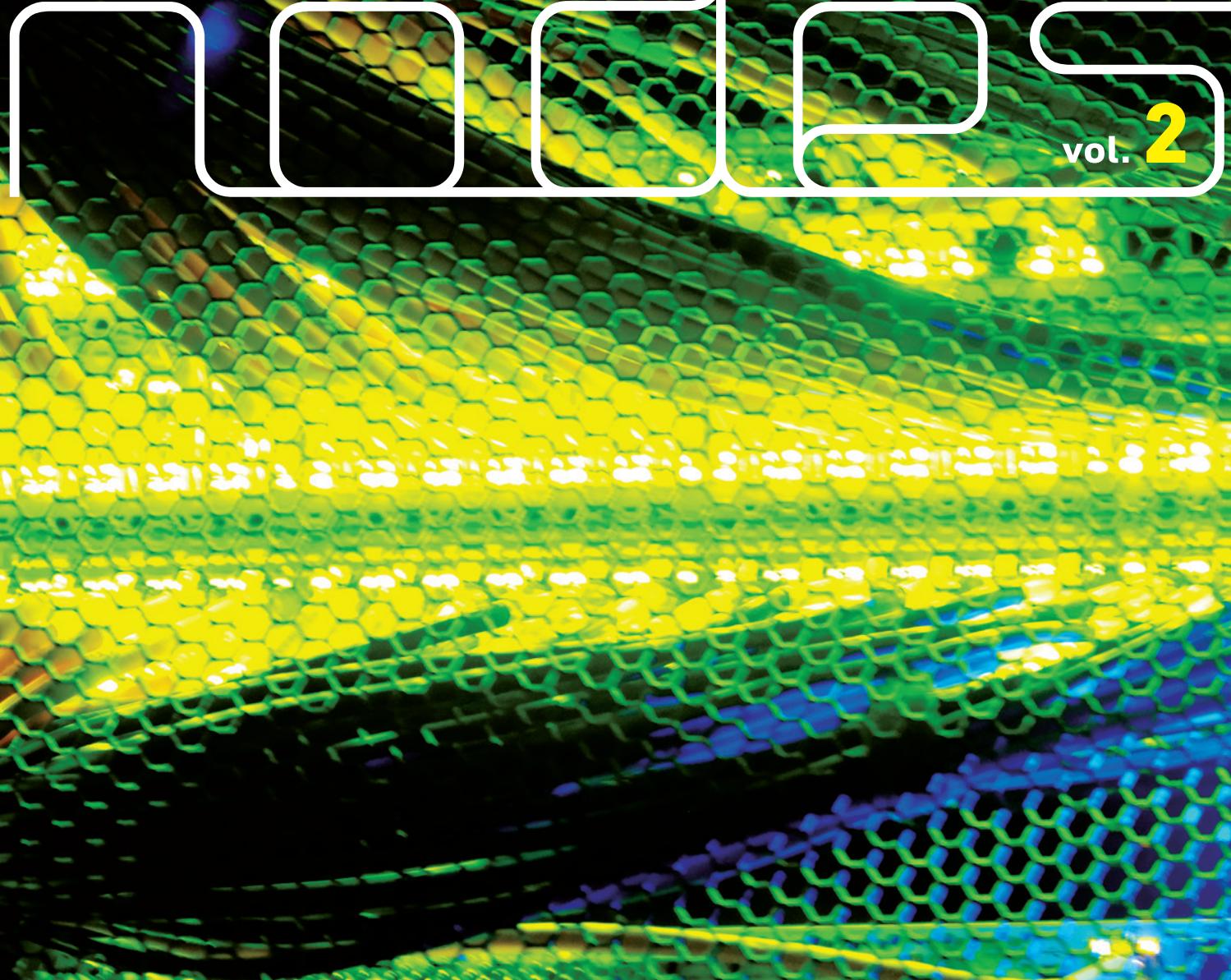


vol. 2



特集

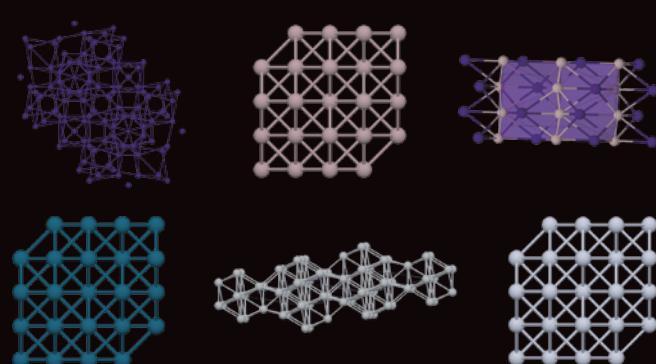
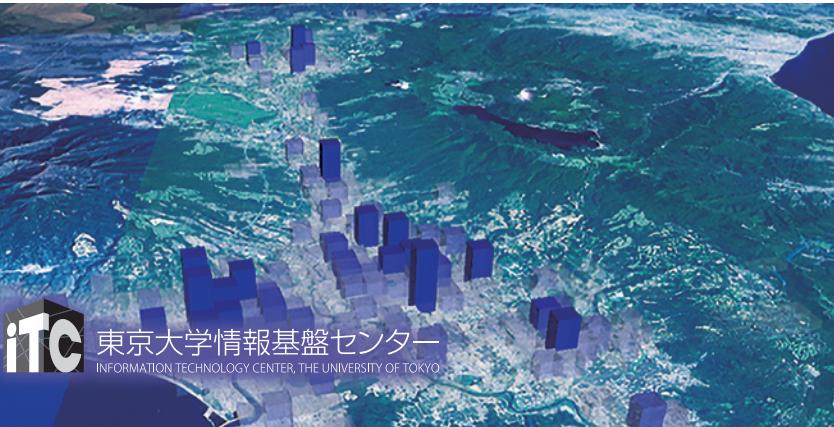
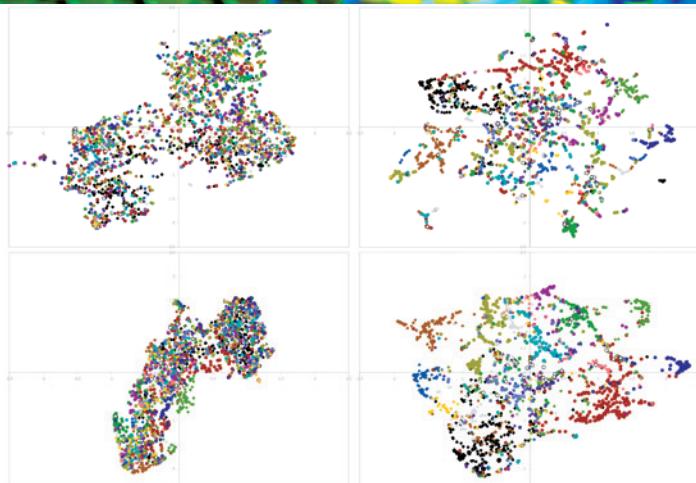
## データ活用社会創成 プラットフォームmdx始動

連載 nodesの光明

情報システム利用者の「困った!」を解決  
——教育のオンライン化に貢献する学生サポーターたち

連載 飛翔するnodes

大型データから解き明かす土星衛星の大気環境



## 卷頭言

データ利活用の波は日常生活にも押し寄せており、DX（デジタルトランスフォーメーション）やビッグデータ、AI、IoTといった用語をメディアでも頻繁に目にします。学術研究の世界では古くからデータの蓄積・利活用が行われてきていますが、莫大な情報から新たな知見を引き出すカッティング・エッジなデータサイエンスの手法を導入することで、無数の研究領域がさらにブーストされる時代になっています。今号の**nodes**では、データ科学・データ利活用にフォーカスした計算機資源である「mdx」を特集します。その設計・運用に携わる計算機・データサイエンスの専門家から、mdx を用いた応用研究に取り組む各分野の専門家まで、幅広いインタビューから mdx が切り拓く学術の未来をお届けします。

加えて、今号では当センターの関わる多数の“**nodes**”の声を拾い上げています。「**nodes**の光明」では、新型コロナ禍で飛躍的に重要性が増したオンライン講義等について、そのサポートという重要な役割を主体的かつ意欲的に担ってくれている学生の皆さんへの思いや活躍をフィーチャーしています。他の連載も、研究・運営などさまざまな立場の方々の活躍を垣間見ることのできる内容になっています。

号を重ねるごとに、重層的な“**nodes**”の姿をお伝えしていきます。ぜひお楽しみください。  
(飯野孝浩)

## Contents

- 03 特集  
データ活用社会創成プラットフォーム mdx 始動
- 10 連載 **nodes**の光明  
情報システム利用者の「困った！」を解決  
——教育のオンライン化に貢献する学生ソーターたち
- 12 連載 飛翔する**nodes**  
大型データから解き明かす土星衛星の大気環境
- 14 **nodes**のひろがり

