

リポジトリタイプの比較

学術コミュニケーションへの奉仕における主題リポジトリ、研究リポジトリ、 国営リポジトリシステム、機関リポジトリの課題と障害

Chris Armbruster, マックス・プランク協会(マックス・プランク・デジタルライブラリ)

Invalidenstrasse 35, D-10115 Berlin - www.mpdl.mpg.de

Executive Director, Research Network 1989 - www.cee-socialscience.net/1989

Laurent Romary, INRIA-Gemo & ベルリンフンボルト大学(ドイツ語・言語学研究所)

Dorotheenstrasse 24, D-10117 Berlin - www.linguistik.hu-berlin.de/

概要

過去20年間にわたるリポジトリの発展から、デジタル学術コミュニケーションのためには、ひいては新しい知識の産出のためには、どのタイプのリポジトリとどのようなサービスが最良であるかという問題について、結論を出すことができるように思われる。出版物リポジトリは次の4つに分けられる。すなわち、主題リポジトリ、研究リポジトリ、国営リポジトリシステム、機関リポジトリである。リポジトリの役割においては、2つの重大な転換が起こっていることに留意すべきだろう。コンテンツについては、高品質で明確な輪郭を所蔵物が不可欠である。このことが意味しているのは、リポジトリサービスは、ユーザーと読者を第一に念頭に置いて構築されたものが最も成功するということである。サービスに関しては、専門分野の研究者コミュニティに重きを置くことが不可欠である。これが意味するところは、研究者にとって最も有用なリポジトリとは、新しい知識の産出を支援する専用サービスを提供するものだということである。

以上のことに付随して、リポジトリの発展に関わる課題と障害も示すことができるだろう。それは次の3つである。(a)コンテンツの特定とデポジット、(b)アクセスとサービスの利用、(c)コンテンツの保存とサービスの持続。結論では、世界のいくつかの代表的な地域、ヨーロッパ、北米、東アジア、オーストラリアにおける課題と障害の比較を示す。

キーワード

リポジトリ基盤、学術コミュニケーション、科学情報、主題リポジトリ、研究リポジトリ、国営リポジトリシステム、機関リポジトリ、デポジットの義務化、オープンアクセス

取り上げられているリポジトリとサービス

ARROW, arXiv, BASE, CiteSeer X, DRIVER, HAL, HighWire Press, Japan DRF, MIT OCW, NARCIS, OAIster, PMC, PMC Canada, Recolecta, RePEc, SSRN, UK PMC, Webometrics

Ranking of Repositories

過去20年間のデジタル化の普及によって、我々はさまざまなリポジトリソリューションの発達を目の当たりにし、役に立つものとそうでないものについて、理解を深めてきた。主なリポジトリソリューションは以下のように分けられる。

- ・ 通例、主題リポジトリ(営利的/非営利的、単体/連携型)は、あるコミュニティのメンバーによって始められ、より広範囲に受け入れられるようになったものである。このようリポジトリと同様に、自主的なセルフアーカイブも研究者の内発的な有用性ゆえに広まった。著者にとって内発的に有用だというのは、ワーキングペーパーやプレプリントの形で着想や結果を早期に公けにする機会を作ることにより、さまざまな利益が得られるからである。それは例えば、先行性を主張したり、着想や結果を考査したり、提出に先立って出版物を改善したり、認知度を高めたり、国際的に注目を集めることなどである。このように、主題リポジトリはテーマが明確で、アラートサービスや利用統計がコミュニティユーザーに有益である。
- ・ 通例、研究リポジトリは、研究助成機関や研究成果の取得を意図する運営母体による資金援助を受けている。研究成果の取得は普通、デポジットの義務化という形でなされる。書籍も含めた出版物だけでなく、データもまた取得すべき成果と考えられるので、コレクションの内容はバラエティに富んだものとなる。これらの内容が科学の記録となるのであるから、デポジットと保存の基準は厳格でなければならない。リポジトリの後援者は、記録機能とデポジット義務とを結びつけるだろう。例としては、助成を受けた研究者が助成機関に報告書を提出するとか、年次報告書に研究成果を掲載するということである。研究リポジトリは、高品質のアウトプットを所蔵するであろう。なぜなら、そのコンテンツは複数回の査読を経たもの(例えば、研究助成申請書、ジャーナル掲載論文、研究評価書)であり、十分な資金援助を受けて産出された成果だからである。研究協力者や競争者や新たな研究プロジェクトを構想している人がユーザーならば、関心領域のコレクションを発見するだろう。
- ・ 国営リポジトリシステムには協調体制が必要である。連携型ではより一層そうであるが、単体であればそれほどではないけれども。国営システムは、より広範囲に学術的なアウトプットを取得できるように企画されており、単に記録の保存のみならず、高等教育の教授や学習といった支援の観点も取り入れている。実際、国益に叶うことのみが、国家的投資の正当化の根拠たりえるのである。このようなシステムであるので、学術的アウトプットは当該国の国語で表示される傾向があり、高名な学者の出版物に重きが置かれたり、学位論文の記録のシステムが開発されたりする。こうした国営システムは、当該国の国語で交わされる学術コミュニケーションに裨益する国立の研究図書館の一環と見なすことができるだろう。それは、国際的にはその国のアウトプットのショーケースであり、当該国の公共政策、すなわち高等教育や知識に対するパブリックアクセスなどを支えるものでもあるだろう。
- ・ 機関リポジトリは、当該機関の多岐にわたるアウトプットを所蔵するものである。そのようなアウトプットのなかでも研究成果が重要ではあるが、資格取得論文や教育・学習教材もまた重

要である。機関リポジトリがあらゆるアウトプットを取得するならば、それは図書館であり、ショーケースでもある。コレクションを所蔵しているから図書館であり、オンラインでコレクションをオープンアクセスで公開して、例えば当該機関の同窓生や研究仲間と繋がりを持つという意味でショーケースである。機関リポジトリはまた、内外部の評価や戦略の企画の際の道具にもなりうる。加えて、機関リポジトリは、地域の発展においても重要な役割を果たすことができるものである。企業や公共機関、市民団体などがリポジトリを利用すれば、その地域にどんな専門家がいるのかが直ちにわかるからである。

