SPARC Japan NewsLetter
ISSN 1883-826X
NO.40 2020 年 3 月

SPARC Japan ニュースレターでは、各回セミナーの報告に講演やパネルディスカッションを書き起こしたドキュメントを加え、さらにそのほかの SPARC Japan の活動をご紹介します。

※所属、肩書はすべて開催当時のものです。

CONTENTS

■ SPARC Japan 活動報告

学術情報流通推進委員会

学術情報流通推進委員会の会議資料をウェブサイトで公開しています。
http://www.nii.ac.jp/sparc/about/committee/

CLOCKSS の活動について

CLOCKSS は世界の主要な出版社及び図書館による非営利の共同事業で，持続可能で地理的に分散されたダークアーカイブ（アクセスが限定されたアーカイブ）を構築し，ウェブベースの学術文献の長期保存を行っています。出版社から電子ジャーナル等のコンテンツが提供されなくなった場合に，CLOCKSS を通して誰もが無料でそのコンテンツを利用することができます。

日本の支援体制

- 国立情報学研究所は，アジアにおけるアーカイブの「ノード」の一つになっています。2010 年度に CLOCKSS Box と呼ばれるアーカイブシステムを構築し，コンテンツの保存を行っています。また，日本での実務レベルの窓口を担い，日本の参加機関と CLOCKSS との仲立ちを行っています。
- 理事会(Board of directors)メンバーとして国立情報学研究所の武田英明教授が参加し，運営方針等に関する協議・採決に加わっています。
- 大学図書館コンソーシアム連合（JUSTICE）は，CLOCKSS の理念と活動を日本に広めるためのアドボカシー活動を行っています。 https://www.nii.ac.jp/sparc/about/international/

■ SPARC Japan セミナー報告

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～」

2019 年 10 月 24 日（木） 国立情報学研究所 12F 会議室 参加者：54 名

今回は、2018 年度第 4 回セミナーに引き続き人文社会系分野におけるオープンサイエンスをテーマとし、オープンな研究活動が既に展開されている取組に注目しました。事例として、研究者による当該分野の基盤的データの構築と普及や、新たな研究データ作成や研究基盤の構築を目指す市民科学の活動を取り上げ、また様々なかたちで研究データを外部へ繋いでいく役割を担いうる URA の実践についても紹介し、人文社会系分野のオープンサイエンスを幅広く安定的に展開するためのヒントを共有できる企画といたしました。

次ページ以降に、当日参加者のコメント（抜粋）、企画後記およびドキュメント全文（再掲）を掲載しています。その他の情報は SPARC Japan の Web サイトをご覧ください。
(https://www.nii.ac.jp/sparc/event/2019/20191024.html)

企画概要

近年、オープンサイエンスの推進が分野を問わず学術界全体に求められており、人文社会系分野においても今後ますますオープン化を巡る動きが重要になってくると考えられます。当セミナーではこれまでも人文社会系分野に注目したセミナーを実施してきましたが、今年度は実践的かつオープンな研究活動が既に展開されている取組に注目しました。

今回のセミナーでは、言語学分野における基盤データの構築とそれを用いた研究活動振興等の実績がある国立国語研究所の取組と、研究者と市民が協働し新たな研究データ作成や研究基盤の構築を目指す市民科学の実践事例として「みんなで翻刻」の取組を、それぞれご講演いただきます。そして、研究者によるデータ構築と市民科学との間を繋ぐ媒介者としての役割を担いうる URA（リサーチ・アドミニストレーター）の取組についても論じていただきます。

これらの報告から、人文社会系分野のオープンサイエンスを幅広く安定的に展開していくための情報共有を試みます。また、オープンサイエンスと言うと研究者のものと思われがちな面もありますが、図書館職員や大学職員、出版関係者など多くの関係者が、それぞれの立場から「オープンサイエンスの実践」について具体的に考えられるような検討を行っていきます。

パネルディスカッション（左から鈴木氏、小木曽氏、加納氏、中村氏）

参加者から

（大学/図書館関係）
・オープン化したものやそのプロセスを社会の接点として広めていくという話をうかがって、人文系もオープンサイエンスにコミットしていける可能性を実感できました。

・「オープンサイエンス」自体は知っていたが、主に自然科学系向けのものだと思っていました。今回のセミナーで基本的理念（小野氏の講演）と具体的事例（小木曽氏、加納氏）の双方を聞くことができ、人文・社会科学系にも十分に適用できると思いました。

・人文系のオープンサイエンスの可能性を感じました。シチズンサイエンスへの取り組みは、大学にとって必要な活動になるように感じました。

（企業/その他）
・シチズンサイエンスという切り口でのオープンサイエンスについて知見を得られた。

（その他/図書館関係）
・今後、自機関において研究データベースの構築や資料公開を進めていく上で、貴重な実践報告であった。オープンアクセスに対する捉え方や、目的に向けた解決法に関して、多様な経験を知ることができた。

（その他/大学・教育関係）
・市民科学のニーズについて理解できた。

企画後記

2 年連続で人文社会科学系をテーマとした企画をしました。「人社系は遅れている」「独特の文化があるので難しい」といった紋切り型の議論を抜け出す、具体的な動きを紹介できるセミナーになったと考えております。更に一歩進んで、どう先進的な動きにキャッチアップしていくのか、人社系の独自性を新しい動きにどう持ち込んでいくのか、といった議論も進めていきたいですね。登壇・参加いただいた皆様に改めてお礼申し上げます。

鈴木　親彦
（国立情報学研究所 /人文学オープンデータ共同利用センター）

昨年に続き、人文社会系分野におけるオープンサイエンスの回を担当しました。オープンサイエンスについては、私自身まだまだ茫洋としている感じがあります。だからこそ様々な事例を共有することで、オープンサイエンスを少しでも自分のものとして具体的に考えられるセミナーになればと思い企画しました。今回も講師の皆様から多くの情報を頂けたので、今後に活かしていきたいと思っています。

中村　美里
（東京大学附属図書館）

本誌についてのお問い合わせ
国立情報学研究所 SPARC 担当
E-mail co_sparc_all@nii.ac.jp FAX 03-4212-2375
https://www.nii.ac.jp/sparc/

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～」

開会挨拶／概要説明

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)

鈴木 親彦

2019年度SPARC Japanセミナー企画ワーキングメンバー。

情報・システム研究機構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（CODH）および国立情報学研究所（兼務）特任研究員。美術史学・文化資源学・人文情報学を修め、東京大学大学院人文社会系研究科博士課程満期退学後、2017年より現職。研究対象は情報学の成果およびオープンデータの人文学への応用。現在は特にIIIF画像の活用に重点を置いている。

https://researchmap.jp/chsuzuki/

第１回セミナーの趣旨

SPARC Japan ではこれまでも何度か、人文学および社会科学をテーマにしたセミナーを開催してまいりました。昨年、数年ぶりに人文社会系分野をテーマとしたセミナーを開催しましたところ大変好評で、これまでのように年を空けずに毎年動きをウオッチすべきだという声も多く頂きました。われわれ企画 WG といたしましても、人文社会系分野で今後ますますオープン化を巡る動きが重要になってくると考え、昨年に引き続き人文社会系をテーマとしたセミナーを企画しました。

昨年のセミナーでは、人文社会系においてオープンサイエンスを進めるためのインフラをどのように構築していくかという面に重点を置きました。国の政策などの大きな動きから、個別の研究に寄り添った細かな動きまで幅広く講演を頂き、ディスカッションを行いました。今回はもう少し行動の方に軸を移して、具体的な研究活動の蓄積に軸を置き、実践的かつオープンな研究活動が既に展開されている取り組みについて、講演とディスカッションを行ってまいります。

実践に向けて

ごく簡単にこの後の講演の流れをご説明いたします。まずは、研究者によるデータ構築と市民科学の間をつなぐ媒介者としての役割ということで、リサーチアドミニストレーターについての取り組みをご紹介いただきます。続きまして、言語学分野における基盤データの構築とそれを用いた研究活動について、さらに、研究者と市民が協働し新たな研究データを作っていく、「みんなで翻刻」の取り組みについて、それぞれご講演いただくことになっております。

講演の最後に、講演者の皆さまと企画ワーキンググループのメンバーが登壇して、パネルディスカッションを行い、人文社会系分野のオープンサイエンスを幅広く安定的に展開していくための情報共有と問題の共有を行いたいと思っております。オープンサイエンスというと、研究者のものと思われがちな面もありますが、むしろこの場に集まっていただいたさまざまな関

係者の皆さま、図書館職員の方、大学職員の方、または出版関係者の方まで含めて、多くの学術情報に携わる皆さまが、それぞれの立場から人文社会系の「オープンサイエンスの実践」について考えていただける場、そして、ご自身の取り組みにつなげていただける場にしたいと考えております。

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～」

オープンサイエンス的市民協働のために大学ができること

小野 英理

（京都大学情報環境機構 IT 企画室）

講演要旨

オープンアクセスやオープンデータなどの学術情報基盤の変容に見られるようなオープンサイエンスが進展する現在、様々な市民参加型研究プロジェクトが行われている。米国では複数のプロジェクトを包括するオンラインプラットフォームが作られ、環境面で比較的整備が進んでいる一方、日本では研究者の個人的努力に委ねられることが多い。本発表では、日本における市民参加型研究プロジェクトを進める場合に、大学組織としてどのような方策をとり得るのか私論を交えて述べる。特にURA、博物館、図書館、といった研究環境を整備・支援するプレーヤーの役割について考えたい。

小野 英理

霊長類研究の傍ら、ウェブサイト等のデザイン制作を経験。2015年より京都大学次世代研究創成ユニットの研究支援職（URA）に着任し、若手研究者を対象にした支援を行う。2016年よりKYOTOオープンサイエンス勉強会を主催し、市民参加型研究プロジェクトの調査等を実施。2018年より現職。Webを通じた情報発信の実務やその体制構築を進める一方で、科研費研究計画調書を題材にした学術情報のデザインに関する講演を行う。博士（理学）。

本日は、「オープンサイエンス的市民協働のために大学ができること」というタイトルで発表させていただきます。初めに、そもそも私がどういう人間かということについてお話ししたいと思います。

1.自己紹介

専門は元々サルの研究をしていました。サルのお尻がなぜ赤いのかという生物学の研究です。それで学位を取りました。一方で、在学中から二足のわらじでウェブデザインの仕事もしていましたので、経験として研究とデザインを二本柱で持っておりました。研究だけではない経験があったので京都大学で URA（ユニバーシティ・リサーチ・アドミニストレーター）という研究支援職に就き、主に若手研究者支援を行っていました。2018 年から、Web 戦略室という大学の情報インフラを整えるような情報基盤の部署にいます。

一方で、2016 年より KYOTO オープンサイエンス勉強会を主催しています。3 年前になりますが、「オープンサイエンス」というような言葉が出てきて、みんな、「何やそれ」という感じになっていたところで、私自ら勉強したいなと思って、同じような意思を持った方々と一緒になって勉強会を立ち上げました。特に、市民が研究に参加するようなプロジェクトを運営したり、その調査をしている人たちに毎月来ていただいて、話を聞いて自分たちで勉強していこうということで始めています。現在まで 30 回以上、開催しています。

2.オープンサイエンス的市民参加型研究

その月 1 回の講演のキーワードを図 1 に並べてみました。改めて私自身が振り返ると四つぐらいのカテゴ

リーに分かれるのではないかと思いました。

一つ目はボランティアです。私は市民参加型の研究について興味を持ってやっていたのですが、市民ボランティアという人たちがいます。Wikipedia は市民ボランティアによって従来行われている活動です。

二つ目はアカデミアです。アカデミア寄りの活動をされる方々がいらっしゃいます。「みんなで翻刻」もそうです。

三つ目は政策・体制インフラです。政策的なことを考える方にも来ていただきましたし、研究データマネジメント、つまりオープンにする一方でデータマネジメントをどうするのですかというお話をする方にも来ていただきました。

四つ目は企業です。市民がアプリで自分のところの天気を知らせるという、ウェザーニューズの人たちにもお話を頂きました。

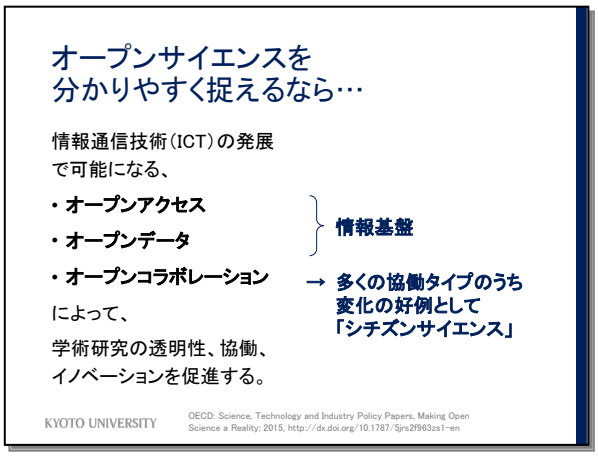
「オープンサイエンス」とは、ものすごく幅の広い言葉で、一言でなかなか説明できないと思います。概念の包含関係も人によって考え方が違うので、どう定義するのかどれが正しいというわけではないのですが、いろいろな意見があります。私は OECD の資料をもとに、オープンサイエンスを分かりやすく捉えるならば、情報通信技術（ICT）の発展で可能になる、「オープンアクセス」「オープンデータ」「オープンコラボレーション」の 3 本柱によって、学術研究の透明性、協働、イノベーションを促進することと言えば分かりやすいのではないかと考えています（図 2）。

ここで図書館関係の方は「オープンアクセス」が親しいところだと思います。「オープンアクセス」「オープンデータ」は、情報基盤をどう整えるかという話かと思います。そして「オープンコラボレーション」は、整えた情報基盤の上でいろいろなステークホルダーがコラボレーションするということが想像されます。今日の話は、市民が参加するというところに興味を持ってお話ししますが、別にそれだけではなく、従来の研究者同士のコラボレーションも入りますし、行政やいろいろなステークホルダーが関わってくるので、そういうものを全部「オープンコラボレーション」と言うということです。

なぜ「シチズンサイエンス」、市民が関わる部分に興味を持っているかというと、こういうオープンサイエンスのようなことが進んでいった一つのグッドプラクティスとして、今までほとんど研究の世界から切り離されていたような人たちが研究に参加できるようになるということが、変化の好例として面白いなと思ったたからです。本日、加納先生からお話しいただく「みんなで翻刻」は、かなり人気も出て目覚ましい成果を上げていると思います。

では、市民が学術研究に関わると何が変わっていくのでしょうか。研究サイクルは、構想、資金獲得、実験・調査・データ収集、発見、公開の五つに分けられると考えています。従来は研究者がどういう研究をするかを構想し、その構想について、科研費なり申請書を書いて資金を獲得しました。税金を使うので間接投

