データ活用社会創成プラットフォーム mdx の設計・実装・運用
～多様な学際領域における共創に向けて～

鈴村 豊太郎[1)]，杉木 章義[2)]，滝沢 寛之[3)]，今倉 暁[4)]，中村 宏[1)]，田浦 健次朗[1)]，工藤 知宏[1)]，塙 敏博[1)]，関谷 勇司[1)]，小林 博樹[1)]，松島 慎[1)]，空閑 洋平[1)]，中村 遼[1)]，姜 仁河[1)]，川瀬 純也[1)]，華井雅俊[1)]，宮嵜 洋[5)]，石﨑 勉[5)]，下徳 大祐[5)]，関本義秀[6)]，樫山武浩[6)]，合田 憲人[7)]，竹房 あつ子[7)]，政谷 好伸[8)]，栗本 崇[9)]，笹山 浩二[9)]，北川 直哉[9)]，藤原 一毅[10)]，朝岡 誠[10)]，中田秀基[11)]，谷村 勇輔[11)]，青木 尊之[12)]，遠藤 敏夫[12)]，森 健策[13)]，大島 聡史[13)]，深沢圭一郎[14)]，伊達 進[15)]，天野 浩文[16)]

1）東京大学 情報基盤センター
2）北海道大学 情報基盤センター
3）東北大学 サイバーサイエンスセンター
4）筑波大学システム情報系
5）東京大学 情報システム部情報基盤課
6）東京大学空間情報科学研究センター
7）国立情報学研究所 アーキテクチャ科学研究系
8）国立情報学研究所クラウド基盤研究開発センター
9）国立情報学研究所学術ネットワーク研究開発センター
10）国立情報学研究所 オープンサイエンス基盤研究センター
11）産業技術総合研究所 デジタルアーキテクチャ研究センター
12）東京工業大学 学術国際情報センター
13）名古屋大学 情報基盤センター
14）京都大学 学術情報メディアセンター
15）大阪大学 サイバーメディアセンター
16）九州大学 情報基盤研究開発センター

suzumura@ds.itc.u-tokyo.ac.jp

The *mdx* Project Report 2021
A Large-Scale Platform for Accelerating Cross-Disciplinary Research Collaborations Towards Data-Driven Societies

Toyotaro Suzumura[1)], Akiyoshi Sugiki[2)], Hiroyuki Takizawa[3)], Akira Imakura[4)], Hiroshi Nakamura[1)], Kenjiro Taura[1)], Tomohiro Kudoh[1)], Toshihiro Hanawa[1)], Yuji Sekiya[1)], Hiroki Kobayashi[1)], Shin Matsushima[1)], Yohei Kuga[1)],Ryo Nakamura[1)], Renhe Jiang[1)], Junya Kawase[1)], Masatoshi Hanai[1)], Hiroshi Miyazaki[5)],Tsutomu Ishizaki[5)], Daisuke Shimotoku[5)],Yoshihide Sekimoto[6)],Takehiro Kashiyama[6)], Kento Aida[7)], Atsuko Takefusa[7)], Yoshinobu Masatani[8)], Takashi Kurimoto[9)], Koji Sasayama[9)], Naoya Kitagawa[9)], Ikki Fujiwara[10)], Makoto Asaoka[10)], Hidemoto Nakada[11)], Yusuke Tanimura[11)],Takayuki Aoki[12)], Toshio Endo[12)], Kensaku Mori[13)], Satoshi Ohshima[13)], Keiichiro Fukazawa[14)], Susumu Date[15)], Hirofumi Amano[16)]

1) Information Technology Center, The University of Tokyo
2) Hokkaido University Information Initiative Center
3) Cyberscience Center, Tohoku University
4) Faculty of Engineering, Information and Systems, University of Tsukuba
5) Division for Information and Communication Systems, The University of Tokyo
6) Center for Spatial Information Science, The University of Tokyo
7)Information Systems Architecture Science Research Division, National Institute of Informatics
8)Center for Cloud Research and Development, National Institute of Informatics
9)Research and Development Center for Academic Networks, National Institute of Informatics
10)Research Center for Open Science and Data Platform, National Institute of Informatics

11) Digital Architecture Research Center, National Institute of Advanced Industrial Science and Technology
12) Global Scientific Information and Computing Center, Tokyo Institute of Technology
13) Information Technology Center, Nagoya University
14)Academic center for Computing and Media Studies, Kyoto University
15) Cybermedia Center, Osaka University
16) Research Institute for Information Technology, Kyushu University

