高品質・高信頼なデータ流通でデータ中心社会を実現する次世代データハブ技術

NTTソフトウェアイノベーションセンタでは，IOWN（Innovative Optical and Wireless Network）構想の一環として，高頻度かつ大容量なデータトラフィックに対応し，高度なデータ保護により機密データを含むさまざまなデータを安全に流通させる「次世代データハブ」の研究開発に取り組んでいます．本稿では，データ流通における課題とそれらを解決するデータハブの概要，およびデータハブの主要機能であるデータガバナンスを構成する技術について紹介します．

持田（もちだ）　誠一郎（せいいちろう）
長田（ながた）　孝彦（たかひこ）
三原（みはら）　淳慎（あつのり）
NTTソフトウェアイノベーションセンタ

NTTがめざすデータ中心社会

近年，センシング技術の向上によるIoT（Internet of Things）デバイスの普及や5Gの普及による広帯域のコネクティビティを持ったデータソースの増加，それらのデータを人間の認知・処理能力をはるかに凌ぐ速度で処理可能とするAI（人工知能）技術の進展により，世界中で生み出されるデータ量は増加の一途をたどっており，今後もその傾向は加速しながら継続していくと予想されます．

NTTではこれらの膨大なデータが，企業などの閉ざされた組織内で利用されるだけにとどまらず，AIにより自律的に動作するシステム間で超高速に，業界や分野の枠を超えて幅広く流通され，従来は出会うことのなかったデータやノウハウを掛け合わせて新たな価値を生み出したり，社会課題を解決したりすることを可能とするデータ中心社会の実現をめざしています．

データ中心社会の実現に向けての課題

しかし，このようなデータ中心社会を実現するためには，大きく分けて次の2つの課題があります．

■現在のデータ処理アーキテクチャの限界

現在のデータ処理は目的・処理方法ごとにサイロ化された個別のシステムで行われるため，その間で大量にデータのコピーが生成されています．現状をはるかに凌ぐ量のデータを，現状よりはるかに多い主体（データを流通する人間・システム・デバイス等）間で超高速にやり取りするデータ中心社会においてはこの状況がますます加速されるため，現在のデータ処理アーキテクチャのままでは以下の問題が発生します．

- ストレージやネットワークの性能・容量を圧迫する
- 管理しなければならないデータ処理フローの数が増大し管理が困難になる
- 取り扱うデータの数・種類が増大し，生成元・変更履歴の管理が困難になる

・元データの提供者による権限やライフサイクル管理がおよばない複製・派生データが増加する

■機密情報やノウハウを他社と共有することに対する抵抗感

現在は企業などの組織を超えて機密情報やノウハウを共有する場合，機密保持契約によって共有内容の二次流通や合意した目的以外での利用を制限するのが一般的です．一方で，二次流通や目的外利用を防止するために有効な技術的な仕組みが存在せず，契約履行の強制力には限界があります．これが原因で将来的にも業界や分野の枠を超えたデータ流通が活性化しない可能性があります．

課題解決に向けた取り組み

NTTソフトウェアイノベーションセンタでは，NTTセキュアプラットフォーム研究所をはじめとしたさまざまな研究所と連携して，次のような特徴を持つ次世代データハブ（**図1**）を開発することによりこれらの課題を解決し，データ中心社会を実現することをめざしています．

■Data Omnipresence

・データ提供者がデータハブ上にフォルダもしくはキューを作成して，そこにデータを投入した瞬間から，権限のある利用者が世界中どこからでも利用可能となる．

・データ利用者はデータハブ上のデータに対し，さまざまなワークロードの処理や長期保存をデータの移動を意識することなくリーズナブルなコストで実施可能となる．

・データ利用者はデータハブ上のデータを処理するために，データハブにネットワーク接続可能なさまざまなコンピューティングリソースをデータの移動・複製を意識することなく利用可能となる．

