

NII Today

National Institute of Informatics News

93

Dec. 2021

インタビュー | オープンサイエンスの
新拠点がスタート

研究データの管理・共有のハブとしての役割を担う柏分館
合田 憲人

インタビュー | mdxがめざすもの: 田浦 健次郎

座談会 | 新たな知の基盤を築く
安浦 寛人、田浦 健次郎、合田 憲人

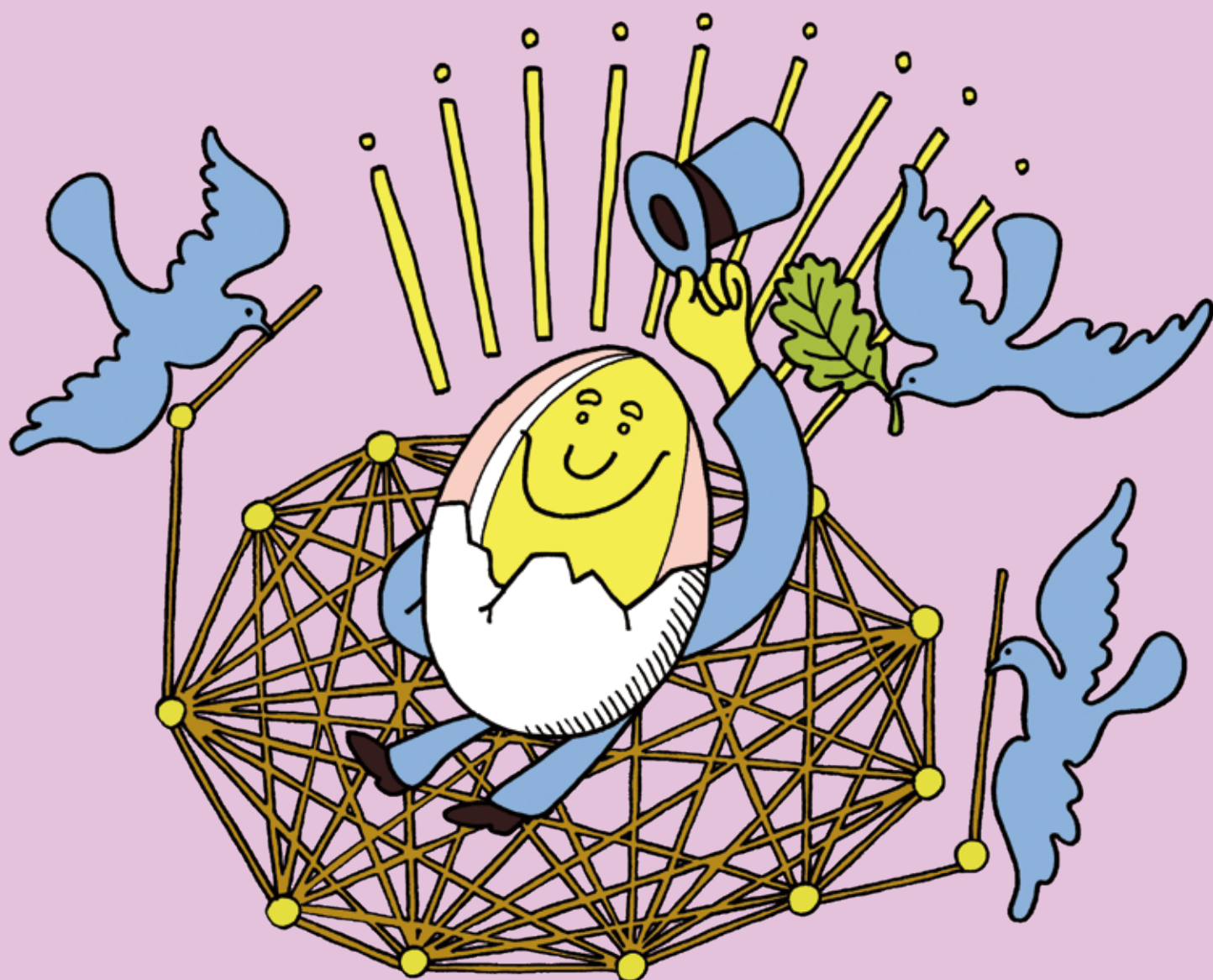
共同研究 | 地球観測データプラットフォーム「DIAS」
生駒 栄司氏、北本 朝展

Essay | 暗闇の中での対話: 稲邑 哲也

特集

新たな知の拠点、柏分館

学術情報基盤サービスの拡充へ



新たな知の拠点、柏分館

学術情報基盤サービスの拡充へ

特集

2021年3月、国立情報学研究所(NII)は、学術情報基盤サービスの拡充を目的に、柏分館(千葉県柏市)を東京大学(東大)柏IIキャンパス内に開設した。学術情報ネットワーク「SINET」や学術情報基盤サービスのための設備を備える柏分館は、東大との研究・連携を強化する場としても機能する。

東大には、データ活用社会創成プラットフォーム(mdx)が導入されており、NIIも運営機関の一つとして参加している。NII柏分館は次世代学術研究データプラットフォームとしての一翼も担うことになる。

本特集では、柏分館の概要、mdxにおける役割、東大との共同研究について紹介する。



合田 憲人 AIDA, Kento

国立情報学研究所 アーキテクチャ科学研究系 教授
総合研究大学院大学 複合科学研究科 教授
学術基盤推進部 部長

聞き手

山田 哲朗氏 YAMADA, Tetsuro

読売新聞論説委員



インタビュー

オープンサイエンスの新拠点がスタート 研究データの管理・共有のハブとしての役割を担う柏分館

日々、膨大な研究データが生まれ蓄積されている。眠れるデータを活用すれば、新しい知見を導き出せる。人工知能(AI)と組み合わせれば、新たな発見ができる。こうした研究の拠点となるべく、2021年3月に国立情報学研究所柏分館が発足した。データ活用社会の推進を担う新拠点の意義について、柏分館の発足に尽力した国立情報学研究所(NII)の合田憲人教授に聞いた。

データ管理や共有に役立つ ストレージサービスを開始

山田 東京大学柏IIキャンパスにできた柏分館にはどんな設備がありますか。

合田 NIIは、所内または所外の研究者向けに、さまざまなサービスを提

供していますが、そうしたサービスに必要な計算機やネットワークなどの設備が柏分館に設置されています。

所内向けサービスとしては、NIIが運用している研究用クラウドの設備が置かれています。能力の高いサーバが100台ぐらいという規模でしょうか。

また、所外向けの共同利用の事業サービスとしては、昨年度に新しくつ

くった、研究者が研究データを管理する「研究データ基盤(NII Research Data Cloud, NII RDC)」(p.7の※2参照)のストレージがあります。

山田 ストレージの設置には、どのような意味がありますか。

合田 NIIはこれまでストレージサービスは手がけていませんでした。データ管理ということでは、管理のた

めのインターフェースやサービスは提供していましたが、「データ自体は、それぞれの大学内のストレージや大学が契約したクラウドなどに置いて下さい」という形でした。NIIの研究データ管理基盤がこれらの大学のストレージとつながっており、「研究データ管理基盤を利用することで大学に保存された研究データを管理ができます」というスタンスでした。

