

国立情報学研究所 所長／東京大学生産技術研究所 喜連川 優 教授インタビュー： シリーズ「オープンサイエンスの展望」の開始によせて

聞き手：科学技術動向研究センター 上席研究官 林 和弘

科学技術イノベーション政策において、イノベーションを生み出す仕組み、環境作りは重要なテーマであり、近年、研究成果の活用・再利用によるイノベーション創出を加速する情報基盤作りとして、オープンアクセス、オープンサイエンスに注目が集まっている。特にオープンサイエンスに関しては、公的資金を得た研究成果をよりオープンにして利活用を促進し、新しい科学技術や産業の可能性を模索するという文脈で、内閣府のオープンサイエンスに関する検討会の報告書が今年3月に公開された。それを受けて第5期科学技術基本計画の答申素案、並びに文部科学省の第8期学術情報委員会の中間まとめが公開された。あるいは、平成27年版科学技術白書においてもオープンサイエンスの可能性が紹介されており、日本学術会議でも検討が行われている。当研究所ではSTI Horizon誌の前身の科学技術動向誌において、この動きを先導するシリーズレポートを公開し、内閣府の報告書にも引用されている。

一方、オープンサイエンスをめぐる状況はいまだ不明確な点が多く、研究者を中心とした関係者の認識もまだまだ不足しており、一部を除いて具体的な施策につなげ、予算化に結びつけることが難しい状況である。この状況を受け、STI Horizon誌では引き続き「オープンサイエンスの展望」と題した一連のレポートを企画し発行することにした。そこで今回、シリーズ開始に当たって国立情報学研究所所長であり、内閣府のオープンサイエンスフォローアップ検討会（副座長）、オープンサイエンスをメインの議題に取り上げた第8期学術情報委員会（委員）、そして、日本学術会議のオープンサイエンスの取組に関する検討委員会（委員）の全てに関わっている唯一の人物である喜連川優先生に、オープンサイエンスの展望について、インタビューをさせていただいた。

オープンサイエンスが実現する研究の地平線

まず、オープンサイエンスが果たす役割を簡単に考えてみますと、少なくとも論文になった研究については、それに至るまでのデータをオープンにしてシェアすることにより、1. 成果を多くの人がいち早く利用できるため研究開発が加速され、2. 同様に「同じ轍を踏む」「車輪の再発明」といった手戻りを防ぐことができ、3. 研究者自身にとっては、第三者の力も借りながら誤りを効率よく見つけることができるとともに研究の正当性が担保される、といった点があげられると思います。つまり1. R&Dやイノベーションの加速、2. 冗長性の排除、3. 研究者保護、の3点です。

オープンサイエンスの現状： オープン（リサーチ）データの日の出

オープンサイエンスの理念が達成された暁には、これらのメリットを最大限に享受できるようになるわけですが、現在はようやくこれらの取組が緒に就いたところで、まずはオープンサイエンスの第一歩として「オープン（リサーチ）データ」の取組が進展しつつある、といったところでしょうか。

しかしながら、それすら簡単に進んでいるわけではありません。例えば人間が関わることですから理念以前に感情の問題があります。サイエンスというものに真摯に取り組んでいる研究者であればあるほど、自身の研究やそのデータというのは、真剣に大切に育ててきた宝物です。その宝物を自分がよく理解していない、世間でも評価が定まっていないような枠組みにのせて、そして誰でも見られるような場所に置く、というのは感情的にまだしっくりこないところがあります。もちろん、相手は研究者・科学者ですし、研究遂行には少なからず公的資金が投入されていることでもありますので、前述したようなオー

オープンサイエンスの理念やその枠組み、メリット・デメリットを理解いただければ、協力を取り付けることは比較的容易です。だからこそ、きちんと内容を伝えていくことや、オープンな議論を積み重ねて細部を詰めていくことについての、より一層の努力が求められていると考えます。

オープンサイエンスに向けたオープン（リサーチ）データの課題

さて、ここで翻って「オープンサイエンス」を目指して進んでいる「オープン（リサーチ）データ」について考えてみましょう。オープンサイエンスやオープン（リサーチ）データの理念自体は良いと思うのですが、具体的・実務的な枠組みについてはまだ十分に議論が尽くされていない部分もあげられます。

