

第2回 SPARC Japan セミナー2015

「科学的研究プロセスと研究環境の新たなパラダイムに向けて
- e-サイエンス, 研究データ共有, そして研究データ基盤 -」

ディスカッション 「研究者の期待と研究データ共有への関わり」

蔵川 圭	(国立情報学研究所)
加藤 斉史	(科学技術振興機構)
田中 良昌	(国立極地研究所)
大山 敬三	(国立情報学研究所)
星子 奈美	(九州大学附属図書館)

●蔵川 まず、4名の方から伺った具体例や周辺の状況は、だいぶコンテキストが違うということが分かりましたが、何か質問はありますか。

●フロア1 大変興味深くお話を伺いました。私は社会科学系をベースにした歴史的思想史が専門なので、人文社会科学側の事情は詳しいのですが、自然系でこれほどオープンアクセス、オープンデータを大事にしている分野があるということを知りました。

私は勤め先で図書館長をしていますが、オープンアクセスを進めることについて、学内でなかなか説得できません。特に自然系の人々は、自分の実験データを中途半端なところで出したら横取りされるという警戒心を持っています。そのような中で、むしろ共有することではじめて考えたいことが分かるという大きなフィールドがあることを知って、非常に励みになりました。

オープンアクセスには二通りの意味で混乱があります。昨年以來、オープンアクセスジャーナルは大事だから、科学研究費の出版助成で日本でもオープンアクセスジャーナルを育てようという話が出ています。ところが、研究者のごく一部の事情に詳しい人は、ジャーナルに投稿して載らなかったものは順繰りにオープンアクセスに載るらしいということを知っています。その中で、内閣府がオープンアクセスは大事だという

報告を出して、今、みんなでやろうとしています。現場では、オープンアクセスは良いものか悪いものか、研究者にこれをどう説明するかという難しい問題があります。これを皆さんはどう説明しているか、これが最初の質問です。

もう一つ、今はクレジットが大事にされていて、知識は誰かのものであるということが大前提で研究が行われています。しかし、科学というのは、神の知識を人間が頑張って分かってほしいというだけの話で、自分の知識などと言うのは傲慢なことです。そんなことより、ここまで分かったけれど、ここからは分からないから、みんなで考えようというのが健全な科学ではないでしょうか。オープンアクセスからオープンナレッジへ、分からないことをみんなで文殊の知恵で頑張って研究しようという機運を醸成することが求められているのではないかと思います。

ただ、残念ながら研究費も研究ポストも、自分の成果が認められてありつけるという状態ですから、非常に難しいです。今日の講演を聴いて、オープンアクセスが進んでいたのは、オープンだからできるという分野であって、そうでない分野の人をどうやって説得したらいいのでしょうか。星子さん、九州大学は試みを進めているので、どのように九州大学で頑張っているのかを教えてくださいと大変励みになります。

●星子 オープンアクセスジャーナルが査読誌に掲載されなかったときの代替手段として用いられているというのは、オープンアクセスの理念自体というより、現存の査読システムに内在する問題なのではないかという印象を受けました。オープンアクセス自体は、オープンアクセスジャーナルを用いずとも、既に査読された論文の著者最終稿を機関リポジトリに登録することによっても実現される場合もあります。

●蔵川 このような視野の広い話はぜひ後半に持っていきたいので、ディスカッションの前に知っておきたい、発表でアンクリアだった部分について聞いていただけるとありがたいです。

●フロア2 田中先生に伺います。「データベースをつくる仕事は損した気分になる」というところで、私は図書館員なので少し気持ちが分かるところもありました。実際にデータベースをつくる仕事は誰に振られるのでしょうか。基本は研究している人がデータベースをつくると思うのですが、教授になった方々がつくっているとはあまり思えないので、データベースはこの人をお願いしたいと仕事の分担が分かれてくると思います。どういう人に向かっていくのか、その辺の内部事情をお聞かせいただければと思います。