編集・発行	東京大学情報基盤センター
編集委員長	飯野孝浩
編集委員	関谷貴之、川瀬純也、中村 遼、下川辺隆史、有馬和美、大林由尚
編集・執筆協力	サイテック・コミュニケーションズ
撮影	盛 孝大
デザイン	日向麻梨子(オフィスヒューガ)
印刷	株式会社 DPP アークス

【扉写真】mdx (部分)



日本社会の未来の形「Society 5.0」。

その実現には、現実世界から集めたデータをいかに活用するかが鍵となります。

そこで、さまざまな分野の研究者が「データ活用」を核として集い、  
研究する場となる「データ活用社会創成プラットフォームmdx」が

構築され、運用が始まりました。

どのようなコンセプトで構築され、  
どのような研究が行われているのでしょうか。  
この特集でご紹介します。

# データ活用社会創成 プラットフォームmdx始動

特集



# 幅広い分野の 研究者ニーズを満たす データ活用社会 創成プラットフォーム mdx

2021年9月、「データ活用社会創成プラットフォーム mdx」の試験運用が始まりました。

mdx は東京大学柏IIキャンパスに設置されており、

東大を含む9大学と2研究所<sup>\*1</sup>が共同運営しています。

世界初の「産学官連携のためのプラットフォーム」として mdx が目指すのは、「誰もがデータを利用できる計算環境を提供すること」。

そのハードウェアの構築を担当した東大情報基盤センターの塙 敏博教授と、サービス設計・運用に関わる大学院情報理工学系研究科の鈴村豊太郎教授に、mdx の特徴や現在の利活用のされ方、

期待されるアカデミアへのインパクトについてうかがいました。

東京大学  
大学院情報理工学系研究科 教授

## 鈴村 豊太郎

専門は人工知能、大規模グラフ深層学習。東京工業大学大学院情報理工学研究科博士課程修了、博士（理学）。IBM Research に17年間勤務し、MIT-IBM Watson AI Lab プロジェクト代表も務めた。2021年より現職。情報基盤センター教授を兼任。

東京大学情報基盤センター 教授

## 塙 敏博

専門は高性能計算システム。慶應義塾大学大学院理工学研究科博士課程修了、博士（工学）。筑波大学大学院システム情報工学研究科准教授、東京大学情報基盤センター准教授などを経て、2020年より現職。

Society 5.0 時代に求められる  
データ活用プラットフォームを目指して

「Society 5.0」は、サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムによって、経済発展と社会的課題の解決を両立する、人間中心の社会のことです。2016年に国が策定した「第5期科学技術基本計画」<sup>\*2</sup>の中で日本の未来社会のコンセプトとして提唱され、以来、アカデミアと産業界は Society 5.0 の実現に向けた研究開発を加速しています。東京大学も、今後アカデミアがどのような計算資源を必要とするかを議論し、その答えの1つとして、データ利活用の推進を支援する「データ活用社会創成プラットフォーム mdx」を、国立情報学研究所（NII）との密接な連携のもとで構築しました。そして、2021年9月、試験運用が始ま

まつたのです。

mdx は、計算資源、ストレージ、ネットワークからなっており、構成要素はスパコンと同じです。しかし、「スパコンが大規模計算を行うために大きな計算資源を提供することを主目的に構成されているのに対し、mdx は十分な計算資源と同時に、一人ひとりのユーザが研究しやすい環境を提供する点がまったく違います」と、塙教授は言います。これまでスパコンを使っていなかった人を含め、データを利用する幅広い分野の研究者に使ってもらうことを目指しているのです。

mdx を利用できるのは、おもにアカデミアの研究者です。プロジェクトを申請し、承認されると、計算資源やストレージ、ネットワーク構成が割り当てられます。そこに、OS やソフトウェアをインストールすれば、プロジェクトに合った仮想環境を短時間で構築できます。そして、この仮想環境を用いることで、データを収集・保存したり、解析を行ったり、その結果を公開したりといった研究活動を mdx 内で完結できます。

このように、ユーザにとって使い勝手のよいプラットフォームであることが、mdx の最大の特徴です。さらに、「スパコンでの計算やデータの利用に慣れていない研究者がプロジェクトを申請する際には、私たちが相談に応じています」と塙教授。使用中のサポート体制も充実しており、ユーザのハードルはぐっと低くなっているようです。

### 研究者ならではのニーズに応える3つの特徴

mdx は、研究者用のプラットフォームという性格から、使いやすさに加えて以下の3つの特徴を備えています（図1）。

1つめは、研究者ごとの仮想環境がセキュアに保たれていることです。研究者が扱うデータには、個人情報や、企業の技術情報など秘匿性の高いものが多くありますが、mdx ではデータが漏れる心配をせずに研究することができます。さらに、外部からセンサーなどを使ってデータを収集する際にも、その通信までセキュリティが担保されています（後述）。

2つめは、スパコン並みの計算資源を備えており、データ科学と計算科学の手法を融合した研究が行えることです。

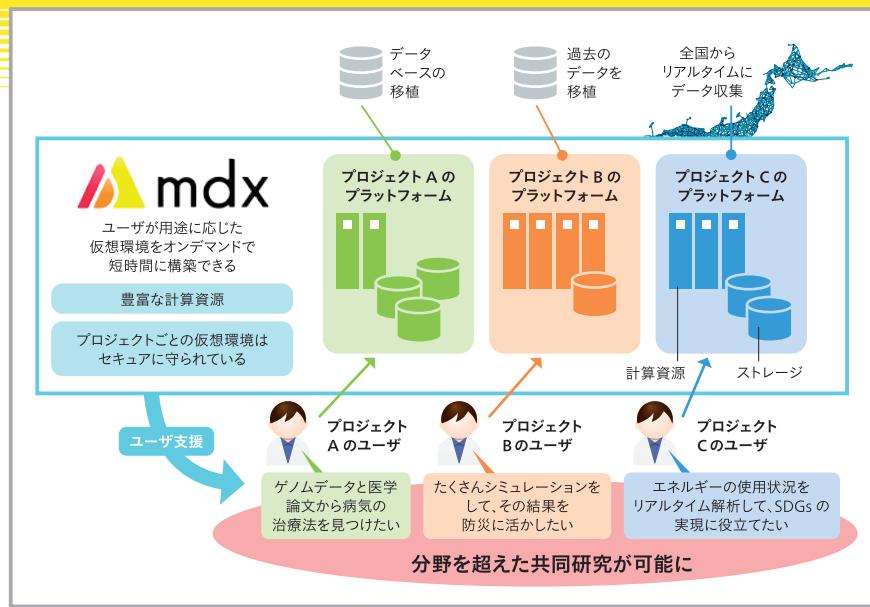


図1 mdxの特徴

mdx は、アカデミアにおけるデータ科学・データ活用に特化したデータプラットフォームである。ユーザが研究内容に合った構成のセキュアな仮想環境を構築でき、リアルタイムでデータを収集し、解析することもできる。この特徴を生かし、医療、防災、エネルギー、農水産業などへのデータ利活用研究が行われることが期待される。

3つめは、研究者間の連携を促す場となることです。各研究者の仮想環境はセキュアに保たれていますが、必要に応じてデータを共有し、利活用するための機能も提供されます。これにより、研究者は mdx 上で異分野の研究者と学際的な共同研究を展開することができます。

### 柔軟に構築できるハードウェアと仮想環境による高いセキュリティ

このような mdx の特徴は、どのようにして実現されたのでしょうか（図2）。「mdx の使い方はユーザに全面的に委ねられますから、ユーザが柔軟に仮想環境を構築できるようなハードウェア構成と、セキュリティの確保に特に気をつかいました」と塙教授は説明します。

まず、計算資源としては、最新の CPU を多数備えている他、ディープラーニングのアクセラレータなどとして使われる GPU も搭載しています。ストレージは、「大容量 HDD ストレージ」と「高速 NVMe ストレージ」の他、研究で得られた成果やデータベースを外部に公開することを想定して「外部共有オブジェクトストレージ」も設けられています。

