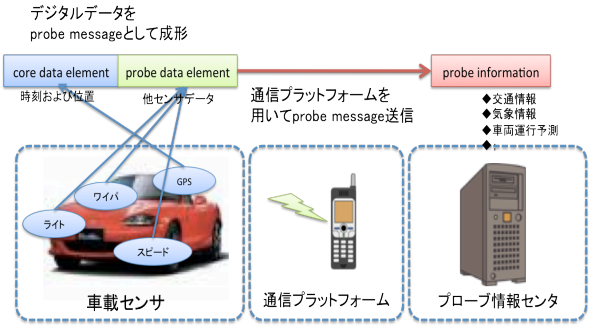

図 1 プローブ情報システム概要

[4] http://itsworldcongress.org/sessions/sharing-of-road-traffic-information/

ビッグデータを用いたサービス実用例としても早い段階から着目されており、車両メーカによる独自開発されたテレマティクスサービスとしては、HONDA のインターナビ・フローティングカーシステムや、TOYOTA の G-book、BMW の XFCD（Extended Floating Car Data System）等が挙げられる。車両や交通情報に特化した情報サービスだけでなく、2001 年に発足したインターネット ITS 協議会（発足当時はインターネット ITS コンソーシアム）においては、当時名古屋や横浜においてタクシー業務支援や駐車場管理、商店街との連携や運転者支援など、多角的かつ大規模な実証実験が行われた[5]。また、国内では一般的な利用例として VICS（Vehicle Information and Communication System）情報に連動させる等有効活用されている。しかし、各社が開発・提供するプローブ情報システムで取り扱う情報は、利用者との契約の関係上、原則として事業者内で閉じて管理運営しており、通信手段やデータ管理方法等も形態は様々である。これらのサービスの中でも共有可能な情報（プローブデータ）は、共通の形式で情報通信基盤に流通させることで緊急時等の情報提供協力や広範囲かつ細密な道路・交通管理の提供が実現可能となるため、イノベーションが期待される重要な技術領域として認識されている。プローブデータのフォーマットや構成等は ISO において、IS22837[6]として国際的に標準化されている（図 2）。

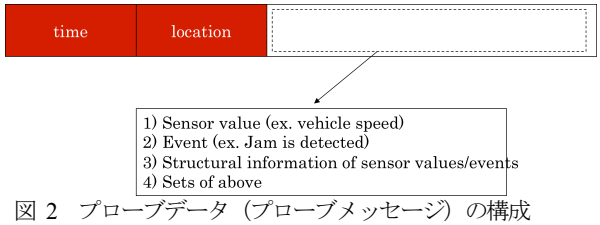

図 2　プローブデータ（プローブメッセージ）の構成
（ISO IS22837 より）

標準化されたプローブデータは、その情報を取得した時刻および位置情報を示すコアデータエレメント(図 2、赤い部分）と、センサーやイベント情報を格納するデータエレメントとで構成される。

3.　プローブ情報システムのプライバシー保護に関する取組

3.1　プローブ情報システムにおけるプライバシー侵害の懸念と対策

IS22837 におけるプローブデータ（プローブメッセージ）は、データ送信時にその情報を取得した位置情報および時刻情報を組み合わせて送信することになっており、ここで定義されているプローブデータ[7]自体は統計情報、つまり匿名のデータとして取り扱われる。従って、このプローブデータ自体には定義上、個人情報は含まれない。しかし通信に用いる ID やプローブデータの性質上、一定時間の観測や複数情報の統合により、プローブデータ取得または収集時に個人やプライバシーに関わるパーソナルデータが生成または類推可能となる。たとえば時刻と位置情報を取得するだけであっても一連の通信を捕捉していけば速度超過をした区域が特定可能となり、データ送信の起点や終点が私有地であればその車両所有者の居住地である可能性が高いなどの類推が可能になる。また、車両の保険や車検情報は、保険会社やガソリンスタンド、整備工場など関連業界にとっても魅力的な情報となる。このように車両から発する情報の一部は、それ自体が匿名であっても時系列または他の情報と組み合わせることでその所有者または運転者の情報と関連性が発生するため、広くサービスを開発・展開するにはシステム自体の安全性だけでなく情報管理や保護への対策が不可欠となる。プローブ情報システムにおけるプライバシーや個人情報保護に関する技術的および運用に関する対応と議論を重ね、脅威分析等を行ったことについては、(1),(2),(3)等で論じている。なお、これらの議論と OECD プライバシーガイドライン等を元に、プローブ情報システムにおけるパーソナルデータの取り扱いに関する基本原則が IS24100[8]としてまとめられている。

