第3回 SPARC Japan セミナー2013 「オープンアクセス時代の研究成果のインパクトを再定義する：再利用とAltmetricsの現在」

ディスカッション

林　和弘	（科学技術・学術政策研究所）
池内　有為	（筑波大学大学院）
Mark Hahnel	（figshare）
Jason Priem	（ImpactStory）
坊農　秀雅	（ライフサイエンス統合データベースセンター）
大園　隼彦	（岡山大学附属図書館／DRF）

●**林**　皆さん、こんにちは。オープンアクセス・ウイークの大イベントが 2008 年から続いており、SPARC Japan でもセミナーを都度開催していますが、今回で 6 年目となり、しかも SPARC Japan 自身は今年で 10 周年です。10 周年記念のオープンアクセス・ウイークにこのような刺激的な、最先端の話題でセミナーができることを非常にうれしく思っております。西薗さんと私と松本さんの 3 人で企画したセミナーにこれだけの反響を頂き、ありがとうございます。

始める前に一つだけアピールしたいのですが、「オープンアクセス時代の研究成果のインパクトを再定義する」というタイトルは、われわれがオリジナルに考えたものです。その後、オープンアクセス・ウイークのタイトルが「インパクトを再定義する（Redefining Impact）」とされ、オープンアクセス・ウイークのタイトルを先取りしてわれわれがセミナーを企画していました。その点は日本人として、日本のプレゼンスが多少でも高まったのではないかと勝手に思っている次第です。

しかも今回、お忙しい中、まさに Altmetrics の提唱者である Jason さんと、研究論文以外の研究成果の公開とそのインパクトを計測するサービスの展開を一番進めている Mark Hahnel さんのお二人と一堂に会してお話しできることは、非常にありがたいことです。実際に話していただいたら、こんなに話が合うとは想像もしていなかったので、それもうれしいですし、それに対して日本の研究者の活動の事例として坊農先生からお話がありました。図書館も負けずに Altmetrics を既に導入しているのだというお話が大園さんからあり、Altmetrics の利用はまだまだこれからですが、反応が薄い段階で始めておくから今後伸びていくことが確認できるわけで、例えば 10 年後の大園さんが楽しみになるわけです。

そして、大事なポリシーや物事を決めるのは大抵イギリスであるというのが出版業界の定説となっているので、池内さんからお話しいただいたという格好になっています。これまでのご講演を受けて、ディスカッションに入っていきたいと思います。

今回は「研究成果のインパクトを再定義する」というタイトルで、図書館職員（ライブラリアン）が今後どういう活動をしていくべきか、また、インフォメーションプロバイダー、ジャーナルパブリッシャーとしての学会、出版社、印刷業者、ベンダーがどう考えていくかを中心に議論ができればと思います。さらに、今回は珍しくありがたいことに参加者に研究者がたくさんいらしていますので、ご意見を頂くとともに、研究者の方に対してどうアピールできるかという議論も加えることができればと思います。

それでは早速ですが、全体的な質問でも個別の質問でも構わないので、どんどん進めていきたいと思いま

す。

●**高久**　筑波大学の高久です。どの方のお話も面白くて参考になったのですが、2 点お聞きしたい点があります。

研究をしているとデータが大量に出てきます。例えば共同研究で 1 回実験をやると、ビデオデータなどだと 1TB のポータブルディスクが 1 個、2 個と必要になっていきます。自分で取ったデータなので自分自身はもちろんよく分かるのですが、これを他人と共有するというのは非常に難しい仕事になります。他人と共有するためには、そのデータが何に由来して、どういう条件で取られたのかをきちんと書いておかないと、自分自身も分からなくなるし、相手の人も分かってくれません。すなわち、共有するためには、それがどういうデータなのかを記述しなければいけないのです。

では、それを誰がやるのかというのが一番難しい問題です。坊農先生がおっしゃったような新着論文レビューなど、ライフサイエンス分野のデータベースの標準化されたデータやツールの世界で、比較的うまく回っているのではないかと思うのです。しかし、そうした標準がない分野ではどうすればいいのか、常に私も頭を悩ませています。データを作って公開して、ツールを提供しても、それ自体は研究業績にはならないので、誰がやるのかという問題と、それをどの人のどういう業績として認めるのかという点について、お考えをお聞かせいただければと思います。

●**林**　この質問に対して、坊農さんからはある特定の領域としてのお答えをいただき、汎用的なデータに関しては Mark から、もし池内さんが可能なら、ライブラリアンはどうしたらいいかということに関してコメントをお願いします。