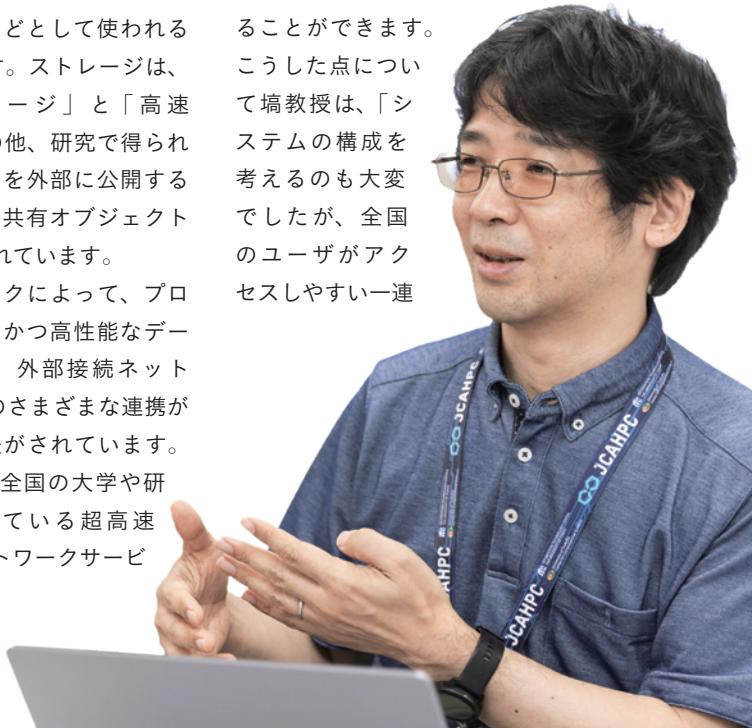
内部高速ネットワークによって、プロジェクト内でセキュアかつ高性能なデータ処理が行える一方、外部接続ネットワークは、SINET6 とのさまざまな連携が実現できるように工夫がされています。SINET6 は、NII が日本全国の大学や研究機関向けに提供している超高速（400Gbps）の学術ネットワークサービ

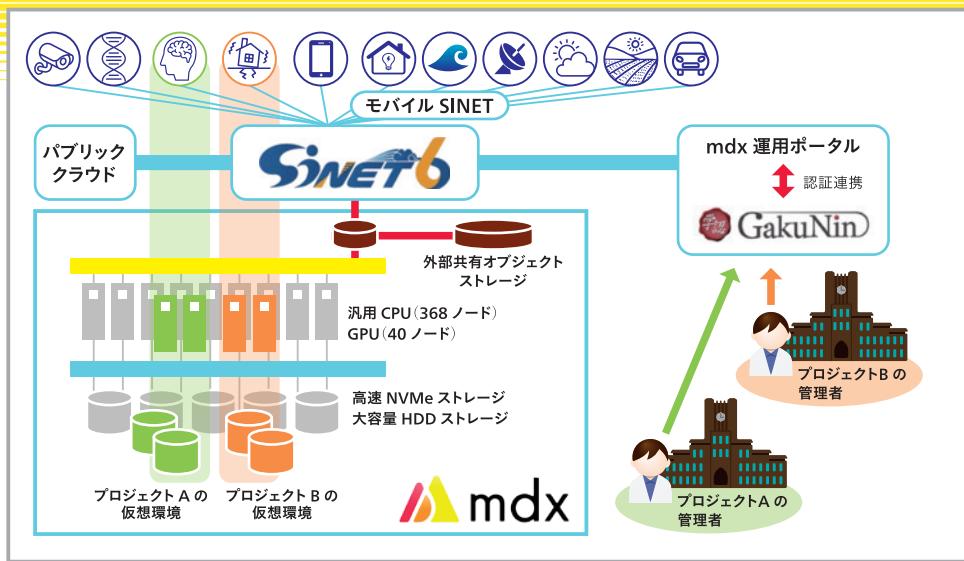
スです。SINET6 自体が一般的のインターネットとは切り離されており、セキュリティが高いのですが、さらに、内部と外部接続のネットワークを分けたことにより、万一外部から侵入を受けても内部のストレージなどを守ることができます。

また、SINET6 は、モバイル SINET（携帯キャリアの基盤を利用した閉じたネットワーク）と接続しているため、外部のセンサーヤや遠隔地の装置等からデータをセキュアに収集することができます。プロジェクトごとの仮想環境は、モバイル SINET まで含めて完全に分離されており、秘匿性の高いデータを扱う研究も安心して実施できます。

一方、ユーザは NII が運用する「Gaku Nin」<sup>※3</sup> というシステムを経由することで、mdx を容易に利用す

ることができます。こうした点について塙教授は、「システムの構成を考えるのも大変でしたが、全国のユーザがアクセスしやすい一連





#### ▶ 深く学ぶには

「データ活用社会創成  
プラットフォーム mdx」ウェブサイト  
URL: <https://mdx.jp/>



図2 mdxの特徴 mdxのハードウェアの構成と利用イメージ

ユーザが使いたいときすぐにアクセスできる一方で、プロジェクトごとの仮想環境は高度にセキュリティが守られている。

の仕組みをつくり込んだり、運用のやり方を考えるのはさらに大変でした」と振り返ります。

#### 試験運用に参加する70余の課題と今後期待される学際的研究分野の誕生

鈴村教授は、2022年6月30日時点のmdxの利活用状況について「プロジェクトは試験運用開始以来、徐々に増え、現在70以上の課題が動いています。計算資源の使用状況は時々刻々変動しますが、GPUは使用率が90%に迫った時期があるほど大いに活用いただいている」と話します。

順調に試験運用が進むmdxですが、利便性向上のためにさらなる機能の拡張が計画されています。1つは、NIIが運用する研究データ管理基盤サービス「GakuNin RDM<sup>※4</sup>」との連携です。近年、研究者は研究データの適切な管理を強く求め

られます。研究データを常時、収集し構造的に管理し、必要なときに取り出すことのできるGakuNin RDMと連携することで、より研究しやすい環境が整えられます。また、ハードウェア

の面では、mdxを共同運営する11機関が所有するスパコンと接続し、計算資源の強化を図ることが考えられています。さらに、運用を円滑に続けるための適切な課金制度も検討されています。

新サービスとして模索を続けるmdxについて、鈴村教授は「欧米でも、商用ではない“みんなが使えるデータプラットフォーム”的必要性が言われていますが、国レベルでmdxのようなシステムをつくったのは日本が世界で初めてです。日本独自のシステムとして十分に活用していただきたいです」と言います。

柔軟な使い方ができるmdxには、期待通りさまざまな分野のプロジェクトが集まっています（表1）、日本でのどのようなデータ利活用研究が行われているかの全体像も見えてきそうです。そして、その先に

は、これまで思いもよらなかった分野間の学際的な研究が生まれることが期待されています。「材料科学と情報科学の学際研究が始まり、新素材を発見するための手法を構築しようという研究もすでに始まっています」と鈴村教授。塙教授とともに「日本ならではの新しい研究分野の誕生に積極的に関わっていきたい」と決意を新たにしています。

（取材・構成 池田亜希子）

※1 北海道大学、東北大大学、筑波大学、東京大学、東京工業大学、名古屋大学、京都大学、大阪大学、九州大学、国立情報学研究所、産業技術総合研究所

※2 詳細は内閣府ウェブサイト参照 <https://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>

※3 NIIが全国の大学と連携して運営している学術認証フェデレーション。詳細はNIIウェブサイト参照 <https://www.gakunin.jp/>

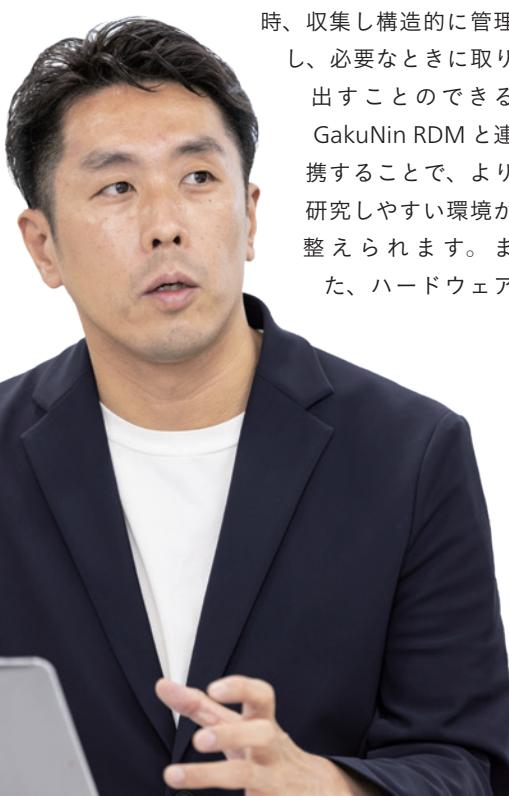
※4 詳細はNIIウェブサイト参照 <https://rcos.nii.ac.jp/service/rdm/>

※5 詳細はJHPCNウェブサイト参照 <https://jhpcn.kyoten.itc.u-tokyo.ac.jp/ja/>

表1 mdxで実行中の研究課題の例

実行中の70余の課題のうち、学際大規模情報基盤共同利用・共同研究拠点（JHPCN）<sup>※5</sup>で2022年4月に採択された15課題を例としてあげた（うち3課題を次ページ以降で紹介する）。