上述のような4つのリポジトリタイプの理念型は、実際のリポジトリの歴史と論理的思考から生まれたものである。そこには、学術コミュニケーション、オープンアクセス、リポジトリに関連した文献から得たものが含まれている。だが、これから展開する議論は、文芸批評ではなく、抽象的な理念と個々の事例との間を行きつ戻りつする内容となる。¹ 理念型タイプは分類と混同されるべきではない。分類では、各リポジトリがそれぞれの属性に従ってカテゴリ分けされるが、この場合はそうではない。むしろ理念型タイプを作る目的は、リポジトリについての理解を深め、その発展の分析ツールを提供するところにある。

いずれかの理念型タイプにぴったり当てはまるように思える出版リポジトリがある。伝統のあるリポジトリのいくつか、例えば、arXivやRePEc² は、とりあえず主題リポジトリと呼べるだろう。他のタイプの特徴をほとんど備えていないからである。しかし、より興味深い問題として、では、他のタイプの要素がどれほど小さなものなのかということ、それが何を意味するかということがある。ArXivを例にとれば、これは高エネルギー物理学分野の発展と成功に関わる重要な研究を支援している機関の研究リポジトリでもある。これに比して、RePEcはいつそう機関的な要素が大きい。このリポジトリは多様な機関や施設からのインプットやサービスに依存する連携システムの一環だからである。

別の例を挙げれば、PubMed Central (PMC)は、一見すると主題リポジトリのように見える。しかしながら、そのコンテンツの取得は、(NIHの)助成を受けた研究のアウトプットを所蔵する研究リポジトリであると宣言された後に開始された。さらに、米国議会がデポジットの義務化を法律として制定したので、PMCは国営リポジトリとなった。これと同様なりポジトリが、英国(UK PMC)やカナダ(Canada PMC)に、統合的な形とはいえ作られたことは、それゆえ驚くべきことではない。このように、上記の理念型タイプをPMCの発展の分析に用いることは有益であろうし、また恐らくは、例え

¹ Charles W. Bailey 編の総合的かつ最新の文献案内がある。以下を参照。Scholarly Electronic Publishing Bibliography, The Open Access Bibliography, and the Institutional Repository Bibliography.

参考文献情報は以下にアクセスのこと。 <http://www.digital-scholarship.org/>

² <http://arxiv.org/>; <http://repec.org/>

ばヨーロッパPMC³ の創造のような、将来のPMCの可能性を論じる際にも意味があるだろう。

国家事業として乗り出す例は次第に増えている。(形式としては地域主体ということもあるだろう)。しかし、研究のアウトプットを重要視するのか、機関を重要視するのか、どちらかということでは、国によってさまざまに異なる。フランスのHALシステムは、国営の有数の研究組織であるCNRS (フランス国立科学研究センター)が全面的に後援しているので、研究成果の普及に強い。日本では、国立情報学研究所(NII)が87機関の加盟するデジタルリポジトリ連合を後援している。主に図書館員がそのシステムの運営にあたっている。スペインでは、アグリゲータで検索ポータルであるRecolecta が数ある機関リポジトリの頂点に位置し、多様な著作物を取り揃えている。オーストラリアでは、機関リポジトリは何をおいてもまず国による研究評価活動に直結しており、査読済論文が重視されている。⁴

リポジトリは、インターネットと共に進化してきた。ゆえに、その特徴は公開性にある。すなわち、オープンソースの基盤の開発とコンテンツのオープンアクセスである。だが、現行のリポジトリにとって出費が許されるのは、それが学者達、言い換えれば研究者や講師によって受け入れられ、利用される場合にのみ限られる。それゆえリポジトリは、出版やデータの収集などにおいて、他の形態の径路と競合することになる。だから、コンテンツとサービスは、相互に価値を高めあうような形で組み合わせられるべきであり、それは書齋や実験室や教室でなされているがごとくに、新たな知の創造へと向かう学者を支えるものでなければならない。

本稿では、リポジトリがどのような形で学術コミュニケーションに奉仕するのが最良かということを考察するために、4つの理念型タイプの差異を利用する。⁵ ものの道理として、リポジトリは多くの機能を備えるにしても、まず何よりも学術コミュニケーションに役立つものでない限り、学術コミュニティに受容され、かつ長期にわたって利用されるようにはならないであろう。学術コミュニティに

³ <http://www.ncbi.nlm.nih.gov/pmc/>;

<http://ukpmc.ac.uk/>; <http://ukpmc.blogspot.com/2009/07/pmccanada-will-launch-in-autumn-2009.html>; on Europe PMC については Wellcome Library, The Year in Review (2008), p. 13 を参照。
- <http://library.wellcome.ac.uk/assets/wtx055651.pdf>

⁴ <http://hal.archives-ouvertes.fr/>; <http://drf.lib.hokudai.ac.jp/drf/>; <http://www.recolecta.net>;
オーストラリアについては、Kennan, Mary Anne, and Danny A. Kingsley (2009) 'The State of the Nation: A Snapshot of Australian Institutional Repositories', *First Monday* 14(2) を参照。

⁵ 本稿は前論稿において試みた、機関リポジトリとより大規模な解決策の比較考察をさらに進め、詳述するものである。以下を参照。Romary, Laurent and Chris Armbruster (2010) Beyond Institutional Repositories. *International Journal of Digital Library Systems* 1(1), forthcoming; available at SSRN: <http://ssrn.com/abstract=1425692>

よる受容と利用は、存続に関わる大問題である。サービスの改善のための課題と障害を特定し、どのタイプのリポジトリが、またどのようなサービスが将来的に必要とされるかを考察する。次のように議論を進めたい。第一に、デジタル学術コミュニケーションにおける2つの重大な転換とリポジトリに対するその影響を分析する。とりわけ、(a)ユーザーにとって関連性があり、興味深く重要な情報を入手できるような形で、増大する一方の出版物をいかに組織化するという課題、(b)特定の学術コミュニティに対する高度に有用なサービスをする必要について。次に、リポジトリの発展に立ちはだかる課題と障害を次の3つの局面において論じる。(a)コンテンツの特定とデポジット、(b)アクセスとサービスの利用、(c)コンテンツの保存とサービスの持続可能性。結論部では、世界の主要地域におけるリポジトリの課題への取組と障害の克服の努力について、比較検討を行なう。

