



e-プリント・アーカイブの衝撃

理学研究科 教授 九 後 太 一

1. インターネット

情報革命とかITと言われて久しい。しかし私がそれを本当に実感したのは次のような「小さな体験」を通してであった。

私は今春、専門外の一般の人々を対象に、「素粒子論の発展」について話をするよう依頼された。「理論物理学の特質は、『何か未知のものの存在』を論理的に予言するところにある。」ということをお話そうと思った。歴史的に分かりやすいところからの例として、その時に思いついたのが冥王星の話である。海王星が何か奇妙な動きをするというのでその原因として冥王星の存在が予想された。海王星のデータから理論的な計算のすえ冥王星の軌道が予言され、その予言にかなり近いところに実際に発見された、という話をどこかで聞いたのを思い出したのである。ところが人に話をするにはそれが一体いつの話で、理論的な予言をしたのが誰で、そして実際の発見者は誰なのかくらいは知らないといけな。そこで私は理化学事典や物理学事典を調べてみたのだが、どれにも載っていない。大変困ってしまった。

ところがそれを調べていた部屋に、私の研究

室のある大学院生が来たので「何で調べれば載っているだろう」と話すと、彼は即座に「そんなのは簡単ですよ。インターネットで調べれば一発です。」果た



して、Googleの検索画面を開き、「冥王星 海王星」というキーワードをいれてたただけ、たちどころに横浜こども科学館ほか1万5千件が引っかけた。トップの横浜こども科学館につなぐと、NASAやローウェル天文台のページへリンクがあり、それこそあつという間に「冥王星（Pluto）は、海王星（Neptune）の軌道を横切る惑星としてP. LowellとW.H. Pickeringによって予言され、1930年に、Lowell天文台の天文学者C.W. Tombaughによって発見された。」という確かな情報が得られたのである。

この「事件」は私にはショックだった。私には先ず事典を調べるという旧来の発想しかなかったのだが、時代はもうはるか先に行っている

のである。ちなみにこの大学院生は、論文の自分の英語の使い方が正しいかどうかを調べる時にも、英和辞典は使わず、Googleにそのワードを打ち込んでいっぱい引かかる文例を見るとのことである。

2. 論文作成とプレプリント

私が大学院に入学したのは1971年であるが、そのころの素粒子論分野での最新の研究情報というのはプレプリントであった。新しい研究成果が出て論文を書きあげると、先ずもちろん学術雑誌に投稿する。同時に、その原稿を200～300部（大きな研究所では500～1000部）印刷してそれを全世界の主だった研究所や関連の研究者に直接郵便で送るのである。この印刷物が「プレプリント」と呼ばれていた。

このプレプリント作成作業が大変だった。先ず、印刷するためには「十分美しいタイプ原稿」を用意しなければならない。それにふつうの研究者はタイプを打てなかったので、先ず手書きの原稿を用意して秘書さんにタイプしてもらう。我々の分野ではギリシャ文字などの特殊な文字や数式が多く含まれているのでその部分はタイプ原稿では適当な空白を空けておいてもらい、後で手書きで入れるのである。また論文が図を含んでいたりすると、製図用の烏口や雲形定規などを使ってインディアンインクで美しく描かねばならない。

このようにしてできた原稿を、次に印刷し「製本」（頁をそろえてホッチギスで止める作業）しなければならないが、この部分は当初は印刷会社に外注していた。しかしこれは時間もお金もかかるというので、その後、印刷は教室内の事務の人にやってもらうようになり、オフセット印刷の美しい仕上がりのものになった。「製本」は著者自身が手作業でやっていた。一時コレクターと呼ばれる割合高価な機械が教室に入ったこともあるが、よく調子が悪くなり、機械を直しているより結局手でやるのが速かった。そ

