# II-(5) 研究者のアクセス手法 I:自然科学の研究者と図書館

筑波大学大学院数理物質科学研究科教授 学長特別補佐(研究戦略、情報化戦略担当)

宇川 彰

### §1. はじめに

大学教員と図書館が切っても切れない関係にあることは言うまでもないが、その有り様は学問の分野によってずいぶんと異なり、また、同じ分野であっても、時とともに大きく変わって来ていると思われる。私自身は、素粒子物理学の理論的研究という分野に身を置いて過ごしてきた。学生であった1970年代には、図書館にこもって学術雑誌を調べ、重要な論文だけは、当時は高額であったゼロックスコピーに取って丁寧にファイルに整理し、といったクラシックな関係を楽しんだが、現在では、図書館に足を運ばずとも、また自分の研究室に居なくとも、何処にでも持ち歩くノートPCからe-printのarchiveにアクセスすれば最新の論文が手に入る。また、旅先からでも archive に論文プレプリントを投稿すれば、翌日には世界の研究者が見て、賞賛や批判の電子メールが飛び込む時代になった。

素粒子物理学は学術情報の伝達に関して先端的だが特異な分野との印象を持たれているようである、実際、この分野では20世紀を通じて早いペースで進歩が続いたが、そのことに付随する強い競争意識を反映して、最新の研究成果の伝達には異常なほどの努力が払われた分野である。また、素粒子物理学の実験的研究には、大規模な加速器や、電子部品のかたまりのような素粒子検出器が活躍するが、これらは時代時代の最先端の計算機技術を駆使しており、このような技術を情報伝達に応用する試みも積極的に行われた。たとえば、素粒子物理学の分野では、インターネットの揺籃期である1980年代の末から電子メールの利用が常態化していたし、現在、インターネットの基本技術となっている world wide web (いわゆる "ホームページ") は、素粒子物理学の世界拠点のひとつである CERN(Centre Européenne pour la Recherche Nucléaire スイスのジュネーブ郊外にある)で開発された。また、プレプリント(学術雑誌に発表される前の段階の論文草稿)を電子的に蓄積し自由に閲覧できる e-archive も、この分野で発生したものであり、現在、内外の図書館で構築の努力が行われている "リポジトリ"の原型のひとつにもなっていると思われる。

本稿では、自然科学分野での学術情報の蓄積と伝達について、極端例のひとつとして素 粒子物理学分野の状況を概観し、併せて、今後に図書館が果たす役割について考えてみた い。

### § 2. 研究方法

自然科学の多くの分野と同様に、素粒子物理学の伝統的な研究方法は、実験的な方法と 理論的な方法である。実験的な方法の基本は、素粒子同士を高いエネルギーで衝突させて どのような現象が起こるかを調べる。このために素粒子を高エネルギーに加速する装置が 加速器であり、衝突の結果起こる現象を調べる装置が検出器である。現代の加速器は数千 億円の予算と10年に及ぶ建設期間を必要とする巨大科学の典型となっている。

理論的な研究は、いわゆる「紙と鉛筆」の研究である。実験結果をもとに、数学的な分析によって、その現象の背後に潜む法則を研究し発見しようとする。世界で最も有名な理論物理学者は、相対性理論のアインシュタインであろう。わが国の理論物理学の開拓者には、京都大学で活躍した中間子論の湯川秀樹や、筑波大学の前身である東京文理科大学・東京教育大学で活動したくりこみ理論の朝永振一郎がいる。

20世紀後半のコンピュータの発達は目覚しい。自然科学の分野での重要なインパクトは、1970年代に始まるスーパーコンピュータの発達と、それを使ったシミュレーションの発展であり、「計算による方法」が実験、理論と並ぶ、科学の第三の方法となっている。スーパーコンピュータは科学技術計算を高速に実行するように設計され、普通のパソコンの数万倍以上にも及ぶ計算性能を持つ。素粒子の衝突のような複雑な現象は「紙と鉛筆」では解けない場合が殆どである。また、宇宙は 137 億年前の "ビッグバン"で始まったと考えられているが、その直後に物質がどのような状態であったかを調べたいと思っても実験は不可能である。スーパーコンピュータは、理論モデルに基づいて、このような現象のシミュレーションによる研究を可能として、素粒子物理学だけでなく、自然科学全般において大きな進歩をもたらしている。

