

第1回 SPARC Japan セミナー2019

「人文社会系分野におけるオープンサイエンス ～実践に向けて～」

パネルディスカッション



- 鈴木 親彦** (国立情報学研究所 / データサイエンス共同利用基盤施設 人文学オープンデータ共同利用センター)
- 小木曾 智信** (国立国語研究所)
- 加納 靖之** (東京大学地震研究所 / 地震火山史料連携研究機構)
- 中村 美里** (東京大学附属図書館)

話題提供：大学図書館とオープンサイエンス —個人的な経験—

●中村 パネルディスカッションの冒頭は話題提供ということで、私から話をさせていただきます。

今日は3件の講演をしていただきましたが、大学図書館という視点で、私の個人的な経験から話をさせていただきますと思います。

オープンサイエンス関連の主な大学図書館業務、取り組みなど

まず、オープンサイエンスに関連した大学図書館の業務、取り組みなどについてざっと見てみると、まず、機関リポジトリの運用、オープンアクセスリポジトリ推進協会 (JPCOAR) の取り組みや大学図書館コンソーシアム連合 (JUSTICE) が電子ジャーナルを中心とした電子コンテンツの安定利用のためのさまざまな活動などがあります。近年は、各大学でオープンアクセス方針が策定されていたりします。

また、私は国立大学にいますので、例えば国立大学図書館協会が、「国立大学図書館機能の強化と革新に向けて」というビジョンを出しているのですが、それはかなりオープンサイエンス時代の図書館の在り方を意識して作られていますし、2019年3月には「国立大

学図書館のオープンサイエンスへの取り組み」という文書も出されています。所蔵資料のデジタル化・公開、デジタルアーカイブの構築・運用も図書館で盛んに行われています。

次に、研究データの管理・保存については、機関リポジトリへの研究データ登録など、少し例はあるようなのですが、大学図書館全体で見るとまだ様子見というか、大学図書館が大きく関わっている状況ではまだないのかなと思っています。

その中で、今日は人文社会系ということもあり、私が今まで経験した、所蔵資料のデジタル化・公開、デジタルアーカイブの構築・運用からオープンサイエンスというものを考えてみたいと思っています。

個人的な経験 1：国文学研究資料館（国文研）での取り組み

私は2013年度から2016年度まで国文学研究資料館（国文研）にいて、「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」に携わっていました。国文研と国内の大学の幾つかで、古典籍のデジタル化、30万点の画像データベースを作るというプロジェクトだったのですが、その一環で、2015年11月に国立情報学研究所 (NII) との協働により、国文研が持っていた古典籍350点の画像・書誌データなどをデータ

セットという形にして公開しました。利用条件は CC BY-SA として、オープンデータとして出しました。

1年後に、引き続き NII と連携しつつ、その当時発足した人文学オープンデータ共同利用センター (CODH) と新たに連携して、そこで各種データセットの公開を行いました。この取り組みは現在も続いていて、登録データはどんどん増えている状態です。

このデータの公開によって、さまざまな活用例が生まれました (図 1)。「くずし字学習支援アプリ KuLA」の用例画像の一つに、国文研のデータセットが使われています。次に、このデータセットの活用を考えるアイデアソンをしようというイベントが行われ、そこで、江戸料理を古典籍から再現してみたら面白いのではないかという意見が出され、そこからとんとん拍子に話が進んで、クックパッドで江戸ご飯というページを作り、そこで本当に江戸料理の再現を行いました。

また、一文字ずつ画像を切り出して、文字のデータセットを公開することも行いました。単なる画像ではなく、その中の一文字ずつを切り取って、データセットとして出しました。

画像をデータセットとして公開するという話が来たとき、最初は古典籍の画像をデータセットで出して一体誰が使うのだろうと、実は内心思っていました。また、今までデータベースの中で出してきたものを、データセットという、プレーンとまではいかないのですが、そういう形で出して、何かあらが見えてきて怒られるのではないかと、非難されるのではないかとという心

配をしていたのですが、先ほどの活用例にあったとおり、予想以上の反響があり、さまざまな分野の人が関心を持ってくれるということを実感しました。あるシンポジウムでこの取り組みを話したときに、「国文研のデータ公開はオープンサイエンスのグッドプラクティスになりそうだと思う」と司会の方から言われて、「これがオープンサイエンスと言われるものなの？」と、人から言われて気が付きました。これが 2016 年ぐらいです。

この国文研の取り組みがうまく進んだ理由として、古典籍なので著作権処理を気にしなくてもよかったこと、国文研の大型プロジェクトとしての枠組みがあったので、図書館だけではなく研究者の方と一緒に進められたということが大きかったと思います

先ほど述べたように、最初はデータを出すというのはどうなのだろうと思っていました。私が心配するより、世間や技術はもうずっと先に進んでいて、データをきちんと出せば何か面白がってくれる人がいる、何か利活用してくれる人がいるのだなということを感じてきたことは自分の中で非常に大きな経験でした。

個人的な経験 2：東京大学デジタルアーカイブズ構築事業

国文研への出向時期が終わって東京大学に戻り、今は「東京大学デジタルアーカイブズ構築事業」に携わっています。これは東大の「ビジョン 2020」という行動指針に基づいて始まったもので、図書館だけではなく、博物館、文書館、情報基盤センターという 4 部局が連携して実施しています。主には、公募によるデジタル化の予算配分、画像公開の支援、さまざまなデジタル化資料の活用促進を行っています。図 2 にあるとおり、これまで東京大学では学部や研究所、研究室、科研費単位でデータベース、デジタルアーカイブが作られていて、それぞれ特徴を持ったデータベースは必要なのですが、それらを横断的に検索できる仕組みがありませんでした。

