

第3回 SPARC Japan セミナー2020

「初めての研究データ」

研究データ公開実践のための課題を探る： 北海道大学での実例を通じて

三上 絢子

(北海道大学附属図書館)

講演要旨



国内の大学において、研究データ公開を支援するための体制は未だ十分に整備されていない。研究者の意向を反映した形での支援体制を構築する上では研究データ公開の実践事例の蓄積と分析が必要と考えられる反面、研究データ公開希望に対する国内大学での実践事例はこれまでほとんど共有されてこなかった。本講演では、2017年から2020年まで北海道大学の機関リポジトリ（HUSCAP）上でデータを公開した14件の事例について、研究者がデータ公開を希望した背景とデータ公開時に機関リポジトリ担当部署側が行った対応を紹介する。また、これまでの研究データ公開に関する議論を踏まえてこれらの事例を分析することで、現在の機関リポジトリの運用体制では対応が不十分な点やこれから改善すべき点を発見する。



三上 絢子

2016年より北海道大学附属図書館勤務、現在は研究支援課研究支援企画担当係員。2020年度オープンアクセスリポジトリ推進協会（JPCOAR）研究データ作業部会員。2019年度国立情報学研究所実務研修において研究データ管理・公開業務に携わる。2019年より現在まで北海道大学の機関リポジトリ業務を担当。

自己紹介

私は北海道大学附属図書館の職員です。研究支援企画担当ですが、3年ほど前まではシステム担当という名前でした。その名のとおり、主な業務は図書館の各種オンラインサービスの管理です。このサービスの中には、図書館業務に必要なオンラインカタログや、電子ジャーナルに学外からアクセスするためのシステム、その他、機関リポジトリも含まれており、実質的に機関リポジトリ担当ということになります。今回はこの機関リポジトリ上での実践を中心にお話しします。

北海道大学学術成果コレクション（HUSCAP）の概要

北海道大学学術成果コレクション（HUSCAP）は、

学術成果公開を目的とした機関リポジトリです（図1）。運用は2005年からで、当時から機関リポジトリ構築用のオープンソースソフトウェア DSpace をベースに、機能を継ぎ足して運営しています。主な公開資料は北海道大学所属の研究者の論文や学会発表資料、北海道大学発行の紀要ですが、ここ数年はごく少数な

北海道大学学術成果コレクション(HUSCAP)の概要

66,910件の学術資料を公開する機関リポジトリ(2021/02/04現在)
オープンソースソフトウェア「DSpace」をベースに構築している(運用は2005年から)
主な公開資料:
・北大所属の研究者から公開依頼のあった資料(論文、教育資料、学会発表資料など)
・北大発行の紀要
・研究データ(少数)

北海道大学

(図1)

がらも研究データを公開する事例が出てきています。

研究データ公開事例の概要

2017年から2020年までにHUSCAP上で行われた研究データの公開事例を紹介します。今回の事例紹介で扱う研究データの範囲は、HUSCAP上で公開された資料のうち、メタデータの資料種別がdatasetのものに限定しています。実際のデータの形式としては、数値データを集めたExcelファイルや画像ファイルが大半でした。対象の範囲は2017～2020年で、この間に14件の依頼がありました。これはファイル数ではなく依頼件数ベースでカウントしています。そのうち北海道大学発行の紀要が1件で、残りの13件は雑誌投稿論文に関連したデータです。ちなみに、2016年以前には公開したデータセットはありませんでした。

データ公開は北海道大学所属の研究者の依頼により行っています。図2の左側に依頼者の所属部局、右側に、部局ごとにデータに関連する論文が投稿された雑誌名をまとめました。今回の場合は、やや分野に偏りが見られ、大まかに言えば地球科学分野や医学分野が多かったことが分かります。

また、今回の実例では、研究データ公開が依頼される時期が、データに関連する論文が出版される前と出版された後の二つに大きく分けられました(図3)。件数としては、出版前に依頼された件数が若干多く、特に低温科学研究所所属の研究者の依頼は出版前、医学系所属の研究者の依頼は出版後が多い傾向にありました。出版前に公開する理由としては、論文のアクセプト条件として事前に研究データを公開するという要

件が課せられていたからというものがほとんどです。出版後に公開したものは、既に他の場所で公開されているオープンアクセスの論文をアーカイブする際に、関連して公開されていた研究データも一緒にアーカイブしたものです。

