

7

電子書籍化する学術論文

— CiNii Articles の展開を中心に —



大向一輝 (国立情報学研究所)



学術情報流通の現状

一般向けの書籍に先駆けて、学術論文の電子化がきわめて早いペースで進行している。1995年に世界初の電子ジャーナルが発行されて以来、論文作成・出版プロセスのデジタル化や流通におけるインターネットの利用は着実に進み、もはや電子的な情報なしには研究教育が成り立たない状況にあると言ってよい。本稿では、学術情報流通の現状や課題について、筆者が開発・運営に携わっている CiNii Articles を例に取り上げながら述べる。

一次情報と二次情報

著者が論文を執筆し、学会や出版社を通じて公開・提供が行われ、読者がそれを入手するまでの一連の過程を学術情報流通と呼ぶ。情報技術の発展によって、学術情報流通は急速に電子化され、研究活動そのものにも大きな影響を与えている。学術情報流通は多くのプレイヤーによって支えられている。一般に、学術情報は論文や研究データなどの一次情報と、それらに関する詳細な説明、所在、入手方法などを記述した二次情報に分けられる。

学会や出版社は、著者が執筆した論文を編集し、一次情報として入手可能な状態にするとともに、二次情報を提供することで読者の入手を支援する。電子化以前には冊子体の雑誌を発行することがこれらの組織の主な役割であった。電子化によって、学術情報流通の舞台はインターネットに移行しつつある。著者は自らの PC で作成した論文を電子ファイルとして学会に送付し、学会は編集を行った後

Web 上で電子ファイルを公開する。組版や印刷などの発行プロセスが大幅に簡略化されたことで、定型的な公開作業を支援する電子ジャーナルプラットフォームを利用する学会が増加した。Elsevier 社の ScienceDirect や Springer 社の Springer Link など、出版社自らが電子ジャーナルプラットフォームを提供し、二次情報の作成も請け負うことで多くの学会を集める例もある。冊子体でのみ発行されていた過去の論文や、電子的な出版プロセスを持っていない中小規模の学会に対して、デジタル化作業と公開を代行するサービスも存在している。また、近年では大学が主体となって構成員の研究成果を収集し、公開する機関リポジトリの普及が進んでいる。あるいは研究者個人が自身のサイトで論文を公開するセルフアーカイビングの動きもある。

二次情報には、論文のタイトルや著者名といった書誌情報や、どこで入手可能であるかという所在情報が含まれる。これらの情報は、出版元と購読者が分担して作成してきた。冊子体の時代には、大学における購読主体である図書館が手書きで作成する目録カードに配架位置などが記述されていた。また各機関の連携協力によって、どの機関が何を購読しているかを知るための総合目録が作成されてきた。1970年代からは目録のコンピュータ化、ネットワーク化が始まり、機関の枠を越えて雑誌や論文を検索し、図書館の相互貸借を通じて迅速に入手する体制が作られた。

一方、電子化された論文は、電子ジャーナルプラットフォームで公開され、読者はそこにアクセスする形態を取る。そのため、アクセスに必要な二次情報はプラットフォームが提供することになる。

Elsevier 社の Scopus など、各プラットフォームの二次情報を広く収集し、横断的な検索に特化した二次情報データベースや、論文間の引用・被引用に関する情報を独自に構築し、雑誌あるいは論文の評価指標を付加価値として提供する Thomson Reuters 社の Web of Science のようなサービスもある。近年では、Web 検索エンジンを運営する企業が、プラットフォームだけでなくセルフアーカイビングの論文まで広く収集する Google Scholar や Microsoft Academic Search などの検索サービスがある。

冊子体とは異なり、電子ジャーナルではファイル自体を販売するのではなく、プラットフォームへのアクセス権に関する契約を行うことが多い。この契約モデルによって、各研究機関は比較的安価に多くの論文を入手することができるようになった反面、契約が終了すると一切のアクセス権が失われる。このような購読方法の違いが、後に述べるような価格問題の一因になっている。契約モデル以外にも、論文単位で電子ファイルを販売し、所有権が確保される Pay Per View (PPV) や、従来通りの冊子体の購読契約に電子版のアクセス権を負荷するサービスなど、販売・購読のモデルも多様化している。

デジタル情報はコピーが容易であるため、長期的な保存に適している。著作権を侵害しない形で各国の機関が電子ファイルを持ち合い、アーカイブして不測の事態に備える CLOCKSS プロジェクトが国際的に進められている¹⁾。また各国の国立図書館でもデジタル情報の保存が進められるなど、電子ジャーナル時代への適応が徐々に進んでいる。

