

戦略プロポーザル

材料創製技術を革新する プロセス科学基盤

～プロセス・インフォマティクス～

STRATEGIC PROPOSAL

Process Science Platform for Innovation in Materials Creation Technology

- Process Informatics -

研究開発戦略センター（CRDS）は、国の科学技術イノベーション政策に関する調査、分析、提案を中立的な立場に立って行う公的シンクタンクの一つで、文部科学省を主務省とする国立研究開発法人科学技術振興機構（JST）に属しています。

CRDSは、科学技術分野全体像の把握（俯瞰）、社会的期待の分析、国内外の動向調査や国際比較を踏まえて、さまざまな分野の専門家や政策立案者との対話を通じて、「戦略プロポーザル」を作成します。「戦略プロポーザル」は、今後国として重点的に取り組むべき研究開発の戦略や、科学技術イノベーション政策上の重要課題についての提案をまとめたものとして、政策立案者や関連研究者へ配布し、広く公表します。

公的な科学技術研究は、個々の研究領域の振興だけでなく、それらの統合によって社会的な期待に応えることが重要です。「戦略プロポーザル」が国の政策立案に活用され、科学技術イノベーションの実現や社会的な課題の解決に寄与することを期待しています。

さらに詳細は、下記ウェブサイトをご覧ください

<https://www.jst.go.jp/crds/>

エグゼクティブサマリー

「材料創製技術を革新するプロセス科学基盤 ～プロセス・インフォマティクス～」とは、目的材料の合成プロセスを効率的かつ統合的に探索する技術基盤を確立するための研究開発戦略である。ここでは、プロセス・インフォマティクスを「従来からの実験科学、理論科学、計算科学と、近年の進展が著しいデータ科学を、統合的・融合的に活用することにより、目的材料の合成プロセスを効率的かつ統合的に探索する方法」と定義する。

プロセス・インフォマティクスを、所望の機能を有する物質・材料を効率的に探索するマテリアルズ・インフォマティクス、プロセスの内部状態や生成物をリアルタイムに観測する計測インフォマティクスと組み合わせることで、物質・材料創製技術を次ステージに進化させることができる。また、データ科学活用による熟練者の経験知（勘・コツ）の取り込みや、物理化学的解析およびデータ科学的解析との組み合わせによって、各論的・局所的な議論に落ち込みがちであった合成プロセスに対する考え方を刷新し、汎用性の高いプロセス科学基盤へと拡充することをめざす。

現在、ありとあらゆる場面で“材料”への期待が高まっている。SDGsの達成、カーボンニュートラル社会の構築、資源・物質循環の実現、わが国が目指すSociety 5.0社会の実現や、新型コロナウイルス感染症の猛威によって加速されるデジタル化の流れ、いずれにおいても、新たな材料の創出が決定的に重要となっている。

マテリアルズ・インフォマティクスが新材料開発において強力なツールとなることがさまざまな例で実証されてきているが、そこでは候補となる新材料の組成・構造が予測されているものの、その材料が「実際に作れるのか」、「どう作るのか」まで示された例は少ない。

材料合成プロセスは、材料ごとに手法のバリエーションが多く、それを制御するパラメータも複雑であるため、統一的に扱うことは難しい。このため、個別プロセスの改良・最適化が主となり、最適なプロセスを科学的に探索するアプローチはとられてこなかった。しかし、近年のさまざまな要素技術の進化（データ科学の進展、シミュレーション技術の高度化、プロセスをリアルタイムに観測するオペランド計測技術の開発、ハイスループット実験技術の確立など）や進展著しいデータ科学の活用も期待され、また2021年3月から本格運用が始まったスーパーコンピュータ「富岳」が利用可能になったこともあり、材料合成プロセスを効率的かつ統合的に探索するプロセス・インフォマティクスに取り組むための環境が整いつつある。

本提言において取り組むべき研究開発課題としては、**①各材料領域におけるプロセス・インフォマティクス手法の構築**、**②プロセス・インフォマティクス共通基盤構築**、**③プロセス科学基盤の拡充を可能にする新たな指針やコンセプトの創出**があげられる。

① 各材料領域におけるプロセス・インフォマティクス手法の構築

ここでは、有機材料、無機材料、複合構造材料などそれぞれの領域から中核になるプロセスを選び、そこから材料領域ごとにプロセス・インフォマティクス手法を展開していく。

有機材料系合成プロセスでは、マイクロフロー化学での有機合成プロセスが候補例としてあげられる。計算科学やデータ科学を活用した合成経路探索手法と、マイクロフロー化学のような理想系に近い反応空間での合成実験を組み合わせることで、目的化合物の合成経路を効率的に確立するための研究開発を行う。

無機材料系合成プロセスでは、結晶成長プロセスが候補例としてあげられる。精緻なシミュレーションが可能であるが、計算に長時間を要するため現状ではプロセスの設計には使いにくい。実測可能なデータを

使った機械学習モデルを構築することにより、プロセス状態の高速な予測を可能にし、プロセス設計に活かすための研究開発を行う。

複合構造材料の合成プロセスは、材料自体とその合成プロセスの両面で複雑な材料であり、必要なパラメータ数が非常に多い。機械学習の分野でオートエンコーダーとして知られる次元削減操作において、中間データとして物理的に意味のある値を用いるなどの新たな手法によって複雑なプロセス全体の最適化をするための研究開発を行う。

② プロセス・インフォマティクス共通基盤構築

個々のプロセスの要求にこたえるだけでなく、プロセス・インフォマティクス全体を加速する基盤的研究が必要である。そのためには、プロセスのパラメータ空間が広大であることや、多段階プロセスにおいて前後のプロセスへの影響があるような材料合成プロセスの特徴に対処できる、プロセス・インフォマティクスに適した機械学習アルゴリズムの開発が必要になる。

また、材料合成プロセスのシミュレーション・モデリングにはa) 実験データに基づく経験パラメータを用いたモデリングや、b) 第一原理計算および分子動力学によるシミュレーションがあるが、それらをつなぐ手法であるマルチフィジックス・シミュレーションも重要である。これらの手法を用いて第一原理計算と統計熱力学を融合することにより、各パラメータの寄与を解析可能なモデルを構築する研究も重要である。

さらに、実験データの収集方法、ハイスループット実験手法、ロボットの単位操作の共通化・標準化も重要な課題である。

③ プロセス科学基盤拡充

データ科学によってプロセス特性を適切に表す記述子を設計することにより、各論的な取り扱いであったプロセスを新たなカテゴリーに分類して議論することができるようになると考えられる。これにより、個々のプロセス解析では理解できなかった、プロセスを制御する重要な共通因子の把握が可能になり、従来は活用が難しかった複数のプロセス間でそれぞれの知見・データなどを活用できるようになる可能性がある。このようなアプローチにより、各論的・局所的であった合成プロセスの議論を刷新し、プロセス科学の基盤拡充につなげる。

上記で述べた研究開発課題を推進していくために、対象プロセスごとにプロセスセンターを設置し、そこに合成プロセス装置、評価・計測装置などを整備することが望ましい。プロセス・インフォマティクス共通基盤は、理論科学、計算科学、データ科学、プロセス技術、計測技術など多岐に渡る専門家の参画が必要であるが、別々の場所で活動するバーチャル拠点であっても機能させることができる。ただし、プロジェクトごとの一時的な連携でなく、強力な連携を継続する拠点として設置する必要がある。その上で、プロセス科学基盤拡充には、各プロセスセンターと、共通基盤の拠点を強く連携させることが求められる。また、それぞれの研究開発課題の成果を活かし相乗的に進展させるために、全体を束ねる仕組み（ガバニングボードなど）が必要である。人材育成としては、材料研究者とデータ科学者の知識・経験を融合させることが大切になる。産業界が参画する仕組みの整備、データ取扱ルールの設定なども重要な課題である。

Executive Summary

“Process Science Platform for Innovation in Materials Creation Technology -Process Informatics-” is a research and development strategy to establish a scientific framework for the efficient and integrated exploration of synthesis processes of target materials. Here, Process Informatics (PI) is defined as “a method for efficient and integrated exploration of synthesis processes of target materials through the integrated use of experimental, theoretical, and computational sciences, as well as data science, which has made remarkable progress in recent years”.

By combining PI with materials informatics (MI), which efficiently searches for materials with desired functions, and measurement informatics, which observes the internal states and products of the process in real time, it is possible to advance the materials creation technology to the next stage. In addition, we aim to expand the process science platform by reforming the way of thinking about synthetic processes, which have tended to be discussed locally and individually, through the incorporation with tacit knowledge (intuition and tips) of skilled people and the combination of physicochemical analysis and data science analysis.

Today, expectations for "materials" are rising in all sorts of situations: the achievement of the SDGs, the establishment of a carbon-neutral society, the realization of resources and materials recycling, the realization of the Society 5.0 society that Japan is aiming for, and the trend toward digitalization accelerated by the onslaught of the new coronavirus infection (COVID-19). The creation of new materials has become critically important in all these areas.

Various examples have demonstrated that MI can be a powerful tool in new materials creation, but although the compositions and structures of candidate new materials are predicted, not all of them are shown to be "actually possible to make" or "how to make".

Material synthesis processes are difficult to handle in a unified manner because there are many variations of methods for each material and the parameters that control them are complex. For this reason, the improvement and optimization of individual processes have been the main focus, and scientific approach has hardly been taken for the process optimization. However, with the recent evolution of various technologies (e.g., data science, simulation technology, high-throughput experimental technology, operando measurement technology for real-time observation of processes, etc.) and the use of a supercomputer “Fugaku”, which is compatible with data science, the environment for working on PI is being prepared.

The R&D issues to be addressed in this proposal include (1) development of PI methodologies in each material area, (2) construction of a common infrastructure for PI, and (3) creation of new guidelines and concepts that will enable expansion of the process science platform.

(1) Development of PI methodologies in each material area

We will select core processes from each area, such as organic materials, inorganic materials, and composite materials with modular structures, and start development of PI methodologies in each material area.

For organic materials synthesis processes, the micro-scale continuous flow reactor system is a candidate example. R&D will be conducted to efficiently establish synthetic pathways for target compounds by combining synthetic experiments in a reaction space close to an ideal system, such as micro-flow chemistry, with synthetic pathfinding methods that utilize computational science and data science.

For inorganic materials synthesis processes, the crystal growth process is a candidate example. Although it is possible to conduct precise simulations, it is difficult to use for process design because of the long calculation time required at present. R&D will be conducted to enable fast prediction of process states by building machine learning models using actual measurable data, and to utilize them in process design.

The synthesis processes of composite materials with modular structures are complex, both in terms of the materials themselves and their synthesis processes, and the number of required parameters is very large. To optimize the entire complex process, R&D will be conducted to develop new methods such as using physically meaningful values as intermediate data in the dimensionality reduction operation known as auto-encoder in the field of machine learning.

(2) Construction of a common infrastructure for PI

Fundamental research that not only meets the requirements of individual processes, but also accelerates the entire PI, is important. For this purpose, it is necessary to develop machine learning algorithms optimized for PI that can deal with the characteristics of materials synthesis processes, such as the vast parameter space of the process and the influence of the multi-step processes effected each other.

Simulation and modeling of material synthesis processes includes a) modeling using empirical parameters based on experimental data, and b) simulation using first-principles calculations and molecular dynamics. In addition to these methods, multi-physics simulation is also important as a method to connect them. Using these methods, it is also important to integrate first-principles calculations and statistical thermodynamics to construct models that can analyze the contribution of each parameter.

The standardizations of experimental data collection methods, high-throughput experimental methods, and unit operations of robots are also important.

(3) Enhancement of process science platform

By designing descriptors that appropriately represent process characteristics using data science, it will be possible to classify processes that have been handled individually into new categories and discuss them. This will enable us to understand important common factors that control processes, which could not be understood through individual process analysis, and may enable us to utilize knowledge and data of each process among different processes. Such approaches will renew the discussion of synthetic processes, which has been held individually and locally, and lead to the enhancement of the process science platform.

In order to promote the R&D issues described above, it is desirable to establish a process center for each target process, and to equip it with synthetic process equipments, evaluation and measurement equipments, etc. The common infrastructure for PI requires the participation

of a wide range of experts in theoretical science, computational science, data science, process technology, measurement technology, and so on, but it can function even if it is a virtual center operating in separate locations. However, they need to be established as a center for continuous powerful collaborations, rather than temporal collaboration for each project. In such a situation, the expansion of the process science platform requires a strong linkage between the process centers and the center for common infrastructure. In addition, a mechanism that unites the entire group (governing board, etc.) is necessary to make synergistic progress by utilizing the results of the R&D projects. As for human resource development, it will be important to integrate the knowledge and experience of materials researchers and data scientists. It is also important to establish a mechanism for the participation of industry and to set rules for data handling.

目次

1	研究開発の内容	1
2	研究開発を実施する意義	6
2.1	現状認識および問題点	6
2.2	社会・経済的効果	8
2.3	科学技術上の効果	11
3	具体的な研究開発課題	13
3.1	各材料領域における プロセス・インフォマティクス手法の構築	13
3.2	プロセス・インフォマティクスの共通基盤構築	20
3.3	プロセス科学基盤拡充	26
4	研究開発の推進方法および時間軸	28
4.1	研究開発の推進方法	28
4.2	時間軸	31
付録 1	検討の経緯	32
付録 2	国内外の状況	35
付録 3	専門用語説明	44

1 | 研究開発の内容

「材料創製技術を革新するプロセス科学基盤 ～プロセス・インフォマティクス～」とは、目的材料の合成プロセスを効率的かつ統合的に探索する技術基盤を確立するための研究開発戦略である。ここでは、プロセス・インフォマティクスを、「従来からの実験科学、理論科学、計算科学と、近年の進展が著しいデータ科学を統合的・融合的に活用することにより、目的材料の合成プロセスを効率的かつ統合的に探索する方法」と定義する。

プロセス・インフォマティクスを、所望の機能を有する物質・材料を効率的に探索するマテリアルズ・インフォマティクス、生成物やプロセス内部の状態をリアルタイムに観測する計測インフォマティクスと組み合わせることでデータ駆動型物質・材料創製を次のステージに進化させることができる。

本提案では、現有プロセスを前提として最適条件を効率的に探る方法にとどまらず、目的材料の最適な合成プロセスを探索する科学的手法の確立を目指している。また、データ科学を活用した熟練者の経験知（勘・コツ）の取り込みや、物理化学的解析およびデータ科学的解析を組み合わせることにより各論的・局所的であった合成プロセスの議論を刷新し、汎用性の高いプロセス科学基盤を拡充することも目指す。

具体的には、①各材料領域におけるプロセス・インフォマティクス手法の構築、②プロセス・インフォマティクスの共通基盤としての技術体系構築、③プロセス科学基盤の拡充を可能にする新たな指針やコンセプトの創出を行う。

本研究開発戦略を実施することで、材料合成プロセスを効率的に探索できるようになり、所望の機能を持つ新材料を他者に先んじて創出することが可能になるため、材料開発スピード競争に勝てるようになる。また、プロセス開発に要する費用の削減や既存プロセスの効率化（省エネルギー、副生成物抑制）などを達成することができ、日本の材料開発の優位性の確保・強化に貢献する。

図1-1に提案の研究開発の概略を示す。昨今、材料技術への要求はますます高度になっており、さらにトレードオフの関係にあるような機能を同時に、かつ低コスト・低環境負荷で実現することが求められている。このような材料への高度な要求に応えるために、従来、日本が得意としてきた技術蓄積を基礎にした材料開

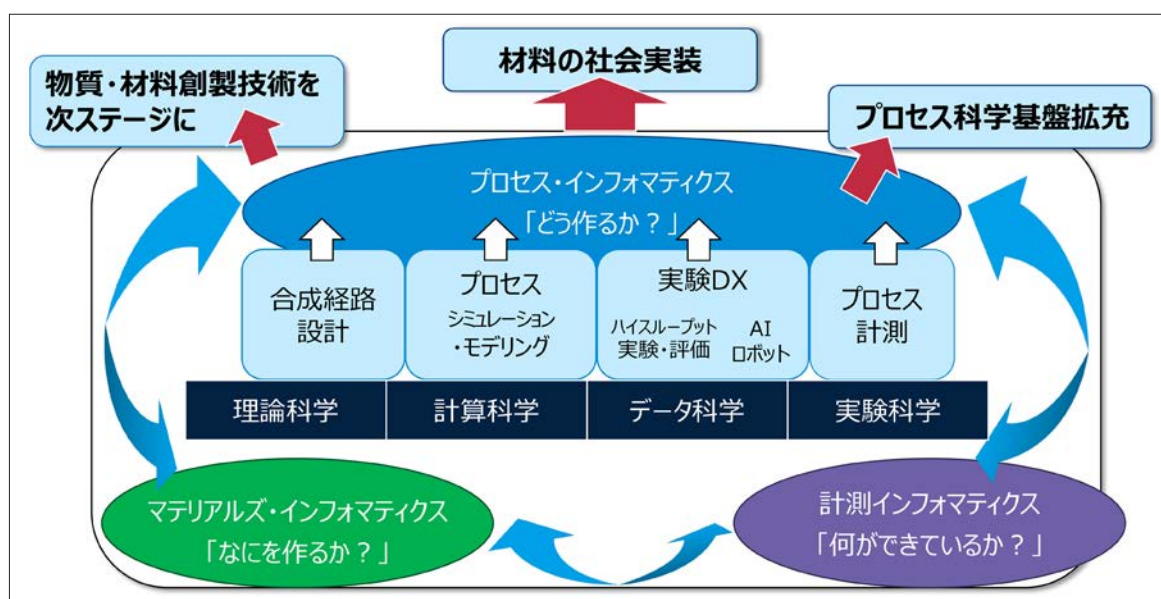


図1-1 本提案の概略

発手法に加え、高度なシミュレーション技術やデータ科学などコンピュータを用いた新たな材料設計手法であるマテリアルズ・インフォマティクスが活用され、すでにさまざまな成果が生み出されている。しかし、設計・予測された物質・材料を合成・作製する具体的な手法がわからず、実証できていない場合も散見されている。したがって、設計・予測された材料をどう合成・作製するのかという課題を解決することが必要になってきている。一方で、合成経路設計技術、プロセスのシミュレーション・モデリング技術、ハイスループット実験やAIロボットなどの実験DX技術、プロセス可視化・計測技術、データ科学的手法などの進展が著しく、材料合成プロセスの世界へデータ科学的側面から足を踏み込む環境が整ってきていると言える。

以下に本提案において取り組むべき研究開発課題について述べる。

① 各材料領域におけるプロセス・インフォマティクス手法の構築

材料合成プロセスは取り扱う材料の種類やプロセスが行われる環境によって全く様相が異なるため、プロセス・インフォマティクスの実現には、まず個別プロセスごとの探索・最適手法を模索・構築していく必要がある。

ただし、プロセスデータの収集・蓄積は共通的な課題である。データ科学的な手法を適用するためには、まず、プロセスデータを収集しなければならない。プロセスのパラメータ空間は材料のパラメータ空間に比べてはるかに広大であり、効率的なプロセスデータ生成/取得手法の確立はプロセス・インフォマティクスを実現するために不可欠な要素である。プロセスデータ生成法としては、大きく分けて、実験のDX化による直接取得、シミュレーションによる生成、文献情報からの抽出の3種が考えられる。ただし、文献情報にはすべてのプロセスデータ（実験条件等）が記載されているわけではないため、プロセス・インフォマティクスでは、実験のDX化による直接取得、シミュレーションによる生成の重要度が高くなる。

個別プロセスとしては、有機材料、無機材料、複合構造材料それぞれの領域から、産業界の具体的なニーズがありプロセス・インフォマティクスによって合成プロセス開発に効果があがると期待されるものであって、かつ、プロセス・インフォマティクス手法構築に適するプロセスであるものを中核として選び、そこから材料領域ごとに横展開していくことが望ましい。

●有機材料系合成プロセス

取り組むべきプロセスとしては、たとえばマイクロフロー化学が有望である。マイクロフロー化学は、精密かつ迅速な温度制御性が可能、反応時間の制御がしやすい、混合が短時間で済み濃度分布などの影響が少ないなどの特徴があり、いわば理想的な反応空間といえることができる。つまり、マイクロフロー化学は、温度や時間などの反応条件の制御・変更が容易であり、様々なプロセス条件を幅広く振って最適値を探すのに適している。また、フロー系であるので、温度・流速や生成物濃度などのインライン計測が可能であり、ハイスループット実験やロボット実験に適した手法といえる。計算科学やデータ科学を活用した合成経路探索手法と、マイクロフロー化学での合成実験を組み合わせることで、目的化合物の合成経路を効率的に確立することができるように期待される。