概要

我が国が目指す Society 5.0 はデータ利活用の恩恵をだれもが安心して享受できるインクルーシブな社会である。このような社会の実現には、幅広い用途に使える情報基盤の整備と、知識集約の中核を担う大学・研究機関をハブとしたデータを解析したい人と解析技術・公開データを結ぶ人的環境の形成を、日本全体で進めることが不可欠である。このような社会を支えるプラットフォームとして、データ活用社会創成プラットフォーム mdx が設計・実装され、2021 年 9 月にテストリリースとして、9 大学 2 研究機関での共同運営が開始された。本稿では、ハードウェア構成、提供するソフトウェア・サービスを含めた mdx プラットフォームの概要、そして、2021 年度に開始された連携プロジェクトの一部を紹介する。

1 はじめに

Society 5.0 は、地域、年齢、性別、言語による格差等の課題を解消し、地域の特色を活かした多様な産業の活性化に貢献する社会を目指している[1]。その実現にむけてこれからの社会ではあらゆる分野でデータ活用が必須であり、学術コミュニティにおいても、社会実装を指向した研究分野が重要である。実際に様々な研究分野でデータが研究における重要な資産となってきた[2]。本稿では、これらのデータを利活用し、データを中心とした様々な学際的な共創を促進するためのプラットフォーム mdx の概要、ハードウェア構成、提供するシステムおよびサービスに関して述べると共に、材料科学、空間情報科学、気象、地域経済・資源循環など、mdx 上で開始されている連携プロジェクトの一部を紹介する。

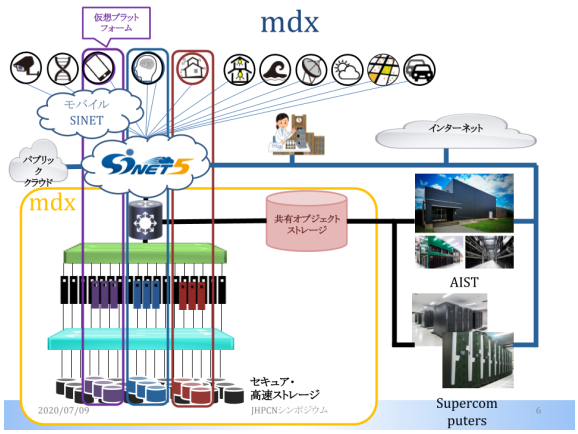

図 1： mdx 概要図

2 mdx プラットフォームとは

mdx プラットフォームは、データ中心的な研究分野や、データ解析・データ科学的手法への期待が高い分野に情報基盤を提供し、分野・産学間の連携を促進する取り組みである。国立情報学研究所(NII)との綿密な連携の元、2018 年初頭から構想を始め、全国の大学との連携・協力関係を築きつつある。全国利用を前提としたデータプラットフォームの先導的システム整備が 2020 年度末に東京大学柏Ⅱキャンパスに実施された。mdx は次の 3 つの特徴を有している [3]。

① データ収集機能：IoT データや大規模リアルタイムデータを円滑に扱えるセキュアな大容量通信回線を提供する。現在、国立情報学研究所では各都道府県を 100Gb/s 以上の帯域で接続した学術ネットワーク SINET 5（以下「SINET5」という）を運用している。また SINET 5 に接続したモバイル網を「SINET 広域データ収集基盤」として提供している[18]。これは、各プロジェクトごとに、モバイル仮想閉域網(Mobile Virtual Private Network; Mobile VPN)を提供し、その VPN をクラウドなど計算基盤まで延伸することを可能にしたものである。データプラットフォームもこの広域データ収集ネットワークと連携し、日本全国からのデータの収集から蓄積・処理までをプロジェクトごとに閉じたセキュアな環境で行うことを可能にする。

② データ解析機能：高度・高速な解析を実現する高性能計算環境やストレージを提供する。mdx では、IaaS 環境と同様に、仮想化技術を用いて複数のプロジェクトに用途ごとに分離された管理者権限で自由にソフトウェアの設定が可能なプライベート環境であるテナントを提供する。個々のプロジェクトには一つまたは複数のテナントが割り当てられる。mdx は、広域ネットワークと連携

し、共通のインフラを用いて使いたいときに短期間で広域ネットワーク、計算機、ストレージなどから構成される広域にまたがるテナントをプロジェクトに割り当てる。これは、利用する個々のプロジェクトから見ると、専用のインフラが整備されたかのように使える。mdx 全体の管理を行う管理者を「システム管理者」、mdx に資源を要求し、テナントの割り当てを受けて利用するユーザを「テナント管理者」と呼ぶ。利用者の管理はテナント管理者に任されており、例えば利用者になんらかの形でテナント上に構築した情報システムへのログインを許す運用もあり得る。テナント管理者がポータルを介してテナントを要求すると、資源管理ソフトウェアとシステム管理者によりテナントが割り当てられる。仮想インフラ上の VM へ他のネットワーク（インターネットや、他のテナント、SINET5 上の VPN など）からのアクセスを許すかどうかは、テナント管理者がポータルを介して設定する。