●**坊農**　論文のデータや DNA の配列データなどに関しては割ときちんと決まっているのですが、それは長年、そういうものをどんどんアーカイブしてきて、「これがあると再利用しやすい」といった蓄積があるからで、全ての生命科学分野のデータがそうなっているわけではありません。特に動画のデータ、蛍光で測定した発生過程のデータ、MRI や NMR のデータ、波形のデータなどは、各研究者のローカルなコミュニティーごとにはどんどんシェアしようという動きがありますが、どういう基準でメタデータを付けたらいいのか、どういう基準で配布したらいいかということに、皆さん悩んでいます。

本当は CC0 が一番楽なのですが、できる限り制約のないライセンスで交換するようにするのがいいと私は言っています。それはやはり使いたい人間同士が取り決めていく。やはり他人のデータも使いたいので、その人のデータがよく使われているとなれば、もっとその人がリスペクトされることになります。研究者の間ではそういう取り決めがあると思いますが、皆さんは数値化するのが好きだから、どうしてもインパクトファクターの高い雑誌の査読を通ったものとなってしまうのですが、その辺が Altmetrics 関係で数値化されていくといいと思います。

また、論文の中でどれだけそのデータが使われたかというのは、引用されにくいのです。ですから、そういうものがうまくカウントされるような仕組みをつくり、実際にカウントしていく。それを指標に変えていくといいのではないかと思います。私が研究している遺伝子発現データはちょうど真ん中ぐらいで、10 年ぐらい前からどんどん規格化しているのですが、今も新しい種類のものがどんどん出てきて落ち着かない状況なので、そのように思っています。

●**Hahnel**　今のお話に 2 点ほど付け加えたいと思います。一つは、誰がデータセットを書き、誰がそのデータセットの背景にあるメタデータやなぜそういうデータになったかという理由を書くのかということです。出版社側も、論文を出版すること、そして今はまだ論文が主流であることと、新しく出てきたデータセットというものの間にギャップがあるという認識は持って

います。そして最近は、データジャーナルを発行する出版社もたくさん出てきました。午前中の講演でもNature Pulishing Group から Scientific Data というデータジャーナルが発行されるという話が出ました。Nature Publishing Group は、メタデータをジャーナルにおける論文に当たるものとして、データセットに焦点を当てたジャーナルを創刊しようとしているわけです。データセットとメタデータのセットで通常の論文と同じように扱うのです。私たちもこのジャーナルにかかわっていますが、そのほかにテラバイト単位のビッグデータに関しては Amazon Web Services を利用しています。大学ではたいてい大容量のデータを保存するスペースを保有しています。ですから、データがオープンになってさえいれば、サイズはそれほど問題ではありません。ストレージの予算がないとなると、それは問題になってきますが。

●**池内**　まず理想を言ってしまえば、研究者がメタデータを付けることです。私が高久先生に代わってメタデータを記述するよりも、そのこと（研究）を一番よく分かっている高久先生が付けてくださるのが理想です。

ただ、今はまさに過渡期にあって、データの評価自体が定まっていません。先ほど、NSF では今年 1 月から研究業績として論文や特許と同じように生データやソースコードが書けるようになったという話をしました。日本においても、論文と同じ比重でデータが評価されるようになれば、恐らく研究者も時間を割いてデータに情報を付けてくれるのではないかと思うのですが、今はまだそうはなっていません。この状況では、研究者が次の研究をせずにそちらに時間を割くということは難しいでしょう。

特に日本においては、論文の生産性が伸び悩んでいることが問題として指摘されています。その原因はいろいろありますが、とにかく日本の研究者は研究時間が足りないのだと言われていますし、元図書館員としては、大事な研究時間を業績の評価対象にならないデータのアーカイブに使うぐらいならば、次の研究に使ってほしいというのも本音としてあります。ですから、今後このデータの評価がどうなっていくかが一番のポイントではないかと思います。

●**林**　インセンティブですよね。データを共有、評価することで楽しくなる、お金がもらえる、研究費がもらえる、教授になるなどの、便益があればいいと思うのですが。

BioMed Central（BMC）はオープンアクセス・プラットフォームの先駆けとして、データのことも無視していないと思うので、コメントを頂けませんか。

●**フロア 1**　BioMed Central に勤めております。論文に付随するデータの取り扱いに関して、BioMed Central では、今年 9 月から出版される論文の中のデータは、図表も含めて全て CC0 で出版されるように変わりました。そこは大きなところだと思います。

また、GigaScience という、BGI がスポンサーしている、ビッグデータにフォーカスしているオープンアクセス・ジャーナルがあるのですが、そちらで発表されるデータには全て DOI が付いていて、誰でも再利用できる状態で公開されています。その意味では、データの再利用に関しては早くから取り組んでいると言えるかと思います。

