

第2回 SPARC Japan セミナー2016

「研究データオープン化推進に向けて：インセンティブとデータマネジメント」

超高層大気観測データの メタデータ作成実験経過報告

南山 泰之

(国立極地研究所)

講演要旨



本報告では、研究者と図書館員における協働の一事例として、超高層大気分野における観測データのメタデータ作成実験の経過報告を行う。本実験は、大学図書館員の持つメタデータ運用スキルが、特定分野にフォーカスしたメタデータの作成に応用できるか、の検証を通して、①メタデータ作成支援による研究者の負担軽減、②図書館員による流通面での知見提供、といった具体的なインセンティブ付与、及び③図書館員による研究データマネジメントの今後の可能性を探るものである。幅広い関係者からのコメントをいただくことで、今後の両者の協働のあり方を考えるための一助としたい。



南山 泰之

2005年より国立極地研究所情報図書室に勤務。2007年～2008年、第49次日本南極地域観測隊に参加。その後、東京大学駒場図書館（2011年～2014年）を経て現職に戻る。2014年8月より機関リポジトリ推進委員会協力員。2014年第4回SPARC Japanセミナー企画ワーキンググループ（WG）メンバー、2016年SPARC Japanセミナー企画WGメンバー。

私は 2005 年に国立極地研究所に配属になり、2007 年から 2008 年にかけて、南極に観測隊として参加していました。帰ってきてしばらく極地研究所にいたのですが、その後、東大の駒場図書館に少しお世話になり、2014 年から極地研究所に戻ってきて、また仕事をしています。今は機関リポジトリ推進委員会をメインに活動しています。他に、図書館系の雑誌の編集や SPARC Japan の企画ワーキングを仰せつかっています。

機関リポジトリ推進委員会は、オープンアクセスリポジトリ推進協会と 2016 年 7 月から名前が変わりつつあるようですが、大学図書館、国立大学、公立大学、私立大学が連携した連携・協力推進会議が親組織です。大学図書館コンソーシアム連合や「これからの学術情

報システム構築検討委員会」も連携・協力推進会議を親組織としています。私はオープンアクセスリポジトリ推進協会の中で、メタデータや、研究データに図書館がどう関わるか、を検討するタスクフォースに参加しています。

研究データ管理に関わる背景

国際的動向を踏まえたオープンサイエンスに関する検討会から、報告書「我が国におけるオープンサイエンス推進のあり方について」が公表されてもう 1 年以上たちます（図 1）。ここから「オープンサイエンス」という単語がメジャーになり、図書館は何をしようかという話が始まったと言ってもいいと思います。

今回の主題であるデータの話は、「公的研究資金による研究成果」の部分で少し触れられています。図書館の役割については、「研究成果等の収集、オープンアクセスの推進、共有されるデータの保存・管理を行う基本機能」を持つべきと言われたのか、持つとありがたいと言われたのかは定かではありませんが、そのように書かれていました。

これを受けて、学術情報委員会で審議のまとめが出されました。そこから大学図書館の役割を幾つか抜粋して図2に書いてあります。論文のオープン化、研究データのオープン化、研究成果の散逸等の防止、人材育成です。今回、私の発表内容に関係するところしか赤字にしていますが、研究データのオープン化に関しては、論文のエビデンスとしての研究データの公開に機関リポジトリを活用するようというようことが書かれています。

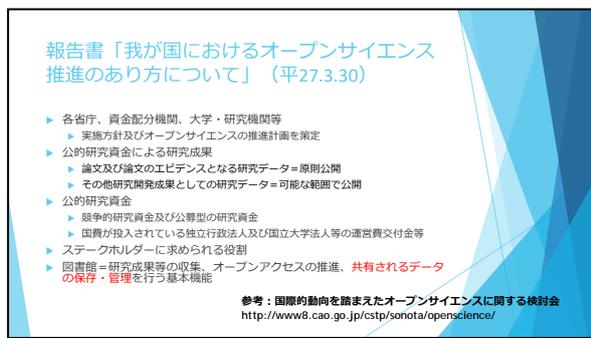
人材育成については、技術職員、URA および大学図書館職員を中心にデータ管理体制を構築すること、あるいは、機関リポジトリの構築を進めてきた経験等から研究成果の利活用促進を担うことが役割として挙

