

第1回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

パネルディスカッション

- 八塚 茂 (バイオサイエンスデータベースセンター)
- 林 賢紀 (国際農林水産業研究センター)
- 池内 有為 (文教大学/研究データ利活用協議会 研究データライセンス小委員会)
- 海老沢 研 (宇宙航空研究開発機構)
- 上島 邦彦 (株式会社日本データ取引所)
- 三橋 信孝 (バイオサイエンスデータベースセンター)
- 桂樹 哲雄 (農業・食品産業技術総合研究機構 (農研機構))
- 篠田 陽子 (物質・材料研究機構)
- 仲 修平 (東京大学社会科学研究所)



●八塚 このディスカッションでは、池内さんをご講演で示された三つの論点を中心に議論していきます。それから、皆さまから頂いているご質問に回答したいと思います。

まず、皆さまのご講演の振り返りを兼ねて、二つの論点についてお一人ずつコメントを頂きます。一つ目は、フルオープンではなく制限公開をする場合の根拠と、制限公開に区分を設ける場合、どのような基準で区別しているかということです。二つ目は、制限公開をしている場合はアクセス申請を受け付けることとなりますが、その妥当性の判断をどのようにされているかということです。

●仲 一つ目の制限するときの根拠は、利用資格に合致しているか、過去の利用状況はどうか、不正な使い方をしていないか、寄託していただいている大切なデータを扱っていますので、それに対して誠実に扱ってもらっているかという点で検討しています。

二つ目の妥当性の判断に関しては、個別に一つ一つ検討しているのが実際のところです。基本的には多くのデータを使って新たな学術的価値をつくっていただきたいと思っているので、どんどん公開したいのですが、制限をかけるときには利用条件や使い方の点で判断しています。その判断が割れるときには、上長に確認したり、あるいは内部の委員会にかけたりして、制

限をかけるときには慎重に対応しています。

●**篠田** 一つ目の制限公開と公開区分の基準の根拠については、材料分野はどちらかというとオープンしにくい状況があるのですが、その中でも論文発表や外部発表をして一般に公表したデータは公開リポジトリに入れやすく、また、外部発表をするときには機構の中での手続きもありますので、そういうところを公開の一つの根拠としています。

利用者のアクセス申請の妥当性については、例えばMatNaviですと、ユーザー登録をしていただいた方は承認されるという形になっています。申請は必要ですが、非常に制限が緩やかな制限公開という形です。あるいは、もう一つの例としてご紹介した共有データリポジトリなどは組織契約を基にしますので、相手方に内部で共有利用の責任を持っていただけるかなどの審査フローを経た上で、妥当性を確認することになる予定です。

●**桂樹** 一つ目の根拠については、個人に属するデータ、例えば腸内細菌の細菌叢を調査する場合には、これだけで個人のデータになってしまいますので、そういったもの。あるいは国として、または県などでもあるのですけれども、他の県には見せたくないといったデータは見せないという形で考えています。

二つ目のアクセス申請に対する判断は、一般的な回答になってしまいますが、活用目的によって個別に判断しています。ビジネス目的よりはアカデミックの方が緩くできるのかなという感じです。

●**三橋** 制限公開の根拠は、ヒトのゲノムデータ、病歴データ、投薬データなどは個人情報保護法や政令で基本的には個人情報となっていますので、制限公開にせざるを得ないという判断です。

申請の妥当性については、ヒトに関する研究に関しては各研究機関で倫理審査委員会が確立しており、そこで了解を頂かないといけないので、われわれとして

は基本的にはそれを確認するという仕組みでやっています。

●**上島** 企業のデータの扱いに関して言いますと、各組織共通の事情として、逆に公開することに根拠が求められます。今回ご登壇された皆さまは、皆さんが指摘されているような法的根拠や、コミュニティにおける貢献が重視される場所、また各分野の研究動向や学術的な潮流を踏まえて、機関として公開していかないといけないという意味決定をされているかと思いますが、(企業でも) そのこのところの理由付けが必要で、取引先がやっているから、社会的要請として必要だから、あるいはパートナーシップを増やすために、例えば研究者に使っていただくためのデータを公開していくといった根拠づけが求められます。