3.2　システム上の ID とプライバシー問題

前述のように、パーソナルデータは様々な分野で急速に利活用、実用されてきており、ITS 分野においても新しく利便性の高いサービスが誕生する可能性が極めて高いと考えられている。その一方で、一般の利用者からもプライバシー保護の観点から不安視・問題視される部分でもある。その不安の原因は「パーソナルデータ利活用のルールが明確でないため」とされており、産業界をはじめとする多くの分野、あるいは学問領域において問題の整理と適正な利用・流通促進に向けた議論が続いている。プローブ情報システムにおいても、2006 年 12 月に開催された経済産業省基準認証事業「プローブ情報システムにおける個人情報保護に関する標準化」の活動内容をと策定したガイドラインの紹介を行うワークショップ[9]の参加者を対象に、プローブ情報システムの利用に関して不安原因のアンケート調査を

---

[5] http://www.internetits.org/

[6] ISO IS22837:2009 "International Standardization Organization: Vehicle probe data for wide area communications"

[7] ISO IS22837 におけるプローブ情報システムの定義は、「プローブ情報システムは自動車の持つセンサ情報を、情報通信基盤を用いて収集し、統計処理等を施すことによって交通情報や環境情報、気象情報等、有益な価値ある情報を生成し、情報の共有・提供を行うシステム」となっている。

[8] ISO IS24100:2010 "Basic principles for personal data protection in probe vehicle information services"

[9] 2006 年 12 月 19 日（火）14:00-17:00 「自動車と個人情報保護に関するワークショップ」
慶應義塾大学 SFC 研究所主催で三菱商事ビルディングで開催された。
https://www.kri.sfc.keio.ac.jp/ja/event_file/20061219_its.html
日本自動車研究所が取りまとめたアンケート結果によると、ワークショップ出席者は 81 名（一般 64 名、関係者 17 名）、アンケート回収率 88%（56 名）、アンケート回答者はこのワークショップに興味関心のある一般社会人であり、電気通信メーカや自動車部品メーカ、大学等の研究教育機関に従事するものがほとんどである。

行っている。このアンケートの回答数は 55 名であり、参加者の関心は図 3 に示す通り、プローブ情報システム（31 名）やその個人情報（32 名）となっている。プローブ情報システムの事業者または利用者としてどのような関心・問題を抱えているか調査するためのアンケートの結果としては、例えば、車両のワイパやヘッドライト、ABS などの情報は提供しても良いが、カメラ（画像）や経路情報は求められても提供に同意出来ないという傾向が出ていた（図 4）。カメラ画像や経路情報は、車両の走行速度やワイパ稼働状況等のセンサー情報に比べるとパーソナルデータとしての個人の行動履歴に深く関係があると認識されていると考えられる。なお、これらの情報提供に同意出来ない主な理由は「どのように利用されるか分からないから」「プライバシー情報だから」であり（図 5）、これらの情報の提供同意に必要な条件が「不愉快な思いをしないことが保証される」「社会的な効用」「事業者が信頼できる」であった（図 6）。これは、例えばある区間を速度超過で走行した場合に後からこのデータを論拠として不利益を被ることが心配されたり、何に使われているか分からないという不安があったりすると考えられる。

なお、事業者としては、個人を特定可能なプローブ情報に需要があるかどうかという設問に対して、マーケティング等に有効であるため一部「個人が特定できる」あるいは「年齢と性別が特定できる」データに需要があるとしているが、多くは個人が特定できないプローブデータで十分利用に関心があるという結果になっている。

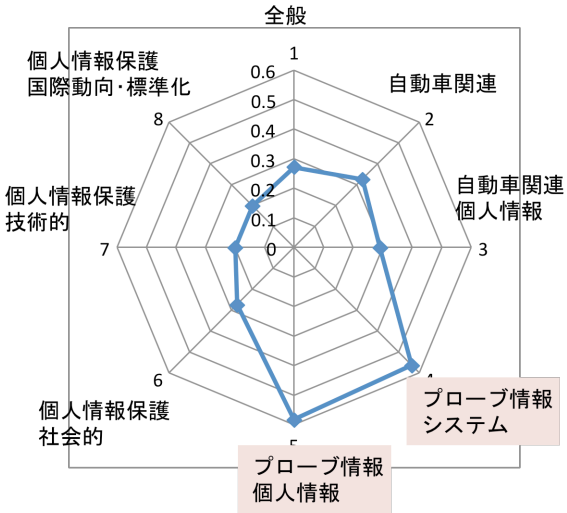

図 3　産業界からの期待（戦略的国際標準化推進事業／標準化先導研究「ITS プローブ情報システムのサービスアーキテクチャ構築に関する標準化」平成 21 年ワークショップにおけるアンケート結果より）