（図 1）



（図 2）

資になります。お金を得ることができれば、実験・調査、データ収集をします。そして何かしらの科学的発見を行い、成果を有料の学術誌に公開していました。

これが現在、資金獲得の部分ではクラウドファンディングのような方法が出てきています。実験・調査・データ収集でも、市民が調査・収集して、クラウドソーシングとも言いますが、オープンデータ化が進んでいます。一方で、科学的発見も市民が実際にしていることがあります。それをオープンアクセスの無料学術誌に載せるということで、サイクル自体が変化しつつあるという状況です。

実際に市民が研究に参加しているようなプロジェクトは何があるかというと、Galaxy Zoo は非常に有名で、天文学の分野のプロジェクトになります。銀河の画像を市民がウェブサイト上で見て、渦がどう巻いているか、銀河の形が何かなど、ウェブサイト上でボタンをクリックして、データをどんどん蓄積していくというものです。eBird というのは、鳥の観察のデータを蓄積しているデータベースです。

「みんなで翻刻」は、昔の資料をみんなで翻刻するというものです。「花まるマルハナバチ国勢調査」は、マルハナバチという種に着目して、全国の目撃情報を集めているプロジェクトです。「ナメクジ捜査網」も同じことですが、外来種のナメクジを探すために市民の協力を得ているプロジェクトです。Safecast は少しだけ性格が違うのですが、福島の原発事故以降に空間の放射線量を測る活動を市民と一緒にやっているものです。

Galaxy Zoo は Zooniverse というプラットフォームに入っているのですが、Zooniverse というプラットフォーム自体は理系に限らず、人社系のプロジェクトも多数扱っております。そのうちの一つが Shakespeare's World というもので、実はつい最近、全てのタスクが終わったところです。

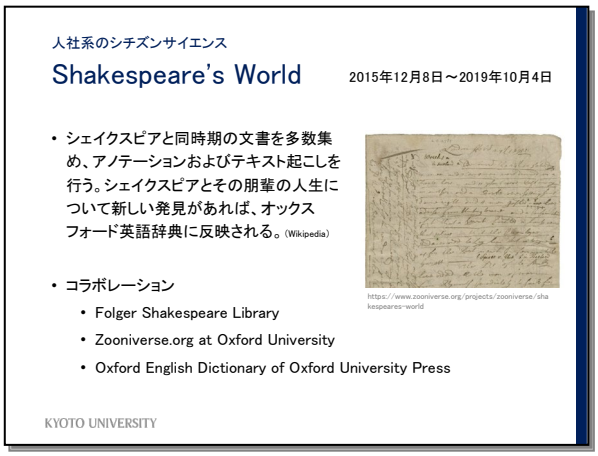
どういうものかというと、Wikipedia には、「シェイクスピアと同時期の文書を多数集め、アノテーションおよびテキスト起こしを行う。シェイクスピアとその朋輩の人生について、新しい発見があればオックスフォード英語辞典に反映される」と書かれています。市民が、図 3 の右にある画像のようなものを読んで、何が書いてあるのかを書き起こしていくようなプロジェクトです。コラボレーションとして、Folger Shakespeare Library という図書館と協力したり、Oxford English Dictionary とも連携したりしています。

3.シチズンサイエンス（CS）に係る政策

こうやっていろいろプロジェクトが動いているという現在ですが、政策的にはどうなのかということについてお話しします。日本では第 5 期科学技術基本計画（2016～2020 年度）が進められています。その中の一つとして、「オープンサイエンスの推進」が挙げられており、「市民参画型のサイエンス（シチズンサイエンス）が拡大する兆しにある」、それをイノベーションの基盤としたいということが科学技術基本計画には書かれております。

2019 年の G7 で、日本学術会議が一緒になって、「G サイエンス学術会議共同声明」というものを出しました。「インターネット時代のシチズンサイエンス」がこの声明の中に含まれており、今後日本においてシチズンサイエンスを推進するための環境整備の必要性を指摘しています。

一方で世界を見ますと、欧州では、Horizon 2020 で、総額 800 億ユーロ、日本円にして 10 兆円の大規模な研究助成の資金源の一部をシチズンサイエンスの助成



（図 3）

に充てるということがうたわれております。

Horizon 2020 の次が Horizon Europe というプログラムなのですが、そこでも、シチズンサイエンスと書いていたか、オープンサイエンスと書いていたかは失念しましたが、関連のものを 3 本柱の一つとして進めるそうです。特に欧州での政策的な動きは活発です。

4.三つの論点

本日のセミナーのタイトルは、「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～」なので、私は実践に向けて何かお話ししなければいけないと考えています。また、ウェブサイトのセミナー概要には、「研究者によるデータ構築と市民科学との間をつなぐ媒介者としての役割を担い得る URA（ユニバーシティ・リサーチ・アドミニストレーター）の取り組みについても論じていただきます」と書いてありました。ということで、「市民科学」ということが一つキーワードになっており、どういう人たちが研究者によるデータ構築と市民科学との間をつなぐことができるのかが主題ではないかということで、論点として三つあると思いました。

まずは、そもそも大学のような組織がシチズンサイエンスに本当に取り組むべきなのかどうかということです。取り組むべきだとしたとして、では誰がそれを担うのかというのが次の論点です。さらに、その人たちはどうやって実践するのかということも論点になってくるかと思います。

一つ一つの論点がかなり幅広に議論できるところかと思いますので、これから私がお話しする話もなかなか全部を包括できるとは言えないのですけれども、私見を交えてお話しさせていただきたいと思います。

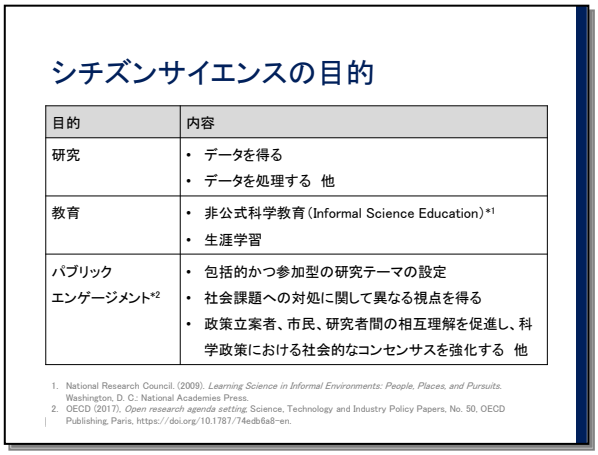
4-1.Why: 大学として CS に取り組むべき？

では、大学としてシチズンサイエンスに取り組むべきなのでしょうか。図書館の方もこれに含まれるかと思いますが、URA をはじめとした研究推進／支援人材が業務として取り組むためには、シチズンサイエンスが組織の方針に合致するべきだと思います。研究者個人が勝手に好きにやるのであればいいのですが、組織として揃えている支援人材が動くためには、その組織の方針に合致すべきだということです。

では、そもそもシチズンサイエンスは何のためにあるのかというと、いろいろな目的があると思います。研究のためにやる、市民に一緒になってもらってデータを得る、データを処理するということがあると思います。教育的な意味もあると思います。非公式科学教育（Informal Science Education）といった言葉がシチズンサイエンスで語られたりしています。私はこの専門ではないので、ご関心のある方は下の参考文献を見ていただければと思います（図 4）。生涯学習だという言い方もされます。

一方で、組織的にはパブリックエンゲージメントというところで、右に書いたようないろいろな政策立案でシチズンサイエンスを取り入れるということもあるかと思います。あとは単純に企業の CSR 的に、大学でも市民と一緒にやるということが良く見えるのであれば、それをやるということもあると思います。

研究に関してもう少し詳細に説明すると、どういう市民の関与の仕方があるのかということで、いろいろなフレームワークがあるのでこれは一例なのですが、この上から下に研究のステップが進むと考えていただいて、市民がどのステップで関与するのかというのが、「貢献型」「協働型」「共創型」で三つぐらいに分かれます（図 5）。今、主に動いているのはサンプル収集



（図 4）

や分析というのが想像しやすいものかと思いますが、「共創型」に行くと、全てのステップで市民の関与があるということも想定できます。

もう一つ似たような分類ですが、四つのタイプがあると言う人もいます。1 番目は「センサーの役割」、単純にセンサーを持って動き回るとか、あまり頭を使わなくてもできるようなこと。2 番目は「基礎的な解釈」を行うもの。「みんなで翻刻」も 2 番に該当します。3 番目は「問題設定やデータ収集」で、問題設定から関わるというものです。4 番目が「問題設定やデータの収集・分析」で、さらに分析も行います。ただ、参加程度が高いほど良いという意味ではないので、それぞれの研究プロジェクトに応じてどのレベルでやるかを意識してやることが大事だということです。

パブリックエンゲージメントについて言うと、最近、「Citizen science and the United Nations Sustainable Development Goals」という論文が出ました。SDGs との関連が述べられているものです。内容は割愛しますけれども、今こういう話も出ております。

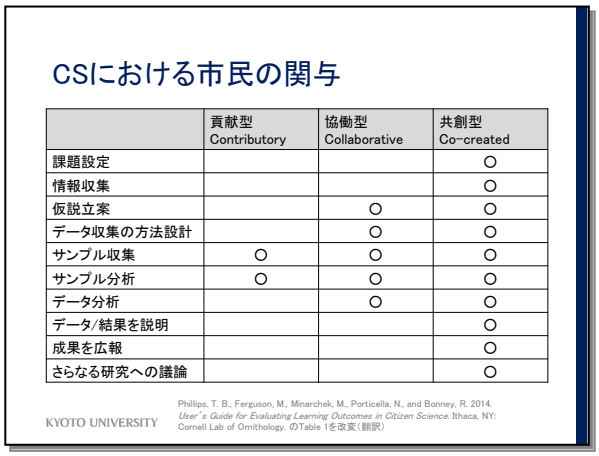
アイデアとしては、そもそも「市民協働」といったときに、市民と本当に協働するということ、それはそれで大事だと思いますが、市民協働を「学術研究のユニバーサルデザイン」として考えるということが大事だと私は考えています。市民といってもなかなか想像しにくいのですが、市民というのはイコール、その分野に関しての非専門家ということなのです。そういった人たちも含めて、誰もが自分の研究領域に参加しやすい環境を目指し、データをどう配布したらいいのか、そのデータフォーマットはどういうものなのか、そういうことを全て考えることが、実は自分の研究の裾野を広げて盛り上げていくことになっていくのではないかということです。市民協働だからといって、市民みんなでパブリックに何でもやっていこうということだけではないと思います。

図 6 の右の写真はユニバーサルデザインの有名な例です（著作権保護のため公開資料から割愛）。ボトルに線が入っている方がシャンプーで、そうでないのがリンスです。目をつぶっていても分かるということで、元々は視覚障害を持たれている方に分かりやすいようにしたデザインでしたが、健常者でも、目をつぶっていても、ぱっと取って分かるということは便利なことです。今まで不自由を感じていた人たちを仮想のターゲットと考えてデザインを検討したらみんながより使いやすくなるという感じで、市民協働を考えてもいいのかなというのが私のアイデアです。

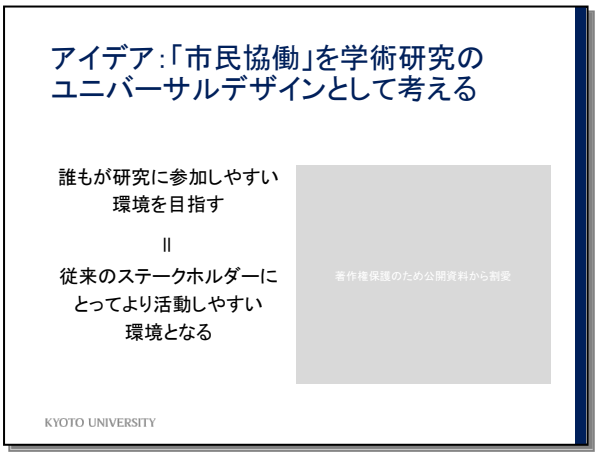
ここまでをまとめます。現状としては、シチズンサイエンスの目的は、研究や教育、パブリックエンゲージメントなど、多様です。将来像は、研究者（運営者）は計画時に目的を見極めることが望ましいと思います。それが各機関の目的に合致するならば、組織として推進／支援することは妥当かと思います。

4-2.Who: 誰が担うのか？

では、シチズンサイエンスを実際に組織としてやっ



（図 5）



（図 6）

ていきましょうとなったときに、URA といわれる人たちが媒介者になるのかどうかというところです。京都大学の状況で言うと、URA は研究力強化のために存在していると考えられています。研究大学強化促進事業で、URA の人数が増えたりしたので、とにかく研究力の強化のためにいるのだということです。

では、シチズンサイエンスを進めることが研究力強化につながるのかというと、まだそこは自明ではないという空気感があるかと思います。従って、現在の URA 業務に市民協働の構築／実践は含まれていないというのが実情です。個別ケースとして、「みんなで翻刻」の広報を支援したという事例はあるのですが、ルーチンの業務としてやるかというと、そうでもないです。では、そもそもシチズンサイエンスを含めてオープンサイエンスの担い手は誰なのかということです。

OECD レポート（2015）では、担い手として、研究者、省庁、助成機関、大学等教育研究機関、図書館／リポジトリ／データセンター、NPO／NGO、出版社、産業界が挙げられています。今日の話に関わるのは大学等教育研究機関、図書館です。

そのレポートに書かれている、大学等教育研究機関の役割は、オープンサイエンスに関するポリシー策定、必要なスキルについて研究者をトレーニングすること、リポジトリの使い方、データの扱い（クリーニング、キュレーション、管理）について教えることです。

レポートに書かれている、図書館／リポジトリ／データセンターの役割は、かなり幅広いです。これらはオープンサイエンスに向けて重要かつ実現に不可欠なアクターだと指摘されています。デジタルな研究資源、つまり出版物やデータ等の研究関連のコンテンツを保存し、キュレーションし、公開し、広めるための活動をすべきであるとされています。研究者が成果をシェア、利用、再利用するためのインフラを構築し、オープンサイエンスを実現させるということで、実際に動くという部分でかなり重要視されています。

このようなアクターが OECD のレポートでは想定されていますが、日本でシチズンサイエンスというときにどういう体制でやられているのでしょうか。「ナメクジ捜査網」は、日本全国のナメクジ、特に外来種を探すようなプロジェクトで、2016 年から開始されています。宇高寛子先生が運営されており、ウェブサイトは先生が自分で作りました。Twitter やメールで市民からの情報を集めています。全て宇高先生一人がチェックし、返信もしています。結構大変なのですが、現状一人で可能な範囲で実施されているということです。

もう一つは「みんなで翻刻」です。これはちょっと面白かったのですが、つい最近ウェブサイトにアクセスすると、「お詫び」と書かれていまして、「開発者の多忙につき不具合の修正が滞っております」ということです。一人の研究者、橋本雄太先生が開発されていて、どうしても一人でやっているとプロジェクト運営に影響が出ることもあります。