●田中 損した気分になるとは言いましたが、それは面白く言っただけで、極地研としては、まずデータを使ってもらうことが非常に重要であることは皆さん認識しています。極地研はデータをとにかくたくさん取るのですが、使われずに眠ってしまうデータもこれまでたくさんありました。国の予算を使ってデータを取っているのに、使ってもらいたいという意識も強くありました。しかし、研究者が観測して、しかもデータベースもつくるといことになると、時間もなくてなかなか難しかったという面がありました。

おっしゃるとおり、教授は忙しくて時間がないということもありますし、新しいスキルが必要なので、若

手のポストドクレベルの方がやることが多いと思います。大学では、学生がやっていることも多いのではないかと思います。これまではデータベースの専門家がいなかったのも、データベースをつくるといっても、個人個人で好きなデータベースを勝手につくるためデータベースが乱立してしまい、また、データフォーマットもばらばらで、使い勝手の悪いものになっていました。IUGONETのプロジェクトが立ち上がって仕事をしていく上で、データベースをつくるには人手もお金もデータベースの専門知識も要するというのをあらためて感じました。お答えになっているかどうか分かりませんが、基本的には、これまではやはり若い人、新しく入ってきた人をお願いする形が多かったのではないかと思います。

●フロア2 ありがとうございます。先ほど星子さんから共同分担目録という単語が出てきたのですが、図書館でも、データベースをきちんとつくりようとする上で同じような問題を抱えているということだったので、質問させていただいた次第です。

●フロア3 オープンで使えるデータをいろいろ見ていたら、やはり使いにくいところが多くて、データにはアクセスできるのに、それをどう使えばいいか全然分からない状況が今ではないかと思っています。アメリカやヨーロッパの場合は、オープンソースになっていたり、コミュニティーが一緒になったりしていて、データのユースケースを共有することも考えていますが、日本はまだそこがあまり見えていません。なぜそうなのか、お考えを知りたいです。IUGONETは6年ほどの長い経験を持っているので、どこでデータがよく使われたか、それで少しくまってきたところとうまくなかったところは何があるか、教えていただければと思います。

●田中 IUGONETはデータをかなりオープンにしてきたと思うのですが、同じ分野のコミュニティーに対

してオープンという部分が多いです。もちろんデータは完全にオープンになっているのですが、一般の人がそれにアクセスして、何か新しいことができるかという、それは結構難しいと思うのです。前半に北本先生から、本来、データベースは一般向けと研究者向けに分けない方がいいという話がありました。理想はそうだと思うのですが非常に難しいです。コミュニティにとってはかなり専門的なデータというか、サイエンスに使えるデータが必要なので、それを一般の人が使うとなると、もう一段優しい、使い勝手のいいものにしていかないと使っていただけないと思います。

その意味で、IUGONET が公開しているデータはオープンなデータなのですが、一般人向けにはオープンになっているとは言い難いと思います。そのギャップを埋めるのはこれからだと認識しています。われわれの一つの取り組みとして、主にコミュニティに対してですが、データベースや解析ソフトの使い方の講習会を年に 2 回行っています。20~30 人集まって、研究集会のような形式で行っています。そこで実際にデータをプロットしてもらって、このデータはこのように使うのだなということを体験してもらいました。これは研究者にとってはとても好評で、われわれのコミュニティの中ではかなりデータを使ってもらえるようになったと思います。

これらのデータを一般の方に使ってもらうのは今のところは難しいので、もう少し一般向けに、ウェブサイトをもっと親切にする、マニュアルをもっと整備する、YouTube にもっと分かりやすい動画をアップするなど、大変かもしれませんが、そのような取り組みが必要かと思えます。

●**蔵川** 他に発表についてクリアにしたいところがあれば受け付けますが、そうでなければ、私から聞きたいことを質問します。

第 3 部のテーマは、研究データ基盤をどう担うかです。日本では、特に図書館という今まで研究データに関わっていなかったところがこれから関わろうとして

います。そのような中で、日本の研究データ基盤は歴史的にどのようにつくられてきて、これからどうなっていくかということが知りたいわけです。

パネリストの方のお話で分かったことは、データのユーザーのコンテキストが全く異なるということです。「研究データ」と一言で言っても、もともと研究するためのデータでなかったり、個人的な研究のためのデータであったりと、さまざまなパターンがあります。今、ユーザーにどのような要求があつて、これからどのような要求に変化していくかということを、それぞれのドメインの担当の方に質問したいと思います。

