

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」

ディスカッション 「インセンティブとデータマネジメントの今後のあり方」



蔵川 圭	(国立情報学研究所)
仲里 猛留	(情報・システム研究機構ライフサイエンス統合データベースセンター)
下田 研一	(長崎大学附属図書館)
南山 泰之	(国立極地研究所)
青木 学聡	(京都大学情報環境機構)
武田 英明	(研究データ利活用協議会/国立情報学研究所)

●蔵川 先生方がいろいろなテーマで、研究データ共有について情報を提供してくださいました。仲里先生からは、バイオインフォマティクスの連綿と連なる歴史的なデータベースの共有、またそれにつながる研究データについて、下田様からは、研究データ共有やオープンサイエンスと言われる前からずっと続けられている古写真の活動について、南山様からは、図書館の新しい試みとしての研究データの登録について、青木先生からは、京都大学の研究データ共有に関する今後の取り組み、それにまつわる世界の活動について、武田先生からは、特に世界における研究データのトレンド、または国内の活動として、Research Data Alliance (RDA) あるいは研究データ利活用協議会についての今後の方向性について発表を頂きました。

今回のパネルディスカッションは、それらの情報をベースにしながら、われわれ日本における研究データのオープン化というものを、図書館員と研究者の協同という観点から今後どのように推進していくことができるかを考えてみるのが趣旨です。

研究データそのものをシェアする、共有するということは、広い観点からすると恐らく全員が得をするだろう、幸せになるだろう、イノベーションも生まれて、研究者だけでなく、人類全体が得をするだろうという総意の下で動きたい、動こうということで号令をかけて進もうとしているわけです。

個々の点からすると、単純に総意でもって研究データのオープン化に動いてしまったときには、不幸にも泣いてしまうとか、「それは私の利益とはつながらないのでやめます」ということがあったりするでしょう。そういうときに特に最初に問題になるのは、研究データそのものは研究者がつくっているのに、研究者や科学コミュニティに対するインセンティブはどういうものが存在するのか、インセンティブがそもそもあるのか、あるとすると何か、何が不足で、今こういう状態になっているのかを考えなければいけないということです。

よく言われることとしては、新たな知見や価値が生み出せるということが一つのコミュニティにとってのインセンティブですし、または、今はありませんが、研究データを出した者にオープン化の成果に見合った処遇を与えるということもインセンティブでしょう。このようにインセンティブがなければ、最初に生み出されるデータ共有の点というものが存在しないということになるので、実はこれは非常に大事だということが確認できます。

それを受けて、では世界の中で研究データマネジメントというのは、どういうものが必要で、今どのくらいできていて、実際に何かに困っているのであればこれからどういうものをつくればいいのかという話に展開すると思います。

このパネルディスカッションでは、主に二つのテーマを設けて、インセンティブには一体どういうものがあるのか、なぜそれが議論の点として必要なのかということ踏まえて議論をしていただきたいと思います。また、インセンティブを踏まえて、研究データマネジメントとしてどういうものが必要で、これからどういうものができるのか。そして、今回は SPARC Japan セミナーで、参加者の皆さまの中には図書館員の方も多いと思うので、図書館員としてそういう研究データマネジメントをするときに、研究者にどういうことが必要とされ、また自分たちはどういうことができるのかを明らかにしていきたいと思います。

まず、インセンティブですが、今回の企画は World Data Center for Geomagnetism, Kyoto (地磁気世界資料解析センター京都) で活動されている能勢先生が主導されました。研究データを実際につくっている側として、研究者にとって研究データの作成とは研究生活の中でどういう位置付けなのかということと、あまりデータをつくっていない人からするとよく分からないので、研究者のインセンティブとは何なのかということも含めてご紹介いただきたいと思います。能勢先生はパネリストではないですが、そこから始めたいと思います。

●**能勢** 私は京都大学理学研究科の地磁気世界資料解析センターという、地磁気のデータ、主に地球の磁場のデータをデータセンターとして扱っている部局にいます。大学の機関なので、研究と教育も大きなファクターを占めるのですが、それに加えて地磁気データを管理する、それからサービスをすることが一つの業務になっています。

そういう立場で、ご質問のありました研究データの作成についてですが、もちろんデータを作成したり、データを加工して他の研究者へ提供したりすることは喜びではあります。ただ、本業の教育・研究以上に時間が取られるので、使った時間、使ったコストに対して何らかの処遇を今後考えていけないのではないか、

