

国立国会図書館 調査及び立法考査局

Research and Legislative Reference Bureau
National Diet Library

論題 Title	第2章 マテリアル科学とデジタルトランスフォーメーション
他言語論題 Title in other language	Chapter 2, Materials Science: Digital Transformation
著者 / 所属 Author(s)	門平 卓也 (KADOHIRA Takuya) / 国立研究開発法人物質・材料研究機構技術開発・共用部門ユニットリーダー
書名 Title of Book	マテリアル科学—最先端と未来への選択肢— 科学技術に関する調査プロジェクト報告書 (Materials Science: The State of the Art and Future Options)
シリーズ Series	調査資料 2023-6 (Research Materials 2023-6)
編集 Editor	国立国会図書館 調査及び立法考査局
発行 Publisher	国立国会図書館
刊行日 Issue Date	2024-3-27
ページ Pages	24-40
ISBN	978-4-87582-924-9
本文の言語 Language	日本語 (Japanese)
摘要 Abstract	マテリアル科学におけるデジタルトランスフォーメーションの現状、国際動向を踏まえ、データ記述法の発展、データ管理・利用の共通ルールの策定、人材育成等の重要課題について概観する。

* この記事は、調査及び立法考査局内において、国政審議に係る有用性、記述の中立性、客観性及び正確性、論旨の明晰（めいせき）性等の観点からの審査を経たものです。

* 本文中の意見にわたる部分は、筆者の個人的見解です。

第2章 マテリアル科学とデジタルトランスフォーメーション

【要旨】

本章では、マテリアル科学におけるデジタルトランスフォーメーション（DX）の現状、国際動向を振り返り、未来への選択肢として重要な課題について概観する。日本が世界と比較しても競争力を有するマテリアルは、新しい産業や付加価値が生まれる分野として重要であり、データ活用とDXにより競争力を更に高められる。鍵はAIの活用とそのためデータの管理にあり、それを実現するための不可欠なインフラストラクチャとしてデータプラットフォームが位置付けられる。その機能は「つくる」「ためる」「つかう」の大枠で分類され、さらに、マテリアル分野で広く受け入れられている共通見解に基づいてデータ共有や活用が効率的に進められる。各国の動向として、データプラットフォーム構築に関する活動が活発化しており、データ記述法の発展、データの管理と利用にかかる共通ルールの策定、データを使いこなせる人材の育成が共通する重要課題である。

はじめに

前章で、社会の基盤であり、日本が世界と比較しても競争力を有するマテリアルは、新しい産業や付加価値が生まれる母体として、我が国の研究開発上、極めて重要であることを見てきた。

そうした中、マテリアルの研究開発の在り方を革新しているのがデータ活用であり、デジタルトランスフォーメーション（DX）と言われる所以（ゆえん）である。結論めいたことを先に述べてしまうと、DXは研究開発の効率化やスピードアップのみをもたらすのではなく、現在のマテリアル科学そのものの在り方・考え方を根本的に変え、更なる競争力を獲得する契機となり得るものである。

本章では、マテリアル科学におけるDXを概観するために、データ駆動型研究を実施するための必須インフラとなるデータプラットフォームの機能と役割について整理する（I節）。さらに、関連する取組を海外動向も含めて振り返り（II節）、未来に向けて解決すべき課題にはどのようなものがあるかを見ていく（III節）。

I 材料データプラットフォームの機能と役割

1 「つくる」「ためる」「つかう」

本節では、まず、現状認識として、マテリアル分野におけるデータ活用のためのプラットフォーム（データプラットフォーム）とはどのようなものなのかについて概観する。世界中でこうしたデータプラットフォームの必要性が認識され様々なシステムが構築されているが、ここでは日本で推進されている「マテリアルDXプラットフォーム構想」⁽¹⁾の下、整備が進められている材料データプラットフォーム⁽²⁾構築を一例として取り上げ、その機能と役割について

* 本稿におけるインターネット情報の最終アクセス日は、令和5（2023）年10月24日である。

(1) 内閣府統合イノベーション戦略推進会議「マテリアル革新力強化戦略」の中で、「データを基軸とした研究開発プラットフォーム（マテリアルDXプラットフォーム）の整備とマテリアルデータの利活用促進」が目標として掲げられている。

(2) 「DICEとは」国立研究開発法人物質・材料研究機構ウェブサイト <<https://dice.nims.go.jp/about.html>>

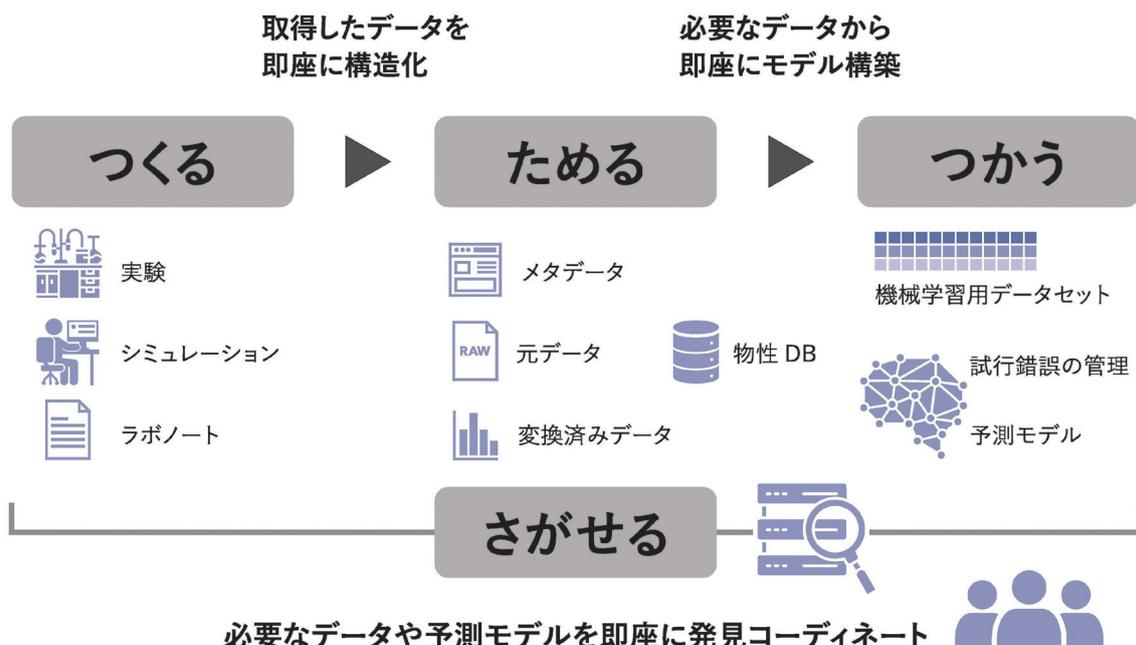
で説明する。

一言で言えば、プラットフォームの主要な機能は、データを作成する（データをつくる）、データを保存する（データをためる）、そしてデータを利用する（データをつかう）ことである（図1）。これらの機能は、マテリアル科学におけるデータ駆動型研究開発をサポートするために設計されている。