③ 応用開発基盤：多様な応用を実現する基盤ソフトウェアと共用データを提供する。mdx はこれまでのスーパーコンピュータとは大きく異なる分野、異なる形態での利用が想定される。スーパーコンピュータは主に大規模な並列計算を高性能に行うためのものであり、大規模データ処理や深層学習にも適している。そのレベルで計算機の構成を抜本的に変える必然性はないが、これまでの運用方法ではサポートできない利用形態が存在する。

一つの形態は「プラットフォームのプラットフォーム（メタプラットフォーム）」である。それは、mdx の「ユーザ」とは、単にデータ処理をするための計算資源としてプラットフォームを利用することにとどまらない。分野のデータレポジトリを整備し、それを分野研究者に公開する、つまり「プラットフォーム構築」をするユーザであり得るということである。このようなユーザをサポートするには、これまでのスーパーコンピュータの設計・運用とは異なる要素を取り入れる必要がある。外部ネットワークへの接続、恒久的な資源の割当て、分野プラットフォーム構築のために柔軟にシステムソフトウェアを構成できることが必要で、それと合わせて大規模機械学習処理やデータ同化シミュレーションなどのために、高性能な計算資源と連携できる必要がある。

3 mdx を支えるハードウェア構成

本章では、mdx のシステム構成を述べる。mdx システムは、スーパーコンピュータと同等のハードウェアを備えているが、マルチテナントのクラウドサービスと同等以上のセキュリティを実現できるように設計されている。高速なネットワーク、大容量のストレージを持ち、柔軟かつセキュアで高性能な計算環境を提供する。

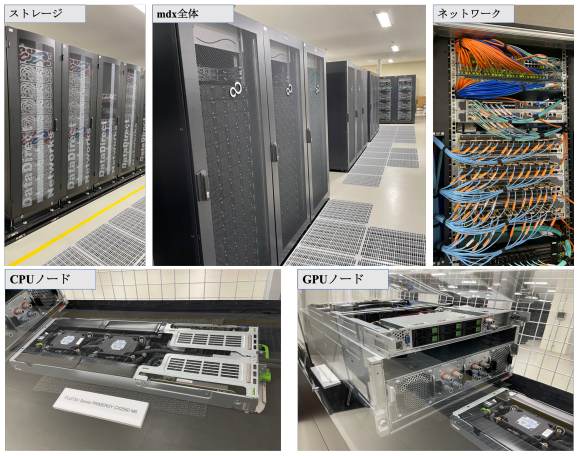

図 2：mdx を構成するハードウェア s

mdx の仕様を表 1 に、実際のハードウェアの一部を図 2 に示す。計算ノードは、高性能 CPU を備えた汎用 CPU ノード群と、CPU に加えて高性能 GPU を備えた GPU 演算加速ノード群からなり、利用者には隔離されたプラットフォームを提供するため、VMware vSphere により仮想化環境を実現する。ストレージは、Lustre ファイルシステムによる高速内部ストレージ(NVMe)、大容量内部ストレージ(HDD)と、仮想マシン用には別途ストレージを用意している。さらに、外部とのデータ共有のために、Amazon AWS S3 互換の共有オブジェクトストレージも備える。

ネットワークには、外部接続を実現するサービスネットワークと、ノード間の高速な RDMA 転送とストレージ転送を担う内部高速ネットワークの 2 種に分離している。これは、サービスネットワーク経由で外部から侵入されたとしても、ストレージ領域を保護するためである。いずれのネットワークも Virtual eXtensible LAN (VXLAN)によるオーバーレイネットワークを用い、内部高速ネットワークにはさらに RDMA over Converged Ethernet (RoCE) v2 を用いている。

内部高速ネットワークへの接続方式として、(1)VMware デフォルト方式(TCP 接続)、(2)準仮想化 RDMA(PVRDMA)、(3)SR-IOV を提供している。ノード間の RDMA 転送は(2)(3)で可能であるが、Lustre への RDMA 転送は(3)のみで可能である。また、各 VM から GPU へは PCIe パススルー方式で接続される。つまり、通常のベアメタルサーバと同等の直接アクセスを実現している。

利用者が mdx を使用する際、ポータルを通じて、必要な計算資源、ストレージ資源、ネットワーク構成を要求する。このとき、VXLAN を用いてプロジェクト毎にサービス、ストレージ及び RDMA ネットワークに VLAN を割り当て、隔離されたテナントが構築される。各テナントは SINET に延伸が可能で、SINET の L2VPN サービスと併用することで、SINET に接続されている研究機関であれば、研究室等の計算機と mdx 上の計算・ストレージ資

