九州大学学術情報リポジトリ
Kyushu University Institutional Repository

機関リポジトリQIR の利用動向調査 : 機関リポジトリの活用促進のために

伊東, 栄典
九州大学情報基盤研究開発センター

馬場, 謙介
九州大学附属図書館研究開発室

吉松, 直美
九州大学附属図書館e リソースサービス室リポジトリ係

星子, 奈美
九州大学情報システム部情報基盤グループ

https://doi.org/10.15017/18320

出版情報：九州大学附属図書館研究開発室年報. 2009/2010, pp.16-18, 2010-03-31. 九州大学附属図書館
バージョン：
権利関係：

論文

機関リポジトリ QIR の利用動向調査
－機関リポジトリの活用促進のために－

伊東 栄典[†]　馬場 謙介[‡]　吉松 直美[§]　星子 奈美[**]

<抄録>
学術文献の電子化に伴い，電子ジャーナルが普及している．電子ジャーナルは便利かつ有用であるものの，価格高騰が問題となっている．この問題に対しオープンアクセスジャーナルとセルフアーカイブが提唱され，その運動が広がっている．セルフアーカイブのためには，図書館等によって機関リポジトリが運営されている．研究者は所属機関の機関リポジトリに論文データを登録することで，その論文データが機関リポジトリから全世界に公開される．機関リポジトリの文献参照に対する効果は定量的に評価する必要がある．本論文では，九州大学学術情報リポジトリ（QIR）の利用動向データに基づき，機関リポジトリの有効性を検証する．

<キーワード>　機関リポジトリ，オープンアクセス，ログ解析

Quantitative Evaluation of QIR
－Toward Next Generation Institutional Repository－

ITO Eisuke　BABA Kensuke　YOSHIMATSU Naomi　HOSHIKO Nami

1. はじめに

大学図書館における資料費のうち電子ジャーナル購読費の占める割合が高くなり，その経費捻出の努力や，購読する電子ジャーナルの制限を強いられている．この問題に対し，学術研究成果を無料で制約なく公開する「オープンアクセス」の概念に基づく活動が行われつつある[9]．そのひとつとして，大学や研究機関における研究成果を保存・発信する機関リポジトリの運用が開始されている．機関リポジトリで公開される論文数の増加は，間接的に電子ジャーナル購読料の問題の解決に繋がることが期待され[2]，これを目標とする機関リポジトリの整備が早急に行われるべきである．

機関リポジトリの効果的な整備のためには，機関リポジトリの研究活動に与える影響を定量的に調査する必要がある．前述の購読料問題解決への貢献を考えるためには，論文のダウンロード数や参考文献としての被引用数がひとつの目安になると考えられる．特定の論文誌についての調査では，ダウンロード数と被引用数の有意な関係がいくつか報告されている[5, 6]．これに対し，佐藤ら[8] は，いくつかの国内の機関リポジトリについて，ふたつの数の間に相関関係はないと結論づけている．

本研究の目標は，論文のダウンロード数の時間的な推移による，論文の何かしらの分類の可能性を明らかにすることである．論文のダウンロード数と被引用数の関係についての，時間的な因子を含んだ上での分析方法は文献[8] に整理されている．ここでは，「ある時点でのダウンロード数」は考慮されているが，「ダウンロード数の時間的な変化」は考慮されていない．我々は，まず，九州大学の機関リポジトリで収集されたデータについて調査を行い，論文の被引用数との関係を調べる．本論文では，そのための準備として，ダウンロード数の基本的な解析を行い，調査の方針を検討する．

2. 九州大学学術情報リポジトリ（QIR）

九州大学学術情報リポジトリ（Kyushu University Institutional Repository，QIR）[1] は，学内で生産された知的生産物を保存・公開することを目的とした学術情報資源管理システムであり，平成 18 年 4 月より，主に九州大学附属図書館によって管理･運用されている．QIR では，学術論文の他，会議発表委資料等の著作物も蓄積し，原則として無償で配布している．規模の目安として 2010 年 1 月 4 日現在の QIR に登録されているコンテンツ数を表 1 に示す．また，図 1 に文献種別による割合も示す．