CiNii Articles の概要

CiNii Articles は国立情報学研究所 (NII) が 2005 年から運営する国内最大規模の学術論文検索・提供サービスである²⁾。これまでは単に CiNii と呼ばれてきたが、2011 年に大学図書館の図書・雑誌検索サービス CiNii Books を開始した際に改称した。CiNii Articles を前述の枠組みに位置づけると、電子化代行と二次情報データベースを兼ねたサービスで

あると言える。

論文の電子化代行は、NII の前身である学術情報センター (NACSIS) が 1990 年代初頭から提供してきた電子図書館サービス (ELS) をその起源としている。学協会の論文誌など、冊子体の出版物をスキャンし、書誌情報を作成して Web 上に公開する。公開に際しては無料・契約モデル・PPV を選択することができ、それに応じて契約や課金の代行を行う。ELS には 2012 年 9 月時点で 406 の学協会が参加しており、延べ 1,297 誌、342 万件の論文が公開されている。

二次情報データベースとしては、ELS のほか、国立国会図書館の雑誌記事索引、科学技術振興機構の電子ジャーナルプラットフォーム J-STAGE、各大学・研究機関が運営する機関リポジトリ、各種学会・出版社の記事データベースなどから提供される書誌情報計 1,500 万件を対象とした検索を行うことができる。検索結果からは本文の入手先へのリンクを表示する。

同一の論文に関する書誌情報が複数の情報源から提供されることは珍しくない。たとえば、ELS で電子化されている論文の大半は、国立国会図書館の雑誌記事索引の採録対象となっている。これらのデータを無加工でデータベースへ投入すると、検索結果に重複が生じ利便性が低下する。このため、同一の論文に関する情報であると認められるものについては統合作業が必要になる。1,500 万件を超える書誌情報について、手作業で同一性判定を行うことは現実的でない。CiNii Articles では、機械学習を用いた名寄せと、人手によるチェックを組み合わせることで信頼性の高い情報を生み出すワークフローを構築し、この問題に対応している³⁾。

学術情報のファインダビリティ

論文は学術的なコミュニティの中での情報流通を目的として出版されるものである。それゆえに一定の需要は確実に存在するものの、その需要がコミュニティの規模を著しく超えることは考えにくい。実

