ZS Project: Zoological Science Meets Institutional Repositories

Sho Sato
Graduate School of Library, Information and Media Studies, University of Tsukuba
Yuko Nagai
The Zoological Society of Japan
Takashi Koga
Research and Development Laboratory, Kyoto University Library
Shigeki Sugita
Library & Academic Information Section, Otaru University of Commerce Library
Mika Saito
Electronic Resources section, University of Tsukuba Library and
Hiroshi Itsumura
Graduate School of Library, Information and Media Studies, University of Tsukuba

Abstract

To evaluate how the deposition of journal articles in Institutional Repositories (IRs) affects the number of citations and e-journal usage, we placed some articles published in *Zoological Science* in two IRs, and compared their use in IRs with e-journals, as well as with the number of resulting citations between 2008 and 2009. The results reveal that deposit in IRs did not reduce e-journal usage. Moreover, whereas the journals gained new readers, this did not have an effect on the number of citations.

Keywords

Open Access, Institutional Repositories, usage analysis, citation analysis, scholarly publishing, zoological science

1. Introduction

The objective of this study was to determine how article deposition in Institutional Repositories (IRs) affects both citations and e-journal usage. It has been said that open access will make research output more visible and accelerate research progress. For example, significant attention has been paid to the question of whether open access will increase the number of citations, including studies by Lawrence (2001) and Harnad and Brody (2004). Some of these studies have focused on open access papers deposited in Subject Repositories (SRs) such as arXiv (Moed, 2007; Henneken et. al., 2006), and others have used a comparative approach by making e-journals partially available to the public (Davis, 2008; Davis, 2009). Moreover, whereas IRs play an important role in open access, their effect on citations has not been sufficiently studied.

It has also been suggested that making journal articles public through the web affects the numbers of readers of the commercial publishers. In this regard, Davis and Fromerth (2007) revealed that deposit in arXiv reduced publisher downloads for mathematics articles. There has been no sufficient analysis as to whether IRs would reduce publisher e-journal usage, however. It may be posited that IR users differ from SR users, and consequently the deposit of an article in IRs will have a different effect on e-journal usage than when deposited in SRs.

To evaluate how the deposition of journal articles in IRs affects citations and e-journal usage, we initiated the «*Zoological Science* meets Institutional Repositories Project» (ZS Project). We placed articles published in *Zoological Science* (the peer-reviewed journal of the Zoological Society of Japan) in two IRs to study their usage in each IR, the number of resulting citations, and any differences in usage when comparing them with e-journals offering publisher versions. The project period was between 2008 and 2010, and we describe interim results in this paper.

2. Data and Methods

This study is based on article-level data of *Zoological Science,* which has been published electronically on BioOne.2 (http://www.bioone.org/loi/jzoo) since 2008 and on J-Stage (http://www.jstage.jst.go.jp/browse/zsj/) until 2009. We secured cooperation from the authors of several articles and deposited those articles in two IRs, the Hokkaido University Collection of Scholarly and Academic Papers (HUSCAP) and the Kyoto University Research Information Repository (KURENAI), in 2008.

The four variables analyzed in this study were the number of ISI citations, the number of BioOne.2 full-text downloads, the number of J-Stage full-text downloads, and the number of IR full-text downloads. While the IRs and J-stage

usage logs have been made available since 2008, the statistics available on BioOne.2 began in 2009. Access by programs, including search engine bots, continuous accesses, and double clicks, were removed from the download statistics based on the COUNTER code of practice (2008).

Using these data, we analyzed the IR usage statistics, including access path and user domains, the conditions of full-text download from both IRs and publisher websites, and the relationships between the number of full-text downloads and citations.

3. Results

From a total of 3,281 articles published in *Zoological Science* between 1984 and 2008, 1,718 were in BioOne.2, 1,376 in J-Stage, and 171 in the IRs.

3.1 IR usage statistics

The total number of IR full-text downloads between 2008 and 2009 was 10,099 (average=59.0/median=43). Table 1 shows the access path for articles deposited in IRs.

