第 7 回 SPARC Japan セミナー2010 「著者 ID の動向」

ディスカッション

司　　会：　**林 和弘**　（日本化学会 学術情報部課長）

パネリスト：　**谷藤 幹子**　（物質・材料研究機構 科学情報室 室長）

武田 英明　（国立情報学研究所教授／学術コンテンツサービス研究所開発センター長／ ORCID Board Member ）

蔵川 圭　（国立情報学研究所 学術コンテンツサービス研究開発センター特任准教授／ ORCID Technical Working Group Member ）

●**林**　ディスカッションに入りたいと思います。先ほどの続きで、そもそも押さえなければいけない著者の数について、トムソン・ロイター、もしくは Scopus であれば、毎年どのぐらいの論文数を収録しているかが分かると思います。そこから大体の人数が分かるのですが、堀切さんは Web of Science に毎年収録される論文数もしくは著者数のデータをお持ちですか。

●**堀切**　トムソン・ロイターの堀切です。著者数についてはしばしばご質問をいただくのですが、正直に申し上げますと、実はわれわれも正確な数字は分からないのです。それは最終的には本人確認をしなければ分かりません。技術的に例えば引用や共著者関係で著者と業績の同定を行って人数を特定するような研究は出されてはいるのですが、世界的に人数を把握しているという報告については、私自身は存じておりません。

●**林**　重複があってもいいので、とにかくこれ以上はないという上限値、延べ数はお持ちではないですか。そこから重複などを除いていくと絞られてくるので、マックスのけたが例えば何百万なのか、1000 万のオーダーなのか、それぐらいは類推で分かりませんか。5000 誌採録していて、1 誌が 100 論文出すと 50 万論文ということになりますが。

●**谷藤**　科学技術政策研究所が 12 月に出したサイエンスマップによると、世界で生産される論文数は大体 100 万ですね。

●**林**　そうすると、平均の著者人数が 5 人ならば、延べ数は 500 万人ということになりますか。High Energy Physics のような 2000 人は特殊な例として、1 論文当たり 10 人いくこともないとすれば、規模感的には 500 万～1000 万人ぐらいでしょうか。同姓同名による重複と、同じ人がまた次の年にも書くといった経時による重複があり、そのぐらいの著者をどう同定するかという話になりますね。無理やり基礎スペックの情報をどのようにつかむかというディスカッションをさせていただいたのですが。

●**谷藤**　それに色を付けるとしたら、論文を出している間にもポスドクの期間が終わって次に行くといったことがありますので、その 100 万という論文に付随する著者の数が変動します。そこを追跡することも必要ですね。そう考えていくと、延べの数は大層な数になりそうだと思って先ほど質問したのですが、その著者と出版雑誌とのコミュニケーションで、「ちゃんと確認してね」「違ったら直してね」「移ったら加えておいてね」という関係が成立するのかどうか。研究者総覧と

いうすごくミニマムな世界での職員と同僚という関係の経験からすると、それはとても想像ができないくらい大変なことのように思えます。かつ、コンピュータに近い分野だけでなく、違う分野の人もいることを考えれば、ちょっと難しいのではないでしょうか。税金ではありませんが、投稿するときに「ID を入れなければ投稿できません」とするのであれば別ですが、学会も含め、そういう方向に行くのですかというのが私の質問でした。

●**林**　逆に、物質・材料研究機構（NIMS）の中にもいろいろな分野の方がいらっしゃるということでしたが、明らかに分野による違いはあるという理解でよろしいでしょうか。

