

第1回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

「研究データの公開・利用条件指定ガイドライン」と 制限公開

池内 有為

(文教大学文学部/

研究データ利活用協議会 研究データライセンス小委員会)

講演要旨



2019年に研究データ利活用協議会(RDUF)研究データライセンス小委員会は、研究者やステークホルダーが研究データのライセンスに関する共通理解を深め、適切な公開方法を決定するためのツールとして、「研究データの公開・利用条件指定ガイドライン」を公開した。ガイドラインは、文献調査、インタビュー調査、質問紙調査、イベントでの公開ディスカッションなどに基づいて策定された。講演では、ガイドラインの概要、およびフルオープンと制限公開を判断するプロセスについて紹介することによって、本日の議論の枠組みを示したい。

池内 有為

文教大学文学部専任講師。博士(図書館情報学)。慶應義塾大学法学部卒業(1995年)、同文学研究科図書館・情報学修士課程修了(1997年)の後、1997年から2005年までフェリス女学院大学附属図書館勤務。主婦、筑波大学大学院博士後期課程を経て、2019年より現職。専門領域は研究データ共有とオープンサイエンス。2017年に研究データ利活用協議会にて研究データのライセンス検討プロジェクト小委員会を立ち上げ、2019年から研究データライセンス小委員会としてガイドラインを策定。2020年度SPARC Japanセミナー企画ワーキングメンバー。



本日は研究データ利活用協議会の研究データライセンス小委員会として参りました。小委員会では「研究データの公開・利用条件指定ガイドライン」を作成したのですが、その中で、本日のテーマであるフルオープンと制限公開の境界線に関する課題がいろいろと見えてきましたので、お話しさせていただきます。

研究データ利活用協議会と研究データライセンス小委員会の概要

研究データ利活用協議会(RDUF)は2016年6月に設立されました。研究データを公開して利活用しようとしたときに、ライセンスの問題や、データの引用が

進まない、あるいはリポジトリをどうするかなど、さまざまな問題が出てきます。それらについて、それぞれ関心を持つ産官学の人たちが集まって小委員会を作り、課題解決に向けてボランティアベースの活動を行っています。

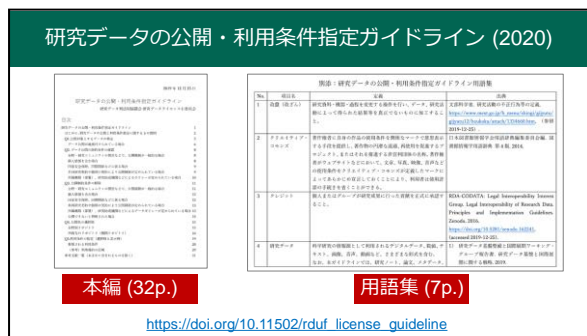
研究データライセンス小委員会は2017年10月から活動しています。研究データを公開する際に、制限公開したいとき、何か条件を付けたいときに、どのようにライセンスを付ければいいのか、あるいは付けられているライセンスを正しく理解するにはどうしたらいいのかということがなかなか難しいので、分かりやすいガイドラインを作ろうという小委員会です。産官学

から多種多様な愉快的なメンバーが集まって活動して
ますので、ご関心のある方はぜひご参加ください。本
日この後ご講演される海老沢先生や上島さんにもメン
バーに加わっていただいています。

「研究データの公開・利用条件指定ガイドライン」
は2019年12月にできました(図1)。当初は1年半く
らいでできるかと思っていましたが、いろいろと難し
くて、討論に討論を重ねて、最終的には時間をかけて
作りました。今年の頭から公開していますので、ご関
心のある方は、ぜひご覧ください。シンプルに作った
つもりですが、いろいろ盛り込んで、本編が32ペー
ジとなりました。それに加えて7ページの用語集を付
けています。そもそもデータの公開とはどういうこと
か、というところから非常に難しいのですけれども、
こうした「データ公開」などの基本的な用語について
も定義を行っています。

ガイドライン作成の経緯と得られた知見

そもそもの前提として、データを公開するなら、理
想はOpen by defaultでフルオープンになっていると、
使う側としては非常に安心して使えます。しかし、私
たちが調べた中でも、分野によって慣習があったり、
研究機関や共同研究者の事情でどうしても制限せざる
を得ない、機密情報やプライバシー情報、個人情報や
DNA情報が含まれているからフルオープンにはでき
ない、特許や商業的利益に関わる、あるいはフルオー
プンにして誤解を生んだり剽窃されてしまったらどう
しようなど、なかなかフルオープンにはできない、制
限で公開したいというニーズがあります。



(図1)

では、制限を付けて公開すればいいというように話
は簡単に進まず、制限公開する際にもいろいろな問題
があることが分かりました。まず、利用条件が複雑で
す。多くの研究データは著作物として保護されません。
公開者としては、自分の望む利用条件で使ってほしい
のですが、その契約なり条件付けに非常にコストがか
かります。再利用しようとする利用者としても、ライ
センスや契約が付いているけれども何をすればいいの
か確信が持てない、利用条件や契約内容がよく分ら
ないとなると、やはりデータの公開や利活用が進みま
せん。