それらのメタデータを集約して、「東京大学学術資



(図 1)

産等アーカイブズポータル」というものを 2019 年 6 月に公開して、運用を行っています。また、画像はあるけれどもなかなか公開システムの構築ができないといった画像をお預かりして、IIIF で公開する公開支援も行っています。全学的な取り組みとしては、「東京大学学術資産等アーカイブズポータル」の公開により、東京大学で公開しているデジタルコレクションがかなり横断的に検索できるようになりました。

次に、総合図書館に特化した取り組みなのですが、2018 年 1 月から IIIF に対応した画像公開を開始しました。それから少し遅れて、2018 年 6 月に画像データの二次利用の条件を緩和しました。せっかく IIIF で公開するので、ライセンス的にも使い勝手の良いものにしようということで、クリエイティブ・コモンズ・ライセンスは付けていないのですが、CC BY 相当の条件で公開を始めました。今、この動きは総合図書館以外の各図書館・室にも少しずつ広がってきています。

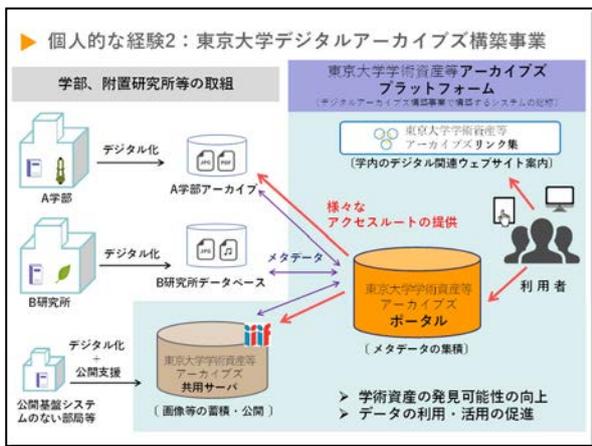
もう一つ、最近の面白い取り組みとして、2019 年 9 月に亀井文庫『ピラネージ版画集』を再公開しました（図 3）。これは元々、総合図書館の貴重図書で、2003 年ぐらいに、ある科研費で画像データベースが構築されたのですが、システム運用上の問題で公開をいったん停止していました。図書館はこのデータベースの構築やデータ管理には全くノータッチでした。

昨年、デジタル化すべき所蔵資料を検討していたときに、「亀井文庫のピラネージ」という意見が出たの

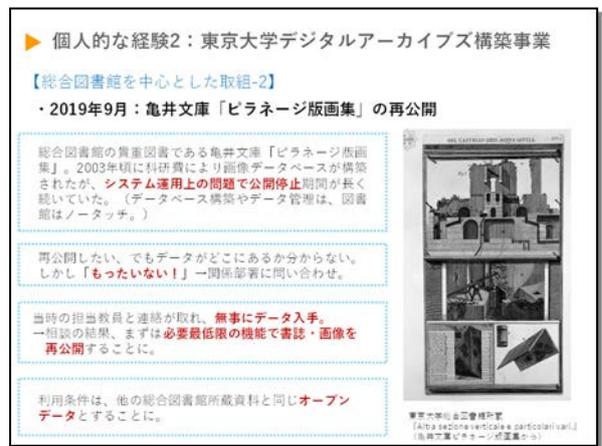
を機に、実はもうデジタル化はされていてデータベースが止まっている状況ということが分かり、それはもったいないということで、関係部署、あるいは当時の担当者と思われる方にいろいろ問い合わせをしました。その結果、当時担当されていた先生と連絡が取れ、こちらの意図を話したところ、もちろん喜んで協力しますと言っていただいて、無事にハードディスクを預かりました。そこからデータを救済でき、公開していた当時のまま再公開できたわけではないのですが、画像と書誌をきちんと見られる必要最低限の形でこの 9 月に再公開することができました。利用条件としては、総合図書館所蔵資料と同じオープンデータとして公開することができました。

これも数年前でしたら、「オープン化で出しませんか」と話をしても駄目だったかもしれませんが、先生方の方が「今はそういう時代だから使ってもらった方がいい」と言われて、誰も反対することなくオープンデータとしてリニューアル公開することができました。

さて、大学図書館は、オープンサイエンスといわれ始めたここ数年で急にデジタルアーカイブを作ってきたのかというと全くそんなことはなくて、2000 年ごろから所蔵資料のデジタル化を行ってきました。ではこれからのデジタル化、デジタルアーカイブ構築には何が必要なのかと考えたときに、単なるお宝資料を、1 点の画像を大事に出すというより、活用を見据えたデータ公開というものが become 必要になるのではないかと思います。



(図 2)



(図 3)

IIIIF やデータセットなど、利用しやすい形式で提供すること、また、ライセンスです。国語研究所の話でもあったように、オープンデータにすることが一番いいとは思いますが、きちんとライセンスを付ける、こういう条件で使ってくださいということを少なくとも明示するというような利用しやすい条件、あるいは見やすい条件で提供することが必要です。また、教員や研究者となるべく協働して出すことも必要だと思います。図書館だけで画像を出しましたと言っても、なかなか不十分なこともあるので、できれば研究者と話しながら、一緒に何か画像公開やデジタル化をしていけばいい、それがおのずとオープンサイエンスにつながっていくのではないかと、ここ数年この仕事をしながら思っています。

