

Christine Merk, Nils K. Windisch

最終報告書

JISC 利用統計レビュー

2008年9月24日

プロジェクト参加機関:

コンスタンツ大学図書館

ゲッチンゲン大学・州立図書館

シュトゥットガルト大学図書館

ザールランド州立・大学図書館

ベルリン・フンボルト大学コンピュータ・メディアサービス

要旨

JISC 利用統計レビュー・プロジェクトは、リポジトリのログファイルに使用する基本スキームの策定と、研究論文や科学リソースなどの電子文書のための有意義で比較可能なアイテムレベルの利用統計を提供するログファイル収集用の標準の提供を目的としている。プロジェクトの一環としてベルリンで開催された関係者によるワークショップにおいて、利用イベントを記述する要素として以下の要素が合意された。

必須要素:

- 誰が: ユーザ/セッションの識別
- 何を: アイテム識別
- 何を: 実行されたリクエストの種類(たとえば、フルテキストやフロントページへのリクエストで、失敗した場合や一部成功した場合を含む)
- いつ: 日付と時間
- 利用イベント ID

任意要素:

- どこから: リファラ/参照体
- サービスの識別

このように記述された利用イベントは OAI を使って OpenURL Context Object 形式で交換されるべきである。自動化アクセス(ロボットなど)はタグ付けされるべきである。自動化アクセスの定義は徐々に改良できるように単純なものでなければならない。ユーザは明確に、しかし、プライバシー法に抵触することを避けるため個人情報を記録することなく識別されるべきである。

利用統計のための標準の策定を行い、将来その実装を実現しようとしているプロジェクトが 2 つ存在する。JISC の助成による PIRUS (Publisher and Institutional Repository Usage Statistics) と DFG の助成による Open-Access-Statistics である。幅広い比較を可能にするために、各国の取り組みは 1 つにまとめられるべきである。中央機関 – たとえば、DRIVER (Digital Repository Infrastructure Vision for European Research) が考えられる – が利用データを集約するべきである。アグリゲータとして、中央機関は複数のコンテンツプロバイダで公開されているアイテムの重複を除外しなければならない。人以外のアクセスのタグ付けもこの機関の仕事である。また、データレコードの品質管理も行う必要があるだろう。

JISC 利用統計レビュー

統計に関するポリシーは、リポジトリコミュニティだけでなく出版コミュニティも考慮して策定されるべきである。統計ポリシーに関する情報は OpenDOAR や RoMEO のようなサービス上で公開されるべきである。

目次

1	はじめに	1
2	現在行われている活動のレビュー	2
3	基本スキームの制定	7
3.1	データソース	7
3.2	基本スキーム	8
3.3	基本スキームの正規化	10
4	利用データの交換	11
4.1	利用データハーベスティング用の共通プロトコル.....	11
4.2	利用データのマッピング	11
5	基本スキームと他の試みとの比較	13
6	ログファイルの記録と集約に関する法的制約	16
7	利用統計ポリシー	17
8	利用統計サービスの今後の展開	19
	参考文献	22
	付録	23

1 はじめに

JISC¹利用統計レビュー・プロジェクトは、リポジトリのログファイルに使用する基本スキームの策定と、研究論文や科学リソースなどの電子文書のための有意義で比較可能なアイテムレベルの利用統計を提供するログファイル収集用の標準の提供を目的としている。

プロジェクトはJISCからの助成を受け、ベルリン・フンボルト大学コンピュータ・メディアサービス、ゲッティンゲン大学・州立図書館、コンスタンツ大学図書館、ザールランド州立・大学図書館、シュトゥットガルト大学図書館が参加する研究コンソーシアムにより実施される。制定される勧告の検証と実装はこのプロジェクトの活動範囲には入っていない。ドイツに限って言えば、その結果は、ドイツ研究協会 (DFG: German Research Foundation)²の助成を受けたドイツ・オープンアクセス統計プロジェクト (OA-Statistics: German Open Access Statistics)³による実装に受け継がれる予定である。また、プロジェクトはアイテムレベルの利用統計を記録・収集するための国際標準を実現するために世界中の読者の注目を喚起したいと考えている。

利用統計の利点は特にその即時性にある。引用ベースの統計とは異なり、集計やランキングは実際の利用イベントがあり次第更新することができる。利用統計は電子リソースの可視性を計る指標である。その数値は、リポジトリにある研究成果がどの程度可視性があるかを著者により分かりやすく示すことができる。様々なリポジトリを比較することにより、著者はより高いインパクトを得るために公開場所を選択することができる。ユーザや科学コミュニティはその数値で論文や著者を評価することができる。関係行列に基づくリコメンドシステムなどの付加サービスはユーザに対するリポジトリの魅力を高めるだろう。

本レビュープロジェクトの中心事業は関係者によるワークショップである。ワークショップは2008年7月7-8日にベルリンで開催された。ワークショップにはリポジトリと図書館の代表に加えて、COUNTER、IRStats、JISC、LogEc、MESUR、OA-Statistics、OA-Network (Network of Certified Open Access Repositories) プロジェクト⁴の代表が参加した。ワークショップ期間中に、ログファイルの記録と交換のための基本スキームが作成された。その過程で、収集データの正規化が議論され、生データの交換方法が検討された。さらに、個人データの記録に関する法的制約や統計

¹ 英国情報システム合同委員会, <http://www.jisc.ac.uk/>

² <http://www.dfg.de/en/index.html>

³ <http://www.dini.de/oa-statistik>

⁴ 参加者リストを付録に示した。ワークショップで使用されたスライドは次の URL で公開されている。
<http://www.dini.de/veranstaltungen/workshops/jisc-workshop-2008/agenda/>

データに対する著作権規定が評価された。最後に、統計データの公開ポリシーやアグリゲータの必要性が議論された。ワークショップの成果は文章化して公開され、参加者や他の関係者から意見を求めた。本報告書はプロジェクトの最終報告書であり、ワークショップの成果と現在行われている活動のレビューおよびアイテムレベルの利用統計の将来に向けた勧告を合体したものである。

報告書ではまずオンライン利用統計分野におけるいくつかの有名なプロジェクトやサービス、イニシアティブの概要を手短にまとめる。次に、ワークショップの成果を記述する。ここでは、スキーム案の技術的側面と OpenURL Context Object への対応付けを扱う。第3に、個人データの記録に関する法的問題を記述し、最後に、ポリシーの問題と必要となる機関からの支援の分析を評価する。

2 現在行われている活動のレビュー

欧州 17 カ国の 114 のリポジトリを調査した DRIVER の研究 (Eindhoven and Graaf 2007) は、2006 年において、これらリポジトリの約 70% はデータのダウンロードとアクセスを記録しているが、エンドユーザにアイテムの利用統計を提供しているのは 30% に過ぎないことを示した。そのようなサービスを行っているリポジトリのおよそ半分は英国のリポジトリであり、全体の 1/4 はこのサービスの導入を計画していた。

2008 年には、様々なプロジェクトやイニシアティブ、関係者によりオンライン利用統計の集計や分析が行われた。本プロジェクトの目的から見ると、利用イベントに対する定義の違い、マルチクリックの重複除外、人以外によるアクセスの識別、利用データの収集が興味深い。本プロジェクトでは、各自が測定に関して極めて異なるコンセプトを持っている有名なプロジェクトをいくつか選択して、その活動を調査した⁵。

