

# パーソナルデータと AI

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

AIの効能は、知的な業務を自動化して生産性を高めることである。業務の自動化には、業務のデータ（業務規則と業務文書）を機械が理解できて、そのデータが社会的に共有される必要がある。それには、業務データを機械理解可能な形で表現するためのオントロジーと、データを共有するための社会的な基盤が必要である。特に、B2Cサービスの価値をAIによって高めるには、オントロジーにより構造化したパーソナルデータを安全・公正に流通させなければならない。以下では、このような観点から、パーソナルデータの管理と活用により価値の高いB2Cサービスを実現する方法を論ずる。

## 分散PDS

ヘルスケア、観光、教育・学習、就労など、さまざまな領域において、各個人へのサービスを特定の事業者が丸抱えするのは不可能である。たとえばパック旅行に参加した観光客もそのパックを提供する旅行代理店と無関係なレストランや土産物店で食事や買物をするだろう。したがって、図1のようにサービス受容者である個人のデータを複数のサービス（を担う他の個人や事業者）が共有して連携すべきことが多い。たとえば、複数の医療機関の間で診療記録が共有されれば、より安全で効果的な治療が可能になるだろう。また、食べ物の好みや宗教上の制約をレストランに開示するとそれに合わせた食事が提供されるかも知れない。学習履歴や志望に応じて学習の対象と方法についてアドバイスを受けたり、職歴や資格や家庭の事情に即して就職先を探してもらったりすることもできるだろう。

このように連携すべき複数のサービスはしばしば複数の事業者などによって提供されるので、それらのサービスを連携させるにはそれらの事業者などが相互連携する必要がある。たとえば観光においても、各々の観光客に対する宿泊、移動、飯食、購買、その他のアトラクションなどのサービスを単一の事業者がすべて丸抱えするのは一般には明らかに不可能であるから、各観光客のデータを複数の事業者が共有せねばなら

ない。ヘルスケアや教育など他のサービスの領域においても同様である。

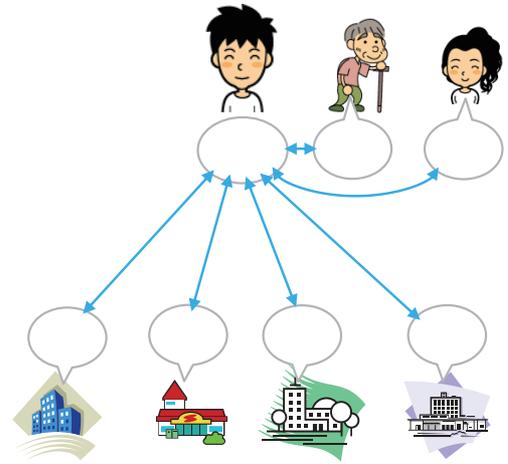


図1 パーソナルデータの共有

多くの人々のパーソナルデータを複数の事業者などが共有するため、これまでは、図2のように多人数分のパーソナルデータを集中管理する仕組みが用いられてきた。たとえば従来のEHR（Electronic Health Record; 医療データを医療機関の間で共有するサービス）ではEHR事業者が多数の医療機関にわたる多数の患者のデータを集めて管理する 경우가多く、MS HealthVault<sup>\*1)</sup> や「どこでもMY病院」のようなPHR（personal health record; 個人が自分の医療・健康データを集約して活用する仕組み）においても大勢の個人のデータを特定の事業者が集中管理していた。

しかし図2に示したように、この集中管理方式はサーバーを運用するためのコストがかさむだけでなく、データを集約することによってそのデータがまとめて漏洩するリスクをわざわざ生み出してしまう。また、同図の左下に示すような同種の集中管理の仕組みが他にも現われることが多いが、集中管理事業者同士はしばしば互いに競合するので直接的なデータ共有が困難である。ゆえに集中管理方式はデータ共有のための一般解になり得ない。

