PLRとパーソナルデータエコシステム

PLR and Personal-Data Ecosystem

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。日本認知科学会会長、言語処理学会会長等を歴任。

hasida.koiti@i.u-tokyo.ac.jp

1　データポータビリティと PLR

ほとんどの国々で GDP の大半を家計消費（小売を含む個人生活者向けサービス）が占め、さらに GDP にカウントされない個人向けサービス（勤労者としての個人に対するサービス、および家事や近所付き合いなど無償の C2C サービス）の価値の合計も同程度と考えられるから、世の中のほとんどの価値は個人向けサービスに由来すると言えよう。したがって、個人の幸福と社会の発展のため、社会全体でパーソナルデータの活用を促進することにより個人向けサービスの価値を高めることが最重要課題である。

パーソナルデータの使用には原則として本人同意が必要であり、またパーソナルデータの活用による最大の受益者は本人である。ゆえにデータポータビリティ（data portability；本人がパーソナルデータの管理権限を持って自由に使えるようにすること）によってパーソナルデータの活用が促進され、個人向けサービスの価値が高まる。さらに、多数の個人から本人同意だけでパーソナルデータを簡単に収集できるようになり、そのデータの二次利用によって技術やサービスの価値を高めるのも容易になる。

2018 年 5 月末にヨーロッパで施行され事実上の世界標準になりつつある GDPR だけでなく、GDPR に先立って 5 月初めに中国で採用された個人情報セキュリティ規準においてもデータポータビリティの権利が明文化されている。日本の個人情報保護法も 2020 年の再改正の際にデータポータビリティ権を明記する可能性が高いと思われる。公的機関が保管するパーソナルデータを 2020 年からマイナポータルで本人に開示するという日本政府の方針もそれと符合する。これによってデータポータビリティの普及が期待される。

GDPR は、人権を守るために企業活動を制限する側面のみが注目されがちで、データポータビリティが個人の利便性のみならず事業者の収益も向上させる点は残念ながら世間ではあまり認識されていない。データポータビリティが普及すれば、事業者がパーソナルデータの管理にまつわるリスクとコストを免れるだけでなく、個人に集約され名寄せされて価値の高まったパーソナルデータが本人同意で容易に活用される。それにより、社会全体でパーソナルデータの活用が盛んになって全事業者の収益の合計が増大する。後述のように、パーソナルデータの本人への提供等によってその増収に貢献した事業者にも収益を分配することにより、あらゆる事業者の収益が増えるはずである。そうすれば、有用なパーソナルデータを作って本人に提供することが持続的な収益につながるから、良質のパーソナルデータの生成・活用が促されるだろう。

データポータビリティに応じて、個人が本人のパーソナルデータを自由に活用するツール、つまり PDS（personal data store）が必要である。しかし、PDS の多くは集中管理型であるため、データポータビリティ、安全性、経済性を十分確実に満たすことが難しい。