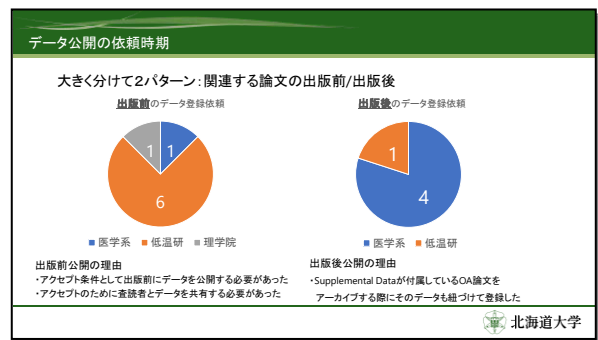
次に、これらの研究データの公開依頼に対して実際に行った作業を紹介します。ここでは公開事例をデータ公開のタイミングごとに分けて順に解説していきます。

パターン1：投稿前の論文の根拠データを公開

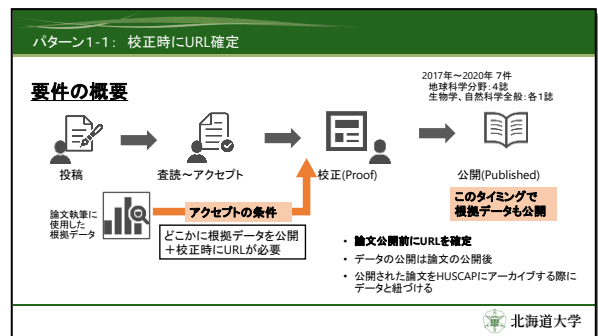
パターン1-1は、校正時に研究データのURLを確定させる必要があったケースです(図4)。実際に研究データをオープンにしたのは、論文投稿の流れ図の右側、最後の論文公開時ですが、論文誌面上にURLを掲載する必要上、校正の段階で事前にURLだけ確定させておきたいというケースでした。これが今回の事例の中では最多で、5誌7件の依頼が該当しました。ポイントとしては、論文の公開前にURLを確定させておいて、論文が公開されるタイミングでデータを公

依頼元の研究者が所属する部局		投稿雑誌
部局	件数	
低温科学研究所	7	Journal of Geophysical Research: Atmospheres Mammal study
医学研究院	3	低温科学 研究所 Earth and Planetary Science Letters
その他医系部局	2	PNAS Polar Science
理学院	1	医学研究院 Behavioural brain research
水産学部紀要	1	Scientific data Molecular Cell
		その他医系部局 The Journal of Clinical Endocrinology & Metabolism
		理学院 Tectonics
		水産学部紀要 海洋調査漁業試験要報

(図2)



(図3)

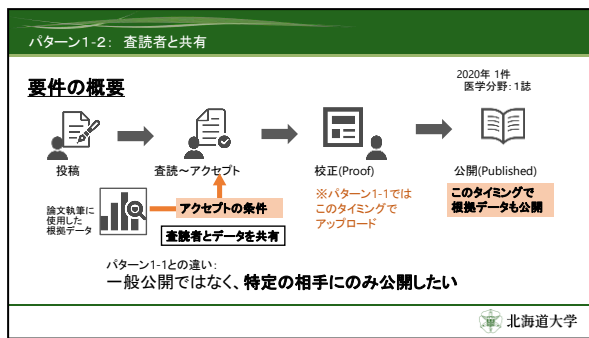


(図4)

開することと、論文の公開後にデータと論文とのひも付けを行うことです。機関リポジトリ側の作業としては、まず非公開状態でデータをアップロードし、この時点で URL を確定させました。その URL を校正用にエディターに送り、論文が公開された後に、研究者から公開された旨の連絡をもらってデータを一般に公開しました。公開の際に、研究データのメタデータに関連する論文の情報を入力しました。特に、データの公開は論文が公開された後に行いたいという要望が多かったため、データの公開前に URL を確定できる機能があったことは作業していて助かりました。