・データ利用者はデータハブへのアクセスにメジャーなストレージ系，メッセージブローカー系の製品やサービスのAPIを利用可能となる．

■データガバナンス

・データ提供者はデータハブに投入したデータおよびその複製・派生データに対し，アクセス制御や確実な削除などの管理権限（ガバナンス）を維持し続けることが可能となる．

・データ提供者が合意した目的以外でのデータの利用を防止することが可能となる．

・データ提供者が提供したデータが誰に何の目的でどのくらい利用されたかの確認が可能となる．

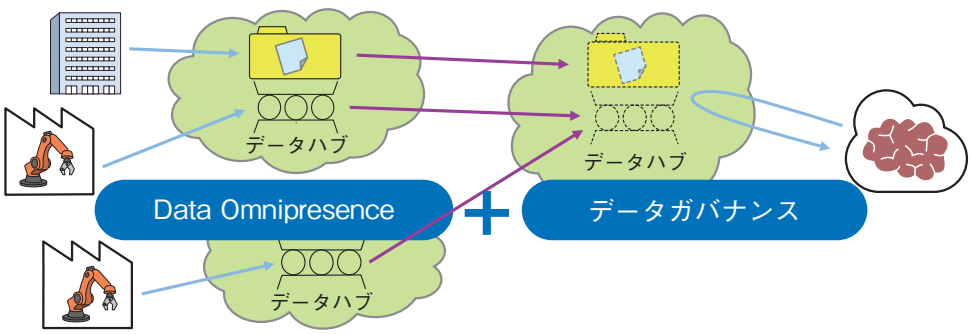

図1　次世代データハブ

次章では，現在開発中の次世代データハブが備える「Data Omnipresence」「データガバナンス」の2つの主要機能のうち，開発が先行しているデータガバナンスを構成する技術について主に説明します．

データガバナンスを構成する技術

データガバナンスを構成する主要な技術のうち，本章ではNTTソフトウェアイノベーションセンタとNTTセキュアプラットフォーム研究所が共同で開発している「データサンドボックス技術」について主に説明し，NTTセキュアプラットフォーム研究所が開発している「相互認証・鍵共有技術」「秘密計算AI技術」については概要を後半で説明します．

■「データサンドボックス技術」

企業などの組織の枠を超えてデータやノウハウ（データを価値化するための処理アルゴリズム）を流通し，それらをかけ合わせて新たな価値を得る，というデータ中心社会を実現するにあたり，データやノウハウを流通し合う当事者には，以下のような心理が働くことが想像できます．

・他人のデータやノウハウは活用したいが，自分のデータやノウハウは共有したくない．
・きちんとマネタイズできるのであれば自分のデータやノウハウを共有したいが，共有したデータやノウハウを約束した目的（どの種類・範囲のデータをどのアルゴリズムで処理するか）以外に利用したり，第三者に横流ししたりされたくない．

こうした懸念に対し，これまでは個社間で時間をかけて機密保持契約を締結し，相手を信頼するしか対策がなく，それが組織を超えたデータ流通を阻害する大きな要因になっていたと想定されます（**図2**）．

こうした懸念に対して，契約や相手に対する信頼による防御ではなく，システム的・技術的な防御手段を提供するのが，NTTソフトウェアイノベーションセンタが開発した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の動作概要は以下のとおりです．

① データ流通の当事者ではないクラウド事業者などの第三者プラットフォーム上に，データサンドボックスという隔離処理実行環境を作成します．データサンドボックスでは外部との通信が適切に制限され，メモリ・ディ