ただ、規模の大きな大学などは自前でストレージを持ったり、クラウドを借りたりすることが可能ですが、それができる大学というのは、予算面や、情報スキルのある人材の面から非常に限られます。中小規模の大学になると、自分でストレージを持つことができない場合が多いので、そうした部分をカバーするため、NIIが標準的なストレージサービスを用意し、そこにデータを置けるようにしたわけです。これはまったく新しいサービスになります。

山田 ストレージを使いたいというニーズが相当あったわけですね。

合田 スパコンを持っているような大学はかなりのことが自前でできるのですが、SINETでつながっている機関は全国に1000弱ぐらいもあり、そのうちスパコンを持てるような技術がある大学は少ない。ほとんどは設備も人材も持っていないので、NIIが大学共同利用機関として、きちんとサポートしていこうという考え方です。

研究データを管理・共有することで新たな研究へ

山田 新しいストレージサービスを使って、どんな研究を期待していますか。

合田 研究の内容はさまざまでしょ

うが、まずは、研究した後のデータの管理や、データの共有に役立ちたい。「研究データを研究室のPCに置いておいたら、いつの間にか消えてしまった」などということを防ぐとか、さらには、研究データも最近は機微な情報を含むものもありますので、きちんとアクセスを制限するなど、しっかり管理しながら進める必要があります。研究データをきちっと管理できる文化というか、研究体制をつくるのが目的ですね。

次に、研究データをきちんと管理できるようになったら、今度はこれまでのデータを利用して新しい研究をしようという動きが、ほかの研究者を含めて出てくるので、データを使った研究が飛躍的に広がっていくのではないかと思います。

ですから、データ管理とその利活用という、2つのことを狙っているのが柏分館です。こういう共通のインフラをつくるのがNIIの仕事でもあります。

山田 データを活用した新しい研究というと、具体的にはどういうものがイメージできますか。

合田 私が別のプロジェクトで経験した例を紹介すると、例えば医療ビッグデータがあります。NIIの「医療ビッグデータ研究センター（Research Center for Medical Bigdata, RCMB）」は、国内の6つの医療系分野の学会と連携し、全国の病院から集めた画像データをAIで解析するためのクラウド基盤をつくりました。データを集める医師や、データを解析するAIや画像解析の専門家らが、同じ基盤を使って研究をするというプロジェクトです。

ここでは、30ぐらい共同研究テーマが立ち上がり、大きな学際的コミュニティができました。データが集まってくると、それを提供する人、分析する人も

集まって、いろいろなアイデアや研究課題が出てきます。1つの共通基盤があったからこそできた学際的研究の良い例だと思います。同じようなことが柏でも起こることを期待しています。

山田 NIIが「こういう研究をやってくれ」と言うより、いい舞台を用意すれば、自然に面白い研究が生まれてくるわけですね。

合田 いえいえ、そう甘くはなくて、「設備を用意したから後はよろしく」ではうまくいかないと思います。やはりNII自身がコミットするというか、共同研究に入り込んで一緒に考えていかないと、箱ができただけで終わってしまう。医療ビッグデータの例にしても、NIIに医師はいなくても画像処理やAIの専門家はたくさんいるので、我々の側がハブになるという意気込みで臨んでいました。「基盤を用意したのでどうぞ」ではなくて、他分野のことも勉強しながら、共通基盤の上で人と人を結びつけるような努力が必要です。

山田 NIIも一緒に舞台をつくり、他分野の方たちとともに演じなければ成果にはつながらないのですね。

NIIがハブとなり研究のマッチングに貢献したい

山田 柏分館では、NIIが支援したり調整したりする体制はありますか。

合田 それをつくらなければいけないと思っています。柏分館が動き出したのは今春で、まだそこまでいっていません。実際にデータを集めたい人は、計算機の専門家ではあまりせん。「私はこんなデータを持っています」という人と「私はこんな解析をしたい」という人をうまくマッチングする仕組みをつく

らなければいけないと考えています。そういったプロジェクトを探し、「こんなデータがありますけど、こんな研究はどうですか」という感じでこちらからアプローチしながら研究を進める形をとっていかなければいけないでしょう。

山田 NIIがアプローチして発掘するぐらいでないといけなわけですね。例えば、モバイル端末を使って、いままでにないデータを集めたいといった要望はあるのでしょうか。

合田 それはいろいろあります。自動車や人の移動情報などのほか、農業系では、放牧している牛の情報を集めています。医療系では、人の生体情報を集めて解析する動きもあります。

共通の基盤があれば、そこをベースにさまざまな新しい共同研究や連携が始まると思うので、NIIはそういった部分のハブになりたいと思っています。柏分館では同じ建物内のしかも同じフロアに東京大学の情報基盤センターがあることが強みです。そこにはmdxというデータ活用社会創成プラットフォームがあり、全国の大学や研究機関とSINETで高速接続しているので、ネットワーク的にもハブになります。柏を経由して、さまざまな研究が生まれるのではないかと期待しています。

山田 今後、ボトルネックになりそうな弱点や、心配な点などありますか。

合田 先ほどお話ししたように、機械を置いただけではダメで、人がいろいろとサポートしなければいけません。学際的な支援・調整ができる人材というのは限られますので、いかにそうした人材を育て、集めていくかが課題でしょう。

計算機のインフラ構築から、他分野の研究者との共同研究まで、幅広

くこなせる人材は少ないものです。学術の世界では研究論文がどれだけあるかというのが評価の中心になりがちですが、データインフラをつくり維持する仕事からは、どうしても論文の数は出ません。そういう人の能力をきちんと評価し、アカデミアで正当なポジションを与えることも大切だと思います。

ます。

柏分館の成功は、人がどれくらい入りするかというより、ここを介して研究データの利活用がどれだけ進むかという点にかかっています。研究者から、計算機の設備が足りないという声が聞こえるくらい活用されるとうれしく思います。 (写真=盛 孝大)

インタビューからのひとこと

データの利活用が研究力を左右するようになるというのは、材料科学や生命科学に限った話ではなく、いずれ人文・社会科学でもビッグデータを使った思いがけない調査手法や研究成果が登場するのではないだろうか。日本は、行政手続きや企業活動の面ではデジタル化の遅れが指摘されているが、幸い、学術面ではまだそうした状態ではないらしい。柏分館が中心になり、新時代のデータ駆動研究をリードしてもらいたい。

山田 哲朗 読売新聞論説委員

1990年、東京大学卒、読売新聞入社。2006年、マサチューセッツ工科大学(MIT)ナイト科学ジャーナリズム・フェロー。経済部、科学部、ワシントン支局特派員などを経て、2018年、科学部長。2019年から論説委員(科学技術担当)。



NII柏分館(2021年3月開所)

NII柏分館 施設情報		
	内訳(参考)	
機器室	1階 480.80㎡	研究系開発室(NW): 194.40㎡
		研究系サーバ室1: 286.40㎡
	2階 669.60㎡	研究開発室(クラウド): 334.80㎡ 研究系サーバ室: 334.80㎡
実験研究室	5室(2階1室、3階4室(64.80㎡/室))	
共同研究室	4室(3階(129.60㎡/室と121.60㎡/室が2室ずつ)) *うち1室は簡易パーティションで16区画に区切られ、シェアオフィスとして利用可	
ミーティングルーム(会議室)	1室(4階49.36㎡)	
その他「施設情報」	・非常用発電機(500KVA)・管理員(施設)が24時間常駐	