●**フロア 2**　Nature Publishing Group に勤めております。先ほど Hahnel から Scientific Data の話が出ましたが、NPG から来年 4 月に Scientific Data というパブリケーションが計画されており、もうサブミッションが始まっています。これはデータのパブリケーションで、メタデータを記述した上で、それがピアレビューされて、結果的に出版物として発表されるというものです。

●**林**　補足すると、高久さんの質問に答えたジャーナルを既に創刊しようとしているということなのですね。データの投稿規定のようなものがあるわけだから、そ

のルールに沿って投稿すると、その後の査読も含めて、ある程度フィルタリングがかかって整った体裁でオープンになるので、引用が可能にもなり、評価にもつながる可能性があるということです。

●**フロア 2**　データのプラットフォームとして、figshare も使われていますが、標準的に使われているデータのプラットフォームであれば、どれでも Scientific Data のデータのプラットフォームとして使われるようになります。

●**林**　このデータを保持して公開する話は機関レポジトリを進めている人にも非常に大事です。つまり、最近では機関リポジトリに紀要を載せていることに関して一定のルール作りと運用の経験をお持ちになっていますから、ライブラリアンにとっても無縁ではない話だと思います。

●**フロア 3**　WDS に関わっております。ICSU-WDS が東京にインターナショナル・プログラム・オフィスを置くことが決まって、2 年前からこのフィールドの勉強をさせていただいている最中で、私自身のバックグラウンドは地球物理学です。

2 年間、いろいろ国際会議や WDS のサイエンスコミッティーや CODATA などで何が議論されてきたのかを見させていただいた中で学んだことが、例えば Data Citation が、データの DOI をリファーすることで、Citation Index がメトリックスになっていく。彼らのもくろみとしては、データがたくさん引用される良いデータのプロデューサーは良い研究者として評価され、論文と共に研究者の評価軸になっていく。同様に、データジャーナルがメタデータの出版プロセスとして確立していくことで、新しいデータ・ソーシャル・ブリッジのようなものをつくりたいということだと思います。そういう理解でよろしいでしょうか。

また、そういう流れの中で、Altmetrics の中で、データというものが普通の論文と同じように統合されて研究評価になっていくものなのでしょうか。

●**林**　誰が評価すればいいでしょうか。Jason、今の評価に関する質問に対して何かコメントはありますか？

●**Priem**　質問の内容を完全に理解できていないかもしれませんが、まずはイエスとお答えしておきます。DOI は、データサイテーションを活用するための有用なツールだと思います。DOI の使用とは関係なく、データサイテーションを追跡する試みが広く行なわれていて、ある程度の成功は収めています。しかし、それが大規模に行なわれるようにするためには、一つ一つのデータオブジェクトに一貫した識別子をつけることが絶対に必要となってきます。データオブジェクトが科学者の評価や報酬のあり方を短期的、あるいは中長期的にどう変えていくのか、非常に興味をそそられるところです。

短期的に見れば、データは指標のような役割を担うものとして認識されていくと思います。あなたは優れた論文をたくさん持っている。彼女も優れた論文をたくさんもっている。でも、あなたには質の高いデータセットもあるとなれば、それはあなたのアドバンテージになります。政府をはじめ、助成団体は非常に現実的です。中長期的にオープンデータの価値がきちんと確立されていき、私たちがデータ再利用の価値を示す事例や大規模な定量的エビデンスを蓄積していけば、データ収集専門の科学者が活躍する未来というのも容易に想像することができます。データ収集専門の科学者は、データセットにきちんとしたメタデータをつけて公開する。それが彼らのメインの仕事となり、それに対して報酬を得るというようになっていくでしょう。

私たちは、ImpactStory でデータセットのインパクトを測定するメトリクスを多く集めることによって、今、申し上げたような未来をより早く実現できるのではないかと思っています。なぜなら、助成団体にそのデータを提示することができるからです。当然、それ

は論文ではありません。しかし、それが真に重要な論文 10 本につながればいいのです。ImpactStory を使えば、そのことを示すことができますし、恐らく ImpactStory によってこの流れは加速されていくでしょう。科学者はスクープを心配しますが、それでも論文を発表します。論文を発表して他人に自分の発見を知ってもらうことは心配しないのです。なぜなら、論文は自分の評価につながるものだと知っているからです。データの場合もその延長で考えればいいことですし、むしろより良いものにできる可能性があります。今はまだデータセットの活用事例は限られていますが、これからはもっと広い範囲で使われるようになっていくと思います。

●Hahnel　天体物理学の分野や CERN の Higgs 粒子研究のデータでは既にそれが行なわれています。データは公開され、他の研究者による再分析が行なわれています。既に同じデータを使った多くの論文が発表されていますが、それでもノーベル賞を受賞したのは Higgs 博士でした。