「花まるマルハナバチ国勢調査」も同じような話です。富士通の携帯フォト・クラウドシステムや携帯アプリ「ここピン！」を使っていたのですが、それが終了するので、どうするかを研究者がまた考えなければいけません。図 7 の右下は、ホームページに上がっているスライドをそのまま持ってきたものですが、「研究者だって、市民の皆さんの期待に応えたい！　でも、できない！」「研究者は、研究に特化した人間です。マネジメントや普及活動に長けているわけではありません。決して怠けているわけではありません」。やりたいのだけれども、なかなか研究者だけでは難しいの

（図 7）

です。

一方でアメリカではどうかということなのですが、Galaxy Zoo を含んでいる Zooniverse というプラットフォームは、簡単な画像分析であれば、分野を問わず構築可能なサービスを提供しています。Project Builder といって、画像等のデータを見て市民がボタンをクリックするだけでできるようなものであれば、どんな分野でもできます。運営は Citizen Science Alliance というもので、オックスフォード大など九つの機関が参画していて、スタッフは 34 名だとホームページに書いてありました。

eBird を運用している Cornell Lab of Ornithology は、市民による世界中の鳥類の観察記録を集約しています（図 8）。コーネル大学と全米オーデュボン協会によって 2002 年に設立され、年間予算が約 35 億円で、収入が約 36 億円です。シチズンサイエンスにこれくらいのお金が費やされているということです。スタッフも 200 名以上います。右の図は収入の内訳ですが、こういう内容で収入約 36 億円を毎年維持しているということです。

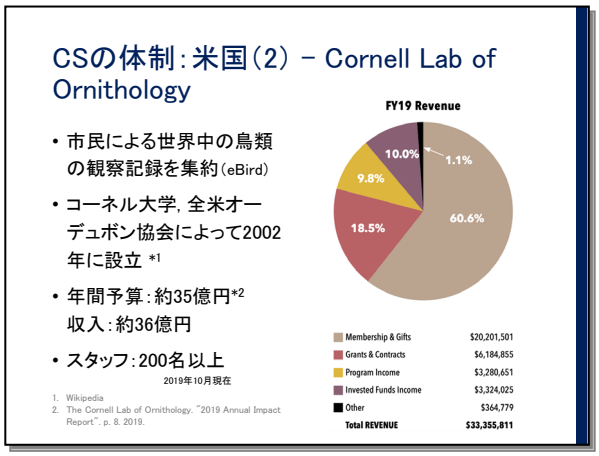
職種もたくさんありまして、ディベロッパー、デザイナー、プログラマー、こういう人たちが充実しています。図書館・博物館との連携に関しては、データをマコーレー図書館にアーカイブしたり、標本をコーネル大学の脊椎動物博物館に保存したりして、周りの組織ともよく一緒に連携しています。

ここまでを総括すると、シチズンサイエンスは、日本では研究者の個人的努力に委ねられているものが多いかと思います。海外でもそういうものがあるのかもしれませんが、少なくとも大きく成功している事例は、豊富な資金と組織的な活動が基盤になっています。従って、大学として、組織として、これからシチズンサイエンスを本当にやるのであれば、URA といっても大学によって意味合いが違うこともありますので、URA のような組織／部署間連携を推進する者が組織的基盤を構築することは妥当かもしれないと思います。また、情報インフラやウェブサイトの構築や維持において、図書館／リポジトリ／データセンターの役割は重要です。一方で、資金の問題をどうやって解決するかということは残ると思います。

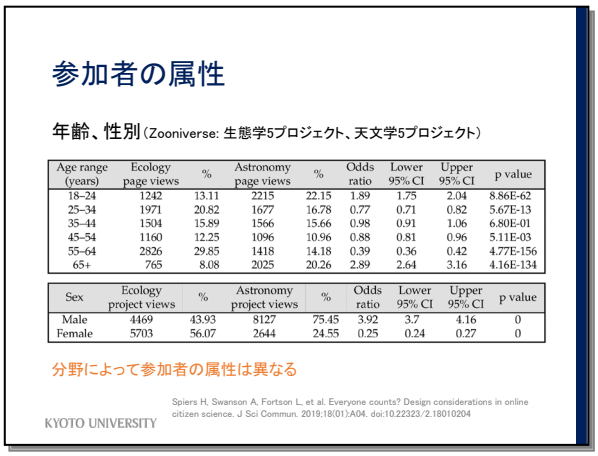
4-3.How: どうやって実践する？

実際にどうやって実践するのかというところで、How については本当にいろいろな角度で話ができるので、三つだけお話しします。まずは、どういう人が参加するのかを知る、その市民が関わったデータの質を担保する、プロジェクト設計のために評価を行うというところが一つの論点かと思います。

図 9 は、Zooniverse に入っている生態学のプロジェクトの年齢別ページビュー、ウェブサイトがどれくらい見られているかというものです。天文学と比較すると、年齢にかなりばらつきがあります。天文学では 65 歳以上が 2,000 ビュー以上である一方で、生態学は 765 ビューと、パーセンテージから見ると全然違いま



（図 8）



（図 9）

す。性別に関しても、生態学と天文学では結構違いがあり、天文学は男性に人気ということがあります。実際やるときには、分野によって属性が違うということも考慮する必要があるかと思います。

では、どういう動機で参加するのかということもありますが、Zooniverse はいろいろなプロジェクトがあるので分野は関係ないですが、どういうモチベーションで参加しているのか参加者に聞いた結果、「貢献」や「興味」が大きかったです。

われわれがその調査をしている中で、スーパーボランティアが重要な役割を果たすのではないかということが分かってきました。スーパーボランティアというのは、通常の参加者よりも熱心に参加して、いろいろなタスクをこなす人たちです。

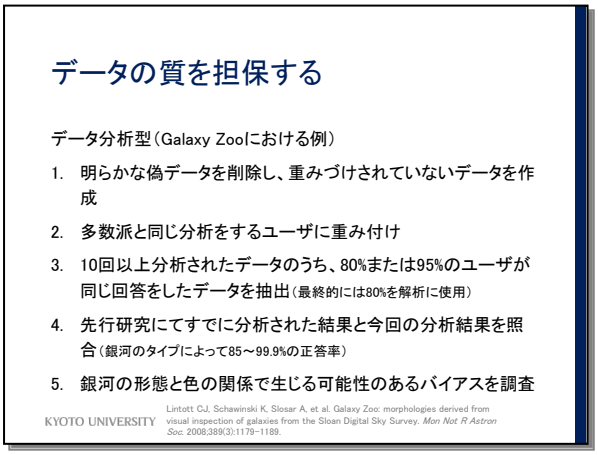
例えば、Galaxy Zoo では 4〜7%のユーザーが全体の 85%のタスクを実施しているというデータがありました。「みんなで翻刻」でも文字数を見ると、約 3%のユーザーが全体の最大 88%を翻刻しているということです。スーパーボランティアの中には、初心者への解説、オフラインでのワークショップ、入門講座のようなものを自分たちでモチベーションを持ってやっているような人たちがいます。こういう方々をきちんと呼び込むというのが、プロジェクトを実施する上では重要かと思います。

データの質を担保するというところでは、上位 5 項目、いろいろな質を担保する取り組みがあります（図 10）。面白いのは、オンラインでやっている割には、紙でデータを提出して、データの質を担保するような取り組みもあるということです。

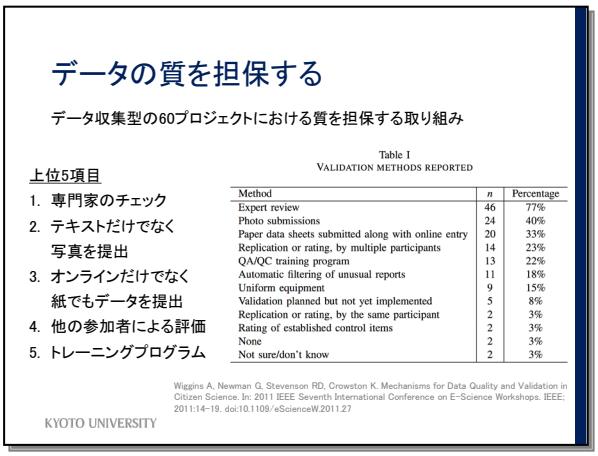
Galaxy Zoo の例では、かなり詳細にデータの質を担保するための検討がされています（図 11）。

最後に、評価というものが大事だと思っています。多くのプロジェクトで、特に日本でも行われていないと思います。評価の目的は、プロジェクトの強み、弱みを見つけること、参加者のニーズを知ること、プロジェクトの成功をステークホルダーに示すこと、さらなる資金獲得につなげることで、プロジェクトを運営するいろいろなステップでこの評価を行うことが重要です。その場合、評価を研究推進／支援職が担うことは可能かもしれないと個人的には思います。初期、実施中、事後、それぞれのタイミングで評価することができます。

図 12 の右手の、コーネル大学の、鳥類のデータベ



（図 11）



（図 10）

（図 12）

ースを作っているところが出しているエバリュエーションに関する資料が非常に参考になります。どういう流れでやるかということを資料としてまとめていただいています。

5.まとめ

最後に、実践に向けてどうしたらいいのかという個人的な考えをまとめます。シチズンサイエンスの多様な目的が組織の方針に合致すれば、業務として URA のような人たちが推進することは妥当ではないかと思います。組織間連携に URA のような人材が貢献できそうです。一方で、実際に情報インフラやウェブサイトの構築／維持は、図書館／リポジトリ／データセンターに含まれる人たちが重要な役割を担うのではないかと思います。どうやって実践するのかについては、参加者の属性や動機を見極める必要があり、データの質を担保するための工夫もきちんと考えなければいけません。そういうことを全部包括して、プロジェクトを設計するために、きちんと始める前から評価を行っていくことがいいのではないか。そのときに URA のような人たちが評価を手助けするということはあり得るのではないかと考えました。

●**フロア 1**　アメリカのシチズンサイエンスの状況は非常に予算があって、大学にいる以外のブレーンがかなり動いていると感じられます。しかし、日本の場合はそういう人たちが非常に多くいるはずだとは思うのですが、メディアに出てきたり、限られた範囲での活躍は見られますが、理想的な形にはなっていないと思うのです。

アメリカのようにできるとは思えないのですが、これからどのように予算を付けたりするのでしょうか。産官学でやることが必ずしもいいことだけではなかったり、いろいろ検証しなければいけない状況にあったり、それこそ本当の意味での第三者的な立場での検証が必要だったりします。具体的にこれからどう進めていく予定か、お願いします。

●**小野**　クリティカルな質問をありがとうございます。日本の状況とアメリカの状況をご紹介しましたが、その間にはかなり隔たりがあると思っています。そもそも近づけていくことが良いのかということも議論はあると思うのですけれども、いろいろなステークホルダーが一緒になって活動していくためにどうしたら良いのかというところについては、私も知りたいところです。

ただ、大学で URA を一時期やっていた人間からすると、少なくとも同じようなニーズを抱えている研究者、それを束ねている組織をつなげていくことはできるのではないかと。そこにお金と人が本当に付いてくるのかということになると、また問題があるかもしれません。

今は日本では、個人的活動がベースになっていて、例えば Zooniverse だったら画像を使って分析するものであれば、どんな分野でもそこに放り込んでいくことができるようなプラットフォームが作られています。そういうものがあれば、自分の研究で市民に画像分析をしてもらいたいと思ったら、すぐ使えるわけです。同じようなニーズを持った研究者が複数いたら、その人たちはみんなそのプラットフォームが使えるという仕組みができています。せめて、いろいろなニーズがある中で、近いニーズを持っている人たちをつなげていって、そこにできれば人とお金が付いてというのが理想かと思っていますが、実際に具体的にお金と人を集めてくるときにどうしたらいいのかということは、どうしたらいいのでしょうねというところです。

●**フロア 1**　アメリカでどうやって生じたかという、そのプロセスの研究をなさいましたか。

●**小野**　それはぜひやりたいと思っているのですが、まだやっていません。例えば Galaxy Zoo は Citizen Science Alliance というものが基本的には母体になって

いるのです。オックスフォード大など、幾つかの大学が一緒になってアライアンスをつくっているので、そういうときに何を最初に各大学が考えてつくりはじめたのかというのはこれから調査したいと思っています。ありがとうございます。

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～」

国立国語研究所の言語資源とオープンデータ・オープンサイエンス

小木曽 智信
（国立国語研究所）

講演要旨

国立国語研究所では、コーパスや電子化辞書、言語地図や方言の調査データや音源などさまざまな言語資源を自ら構築し、保有し、公開している。オープンサイエンスの潮流の中で、これからの国語研では、これらの資源をオープンデータとして公開し、研究者のみならず広く一般に利用できるようにすることを計画している。しかし、コーパス等の言語資源のオープン化にはいくつかの課題がある。その一つは、自己収入の確保の観点から、安易にコーパスをオープン化することができないことである。コーパスは高いコストを払って内製したものであり、かつ IT 企業からの強い需要があるデータである。今日研究機関に求められる自己収入の確保や、研究機関の存在意義にもつながる内製データの保持ということと、研究資源をオープンデータとして広く公開することの間でどのようにバランスを取るべきなのか。国語研の取り組みの現状を紹介するとともに、問題提起としたい。

小木曽 智信

東京大学大学院人文科学研究科修士課程終了、博士課程単位取得満期退学。奈良先端科学技術大学情報科学研究科修了、博士（工学）。明海大学講師、独立行政法人国立国語研究所研究開発部門研究員等を経て、2017年より現職。日本語学会評議員。専門は日本語学（国語学）、コーパス言語学、自然言語処理。バックグラウンドは日本語の歴史の研究で、コーパスを活用した研究を行っている。国立国語研究所でコーパスの開発に携わり、在職中に奈良先端科学技術大学院大学で自然言語処理を専攻。現在は日本語の通時的な研究を可能にする「日本語歴史コーパス」の構築プロジェクトのリーダーを努める。

私の講演の元々のタイトルは「国立国語研究所の言語資源とオープンデータ」でよいのですが、今回の全体テーマに合わせて「オープンサイエンス」という言葉を最後に加えました。データの話だけではなく、研究方法の実践としてのオープンサイエンスの話を少し加えてお話をさせていただきたいと思います。

私は普段は国立国語研究所の言語変化研究領域で、日本語の変化・歴史を研究しています。そこで「通時コーパスの構築と日本語史研究の新展開」というプロジェクトのリーダーを務めています。コーパスというものを作る、そして、使うということは、オープンデータ、オープンサイエンスに少し関わってくるところがあります。

元々の専攻は文学部で人文社会系で、日本語学の人間なのですが、社会人になってから今度は情報系の大学院に行って、自然言語処理について学びました。大学院時代から、古い時代の日本語のコーパス、用例データベースのようなものを作る仕事を行ってきました。