区分の基準として、大きいところでは、やはり秘密保持契約のように、相手方との取引関係において情報が漏れない約束事があるかが焦点になりがちです。利用規約に同意されているか、実際の利用実態はどこまで調べていくのかということ、秘密保持契約にさらに追加している課していくことが一般的に多いと感じます。

●**海老沢** JAXA 宇宙科学研究所の場合、データセンターが利用者の可否判断をすることはありません。公開されていないデータがたくさんありますが、衛星や探査機計画に参加した人たちが、1年程度、優先的に使うための非公開です。プロジェクトごとに、装置をつくった人、運用計画に参加した人、科学的アドバイスをした人などを集めてチームをつくっていて、チームの中で、こういう人はデータを使える、使えないという判断、1年間はデータを公開しない、などの判断をします。各プロジェクトやチームごとの判断であり、所として、あるいはデータセンターとして、誰が使っている、使ってはいけないという判断はしていません。誰がデータの優先権を得るかは、衛星ごとの研究チームによって決まります。

●池内 今、改めてお話をお伺いして、それぞれの事情があることがよく分かりました。データを公開するとなると、分野を越えて企業のデータを研究者などいろいろな人が使っていくときに、その分野の暗黙知のようなものを理解していない、あるいは知らない人が使うことになります。そのときに、それをどう伝えるかということも、その先の議論としては重要だと感じました。

●林 池内さんからコメントを頂いたのですが、例えば桂樹さんのご発言で、品種の情報はライバルである他県にはあまり見せたくないというのは農業分野では分かりやすい話ですが、他分野の方から見れば、別に県同士だからいいのではないかと思われるかもしれません。文化が違うという言葉でいいかどうか分かりませんが、分野によっても事情や意識は違うのではないかと思います。このあたりは Slido でも質問が来ていましたので、後で広げられればと思います。

●八塚 ありがとうございます。ここからは、Slido で頂いた質問を交えながら話をしていきたいと思えます。

桂樹さんに対するご質問として、メタデータの一覧や、公開・非公開の表がスライドであったと思いますが、そのメタデータの定義にたどり着くまでにはどのようなディスカッションがあったのか。それから、カスタムメタデータのお話がありましたが、統一しにくいメタデータもあると思います。それをどのような方法で記録したり、管理したりすることを想定されているか。桂樹さん、いかがでしょうか。

●桂樹 メタデータの考え方ということだと思うのですが、最初、ゼロからどうやってメタデータを付けていくかという議論をした中で、世の中に存在している標準的なものをいろいろ勉強して、この辺かなということで、一番標準的な Dublin Core の項目を見たりしながら、農研機構ではどのようなデータを必要として

いるか、メタデータとして記録しているかを考えながら設計していきました。その中で、基本的に農研機構内でデータを整理するという事なので、どの予算を使ったか、どんなプロジェクトでやったかということを入れていくなど、そういうことを基準に考えています。

カスタムデータについては、分野横断的に、例えばドローンデータ、画像データを扱うプロジェクトが幾つか立ち上がって、その方たちと相談して、どんなメタデータを付けていけばいいか、検索したいとしたらどんなデータか、どういう項目が付いていたら便利かということディスカッションしながら、それぞれの分野で必要なものを付けていこうという形です。そのデータを使っている人々、幾つかのグループに聞いて回って、では、この辺を付けましょうかということをしました。

●八塚 なるほど。では、カスタムと言いつつ、実はある程度は調停したというか、ここに使うということでは考えられていたわけですか。

●桂樹 データを後で検索するときに、なるべく統制をかけたいということがありまして、それでかけています。

●八塚 分かりました。ありがとうございます。

では、他のご質問も紹介したいと思います。他から取ってきた 1 次データを加工して、2 次データ、3 次データを作った場合、その共有化を義務付けることについてはどう思われるか。各領域では、そもそもどのようなルールがあるのか、ないのか。あるいは皆さん自身、規範的にどう考えられるかというご質問が来ています。これはお一人ずつ伺いたいと思います。