データ駆動型研究開発では、従来の学問的な知識だけでなく、データを活用して新しい材料や物質あるいはそれらを制御するプロセスを発見する。このようなデータを活用するアプローチは、機械学習などの技術の進歩を背景に様々な分野で成功を取めている。つまり、マテリアル DX プラットフォームが目指すのは、データを活用して新しい材料の創製や物性の発現をシステムティックに予測し、研究活動を効率化することである。

この目的を達成するために、データを利用する（データをつかう）段階に先立って、データを生成し（データをつくる）、それらを適切に保存し（データをためる）た上で、最終的に利用するプロセスを整備する必要がある。

図1 材料データプラットフォームの機能の分類



(出典) 筆者作成。

データを生成する段階では、実験結果や計算結果を得ることが含まれる。これらは、物質の性質や挙動を理解するための基本的な情報源となるものである。また、ラボノートなどの記録もデータ生成の一部であり、管理の対象となる。ほかに研究活動において生成されるデータとして、研究の成果を最終的にまとめて文書化した「論文」がある。

生成されたデータは、保存する段階に進む。実験結果、計算結果、ラボノートなどは、日々の研究活動の記録であり、後から探し出せる形にして保管することが重要である。ここでいう「後から探し出せる」とは、いつ、誰が、どんな研究テーマで、といった情報を基に探し出せることのほかに、この後に控える「データを利用する」段階の観点から必要なデータを同定で

きることも含まれる。例えば、対象物質の構造を知るためのX線回折実験において、実験結果として得られるスペクトルの数値列などが探索対象となり、それらを指定して探せるようにしておくことが重要である。こうした探索を可能にするには、そのデータ内容を説明する「メタデータ」をあらかじめ整理しておく、さらに、検索後アクセスしたときにデータを利用しやすい形にあらかじめ変換しておく、などの前処理が必要であり、材料データプラットフォームはそれらを容易に行うための機能を有している。

ほかにも、生成されたデータからよく参照される値を抽出し、それらの値をデータベース化することも、データを保存する段階に含まれる。例えば、物質ごとの特性を抽出してまとめた「物性データベース」はその一例であり、物質ごとにその特性をまとめて検索できるため、マテリアル研究者が情報を取得するための便利なツールである。また、最終成果としてまとめられる論文は、出版を通じて全世界に発信されるほか、研究機関や分野ごとに成果を集約・管理する「学術成果リポジトリ」に登録されて、世界中からその成果にアクセスし利用できるように保管されることが一般的になりつつある。

こうしたデータをつくる・データをためる工程を経た後、最終的に、データを活用する段階では、材料データプラットフォームはデータを効果的に利用するためのツールとして振る舞う。提供される機能については次節で詳しく述べるが、データ駆動型研究開発を支え、新しい材料や物性の発見に貢献する役割を果たす。

以上のように、材料データプラットフォームはデータの生成、蓄積、利用といった重要な機能を提供する。そこでとられるデータ駆動型のアプローチは、新たな材料や物性の探求において、従来の学問的なアプローチを補完し、より効率的な研究を実現する道を拓き、マテリアル科学の発展に寄与している。

2 機械学習の援用にフォーカスした「データをつかう」の深掘り

次に、「データをつかう」について、機械学習という観点から深掘りしてみよう。

機械学習は、データの傾向やパターンを学び、その情報を予測や判断に活用する技術を指す。この技術を適用するためには、関連性を保ちながら同じ種類のデータを多数収集する必要がある。具体的には、図2に示すようにデータが行列の形で整理されていることが望ましい。ここで、各行は1回の計測や計算に対応する一連の関連した情報を表し、各列はその計測や計算の内容を説明するデータの種類を示している。

このようなデータ行列を作成するために、同じ方法で条件を変えて何度も計測や分析を行うことがある。一方で、マテリアル分野の計測はナノメートルスケールに至る精緻な観察を行うための特殊な装置を利用するなど高コストな計測となるものも多く、結果、多くのデータを得ることができない場合も多い。つまり「ビッグデータ」ではなく、「スモールデータ」となる。こうしたデータはそのまま機械学習にかけられず、測定データをオリジナルとして、様々な工夫を凝らして学習用のデータを構築する必要がある。具体的な工夫としては、シミュレーションや物理法則などの既存知識を活用してデータを補完する方法などが含まれる。どの方法を用いてデータを準備するかは研究者の選択にかかっており、試行錯誤する必要がある。

図2 機械学習に適した行列形式のデータとマテリアル分野における課題



(出典) 筆者作成。

この試行錯誤には、データの準備の先にある、そのデータを機械学習にかけて「予測モデル」を構築する段階も含まれる。例えば、どの入力データを使うか、どのパターン認識手法を採用するか、ハイパーパラメータ（モデルの挙動を調整するパラメータ）をどう設定するかなどが試行錯誤のバリエーションを生み出す。そして、構築したモデルごとに、十分な予測能力を有するかの評価も重要であり、予測の精度や安定性の確認が行われる。

予測能力を向上させるためには、複数のバリエーションを試して最適なモデルを選択することや、新たなデータを収集してモデルを改善することが必要である。新たなデータを収集する際には、機械学習を活用して次にどのデータ点を計測すべきかを予測する場合もある。

以上のとおり、機械学習を想定したデータ活用の段階では、モデルの構築や評価、改善などを迅速に行い、その試行錯誤の内容を簡潔に整理できる場が必要となる。こうした場合は、マテリアル分野であるか否かによらず、モデルの構築 (Machine Learning: ML) から運用 (Operations: Ops) までを一体的に行える場として、MLOps の名で一般的に知られており⁽³⁾、「データをつかう」段階では、こうした好環境の実現がデータプラットフォームには求められる。

3 マテリアル分野において「つかう」、そのための「つくる」「ためる」の工夫

材料データプラットフォームには、これまで概観してきたような「データをつくる」「データをためる」「データをつかう」に関するシステム機能が備わっている⁽⁴⁾。まず、「データをつくる」「データをためる」については、RDE と MatNavi がある。RDE は日々のデータを「機械学習しやすい」形に自動的に整えて保存し、MatNavi は分野エキスパートの人的作業により文献から抽出された物性データを収集・整理する。一方、「データをつかう」に関しては、pinax

(3) Dominik Kreuzberger et al., “Machine Learning Operations (MLOps): Overview, Definition, and Architecture,” *IEEE Access*, Vol.11, 2023, pp.31866-31879. <<https://ieeexplore.ieee.org/document/10081336>>

(4) 「データ戦略の中核を担い、日本のマテリアル革新力を強くする！」『NIMS now』 Vol.23 No.3, 2023, pp.12-13. <https://www.nims.go.jp/publicity/nimsnow/vol23/g66ep20000007ub0-att/NN_JP_2023_no3_WEB.pdf>