●林　一つ追加すると、Linked Open Data という形で、学術の世界ではない一般の世界でもいろいろなデータにリンクを張ってオープンにするという動きがあり、学術コミュニケーションのさらに外側にオープンなデータが存在するので、そのデータも活用することで新しいサイエンスが生まれる可能性もあるかもしれないということです。

●フロア 4　学術出版社に勤めております。Jason さんに質問です。私は出版社として、ソーシャルメディアの中で論文をどうやって正確にキャプチャーするのかということに興味があり、DOI があればもちろんなのですが、出版社独自の URL、PubMed の URL もあれば、論文のタイトルだけを言及している場合もあります。例えば日本の新聞などでは、オリジナル出版社のウェブサイトへのリンクなどを付けることはまずないと思います。Altmetrics が重要になってくると、書き手側のマナーに関する教育も必要になってくるかと思いますが、その辺りをどうやってキャプチャーしているのか、お伺いします。

●Priem　とてもいい質問です。確かにそれは Altmetrics にとって大きな課題となっています。短期的には、幸運なことに十分な量のデータがありますので、一つの論文に対するメンションの多くを無視しても、まだ見つけることができます。さらに、論文の参照のされ方に偏りがあることを示す適切なデータはありません。二つの論文を比較していて、片方は 100 ツイート、もう片方が 2 ツイートだったとします。その中で、私たちは二つに一つくらいしかツイートを見つけられないでしょう。しかし、それはどちらの論文にも言えることです。私たちが持っているデータが欠けていたとしても、恐らく二つの論文の関係に関して、組織的バイアスは持ち込んではいないはずです。長期的には、タイトルのテキストマイニングに関して非常に興味深いことが行なわれるようになっていくのではないかと思います。さらに高度なテキストマイニング、例えば「先週、日本の研究者グループがカンガルーに関する発見をした」というような一文のテキストマイニングが可能になるかもしれません。そうしたデータがあれば、恐らくその論文を見つけることは可能になりますが、そのためにはもっと技術を洗練させていく必要があります。

●フロア 5　情報発信に関わる仕事をしています。今、誰でもデータセットを使えるようにならなければいけないという風潮があります。それに対して、歴史的な学術論文のコピーライトを出版社が全部持っていってしまい、それですごく金儲けをしているという批判があり、それがオープンアクセスにつながっていったというか、SPARC ムーブメントはそういうことから始まっているように私には見えます。

　そう考えると、先ほどのデータセットやデータベー

スを登録するのはいいけれども、そこにまたコピーライトを付けてという話にはならないのだと思います。その辺の流れやバランスについて、パネリストの方で何かアイデアがあれば、お聞きしたいと思います。

●Hahnel　データセットのコピーライトに関しては、国によって違いはありますが、「空は青い」といった事実にコピーライトを付けるのはかなり難しいと言えます。事実にはコピーライトは付与できないのです。ただ、究極的なことを言えば、研究者が研究するのは金銭的な報酬のためなのか、それとも科学の発展と自分の研究に対する他の研究者からの賞賛を見返りとするのかということになってくるのではないかと思います。その好例が、最近起きた NASA の閉鎖でしょう。人々は NASA に納税者が収めた税金はどうしたのかと問いました。つまり、アメリカ国民は NASA から何を得たのかということです。NASA からは多くの技術が生まれました。NASA が直接開発したわけではありませんが、NASA があったからこそ誕生した技術がたくさんあるのです。例えば、衛星テレビなども NASA の技術を元に生まれたものですが、NASA がそのコピーライトを保有しているわけではありません。なぜならその研究の資金は納税者が納めた税金でまかなわれていたからです。民間企業はその技術を利用することができますし、それは素晴らしいことだと思います。民間資金で行なわれた研究であれば、そのデータに好きなようにコピーライトをつけてもいいでしょう。ですが、公的資金で行なわれた研究では、コピーライトをつける権利はないのです。

●林　坊農先生、著作権上の扱いで苦労された点など、ご披露いただける情報はありますか。

●坊農　われわれも統合データベースプロジェクトをかれこれ 6 年ほどやっています。当初は、皆さんにどのようにデータベースを使ってもらったらいいのだろうと、われわれも勉強しました。そのときに一番見習うべきものとして、NLM（アメリカ国立医学図書館）の PubMed などの活動がありました。あれは完全にパブリックドメインで公開されていますが、それを日本語で出されたデータにやるということには、あまりにも抵抗が大きかったのです。

そこで、NII の弁護士さんなどを兼任されている人などと相談し、結果として CC BY をわれわれの提供するデータの最もベーシックなライセンスにしようということになりました。もちろん、シェアライクでないと困るものは個別に CC BY-SA にしたりもしています。われわれは公的資金でやっているので、コマーシャルなどはなく、アカデミックオンリーという制限などはせず、誰でも自由に使っていただけるようにしています。