学術研究成果等を無料で公開する「オープンアクセス」の活動が世界中に広まりつつあり，多くの大学や研究機関で機関リポジトリの運用が開始されている．

[†] いとう えいすけ　九州大学情報基盤研究開発センター　E-mail: ito.eisuke.523@m.kyushu-u.ac.jp
[‡] ばば けんすけ　九州大学附属図書館研究開発室　E-mail: baba@lib.kyushu-u.ac.jp
[§] よしまつ なおみ　九州大学附属図書館 e リソースサービス室リポジトリ係　E-mail: nyaomi@lib.kyushu-u.ac.jp
[**] ほしこ なみ　九州大学情報システム部情報基盤グループ　E-mail: nhosh@lib.kyushu-u.ac.jp

そして，その効率化のための工夫や，論文引用数等で表される研究活動への影響が研究されている. 我々は，機関リポジトリに蓄積される論文の質の確保を目的とした効果的な査読システムを検討している[4]．このシステムの有効性の検証についても，ダウンロード数や被引用数を用いる予定である[3].

表１　QIR のコンテンツ数（2010 年 3 月 31 日現在）

学術雑誌論文	1, 142
学位論文	95
紀要論文	10, 443
会議発表論文	731
会議発表資料	137
図書	95
テクニカルレポート	395
研究報告書	105
一般雑誌記事	164
プレプリント	146
教材	32
その他	463
合計	13, 948

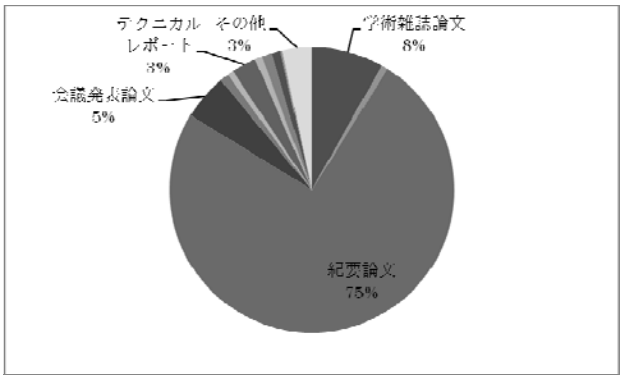

図１　QIR コンテンツの文献種別割合

3. 有効性の検証

QIR のログおよび，QIR に格納されている文献についての被引用数などのデータから，機関リポジトリの有効性を検証する.

3.1. アクセスログ解析

QIR のアクセスログから，QIR が保持する文献のダウンロード数を調査した. 期間は 2008 年 6 月 1 日から 2009 年 12 月 31 日までである．解析対象は PDF 形式のものだけとした．これは，QIR の保持する論文は，ほとんど PDF 形式であるためである.

QIR の保持する各文献は，それぞれ一意な ID が割り振られている．ID は，「2324/2970」のように，2 つの数字をスラッシュ「/」でつないだものである．この ID は，文献のハンドルとして「http://hdl.handle.net/2324/2970」の形式で使われている.

QIR のアクセスログは Apache Web サーバの形式になっている．たとえば，次の様な形式をしている.

66.249.68.172 - - [22/Dec/2009:03:17:08 +0900] "GET/dspace/handle/2324/15958 HTTP/1.1" 200 12992 "-" "Mozilla/5.0"

ダウンロード数調査のため, アクセスログで,「GET メソッドのアクセス」かつ「リソースの path が pdf か PDF で終わっているもの」かつ,「リソース指定の文字列が 2324/2970 のような文字列を含むもの」という条件で数え上げた．2008 年 6 月から 2009 年 12 月末までのアクセスログから，13,214 件の文献 ID を抽出した. QIR の全登録数は 2009 年 12 月 10 日現在で, 13,656 件であるため 97%の文献がアクセスログに出現している.

3.2. 文献ダウンロード数の解析

上位 10 位までのダウンロード数を表 2 に示す. 表 2 のダウンロード数は，検索エンジンのためのデータ収集ロボットのアクセス数も数えている．各文献のダウンロード数を順位順でプロットした．その結果を図 1 に示す．縦軸はダウンロード数を，横軸は順位（ダウンロード数で並べた場合の）を表している．左上のものが，最もダウンロード数が多いものである.

図 1 を見ると，ダウンロード数は 9,000 位くらいまでは Zipf の法則（Zipf's law または power law）を満たす形になっていることが分かる. 9,000 位以下のダウンロード数が少ない文献部分は，急激にプロットの線が落ち込んでいる. このようなアクセス頻度の図は, Web 上のコンテンツには頻繁に出現する特徴である．論文の IF 値（Impact Factor 値，被引用数）について考察している文献[7] では，IF 値が図 1 と同様の傾向を持つことを示しており，この傾向は Lavalette の法則（Lavalette's law）であると示している.