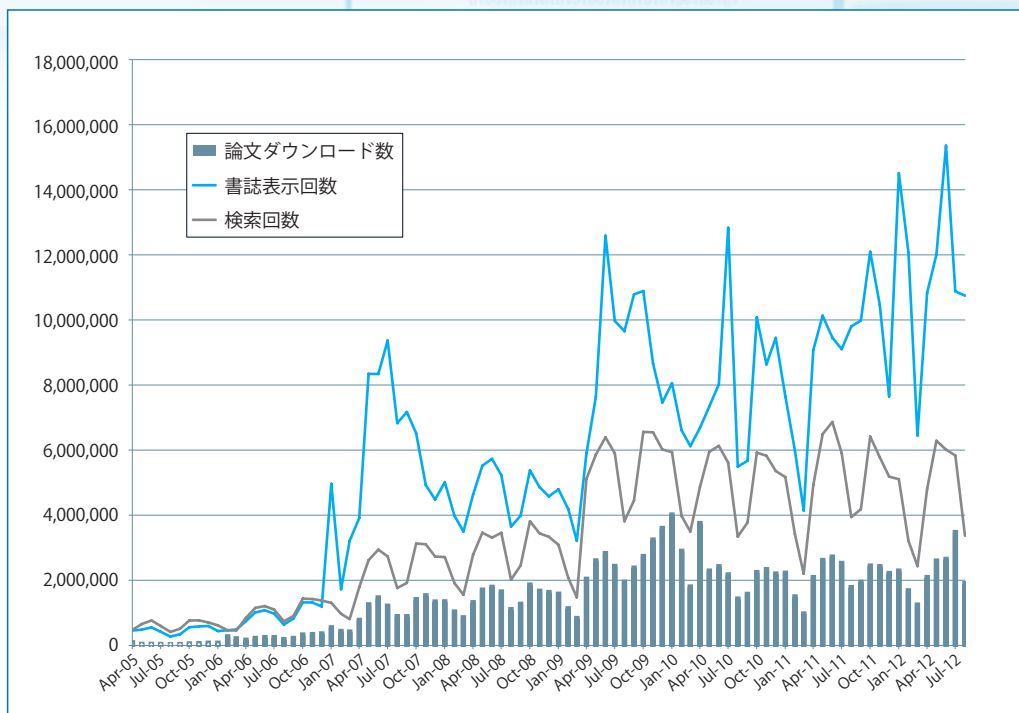


図-1 CiNii Articles のアクセス数の推移

際に、多くの学術系データベースは契約者だけが利用可能であるクローズドなモデルで運営されてきたが、契約者数の推移を見ても大きな変化はない。

一方、Web はわずか 20 年ほどで世界中を席卷するようになった。その成功の要因についてはさまざまな議論があるが、Web 特有のオープンさによって情報が多様なユーザ層に利用され、予期しない評価がなされることで新たな情報公開に対するインセンティブが高まるといった正のフィードバックが働いたことも一因であろう。結果として、さまざまな情報が Web に集まり、逆に Web で発見できない情報は存在していないものと見なされるようになりつつある。こうした状況の中では、学術情報であっても従来のクローズドな運営を続けていると同じように認識される恐れがある。

そこで、CiNii Article では 2006 年からサービスのオープン化に着手した。まず、従来は契約者に限定されていた書誌情報へのアクセスを一般ユーザにも開放し、一意な URL を定めた。これによって、1,000 万ページを超える書誌情報が誰からでも参照できるようになった。2007 年にはこれらのページを Google の Web 検索ならびに Google Scholar の

検索対象とする連携を開始した。2009 年にはユーザインタフェースを一新したシステムへの切り替えを行い、2010 年には Yahoo! Japan と連携し、論文検索機能の提供が開始された。

CiNii Articles がスタートした 2005 年 4 月から 2012 年 8 月までのアクセス数の推移を図-1 に示す。年々の増加傾向とは別に、オープン化の施策を行うごとに非連続的な伸びを記録していることが分かる。現在では、季節による変動があるものの、合計で月間 1,500 万～2,000 万アクセスを記録している。また、アクセスログの分析より、全アクセスの 60% 程度が外部の検索エンジンから書誌情報へ直接遷移していることが明らかになった。

論文のダウンロード数も年を追うごとに増加している。誰もが入手可能な無料論文のダウンロード数が増加していることは、オープン化によるユーザ層の拡大で説明ができる。しかしながら、契約モデルで提供されている論文についても同様の傾向が見えるのが興味深い。契約モデルの対象となっている機関数が変わらないにもかかわらずダウンロード数が増えたことは、オープン化によって研究者や学生にとっても論文を目にする機会が増え、読まれるよう

になったことを示唆している。

たとえ専門家であっても、広範囲に及ぶサーベイや他分野の調査をする際の行動様式は一般ユーザと変わらない。情報そのものが持つ見つけやすさ、すなわちファインダビリティを高め、想定ユーザの動線の中に情報を組み入れることで利用率は大幅に上がることが示された⁴⁾。

2009年からは、CiNii Articlesの書誌情報を機械可読なフォーマットで提供するWeb APIを提供している。これを利用した派生サービスも徐々に増えており、WebAPIへのアクセス数はユーザ向けサービスとほぼ同等の数値を記録しており、さらに増加する傾向にある。

ボーン・デジタル時代の 学術情報流通

デジタル化による出版コストの劇的な低下によって、誰もが学術情報流通の担い手になることが可能になりつつある。その過程で、電子化代行など、過渡期に必要とされていた機能は遠からず整理されるものと思われる。一方で、学術情報流通の多様化によって、学会や電子ジャーナルプラットフォーム、二次情報データベースが捕捉できない情報も増加している。特に、国際会議の予稿集などでは、速報性や柔軟性を求めて独自に論文を公開する例が増えている。これらの論文に関する情報を広く収集するには、Web検索エンジンで用いられるようなクローラが必要となる。このことから、今後の学術情報流通の起点はWeb系の検索エンジンに移り、それと同時に既存の仕組みの地位が低下するとの観測もある。

しかし、学術情報流通においては二次情報の重要性がきわめて高い。一般のWebページと比較して、論文では書誌情報の構造化がなされており、文献を引用する際には書誌事項を記載すればほぼ一意に特定できる。また、プラットフォームの変更に左右されない恒久的なIDを提供するDOI (Digital Object Identifier) のような機構も整備されている。これ