PDS に限らずほとんどの情報システムは集中管理されている。つまり、図 1 のようにシステム全体の管理者がシステム内の全データを集中的に管理する。この方

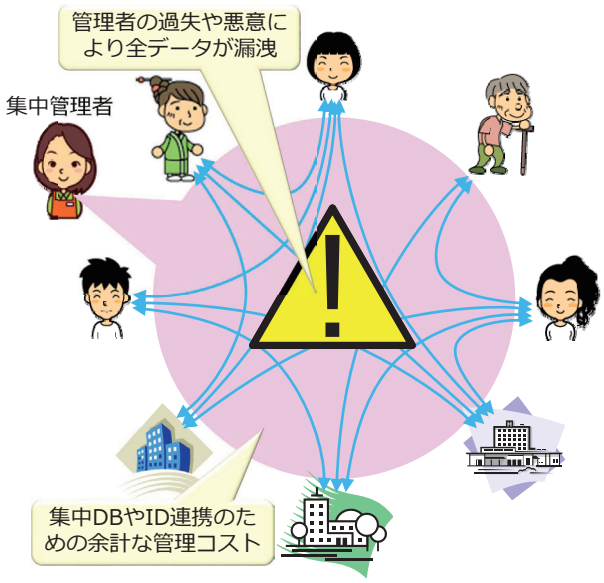

図１　集中管理のコストとリスク

式は、集中管理のためのハードウェアやソフトウェアの導入・運用コストを要するだけでなく、管理者がシステム内の全データにアクセスすることが（契約や業務規則によって禁じられていても）技術的には可能なので、管理者の過失や悪意によって全データが漏洩するリスクを生み、実際にそのような情報漏洩事件は枚挙に暇がない。すなわち、集中管理はわざわざコストをかけてリスクを高める。

分散管理は管理コストをかけないことによりデータ漏洩等のリスクを防ぐ。各データの管理者はそのデータ主体（個人や事業者）のみであり、データ主体の意思に反してデータにアクセスすることが技術的に可能な管理者がいない。つまりデータ主体の許可がなければデータへのアクセスが技術的に不可能であり、ゆえにシステム内のデータがすべて漏洩することはあり得ない。また、データ主体によるデータの活用や共有を妨げる者も存在しない。この方法は、パーソナルデータや企業秘密等の非公開データの管理・運用に適している。

PLR（personal life repository）は、数少ない分散管理に基づく PDS のひとつであり、個人でも事業者でも同様に利用できるおそらく世界で唯一の PDS でもある。その仕組みの概要を図２に示す。

PLR の本体は一種のミドルウェアであり、利用者（個人でも事業者でもよい）の PLR アプリ（PLR 本体と直接連携して PLR のデータにアクセスするアプリ）と連携しつつ PLR クラウド（Google ドライブ等の出来合いのオンラインストレージの寄せ集め）を経由して他

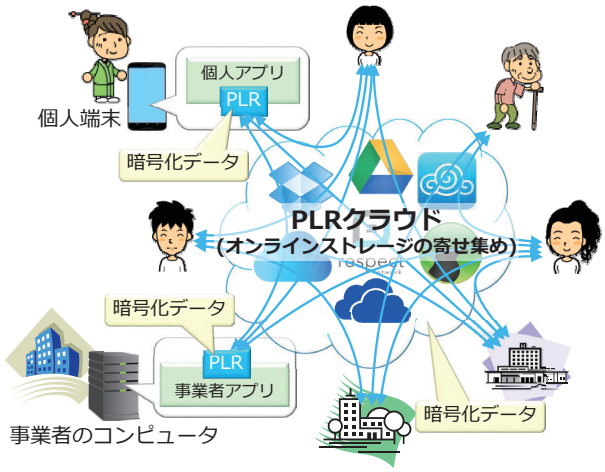

図２　PLR の仕組み

の利用者の PLR とデータを共有する。PLR クラウドにおいても利用者の端末（スマートフォンや PC やサーバコンピュータ）においても保存される非公開データはすべて暗号化され、暗号を解く鍵は本人が明示的に指定した者にしか開示しないので、通信の秘密やプライバシが守られる。

さらに、暗号を解いた状態の平文データをファイルに保存したり外部に送信したりできないように PLR アプリの機能を限定するので、まとまった平文データを取り出すにはそのような限定のない不正なアプリをインストールしてそれに PLR のパスフレーズを入力する必要があるが、それには明確な意思が必要である。つまり、利用者の過失によるデータ漏洩はあり得ない。事業者が PLR を利用する場合は、正しく機能限定されたアプリに署名を施しておき、OS が署名を検証して不正なアプリを排除すれば良い。すると、まとまったデータを平文で取り出すには不正な OS をインストールする必要があるが、通常の管理の下で OS を人知れず入れ替えるのは不可能だから、セキュリティを確実かつ安価に担保できる。