\*1) HealthVault は、Microsoft Corporation の商標または登録商標です。

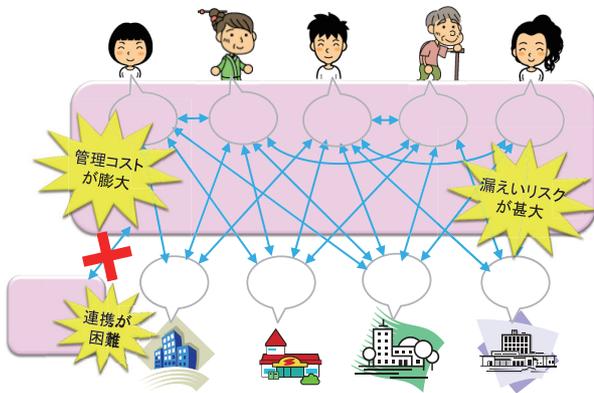


図2 集中管理によるデータ共有

一方、図3のように各個人（または代理人など）が本人のデータを（後述のPLRのようなツールを用いて）管理してそのデータのみについて必要十分な共有を行えば、情報漏洩リスクが極小であり、後述のようにコストも低い。しかも、各事業者と顧客とのデータ共有は事業者同士のデータ共有よりもはるかに容易だから、競合する事業者同士でも顧客を介してデータを共有し間接的に連携することができ、こうして一般的なサービス連携が容易になる。

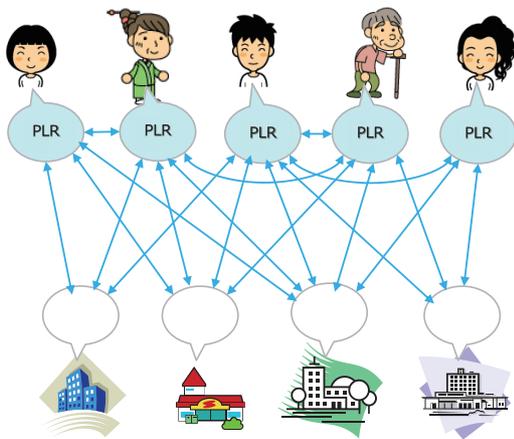


図3 分散管理によるデータ共有

さらに、図4に示す通り、パーソナルデータを本人が管理していれば、そのデータを本人同意に基づいて収集することも簡単である。パーソナルデータを本人が管理していれば、データを収集・分析する者は、分析しないかも知れないデータを収集・保管する必要がなく、必要が生じたときにデータを収集して分析が終わったら結果を残してデータを消去することにより管理コストと漏洩リスクを最小化することができる。たとえば疫学調査や治験でも10万人以上のデータを分析することはまれであり、たいていは1,000人程度以下のデータで

\*2) Google ドライブは、Google Inc. の商標または登録商標です。 \*3) Dropbox は、Dropbox, Inc. の商標または登録商標です。

足りるから、10万人のデータを常時集約して保管しておくのは無駄であり、コストとリスクを徒に高めてしまう。

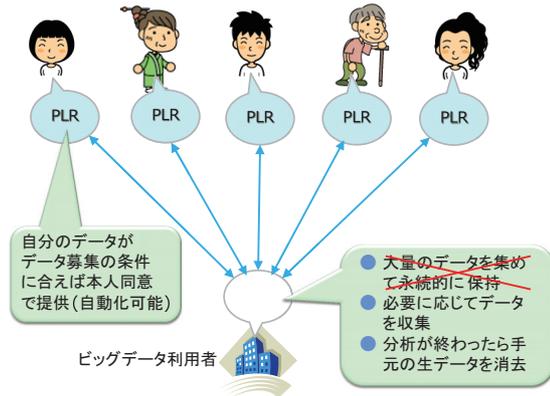


図4 分散管理に基づくビッグデータの活用

以上のように各個人（代理人）が事業者に依存せずに本人のデータを管理し自由に他者と共有して活用するための仕組みを分散PDS (Decentralized Personal Data Store) と呼ぶ。PLR (Personal Life Repository)<sup>1), 2), 3), 4), 5), 6)</sup> は分散PDSの一種であり、図5に示すように、個人が本人のデータを個人端末のアプリで管理しクラウド経由で他者（他の利用者や他の端末）とのデータ共有を可能にする。専用のクラウドは不要であり、Googleドライブ<sup>\*2)</sup> やDropbox<sup>\*3)</sup> などの基本無料のパブリッククラウドストレージをそのまま使えるので、PLRの運用コストは利用者数に依存せず、アプリの保守費用だけである。またPLRでは、単に多要素認証を用いるのみならず、クラウドにおいても端末においても暗号化などによってパーソナルデータを秘匿し、かつPLRのアプリは平文のデータをファイルに書き出したり外部に送信したりする機能がないので、利用者が間違ったり騙されたりしてもまとまった量のデータが一挙に漏れることはあり得ず、たとえばパスワードなどが洩れたとしてもアプリが偽造されなければ大量の平文データが盗まれることはない。

