論　文

Usage Log Analysis of Articles in Five Japanese Institutional Repositories: the Relationships between Users, Access Paths, and Accessed Articles.

Sho Sato, Minako Nishiura, Yuko Nagai, Hiroshi Itsumura

Abstract

Background

As Institutional Repositories (IRs) have become major, the opportunity to gain access to academic articles has increased not only for researchers, but for literally everyone. However, it has not yet been clarified what type of people tend to find these academic articles and using which access paths, and which academic fields are the most popular. In this study, we analyzed the usage logs of IRs to explore the relationships between users, access paths, and accessed articles.

Methodology/Principal Findings

For our research sample we used the access logs of five Japanese IRs for the year 2008. We analyzed the user domains, the academic fields of the articles and the access paths to them. The results of this analysis found that there were many non-researchers amongst the IR users, and 40.8-58.1% of accesses were from users' homes or other private sites. These private users accessed articles in applied sciences more frequently than those in pure sciences. By contrast, academic users such as researchers accessed articles equally across all academic fields. Since the main access path to the articles in IRs was from search engines, non-machine editable articles were less frequently accessed than machine-editable ones.

Conclusions/Significance

This study found that articles deposited in IRs were accessed not only by researchers but also by the general public. Although it has been pointed out in some studies that the general public is able to easily access academic articles, this study is the first to reveal that such articles are actually used quite often by the general public. However, non machine-editable articles were not commonly used because the main access path to IRs is through search engines, which cannot index non-machine editable text files. To increase the accessibility and to realize truly open access, it is important to have article files in machine-editable text format.

Introduction

The purpose of this study is to identify who uses the articles deposited in Institutional Repositories (IRs), what kind of articles tend to be used, and how those articles are discovered by users.

An IR is a system or a service which collects, preserves, and disseminates the intellectual output of academic institutions by using the Internet. According to Open DOAR (http://www.opendoar.org/), there were 1,642 IRs worldwide as of July 2011. In Japan, IR statistics are provided by the NII Institutional Repositories Database Contents Analysis System (http://irdb.nii.ac.jp/analysis/index e.php). As of July 2011, 181 Japanese institutions have established their own IRs and more than 810,000 articles have been published in them. The use of IR articles has been receiving increasing attention in-line with the increasing the number of IRs.

Researchers and students are most likely to use the articles registered in IRs. This is because most of the papers in IRs are academic articles, which were written mainly for the authors' colleagues and students studying their disciplines. Open access achieved by IRs, or by other ways, plays a role in removing barriers to scholarly communication, particularly in developing countries, where inaccessibility to scholarly articles due mainly to financial restraints is a crucial problem.(1)

However IRs are not only used by researchers and students. The U.S. National Institutes of Health's (NIH) Public Access Policy on the results of research that it funds requires scientists to submit their final peer-reviewed journal manuscripts to PubMed Central (http://www.ncbi.nlm.nih.gov/pmc/) upon acceptance for publication. NIH states that the reason for this is because not only the researchers and students but also "Clinicians, patients, educators, and students can better reap the benefits of papers arising from NIH funding by accessing them on PubMed Central at no charge", and "the Policy allows NIH to monitor, mine, and develop its portfolio of taxpayer funded research more effectively" (http://publicaccess.nih.gov/FAQ.htm). A Japanese government report released in 2009 also suggests that accountability can be ensured if people get access to research results made possible through use of taxpayers' money.(2) In focus group interviews (FGIs) conducted in the Netherlands, Zuccala (2010) found that ordinary citizens (people who are not researchers) also see open access literature as a useful resource for their personal decision-making.(3) In addition, the U.K. Joint Information Systems Committee (JISC) stated that open access would help small and medium sized enterprises (SMEs) to access research papers which could be useful for their work.(4)

As can be seen from the above, the users of articles registered in IRs include not only researchers and students, but also the general public and enterprises. Zuccala (2010) pointed out that FGI participants expressed their interest in health sciences and psychology, business and economics, and earth and environmental sciences, but not in chemistry, mathematics and statistics.(3)

Thinking about ways of accessing materials, there are three major types of access path to IRs. The first type is browsing and searching within IRs themselves. The second is through external databases connected to IRs using the OAI-PMH metadata harvesting protocol. For example, in Japan, the CiNii (http://ci.nii.ac.jp/en), the Scholarly and Academic Information Navigator managed by the National Institute of Informatics, has been connected to Japanese IRs. The third type is through search engines such as Google.

