

PIRUS—出版者・機関リポジトリ利用統計

PALS 3 プロジェクト

機関リポジトリ、出版者などがホスト（所蔵・提供）する個別のジャーナル記事についてのオンライン利用統計の記録、報告、集約を可能とする世界的標準の策定

（出版者メタデータ・相互運用性プロジェクト 3）

最終報告

2009年1月

プロジェクトチーム：

Tim Brody（サウサンプトン大学）

Richard Gedye（オックスフォード大学出版局）

Ross MacIntyre（マンチェスター大学）

Paul Needham（クランフィールド大学）

Ed Pentz（CrossRef）

Sally Rumsey（オックスフォード大学）

Peter Shepherd（プロジェクトマネージャー）



このコンテンツは[クリエイティブ・コモンズ表示 2.0 イングランド&ウェールズ](https://creativecommons.org/licenses/by-nc-nd/2.0/)のもとでライセンスされています。

目次

	ページ
1. 謝辞.....	3
2. 概要.....	3
3. 背景.....	5
4. 目的と目標.....	6
5. 手法.....	6
6. 実行.....	7
7. 成果物.....	16
8. 結果.....	17
9. 結論.....	17
10. 含意.....	18
11. 提案事項.....	19
12. 参考資料.....	19
13. 付録.....	20

1. 謝辞

本プロジェクトは、出版者メタデータ・相互運用性プロジェクト3 (Publisher Metadata and Interoperability Projects 3) の公認を受け、JISC (英国情報システム合同委員会) から資金提供を受けています。そのご支援に厚くお礼申し上げます。とりわけ Alastair Dunning 氏には、本プロジェクトへの指針をいただけたことに感謝の意を申し上げます。また、必要なテストの構築と実行に際してクランフィールド大学、サウサンプトン大学、BioMed Central から協力をいただけたこと、各出版者やリポジトリから様々なシナリオに対するフィードバックを提供いただくという形で協力をいただけたことについても、ここに謝意を申し上げます。

2. 概要

PIRUS (出版者・機関リポジトリ利用統計) の目的は、オンラインジャーナル記事をホストする全ての機関 (出版者、アグリゲータ、リポジトリなど) が導入できる形で、かつ研究成果の利用状況を世界レベルで標準化して記録、報告、集約できるような形で、個別記事レベルの COUNTER 準拠の標準規格と利用レポートを策定することであった。

PIRUS を進める過程でその主たる目標が変わることはなかったが、リポジトリの技術的システム、組織、管理、資料内容が極めて多様であることを考慮して、当初提案されていたアプローチを一部修正する必要があることがプロジェクトのフェーズ1で実施した調査や机上研究の過程で明らかになった。万能なアプローチは存在せず、各リポジトリが COUNTER 準拠の利用統計を個別記事レベルで作成できるようにするためには、複数のアプローチを提供することが必要となる。そのため複数のシナリオを作成し、プロジェクトチームとしては、これによって大多数のリポジトリと出版者が COUNTER 準拠の利用統計を個別記事レベルで作成することができるようになると考えている。

本プロジェクトには、以下の4つの主要な成果がある：

- a. 個別記事利用報告「記事レポート1：フルテキスト記事ダウンロード成功数」用 COUNTER 準拠 XML 概念実証プロトタイプ。リポジトリと出版者が共通に利用できるもの。理想的には、このレポートは、著者個人と機関の両者に対して作成できる。しかし実務的には著者個人についてのレポートは生成が容易で短期的に現実性のある目標となる一方、機関など (例えば、助成機関) についてのレポートはずっと複雑であり、長期的な目標と見なすべきである。
- b. トラッカーコード。リポジトリが実装する。利用統計の作成と集約の責任および集約を行う関係出版者への送付の責任を担う外部当事者へ、またはローカルのリポジトリ・サーバーへ、メッセージを送る。
- c. 個別記事利用統計の作成、記録、および集約のための様々なシナリオ。現行のリポジトリの大部分をカバーする。各リポジトリは、自組織の技術と実装に合ったシナリオを選択することができる。
- d. 必要に応じて (一部のカテゴリーのリポジトリについて) 利用統計を作成し、他者のために利用統計を収集、集約する中心機構についての要件の決定。

本プロジェクトチームによる提案事項は、以下の通りである：

a. JISCに対して：PIRUSは、リポジトリおよび出版者からのデータを用いて個別記事の利用統計を作成、記録、集約することが**技術的に可能**であることを実証した。これを基に実装可能な新たな**COUNTER**標準およびプロトコルを作成するためには、さらなる研究と開発が具体的に以下に示す領域で必要となる：

- ・技術面：提案されているプロトコルとトラッカーコードにスケーラブル性、拡張性があり、主要なリポジトリの環境で問題が生じず、記事以外のアイテムにも応用可能であることを確認するために、より多くの多様なリポジトリと大量のデータによるテストが必要である。
- ・組織面：提案されている中心的クリアリングハウスの性格と使命を明確化し、候補組織を特定、テストする必要がある。
- ・経済面：必要となる利用レポートを生成する上でリポジトリと出版者にかかるコストおよび中心的クリアリングハウスのコストを評価し、こうしたコストを関係者の間で配分する方法について検討する。
- ・政治面：すべての主要関係者集団（リポジトリ、出版者、著者）による幅広い支持が必要となる。PubMed Centralのような主題リポジトリからは本プロジェクトの現段階においては積極的な関与を得ていないが、今後参加を得る必要がある。知的財産、プライバシー、財務面の問題についても対応する必要がある。

PIRUSプロジェクトチームは、上述した諸問題に対処するためのさらなる研究を助成することの検討を JISC に提案する。

- b. COUNTERに対して：COUNTERの使命を拡大してリポジトリからの利用統計を含めるようにすること。新たな記事レポート1を任意追加レポートとして導入することを検討すること。既存の独立**COUNTER**監査を修正して新たなレポートやプロセスをカバーすること。
- c. リポジトリに対して：主題リポジトリが本プロジェクトの次の段階に参加すること。すべてのリポジトリが記事のバージョンなどに対して標準的データ記述を使用すべきである。
- d. 出版者、ベンダーに対して：個別記事レベルで信頼できる利用統計を提供することが好ましいことを理念として認めること。

以上をまとめると、リポジトリと出版者が共通の技術標準に従ってオンライン利用を計測することには、その事業を担う組織と技術環境の多様性にもかかわらず現実的可能性があることが PIRUS によって示されている。しかしまた、この実現可能性研究の結果を基に現実的かつ実装可能でスケーラブル性のある解決策をまとめるため、さらなる研究が必要となることも示された。

3. 背景

COUNTER は、オンライン出版物について利用できる利用統計の信頼性を改善するために開始された。この目的を実行するため、COUNTER はベンダー利用統計の記録、報告、送達についての標準を定める実務指針（Code of Practice）を策定した。ジャーナルおよびデータベース用の COUNTER 実務指針リリース 1 は、2003年1月に発行された。2005年3月に発行されたリリース 2 は以後ベンダーの間で幅広く採用され、それによって作られる利用レポートは図書館で幅広く利用されている。2006年にはオンライン書籍および参考図書

用の実務指針が発行され、COUNTERの対象範囲はさらに拡大した。COUNTER 準拠利用統計は、現在 100 を超えるベンダーから 15,000 本を超えるオンラインジャーナルについて入手可能となっており、統計対象とされるオンライン書籍や参考図書の数も増加している。ジャーナルおよびデータベース用の実務指針リリース 3 は、現在 COUNTER のウェブサイト (http://www.projectcounter.org/code_practice.html) で利用可能であり、ベンダーは 2009 年 8 月 31 日までにこれを実装する必要がある。このリリースには、XML フォーマットによる利用レポートの義務化や SUSHI プロトコルの実装など、いくつかの微調整された点がある。

現在までのところ COUNTER によって利用報告が要求されている最も細かいレベルは、個別のジャーナルのレベルである。個別の記事レベルでの利用統計を求める利用者からの要求はこれまでのところ小さい上、エクセル環境で利用報告を行うことの不便さもあって、COUNTER ではこれまで個別記事レベルでの利用レポートが重視されることはなかった。しかし、最近のいくつかの状況の変化により、個別記事レベルでの利用統計の記録と報告についての COUNTER 標準の策定に力を入れる必要が出てきている。中でも、以下の変化が特に重要である：