げられています。

大学図書館はこれらの潮流を受けて、どのような動きをしているのかというと、2016年6月に国立大学図書館協会の総会があり、そこで「国立大学図書館協会ビジョン2020」が策定されました。その中で重点領域とされたのが、重点領域1「知の共有：＜蔵書＞を超えた知識や情報の共有」、重点領域2「知の創出：新たな知を紡ぐ＜場＞の提供」、重点領域3「新しい人材：知の共有・創出のための＜人材＞の構築」の三つです。今回のデータの話に絡むのは、特に重点領域1と重点領域3です。

重点領域1では、学習教材やデータといった教育研究成果を対象として、知の共有のための方策を検討し、実現することが求められています（図3）。

重点領域3には、これまで培ってきた学術資料に関する専門的知識やメタデータ運用スキルに加え、新たな知識やスキルを習得することによって、国立大学図書館に期待される新たな機能を実現することが書かれています。

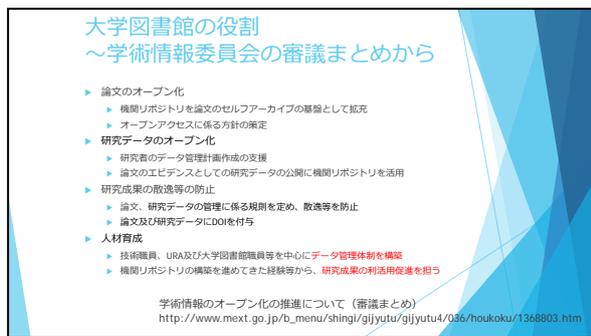


(図1)

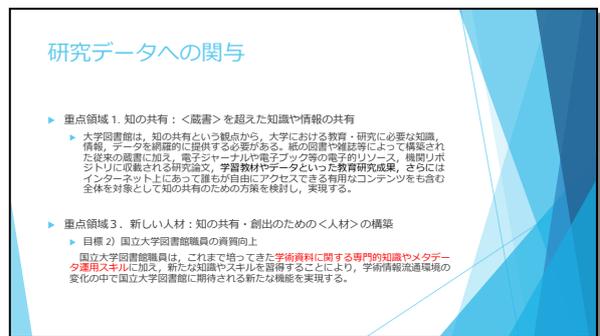
図書館ができること

ここまでの内容から、「図書館ができること」として、図書館の外に対して何をアピールできるか、私なりにまとめました（図4）。

一つが、ネットワークの活用です。図書館は横のネットワークが非常に強いのです。委員会をつくって、タスクごとにその都度、最新のノウハウを共有し、検討してきたという経緯があります。図書館員はジョブローテーションがあり、3～5年ごとに異動して違う



(図2)



(図3)

仕事を始めますが、そのような人事異動にも、委員会活動での知の共有があるので対応できるようになっています。

もう一つはメタデータ運用です。図書館員はメタデータが得意な人が多いです。日本はリポジトリ大国とよく紹介されますが、9月末現在のデータを見ると、(構築中のものも含めれば) インスタンスを全部で約744 持っていて、公に出ているデータとしては世界一リポジトリを持っている、すなわち、それを運用する図書館員は、一般的にメタデータに強いと言えると思います。

(日本のリポジトリ業務における)メタデータ運用で使われるスキーマは、junii2、Learning Object Metadata、LIDO などいろいろありますが、標準的な図書館のリポジトリで使われているのは junii2 です。LIDO は博物館関係のスキーマです。余談ですが、最近のアップデートでは今回お話しする SPASE というスキーマを、なぜか JAIRO Cloud に搭載してくださったので、非常にお話ししやすく助かります。何が言いたいかと言えば、図書館にはこのように複数のスキーマを横断的に扱う人材がたくさんいます。(スキーマを横断的に扱うことが)業務の一環となっているので、メタデータ運用スキルは高いと言えるように思います。昔ながらの目録もスキーマの一つと言ってもいいですね。

実験の概要

ここまでは前振りで、次にわれわれが行った、図書館員による IUGONET メタデータ作成(超高層大気観測データのメタデータ作成)の実験についてお話し



(図 4)

