

データから価値を連鎖的に生み出す トラステッド・データスペース

「スマートな世界」では、さまざまな組織が企業や業種の壁を超えてデータを相互に持ち寄って分析し、新たな目的を持つデータを連鎖的に生み出すことで、社会全体でデータの価値を最大化できる可能性を持っています。現実には、組織間のデータ利活用は「限定した相手」に対して「限定的なデータ」を提供するにとどまり、価値の連鎖には至っていません。本稿では、この問題の解決に向けた取り組みとして、「トラステッド・データスペース」というデータ流通の新しい仕組みと、これを実現する要素技術を紹介합니다。

わしお 鷲尾	ともあき 知曉 ^{†1}	いとう 伊藤	ひろき 宏樹 ^{†1}
みたに 神谷	こうき 弘樹 ^{†1}	もろはし 諸橋	げんぶ 玄武 ^{†1}
うまこし 馬越	けんじ 健治 ^{†1}	おくだ 奥田	てつや 哲矢 ^{†1}
たかや 高屋	かずゆき 和幸 ^{†2}	おおむら 大村	けい 圭 ^{†2}
たかはし 高橋	げん 元 ^{†1}		

NTT社会情報研究所^{†1}

NTTソフトウェアイノベーションセンター^{†2}

業種横断のデータ流通が実現するスマートな世界

近年、製造ラインや化学プラントといった実世界のシステムをサイバー空間上で再現し、システムの動作を分析、予測するのみならず、組織や業種・業界を超えたデータ共有、データ分析を実現する技術の研究開発がさかんに行われています。

例えば、スマートシティでは、フィジカル空間（物理空間）に設置された膨大な数のセンサや動画像等をデータ化し、サイバー空間で分野横断的にヒト、モノの動きの分析を行い、得られた分析結果を基にフィジカル空間である都市におけるヒト、モノの行動を制御します。このようなフィジカル空間とサイバー空間とが高度に融合した「スマートな世界」が到来します。

スマートな世界ではさまざまな個人や企業から、世界規模でこれまでにない質と量を持つデータが生成され続けるため、これらのデータを有効に活用する仕組みが求められます。私たちは、誰もが企業や業種の壁を超えてデータを相互に持ち寄って分析し、新たな目的を持つデータを連鎖的に生み出すことで、互いの持つデータの価値を互いに見出し、社会全体で価値を最大化していくデータの市場（いちば）が必要だと考えています。この仕組みを「トラステッド・データスペース（データスペース）」と呼んでいます。

データスペースには提供者と利用者の両方が参加します。データスペース上のデータは、提供者の管理下に置かれながらも、仮想的に1つの巨大なデータレイクのように利用でき、自由にデータを検索することができます。

提供者はデータの利用条件（利用期間、実行可能な処理、二次利用条件等）を提示し、利用者はその条件に合意することで許された範囲でデータを使用することができます。このような仕組みを提供することで、ある目的で収集したデータが新たな目的にも利用可能となり、データから新たな価値を持つデータが連鎖的に生まれることが期待できます。例えば、娯楽施設、交通事業者、飲食店等が協力してユーザの1日の行動をレコメンドする目的で集めたマーケティングデータを、自治体が防災や防犯、住民の健康増進施策や街づくりに活用するといった、特定の事業者のみでは達成できないことも可能となります（図1）。