それがどのように使われたかは示してくださいと言っていますが、今のところトラッキングは特にしていません。ただ、昔、サイエンス・コモンズをやられていた方を日本にお呼びしたときに、そういうものはしっかりした方がいいと言われていたのを思い出しました。どれぐらい使われたかをトラッキングすることは大事だと思いますが、なかなかできないので、これからは DOI を活用して、どれぐらい使われているかという情報もクロールした方がいいのではないかとは思っております。

●林　今、Creative Commons の 2.1 を使われていますが、3 に上げない理由は何でしょうか。

●坊農　ずっと 2.1 でやってきたからです。3 に上げて、法律的な解釈などがいろいろ変わると、弁護士さんに相談するなどして、また一から調査しないといけないのですが、お金がないので上げることができないのです。

●フロア 6　農林水産省の農林水産研究情報総合センターにて、図書館職員として図書館システムの管理をしています。

私は農業が研究のフィールドですので、論文も書いていますが、例えば新しく作った品種について紙 1 枚で説明するような、論文を農家の方に分かりやすく解説するペーパーなども作って公開しています。研究成果のインパクトという意味では、もちろん研究者に評価された論文のインパクトも当然ありますが、例えば熊本で「森のくまさん」という品種が売れるようになっていて、その品種を普及したということが研究成果につながることがあります。

そうすると、一概に論文のインパクトだけではなく、ウェブで公開しているもののインパクトをカウントする必要が出てきます。これまではページのアクセス数などで見ていたけれど、先ほど、ImpactStory のお話でページのインパクトも測定できると言われましたが、今後はそういうところが重要になってくるかと思っています。

そういうものを、個人ではなく組織的に把握する方法はあるのでしょうか。もちろん研究者個人も見たいだろうし、それを評価する組織なり研究所、機関という単位でも見たいだろうと思うのです。

●Priem　いい質問ですね。私の答えはごく短く、ノーということになります。来年あたりには可能になると思いますが、今はまだできていません。

●**フロア 6**　今、API のドキュメントを読ませていただいていて、ページに張っておけば誰でも分かるので、当面はそれで対応したいと思います。今日はとてもいい情報を頂けて、感謝しています。

●**大園**　個人的な質問をしてもいいでしょうか。僕の発表でもそうでしたが、国内で Altmetrics はまだまだ普及しておらず、今はあまり利用できないというのが正直な感想なのです。海外の研究者や大学・機関の間では、Altmetrics は評価ツールとしてどれぐらい認知され、使えると思われているのでしょうか。

●Priem　私は、今はまだ Altmetrics の黎明期だと考えています。今日、この会場にいる皆さんは Altmetrics の最先端にいる方々なのです。思い出してください。Bibliometrics や citation mining が提唱されたのは 50 年代後半ですが、それらでさえ 20 年から 30 年の間、それほどの影響力は持っていませんでした。ですから、新しいメトリクスが広がっていくスピードについても過度な期待を寄せるべきではないと思います。とはいえ、Altmetrics を早期に導入した人たちは、Altmetrics を熱心に、そして非常に効果的に活用しています。例えば、ワシントン大学の水産生態学者スティーブン・ロバーツ博士は、準教授から教授に昇進する際、自分の査定資料に ImpactStory による数値を入れたそうです。そうしたところ、彼が作成したスライドのダウンロード数、さらには全スライド分を合計したトータルダウンロード数を見た外部レビュアーの一人から非常に熱心な返事を受け取ったのです。トータルダウンロード数は査定資料には入っていなかったのですが、そのレビュアーは、以前であればこうした論文を読むことがなかったような人々にまで広く読まれているということに興奮と感動を覚えたそうです。これはごく一例ですが、われわれは現時点で既に Altmetircs が実際にキャリア形成に影響を与えているというエビデンスを蓄積しつつあります。

●**林**　Mark の講演で、論文が引用される前に figshare のアクセスが 1000 回を超えているという話がありました。このように、論文の引用よりも早く研究成果のインパクトを測れるようになっていることが重要ですし、あるいは、これがもしかしたら引用数の先行指標になるかもしれない。これが Altmetrics の持つポテンシャルです。いろいろなケーススタディーを積み重ねていくと、そのうち Altmetrics の効用が分かるのではないかという議論だと思います。

