

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9回）
議事要旨

1. 日 時：平成28年4月25日（月）10:00～12:00
2. 場 所：日本学術会議 5階 5-C（2）会議室
3. 出席状況
出席者：土井委員長、大杉幹事、吉川委員、高木委員、喜連川委員、村山オブザーバー（情報通信研究機構）、渡邊参考人（文科省学術基盤整備室長）、安達参考人（国立情報学研究所副所長）（8名）
欠席者：杉田委員、戸田山委員、岡委員、三成委員（4名）
事務局：石井参事官、松宮補佐、大橋専門職付、鈴木専門職付、漆畑上席学術調査員
4. 配布資料：
資料1 第8回議事要旨（案）
資料2 学術情報のオープン化の推進について（審議まとめ）
資料3 オープンサイエンス推進のための研究データ基盤
資料4 提言案「オープンイノベーションに資するオープンサイエンスのあり方に関する提言」
参考1 委員名簿

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

渡邊参考人より、資料「学術情報のオープン化の推進について（平成28年2月26日科学技術・学術審議会・学術分科会・学術情報委員会審議まとめ）」に基づき、文部科学省のオープンアクセス、オープンデータの取組及び課題について説明が行われた。説明について議論された後、安達参考人より、資料「オープンサイエンス推進のための研究データ基盤」に基づき、国立情報学研究所（NII）によるオープンサイエンスへの取組み、データ基盤整備に関する説明が行われた。

【学術情報のオープン化の推進について（渡邊参考人のプレゼン資料）】

渡邊参考人によるプレゼン資料の構成は次の通り。

「学術情報のオープン化の推進について」

- ・検討の背景
- ・基本的考え方
- ・研究成果の公開についての基本的方策（骨子）
- ・研究データ基盤整備の方向性
- ・継続的な検討の必要性

（参考）審議まとめにおける関係機関の役割

本プレゼンに基づく主な議論は次の通り。

【資料に関する確認】

- ・JSPSは研究資金配分機関に含まれているのか。
→そうである。
- ・資料2-1（3頁）の「グリーンOA」とは何か。
→「グリーン」は、出版社から雑誌に掲載されているものを二次的に許諾を得てアーカイブする、そしてフリーに公開するもの、「ゴールド」は最初からオープンアクセス・ジャーナルとして刊行と同時にフリーアクセスになるもの。
→資料同頁の「質の高いOA」はどちらの意味か。

→ゴールド OA のこと。

- ・これは、研究者が作成した論文とそれの基になるデータのオープン化の範囲ということか。誰でも利用可能な基本データではなく、論文化されたデータということか。

→そうである。今回学術情報委員会は、論文とエビデンスデータに絞り提言をまとめた。

→資料 2-2 に「研究成果の利活用」とあるが、誰が利活用するのか。論文の筆者でなく、そのデータを見た二次利用者のことか。

→そうである。

→ジャーナルに論文を出した時は、論文のデータと元データの作成は同一人物が行うのが普通だろうが、異なる場合はそれを明記するということか。

→同じ場合でも、公開データだけを利用する場合もあるだろう。自分たちが使えるものがあれば、データを引用した時にそのデータ作成者を明記するということ。

【データ管理の専門人材】

- ・同資料同頁の「人材の育成及び確保」のところ、「研究データを専門的に取り扱える新たな専門人材の育成」は、このように範囲を限定した際にも必要になるのか。

→誰が実際に支援をするのかという議論が出ていた。ただ、日本ではそういった専門人材がいないということで、こういう提言が出てきた。

- ・OECD の国際的なデータインフラの整備に関するポリシーについて会議があり参加した。そこではヒューマンリソースがないとデータのマネジメントができない、そういう専門人材が必要という議論があった。二次利用できる基礎的データを国際的に共同利用したいという方向だが、データのエコシステム、ライフサイクルというようなものを回していこうとする場合に、インフラ装置があっても専門研究者が次の人のためにデータを整備する余裕はない。専門人材がいて次の人のために整備する、つまりインフラというのはヒトとともに構成されるべきという議論であった。

→ご指摘の点は重要。ここは、大学等の連携の枠組みの中でアカデミッククラウドにより基盤を作っていこうということ。一方で人の手当てという話は、財政的に厳しい現在の大学のおかれている現状を見ると、今いる方を能力開発して新たな機能・役割を担ってもらおうと同時に、大学に対する期待としては量的な拡大も考えないと、今後のこういうデータを中心とした流れには対応できないのではないかとということで人員の拡充も提起した。