量から質へ：ユーザーと読者を優先すること

1世紀以上前から、研究者・ジャーナル・出版物の数は確実に増えてきた。留意すべきは、出版された知識がかつてないほどの速さで増大していることである。対応策の一つに、専門分野の特化があるが、ユーザーの読む範囲がより限定的になる。別の対応策は、読み物の選択性を高めることで、例えば、ジャーナルのインパクトファクターを見て何を読むか決めるといったことだ。しかし一方で、科学情報を探す読者を、その関心領域、興味、クオリティ(例としては、article-level metrics やテキストマイニング)に従って、ナビゲートするサービスの開発にいそむステークホルダーもいる。

インターネットによって、出版される知識は留まることを知らずに増大しつづける。かたや、デジタル化によって、過去に出版された知識も同時に入手可能になっていく。加えて、ワーキングペーパーやコンファレンスプレゼンテーション、補完的データ、教育教材などの付加的な資料も、オンラインで発信されている。それでもなお今までは、増え続ける出版物の費用を賄うことが可能であった。財政的にそれが許されなくなるとしたら、知の氷山の頂上の一隅しか出版されなくなるという事態も起こりえるが、一方で知識量の増加を可能にする効率的な技術刷新は進んでしまっている。このような意味で、コンテンツへのナビゲートにおいてユーザーや読者の立場を優先することは、肝要であるという以上に、早急に求められていると言ってよい。リポジトリはどのような対応策をとってきたのだろうか。

主題リポジトリは、特定の主題カテゴリ(または研究領域)に関心を持つユーザーが購読するかもしれない配信サービスの運用実績を持っている。新規掲載論文やインパクト統計を知らせるアラートサービスが、主題別に総合的かつ専門的に配信されている。これらのサービスは出版社と比べるとより総合的である。というのは、出版社のサービスはジャーナルのタイトルしか網羅していないからだ。あるリポジトリが一つあるいは複数の主題カテゴリを扱っている場合、それはワンストップショップになりえる。そうなれば、出版社が新しい出版物の書誌情報をそこにフィードするよう

になるかもしれない。対照的に、機関リポジトリは読者に対してほとんどまったく何のサービスもしていない。大半の機関リポジトリは、読者がアグリゲータか汎用的な検索エンジン経由で検索してたどり着く方法に依存しているのである。ハーベストや連携検索などの努力も、今のところ、明確な輪郭をもった所蔵物にアクセスできるようなポータル構築にまで至っていない。

国営リポジトリシステムは、コレクションがある程度の大きさに達したら（そして、投稿率が十分に上がった）、必要とする人々に対して整備されたサービスを提供することができるだろう。例えば、フランスのシステムHALは専門を越えて論文や書誌情報の購読サービスを提供している。2008年には、年間投稿本数は5万を超えた（ちなみに、arXivも2008年に5万を超えている。）対照的に、オランダのリポジトリネットワーク(DAREnet)に連携している国営ポータルNARCISは、このような購読サービスを展開していない(Agricultural Repository News Exchangeのような過去のプロジェクトはあったが、こうしたサービスは採用されなかった)。DARnetはもともとそれ自体、(オランダの大学の)機関リポジトリのネットワークとして構想されたものであるが、NARCISは、いまや国営ポータルである。HALは、大規模なフランスの研究組織に支えられ、その主題リポジトリと連携協力してきた実績（例えば、コンテンツの相互交換など）がある。⁶特に、研究リポジトリは、それが主題リポジトリや国営システムの一部である場合（あるいは、そのように位置づけられている場合）、サービス提供を行なっている（例として、HALのなかのCNRSやINRIA）。しかし、もし研究リポジトリが単一の研究機関にのみ提携しているものだとしたら、そのようなサービスの提供が意味をなすほどのクリティカルマスに達しない、または所蔵物が明確な輪郭を持つに至らないという可能性もある。

つまり、主題リポジトリと国営システムの場合には、それらが膨大な量の出版物のなかで研究者をうまくナビゲートすることができると思えるに足る実績がある。機関リポジトリはそうではない。それらは、国営システムか、あるいは主題ポータルのなかにも、またはその両者に組み入れられない限り、その構造的な編制ゆえに上述のようなサービスを提供することができなかつたし、これからもできないだろう。⁷研究リポジトリは、高品質なコンテンツを発信することで、主題リポジトリや国

⁶ <http://www.narcis.info/>; <http://hal.archives-ouvertes.fr/>

⁷ 本稿の主題は、リポジトリの比較であって、ことさらに機関リポジトリに対する批判をすることではない。とはいえ、批判の声が強くなっていることを紹介しておく。以下を参照。Salo, Dorothea (2008) 'Innkeeper at the Roach Motel', *Library Trends* 57:2; Basefsky, Stuart (2009) The End of Institutional Repositories and the Beginning of Social Academic Research Service: An Enhanced Role for Libraries. Available at <http://www.llrx.com/authors/1133>; Albanese, Andrew R. (2009) 'Institutional Repositories: Thinking Beyond the Box', *Library Journal* (March 1); Romary, Laurent and Chris Armbruster (2010) Beyond Institutional Repositories. *International Journal of Digital Library Systems* 1 (1), forthcoming; available at SSRN: <http://ssrn.com/abstract=1425692>

営システムの充実化を支援することができる。

高価値への転換: 専用サービス

学術コミュニケーションは、第一に知のさらなる進歩に貢献するものである。次に重要な機能は、次世代の訓練である。しかしながら、学術コミュニケーションのシステムはまた、別の目的のためにも利用されうる。例えば、教育と学習である。リポジトリは、これらの多様な種類の知の生産（および、その観察、例えば研究評価）を支えるために、サービスをどのように多角化させねばならぬのだろうか。どのタイプのリポジトリがどんな専用サービスを提供するだろうか。