●**高橋**　NII の高橋です。数年前までは大学図書館に勤めていました。

今日のお話で、Jason さんがおっしゃった、今や全てがウェブに移っていくことを前提に考えなければいけないというのがとても印象的でした。では、図書館は全てがウェブに移って行ったときに、どこを担うことになるのだろうかと私はとても気になっています。それは恐らく午前中に Mark さんや池内さんがおっしゃったように、研究データの管理や、そのための教育というところに図書館に役割が残っているのだろうという気はしたのですが、他に何か思い付くものがあれば教えていただければと思います。また、大園さんは何がやりたいかということも聞いてみたいです。

それから、Altmetrics という評価の部分ですが、従来、図書館は評価には関わらず、データを集め、それを研究者に提供するということに徹してきたわけです。そこで、今、日本では機関リポジトリの多くを図書館が担っていて、そこに Altmetrics を付けるというのは、図書館が評価に足を踏み入れることになるのではないか。研究者として見たときに、図書館が評価の世界に乗り出すことについてどう思うかを研究者の先生方にお聞きしたいと思います。

●Hahnel　まず言えるのは、図書館は今後、恐らく大学で最も重要な施設になっていくだろうということです。それは、研究者が研究に集中することが許されれば、より多くのコンテンツ、より多くの成果を生み出せるようになるからです。図書館は研究成果およびそれらへのアクセスを管理し、それを適切な場所に移動・保存し、適切な利用者に提供する場所です。例えば、私は博士課程に在籍中、図書館に実際に足を運んだことは一度もありませんでしたが、図書館は毎日利用していました。

●**池内**　私も Mark とほぼ同意見です。大学図書館は、ここ十数年、資料がどんどん電子化され、その動きはもはや止まりそうにありません。研究者も学生もどんどん図書館に来なくなっています。つまり、図書館という建物に来なくとも、研究室や教室にいながら、自分のリサーチに必要なデータを得られるようになってしまったのです。「場としての図書館」という議論も盛んに行われ、ラーニング・コモンズなどが建てられていますが、あれが図書館の中にあるという必然性も、個人的にはまだ見いだせないのです。図書館外にラーニング・コモンズがつくられているという例もかなりあり、非常に危機感を抱いたというのが、私が図書館界に研究者として戻ってきた理由の一つでもあります。

データの話に私が飛び付いたのはまさにそこで、たくさんのデータを図書館に集めよう。これは先ほどのプレゼンとは若干矛盾するのですが、外部リポジトリ、それこそ CERN などの大きなリポジトリなどはもちろん活用するし、figshare もどんどん活用するのですが、図書館は全てを外に出してはいけないと個人的には思っています。もしデータの価値が今まで以上に、社会的にも研究界でも評価されるようになったとき、それを図書館が握っていたら、大学から大学図書館がなくなるということは絶対にないと思います。逆に、大事な研究のデータというものを図書館が全て外に任せて出してしまったら、それこそ図書館のアイデンティティーは危機に瀕するのではないかというのが私の意見です。

ということで、高橋さんの「さらに他に役割はないか？」という質問に対する答えではないですが、データは大学図書館が一部でもいいから握っていよう、というのが私の本音です。

●Hahnel　教育に関して、1 点付け加えたいと思います。Jason が言ったように、この場にいる人たちは全員 Altmetrics やオープンリサーチに関して日本の最先端にいる人たちです。挙手していただいたのを見ると、ここにはかなりの数のライブラリアンがいらっしゃいます。ライブラリアンは知識を持っていますから、研究者を教育する役目を担っていただきたいと思います。

●**大園**　池内さんから「場としての図書館」という話が出ました。僕は今、医学分館に勤めているのですが、

ちょうど耐震改修中で、非常に狭い仮設の図書室に引っ越しているのです。その図書室には本や雑誌だけ置いているのですが、1 日の利用者はたった 100 人前後なのです。やはり研究者や学生は、電子的な環境が備わってなければ図書館には来ないのだということを実感しました。

今回 Altmetrics に触れてみて、新たに研究評価の業務に図書館が関与できるようになることで、研究成果の公開につなげていくことができるのではないかと感じました。図書館員もウェブのスケールの研究を意識してサービスを考えていく必要があるのではないかと思います。

●**坊農**　今は、研究の評価は外部の人にしてもらっている状況です。一方、大学図書館の人というのは内部の人ですから、一部のよく分からない人に評価されているよりはいいのではないかと思います。なぜそれが嫌だと思うのか分かりません。

●**林**　もっと積極的にということですね。

●**坊農**　むしろ権力を握ってうれしいのではないですか。

●**フロア 2**　私は最初の仕事が図書館員で、1999 年まで 10 年ほど勤めましたが、それまでの仕事と、今の図書館員の皆さんの仕事が相当変わっているということを前提として、あえて申し上げます。

