

第2回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

機関リポジトリによるプレプリント公開

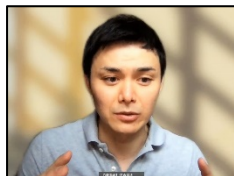
河合 将志

(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)

講演要旨



本報告では、機関リポジトリによるプレプリント公開の可能性について検討する。具体的には、プレプリントサーバーとしての機能が求められるようになった背景や、実装にあたって見込まれる問題などについて論じる。加えて、機関リポジトリのこうした機能変化が、学術情報流通における書誌多様性に与える影響についても言及する。



河合 将志

大阪大学大学院国際公共政策研究科博士後期課程修了。オープンサイエンス基盤研究センターではデータ解析を担当しており、システムの利用統計処理などに携わっている。

報告内容

本報告では、近年、プレプリントサーバーへの関心が高まっていることを受けて、機関リポジトリでプレプリントの公開が可能なのかという点について検討します。まず、機関リポジトリとプレプリントサーバーの違いについて確認した上で、機関リポジトリによるプレプリント公開の可能性について検討します。また、プレプリントと書誌多様性の関係が本日のパネルディスカッションのテーマですので、機関リポジトリによるプレプリント公開が可能になった場合における書誌多様性への影響についても最後に言及します。なお、書誌多様性という概念は、単純に学術情報流通の各側面における多様性ということかと思えます。

機関リポジトリとプレプリントサーバーの違い

まず、機関リポジトリとプレプリントサーバーの位置付けを確認します。

公開基盤であるオープンアクセス (OA) リポジトリの中の大きいカテゴリーの一つに、機関単位で運営

される機関リポジトリがあります (図 1)。国内の機関リポジトリとしては JAIRO Cloud や紅が挙げられます。もう一つの大きいカテゴリーとしては、個別の研究分野ごとに運用される分野リポジトリがあります。例えば PubMed Central (PMC) などです。この分野リポジトリのうち、プレプリントに特化したものがプレプリントサーバーです。arXiv や Social Science Research Network (SSRN) が例として挙げられます。

次に、機関リポジトリとプレプリントサーバーのコンテンツの違いを、学術雑誌論文を例に見ていきます (図 2)。まず、機関リポジトリの場合は、出版社に



(図 1)

よる体裁調整がなされた出版社版がコンテンツになる場合もありますが、出版社版の一つ前のバージョンで、査読はなされているものの体裁調整がなされていないポストプリントが主なコンテンツです。

一方のプレプリントサーバーの場合は、ポストプリントのさらに一つ前のバージョンで、査読も体裁調整もなされていないプレプリントがコンテンツです。

次に、プレプリントサーバーが着目される背景を把握するためにも、機関リポジトリのデメリットについて確認します(図3)。機関リポジトリのコンテンツは査読がなされたポストプリントなので、機関リポジトリが公開の迅速性や更新性に関して優れているというわけではありません。従って、研究者から見た場合には、査読が済んで掲載が決定しているのに機関リポジトリにわざわざ登録するメリットは何なのだということになるかと思います。

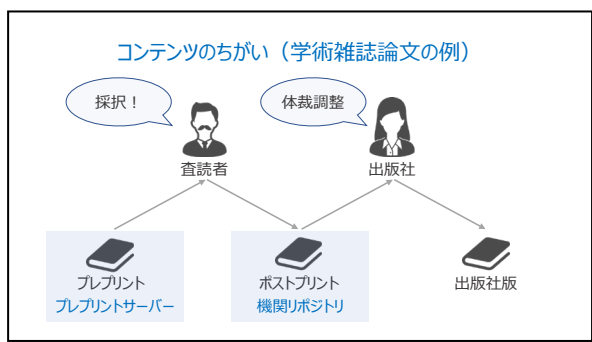
研究者が機関リポジトリにメリットを見いだせていないということは、私が行っている解析の結果にも表れています。図4は機関リポジトリの学術雑誌論文コンテンツを増やす上での各活動の効果の推定値を示し