- ・機関リポジトリなどのリポジトリがホストするジャーナル記事の数の増加。これらの利用統計については、広範に認知された標準が策定されていない。
- ・JISC が 2007～2008 年のデジタルリポジトリ事業で実施した利用統計レビュー。2008 年 7 月にベルリンで行われたワークショップを受け、デジタルリポジトリが保持する電子文書についてのアイテムレベルの利用統計を提供するためのアプローチが提案された。(1)
- ・オンライン利用統計が、記事およびジャーナルの価値を測る代替的尺度として認知されるようになってきたこと。利用ベースの指標が、英国研究評価フレームワーク(2)などで利用するツールの 1 つとして検討されている。
- ・著者と助成機関の間に、世界的な規模での個別記事の利用状況を信頼できる形で把握することへの関心が高まっていること。
- ・COUNTER が XML ベースの利用レポートを策定したことで、より細かいレベルでの利用報告が現実的になったこと。
- ・COUNTER が SUSHI(3)プロトコルを策定したことで、様々な情報源からの利用データを自動的に集約することが容易になったこと。

COUNTER は、その立場故に、出版者、リポジトリ、リポジトリシステム供給者、著者、図書館の関係専門家の協力を得て、既存の COUNTER 標準を基礎として、様々な場所にホストされている個別記事についてのオンライン利用統計の記録、報告、集約を律する新たな標準を、実施可能で幅広く受け入れられる形で策定することができる。

4. 目的と目標

本プロジェクトの目的は、オンラインジャーナル記事のフルテキストをホストする全ての機関（出版者、アグリゲータ、リポジトリなど）が導入できる形で、かつ研究成果の利用状況を標準化して記録、報告、集約できるような形で、個別記事レベルの COUNTER 準拠の標準規格と利用レポートを策定することであった。

PIRUS を進める過程でその主な目的が変わることはなかったが、リポジトリの技術的システム、組織、管理、資料内容が極めて多様であることを考慮して、当初提案されていた具体

的な目標の一部を修正する必要があることがフェーズ 1 で実施した調査や机上研究の過程で明らかになった。万能なアプローチは存在せず、各リポジトリが COUNTER 準拠の利用統計を個別記事レベルで作成できるようにするためには、複数のアプローチを提供することが必要となる。そのため複数のシナリオを作成し、プロジェクトチームとしては、これによって大多数のリポジトリと出版者が COUNTER 準拠の利用統計を個別記事レベルで作成することができるようになると考えている。こうしたシナリオについては、下記セクション 6 の図 1 に示す。

この目標については、出版者やリポジトリなどに対して、自らが出版するすべての記事を列挙してそのダウンロード数を記録した月次レポートを定常的に作成することを要求するようなものであるべきではないという合意があった。そのような方法では、データ量が莫大になり、レポートは扱いにくい巨大なものとなるだろう。そのようなことを目標とするのではなく、個別記事の一つ一つやその集合についての利用データを必要な時にそれに応じて取りまとめて入手できるようにすることを目標とする。その方が、より実際的なアプローチである。

5. 手法

本プロジェクトは 3 つのフェーズに分けられた。詳細は以下の通りである。

フェーズ 1 (2008 年 8～9 月) : 調査・机上研究。個別記事識別子やその他のメタデータの利用についての現状の実務慣習や個別記事の複数のバージョンを区別する方法などを評価した。これは、PALS 2 事業ですで行われた記事識別子についての研究を基礎としている (<http://www.iisc.ac.uk/whatwedo/programmes/pals2/counter.aspx>)。2 つの研究が実施された。P Needham によるリポジトリについての研究と P Shepherd による出版者についての研究である。本プロジェクトのこのフェーズは、予定通りに完了した。

フェーズ 2 (2008 年 9～11 月) : 個別記事の利用の記録および報告についての利用レポートとプロトコルのドラフトを策定し、これを出版者とリポジトリが持つ利用データでテストした。フェーズ 1 の結果、リポジトリが DSpace などの同一のソフトウェアを使用している場合であってもその構成には様々なものがあることが示されたため、オンライン利用統計の記録、収集、報告に対して単一のアプローチを指定することは実際的ではないと考えられた。そのため、複数のアプローチを開発、テストすることが決定された。どのアプローチも、妥当で比較可能な利用統計を提供するものである。結果的に、本プロジェクトのこのフェーズの完了は 12 月になった。

フェーズ 3 (2008 年 12 月) : テストの結果を考慮して COUNTER 準拠利用レポートの最終的なフォーマットをそれを使用するプロトコルと合わせて提案し、COUNTER が採用および維持する新たな標準としての承認を求めて COUNTER に提出した。フェーズ 2 が 12 月まで完了しなかったため、COUNTER などの組織への最終報告と提案が終わったのは、2009 年 1 月になった。

本プロジェクトの過程で、プロジェクトチームは計 10 回の会合を持った。通常は、電話会議の形を取った。

6. 実行 本プロジェクトの目標は、以下の通りであった :

- a. リポジトリと出版者が個別記事レベルでオンライン利用を特定、記録、報告している現行の方法を理解すること。

- b. オンラインジャーナル記事をホストする全ての機関（出版者、アグリゲータ、リポジトリなど）が導入できる形で、かつ研究成果の利用状況を世界レベルで標準化して記録、報告、集約できるような形で、個別記事レベルの COUNTER 準拠の標準規格と利用レポートを策定すること。

PIRUS は、上記セクション 5「手法」で述べたように、3つのフェーズで実行された。各フェーズの実行内容を、以下に説明する。

フェーズ 1：調査・机上研究。 個別記事識別子やその他のメタデータの利用についての現状の実務慣習、個別記事の複数のバージョンを区別する方法などを評価した。

出版者の調査

アンケートを 15 の出版者、ベンダー、ホストに送付した。合計で 12 の回答があり、回答は書面によるか電話での聞き取りによって得た。回答のあった組織には、下記の文中で星印を付した。

調査対象となった出版者・ベンダー：American Chemical Society*、American Institute of Physics、Atypon*、BioMed Central*、EBSCO、Elsevier*、Ingenta*、Institute of Physics Publishing*、Nature Publishing Group*、OUP*、Ovid*、Sage*、Springer*、Taylor & Francis*、Wiley Blackwell。これらの出版者・ベンダーは、すべて現在 COUNTER に準拠しており、調査対象が領域、規模、地理的所在地の面で業界の代表的サンプルとなるように選択した。

質問と回答の完全なリストについては、付録 A を参照。

調査に回答した出版者の多くは、理念としては個別記事レベルでの利用状況を報告することに価値があると考えているが、若干の出版者は、これが本当に有益であるのかと依然として疑念を抱いている。ほかにも、関係するデータの量が膨大になる可能性やその処理の実務面や費用についての懸念が存在している。このような懸念の背後には、個別記事レベルのレポートが従来の COUNTER 利用レポートを踏襲したものになるという前提があった。従来のレポートでは、各顧客ごとに購読しているすべてのジャーナルについて毎月報告する必要がある。プロジェクトチームでも、出版者に対してそのような報告を個別の記事レベルで提出することを出版者に求めることが過度の負担となるだけでなく、関係するデータの量があまりにも膨大になるために顧客による管理も容易ではなくなるということを確認している。出版者にとって負担が少なく機関や著者にとって価値の高い方法は、個別記事利用データを著者や機関が必要とする時にそれに応じて提供することである。従って、目標は、出版者とりポジトリが著者やその機関についての個別記事利用統計を生成、処理することができるようにし、その統計が COUNTER 標準と比較可能で、信頼でき、整合性を持つようにすることである。

回答者のほとんどは、一意的な記事識別子を使用している。これは記事の恒久的な属性であり、すべてではないもののほとんどが DOI を使用している。バージョンの追跡と特定の方法には、様々な慣習がある。

運用面での多少の違いがありながらもジャーナル出版者の間で DOI が一般的に使用されている一方、アグリゲータの間では DOI の使用はそれほど一般的ではない。アグリゲータがホストするフルテキスト・アイテムの多くがジャーナル記事ではないからである。しかし、ジャーナル記事のみに限れば、DOI が一般的に使用されているようである。個別記事の利用の報告には記事識別子の標準を定めることが必要であり、少なくとも出版環境においては、DOI が最も有力な候補であろう。その使用を要求するためには、それを出版プロセスの中に

導入するためのプロトコルを定めることも必要となる。このプロトコルでは、次の問題を扱う必要がある：

- ・ DOI を使用する記事のバージョン
- ・ 出版プロセスの中で DOI が使用される段階

機関リポジトリ (IR) の状況

IR に関する状況の調査は、机上研究、PIRUS チーム内外での討論、メーリングリストの JISC-REPOSITORIES で送られた電子メール調査を組み合わせで行った。