それはわれわれだけではなくデータセンターを管理されている方は必ず持つておられる意識だと思います。

かつ、先ほど仲里先生からお話がありましたように、持続的にデータセンターを発展させていく、維持していくというのもやはり問題です。組織を維持していくためには人や費用が必要なので、何らかビジネスモデルと言うと少し大きくなりますが、そのようなものが考えられないのかというのが今回、このセミナーを企画させていただいた理由の一つです。

私が最初に趣旨説明で、馬の折り紙の写真をお見せしましたが、これが今、私が考えているインセンティブです (図 1)。ご講演を伺っていて感じたのは、局所的な成果に見合った処遇・研究費が得られるという即物的なものだけではなく、仲里先生から、データを公開することで、データが使われて世界を動かせるというような言葉がありましたが、まさにそのようなことが理想だと思います。そういう見方もあるのではないかと、もしそれでコミュニティとして動いていけるのであれば幸せではないかと感じました。

インセンティブとは一体何かというご質問もありました。研究者がなぜ論文を書くかということ、もちろん出てきた結果をきちんと世に問うという理由もあると思いますが、自分の業績になるから書くわけです。それと同様な理由で、データセンターを維持していく、データサービスを恒常的に維持していくために使った時間、コストに対して、何らかのインセンティブを考えていく必要があるのではないのでしょうか。



(図 1)

●**蔵川** そもそも研究データ作成は研究者にとってやりたくないことなのか、やりたいけれども人に任せたくないことなのか、いろいろな位置付けがありますよね。研究論文を人に書かせているようでは研究者ではないので自分で書きます、でも後進のペーパーをチェックしながら書いていますなど、いろいろあると思います。そういう見方からするとどうなのですか。

●**能勢** 人によって感じ方は大きく違うと思うのですが、例えば生物や医学など、自分で実際に手を動かして実験をされて取られたデータは、もちろんその方にしか分かりませんし、その方が自分のデータとして管理する、処理するのは普通のことだと思います。データセンターだと、それ以外の他人が取ったデータの管理もあるので、そういうものの管理はその方に問い合わせたりして、時間が取られるのではないかと思います。

●**蔵川** 今のは地磁気の例だと思いますので、そこからするとかなりルーチンワーク化されていて、むしろデータを取ることが研究の中では当たり前であって、あえて自分から新しいデータを取ろうという意思が働かないような状態のルーチンワークさえあり、それをこなしていくという義務の中で今やっているという感覚ですか。

●**能勢** もちろん自分自身の研究をするためのデータもフィールドに出て取るのですが、それ以外にも世界各地の地磁気観測所で取られたデータ、もしくは地磁気に限らず、われわれの分野であれば、他の観測所で、ある国の機関が取ったデータを集めてきてデータとして整形してサービスするので、その意味では自分が直接取ったデータではないけれども、データセンターとしてデータサービスを続けていく上で必要なデータの管理が出てくるという状態です。

●**蔵川** そこにはインセンティブがちょっと足りない

ということですね。今は地磁気の例でしたが、ご発表者の中で、データをもうつっているという方は仲里先生が挙げられると思います。そもそも研究データというものの研究する上での位置付け、価値、これはぜひ自分でやらなければいけないのか、人にできればやらせたいのか。また、やりたいという場合は、インセンティブがあるからそうなるのだと思います。その辺はご自分の分野の中ではどんな感じですか。

●**仲里** 私は最近ではデータを取っているわけではないのですが、学生時代はデータを取っていました。今の議論を聞いていて思ったのですが、研究活動自体がデータを取ることなのです。測定しないと新しい知見が得られません。測定するというはもうデータが出てしまうのです。研究者だったら、測定したデータを煮たり焼いたり切ったり貼ったりというプロセスを経て、論文まで仕立てていきます。

私は今、日本 DNA データバンク (DDBJ) という生命科学系のデータを集めるところの隣にいて見ているのですが、「さあ登録するぞ」と言って登録しようとすると、少しパワーが要るのです。測定したそばからデータがデータベースに入って、簡単な操作をするだけで、「登録されました」となれば楽なのですが、さあ最後に登録ですとなると、今まで自分のハードディスクに死蔵されていた、もしかすると前のパソコンに入れっ放しだったかもしれないものまでひっくり返してきて、それをまとめて登録しなければいけないので、そこが研究者としてマイナスで、面倒くさいと思われているところではないかと思っています。

●**武田** 一研究者として見た場合、私も社会のネットワークデータのようなものを自分でつくることがあるのですが、先ほど偉そうにプレゼンしておきながら、自分のデータがどこに行ったか分からなくなっています。もう一回あのデータを使って別の研究者と共同研究をしようとなったときに、「ごめん、そのデータがない。前の共同研究者の手元にあるかもしれない」な

