IOWN時代のデータガバナンスを実現するトラステッド・データスペース技術

IOWN（Innovative Optical and Wireless Network）が普及し，すべての情報がデジタルツインや AIによって活用される時代を支えるデータ流通基盤であるトラステッド・データスペースについて，その構成要素を説明します．また，暗号化した状態で計算処理を行うことでデータ処理に関するガバナンスを実現する技術群（サンドボックス技術，秘密計算技術，セキュアマッチング技術）を説明します．

井上（いのうえ）　知洋（ともひろ）　森田（もりた）　哲之（てつし）
NTT社会情報研究所

背　景

IoT（Internet of Things）やAI（人工知能）技術の進展に伴い，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，予測可能とするデジタルツインの構築が進んでいます．現在は，特定目的のためにつくられたデジタルツインの構築が先行していますが，近い将来にデジタルツインどうしが相互接続され，組織や業界の壁を越えたデータ共有やデータ分析が活発化すると考えられます．例えばスマートシティで期待される防災，防犯や魅力的な街づくりなどのさまざまなユースケースの実現には，公共のデータに加えて交通事業者や飲食店，誤楽施設等から得られるデータを相互に連携させて活用する必要があります．

私たちは，将来さまざまな個人や企業が生成した多種多様なデータが組織や業種の壁を越えて有効に活用される世界，誰もが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世界をめざして，「トラステッド・データスペース」の研究開発に取り組んでいます．

トラステッド・データスペース（**図1**）は，データスペース*1ごとに管理・共有されたさまざまなデータを相互に接続しデータスペースをまたいだデータ流通を実現する機能や，各々のデータ所有者によって管理されるデータを仮想的に統合する仮想セキュアデータレイク機能，データのカタログ化や信用評価等を通じてデータ所有者とデータ利用者の間のマッチングを行いデータ流通による価値創造を支援する機能，データのアクセス権に加えて利用条件に関する合意を管理しその利用条件に基づいてデータ処理の方法を制限する機能等で構成されます．本稿では，最後のデータの処理方法を制限する技術について詳細を説明します．

データ処理方法に対するガバナンス技術

トラステッド・データスペースで想定する社会では，データは一極集中された事業者が管理するのではなく，各企業，組織や個人等が自らのデータをそれぞれの管理下に保っている状態を想定しています．このためには，データは所有者が望む場所（信頼できる機器やデータセンタ等）に配置でき，他者に対しては必要な範囲で必要なときだけ共有し，関係者の間で定められた条件下でのみ活用されることが理想です．トラステッド・データスペースの仮想セキュアデータレイク機能では，データ所有者の管理下に置かれたままのデータを仮想的に1つの巨大なデータレイクのように統合しデータを検索することが可能です．データ所有者は共有するデータの取り扱い方針（例えば利用期間，複製の許可，実行可能な処理等）を提示し，データ利用者はその方針に合意することで許された範囲

＊1　データスペース：安全性の高いデータ共有の実現を目的としたシステムや仕組みを持つコミュニティ．詳細については本特集記事『IOWN 時代のデータ流通を実現するデータガバナンス』を参照．