主題リポジトリへの投稿本数は飛躍的に増大している（一ヶ月に数千本、一年で5万本が普通である）。その結果、読者向けの適切なサービスの保持と品質の維持に多くの労力を費やすことになっている。解決策はさまざま。例えば、arXivはどの研究分野でも初投稿の著者には、すでに実績のある著者の推薦を受けることを義務付けている。SSRNは、研究分野の数を増やし、沢山の学術編集者がアラートサービスに載せるべきものを選び分けている。また、利用率と引用指数の拡大に向けた努力もしている。⁸ 一般的に、主題リポジトリは専門サービスを提供する大規模なプロバイダーになるという課題の克服に成功した。そのような専門サービスとは、あらゆる分野の著名な研究者を含む学術コミュニティに対し、関心領域に関連した重要な情報を配信するサービスということである。

対照的に、機関リポジトリは、どのような専門分野のコミュニティに対しても専門サービスを提供していない。これはつまり、それら自身では、学術コミュニケーションに何らかの意味ある機能を果たすことがないということである。せいぜい良くて、次第に盛んになってきた評価活動に貢献することぐらいである。例えば、機関リポジトリは当該機関や部門のアセスメントで役立つことなどである。この局面では、デポジットの義務化という機関のポリシーは意味がある。次に有用な機能としては、教育・学習教材の活用、とりわけ遠隔地における利用である。もし出版物やコース用教材が（また、ブログのような他のオンラインツールも）連結されれば、全世界で利用可能な形で配信できる。けれども、いくつかの名門大学のサービス（例えば、MIT OpenCourseWare のような）しか、多数のユーザーを獲得しそうにない。⁹ もう一つの機能は、機関リポジトリは研究者が著作物をデポジットする手伝いをするということである。図書館員や運営担当者が投稿物の質をアセスメントに適合するように、あるいは教育・学習教材として使用可能なものになるように保全することができるだろう。最後に、機関リポジトリが将来とりうる道として、新しいコンテンツ（例えば、RePEc モデルを他の分野でも採用することなど）を提供することで、特化した学術コミュニティ（またはポータル）の支援をしてゆくということがあるだろう。

⁸ <http://ssrn.com/>; <http://arxiv.org/>

⁹ <http://ocw.mit.edu/>

研究リポジトリは、研究の最先端を研究者（そして、他の利害関係者、例えば研究助成者など）が、追跡できるようにするという（潜在的に）最重要の機能を担っている。しかしこのため、単に科学の記録を保管するというのと、最新のコミュニケーションツールであるということの間に緊張が生じるのを免れない。もし研究リポジトリが主に最終的な研究のアウトプットを保存するというのであれば、それは査読済出版物であるから、公式な出版を経た後にしか所蔵されず、つまり出版社が設ける配信停止期間が終了してから後ということになり、それが実質的には科学の公式な記録ということになる。だが、最初に採用された研究助成申請書（その判定は最先端のもの）とその研究成果がリポジトリで公開されて読めるようになるまでの期間がとにかく長すぎる。これでは、リポジトリは学術コミュニケーションにおいて有益な機能を果たすことはできない。求められていることは、リポジトリが研究情報サービスをも兼ねることだろう。例えば、Grant-Lookupツールや、早期の結果の表示や研究動向の追跡を可能にするような新しいツールを搭載したサービスである。特に、UK PMCはこのようなサービスを展開している。UK PMC が主題リポジトリでもあることを考えると、このようなサービスは汎用的にすべてのリポジトリ（単一の分野のものでも、複数の分野にわたるものでも）が採用することは可能ではないだろうか。¹⁰

国営リポジトリシステムは、非常にたくさんの表示を提供していて、それらは機関別であったり、主題や出版物の種類別（学位請求論文等）であったりする。しかし、主題リポジトリや研究リポジトリに匹敵するような専門サービスを開発しているようには思えない。問題は、そのようなサービスの開発には、クリティカルマスを超えている十分に明確な輪郭を持った所蔵物の実体が不可欠だということだ。かなり広範囲にわたるコレクションポリシーをとっている国営のシステムであるとしたら、サービスが実施されるには、第一に利用可能な著作物が大量に存在すること、かつデポジット数が継続的に高いことが必要である。HALの年間デポジット数が5万に達しているならば、フランスのシステムは、これまで以上のサービスを開始する勇気があって然るべきである。さらに、オランダのコレクション（NARCIS）は、18万以上の著作物を所蔵しており、オーストラリアの諸大学は20万本以上の所蔵を誇っている（もっとも、フルテキストは3万本ほどである）。スペインのアグリゲータ（Recolecta）は45万本以上を所蔵している（フルテキストの数は不明）。¹¹ こうしたことを見ると、さらなるサービスは可能であろう。

上述の議論は、ユーザーと読者に焦点を当てたものであり、とりわけ読者としてのユーザーに着目したものであった。だが、著者もやはりリポジトリのユーザーである。例えば、著作をデポジットした人は、その著作のインパクトを追跡するだろう。とりわけデポジットが大規模に組織化されている場合には、彼らのニーズに対しても専用サービスが求められるだろう。例えば、経験者がデポジ

¹⁰ <http://ukpmc.ac.uk/grantLookup/>; http://ukpmc.ac.uk/ppmc-localhtml/future_plans.html

¹¹ <http://www.narcis.info/>; Kennan, Mary Anne, and Danny A. Kingsley. "The State of the Nation: A Snapshot of Australian Institutional Repositories." *First Monday* 14, no. 2 (2009).; <http://www.recolecta.net/buscador/results.jsp>

ットに支援するとか、第三者によるデポジット(例としては、出版者がリポジトリにデポジットする)などが考えられる。実際、この問題は、リポジトリの発展に関わる第一の課題であり障害なのである。