Literature Review

The number of studies using usage log analysis to explore the behavior of academic article users has increased along with the growth of e-journals. David Nicholas et al. at the University College London's Centre for Information Behavior and the Evaluation of Research (CIBER) have been conducting user behavior studies using log analysis methods. They found that there were significant institutional and subject differences in information seeking behavior, and when the contents were opened to Google, a large amount of traffic came from Google searches.(5-8) 2005 usage statistics from HighWire also indicated that Google provided the majority (56.4%) of the referrals from search engines.(9)

The use of articles in IRs has been receiving increasing attention in recent years. In particular, the standardization of usage statistics is one of the most noteworthy ongoing studies. The Publisher and Institutional Repository Usage Statistics (PIRUS) project was funded by JISC to develop COUNTER-compliant standards and usage reports at the individual article level that can be implemented by any entity (publisher, aggregator, repository, etc.) hosting online journal articles.(10) The Interoperable Repository Statistics (IRS) project designed an application programming interface (API) for gathering IR usage data and implemented it for data providers using common IR software (http://irs.eprints.org/about.html). In Japan, Chiba University has initiated the "Standardization and upgrading of institutional repository output assessment" project. The project team released the Repository Output Assessment Tool (ROAT), which automatically collects usage logs from the IRs of project member universities and processes them according to a standardized method (http://roat.l.chiba-u.ac.jp/xoops/). The Open Access Statistics (OAS) project in Germany also collects usage logs from several universities and produces standardized statistics.(11)

There have been some studies using usage logs to analyze user behavior in IRs. In terms of access paths, Organ (2006) analyzed usage statistics for the IR at the University of Wollongong, Australia over the six-month period from January to June 2006. He found that 95.8% of the measurable full-text downloads were directed via Google.(12) Ikeda & Inoue (2009) analyzed the usage logs of Kyushu University IR from July 01, 2008 to January 31, 2009. They showed that more than half of total accesses came from search engines such as Google, and only few accesses from harvesters such as

OAIster. In addition, they found that other services related to academic articles, such as CiNii, provide strong and stable accesses to the IR.(13) Robinson (2009) analyzed the usage of the IR in the Hong Kong Institute of Education (HKIEd) from April to May 2009 using Google Analytics. In both months, over 75% of the accesses were via search engines, with between 11% and 14% via referral sites and 10% or less from direct access to the IR.(14) Connell (2011) analyzed the usage data of contents deposited in the IR of Ohio State University from January 01, 2007 to March 31, 2009 and examined the hypothesis that there would be some differences in the frequency of use among different types of content. She found that text content was more used than images, and that within the text content, journal articles and theses were the most commonly used.(15) Bonilla-Carelo (2008) analyzed statistics and citations of physics articles download between 2006 and 2007 from the IR of the University of Strathclyde. He found that there was a positive correlation between the number of citations and downloads and the number of distinct countries cited and downloaded to. The most frequently cited and downloaded documents were post-prints of journal articles.(16) However, Royster (2007) suggested that materials not previously published elsewhere were used many times from top-downloaded articles in the IR of the University of Nebraska-Lincoln.(17) In relation to user location, Robinson (2009) found that while the majority of accesses to HKIEd were from Hong Kong itself, there were some from Mainland China, the U.S., Taiwan and Australia. In addition, many Chinese articles were accessed via Baidu, a Chinese search engine.(14)

While there have been many case studies, few studies analyzed the usage logs taken from two or more IRs and compared the results. The present study analyzed the usage logs of five Japanese IRs to explore the details of relationships between user groups, access paths and article academic fields. Although limited to Japan, general tendencies in the usage of IRs can be identified by comparing usage logs from these several IRs.