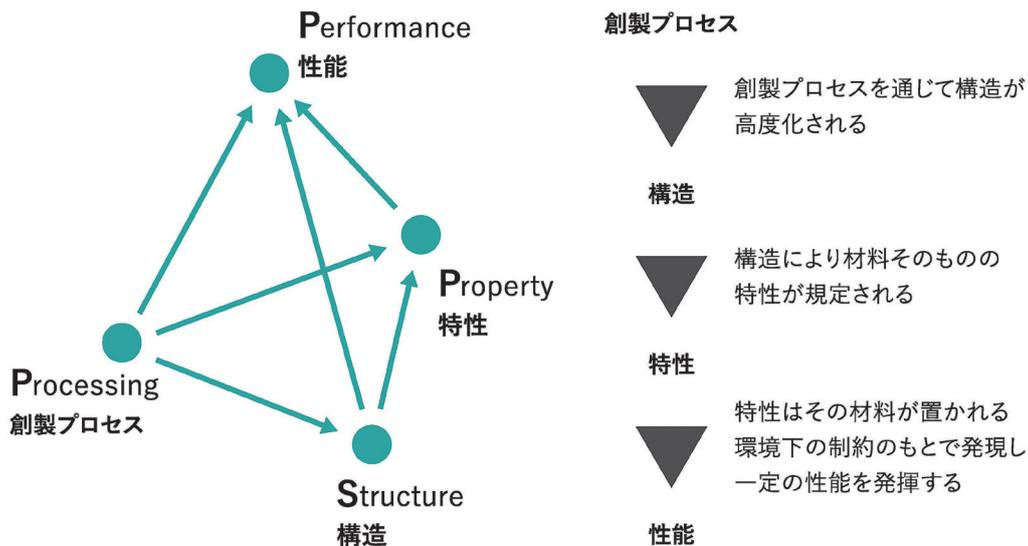
(構築中)とMIntがある。前者はためたデータを「材料工学的な観点で」探索し機械学習に適したデータを構築する場を提供し、後者は構築した予測モデルを利用して実際に予測を行う(運用する)ための場である。これらは個別のシステムとして設計・実装され、さらに相互の連携を通じて、データの有効活用が可能となる。

加えて材料データプラットフォームでは、マテリアル分野におけるデータの記述上の工夫を様々に試行している。その一つに、PSPP (Processing-Structure-Property-Performance) と呼ばれる、マテリアル分野の研究者の間で共有されている素朴な因果関係の活用があり、現在、MIntシステムに実装されている。PSPPとは、具体的にいうと、

- ① 創製プロセス (Processing) を通じて構造 (Structure) が高度化される、
- ② 構造 (Structure) により材料そのものの特性 (Property) が規定される、
- ③ 特性 (Property) はその材料が置かれる環境の制約下で発現し一定の性能 (Performance) を発揮する、

という一連の関係性である(図3)。この関係性に従ってデータが記述されれば、マテリアル研究者がデータを容易に把握することが期待される。

図3 PSPP (Processing-Structure-Property-Performance)



(出典) 筆者作成。

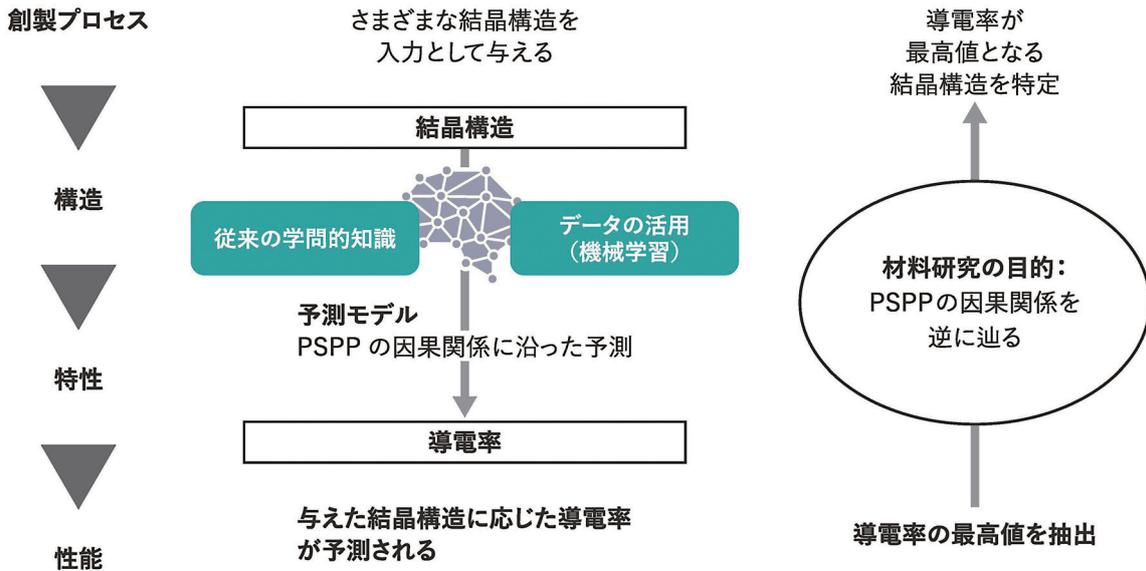
例えば、構造 (Structure) の情報である結晶構造を入力として、特性 (Property) の指標である導電率を予測するモデルを考えてみる。まず重要なのは、このモデルに対する入出力の関係性が上記のPSPPを背景に想起されていることである。これにより構築されるモデルは、作った本人以外のマテリアル研究者も比較的容易に把握できる。

さらに、図4に示すように、構築したモデルを使って様々な結晶構造に対して導電率を予測し、それらの結果を総合すると最も高い導電性を持つ結晶構造を選び出せる。これは「導電率の最高値を達成するための構造を特定する」方法を示しており、PSPPの因果関係を逆に辿(たど)るような解析を行うことに相当する。

簡単に書いたが、ここで示した例のようなPSPPの因果関係を逆に辿るような知見の獲得は極めて重要である。なぜなら、それは材料研究の目的そのものだからである。すなわち、目標

の性能（Performance）を達成するための材料の創製プロセス（Processing）の確立が求められるのであり、そのためには、性能を達成するための特性、その特性を発現できる構造、その構造を生み出せる創製プロセスを、因果関係を逆に辿って明らかにしたいのである。例に示したように、PSPPの因果関係に沿ったモデルによる予測は順問題予測と言われ、因果関係を逆に辿る予測は逆問題予測と呼ばれる。ここで示した例は、順問題の予測モデルの構築・運用を通じて逆問題を解く方法により、材料研究の目的達成をサポートできることを示している。

図4 PSPPに沿った機械学習の材料研究への活用



(出典) 筆者作成。

「データをつくる」「データをためる」段階においては、扱う情報をPSPPでラベル付けして分類することで、研究者は予測モデル構築に必要なデータを効率的に探せるようになる。

4 I 節の総括

I 節を総括すると、まずマテリアル DX プラットフォーム構想の下、構築が進められる材料データプラットフォームを例として、マテリアル分野におけるデータプラットフォームの機能を概観した。機械学習を想定したデータを効果的に活用する（データをつかう）ための様々な機能を提供し、活用の前段階の「データをつくる」「データをためる」における機能も含むものである。さらに、マテリアル分野に特徴的なPSPPの関係性をデータ管理に適用することで、関係者がデータを把握し、利用しやすくなることを示した。これらの機能群により、マテリアル分野の研究者はデータをよりどころとして研究を進め、新たな発見やイノベーション促進を活性化できる。