英国の IRStats プロジェクトや経済学データベース RePEc のために設計された LogEc のように、科学コンテンツを扱うイニシアティブが存在する。ドイツでは、最近、DINI が OA-Statistics プロジェクトと OA-Network プロジェクトを開始した。さらに、オーストラリアでも現在 Benchmark Statistics Service (BEST) という活動が行われている。利用データの構造的測定の先駆者はロスアラモス国立研究所 (LANL) で行われた MESUR プロジェクトである。リポジトリコミュニティでは、オープンソースの分析ソフトウェアパッケージ AWStats がよく利用されている。広く知られている COUNTER 実施指針 (Code of Practice) は雑誌とデータベースの利用しか規定していない。IFABC は例えば新聞オンラインサービスなどのページインプレッションの利用を測定する標準を策定している。

⁵ これが利用統計分野で現在行われているすべての活動の包括的リストだと主張するものではない。

JISC 利用統計レビュー

本プロジェクトではノルウェーとオランダの国家的イニシアティブも調査した。ノルウェーでは、現在のところ利用統計を扱ったプロジェクトは存在しない。一方、オランダでは、利用データを収集・提供するプロジェクトがSURFfoundationによりおそらく開始されるはずであるが、まだ計画は明確な形にはなっていない。国際的なプロジェクトでは、2007年のKnowledge Exchangeで利用統計が話題になった。その報告書には当時の利用統計に関する戦略上の課題や問題点が挙げられている(Knowledge Exchange 2007)。Knowledge Exchangeは、DEFF(デンマーク電子研究図書館)、DFG(ドイツ研究協会)、JISC(英国)、SURFfoundation(オランダ)が高等教育と研究のための情報通信技術基盤を共同で開発するためのプラットフォームである⁶。

IRStats(Interoperable Repository Statistics)⁷はサウサンプトン大学で行われている英国のプロジェクトで、EPrintsまたはDSpaceを使用しているリポジトリ用の利用統計モジュールを設計することを目的としている。プロジェクトは2005年6月に開始され、JISCから24ヶ月間の助成を受けた。プロジェクトの成果は統計ツールの試行版であり、サウサンプトン大学電子工学・計算機科学科のEPrintsリポジトリ⁸で稼動している。このプログラムは24時間以内のマルチクリックを除外し、AWStatsのロボットリストを使用して人以外のアクセスを除外する。今のところ、パッケージは基礎となるAWStatsと一緒になければ動作せず、幅広く実装することができない。

LogEcは、メタデータアグリゲータRePEcを補完する無料のオンラインサービスであり、経済学文献に特化したものである。LogEcは、RePEcに参加しているIDEAS、EconPapers、Socionet Personal Zone、Inomics⁹の各サービスで公開されているアイテムの統計を提供する。また、抄録ページ閲覧数とダウンロード数によるアイテムのランキングも作成している。すべての公開サービスについて、トップページから統計(各アイテムの抄録ページ閲覧数とダウンロード数)にアクセスすることができる。LogEcはスウェーデンのオレブロ大学ビジネススクールにより運用管理されている。マルチクリックは1ヶ月以内の再訪がIPアドレスを使って除外される。ログ分析は参加サービスが各自処理し、利用データがLogEcにアップロードされる。人以外のアクセスは次の2段階処理で除外される。まず、ロボットに対するインデックス化を許可するディレクトリに関する規則をrobots.txt¹⁰で設定し、既知のロボットホストを除外する。次に、1ヶ月以内にRePEcの全アイテムの10%以上にアクセスしたユーザを除外する¹¹。

⁶ <http://www.knowledge-exchange.info/>

⁷ <http://irs.eprints.org/proposal.html>

⁸ http://eprints.ecs.soton.ac.uk/index_stats.html

⁹ <http://repec.org/docs/RePEcIntro.html>

¹⁰ <http://www.robotstxt.org/>

¹¹ <http://logec.repec.org/about.htm> とSune Karlssonからの追加情報による。

DINI e.V.¹² (Deutsche Initiative für Netzwerkinformation e.V.) は、ドイツにおける情報・コミュニケーションサービスの連携と標準化を支援している。リポジトリに関してはDINI-Certificateに関する標準を定めている。今のところ、その認証基準にリポジトリ統計に関する包括的なチェックリストは含まれていない。しかし、現在、DINIはOA-StatisticsプロジェクトとOA-Networkプロジェクトによりオープンアクセスリポジトリネットワークのための試行版の実装を行っている。

BEST (Benchmark Statistics Service)¹³ は、2007年9月から12月にかけてオーストラリアのAPSR (Australian Partnership for Sustainable Repositories) により実施されたプロジェクトである。その目的は、利用統計の「扱い方を決め、さらなる開発のための枠組みを提供して試行サービスの設計を開始する」ことであった (Benchmark Statistics Project 2007: 3)。その活動範囲は本プロジェクトと極めて似ている。BESTはオーストラリアにアイテムレベルの統計を導入するための勧告を行っている。利用統計データの交換を行うために、BESTはOAI-PMHを採用した。人以外のアクセスを除外する条件はLogEcの基準と似ており、robot.txtを適用し、コンテンツの約10%以上に対する動的アクセス基準を採用しているが、さらに、AWStatsクローラリストの使用が追加されている。また、BESTは、2つのデータソースについて、集計値だけでなくアイテム統計も表示できる試行サービスを開始している¹⁴。

MESUR (Metrics from Scholarly Usage of Resources)¹⁵ は、学術研究成果に関する利用統計の分野では最も進んだプロジェクトである。この研究プロジェクトは、頻度論学派的測定から構造的測定へと一歩踏み込んでいる。MESURは、約10億件の利用イベントデータを持つデータベースを使用し、様々な利用統計値の検証や、雑誌ネットワークやアイテムネットワークの生成を行っている。ドキュメントからドキュメントへと続くユーザのクリックストリームを追跡すると関係データが作成される。このデータを使って関係行列を作成し、論文レベルのリコメンドシステムのプロトタイプを設計するのに使用している (Bollen and Van de Sompel 2006: 304)。

2002年から2007年にかけて、出版社やアグリゲータ、研究所など、様々なソースからデータが収集された。イベントの識別、ユーザまたはセッションの識別、アイテムの永続識別子、イベント発生時間が生データの必須項目である (Bollen et al. 2007: 5)。固有識別子が存在しない場合、MESURは雑誌の重複を避けるために「識別子バッグ」法を使用している。まず、レコードはISSNが比較され、次に雑誌タイトルの様々な形や略称が含まれている識別子バッグと比較される。雑誌レベル以下のアイテムは、出版年とタイトル冒頭の25文字を使って突合せが行われる。人以外

¹² <http://www.dini.de/ueber-dini/>

¹³ <http://www.apsr.edu.au/best/index.htm>

¹⁴ <https://devel.apsr.edu.au/cosi/best/reports/index.php>

¹⁵ <http://www.mesur.org>

のアクセスは、その行動の典型的な特長により識別される。カリフォルニア州立大学のリンキングサーバシステムのログファイルデータが OpenURL Context Object 形式にパックされ、OAI-PMH でハーベストされた (Bollen et al. 2007: 5)。

オープンソースのログ分析ソフトウェア **AWStats**¹⁶ は、リポジトリの世界で広く使用されている。同一 IP アドレスによる 1 時間以内のマルチクリックは除外される。クリックストリームは記録されず、各自のリポジトリの統計だけが利用できる。人以外のアクセスは AWStats が持つロボットリストを使って除外される。このリストは IRStats や BEST でも使用されているのである種の標準となっているが、様々なバージョンのリストが出回っているのでベンチマークとしては極めて曖昧なものである。