Access path	2008 (N = 171)		2009 (N = 171)		Total (N = 171)	
	downloads	%	downloads	%	downloads	%
Direct access	1,112	22.0%	648	12.8%	1,760	17.4%
Through other IR pages	449	8.9%	563	11.1%	1,012	10.0%
Through search engines	3,361	66.6%	3,656	72.4%	7,017	69.5%
Through other pages	122	2.4%	184	3.6%	306	3.0%
Total	5,044		5,051		10,095	

Table 1. Access path to articles deposited in IRs

The majority of users (66.6% – 72.4%) found articles using search engines, especially Google. Only a few people found articles through other paths, including IR interfaces.

Table 2 shows the types of user groups defined by domain: .ne and .net (private users and network operators); .ac and .edu (academic users); .co and .com (corporate users); and others. We identified user domains by IP address (unfortunately some of these could not be resolved).

User groups	2008 (N = 171)		2009 (N = 171)		Total (N = 171)	
	downloads	%	downloads	%	downloads	%
Private users (.ne or .net)	1,121	34.7%	1,200	38.7%	2,321	36.6%
Academic users (.ac or .edu)	558	17.3%	512	16.5%	1,070	16.9%
Corporate users (.co or .com)	572	17.7%	462	14.9%	1,034	16.3%
Others	983	30.4%	928	29.9%	1,911	30.2%
Total	3,234		3,102		6,336	

Table 2. Types of user groups for articles deposited in IRs

The largest user group was private users (34.7%-38.7%). The private and corporate users together accounted for more than 50% of all repository users. While *Zoological Science* is a refereed scientific journal, it indicates that researchers and students as well as the general public read the articles that are free on the web.

Table 3 shows the number of full-text downloads by Japanese (.jp domain users) and foreigners (other domain users).

Domains	2008 (N = 171)		2009 (N = 171)		Total (N = 171)	
	downloads	%	downloads	%	downloads	%
Japanese (.jp)	440	13.6%	553	17.8%	993	15.7%
Foreigners (not .jp)	2,794	86.4%	2,549	82.2%	5,343	84.3%
Total	3,234		3,102		6,336	

Table 3. IRs full-text downloads by Japanese and foreigners

Although the IRs HUSCAP and KURENAI are managed by Japanese universities, and *Zoological Science* is published by a Japanese society, the majority of users were from outside Japan (82.2%-86.4%). This may be because most of the articles in *Zoological Science* are written in English. Sato et. al.'s 2009 analysis of the usage patterns in four Japanese IRs revealed that most of the users for English-language papers were from outside Japan. IR users from Japan tend not to read papers written in English.

On the other hand, many of the users of publisher websites and authors who cited *Zoological Science* were Japanese. Table 4 summarizes the top 10 countries that downloaded full-text from IRs, from J-stage, and that cited *Zoological Science* between 2008 and 2009. We calculated the number of J-stage users and ISI citations by using articles deposited in IRs (N=171). User domains are not available from BioOne.2 (BioOne's Privacy Policy does not allow us to share user domains).

Rank	IR downloads		J-stage downloads		ISI citations	
	Country	Downloads	Country	Downloads	Country	Citing authors
1	USA	2,388	Japan	2,052	Japan	383
2	Japan	993	China	998	USA	178
3	Germany	183	USA	851	Canada	56
4	India	183	EU	576	China	52
5	Russia	157	Indonesia	386	England	37
6	Indonesia	153	India	287	Germany	34
7	Italy	151	Thailand	190	Italy	31
8	Australia	150	Korea	184	France	29
9	England	147	Taiwan	182	Taiwan	27
10	Brazil	117	Italy	178	Argentina	23

Table 4. Top 10 countries that used IRs, J-stage and cited *Zoological Science*

The majority of J-stage users and authors who cited *Zoological Science* were from Japan. However, in IRs most of the users came from the USA. As for user location, there were different tendencies between countries depending on the frequency of access to IRs or to J-stage, and the number of citations.