●**谷藤**　はい、もう学会の名前で言えるくらい違いがあります。典型的な振る舞いの違いとしては、EndNote のような文献管理システムに自分や気になっている人の論文リストを書き出しておいて、次に自分が論文を書くときの引用文献にさらっと入れるというのは比較的、化学や生物工学に近い人たちですし、「そういうことなんかは、なぜ TeX でやらないんだ」というのは、少し物理あるいは計算科学に寄っている人たちです。それから、例えば金属物理など、そもそもそういうこと自体をあまり必要としない分野の方もおられます。それだけ考えても、NIMS の正規の研究職員 500 人に、ORCID を「全員やりましょう」と声をかけたとして、実際にしてくれるのは 1%ぐらいではないでしょうか。何か別のインセンティブと結び付いていれば、マインドはもう少し違った動き方をするだろうと思います。

●**林**　その議論を拡張すると、物質という、テーマがある程度決まっている研究機関ですらこの状況であるときに、大学等で ORCID を進めていくとなると、さらに広範囲な分野の研究者を相手にすることになります。機関リポジトリで既に図書館の方々が相当苦労されていると伺っていますが、その辺で、図書館の方からコメントや、ORCID に対する課題、期待など、ありませんでしょうか。

●**杉山**　静岡大学の杉山です。図書館代表という視点とはちょっとずれてしまうのですが、機関リポジトリにしても多言語の資料を扱っており、英語論文や日本語論文が混在しています。しかし、蔵川先生の ORCID のユーザー入力データ項目などを見ても、これだけでは日本語には対応しないのではないかと思われます。特に著者の表記については、日本語の場合、英語で書いたり、漢字で書いたり、あるいは読みも可能性としてあるのかと思うのですが、その解決はどうなっているのでしょうか。

●**蔵川**　ワーキングのメンバーからは、例えばウェイ・ワンという中国人の名前には 10 とおりぐらいの表記の仕方があるのですが、読みとしては全部ウェイ・ワンになってしまって困るという話がよく出てきます。そういうことにも対応するという意思の表れとしての例なのだと思います。ORCID の Technical working group の議論では、基本的には著者は ID で表現する。名前は、UTF-8 で表現できれば何でもいいのですが、プライマリーの名前と、もう一つ other name という項目があり、ペンネームを含めてすべての別表記はそこに書く。そして ORCID の ID がいろいろな名前のバリエーションを持っていることを示すために、それをリストとして表示するということになっています。読みについてはまだ触れていません。ただ、日本語の名前は表記と読みをセットにしたいというところもあるでしょうから、それについては考慮する必要があるとは思います。

●**林**　付け加えると、日本の研究者が世界に向けてプロファイルを出したいときと日本に向けてプロファイルを出したいときとで、ユーザーインターフェース的に使い分けさせるような理論はまだないのでしょうか。

今のは ID にひも付けられる一意の名前とサブの名前という概念ですが、そうではなくて、エンドユーザー側から見ると、例えば文科省に業績を出すときに英語のものを出しても嫌がられると思うのです。そういった観点での議論はまだされていないのでしょうか。

●**蔵川**　まだありません。これから出てくると思います。

●**林**　それは多分、スキン的に別な議論として行われて、どう対応させていくかという、ローカライズの観点ですね。

●**蔵川**　むしろそれよりも解決しなければならない問題が山のようにあります。

●**林**　ほかに何かありますか。

●**杉田**　国立情報学研究所（NII）の杉田です。3 人の方、今日はどうもありがとうございました。谷藤さんのお話の中で、機関、図書館、研究者により近い立場から、ORCID に期待されるものとして、例えば投稿査読システムで著者 ID を使えたらいいという具体的なお話がありました。ほかにはどのような具体的サービスに使われることを想定して ORCID は設計されつつあるのか教えていただきたいというのが一つです。

また、ORCID のサービスは無料のものも有料のものもあり得るというお話がありましたが、メンバーシップフィーを支払うインセンティブはどの辺にあるのか、持続性ということも含めて蔵川先生と武田先生にそれぞれ伺いたいと思います。

