

第 7 回 SPARC Japan セミナー2010

著者 ID の動向

ORCID とは何か

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)

講演要旨

ORCID (Open Researcher & Contributor ID) は今年から動き出した学術論文の著者 ID を統合的に管理しようという試みである。Elsevier 等の主要な学術出版社、Thomson Reuters、いくつかの大学図書館、CrossRef、OCLC などが主体となり、米国 NPO として発足した。本稿では、ORCID の目標や組織構成、現在の活動状況などの概要について述べる。



武田 英明

経歴	2006 年～2007 年	国立情報学研究所 教授
	2005 年	国立情報学研究所 教授 実証研究センター
	2004 年	情報 国立情報学研究所 システム研究機構 国立情報学研究所 実証研究センター教授
	2003 年	国立情報学研究所 助教授 実証研究センター 教授
	2001 年～2002 年	国立情報学研究所 知能システム研究系 助教授
受賞	2009 年 9 月	財団法人船井情報科学振興財団 FIT2009 船井ベストペーパー賞
	2007 年 10 月	IADIS Best Paper Award, IADIS International Conference WWW/Internet 2007, Vila Real, Portugal
	2007 年 6 月	人工知能学会 人工知能学会 功労賞

今日は私と蔵川先生で ORCID (Open Researcher & contributor ID) の話をします。私からは、これまでの経緯や組織、そもそもなぜ ORCID なのかということについてお話しし、技術的な内容については蔵川先生から詳しく説明します。ORCID は既にサイトがあり、<http://orcid.org> を見ていただくと今日私が話す内容は結構載っていますが、今から順に説明していきます。

世界の論文、国際共著論文、 主要国の論文数の変化

まず、なぜ今なのかという話が出ました。科学技術政策研究所の報告書に載っているグラフで世界の論文数の変化を見ると、2000 年あたりから非常にカーブが急になっていることが分かります (図 1)。これは多分、皆さんも何となく認識されていたと思います。また、国をまたがった論文が急増していることも分かるでしょう。つまり、今まさに研究自体がグローバルになっており、著者数も劇的に増えていることが容易に想像

できると思います。論文数がこれだけ増え、しかも今までは、ある国の論文はある国の著者が書いていたのが、複数の国をまたいだ国際研究となると、ますます相手が誰か分からなくなります。そういう時代が来ているということです。

ちなみに、私の世界ではあまりないのですが、高エネルギーの研究者で ORCID 理事の Salvatore Mele が挙げた例では、ある一つの論文に著者が約 2000 人、組織も 200 ほどかかわっています。高エネルギーでは国際共同研究が多いので、このような現実があるわけです。こういう時代になると、確かに著者問題が大変であることが分かります。

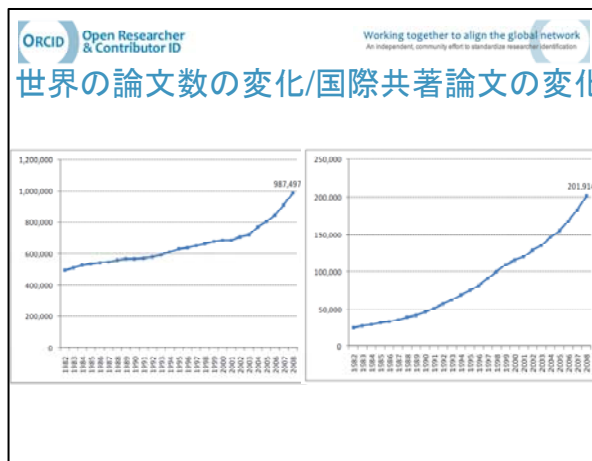
国別の論文数の変化を見ると、中国が急激に増えています (図 2)。この辺が論文数の増えている理由でもあります。これについては次回の SPARC セミナーで詳しい解説があると思います。とにかく、こういう状況の中で ORCID が作られました。

ORCID Inc.について

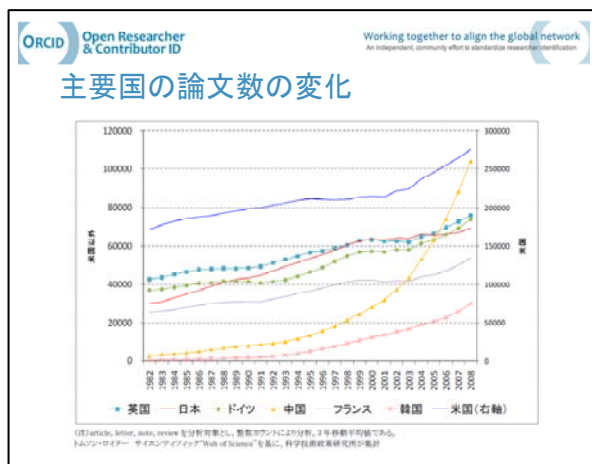
ORCID は、「学術研究における名寄せ問題 (name ambiguity problem) 解決のために設立された非営利機関であり、世界中の影響のある大学、助成機関、学会、出版社および企業と共同で運営されるものである。理想的な解決はコミュニティー全体でデファクト標準として採用・受容されるレジストリを構築することである。システムティックな名寄せ問題解決としては、個々の学術成果にリンク可能なユニークな識別子を付与することであり、それにより科学的発見のプロセスを強化し、助成や共同研究の効率性を上げることができる」ということで、これがミッションになっています。ポイントは、大学や助成機関、学会、出版社、および学術流通に関係する企業がみんなで一緒にやりましょうということと、NPO であるということです。