Materials and Methods

Target data

In this study, we analyzed the usage logs of five Japanese IRs: the Academic Research Repository at the Institute of Developing Economies (ARRIDE), the Hokkaido University Collection of Scholarly and Academic Papers (HUSCAP), the Kyoto University Research Information Repository (KURENAI), the University of Tsukuba Repository (Tulips-R) and the Hitotsubashi University Repository (HERMES-IR). The research period was from January to December 2008. Table 1 presents the details of each repository and Table 2 presents the details of the articles in each repository.

We analyzed only accesses to the full-text files of the articles and excluded accesses to the repositories' main pages or metadata pages. In addition, before analyzing usage logs, we excluded accesses considered as noise such as those by web-crawlers and continuous accesses such as double clicks from the same user. To exclude such noise from IR usage logs, Yoshinori Sato (2008)(18) proposed a log filtering method based on the COUNTER code of practice.(19) In the present study, we employed the same method to exclude noise, as shown below.

a) Extraction by HTTP status code

Log data only with successful file accesses (specifically, only those with HTTP status codes 200 or 304) were extracted.

b) Handling of duplicate accesses

Continuous accesses by the same user, such as double clicks, were removed. All continuous accesses from the same IP address within a 30-second time frame were counted as a single access.

c) Cutting off web-crawlers

Accesses by some programs, such as search engine crawlers, were removed. Those programs were identified by using the User Agent and IP addresses data.

d) Removal of internal access data

Accesses from IR administrators were removed. We obtained lists of IP addresses for internal use from the contributing institutions and removed the relevant accesses based on this information.

Table 3 shows the results of excluding the accesses counted as noise in the five IRs. The majority of the accesses to the full-text files of the articles were by web-crawlers. The log data after excluding that noise (19-50% of the total) were analyzed in our study.

Table 1. Details on the five Japanese IRs studied

IR name (abbreviated name)	Academic Research Repository at the Institute of Developing Economies (ARRIDE)	Hokkaido University Collection of Scholarly and Academic Papers (HUSCAP)	Kyoto University Research Information Repository (KURENAI)	University of Tsukuba Repository (Tulips-R)	Hitotsubashi University Repository (HERMES-IR)
Organization name	Institute of Developing Economies, Japan External Trade Organization	Hokkaido University	Kyoto University	University of Tsukuba	Hitotsubashi University
Organization type	Research institute	National university	National university	National university	National university
Release date	Aug 15, 2006	Apr 01, 2006	Oct 02, 2006	Mar 23, 2007	May 01, 2007
Software	DSpace	DSpace	DSpace	DSpace	DSpace
URL	http://ir.ide.go.jp/	http://eprints.lib.hokudai.ac.jp/	http://repository.kulib.kyoto-u.ac.jp/	http://www.tulips.tsukuba.ac.jp/dspace/	http://hermes-ir.lib.hit-u.ac.jp/rs/

Table 2. Details on the articles in the five IRs

	ARRIDE	HUSCAP	KURENAI	Tulips-R	HERMES-IR
Total number of pieces of literature*	640	25,542	28,356	7,899	7,821
Language					
Japanese	398	15,086	16,394	6,922	5,971
English	242	9,871	10,612	923	1,713
Other languages	0	585	1,350	1	128
Type of literature**					
Journal articles	242	2,719	1,413	929	222
Theses and dissertations	0	343	300	6,687	27
Departmental bulletin papers***	0	22,026	20,750	6	6,116
Conference papers	0	132	174	48	7
Presentations (slides or handouts)	0	128	52	5	26
Books or book chapters	8	13	51	26	16
Technical reports	178	0	1	0	1,347
Research papers	1	5	190	133	38
Articles	7	65	2,728	2	13
Preprints	0	0	0	42	0
Learning materials	0	24	4	7	0
Data or datasets	0	0	6	0	0
Software	0	0	2	0	0
Others	204	87	2,685	0	0

*The difference between the total number of pieces of literature and the number for each category is due to missing data.

