

# Usage log analysis of articles in six Japanese institutional repositories: Which region do users access articles from?

Sho Sato<sup>1</sup> and Mitsuo Yoshida<sup>2</sup>

<sup>1</sup>Graduate School of Library, Information and Media Studies, University of Tsukuba

<sup>2</sup>Graduate School of Systems and Information Engineering, University of Tsukuba

**Abstract:** To identify from which country or region users accessed articles, and what their access paths were, we analyzed six Japanese Institutional Repositories' (IRs) usage logs. The results found that articles written in English were used by non-Japanese people, mainly from Asia, Europe and North America. The largest user group was in Asia, and they tended to use STM articles more than other groups. IRs play a significant role in scholarly communication in Asia, especially in STM fields.

**Keywords:** Institutional Repositories, Open Access, Usage Log Analysis

## Introduction

The objective of this study is to explore the current state of international dissemination of articles deposited in Japanese institutional repositories (IRs). IRs have been expected to play a significant role in open access in terms of "the world-wide electronic distribution of peer-reviewed journal literature" (Budapest Open Access Initiative, 2002) as self-archiving storage sites. However, despite this, there has been little research carried out looking into whether articles deposited in IRs are used globally. Robinson (2009) analyzed the use of the Hong Kong Institute of Education's repository, and found that it was used not only from within Hong Kong itself, but also from Mainland China, the United States, Taiwan, Australia and a number of other countries. Sato et al. (2009) analyzed four Japanese IRs' user domains and found that articles written in Japanese were used mainly by Japanese users, whereas articles written in English were used mainly by non-Japanese users. However, Sato et al. did not analyze the details of usage by non-Japanese users.

To identify from which country or region users accessed articles and what their access paths were, we analyzed six Japanese IRs' usage logs. Based on the results thereof, we discuss the significance of IRs in international scholarly communication.

## Data and Methods

In this study, we analyzed the usage logs of six Japanese IRs: the Academic Research Repository at the Institute of Developing Economies<sup>1</sup> (ARRIDE/ the number of articles was 688), the Community Repository of Fukui<sup>2</sup> (CRFukui/ 2,171), the Hokkaido University collection of Scholarly and Academic Papers<sup>3</sup> (HUSCAP/ 29,992), Kyoto University Research Information Repository<sup>4</sup> (KURENAI/ 47,625), Kyushu University Institutional Repository<sup>5</sup> (QIR/ 12,166), and the University of Tsukuba Repository<sup>6</sup> (Tulips-R/ 8,919). The research period was from January to December 2009.

Access by programs, including search engine bots, continuous accesses, and double clicks, were removed from logs, and we calculated the number of full-text downloads based on the COUNTER code of practice (2008).

We determined users' geographical information from IP addresses using the free database *GeoLite country*<sup>7</sup>. The metadata records of the object articles were harvested using OAI-PMH.

---

<sup>1</sup> <https://ir.ide.go.jp/> (Retrieved 28 September 2010)

<sup>2</sup> <http://crf.flib.u-fukui.ac.jp/> (Retrieved 28 September 2010)

<sup>3</sup> <http://eprints.lib.hokudai.ac.jp/dspace/> (Retrieved 28 September 2010)

<sup>4</sup> <http://repository.kulib.kyoto-u.ac.jp/dspace/> (Retrieved 28 September 2010)

<sup>5</sup> <http://qir.kyushu-u.ac.jp/dspace/> (Retrieved 28 September 2010)

<sup>6</sup> <http://www.tulips.tsukuba.ac.jp/dspace/> (Retrieved 28 September 2010)

<sup>7</sup> [http://www.maxmind.com/app/geoip\\_country](http://www.maxmind.com/app/geoip_country) (Retrieved 28 September 2010)

## Results

### Countries and regions

Table 1 is a listing of the top ten countries of origin for visits to each repository. The highest number users were from Japan, but this was because the majority of articles were written in Japanese. Articles written in English were accessed by more non-Japanese than Japanese users. Excluding Japan, the highest number of accesses was from the United States, China and India. In addition, Germany, the United Kingdom, France, Republic of Korea, and Taiwan featured in the locations from which access was most common.