します。概要と問題意識、詳細、ここまでの検証についてお話しします。

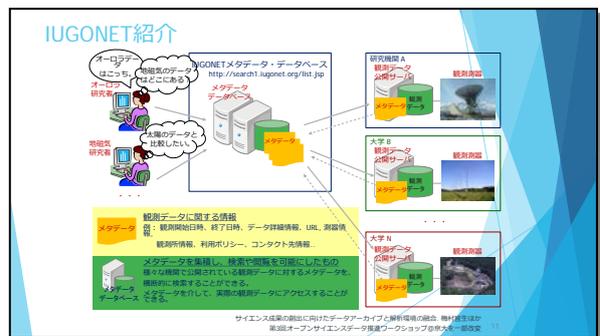
一つ目は概要です。今回の実験を通じて探りたかったことの一つは、研究者へのインセンティブの可能性です。研究者、特にメタデータ作成業務従事者の負担軽減、図書館が持つ流通面での知見の提供ができるかをこの取り組みを通じて考えたかったのです。

もう一つは、図書館による新規サービス展開の可能性です。図書館が研究支援の機能を持ってほしい、海外の機関と比べてこのようなことをやってほしいという声はいろいろなところで聞くのですが、実際にそれにどのように応えていくのかを探りたい、研究者との協同による直接的な研究支援を考えたいと思っていました。従って、オープン化というより、メタデータを通じて、研究データのありかが分かるようにして、オープン化に寄与しようという取り組みになります。

IUGONET 紹介

IUGONET は正式名称が Inter-university Upper atmosphere Global Observation NETwork です(図 5)。私が所属している国立極地研究所、それと東北大学、名古屋大学、京都大学、九州大学の 5 機関が連携して、観測データからメタデータを抽出し、それをネットワーク上で広く共有するシステムを構築するプロジェクトです。

IUGONET は 6 年以上進んでおり、やはりやってくうちに課題が挙がってきます。2015 年の SPARC Japan セミナーで、極地研究所の田中良昌先生が現状を発表されました。図 6 はそのときのスライドです。



(図 5)

問題の一つは、データベースが個人レベルでのメンテナンスにどうしても依存してしまうので、研究者のモチベーションが尽きたら終わってしまう面があることです。

もう一つは、ドメインの研究者がデータベース構築、管理をしなければいけないことです。研究者は専門分野であるドメインの研究をしたいのですが、データベースの構築はドメインとはやや離れます。研究者は素人ではないので、ご自身で何とかしてしまうのですが、負担が大きいのは確かです。

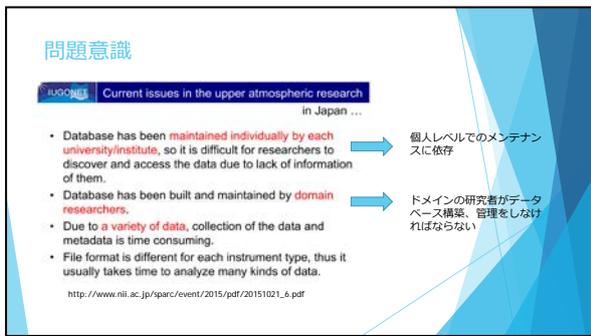
そこで、個人レベルのメンテナンスに依存するのであれば、図書館のネットワークを利用できるのではないかと。ドメインの研究者が構築・管理をしなければいけないのであれば、(サーバーの運用管理の話は少し置いておいて) 少なくともメタデータ運用に関しては図書館側にこれまでの蓄積があるので、スキルをうまく共有できるのではないかと考えました。これが今回の実験につながっています。

IUGONET で使っているメタデータスキームは SPASE というものです (図 7)。なぜか JAIRO Cloud

に搭載されています。SPASE は NASA やアメリカの研究機関から成るコンソーシアムで作成しているメタデータスキームで、太陽・惑星間空間・地球地磁気圏の人工衛星観測データを念頭に置いたメタデータフォーマットです。ご存じない方がほとんどだと思うのですが、私も半年前まではあまり分かっていなかったのですが、ここではこのようなものだと見ていただければと思います。

2016 年 6 月からパイロットを開始しました (図 8)。図書館職員とのコラボレーションとして行って内容を決め、その難易度も設定して行っています。現在は 3 番まで進んでいますが、重要な点として、私個人ができてあまり意味がなく、水平展開をいらいらと、あくまでも図書館のベースの知識の中でどこまでできるのかを探るのが今回の取り組みのメインです。図書館コミュニティの中で共有されていると思われる知見で、できる範囲を見極めようとしています。