どと言って、やっと手に入ったりすることがあります。

結局、自分も研究データが大事だと言いつつ、自分の研究データをきちんと管理できないのです。理由はほとんど同じで、研究をやっているときには自分にとって大事なデータなので一生懸命やるのですが、終わった瞬間に、論文を書いた瞬間に、そのデータをどのハードディスクに置いたか、ほとんど記憶から飛んでしまうということはよくある話で、これは研究者あるではないかと思います。

●**蔵川** 青木先生は学内で研究データを出せ、出せと言っている立場、システムをつくっていたという立場ですが、そういう反応はありましたか。

●**青木** まだ試行段階なので、あまり研究データを出せと宣伝ができていないのです。それぞれの研究者に対して、「あなたのところはデータをどうやって保存しているのですか」とインタビューして回ってつり上げている状態ではないという状況です（笑）。

個人的なお話をすると、私はシミュレーションが専門なので、20 テラぐらいハードディスクを持っていて、そんなものを保存しろと言われてもできようがないということが一つあります。

今は、コンピューターサイエンス系では研究データをどう保存するかはまだかなり模索状態ですが、バイオの実験系では、きちんと実験計画を立ててそれに従ってデータを取りなさいということを事細かに最初に教育されるという話を聞いています。それに従ってデータの蓄積が進むということであれば、研究のプロセスが自動的にデジタル化され、ラボラトリー・研究室の中できちんとアイデンティファイできる形で整理されているというルールが構築されるのであれば、それはデータマネジメントの第一歩になるのではないかと思います。

●**蔵川** ルール化されればみんなやりますよということでしょうか。RDA のインターナショナル・デー

タ・ウィークが 2016 年 9 月にありました。RDA には何回か出席したのですが、そこに科学技術データ委員会（CODATA）も併設されていて、データのインセンティブについて語り合うセッションがありました。データ共有はこれからどんどん拡大していく方向に向かいたいのですが、現状ではデータ共有のインセンティブがありません。

「あなたはこれをしなければ失職します」と言われたらやるかもしれませんが、そうでない限り、実質自由ですと言われた瞬間に、インセンティブがなければ研究者は誰も動きません。では、どんなインセンティブが大事かを皆さんで議論しましょう、もしそれを研究データを出す側として受け入れられるのであれば、それを仕組みとしてつくってくださいというようなことが議論されていたのです。

能勢先生にしても、これ以上何のインセンティブもなければ私は何もしませんという意見も含めて企画されたと思うのですが（笑）、どんなインセンティブがあれば「私は動きます」となるのでしょうか。

●**能勢** 難しいですね。何か物事を動かすためにはインセンティブは非常に重要で、特に内的・外的とよくいわれますが、研究者自身が、データを公開することにより、自分のデータが学術の発展や人類の知に貢献するという意識を持ち、内的なインセンティブによって公開していただくことが理想的な状況ではあると思います。

データベースの維持には非常に手間が掛かります。CODATA でも、データベースを維持していくビジネスモデルについての議論があり、インセンティブはそれにも関わってくると思うのです。インセンティブは外的なものもある程度必要かなと思います。その際にどういう仕組みが可能なのかということが、現在自分でも考えているテーマです。

自分のデータを出すことによってたくさんの方に使われて、学術の発展、流れをリードしていけるのだという内的インセンティブもあると思いますが、投入したコストに対して見返りがあるようなインセンティブ

をつくっていけないかと考えています。

●**蔵川** 非常に奥ゆかしく表現されるので分かりにくい気がしなくもないのですが、仲里先生、インセンティブについてどうですか。

●**仲里** 先ほども言いましたが、私はDDBJというデータを集めるところの隣にいて、人のデータが集まってくるのを見ているわけです。DDBJはデータを集めることが仕事で、私たちはそれをリサイクルしてユーザーに使ってもらうことが仕事です。そのために、自己紹介でお見せしたような検索エンジンをつくっているのですが、検索エンジンをつくることは、研究者的には別にプラスにならないのです。それに対して何か論文を書くとはじめて研究者として評価されますが、データベースをつくりました、ツールをつくりただけでは、今までは研究者として評価されてきませんでした。

でも、今はデータジャーナルなどが出てきて、そういうところに、つくったデータベースそのものについて、「こんなものをつくりました」ということを記述・報告できるようになり、それがちゃんと他の論文と一緒にリファーできるようになってきていて、前よりは良くなったのかなという気はしています。