Table 1 Users' source access countries for each repository

	ARRIDE ( <i>N</i> * = 688)			CRFukui ( <i>N</i> = 2,171)			HUSCAP ( <i>N</i> = 29,992)		
Rank	Country	Traffic**	%	Country	Traffic	%	Country	Traffic	%
1	Japan	10,061	35.5%	Japan	19,283	58.1%	Japan	387,151	71.8%
2	United States	2,839	10.0%	United States	3,046	9.2%	United States	25,604	4.7%
3	China	2,304	8.1%	India	1,091	3.3%	China	15,201	2.8%
4	India	1,537	5.4%	China	1,008	3.0%	India	9,785	1.8%
5	United Kingdom	1,036	3.7%	Ukraine	493	1.5%	Germany	8,630	1.6%
6	Indonesia	807	2.8%	United Kingdom	486	1.5%	United Kingdom	6,932	1.3%
7	France	721	2.5%	Germany	454	1.4%	Taiwan	6,917	1.3%
8	Malaysia	689	2.4%	Iran	424	1.3%	Korea, Republic of	5,594	1.0%
9	Thailand	537	1.9%	Saudi Arabia	415	1.2%	France	5,323	1.0%
10	Philippines	474	1.7%	Canada	382	1.2%	Canada	4,053	0.8%
	Others	7,328	25.9%	Others	6,127	18.4%	Others	64,379	11.9%
	Total	28,333		Total	33,209		Total	539,569	
	Unknown	648		Unknown	4		Unknown	17	
	KURENAI ( <i>N</i> = 47,625)			QIR ( <i>N</i> = 12,166)			Tulips-R ( <i>N</i> = 8,919)		
1	Japan	505,723	61.7%	Japan	373,230	76.0%	Japan	187,647	81.3%
2	China	54,029	6.6%	United States	18,146	3.7%	United States	8,947	3.9%
3	United States	50,203	6.1%	China	12,169	2.5%	China	8,438	3.7%
4	India	21,013	2.6%	India	10,368	2.1%	India	2,857	1.2%
5	United Kingdom	11,501	1.4%	Germany	5,690	1.2%	Korea, Republic of	1,646	0.7%
6	Germany	10,955	1.3%	United Kingdom	4,912	1.0%	United Kingdom	1,551	0.7%
7	Indonesia	10,071	1.2%	Korea, Republic of	4,302	0.9%	Taiwan	1,433	0.6%
8	France	9,788	1.2%	France	4,175	0.8%	Germany	1,388	0.6%
9	Korea, Republic of	9,171	1.1%	Iran,	3,919	0.8%	France	1,239	0.5%
10	Taiwan	8,681	1.1%	Taiwan	3,513	0.7%	Iran	1,140	0.5%
	Others	129,138	15.7%	Others	50,809	10.3%	Others	14,397	6.2%
	Total	820,273		Total	491,233		Total	230,683	
	Unknown	16,365		Unknown	234		Unknown	1	

\* The number of articles deposited in each repository. \*\* The number of full-text downloads.

Table 2 shows the number and percentage of users grouped into seven regions. Excluding Japan, most users were from Asia for all repositories. In Tulips-R and CRFukui the second largest visitor group was from North America, and in others it was from Europe. Users from South America, Africa and Oceania were notably fewer than those from other regions.

Table 2 Users' source access regions for each repository

	ARRIDE		CRFukui		HUSCAP		KURENAI		QIR		Tulips-R	
	Traffic	%	Traffic	%	Traffic	%	Traffic	%	Traffic	%	Traffic	%
Japan	10,061	35.5%	19,283	58.1%	387,151	71.8%	505,723	61.7%	373,230	76.0%	187,647	81.3%
Asia (without Japan)	9,382	33.1%	5,591	16.8%	62,063	11.5%	158,504	19.3%	54,873	11.2%	21,305	9.2%
Europe	3,818	13.5%	3,175	9.6%	40,815	7.6%	66,012	8.0%	28,422	5.8%	8,583	3.7%
North America	3,282	11.6%	3,550	10.7%	30,586	5.7%	58,720	7.2%	21,968	4.5%	10,054	4.4%
South America	356	1.3%	540	1.6%	5,950	1.1%	7,651	0.9%	3,324	0.7%	770	0.3%
Africa	934	3.3%	676	2.0%	5,387	1.0%	15,715	1.9%	5,766	1.2%	1,543	0.7%
Oceania	466	1.6%	355	1.1%	3,108	0.6%	6,179	0.8%	2,801	0.6%	642	0.3%
Unknown	34	0.1%	39	0.1%	4,509	0.8%	1,769	0.2%	849	0.2%	140	0.1%

### Access path

Table 3 represents the average and median number of non-Japanese users' accesses through four access paths: direct access (accesses without referral), through metadata pages, through search engines and through other websites.

