

「信頼できるAI」の実現は「データ基盤」から
AI時代の
「データの主権」「信頼性」「ガバナンス」とは何か
東大・越塚教授と考えるデータ活用の未来像



生成 AI 活用が進む今、企業のビジネス価値を高める鍵にも変化が生まれ、「信頼できるデータ基盤」の重要性が高まっている。AI 時代に求められるデータ基盤に必要な要素は今までと何が違うのか。東京大学 越塚登教授とエンタープライズ ストレージ に詳しいネットアップ 神原豊彦氏の対談から答えを探る。

生成 AI の活用が企業の競争戦略として加速する中、データの価値や扱い方が大きく変わろうとしている。そこで焦点となるのがデータ環境の整備だ。すでに国際間のデータ流通を目指す DFFT (Data Free Flow with Trust) 構想を実現するための環境が整備されつつある。AI が人と同じようにデータを自律的に参照する AI エージェントの存在感も増している。これからのデータ基盤には、従来にない要件が求められるようになる。

本稿では、DFFT 構想を念頭にデータ流通基盤の整備をリードする東京大学大学院の越塚登氏と、組織の中のデータ基盤整備に詳しいネットアップの神原豊彦氏に、生成 AI 時代のデータの価値の変化とそれに対応するデータ基盤の要件について語ってもらった。

※以下、敬称略

「データ主権」と、「必要なときに、必要な人に、必要なデータ」を届ける仕組み

神原 先生は、国際的なデータ流通の仕組みを産官学横断で取り組まれています。その取り組みにおいて「データ」はどの位置付けられているのでしょうか。

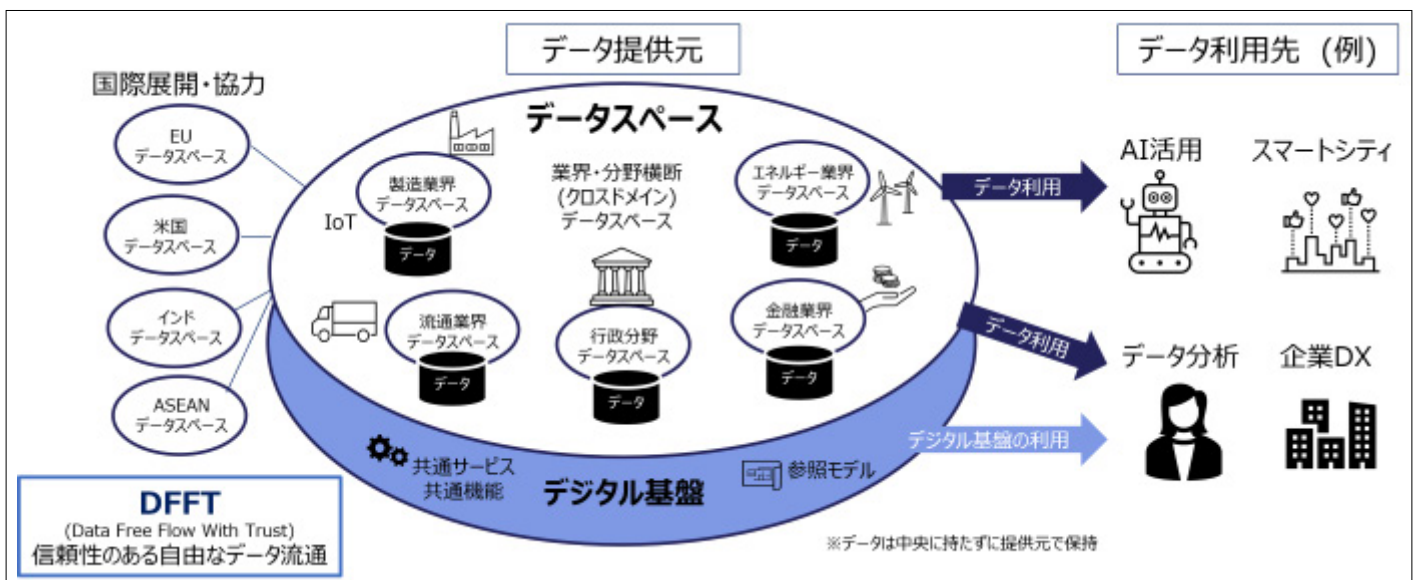
越塚 AI の急速な普及によってデータの重要性が急速に高まっています。どれほど優れた AI モデルを持っていても、質の高いデータがなければポテンシャルを発揮できません。しかし、オープンなデータは数年で AI が学習し尽くしてしまうといわれています。では、非オープンなデータはどうか。データの重要性は石油によく例えられますが、コピーできるため石油のように商品として流通させるのは難しいでしょう。