II 国内外のDX活用の動向

1 マテリアル分野におけるデータ駆動型研究の勃興期

前節において、マテリアル分野におけるデータ駆動型研究の現状として、それらの研究開発

をサポートするデータプラットフォームの機能を取り上げ概観した。本節では、そのようなデータプラットフォームの重要性が認識されるようになった背景及び経緯、また、現在進んでいる取組についてまとめる。

まず背景として重要なのは、ICT技術の進化と、それを礎においた「インダストリー 4.0」⁽⁵⁾の概念である。2011年にドイツで提唱されたこの概念は、製造業をデジタル化し、製品の生産と流通の効率化を目的としていた。提唱された当初は、IoT (Internet of Things)、クラウドコンピューティングなどのICT技術を中心に、製品の「製造」や「配送」に関する情報をリアルタイムで共有し、製造プロセスを改善し、新しいビジネスモデルを創出することを重視していた。この取組の一環として、「マスカスタマイゼーション」という、大量生産の仕組みを活用しながら、個別の顧客ニーズに合った製品を提供するアイデアも浮上した⁽⁶⁾。

当然のことながらこの概念は製造業にとどまるものではなく、マテリアル科学分野においても波及し、まずはシミュレーションを中心にデジタル化とデータ活用の取組が展開された。取組を振り返るときに、政策的に特に重要なのは、2011年に米国オバマ (Barack Obama) 政権下で発表された Materials Genome Initiative (MGI) であり⁽⁷⁾、イニシアチブを推し進める際の「成功例」として取り上げられた、Materials Project と呼ばれる新しい物質の発見に役立つ計算結果データベースである⁽⁸⁾。このデータベースには、理論のみに基づいて安定的な原子配列と各種物性を計算するシミュレーション (第一原理計算) 結果を収録しており、実測されていない仮想的なものも含めて、様々な原子配列に対する各種物性を計算する手続 (ワークフロー) を自動化し、次から次に計算を自動実行し、得られた結果を後から検索できるように一定の手続に従って整理した後データベースに収録している。このデータベースを使えば、例えば、蓄電池の電極材料として適切な原子配列を探し出す、という検索が可能になり、データへの注目が進んだ。同様のデータベースが各種開発されており、AFLOW⁽⁹⁾、OQMD⁽¹⁰⁾、AiiDA⁽¹¹⁾、NOMAD⁽¹²⁾などがその一部で、これらのデータベースは新しい材料の特性を予測するための計算に役立っている。

このような計算結果データベースが一定の成果を収める一方で、MGIの目標は、あらゆる材料研究のデジタル化を想定していた。Integrated Computational Materials Engineering (ICME) はそうした認識の下で提唱された概念であり⁽¹³⁾、航空機・自動車・船舶などの製造を、材料

(5) Andreja Rojko, "Industry 4.0 Concept: Background and Overview," *International Journal of Interactive Mobile Technologies*, Vol.11 No.5, 2017, pp.77-90. <<https://doi.org/10.3991/ijim.v11i5.7072>>

(6) Yi Wang et al., "Industry 4.0: a way from mass customization to mass personalization production," *Advances in Manufacturing*, Vol.5 No.4, November 2017, pp.311-320.

(7) Office of Science and Technology Policy, "Materials Genome Initiative for Global Competitiveness," June 2011.

(8) Anubhav Jain et al., "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, Vol.1 No.1, July 2013. <<https://doi.org/10.1063/1.4812323>>

(9) Stefano Curtarolo et al., "AFLOW: an automatic framework for high-throughput materials discovery," *Computational Materials Science*, Vol.58, 2012, pp.218-226.

(10) Scott Kirklin et al., "The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies," *npj Computational Materials*, Vol.1, 2015. <<https://doi.org/10.1038/npjcompumats.2015.10>>

(11) Sebastiaan P. Huber et al., "AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance," *Scientific Data*, Vol.7, 2020. <<https://doi.org/10.1038/s41597-020-00638-4>>

(12) Markus Scheidgen et al., "NOMAD: A distributed web-based platform for managing materials science research data," *Journal of Open Source Software*, Vol.8 No.90, 2023. <<https://doi.org/10.21105/joss.05388>>

(13) National Research Council, *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security*, Washington, D.C.: National Academies Press, 2008. <<https://doi.org/10.17226/12199>>

の創製プロセスから考えて最適化するものである。具体的には、前述の計算結果データベースと同様の考え方で、シミュレーションを利用した網羅的な探索により、目的となる製品を実現するための材料創製プロセスを見つけ出すことを目指す。

ICMEの実証例として有名なのは、2014年に米国国立標準技術研究所（National Institute of Standards and Technology: NIST）からのファンディングによりノースウェスタン大学を中心として開始されたCHiMaDプロジェクト⁽¹⁴⁾である。ここではPSPPの関係をベースとしたシステムティックな材料設計と最適化の実現を目標にプロジェクトが推進されている。このプロジェクトの下で、参加組織の一つであるQuesTek Innovationsというベンチャー⁽¹⁵⁾が、2023年から材料設計アプリケーションをクラウドサービスとして提供している⁽¹⁶⁾。あわせて欧州ではICMEのコンセプトの下で欧州が強みをもつ材料シミュレーションソフトウェアを材料の創製プロセスと関連付ける取組が進んだ⁽¹⁷⁾。その成果は書籍としてまとめられており、多くのソフトウェアが連携して材料設計を効率的に進める姿が示されている⁽¹⁸⁾。

さらに、日本でも内閣府の戦略的イノベーション創造プログラム（SIP）事業を通じてMIntという計算システムが開発され、ICMEのコンセプトを背景として構造材料分野における計算機援用による材料設計を可能にしている⁽¹⁹⁾。

これらの勃興期における取組は、マテリアル分野でのデジタル化とデータ駆動型研究の発展に寄与しており、新しい材料の開発や特性の予測に大きな影響を与えた。特にシミュレーション技術を応用し、大量の計算を網羅的に行いその結果を分析することで、必要とする特性を満たせる構造や製造プロセスを見つけ出す、という方法論が世界的に認知されるようになった。その方法を具体的に実行するためには、網羅的な計算と結果を間違いなく管理できるインフラストラクチャが重要となるのは言うまでもない。その後、現在に至るまでに、AI技術の爆発的な発展がもたらされ、データの重要性はますます高まった。当然のことながら、それに呼応するようにインフラストラクチャへの注目度も上がっている。その現状についてはこの後に述べるが、その前に、データが注目を集めることになった別の文脈であるオープンサイエンスについて次項で概観する。

2 オープンサイエンスとデータ

マテリアル分野における研究開発のデジタル化を考える上で、同じ時期に起きた科学一般におけるデータの重要性の高まりについても触れておく。具体的には、オープンサイエンスと呼ばれるコンセプトであり、科学研究の成果へ誰もがアクセスできることを目指した、オープンアクセス、オープンデータ、オープンコラボレーションの原則に基づく各種取組であり、世界

(14) “NIST announces center to advance materials by design,” *MRS Bulletin*, Vol.39 No.1, January 2014, p.10. <<https://doi.org/10.1557/mrs.2013.324>>

(15) QuesTek Innovations LLC Website <<https://questek.com/>>

(16) “ICMD Materials Design Platform.” QuesTek Innovations LLC Website <<https://questek.com/products-client-services/icmd-materials-design-platform/>>

(17) Georg J. Schmitz and Ulrich Prael, “ICMEg: the Integrated Computational Materials Engineering expert group: a new European coordination action,” *Integrating Materials and Manufacturing Innovation*, Vol.3, 2014, pp.20-24. <<https://doi.org/10.1186/2193-9772-3-2>>

(18) Georg J. Schmitz and Ulrich Prael, *Handbook of Software Solutions for ICME*, Weinheim: Wiley-VCH Verlag GmbH & Co., 2017.