また、PLR と PLR アプリの提供者にとって PLR クラウドの運用コストはゼロだから、PLR は利用者数が何億になってもアプリの保守費用だけで運用できる。各 PLR 利用者のデータの量がオンラインストレージにおける所定の上限を越えると課金が発生するが、それほど PLR を使い込んでいる利用者であれば料金を支払うだろう（PLR の提供者にはコストがかからない）。また PLR 利用者は一定の割合でオンラインストレージの有料利用者になるだろうから、PLR はオンラインストレー

ジ事業者の収益にも貢献する。

PLR クラウドはオンラインストレージの寄せ集めなので、オンラインストレージの利用者数の上限が PLR の利用者数の上限となる。現在のオンラインストレージの主流である Google ドライブや OneDrive 等のパブリッククラウドストレージは全体として数十億人が利用可能なので、PLR も数十億人が利用可能で、さしあたりはそれで十分である。パブリッククラウド全体の容量は増大し続けているが、将来は個人用情報機器の記憶容量と処理能力が高まることにより、IPFS 等の P2P のファイル共有システムが PLR クラウドとして永続的に使えるようになる可能性が高い。いずれにせよ PLR の運用は持続可能である。

PLR の本体はすでに Android、iOS、Java の環境で稼働している。また、PLR の基本機能（認証、暗号化 / 復号、通信）に加えてデータを作成・共有・活用する機能とオントロジーをカスタマイズする機能を備え、名簿管理や SNS が可能な PLR 統合アプリの Android 版と iOS 版をそれぞれ Google Play ストアと App Store から 2018 年内および 2019 年初頭に無料で一般公開する予定である。一方、さまざまなサービスに PLR を用いる実証実験や商用化が進みつつある。

そのうち教育への応用について紹介する。東京大学・理化学研究所・埼玉県の共同研究により、PLR による e ポートフォリオ（電子学習録）の仕組みを開発して埼玉県等の高校で運用する予定である。2020 年度以降の大学入試では、受験生が高校在学中の課外活動等の電子データを e ポートフォリオで予め作っておいて出願時に大学にそのデータを提出し、大学は入試の成績だけでなく e ポートフォリオのデータ等も勘案して合否を決めるとの方針が文部科学省から示されている。そこでは以下の条件を満たす必要がある。

（1）データポータビリティ
（2）e ポートフォリオと校務系システム（成績や出欠を管理するシステム）との連携
（3）校外から校内の情報システムへの不正アクセスの防止
（4）生徒による校務系システムへの不正アクセスの防止

学習者の学習や進学や就業の機会を最大限に確保するために e ポートフォリオ等の教育・学習データのポータビリティが必須であることは自明だろう。図 3 のように PLR を用いて e ポートフォリオと校務系システムを連携させる（e ポートフォリオが校務系システムと同じく校内のシステムである場合には、e ポートフォリオを校務系システムと同じく PLR で PLR クラウドとつなぐ）のが、（1）～（4）を最も明確に満たす方法であり、しかもそれは想定される他の方法よりはるかに安全かつ安価と考えられる。したがって、この方法が全国の高校生約 330 万人とその保護者の多くに普及する可能性が高い。

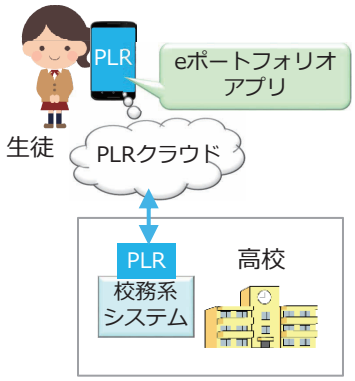

図 3　PLR による e ポートフォリオの運用

また、e ポートフォリオを「スタディ・ログ」に拡張し、小学校から大学を経て就職後も各個人が自らの学習等のデータを持ち続けて生涯学習やキャリア形成に活用するという文部科学省の構想も、この e ポートフォリオの仕組みの拡張として PLR で安全かつ安価に実現できる。もちろんその利用者は高大接続 e ポートフォリオの場合よりはるかに多い。