たものです。縦軸は各活動、横軸はコンテンツを増やす上での活動の効果を表しています。図書館が研究者にポストプリントの提供を依頼する学術雑誌論文提供依頼や、登録されたコンテンツのダウンロード件数を通知するダウンロード件数通知の効果が高い一方で、研究者自身がコンテンツを機関リポジトリに登録するセルフアーカイブの効果は非常に低いことが示されています。その理由としては、研究者からのセルフアーカイブの要請がないためにそもそも実施していないことが多いようです。つまり、研究者は図書館から頼まれたらポストプリントを提供するけれども、セルフアーカイブをするほどの魅力を機関リポジトリには見いだせていないということです。

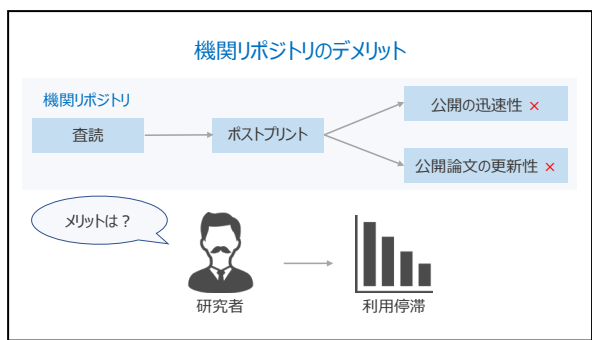
一方で、プレプリントサーバーのコンテンツは査読がなされていないプレプリントなので、プレプリントサーバーは公開の迅速性や更新性に優れていると言えます。従って、研究者からプレプリントサーバーを見た場合には、先取権を確立するためのタイムスタンプを気軽に押せる魅力的な基盤に見え、そのことが利用評価につながっていると思います。

以上のように、プレプリントサーバーのメリットは、プレプリントサーバーそのものの性能というよりは、査読がなされていないというプレプリントの性質に由来していると言えます。

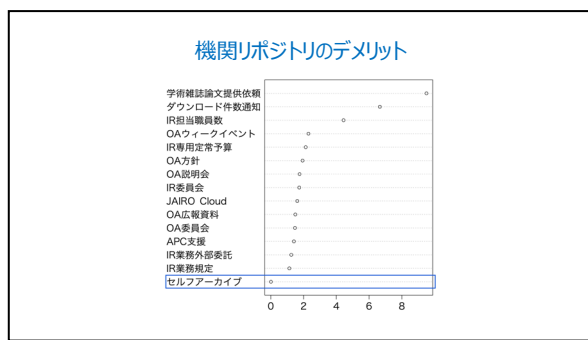
そう考えたときに、プレプリントサーバー以外のOAリポジトリによってプレプリントを公開する可能性はないのかという疑問が当然ながら生じてきます。わざわざプレプリントサーバーによって公開するのではなく、日本において充実した整備状況にある機関リ



(図2)



(図3)



(図4)

ポジトリによってプレプリントを公開することが可能かどうか、検討する必要があると思っています。より端的に言えば、既にある機関リポジトリを活用すれば済むのではないかという疑問が生じてくるわけです。

機関リポジトリによるプレプリント公開の可能性

疑問点を確認したところで、機関リポジトリによるプレプリント公開の可能性を検討します。

日本において機関リポジトリの数はリニアに増えており、現在 800 以上の機関リポジトリがあります。この数は世界的に見ても非常に多いものとなっています。そのコンテンツとしては、紀要論文が最も多く 54%で、次いで学術雑誌論文が 14%となっています。また、博士論文や会議発表論文、データベースなどもあり、多様なコンテンツが登録されています。

これらのコンテンツのうち、主要な幾つかを研究フローに対応した形でソートしてみます（図 5）。当然、分野によって順序は変わってきますが、大体このようにソートできるかと思います。これを眺めると、研究フローをおおよそカバーするコンテンツが登録されていることが分かります。このソートしたコンテンツの並びにプレプリントを落とすと、データベースと会議発表論文の間あたりになるかと思います。ここで言いたいのは、機関リポジトリの既存コンテンツの中にプレプリントを位置付けることに不自然さが無いということです。機関リポジトリは単に整備状況がいいだけではなく、プレプリントとの相性もいいので、機関リポジトリによるプレプリントの公開は可能なのではな