図書館業務はいろいろあるのですが、時代がオープン化に進んでいること、あるいは今日の講演で聞いたような話を念頭に置いて、それに役立つデータの出し方をしていくことを図書館で考えることが必要なのではないかと考えています。それぞれの大学の規模や分野などによって、これは図書館の仕事、これはURA の仕事、これは研究者の仕事と一つに決められるものではないと思うのですが、一人一人の職員、図書館の人が、研究者や非専門家、私たち図書館員が、業務や研究支援をするときにどうすれば便利にデータが使えるかを考えていくことが大事なのではないかと思っています。

大学図書館からの視点ということで、自分の経験に基づいた思いをここで述べさせていただきます。

それではここからパネルディスカッションに入ります。この時間からはモデレーターを鈴木先生にお願いしたいと思います。

●鈴木 それでは、ここからパネルディスカッションに移りたいと思います。パネリストはご登壇いただいた小木曾先生と加納先生、そして話題提供いただいた東京大学附属図書館中村さんの3名で、モデレーターは私、鈴木が務めさせていただきます。よろしくお願

いいたします。

前半の皆さんの発表をごく簡単にまとめておきましょう。シチズンサイエンス系のお話として加納先生のお話があり、研究者寄りのコミュニティの話として小木曾先生の話がありました。また基盤をつくるところに市民の方々が関わっているという加納先生のお話と、既に研究コミュニティの中であがり出来上がった基盤をオープンに活用していくという小木曾先生の話があったと思います。これら二つは対立構造として市民対専門家というように分けられる話ではなく、相互に補い合う話だということは皆さんお分かりになるかと思います。そのあたりを踏まえつつ、SPARC Japan としてはもう一つ重要なプレーヤーとして、今回ご登壇いただいた図書館という三つ目のプレーヤーを置きたいと考え、こういう形のパネルをつくりました。

最初に、お互いの発表を聞いた上でヒントになったことや、共有できそうな問題点がありましたら、順番に話していただければと思います。

●小木曾 われわれ国語研究所が今、考えているのは、まだしょせん研究者コミュニティの中ですが、本当は今後はシチズンサイエンスという全体の話まで広げていかなければいけないと思っています。

特に展開としてあり得るのは、方言など話者に密着した情報になってくると、市民を巻き込むことはとても重要になるのではないかと思っています。そのときに、「みんなで翻刻」のような、うまくオーガナイズされている例は参考になります。また、本当は、今日お話したような「中納言」上でアノテーションという段階でも一般の人にどんどんやってもらいたいなという思いはあるのです。ひょっとするとやってもらえるのかもしれないのですが、先ほど考えてやはり駄目かなと思ったのは、「みんなで翻刻」は、翻刻すると達成感があるし、「これはこの字だったんだ」と思えるのですが、「みんなで品詞分解」はそんなにやってくれるかなと（笑）、少し難しいかもなと思ったのです。

私は、昔は作業として品詞分解的なことをやって直していたのですが、言葉が好きな人間、昔の本が好きで人間にとってはそれは面白い作業です。「コンピュータはこんな間違えてるわ」と直す、「こんな書き方もあるんだ」「こんな漢字でこう読ませるんだ」と、そういう面白さもあるので、うまくそういうことが伝えられるようになったらその可能性はあるのではないかと考えた次第です。もっと研究者の中にとどまらず進められるようになったらと考えております。

●加納 今の小木曾さんの話を受ける形でお話すると、この間、東京大学の学生向けに、「みんなで翻刻ゾン」という、3日間ひたすら「みんなで翻刻」に関わってもらいイベントをやりました。最後に感想を聞くと、「これは面白い」「入試に出た資料、文献が原文で読めますとやったら、高校生が参加してくれるのではないか」という意見をもらって、なるほど、そういう広げ方もあるなと思いました。品詞分解だときっと入試に、そこまで専門的なものは出ないかもしれませんが、例えばそういう楽しみだけではなくて、勉強したいということはどう仕掛けるかということの一つヒントになるのではないかと思います。この間まで受験勉強をやっていた人たちのアイデアは新鮮で面白かったです。

今日、小野さんの話を聞いていて思ったのは、やはり評価してもらうことは大事で、「みんなで翻刻」はうまくいっていて、たくさんの方に参加していただいているのですが、どうしてうまくいったのかをきちんと言語化するとか、きちんと残さないといけないプロジェクトにつながっていかないし、また、自分が今後もっと広げていくためにどうしたらいいかが分かってこないで、それを第三者、少し外側から見ていただいて「ここがうまくいっている」とか、ただ「頑張ったね」ではなくて、批判とか、マイナスのところも含めてきちんと評価してもらえたい人があるなと思いました。

中村さんの話も伺っていて、まさにデータベースと

して公開されているものをどんどん翻刻したい、あるいは、そういうものが、別に地震や災害にかかわらず、目に触れるような世の中になるといいなと思っているのです。元々図書館はいろいろな人が本を借りに来たり、資料を調べに来たりして、人が集まる場です。私たちが研究のための資料を調べに行く場なのですが、そういう場所をどううまくつくっていくか、オープンサイエンスの中に位置付けていくかということは、「みんなで翻刻」などをやっても何かできることがあるのではないかと思います。

●中村 私も講演を聞いて、研究機関が作るデータ、あるいは市民科学、非専門家に、図書館がどういう形で関わっていけばいいのかがまだよく見えてこないで、何かこういうところが困っているということがあればぜひ聞きたいです。

