

Simultaneous Registration of Scholarly Papers to a Researcher Database and a Repository

Baba, Kensuke
Research and Development Division, Kyushu University Library

Mori, Masao
Institutional Research Office

<https://hdl.handle.net/2324/25182>

出版情報 : Proc. the International Conference on E-Learning and E-Technologies in Education (ICEEE 2012), pp.200-203, 2012-09. IEEE

バージョン :

権利関係 :



Simultaneous Registration of Scholarly Papers to a Researcher Database and a Repository*

Kensuke Baba[†]

Masao Mori

Abstract

This paper introduces a system that realizes a simultaneous registration of scholarly papers for both a researcher database and a repository. An institutional repository is a method to realize free access to research outputs. However, the number of scholarly papers archived in institutional repositories is extremely small compared to the papers practically produced by researchers. The authors considered that an obstacle to increase the number is the efforts of researchers to register their papers to institutional repositories, and they developed a system that realizes a simultaneous registration of papers to the researcher database and the institutional repository in their university. This paper introduces the main idea of the system.

Keywords: digital library; Open Access; repository; Web database.

1 Introduction

Returning research outputs to society is an obligation of public research institutions. A *repository* is a system for archiving and opening research outputs such as scholarly papers, and a repository for outputs in an institution is called an *institutional repository (IR)*. Giving access to scholarly papers in IRs can be a method to return research outputs to society.

An essential prerequisite to a useful IR may be a sufficient number of archived papers. In most IRs, however, the number of archived papers is estimated to be extremely small compared to the papers practically produced by researchers [8]. The first step to increase the number of papers in IRs is to encourage researchers to register their “buried” papers and prevent burying current papers. A simple approach is to reduce the efforts of researchers for the registration of a paper to an IR. Researchers who belong to an institution often have to repeat the adding of the metadata of their research outputs for research reports and so on. This situation may put researchers off.

We practically faced the problem with the IR [2] of Kyushu University. Our solution lays in the creation of a connection between the IR and the researcher database [1] in the university. In Kyushu University, registration of research activities (include the metadata of research papers) to the researcher database is a mandate for the researchers. By re-using their already registered metadata in the IR, the efforts of researchers are expected to be reduced. Cooperation of a researcher database and a repository is a simple idea to improve the usability of the systems, actually there exist some integrated systems [10, 12]. In our development, the cooperation was conducted to improve the efficiency of the data entry in addition to the usability. Compared with a development from zero, the simple idea is difficult to be implemented especially in the situation that each system is being used heavily and managed by separated divisions in the institution.

On the basis of this idea, we have developed a system to realize a simultaneous registration for both the researcher database and the IR. In this paper, we describe the outline of the developed system and the details of its implementation. The number of the papers (full texts) in the IR is estimated to be at most 30% of the number of the papers (metadata) in the researcher database [6]. Therefore, a dramatic increase is expected if the system works well.

The problem of the number of archived papers in the IR is not a special case [9]. Most universities and research institutions maintain their researcher databases in addition to their IRs. Therefore, our trial is worth as a case study and the main idea can be applied to the IRs of other institutions.

2 Problem

This section describes the current situation of the IR and the researcher database in Kyushu University in order to make clear the problem we tackle.

2.1 QIR

Kyushu University Institutional Repository (QIR) is operated by the library of Kyushu University. Generally, an IR archives the full text of each paper in addi-

* An edited version of this report was published in: *Proc. the International Conference on E-Learning and E-Technologies in Education (ICEEE 2012)*, pp. 200–203, IEEE, Sep, 2012.

[†] Research and Development Division, Kyushu University Library, baba@lib.kyushu-u.ac.jp

tion to its metadata. The total number of the archived papers in QIR is about 20,000 as of July 2012. The number of papers in an IR is an important factor to evaluate it. Actually, Ranking Web of World Repositories [3] is taking account the number of full text files as an element of the ranking. The number of QIR ranks 68th in about 2,000 IRs as of April 2012.