いかと考えています。

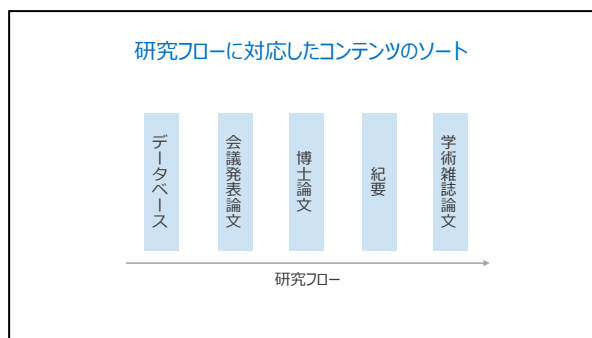
少し話がそれるかもしれませんが、より将来的な話としては、プレプリントだけではなく、研究フローをさかのぼって研究のアイデアなどもコンテンツ化し、なおかつ、公開後に SNS のようにフィードバックも得られるようにすることが考えられます（図 6）。もちろんフィードバックは欲しいけれども、コンテンツによっては、まだアイデア段階なので完全公開まではしたくないというニーズもあると思うので、公開の範囲を限定できる制限公開機能なども併せて使うことが考えられます。この制限公開機能については、国立情報学研究所（NII）の朝岡先生が現在開発中です。

研究者は機関リポジトリにメリットを感じていない現状にありますが、プレプリントに限らず、望むコンテンツを望む範囲で公開でき、なおかつフィードバックも得られるとなれば、プレプリントサーバー以上のメリットを機関リポジトリに感じられるようになると思います。

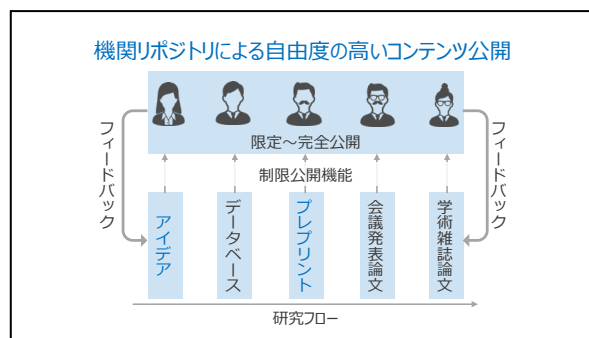
書誌多様性への影響

最後に、機関リポジトリによるプレプリント公開や、その発展版である、より自由度の高いコンテンツ公開が可能になった場合の書誌多様性への影響にも言及したいと思います。

先ほどのような自由度の高いコンテンツ公開を可能とする基盤があると、各研究者は公開コンテンツと公開範囲を自由に選択できるようになります。その選択が堆積していくと、結果として書誌多様性が生まれる確率が高くなります。ただ、多様性がないということ



(図 5)



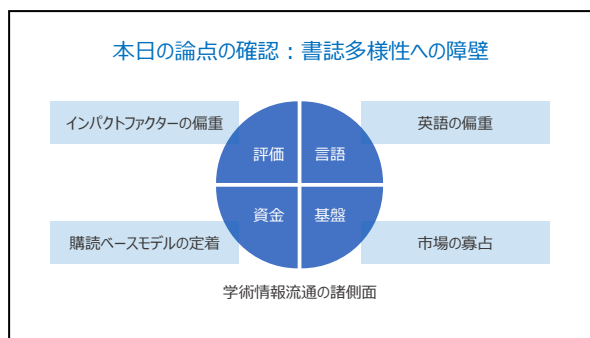
(図 6)

は、例えばルールが明確であるとも言えますし、多様性がある方が良いということが必ずしも自明なわけではありません。従って、そこが不明確なまま、多様性の構築そのものが目的化しないように注意する必要があります。

本日の論点の確認：書誌多様性への障壁

本日のパネルディスカッションの論点である書誌多様性への障壁について簡単に確認しておきたいと思います（図 7）。学術情報流通の諸側面として「言語」「基盤」「資金」「評価」があるとして、言語面では英語の偏重、基盤面では市場の寡占、資金面では購読ベースモデルの定着、評価面ではインパクトファクターの偏重という問題があり、各側面において多様性が失われていることがレポートで指摘されています。パネルディスカッションでは、それが良いことなのか悪いことなのかという根本的な点も含めて議論していくことになると思っています。

