

英国における機関リポジトリ：グーグル利用者はそこで何を見つけることができるか (Margaret Markland)

Markland, Margaret. Institutional repositories in the UK: What can the Google user find there? *Journal of Librarianship and Information Science*, 38(4) December 2006, p.221-228.

本調査では、広範な主題分野を扱う 26 の英国の機関リポジトリから収録資料 (item) を検索する際のグーグル検索エンジンの効率を調査する。各リポジトリから収録資料を 1 つ選択し、4 つの検索を実施した。それは 2 つのキーワード検索と 2 つの詳細タイトル (full title) 検索で、それぞれの検索でグーグルとグーグル・スカラーの両方を使った。さらに、リポジトリの検索画面 (interface) から収録資料の検索試験を行った。検索結果の解釈では、既知情報探索行動 (known information-seeking behaviour) を考慮に入れた。これらは、グーグルとグーグル・スカラーの利用者が、特に詳細タイトルを知っている場合、リポジトリから収録資料を検索するであろうことを示している。しかしながら、検索した本文のさまざまなバージョンを評価するには、ある程度のスキルが必要とされるだろうし、より精緻なハーベスティング・サービスを介してリポジトリにアクセスすることが、学者にとってもっとよい代替手段であることを証明するかもしれない。

キーワード：情報検索，情報探索，機関リポジトリ，検索エンジン

はじめに

電子環境下では、印刷物の世界から移動してきた断片的なパッケージの代わりに、検索エンジンが情報への事実上のインターフェースになっている (Hitchcock et al., 2003: 255)。

この小規模な調査を遂行するアイデアは、二つの研究上の関心の融合結果である。第一は、合同情報システム委員会 (JISC) が助成した 2004 年から 2005 年にかけて、1 年間のプロジェクト評価へ参加し、その間英国および世界中で注目を浴びている学術コミュニケーションについて議論したことであり、第二は、学術研究用の情報資源の検索ツールとしてのグーグルの学生利用についての長年にわたる関心である (Markland, 2005)。

SHERPA は英国の機関リポジトリに特に焦点を当てており、機関がそのようなリポジトリの作成と存続 (populate) のために資源—時間、資金、努力—を投資し始めていることから考えると、それらがどのように使われているかについて点検するのに適切な時点であるように思われる。SHERPA プロジェクトに関する資料は多く出版されているが、本調査でとられた特別な観点は、学術コミュニティや実際にはより広いコミュニティにおける情報探索者で、それはプロジェクトの問題から離れている。英国の機関リポジトリにどのような種類のコンテンツが存在するのか。また、それをみつけるのはどれくらい簡単か。

背景

最近の論文で、Van Westrienen and Lynch (2005) は、現在英国には 31 の学術機関リポジトリがあると述べている。これらのリポジトリのコンテンツの形態の検討によって、「現在、機関リポジトリの大半は、従来の（印刷物指向の）学術出版物と灰色文献、すなわち雑誌論文や学位論文や研究報告書を保管している」という結論を彼らは導いている。彼らはさらに踏み込んで「学術出版におけるオープンアクセスの諸課題は、e サイエンスや e リサーチに関連する学術コミュニケーションについての新しい要望というよりは、少なくとも短期間に機関リポジトリの発展を推進させるための重要な要素である」と推測している。

「雑誌の危機」や図書館助成の不足や著作権の保有のようなオープンアクセスの諸課題についての議論は、文書で裏付けられつつある。本調査に関連するさらに深い課題は、最近いくつかの議論を誘発しているが、それは公的資金による助成研究の成果は、公衆が無料で利用できなければならないということである。これらの諸課題のすべては、「オープンアクセス」の意味内容の洗練と明確化を導き、代表的な最近の定義は「査読済みの科学・学術文献をインターネット上に置き、それを無料で著作権やライセンスの制約からほぼ自由に利用できるようにし、本格的な研究への障壁を除去すること」かもしれない (Suber, 2005)。

これを達成する方法の一つは「セルフアーカイビング」であり、機関リポジトリが、著者がそうすることができる場所の用意によりこの目的を支援することは明らかである。だが、本当のオープンアクセスを達成するためには、これは方程式の一部にしか過ぎない。それはまた、学術コンテンツを発見し、検索し、利用することができなければならない。2003 年のブダペスト・オープンアクセス・イニシアティブ (BOAI: Budapest Open Access Initiative) は、「第一に、学者は、セルフアーカイビングと通常呼ばれる慣行である、オープン電子アーカイブに査読雑誌論文を保管するツールと支援を必要とする。これらのアーカイブが、オープン・アーカイブ・イニシアティブが作成した標準に従う場合、検索エンジンや他のツールは、ばらばらのアーカイブを単一のものとして扱うことができる。利用者は、コンテンツの発見や利用のためどんなアーカイブが存在し、それらがどこにあるのかを知る必要がない。」と述べた。