●仲 2 次データや 3 次データの共有化を義務とするかというのは判断に迷うところかと思えます。ただ、学術目的で集められたデータは、いろいろな人に使わ

れて初めて価値が出てくると個人的に考えています。ですので、可能な限りデータの公開と共有をできるように、調査の段階で合意を取って、将来的には公開されるという流れを想定しています。例えば科研費などの公的な資金を得て調査をしている場合には、データそのものにはより公共的な価値があると考えられます。そういった場合には、2次データとして使えるようにしていく形がいいのではないかと考えています。

●篠田 これは大変難しい問題だと思っています。特に材料分野では、今、マテリアルズ・インフォマティクスや機械学習による材料設計のようなものがあり、そういうところに使うためにデータセットを作ったり加工したりすることがあります。基本的には元のデータがどういうライセンスをしているかに関わってくると思いますが、そういうデータセットを作るときに、ライセンス条件が違うものが混ざったときにはどうしたらいいのかということがあります。ですので、一律に共有化や公開がなかなかできないところだと思っています。

●桂樹 皆さんがおっしゃっているとおりだと思います。われわれの場合には、まだデータが集まってきていない部分もありますので、これから考えていくのですが、そんな中でも、取ったときのライセンスが同じプロジェクトの中でも違う、あるいは継続して次のプロジェクトに移って、今回のプロジェクトはこのように皆さんでデータを共有しましょうといったときに、うっかり前のものを一緒に枠組みの中で公開してしまいそうになったりします。時間軸も考えなければいけないところが難しいと思っています。

●三橋 ヒトのデータの場合は、研究参加者から同意をベースに試料を提供していただきますが、その場合には同意撤回という制度があります。同意撤回されると、データベースからそのデータを削除しないといけません。そうすると、あまり2次加工、3次加

工してトレースできないものが増えてしまうと困ります。基本的にはデータを使って論文発表ができないと意味がないので、論文発表で必要なデータは加工して載せていただくということまではある程度認めているのですが、それ以上の流通は実はあまり認めていないというのが実情です。

●上島 どこまでご参考になるか分かりませんが、コンテンツビジネスでは企業が2次利用やファンフィクションのためのガイドラインを公表することがあります。オープンデータに近いライセンス表示と、著作権表記 (all rights reserved) による権利保持を併用する運用も見たことがあります。フェアユースの範囲をどう設定していくかという、著作物の扱いと近い枠組みで考えられると思っています。

実務的には、弊社の例でいうと、あるデータ作者が作ったデータを別の方が加工して公開したいといった場合に、出どころの表示をどう書くか、その方がどのような加工をしたかどこまで明記すればいいのかということが、細かいところですが論点になっています。そのあたりは、産業界、学術会、行政も含め、何となくの慣習ができてくると、心理的な抵抗感がなくなってくるのかなと感じています。

●海老沢 これは大事な課題だと思います。私たち宇宙科学研究所のデータポリシーは、論文を出したときのエビデンスデータは公開すべき、というのですが、実はそこまで至っていません。宇宙科学の分野では、幾つかのジャーナルは、必ずエビデンスデータを公開して URL を示さないと受け付けません。「Nature」「Science」も厳しいですし、American Geophysical Union (AGU) もそうになっています。しかし、JAXA も含め、国内の大学等にそういうデータリポジトリがないのが現状です。

一方で、海外では、誰でもデータを置いて公開できるリポジトリを運用しているところがあります。国内の研究者が海外の著名なジャーナルに論文を出すため

にはデータを公開しなければいけないので、そういう優れたデータがどんどん海外に置かれるという状況が見られ始めており、これはよろしくない状況だと言われています。これは国内で私たちが真剣に考えないといけない課題だと思います。それと関連して、今、NII が主導して JAIRO Cloud を始めています。そういうところが国内の研究者のデータ公開のリポジトリになっていけるのではないかと注目して見守っているところです。