して最後は、発送作業である。これも切手貼り、研究機関宛名のラベル貼り、個人的な送り先の宛名書き、などひと仕事である。

3. プレプリント・システム

プレプリントは、このように作成も大変で、いろいろ手間や費用がかかったが、自分の研究をいち早く世界の研究者に伝え、また他の研究者の最新の研究成果を知るために、もっとも効果的な方法であった。もちろん学術雑誌に投稿された論文はゆくゆくは雑誌に掲載されるのだが、査読者の閲読・判定や印刷・校正などに時間がかかる。実際に刊行された雑誌で研究者が論文を見ることができるのは、半年後くらい（きわめて早い場合でも3ヶ月、遅いと1年以上）になるので、発展の速い我々の分野ではプレプリントの果たす役割は絶大であったのである。

このように世界中に出回っているプレプリントの情報については、それぞれの研究機関が1週間に受け取ったプレプリントを毎週リストにし現物を書架に展示するというをしていた。例えば私のところでは、基礎物理学研究所と物理学教室がそれぞれでリストを作り、お互い情報を交換していた。このプレプリントの管理およびリスト作りは、論文のタイピングの仕事と並ぶ教室の秘書さんの大きな仕事の一つであった。いつの頃からか正確ではないが、1980年代にはSLAC（Stanford線形加速器センター）が、受け取ったプレプリントをリストにし毎週毎週、全世界の研究機関に送付するというサービスをボランティアにやり始めた。このリストには著者名、研究機関、タイトル、ページ数が小さな字でぎっしり書き込んであった。興味がある題名の論文をリストに見つけると、自分の研究機関に来ていなければ、直接著者の研究機関にプレプリント請求のはがきを出すのである。

このようなプレプリントを中心にした最新研究情報交換のシステムはずいぶん長く維持され、結局1992-3年ごろまで続くのである。しかし、転換の兆しは1987年頃にやって来た。数学や物理学の分野での新しい論文清書システム—TeX の登場である。

わが物理学教室の理論グループでは、上述のプレプリント・システムでの論文作成をより効率的にするため、1980年頃に一台目の、1985年頃に二台目のIBMのワードプロセッサを導入した。どちらもその時代のかかなり最先端のもので、数百万円もしたと記憶している。しかし、どちらもギリシャ文字を打つには印字部の「ボール」や「カエデ」をいちいち交換せねばならず、数式をきれいに打つことはさらに難しかった。だからこれらの高価なワープロを購入しても、数式部分は依然として手書きで書き加えていたのである。

ところが、1987年の中頃から研究室にTeX のシステムが導入された。これは、パソコンの上で5インチフロッピー1枚に入ったプログラムで動き、ギリシャ文字はもちろん数式も全く自由に書け、アウトプットも極めて美しい画期的なフリーソフトウェアだった。このTeX システムは、前年1986年末に私が米国のSanta Barbara研究所に滞在した折には、既にその研究所では各部屋で使えるようになっていた。その後、TeX は、日本にもヨーロッパにも急速に普及した。

4. e-プリント・アーカイブ

このTeX の普及が情報交換に本質的な変化をもたらすことになった。すなわち、1991年8月からLos Alamos国立研究所でPaul Ginsparg が始めた e-プリント・アーカイブである。これは最初、Ginsparg が自分の専門の素粒子論のストリング理論や2次元量子重力の分野で、200人程度のユーザーを想定して始めたプレプリント・データベースで、高エネルギー物理学の理

論 (High Energy Physics -- Theory) という意味で hep-th と名付けられた。その分野のプレプリントを、ハードコピーではなくTeX ソースファイルの形でコンピュータ上にストアしておき、世界中どこからでもLos Alamosにアクセスしていつでも瞬時に取り寄せられるというシステムである。名前を登録した研究者には、一定期間毎に (現在ではweekday毎日) アーカイブに新しく投稿された論文の、題名、著者名、アブストラクト、がe-メールで送られてくる。もし本文を読みたい論文があればLos Alamosにアクセスして取り寄せるのである。

ところが始めるや、hep-th のユーザーは2,3ヶ月後には1,000人に、2,3年後には4,000人にもなった。程なく、hep-th以外にも、高エネルギー現象論 (hep-ph) 原子核理論 (nucl-th) 物性物理学 (cond-mat) 数理物理学 (math-ph) 数学 (math) などを初めとした物理学や数学の数々の分野 (現在15) のe-プリント・データベースが次々と開設され、同じく数多くのユーザーを抱えるようになった。