市民科学について言うと、そこを図書館が手伝うというよりは、市民として参画していくということが一つあるのかなと。比較的、図書館の人はそういう作業は好きだと思いますので、例えば土曜日に休みが一日あって、「品詞分解しましょう」と言われると私は結構うれしい方だと思います。それは私の独特のところかもしれませんが、そういうことが好きな人は多いような気がします。

場所の提供など、そういうことももちろんあると思いますが、こういうものが必要だよとか、こういうことを助けてほしいよねということがもう少し見えてきたときに、図書館が得意なところを生かす。あるいは、それは全然別の部署であったり、絶対に図書館でなければいけないということはないと思うので、そういう情報がもう少し分かりやすく見えてくると、もう少しつながりが出てくるのかなと、話を聞いていて思いました。

●鈴木 ありがとうございます。今回、加納先生のタイトルで「非専門家」という言葉を使っていたことが大変良いと思っています。これまでオープンサ

イエンスやシチズンサイエンスの関係では、そこに「市民」というかなり限定した言葉が入ってしまっていたのですが、「非専門家」というと、地震であれば、私は地震の非専門家ですし、国語の問題であれば、小木曾先生以外の他の3人は非専門家なわけです。その切り口からは、オープンサイエンスやシチズンサイエンスでの協力の在り方について、また少し変わった風景が見えてくるのではないかと思います。

小野先生の講演で、「ユニバーサルデザインとしてのオープンデータ」ということがありましたが、他の非専門家にとっても使いやすいという点は重要だと思います。学術的な訓練という意味での非専門性ではなくて、その分野の知識を持っていないという非専門性というところでのオープンデータを考えると、大変面白い発想になるのではないかと思います。

そういう意味で言うと、実は「中納言」などのコーパス系も、非専門家に向けてよりオープンに使っていく、研究者だけれども国語の研究者ではない方が使っていくというような形での、インターディシプリナリーとしてのオープンという可能性もその言葉で見えたと思います。そういうあたりから小木曾先生、何かありましたらお願いします。

●**小木曾** 元々、日本語研究者向けに作ったものなのですが、古典のデータがあれだけ入ってくると、当然、文学の方（かた）にも使ってもらいたいということを考えています。最後に少しお話しした、「中納言」にアノテーション機能を追加する、科研・挑戦的研究（開拓）「日本語コーパスに対する情報付与を核としたオープンサイエンス推進環境の構築」のメンバーの中には、歴史民俗博物館や国文学研究資料館の人も入っているのですが、他に国語教育の人に入っているのです。アノテーションで「それ」「これ」などが何を指しているか、それはほとんど古文の問題なわけで、また、古文の問題をコーパス上である意味再現できる部分もあるわけです。ですから、そんなところからいろいろ広げていけないか考えています。

古典教育や言語文化の教育などにまずコーパスを使ってもらるところから始めようかと思っています。そうすると、中高生ですから、一般の方にも同じようなシステムを使っていただけのではないかなと考えています。

●**鈴木** そういう形でさまざまな専門家に使ってもらうためのデータを整備していくことは、今まで図書館がデータを公開する中で、メタデータを標準化したり、発見可能性を上げたりするという事の中でやってきたことだと思います。そういう形で、図書館と進んできた人文科学系のデータに関わっていく可能性も当然あると思います。そのあたり、何か中村さんからご意見がありましたら。

●**中村** 確かにその点は、図書館が今まで得意にしてきたところだとは思いますが。この仕事をしていて悩むのは、どういう画像、あるいはどういうメタデータを出すと研究者の人は役に立つとってくれるのだろうかということです。それは分野によって、先生によってさまざまだと思いますが、こちらは何となくとてもきれいな貴重書を出したいとか、一点物を出したいとか、そういうニーズももちろんあると思うのですが、何かこういうデータを出してほしいとか、この辺を頑張ってもらいたいということがあれば、対話をしながらデジタル化の事業を進めていけるといいのかなと思っているのですけれども。地震学と国語学の分野でもいいのですが、お聞かせいただければ。

●**小木曾** 直接のお答えではないかもしれないのですが、逆に今、こうやって歴史コーパスができてきているのは、図書館の方々がオープンデータとしてたくさん画像などを公開してくださっているおかげの部分があるということ、まず先に申し上げたいと思います。古い時代のものは出版社のデータが多いのですが、江戸時代以降のくずし字については原本から書き起こしているのです。そのデータの大部分は、早稲田大学や

東京大学、大阪大学など、たくさんの大学図書館のデータを使わせていただいているのです。そういうものでできてきていて、コーパスを使ってその元データを見に行くことができるようになっていて、そういう関係にまずあるので、既にこれまでの取り組みのおかげでわれわれは支えられていると申し上げたいです。

その上で、コーパスとメタ情報などをどうつなぐかは、まだ私もすぐ言えないのですが、IIFなどで画像公開していただいているものについては、一部分ですが、京都大学の鈴木本、『今昔物語集』などはIIFでリンクを取っています。われわれの方はまだ追いつかないのですが、ちゃんとやれば、もっと行単位や単語単位でリンクすることもできるはずですよ。これからまだまだ発展の余地があって、われわれが逆に勉強しなければいけないことが多いのですが、そういう可能性があるのではないかと考えています。

●加納 例えば、画像の資料、紙の過去の記録などいろいろなデータを公開するとき、数値データを公開するときには、メタデータが必要なのですが、そもそも私はメタデータに関してあまり詳しくない、いわゆる非専門家になると思います。常識的には、最低限何を付けなければいけないかというところから勉強しなければいけないので、逆に教えていただけるといいなと思っています。