特定とデポジットにおける課題と障害

リポジトリは書棚なしのソフトウェアであり、多くの人々ができるだけ多くのリポジトリを作りネットワーク化することが進むべき正しい道なのだと信じてきた。だが、驚くほど多くのリポジトリが最初のハードルの前ですでに立ち往生しているのだ。それは、学術コミュニケーションにおいて関連と関心の高い著作物の特定とデポジットである。リポジトリ基盤が良く整備されている国々(英国、オランダ、ドイツ、オーストラリア等)¹²でも、デポジット率は低く、また押しなべてリポジトリが明確な輪郭を持ってはいない。唯一の例外は、オランダのプロジェクトCream of Science (DAREnet) で、オランダで最も重要な200人の研究者の大半の論文をオンライン上に掲載している。とはいえ、その大部分は古典ともいべきもので、主たる価値は歴史的なものではあるが。

特定とデポジットは別物であるが、リポジトリは学者にとって価値を生み出すために重要なのはどの著作物を特定しなければならない。ところが、多くのリポジトリにとってデポジットの義務化は一番の希望に留まってしまっている。対照的に、主題リポジトリはコンテンツの任意投稿を受け取り続けている。成功するリポジトリとは、サービスとセルフアーカイブ、毎月何千もの投稿、(arXiv, RePEc, SSRN)、通常はプレプリントから出版物のポストプリントという流れの良循環を作り上げているものだ。デポジットの義務化は機能的に相当するものであるが、むしろ重視されているのは、著者の最終版査読済原稿である。

デポジットの義務化は、研究助成者や著名な機関にとっては効果がある。なぜならば、コンテンツが一般に高い関連性をもつものであるばかりでなく、クオリティも高く、ユーザーの興味を刺激するからである。このことから次のような考察が導き出される。

- デポジットの義務化は望ましいコンテンツの特定に役立つ。それは主に査読済出版物である。
- 義務化は研究者の遵守を求め、管理を必要とする。その意味で、このタイプのデポジットの義務化はセルフアーカイブと異なる。
- 機関リポジトリは、デポジットの義務化の対象が主に研究成果である限りにおいて、その性格を変えることがあったのかもしれない。

デポジットの義務化が完璧な解決策というわけではない。所蔵物を集積する機会であり、質の高

¹² e.g. <http://www.sherpa.ac.uk/>; <http://www.rsp.ac.uk/>; <http://www.wrn.aber.ac.uk/en/>; <http://www.narcis.info/>; <http://www.dini.de/wiss-publizieren/>; <http://www.arrow.edu.au/>;

いコンテンツの特定に有益だというに過ぎない。デポジットは決して小さな問題ではない。経験を積んだ学者や業績の多い著者は多忙な人々だ。リポジトリが義務化の規則の遵守に依存しているなら、支援が必要だし、可能なら自動的なデポジットが求められている。例えば、UK PMCIは、デポジットされる版が最終版でフルメタデータを載せている所蔵版である場合には、出版社が著作物の取得に応じて、直接デポジットする方が容易になることを見出した。¹³ オープンアクセス出版の場合、出版社によるデポジットは普通、問題にはならない。伝統的な形態の出版社の場合、ことがらはより複雑になる。こうした出版社は配信停止期間が守られることを声高に主張しているだけでなく、デポジットされる版も最終版ではなく、著者の最終査読済原稿なのである。つまり、それには出版版にはなされている頁付けや編集がなされていない。それゆえ、デポジットの義務化が高品質のコンテンツをもたらすとはいえ、それは権威があると見なされうるやり方で(まだ)なされているとは言えないのである。この意味で、出版物リポジトリは、出版版が入手可能であり配信停止期間も守られていて、より高い価値を有している。例としては、HighWire Press がある。これは多くのジャーナルがオープンアクセスで公開する前に12ヶ月の配信停止期間を設けているプラットフォームである(いくつかのジャーナルは、配信停止期間を3ヶ月から6ヶ月としているし、ムービングウォールを数年間としているものもある)。¹⁴ 現在のところ、190万本の論文を読むことができる。HighWire Pressは小規模のSTM誌出版者、しばしば学協会の出版者でその多くは生命科学を専門とするところに好まれている。出版者のリポジトリは、とりわけそれがコミュニティ志向である場合、デポジットの義務化にコンテンツの獲得を依存しなければならないリポジトリにとっては、興味深い選択肢となるのではないだろうか。

さらに、いわゆる著者の最終原稿のアーカイビングにも支援が望ましい。例えば、出版社による直接デポジットによってプロセスは単純化される。そうでなければ、出版社は出版前の最終原稿を著者にデポジットのために返却することになる。もちろん、このシナリオでは出版社は配信停止期間を定めていることとする。加えて、最終査読済原稿の管理された版を返却するにあたってのサービス料も生じるかもしれない。このタイプの解決策は、デポジットの義務化に頼ることができない機関リポジトリにとっては魅力的であろう。これは、大きなSTM誌と選ばれた機関リポジトリによって連携プロジェクトとして試みられている最中である(Publishing and the Ecology of European Research - PEER - 欧州連合による共同助成を受けている)。¹⁵ 大量の査読済論文がオープンアクセスアーカイブによって読めるようになっている。そのうち半分は出版社からの直接デポジットによるものであり、後の半分はセルフアーカイブの許可をつけて著者に返却されたものである。我々の期待としては、セルフアーカイビングがまばらになること、つまり、支援を受けたデポジット、例えば図書館員による組織的な支援を受けたデポジットが増えて、オープンアクセスデポジットが系統的になされるようになることである。

¹³ <http://ukpmc.ac.uk/ppmc-localhtml/about.html>

¹⁴ <http://highwire.stanford.edu/about/>;

¹⁵ <http://www.peerproject.eu/>

規模、質、サービスが問題である。大規模な主題リポジトリは、当該分野のトップ研究者からコンテンツを取得できるため、高品質なコンテンツの自主的なデポジットにほとんど困っていない。それ以外では、デポジットの義務化は研究助成者や研究組織(国立アカデミー等)、名門大学によって実施されるならば、効果があるものである。対照的に、国営システムはコレクションポリシーに裏打ちされ計算された戦略や、主題ベースのサービス、そしてより好ましくは研究助成者や国立機関による何らかの義務化なしには、難しい。さらに、機関リポジトリとなると、一つ以上の他のタイプのリポジトリ(国営システム、主題ベースサービスのフィード等)に統合されていること、あるいは、最終査読済原稿か最終出版版をデポジットしている出版社との提携体制に依存しているように見える。しかし、もしアクセスが配信停止期間を守ってなされるならば、出版社のリポジトリも権威のあるコンテンツを備えたワンストップショップとしてより効率的なソリューションとなりえるかもしれない。

