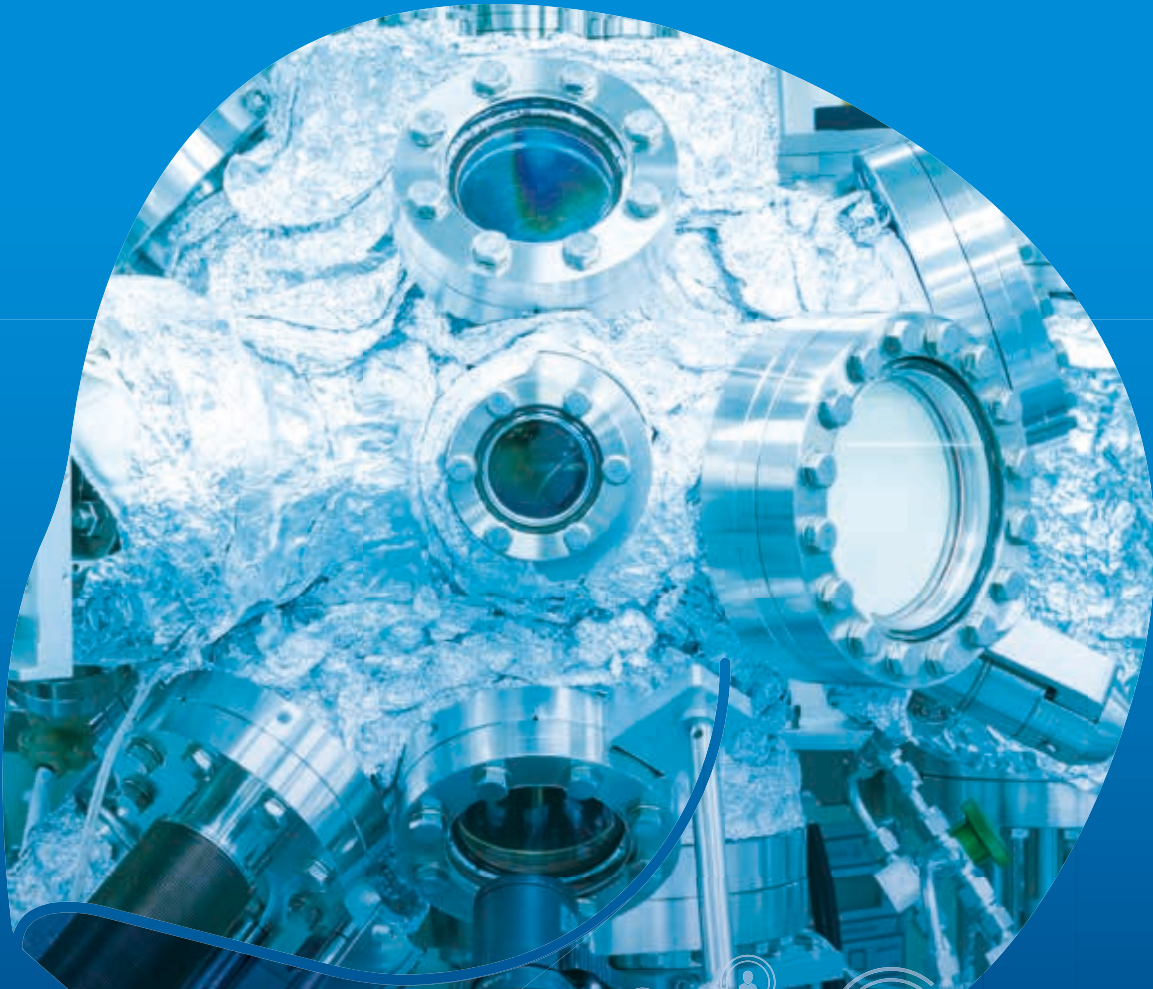


NIMS NOW 5

No.



マテリアルDX の真価

～材料開発の現場が、研究のあり方が変わる～

マテリアルDX の真価

～材料開発の現場が、研究のあり方が変わる～

新材料の探索にAI・データ科学を取り入れる。
NIMSが先導してきたマテリアルズインフォマティクスの取り組みは、
材料開発のDX(デジタルトランスフォーメーション)化を促進してきた。

その中であらためて浮き彫りになったこと。
それが、データ共有の重要性だ。

日々量産される多様で膨大な研究データ。
それを適切な形で迅速に共有するためには、分野の壁、組織の壁を越えた、
横断的かつオープンなプラットフォームの構築が不可欠だ。
NIMSはこれまで培ってきたデータ収集や活用の手法、そして先端機器共有のネットワークを生かし、
オールジャパン体制でのプラットフォーム構築へと今まさに舵を切った。

材料開発だけでなく、研究のあり方そのものをも変える可能性を秘めたマテリアルDX。
その真価を引き出せるか——改革の現場に迫る。



Masahiko Demura
国立研究開発法人物質・材料研究機構
統合型材料開発・情報基盤部門 部門長

出村
雅彦

江頭
基

Motoi Eto
文部科学省 研究振興局参事官
ナノテクノロジー・物質・材料担当



特別対談

マテリアルDXがもたらすもの

2020年度から文部科学省主導で推進されている「マテリアルDXプラットフォーム構想」は、産学官の材料データの戦略的な収集・蓄積・流通に加え、データが効率的・継続的に創出・共用化される仕組みを持った、日本全国が連携する材料研究のプラットフォームだ。NIMSでは従来のデータ収集・活用の手法をさらに発展させ、データ中核拠点の構築に取り組んでいる。我が国が目指すべきマテリアルDXプラットフォームとは何か。その中でNIMSが果たすべき役割とは？ プロジェクトをけん引するNIMSの出村雅彦部門長と文部科学省研究振興局の江頭基参事官がマテリアルDXの真価について話し合った。

「マテリアルDXプラットフォーム」の整備に取り組むNIMS

江頭 AIによるビッグデータを使った材料開発が世界的に進む中、NIMSは世界最高水準の材料データベースの運用をされています。それが素材産業を大きく支えていますよね。

出村さんは、内閣府のSIP*「統合型材料開発システムによるマテリアル革命」の

中の研究課題「先端的構造材料・プロセスに対応した逆問題MI基盤の構築」で研究責任者をされています。このプロジェクトでは、NIMSの持つデータベースをさらに強化し、単なるデータ効率化というだけでなく、人の扱いきれないほどのデータ量の掛け合わせが可能なデジタルトランスフォーメーション (DX) へと飛躍させようという狙いですね。