【データ作成・管理に係る研究者の負担・業績評価】

- ・同資料 2 頁に「継続的な検討の必要性」とあるが、どこまで踏み込んでのことなのか。内閣府にもフォローアップ検討会があるので、それとの関連も含め説明してほしい。

→この課題は中長期で考えるべきこと。一方、分野ごとにどうするのかという話は必ず出てくる。コミュニティの検討状況や日本学術会議の提言も踏まえて丁寧にフォローする必要がある。内閣府での議論、学術会議での議論が進んだところで検討の内容を示したい。

- ・研究者に負担にならないよう気軽にデータを作成できるようにしておかないといけない。負担は最小限にという趣旨のことはあまり書いていないが。

→ストレートではないが、支援策の中で研究者負担を減らそうという趣旨の部分はある。

- ・2 頁の (4) のところで、「データ作成を業績として評価する」とあるが、文科省としてはこの問題でリーダーシップをとろうとは考えているのか。

→データのサイテーションは行政がどこまで対応するのは難しいが、研究者が研究の中で実行していくことが基本と考えている。一方でデータの引用は、論文のように仕組みが整っているわけではない。データセットの識別子をつける取組みは試行されており、それを基盤に仕組み作りをする中で、研究者の行動に期待しているという段階。

→データ・サイテーションを業績評価につなげるために、学術ジャーナルのポリシーにデータを引用するように、強制ではなく“encourage”することが重要である。

→研究者のホルダーは大学であり、大学が教員をどう扱うかということだ。

【オープンサイエンス推進のための研究データ基盤（安達参考人のプレゼン資料）】

安達参考人によるプレゼン資料の構成は次の通り。

「オープンサイエンス推進のための研究データ基盤」

- ・オープンサイエンスの意義
- ・政策によるオープンサイエンス推進
- ・オープンサイエンス推進に不可欠な基盤整備
- ・オープンサイエンスの実現に向けて
- ・Open science のための共通基盤整備
- ・研究データ基盤とは
- ・NII が提供する学術情報流通基盤
- ・目的と実施内容
- ・管理基盤（バックエンド、フロントエンド）
- ・公開基盤
- ・ディスカバリサービス基盤
- ・実施機関と実施体制
- ・年次計画
- ・オープンサイエンスを支える研究データ基盤

本プレゼンに基づく主な議論は次の通り。

【データ識別子の国際化について】

- ・利用者の識別子、研究者の識別子を国際化していくという文脈は今後あり得るのか。例えば、シンガポールの大学では全学のアカデミック情報管理のための国際研究者 ID (ORCID-ID) 取得の動きがある。(ORCID: Open Researcher and Contributor ID)
- 欧米の政府機関、出版社などが参加するコンソーシアム活動があり、NII もフォローしている。わが国でもいくつかの大学が大学単位で ID を取ろうとしているし、先生方によっては、出版社で論文を出した時に ID を出ささいといわれたこともあるだろう。それは、NII の立場ではコミュニティが決める話。現実的な問題としては ID がひとつに決まるのが望ましいが、科研費の研究者番号など様々な ID がありそれらをうまく紐づけることの方が重要で(論文を特定するより研究者を特定する方が難しい)、台帳のようなデータベースがなければならない。
- マッピングのようなものを導入することで、事実上国際的なものとの相互流通することのバリアは除去できるということか。
- 日本語で書かれた論文をうまく外に出すのは難しい。英語の論文誌を作成する学会ではそういう問題は徐々に解決できるのでは。

【基盤に入れるデータの取扱い】

- ・欧米で基盤ができてそれと協調する上で、現在検討中のシステムで問題ないのか。
- 欧米の動向との整合性を見ながら設計している段階。NII としては、研究者の ID 問題などを日本の多様性を想定して作る必要はあると考えている。
- 研究者としては、管理基盤に入れるデータを公開するか否かは自ら決めていいのか。
- 状況による。例えばある分野では、論文投稿の段階で使用した論文を全て記載しなければならず、査読の時にはデータを見ながら査読する場合は、草稿の段階でそういうデータを作成し、管理基盤に入れて査読者が見られるようにしなければならない。再読された時にそのデータの公開が求められるので、管理基盤から勘定は移って例えば大学のリポジトリから見られる、それが有用データと見られれば、検索サービスなどで不特定の研究者によるアクセスが可能となる、といったことが考えられる。そういうスタイルを目指してコミュニティが動く、これがオープンサイエンスにつながる。一方、分野によっては共同で大きなデータを持っていて、それをコミュニティで共有して研究するといったストレージを確保し、データを作成するために苦労する研究者たちがいるというスタイル。
- そのようなスタイルの場合はどうするのか。
- そこが頭の痛いところだが、ストレージを用意すると一部の大きなデータを持つ研究分野がそれを使わせてほしいといってくるのが考えられるので、それを調停する仕組みが必要。例えば、スパコンを使ってシミュレーション結果を保存したいという研究者もおり、ストレージが