雑誌およびデータベースのための **COUNTER** (Counting Online Usage of Networked Electronic Resources)¹⁷ 実務指針は、既に幅広く使用されている雑誌利用統計のための標準である。2008 年 8 月にはリリース 3 が公開され、COUNTER 準拠となるために各ベンダーは 2008 年 9 月までにこれを実装しなければならなかった。このリリースの主な変更点は、本報告書にも関係するが、COUNTER レポートのハーベストに SUSHI を使用することが必須になったこと、XML 形式によるレポートの提供が要件になったこと、レポートから人以外によるアクセスが除外されたことである (COUNTER 2008: 2)。最後の件について、COUNTER のリストには現在 36 のロボットとクローラが含まれている (COUNTER 2008: Appendix K) が、このリストは常に更新されることになっている。このリストは最小限の基準に過ぎないが、除外リストに追加したホストをベンダーが文書化する必要があるか否かは明らかでない。また、横断検索や LOCKSS¹⁸ によるアクセスは除外するか、少なくとも別個に集計しなければならない。同一ユーザによる HTML 文書のマルチクリックは 10 秒以内に行われた場合はカウントしない。PDF 文書は 30 秒以内のマルチクリックはカウントしない。このマルチクリックの判定において、ユーザは IP アドレスまたはセッションクッキーなどの何らかの識別子により識別される。カウントは成功したアイテムリクエスト (すなわち、リターンコードが 200 と 304 のもの) だけ行う (COUNTER 2008: 32; Appendix D)。

COUNTER における統計の最低レベルの粒度は雑誌である。特にリポジトリにあるドキュメントにとって意味のある指標はアイテムレベルの指標である。学際的なリポジトリにおいて、集約された統計は可視性と人気についてはある程度伝えることができるが、ある主題分野においてそれが特に重要であるか否かについては何の情報も与えない。現在、COUNTER は雑誌論文を対象とするアイテムレベルに対する標準の拡張を行っている。PIRUS (Publisher and Institutional Repository Usage Statistics) プロジェクトは、出版

¹⁶ <http://awstats.sourceforge.net/>

¹⁷ <http://www.projectcounter.org/>

¹⁸ LOCKSS (Lots of Copies Keep Stuff Safe): <http://www.lockss.org>

社とリポジトリコミュニティを統合し、両環境で適用可能なアイテムレベルの利用統計を記録する共通規格を策定しようとしている。その成果は2008年末までには現れることが期待できる¹⁹。

UKSG (United Kingdom Serials Group)²⁰は、出版社と図書館の利益集団としてPIRUSにも参加している。それ以外に、COUNTER代表のPeter Shepherd (2007)はUKSGの委託により2007年に、雑誌に利用因子(UF: Usage Factor)を導入することに関するステークホルダー調査を行った。UFは、COUNTERレポートの雑誌利用統計をオンラインで公開されている論文の数で割ったものであり、両変数とも指定の期間における数値が使用される(Shepherd 2007: 4)。近い将来、UKSGはUFのフレームワークを示す提案依頼書を発行する計画である²¹。

IFABC (International Federation of Audit Bureaux of Circulations)²²は、研究には関係しない商用プロバイダのオンラインコンテンツの利用を測定する標準を策定している。この標準は学術コンテンツを扱っていないが、インターネットトラフィックを測定する際立った方法をガイドラインで提供しているのでこのレビューに含めた。この標準ではドメイン別のページインプレッション数と訪問数がカウントされる。ここで、ページインプレッションとは「ユーザのリクエストの結果としてユーザに送られるファイルまたはファイル群」である²³。これ以外に、チャットインプレッションやストリームインプレッションなど、その他の種類のオンラインコンテンツの利用も記録される。

IFABCは最低限の基準を設定しており、様々な測定方法のガイドラインを提供している。例えば、ユーザはIPアドレスやユーザエージェント、クッキー、登録IDにより識別することができる。ここでは、訪問を再現できることだけが重要である。IFABCのドイツの会員組織の規格IVW²⁴はもっと具体的である。IVWはバグトラッキングの使用を規定している。マルチクリックは30分以内のものが除外される(IVW 2008:6)。理論的に構成したユーザ識別の技術的実装についてはそれほど強く規定していない。これは、COUNTER実務指針と似ている。イギリスのABC electronic²⁵は、IFABC規則を実装し、ロボットリストを使って人以外のアクセスを除外している。IFABCとその各国会員組織は会員だけに文書を提供する商用サービスであるので、除外するロボットや利用データのハーベストに関する文書は公開されていない。

¹⁹ PIRUS に関する情報は Paul Needham と Richard Geyde により提供された。

²⁰ <http://www.uksg.org>

²¹ <http://www.uksg.org/usagefactors>

²² <http://www.ifabc.org/who.asp>

²³ <http://www.ifabc.org/standards.htm>

²⁴ 貸出識別研究部会 (Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e.V.): <http://www.ivw.de>

²⁵ Audit Bureaux of Circulation electronic: <http://www.abce.org.uk>

3 基本スキームの制定

理想的には、利用統計は将来の出来事に基くことになる。ユーザのクリックストリームはオンラインコンテンツを経由して追跡され、これらの経路はアイテム間に関連ネットワークを構築する。同一の読者を惹きつけるアイテムは互いに密接に関連しているからである。しかし、リポジトリはまだそのようなデータを提供することはできない。ランキングのほか、論文のプレステージや媒介中心性 (Betweenness Centrality) などの各種指標を算出できるようになるには、まだ何段階かの過程を経る必要がある。この目的に向けた第一歩は、既に多くの組織で採用されている電子ジャーナルのためのCOUNTER統計に比較しうる統計を提供することである。ただし、これはCOUNTER²⁶とは異なり、雑誌レベルではなくアイテムレベルの統計であるべきである。本章では主に、ベルリンで開催された関係者によるワークショップにおける議論の結果を述べる。

3.1 データソース

ドキュメントに対するリクエストそれ自体は、その実際の利用法、すなわち、それを読むのか引用するのか、については何も語らない。従って、より正確に言うなら、それはアクセスデータと呼ばれるべきである。しかし、すでに「利用データ」という用語が広く使用されている。ワークショップでは、アクセスの定義をアイテムまたはそのトップページに対する成功したリクエストとすることが合意された。検索結果リストに現れるアイテムは利用イベントの構成要素ではない。これは、そのアイテムへのユーザの関心をそれほど明らかに示しているわけではないからである。

ログファイルはそのソースにより様々な形式をしている。一般的な Web サーバの間だけでなく、ライセンスサーバやリンクリゾルバ、リポジトリソフトウェアパッケージの間でも違いが存在する。これらを比較可能にするためには、ログファイルをパースし、標準フォーマットに変換する必要がある。ワークショップの参加者は、OpenURL Context Object 形式の XML フォーマットを標準フォーマットの構文とすることに合意した。XML は COUNTER、IRStats、MESUR でも使用されている。

Web ページへのアクセスはまず Web サーバのログに記録されるが、リンクリゾルバやライセンスサーバ、リポジトリソフトウェアパッケージのログからも抽出することができる。これら 4 種類のログファイルはすべて有効な選択肢であり、組み合わせることもできる。ただし、全てのリポジトリソフトウェアパッケージがリクエストをログ出力するわけではない。

リンクサーバを利用データの記録に使用することはできるが、これを実現するには、リンクサーバが利用の大部分を記録できるようにデジタルライブラリシステムを構築しなければならない。これは欧州における多くの図書館や研究機関には当てはまらない。別の障害も存在する。例

²⁶ <http://www.projectcounter.org>

例えば、リンクサーバのログファイルは無料で利用できない可能性がある。ライセンスされた出版物は簡単に追跡することができるが、リポジリのコンテンツは常にリゾルバシステムに含まれているわけではない。さらに、ドキュメントはリンクリゾルバ経由でアクセスできるだけでなく、例えば、永続識別子を經由して直接出版社サイトのドキュメントにアクセスすることもできる。