研究課題名	代表者	所属
ハイブリッドクラウドを用いたゲノム情報に基づく構造多型パネルの構築とアノテーション	長崎正朗	京都大学
超高解像度の即時予測の実現に向けた都市街区間風況データベースの構築	小野寺直幸	日本原子力研究開発機構
Distcloudの拡張、継続性の向上、およびその応用研究	柏崎礼生	国立情報学研究所
次世代演算加速装置とそのファイルIOに関する研究	塙 敏博	東京大学
合成人口プロジェクト：mdxによる合成人口データベースの構築	村田忠彦	関西大学
統合機械学習分子動力学システムの構築	奥村雅彦	日本原子力研究開発機構
次世代学術情報基盤に向けた基盤ソフトウェアの実践的な研究・開発・評価	杉木章義	北海道大学
財務ビッグデータの可視化と統計モデリング	地道正行	関西学院大学
医療・介護領域の人材マッチングに最適化された大規模グラフニューラルネットワーク	鈴村豊太郎	東京大学
エージェントモデルと統計データを用いた全国規模の疑似人流データの開発	関本義秀	東京大学
大規模な日本語モデル構築・共有のためのプラットフォームの形成	相澤彰子	国立情報学研究所
グラフニューラルネットワークとマルチタスク学習による汎用的物性予測モデルの構築	華井雅俊	東京大学
ビヨンド・“ゼロカーボン”を目指し地域と技術をつなぐ情報基盤の構築	兼松祐一郎	東京大学
単語間に区切りのない手書き言語における係り受け解析エンジンの開発	安岡孝一	京都大学
多次元高精細地表情報（MHESD）の地球科学・歴史考古学における高度利活用	早川裕式	北海道大学





# ゼロカーボンを目指す地域と技術をつなぐ情報基盤の構築

東京大学総括プロジェクト機構「プラチナ社会」総括寄付講座 特任講師

兼松祐一郎

## 深く学ぶには

「ビヨンド・“ゼロカーボン”を目指す“Co-JUNKAN”プラットフォーム研究拠点」は、JST共創の場形成支援プログラムの一拠点(JPMJP2003)。技・知・人を産学公で循環させるプラットフォームを実装し、地域がゼロカーボンの先を見据えて持続的発展を目指せる社会の実現につなげる。

**URL** <https://cojunkan.ifi.u-tokyo.ac.jp/index.html>

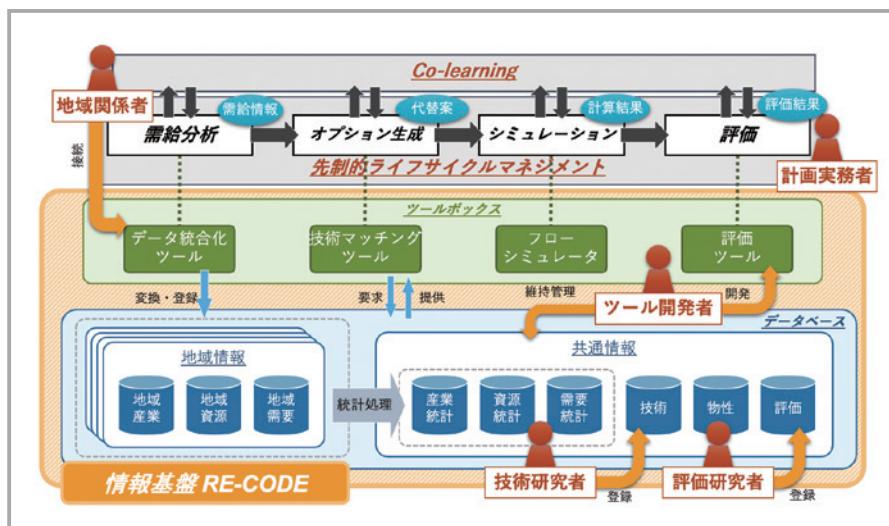


図 地域と技術を繋ぎ実装を加速する情報基盤（RE-CODE）の構想

## 自治体の悩み

気候変動への対応が急がれる中、日本でも総人口の9割近くにあたる自治体が「2050年までに二酸化炭素排出実質ゼロ」を表明しています。しかし、実現のための具体的な施策や計画を策定している自治体は多くありません。その大きな理由として、新しい技術や仕組みの導入が必要なことはわかっていても、技術へのアクセスや評価が難しいことがあります。そこで私たちは、技術導入を支援する情報基盤を構築しようとしました。まだ、開発は始まったばかりですが、その構想と研究の現状をご紹介しましょう。

## 必要なツールとデータを1カ所に

この情報基盤「RE-CODE」は、自治体にとっては地域の実情に合った技術を発見・評価できる場、企業にとっては技術を普及させる場となることを目指しています。そのため、①需給分析、②オプション生成、③シミュレーション、④評価のそ

れぞれを行うツールを搭載するとともに、必要なデータを集積する計画です（図）。

①は、地域資源の需給状況の把握と将来予測をする段階です。ここではオープンデータを活用できますが、各省庁や地域に分散していて探すのも集めるのもたいへんです。そこで、データを統合・可視化して一目で把握できるツールをつくりました。すでに一部のデータは公開しており、データの拡充も進めています。

この需給分析で、例えば、「この地域はバイオマス資源に強みがある」とわかったとしましょう。すると次は、バイオマスをエネルギーや製品に変換するための技術オプションを探す②の段階になります。これを支援するのが技術マッチングツールです。

③のシミュレーションでは、②で見つけた各技術を導入したときの物質とエネルギーのフローを推算します。それを計算するためのツールとして、さまざまなシミュレータを開発しています。

最後の④は、ライフサイクル評価や社会経済性評価です。③で得たフローの推算結

果と既存の評価用データベースを組み合わせることで、各技術が環境と経済に与える影響を評価します。

## mdxの特徴が生きる研究課題

まずは、RE-CODEの一部をmdxに移植し、その後、ツールやデータを拡充していくきます。mdxは、RE-CODEのコンセプトと相性のよいプラットフォームだと思います。例えば、②の技術マッチングと③のシミュレーションには地域や企業から提供いただく情報も使いますが、そのとき、mdxでデータを秘匿化しながら間接的に参照する仕組みとすることで、安心してデータ提供に協力していただけると思います。また、シミュレーションには大量の計算が必要なので、mdxの豊富な計算資源にも期待しています。

このような情報基盤を実現したいという思いは数年前から抱いていましたが、2022年度から東京大学で始めた「ビヨンド・“ゼロカーボン”を目指す“Co-JUNKAN”プラットフォーム研究拠点」というプロジェクト内の研究開発課題の1つとなったことで、本格的に研究できるようになり、張り切っているところです。プロジェクト内の他のチームが行っている技術の実証実験や、人材育成などの課題とも協働することで、RE-CODEを地域と技術をつなぐ場として発展させていきたいと思っています。

(取材・構成 青山聖子)



兼松祐一郎／専門はプロセスシステム工学、ライフサイクル工学。東京大学大学院工学系研究科修士課程修了。博士（工学）。株式会社菱化システムに勤務後、東京大学総括プロジェクト機構「プラチナ社会」総括寄付講座 学術支援専門職員、特任助教を経て、2022年より現職。



▶深く学ぶには

「人の流れプロジェクト」

URL <https://pflow.csis.u-tokyo.ac.jp/>

# オープンデータから 日本全国の人流をつくりだす

東京大学デジタル空間社会連携研究機構 機構長  
東京大学空間情報科学研究センター 副センター長・教授

関本義秀

## 早くから「人流」の重要性に着目

新型コロナウイルス感染症との関連で「人流」という言葉が広く知られるようになりました。しかし、私たちが「人の流れプロジェクト」を始めたのは、それよりずっと前の2008年です。人々が時々刻々移動していく有様をとらえた人流データは、防災や防犯、マーケティング、交通・都市計画などに役立つだけでなく、さまざまな分野の研究者にとって重要なデータインフラとなります。そのため、人流データを作成・提供するプロジェクトを開始したのです。

これまでおもに、パーソントリップ(PT)調査(国や地方自治体などが行う交通アンケート調査)のデータを処理することで人流を再現してきました。この人流データは共同研究を通じて提供しており、その件数は400件近くに上ります。

しかし、PT調査は大都市圏などでしか行われておらず、日本全国の人流を描き出すことはできません。一方、携帯キャリアとの共同研究で、スマホの位置データから人流を得る試みもしましたが、位置データ

は個人情報のため、得られた人流データを他の研究者に提供するわけにはいきません。こうしたフラストレーションを抱える中で、「オープンデータを利用することで、日本全国の人流を擬似的につくりだしたい」と考えるようになり、mdxを使って実行しました。

## オープンデータから 必要な情報をつくりだす

擬似人流をつくりだすために、「エージェントモデル」という方法を使いました。大まかな手順は以下の通りです(図)。

①まず、コンピュータの中に日本の総人口(約1億3000万人)分の仮想的な人間(エージェント)をつくり、各人の年代や性別、家の場所を決めます。②次に、各人の通勤・通学の有無と平日の行動パターン(例えば、通勤者の場合、「家→勤務先」「家→勤務先→第3の場所」など)を決めます。③さらに、各人の具体的な行き先を決めます。④その上で、行き先までの距離、最寄り駅へのアクセスなどから交通手段を選び、各人の1分ごとの位置を割り出していくします。これをすべてのエージェントについて