最初の課題は、世界中で使用されているアプリケーションソフトについて明確な理解を得ることだった。

リポジトリソフトウェア

すでに 2、3 年が経過しているものの、ブダペスト・オープンアクセス・イニシアティブ (BOAI) による報告書「A Guide to Institutional Repository Software v 3.0」(4)でリポジトリソフトウェアについての優れた紹介が行われている。オープンソース・ライセンスで利用可能な 9 本のソフトウェア (Archimede、ARNO、CDSware、DSpace、Eprints、Fedora、i-Tor、MyCoRe、OPUS) についての説明と比較が行われている。中でも System Feature & Functionality Table という表が有益で、9 本のソフトウェアの相違点がまとめられている。

これらのオープンソース・ソフトウェアだけでなく、BePress の Digital Commons や Ex Libris の Digitool といった商用システムを利用することもできる。

付録 B に示す最も普及しているソフトウェアの表は、BOAI の「Guide to Institutional Repository Software」、ROAR の「Registry of Open Access Repositories」、および各ソフトウェアについての文書からの情報を組み合わせて作成した。

この表の情報を分析すると、世界的に見て、掲載されている IR の 5 分の 4 (80%) が使用しているソフトウェアは、わずかに 5 つしかない。

DSpace (Open Repository を含む)	37.9%
Eprints	27.3%
Digital Commons	8.3%
OPUS	3.9%
DiVA	2.7%

IR の使用ソフトウェアの 3 分の 2 が DSpace と Eprints の 2 つに集中している。

ただし、ROAR のリストでは、何らかの理由で Fedora リポジトリの使用率が低めに評価されているように思われる。

IR のコンテンツタイプ

IR が扱うコンテンツは多様なタイプが混在していることが普通であり、ジャーナル記事、大会発表論文、学位論文、紀要、技術レポート、プロジェクトレポート、書籍の章、プレゼンテーション、データセット、画像などがある。

従って、どのアイテムが記事であり、どのように各記事の異なるバージョンが特定されているかを理解するためには、IR で利用されているメタデータについて詳しく見る必要がある。

メタデータ

リポジトリソフトウェアのほとんどは限定子付きダブリンコア (QDC) をサポートしているか、そうでなければ QDC に対応しているか非常に簡単に QDC に対応づけることができるメタデータを保持している。

IR が記事をカタログ化する際に通常使用するメタデータの要素は、以下の通りである：

- ・タイトル
- ・著者
- ・アブストラクト
- ・ジャーナルタイトル
- ・巻 (号)
- ・ページ
- ・ISSN
- ・DOI
- ・書誌引用
- ・リソースタイプ
- ・ローカル識別子

どのリポジトリも、タイトル、著者、リソースタイプのメタデータを使用している。机上研究 (付録 C) とその補完としての電子メール調査 (付録 D) によって、出版された記事のバージョンを記録するための引用情報を加えているリポジトリが多いことが確認されている。

以下、最も普及している 2 つのソフトウェアについて見てみる：

DSpace

「タイトル (Title)」フィールドは、フリーテキスト。

「著者 (Author)」フィールドは、フリーテキストの「姓 (Last name)」と「名 (First name(s))」の 2 つのサブフィールドから成る。

「引用 (Citation)」フィールドは、フリーテキスト。入力内容が制限されていないのは、このフィールドの内容がリポジトリによって異なり、予想できないからである。

「タイプ (Type)」フィールドはリスト選択で、リストはインストール単位で設定することができる。このフィールドが保持する値は非常に多様であり、その例を以下に示す：

- ・記事
- ・ジャーナル記事
- ・ポストプリント
- ・研究論文
- ・出版ジャーナル論文 (審査付き)

ここから分かるように、「タイプ」フィールドは以下の情報を盛り込むために多目的に使用されることがある：

- ・ リソースタイプ (記事)
- ・ 学術ステータス (審査・査読付き)
- ・ 出版ステータス (出版済み)

Eprints

「タイトル (Title)」フィールドは、フリーテキスト。

「著者 (Author)」フィールドは、フリーテキストの「姓 (Family name)」と「名・イニシャル (Given name/Initials)」の2つのサブフィールドから成る。

「引用 (Citation)」フィールドは、著者 (Author(s))、出版日 (Publication date)、タイトル (Title)、ジャーナルタイトル (Journal Title)、巻 (号) (Volume (Number))、ページ (Pages)、ISSN など、複数のフィールドの合成である。

このうち、著者、タイトル、ジャーナルタイトルは必須フィールドであるが、その他はそうではない。

「タイプ」フィールドはリスト選択で、リストはインストール単位で設定することができる。しかし、このフィールドには (ほとんど) 常に初期状態の値である「**article** (記事)」が与えられる。

査読ステータスと出版ステータスは、別個のフィールドとされている。

フリーテキストで入力される多くのメタデータは人間が記事を特定するためには有用だが、自動化された機械的特定のためには有用性が劣る。個別記事レベルでの利用統計を提供するためには、集約や重複チェックなどが可能となるよう、個別記事を正確に特定できることが必須である。そのため、信頼できる形で認識できる識別子を採用することが必要である。

識別子

すべてのリポジトリソフトウェアで、レコードが作られた時にローカル識別子が割り当てられる。現行の慣習は、以下の通りである：

Digital Commons

Digital Commons は、各レコードに独自の識別子を付与する。

DigiTool

Digitool は、各レコードに独自の識別子を割り当てる。

DiVA

DiVA は、各レコードに識別子として URN:NBN を付与する。URN は、スウェーデン王立図書館のリゾルブサービスを利用して読み出すことができる。

DSpace

DSpace は、初期状態では CNRI のハンドル (Handle) をプライマリー識別子として使用し、DSpace を使用しているリポジトリの大多数はハンドルを利用している。

Eprints

Eprints は、各レコードに独自の識別子を付与する。

Fedora

「Fedora のデジタルオブジェクトは、Fedora 内では PID (持続的識別子) を用いて特定される。PID は大文字小文字を区別し、名前空間の接頭辞と単純な文字列識別子から成る。」

[参照 : <http://www.fedora.info/definitions/identifiers/>]

OPUS

OPUS は、各レコードに URN を付与する。URN は、ドイツ国立図書館のリゾルブサービスを用いて読み出すことができる。

arXiv

「2007 年 4 月 1 日 (0704) 以後、新しい論文はすべて arXiv:0706.0001 という形式の識別子を持つ。個別のバージョンへの参照には、arXiv:0706.0001v1 のようにバージョン番号を付加する。一般的形式は arXiv:YYMM.NNNNvV であり、YY は 2 桁の年号 (07=2007 年から 99=2099 年まで、06=2106 年までとする可能性もあり) である。」

[参照 : http://arxiv.org/help/arxiv_identifier]

1991 年から 2007 年 3 月までの識別子は、以下の形式に従っている :

アーカイブ.主題クラス (該当する場合) /YYMM 番号 (例 : math.GT/0309136)

これら識別子の一部 (URN、ハンドル、PID) は認知されている持続的識別子だが、そうでないものもある (もともと、リポジトリのベース URL が同一である限りは持続的だが)。これら識別子はどれもリポジトリの内部でのアイテムの信頼できる特定とその読み出しのために重要である。しかし、複数のリポジトリ間でのアイテムの容易な特定という観点からは、その利便性は限定的である。

幸いなことに、ジャーナル記事については、異なるホスト間で信頼できる形でアイテムを対応づけるために使用できる可能性のある世界的な識別子が存在する。

DOI

DOI は、出版界で記事の特定に世界的に使用される識別子として最も普及しており、どのリポジトリソフトウェアも DOI を記録する機能を持つ。机上研究と電子メール調査の結果、多くのリポジトリでは、リポジトリが保管するアイテムを出版者のウェブサイト上の記事にリンクする DOI を付加している (DOI が利用可能で、時間的にそれが可能な場合)。

一例として、2009 年 1 月 14 日時点で、Cranfield CERES はジャーナル記事に対応する 707 個のアイテムを保持していた。これら 707 個のアイテムのうち、468 個 (66%) のメタデータに DOI が含まれていた。

これは希望を感じさせる状況だが、リポジトリが保管する記事の中で DOI が含まれる割合を高めるために何らかの作業が必要である。この点については、CrossRef がツールを提供していることを指摘しておきたい。Simple Text Query と呼ばれるこのツールは <http://www.crossref.org/SimpleTextQuery/> で利用することができ、テキスト引用や引用リストから DOI を得るために役に立つものである。

DOI がジャーナル記事以外のリソースにも利用できることも付記しておきたい。従って、リソースタイプも一意的に特定できることが不可欠である。

記事特定のためのソリューション

すでに述べたように、Dspace のシステムでは「タイプ」フィールドが多目的に使用される傾向があり、リソースタイプ（記事）、学術ステータス（審査・査読付き）、出版ステータス（提出済み、出版済み）といった情報が 1 つのフィールドに詰め込まれている。これに対して Eprints では、これらを別個のフィールドとしている。