*戦略的イノベーション創造プログラム

出村 NIMSでは長年にわたり材料データベースを構築し、運用してきました。2017年からは新しい時代の材料データプラットフォームの整備を開始し、データを蓄積する場所の整備、効率的にデータを集める方法、そして集めたデータの利活用方法の開発という3本柱で進めてきました。2020年度からは、文部科学省主導の「マテリアルDXプラットフォーム構想」のデータ中核拠点構築にも取り組んでいます。また2021

年4月には、「マテリアル先端リサーチインフラ」事業が始まり、NIMSはセンターハブとして採択されました。材料開発の現場では、AIでビッグデータを解析し、開発の高速化を行うことは当たり前になっています。産官学の研究データをより広く、より使いやすくすることがDX推進において重要です。**江頭** 私が参事官として担当しているSIPの課題でも、同じくデータを収集して利活用しようという課題がいくつかあります。その1つが「スマート物流サービス」です。IoT技術により集荷に関するデータを集め、クラウドサーバーで計算処理して配送の最適化をするというものです。同じエリアに荷物を運ぶのであれば、異なる事業者同士が1台のトラックに混載するだけで、かなりの効率化が図れます。SIPで実証実験をしたところ、労働時間やガソリン代、CO₂排出量などあらゆる面で数十%は効率化できることがわかりました。物流の流れは企業秘密という考えが一般的です

が、一方で物流業界は深刻な人手不足に陥っており、このような状況を乗り切るには集荷情報をデータ化して共有し、配送業者同士、協力するべきという機運が高まっています。これでも企業間の壁に風穴を開けるものとも言えるかもしれません。NIMSには材料の研究者が集まっていますが、一口に材料と言っても種類は多く、さらに同じ材料の種類であっても、機能性材料から構造材料まで多岐にわたります。また、研究の観点も物理、化学、金属学、機械工学まで実にさまざまです。最近では、一層細分化しており、専門性が深まっているともいえますが、隣でやっていることが見えなくなっています。タコつぼ化が進んでいるのかもしれませんが、NIMSでもDXにより、

研究者間の障壁に風穴を開け、専門領域に横串を刺すことが重要であると考えています。**江頭** 日本の素材産業はまだまだ強いですが、日本の材料メーカーは、研究開発を通じて蓄積してきたデータを企業間で共有して、連合体として研究開発を加速させていかなければ、今後海外の材料メーカーに勝っていけないという強い危機意識をお持ちだと思います。ぜひDXにより産官学や研究者間の障壁に風穴を開け、素材産業全体のさらなる国際競争力の向上を目指していただきたいところであり、国としても強力に支援していく方針です。**データ構造は研究に対する研究者の見方そのもの**
出村 さきほど、物流業界には「集荷に関するデータは企業秘密」という考えがある

**研究者同士の分野横断のコミュニケーションが活性化し、
新たな概念が生まれやすくなるような
マテリアルDXプラットフォームを構築していきたい——出村**

と伺いましたが、実はデータが共有されなかった理由は、案外ポリシーの問題ではなく、データをやり取りするための共通のデータフォーマットがないことが本質的な原因だったのではないかと私は推測しています。逆にいうと、データフォーマットが共通化していれば、データ共有の方が業務を最適化して利益率を上げることができるのですから、自然とそちらに向かっていくことになる。つまり、データフォーマットがブレイクスルーの起点になるのではないかと考えています。

江頭 おっしゃる通りだと思います。たとえば先ほどのスマート物流サービスプロジェクトでは、業者間で異なる配送車両番号を運輸局に車両登録する際の番号に統一するといった共通ルールを決めました。また、同じ商品でも配送業者によって登録名称などが異なっていたため、各社データを共通データに変換し、物流プラットフォームを構築したのです。その結果、実証実験での効率化が裏付けられ、データ共有のメリットを実感していただきました。

出村 データフォーマットの共通化やその変換手法は私たちも大きな課題のひとつで、取り組みを強化しているところです。

江頭 マテリアルDXプラットフォームに関しては、単にデータフォーマットを共通化するだけでなく、研究者がなぜこのような測定をしたのかわからなければ、他の人にとっての利用価値は低いのではないかと思います。いかがでしょうか。

出村 おっしゃる通りです。私は、研究は3つのレイヤーで構成されていると考えています。まずはデータそのものです。研究成果はデータに基づいて出てきますのでデータが基本です。しかし、データを取得する前提として、どのようなデータをどのように、どこに着目して測定し、その測定結果をどう表現するかを示すデータ構造というレイヤーがあります。先の例で言うとデータフォーマットがデータ構造にあたります



**DXにより産官学や研究者間の障壁に風穴を開け、
素材産業全体のさらなる国際競争力の向上を目指していただきたい——江頭**

ね。日時、研究者名はもとより、研究目的、対象とする材料名、材料組成、研究手法などの基礎データがきちんとデータテーブルに添って構造化されているということです。データ構造があって初めてデータを測定することができます。いわばデータ構造は研究対象に対する研究者の見方そのものです。また、データ項目同士の物理的な関連付けを行うためのレイヤーも大切で、それをモデルと呼びます。

江頭 データ構造、データ、モデルという3つのレイヤーですね。

出村 その通りです。研究論文には必ずこれら3つのレイヤーが含まれています。ただ、論文は著者の主張が読者に伝わりやすいように書かれているため、3つのレイヤーが入り混じって構成されています。これを3つのレイヤーに整理して機械が読める形で投稿することが習慣化されると、マテリアルDXプラットフォームが一層進展すると思

います。そして、これは、単に研究が効率化するだけではなく、研究者同士の新たなコミュニケーションを生む仕組みになるのではないかと考えています。

**目指すべきマテリアルDX
プラットフォームの姿とは**

出村 収集して整理することをキュレーションと呼びます。データを活用するには、データ構造を読み解く必要がありますから、データだけでなく、研究毎に違うデータ構造から、データ収集用にテーブル構造をどう構築するかのキュレーションも大切です。わかりやすくいうとエクセルの表組みの定義ですね。こうしたデータ構造は研究者の物の見方そのもののわけですが、よく考えてみると、科学を発展させるのは新しい物の見方です。そのため、データ構造がダイナミックに変わっていくことも当然認め

なければなりません。生き生きと変容していくデータ構造をどのようにデジタル化するか、この点が基礎研究におけるDXのポイントでもあります。

江頭 研究者の発想と研究者同士のコミュニケーションを妨げるようなプラットフォームになってはいけません。研究者によって研究スタイルや文化は異なりますし、社会的ニーズも変化していきますので、データ構造が変わっていくことを許すのは当然のことでしょう。