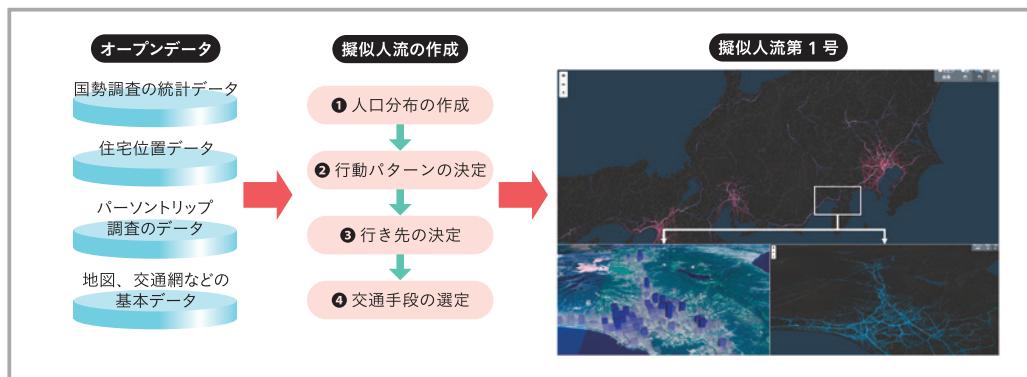
集めれば、日本全国の擬似人流を描きだせるというわけです。

個人情報が含まれていないオープンデータから、どうやって各人の属性や行動を決めたのでしょうか。私たちは、使えそうなデータを探し集め、その処理方法を徹底的に工夫しました。例えば、国勢調査には地区ごとの年齢・性別構成の統計がある一方、地図情報会社が提供している全国の住宅位置データ<sup>※1</sup>があります。これらを組み合わせることで、どの家にどの年代・性別の人人が住んでいるかを推定しました。また、私たちはPT調査から人流を導いた際に、人々の行動パターンを分類し、それぞれの割合を明らかにしていました。②の段階では、このデータを用いて各人の行動パターンを確率的に決めました。こうしたことの積み重ねで、かなり現実に近い属性の1億3000万のエージェントをコンピュータの中につくりだし、行動させることができたのです。

手順のうち④の段階は多くの計算資源を必要とするのですが、mdxを使ったことで順調に計算でき、2021年度に擬似人流第1号ができあがりました。この擬似人流は、大きなメッシュで見れば、スマホの位置データからの人流と比べても遜色ないもので、すでに研究者へのデータ提供も開始しています。今後もmdxを使い、②の行動パターンをきめ細かくするなどして、より精度を高めていきたいと考えています。

(取材・構成 青山聖子)

※1 有償だが安価であり、誰でも利用できる。



関本義秀／専門は人間都市情報学。東京大学大学院工学系研究科博士課程修了、博士(工学)。国土交通省国土技術政策総合研究所研究官、東京大学空間情報科学研究センター特任准教授などを経て、2020年より同センター教授、2022年より副センター長。2021年より東京大学デジタル空間社会連携研究機構機構長。



# 医学分野の論文から日本語言語モデルを構築する

国立情報学研究所コンテンツ科学研究系 教授

相澤彰子

▶深く学ぶには

情報処理学会研究報告

「学術分野に特化した事前学習済み

日本言語モデルの構築」

URL <https://bit.ly/3KrMpNO>

## 言語モデルとは

言語モデルは、膨大な量の文書に含まれている情報をなるべく効率よく圧縮して、いろいろな処理で活用できるようにしたもの。最近、文書の自動要約、対話システム、機械翻訳などの性能が格段に向かっていますが、その背景にはここ数年で言語モデルが急激に発展したことがあります。

実際、最近の巨大言語モデルには、文脈を考慮した単語の意味、文法、単語と単語の関係（日本の首都は東京）など、さまざまな知識が埋め込まれています。ただし、そのような巨大言語モデルをつくるには、莫大な計算資源を使って大量の文書を学習させる必要があるため、誰もが簡単に構築できるわけではありません。

## なぜ日本言語モデルをつくるのか

巨大言語モデルが1つあれば、他の言語モデルはいらないのではないかと思われ

るかもしれません、やはり日本語には日本語の文書で学習させたモデルが必要です。しかし、英語では新しい巨大言語モデルが続々と発表されるのに対して、日本語のモデルは手薄だという現状があります。さらに、例えば門外漢である私が医学論文を読んでも知らない単語ばかりで意味がわからないのと同じで、モデルも学習していない新しい分野は得意ではなく、その分野に合ったモデルを構築する必要があります。

こうした背景から、私は mdx のプロジェクトで、日本語の医学分野の言語モデルを一からつくることに取り組んでいます。実は、カルテを教材とした日本語の言語モデルはすでに存在し、診療データから病名を抽出するといった用途が考えられています。しかし、論文を対象とした言語モデルはまだありません。日本語の症例報告論文は知識として価値が高いのにこれは残念なことです。このため、私の言語モデルは、日本語の医学論文を教材として学習させることにしました。論文で学習させたモデルがあれば、論文を医学の分野ごとに分類したり、論文から病名を抽出したりすることが可能になり、医学研究に役立つと考えたのです。

## mdxの利用で スムーズに研究が進んだ

学習に用いたのは、おもに論文の抄録で分量は約 1160 万文です。学習させる前の単語列の処理の仕方や、専門用語の辞書の組み込み方を変えて学習させ、最終的に 8 個のモデルができあがりました。しかし、言語モデルは「つくれば終わり」ではなく、公開して使ってもらう前に性能をテストしなければなりません。

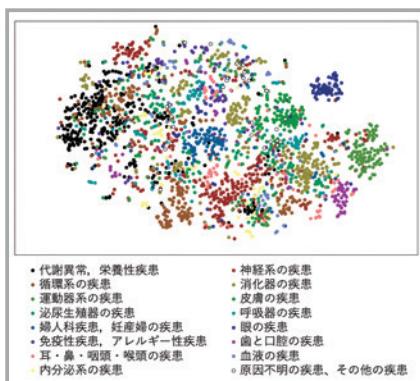


図 今回作成した言語モデルのテスト結果の例

言語モデルが単語を病気の種類の違いで分類した結果を可視化したもの。点は単語を表し、色の違いは病気の違いを表す。同じ色の点が集まっているかどうかで、言語モデルの性能がわかる。

テストをするには、解きたい問題の「正解データ」を用意しておき、モデルに問題を解かせて正答率を見ます。しかも、タイプの異なる複数の問題を用いる必要があります（日本語のモデルを発展させるには、こうしたリソースづくりも重要なことです）。現在はそのデータを使ってモデルをテストしており、成績がよかつたものを 2022 年度中に公開できればと思っています。

言語モデルを一から構築するには、大量の文書を学習させるプロセスを、条件を変えて何度も繰り返す必要があるため、割り当てられた計算資源を自由に使える mdx がとても有効でした。まだテストの途中ですが、医学論文の解析に適したモデルができたのではと手応えを感じています。

また、ここでは私の研究だけをご紹介しましたが、同じプロジェクト内では早稲田大学の河原大輔教授も mdx を使って汎用型日本語モデルの構築を進めていますし、他のグループも mdx の利用に意欲を示しています。メンバーが mdx 上に集うことで、ノウハウや、モデルの構築・評価に必要なリソースの情報の共有が進むことも、mdx の効果として期待しています。

（取材・構成 青山聖子）



# 情報システム利用者の「困った！」を解決 教育のオンライン化に貢献する 学生サポーターたち

東京大学は、コロナ禍前からオンライン授業やウェブ会議などのための情報システムを整備してきました。「utelecon」は、それに関する情報を一元的に提供するポータルサイトで、2020年3月に開設されました。uteleconのサポート窓口には、学生や教職員からさまざまな問い合わせが寄せられますが、その対応の主軸となっているのは学生サポーターの皆さんです。指導にあたる情報システム本部の玉造潤史准教授、技術補佐員で元学生サポーターの竹内朗さん、学生サポーターの上野爽さんと徳永紗英さんに活動の実際をうかがいました。



学生のサポート活動を見守ることも、  
教育の一環だと考えています



東京大学情報システム本部  
副本部長・准教授

玉造潤史

Q utelecon での玉造先生のお立場は？

A 学生サポーターの皆さんには、次々にやりたいことを提案してくれます。それを思う存分できるように自主性を尊重しつつ大学側の責任者として支援しています。

Q サポート業務の運営上、心がけていることは？

A 教育の一環ととらえ、学生サポーターの皆さんに、正しいIT知識を身につけるだけでなく、社会に役立つ経験を積み、皆で協力して成果をあげていなくなまで成長できるように心がけています。例え

ば、徳永さんには、大学とZoom社の定期ミーティングに参加して意見交換をしてもらっています。

Q 学生さんの活動への感想は？

A コロナ禍でも、Slack（メッセージングアプリ）を活用してすばらしい成果を上げています。担当者の1人が、「こんな相談が来て困っている」と発信すると、瞬時にほかの担当者が反応し、ときには50件近いチャットをやりとりしてあっという間に解決するのです。「これがデジタルネイティブの問題解決方法なのか」と感心します。自主性に任せると、学生はしっかりと考えてくれますね。上野さんをはじめ、要所要所で大学の判断や許可、情報提供を求める連絡をくれるなど、みんな責任感をもって仕事に取り組んでいます。