●池内 今お話を伺っていて、2 次利用、3 次利用されていった場合に、トラッキングの問題はデータを出す側にもあるし、やはり利用する側にもあると感じました。孫データやひ孫データになると追いつけなくなってしまうという状況も出てくるでしょうし、もちろんコストも莫大になっていくと思います。使う側としても、そのデータが 1 次公開されたデータなのか、2 次データなのか、3 次データなのかという判断が非常に難しく、そこを気にするとデータの再利用がしづらい状況になりそうだと思います。上島さんがおっしゃったように、やはり文化の醸成が必要であることと、出典や加工の詳細をしっかりと書くという方向に行くのか、それとも、ある程度もういいだろうということにだんだんっていくのかは、データの種類や分野の慣習によって異なると思いますが、今後の展開がどうなるのかなというコメントをさせていただきます。

●八塚 ありがとうございます。私自身はリポジトリを運営していて、もちろん 2 次データ、3 次データを作ってもらうためにリポジトリを公開していますので、そこは非常に意識しています。逆に、1 次データこそまさにオープンでないと、それを追跡するという話になると大変なコストがかかってしまいます。そういう意味では、1 次データにどのようなライセンスを付けるのか、どのような制限をかけるのかということは非常に大きな問題になってくると思います。林さん、何かありますでしょうか。

●林 そもそも他のデータを使ったときに、どのようにそのことを表現するのか。論文を書いて、引用文献として書く、あるいは謝辞という形で何かしらの貢献を示すという方法もあるかと思います。しかし、こんなデータを使ってすごく助かった、参考になった、あるいはデータの誤りがあった、ここのデータがもっとこうだったらいいという要望など、データを使ったときの表現や貢献を示すに当たっての慣習が皆さまの分野であればコメントを頂けますか。

●海老沢 他の分野でもそうかもしれませんが、私たち天文学・宇宙科学に関わる分野で、データアーカイブを作ると、みんなに使って喜んでもらえるのだけでも、それが自分の研究評価として評価されないという、若手の不満を聞くことがあります。そこで、データを作って DOI を付けて、データサイテーションが自分の業績になる枠組みができればいいという話があります。最近、宇宙科学の論文では、DOI があるときは必ず DOI を示しています。その DOI に行くと、そのデータを作ったのは誰かということが分かって、データを作ることが評価される。これが、データサイテーションが進んでいくということになるのではないかと思います。

●八塚 ありがとうございます。まさに DOI を使ったデータサイテーションは、RDF でも専門の小委員会がありますけれども、本当にこれから重要になってくると思います。

●仲 海老沢先生がおっしゃったように、成果物に対して、引用や謝辞など、DOI をフォローするようにしています。特に論文の中で、データの利用に対する謝辞の記載があるかどうかを確認しています。基本的には利用が終わった段階で成果物を登録していただくのですが、その際に論文を現物あるいは PDF で送っていただいて、一つずつ謝辞を確認するという形を取っています。

もう一つは、一連のデータを寄託していただいたり、寄託いただいたデータの利用件数が多い場合には、年に一度、該当の寄託者様を表彰する仕組みをつくっています。SSJDA を利用する側と寄託する側の双方がよりよく結びつくように取り組んでいます。

●**桂樹** われわれはデータがまだ全然集まっていない段階ではあるのですが、これからデータを作った人をきちんと評価できるように、まさにデータごとに DOI が取れる仕組みをつくっていかうとしてるところです。それでデータを作る人のモチベーションも上げていきたいと考えています。

●**三橋** 塩基配列のデータベースに関しては、公的 DB に登録するとアクセス番号という番号が発行され、それを論文に書くと、そのデータを出した人が評価されるという仕組みになっています。あとは、ヒトデータベースの利用者のガイドラインに必ず年に 1 回、利用報告を出すように決めており、その中で、論文発表をしたらそれを報告するよという決まりをつくっています。