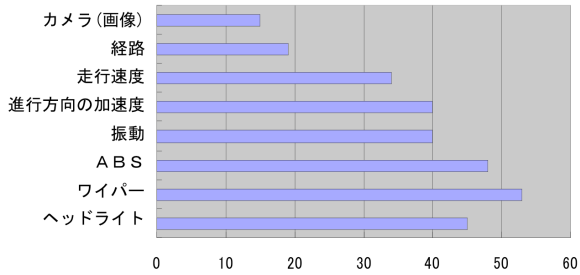

図 4　提供可能なプローブデータ（戦略的国際標準化推進事業／標準化先導研究「ITS プローブ情報システムのサービスアーキテクチャ構築に関する標準化」平成 21 年ワークショップにおけるアンケート結果より）

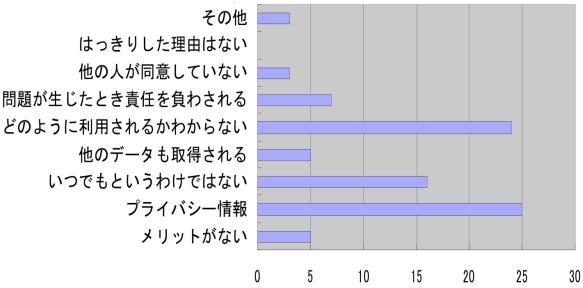

図 5　提供できない理由（戦略的国際標準化推進事業／標準化先導研究「ITS プローブ情報システムのサービスアーキテクチャ構築に関する標準化」平成 21 年ワークショップにおけるアンケート結果より）

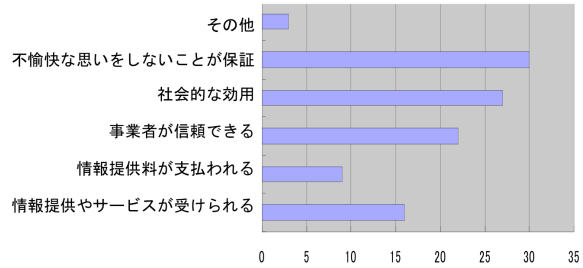

図 6　情報提供同意の条件（戦略的国際標準化推進事業／標準化先導研究「ITS プローブ情報システムのサービスアーキテクチャ構築に関する標準化」平成 21 年ワークショップにおけるアンケート結果より）

このアンケート結果詳細については文献(4)を参照されたい。なお、このアンケートは東日本大震災前に行っている。震災後多くの人の防災意識が変わっていることを考えると、同様のアンケートを再度調査し、その差を比較すること等は今後の研究課題としたい。

プローブ情報システムでは、プローブデータを情報流通基盤としてインターネットを用いて送信し、インターネット上にあるサーバやサービス事業者がそれらのデータを収集・管理・加工・提供している。インターネット上に発信される情報には、メタデータとして情報発信者が用いる ID や IP アドレスなど、何らかの識別情報（ID）や痕跡が残るのが一般的である。プローブ情報システムでは、通信 ID だけでなく会員番号やユーザ ID、機器 ID 等、いくつかの識別子を適宜用いる。通信 ID に関しては、プローブデータの送信（車両や路側からインターネット上のサーバ、サーバ間、サーバからサービス利用者へ）に用いる通信路の暗号化や集