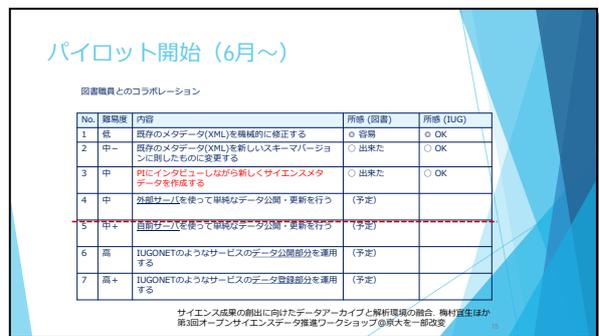
赤い線は業務境界線で、今は 4 番と 5 番の間に引いています。5 番以下は、国内の図書館の中で行っている例、自前サーバーを使って公開したりする例もありませんのですが、図書館員の業務負担が非常に大きくなります。また、機関リポジトリとの競合の問題もありますし、自前サーバーではなくクラウドにしたらどうかという話が出てくるので、共通化するのはこのあたりまでが一番いいのではないかと IUGONET の担当者と話しています。



(図 6)



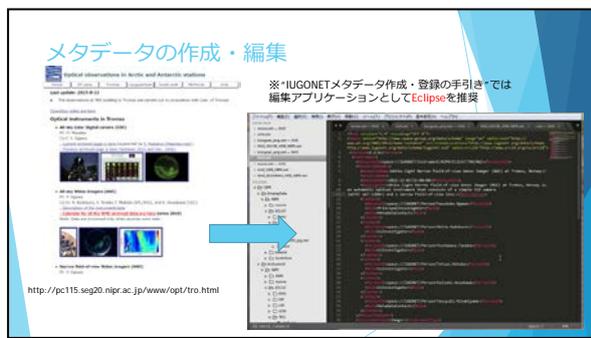
(図 7)



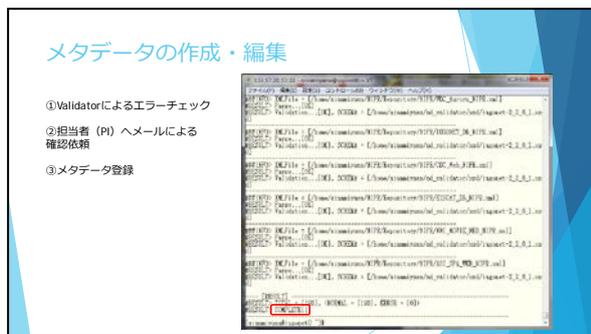
(図 8)

実際の作業はどのようなことをやっているかということで、メタデータの作成・編集です(図9)。今、メインに行っている作業ですが、(前提として)ウェブ上に既に研究者が情報を載せています。「このようなことをやっています」というようなページをつくって、そこにデータを置いているのですが、それをスキーマに落とし込む作業をしています。IUGONETのメタデータ作成・登録の手引きは、Eclipseというソフトの使用を推奨していますが、私は自分で使いやすいテキストエディタでXMLを直接編集してメタデータをつくっています。情報源をメタデータに起こすという作業は、通常の図書館の作業(目録や、リポジトリのメタデータ登録)に近いので、そのイメージで見ただけだと思います。マニュアルがあるほか、他のデータベースにも(同じデータに対する)簡単なメタデータが登録されているので、そのようなものも参照しながらやっています。

Validatorという、XML Schemaの構造にきちんと合っているかをチェックするソフトがあるので、そちらに掛けて問題がなければ、「このメタデータで本当に



(図9)



(図10)

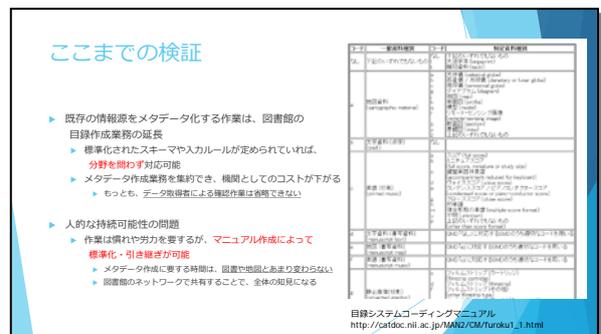
よろしいですか」と担当のPIにメールで確認します(図10)。当然Validatorを通ったといっても、あまり薄々のデータでは使い勝手が良くないというか、最終的にきちんと確認してもらわなければいけないこともあるので、それを確認していただきます。それで問題なければ、データベースへのメタデータの登録に進みます。

