

# パーソナルデータ

Personal data

東京大学大学院情報理工学系研究科ソーシャル ICT 研究センター教授

橋田 浩一

1986 年より 2001 年まで電子技術総合研究所。その間 1988 年から 1992 年まで（財）新世代コンピュータ技術開発機構に出向、2001 年から 2013 年まで産業技術総合研究所。2013 年から現職。専門は自然言語処理、認知科学、サービス科学など。日本認知科学学会会長、言語処理学会会長等を歴任。

✉ hasida.koiti@i.u-tokyo.ac.jp

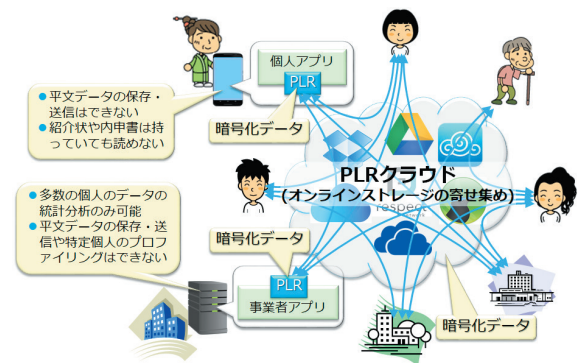
日本を含むほとんどの国で GDP の 60 ~ 70% を家計消費（小売を含む生活者向けサービス）が占める。勤労者としての個人に対するサービスは GDP に含まれないがその売り上げは GDP の 30% ほどと思われる。さらに家事や育児など無料の C2C サービスも貨幣価値に換算すると GDP の約 20% に相当すると考えられる。したがって、個人向けサービスは社会の価値のほとんど（GDP の 110 ~ 120%）を生み出していると言える。その価値をさらに高め生活や業務の質を向上させるには、サービス受容者のリッチなパーソナルデータ（個人に関するデータ）を簡単に利用できるようにする必要がある。

AI の運用と開発においてもパーソナルデータが簡単に利用できる必要がある。AI の運用では、ほとんどの場合 AI によるサービスの相手は個人であり、その個人に関するリッチなデータが簡単に得られないと AI はその人に合ったサービスを提供できない。また、そのような AI の開発においても多数の個人のパーソナルデータを収集して分析する必要がある。

個人向けサービスの価値を最大化するには、パーソナルデータの活用を最大化するためにパーソナルデータの分散管理が必要である。分散管理とは、管理者を本人（または代理人）のみとする、すなわちデータの管理を個人に分散することである。集中管理方式では管理者が多数の個人のデータに（法律や契約に反して）アクセスすることが技術的に可能であるため情報漏洩のリスクが大きくそのリスクを管理するコストを含む運用経費が高いため、拡張性が低い。実際、集中管理システムからの大規

模な情報漏洩事件等が頻繁に生じている。また、たいていの人には他人に知られたくない秘密があり、そのような機微な情報の活用は特に高い価値を生む可能性がある。この可能性を生かすには本人がその情報を他者に開示せずに活用できる必要があり、それにも分散管理が必須である。

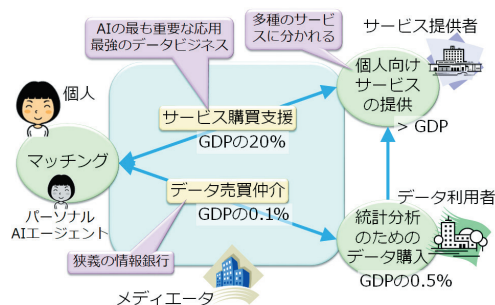
パーソナルデータを本人が管理し活用するための仕組みを PDS (personal data store) と呼ぶが、そのうち分散管理に基づく PDS がパーソナルデータの活用を最大化するために必須であることがわかる。そのような分散 PDS の一例である PLR (personal life repository) の仕組みを下図に示す。



PLR は個人アプリや事業者アプリに組み込まれるミドルウェアである。PLR が PLR クラウドとデータを共有することによって、PLR 利用者（個人や事業者）同士がデータを共有できる。ここで PLR クラウドは Google ドライブ等のオンラインストレージの集合であり、各 PLR 利用者が自分のデータ領域を保有するので、PLR によるサービスは利用者数が何十億になってもアプリの保守費だけで運用可能である。個人端末で