■組織横断的なデータ利活用の課題

しかし、こうした社会の期待の高ま

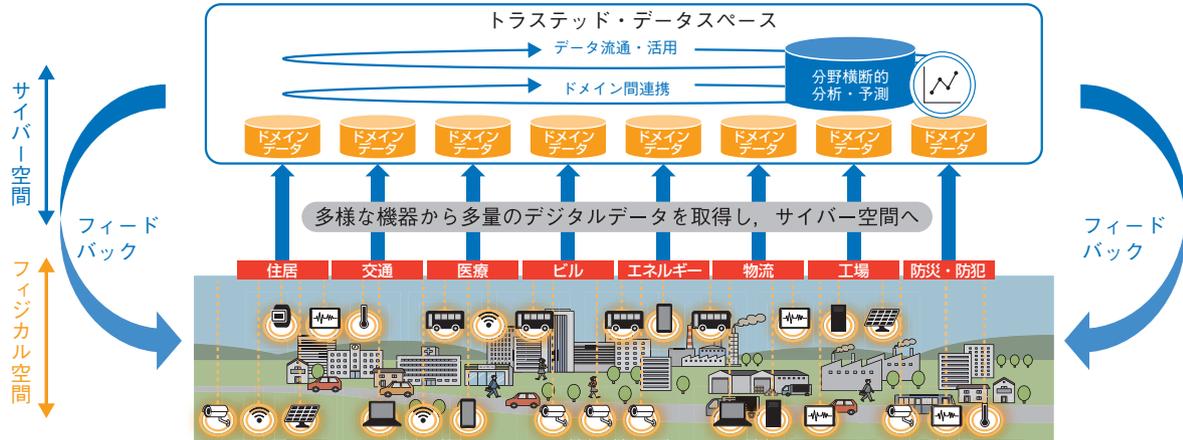


図1 スマートな世界とトラステッド・データスペース

りにもかかわらず、現時点のデータ共有・分析技術の社会実装は、局所的な価値発見にとどまっています。組織が収集したデータはそのほとんどが「集めた企業」が「そのデータを集めた目的」でしか活用されておらず、組織間のデータ利活用は「限定した相手」に対して「限定的なデータ」を提供することとどまり、価値の連鎖には至っていません。私たちは、これを「データ流通の壁」と呼んでいます。

「データ流通の壁」を克服するには、適切なデータ提供先や最適なデータの発見に関する課題、データ利用の合意を形成する際の課題、合意形成に基づいてデータ共有・利用する際の課題があり（表）、これらの課題を解決する技術が必要になります。

私たちは、これらの課題を解決する技術を提供することで、新たな価値が連鎖的に生み出されるデータ流通を実現し、ひいては社会全体でデータの価値を最大化できると考えています。

世界のデータ流通動向

欧州では、製造業を中心にデータ流通が活性化しつつあります。代表的なプロジェクトとして、欧州のデータ流通基盤をめざす Gaia-X⁽¹⁾ というプロ

表 組織横断的なデータ流通の課題

適切なデータ提供先、およびデータの発見に関する課題	・望まない相手どうしてデータの授受が行われない ・データ利用者が、自らが求める条件に適するデータを発見できる
合意を形成する際の課題	・データ提供者とデータ利用者との間で、データ処理条件の合意形成を行い、条件に基づくデータ処理を行う
合意形成に基づきデータ共有・利用する際の課題	・提供者が保有するデータが、そのデータの中身や加工方式が相手に開示されることなくデータ処理され、データ利用者に提供される

ジェクトがあります。Gaia-Xはドイツ政府とフランス政府が2019年10月29日に発表した欧州規模でのデータの共有や利活用を支援する構想で、その実現に向け2021年1月に Gaia-X AISBL という非営利団体が発足しました。法規制や契約に基づいてデータへのアクセスを制御し、データ主権*を保護しつつ、さまざまなクラウドサービスとの相互運用性を確保する技術的な仕組みを提供する基盤を構築していく予定です。また、この動きに追従して、日本では分野を超えたデータ連携をめざすプラットフォーム「DATA-EX」⁽²⁾ の構築を担うデータ社会推進協議会が2021年4月1日に発足しています。欧州の Gaia-X を用いたデータ流通の取り組みとして、サプライチェーン等の信頼関係にある企業どうして構成されたデータ流通基盤を構築するプロジェクトが複数発足しています。例えばド

イツでは、Mobility as a Service (MaaS) を実現するための「Mobility Data Space」や、自動車製造サプライチェーンのための「Catena-X」、オランダでは製造業の中小企業のための「Smart Connected Supplier Network (SCSN)」等があります。2022年に、Catena-Xは1000社規模、SCSNは3000社規模のデータ流通基盤を構築する計画です。

このように、業種を超えたデータ流通基盤の構築はすでに始まっています。しかし、現時点では Gaia-X を中心としたルール整備の途上であり、データ流通の相手は信頼関係のある企業間に限られています。また、サイバー攻撃からデータを守るためのセキュリティや、データ提供者の権利を守りな

