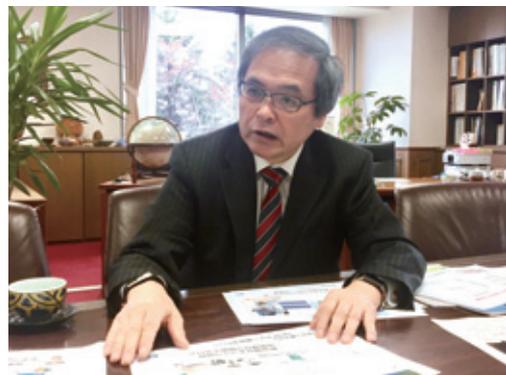


特別インタビュー

国立研究開発法人物質・材料研究機構 理事長／総合科学技術・イノベーション会議 議員 橋本 和仁 氏 インタビュー —マテリアルズ・インフォマティクスがリードする材料研究・開発、そして NIMS の戦略—

聞き手：科学技術・学術基盤調査研究室 研究員 松本 久仁子
科学技術予測センター 特別研究員 蒲生 秀典、上席研究官 林 和弘

第5期科学技術基本計画における Society5.0 はデータ駆動型社会への移行を示唆している。国立研究開発法人 物質・材料研究機構 (NIMS) が取り組むマテリアルズ・インフォマティクス (MI) は、材料科学とデータ科学、そして人工知能 (AI) の融合という新たな研究手法である。これは、従来の研究にどのような変化をもたらすのか、材料開発が加速する未来とはどのようなものか、日本の政策への期待とともに、NIMS 理事長 橋本 和仁氏に、MI がリードする NIMS の研究開発戦略についてお話を伺った。



橋本 和仁 国立研究開発法人物質・材料研究機構 理事長／
総合科学技術・イノベーション会議 議員

I. MI の進展で変わる材料科学研究の姿

材料科学研究は大きな変革期に来ていると感じています。2017年頃からと言ってよいと思いますが、MI に対する材料分野の研究者や技術者の認識が大きく変化し、その有効性が現実感を持ってとらえられるようになってきました。特に、日本の材料関連メーカーでは、このままでは欧米の AI やデータ関連企業の下請企業となってしまうのでは、といったほどの危機感を持つところも現れてきています。この背景には、次の2つの材料データをめぐる潮流（国際競争の激化）があります。

一つは IBM や Google を代表とする AI に強みを持つ企業の材料開発分野への参画です。彼らは材料データを企業などから収集し、それらを用いて高度な AI 解析を行う。そして、データの価値を高めて高額で販売し、さらに企業が得た新たなデータの提供まで義務付ける、といったビジネスモデルの構築を狙っています。

もう一つは、エルゼビアを代表とする出版社や欧米化学会の傘下にあるデータ関連企業です。彼らは論文投稿時に著者からデータを提出する仕組みを構築することにより大量の学術データを保有し、専門家によるデータベース化や MI 用データセットの作成など、

データの利用価値を高め、データベース会社や IT 企業と連携してビジネスにつなげようとしています。日本の学術界はデータの囲い込みを感じつつも、企業ほどは危機感を持っていないのが現状です。

II. MI 研究における NIMS の強み

マテリアルデータのプラットフォーマーとして、NIMS は前述の AI 関連企業や、出版社等のデータの囲い込みを進めている企業との競争に挑んでいます。これらの企業に対する NIMS の優位性は、長年にわたる質の高いデータの蓄積と、実験研究者と MI (AI) 研究者の密接な連携にあると考えています。NIMS と同様の戦略を考えている海外機関（例えば米国国立標準技術研究所 (NIST)）もありますが、NIMS も十分に競争できるレベルにあると考えています。

NIMS では2種類のデータ収集活動によって、精度の高い世界最大級のデータベースを構築しており、MI の開発も同時に進めています。一つは数十年前から取り組んでいる、論文から材料情報をデータとして収集する方法です。専門家の目利きによる質の高いデータの蓄積が特徴です。もう一つは、論文にする前

の実験データを直接ストックしていく方法です。最新のデータがどんどん蓄積されていきます。

また、NIMS では、材料開発の研究者と情報科学の研究者が共同で研究することにより、MI で設計したものを即、実験で確認することができ、更に実験結果を MI 研究者にフィードバックするという、MI と実験の迅速な連携によるデータの構築が可能となっています。これは、NIMS ならではの強みであると考えています。

