

日本学術会議
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8回）
議事要旨

1. 日 時：平成28年3月18日（水）13:00～15:00
2. 場 所：日本学術会議5階5-C（2）会議室
3. 出席者：土井委員長、杉田副委員長、大杉幹事、高木委員、岡委員、喜連川委員、三成委員（7名）
欠席者：戸田山委員、吉川委員（2名）
事務局：石井参事官、松宮補佐、熊谷専門職付、鈴木専門職付、漆畑上席学術調査員
4. 配布資料：
資料1 第7回議事要旨（案）
資料2 提言「オープンイノベーションのためのオープンサイエンスのあり方」（骨子案）
非公開資料 アンケート結果
参考1 委員名簿
レジュメ「（生命科学分野から考える）オープンイノベーションに向けたあり方（に関する話題提供）」（高木委員）

5. 議 事：

（1）前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）論点まとめ

高木委員より、資料「（生命科学分野から考える）オープンイノベーションに向けたあり方（に関する話題提供）」に基づき、生命科学分野におけるデータ共有の現状及び問題点・課題について説明が行われた。本資料の主な構成は次の通り。

- ・生命科学におけるデータ共有の意義、原動力
- ・生命科学における研究データ共有の歴史
- ・データ提供者の義務とインセンティブ付与
- ・データの利活用に関する障害
- ・わが国の生命科学DB統合プロジェクト
- ・NBDC（バイオサイエンス・データベースセンター）ヒトデータベース
- ・GA4GH加入組織
- ・GA4GH運営体制
- ・残された課題と対応策（1）
ーデータ量増大への対応、データ提供者のインセンティブ等
- ・残された課題と対応策（2）
ーデータ共有ポリシー、ガイドライン策定、データの品質管理、データの権利等
- ・残された課題と対応策（3）
ーデータキュレーター及びサイエンティストの育成、データ統合解析環境の整備等

プレゼンを踏まえ、引き続き行われた主な議論は次の通り。

【わが国と米国のデータ共有】

- ・資料の中のNBDCヒトデータベースは、科学技術振興機構（JST）が持っているのか。
- JSTの中の1つのセンター（NBDC）が構築・提供しているデータベースである。ただ、組織の財政が苦しくなればデータベースをやめるという危険性がある。ミニマムな部分は確保しておかなければいけない。
- 米国の場合は一応法的に担保されていることだが、それは法的にどういう形でなのか。
- 具体的にはデータベースセンターが根拠法に基づいて作られているということだが、その他にも、例えば、電子情報公開改正法のように国民に研究成果を知らしめるいくつかの法律がある。

それらが基盤にあるので簡単に資金不足だからやめてしまおうということにはならないようだ。日本よりは状況はよい。

→（米国では、）州の自立性とか情報と個人、企業とかの自立性とか、連邦法で一括して情報の統合化は法的に規定されているのか。

→電子情報公開改正法は今から 20 年ほど前に通った。たしかアル・ゴア（元米副大統領）が生命科学分野文献データのインターネット公開のきっかけを作った。

【企業サイドの動き】

→その共有データがある企業にとって重要な時、企業がそのオープンデータを自分のところで加工するようなことがなされるのではないかと。企業サイドの動きはどうか。

→共有や統合によりどの程度の成果があったのかは、シークレットでなかなか見えない。データアクセスへのトレースはしないようにとどの製薬会社も強くいうことで、どの企業が開発したか、どの検索ワードを使ったか、何をダウンロードしたのかといった足跡は消すようにいわれている。そのため企業への貢献ははっきりしない。

【サステイナブルなコスト負担】

- ・（共有データを）サステイナブルにするコスト負担モデルは、本日プレゼンの一番の要だろう。データ共有のポジティブサイド（論文発表が増加するなど）、ネガティブサイド（研究者の負担増等）のバランスをどう図るかは、ROI（費用対効果）を見るしかなく、何を入れるか捨てるかという「目利き」が重要ということだろう。

→そここのところは、企業に聞いてもなかなか伝わってこない部分だ。

- ・米国でたびたびデータベースのための財源が少なくなっているという話があったが、これは単に財源がないということなのか、その有用性について疑問が出ているということなのか。

→有用性も入ってくるかもしれないが、これまでは財政的な事情が大きかった。

【分野ごとのデータ共有に向けた取組み】

- ・植物頑健性について、クレストのところで始めているとのことだが、植物頑健性のクレストが終わった後に集めたデータベースはどうするのか。

→一般に公開する予定だと聞いている。

→喜連川先生のビッグデータのプロジェクトでも、何かそういうものはあるのか。

→（私たちは、）「データさがし」 というものを作っている。日本の中でデータをきっちりキュレートして、分析の前段階で統合化したしっかりとしたものを作れる人もそれ相当集まった。それはピュアオープンで、腸内細菌、脳波、農業など結構いろいろなものがある。