柏分館の研究室にて、PCセッティング中のNII 河原林健一教授



インタビュー

mdxがめざすもの 東京大学とNII柏分館の 連携相乗効果に期待

田浦 健次郎 TAURA, Kenjiro

東京大学情報基盤センター長
東京大学大学院情報理工学系研究科 教授
国立情報学研究所 客員教授

聞き手

山田 哲朗氏 YAMADA, Tetsuro

読売新聞論説委員

東京大学(東大)柏IIキャンパスの新拠点で、東大など9大学とNIIなど2研究所が共同運営する「データ活用社会創成プラットフォーム(mdx)」が始動した。研究データの利活用に狙いを定めたmdxの特徴や可能性について、プロジェクトを率いる田浦健次郎 東大情報基盤センター長に聞いた。

より学際的な、ビッグデータ 研究に資するmdx

山田 柏IIキャンパスに学術データの拠点ができた経緯を教えてください。

田浦 東大の五神真前総長とNIIの喜連川優所長が、データ駆動型社会に向けて、公的な新拠点をつくろうと意気投合されたことがきっかけです。東大側も、情報基盤センターを移転し、データ活用社会創成プラットフォーム(mdx)を新設すると決めました。

山田 東大情報基盤センターはすでに、8大学を結ぶ「学際大規模情報基盤共同利用・共同研究拠点(JHPCN)」(※1)の中核になっています。

田浦 情報基盤センターのミッションは、東大にとどまらず、全国的なミッションとしてスパコンの基盤を整備しています。スパコンは、情報学の研究

に加え、情報関係の分野ではない理工系の研究者にも計算のために広く使われてきましたが、近年ではこうした情報基盤をハブとして学際的な研究に役立てたいという要望も増えています。1つの大学だけでは持てないような設備を広く全国に提供し、情報科学や情報基盤の専門家でない研究者にもスパコンを使ってもらって、共同研究に貢献するのが目的です。

これまで、スパコンについてはいわゆる理工学系のシミュレーションのための利用が中心でした。それを、データ科学やデータ利活用を軸とする分野に拡大し、幅広いユーザに使ってもらうためのプラットフォームとして新設したのがmdxです。

山田 より学際的であり、ビッグデータを扱う研究者全般に利用が広がるイメージでしょうか。

田浦 はい。従来スパコンを使って

いた研究者の中にも、計算の結果生まれたデータや実験の結果生まれたデータを蓄積し、後から解析できるようにデータベース化して、他の研究者に共有したいと考えている人もいます。これまでのスパコンではできなかった部分を補い、可能にするのがmdxの狙いです。

NII RDC(研究データ基盤)との連携で学術研究を支える

山田 そもそもデータが存在すること自体を知らないとか、どこにデータがあるかわからないといった、研究者もいるかと思います。

田浦 さまざまな深さのレイヤーで取り組まなければいけません。mdxという基盤を1つつくれば解決する問題でもありません。全国的なデータをカタログ化し検索できるような世界をつ

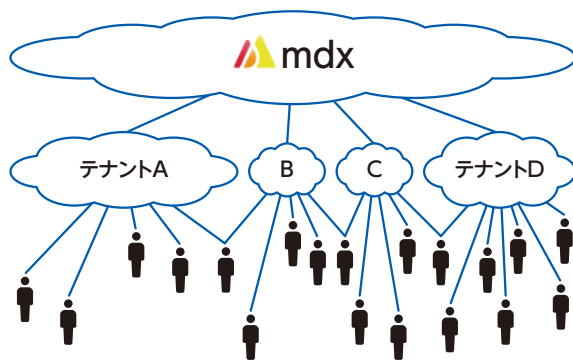


図 mdx上に構築される仮想化環境と利用の仕方
仮想化された高性能環境により、利用グループごとに分離した高いセキュリティ環境を提供する。

くっていかないといけないと思います。

例えば投薬や医療行為のデータ、IoTのセンサーや実験から出てくるデータ、気象データのような観測データなど、それぞれの分野で集めたものをお互い効率よく共有でき、簡単に処理や計算もできるような、そういう基盤が全体として必要です。

山田 NIIとどう連携しますか？

田浦 2021年度からNIIの「研究データ基盤(NII Research Data Cloud, NII RDC)」(※2)事業が本格運用されています。その基盤の1つ「GakuNin RDM」を活用して研究データを管理し、研究者間でデータを共有します。さらに研究者が自分の管理するデータを公開基盤である「JAIRO Cloud」(※3)につなげば、データを簡単に公開することができます。

mdxの役割の1つは、NII RDCを通じて共有、公開するデータをきちんと格納するストレージを提供することです。もう1つは、見つかったビッグデータを実際に処理しようという時に必要な計算機の機能を提供することです。NIIのRDCと我々のmdxは、全体として、誰に対しても優しいサービスをいかに提供できるかを問われます。情報系のスキルの高い専門家から、それ以外のWebブラウザでデータを手早く分析したい人まで、いろいろな人に使ってもらえるものを整えていきます。

各研究機関をつなぎ 全国的な学術基盤へ

山田 コンピュータ技術、ネットワーク技術の観点では、どんな利点がありますか。

田浦 これまでのスパコンというのは、1つの大きな部屋を用意して、そこに「みんな一緒に入ってきて使ってください」という仕組みで、ある程度は他のユーザから勝手にデータを見られたりしないようにはなっているものの、分離の度合いは弱く、セキュリティ上、注意が必要な個人データや医療データなどをスパコンに上げるのは躊躇されるケースが多くありました。

これに対し、mdxはユーザごとに、1つの完全に独立した仮想化環境をつくることができます(図)。セキュリティが重要な用途にも、安心度が高い分離した環境をつくることができ、その中は基本的にユーザが好きにいじってカスタマイズしていいという仕組みです。ここが今までといちばん違う点です。アカデミアのクラウドとして

は、かなり先駆的な取り組みと言えるでしょう。

山田 東大とNIIの相乗効果は期待できますか。

田浦 NII柏分館と東大情報基盤センターは同じ建物にありますので、密な人的交流に期待しています。NIIは学術情報ネットワーク「SINET」で日本中の大学をつないでいます。さらに、先述のようにNII RDCを通じて、研究者が自分の研究データを共有しています。これらとmdx、さらには各大学にあるスパコンなどをつないだ全国規模の一体的な学術基盤をつくっていくにあたり、NIIと情報基盤センター群の密接な協力が不可欠だと思っています。今後の展開に期待してください。(写真=盛 孝大)

【用語解説】

※1=JHPCN:「学際大規模情報基盤共同利用・共同研究拠点」:北海道大学、東北大学、東京大学(中核)、東京工業大学、名古屋大学、京都大学、大阪大学、九州大学の基盤センターが運営する共同研究のための組織。全国から共同研究を募集し、採択された課題に計算設備(これまでは主にスパコン)の利用時間を割り当てている。

※2=NII RDC:NIIが提供する、研究データのライフサイクルを支える基盤として、2021年に本格運用を開始。研究データの管理基盤(GakuNin RDM)、公開基盤(WEKO3)、検索基盤(CiNii Research)から構成されている。

※3=JAIRO Cloud:NII開発のWEKOを機関リポジトリソフトウェアに採用した、クラウド型の機関リポジトリ環境提供サービスとして2012年度に運用開始。2016年7月からはJPCOARとNIIとの共同運営に移行。