いくらあっても足りないという分野もある。そうしたデータの中には公開基盤に入ってくるケースもある（公開基盤に入っていないものも多くある）。

- セミ・オープン、コミュニティ・オープン（コミュニティの中でだけまずは使用する）という感じだろう。一枚岩でくくられているように見えるが、比較的大きなものはまず自分だけのものを入れておく（研究過程に係るもの）、それが一定程度できてコミュニティに広がっていくが、そこには制限がつくことになる。例えば、気象庁のデータは国交省が管理しているように重層的な権利制限がかかってくる。そういうことはあまり資料には書いておらず、まずは論文、エビデンスからまずオープンにするところから始める。そこをシードにしながら周辺のデータを維持することがないと、研究者にも有益ではない。ゴールは、研究者がいかにデータを使いやすいようにして学問発展につなげるかということだが、何もなければできないので、まずは論文、エビデンスといったコアの部分から広げていくということ。
- 欧米との協調は、まずコアの部分から何年計画で対応していくことで大丈夫なのか。
- 英国で高等教育委員会情報分野の会議があって参加した。先方は様々なプロジェクトの仕様書を作っており、それを基にNIIとして先方と直に議論している。欧州の方向性を見ながら、当方の考え方とのすり合わせをしている。オープンサイエンスに関し、向いている方向は日本とは違ってはいない（例えば、オープンサイエンス・クラウドを学界に提供するなど）。
- オープンサイエンス・クラウドについては、EU研究イノベーション総局が予算を出そうとしているが、他の研究データ基盤整備には別の総局が予算を出すという状況で、それぞれのポリシーで動いてきた状況である。そうした状況をウォッチすることで、今後の戦略を考えられる時間はある。
- 欧米も含めてまだ方向感が安定しているわけではない。その状況にわが国も入って行ってオープンサイエンスの方向性を見ながら、日本の状況を踏まえて対応するのが、情報システム設計の腕の見せ所。
- 日本は、アジアのリーダーシップを取ることが期待される立場で、日本が音頭を取って日本発・アジア発の方式・設計を現状の欧米の議論に反映させられるようにするのが望ましい。

【「学術情報」の意味】

- ・学術情報という縛りでデータを集めるのと、分野を超えて（例えばライフサイエンスは厚労省とか）様々なデータが一緒に集まるのが分野としては有難い。そのあたりの構想を考えると、学術分野として集めるというのはどういう考えなのか。
- NIIは「学術情報」を「（研究のために）大学が必要とする情報全て」と定義している（大学の研究に必要なならば、政府の保有情報も「学術情報」とみなす）。省庁の縦割りで分けているわけではない。ただ、際限なく広げるのも問題なので、そうした曖昧な線を引いている。
- 「学術情報」は、研究者の立場から見た時にシームレスに使えるようなものという意味ととらえている。学協会にはこの構想に対し動いているところ、動けていないところがある。データベースの構築など難しいことではなく、このポイントを押さえれば参加できるというものを作りたい。もう一つは、データ基盤を有さない小規模大学や、どうしていいかわからないという人文社会科学のために、そういう団体が参加してもらえるために参考となるものをまとめた。各々のポリシーを定めてもらい、それによって自分たちが作成したデータがオープン化でどうなるのか、どう使われるのかといったことを書いておけばいいのではと考えている。
- 文科省は「学術情報」をそういう意味で使っているのか。
- 「学術情報」は、資料本文にあるように「主に教育研究活動の成果として生み出される論文、研究データ、教材等」としている。
- 文科省は科研費を含め論文ベース、NIIは「大学」というキーワードで「情報」ととらえている。省庁のように情報を持っているが情報管理も行うような組織体（各省庁のデータを融合して関係者のディポジットをするといった仲介役）は、文科省、NII以外にないのか。
- オープンデータとガバメントデータ（国が作成したデータ）は性格が違う。例えば、米国はオープンガバメントということで、データの公開に取り組んでおり、日本も考え方の上で影響を受けているが、それを組織的にどうするかということまでは議論が至っていない。ただ、現実には個別にそういう話は出てくるので、いろいろな形での協力は考えられる。
- 内閣府オープンサイエンス検討委員会の議論はやや自然科学データ系が多く感じるが、日学は