パターン 1-2 は、査読者とデータを共有したい場合です (図 5)。先ほどのパターンでは、流れ図の最後の公開までは完全にデータ非公開状態で構いませんでしたが、このパターンでは、一般公開よりも前の査読の段階で特定の相手にだけ公開したいという要件が入ってきます。これも、雑誌の規定でこれがアクセプト条件だったというものです。まずは先ほどのパターンと同じく、研究データを非公開状態でアップロードしましたが、HUSCAP とその大本になった DSpace の場合、特定の非公開アイテムだけが読み取り可能なアカウントを元々の機能として作成することができるので、査読者にこの研究データのアクセス権限付きのアカウント ID を渡しました。これは査読者と共有するところ以外は最初のパターンと同じで、論文が公開されたタイミングで依頼元の研究者から連絡してもらってデータを一般公開しました。

ここまでの 2 パターンは、論文が公開された後に研究データも一般公開しましたが、次のパターン 1-3 は、

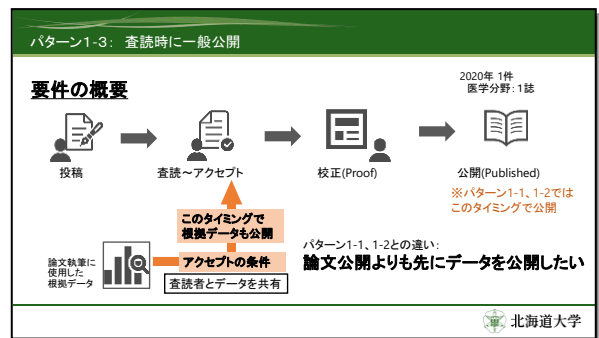


(図 5)

査読時にデータを一般公開したいケースです (図 6)。パターン 1-2 では、査読のタイミングでは査読者のみとの共有でしたが、今回は雑誌投稿規定により、データだけ先行して一般に公開する必要がありました。これは、システム的には非公開を使わずにそのまま公開すればいいので逆にシンプルなのですが、HUSCAP で公開する場合、システム以外の問題で一つ手続きが必要になりました。現在の HUSCAP の運用規定では、原則として公表済みの資料を公開することを前提としています。これは論文の剽窃の疑惑を避けるためや、機関として公式に公表できる資料かどうかについて、既に学術資料として一般に公表済みであるということが一つの判断基準として有用であるためという理由が挙げられます。従来の登録資料は公開済みの論文が大半なので、この基準で特段問題はありませんでした。しかし、今回のパターンでは、論文がまだ公表済みではないため、付随するデータも未公表資料と考えられます。これは規定を明文化した北海道大学学術成果コレクション資料提供・公開細則に資料の種別として明示されている「公表済資料」に当てはまらないため、このパターンの場合は「その他、特に館長の認めた資料」として都度、館長が決裁して掲載しています。ただ、このポリシーについては今後改定して、もっとスムーズに研究データを公開できるようにする予定です。

パターン 2 : 投稿後の論文に付随するデータを公開

パターン 2 は、既に公開されている論文のアーカイブに伴って付随データを公開する場合です (図 7)。



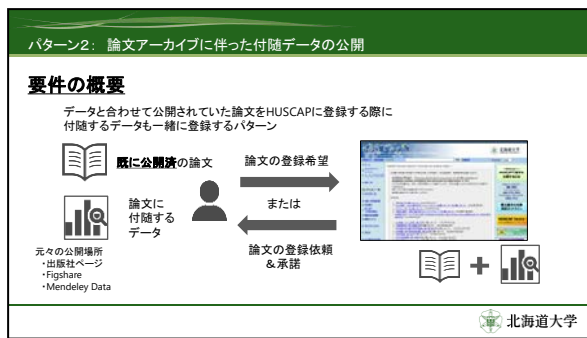
(図 6)

こちらは、データ公開の依頼時期は必然的に論文出版後になります。公表済みの資料をアーカイブするという点では、これまで機関リポジトリが行ってきた論文アーカイブ業務とほぼ同じです。従って、論文同様に研究データにもメタデータを付与してファイルを登録し、公開します。こちらは特段、非公開処理や運用規定などの判断が必要ないので、シンプルなパターンと言えます。

ちなみに、北海道大学の場合、論文アーカイブの経緯は主に二つあります。一つは、研究者から登録の依頼がある場合です。もう一つは、図書館が北海道大学の研究者が主要な著者となっている論文を定期的に検索し、依頼を出し、研究者から承諾を得てアーカイブをする場合です。