3.2 Relationships between IR deposit and publisher downloads

Table 5 provides descriptive statistics comparing J-stage full-text downloads that were deposited in IRs with those that were not between 2008 and 2009.

	deposited in IRs (N=126)	not deposited in IRs (N=1,192)
Average	57.9	51.6
Median	50.5	36.0
Maximum	184	935
Minimum	8	0

Table 5. Descriptive statistics of J-stage full-text downloads

The average number of full-text downloads of articles deposited in IRs was higher than that of those not deposited. A Mann-Whitney test shows that the difference between downloads of those articles deposited and not deposited is statistically significant ($p<0.001$). Accordingly, IR registration did not reduce the number of J-stage users.

Table 6 illustrates BioOne.2 full-text downloads of articles deposited in IRs and those not deposited in 2009. Here, the average number of full-text downloads of articles not included in IRs was higher than that of those that were deposited. However, a Mann-Whitney test shows that the difference between articles deposited and not deposited in IRs is statistically insignificant ($p>0.1$). It has been demonstrated that articles published in recent years are used more often than older articles (Tenopir and King, 2009). There were many recent articles (published in 2008) included in the articles not deposited in IRs because we started this experiment in 2008 where *Zoological Science* had a one-year embargo. This inflated the average number of full-text downloads of articles not deposited in IRs. Table 7 presents descriptive statistics of BioOne.2 full-text downloads, not including the articles published in 2008.

	deposited in IRs (N=135)	not deposited in IRs (N=1,629)
Average	52.4	62.1
Median	35.0	43.0
Maximum	287	840
Minimum	3	0

Table 6. Descriptive statistics of BioOne.2 full-text downloads

	deposited in IRs (N=135)	not deposited in IRs (N=1,476)
Average	52.4	54.6
Median	35.0	39.0
Maximum	287	793
Minimum	3	0

Table 7. Descriptive statistics of BioOne.2 full-text downloads (not including articles published in 2008)

Compared with Table 6, there is some difference in use between the articles deposited and not deposited in IRs in Table 7. Thus we can say that IR registration did not reduce the use of BioOne.2 articles.

Figure 1 showing the scatter plots of full-text downloads further supports our assumption.

There were significant positive correlations between IR full-text downloads and J-stage (Spearman's $\rho = 0.363$) and BioOne.2 ($\rho = 0.519$). However, the correlation coefficients were not very high. The scatter plots also tell us that, whereas some articles were downloaded many times in IRs, this was not the case on publisher websites. These analyses therefore indicate that IR registration does not reduce the number of users on publisher websites. Instead,

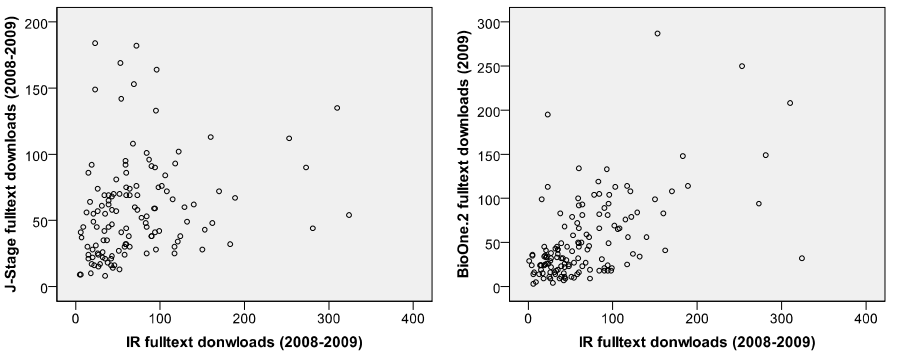

Figure 1. Scatter plots of IR full-text downloads and J-stage / BioOne.2 downloads

deposition in IRs tends to attract new users, some of whom read articles that are not read by academic readers.