そこで重要になるのは「データの価値」に対する考え方です。本来、企業の資産は外部に出さないように厳重に管理しなければいけません。しかし、データには「オープンにして外部のデータと組み合

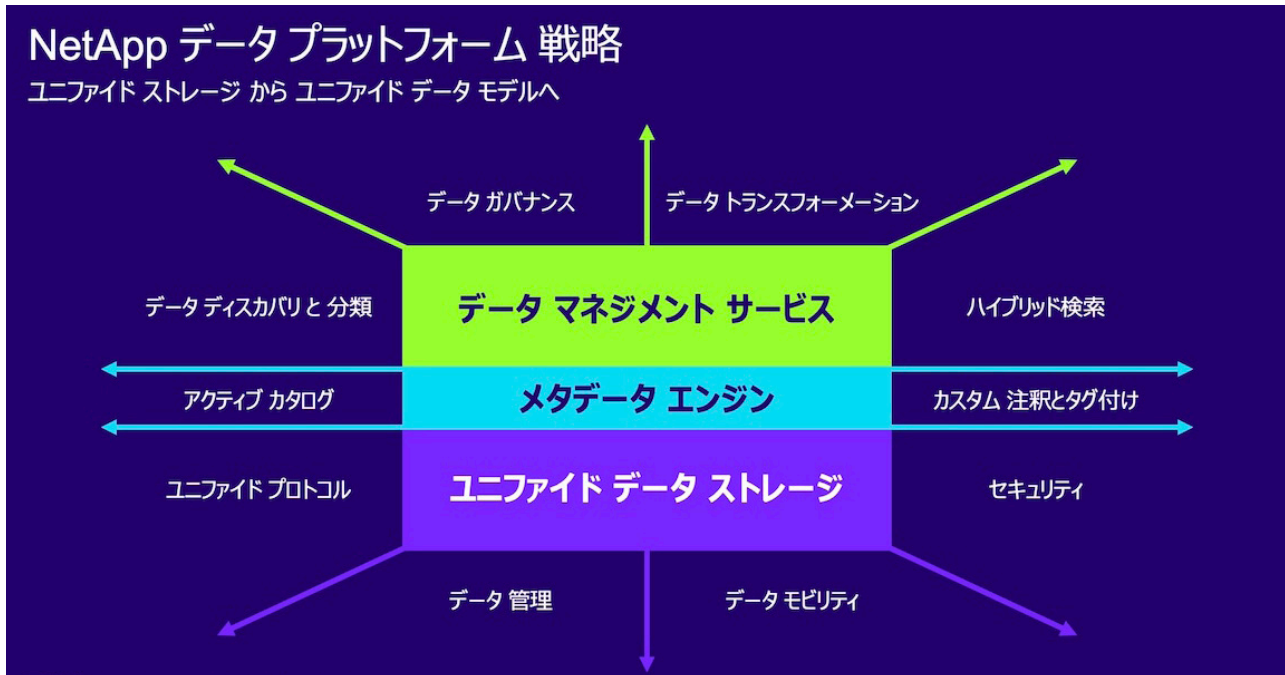


わせることで大きな価値を生む」という特性があります。大事なもののだけオープンにしないといけない。この相反する要求を技術で解決するために、私が取り組んでいるのが「データスペース」です。
越塚 データスペースの核心にあるのが「データ主権」です。インターネットは「いつでも、誰でも、どこでも」アクセスできることを理想として成長してきました。しかしビジネスの世界ではそれだけだと競争力につながりません。「必要なときに、必要な人に、必要なデータ」を届けられるように、適切にコントロールしようというのがデータ主権の考え方です。