In ARRIDE, most users found articles through other websites, especially with RePEc (Research Papers in Economics), which is an economics database. In other IRs most users found articles using search engines, especially Google. There were no regional differences in access paths.

Table 3 Non-Japanese users access path to IRs

		Direct access	Through metadata page	Through search engines	Through other websites
ARRIDE	Avg.	9.3	3.9	9.0	36.8
	Med.	5.0	1.0	4.0	23.5
CRFukui	Avg.	5.1	0.8	44.6	0.5
	Med.	2.0	0.0	15.5	0.0
HUSCAP	Avg.	1.1	1.6	5.5	0.2
	Med.	0.0	1.0	1.0	0.0
KURENAI	Avg.	1.8	1.6	8.8	0.3
	Med.	0.0	0.0	1.0	0.0
QIR	Avg.	3.4	1.6	22.7	0.4
	Med.	1.0	1.0	11.0	0.0
Tulips	Avg.	2.1	0.4	10.8	5.5
	Med.	1.0	0.0	4.0	2.0

### Discipline

Table 4 shows the relationship between articles' fields and the number of accesses to articles written in English in HUSCAP and KURENAI (only these two repositories had metadata on articles' fields in object repositories).

In KURENAI, Arts & Humanities articles attracted many users in Asia because the majority of the articles (549 in 728) were published in the journal *Southeast Asian Studies*. Excluding those, users from Asia tended to use articles in STM fields more so than did users in Europe or North America.

Table 4 The relationship between articles' fields and the number of accesses

HUSCAP		Non-Japanese Total	Asia	Europe	North America
Arts & Humanities (N = 388)	Avg.	16.6	4.3	5.2	5.4
	Med.	3.0	0.0	1.0	1.0
Social Sciences (N = 654)	Avg.	13.1	4.5	3.7	3.0
	Med.	3.0	1.0	1.0	1.0
Science, Technology and Medicine (N = 10,806)	Avg.	8.2	3.2	2.1	1.7
	Med.	3.0	1.0	1.0	0.0
Others (N = 61)	Avg.	14.7	4.2	4.1	3.6
	Med.	6.0	2.0	2.0	1.0
KURENAI		Non-Japanese Total	Asia	Europe	North America
Arts & Humanities (N = 728)	Avg.	41.0	24.4	7.9	5.7
	Med.	18.0	8.0	3.0	3.0
Social Sciences (N = 690)	Avg.	35.1	7.8	9.6	7.6
	Med.	16.5	3.0	4.5	3.0
Science, Technology and Medicine (N = 12,161)	Avg.	9.0	4.4	2.1	1.5
	Med.	3.0	1.0	1.0	0.0
Others (N = 39)	Avg.	46.5	22.7	12.2	8.2
	Med.	21.0	9.0	9.0	5.0

\* The number of articles written in English in each fields.

### Conclusion

In this article, we analyzed users' locations and tendencies of accessing articles deposited in six Japanese IRs. The results found that articles written in English were used by non-Japanese users, mainly by those from Asia, Europe and North America. The largest user group was from Asia and they mainly found articles in IRs by using search engines. In addition, users from Asia tended to use STM articles more than others. It may be said that IRs play a significant role in scholarly

communication in Asia, especially in STM fields.

### **Acknowledgements**

This work was supported by a MEXT Grant-in-Aid for Scientific Research. (No. 30026348). 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 and Prof. Hiroshi Itsumura, University of Tsukuba for their support and encouragement.

### **References**

- Budapest Open Access Initiative. (2002). Retrieved 12 September, 2010 from <http://www.soros.org/openaccess/read.shtml>
- Counting Online Usage of Networked Electronic Resources. (2008). *The COUNTER Code of Practice. Journals and Databases. Release 3*. Retrieved 12 September, 2010 from <http://www.projectcounter.org/r3/Release3D9.pdf>
- Robinson, M. (2009). Promoting the visibility of educational research through an institutional repository. *Serials Review*, 35(3), 133-137.
- Sato, S. et al. (2009). *Usage log analysis of the contents of institutional repositories: user domains, types of referrals and content attributes*. Poster presented at the DRF International Conference 2009, Tokyo, Japan. Retrieved 12 September, 2010 from <http://hdl.handle.net/2433/88082>