出村 データ構造を通して、「この人はこんな物の見方をしているのか」「あの人はこんな表現の仕方をしているのか」といった研究者の思想や哲学を知ることができません。また、データ構造を基に「これとこれは同じだね」「これはあれの上位概念だね」といった整理をする行為は、実は学術活動そのものです。新しい概念が出されたとき、それを検証することこそが学術活動の本質

です。

学術研究では、これまで論文や学会での発表を介して研究者同士のコミュニケーションが行われてきました。私が若い頃は、論文を、主に、月々届けられる雑誌を通じて読んでいて、自然と隣の領域にも目がいったものです。学会も今ほどは数がなく、一つの学会に出ておけば、材料分野を広く見渡すことができました。現在では、日々大量の論文が発行されており、インターネットで検索するのが日常化しています。学会も多様化していて、全てを見渡すことは到底叶わない。これでは自分の関心事だけに情報が偏ります。この状況に風穴を開けるDXでありたい。

江頭 先ほどの、研究者間の壁に風穴を開けること、専門領域に横車を刺すことですね。

出村 はい。私はデータ構造がDXを推進する切り札だと思っています。データ構造

の収集と整理、つまりキュレーションを通して、隣接領域のデータも活用できるようになると、研究が効率化するだけでなく、研究者に気づきを与えることができるのではないかと考えています。新しい概念がより生まれやすくなるともいえます。私がデータそのものよりも、データ構造をキュレーションすることの方が重要だと考える理由はそこにあります。私は今後10年間をかけて、研究者同士の分野横断のコミュニケーションが活性化し、新たな概念が生まれやすくなるようなマテリアルDXプラットフォームを構築していきたいと考えているのです。

江頭 当初マテリアルDXプラットフォームとは、大雑把に言えば、計測データを効率よく収集してデータベースを整備することかなと考えていましたが、目指すべきゴールはもっとずっと先にあるのですよね。単に色々なところに存在しているローカルデータを集

めてきて、共有化できるようにしようということではなく、DXにより研究者同士のコミュニケーションが進み、研究の幅や発想が広がり、新しいものが生まれる可能性が高まるということなのですね。

出村 そうですね。そのためには平行して進めている、全国の大学等の共用設備から研究データを収集するスキームであるマテリアル先端リサーチインフラもますます重要になってくるでしょう。

江頭 出村さんのお話を伺っていると、今後DXによって研究機関の在り方そのものも大きく変わっていくのではないかと感じました。DXにより、研究者間だけでなく、時間や場所という障壁も取り払われ、特定の研究機関に属する研究員といった研究スタイルも唯一ではなくなっていくのではないのでしょうか。今回はDXについてじっくりと対談できてよかったです。どうもありがとうございました。

出村 こちらこそどうもありがとうございました。(文・山田久美)



データ循環の要となる 材料データプラットフォーム 「DICE」、遂に始動!



谷藤 幹子
Mikiko Tanifuji

統合型材料開発・情報基盤部門
材料データプラットフォームセンター センター長

使いやすいデータベースにこそ、よりよいデータが集まる——データ利用の好循環を生み出すために、NIMSは材料データプラットフォーム『DICE(ダイス)』を2020年6月に公開した。マテリアルDXの核としての活用が期待されるDICEについて、材料データプラットフォームセンター (DPFC)の谷藤幹子センター長に聞いた。

データ駆動の 入り口から出口までを担う

DICEは、データを効率的に収集・蓄積するデータ駆動型研究の入り口としての機能と、収集したデータをAIが解析しやすい形式に整え、解析のためのツールもあわせて公開する出口としての機能を兼ね備えた画期的なシステムだ。収集するデータは、実験・計測データ、計算データなど研究現場で生まれるデータに加えて、論文や特許などの公知データや、研究連携などによる外部データベースなど幅広い。また、データによって形式がバラバラであったり、専用のソフトウェアでないと読めないものもあるため、機械学習にかけられるように変換するツールを備え、さらに簡易的な解析ソフトウェアや可視化システムから、ビッグデータを用いた高度な材料性能の予測システムまでを提供する。まさにデータ創出から、データ活用、データ公開までが集約される、一貫した材料データプラットフォームである(右図参照)。

多機能なDICEを、いかにユーザーが使いやすい形で公開するか。重視しているのは「FAIRの原則」だと谷藤はいう。「FAIRとはデータの適切な公開方法を示した言葉で、Findable(発見)、Accessible(アクセス)、Interoperable(相互運用)、Reusable(再利用)の略です。データの存在、データベースの存在が発見できて、そこにアクセスできなければ利用できません

ん。また、たとえば企業が自社のデータベースと連結して解析したいと希望する場合、連結する方法が提供されて、かつデータ形式の互換性がなければ利便性が著しく低下します。そもそも再利用できる形でデータを保管しなければ意味がありません。データは人類共通の資産として、広く公開するという考え方が世界的な潮流となっており、FAIRの原則に沿ったデータベースの標準化が進んでいます」。

DICEでも、発見しやすくするために一意で永続的な識別子をデータに付与し、オープンな通信プロトコルを使用してアクセス性

を確保、相互運用のための共通した記述言語の採用、再利用のための明確なライセンス付与などを行うことでユーザーの利便性を高めている。一方で、材料研究はその国の競争力にも直結するため、セキュリティの担保も必要となる。「学術的な発展にはFAIRの原則を重視して、多くのデータを使いやすい形で公開すべきです。一方で、国の研究機関として産業競争力を高めるために、企業を含めたユーザーが安心して利用できることも重要です。このバランスを慎重に見極めてシステムを運用しています」(谷藤)。