PIRUS チームは、記事のリソースタイプ、バージョン情報、査読ステータスを、別個のメタデータ要素として提示するとともに可能な場合には DOI を付加するという運用を採用することをすべての IR に奨励するよう提案する。

リソースタイプ

リソースタイプに対して提案する値は、「article（記事）」または「journal article（ジャーナル記事）」である。

記事のバージョン

プロジェクトチームによる決定では、受理された原稿とそれ以後のバージョンのみを利用カウンターの対象とする。これは、記事が公式の学術記録に含まれるのがジャーナルでの出版が受理された時点からだからである。また、プロジェクトチームでの合意として、PIRUS は JISC VERSIONS プロジェクト

(http://www.lse.ac.uk/library/versions/VERSIONS_Toolkit_v1_final.pdf) が使用している用語と一貫性を持たせる。このプロジェクトでは、記事の変遷に 5 つの主要段階が定義されている。また、記事のバージョンについて最近合意された NISO/ALPSP 提案

(<http://www.niso.org/publications/rp/>) と一貫性を持たせることとする。この提案では、ジャーナル記事に 7 つの段階が定義されている。

表 1：記事出版の段階

NISO/ALPSP の定義	VERSIONS の定義	PIRUS プロトコル
著者オリジナル (AO)	ドラフト	カウントせず
審査中の提出原稿 (SMUR)	提出バージョン	カウント対象
受理済み原稿 (AM)	受理バージョン	カウント対象 (バージョン A)
プルーフ (P)		カウント対象 (バージョン A)
記録バージョン (VoR)	出版バージョン	カウント対象 (バージョン B)
修正記録バージョン (CVoR)	改訂バージョン	カウント対象 (バージョン B)
拡張記録バージョン (EVoR)	改訂バージョン	カウント対象 (バージョン B)

しかし、PIRUS の目的のためには、NISO/ALPSP と JISC のいずれの定義であれ、そのすべての段階での利用を別個に記録、報告する必要はないと合意された。利用度を知るためには、受理された原稿やプルーフ原稿の利用と記録バージョンの利用とを区別することが望ましいだろう。様々なバージョンを大まかに区別するこれら 2 つのカテゴリー（表 1 第 3 列のバージョン A と B）の利用度については個別記事ごとに別個に記録、集約、報告するようにする

ことが好ましいが、当面の間はほとんどの出版者とリポジトリにとって現実的ではない可能性が高い。当面は、**A**と**B**を合わせた利用レポートが許容される。

解決できていない問題として、この情報を提示するためにどのメタデータ要素を使用すべきかという問題が残されている。標準は存在してしない。

査読ステータス

これについても、解決できていない問題として、この情報を提示するためにどのメタデータ要素を使用すべきかという問題が残されている。標準は存在してしない。

主題リポジトリ

この調査では機関リポジトリのみを対象としたが、記事のオンライン利用に占める主題リポジトリの割合が増加していることをプロジェクトチームは認識していた。**PubMed Central**は、主題リポジトリの重要な一例である。従って、そのような主題リポジトリに対して**PIRUS**を周知すること、そして今後生まれる新たな標準が主題リポジトリにとって意味を持つものとなり主題リポジトリの支持を得るものとなるようにすることが重要である。本プロジェクトの過程では、3つの主要な主題リポジトリ（**PubMed Central**、**ArXiv**、**Social Science Research Network**）に諮問し、以下のようなフィードバックを得た：

- **PIRUS**が提案する技術アプローチが問題となるリポジトリはなかった。
- 2つのリポジトリはプライバシーの問題に懸念を持っていた。ただしこれは記事の利用者個人が特定できる場合のことで、本プロジェクトではそのような提案は行っていない。
- 3つのリポジトリはいずれも、**PIRUS**が行う次の段階についての情報の連絡を希望しており、理想的には関与していくことに関心を持っている。

フェーズ 2 (2008年9～11月)：個別記事の利用の記録および報告についての利用レポートとプロトコルのドラフトを策定し、これを出版者とリポジトリが持つ利用データでテストした。

出版者の状況

オンラインジャーナル出版者の多数はすでに**COUNTER**実務指針に準拠してオンライン利用統計をジャーナルレベルで提出しているため、個別記事レベルでオンライン利用統計を提出するために次のような段階的行動が可能なことは明らかである：

- すべての出版者が一意的な記事識別子として**DOI**を使用することを義務づけること。
- すべての出版者において標準化された形で**DOI**を実装して出版プロセスの同一段階で記事の恒久的属性として付与することを義務づけること。
- 利用をカウントできるようにする記事のバージョンを指定すること。これは**PEER**プロジェクト(**REF**)の定義による第2段階(ジャーナルでの出版が受理された著者原稿)と第3段階(最終的出版原稿)となる可能性が高い。
- 個別記事レベルでの利用の記録と報告のためのフォーマットと関連するプロトコルを規定すること。(下記セクション7の記事レポート1を参照)

リポジトリの状況

利用統計に関するリポジトリの状況は、昨年大きな注目を集めた領域である。PIRUS にとって特に重要な点は、以下の 2 つの報告が取り上げた調査結果である：

- The JISC Usage Statistics Review Project(1)
- The Knowledge Exchange Institutional Repositories Workshop Strand on Usage Statistics (5)

どちらのプロジェクトも、リポジトリソフトウェア間の多数の相違点に対応するためにリポジトリログファイルについての基礎的なフォーマットやスキームが必要となるという点で一致していた。このため、どちらのプロジェクトも、OpenURL コンテキストオブジェクトを XML 形式で提示することを、標準化されたフォーマットの理想的候補の 1 つとして提案している。OpenURL の最も一般的な利用法は、適切なバージョンのリゾルブ手段を提供することであるが、ここでの目的はそうではないことに注意しておきたい。ここではむしろ、統計目的に必要なメタデータを保持して提示する機能をすでに有している既存の受け入れられた標準フォーマットを活用し、恣意的な標準をまったく新たに作り出すことを避けることが主旨である。OpenURL を使用したリンクのリゾルブは複雑な問題だが、OpenURL の構築自体は単純な作業である。

PIRUS チームは、この提案に完全に合意する。しかし、現在使用されているリポジトリソフトウェアの多様性とその運用方法における違いを考慮すると、すべての状況で実用性を持つ単一のアプローチを提案することは実際的ではない。例えば、現在のところ、フルテキスト記事がダウンロードされる時には、以下のようになる：

- DSpace では、java のサーブレット (BitstreamServlet.java) が呼び出され、それが要求されているファイルを返し、DSpace のログにエントリーを生成する。
- Eprints では、Perl のモジュールが呼び出され、それが人間にとって整った URL を内部的に扱いやすい形式に書き替え、それが要求されているファイルを返す。そして、データベースアクセスログにエントリーが生成される。

こうした違いを克服するための実際の解決方法として、プロジェクトチームは以下のシナリオ (下記図 1) を提案する。この提案では、標準化された利用統計を生成するために 3 つの経路が存在し、想定されるリポジトリの状況をほとんど網羅する。

	ステップ 1：フルテキスト記事がダウンロードされる。	
	ステップ 2：トラッカーコードを呼び出し、OpenURL のログエントリーを生成。	
シナリオ A	シナリオ B	シナリオ C
ステップ A1：OpenURL のログエントリーを利用統計の作成と集約を担当する外部当事者に送付。	ステップ B1：OpenURL のログエントリーをローカルサーバーに送る。	ステップ C1：OpenURL のログエントリーをローカルサーバーに送る。
<u>ステップ A2：COUNTER ルールによるログのフィルタリング。</u>	<u>ステップ B2：OpenURL のログエントリーを利用統計の作成と集約を担当する外部当</u>	ステップ C2：COUNTER ルールによるログのフィルタリング。

	<u>事者がハーベスティングする。</u>	
<u>ステップ A3:COUNTER 準拠利用統計を収集し、記事 (DOI) ごとに XML フォーマットにまとめる。</u>	<u>ステップ B3 : COUNTER ルールによるログのフィルタリング。</u>	ステップ C3 : COUNTER 準拠利用統計を収集し、記事 (DOI) ごとに XML フォーマットにまとめる。
<u>ステップ A4 : 承認を受けた当事者が COUNTER 準拠利用統計を中心的組織から入手可能になる。</u>	<u>ステップ B4 : COUNTER 準拠利用統計を収集し、記事 (DOI) ごとに XML フォーマットにまとめる。</u>	ステップ C4 : 承認を受けた当事者が COUNTER 準拠利用統計を中心的組織から入手可能になる。
	<u>ステップ B5 : 承認を受けた当事者が COUNTER 準拠利用統計を中心的組織から入手可能になる。</u>	