中学校や高校でも  
生徒サポーターが活躍できるはずです



東京大学情報システム本部  
技術補佐員  
竹内朗

Q utelecon立ち上げ時に学生サポーターとして参加したのですね。それはなぜですか？

A uteleconができる前から、ECCSシステム相談員として支援活動をしていました。これは、キャンパス内にある共用のパソコン室で、利用者の相談に応じるコンシェルジュのような仕事で、学内のシステムに関するノウハウもウェブサイトで発信していました。自分が相談員になったのも、大学入学時にこのサイトで役に立つ情報を見たことがきっかけでした。コロナ禍で、私も含め多くの相談員が utelecon のサポーターになりました。

Q 職員になって感じていることは？

A 大学のシステムを整備する立場になってみて、「学生がシステムの運営に参加することが、大学の教育・研究の発展につながる」という学生サポーターの意義を改めて実感しました。中学校や高校にも、国のGIGAスクール構想などでICT教育が導入されているので、似た取り組みができるでしょう。学校で学ぶ主体である学生や生徒がサポーターとして運用に参加すれば、ICT教育の効果は桁違いに高くなると思います。

Facts

## uteleconの学生センター

サポート窓口は、学生や教職員が東大の情報システムを使うときの困りごとを、ワンストップで解決することを目指しています。ここでは、2022年8月現在、質問にメールやZoomで対応する「コモンセンター」が37名活動中で、文系・理系を問わず、学部1年生から博士課程まで幅広い学年の学生が参加しています。これまでに、のべ88名が活動し、約4000件の問い合わせに対応しました。また、システムに関する情報を発信する「utelecon サポート」など、ほかの学生センターも活動しています。

The screenshot shows the utelecon support window. At the top, there's a header with the utelecon logo and links for Google Translate, TOP, About, and English. Below the header, a sidebar titled "サポート窓口" lists sections like "目次", "注意事項", "チャットで相談する", "Zoomで相談する", and "メールフォームで相談する". The main area shows a conversation between a user and a chatbot. The user asks "こんな時は、何で相談窓口を使います?", and the bot responds with "Hello! How can we help you?". There's also a QR code and some text about the operator's availability ("平日 Weekdays 8:30~17:00").

### utelecon サポート窓口

**URL** <https://utelecon.adm.u-tokyo.ac.jp/support/>  
問い合わせには、まずチャットボットが自動応答する。それでは解決しない場合に、学生センターが対応する。

充実したサポートがあれば  
未知のシステムもすんなり活用できます



東京大学大学院工学系研究科  
修士課程2年／コモンセンター

上野 爽

#### Q どんな問い合わせが多いですか？

上野：「なぜか、急にログインできなくなったり」といった相談は多いですね。何かのきっかけで特殊な条件が揃ってしまい、ログインできなくなるのです。相談している方も自分の状況を説明できないので、こちらが質問を重ねて推測していきます。

徳永：多くの相談に対応する中で、エラーメッセージから状況を推測できるようになってきたため、私たちは「エスパー」だと自称しています（笑）。チャットで話し合いながら、状況を確認し、解決方法を探します。

#### Q サポートで心がけていることは？

徳永：春は、新入生からの問い合わせが増えます。上野さんがある新入生に対して回答を作成された際、問題の解決方法を説明する前に「ご進学おめでとうございます」という一言を添えていました。私が素敵だなと思う回答です。正しい方法を伝えるのはもちろんですが、こういう心配りも大切だと思っています。入学時に、いきなりシステムの設定でつ

まずいたら、きっと不安で暗い気持ちになってしまうでしょう。

上野：相手の理解度に合わせた伝え方も大切にしています。例えば、東大にはメールシステムが複数あるため、名称を間違って質問してくれる方もいます。そういうときも、相手の意図を想像し、説明の難易度を相手に合わせてコミュニケーションします。

徳永：オンライン講義をする先生から「学生の意見を集めて議論するためのツールを紹介してほしい」という問い合わせがあったときには、よりよいツールを選べるように、学生センターの経験や調査から、複数のツールの機能比較をして回答しました。

上野：回答するために、よくシステムの公式HPを参照します。ただ、日本語訳の情報は古い場合もありますから、英語の公式ページも参照します。それでも情報がない場合には、自分の手元で問い合わせの状況を再現して、いろいろ試し、解決法をお伝えする場合もあります。なんとかして解決の糸口を探すようにしています。

徳永：私はZoom社に直接問い合わせることもよくあります。何度も問い合わせている

うちに、玉造先生のはからいで、Zoom社との定例ミーティングにまで参加するようになりました。Zoomはおもに企業で使われてきたツールで、大学での使い方を想定していないところもあるため、大学に適したZoomの機能改善を提案したり、バグを報告したりしています。

#### Q 学生センターをしていてよかったことは何ですか？

上野・徳永：私たちが対応したことで「とても助かりました」と言っていただき、システムを活用していただけることです。私たちの活動が大学の一助となっているなら、とてもうれしいです。



東京大学教養学部4年／

コモンセンター・utelecon サポート

徳永紗英

遊ぶように夢中になって  
活動するうちに、大学や企業を  
動かしていました

#### ▶ 深く学ぶには

uteleconの活動は以下で紹介されている。

##### ▶ 学内広報 no.1562

「utelecon のセンターをご存じですか？」

**URL** <https://bit.ly/3L5eWca>



##### ▶ nodes vol.1 「全学授業オンライン化で『ユーザー目線の大切さ』を痛感」

**URL** <https://bit.ly/3QWRDU1>



##### ▶ Business Insider「なぜ、東大はどこよりも早くオンライン化できたのか？」

**URL** <https://www.businessinsider.jp/post-213860>

#### 取材を終えて

すばらしいサポートができるのは、東大生の皆さんのが優秀だからだという一面もあると思います。しかし、他大学、高校、中学にもデジタルツールを楽しげに使いこなし友人に教えているセンター候補は大勢いるでしょう。本記事が「情報システムの運用に教育を受ける側も参加する」という取り組みが広がるきっかけになればと願っています。

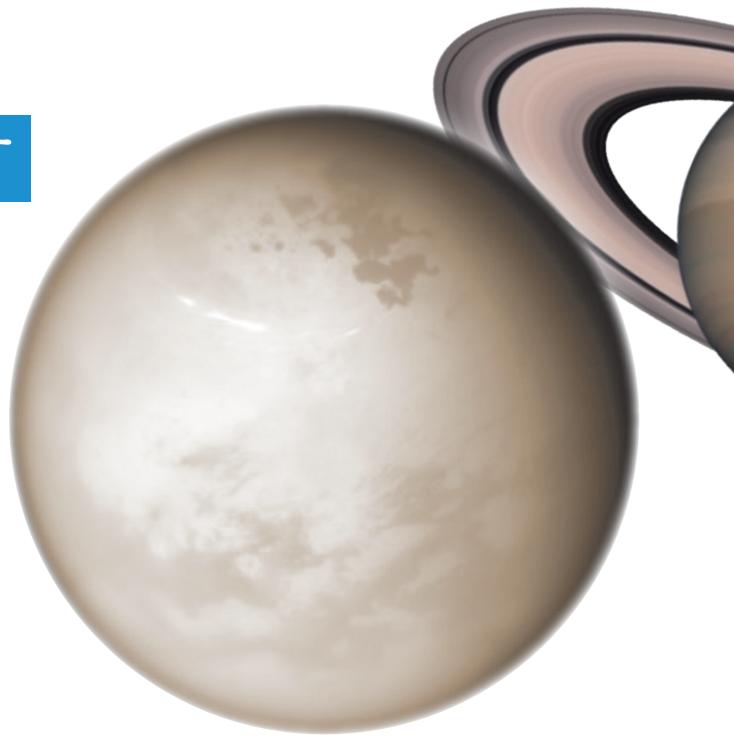
（取材・構成 大石かおり）

# 大型データから解き明かす 土星衛星の 大気環境

東京大学情報基盤センター 特任准教授

飯野孝浩

土星の衛星「タイタン」は、地球と同様、分厚い大気に覆われています。窒素を主成分とする大気には、生命の材料となる物質がどうつくれたかを知る手がかりが隠れています。飯野孝浩特任准教授らは探査機に頼らず、地上望遠鏡の観測データを解析することで、タイタンの大気における化学過程の一端を明らかにしました。いったいどのようにして遠く離れた天体の大気の成分を調べたのでしょうか。苦労したことなども含めて話をうかがいました。



## 【左】タイタンの大気は 生命材料がつくられる出発点

——まず、タイタンの大気を調べる意義について教えてください。

**飯野** | 多種多様な分子を含む分厚い大気に覆われているタイタンの環境は、昔から地球上に似ていると言われてきました。タイタンの大気は主成分が窒素で、その次に多いのがメタンです。大気中ではこの窒素とメタンを起点としてさまざまな化学反応が起