Presently, there exist two ways to register a paper to QIR:

- Create a QIR account, then input the metadata and upload the full text by the registration form,
- Send the metadata and the full text to the QIR manager by e-mail.

In the registration form of QIR, authors have to complete the metadata including title, authors' name, and so on. Usually the efforts to add the metadata of papers have to be repeated when researchers upload the papers on their web site, submit a list of their papers as a report to their institute, and so on, which may be an obstacle to paper registration to an IR.

2.2 DHJS

Kyushu University Academic Staff Educational and Research Activities Database (DHJS) is the researcher database of Kyushu University which has been developed by the institutional research office of the university. The database contains various kinds of data relating to the researchers including the scholarly papers they have produced. DHJS consists of the two subsystems, the data-entry system and the viewer system. The data-entry system supports researchers to register their research activities to DHJS. A user authentication is required for a log-in to the system. The viewer system shows the research activities registered by the data-entry system into DHJS. Fig. 1 is an example of the list of papers shown on the viewer system of DHJS.

In Kyushu University, any researcher has the duty to register his research activities, including the metadata of published scholarly papers, into DHJS. The number of the "metadata" of scholarly papers registered in DHJS is about 100,000 as of July 2012. The ratio of duplicate data (that is, metadata for the same paper) is estimated to be at most 20% [6]. On the other hand, QIR has only 20,000 "full texts" as mentioned in the previous subsection. That is, potentially, there exists a large number of research outputs that are produced in Kyushu University but are not archived in QIR.

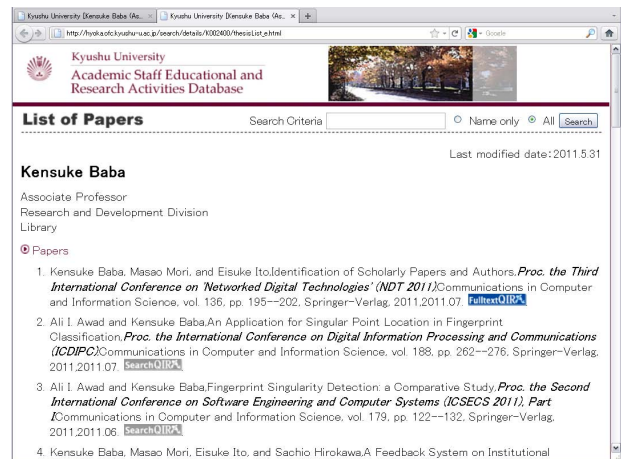


Figure 1: An example of the list of papers in DHJS shown on the viewer system.

3 Link from the Viewer System to QIR

As the first trial, we developed a system that connects QIR with the viewer system of DHJS and evaluated the effect of the system by analyzing the access log.

3.1 Outline of the System

We developed a system that connects QIR with the viewer system of DHJS in order to encourage researchers to register their papers to QIR [5]. By the system, the metadata of papers shown in the viewer system are linked to the corresponding full text in QIR. The link is realized by two kinds of icons (see Fig. 1) in the viewer system that distinguishes whether the paper is archived in QIR or not. For each paper in the list, there can be three situations:

- A dark-colored icon "FulltextQIR" is added,
- A light-colored icon "SearchQIR" is added, or
- There is no icon.

The researcher can decide whether he put the icons to each paper in the list when he register papers to DHJS by the data-entry system. The first case means that the full text of the paper exists in QIR, the second case means that there is no full text (although the researcher wants to register), and the last case means that the researcher does not want to link to the full text. In the second case, the link system returns the result of a search by the author name in QIR. Additionally, if the user is an author of the paper, the user can register the paper to QIR by the registration form of QIR. (A user authentication is required for the registration to QIR.) At the time, the metadata of the

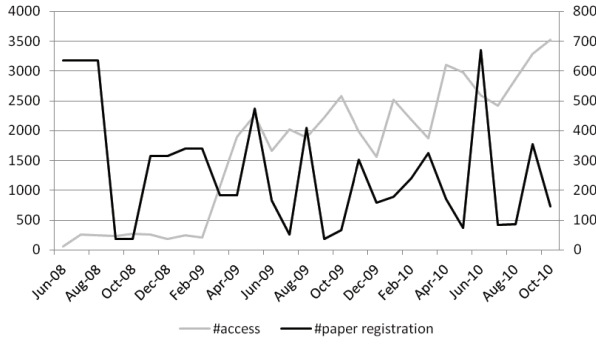


Figure 2: The number of access to QIR from the viewer system of DHJS and the total number of paper registration to QIR.

paper in DHJS is automatically used to complete the registration form, therefore this function can reduce the efforts for registration.

