分散 PDS: 個人の尊厳と公共の福祉

橋田 浩一（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）

1. 個人データの管理

データの管理とは、データを保管しつつ、データの利用の可否を決定するということである。管理者の意思または過失によってデータが利用されたり漏洩したりするわけである。

個人データの**集中管理**(centralized management)とは、管理者が多数の個人のデータをまとめて管理することである。ゆえに集中管理の下では、多数の個人のデータを一挙に利用することができ、また多数の個人のデータが一挙に漏洩することがあり得る。集中管理においては、定義により、個人データの利用に際してその都度本人の同意を求めないので、多人数のデータを活用するのが簡単である。したがって集中管理は、本人たちのメリットがあまり明確でないデータの利用に適している。それは典型的には多数の人々のデータの分析であり、たとえばある疾患の治療法を発見するために多数の患者から集めたビッグデータを分析する場合などが考えられる。

逆に、本人のメリットが明確なデータの利用には集中管理は適さない。これは、管理者の利害と本人の利害が一致しないことが多いからである。たとえば、病院が治療のデータを他の病院や診療所と共有すれば患者のメリットになるはずだが、従来はデータを共有しても病院が儲からなかったので、医療データの共有はほとんど進んでいない。

一方、個人データの**分散管理**(decentralized management)とは、管理者が 1 人分の個人データのみを管理することだが、その管理者は典型的には本人または代理人や後見人であろう。分散管理の下では一挙に利用されたり漏洩したりし得るデータは 1 人分のみである。したがって、たとえば個人データが数千万人分あるとすると、集中管理よりも分散管理の方が数千万倍安全である。また、分散管理においては個人データの利用が本人(代理人)の同意に基づくので、分散管理は本人のメリットが明確なデータの利用に適する。たとえば、ある病院での治療の記録を別の病院で開示することによって安全で効果的な治療を受ける場合などが考えられる。逆に、上述したビッグデータの分析のように本人のメリットが不明確なデータ利用のためには分散管理は不適切かも知れない。

このように、個人データの集中管理と分散管理はそれぞれに有用で相補う関係にあり、いずれも必要だが、問題は分散管理が実際にはほとんどなされていないことである。前記の通り、個人データの分散管理は本人のメリットを高めるために有効だが、現在は個人が自分のデータを体系的に管理できていないことが多く、過剰な集中管理と相俟って、それが個人のプライバシを脅かすのみならず、B2C サービスの価値の向上や市場の拡大を阻害している。前述の医療の例のように、個人が本人のデータを自らの判断に基づいて自ら指定する事業者と簡単に共有できれば、各 B2C サービスの価値が高まり、市場全体のパイが

大きくなるはずである。さらに、商品やサービスの利用等に関するデータを個々の消費者が管理し社会的に共有し分析して事業者を比較評価することで競争環境を整え、その産業の国際的な競争的および他産業との競争力を高めることによっても、当該産業のパイが拡大するだろう。

2. PDS

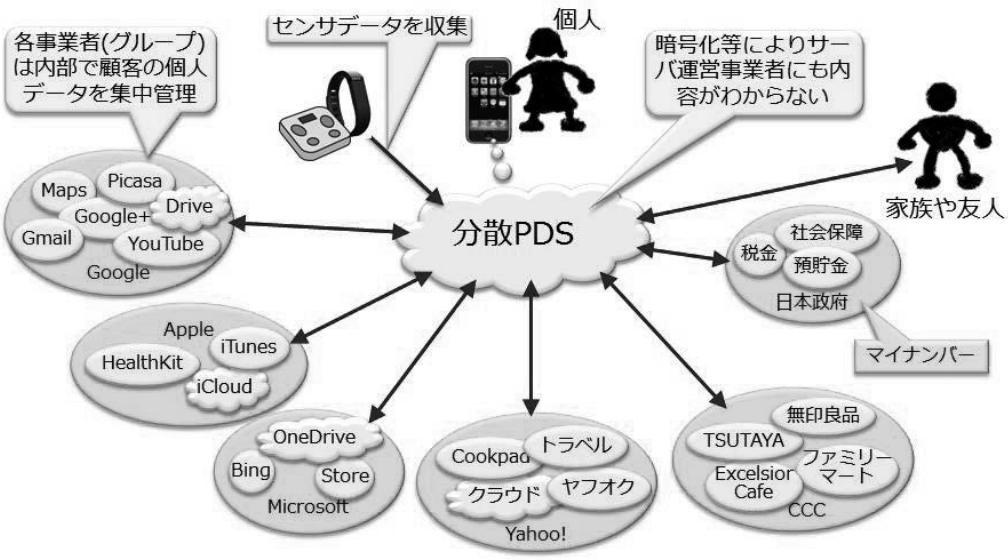
個人が本人のデータを電子的に蓄積・保管して他者と共有し活用できるようにする仕組みを **PDS** (personal data store、personal digital store、personal data service、personal data vault など) [Bell 01]と言う。いわゆるデータブローカは、個人データを本人のために活用する機能を持たないので、本稿では PDS から除外して考える。