田浦 健次郎

1997年東京大学理学系研究科博士課程を修了。博士(理学)。1996年より同大学で助手、2001年より講師、2002年より准教授、2015年より教授。2018年より情報基盤センター長。国立情報学研究所客員教授を務める。専門は並列処理、高性能計算、プログラミング言語、システムソフトウェア。高性能、大規模並列処理とプログラミング、利用の容易さを両立させる研究に従事している。

座談会

新たな知の基盤を築く 科学技術の進展を担うインフラの構築



安浦 寛人
YASUURA, Hiroto

国立情報学研究所 学術基盤
チーフディレクター／文部科
学省 科学技術・学術審議会 情
報委員会主査



田浦 健次郎
TAURA, Kenjiro

東京大学情報基盤センター長
／東京大学大学院情報理工学
系研究科 教授／国立情報学
研究所 客員教授



合田 憲人
AIDA, Kento

国立情報学研究所 アーキテク
チャ科学研究系 教授／総合研
究大学院大学 複合科学研究科
教授／学術基盤推進部 部長

柏分館が提供するサービスは全国を対象にしており、各大学や研究所との連携や協力が必須となる。また研究の発展を促すべく、ユーザ同士を結び付ける役割も担っていく。これまでばらばらだったデータ活用のあり方をまとめて最適化し、科学の進展に貢献する。この大きな課題に挑む3人のキーパーソンに、現状と展望を語ってもらった。

データ科学の進展を支える

田井中 新たに情報基盤が構築された背景を教えてください。

田浦 大きな情報基盤やデータ基盤は、これまで主に理工系分野のシミュレーションに使われてきましたが、いまは人文系のデジタルアーカイブをはじめ、社会科学、医療、薬学、教育学など、さまざまな分野で情報技術やデータ活用が進展しています。

データ科学が発展した理由の1つはAIの登場です。特に深層学習によるAIは、現象の根本的な原理はわからなくても、データを元に未知の結果が予測できます。つまり、方程式がなくてもデータがあればいろいろなことが解明できる。もう1つは、取得されるデータ量が増えたこと。センサーの発達などにより、さまざまな観測データを簡単に入手できるようになりました。

これまで、情報基盤センターや国のスパコンが対象にしてきたのは従来のシミュレーションですから、使い方にミスマッチが出てきました。ユーザ層の広がりに応じて、新たな研究手法に合った基盤をつくる必要があったわけです。それが根本にあります。

安浦 ある事象を観察し、そこから仮説をつくり、数学を中心に定式化して結果を予測する。それこそがギリシャ・ローマ時代からの科学でした。特に18世紀以降の物理学では、理論物理学と実験物理学が相互に作用しながら学問を発展させてきました。その後、20世紀半ばにコンピュータが登場し、数式さえあれば、計算によりさまざまなことが予測できる計算科学が現れました。天気予報などはまさにそうですね。

さらに20世紀の終わり頃からはデータ量が急増し、現在、パソコンではギガバイト(ギガ=10⁹)が当たり前

になりました。私が情報工学を学んだ1970年代は、大学で一番大きな計算機の記憶容量は1～2メガ(10⁶)でした。いまやエクサ(10¹⁸)、ゼタ(10²¹)といった大容量のデータが使われ始めています。

かつての理論科学と実験科学がそうであったように、人間が式を立ててシミュレーションをする計算科学と、AIなどによるデータ科学が相互に影響し合いながら発展しつつあります。柏分館およびデータ活用社会創成プラットフォーム(mdx)は、このデータ科学を支えるための仕組みをつくっているわけですね。

合田 これまでデータは研究者が自分たちで管理していましたが、あまりにも巨大化したためにそれができなくなり、共通基盤が必要になったということも背景にあります。データを共有するためには、アクセス権や共有できる範囲などをきちんと管理すること

が必要で、そのための基盤を個々の研究グループごとに用意するのは難しい。データの置き場所はユーザ自身の判断になりますが、NIIの研究データ基盤では、研究データ管理の仕組みに加えてストレージも設置して、データ利活用を図れるようにしています。

田浦 研究に必要なデータを自分だけで集めることは難しいため、データを共有する価値は大きいと思います。ただ、個人情報など機微性の高いデータや、材料系や医薬系などの商業的な開発に絡むデータは、公開が難しいのが実情です。セキュリティを高めたうえで、活用する価値をきちんと示すことが重要だと考えています。

田井中 国としての戦略が必要でしょうか。

安浦 情報関係の科学技術政策としては、1970年代から主要な大学に大型計算機センターをつくりました。当時は何十億円もするような設備でしたが、現在のパソコンより能力が低いものでした。さらに地球全体の大気の状態をシミュレーションするような高性能の計算機をつくろうと、90年代後半にスーパーコンピュータである「地球シミュレータ」が開発されました。その後、さらに高速なスパコンの構築をめざして国家プロジェクトが発足し、「京」や「富岳」につながっていきました。計算科学では世界最先端のトップレベルの計算ができる技術と、それを動かせる人材を国としてしっかり持っていることが非常に重要です。

今後はデータの分野でもそれを行わなければなりません。極めて大きなデータを安定的に維持するために、データに関してスパコンに匹敵するような仕組みをどうつくるかが1つの課

題になっています。ただし、データ量だけではなく、どのように使われるかが問題です。セキュリティの程度や公開の有無など、それぞれのデータの特性に合わせてハンドリングする技術を、国として持つ必要があるでしょう。mdxの導入をきっかけに学術データ基盤のあり方を手探りで考え始めている段階です。

データ科学の面白さは、異なる分野のデータを掛け合わせることにあり、新しい発見の可能性があることにあります。異なるコミュニティが互いにデータを公開し合って一緒に考えることにより新しい発見がある。その可能性を試す場をどうつくるかということも、大きなチャレンジでしょう。

プラットフォームの設計と運用

田井中 データプラットフォームの設計のポイントを教えてください。

田浦 人と人をどうつなげるかが重要だと思っています。それぞれのデータは量も違うし、数値の意味するところも全く違うので、信頼性を含めてその道の専門家にしか解釈できません。そのような、データに詳しいさまざまな分野の人たちと、計算の仕方やコンピュータに詳しい私たちが、問題そのものを共有して一緒に議論するためには、きめ細やかな対応が必要だと思います。

合田 データの集め方も重要です。これまでは、ユーザが自分のデータをスパコンにドンと投入して計算するスタイルでした。これからはセンサーから直接データが送られることが想定されるため、スパコン側も計算の

ためのリソースだけでなくデータを受け取る部分が必要になります。送られてくるデータの形式もまちまちですし、プログラム(ソフトウェア)もアプリケーションごとにばらばらです。さらに、機微情報の場合は匿名化して送らなければならないなど、それぞれの状況に応じたバリエーションも考えなければなりません。

単にネットワークとサーバが速ければいいのではなく、そのようなソフトウェアを安全に動かすための仕組みが必要です。これもまたユーザとともに議論をしながらつくらなければなりません。

田井中 運用についてはいかがでしょうか。

田浦 mdxは1つの共通基盤を全国の11機関で共同運営します。それぞれの機関の得意な分野や強みを生かして、いろいろな分野のグループとつながり、mdxでの成果を最大化していきたいと考えています(図)。1機関が新しいマシンを買ったら、それを11機関でシェアするような新しいサービスの形態もつくっていきます。成果を出して、データ科学の基盤を広げたいですね。