実際の登録は幾つかしています。Keogram dataで検索すると、検索結果として私が元を作成したメタデータが出てきます。対象のメタデータの中には、私の名前もメタデータ・コンタクト・パーソンとして入れていただいています。

ここまでの検証

メタデータをつくる、機械的に修正するなどの作業を幾つか10月までにやってきました。ここまでの検証として分かったことの一つは、既存の情報源をメタデータ化する作業は、図書館の目録作成業務の延長だということです(図11)。あえて「リポジトリ」と書かずに「目録」と書いたのは、リポジトリをやっているなくても、昔からそのような知識を持っている方、昔からそのようにやっている方であれば大体対応できるだろうと思っているからです。

標準化されたスキーマや入カールールが、きちんと研究者との協定の段階で定められていれば、分野を問わず対応可能と言っていると思います。分野を問わずと書いたのは、私はもともと法律系の人間で、そもそも理系ですらないですが、(IUGONETのメタデータにも恐らく対応できているため)大体何とかなるという



(図11)

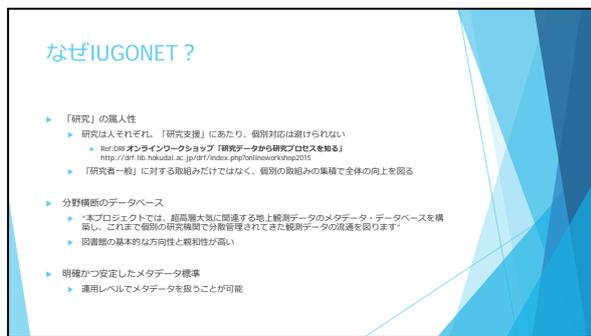
話をしたかったのです。右側に目録システムで取り扱ってきたコンテンツの表をあえて持ってきましたのも同じ意味ですが、地図・楽譜・静止画像など書籍ではないものであっても、最低限のレベルのメタデータは何か作成できるということが言いたかったからです。

(図書館が研究データのメタデータ作成に関与できる、という前提からは) 研究者が実際にやっている、あるいは研究者が非常勤の方を雇って、その方にやってもらっているようなメタデータ作成業務を、図書館の基幹業務として集約することができることになるので、機関としてはコストが下がるというメリットが挙げられると思います。もっとも、データ取得者による確認作業は省略できません。

もう一つは、人的な持続可能性の問題についてです。作業を実際にやってみたところ、慣れや労力は要しますが、マニュアル作成によって標準化・引き継ぎは可能です。作業に時間がかかり過ぎたらあまり意味がないと言われるのですが、普通の目録、リポジトリのメタデータ登録の作業とほぼ変わりませんでした。慣れれば 30 分ぐらいで一つできます。この知見をこれから図書館ネットワークで共有しようと思っているので、これは図書館全体の知見として共有されます。すぐにとは確約できませんが、誰でもできる体制になってくると思われます。

研究支援についての展望

IUGONET の研究集会在 1 週間前にあり、そこで質問として、なぜ IUGONET に協力したのか、個別のプロジェクトに対して図書館員がサービスする意義、



(図 12)

スタンスは何かと聞かれたので、研究支援について私が考えることを図 12 に書き出してみました。この点、皆さまからもぜひご意見を頂きたいと思っています。

当然、プロジェクトに関わるに当たっては、その中の人とやりたいかどうか、楽しく仕事ができるかどうかが初めに来るのですが、それは置いておきます。

一つ目は研究の属人性です。昨年度、DRF でオンラインワークショップ「研究データから研究プロセスを知る」が開催され、私もファシリテーターとして関わりました。自分の担当でもそうでしたが、報告書を見ると、やはり研究者の研究のやり方はばらばらでした。ばらばらなのが研究ですから、それに文句を付けるつもりで書いたのではなく、研究支援業務を一般化するのはかなり難しいのではないかと感じました。

一般化できないのですから個別対応は避けられないというスタンスに立つと、(研究者一般に対する取り組みのやり方を今まで図書館は考えてきた面があるのですが) どこの分野に取り組むとか、この先生とタイアップするというような個別の取り組みを集積させることで、全体の向上を図るような取り組みに変わっていけばいいのではないかと、思って取り組みを進めています。