●**加藤** 研究データの DOI 登録に当たっては、われわれではコンテキストが分からないところもあって、モデルとなる先輩の DataCite の状況も見ながら始めて、最低限、DOI を付けたものが識別できるようにすることを目指しました。メタデータも、それほど複雑でもなく、シンプルで一般化したものであり、逆に言うと、その分野では少しもの足りないかもしれないものになってきているのが現状です。

今は実験プロジェクトを通して、だいぶコンテキストも明確になってきました。現在は、研究データに DOI が振られている数もまだ少ないので、われわれもだいぶ見えてきたとはいっても、ユースケースはまだ集まりきっていません。その中で、一番近い目標としては、分野の中の細かい活用については、ドメインのデータベースに委ねていきたいと思っていますが、研究データと論文のように種別を超えたところをつなぐ仕組みとしての役割を果たしていきたいと思っています。

●**蔵川** JST の JaLC は日本の中ではかなりビッグインフラストラクチャーだと思うのです。今、われわれがどういう基盤を提供していくかというよりは、むしろユーザー側にどういうニーズがあるのかを伺いたいです。例えば、文献になぜ DOI を付けたいというニーズが湧いてきたのか、なぜ次は研究データだと言いだしたのかということです。プロビジョニングとして、

このようなシステムがこれから必要であろうからといって出しているというよりは、データにも DOI を付けたいというユーザーの声の変化があったからこそ、今ここでプロジェクトが展開されていると理解しているのですが、そのような声はあったのですか。

●加藤 ビッグインフラストラクチャーというお話がありました。研究データと論文の両方を扱い、さらに CrossRef と DataCite を仲介して、日本の中から出ていくものを逃がさないで集めるという役割を Japan Link Center として担おうという考えがあります。

また、論文から研究データに至る中で変わってきたこととしては、研究データでも論文でも、どちらでも変わらないといえば変わらないと思うのですが、識別してそれを評価に使いたい、つまり、どこで引用されているかをきちんと把握したいというニーズがよく届いていました。標準的な識別子 DOI を付けることで、引用の定量的な把握をしやすくしたいというのが一つのニーズだと思っています。

●蔵川 ありがとうございます。次は田中先生から、研究者の研究データに対するニーズの変化、昔はどのようなものがあって、今はどうなっているかを教えてください。例えば、もともとは研究データは個人が収集していて、「IUGONET のようなものは必要なかったけれど、今の状況になって必要になった」というお話をされたと思います。基本的には、研究の現場や研究者のニーズに合わせて IUGONET のような研究基盤をつくっていると思うのですが、研究データの利用者のニーズを考えたときに、どのような変化がありましたか。

●田中 昔はネットワークもインターネットも発達していなかったんで、自分のところのデータを解析すればそれで論文になっていたのですが、今は一つのある現象があったら、それをいろいろなデータを使って見なければ全体が見えません。能勢さんの講演でもあり

ましたが、ライトで1カ所を照らすだけでは全体像は全く分からなくて、いろいろな人がライトを照らさなければ見えないという世界です。今や外国のデータ、例えば、NASA の衛星データなどがどんどんオープンになっているので、とにかくたくさんデータを一度に解析した方がより成果が出るという傾向になっています。

その結果、とにかくたくさんデータを早く集められる人がより多く成果を出せるので、一つのソフトでデータがたくさん集められて、一度にプロットできることへのニーズが非常に大きい。われわれの解析ソフトはそのニーズに合っているのだと思います。

また、次のような解析ソフトの予想外の使い方もありました。われわれは発展途上国などに行き現地の人々の協力を得てデータを取ることが多いのですが、その現地の人たちも、最近はこのデータが何か知りたいというニーズを持っていて、「言われるがままにデータを取っているけれど、これは何のですか」と聞かれることが多く、その場合に、彼らも解析ソフトの使い方を覚えて、その場でデータをプロットしてデータを理解すると、モチベーションが上がって、より協力が深くなります。このように、キャパシティービルディングとしてのメリットも非常に大きく、とにかく簡単にデータがプロットできるというのは重要なことだと思います。