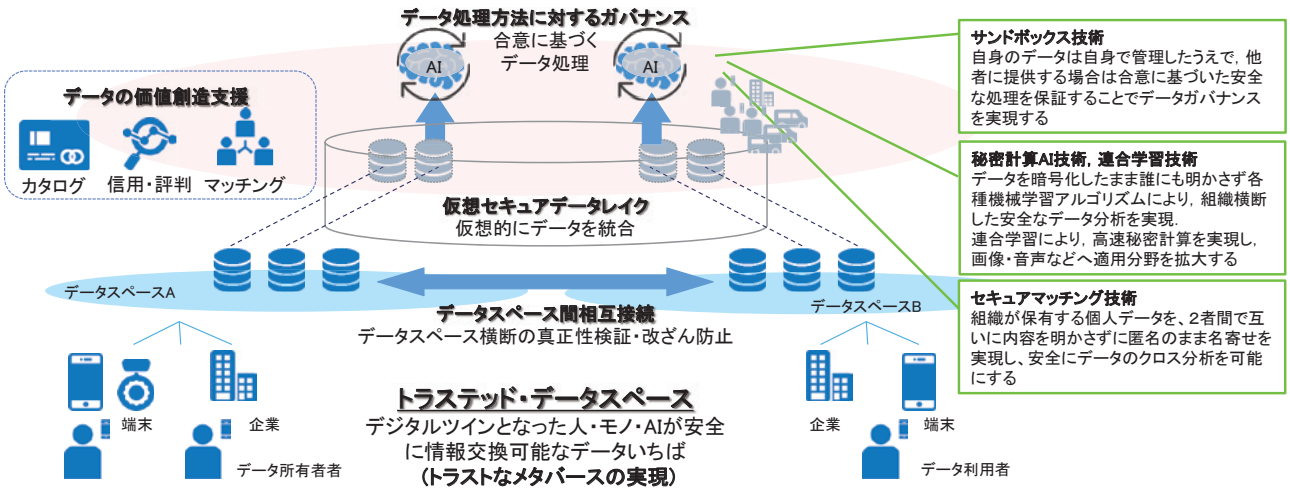

図1　トラステッド・データスペース

でデータを使用することができます．私たちはこのような仕組みを整えることによって，今までは他者に共有が難しかったような機密性の高いデータを活用しやすくし，またデータの二次的な利用を促進し，価値を連鎖的に生み出せる世界を実現できると考えています．

データガバナンスに関する機能のうち，暗号技術を用いてデータ処理方法に一定の制約を課す技術については，近年の発展が目覚ましい領域です．従来，暗号技術はデータの転送や保存の際に第三者から窃取を防ぐために主に用いられてきました．最近では，データの計算処理の過程にあるデータについても暗号化したまま処理可能な技術が発展し計算速度や実用性が高まってきたことにより，計算過程のデータが窃取され本来の目的外に利用されることも防止可能となってきています．さらに，個人のパーソナルデータや企業の営業秘密を用いる分析業務で，データを漏らさないだけでなく「データの中身を見ない」運用も可能になります．これにより，より安全なデータ処理はもちろんのこと，今まで他組織に開示することが難しかったデータを持ち寄った，企業や業界の枠を越えた新しい統合分析が可能になります．

以降では，私たちが取り組んでいる技術についてそれぞれ説明します．

■データサンドボックス技術

前述のとおり，データの活用による連鎖的な価値創出には組織を越えたデータ連携が必要になりますが，実際のビジネスにおいては一度他者に共有されたデータが複製されることや目的外に利用されることに対する懸念は大きく，そのような活用は大きく広がっていません．私たちはこの懸念の解消をめざして，各企業・組織が管理するノウハウ（本稿ではデータやアルゴリズムといった組織が機密にしたい具体的な情報を指します）をお互いに秘匿しつつ，それらを組み合わせた価値を活用可能とする技術（データサンドボックス技術，**図2**）を開発してきました．この技術では，最近のCPUに備えられた特別なセキュア計算領域（TEE：Trusted Execution Environment）の中で処理を実行することで，共有されたノウハウの複製や悪用を防止することが可能です．

本技術では，プラットフォーム上にデータサンドボックス（DSB）というTEEで構成された隔離処理実行環境を作成し，その中でデータ処理を行います．①データ所有者とアルゴリズム所有者がお互いのノウハウの共有に合意しプラットフォームにポリシーを登録すると，DSBが生成されます．DSBは外部との通信ができないように制限され，メモリ・ディスクは暗号化されOSや運用者も中を見ることはできません．②データ所有者およびアルゴリズム所有者は，それぞれがDSBとの間で共通鍵の生成・共有を行い，それぞれ固有の共通鍵で暗号化したデータおよびアルゴリズムをDSBに配置します．このとき各所有者は，システムの提供するアテステーションレポート*2を参照することで，DSBが事前に合意したポリシーどおりに作成されたかどうか，すなわち悪意のあるデータやアルゴリズムにすり替えられていないかどうか自ら検証可