2005 年のベルリン 3 会議 (Berlin 3 Conference) の代表者は、機関が「研究者にオープンアクセス・リポジトリにすべての出版論文のコピーを保管するように要求」すべきであると勧告し、Harnad はこの宣言をさらに洗練し、選択のリポジトリとしての機関リポジトリに特に言及して、次のようにいう。

大学や研究機関や研究助成機関は、一機関の政策上の問題として一被雇用者・被助成者が所属機関の OAI 準拠リポジトリに、出版された研究雑誌の各論文の補助的なコピーを保管することを要求すべきである。(Harnad, 2005a)

義務的保管の支持者は、保管が引用インパクトを増加させると主張することによって彼らの意見を補強する。Harnad and Brody (2004) は、例えば、論文がセルフアーカイビングされる場合、「論文の引用頻度の点で、実際に「明確に識別できる相違」があ

ることを示している。つまり、著者が OA にした論文を支持するという飛躍的なメリットがある。」

オープンアクセス運動の目的が機関リポジトリ—査読雑誌論文がそこに出現し、無料で全文のコピーが検索できる—によって一層推進されるという確実な兆しがある。だが、世界中で毎年出版される論文数を考えると、それらの存在は今のところ少ない。世界で 1 年間に出版されている論文数は 250 万件であると見積もられることが多い (Harnad et al., 2004; Harnad, 2005b) 一方で、Van Westrienen and Lynch (2005) は英国のリポジトリ当りの平均的な全種類の資料数は 240 件であると見積もっている。

世界中の機関リポジトリやその他のリポジトリの数が増加しているため、それらについての情報は機関アーカイブ台帳 (Institutional Archive Registry) で照合され、区分されている。これは現在、「英国や研究機関や部局」アーカイブを含む多くの種類のリポジトリについての情報を含んでいる。本調査の時点でこれらの数は Westrienen と Lynch の状況の査定と大体一致していた。

英国のリポジトリのコンテンツのフォーマットの多様性への論評や「e サイエンスや e リサーチに関連する学術コミュニケーションについての要望」への視点の欠如は興味深い。英国のリポジトリがこのように発展した理由の 1 つは、もちろん、それらの多くが SHERPA プロジェクトの後援の下に始まったからである。SHERPA は「メタデータハーベスティングのためのオープン・アーカイブ・イニシアティブ・プロトコル (OAI PMH: Open Archives Initiative Protocol for Metadata Harvesting) に準拠した 20 の連携機関に、機関オープンアクセス e プリント・リポジトリを設置すること」を特に目的としていた。事実、7 つの SHERPA 開発連携大学や英国図書館 (British Library) や 12 の賛助連携大学は、本調査の時点で Registry に記録された英国の機関リポジトリの集団の 2/3 を構成する。

リーズ、ヨークおよびシェフィールドの白バラ (White Rose) コンソーシアムのように SHERPA 開発連携リポジトリの大半は、基本的な科学データやそれ以外のデータの保管というよりも「オープンアクセス e プリント資料」、特に学術雑誌論文全文に対して焦点があるが、リポジトリのすべてがそのような論文しか含んでいないと考えるはならない。SHERPA のリポジトリは、雑誌論文 (利用者が購読機関に所属していることや発見の都度論文を購読することが必要である) の引用やレポートや講義録やニュースレターなどへ引用へのリンクを提供している。プレプリントや査読雑誌論文のポストプリントのような学術資料の全文があるリポジトリを存続してゆくという目的は簡単に達成できない。Markland and Brophy (2005) が報告しているように、著作権問題に関する誤解や「学問の慣性」や機関の政策の欠如によって学者の一部が保管をいやがるようになり、これは、そのようなコンテンツを持つリポジトリの急速な増加を妨げている。しかしながら、これと釣り合いをとって、例えばエジンバラやグラスゴーのような連携機関の一部は、他の機関イニシアティブと一緒に「e プリント」リポジトリをどのように発展させることができるかを示している。それらの機関のウェブサイトは、機関リポジトリが何であるかについてのかかなり広い見方かもしれない—最も広い意味で、機関の一連の文献や学術データへのアクセスが提供される仮想の場所—と区分された部分としての得意な分野 (niche) を発見していることを示す。