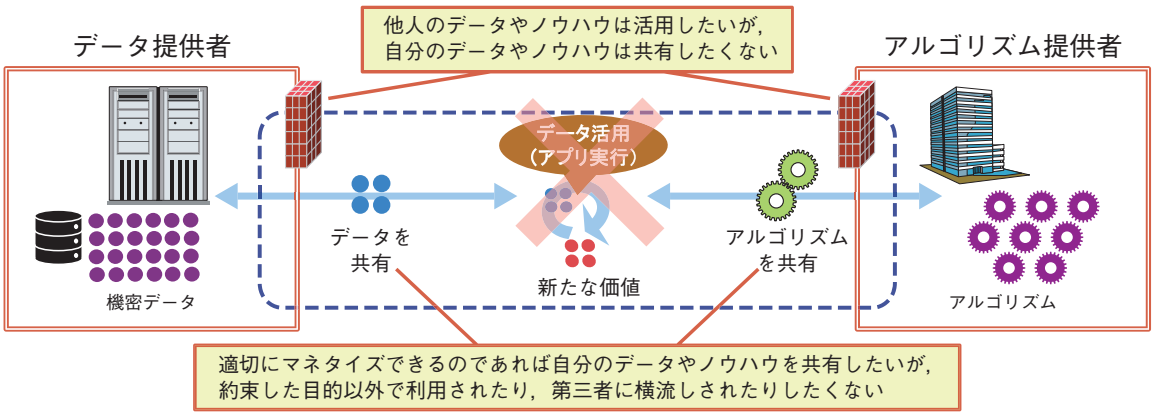

図2　企業間データ流通の課題

スクが暗号化されています．この暗号化はプラットフォーム事業者さえも解除できません（**図 3**）．

②　他社のアルゴリズムを用いて分析したいデータを有するデータ提供者，および他社に提供してマネタイズしたいアルゴリズムを有するアルゴリズム提供者は，それぞれがデータサンドボックスとの間で共通鍵の生成を行い，それらの共通鍵で暗号化したデータおよびアルゴリズムをデータサンドボックスに配置します．データ提供者とアルゴリズム提供者の共通鍵は異なるため，互いのデータやアルゴリズムを見ることはできません．また，データサンドボックスは外部との通信が制限されているため，データ提供者およびアルゴリズム提供者はデータおよびアルゴリズムを配置できるだけで，データサンドボックスの中を見ることはできません（図 3）．

③　データサンドボックスはデータ提供者およびアルゴリズム提供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ます．ただし，データサンドボックスのメモリ・ディスクは暗号化されているため，データおよびアルゴリズムをプラットフォーム事業者が見ることはできません（**図 4**）．

④　データサンドボックスはデータおよびアルゴリズムを用いて処理を実行します．この際メモリ・ディスク内のデータおよびアルゴリズムはCPU内に限り復号されて平文で処理されるため，高速な処理が可能です．処理結果がCPU外に出る場合は再びメモリ・ディスク上で暗号化されます（図 4）．

⑤　データサンドボックスは，処理結果のみをデータ提供者に返却します．データサンドボックスは外部との通信が適切に制限されているため，アルゴリズム提供者の故意・過失により不正なアルゴリズムが配置されたとしても，アルゴリズム提供者が元データや分析結果を入手することはできません（**図 5**）．