(19) Satoshi Minamoto et al., “Development of the Materials Integration System for Materials Design and Manufacturing,” *Materials Transactions*, Vol.61 Issue 11, 2020, pp.2067-2071. <<https://doi.org/10.2320/matertrans.MT-MA2020002>>

的に活性化している。ここには、科学の在るべき姿であるグローバルな公共善としての科学⁽²⁰⁾が投影されている。

オープンサイエンスを推進するこれらの原則のうち、オープンデータの中心には研究データの安定的な保存と管理がある。これを支えるデータ管理の基本原則として知られるのが、FAIR原則⁽²¹⁾である。FAIRは、Findable（見つけやすい）、Accessible（アクセス可能な）、Interoperable（互換性がある）、Reusable（再利用可能な）の頭文字を取ったもので、データの有効な利用を実現するために管理するデータが満たすべき条件を述べている。

このFAIR原則を満たせるようにするためには、データインフラが重要である。各国では異なるアプローチが取られており、米国では政府主導の共通インフラの計画は進行していないが、欧州ではEuropean Open Science Cloud（EOSC: 欧州のサイエンスクラウド）⁽²²⁾、ドイツではNational Research Data Infrastructure（NFDI: 研究データ管理コンソーシアムのネットワーク）⁽²³⁾などが政府主導で構築が進められているデータインフラの例である。これらのインフラは科学研究全般をサポートし、分野ごとのデータ駆動型研究を推進する役割を果たしている。例えば、ドイツNFDIの下では、MatWerk⁽²⁴⁾、FAIRmat⁽²⁵⁾など材料関連のコンソーシアムが複数推進されている。日本でもNII Research Data Cloudなどがデータ管理基盤として整備されており⁽²⁶⁾、科学の進歩とオープンサイエンスの実現に向けた取組が行われている。データの整備と共有は、科学研究の効率性を向上させ、新たな発見を促進する重要な要素となっている。

オープンデータの現状として、AI技術の爆発的な発展によるデータの価値の高まりを背景に、データを適切に管理し、保護するための取組であるデータガバナンスの重要性が高まっている⁽²⁷⁾。データガバナンスを具体的に考えたときに、攻めの文脈と守りの文脈の二つの側面から捉えられる⁽²⁸⁾。攻めの文脈では、データ駆動型研究や学際領域研究、イノベーションの創出などが重要である。一方、守りの文脈では、公的資金で支えられた研究成果を社会に還元し、研究の透明性や説明責任の下で、研究の公正性と再現性の確保が重要である。

オープンサイエンスの背景にある「グローバルな公共善としての科学」の立場からは、特に守りの文脈が強調され、研究の公正性やアクセス可能性が重視される。一方で、データ活用はイノベーションの創出と結び付いており、データの適切な取扱いが、ウェルビーイング（幸福

(20) Geoffrey S. Boulton, *Science as a Global Public Good: a Position Paper of the International Science Council*, Paris: International Science Council, 2021. <<https://council.science/wp-content/uploads/2020/06/ScienceAsAPublicGood-FINAL.pdf>>

(21) Mark D. Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, Vol.3, 2016. <<https://doi.org/10.1038/sdata.2016.18>>

(22) European Commission, *Realising the European open science cloud: First report and recommendations of the Commission high level expert group on the European open science cloud*, Luxembourg: Publications Office of the European Union, 2016. <<https://data.europa.eu/doi/10.2777/940154>>

(23) German National Research Data Infrastructure (NFDI) <<https://www.nfdi.de/?lang=en>>

(24) NATIONALE FORSCHUNGSDATENINFRASTRUKTUR FÜR MATERIALWISSENSCHAFT & WERKSTOFFTECHNIK (NFDI-MatWerk) <<https://nfdi-matwerk.de/>>

(25) FAIRmat, funded by the Deutsche Forschungsgemeinschaft (DFG: German Research Foundation)-project 460197019 <<https://www.fairmat-nfdi.eu/fairmat>>

(26) Masaru Kitsuregawa et al., “Activities of National Institute of Informatics in Japan,” *Communications of the ACM*, Vol.66 No.7, July 2023, pp.58-63. <<https://dl.acm.org/doi/10.1145/3589736>>

(27) 『研究データ共有（オープンデータ）の動向』科学技術振興機構研究開発戦略センター, 2023. <<https://www.jst.go.jp/crds/pdf/2022/FU/TP20230203.pdf>>

(28) 船守美穂「大学はオープンサイエンスにどのように向き合うか—社会に開かれた大学に向けて—」『科学』92巻8号, 2022.8, pp.703-707.

と福祉)の追求につながっている。攻めの文脈と守りの文脈は二項対立するものではなく、その両立こそが、科学研究の発展に貢献し社会全体の利益に結び付くことを理解しておく必要がある。

3 マテリアル分野におけるデータインフラストラクチャに関する取組

前節までの背景整理の下で、マテリアル分野におけるデータインフラストラクチャに関する取組の現況について概観する。米国、欧州、中国・韓国の状況についてまとめた後、日本の立ち位置を分析する。

(1) 米国

米国では、まず、ICMEのコンセプトに基づいたデータインフラストラクチャに注目が集まっており、鉱物金属材料学会(The Minerals, Metals & Materials Society: TMS)の材料データインフラストラクチャに関するレポート出版⁽²⁹⁾や研究データ同盟(Research Data Alliance: RDA)のInterest Group活動⁽³⁰⁾を通じて取組が進められている。

こうした活動が継続される中、2021年のバイデン(Joe Biden)政権発足に伴い、MGIが再びイニシアチブとして復活し、戦略目標が見直された⁽³¹⁾。データインフラストラクチャの整備、コミュニティ構築及びAIを活用した研究推進に重点が置かれている。

まず、データインフラストラクチャの整備に関して、特定のプラットフォームを重要視せず、ユースケースの明確化が強調されている。実際に構築されているプラットフォームも多数あり、Materials Data Facility⁽³²⁾やMaterials Commons⁽³³⁾、NanoMine⁽³⁴⁾などがその一部である。また、民間による材料設計プラットフォームの提供も進んでおり、早くからAI援用による材料設計のコンセプトを提唱したCitration⁽³⁵⁾を始めとして、QuesTek ICMD⁽³⁶⁾、Ansys Granta MI⁽³⁷⁾などが存在し、これらは材料研究を支援するためのツールとして活用されている。さらに、II1で述べた第一原理計算の結果を集約・蓄積するデータベースも稼働しており、継続的に利用されている。

これらのデータプラットフォームを活用するコミュニティ構築も重点課題の一つである。こ

(29) The Minerals, Metals & Materials Society (TMS), *Building a Materials Data Infrastructure: Opening New Pathways to Discovery and Innovation in Science and Engineering*, 2017.