### § 3. 研究資料

以上の述べた実験・理論・計算、いずれの研究方法であろうとも、その研究の成果は最終的に学術論文として発表され、学術雑誌に掲載される。そこに至る過程で、さまざまの国際会議で発表が行われ、途中段階の速報や、最終結果の要約が会議プロシーディングスに掲載される。また、最近の傾向として、会議発表資料が会議のweb page のプログラムにリンクされることも多い。さらに、これらの論文や会議プロシーディングスを蓄積するe-archive やインターネット上のデータベース、データそのものを蓄積し検索・ダウンロードを可能とする各種のデータグリッドがあり、それらを検索するための情報検索エンジンがある。以下、素粒子物理学を例に取って、それぞれの状況を述べてみよう。

# (1) 学術論文の作成と TeX 革命

学術論文の形態については、1980 年代から 1990 年代にかけて二つの革命があった。第一は Donald Knuth による document 作成ソフトウェア TeX の開発と普及(1980 年代前半)であり、第二は preprint の蓄積・公開装置である e-archive の開発と普及(1990 年代前半)である。

物理学を中心とする基礎科学は自然の数学的な法則性を探求する学問であるから学術論 文には必然的に数学式が多数登場する。これをどのように処理するかは論文作成上の大き な問題であった。手動タイプライタしかなかった時代には、原稿作成の時に、通常の文章 だけを打っておき、数式は後ほど手で記入することが普通であった。また電動タイプライタの時代には、通常の文字へッドに加えて、ギリシャ字や数学シンボルの為のヘッドが特別にあり、数式を打つたびに両者を取り替えて使用したりした。

TeX は通常のワードプロセッサソフトに比べると、多少の学習と習熟を必要とするが、ひとたび慣れてしまえば、印刷水準の文書を、研究者個人が作成できる。(あまりに綺麗に出来るので、それだけで何かを達成したような気になって、内容を練り上げることが疎かになるとの批判はあるが)。学術雑誌に掲載された出版論文は、タイトル、要旨(アブストラクト)、複数の章からなる本文、その中に埋め込まれた数式や表や図、最後に参考文献から構成されている。LaTeX を使えば、それぞれの部分のフォントの種類や大きさ、章番号、式番号、図番号、引用文献の番号などが自動的に管理され、さらに、ps ファイルで図を作成すれば、それも指定した大きさにスケールされて望みの位置になるべく近く取り込まれる(最後の点はしばしば時間がかかるが)。原稿は直接に PC 上でタイプインすれば良く、何度校正を繰り返そうが、章番号や式番号は自動管理されているから気を使う必要はない。これによって、素粒子物理学分野では、1980 年代半ばから後半にかけて、論文作成は、タイプライタ印字と鳥口による作図の時代から、完全にコンピュータベースの TeX による文書整形と postscript による図の作成挿入に切り替わった。図1 に TeX 原稿と整形コマンド処理後の出力例を示す。