図 1 : リポジトリでの利用統計の記録と報告に対するアプローチ案

図 1 の各段階のうちで文章に下線が引かれていないものは、リポジトリをホストしている機関の中で行う。文章に下線が引かれているものは、外部の当事者が処理する。

ステップ 1

上記図 1 に示されている各シナリオは出発点が共通で、エージェント（例えば、人間のユーザーやロボット）が記事のダウンロードのためのリンクにアクセスすることから始まる。

ステップ 2

ダウンロードのイベントが、利用するソフトウェアに応じて異なる活動を引き起こす。例えば以下ようになる：

- **DSpace** では、フルテキスト記事のダウンロードのために **java** のサーブレット (**BitstreamServlet.java**) が呼び出された時、サーブレットは **DSpace** のログエントリーだけでなく **OpenURL** コンテキストオブジェクトのログエントリーも生成する。
- **Eprints** でもデータベースアクセスログにエントリーが生成されるが、ポーリング方式で実行されている別のスクリプトが最新のログエントリーを確認し、フルテキスト記事のダウンロードに対して **OpenURL** コンテキストオブジェクトのログエントリーを生成する。

その他のソフトウェアについても詳細は異なるが、結果は同じことになる。つまり、いずれの場合もフルテキスト記事のダウンロードについては **OpenURL** コンテキストオブジェクトのログエントリーが生成される。

関係する組織や機関によって異なる要件や能力に対応し、使用されているソフトウェアの多様性にも対応するため、**OpenURL** の処理に対して 3 つのシナリオが提案されている。以下に簡単に説明する：

シナリオ A

シナリオ A では、生成された **OpenURL** が第三者当事者（中心的クリアリングハウス）がホストするサーバーに直接送信される。この方法は **Google Analytics** のモデルと類似している

が、**Google Analytics**とは異なり、**JavaScript**ではなくサーバーサイドのコードが使用される。**JavaScript**では、**PDF**などの様々なファイルタイプのログ記録の問題に十分対応できないからである。

受信した外部当事者は、エントリーを **COUNTER** ルールに従ってフィルタリングし、承認を受けた当事者の利用に供することができる **COUNTER** 準拠レポートを作成する。

プロジェクトチームは、このシナリオを **DSpace** (克蘭フィールド大学にて) と **Eprints** (サウサンプトン大学にて) の両方でテストした。いずれの場合も、ログエントリーを **BioMed Central** のサーバーに送信することに成功した。

シナリオ B

シナリオ B では、生成された **OpenURL** のエントリーが機関内でローカルにホストされているサーバーに送られ、それがこれらのエントリーを **OAI-PMH** 経由で公開し、外部の第三者当事者がハーベスティングする。

受信した外部当事者は、エントリーを **COUNTER** ルールに従ってフィルタリングし、承認を受けた当事者の利用に供することができる **COUNTER** 準拠レポートを生成する。

プロジェクトチームはこのシナリオをテストしていないが、**OAI-PMH** は機関リポジトリの間で十分に理解されており、実装は比較的単純であろう。

シナリオ C

シナリオ C でも、生成された **OpenURL** のエントリーが機関内でローカルにホストされているサーバーに送られる。しかしこのシナリオでは、エントリーのフィルタリングと **COUNTER** 準拠レポートの作成に必要な処理のすべてがローカルに行われる。

作成された **COUNTER** 準拠レポートは、**SUSHI** 経由で承認を受けた当事者の利用に供することができる。

プロジェクトチームは、克蘭フィールド大学でインストールされている **DSpace** を対象としてこのシナリオのプロトタイプとテストを行った。テストの結果、**COUNTER** 準拠の **XML** レポートを作成することができることが実証され、その結果は <http://cclibweb-1.dmz.cranfield.ac.uk/pirus/AR1.xml> で見ることができる。(このシナリオを完結させるために残された作業は、**SUSHI** のラッパーをレポート周りに追加することだけである。残念なことにこの作業は本プロジェクトの期間内には不可能だったが、十分に完了できる作業である。)

図 1 に説明されているシナリオによって大多数のリポジトリに対応することができる。リポジトリは、その 1 つを実装することで個別記事のオンライン利用統計を生成することができ、それは出版者が生成するものと整合性のあるものとなるだけでなく、集約することで記事ごとの合計オンライン利用数を得ることもできるようなものとなる。

チームは、図 1 のシナリオで使用されるトラッカーコードが作成する **OpenURL** コンテキストオブジェクトが、理想的には以下の情報を含むものとすることを提案する：

- アクセスの日付と時間
- DOI
- 記事フォーマット (例えば、**html** や **pdf**)
- 記事のバージョン (ただし、現時点ではすべてのリポジトリがこのデータを提供できる状況にはない)

- ・不正利用データ（不正利用の検出と排除のために収集されるデータ）
 - ・IPアドレスとそのハッシュ
 - ・ユーザーエージェント
 - ・その他同一セッション中に発生した活動

しかし、関係組織に対して強調しておきたい重要なこととして、IPの収集は統計のみを目的とするもので、そのデータは安全に保管され、外部に公開することはない。（この点は、製薬企業や多くの米国政府部局にとって特にデリケートな問題となる。）このような懸念にはIPをハッシュすることで対応し、利用パターンを検出可能としつつも実際のIPは匿名化される。

COUNTERが幅広く導入されてきた理由の1つに、参加する出版者に対して資金面、組織面で過度の負担がかからないということがある。この思想をリポジトリにも拡大することで参加を促すことが重要である。このため、リポジトリには、ISSNのようなフィールドを埋めることは義務づけられない。ジャーナル単位での利用の集約は主として出版者が関心を持つことであり、このような情報は必要に応じて「下流」工程で出版者自身が付与することができる。

中心的クリアリングハウスについての要件

図1に説明したシナリオから明らかなことだが、中心的クリアリングハウスには、リポジトリから利用データを受け取り、結果としての利用統計を生成し、出版者が自身の利用統計と合わせて集約することで合計利用数を得ることができるようになることが要求される。現段階で具体的な組織を特定することは尚早だろうが、プロジェクトチームとしては、既存の確立された組織の中に、要求される能力のほとんどをすでに有しているところがあると考えている。中心的クリアリングハウスは、最低要件として以下の基準を満たす必要がある：

- ・独立性を持ち、主要な関係集団（著者、図書館、出版者、助成機関）から信頼されていること。
- ・関係するメタデータを受領、蓄積、処理し、利用統計を生成する能力の実績があること。
- ・大量のメタデータと利用統計を扱うことができること。

7. 成果物

本プロジェクトの4つの主要な成果物は、以下の通りである：

- 個別記事利用報告「記事レポート1：フルテキスト記事ダウンロード成功数」用 COUNTER準拠XML概念実証プロトタイプ（付録E）。リポジトリと出版者が共通に利用できるもの。このプロトタイプはCOUNTER実務指針リリース3(6)と整合性があり、<http://cclibweb-1.dmz.cranfield.ac.uk/pirus/AR1.xml>で閲覧できる。理想的には、このレポートは、著者個人と機関の両者に対して作成できる。しかし実務的には著者個人についてのレポートは生成が容易で短期的に現実性のある目標となる一方、機関など（例えば、助成機関）についてのレポートはずっと複雑であり、長期的な目標と見なすべきである。
- トラッカーコード。リポジトリが実装する。利用統計の作成と集約の責任および集約を行う関係出版者への送付の責任を担う外部当事者へ、またはローカルのリポジトリ・サーバーへ、メッセージを送る。

- c. 個別記事利用統計の作成、記録、および集約のための様々なシナリオ（上記図 1 を参照）。現行のリポジトリの大部分をカバーする。各リポジトリは、自組織の技術と実装に合ったシナリオを選択することができる。
- d. 必要に応じて（一部のカテゴリーのリポジトリについて）利用統計を作成し、他者のために利用統計を収集、集約する中心機構についての要件の決定。

こうした成果は、本プロジェクトの当初の目標と整合的であり、リポジトリの間に現在存在している技術面や組織面の構成の多様性が考慮されている。COUNTER によって導入されれば、情報源を問わず同一基準に従って個別記事のオンライン利用統計の作成、報告、集約を世界レベルで行うことが可能となる。

その結果、COUNTER データをより細かいレベルで大幅に充実させ、著者、出版者、機関、助成機関に対して初めて個別記事レベルでの比較可能な利用統計を提供することができるようになる。

8. 結果

本プロジェクトは、オンライン利用の測定に関して出版者とリポジトリの両者に適用される標準を定めようとする初の試みである。個別記事の利用の記録と報告のためにそのような測定の標準を定めようとする初の試みでもある。これは COUNTER によってすでに行われた進歩なしには不可能だったはずであり、本プロジェクトは新たな領域を開拓した。技術面と組織面で現在多様な構成が見られるリポジトリに対する標準策定に伴う実務的な課題について、重要な教訓も得た。

