

材料開発の DX に向けた協調戦略

Teiichiro KONO 河野禎市郎 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



材料開発の DX とそのインパクト

マテリアルズインフォマティクス (MI) は、材料開発を加速するデータ駆動型的手法として近年注目を浴びている。これは材料の研究開発力で高い国際競争力を有する日本にとってさらにリードを広げるチャンスであるが、一方で MI のアルゴリズムのさらなる進化や、データベースの整備、短時間に大量のデータを生成する High Throughput Synthesis (HTS) や富岳などの High Performance Computing (HPC)、さらには量子コンピューターなど、数々の破壊的技術の急激な進歩により、いわゆる材料開発のデジタルトランスフォーメーション (DX) が起こり、従来とは比較にならないスピードでの材料開発が一般化してしまう可能性もある。DX の波に乗り遅れば、日本の材料開発の優位性が一気に失われるリスクも無視できない。材料開発の DX に向けた取組をできるだけ早期に産官学が協調する形で進めて行く必要がある。

材料開発の DX における材料データ基盤の重要性

材料開発の DX においては MI などのアルゴリズムの開発力を高めることも重要だが、材料開発のデータはビッグデータとは対照的に規模が小さい場合が多く、データ件数を強化する効果はより顕著である。

データ件数の強化には上述した HTS や HPC などで大量にデータを生成する取り組みが有効である。特に近年普及しつつあるベイズ最適化や能動学習、強化学習とのコンビネーションによるインパクトは計り知れず、国を挙げて取り組むべき革新的技術と言えよう。しかしスタートラインは他国と同じであり、これだけでは競争戦略としては十分とは言えない。

一方日本にはこれまでに蓄積された大量の材料開発に関するデータという資産がある。したがってこれらを武器に早い段階でリードを広げる戦略こそが、日本の材料開発が圧倒的優位性を築く上でのキーになると筆者は考える。そのためには、公開データや各機関が保有するデータ相互活用を可能にする“材料データ基

盤”の整備が重要となる。

材料データ基盤の課題と戦略

材料データ基盤は、図の赤枠に示した、①A~Cの様々なデータソースからデータを統合する機能、②辞書により用語や単位を統一する機能、③構造化^{*1}を行い MI 用データセットとして出力する機能からなる仕組みで、システムだけでなく運用やルールも含めて機能するトータルとしての仕組みである。

現在これらの機能の開発に向けいくつかの公的資金による研究プロジェクトや業界団体主導の取組が進められている。しかし具体的な仕様が描けず迷走気味な状況も見受けられる。仕様策定には、まずはデータ基盤の利用イメージの理解や、現状の技術で何がどこまでできるのか、すなわち技術のポテンシャルやフィージビリティの理解が不可欠であるが、データ基盤の先例がないこともあり、これらの理解がない状況で検討が進められていることが大きな要因と考えられる。

このような取組には、DX に広く用いられるアジャイル開発^{*2}が必要と筆者は考える。これにより上記の理解を深めつつ、仕様を徐々に具体化詳細化しながら検討と開発を進めることができる。

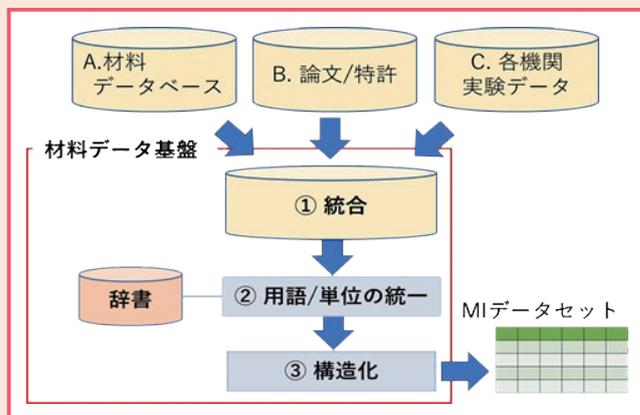


図 材料データ基盤とデータの流れ

*1 テキストや機器の生データ、多様な書式・様式のデータ等を、統一された表形式などの形式に整理・変換すること。

*2 小さな題材を設定し様々な技術の実装とテストをクイックに繰り返す開発手法。

以下そのアジャイルを軸にした実現への取組について、筆者の考える戦略を述べて行きたい。

ステップ 1: データ基盤の全体像の理解 —材料データベースを用いた試行—

最初のステップは、図の一通りの流れの理解や、技術や知識の習得も狙いとなるので、まずは試行を素早く回すことが理想であり、比較的整理されたデータソースである A. 材料データベースを題材とし、簡易なプロトタイプを構築するのが好ましい。