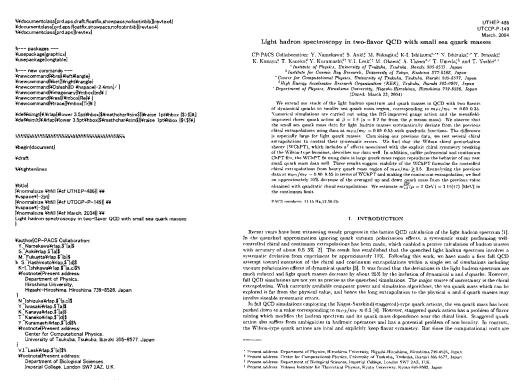


図 1 学術論文の TeX 原稿(左)と出力 postscript ファイルの例

TeX 革命は二つの意味がある。第一は、研究者にとって、論文作成の手間が大幅に軽減され、同時に PC さえあれば何処でも何時でも作成が可能になったことである。

第二は、論文媒体のすべてが電子化されるために、研究者間の論文交換が電子メールで 行えるようになり、瞬時に最先端の成果が伝わることになったことである。

現在では、素粒子物理学分野の学術論文や会議録はすべて TeX により書かれているといって間違いではない。また、素粒子物理学分野では、米国物理学会出版の Physical Review D, Physical Review Letters, Elsevier 出版の Nuclear Physics, Physics Letters 等が主要な学術誌であるが、こられの雑誌の多くは論文受付に TeX 及び ps ファイルを標準フォーマットとして投稿受付を行っている。

## (2) e-archive の出現

素粒子物理学分野の学術情報流通の第二の革命は 1991 年の e-print archive の誕生である。e-print archive とは、研究者が自らの論文をインターネット上のサーバーに送り、サーバー側ではこれを受け付けて保管し、世界中の研究者に自由に閲覧・ダウンロードさせる、完全にオープンで自動化された仕組みである。

e-print archive に投稿する研究者は、TeX と postscript 形式で論文テキストと図を作成し、そのファイル一式をサーバーに送る。サーバー側はこれを受け付けて TeX により整形された論文を postscript 形式で出力し、電子メールで著者に確認を求める。サーバーのデータベースは 24 時間毎に更新されるので、著者の確認が終わった論文は遅くとも 24 時間後の更新に反映されて世界に公表される。新しい論文のリストは登録したユーザには電子メールで毎日送付される。以上の仕組みは完全に自動化されており、ユーザは、収録論文リストを見たり、附属の検索エンジンを使って論文検索を行い、興味ある論文を閲覧したり、postscript やpdf 形式でダウンロードできる。

e-print archive は、e-journal 誕生以前の 1991 年 8 月に、素粒子理論の研究者 Paul Ginsperg が中心となってソフトウェアを開発し、素粒子物理学の超弦理論と呼ばれる分野の研究者グループを対象とした活動として始まった。瞬く間に物理学、数学、計算機科学に拡がり、最近では生物学の一部にまで拡大している。当初、米国 Los Alamos 研究所にサーバーが設置されていた (http://xxx.lanl.gov 現在は http://lanl.arxiv.org/と名前を変えている)が、現在では米国 Cornell University に管理が移り、世界 13 ヶ国 15 ヶ所にミラーサイトが置かれている。我が国のミラーは京都大学基礎物理学研究所にある (http://jp.arxiv.org/)。蓄積される論文数は、アーカイブが始まって以来増加を続け、現在 1ヶ月平均で約 3500 編、またサイトあたり接続数は 1ヶ月平均 100 万件に及んでいる。

図2の左側にCornell Universityで運用されているe-print archiveのトップページの一部を示す。物理だけで14におよぶ分野が運用されている。右側には、この中のHigh Energy Physics-Latticeに入り、簡単な著者検索を行った結果を示す。hep-lat が分野の名称、その下の7桁の数字が登録番号で、西暦の最後二桁、月、各月での登録番号が並んでいる。ダウンロードする形式は、論文全体のpostscript形式、pdf形式以外に、論文原稿であるTeX file や挿入されている図面のpostscript file もダウンロードすることができる。