図書館員がどういう仕事をしていくかというのは、図書館員の方々が選ぶこともちろんあると思いますが、一方で世の中のデマンド、組織としてのデマンドがあってしかるべきだと思います。これは国によって大きく違いますが、例えばオーストラリアの図書館員は、相当の数の人たちが、もともとデータに明るいという強みを生かし、リサーチ・アドミニストレーターに転向していっています。そのように非常に積極的に動いている国もあります。オーストラリアが特別だったのは、恐らく ERA という枠組みに組織としてアダプトするために、どうしても内側にデータに明るく、それを担当する人間を持たなければならず、そこを学内で手当てするために図書館員の方が動いたという背景があると思います。

ですから、必ずしも日本の図書館員の皆さんがどうしなければいけないということではないのですが、必ずしも自分がどうしたいということだけで図書館員の仕事が決まっていくものではないと思います。

●**フロア 7**　学術出版社に勤めております。データのシェアに関して、今、出版業界では出版倫理が一つ大きなトピックスとなっていて、plagiarism（盗用）やデータパブリケーションの問題がよく取りざたされます。

日本でも最近、臨床試験論文のデータ改ざんがあり、論文が取り下げられる事件がありました。元のデータがきちんと登録されていれば、アナリシスのところで改ざんが防げたのかもしれませんが、元のデータ自体を改ざんされてしまってはどうしようもありません。こうした研究者の不正に対抗し得るようなアイデアがあれば、ぜひお聞かせいただきたいのですが、いかがでしょうか。

●**Hahnel**　まず、データが公開されるようになれば、それを再分析することができるので、不正は見つけやすくなります。データが改ざんされている場合、一定のパターンが見られます。データが公開されていなければそれをテストすることはできません。グラフだけからは、そのデータが虚偽かどうかのテストを行なうのは不可能です。これは、正しい方向に向かうための第一歩です。データの検証はデータが公開され、誰でも見られるようになっていて初めて可能になるのです。もしデータの改ざんがあれば、通常とは違うパターンが現れてくる可能性がありますが、実際には改ざんはかなり難しいでしょう。

●**Priem**　私からは二つ申し上げたいことがあります。

一つは、オープンソース・ソフトウェアのコミュニティーで知られている「リーナスの法則」のことです。これは「たくさんの目があれば、どんなバグも深刻なものにはならない」という法則です。あなたが書いたコードを 100 人、1000 人、10 万人という人が見るようになれば、以前は見つけるのがむずかしかったバグも簡単に見つけられるようになります。これはデータについても同じです。

二つ目は、先ほど Mark が話していた、数値で検証できるということの実際の例です。「ベンフォードの法則」と呼ばれるもので、自然界に出現する数値の分布は一様ではないという法則です。自然界の数値は一定の法則に従っていて、面白いことに 9 よりも 1 の方がはるかに出現率が高いのです。これはさまざまな種類の自然数に共通する、確立された知見です。もし数値を操作していたとすれば、この自然数のパターンとは異なったものが出てくるでしょう。このパターンは極めて強固な法則であり、アメリカの法廷では証拠として認められているほどです。会計上の不正を見つけるためによく使われています。

想像してみてください。世界中のあらゆる科学データがオープンになり、自動マイニングにかけられ、こうした不正が見つけられるようになったら、素晴らしいと思いませんか？　そういうものが見つかるようになって、私たちは失望するかもしれません。恐らくこの技術はたくさんの不正を掘り起こすことになるでしょう。でもそうすることによって、私たちは、より強力でより真実に近い科学を手にすることができるのです。

●**林**　もう一つ付け加えると、Altmetrics が持っている可能性として、ネガティブインパクトも測れるので、データライブラリーに何らかの不正があったときには、Twitter などで炎上するわけです。そうすると、前よりは早くそのような動き、不正を掴むことができる。Jason の方がカッコイイ手法ですが、より現実的にはそういう形で、不正を事前に止めるというより、起きたときにすぐ把握できる状態にすることで抑止になるという世界が来るのではないかと思っています。

●**フロア 8**　今日の話は未来のことだと思うのですが、過去の歴史から何が学べるかということも考えたいのです。先ほど Jason さんが昇進の際の Altmetrics の利用を新しいこととして紹介されましたが、恐らく 50 年前に誰かがインパクトファイターを書いたら、それは非常にインパクトのある昇進書類として受け取られたかもしれません。しかし今、このインパクトファクターがいかに正当な評価を歪めているかというのは、誰もが認識していることだと思います。

Altmetrics に関しても、同じことが恐らく起こるでしょう。特にこういうメジャーメントというのは、その仕事が今、どれだけはやっているかということに影響されます。しかし、新しい人を採用するときには、その人がどれだけ論文を書くか、お金を取ってくるかではなく、10 年後、20 年後にどのような爆発的発展をするかという可能性を見て採用するのが本来の人事だと思います。