●無機材料系合成プロセス

取り組むべきプロセスとしては、結晶成長プロセスが有望である。この材料合成プロセスの特徴としては、(a) 結晶成長界面が熱平衡に近い条件下での反応である、(b) ゆっくりとした反応である、の2点があげられる。これらの特徴は、物理化学にもとづく比較的精緻なシミュレーションが可能であることを意味するが、計算に長時間を要する現状ではプロセスの設計には使いにくい。このため、実測可能なデータからシミュレーション結果を高速に推測する機械学習モデルを構築する方法などが検討されている。ひとたび訓練済みの機械学習モデルが構築できると、プロセスの状態を高速に予測できるようになり、プロセスの設計に活かすことができる。将来的には、結晶成長プロセスで得られたノウハウを活用して、複雑性が高く理論モデ

ルの構築が困難な多結晶体やガラスなどの系に拡張していくことが期待される。

●複合構造材料合成プロセス

複合構造材料は、複数の異なる材質の混合物からなる材料であり、材料自体とその合成プロセスの両方の意味で複雑な材料であることが特徴である。燃料電池の電極は、集電体上に触媒・導電性付与材、バインダーなどが塗布された構造であり、複合構造材料の例として挙げられる。このような系では、理想的な組成や構造が明確になっていないため、基本原理が解明されなくても最適な組成や構造を探索できる可能性のあるデータ科学的手法が適している。ただし、その材料とプロセスの複雑さのため、必要なパラメータ数は非常に多くなり、通常の機械学習アルゴリズムがうまく適用できない可能性もある。機械学習の分野でオートエンコーダーとして知られている深層学習の中間層での次元削減操作を、中間データとして物理的に意味のある値を用いて適用することで、プロセスの最適化が行えるという研究例もある。このようなパラメータ数が多いプロセスに適した新しい学習アルゴリズムの探索・構築が強く望まれる。

② プロセス・インフォマティクス共通基盤構築

プロセス・インフォマティクスの共通基盤構築としては、個々のプロセスの要求にこたえるだけでなく、プロセス・インフォマティクスを加速する基盤としての共通化・標準化技術の開発が重要な課題である。

プロセス・インフォマティクス用機械学習アルゴリズムとしては、プロセスのパラメータ空間が極めて広大であることや、多段階プロセスにおいて前後のプロセスの関係が複雑であることなどの特徴に対処するため、中間データを用いた次元削減をするオートエンコーダーを活用する機械学習アルゴリズムなどの開発が必要となる。

実験データの効率的な収集方法としては、IoT(Internet of Things)を利用して、測定装置から測定データをデータベースに自動的に収録するとともに、計測条件や試料情報等のメタ情報を抽出し記録する技術などは共通化・標準化することが望ましい。また、たとえば、成膜プロセスにおける基板ホルダーなどの研究ハード面の共通化を進めることによって、プロセスデータを集めやすくなるような仕組みも重要である。

また、多くのプロセスデータが収集された後に、それを適切に蓄積・運用していくことも重要である。オープンにする範囲、クローズにする範囲、限られたメンバーだけで共有するいわゆるシェアデータの範囲などをあらかじめ明確に定義することが必要である。また、蓄積したデータ自体を秘匿したまま機械学習モデルを構築する手法（たとえば秘密計算技術）が開発できると、蓄積したデータをより幅広く活用することが可能になる。

材料合成プロセスのシミュレーション・モデリングには、大きく分けてa) 実験データに基づく経験パラメータを用いたモデリングと、b) 第一原理計算および分子動力学によるシミュレーションがある。しかし、a) は、実験結果をベースとしているため実験条件の範囲外の予測をすることは困難である。また、b) は、計算に用いる前提が現実系とは異なるため、現実空間におけるプロセス結果との乖離が生じたり、計算コストの制約を強く受けるなどの問題がある。最近研究が進んでいるマルチフィジックス・シミュレーションは第一原理計算と統計熱力学を融合させることによってa)とb)をつなぎ、プロセスにおける各パラメータの寄与を解釈できる形でモデル化できる可能性があり、将来、プロセス科学の共通基盤として幅広い活用が期待できる。

③ プロセス科学基盤拡充

材料合成プロセスの素過程の解明は、プロセスの最適化や新プロセス設計、ひいてはプロセス科学の学理構築の上で重要であるが、素過程の記述・制御に必要なパラメータは実はよくわかっていないのが現状である。そのため多くの場合、これまでの経験に基づいてプロセスパラメータを選択し調整しながら、素過程を推定してきたのが現状である。ここにこれまでのプロセス科学の限界があった。

この課題を克服するためには、これまでの手法に加えて、データ収集を行い、データ科学を活用していくことが重要である。直接得られるプロセスパラメータを使ってデータ科学の手法によりプロセス特性を適切に表す記述子（プロセス記述子）を導出し、物性科学の知識や熟達者の経験知（勘・コツ）に照らし合わせることで合成プロセスにおける重要な素過程を推定することができるであろう。

プロセス記述子によって素過程が記述・表現されると、装置固有であるため各論的な取り扱いしかできないとされていたプロセスを新たなカテゴリーに分類して議論することができるようになる。これにより、ひとつひとつのプロセスの解析ではわからなかったプロセスを制御する重要な共通因子の把握が可能になり、従来は活用が難しかった複数のプロセス間でそれぞれの知見・データなどを活用できるようになる可能性がある。また、熟達者はこのような共通因子を経験知として体得しているがゆえに、適切なプロセスやその条件の選択ができていると思われるが、プロセス記述子をうまく活用することで、暗黙知を形式知化することのみならず、新たな視点や気づきを与えることができるようになると思われる。

このようなアプローチにより、各論的・局所的であった合成プロセスの議論を刷新し、汎用性の高いプロセス科学の基盤拡充につなげることができる。（プロセス記述子の詳細については第3章に記載）

上記で述べた研究開発課題を推進していくために、対象とするプロセスごとにプロセスセンターを設置し、そこに合成プロセス装置、評価・計測装置を整備し、また、研究者だけでなく、プロセス技術者、計測技術者、技術支援員、データ管理者なども配置することが有効であると考ええる。プロセス・インフォマティクス共通基盤については、理論科学、計算科学、データ科学、ロボット技術、プロセス技術、計測技術など非常に多岐に渡る専門家の参画が必要であるが、デジタル・トランスフォーメーションが進んでいる状況では、同じ場所に集まる必要はなく、各々の研究場所で参画しながら連携するバーチャル拠点方式で十分に機能させることができる。ただし、プロジェクトごとの一時的な連携でなく、強力な連携を継続する拠点として設置する必要がある。さらに、プロセス科学基盤拡充のためには、プロセスセンターと共通基盤の拠点を強く連携させることが重要である。また、それぞれの研究開発課題の成果を活かし相乗的に進展させるために、全体を束ねる仕組み（ガバニングボードなど）や学会・シンポジウムが必要である。

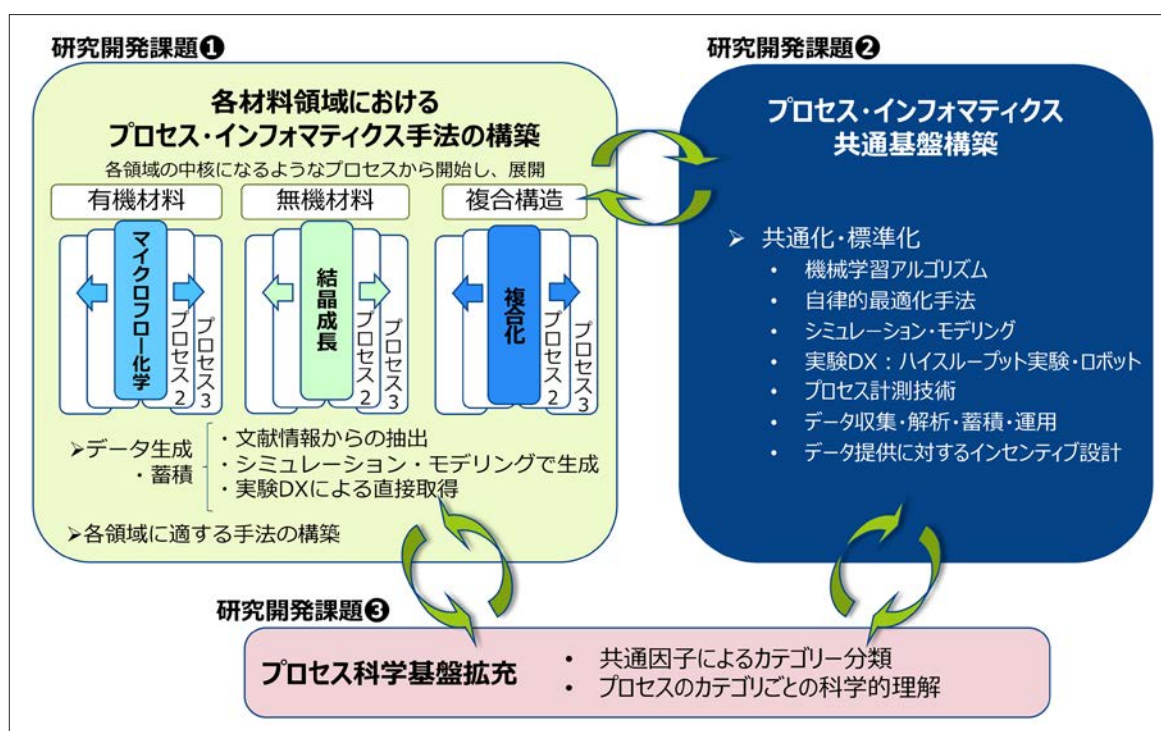


図1-2 具体的な研究開発課題

また、人材育成はきわめて重要である。材料・研究者とデータ科学者の知識・経験を融合させることが大切になる。材料科学・データ科学両方の視点を持った若手人材の育成がこれまで以上に重要である。産業界が参画する仕組みの整備、データ取扱ルールなども重要な課題である。

2 | 研究開発を実施する意義

2.1 現状認識および問題点

現在、ありとあらゆる場面で材料への期待が高まっている。SDGsの達成、温室効果ガス削減のためのカーボンニュートラル社会の構築、低環境負荷の資源・物質循環の実現などに対し、新たな材料の創出は決定的に重要となっている。また、新型コロナウイルス感染症の猛威によって加速されるデジタル化の流れ、わが国がめざすリアルとサイバーが高度に融合した社会（Society 5.0）の実現といった潮流の中でも、ハードウェアやその基盤をなす材料への要求は高まる一方である。

2019年にCRDSが発行した戦略プロポーザル「未来材料開拓イニシアチブ～多様な安定相のエンジニアリング～」(CRDS-FY2019-SP-02)で述べられているように、場合によっては相反しているとさえいえるような複数の物性、例えば、固体電解質における高速Liイオン伝導と広い電位窓、蓄電池正極材料における高Li出入り量と安定性、磁石材料における高保磁力と高飽和磁化、触媒材料における高い触媒活性と安定性・低コストなどを両立することが求められてきている。上記プロポーザルでは、こうした相反する要求に応えるために、材料の多元化や準安定状態を含む未利用安定相の活用を通して、材料探索の範囲を従来よりも広げ、これまで未開拓であった材料創製空間を探索し、新材料を創製することが提案されている。しかし、材料創製空間を拡大するということは、今までに使われていなかったような条件で材料合成をするということを含み、多くの場合、プロセスパラメータをこれまでより広い範囲で振って探索したり、プロセスパラメータの種類を増やしたり、さらにはそれぞれのパラメータをより精密に制御することが必要となる。

このような背景のなか、これまでわが国が得意としてきた技術蓄積をベースにした材料に関する経験知（勘・コツ）を頼りに材料開発を進めることは困難になってきており、高度なシミュレーションやデータ科学などの新たなツールを用いた材料設計が必要となっている。実際、AIやビッグデータの研究開発手法を活用し、研究開発期間の短縮、低コスト化をめざすデータ駆動型材料開発の取り組みであるマテリアルズ・インフォマティクスが世界的に進展している。

マテリアルズ・インフォマティクスは一般的に、計算科学やデータ科学を用いた物性予測、ハイスループット材料合成・評価、データマイニングによる特徴抽出などにより、新材料の設計、探索、発見を飛躍的に加速することを可能にする取り組みを総称した技術体系を表す。

マテリアルズ・インフォマティクスが世界的な潮流となる契機となったのは、2011年に米国が「Materials Genome Initiative (MGI)」という国家プロジェクトを開始したことにさかのぼる。MGIは、物質・材料研究にかかわる計算手法と情報解析手法を統合することによって、新材料の発見から製品化まで20年以上といわれる開発工期を半減させることを目標に掲げた。National Science Foundation (NSF) や National Institute of Standards and Technology (NIST) のファンディングと拠点整備、さらには物質・材料データの豊富なプラットフォームの整備を通じてマテリアルズ・インフォマティクスの研究開発が進んだ。そしてMGIに追随するように、欧州、アジアでも同様の取り組みが急速に進展した。日本では、2013年にCRDSが発行した戦略プロポーザル「データ科学との連携・融合による新世代物質・材料設計研究の促進（マテリアルズ・インフォマティクス）」において、物質・材料の設計に必要な新たな方法論としてマテリアルズ・インフォマティクスの重要性が提案され、その後、日本国内で様々な研究開発プログラムが実施されてきた。

マテリアルズ・インフォマティクスが材料開発において非常に強力なツールとなることは様々な研究事例を

もって実証されてきており、もはや材料研究にデータ科学を活用することは欠くことができなくなっている。近年では国内外の多くの事業者からデータベースやデータ解析、コンサルティング、自動分子設計などのサービスが展開されている。

一方で、現状ではマテリアルズ・インフォマティクス、特に、第一原理計算を併用する手法を通じて発見された新材料・新物質については、多くの場合、候補となる新材料の組成・構造が予測されているだけで、その材料が「実際に作れるのか」、「どう作るのか」、までが示された例はほとんどない。これは、物性物理学の豊富な知見の蓄積により、材料の組成・結晶構造からその物性をかなり正確に予測できるようになっているのに対し、製造プロセスにより材料がどのように合成されていくかに関する予測可能性は非常に限定されていることに起因している。データ科学をプロセス発見につなげるためには、マテリアルズ・インフォマティクスで確立された手法をさらに高度なものに発展させていく必要があると思われる。

材料の合成プロセスは、材料ごとに手法のバリエーションが多く、それを制御するパラメータも非常に複雑であるため、統一的に扱うことは難しい。このため個別プロセスの改良・最適化が主となり、最適なプロセスを科学的に探索するアプローチはとられてこなかった。

しかし、AI・機械学習にみられるようなデータ科学の最近の進展、シミュレーションの高度化、オペランド計測のような合成プロセスをリアルタイムに観測・計測する技術、さらにはロボットを使ったハイスループット実験によるプロセスデータの大量自動取得など、材料プロセスとデータ科学をつなぐためのツール群として以下のような要素技術の進展が著しい。

- | | |
|------------------|---|
| (1) 自動・自律実験技術 | ・ハイスループット合成・評価
・コンビナトリアル合成
・自律実験ロボット |
| (2) シミュレーション技術 | ・反応経路自動探索
・第一原理計算
・3Dモデリング（デジタルツイン） |
| (3) プロセス可視化・計測技術 | ・オペランド計測
・推定制御（ソフトセンサ、バーチャルメトロロジー） |
| (4) データ科学手法 | ・プロセス・インフォマティクス向け機械学習アルゴリズム
・実験データとシミュレーションデータのデータ同化 |

特に、近年スーパーコンピュータの進化が著しい。2021年3月から本格運用が始まった「富岳」においては、AIによる効果的なモデル構築と実装や、高解像度・大規模・長時間のシミュレーションなどが想定されており、プロセス・インフォマティクスでの活用も期待される。また、スパコンのような先端的な計算資源を用いなくても、手元のGPUベースのコンピュータやオープンソースの機械学習ソフトを用いることで材料科学者がデータ科学的アプローチを活用できる時代になってきており、プロセス・インフォマティクスに取り組むための環境は整いつつある。

2.2 社会・経済的效果

本戦略プロポーザルの研究開発を実施することにより、目的材料の合成プロセスを効率的かつ統合的に探索する技術基盤を確立することができる。このことにより、目的材料の合成プロセス開発の成功率を改善することができ、また開発に要する期間を短縮することができる。また、合成プロセスにおける副生成物の抑制、高価な原材料・触媒・反応助剤などの使用量削減、エネルギー効率の改善を図ることも可能になる。さらには、今後想定される規制物質の使用を回避したプロセスの開発など、環境への影響を最小限にする社会的要求にも貢献することが可能になる。

経済的效果を定量的に予測することは難しいが、さまざまな形で具体的効果が想定できる。

たとえば、プロセス開発の期間短縮は、直接、開発費用の削減につながる。米国のMGI（Material Genome Initiative）では、新材料発見から製品化まで20年以上といわれる開発期間を半減させることを目的に掲げた。マテリアルズ・インフォマティクス研究の成果により、目的とする性能を持つ新材料を効率的に探索することができるようになってきたため、最近では合成プロセスの開発が律速になりつつあると考えられる。また、新エネルギー・産業技術総合開発機構（NEDO）の超先端材料超高速開発基盤技術プロジェクト（超超プロジェクト）では、従来の材料開発と比較して、試作回数・開発期間を1/20に短縮することを目標としている。このように、開発期間短縮は重要な目標であるといえる。プロセス・インフォマティクスの活用によるプロセス開発期間短縮の具体的な研究実績としては、高品質のシリコンカーバイド単結晶の合成プロセスにインフォマティクスを活用することで、約20mm径から150mm径までの大径化を約2年間で達成した例があげられる。シリコン単結晶の大径化では、20mm径から150mm径までに約20年かかっている。研究環境の変化などがあるため、単純に比較することはできないが、開発期間を大きく短縮できていることは明らかである。また、シリコン単結晶の大径化には、多くの研究者が携わったことを考慮すると、開発費用の削減という意味では、上記の開発期間短縮以上に大きな効果があるといえると考えられる。

また、合成プロセスにおける副生成物抑制、高価な原材料・触媒・反応助剤などの使用量削減、エネルギー効率の削減は、製造プロセスにおける運用コスト（OPEX：Operating Expenditure）につながる。また、必要最小限の機能を持つ製造装置を設計することが可能になり設備投資額（CAPEX：Capital Expenditure）の抑制も可能になると期待できる。

本戦略プロポーザルの研究開発の最大の経済的效果は、今後の成長が期待される材料市場への早期参入機会の獲得により、世界市場規模の増大と日本のシェア向上の両面において日本の材料産業が成長を拡大できることにある。

プロセス・インフォマティクスを、所望の機能を有する物質・材料を効率的に探索するマテリアルズ・インフォマティクス、プロセスの内部状態や生成物をリアルタイム観測する計測インフォマティクスと組み合わせることで、所望の機能を持つ新材料を他の国よりもいち早く合成することが可能になり、材料開発競争に勝てるようになる。

最終製品の性能向上を実現するため、材料への要求もますます高度化している。それに加え、最近の環境変化によって、材料にとって開発スピードの重要性が高まっている。これは、最終製品に対するニーズの変化が早くなっていること、また、従来、わが国が得意としてきたすり合わせ型開発の環境が変わってきていることなどが背景にある。

材料への要求が高度化したことにより、新規物質・材料創製の難度が上がり、開発期間の長期化や開発コスト上昇が引き起こされている。それにもかかわらず、最終製品に対するニーズの変化は早く、製品寿命が短くなる傾向にあり、そのため、製品性能を支える材料も、短期間で置き換えが起こる状況になってきている。このような環境にあるため、材料開発スピードが強く求められるようになっている。

また、従来、わが国が得意としていたすり合わせ型開発の環境の変化によっても、開発スピードの重要性が増加している。これまでは、強い最終製品を提供する川下企業との協力関係のもとで新規材料を開発することができ、材料に対する将来の要求性能を比較的早い段階から知ることができる状況にあった。しかし、現在では、要求性能が開示されてから採用までの期間がますます短くなる傾向にあり、これに対応するためにも材料開発スピードは重要となっている。