アクセスと利用における課題と障害

著作物がデポジットされたからといって、それをユーザーが発見して利用すると考えるのは早計である。よく知られていることだが、リポジトリにとって検索と発見は課題なのである。なぜなら、専門的な検索サービスがカバーしている範囲は限られており、汎用的な検索エンジンではリポジトリの所蔵物はしばしばリストの下のほうにしか表れないからである。明らかに、リポジトリのコンテンツの質と視認性がアクセスの鍵を握っている。

アクセス数の外部統計は、Spanish Research Council (CSIC) が運営するCybermetrics Lab によるRanking Web of World Repositoriesで得られる。¹⁶ このランキングは、リポジトリのコンテンツの視認性と質(バックワードリンク、フルテキスト論文の数、学術的な質の高さ等)について良質の情報を与えてくれる。総じて、かなりの数のリポジトリがコンテンツの視認性を高めること、とりわけ検索エンジンで表示されることにおいて難題を抱えている。非常にしばしば、リッチファイル(テキスト、データ、写真)は見つかりづらい。なぜならそれらには、例えばCiteSeer、PubMed Central、RePEc でのように、標準の拡張子がついていないからである。これ以上の大規模な主題リポジトリは高くランキングされている。HALは、その機関領域とともに国営システムとして高く評価されている。機関リポジトリについては、機関そのものの学問的名声とリポジトリのランクには何の関連性もない。例えば、多くの名門大学や研究機関は高くランキングされているリポジトリを持っていない(顕著な例外は、恐らくカリフォルニア大学とマサチューセッツ工科大学だろう)。

アクセス数の内部統計は、利用率と引用統計から得られる。利用率はある著作物がどのくらいの頻度で見られ、ダウンロードされたかを示す。利用統計はあらゆるリポジトリで収集されるだろう(とはいえ、真の数字は数多くのソースを集合しないと得られない。例えば、出版社、リポジトリ、仲

¹⁶ <http://repositories.webometrics.info/>;

介サイトやキャッシュなど)。¹⁷ 引用統計は、領域を超えて参照箇所を追跡する必要がある。それには、どれを引用と見なすかという判断も含まれる。しかし、明確な輪郭を備えたリポジトリでさえも(例えば、arXiv、SSRN、RePEc等)、自身の所蔵物からのみでは信頼するに足る引用指数を出すことは難しい(ハーベストされたものでも、連携している場合にも、中央化されている場合でも関係ない)。今のところ、唯一の可能な解決法は、Google Scholarから引用統計を得ることである。¹⁸ 大規模な主題リポジトリが所蔵物を集積すれば、そこから有意義な統計やランキングが得られるかもしれない。だが、他のタイプのリポジトリでは研究者に価値のある情報サービスとしての利用統計は直接には得られないだろう。そのためには、国際的な協力体制が必要である。利用状況については、諸々のリポジトリから信頼され、利用状況を客観的にカウントし比較する独立主体が必要である(例えば、著作物ごとに測定するProject COUNTER に基づくなど)。¹⁹ 引用サービスはそれ自体で成立するが、ただそれらが大規模で意義のある所蔵物について横断的になされた場合にのみ独立したものになりえるだろう。例えば非常に大きな主題リポジトリや国営システムなどの場合である。現在のところ、RePEのみが意義のある引用統計を持っている。²⁰ 対照的に、出版社は共同の基準を作り、テクノロジーを施行して、ユーザーが引用箇所を追跡したりサーフできるようにしている(CrossrefやWeb of Knowledge、Scopusなどのデータベース等)。²¹ リポジトリのために、国営リポジトリポータルやアグリゲータのいくつか(スペインの Recolectaやthe Bielefeld Academic Search Engine、OAisterなど)²² は、引用の追跡、サーフィン、集計ができるようなサービスの構築を開始できるだろう。

リポジトリのユーザーは次第に、コンテンツ・ダウンロード型に代わるアクセスサービスを期待し始めている。例えば、リポジトリのコンテンツの表示を多角化して、ユーザーの関心と関連コンテンツの発見につながるようにすることは重要である。しかしながら、これはかなり広がりを持つ話になる。例えば、著作物を一覧するのに、新規性やランキング(全分野、あるいは特定の分野)を勘案して見たいとか、ある一人の著者とか研究チームが投稿したものをすべて見たいとか、ある機関や部門の挙げた成果を評価したいとか。リポジトリにおける表示の多角性の必要は、研究者、講師、学生、査読者、学部管理者、研究助成者などといったユーザーの多様性にも連動している。別の例では、より進んだユーザーなら、新たな知の創造のために、アクセスだけでなく、コンテンツのマイニングやリユースの権利も求めるかもしれない。これらのサービスの多くは、オープンアク

¹⁷ e.g. <http://ssrn.com/> (top papers, top authors, top institution)

¹⁸ e.g. <http://www.citebase.org/>; <http://ssrn.com/update/CiteReader.html>; <http://scholar.google.de/>

¹⁹ <http://www.projectcounter.org/>

²⁰ Armbruster, Chris (2008) Access, Usage and Citation Metrics: What Function for Digital Libraries and Repositories in Research Evaluation? *Online Currents* 22(5) 168-180; available at <http://ssrn.com/author=434782>

²¹ *ibid*

²² <http://www.recolecta.net/buscador/>; <http://base.ub.uni-bielefeld.de/>; <http://www.oclc.org/oaister/>

