5 データサンドボックス技術

企業や業種の壁を超えて 安全にデータを持ち寄り利活用可能に

業種・業界を越えたデータ利活用の取り組みが活発化するなか、NTT 社会情報研究所（以下、社会研）は企業が安心してデータを持ち寄れる「トラステッド・データスペース」の実現に向けた研究開発を進めている。本稿では要素技術の１つである「データサンドボックス技術」、および世界各国で導入が進むデータスペースとの相互接続に関する取り組みを紹介する。

データとアルゴリズムを秘匿したまま共有し利活用

企業間でデータを共有・流通し新たな価値を創出することへの期待が高い。欧州を中心に、データを持ち寄る場として「データスペース」を構築する動きも進んでいる。実際にデータを共有する際にはプライバシーに関わる情報やアルゴリズムを含めた機密情報を秘匿することが重要になるため、データを暗号化した状態でデータ処理を行う「コンフィデンシャル・コンピューティング」にも関心が集まっている。それを可能にするTEE（Trusted Execution Environment）と呼ばれる機能を実装したCPUも登場している。

こうしたことを背景とする社会研の研究開発について、神谷氏は次のように述べている。

「もう１つ重要な観点が『データ主権』です。データの提供者が開示範囲や用途をしっかりコントロールできる必要があります。この観点からの対応も含め、企業間データ流通の実現に向け研究開発しているのがデータサンドボックス技術です。」

データサンドボックス技術

データサンドボックス技術の概要を図１に示す。データ提供者、プラットフォーム事業者、アルゴリズム提供者という複数のステークホルダーを想定するコンフィデンシャル・コンピューティングは世界的にも珍しい。前述のTEEを応用してメモリ暗号化によりデータ／アルゴリズムを保護しながらデータ処理を実行することにより、実行環境を提供するプラットフォーム事業者も含め、お互いが入力データやデータ処理の結果を知ることができない仕組みを実現する。

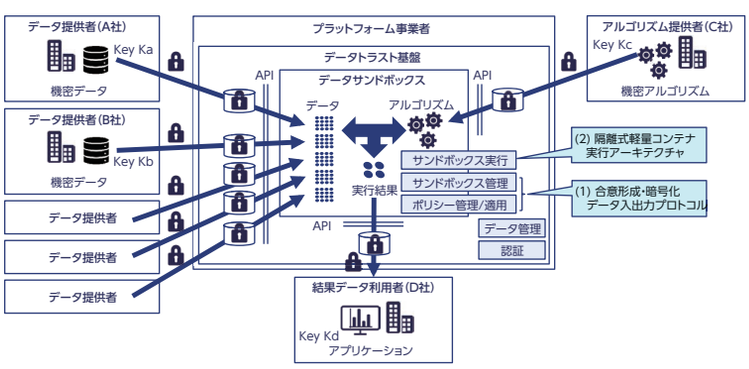
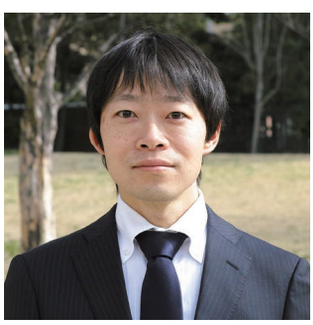


図１　データサンドボックス技術


NTT 社会情報研究所
社会情報流通研究プロジェクト
主任研究員　**神谷 弘樹**氏

データ／アルゴリズムを暗号化した状態で入出力する機能、ステークホルダー間での合意形成および合意に基づき適切なデータ管理を行う機能の実装が進んでいる。データやアルゴリズムを秘匿したまま処理を行うものの、そのことによる処理速度の低下はほとんどないという。

データサンドボックス技術のユースケース例

エンジニアリングチェーン

複数の部品メーカーが部品の

CADデータを、また航空機メーカーがそれらの部品で組み上げる航空機の空力シミュレーションアルゴリズムをデータサンドボックスへ入力する。部品メーカーが許可したアルゴリズムでのみCADデータによるシミュレーションを実行するよう合意を形成できるため、各社のノウハウがお互いに漏洩する心配がない。

秘匿データ翻訳サービス

機密情報を含むドキュメントを安全に翻訳するクラウドサービスの実現にデータサンドボックス技術を活用できる。翻訳サービスの提供事業者にも、翻訳アルゴリズムが盗まれる心配がないというメリットがある。