このように、プロセス・インフォマティクス導入による所望の機能を有する物質・材料の実現、市場参入機会の早期獲得による世界市場におけるシェア拡大により大きな経済効果が期待できる。

また、既存材料についても、環境規制などの変化に対応するための原材料転換や副生成物・廃棄物の削減、また、カーボンニュートラル社会に貢献するためのエネルギー効率改善などの要求に応えるために、合成プロセスの変更が必要になるケースがある。このような場合にも、プロセス・インフォマティクスを活用することが有効である。

本戦略プロポーザルで対象とする材料合成プロセスが適用できる有機材料、無機材料、複合構造材料の市場は、下記の〔参考：関連市場の規模・成長率〕に示されるように、現時点においても合計10兆円を優に超える規模であり、さらに今後の市場の伸びも予想されている。以上述べてきたように材料開発においてボトルネックとなりつつある合成プロセス開発にプロセス・インフォマティクスを活用することは大きなメリットをもたらす。

〔参考：関連市場の規模・成長率〕

(1) 有機材料

有機化合物については、医薬品およびその原料を中心に、求める薬効を有する化合物の構造設計やその合成経路の探索にインフォマティクスが活発に使われてきている。ただし、医薬品とその原料以外で取り組まれている例はまだ多くない。有機材料分野には、半導体用フォトレジストや液晶用材料、有機EL用発光材料、印刷用インクなど高付加価値のものが多くあり、新材料の開発も盛んである。今後、有機材料分野でも新材料開発にデータ科学の活用が広がっていくことが期待されている。

NEDOが実施している主要先端製品・部材の売上高と世界シェア分析¹によると、これらの有機材料の世界市場は以下のようにになっている。()内の数字は日系メーカーのシェアを示す。

半導体製造用フォトレジスト（ArF用、KrF用、UV用の合計）は、2014年1027億円（83%）、2018年1257億円（83%）である。また、液晶用材料（液晶用材料、フォトレジスト、カラーレジスト、ブラックレジスト、配向材料の合計）は、2014年4421億円（55%）、2018年4025億円（53%）、印刷用インクは、2014年1兆4150億円（49%）、2018年1兆3531億円（49%）、また、有機EL低分子発光材料は、2014年345億円（21%）、2018年522億円（16%）である。以上に例示した市場だけでも市場規模は、約2兆円である。

また、それぞれの年平均成長率（CAGR）としては、半導体プロセス用フォトレジストは10%（2024年までの5年間、富士経済）、印刷用インクは5%（2024年までの5年間、TechNavio社）、有機ELディスプレイは17%（2025年までの5年間、富士キメラ総研）が予測されており、これらは引き続き大きな成長が期待される市場である。

1 新エネルギー・産業技術総合開発機構（NEDO）「2019年度成果報告書 日系企業のITサービス、ソフトウェア及びモノの国際競争ポジションに関する情報収集 情報収集事業(2)「モノを中心とした情報収集と評価」」令和2年3月（委託先 株式会社富士キメラ総研）

(2) 無機材料

無機系材料のプロセス技術としては、磁性材料や誘電体材料などセラミックス電子材料の製造で用いられる焼結プロセス、半導体デジタル・アナログ素子や熱電、発光素子などで利用される結晶成長プロセスが挙げられる。これらのプロセスに関しては様々な材料に対して最適化研究、実用化研究が進められ、それらの材料に応じた独自プロセス技術が開発されてきた。一方で、マテリアルズ・インフォマティクスの進捗により多くの新材料が提案されているが、従来の製法では作製が困難な材料、プロセス最適化に膨大な時間を要する材料も多く、また、研究者・開発者の経験知（勘・コツ）が要求されるプロセスも存在しており、プロセス・インフォマティクスの早期導入、構築が望まれている。

[焼結プロセス]

NEDOによれば、世界市場は以下のようになっている。なお、() 内の数字は日系メーカーのシェアを示す¹。
焼結磁石（レアアース）は2014年0.45兆円（22%）、2018年0.51兆円（23%）、セラミックスコンデンサーは2014年0.8兆円（53%）、2018年1.0兆円（56%）、自動車用触媒は、2014年1.2兆円（24%）、2018年1.3兆円（24%）、である。

永久磁石合計の世界市場は2019年で2.1兆円、2027年には14.5兆円（グローバルインフォメーション(株)）と予測されており、例えば、自動車用途では、現在1千万台/年であるxEV（HEV/EV、FCVなど）も2050年は1億5千万台/年になると推計され、EV一台あたりの磁石使用量を2kgとすると、現在の磁石世界生産量の約5倍がモーターに必要と言われている²。また、環境触媒についても、2017年で1.6兆円、2030年には2.2兆円（富士経済）と予測され、市場は拡大傾向にある。新たな代替材料の開発、市場への早期投入によるシェア拡大により、極めて大きな経済効果が期待される。

[結晶成長プロセス]

NEDOによれば、世界市場は以下のようになっている。なお、() 内の数字は日系メーカーのシェアを示す¹。
LEDは2014年1.6兆円（22%）、2018年2.1兆円（14%）、また、富士キメラ総研によると2025年で2.1兆円と予測されている。Si太陽電池は、2014年3.5兆円（10%）、2018年4.7兆円（2%）であり、太陽電池全体の市場は2030年で6兆円（予測）と予測されている³。

これらの材料・デバイスでは市場は微増傾向であるが、いずれもこの4年間で日本はシェアを大きく下げており、プロセス技術革新による低コスト化、競争力強化を実現し世界シェアの巻き返しが求められる。

(3) 複合構造材料

ここでの複合構造材料は、最終製品/生成物となった時点で、複数の異なる材料・構造の混合物からなる材料のことを指す。各種の構造材料やリチウムイオン電池（LIB）などの二次電池電極、燃料電池電極などが例にあげられる。

LIB正極材料では、2014年0.35兆円（シェア29%）、2018年0.87兆円（18%）である¹。また、LIBそのものの市場も2020年4.7兆円から2024年予測9.5兆円とほぼ2倍に拡大する見通しである（富士経済）。xEV（HEV/EV、FCVなど）への移行、生活スタイルの変化による新たな需要喚起で更なる需要拡大が見込まれている。

炭素系複合材料では、2014年0.41兆円（41%）、2018年0.77兆円（シェア42%）であり¹、PAN系炭素繊維複合材料の世界市場動向では、2020年1.6兆円から2030年予測では3.2兆円（富士経済）と報告されており、航空機、風力発電、自動車への展開が継続すると予想される。

2 新エネルギー・産業技術総合開発機構（NEDO）材料・ナノテクノロジー部 「次世代自動車向け高効率モーター用磁性料技術開発」（2019年8月）

3 技術戦略研究センターレポート TSC Foresight「太陽光発電分野の技術戦略策定に向けて」（2016年6月）

2.3 科学技術上の効果

本戦略プロポーザルの研究開発を実施することにより、材料合成プロセスの科学的アプローチの進展、マテリアルズ・インフォマティクスとの融合による材料創製技術の発展、材料創製から製品化へのシームレスな橋渡しなどの科学的効果が期待できる。以下、それぞれについて述べる。

(1) 材料合成プロセスの科学的アプローチの進展

材料合成プロセスや合成条件の探索・選定は、これまで、経験や暗黙知をもとに行われ、その探索範囲は、個々の研究者あるいは研究室や研究機関の経験により決められる場合が多かった。

このような経験に基づくプロセスの最適化は、実験を繰り返し行い、特性・性能がより高くなるようにプロセス条件を絞り込んでいくことで導き出されているため、局所最適化になっているケースも多く、また必ずしもその背景にある科学的現象を十分に理解して行われている訳ではなかった。本プロポーサルで掲げるプロセス・インフォマティクスでは、ハイスループット実験やAIロボットなど効率的な実験手法で解析用の大量のデータを習得し、データ科学的手法を用いて多次元のパラメータ空間から重要な因子を抽出することができ、その結果としてプロセスの理解に貢献できると考えられる。

また、これまで材料合成プロセスは、プロセスごとに各論的・局所的に議論されてきていたが、データ科学的手法を用いることで異なるプロセス間の共通因子を見つけられる可能性がある。このことにより、ひとつひとつのプロセスの解析ではわからなかったプロセスを制御する重要な共通因子の把握が可能になり、プロセスの理解がさらに進むと予測される。このように、実験科学、理論科学、計算科学、データ科学の4つの科学を融合的・統合的に活用することにより、プロセスの科学的理解を深めることに貢献できると考えられる。

プロセスの背景にある科学的な現象の理解を深めることで、従来よりも優れたプロセス候補を提案することも可能となる。従来のプロセスの延長線上ではなく、別の方法、例えば従来はスパッタ法が最良と認識されていた合成方法に対し、CVD法やMBE法がより良い方法として提案できることも考えられる。また、同じ特性・性能をより簡潔な方法で合成するプロセスの提案も可能となる。さらに、異なるプロセス間で、従来は活用が難しかったそれぞれの知見・データなどを活用できるようになる可能性もある。このように、これまでの経験や暗黙知の活用では難しかった、新たなプロセス手法を生み出す可能性があり、材料創製分野の飛躍的な発展が期待できる。

(2) マテリアルズ・インフォマティクスとの融合による材料創製技術の発展

物質・材料創製の分野では、マテリアルズ・インフォマティクスが積極的に活用され始めている。データ科学を駆使することで、試行錯誤的に実験を行うことなく、目的とする特性・機能を有する物質・材料の構造・組成を効率的に見いだすことも可能となってきており、国内外で多くプロジェクトが進行中で、成果も出始めている。しかしながら、マテリアルズ・インフォマティクスでは、目的とする特性・機能を有する物質・材料を提案することはできても、どうやって合成するか情報は提供されない。このため、実際の合成に多くの時間と労力を要し、合成することができない非現実的な材料が提案された例もある。対象とする物質・材料の最適な合成方法を提案することを目的とするプロセス・インフォマティクスの研究開発は、マテリアルズ・インフォマティクスと相補的な関係にある。両者が融合することで、「目的とする特性・機能の実現のためには、どのような物質・材料をどうやって合成するか?」という物質・材料創製の一連の流れを統合的に進めることが可能となる。マテリアルズ・インフォマティクスで提案された物質・材料の合成プロセスが効率的に確立できるようになり、データ駆動型物質・材料創製技術の統合的な発展が期待できる。

(3) 材料創製から製品化へのシームレスな橋渡し

プロセス・インフォマティクスは、マテリアルズ・インフォマティクスによる分子・材料設計と、量産プロセスの構築・確立を結びつけ、材料設計から製品化までをつなぐ役割を担うことができる。研究室レベルの実験装置を対象とした材料合成プロセスの最適条件の導出にとどまらず、製品を製造する量産装置におけるコントロールまでを視野に入れたプロセス提案が可能となる。

産業界における製造においては、高い生産性だけではなく、プロセス条件の変動に対しても安定に生産できるロバストなプロセスであることが重要である。また、製品の品質管理のためにはプロセスのオンライン監視・制御がしやすいプロセスであることも必要とされる。プロセス・インフォマティクスを活用することで素過程の科学的理解が進み、製品品質や収率などへの各プロセスパラメータの影響度合いも予測可能になることが期待される。このように、量産プロセスに必要な要素を含めてプロセスを最適化できるようになると期待される。

プロセス・インフォマティクスとマテリアルズ・インフォマティクスが融合することで、安定生産しやすさを考慮した物質・材料の提案も可能となると考えられる。また、材料設計段階において、製品の量産プロセスでの品質管理手法までを同時に考慮することが可能となり、材料創製から製品化へのシームレスな橋渡しが実現できると期待される。

3 | 具体的な研究開発課題

ここでは、4つの科学（実験科学、理論科学、計算科学、データ科学）を融合的・統合的に活用して目的材料の合成プロセスを効率的かつ統合的に探索する方法であるプロセス・インフォマティクス技術基盤を確立し、材料創製技術を革新するプロセス科学基盤を拡充するために重要な研究開発課題について述べる。具体的には、①各材料領域におけるプロセス・インフォマティクス手法の構築、②プロセス・インフォマティクスの共通基盤、③プロセス科学基盤の拡充を可能にする新たな指針やコンセプトの創出である。

3.1 各材料領域におけるプロセス・インフォマティクス手法の構築

3.1.1 プロセス・インフォマティクスに共通する課題

材料合成プロセスは取り扱う材料の種類や、プロセスが気相/液相/固相のいずれで行われるかによって大きく様相が異なる。したがって、プロセス・インフォマティクスの実現には、まず技術領域ごとに探索・最適手法を模索・構築していく必要があるだろう。しかし、具体的手法に差異があっても、プロセスデータの生成・利用・蓄積・活用などの要素はプロセス・インフォマティクスを実行するにあたっては共通に必要となるため、まずはその共通する課題について述べる。

(1) プロセスデータの生成

データ科学的な手法を適用するためには、まず、プロセスデータを収集しなければならない。マテリアルズ・インフォマティクスで問題になったのは、材料科学においては、データ科学が前提としている十分な数のデータが存在することが稀であるという点であった。この事情はプロセス・インフォマティクスでも同様であり、プロセスのパラメータ空間が材料のパラメータ空間に比べてはるかに広大である点が、問題をより深刻化させている。効率的なプロセスデータ生成/取得手法の確立はプロセス・インフォマティクスを実現するために不可欠な要素となる。

プロセスデータ生成法としては、大きく分けて、実験のDX化による直接取得、シミュレーションによる生成、文献情報からの抽出の3種が考えられる。インフォマティクスに必要とされるプロセスデータ量の膨大さから、昔のように人間がデータ読み取りやデータ入力を行うことでは対応できなくなっていることは言うまでもない。データ取得の段階からデータの整理・編集などのキュレーションや、データ格納に至るまでを一貫してコンピュータの中で実施する必要がある。

一番目の実験DX化に関して必要な要素としては、多数回の実験を並列高速に行うハイスループット実験（合成・計測）や、従来人間が行ってきた複雑な実験操作を人間に代わって休みなく行うロボット、プロセスから必要なデータを取得するための新たなプロセス計測技術の開発などがある。ハイスループット実験装置やロボットの開発には、材料合成の専門家と機械装置の専門家の高度な協調が求められ、また、プロセス内のこれまで把握するのが難しかったデータをリアルタイムで取得するためには材料合成の専門家、計測専門家との緊密な連携が必要である。

プロセス・シミュレーションに関しては、非平衡でかつ、多くの場合、非定常でもある材料合成反応場を高速かつ高精度に予測するシミュレーション・モデリング技法の開発が望まれる。サブナノスケールで起こる化学反応から、マクロスケールで起こる原料・生成物の輸送に至る複雑系の長時間（実時間）シミュレーションが実行できるようにすることが最終目標である。

文献情報に関しては学术论文や特許だけでなく、過去の実験記録、便覧・データブックなどの幅広い言語・数値情報から、プロセスパラメータの収集ができるようにする必要がある。その際、データの整理・編集が必要であり、当面の間はその領域に関する知見（領域知：domain knowledge）を持つ人間の関与が必要だが、最終的にはほぼ自動に収集できるようになることが理想である。

(2) プロセス構築基盤の発見・確立

蓄積されたプロセスデータは、それ以降の材料合成に活用されることになる。プロセスデータのうち、直接制御可能なプロセスパラメータ（反応時間、原料投入速度、ガス組成・分圧、ヒータ温度、投入電力など）と、直接的には制御できないが反応中にモニタリングを行いながら、フィードバックすることで制御する値（たとえば炉内温度分布、融液対流速度、成長中の結晶直径など）は、いずれも、プロセスのレシピとして再利用できるプロセス制御パラメータとなる。こうしたプロセス制御パラメータの再利用は、通常のプロセス管理/設計で行われてきたものであり、それらのデータ自体は、各製造現場においてノウハウ的に維持・蓄積されている。「同じ場所で同じ装置を用いる限り、製造される材料デバイス構造が許容誤差範囲内で同一である」ことを保証するような制御パラメータ群、すなわち「その通りに実施すれば必ず所望の製品が得られるレシピ」こそが、企業の競争力の要である。特に、半導体プロセスのような非常に長い工程を持つ製造プロセスにおいては、個々のプロセスの再現性・制御性の高さが、最終生成物の良品率（歩留まり）、製造コストに大きな影響を与えるので、各社の重要な財産であるといえる。企業においてはプロセス制御パラメータの蓄積のために多大なリソースが投入されているため、厳重に管理され、外部に公開されるようなものではない。

そのようなプロセス制御パラメータを生み出すためには、以下のように必ずしも科学的に説明しきれない方法が利用される場合がある。たとえば、最終的なプロセス最適化のために、パラメータセットに対しての網羅的探索を行うべきであるが、そのパラメータ選定のために調べるべき空間は、一般的に広大であり、目的とする製品が従来よりも複雑化・精緻化するにつれ、その広さはますます膨大なものになってくると考えられる。広大なパラメータ空間全域について網羅的探索を行うことは不可能であるため、現実的には集中的にパラメータ探索を行う場所を限定せざるをえない。そうした「各段階でどの手法（装置、プロセス）を使うのか」とか「パラメータ空間のどのあたりで探索するのか」とかといった判断は、現状では熟練者の勘・コツ・経験などを根拠に決められている。このようなプロセス設計の判断の前提となる知見をプロセス構築基盤とここでは呼ぶことにすると、そのような、プロセス構築基盤を発見・確立することが、プロセス・インフォマティクスの究極の目的である。プロセス構築基盤の実体が、「どのようなものになるのか」、「なるべきなのか」を現時点で見定めることは難しいが、熟練者の勘やコツといった数値化が難しいものを伝承し発展させるためには、これを再利用可能な形で記述・収集することが重要となる。

上に述べた課題の多くはデータ科学によってある程度は解決することが期待できる。どのようなプロセスパラメータで合成実験を行うかについては、確率分布にもとづくベイズ最適化を用いた自動探索手法が知られており、この手法で制御される合成ロボットも実現しつつある。プロセス中の直接測定しにくい値を、容易に測定可能な他の測定値から推測するソフトセンサー技術は熟練者の経験知（勘・コツ）の代替になる可能性がある。過去の大量の文献情報からプロセスデータを抽出することができるテキストマイニングにより、領域知や暗黙知からの推論と類似のことが可能になることも期待される。プロセス構築基盤全体が情報科学の高度な利用なしには構築が難しいと考えられる。このような点からも、材料科学者、プロセス技術者と、情報科学者のコラボレーションが、不可欠であることがわかる。

3.1.2 各材料領域におけるプロセス・インフォマティクス課題

次に個々のプロセスごとの課題について述べる。材料合成プロセスは取り扱う材料の種類によって大きく様相が異なるので、有機材料、無機材料、複合構造材料、それぞれの材料領域から、**産業界の具体的なニーズがあり、プロセス・インフォマティクスによって合成プロセス開発に効果があがると期待されるものであって、かつ、プロセス・インフォマティクス手法構築に適するプロセスを中核として選び、そこから材料領域ごとに横展開していくことが望ましい。**これによりそれぞれのプロセスで得られた技術・手法は、他の材料領域でも応用されていくことが期待できる。

(1) 有機材料系合成プロセス

有機材料系合成プロセスでは、取り組むべきものとして、たとえばマイクロフロー化学による有機合成プロセスがあげられる。

有機合成プロセスの特徴としては、かなりの規模の反応データベースが整備されていることが挙げられる。過去の文献記載の反応などを収録したデータベースとしてはSciFinder、Reaxys、Synthia、のようなものが欧米を中心にすでに整備されており、発表される文献データなどを随時取り込み更新され続けている。しかし、論文などには、反応温度や反応時間、反応物や触媒の量などのような基本的な条件は記載されているが、プロセスのすべての条件が記載されているわけではないことは注意が必要である。特に、バッチ反応における反応容器内の温度分布やフロー反応における流速分布など、重要であるにもかかわらず多くのケースで記載されていないパラメータも多く、これらのデータベースのみで目的とする化合物を合成する経路を探索・設計するのは困難である。

既存の反応データベースを用いると、候補となる合成ルートを多数見つけだすことはできるが、それらの中から適切な合成経路を選択するのは、これまでは有機合成化学者の経験や知識、またそれらに裏打ちされた直感によるところが大きかった。しかし、近年の計算科学の進歩による反応経路自動探索（GRRM法）や遷移状態高速解析、また、データ科学を活用した逆合成手法（コラム1参照）など、多数の合成経路を比較し最適な合成経路を探し出す手法が提案されており、実際の有機合成ルートの探索に活用できる段階になっている。