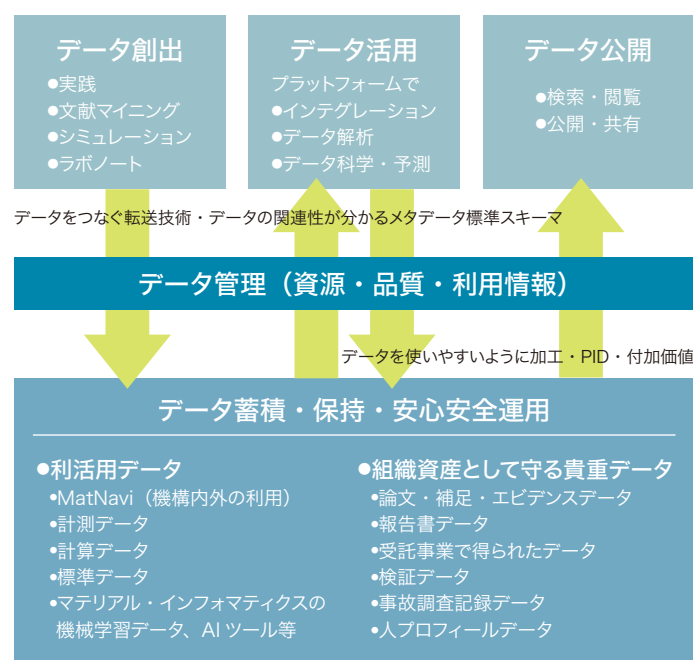


図. 材料データプラットフォームDICEの主要機能

DICEの核となる 主要4システムとは

1 進化する世界最大規模の 材料データベース:MatNavi

MatNaviは2003年から公開している世界最大規模の材料データベースである。高分子データベース(PoLyInfo)、無機材料データベース(AtomWork、AtomWork-Adv)、金属材料データベース(Kinzoku)や、NIMSの構造材料データベース(クリープデータシート、疲労データシートなど)など十数種類のデータベースに加え、金属偏析予測システム(SurfSeg)、界面結合予測システム(InterChemBond)などのアプリケーションシステムが提供されている。「プラットフォームを使っていたためには、まず高品質なデータベースがなければ始まりません。MatNaviは専門家の目で見極めた非常に質の高いデータを、20年近くにわたって収集・蓄積してきた世界でも類を見ないデータベースです」と谷藤。

しかし課題もあった。MatNaviを構成するデータベースは基本的にひとつひとつ独立しており、これまで横断検索はできなかったのだ。2003年当時を谷藤はこう振り返る。

「データをデジタルにして基本的に無償で公開することだけでも画期的なことで、同じ無機材料でも電子材料なのか触媒なのか等の分野ごと、研究テーマやターゲットごとにデータ形式や記述方法、構造が違うのは普通なことだったと思います。他人のデータや、他のプロジェクトとの互換性を意識することや、後から発見しやすくするために自分のデータに永久識別子としてのIDを振るという考え方は、この10年くらいで盛んに議論されるようになりました。FAIRという概念が世界で受け入れられ、研究者自身によって語られるようになったことが、データをデジタル情報として捉え、コンピュータで処理する対象へと意識が変わったことの象徴だと思います」。

そこでFAIRの原則に沿って、MatNaviにおいて横断検索を可能にするため、永続

識別子(ID)の付与が進められている。さらに最近では、企業がAIでの予測精度を上げるために、社外秘のデータベースにMatNaviを接続したいという要望が増えている。その要望に応えるために、データベース同士を連結するためのAPIの開発も進めている。「すでに高分子データベースは、機械学習に利用する用途を想定して、APIを開発しました。所外へ公開できるよう、共同研究などのスキームと合わせ、API認証方式の技術的な検証を進めているところです」(谷藤)。

2 研究ルーティンを 自動化するシステム:RDE

かつて、手書きのラボノートにしかなかった研究データは、いまリアルタイムでサーバに蓄積され、自動で解析・予測される時代になろうとしている。その先駆けとなるシステムが、RDE(Research Data Express)である(pl1も参照)。

実験装置で測定したデータを、自動で安全にデータサーバへ送信するシステムに加えて、実験の材料や条件を記録するラボノートも電子化することで、研究現場で生まれるあらゆるデータを自動収集する。ただし、装置から出てくる生データは、専用のソフトウェアでなければ読めないバイナリデータであることがほとんどだ。そのため、装置ごとに異なる用語やデータ形式を共通のフォーマットに変換するためのツール「M-DaC」も開発。さらに、データの簡易的な解析も自動で行うことで、研究者はデータを測定後、すぐに結果を分析することができる。希望に応じて後述するリポジトリ「MDR」でデータを公開し、論文出版時の参考資料としても引用可能だ。「質の高いデータをいかに効率的に、かつ継続的に集めるか。NIMSの強みは、材料開発を行っている現場があり、その現場のニーズを知っていることです。RDEが、研究現場でルーティン化しているデータの収集からグラフ化(簡易ビューアー)、数理統計解析までを行う一貫通のデータシステムとして、研究のスマートラボラトリ

化をサポートする「場」となることを目指しています」(谷藤)。

3 FAIRの原則を体現した リポジトリ:MDR

リポジトリとは、データを共有・公開する保管場所(オンライン・ストレージ)のことだ。NIMSは、実験から生まれる計測データや解析データから、解析に使われるツールやプログラムまで、様々なタイプのデータを共有・公開するリポジトリとして、MDRを運用している。共有の範囲も、共同研究者間で共有する方式(MDR Closed)と、DOIを付与してパブリックに公開する方式(MDR Open)の2つを用意し、さらに、出口での使いやすさを考えたデータ構造になるよう収集方法も設計されている。

「機械学習や試料合成など、用途や目的によってデータに必要な情報のニーズは異なります。そのため、ニーズに合ったデータ構造に変換して登録できるように、データキュレーターやデータアーキテクトといった専門職にも参画してもらっています。彼らの協力を得ることで、データを再利用しやすいように、世界で進むデータ通信方式、メタデータの記述形式に準拠した公開もスムーズに進み、まさにFAIRの原則を体現したりリポジトリシステムを構築することができました」(谷藤)。

4 産学官連携での データ活用を加速する:MInt

MInt(Materials Integration by Network Technology)とは、金属系構造材料の様々な課題について、AIや材料工学理論・経験則に基づくモジュールと呼ばれる多様な計算ツールを実装し、さらにモジュールを自在に接続したワークフローを構築することにより、プロセスから構造、特性、性能までを一貫通貫に予測するシステムである。例えば、火力発電所で使用される耐熱鋼の溶接部材について、高温高圧力下で使用される際の破断寿命を高い精度で推定することが可能となっており、何千時間もかかる実験を



数時間の計算で代替しながら最適な溶接施工条件を見出すことができる。

内閣府SIP（戦略的イノベーション創造プログラム）第1期「革新的構造材料」（2014年～2018年）においてコンセプトが提案・実証され、第2期「統合型材料開発システムによるマテリアル革命」（2018年～）においてMIntを産学官で活用していくために「マテリアルズインテグレーションコンソーシアム（MIコンソ）」が2020年12月に設置された。2021年10月現在、企業会員7社、アカデミア会員16研究室・センターが参画しており、MIコンソ会員はMIntを自由に利用して自社内での研究開発を加速できる。さらに、新しい課題の解決のために産学の共同研究が実施され、新規に開発されたモジュールやワークフローが搭載されていくことで、MIntシステムが進化することになる。「産学官で連携したデータ活用」の場としてだけでなく、データを活用することでプラットフォームが高度化する、そういっ