このe-プリント・アーカイブに保存されるのは、ふつうの場合、論文 (プレプリント) のTeX ソースファイルと、論文中に現れる図のポストスクリプト形式のファイル、などの一組である。物理学や数学の論文をこのようなe-プリント・アーカイブに保存することが可能になったのは、コンピューターとネットワークの発達とともにTeX のおかげなのである。TeX によって手書き部分の無い完全原稿がデジタルデータとして用意でき、コンパクトなサイズのe-プリント・アーカイブが可能になった。スキャナーで画像情報として取り込んだ場合はファイルが馬鹿でかくなるだけでなく、情報の2次利用にも使えない。

e-プリント・アーカイブは、最初の2年ほどは、研究者からの論文投稿はe-メールで送付し、アーカイブから読みたい論文はftpで取り寄せる、という方式だった。しかし、1993年の12月

には WWW (World Wide Web) のインターフェイスが用意され、以後論文取り寄せは Web から download する方法が主になり、1996 年 6 月には、論文投稿の方も Web Upload ができるようになった。また、全世界に 17ヶ所のミラーサイトが用意され、世界のどこからでも素早く download できるようになった。

我々の分野の研究者には、この e-プリント・アーカイブと同じくらいよく利用しているデータベースがもう一つある。SLAC が提供している SPIRES (Stanford Public Information REtrieval System) の 高エネルギー物理学 (HEP) 文献データベースである。これは先に述べた 1980 年代のプレプリント・システムの時代から SLAC が DESY (ドイツ電子シンクロトロン研究所) と共同プロジェクトとして始め、1974 年以來の高エネルギー物理学分野のほぼ全てのプレプリント、学術誌掲載論文、会議録などの、50 万件以上の情報を集めた巨大なデータベースである。収録されているのは、論文の題名、著者名、プレプリント番号、e-プリント・アーカイブでの登録番号、掲載済みの場合は、掲載雑誌名、年号、巻、ページ数、等の情報である。1993 年 6 月にはいち早く WWW で利用できるようになり、現在では、e-プリント・アーカイブや掲載雑誌のオンラインページへのリンクが張っており、直接その論文の内容を見ることもできる。



e-プリント・アーカイブの日本のミラーサイト (<http://jp.arXiv.org/>)

SPIRES は、この分野のほぼ完全なデータベースで極めて便利なものである。昨今、自分の論文の被引用回数を調べさせられる機会が増えたが、これなど朝飯前で、著者名を打ち込むだけで 1974 年以降の全論文に関してすぐさま被引用回数を答えてくれるし、被引用回数でランク分けした論文の度数分布 (成績表!) まで出すこともできる。もっと創造的な使い方は、例えば、何か人の話を聞いたり、論文を読んだりして新しいテーマの研究を始めようとする時、先ずその人の当該の論文を SPIRES で検索する。その検索結果の下のある部分をクリックすれば即座に「その論文を引用している論文」を全てリストアップしてくれる。それを調べればそのテーマの最新の状況がわかるのである。これは「未来への検索」であり、データベースが無ければ決してできなかったことである。また、昔は論文を書いた時、最後に参考文献の情報を完全にする作業が面倒だった。昔読んだ文献等の場合、うる覚えの年号等を頼りに図書館の書庫で、雑誌の巻末の索引を手当たり次第に探すという作業をよくやった。しかし、今は SPIRES を立ち上げるだけでこの作業がすぐにできるのである。

5. 新しい図書館へ

この e-プリント・アーカイブのシステムは、非常に大きなインパクトを持っている。

1) 先ず、先行する上述のプレプリント・システムは、この e-プリント・アーカイブの普及に伴い、完全に取って代わられた。すなわち、上で説明したような、印刷、発送の手間と費用から完全に解放されたのである。