●蔵川 活動のビジビリティやインセンティブというのはそのようなところでもあって、それが研究基盤の仕組みに対してもフィードバックが掛かって、そのような仕組みが必要だということになったのですか。

●田中 そうですね。

●蔵川 分かりました。次の大山先生は全然違うコンテキストだと認識しています。情報学のデータ自体は、1970年代から、テキストコーパスのような、テキストのデータベースはあったと思っています。最近では

Yahoo!のデータセットなど、それとは全く別の活動から出てきたようなデータも扱っていて、さまざまなニーズを一手に受けているような感じがしています。それは情報学を中心に考えたときに、研究データのニーズがかなり変化しているということで、次の時代には予測もつかないニーズが出てくるような気もしています。そのような観点から、情報学の研究データに対する消費者のニーズの変化はどういう感じになっているのでしょうか。

●**大山** しばらく前までは技術がそれほど進んでいなかったこともあって、自然言語処理、画像、映像処理といった分野で、こつこつと積み上げてきた技術が研究の中心だったと思います。そのような研究は、きれいなきちんと整備されたデータがないとできないところがあって、逆にそのようなデータは、大きければ大きい方がもちろんいいのだけれども、小規模でもある程度研究はできました。データをつくるのにコストが高くつくので、小規模でもうまく研究にフィットする形のデータを丁寧につくっていききました。しかも技術的にまだ十分ではなく、どんな環境でも動くということは考えられないので、限られた条件下で動くことを目標にやっていたのだと思うのです。そうすると、そのような条件を反映したデータが求められて、研究者あるいはそれに近い人たちが、特定の研究目的でデータを準備して研究するという形が多かったのではないかと思います。

最近ではディープラーニングなど、人工知能とひとくくりにするると語弊があるかもしれませんが、統計的な手法が発達してきたことによって、とにかくデータさえあれば何か結果が出るという傾向がかなり大きくなってきました。特定の目的のためにこつこつつくったものはそれにしか使えませんが、専用のデータではない生のデータは、とにかく大規模であれば、その切り口を変えるといろいろな使い方ができます。逆に、新しい使い方を見つけるとそれが研究になるという側面もあって、そのようなデータに対しての注目が大きく

なってきたというのが一つです。

もう一つ、従来は割と一つの狭い分野の技術の研究で閉じていました。例えば自然言語処理なら、自然言語のコーパスがあれば研究ができる、画像処理、映像処理なら、それらのデータベースがあれば研究できるということだったのですが、最近の情報学分野では、一つのメディアに閉じた研究から、だんだんメディアをまたがる研究や、人や社会との関わりに関する研究テーマというように、大きくなってきたと思うのです。そのようなものは人工的なデータでは捉えきれませんし、人間や社会が関わる環境の中でつくられたデータを使わないとそのような研究はできないということがあって、実システムから生成されたデータが求められてくるようになったのではないかと思います。

現状では、われわれが提供しているデータの中にログデータはほとんどありません。人間の行動が反映されたデータは取扱が大変難しいからです。そのようなデータを使いたい研究者は多いのですが、持っている方が出したくても出せないという状況で、そこがジレンマです。他にも取扱を難しくしている要素があるので、三すくみ、四すくみになっています。これはオープンデータでも、クローズドデータでも、データを大きな固まりとして扱っているだけでは解決できない問題なので、そこをどう解決していくかが今後の課題になると思っています。

●**蔵川** 超高層大気の場合は、ある研究をしたい、知識を創成したい、そのためにこのようなデータをつくるというストーリーが順序立てられているのですが、情報学の場合はそのようなわけではないですね。例えばブログのような、研究のためにつくられているわけではないデータがあって、それがなぜかいつの間にか研究になっているという、逆の展開のニーズもあると思うのですが、その辺は何かありますか。