源が直結された環境も構築できる。また、モバイル SINET 経由で IoT デバイスからも mdx 上の計算環境にセキュアにデータ転送・収集が可能である。

表 1：mdx のハードウェア仕様

汎用 CPU ノード	ノード	富士通 PRIMERGY CX2550 M6
	CPU	Intel Xeon Platinum 8368 (IceLake, 38 cores/76 threads, 2.4 GHz) ×2 socket
	メモリ	256 GiB
	ネットワーク IF	Intel XXV710 25Gb/s Ethernet +Mellanox ConnectX-6Dx 100Gb/s Ethernet with RoCEv2
	ノード数	368 ノード
	総理論演算性能	2.1 PFLOPS (FP64), 4.2 PFLOPS (FP32)
GPU 演算加速ノード	ノード	富士通 PRIMERGY GX2570 M6
	CPU	Intel Xeon Platinum 8368 (IceLake, 38 cores/76 threads, 2.4 GHz) ×2 socket
	メモリ	512 GiB
	GPU	NVIDIA A100 Tensor Core GPU 40GiB x8
	ネットワーク IF	Intel XXV710 25Gb/s Ethernet x2 + Mellanox ConnectX-6Dx 100Gb/s Ethernet with RoCEv2 x4
	ノード数	40 ノード
	総理論演算性能	6.5 PFLOPS (FP64), 6.7 PFLOPS (FP32)
ストレージ	高速(NVMe)	Lustre File System 1.0 PB, 252 GB/s
	大容量 (HDD)	Lustre File System 16.3 PB, 157.5GB/s
	外部オブジェクト	DDN S3 Data Service with Lustre File System 10.3 PB, 63.0 GB/s
	仮想ディスク	NFS 0.55 PB
ネットワーク	対外接続	100 Gb/s (2022 年春以降 SINET6 の移行時には 400 Gb/s 増速予定), Wisteria/BDEC-01: 400 Gb/s, 産総研: 100Gb/s
	サービス	25 Gb/s Ethernet NIC
	ストレージ& RDMA	100 Gb/s Ethernet NIC
	仮想化	EVPN-VXLAN
仮想化ソフトウェア		VMware vSphere

4 mdx プラットフォームの設計・実装

本章では mdx プラットフォームにおけるソフトウェア面、特にユーザが使用するポータルの概要を述べ、次に、計算資源量の割当ポリシー、仮想マシンテンプレート、そして 2022 年度の本番運用に向けた現状を述べる。

4.1. mdx ポータルの概要

まず、ユーザ目線で mdx をどのように用いていくかを述べる。機関管理者用のポータル画面も実装されているが、本稿では、mdx プラットフォームを使用するユーザが VM を使用するまでのステップの概要を述べる。

Step 1: プロジェクト申請：プロジェクト代表となるユーザは、学認(学術認証フェデレーション)[14][15]の認証方式を用いて、「申請ポータル」にログインする。申請ページでは、プロジェクトの目的、計算資源を要求する連携機関、リソース量（CPU 量、GPU 量、ストレージ量、グローバル IP アドレス、使用期間等）を指定する。申請は指定した連携機関の管理者が承認した後、ユーザポータルにログインすることが可能になる。

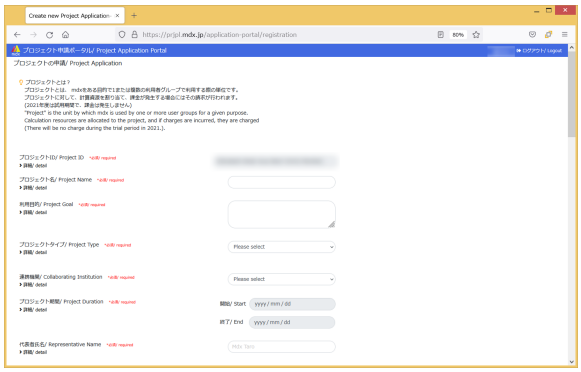

図 3： ログイン画面

Step 2: VM 配備と起動：プロジェクト代表者はユーザポータルにログインした後は、mdx の管理者もしくはユーザによって提供された仮想マシンテンプレートもしくは ISO イメージを元に VM を構築する。4.3 章において詳述するが、GPU ドライバーや Lustre の設定など、mdx プラットフォームの機能を使用するための推奨仮想マシンテンプレートが用意されており、独自の環境を構築する必要がない場合はこの仮想マシンテンプレートを選択する。

