

## Measuring the Impact of Use of External Databases at Institutional Repositories for Promotion of Open Access

Chifumi Nishioka (Kyoto University Library, Japan)

Jun Maeda (Hokkaido University Library, Japan)

Masashi Kawai (National Institute of Informatics, Japan)

### Abstract

In this paper, we investigate the impact of system linkages between deposit systems of institutional repositories (IRs) and external databases including current research information systems (CRIS). Specifically, we compare the state of Open Access (OA) at institutions whose deposit systems fetch information of unregistered publications from external databases and those whose deposit systems do not use any information from external databases. The result of a quantitative analysis confirms that the system linkage promotes deposits of publications to IRs with a significant difference. On the other hand, most of publications on IRs are also freely available on other places (e.g., publisher sites) as well. Furthermore, questionnaires and interviews to institutions show that the primary factor for increasing the number of publications on IRs is direct e-mails, asking for digital copies of publications to be registered. To increase the number of publications that are freely available only on IRs, we should directly send e-mails for publications, which are not OA in anywhere, with priority.

### Keywords

Open Access; institutional repository; current research information systems; scholarly communication

### 1. Introduction

Open Access (OA) to scholarly articles refers to the removal of barriers including price barriers and permission barriers (i.e., copyright and licensing restrictions) from accessing them (Suber, 2004). The movement of OA has been evolved in these decades and a lot of institutions and funders have adopted OA policies. Following the movement, institutions in Japan have adopted OA policies, in which researchers are mandated to deposit a digital copy of a publication at their institutional repository (IR).

To promote OA, several IRs link their deposit systems to external databases including current research information systems (CRIS) to facilitate researchers to deposit publications. The deposit systems extract information of publications that have not been deposited yet from external databases such as Scopus and researchmap<sup>12</sup>. Figure 1 presents an example of system linkage between a deposit system and external databases. The green rectangle provides a list of a user's unregistered publications, so that he/she can see publication that should be uploaded to IRs.

Here in this paper, we aim to reveal the impact of external databases on IRs' content acquisition. It has remained quantitatively unclear if the collaboration with external databases will enhance the content coverage. To this end, we take an analysis, which enables us to further clarify the major challenges, driving factors, and possible solutions for

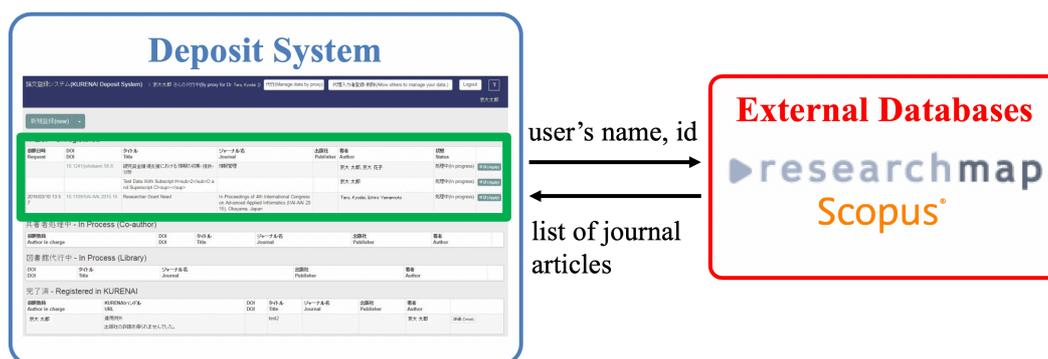
---

<sup>1</sup> researchmap is a database of researchers in Japan, in which they register their publications and other research outcomes.

<sup>2</sup> <https://researchmap.jp/?lang=en> (last accessed on 02/06/2020)

the OA promotion by IRs through a system linkage with CRIS. Specifically, we compare the state of OA at institutions whose deposit systems fetch information of unregistered publications from external databases and those whose deposit systems do not use external databases. The result of a quantitative analysis confirms that the system linkage promotes deposits of publications to IRs with a significant difference. However, most of those publications are also freely available on other places such as publisher sites, as well.

This paper is organized as follows. In the subsequent section, we review related works. Thereafter, Section 3 describes methodologies of the analysis to investigate the impact of external databases on IRs' content acquisition. Section 4 presents the result of the analysis, before concluding the paper.