2　マッチングと VRM

個人向けサービス（商品の小売りを含む）の販売には、個人のニーズとサービスとのマッチングが必要である。Amazon の推薦や Web 等のターゲティング広告は事業者側でのマッチング（CRM：customer-relationship management）によるが、どの事業者も各個人のデータのほんの一部しか使えないのでマッチングの精度が低く、しかもそのデータを管理するコストが高い。一方、データポータビリティが普及してパーソナルデータが本人に集約されると、個人の側でのマッチング（VRM：vendor-relationship management）の方が望ましい。個人はどの事業者よりも網羅的に本人のデータを使うことができ、他人に知られたくない機微な個人情報等も使

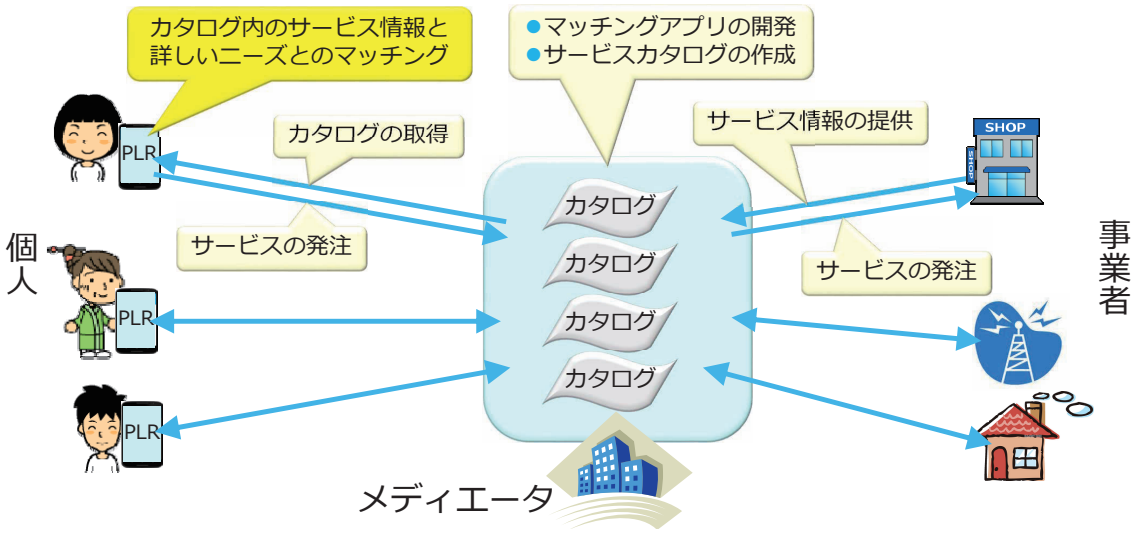

図４　メディエータによる VRM の支援

えるのでマッチングの精度が高い。また個人情報を本人が他者に開示したり事業者が保管したりしないので安全かつ低コストである。

図４のように VRM を支援する仕組みをメディエータと呼ぶ。メディエータは一部の個人からパーソナルデータを取得し、図の右側のサービス提供事業者からサービスの情報を取得することにより、マッチングアプリを開発しそれに入力するサービスカタログを作成して個人に提供する。

メディエータはパーソナルデータを保管しないので個人を囲い込むことができず、またサービスの情報を囲い込むことにより事業者を囲い込むこともできない。したがって、サービスの種類や地域に応じて多数のメディエータが現われるはずである。

マッチングの対象である個人のニーズとサービスの情報は、個人とサービス事業者が取引（サービスの授受）に参加するための条件（取引条件）である。一般的なマッチングは記号的なマッチングと統計的なマッチングの組み合わせであり、取引条件もこれら２種類のマッチングに用いられる２種類に分かれるだろう。