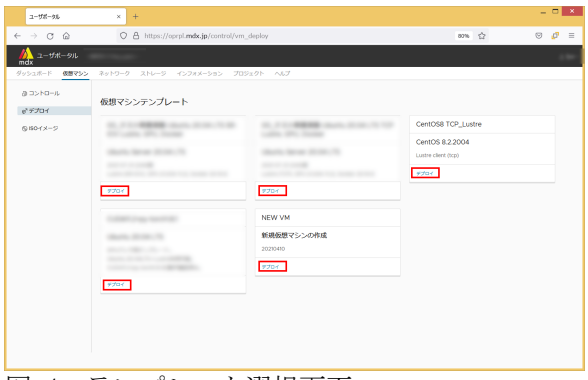

図 4：テンプレート選択画面

仮想マシンテンプレートを選択した後は、プロジェクト申請時に申請した資源量を上限として、CPU 数、GPU 数、ストレージ量を指定して、VM 配備を実行する。

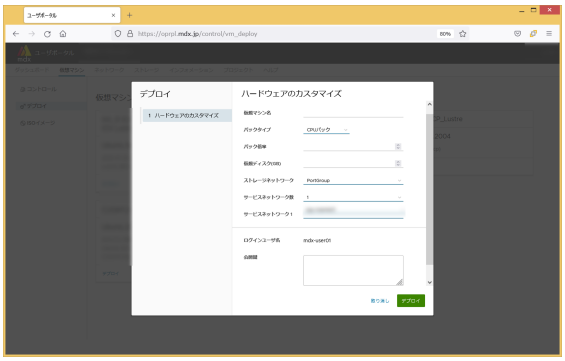

図 5：VM 配備設定画面

VM 配備のリクエストを送信した後は、DHCP によって IP アドレスの割当を待つ。ローカル IP アドレスが割り当てられた後は、グローバル IP アドレスとそのローカル IP アドレスのマッピングを DNAT で設定する。また、ACL の設定画面において、どの IP アドレスからアクセスを許可するかを指定する必要がある。この設定が完了した後、SSH でその VM にログインする事が可能になる。

上記の 2 ステップが利用者目線で見た時の利用ステップである。この他に、VM のクローン複製や複数 VM の配備なども実装されている。

4.2. 計算資源量の割り当てポリシー

mdx は 11 機関で共同運用され、各機関には一定量の計算資源が割り当てられる。利用者から各機関に対して行われた申請に対して、各機関は審査等を行い、割り当てられた計算資源を利用者に割り当てることになる。これにより、各機関が保有す資源量の範囲においては、各機関の判断でプロジェクトの受け入れが可能となり自由な運用が可能となる。

利用者に割り当てる資源は、以下の単位(パックと呼ぶ)となっている。利用者は利用申請ポータルからパック単位で自由に資源量を設定し、申請(利用)することができる。

- **汎用 CPU ノード**：vCPU 1 コア単位で指定、コア当たりのメモリ量を比例配分(1 コア、vMem 1.64 GB/コア)
- **演算加速 GPU ノード**：GPU 単位で指定、コア数、メモリ量は演算加速 GPU ノード全体の 1/8 ずつ(1GPU、18vCPU コア、vMem 62GB/GPU)
- **ノード占有**：ノード単位での割り当て(表 1 に記載された CPU ノードもしくは GPU ノードを専有する。ハイパーバイザー等の管理領域を除く。)

現時点では mdx 全計算資源の約 50%を各機関が利用できる資源として配分しているが、2022 年度以降はプロジェクトへの割り当て実績をもとに運営委員会にて資源配分量を調整していくことになる。なお、残りの資源量については利用者が一時的に申請以上の計算資源量を利用できる環境を構築、利用できるようにする予定である。

4.3. 仮想マシンテンプレート

mdx のような仮想マシンを用いた計算機基盤の利点は、利用者が自身の仮想マシンの管理者権限を持ち、自由に OS 環境構築が可能な点である。一方で、利用者は、デプロイした仮想マシンごとに OS の設定であったり、ソフトウェアのインストールを自身で実施する必要がある。mdx では、そのような仮想マシンの環境整備を支援する目的で、仮想マシンテンプレートと仮想マシンの構成管理スクリプトを提供している。

はじめに、現在 mdx では、仮想マシンを準備する方法として、(1) ISO イメージを用いた手動による OS インストール、(2) OVF (Open Virtualization Format)フォーマットによる仮想マシンイメージのインポート機能、(3) mdx が事前に用意したセットアップ済み仮想マシンイメージを用いた OS のデプロイ、の 3 つの方法が利用可能になっている。そのうち、(3)の事前に用意したセットアップ済み仮想マシンイメージを mdx では「仮想マシンテンプレート」と呼んでいる。

現在、仮想マシンテンプレートでは、Ubuntu Linux(Server と Desktop イメージ)を提供している。仮想マシンテンプレートは、mdx の利用に必要なネットワークドライバなどのソフトウェアのインストールと、ネットワークや NTP (Network Time Protocol)といった OS 設定が事前に用意された仮想マシンイメージである。利用者は、仮想マシンテンプレートを用いることで、OS セットアップの時間を短縮できる。今後、mdx では Ubuntu 以外の仮想マシンテンプレートについても提供を予定している。また、mdx では、仮想マシンテンプレートに加えて、複数台の仮想マシンを一括してセットアップする仮想マシンの構成管理スクリプトである machine-configs を Github 上[5]で提供している。machine-configs は、mdx 用に用意