基本的な姿勢としてデータを出すこと、オープンにすることはもちろん正しい方向だと思いますが、その先をどうするのか、どうすべきなのか、そろそろきちんと整理しておくべき時期に来ているのではないのでしょうか。例えば情報系の分野では「データそのもの」もさることながら、データを生成したプログラムも重要である場合があり、プログラムがセットになっていなければデータの価値が損なわれてしまう、といったこともありえます。こうなると単にデータだけを考えていたのでは間に合いません。

現状では、まずは姿勢を養うという意味もあってか「とにかく最終的なデータだけでも」といった形で進んでおり、「データジャーナル」のようなものも出てきていますが、“データ（だけ）を記録しましょう、公開しましょう”という方向に強く偏りすぎてしまうことについては危惧を抱いています。

上述したデータの解釈、変換を司るプログラムの関係性などを勘案すると、今後方向性を変えていく必要があるのではないのでしょうか。つまり“研究の何をどこまでオープンにすると再現可能になるのか”といった、より本質的な議論が必要ではないのでしょうか。



喜連川 優 国立情報学研究所 所長／
東京大学生産技術研究所 教授

オープンサイエンスが分野間の対話を促す

この議論は研究分野間の違いや、各分野の在るべき方向性を浮き彫りするため、研究の本質を見つめ直して今後の加速を促す重要なきっかけにもなるかもしれません。データは基本的に使い捨てでデータそのものの価値は低いという分野もあるでしょうし、データのリネージ（血統）を丁寧に管理して一緒に提供しないと意味が見えてこない分野というものもあるでしょう。研究分野固有の特性が明らかになってきますと、今後の研究の在り方を大きく変える動きというものも出てくるかもしれません。

例えば人文社会系では、非常に細かいたくさんのデータを丁寧に収集・分析して研究を推進されると聞きますが、一方でそういったデータは紙などの形で管理されており、その上で一度、論文なり形として成果にまとめてしまった後は、それらのデータは散逸してしまうことが多いというお話も耳にするとところ。このようなデータの散逸は研究の構造が見えれば、現状のITでもかなり支援できる場所がありそうです。

ともあれ、「データ」といっても分野によって感覚が随分違うことは明らかです。データを取ること、ためることのメリット、これによってその分野ではどのような研究の加速が見込めるのか、ということを丁寧に議論していきたいと思います。

オープンサイエンスの功罪

ここまで、基本的に研究を拡大する方向性のお話をしてきましたが、やはりオーバーヘッドの議論も避けて通れません。科学技術が一般社会への接近を深めている一方で、近年様々な研究不正が生じていることも事実であり、研究者保護や説明責任の観点からオープン（リサーチ）データの更に手前で「研究ノート」の厳密な記録・管理も求められているところ。これ自体は推進すべきですが、「オープン（リサーチ）データ」と結びついて永続性のある空間に置くということになってくると、これも議論が必要な箇所と思われる。

オープンサイエンスは一義的には研究開発やイノベーションを加速します。一方で、研究者にとって“第三者が見ても分かるような形で”整理してデータを入れたり、補足を書き加えたりするという作業は極めて大きな負担です。これによって本来の研究自体が大きく減衰するようなことが起きるのであれば、本末転倒と言わざるを得ません。情報系の研究者はプログラミングを行う際に「バージョン管理システム」などを利用して、コードを変更するたびにコメント付きで記録をつけることも多く、工夫をす

ることで負荷が軽減できることも予見されますが、多様な研究分野でどのように実現していくかは大きな課題だと思います。とにもかくにも「研究者の負担を最小化する」にはどうすればよいかを真剣に考える必要があります。

記録に関わるコストとベネフィット、ここはしっかりと議論が必要です。「研究ノート」など研究者保護や説明責任の議論は行政主体で行われることも多いですが、研究関連のデータまで「とにかく取りましよう、記録しましよう」となると、その手間が心配です。研究者主導で現場感を持って「こういった種類の研究であれば、ここまでは記録をするべき」という最低ラインを早急にまとめることが重要となるでしょう。米国もこの手間を最小限とする点について丁寧に取り扱っていると感じますが、我が国も先端のITを駆使してとにもかくにもオーバーヘッドを最小限にすべきです。これは学術ではありませんが、SOX法（米国で2002年に制定された会計不祥事を防ぐための法規）導入時にも大きな負担が問題になりました。留意することが必須だと感じます。