*2　アテステーションレポート：TEEの持つ機能であり，ハードウェアのセキュリティチップの信頼性の裏付けとしてTEE内のバイナリ等の状態を証明する機能．

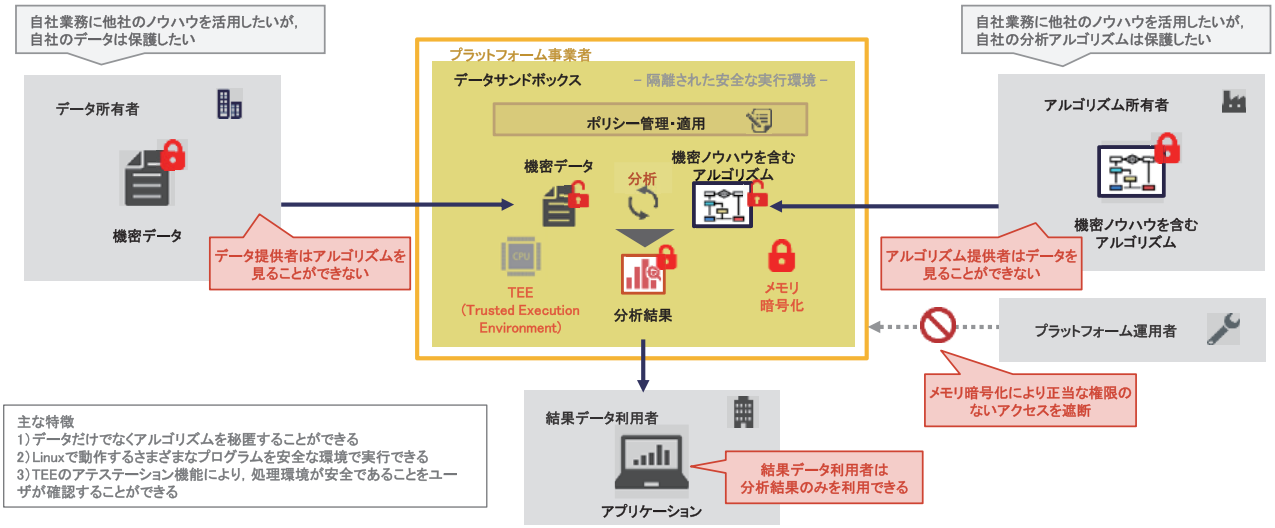

図２　データサンドボックス技術

能です．③DSBはデータ所有者およびアルゴリズム所有者との間の共通鍵を使ってデータおよびアルゴリズムを復号し，処理を実行します．④DSBは，処理結果をデータ利用者との間で作成・共有した共通鍵で暗号化して返却します．⑤処理の終了後は，データやアルゴリズムはDSBごと削除されます．このような仕組みによりデータサンドボックス技術では，入力されたデータやアルゴリズムだけでなく，処理過程や処理結果のデータに対しても誰もアクセスすることなく，データの活用が可能になります．

■秘密計算技術

秘密計算とは，データを終始一貫して，CPU内ですら暗号化したまま計算できる技術です．データの通信中・保存中の暗号化に加えて，秘密計算はさらにデータの計算過程でもデータを一度も復号することなく実行することができ，高いセキュリティを確保することができます．

NTTの秘密計算は，暗号化の仕組みとしてISO標準準拠の秘密分散を採用し，秘密分散をベースにしたマルチパーティ計算を採用しています．マルチパーティ計算とは，複数のサーバがあらかじめ定められた手順に従って暗号化データの演算と交換を行うことで，暗号化したままのデータ処理を実現するものです．これらの手順を実行している間，データは常に秘密分散のシェアと呼ばれる断片として暗号化された状態で扱われるため「データの中身を見ずに」処理が実行されます．現在，秘密計算商用クラウドサービス「析秘（セキヒ）」としてNTTコミュニケーションズから提供され利用可能となっています．研究所では，本技術を高度化し複数のAI（人工知能）の学習，推論が高速に行える技術の研究開発に取り組んでいます．

■セキュアマッチング技術

トラステッド・データスペースで想定しているデータ活用の中でも，共通の対象に対して業界や組織を横断してデータを持ち寄り分析したいというニーズは特に重要で期待されています．しかし現在，機密保持や個人情報保護の観点から，組織内のデータをそのまま他者と自由に共有し統計分析することはできません．セキュアマッチング技術は，組織が保有する個人情報・作業データを２者間で互いに内容を明かさずに匿名のまま名寄せし，安全にデータのクロス分析を可能にする技術です．前述のデータサンドボックス技術および秘密計算技術では任意のデータ処理が可能であるのに対して，本技術は２者間でのデータ交換，かつ集計処理に絞ることにより，お互いのデータ所有者がシンプルなシステムを自社に導入してデータ交換するだけで安全に統合分析ができるようになっています．

技術のポイントは２つあります．１つは，可換ハッシュ関数，準同型暗号などの高機能暗号を駆使し，暗号化したまま安全にデータを結合・集計するプロトコルです．もう１つは，集計結果からも元データが分からないようにプライバシーを保護するため，暗号化したまま高速にノイズを付与する差分プライバシー技術です．今後，NTTが開発した他のセキュアデータ流通技術と掛け合わせ，信頼できるデータ流通を実現するトラステッド・データス

