

日本学術会議 課題別委員会
オープンサイエンスの取組に関する検討委員会
(第23期・第2回)
議事要旨

1. 日 時：平成27年5月21日(木) 10:00~12:00
2. 場 所：日本学術会議 6階 6-A(1) 会議室
3. 出席状況
出席者：土井委員長、杉田副委員長、大杉幹事、戸田山委員、岡委員、喜連川委員(6名)
欠席者：吉川委員、三成委員(2名)
参考人：高木利久 東京大学大学院理学系研究科教授
藤田大介 国立研究開発法人 物質・材料研究機構 先端的共通技術部門 部門長
オブザーバー：渡邊堯 名古屋大学太陽地球環境研究所・客員教授、情報通信研究機構・招聘専門員
真子 博 内閣府政策統括官(科学技術・イノベーション担当) 付参事官補佐(国際統括)
事務局：盛田参事官、松宮補佐、大西専門職、鈴木専門職付、辻上席学術調査員
4. 配布資料：
資料1 第1回議事要旨(案)
資料2 ライフサイエンスにおけるデータ共有の現状と課題
資料3 材料分野におけるオープンサイエンスについて
参考1 委員名簿

5. 議 事：

(1) 第1回議事要旨(案)の確認

資料1に基づいて、前回議事要旨(案)が確認され、了承された。

(2) 先例紹介：高木参考人による説明

高木利久東京大学大学院理学系研究科生物化学専攻教授より、資料2に基づいて、「ライフサイエンスにおけるデータ共有の現状と課題」について説明が行われた。

その後、議論が行われた。主な発言は次の通り。

【データベース、コスト】

- ・ライフサイエンスのデータベースの統合推進については、各省(文部科学省、経産省、農水省、厚労省)で行ったものが、現在では文部科学省のデータベースに統合されているのか。
- 完全に統合されているものと、部分的に統合されている、すなわち、インデックスは統合されているがデータそのものは各省で管理しているものがある。いずれは統合され一か所に集まる見込みである。
- ・人員や予算の圧倒的な不足を検討する場合には、予算・データ・人員の規模を国際比較して、議論することが必要ではないか。
- 例えば、アメリカのゲノムと文献にかかるデータセンターであるNCBI: National Center for Biotechnology Information等は、100億規模の予算で500~700人程度のスタッフを抱えて行われている。このほかにもガンや、微生物などそれぞれのデータセンターが存在している。どの範囲までカウントするかにもよるが、それらを合算すると大きな違いとなる。日本全体で、ライフサイエンス関係の研究予算が4000億円程度あると考えると、最低でも、その1%程度はこうしたことに使わないと上手く回らない。ただデータが保管されていれば良いのではなく、「雑多なデータ」「使い勝手の悪いデータ」を、「共通のフォーマットになっているデータ」に変える作業が必要である。
- ・コスト構造はどうなっているのか。pay-per-viewのようなものの可能性はあるのか。
- 研究そのものが欧米と連携している場合もある。その際に、欧米ではフリーで活用できるのに日本だけが課金するという事は難しいのではないか。

【人材育成、求められるスキル】

- ・バイオインフォマティクス人材の育成状況についてはどうか。
- バイオインフォマティクス人材の不足は、日本で顕著であるがアメリカでも不足している。アメリカではこの問題に対応するために、数十もの研究機関においてバイオインフォマティクスのコアファシリティを設置して解析ニーズに対応している。人材が不足する理由は、パラメータのチューニングやデータのごみ取りなど論文にならない下働きの仕事が多いためである。キャリアパスもつながらず、人も来ない。人材を集めるためには、評価システムを変えたり、インセンティブを変えなければならない。データをオープンにすることと、それにかかる人を育成し増やすことを両輪で進めていかないとならない。
- ・バイオインフォマティクス人材に求められるのは、専門的な知識なのかそれとも、ある程度分野横断的に共通する汎用性のある知識なのか。
- 両方が必要である。専門知識がないとプログラムが書けない。常に新しい分野・種類のデータが出てくるという状況により、必要な専門的な知識もどんどん進んでいくということが問題である。
- ・バイオインフォマティクス人材の育成プロセスの日米の違いはどうか。
- 欧米には育成する学科が存在し（300程度）、サマースクールなども豊富であり、大学内に教育する仕組みがある。一方で日本には数個しかない。新しい学科を作ることができないので、根本的な手を打たなければならない。その他にも、日本の場合は生物系の人材は、プログラムを書くことが苦手であるという傾向がある。また古い分野から新しい分野に人を移行させるような、人員の整理・採用の仕組みがないということもある。さらに学生にとってみれば、バイオインフォマティクスの採用数が見込めないことも、こうした分野を学ぶインセンティブがないという状況である。
- ・バイオインフォマティクス人材は、ライフサイエンス分野の解析人材を自ら育成していると考えられることもできる。データにかかる政策を全体で考えて、ライフサイエンスに限らず他の分野でも活動できるスキームの構築を考える必要があるのではないか。
- 分野共通の部分と分野固有の部分をきちんと描いて進めていくことが良い。