3.2 Evaluation

We evaluated the effect of the link system by analyzing the practical access log of QIR [7]. Fig. 2 shows the number of access to QIR from the viewer system and the total number of paper registration to QIR. We analyzed the access log from June 2008 to October 2010 for each month. The link system is working from April 2009 and was improved in April 2010. In the figure, there is no significant change in the number of paper registration, while there are two points of increases at March 2009 and April 2010 in the number of access. Therefore, we can conclude that the system is not effective to the number of paper registration. The result implies that the amount of the practical efforts for paper registration (such as adding metadata) is not the only factor for an improvement of the number.

4 Paper Registration from the Data-entry System to QIR

We developed a function that realizes a simultaneous paper registration to QIR and DHJS. This section describes the outline and some anticipated difficulties in the implementation.

4.1 Main Idea

The main idea of our solution is that a paper registration to QIR is performed simultaneously with the registration to DHJS at the data-entry system. To be concrete, we developed a function to upload the full text file of a paper in the data-entry system, and other metadata of the paper is re-used from the data

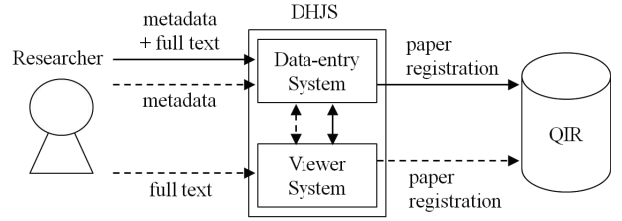


Figure 3: Overview of the system with the developed function of a simultaneous paper registration. The solid lines express the flow of data in the system with the developed function, and the dotted lines for the old version.

submitted for the registration of research activities to DHJS. Fig. 3 is the overview of the system with the developed function.

By the result of the link system in Subsection 3.2, we considered that the trigger of paper registration to an IR is an important factor for researchers in addition to the amount of the efforts. Most researchers are not negative about the ideal of IR. However, they don't want to take their time even if the efforts is small. In the link system, after a researcher register his research activity (the metadata of a paper) to DHJS and put the icon in the data-entry system, he has to conduct the registration of the paper (the full text) to QIR in the viewer system on another occasion. As we mentioned, registration of research activities to DHJS is a mandate, while paper registration to QIR is conducted spontaneously. We expected that the requests of paper registration to QIR will increase if it can be conducted simultaneously with registration of research activities to DHJS.

4.2 Implementation

On the basis of the idea in Subsection 4.1, we developed the function for paper registration to QIR in the data-entry system, and we are going to start the implementation of the function. In the rest of this paper, we describe some anticipated difficulties in the practical implementation.

The simultaneous registration of a paper to DHJS and QIR cannot be achieved by a straightforward transfer of the data. For the developed function, the difficulties were the following two points.

One of the difficulties is a check of duplicate papers. Any paper requested for registration to QIR should be confirmed that the same paper have not been registered yet. In DHJS, any paper should be registered for each of the authors because a paper can be a research activity for every authors. On the other hand, duplicate requests (even different versions) of a paper should be unified in QIR. If the papers in DHJS are requested to register to QIR, the number of duplicate

requests will increase drastically. The confirmation is currently conducted by simple string matching of the title and the authors' name. The efficiency of the process can be improved by putting identifiers of the authors on the paper and using a suitable data construction in QIR. Moreover, this formalization in QIR can identify duplicate data by coauthors in DHJS, which realize a link to the full text in QIR from a metadata in DHJS also for a researcher who is not the main author (that is, has no full text of the paper).