具体的な例で思うことは、例えば地震研究所の画像を「みんなで翻刻」に載せたときに、どの資料にどの地震のことが書かれているか、書誌情報の中に入っている情報を利用して、ある地震に関する資料だけをソートできるような仕組みにできたりします。それは地震研究所の持っていた資料なので、ざっと読んで、「この地震」というふうにタグとか、書誌情報を付けているのですが、もし、そういう情報があれば、あればあっただけ使えます。それは図書館の方が付けるのか、研究者が協力して付けるのか、そこは資料によっていろいろだとは思いますが、一番身近な例として、研究された例、例えば論文などがあって、そこで

どういう情報を重視して研究されているのかを見ると、そのヒントがあるのではないかなとは思っています。

●鈴木 ありがとうございます。大量の質問をSlidoで寄せていただいて、結構クリティカルな質問が幾つかありますので、ここからしばらくSlidoを映していただいて、それを中心に話を進めたいと思います。

小野先生に対して、「シチズンサイエンスで、ボランティア側面は別として、市民が得られるメリットは何だとお考えでしょうか」という質問が来ています。これは加納先生に関する質問にもつながると思いますし、さらにオープンなシチズンサイエンス的な人文学を進める上で非常に重要だと思います。

加納先生も幾つかアンケートの結果などから、こういうメリットがあるというようなことはおっしゃっていました。逆に言うと、お金がもらえないからやらないというようなことをおっしゃっている場合もあるし、元々それを有料でやっていた人の職が奪われるのではないかというお話もありました。

オープンであるということと同時に、長期的にサステイナブルに学術資源を使い続けることも考えると、その辺のバランスはすごく重要だと思います。そのあたりで「みんなで翻刻」で何か議論があったかどうか、加納先生にお聞かせいただければと思うのですが。

●加納 議論をしたというよりも、走りながら考えていたという面はあると思います。いろいろなモチベーションで参加されている方がいて、自分の勉強になるからとか、単に楽しい、オンラインでできるので会に集まって2時間ずっとそこにいなくても、自分の5分なら5分の「隙間時間でできる」ことが面白いというコメントも頂いたりしています。「みんなで翻刻」に参加する人の思いはいろいろなわけですよ。けれども、続けてくださる方は、楽しいでも、勉強になるでもいいですし、研究データを作ることに貢献できるということもあると思いますが、自分に何らかのメリットを感じて参加してくださっています。よりそういうふう

に思ってくださいる方、そういうきっかけで参加してくださる方が増えるためにはどうプロモーションしていけばいいか、どう宣伝していけばいいかと考えたという感じですかね。

小野さんも答えておられますが、「プロジェクトが扱う分野への好奇心を満たすこと」「参加してタスクをこなすことが娯楽になる」ということは、私も賛成です。

●鈴木 これは「いいね」がたくさん付いている質問なのですが、小木曾先生に対する質問として、「小木曾先生が提案された、データがオープンデータであるということとは別に、オープンな形で研究を進める、オープンなサイエンスを進めるということはすごく重要だと思います」。これは私もそう思います。「一方で、研究分野全体にとって基盤となることが既に自明な例なので」、これは小木曾先生の発表を聞くと確かに私も自明だと思うのですが、「コミュニティ全体の取り組みとして資金を獲得してオープンにする、ヨーロッパ型の大きな資金獲得による、社会基盤としてのオープン化のようなものに進む考えはありますか」というご質問が来ています。このあたりをご説明いただけますか。

●小木曾 非常にもっともお話だと一方では思います。オープンにしていく、どこかで買い取ってもらえる話かと思うのですが、そうすることで本当に自由に使えるようになる良さがあると思うのですが、一つだけ、研究機関という観点から、ここは誤解されると嫌なのですが、少しいやらしいことを言いますけれども、オープンにする元のデータはわれわれにとって資産のようなものでもあるわけです。大学図書館がたくさんオープンデータ、貴重書の画像を出してくださるのですが、まさか貴重書本体を出すわけではないです。ところが、デジタルデータとして作った資産をオープン化してしまうというのは、研究所の資産がそのまま流出するわけです。別にそれで構わないと、研究コミ

ュニティとしてはその方がいいかもしれないのですが、その結果、コーパスの出がらしになった国語研究所は要らないと言われると困るのです。

大学のような学生がどんどん入ってくる機関とは違って、研究所が存続していくためにはいろいろな部分が必要でして、オープン化したときに、誰が使っているかという情報が分からなくなり、評価を受けるときに、何件使ってもらったかを言いにくくなるということも一つあるのです。ですから、例えばほとんど無料で出すけれどもオープンではないというようなことでもよければ、割とやりやすいのかもしれないな思ったりしました。

その一方で、私は言語処理もやっていたので、とにかく手元に全部データがないと嫌ですし、歴史コーパスのようにオンラインでしか使えないのは困るという気持ちもあります。ああいうものについては元々の資料が著作権切れのものも多いので、歴史コーパスを CHJ と呼んでいて、オープン CHJ と個人的には思っているのですが、そういうものが、本文は別でいいので、小学館の本文でなくても、いろいろな人が作った本文をかき集めてオープンなものを作るという手もあるのではないかと考えたりはしています。