【データ登録】

- ・論文投稿時に出版社からデータベースへのデータ登録の義務付けがされるというのはどのような仕組みか。
- 論文投稿時に、公的データベースに登録しその番号を示さないと、投稿そのものが受け付けられない。データベースが指定されている場合と、公的なデータベースなら認められる場合がある。
- ・ライフサイエンス以外の分野における投稿にかかるデータベース登録の状況はどうか。
- アメリカの GEO では求めている。ファンディングがそのように要求していることが影響していると言える。

【ノウハウ、活用方法】

- ・プログラムチューニングのナレッジ、ノウハウ部分についてはどのように共有化されているのか。
- その点については特に共有化されていない。生データがあっても、プログラムが変化しており、再現することはできないというのが現状である。それを解決するために、仮想化ソフトウェアを使ってデータとプログラムをまとめて再現できるようにする取り組みが始まっている。
- ・エコシステムで考えたときに、データ提供の部分については比較的進んでいるが、しかしそのデータをさらに活用するアウトカムについては、アメリカにおいても手がついていないということか。
- そのような状況である。
- ・データベースで保存されている生データの活用イメージについて説明してほしい。メタ解析などで新しい知見を生み出すといった流れはあるのか。
- メタ解析は欧米では進んでいる。一方で日本では数件しか行われていない。それはデータの圏

い込み、バイオインフォマティクス不足という問題が背景にある。

【国の資産としてのデータ】

- ・アメリカのデータセンターに任せるといえるのか。
- 現実に、欧米のデータセンターに登録する人もおり、日本のデータベースをやめても良いのではないかという議論はある。またイネや日本人のゲノム等重要なものに絞ってデータベースを構築してはどうかといった議論もある。重要なことは、データベースのマネジメントをしないと、データを使うノウハウが蓄積されず研究ができなくなるということである。新しいデータが出てくるときに、それをデータベースとして蓄積することを行わないとそれに伴う技術者がいなくなってしまう。日本として何をすべきか、また他の国と重複しても行うかどうかといったことについてきちんと考える必要がある。
- ・内閣府の議論はどうか。
- 公的研究資金を使って、海外のデータ登録等が行われている現状については、これを国内のデータベースにきちんと登録し蓄積していくべきであるという議論が行われた。
- ある種の国防的な論理で、三極（アメリカ、ヨーロッパ、日本）でデータベースをそれぞれ持っていることについては意味がある。特にヨーロッパは、仮にアメリカが方針を変更したとしてもオープンデータ利用を維持できるというところを狙っていると思う。日本についてはまだ議論は進んでいないのが現状である。
- ・受け皿としてどこに置くべきかということと、国の資産をどうすべきかというのは別の議論ではないか。研究がデータを基礎として展開していく時代において、海外にデータを置くということはもはやありえないのではないか。その中でいかにローコストにデータベースを設置できるのかということが重要である。そう考えるとやはり、ライフサイエンスをはじめとする個別の分野ごとに考えるのではなく、分野を超えてデータベースについて考えなければならないのではないか。

【その他】

- ・ツールセットのローカリゼーションが進んでいるとも言えるだろう。そうした仕事については、大学の教員にはインセンティブがない。しかし、プロダクトニーズがあるということならば、企業は全体のプロセスをパッケージ化することに取り組むのではないか。
- 10年ぐらい前に、そのように思う企業があつて、取り組んでいた。しかし、パラメーターチューニング等が多くあるのでパッケージ化できずに結局撤退してしまった。製薬会社の規模もそう大きくない日本では、企業の商品化ニーズも弱い。
- ・文化遺産的なものをどのように共有化していくかという問題もある。

(3) 先例紹介：藤田参考人による説明

藤田大介国立研究開発法人物質・材料研究機構（NIMS）先端的共通技術部門長より、資料3に基づいて、「材料分野におけるオープンサイエンスについて：材料イノベーションを加速するオープンサイエンスの動向について」の説明が行われた。

その後、議論が行われた。主な発言は次の通り。

【計測データと計算データ】

- ・計測データ（実験データ、実測データ）と計算データが異なってしまった場合はどのように処理するのか。
- もはや、計算データは計測データに匹敵する段階になっていると考えられる。誤差はあるものの、膨大な計算データを集めることにより物性の誤差推定の精度が高まっている。とはいえ、計測データは不可欠である。