また、データをデータベースに入れることは、自分のデータに客観的に見てどのような価値があるのか評価してもらえる土俵に乗るといってもあります。具体的に言うと、私が手伝ったデータベースでも、「当然このデータはオープンだろう」と考えていたら、いきなり横から知財部が出てきて、「それはすごく重要なデータですからオープンにしないでください」と頼まれることも若干はあるのです。

オープンにするかクローズにするかは別として、データをデータベースにためるといことは、評価してもらえない土俵に乗るといことなので、「あなたのデータが知財として売り出せるかどうか分かるかもしれないので、入れてみませんか」といことはインセンテ

ィブになるのではないかと個人的には感じています。

●**武田** 研究者がデータをつくるインセンティブにどのようなものがあるかというメニューは既に分かっていると思います。それは各分野でウエートが違うと思いますが、データサイテーション、データのオーナーシップ、マンドトリーなどです。逆に言うと、メニューにないような、今までにないインセンティブが降って湧くとは思えません。

掛けるコスト・労力とのバランスが取ればインセンティブになるし、取れなければインセンティブにならないわけです。コストを下げることは研究者単独ではできないので、大学・研究機関、インフラ側で頑張らなければいけないと思います。研究者のインセンティブは、それほど頭を悩ませて、今さらひねくり出すようなものではないのです。データサイテーションの仕組みやデータジャーナルというものも一応できたので、最低限の仕組みはできたと思っています。

そこで問題になるのは、今までのア－ティクルと同じようにやればうまくいくのかということ、うまくいかないであろうということです。論文は読めば良いか悪いか分かります。ところが、研究データになるとそれが簡単ではありません。そのデータが本当に価値がある良いデータなのか、クオリティが高いデータなのかは、データそのものを見ても分からないのです。それがむしろ、研究データに関わるかなり本質的な問題点ではないかと思います。

たとえサイテーションの仕組みができて、データジャーナルができて、それはある種の評価システムではあるのですが、本質的に中身を見られない、中身を評価できないというところがやはり論文と違います。それをどうしようという答えは私にもありません。

FAIR原則で言うと、Re-usable（再利用できる）の項目で、「データは由来を付けるべき」とされていますが、まだ良い解決法はないのではないかと感じます。むしろ他の分野で、データのクオリティに関して何か活動があればご紹介いただければと思います。

●仲里 私は次世代シーケンサーのデータを扱っていますが、小さくても数ギガで、大きいとテラぐらいになってしまいます。前はダウンロードするのに一晩かかるくらいで、そういうデータを輸送するのに何がいか測定した人がいて、ハードディスクを宅急便で送り付けるのがいいという結論が出たぐらいのデータでした。

今、検索エンジンをつくっていると、「じゃあ、どれを使えばいいの」と言われるのです。私たちはそれに対して二つの取り組みをしています。一つは、データクオリティをチェックするプログラムがあるのでそれにかけることです。そうするとリユースするときにあらかじめクオリティチェックがかかっているので、それを見てこれは使えるなどと思って、落とします。もう一つは、データを登録することと論文を出すことは独立なので、論文がきちんと出ているデータはこれですというマークを付けることです。そうすると、論文が出るぐらいクオリティがいいようなデータだろうとユーザーは分かる、そのような仕組みを入れています。

●南山 今、データジャーナルなどの品質管理の問題が少し出たと思うのですが、Nature の「Scientific Data」などでは、品質管理というほどかは分かりませんが、一つ一つバリデーションはどうやった、メソッドはどうやったという項目がきちんと書かれています。これを書いてあるのだから、きちんとしたデータだろうというような、形式面から品質を担保するような取り組みは行われていると思います。

●蔵川 品質という話が出ました。青木先生、死蔵されたデータという話もありましたが、死蔵されたデータの品質についてコメントはありますか。

●青木 実験系の場合はほとんどが失敗データですよ。何らかのはずみで一度掘り出して、何かを見つけたというときは追試して、もう一度確認を取るわけです。死蔵されたデータというのは、研究者自身しか気

が付かないこと、あるいは測定条件などを残してはいますが、そこから書き漏らした何かが入っているようなデータだと思うのです。

少し話がずれてきましたが、研究データをオープンにすることには、何か心理的な抵抗があるのです。例えば、ストレージの容量が足りないのにこれをずっと残しておいていいのか、データを取ったけれどどんな名前を付けようか、(仮)のような名前のままデータリポジトリに上げるわけにはいかないといった心理的な抵抗と物理的制約によって、自分がやっている実験そのものも上手にマネジメントできない状況があります。それで、何かの弾みで再利用しようとしてもできないというところがあるのかなと思います。