積したデータに関する匿名暗号技術の適用、統計処理の一環としてのk-anonymityの検討など、複数のレイヤで技術的また運用的な対応を検討している。

3.3 プライバシー問題の対応としてのIDの匿名化

統計処理の一環としてのk-anonymityは、プローブデータを送信する際のIPアドレスのような通信IDが永続的に固定であればそれ自体は個人情報でなくても比較的短時間で車両を特定可能であることから必要な対策となる。例えば、プローブデータを取得した位置と時刻情報を同じ通信IDで同じサーバに送信し続ければ、同IPアドレスからの送信情報という時系列データを見るだけでその車両の走行情報が判明し、かつその位置に私有地や特定の時間（その時間にその地域を走行した車両は1台だけだった、等）が含まれていれば個人を特定できる可能性も含まれる。つまり、プローブ情報システムとして「パーソナルデータ利活用のルールが明確に」していたとしても、匿名であるはずのプローブデータが匿名でなくなる、つまり個人の行動履歴、パーソナルデータとなり場合によっては深刻なプライバシーを侵害する事態にも発展する。

そこで、プローブ情報システムにおける匿名性は、技術的・概念的な意味での”厳密な”匿名性ではなく，社会一般的に受け入れられる匿名性を採用することが妥当であるとし、プローブ情報システムにおける「匿名性」は『*情報の発信源（主体：subject）となる人物が「誰であるのか分からない（実名や実体を特定されない）」状態*』であると定義し、同じIDを用いて、どのような場合には実名や実態が特定され、どのような場合には特定されないのかを検討・分類することで、「匿名性」を区別して扱うことを検討している。文献(5)では、この定義と分類に従い、プローブ情報システムにおける匿名性を「実名(identify)」、「解明可能な匿名性(resolvable anonymity)」、「完全な匿名(total anonymity)」の3つに分類して取り扱っている。一般的な匿名は、「解明可能な匿名」のうち、個人の同一性が認められない状態であるとし、情報、および情報群の中に常に異なる識別子が含まれている場合や匿名認証機構等を用いて情報同士の結びつきを切断処理した後の交通情報等がこれにあたるとしている。

なお、匿名性に関する評価指標についても、文献(4)を参照されたい。

プローブデータや車両情報インタフェースの統一と同様に、プローブ情報システムにおける個人情報保護やプライバシーについても、国際的な基準としてISO（具体的にはISO/TC204/WG16）で、日本主導で議論されている。現在は、OECDプライバシーガイドラインを元にプローブ情報システムとして留意すべき項目を策定したISO 24100をベースにして、前述の匿名性に関する分類を中心に議論した個人情報保護およびプライバシー関する国際標準策定が行われている。

4. 議論と考察

ビッグデータのサービス事例としてプローブ情報システムにおけるプライバシー問題への対応については、個人や車両を識別可能な情報としてのIDを起点に整理してきた。この整理によって同じ「プローブ情報システム」と云えども多種多様なサービスが展開されており、統一して共有できる部分とそうでない部分があること、データ自体が匿名であってもシステムとして匿名が保てるとは限らないこと、利用者にとっては、他愛もない（と判断できる）情報は提供しても良いが、利用目的や不利益を被らないことの保証が明確でなければ提供したくない情報もあること、等が明確になった。そこで、ここでは、ビッグデータとして情報を共有可能な形にするにあたり、そのデータがどう使われるのか（公的なサービスか、私的に差別化されるものか）、および、データの種類と利用目的等、議論を分けて検討すべき項目を挙げる。

4.1 議論を分けるべき2種類の「サービス」

プローブ情報システムにおいては、そのサービスを「公的」または「私的」なものの二つに大別して議論ができる。公的・私的、いずれの場合も、情報システムとしては通信形態や利用機器が多岐にわたり、求められるセキュリティレベルも異なる。そのため、必要な技術や設計上の工夫、匿名性確保のための仕組みの組み合わせも多様になり、これらを画一的・定量的かつ客観的に評価するたけの指標を設定することは現実的ではない。しかし、少なくとも扱う情報が単純かつ匿名な統計データで良いのか付加価値のついた情報が必要なのかで、ある程度の安全性の評価が可能であると考える。

公的なサービスとしては、道路交通情報、気象情報、公共交通の運行情報、塵芥車の位置情報等のような主に行政や自治体が提供する情報サービスを想定している。これらのサービスにはプローブ情報システムではプローブデータとして各車両や路側から発信される情報を扱うが、個人（車両）情報やパーソナルデータは必要ないため、車両を特定するようなIDも必要なく、単純に統計的なデータとして扱うことができる。また多くの場合、時系列データが必要であってもそのIDを適宜変更させ、k-anonymityを保った状態でデータをバラバラに扱うことで車両の特定を防止することが可能である。