お答えになったか分かりませんが、こちらからは以上です。

●鈴木 今の最後の方に出た議論は、データをどう引用するかという動機付けの話にもつながってきます。人文学においてもデータ引用の作法というものが、ある意味まだきちんと確立されていません。小木曾先生がコーパスを引用として付ける人はいませんよねとおっしゃっていたのと同じですが、これを本当にオープンにするためには、きちんと、例えば DOI のようなものが振られた上でこのコーパスを使いましたという形で引用ができるのが理想かもしれません。そうするとその引用回数から、これだけ引用されているのだからいいですねという評価にもつながる。小木曾先生が毎日大量に論文集を集めて、1 ページ、1 ページめく

って探さなくてもいいという状況をいかにつくるかという、より大きな話にもつながる問題提起かなと思います。そのあたりは「みんなで翻刻」の出来上がったデータはどう使われていくかというようなところにも関わってくる問題だと思うのですが。

●加納 「みんなで翻刻」での成果物というか、テキストは、CC BY でどうぞ使ってください、ただ、「みんなで翻刻」で作られたテキストですということは明示してくださいとしています。

Slido のご質問の中に、「品質をどう高めていくか、どれぐらいの品質なのか」という質問がありました。 「みんなで翻刻」品質のものであるということを明示して使ってください、あるいは、研究者であればもっと正確なものが必要なら、きちんと自分で見直してから使ってくださいという立場を取っています。今は別に DOI を付けていないので、せめて「みんなで翻刻」と名前を入れていただいて、URL を書いていただくと。もっと言う場合は、例えば橋本さんの博士論文を引用していただくというような形かなと思います。

●鈴木 データ引用に関してはこの SPARC Japan でも繰り返し議論がされていますし、日本でもデータジャーナルが定着していく中で、人文学においてもこれから整理が必要になっていくという一つの証左かもしれないと思います。

次は小木曾先生へのご質問です。「コーパスというのは、それぞれ別個に作られたコーパス内の語彙の横断的な検索（データベース間の横断検索）というのはニーズがあるものでしょうか。また、維持について、関係組織とコンソーシアムを形成するなどして、機関を超えた体制を持つ話などはあるのでしょうか。」

●小木曾 後半については、話はないですね（笑）。ないですけども、とにかく維持していくのに一定のお金が必要で、しかしインフラとしてもう機能してしまっている以上止められないというものになってくる

と、一研究機関でいつまで維持できるのでしょうかねという話があります。運営費交付金は毎年 1.6%減っていますので。そういうことははっきりありまして、いろいろ考えなければいけません。ユーザーからお金を取らなければいけなくなってしまうのか、そうならないように何か機関を超えた体制をつくるのかということは考えなければいけないでしょうね。現時点にはまだ具体的なそういう話にまでは至っておりません。

前半部分の質問「コーパスというのは、それぞれ別個に作られたコーパス内の語彙の横断的な検索（データベース間の横断検索）というのはニーズがあるものでしょうか」については、国語研究所でもそういうことを少し考えています。「中納言」の包括的検索系というベータ版では、時代別、話し言葉、書き言葉別にどんな用例があるか、複数のコーパスを検索できて、現代語のコーパスもいろいろ入っているのです。そのように、横断検索というのも実は作っています。

ただ、横断検索は、それぞれのコーパスの特徴、どういう設計になっているか、どういうデータが入っているのかを知らないで使ってしまうと非常に危険だという側面はあります。「地震」という言葉が、江戸時代では増えていそうなのになぜ検索結果がゼロなのかというと、まだ人情本と洒落本しか入っていないからなのです。そういうところに「地震」という言葉はあまり出てきません。「江戸時代は地震のことに関心がなかった」というのは違います。こういうツールも作っているということで、ちょうど質問に対するお答えかと思えます。

●鈴木 次に、加納先生に対する質問をまとめて二つお答えいただければと思います。『みんなで翻刻』の拡張の際の話を、シチズンサイエンスのスタートと目的に絡めてお話してください」という質問と、「スライド最終ページの『専門家・非専門家の相互のコミュニケーション』という部分（図 4）に関して、現状では、研究者検索サイト等はさまざま存在していますが、双方向のコミュニケーションが可能なツール（仕組み）

がないように感じています。シチズンサイエンスや産学連携などの観点からそのような仕組みがあると有用だとお考えでしょうか」という質問が来ています。この二つは相互に関係する話かと思うので、まとめて回答いただければと思います。

●加納 現在の「みんなで翻刻」に拡張したことに対するご質問については、最初は研究者同士が共同研究するためのツールを想定していたのですが、これはまさに「オープンサイエンス革命」で、今日小野さんが紹介されていた Galaxy Zoo、あるいはベンサムテキストを翻刻するプロジェクトがあるということを橋本さんが持ってこられて、翻刻も、研究者だけではなくもっと広い人々とつながるプロジェクトとしてやっていくのはどうかということになりました。最初の研究者同士というアイデアは中西さんが言われて、オープンサイエンス的なアイデアをそこに橋本さんが加えて今に至るという感じだったと思います。

それは非常に面白いと思ったのと、私は、地震の研究はどこかで社会と触れる部分があるし、それをどんどん増やしていかなければいけないと思っていたので、意味があるのではないかと思って進めていきました。

もう一つは、橋本さんは元々、人文情報学といって、人文学にどう情報学を応用していくかという研究をされている中で、このオープンサイエンス的な取り組み、あるいはクラウドソーシング的な取り組みが研究テーマになると言ったので、それはぜひみんなで協力して

研究データやプロセスの共有

- 同床異夢だとしても、興味・関心を共有できる対象（プロジェクト）の形成が肝心なのは？
 - 大義だけでは動かない 「研究」はそんなに大事？
- 専門家・非専門家の相互のコミュニケーションをうまくデザイン/プロデュースできないか