では、標準的な、みんなが共通理解を持っているよ
うなライセンスがあれば、それに乗っかっていけば進
みそうなのですが、なかなかそうもなっていません。
既によく普及していると予想されるクリエイティブ・
コモンズ・ライセンスなども、まだ認知度はそれほど
高くありません。また、クリエイティブ・コモンズ・
ライセンスの場合は、意思表示に使うことはできるの
ですが、データそのものに著作物性が認められない場
合などは、それを侵された場合も法的な保護を受けら
れない可能性があるという問題があります。

国際的には、2016年にResearch Data Alliance (RDA)
とCODATAが「研究データの法的相互運用性：指針
と実施のガイドライン」を公開しています。これをそ
のまま翻訳して公開することも考えましたが、大部で
あり、また日本の状況にどれだけ当てはまるか検討が
必要でした。

日本版ガイドラインの策定

そこで作成したのが日本語版のガイドラインです。
データを公開したい側も、データを利用したい側も、
すぐにライセンスを理解し、分かりやすくライセンス
を付けられるようなものを目指しました。

このガイドラインを作る上でのプロセスは図2のと
おりです。まず、基礎文献の収集・翻訳から始めて、
インタビュー調査をあちこちの機関に対して行いまし
た。2、3年前になりますが、まさに本日講演をして

くださっているような機関にメンバーが足を運び、靴底を減らしてお話を伺ってきました。アンケート調査は、研究者の方や、研究の支援をされる方、図書館員の方などに対して行い、その結果をもって Japan Open Science Summit (JOSS) などのイベントでセッションを開催し、ステークホルダーの方や会場の方とも話しながらいろいろな事例を集めました。そしてドラフト版を作り、関係者の方からレビューを頂き、ガイドラインが昨年末に完成しました。このあたりの話は南山委員長が記事で紹介していますので、もしよろしかったらご覧ください。

制限公開のニーズ

図3はアンケートの結果です。データを公開する際に何か条件を求めたいと思われませんかということで、インタビュー調査で聞いてきた項目を挙げてお伺いしたところ、やはり全体的に制限公開したいというニーズがあることが分かりました。「成果にクレジットを表示したい」「不適切な利用をした場合は利用を禁止したい」「改変禁止」といったことを研究者の方は求

めています。「何も求めない」はやはり少なく、公開に当たっては何らかの制限を求めたいというニーズがあることが分かりました。

そして、既存ライセンスの認知度についても伺いました(図4)。クリエイティブ・コモンズ・ライセンス、オープン・データ・コモンズ、政府標準利用規約についてそれぞれ聞きましたが、認知度の「はい」の比率はあまり高くなく、ご存じの方であっても、実際に使っている方はごく限られていることが分かりました。

また、求める条件が遵守されるならば、自身のデータを公開してもよいと思われませんかという問いに対しては、「公開してもよいと思う」「やや公開してもよいと思う」は合わせて88.5%でした。ライセンスあるいは制限公開の条件がきちんと整理され、かつそれがうまく遵守されることが担保されれば、データの公開や利活用が進むのではないかとことが示唆されました。

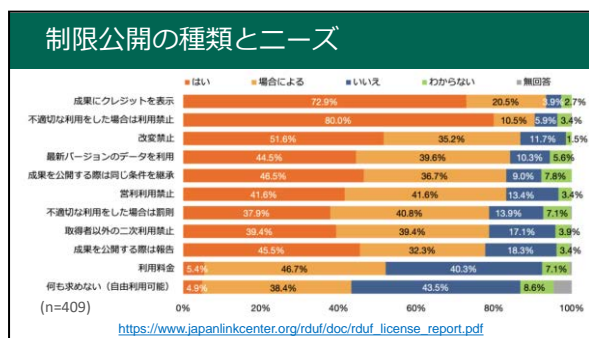
本日の論点：制限公開とフルオープン境界線

本ガイドラインでは、研究データの公開・利用条件を指定するときのワークフローを用意しています(図5)。まずは対象のデータを選び、公開データと非公開データの区別、条件の解除を一定期間後にするかどうか、公開先、そして利用条件を考え、データの公開に至るというものです。

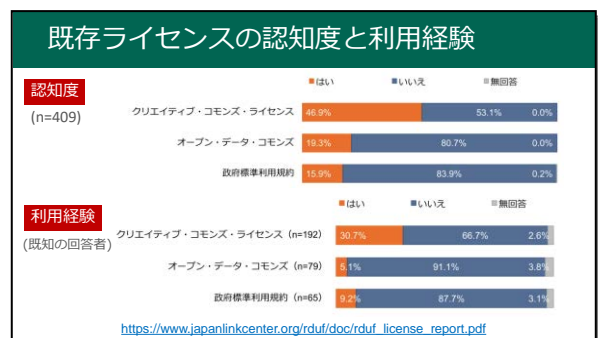
その中で、聞き取り調査やディスカッションをしていて、やはりここが難しいという話になったのが、まさに本日のテーマである公開と制限公開の部分です。