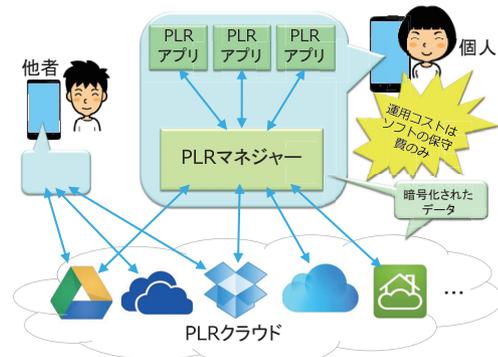


図5 PLRの仕組み

## 自律分散協調サービス

各個人が本人のデータをPLRで管理し多様な事業者などと共有することによってサービスの質が高まるといふ、いわば自律分散協調サービスの概要を、観光を想定して図6に示す。たとえば、食べ物や行き先の制約や好み(ハラルでないとかだめとか高所恐怖症だとか歴史が好きか音楽が好きかなど)や行動履歴(いつどこに行ったとかどんな経路で移動したとかいつどのホテルに泊まってどのレストランで何を食べたかなど)の情報をPLRで管理しておいてホテルやレストランや土産物屋や医療機関に簡単に開示すれば自分に合ったサービスを受けることができるわけである。また、持病や体調のデータがPLRに入っていて緊急時にそれを簡単に開示できる(たとえば本人の意識がなくても予め登録された医師や救命士が本人の健康データを閲覧できる)とすればリスク管理として有効だろう。

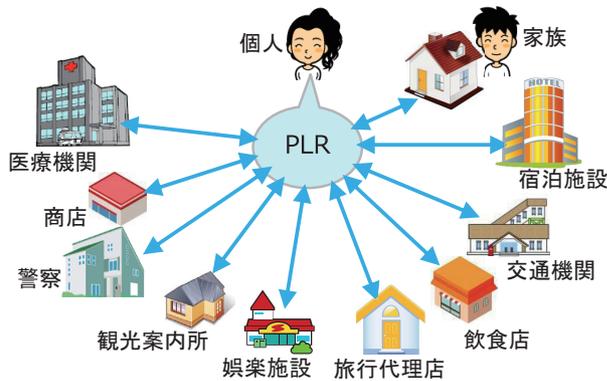


図6 自律分散協調サービス

しかし、位置や購買や健康など個人の過去の生活行動や属性だけに基づいて当該個人のニーズを予測するのは難しい。たとえば位置情報などに基づいて商品やサービスを推薦するサービスが事業として成功したという話は聞いたことがない。個人のニーズを知るには本人の意思(計画や予定)に関する情報の方がはるかに有用であることが多い<sup>7)</sup>。

このように個人のパーソナルデータを活用することにより当該個人のニーズに適合したサービスが可能になると期待されるが、そのデータを特定の事業者が管理していると、他の事業者のサービスを個人のニーズに適合させるためにパーソナルデータを用いるのが難しい。まず、ある事業者が保管している顧客のパーソナルデータを別の事業者に渡すには当該顧客の同意が必要だが、同意の取得には手間がかかる。また、そもそも事業者同士が競合関係にあれば、顧客のデータを

共有しようとはまず考えないだろう。

しかしそれでは、各個人がさまざまな事業者から常に良いサービスを受けることが難しい。事業者同士が競い合うのはサービスの質や産業の国際競争力を高める上で望ましいが、顧客のデータは競合する事業者の間で共有すべきである。ところが顧客を奪い合う事業者同士が顧客のデータを直接共有するのは難しいから、各個人がいつでもどこでも自分に合った良いサービスを受けるには、本人が分散PDSで本人のデータを管理して任意の事業者などに開示し、それにより競合する事業者同士でも図5のように顧客を介して間接的にデータ共有できるようにすべきである。そうすれば、多くの事業者のサービスの質を高め、産業全体を振興することができる。

顧客データの共有には、事業者同士の直接連携よりも、事業者と顧客との直接的連携に基づく事業者同士の間接的連携の方が有効である。たとえ各事業者の規模が小さくても、顧客を介して情報共有するオープンシステムを構成することにより、多くの事業者が全体として多数の顧客に良質のサービスを提供できるだろう。すなわち、図6に示した自律分散協調サービスにおいては、個人だけでなくサービス事業者もまた自律分散協調的に働き、異業種にわたる複数の事業者が間接的に協調して一貫したサービスを提供するわけである。