も事業者のコンピュータでも PLR クラウドでも、保存される非公開データはすべて暗号化し、DRM (digital rights management; 暗号化されたデータにアクセスする手段の限定) を適用する。したがって、図のフキダシのように、たとえばデータ主体本人も暗号を解いた平文のデータをファイルに保存したり外部に送信したりする手段を持たず、ゆえに利用者の過失によるデータ漏洩はあり得ない。また、医師の紹介状や学校の内申書のデータを患者や生徒本人が物理的には保有していてもその内容がわからないようにしたり、医療機関由来のデータを患者本人が開示する相手を医療機関が制限したりすることも可能である。同じく DRM によって、権限定義 (データの各部に対する各利用者のアクセス権限を定義してそれが改竄されたりしないように管理すること) が可能であり、どのデータを誰がどう使ったかに関する追跡可能性 (traceability) も成立する。データの安全性 (security) のうち機密性 (confidentiality) と完全性 (integrity) も DRM で保証され、また PLR クラウドは通常の情報システムよりはるかに可用性 (availability) が高い。

下図に示すように、パーソナルデータの活用は、マッチング、特定個人へのサービスにおける本人のデータの1次利用、および統計分析 (2次利用) の3種類に分類できる。マッチングとは、パーソナルデータに現われる本人のニーズに合うサービスを選定すること (個人が持つデータを使う統計分析の選定も含む) である。このマッチングは個人とパーソナル AI エージェントとの対話によってインタラクティブになされることが多いだろう。マッチングのための計算は個人端末が担っても図中のメディアータ (個人とサービス提供者等の間を仲介する事業者) のサーバコンピュータが担っても良い。後者の場合も前述の DRM により、メディアータの社員が顧客の個票データを閲覧したりすることは技術的に不可能とする。



パーソナルデータの以上3種類の活用法のうち最も収益性が高いのはマッチングである。前述のように個人向けサービスは GDP を越える価値を生むから、パーソナル AI エージェントによるサービスのマッチングに基づく購買支援の手数料の合計は、GDP の約 20% であり、国家予算に匹敵する金額になるだろう。一方、個人向けサービスは医療や教育やモビリティなど多くの種類からなり、種類ごとに異なるデータ処理の仕組みが必要で、それらのサービスの価値の大半はデータ処理以外によるものだから、各種のサービスにおいてパーソナルデータが生む価値はさほど大きくない。また、統計分析のためのデータ収集を仲介する事業 (狭義の情報銀行) の市場規模は GDP の 0.1% 程度だろう。たとえば日本の場合、各個人が自分のデータを企業等に売った場合の1年間の平均的な売り上げは3万円未満で、全国では4兆円未満だろうが、その売買を仲介する手数料の合計は高々5~6千億円と推定される。

この購買支援に関与した個人や事業者にその貢献度に応じて購買支援手数料の収益を分配するのが自然である。とりわけ、マッチングに用いたパーソナルデータを本人に提供した事業者に収益を適切に分配すれば、各事業者はより良質のデータをより多く作成して本人に提供することでより多くの収益が得られる。こうして、リッチなパーソナルデータの生成とポータビリティが促進され、データの活用が拡大することにより、分配される収益がさらに増大し、なおさらデータの生成とポータビリティが促されるという、パーソナルデータエコシステムが成立する。サービスの購買支援の方がデータ収集仲介より分配可能な収益が圧倒的に大きく、ゆえにパーソナルデータエコシステムの基盤として強力である。

個人向けサービスの購買を支援するためのマッチングを担うパーソナル AI エージェントは AI の応用のうちで経済的利益が最も大きく、ゆえに急速かつ広範に普及する可能性が高い。それは、各個人の行動を少なからず左右して GDP を上回る価値の交換と国家予算規模の富の分配を司り、さらにデータの2次利用も促進することによって、生活や産業や政治や文化にきわめて大きな影響を及ぼす。そのようなエージェントには高度な中立性と倫理性が要求され、AI 関連技術の中でとりわけそのガバナンスが国際的にも重要な課題になるだろう。