記号的なマッチング（おそらく Pareto 最適解の導出のようなゲームの求解）にかかる取引条件としては、たとえば図５のように、個人が健康管理をしてもらいたいとかそのために病歴や服薬のデータを開示しても良いとか月に 300 円まで払っても良いなどということが考えられる。事業者のサービスとしても、病歴やバイタルデータを開示してもらいたいとか月額 200 円で良いとかいうことが考えられる。

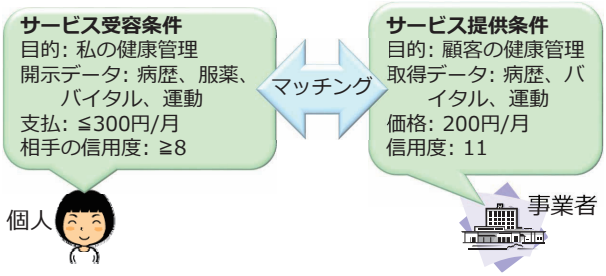

図５　記号的マッチング

保険の約款やスマホのアプリの利用条件のような複雑な取引条件を各個人がいちいち吟味した上でサービス受容に同意するか拒否するかを判断するのは現実的でない。実際、そうした利用条件等はほとんど読まれない。したがって一般に、図４のようなマッチングは自動化すべきである。

統計的なマッチングに用いられる取引条件は、個人の体質などの属性やサービスの利用履歴だろう。それらに関するマッチングは、たとえば図６のようなニューラルネットワークによるだろう。

図６の制限ボルツマンマシンの可視層の各ユニットは、ある個人による１つのサービスの評価または当該

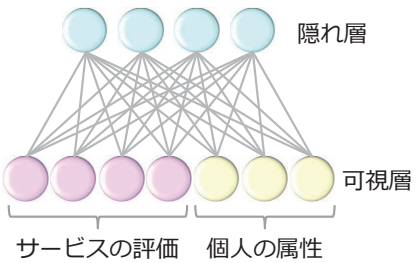

図６　統計的マッチングのためのニューラルネットワーク
（連続値の制限ボルツマンマシン）

個人の 1 つの属性（年齢、性別、住所、バイタルデータなど）を表わす。ある個人に対するマッチングの実行は、当該個人によるサービスの評価と当該個人の属性のうち既知のものを可視層に入力して未知のものを推定する処理である。

サービスの種類は非常に多いので、全サービスを複数のカテゴリ（たとえば成人男性用スーツ、整髪サービス、健康食品など）に分け、複数のニューラルネットワークの各々が 1 つのカテゴリに属するサービスのカタログの分冊を表現する必要がある。制限ボルツマンマシンは O（可視層のユニット数×隠れ層のユニット数）サイズのデータで表現できるので、たとえば隠れ層のユニット数が数十程度でサービスが 1 万種類程度ならカタログの分冊のサイズは 1MB 程度以下になるから、個人端末にカタログをダウンロードしてサービスとのマッチングを個人端末内で実行することが現実的に可能だろう。そのようにサービスのカテゴリを予め設定しておけるかどうかは今後の研究テーマである。

3 パーソナルデータの活用と価値

パーソナルデータの活用法には下記の 3 種類がある。

(a) 個人のニーズとサービスとをマッチングする

(b) 通常の個人向けサービスの質を本人のデータによって高める（一次利用）

(c) 多数の個人からデータを収集して統計分析や機械学習に用いる（二次利用）

これらの間の関係と関連事業の市場規模を図 7 に示す。

自動マッチングを行なう個人アプリによってマッチングしたサービスを購入することもできるとすれば、そのアプリを個人に提供することによって個人向けサービスの販売代行業が成立する。個人向けサービスの末端価格の総和は家計消費と（GDP にカウントされない）勤労者向けサービスの市場規模の合計（国内ではおそらく 500 兆円/年程度）であり、また家事や育児など無料の C2C サービス（貨幣価値に換算すると 100 兆～140 兆円/年）も合わせると個人向けサービス全体の価値は 600 兆円/年以上と考えられる。販売代行手数料はサービスの末端価格の 10% 以上だから、無料サービスのマッチングにも課金できれば、この販売代行業の国内市場規模は 60 兆円/年以上となる。