- ・結局、たまたま植物頑健性の場合は JST にセンターがあるから残せるということか。

→他のプロジェクトでもデータを受け取る仕組みはあるので、いただければ私どもが預かって公開する。ただ、データをコンバートするなどの費用負担はできない場合がある。

→植物頑健性の場合は、最初からそういうことを段階的にやっていったということか。

→植物頑健性は今年始まったばかり。ただ、その受け皿はライフサイエンスでは私どものセンターしかない。それ以外の分野はどうしていくのかは別の問題である。

- ・米国の場合、データキュレーターとかデータサイエンティストといった人たちは、それなりにポジションやキャリアパスがあってというシステムができていると考えていいのか。

→日本よりはだいぶある。米国や欧州のセンターでは 100 人、200 人の構成で、Ph.D を持ったキュレーターが多く働いている。パーマネントポジションも得られる仕組みもある。

【学協会へのアンケートについての説明】

- ・アンケートについて説明の後、委員長によりこれまでの議論の途中経過が整理された。

- ・今の議論を踏まえて、そろそろ提言を考えたい。その前に、事務局が学協会にアンケートを行っている。完全に内容を見たわけではないが、回答をもらった学協会を見る限り、感触としては、データ公開をしているところは意外に多い。シンポジウムとかデータ管理をやっていることから見ると、そういう意識はどこも持っているようだ。

【何を残すかを定める機関】

- ・何を残して何を残さないかは学協会で決めなければならないということだが、それに実効性を持たせることは、ファンディングエージェンシーがお金を出して行うということか。
- それを義務付けるのは、ファンディングエージェンシーでないとできないということ。学協会が意見をいっても何らかの義務をかけられる権限を持つところしかできない。何が必要かという議論とどうやって実効性を持たせるかという議論は別に考えた方がよいのでは。
- 実効性の話は米国がいったんやってしまっていることなので、米国が今問題となっていることを踏まえて何をいうかが問題。
- オープンクローズの議論は分野別に全然違うので、学協会レベルである程度やるしかない。

【データプラットフォームの構築】

- ・提言には、実際に誰が責任もって残すのかが問題。学協会に対してどういうメッセージを出すかが提言 2 に関わる。提言 3 は、データを作成、キュレーションする人材育成の話。提言 4 はレポジトリ機関の利用に関するもの。ほかに提言に入れることがあればいい。
- このデータシステムが今後新たに国家として重要な役割を持つときに、キュレートして格納するというファンクションを実現するというのは各領域がやらなくてはならない。しかし入れる器自身は、集約化した方が圧倒的に効率化できるしサステナビリティも高まる。
- 国がオープンデータの枠組みを用意するから、それを皆で使おうということか。
- このサイトにこれだけ入れなさいというサービスを国家として行うが、障壁が少なく自然な形で進むことにつながる。自然な形で研究しながらデータが入っていくというプラットフォームを作っていく、その枠とかデータ形式はドメイン・バイ・ドメインなので、それは各々にやっていただくというそんなメカニズムがいいのではないか。大きなデータを出す分野が一定の負担をするとなると、残りの部分はそこの余剰を使った方が国益的にも圧倒的に得だ。

【個人情報の取り扱い】

- ・個人情報の扱いについてどこまでハードルを下げさせるかについてすぐに検討しないと、文科系では本当に利用できないデータとなってしまう。これも提言に入れた方がよい。
- ヒトとノンヒトは根源的に異なる。個人に紐づくような医療のようなデリケートなものは、国が用意する器の中で扱うのは難しい。ヒトに接する人文社会の中でのデータは医療のデータではないので、人文社会系のデータはそこで集約して入れるというのがいいのでは。

【データセンターの維持】

- ・最近の EBI (European Bioinformatics Institute) レポートによれば、データオープンについて、100 億規模の費用に対し経済的に 40~50 倍の効果があつた。研究者が同じデータを産出する手間が省けた、データへのアクセススピードがあがったなど、大きな効果があつたということが書かれている。そのセンターを維持する根拠としては、そうしたものが参考になるのでは。
- 先ほどのご指摘の論文だけではないという指標を明確にしておかないと、集約してもそんなにお金をかける必要があるのかという意見に対抗できるようにしておかないと、そういうセンターのサステナビリティの根拠としては不十分。
- 「入れ物」は国が用意するかもしれないが、それをどう維持するのか。
- 研究費から一定%納入するという考え方と、そこに国家予算をつけようといういい方もあるだろう。議論は分かれるが、提言にはまず国家としてデータを蓄える、でないとコミュニティによって足並みがそろわない懸念があるということ盛り込むということだろう。また、全学問分野を対象とした場合、オプトイン（一定のコストを負担してデータを入れることについて、その学問分野の許可を得る）とするべきかもしれない。その学問分野がやりたいと思った時には国家がそのインフラを用意する、など少し穏やかめにするのもいいのでは。