たとえば「おもてなしプラットフォーム<sup>8)</sup>」はインバウンドの旅行者のパーソナルデータを複数の事業者が共有することにより異種のサービスや複数地域のサービスを連携させる仕組みの構築を目指すプロジェクトであるが、これも分散PDSとの連携によるオープン化と拡張が必要だろう。旅行者へのサービスの価値をさらに高め、観光産業の一層の振興を図るには、関東、関西、九州で運用する3つのシステムの各々にさらに多くの事業者を参画させるとともに、これら3つのシステムを相互に連携させる必要があろうが、すでに参画している事業者と競合する事業者の参画は困難であり、また遠く離れた地域の間での連携に大きなコストをかけるのも難しい。しかし、各システムを旅行者の分散PDSと連携させれば、各システムの拡張とシステム間の相互連携が安価かつ安全に実現できる。

## メディアーター

どの事業者からいかなるサービスを受けるべきかが自明でない場合は、各個人が分散PDSで管理している

自分のパーソナルデータをさまざまなサービスの情報と照合（マッチング; matchmaking）することによって自分に合ったサービスを特定できることが望ましい。だが、サービスの種類が多い場合には、それらが各個人に向いているかどうかを判断するのに必要なサービスのデータが膨大であるため、個人端末でマッチングを実行するのは非現実的である。したがって、何らかの集中管理の仕組みにより多数の事業者からサービスの情報を集め多数の個人からパーソナルデータを取得してそのようなマッチングを行い、各個人に合ったサービスを推薦する必要がある。

この仕組みをメディエーター（mediator）と呼ぼう。医療、介護、観光、飲食、教育などのサービスの領域や地域ごとにメディエーターがあり得るが、各サービス利用者に最大の利益をもたらすようなマッチングを行い、かつ利用者から預かったパーソナルデータの利用を本人同意の範囲に限る、という意味での公正性がメディエーターには要請されるので、図7のように、各領域や各地域において複数の競合するメディエーターが存在し、公正でないメディエーターは淘汰されるべきである。

多くの場合、上記のマッチングは数万人以上の個人と数万件以上のサービスを対象とするので、AIによって自動化する必要がある。すると、メディエーターもPLRを用いれば、メディエーターのサーバーに保存されるデータは暗号化（または他の手段によって秘匿化）されるので、マッチングの計算に用いる平文のデータはインメモリのみが存在し、マッチングの結果を見る

人間はマッチングサービスの利用者本人だけ、という運用が可能である。さらに、サーバーのOSがサーバー内のPLRアプリを認証して不正なアプリを排除すれば、不正なOSをインストールしない限りパーソナルデータを盗み出すことはできない。また、サーバーマシンへの不正なOSのインストールを防ぐことに絞り込んだ管理体制の構築と運用は容易かつ安価であり、小規模な観光地などでも実現可能だろう。

メディエーター事業の収益性は非常に高いと期待される。たとえば観光メディエーターは観光市場における間接業務を一手に引き受けることにより、観光関連事業者からの登録料、旅行者と事業者との個別の取引の仲介手数料、ビッグデータ分析の結果（御社のこのサービスはここをこう変えればもっと使われそうですよ、など）の提供料、その他個人へのサービスの料金などを得ることになる。この業務を自動化し上記のように安価で安全な管理体制を採用すれば、メディエーターの利益率はかなり良くなりそうである。また、そのような事業であるならば、メディエーターには前記のような公正性がなおさら強く求められる。日本政府が構想している「情報銀行」<sup>9)</sup>は以上のような要件を満たすメディエーターとして設計すべきだろう。