●**大山** 情報学が多くの自然科学と違うところは、人間や社会も研究のターゲットに含まれることです。情

報はそのようなところから生成されてくるからです。ですから、情報を抜こうとすると、必然的にそのような発生源までさかのぼる必要が出てくると思います。例えば、人間はこうあるべきだというモデルをつくることは考えられるのですが、観察によって何かそこから見つけるというストーリーは極めて自然にあれども、モデルを人間に当てはめるという研究のアプローチはあまりありません。

ですから、どちらかというと、あるがままのデータを使って人間の行動を観察し、それをどう有効に使うかという2段階になると思うのです。最初の段階は、実際に使えるデータが設計して取れるようなものではないので、自分はこのようなことをやりたいのだけれども、データはこのようなものしかないから、その中でこのような切り口だったら近いことができるというように、自分がやりたいことと使えるデータの整合性が高いデータを見つけるとうまく研究ができるということです。そこが逆にセンスといいますか、いいテーマの見つけ方という話になってくるのではないかと思います。

●**蔵川** ありがとうございます。次は星子さんに伺います。まだ活動されていないということなのですが、少なくとも自分の周りで、研究データといろいろな人が言っているわけですよね。今までは情報学でも地磁気でもかなり分野として確立されていて、古い時代からコミュニティとしてデータをつくっていて、このようなユーザーの変化があったと分かったのですが、図書館から見て、学内の研究者をユーザーと見立てたときに、研究データに対してどういう変化があって、今はどういう状況になっていますか。

●**星子** 図書館に対し研究者からデータについてご相談を頂く事例はかなり少ないのですが、10年ほど前に「長期間にわたるカイコの飼育記録のデータを持っているので、図書館でデータベース化してもらえないだろうか」という相談を受けたことがあります。この

データは図書館が管理するデータベースを用いて公開しましたが、その後システムの更新に伴い維持が難しくなりました。現在そのデータは研究プロジェクトの方で管理されています。

●**蔵川** 図書館は本を扱うところなので、紙をベースにした資料というイメージがあるのですが、カイコのデータベースは、ハードディスクの上にあるものなのでしょうか。明治時代の人がコンピューターを持っているとは思えないのですが、そのようなものも研究データとして保存してあるという話もよく聞きます。そのようなものは図書館の中では何かニーズの変化があったりするのですか。

●**星子** 先程申し上げたカイコの飼育記録の場合は紙ベースの記録があり、それをデータ化したものを公開していました。現在では、記録形式自体が電子化されている可能性もあるかと思います。そのような意味で、データの残し方が変わってきているというのはあるかもしれません。

●**蔵川** それは10年前のお話ですが、紙ベースの話で、恐らく画像としての電子化ですよね。特に今、オープンデータ、オープンサイエンスといわれて、図書館の役割がかなり期待されていると思うのですが、そういう声はありますか。

●**星子** 現在の時点では、研究者自身というより、学内の事務部門から今後の見通しに関し質問が出始めているといった状況です。

●**蔵川** これまで各方面で研究基盤を担われている方々から、ユーザーの挙動やニーズについてお話をさせていただいたのですが、実際にユーザーがいる状況の中で、研究データは、電子化して公開すればそれでいいわけではないと思っています。例えばライセンスの問題、受け入れるときに人をきちんとチェックしなけ

ればいけないという話など、実務だからこそ、これは気を付けないといけない、その穴を埋めてからでないと前には進めないというものはかなりいろいろあると思うのです。それは全部ですといえば全部なのでしょうけれども、それぞれが関わっている実務において、表現できるものがあればしていただけませんか。そうすると、共通項と差分が見えるのではないかと考えて質問しました。

まず加藤さんからお願いします。加藤さんはずっと実務をされていると思います。JSTさんはすごく期待されているし、研究データにもDOIを入れると言っていますが、例えば無料化しないと駄目だ、有料にしないと駄目だなど、ライセンスの問題はあまりなさそうですが、データがコレクションされると困るとか、クレームがいろいろありますよね。そのような観点で、これは実務を進めるに当たって気を付けておかないといけないというエピソードはありますか。

●加藤 まずユースケースが少ないので、どういうメタデータを入れてほしいかなど、目的がきちんと見えていないままに決めているところがあり、入れる側にしても、こちらが欲しいと思っているものがきちんと伝わりきっていないで、登録する人たちにもとまどいを生んでいるのではないかとこのところが一つです。