オープンサイエンスがもたらす研究評価の変化

何をどこまで記録するか、という話題に関連して評価、特に研究評価という話題も避けて通ることができません。オープンサイエンスは研究の評価も大きく変えていく力を有しています。

現状、論文の世界の中で質がどれだけ担保されているでしょうか。また、その質を責任を持って保証してくれる人や機関はあるのでしょうか。現状の論文評価の仕組みはピアレビュー（査読）という専門家同士の相互確認が基盤となっていますが、個別の論文についてレビュー担当者はせいぜい数人までですから結局極少数人でしか確認ができていない、インコンプリートな世界です。そういう状況ですので、とにかく何かに依拠したいという一心で雑誌評価用基準のインパクトファクターを研究評価に使おうという気持ちは分かりますが、限界もあります。

オープンサイエンスはこういった問題にも一石を投じ、大きく変えていく可能性があります。もちろんピアレビューは民主主義と同様に、問題のある不完全な仕組みながらも決定的な代替案もなく、これまで十分に機能してきた制度ですから、廃止して切り替えるという話ではなく、多面的な軸の一つとして考えるのは妥当でしょう。質と評価は表裏一体で、それを規定する価値観というものをどう集約し、表現していくかというエンドレスな世界でもありますので、多様な軸が必要であり、データの価値は新しい価値基準を与える可能性があります。データが論

文よりもっとたくさん参照（利用）される可能性もあり、ワクワクします。

オープン（リサーチ）データを促進するためのエコシステムの醸成

話が戻りますが、記録に関わるコストとベネフィットの議論などの大枠が定まると今度は質を担保するためにもデータや論文をとにかく気楽に出せるような仕組みが重要度を増してくるでしょう。出す方も気楽にデータや論文を出すことができ、見る方もこのデータやロジックはおかしいのではないかということを楽しめるような仕組みが必要です。

ITを用いれば最低限のアノマリーディテクション（書き損じや、図の使い回しなどの異常検知）はできるので、その上で間違いをバッシングするのではなく、うまくエンカレッジできる雰囲気醸成できると、「とりあえず完璧ではないかもしれないけれど、問題はあってもいいかもしれないけれど、使ってみてください。問題があれば教えてください。」という感じで、データも出しやすくなってきますし、それによって次の研究が進展する好循環が回ります。プログラミングの世界などでは「github」や「Qiita」といったサービスが正にそのような世界を実現しつつあります。

基本的に完璧なデータというのはまず存在しません。世界にあふれているデータというのは基本的にエラーを含んでいて、多くの場合、ダーティなものです。それでも、今あるものを使うだけでも大きな利益が出てくることもありますし、出してくれば誰かがエラーを修正するチャンスも出てきます。データ流通を促進する機械的・心理的な仕組みは鍵になります。

いつまで 誰が永続保存の費用負担をするか？

データをどう集めてくるか、集まったデータをどのように活用するか、これらと並んで、データやプログラムをいつまで、どうやって保存するのか、という課題も困難な壁の一つとして存在します。

米国では最近Open Accessに変えてPublic Accessという言葉を使うようにしていると聞き及んでいます。その背景にはOpen Accessという言葉が発行主負担を想起させるため、それを避けたいという動機があるそうです。さてこのオープンアクセスに関わる負担ですが、これは原則、それぞれの研究コミュニティが判断せざるを得ないでしょう。これについて米国のNSFでは、研究申請書に「データをどう保管するかの計画を書かせる」ことにしているそうです。これは研究者からデータ保管に関する良いアイデアを募るといった行為とも捉えられ、我が国でも見習うべきところがありそうです。また、研究費がある

期間は負担するとして、その予算が切れたとき、価値があるものは皆で上手に負担し、価値がないと思われるデータは思い切って棄却する、ということも考えざるを得ません。この取捨選択の意思決定は当該研究コミュニティの判断に頼らざるを得ないでしょう。