人文社会科学を含めて学術全分野について議論をしようとしている。総合的に実際のマネジメントができる組織体を作るのはまだいくつか関門がある。

→人文社会科学系についても、例えば、国文化研究資料館が作った古典籍データベースに対し、NIIはシステムとクラウドで協力している。また、土木工事の際に発見された遺跡に係るデータベースをデジタル化し共有するプロジェクトを、いくつかの大学と連携し進めている。

【オープンサイエンス検討委員会の提言について】

土井委員長から、提言に関する資料4の内容、提言案に対する委員の検討状況などについて説明があった後、構成、追加事項等提言内容について、委員間で次のような意見交換が行われた。

- ・提言案は査読が入り、意見をもらってから修正ということか。親部会はどうなるのか。
- 科学と社会委員会に査読してもらい、幹事会にあげる。同委員会には5月中に提出し、そこから査読してもらい6月24日に間に合わせる（6月頭に査読が終了している前提）。
- 提言案の要旨のところで、「研究データを蓄える学術情報基盤の重要性を指摘」と、提言の「研究データ基盤の整備」は異なる意味なのか、重なる部分はあるのか、似たものなのか。
- 方向性はかなり似ている（今回の提言の方がより具体的）。
- 提言の中の「欧米の学会や出版社が... オープンアクセスを進めてきている」について。欧米と日本の方向性が似ている日本がリーダーシップを取るということと思うが、「欧米データの囲い込み」の部分は必要か。日本はネガティブな話はあるが、欧米に対抗するためにオープンにするということでもあり、「大きく遅れをとっている」という表現はいかがなものか。
- その部分は再検討する。
- 背景としては、学術誌がほぼ100%欧米出版社に握られ日本が遅れをとったという反省がジャーナル誌問題にあったこと。それに対し、データはまだ国際的に議論が続いており、欧米が枠組みを作る前に日本が介入する余地がある、今のうちに問題解決の一步とするということ。
- ここは丁寧に書かなくてはならない部分。学術誌問題もあるしオープン化の流れもあるが、オープン化ではまだ介入できる余地があるということ。
- ・京都大学の先生が、段階的にデータ整備・公開を進めるという提案をされていた。「オープン」という言葉に拒絶感を出させない、あるいは「オープン」という言葉にこだわりすぎないような工夫が必要。
- ・文章と図表の整合性がわかりにくいし、細かすぎる印象。わかりやすさを意識しては。
- これは一般市民に対し出すものではなく、研究データ基盤が必要で欧米でもオープンサイエンスの取組みは評価されていることをいうためのもの。あえてこのレベルで作成した。
- ある限定された人に向けて発信するのなら問題ないが、日学から発信するものは一般向けというのが前提なので、わかりやすさは意識した方がよい。
- 誰に向けて提言を発信するかということだが、資金を出してもらおうためのものだとすると、内閣府、文科省の責任者に説明することを念頭においている。
- 科学外交として、オープンサイエンス戦略で日本がアジアのリーダーとして立つとしたら、日本の立ち位置、情報と基盤をアジアにどう転換していくかという視点も必要。
- 提言を英訳して出すという話もある。全部でなく要旨と研究データ基盤の部分の英文訳も付録につけるといのもひとつ。あるいは提言の部分だけを1頁にまとめてもよいだろう。
- ただ、それは時間ばかりかかるので、今の政府の動きからすると早い方がよい。
- その意味では、査読に出した時点で原山議員に説明することは必要だろう。

【今後の進め方について】

できるだけ連休前に意見をもらい、その結果を連休明けにお知らせする。並行してフォーマット化し、5月中旬に科学と社会委員会に提出する。できるだけ6月上旬に査読を終わらせる（6月の幹事会にかけて、6月24日に間に合わせる）といったスケジュール。次回委員会を開催するのは難しいので、委員長に一任いただきたい。

以上