もう一つは、登録の流れなども、大学の図書館からNIIを経由してIRDBに参加している流れもあれば、直接、大学の図書館が入るような流れもできてきます。そのような流れは確立しつつあるので、そこについてはさらにスムーズにいけるようにしていきたいと思っています。2点ほどが実務的なところで気になっているところだと思います。

●蔵川 いろいろな人の話を聞いて、スムーズに動けるようにするということですね。分かりました。田中先生は。

●田中 前の部でかなり言われていたのですが、ライ

センスやクレジット、謝辞の部分はIUGONETとしても常にかなり気にしている部分です。IUGONETはデータプロバイダーのグループの集合体なのですが、データの使用ポリシーはIUGONETグループ全体のポリシーがあるわけではなく、各データのポリシーを各PIが決めています。つまり、われわれはIUGONETで何かクレジットやデータポリシーを決めるということはやりませんでした。

データは全てオープンにしているのですが、使ったときに、「このデータを使うときにはこの人にコンタクトしてください」と表示されるようなシステムはつくっています。しかし、自由にプロットもできるし、論文にも出してしまう状況です。それは研究者の良心に任せるというスタンスでやっています。ただ、今までのところ例はないのですが、PIの方からは、データを勝手に使われたらどうするのかと常に言われていて、それはかなりプレッシャーになっています。また、その人のデータを使ってどれぐらい論文が出たのかというのは、今はインターネットで検索するしかない状態です。実際、勝手に出版されていても気付かない状況なのです。自分たちの努力で調べないと分からないので、その辺はすごくプレッシャーになって、何とかできないものかとは常に思っています。

●蔵川 気になるが良心に任せるという感じですね。大山先生の情報学だと、紛争になるというか、裁判でも起こしそうな感じがあるのですが、何か実務的なところでありますか。

●大山 やはりその恐れがあるので、データを提供してもらったときの条件はかなり慎重に詰めています。逆に、提供してくれるという民間企業の担当者が非常にナイーブで、気楽に「出せるよ」と言ってくるのですが、われわれの経験で、それはあまりにも危ないということで、むしろこちらが抑制するようなことも現実問題としてはあります。

さすがに最近はないと思いますが、以前では、ユー

ザーから投稿されたデータをテキストとして頂く場合に、そもそもサービスの利用規定に著作権の規定がないという事例も実はあって、そのようなことはデータを提供する前に、「利用規定を変えた方がいいですよ」というような話をすることがあります。

ですから、単に提供者とわれわれの間の関係だけではなくて、利用者との関係がどうなっているかという利用規定までチェックして、大丈夫かどうかを確認することもやっています。そのように保守的にアプローチをしていることもあってか、紛争になったことは今まではありません。

●**蔵川** ありがとうございます。星子さんは、図書館で研究データと言いながらも紙の本がユーザーからリクエストされていて、ライセンスの問題で実はもう一回調べたことがあるということは何かありますか。

●**星子** ディスカバリーサービスの運用に伴い、図書館の所蔵資料だけでなく、貴重資料の画像も統合的に検索可能になり、画像の二次利用申請が以前より増加しています。こうした申請に迅速に対応するため、二次利用に関する規程やデータの保管場所、データ形式といった情報をより系統的に管理・共有することの重要性が高まっていると思います。

●**蔵川** 研究データの共有が既にそれぞれのドメインで行われており、今の状況をより広い分野の人たちに開放するという形で研究データ共有を進めていこうとする流れになっていると思います。そのようなドメイン利用者の研究分野の拡大を踏まえて、これからわれわれがどういうことをやっていかなければいけないのかということ、会場も含めて残りの時間でディスカッションしたいと思います。

そのような観点から何か質問はありますでしょうか。特に会場からあるとうれしいです。前向きな発想で、私だったらこうする、うちの分野だったらこうする、隣の分野の研究データのこういうことが情報交換でき