**Figure 1:** An example of a deposit system in which unregistered publications are extracted from external databases.

## 2. Related Works

So far, a lot of institutions have implemented a system linkage between CRIS and IR for multiple purposes such as improving publicity for research outcomes and increasing the number of publications in IRs (Jeffery and Asserson, 2009; Nixon, 2010; Wenaas et al., 2012). Ribeiro et al. (2016) conduct a survey to collect information on CRIS-IR interoperability. The survey reveals that two systems are clearly complementary and they do not think that CRISs gradually replace IRs. 65% of the institutions implement a system linkage between CRIS and IR. A survey by Serrano-Vicente et al. (2018) shows that 32% of Spanish institutions have implemented a system linkage. Regarding impact of use of information from external databases on IRs' content acquisition, de Castro et al. (2014) report that the number of publications in the IR has increased resulting from CRIS-IR integration at the Polytechnical University of Catalonia (UPC).

Different from the works mentioned in above, we compare the percentages of OA publications between institutions using external databases and those not using any external databases, in order to quantitatively evaluate the impact of the system linkage. In addition, we analyze the state of OA in each institution with respect to different OA types (e.g., Green and Gold). Thus, the analysis contributes to understand to what extent the system linkage promotes each OA type.

## 3. Methodology

We analyze the state of OA of 178 Japanese institutions, at which at least one publication with a DOI is deposited to their IRs. Among them, three institutional repositories use information from external databases. These three institutions differ a lot in terms of their sizes, locations, and so on. The analysis explores differences regarding the

state of OA between the three institutions and the others.

For the purpose, we first extract the list of publications with a DOI for each of 178 institutions using Scopus Affiliation Retrieval API<sup>3</sup>. Since most of the deposit systems started to use information from external databases in around 2016, the analysis focuses on publications that are published 2016 or later. From each list of publications, we extract publications that are included in Unpaywall dataset<sup>4</sup> (version 2019-11-22)<sup>5</sup>. Unpaywall is an application that links every publication with a Crossref DOI to a URL where the fulltext of the publication can be freely accessible. It gathers data from over 50,000 journals and OA repositories (Piwowar et al., 2019).

The Unpaywall dataset provides metadata and information regarding OA (e.g., OA type, OA location URL) for each publication. In terms of OA type, each publication is labeled either Gold, Hybrid, Bronze, Green, or Closed as described in Table 1. The details regarding how to identify the OA type of a publication is described in (Priem, 2019). Although Unpaywall crawls a lot of OA repositories, we find that some of Japanese institutional repositories are not included. Thus, we add an OA type “Uncaptured Green”, which refers to a publication whose digital copy is available at a Japanese IR, but not identified by Unpaywall.

In addition, we conduct an analysis focusing on repositories. As Green labeled by Unpaywall does not distinguish OA using IRs from OA using other repositories such as preprint servers. Thus, it is difficult to observe to what extent IRs contribute to OA. To this end, we label each publication as described in Table 2. Please note that a publication can be given more than one OA types different from Table 1.

**Table 1: OA types**

OA type	Publications
Gold	published in a full-OA journal
Hybrid	published in a toll-access journal, available on the publisher site, with an OA license
Bronze	published in a toll-access journal, available on the publisher site, without an OA license
Green	published in a toll-access journal and the only fulltext copy available is in an OA repository
Uncaptured Green	published in a toll-access journal and the only fulltext copy available is in an OA repository that is not recognized by Unpaywall
Closed	everything else

**Table 2: OA types for the analysis focusing on repositories**

OA type	Publications
Repository All	available on at least one OA repository
Non-IR	available on at least one OA repository other than Japanese IRs
IR	available on at Japanese IR
Only IR	Available only on Japanese IR

<sup>3</sup> [https://dev.elsevier.com/sc\\_apis.html](https://dev.elsevier.com/sc_apis.html) (last accessed on 02/06/2020)

<sup>4</sup> <https://unpaywall.org/products/snapshot> (last accessed on 02/06/2020)

<sup>5</sup> As the Unpaywall dataset contains all publications assigned a Crossref DOI, it is equivalent to extract publications with a Crossref DOI from each list of publications.

#### 4. Result and Discussion

Table 3 presents mean and standard deviation of the percentages of each OA type for IRs using external databases and IRs not using external databases. We conduct a Mann–Whitney U test to verify significant differences between the two sets, but we do not observe any significant difference at all OA types. Table 4 shows mean and standard deviation of the percentages of each OA type focusing on OA made by repositories. A Mann–Whitney U test reveals a significant difference at IR and Only IR. Thus, the use of external databases promotes and facilitates deposits of publications. Compared to the percentage of IR, the percentage of Only IR is small. It indicates that over 80% of publications deposited to IRs are also available at other places such as publisher sites or other OA repositories. We need to increase the number of publications that are available only in IRs, since this number contributes to availability of OA publications in the entire scholarly communications.