●**八塚** ありがとうございます。大体この話題については出尽くしたかと思いますので、次のご質問にいききたいと思います。

これも組織の中の話になるかと思いますが、データ公開において、例えば知財の部署との相談はされていると思いますが、それをどのように進めているのでしょうか。個別に相談すると件数が多くなったり、時間がかかったりする場合があります。かといって、知財部署との調整なしでやるわけにもいかないと思います。この辺で、こういう苦勞をしたとか、こういう方法がありますという方はいらっしゃいますか。

●**海老沢** 私たち宇宙科学研究所のデータポリシーの中で、商業利用も OK というポリシーにするときに、JAXA としてそれが許されるのかどうか気がになりま

した。そこで JAXA の広報部や新事業促進部などの意見を伺ったのですが、特に問題はありませんでした。科学データに関しては、知財などの問題はなく、商業利用目的ということで公開できました。意外とみんな知財や商業利用に関して臆病というか、びくびくしているような感じがして、もっとオープンにデータを公開しても良いのでは、という印象を持ちました。一方、地球観測などに関しては、宇宙観測と違って、知財などの難しい問題も出てくるのかなという印象も持っています。

●**上島** どこまで参考になるか分かりませんが、私の関わった事例ですと、組織内の関連する部門——広報、法務、知財あるいは情報システムからご担当にお集まりいただいてタスクフォースを結成して、そこで定期的に会合を持つ形で、寄せられてきた案件をさばいたり、集中審議をかけたりする。組織としてそういう枠組みをつくっているところはございました。

●**八塚** ありがとうございます。やはり個別に処理するというよりも、ポリシーや仕組みをつくることの方がとても大事だと思います。

それでは、次のご質問にいききたいと思います。実際的な話ですが、どのくらいのスタッフで運用しているのか、それから、どんな分野のスタッフで構成されているのかというご質問です。これは実際にデータを提供されているところということで、海老沢さんからお話しいただけますか。

●**海老沢** 私たちの研究所では、データを作る立場と公開する立場を分けていて、基本的にデータを作るのは衛星などの「プロジェクト」です。プロジェクトは、どこかでスタートして、どこかで終了します。しかしデータセンターは定常組織なので終わりがなく、ずっと継続するわけです。私たちのデータセンターは非常に小さくて、私たちの部署で DARTS というデータアーカイブに関わっているのは数人の正職員のほとんど

が兼任、専任の派遣職員が数人、トータル 10 人足らずです。一方で、データを作る方は集中的に短期間で、例えば 1 年間に 10 人で、10 年の間に延べ 100 人ぐらいという形になります。データを作る側とデータを維持する側が違うというところがポイントだと思います。

●三橋 われわれの場合、DDBJ と NBDC が共同で運営しているのですが、それぞれに少なくとも 1 人ずつ、バイオロジーで Ph.D.を取って、なおかつ倫理等のことをきちんと勉強している、ほぼ専任のコアなメンバーがいます。それにプラスして、他をサポートする私のような IT 系の間人、外注などのスタッフで何とか動かしているという形です。

●桂樹 私はデータ戦略推進室という部署なのですが、上長が 1 人いて、Ph.D.を持つ研究員が 3 人、システムをやる人が 2 人、SE が 1 人です。すごくこぢんまりとした状況で、あらゆるデータを集めてこいと言われて四苦八苦しているところです。倍ぐらいの人がいたらいいのと思うことがよくあります。

●篠田 私の所属する材料データプラットフォームセンターは、材料プラットフォームを構築するというミッションの下で動いていますので、非常に大所帯で、正確には把握していないのですが、あるときには 60 人ぐらいいると聞いていました。その中でも、基盤の構築とサービス提供を両方していますので、構築の方でシステムエンジニアなどがかなり多くいます。ただ、要素技術の開発も行っていますので、研究者も参加しており、実際に材料分野の研究者でセンターでサービス構築を進めている方もいます。