Web サーバのログはより広範囲で、全ての利用データをカバーしている。アイテムがリクエストされる度に常に作成されるからである。しかし、一方で、アクセスライセンスを提供するだけでコンテンツは配信していない機関は、必ずしも常にログが利用できるわけではない。これらの利用イベントはリンクリゾルバが記録することができる。これら 2 つの選択肢が抱える問題により、これらの組み合わせが最も有望な方法になっている。これは、Web サーバとリンクリゾルバ両者のログから利用データがマージされることを意味する。

3.2 基本スキーム

レコードを収集する前に、様々な Web サーバやリポジリの間でレコードを比較可能かつ交換可能にする必要がある。そのため、利用データの記録は調整されなければならない。本報告書では、これを「基本スキーム」と呼ぶ。ワークショップにおいて、参加者はログファイルの必須項目について合意した。それは、ユーザ、セッション、アイテムの識別、リクエスト種別、リクエスト日時である。これらの要素がログファイルの不可欠な部分となるべきである。これらの要素は、基本的な質問「誰が」「何を」「いつ」に関する情報を提供する。これらの項目以外は、このスキームを自由に拡張できるようにするべきである。特に、サービス、リファラ、参照実体の識別は、フォーマットを拡張する際の候補になるだろう。しかし、基本サービスには、これらは不要だと思われる。

ユーザ ID/セッション ID/利用イベント ID

マルチクリックの除外を可能にし、オンラインコンテンツをたどるクリックストリームの追跡を容易にするためには、リソースをリクエストするユーザを識別しなければならない。このための第 1 の選択肢は IP アドレスである。しかし、これはユーザにまで遡ることができるので、プライバシーの問題が発生する。IP アドレスをまったく記録すべきでないかどうかは疑問の余地がある。非常に制限の多いドイツ法の下では、ユーザの識別を可能にする IP アドレスやその他のデータの処理は強く制限されている。もう 1 つの問題は、プロキシサーバや同一の IP アドレスを共有するコンピュータネットワークである。この場合、2 人の別のユーザを区別することがほとんど不可能である。

IP アドレスの記録を不要にし、より洗練されたエンドユーザサービスを可能にする別の戦略が存在する。想定したサービスでは、ユーザは IP アドレスの代わりに自動的に付与されるセッション ID により識別することができる。以下に述べる 2 つの選択肢は、セッション統計と読んだ方がふさわしいかもしれない。

第 1 の選択肢は、連続的なクリックストリームのために、訪問する度にセッション ID、すなわち、UUID (Universally Unique Identifier) を付与することである。これにより、ユーザが 1 セッション中に発行するドキュメントへの複数回のリクエストをコントロールすることができる。また、1 つのセッション ID で行われたリクエストをタイムスタンプ順に並べることにより、あるドキュメントから別のドキュメントへと移動したユーザの軌跡を示すことも可能である。

第 2 の選択肢は、セッションクッキーやクッキーの利用である。これらはユーザの追跡を少し容易にする。セッションクッキーはブラウザがクローズされる際に削除される。クッキー (例えば、1 ヶ月後に期限切れとなるクッキー) は、この期間中の再訪の識別を可能にし、記録されるクリックストリームの長さを延長する。欠点はクッキーを無効にしているユーザが無視されることである。2005 年には、およそ 5% のユーザがクッキーを無効にしていた²⁷。クッキーのもう 1 つのマイナス面は、使用されている様々なソフトウェアプログラム用に設定しなければならないので、実装コストが高いことである。これはリポジトリ予算を超える可能性がある。かなり大規模なアグリゲータを除けば、これを実装する価値はないかもしれない。小さなリポジトリを訪問するユーザは、単一のアイテムをダウンロードするだけだからである。ユーザのセッションが複数のコンテンツプロバイダ (すなわち、リポジトリコンソーシアム) を通じて追跡可能な場合にのみ、クッキーは有用になるだろう。PDFドキュメントへのアクセスに対するクッキーの実装はさらに複雑であるが、全般的に問題は少ない。ユーザのリクエストを PDF にリダイレクトする前に、リクエストを記録するトラッキングスクリプトにリクエストをリダイレクトすることができるからである。

全てのイベントまたはセッションに付与する固有識別子も必須の要素である。識別子は、重複除去を容易にし、イベントを二重にカウントすることを防止する。これは、リンクリゾルブが行われることにより 2 回記録されることになる場合である。今のところ、リンクングサーバにセッション ID を実装することは技術的に難しく、ほとんど利用されていない。MESUR プロジェクトは SFX 用の拡張機能を使用した。技術的に他のリンクリゾルバサービスがセッション ID を実装できるかどうかは明らかでない。

IP アドレスから、リクエストが発行された国を識別することができる。これは面白い情報である。この情報はユーザの地理的な広がりを示すので、一般にリポジトリだけでなくドキュメントのインパクトに関しても多くを語るからである。しかし、これはプライバシーの問題を引き起こす。これについては、6 章で検討する。

アイテム

²⁷ http://www.web-analytics.org/index.php/webanalyseartikel/weiter/ohne_cookies_kein_korrektes_tracking/

アイテムの収集と重複削除、そして、イベントIDが利用できない場合はイベントの重複削除のために、アイテムを識別可能にする必要がある。画像を含み、そのためログファイルに複数回記録される HTML 文書のマルチヒットも調整されなければならない。アイテムはそのメタデータに基づいて識別されなければならない。

理想的には、全てのドキュメントに、DOI (Digital Object Identifier) や URN (Universal Resource Name) または同様な永続識別ソリューションによる固有識別子が付与されるべきである。残念ながら、各リポジトリはアイテムをユニークに識別するために様々な標準を使用している。しかし、全てのアイテムに固有識別子を付与することが利用データ交換のための必須条件にならない。これが不可能な場合は、MESUR や OA-Network でも行われたように、メタデータに基づくファジーマッチングがアイテムのイベントを集約するための選択肢となる (Bollen and Van de Sompel 2005: 305)。この方法は、複数のリポジトリに含まれており、そのため、異なる固有識別子が付与されている可能性があるドキュメントを識別する際にも有効である。アイテムやイベントの重複削除は、その一貫性を保つためにアグリゲータが行わなければならない。

3.3 基本スキームの正規化

ログファイルは人によるアクセスの記録だけでなく、スパイダーや Web クローラによる活動の記録でもある。後者は利用データを歪曲し、不正なものにする。ログファイル中のスパイダーの記録は、既に第 2 章で記述したように、様々な方法で除外することができる。

- robots.txt
- AWStats や IFABC のロボットリストのようなリスト
- 指定した時間内に行われた全コンテンツのある割合へのアクセス
- 特殊な利用パターンを持つユーザの除外

数人の参加者は、人以外のアクセスレコードの除外に反対すると述べた。このデータはロボットの行動について多くを語るだけでなく、MESUR が実装したようなロボットを除外する代替方法を開発する機会を与えるからである。人ユーザによるクリックストリームは極めて短く、意味のあるパターンを示すように思われるが、自動化アクセスは全てのドキュメントをランダムに抽出する。これにより不完全と思われるロボットリストへの依存を避けることができる。人以外のアクセスによるログファイルエントリはローカルレベルまたはアグリゲータレベルでタグ付けすることができると思われる。ローカルレベルのタグ付けは、最終的にデータを分析するために新たなリポジトリ管理作業を招くことになる。それゆえ、人以外のアクセスの除去やタグ付けは、中央レベルで行うことが最も効率的である。これにより一貫性のあるロボットの除去も可能になると思われる。