渋滞予測サービス

渋滞予測を行うには予測アルゴリズムに加え都市モデル、地図データ、交通情報などのさまざまなデータが必要になる。データサンドボックス技術を使えば機密情報を公開することなく、幅広いデータを利用することが可能になる。

グローバルなデータ相互流通にも対応するデータ流通プラットフォームの開発

持続可能な社会の実現に向けたグローバルなデータ流通のニーズが高まっている。たとえばカーボンニュートラルの達成を目指す取り組みではCO_2の排出量データをグローバルバリューチェーン全体で可視化することが重要だ。そのためには排出量データが改ざんされることなく、信頼性が保証される仕組みが必要となる。

欧州では“Gaia-X”というデータ流通構想に基づいた取り組みが活発化しており、ドイツの自動車メーカーなどが主導するデータスペース“Catena-X”もその1例だ。今後、欧州の自動車関連企業と取引する日本企業にもCatena-Xによるデータ流通への参加が求められることが予想される。しかしCatena-Xでは欧州のポリシーでデータが管理されるため、日本のポリシーによりデータを保護することは難しい。また、米国ではITメガプラットフォーマーが企業間データ連携を推進し、中国ではそれに加え政府が規制強化やアジア展開を推進するなど、国・地域のレベルでデータ活用による産業基盤構築の取り組みが始まっている。

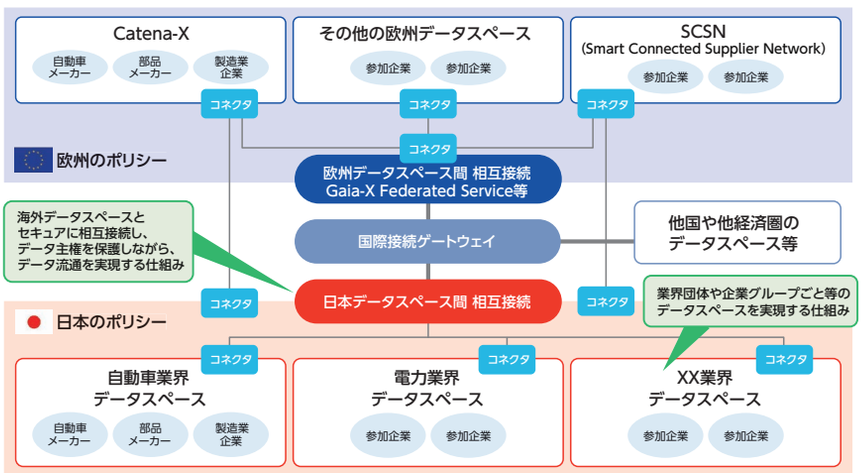


図2　データ主権を保護できるデータ流通プラットフォームの全体像

このような背景から、2022年4月、NTTコミュニケーションズ株式会社と株式会社NTTデータは共同でデータ流通プラットフォーム（図2）を開発すると発表した。日本のポリシーでデータを保護できることに加え、必要に応じてCatena-Xのようなデータスペースと相互接続できるようにする方針だ。

「Catena-XではGaia-Xとともに提唱されているIDSという技術に基づきデータ交換が行われます。このIDSを用いてスイス、ドイツ、日本の間でデータを相互流通させる実証実験を実施したほか、業界を横断する国際的なデータ流通テストベッドを構築するなど、さまざまな取り組みを進めています。」（神谷氏）

これらの取り組みにおいてデータの公開範囲や用途に関する合意形成や制御を行う機能の実現には、データサンドボックスで実装したものと同じ技術を用いている。今後はデータサンドボックス技術のもう1つの特徴であるデータやアルゴリズムを秘匿したまま処理を行う機能も活用していく方針だ。

企業間のデータ流通促進に貢献していく

「企業間のデータ流通を促進したいという思いでデータサンドボックスの技術的な課題解決に取り組んでいます。同時にNTTグループの事業会社と協力しユースケースの発掘も進めています。各国の取り組みを参考にしつつ、国境をまたいだデータ流通が可能なデータスペースの構築を目指して、いろいろな企業の方と一緒に対応を考えていきたいと思っています。」（神谷氏）