**The type of literature is based upon junii2 (http://www.nii.ac.jp/irp/archive/system/pdf/type NII.pdf)

***A departmental bulletin paper (Kiyo) is a kind of journal published by universities or research institutions, usually without a referee system.

User domains

To identify the users of IRs, we analyzed user domains based on IP addresses. Within the logs, after excluding the types of noise shown in Table 3, 62-78% of domains were successfully identified (Table 4).

We identified the types of user groups from these domains. The details of each group are shown below.

a) Private users

Accesses from .ne for Second Level Domain (SLD) or .net for Top Level Domain (TLD)

b) Academic users:

Accesses from .ac for SLD or .edu for TLD

c) Corporate users

Accesses from .co for SLD or .com for TLD

d) Others

Accesses from anywhere except for those above

Table 3. Access logs with noise data removed

	ARRIDE		HUSCAP		KURENAI		Tulips-R		HERMES-IR	
	Logs	%	Logs	%	Logs	%	Logs	%	Logs	%
Raw logs to full-text	95,824	-	790,653	-	1,872,902	-	595,811	-	505,702	-
Extraction by HTTP status code	83,209	86.8%	756,254	95.6%	1,845,176	98.5%	591,146	99.2%	487,386	96.4%
Handling of duplicate accesses	77,834	81.2%	684,701	86.6%	1,660,565	88.7%	510,139	85.6%	449,482	88.9%
Cutting-off web-crawlers and internal accesses	18,516	19.3%	397,698	50.3%	600,103	32.0%	142,650	23.9%	183,231	36.2%

Table 4. Results of examining user domains

	ARRIDE		HUSCAP		KURENAI		Tulips-R		HERMES-IR	
	Logs	%	Logs	%	Logs	%	Logs	%	Logs	%
Resolved	11,408	61.6%	293,865	73.9%	452,082	75.3%	109,833	77.0%	143,112	78.1%
Unresolved	7,108	38.4%	103,833	26.1%	138,021	24.7%	32,817	23.0%	40,119	21.9%
Total	18,516	-	397,698	-	600,103	-	142,650	-	183,231	-

User groups and academic fields of the articles

To identify the relationships between user groups and the academic fields of the articles in IRs, we divided the articles into six academic fields, namely pure sciences, applied sciences, medicine, social sciences, humanities and others, based on their authors' departments. We then calculated the means and medians of the number of accesses from each user group in each academic field. In this analysis, we used only journal articles, theses or dissertations and departmental bulletin papers in HUSCAP and KURENAI. This was due to the fact that only these two IRs had metadata on authors' departments. Considering the impact on the research results, only files with machine-editable texts were analyzed. We excluded papers from “Surikaiseki kenkyujo kokyuroku” (Research Institute for Mathematical Sciences Kôkyûroku), which occupies 40% of the files deposited in KURENAI, from this study to avoid confusion in the research results, since it gets considerably fewer accesses than other journals, while it has 5 times as many articles as those from other basic science fields. Hereafter, unless specifically mentioned, all the analysis of the often-used articles excludes “Surikaiseki kenkyujo kokyuroku”.