国語研とオープンサイエンス・オープンデータ

国立国語研究所では将来計画委員会というのをやっ

ていて、そこで、オープンサイエンス、オープンデータを次の期の研究所の基盤として進めていこうと言っています。調査・収集してきたデータは公開を原則としオープンデータにする、所定の手続きで誰でもアクセス可能なデータにすることを考えています。ここで、本当のオープンデータであれば当然無料、無制限の利用ということになってくるのですが、そこはこの後、お話しするような事情で完全オープンではありません。

それから、方法の面でもオープンにしよう、検証可能にしようということで、主観を排するということもそうですが、研究に用いた中間的なデータやプログラムもオープンにしていこうとしています。小野さんからお話があった市民参加という意味でのオープンサイエンスもあるのですが、それ以前の話として、研究者間でのデータ共有、データ・方法をオープンにしていくということもオープンサイエンスの重要な側面だと思っています。

それから、コーパスとアーカイブを核とした研究をしていこうということも言っています。今の研究所の中で一番中心になっていることが「多様な言語資源に基づく総合的日本語研究の開拓」です。たくさんのいろいろなコーパスを作って、それを基に研究を進めていくことが中心となっていて、また、外部からも高い評価を得ていることから、そのようなことを考えています。また、今できていないこととして、危機言語、危機方言等の音声データ、録音データをアーカイブ化して、それもオープンにしていきたいと考えています。

今日はオープンサイエンスの実践を進めている方々からのお話ということでしたが、まだそんなにできているわけではないのです。だからこそ、次の期にオープンサイエンスを進めていきたいという話をしています。人文系の研究だと昔からある話ですが、データの囲い込みのような問題、昔で言えば本を見せないというような話から始まって、個人で作ったデータは出さない、カードは見せないというような状況をどんどんオープンにしていきたいということです。

また、言語研究だとしばしば問題になるのが文法性の判断で、こういう言い方は言える、言えないというのが文法の記述で重要になってくるのですが、それが主観的にしか見えないと言われることがあります。これは言える、言えないと言っているのですが、その根拠はというときに、何かが見えない。それはやはり実験など、いろいろな方法でもっとエビデンスを出せるようなものにしていく必要があります。また、それらに限らず、論文の元となった中間的なデータなどが公開されないので、そう言うならそうなのだろうという話になってしまって、検証可能性という面で乏しさが残るといったことがあります。

そういうことを何とかしていきたいということが研究所としての方針でもあり、それを何とか変えていくための取り組みとして、オープンデータ、オープンサイエンスということを言っていきたいということです。

国語研のコーパスとオープンデータ

国語研究所の言語資源と最初に申しましたが、ここではコーパスのことを主にお話ししていこうと思います。言語資源と言う場合、いろいろなものがあり、公開されているものでも、言語地図、社会調査型の調査結果、音源などいろいろありますが、一番中心となるコーパスのことをお話ししていきたいと思います。

「コーパス」という言葉になじみのない方もいらっしゃると思うので、簡単にお話しします。これは、言語を分析するための基礎資料として、書き言葉や話し言葉の資料を体系的に収集し、研究用の情報を付与した大規模なデータベースです。要は、実際の言葉の用例をたくさん集めてきて、研究に必要なだけの情報を付けたものということになります。ただ何でもかんでも集めるのだったら、ウェブをクロールすればいいのですが、そうではなくてバランス良く集めたり、日本語の実態を反映できるように設計の上で持ってきたりします。

また、研究用の情報としては、誰が、どんな人がいつ発話しているというような情報から、全てのテキストに単語の情報を付けて、品詞分解のようなことをし

て、全部に付けるということをしています。今、これが日本語研究、言語研究の中で非常に重要な位置を占めていて、研究するならこういうものが必要だということが今世紀に入ってから常識化してきています。

研究所で出してきたものとしては、2004 年の「日本語話し言葉コーパス（CSJ）」、2011 年に公開した「現代日本語書き言葉均衡コーパス（BCCWJ）」、こちらは 1 億語の日本語の書き言葉を、新聞・雑誌・書籍などからバランス良く集めてきたものです。今、私が中心になっているのは「日本語歴史コーパス（CHJ）」で、これは 2013 年から公開しています。奈良時代の『万葉集』から始まり、明治・大正時代までの新聞・雑誌等を集めてきて、千数百年分の日本語の歴史を検索できる通時コーパス、縦に時代を調べられるコーパスを作っています。他にも、「国語研日本語ウェブコーパス（NWJC）」はウェブをクロールしてくるタイプの大規模な 100 億語のコーパスです。

それから、「多言語母語の日本語学習者横断コーパス（I-JAS）」という、外国人などの日本語を学習している人の日本語を集めたものがあります。さらに「日常会話コーパス（CEJC）」という、日常どんなことをしゃべっているのか、カメラとマイクを入れて食卓・居酒屋・職場などで録音してきて、それを書き起こしてコーパスにするというようなものなども作っています。また、日本各地の方言のコーパス化なども進めています。

これらのコーパスは、国語研究所ではコーパス開発センターというところがあり、ここで公開等を行っています。全てオンラインで利用可能になっています。オンラインですから、オープンといえばオープンで、いずれも無料で基本的には使えるようになっているのですが、この後お話しするような事情で、そうではない部分もあります。

こういうコーパスが日本語研究でどれぐらい使われているかということなのですが、「現代日本語書き言葉均衡コーパス」は、今、登録ユーザーが 2 万人になっています。年間のクエリ数が 50 万件ちょっとです。この辺はぴんとこないかもしれませんが、言語研究者の数は、多く見積もってもそもそも日本に数千人しかいないはずです。ですから、かなりたくさんの人が使ってくださっているということは間違いありません。また、これを利用した論文が年に約 70 本出ています。日本語の研究論文の数はそんなに多いわけではないので、かなりの割合ということです。

同じように、「日本語歴史コーパス」は今、登録ユーザー数が 1 万人になっています。年間のクエリ数、検索の数が 26 万件、利用した論文が大体年に 50 本出てくるようになっています。これも日本語の歴史という非常にニッチな部分での数なので、この分野においては研究に欠かせないものになっていると言っていいのではないか、そういう意味でインフラと言ってもよいのではないかと考えています。

このコーパスは、基本的には無料で、オンラインでの利用という形で公開しています。「中納言」というアプリケーションの中で、マウスでクリックしながら、調べたい言葉を入れていくと検索できます。『源氏物語』に出てくる形容詞を全部持ってくるとか、『枕草子』の中のこれを探そうとか、ある言葉の次に来る助動詞にどんなものがあるかとか、そういう非常に細かい検索ができるので、日本語の研究にとって欠かせないものになってきています。オンラインで無料だけれども、登録が必要な形で公開しています。

1 億語のコーパスやオンラインでの検索環境の提供ということで、コーパスの構築と公開には大変コストがかかります。1 億語の「現代日本語書き言葉均衡コーパス」の場合は、特定領域研究で、科研費で全体で 8 億円、プラス、国語研究所の運営費交付金が、人件費を入れると同額ぐらいになってしまうのではないかという気もしますが、かなりの額をつぎ込んで、また、5 年間丸々かけて作ったものになります。書籍等からバランスを取ったサンプリングをする、このときにはいろいろな図書館にお世話になったのですが、J-BISC からランダムサンプリングして、バランスを取るということをしています。それを電子化して単語の情報を

付けるということをしたものです。

「日本語歴史コーパス」の場合は、2013 年から私が関わっているので、大体、把握しているのですけれども、何しろ奈良時代から明治・大正時代までの日本語の全てに単語の情報を付けなければいけません。つまり、品詞分解、まさに古文の品詞分解そのものを 1,000 万語以上のものに対してやるのです。

もちろん人間ではできませんので、形態素解析という自然言語処理の技術を用いるのですが、大変な手間をかけて、それを後で修正してやるということをしています。それから、外部の画像等にリンクして使えるようにしています。最近、各地の大学図書館等がオンラインで貴重書をどんどん公開してくださっているので、歴史コーパスで検索すると、その言葉が出てきた原本を確認できるという体制が整ってきています。ですから、コーパスというのは、一面ではそれだけで独立しているようでもあるのですが、考えようによっては原本を読むためのツールとしても使うことができて、言葉を探して、それがどう出てきているかというような形で原本を読んでいくという利用の仕方も可能になっています。こちらは年間 3,000 万円弱の国語研究所の予算に加えて、科研費の基盤（A）などを加えて、やはり年間数千万円のコストをかけて作っています。

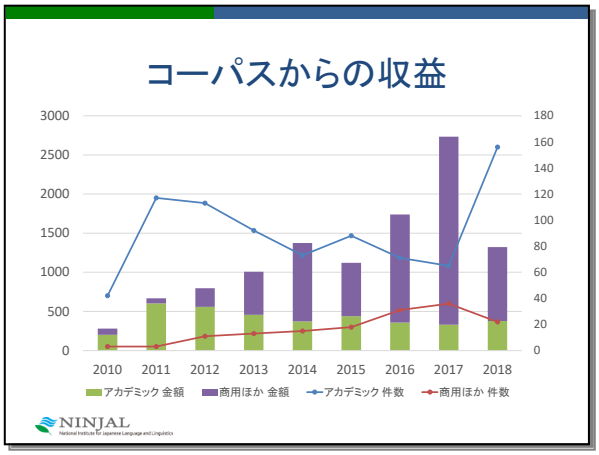
そして、サーバーで「中納言」などを公開していますが、人件費や電気代を置いておいても、サーバーのリプレース等で 1,000 万円近くかかってきます。新しいコーパスを作るというときに、新規性をもって予算獲得をするということはできるわけですが、インフラ化してしまうと、むしろそういうお金が取りづらくなってきて、必要なものであるにもかかわらず、新しく取ることは難しいということがあります。

研究所で、コーパスからの収益は上がっています（図 1）。最初に申し上げたとおり、完全にフリーではありません。「現代日本語書き言葉均衡コーパス」については、特定の社名は出しませんが、GAFA のような IT 企業数社と契約を結んだりしています。ここ 5 年以上の間、年間 1,000 万円以上、全体で収益が上がっています。こういうものがコーパス公開の原資にもなっていくことから、コーパスをオープンデータ化してほしいという話とは相反する部分が出てきます。

コーパスは、コーパスの構築を目的として研究を組むということが行われます。元々、多目的なものなので、一つの目的のためでなく、一回作ってしまえばいろいろできるため、コーパスの構築を目的とした研究がたくさんあります。ですから、副産物ではなく、研究データといっても、研究のために作られたものなのです。それ自体が目的なのです。しかも、コーパスは元々は他者の著作物であるものもあるのですが、それを集めてきて、単語レベルでたくさんの情報を付けるということをしているので、かなり高度な編集著作物ということにもなってきます。

そして、自己収入を生み出すような経済的な価値を持っています。現代語のコーパスの場合は、現代語の自然言語処理のベースになっています。話し言葉コーパスは、今日使われている日本語の音声認識のベースになっていて、Alexa、Siri などの基礎にもなります。

というわけで、なかなかオープンデータにはそのままつながっていかない部分があります。では研究所としてはどうするかというと、所定の手続きを踏んでいただければ誰でもアクセスや利用は可能ですが、無償・無制限であることは意味しません。なぜなら、たくさんの企業から引き合いがあるからというスタンスで、コーパスについては扱っているからです。つまり、オープンデータではないのです。ただ、アカデミック



（図 1）

な料金を設けるなどして、もちろん使いやすいようにしていますし、先ほどのオンライン検索のような場合には完全に無料で利用できるようにしています。これは頑張っているところで、サーバー等はこちらで負担して、でも、学術の基盤としては公開し続けなければいけないだろうと、そういうことをしているわけです。

ですから、オープンデータとしてコーパスを語ることは非常に難しくなってしまうのですが、そうではなくて、コーパスに対して情報を付けていく、「アノテーション」は全然構わないのです。どこにどういう情報を付けるかという、付けただけの情報については本体とは切り離して捉えることができるので、これは無料です。

例えば、コーパスの中のこの単語はこうであるという情報を付けるとか、この文はこうだとか、この発話は誰がしているとか、これはどういう意図で言っているとか、そういう情報を付けていくということは言語研究ではよく行われます。それはやっていいし、付けたデータはむしろどんどん公開してほしいわけです。コーパスの利用価値の向上にもつながるからです。

さらに、もっと高度なものだと、意味情報、全部の単語にこれは食べ物だとか、こういう意味だという情報を付けるとか、統語情報、係り受けや句構造というような文法情報を付けるとか、メタ情報としてさまざまな書籍のレベルの話から人物の話など、いろいろなことを付けていくことも可能です。従って、アノテーションを基盤にしてコーパスに依拠したオープンサイエンスということが可能なのではないかと考えています。そのことをこの後お話ししていきたいと思います。

コーパスとオープンサイエンス

「日本語歴史コーパス」を例にして、コーパスを基礎に、アノテーションということを使いながらオープンサイエンスに近づけるような話をしていきたいということです。

人文学をやっている方、特に言語研究をしている方はよくお分かりだと思いますが、日本語研究をする場合、必ずデータを研究者は持っているはずです。それなしに直感だけで何かを言うことはあり得ないはずで、何かあります。特に今は、先ほどの人数からも分かるように、みんなコーパスを使っているのです。

ですから、コーパスを使って用例を分類したという場合は、研究者の手元には大体 Excel データがあって、用法分類をしたものがあるはずです。ところが、それは個人が持っていて公開されることがないのです。論文にまとめると、論文が完成品なので、基礎となったデータは捨ててしまう、少なくとも出てこないということが非常に多いです。これが非常にもったいないと昔から思っていました。

もう一つは大学院の授業です。演習で何かを読むということをやる場合は、みんなでかなり徹底的に読んで、情報付与のようなことをするはずです。そういったものが消えていってしまうということもありました。

まずは、用法分類のようなデータについて、Excel かどうか分かりませんが、手元にある Excel を共有することができれば、まずは出してもらった研究が検証可能になります。エビデンスが出てくるということになります。もちろん論文の中にも載っているのですが、さらにベースのローデータに近い部分を出してほしいのです。それが出てくると、他の人も共有することができて、新しい研究に利用できるのではないかということです。これをアイデアとして、コーパスをベースにしたオープンな研究を進めていきたいと思っています。このように、コーパスを対象にした研究のデータはほとんどがアノテーションという形でまとめることができるはずなのです。

では、そのためにどういう仕掛けをしておく必要があるかということで、これは非常にプリミティブなところから始めているのですけれども、今までの日本の人文系の研究では必ずしもできていなかったことで、まずはこういうことからやりましょうということで、1 から 5 まで挙げてみました（図 2)。「検索条件式」というのが出てくるのですが、「中納言」で検索するときの式の話、それから、用例にパーマリンクを付け