【国際的な情報共有】

- ・オープンにする範囲の問題だが、例えば日本にいる外国人はどう考えるのか。国際的に協力している部分もあるわけだが、それについて欧米はどう制限しているのか。
- ライフサイエンス分野では、基本的には制限なくできる。ただこれは、米国国内でしかできな

- いというようなことも一部はあるようだが、だいたい国は国の制約なくできるようになっている。
- 今の内閣府の報告書では、特に国費を投じたものについてはきちっとできるのだからやれという主旨のことが書かれている。恐らく国の研究費を使ったものは原則入れる方向でやりなさいというのがむしろ広すぎるくらい広がっているというが、やや懸念事項ではあるが、対象にはなっている。だから当然、海外には全部出す前提で、ただ今デジタル情報であるものを全部出してどれだけ実効的かという問題は別途ある。
 - 原則一番この大きな問題は中国。R&D も一番困っているのは中国対応だ。ただそれをいろいろなわけにもいかないので、原則グローバルなアクティビティとして日本もそれに先導的に取り組んでいくとせざるを得ないだろう。
 - それをオープンにすることが国益につながらないような話になるとまずい。その意味で人材育成が負担になってくる。データをオープンにした段階で、(植物頑健性プロジェクトのように、)日本も欧米に伍してそれを使えるのだという部分がないといけない。入れ物だけ大きなものを作ってグローバルに利用してもらおうというメッセージだけではなかなか大変だ。
 - 国益に関連して、分野によっては少しずつポーズを見せつつ戦略で持っていくしかなく、初めから全部オープンという話ではない。そこは表現を少し緩やかにしておくのがよい。

【データ管理の労力コスト】

- (データ管理を)集約すると、一声コストは必ず半分くらいになる。その浮いた部分をもっと活用する側に投資するというのはとてもいいことだ。
- 内閣府の報告書では、分野別の戦略とは別に機関の役割としてどこまで守るか、「ゴミ箱」まで守るのかという議論になっている。それを入れるとコストの話も相当大変。そこは機関に任せますとした上で、コンバートした後の流用性の高いものをどうやって使うかというところに特化するか否かについてここで方針として示した方が議論は楽なのでは。
- イメージとしては、研究者が持っている生データがあり、それが学協会の捨てる捨てないという何らかの判断があるものがコンバートされて大きな器に入っていく。そこに各々の学協会が何を捨て何を残すかというポリシーに基づき、どういうフォーマットにするかとか決めてもらって入れていくという、大きな絵をとりあえず描けばいいのではないかと。人材という意味では、研究のプロセスとそのキャリアを明らかにする絵が一枚あった方がいいとは思う。
- サイエンス 2.0 では、データを軸にした科学方法論の効率化が、今後の大きな視点のひとつになると随分いわれていた。データ共有の労力・コストは大きいので、全部用意する方向が今後重要で、そのハイライトがデータオープンだというと学術会議的には美しいのでは。
- レポジトリとダブルでやるということか。
- 現実には、この大きな空間の 7 割くらいは機関が持つレポジトリ。そこには中間プロセスのようなものが随分入っている。その 7 割は、何か文句をいわれた時の説明証拠としておいておくもの。あとの 3 割は発信側の上等なストレージにして分ける。技術的にはそういうソリューションになってくるのでは。
- 学術会議としては、器を用意してくれればいいよと。学協会も苦労していると思う(サーバーを設立したり、クラウドを使うとか)ので、それがなくなるだけでも随分違ってくる。人材育成についても、高木先生のレポートなどを根拠にそれを目指した方がいいということもできる。匿名化についても、そこに入ればちゃんと使えるというイメージになれば違うと思うので、そういうところを目指してもらいたいと、書いておくことが重要だろう。

【今後の進め方について】

- ・学術会議の意見として出していくことが重要だと思うので、今日の意見を基に要旨(2 頁程度)をまとめてみるのでそれに対して議論してもらい、高木先生からも資料や EBI 報告書をもらい、文科省からも学術委員会報告書ももらいながら、提言としての体裁をまとめることにする。事務局にはアンケートを付録でまとめる時のたたき台を出してもらいたい。9 月までには提言を出すので早くまとめたい。

以上