(30) “RDA/CODATA Materials Data, Infrastructure & Interoperability IG.” Research Data Alliance Website <<https://www.rd-alliance.org/groups/rdacodata-materials-data-infrastructure-interoperability-ig.html>>

(31) *Materials Genome Initiative Strategic Plan, A Report by the Subcommittee on the Materials Genome Initiative Committee on Technology of the National Science and Technology Council*, Washington, D.C.: Executive Office of the President of the United States, November 2021. <<https://www.mgi.gov/sites/default/files/documents/MGI-2021-Strategic-Plan.pdf>>

(32) Ben Blaiszik et al., “The Materials Data Facility: Data Services to Advance Materials Science Research,” *JOM*, Vol.68 No.8, July 2016, pp.2045-2052.

(33) Brian Puchala et al., “The Materials Commons: A Collaboration Platform and Information Repository for the Global Materials Community,” *JOM*, Vol.68 No.8, July 2016, pp.2035-2044.

(34) He Zhao et al., “Perspective: NanoMine: A material genome approach for polymer nanocomposites analysis and design,” *APL Materials*, Vol.4 No.5, May 2016. <<https://doi.org/10.1063/1.4943679>>

(35) Jordan O’Mara et al., “Materials Data Infrastructure: A Case Study of the Citration Platform to Examine Data Import, Storage, and Access,” *JOM*, Vol.68 No.8, June 2016, pp.2031-2034.

(36) “ICMD Materials Design Platform,” *op.cit*(16)

(37) “Ansys Granta MI Enterprise.” ANSYS Inc. Website <<https://www.ansys.com/ja-jp/products/materials/granta-mi>>

れに関連して、Materials Research Data Alliance (MaRDA) と呼ばれる研究者同盟の立ち上げにファンディングがなされ、2回の年次総会を成功裏に開催し、現在米国及び海外から300名を超える会員を擁するコミュニティに成長している⁽³⁸⁾。現在、MaRDAはジョンズホプキンス大学に事務局を置き、マテリアル分野におけるデータ活用に関する草の根活動を行う母体となっており、マテリアル分野における研究開発のデジタル化に関しての課題共有と活動の集約が進められている⁽³⁹⁾。

(2) 欧州

欧州におけるマテリアル分野のデータインフラストラクチャとして、第一原理計算データベースの整備が米国と同様に継続されている。重要なものは、NOMADやAiiDA/Materials Cloudであり、ワークフロー定義による第一原理計算の自動実行結果の収蔵先として活用されている。蓄積されたデータは系統的に整理され、検索可能にする基盤として機能している。

この動きと並行して、特にドイツではMaterialDigital⁽⁴⁰⁾というプロジェクトが進行している。このプロジェクトは、マテリアル分野の研究開発をデジタル化し、データの管理の強化を目的として、様々な研究機関の連携により推進されている。サービスプロバイダとして統一的なデータインフラストラクチャの構築・提供はせず、個別のデータインフラストラクチャに関する取組のネットワーク・インターフェースとしての役割を果たすことを目指している。特に重要なのは、データの記述方式の統一である。つまり、各機関においてデータが生成されると、統一の記述方式に基づいて保存され、分散する記憶装置に保管しつつ、様々なアプリケーションが定義されたワークフローに基づいて連携できるよう、標準化されたデータアクセスと処理の実現を目指している。また、試作の位置付けでインフラストラクチャ構築も進めている⁽⁴¹⁾。また、これら複数のデータインフラへの相互アクセスの提供もサポート範囲であり、それぞれの機関が有する固有のデータに対して機関外のパートナーがアクセスできる仕組み作りが進められている。ただし、いずれの取組も分散しており、統合はこれからという状況である。

なお、MaterialDigitalへの参画機関は、カールスルーエ工科大学 (KIT)⁽⁴²⁾、ドイツ連邦材料試験所 (BAM)⁽⁴³⁾、フラウンホーファー材料メカニズム研究所 (Fh-IWM)⁽⁴⁴⁾、マックスプラ

(38) “NSF Award # 1933640, Collaborative: Summit on Big Data and Cyberinfrastructure in Materials Research.” The U.S. National Science Foundation Website <https://www.nsf.gov/awardsearch/showAward?AWD_ID=1933640>

(39) MaRDA Website <<https://www.marda-alliance.org/>>

(40) MaterialDigital Website <<https://www.materialdigital.de/>>

(41) Jan Janssen et al., “pyiron: An integrated development environment for computational materials science,” *Computational Materials Science*, Vol.163, June 2019, pp.24-36; Zoe Landgraf et al., “SIMstack: A Generative Shape and Instance Model for Unordered Object Stacks,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp.12992-13002. <<https://doi.org/10.48550/arXiv.2103.16442>>

(42) ドイツ最古の工科大学であり、2009年に大型研究施設の運用をミッションの一つとするヘルムホルツ協会の原子力研究機関であったカールスルーエ研究センター (Forschungszentrum Karlsruhe) を吸収合併し、大学と公的研究機関の二面性を持つ世界でも類を見ないユニークな教育研究機関として運営されている。

(43) ドイツ連邦経済・気候保護省が所管する、研究開発、試験、分析、承認、アドバイス、情報に関して、技術と化学の安全性を向上させることを目的とした公共の技術的安全と計量業務を担当する。また、ドイツの技術標準であるDINにおける材料試験を担当する。

(44) 産業技術を研究開発するフラウンホーファー協会が擁する研究所の一つ。製造工程における材料の特性やコンポーネントの挙動を様々なスケールで評価し、材料とコンポーネントの開発、製造、使用を総合的に検討し、新しい機能や製造工程の実現に貢献する。

ンク鉄鋼研究所 (MPIE)⁽⁴⁵⁾ などであり、それぞれ、ドイツにおける基礎研究振興、産業応用、標準化など異なるミッションを有しており、プロジェクト成果に様々な応用が期待されていることを反映している。

(3) 中国・韓国

中国では、研究開発プロジェクトの一環としてデータプラットフォームの構築が進行中である。プロジェクトとその実施拠点は、上海大学が運営する Materials Genome Institute (MGI)⁽⁴⁶⁾、及び北京科技大学 (USTB) の Materials Genome Engineering (MGE) センター⁽⁴⁷⁾ である。これらの拠点においてデータインフラストラクチャの整備が進んでいる⁽⁴⁸⁾。

韓国では、マテリアル分野におけるデータ駆動型研究のキャッチアップが盛んに行われている。K-MDS⁽⁴⁹⁾ と呼ばれるプロジェクトでは、韓国科学技術情報研究院 (KISTI) がデータプラットフォームを運用しており、「国家戦略技術を支える未来材料の確保に向けた戦略 (Strategy for Securing Future Materials to Support National Strategic Technologies)」の下で、データ駆動型材料研究やロボット活用などと組み合わせた材料開発のパイロットプロジェクトを推進している。さらに、韓国標準科学研究院 (KRIS) は、材料研究データベースセンター (National Center for Materials Research Data: NCMRD) を運営し、データインフラとネットワーキングに力を入れており、国際シンポジウムを主催して欧米とのネットワーキングも積極的に進めている⁽⁵⁰⁾。