セスの「自由」を必要とする。それは、科学情報の再利用を可能にする法的なパラメータである。

配信停止期間もまた、アクセスと利用状況に対して影響をもたらす。いくつかの伝統的な主題リポジトリの強みは、それらが早期の投稿(ワーキングペーパーやプレプリント等)を惹きつけるところにある。しかし、義務化に依存しているリポジトリは配信停止期間を遵守しなければならない。それは通常6ヶ月であるが、それよりも長い場合もある。結果的に、情報が公開されるのはずっと後のことになる。プレプリントの段階から査読を経て、編集・出版まで時間がかかる。それに配信停止期間が加わるとなると、たとえそれが6ヶ月だとしても、公開まで少なくとも1年はかかることになる。悪ければ2年である(これは、最初の投稿の日付から数えて、査読、校正、出版までのことだ)。また、オンラインの第一の機能によって、多くのコンテンツが出版社のプラットフォーム上でジャーナルの公式発行日より前に入手可能になることも、配信停止期間の長期化を間接的に助長している。オープンアクセス出版は出版されたら直ちにアクセス可能になるという利点を備えている。出版された知識が新たな知識の探求に力を与えるものである限り、アクセスの時宜性は利用率を左右する。ワーキングペーパーやプレプリントの流通こそが何にもまして、アイデアやデータの早期公開の場の水準を決定する。もしリポジトリが配信停止期間を受け入れるならば、有料アクセス出版が一番アクセスの多いサイトであり続けるだろう。そして相変わらずアクセス権を持ってない人は不利なままである。なぜならアクセスできるようになるのは遅すぎるからだ。このシナリオでは、オープンアクセス出版は、研究者により多くの価値をもたらすことだろう。

持続性と保存における課題と障害

持続性はつまるところ、利用率とユーザーの満足度による。Directory of Open Access Repositoriesは、1500以上のリポジトリをリストにあげているが、そのうちの1100以上が機関リポジトリである。²³ 過去何年かの間増加は顕著ではあるが、いまだ大半の研究機関や高等教育機関はリポジトリを持っていない。とはいえ、運営能力はどんなリポジトリにとっても課題である。無制限で高速のアクセスを常に可能にするためには、強力なコンピュータ施設が必要であるし、とりわけリポジトリが成長しつつあり、アクセスがそれに伴って増えている場合にはそうである。加えて、ユーザーと著者のためのサービスも必要に応じて拡張せねばならない。特に、主題ベースであったり、国営システムとしてデザインされているならば、スケーラビリティとパフォーマンスの問題に直面する。一方で、機関リポジトリはデザインによって限定されている。しかしながら、小規模なりポジトリがそのように規模の経済性の恩恵を受けることができないのであれば、各機関のリポジトリは戦略に最もお金をかけるべきだということになるだろう。

主題リポジトリは、研究者からの貢献と支援を目に見える形で受けるほど十分に魅力的であることを誇示してきた。貢献と支援とは、時間とリソースの供与であり、例えば編集サービス等である。

²³ <http://www.openoar.org/>

対照的に、機関リポジトリは研究者に対してデポジットの勧誘にも苦労している有様である。持続性にはコスト管理が重要であるが、開設初期のリポジトリの場合は研究者に受容され利用されているということをまず広く知ってもらう必要がある。もちろん、機関リポジトリは学生の論文の所蔵庫という機能を常に担っているわけだが、さらなる発展のためのリソースの活用可能性は、それが学術コミュニケーションと知識の生産に対してどれほどの価値を持っているかを示すことにかかっているのである。ただ主題リポジトリだけが、とはいえそのすべてがというわけではないが、自他共に認める価値と受容を誇ってきた。国営システムと研究リポジトリは、この課題から脇にそれ、パブリックアクセスと経済的価値に焦点をあてようとしているようだ。それは重要なことだが、しかしそれによって学術コミュニティにおけるリポジトリのあり方が正当化されるわけではない。

コスト管理については、大規模なリポジトリは規模の経済性の恩恵を受けることができるはずだが、現状では、大規模な主題リポジトリのビジネスモデルは破滅的である。商業的な要素が強すぎるし、明らかに無償の労力の上に成り立っている。研究リポジトリと国営システムは、総研究費のなかで僅かな割合を占めるに過ぎないが、研究成果のオープンアクセスによる普及がパブリックアクセスを可能にし、より広範な影響を与えるという意味で正当化されうるだろう。機関リポジトリは独立型であるか、図書館のような部門の一部であったりするが、所属機関の財務健全性に大きく依存している。

もしリポジトリが長期にわたって価値を保ちたいと願うならば、品質管理システムは必ず実施されねばならないし、所蔵物の統一性も保持されねばならない。広く流布している誤解の一つにリポジトリは「なんでも」所蔵するところだというものがあるが、現実には、ユーザーの関心はより一層研究成果の関連性や価値へと向かっているのである。これは、博士論文を検索したところで面白い結果は得られないだろうなどと言いたいわけではなく、リポジトリのコレクションと展示のポリシーが、そのリポジトリが受容され利用されるかを大きく左右すると言いたいのである。研究リポジトリにとって、これは容易なことだ。というのは、成果は複数回の査読を経たものであるからだ。主題リポジトリの多くは、品質管理と標識を備えている。例えば、デポジットする者を管理し、著名な機関からのシリーズ物を揃え、利用状況を示す統計を提供する等である。機関リポジトリは通常、品質や関連性がばらばらな著作物を収集しているが、同時にコレクションポリシーによって、大学卒業論文のような研究出版物は明確に区別していることもある。

リポジトリにとって長期保存とアクセスは未解決の課題ではあるが、喫緊の課題というわけではない。今のところ、ほんの少数の主題リポジトリだけが保存に値する。もう少し視野を広げると、国営リポジトリシステムは、国立図書館や研究リポジトリのコンテンツが科学の記録として不可欠な価値を有すると認められる限りにおいて、そうであろう。もちろん、デジタル化保存の努力は続けられており、図書館員と出版社は、例えばCLOCKSSのようなコンテンツの保存に向けての連携事業を

構築している。とりわけ、コンテンツがもはや出版社側に存在していない場合がある。²⁴ リポジトリは持続性の問題を最優先にしなければならない。それはすなわちサービス、利用率、費用の問題である。これまでの議論が示すことは、すべてのタイプのリポジトリが、可能であれば提携や統合によって効率性を高めることで、いかにサービスを改善することができるのかを考えなければならないということである。