SHERPA コミュニティを超えて、コンテンツ方針については混合アプローチが普通であるという証拠がある。たとえば、クランフィールド大学のリポジトリは、出版された雑誌論文や教員の履歴書 (CV) や内部刊行の報告書や学位論文を含んでいる。その目的は、学術コミュニケーションの改善としてのみならず、公衆に対して機関の視認性 (visibility) を増加させ、一連の生産物を発見し、分析し、捕捉するような管理プロセスを容易にする手段として明確に定義されている。リポジトリは、機関や機関内の人々の事業ニーズに応じて定義され、存続され、その過程で従来は隠れたままであった一連の文書やその他の資料までアクセスが広げられた。このパターンは国際的なリポジトリの展開 (deployment) に反響している。米国の経験についていえば、Van Westrienen は「少なくとも短期間には、いくぶん強調点が異なる機関リポジトリについての2つの見解」(Van Westrienen, 2005: 33) があり、それは「主として教員の著作の e プリントの様々な形態への取り組み」と「機関の知的著作の広範囲な保管」であると述べている。

そのような活動は、むしろ、実際には機関の知的生産物の「ショーケース」かもしれないが、それ自体、「外部の」だれもがそこにあるものに目を向けることを意味しない。学術コミュニティにおける個人は、コンテンツ本体の「機関の展示」を気にするよりも学術著作や主題領域に関連するコンテンツを探索する傾向がある。これらの潜在的な利用者、学生および研究者の両方とも、彼らの学問的生活から自然に依存することになる情報探索のパターンに従うように思われ、利用者の多くはいうまでもなく、引用索引や文献データベースやハーベスティング・ツールを利用するだろう。

しかし、利用者について行われたいくつもの初期調査によって、彼らが、また、研究に関連のある、無料で利用できるコンテンツを発見するためにそんなに定型的ではない探索戦略を頻繁に使うことがわかっている。彼らは、簡単で、迅速で、利用がわずらわしくないのに特に検索エンジン (ことにグーグル) を好んでいる。例えば、みたところ他の検索ツールよりも、グーグルを好む理由を説明するように質問された利用者の一人は「率直に言って、例えばグーグルのような複数サイトがあり、それらが私が望むものに即座にアクセスを提供する場合、入り込むのに沢山の「努力」をしなければならないサイトを見ると、私は他をクリックする (Brophy et al., 2004a: 17)。彼らはまた、検索エンジンが「世の中に」あるものを網羅的に提供すると感じている。すなわち、「検索し、発見したものからいえば、私はグーグルの方がたぶんより多くの結果に到達し、より豊かな資源を生み出すと考えている」か (Brophy et al., 2004a: 17), 「それは迅速で、簡単で...網羅的で、驚くほどに十分であり、そこで学術的なものを発見できる。」

(Brophy et al., 2004b: 47) さらに、探索者は戻ってきた大量の「ヒット文献」を徹底的に調査するのを嫌う傾向があり、検索結果を 3 ページ以上ざっと見ることはめったにない。実際に、利用者の情報探索行動についての調査で、Craven and Griffiths (2002: 183) は、「圧倒的な数の両方のグループの参加者が検索結果の最初のページしか見ない」ことを発見した。そして、そのような情報探索者の個人的な情報風景の場面で、グーグルはいかに効果的に機関リポジトリのコンテンツを提供するのか。

本調査の目標と目的

本調査の目標は、したがって、グーグルの検索エンジンを利用して英国の機関リポジ

トリのコンテンツから収録資料の効果的な選択ができるかを評価することであった。

本調査の目的は以下のとおりであった。

- ・ 現行の英国の機関リポジトリを識別すること
- ・ 各リポジトリから収録資料を1つ選択し、リポジトリの検索画面 (interface) やグーグルを使ってそれがいかに簡単に検索できるかを確かめること

方法

Eprints.org の「機関アーカイブリポジトリ (Institutional Archive registry)」を使用し、さらに英国でかつ「研究機関」または「部局 (departmental)」のオプションを選択し、36 のリポジトリを識別した。主題分野別部局リポジトリと授業用資料しか含んでいないリポジトリを除く、26 の英国の機関リポジトリを本調査目的のために選択した。これらには3大学のコンソーシアムである白バラリポジトリを含むが、この調査のために単一の1つのリポジトリとして計上した。

このように識別した26のリポジトリについて、各リポジトリから収録資料を1つ選択した。リポジトリの検索画面のブラウザ機能を使うとたちまちすべてのリポジトリを横断して広範な主題領域が検索できることが示される。これらは、多様な主題の楽譜・録音資料や産科学・婦人科学や東洋学 (Oriental Studies) や思弁哲学や財務数学 (Financial Mathematics) とともに芸術や人文科学や社会科学や工学や科学に及んでいた。26 のさまざまな主題領域を選択することを決定した。それで、最も小規模のレポジトリを手始めに各リポジトリからさまざまな主題が選択された。