パターン1とパターン2に共通する対応

論文公開後の作業として、これまでのパターンの全てにおいて、メタデータを使って論文と研究データの関連付けを行っています。こちらのメタデータはJPCOARスキーマに沿っていて、その中で規定されているリレーション要素の属性 isRequiredBy で、元論文の DOI と研究データの関連付けを示すことができます。ここに入力することによって、出版社や学会のサイトで掲載されている論文へのリンクを貼ることができます。加えて、HUSCAP 内にも論文をアーカイブしている場合には、HUSCAP 内へのリンクを別に貼って関連付けています。



(図 7)

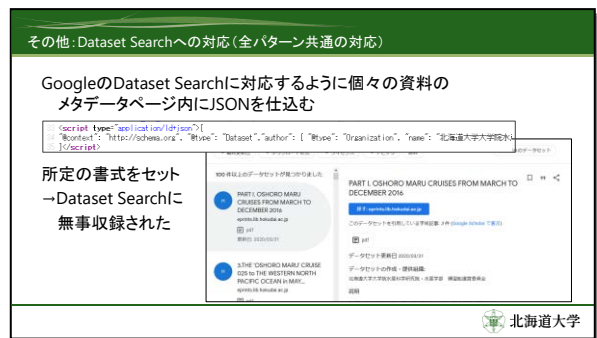
その他

その他に水産学部紀要「海洋調査漁業試験要報」を1冊データセットとして公開しています。元々冊子体で発行していたものをPDF形式でアーカイブしているもので、内容としては、北海道大学水産学部が所有する練習船「おしよる丸」から得られたデータを記録しています。

また、研究データを公開する際の作業として、GoogleのDataset Searchという検索エンジンに収録されるよう、メタデータページ内にJSON形式でデータを入れています。所定の形式でメタデータを入れておけば、Googleによってクロールされる際に、その研究データがDataset Searchに拾われて収録されます。これは、Google Scholarの場合にクロールしてもらったためのデータとも多少形式が違い、Dataset Searchに対応したものを新しくコードを書いて出力したものと聞いています。図8のとおり、北海道大学水産学部附属練習船「おしよる丸」のデータを掲載した紀要がDataset Searchに登録されていることがわかります。

研究データ公開例のまとめ

ここまでのまとめです(図9)。2017年からの4年間で14例のデータを公開しました。論文の根拠データの公開手順は、いつ登録・公開するか、誰と共有するかによって幾つかのパターンに分けられました。公開手順に関係する事柄としては、非公開状態やIDによる制限公開が思いのほか役に立ったこと、従来からの運用要綱に当てはまらないパターンもあること、デ



(図 8)

ータと論文を関連付けたり、検索エンジンに登録したりといった発展性を高める手法をいろいろと試してみたことなどが挙げられます。

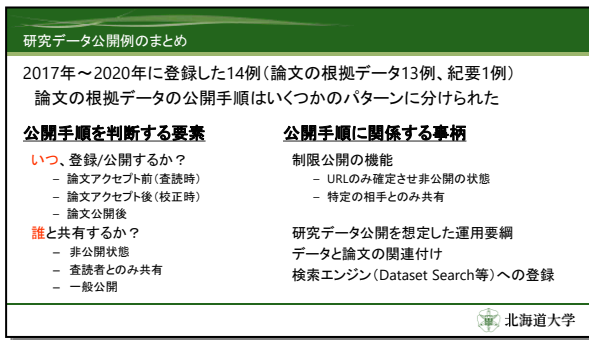
研究データ公開事例から考える今後の課題

ここからは、今回の事例を踏まえて、今後の研究データ公開の課題を考えていきます(図 10)。今回紹介した事例は、割と限られた範囲の実例であるとも言えます。一方で、望ましい研究データ公開支援の先行研究や議論の中では、もっと幅広い事例でのデータ公開を想定しています。今回の事例でカバーできていなかった範囲についての対応が今後の課題と言えます。例えば今回の事例における公開理由は、雑誌のアクセプト要件だからというものがほとんどでしたが、今後の研究データ活用推進においては、その他の公開理由も出てくると想定されます。多彩な公開理由を想定した公開基準の策定、また、今回は特に考えていなかった再利用を促進するドキュメントの整備や、公開場所としての機関リポジトリの位置付けも、今後検討すべき課題と言えます。これらの課題について、一つずつ詳