神原 データ主権はお客さまの多くが直面している課題です。多くの企業が AI を利用したいという強い意欲を持っています。しかし、自社のデータがどこに保管され、誰がアクセスできるのかを適切に統制できなくなる懸念もあります。近年はクラウドサービスの普及により、データの物理的な所在地や管理権限があいまいになりました。「クラウドに預けたデータの所有権はどうなるのか」や「海外のクラウドサービスを使うと、データに外国の法律が適用されるのではないか」といった問い合わせが非常に増えています。



データスペースのイメージ (出典：情報処理推進機構「データスペース」)。異なる組織が互いに信頼性を確保しつつデータを自由に流通させるための枠組み。データのアクセス制御や利用条件の管理などの機能を組み込んだ、信頼できるデータ流通の仕組みの実現を目指す



NetApp のユニファイドデータモデル戦略の全体像 (提供：ネットアップ)

われわれは、オンプレミスやクラウドに分散したデータを統合管理するユニファイド データストレージにメタデータ エンジンやデータ マネジメント サービスを統合して、データ活用に必要な機能を包括的に提供する「ユニファイド データ モデル」を提唱しています。

この戦略の先には「メタデータ ファブリック」という構想があります。複数のストレージのメタデータ エンジンやデータ エンティティの相互の関連性を統合管理するアプローチです。先生が取り組まれているデータスペース整備が組織間で信頼性の高いデータを流通させる仕組みです。一方、われわれの取り組みは組織の中でデータを信頼できる状態で「必要なときに、必要な人に、必要なデータ」を届ける仕組みです。レイヤーは異なりますが、目指す方向は共通しています。

越塚 データ主権は、単なる技術的な問題だけではなく、経済安全保障や国家戦略にも関わる重要なテーマになっています。各国がデータ主権を重視する中、日本企業もこの問題に真剣に取り組む必要があります。

今後は人間や従来型のシステムだけでなく、自律的に動作するAI エージェントの利用も想定しなければなりません。AI モデルが参照するデータの信頼性やセキュリティを正しく管理する必要があります。メタデータを相互接続する手法は、これを実現する有効な手段になるでしょう。

「正しい問題の解き方を知っているAI」にはデータの主権、信頼性、ガバナンスが重要

越塚 組織の中のデータは今まで以上に重要になるでしょう。LLM (大規模言語モデル) の多くはオープンなデータを学習していますが、そこからは汎用 (はんよう) 的な回答しか得られず、必ずしも企業にとって価値のあるアウトプットが得られません。ビッグテックはAGI (汎用人工知能) の実現を目指していますが、それ以外の



東京大学大学院の越塚登氏 (情報学環教授 コビキタス情報社会基盤センター長)

アプローチもあると思います。必要なのは問題の「正しい解き方」を知っているAIです。

神原 今のLLMは一般常識には強いものの、各業界、各企業の保持する固有の情報や知識まではカバーできません。これまでに蓄積してきたデータをAIと組み合わせる必要がありますね。

越塚 そのためにはデータ主権に加えて「データの信頼性」の確保が欠かせません。データの信頼性には複数の側面があります。データの正確性や完全性といった品質。次に、データがどこから来て、どのように加工されたかという履歴の管理。そして、データへのアクセス権限や利用目的が適切に管理されているかどうかというガバナンスです。これらを技術と制度の両面から支援することが重要です。

神原 AI開発の現場にも多くの課題があります。AIモデルの学習には膨大なデータが必要ですが、そのデータが分散していたり、アクセス権限が複雑だったり、データの品質が低かったりするとプロジェクトは前に進みません。データサイエンティストの作業時間の81%がデータの準備に費やされているという調査結果もあります。