これらの手法で探し出した合成経路を検証し、最適化する実験的手法としては、マイクロフロー化学の活用が適している。マイクロフロー化学は、数mm径以下の流路、またはその流路内の数 μ L程度の液滴で化学反応を行うものであるが、精密かつ迅速な温度制御が可能、反応時間の制御がしやすい、混合が短時間で済み濃度分布などの影響が少ないなどの特徴があり、いわば理想的な反応空間といえることができる。つまり、マイクロフロー化学は、温度や時間、反応条件の制御・変更が容易であり、様々なプロセス条件を幅広く振って最適値を探すのに適している。また、フロー系であるので、温度や流速などのインライン計測が可能であり、また、直接の測定が難しい反応生成物濃度についても赤外吸収スペクトルの特定吸収波長などからリアルタイムに推定できるソフトセンサー技術の活用も検討されており、ハイスループット実験やロボット実験に適した手法といえる。さらに、使用する原材料も微量で済むというメリットもある。

また、マイクロフロー化学でのハイスループット実験やロボット実験では、目的物の収率が低いケースなどのいわゆるネガティブデータを含めて、多量の実験データを得ることができる。これらをデータベースに蓄積することで、より実用性の高いデータベースを構築することができる。計算科学やデータ科学を活用した合成経路探索手法と、マイクロフロー化学での合成実験を組み合わせることで、目的化合物の合成経路を効率的に確立することができるようにと期待される。

マイクロフロー化学での検討では、例えば、わずかに温度が変化した場合や流速が変化した場合の収率変化を比較的容易に把握することができる。このようなデータを活用することで、スケールアップ時の反応容器内の温度分布や不均一流速の影響を見積もることも可能になる。このようにマイクロフロー化学での検討結果

は、スケールアップ検討の際の重要な指針を与えてくれるものであり、量産スケールの有機化学品製造にスムーズに橋渡しするためにも重要である。

最近、マイクロフロー化学での高分子精密重合研究なども行われており、マイクロフロー化学におけるインフォマティクス適用の範囲は今後広がっていくと期待される。

(2) 無機材料系合成プロセス

無機系材料プロセスで取り組むべきものとしては、たとえば結晶成長プロセスが挙げられる。

この材料合成プロセスの特徴としては、(a) 結晶成長界面が熱平衡に近い条件下での反応である、(b) 非常にゆっくりとした反応である、の2点があげられる。これらの特徴は、物理化学にもとづく比較的精緻なシミュレーションが可能であることを意味する。融液、溶液からの結晶成長にしても、気相輸送による結晶成長にしても、熱エネルギーや原料物質の輸送を記述する理論（熱力学、流体力学）は確立しており、十分な計算資源があれば、かなり正確に実験結果を予測することができる。また、よく取り上げられる材料系については結晶の原料物質の熱力学データベースや物性データも入手しやすく、シミュレーションの精度を上げやすくなっている。

こうした特徴は、プロセス・インフォマティクスを早期に実現するのに有利に働くが、それでもなお、現状はコンピュータ上でプロセス状態の再現を可能にするデジタルツインを構成できている状況とは言い難い。問題点としては、(i) 熱や物質の流れを正確に予測するためのシミュレーションにはかなり長い計算時間を要する、(ii) 反応に関与するすべての物質の物性データが知られているわけではない、(iii) シミュレーションに必要ながそもそも測定自体が困難なパラメータが存在する、などがある。

(i) については、正確なシミュレーションが、デジタルツイン形成の根幹的要請であるのはもちろんだが、膨大なパラメータ空間に対しての予測が必要なプロセス・インフォマティクスにおいては、ひとつの計算に長時間を要することは可用性の面で致命的である。このため、いくつかの正確なシミュレーション結果を元データとして、実測可能なデータからシミュレーション結果を高速に推測する機械学習モデルの構築などが検討されている。ひとたび訓練済みの機械学習モデルが構築できると、シミュレーションよりも桁違いに速く熱や物質の流れを予測することができ、この予測を用いてプロセス・インフォマティクスによるプロセスの設計が可能になる。もちろん、正確なシミュレーション自体をより高速化するような、シミュレータ技術の開発も不可欠である。たとえ、機械学習モデルを利用する場合であっても、その元となるシミュレーションが高速化すれば、モデル自体の精度が向上することとなり、正確なデジタルツインの構築に有用である。

(ii) については、オペランド計測の手法を拡充し、これまで取れていなかったプロセス中の微小な領域の物性値を正確に測定することが、まずは王道となる。材料合成の反応空間は一般には、計測環境とは異なる閉鎖空間である場合が多く、また計測のための仕組みが反応環境に影響を与えてしまう場合が多いが、それらの課題を乗り越える計測手法が編み出されれば、実測結果ほどシミュレーションの役に立つパラメータは他にはない。今後も、実環境下で実時間測定を行ううえで、オペランド計測手法の研究開発に期待が持たれる。一方で、(iii) にあげられるパラメータ、例えば、成長表面における固液界面の熱抵抗とか、成長表面の界面エネルギーのような、実環境での計測が難しい量や、(ii) で挙げられたまだ測定されていない材料の物性パラメータなどについては、モデルから推定する方法も行われ始めている。これは、シミュレーション結果を再現するように機械学習モデルに使われる物性パラメータを変えることで、真の物性パラメータを推定する方法である。こうしたやり方で、本質的に測定が難しい量を推定できるように、活用できるデータを増やしていくことで、高速な予測が可能なデジタルツイン構築に近づけていくことができる。

さらに将来的には、結晶系で得られたノウハウを活用して、これらの結晶成長に比べると複雑性が高く、完全な理論モデルの構築が困難な多結晶体やガラスなどの系についても、拡張していくことが期待される。

(3) 複合構造材料系合成プロセス

ここで述べる複合構造材料は、最終製品/生成物となった時点で、複数の異なる材料・構造の混合物からなる材料のことを指す。

前項の結晶成長のケースの場合は、成長中には異なる化学組成の相が共存する場合があるが、最終的な生成物としては異種材料が混在した状態を考える必要はない。生成物である結晶の品質を記述するための不純物濃度、転位密度、残留応力といったよく定義された評価指標が存在するうえ、その理想状態としての完全結晶が想定できるので、製造プロセスの良さ・悪さを生成物の評価から決定することができる。得られた結晶がどれだけ完全結晶に近いかという評価指針が存在しているともいえる。

これに対して、複合構造材料は、複数の材料（相）が特定の構造（組織）をもって共存していることによりその機能が発現しているため、一般にその材料における理想的状態が明らかではなく、目的の機能を実際に測ること以外にプロセスの良否を評価する方法はない。

固体高分子型燃料電池の電極を例にとると、電極は、外部に電流を取り出す集電材料、触媒である白金粒子、白金粒子を保持するとともに電極に電子を伝える役割のカーボン粉末、電極全体の構造を固定するバインダーと電解質にイオンを伝えるイオン導電体を兼ねるイオノマーからなる。電極に期待される機能要件は、触媒表面上での燃料や酸素の酸化還元反応が速やかに行われること、イオン伝導や電子伝導による電気抵抗が小さいこと、燃料や酸素が十分に速く触媒表面に到達できること、などがあるが、それらを実現するための、理想の組成や構造が明らかにされているわけではなく、むしろ、これらの要求の間には、それに有利な組成や構造の点で相反するものが多く含まれており、すべてを満たす目指すべき姿を設定しにくい。

また、仮に理想の組成や構造が見つかったとしても、10 nm～10 μm にわたるスケールの構造を持つ有機物、無機物、金属の混合物を望んだとおりの構造に組み上げることも容易ではない。一般的な複合材料の状態が、完全結晶のように絶対零度での熱力学的平衡状態として定義される理想状態ではなく、多くの場合は無数に存在する準安定状態の一つであるためである。

複合構造材料の合成プロセスとしての難しさには、(a) 生成物が多数存在し、それらの相互作用が完全には理解されていないこと、(b) 生成物が熱平衡から遠いため単結晶のように「平衡状態に近づけるほど良いものができる」という戦略が取れないことなどの生成物自体の複雑さからくる要素と、そのような材料を何とか合成するために、(c) 既存の合成プロセスが必ずしも平衡状態を用いたものではなく、速度論的アプローチに頼らざるを得ない非平衡・非定常であるケースが多いこと、(d) そのようなプロセスにはパラメータが多いこと、非線形性が強く直感が働きにくいことなどの既存プロセスが共通に抱える課題がある。言い換えると、(a)、(b) は、材料として機能の発現を完璧に取り扱う理論・モデルの構築ができていないこと、(c)、(d) は、そういう状態であるために、既存の合成プロセス自体が、十分な科学的根拠に基づかず、試行錯誤や熟練者の経験知（勘・コツ）によって構成されていることに起因する。複合構造材料は、材料自体とその合成プロセスの両方の意味で複雑な物質であるといえる。

このような機能原理の解明自体が難題となるような複雑な材料の合成プロセスの検討においては、おおもとにある基本原理が解明されなくても、最適なあるいは最適に近い解を探索できるデータ科学的手法が適している。実際に、入力としてのプロセスパラメータから、目的性能を予測する機械学習モデルを作り、それによりプロセス最適化を目指す試みも始まっている。理論に基づいた信頼性の高いシミュレーションができない複雑系の場合、機械学習に使用するデータは実験から得る必要があるため、ハイスループット実験や、人手を介さないロボット実験環境などの実験 DX 構築が、複雑系材料のプロセス・インフォマティクス構築には非常に有効な手段となる。また、一般的には非定常/非平衡で行われる材料合成のパラメータを可能な限り多く収集するための、様々なプロセス計測手法の開発も重要な要素となる。

さらに、実験 DX でデータを大量に収集する工夫を行ったとしても、非平衡プロセスを特徴づけるのに必要なパラメータ数は平衡状態で行われるプロセスよりも桁違いに多いために、普通の学習アルゴリズムがうまく適用できない可能性もある。機械学習の分野でオートエンコーダーとして知られている深層学習の中間層での

次元削減操作を、中間データとして物理的に意味のある値を用いて適用することで、あるプロセスの最適化が行えたという報告例もある。このような、パラメータが多く、学習データが少ない場合にも適応可能なプロセス・インフォマティクスに適した新しい学習アルゴリズムの探索・構築も強く望まれる。

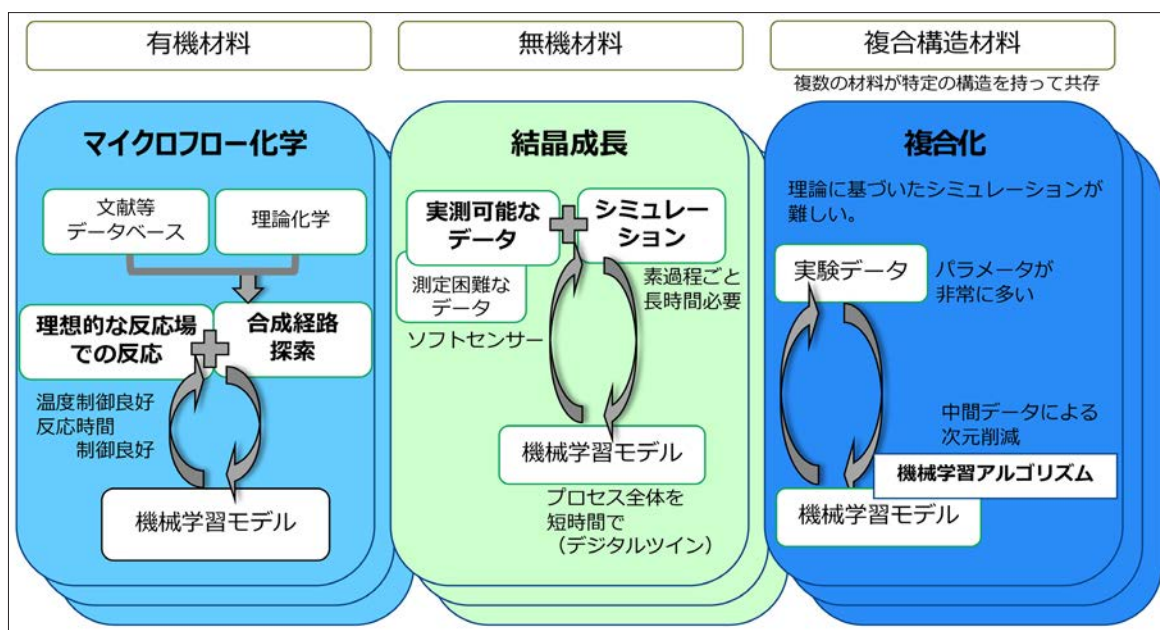


図3-1-1 各材料領域におけるプロセス・インフォマティクス課題

コラム1

機械学習を用いた逆合成

所望のターゲット物質の効率的な合成経路の設計（逆合成）は、間違いなく最も複雑な問題の1つであり、プロセス・インフォマティクスの中心課題である。従来は、ルールベースの方法で、ルールを人手で列挙しながら面倒な手動エンコーディングすることで進歩してきたが、データセットサイズの増加に伴い格段に手間や難易度が上がるとともに、既存のすべてのルールと新しいルールの間の論理的な一貫性を評価する際の複雑さは、コード化されたルールの数とともに増加し、問題が手に負えなくなるレベルに達するのは時間の問題である。Google DeepMindが開発した囲碁AI（人工知能）「AlphaGo」が登場する2016年頃から、化学合成経路探索の機械学習のアプローチが出てきており、米国特許明細書から化学反応を集めて整備されたUSPTO（米国特許商標局）データセットの公開（2017年）とともに、ルールベースではない機械学習を活用した手法が急激に進んでいる。

Coley（MIT）らは、目標とする生成物と既存の反応データベースのマッチングを行い、原材料と生成物の類似度をそれぞれ計算することで、その類似度に基づいて

て一段階前の反応候補のランキングを出力する方法を開発している[1]。彼らの手法は、確かに、既存の方法に比べておおむね高精度であるが、既存の反応データベースとのマッチングに基づくため、探索空間は既存の反応に限定される。そこで、福島（トヨタ）と本山および吉見（東大）は、敵対的生成ネットワーク（Generative Adversarial Network：GAN）を用いた逆合成の経路探索を試みている[2]。彼らの狙いは、GANに基づく反応物の生成モデルを学習し、反応物を生成した後に順方向の反応予測を行って反応式を生成することにより、探索空間の拡大が期待できる。

Segler（WWU Münster）らは、囲碁ソフト AlphaGo に倣って、自動逆合成解析ソフト AlphaChem の開発を近年試みている[3]。化学反応データセット Reaxys を使用して、生成物の反応分類をニューラルネットワーク（Neural Network：NN）で行い、AlphaGo と同様にモンテカルロ木探索で探索することで、既知の合成経路を高精度・高速で見つけることができ、高精度・高速な逆合成経路探索を可能にした。ただし、AlphaChem は、化合物構造式の部分構造の有無を数値化したベクトルで表現した ECFP（Extended-Connectivity FingerPrint）で特徴量抽出していることから、化合物構造式をグラフと見なしてグラフ畳み込みネットワーク（Graph Convolutional Networks）など使って精度上げる方法も考えられ、まだまだ発展の可能性を秘めている。

Schwaller（IBM Research Zurich）らは、人間の介入なしに逆合成を設計するためのハイパーグラフ探索戦略と組み合わせた分子トランスフォーマーアーキテクチャの拡張を提示している[4]。ここで、分子トランスフォーマー（Molecular Transformer）とは、時系列データを逐次処理する必要がない深層学習モデル（Transformer）に基づいて、データセット内に存在する反応物、試薬、および生成物における化学モチーフの存在と非存在の間の相関を推測することにより、正しい化学選択性、位置選択性、および立体選択性に関して不確実性も含めて予測を行うモデルである。彼らの訓練済み機械学習モデルは、クラウドベースの「IBM RXN for Chemistry」プラットフォームとして無償公開されている。

《参考文献》

- [1] Connor W. Coley et al., “Computer-Assisted Retrosynthesis Based on Molecular Similarity,” ACS. Cent. Sci. Vol. 3 (2017) 1237-1245.
- [2] 福島真太郎, 本山裕一, 吉見一慶, 敵対的生成ネットワークを用いた逆合成の経路探索, 第41回ケモインフォマティクス討論会（2018年10月26～27日, 熊本）講演要旨集 p. 2C12.
- [3] Marwin H. S. Segler, Mike Preuss, Mark P. Waller, “Planning chemical syntheses with deep neural networks and symbolic AI,” Nature Vol. 555 (2018) 604-610.
- [4] Philippe Schwaller et al., “Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy,” Chem. Sci. Vol. 11 (2020) 3316-3325.

3.2 プロセス・インフォマティクス共通基盤構築

プロセス・インフォマティクスの共通基盤としては、個々のプロセスの要求にこたえるだけでなく、プロセス・インフォマティクスを加速する基盤としての共通化・標準化技術の開発が重要な課題である。その例としては、プロセス・インフォマティクス用機械学習アルゴリズム、自律的最適化手法、シミュレーション・モデリング、ハイスループット実験、AIロボット技術、プロセス計測技術（ソフトセンサー）、データ収集・解析手法、データ蓄積・活用、インセンティブ設計などが挙げられる。

●プロセス・インフォマティクス用機械学習アルゴリズムおよび最適化手法

医薬品またはその中間体の逆合成を中心に、機械学習アルゴリズムや自律的最適化法が検討・開発され、それによって、より効率的に合成ルートの探索ができるようになってきている。しかし、材料合成プロセスにデータ科学を適用するにあたっては、プロセスのパラメータ空間が極めて広大であることや、合成プロセスの多くは多段階プロセスであり前後のプロセスとの関係が複雑であるといったプロセス固有の特徴があるため、有機化合物の逆合成やマテリアルズ・インフォマティクスで開発された機械学習アルゴリズムや自律的最適化手法をそのまま適用するのでは、十分に機能しないことが予想される。そのため、プロセス・インフォマティクス用に適した機械学習アルゴリズムや自律的最適化手法を開発する必要がある。

プロセス・インフォマティクス用に適した機械学習アルゴリズムとしては、例えば、中間データを用いて次元削減をするオートエンコーダーの活用、さらにはその中間データに物理的な意味を有するものを使用することにより効率的に条件探索する手法が挙げられる。また、前段である材料合成プロセスの機械学習モデルと後段である材料加工プロセスの機械学習モデルをそれぞれ用意しておいてそれらを直列につなぐことで、加工プロセスに適する材料の合成プロセスを探索する手法などが考えられる。

また、プロセス・インフォマティクスでは、材料組成や合成条件の探索範囲が格段に広大化するため、試料作成・物性評価・パラメータ最適化を機械のみによって行う、これまでの自律的最適化手法（Closed-loop）を使っても網羅的な探索は困難になってきている。このため、研究者・技術者の経験知（勘・コツ）を活かして最適化を進める「匠・イン・ザ・ループ」（コラム2参照）への取り組みが必要であると考えられる。

●シミュレーション・モデリング

材料合成プロセスのシミュレーション・モデリングは大きく分けて、a) 実験結果に基づく経験パラメータを用いたモデリング、b) 第一原理計算および分子動力学によるシミュレーションがあるが、それぞれ課題がある。経験パラメータによるモデリングa) では、実験結果をベースとしているため、パラメータ構築に用いた実験条件の範囲外の領域での予測をすることは困難である。第一原理計算や分子動力学によるシミュレーションb) では、経験的パラメータを用いないため未知のプロセスでも予測・検討できるが、シミュレーションの前提条件と現実空間での条件に乖離が生じたり、計算資源の制約を強く受けたりする。最近では、それらをつなぐ手法であるマルチフィジックス・シミュレーションの研究が進んでいる。たとえば第一原理計算と統計熱力学との融合により、プロセスの各パラメータの寄与を解析可能なモデルを構築することができ、コンピュータ上でのプロセス状態の再現を可能にするデジタルツインを整備する可能性が検討されている。現状では、個々のプロセスに関する研究を推進している状況であるが、このようなアプローチは、将来はプロセス科学の共通基盤として幅広い活用が期待できる。

●ハイスループット実験、AIロボット技術

プロセス・インフォマティクスに必要な十分な数の斉一なデータを収集するために、ハイスループット実験やAIロボット技術が活発に使われるようになってきている。合成プロセスは多岐に渡るものであるが、それぞ