また、マテリアルの実験データは、量としては限られているため、MI によって優れた材料を開発するためには、量だけでなく質の高いデータが重要になります。例えば、NIMS では、40 年間かけて得られた高品質試料と高信頼解析データをもとに得られた予測精度の高い合金設計プログラムを開発しており、耐熱性の高い Ni 超合金の開発が可能となっています。

さらに、NIMS では自機関の強みを活かしながら、その一方で AI とデータ科学の手法を積極的に進め、外部機関と連携しながら様々な取り組みを行っています（図表 1 参照）。

データ構築の部分については NIMS 独自の交付金や公的資金のプロジェクトとして行うとともに、また、共同研究として理化学研究所（理研）や大学等の外部機関との連携もしています。

具体的には、国立研究開発法人科学技術振興機構（JST）のイノベーションハブ構築支援事業のプロジェクトとして情報統合型物質・材料開発イニシアティブ（Mi²i）、内閣府のプロジェクトとして戦略的イノベーション創造プログラム（SIP）「統合型材料開発システムによるマテリアル革命」における研究開発課題「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」、更に複数の化学メーカーと水平連携で取り組んでいる MOP（Materials Open Platform）があります。

Ⅲ. MI を支える NIMS のデータ収集活動

1. 論文からの材料データの収集と課題

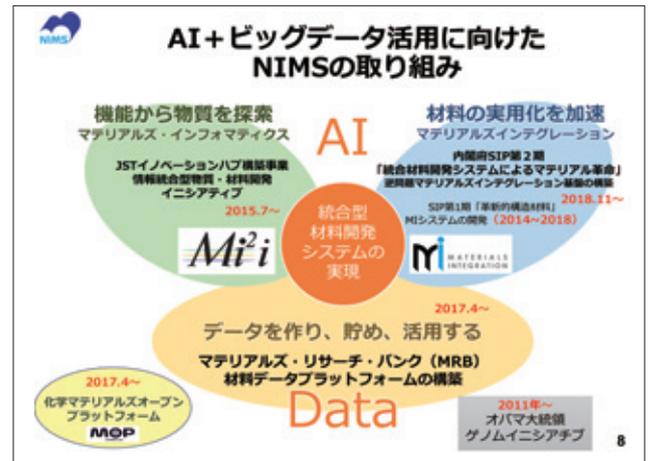
NIMS の高分子データベース PoLyInfo の構築に当たっては、これまでは論文から材料データを人が手動で入力していましたが、論文を AI を使ったテキストマイニングで自動で入力することも始めています。図表 2 は、61,000 の論文から収集した、ガラス転移温度と融点のデータをプロットしたものになります。人が読んだデータのプロット（灰色）と自動入力したデータのプロット（オレンジ色）はほぼ重なっており、同じように読めていることがわかります。自動入力を進めることにより格段に速いペースでデータを増やすことができます。NIMS のデータベースは

質が高いため、世界的にも注目されており、また、日本の産業界からも非常に興味を持たれています。

このような大規模かつ高品質なデータベースを構築していく上で、費用面での大きな課題があります。テキストマイニング研究のために、人が読む論文データ (pdf) に加えて機械が読む論文データ (xml 等) を多額の費用をかけて購入、整備しています。論文データは出版社や学会から購入しますが、NIMS 研究用途に限られます。企業との共同研究に活用していこうとすると、恐らく桁違いの費用を出版社に払う必要が出てきます。現在、PoLyInfo のデータの一部は無料公開していますが、基本は有料で利用していただくシステムになっています。今後、更にテキストマイニングによるデータを公開していく際には、別の枠組みを作り、論文データ購入分は企業側で御負担していただくようなシステムを作る必要があります。

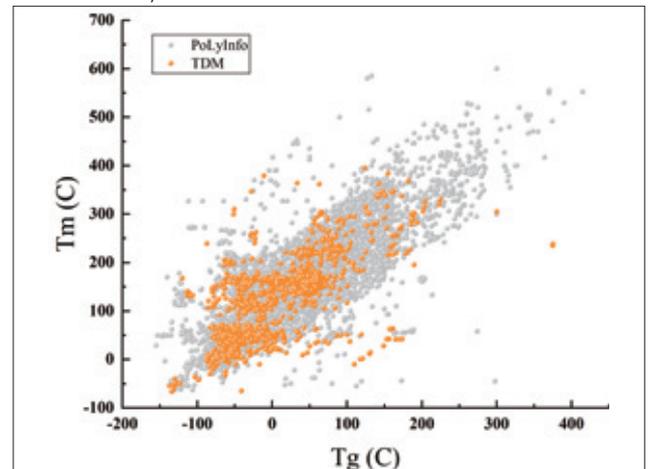
また、データベースの維持にも、膨大な費用がかかります。現在は文部科学省の予算で、10 ペタバイトの