(図4)

橋本さんの研究としても進めていきたいと思いますという背景もあったと思います。

もう一つは双方向性ですね。例えばここで「皆さんから質問を受け付けます」と言ってコメントを頂く形というよりも、「みんなで翻刻」の中で一緒に活動している皆さんの、Twitter での感想や、機能を改善してほしいというコメントを受けるという双方向性です。双方向の取り方が少し違います。こちらは研究成果やシステム「みんなで翻刻」を提供して、参加者がどうという思いで、どういう参加の仕方をされているかを見るというつながり方をしているということかと思いません。

また、「みんなで翻刻」はフォーラムといって掲示板のようなものがあり、そういうところでもコメントいただけるようになっていたので、それを見ながら、そういう意味での双方向性かなと思っています。バージョン2を公開したときに、どんどんバグ、不具合報告が上がってきて、橋本さんがどんどん対応されました。コーディングするのは橋本さん、運営側かもしれませんが、いろいろな問題点を参加者にも指摘していただいている、システムも実はみんなで作っているという双方向性も生まれていたと思います。

●鈴木 ありがとうございます。次は、いかにも SPARC Japan らしい質問でここで議論すべき質問だと思います。「いいね」が5件集まっています。「大学図書館は人文社会系オープンサイエンスのインフラになり得る」という話題が昨年あったかと思いますが、今回の講演内容ではあまり図書館が登場しませんでした。話題提供では少し触れられましたが、実践的な研究と図書館の関係はどうなる可能性があるか考えをお聞きしたいです」。

昨年、ほぼ同じテーマでインフラの面に注目した SPARC Japan のセミナーを行いました。そのときには、図書館というのは非常に重要なインフラになるという確認をしました。でも今回の皆さんの講演ではあまり図書館が登場しなかったではないかということで、実

実践的な研究レベルとなると、図書館との関係はどうなる可能性があるか、お考えをお聞きしたいということです。

これはまず、図書館の内側にいる中村さんからお話しいただいた上で、研究者側にも聞いていきたいと思うのですが、いかがでしょうか。

●中村 正直なところ、まだ誰もよく分からないというか、解があるわけではないと思います。私も大学図書館に長くいますが、研究と図書館の関係というのは、研究支援というような形になると思うのですが、さまざまな支援の形があって、先ほど言っていたように、画像を出すことでそれが研究の一つの基盤になっているということもあると思いますし、例えば学部生へのリテラシー教育が研究支援の一つであるという言い方もできると思うので、うまくまとめられないのですが、いろいろな関係があり得ると思っています。

私は国文研にいて、今は東京大学にいて、同じ大学図書館でくくれるとは思いますが、そこでも全然違います。最近では研究を支援するのが URA かもしれないですし、図書館職員かもしれないですし、例えばメタデータの生成や管理、研究データの管理で言うと情報基盤センターの職員かもしれないですし、図書館が絶対これというのは解としてないのかなと。それはそれぞれの組織がそれぞれの体制の中で何ができるのかということを考えていく必要があるのだろうと思っています。

ただ、これまで図書館が所蔵資料の管理とデジタル化を行ってきたり、機関リポジトリを運営したりしている大学はとても多いので、そのノウハウがお役に立てるのではないかと考えている人は多いと思います。ですから、その辺の、図書館だから絶対ここが役に立ちます、こうしますということではないのですが、なるべく組織全体のミッションや方向性を見ながら、自分たちが何ができるのかを考えていくしかないのではないかと考えています。その中でなるべくであれば前向きに何かに取り組んでいきたいなと。「できません」

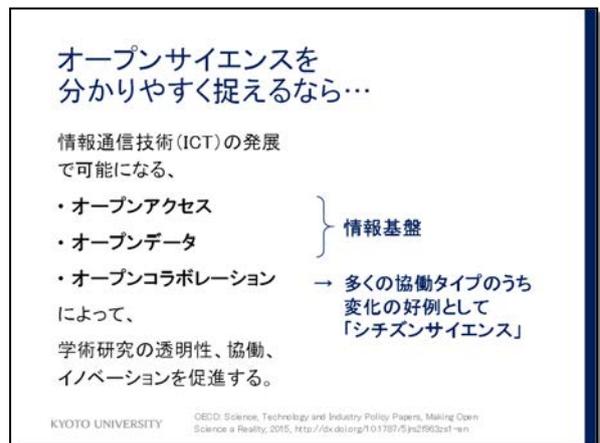
と言うのはつまらないので、「これだったらできます」「一緒にやってみたいのでぜひやらせてください」というような姿勢でいることがいいのではないかと考えています。

●小木曾 先ほど言ったことなのですが、コーパスの基になるデータとして、図書館のオープンデータが十分に役に立っているのだということを改めてお伝えしたいと思います。

また、コーパスの講習など、広めるときに、今われわれは一分野で言語研究なので、直接、研究室の先生たちと話をしますが、応用先が広がれば広がるほど、図書館の役割が増えてくるのではないかと考えています。元々、多目的なものなので。

●加納 オープンサイエンスに対して何をするかとあまり構えるのではなくて、今までどういうふうをやってこられたか。私はインターネットが全盛になる前から図書館にはお世話になって、論文もコピーしに行かなければいけないし、昔の教科書も読まなければいけないしということをやっていて、それがだんだんオンラインになっていったというか、ウェブにのっかっていったという面はあると思います。