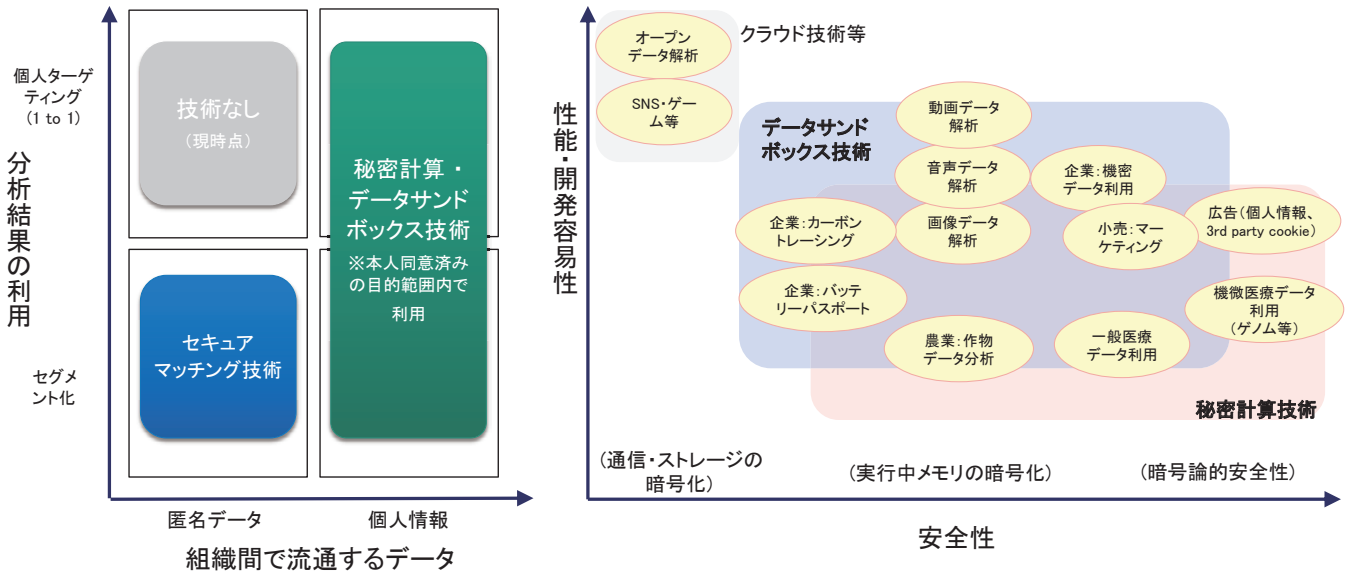

図 3　各技術の適用領域

ペースを創出していきたいと考えています．本技術は，JAL，JALカード，NTTドコモでの顧客体験価値向上と社会課題の解決に向けたデータ活用の実証実験でも用いられています．

■安全なデータ活用のための技術選択

上記の技術が対象とするユースケースには多くの重なりがあります．究極的には，理論的に絶対安全な方法で機密データの処理を高速に行い，必要な結果だけを必要とする人に渡すような技術があればすべてのユースケースに対応可能ですが，現時点ではそのような技術はなく，対象とするデータの特性や関係者および法的な規制等の要請に応じて，最適な技術を適宜選択し適用する必要があります．**図 3** は，私たちが考える技術の使い分けを示します．匿名データセキュアマッチング技術は，名寄せ可能なIDを持つデータの集合に対して，他組織のデータと組み合わせた統計分析を行いたいときに利用可能な技術です．個人情報の匿名性を保ったままマーケティング分析を行う場合などに特に有効です．秘密計算は暗号理論的安全性に基づいた処理を行うため安全性は高くなりますが，サンドボックス技術はハードウェア機能を活用するため処理速度が速く，また既存のアルゴリズムをほぼそのまま実行可能であるという特徴があります．このため，ユースケースに要請される安全性と性能のバランスの中で処理方法を選択する必要があります．

今後の展開

新たな価値が連鎖的に生み出されるデータ流通をめざすトラステッド・データスペースの実現に向けては，本稿で紹介した技術だけでなく，信頼できるデータや分析者を検索しマッチングさせる技術，データフォーマットの標準化や既存データスペースとの相互接続の推進，AIによる価値創出の技術等，さまざまな領域での進展が必要です．私たちは幅広いパートナーの皆様との協力や技術検証を進め，これらの研究開発を推進していきます．

（左から）井上　知洋／ 森田　哲之

データドリブンな社会の進展に伴い，データ利用にかかわる処理全般における信頼性の向上は今後ますます重要になります．特に暗号化した状態でのデータ活用については期待が大きく，NTTではこの分野の研究開発を加速させていきます．

◆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