そういうところをうまくサポートできるようなパーソナルな研究データマネジメントをして、そのうちの一部をアーカイブとしてパブリッシュできるというようなサイクルが描けたら、研究者自身にとっては非常に有用なのではないかと思います。

その場合は物理的制約はできる限り外したいのです。アップロードは1日1回までとか、あなたが使えるファイル容量はこれぐらいとか、自分のハードディスクのバックアップも取れないような量しかもらえないという制約はできる限り外すことが、インフラを提供する、気持ち良くサービスを使ってもらう側のやるべき内容ではないかと考えています。

●蔵川 結構いい感じで、研究者にとっての研究データマネジメントについてスライドしていただきました。仲里先生は分野としてもかなりされていますよね。どんなことが今問題になっているのか、青木先生が言われたような話は既に終わっているのか、これからもやりたいのか聞かせてください。

●仲里 まさに同じ問題に直面しているのではないかと思います。次世代シーケンサーのデータはとて大きいので、例えばアメリカの NCBI (国立生物工学情報センター) などでは集めているのですが、一度、予

算が付けられないからやめるというアナウンスが出た
ことがありました。その後やはりやることになったの
でよかったのですが、やはりデータを持って運用する
側は、ディスクを食って、いくらディスクがあっても
足りないという状況に直面していて、そうすると要ら
ないデータは消そうとか、非可逆圧縮で、クオリティ
が悪いところは切って圧縮するといった話がこの界
隈では出てきます。

●**蔵川** 今度は図書館にとっての研究データマネジ
メントの話を下田様に振りたいと思います。長い間古写
真を展示されていて、研究データという観点からして
も何かマネジメントの必要性はありますか。

●**下田** 古写真について言えば、研究データと言え
るレベルのデータは扱っておらず、あくまでもメタデー
タの一要素としてのデータを扱ってきました。だから
こそ研究者の間で共有できました。研究者はもともと、
これぐらいの古写真に関する解説は研究成果にならない
と思っていたと思います。むしろ、そういうものを
たくさん集めて、それを自分が通覧することで、そこ
から一歩踏み込んで自分の研究成果を出していくとい
うスタイルだったのではないかと思いますので、あまり
研究データそのものを図書館で扱ってきたというこ
とはありませんでした。

●**蔵川** 南山様に聞きたいのですが、既に研究データ
について、新しい IUGONET という組織でいろいろ
やろうとしている、その経験から言って研究データ管
理はどんな感じですか。

●**南山** 研究データ管理そのものは、研究プロセスを
通じて研究データがどのように生成されていくかとい
う、全体のもっと大きな話だと思うのですが、今回は
その中でメタデータに関するところを取り出して研究
データ管理にコミットしはじめたというスタンス、位
置付けで考えています。

図書館員が研究データ管理全体にどこまで手を出せ
るかというのはこれからの話だと思うのですが、研究
データ管理にはいろいろなプロセスがあります。そも
そも管理のトレーニングをしなければいけない、ある
いは実際にメタデータをつくる時に検証まで手伝え
るのか、メタデータをつくっても実データの保存は誰
がやるのか、そのようなところがあると思います。

こちらからはやれることをとにかく提示していく形
ですが、(協同に対する)私の感覚としては、研究者
の方々が研究データ管理をどこまでやりたいか、その
中で図書館員に何を求めてくれるかです。分野によ
っては、これは業績として認められるし、きちんと若手
の育成に使っているのでもそんな手数は不要ですとい
うところもあるかもしれませんし、そうではないところ
もあるかもしれません。まだ全くやっていないのでぜ
ひ力を借りたいと言ってくれるところもあればうれし
いというスタンスです。これで回答になったでしょ
うか。

●**蔵川** また私が長々質問すると申し訳ないのですが、
図書館から見たときの研究環境をつくるということと、
研究者から見たときの研究環境をつくるということは、
実は違うことが多いのです。研究者の人が研究環境を
つくっているというときに、何をやっているのか聞くと、
「学生にどれだけ論文を読ませるか、学生にどれ
だけ実験をさせるかという研究室の環境をつくってい
る」と答えます。例えば、本の購入をしたり、学習の
場を与えたりするような、図書館でやっているような
ことは、「あれはインフラであって、研究環境ではな
いですよ」とまで言うのです。

一方で図書館は、「私たちは根本的な研究環境をつ
くっています」とよく言われるのですが、そういうギャ
ップは感じましたか。あるとしたらどんな感じですか。

●**南山** おっしゃるとおりのギャップがずっとあると
感じています。私は今は研究所勤務なので、まだ研究

者の方々と話す機会が多く、近い立場にあるとは思いますが、前に東京大学にいたときや、一般の大学図書館では、図書館は図書館として独立していて、「図書館の」仕事をやっている人たちというスタンスがあり、研究者は使えるところだけ使ってくださいという感じで、そもそも研究としての枠組みとだいたいぶずれてしまっていると感じています。