●**武田**　それに答える前に、今日は私や蔵川先生からいろいろな情報を出してしまったので、逆に混乱するかもしれないのですが、実は ORCID の目的は研究者に ID を振るというところに尽きるのです。それをどう実現するか、どうメンテナンスするか、どう生かすかというのが今日の話です。それが漢字で表記されるのか、アルファベットで表示されるのかというのは、まず ID があっての議論です。谷藤さんが言うように、今まであった名前のように曖昧なものよりも、もっと一意に同定できるもの、サイエンスコミュニケーションで名前と一緒に使われるもの、それが ID です。名前と一緒に使われるのだから直接名前との関係がなくてもいいのです。その ID を振るということが、一番根本のやりたいことです。

そこまでは全員が合意するところですが、この ID の中身を知りたかったらお金を払えというのでは ID としては機能しないのですから、当然オープンでなければ価値がないわけです。まずは、とにかく研究者に番号を振るということです。

ただ、そこから先は、それをどう利用するかを利害関係者がいろいろ考えるわけです。出版社は出版社で、自社のサービスとどうひも付けたいのか、あるいはどうすれば楽をしてそれを作れるかということになって、出版社の情報をそのまま使いたいなどという話になるわけです。

日本の場合は科研費番号が既にあって、世界にはありません。基本的にはあれが欲しいわけです。ただ、世界政府があって割り振ってくれるわけではないので、現状はボランティアベースですが、それを作ってしまおう、そのために利害関係者がたくさん集まったというのが ORCID の現状だと思います。今、ID を登録する話と一緒に議論してしまっているので訳が分からなくなっていますが、逆に、例えば ID があったら何ができますか、自分の組織の人が全員 ID を持っていたら何ができますかというように考えるべきではないか。これが、今の質問に対する逆の答え方です。つまり、ORCID のサービスは、根本的には ID がユニークであって、ある基本情報と結び付いていることを保証するだけです。そういうものがあると今までと違うどんなことができて、それをしやすくするためには ORCID に何が必要なのか、どんな付加的なサービスを出してくれたらうれしいのかと各組織で考えた方が、問題が

整理しやすいかと思います。これが基本の理念的なことで、会費を払うモデルがうまくいかなくなったというのはその辺があります。

ただ、お金の問題に関しては、機関ごとの入力・出力のバルクサービスあるいは評価などは各機関・大学もお金をかけていいと思うだろうからお金を取る余地があるのではないか、要するに金に見合うサービスをしたいというよりは、そこぐらいしかお金が取れないからそういうことを考えているというのが ORCID 側の立場だと思います。

機関ごとにきちんと整備されたデータを作ると、ORCID 自身の付加価値も高まるし、利用する各機関にとっても付加価値が高まります。それをもう少し詰めて、そういうサービスをどうするかを考えていくのだろうと私は思っています。答えになったでしょうか。

●**林**　若干補足すると、今はまだ黎明期の段階で情報を得られているので、大げさなことを言うと、逆にわれわれが欧米に先んじてサービスを考えることすらできるようなタイミングではあります。今回非常に感慨深いのは、CrossRef のときは、その元である PILA（Publishers International Linking Association）の立ち上げに日本の関係者が誰一人おらず、最初のプロジェクトの段階で日本の貢献はなかったのですが、ORCID の場合は、この早い段階でもう 2 人の先生が入っていらっしゃって、世界のイニシアチブの黎明期の段階でこういう情報を得る機会が得られている状況にあることです。それは裏を返すと、世界の動向を自分たちで真剣に考えなければいけないということです。それがイニシアチブを取る人たちの義務でもあるので、むしろ積極的に何ができるのかというディスカッションをしていかなくてはいけないと思います。ここでそれをいきなりやれといっても無理なのは分かっていますので、帰ってから、近くの人としてもいいでしょうし、あるいはもう一回こういう場を設けてもいいと思います。確かに今回はまだ、こうもある、こうもあるという話なので、もう少ししてから実際にこういうものができるという議論をしてみてもいいでしょう。でも、谷藤さんのお話などは、もう著者 ID を使っているわけですから、実例ですね。そういったものをどんどん拡張していくということになるのではないでしょうか。