(1) 集中 PDS

PDS にも個人データを集中管理するものと分散管理するものとがある。**集中 PDS** (centralized PDS)は、個人データを本人の役に立てるだけでなく、多数の個人のデータに匿名化等の処理を施した上で事業者に提供するデータブローカの機能も備える場合が多い。ヘルスケアに関する集中 PDS として、Google Health や MS HealthVault や日本の「どこでも MY 病院」構想の下で開発されたシステムなど、集中管理に基づく PHR (personal health record; 個人が本人の医療データを管理しヘルスケア事業者と共有して活用する仕組み)が挙げられる。これらはヘルスケアのための民間の PDS だが、デンマークの Borger [Borger]などは、ヘルスケアに限らない多様な個人データを政府が集中管理して本人の役に立てたり企業等に提供したりすることによって産業や文化の振興を図っている。米国政府が運用する Blue Button [HealthIT]と Green Button [GreenButton]はそれぞれヘルスケアと電力エネルギーに関する集中 PDS と言えるだろう。現在開発が構想されている日本の情報銀行[iBank]もデータブローカの機能を備える集中 PDS の一種と言えよう。

(2) 分散 PDS

分散 PDS (decentralized PDS または distributed PDS)は、次の図のように、集中 PDS を含む多くの集中管理型サービスを個人が自由に組み合わせて利用することを可能にする。この図は、Google や CCC や日本政府がそれぞれ ID 連携等の仕組みを用いたデータの集中管理に基づくサービスを行ない、個人がそうした複数のサービスを利用している様子を表わす。たとえば、YouTube の利用履歴と iTunes の利用履歴を統合して分析したいとか、日本政府がマイナンバーで管理する預貯金のデータと CCC が T-ID で管理する購買のデータを統合して分析したいとか言っても、Google と Apple が顧客のデータを共有するとか、日本政府と CCC が個人データを融通し合うとかいうことはあり得ないし、あってはなるまい。YouTube と iTunes の利用履歴を統合するとか預貯金のデータと購買のデータを統合するとかいうことは、個人が本人の意思に基づいて分散 PDS で行なうしかない。さらに、この統合を多数の個人のデータに拡張するにはその人々から同意を取得する必要があるが、それも分散 PDS によって容易になる。



多数のサービスを相互連携させる機能の一環として、分散 PDS はそれらのサービスに関するシングルサインオンの機能を備え、それによってパスワードの使い回しを防ぐなど、個人レベルでのセキュリティを向上させる。多くの人々のデータは分散させた方がセキュリティが高いが、1 人分のデータをあまり多数に分けると、管理者である本人や代理人の認知限界を越えてしまい、アカウントを忘れたりパスワードを使い回したりするので却って危ない。1 人分のデータはひとまとめにして管理するのが実際には最も簡単で安全と思われる。ただし、データの開示は一般には 1 人分の全データを対象とするのではなく、各々の場合の目的に応じたごく一部のデータに限ることは言うまでもない。

分散 PDS には、個人用端末同士の P2P 通信によって利用者間でのデータ共有を実現するものと、個人用端末以外にサーバ(ホームサーバや事業者が運営するサーバ)を用いてサーバ同士またはサーバと個人用端末との通信によりデータを共有するものとがある。前者の P2P 型分散 PDS としては Personal Server [Want 02]などが提案されている。しかし、その後はいまのところ実験的にでも稼働しているものはなさそうである。データ共有にサーバを用いる分散 PDS は、データ共有以外のさまざまな情報処理におけるサーバと個人端末との役割分担の観点から分類することができる。そこでサーバが主要な役割を果たす方式の分散 PDS としては、Persona [Baden 09]、VIS [Cáceres 09]、PDV [Mun 10]、PrPl [Seon 10]、openPDS [deMontjoye 14]、Respect Network [RespectNetwork]などがある。