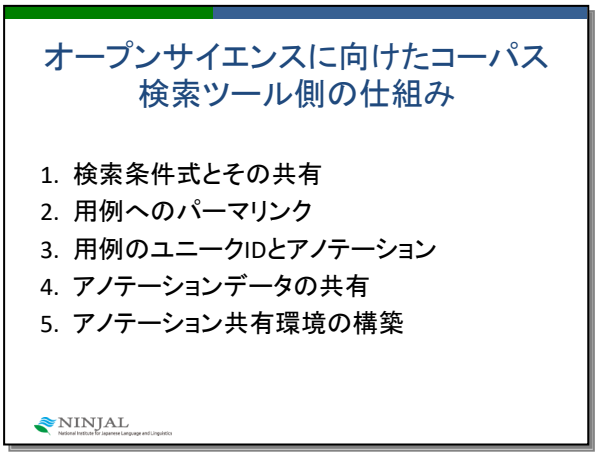
てやる、そして、ユニーク ID を付けてやって、それを使ったアノテーションができるという話になるのですが、順番に見てまいります。

まずは検索条件式です。図 3 は「中納言」というコーパスを検索するアプリケーションで、品詞が形容詞で、活用形が連体形、つまり、形容詞の連体形の次に「言葉」という語が来ている、そういうものをここで探しています。キーが形容詞になっているので、「言葉」というものの前にどんな形容詞が来るか、美しい言葉、優しい言葉、きつい言葉とか、どんな形容詞が使われるかというのを探そうという例です。こういうことが「中納言」だとできます。

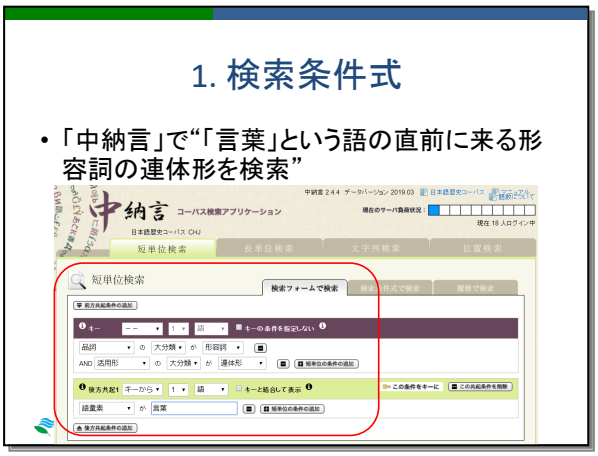
画面で説明するのは大変ですが、内部的には検索条件式というのがあって、これをすぐ表示することができるようになっています（図 4）。一見すると、これはちょっとやっかいなのですが、そんなに難しいものでもありません。コピペしてメールなどで送って、「これでやると用例が取れますよ」とか、「これだとこんな結果だけれど、あなたのと用例数が違うのですが」ということに使えます。研究の再現性、用例の共有のベースになると思います。ですから、まずこれを出すようにしました。

「中納言」の講習会では必ずこの話をして、研究の再現性のために、論文に表示するときには検索条件式を貼りましょう、そうすれば読んだ人が研究を再現できますからとよく言っています。コーパスを使っている時点で研究データの共有は一定のレベルであるわけです。それをどう使ったかという検索条件式が加わると、かなりのレベルでの検証可能性が出てきます。とんでもない間違いをしているというのはあり得る話で、それが抑止できるし、良いものにしていけるのではないかというのが一つ目の話です。これは実際に、かなり以前から「中納言」に組み込んだ機能で、使われるようになってきています。

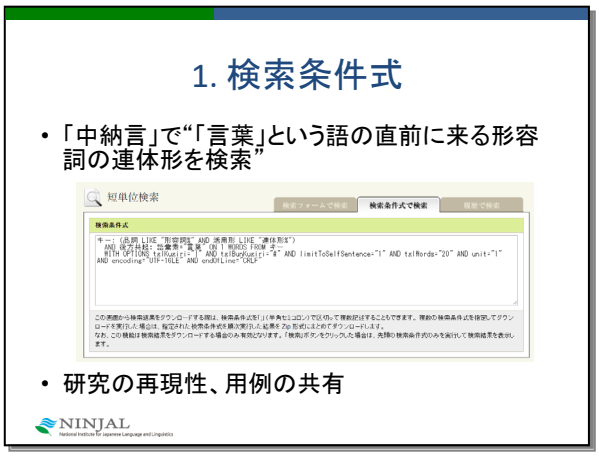
さらに、用例へのパーマリンクを作りました。個々の用例、つまり全部の作品が品詞分解されていますから、『源氏物語』の「桐壺」の最初のところに出てくる「やんごとない」という形容詞に ID が付いていて、それをクリックすれば「中納言」上で表示できるというパーマリンクを用意しました。これがあれば QR コードなどで出せます。「中納言」のアカウントがないといけないのですが、それさえあれば、リンクをクリックすれば用例が表示されます。



（図 2）



（図 3）

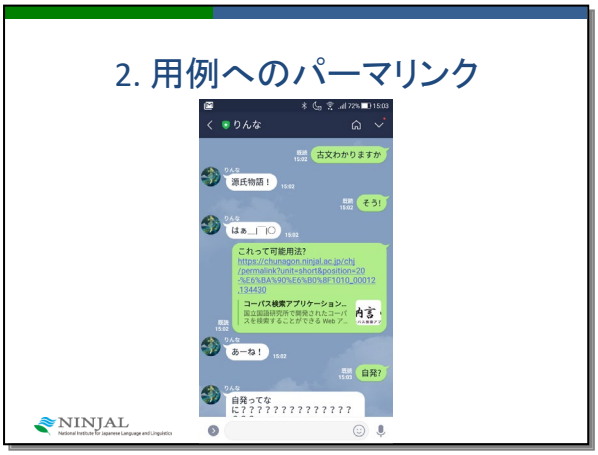


（図 4）

用例が表示されると何がうれしいかというと、まず単語の情報が付いています。品詞など、そういうものが付いていて、文脈が分かるだけではなく、ジャパンナレッジや各大学の原本データへのリンクが付いています。これを使うことで、「この用例なのだけど」という話ができるわけです。SNS でもできるので、みんなやろうよと日本語学会ではよく言っています。

図 5 は LINE の画面ですが、私は友達がいないので、マイクロソフトの AI 女子高生とやりとりをしたところを表示しているものです。要するにこうやると、「この用例はどうなの」という話がお互いにできるのです。この基盤が今までなかったわけです。もしやろうとすると、「『源氏物語』の新編全集の何ページの 3 行目なのだけれども」という話をしなければいけなかったかもしれませんが、これがあれば、そこを出して、「これはこうじゃないの？」という話ができます。基礎になるようなものがまずなければいけないだろうということで、作った機能です。このパーマリンクは何とか維持していかなければいけません。維持するのは大変で、勝手にずらしたりしないようにしなければいけないのですが、そういうものを頑張ってやりはじめました。

それから、今の ID にそのまま使われていることなのですが、パーマリンクに組み込まれているのはユニークな ID です。われわれは「サンプル ID と開始位置」と呼んでいるのですが、これは先頭から何文字目かという情報です。ですから、本が変わらない以上、



（図 5）

ずれないのです。「これは用例のマイナンバーだから、みんな使ってね」ということを日本語学会へ行ってよく説明しています。先ほどの「中納言」の検索結果で表示される部分です。ID さえあれば他の人は表示できるから、文脈や品詞などがなくてもいいからこれは出してほしいと言っています。これがあると何がいいかというと、必要な用法を番号だけ並べてやれば、特殊な用例集のようなものができることです。

ちなみに、先ほどの AI 女子高生とのやりとりは、『源氏物語』の一部分の「海見やらるる廊に」で、「るる」「られる」なのですが、受け身、尊敬、自発、可能のうち、「海を見やることができる」と訳せるから、可能に見えるのです。しかし、中古に可能の肯定用法はないはずだといわれているので、どう考えるかというと、「自然と見渡される」という自発ではないかといわれています。議論の対象になるものをリストアップするだけで、日本語研究者にとって貴重なデータになるのです。

ちょっとおかしな自発用法のリストとか、さらにその横にこれは自発だと一言付けてやれば、それはアノテーションとして使えるようになって、「られる」にこういうものを並べて、受け身、尊敬、自発、可能のどれかを自分で付けましたというデータができれば、結構使えるデータになるのではないかと思います。

コーパス上では助動詞の「る」「らる」という情報しか付けていないのです。それが受け身、尊敬、自発、可能かというのは研究者によってもかなり議論が分かれる部分があるのですが、これを付けてやると、例えば小木曽（2019）のデータで可能用法は何例あるとか、小木曽（2019）のこの ID の可能は間違っているとか、そういう議論につなげていけます。このデータを共有することで、他の人はその次に来る可能用法の動詞だけを持ってくるなど、そういうことができるので、みんなでどんどん共有しようよと言い歩いています。

要は、Excel のデータなのです。日本語の研究者はみんな多分、これを作っていたのです。そうでないと、可能用法が何例あっても出せないはずなので、これを

集計しているに違いないというか、私も実はやっていました。ですから、これを表に出しましょうということを一生懸命みんなに勧めているところです。

できたらそれを公開しましょうということも言っていて、自分もやらないわけにいかないので、隗より始めよということで始めました。researchmap の「資料公開」のところで実際にこのデータを公開してみました。そんなにどんどん利用が進むようなものでもないのですが、まずはそういうところから始めることで、研究用のデータの流通、共有、再利用が進むのではないかと考えています。

アノテーションのデータをオープンデータとして公開しようとしています。アノテーションを引用しようというのも、先ほどから小木曽（2019）と言っているのですが、これを作るのは結構大変なのです。何しろ、平安時代後の全部の「れる」「られる」について、受け身、尊敬、自発、可能を付けるという話になるので、これだけで一つのプロジェクトになってもおかしくないような話なのですが、そうやって作ったデータはちゃんと引用しましょうということも言っているわけです。そうしないと、オープンにしてデータを出してくれる人がモチベーションを保てませんし、作ったデータを再検証していく、再利用していくときに他の人が見られないということです。こうやっていくことで、コーパスをみんなで育てることができるのではないかと考えています。

コーパス本体が必ずしもオープンでなくても、アノテーションをベースとして公開していくことで、それが可能なのではないかということを言いたくて、こんなことを春の日本語学会で一生懸命主張してきて、まだそれほどでもないかもしれませんが、それなりの反応は得ているところです。コーパスを育てていくことをしたい、アノテーションを共有して、学会の共有財産になるようにしていきたい。この作成・公開は研究の業績として正当な評価を得られるようになってほしいということです。

そういう話をしていると、これはもっと簡単にできるようにしないと駄目なのではないかという気がします。ダウンロードしたデータで Excel 上に横に入れてやる分にはいいのですが、それを共有したとしても、それを再利用するのは大変なのです。先ほどの Excel データのような一工夫が要ります。VLOOKUP ぐらいは使えないと駄目、Access が使えればいいのだけれどとなってくると、いきなり人文系の研究者は「それは困る」という話になってしまいかねないのです。だから、それはアプリケーション上で実現したいし、もっとサポートできないかということで新しく始めたいと思いました。

そして昨年の今頃、一生懸命、科研費の書類を書いて、この 7 月に幸い採択されたので、そのお話を最後にしようと思います。

「挑戦的研究（開拓）」というものです。3 年間かけて、「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」と、すごく大きく出たタイトルなのですが、つまり、先ほど私がお話ししたようなことで、研究者間のアノテーションとしてのデータ流通を行いたいのです。先ほどから話が出ている「みんなで翻刻」の橋本さん、人文情報学研究所の永崎さん、歴史民俗博物館の後藤さんなど、いろいろな分野の主立った方に入ってもらって始めました。

全く新しいものを作っても難しいので、既に 1 万人、2 万人ユーザーがいる「中納言」に新しい機能を追加するという形で、アノテーション機能を追加するということを始めたいと思っています。本当はもう少し画面など、できたものがあるといいのですが、現時点では、お金が来ただけで、頑張ってこれからやるところなので、まだ何もありません。例えば「中納言」上で間違っている部分を指摘するということでもいいと思うのです。それをきちんと使えるようになれば、「みんなで翻刻」ではなくて、「みんなで品詞分解」という話になります。

それから、先ほどのような情報をさらに追加することもできるといいなと思っています。付けたら、それを他の人が画面表示できるようにして共有できるよう

に、引用できるようにしていく、そういう仕組みを作りたいと考えています。こういうものが実践できるようになると、もっとコーパスを使ったエビデンスベーストな、オープンな研究環境ができていくのではないかと期待しているところです。これから頑張ります。

まとめ

最初、オープンデータの話と言ってしまったのですけれども、実は日本語コーパスはあまりオープンではありません。CC ライセンスで言うと、一番緩いものでも SA が付いているぐらいです。ND が付いているものも多いですし、そもそも CC ライセンスを付けていないものが多いです。それは先ほど申しましたような理由です。

だから、コーパス本体はなかなかオープンとして出せないのですけれど、アノテーションという切り離した関係にはなるのですが、それでオープンサイエンスを推進する基盤となし得るのではないか、それを拡張して、環境の整備を今後進めたいと考えているところです。

●**フロア 1**　勉強になるお話をありがとうございました。東京外国語大学附属図書館の職員です。

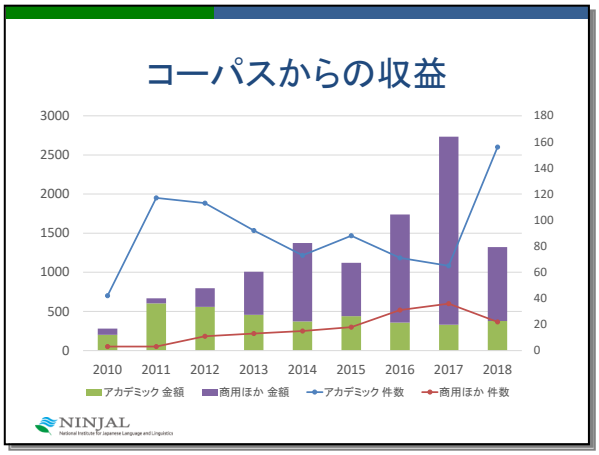
当館でもこの間、大学院生と「現代日本語書き言葉均衡コーパス」の使い方のガイダンスを作ったばかりで、本学の研究分野では本当に欠かすことができないインフラだと思ってお話を聞いておりました。

どうしても商用利用というか、有料化は外すことはできないというお話だったのですが、「コーパスからの収益」のスライド（図 1）を拝見すると、アカデミックユースが一定数であって、最初はアカデミックが多いのですが、2017 年から商用ユースがものすごく増えています。先ほどの登録者のお話でも、2 万人だけれど研究者は数千人しかいなくてというお話をされていました。アノテーションという参加をするのは多分、研究者がメインだと思うのですが、コーパスを育てている中のコミュニティの人からも現行ではお金を取っていて、商用ユースの人からも取っている、多分、価格の差はあると思うのですけれども、そういう状態かと思うのです。