Access paths

To analyze users' access paths to IRs, we identified four access paths, shown below, based on the referrers.

a) Direct access

Direct access is access with no referrer. It includes the case that a user inputs a URL directly into the address bar.

b) Access through metadata pages

This is access from links on web sites which only contain the metadata of articles. It includes the case in which IR search functions are used, and the case that access is from external databases such as CiNii.

c) Access through search engines

This includes the case where a user clicks a link from a major search engine. If a user clicked a link to a metadata page from a search engine first and then clicked a link to a full-text article from the metadata page, it was treated as b) Access through metadata pages.

d) Others

This includes access from a weblog, Twitter, SNS services and from other web pages. If a user clicked a link to a metadata page from another page first, then clicked a link to a full-text article from the metadata page, it was treated as b) Access through metadata pages.

Accessibility for search engines is related to the issue of whether articles have machine-editable text or not. However, how much of an impact do machine-editable texts have on the number of accesses through search engines remains unknown. Whereas most of the articles analyzed in this study had machine-editable text, there were a number of departmental bulletin papers without them in HUSCAP and KURENAI. Thus, to investigate the impact of machine-editable text on the frequency of accesses, we analyzed the relationship between the number of accesses through each path and departmental bulletin papers in a machine-editable text format in HUSCAP and KURENAI. Again, we excluded "Surikaiseki kenkyujo kokyuroku" (RIMS Kôkyûroku) in KURENAI for the same reason as in the analysis of the relationship between the user groups and often-used articles.

Results

User groups

Table 5 shows the numbers and percentages of accesses from each user group, namely private users, academic users, corporate users and others. In all five IRs, the largest user group was that of private users (40.8%-58.1%). The second largest group was that of academic users (12.8%-24.0%) and the third largest was that of corporate users (9.1%-18.1%). Accesses from the inside the university itself made up 6.5%-20.5% of academic users, and most accesses by academic users were from external academic users.

Table 5. Numbers and percentages of accesses by each user groups

Types of group	Domains	ARRIDE		HUSCAP		KURENAI		Tulips-R		HERMES-IR	
		Accesses	%	Accesses	%	Accesses	%	Accesses	%	Accesses	%
Private users	.ne and .net	4,653	40.8%	127,693	43.5%	227,165	50.2%	48,512	44.2%	83,173	58.1%
Academic users	.ac and .edu	2,027	17.8%	59,764	20.3%	72,576	16.1%	26,407	24.0%	18,363	12.8%
Corporate users	.co and .com	2,062	18.1%	36,723	12.5%	55,195	12.2%	12,796	11.7%	12,965	9.1%
Others		2,666	23.4%	69,685	23.7%	97,146	21.5%	22,118	20.1%	28,611	20.0%
Total		11,408	100%	293,865	100%	452,082	100%	109,833	100%	143,112	100%

User groups and the articles' academic fields

Table 6 shows the relationships between the user groups and the academic fields of accessed articles (the mean and median number of accesses from each group) in HUSCAP and KURENAI. The article types were journal articles, theses or dissertations and departmental bulletin papers. Articles which did not have machine-editable text were excluded from the analysis process. The common tendency seen in HUSCAP and KURENAI is a greater number of accesses by private users to articles in the humanities and fewer accesses to those in the pure sciences. Corporate users most frequently accessed articles in the field of engineering. Academic users tended to access articles in all fields equally.

Table 6a. Relationships between user groups and the academic fields of articles in HUSCAP

	Pure science (N=888)		Engineering (N=1,098)		Medicine (N=523)		Social Sciences (N=891)		Humanities (N=201)		Others (N=91)	
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
Private users	9.5	5.0	17.1	9.0	12.6	8.0	20.1	9.0	20.1	11.0	31.8	25.0
Academic users	7.8	4.0	7.9	5.0	6.4	4.0	8.4	4.0	6.8	4.0	16.7	14.0
Corporate users	3.9	2.0	7.3	4.0	7.0	4.0	3.6	1.0	6.0	1.0	4.7	3.0

Table 6b. Relationships between user groups and the academic fields of articles in KURENAI