一方、私的なサービスとしては、各社のプローブ情報システムのようにコンシェルジュサービスやメンテナンス情報提供などより魅力的なコンテンツを提供するため他との差別化を目的とし、より個人の特化したロングテール的なサービスが必要となる。これには個人（車両）情報や個人の行動履歴などのパーソナルデータが不可欠である。

そこで、まずは、そのビッグデータを使って提供するサービスが「公的」なのか「私的」なのかを設定した上で、それが私的な情報サービスである場合、システム設計時点でプライバシーや安全性のことを組み込む「プライバシー・バイ・デザイン」の概念が必要であると考える。この概念に基づき、プローブ情報システムにおける匿名性の分類のような、いわば「**匿名性の強度**」のような指標と、その「**達成深度**」を測ることができるシステムとして設計する。この 2 次元の匿名性評価指標は、必要となる匿名性の強度とシステム規模に応じて採用する技術や運用上の留意点をあらかじめ達成深度として設定（公開）する。この指標に基づき、利用者が適宜システムの匿名性を検証可能な形にすることで、3.2 節で挙げた様な不安を払拭し、ある程度安心してビッグデータのために情報提供し情報サービスを利用することが可能になると考える。

4.2 議論を分けるべき 2 種類の「透明性」

ビッグデータの利活用の前提は、高解像度でフレッシュな多種多様なデータが多量に必要である、というものであった。この大量のデータをソーシャルメディアやウェブサイト、センサ・ネットワーク、各種 ICT サービスのログやオペレーションデータを組み合わせることで得られる「人間の行動履歴・属性情報」や「モノの状態」のデジタルデータで実現している。これに対する不信感、プライバシーや個人情報保護に関する不安感を除くために、ビッグデータにおける「透明性」として、「公的」および「私的」情報サービスいずれの場合にも、利用する「**データの種類**」とその「**利用目的**」を明らかにする必要がある。つまり、「人間の行動履歴・属性情報」や「モノの状態」のデジタルデータの「何の」情報がどのように集められ、それが「何に（どのような調査・解析に）」使われたのか、ということを明らかにする必要があるということである。これらは既に、ビッグデータの内の個々のデータはデジタル化し収集・管理される前に「データを収集すること」に関する規約を事前に約款等に記述・説明し、利用者は（意識的または無意識的に）同意しているが、さらに簡潔に利用者が自覚できる様に明示することで、自分が提供するデータが他の情報サービスとして有効に活用されていることが分かれば、利用者も安心して積極的にデータ提供をすることが可能になる。

なお、収集する「データの種類」はデータ収集時に必要なものが判明しているが、それらを用いて何をするか（「利用目的」）は、当初想定していたもの以外に利用価値を発現したり、他データと組み合わせることよる付加価値に気づいたりした時点で変わる可能性がある。したがって、この二つは明確に議論を分け、一度収集したデータを用いて、同意した目的外にデータが利活用されていないことを技術的または社会的に検証することが可能な第三者機関等の存在や、目的外利用の事実があった時点でデータ収集の停止や提供したデータの削除を実現できる法令など社会的環境の整備に関する議論も必要であると考える。

4.3 今後の課題

プローブ情報システムは、その基本的なデータ形式やシステムの構成を国際標準という形で共有し、国内外の関連業界や政府・自治体がそれに準拠した形で個々のサービスを想定または展開している。他のビッグデータにおいても、玉石混合のデータの中から必要なデータを抽出すること自体が大変な作業になることも多いため、なんらかの形でデータ形式を共有化する必要があると考えられる。たとえば、政府や地方自治体等が所有するデータを広く一般に公開することで、情報の有効活用に資することを目的としているオープンデータでは、「機械判読に適したデータ形式」かつ「二次利用が可能な利用ルールで公開されたデータ」と定義されており、人手を多くかけずにデータの二次利用を可能となるものとされる。行政機関が保有する地理空間情報や防災・減災情報、統計情報などが例として挙げられ、地方公共団体も公共データの整備とオープンデータ推進のための環境整備を積極的に進めている[10]。

さらに、社会的なビッグデータへの期待が増大する一方、そもそもその「データ」は誰のものなのか、という議論もデータのオーナーシップに関する研究として展開されている。利用者が自分の行動履歴データは自分のモノだ、として、プライバシーに関連する「自己情報制御権」を主張する場合もあるが、多くの場合、必要な種類のデジタルデータを目的に応じて収集する事業者がなければビッグデータとして有効な情報にはなり得ない。では、その事業者がデータの所有者と認識されるのか。オーナーシップには「責任」が伴うが、事業者がサービスを継続できなくなった場合、既に収集したデータは誰にどのように委譲されるのか、不正に利用されたり流出したりしたデータには、どのような対応が可能なのか等、収集データに対する責任に関する議論が必要である。