た好循環が生まれることも目指しています」（谷藤）。

データ共有を評価するカルチャーが欠かせない

研究者が安心してデータを預けられる。データを活用しやすい環境で提供する。データの活用が進むことで、さらに高品質なデータ蓄積がされていく。そうしたプラスのサイクルを加速するためには、データを提供した研究者を評価する制度と、データ共有を楽しみ歓迎するカルチャーの醸成が重要だと谷藤はいう。

「日本の研究現場は自分の研究データは自分だけに属する、という意識が強いですね。この意識を変えるために、たとえばデータ共有をあたりまえに行っているデータサイエンティストやデータアーキテクトに積極的に材料研究に関わってほしいと思っています。彼らが材料データを使ってなにか面白

いことを始めてくれば、データ共有への関心が高まり、プラットフォームの利用が活性化するのではないかと期待しています。そのためにも、このDICE自体が魅力的なコミュニティ、人が集まる舞台に育っていくことも大事です。そうやってプラットフォームが成長することで、特に若い人は、データサイエンス、マテリアルズ・インフォマティクスといった仕組みに対し、いまの研究者よりもっと柔軟に対応できるようになっていくでしょう。それが、データ共有を評価するカルチャーの醸成につながっていくと思うのです。私たちはそうした未来を担う若い研究者たちが興味を持つ空間をつくっている感覚があります。

自らの実験・計算データを自分以外の誰かに、そしてAIが学習できるデータとして登録し、FAIRで適切な運用のもとで、機械学習やデータ駆動の研究が当たり前になる世界はすぐそこだ。
(文:小森岳史)

材料研究の現場はこう変わる

データをリアルタイム収集し、解析まで自動化! RDEで研究スタイルが変わる



吉川 英樹

Hideki Yoshikawa

統合型材料開発・情報基盤部門
材料データプラットフォームセンター副センター長

データの質とともに、データ駆動型研究の成否を左右するデータの量。NIMSは、研究現場で日々生み出される膨大なデータをリアルタイムに収集し、データベースに蓄積するシステム「RDE (Research Data Express)」を開発した(下図参照)。RDEの導入は、研究現場のDX化推進だけでなく、研究スタイルそのものを変える可能性も持っているという。開発にあたったDPFCの吉川英樹副センター長に聞いた。

古い装置でも安全な通信を確保

材料に関する各種データは、基本的に材料の研究者が個々に実験装置を使って得ている。このデータを自動的にリアルタイムで収集できれば、データベース構築が容易となり、研究現場のDX化が推進される。データは装置に付属するPCに保存されることが多いため、そのPCからデータをデータサーバに送信することは一見簡単そうに見える。しかし、その実現には研究現場ならではの課題が立ちだかっていた。それが、「OSの古さ」である。

実験装置は高額なものが多く、購入から十年を越えても使用されるものも多々ある。装置は専用のPCで制御するが、そのPCのOSをアップデートすると装置制御に不具合

を生じる可能性があるため、多くのOSが購入当時の古い状態のまま使用されている。するとセキュリティの問題があるため、ネットワークには接続できない。そのため研究者は、測定データを自分のPCに移す際に、USBメモリなどに書き込んで取り出していた。

そこで吉川らは、OSの古いPCで管理している装置であっても、ネットワークに安全に接続できる方法を模索。外部から攻撃を受けず、かつPCがネットワークに接続されてOSのアップデートが自動で始まってしまうよう、装置からデータを蓄積するサーバへの一方方向の通信のみを行う「IoTセキュリティデバイス」を使ったデータ転送システムを開発した(次頁写真参照)。「このデバイスは、装置を制御するPCのLANポートに

接続するだけで一方方向の通信が可能になります。しかも、PC上の指定されたフォルダにデータを保存するだけで、自動的にサーバにデータが送信されるため、通信セキュリティを担保しつつ、ネットワークを介して各種装置から測定データを自動的に収集できるようになりました」(吉川)。

装置ごとの特殊な言葉を翻訳

自動収集できても、それだけでは“使える”データベースにならない。なぜなら装置から出力される各種データの用語やフォーマットは、装置の種類やメーカーによって異なり、しかも、各装置で測定されたデータは、装置専用PCの付属ソフトウェアでなければ読めない場合もあるからだ。そこで開発されたツ



写真: DICE を構成するサービスシステム群。各システムの詳細はこちら → <https://dice.nims.go.jp/>

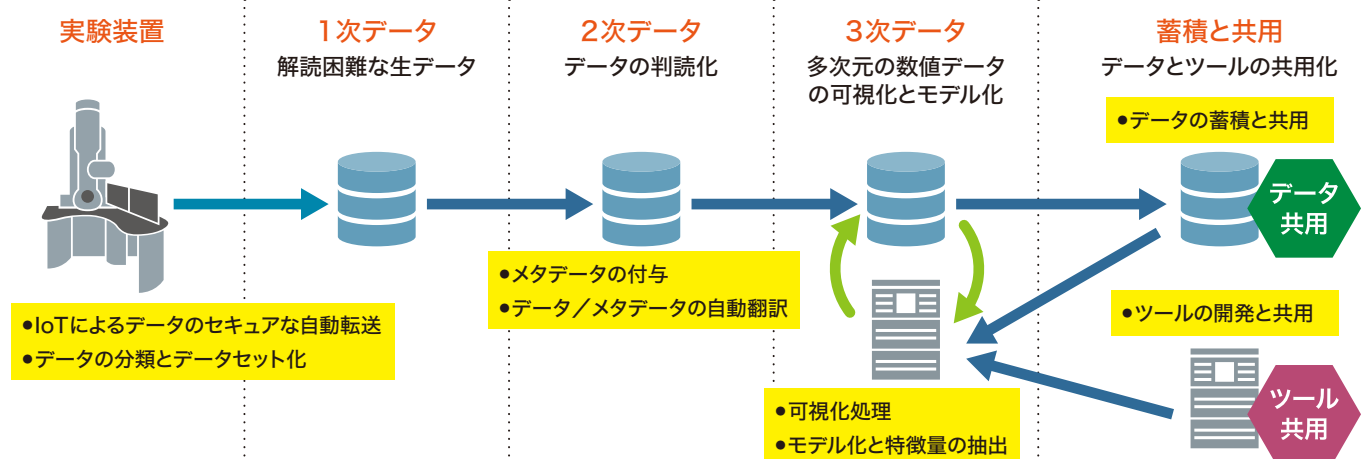


図: RDE におけるデータ収集から蓄積共有までの流れ



© 2021 Toshiba Infrastructure Systems & Solutions Corporation

一方向の通信を実現するIoTセキュリティデバイス

ルが「M-DaC」である。装置メーカーの協力も得て、データを専用ソフト以外でも読める形式に変換し、装置ごとに異なる用語とフォーマットを判読性と共通性に優れたものに翻訳することで、AIが利活用しやすい形で材料データプラットフォームに蓄積することができる。