【材料分野のオープンクローズ戦略、ノウハウ、技術の成熟度】

- ・国家の資産としての材料科学のオープン化については、戦略として考えるとデリケートなのではないか。
- その通りである。すべてをオープンにするのではない。オープンにするところとクローズにする

るところをはっきりさせることが必要である。計算データはオープン、ノウハウはクローズにする。材料の競争力を維持するために必要なノウハウはオープンにはしない。

- ・ノウハウは、オープンサイエンスとどのような影響にあるのか。またノウハウについて考えると、ノウハウを開発する人材はどのように育成されるのか、これまでかんとひらめきだったノウハウの蓄積はオープンサイエンスで変わっていくのか。
- ノウハウの重要な部分は製造プロセスである。この製造プロセスの部分までオープンにする必要はない。アメリカの取り組みでは、新しい材料を作っていくときにどの組成が適切なのか「あたり」をつける。膨大な計算データをもとに、有望な物質群を見つけていくのである。ある程度予測したうえで実際に実験を行い、そこで良い材料ができれば特許を取得し、後にデータをオープンにするという方法をとっている。こうした効率的な取り組みを日本でも行いたいと考えている。
- ・アメリカの場合は、有望そうなものもそうでないものもオープンになっているのか。ノウハウは、技術の熟度が進むと必要なくなるのではないのか。技術の成熟度にあわせて、オープンとすべきかクローズとすべきか検討するのではないのか。→アメリカの場合でも、有望なものについてはすぐオープンにするのではないと思われる。彼らの取り組みで重要なのは、最終的にはオープンにしましようと言っていること、そしてそれによってデータが集まりデータの中心にいることである。
 - ・国際競争だけではなく、国内にも研究機関の間の競争はある。国内ならばノウハウが連携されるわけでもないし、海外の企業等との連携もありうる。そう考えるとナショナルに限定するということがわかりにくい。
- 欧米では自国内の企業に閉じているわけではない。一方で日本は国内企業に限定している場合が多い。NIMSについてみると、海外企業も参加しているが、オープンなシステムというわけではない。
- ・ライフサイエンスのデータについては公的資金によるものはオープンにするというルールであったが、材料についてはどうか。
- データベースに入れようとしているデータの種類は計算データである。日本でこれを行う場合は、無償で提供する仕組みを作らないとならない。関連してNIMSの計算リソースを提供することもありうる。ただしすべてにおいてオープンにするのではない。知財権を取得するまではオープンにはされない。
- ・潜在物質を計算機で作ってから、製造プロセス（ノウハウ）に入るという仕組みを前提とした場合に、何をどこまで共有すると活性化した研究につながるのか。オープンサイエンスのプラットフォームを考える際に、オープンクローズ戦略をどうすべきなのか。どのようなメッセージが必要なのか。
- NIMSのマテリアルズデータベースのプロジェクトでは、オープンサイエンスの概念がまだ入っていない。しかし、次期の科学技術基本計画にオープンサイエンスが乗ってくれば、状況は変わるだろう。

【オープン化のメリット】

- ・アメリカで計算データのデータベースが増えていったときに、知財を経たものがオープンにされるメリット、その中心にいるメリットは何か。
- アメリカでオープンデータから知財化するノウハウについてはわからない。データマイニングの技術はクローズドである。この部分が重要である。

【「オープン」の意味】

- ・「オープン」と言っても色々な意味がある。分野による違いがある。ビジネスを念頭に置いて特許を取得してからオープンにするところもあれば、研究に重きを置いてすべてを直ちにオープンにするところもある。前者はライフサイエンスや材料の分野で、後者は天文の分野がそれにあたる。ライフサイエンスはデータがオープンであってもチューニングして計算する方法の部分はノウハウでクローズドであり、材料分野も計算データがオープンであっても製造プロセスの部分（ノウハウ）はクローズドである。一方で、天文分野のようにすべてみんなまで共有してやっていくという世界もあるが、これはライフサイエンスや材料とは異なる。

- ・データマイニング・そのデータをどう使うかということがポイントとなる。データマイニングが弱いのに、データのオープン化・標準化の部分だけ国際的にリードしてもどうなのか。

【その他】

- ・ISOの標準化、VAMOSとはどのような意味か。
- VAMOSはオープンな国際標準のルール化のことである。具体的には、材料の分野では計測方法とその誤差などについて議論を国際的に行っている。こうした議論を行う際に、良い人材を投入し、日本の企業に有利な交渉をしていかなければならない。

(4) その他

【次回の開催について】

今回は、6月15日(月)16時より開催し、大石雅寿 大学共同利用機関法人自然科学研究機構国立天文台天文データセンターセンター長、黒川顕 東京工業大学地球生命研究所教授の2名を参考人として招きヒアリングを行う。

【今後の審議の進め方について】

7月以降、論点の整理を行い、必要に応じて調査等を実施することとした。

以上