このように、新しい情報端末やサービスが次々と発表され、様々な形でデジタルデータが増大している中で、社会変化や利用者意識の変革も加味して議論すべき項目は多い。また、欧米における「忘れられる権利」や「プライバシー・バイ・デザイン」、またはアジア環太平洋地域においては、日本政府が 2013 年 6 月に参加申請をした APEC 越境プライバシールールシステム（CBPR: Cross Border Privacy Rules System）のように、基本的人権の保障としてのデータ保護に関する新たな規則の提案や環境整備が進む中で、ビッグデータに紐づく個人のプライバシー情報はどう扱われるのか、丁

[10] 2013 年 6 月の主要 8 カ国首脳会議で各国首脳が「オープンデータ憲章」に合意、日本では「電子行政オープンデータ戦略」（2012 年 7 月 IT 戦略本部決定）に基づいて行政機関等が保有する公共データの活用推進を目的としたオープンデータの取組を推進している。2014 年 2 月現在、データカタログサイト試行版（http://www.data.go.jp/）や Open Data METI β版（http://datameti.go.jp/）等が公開されている。http://www.kantei.go.jp/jp/singi/it2/densi/dai4/sankou8.pdf。

寧な議論を重ねる必要があると考える。

5. おわりに

本論文では、ビッグデータへの期待と、それに伴う個人情報およびプライバシーに関する問題として、「本人の同意がなくてもデータの利活用を可能とする枠組み」が導入された場合、個人を特定する情報（ID 等）やプライバシー問題はどのように扱われるべきか、技術だけでなく教育・啓蒙や社会・法整備、情報倫理等の多方面から検討が必要であると問題提起した上で、ビッグデータに紐づいているIDとそれに伴うプライバシー問題を考察するために、ビッグデータを用いたシステムの一つであるプローブ情報システムの今までの取り組みと方向性を紹介した。

プローブ情報システムにおける個人情報およびプライバシー保護は、技術的には通信路の暗号化や集積したデータに関する匿名暗号技術の適用、統計処理の一環としてのk-anonymityの検討など、複数のレイヤでの対応を検討しており、運用上もISOにおける国際標準の場でガイドラインを作成して共通認識を広めている。

同様に、ビッグデータ関連システムのIDやプライバシー問題として、個人情報やパーソナルデータを保護するためには、単純に流通するデータが匿名であれば良いという訳ではなく、プローブ情報システムのように、通信やシステム上のID匿名化を慎重に検討するのと同時に、利活用するサービスが公的なものか私的なものか、また、個人の「何の」情報がどのように集められ、それが「何に（どのような調査・解析に）どう有効に使われた実績があるのか、ということが利用者に検証可能であることが望ましいこと等を考察し、今後検討すべき課題について触れた。

謝辞

本研究は、WIDEプロジェクトiCAR WGの活動の一部、および基準認証事業受託研究の結果等を元に展開したものである。国立シンガポール大学・慶應義塾大学の佐藤雅明博士を始め、関係各位に深く感謝する。

また、本研究の一部は、情報科学研究所研究助成によるものである。

参考文献

(1) Michiko IZUMI, Masaaki SATO, Hideki SUNAHARA, ″Requirements for protection methods of personal information in vehicle probing system″, SAINT2007, 2007.

(2) Masaaki SATO, Michiko IZUMI, Hideki SUNAHARA, Keisuke UEHARA, Jun MURAI, ″Requirements for protection methods of personal information in vehicle probing system″, ICWMC07, 2007.

(3) Masaaki Sato, Michiko IZUMI, Kanae MATSUI, Hiroshi ITO, Keisuke UEHARA, Jun MURAI, ″CRITERIA FOR PRIVACY AND INTEGRITY PROTECTION IN PROBE VEHICLE SYSTEMS″, 18th ITS World Congress, 2011.

(4) 和泉 順子, 佐藤 雅明, 砂原 秀樹,″プローブ情報システムへの匿名性評価手法導入のための一考察″, 情報処理学会第45回EIP研究会 , 2009,

(5) 佐藤雅明, ″インターネットと自動車情報の融合″, コンピュータソフトウェア, Vol.31, No.3, p.17-31, 2014.