あとは、サービスを提供する側として、私どもは併任も多いのですが、サービスチームというものがあります。こちらは 14~15 人のメンバーがいて、構築されて運用が始まったものをユーザーに提供していくところを担当しています。

●仲 東京大学社会科学研究所附属社会調査・データアーカイブ研究センターでは、任期付きの教職員を含めて 15 人程度で運用しています。その中でも、今日ご紹介したデータ提供のプロセスでは、実質的には 2 人で対応しています。もちろんデータを作る部門にはデータアーカイブのマネージャーがいて、短時間で働いてくださっている方も数人います。思ったよりも小さい組織で運用しています。

●林 先ほどの桂樹さんのお話の中で、例えばメタデータを作るのに苦勞されているということがありました。また、篠田さんのご発表の最後にも関連しますが、スタッフの中にライブラリアンというか、メタデータの扱いが得意な図書館の人が入っていると、お手伝いになるのではないかと思います。

●八塚 ありがとうございます。では、他の質問にききたいと思います。

先ほど、2 次データや 3 次データも含めたデータの利用について話題が出ましたが、それを実際どれくらいチェックされているのか。例えば、利用受託を守っているか、引用と謝辞を行っているかということは先ほどのコメントで少しあったと思うのですが、他にこういう調査をしているということがありましたら、どなたかコメントを頂けますか。

●篠田 直接の利用については、やはりデータに DOI を振って引用状況を見たり、謝辞は確認が難しいので、皆さんから「こういうものを載せたよ」という報告を頂いて情報収集する形を取っています。2 次利用がされて、それが成果につながったかは別にして、直接的に言えるデータのダウンロード件数やページへの訪問者数を一つの指標として、データの利活用状況などをモニターしている状況です。

●海老沢 私たちは衛星プロジェクトごとに、どれだけその衛星データが使われたかということでカウント

しています。たくさんの論文が出れば、良い衛星であり、良い成果ということです。一方で、データセンターの貢献と衛星の貢献の区別は難しいのです。私たちデータセンターは、いろいろなデータを集めて使いやすくしてデータ利用を促進しています。DARTS というデータアーカイブをしているのですが、どの部分で貢献したかを判断するのが難しく、データセンターとしての貢献を測るのは難しいです。衛星のデータは、宇宙科学研究所からダウンロードされることもあるし、例えば NASA から同じデータがダウンロードされることもあるのです。ですから、データセンターの貢献を評価してもらう指標は、なかなか難しいです。

●八塚 ありがとうございます。それでは、ここでいったん質疑から離れまして、ディスカッションの論点に戻ります。

制限公開の今後がどうなっていくのかということについて、皆さんのご意見を伺いたと思います。例えば、分野を越えた共通のフレームワークが必要になるのか。あるいは、既に商用もある程度あると思います。そういった学術以外の利用が進んでいくのかどうか。それから、完全公開のデータに比べると、当然、制限公開の場合はコストがかかるのですが、そのコスト負担は一体誰がすべきなのか。これは池内さんからお願いできますか。

●池内 まずコストの話からすると、やはり現時点でもトラッキングのコストや判断のコストはすごく大きいと思いました。一方で、DOI で追っていくのは、技術的に解決できる部分が今後出てきそうだと思いますので、まずは DOI を振るという慣習を根付かせることが、すごく遠回りみたいですけれども、やはり必要だという気持ちを新たにしたところです。各データアーカイブの機関においても、登壇されている皆さまは十分されているかと思うのですが、そういうことが大事だということを改めて感じました。

制限公開に関しては、できるだけ安全側に倒して制

限しているという部分もあるかと思うのです。例えば公開するという意味では、論文を公開するということについては、私たち研究者は抵抗感が少ないと思います。論文であっても剽窃されてしまう可能性はあります。つまり、データも論文もそんなに変わらない部分もあるはずですが、やはりデータに関しては、まだまだ論文と比べると公開している歴史が短い分、恐れてしまう部分はあるのかと思います。