(図書館からの研究支援が) 研究者一般に対する取り組みとしてある以上は、研究室の環境を整えることには全く影響しないというか、関われないような立場だと思うのですが、今後関わっていくのであれば、そもそも研究のプロセスを知った上で図書館の人がそこに何ができるかという視点で取り組んでいかなければいけないのだろうことが、今回の IUGONET の動機の一つです。

(機関リポジトリで実データを扱うこととの関係について、) インフラを別に図書館の方で持ちたいと思って IUGONET との協同をやっているわけではなく、IUGONET でできている研究環境の中で図書館員が役に立つスキルとは何か、ということをまず検証したいのです。その上で、今まで既存でやってきたものにプラスアルファでできるものがあれば提供したいというスタンスで考えるべきだと思っています。

●**蔵川** 下田様は長い間、研究とは無縁の世界で活動してきたと言われていますが、そういうギャップのようなものはありますか。

●**下田** 古写真以外ではあまり経験していないので、ギャップというのはよく分かりません。長崎大学で古写真の研究に関わる先生方は複数おられたのですが、それぞれが写真研究のプロということではなく、いろいろな別の研究分野を持っておられたので、本当の研究フィールドの深いところで古写真に関わるというよりは、むしろ自分の専門分野から見たときに古写真をどんなふう利用できるかということでした。

研究者との間で流れているデータは、研究上それほ

ど深いものではなかったと思います。むしろ、新聞記事になったり、街に出かけて行って古写真展をしたり、テレビに出たり、本を出したりするのに使われていました。図書館が研究者に期待されたことは、書いた新聞記事の原稿をひととおりチェックすること、本を出版するときの多少編集的な仕事などであって、サイエンスだとか、そういうものではなかったように思います。むしろ一般の人たちに公開するということだったので、一緒にできてきたのではないかと思います。

●**蔵川** ちょうど研究者と図書館員のコラボレーションという観点からいろいろなストーリーをご紹介いただいたのですが、南山様にまた聞きたいことは、IUGONET を通して、研究者のニーズが何だと理解して、それにどれだけ応えたのか、応えられたのかということですか。本質的なディスカッションに向かっていきたいと思います。研究データマネジメントではハードディスクがふんだんにあればいいなど、情報システムには関係あるけれど、図書館員との絡みでは難しい点があったので、その辺の接点という意味で。

●**南山** 研究者が困っていることに対して、図書館はどのように貢献できるかという視点から私が今考えている答えの一つが、スライドでもご紹介したネットワークの活用とメタデータ運用です。

研究者のニーズに対してどれだけ応えられたかというと、私としては、作業が減って役には立つのだろうなというスタンスで関わってきているのですが、実際にやった人、IUGONET 側の人に聞くのが面白いのではないかと思います。梅村(宜生)さん(名古屋大 ISEE)、いらっしゃいますか。僕は役に立ちましたか?(笑)

●**梅村** 非常に役に立っております。というのは、研究の現場では死蔵化されつつあるデータ、つまり、まだメタデータが付与されていないデータがまだまだたくさんあります。では、研究者サイドではどういう状

況かという、「そんなところ構ってられないよ」という、手を付けられない状況なのです。

そんな中で南山さんから、「ちょっとやらせてください」とトライアルのアプローチを頂いて、われわれの期待以上に成果を出していただきました。われわれの期待以上に図書館側ができそうだという考えを持っています。そういう意味では非常に役に立っています。

今回はトライアルで体制を組ませていただきましたが、実は内々で、オフィシャルパートナーとしてやったらどうなのかという意見もちらちら出ています。現場と図書のサイドでそういう体制が組めればいいなと思います。

まだまだ現場では死蔵されつつあるデータがあるのですが、まさに蔵書管理ですから、図書館の人はそういったものの整理整頓が得意だと思うのです。「現場はこういうデータを持っていますよね。図書にらせてください。メタデータをつくりますよ」というリーダーシップをとってぜひやってもらえればと思っています。

●南山 どうもありがとうございます。大変うれしいです。協同の話はまた別として、今でお答えになったでしょうか。

●蔵川 役に立ったということで、トンネルの遠くの方に光が見えている感じがしてはいるのですが、まだコラボレーションしていない研究者からすると、「本当に図書館員にできるのですか、私の仕事の何ができるのですか」という意見を言う人が多いのです。仲里先生はまだコラボレーションされていないと思うので、図書館員に何を任せればいいのかという意見はありますか。