こり、複雑な分子がつくられています。やがてこれらは雲となり、雨となって地上に降り注ぎ、生命の材料になっている可能性があります。つまり、タイタンの大気中の化学過程は、生命の材料がつくられるプロセスの出発点と考えられるのです。タイタンの大気中でどんな分子がどのような化学反応で生じているのか。それを順に明らかにすることで、生命につながる化学過程の全貌が見えてくると世界中の科学者が期待しています。

## 【右】地上望遠鏡と解析技術により 探査機にも比肩する成果

——どのようにしてタイタンの大気の成分を調べるのですか？

**飯野** | 過去には探査機「ボイジャー」や「カッシーニ」が送り込まれ、土星を周回したりタイタンに小型機を降下させたりして詳細な調査が行われましたが、いずれもミッションが終了しており、次に土星に探査機が送られるのはいつになるかわかりません。

一方で近年は、地上大型望遠鏡を用いることでも遠く離れた天体を詳細に観測できるようになってきました。私たちは南米チリにあるアルマ望遠鏡の観測データを使って、タイタンや海王星など太陽系の中でも特に遠い天体の大気環境を調べています。

——研究からどんなことがわかりましたか？

**飯野** | 2020年2月に発表した研究では、タイタンの大気中にごくわずかに存在するアセトニトリル( $\text{CH}_3\text{CN}$ )という分子がどのように生成されているかを調べるため、アルマ望遠鏡を用いてタイタンの大気中のアセトニトリルとその窒素同位体が放つ微弱な電波を検出し、解析を行いました。その結果、アセトニトリルは成層圏下部で銀

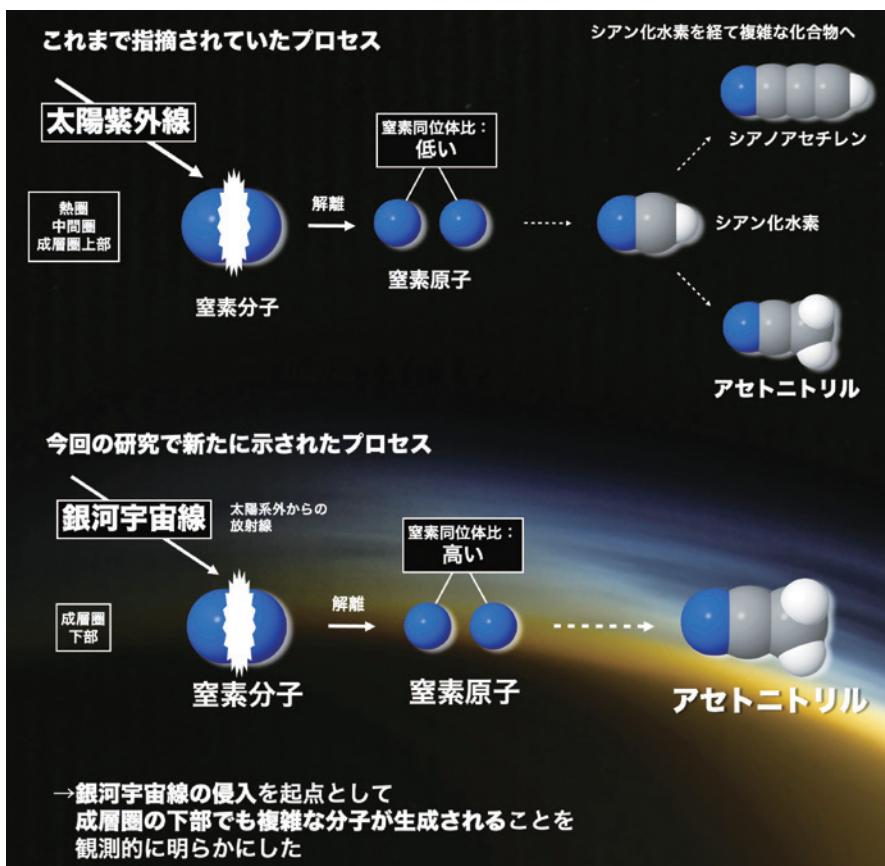


図 タイタンの大気中におけるアセトニトリルの生成プロセス

アルマ望遠鏡のデータを解析したところ、タイタンの成層圏下部でアセトニトリルの高い窒素同位体比( $\text{CH}_3\text{C}^{15}\text{N}$ に対する $\text{CH}_3\text{C}^{14}\text{N}$ の存在割合)が観測された。これは太陽系外から飛来する銀河宇宙線により窒素原子が解離してアセトニトリルを生成するというシミュレーション予測の結果と一致した。

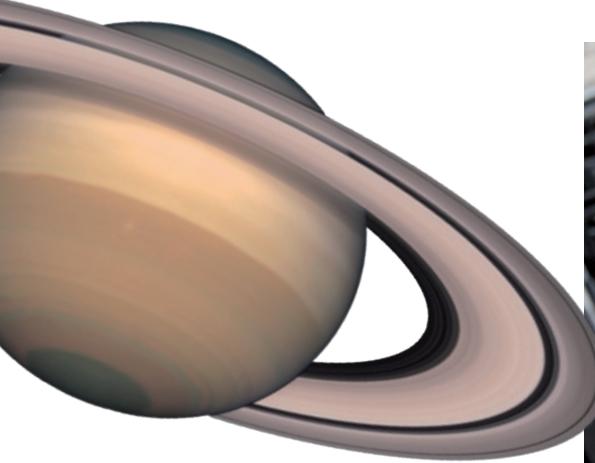


写真 観測データの解析に用いるハードディスク

48台のハードディスクをつないで、家庭用パソコンの100～1000倍に相当する624テラバイトもの巨大なストレージを構築した。

河宇宙線によって解離した窒素分子から直接生成されていることがわかりました（図下）。先行研究でこうしたシミュレーション予測はされていましたが、観測で明らかにしたのは世界で初めてです。「アセトニトリルは成層圏上部より上空で、太陽紫外線によって解離した窒素分子からシアノ化水素を経て合成される」という説（図上）もあったのですが、そうではありませんでした。この成果は、地上望遠鏡と解析技術を組み合わせることで、探査機に比肩する科学的成果が挙げられることを示したといえます。

当初は小さなハードディスクで  
膨大なデータを1つずつ解析

——研究を進めるうえで苦労した点は？

飯野 | アルマ望遠鏡は世界中の多くの研究者が使いたがっているので、観測提案を出してもそう簡単には採択されません。そこで私たちは、誰でも利用できるキャリブレーション用のデータ（目的とする天体の観測データを補正するために取得する観測データ）入手し、その中から研究に使えるタイプで質の高いものを選び出して解析を行っています。

ただし、アルマ望遠鏡がキャリブレー

ション用に取得しているタイタンの観測データは、年間で約200件、データ量になると約35テラバイト（1テラバイト=1000ギガバイト）の量になります。これを保存するだけでも大きなハードディスクが必要ですが、科学研究用のデータに変換するためには、さらにこの5倍の容量のハードディスクが必要になります。

研究を始めたばかりのころは資金がなかったので、家電量販店で購入した小さなハードディスクを使い、データを1件ダウンロードして処理したらデータを捨て、また1件データをダウンロードして処理しては捨て…ということの繰り返しで、大変な手間でした。1件のデータを処理するのに1日かかるので、1年分のデータを処理するのに半年以上もかかりました。

次第に研究費をいただけるようになり、624テラバイトの巨大なハードディスクを整備することができました（写真）。さらに、複数のCPUで効率的にデータ処理を行うソフトウェアも開発し、そのお陰で、今回の研究では約3000件のタイタンのキャリブレーション用データを1カ月足らずで解析できました。また、アルマ望遠鏡の観測データを解析するための一連のプロセスも確立できたので、その後はさらに

効率的に研究を進められています。

——今後の展望はいかがですか？

飯野 | 木星や海王星など、窒素を含む大気をもつ惑星は他にもあります。そうした天体でも、今回のように地上望遠鏡を用いた窒素同位体の観測から、大気の化学過程と惑星誕生後の歴史を解明できる可能性があります。今回構築した解析技術を他の惑星にも広げ、さらなる観測研究に発展させていきたいと考えています。

（取材・構成 秦 千里）

▶深く学ぶには

プレスリリース「地上大型電波望遠鏡により、  
土星の衛星タイタンの大気成分の詳細な観測に成功」

URL [https://www.u-tokyo.ac.jp/focus/ja/press/z0310\\_00018.html](https://www.u-tokyo.ac.jp/focus/ja/press/z0310_00018.html)