最初の資料はそこで各主題領域リストから選択された。各資料はこれがどのような種類のものか、そしてリポジトリのその資料へのリンクが「生きている」かをみるためにリポジトリのリンクからチェックした。資料が雑誌論文の場合、これが出版されたバージョンなのか、プレプリントや著者の最終バージョンとして表示されるのかをさらにチェックした。

リポジトリから資料がどのように簡単に検索できるかどうかをみるために次の探索戦略を使って検索を行った。

- ・ リポジトリの検索画面 (interface) からの検索
- ・ タイトル中のキーワードや句を使ったグーグルの検索
- ・ 詳細タイトル (complete title) を使ったグーグルの検索
- ・ タイトル中のキーワードや句を使ったグーグル・スカラーの検索
- ・ 詳細タイトル (complete title) を使ったグーグル・スカラーの検索

キーワード検索のために選択された検索タームは、通常は資料を検索するのに充分詳しいと判断された、資料のタイトルから抽出された句であった。だから、例えば、'QT interval prolongation related to Psychoactive drug treatment: a comparison of monotherapy versus polytherapy' では、選択された検索タームは 'QT interval prolongation' であった。プロセスのこの部分は、ある程度の主題知識を持っている情報探索者の役割を研究者が仮定していることを前提としているが、必ずしもそうとも限ら

ない。研究者が特定の検索の句の重要性を間違っって判断するのを防ぐために、検索がうまくいかなかった場合、別の句が選択され、検索が再実行された。事実、句の選択は非科学分野の場合に最も難しいことが証明された。その分野では、例えば、"Time for a threesome: European leadership and America"のように、想像力に富んでいるタイトルには非常に詳しい専門用語がないが多かった。

機関リポジトリから資料が検索されたかどうか、また他の無料のバージョンや購読バージョンが掲載されているかどうかをみるために「ヒットした」最初の3ページを吟味した。リポジトリのバージョンが購読バージョンよりも上位にランクされたかどうかを注記した。

結果

1. 26の機関リポジトリで何か見つかったのか

各主題リストの最初の資料を選択することで、一連のさまざまな種類の資料が見つかった。これらを表1に示す。

表 1. 26の機関リポジトリで発見された収録資料の範囲	
Type of article found	No. found
Full text publisher's version of journal article	5
Full text preprint journal article marked as author's final version	4
Full text pre-print journal article	6
Newsletter/magazine article	2
Citation with link to publisher's website	2
Citation only	1
Conference paper published	3
Conference paper unpublished	1
Lecture transcript	1
Login to repository required, so unable to identify	1

2. レポジトリの検索画面の検索ツールを使用した選択資料の検索

リポジトリのコンテンツからの検索に含まれるはずの、選択されたすべての資料について、実際にリポジトリから簡単に資料が検索されるかどうかをチェックするために、提供されている検索ツール、選択したキーワードおよび収録資料のタイトルを使ってリポジトリの検索画面から検索を実施した。この試験結果を表2に示す。

表 2. リポジトリの検索画面での検索結果	
Number of searches	
Found in the repository	23
Not found	3

3. グーグルおよびグーグル・スカラーを使用した選択資料の検索

タイトル中の句と詳細タイトルで検索するために、グーグルおよびグーグル・スカラーを使用し、選択した各資料について 4 つの検索を実施した。この活動の結果を表 3 に示す。

議論

1. 検索の単純な数値比較

レポジトリから選択されたすべての検索資料について、ホームページ上の「検索」装置を使用してレポジトリの検索画面から資料のタイトル中の句やメタデータ・レコードのキーワードで検索すれば、それらすべてが簡単に検索できると予想していた。しかしながら、これは事実ではなかった。検索で 23 件のみが当該資料にすぐに行き着いたに過ぎない。1 つのレポジトリの検索装置は、単純検索を行うため、論文タイトル中の句あるいは著者名を使うと、会議論文を見つけることができなかった。著者名を詳細検索の「著者・編者」欄に入力した場合に限って資料が見つかった。タイトルの句、メタデータ・レコードのキーワードの 1 つ、著者名あるいは詳細タイトルによる 2 つの検索では雑誌論文を見つけることができなかった。検索システムに内在する誤りは、当然のことながら、レポジトリの管理者が監視しなければならない重大事である。