一方、個人端末が主要な役割を果たす分散 PDS として PLR (personal life repository; 個人生活録) [橋田 13][Hasida 14]がある。PLR では、すでにコモディティになっている Google Drive や Dropbox 等のクラウドストレージサービスをそのままデータ共有用のサーバとして使い、データ共有以外の情報処理をすべてスマートフォン等の個人端末が担う。PLR は、非公開の個人データを暗号化してからクラウドストレージに送信し、クラウドストレージからデータを取得した後に手もとで復号する。その復号に必要な鍵を、クラウド

ストレージ事業者にもPLRを開発する事業者にも原則として開示せず、利用者が自ら指定した他者にのみ開示することにより、個人データの分散管理を実現する。まだコモディティになっていない類のクラウドストレージやID連携などのサーバ機能を(P2P型以外の)他の分散 PDS が必要とするのに対し、PLR 本体はスマートフォン等のアプリに過ぎずサーバとしては既存のコモディティをそのまま活用するという意味において、PLR はきわめて簡便で運用コストの低い分散PDSと言えよう。

3. 展望

個人情報漏洩のリスク管理や医療制度改革が分散PDSを普及させるきっかけになる可能性が高い。分散PDSを普及させる他のきっかけとしては、スーパーハイビジョン(SHV)放送やマイナンバーや電力小売の自由化も分散PDSの普及に貢献するだろう。たとえば、自由化された電力小売市場での適正な競争を促すには、各需要家が自分のエネルギー消費に関するデータを持ち、それに基づいて電力小売事業者を選択できるようにする必要がある。そのために米国のGreen Buttonのような集中PDSを構築するよりも、分散PDSを用いた方がはるかにコストが安く、かつデータ流通の自由度が高くなると考えられる。個人情報漏洩の防止、ヘルスケア、SHV、マイナンバー、エネルギー管理、教育等に関する分散 PDS の普及が相互に促進し合うことは言うまでもない。これらのうちいずれかの領域において分散 PDS が普及すれば、ドミノ倒しに他の領域でも分散 PDS が普及することになるだろう。そのドミノ倒しは2020年の東京オリンピックまでに起こると考えられる。

参考文献

[Baden 09] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin: Persona: an online social network with user-defined privacy. ACM SIGCOMM Computer Communication Review. Vol. 39, pp.135-146. (2009)

[Bell 01] Gordon Bell: A Personal Digital Store. Communications of the ACM, 44: 86–91 (2001)

[Borger] borger.dk: https://www.borger.dk/

[Cáceres 09] Ramon Cáceres, Landon Cox, Harold Lim, Amre Shakimov, and Alexander Varshavsky: Virtual individual servers as privacy preserving proxies for mobile devices. Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds. ACM, pp.37-42. (2009)

[deMontjoye 14] Yves-Alexandre de Montjoye, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland: openPDS: Protecting the Privacy of Metadata through SafeAnswers. PROS ONE, 9(7): e98790 doi:10.1371/journal.pone.0098790. (2014)

[GreenButton] Department of Energy: Green Button. http://energy.gov/data/green-button

[橋田 13] 橋田 浩一: 分散 PDS による個人データの自己管理. 人工知能学会誌, 28(6), 872-878. (2013)

[Hasida 14] Kôiti Hasida: Personal Life Repository as a Distributed PDS and Its Dissemination Strategy for Healthcare Services. Big Data Becomes Personal: Knowledge into Meaning, 2014 AAAI Spring Symposim Series. (2014)

[HealthIT] HealthIT.gov: Your Health Record. http://www.healthit.gov/patients-families/your-health-records

[iBank] 東京大学 空間情報科学研究センター/地球観測データ統融合連携研究機構: 情報銀行. https://ibank.csis.u-tokyo.ac.jp/ibank/index

[Mun 10] Min Mun, Shuai Hao, Nilesh Mishra, Katie Shilton, Jeff Burke, Deborah Estrin, Mark Hansen, and Ramesh Govindan: Personal data vaults: a locus of control for personal data streams. Proceedings of the 6th International Conference. ACM. (2010)

[RespectNetwork] Respect Network. https://www.respectnetwork.com/

[Seon 10] Seok-Won Seong, Jiwon Seo, Matthew Nasielski, Debangsu Sengupta, Sudheendra Hangal, Seng Keat Teh, Ruven Chu, Ben Dodson, and Monica S. Lam: PrPl: A Decentralized Social Networking Infrastructure. ACM Workshop on Mobile Cloud Computing and Services: Social Networks and Beyond. (2010)

[Want 02] Roy Want, Trevor Pering, Gunner Danneels, Muthu Kumar, Murali Sundar, and John Light: The personal server: Changing the way we think about ubiquitous computing. Ubicomp 2002, 223-230 (2002)