また、前述した分野間でのデータの価値の違いとも関連しますが、生み出す研究データの量についても分野によって大きな差があります。高エネルギー物理、天文、ゲノムなどのビッグサイエンス分野のデータ量はかなり膨大なもので、例えば一般的なPCのハードディスク容量などでは全く歯が立たない膨大な量になります。このような違いを考えると分野ごとに別々にストレージ空間を調達することも考えられますが、よりコストパフォーマンスを上げるために、国家規模でオープンサイエンスリポジトリを用意するという大胆な戦略もあり得るでしょう。そうすると規模のもたらすメリットによって圧倒的なコスト削減が可能となり、しかも人文社会などの比較的小容量のデータ保持コストは実質的にカバーできるかもしれません。研究成果につながる実物の保存は別として、デジタルなデータの保存には、先進的なITの利活用が国益上も極めて重要になると感じています。

今後に向けて

何にせよ、オープンサイエンスもそうですが、良

い研究や優れた仕事は短期で評価できるものではありません。私どもの分野では、10年前の論文を今評価しようという取組がなされています。その年のベストペーパーが再び選ばれることも有り得ますが、通常は別の論文が選ばれます。トップカンファレンスに採録された論文であっても、10年後に真にインパクトを与える論文はわずかでありません。そもそも、そういったものは後になって分かってくるもので、事前の評価は困難ですし、引用されないようなものもたくさんあります。ノーベル賞を取られた先生方の研究も、即物的なはやりものに乗ったものではなく、遠い将来を見通した長期的な研究に基づいていることは明らかです。

オープンサイエンスもバズワードとして即物的に取り組むのではなく、長期的視野にたって、10年後納税者の方に国費を投じたかいがあった、我が国の科学技術・学術推進に効果があったと言っていただけのような、本質的な活動として、着実に進展していくことを望みますし、そのためにも本質を捉え、研究活動の地平を見据えたオープンサイエンスの議論を進めていただきたく思います。

STI Horizon 誌は次号よりオープンサイエンスに関連した新しい科学技術やイノベーションの兆しとなるトピックについてレポートする。

参考情報

- 1) 国際的動向を踏まえたオープンサイエンスに関する検討会報告書：我が国におけるオープンサイエンス推進のあり方について ～サイエンスの新たな飛躍の時代の幕開け～：<http://www8.cao.go.jp/cstp/sonota/openscience/index.html>
- 2) 「第5期科学技術基本計画」 答申素案：<http://www8.cao.go.jp/cstp/pubcomme/kihon5soan/m001.pdf>
- 3) 第8期学術情報委員会 学術情報のオープン化の推進について（中間まとめ）：
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1362564.htm
- 4) 平成27年版科学技術白書：http://www.mext.go.jp/b_menu/hakusho/html/hpaa201501/detail/1358751.htm
- 5) 日本学術会議オープンサイエンスの取組に関する検討委員会：
<http://www.scj.go.jp/ja/member/iinkai/openscience/openscience.html>
- 6) 村山 泰啓, 林 和弘, オープンサイエンスをめぐる新しい潮流（その1）科学技術・学術情報共有の枠組みの国際動向と研究のオープンデータ. 科学技術動向, 2014, 146, p.12-17 : <http://hdl.handle.net/11035/2972>
- 7) 村山 泰啓, 林 和弘, オープンサイエンスをめぐる新しい潮流(その2)オープンデータのためのデータ保存・管理体制. 科学技術動向, 2014, 147, p.16-22 : <http://hdl.handle.net/11035/2990>
- 8) 林 和弘, 村山 泰啓, オープンサイエンスをめぐる新しい潮流（その3）研究データ出版の動向と論文の根拠データの公開促進に向けて. 科学技術動向, 2015, 148, p.4-9 : <http://hdl.handle.net/11035/2999>
- 9) 野村 稔, オープンサイエンスをめぐる新しい潮流（その4）研究コミュニティに向けた協働データインフラの開発動向—欧州のEUDATの取組から—, 科学技術動向, 2015, 149, p.11-18 : <http://hdl.handle.net/11035/3039>
- 10) 林 和弘, オープンサイエンスをめぐる新しい潮流（その5）オープンな情報流通が促進するシチズンサイエンス（市民科学）の可能性. 科学技術動向, 2015, 150, p.21-25 : <http://hdl.handle.net/11035/3046>