グーグルおよびグーグル・スカラーの検索結果を、特定の学術活動を支援するために情報を探している学生、学者あるいは研究者の視点から検討してみると、一目見ただけで、検索結果からこれらのツールを利用して機関レポジトリから役に立つものを見つける最善の戦略が、単純なグーグル検索を行うことであることがわかる。[グーグル・スカラーの]12 に対して[グーグルでは]18 の資料が見つかった。しかしながら、率でいうと、グーグル・スカラーはレポジトリから独占的に高い率の資料（グーグルの 44% に対して 50%）を検索し、検討の対象となるきわめて少数の検索結果が返ってきた。それはたぶんすべてのレポジトリを（ひょっとすると「まだ」）対象にしていなかった。レポジトリが混合している場合のグーグルのキーワード検索を検討してみると、無料バージョンや購読バージョンが検索され、これらは検索結果の最初の 3 ページに分散していた。たぶん、この観点からすれば、レポジトリから独占的に多くの資料を検索するグーグル・スカラーが使用ツールとして優れているといえよう。

表 3. グーグルおよびグーグル・サーチの結果		
Searches using a phrase/keyword from the title <i>n</i> = 26	Google	G Scholar
Average number of hits	151,973	832
Repository version found in the first 3 pages of results	17	8
Repository version not found, but item found elsewhere	1	4
Item not retrieved	8	14
Total items found	18	12
Only repository version found (of which subscription required to access)	8 2	6 0
Only other subscription version found	1	2
A mix of repository, free or subscription versions found (of which subscription version ranked higher than free)	9 2	4 3
Searches using the full title <i>n</i> = 26	Google	G Scholar
Average number of hits	32	3
Repository version found in the first 3 pages of results	25	17
Repository version not found, but item found elsewhere	0	3
Item not retrieved	1	6
Total items found	25	20
Only repository version found (of which subscription required to access)	10 2	10 0
Only other subscription version found	2	2
A mix of repository, free or subscription versions found (of which subscription version ranked higher than free)	13 3	8 3

利用者が既に雑誌論文や資料のタイトルを知っている場合、グーグルやグーグル・スカラーは、多数の資料を検索し、うまく機能するように思われる。これまでのとおり、率からいうと、グーグル・スカラーがリポジトリから独占的に多くの資料を検索するように思われる。しかしながら、その数はすべてではない。発見された資料をもっと詳細に調査すると、多数の興味深い課題が明らかになるだろう。

2. バージョンの多様性からの選択

全体としてグーグル・スカラーが、数からみればグーグルよりもわずかに効率的ではないことが証明されたにもかかわらず、詳細タイトルがわかっている場合、グーグル・スカラーが行うのは、1回のヒットで同じ収録資料のさまざまなバージョンを一緒に集めることであり、その結果、無料ないしは「よりよい」バージョンを探すために検索者が列挙されたページを徹底的に探す必要が取り除かれる。グーグル・スカラーによって検索された8つの「混合 (mixed)」バージョンには、それぞれ1回のヒットで、3つのリンクとさらに16のサブリンクがある。さまざまなバージョンの「コレクション」は注意深く評価しなければならない。

例として、タイトル中の“Time quantization and q-deformations”という句を伴う物理学論文のグーグル・スカラーによる検索を取り上げる。

検索された最初のヒットは利用者を *Journal of Physics* の論文の引用に導く。ここからのアクセスのオプションは、当該雑誌の機関購読、論文の購入、当該雑誌への個人購読によるアクセス、あるいは利用者がこれを入手可能な場合は Athens のログインとなる。

しかしながら、本調査の時点でグーグル・スカラーの検索結果は、「他のオプションを参照」を経由して論文の16の他のバージョンに膨大な重複とともにリンクされている。これらの1番目は雑誌、2番目はデッドリンク、3番目は機関リポジトリのプレプリント・バージョンであった。この下に雑誌への1つのリンク、機関リポジトリのバージョンと同じかどうかわからないが ArXiv 分野別リポジトリのプレプリント・バージョンへの2つの直接リンク、物理学抄録サービス経由の ArXiv リポジトリへの1つのリンク、Ingenta サービスへの2つのリンク、韓国に本拠がある（と思われる）分野別リポジトリのプレプリント・バージョンへの1つのリンク、米国の分野別リポジトリへの1つのリンク、機関リポジトリのプレプリント・バージョンへの2つのリンク、プレプリントであると思われる第一著者のホームページへのリンク、第二著者のホームページへの2つのリンク、出版雑誌のバージョンの複製であると思われるもの1つ、その他機関リポジトリのあると思われるプレプリント・バージョンがある。沢山の選択肢がある。物理学者は、可能であれば出版雑誌のバージョンを選択すると推測されるが、物理学コミュニティはプレプリントを共有する文化であることが知られており、ここでは考えようによっては2つの異なったプレプリント・バージョンと出版状態が不明な著者のウェブページからのバージョンがあることになる。ArXiv バージョンは実際の文献の日付が明確であり、機関リポジトリバージョンや他のバージョンはそうではない。それらは、もっと古いかあるいはもっと新しいバージョンの可能性もある。したがってここでは、主題別リポジトリバージョンは重要な価値を加えるのか。どのバージョンが情報探索者