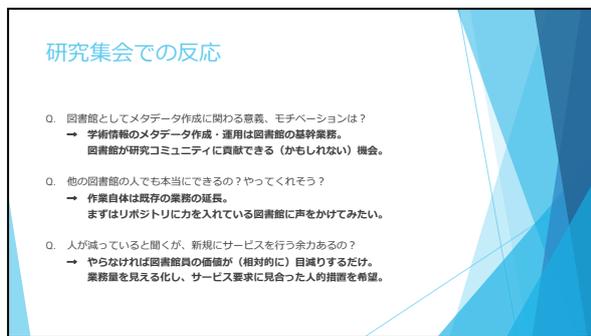
二つ目は分野横断のデータベースということです。IUGONET が分野横断のデータベースであったことが、やりやすさの一つでした。超高層物理という specific な分野ですが、その中でもさらに細かくデータの分野が分かれています。それを横断させるようなデータベースをつくっているということで、図書館は最終的な集合体として全部を大きく横断させるのが目的ですから、基本的な方向性として親和性が高く、図書館の業務としては触りやすいと感じていました。三つ目は明確かつ安定したメタデータ標準です。研究でメタデータを扱うときは、「このようなメタデータがあれば便利だ」という考えが研究を進めていく段階でどんどん出てくると思いますが、アイデアに一つ一つ図書館員の方で付き合っていくというのは、(運用担当として) なかなかやりづらい面があります。しかし、明確

に「これをフォーマットにしましょう」「当面これでやりましょう」というように定めていただくと、運用のレベルでメタデータを扱うことが可能になります。従って、図書館員にとってやりやすいです。

研究集会での質問と私の回答を図 13 にまとめました。「図書館としてメタデータ作成に関わる意義・モチベーションは何か」については、学術情報のメタデータ作成・運用はそもそも図書館の基盤業務であり、この仕事をきちんとしなければ、図書館は（オープンサイエンスの潮流の中で）最終的に何をやるのかという話になりかねないと思っています。もう一つ、前向きな回答としては、図書館が研究コミュニティに貢献できるかもしれない機会だからぜひ進めていきたいとお答えしました。

「他の図書館の人でも本当にできるか、やってくれそうか」については、正直、人によるという回答をしたと思います。ただ、各図書館のトップの方がどのように考えるかはともかく、作業自体は既存の業務の延長なので、やってできないことはありません。まずは、リポジトリに力を入れている図書館や、プロジェクトに関連する大学の図書館、具体的には IUGONET は京都大学、東北大学の方が多いので、そのあたりにぜひ声を掛けてみたいと思っています。

「人が減っていると聞くと、新規にサービスを行う余力があるのか」という質問については、初めの質問とも関連しますが、やらなければ図書館員は紙だけ扱うのかということになって、図書館員の価値が相対的に目減りするだけ、と考えています。業務量が多いから手を出さないという方向ではなく、業務量を見える



(図 13)

化して、このサービスが本当に必要なら、（業務量に応じて）人員を増やしてほしいという話に持っていきのが筋だと考えます、という話をしました。

今後の展開

IUGONET で今後やりたいことの一つ目は、運用レベルでの支援（の続き）です。スキーマのバージョンアップは今後も起こり得るので、それに対応したアップデートを行ったり、そのときについてメタデータクリーニング作業をしたり、あるいは新たにメタデータをリッチにしたりするような支援ができればと思っています。

二つ目は、水平展開です。これは特に重要です。複数人による安定的なメタデータ作成サービスでなければ、私が初めに提案したメリット（図書館のネットワーク活用）が薄れてしまうので、プロジェクト参加機関の図書館との連携を今後やっていきたいと思っています。

三つ目は、図書館側の知見の提供です。これは図書館側としては重要で、「協同」ですから、ただお手伝いするだけではなく、私の方からも何かしら提供しなければ、図書館側としてはあまり面白みがないのです。一番やりやすそうなのは機関リポジトリとの連携です。メタデータをハーベストして、図書館側のデータベース（IRDB）にデータを流して、CiNii などいろいろなところで検索できるようにしてみたりすることのほか、ライセンスに関する情報提供、識別子の話、あるいはデータベースをどのようなクラスターで検索させるか、検索しやすいためのファセット構造を考えるなどの話は、こちらから提案していければと思っています。

何にせよ、研究者の方々が、一番研究が進むやり方、あるいはやりやすいやり方を検討していければと考えています。