さらに各装置から出力されるデータファイルには、測定結果の数値データに加え、

実験条件や装置の種類といった「メタデータ」も含まれている。メタデータはデータを検索し選別するのに必要で、AIでの解析には不可欠である。「データファイルから自動で収集できないメタデータについては、必要に応じて実験者が手入力する必要がありますが、人間が行う作業のためどうしても入力ミスや入力漏れが発生してしまいます。そこでそれらを防ぐために、ユーザーインター

フェスとしてHTML形式のテンプレートを作成し、これにプルダウンメニューを使用した値の選択や異常値を手入力した際のエラー表示などの機能を組み込み、メタデータ入力ミスや漏れの防止を図っています」(吉川)。

またRDEでは、データの収集だけでなく、処理や解析を効率化するシステムも提供する。たとえば、数値データの羅列を見た

だけでは、データの良し悪しの判断がむずかしい。そこで、研究者がその場で判断しやすいように、収集したデータをリアルタイムにグラフや画像として可視化している。さらに分光装置のスペクトルデータや顕微鏡の画像データなどの複雑なパターンを持つデータが大量にあり、研究者による迅速な判断が困難な場合、その複雑なパターンを数式で表現される数理モデルのパラメータに変換して保存することで、大量のデータを高速に検索し、かつ研究者の判断を自動化できるよう情報圧縮技術の開発を進めている。「今後、現場の声を聞きながら、機械学習やAIによるデータ解析ツールの開発も推進していき、研究者が材料の性能向上により専念できる環境を整えていきたいですね」(吉川)。

企業の研究現場にもデータ駆動型を浸透させる

2021年度からは文部科学省の「マテリアル先端リサーチインフラ (p.14参照)」事業の下、他の研究機関や大学、企業が取得し

た材料データも材料データプラットフォームに蓄積していくこととなった。特に企業では、自社で測定したデータは社外秘とされるのが通例だが、マテリアル先端リサーチインフラ事業の装置利用者に対しては、積極的に提供してもらうことを求めている。同時に翻訳ソフト「M-DaC」についても対象装置を拡充すべく、文科省の支援を受けながら、本事業で使用する装置のメーカーに対し、メーカー独自の用語やフォーマットで記述されているデータファイルの共同利用に関する合意に向けた交渉を進めている。

「材料開発における国際競争が激しさを増す中、基盤的なデータに関しては共用するメリットの方が大きいと認識する企業が増えてきています。より多くの企業にデータ共有のメリットを実感していただくためにも、データベースを充実させて、利便性を上げることが重要であり、現場の装置とシステムをシームレスにつなぐことの意味は増えています」(吉川)。

また吉川は、データを研究現場から自動収集し、メタデータを付与し翻訳した上で共用することの重要性として、そのデータを測

定した研究者の研究分野と関係がない第三者もデータを扱えるようになることで、従来は見逃していたような発見が期待できることもあると強調する。

「たとえば超伝導材料の研究者は、新たな材料を発見したとしても、超伝導性を示さない材料には興味を示しません。しかし、誘電体材料の研究者にとっては、それが極めて画期的な材料かもしれません。このように、セレンディピティが起きる可能性が高まるのが期待されるのです。自分の実験データに加え、他の研究者の実験データも同じプラットフォームで扱えるようになるため、RDEによって研究スタイル自体も大きく変わっていくことでしょう」(吉川)。

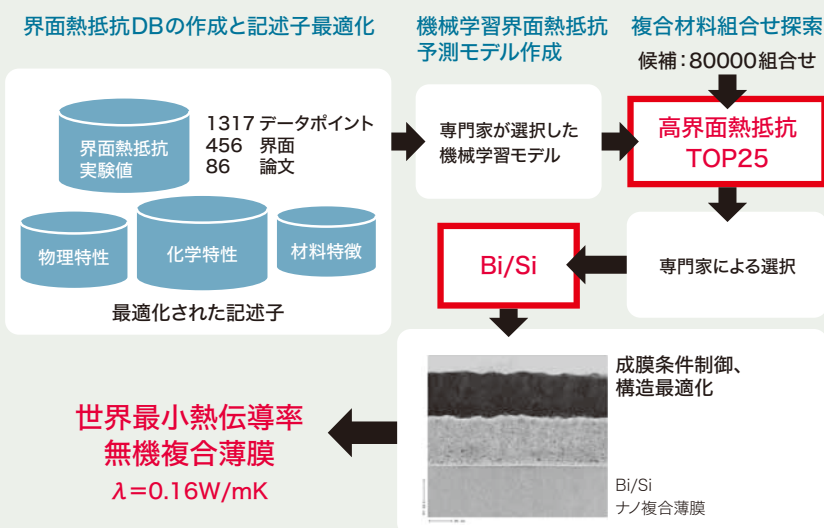
(文:山田久美)

豊富なデータで新材料探索①

機械学習を用いた世界最小熱伝導率ナノ複合材料の設計と開発

徐一斌
情報統合型物質・材料研究拠点 副部門長

エネルギー効率の向上のため、断熱材料の高性能化が求められている。徐は、既存のデータベースに加えて、断熱材料に関する80以上の論文から複数の界面熱抵抗に関するパラメータを抜き出し、AIに入力。どのパラメータが断熱性と関係があるか機械学習で探索させて、約80,000の候補材料の中から、熱抵抗の高い25種類を絞り込んだ。その後、合成のしやすさなどから、Bi/Siを選択。成膜条件を変えながらBi/Siを作成することで、世界最小の熱伝導率を持つ無機複合薄膜を実現した。