私も個人的にもどきどきしながら公開するので、そういった気持ちも十分理解できますが、実はそんなに恐れることはないかもしれないという文化ができていたり、公開したことでこんなに良いことがあったというベストプラクティスが共有されたりしていくと、だんだん制限公開に対する考え方も変わってくるのかもしれないと思いました。分野特化ではなくて全般的な話なので、ふわっとしていますが、そのように考えました。

●海老沢 私が専攻している天文学、宇宙科学は、基礎科学の中でも特にオープンな分野だと思います。宇宙科学のデータを公開しない、地上天文台のデータを取って公開しないということは、ほとんどなくなってきています。逆に公開しなくて悪かった例はたくさんあり、過去に天文学の分野で間違った論文がたくさん出ていますが、それらはほぼ 100%非公開データから出ています。データを公開しないと信頼されないのです。

これは実は恥ずかしいのですが、私の研究所で、以前、あるロケットを打ち上げて大きな成果を上げて、大きな物理学の賞を取った人がいます。この研究所の所長にまでなりました。一方では、他の大学で、地上の望遠鏡である発見をして同じ賞を取った人がいます。どちらの発見も、今では否定されています。それらのデータは公開されていませんでした。これも名前を出すと思いが流れるかもしれませんが、現在でも、別の地上の望遠鏡ですと何十年も観測を続け、大発見をしたと言っている先生がいます。しかし、データを公開

しないのです。データを公開しないと信頼されないという流れは、はっきりしています。

私たちの分野で課題になっているのは、人工衛星をつくった人たちのデータの優先権をどのくらいにするかです。大体1年が目安ですが、それをもっと短くした方がいいのではないかと、あるいは優先権は全くなくてもいいのではないかとという議論もあります。

いずれにしろ、宇宙科学の分野に限って言えば、データ公開の流れは止まりません。過去にはデータを公開せずに間違った結果が報告されたという科学的な損失もありました。現在も残念ながらあります。

●上島 産業界から見て、アカデミズムの引用のやり方や、DOI がきちんと付されていてそれを業績としてカウントしているということは、まだまだ知られていないところがあります。世の中の的にもう少し知られていくといいのだろうと感じます。

また、コストをどう負担するかについては恐らく二つしかなくて、一つは、コストに見合う収益をどう稼ぐか、もう一つは、コストをいかにスリム化するかです。コストのスリム化に関して言うと、ある程度の利用手数料を頂いたり、何らかの有償化する方策が、(研究活動の) 持続可能性を考えると必要だろうと、皆さんのお話を伺っていて感じました。収益化に関して言うと、企業のデータであれば、製品化して商売として売りましょう、値段を付けましょうということで解決するのですが、オープンなデータを研究成果として公開するとなると、なかなか難しいと重ねて感じます。

今後の安心感をどう醸成していくかについては、リスクをどう考えるかに尽きると思います。リスクにもコントロール可能なリスクとコントロール不可能なリスクがあります。コントロール可能なリスクは、審査制など、マニュアル(人的) 対応の中で皆さんコントロールしようとされていて、企業の取り組みともかなり重なるところがあります。コントロール不可能なリスクは、恐らくコミュニティとして負うことは難しいので、抽象的な言い方になるのですが、何らかの「保

険」のような仕組みが学術界でも必要ではないかと感じました。

●三橋 提供者が獲得した研究費でデータを集めて研究論文を書くまではコスト負担者と受益者が一致しています。そうなると、論文を書くまでは研究者がある程度、独占的に使ってもいいけれども、論文を書いたら、やはり公開してもらおう。その公開のコストに関しては、今もそうですが、提供者ではなく、われわれのデータベースセンターで負担するのがいいのではないかと感じています。