●**武田**　もちろんできることとできないことがあると思うのですが、ここで、むしろこういうことがしたいという話を大いにしていただいて、運営の方針についてもこうしてほしいということがあれば、私もボードメンバーで言うチャンスはありますので、言ってください。ちなみに、もともと出版社ベースで動いていたものが大学ベースに舵が振れたのは、ボードメンバーの CERN の Salvatore Mele がコミュニティーに貢献しろと強く言い続けた結果なのです。まだオープンな状態なので、この場でもいいですし、後でもいいので、ぜひ大いに意見を言ってください。

●**蔵川**　ORCID での議論や、そのメンバーのこれまでの活動が論文になっていたり、ブログに書かれたりしています。そういうものを見ていると、小さい各機関や論文サービス、データベースを含めて、著者をアイデンティファイすることによってどんなサービスができるかは、基本的には既に見えているし、やってきたのです。けれども、その閉じられた世界の中では不十分で、精度を上げるためには機械処理では無理だという経験が、この 10～20 年の中であるのですね。で

は何が必要かというと、やはり手で各研究者に正確に、重ならないように ID を付けたデータベースを世界に一つ作ることです。それにはどうすればいいのかということを考えているような気がします。

そのためには、その ID を表すデータベースは一つの組織に依存する形ではなく、全体の中で一つのインディペンデントな形であるべきだし、それぞれの既存のサービスや構想の中での触媒となるような位置付けに ORCID があればいいと考えているように見受けられます。そのためには、このコミュニティー全員が参加して、全員で一つの ID データベースを作るのだというムーブメントが今ここにあるような感じがします。だから、ここにいる皆さん、関係者が全員参加することがゴールだと思います。

●**杉田**　ORCID に参加する方法にも幾つかあると思うのでお聞きするのですが、例えば金沢大学は直接 ORCID のメンバーとして入っていますね。グローバルな枠組みができたときに各機関レベルで手を挙げて入るのはすごくいいと思うのです。ただ、個々の大学が全部そうできるかというと、難しいと思います。日本では NII や科学技術振興機構（JST）が入っているので、例えば NII の研究者リゾルバーID をリポジトリに入れておけば、それを介していずれは ORCID にもつながっていけるというイメージを持てるといいのではないかと思います。研究者リゾルバーID と ORCID は、今後どのような関係になっていくでしょうか。また、身近な例として、例えばリポジトリの担当者が今すぐできることはあるでしょうか。

●**武田**　アメリカだと、国立衛生学研究所（NIH）と ORCID は少し相談をしているようです。NIH は非常に大きく、ファンディングエージェンシーでもあり、PubMed の運営機関でもあるわけです。ファンディングエージェンシーも ID を管理しているようですが、それと PubMed の ID はまた別らしい。今 ORCID はそういうところと相談しているようです。ファンディングエージェンシーやナショナルワイドの組織がどう関与するかに関しては、これからいい関係を探していこうというのが全体の流れです。

では、日本の状況においてどうなのかという問題は確かにおっしゃるとおりで、例えば四大学術出版社のジャーナルに論文を投稿するときは必ず ORCID の ID を付けるというモデルがあれば、メジャーな英文雑誌に載せる人にはほぼ自動的に ORCID の ID が付くでしょう。これは割と考えられる未来なわけです。ただ、困るのは、主に日本語で活動している人たちはそういうことに触れるチャンスがないので、ORCID ID を取るチャンスもなくなってしまう。気付くことすらないかもしれません。今までほとんど英語圏の人たちの議論だったので、そんなことはあまり問題にしていなかったのですが、われわれのように固有の言語の学会やジャーナルを持っている場合は少し事情が違います。単純にジャーナルに投稿するときに ID を付ければ、そのうち黙っていても研究者が全員 ID を持つかというと、そうとも言えません。そのときにどうするかということは確かに問題になると思います。