田井中 mdxは2021年9月に公開されたところですが、現状を教えてください。

田浦 mdxのホームページから利用申請ができます。NIIが全国の大学と連携して運営する「学認」の認証基盤から入ることができますが、多くの大学ではまだmdxを利用可能にする設定をしていませんので、周知していきたいと思っています。

合田 「学認」で使うサービスは、セキュリティ上、各大学で利用許可の

設定を行ってから利用する仕組みにしています。これまでは電子ジャーナルなどの全学的な利用が主でした。mdxのような計算機システムが提供するサービスに使うのは、おそらく今回が初めてで、大学側への啓蒙も必要ですね。

田浦 一般ユーザへ参加の呼びかけも行っており、研究室単位や個人などで順次利用が始まっています。分野は、AI、画像処理、スマートシティに使われる空間情報、ゲノム、マテリアル材料、経済関係の人流データなどさまざまです。

安浦 まだまだ試験的な運用で、いろいろな問題点に対応している状況だと思います。大きな計算機やデータ基盤を持ってない研究者が気軽に申請し、mdxにデータを入れて研究する。5年後くらいにそのようなことがスムーズにできるようになれば、サポート側としては成功と言えるでしょう。

国家的なインフラとして 実働させるために

田井中 商用クラウドとの使い分けはどのようになるでしょうか。

合田 データをどこに保管するかはコミュニティごとに考え方が違いますが、使い分けの方針を決めてうまく運用すればいいと思います。

田浦 商用クラウドの場合、費用が高額になりますし、ほとんどのベンダーは海外の企業ですから、皆、やむを得ず使っているのが本音だと思います。ほかに選択肢がないから、スパコンでできないものは商用クラウドで、となるわけです。そこに、国のインフラとしてどれだけのものを整備でき

るか。使いやすい環境をつくっていかねばなりませんね。

安浦 研究は競争です。先に見つけた人が発見者ですから、1番でないといけない。国のサポートが不十分だったために結果が出せなかったということがないように、データ分野でもしっかりした基盤をつくる必要があります。

私が主査を務める、文部科学省 科学技術・学術審議会の中の情報委員会では、すべての学術分野に対して、学術情報基盤をどのようにつくり、維持していくかについて議論しています。富岳や学術情報ネットワーク「SINET」に加え、今後はmdxを中心としたデータ基盤も対象になります。

この分野は6～7年経つと技術が古くなるので、予算要求も常に先を見据えて準備し続けなければなりません。特にサービスの提供体制を持続可能な仕組みにしておくことが重要です。すなわち、人件費や電気代などいろいろな経費を見込んで、安定した予算をつける必要がある。もしmdxが成功すれば、こうした基盤を複数力所に分散させるなど、予算的に1桁上の仕組みを考えることもできるでしょう。そのためにも、実績が必要です。

NIIの果たすべき役割

田井中 柏分館に何を期待されますか。

田浦 NIIへの期待はとにかくたくさんあります(笑)。現在の学術情報基盤は各大学にばらばらに分散していて、ユーザは複数の大学のアカウントを使い分けなければなりません。NII

には、物理的に分散しているものをまとめて、ユーザから1つのシステムに見えるようなデータ基盤サービスを構築していただきたいと思います。

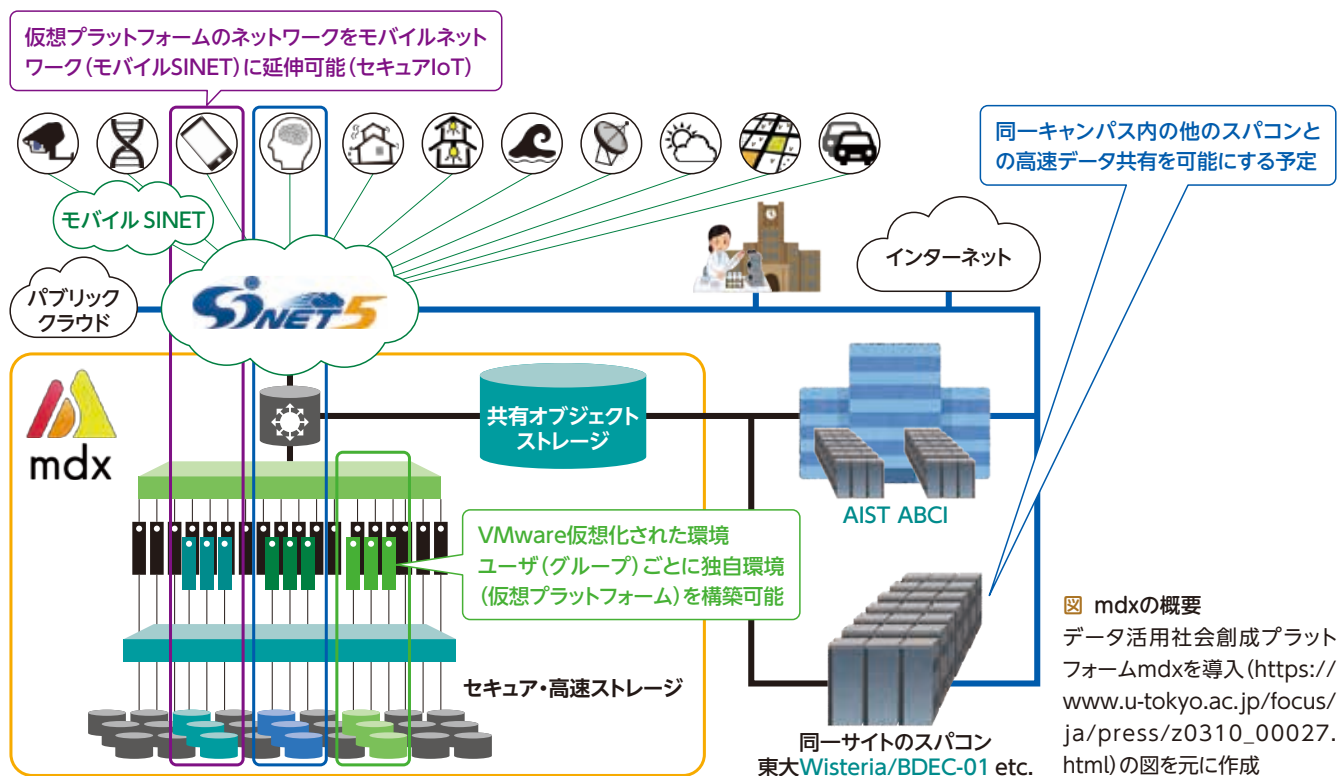
また、私たちにとって一番大事なのは、合田さんをはじめとする柏分館の方々との協働です。定期的に作戦会議を開いて、mdxやSINET、研究データ基盤システム(NII Research Data Cloud)など身近な話題を話し合いたいですね。困難な問題が出てきた時には、一緒に予算を獲得して共同研究を始めるなど、緊密な関係を望んでいます。

合田 しっかり受け止めたいと思います。データが生成され、ネットワークを経由して解析され、その結果が世の中で役に立つ。それが一気通貫でできる世界を柏分館でつくりたい。それを実現するインフラが柏にできたことは、たいへん意義のあることだと思います。

SINETには有線だけでなくモバイル網があり、これを介してデータを自動的に集めて計算機に安全に流し込む仕組みもつくっています。そういう技術的な仕掛けをパズルのように組み合わせ、計算パワーのあるmdxにつなげていきたいですね。