(2014年9月30日受付)
(2014年12月3日採録)

6. 付録（用語定義）

6.1 ビッグデータ

ビッグデータとは、総務省の HP[11]では鈴木良介氏の著書[12]から「ビッグデータを『事業に役立つ知見を導出するためのデータ』とし、ビッグデータビジネスについて、『ビッグデータを用いて社会・経済の問題解決や、業務の付加価値向上を行う、あるいは支援する事業』と目的的に定義している例がある。」とした上で、「ビッグデータは、どの程度のデータ規模かという量的側面だけでなく、どのようなデータから構成されるか、あるいはそのデータがどのように利用されるかという質的側面において、従来のシステムとは違いがあると考えられる」ものであると記述している。ビッグデータを構成するデータ群は、ソーシャルメディアデータ、ウェブサイトデータだけでなく、センサデータ、オフィスデータ、ログデータ、オペレーションデータなど出所が多種多様であり、各データを連携させることでさらなる付加価値の創出が期待されている。

6.2 プライバシー

「プライバシー」という言葉は、分野によってその定義や認識は異なる未整理の概念である。1880 年にハーバード大学法学紀要に発表した論文 "The Right to Privacy"で初めてプライバシー権の法理を主張したアメリカ最高裁判事のルイス・フランダイスは、プライバシーを「もっとも包括的な権利であり、かつ文明人によって最も価値を置かれている権利」であると宣言しているが、現在ではこの権利だけに留まらない。思想や言論、監視からの自由、自己情報コントロールの権利、自分自身の評判や検索からの保護等も含まれており、プライバシーという概念の定義はなおいっそう困難になっている。

法律分野においては、ダニエル・J・ソローヴによると「ある行為がプライバシーを侵害しているかどうかアセスメントする際、法定や政策立案者はしばしばプライバシーをただ一つのものと考えがちであった。その結果、彼らは重要な差異があるにもかかわらず複数のプライバシー問題をごっちゃにするか、問題を全体として認識するのに失敗してきた。要するに、プライバシー問題は、法律においてしばしば誤解されるか不整合な理解をされてきたのである。」とされている[13]。なお、平成 25 年度版情報通信白書第一部第一節によると、「プライバシーについて一般的に規定した法律は存在しないが、判例法理上、プライバシーは法的に保護されるべき人格的利益として承認されてきた。また、最近ではプライバシー保護の対象となる情報は拡大傾向にある。」とされる[14]。

Geek なページ[15]では、情報のリテラシとデリカシーを分けて考察した上で、「人は他人の情報は根堀葉掘り知りたい反面、自分の情報は知られたくありません。例えば、芸能人のプライバシーは知りたいけど、自分のプライバシーを新聞などに書かれると恐らく怒り狂います。他人の情報を暴露したいという魅惑は、そこら中に転がっています。 その魅惑とどう付き合うかや、書きたくなったら相手の了承を得たりするというステップを踏むか踏まないかなどが「情報デリカシー」と言えるのかも知れません。」と記述している。

6.3 個人情報保護法の目的

個人情報保護法の目的（第 1 条）は、「この法律は、高度情報通信社会の進展に伴い個人情報の利用が著しく拡大していることにかんがみ、個人情報の適正な取扱いに関し、基本理念及び政府による基本方針の作成その他の個人情報の保護に関する施策の基本となる事項を定め、国及び地方公共団体の責務等を明らかにするとともに、個人情報を取り扱う事業者の遵守すべき義務等を定めることにより、個人情報の有用性に配慮しつつ、個人の権利利益を保護することを目的とする。」と定義されている。

---

11 http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc121410.html（2014 年 8 月現在）

12 鈴木良介「ビッグデータビジネスの時代」（平成 23 年 11 月）p.14

13 「プライバシーの新理論 –概念と法の再考–」ダニエル・J・ソローヴ著、大谷卓史訳、みすず書房、2013 年 6 月。

14 ここでの判例は「宴のあと」事件（東京地裁昭和 39 年 9 月 28 日判決）が参照されており、また、拡大傾向については早稲田大学江沢民講演会名簿提出事件（最高裁平成 15 年 9 月 12 日第二小法廷判決）が参照されている。

15 http://www.geekpage.jp/blog/?id=2009/11/9/1 「情報デリカシー」