表２　上位１０位までのダウンロード数

順位	文献ハンドル	ダウンロード数
1	2324/8838	4295
2	2324/54	4395
3	2324/1327	3731
4	2324/12542	3762
5	2324/15004	3230

6	2324/9836	5048
7	2324/9082	2607
8	2324/10325	3215
9	2324/277	2167
10	2324/14208	2039

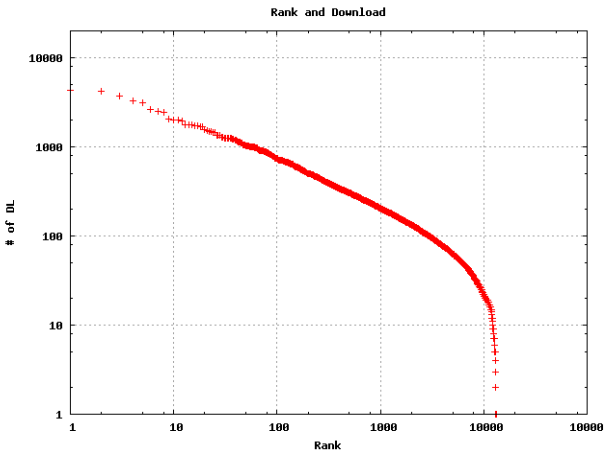

図２　順位およびダウンロード数

3.3. 被引用数とダウンロード数の相関

前節で述べたように，IR に格納された論文のダウンロード数は，論文における被引用数（IF 値）と傾向が似ている．IR 上の文献は Web 上に公開されているため，内容の良い文献は良く参照されるはずであり，そうであればダウンロード数も多くなると考えられる．しかしながら，佐藤らは文献[8]の中で，現在のところ，IR による文献の公開と被引用数との間には，明確な相関は発見できないと主張している．日本の IR については，文献データの公開自体が比較的最近始まったものであり，重要論文の掲載数も多いとは言えないため，今後の詳細な検討が必要である．

そこで，ダウンロード数が上位，中位，下位の論文について，被引用数との相関関係を調べる予定である．具体的には，論文のタイトルをキーにして，Google scholar で被引用数を抽出する．問題は，QIR に限らず，大学の IR に掲載されている論文は論文誌等で公開されていないものを含むので，全体的に被引用数が少ないことである．QIR が持つ 13,000 の文献から，何らかの方法で被引用数の多いもの，中程度のものを探し，相関を調査する．

4. おわりに

本論文では，論文のダウンロード数の時間的な推移と被引用数の関係を明らかにすることを目標として，九州大学の機関リポジトリで収集されたデータについて解析を行うための準備となる調査を行った．今後は，ダウンロード数の時間的推移や，ダウンロード数による順位を考慮して，被引用数との関係を調べる予定である．

参考文献

[1] 九州大学学術情報リポジトリ，https://qir.kyushuu.ac.jp/dspace/.

[2] 文部科学省，“大学図書館の整備及び学術情報流通の在り方について（審議のまとめ）”，http://www.mext.go.jp/bmenu/shingi/gijyutu/gijyutu4/toushin/attach/1283023.htm, August, 2009.

[3] 馬場謙介，伊東栄典，吉松直美，星子奈美，“機関リポジトリの有効性分析”, Proc. of DEIM2010 (第 2 回データ工学と情報マネジメントに関するフォーラム)，F7-3, February, 2010.

[4] Baba, K., Ito, E., Yoshimatsu, N., Hoshiko, N., and Murakami, K., A Model of Publication of Scholarly Papers on Institutional Repositories, DRF International Conference 2009 Conference Proceedings, December, 2009.

[5] Mansillaa, R., Köppena, E., Cochob, G., and Miramontes, P., “On the behavior of journal impact factor rank-order distribution”, Journal of Informetrics, vol.1, issue 2, pp.155-160, 2007.

[6] O'Leary, D. E., “The relationship between citations and number of downloads in Decision Support Systems”, Decision Support Systems, vol.45, no.4, pp.972-980, 2008.

[7] Popescu, I., “On a Zipf's Law Extension to Impact Factors”, Glottometrics 6, pp.83-93, 2003.

[8] 佐藤翔，冨本壽子，逸村裕，“論文の被引用数と機関リポジトリにおけるダウンロード数の関係”，図書館情報メディア研究, vol.7, no.1, pp.53-65, 2009.

[9] 土屋俊，“学術情報流通の動向”，現代の図書館，vol.42, no.1, pp.3-30, 2004.