●仲里 個人的な意見というよりは全体的に聞いて思ったことですが、同じ話を何度もしていますけれど、DDBJ では生命科学のデータ、例えば遺伝子配列や次世代シーケンサーのデータを集めています。「集めて

います」と言いますが、それは研究者が Excel などのフィールドを埋めてサブミットするのです。すると、DDBJ にアノテーターという人がいて、それを眺めて、きちんと書いているのかどうかチェックします。例えば、これは絶対にオスのデータなのにメスになっているとすると、はねます。

そういうプロセスは、南山さんがおっしゃったメタデータをきちんと付けるというプロセスとまさに同じではないかと思っています。職業は違いますが、結局やっていることは同じなので、生命科学のアノテーターが回している仕事を図書館の人が同じように回しているということを感じました。

●南山 今聞いていて、発表で話し損ねたことを1点思い出しました。実際に研究者の方々にデータベースを運用されていても、自分の専門分野以外の、超高層物理の中でも幾つか細かく分野があって、例えば地磁気のデータ、宇宙系のデータなどがあるのですが、当然、研究者間でも（細分化された）自分の分野でなければ分からない。どのようにそれ（アノテーション）を対応するのか聞くと、「担当の人（PI）に聞いてやっている」と話されています。「担当の人に聞いてメタデータをつくるのであれば、それを図書館の人がやっても同じ仕事ができると思います」という殺し文句で今回のお話を進めてきたのですが、そういう関わり方も図書館員としてはありなのではないかと思っています。図書館員は「人と人をつなぐのが専門性である」と表現されている方もいらっしゃるのです。

●蔵川 青木先生は今度全学システムをつくらうと言われていますが、図書館員とのコラボという点で、何か意見はありますか。

●青木 全学のプロジェクトで意見交換を行うまで、私は恥ずかしながら、図書館の方々の、オープンアクセスやオープンデータなどデジタルデータに対する力の入れようを実は全く知りませんでした。ここ数カ月

いろいろ話をしている、図書館の方が情報を扱うことに関する専門性を有している、その成果の一つとして各機関リポジトリが存在するのだという認識に至っています。ですから、情報を上手にアノテートして整理してもらおうということについて、今後図書館の方とコラボレートする機会は存分にあると思っています。

ただ、図書館側のヒューマンリソースが限られている状況で、山のようにある研究データをどう切ったり貼ったりするかは非常に難しい問題です。南山さんが今回、一例としてデータキュレーションを採用されたということですが、いろいろなところにデータキュレーションは必要なので、データキュレーションのスキルを持った人を、図書館の専門職員、学生、研究者への教育といったものとセットにして、リテラシーの底上げができるような体制ができればそれがベストではないかと考えています。

●**蔵川** 武田先生、研究データ利活用協議会の中で、図書館とのコラボレーションについて何かありますか。

●**武田** 研究データ利活用協議会はどちらかというと、データを直接扱う人をまずターゲットにするものなので、それほど図書館を強く意識はしていません。しかし今回、キックオフの時点で千葉大学のアカデミック・リンク・センターに入ってくださいました。アカデミック・リンク・センターは自分自身がデータをつくるというよりは、教員などからもらっているデータをキュレーションするような立場にあるので、そういう意味ではそういうメンバーがいないわけではありませんが、DOI を付けるというスタンスで言うと、あまりそこに重きを置いていなかったところはあります。

でも、今日の議論を聞いていて、そこはやはり重要だと思いました。結局、見えないデータを見える化するためには、豊富なメタデータを付けなければいけません。単に形式的なメタデータだけではなく、中身まで入ったメタデータを付けなければ、データの中身のクオリティが保証できないからです。それはデータ

の価値を高めませんが、それをやることは研究者にはものすごく大変です。そこでキュレーションはある程度別のセクターが担ってくれたら、先ほどの研究者側のコストを下げることに繋がります。

図書館、キュレーションセンターのようなところ、アノテーション専門のセクターが、実は同じ役割をしていることを理解して、役割分担が明示的にできると、研究者も、図書館も、キュレーションセンター的なところも、お互いにハッピーになれるのではないかと今の議論を聞いていて思いました。お互いに役割を認識し合うことができれば、もう少し前向きに進めそうな気がしてきました。