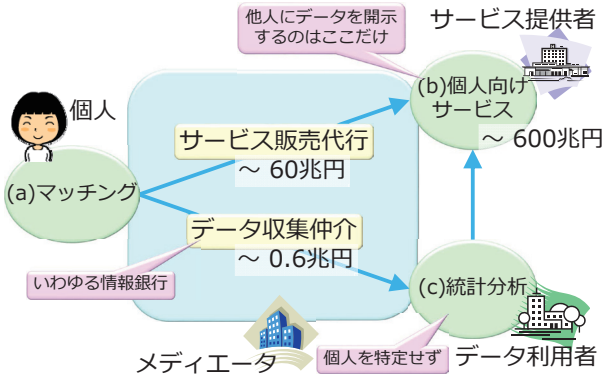

図 7　パーソナルデータを用いる事業の市場規模

これに対し、(b)の一次利用を伴う個人向けサービスそのものは多数の具体的なサービスの種類に分かれ、各種のサービスにおけるパーソナルデータの一次利用の市場規模は 60 兆円/年をはるかに下回るだろう。たとえば国内の自動車産業全体の規模が 60 兆円/年程度だが、自動運転等におけるデータ活用の価値はそのごく一部に過ぎない。一方、(3)の二次利用のためのデータ収集は狭義の情報銀行（三菱 UFJ 信託銀行等が近々始めると言われている事業）の業務である。そのデータ収集に係る国内でのキャッシュフローは 4 兆円/年程度以下（3 万円/人年以下×1 億 3 千万人）であり、情報銀行の収入はその仲介手数料なので高々 6 千億円/年程度だろう。なお、IDC の調査によれば国内の（c)の市場規模は 3 千億円/年程度とのことである。

以上をまとめると、個人向けサービスの販売代行業は市場規模が 60 兆円/年に達し、前述のマッチング技術（記号的マッチングと統計的マッチングの組合せ）がその販売代行の共通基盤として経済的価値が最大の人工知能技術である。その意味において、自動マッチングは人工知能の最も重要な研究テーマと言えよう。

メディエータは個人からパーソナルデータを集めて機械学習に用いることによってマッチングアプリ（およびサービスカタログ）を開発するが、前述のように OS がアプリの署名を検証する等の方法により、そのデータの用途をこの機械学習に技術的に限定することが可能である。つまり、そのデータを第三者に提供したり、人間が生データにアクセスしたり、プログラムが特定の個人をプロファイリングしたりすることを技術的に禁止できる。これにより、メディエータのコストとリスクが最小限に抑制される。

メディエータの収入は販売代行手数料とデータ収集仲

介手数料からなる。メディエータは、マッチングに使われたデータを個人に提供した事業者に対し、そのデータがマッチングに貢献した度合い応じて、その手数料の一部を分配する。その様子を図８および図９に示す。図８は販売代行、図９はデータ収集仲介の場合である。

