

## 第 7 回 SPARC Japan セミナー2010

著者 ID の動向

### 趣旨説明

林 和弘

(日本化学会 学術情報部 課長、SPARC Japan 運営委員)



1994年東京大学大学院に大学院生として在籍時、日本化学会の論文誌査読管理データベースを開発したことがきっかけとなって、日本化学会の論文誌の電子ジャーナル化に取り組む。化学の研究者の立場がわかり、かつ IT スキルを持った氏は日本発の学術情報発信について、電子投稿査読、XML 出版、電子ジャーナルプラットフォーム構築、宣伝活動の広いフェーズで実務に基づき考察と改善を加えた。その活動は、結果として日本化学会 Chemistry Letters 誌を一般化学誌として世界最速クラスで発行する電子ジャーナルに整え、化学系学会出版としては世界でもいち早いオープンアクセス対応を開始するなど、数々の実績に反映されることとなった。また、そのノウハウは SPARC Japan、J-STAGE 等各種のプロジェクトに反映され、科学技術・学術審議会下の WG 専門委員も務めて日本の学術論文誌のあるべき姿を提言し、現在も日本発の論文誌をより魅力的にするための考察と改善を繰り返している。さらに、文部科学省科学技術政策研究所客員研究官他を兼任し、電子ジャーナルの将来とインターネット時代の科学コミュニケーションのあり方そのものについても興味を持つ。

今回のセミナーは、「著者 ID の動向」についてで、特に ORCID プロジェクトにフォーカスを当てています。まず、著者 ID の動向をトピックとして SPARC セミナーで取り上げようと思った背景について簡単にご説明します。私は科学技術政策研究所 (NISTEP) の客員研究官を仰せ付かっており、そこで ID 問題についてレポートしている関係もあるので、個人的な考えも含めて紹介しながら、ORCID プロジェクトのセミナーの導入にしたいと思います。

#### 著者 ID=人の ID

そもそも著者の ID とは結局、人の ID のことで、この問題は全く新しくも何ともない問題といえます。学校には名簿があり、学生には学籍番号があるように、名前をキーとした表やデータベースはインターネット

やデジタル化の前からありました。それが学協会などの場合は特に学会の会員名簿としてそのデータベースをどうしていくか、どのようにオープンしていくかということが、ここ 10 年ほどのトピックでした。似たような話は図書館など、データベースがあるところには存在しています。

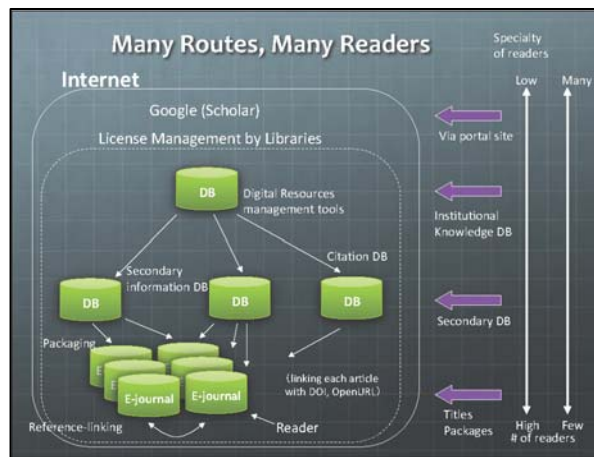
インターネットの時代になると、インターネット上の ID という話が出てきました。ショッピングサイトでは Email をベースにしたもの、BBS でのハンドルネーム、最近では Twitter のアカウントなどがあります。これまでの比較的閉じた世界の中では、人の ID をどうするかという話はその閉じた中で完結しているものとして沢山あります。しかしその連携はまだそれほど進んでいません。強いて言うと、インターネット上の ID の問題に関しては、例えば Twitter と同じア

カウントで mixi に入るといような Open ID と呼ばれるトピックがあるのですが、それは今日のテーマからずれるので、学術の著者 ID の方に話を進めたいと思います。