(図2)



(図3)

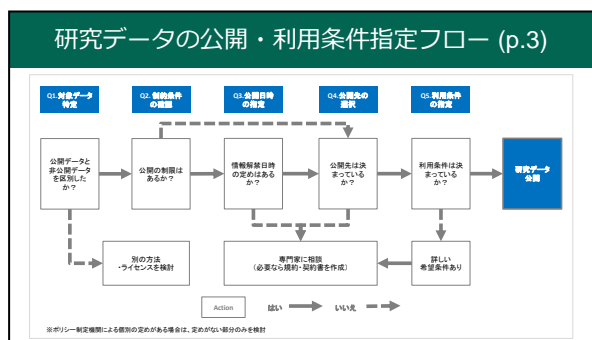


(図4)

公開・非公開をどう線引きするか、その上で、フルオープンにするのか、それとも一定の条件を設けるのか、機関などでのコンセンサスを得ることが非常に難しいというのが、これまでの調査の感触です。では、具体的にどう制限を設けるのかということも、また一つ判断が困難な部分です。一つの機関や個人の研究者が考えていくことは非常に難しいですが、本日のようにいろいろな機関の間で話し合う機会があると、相互に大変役立つのではないかと考えます。

さまざまな区分と制限

この話をするときに、四つのレイヤーがあるというところが少しややこしいので、改めてご説明します（図 6）。公開に関しては、フルオープンと制限公開があります。そして制限を付けつつ、条件を満たしたとしても外部の人（機関外、グループ外の第三者）には見せないというのが制限共有です。こういう線引きが割と多くなされているので、そこで線を引き、基本的にはこの四つに分けて考えていただければと思います。



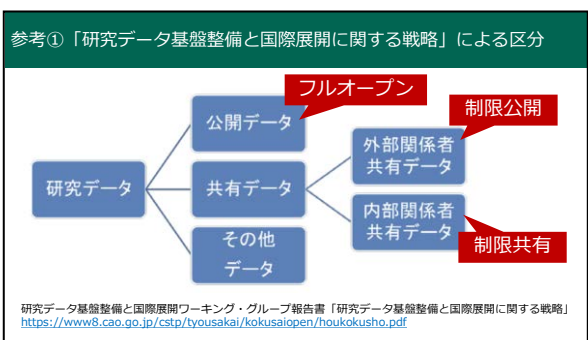
(図 5)

例えば、内閣府の「研究データ基盤整備と国際展開に関する戦略」では、公開データ・共有データ・その他と分けています（図 7）。公開についてはフルオープンです。共有データの中に外部関係者にも共有するものがあり、これは本セッションでいう制限公開に当たります。共有データの中でも内部の者にしか共有しないものが制限共有です。

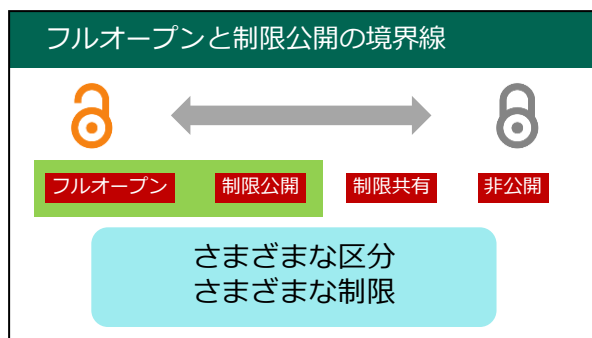
図 8 は国立研究開発法人日本医療研究開発機構（AMED）による区分です。公開と共有は、公的データベースに載せるか、一機関のデータベースに載せるかで線を引いています。このように、いろいろな線の引き方があるのが現状です。

フルオープンと制限公開の課題

まとめとして、フルオープンと制限公開の課題について3点挙げさせていただきます。一つは、フルオープンと制限公開を分けるときに、条約や法律などのハードローによる制約もあれば、分野ごとのソフトロー（規範や慣習、文化）もあります。これが十分に醸成されている分野もあるものの、そうでない分野も非常



(図 7)



(図 6)

参考② AMEDによる区分

公開レベル	手続き	リポジトリ
公開	不要	公的データベース
制限公開	アクセス申請→承認	
制限共有	アクセス申請→承認	自機関データベース等
非公開	(なし)	

ゲノム医療実現のためのデータシェアリングポリシー (AMED)
<https://www.amed.go.jp/content/000060867.pdf>

(図 8)

に多く、まだ混沌としています。ここの線引きをどうするかが大きな課題です。

二つ目は、制限公開に対するアクセス申請をして、OKであれば公開というときに、その妥当性をどう判断するかです。この線引きの難しさもあります。

三つ目は、制限公開の今後の可能性です。どのような方法があるのかというところまで、本日の議論で踏み込めればと考えています。