

## アウトプット評価のための機関リポジトリガイドライン (第2版)

「機関リポジリアウトプット評価の標準化と高度化」プロジェクト  
(代表機関：千葉大学 / 連携機関：東北大学, 筑波大学)

### \* はじめに

「機関リポジリアウトプット評価の標準化と高度化」プロジェクトは、平成 18-19 年度の「機関リポジトリの評価システム」プロジェクト (当大学：千葉大学, 三重大学), 平成 20-21 年度の「機関リポジトリ評価のための基盤構築」プロジェクト (代表機関：千葉大学 / 連携機関：東北大学, 金沢大学, 北海道大学, 大阪大学) を引き継ぎ、機関リポジトリを単にコンテンツ数の多寡で評価するのではなく、有用なコンテンツはアクセスが多くなりやすいことから、アクセス数で評価することの検討を重ねてきた。

その間、平成 21 年 1 月には「機関リポジリアウトプット評価システム (ROAT: Repository Output Assessment Tool)」を構築し、平成 22 年 2 月には「アウトプット評価のための機関リポジトリガイドライン」を作成した。

さて、近年、大学図書館に関する統計調査に機関リポジトリに関する項目が設けられるようになりつつあり、ダウンロード数と称してアクセス数が調査対象になっているが、その数値は必ずしも集計の基準が統一されているようには見えない。

各機関の機関リポジトリを各機関で正当に評価するためにも、統一された基準で集計することの重要性は明らかであり、「アウトプット評価のための機関リポジトリガイドライン」を改訂し、統一された基準の構築に寄与することを図る。

### 1 カウント対象

機関リポジトリのアクセス数を集計する場合、Web サーバとしての機関リポジトリのアクセス・ログを基礎とすることが一般的と考えられるが、アクセス・ログを基礎とするか否かに関わらず、英国の合同システム情報委員会 (JISC: Joint Information Systems Committee) の利用統計に関するレビュー<sup>\*1</sup>にも記述されているとおり、ユーザ、セッション、アイテムの識別、リクエスト種別、リクエスト日時が必要である。

#### 1.1 ROAT におけるカウント対象の扱い

ROAT では、Apache の Combined 形式のアクセス・ログを使用することを原則とすることにより、上記の JISC のレビューに記載された項目を取得している。ユーザをアクセス元の IP アドレス、アイテムの種別を国立情報学研究所が運営する JAIRO にハーベストされている本文の URL と一致したアクセス・ログに残るアクセス先の URL としている。

## 1.2 同一の著作が複数のファイルになっている場合等の扱い

同一の著作が、例えば HTML と PDF といったように、複数の形式で機関リポジトリに登録されるということや、異なる複数の機関に所属する研究者の共同研究がそれぞれの機関の機関リポジトリに登録されるということがありえるが、著作単位のアクセス数を集計し、コンテンツの利用度を計る場合には、同一の著作の別ファイルへのアクセスを合算が必要である。

一方、ファイル・サイズその他の事情により、1 つの著作が複数のファイルに分割して登録されるということがありえるが、この場合は分割された全てのファイルへアクセスがあった時点で1回のアクセスとするなどの操作が必要である。

## 1.3 同一のファイルに複数の URL が存在する場合の扱い

リポジトリのソフトウェアの中には、同一のファイルに複数の URL が付与されるものが存在することが確認されているが、この場合も、それぞれの URL でアクセスされた回数  
の合算が必要である。

## 1.4 いわゆる共同リポジトリの扱い

1 つの機関リポジトリを複数の機関で使用する共同リポジトリが構築されているが、この場合、特別な構築方法を採用しない限り、個々の機関ごとにアクセス・ログが保存されない。したがって、機関ごとに別々のディレクトリにコンテンツを登録したり、ファイル名に機関を示す記号等を付与しておくなどの措置が必要である。

## 2 カウント方法

機関リポジトリへの主たる登録対象を雑誌等に掲載された論文としていることから、カウント方法については、電子ジャーナルの利用統計で一般的な COUNTER 実務指針 (COUNTER Code of Practise. Release 3) に準拠することが望ましい。

### 2.1 ROAT におけるカウント

ROAT においても COUNTER 実務指針に準拠し、下記のとおり処理を行っている。

- (1) 同一のアクセス対象に対する同一ユーザ (IP アドレス) からのアクセスについて、HTML ファイルは 10 秒以内、PDF ファイルは 30 秒以内である場合、ダブルクリックによって生じた重複アクセスとし、利用統計から除外する。
- (2) HTTP Status Codes が 200 および 304 であるもののみを集計する。
- (3) image, gifs, style sheets 等に対するアクセスは利用統計から除外する。
- (4) 機関リポジトリを含む Web サーバへのアクセスには、検索サイト等が情報を収集するための、いわゆるロボットによるアクセスが含まれるが、ロボットによるアクセスは利用統計から除外する。

### 2.2 ROAT によるカウントと ROAT によらないカウントとの関係

ROATによってカウントするか否かに関わらず、カウントに当たっては2.1の(4)の処理を行うためのロボットのリストは同一でなければならない。

### 3 統計の精緻化等について

1.2に掲げた課題については、DOIやURNのような識別子を利用することによって解決できる可能性がある。現在、ジャパンリンクセンターによる国内文献への識別子付与の計画が進んでおり、各機関においてはメタデータに識別子付与を進めるとともに、識別子の利用によって、アクセス数の合算が自動的に行えるようなシステムを検討する必要がある。

こうした課題の解決に向けては、出版社、機関リポジトリ、分野別リポジトリ等の論文へのアクセス回数の総計を目標としたPIRUS2 (Publisher and Institutional Repository Usage Statistics Project 2) の最終報告書<sup>\*2</sup>の内容を含め、国際的な視野に立って進めなければならないだろう。同報告書においては、主として経済的側面で出版社からの賛同が得られず実現には時間がかかる見通しではあるが、機関リポジトリの利用統計部分についてIRUS (the Institutional Repository Usage Statistics service)として独立させ、JISCが支援することを提案していることが注目される。なお、PIRUS2における機関リポジトリの統計データ収集方法については、別紙を参照されたい。

### 4 その他

現在、ROATではアクセス元のIPアドレスによって同一のユーザであるか否かを判断している。しかし、近年、プライベート・アドレスが多用され、同一の機関から複数の端末を使用した場合も、アクセス先では同じIPと認識されることが多い。このため、当プロジェクトではクッキーを利用して、プライベート・アドレスが使用されることの影響を調査した。この結果、現状においてはクッキーを利用しない場合と比べて有意な差は確認できなかったが、今後さらにアクセス数の多いサンプルを用いて検証を進める必要がある。

\*1 Merk, Christine & Windisch, Nils K. *JISC Usage Statistics Review: Final Report*. JISC, 2008.9, 24 p.

[http://ie-repository.jisc.ac.uk/250/1/Usage\\_Statistics\\_Review\\_Final\\_report.pdf](http://ie-repository.jisc.ac.uk/250/1/Usage_Statistics_Review_Final_report.pdf)

\*2 Shepherd, Peter & Needham, Paul. *Publisher and Institutional Repository usage Statistics: The PIRUS2 Project Final Report*. JISC, 2011.10, 74 p.

[http://www.projectcounter.org/News/Pirus2\\_oct2011.pdf](http://www.projectcounter.org/News/Pirus2_oct2011.pdf)