### おわりに: ソーシャルサイエンス

パーソナルデータの分散管理とそれに基づくサービスについて、AIとの関連において論じた。世界中でAIの研究開発競争が巻き起こっているが、その研究開発

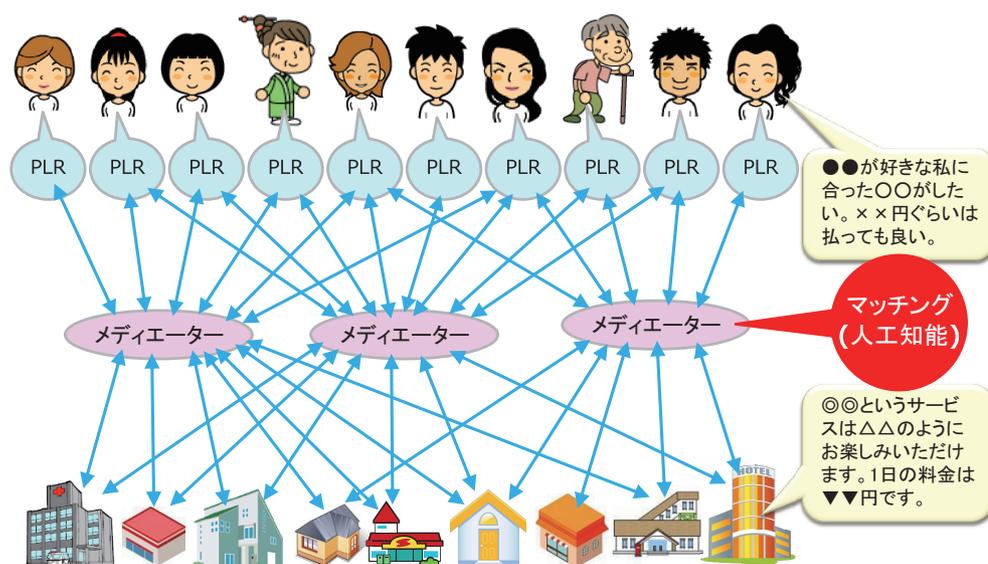


図7 メディエーター

のコストのほとんどがデータの整備にかかっている。したがって、AIにおける最重要課題は、機械学習のアルゴリズムの開発などではなく、データ整備の効率向上と規模拡大と持続可能性の担保であり、分散PDSやメディエーターはそのために必須である。

データ整備というのはデータの流通と構造化、つまり、標準的な仕様に基づいて構造化された良質のデータが潤沢かつ円滑に流通し活用されるための社会基盤を整備することである。そのような基盤が整った社会は、AIの研究開発のためのフィールドになるだけでなく、その成果を用いた事業の市場にもなり、それによって社会の共有価値が飛躍的に高まるだろう。それは、AIの研究成果を社会に還元するというよりもむしろ、AIと社会を融合して社会全体の知能を高めるということであろう。

社会と融合したAIは「ソーシャルサイエンス」とでも呼ぶべき、図8のようなものだろう。「ソーシャルサイエンス」というのは社会科学に関するサイエンスではなく、サイエンスの社会化である。

eサイエンスとは、知識やデータや人材や装置を情報ネットワークで共有して協働することによる科学研究のことである。多分野の研究者の間での電子的なコミュニケーションを通じて知識を共有・共創したり、実験装置を共有して遠隔地から操作したり、巨大なデータを共有してスーパーコンピュータで分析やシミュレーションを行ったりするというグループワークによる科学・技術の研究が、ICTの進歩と普及に伴って可能になっている。その情報インフラを利用して、理論科学、実験科学、及び計算科学（computational science; 大規模なデータのコンピュータ処理に基づく科学研究）を統合しようという企てがeサイエンスだと言えよう。

これまでのサイエンスの典型的なテーマは、原子核物理学や分子生物学やナノテクノロジーなどだった。これらの研究分野では、素粒子の加速機などの非常に高度な実験・計測装置が使われ、普通の人の日常感覚による想像を絶するような実験が行なわれ、素人にはまったくわけがわからないデータが分析されている。そこで用いられる知識や人材やデータなどの研究資源は一般の人々の日常生活や産業とはほぼ隔絶した世界のものだった。ちなみに、そのような分野に関するサイエンスにおいて主に用いられてきた情報関連技術は、スーパーコンピュータやグリッドコンピューティングなどのハイパフォーマンスコンピューティング系のものである。

ソーシャルサイエンスとは、情報ネットワークによって共有される知識やデータや人材や装置が科学研究にも日常生活や産業にも普通に使われる、という意味で、社会と密着し、社会全体に拡張されたサイエンスのことである。その研究の対象は、主として社会のさまざまな機能（広義のサービス）であり、それは政治学、経済学、経営学、社会心理学などを含み、情報技術や認知科学との関係が深い、それだけに限定されるものではなく、たとえば疫学や気象学や生態学にも及ぶだろう。