れのプロセスも単位操作レベルでみると共通となっているものが多い。それら共通的な単位操作ごとのモジュールを共通基盤として用意しておき、それを組み合わせて使うようなアプローチも有効であると考え。

●プロセス計測技術（ソフトセンサー）

ソフトセンサーは、データ科学やシミュレーション技術を活用することで、直接測定が困難なプロセス変数をリアルタイムで容易に測定できる変数から予測する技術であり、現在でも製品品質の安定化およびプラント運転の効率化などに使われている。この技術を活用することで、新規材料の合成プロセスの検討においても、直接測定が困難なプロセスデータをリアルタイムで把握することができるので、プロセスデータ収集の重要な基盤技術として位置づけられる。

●データ収集・解析手法

実験データを効率的に収集する体制の整備が必要不可欠である。IoT（Internet of Things）を利用して、測定装置から測定データをデータベースに自動的に収録するとともに、計測装置より出力されるデータから計測条件や試料情報等のメタ情報を抽出し、機械可読性の高いXMLファイルへと変換するツールが必要となる。すでに物質・材料研究機構（NIMS）においては、M-Dac（NIMS Materials Data Conversion Tools）が開発・公開されているが、さらに高度なツールの開発が望まれる。データ収集においては、それぞれのプロセスにおいてデータ収集・解析、データベース化を個別に進めるだけでなく、将来的に、異なるプロセス間においても共通的にデータを扱い、データ科学を活用したプロセス科学の適用する範囲を拡げるためにも、共通基盤として整備することが欠かせない。また、たとえば、成膜プロセスにおける基板ホルダーなどの研究ハード面の共通化を進めることによって、プロセスデータを集めやすくするような仕組みも必要である。

●データ蓄積・活用

多くのプロセスデータが収集された後に、それを無形資産として適切に蓄積・活用していくことが重要である。膨大で常に増え続けるプロセスデータを安全に蓄積し必要に応じて高速にアクセスできるようにする方法、蓄積されたデータの適切な利用範囲のマネジメントなど、データ蓄積の初期の段階から適切に設計されるべき課題が多く存在する。うまく運用すれば、企業の保有するプロセスデータ自体を非公開にしたうえで、それらの集合から作り出されるプロセス構築基盤を共有資産とすることも可能になるだろう。それをいけば、新たな材料を合成するための新たなプロセスの開発が加速され、日本の基幹産業の国際競争力の向上が期待できる。一方で、プロセスデータは企業などにとっては競争力の源泉になっていることが多く、その部分を秘匿する必要も生ずる。予めオープンにする範囲、クローズにする範囲、限られたメンバーだけで共有するいわゆるシェアデータの範囲などを明確に定義することが求められる。また、蓄積したデータ自体は秘匿したまま機械学習モデルを構築する手法（例えば秘密計算技術）の開発などが求められる。

●データ提供などに対するインセンティブ設計

プロセス・インフォマティクスは、研究者の生み出したデータや経験知（勘・コツ）を活用し、コミュニティ全体の材料開発力を飛躍的に高める試みでもある。それにはこの潮流に参加することのインセンティブ・メリットを明確にすることが重要である。そのためプロセスデータや経験知を提供することに対するインセンティブの設計が必要となる。（コラム3参照）

また、企業の保有する実験データの提供などを推進するには、データ利用権限の取引のような新たな仕組みも必要になると考えられる。

コラム2

匠・イン・ザ・ループ（Takumi-in-the-loop）への挑戦

材料組成や合成条件の広大なパラメータ空間を探索するために、試料作成、物性評価、パラメータの最適化を、人間が介在せずに、機械のみによって行う自律的最適化手法（Closed-loop operation）が広く使われ始めている。研究者が目標とする物性を指定すると、ロボットが材料を自動で合成・評価し、その結果をもとにベイズ最適化により次の実験条件を自動的に設定し、またロボットが自動的に合成・評価を繰り返して、自律的に最適化を行う方法である。この方法の利点は、局所解にならずに大域的に最適化ができることにある。また、ロボットの生産性の高さにより、より多数の実験（ハイスループット実験）を繰り返すことができる利点もある。

しかし、現在の物質・材料創製においては、材料組成や合成条件の探索範囲が格段に拡大しており、自律的最適化手法を使っても、すべてのパラメータを網羅的に探索することは困難である。このため、さらに一歩進んだ形である Human-in-the-loop（人間参加型）機械学習が提唱されている。研究者・技術者の経験知（勘・コツ）を活かして優先的に探索する範囲を絞り込んだり、理論やひらめきを活かして探索範囲を大きく振ることによって、人間が積極的に絶え間なくフィードバックループに関与して効果的な機械学習アルゴリズムを生成する方法である。

Willis（CAS）[1]らは、機械学習による分子構造に、化学者を関与させる効果の検証結果を報告している。専門の化学者による生物活性への影響についての考察を加えることによって、よりよい分子構造を選出することができている。化学者は文献にはあまり報告されていない特徴に注目していることも明らかになっており、研究者に宿る暗黙知の重要性を示している例であるといえる。

そこで、我々は、本戦略プロポーザルにおいて、Human-in-the-loop からさらに一歩進んだ形として、データ科学を活用することで研究者・技術者の能力・知の拡張を目指す、匠・イン・ザ・ループ（Takumi-in-the-loop）というコンセプトを提唱したい。

研究者・技術者が着目している実験の観察内容や測定結果を数値化・データ化して、機械学習の目的変数や説明変数、または中間データとして活用したり、研究者の判断基準を機械学習に予備学習させることによって、より効率的に最適化できるようになることが期待される。データが少ないときでも、熟練者（匠）は「こういう考えが良い」等の（ベストプラクティスとしての）共有可能なデザインパターンを持っており、匠・イン・ザ・ループに活かすことができるはずである。また、機械学習によって得られた相関関係を、研究者が分析し因果関係を解明することで、新たな気づきへの繋げることも期待できる。新たに見つかった相関関係から、因果関係の解明、理論への展開のためには、数理科学、説明可能AI、Symbolic regression（関数同定問題）などの技術の進展も必要になるであろう。AIと人間が協奏することによって研究者・技術者の能力・知を拡張させることに繋がるもの

であり、匠・イン・ザ・ループは是非取り組むべき戦略的アプローチだと考える。

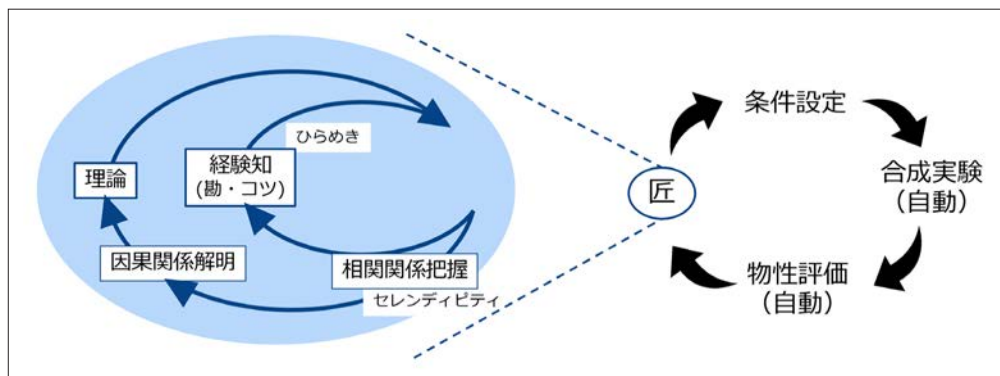


図 3-2-1 匠イン・ザ・ループの概要

《参考文献》

- [1] Todd J. Wills, Dmitrii A. Polshakov, Matthew C. Robinson, and Alpha A. Lee, Journal of Chemical Information and Modeling Vol. 60, pp.4449–4456 (2020) . <https://doi.org/10.1021/acs.jcim.0c00193>
- [2] 一杉太郎,「スパコン・ロボット・ビッグデータを活用した材料研究戦略 ～理論家が自分で材料合成する時代がやってきた～」, HPCI 主催第 11 回材料系ワークショップ「マテリアル DX を加速する「富岳」を活用した成果の創出に向けて」(2021 年 2 月 10 日) 講演資料, https://www.hpci-office.jp/invite2/documents2/ws_material_210210_hitosugi.pdf (2021 年 5 月 24 日アクセス)
- [3] 長藤圭介,「粉体プロセス開発のハイスループット化のためのプロセス・インフォマティクスと材料ハイスループット化との関係」, 科学技術未来戦略ワークショップ報告書「材料創製技術を革新するプロセス・インフォマティクス」, <https://www.jst.go.jp/crds/report/report05/CRDS-FY2020-WR-12.html> (2021 年 5 月 24 日アクセス)

コラム3

データ収集・活用における課題～優れたインセンティブ設計の重要性

昨今、オープンサイエンスの観点から、研究論文と同様に研究データも広く公開していくことが国際的な潮流となっている。オープンサイエンスの振興は当該研究領域の深化はもとより、学際領域研究や産業界における製品開発、ひいてはイノベーション創出に貢献するものと期待されている。特に、データ科学の立場では、十分な量のデータが利活用可能であることが重要である。このためには、オープンサイエンスの流れを踏まえてデータの収集をより積極的に進める必要があるが、データの創出元は各研究者の日々の研究活動にある。したがって、各研究者から、研究成果としてデータ提供を受ける仕組みが必要である。この仕組みの一つに機関リポジトリがあるが、現状では研究者の成果は、主として学術論文の質と量で判断されるので、機関レポジトリ等を利用して、研究データをどれだけコミュニティに提供したかが研究成果として評価されることはあまりなかった。とくに研究当初の予想や期待と異なるデータや、論文の論旨にとって重要ではない多くのデータ（いわゆるネガティブデータ）は公表されずに、研究者本人のみが把握している状態になっていた。データ科学の観点からはネガティブデータは論文等に掲載されるデータ以上に解析価値のあるものであることが多い。

このような状況を変え、各研究者からのデータ提供を加速するには、データ提供のインセンティブを制度設計する必要がある。この取り組みは、世界的にもまだ始まったばかりだが、いくつかの参考になる例も出始めている。

たとえば、オーストラリアでの取り組みでは、Australian National Data Service（ANDS）が、研究者の生み出したデータセットに対して、学術論文と同様に Digital Object Identifier（DOI）を付与している。これにより、関連する研究者が論文の元になるデータセットにアクセス可能になっており、そのデータセットを活用して新たな研究成果を生み出した際には、その論文で元のデータセットの DOI を引用する仕組みとしている。また、現時点で唯一のデータ引用に関するデータベースである Data Citation Index（Clarivate 社）に登録するためのサービスも提供している。これによって研究者の評価には、論文と同様にデータセットの引用も参照され、客観的な研究成果として評価することが可能となり、研究者の研究資金やポスト獲得の際にも考慮される。また、データ等を出版する学術雑誌（データジャーナル）の存在も重要である。そのような学術雑誌として、材料分野では Science and Technology of Advanced Materials（STAM）が、コンピュータプログラムに関しては Computer Physics Communications（CPC）が知られている。このような仕組みは、今後、日本においても検討を進める必要があろう。たとえば、競争的資金の審査において、これまでのデータ提供・公開実績を学術論文と同等に扱うといった方法も考えられる。

一方で、データ提供・公開を研究評価に用いることの弊害にも注意する必要があ

る。学術論文の評価については、ピアレビューによる査読制度や、論文の引用件数、さらには掲載学術雑誌のインパクトファクターなどにより、コミュニティでのコンセンサスがとれているのに対し、データに関しては、基礎物性データや大規模データベースを別とすれば、その質および量に関する評価方法が確立していない。とくにデータの質の評価・担保は物質科学においては重要であり、高品質データの提供には高いインセンティブを与える必要がある。

データが研究者の日々の研究活動から創出されるとすれば、それを広く利活用に供するためのインセンティブとして考慮すべき重要な点は、研究の評価以上に、データ提供・公開によって自分の研究がより進展することであろう。研究費獲得といった研究環境の改善のみならず、データ提供者の研究そのものが直接的により進展するインセンティブ設計が望まれる。たとえば自分のデータを登録すると、関連する他者のデータが登録者に提供される仕組みが用意されれば、登録者の研究に資するであろうし、公開データとして登録すると、さらに多くの関連データが得られれば、積極的に公開登録するだろう。

データ提供者の研究の進展と深化をもたらす優れたインセンティブ制度設計はデータ収集・活用の活性化につながり、ひいては研究者の意識を変え、コミュニティのコンセンサスが醸成されていくと期待できる。

《参考文献》

- [1] 国立情報学研究所オープンサイエンス基盤研究センター,
<https://rcos.nii.ac.jp/openscience/> (2021年5月24日アクセス)
- [2] Australian National Data Service,
<https://www.andis.org.au/> (2021年5月24日アクセス)
[https://www.andis.org.au/_data/assets/pdf_file/0003/383025/
data_citation_poster.pdf](https://www.andis.org.au/_data/assets/pdf_file/0003/383025/data_citation_poster.pdf) (2021年5月24日アクセス)
[https://www.andis.org.au/working-with-data/citation-and-
identifiers/data-citation/citation-metrics](https://www.andis.org.au/working-with-data/citation-and-identifiers/data-citation/citation-metrics) (2021年5月24日アクセス)

3.3 プロセス科学基盤拡充

材料合成プロセスの素過程の解明はプロセスの最適化や新プロセス設計、ひいてはプロセス科学の学理構築の上で重要である。たとえば薄膜材料を得る代表的な合成プロセスとしてMBE法がある。MBE法では材料ガスの選択、流量、濃度、基板の種類、シャッター時間、真空度、クヌーセンセルの加熱温度等々、素過程に関係すると思われる多くのプロセスパラメータがある。しかしながら、素過程の記述・制御に必要な十分なパラメータは実はよくわかっていないのが現状である。そのため多くの場合、これまでの経験に基づいてパラメータを選択し調整しながら、素過程を推定してきたのが現状である。ここにこれまでのプロセス科学の限界があった。

この課題を克服するためには、これまでの手法に加えて、当該プロセスに関するデータを可能な限り大量に収集しそれらを記述子として扱うことでデータ科学を活用していくことが重要である。（記述子とは機械学習モデルの説明変数として用いるために数値として取り出される情報を示す。）通常使われているプロセスパラメータを一次記述子として、プロセス特性を適切に記述する二次記述子（プロセス記述子）設計を行うことはプロセス・インフォマティクスの課題の一つである。その結果得られる記述子を、物性科学の知識や熟達者の経験に照らし合わせることで合成プロセスにおける重要な素過程を推定することができるとされる。これはプロセス記述子を活用したデータ駆動型の素過程設計手法ということができる。

プロセス記述子によって素過程が記述・表現されると、装置固有であるため定量的な記述が困難でかつ各論的な取り扱いしかできないと思われていたプロセスを計量化することができる。

このようなプロセス記述子が各プロセス、たとえば薄膜成長でいえば、MBE法、スパッタ法、あるいはCVD法等で構築されると、それぞれの素過程のプロセス記述子間の関係を論じることができ、プロセスをいくつかのカテゴリーに分類することができるようになる。従来の物理化学的な知識による分類わけに加えて、プロセス記述子によってデータ科学的にクラス分けすることが考えられ、たとえば、物質そのものではなく、合成プロセス中の平衡状態からの外れ具合、といった軸でクラス分けできる可能性がある。

そのようなプロセス記述子のアナロジーを、すでに知られている事例から探してみると、結晶成長の様子の説明に使われる過飽和度がそれにあたる。大気上層部の低温領域で起こる気相からの氷（雪）結晶の成長であろうと、1500℃以上の高温融液から引き上げるシリコンの結晶成長であろうと、固相/気相、固相/液相の物質の化学ポテンシャル（単位量当たりのギブス自由エネルギー）の差から定義される過飽和度というパラメータで整理することで、結晶成長の様子を決める重要な因子である、核生成密度、結晶成長速度、ファセット成長などに関して、統一的理解をすることができる。つまり、これらの結晶成長は温度、圧力、物質密度などの点では、非常にかけ離れているに関わらず、それらを一次記述子として出発することで定められる二次記述子＝過飽和度を介することで、統一的に理解できたことになる。この例の二次記述子としての過飽和度は、確立した熱力学の体系を通してすでによく知られている量であり、それを通して現象の解釈を容易にしている。本研究開発課題で目指しているものは、データ科学を高度に駆使することで、数多くのプロセス一次記述子から異なるプロセスの間に共通に利用できる二次記述子、人類のまだ知らないプロセスの支配パラメータを見いだすことにある。

このようなアプローチにより、個別プロセスの解析ではわからなかった重要な共通因子を推定できる可能性がある。薄膜成長実験の熟達者はこのような共通因子を経験知として体得しているがゆえに、ある場合にはMBE法を採用し、別の場合にはスパッタ法を選択していると思われるが、その暗黙知を形式知化するのみならず、新たな視点や気づきを与える可能性がある。

各プロセスのプロセス記述子は特性（合成容易性、高品質性等）とも関連付けられているので、プロセス間のプロセス記述子のデータ科学的解析、および反応素過程に基づく物理化学的解析によって、必要特性を実現する合成プロセスの設計・推定が可能になる。

新たに得られたプロセス記述子の解釈は、そのままでは人間による解釈に適さない（しかし、プロセス設計には使える）ものであるかもしれない。しかし、それを解釈していこうとする試みによって、新たな学理が構築され、各論的・局所的であった合成プロセスの議論を刷新し、プロセス科学の統合的な基盤拡充につながる可能性もある。

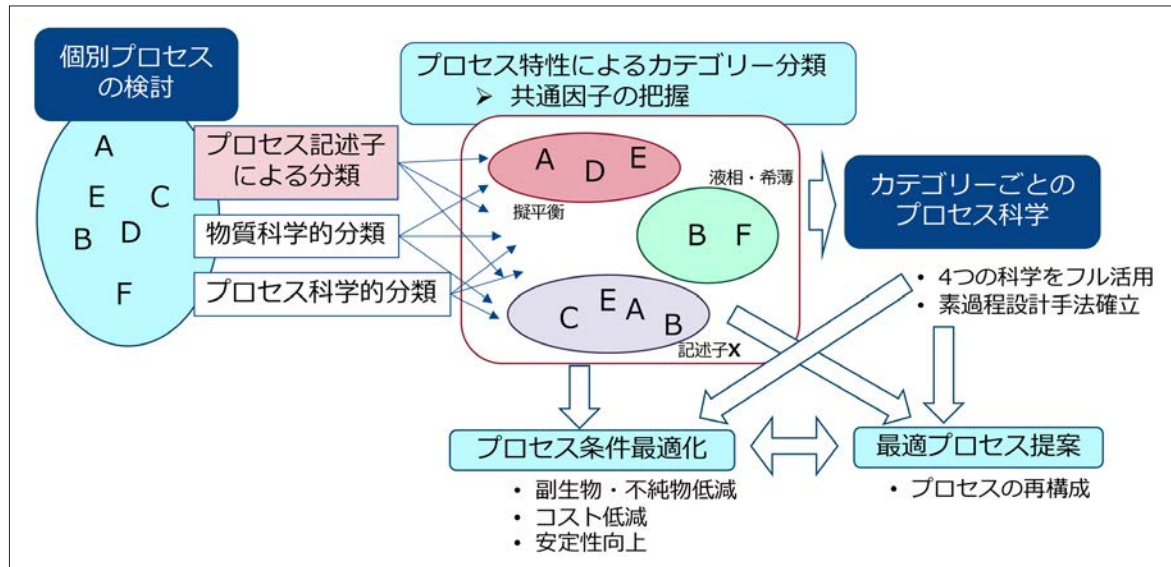


図3-3-1 プロセス科学基盤拡充

4 | 研究開発の推進方法および時間軸

4.1 研究開発の推進方法

本戦略プロポーザルで述べる研究開発課題は、従来とは異なる研究開発体制・仕組み、構成メンバーで取り組むことが必要になると考える。本章では、①各材料領域におけるプロセス・インフォマティクス手法の構築、②プロセス・インフォマティクス共通基盤、③プロセス科学基盤拡充、それぞれに適した推進方法を述べ、また全体に共通する推進方法について述べる。