したAnsible[6]のスクリプト集であり、データ科学用に必要な計算機環境の設定、例えば、NFS でのファイル共有、Lustre ファイルシステムのマウント、LDAP によるユーザ管理、CUDA Toolkit、Jupyter などを、複数仮想マシンに対して一括インストールと設定ができる。mdx では、今後も仮想マシンテンプレートや machine-configs のような、利用者支援の機能開発を継続していく。

4.4. 運用に向けて

mdx プロジェクトでは、11 機関が mdx 立ち上げワーキンググループにより協定書案を準備し、正式に「データ活用社会創成プラットフォーム共同研究基盤の設置及び運営に関する協定」を 2021 年 6 月 1 日に締結した。8 月末には国立情報学研究所が運営する学認を用いた認証の正式運用が可能になり、そして利用規約を 9 月 8 日に制定、9 月 22 日にテストリリースされた。10 月 11 日には「データ利活用社会創成シンポジウム 2021」[13] を開催する。なお、10 月下旬以降に運営委員会を設置、2022 年度からの本格運用を目指している。

5 mdx を用いた共創領域の一部紹介

現在、mdx 上での様々な共創の取り組みが検討されており、一部の取り組みは開始している。本章では、空間情報科学、材料科学、気象、地域経済・資源循環の 4 領域における取り組みを報告する。あくまで、本稿では一部のみの紹介であり、より広範囲な共創の議論がされていると共に、トップダウンの共創だけではなく、ユーザからのボトムアップの共創プロジェクトが期待される。

5.1 空間情報科学

人々の移動データは、様々な都市政策の共通なデータ基盤である事は明白であるものの、民間ベースの携帯端末データをもとにしたものは、集計ベースであっても依然高価である。さらに個人情報保護などの規制の側面などもあり、研究者の観点から自由な発想に基づき、人間の行動変容などの科学的な解明のために利用することは難しい状況にある。そこで、我々は、2008 年に人の流れプロジェクト[21]を立ち上げ、PT 調査ベースの移動データ（海外含め 36 都市圏、延べ約 700 万人分）を研究者に公開する取り組みを開始した。現在では、国や地域の政策現場で汎用的に活用できることを念頭に置いて、PT 調査エリアの制約を受けることなく、シームレスな移動データの提供を目指して、全国規模の擬似人流データの開発を進めている。これまでの研究成果として、pre-trip 型のアクティビティモデルを構築することで、全人口分の典型的な 1 日の行動を表現するトリップチェーンデータの試作を行った。現状のモデルでは、すべての行動パターンを再現できているわけではないが、それでも試作データは２億３千以上ものトリップを含んでいる。ここから、時空間内挿処理によって疑似人流データ生成するのだが、これには膨大な計算処理リソースが必要となることから、今回、mdx プラットフォーム上に構築したインスタンス（100vCPU、160G メモリ）を利用した。時空間内挿処理について説明すると、図 6 に示すように、トリップごとの起点、終点間の最短経路をダイクストラ法により計算し、探索された経路に対して、距離情報をもとにタイムステップごとの位置を決定している。最終的に作成された疑似人流データの可視化結果を図 7 に示す。今後もモデルの改良や交通シミュレータの導入を計画しており、その中では、mdx が提供する豊富なリソースを活用していく予定である。

図 6：時空間内挿の概要

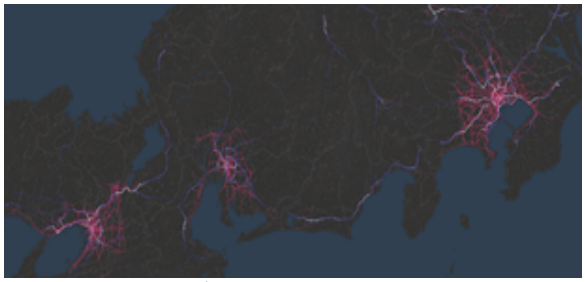
図 7：疑似人流データの可視化結果

5.2 材料科学

近年、材料開発において材料科学と情報科学・機械学習を組み合わせた効率化の試み、いわゆるマテリアルズ・インフォマティクス（MI: Materials Informatics）が注目されており、今後のデータ活用社会において重要な役割を担うと考えられている。MI においてデータとは高性能計算機による長時間計算によって得られるシミュレーション結果（第一原理計算など）や実験によって得られた計測データであり、その集約化・オープン化が進められている[7-9]が、品質や種類は未だに限定的である。材料データの収集や利活用において、材料科学と計算科学・機械学習の双方に研究課題が多く、分野や組織を横断した連携が不可欠である。