The other difficulty is confirmation of the copyright policy of publishers. The object of the copyright is the full text of a paper, therefore this problem is caused in QIR rather than DHJS. Kim [11] reported that one of the obstacles to self-archiving for researchers is concern about the copyright. Some publishers and academic societies that publish scholarly papers have copyright policies concerning whether a paper can be open at an IR. The policy depends on the versions of the paper: pre-print, post-print, and publisher versions. The pre- and post-print versions are held by the author and classified by the situation whether a review has been done or not. The publisher version is the exact version published by the publisher. For this matter, repository managers have to have a contact with the author in some cases [4]. In order to solve the problem, first of all, a good understanding of copyright policy is necessary for researchers. Practically, we formalized and automated partially the confirmation process of the copyright policy [13]. Additionally, since most publishers allow opening the pre-print version, we are going to add a notification to recommend researchers to upload the pre-print version in the data-entry system.

5 Conclusion

The cooperative system between the IR (QIR) and the researcher database (DHJS) in Kyushu University was introduced. We developed the system that links the viewer system of DHJS to QIR and evaluated by analyzing the access log in terms of the number of access and the number of paper registration. As the result, we could not find any significant effect for the number of registration. We also developed a function in the data-entry system of DHJS that realizes a simultaneous paper registration to QIR and DHJS. An evaluation of the developed function is one of our future work.

References

- [1] DHJS: Kyushu University Academic Staff Educational and Research Activities Database. Retrieved August 20, 2012, from <http://hyoka.ofc.kyushu-u.ac.jp/search/index.e.html>.
- [2] QIR: Kyushu University Institutional Repository. Retrieved August 20, 2012, from <https://qir.kyushu-u.ac.jp/dspace/>.
- [3] Ranking Web of World Repositories. Retrieved June 25, 2012, from <http://repositories.webometrics.info/>.
- [4] K. Baba, N. Hoshiko, E. Kudo, N. Yoshimatsu, and E. Ito. Semi-automated paper-registration system for institutional repository. In *Proc. the Third International Conference on Awareness Science and Technology*, pages 431–434. IEEE, 2011.
- [5] K. Baba, M. Mori, and E. Ito. A synergistic system of institutional repository and researcher database. In *Proc. the Second International Conference on Advanced Service Computing*, pages 184–188. IARIA, 2010.
- [6] K. Baba, M. Mori, and E. Ito. Identification of scholarly papers and authors by connecting databases. *Journal of E-Technology*, 2(3):120–125, 2011.
- [7] K. Baba, T. Tanaka, E. Ishita, M. Mori, E. Ito, and S. Hirokawa. Evaluation of link system between repository and researcher database. In *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, volume 7008 of *Lecture Notes in Computer Science*, pages 381–382. Springer, 2011.
- [8] B.-C. Björk, A. Roos, and M. Lauri. Scientific journal publishing: yearly volume and open access availability. *Information Research*, 14(1):391, 2009.
- [9] S. Harnad, L. Carr, A. Swan, A. Sale, and H. Bosc. Open access repositories - maximizing and measuring research impact through university and research-funder open-access self-archiving mandates. *Wissenschaftsmanagement*, 4(4):36–41, 2009.
- [10] K. Ishizaka, M. Iwai, M. Gokan, H. Ohba, and R. Sakaguchi. Development of the Shinshu University online system of general academic resources (SOAR). *Progress in Informatics*, (5):137–151, 2008.
- [11] J. Kim. Faculty self-archiving: Motivations and barriers. *Journal of the American Society for Information Science and Technology*, 61(9):1909–1922, 2010.

- [12] D. T. Palmer. Knowledge exchange: A strategy for open access success at the University of Hong Kong. *Chinese Journal of Library and Information Science*, 4(1):81–94, 2011.
- [13] K. Sakaguchi, S. Nakayama, and K. Baba. Formalization of the paper registration process to institutional repositories. In *Proc. 2012 IIAI International Conference on Advanced Applied Informatics*, 2012.