① 各材料領域におけるプロセス・インフォマティクス手法構築の推進方法

各材料領域において、プロセス・インフォマティクスの中核となるプロセスを選び、そこから各々の材料領域において横展開していくことが望ましい。その際、プロセス・インフォマティクスの必要性が高く、実施することによって効果が見込めるプロセスから取り組むのがよいと考える。このようなプロセスは、すでにプロセス・インフォマティクスの活用を検討が始まっているものもあるが、それらは個別プロセスの最適化を目指した研究であることが多い。先行するプロセスの研究事例から得られる知見をそれぞれの材料領域に上手に活用することで、当初は難しく思えたプロセスにも適用を拡げていくことが期待できる。そのためにも、先行するプロセスで成功事例を作ることと、材料領域に拡げるための知見を得ることの両方が重要である。

プロセスそれぞれにおいて、十分な数のデータを生成・収集・蓄積し、データベース化することが必要である。データとしては、実験データのみならず、シミュレーション、モデリング、論文・特許などの文献データも重要である。しかし、論文・特許に掲載されているプロセスデータは基本的に抜粋されたものであり、プロセス条件をすべて把握できることはまずない。特に、プロセスは多くのパラメータの影響を受けるため、再現するために必要な情報をすべて含むトレーサビリティの確保されたデータが中心的役割を担う。したがって、実験、シミュレーション、モデリングによるデータを積極的に生成・蓄積することが重要である。

また、予測や期待に反した結果であるネガティブデータもプロセスの科学的理解やデータ科学への適用のためにも非常に重要であり、この意味でも、文献・特許データに頼るだけでなく、実験、シミュレーション、モデリングによるデータ生成は重要である。

実験により、トレーサビリティの確保されたデータを大量に生成し蓄積するためには、実験 DX の活用が有効である。特に、最近進展がめざましいハイスループット実験やロボット技術を積極的に活用するべきであると考えられる。また、プロセス内の状態を正確に把握するための、プロセス計測技術を開発することも必要になると考えられる。さらに、設定条件や計測データを各装置から効率よく収集しデータベースに蓄積することも重要である。

これらを実現するためには、合成プロセス装置・計測装置などの整備が必要である。したがって、個別対象プロセスについては、プロセスごとに物理的拠点であるプロセスセンターを設置し、そこに合成プロセス装置、評価・計測装置を整備することが、効率的であり、実効があがる方法であると考えられる。

プロセスセンターには、研究者だけでなく、プロセス技術者、計測技術者、技術支援員、データ管理者なども配置するべきである。

一方、シミュレーション、モデリング、データ科学については、一つのプロセスに一時的に集中的に取り組むことが必要になることもあるが、他プロセスとの共通性を見つけだしたり、より汎用的な手法へと発展させたりする必要もあるため、必ずしも、物理的にプロセスセンターに集約するのではなく、別々な場所に

いたまま連携することにより推進する方法も可能である。

また、各プロセスに対するインフォマティクスの適用では、その成果を具体的な問題解決に活かし社会実装に繋げることが非常に重要である。このため、産業界からも、具体的な検討テーマを持って、プロセスセンターに参加することが強く望まれる。

さらに、前述のように、トレーサビリティの確保された十分な量のデータを確保することが重要であるので、プロセスセンターでの研究開発によって生成されたデータは、企業の参加した研究であっても、プロセスセンターが利用権限を持つことを原則とすることが望ましい。

② プロセス・インフォマティクス共通基盤構築の推進方法

プロセス・インフォマティクス共通基盤では、プロセス・インフォマティクス用に最適化された機械学習アルゴリズムの開発や、プロセス・シミュレーションとモデリング、ハイスループット実験やロボットなどの実験DX、プロセス計測技術、データベース構築などに取り組む。このためには、理論科学、計算科学、データ科学、ロボット技術、プロセス技術、計測技術など非常に多岐に渡る専門家の参画が必要である。

デジタル・トランスフォーメーションが進んでいる状況では、プロセス・インフォマティクス共通基盤は、物理的な拠点に集結することで推進する必要は必ずしもなく、別々の場所で連携活動を実施するバーチャル拠点であっても十分に機能させることができると考える。しかし、ここでの拠点は、プロセス・インフォマティクスを強化・拡充するために必要な機能であるので、プロジェクトごとの一時的な連携でなく、強力な連携を継続する拠点として設置する必要がある。

また、プロセス・インフォマティクスの確立・普及には、研究者の意識変革が重要である。個別プロセスのインフォマティクス活用研究の事例をつかって、学会やコンソーシアム、ワークショップなどで啓蒙し、プロセス・インフォマティクスに取り組もうとする研究者の裾野を拡げていくことも、共通基盤にとっても重要である。このような活動をするによって、共通基盤として真に必要な情報が集まってくるようになる。

③ プロセス科学基盤拡充の推進方法

各プロセスにおいて、当該プロセスに関するデータを可能な限り大量に収集しデータ科学に活用していくのは、各プロセスセンターを中心に行っていくのが適当である。その上で、データ科学を用いてプロセス特性を適切に表す記述子を設計し、異なるプロセス間の共通因子を把握することでカテゴリーに分類をし、カテゴリーごとの共通因子の科学的理解を進め、素過程設計手法を確立していく。そのためにはプロセスインフォ共通基盤の拠点と各プロセスセンターが強く連携しながら進める必要がある。

全体として考慮すべき事項

①各材料領域におけるプロセス・インフォマティクス手法の構築、②プロセス・インフォマティクス共通基盤、③プロセス科学基盤拡充、の各研究開発課題は相互に関係が強く、それぞれの成果を活かして相乗的に進展していくものであるため、全体を束ねる仕組み（ガバナリングボードなど）や学会・シンポジウムの活用などが必要である。

プロセス・インフォマティクスを推進するためには、人材育成がきわめて重要である。材料研究者が機械学習を理解して材料研究者の視点でテーマ設定をする、あるいは、データ科学の研究者が材料の基礎知識・ノウハウを理解するなど、双方を融合させることが大切になる。材料研究者とデータ科学研究者の単なる共同研究だけではなく、両方の視点を持った若手人材の育成が今まで以上に重要になってくる。

また、産業界が参画する仕組みを整えることも重要である。本戦略プロポーザルでは、材料創製技術を中心に述べているが、材料は実際にスケールアップし量産できてこそ世の中で使われていく。そのためには、産学連携を従来以上に進めることが必要である。プロセス・インフォマティクス推進を契機に、アカデミアと産業界がもっとシームレスかつ一体的になることも期待される。産業界とアカデミアが共同でベンチャーを設立

して新プロセスの実用化を進めるなどの方法も考えられる。

産業界の参画や、複数グループの共同研究を促進するためにも、データ取扱ルールを早期に明確にすることが重要である。インフォマティクスに共通のことであるが、データ科学を有効に機能させるために、できるだけデータをオープンに利用できる環境にすることが望ましい。類似の実験や、シミュレーション、モデリングを繰り返す無駄を減らし、一定の投入資源でより多くの成果が得られるためでもある。一方で、プロセスデータは企業などにとっては競争力の源泉になっていることが多く、その部分を秘匿する必要も生じてくる。予めオープンにする範囲、クローズにする範囲、限られたメンバーだけが利用可能とするいわゆるシェアデータの範囲などを明確に定義し、適切に管理・運用することが求められる。また、蓄積したデータ自体は秘匿したまま機械学習モデルを構築する手法の開発、活用など技術流出への対策も必要となる。学習データは非開示のまま、学習済みの機械学習モデルを開示し、幅広く活用していく方法も有効と考える。

プロセス・インフォマティクスによって獲得される合成プロセス開発における競争力を維持し続けることも重要である。各プロセスセンターは、共通基盤拠点と連携をしながら、具体的なテーマに継続的に取り組み、プロセス・インフォマティクス関連の技術の高度化を継続するとともに、データやノウハウを蓄積させることによって、当該合成プロセス開発における中核拠点として優位性を維持することが必要である。このため、合成プロセスに関する具体的な開発ニーズを集める仕組みや、共同研究・委託研究の費用や、データ・機械学習モデルなど利用対価の設定などによって、プロセスセンターは、最終的には、できるだけ自立的な運営ができることを目指すことが望ましい。

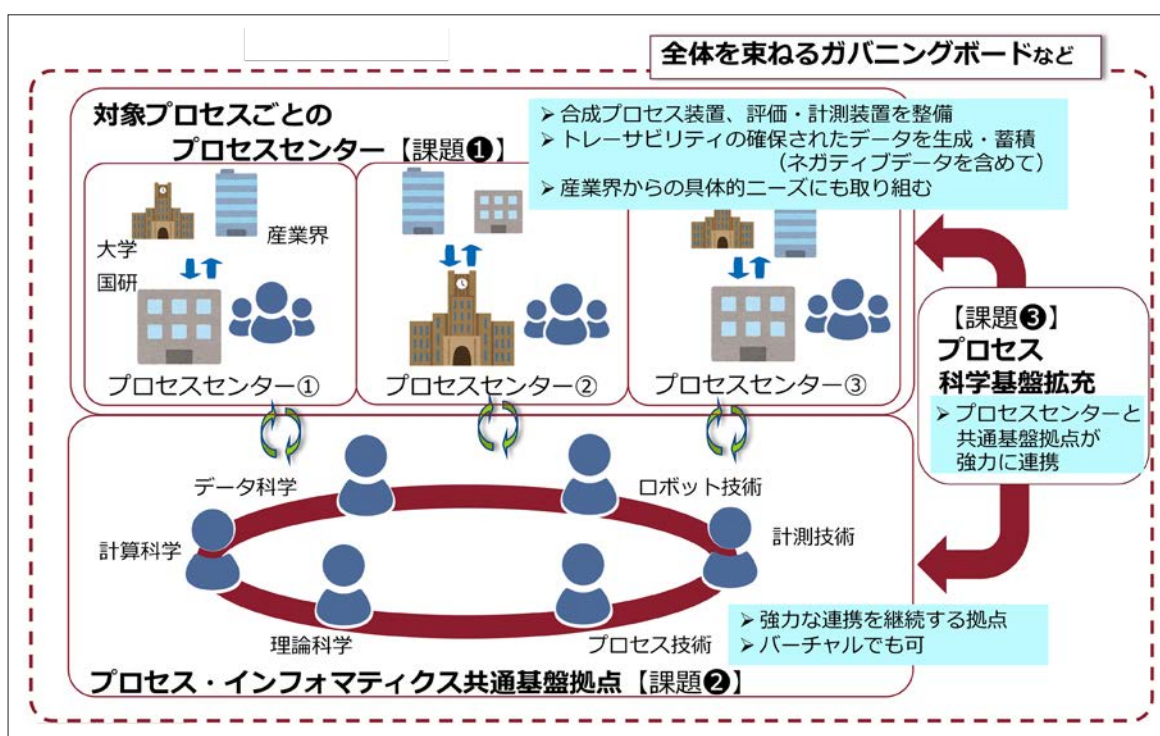


図4-1 研究開発の推進方法

4.2 時間軸

本提言では、今後10－15年程度の期間、3章に示した3つの研究開発課題について継続的な研究開発を行い、プロセス構築基盤を確立していくことを目指す。それぞれの研究開発課題の時間軸を図4-2に示す。

①各材料領域におけるプロセス・インフォマティクス手法の構築については、すでに萌芽的な研究がみられていることもあり、そうした研究を核にしながら、強化・拠点化を速やかに開始していくことが望まれる。この課題については、早い段階から産業応用の検討が開始できるような成果が生み出されてくる可能性もあるため、そうした場合に備え、産業化を目的とするプログラム等の受け皿も考慮しておく必要がある。例えば、プロジェクトの期間途中であっても、産業界などと合同で実用化検討が行えるような柔軟な体制にしておくなどが考えられる。また、この課題で得られる知見やプロセスデータを蓄積し、プロジェクト間で利用できるようにし、②プロセス・インフォマティクス共通基盤構築のスタートと同時にそれらが利用できるようにする。②は、①からのプロセス情報や知見のインプットを受けながら、新たなデータ科学的手法構築を目指すとともに、その成果物を①に返すことを繰り返して協調しながら並行して推進していく。上述の産業化を目的とするプログラムがスタートする際には②の担当機関の一部もそのプロジェクトに合流できるようにしておく。

③プロセス科学基盤拡充に関しては、科学・技術的難易度が高く、研究初期から多くの成果が出続けるものではないと思われる。また、②に比べると具体的なプロセスデータを大量に使うというより、その手法についての基礎的な検討から開始することになるため、①、②と独立に同時期に開始しても良いし、あるいは、①、②である程度の知見や方向性が見え始めた時期（1,2年後）からスタートし、より長期的な視野で基盤構築を目指してもよい。10年スケールで、いくつかのカテゴリーについて、重要な記述子を発見し、ある程度の実用性を持ってプロセス間をまたがる最適化が可能になっていることを目指す。

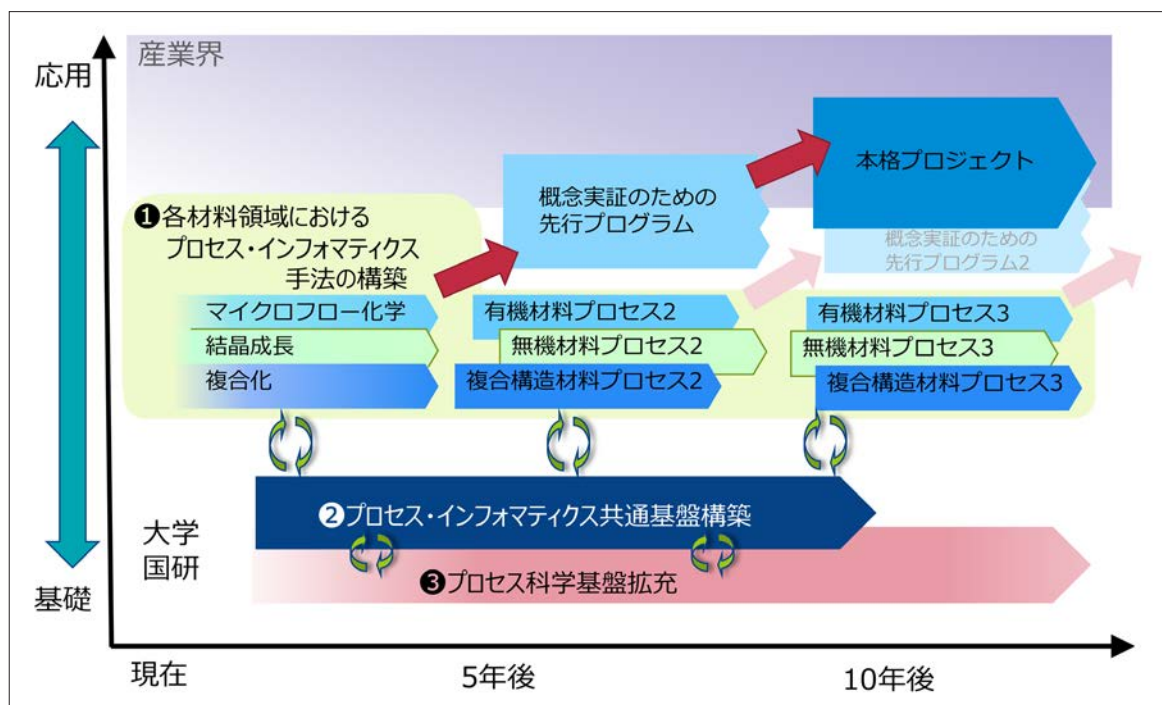


図4-2 研究開発の時間軸

付録1 検討の経緯

- ・JST 研究開発戦略センター（CRDS）では、令和元年度に戦略プロポーザルを作成すべきテーマの候補をCRDS戦略スコープ2020検討委員会を経て指定し、令和2年5月に検討チームを発足させた。その後、検討チームにおいて提言作成へ向けた調査・分析・検討を重ねた。
- ・チームの活動では、調査によって国内外の研究開発動向・技術水準を明らかにしながらスコープの焦点を絞り、その過程において提言の方向性を検討するため、以下の有識者へのインタビュー・意見交換を実施した。
- ・その上で、プロセス・インフォマティクスの研究開発に関してCRDSが構築した仮説を検証する目的で、科学技術未来戦略ワークショップを開催した（詳細次ページ）。ワークショップの結果は報告書として令和3年3月にCRDSより発行している（CRDS-FY2020-WR-12）。
- ・CRDSでは以上の調査・分析の結果と、ワークショップにおける議論等を踏まえて、令和3年6月に本戦略プロポーザルを発行するに至った。

意見交換・インタビューを実施した識者（敬称略、所属・役職は実施時点）

氏名	所属・役職
青柳 岳司	産業技術総合研究所 機能材料コンピューテーショナルデザイン研究センター 総括研究主幹
池端 久貴	旭化成株式会社 研究・開発本部 インフォマティクス推進センター リードエキスパート
石田 真彦	日本電気株式会社 NEC-産総研量子活用テクノロジー連携研究ラボ 主幹研究員
伊丹 健一郎	名古屋大学 大学院理学研究科 教授
宇治原 徹	名古屋大学 未来材料・システム研究所 教授
岡部 朋永	東北大学 大学院工学研究科 教授
金子 弘昌	明治大学 理工学部 准教授
寒川 義裕	九州大学 応用力学研究所 教授
菊川 豪太	東北大学 大学院工学研究科 准教授
河野 禎一郎	旭化成株式会社 研究・開発本部 インフォマティクス推進センター センター長
甲村 長利	産業技術総合研究所 触媒化学融合研究センター 研究チーム長
清水 亮太	東京工業大学 物質理工学院 准教授
袖山 慶太郎	物質・材料研究機構 統合型材料開発・情報基盤部門 グループリーダー
高橋 啓介	北海道大学 理学研究院化学部門 准教授
瀧川 一学	理化学研究所 革新知能統合研究センター 研究員/北海道大学 化学反応創成研究拠点 主任研究員
谷池 俊明	北陸先端科学大学院大学 先端科学技術研究科 教授
富谷 茂隆	ソニー株式会社 R&D センター 主幹研究員
内藤 昌信	物質・材料研究機構 統合型材料開発・情報基盤部門 グループリーダー
永木 愛一郎	京都大学 大学院工学系研究科 准教授
長藤 圭介	東京大学 大学院工学系研究科 准教授

中山 将伸	名古屋工業大学 生命・応用化学専攻 教授
畑中 美穂	慶應義塾大学 理工学部 准教授
一杉 太郎	東京工業大学 物質理工学院 教授
船津 公人	東京大学 大学院工学系研究科 教授
宝野 和博	物質・材料研究機構 磁性・スピントロニクス材料研究拠点 理事・拠点長
前田 理	北海道大学 化学反応創成研究拠点 教授・拠点長
町田 友樹	東京大学 生産技術研究所 教授
右田 啓哉	株式会社日本触媒 データサイエンス&インフォマティクス推進室 室長
水野 洋	パナソニック株式会社 マテリアル応用技術センター デジタル融合総括
村岡 恒輝	カリフォルニア大学バークレー校 日本学術振興会海外特別研究員
山内 悠輔	クイーンズランド大学 化学工学科 教授
吉川 英樹	物質・材料研究機構 材料データプラットフォームセンター 副センター長
吉田 亮	情報・システム研究機構 統計数理研究所 教授
柳 日馨	大阪府立大学 特認教授

科学技術未来戦略ワークショップ

「材料創製技術を革新するプロセス・インフォマティクスの研究開発」

(敬称略)

開催日時：2021年1月24日（日）9：30～17：00

開催会場：リアル（TKP市ヶ谷カンファレンスセンター 8A）とオンライン（Zoom）のハイブリッド開催

9：30～9：35 開会挨拶 曽根 純一（JST-CRDS）

9：35～9：50 ワークショップの趣旨説明 福井 弘行（JST-CRDS）

9：50～12：10 （話題提供①個別プロセスにおけるインフォマティクス活用研究）

9：50～10：15 「結晶成長デジタルツインに向けた取り組み（SiC 溶液成長を中心に）」

宇治原 徹（名古屋大学）

10：15～10：40 「ロボット-機械学習による全自動・自律的な無機機能性薄膜合成の現状と展望」

清水 亮太（東京工業大学）

10：55～11：20 「精密有機合成とプロセス・インフォマティクス」

甲村 長利（産業技術総合研究所）

11：20～11：45 「粉体プロセス開発のハイスループット化のためのプロセス・インフォマティクスと材料探索ハイスループット化との関係」

長藤 圭介（東京大学）

11：45～12：10 「データ駆動型高分子材料開発とプロセス・インフォマティクス」

内藤 昌信（物質・材料研究機構）

13：15～14：55 （話題提供②プロセス・インフォマティクスに必要な技術群）

13：15～13：40 「プロセス・インフォマティクスにおける技術紹介および課題と将来展望」

金子 弘昌（明治大学）

13:40～14:05 「窒化物半導体MOVPEプロセスのマルチフィジックス結晶成長シミュレーション」

寒川 義裕 (九州大学)