実は田浦先生と私は学生の頃、同じコミュニティで活動していたのです。この数年はお互いにインフラを構築・運用する立場になっているわけですが、新しい日本の基盤を一緒につくれることを非常に楽しみにしています。

安浦 研究資源の共用は重要な課題ですが、コミュニティや大学にはそれぞれ固有の文化があって、問題はそれほど単純ではありません。NIIはそういうものを乗り越えながらSINETな



どを手掛けてきた。今度はそのデータ版を粕でやろうとしています。運用において、重要な役割を果たすべき立場にあると思います。

田井中 人材の育成についてはいかがでしょうか？

安浦 非常に大きな問題です。この仕事の重要性がもっと理解されなければいけません。学術論文では、執筆者と同様に、データをつくる人や整理する人が重視される方向に流れは変わりつつあります。研究する人とそれを支える基盤を担う人、この両者の共同事業として研究成果が生まれる。そのような文化を醸成していくことが大事だと思います。

NIIには先端的なサービスを常に維持し、改良できる人が必要です。大学でも、研究より実務のほうが合っている優秀な人はいます。そういう人たちが給与面も含めて研究者と同じように待遇されるような環境をつくってい

く。そうでないと、日本の科学技術自体が危うくなっていると思います。

田浦 情報系の研究者として論文を書くことと、他分野の方と協力して研究結果を出すためのヘルプをすることは、往々にして両立が難しいところですね。情報技術に精通する者として、他分野の研究者を助けなければいけない場面は少なくありません。そういう大事な貢献が見過ごされないよう、世の中を少しでも変えていきたいですね。

コロナ禍では多くの授業がオンライン化されました。そのとき情報技術の得意な学生たちが自主的に立ち上がり、困っている先生や学生を助ける仕組みをつくり出したのです。東京大学ではいま、それが制度として定着しつつあります。このように意欲的で、かつユーザの気持ちがわかる学生たちが、よいタイミング、よい仕組みでこの仕事に触れてくれば、若い力が

育ってくるのではないのでしょうか。

合田 データが集まってくると、不思議とそれに関連するいろいろな人たちも集まってきて、つながりができるものです。この活動は、そのようなつながりの上に広がっていくはずですね。インフラ屋として応用分野の人たちとたくさんの会話をし、お互いの課題を解決できるよう、努力をしたいと思います。

安浦 データ科学の基盤整備は科学技術の成果に直結することから、世界で競争になっています。mdxはその第一歩ですが、日本の今後10年以上にわたる科学技術の進展を背負っていると言っても過言ではありません。しかもこの取り組みは海外と比べても先進的なものと言えます。この事業を皆さんに広く知っていただき、研究者にはぜひ使っていただきたいと思っています。

(取材＝田井中麻都佳、文＝平塚裕子)

共同研究

地球観測データプラットフォーム「DIAS」

社会課題の解決に資する研究を柏分館から



生駒 栄司氏
IKOMA, Eiji
東京大学
地球観測データ統融合連携研究機構
特任准教授



北本 朝展
KITAMOTO, Asanobu
国立情報学研究所
コンテンツ科学研究系 教授
総合研究大学院大学 複合科学研究科 教授

柏分館の重要な目的の1つに、東京大学(東大)との連携研究の強化がある。そのなかで、昨今の気候変動に対する注目度の高まりから、急速にその重要性が認知されつつある国家プロジェクト「DIAS(Data Integration & Analysis System)」について紹介する。システム面からリードする東大地球観測データ統融合連携研究機構 生駒栄司特任准教授と、オープンサイエンスの観点から研究の高度化を図る国立情報学研究所(NII)の北本朝展教授に話を聞いた。

地球環境研究の基盤である「DIAS」

地球温暖化に、科学はどう解決への緒いとぐちを示すのか。DIASはこれらの課題に取り組む研究者たちを支援するための、ビッグデータプラットフォームだ。DIASには約50ペタバイト(ペタ=10¹⁵)もの地球環境データの格納が可能だ。さらにそれらのデータを統合・管理するミドルウェアや、各種データを利用するための共通APIやアプリケーション開発基盤、研究者が直接利用するサービスまでもが連携。予測分析やシミュレーションに使いやすい形でデータを統合的に管理し、研究活動を支えてきたところにDIASの特徴がある。

第1期プロジェクトがスタートしたのは2006年(※1)。現在の第4期「地球環境データ統合・解析プラットフォーム事業」(2021～)は、海洋研究開発機構(JAMSTEC)を中心に、東大やNIIも分担機関として参画する

かたちで進められている。もっともDIASは2006年に突然始まったものではない。原点は、約40年前に東京大学生産技術研究所(東大生研)が手がけた地球観測衛星データのアーカイブに遡る。手づくりの衛星データ受信アンテナ、自前の記憶装置(テープアーカイブ)など、データの収集から格納まですべてを自分たちで地道に行ってきた活動の蓄積が、現在のDIASへつながっている。

「東大生研では40年前から、『データによる公共的な価値の創出』をめざし、地球観測のための各種の衛星データなどを受信し、収集したデータを格納するシステムを自前で構築、研究に使えるデータとして統合的に管理してきました。現在も、やっていることは基本的に変わりません」と、生駒特任准教授はこれまでの活動を振り返る。

膨大なデータを扱うのに最適な柏分館

DIASのシステムは現在、東京・駒場

にある東大生研と東大柏IIキャンパス、北海道大学、北見工業大学の4カ所に分散しており、柏IIキャンパスには東大とNII柏分館の2施設が含まれている。拠点間の通信には、NIIの学術情報ネットワーク「SINET」が使われており、気象衛星ひまわりやXRAIN(国土交通省が運用するリアルタイム雨量観測システム)などの外部データもSINET経由でDIASに格納されるしくみだ。

「DIASにはシミュレーションデータ以外に、ひまわりなどの観測データも毎日、リアルタイムでテラバイト単位(テラ=10¹²)で流れてきます。安定運用には、大容量データ的高速伝送が可能なSINETのパワーが欠かせません。また、大規模なデータシステムを大学だけで運用するのは難しい。SINETに直結した柏IIキャンパスの利用はDIASにとって必須でした」と、生駒特任准教授は柏分館との連携の意義を話す。

プロジェクトの初期段階からDIASの活動に関わる北本教授も、「柏分館には大規模な設備を置くスペースがあり、電力設備も充実している。大規

模システムを支えるのに十分な可用性を備えています」と強調する。

なお、DIASの登録ユーザ数は2021年8月時点で9,416名。2015年度の約5倍増となった。データ量だけでなくユーザ数、アクセス数ともに激増するDIASのバックボーンを、柏分館の先進的な設備が支えているのだ。

オープンサイエンスや人材など、今後の課題

柏分館ができたことで、ハード面のサポートが拡充されたDIAS。今後はオープンサイエンスに、いっそう注力していくという。北本教授は、「地球環境データはプライバシーなどに関する問題が比較的少なく、環境や防災など公共性が高いものが多いため、研究者間でのデータ共有の意識が比較的高く、オープン化しやすい。ただこれまではデータのライセンスが必ずしも明確ではなく、オープン化の妨げになる場合もありました。現在はクリエイティブ・コモンズ(CC) (※2) ライセンスをデータセットに付けるなど、改良を重ねています」と語る。