* データ主権：データの開示範囲や利用用途などをデータ提供者の意思で決めることができる権利です。

がらデータの共有を可能とするデータ主権の保護について、どのような技術を取り入れて実装し、社会インフラとして構築していくかが今後の大きな課題となってくると考えられます(図2)。

トラステッド・データスペースの構成技術

「業種横断のデータ流通が実現するスマートな世界」で説明した課題を解決するには、以下の3つの機構を考える必要があります。

- ① 信頼できるデータやアプリ、取引相手の発見機構：データやアプリ、取引相手をカタログ化し、加えてこれらが妥当であるかの信用情報を可視化します。可視化された情報を基に、信頼できる取引相手とマッチングを行います。
- ② データ処理条件の合意形成機構：データ提供者とデータ利用者との間で、カタログ化された情報を基にデータ処理条件や開示条件、加工方式の合意形成を行います。合意した内容に従いデータ処理の実行可否を判断します。
- ③ 合意に基づくデータ処理機構：データ提供者がデータを自ら管理し、必要な場合のみ仮想的に共有、統合します。データのみならずデータの処理・加工方法までも秘匿したまま、データ利用者が必要とするさまざまな分析・加工等の処理を実行します。

現状のデータ処理は、NDA(秘密保持契約)等の契約や運用管理策でリスクマネジメントを行っていますが、③のように技術による裏付けが必要です。さらに、取引関係のない未知の相手との取引を行い、価値を連鎖的に生み出すためには、②のデータ処理条件の合意形成機構や、最終的には①の信頼できる相手・データ・アプリの発見



図2 業種を超えたデータ流通基盤

機構が必須となります。今回は、データスペースの中心となる合意に基づいたデータ処理機構について紹介します。合意に基づいたデータ処理機構では、仮想データレイクにより組織や個人に散らばるデータを暗号化したまま集めて仮想的に統合し、データサンドボックスもしくは秘密計算という二通りの方法で、暗号化したままデータ処理します。さらに、データ処理条件の合意形成機構の先行的な取り組みについて紹介します。

■仮想データレイク

自らが持つデータを他者に共有する場合、あらかじめ定めたデータ利用条件に従い、最低限必要なデータを提供し、情報の保護・管理が行えることが求められます。仮想データレイクでは、広域に分散する管理主体の異なるデータを、管理主体のガバナンスを維持して仮想的に統合するとともに、データ利用者の要求に基づき最低限のデータのみを転送する技術を実現します。そのうえで、データが生成開始されたことを速やかにデータ利用者に伝え、データ生成が完了する前からデータ利用者が一部のデータを利用開始できる仕組みを実現します。これにより、大容量データであっても早期かつタイムリーに利用でき、データ流通を促進できます。例えば、企業でデータが生成されると同時に、データ管理情報(データカタログや制御ポリシー等のメタデータ)も生成・更新されてプラットフォー

ムに通知されます。プラットフォームはそのデータ管理情報に基づき、許可された利用者に対してデータの存在を知らせ必要分のデータを送信し、当該データの流通を適切に制御します。利用者は、仮想統合されたデータ一覧を参照することで所望のデータに効率良く辿り着くことができ、かつ早期にデータを利用開始できます。これらの仕組みにより、組織を超えた安全・便利なデータの活用を可能にします。

■データサンドボックス技術

データサンドボックス技術は、データを持っているが分析技術を持っておらずデータを他社には共有したくない組織と、分析技術を持っているが他社には共有したくない組織を結び付け、データおよび分析技術を相互に共有することなく分析結果を得るための技術です。

データ所有者と分析技術所有者がそれぞれデータサンドボックスの利用に合意すると、合意した内容に基づくデータと分析技術を含む専用のデータサンドボックスが作成されます。データサンドボックスは外部との通信が断絶されており、また、内部の通信・ストレージ・メモリはデータサンドボックス提供者も復号できないように暗号化されています。これにより、データ所有者、分析技術所有者、データサンドボックス提供者の誰もが、復号されたデータ・分析技術にアクセスすることなくデータを分析可能です。