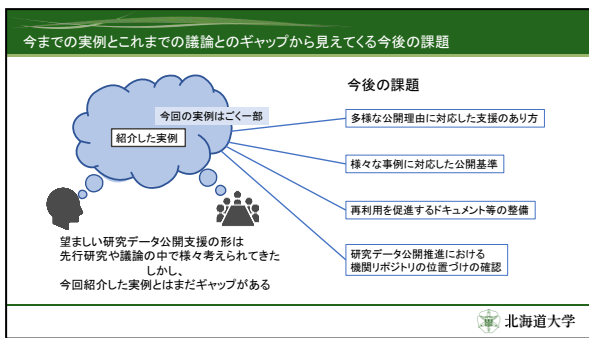
しく見ていきます。

まずは公開理由についてです(図 11)。先にも述べたとおり、今回の事例では、ほぼ全てのデータの公開理由が、雑誌の投稿規定にあるからというものでした。一方で、2017年に科学技術・学術政策研究所(NISTEP)科学技術予測センターが行った実態調査の結果によると、データ公開を行った経験のある研究者の公開理由としては、「雑誌のポリシーだから」が全体の中で2番目に多かったです。一番多かった理由は「研究結果を認知してもらいたいから」で、その他に「科学研究や成果実装を推進したいから」「他の研究者のリクエストに応じて」などが続きます。このように、雑誌の投稿規定として義務化されているからというだけでなく、研究者側のモチベーションによってデータ公開を希望する事例も想定されます。この場合の疑問としては、そういった理由やモチベーションに対応したリポジトリの運用方針や公開基準がまだ明確ではないのではないか、それから、今回の事例では雑誌の投稿要件を満たしていればそれでよかったです。が、その他の理由で公開を希望する研究者については、また違った要件があるのではないかとあります。

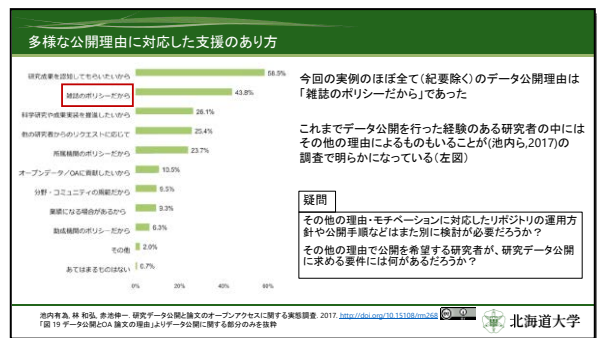
続いて、さまざまな事例に対応した公開基準についてです(図 12)。今回の事例では、全て論文に関連するデータで早急に公開する必要があり、なおかつ、個人情報や知的財産権などの問題がないと考えられたため、データの公開可否を特段問わず登録・公開しました。しかし、研究データの中には、公開に対して何かしらの問題が出てくるものも予想されます。例えば、



(図 9)



(図 10)



(図 11)

著者名が明記される論文と違って、データの場合、特にデータが単体の場合は、権利表記において誰がクレジットされるべきか明確に答えが出ません。また、共同研究者との間で契約上の問題がある場合や、分野内で別途ガイドラインが作成されている場合、個人情報や特許権などの法律的な判断が必要な場合も想定されます。

以前、私はNIIの実務研修において、公開可否判断ワークフローというものをつくったことがあり、このような問題についてどういった順番で判断していけばよいかということフローチャート図にまとめました。しかし、実務において、このフローチャート図は誰か1人、あるいは単独部署でたどれるものではないことに気がきました。権利表記や共同研究者との合意は研究の関係者しか知り得ない情報ですし、例えばデータ公開が特許申請の何かしらの妨げになり得るかどうかの判断には、知的財産権の専門知識が必要になります。このように、公開に際しての問題を誰が判断し、解消するかについては、図書館などの主な支援部署単独では難しく、研究者や関連部署との連携が必要と言えます。

また、そのときに、研究データ公開に取り組む意義や問題意識をどうやって共有し、協働するかについては、現状集まっている事例だけ考えることは難しいと言えます。これは、複数の利害関係者のそれぞれの意見を聞きながら進めていくべき事項だと思います。