4 現状分析

これまで見てきたように、科学研究一般に対して、データ管理とそのためのインフラの大切さは、世界中で認識されている。データ管理を強化するための目標は、どの分野でも同じであり、ワークフローの自動化とデータの統合である。これらの目標は、マテリアル分野においても同様である。

一方で、データ管理の目的は、幾つかの側面があることを見てきた。一つ目は、イノベーションの促進である。これは、将来の利益を追求し、研究開発のデジタル化による効率向上を意味する。また、オープンサイエンスの一環として、コラボレーションによる新しいアイデアの創出も含まれる。二つ目は、科学のグローバルな公共善としての達成である。これは、オープンサイエンスの守りの側面で、長期的で平等なアクセスの確保を目指すものである。この目的に向かうとき、当初の段階では、データを蓄積して検索できるようにすることまでを達成目標と

(45) 基礎科学振興を目的とするマックスプランク協会が擁する研究所の一つ。民間出資の研究所である鉄鋼協会 (VDEh) との共同出資により運営され、産業応用もその活動範囲としたユニークな運営が行われている。特に化学、物理学、工学、材料科学を組み合わせた高度に学際的なアプローチを追求し、ハイテク構造及び機能コンポーネント用の新しい高性能材料の開発に貢献する。

(46) Materials Genome Institute of Shanghai University Website <<https://mgi-en.shu.edu.cn/index.htm>>

(47) Beijing Advanced Innovation Center for Materials Genome Engineering Website <<http://en.bjmgc.ustb.edu.cn/>>

(48) Zhuo Wang et al., "Integrated materials design and informatics platform within the materials genome framework," *Chinese Science Bulletin*, Vol.59, March 2014, pp.1755-1764; Shilong Liu et al., "An infrastructure with user-centered presentation data model for integrated management of materials data and services," *npj Computational Materials*, Vol.7, June 2021. <<https://doi.org/10.1038/s41524-021-00557-x>>; Shuling Xiong and Luohan Wang, "Research Progress and Development Trends of Materials Genome Technology," *Advances in Materials Science and Engineering*, Vol.2020, Article ID 5903457. <<https://doi.org/10.1155/2020/5903457>>

(49) Korea Materials Data Station (K-MDS) Website <<https://kmds.re.kr/en/web/guest>>

(50) Second International Symposium on Materials R&D Data Website <<https://www.ncmrd-symposium.net/>>

する場合が多く、その後にある「データをつかう」段階に関しては、ユーザが工夫しなければならないことが多い。

日本においては、マテリアル分野におけるデータプラットフォームは、特に、AIを活用して新しい素材の創出を見据えたマテリアル強化戦略に沿って構築が進められている。この戦略において「データをつかう」段階におけるAI活用が求められ、そのために必要なデータ管理が鍵となる。そのためにプラットフォームが備えるべき機能については、I節で概観したとおりである。

マテリアル分野におけるデータプラットフォームはプロジェクト単位で分散しており、「データをつかう」ことまでを意識して分散するデータを統合する取組に関しては具体的な成果にまで至った例は極めて少ない。必要なワークフローやデータの記述形式が見定められていない現段階では時期尚早と考えている国もある。現段階において、統合的なプラットフォームを稼働させるのは挑戦的なプロジェクトであり、成功すれば世界で先駆的な存在となることができる。そうした中、政府主導によるイニシアチブから統合システムを追求し、その構築を中央集権的なアプローチで進める日本の取組は、全国の共用機器からのデータ集約を始めとしてほかの国々と比べても先行している部分も多く、国際的にも注目されている。

Ⅲ 未来への選択肢—マテリアル科学研究のエンジニアリング—

前節で見た取組を、今後、更に発展・強化していくにはどうすべきか。その考慮すべきポイントとして、本節では、データ記述法、データフロー制御、人材育成、の3点を取り上げ考察する。

1 エンジニアリング対象の記述

マテリアル分野におけるデータプラットフォームを構築して行おうとしているのは、材料研究そのものをデータとプロセスとしてコンピュータ上で表現し、自動化と機械学習のサポートを得て効率的な研究開発とイノベーション創出を図ることと捉えられる。現在、このアプローチは、現実として、AIの活用を呼び込み効率的な研究開発の実施と具体的な成果創出につながっているが、まだ中途である。

未来に向けての課題は、これまで以上に材料開発の段階から「適用する・利用する」ことを考慮した開発である。インダストリー4.0で注目を浴びたマスカスタマイゼーションの実現のみならず、サーキュラーエコノミー（資源循環型経済）の考え方で知られるように、持続可能性の観点も重要性を増している。単に材料を創製し大量生産するまでを材料研究の役目とするのではない多角的な検討が、材料開発の段階においても必要になりつつある。

こうした課題解決に向けてデータ駆動型研究開発のアプローチを重ねると、必要なのは材料創製の帰結としての「適用する・利用する」ことのコンピュータ上での表現だが、現段階において、その記述方法は確立されているとは言い難い。

一つのアイデアは、前節まで述べてきたPSPPの関係性に立ち戻ることである。ここまで見てきたように、この関係性は、材料研究者にとって有用な概念でデータ駆動型材料研究と整合性が高い上に、「適用する・利用する」の可否を判断する重要な指標としてPSPPの最後のPであるPerformance（性能）を観点として有している。一方で、性能の定義が曖昧なままデー

タ駆動型材料研究が進んできたのも事実であり、「適用する・利用する」の記述方法が確立されていない現状をそのまま表している。

性能の記述方法を考えるためには、「適用する・利用する」を深掘りする必要がある。この議論は続く章に譲るが、ここでの結論として、材料の効率的な開発と社会適用、そのためのカスタマイズを実現するためには、データ時代におけるエンジニアリングの対象である「マテリアル科学」のコンピュータ上での体系的な記述が不可欠であり、PSPPの関係性の記述をいかに発展させるかが未来への選択肢であることを強調したい。

2 目的に結びついたデータフローの制御

データプラットフォームの構築により明らかになっている課題として、研究開発において扱われる知的資産があまねくデジタルデータとして表現されるため、これまで以上にその取扱い方に対する深い見識が求められることが挙げられる。具体的には、「データをつくる」「データをためる」「データをつかう」というそれぞれのフェーズは互いに連携して、場合によってはサプライチェーンと結び付きながらデータの流れを生み出すが、このとき、その流れがどこから来ているのか、利用可能なのか、出力できる範囲はどこからどこまでなのか、など、データの流れを認識し、目的に合わせたデータの選択や保護が極めて重要な課題となる。

課題の解決に向けては、マテリアル科学者の考え方そのものの変革が求められるかもしれない。まず、データの選択においてであるが、通常、マテリアル研究者は材料の創造に焦点を当て、それを適用・利用する段階については、製品製造で考慮されるものとして研究を進める場合が多い。しかし、II節で述べたように、データ駆動型研究は、インダストリー 4.0で掲げられたマスカスタマイゼーションや、持続可能性への配慮など新しい価値観を実現する契機となり得るものである。このような新しい価値観へ対応していくには、新しいものを作り廃棄するリニアモデルから脱却し、マテリアル科学者がデータ駆動型研究を推進する中で「適用する・利用する」の段階までを考慮できることが重要である。当然のことながら、データの選択においてもこうした視点が必要である。