インテリジェントデータインフラストラクチャの価値と効果



インテリジェントなデータストレージを実現

データ基盤 そのものとなる
オンプレミスとクラウドの
ストレージを賢くします

- サイロの解消: オンプレミスと主要クラウド (AWS, Azure, Google Cloud) で共通のOS (ONTAPなど) を利用し、環境の壁を超えて一貫したデータ管理を可能にします。
- AI対応: NVIDIAとの強力なパートナーシップにより、大量のデータをAIモデルに効率よく供給できる性能と拡張性を備えています。



インテリジェントなデータサービスを実現

データの保護、セキュリティ、
ガバナンスを自動化し、ユーザーの
データアクセスを便利にします

- サイバーレジリエンス: ランサムウェア攻撃をAIでリアルタイムに検知し、感染前の状態へ即座に復旧させる機能を統合しています。
- データ保護: ポリシーに基づいた自動的なバックアップやレプリケーションにより、管理負担を軽減します。



インテリジェントなデータオペレーションを実現

オペレーバビリティ (観測性) と
AIを活用し、限られた人的リソースの中で
データ運用を最適化します

- 可視化: Data Infrastructure Insightsなどのツールを通じ、マルチクラウド環境全体のパフォーマンスやコスト、使用状況を統合的に把握します。
- 適応型運用: 予測分析を用いて、問題が発生する前にリソースを自動調整するなど、ビジネスの優先目標に合わせて動的に変化する基盤を提供します。

IDI がもたらす 3 つの価値 (提供: ネットアップ)

AI で価値を生むには、この非効率を解消してデータをすぐに使える状態にすることが不可欠です。

「Intelligent Data Infrastructure」(IDI) における「データの主権」「信頼性」「ガバナンス」

神原 従来のデータストレージ装置は業務システムのデータ保管が主な用途であり、AIに最適化されていません。そこでわれわれが提唱しているのが「Intelligent Data Infrastructure」(IDI) というコンセプトです。IDIは、企業が保有するさまざまなデータをAI用に最適化する、新しいデータ基盤像です。

オンプレミスとクラウドのストレージを賢く使い分け、AIワーク



ネットアップの神原豊彦氏 (チーフテクノロジー エヴァンジェリスト)

ロードに最適化された分散ストレージアーキテクチャによる「インテリジェントなデータストレージ」、データの保護やセキュリティ、ガバナンスを自動化し、ランサムウェア攻撃をAIでリアルタイムに検知する「インテリジェントなデータサービス」、オペレーバビリティとAIによって限られた人的リソースでデータ基盤を運用できるように支援する「インテリジェントなデータオペレーション」という価値を提供します。地球上で最もセキュアなデータ基盤を目指しています。

IDIを実現する製品のひとつが「AI Data Engine」です。データ基盤の内部でAIデータパイプラインを実現します。先ほども言ったように、AIプロジェクトではデータの準備に多くの時間を割いているという課題があります。このプロセスをデータ基盤が担うべきだ、というのがわれわれの考えです。

メタデータエンジンによるデータの発見と理解、データシンクによる自動更新、データガードレールによるセキュリティ確保、データキュレーターによるAIアプリケーション用のデータ変換といった機能を提供します。AIアプリケーションやAIエージェントが参照しやすいデータ基盤を整備することで、企業が蓄積してきたデータを「AI-Ready」な状態にできる点が大きな特徴です。

越塚 既存のデータ基盤をAI時代に適応させることで、企業資産を効率的に活用できるでしょう。日本企業が持つ品質の高いデータと、うまく組み合わせられる基盤が整えば、日本の勝ち筋も見えてくるはず。データ基盤製品を提供するベンダーの活躍に大いに期待しています。

●お問い合わせ

ネットアップ合同会社

東京都中央区京橋 2-1-3 京橋トラストタワー

<https://www.netapp.com/ja/>

<https://www.netapp.com/ja/forms/sales-contact/>

※この冊子は、ITmedia エンタープライズ (<https://www.itmedia.co.jp/enterprise/>) に 2026 年 3 月に掲載されたコンテンツを再構成したものです。
<https://www.itmedia.co.jp/enterprise/articles/2602/24/news003.html>

copyright © ITmedia, Inc. All Rights Reserved.