たらより論文が書けるなど、いろいろな状況があると思います。そのためにはさまざまな問題をクリアしなければならないということが分かりました。系統的にどういう基盤をつくることによってそれが楽になるのか、それは誰が担うのか、そのコストは誰が負うのかということも、ヘテロジニアスな観点から事象が複雑に絡み合っ、全体として一つのデータを扱うリサーチコミュニティが前に進むという状況ではあると思うのです。そのような観点の中で、今後われわれがどうしていけばいいのか、コメントでも自分の思いでもいいですし、来年3月にRDAが東京で行われるので、活動を始めるところのインセンティブになるような話がもしありましたら、お願いします。

●**フロア1** それぞれで予算を立てられる人々がどう振る舞うかというのが最終的には一番大きいことだと思うのです。似たようなフィールドで、あちらでもこちらでも類似の実験をしていて、誰が最初に結論にたどり着くかという状況では、うかつにはオープンにできません。最終的にはそこになるのですが、昨年来、そこで不正が起きたから、データをオープンにしなさいといわれています。オープンアクセス、オープンデータといっても、後ろ向きの、不正をなくすためのオープンアクセス、オープンデータという話の一部になってしまっています。

しかし、今日お話しになったのは、後ろ向きではなく前向きで、それぞれが持ち合わせている情報から何が分かるかやってみようというオープンアクセス、オープンデータでした。大掛かりな実験をした、さまざまなエネルギーを注ぎ込んだ、しかも若手の力を多く借りながら行った調査であつてもなおデータが足りないのだから、みんなで持っているデータをこれだけ集めようということは、手書きで原稿を書いて、タイプで印刷して、それを配ってということをやっている時代にはできませんでした。それが今はネットワークを通じて大いに進められるようになって、科学のコンテキストと手法が変わってきています。そして今、変わ

ったことをどう生かして、振る舞って、かつ、それだけ大きな予算を頂くのだから社会に対しての説明が付けられるのかというところにたどり着いているのです。

人文社会科学の場合、伝統的には個人事業で、自分が好きなフィールドを一人でこつこつ十何年かけて調査したら何かがあったという研究スタイルだったのですが、最近では大幅に変わっています。十何年ほど前までは、まだ日本の大学でも貴重コレクションはお宝で、劣化しないように大事に管理するのが図書館の機能でした。それにのめり込んだマニアがひたすら読んで上で理解していました。しかし最近はそのようになって、博物館、美術館、図書館も持っているお宝を画像にして、オープンにした上でみんなに提供するというところまで来ています。問題はそれがまだランダムであることです。国文学研究資料館のプロジェクトは、それを国家的予算で支えて、貴重な資料をみんなが見られるようにした上で、研究者だけではなく、国文学をかつて学んだ在野の方々にも大いに興味を持ってもらって、解説して新しい文化をつくろうというものだと思うのです。そうすると、人文社会科学系でも実はオープンサイエンス、オープンデータは意味があるのです。ものによっては興味を持つ人々が世間にも非常に多くいらっしゃるので、眠っているデータを持ち寄ることによって新しい世界が開けそうです。

また、公式発表のデータだけでは分からないことがあります。国際的なデータでは、公式発表をしている国のデータはよく分からないけれども、社会科学系、特に経済学が調査畑の人が、さまざまな統計データ、例えば企業の活動、貿易データ、為替の動きなどをつじつま合わせして、どうもこのデータはうそで、このようにバイアスを掛けて統計的に処理すると、本当はこんなことであるに違いないという解析を行っています。これはそれぞれのところでやって意味が分かるのではなくて、同じような、しかし、手元に入るさまざまな断片的データを修正したものを持ち寄ることによってはじめて分かってくるのです。これは講演で出てきた、気象学や遺跡と似通った話です。そうすると、

オープンデータにした上で分かってくることは非常に多そうだということで、何ができるかはよく分かっていないのですが、少しは希望を持って励みましょうということですか。

●蔵川 ありがとうございます。大変前向きな形で、アトラクティブにデータ共有がなされるということがよく分かりました。

残念ながら時間が来ましたので、ここでパネルディスカッションは終わりにしたいと思います。登壇の皆さま、質問していただいた皆さま、どうもありがとうございました。