●**蔵川** そろそろ時間ですので、会場を含めて質問をしていただける方、ご意見がある方はいらっしゃいますか。

●**フロア** WDS-IPO (World Data System-International Programme Office) の渡邊です。私自身も研究データを取って論文を書いています。私や能勢さんが当てはまるようなデータセンター型研究者は、データアーカイブを大事にしている、データアーカイブで自分の論文が書けるし、データベースもつくり、データ解析システムまでつくってしまいます。そうするとそれが共同研究になり、一般公開しようという話になります。IUGONETなどはまさにそれです。データベースがないと自分の研究が進まないというタイプの研究をしている人たちは、データの公開に関して比較的問題が少ないです。もちろんデータセンターをどう維持するか、金の問題、人の問題はありますが、それこそ WDS が取り組むべき問題で、これははっきりしています。

研究者が論文を書いて、そのデータを公開するのはある程度義務でしょう。裏付けデータを何も公開しないで、「書きました」ではおかしいわけです。やはりそれは何らかの形でアーカイブするというか、いつも見られるように公開しておくことが研究者の義務であり、モラルの問題です。学会やアカデミーなどの議論

によって、それは進められるでしょう。

一番問題なのは研究データです。先ほど失敗データとありましたが、時系列データ、例えば太陽面爆発のデータなどはいつ起きるか分からないので延々とデータを取っています。論文を書くときは、何かあったところのデータだけを使います。他の延々と取ったデータも使い道があるかもしれませんが、あまり興味のない人が多いので、放ってあるデータも結構あるのです。

ところが、何も無い、延々と取ったデータが大事だという人もいます。自分が考えていることとは違う使い方をされる可能性もあります。私は電波天文学で、太陽から来る風をリモートセンシングで見えていたのですが、先駆けて公開しました。「そんなデータを公開していいのかわかるか」と言われ、「それで誰かに論文を書かされたらどうするのか」とさえ言われました。でも、われわれは、公開することによって共同研究が広がるのではないかと、自分たちが考えづらいような使い方をする人もいないかと考えて公開を始めたのです。

そこで大事なことは、自分が研究に使わなかったデータでも、それを自分は持っている、こういうデータがどこにあるという所在情報だけでも公開してもらうことです。そうすれば新しい共同研究ができるかもしれません。データを取った人に断りもせず、名前も入れずに研究論文を出す人などいないのです。それが研究者としての一つのインセンティブにもなるのではないのでしょうか。

あまり最初から、エラーも含めて全部保存しなさいというのは無理ですが、とにかくデータの所在情報だけでも公開する、それが手始めではないのでしょうか。そういう印象を持ちました。図書館あたりにもそういうところからまずやってもらうと、研究者としてもインセンティブになると思います。

●武田 インセンティブはいろいろあり得ますが、今は人的コスト、実際のコストも含めて、掛かるコストをいかに下げるかということにかかっているのではな

いかと思います。

今、渡邊先生がおっしゃったことも、取っているデータが、研究者がわざわざ触らなくても、そのままデータアーカイブなどに取り込まれるのであれば研究者はやると思います。黙っていてもできてしまうのだから、イエスカノーかと聞けば、きっとイエスと言います。そこで「これはぜひ隠しておきたい」と言う人は、今時は多分いないでしょう。

ただ、「あなたがコピーして、どこそこに置いて、名前を付けて登録してください」と言われると、優先度が低いからやらないのです。小さなインセンティブ、もしかすると1万人に1人の研究者が利用するかもしれないという程度なら、普通の人はそんなことをしないのです。でも、それをボタン一つでできるようにして、コストが下がればやるでしょう。

インセンティブの種類よりもむしろ、それを実現するためのコストをいかに安くできるかに尽きるのではないかとというのが僕の思うところです。

●蔵川 最後に南山さん、いいインセンティブの話がありました。その中に図書館はどうやって入ってきますか。

●南山 今の武田先生のお話を受けると、公開コストをいかに下げるかということにまず図書館がコミットしていくのが良いように思います。いろいろな研究者に聞いても、(図書館業界の)大御所の方に聞いても、メタデータの仕事はもともと図書館が専属でやってきた、それは専門性であると言うので、それならばメタデータの仕事を図書館で引き受けて、図書館に全部まとめることで機関としてのコストを下げることで、貢献できればと考えます。

(実際の運用では)ボタン一つで図書館の人にデータが飛び、図書館の人はそのメタデータを必要な分だけつくる。その中から公開したいデータは公開する、というような理想的なワークフローがつかれるのではないかと思います。この辺はインフラがあつて

こその話ですが、インフラがあれば図書館としては絡みやすいですし、絡むべきだと思っています。

●**蔵川** ありがとうございます。時間になりましたので、登壇者の方々へ拍手をもってこのパネルを終わりにしたいと思います。