この方法は、人以外のアクセスによる別の問題を発生させる。ログを記録するストレージ容量が必

要になることである。Tim Brodyは、ヒットの約 90%は人以外のアクセスに帰することができるかと報告している。2007年7月には、RePEcの抄録ページの閲覧件数の74%がロボットによるものだった²⁸。ロボットによるアクセスを含むログファイルのサイズは、多くのリポジトリのストレージ能力を超えている。これらのリポジトリは、ログファイルを長期間保存しておくことがほとんど不可能なのですぐに削除する。その解決策の1つは、比較的安価なストレージを提供しているAmazon S3²⁹などの代替ストレージを検討することだと思われる。もう1つの選択肢は、利用データのハーベスト回数を増やし、収集後、リポジトリが直ちにログファイルを削除できるようにすることだろう。

全員の意見が一致したのは、常に人以外のアクセスが存在すること、その全てが識別できるわけではないことである。しかし、データを比較可能にするにはこれらを除外するための共通のガイドラインを持つことが不可欠である。そうでなければ、全ての指標は無意味である。識別は(RePEcの例のように)実用的で幅広いシステムで簡単に実装できるものであるべきである。また、識別粒度の精緻化や改変が可能であるべきであろう。

4 利用データの交換

4.1 利用データハーベスティング用の共通プロトコル

利用データのハーベストに使用できる2つの競合する方法がOAI-PMHとSUSHIである。ハーベストはコンテンツプロバイダからアグリゲータの一方向に行われるだけでなく、逆方向にも行われる。コンテンツプロバイダ(リポジトリなど)は、集約された統計や各種指標をアグリゲータからハーベストして、必要とされる情報をユーザに提供することができる。ローカル機関はOAIデータプロバイダとして機能し、中央アグリゲータはOAIデータプロバイダ兼OAIサービスプロバイダとして機能する。

SUSHIプロトコルは非常に良く定義されているが、その実装による評価は曖昧であり、リポジトリコミュニティにおけるOAI-PMHほどには広く使用されていない。SUSHIは出版コミュニティにおける統計の交換用の標準である。ワークショップの参加者はOAI-PMHを支持した。SUSHIの代わりにOAI-PMHが実装される場合は、両標準間のクロスワークが必要となる。これはセマンティックレベルにおいては特に重要である。それ故、PIRUSプロジェクトによる雑誌論文用のアイテムレベルの標準拡張を行っているCOUNTER策定作業に、このワークショップの結果が取り込まれたことは大いに評価されることである。

²⁸ <http://logec.repec.org/about.htm>

²⁹ <http://www.amazon.com/gp/browse.html?node=16427261>

もっと簡単な方法が FTP (File Transfer Protocol) によるデータの交換であり、これも議論された。FTP は OAI-PMH に比べて技術的に実装することは容易であるが、ハーベットの正確性は劣っている。FTP はファイルの標準命名法を提供していないので、既にハーベットされたファイルをタグ付けすることができず、ファイルの削除も統一的には記録されない。

XML ペイロードのコンテナには OpenURL Context Object を使用することが合意された。OpenURL Context Object の利点は、拡張が容易なこととデータの圧縮率が高いことである。

4.2 利用データのマッピング

利用データの OpenURL Context Object へのマッピングは既に Bollen と Van de Sompel (2006) により提案されている。本プロジェクトはこの提案を採用し、基本スキームのニーズに合わせるために、OpenURL 標準 (ANSI/NISO Z39.88-2004) に当たって少しだけ調整した。表 1 の左の欄に示したログファイルの項目から順に Context Object の該当する要素を特定し、右側の欄に示した。

表 1: 利用データの OpenURL Context Object データ構造へのマッピング

ログファイル項目	OpenURL Context Object
文書識別子	被参照体
アクセス時間	タイムスタンプ
イベント識別子	Context Object 識別子
IP アドレス	リクエスト
セッション識別子	リクエストの値指定のメタデータ
ユーザエージェント	リファラ
HTTP 状態コード	サービス種別

Context Object は次の 6 つの項目を含むことができる: 被参照体 (*Referent*)、参照体 (*ReferringEntity*)、リクエスト、サービス種別、リゾルバ、リファラ。これらの項目は、識別子、値指定 (*By-Value*) のメタデータ、参照指定 (*By-Reference*) のメタデータ、独自データ (*Private Data*) で記述することができる。

被参照体はリクエストされたアイテムであり、何らかの永続識別子 (たとえば、URN) で識別されるべきである。アイテムの重複削除をしなければならない場合は、被参照体に値指定のメタデータを追加することができる。これは、ペイロードのサイズを大幅に増加させる可能性がある。代わりに

参照指定のメタデータを使用すると、イベントデータを必要に応じてメタデータに結合することができる。Context Object に追加されるのはメタデータへのリンクだけであるからである。アクセス時間とイベント識別子は XML Context Object のヘッダに設定される。

ユーザの IP アドレスまたはそれを符号化したものはリクエストフィールドにマッピングすることができる。セッション ID のための独立した項目は存在しないが、リクエストのメタデータに含めることができる。この項目のフォーマットはいずれ定義されなければならないと思われる (Bollen and Van de Sompel 2006: 301)。ユーザエージェントはリファフィールドに設定される。

HTTP状態コードをサービス種別へマッピングすることは、今のところ不可能である。サービス種別に設定できる値は、OpenURLレジストリによれば、“fulltext”、“abstract”、“citation”、“holdings”、“ill”、“any”だけである³⁰。HTTPレスポンスコードの情報を含めるには、新たな値を定義する必要があると思われる。被参照体がフルテキストか、抄録かを示す情報もContext Objectに含めるべきである。ただし、これはログファイルからは得られない可能性がある。コンテンツプロバイダは、この情報を提供しなければならない。また、これを記録後にログファイルに追加するか、Context Objectのサービス種別の値指定のメタデータに追加するか、のいずれかをしなければならない。これら2つのサービス種別を持つために、Context Objectには2つのサービス種別項目(各々の種類に1つ)を指定することができる。

Context Object は OAI-PMH を使ってハーベストされるが、Context Object は XML フォーマットなので、SUSHI でもハーベストできるはずである。その場合、OAI ヘッダを SUSHI ヘッダに置き換えなければならない。SUSHI 自体は、XML ペイロードが特別なフォーマットを持つことを要求していない (ANSI/NISO Z39.93-2007)。現時点で、技術上の相互運用性を評価することはできないが、今後、出版コミュニティによる仕様策定に沿って作業を行うべきである。

5 基本スキームと他の試みとの比較

表 2 にワークショップで起草されたスキームの特徴と COUNTER、LogEc、IFABC の特徴をまとめた。COUNTER、LogEc、IFABC の各統計と比較可能にするために本報告書が提案する統計をどのように調整しなければならないのか、また、それはまったく意味がないのか否かについて、以下で手短かに評価する。

ワークショップの提案と実施指針リリース3による COUNTER 指標を比較することはできない。現時点では粒度があまりにも違うからである。COUNTER は雑誌とデータベースの利用しか捕捉していないのに対して、ワークショップ提案はアイテムレベルを目指しており、リポジトリコンテンツに焦点

³⁰ info:ofi/fmt:xml:xsd:sch_svc

を絞っている。そのため、たとえ集約した形であっても両者の比較は意味がないと思われる。現在 COUNTER(2008)は論文レベルに対する指針の拡張を行っているので、この新しい成果に適応するために、PIRUS プロジェクトによる策定作業をしっかりと監視するべきである。

LogEc のスキームはワークショップ提案とはほんの少し異なっている。基本スキームは 2 つの点 (データの正規化とマルチクリックを除外する時間基準)がまだ詳細に定義されていないからである。したがって、両者を比較することがどの程度難しいかを完全に評価することはできない。ドイツの法的環境においては、たとえ一方向ハッシュを使用したとしても 1ヶ月以内に再訪したユーザを識別することはできないと思われる。LogEc はプル機構を使ったハーベストを行っておらず、ローカルで分析された利用イベントが FTP を使ってアップロードされている。しかし、これは指標にとって大きな違いではない。何をカウントするかという定義には影響を与えないからである。