2) (我々の分野では) ほぼ全ての論文が e-プリント・アーカイブに投稿されるため、従来の学術雑誌に載るよりはるかに早く、世界中のあらゆる場所でリアルタイムで読むことができるようになった。

3) しかも、学術雑誌に載った後でも当該の論文はアーカイブに残っているので、掲載雑誌を所属大学・研究所が購入していなくとも、ネットにつながっているパソコンやコンピューター端末がありさえすれば、部屋にいながらにしていつでも即座に読み、プリントすることができる。

このことは従来の学術雑誌のあり方、存在意義に対して深刻な挑戦を突きつけている。従来の紙版の学術誌というのは、その論文の印刷・製本・配布・蓄積・コピーということが難しかった時代に生まれてきたものである。それゆえ、投稿された論文は同じ専門の研究者に慎重に査読され、パスしたもののだけが印刷に回される。査読という点は、論文の質に対する一定の保証を与えるもので現在の電子化の時代でも従来の学術誌の果たす「役割」の重要部分と見なされている。が、まさにその点が出版に時間がかかる元凶でもある。e-プリント・アーカイブは、(コンピューターによる自動的なチェック以外)この査読部分を全く無くし、また、論文の受理から公開まで徹底して自動化したシステムであり、それを維持するコストは従来の学術誌出版などに比べて2桁も安い。

私は、実は、湯川博士の始めた「Progress of Theoretical Physics」という雑誌の編集委員をしており、その関係で、2000年に発足した物理系学術誌刊行協会IPAPの理事の一人である。アメリカでは、Physical Reviewを初め物理学関係の10学会⁸⁶誌の学術誌が合同でAIPという組織を作っている。日本でも物理学関係の学術誌がIPAPという合同組織を作り、この電子化の流れの中でリーダーシップをとり、日本を物理学の分野でもアメリカやヨーロッパに並ぶ情報発信の第三極とすべく設立された。IPAP傘下の学術誌も、AIP傘下のそれと同じように従来の紙版の雑誌のオンライン化を強力に進めている。しかしながら、それを進めれば進めるほど自身の経済的存立基盤を掘り崩すディレンマを抱えている。Web上で雑誌の論文を読める

ようにすると先ず「課金するのだろうか?」という難問に遭遇する。課金する方法という技術的な問題以上に、「課金すればその雑誌を読んでもくれなくなり、ひいては投稿も減る。」という、より深刻な問題である。上述のe-プリント・アーカイブのような無料のアーカイブにほぼ同じ論文のコピーがある時に、わざわざ有料のWebに誰がアクセスするだろうか?

図書館も、同じく、その存在理由を問われている。我々の分野の多くの若い研究者は今や、学術雑誌を見に図書館に行くことは減多にない。先ず、最近の参考文献に関してはそれが学術誌に掲載されていようがいまいが、全てe-プリント・アーカイブからdownloadする。e-プリント・アーカイブに投稿されていないような古い論文の場合にも、掲載雑誌のWebに置かれているPDFファイルをとってくる。あるいは、SPIRESのデータベースに昔のプレプリントのスキャンイメージがリンクされていないか見る。それでもだめな場合に初めて、図書館に行くのである。こういう状況では、従来の学術誌を従来通り書庫に並べただけの図書館では利用がますます減るだろう。大学全体で学術誌をオンラインで見られるようサイトライセンスをとる手配をし環境を整える、というのも重要な仕事であろうが、それだけでは図書館に将来はないだろう。

図書館は、いまや、より積極的にローカルな情報の収集・蓄積という仕事に取り組むことが重要であろうと思う。例えば、京都大学の博士論文・修士論文の収集である。博士論文などは論文として学術誌に発表されている内容もあるが、全く独立に書き下ろされたものもあり、外部からは入手が難しい。いわんや修士論文などはそうだろう。そういうものを収集して外部からオンラインで検索・閲覧ができるようにするなど、創意工夫をもって積極的に情報を発信する基地になる必要があるだろう。いま時代は動いている。

(くご たいち)