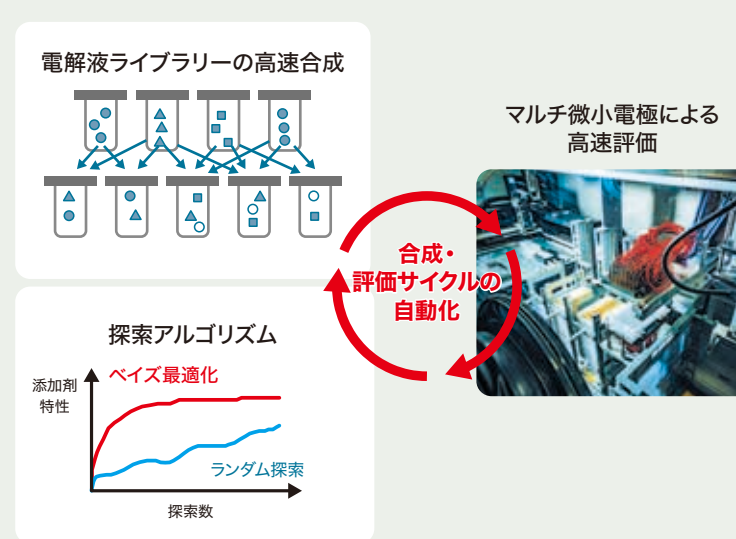


豊富なデータで新材料探索②

ハイスループット電解液探索システム

松田 翔一
エネルギー・環境材料研究拠点 二次電池材料グループ 主任研究員

脱炭素社会実現に向けたEVシフトが進むなど、蓄電池の高性能化への期待が高まっている。正極、負極とともに蓄電池の性能を左右する電解液だが、その候補は膨大で、人の手で試行錯誤を繰り返すアプローチにはすでに限界が来ている。そこで、松田は電解液探索の流れを自動化したシステムを構築。ロボットアームがサンプルを作成し、性能を自動で評価。そのデータを探索アルゴリズムに反映させ、導かれた候補材料から新たなサンプルをまた自動で作成・評価する。このシステムにより一日400サンプルもの評価が実現。リチウム負極用高機能電解液では16種の化合物から5種を選定し、新規の電解液組成を見出した。(NIMS NOW Vol.20 No.2も参照)





装置共用ネットワークをアップデートする

データの収集拠点を全国へ展開 「マテリアル先端リサーチインフラ」



小出 康夫
Yasuo Koide

物質・材料研究機構 特命研究員
マテリアル先端リサーチインフラ 運営機構長

研究現場で生まれる高品質なデータを収集・蓄積・共用する…この一連の流れを全国規模で展開するのが「マテリアル先端リサーチインフラ」だ。この事業にはこれまで20年間で蓄積してきた装置共用のノウハウが活かされている。

高品質なデータを、なるべく広範囲に、かつ効率的に収集する。その基盤となるのが、全国の大学、国立研究開発法人などの先端施設・設備を共用するネットワーク「ナノテクプラットフォーム」である。2012年に開始したナノテクプラットフォームは、装置利用は年間3,000件を超え、ユーザーも2万人を超えている。この共用基盤、人材、ノウハウを最大限活用して新たに立ち上がった「マテリアル先端リサーチインフラ」では、装置の測定データをリアルタイムで収集する技術 (p.11-13 参照) を活用し、全国の共用装置から生み出されるデータ群を収集・変換・蓄積し、データベ

スの拡充を図る。

「全国の共用装置を通じてデータを集めるというプロジェクトは装置の共用ネットワークが発達しているアメリカなどでも実はやっていない、世界で初めての取り組みです」とマテリアル先端リサーチインフラの小出康夫運営機構長は話す。

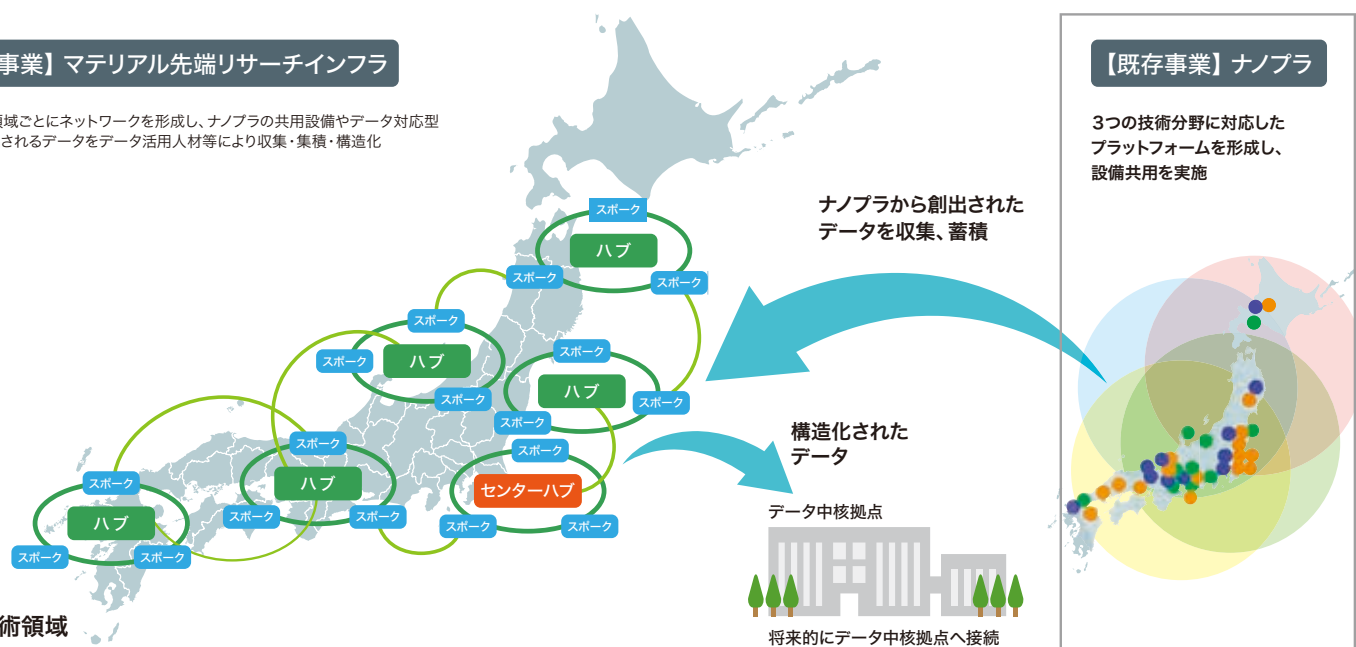
立ち上げの背景には、NIMSが進めてきた材料データの収集と活用、およびナノテクプラットフォームの実施機関および統括事務局として20年にわたって先端装置の共用事業を運営してきた歴史がある。最新鋭の装置を使うユーザーのデータを集めることで、非常に質が高く、幅広い