	Pure science (N=2,317)		Engineering (N=1,259)		Medicine (N=1,295)		Social Sciences (N=3,967)		Humanities (N=1,359)		Others (N=1,104)	
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
Private users	11.2	6.0	10.7	5.0	26.7	9.0	14.2	7.0	20.0	8.0	11.2	2.0
Academic users	4.2	2.0	4.9	3.0	5.0	2.0	3.8	2.0	4.5	2.0	4.0	0.0
Corporate users	5.4	2.0	4.2	2.0	2.8	1.0	2.8	1.0	2.3	1.0	5.1	1.0

Access paths

Table 7 presents the number of the accesses through each of four access paths, and their proportions. In ARRIDE, most users found the articles through websites included in “Others” (46.7%), most of which were accessed through an economics database, RePEc (Research Papers in Economics). In the other four IRs, most users found articles via search engines (48.7%-79.8%), and almost all these users accessed articles via Google. Only 4.2%-33.0% of all accesses were from metadata pages, implying that most users accessed articles in IRs without seeing any metadata.

Table 7. Numbers and percentages of accesses through each access paths

Types of referrals	ARRIDE		HUSCAP		KURENAI		Tulips-R		HERMES-IR	
	Accesses	%	Accesses	%	Accesses	%	Accesses	%	Accesses	%
Direct access	2,505	17.8%	61,447	15.5%	94,520	15.8%	13,194	14.4%	25,163	13.8%
Through metadata pages	2,477	17.6%	131,036	33.0%	135,909	22.7%	3,842	4.2%	28,917	15.8%
Through search engines	2,506	17.8%	193,010	48.7%	334,188	55.8%	73,263	79.8%	124,184	68.0%
Others	6,560	46.7%	11,196	2.8%	34,674	5.8%	1,520	1.7%	4,434	2.4%
Total	14,048	100%	396,689	100%	599,291	100%	91,819	100%	182,698	100%

The use of access paths varied depending on user groups. Table 8 shows the number of accesses by each user group from the metadata pages and search engines, which are used more often than the other two access paths, as well as the ratio between these two main access paths.

Table 8. Relationships between the user groups and the two main access paths

		Private users (.ne, .net)		Academic users (.ac, .edu)		Corporate users (.co, .com)	
		Accesses	(b) / (a)	Accesses	(b) / (a)	Accesses	(b) / (a)
ARRIDE	(a)	594	1.4	409	0.7	136	2.4
	(b)	819		271		325	
HUSCAP	(a)	44,547	1.4	30,030	0.8	7,898	2.5
	(b)	61,988		22,767		19,954	
KURENAI	(a)	51,208	2.7	27,417	1.4	7,765	4.6
	(b)	138,083		37,243		35,489	
Tulips-R	(a)	1,051	22.6	1,023	12.9	227	29.9
	(b)	23,718		13,210		6,791	
HERMES-IR	(a)	10,506	5.6	4,322	2.8	918	10.4
	(b)	58,369		12,117		9,527	

(a) : Through metadata pages, (b): Through search engines

*The reason why the ratio of (b)/(a) in Tulips-R is considerably higher than for the others is because the total number of accesses from the metadata pages was smaller than it should have been, due to some errors in its logs.

All the IRs had the tendency that academic users accessed articles through metadata pages more than did those in other groups. Apparently, scholars often access materials by using academic databases, many of which are not directly linked to an article but to its metadata. Private users and corporate users are more likely to access articles through search engines. It can be assumed that search engines have become their main tools for finding articles deposited in IRs instead of using academic databases.