そこで我々は、「マテリアル先端リサーチインフラ」事業（文部科学省が進めるマテリアル研究のデジタル革新より[10]）との連携を開始し、mdx を用いた MI 研究に着手する。具体的にはその先駆けとして、熱伝導材料を対象にした高精度物性予測モデルの研究開発と高品質なシミュレーションデータセットの収集を開始した。材料の熱伝導率は他の物性値（例えば弾性率）に比べその計算コストが高く、既存のオープンデータベースにおいてその高品質な計算データは非常に限られている。例えば、本分野において人気の高い大規模オープンデータベース Materials Project では、弾性率のデータ登録数は 13 万であるのに対し熱伝導率のデータ数は 100 に満たない[7]。高精度の物性予測モデルを構築するために、少データの有効活用と高品質なデータの収集を同時に進めることが急務である。本研究をベースに今後マテリアル先端リサーチインフラとの連携をより強化し、mdx の材料科学への全般的な貢献を予定している。

5.3 気象

近年、豪雨・台風などによる深刻な気象災害が顕著になっており、今後の地球温暖化の影響による災害リスクの増大が懸念されている。そこで、過去の日本域における気候変動や異常気象に関して観測データを最新の数値予報モデルに取り込む「日本域気象再解析」により、地域的な大気状態の全体像を、長期間にわたり均質に矛盾なく 4 次元的に再現する。「地域気象データと先端学術による戦略的社会共創拠点(ClimCORE)」（代表機関：東京大学先端科学技術研究センター）[11]の下で、過去の観測データと最新の数値予報モデルにより、スーパーコンピュータを用いてデータ同化と気象予報の組み合わせにより、詳細かつ長時間にわたって再解析を行い、データベースを構築する。構築されたデータベースを用いて、今後の防災・現在対策の立案や、農林水産業や再生可能エネルギーなど、多様な分野への産学官公連携による社会応用を目指している。我々は mdx を通じたデータの利活用と基盤の構築に関して連携していく。オンデマンドで常に利用できる mdx の計算環境と、必要に応じてバッチスケジューリングによる大量の高性能計算を実現するスーパーコンピュータに処理をオフロードすることで、処理要求量の変動に対応しつつデータ処理を効率的に行う仕組みを構築する。

5.4 地域経済・資源循環

近年、自治体や地域において、脱炭素などの持続可能な開発目標(SDGs)などの目指すべきビジョンや、再生可能エネルギーに関する社会経済的な仕組みなどが提案・実施されているが、具体的な対策の検討や計画は困難である。「資源を循環させる地域イノベーションエコシステム研究拠点」（代表機関：東京大学未来ビジョン研究センター)[12] では、産学公の共創により、地域資源の循環利用で到達できる物質・エネルギーシステムの設計、地域経済循環の可視化に基づく技術と地域システムのマッチング、最先端知に基づくビジョンと地域の Co-learning、論理・論拠・情理に基づく地域資源を活用する物質・エネルギーシステムの実証・実装を目指している。我々はこの拠点活動を加速するため、mdx において秘匿性の高いデータ利活用システムを構築する。5.3 節でも述べたスーパーコンピュータとの連携機構を活用して効率の良いデータ処理基盤を実現する。

6. mdx 専用データ公開基盤に向けて

mdx において整備を進めているプラットフォーム機能（いわゆる Platform as a Service）のひとつとして、データ公開・共有とデータ分析とをユーザが一体的に実施できる機能がある。様々なデータに様々な研究者がアプローチすることを可能にし、これまでになかった研究者同士の協働を生み出すことを促進する狙いがある。

まず、mdx において取集、解析されたデータを、広くあるいは目的に応じて限定的に公開・共有することができる mdx 専用のデータ公開基盤の開発を進めている。昨今、社会には様々な種類の大量のデータが溢れており、その全てが十分に活用されているとは言い難い。これは学術研究分野に限った話ではなく、産業界や政府・地方公共団体等でも同様の状態であると言え、Society 5.0 を目指すうえで大きな障害になると考えられる。こうしたデータを十分に活用していくためには、データと、そのデータを活用するノウハウを持った研究者とが効率的にマッチングされることが望ましい。そのために mdx では、自分たちが保有する様々なデータを他の mdx ユーザに安全に公開・共有し、分析してもらい、幅広く研究に活用される仕組みづくりを行っている。ユーザは自由にデータを検索、閲覧し、それぞれに設定された利用ポリシーに従って自身の研究に活用することが可能になる。またそれぞれのニーズに合わせて、共同研究を開始できるような環境も備える。あるいは、それぞれのデータの指向や機密性に合わせて、データ提供者が公開・共有範囲を指定できる仕組みも整備している。将来的には mdx 以外のデータリポジトリとも連携し、学術研究データに限らず様々なデータを活用していくための仕組みづくりを進めていく。また、mdx の計算資源を活用したデータ分析を、このデータ公開基盤から直接開始できる機能の開発も進めている。データの検索から利用開始、分析開始までのシームレスな実行を実現する。