各種統計をIFABC標準と比較可能にすることはさらに少し難しい。ロボットリストが一般に公開されていないからである。また、抄録ページの閲覧に相当すると解釈することができるスプラッシュページがIFABCの標準ではカウントされないので、何をカウントするかという定義をフルテキストへのアクセスに縮小しなければならないと思われる³¹。

³¹ <http://www.ifabc.org/standards.htm>

表 2: JISC 利用統計レビューの提案と他のスキームの比較

基準	基本スキーム	COUNTER(雑誌用) ^a	LogEc ^b	IFABC ^c
粒度	アイテムレベル	雑誌レベル	アイテムレベル	ページインプレッション、 訪問
利用イベントの定義	成功した抄録ページの閲覧 とダウンロード	成功したフルテキスト リクエスト	抄録ページの閲覧と ダウンロード	ドメイン別ページインプレ ッション数、訪問数
マルチクリックの重複除外	規定なし	HTML は 10 秒、 PDF は 30 秒	1 ヶ月	1 時間
人以外のアクセスの識別	規定なし	ロボットリスト	robots.txt、動的基準	ロボットリスト
ハーベスティング	OAI-PMH	SUSHI	ローカルで分析した利用デー タを FTP でアップロード	-

^a COUNTER 2008

^b <http://logec.repec.org/about.htm> とSune Karlssonからの追加情報による

^c <http://www.ifabc.org/standards.htm>

6 ログファイルの記録と集約に関する法的制約

ワークショップの法律セッションはベルリン・フンボルト大学図書館・情報科学研究所のMichael Seadleとミュンスター大学情報・通信・メディア法研究所のHannes Obexが司会を務めた³²。

利用統計については様々な法律上の懸案事項が存在する: (1) ユーザのプライバシー、(2) 論文の利用統計が公開される著者や出版社のプライバシー、(3) 統計指標の著作権。

プライバシー法は IP アドレスの記録・処理やクッキーの使用により侵害される可能性があるが、その規則は国により大きく異なっている。プライバシー法は各国で制定されているが、欧州法の支配下にある。欧州法は米国の規制より厳格であるが、ドイツ法よりは厳格でない。英国の状況は大部分欧州法により支配されている。

米国では、連邦プライバシー保護法(1974)は公共企業体にしか適用されない。通信法(1996)は主にアクセスを提供する文脈に沿って設計されており、コンテンツプロバイダには適用されない。ユーザのプライバシー保護はもっぱら自主規制に依存しており、私的電子通信部門のためのプライバシー保護法は米国には存在しない。この規制レベルの低さが EU における当事者間のデータ交換を難しくしている。EU から米国への個人データの転送は欧州法では許されていない。唯一の例外はセーフハーバーシステムに所属する企業である。これに参加する米国企業は EU の基準に従わなければならないが、これにより欧州共同体内のデータ転送が可能になる。

ドイツは、プライバシー保護に関して非常に制限の多い規則を持っている。憲法は情報の自決権を与えている。個人データの記録と取り扱いは厳しく制限されている。利用統計の文脈における個人データは、ユーザを識別可能にする IP アドレスである。ドイツ電気通信法(TMG)はその境界を定義し、サービスの提供または料金徴収の目的を除いて、個人情報を収集または処理してはならないと規定している。それゆえ、無料サービスを提供するプロバイダは IP アドレスの処理が許されていない(TMG § 15 (1))。この規則が、データを仮名化または無名化する前に IP アドレスから国コードを得ることに適用されるか否かは明らかでない。サービスプロバイダは、市場調査、広告、オンデマンドサービスの設計のために仮名化されたユーザプロフィールを作成することが許されている。クライアントの国名はユーザプロフィールの要素であると解釈できるので、データが後に仮名化されるのであれば法的な問題は発生しないかもしれない。仮名化したデータを個人情報にマーキングすることは禁止されている。

アクセスデータの記録にユーザが異議を唱える権利は一般に上に示した規則の影響を受けない。

³² 法的制約に関するこの節の内容はほとんどが Obex と Seadle による講演と両者の専門知識に拠っている。彼らのスライドは次の URL で公開されている。

http://www.dini.de/fileadmin/workshops/JISCworkshop/Legal_Session.pdf

コンテンツプロバイダは、事前にデータを記録・処理する範囲と意図をユーザに知らせなければならない。提供されるサービスが唯一のもので他では利用できないものである場合は、たとえデータを記録することに同意が得られなかったとしても、そのサービスはユーザを拒絶することはできない。利用データは交換または集約する前に無名化されなければならない。また、ユーザはそれを知らされなければならない。コンソーシアム参加機関の間のデータの交換が法律上のデータ交換の定義に含まれるか否かは、今のところ明らかではない。クッキーの使用は、それが個人データを含まない場合を除いて、同じ制約の対象となる。

ドイツの法環境においては、IP アドレスの記録は厳しく制限されているが、この法律の現行の解釈はあいまいである。法的問題を避けるために、利用イベントのすぐ後に IP アドレスを仮名化するか、IP アドレスをまったく使用せず、IP アドレスを記録しない何らかのセッション識別の実装を促進することがベストだと思われる。

一方、著者と出版社のプライバシーに関しては、統計の公開により法的な問題を引き起こすことはないはずである。これらは個人データの構成要素ではないからである。同じことが出版社にも当てはまる。しかし、様々な研究分野の中には利用統計の公開に対して大きな反対を引き起こす可能性のある社会規範を持つ分野もあるかもしれない。また、利用統計の一般公開は出版社の経済的利益を損ねる可能性もある。

利用統計の著作権も議論の余地がある問題である。その基本的な疑問は、元々利用統計が保護するに足るオリジナリティを持っているか否かである。もしオリジナリティを持っていないなら、それは真実として分類されるので、著作権法により保護されないと思われる。著作権法は自動的に行使されず、権利保持者は侵害行為を告発しなければならない。しかし、そのような告発の可能性は、統計から得られるかもしれない想定上または実際の経済価値にしたがって増加する。

7 利用統計ポリシー

ワークショップのポリシー・セッションでは、利用統計の公開、品質標準、利用統計に対する機関の支援が取り上げられた。その他、提案された利用統計の実装を開始するには何をすべきかについても関心が集まった。

利用統計をオープンアクセスにするべきか、もしそうするなら、どの程度オープンアクセスにするべきかが議論された。プライバシーが容易に侵される可能性があるため、生データは一般公開するべきでないという共通理解があった。研究目的の厳しく制限したアクセスだけが可能になるべきである。

利用統計の状態については十分な合意が得られなかった。多くのリポジトリはオープンアクセス・ムーブメントの一部であるのでその考えを否定したくないと考える。しかし、一方では、サービスのインフラには資金が提供されなければならない。利用統計は価値のあるサービスであると思われる。利用統計は研究評価に利用できるし、リコメンドシステムを導入する際の前提条件でもある。無料公開と有料サービスのほかに、第三の選択肢として部分的に一般公開するサービスが考えられる。基本的な統計はオープンアクセスにし、詳細な統計やリコメンドシステムへのアクセスを制限する方法である。これは、特に利用統計を導入した初期段階において、利用可能なリソースに合わせてリポジトリのサービスを調整する方法でもある。この階層的アプローチにより、コンテンツプロバイダは提供する統計レベルを決定することができる。たとえば、アイテムレベルの COUNTER 統計のみを提供したり、単純な集計以上の統計を提供したり、構造的統計を提供したりすることができる。