14:05～14:30 「触媒インフォマティクスの実践に見るプロセスの重要性」

谷池 俊明 (北陸先端科学大学院大学)

14:30～14:55 「計測インフォマティクスを目的とした計測スペクトルのスパースモデリングの自動化」

吉川 英樹 (物質・材料研究機構)

15:15～16:55 総合討論 ファシリテーター
コメンテーター

伊藤 聡 (JST-CRDS)

水野 洋 (パナソニック)

富谷 茂隆 (ソニー)

右田 啓哉 (日本触媒)

畑中 美穂 (慶應義塾大学)

柳 日馨 (大阪府立大)

- 【論点】 1.「PI 共通基盤構築研究」の拠点で取り組むべき研究開発課題。研究開発体制、制度面などで検討すべき事項、参画が望まれる研究者・技術者の専門分野。
- 2.「個別対象プロセス」として取り上げるべきプロセス。プロセスセンターに用意すべき人材・技術・装置、制度・仕組み。参画が望まれる研究者・技術者の専門分野。
3. アカデミアと産業界の連携の仕組み
4. その他 (コミュニティ形成、人材育成、海外連携)

16:55～17:00 閉会挨拶

曾根 純一 (JST-CRDS)

招聘識者（敬称略、所属・役職は開催時のもの）

（発表者）

- ・宇治原 徹 名古屋大学 未来材料・システム研究所 未来エレクトロニクス集積研究センター 教授
- ・金子 弘昌 明治大学 理工学部 准教授
- ・寒川 義裕 九州大学 応用力学研究所 教授
- ・甲村 長利 産業技術総合研究所 触媒化学融合研究センター フロー化学チーム チーム長
- ・清水 亮太 東京工業大学 物質理工学院 准教授
- ・谷池 俊明 北陸先端科学技術大学院大学 先端科学技術研究科 教授
- ・内藤 昌信 物質・材料研究機構 データ駆動高分子設計グループ グループリーダー
- ・長藤 圭介 東京大学 大学院工学系研究科 機械工学専攻 准教授
- ・吉川 英樹 物質・材料研究機構 材料データプラットフォームセンター 副センター長

（コメンテーター）

- ・富谷 茂隆 ソニー株式会社 R&Dセンター TL27 (Materials Analysis) 主幹研究員
- ・畑中 美穂 慶應義塾大学 理工学部化学科 准教授
- ・右田 啓哉 株式会社日本触媒 データサイエンス&インフォマティクス推進室 室長
- ・水野 洋 パナソニック株式会社 マテリアル応用技術センター デジタル融合総括
- ・柳 日馨 大阪府立大学 研究推進機構 特認教授

付録2 国内外の状況

ここでは、プロセス・インフォマティクスを含むデータ駆動型の材料開発に関連する国内外の政策・施策動向、およびプロセス・インフォマティクスの研究事例等について記載する。

1. 政策・施策動向

計算科学に基づいたマテリアルズ・インフォマティクスの概念は、2000年頃から提唱されていたが、当時は計算機で取り扱える原子数や精度が不十分であり、また大量データを統合し材料科学の問題をマイニングする手法也未開拓であった。しかしながら、近年の計算機能力の飛躍的向上と大量データを取り扱うデータ科学の進化により状況が劇的に変わった。2011年に米国が「Materials Genome Initiative: MGI」を発表したことや2012年頃の深層学習の成果による第三次AIブームをきっかけに、計算科学とデータ科学を融合させたマテリアルズ・インフォマティクスが注目され、世界中で精力的な研究がなされている。かつて、理論、実験に次ぐ第3の科学として計算機科学が物質・材料研究における重要な位置を占めるようになったことと同様に、最近では「第4の科学」としてデータ科学の重要性が指摘されるようになった。

〔日本〕

2013年8月にJST研究開発戦略センターが戦略プロポーザル「データ科学との連携・融合による新世代物質・材料設計研究の促進（マテリアルズ・インフォマティクス）～物質・材料研究を飛躍的に発展させるための新たなパラダイム～」を発行し、マテリアルズ・インフォマティクスの重要性について政府に提言を行った。翌2014年3月には、産業競争力懇談会（COCON）から報告書「シミュレーション応用による新材料設計手法」が発行されている。ここでは、ものづくり現場における材料シミュレーション活用のあるべき姿の実現に向けた第一歩として、材料計算科学と情報科学の融合によるハイスループット材料設計システムの開発に取り組む必要がある、としている。

さらに2015年からJSTにおける「国立研究開発法人を中核としたイノベーションハブの構築支援事業」が開始され、物質・材料研究機構（NIMS）を中核機関とする「情報統合型物質・材料開発イニシアティブ（"Materials Research by Information Integration" Initiative: MI²I）」（PL：寺倉清之（2015～2016年度）、伊藤聡（2017～2019年度））が発足した。MI²Iでは、蓄電池材料、磁性材料、伝熱制御・熱電材料を具体的テーマとした研究開発に取り組むとともに、データ駆動型研究手法の開発にも取り組んでいる。また、物質・材料科学から情報・数理科学にわたる産学官の協働体制を構築し、コンソーシアムを設立することで広範な企業の参画を促し、オープンイノベーションに繋がるハブ拠点化を目指している。

それ以前の2014年に開始した内閣府戦略的イノベーション創造プログラム（Cross-ministerial Strategic Innovation Promotion Program: SIP）「革新的構造材料」（PD：岸輝雄（2014～2019年度））の中の研究開発項目の一つ「マテリアルズインテグレーション」（領域長：小関敏彦）では、材料科学の成果を活用するとともに、理論、実験、解析、シミュレーション、データベースなどの全ての科学技術を融合して材料の研究開発を工学的な視点に立ち支援することを目指す統合的な材料技術ツールを開発している。

また、2016年より開始した経済産業省/NEDO「超先端材料超高速開発基盤技術プロジェクト（超超プロジェクト）」（PL：村山宣光（2016～2021年度））では、有機系機能材料を対象にマルチスケールシミュレーション等の計算科学を活用し、材料の「構造」と「機能」を結びつけたデータ群を作り出し、マテリアルズ・インフォマティクスと融合することで革新的な機能性材料の創製・開発の加速化を目指している。

上述の内閣府のSIP「革新的構造材料」、文部科学省/JSTのMI²I、経済産業省/NEDOの超超プロジェクトは、第5期科学技術基本計画（2016年1月22日閣議決定）におけるSociety 5.0の実現に貢献する11

のシステムの1つ「統合型材料開発システム」に関する3府省連携施策として位置づけられ、これら3府省のプロジェクトが補完的に研究開発を実施していく体制が、総合科学技術イノベーション会議ナノテクノロジー・材料基盤技術分科会を通じて構築された。さらに、2018年度からは内閣府においてSIP第2期「統合型材料開発システムによるマテリアル革命」(PD:三島良直(2018～2022年度))が開始され、マテリアルズインテグレーション(MI)の素地を活かし、欲しい性能から材料・プロセスをデザインする「逆問題」に対応した次世代MIシステムを世界に先駆けて開発するとともに、MIを活用して競争力ある革新的な高信頼性材料の開発や設計・製造・評価技術の確立に取り組み、発電プラント用材料や生体用材料、航空用材料等を出口として、先端的な構造材料・プロセスの事業化を目指している。

一方、JST戦略的創造研究推進事業においては、2015年にさきがけ研究領域「理論・実験・計算科学とデータ科学が連携・融合した先進的マテリアルズインフォマティクスのための基盤技術の構築(略称:マテリアルズインフォ)」(研究総括:常行真司(2015～2020年度))、2016年よりCREST・さきがけ複合領域「計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用(情報計測)」(研究総括:雨宮慶幸、副研究総括:北川源四郎(CREST:2016～2023年度、さきがけ:2016～2021年度))、2017年よりCREST研究領域「実験と理論・計算・データ科学を融合した材料開発の革新(革新材料開発)」(研究総括:細野秀雄(2017～2024年度))が相次いで発足した。さきがけ「マテリアルズインフォ」では、実験科学、理論科学、計算科学、データ科学の連携・融合による先進的マテリアルズインフォマティクスの確立およびそれを牽引する次世代研究リーダーの輩出を目的としている。CREST・さきがけ「情報計測」においては、多様な計測・解析技術に最先端の情報科学・統計数理の研究の融合によって、これまで捉えられなかった物理量・物質状態やその変化あるいは潜在要因等の検出、これまでは困難であった実動作下のリアルタイム計測等を実現するインテリジェント計測・解析手法の開発とその応用を目指している。CREST「革新材料開発」では、高い社会ニーズがあるものの未だ達成されていない材料や機能をターゲットにし、物質・材料開発の基本である実験科学と、理論・計算・データ科学を融合させることで革新的な材料開発手法を提示し、わが国の産業競争力の向上に貢献することを目的としている。また、2021年に発足したCREST「未踏探索空間における革新的物質の開発(略称:未踏物質探索)」(研究総括:北川宏(2021～2028年度))では、計算科学/データ科学/高スループット評価/非平衡プロセス/プロセス・インフォマティクスに直結させたその場計測などを含む材料創製手法を開発することで、元素の潜在能力を最大限に引き出すと共に、元素の組み合わせがもたらす膨大な物質探索空間を開拓することにより、複数元素のシナジー効果による革新的な機能・物性を有する新材料を創出することを目的としている。

2020年6月に公表された「マテリアル革新力強化のための政府戦略に向けて(戦略準備会合取りまとめ)」では、新型コロナウイルス感染症の世界的流行に伴い、データやAI、ロボットを活用した新たな研究開発手法や研究開発現場の本格導入の必要性の高まり、マテリアルの研究開発現場や製造現場全体のデジタルトランスフォーメーション(DX)が急務であることを受けて、マテリアル研究開発の川上から川下までのデータが持続的・効果的に創出、共用化、蓄積、流通、利活用される「マテリアルDXプラットフォーム」の必要性についても言及がなされた。

文部科学省では、これらを受けてマテリアルDXプラットフォーム構想実現に向けた取り組みを開始し、「データ中核拠点の形成」「データ創出基盤の整備・高度化」「データ創出・活用型マテリアル研究開発プロジェクト」の3つを主軸とする事業を計画している。そのうちの1つであるマテリアル先端リサーチインフラについては2020年12月に公募を開始し2021年3月に採択機関が公表された。同事業は2021年度から10年間の計画で、大学・国立研究開発法人等において、広範に充実した最先端設備群及び技術・ノウハウを有するハブ機関と、一定の領域で特徴的な設備・技術を有するスポーク機関からなるハブ&スポークによる全国体制の構築を行う。また、各機関が保有する設備・技術・ノウハウ・データを共用することで、わが国におけるマテリアル研究開発を先導し支える研究インフラ・プラットフォームを目指す。今後、マテリアル先端リサーチインフラの利用を通じて創出されるデータを、物質・材料研究機構が担うデータ中核拠点と接続させることが計

画されている。

さらに材料創製・計測・理論計算に加えてデータサイエンスが有機的に連携することでマテリアル革新力を強化することをめざした「データ創出・活用型マテリアル研究開発プロジェクト」の検討が進んでいる。事業期間は同じく2021年度からの10年間の計画であり、開始の2021年度をFS期間として、2022年度に本格移行する予定である。FSでは、「マテリアル革新力強化のための政府戦略に向けて（戦略準備会合取りまとめ）」において特定された8つの重要技術領域（エネルギー変換マテリアル、マテリアルの高度循環技術、高度デバイス機能マテリアル、量子・電子マテリアル、極限機能マテリアル、バイオ・高分子マテリアル、ナノスケールマテリアル、マルチマテリアル化技術）において、特に社会ニーズが高く、革新的な成果が期待され、DXによるインパクトの高い材料課題を特定する。そのうえで、材料・設備コミュニティの連携体制を確立するとともに、各材料課題における最適なDXの方法論を具現化することとしている。

一方、経済産業省においては、プロセス・イノベーション・プラットフォームの構築を計画しており、2021年度を目途に重点領域（日本の国際競争力の高いマテリアル）を同定し、産業技術総合研究所地域センター（つくば、中部、中国）に高機能材料（触媒、セラミックス、セルロースナノファイバー等）の製造プロセスデータを一通貫、ハイスループットで収集できる設備環境（プロセス・イノベーション・プラットフォーム）を整備、運用を開始するとしている。また、NEDOにおいては2021年6月から「マテリアル革新技術先導研究プログラム」として「データを活用した革新的マテリアル製造プロセス・インフォマティクス技術の開発」が開始予定である。

〔米国〕

2011年より国家的な取り組みとして開始されたMGIは、材料開発に要する期間を2分の1に短縮するという目標を掲げ、「計算ツール」「実験ツール」「デジタル・データ」を材料イノベーション基盤として整備するべきであるとするものであった。このプロジェクトでは、以下に上げる3つの方向で研究が推進された。

1つ目は、第一原理計算を主体として物質の構造や物性に関するデータを作り出し、所望の性能や機能を持つ物質・材料を見つけ出そうとする流れであり、実験と計算科学に基づくデータベースと、データ科学に基づく各種分析ツールを融合した物質の網羅的スクリーニングを行うものである。その例として、マサチューセッツ工科大学（MIT）のグループ（現在はUC Berkeley）とローレンス・バークレー国立研究所による“Materials Project”がある。ここでは、無機結晶構造データベースに収録された結晶構造データをもとに、網羅的に第一原理計算を行った結果を、熱力学データベースをリンクさせて、状態図や構造予測、化学反応予測などを収録したデータベースを公開している。さらに、これらのデータを活用できる各種アプリケーションソフトウェアも公開しており、リチウムイオン電池の正極材料の探索などに用いられている。同様のデータベースの構築と公開は、デューク大学、ノースウエスタン大学をはじめとする、大学・国研で行われている。

2つ目は、コンビナトリアル合成・計測によるハイスループットデータ収集であり、実験家を中心に、高速に物質を合成し、物性を計測し、データベース化するとともに、視覚化ツール、分析ツールを用いて最適物質、構造を予測するというサイクルを構築するものである。こうしたコミュニティの活動は、エネルギー省傘下の国立研究所ネットワークのもとで、High-Throughput Experimental Materials Collaboratory（HTE-MC）のプログラム推進に結実している。

3つ目は、統合計算材料工学（Integrated Computational Materials Engineering：ICME）の流れであり、構造材料を中心に、結晶構造（第一原理計算）から高次構造まで様々なスケールのモデリングをリンクさせ、また、加工プロセスと組織構造の相関もデータにより統合していくものである。2014年、NISTが立ち上げたCenter for Hierarchical Materials Design（CHiMaD）は、ノースウエスタン大学、シカゴ大学を中心としたもので、熱力学・状態図計算など、個別のニーズに合わせて速度論のシミュレーションを行って材料特性の予測、材料開発の支援を実施している。

MGIに関する最近の注視すべき動向として「Materials Research Data Council：MaRDaC」が挙げら

れる。この評議会は、2019年11月にスタートし、オープン・アクセス可能・相互運用可能な材料データを実現することを目的として設立された「材料研究データアライアンス（Materials Research Data Alliance：MaRDA）」の運営委員会として位置づけられている。具体的な動きとして、最初の会合である「Materials Data Summit」が2019年11月にシカゴで開催された他、NSFから研究コミュニティに向けて2019年10月に発信されたCSSI Dear Colleague Letter：「データ集約型の科学・工学研究をサポートするサイバーインフラストラクチャに関する情報のリクエスト」に呼応して、材料コミュニティからの返答書を2019年12月に返信している。この返信に記述された内容は、(1) サイバーインフラストラクチャ、(2) データベース技術、(3) 相互運用可能システムとユーザビリティ、(4) 自動データキャプチャ、(5) 機械学習、(6) コミュニティ(文化、持続可能性、人材、トレーニング)に関する材料コミュニティからの現状の説明となっている。

もう1つ特筆される取り組みは、NIST-CoE プログラムの1つとして推進されていたCHiMaDの継続である。前述のように2014年からノースウエスタン大学を中心とする研究チームの5年プログラムのファンディングが開始し、NISTのMGI研究チームとの密接なコラボレーションのもとシカゴ地区でのデータ活用型材料研究を牽引してきたが、2019年にさらに5年間にわたるファンディングの継続が決定された。研究チームの中心人物であるOlson教授は2019年12月にMITに移籍しており、CHiMaDセンターとの今後の関係が注目される。

データインフラストラクチャに関しては、MGIの黎明期から国家プロジェクトで開発が進められてきた米国におけるデータプラットフォームシステムのいくつかにおいては、データ公開のサービス運用が開始されている。

Materials Data Facility (MDF) は、CHiMaDのデータツールの1つとしてアルゴンヌ国立研究所を中心に開発が進められてきたものであり、登録されるデータにDOIを付与し、グローバルな学術コミュニケーションの中でデータを流通させることができるほか、組織、登録年、キーワード、関連する元素などに注目した検索機能を有している。

その他としては、ミシガン大学のPRISMS (PRedictive Integrated Structural Materials Science) センターが開発したMaterials Commonsも、同じようなデータ公開機能を有するほか、民間ではCitrine informaticsが旧来からのプラットフォームサービスを継続し、AI活用可能なデータ管理機能を強みとして、製造業との連携を強めている。

さらに、MGIの後継策の一つとしてHTE-MC (High Through-put Experimental Material Collaboratory) が動き始めている。HTE-MCは、高品質な実験データを迅速に大量取得してデータベースに登録し、材料設計を大幅に改善するための材料合成・特性化・データ管理サービスの統合型ネットワークである。サンプル・ライブラリとデータ/メタデータ・レコードを紐づけ、永久的なコードを付与している。サンプルそれぞれと各設備には、クイックレスポンス (QR) コードが付いている。これにより、サンプル・ライブラリに関連付けられたすべてのデータおよびメタデータをワンステップで発見することが可能になっており、また、バーチャル研究所内の様々な測定装置でサンプル測定する際に、新しく測定したデータとサンプル・ライブラリとを自動的に関連付けることができるようになってきている。“Collaboratory”は、Collaboration + Laboratory の造語で、複数の研究所の壁を取り除いた組織、バーチャル研究所を意味している。天文台の国際アライアンス International Virtual Observatory Alliance のデータ共有のあり方を参考にしたものと言われており、データの取り扱いについて求められる FAIR (Findability, Accessibility, Interoperability, and Reusability) ガイドラインを実践している。また、初期は公的資金が必要としつつ、その後は官民協力により自律的な運営ができる体制を目指して活動している。

[欧州]

FAIR原則の重要性が材料科学分野においても認識されつつあるなかで、2018年9月に「FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V. (FAIR-DI)」がドイツ・オランダを中心とする研究機関連携の母体として発足している。この団体は、Horizon 2020 で2015年から

ファンディングされているNOMAD-CoEを主軸に置くが、材料分野にとどまらずバイオ分野・天文学分野とも横串を指す形でFAIR原則に従うデータ管理の実現と、そのための世界的なデータインフラストラクチャを構築することを目的としている。NOMAD（NOvel MAterials Discovery）におけるインフラストラクチャの整備は系統的に進んでおり、生データを登録・管理するリポジトリサービスのほか、解析ツール、可視化、知識ベース、メタデータの統合と合わせて合計5つのサービスを柱として発展している。2020年3月に3年間の追加ファンディングが決定し、2020年6月には材料分野におけるFAIRデータ管理を主題とした国際ワークショップを開催している。

FAIR-DIがデータインフラストラクチャに焦点を当てている一方で、「MAterials design at the eXascale (MAX) a European centre of excellence」はエクサスケールのHigh Performance Computingのためのインフラストラクチャに焦点を当てている。このセンターでは、5カ所のHPCリソースをネットワークし、それらの上で動くcode開発をテクノロジーパートナーとともにサポート、さらにはHPCリソースを必要とするコミュニティとの連携などを主な活動項目としている。このセンターのデータ管理プラットフォームとして機能するのが、Materials Cloudである。PythonをベースとするオープンなインフラストラクチャシステムであるAiiDAをベースとして動作するもので、特に計算データの生成ワークフローを管理し、これに伴うデータ来歴（provenance）の管理も行えることが特色である。AiiDA、Materials Cloudのいずれもスイス連邦工科大学ローザンヌ校（École polytechnique fédérale de Lausanne：EPFL）において開発が進められてきたものであり、必然的にEPFLはMAXの中心的な推進機関でもある。