それをせめてアカデミックユースは無料化するということで、アノテーションを付けるコミュニティとしては、一種、国語研究所を超えた、コンソーシアムではないですが、共同体として研究者の世界ではオープン化する、そういう発想はあるでしょうか。

●**小木曽**　実は、先ほどのアカデミックなのに有料という部分は、「現代日本語書き言葉均衡コーパス」と「日本語話し言葉コーパス」だけのケースです。これはどういうものかというと、「中納言」のようなオンラインではなくて、ディスクに入れて生データをそのままお渡しするというタイプのものなのです。これについては配布の費用もかかるということもあるのと、そもそもこれのアカデミックライセンスは相当安いです。しかも、研究室単位、大きな単位で使えるものなので、企業などと比べると相当廉価にはなっていると思います。

そうはいってもオープンにできないかというときに、結局のところ、オンラインに置けないのです。パッケージで配るということにはなってしまっていて、アカデミックにオープンということは完全にはしづらいかと思っているのですけれども、お答えになったでしょ



（図 1）

うか。

●**フロア 1**　ありがとうございました。パッケージのものをオンラインに置けないというのは、容量的に大き過ぎてということですか。

●**小木曽**　流通してしまうのではないかということが基本だと思います。二つのコーパスについてはそのようなことなのですが、近代語のデータなど、一部のものは CC ライセンスで、オープンとは言えないかもしれませんが、公開しているものもございます。

歴史コーパスの方は、また別の理由で完全公開ができなくて、中世以前のデータの大部分が小学館の『新編　日本古典文学全集』のライセンスといいますか、契約の下で使わせていただいているということがあって、われわれにもできないということがあります。同じようなことで言うと、「現代日本語書き言葉均衡コーパス」については、これも原著作者がいるわけですけれども、その人たちに許諾を得たときに有料での配布ということも書いてしまっていて、いろいろと権利関係、原著作権者との関係があるということです。

●**フロア 1**　大変よく分かりました。ありがとうございました。

●**フロア 2**　東京大学の教員です。コーパスで人文系では最大級のデータセットということで、非常に素晴らしい活動だと思っております。毎年 50〜70 本ぐらいの論文が出ているというお話だったのですけれども、これは一体どのようにして実際に使われているかというのを捉えられているか、そのプロセスが分かりましたら教えてください。

●**小木曽**　本当はコーパス利用の契約書の中に、論文を書いたら送ってくれ、少なくとも情報を送ってくれということを書いているのですが、送ってくれません。来ることはありません。また、コーパスのようなものは、特に人文系の研究では参考文献とか、引用するという習慣があまりないのです。本文中に書いてあるとか、脚注に付いているとか、言及があればいい方です。中には、「国語研の現代語のコーパスで検索したところ」などと、固有名詞なしの書き方しかないこともあります。われわれは仕方がないので、基本的には Google Scholar のようなところはもちろん、国立情報学研究所のデータベースの他に、論文集など、そういう主立ったものは見て、それでリストアップするという作業を毎年、年に 2 回やっています。その数字ということになります。

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～」

「みんなで翻刻」にみる歴史地震研究への非専門家の参加

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）

講演要旨

市民参加型のオンライン史料解読プロジェクトである「みんなで翻刻」（https://honkoku.org/）は 2017 年 1 月のリリース以来、5000 人を越える参加者を得て、東京大学地震研究所が所蔵する史料のうちデジタル公開されているもののほとんどを翻刻するなどの成果を挙げている。2019 年 7 月にはシステムを更新し、さらに多様な史料の解読により気軽に参加できるようになっている。「みんなで翻刻」には、学習や楽しみをベースとしたシステム設計など、参加と継続のためのさまざまな仕組みが施されている。これらを紹介しながら、歴史地震研究への非専門家の参加について考察する。

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准教授。京都大学大学院理学研究科地球惑星科学専攻博士後期課程修了。博士（理学）。京都大学大学院理学研究科21世紀COE研究員、京都大学防災研究所助教を経て、2018年7月より現職。主な研究テーマは歴史地震や歴史災害。近著に『京都の災害をめぐる』（小さ子社刊。大邑潤三氏との共著）。ツイッター@KanoYasuyuki。

1.はじめに

本日、主にお話をする「みんなで翻刻」というのは、私だけではなく、小野さんのご発表で「たった一人の開発者」と言われた、今は佐倉にある歴史民俗博物館におられる橋本雄太さんや、私たちが京都大学でやっている古地震研究会のメンバー、「みんなで翻刻」の最初のアイデアを出した中西一郎さん、そういった皆さんとずっとやってきているプロジェクトです。今日は私がお話ししますが、こういう皆さんや、「みんなで翻刻」に参加いただいているたくさんの方たちを代表してお話しすると考えていただければと思います。

私は今、東京大学の地震研究所に所属し、地震火山史料連携研究機構に兼務しております。連携研究機構というのは、東京大学の中で部局を超えた研究連携、共同研究をするためにつくられている仕組みです。この地震火山史料連携研究機構は、私たちの地震研究所と史料編纂所の教員が一緒に研究をしている機構です。

私の専門は地震学です。中でもここ数年は「歴史地震」という、歴史上の地震、過去に起きた地震について調べています。1 年半ぐらい前までは京都におりまして、京都大学の防災研究所に勤めていました。京都にいた頃から古地震研究会という、昔の地震を調べる、あるいは昔の地震を調べるために必要な史料の解読の勉強をする会をやっています。その活動の中から「みんなで翻刻」が出てきて、それの運用にも関わっています。

今日のお話としては「みんなで翻刻」が中心になるのですが、それ以外にもオープンな取り組みに興味を持っていて、敢えて自分の周りにあるものでオープンと言えるかなというものを幾つか紹介したいと思いま

す。それから、市民科学（シチズンサイエンス）、オープンサイエンスといったときに、いきなり市民の方というだけでなく、非専門家、自分の専門以外の方と一緒に研究する、あるいは研究活動をするということがどういうことかを考えてみたいと思います。

ちなみに、客層の把握も含めて皆さんに聞いてみたいのですが、「みんなで翻刻」をご存じだという方はどれぐらいいますか。ありがとうございます。では、ご存じで、翻刻もしたことあるよという方はどのぐらいですか。はい。それから、専門家・非専門家という意味で、私は地震の研究をしているのですが、地震や災害の専門家はこの中におられますか。いない、しゃべりやすいですね。逆にもう少し広く、人文科学は専門と言ってもいい、昔勉強したことがあるなという方はどのくらいおられますか。これもそうでもないですね。翻刻、昔の人が書いたものが読めるよという方はどれぐらいおられますか。あまりいないですね。分かりました。ありがとうございます。

今日お話しする「みんなで翻刻」に、皆さんがもし参加されるとしたら、地震学の専門家でもないし、翻刻や解読するということも別に専門でもないし、習ったこともないという立場で、非専門家として参加することになります。そういう意味で今日は、市民と研究者ではなく、専門家・非専門家という言い方をします。

2.身の周りのオープンサイエンス的な取り組み

さて、いきなり宣伝で恐縮なのですが、図書館関係の方が多いと聞いたので、図書館でぜひ買ってくださいという話です。『京都の災害をめぐる』という本をつい最近出しました（図 1）。写真を載せて、京都のいろいろな場所、災害にゆかりがある場所、ありそうな場所を取り上げて、観光ガイド風に紹介しています。ほぼ全ての地点に写真が付いています。

この本の中には、私たち著者、あるいは編集に関わった出版社の方以外の方が撮られた写真が数枚載っています。今年 5 月にまち歩きのイベントをやって、参加者の方にいい写真が撮れたら送ってくださいと頼みました。あるいはまち歩きの範囲に入っていないけれど、遠いから写真を撮りに行くのが大変だと思っていたようなところに行って写真を撮ってきてくださるなど、そういういいことがないかなと思って、投稿を募りました。そうしたら何枚か送っていただけました。

こういうものを撮りに行っていただくことで、私たちが書いた本、あるいはこの本の中に書かれている京都の災害に興味を持っていただけるといいなと思って、こういう取り組みをしてみました。ゆくゆくはこの本がきっかけになって、本はページ数の制約があるので、もっと知らない、面白い地点があるかもしれないのですが、そういうところを口コミ的に、あるいは投稿のような形で教えていただいてまとめることができたらと思っています。

「古典オーロラハンター」は、私は企画に携わったわけではなく、ちょっと呼ばれて行って、昔の地震の解説をしました（図 2）。国文学研究資料館と国立極地研究所で主にやっておられる、古典籍からオーロラを探すというプロジェクトに、市民の方に参加していただくというものです。京都大学の図書館にある会議室に研究者と非専門家の方に集まっていただいて、昔の人が書いた古典籍、ここは翻刻をするというよりは、本になったものの中から、オーロラっぽいもの、あるいは地震について書いてあるのではないかというところをぱらぱらめくって見つけていただいて、書かれていることはどういう天文現象だったのか、どういう地

（図 1）

震だったのか、例えば、僕は地震について解説するというようなイベントでした。

図書館という場もいいのです。大学の図書館なので、一般の方は少し敷居が高いかもしれませんが、図書館は本が好き、何かを読むのが好きな人が普段からおられて、あるいはイベントの開催場所としていろいろな講演会も行われています。そういうところに、宣伝をするとたくさん人が集まってくる、そういういいことがあるなと思って、こういう取り組みにも顔を出しました。

「満点計画」は、私が京都大学にいた頃の同僚がやっているプロジェクトです（図 3)。1「万」カ所の地震観測「点」をつくりたいということと、「満点」が掛けてある計画です。例えば 1 万点の地震の観測をしようと思うと、なかなか研究者、専門家の力だけでは大変です。ものすごくお金をかけて外注してできるかもしれませんが、これはそうではなく、ボランティアの方に参加していただいて、専門家と共に地震観測に参画していただき、山に行って地震計の装置を置きます。これは非専門家の方とやるのですが、装置自体はプロ仕様、専門家が使うものと同じものを使います。同じやり方をして、もちろん練習もして設置しています。

実際にやっておられる方は、あまりオープンとか、市民参加とか言われないのですが、端から見ていると、いわゆるシチズンサイエンス的な取り組みだと思って、ご紹介しました。最近は設置するだけではなく、この地震計で取れたデータを解析する、分析するというところにも非専門家、元々は専門でなかった方、ボランティアの方が参加してくださるということになってきています。

こうなってくると、例えば学部生で研究室に来たばかりの人より、この人の方が地震のことがよく分かるわけです。そういう意味では、専門家と非専門家とは何なのだという話にだんだんなってきます。こうやって一緒に研究活動に参加しているうちに、ある部分に関しては本当に研究者に近いところまでできるようになったり、あるいは地震に関して、災害に関して深く考えていただくようになるということかと思っています。

「地震計記録のデジタル化プロジェクト」は、ハーバード大学で地震学の研究をしている石井水晶さんがやっておられるプロジェクトです。ハーバード大学に限らず、京都大学でも東京大学でも過去の地震の記録はたくさん持っています。過去の記録は、紙に記録されています。何らかの形で、ペン書きでも何でもいいのですが、地震の波形が紙に書かれています。今だったら全部デジタルで記録されて数値データになるのですが、ある時期以前は全部紙です。紙の状態だと、今の現代的な分析をするのはなかなか大変なので、これをトレースして数値データに変換する作業が必要になります。機械でもある程度、画像認識でできるのですが、地震波形の中に飛びがあったり、線が重なっていたりしていると、なかなかうまくいかないのです。き

（図 2）

（図 3）

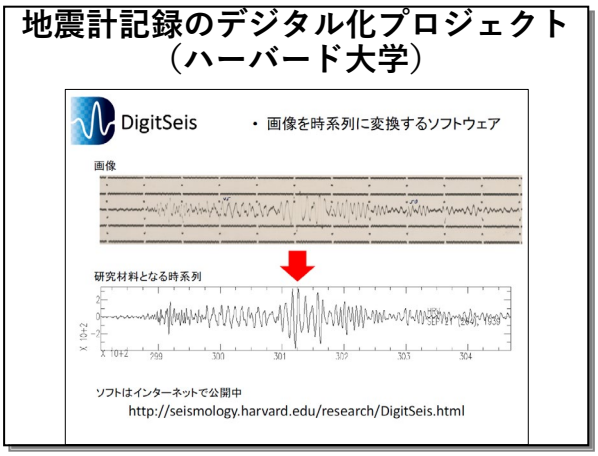
ちんとした地震波形として数値情報に変換するには、やはり人手の作業が必要であるということです。

ここを石井水晶さんたちのグループは日本の高校生と協力して、数値化するというプロジェクトをやっています。専用のアプリを使って高校生が作業すると、きれいな数値の波形データになります（図 4）。これのいいところは、高校生に地震の勉強をしてもらおうというだけでなく、出来上がった数値データをどんどん積み重ねていって、ダウンロードできるようにされています。地震の専門家がダウンロードして解析に使える、昔起きた地震の数値的な分析に使える状態になる、こういうところまでやっている取り組みです。これも非常にオープンなというか、研究室だけでやるのではなく、外側の人たちと一緒にやるということです。

このためにこれがうまくできるようなアプリを開発したり、いろいろな工夫はもちろんしているのですが、こういうプロジェクトが行われています。これは私も聞いてすごいなと思って、皆さんと共有したいと思って今日ご紹介しました。

さて、今度はまた私がやった話に戻るのですが、今年 9 月に地震学会の秋季大会がありました。そこで「オープンデータと地震学」という特別セッションを提案して、採択されてセッションをしました。口頭発表 17 件、招待講演もあって、ポスター発表 4 件と、地震学会の中では結構大きなセッションになりました。

元々、地震学の中でもオープンデータを進めていきたいと思ったのは、アメリカの地震観測網に関わっておられる方の論文がきっかけです。データをオープンにすることで、しかも例えば DOI、識別子を付けてデータを公開することで、生産したデータを、再現性はもちろんなのですが、引用できるようにして、論文と同じようにデータも引用して、データを生産した人の成果としてきちんと評価できるようにすることで、将来にわたってより良いデータを生産していけるようになる、生産する立場の人が頑張ろうという気持ちになるだけではなくて、実際に仕事としてきちんとできるようになります。



（図 4）

最近、地震学ではだんだん分業が進んできて、観測する人と解析する人は別々の人物ということはざらにあるのですが、データを使う人もきちんとデータに対してアクノレッジして使えるようにしよう、そうすることが地震学全体の発展につながるというのは、もっともだと思っています。

私が京都大学にいた頃はデータを生産する方で、24 時間 365 日、きちんとデータが来ているかを監視するのが仕事で、いわゆるエフォートの半分ぐらいを占めていました。その経験もあり、そこの部分がきちんと評価されるようになったらいいなという思いがあって、このような取り組みをしています。やってみると、それぞれの研究者がどのようにしてデータを生産し、それを公開しているのか、非常に先進的な事例もあります。