Furthermore, the number of accesses varies depending on whether there are machine-editable texts available through search engines. Regarding the departmental bulletin papers in HUSCAP and KURENAI, both of which have data related to their machine-editable texts, Table 9 shows the correlation between papers with or without the machine-editable text and the number of accesses through metadata pages and search engines. The mean and median values are shown in Table 10. As seen in Table 9, there were more accesses of machine-editable articles through search engines than through metadata pages, whereas the non-machine editable articles were accessed less often through search engines than through metadata pages. Table 10 shows that the mean number of accesses to machine-editable articles through search engines was about 16 times higher than that to non-machine editable articles, and the medians of accesses to machine-editable articles through search engines were about 10 times higher those to the non-machine editable articles both in HUSCAP and KURENAI. However, there was a 4-5 fold difference in both the mean and median values of the number of accesses to the machine-editable files and to the non-machine editable files through the metadata pages in HUSCAP. In KURENAI, the mean and median of the number of accesses to machine-editable articles through metadata pages were lower than for the non-machine editable articles.

Table 9. Correlation between articles with/without machine-editable text and the number of accesses

		Machine editable		Non-machine editable	
		Accesses	(b) / (a)	Accesses	(b) / (a)
HUSCAP	(a)	25,180	2.0	69,621	0.6
	(b)	49,674		39,462	
KURENAI	(a)	55,151	4.6	2,233	0.3
	(b)	256,225		560	

(a) : Through metadata pages; (b): Through search engines

Table 10. Mean and median of accesses to articles with/without machine-editable text

		Machine editable		Non-machine editable	
		Mean	Med.	Mean.	Med.
HUSCAP	(a)	15.8	8.0	3.4	2.0
	(b)	31.2	10.0	1.9	1.0
KURENAI	(a)	5.7	2.0	6.3	3.0
	(b)	26.4	10.0	1.6	0.0

(a) : Through metadata pages; (b): Through search engines

Discussion

In this study, we analyzed five Japanese IRs' usage logs and tried to identify the relationships between user groups, the academic fields of the accessed articles and access paths. From the user group analysis we found that the majority of Japanese IR users were private users (.ne or .net domain users). The private and corporate users together accounted for more than 50% of all IR users. Although we have to make allowance for the possibilities of faculty members and students accessing the articles from their own homes or accessing them from the research divisions of corporations, it can be assumed that IR users include many people who do not belong to the traditional academic world.

According to Zuccala's (2010) research on citizens' perceptions of open access, the general public was interested in the research produced in "health sciences and psychology" and "business and economics". They were not so interested in pure sciences such as "mathematics and statistics", "chemistry" and "physics and astronomy".(3) The usage statistics in the present study support the validity of Zuccala's results. While academic users tended to access articles in all academic fields almost equally, private users tended not to access articles in the pure sciences.

As seen in many other prior studies on access paths to IRs, our study also has found that the majority of users find articles in IRs mainly by using search engines. Observing each group, accesses by the general public and corporate users are mainly from search engines, whereas academic users used some databases to access IRs. For those who do not use academic databases, such as the general public, search engines are an essential means for discovering articles deposited in IRs. However in these IRs, there are many materials that cannot be found via search engines and have no machine-editable text. Such materials are accessed less often than those with machine-editable text. At the Budapest Open Access Initiative (2002), the open access to the academic materials was defined as "free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself" (http://www.soros.org/openaccess/read.shtml). However, a number of materials deposited in the IRs cannot be crawled for indexing for technical reasons, which has been preventing people from accessing some articles. To increase the accessibility of such articles and to realize true open access, it is important to have machine-editable texts on article files.

Conclusion

The results of this study show that not only academic researchers but also a wider range of users, such as the general public and corporate users, have been able to access materials in IRs via search engines since IRs enabled them to access academic articles freely available online. Since there are more accesses in the field of applied sciences than in pure sciences, it can be pointed out that the general public uses academic materials for their own learning or getting solutions to everyday issues they face. In this way, the implications of these results are that open access has crossed the boundary from academic information distribution and has become useful for society at large. However, there remain obstacles to access, such as certain technical barriers to accessibility from search engines. Therefore, removing such barriers will be essential to enhancing the impact of IRs in the future.

Acknowledgments

The authors would like to thank all involved partner institutions. We would also like to show our gratitude to Prof. Yoshinori Sato, Tohoku Gakuin University for his support and encouragement.