ORCID のこれまでの流れ

時系列で見ると、2009 年 12 月 1 日に ORCID イニシアチブが作られました。この前月あたりにトムソ



(図 1) 世界の論文数の変化/国際共著論文の変化



(図 2) 主要国の論文数の変化

ン・ロイターと Nature Publishing が名寄せ問題に関するカンファレンスを行い、それを受けた形で、「イニシアチブ」という、組織のない集まりとして作られたのが、ORCID の始まりです。そして 2010 年の前半あたりからワーキンググループを幾つか作って進めてきました。いろいろなワーキンググループがありますが、Business working group と Technical working group が主たるもので、Business working group はその後理事会に変化しました。Technical working group はそのまま続いていて、これについては後の蔵川先生がお話しになります。

参加者会合を 2010 年 4 月にボストンで開きました。このときに集まったのは 30 人ほどと大した数ではあ

りませんでした。2010年12月にロンドンで行った参加者会合のときは60~70人になっていました。この間、1~2カ月に1度ぐらいミーティングを行い、2010年8月にアメリカのデラウェア州登録のNPOとして法人が発足します。このときをもってORCIDイニシアチブがORCID Inc.になり、理事会メンバーが確定しました。その後、運営メンバーを確定し、一番最近では今申し上げた2010年12月にロンドンのWellcome Trustで参加者会合を開いたという流れです。実際に動き始めてからほんの1年ちょっとです。

最初のころの参加メンバーとしては、出版社が結構入っていますが、出版社以外でも入っているところがあります(図3)。国立情報学研究所(NII)は、たまたま2月か3月にトムソン・ロイターを訪問したときに、日本からも参加しないかという話が出て、そのときから参加しています。現在は141機関が登録していますが、実は今のところ、登録といっても特に何もなく、単に興味があるというだけの登録です。

理事会メンバー、運営メンバーの構成

理事会メンバーを、Commercial sectorを赤、出版社をオレンジ、学会を緑、Funding Agencyを青緑、大学を青、研究所と紫に色分けしてみると、純粋な出版社としてJohn Wiley & Sons、Nature Publishing、Elsevierの3社、Commercial sectorとしてトムソン・ロイターが入っていますが、あとは青系統が多いことが分かります(図4)。実は先ほどの宣言にも書いてあったように、この団体自身がNPOですが、運営もNPOの人が過半数になるようにという規約を設けて明記しています。青系統が多いのは、過半数を必ずNPOで占めなければいけないからなのです。

詳しく見ていきますと、Wellcome Trustはイギリスの助成機関です。Harvard University Library、John Wiley & Sons、Hannover Medical School Library、OCLC(Online Computer Library Center, Inc.)、トムソン・ロイター、CERN(European Organization for Nuclear Research: 欧州原子核研究機構)、Publishers



(図3) 最初のころの参加メンバー



(図4) 理事会メンバー

International Linking Association、Nature Publishing Group、ACM(Association for Computing Machinery, Inc)、Elsevier、MIT Library、そしてNIIからは私、Cornell University LibraryからはarXivを手掛けている人が入っています。欧米系が非常に多く、割と英語圏に偏っている感じが強いです。

運営メンバーになるとだいたい色が変わって、チェアマンはNature PublishingのHoward Ratnerです(図5)。

ORCID Principles

つい最近、ORCIDでは、ORCID Principlesという10の原理を定め、この原理に基づいて運営していくこ

としました。一部は結構ディスカッションしましたが、残りは割とあっさり決まっています。

重要なところだけ紹介しますと、3 番目に「ORCID は学術コミュニケーションに興味を持つすべての組織に対して開かれている」、つまり、どこにも向いていない、全方位であるということ、オープンであるということ宣言しています。5 番目は、「研究者は ORCID ID とプロフィールの生成、編集および維持を無料で行える」、研究者（ユーザー）からはお金を取らないと宣言しています。6 番目の「研究者は ORCID の自分のプロフィールデータのプライバシー設定で制御できる」は、ヨーロッパはプライバシーに厳しいので入れました。