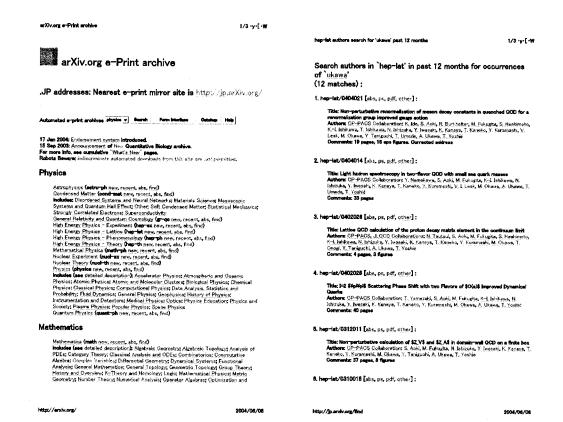


図 2 e-print archive (<a href="http://arxiv.org/">http://arxiv.org/</a>) のトップページ (左側) と、検索結果を表示したページの例(右側)。

### (3) 電子 journal

e-journal は今ではどの図書館にとっても重要なアイテムとなっているので詳しい説明は必要ないであろう。e-journal は 1990 年代半ばから一部の学会誌により先鞭が付けられたが、1990 年代後半に Elsevier、Springer、John Wiley 等の大手出版社が出版雑誌の電子化を推進するに至ってタイトル数が急速に増加した。学術雑誌の総数は世界で2万タイトルとも3万タイトルとも言われるが、現在約1万タイトルが e-journal 化されているようである。

最近発刊した学術誌には、電子版のみの完全 e-journal も増えているが、e-journal の多くは従来から発行されていた冊子体雑誌が、電子版へと拡張されたものである。購読料を払い、大学の IP アドレスによりアクセスの認証を行うサイト・ライセンス形式が多い。バックナンバーの電子資料化もかなりのスピードで進んでいる。私の分野では、アメリカ物理学会(American Physical Society)の発行する Physical Review は世界的にも最重要雑誌の一つであるが、1893 年 7 月発行の第一巻から現在までに発行された全ての巻と論文がGIF 及び PDF、場合によっては postscript 形式で電子化され、読むことができる。

e-journal の特徴は、発行と同時に即時に論文を読める速報性、学内との制限はあるにせよ、図書館の所在や開館時間に関係なく、何処でも・何時でも読めること、文献検索をその場で行えること、などである。いずれも、従来の冊子体の雑誌には望めない機能であり、e-journal の普及の大きな要因になった。

これに対して、e-archiveでは、ユーザは新しい論文の存在を登録後1日以内に知ることが出来る。また、完全にオープンなシステムであるから、インターネットさえあれば、誰もが何時でも何処からでも自由にアクセスでき、検索も容易である。また、サーバーと技術的知識・経験さえあれば少人数で運用できる。

## (4) データベース

素粒子物理学分野では、すでに 1970 年代からプレプリントの収集とデータベース化の努力が始められていた。Stanford 大学付属の Stanford Linear Accelerator Center が中心となり、1974 年以来、素粒子物理学分野で発表された学術論文や会議発表論文のデータベースを構築しており、現在 Spires と呼ばれている。



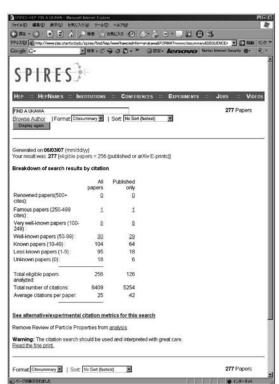


図3 素粒子分野の文献データベース Spires http://www.slac.stanford.edu/spires/hep/ 画面

素粒子分野に限れば、Spires はほぼ完全な文献データベースと言ってよい。個人の論文はすべて簡単なコマンドにより検索することができる。また、10年ほど前から、Spiresで

登録されている文献間の引用情報も整備されており、著者を指定すればそのすべての論文の被引用数、引用している論文のリスト、被引用数の多い論文の一覧、等をリストアップさせることができる。近年は個人の研究活動を示す web page を作成することが多いが発表論文リストを一々アップデートするのは面倒である。簡単な cgi を書くことにより、Spires と連動させて最新の文献リストを表示させることもできる。図 3 に Spires のトップページと文献検索を行い被引用数統計を表示させたところをしめす。