分析結果はデータ所有者と分析技術所有者の合意に基づいて、データ所有者および・または分析技術所有者が閲覧可能な場所に保存されます。分析結果をデータ所有者が閲覧可能な場合、データサンドボックスの利用によってデータ所有者は「データを他社に共有することなく」解析結果を入手可能となり、また分析技術所有者は「アルゴリズムを他社に共有することなく」他社に自社のデータ分析技術を利用してもらうことによる対価を得られます。分析結果を分析技術所有者が閲覧可能な場合、データサンドボックスの利用によってデータ所有者は「データを他社に共有することなく」他社の解析に自社のデータを利用してもらうことによる対価を得られ、また分析技術所有者は「分析技術を他社に共有することなく」自社が持たない他社のデータを利用した分析結果を入手可能となります。

■秘密計算

秘密計算技術は「データを暗号化したまま一度も元データに戻さずに処理を行う」ことを可能にする高度な暗号技術です。2019年には、AI（人工知能）で利用されるデータに対するセキュリティ・プライバシーへの対策として、秘密計算ディープラーニング、つまりデータを暗号化したまま一度も元データに戻さずにディープラーニングの学習や予測をする技術を開発しました。従来の方法では性能面が課題となるため、（暗号化しない）通常のディープラーニングの学習・予測手法よりも単純な処理で代替していました。NTTの秘密計算ディープラーニングでは、世界トップクラスの秘密計算処理性能を活かし、ディープラーニングで行われる標準的な最適化処理を利用した学習処理を、世界で初めて秘密計算処理によって完全に再現しています。

つまり、ディープラーニングでのデータ活用に必要な、①データ提供、②データの保管、③学習処理、④予測処理、のすべてのステップが暗号化した状態で実施可能となるわけです。常にデータは暗号化されたままであり一度も元データに戻すことがないため、従来よりもユーザや組織が安心してデータを提供でき、学習に利用できるデータ量や種類が増えることにつながります。このデータの拡充こそが、より精度が高く高度な分析を可能とするAIを実現します。

■合意形成機構

データ利用においては、提供者がデータ利用を認める条件と利用者がデータに求める要求のそれぞれが存在し、提供者、利用者によって条件や要求の内容は異なります。提供者は、利用を許可する相手、利用目的、利用範囲、利用期限等を条件として指定することが考えられます。一方、利用者は、対象のデータ、利用目的、行いたい処理等を要求として指定することが考えられます。

両方が納得するかたちでデータ利用を行うためには、提供者と利用者間でデータ利用に対する条件と要求とを突き合わせ、合意を形成する仕組みが必要になります。

通常は提供者、利用者がそれぞれ提供条件、利用条件をポリシーというかたちで表現し、それらの申請内容を人手で確認し、合意を形成しますが、将来的には利用者から利用要求が行われたタイミングで要求内容とポリシーを突き合わせて合意の可否を自動的に判断するやり方も考えられます。さらに、要求と条件との間に差異がある場合に、提供者と利用者間で要求および条件を動的に調整することができれば、より柔軟な合意形成が可能になります。企業や組織間でのデータ利活用

を促進するためにもこのようなアプローチについても今後検討が必要と考えています。

今後に向けて

トラステッド・データスペースによりこれまで困難だった企業や業種の壁を超えたデータの共有が加速され、新たな価値が連鎖的に生み出されるデータ流通が可能になると考えています。その実現に向け、私たちは、要素技術の研究開発だけでなく、パートナーの皆様との技術検証を加速していきます。

■参考文献

- (1) <https://www.gaia-x.eu/>
- (2) <https://data-society-alliance.org/about/vision-mission/>



（上段左から）鷺尾 知暁 / 伊藤 宏樹 / 神谷 弘樹

（中段左から）諸橋 玄武 / 馬越 健治 / 奥田 哲矢

（下段左から）高屋 和幸 / 大村 圭 / 高橋 元

私たちは、本稿で紹介したトラステッド・データスペースの研究開発を通じてデータから価値を連鎖的に生み出すスマートな世界の実現に貢献していきます。

◆問い合わせ先

NTT 社会情報研究所
企画担当
E-mail solab@hco.ntt.co.jp