すでに述べた統計の状態や関係者の記述の他に、リポジトリポリシーにはデータの収集や処理に関するガイドラインを含めるべきである。高等教育機関のランキングに関するベルリン原則 (Berlin Principles on Ranking of Higher Education Institutions)³³は、利用統計に関するポリシーを策定する際のたたき台として利用できる。ベルリン原則は 2006 年に UNESCO 欧州高等教育センター、高等教育開発センター大学ランキング (ドイツ)、高等教育政策研究所 (米国) により策定された。特に、指標の設計と重み付け、データの収集と処理、ランキング結果の表現に関するガイドラインは利用統計に適用できる。これらの原則から導かれるポリシーには、データがどのように標準化されているかだけでなく、指標がどのように構成され重み付けられているかについても包み隠さず記述するべきである。

データの集約と処理だけでなくその品質保証も信頼のある中立的な第三者機関で行われなければならない。データの正規化と処理を集中化すれば、統計の信頼性と正当性に対する脅威は最小になる。ローカル処理に分割することに起因するエラー発生要因の多くが排除されるからである。また、中央機関はデータ記録が高品質であることを保障するために、COUNTER や IFABC がすでに行っているような監査プロセスを組み込むべきである。この中央機関は必要に応じて随時ポリシーを変更できる権限を持つべきである。このアグリゲータの役割を与えるべき既存の機関には、DRIVER や DINI、JISC Collection などがある。最終的にどこを選択するかはイニシアティブの全体的な適用範囲による。それにより、全国利用統計か国際利用統計のいずれかになる。

8 利用統計サービスの今後の展開

³³ http://www.che.de/downloads/Berlin_Principles_IREG_534.pdf

英国では、最近開始されたPIRUSプロジェクトにより、比較可能なアイテムレベルの利用統計に向けた次なる一歩が踏み出された。その目的は、雑誌論文あるいはより広い意味でのアイテムレベルの利用を測定する出版社およびリポジトリ向けの COUNTER 準拠の標準を策定することである。この標準は、リポジトリの環境にも適用できるように設計される予定である。このプロジェクトは、JISC による 2008 年 9 月から 12 月までの助成を受け、出版社とリポジトリの代表者による共同作業が特徴となっている。利用統計の実装はその適用範囲に入っていない。

英国におけるさらなる展開の集成者または創始者としては、それに必要なリソースを持っている JISC がおそらく最もふさわしい機関である。一方、UKSG は主に図書館と出版社を代表している。UKSG はリポジトリの世界における利用統計の第 1 の理想的な指導者ではないが、PIRUS と COUNTER の重要なパートナーである。

ドイツにおいては、DFGが助成しているOA-Statisticsプロジェクトが本報告書の結果を最初に取り上げる予定である。このプロジェクトは参加機関のリポジトリのために統計サービスの試行版を開発する。ただし、その焦点は主にインフラの設計に当てられており、指標そのものではない。最新の作業パッケージの 1 つとして、OA-Statisticsはネットワーク分析法を用いてデータの解析を行い、エンドユーザーサービス案を特定する予定である。論文の重複削除などその他の重要な作業は主としてOA-Networkにより行われる。これら 2 つのプロジェクトは密接に協力して、リポジトリコンテンツを検索し、利用統計を提供するための共通ユーザインターフェースを開発する計画である(図 1)。プロジェクトは 2008 年半ばに開始され、最終結果は 2009 年第 4 四半期までには提出されることが期待できる。OA-Networkにより提供されるサービスはプロジェクト期間を超えて公開され、調整されることになっている。プロジェクトの一部として、運用のためのビジネスモデルの策定がある。これにはおそらくサービスの提供におけるDINIの継続的な関与が含まれると思われる³⁴。

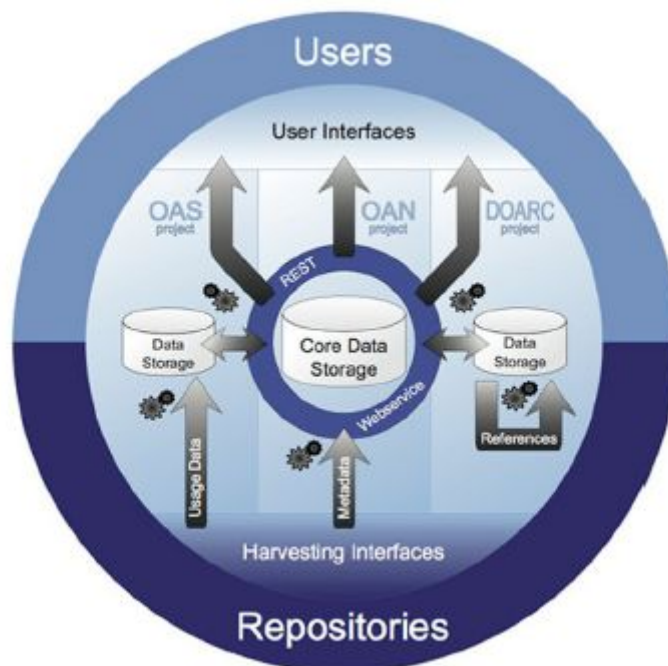
図 1 の右側はDOARC (Distributed Open-Access Reference Citationサービス)プロジェクト³⁵を示している。DOARCはOA-NetworkとOA-Statisticsを補足するサービスとして計画され、参照された研究成果を解析してリンクを張るサービスであるが、プロジェクトはまだ開始されていない³⁶。

図 1: ドイツの OA-Statistics、OA-Network、DOARC プロジェクトの構造と相互関係

³⁴ <http://www.dini.de/projekte/oa-netzwerk/projektbeschreibung/>

³⁵ <http://doarc.projects.isn-oldenburg.de/index.php?site=doarc>

³⁶ 本報告書の最終版公開の時点で DOARC の実装は依然として不透明である。



出典: <http://www.dini.de/projekte/oa-statistik/>

ドイツ、オランダ、英国で現在行われている活動は重要なイニシアティブであるが、リソースをより効率的に使用し、共通の欧州標準を育成するために、これらのプロジェクトはできるだけ密接に協力するべきである。JISC 利用統計レビュー・ワークショップは、これらプロジェクト間の対話と協力を促進する意味で重要なイベントであった。

既に、2007 年の Knowledge Exchange において、各国の全国組織である DEFF、DFG、JISC、SURF は、利用統計に関する各自の見解を共有していた。これは継続されるべきである³⁷。しかし、さらに重要なことは、各国のプロジェクトの成果を DRIVER の世界に取り込むことである。OA-Network は DRIVER プロジェクトに対するドイツの貢献である。そのモジュール(結果として OA-Statistics も)は、DRIVER プラットフォームと互換性があるように設計されている。それゆえ、OA-Network と OA-Statistics の成果を DRIVER に統合することはきわめて魅力的であり、実現も可能である。OA-Statistics ガイドラインを DRIVER に適用することも可能だと思われる。互いの連絡窓口は既に確立されている。各国のプロジェクトが協力することにより、DRIVER コミュニティ内での幅広い支援も約束されると思われる。

技術的・組織的実装以外に、利用統計を扱う際にはリポジトリや著者に情報を提供し支援することが重要である。これにより、利用統計の受容度が高まるだけでなく、標準に対するコンプライアンス