にとって最も価値があると思われるのだろうか。グーグル・スカラーの利用者は、さまざまな情報源やバージョンを評価するために何を理解する必要があるのか。

3. リストの最初に掲載されているものが最善とは限らない

グーグルの利用者にとって課題が、かなり違う可能性がある。例えば、特定の雑誌論文を検索するために、'spell checking algorithms'というタームを使って、グーグルの検索を実行した。グーグルの検索結果の最初のページには論文への5つのリンクがあった。これらの最初と2番目のものはIEEE デジタル図書館 (IEEE Digital Library) の論文にリンクされていた。大学キャンパスから離れた場所で検索した場合に論文に到達可能なオプションは19ドルで購入することである。グーグルのリストの3つめのリンクは、しかしながら、機関リポジトリ中の論文の全文出版バージョンに導いた。この例の場合、論文の別のバージョンが非常に隣接していたが、これはいつもそうとは限らない。それらは最初の30件の検索結果にちらばっていることが多く、無料の全文バージョンの前に資料（特に雑誌論文）の購読バージョンをリスト化しているグーグルとグーグル・スカラーの例がいくつもある。これは「満足化 (satisficing)」問題を引き起こす。以前の研究で、探索者が「最低限度の条件に満足する」ことが多く、検索された最初の資料が、彼らが必要とする情報を提供している場合にそれで間に合わせ、さらに、よりよいオプションを探すというよりは「検索結果の質を検索に費やす努力や時間と引き換えにするかもしれない」(Brophy et al., 2004a: 25) ことが知られている。検索結果で購読バージョンがリストの最初に押し上げられるようなやり方で順位付けされた場合、利用者はデリバリー・サービスのために不必要な支払をするかもしれない、論文を無視するかもしれない、悪いことに無料で利用できるものがないと推測してしまうかもしれない。

4. 機関リポジトリで見つけたものの出版状態は必ずしも明確ではない

繰り返し発生する課題の1つは、リポジトリで表示される文献の正確な状態を識別することが難しいことである。「本論文は著者がリポジトリに投稿したものである」、「eプリントタイプの論文」あるいは単に「eプリント」のような記述は、文献の最初の原稿、著者の最終バージョンあるいは出版バージョンのある種類のコピーを探している利用者にとって実際に充分ではない。資料がリポジトリに登録された日付があるかもしれないが、日付のある文献はわずかである。

リポジトリの文献を確認できる大変よい例がある。エジンバラ大学のエジンバラ研究アーカイブ (ERA: Edinburgh Research Archive) から検索された資料の標題紙の通知は、例えば、「これは、査読を経た John Wiley and Sons が出版する...という雑誌に受理された論文のコピー編集前の著者作成 PDF である」と疑問の余地なく述べ、それは文献に付随するメタデータ・レコードに埋もれた表示よりもすぐ目につく。

文献についての明快な引用情報も役に立つ。プレプリントの回覧は、ある学問分野にとって目新しい現象なので、学生や研究者は、リポジトリ文献を引用する際に、バージョンを正しく区別する方法をよくわからないのである。

5. タイトルで使われているタームは検索エンジンによる検索の機会を少なくする

句による検索で、機関リポジトリから検索できなかった資料は、タイトルやコンテンツ中の専門語や主題限定語が少なかったものである。これらのいくつかの資料はタイトル全体をキーインしないと検索できなかった。全体的に、科学や医学や工学分野の方が芸術や人文科学や社会科学に比べて検索がうまくいった。雑誌論文や会議論文の主題が'glutamatergic synapses'の場合、'fallen woman'や'perfect pitch'が主題の場合に比べてリポジトリバージョンがリストの高い順位にあることに驚いてはいけな。この問題をどのように取り扱えばよいかを知るのはむずかしい。それは、主題語彙に特定の用語か、一意の用語か、「専門的な (technical)」用語がわずかにしかない場合、著者が検索エンジンに著作をどのように目にふれさせるかということだ。

学術コミュニティのこれらの利用者に対し、リポジトリが OAI 準拠であることを保証し、学生や学者や研究者が、ハーベスティング・ツールを使用することを推奨することは、絶大な成功を収める経路であることを証明するかもしれない。もちろん、これは強力な検索ツールの開発のための助成とそれらの利用に際して大学の利用者が広範な訓練を受ける必要があることを示している。