⑥　データサンドボックスは処理終了後にデータやアルゴリズムごと削除されます（図 5）．

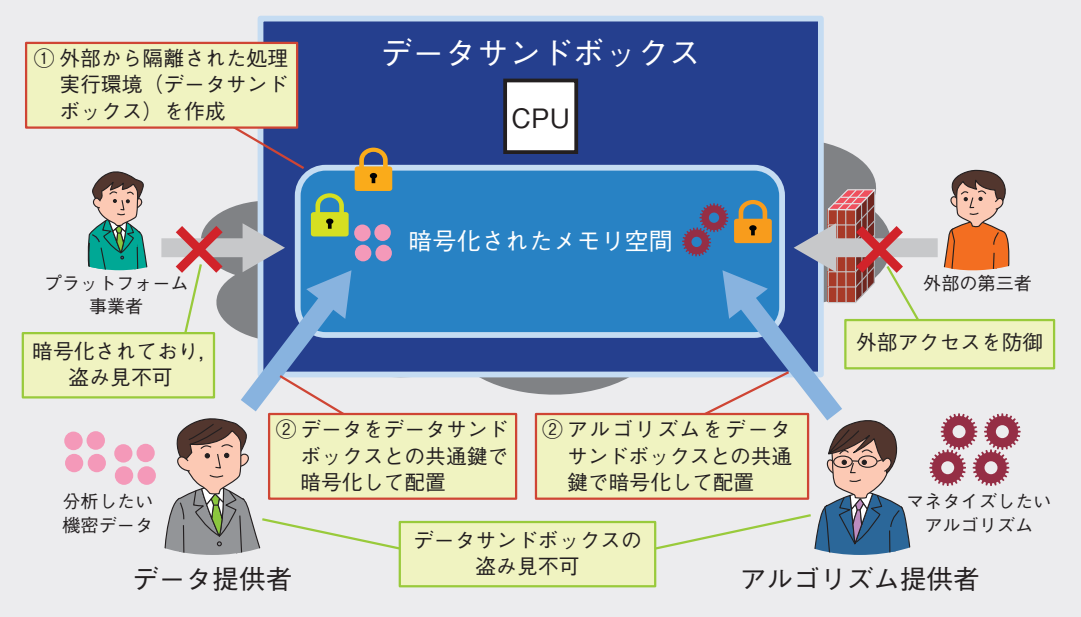

図 3　データサンドボックス（動作概要①）

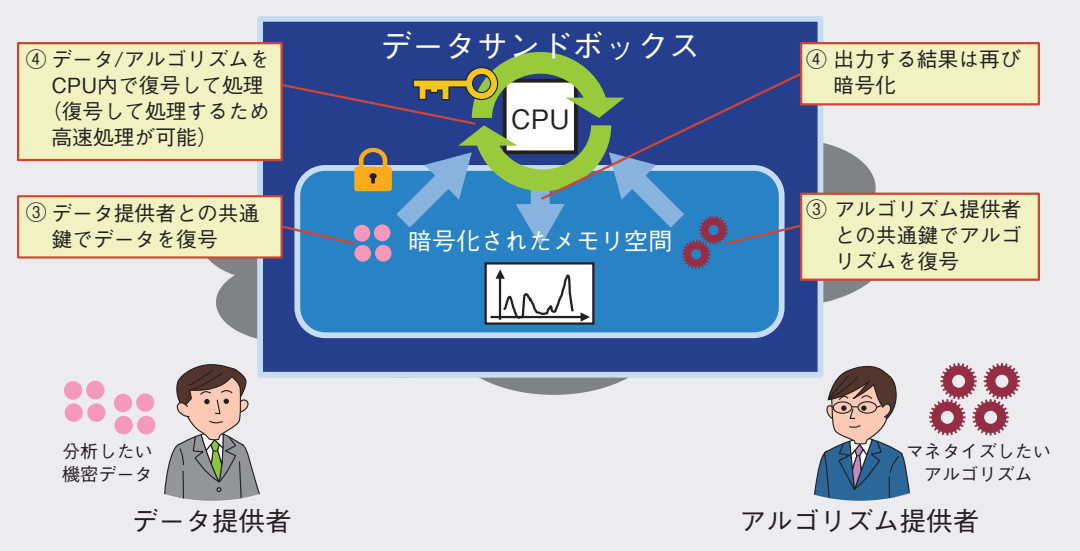

図 4　データサンドボックス（動作概要②）

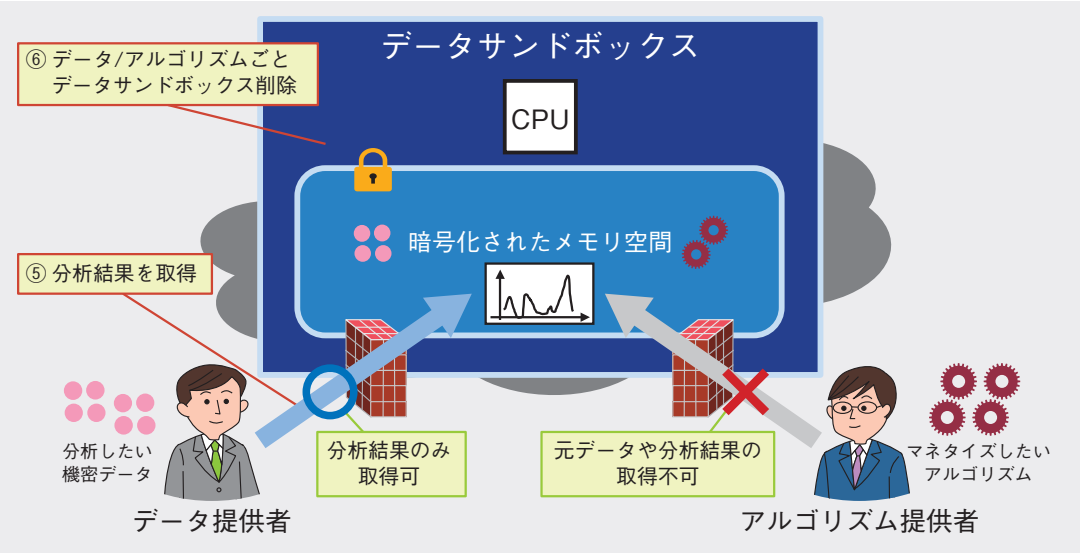

図 5　データサンドボックス（動作概要③）

このような動作をするデータサンドボックスを利用することにより，データ流通の当事者は以下のような効能を得ることができます．

・自分のデータやノウハウを相手に共有することなく，相手のデータやノウハウを活用し，新たな価値を得ることができる．
・自分のデータやノウハウをあらかじめ合意した目的以外に利用されたり，第三者に横流しされたりすることをシステム的・技術的に防止しながら，マネタイズすることができる．

■「相互認証・鍵共有技術」

次世代データハブにより世界中のデータの即時共有が可能になる一方，データ提供者が認めたデータ利用者のみがその内容を参照できるよう暗号化して授受する必要が出てきま

す．相互認証・鍵共有技術は，データ提供者とデータ利用者との間でお互いの識別子や属性を確認したうえで，授受するデータの暗号化や復号に用いる鍵を他者に見られずに共有する技術であり，所望の相手のみとの安全なデータ共有を実現します．

次世代データハブにはこれまでより極めて多くのIoTデバイスが接続し，大量の実世界データを供給するようになると考えられます．そのため，これまでより必要な計算リソースや通信帯域が少ない相互認証・鍵共有技術を開発しています．また，次世代データハブにつながる多数の者の間で互いにデータを提供・利用し合うことが想定されます．そのため，一対一ではなく，大人数の間での相互認証と鍵共有を効率的に行え，また，データを提供・利用し合う者の増減に応じた鍵の更新が柔軟に行える相互認証・鍵共有技術を開発しています．