飯野孝浩／専門は電波天文学、惑星科学。名古屋大学大学院理学研究科博士課程修了、博士（理学）。東京農工大学科学博物館特任助教を経て2019年より現職。



皆さんは2000年から2005年にかけてNHKで放送されていた「プロジェクトX～挑戦者たち～」をご存知でしょうか。仕事に対して熱い情熱と使命感を抱き、前人未踏の「夢」を実現させてきた「名もなき人々」に焦点を当てたドキュメンタリー番組です。私が特に魅了されたのは、この番組が1人の異才・天才を取り上げるのではなく、さまざまな職種・立場の面々が一丸となって、いくつもの困難を乗り越えていく姿を丁寧に描いていた点です。

今号で特集されているmdxには、教員、研究員、事務方、ベンダーなど、多くの方々が関わっています。日本の研究レベルの底上げ、複合領域にわたる融合研究の活性化に資する基盤とすべく、目下運用体制の構築やシステムの機能強化に取り組んでいます。mdxが有効に利活用されるためには、その利便性向上はもちろんのこと、アカデミアのみならず広く社会に認知されるための広報や、利用を支援する体制

が不可欠です。自分1人の力は微力ながらも、チームが一丸となって、「プロジェクト md“X”」を成功させたいと願っています。

**有馬和美**／2012年9月に東京大学物性研究所計算物質科学研究中心に事務補佐員として着任。同センターで学術支援職員、学術専門職員を経て、2021年9月より現職。mdxの運用支援を担当。



息子（当時高1）が生地から自作したクリスマスケーキ。家事も「目指せ！ チーム一丸」です。

## プロジェクト md“X”

**有馬和美**（情報システム部情報基盤課データ利活用チーム 特任専門職員）

**nodes**の

## スパコン導入の舞台裏

### 中張遼太郎

（情報システム部情報基盤課  
スーパーコンピューティングチーム 一般職員）

「富岳」による新型コロナウイルス感染症関連のシミュレーション動画をご覧になった方も多いと思います。最近、スパコンについて報道される機会が増えました。しかし、スパコンがどのように導入されるかはご存知でしょうか。

東京大学でスパコンを導入する場合は、公的機関を対象とする政府調達の手順（仕様書作成→公告・入札→開札・納入）に従います<sup>\*1</sup>。仕様書には導入が必要な背景、ハードウェアなどの技術的要件、保守など性能・機能以外の要件などを詳細に記します。その際、実現不可能な仕様になつてはいけないので、念入りに市場調査を行い、納入時期に入手可能な機器とその価格を想定します。さらに、パンデミックによる物流への影響や世界的な半導体不足などの世界情勢にもアンテナを張り、調達への影響を考慮します。

無事に落札者が決まれば、サービス提供



2021年春に導入が完了したWisteriaの設置作業

に向けて詳細な仕様を決めていきます。規模などにもよりますが、開札から納入まで約5ヶ月、仕様書作成開始からは約2年を要し、構想期間も含めるとさらに長くなります。

動き出せばさまざまな分野で活用されるスパコンですが、その裏では運用はもちろん導入にも多くの労力が投じられているのです。

<sup>\*1</sup> [https://www.cas.go.jp/jp/seisaku/chotatsu/pdf/r2\\_siryou2.pdf](https://www.cas.go.jp/jp/seisaku/chotatsu/pdf/r2_siryou2.pdf)

**中張遼太郎**／東京大学工学系研究科修士課程修了、修士（工学）。学士課程では麻雀のコンピュータプレイヤー、修士課程では電気的筋肉刺激が重量知覚に及ぼす影響の研究に従事。大学院修了後、Sler企業を経て2017年10月より現職。



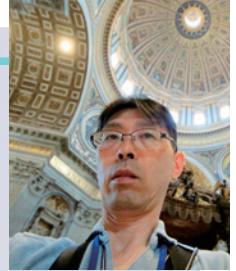
HPC（高性能計算機）向けシステムソフトウェア研究開発の醍醐味は、0と1の羅列で構成されるソフトウェアで最先端の超巨大システムを自在に操れること。思う通りにシステムが動作した時の達成感は格別で、モノづくりのおもしろさを実感するひと時です。

着任後、最初の任務は、Wisteria/BDEC-01のシミュレーションノード群とデータ・学習ノード群をつなぐ通信ソフトウェアWaitIOの研究開発でした。プロセッサも結合ネットワークも異なる2つのシステムを有機的に結合するWaitIOの実装により、複数システムにまたがるジョブ群を一体化して協調実行できるようになりました。2022年6月より試験運用が開始され、例えば、大量の気象シミュレーションの結果を機械学習してすばやい気象・気候予測を行うなど、実社会に役立つアプリケーションへの適用も始まっています。

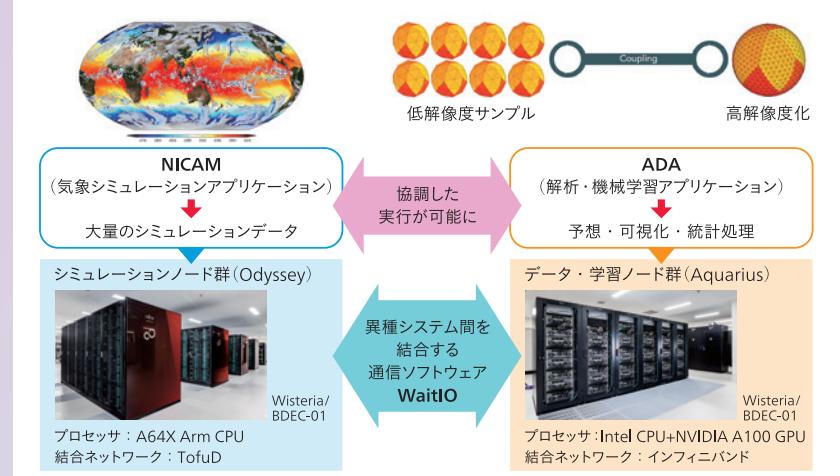
大学時代に8ビットパソコンでのプログラミングに魅了されてから、はや40

年。今までできなかったことをシステムで実現可能にすること、そのシステムで世の中をよりよくすること、それがシステムソフトウェア研究者のなすべき仕事と考えて研究を続けてきました。本職でも革新的なシステムソフトウェアで新しい世界を切り拓いていきたいと思っています。

**住元真司**／専門は高性能計算機向けシステムソフトウェアおよびアーキテクチャ。慶應義塾大学大学院理工学研究科・博士（工学）。前職は富士通株式会社、1997-2001年新情報処理開発機構出向後、国内ハイエンドHPCシステム（理研RSCLC、筑波大PACS-CS、「京」コンピュータ、富岳等）の研究開発に携わる。2022年4月より現職。



WaitIOの機能とアプリケーションへの適用例



## 0と1で超巨大システムを操る

**住元真司**（スーパーコンピューティング研究部門 特任教授）

ひろがり

## モノとその繋がりの科学

**華井雅俊**（データ科学研究部門 特任助教）

東京都の道路交通網をグラフ分割した図



大学で新しく学ぶ数学分野の1つにグラフ理論があります。グラフというと表計算の結果をまとめた図が思いつきますが、ここでの「グラフ」は別の意味で、「モノとその繋がり」を表します。「ネットワーク」と同じ意味で、こちらのほうがなじみ

があるかもしれません。日頃よく目にするグラフは例えば電車の路線図で、駅（モノ）と路線（繋がり）を表しています。

グラフ理論は約300年の歴史がある分野ですが、長い間研究されている理由の一つに遍在性（どこにでもあること）がありま

す。世の中には、道路交通網、インターネット網、金融ネットワークなど、さまざまなグラフが存在し、実は私たちの体内にも、脳神経網やタンパク質相互作用ネットワークなどのグラフが存在します。最近のものだと、LINE、Twitter、Instagramなどのユーザー・友だち関係は全世界規模の巨大なグラフです。

まずは身の回りのグラフを見つけてみましょう。そして300年続くグラフ研究の世界へ、第一歩を踏み出してみてください。学生の皆さんももちろん、誰にでもおもしろい発見があると思います。

**華井雅俊**／専門は大規模グラフ解析、Graph Neural Network、大規模離散系シミュレーション。東京工業大学大学院修了後、NTU（シンガポール）、SUSTech（中国深セン）を経て、2021年より現職。





# nodes

2022年9月発行

〔ノーズ〕  
vol. 2

編集・発行

東京大学情報基盤センター

〒277-0882 千葉県柏市柏の葉 6-2-3 柏II キャンパス内

<https://www.itc.u-tokyo.ac.jp>

©Information Technology Center, The University of Tokyo

ISSN 2436-9543(Print)、ISSN 2436-9551(Online)



本誌へのご感想、ご要望をお寄せください。  
印刷版の送付のご要望もこちらで承ります。

<https://www.itc.u-tokyo.ac.jp/public/nodes/>

〔表紙写真〕

mdx (部分)

mdx を用いた研究例：医学分野日本言語モデルの作成（モデルによる単語の疾患別分類）／擬似人流データの作成（可視化例）／機械学習による分子モデルからの物性値予測（分子モデルのイメージ、Material Project より取得 <https://materialsproject.org>）

〔裏表紙写真〕

mdx