この検討で①～③に投入する様々な候補技術を試行し、どんな機能をどのレベルまで実現可能か、人手で補うべき作業は何か、さらに今後の検討課題は何かを明確にし、仕様を徐々に明確化できる。

また完成したプロトタイプを広く公開することで、多くの研究者にデータ共有基盤を用いた MI の有効性と重要性を具体的に認識してもらい、本取組への共感を得ることができるのではないかと考える。これにより自発的な参画メンバーが一気に増え、以降のステップが一気に加速する正のスパイラルが実現できると考える。

ステップ 2: 対象データの拡大 —特許・論文等の非構造化データの活用—

材料データベースは比較的構造化されており活用しやすいが、対象とする材料や情報は限定的である。一方、論文や特許はテキスト形式の非構造化データであり、構造化の難易度は高い一方、対象は幅広い。これらの構造化には多大な労力を要することから、構造化を自動化する AI の開発が NEDO プロジェクトなどで行われているが、実現にはまだ時間を要する見込みで、当面は手作業で進める必要がある。

米国では CAS が多くの専門人材を擁し構造化で先行しており、日本も早期に構造化を担う人材を育成する必要があるが、幸い日本は優秀なシニア材料技術者を擁する。シニア技術者を対象とした構造化トレーニングを行えば、キャッチアップも十分可能と考える。

なお、論文データの活用においては、出版社との契約問題が壁となっており、大量の論文からデータを手する MI ではライセンス料が莫大になる。まずは出版社に新たな価格体系を整備してもらうべく、産官学が結束して交渉を進める必要がある。一方で日本の成果である論文が海外の出版社の権利になってしまう不条理な問題についても、議論が必要な時期にあらう。

ステップ 3: クローズデータの共有とエコシステムの形成 —制度やガイドライン等の整備—

MI では失敗データや研究機器から生じる生データが学習モデルを構築する上で重要になるが、これらのデータの多くはクローズデータとして各機関が保有す

る。これらは管理や整理が不十分ですぐの共有化が困難なケースが多いが、ステップ 1, 2 で整備した技術や手順などを適用し、育成した人材を活用することで、効率的に解消して行けるであろう。

また、そもそも機密情報の共有には大きな壁がある。この問題に対しては、データを提供するメリットやインセンティブの設計、評価制度との連動など新たな施策が必要と考える。

またクローズデータの共有には、関係者以外への漏洩を防ぐ暗号化やトレーサビリティを担保するブロックチェーン、取引のルールやガイドライン、さらにはデータ取引における収益モデルや価格設定の考え方、取引市場などを含めたエコシステムの整備など、様々な議論と検討を重ねて行く必要があると考える。

さらに当面はオープンデータとクローズデータを組み合わせた活用形態が続くと想定され、その仕組みの構築も必要となろう。こちらもアジャイル的に検証できると良い。例えば、材料の構造に関する分析データは、材料特性や合成条件の MI 解析に有効で、幅広い活用が期待される。材料特性や合成条件に関するデータは機密に該当するケースが多く当面はクローズにせざるを得ないが、材料構造に関するデータの共有だけでも効果はある。データ登録を利用条件にすることでデータの蓄積も進むと考えられ、運用しながらデータベースの持続的な成長を可能にするオープンクローズのエコシステムの実証にもなると考える。

おわりに

上述したように、データ基盤構築には多岐にわたる様々な取組課題がある。これらには産官学が協調してアジャイル的に取組む必要があるが、既存の組織や公的資金による研究プロジェクトなどの枠組み・制度での主導は難しいであろう。既存の枠組みでは緻密なゴールイメージと計画策定、厳密な予算管理が前提となっており、試行錯誤的にゴールを描いて行くアジャイル開発の運用管理にそぐわない。国家予算を投じる以上ある程度の具体的な成果イメージと計画管理、予算管理は当然必要ではあるが、DX 時代に対応するにはこれらの仕組みを見直し、どのような組織や枠組み・制度が新たな協調戦略に必要なか、産官学で自由闊達に議論を行える場が必要であろう。

© 2020 The Chemical Society of Japan

ここに載せた論説は、日本化学会の論説委員会が依頼した執筆者によるもので、文責は基本的には執筆者にあります。日本化学会では、この内容が当会にとって重要な意見として掲載するものです。ご意見、ご感想を下記へお寄せ下さい。
論説委員会 E-mail: ronsetsu@chemistry.or.jp