データの標準化に関しては、European Materials Modeling Council（EMMC）がHorizon 2020ファンディングのもとで進めてきたオントロジーを中心としたものになっている。EMMCはその目標として、材料分野において産業界、とくに中小企業をカスタマーとするデータ市場の立ち上げを掲げており、市場でデータを流通させるために必要な共通記述様式の策定を進めてきた。2018年にロードマップをまとめ上げてHorizon 2020プロジェクトを終了したが、2019年7月に非営利法人化してその後も活動を継続している。データ市場実現に向けた2つのプロジェクトが並行して推進されており、いずれもMODA（Materials Modelling Data）と呼ばれる材料モデリングにおけるデータ記述のための標準規格に準拠した標準オントロジーを基礎においてデータ構造化を進めている。

一方、入出力するデータの形式を揃える別の手段として、入出力結果へのアクセスを全てAPI化し、そのAPIの仕様を標準化するという方法がある。この方式のもと、コミュニティ主導で進められている活動が「OPTIMADE（The Open Databases Integration for Materials Design）」である。OPTIMADEは、2018年から一年に一度のワークショップを毎年主催し、材料分野におけるデータ入出力をREST API化することについて議論を進めてきた。様々なプログラムが入出力を仕様に合わせるように調整を行い始めている。その先鋒となっているのが、NOMADおよびMaterials Cloudの両リポジトリである。加えて、米国の代表的な第一原理計算データベースを運営するAFLOW、Materials Project、OQMDはいずれもこの取り組みに関与しており、16のAPIプロバイダーがOPTIMADEに準拠するとしている。今後、第一原理計算コミュニティにおいて、OPTIMADEはデータ入出力のデファクト標準になる可能性がある。

【その他の国】

中国においては、2014年に上海市と上海大学が共同で進める「上海マテリアルゲノム研究所（Shanghai Materials Genome Institute）」、2016年に上海交通大学において「マテリアルズゲノム連合研究センター」を設立し、近年急速にデータ科学を活用した材料開発研究を進めている。また、2016年には中国科学院物理研究所と北京科技大学が共同で「北京マテリアルズゲノム工学イノベーションアライアンス」を設置し、翌2017年には、中国鋼研科技集团有限公司等も参加する「北京マテリアルゲノム工学イノベーションセンター」を設立するなど、国を挙げてマテリアルゲノム研究に力を入れ始めていると同時に欧米の研究者との連携を強くしている。2016年3月に発表された科学技術イノベーション第13次五カ年計画においても、中国産業の国

際競争力向上のための重点技術の1つ「新素材技術」のなかに「マテリアルズゲノム工学（目標：新材料の開発期間・コストの半減）」と明記されている。さらに、国家重点研究開発計画の1つとして「材料ゲノム工学のキーテクノロジーと支援プラットフォーム」（2016～2018年の3年間に計44課題、総額約8億円）が推進されている。

韓国においては、2015年から10年計画でCreative Materials Discovery Projectが開始している。また、最近、韓国科学技術研究所（Korea Institute of Science and Technology：KIST）において計算科学を中心としたMaterials Informatics Database for Advanced Search（MIDAS）が設置された。

カナダにおいては、2018年11月、カナダ天然資源省は人工知能を活用してエネルギー効率の高い新素材の開発プロセスを加速する世界初のロボットプラットフォーム「ADA」の開発に800万ドルの投資を行うことを発表。エネルギー・イノベーション・プログラム（Energy Innovation Program：EIP）の支援を受け、ブリティッシュ・コロンビア大学は「Autonomous Discovery Accelerator for Materials Innovation」と呼ばれるプラットフォームを構築している。

2. プロセス・インフォマティクスに関する先行研究事例

〔日本〕

- ・東京工業大学のグループは、ベイズ最適化と自動合成、自動物性評価を組み合わせたシステムを開発した。試料搬送用ロボットアームを中央に配置し、その周辺に自動成膜装置、自動評価装置がサテライト状に配置されている。ベイズ最適化を用いて、それまでの成膜・評価結果から次の最適な合成条件を導出し、成膜条件の最適化を行うことを可能にした。これらの成膜・評価・最適化をコンピュータにより一括制御することで、全自動で物質探索ができるシステムとなっている。全固体電解質の成膜プロセスの最適化実験では、24時間で10回程度の成膜がコンスタントに可能であり、人間が実験する場合の約10倍のスループットが実現できるとしている。また、現在の装置を拡張した「マテリアル・ドック」というコンセプトを提唱している。新規物質を合成した際に、目的とする物性だけでなく、電気特性や発光特性、熱電特性、イオン伝導性などの物質の特性を徹底的に調べてデータベース化するものであり、別目的で別の材料を作り出そうとする際に、これまでに蓄積されたデータを活用できるため、効率的な研究開発が可能になることが期待されている。
- ・JSTのERATO前田化学反応創成知能プロジェクトでは、有機合成化学のリモート化に着手している。有機合成化学のリモート化には様々なメリットがあるが、一般的に、危険性を有する化学薬品を使用し、種々の特殊な装置を直接操作する必要があることから、リモートで研究を遂行するのは困難とされる。既にサンプリング等を自動で行うことのできる市販の有機合成装置が導入されている例もあるが、基本的に人間が作業することを前提としたシステムであるため、完全にリモートでの実験は難しい。しかし、操作命令を入力できるセキュアなネットワーク環境や、試薬の自動秤量・分注システム、反応生成物の自動分析システム、広角・多角度・複数のネットワークカメラからなるモニタリングシステムを総合的に組み合わせることにより、リモート実験が可能となる。同プロジェクトでは、2020年6月より「新しい生活様式」に対応したシステムについて構想し、並行して事故発生時の対応方針などの安全管理体制の検討を進めている。市販の自動有機合成装置を導入後、チューニングを経て試運転を開始し、徐々に複雑な構造をもつ有機化合物の自動合成を試みた。また、自動有機合成装置に分析機器を組み合わせ、さらに、遠隔モニタリングのための複数のネットワークカメラを設置し、同年12月現在、全システムを稼働させつつある。新型コロナウイルス感染症の再拡大が懸念されるため、研究遂行に役立てられるよう可能な限り早期のシステム活用を図るとしている。本リモート有機合成システムにより、感染症への対策、作業者の危険な薬品への曝露の防止（安全性の向上）に加え、実作業時間が大幅に短縮する効果が期

待される。

- ・JST未来社会創造事業（探索加速型）「共通基盤」領域の令和3年度本格研究課題「マテリアル探索空間拡張プラットフォームの構築」（研究開発代表者：長藤圭介）では、電池材料を題材として材料探索スループットを1000倍に向上させることをPOC（Proof Of Concept：概念実証）に掲げ、①ハイスループット自律探索システム、②データ駆動/仮説駆動ハイブリッド型研究、③ナレッジシェアリングの3つの柱でマテリアル探索空間拡張プラットフォームの構築を進め、マテリアル研究開発エコシステムの実現を目指している。

〔米国〕

- ・IBMは、AIとロボット、クラウドを組み合わせ、"RoboRXN"と呼ばれる新たな化学研究所をクラウド上に構築した。研究者がリモートで新たな分子を設計・合成するのを支援する。研究者はwebブラウザ経由でRoboRXNにログインし、分子化合物の骨格構造をキャンバス上に描くと、RoboRXNが機械学習を用いて必要な材料と調合の手順を予測し、遠隔地の研究所のロボットに指示を送ってその内容を実行させる。実験が完了したら、分析結果が研究者に送信される仕組みである。ウェブブラウザ経由でRoboRXNに接続し、目的の分子構造を描くか貼り付けると、推薦合成ルートが提示され、ロボット合成のプログラムも作成される。
- ・同じくIBMは、2018年に化学反応を予測するための機械学習モデルRXN for Chemistryを開始している。RXN for Chemistryはニューラル機械翻訳アーキテクチャを活用しており、順合成反応予測は90%以上の精度を実現するとしている。2019年には、逆合成予測もRXN for Chemistryに統合された。ロボット合成を実現するためには、化学合成の手順をプログラムすることが必要であるが、RoboRXNでは、文献から自動抽出して構築した化学合成手順のデータベースが統合され、これにより有機合成ロボットのプログラムを自動的に作成することができるようになっている。また、ハードウェアに依存しないため、市販のロボットを活用することも可能である。ユーザーが分子の設計をリモートで送信し、合成した分子の分析結果を受け取るというコンセプトは、新規分子の発見を加速させるための重要な一歩であると考えられている。

〔欧州〕

- ・英国リバプール大学に「マテリアル・イノベーション・ファクトリー」が2018年に開設された。英国研究パートナーシップ投資基金（UK Research Partnership Investment Fund：UKRPIF）、Unilever、Royce等からの合計8,100万ポンド（約110億円）の投資にもとづき、11,600m²の施設で、計算化学とロボットを組み合わせ新しい材料を発見するため研究開発を実施している。従来のロボット実験では、各装置とロボットアームを連動させるためロボット実験プラットフォームをカスタムデザインすることが必要であったが、リバプール大学の研究グループは、一般的な自走式のロボットを改良し、実験室の中を移動し、通常の化学者が扱うものと同じ機器を用いた実験ができるようにした。ロボットによる光触媒の性能向上を目的とした実験を8日間で688回実行し、最初の配合と比べて活性が6倍高い光触媒組成物を見出すことに成功している。このロボットは、充電時間以外は実験を継続することができ、毎日21.5時間の実験をおこなう。結果、このような短期の日数で成果を実現できたものである。同様の作業を従来のように人間が行えば、数か月かかると思われる。この方法は、実験機器ではなく研究者の手足の動きに相当する部分を自動化するものであり、既存の実験設備をそのままに、幅広い実験に適用することが期待される。

〔その他の国〕

- ・カナダ・ブリティッシュコロンビアの大学等の研究グループでは、有機薄膜の合成、加工、特性評価を

自律的に行う、モジュール式でフレキシブルな運転ができる自動運転プラットフォーム“Ada”を開発した。薄膜の特性を最適化することは、原料組成、成膜条件や処理条件などのパラメータの数が多いため、非常に時間がかかるものである。これらのパラメータには相関関係にあることも多く、隣接層の構造や特性に大きな影響を受ける可能性もある。また、組成や構造に乱れがある場合にも特性を予測できる計算ツールはほとんど存在しないため、プロセスの最適化は未だに経験則に大きく依存している。ハイスループット実験は、パラメータ数が多い空間を検証することに適した方法であるが、薄膜形成の全パラメータを検証することは不可能である。Adaには、空気圧グリッパーを備えたロボットアームがあり、液体の注入やシリコン基板の移送をすることができ、さらに種々のユニットと組み合わせることによって、必要な実験を実施できるようにデザインすることが可能である。またAdaは、最大限の情報量を得られるように実験を自律的にデザインし反復的に実施することで、目標とするパラメータを自動的に見つけることができる。有機太陽電池薄膜の最適化実験では、できるだけ高い移動度を持つ薄膜組成を発見することを目的とした自律的反復実験を実施している。原料インクの調整、基板上にスピコート、熱的なアニール時間の最適化、可視光カメラでの撮影、UV-vis-NIRスペクトル測定、電流-電圧曲線測定、移動度の計算、ベイズ最適化による次の実験の設計、というループを自律的に繰り返すことにより、従来は9か月かかっていたものが、今では5日間できるとしている。

3. 学術論文動向

プロセス・インフォマティクスを推進するのに必要となる要素技術に関する論文動向を以下に示す。

図A-1 (a) はプロセス・インフォマティクスを推進するのに重要な要素技術である「ハイスループット実験・ロボット合成」、「材料プロセスのオペランド計測」の世界の論文数の年次推移を示している。また参照データとして、データ駆動型材料開発つまり広義のマテリアルズ・インフォマティクスの論文数の年次推移も併記している。

2010年代からマテリアルズ・インフォマティクスに関する論文数が急激に増加しているなかで、2010年代後半から「ハイスループット実験・ロボット合成」、「材料プロセスのオペランド計測」についても論文数が増加している。特にプロセス・インフォマティクスの要と考えられるハイスループット実験については、論文数自体はここ2、3年での伸びは著しく、プロセス・インフォマティクスに取り組む基盤が整いつつあることがみてとれる。

また、図A-1 (b) に各要素技術の国別の論文数を示す。両分野ともに論文数では米国、中国が牽引している。日本は各要素技術の開発における先進的な取り組みで遅れを取らないよう研究開発を進める必要がある。

図A-2に材料プロセス開発にデータ科学や機械学習などのデータ駆動型アプローチを導入した研究、つまり本プロポーザルで提示するプロセス・インフォマティクスに直接関係する論文数の推移を示す。世界的にみても論文数はまだ少ないものの、ここ数年で論文数が急増している。

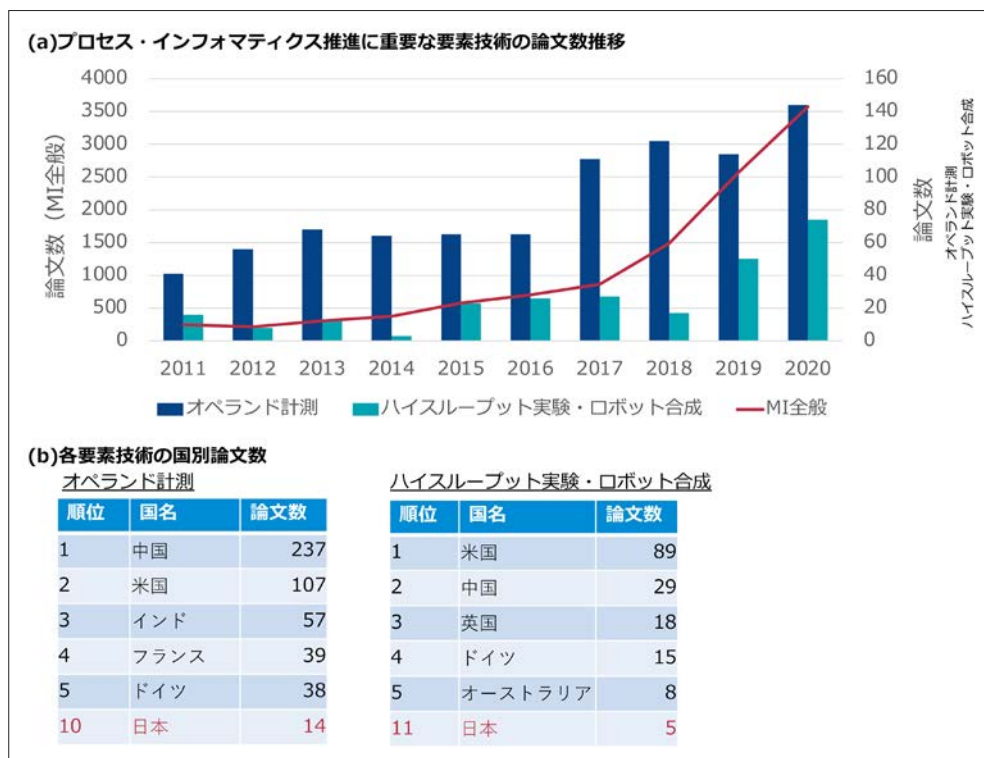


図 A-1 プロセス・インフォマティクス推進に必要な要素技術の世界的論文数推移 (2011～2020年)¹



図 A-2 プロセス・インフォマティクス（材料プロセス×データ科学）の世界的論文数推移 (2011～2020年)²

- 1 データベースはWeb of Scienceを用いて、調査期間は2011～2020年としている。それぞれの分野については以下のキーワードをもとに検索式を作成した。

マテリアルズ・インフォマティクス全般	(データ駆動型 or 機械学習 or AI) AND 材料 AND (設計 or 発見 or 合成)
オペランド計測	(オペランド or in-situ) AND 分析 AND 材料プロセス
ハイスループット実験・ロボット合成	(ハイスループット実験 AND 物質合成) OR (ロボット合成 AND 機械学習・AI)

- 2 データベースはWeb of Scienceを用いて、調査期間は2011～2020年としている。
材料AND (合成 OR プロセス) AND (機械学習 OR データ駆動) をキーワードとして検索式を作成した。

付録3 専門用語説明

機械学習

コンピュータ（機械）を用いて、人間が自然に行うパターン認識や経験則を導き出す活動などを行うものであり、学習の基になる多くのデータを与えることで学習し、学習後に最適な認識や判断を行うことができるようにするもの。

ネガティブデータ

予想・仮説・期待と合致しない実験結果・データ。研究の過程で得られるネガティブデータは通常公表されることがないが、現象を理解し、仮説を検証するためには重要なものである。機械学習の学習データとして使用することで、学習の効果を高めることができる。

オートエンコーダー

ニューラルネットワークの仕組みの1つ。機械学習において、入力されたデータに次元圧縮の処理をして、特徴抽出するアルゴリズム。

デジタルツイン

現実空間（フィジカル空間）にある装置、設備などを仮想空間（サイバー空間）に再現すること。デジタルツインの環境を活用することで、現実空間の材料、プロセスを予測・探索することが可能となる。

ソフトセンサー、バーチャルメトロロジー

直接測定が困難なプロセス変数を、リアルタイムで容易に測定できる変数から予測する技術。プロセスの推定制御技術に活用される。主に化学工業ではソフトセンサー、半導体産業ではバーチャルメトロロジーと呼ばれることが多い。

ベイズ最適化

形状がわからない関数（ブラックボックス関数）の最大値（または最小値）を求めるための手法の一つであり、近年、機械学習において、多次元のパラメータがあっても、網羅的探索に比べて圧倒的に少ない試行回数で最適な条件を探し出す手法として活用されている。

AIロボット

ベイズ最適化などAI手法によって自動的に最適化実験をするロボット

第一原理計算

量子力学の原理に基づき、電子の質量、電荷やクーロン力などの基本物理量から物性を直接導く計算のこと。実験値に合わせるためのパラメータの導入やモデル化などをせず、基本原理から直接物性を導き出す。

作成メンバー

総括責任者	曾根 純一	上席フェロー	(ナノテクノロジー・材料ユニット)
リーダー	福井 弘行	フェロー	(ナノテクノロジー・材料ユニット)
メンバー	伊藤 聡	特任フェロー	(ナノテクノロジー・材料ユニット)
	酒部 健一	主任調査員	(戦略研究推進部)
	佐藤 勝昭	特任フェロー	(ナノテクノロジー・材料ユニット)
	竹内 良昭	フェロー	(環境・エネルギーユニット)
	沼澤 修平	フェロー	(ナノテクノロジー・材料ユニット)
	的場 正憲	フェロー	(システム・情報科学技術ユニット)
	眞子 隆志	フェロー	(ナノテクノロジー・材料ユニット)
	水本 邦彦	主任調査員	(研究プロジェクト推進部)
	宮下 哲	フェロー	(ナノテクノロジー・材料ユニット)

戦略プロポーザル

CRDS-FY2021-SP-01

材料創製技術を革新するプロセス科学基盤

～プロセス・インフォマティクス～

STRATEGIC PROPOSAL

Process Science Platform for Innovation in Materials Creation Technology

- Process Informatics -

令和3年6月 June 2021

ISBN 978-4-88890-744-6

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

Center for Research and Development Strategy, Japan Science and Technology Agency

〒102-0076 東京都千代田区五番町7 K's 五番町

電話 03-5214-7481

E-mail crds@jst.go.jp

<https://www.jst.go.jp/crds/>

本書は著作権法等によって著作権が保護された著作物です。

著作権法で認められた場合を除き、本書の全部又は一部を許可無く複写・複製することを禁じます。

引用を行う際は、必ず出典を記述願います。

This publication is protected by copyright law and international treaties.

No part of this publication may be copied or reproduced in any form or by any means without permission of JST, except to the extent permitted by applicable law.

Any quotations must be appropriately acknowledged.

If you wish to copy, reproduce, display or otherwise use this publication, please contact crds@jst.go.jp.

FOR THE FUTURE OF
SCIENCE AND
SOCIETY



<https://www.jst.go.jp/crds/>