■「秘密計算AI技術」

たとえ安全な実行環境下であったとしても，データが復号されることに不安を感じるデータ提供者や法的制約などにより復号することが許されないデータが少なからず存在します．秘密計算AI技術は，暗号化したデータを一切復号することなく，機械学習における学習と予測を可能とする技術です．データの登録・保管から，学習・予測までの一連の流れを，データの中身を誰にも明かすことなく実行できるため，企業の秘密情報やプライバシに関する情報の安全な流通と利活用が可能となります．また，複数の提供者からのデータや異なる種類のデータを暗号化したまま結合して利用することができるため，元のデータの安全性を向上させるだけでなく，分析対象となるデータの種類や量の増加による新たな価値の引き出しが可能になると期待されます．

今後に向けて

本稿では私たちが開発を進めている次世代データハブについてデータガバナンス機能を中心に紹介しました．今後は次世代データハブのもう1つの主要機能であるData Omnipresenceの開発やData Omnipresenceとデータガバナンスのシームレスな連携，All Photonics Networkとの連携による大容量・低遅延化の実現など，次世代データハブの実用化に向けた開発を加速していきます．

（左から）持田　誠一郎／ 長田　孝彦／
三原　淳慎

NTTソフトウェアイノベーションセンタでは，次世代データハブの実現を通して，データ中心社会の実現に貢献していきます．

◆問い合わせ先
NTTソフトウェアイノベーションセンタ
第三推進プロジェクト
TEL　0422-59-7724
FAX　0422-59-2699
E-mail　info-datahabu-p-ml＠hco.ntt.co.jp