PIRUS が立てていた具体的な目標は達成され、実際のところ、超過達成することができている。しかし、PIRUS が行ったことは、リポジトリの環境が極めて多様な現状においても個別記事についての標準化されたオンライン利用統計を生成することが技術的に可能であるという証明にすぎない。組織面、知的財産面、政治面の諸問題には、まだ完全には対応できていない。また注意すべき点として、フェーズ 1 で行われた調査によれば、すべての出版者とすべてのリポジトリが個別記事レベルでの利用報告の理念を熱心に支持しているわけではないことが明らかになっている。

9. 結論

個別記事利用統計は、研究とその成果の伝達にかかわる重要な関係者にとって価値ある道具となる可能性を持つ。具体的には以下の通りである：

- ・ 研究者／著者。自身の出版物のオンライン利用をモニターし、その意味を理解することに関心を持つ。
- ・ リポジトリ。保持しているアイテムの利用状況に関心を持つ。そうしたアイテムを利用に供することの価値評価の一助とするため、およびリポジトリに対する投資の費用効果を立証するためである。
- ・ 研究機関。研究資金を得るために競争し、研究や自組織の研究者の価値を立証するための研究評価行為などからの圧力にさらされている。
- ・ 助成機関。助成している研究プロジェクトの実績や影響を評価するために定量性と透明性の高い方法を求めている。

著者／研究者ごとに自身の出版物についての個別記事利用統計を提供することは比較的単純だが、そうした利用統計をリポジトリ、研究機関、助成機関についての利用統計として集約することは相当に困難であり、いくつもの技術的（例えば、データの量）、組織的（例えば、スケーラブル性や集約）、経済的（例えば、費用の配分）、政治的（例えば、秘匿性）問題の解決に時間を要する。

本研究の結果として、大きく以下のような結論を導くことができる：

- ・利用を計測するための共通の技術標準を、組織的・技術的環境に多様性がある中でもリポジトリと出版者に対して定めることができる。
- ・この実現可能性研究の結果を基に、すべての関係者集団が実行可能な現実的で実装可能な解決策をまとめるためには、さらなる研究が必要となる。

10. 含意

本研究は、COUNTER およびより広い関係者集団に対して、以下のような含意を有する：

- COUNTER**に対して：**COUNTER**実務指針がさらに改善、拡張される。既存の**COUNTER**実務指針は、専ら出版者やベンダーを対象に設計されている。本プロジェクトの成果を**COUNTER**がさらに発展させて採用することで、リポジトリに対して**COUNTER**が定める最初の標準ができる。**COUNTER**の戦略的役割をこのように拡大するためには、現行の実務指針をレポートの追加や監査の拡大という形で修正することが必要となる。
- リポジトリに対して：リポジトリの間には利用統計を対象とする共通の標準はほとんど存在しないが、リポジトリには利用統計の作成と、さらには公開さえもが要求されている。これらに信頼性を与えるためには、共通の認知された標準に準拠して作成することが必要である。リポジトリには、**COUNTER**が定めるものであれそうでないものであれ、そのような何らかの標準を採用することがメリットとなるだろう。
- 著者に対して：個別記事レベルでの信頼できて透明性のある世界的な利用統計によって新たな指標が提供され、それによって自身の研究成果がどのように利用されているかを知ることができるようになる。
- 出版者、ベンダーに対して：著者の間に自身の記事についての信頼できる世界的な利用統計の入手が可能であることが知られば、そのようなデータの利用が望まれるようになり、当該プロセスに参加するよう出版者に圧力がかかる可能性が高い。個別記事利用統計を提供することは、出版者が著者との関係をさらに強固なものとする機会となる。個別記事レベルでの利用を報告するためのどのような要件も、ベンダーが自身のDOIの実装を標準化することや記事の異なるバージョンを明確に定義して特定することなどの必要性を高めるものとなるだろう。
- 助成機関に対して：研究評価に使用する現行の指標は、引用ベースの比重が非常に高い。UKSGの主導で行われたジャーナル利用ファクター(7)の初期の結果では、例えば英国研究評価フレームワーク (REF) (2)などにおいて引用ベースの指標の補完として利用ベースの指標を使用することへの支持が、著者や出版者の間に幅広く存在することが示されている。個別記事に対する世界的レベルでの信用できる利用統計が入手できることは、助成機関に対しては、研究成果の影響の指標として利用度を考慮するような圧力がさらに増加することにつながる。

- f. 研究機関に対して：改訂REFの中に 1 つの指標として個別記事利用統計が含められれば、研究を中心とする機関には、所属する著者についてのそのようなデータを収集して報告することが要求されるようになるだろう。
- g. データプロバイダに対して：例えばダブリンコアの場合のように、メタデータの定義と蓄積のための標準的方法が要求されるようになる。
- h. 業界全体に対して：個別記事についての利用統計を世界的に集約して報告するとすれば、データを集中的に収集する必要があり、これを行う機能を支える必要がある。業界全体としては、理念としてそのような機能（既存組織の拡張となる可能性もある）を支援することを望むか否かを決定しなければならない。この決定を行うためには、技術面と組織面の仕様を付随する費用と共に詳細に作成する必要がある。

11. 提案事項

本プロジェクトチームの提案は、以下の通りである：

- a. JISCに対して：PIRUSは、リポジトリおよび出版者からのデータを用いて個別記事の利用統計を作成、記録、集約することが技術的に可能であることを実証した。これを基に実装可能な新たなCOUNTER標準およびプロトコルを作成するためには、さらなる研究と開発が具体的に以下に示す領域で必要となる：
 - ・ 技術面：提案されているプロトコルとトラッカーコードにスケーラブル性、拡張性があり、主要なリポジトリの環境で問題が生じず、記事以外のアイテムにも応用可能であることを確認するために、より多くの多様なリポジトリと大量のデータによるテストが必要である。
 - ・ 組織面：中心的クリアリングハウスの性格と使命を明確化し、候補組織を特定、テストする必要がある。
 - ・ 経済面：必要となる利用レポートを生成する上でリポジトリと出版者にかかるコストおよび中心的クリアリングハウスのコストを評価し、こうしたコストを関係者の間で配分する方法について検討する。
 - ・ 政治面：すべての主要関係者集団（リポジトリ、出版者、著者）による幅広い支持が必要となる。PubMed Centralのような主題リポジトリからは本プロジェクトの現段階においては積極的な関与を得ていないが、今後参加を得る必要がある。知的財産、プライバシー、財務面の問題についても対応する必要がある。

PIRUS プロジェクトチームは、上述した諸問題に対処するためのさらなる研究を助成することの検討を JISC に提案する。

- b. COUNTERに対して：COUNTERの使命を拡大してリポジトリからの利用統計を含めるようにすること。新たな記事レポート 1 を任意追加レポートとして導入することを検討すること。独立監査を修正して新たなレポートやプロセスをカバーすること。
- c. リポジトリに対して：主題リポジトリが本プロジェクトの次の段階に参加すること。すべてのリポジトリが記事のバージョンなどに対して標準的データ記述を使用すべきである。
- d. 出版者、ベンダーに対して：個別記事レベルで信頼できる利用統計を提供することが好ましいことを理念として認めること。

12. 参考資料

1. JISC Usage Statistics Review:
<http://www.jisc.ac.uk/publications/publications/usagestatisticsreviewreport.aspx>
2. HEFCE: Research Excellence Framework
<http://www.hefce.ac.uk/Research/ref/>
3. Standardized Usage Statistics Harvesting Initiative (SUSHI)
<http://www.niso.org/workrooms/sushi>
4. Budapest Open Access Initiative, Institutional Repository Software
<http://www.soros.org/openaccess/software/>
5. Knowledge Exchange Institutional Repositories Workshop Strand on Usage Statistics:
http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2Fdownloads%2FIR+workshop+1617+Jan+2007%2FNew+reports%2FKE_IR_strand_report_Usage_Statistics_Sept_07.pdf
6. COUNTER Code of Practice for Journals and databases, Release 3
<http://www.projectcounter.org/r3/Release3D9.pdf>
7. Usage Factor project <http://uksg.org/projects>

13. 付録

付録 A : 出版者の調査

付録 B : リポジトリソフトウェア一覧表

付録 C : 英国の研究機関および部局のリポジトリ

付録 D : JISC Repositories 電子メール調査の回答 (概要)