では、そもそも日本語でしか書かない人が ORCID ID を持つことに意味があるのかというと、ほとんどの情報が英語で書かれていて、例えば組織の名前も英語で書かれるわけですから、そんな情報を持っていてもしょうがないという考え方もあり得ます。あるいは、翻訳されても論文が読まれる時代なので、もうこの世の中では英語で名前を登録することに価値があると考えるのもいいでしょう。私は後者の方が日本の研究者コミュニティーにとってハッピーだとは思っています。自分は日本語でしか書かないから決して英語の登録は要らないと言う必要はないと思います。そのときに日本の機関が助けられるかということに関しては、確かにもっとバルクで入れてしまえば早いという議論はあり得ます。ただ、日本語でしか活動しないなら英語での登録は要らないという人が多数を占めるのであれば、そういうことには多分ご賛同いただけません。逆に、今や世界は狭いのだから日本語で書く人でも英語での

登録があっていいと思う人が多数を占めるのだったら、やる価値があるということになるでしょう。その辺は、これから日本の研究コミュニティーが自問自答しなければいけないのかなという気はします。それは誰が働きかけてもいいですし、われわれが働きかけてもいいのですが、その辺のコンセンサスを得なければいけないでしょう。

今まではインターナショナルに活動するトップの研究者を念頭に置いた議論が多かったわけですが、もう少しすそ野を広げると日本の場合はそういう問題が入ってくるわけです。それに対してわれわれがどういう方向に志向するかが、結構決定的に効いてくるのではないでしょうか。私としては閉じないでオープンにしていってほしいと思います。ただ、それは私の意見であって、日本の研究者コミュニティーの総意ではないかもしれません。そこがむしろ仕組みを始める前の前提条件としてあるかと思います。

●**蔵川**　武田先生はどういったコミュニティーを作るかということを主眼に話されたと思います。どのようにシステムが作られてきたのかを参考にすると分かりやすいかもしれないのですが、例えば、ID を付ける論文システムは既にいろいろあり、arXiv.org も Author Identifier を自前でログイン ID と同じようにする形で URI として ID を付けています。arXiv の Author identifier を作ったときの議論の論文を読んでいると、arXiv の ID は、別の ID との関係を、そこのログインプロセスの中でプロファイルを作ることによって外の世界とつながるということを考えています。

一方、Virtual International Authority File (VIAF) なども各図書館が典拠の ID を持っていて、同じ人物を示すならば VIAF の中の一つの ID を代表して付けることで、それぞれの典拠の ID 同士が同じ人だという ID 体系を作ろうとしています。さらに VIAF は、例えば ISNI の ID とこの人は一緒だということも言うわけです。また、ORCID でも全くそういうことを考えられていないわけではなく、実は ISNI や VIAF とくっついた方がいいのではないかというシステム的なムーブメントはあります。Web というインターネットの中では、データベースを受け持つ係やシステムを担当している人たちがそれぞれ ID を作って、結果としてそれぞれを結ぶことによって一つの ID 体系の整合性を付けようということをしているのです。

一番初めの質問に戻ってしまいますが、研究者リゾルバーや ORCID と何が関係するのかと考えると、自分たちが著者に ID を付けることによって新しいサービスを生み出したければ、まず自分たちで付けることを考えるでしょう。他人が ID 体系を作るのを待てればいいですが、普通は待てないので、早い段階でいいサービスを得たいと思ったら、まず自分のところで作ってしまって、後でメジャーな世界とくっつく道を選ぶというアプローチになるのかなと思います。

●**子安**　日本哺乳類学会の子安です。今日は和文誌編集長の横畑先生と一緒に来ているのですが、実は 2005 年にスイスのローザンヌ大学に 1 年間留学していて、エルゼビア社が Scopus を開発するときのモデル校にいました。日本人としては最初に Scopus に接した 1 人ではないかと思うのですが、Scopus の考え方は、著者のデータベースではなく引用文献のデータベースであるところが非常に斬新だったのと、人文サイエンスのデータベース化に力を入れたという、この二つが特徴的でした。それが Web of Science に強力に対抗するエルゼビア社の戦略だったわけです。