公開に当たって、例えば予算面や組織に対してどういう課題があるか、あるいは将来どうなったらいいかという将来展望についても話がありました。それから、データに DOI を付けるとはどういうことなのか、どのようにやったらいいのか。あるいは、DOI を付けるという方法ではなく、最近はデータジャーナルといって、データ専門で、解析結果ではなく、データそのものを出版する論文誌、雑誌があるのです。そういうものを利用してデータを公開する例もあって、そのようないろいろな方法でデータをオープンにし、それを利用しやすくし、データを作っている人たちもきちんと仕事として認められる、そういう方向にもっと進ん

でいけるといいなと、このセッションの開催が日本の地震学の中での出発点になってよかったと思っています。

自分が最近行った話をすると、例えば 1950 年代、1960 年代のとある地震学に関係するデータが全部紙で、それを一生懸命写真に撮って、今は自分の手元で解析していますが、こういうものを例えば画像のデータベースとして公開すれば他の皆さんにも使っていただけるのではないかと思いました。公開するときに例えばどんなメタデータを付ければいいのか、どういう整理をして公開すればいいか、ある図書館系の職員の方に相談したところ、「研究者がメタデータなど付けたら大変なことになる、やめておけ、ぜひそういうときは相談してくれ」と言われました。どのようにするとうまくこういうデータを公開できるのかということに関しても、今後考えていきたいと思っています。

それを相談した図書館の方は、京都大学古地震研究会のメンバーとして一緒にやっている方です。古地震研究会は、2012 年から、最初は教員と学生で始めました。今は図書館の職員の方、大学にはおられない一般の方にもご参加いただいて、非常に多様なメンバーでやっています。研究分野も地震学や気象学という理系の学問だけでなく、歴史学、地理学、人文情報学など、いろいろな分野の方が参加してくださっています。

基本的には集まって、そこで辞書を片手に、史料の画像やコピーしたものを見ながら解読の勉強、翻刻の勉強をします。この中で、例えばくずし字の辞書や、いろいろなデータを集約するときに情報技術を活用しています。読むための勉強の会と、時々、「情報科」という名前を付けた、情報技術を翻刻や解読の活動に役に立てようとする時間帯も持っています。そういう中から「みんなで翻刻」が出てきました。

古地震学研究会とは別枠で、京都大学にいた頃は図書館職員の方々と勉強会をやっていました。「目録を取るとき、くずし字が読めた方がいいから一緒に勉強会をやりませんか」と声を掛けていただいて、私が東京大学に移ってしまって今はできなくなってしまったのですが、週に 1 回ランチを食べながらやっていました。

3.みんなで翻刻

3-1.取り組みの始まり

「みんなで翻刻」の最初のアイデアは中西一郎さんが出されたように思うのですが、最初は別に、市民参加でやりましょうということではなく、集まらないとできないのではいろいろ具合が悪いということから始まりました。昔の地震に興味を持っている地震の研究者は全国にいるのですが、毎週京都に集まって勉強会をやるわけにはいかないということで、遠く離れたところにいる人と情報・データを共有しながら、翻刻を共有しながら、作業内容を共有しながらやる。テレビ会議で話はしながらできるのですが、それだけではなく、実際、手元のデータを共有しながら、「辞書のこのページのここ」と言いながらできるようにする、そういうことはできないかと、中西さんが橋本雄太さんに聞いて、彼は情報技術のプロなので「できますよ」と言って、そこから始まりました。

そうこうしているうちに、橋本さんや私たちの興味もあって、共同研究の中で閉じた形で共有するのでなく、市民あるいは非専門家が参画する、「みんなで翻刻」という取り組みにつながっていきました。

「みんなで翻刻」で検索していただくと、すぐ出てきます。そこからアクセスしていただいて、どんなことが行われているか見ていただくだけでも結構ですので、一回ログインしてみていただくといいかなと思います。「参加する」を押すと、何らか SNS のアカウントが必要にはなるのですが、ログインして参加していただけるということになっています。

3-2.バージョン 1 とバージョン 2

2017 年に始めた最初の「みんなで翻刻」のことをバージョン 1 と言っています。つい最近バージョンアップして、それをバージョン 2 と呼んでいますが、まずバージョン 1 を紹介します。バージョン 1 は「v1」

という名前で今でも使えるので、こちらものぞいていただくといいと思います。

バージョン 1 は、歴史災害史料の市民参加型の翻刻プロジェクトです。2017 年 1 月 10 日にウェブサイトを公開しました。京都大学古地震研究会が開発・運営しています。プロジェクトの目的は、災害史料の大規模なテキストデータベースを構築すること、翻刻作業を通じた市民の防災意識向上を目指すことです。

2017 年に始めたのですが、ついこの間公開から 1,000 日を超えました。1,015 日までで総入力文字数が 618 万文字、元々8,925 枚の画像、コマが入っているのですが、そのうちの 8,274 枚が翻刻完了し、文字が一通り入っている状態になっています。史料単位では 508 点入っているうちの 498 点、翻刻が完了しました。登録した人は、翻刻はしていなくても、取りあえずログインした人も含めると 5,320 人です。

この史料数の大部分を占めるのが、東京大学地震研究所が所蔵している古文書と分類されているものなのですが、そのうちのデジタル化されている、画像になっているものは全点、翻刻完了という状況になっています。

小野さんの講演で、スーパーボランティア、非常に熱心に参加してくださる方がおられると紹介していただいていました。まさにそのとおりで、この中で実際に文字を入力されている方は 10 分の 1 以下で、しかも、さらに小さな割合の方が大部分の文字を翻刻されているということが実際のデータとしてあります。

開発者の橋本さんが非常に工夫をされたところで、「みんなで翻刻」の特徴は、古文書解読の学習サービスとして設計されていることです。もちろん文字を解読することが目的ではあるのですが、その途中経過として,古文書解読を学習するということが入っています。具体的には、「くずし字学習支援アプリ KuLA」と連携しています。これは「みんなで翻刻」のほぼ 1 年前にリリースされていて、大阪大学の飯倉洋一先生が代表を務めておられたプロジェクトで、橋本さんが開発したアプリです。

スマホのアプリで、くずし字解読の学習ができる。この学習をして、ある程度読めるようになって腕試しをしたいという方は、ぜひ「みんなで翻刻」に参加してくださいという動線をつくっています。あるいは、KuLA はスマホでしかできないのですが、「みんなで翻刻」の中にも「まなぶ」というコンテンツを作っていて、それはそっくりそのまま、このアプリの内容とほぼ一緒なのですが、「みんなで翻刻」の中でも勉強できるというような設計になっています。

それから、参加者相互で学び合いを支援するような設計にしてあります。例えば、読めない場合には「□」で置いておいていい、自信がない場合は「？」を付けておいていいという翻刻のルールになっています。そうすると、後でもっと読める方が来て、ここはこう読むのだよと修正してくれるのです。修正された内容を、最初に「？」を入れた人が後で確認できるようになっています。あるいは、自信がないのでここはぜひ後で添削、確認してくださいと、修正を依頼するようなフラグを立てられるようにもなっています。

くずし字解読を学習しながら、自然に翻刻作業に参加してもらう環境をつくることが目的で、「やりがい搾取」にならないように、参加者に明確なメリットを得られる形で参加してもらうために、このような設計になっています。

参加者の Twitter を見ると、「『みんなで翻刻』は、ちょっと読める人があらかた翻刻すれば、後で達人が誤刻箇所や判読不能箇所を修正してくれる最高の環境なので、とても勉強になる」。それから、「『みんなで翻刻』、前の人が読んだのを見ながら、自分にも新たに読めるところがあって面白いね。国文学をやってた学生時代にこういうのがあったらなあ。時代は変わったね」と言っていただいています。

これはまさに先ほど言っていた目的であって、こういう Twitter でのリアクションを見ながら、私たちもシステムにさらに修正を加えていったり、参加者に呼び掛けるときに、こういうことを踏まえたりしています。

成果物の品質、翻刻がどれだけ正確なのかということも気になると思うのですが、これは橋本さんが博士論文の中で検証されているので、図 5 の表を見ていただくといいかなと思います。

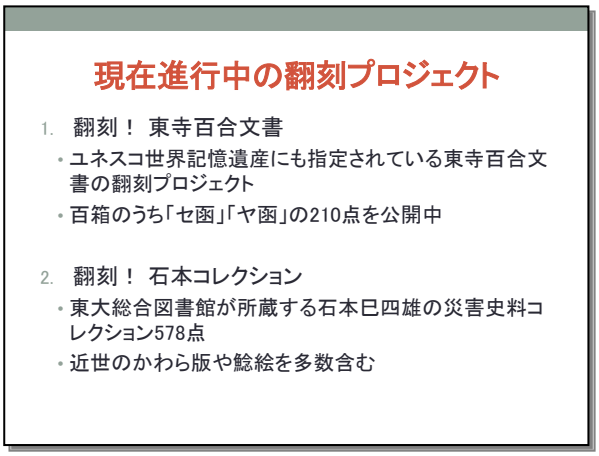
バージョン 1 での課題は、バージョン 1 は東京大学地震研究所のある特定のコレクションを翻刻するというように、狭い範囲の史料だけが対象だったのですが、もっと広い範囲の史料を用いたい、それから、初めて参加するような方にもぜひもっと参加してほしい、参加の敷居を下げたいということです。

このために、バージョン 2 へのバージョンアップのときに新機能を幾つか付け加えました。IIIF への対応、くずし字認識 AI の導入を行いました。IIIF というのは、デジタルアーカイブが公開する画像データの相互運用のための国際標準規格です。これを利用することで、それぞれの所蔵機関がデータベースとして公開しているものを、そのまま「みんなで翻刻」に取り込み、翻刻対象とすることができるようになります。

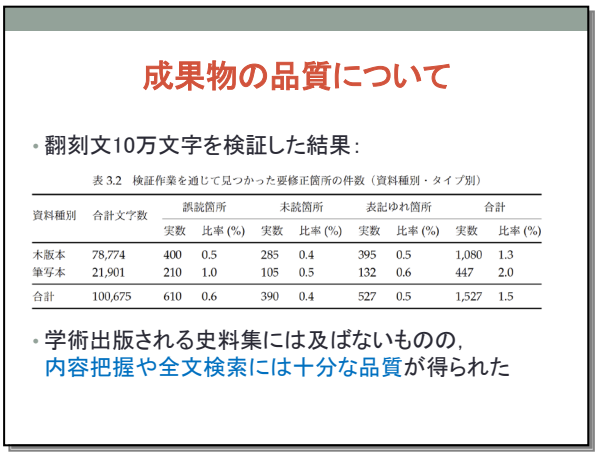
それを利用して、バージョン 2 と呼んでいる、現在の「みんなで翻刻」では、東寺百合文書、また、東京大学総合図書館が所蔵して公開している石本コレクションを翻刻対象として、今プロジェクトを進めています（図 6）。

AI くずし字認識については、「みんなで翻刻」にログインすると図 7 のような画面が出て、右側の史料画像を見ながら翻刻するのですが、読めない字があれば、囲って、AI くずし字認識機能のボタンを押すと、候補が出ます。その候補を見ながら、実際にどう読んだらいいかを最終的に決めるのは参加者なのですが、AI に相談しながら進めることができるということで、参加の敷居がより下がるのではないかと考えて、このような機能を入れています。最初から AI で全部読むというつもりではなく、参加を支援する機能、辞書を引いたりももちろんあるのですが、AI に聞いてみましょうという機能になっています。認識プログラムは、CODH が開発した「KogumaNet くずし字認識」と、凸版印刷の「くずし字認識システム」の 2 種類が選べるようになっています。

バージョン 2 は、7 月に公開して、今は 95 日目で、総入力文字数は 485,000 文字、翻刻が完了した史料は 788 点中 299 点、参加登録者数は 400 名と、順調に進んでいます。こちらの方もぜひ皆さんに参加していただきたいと思っています。



（図 6）



（図 5）

（図 7）

3-3.宣伝・工夫

この「みんなで翻刻」というプロジェクトは皆さんに参加していただいて、翻刻に参加してもらわないと意味がないので、なるべく派手にいろいろな方に呼び掛けるつもりで宣伝をしました。そしてドワンゴとの共同企画が出てきて、ニコニコ生放送で、E テレの教養番組風の仕立てにして、「みんなで翻刻」を紹介しつつ、翻刻とはどういう作業か、昔の地震を調べるとはどういうことなのかという解説する番組を作りました（図 8)。

そうこうしていると、ニコニコ超会議 2017 にも参加させていただけることになりました。ニコニコ超会議は、幕張メッセをほぼ全部貸し切ってイベントを行います。コスプレをされた方もたくさんおられて、例えば平安貴族のコスプレをされている方もおられました（図 9)。私たちがやっていることは机を出して地味に翻刻作業をパソコンの上でやるということなのですが、普段あまり翻刻や歴史地震に興味がなかった方、講演会をしても時間がなくて来られないような世代の方と少しでも交流する、あるいは翻刻に興味を持っていただく機会になったのは非常によかったと思っています。

「みんなで翻刻」の工夫は、学習ベースであるところ、また、ゲーム性も加えてあって、例えば翻刻文字数のランキングが出るようにしてあったり、「いいね！」に当たるものがお互いに送り合えるようになったりしています。あるいは、「200 文字翻刻しました」という毎回の翻刻の成果を Twitter などの SNS でボタン一つクリックするだけで自慢できる機能も付けてあります。

広報は、先ほど述べたドワンゴとのコラボを行いました。また、バージョン 1 を 1 月 10 日に始めたのは、関西では 1 月は災害に対する関心が高まる時期だからです。1995 年の兵庫県南部地震があったからなのですが、その時期に、ある程度関心が高そうなときにリリースしました。そのあたりの広報の戦略は、学術支援室、URA の方とも相談しながら進めました。

3-4.目論見と実際

歴史地震の研究のためにやりたいと私たちは思うわけですが、それは専門家側の思いです。最初のプレスリリースのときには、「過去の地震に関して新しい事実を発見したい、データを再検討したい、テキスト化を加速したい」ということを書いていました。くずし字を読める方を増やしたいということももちろん書いていたのですが、実際にふたを開けてみるとどうかというと、図 10 は参加者にアンケートを取った結果です。研究への貢献というよりは、「作業が楽しい」「自分が勉強になる」という方が結構な割合でおられました。だから、専門家側がこういう取り組みをするときに、こうありたいと思っても、それとは全然違うところで評価されるということが結構あるのだというのが私の感想です。

とはいえ、ご自身の楽しみだけでなく、例えば「自

（図 8）

（図 9）

分の地域の災害に気付くきっかけになった」というコメントも頂いていたので、そういう意味では、私たちが最初に思っていた目的や意図と同じことを感じてくださった方も結構いるのかなと思います。