付録 E : 記事レポート 1 の XML レポート (単一記事用) の例

付録 A : 出版者調査の結果

アンケートを 15 の出版者・ベンダーに送付した。合計で 12 の回答があり、回答は書面によるか電話での聞き取りによって得た。回答のあった組織には、下記の文中で星印を付した。

調査対象となった出版者・ベンダー：American Chemical Society*、American Institute of Physics、Atypon*、BioMed Central*、EBSCO、Elsevier*、Ingenta*、Institute of Physics Publishing*、Nature Publishing Group*、OUP*、Ovid*、Sage*、Springer*、Taylor & Francis*、Wiley Blackwell。これらの出版者・ベンダーは、すべて現在 COUNTER に準拠しており、調査対象が領域、規模、地理的所在地の面で業界の代表的サンプルとなるように選択した。

個別の質問に対する回答

質問

1) 理念として個別記事レベルでフルテキスト記事の利用状況を記録して報告することに価値があると思いますか。

多数の回答者はそうすることに理念として価値があると考えているが、ほとんどの回答者は実務上の問題も予想している。ベンダーの中には、個別記事レベルでの利用をすでに記録して報告していると述べているところや、そうする能力を持っていると述べているところもある。若干のベンダーは、このレベルの情報が大多数の顧客にとって有用とも必要とも考えていない。

a. 理念として、そのような利用を計測するための共通の標準を出版者、アグリゲータ、リポジトリ向けなどに規定することに価値があると考えますか。

理念的にそのような標準に価値があるとする回答がほとんどだったが、多様な情報源からの利用データをまとめることについての実務上の問題を予想し、それを実行する際の実際的な仕組みを導入することが可能かどうかに疑念を呈する回答がいくつもあった。

上記の最初の質問に「いいえ」と回答したところは、共通の標準の作成に価値を見出していない。

2) 一意的な記事識別子

a. フォーマット (PDF、html など) にかかわらず特定のフルテキストのジャーナル記事に同一の一意的な記事識別子を付与していますか。

すべての回答者が、形式にかかわらず 1 つの記事には同一の一意的な記事識別子を付与している。

b. それは、その記事の恒久的な属性ですか。

すべての回答が「はい」。

3) もしそうならば、その目的のためにDOIを使用していますか。そうでなければ、現状の運用慣習はどうなっていますか。

この目的のために DOI を使用しているという回答がほとんどであるが、すべてではなかった。例えば EBSCO の場合、フルテキスト・データベース中の多くの記事に DOI が存在しないため、内部的な受入番号 (Accession Number) を代わりに使用している。

4) 出版プロセスのどの段階で一意的な記事識別子を付与していますか。出版前と出版時など、出版プロセスの異なる段階で異なる記事識別子が付与されますか。

様々なシナリオがある：

- ・最終的に受理された著者原稿に DOI が付与される (最も一般的)
- ・コンテンツがコンテンツ管理データベースに追加された時に DOI が付与される。
- ・オンライン出版時に DOI が付与、登録される。編集プロセスや制作プロセスでは内部的な識別子を使用する。
- ・ある 1 つのケースでは、記事が出版者の制作トラッキングシステムに組み入れられた時点で内部コードが付与される。記事が印刷媒体よりも前にオンラインで出版された時には、これを基に DOI が付与され、この時点で DOI が CrossRef に登録される。その後冊子の中に収載された時には、記事は同一の DOI を維持し、冊子の目次に移動され、CrossRef 中の関連するメタデータが修正されたページ番号や巻番号などが記録される。

いずれの場合でも、1 つの原稿／記事のすべてのバージョンには単一の DOI が使用されている。いくつもの出版者が、この点は CrossRef の指針に従うために要求されていることだとコメントしている。

5) フルテキストのジャーナル記事は、どのフォーマットで出版していますか。

以下のフォーマットが使用されている：html、PDF、enhanced PDF、provisional PDF、SGML (2009 年より後は使用せず)、XML

6) 記事の異なるバージョンは、現在どのように区別していますか。

以下のように、多様な慣習が存在する：

- ・@just-accepted や @ahead-of-print を DOI に付加することで、内部履歴を管理する。こうした DOI は、外部の目に触れることはない。
- ・各バージョンへのアクセスログを別に記録する。
- ・フォーマットフラグを作成する。
- ・最終的な出版されたバージョンのみがユーザーに利用可能となるため、バージョンの識別子は必要がない。
- ・(i) 著者原稿のオンライン出版の後に(ii)組版、編集済みだがページ付けされていないバージョンが出版される。その後、(iii)最終的な冊子バージョンが出版される。以前のバージョンは最新バージョンであるページ付けされた冊子バージョンにリゾルブされる。多くの HireWire の出版者と同様、記事の「コンテンツボックス」中に以前のバージョンが記載される。

- ・ 出版の段階が記事のメタデータ中に記される (xml A++ DTD)。

b) 最近出版されたNISO/ALPSP Best Practices for Journal Article Versions

(<http://www.niso.org/publications/rp/RP-8-2008.pdf>) を導入する予定はありますか。この文書は、記事の様々なバージョンを特定する方法や、その相互関係を明確化する方法について提案しています。

ほとんどの回答者はこの新しい指針について知らなかったが、使用しているソフトウェアで複数のバージョンを扱うことができ、その識別も可能だと考えている。しかし、2つの出版者はこの指針の導入を計画していると明確に述べた。

7) 成功したフルテキスト記事要求数を記事レベルで報告するという新たな報告要件を

COUNTERが定め、各フルテキスト記事にフォーマットにかかわらず恒久的属性を与えるための一意的な記事識別子としてDOIが義務づけられた場合、問題が生じますか。(現在のCOUNTERでは、ジャーナルレポート1によってフルテキスト記事の利用をジャーナルレベルで報告することのみを要求しています。)

これはジャーナルの出版者にとって理想的には問題ではないが、作成する利用レポートを増やすことに伴う作業量の増加を懸念する出版者が多い。これについては、従来のCOUNTERレポートのように毎月機関別に作成する方式を踏襲せず、記事レベルのレポートについては要求があった時にタイトルレベルでのみ作成することとすれば、問題が軽減される。

技術的な課題には、(i)異なるバージョンの利用を合わせてまとめること、および(ii)URLの構造中にDOIが含まれず、ログファイルに含まれるURLを関連するDOIに対応づけることが求められる場合や、ウェブサイトの構造が変更された場合がある。

フルテキスト・アグリゲータにとっては、大量のコンテンツがジャーナルやその他のタイプの資料から生まれ、それらについてはオンライン上に対応するものが存在しない。このような場合、DOIは登録されていない(そして、多くのアグリゲータが同一のコンテンツを処理しているため、アグリゲータがDOIを登録することは合理的ではない)。要するに、DOIを要求しても、本質的に言って、そのようなレポートはアグリゲータのサービスによって供給されるすべての記事を網羅できないことになる。

8) その他のコメントをお書きください。

- ・ 「PMCやIRのような情報源からのそのような利用データを1つの場所に集めた方がよいということに、賛成します。」
- ・ 「こうしたレポートの背後にある理論は正当ですが、そのようなレポートの現実性には注意深い検討が必要です。毎月数百万件のトランザクションを操作して何カ月にもわたって履歴を維持するための処理能力と記録装置についても検討する必要があります。検討すべき1つの問題として、この報告のために出版者が3年分の記事レベルのトランザクションを維持することが期待されているのでしょうか。それとも、出版者側は記事レベルのトランザクションのフィードを提供し、それを受け取る機関が集約と分析を行うということが意図でしょうか。後者の場合、出版者には事後1、2カ月の間フィードを利用可能とすることが要求されるということでもよいでしょう。後者の場合はまた、情

報を受け取る組織がデータの蓄積と分析のための技術を用意する責任を持つことが前提となります。」

- 「私たちの顧客にとっての記事レベルの報告の有用性には、確信が持てません。大量のデータに圧倒されてしまうでしょう。私たちは中規模の出版者で、**30万**の記事を抱えています。レポートの追加によるコストの増加を正当化するためには、顧客にとってのメリットについてさらに詳しく知りたいと思います。」
- 「機関単位の利用から記事へ変更することは、莫大な量のデータを生みます。これを要件とする前に、実装コストと維持コストの見積もりを含む徹底的な費用対効果分析を行うことを提案します。」