7 番目は結構もめたところでもあるのですが、「ORCID に入力されたすべてのプロフィールデータは CC0 に従って年に 1 度改訂され、標準的なフォーマットで無料でダウンロード可能である」、つまり、このデータベースそのものはコミュニティーでシェアしますということを言っています。ちなみに CC0 とは、Creative Commons の中で、普通のライセンスとは別の、いわゆるコピーレフトに当たるようなものです。要するに全く権利を主張しませんということです。議論になったのはオープン化か否かではなく、CC0 という個別のモデルに基づいてオープンにするかどうかの問題になっただけで、総意として公開することに関しては合意が取れていました。

8 番目もちょっと特徴的で、ORCID でソフトウェアを作った場合は、これをオープンソースのソフトウェアにする。もちろん、元のソフトウェアを改造した場合、それがオープンソースでなければその部分は公開できませんが、その元になるソフトウェアもなるべくオープンソースのソフトウェアを使うことを期待すると言っています。これも結構議論になったところですが、一体 ORCID はどのようなシステムを何に基づいて作るべきかということで、結局ここに落ち着きました。

9 番目は、「ORCID ID とプロフィールデータは、無料あるいは有料の API とサービスを通じて提供され



(図 5) 運営メンバー

る」、ORCID はどのようにして維持するのかということになるわけですが、ここでは一応お金も取り得ることが書いてあります。ただ、その具体的な方法はまだ明確化されていません。個人の研究者は無料でアクセスできて、ではどういう場合に有料になり得るか。これも後で議論します。

最後に、過半数は非営利団体で運用するということを宣言しています。

財政モデル

財政モデルについては、最初は出版社ベースで議論が進みました。トムソン・ロイターの David Kochalko が、会費収入を中心として年収 3.5 ミリオンドル（約 3 億円）、会費は CrossRef 型（メンバーフィー+投稿フィー）というビジネスモデルを提案したのです。

ところがメンバーから、これは無理なのではないかという意見が出ました。理由はいろいろあるのですが、論文の場合は出版社に「自分の出版物」という概念があるので、そこにお金を払うのは理にかなっているのに対して、人間のデータは、別に出版社が排他的に持っているものではありません。むしろ出版社が情報を上げることによって ORCID がうまくいくのに、そこにお金が生じるのはおかしいのではないかと、あるいは、財政規模として年間 3 億円を集めるようなスキームは無理なのではないかというわけです。フルスタッフを

何人が雇用し、システムのメンテナンスを行うことなどを全部含めて考えるとそのぐらいになるというのが彼の試算だったのですが、無理ではないかという話が出てきて、ここからだいぶ方向転換をします。

ですから、最初は CrossRef を下敷きにしたような組織を考えていたのですが、趣旨に合わないのと、実際にスケールとしてもうまくいかないのではないかとということで、むしろミニマムコストでできるようなモデルを作ろう、逆にそのためには目いっぱいオープンにしようというのが、先ほどの ORCID Principles の背景です。

ただ、ではどうやって運営するのかという話になると、やはり困るわけです。0 円では運営できないので、今、取りあえず寄付を募っています。スポンサーシップには Gold、Silver、Bronze があり、Gold でも 100 万円ぐらいなので大した額ではありません。2010 年 12 月ごろから募り始めて、2 カ月ほどの間に千数百万の寄付が得られるめどが立ったのですが、これを毎年続けられるかどうかはわかりません。そもそもコストがどれぐらいかもまだ見えていないのです。

現状では、システムそのものについてはミニマムコストを考えています。例えば CrossRef を手掛けている人の一人を半分雇って作ってもらおうという話も出てきています。お金を払うプログラマーも雇うかもしれないけれども、あとはいわゆるオープンソースで、例えば大学の人などに貢献してもらうことで開発費をできるだけ減らそうとしています。

ちなみに、トムソン・ロイターの Researcher ID のシステムを元にするというのが一つのプランですが、これはまだ決定していません。ソフトウェア的には、Researcher ID のソフトが拡張可能なソフトウェアかどうかをいちいち別のプログラマーに検証させ、一応大丈夫そうだったのですが、著作権や権利関係でつまづいています。そこがうまくクリアできないと Researcher ID のソフトが使えないようで困っているのですが、もし使えれば開発費が抑えられます。そして運営費もできるだけ抑えます。

今、実際にお金がかかっているのは法律関係です。トレードマークの世界中での登録と、そもそも NPO として登記する段階、運営の段階、プライバシーポリシーの調査などに結構お金を使っています。これにはそこそこお金がかかるということが分かっています。今は借金状態ですが、その借金は寄付で返せそうです。