●矢吹 プレプリントがリポジトリに入ってくるとなった場合には、その後のポストプリントもある、掲載される論文もあるということになります。多様なバージョンが全て1カ所に入ってくる場合、見ている側にとっては、これはどうするのだという感じにならないかと思いました。限定公開のような仕組みに走ればうまく切り分けられるかと思いつつも、そういうことがあるのではないかと思います。



(図 7)

●河合 例えば、プレプリントを載せた後にポストプリントを載せたら上書きされるという形になれば、紛らわしくないかと思います。

●矢吹 同じ趣旨の質問が来ています。「各機関の方針やシステムによると思いますが、機関リポジトリで、プレプリント版や、査読を経て学術雑誌等で公開された版といった複数のバージョンが公開される場合、バージョン管理にはどのような方法が考えられるでしょうか」。

●河合 最新のもののみを公開したいのか、それまでの変遷を公開していきたいのかも、ユーザーである研究者が選べるようにしていくべきだと考えています。

●矢吹 もう一つ質問が来ています。「内容の質保証についてはどのようにお考えでしょうか」。プレプリントは査読前のものということもあり、そういったものがリポジトリに入ってくるときにどう考えるかという意図でしょうか。

●河合 プレプリントである以上、その公開には研究者のための公開という側面があると思います。フィードバックのための公開といった側面です。

●矢吹 次は大きな質問です。「プレプリントは主にはどの学術分野で行われているのでしょうか」。ご存じの範囲でお願いします。

●河合 どこで特に盛んかという点については、把握できていません。

●矢吹 この後の講演でも幾つかの分野についてのお話が出てきますので、そこでもこの質問の答えが出てくるかもしれません。

もう一つ、これもまた難しい質問です。「機関リポジトリにおけるアイデアの公開は著作権で保護されま

せんが、研究活性化とアイデアの保護のバランスはどのように取っていくことができるでしょうか。

●河合 フィードバックが欲しい一方で、アイデアが取られてしまう可能性は当然あると思いますが、そこはあくまでもユーザーの責任、ユーザーの判断で行ってもらわなければならないと思っています。

●矢吹 次は NII 船守先生からです。「研究者がプレプリントを機関リポジトリに登録するインセンティブが必要ですが、例えば Plan S の権利保持戦略を日本の研究助成機関に採択してもらい、かつプレプリントを公開することを義務化するというのはどうでしょうか。日本の研究のビジビリティの向上にもつながると思います」。

●河合 そういった規定があった場合には登録が加速すると思います。また、ユーザーから見ると、例えば国のリポジトリのような感じで一つのブランドに見えるような仕組みをつくらないと、モチベーションが上がらないと思います。

●矢吹 実は私も、リポジトリは機関別に整備されていて、一方でプレプリントのサーバーは分野別にあるので、その辺の関係はどうなるのかと思っていたのですが、今のお話がそれに当たると思いました。

質問がもう一つ来ました。「機関リポジトリは機関ごとにつくられたものですが、研究者が分野内でコミュニケーションを取れることの方が大事ではないでしょうか。分野ごとに機関リポジトリをグルーピングするような構想はありますか」。

●河合 先ほどの話ともつながると思いますが、ユーザーから見た場合に、そういったくくりがあるように見えるようにしなければいけないと思っています。

●矢吹 深貝先生からコメントです。「休憩時間の前

にでも、フォスタリング文書について皆さんの翻訳をご紹介いただければと思います。パネルに入る前に、関心のある方は休憩時間などにこれにアクセスできるでしょうから」ということです。先ほど翻訳していた文書と元の文書のリンクは、後でチャット欄等を通じてシェアさせていただければと思います。

もう一つ質問が来ました。「リポジトリの登録へのインセンティブとして、登録すると世界的なデータベースへの取り込みも自動的にされるということがあるように思いますが、どうでしょうか」。

●河合 今のところ具体的なイメージは持っていませんが、今後は考えていかなければならないと思っています。