文献データベースと並んで重要なものは、研究成果そのもののデータを蓄積したデータベースである。素粒子分野では素粒子の性質についてさまざまの実験結果を整理し最新の状況を報告する活動が Particle Data Group と呼ばれるグループにより行われて来ている。インターネットの発達以来、取りまとめ結果は web で常時公開されるようになり、研究者にとって重要なデータの蓄積拠点であると同時に、素粒子に関する研究状況をも世界に発信するサイトとなっている。図 4 に PDG のトップページを示す。

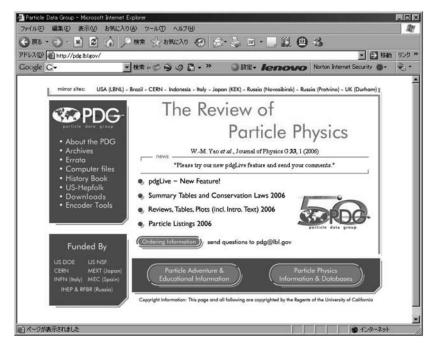


図 4 PDG (Particle Data Group) http://pdg.1b1.gov/ トップページ

## § 4. これからの学術情報

以上の説明に明らかなように、素粒子物理学分野では、学術論文や研究資料の作成、伝達、蓄積、検索といった研究上必要とする作業の多くが図書館の外で行われるようになって外しい。しかしながら、研究発表の最終段階としての学術雑誌への投稿はやはり必要とされ、出版された学術論文は電子 journal という形態ではあれ、図書館によって管理保管されているので、この点が素粒子物理学分野の研究者と図書館との主要な接点となっている。

プレプリントを集積した e-archive があるにも関わらず、何故学術論文が滅びないのかと聞かれることがしばしばある。それに対する答えは、「発表された研究成果がその結果を得る方法を含めて科学的であり、正しいものである」とのコミュニティとしての合意を形成する場が必要だからであろう、ということである。インターネットは、基本的に情報を提供する側が主役の舞台である。逆に、それが正しいものであるのかどうか、従来知られていた以上の新しい結果であるのかどうかは、情報を摂取する側に判断が委ねられている。学術雑誌は、レフェリー制度という仕組みを通じて、研究者コミュニティが新たな成果に対して集団として判断を下し、学問を育てていく場として位置づけを保っている。

それでは図書館はどうだろうか?勿論、今までに蓄積された学術雑誌や文献資料の数々、さらには膨大な数の電子 journal や、様々のデータベースを整理して提供する窓口としての図書館は、学生、研究者を問わず重要な機能を提供していることは疑いない。しかし、一昔前のように、学術雑誌と書籍が知を代表していた時代とは異なり、「知の集積拠点」としての図書館の役割が変容していることは明らかである。

知は体系化し整理することによってはじめて力を発揮する。ばらばらに寄せ集めたままでは意味はない。インターネット上の情報爆発は一言で言えば後者の状況と言える。昔の図書館が、書籍や学術雑誌の形態で、分類された知の集積を研究者や学生に提供して来たように、図書館の大学における役割は、今後も知を体系化し整理して提供することにあるように思える。

従来、このような作業は大学図書館を超えた枠組みで行われてきた部分がある。すなわち、大学における研究成果はそれぞれの分野の学術雑誌に発表され、大学図書館は、紙媒体に印刷された論文や書籍を、ある分類に従って書架に並べて提供してきたのである。このような役割は、電子 journal や e-archive の出現により大きく重要度を低下した。

これからの大学は、そのような迂回路を経ることなく、自らの研究者群がどのような研究を行い、どのような成果を挙げているのか、見やすく検索しやすい形で大学の外に提示できる必要がある。リポジトリの名称の下にそのような努力は既に始まっているが、知を体系化し整理してインターネット上で提示する方法論とその実現はこれからの大学図書館にとって重要な役割となっていくのではないだろうか。