こんな意見もありましたということを紹介すると、「『みんなで翻刻』によって若い院生のバイトや勉強の場を奪うことにならないか心配である」ということを言われました。確かにそれも一理はあるのですが、ただ、翻刻したい、読みたい史料は膨大にあって、もちろんたくさん読みたいとは思っていますけれども、こういう仕事やこういう機会を奪うほどの勢いはまだなく、全部やってしまうということはないので、そこはうまくすみ分けながらできるのではないかというのが私の意見です。

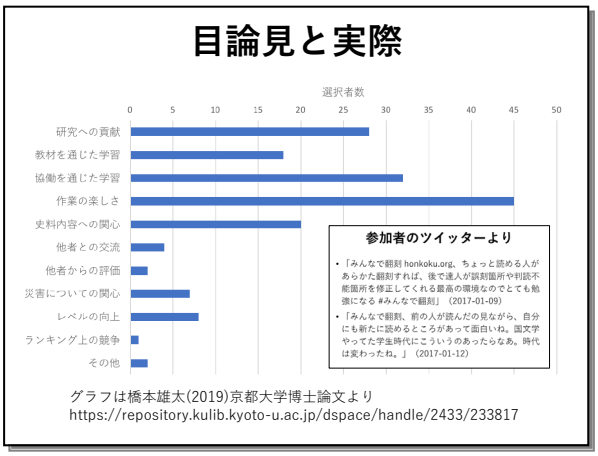
自慢したいと思っているのは、先ほどまでずっとご紹介したような内容です（図 11)。頑張らなければいけないと思っているのは、今テキストがたくさんできてきたのですけれども、それを基に歴史地震の研究をするということ。それから、歴史地震史料コーパスを作るなどして、テキストデータをさらに活用してもらえないか、あるいは、これは地震や歴史災害に関するテキストデータなので、その特徴を生かした活用ができないかと考えています。それから、今は翻刻されたテキストはログインしないと読めない状態のものが多いのですが、これを例えば読み物として、昔の災害について書かれた生の文書として読んでいただくために何か工夫ができるのではないかと思っています。

開発・運用体制については、開発を一人でやっている体制でいいのかということがあります。開発そのものではなく、参加者の方にアドバイスをするという運用の部分は古地震研究会の 10 人ぐらいのメンバーでずっとやってきたのですが、そういうサポートの体制も含めて、大学の研究として、一つの研究室でやるには限界があるのではないかと考えています。

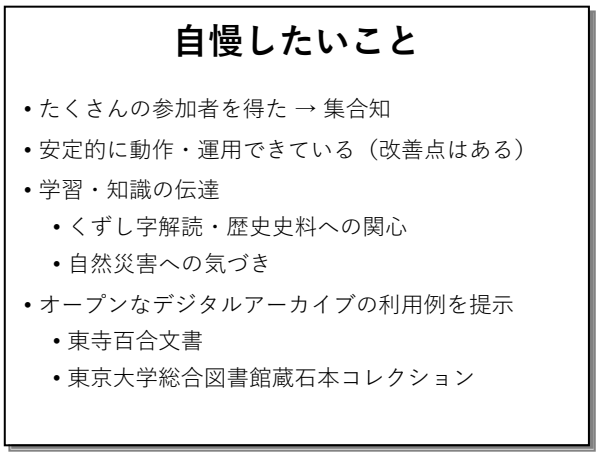
地震学の非専門家と一緒にやることに意味があるのだろうかと自分で胸に手を当てて考えたときに、地震学会が「行動計画 2012」というものを出していて、その中に「社会に対して、“等身大”の地震学の現状を伝えていくべき」という提言があるのです。「地震学をやっています」と言ったら、すぐ、「次の地震はいつ来ますか」「いつ気を付けたらいいですか」と聞かれます。いつ、どこでというのをきちんと予測するのは、地震学の専門家としても難しいのです。そういう地震学の今の実力をきちんと伝えていく必要があります。そのためには、研究活動に参加して、地震の研究とはどういうものかを知っていただくのが非常に有効だと思います。そういうプロセスの一つだと思って、私はやっています。

4.人文社会系分野におけるオープンサイエンスの在り方

皆さんは、オープンサイエンスは自然科学分野の方が進んでいると捉えられているかもしれません。自然科学は数値データを普段扱っているので、何となく進



（図 10）



（図 11）

んでいるような気がするかもしれませんが、そうではなく、非専門家の方と一緒にする取り組みは、どうやったらいいかと悩みながら、でも、やった方がいいのではないかと思っている人たちが取り組んでいるというところかと思います。人文社会学分野については、人文情報学という研究分野がどんどん進んできて、私もそこの成果に学ぶところが非常に大きいのですが、そういう中で今後も進んでいくのではないかと思っています。

今後やりたいことは、「みんなで翻刻」を進めていくのはもちろんなのですが、翻刻は結構大変なので、単に IIIF で公開されている史料画像を見ながら、その中に「地震」という言葉が書かれていたら Yes、書かれていなかったら次というようなものもできると面白いと思っています。これは今までは研究者が図書館や資料館に行って、史料をめくって見ていたことをオンラインに載せるという形です。

最後に、研究データだけでなく、プロセスにも参加してもらい、共有しようというのがオープンサイエンスの流れだと思います。研究者が「こういうふうにありたい」と思ってプロジェクトを始めても、参加される方はそういうつもりでない場合がよくあります。それは設計をうまくするべきだという話かもしれませんが、やってみないと分からない部分もあるので、そこは柔軟に、どこに一緒にできる部分があるのか、どこを一緒にできるとうれしいのかをやりながら考えるといいと思います。

（図 12）

専門家と非専門家の間の相互のコミュニケーションをうまくデザインして、データを共有し、共有したデータについて、「これはいいですね」「いい取り組みですね」「いいデータができましたね」とお互いに評価し合えるといいと思っています。

図 12 は今後行われるシンポジウムの紹介です。

●**フロア 1**　災害史の情報を、行政や司法などの方々とシェアしていないように思われたのですが、その理由は何なのでしょうか。非専門家とは、行政や司法など国家のシステムの中で働いている人も含めた人々のことをいうのか、それとも、地震の知識は関係学問の中だけで収めて、そういう人々にはシェアしないのか、その辺をお聞かせいただけますか。

●**加納**　私はぜひシェアをすべきだと思っていて、シェアするための一つのやり方が「みんなで翻刻」で、市民の方を巻き込んで実際に研究のためのデータを作る、研究のプロセスに参加していただくことだと思っています。

最初に、自分で書いた本を紹介しましたが、あれも、研究者が知っていることを非専門家の方になかなか知っていただけていないなと思って書いたものです。例えば、京都は 100 年か 200 年に 1 回、大揺れをするような地震が起きているのですが、皆さんとお話をしていると、「そんなことあったのですか」という反応をされます。皆さんが生きている間にはなかったかもしれないけれど、あと 3 世代、4 世代さかのぼれば過去に地震があったのですよということをなかなか知っていただけていないというのがあって、それを分かりやすく紹介する本ができないかと考えて、やってみました。

研究者が知っている情報、データなり、過去の地震の歴史なり、地震に関して分かっていることを、どううまく伝えるかはいろいろな方法があると思いますが、

その一つの手段としてこれをやっていると理解していただくといいかなと思います。誰もデータを公開したくないとは思っていないのですが、どういう形で公開するとうまく伝わるのかというのをいろいろ試みているところではないかと思います。

専門家、非専門家は、非常に狭い意味での、地震の専門家とそうでない方々という捉え方です。ですから、行政の方もある意味、非専門家と言ってもいいと思います。

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）
小木曽 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たいと思います。

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、私は国立大学にいるので、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、非難されるのではないか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。

（図 1）

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が、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体感で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

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。

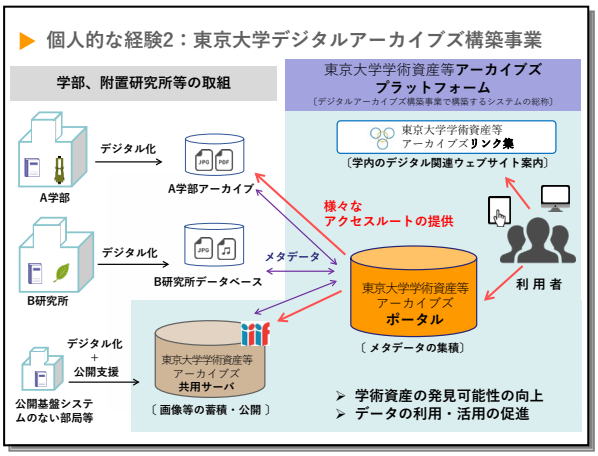
次に、総合図書館に特化した取り組みなのですが、2018 年 1 月から IIIF に対応した画像公開を開始しました。それから少し遅れて、2018 年 6 月に画像データの二次利用の条件を緩和しました。せっかく IIIF で公開するので、ライセンス的にも使い勝手の良いものにしようということで、クリエイティブ・コモンズ・ライセンスは付けていないのですが、CC BY 相当の条件で公開を始めました。今、この動きは総合図書館以外の各図書館・室にも少しずつ広がってきています。

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。

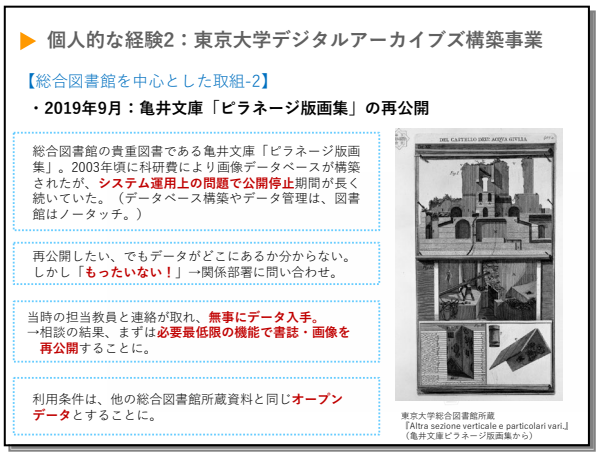
昨年、デジタル化すべき所蔵資料を検討していたときに、「亀井文庫のピラネージ」という意見が出たのを機に、実はもうデジタル化はされていてデータベースが止まっている状況ということが分かり、それはもったいないということで、関係部署、あるいは当時の担当者と思われる方にいろいろ問い合わせをしました。その結果、当時担当されていた先生と連絡が取れ、こちらの意図を話したところ、もちろん喜んで協力しますと言っていただいて、無事にハードディスクを預かりました。そこからデータを救済でき、公開していた当時のまま再公開できたわけではないのですが、画像と書誌をきちんと見られる必要最低限の形でこの 9 月に再公開することができました。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必要になるのではないかと思っています。



（図 2）



（図 3）

IIIF 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うのですが、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をしながら、一緒に何か画像公開やデジタル化をしていければいい、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、それに役立てる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した。

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曽先生と加納先生、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曽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がっち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曽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。

●**小木曽**　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、昔の本が好きな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の間違えてるわ」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**　今の小木曽さんの話を受ける形でお話しすると、この間、東京大学の学生向けに、「みんなで翻刻ソン」という、3 日間ひたすら「みんなで翻刻」に関わってもらう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をどう仕掛けるかということは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いうか、きちんと残さないと他の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ので、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、批判というか、マイナスのところも含めてきちんと評価してもらえる人がいるといいなと思いました。

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。私たちも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」などをやっていても何かできることがあるのではないかと思いました。

●**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ので、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もします。

場所の提供など、そういうこともちろんあると思いますが、こういうものが必要だよねとか、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だ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、小木曽先生以外の他の 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っています。

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曽先生、何かありましたらお願いします。

●**小木曽**　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ただいているのです。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う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る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こと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。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思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は思うのですが、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けれども、何かこういうデータを出してほしいとか、この辺を頑張ってほし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。

●**小木曽**　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を、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る、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、IIIF 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、京都大学の鈴鹿本、『今昔物語集』などは IIIF でリンクを取っています。われわれの方はまだ追い付かないのですが、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。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思っています。

●**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いうか、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このヒントがあるのではないかなとは思います。

●**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、ボランタリーな側面は別として、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。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、小木曽先生に対する質問として、「小木曽先生が提案された、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」、これは小木曽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。

●**小木曽**　非常にもっともな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う的な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と思ったりしました。

その一方で、私は言語処理もやっていましたので、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。小木曽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。小木曽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が、「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。

次は小木曽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」。

●**小木曽**　後半については、話はないですね（笑）。ないですけれども、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はあります。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います。

●**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、あるいはベンサムの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

（図 4）

橋本さんの研究としても進めていきましょう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す。

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るので、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ていました。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

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は思うのですが、さまざまな支援の形があって、先ほど言っていただいたように、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うのですが、そこでも全然違います。最近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。

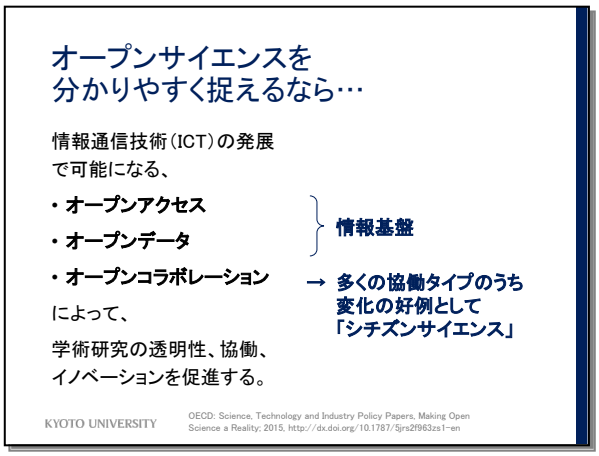
ただ、これまで図書館が所蔵資料の管理とデジタル化を行ってきたり、機関リポジトリを運営したりしている大学はとても多いので、そこのノウハウがお役に立てるのではないかと思っている人は多いと思います。ですから、その辺の、図書館だから絶対ここが役に立ちます、こうしますということではないのですが、なるべく組織全体のミッションや方向性を見ながら、自分たちが何ができるのかを考えていくしかないのではないかと思っています。その中でなるべくであれば前向きに何かに取り組んでいきたいなと。「できません」と言うのはつまらないので、「これだったらできます」「一緒にやってみたいのでぜひやらせてください」というような姿勢でいることがいいのではないかと考えています。

●**小木曽**　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するのですが、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思っています。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にやってこられたか。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います。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たらいいなと私は思っています。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で、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います。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。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います。

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っているのですが。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願いします、というのが私のまとめです。

●**小木曽**　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、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。昨年はインフラの話、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。どちらかというと他の回は、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いかげんにやめて、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。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していただけたのではないかと思っております。

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たいと思います。最後に、登壇者の皆さまに拍手をお願いいたします。