サービスモデル

実は一番大事なサービスモデルにもいろいろ問題があり、まだ確定しているわけではありません。ただ、参加者に対して ORCID サービスに何を期待するかというサーベイを行ったところ、「ORCID ID による名前同定、検索」「プロフィール情報の登録、検索」「他のサービスの ID との連携」が上位三つに入ったので、多分これらが最初に取り組まれるでしょう。

「ORCID ID による名前の同定、検索」は、まさに ID を付与し、その ID によって検索できること、「プロフィール情報の登録、検索」は、所属やそのほかの属性を検索できること、それから「他のサービス ID との連携」は、例えば Web of Science や CrossRef など、ほかの ID と連携させてほしいということです。プロフィールには多分出版情報も含まれると思いますが、一義的には、まずプロフィールが取れるということです。

また、「より広範な研究者情報の登録、検索」、例えばファンディングの情報なども入れたらいいのではないかとということもアンケート項目には含めたのですが、あまり興味を持たれていないようなので多分後回しになるでしょう。

そして、「機関単位の研究者情報の登録、検索」、これは先ほどの財政モデルにも関係してくるもので、実際にお金が取れると思ったのはこの辺なのです。機関単位で、例えばバルクで所属研究者のデータをあげる、あるいは逆にバルクでデータをもらうところでお金が取れるのではないかと議論はしているのですが、システム的には大変な部分なので後回しになるのではないかと気がします。ですから、最初のころはか

なりミニマムで、とにかく ID が登録できるというところでは、

研究者同定の方法

その ID の登録が実際にどのような手順になりそうかという、まず、基本的には本人が登録します。最初のころは、本人の登録なしで出版社同士のデータを持ち合って充てればいいではないかという話があったのですが、本人の同意なしに公開するのはまずいので、自分でログインして「これを公開してよい」と言わなくてはならないことになりました。アメリカの方は割と気楽に考えていたのですが、ヨーロッパ側から EU のプライバシー規定に引っ掛かるのでまずいと言われたのです。もちろん本人が権利を委譲した機関があれば、その機関が行うことも可能です。

それから、各種サービスからの ID 結合を行います。ここには先ほど言った出版社の著者 ID などがデポジットされるので、それとのひも付けをするわけです。それから、プライバシー設定などをします。これが基本です。ですから、システムとしては、むしろ各種サービスからプロフィール情報を取ってきて、参照情報として持っていく、これ自体は公開されないという仕組みで運営することが考えられています。歯にものが挟まったような言い方ですが、実際にシステムがどれだけできるかもまだ確定していないので、あくまでこういうモデルでいきそうだという話をしています。

未解決な問題はいろいろ

問題はたくさんあります。照合に使えるデータは何か。個人が申告するデータは何か。由来を判断するために使えるデータは何か。実際に機械的な照合をどうするのか。既にあるものをどう照合するのか。これには何かアルゴリズムが必要で、既にやっている組織からもらってきたらいいのではないかという議論もありました。

コラボレーション

外的な意味で言うと、ORCID は出版社や大学が参加して運営していますが、似たようなもので International Standard Name Identifier (ISNI) があります。これも最近の動きなのですが、ISO 標準に提案されているものです。ただ、標準と言っていますが、標準だけではなくサービスもちゃんとするというのを念頭に置いており、Online Computer Library Center (OCLC) がかなり主体的に動いています。これは著者名典拠の国際的な名寄せである VIAF (The Virtual International Authority File) の拡張版というイメージです。VIAF はあくまで本の著者ですが、例えばパフォーマーなど、いろいろな世界の人物にも同様のアプローチをすることを狙っているようです。

ただ、ORCID の関係で言うと、そもそもサービスのレベルが違うので、サービスモデルとしては統合できません。ISNI は、名前、生年月日、どんな役割・クラスか (音楽家である等) といったミニマムなデータしか持たないので、明らかに違うサービスになるでしょうが、ID 統合の話は出ています。統合した方がメリットがあるのか、統合しない方が混乱を招かなくていいのかを含めて議論しているところです。

それから、アメリカには、大学などでライブラリーを共有することによってみんなが簡単に成果公開できるようにしよう、ソフトウェアを作り合おうという VIVO と呼ばれるプロジェクトがあります。VIVO のソフトウェアを ORCID で使えるかどうか。つい最近、相互の合意が取れたので、その検証からまず始めようとしています。

今後の展望

私からは、外形的なお話をしました。まとめとしては、現在、組織ができたところで、システムのプロトタイプの完成は 2011 年中を目指しています。サービスモデルはまだ決まっておらず、財政モデルも暗中模索の状態です。ただ、誰でも、どの組織でも参加はできるので、ぜひ日本からもどんどん登録してください。

今のところ日本からは科学技術振興機構(JST)、NII、
金沢大学などが入っているようです。寄付も募集して
いますので、ぜひという方はスポンサーになってくだ
さい。