小野さんの講演で、情報通信技術（ICT）の発展でオープンサイエンス、オープンアクセス、オープンデータ、オープンコラボレーションが進んでいったというスライドがありましたが（図 5）、まさにそういう



(図 5)

ことで、ウェブなり ICT で広がったことによって、今までやってきたことをどう広げられるか、どう面白くできるか、そういうことではないでしょうか。

研究会などのシンポジウムの機会で、今までよく知らなかった分野の人とつながって、面白い研究が思いつくということがやりやすくなっている面があるかと思えます。そこでどううまく、研究者と図書館の人とあまり区別せずに、一緒にやったらどういうことができるかを議論していくというか、相談しながら進めることができればいいなとは思っています。今すぐ「これをやってください」「これをできます」と、いいアイデアがあるわけではないのですが。

●鈴木 この場に小野さんがもしいらっしゃったら、そこに URA の話を絡めてもう少し議論ができればよかったのですが、残り時間があと 5 分しかないということで、そろそろまとめに入らなければなりません。一つ、質問に対して重要な回答を小野さんがリモートでしてくださっているの、ご紹介したいと思います。質問は、「研究は最終的に『社会実装』『社会還元』がキーワードに求められると思っています。また、その成果を定量的に計測できることも今後ますます求められると考えています（予算取得の観点においても）。人文系研究のオープンサイエンスにおける KPI はどのようなものになりますか。非専門家（異分野）とのコラボ数でしょうか」というものです。

それに対する小野先生の回答は、「KPI は目的によるので、『人社系一般の KPI』は存在しないのではないのでしょうか。まずは何のために市民科学を行うのか明確にすることが重要かと思えます。その上で私が図 6 で挙げた資料などを参考に、どの指標で成果を測るのか検討すればよいのではないかと思います」です。これは市民科学の部分をオープンサイエンスに置き換えても同じだと思います。オープンサイエンスというのは前提として進めていくものだということを、この場のわれわれは共有しているはずですが。ただし、何のためにそれを導入していかなければいけないのかを再

度考えなければいけないということは、まさにこの場で共有しておくべき回答かと思えます。

では、最後に一言ずつ感想を言っていただいて、私が最後に締めたいと思います。よろしくお願いします。

●中村 今回の企画を考えたとき、今は図書館でオープンサイエンスのことをよく耳にするけれども、何をしたらいいのか、何から始めていいのか分からない図書館員が多いのではないかという思いがありました。それで、人社系に寄せていろいろ考えてみたのですが、時代はもうオープン化であるので、それに沿って何ができるかを個々に考えていくしかないかなと思っています。そして図書館の中で閉じこもって考えていても仕方がないので、なるべくなら、私は研究者の方と意見交換しながら図書館の活動を進めていければと思っていますので、ぜひ今後も図書館へのご協力をよろしくお願いいたします、ということで締めの言葉にしたいと思います。

●加納 今日はいろいろ議論ができました。ここで全部話ができるわけでもないし、質問も全部片付かないし、来ていただいている方からの質問も聞けなくて、話をしたい方がたくさんいるのではないかと思います。私は、普段、地震学を研究していて、学会や研究会とは全然違う場、他流試合のようなところに出て、いろいろな人と会って、今後一緒にできそうなことを探してきました。そういうこともあって「み



(図 6)

んなで翻刻」も続けてこられています。これをきっかけに、また新しいオープンサイエンス的な取り組みや、別にオープンではなくてもいいのですが、新しい研究の種が出てくると非常にうれしいと思っています。今後ともどうぞよろしくお祈りします、というのが私のまとめです。

●**小木曾** 今回、自分がやっていることをオープンサイエンスという観点から改めて見直してみたつもりで、それなりのつながりはあったかなと思いました。そういう目で見ても、また加納先生のお話を聞くと、研究者がオープンにならなければいけない部分もあって、こういう違う人たちの場に出てきたり、異分野の先生方と交流したりする中からまたいろいろ生まれてくることがある、オープンになっている部分同士がつながってまた何かできるのではないかと感じました。「みんなで翻刻」の地震データのコーパス化の話など、何かできそうなことがどんどん見えてきて、これから楽しみに進めていけたらなと思っています。今日はどうもありがとうございます。

●**鈴木** 皆さま、ありがとうございました。去年はインフラの話、今年は実践の話という形で話をさせていただきました。

SPARC Japan のセミナー全体から言うと、毎回やっている人文学で多少毛色が違うところがあると思います。どちらかという他の回は、欧米の事例や、今後の枠組みを考えるなど、比較的先進事例紹介と現状の課題確認という面があるかと思っています。一方で人文社会学については「取り組み遅れているから頑張ろう」という話ばかりを繰り返してきたことをいかにげんにやめて、その中でどうオープン化を進めていくかという地道な話を中心に組み立てています。大きな話題を求められる人に対しては「何で個別の地味な話をしているのだ」と思われるような話をしてきたのかもしれませんが、しかし、こういった議論こそが人文学のこれからのオープン化にとって重要なのだろうとわれわれ

企画チームは考えていて、登壇者の皆さまにはまさにそういう話をさせていただいたのではないかと考えております。

これから人文社会系のオープン化をどう動かしていくかということは、われわれ一人一人、そしてここに参加している皆さんお一人お一人で持ち帰って考えていただければと思っています。今後の課題は、より大きな学術全体のオープン化の動きの中で、人文社会系研究のオープン化をどう位置付けていくかということになると思います。これを今後の SPARC Japan セミナーへの宿題として頂く形で、今回はここで締めさせていただきます。最後に、登壇者の皆さまに拍手をお願いいたします。