References

(1) Ghosh SB, Das AK (2006) Open access and institutional repositories - a developing country perspective: a case study of India. IFLA Journal 33(3): 229-250.

(2) Ministry of Education, Culture, Sports, Science and Technology Japan (2009) About measures that should take on Grants-in-Aid for Scientific Research currently (summary of a current discussion). Available: http://www.mext.go.jp/component/b menu/shingi/toushin/ icsFiles/afieldfile/2009/08/28/1283490 01.pdf (in Japanese)

(3) Zuccala A (2010) Open access and civic scientific information literacy. Information Research 15(1): paper 426. Available: http://informationr.net/ir/15-1/paper426.html

(4) JISC (2009) Open Access for UK research: JISC's contributions-Summary of achievements. Available: http://www.jisc.ac.uk/publications/documents/openaccesscontributions.aspx

(5) Huntington P, Nicholas D, Jamali HR, Tenopir C (2006) Article decay in the digital environment: an analysis of usage of OhioLINK by date of publication, employing deep log methods. Journal of the American Society for Information Science and Technology 57(13): 1840-1851.

(6) Nicholas D, Huntington P, Jamili HR, Watkinson A (2006) The information seeking behaviour of the users of digital scholarly journals. Information Processing & Management 42: 1345-1365.

(7) Nicholas D, Huntington P, Jamali HR (2008) User diversity: as demonstrated by deep log analysis. The Electronic Library 26(1): 21-38.

(8) Nicholas D, Rowlands I, Huntington P, Jamali HR, Salazar PH (2010) Diversity in the e-journal use and information-seeking behaviour of UK researchers. Journal of Documentation 66(3): 409-433.

(9) Steinbrook R (2006) Searching for the right search: reaching the medical literature.

New England Journal of Medicine 354(1): 4-7.
⑽ PIRUS: Publisher and Institutional Repository Usage Statistics (2009) Developing a global standard to enable the recording, reporting and consolidation of online usage statistics for individual journal articles hosted by institutional repositories, publishers and other entities (Publisher Metadata and Interoperability Project 3) Final Report. Available: http://www.jisc.ac.uk/media/documents/programmes/pals3/pirus finalreport.pdf
⑾ Herb U (2010) Alternative impact measure for open access documents?: an examination how to generate interoperable usage information from distributed open access service. The World Library and Information Congress: 76th IFLA General Conference and Assembly. Available: http://eprints.rclis.org/14920/
⑿ Organ M (2006) Download statistics: What do they tell us?. D-lib magazine 12(11). Available: http://www.dlib.org/dlib/november06/organ/11organ.html
⒀ Ikeda D, Inoue S (2009) Access flows to a repository from other services. 4th International Conference on Open Repositories. Available: http://hdl.handle.net/1853/28422
⒁ Robinson M (2009) Promoting the visibility of educational research through an institutional repository. Serials Review 35(3): 133-137.
⒂ Connell TH (2011) The use of institutional repositories: The Ohio State University experience. College and Research Libraries. In press. Available: http://crl.acrl.org/content/early/2010/07/23/crl-134rl
⒃ Bonilla-Calero AI (2008) Scientometric analysis of a sample of physics-related research output held in the institutional repository Strathprints (2000-2005). Library Review 57(9): 700-721.
⒄ Royster P (2008) Publishing original content in an Institutional repository. Serials Review 34(1): 27-30.
⒅ Sato Y (2008) Trends in institutional repository usage statistics. Current Awareness 296: 12-16. Available: http://current.ndl.go.jp/ca1666 (in Japanese)
⒆ Counting Online Usage of Networked Electronic Resources (2008) The COUNTER Code of Practice. Journals and Databases. Release 3. Available: http://www.projectcounter.org/r3/Release3D9.pdf

（さとう　しょう。
にしうら　みなこ。
ながい　ゆうこ。
いつむら　ひろし。
2015年２月10日受理）