結論

情報探索プロセスの初期段階か、ニーズをうまく特定していないか、あるいはたぶん、特定主題について「役に立つコンテンツ」を探しているグーグル利用者は、うまくリポジトリ資料に行き当たるかもしれない。特定の雑誌論文のタイトルを探している利用者は、グーグルやグーグル・スカラーの検索が論文の無料全文バージョンを提供することを上手に見つけるかもしれない。それは確かに試みるべき探索戦略である。他方、そのような利用者はまた、検索された論文のリストから最善のバージョンをほぐして取り出す技能と専門的知識を必要とするだろうが、現在のリポジトリは各文献の状態を明確に示すことによってこのプロセスに役立っているとはいえない。

機関リポジトリに保管される全種類の文献量が増加するので、2つの検索エンジンのいずれもが、出版された全文査読済みの雑誌論文のみを発見する便宜を現在提供していない。このような細分化のレベルを獲得できるサービスの必要性が学者にとってますます緊急のものとなる。一部のリポジトリの検索画面では、この必要性を認識し、そのような便宜を提供している。例えば、オックスフォード大学図書館サービスのリポジトリである **Oxford Eprints** は、「オンライン」あるいは「頁付けのある雑誌」に加えて「出版済み」や「査読済み」[という条件]で利用者が検索できるようにしているが、でもこれもまた、個々のレポジトリのインターフェースでの検索の習慣として広まっていない。検索のために特定のレポジトリを選択する **ePrints UK** や **OAISTER** のような、学術コミュニティ向けのサービスは、これがどのように達成されるかを示すが、前者はまだ成熟していない。

現在の形式の機関リポジトリが今後とも繁栄するかどうかは現時点では不明である。この時点で、機関リポジトリにおけるさまざまな種類の文献の保管率が増加している明らかな根拠がある。これらのレポジトリに保管されている査読済み雑誌論文の本体は、英国の学樹コミュニティ全体にわたる保管の可能性に比べてまだ大きくない。機関リポジトリに保管している一部の著者は、また、他の分野別またはプレプリントリポジトリ

に資料を重複して保管する道を選んでいるという証拠がある。機関にとってリポジトリは財政投資であり、したがって、それは機関固有の資源としての価値を明らかに証明しなければならないだろう。学者にとって、その価値は役に立つコンテンツがいかに簡単に検索できるかを含む、それ以外のことがらによるものだろう。グーグル、グーグル・スカラーあるいは他の検索エンジンは情報探索者の一部に普及しているというものの、それを行う手段として最善ではないかもしれない。

最後に、検索エンジンには新しい利用者、つまり学術的なメタ検索ツールやサービスに不案内だが、仕事や家庭でグーグルの検索に非常になじんでいる利用者を含む、オープンアクセスの提唱者によって想定された広い公衆に対してオープンアクセス資料への経路を提供するという、興味をそそる可能性がある。グーグルの検索は、学術コミュニケーションの世界で発見されるべき豊かなコンテンツへの糸口になるかもしれない。

引用文献

- Berlin 3 Conference on Open Access to Knowledge in the Sciences and Humanities (2005) *Recommendation*. URL (consulted November 2005): <http://www.zim.mpg.de/openaccess-berlin/recommendation.html>
- Brophy, P., Markland, M., Griffiths, J.R. and Booth, H. (2004a) *EDNER+: Formative Evaluation of the Information Environment: Evaluation of IE Presentation Services* (Deliverable 4, EDNER+ Project). Manchester: CERLIM (The Centre for Research in Library & Information Management). URL (consulted November 2005): http://www.cerlim.ac.uk/projects/iee/reports/wp4_user_testing.doc
- Brophy, P., Fisher, S., Bentley, J.C., Griffiths, J.R. and Markland, M. (2004b) *EDNER: Formative Evaluation of the Distributed National Electronic Resource: Stakeholder Consultation and Analysis. Information Usage in Higher Education* (Deliverable MDA 3a, EDNER Project). Manchester: CERLIM (The Centre for Research in Library & Information Management). URL (consulted November 2005): <http://www.cerlim.ac.uk/edner/dissemination/a3a.doc>
- Budapest Open Access Initiative (2003) URL (consulted November 2005): <http://www.soros.org/openaccess/read.shtml>
- Cranfield QUEprints URL (consulted November 2005): <http://dspace.lib.cranfield.ac.uk/>
- Craven, J. and Griffiths, J.R. (2002) '30,000 Different users, 30,000 Different Needs? The Design and Delivery of Distributed Resources to the User Community', in P. Brophy, S. Fisher and Z. Clarke (eds), *Libraries Without Walls: The Delivery of Library Services to Distant Users: Proceedings of the 4th Libraries Without Walls Conference*, 14.18 September 2001. London: Facet.
- Edinburgh Research Archive (ERA). URL (consulted November 2005): <http://www.era.lib.ed.ac.uk/index.jsp>
- ePrints UK Project. URL (consulted November 2005): <http://www.rdn.ac.uk/projects/eprints-uk/>
- Glasgow Daedalus Project. URL (consulted November 2005):