7. 関連研究

mdx プラットフォームを IaaS つまり VM 提供レベルで見ると、関連システムとしては、Amazon、Google、IBM などの民間企業が提供するパブリッククラウドが挙げられるが、6 章で述べたように、データ保持者・データ解析者・ドメイン専門家等をマッチングさせるコミュニティプラットフォームを、国家レベルで具現化するシステムとしては mdx が先駆的と言える。

米国の XSEDE プロジェクト[16] は、米国の国立研究所が保有するスーパーコンピュータを繋げるプロジェクトであるが、主なターゲットユーザは、シミュレーションなどの計算科学者が対象である。また、2015 年に始動した Cameleon プロジェクト[21]は複数の大学内のクラウド環境を提供しているが、第 1 章で述べた mdx の設計思想で述べたような国家レベルでのプラットフォームを目指したものではない。米国では、このような現状を踏まえて、米国の CCC(Computing Community Consortium)コンソーシアムでは、2021 年 4 月にホワイトペーパーを発表し[17]、AI やデータ科学用の国家クラウドの必要性を述べている。

また、mdx は SINET を通じて、全国の機関と接続されている事もその発展性・拡張性が期待される点であり、将来的には各研究組織の計算・ストレージ資源を連携することも可能である。また、モバイル SINET が配備された今、様々な IoT デバイスからのデータ収集も可能であり、データ駆動型社会 Society 5.0 を支える基盤になると言える。

8. まとめと今後

データ活用社会創成プラットフォーム mdx は、Society 5.0 の実現に向けてデータ活用アプリケーションやデータ科学のために設計された基盤である。これまでのスーパーコンピュータと同様に、高速計算能力に優れるほか、「プラットフォームのプラットフォーム」とでも言うべき、データを活用するための基盤となり、2021 年 9 月にテストリリースを開始した[19]。本番運用を 2022 年春に予定している。それにあたって、課金モデル、ユーザコミュニティの醸成方法、企業連携方法など現在検討中である。mdx 上における更なる学際的な協業が期待される。

参考文献

[1] 田浦健次朗、データ活用社会創成プラットフォーム計画について、東京大学情報基盤センター年報第 20 号, 54-59、2019.

[2] 文部科学省、データ活用社会創成プラットフォームの推進に向けた当面の整備方策

[3] 中島研吾ら、Society 5.0 実現に向けた（計算＋データ＋学習）融合、255-257、大学 ICT 推進協議会 2019 年度、2019.

[4] 合田憲人ら、mdx:データ活用社会創成プラットフォーム、大学 ICT 推進協議会 2020 年度

[5] machine-configs, https://github.com/mdx-jp/machine-configs

[6] Red Hat, Ansible, https://www.ansible.com/

[7] A. Jain, et.al. The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 2013, 1(1), 011002.

[8] 国立研究開発法人物質・材料研究機構, DICE https://dice.nims.go.jp/

[9] AFLOW: Automatic Flow for Materials Discovery http://www.aflowlib.org/

[10] 文部科学省「マテリアル先端リサーチインフラ」2021 年 3 月採択決定 https://www.mext.go.jp/b_menu/boshu/detail/material_research_results_00001.html

[11] 東京大学 先端科学技術研究センター: 地域気象データと先端学術による戦略的社会共創拠点 https://www.climcore.org

[12] 東京大学 未来ビジョン研究センター:資源を循環させる地域イノベーションエコシステム研究拠点 https://coinext.ifi.u-tokyo.ac.jp/

[13] データ利活用社会創成シンポジウム 2021 https://sites.google.com/g.ecc.u-tokyo.ac.jp/dp-sympo2021

[14] Tananun Orawiwattanakul, et.al, User consent acquisition system for Japanese Shibboleth-based academic federation (GakuNin), International Journal of Grid and Utility Computing 2011

[15] Gakunin Web Site, https://www.gakunin.jp/

[16] XSEDE, https://www.xsede.org/

[17] Ian Foster, et.al, A National Discovery Cloud: Preparing the US for Global Competitiveness in the New Era of 21st Century Digital Transformation, 2021, A Computing Community Consortium (CCC) white paper

[18] SINET https://www.sinet.ad.jp/wadci

[19] The mdx platform: https://mdx.jp/

[20] J.Zhen, et.al, Lessons Learned from the Chameleon Testbed, In Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC '20). USENIX Association. July 2020.

[21] People Flow Project: https://pflow.csis.u-tokyo.ac.jp/