データが集まるという考え方が、今回のマテリアル先端リサーチインフラに反映されている。

インフラ全体はセンターハブ、ハブ、スポークが組み合わさった体制となる。スポークとは設備を使って創出された高品質データを収集する最前線である。ハブとは自らもデータを収集するとともに、ハブおよびスポークで創出された高品質データを蓄積し、データ変換、構造化などを担う。スポーク、ハブは重要技術領域ごとに体制が生まれ、各研究に特徴的な装置、データ創出を効率的に行う。NIMSはセンターハブとして、事業全体の運営を最適化する

【新規事業】マテリアル先端リサーチインフラ

重要技術領域ごとにネットワークを形成し、ナノプラの共用設備やデータ対応型設備で創出されるデータをデータ活用人材等により収集・集積・構造化



【既存事業】ナノプラ

3つの技術分野に対応したプラットフォームを形成し、設備共用を実施

重要技術領域

- 高度なデバイス機能の発現を可能とするマテリアル
- 量子・電子制御により革新的な機能を発現するマテリアル
- 革新的なエネルギー変換を可能とするマテリアル
- マテリアルの高度循環のための技術
- 次世代バイオマテリアル・次世代高分子マテリアル
- 次世代ナノスケールマテリアル
- 極限機能を有するマテリアル
- マルチマテリアル化技術

事務局を担うことになる (p.14の図)。

プロジェクトの成功には、データ創出の現場とデータ中核拠点とが手を取りあうことが必須だ。「データそのものをどう扱うのか、データ構造、変換方法、生データの扱いやデータ活用までを熟知した上で先端装置のデータ収集を行わないといけない。これはトップダウンでは絶対うまくいきません。センターハブ、ハブ、スポークの各代表者とのコミュニケーションを密にとり、

データを提供する側、活用する側の意見にしっかりと耳を傾けて事業を進めていきます」(小出)。

今回のプロジェクトの事業期間である10年間を見据え、体制作りが進んでいる。「10年経てばコンピューティングパワーや先端装置自体など、環境も大きく変わってしまう。それに耐える体制やデータ創出をしなければいけない。一方で、社会実装される研究は20年、30年かかるものが当

たり前です。それをいかに短縮させることができるかが、データ駆動型研究の本質であり目的だと思います。10年は長いようで、あっという間。その間に社会に実装できる材料を日本が先陣を切って、このマテリアルDXプラットフォーム全体でつくりあげたいと思っています」(小出)。
(文:小森岳史)

材料創製のプロセスまでを含めたDXには人の力がカギになる

マテリアル先端リサーチインフラは、材料研究をデータ駆動型へアップデートさせるために不可欠なものだ。人の力もうまく取り入れる仕組みを作り、日本の競争力維持のためにも次期インフラづくりに取り組まなければいけないと、マテリアル先端リサーチインフラの曾根純一プログラムディレクターは語る。



曾根 純一 氏
Junichi Sone

国立研究開発法人科学技術振興機構 研究開発戦略センター 上席フェロー
マテリアル先端リサーチインフラ事業 プログラムディレクター

1975年、東京大学大学院理学系研究科修士課程を卒業後、日本電気株式会社に入社、中央研究所研究員。1983年、理学博士。日本電気株式会社基礎研究所長、基礎・環境研究所長を経て2007年に日本電気株式会社中央研究所支配人。2010年、物質・材料研究機構の理事(現・名誉理事)、2015年、科学技術振興機構の研究開発戦略センターの上席フェロー、現在に至る。2012年から2016年まで日本ナノ学会会長。2008年に応用物理学会フェロー受賞。

日本の輸出品目をみればわかるように、素材、材料のウェイトは非常に高い。産業競争力や経済安全保障の意味でも材料研究は重要なものです。一方で、この流れを今後も維持するためには昔の材料研究の手法ではもう限界に来ていることも確かだ、DXはより必須になってきます。

これからのデータ駆動型研究を見据えたときに、より使えるデータを収集・蓄積するためには現場で出てきたビビットなデータをどのように効率よく集めるのか、また、データの変換・構造化とともに、アクセスのしやすさまでを考えた設備共用のありかたが重要になってきます。

ナノテクプラットフォームでは最新鋭の設備を共用することで、日本の材料研究を支えてきました。こうした横断的な設備共用からさらにジャンプアップし、マテリアル先端リサーチインフラでは日本全体の材料研究のDX化に寄与するためのデー

タ創出基盤としての役割が求められています。そのためハブやスポークの体制であり、重要技術領域の設定になります。これらの技術領域は、日本として絶対負けるわけにいかない7つの重要な研究領域となります。

材料設計は材料探索から始まりプロセス設計までできて、はじめて機能します。第一原理計算をして、材料の組成や構造を決める手法が頻繁に使われるようになってきましたが、その材料をどのようにつくるのかのプロセスが一番難しい。プロセスの手法、パラメーターもそれぞれ膨大です。そこまで含めてのマテリアルDXは、とてもチャレンジングになります。

この高いハードルを越えて成功するために、わたしは人材もたいへん重要だと思っています。データとコンピューティング能力だけでなく、DXの世界にHuman In The Loopという言葉があるように、

デジタル技術の世界に人のチェックや考えもうまく取り入れていくことが重要になってきます。

マテリアル先端リサーチインフラでは、ナノテクプラットフォームで培った最新鋭の装置・インフラを最高の形で使いこなすエキスパートが現場にいます。こうした人の技術をマテリアルDXに組み入れることが、データ創出においては非常に重要だとわたしは思います。

データ創出・利活用の取り組みは産業界にとっても大きなメリットになります。企業が持つ自社のデータと我々が集めたデータを合わせることでまた別の視点を持つことができるようになり、新たな材料創製に結びつく。産学連携をより進めることで、日本全体の材料の産業競争力の維持・向上に結びつくことを確信しています。

「未来の科学者たちへ」が DVDブックに!

絶賛
発売中!

No. 5

マテリアルDXの真価

このスプーンは、 結構うるさい

科学映像集
DVD-Book

物質・材料研究の最前線!



佐藤雅彦 + ユーフラテス
NIMS (物質・材料研究機構)

「ピタゴラスイッチ」のスタッフが作る、
楽しくも興味深い科学映像8本 + 特典11分



—— 未来の科学者たちへ

第55回 科学技術映像祭
文部科学大臣賞 受賞

小学館

注文はこちら



厳選映像8本 + 解説ブック、

さらに! この本のためだけに撮り下ろした未公開映像7本!



NIMS NOW vol.21 No.5 通巻190号 2021年11月発行
国立研究開発法人 物質・材料研究機構



古紙配合率 70% 再生紙を
使用しています



植物油インキを使用しています