### 電子ジャーナル化と識別子 (Identifier)

学術論文の電子ジャーナル化で、まず論文誌の識別子の問題が出てきたわけですが、その背景には、コンピュータリソースやネットワークインフラが飛躍的に発展して、今まで閉じた世界でしかやりようのなかったものが繋げられるようになってきたということがあると思います。たかだか 15 年前には、コンピュータの CPU は数十 MHz、メモリは数 M、ハードディスクは良くても数百 M、ネットは電話回線ですらないでしたので、それでよそのデータベースと連携するなど、少なくともコスト面では全く太刀打ちできませんでした。それが、情報伝達スピードが飛躍的に高まり、かつ、多くの世界中の情報と簡単に連携することが可能になりました。これが著者 ID をどうするかという問題が浮き上がってきた大きな要因だと思えます。情報伝達スピードが飛躍的に高まり、かつ、多くの情報と連携できるということはつまり、共有しやすくなり、透明性が高まるということです。少しずつですが、オープンアクセスが進んでいるのも、インターネットインフラの透明性という利便性を利用した活動であるという見方もあります。

ただし、他データベース間で連携するためには、連携したい情報が何であるかを識別しなければならないという問題が出てきます。システム的には、リレーショナルデータベースを作ったときにキーが違ったり一意でなかったりすると混乱しますので、それをどう整えていくかという話になります。少なくとも学術情報流通の世界では、CrossRef に代表される論文をどうやって一意に特定するか、インターネット上のありかをどうするかというところから識別子の話が進んでいき、CrossRef 自身も、Contribution ID として Author Identifier に取り組んできた歴史があります。



(図 1) Many Roots, Many Readers

### Many Roots, Many Readers

電子ジャーナルの一つ一つに識別子が存在することで、その上の二次情報データベースや引用データベースとのリンク、Digital Resources management tools を使った図書館の蔵書管理といった広がりが出てきて、今はさらにその上のインターネットで、Google (Scholar) で検索してユーザーに情報が届くようになっています。ここまで広がると厳密に一意の識別子がなかったりはするのですが、いずれにせよ、今まで専門のジャーナルを読むのは図書館に通う研究者だけだったのが、どんどん広がって、検索上は Google を使ってリンクさえあれば誰にでも情報が届くようになってきている。このような状態において、その論文の所在を確かめるときに、やはりなるべく一意に識別されることが必要です。そのことが、CrossRef が発展してきた背景の一つにあります (図 1)。

### 図書館情報学的苦悩と研究評価の時代

論文が電子ジャーナル化し、特に図書館情報学での研究が進むと、デジタル化ならではの問題が顕在化してきました。誰がどの論文を書いたか、どの機関がどれだけの論文を書いたかということは、紙の時代から研究する人がいたので古くからある課題なのですが、結局、名寄せ問題 (バリエーションや間違い、言語をどうするか)、あるいは経時変化問題 (人の改姓、機関の統廃合の場合にどうするか) が出てきます。分

ないところは昔も今も、ときには力業と主観によって「えいや」と決めているわけです。こういった研究が図書館情報学では古くから行われており、Web of Science や Scopus でデータベース化するとき、彼ら自身も、名前をどう整えるか、機関をどう整えるかという点で非常に苦労していることから、特に Web of Science では Researcher ID の策定と利用を促す流れが出てきていると理解していいと思います。

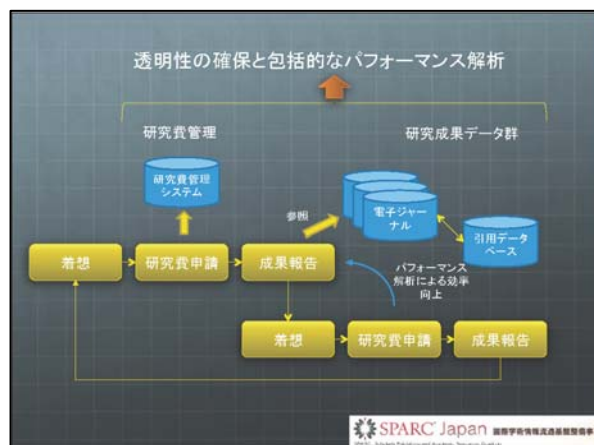
それが、研究評価の時代になると、主に引用データベースに基づく研究パフォーマンス評価が大事になってきました。誰がどの論文を書いたか、どの機関がどれだけ論文を書いたかを正確に知りたいという研究目的と、大学の研究評価に資するというニーズがマッチして、より客観的で精確なデータが必要になってきています。最近では、研究費の適正配分のために使いたいという話にもなっています。