らは、学術研究において漏れのない網羅的な文献調査・引用が必要とされているからこそ継続的に維持管理がなされている。二次情報を作成せずに公開された論文については、検索エンジンで必ず見つかる保証はない。電子ファイルの解析によって論文そのものから書誌情報を自動的に抽出する研究も存在するが、多様なレイアウトの論文があり、確実に抽出することは困難である。Webページ上で書誌情報を記述するフォーマットも標準化がなされていない。現状では、Web系の学術検索サービスにおいても、出版社などの信頼できる情報源から二次情報を得て、その情報に対してWeb上で発見された本文ファイルをリンクして提供されていることが多い。これまで、二次情報の作成コストは投稿料や購読料で賄われてきた。オープン化が進む中でコストを誰がどのように負担するのか、さらなる議論が必要である。

新しい動きがある一方で、既存の電子ジャーナルプラットフォームの大規模化・寡占化も大きな問題になっている。多くの有力学会の出版を大手プラットフォームが担うようになり、それらのアクセス権の契約を包括的に行うビッグディールと呼ばれるモデルが浸透している。研究機関にとっては個別に契約を行うよりもコストが削減でき、またアクセスできる情報量も多くなることから歓迎されてきたが、論文の入手手段がこれに限定されるようになり、契約料が高騰しても負担せざるを得ない状態になっている。中には契約料を払うことができず、PPVによる個別論文の購入に切り替えた機関もある。自由に論文にアクセスできるインフラがなければ研究そのものが滞るため、研究機関は厳しい選択を迫られている。多くの国では複数の機関によるコンソーシアムを結成し、交渉力を高める努力を行っているが、寡占状態にあっては大きな効果は望めないという指摘もある。

こういった商業プラットフォームに対抗して、誰もが論文に自由にアクセスできる論文誌を作るといったオープンアクセス運動が活性化している。オープンアクセス誌は投稿者が掲載料を払うモデルであり、一部の分野では有力な雑誌も登場している。一方、

掲載料の高額化も始まっており、ビッグディールに替わる新たなビジネスモデルとして商業プラットフォームが取り組んでいる例もある。

2008年には、米国国立衛生研究所（NIH）の助成による研究成果はすべてオープンアクセスにしなければならないという義務化の方針が出され、出版界を含めて大きな議論が巻き起こったが、徐々にこの方針は定着しつつある⁵⁾。今後各国の研究資金制度においてもオープンアクセス義務化の議論がなされるものと思われる。

このように、ボーン・デジタル時代の学術情報流通は混沌としており、その行く末を予想するのは困難だが、今後書籍の電子化が進むにつれて同様の事態が生ずることは想像にかたくない。無論、学術情報と一般向け書籍ではさまざまな前提条件が異なるが、印税収入に依拠する職業作家以外の書き手については、学術情報流通と同じような電子化の動きが加速していくと思われる。学術情報流通の形態も、論文のマルチメディア化や研究データそのものを成果と見なす動きが強まればさらに大きな変化を遂げ

ることになるだろう。それにとまって課題も増えていくことになるが、1つずつ解決して知見を積み上げることで、電子書籍の世界にも大きな貢献ができるものと期待される。

参考文献

- 1) CLOCKSS, <http://www.clockss.org/>
- 2) 大向一輝：学術情報サービスのユーザモデルとファインダビリティ, 情報の科学と技術, Vol.58, No.12, pp.595-601 (2008).
- 3) 相澤彰子, 大山敬三, 高須淳宏, 安達 淳：レコード同定問題に関する研究の課題と現状, 電子情報通信学会論文誌 D-I, Vol.88, No.3, pp.576-589 (2005).
- 4) Morville, P.: Ambient Findability - What We Find Changes Who We Become, O'Reilly Media (2005).
- 5) 三根慎二：政策としてのオープンアクセス：NIHパブリックアクセス方針の現状と課題, カレントアウェアネス, No.289, pp.2-4 (2006).

(2012年9月29日受付)

● 大向一輝 (正会員) i2k@nii.ac.jp

1977年京都生まれ。2005年総合研究大学院大学修了。博士(情報学)。同年国立情報学研究所助手, 2007年同助教, 2009年同准教授。2010年総合研究大学院大学准教授を併任, 現在に至る。セマンティック Web, ソーシャルメディアならびにオープンデータに関する研究に従事。