そこで使えるようなシステムがなくならないようにしたいと思いますので、ぜひ、今日のスピーカーの方もそれに関して考えていただければありがたく思います。

●**逸村**　筑波大学の逸村と申します。今のご質問の方に修正を加えたいと思い、マイクを取りました。Altmetrics にどういう可能性があるのかは、これから真剣に考えていくべきことだと思います。ただ、その前にインパクトファクターについての話をあらためてしたいと思います。

来月号の月刊「化学」に池内有為と私とで一緒に書いているのですが、インパクトファクターは 1970 年代に Eugene Garfield が表明したものですが、当時は全く研究者から相手にされなかったのです。日本において研究者がインパクトファクターを認知したのは、1990 年代半ば、大学院の重点化がささやかれるころ

からで、私が調べている範囲では、まず北里大学が取り上げたのです。北里大学は非常にフェアに「インパクトファクターというのは雑誌の評価であって、研究者の評価ではない。けれどもこういうやり方もある」ということをはっきり書いています。それがなぜ日本において研究者評価に使われるようになったのかは、それはそれで問題ですが、また別のストーリーがあったのだと思います。

そういう事実関係を指摘した上で、これからAltmetrics だけでなく、今後、いろいろな SNS 関係のものがまだまだ出てくると思います。それについて、どのような可能性があるかとお考えでしょうか。

●Priem　二つともとてもいい質問だと思います。一つ目は研究者の将来的なパフォーマンスをどのように予測できるかということですね。研究者を数値的に評価してばかりで、その人の重要な資質に目を向けていないということはないのか、人としての資質こそ評価の対象とすべきものではないのかというわけです。これは非常に興味深く、かつ重要な指摘です。しかし、実際はその逆のケースの方がはるかに多いと思います。私たちは科学者として、第一印象や直観は定量的なデータほど信頼されないことを知っています。人は確証バイアスに流されやすいのです。

この流れで思い出したのが『マネーボール』という映画のことです。これはある男が野球チームを作り上げていく物語ですが、過去 100 年、主流だった選手集めの方法は、「あいつが気に入った。あいつはいい顔をしている」というものでした。これは実話に基づいた映画ですが、今までのやり方ではなく、「個人的な感情は抜きにしよう。どんな外見かということも無視しよう。その代わりに信頼するのはデータだ」と言って選手を集め、これが大成功につながっていきます。

良質かつ大量のデータがあり、かつそのデータの解析してくれる賢い人がいたとすれば、その結果は大抵、賢い人が直観で導き出す場合よりもいい結果になるのです。『マネーボール』で描かれていたように、比較的貧しいチームがお金持ちのチームよりも優れた結果を出すことがありますが、それは数値に基づいた方法論を持っているからです。今後、比較的予算が乏しい国や機関が Altmetrics を使ってアカデミック版マネーボールとでも言うべきアプローチを採るようになるでしょう。そして、資金力のある国や機関よりも高い成果を達成するといったことが起きてくると思います。

二つ目の質問に関しては簡単にお答えしておきます。Altmetrics のための新しいデータソースは常に開発されているということです。Altmetrics と従来の引用との極めて重要な違いは、従来の引用は何百年も使われてきたということです。Altrmetics のデータソースは、非常に短期間に現れては消えていきます。何よりも、Altmetrics には非常にオープンで普遍的なアプローチが必要とされます。私たちは、できる限り多くのデータソースを収集し、データを公開して人々がデータを好きなように解釈できるようにします。そうすることによって、何かが欠けていれば、きちんと説明できるようになるのです。

●林　この話題はもう少し掘り下げたいのですが、時間が過ぎてしまっているので、この辺りでクローズさせていただきたいと思います。

今日、オープンアクセス・ウイーク 6 年目のセッションとして最も印象深かったのは、いかにしてオープンアクセス化するかという議論に一度もならなかったことです。過去 5 回はその話をずっとしてきましたが、今回は、これからどうなるというところだけをインテンシブに議論できました。

OA 化そのものを含め、OA と Altmetrics の活用に際して実際にはまだまだ難しい問題がたくさんあります。しかし、Jason と Mark のお二人に言っていただけたように、ここに居るわれわれはカッティングエッジなイケてる人たちなのです。また、先の Jason の講演の質問で出た言葉を借りると、Ninja のように多種多様な（研究成果のインパクトを測る）新しい術を習いましたので、それをご自分のスキルアップやプロジェクト

に生かし、自分や所属のプレゼンスを上げ、アカデミック版『マネーボール』の再現よろしくライバルを越える活動を行ってみるのが良いのではないかと思います。

それでは、これでパネルディスカッションはクローズさせていただきます。演者の皆さん、どうもありがとうございました。