サービス科学が科学的根拠に基づいてサービスの生産性を向上させる方法論だとすると、ソーシャルサイエンスは科学的根拠に基づいて社会の機能を向上させる方法論だという意味でその一般化に相当する。サービス科学では、少なくともこれまでのところ、具体的個別的なサービスを主たる研究の対象とすることが多かったように思われるが、ソーシャルサイエンスにおいてサービスを扱う場合には、複数のサービスからなる社会システムが主要な対象である。

ソーシャルサイエンスはサービス科学を包含するメタレベルの学術領域をなすと考えられる。サービスの限定的な定義の下で経営学や商学などの関連領域とは別にサービス科学に独自の理論体系を構築することは非常に難しそうだが、サービス（つまり生活や業務における社会的相互作用）の現場で得られる客観的データを用いて仮説検証を行なうという方法がサービス科学において重視されていることは、一種のサイエンスとしてのサービス科学の定式化を示唆する。そこで、狭義のサービスではなく社会の機能全般を研究対象と考えれば、客観的データに基づく持続的な仮説検証による知識創造を情報技術の支援により社会全体に拡張したもとのソーシャルサイエンスに到達する。

ソーシャルサイエンスは、日本政府の言う「超スマート社会（Society 5.0）」<sup>10</sup>を包含すると考えられる。また、オープンサイエンスやシチズンサイエンスは市民が参画する科学ということだが、ソーシャルサイエンスは科学研究への市民の参画にとどまらず、市民の日常の生活や業務が全体として科学になっているということである。ここで科学というのはサイクル（仮説検証サイクル）であり、そのサイクルによって創造・維持・向上すべきは社会的に共有される価値である。また、ソーシャルサイエンスの「ソーシャル」は社会全体がサイエンスを実践する主体であるとともにその実践の場でもある。





図8 ソーシャルeサイエンス

## 参考文献

- 1) Koiti Hasida: Personal Life Repository: Distributed PDS for Data-Driven Improvement of Your Welfare, AAAI Spring Symposium 2013 (Data Driven Wellness), Stanford University (2013)
- 2) 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)、 pp. 872-878 (2013)
- 3) 橋田 浩一: 分散PDSと集めないビッグデータ、人工知能学会誌、29(6)、 pp. 614-621 (2014)
- 4) 青木 孝裕 et al.: 個人情報をも本人が管理するPDSシステムモデル・「集めないビッグデータコンソーシアム」における検討報告、マルチメディア、分散、協調とモバイル(DICOMO2015)シンポジウム、 pp. 249-255 (2015)
- 5) 集めないビッグデータコンソーシアム: 平成27年度集めないビッグデータコンソーシアム成果報告書 ～パーソナルデータエコシステムの実現～(2015) [http://www.ducr.u-tokyo.ac.jp/jp/materials/pdf/research/dbd-conso\\_seika.pdf](http://www.ducr.u-tokyo.ac.jp/jp/materials/pdf/research/dbd-conso_seika.pdf) (2016年10月13日)
- 6) 産業競争力懇談会: 2015年度 プロジェクト 最終報告【IoT時代におけるプライバシーとイノベーションの両立】(2016) <http://www.cocn.jp/thema84-L.pdf> (2016年10月13日)
- 7) Doc Searls: The Intention Economy: When Customers Take Charge (2012) Harvard Business Review Press (邦訳 栗原 潔: インテンション・エコノミー(2013) 翔泳社)
- 8) 経済産業省商務情報政策局: おもてなしプラットフォーム — インバウンド消費の拡大に向けた事業者間連携 — (2016) [http://activeictjapan.com/pdf/20160414/jimin\\_it-toku\\_document1\\_20160413.pdf](http://activeictjapan.com/pdf/20160414/jimin_it-toku_document1_20160413.pdf) (2016年10月13日)
- 9) NHK NEWS WEB: 個人情報管理「情報銀行」 年度内に企業向け指針 (2016) <http://www3.nhk.or.jp/news/html/20160916/k10010689491000.html> (2016年10月

13日)

10) 科学技術基本計画. <http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf> (2016年10月13日)

## 筆者紹介

橋田浩一: Koiti Hasida. 1981年東京大学理学部情報科学科卒業。1986年同大学院理学系研究科博士課程修了。理学博士。1986年電子技術総合研究所入所。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、AI、認知科学。現在の主な研究テーマは個人データの分散管理と意味的構造化、及びそれに基づくソーシャルサイエンス。