図表 1 AI+ ビッグデータ活用に向けた NIMS の取り組み



出典：NIMS 橋本理事長御提供資料

図表 2 PoLyInfo の人間と機械によるデータ収集プロット



(注) Tg：ガラス転移温度、Tm：融点
TDM：Text Data Mining

出典：NIMS 石井氏御提供資料

サーバを入れ建屋や維持費も賄っていますが、5-10年後には更新が必要となる見込みです。しかし、今のところ予算の充てはなく、自ら稼がないといけない状況です。

NIMS は国の研究所ですので、金銭的な協力をお願いするとすれば、政府や日本の企業になるでしょう。そのためにも、NIMS へ投資を呼び込める世界に誇れる良いものを作っていくことが必要と考えています。

例えば、テキストマイニングでは材料の合成プロセスも読み取ることができるので、NIMS では単にデータ（組成・物性）を集めるだけでなく、その作成条件、すなわち材料レシピのデータベースも作成して、更に付加価値をつけていくことを始めています。また企業が NIMS のデータと独自のデータを組み合わせる場合は、ライセンス契約を結ぶ等の付加サービスで利益を出すことも考えています。

2. 計測装置からの材料データの収集

論文からではなく、実験データを直接データとして保存し、データベースを作る取り組みも始めています。NIMS の大型測定装置では、セキュリティを担保した上で、実験データを自動的にサーバへ格納するシステムを構築しています。

実験装置によってデータの出力形式は様々ですので、実験データを蓄積していくには、保存する際のデータ形式が重要になってきます。すべての実験装置のデータ出力形式を統一できればよいのですが、世界中の装置メーカーに一律の協力を求めることは困難です。そこで、NIMS では翻訳機（MDAC 翻訳システム）を作成することにより、統一した形式でのデータ蓄積を試みています。

現在、アルバック・ファイ株式会社（XPS^{注1} データ）と株式会社リガク（XRD^{注2} データ）から協力が得られ、データを Web 公開しています。その他のメーカーにも協力をお願いしているところです。そうすれば NIMS の持つ様々なデータと他のデータを合わせて解析して材料開発することも可能となります。そのような利点を示すことで、今後、更に積極的な協力をお願いしていくつもりです。

IV. MI 技術の活用と企業への展開

1. MI 人材の確保

前述のように最先端の MI システムを構築していくには、情報科学の研究者だけでなく、MI での推定

結果をもとに実際に実験で合成するというプロセスが有効となるため、実験研究者と一緒に取り組むことが極めて重要と考えています。

NIMS には元々情報科学の研究者はいませんでした。MI に取り組むに当たって、理論の研究者に転向を促す、共同研究を通じて他機関の情報科学の研究者と組む、あるいはクロスアポイントメントで NIMS に所属してもらう、更に新規に AI 研究者を採用したりと、様々な方法で積極的に人材確保に取り組んでいます。宇宙物理や素粒子物理で大量のデータを扱うデータ科学者の採用にも成功し、成果を上げています。日本ではデータ科学者そのものが少ないと言われていますが、分野や国を越えて見てみると、高いデータスキルを持つ科学者も相当数いると感じています。

2. MI 活用の事例

一般的には、PoLyInfo にあるデータを記述子に直して、例えばデータ量の多いガラス転移点と融点を記述し、深層学習とベイズ推定により候補分子を多数作り、その中から合成可能と推測される候補を選定し、材料開発につなげています。

一方、データ量の少ない物性に特徴を持つ材料の探索も可能です。その一例として、熱伝導率の高い高分子材料の開発があります。高分子材料の熱伝導率は一般的に低いですが、高いものを開発できれば携帯端末などに広く利用できる有望な材料になります。数万ある PoLyInfo のデータの中でも熱伝導率のデータはわずか 8 個しかありません。この場合、転移学習という他で学習する方法で 100 個ぐらいの推定分子を作り、その中から合成しやすいものを選定し、最終的に 3 種類の物質を合成しました。その結果、3 種類合成したうちの 2 つで圧倒的に熱伝導率が高いものが得られました。これは、MI 活用による効果的な材料開発の成功例と言えます。