地域ごとの状況の比較

特にヨーロッパでは、リポジトリのネットワーク化に公的支援がかなり大規模になされており、国内ネットワーク化(オランダや英国)や、欧州の基盤整備、基準設定および実施コミュニティの推進(DRIVER I & II)もそうである。²⁵ ヨーロッパの状況を概観は、DRIVER サーチポータルが提供しており、‘リポジトリによるブラウザ’機能で閲覧できる。²⁶ これを見ると、大半のリポジトリがきわめて限定的なコンテンツしか保持し続けていないことがわかる。記録についてそうなのであるから、フルテキストアクセスは言わずもがなである。このことはヨーロッパにおける機関オープンアクセスの義務化の低さを考慮すれば驚くべきことではない。より重大なのは、欧州レベルのものも含む研究助成機関による義務化である。²⁷ ヨーロッパの多くの地域における不整合はポリシーとインフラストラクチャとの間にあると考えられる。インフラストラクチャは機関デポジットを条件に構築されているのに、ポリシーレベルでは、研究助成機関による義務化が多く、効果が大きいのである。ヨーロッパが課題として取り組むべきは、第一義的に機関リポジトリと位置づけられているすべてのリポジトリを、研究アウトプットを取得するものへと変容させることができるかどうかということである。特に、大半の国営ネットワークは大学リポジトリを呼び物にしているが、それはしばしば全ての大学の中で抜きん出た大学のリポジトリである場合が多く、すべての高等教育機関を網羅してはいない。このため、指標としては研究リポジトリのタイプへの転換が可能であると考えられるのではなかろうか。

Webometrics Ranking of Repositories に拠れば、成功しているリポジトリシステムはフランスの例である。²⁸ 尤も、これに追随している欧州諸国は多くない。フランスのシステムで顕著なことは、国営の研究組織がインフラ整備を主導し、大部分のコンテンツを寄贈していることである。他のヨーロッパ諸国が研究成果の取得に焦点を移せば、同じ土俵ができることになる。最良の組み合わせは、強力な国営システムがデポジットと保存を組織化し、サービスは専門的なコミュニティごとに特

²⁴ <http://www.clockss.org/>

²⁵ 的確な要約は、Vernooy-Gerritsen, Marjan, Gera Pronk, and Maurtis van der Graaf (2009) ‘Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe’, *Ariadne* 59 (April)

²⁶ <http://search.driver.research-infrastructures.eu/>

²⁷ ROARMAP at <http://www.eprints.org/openaccess/policysignup/>

²⁸ <http://repositories.webometrics.info/>

化すること、そして主題リポジトリと両立しうるものにする事だろう。

北米、特にアメリカ合衆国は成功している主題リポジトリのホームであり(arXiv、CiteSeer、PMC、RePEc、Smithsonian、NASA、ADS、SSRN 等)、またいくつかの例外的な機関リポジトリもある(イリノイ大学、カリフォルニア大学、マサチューセッツ工科大学)。ヨーロッパはリポジトリの本部を置くべき地として考えられていない(例えば、RePE はヨーロッパで構想された)。しかし、すべてに積極的に参加しているし、時には補助的なサービス(UK PMC や Europe PMC 等)も構築している。アメリカの研究機関が大学ランキングのなかで上位を独占しているとしても、リポジトリのランキングでは同様ではないことは注目すべきである。主題リポジトリが良いパフォーマンスをしているならば、機関のソリューションとしてはオープンアクセスデポジットの義務化がより魅力的なものになったり(例としては、ハーバード大学)、あるいはデジタルライブラリの構築のようなより専門化した目標を立てたり(例としては、カリフォルニア大学)、全世界的に教育・学習教材の流通を拡大させたり(例としては、MIT OCW)しているからだと思われる。

ヨーロッパとアメリカが主たるプレーヤーである。上述の分析により、これら2者の間で生まれるソリューションが全世界に拡張できるだろう。この2者は主題リポジトリと研究リポジトリのシステムであり、そのうえに様々な国営システム(フランスやオランダ)、機関システム(カリフォルニア大学、マックス・プランク協会など)、そしておそらく単一の名門研究機関(ハーバード大学など)が続くであろう。研究成果の取得に焦点をあて、サービスを通じての学術コミュニティの支援が、学術コミュニケーションにおけるリポジトリの広い受容と利用へと繋がるだろう。それは将来の持続性の鍵を握るものである。

一見したところ、そのようなソリューションは、リポジトリの状況のなかで良いランキングを得ている他のプレーヤー、すなわちオーストラリア、カナダ、日本の機関にも適用可能であるように思われる。²⁹ 事実上、日本とオーストラリアは国営システムを大学の研究成果の取得を基礎にして構築している。カナダは主題リポジトリ(PMC Canada)に参加し、かつ機関リポジトリを支援する二重の戦略をとっていて、カナダ研究図書館協会の5つの会員機関のうち4つがリポジトリを所有している。³⁰ 結局のところ、我々が提案してきた全世界的に適用しうるソリューションは国営システムと並存可能なものであり、そこにはコンテンツを当該国の国語で読めるようにする機能も含まれる。

我々は、全世界に拡張しうるリポジトリ基盤を構築することは第一に可能であるし、学術コミュニケーションにとって価値のあることだと結論づけることができるだろう。しかしまた、上述の議論が示

²⁹ In the Webometrics ranking of July 2009, the first Brazilian repository ranks 57 (University of Sao Paolo), the first Chinese one 88 (National Tsing Hua University), and the first Indian one at 121 (Indian Institute of Science Bangalore).

³⁰ http://www.carl-abrc.ca/projects/institutional_repositories/institutional_repositories-e.html

すように、何らかの画期的な変革も必要であり、とりわけ機関リポジトリのグレードアップや変化が必要である。また、リポジトリ間の提携やアクセスと利用率を高めるサービスプロバイダーとの連携も求められる。こうしたことが起こるのかどうかは予断を許さないが、リポジトリの理念型タイプとそれに続いて展開された議論は将来の発展の分析や考察に活用されうるだろう。