To investigate other factors that promote deposits of publications to IRs, we conduct a combined analysis based on both questionnaires and interviews to 87 institutions in Japan. The results show that the primary factor for higher content acquisition is direct e-mailing to the authors, asking for author manuscript or final version to be registered in IR. It should be noted that OA policy and launch of institutional OA committees have less impact on content acquisition compared to the direct e-mailing. However, most institutions send such emails manually, which becomes a certain amount of labor in IR operation. The system linkage between IRs and CRISs enables to send e-mails automatically, as it identifies publications that are not deposited yet. In addition, content harvesting including originally OA articles has resulted in duplicated contents circulation online. Thus, we propose that initial content control such as online availability check will increase the number of unique content registration, which is only made available through IR. This will lead to Green OA promotion in its essential meaning. Therefore, we should send e-mails for publications, which are not OA in anywhere, with priority.

**Table 3:** Mean and standard deviation of percentages of different OA types.

	IRs using external databases (%)	IRs not using external databases (%)
Gold	33.50 (11.36)	24.54 (10.75)
Hybrid	4.37 (0.81)	5.10 (8.07)
Bronze	6.92 (1.52)	9.06 (8.78)
Green	8.66 (3.70)	8.09 (6.39)
Uncaptured Green	0.51 (0.36)	1.17 (0.33)
Closed	46.19 (7.29)	51.74 (14.31)

**Table 4:** Mean and standard deviation of percentages of different OA types focusing on repositories. \*\* denotes  $p < .01$  and \* denotes  $p < .05$ .

	IRs using external databases (%)	IRs not using external databases (%)
Repository-All	42.25 (6.36)	32.97 (12.33)
Non-IR	37.03 (4.93)	31.36 (12.08)
IR**	9.96 (4.30)	1.94 (4.98)
Only IR*	1.57 (1.11)	0.52 (1.56)

## 5. Conclusion

This paper explores the impact of system linkages between deposit systems and external databases including CRIS. We quantitatively compare the state of OA at institutions whose deposit systems use information from external databases and those whose deposit systems do not use. The result indicates that the system linkage significantly promotes deposits of publications to IRs. However, most of those publications are also freely available on other places such as publisher sites. In addition, questionnaires and interviews to 87 institutions reveal that the primary factor for higher content acquisition of IRs is direct e-mails, asking for author manuscript or final version to be registered. To increase the number of publications that are freely available only on IRs, we should send e-mails for publications, which are not OA in anywhere, with priority.

**Acknowledgement.** This work is supported by JPCOAR (Japan Consortium for Open Access Repositories) and JSPS Grant-in-Aid for Young Scientists (No. 18K13235).

## References

- De Castro, P., Shearer, K., & Summann, F. (2014). The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. *CRIS 2014 Conference*. <http://dx.doi.org/10.1016/j.procs.2014.06.007>
- Jeffery, K., & Asserson, A. Institutional repositories and current research information systems. *New Review of Information Networking*, 14(2), 71-83.
- Nixon, W. (2010). Enrich: Improving integration between an Institutional Repository and a CRIS at the University of Glasgow. *Connecting Science with Society: The Role of Research Information in a Knowledge Based Society*. Toptryk Grafisk ApS, Aalborg, Denmark, pp. 55-64. ISBN 9788773079874
- Piowar, H., Priem, J., & Orr, R. (2019). The Future of OA: A large-scale analysis projecting Open Access publication and readership. *BioRxiv*, 795310.
- Priem, J. (2019). What do the types of oa\_status (green, gold, hybrid, and bronze) mean? *Unpaywall Support Portal*, <https://support.unpaywall.org/support/solutions/articles/44001777288-what-do-the-types-of-oa-status-green-gold-hybrid-and-bronze-mean-> (last accessed on 02/06/2020).
- Ribeiro, L., De Castro, P., & Mennielli, M. (2016). EUNIS-EUROCRIS joint survey on CRIS and IR. <https://hdl.handle.net/10216/82659>
- Serrano-Vicente, R., Melero, R., & Abadal, E. (2018). Evaluation of Spanish institutional repositories based on criteria related to technology, procedures, content, marketing and personnel. *Data Technologies and Applications*. 52(3), pp. 384-404. <https://doi.org/10.1108/DTA-10-2017-0074>
- Suber, P. (2004) Open access overview. <https://legacy.earlham.edu/~peters/fos/overview.htm> (last accessed on 02/04/2020).
- Wenaas, L., Karlström, N., & Vatnan, T. (2012). From a national CRIS along the road to Green Open Access—and back again: Building infrastructure from CRISin to Institutional Repositories in Norway. *CRIS 2012 Conference*. <http://hdl.handle.net/11366/116>