次に、データ管理において「攻めの文脈」と「守りの文脈」の理解も欠かせない。研究インテグリティ、個人情報保護、機密情報保護、ライセンス管理など、様々な観点からのデータ管理が必要で、それを踏まえた上でデータ選択が行われなければならない。

データの流れに関するこれらの観点を考慮して、データプラットフォーム機能が提供されなければならない。その時にポイントとなるのは、利用条件とユーザ管理である。不特定多数のユーザを対象にする場合、利用者像が特定できないために、様々なリスクに備えてシステムの構築と運用が高コストになる可能性がある。そのため、利用条件を課し、その条件を満たすユーザのみを対象とすることで、効率的なシステム構築とデータ利用を実現できる。

以上の課題は全て、各種データへのアクセスを許可するルールの確立と強固に結び付いている。特に、データ駆動型研究においては、取得した目的とは異なるデータ利用や、失敗データさえも利用データとすることも想定され、他人のデータを利用するための仕組みづくりやルールが重要になる。これらの仕組みやルールがないままにシステム化を推し進めたとしても、ルールにそぐわないシステムとなってしまう実際の運用で問題が生じる可能性が高い。言うまでもなく、「グローバルな公共善としての科学」と「イノベーション創出のための科学」を二項対立で整理せず、適切に設定された目的の下で両文脈を満たせる共通点を見いだしルールを

確立していくことが、最重要課題として今後問われるだろう。

3 データ時代の研究人材育成

最後に、データ時代の研究人材育成について考えたい。

ICT技術の発展により、データの重要性が認識され実際に多くのデータを集められるようになりつつある。その一方で、集約したデータの効果的な活用はいまだに課題である。前節まで述べたように、データ記述を理解し、データフローに関する様々な制約を見極めた上で「データをつかう」ことが求められるが、現段階においてシステム機能のサポートは可能であっても、フルオートでの機械的な実施は困難であり、人間が間に入ることから始まる。そうした人材にはデータの「つかいこなし」に関するスキルと知識が必要となる。

今必要なのは、それらのスキルを持った人材の育成である。ツール理解や利用を前提とした教育やトレーニングが行われなければならないが、分野自体が発展途上にあるため確固たる育成プログラムが十分に整備されているとは言い難いのが現状である。

さらには、そうした人材が備えるべきスキルから定型化できるものを切り出してツール群として育てていくことも、教育・トレーニングと併せてコミュニティ全体のスキルセットの高度化につながっていく。言い換えれば、データの使いこなしを熟知した上で、ツールを設計・開発できる技能を備える人材が今後ますます求められるだろう⁽⁵¹⁾。

データ時代には新しいスキルや知識が必要であり、これを持った研究者やエンジニアの育成が大切である。大量生産・大量消費の時代にはなかったマテリアル科学者の考え方の根本的な変革を伴いながら、データを使いこなして新しい知識を生み出しツールを增強する能力を持つ人材をどのように育てるのか、様々な立場からの知恵が求められている。

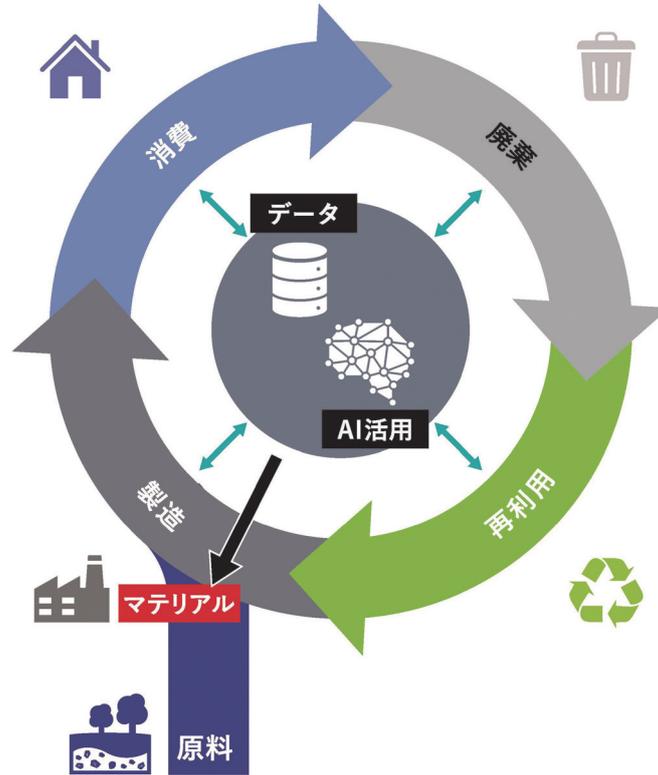
おわりに

本章では、マテリアル分野におけるDXの現状、国際動向を振り返り、未来への選択肢として重要な課題について概観した。社会の基盤であり、日本が世界と比較しても競争力を有するマテリアルは、新しい産業や付加価値が生まれる分野として重要であり、データの活用により競争力を更に高められることを示してきた。

原料—製造—消費のサプライチェーンが複雑化しているとはいえ、その最上流にあるマテリアルにとっての顧客からの要望は、目の前にいる製造段階の「お客」からの要望で済んだのかもしれない。一方で、持続可能性が重要視されるこの時代においては、図5に示すように、サプライチェーンのみを考えれば良いのではなく、その先にある廃棄—再利用等の取組と接続したサーキュラーエコノミーの考え方が広く受け入れられている。つまり、遠い先にある話が、自分自身に直接流れ込んでくるような構造を前提とした取組がマテリアルにも求められることになる。

(51) Chris Woolston, “Why science needs more research software engineers.” 31 May 2022. Nature Website <<https://doi.org/10.1038/d41586-022-01516-2>>

図5 サークュラーエコノミー時代のマテリアルの位置付け



(出典) 筆者作成。

そうした社会においてマテリアル科学に求められるのは、目の前だけを見るのではない、サプライチェーンの先までのみならず、更なる先にある永く続くサイクルを見通した中でマテリアル科学を推進することだろうと考える。データとAI活用は、そうした先を見通すための重要なツールとして位置付けることができる。

さらに、データはこうした円環型の取組の中においては常に蓄積され続ける。新たな蓄積が新たなAI活用によって新たな競争力を生み出すことができるのであり、その効果は時の流れの中で複利的に働くことが期待される。

ここから踏み込んで言えば、サプライチェーンの最上流にあるがゆえに、マテリアル科学はデータと結び付くことで、むしろ積極的に円環型の経済活動の中で競争力の源泉となるようなテクノロジーを生み出し、社会の課題解決、マーケットの課題解決を一層的に図ることができることを示している。そうした取組をより力強く進めるために、様々なセクターが連帯して取組を加速する仕組みや場づくりが、今後強く求められるだろう。その連帯の礎となるのが、データであり、データプラットフォームなのである。

執筆：国立研究開発法人物質・材料研究機構技術開発・共用部門 ユニットリーダー
かどひら たくや
門平 卓也