³⁷ <http://www.knowledge-exchange.info>

も高まると思われる。現在、リポジトリがメタデータやデータ、コンテンツ、投稿、保存に関するポリシーを策定するのを支援するツールにはOpenDOAR³⁸ポリシーツールがある。このツールを利用統計に関するポリシー案を提供するように拡張することができる。OpenDOARは学術オープンアクセスリポジトリの信頼できるディレクトリである。将来的には、リポジトリの様々な利用統計ポリシーの概要も提供できるかもしれない。そうなれば、リポジトリが利用統計を公開しているか否か、その粒度やフォーマットはどういうものかを、ユーザは簡単に調べることができることになる。さらに、出版社のオープンアクセスポリシーを公開しているSHERPA/RoMEO³⁹サービスを拡張して、出版社からどんな統計データがどんな条件で公開されているかを示す情報を表示することもできるかもしれない。

連絡先

windisch@sub.goettingen.de

Christine.Merk@uni-konstanz.de

謝辞

ワークショップに参加されたすべての方々にはその貴重な議論と貢献に対して感謝したいと思います。その準備をしてくださったの方々にも感謝したいと思います。また、追加情報をいただいた Jan-Erik Frantsvag、Richard Geyde、Bjorn Mittelsdorf、Paul Needham、Birgit Schmidt、Maurice Vanderfeesten のの方々にも感謝したいと思います。また、まず第 1 に本プロジェクトを可能にした JISC にも感謝したいと思います。特に、JISC の担当プログラムマネージャーである Neil Jacobs には感謝いたします。

³⁸ <http://www.opendoar.org/>

³⁹ <http://www.sherpa.ac.uk/romeo.php>

9 参考文献

ANSI/NISO Z39.93-2007 (2007). *The Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol*. <http://www.niso.org/workrooms/sushi>

ANSI/NISO Z39.88-2004 (2004). *The OpenURL framework for context-sensitive services*. http://www.niso.org/standards/resources/Z39_88_2004.pdf

Benchmark Statistics Project (BEST) (2007). *Guiding Principles: Version 4*. Australian Partnership for Sustainable Repositories (APSR). http://www.apsr.edu.au/best/best_gp_4.pdf

Bollen, Johan, Rodriguez, Marko A. & Herbert Van de Sompel (2007). *MESUR: usage-based metrics of scholarly impact*. Prepared for the JCDL '07. http://www.mesur.org/Documentation_files/JCDL07_bollen.pdf

Bollen, Johan & Herbert Van de Sompel (2006). An architecture for the aggregation and analysis of scholarly usage data. *JCDL '06: Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, Chapel Hill, NC, USA. 298-307.

COUNTER - Counting Online Usage of NeTworked Electronic Resources (2008). *The COUNTER code of practice: Journals and Databases . Release 3*. <http://www.projectcounter.org/r3/Release3D9.pdf>

Eindhoven, Kwame van & Maurits van der Graaf (2007). *Inventory study into the present type and level of OAI compliant Digital Repository activities in the EU*. <http://www.driversupport.eu/documents/DRIVER%20Inventory%20study%202007.pdf>

IVW (2008). *Anlage 1 zu den IVW-Richtlinien für Online-Angebote: Definitionen und technische Erläuterungen, Version 2.1*. http://daten.ivw.eu/download/pdf/Online_RichtlinienV2_1_Anlage1.pdf

Knowledge Exchange (2007). *Institutional repositories workshop strand report, strand title: Usage statistics*. <http://www.knowledge-exchange.info/Default.aspx?ID=164>

Shepherd, Peter T. (2007). *Final report on the investigation into the feasibility of developing and implementing journal usage factors*. <http://www.uksg.org/usagefactors/final>

10 付録

議事日程

2008年7月7日(11:00-17:30)

- 11:00-11:15 開催の挨拶
- 11:15-12:00 **利用データ**: DINI DFG-JISC プロジェクトの観点から見たワークショップの目的
(司会: Frank Scholze, シュトゥットガルト大学図書館; Nils K. Windisch, ゲッ
チンゲン州立・大学図書館)
- 12:00-12:15 MESUR プロジェクト (Johan Bollen, ロスアラモス国立研究所; LANL)
- 12:15-12:30 COUNTER プロジェクトと Sushi (Peter T. Shepherd, COUNTER 代表)
- 12:15-13:15 **利用データ**: 参加者の見解(参加者による発言)
- 13:15-14:00 昼食
- 14:00-15:30 **法律セッション**: 統計データの著作権状況とプライバシーの問題
(司会者: Michael Seadle, IHU ベルリン図書館情報科学研究所; Hannes Obex, ミ
ュンスター大学情報法・通信法・メディア法研究所)
- 15:30-16:00 コーヒーブレイク
- 16:00-17:00 **技術セッション I**: リポジトリ利用統計交換のための「基本フォーマット」
(司会: Frank Scholze, Nils K. Windisch)
- 17:00-17:15 休憩
- 17:15-17:30 まとめ (司会: Frank Scholze)

2008年7月8日(09:00-14:00)

- 09:00-10:30 **技術セッション II**: リポジトリ利用統計交換のための「基本フォーマット」(司会:
Frank Scholze, Nils K. Windisch)
- 10:30-11:00 コーヒーブレイク
- 11:00-12:00 **ポリシーセッション**: 統計に関するポリシー、統計はオープンアクセスにするべき
か?
(司会: Werner Stephan, シュトゥットガルト大学図書館)
- 12:00-12:15 休憩
- 12:15-13:00 まとめ: リポジトリ利用統計交換のための共通「基本フォーマット」に関するすべて
の要件を我々は持っているか?
(司会: Frank Scholze)
- 13:00-14:00 昼食と送別の辞

表 A-1: 参加者と所属一覧

名前	所属機関
Chris Armbruster	マックス・プランクデジタル図書館
Johan Bollen	MESUR
Tim Brody	IRStats
Susanne Dobratz	ベルリン・フンボルト大学コンピュータ・メディア センター
Fred Friend	JISC
Helmut Hartmann	グラーツ大学図書館
Sabine Henneberger	ベルリン・フンボルト大学コンピュータ・メディア センター
Ulrich Herb	ザールランド大学・州立図書館
Sune Karlsson	オレブロ大学ビジネススクール(スウェーデン)
Anja Kersting	コンスタンツ大学図書館
Frank Lutzenkirchen	デュースブルグ=エッセン大学図書館
Ross MacIntyre	MIMAS
Christine Merk	コンスタンツ大学図書館
Bjorn Mittelsdorf	ザールランド大学・州立図書館
Sebastian Mundt	シュトゥットガルトメディア大学
Hannes Obex	ミュンスター大学情報法・通信法・メディア法研 究所
Peter Schirnbacher	ベルリン・フンボルト大学
Frank Scholze	シュトゥットガルト大学図書館
Michael Seadle	ベルリン・フンボルト大学
Peter Shepherd	COUNTER
Jan Steinberg	バーデン・ビュッテンブルグ図書館サービス
Werner Stephan	シュトゥットガルト大学図書館
Adrian Stevenson	JISC
Fred Vos	NEEO
Nils K. Windisch	ゲッチンゲン州立・大学図書館

プロジェクト参加機関および連絡担当一覧

コンスタンツ大学図書館

名前: Anja Kersting, Christine Merk (プロジェクト・マネージャー)

Email: Anja.Kersting@uni-konstanz.de, Christine.Merk@uni-konstanz.de

ゲオルク・アウグスト大学ゲッティンゲン州立・大学図書館

名前: Nils K. Windisch

Email: windisch@sub.uni-goettingen.de

シュトゥットガルト大学図書館

名前: Frank Scholze

Email: scholze@ub.uni-stuttgart.de

ザールランド大学図書館

名前: Ulrich Herb

Email: u.herb@sulb.uni-saarland.de

ベルリン・フンボルト大学コンピュータ・メディアセンター

名前: Sabine Henneberger

Email: shenneberger@cms.hu-berlin.de