Relationships between full-text downloads and citations

Table 8 shows descriptive statistics for ISI citations of articles deposited in IRs compared with those that were not between 2008 and 2009 (*after* deposition in IRs). Table 9 presents another set of descriptive statistics showing the data between 1984 and 2007 (*before* deposition in IRs).

	deposited in IRs (N=171)	not deposited in IRs (N=3,144)
Average	2.2	1.2
Median	2.0	0.0
Maximum	16	24
Minimum	0	0

Table 8. Descriptive statistics of ISI citations between 2008 and 2009

	deposited in IRs (N=171)	not deposited in IRs (N=3,144)
Average	7.3	6.1
Median	5.0	3.0
Maximum	43	193
Minimum	0	0

Table 9. Descriptive statistics of ISI citations between 1984 and 2007

The average and median number of citations of articles in IRs *after deposit* was higher than those of not deposited. A Mann-Whitney test shows that the difference is statistically significant ($p<0.001$). However, it cannot be determined whether the difference was caused by IR registration. From Table 9, the

average and median number of article citations in IRs *before deposit* were also higher than those of those not deposited. This difference is likewise statistically significant ($p<0.001$), and thus suggests that the difference is not due to IR registration, but rather to quality-bias.

In addition, Figure 2 shows the relationship between the numbers of IR full-text downloads and ISI citations between 2008 and 2009. Although there was a significant positive correlation between IR full-text downloads and ISI citations, it was weak (Spearman's $\rho = 0.334$). The scatter plot shows no relationship between IR full-text downloads and ISI citations. We could not determine whether IR registration affects the number of ISI citations.

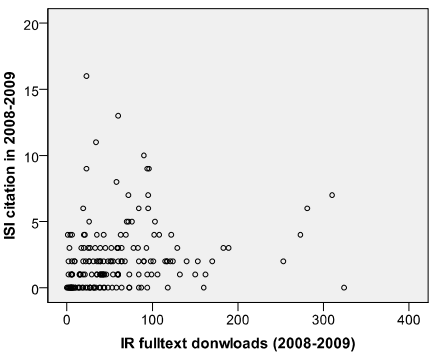

Figure 2. Scatter plot of IR full-text downloads and ISI citations

4. Conclusions

This project revealed that (1) articles deposited in IRs were used by various people including non-researchers who accessed the information through search engines; (2) there are different patterns of use between countries depending on the frequency of downloads from IRs or from e-journals, and the number of citations; and, (3) IR registration did not reduce the number of e-journal users and there were distinctive differences between papers often read in IRs and those in e-journals. Whereas we could not determine whether the presence of an article in an IR affects the number of citations, the results show that depositing journal articles in IRs enhances distribution rather than replaces traditional publication. Depositing journal articles in IRs after a one-year embargo will do more good than harm to scholarly journal publishers.

References

Counting Online Usage of Networked Electronic Resources, (2008). *The COUNTER Code of Practice: Journals and Databases, Release 3*. Retrieved 22 May 2010 from http://www.projectcounter.org/r3/Release3D9.pdf.

Davis, P. M. and Fromerth, M. J., (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles?. *Scientometrics*, Vol. 71, No. 2, 203-215.

Davis, P. M. et. al., (2008). Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ*, Vol.337, a568.

Davis, P. M., (2009). Author-choice open-access publishing in the biological and medical literature: A citation analysis. *Journal of the American Society for Information Science and Technology*, Vol.60, No.1, 3-8.

Harnad, S. and Brody, T., (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, Vol.10, No.6.

Henneken, E. A. et. al., (2006). Effect of e-printing on citation rates in astronomy and physics. *The Journal of Electronic Publishing*, Vol. 9, No. 2.

Lawrence, S., (2001). Free online availability substantially increases a paper's impact. *Nature*, Vol.411, No.6837, 521.

Moed, H. F., (2007). The effect of "open access" on citation impact: an analysis of ArXiv's condensed matter section. J*ournal of the American Society for Information Science and Technology*, Vol.58, No.13, 2047-2054.

Sato, S. et. al., (2009): Usage log analysis of the contents of institutional repositories: user domains, types of referrals and content attributes. Proceedings of the *DRF International Conference 2009:* P-8, 108.

Tenopir, C. and King, D., (2009). Variations in article seeking and reading patterns of academics: What makes a difference? *Library & Information Science Research*, Vol. 31, 139-148.

AUTHOR'S COPY | AUTORENEXEMPLAR