北本教授は、近年、研究者の間で「オープンデータ指向」が急速に強まっていると指摘する。DIASでもこのトレンドに呼応するかたちで、データセットに一意な識別子「DOI: Digital Object Identifier」を付与する活動を2017年に始めた。オープンデータが抱える課題に、「データを使う人(論文の執筆者など)に対し、データをつくる人の貢献が見えにくい」という問題があるが、DOIの付与により、データの利用実績やデータが研究に与える功績を可視化しやすくなる。



■ 平常(usual) ■ 増水中(rising) ■ 危険水位(critical) ■ 氾濫(flood)

図 深層学習を用いた氾濫画像の自動認識(球磨川水系)。ライブカメラの映像をリアルタイムに深層学習で解析し、氾濫状況を色分けして示す。

一方、生駒特任准教授は今後のDIASの活動は、気候変動問題への対策が中心になると語る。なかでも自治体からの要望がとくに強いのが、台風や集中豪雨による河川の氾濫の予測だ。たとえば熊本県球磨川周辺では全国の約2,700カ所に取り付けられた国交省のライブカメラから収集した画像をDIAS上で解析し、河川氾濫の危険度をリアルタイムに判定するプロジェクトが始まっている。観測中のライブカメラの画像を危険度の高い地点の順に並べたり、急速に危険度が高まっている地域をピックアップして、画像をリアルタイムで確認したりすることも可能だ(図)。

今後は、DIAS上の水位等の観測データや気象予報データなどを重ね合わせることで、直近の河川氾濫予測や、自治体のシステムと連携した避難場所の状況確認など、さらなる拡張が期待される。そのほか、浸水予測、災害情報のポータル化、気候データと観光地のサービスとの連携など、さまざま

な分野での適用を計画中だ。

一方、課題は人材不足だ。システム構築からその上で動くアプリケーションやサービスの管理、ユーザである研究者のサポートまで、プロジェクト担当者には幅広い役割が要求されるため、人材獲得が難しいことは否めない。しかし、気候変動やサステナビリティに対する若い世代の関心は数年前に比べて明らかに高くなっている。

「DIASへの参加が、キャリアパスへとつながっていくような道筋を示すことが必要でしょう」と生駒特任准教授。柏分館との協業はDIASを魅力的なデータプラットフォームに成長させるための重要なステップになるだろう。

(取材・文=五味明子)

【用語解説】

※1=文部科学省研究委託事業「データ統合・解析システム」として開始(2006~2010年)。第2期「地球環境情報統融合プログラム」(2011~2015)、第3期「地球環境情報プラットフォーム構築推進プログラム」(2016~2020)と継続。

※2=CC:作者の意思を反映しながら著作物の流通(再利用)を図るための国際的なルール。



生駒 栄司

2000年東京大学大学院工学系研究科 電子情報工学専攻博士課程を修了。博士(工学)。同大学空間情報科学研究センター助教等を経て現在同大学地球観測データ統融合連携研究機構特任准教授。2016年より文部科学省研究開発局学術調査官を兼務。専門は情報工学、データベース工学。

NII NEWS TOPICS



ニュース・トピックス一覧 期間:2021年9月1日(水)～2021年11月24日(水)

ニュースリリース

- 2021/11/24 効率的なスパコン設計につながるグラフ発見を競うコンペ「グラフ ゴルフ」で理論上最小の直径を持つ間接網を発見～ネットワークスイッチを用いてCPUを間接接続する構成にも有効～
- 2021/11/17 人文学・社会科学総合データカタログ「JDCat」本格運用を開始 ～人文学・社会科学データの横断検索が可能に～
- 2021/11/15 自動運転における重大な問題をシミュレーションで検出する技術を開発 ～問題が発生するかを探り、起こりうる問題だけを効率的に自動探索～
- 2021/11/04 制御システムの最適化問題をディープニューラルネットワークで解く ～制約付き非線形システムの離散時間確率的最適制御問題への新たなアプローチを構築～
- 2021/10/20 信頼性が高いガスタービンのシステム設計を自動で効率良く発見する技術を開発 ～ブラックボックス最適化での論理仕様に着目、企業の実製品の設計プロセスに応用～
- 2021/09/22 AIにより生成されたフェイク顔映像を自動判定するプログラム SYNTHETIQ: Synthetic video detectorを開発 ～AI動画の生成、フェイクメディアの検知、メディアの信頼性確保の研究を推進～

受賞

- 2021/11/03 計 宇生教授(アーキテクチャ科学研究系)らの論文が2021 IEEE Internet of Things Journal Best Paper Awardを受賞
- 2021/11/01 水野 貴之准教授(情報社会相関研究系)らの論文が2021年度 人工知能学会全国大会(第35回)優秀賞を受賞
- 2021/11/01 ガラムカリ 和さん(総研大情報学専攻、杉山研究室)、杉山 磨人准教授(情報学プリンシプル研究系)の論文が2021年度 人工知能学会全国大会(第35回)優秀賞を受賞
- 2021/11/01 劉 佳特任助教(サイバーセキュリティ研究開発センター)らの論文が国際会議NaNA 2021の

Best Paper Awardを受賞

- 2021/10/27 井上 克巳教授(情報学プリンシプル研究系)、マニャン モルガン客員教授らの論文が国際会議ILP2020-21のBest Paper Awardを受賞
- 2021/10/12 Paolo Arcaini特任准教授(アーキテクチャ科学研究系)らの論文が13th Symposium on Search-Based Software EngineeringにてDistinguished paper awardを受賞
- 2021/09/17 金子 めぐみ准教授(アーキテクチャ科学研究系)がKDDI Foundation Award 2021 貢献賞を受賞
- 2021/09/09 稲邑 哲也准教授(情報学プリンシプル研究系)らの論文が日本ロボット学会の第9回 Advanced Robotics Best Survey Paper Awardを受賞

NII情報

- 2021/11/15 国立情報学研究所グローバルサイエンスキャンパス(GSC)～2022年度「情報科学の達人」プログラム 受講生募集開始
- 2021/10/27 「教育機関DXシンポ」アーカイブズサイトを公開
- 2021/10/21 情報研シリーズ(丸善ライブラリー)最新刊『ビッグデータが拓く医療AI』を刊行
- 2021/10/20 学認クラウドサービスメニュー拡充／利用申請やサービスメニューの利便性向上
- 2021/10/06 2021年度 国立情報学研究所 市民講座「情報学最前線」の参加受付を開始
- 2021/10/05 広報誌 NII Today 第92号「個人情報からプライバシーへ」を発行
- 2021/10/01 2022年度 国立情報学研究所 公募型共同研究の募集を開始
- 2021/09/01 ウェブサイト「関東大震災映像デジタルアーカイブ」を公開



イベント www.nii.ac.jp/event/2021

- 大学等におけるオンライン教育とデジタル変革に関するサイバーシンポジウム「教育機関DXシンポ」

- 2021/11/19 第43回オンライン開催
- 2021/10/29 第42回オンライン開催
- 2021/10/8 第41回オンライン開催
- 2021/9/17 第40回オンライン開催
- 2021/9/3 第39回オンライン開催
- 2021/11/22 NII-IDRユーザフォーラム2021(オンライン開催)
- 2021/11/17 NIIサービス説明会 ～次世代学術研究プラットフォーム直前情報!何が変わるの?～
- 2021/11/01 第23回 図書館総合展(ブース出展・フォーラム)
- 2021/11/01 総合研究大学院大学 複合科学研究科 情報学専攻 大学院説明会
- 2021/10/24 大学共同利用機関シンポジウム2021 ～宇宙・物質・エネルギー・生命・情報・人間文化～フロントの知を楽しもう(オンライン開催)

- 2021年度 市民講座「情報学最前線」
- 2021/10/10～公開中
 - 第1回 データに隠れた規則性を探せ!～パターンが織りなす構造とAI～ 杉山 磨人
- 2021/11/1～公開中
 - 第2回 AI製品の信頼性をどうやって評価する?～AIのすごさと製品としての難しさ～ 石川 冬樹
- 2021/12 以降も順次公開予定



ご意見募集中!