<http://www.lib.gla.ac.uk/daedalus/index.html>

Harnad, S. (2005a) 'The Implementation of the Berlin Declaration on Open Access: Report on the Berlin 3 Meeting Held 28 February.1 March 2005, Southampton. UK', *D-Lib Magazine* 11 (3). URL (consulted November 2005): <http://www.dlib.org/dlib/march05/harnad/03harnad.html>

Harnad, S. (2005b) 'Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold', *Ariadne* 42. URL (consulted November 2005): <http://www.ariadne.ac.uk/issue42/harnad/intro.html>

[日本語訳]

オープンアクセスへのグリーン・ロードの急速前進：グリーンとゴールドの混同に対する反論 <http://www.nii.ac.jp/metadata/irp/harnad42/>

Harnad, S. and Brody, T. (2004) 'Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals', *D-Lib Magazine* 10 (6). URL (consulted November 2005): <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

[日本語訳]

同一ジャーナルに掲載されたオープンアクセス論文と非オープンアクセス論文のインパクトを比較する <http://www.nii.ac.jp/metadata/irp/harnad/>

Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H. and Hilf, E. (2004) 'The Access/Impact Problem and the Green and Gold Roads to Open Access', *Serials Review* 30 (4): 310.14. URL (consulted November 2005): <http://dx.doi.org/10.1016/j.serrev.2004.09.013>

Hitchcock, S., Brody, T., Gutteridge, C., Carr, L. and Harnad, S. (2003) 'The Impact of OAI-based Search on Access to Research Journal Papers', *Serials* 16 (3): 255.60. Institutional Archives Registry. URL (consulted November 2005): <http://archives.eprints.org/>

Markland, M. (2005) 'Does the Student's Love of the Search Engine mean that High Quality Online Academic Resources are being Missed?', *Performance Measurement and Metrics* 16 (1): 19.31.

Markland, M. and Brophy, P. (2004) EDNER+: Information Environment Formative Evaluation: User testing of SPP Beta Portals (Deliverable Z2) in support of the EDNER+ Project, Manchester: CERLIM (Centre for Research in Library & Information Management). Unpublished report.

Markland, M. and Brophy, P. (2005) *SHERPA Project Evaluation, Final Report*. Manchester: CERLIM (Centre for Research in Library & Information Management). URL (consulted November 2005): <http://www.cerlim.ac.uk/projects/sherpa/finalreport-august.doc>

Oxford ePrints. URL (consulted November 2005): <http://eprints.ouls.ox.ac.uk/perl/search/advanced>

SHERPA Project Homepage. URL (consulted November 2005): <http://www.sherpa.ac.uk/index.html>

- Suber, P. (ed.) (2005) *Open Access News*. URL (consulted November 2005): <http://www.earlham.edu/~peters/fos/fosblog.html>
- Van Westrienen, G. (2005) 'Making the strategic case for institutional repositories. Completed questionnaires. Country update on academic institutional repositories', *CNI-JISC-SURF Conference*, Amsterdam 10.11 May 2005. URL (consulted November 2005): <http://www.surf.nl/en/download/country-update-2005.pdf>
- Van Westrienen, G. and Lynch, C.A. (2005) 'Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005', *D-Lib Magazine* 11 (9). URL (consulted November 2005): <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>
- White Rose Consortium ePrints Repository. URL (consulted November 2005): <http://eprints.whiterose.ac.uk/>

著者略歴

Margaret Markland はマンチェスター・メトロポリタン大学の図書館情報管理研究センター (CRRLIM: Centre for Research in Library and Information Management at Manchester Metropolitan University) の研究員 (research associate) である。彼女は、機関リポジトリを含む合同情報システム委員会の情報環境 (JISC Information Environment) の発展の諸側面を評価する研究プロジェクトに従事している。彼女は現在、EU が助成した、ヨーロッパ全体で授業と学習において児童と教師が電子オンライン技術とモバイル技術をどのように利用できるかを調査する、e ラーニングプロジェクトに関与している。

連絡先

Department of Information and Communications Manchester Metropolitan University
Room 239, Geoffrey Manton Building
Rosamond Street West
Off Oxford Road
Manchester M15 6LL
Tel: +44 (0) 161 247 6142
E-mail: m.markland@mmu.ac.uk