また、今回の公開事例では研究データ公開に際して特別な手順は定めておらず、通常、論文を公開するときと同様にメタデータを付与し、ファイルをリポジ

リで公開していました(図13)。一方で、これまでの研究データ公開の議論の中では、公開時に考慮すべき点がいちいち挙げられています。例えば、国内のJPCOAR 研究データタスクフォース作成の「RDM トレーニングツール」によれば、ファイルの命名規則、ドキュメントの整備、再利用時のライセンスの明示なども考慮すべき点ですが、これらは今回の事例では特に行っていませんでした。理由としては、従来の論文の公開手順をほぼ準用したからということと、単純に出版社の定めた要件、すなわちファイルの公開以上のことは必要なかったからという2点が挙げられます。

ここで考えるべきことは、従来の公開手順では考慮できていない朱色で示した点について、これからどのように対応していくかということです。また、出版社の定めるファイル公開以上の要件、例えば他の研究者に再利用してほしいといった研究データ公開の要望が今後起き得るかどうか、起き得るとすれば、そこにどのようなモチベーションがあり、それに対してどういう公開手順で対応することが望ましいかを考える必要があると思います。

最後に、今回は北海道大学の機関リポジトリ HUSCAP 上での公開事例のみを取り上げましたが、実際には、その他にも、例えば分野別のリポジトリや、Figshare、Zenodo などの汎用リポジトリが既に存在します。今回の事例で、HUSCAP でのデータ公開を選んだ事情が分かるケースについて図14に幾つか示しています。理由としては、既存のリポジトリが第1候補だったが制約を満たすことが難しく HUSCAP を選んだケース、既に他の公開場所で公開されているデー

様々な事例に対応した公開基準

今回の事例の場合：
特段、データの公開可否の判断をしていない
(明確な判断基準も現時点では存在しない)

公開可否判断ワークフロー例：
三上 幹子 研究データ公開ワークフロー、
(2019年度NII実務研修成果集、一部抜粋)
<https://nii.ac.jp/sumino/2019/>

データによっては公開の際に
確認すべき問題があることが予想される

- 権利表記の合意(誰を作成者とするか?等)
- 共同研究者との契約上、守秘義務が生じる場合
- ゲノム情報など分野ごとのガイドラインが存在する場合
- 個人情報を含む場合
- 知財関連(特許等)が絡む場合

判断に専門知識を必要とする問題もある

疑問
誰が問題を判断+解消すべきか?
支援部署だけの対応は難しい?
研究者へあらかじめ注意すべき点を通知?
専門部署(知財等)へ事前に対応可能性を協議?
複数の利害関係者間で問題意識を共有し、協働できるか?

北海道大学

(図 12)

再利用を促進するドキュメント等の整備

今回の公開事例では特別な公開手順は定めていない(論文同様、提供ファイルをリポジトリで公開)

- ファイルの命名規則は特でない(論文公開同様)
- データの内容を説明するドキュメントは特でない
- 論文のAcceptに必要→出版社の定めた要件を満たしていればよい
- データの出身については研究者に任せている
- 再利用については考慮していない

研究データ公開で考慮すべき点 (JPCOAR研究データタスクフォース作成「RDM-トレーニングツール」(2017)より)

- ファイルのフォーマット
- ファイルの組織化(命名規則)
- 再利用のためのドキュメントの整備
- メタデータの付与
- 発見性やアクセス性の向上
- 再利用時のライセンスの明示

※現時点で考慮できていないものを朱色で表記

疑問
従来の公開手順では考慮できない点に対し、どのように対応していくか?
出版社等の要件以上のこと、例えば「再利用の利便性を高めるため、ドキュメント整備を行う」などの場合の研究者/支援者のモチベーションは何か?
どういった手順でそのモチベーションに対応できるか?

JPCOAR研究データタスクフォース RDM-トレーニングツール、2017. <http://rdm.nii.ac.jp/1458/00000002/>

北海道大学

(図 13)

タをアーカイブしたケース、図書館で研究データ公開の相談を受けた流れで機関リポジトリでの公開を選んだケースがありました。

ここで一つ疑問となるのが、研究データ公開における機関リポジトリの位置付けです。論文公開においては、機関リポジトリはオープンアクセスの一つのGreen OA、すなわち許諾の得られた投稿済み論文をアーカイブするところで、長期保存が見込まれ、比較的出版社から許諾の出ることが多い非商用リポジトリという位置付けがありました。しかし、研究データ公開においては、そういった論文のオープンアクセスの事情とは異なるので、機関リポジトリに求められるものが何なのか、あるいは、目的に応じてその他の分野別のリポジトリや汎用リポジトリの活用も視野に入れて支援すべきかということがあります。