日本に戻ってきてから、恐らく私どもは日本で最初に Scopus を導入した大学の一つだと思うのですが、そのときに入れられたのは、Web of Science が入っていなかったからです。つまり、データベースが全然入っていなかったので Scopus を入れることができたのです。それ以来ずっと私は Scopus しか知らないということがありますので、見てきて著者の ID があることも分かったのですが、エルゼビア社は自動的に論文データベースを作るので、引用文献に全部著者 ID を振り分けるわけです。そうすると、著者の数よりも引

用文献の数の方が圧倒的に多いのですから、エルゼビア社が持つ ID は著者 ID より何十倍も多くなります。それらすべての名寄せをしていかなければいけません。それをエルゼビア社はボランティアに頼みました。合わせたい ID と離したい ID を申請するようになっており、その根拠となる URL を送って、基本的にはメールのやり取りでその信頼度を図ることになっています。

私はそれを自分と、スイスのときの指導教官や日本の指導教官の分も含めてボランティアでしました。その一番の動機は、彼らは定年が近かったものですから、定年のときに文献リストを作るのにそれがあると非常に便利だというものでした。そういうインセンティブがあればしますが、たまたまそのスイスの指導教官はドイツ人でフォーゲル、日本の担当の人は織田という、どちらもやたらに多い名前だったので、2 人分をまとめるだけでも大変な努力が必要でした。何とかやってみて、結局、名寄せが一番重要であると痛感しました。その後、今の日本哺乳類学会の役員になって、今回はその立場で参加していますが、学協会からの発言やどう対応するかということが出ていなかったので、提起という形になるかもしれませんが、発言しました。

今、アーカイブ担当という役職に就いていますが、アーカイブ担当者が現在いる 1000 名の会員の ID をどう管理するか、私がボランティア的にやった名寄せを学協会がするメリットは何なのか。例えば賞の選考時に、その人の ID が分かれば過去の業績が全部分かって一覧表が簡単に出るのならば、する価値があるかもしれません。もしそれ以外にも使えそうな例があれば、どなたでも結構なのでお教えいただければと思います。

●林　ほかの学協会の方々はいかがですか。学協会として会員にどうインセンティブを与えて名寄せをしてもらうかという考え方が提示されているわけですが、何かコメントやアイデアはありますでしょうか。

思いつきなのですが、学会や図書館でも、取りあえず鈴木太郎さんや佐藤さんといったよくある名前の人からくどくというのはどうですか。それが一番インセンティブが高そうだなと今一瞬思ったので、つい口に出してしまったのですが。あるいは、谷藤さんの講演にあったように、前職の履歴を引き継ぐことが機関では難しいのですが、学会員である限り、その分野の研究をしている限りは、会員として今までどういう業績を持ってきたかということになるので、実はその問題は存在しないのですね。ですから、学会が著者 ID を積極的に管理するという考え方は、私はあると思っているのです。ところが今まさにおっしゃったように、そのインセンティブは何かというところが非常に難しいと思っています。ここでぱっと答えられる人がいたら、お二人の先生と一緒に加わっていただきたいと思うのですが、今後の議論の一つのキーにしていただければと思います。

というところで、時間になってしまいました。まだまだこれからの話ではありますが、これからだからこそ早めに考えて、一歩もしくは半歩先の手を打てるようにしておくことが大切なのではないかと思います。何か動きが決まってから動くのではなく、話が来たときにさっと対応できるようになっているといいのではないかと、無理やりまとめさせていただいて今回のセミナーを終了させていただきたいと思います。

最後に、お忙しい中、しかもなかなか不確定な中、積極的に情報を出していただいた 3 人の講師の皆さまに、もう一度拍手をお願いいたします。