Nii Todayのさらなる誌面の充実のため、本誌に対する皆様のご意見を下記URLまたは左のQRコードより、お寄せください。お待ちしております。

www.nii.ac.jp/today/iken

情報犬ビットくん (Niiキャラクター)

CiNii Researchを本公開しました

CiNii Research

研究データを論文や図書等と統合的に検索することができるサービス、CiNii Researchを2021年4月1日より本公開しました。(※1)

CiNii Researchは、先立つこと2020年11月にプレ版と呼んでいたバージョンを公開していたところですが、本公開にあたっては、詳細検索や、OR、NOT演算等での検索、RDF等のAPI機能を追加し、より充実した検索をしていただけるようにしています。

また、2021年10月4日より新たに、APIとして、OpenSearch、JSON-LD、ResourceSyncの公開、文献管理ソフトとの連携のための書き出し機能の公開

を行い、機能強化を行っております。(※2)

CiNii Researchで検索が可能となった研究データの数、2021年9月末時点で、約32万件です。論文等に比べるとまだまだ数は少ないですが、本公開開始時点の約18万件に比べると着実に伸びています。今後も、オープンサイエンスの推進を確実に実施するため、収録データベースを増やし、検索対象となる研究データの数の充実に努めます。

さらに、2022年春にはCiNii Articlesを統合し、論文の検索をより一層拡充させ、機能強化、内部情報の高度化を実施していく予定です。

従来、CiNii ArticlesのAPIを利用されていた方、また、CiNii Articlesのサイトにリンクをはられていた方は、CiNii ResearchのAPI利用への切り替え、CiNii Researchのサイトへのリンクの切り替えをしていただきますようお願いいたします。

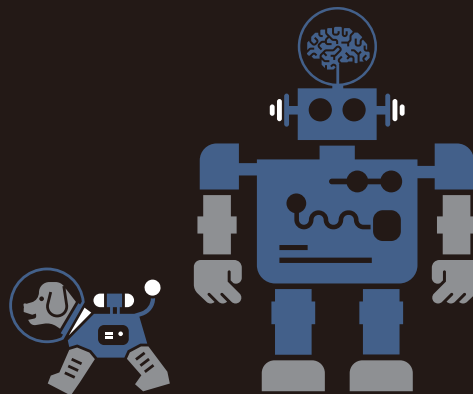


- ※1 CiNii Research (<https://cir.nii.ac.jp/>)
- ※2 CiNii ArticlesのCiNii Researchへの統合について (<https://support.nii.ac.jp/ja/news/cinii/20210706>)

Essay

暗闇の中での対話

稲邑 哲也 INAMURA, Tetsunari

国立情報学研究所 情報学プリンシプル研究系 准教授
総合研究大学院大学 複合科学研究科 准教授

「ダイアログ・イン・ザ・ダーク」というイベントをご存知だろうか？ 視覚に障がいを持つ方々が体験する日常生活を、健常者も同様に体験することができるものである。一切の灯りが無い暗闇に設置された公園の中を散歩したり、店舗で買い物をしたりする。この状況に放り込まれると、聴覚、特に触覚の感覚が研ぎ澄まされ、感覚と行為の関係性が脳内で高速に再構築されている感覚が得られる。

また、感覚の鋭敏化とは別に、もう一つの気づきが得られる。他者の存在の重要性だ。一人では何も行動ができないため、健常者でグループを構成し、視覚障がい者をリーダーとして暗闇の中を進む。聴覚、触覚等によって感じる対象は物体・環境ではなく、他者だ。公園に流れる小川に落ちないように先に進む者が後の者に声をかける。店舗では、何が陳列されているのか分からないので、店員役の視覚障がい者の方に質問をして手渡してもら

う。もちろん壁にぶつからないように物体に対する感覚を研ぎ澄ます必要があるが、ほとんどの注意の対象は他者の存在である。このイベントの名称が「ダイアログ」から始まっている理由が、暗闇の中で初めて腑に落ちる。

人間はこのように激変する状況に対して高度な適応能力で対処することができる訳だが、ロボットだったらどうなるのだろうか？と、研究者視点にスイッチを切り替えながら暗闇の中を歩く。カメラが破損した場合、恐らくロボットは別のセンサーを用いて世界のモデルを再構築し始めるだろう。カメラが使っていた時に学習した世界のモデルを用いて、距離センサーを元に映像を再現する事も可能であろう。暗闇の中でも器用に動き回るロボットが想像される一方で、積極的に対話し始めるロボットの姿が想像できない。対話をしなくても何とかなるからだ。

パンデミックという暗闇の時代の中、人類は対面コミュニケーションとい

うモダリティを失った。他者の声は聞こえるし、存在していることは理解できるが、顔写真のアイコンが存在感の具現化を阻む。普段、握手やハグを頻繁にしていた訳ではないが、肌感覚の喪失感がこみ上げる。暗闇に対抗する手段である代替の感覚モダリティが存在せず、工夫を凝らしているものの決定打が見つからない。そして、この困難に手を差し伸べるのは、皮肉にも対話を必要としないロボットかもしれない。

近年のAIの自然言語処理能力の向上は目覚ましい。しかし依然として、身体に根ざした感覚モダリティ・体験と言語処理系を結びつけ、対話を実現するにはさまざまな課題が立ちほだかっている。その一つが「他者」の存在感・必要性ではないか。他者を必要とするような弱いロボットが、人間にとって強い味方になり、暗闇に対抗する手段になるかもしれない、と妄想しつつ、今日もwebカメラ越しに暗闇の中を手探りしながら対話をする。

情報から知を紡ぎます。



国立情報学研究所ニュース：NII Today 第93号 令和3(2021)年12月
発行：大学共同利用機関法人 情報・システム研究機構 国立情報学研究所
〒101-8430 東京都千代田区一ツ橋2丁目1番2号 学術総合センター

本誌についてのお問い合わせ：総務部企画課 広報チーム
TEL: 03-4212-2028 FAX: 03-4212-2150
EMAIL: kouhou@nii.ac.jp

発行人：喜連川 優

編集委員長：河原林 健一 第93号監修/編集委員：竹房 あつ子、込山 悠介

編集委員：池畑 諭、金子 めぐみ、水野 貴之(五十音順)

外部編集員：田井中 麻都佳

池田 亜希子/片寄正史(サイテック・コミュニケーションズ)

デザインDTP：倉橋 弘(マツダオフィス) 表紙イラスト：市村 譲

この印刷物は、FSC®(森林管理協議会)認証材を使用し、有害な廃液を出さない「水なし印刷」、揮発性有機化合物(VOC)成分を含まない環境に配慮したNon-VOCインキを採用しています。