### 透明性の確保と包括的なパフォーマンス解析

研究者は、着想した研究に対して研究費を申請し、成果報告を出し、そこからまた新しい着想を出して研究費を申請し、成果報告をするというサイクルを繰り返します。今は既に研究費申請のための研究費管理システムが存在し、成果報告は電子ジャーナルででき、そのパフォーマンスについても、少なくとも引用に関してはデータベースができています。それぞれに関して識別子はあるので、研究費の管理、研究費成果データ群と併せて、公的資金（税）で行われた研究に対するアウトプットの透明性の確保、ないしはそのアウトプットのパフォーマンス解析ができるという話に今は議論が進んでいるわけです（図2）。

### 研究費の適正配分と e-Rad

現在、既に e-Rad と呼ばれる府省共通研究管理システムがあります。これは研究資金制度間の情報共有、研究者情報の管理などを行うことにより、不合理な重複の排除や過度の集中を避けるとともに、業務の効率化を図るためのものです。研究費を配分する際の、研



（図2）透明性の確保と包括的なパフォーマンス解析

究者およびその成果報告に関しての一定の連携が日本でも取られつつあります。

e-Rad の裏側では、科学研究費補助金データベース（KAKEN）と国立情報学研究所（NII）の研究者リポーターの連携、科学技術振興機構（JST）の研究開発支援総合ディレクトリ（ReaD）との連携、あるいは経済産業省や厚生労働省などの研究費データベースとの連携もしよう、少なくとも申請段階では重複がないように見ようというところで、このスキームにおいては日本の研究者 ID に関して一定の統制が取られつつある状況です。

ですから、どの研究者がどの研究機関にいて、どこから研究費をもらっているか、その成果がどうであったかということが調べようと思えば調べられる、すぐ近いところまで既に来ているといえます（図3）。これらを包括的ないしは統制的に解析しようとしたときに、ますます識別子が重要になってくるわけです。

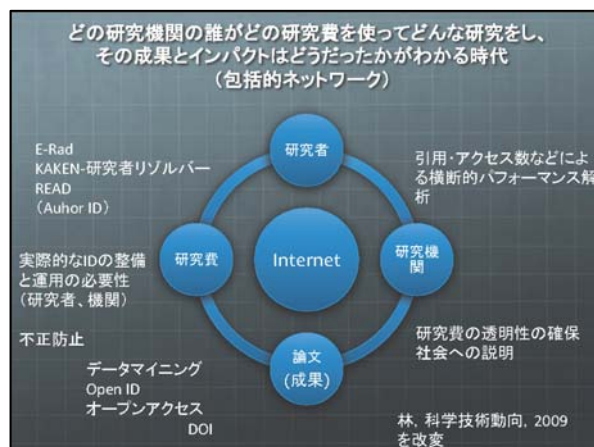
### ORCID—世界レベルの著者 ID の連携

ところが、これはもう日本だけの問題ではなくなっており、世界レベルで著者 ID の連携を取ろうとしています。詳しくは講師の方からご説明いただきますが、理想的には、名だたる大学、図書館、学会、出版、ベンダーの超党派連合を組んで世界中の研究者の情報管理ができるようになれば、日本の研究費がどうこうというだけでなく、グローバルな研究者や研究機関の評

価ができるようになるのではないのでしょうか。

ただし現実的には、Researcher ID に象徴されるように、トムソン・ロイターが既に活動していた歴史があるのですが、なぜ単独では止めてこういう大きな枠組みでするようになったのでしょうか。あるいは、CrossRef が Contributor ID を始めたのはかれこれ 4～5 年前ですが、結局 ORCID で対応するようになったのはなぜなのか、その ORCID のシステムコントロールはどうか、運用体制はどうかという話が出てきます。あるいは、世界の中で日本はどう対応すべきか、また大学、研究機関、図書館、学会、企業として、それぞれどう対応していくかを考えなければなりません。こういった問題提起をするという観点から、今、著者 ID に関してセミナーを行うことがふさわしいのではないかと思った次第です。

より詳しい ORCID の背景とシステムコントロール、各ステークホルダーの対応に関しては、武田先生、蔵川先生、谷藤さんにそれぞれご説明していただきます。



(図3) 研究費の適正配分と e-Rad