3. 企業への展開

現在、NIMS では、材料データプラットフォームの構築に向け、材料メーカーに対して、MI 活用の事例等の説明を積極的に行っています。多くのデータを集められたところが、プラットフォームとして優位な地位を確立できることになるので、大きな費用負担は要求せず、興味を持ってもらうことを優先に考えています。

数年前から日本の主要化学メーカー 4 社と共同研究をしていますが、そこでは互いに切磋琢磨できるオープンな領域づくりを行っています。例えば、各

注 1 X 線光電子分光法

注 2 X 線回析法

社から提供してもらった高分子材料をNIMSが最先端の装置で解析し、データベースを構築することで、各社の材料開発に利用できるような仕組みを構築しています。金属材料の場合と同じで、信頼性の高いデータベースを自分たちで作って、それをもとにMIを行っています。

また、経済産業省の協力を得て、特許データをもとにした高分子データベースの作成を進めようとしています。企業連携を通じて、特許データからデータベースを作っていきます。化学系企業では各社が独自で高分子特許データベースを構築し分析をしているのですが、協力して作成すれば何十倍もの速さでデータベースを構築できるはずで、ただし各社独自のやり方があるので、共通のデータベースの構築が必要です。それについては論文で積み重ねたPoLyInfoでのノウハウを活用できると考えています。

特許は論文以上にノイズの多いデータもあるので、キュレーションをしっかりとやってクオリティが高いものを作ります。これも各社の特許専門家が見たキュレーションであれば、非常に高いクオリティになるはずで、現在、20社程度が賛同していますが、更に参加企業は増えていくことが見込まれます。これらで集めた特許データも、NIMSのデータベースに入れ、PoLyInfoと連携させることを目指しており、2020年のうちにはスタートしたいと考えています。

このようなNIMSの取り組みを、まず日本全体に広げ、更に世界の企業とも価値を認めて協力できることは協力していきたいと考えています。今は高分子材料が対象ですが、成功すれば、他の材料分野にも展開していくロールモデルになることが期待されます。日本全体でのデータサイエンスの取り組みのさきがけとなることを目指しています。

V. 第6期科学技術基本計画に向けて

— 最後に、総合科学技術・イノベーション会議(CSTI) 議員としてのお立場も踏まえて、第6期科学技術基本計画に向けた考えをお聞かせください。

第5期を作成したときとの大きな違いは、変化のスピードにあると考えています。これは科学自体の変化のスピードが速いだけでなく、地政学的にみた国家間の関係の変化も留意する必要があるのではないのでしょうか。もう一つ、情報科学と様々な科学の融合の臨界点のようなものが近づいているのではないかと

も感じています。

このように多くの視点がある中で、私は基礎研究力の強化を緊急の課題の一つとして取り扱いたいと考えています。現在、研究現場で抱えている最大の課題は将来の見通しが不安なことだと思います。研究者自身の将来、そして研究費の見通しの双方に対して不安が渦巻いています。第6期はそれらの不安を解消できるようなビジョンを提示して、具体的な工程を示すことが重要と考えています。

例えば、ファンディングシステムでは、運営費交付金が下がり続けている間は、対症的に競争基金で補ってききましたが、運営費交付金が下げ止まりかけている現状を踏まえ、運営費交付金と競争的資金とを織り交ぜたトータルデザインされた研究開発戦略が必要でしょう。しかし、研究資金は常に動いているものなので、すぐに制度を変えるのではなく10年先のビジョンを見据え、それを共有しながらコンセンサスを作ることが重要になると思います。

また、冒頭申し上げたように、科学自体が大きく変化していることが重要な観点となります。東京大学の進学振り分け^{注3}でも化学系の学科が軒並み定員割れを起こしていると聞きました。それはある意味時代を表していて、その変化を踏まえた科学を牽引する画の設計が必要になります。

その際、情報科学とドメインサイエンスの融合は大前提であり、さらに長期的な視点で、大学も変わり学生も変わることを踏まえなければならないと考えています。

— 橋本理事長、お忙しい中貴重なお話をありがとうございました。

インタビューを終えて、
NIMSの研究本館ロビー（元素周期表）の前にて



左から松本、林、橋本理事長、蒲生

注3 東京大学では、3年次に上がるときに、成績に応じて希望する学部・学科に振り分けられる。