付録 B : リポジトリソフトウェア一覧表

ソフトウェア	研究機関または部局のリポジトリの数 (ROAR より)		メタデータ		備考
	英国	世界	サポートするスキーマ	拡張可能	
Archimede	-	-	限定子付きダブリンコア	不可	
Arno	-	3	ダブリンコア	可	
CDSware	-	6	標準 MARC21	可	
Digital Commons (BePress)	4	50	ダブリンコア	不可	QDC をサポートする開発が進行中。
DigiTool	-	1	ダブリンコア	可	必要に応じて QDC の記録を作成するために使用できる「自由記述」の追加フィールドがある。
DiVA	-	16	内部方式	可	スウェーデン。内部方式のスキーマは、marcxml や QDC など、いくつもの標準的スキーマに対応づけることができる。
DSpace	23	222	限定子付きダブリンコア	可	初期設定では QDC だが、v1.4 以後はその他のメタデータスキーマもサポート可能。
EDOC	-	1	内部形式	可	ドイツ。vCard、DC、OpenURL、AMF、LOM、Ariadne、ODRL、OAI、CLD といった国際標準に基づいたメタデータ。
Eprints	39	163	内部形式	可	内部フィールドが通常 QDC に一対一で対応づけられている。

Fedora	1	5	任意	可	
Fez/Fedora	-	3	MODS、ダブリンコア	可	統計機能あり（例えば、著者別、コミュニティー別、コレクション別、主題別のダウンロード数）
HAL	-	3			フランス。投稿者が関係する書誌情報と DOI を明示することが奨励されている。

ソフトウェア	研究機関または部局のリポジトリの数 (ROARより)		メタデータ		備考
	英国	世界	サポートするスキーマ	拡張可能	
i-Tor	-	1	任意	可	
MyCoRe	-	3	限定子付きダブリンコア	可	ドイツ
Open Journal System	-	1			
Open Repository	4	7	限定子付きダブリンコア	可	BioMed Central によるホスト型 IR サービス。最新の DSpace ソフトウェアに基づく。
OPUS	-	23	限定子付きダブリンコア	可	ドイツ
Other	3	88	様々		
	74	596			

PIRUS—出版者・機関リポジトリ利用統計

PALS 3 プロジェクト

付録 C : 英国の研究機関および部局のリポジトリ

機関	リポジトリ	ソフトウェア	引用	DOI	タイプ	ステータス
リンカーン大学	Institutional Repository	Eprints	あり	あり	記事	出版済み○ 印刷中○ 提出済み× 未出版○
バークベック	Birkbeck ePrints	Eprints	あり	あり	記事	出版済み○ 印刷中○ 提出済み○ 未出版×
バーミンガム大学	ePrints Repository	Eprints	あり	あり	記事	出版済み○ 印刷中○ 提出済み○ 未出版○
ボーンマス大学	Bournemouth University Research Online [BURO]	Eprints	あり	あり	記事	出版済み○ 印刷中○ 提出済み○ 未出版○
ブリストル大学	Bristol Repository of Scholarly Eprints (ROSE)	DSpace	なし	あり	ジャーナル記事プレプリント	出版済み○ 印刷中○ 提出済み× 未出版×
ブルネル大学	Brunel University Research Archive (BURA)	DSpace	あり	あり	研究論文プレプリント	
アベリストウィス大学	CADAIR	DSpace	あり	あり	出版済み審査付きジャーナル論文	
ケンブリッジ大学	CUED Publications Database	Eprints	あり	あり	記事	出版済み○ 印刷中○ 提出済み○ 未出版○
チェスター大学	ChesterRep	Open Repository	あり	あり	記事	

イースト・アン グリア大 学	Digital Repository	Digitool	なし。個 別の引用 エレメン トを利用。 。	あり	ジャー ナル記 事	ポストプリント 整 形済み
----------------------	--	----------	------------------------------------	----	-----------------	------------------

シェフィールドハラム大学	Sheffield Hallam University Research Archive	Digital Commons	あり	あり	ジャーナル記事	
サリー大学	Surrey Scholarship Online	Digital Commons	あり	あり	ジャーナル記事	(該当する場合) 本文書は査読を受けている。

PIRUS—出版者・機関リポジトリ利用統計

PALS 3 プロジェクト

付録 D : JISC-REPOSITORIES 電子メール調査の回答 (概要)

2008年9月2日にメーリングリストの JISC-REPOSITORIES に送られた電子メール調査に対する回答の概要。合計 19 の回答を受け取った。

回答した組織

ロンドン・スクール・オブ・エコノミクス

オックスフォード大学

ロバートゴードン大学

ゲント大学図書館

ユニバーシティ・カレッジ・ダブリン

アバディーン大学

ブラッドフォード大学

エジンバラ大学

ハートフォードシャー大学

レスター大学

サウサンプトン大学

セント・アンドリュース大学

スターリング大学

シュトゥットガルト大学

サセックス大学

タスマニア大学

ウォーリック大学

ウルバーハンプトン大学

トゥエンテ大学

各質問への回答

1) 個別の記事に「持続的」な識別子を割り当てていますか。

どの回答者も、個別の記事に持続的識別子を割り当てていると回答した。

2) Q1 への回答が「はい」の場合、どのような持続的識別子ですか。該当するものをすべて挙げてください。

回答者が割り当てている識別子には、Handle (DSpace) 、PURL、URN、ItemID (Eprints) 、UUID (Fedora) があった。

3) Q1 への回答が「いいえ」の場合、どのような識別子を割り当てていますか。

どの回答者も持続的識別子を割り当てていると述べているため、この質問は無関係となった。

4) レコードにDOIを付加する場合、DOIを記録するためにメタデータのどのフィールドを使っていますか (dc.identifier、dc.relationなど) 。

圧倒的多数の回答者が DOI をレコードに付加している (DOI を利用できる場合) 。

この情報を記録するために使用されるメタデータ要素には、以下のように大きな違いがあった：

- dc.description
- dc.identifier
- dc.identifier type DOI
- dc.identifier.citation
- dc.relation.isreferencedby
- dc.rights
- DOI
- relation

5) フルテキスト記事には引用情報を付加していますか。

多数の回答者が、引用情報を付加している、または入力されたメタデータからソフトウェアが引用情報を生成／合成すると回答した。その他の回答者は、時間的に可能な場合に引用情報を付加しているか、まったく付加していない。

6) フルテキストのジャーナル記事のどのバージョンをIRにデポジットしていますか。(該当するものをすべて挙げてください。)

- a) ドラフトバージョン (プレプリントー審査前初期バージョン)
- b) 提出済みバージョン (プレプリントー審査用提出済みバージョン)
- c) 受理バージョン (ポストプリントー審査後出版前編集)
- d) 出版バージョン (ポストプリントー審査後出版者バージョン)

半数強の回答者が、著作権と出版者の方針による制限に従いつつすべてのバージョンをデポジットしていると述べている。5分の1の回答者は、受理されたバージョンや出版されたバージョンのみをデポジットしている。

7) ドラフトや提出済みのバージョンをIRにデポジットする場合、受理されたバージョンや出版されたバージョンが手に入った時にどのようにしますか。

- a) 記事の以前のバージョンを置き換える（上書きする、削除する）。
- b) 既存のレコードに、以前のバージョンに加えて新しいバージョンを付け加える。
- c) 新しいバージョンのために新たなレコードを作成する。

この質問への回答は様々であり、例として挙げたどのシナリオも可能性がある。

8) デジタルリポジトリには、どのソフトウェアを使用していますか。

以下のようなソフトウェアが使用されている：

- DSpace (10)
- Eprints (6)
- Fedora (1)
- Open Repository (1)
- Opus (1)

付録 E : 記事レポート 1 の XML レポート (単一記事用) の例

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE reports PUBLIC "-//ProjectCounter//DTD reports//EN"
"http://www.projectcounter.org/dtd/2004/reports.dtd">
<reports xmlns="http://www.projectcounter.org/ns/2004/reports">
<article_report id="cranfield.ac.uk_12345" cop_version="1" cop_report="1">
<header>
<title>Number of Successful Full-Text Article Requests by Month and DOI</title>
<timestamp>2008-12-01-T10:03:47Z</timestamp>
<vendor>
<vend_name>Cranfield University</vend_name>
<vend_imprint>Cranfield CERES</vend_imprint>
<vend_site>https://dspace.lib.cranfield.ac.uk/</vend_site>
<vend_contact>cranfieldceres@cranfield.ac.uk</vend_contact>
</vendor>
<customer>
<cust_name>Central Agency</cust_name>
<cust_ref>123-4567</cust_ref>
<cust_ip type="cidr">123.456.78.9</cust_ip>
<cust_username>centralagency</cust_username>
<cust_criteria>institution</cust_criteria>
</customer>
</header>
<article_data>
<article doi="10.1016/j.jairtraman.2005.01.007" format="application/pdf">
<requests start="2008-09-01" end="2008-09-30">108</requests>
<requests start="2008-10-01" end="2008-10-31">201</requests>
<requests start="2008-09-01" end="2008-10-31">309</requests>
</article>
</article_data>
</article_report>
</reports>
```