●桂樹 農研機構では、農業データ連携基盤(WAGRI)に参加しています。データのアグリゲーションとマーケット、集めてくると売っていくことができる仕組みを目指してそういうものを構築しているのですが、それが、データを出すに当たってコストをどうすればいいかという一つの解決策になればと思っています。仕組みとしてはB to B to Cと呼ばれているもので、一つ目のBの人がデータを提供して、そのデータを真ん中のBの人がビジネスにできるような仕組みにします。そしてCの方々に提供する。例えば農薬を買ってもらうことでお金を集めて、データを提供することに対するコストを負担してもらおう。そういう仕組みをつくっていかうとしています。

●篠田 材料分野の方は公開するよりも非公開、制限公開の方に偏りがちで、オープンにしにくいということがあるので、どのようにオープンの方に寄せていくか、戦略的にオープンにしていくかが課題になるのですが、やはり懸念事項として、デジタルデータは一度公開して拡散してしまうと、もうそれを非公開にするのは原理的に不可能ということがあり、オープンにするところはかなり心理的ハードルがあり、慎重にならざるを得ません。こういったところは少しずつ事例を積み上げていく必要があります。一方で、材料分野の中でも、データによってはオープンにして活用しても

raitaiというニーズもあり、そのような研究者やデータの種類もあります。そういう事例を積み上げながら、制限公開と公開をうまく使い分けていければよいと思っています。

コスト負担については、NIMS は基盤なども提供していますが、国研のミッションとしてこういうことをしていますので、お金を頂かずに皆さんの税金できちんと還元できるところに関しては、まずは組織として運用を進めていく。ただ、一方でそれが商用利用などにつながったときには、また別の考え方も必要になってくるかと思っています。繰り返しになってしまうのですが、こういうところは事例を積み上げながらいろいろと練習問題を解いていっている状況です。

●仲 私からは、一人の研究者としての立場と、組織に所属する立場から簡単にリブライしたいと思います。

まず、研究者の立場としては、やはり制限公開であったとしても、利用できる多様なデータがあり、それにいろいろな人が相互に乗り入れるプラットフォームがあるというのは、研究においては本当に大事なことだと思っています。特に1次調査は費用負担が非常にかかってきますので、大型の研究プロジェクトに関わっていない方や学部学生、院生を含めていろいろな方が無償でデータを使えるのは、研究のアイデアで勝負できるという点で可能性が広がると思っています。でするので、どんどん公開できればと思っています。

その一方で、組織としてはデータアーカイブセンターが永続的に続いていかないと、データが散逸してしまうという問題が生じてしまいます。そういう意味では、何らかの形で収益を一定程度稼げる状態をどこかでつくっていくことを視野に入れる必要があると考えています。現時点では完全にフリーで、お金を頂かずにやっています。セミナーに関しては一部費用を頂いていますが、それは講師料などで、データの提供は無償で行っています。

●八塚 ありがとうございます。それでは、最後に林

さんにまとめていただけないでしょうか。

●林 練習問題を懸命に解いているフェーズにある分野がすごく多いのだろうということを、皆さまのご発表や今のディスカッションを通じて感じました。特に、使ったことを例えば DOI で表現するなどで示す、あるいは何か違う方法があるのかもしれませんが、まずはデータに関しても貢献を示していく。巨人の肩に乗るというのが研究だと思いますので、論文と同じように、データについてもきちんと貢献を示すことができるようにする。その上で、公開に当たって一定期間制限があるのか、あるいは最初からオープンになるのか、その最終的なインセンティブや必要性、どうしても公開できないけれどもいずれ公開になるかもしれない、あるいは、これまでは公開に制限なり限定がされていたけれども、もう少ししたら公開が当たり前になるかもしれない。そういう発展もあるのかなということをお話していただき、皆さまとの議論を通じて考えました。

気の利いたまとめになっていないかもしれませんが、こんな感じでどうでしょうか。

●八塚 ありがとうございます。これでパネルディスカッションを終わりたいと思います。パネリストの皆さま、本当にありがとうございました。