まとめ

今回の講演のまとめです（図 15）。前半で事例紹介を行い、後半で今後の課題を挙げました。ここで挙げた課題は支援部署（図書館や URA）が単独で解消できるのではなく、研究データ公開の主体である研究者の意見や、各研究機関の体制や方針等も併せて考えていくべき難しい問題だと思います。今回の実践から見えた今後考慮すべき事柄はたくさんありますが、ひとまずは、今立っている場所を確認することが、望ましい研究データ公開の実践の足掛かりになればと思います。

緑色の丸い小鳥は「はすかつぶちゃん」といって、北海道大学機関リポジトリ HUSCAP のマスコットキ

ャクターです。今後ははすかつぶちゃんと共に、オープンサイエンスの取り組みに注力していきたいと思っています。

●八塚 質問が幾つか寄せられています。一つ目は、「北海道大学で公開された研究データを使って2次分析を行った場合に出てくる研究成果を、リポジトリ側ではどのように扱う予定でしょうか」というものです。

●三上 今回については、特段ライセンス等は付与していませんが、例えば研究者がクリエイティブ・コモンズ・ライセンスで公開したいという場合はそれに従います。ご質問は、明示していない場合の対応でしょうか。

●八塚 恐らくそうだと思います。

●三上 その場合は、論文の転載もそうですが、いったん機関リポジトリ担当に問い合わせさせていただいて、そこからデータの権利を持っている先生に問い合わせ、それにお答えするという形になると思います。

●八塚 続いて、「今回ご紹介いただいた事例には見当たりませんが、例えば外部資金（科研費など）の募集要項においては、公開を推奨しているものが多いと思います。そういった理由が見当たらないというのは、もしかすると、研究者側の公開意識が低いので

(図 14)

(図 15)

はないでしょうか」という質問です。

●三上 確かに今回は、論文の投稿段階で公開したいという事例がほとんどでした。ただ、既に研究データが公開されていたもののアーカイブを行った事例もあり、機関全体で本当に外部資金で公開していないかどうかは分からないところです。機関リポジトリ担当も、相談を受けた分の研究データ公開の事情は把握していますが、それ以外のことは把握していません。実際に研究者がどう考えているかというのは、図書館にとってもなかなか把握が難しいところだと考えています。

●八塚 次の質問です。「研究者にとってはリポジトリが各種存在していて、どれを選択するのが適切なのか悩む場面もあると思います。北海道大学では、図書館を主体とした、研究データに対する相談窓口があるのでしょうか」。

●三上 まだ明示的に研究データに関して図書館にご相談くださいとアナウンスしているわけではありませんが、事実上、機関リポジトリに関連して、図書館の機関リポジトリ担当に相談が来るケースが多くなっています。その中では、機関リポジトリに公開したいと初めから決まっているわけではなく、どこからのリポジトリに公開したいという要望もあります。分野別のリポジトリまでは図書館の担当はあまり把握していませんが、一般的な Figshare、あるいは HUSCAP での公開も可能だという案内はしており、結局 HUSCAP で公開したケースはありました。

●八塚 次に、「三上さんのご経験から、データ公開を支援する部署は何人くらいの人員であることが理想的とお考えでしょうか」という質問を頂いています。

●三上 今、機関リポジトリ担当は4人で行っています。今後、研究データの取り扱いが増えてくるとして、今の機関リポジトリ担当と仕事が被る点もあるので、

図書館内であれば恐らく現状か、研究データの判断の専門人員が要るでしょうか。

●八塚 現状は4人で足りているのですか。それとも人手不足の状態なのですか。

●三上 正確に言うと、4人と、他に部局ごとの図書担当が対応しています。業務内容にもよるので、今の時点で考えるのはなかなか難しいところです。

●八塚 ありがとうございます。次の質問です。「先ほどご説明いただいた中で、パターン 1-3 の査読時に一般公開する場合において、もし論文がリジェクトされてしまった場合は、データもリポジトリから取り下げるとことはあり得るのでしょうか」。

●三上 リジェクトの場合だったか分かりませんが、過去に1度、実際にリポジトリから公開請求を取り下げたものがあります。北海道大学のリポジトリでは、研究データは1度公開したら取り下げ不能という規定は特に設けていないので、先生から取り下げてほしいという要望があれば、そのように対応すると思います。