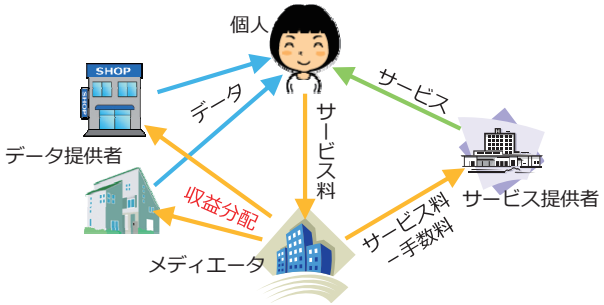

図８　販売代行の収益分配

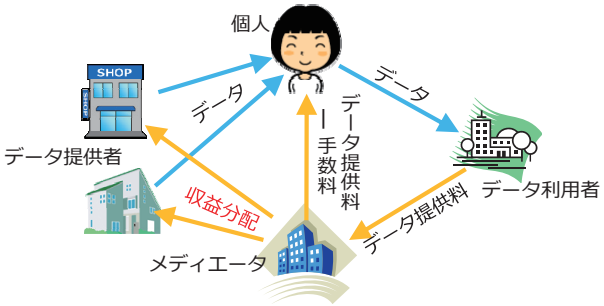

図９　データ収集仲介の収益分配

前述のように、社会全体でパーソナルデータの活用が盛んになれば全事業者の収益の合計が増大するが、パーソナルデータの本人への提供等によってその増収に貢献した事業者にも収益をこのように分配することにより、あらゆる事業者が増益を享受できる。パーソナルデータを顧客本人に提供するとそのデータが顧客から競合他社に渡ったり顧客がそのデータを使って他社からサービスを購入したりする可能性があるため、データポータビリティに対して否定的な事業者が多い。しかし、顧客が自分のデータを売ったり他社からサービスを受けたりすることによって顧客や他社に生じた収益が自社の貢献度に応じて自社に分配されるとすれば、逆に多くの事業者がデータポータビリティに積極的に対応するだろう。

それによってあらゆる参加者が便益を享受できるパーソナルデータエコシステムの概要を図 10 に示す。そこでは、各事業者がパーソナルデータを本人に提供することによってパーソナルデータが本人に集約され、こうして名寄せされ質が高まったパーソナルデータを(a)～(c)に活用することでエコシステム全体として大きな価値(金銭的収益と他の便益)が生まれる。さらに、メディエータ

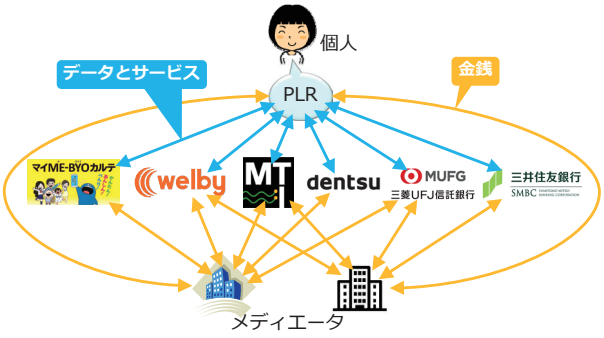

図 10　パーソナルデータエコシステム

がその収益を事業者に適正に分配することで各事業者の収益が増す。事業者がより良質のパーソナルデータをより多く本人に提供して本人が自由に使えるようにすればするほど自らの収益が増えるから、良質のパーソナルデータの生成とそのデータのポータビリティが促進され、それによりエコシステムが持続的に発展する。他の事業者がパーソナルデータを本人に提供した場合も、エコシステム全体の収益が増えるので自社の収益も増える。以上のような意味で、このエコシステムにおけるデータ流通は協調領域に属する。逆にデータを囲い込んでも無意味である。

この図のようにメディエータが通常の事業者と明確に区別され特別な立場にあるのは、メディエータが通常の個人向けサービスを提供せず、また前述の通りパーソナルデータを保管しない（たとえ物理的には保管してもアクセス権限を持たない）からである。パーソナルデータを持たず、顧客に提供するアプリにより顧客の手もとでマッチングすることで、精度が高くリスクとコストが低い販売代行が可能になる。もし仮にメディエータが顧客のデータを保管してマッチングに用いるとすれば、情報漏洩等のリスクが生じるので、機微な個人情報をメディエータに開示しない個人が多く、したがってマッチングの精度が低くリスクとコストが高いだろう。

PLR との連携によるデータ共有はネットワーク効果を持つ。それは、通信サービスのネットワーク効果と同じく、他の多くの利用者とつながるメリットを含む。しかしさらに、多くのデータ提供者が参加することによって、より多くのパーソナルデータが本人に集約され、データの価値が高まる、というメリットをも含む。この意味においてパーソナルデータエコシステムのネットワーク効果は通常のネットワーク効果を越える。その点に関する理解を広めることがエコシステムを普及させるために重要であろう。